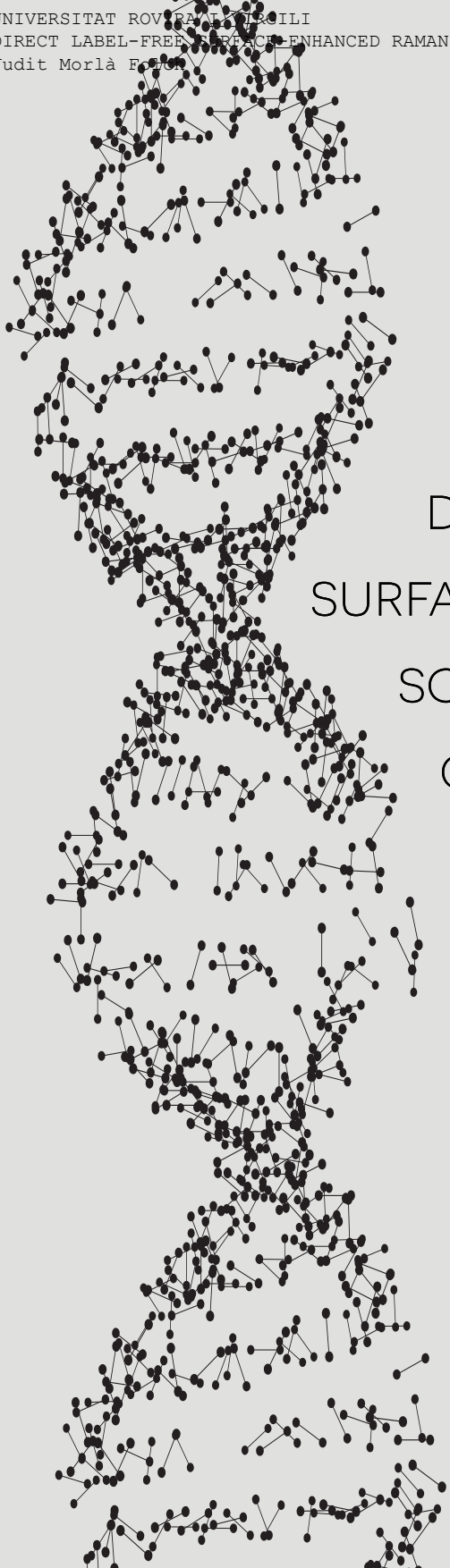
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS

Judit Morlà Folch

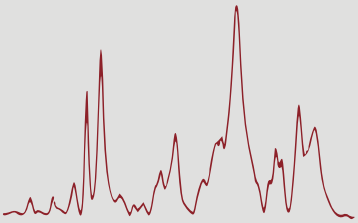
UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch



UNIVERSITAT ROVIRA i VIRGILI

DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS

Judit Morlà Folch



DOCTORAL THESIS

2017

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

Judit Morlà Folch

DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS

DOCTORAL THESIS

International and Industrial Doctorate Mention

Supervised by Dr. Ramón Alvarez Puebla

Dr. Luca Guerrini

Dr. Nicolás Pazos Pérez

FACULTY OF CHEMISTRY

DEPARTMENT OF PHYSICAL AND INORGANIC CHEMISTRY

Tarragona, 2017

UNIVERSITAT ROVIRA i VIRGILI

NANYANG TECHNOLOGICAL UNIVERSITY

medcom advance

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA i VIRGILI

Departament de Química Física i Inorgànica

Prof. Ramón Álvarez Puebla, ICREA Research Professor at Universitat Rovira i Virgili (URV), Dr. Luca Guerrini Postdoctoral Researcher at URV and Dr. Nicolás Pazos Pérez Assistant Professor at URV

CERTIFY

That the study entitled “Direct Label-free Surface-Enhanced Raman Scattering Analysis of Nucleic Acids” presented by Judit Morlà Folch for the award of the degree of Doctor, has been carried out under our supervision at the Department of Physical and Inorganic Chemistry, and that it fulfils all the requirements to be eligible for the International Doctorate Mention.

Tarragona, May 2017

Doctoral Thesis Supervisor	Doctoral Thesis Supervisor	Doctoral Thesis Supervisor
Prof. Ramón A. Álvarez Puebla	Dr. Luca Guerrini	Dr. NicolásPazos Pérez

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

Aquesta tesi s'ha realitzat amb el suport de la Secretaria d'Universitats i

Recerca del Departament d'Empresa i Coneixement de la

Generalitat de Catalunya

Generalitat de Catalunya
Departament d'Empresa i Coneixement
Secretaria d'Universitats i Recerca

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

“ I am among those who think that science has great beauty ”

Marie Curie

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

CONTENTS

LIST OF ABBREVIATIONS (by alphabetical order)

A	Adenine
AgNP	Silver nanoparticles
AgNP@Sp	Silver nanoparticles coated with spermine
CTAB	Cetyltrimethylammonium bromide
C	Cytosine
^{br}C	5-bromocytosine
^{h}C	5-hydroxycytosine
^{hm}C	5-hydroxymethylcytosine
^{m}C	5-methylcytosine
G	Guanine
HOMO	Highest occupied molecular orbital
HPLC	High-Performance Liquid Chromatography
LC-MS	Liquid chromatography-mass spectroscopy
LSPR	Localized Surface Plasmon Resonance
LUMO	Lowest unoccupied molecular orbital
NA	Nucleic Acids
NP	Nanoparticle
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PDMS	Ploydimethylsiloxane
PEI	Polyethylene imine
PLL	Poly-L-Lysine
PLS-DA	Partial Least Squares Analysis
PTFE	Polytetrafluoroethylene
SEM	Scanning Electron microscope

SERS	Surface-Enhanced Raman Spectroscopy
SERRS	Surface-Enhanced Resonance Raman Scattering
SPR	Surface Plasmon Resonance
SSCP	Single-strand conformation polymorphism
T	Thymine
TEM	Transmission Electron Microscope
U	Uracil
UV-Vis	Ultraviolet Visible

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

THESIS SCOPE

The research work described in this thesis has been carried out in the context of surface-enhanced Raman scattering (SERS) spectroscopy for nucleic acids analysis.

SERS is a powerful analytical technique that combines the structural specificity and high experimental flexibility of Raman spectroscopy with the extremely high sensitivity provided by the metal nanostructure amplification of the optical signal. Nowadays, SERS is a mature technology which finds a continuously larger number of applications in diverse fields as analytical chemistry, biomedicine, environmental or materials chemistry.

Characterization of nucleic acids (NAs) has become a major goal in fields such as genetics, medicine or drug discovery. NAs are usually found in low amounts which require extremely sensitive techniques to be detected. Sensitivity represents a key requirement for their implementation into easy-to-use, cost-effective, rapid and high accuracy NA sensors. In this regard, the extraordinary sensitivity, in combination with the rich structural information provided by the Raman spectra, turned SERS into a powerful tool for the analysis of NAs.

SERS analysis of NAs has been largely restricted to indirect approaches, which normally exploit the selective recognition between two complementary strands to identify a specific sequence, and the use of extrinsic Raman labels for

SERS readout. However, these strategies completely dismiss the exquisite structural information provided by the direct acquisition of the biomolecular vibrational fingerprint. Contrarily, direct label-free SERS analysis displays an outstanding potential in terms of chemical-specific information. However, reproducibility issues and/or limited sensitivity have been traditionally limited the successful implementation of direct SERS analysis to NAs. Thus, development of new methods capable of overcoming such limitations would represent an important leap forward for the full exploration of the genetic information by SERS.

On these bases, the *general objective* of the thesis is the implementation of a **novel method for direct SERS analysis of NAs capable of addressing the limitations of previously reported approaches.** This method will rely on the use of positively charged spermine-coated silver nanoparticles as efficient plasmonic substrates for trapping NAs sequences into highly SERS active clusters in suspension via electrostatic interactions. The acquisition of intense and highly reproducible SERS spectra of NAs at an ultrasensitive level will be exploited in multiple relevant applications. In this regard, the following *specific objectives* for this thesis can be identified:

1. **Synthesis and characterization of positively charged spermine-coated silver nanoparticles and their implementation in the SERS of NAs under optimized experimental conditions.**

2. **Detection and relative quantification of chemically-modified cytosine nucleobases in DNA.**
3. **Direct quantification of nucleobase composition in DNA.**
4. **Highly specific structural analysis of different RNAs (conformational classification, identification of nucleobase variants, etc.)**

Accordingly, the thesis is structured in five chapters. **Chapter 1** provides a general introduction with an overview of nucleic acids and SERS. In the first part, not only structure, composition, and lesions on the NAs are discussed, but also the weaknesses of the current techniques to address the desired requirements in terms of sensitivity, reproducibility, or quantification capabilities. In the second part, a brief description of Raman spectroscopy and related issues is presented, which serves as a basis to introduce SERS and its remarkable analytical potential. Different strategies for NAs analysis by SERS are also presented at the ending of Chapter 1.

Cationic silver colloids are the nanostructured substrate used for direct SERS of NAs. **Chapter 2** illustrates the synthesis of positively charged nanoparticles coated with spermine (AgNP@Sp), their optical characterization via several techniques such as UV-Vis spectroscopy or Transmission Electron Microscope and the particular advantages of those colloids for interrogating NAs. Initial studies of direct SERS analyses of NAs with AgNP@Sp are performed to assess the

sensitivity and reproducibility of the method, and adjust the experimental conditions so as to maximize the analytical performance.

On **Chapter 3**, it is described the first application of label-free DNA SERS analysis using the above mentioned AgNP@Sp, which refers to the detection and relative quantification of chemically-modified cytosine nucleobases. Epigenetic modifications play a major role in gene regulation; in fact, they can be considered as information layer beyond the simple genomic sequence. During these last years, an intense research toward the development of novel methods to detect and profile these modifications has been undergone. Specifically, four different cytosine variants are selected among the most relevant chemically-modified nucleobases to be included in both single and double stranded DNA and investigated by SERS. Moreover, the implementation of the SERS method with microfluidics allows for the manipulation and analysis of picogram levels of genetic material.

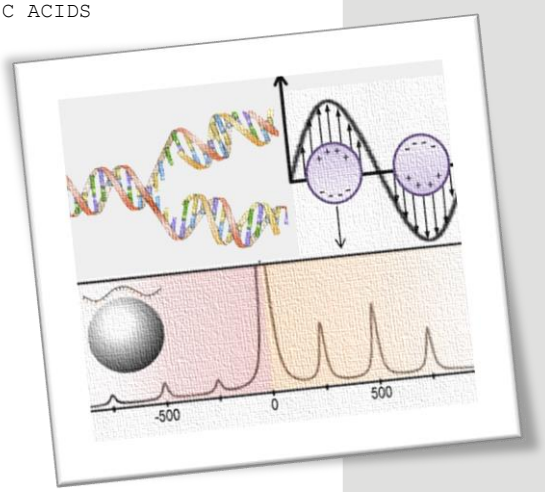
Direct quantification of DNA base composition by SERS is explored in **Chapter 4**. The determination of base composition is a fundamental parameter in genomic analyses, however, conventional methods often suffer from poor sensitivity and they are time-consuming. Herein, the potential of SERS in the quantification of the base composition in short sequences and genomic DNA is investigated in depth.

Additionally, the capacity of label-free direct SERS is not restricted to DNA, highly specific structural information can also be obtained from RNA. Classification of similar RNA sequences with subtle chemical differences or organized into a variety of conformations is explored in **Chapter 5**. The role of RNA in cellular processes is one of the most dynamic fields of biology, particularly, regarding small RNAs. More biological activities carried by small RNAs are continuously emerging and the development of characterization techniques is increasingly needed. Here, direct SERS analyses are performed to screen composition and conformations of RNAs at the ultrasensitive level.

Altogether, the thesis is expected to represent a significant advancement toward label-free NAs SERS analysis and its applicability. The results of this thesis and the impact on the field are then presented in general conclusions.

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch



CHAPTER 1

GENERAL INTRODUCTION

OUTLINE

1.1 GENERAL CONCEPTS OF NUCLEIC ACIDS

Nucleic acids are inarguably the most important biomolecules to all known forms of life. All cells of an organism share the same genetic information, carried by nucleic acids. They are the responsible for the transmission, expression, and conservation of genetic information, ensuring inheritance.[1]

Nucleic acids are divided into two groups; deoxyribonucleic acids (DNA) and ribonucleic acids (RNA). DNA is considered the repository of the biologic information and the responsible for their transmission. The discovery of the DNA structure by Watson and Crick (1953) represented a huge advance in genetics, understanding how information could be passed from one generation to another.[2] The results provided by previous studies on X-ray fiber diffraction patterns[3] and the evidence on base complementarity,[4] made possible the final description of the DNA structure as a double helix with complementary strands (**Figure 1.1**). The same model described the mechanism for the transmission of the genetic information by a semiconservative replication based on two steps (**Figure 1.1**). Firstly, the separation of the two strands, secondly, the synthesis of a complementary strand from each parental one. At 1957, the semiconservative replication was verified by Meselson and Stahl.[5]

At present, the model described 64 years ago, remains true and accommodated with new findings. The fundamental DNA structure was found to be double helix and the differences in the genetic material carried by this sequence rely on the number and order of their building blocks.

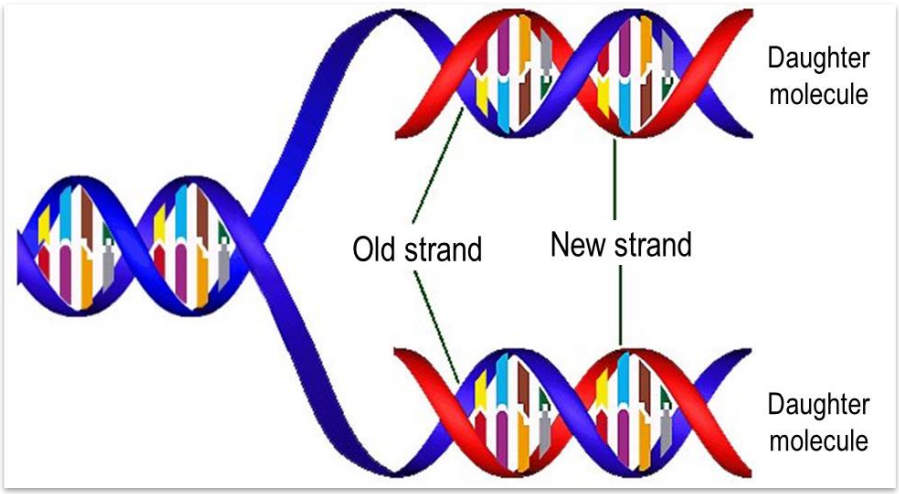


Figure 1.1 The DNA structure during the semiconservative replication. The double helix is separated and a complementary strand for each is synthesized resulting in two DNA molecules.

In nature, however, DNA is not only found as a double helix (dsDNA), but it could be present as well as single strands (ssDNA) (for instance in the chromosomes of small viruses[6]). Moreover, there are DNA sequences that twist the double helix, or some DNA which is linear, whereas others are circular or supercoiled.[7] All that variability shows the wealth of the DNA sequences determining the unique physical and chemical properties necessary as carriers of the hereditary information.[7-8]

The other group of nucleic acids, ribonucleic acid (RNA) appears at first very similar to DNA while differing only in minor aspects of their chemical composition. However, the functions carried by RNA are considered more complex and variable than DNA.[9] Nowadays it is well known the versatility of those molecules being involved in a variety of biological functions[10] and not only working as shuttle

molecules of genetic information supporting the role of DNA as was firstly considered.

In order to fulfil such variety of functions, various types of RNA are found in cells. Specifically, messenger RNA (mRNA) is the carrier of the genetic code from the DNA sequence to the ribosome where the proteins are synthesized. The process in which mRNA is constructed from the DNA sequence is called transcription. Transference RNA (tRNA) plays the role of an adaptor molecule as it translates the genetic code contained on the mRNA to the appropriate sequence of amino acids. Meanwhile, ribosomal RNA (rRNA) is a component of the ribosomes, which is the responsible for driving the proper protein synthesis. This last step in which the proteins are synthesized from an mRNA sequence is called translation. The overall process was suggested by F. Crick in 1970, and it is known as the central dogma of molecular biology, explaining how DNA maintains and transfer the information to encode all our proteins.[11]

In addition, during these last decades even wider and unexpected roles played by RNA have been elucidated. Among others, genetic information can be carried by RNA in the case of viruses or retroviruses,[12] rRNA can develop enzymatic activity catalysing biological reactions[7, 13] and moreover RNA play crucial roles in different pathways related to gene regulation.[10b, 14] Because of all of those characteristics, W. Gilbert hypothesised that early evolution living systems were composed entirely of RNA,[11, 15] a key idea in prebiotic evolution considerations[15] and crucial to understanding the important and diverse roles of

RNA in biological systems. More recently, new forms of RNA were described, especially non-coding RNA; small nuclear and nucleolar RNA (snRNA and snoRNA), micro-RNA (miRNA) and long non-coding RNA (lncRNA). It has been demonstrated that all that non-coding RNA have a key role as regulators of biological processes[16] such as the major involvement of lncRNA and miRNA during cell differentiation and development.[16b, 17] **Table 1.1** summarizes the names and biological functions of the different RNA forms mentioned above.

TYPE OF RNA	ABBREVIATION	FUNCTION
Messenger RNA	mRNA	Contains the gene sequence copy from DNA for the protein synthesis
Transference RNA	tRNA	Transports amino acids to the protein synthesis machinery
Ribosomal RNA	rRNA	Component of the ribosomes (responsible for the protein synthesis)
Small non-coding RNA	sncRNA	Regulates gene expression
Micro-RNA	miRNA	
Short interfering / Silencing RNA	siRNA	
Piwi-interacting RNA	piRNA	
Small nuclear RNA	snRNA	Involved in mRNA processing and intron splicing
Small nucleolar RNA	snoRNA	Involved in processing rRNA molecules and RNA modifications
Long non-coding RNA	lncRNA	Non-protein coding transcripts* longer than 200 nt (unknown function)

Table 1.1 Names, abbreviations, and biological functions of different forms of RNA. (*) In 2001 it was published the first study evidencing that some lncRNA do encode proteins.[18]

NAs composition

The fundamental building blocks of the nucleic acids are the nucleotides. They are formed by three components, *(i)* five-carbon sugar (2'-deoxyribose for DNA or ribose for RNA), *(ii)* nitrogenous base (as well known as nucleobase) and *(iii)* at least one phosphate group (**Figure 1.2A**). When is referred to the nitrogen base and the five-carbon sugar (without the phosphate group) is called nucleoside.[7] In a nucleotide, the base and the sugar are linked by a glycosidic bond, sugar and phosphoric acid make phosphoester bond (**Figure 1.2B**). In addition, nucleotides are linked to each other by phosphodiester linkage forming the characteristic polymeric macromolecules by repeating sugar-phosphate backbone.[8, 19] This backbone structure is a conserved feature of the DNA, while the order and the composition of the nucleobases provide variability allowing the tremendous repository of the genetic information.[8]

As it was mentioned, one component of a nucleotide is the nitrogenous base. Nitrogenous bases consist of heterocyclic rings composed of nitrogen and carbon atoms and they can be divided into two main groups: purines and pyrimidines. The two main purines present in both DNA and RNA are adenine (A) and guanine (G), while the pyrimidine bases cytosine (C) and thymine (T) are the most common in DNA. It is important to note that thymine is an exclusive nucleobase of DNA which is replaced by uracil (U) in RNA. **Figure 1.3** shows the molecular structures of purines and pyrimidines bases.

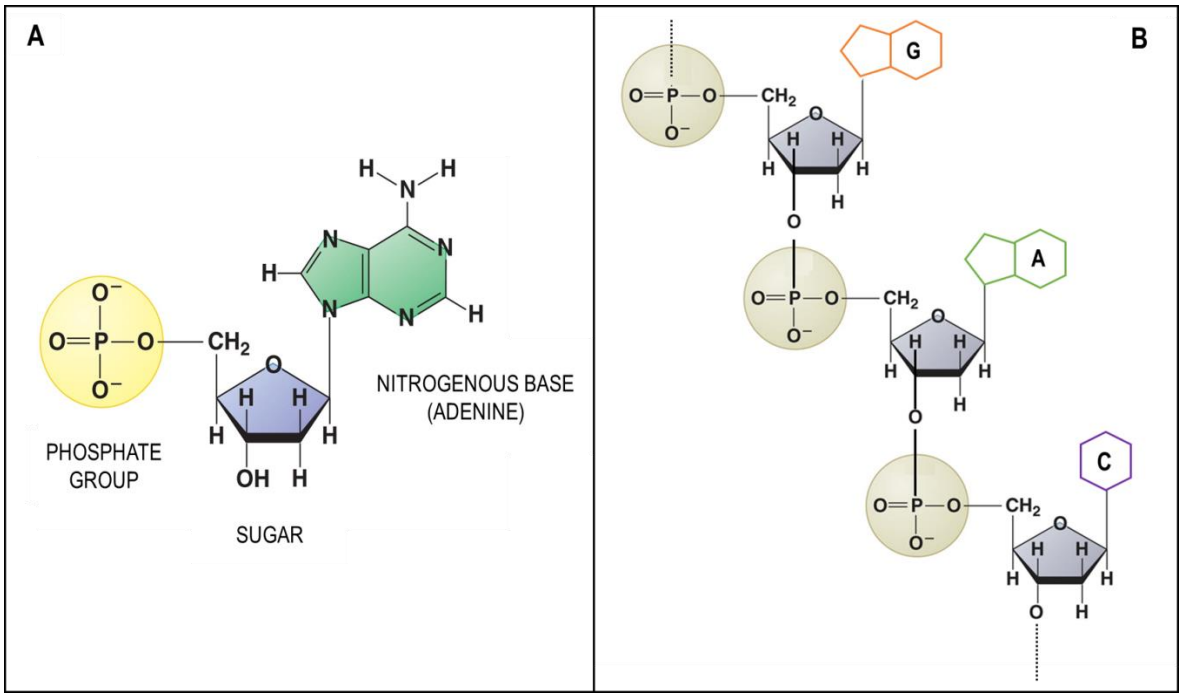


Figure 1.2 A) Nucleotide structure and B) Phosphodiester bonds linking nucleotides to each other to form the overall strand structure. The chemical structure of nitrogenous bases was simplified for clarity.

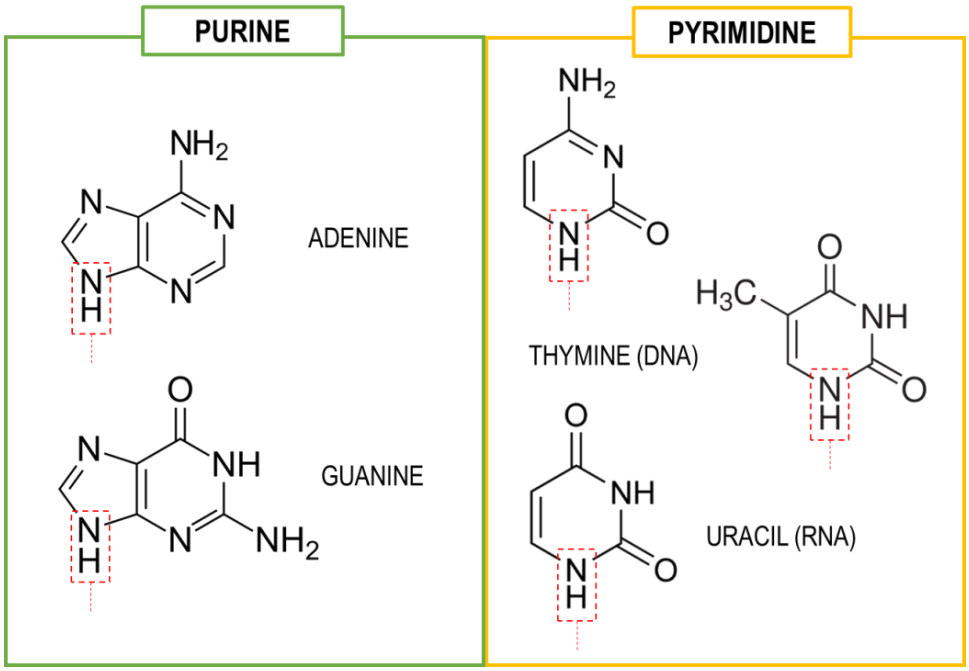


Figure 1.3 Detailed structure of common purines and pyrimidines bases found in DNA and RNA. The groups highlighted in red dotted lines indicate the sites of attachment of the sugar ring.

It is worth noting that even though A, C, G and T/U are the most common nucleobases in DNA and RNA respectively, other minority species are found in natural NA. Notably, products resulting from methylation process are the most common minor components (5-methylcytosine, N^6-methyladenine, N^4-methylcytosine or 5-hydroxymethyluracil).[7, 20] The methylation occurs after the synthesis of the DNA and, generally, plays an important role in gene regulation or protecting the genetic information. The importance of those nucleobase alterations will be further discussed in section 1.1.1.

Previously to 1953, when Watson and Crick elucidated the structure of double helix, several studies had been published giving information about the structure and composition of the DNA. Particularly, the quantification studies carried out by Erwin Chargaff (1949) were crucial for the understanding of the composition of the DNA.[21] *Chargaff et. al* studied the DNA composition of several organisms and they exposed decisive conclusions, nowadays known as *Chargaff's rules*. Those rules establish quantitative relations between DNA bases and they can be summarized as follows:

- Base composition is preserved in different tissues from the same organism. Moreover, it does not change over the time, nutrition or other environmental factors.
- On contrary, DNA base composition varies between different species.
- In all cellular double-stranded DNA of any organism (independently of the species), the number of adenine and thymine is the same.

Similarly, is repeated by the number of cytosine and guanine, then A=T and C=G (i.e., Parity Rule)

- From the Parity Rule was deduced that in a dsDNA the number of purine bases is equal to pyrimidine bases, therefore; A + G = T + C.

These rules were later supported by multiple studies and were crucial for the definition of the DNA structure by Watson and Crick.

Nucleic acid structure

DNA structure was defined as a double helix composed of two polynucleotide chains having the same helical geometry but aligned in opposite orientations. Hydrophilic chains composed of the sugar ring and the phosphate groups are pointing out the structure, in order to be in contact with the aqueous medium. On contrary, hydrophobic parts (purines and pyrimidines) are located flattened and stacked at the internal part of the helix[7] (**Figure 1.4**). Importantly, DNA chain shows an inherent polarity due to phosphodiester linkages (asymmetry of joined nucleotides), therefore free 5'-phosphate or 5'-hydroxyl at one end and 3'-phosphate or 3'-hydroxyl are found at DNA chains (**Figure 1.4**). The conventional way to write DNA sequences is from 5' to 3' end, generally with a 5'-phosphate group and 3'-hydroxyl.

The connection of both chains and the stability of the double helix are explained by two factors: base-stacking and base-pairing. Bases tend to be flattened and perpendicular to the helical axis direction (**Figure 1.4**), thus

favouring electron cloud interactions (π - π). This interaction, known as van der Waals forces, ensures the base stacking and contributes to the overall stability of the duplex.[19, 22]

The base-pairing is what explains the complementarity of the dsDNA and a fundamental feature for the stability of the double helix. Hydrogen-bonding is the responsible for the interactions between A:T and C:G. **Figure 1.4** shows how adenine and thymine match by 2 hydrogen bonding, while guanine and cytosine bases are connected by 3 hydrogen bonds.

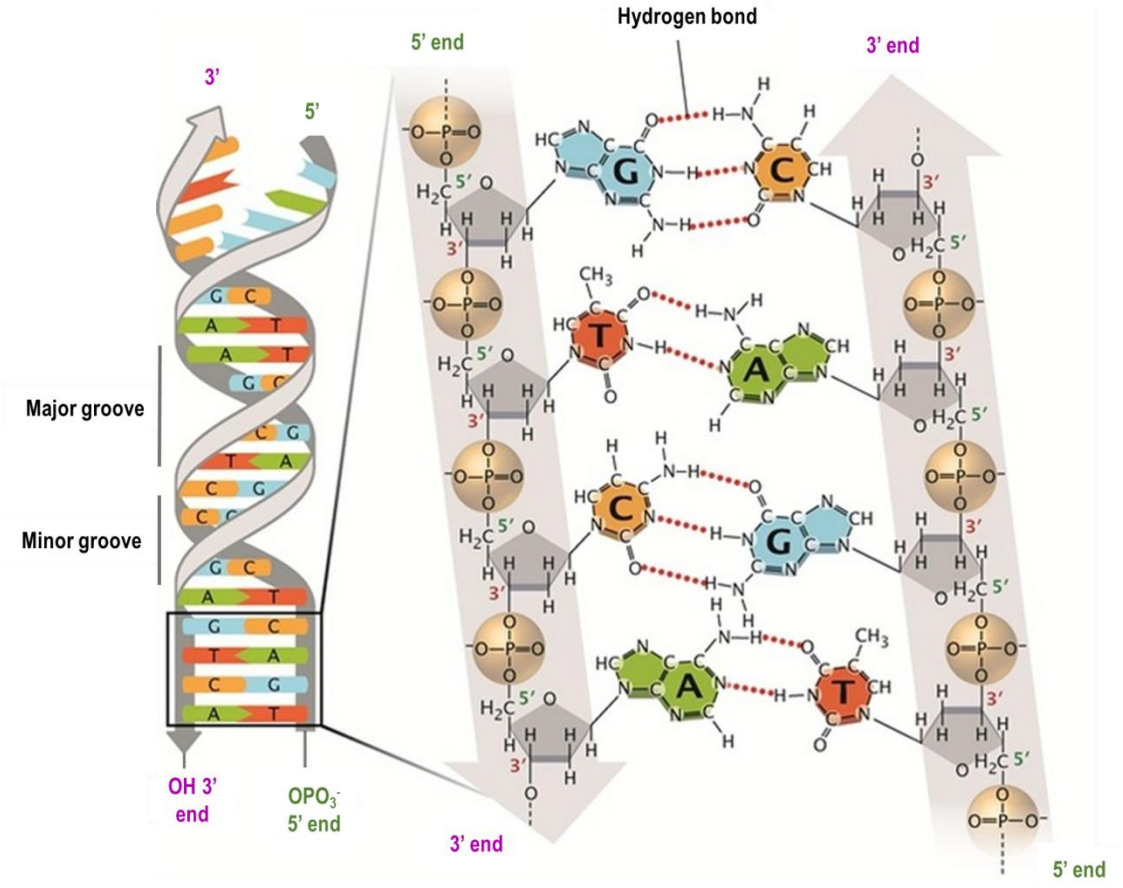


Figure 1.4 Double stranded DNA helical structure of the complementary chains in opposite directions. A portion of the sequence is amplified showing hydrogen bonds between complementary bases.

Hydrogen bonds established between bases are not only offering complementarity and stability to the duplex, but they are also responsible for the specificity of base pairing.[8] It is important to note that a base pair composed by adenine and cytosine or guanine and thymine would be unstable due to the inability to form proper H-bonds.

Moreover, another characteristic of the double helix is the presence of two grooves with different size. As a consequence of the base-pairing geometry produced along the same plane, the sugar ring (2'-deoxyribose for DNA or ribose for RNA) protrudes from the base pair in different angles. Larger angles of the sugar at the edge of the base pairs generate a major groove. On contrary, narrower angles lead a minor groove (**Figure 1.4**).

Nucleic acid 3-D structure

NA is a flexible molecule that can suffer changes in its structure such as rotations or curvatures. Moreover, if thermal treatments are applied, elongation or separation of the chains can occur.[19, 23] NAs are able to adopt different three-dimensional conformations and, generally, this variability plays an important role in the DNA metabolism.[7] The first forms of DNA described were A-form and B-form, discovered by Rosalind E. Franklin and R.G Gosling at 1953 by using X-ray fiber diffraction.[3a] Years later, a new DNA form was disclosed by Alexander Rich who named it as Z-form, due to the zigzag arrangement of the backbone of the molecule.[24] All three DNA forms are represented in **Figure 1.5.**

The vast majority DNA in cells exists in B-form, which is the most stable form and the one described by the model of Watson and Crick. B-DNA contains 10.5 base pairs per turn of the helix.[7, 19] On the other hand, the helix in an A-form DNA has an average of 11 bp, leading to a broader minor groove and narrower major groove as compared to the B-form.[7] The A-form is commonly found in low humidity conditions, contrary to the B-form that is usually observed at high humidity. Moreover, the A structure is adopted by the DNA in certain complexes with proteins.[25]

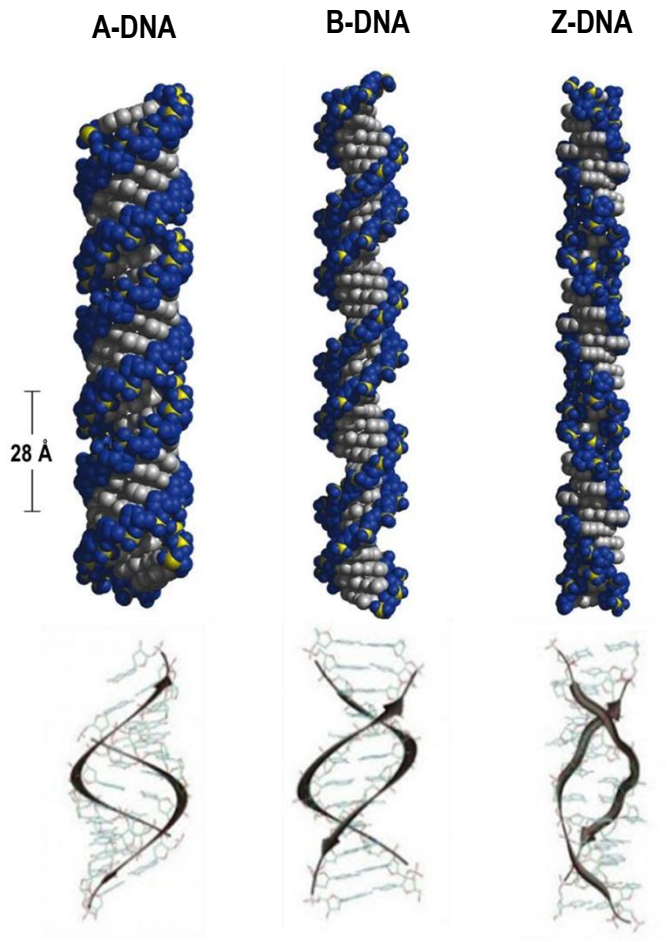


Figure 1.5 Three-dimensional structures of A-DNA, B-DNA, and Z-DNA.

The Z-form is the most distinct structure due to the left rotation of the helix instead of right rotation as A- and B-form. In that structure, the helix is thinner and longer (12 bp per turn).[26] The presence of cellular DNA in A-form is rarely observed, such as in the case of bacteriophages in which DNA is highly packaged.[25, 27] On the contrary, it has been demonstrated that small portions of DNA show Z-forms which could play an important role in gene regulation.[19, 28]

Particularities of the RNA

RNA differs from DNA in 3 main aspects:

- Ribose: The sugar ring of the nucleotides is ribose instead of 2'-deoxyribose (therefore at the position 2' there is a hydroxyl group).
- Uracil: The most common nucleobases are adenine, cytosine, guanine and uracil. Thymine is replaced by the structural analogue (**Figure 1.3**).
- ssRNA: RNA has the capability of forming duplexes, nevertheless in nature is more commonly found as a single strand (ssRNA), while DNA natural state is double helix (dsDNA).

All these differences are not merely structural or compositional, they affect the overall function of the RNA, turning it in a more versatile and complex molecule than DNA. The three-dimensional structures that RNA can adopt are much wider than DNA, and in most cases, the RNA function is fully dependent on its structure and dynamics.

RNA, similarly to DNA, is stabilized by weak interactions such as base-stacking. Nevertheless, many regions of RNA strands are folded on themselves when complementary sequences are in close proximity. Depending on the distribution of the base-paired segments, the structures that RNA can adopt are different. Stem-loop structure, internal loops, bulges or junctions (all represented in **Figure 1.6**) are some examples of the structural complexity and variability of RNA molecules.[29] Moreover, when RNA molecules are assembled in a double helix, the A-form is the predominant one due to the presence of the 2'-hydroxyl in the RNA backbone which prevents the adoption of the B-form conformation.

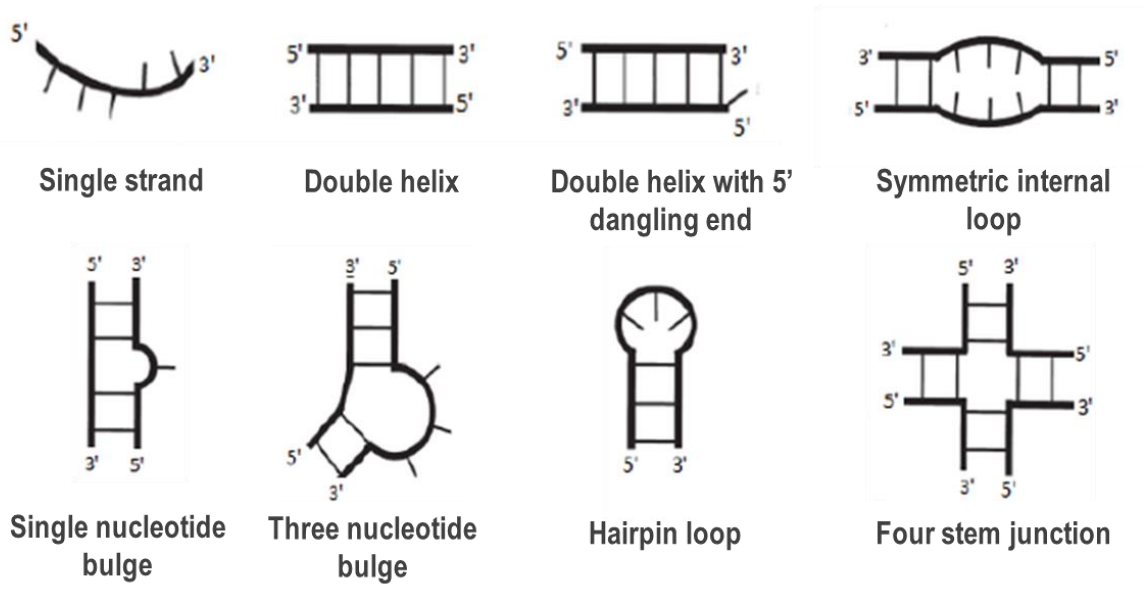


Figure 1.6 Schematic diagram of RNA stem and loop structural elements.

1.1.1 Epigenetic modifications in Nucleic Acids

Epigenetics variations are defined as the changes in the nucleic acid sequence that are inherited but do not trigger a change in the nucleotide sequence. These changes may affect gene activity or expression and, generally, they are induced by environmental factors or being part of the normal development program.[20, 30] It is already known that not all genes in an organism are active in each cell and that there are mechanisms to regulate their activation and expression. Modified nucleobases, as well known as adducts,[31] play a crucial role in that regulation, introducing a second layer of genetic information. Then, epigenetics could be understood as the study of the regulation and control of gene expression in a potentially heritable way.[32]

Epigenetic modifications could be grouped into three groups; *(i)* nucleic acids modifications (DNA and RNA), *(ii)* histone modifications, and *(iii)* nucleosome positioning[32] (**Figure 1.7**). DNA molecules wrap the histones which work as packagers for the genetic material at the cell nucleus. They are usually subjected to post-transcriptional modifications, where the most common ones are methylation or acetylation occurring at histone tails. Therefore, histones can be considered key players in epigenetics.[33] In addition, nucleosomes positioning could interfere with the DNA transcription by blocking the access of transcription factors and activators. Nucleosomes are the basic units of DNA packaging (composed by 8 histones surrounded by DNA sequences). Conformational

changes due to epigenetic modifications can significantly affect the accessibility to the DNA sequences during the transcription process.[33-34]

In this section, the attention will be focused on the epigenetic modifications in DNA and RNA. Nevertheless, it is important to take in consideration that epigenetics is not only affecting the genetic material, but also nucleosomes and histones.

DNA methylation and particularly, cytosine methylation at 5' position represents one of the most critical epigenetic modification in eukaryotic DNA.[35] Methylation process is typically associated with a gene repression and is fundamental for the normal regulation of gene expression.[35] 5-methylcytosine is widespread in both eukaryotes and prokaryotes and it has been widely studied as the most common epigenetic modification in humans.[35-36] The importance of human methylation was elucidated because of their relation to cancer development. Many evidences show that alterations on the methylation pattern play a critical role in the regulation of oncogene activation, tumour suppression gene silencing and chromosomal instability.[32, 37] Demethylation process of 5-methylcytosine, leading to 5-hydroxymethylcytosine is considered as well an important epigenetic regulation system.[38] It is worth mentioning that epigenetic modifications are not only methylation-associated processes, other modifications such as the addition of Bromo atoms[39] or oxidative reactions on DNA bases[40] produce as well epigenetic changes on DNA.

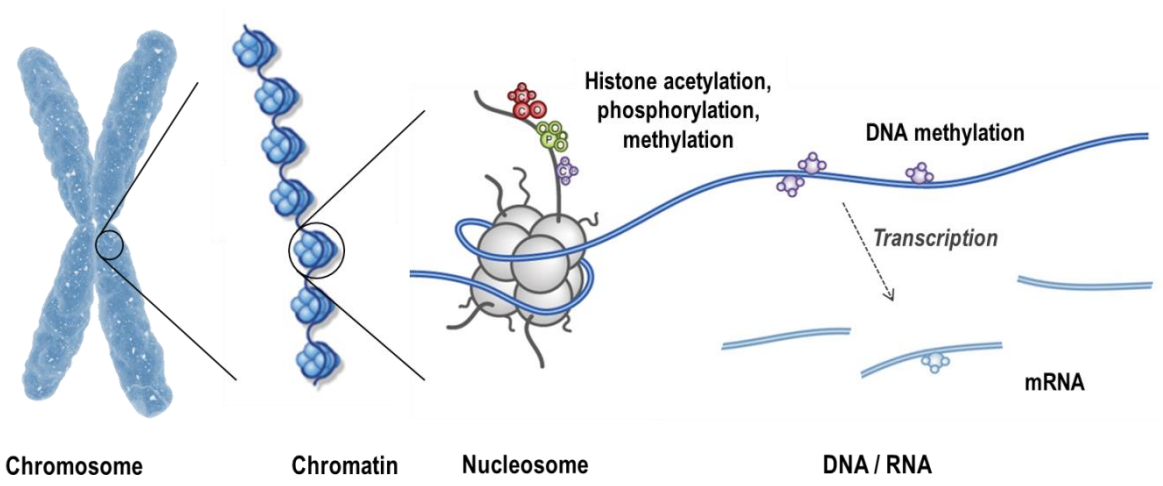


Figure 1.7 Overview of the epigenetic modifications mechanism. DNA strands are wrapped around histone octamers which form the compact chromatin of the chromosomes. Modifications on histones occur at multiples positions as well as in DNA where methylation is the most common modification.

Epigenetic modifications are not only localized in DNA sequences, chemical alternations of the RNA nucleosides have been widely described (more than 100 RNA modifications have been reported).[20] The largest diversity of modified RNA is found in tRNA, even though all RNA species present modified nucleosides, such as pseudouridine (ψ) which is present in almost all types of RNA. Similarly to DNA, one of the most common modifications in RNA is methylation which occurs nearly in all types.[41] Because of the abundant functions in which RNA molecules are involved, there are many methylation processes whose biological role remains unclear.[41] The most abundant modification in mRNA is N^6-methyladenosine which is implicated in stress responses, showing dynamic changes depending on the stress conditions.[42]

During these last years, epigenetics has experienced an increasing interest and awareness. New adducts have been discovered considering their key role not only in tumorigenesis but also in other human diseases such as neurodevelopmental disorders.[30, 32] The study of such modifications in DNA and RNA, and their detection and characterization is crucial to understand the development of certain diseases as well as for the improvement of their diagnosis.

Conventional techniques

Several methods for the analysis of DNA modifications are currently employed. Liquid chromatography-mass spectroscopy (LC-MS) is the most frequently used for characterization and quantification.[31] This technique is based on the ionization of the analytes to separate them accordingly their mass to charge ratio (*m*/*z*).[43] However, it requires laborious sample preparation comprising several steps (denaturalization and hydrolysis of the DNA chain to monomers, adducts enrichment, removal of the unmodified bases and inclusion of an internal standard).[43] Even though instrumentation improvements have been introduced to optimize the chromatography and sample preparation,[38b] this technique remains costly and time-consuming. In addition, exceptional care must be paid to avoid the generation of artificial adducts.

The coupling of High-Performance Liquid Chromatography with electrospray tandem mass spectroscopy (HPLC-MS/MS) is an example of a successful technology with new analytical abilities for the identification of DNA modifications.

A good sensitivity can be achieved, allowing background levels measurements,[44] even though the main disadvantages are its complexity and high costs. Moreover, as DNA adducts can be thermally unstable, other methods such as gas chromatography-mass spectroscopy (GC-MS) are less frequently used.

^{32}P-Postlabelling assays are widely used from the 1980s because of the capability to simultaneously monitor several modified nucleotides.[45] However, as LC-MS is improving providing more selective and quantitative measurements, the application of ^{32}P-Postlabelling methodology is decreasing due to their intrinsic limitations, mainly related to the requirement of significant amounts of radioactive phosphorous.[31, 45]

Immunological and fluorescence assays are widely used in genomics, but less common for the detection and purification of pure adducts.[42b] On contrary, next generation sequencing (NGS) technologies are gaining importance as a genome-wide profiling method.[46] NGS provides a single-base resolution sequencing, useful for the elucidation of new adducts in DNA. Moreover, it allows for the analysis of large sample data in a low-cost fashion as compared to other techniques. NGS technologies have a broad range of applications, however, advances are needed for the efficient isolation of DNA in its original form, and the development of easy and fast methods for accurately reading the substantial content of produced data.

1.1.2 The importance of miRNA detection

MicroRNA (miRNA) is a non-coding protein RNA, single-stranded and generally composed of 19 to 23 nucleotides. It plays critical regulatory roles in a wide range of biological processes such as early development, cellular differentiation, proliferation or apoptosis.[14] miRNA is one of the more abundant classes of gene regulatory molecules in multicellular organisms. Over 15000 mature miRNA are listed in miRBase 15.0 database[47] (*http://www.mirbase.org/)*, particularly in humans ~1000 miRNA have been identified which can target >30% of the human genome.[14, 47] Specifically, miRNA molecule with RNA-induced silencing complex (RISC) can pair with the target messenger RNA (mRNA) molecule (**Figure 1.8**) guiding different mechanisms; *i)* if there is a total complementarity, cleavage of the target mRNA is induced (**Figure 1.8A**), *ii)* when the recognition between mRNA and miRNA is limited to partially complementary sequences, repression on the translational machinery is triggered (**Figure 1.8B**).[14, 17] The repression of that mechanism leads to reduced protein levels, and to further consequences on the next biological processes. Because of the imperfect pairing, a single miRNA might have several targets and a single gene can be regulated by numerous miRNA molecules.[48]

Published results show the multiple functions in which miRNA is involved in specific tissues[49] or at determined situations, such as stress conditions.[50] Alternations on the expression pattern of miRNA have been associated with numerous diseases such as neurodegenerative diseases,[51] autoimmune

disorders,[48] diabetis[52] or cancer.[53] In particular, it is worthy to note that miRNA can play as both oncogene and tumor suppressor.[53] Thus, miRNA expression profiles can be used as biomarkers for determining the state of certain diseases even is possible to use it for gene therapy or as potential drug targets.[17]

The research on the miRNA field is continuously evolving. It is known that miRNA performs gene regulation through mRNA cleavage or translational repression, but the specific mechanisms are not fully understood yet. Their roles in regulatory pathways and their analysis as new targets for miRNA-based therapies need more exploration. Nonetheless, their intrinsic characteristics such as their small size or low abundance hinder their analysis by conventional techniques.[17]

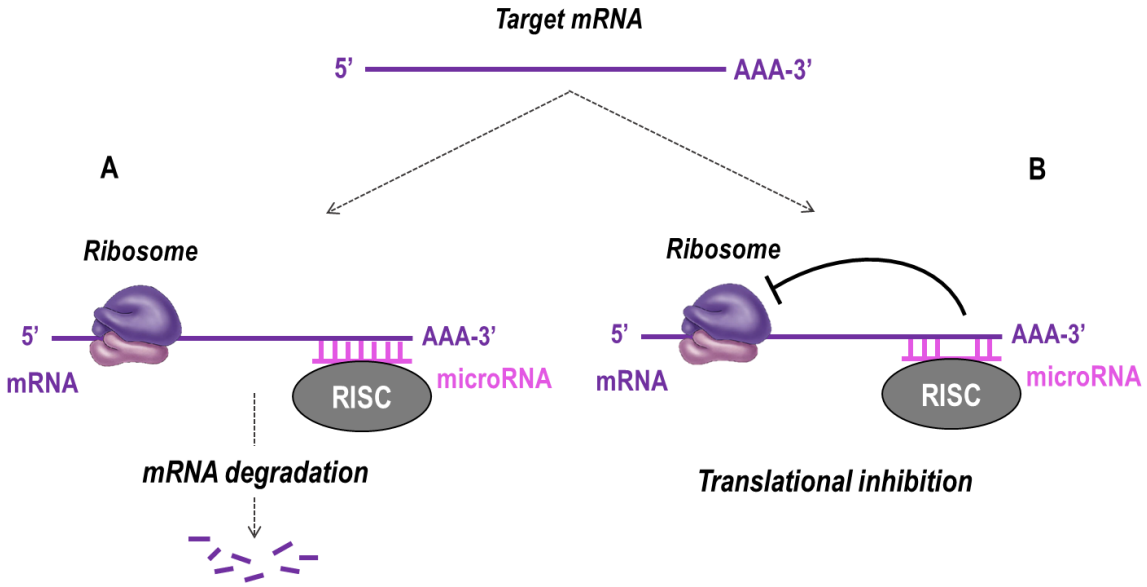


Figure 1.8 Diagram showing gene regulation mechanisms of miRNA with RISC complex. A) degradation of mRNA transcripts through cleavage, B) repression of mRNA translation.

Conventional techniques and emerging of nanotechnological approaches

Many efforts have been devoted to the development of efficient and reliable techniques trying to overcome the complexity of miRNA detection, which can be outlined as follows:

- The small size of miRNA (19-23 nucleotides) is a limiting factor for the employment of certain techniques such as Polymerase Chain Reaction (PCR) methods, widely used for the analysis of genetic material.
- miRNA is found in low abundance representing only a small fraction of the total RNA (ca. 0.01% of the total mass in RNA sample).[54] This low amount of target requires a highly sensitive detection technique. Moreover, miRNA of the same family could be very similar in sequence between them, making the specific detection even more complicated to achieve.
- miRNA are very dynamic molecules having more than one target each, meaning that the same gene can be regulated by multiple miRNA molecules.[48] That property requires a technique allowing multiplex detection in order to understand the complex regulatory function of those molecules.

Considering the key role that miRNAs develop in biological functions and all those characteristics mentioned, the advancement on rapid, sensitive and quantitative miRNA detection is imperative.

PCR methods are very helpful to amplify the miRNA content in samples with minimal amounts by annealing and extension events of complementary oligonucleotides primers. In a conventional PCR, small sequences are needed to work as primers which are essential for the correct amplification. The primer length is approximately the same of miRNA sequences.[55] Therefore, for the design of the PCR analysis of miRNA, shorter primers than the conventional ones would be needed. Although the design of shorter primers is possible, it negatively affects the PCR efficiency mainly in two ways: *(i)* it reduces the selective detection abilities due to the difficulty of labelling shorter primers, *(ii)* the hybridization binding between primer and target weakens therefore increasing the risk of cross-hybridization which can be easily perceived as false positive.[56]

Nowadays, the standard method for miRNA detection is Northern blotting, widely used for expression profiling.[55, 57] However, northern blotting is a laborious technique comprising many steps and a radioactive labelling (typically ^{32}P).[57] In particular, the main drawbacks are: *(i)* the lack of sensitivity (usually requiring up to 20 μg of RNA sample)[57] and failing to detect miRNAs in low abundance, *(ii)* time-consuming because often it takes days for culmination, and *(iii)* limited dynamic range (2-3 orders of magnitude). Even though Northern Blotting allows the analysis of multiple samples, only one miRNA sequence can be analyzed at the same time.[56-57] Considering that multiple targets can be recognized by the same miRNA sequence, the capability to perform the global analysis is becoming increasingly important in the design of the optimal technique.

Alternatively, miRNA microarrays and quantitative Real-Time PCR (qRT-PCR) are also used for the study of expression patterns and to acquire quantitative information, respectively.[58] miRNA microarrays are commonly used as screening techniques for profiling large number of miRNA sequences but they tend to have lower sensitivity and smaller dynamic range than Northern Blotting.[17] They have the advantage of being cheaper allowing numerous parallel measurements, however, in some cases imperfect specificity is found. On the other hand, qRT-PCR has the advantages of covering wide dynamic ranges and high accuracy, while providing absolute miRNA quantification.[56] The technique relies on the reverse transcription of miRNA to complementary DNA (cDNA) which is amplified and monitored by qPCR, providing precise profiling results.[59] The sensitivity of qRT-PCR is very low (down to nanogram levels of the total RNA),[59-60] nevertheless some parameters such as RNA integrity control or primer design have to be taken carefully into account to yield meaningful results. Generally and because of the associated costs, microarrays and Northern Blotting are firstly used for a miRNA screening and, then, the results are validated by qRT-PCR.[58]

Exploration of new strategies capable of overcoming the limitations of the established techniques is a major topic in science. The use of nanotechnology and, particularly, nanoparticle-based detection are of increasing interest because of their advantages of the capability to work at a lower limit of detection, to enhance the sensitivity and accurateness, and the ability to perform rapid analysis.[56, 58] These strategies comprise a wide variety of sensing schemes. In

some cases, magnetic nanoparticles (NPs) are used for sample concentration.[61] Other examples are the use of NPs as carriers of miRNA for achieving cell transfection, exploiting gold NPs due to their unique optical properties and biocompatibility.[62] Furthermore, NPs can be used in conjunction with established techniques such as microarrays, engineering detection platforms with high specificity and reproducibility for miRNA detection at low amounts.[63]

In addition, surface plasmon resonance (SPR) from the metallic nanoparticles can be exploited for miRNA detection and classification. For instance, a platform consisting on silver nanorods array was reported to classify different miRNA sequences.[64] However, most of those strategies based on SPR effect from metallic NPs rely on indirect strategies or they have limited multiplexed and quantification capabilities. Therefore, more efforts have to be devoted to obtaining approaches which provide sensitive, reproducible and direct signals of miRNA without compromising their specificity.

1.2 RAMAN SCATTERING AND SURFACE-ENHANCED RAMAN SCATTERING

1.2.1 Raman Spectroscopy

Raman spectroscopy is a spectroscopic technique based on the Raman effect that directly informs about vibrational or rotational modes, or other low-frequency modes, in a molecular system. When light interacts with matter, two main mechanisms can occur: absorption or scattering. If the substance absorbs the light is because the energy of the incident photon corresponds to the energy gap between the ground and the excited state of an atom or molecule. On contrary, scattering can occur whether or not there is a suitable pair of energy levels that allow absorption of the radiation.[65]

Scattering processes (in which the emitted photon is called scattered photon) can be classified into two groups:

- Inelastic Scattering or Raman Scattering (RS): It takes place when a molecule is illuminated and a part of the light is scattered with a different wavelength. The energy difference between this scattered light and the incident light is due to the interaction of the photons with the vibrational states of the molecule under study. The phenomenon was predicted by A. Smekal[66] at 1923 and later demonstrated by C.V. Raman[67] at 1928.
- Elastic Scattering or Rayleigh Scattering: It occurs when the excited molecular system relaxes without any noticeable change in energy (elastic process). The incident and the scattered photons have the same

energy (with a different direction and/or polarization). It is often referred to Rayleigh scattering because it was described by Lord Rayleigh on 1871.[68]

There is a remarkable difference in the efficiency of Rayleigh and Raman scattering since one photon out of 1000 is re-emitted as Rayleigh scattering, whereas only 1 out of 10^6-10^8 undergoes Raman scattering.[65a] **Figure 1.9** shows a simple diagram illustrating Raman and Rayleigh scattering.

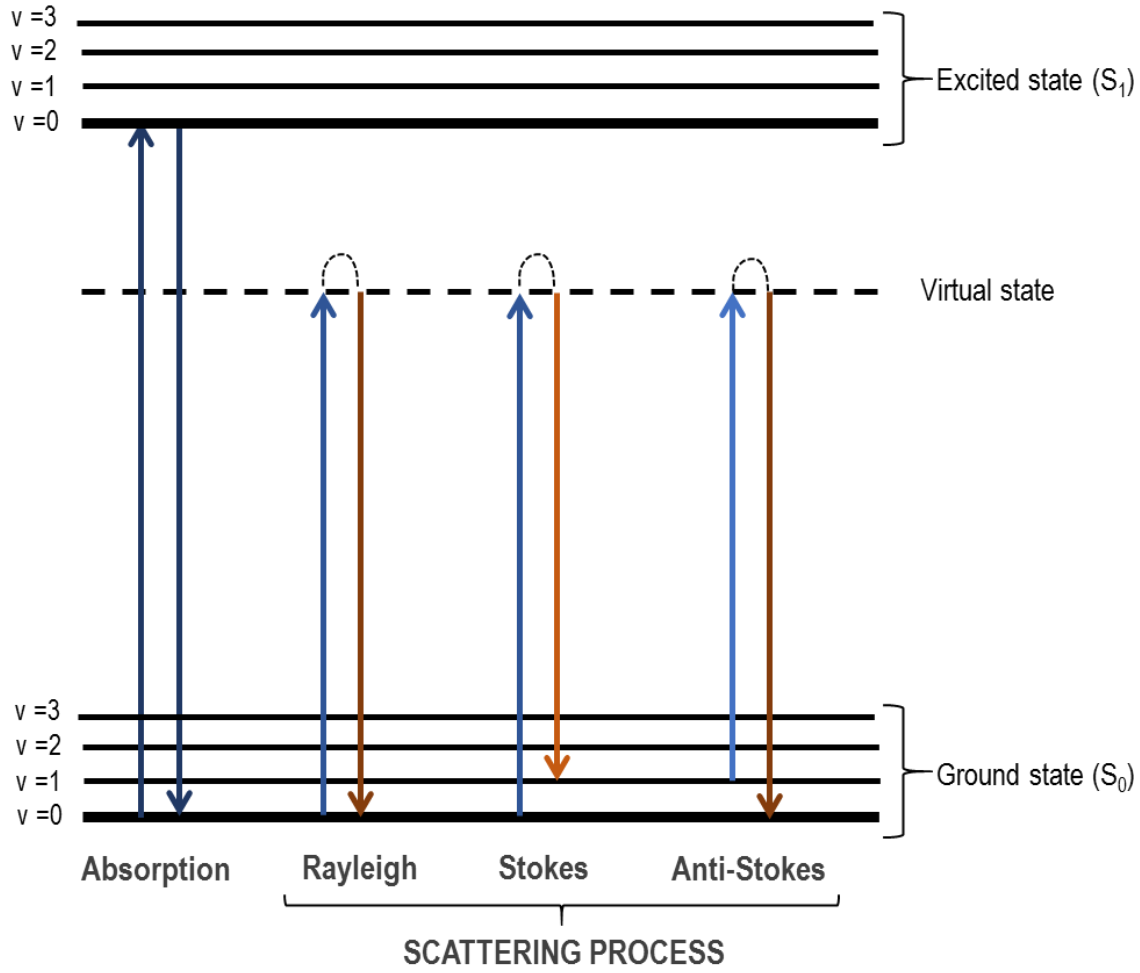


Figure 1.9 Simplified Energy diagram showing transitions involved during absorption, Rayleigh, Raman Stokes and Anti-Stokes processes. (v) refers to the vibrational state.

The Raman scattered photon has either lower or higher energy than the incident one. When the scattered photon has less energy than the incident is called *Stokes process* (**Figure 1.9**). In that case, the molecule is excited from the vibrational ground state (v=0) to a higher state (v=1). On contrary, the scattered photon can have more energy than the incident photon producing a relaxation of the molecule, from an excited vibrational state to its ground state. That process is called *anti-Stokes* (**Figure 1.9**) and is normally much weaker than *Stokes process,* mainly because at normal temperature (T), most of the molecules exist in the ground vibrational state.

Raman Spectrum and Raman shift

Raman shift is described as the energy lost or gained by the photon in a scattering event, therefore it is positive for Stokes and negative for an anti-Stokes process (**Figure 10**). In terms of energy, Raman shift is described as $\Delta E_R = E_L - E_S$ where E_L and E_S are the energy of the incident photon and the scattered, respectively. It is commonly expressed in wavenumber (cm^{-1}). At a given incident wavelength, the Raman spectrum corresponds to the wavelength dependence of the Raman scattered intensity, as can be shown in **Figure 1.10**.

The information given by a Raman spectrum can be considered as the unique fingerprint of the molecule, where each peak in the spectrum corresponds to a vibrational mode of the molecule under analysis. The vibrational energy of each mode is represented by their Raman shift varying from mode to mode.

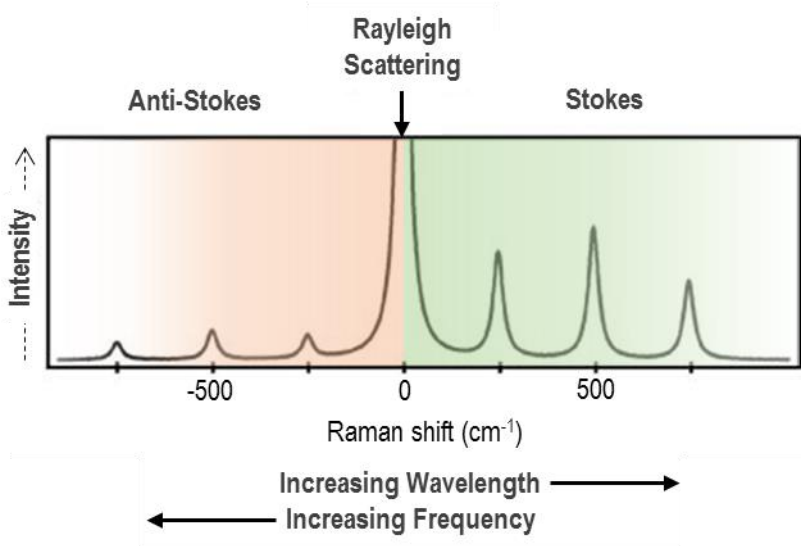


Figure 1.10 Typical Raman spectrum (Raman intensity *vs* Raman shift) with several Raman peaks on the Stokes side and their weaker anti-Stokes counterparts.

Raman cross-section

In addition, to describe qualitatively the optical processes, there is a quantitative variable that can measure how many incident photons are scattered (through Raman or Rayleigh). This proportionality constant is called Raman cross-section (σ) and is a function of the excitation light energy, having a dimension of area (typically expressed as m^2). *Equation 1.1* shows the simplest description of the cross-section of a given linear optical process underwent by a molecule, where P is the power of the signal produced and S_0 corresponds to the incident power density.

$$P = \sigma \times S_0 \qquad 1.1$$

Nevertheless, scattering is not a linear optical process, therefore to fully describe the process where many photons are scattered in specific directions this same equation is developed to a differential scattering cross-section.[69]

It is worth noting that scattering process depends on the orientation of the molecule with respect to the incident field polarization and, in a practical case where large numbers of molecules are measured with randomly orientations, the signal obtained is an average of the overall orientations. Therefore, Raman cross-section definition is referred to molecules with randomly-averaged orientations.

1.2.2 Surface-Enhanced Raman Scattering (SERS)

Surface-Enhanced Raman Scattering (SERS) is based on the enhancement of the Raman signal for molecules adsorbed on nanostructured plasmonic materials. The presence of metal nanostructures is a requirement for SERS. When a nanomaterial is irradiated by the appropriate light, a strong field can be generated at the metal surface which couples with the vibrational modes of the molecules in close proximity, leading to a giant amplification of the Raman scattering. Nowadays, enhancement factors up to 10^{13} - 10^{14} fold have been reported, providing SERS with single-molecule sensitivity.[70] The main mechanism of SERS enhancement is electromagnetic in nature, while additional intensification may arise from the so-called chemical mechanisms.

Differently from normal Raman scattering, which is restricted to the interaction between light and matter, in SERS the electromagnetic interaction between light and metal nanostructures is at the heart of the signal enhancement. The adequacy of the metallic material used for SERS strongly depends on the metal nature itself. The dielectric function of a metal (ε), expressed by a real (ε') and an imaginary (iε'') part is an indicative of the metal behaviour with light (*Equation 1.2*). The real part is related to the scattering efficiency and the imaginary part is associated with the absorption efficiency. The most suitable metal for SERS should have a negative real part and small imaginary part of the ε. Among different metals, those which fulfil these requirements are the noble metals, such as gold and silver (widely used in SERS spectroscopy).[65b, 71]

$$\varepsilon = \varepsilon' + i\,\varepsilon'' \qquad 1.2$$

The remarkable analytical potential of SERS, its versatility and the rich structural information that can provide, paved the way to its applications in a wide range of fields.[72] Moreover, it requires minimal sample preparation and it can be applied in almost any analytical matrix.

1.2.3 Electromagnetic mechanism

The main contribution to the increment in the Raman signal is electromagnetic (EM) in nature. When photons of appropriate energy impinge on a plasmonic nanostructure, an intense local electromagnetic field at the metallic surface is generated. Specifically, when a small spherical metallic nanoparticle is

irradiated by light, the oscillating electric field originates a coherent motion of the conduction electrons generating a dipole (**Figure 1.11A**). The collective oscillations of the surface conduction electrons that propagate along the metal surface are known as plasmons. As a result of the confinement of surface plasmons with a wavelength of light used to excite, Localized Surface Plasmon Resonances (LSPRs) are induced. LSPR is the responsible for the remarkable optical properties of nanosized metals. It is important to note that LSPR is highly dependent on the nanoparticle morphology, as the oscillation of the electron cloud can occur along different symmetry axes depending on the morphology. For this reason, as illustrated in **Figure 1.11B** and **C**, the LSPR bands represented in the UV-Vis spectra for a nanosphere or a nanorod are different.

Therefore, the EM enhancement relies on field enhancement of both incident and scattered fields as a consequence of the LSPR excitation. To profit from such near-field enhancement, the molecule must be close to the surface (within ~10 nm maximum)[73] or, as in most practical cases, the molecule should be directly absorbed on the surface.

A variety of terms exist to describe different aspects of plasmons, here it is only presented the most basic theory needed to outline the SERS effect.

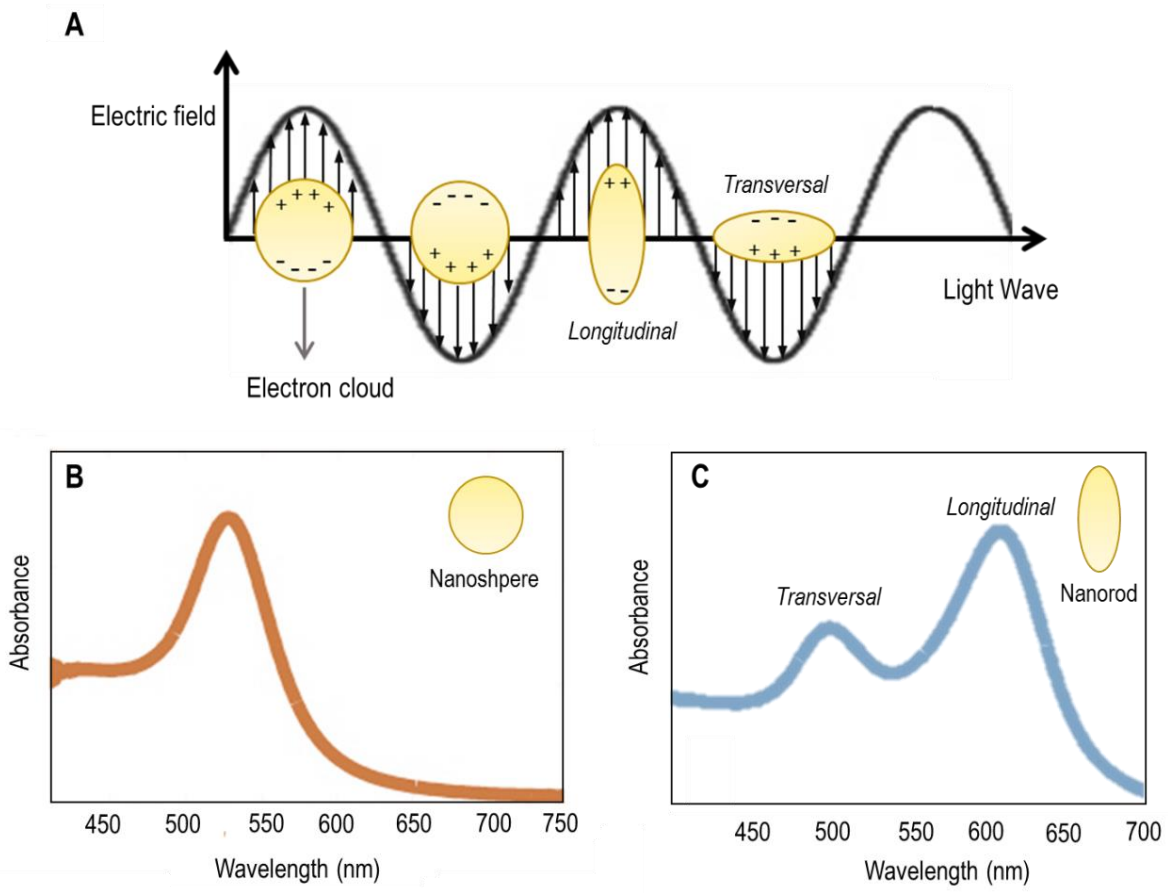
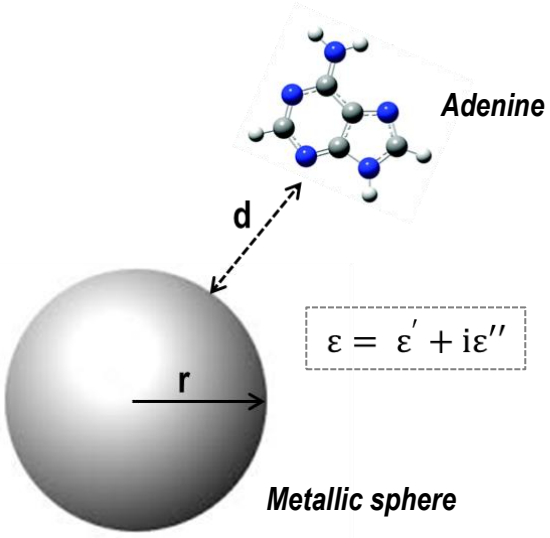


Figure 1.11 A) Representative scheme depicting the induction of LSPR in a metallic sphere and an elongated (nanorod) particle upon excitation with an electromagnetic radiation. An electric dipole is created in the small spherical particle, whereas the oscillation of the electron cloud can occur along any of the two symmetry axes of the ellipsoid. Experimentally, this is reflected as one or two (transversal and longitudinal) LSPR bands in the UV-Vis-NIR spectra of Au nanospheres (B) or nanorods (C).

The simplest description of the electromagnetic mechanism is based on small metallic sphere models. When the light is resonant with the LSPR, the metal sphere radiates its own dipolar field (E_{SP}) (**Figure 1.12**). The extension of which E_{SP} perturbs the molecule near to the surface depends on several factors such as the radius of the metallic sphere (r), the distance between the molecule and the

metallic surface (d), the dielectric constants of the metal (ε) and the surrounding medium (ε_0) and the incident field strength (E_0) *(Equation 1.3).*



$$E_{SP} = r^3 \frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0} E_0 \frac{1}{(r+d)^3} \qquad 1.3$$

$$E_M = E_0 + E_{SP} \qquad 1.4$$

Figure 1.12 Schematic representation of the electromagnetic effect (not represented at real scale).

The molecule perceives an enhanced local field (E_M) which includes both electric field magnitudes: $E_0 + E_{SP}$ *(Equation 1.4).* The ratio between the field at the position of the molecule *(E_M)* and the incoming field *(E_0)* explains the light field enhancement *A(v)* described in *Equation 1.5*.

$$A(v) = \frac{E_M(v)}{E_0(v)} \approx \frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0} \left(\frac{r}{r+d}\right)^3 \qquad 1.5$$

According to this equation, the most favorable conditions for the resonant excitation of the metal surface plasmons (which approximates the largest EM SERS enhancement) would be; *i)* when the real part of ε(v) is $-2\varepsilon_0$, and *ii)* when the imaginary part of the dielectric constant is small. Additionally, the scattered field will be enhanced if it is in resonance with the particle LSPR. Thus,

considering the enhancing effects for the laser and the Stokes field, the electromagnetic enhancement for the Stokes signal intensity (G^{SERS} (ν_s)) can be expressed as:

$$G^{SERS}(\nu_s) = |A(\nu_L)|^2 |A(\nu_S)|^2 \approx \left|\frac{\varepsilon(\nu_L) - \varepsilon_0}{\varepsilon(\nu_L) + 2\varepsilon_0}\right|^2 \left|\frac{\varepsilon(\nu_S) - \varepsilon_0}{\varepsilon(\nu_S) + 2\varepsilon_0}\right|^2 \left(\frac{r}{r+d}\right)^{12} \quad 1.6$$

The *Equation 1.6* describes important and specific properties of the EM in SERS, showing that the enhancement scales as the fourth power of the local field of the metallic nanostructure. This is recognized as the *E^4 approximation* and explains how particularly strong is the enhancement when the excitation and scattering fields are in resonance with the surface plasmons.

Enhancement Factor (EF)

SERS electromagnetic enhancement factor (EF) can be elucidated from the previous model (*Equation 1.6*) by separately taking into account the enhancement of both the incident and the emitted radiation. EF can be simply expressed as in *equation 1.7*, where M_{Loc} (ω_L) is the local field intensity enhancement factor and M^d_{Rad} (ω_R) the directional radiation enhancement factor.

$$EF = M_{Loc}(\omega_L) \times M^d_{Rad}(\omega_R) \quad 1.7$$

In order to avoid complications, it is often assumed that M_{Loc} (ω_L) ≈ M^d_{Rad} (ω_R). Moreover, in many cases, the Raman shift is also small and the additional approximation; $\omega_R = \omega_L$ is considered,[74] leading to the more popular expression *(Equation 1.8)* of the SERS enhancement for zero-Stokes shift;

$$EF(\omega_L) = \frac{|E_{Loc}(\omega_L)|^4}{|E_{Inc}|^4} \qquad 1.8$$

Where E_{Inc} is the incident field yielding a local field E_{Loc}. That approximation is generally sufficient to obtain the right order of magnitude of the EF. However, it is important to keep in mind that this approach only approximates the radiation enhancement and does not account for polarization effects or surface selection rules in SERS.

Maximization of the EM enhancement (Hot-Spots)

The *E^4 approximation* model above mentioned is commonly used to explore the EM enhancement of a particular metallic surface. In the case of isolated nanoparticles, EM enhancement in the order of 10^6 is the maximum value reported.[70a] Nevertheless, larger EM enhancements can be achieved in other conditions, for instance, the electromagnetic SERS enhancement factor can be increased up to 10^{11} when the morphology changes from a sphere to structures with sharp tips or edges.[75] In addition, closely spaced interacting particles can provide extra field enhancement (EM enhancement factors up to 10^{11} have been estimated for dimers of silver particles).[76]

These large enhancements are normally attributed to highly concentrate electromagnetic (EM) fields associated with strong localized surface plasmon resonances at interstitial sites (so-called EM or SERS **hot spots**).[77] The size of the hot spots is about few nanometers and the corresponding field concentration at

these gaps depends strongly on the geometry of the nanostructured site where the probed molecules reside, the wavelength and the polarization of the incident line.[73] However, hot spots are much less frequent than sites with moderate enhancement factors, therefore the probability of finding a molecule positioned exactly at the hot spot is much more complicated.

1.2.4 Additional Chemical mechanisms

Additional complementary mechanisms to EM can arise from chemical or charge transfer processes (CT). CT is related to the chemical complexation of the analyte with the metallic surface which may alter the polarizability of the adsorbed molecule. Due to the new transit (chemical transfer) in the metal-molecule system, the vibrational pattern may be affected due to the change in the molecular symmetry, molecular orientation on the surface and light polarization, in comparison to a free molecule.[77]

Therefore, in the case of SERS, the laser excitation enables a CT excitation, which could be molecule-to-metal or metal-to-molecule, consisting of a transition between the Fermi level (E_F) of the metal nanostructure and the lowest unoccupied molecular orbital (LUMO) of the molecule. **Figure 1.13** shows the three different contributions to the chemical mechanism, where Fermi level of the metal is symmetric to HOMO and LUMO, highest and lowest occupied molecular orbital, respectively.

The laser energy can be directly in resonance with an electronic transition of a molecule-metal complex (**Figure 1.13A**). This case can be described as a normal Resonance Raman event. Moreover, the laser can also profit from an indirect coupling (charge transfer) through the metal (**Figure 1.13B**). This second situation is more sophisticated and occurs when the difference between E_F of the metal and HUMO or LUMO energies is matched by the laser.[69] Enhancement in the metal nanoparticle (**Figure 1.13C**) can occur as well due to a very strong local field when the excitation wavelength is resonant with the plasmon excitations.

Chemical enhancements normally provide a much lower contribution to the overall enhancement (up to two orders of magnitude). However, under typical SERS conditions, it is always difficult to separate the chemical effects from possible contributions of electromagnetic origin.

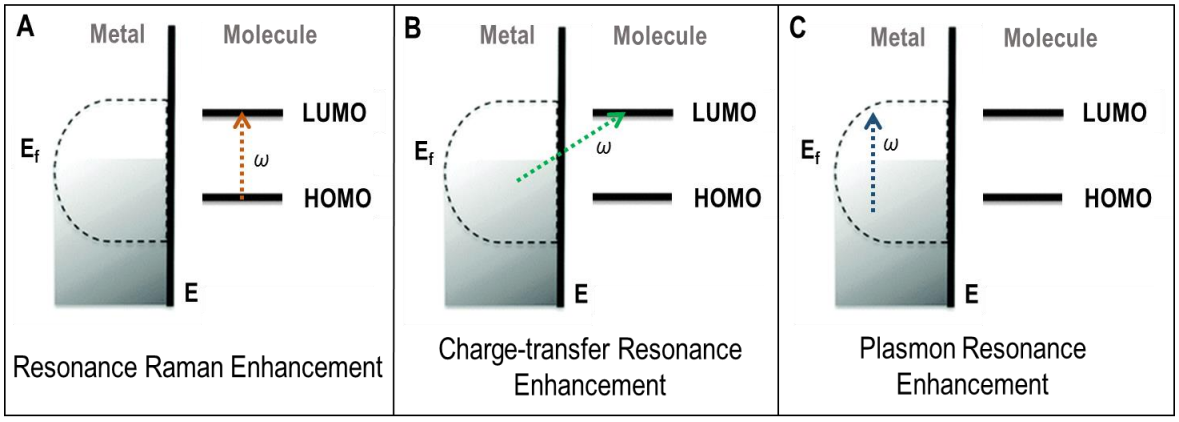


Figure 1.12 Illustration of the three different types of chemical enhancement mechanism in SERS.

1.2.5 SERS active Substrates

The term *SERS substrate* is generally referred to any metallic structures that produce SERS enhancement. The prepared SERS-active substrates should accomplish several requirements including large enhancement factors (EFs), good reproducibility from one sample to another, stable SERS signal with time, simple preparation process and easiness to be applied to many analytes. An ideal SERS substrate then possesses the following features.[78]

- The substrate should have high SERS activity and therefore provide high sensitivity. By tuning the LSPR frequency is possible to achieve the coupling and induce the maximization of the enhancement.
- The substrate must show good stability and reproducibility. Even after a long shelf time, the enhancement effect should still be maintained.
- The substrate should have a clean and easily modifiable surface so that it can be applied not only for strong but also weak adsorbates or even unknown samples.

Unfortunately, nowadays is still difficult to obtain SERS substrates that can simultaneously fulfill all of the above requirements. The task of finding '*the universal SERS substrate'* is not trivial and may not even be the right approach for implementing SERS as an analytical tool. Chemical and electronic effects have to be carefully controlled in order to bring SERS as close as possible to the basic analytical requirements. SERS can be maximized mainly through control of two

main factors; *i)* careful design of the optical substrate and *ii)* improvement of the absorption of the analyte under study. During the last years, the preparation of SERS substrates has become more controllable, benefiting from the development of nanoscience and nanotechnology. Many improvements in the synthesis and employment of those substrates have been elucidated and even new ones have been reported.[79]

From a general point of view, SERS substrates can be divided mainly into three classes; metallic nanoparticles (colloidal solutions); planar metallic structures supported on planar substrates such as silicon or glass, and metallic electrodes. In this section, the attention will be focused on metallic colloids and their particular relevance for SERS experiments.

Metallic colloids for SERS

Solutions of metallic colloids, predominantly made of silver (Ag) or gold (Au) represent one of the most common and simplest ways to generate highly active SERS substrates. Colloidal nanoparticles are advantageous for many reasons. Their versatility and ease of synthesis with a high degree of control over the composition, shape, and size, make them by far, the most commonly used substrates in SERS. Moreover, the utilization of nanoparticles allows the direct SERS analysis within analyte natural solution mediums.[80] The presence of the solvent and the Brownian motion of the analyte-particle complexes minimize damage to the sample, even when using more energetic laser lines and higher

power at the sample for excitation. Finally, colloidal metals can also be used for the preparation of thin films,[81] which add portability and versatility to “on-field” SERS analysis. On the other hand, regular physical fabrication techniques such as sputtering, physical vapour deposition or electron beam lithography, are often difficult to find in conventional laboratories.

Metallic nanoparticles are also of historical significance related to the development of SERS, as the first single molecule SERS detections were reported using colloidal particles.[70a] Colloids are typically produced by a reduction reaction (mostly in water) and via different chemical routes. Through careful control of the three key factors (composition, size, and shape), the optical and plasmonic properties can be optimally tuned in order to maximize the SERS effectiveness for specific applications like genetic diagnostics, or immunoassay labelling.

Despite all the advantages mentioned above, suspensions of metallic nanoparticles present some limitations related to the colloidal and optical stability. The concentration of the electromagnetic field is not homogeneously distributed around the whole nanoparticle surface as it has been experimentally observed at different nanostructures.[82] Specifically, in **Figure 1.14** is shown the case of gold nanostars where hot spots are clearly localized at the tips.[75b]

The presence of hot-spots in the SERS substrate can dramatically increase the SERS signal in an experiment. In the case of spherical nanoparticles, hot-spots are mainly generated when nanoparticles aggregate into clusters where the

individual oscillating dipoles of the individual nanoparticles are coupled. As consequence, new plasmonic responses may be generated within the aggregate over a broad range of frequencies, from the visible region to the near-infrared (NIR). Remarkably, largest plasmonic field enhancements are normally spatially localized at the interparticle junctions (further discussed in chapter 2).

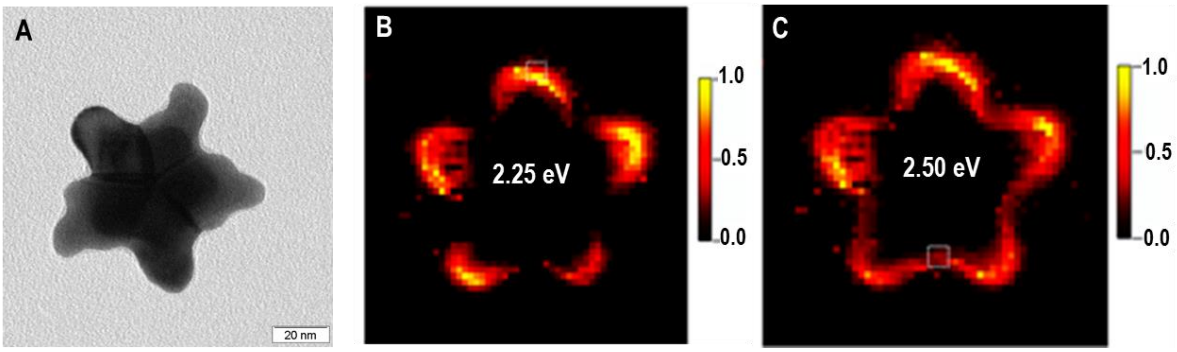


Figure 1.13 A) TEM image of a gold nanoshuriken (five-tipped nanostar). B and C) Intensity maps extracted from the EELS spectrum-imaging after removing the zero-loss peak. The intensity maps show the spatial distribution of the electromagnetic field.

In addition, it is important to note that not all functional groups present the same affinity for metallic surfaces. As the proximity of the analyte to the nanostructured surface is a key condition in SERS, promoting the efficient adsorption of the molecules remains a central issue in SERS experiments. Appropriate modification of the surface properties of the plasmonic enhancer represents the main solution to overcome issues associated with limited affinity of the analyte towards the nanostructured material.

Stabilization of metallic nanoparticles

The colloids in solution are stabilized by Coulombic or steric repulsions among particles, which usually requires the presence of a stabilizing agent coating the nanoparticle surface to prevent irreversible aggregation. Cetyltrimethylammonium bromide (CTAB) or poly(vinyl pyrrolidone) (PVP) are surfactants commonly employed in the fabrication protocol as capping agents. Additionally, these capping agents may also work as shape directing agents because of their different affinity for the different facets of a given nanoparticle.[75b] In some cases, such as sodium-citrate-reduced colloids, a single chemical compound plays both the role of reducing agent and stabilizer (one of the most used types of colloids in SERS).[83]

Silver *vs* Gold

The predominant metallic nanoparticles are made of silver or gold because of their intrinsic characteristics as a good SERS substrate among the other noble metals. Colloids are typically produced by a reduction reaction in solution through several possible chemical routes.

Au colloids normally display higher chemical stability (gold is less susceptible to oxidation), ease of functionalization and biocompatibility. Further, the extent of control over size and shape that can be exerted on Au colloids is generally much higher than on Ag, due to the higher silver reactivity.[84] On the other hand, the silver is a much more efficient plasmonic material than gold (at least in the visible

spectral range). In fact, the optical properties of gold are strongly affected at a shorter wavelength ($\lambda < 600$ nm) by a large optical absorption which significantly damps the LSPRs.[85]

1.3 NUCLEIC ACID ANALYSIS USING SERS

NA sensing is a very active field with relevant impact in pathology, genetics, and clinical diagnosis, but also food and drug industry or environmental monitoring. SERS offers several advantages for the detection of biomolecules such as DNA or RNA, which can be summarized as follows:

- In addition to the exploitation as a pure sensing method, SERS is a very powerful analytical technique to be used in the structural characterization of the biomolecules giving chemical-specific information (i.e., for DNA, electrostatic, hydrophobic, and hydrogen bonding interactions between nucleotides can be reported).
- There is the possibility of acquiring SERS from aqueous solutions (biomolecules can be investigated in their natural environment).
- It produces a molecule-specific vibrational spectrum, (i.e., structural fingerprint of the DNA).

For all those reasons, SERS is considered a powerful technique for the DNA and RNA analysis. There are different strategies for SERS-based NA detection, which can be classified mainly into two groups. The large majority of strategies

rely on indirect approaches, such as those using SERS-encoded nanoparticles.[73] A second class of SERS methods is based on the label-free direct SERS analysis of NAs (i.e., the intrinsic SERS spectra of NAs is acquired). In the very recent years, an upsurge of interest in direct label-free SERS NAs detection has been observed by the efforts of different groups addressed to overcome the main limitation; the reproducibility issues associated with the NAs interaction with plasmonic metal colloids.[89]

1.3.1 Indirect SERS analysis of nucleic acids

The vast majority of SERS applications to large biomolecules, such as NAs detection, rely on the indirect sensing approach in which a Raman label is used for the SERS readout. Mostly, they are based on the selective recognition between two complementary strands to identify a specific sequence. The different chemical or physical properties of NA before and after hybridization are compared by the change on the Raman intensity label. Normally, oligonucleotides are modified with a thiol group to ensure the binding with the plasmonic surface and provide the selective biorecognition element to the complementary target sequences.[90] By the addition of a thiolated group, a covalent bond between surface and oligonucleotide is achieved (**Figure 1.15A**). This results in a more titled orientation of the strand, preventing the interaction of the bases with the metallic surface.

Particularly, Mirkin and co-workers[91] developed an indirect SERS strategy for NAs detection based on the idea of a microarray where immobilized sequences hybridize with target sequences attached to nanoparticles having a reporter (**Figure 1.15A**). The SERS signals were obtained from the dye (attached to the nanoparticle) and further enhancement was produced by a silver coating, achieving detection in the femtomolar regime.[91]

SERS detection of nucleic acids via indirect strategies has consistently proven high-throughput screening and multiplexing capabilities, enabling discrimination of a mixture of probes by the use of multiple Raman reporters.[92] Moreover, in comparison to traditional NAs analysis methods, it offers a lower detection limit and rapid and reliable detection in complex media.[93]

However, in this method is necessary the modification of the NAs with the Raman label which requires complex and costly chemistry. In addition, indirect strategies do not allow the acquisition of the chemical and structural-specific information of the NAs contained in the specific SERS spectral fingerprint of the target analyte.

1.3.2 Direct label-free SERS analysis of nucleic acids

Direct label-free SERS strategies are based on the detection of the NAs strands directly absorbed onto the nanostructured surface (**Figure 1.15B**). This methodology provides richer structural and chemical information, removing the

need of any NAs modification and shows outstanding potential in terms of sensitivity and selectivity.[94]

To date, however, direct SERS analysis still poses significant difficulties and challenges due to the several interdependent multiple factors that determine the degree of affinity of the molecule for the plasmonic substrate, as well as, the final property of the resulting surface complex. For this reason, direct acquisition of the SERS spectra of nucleic acids has historically suffered from limited reproducibility and/or sensitivity and mostly restricted to single-stranded DNA sequences.

In recent years, many efforts have been devoted to overcoming these limitations such as the development of novel SERS substrates with high and uniformly distributed enhancements[64, 95] or the implementation of multivariate statistical approaches in the spectral analysis.[96] Nevertheless, the main challenge for label-free detection is related to the surface of the SERS substrate. In fact, the metal surface chemistry largely defines the binding affinity of the target for the plasmonic substrate as well as the final structural features of the adsorbed molecule.[97]

One alternative to achieve the location of NAs near to the plasmonic surface is the use of SERS substrates where the solution of NAs is deposited.[64a, 94b] However, in that method the resulting SERS spectra are coming from dried oligonucleotides with the associated problems of reproducibility and constraining the NA structure.

A more exploited alternative is the use of nanoparticles in solution. In that case, many issues have to be considered. Firstly, the colloids usually have a layer of surfactant or capping agent that provides stability after the synthesis. Those compounds are usually unstable leading to variations in SERS signals shown as impurities. Secondly, the metal surface is negatively charged and considering that NAs are surrounded by phosphate groups (negatively charged) the electrostatic interaction between them would be very weak.

In order to overcome such limitations, many strategies have been developed focused on the achievement of a reproducible and reliable technique for label-free SERS NAs analysis. Most of them employ an aggregating agent in order to obtain higher SERS intensity coming from the hot-spots generated by the aggregates.[98] On the other hand, there is the possibility of tuning the surface chemistry of the colloidal particles to improve the affinity for nucleic acids.[99] An extensive discussion of different strategies using colloids for label-free NA detection is exposed on chapter 2.

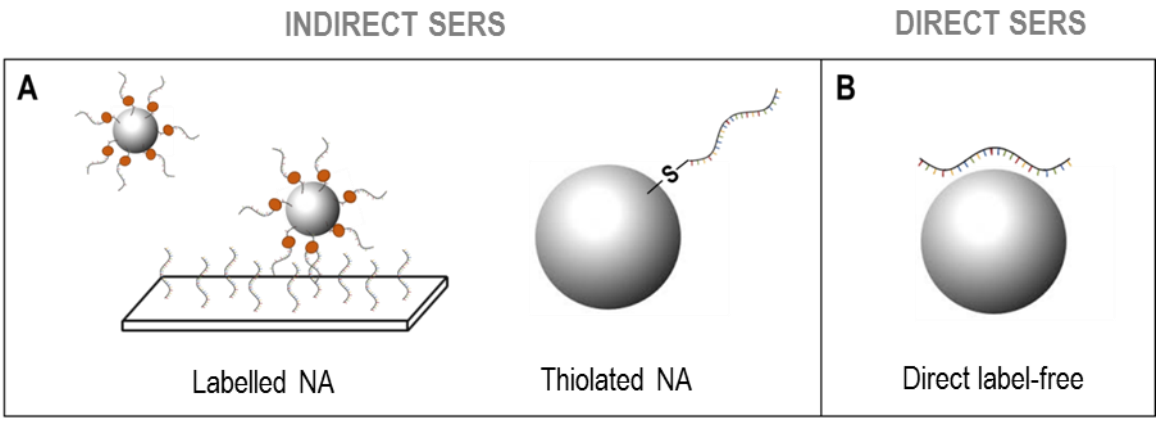


Figure 1.14 Representation of the strategies employed for NA detection by SERS. A) An example of an indirect SERS strategy in which the signal is obtained from the label or their associated change. Thiol modification is widely used to achieve the attachment of the NA to metallic surface. B) Representation of the direct label-free NA detection using nanoparticles.

REFERENCES

[1] G. Cooper, *The Cell: A Molecular Approach*, Sunderland (MA), **2000**.

[2] J. D. Watson, F. H. C. Crick, *Clin. Orthop. Rel. Res.* **2007**, 3-5.

[3] a) R. E. Franklin, R. G. Gosling, *Nature* **1953**, *171*, 740-741; b) M. H. F. Wilkins, A. R. Stokes, H. R. Wilson, *Nature* **1953**, *171*, 738-740.

[4] E. Chargaff, *Experientia* **1950**, *6*, 201-209.

[5] M. Meselson, F. W. Stahl, *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 671-682.

[6] a) D. P. Martin, P. Biagini, P. Lefeuvre, M. Golden, P. Roumagnac, A. Varsani, *Viruses-Basel* **2011**, *3*, 1699-1738; b) F. E. D. Lima, S. P. Cibulski, H. F. dos Santos, T. F. Teixeira, A. P. M. Varela, P. M. Roehe, E. Delwart, A. C. Franco, *PLoS One* **2015**, *10*.

[7] V. A. Bloomfield, D. M. Crothers, I. Tinoco, *Nucleic Acids. Structures, Properties, and Functions*, Suasalito, CA, **2000**.

[8] J. D. Watson, T. A. Baker, M. Gann, M. Levine, R. Losick, *Molecular Biology of the Gene*, Glenview, **2014**.

[9] W. Gilbert, *Nature* **1986**, *319*.

[10] a) S. Clancy, *Nature Education* **2008**, *1*, 60; b) L. He, J. G. Hannon, *Nat. Rev. Genet.* **2004**, *5*, 522-531.

[11] F. Crick, *Nature* **1970**, *227*, 561-&.

[12] a) P. D. Yin, W. S. Hu, *Journal of Virology* **1997**, *71*, 6237-6242; b) N. Bazhanov, O. Escaffre, A. N. Freiberg, R. P. Garofalo, A. Casola, *Scientific Reports* **2017**, *7*.

[13] N. Vaidya, N. Lehman, *Nucleic Acids Res.* **2009**, *37*, 3981-3989.

[14] D. P. Bartel, *Cell* **2004**, *116*, 281-297.

[15] R. F. Gesteland, T. R. Cech, J. F. Atkins, in *Rna World, Third Edition, Vol. 43* (Eds.: R. F. Gesteland, T. R. Cech, J. F. Atkins), **2006**, pp. IX-X.

[16] a) J. S. Mattick, *Embo J.* **2012**, *31*, 515-516; b) S. Y. Ng, R. Johnson, L. W. Stanton, *Embo J.* **2012**, *31*, 522-533; c) A. M. Khalil, M. Guttman, M. Huarte, M. Garber, A. Raj, D. R. Morales, K. Thomas, A. Presser, B. E. Bernstein, A. van Oudenaarden, A. Regev, E. S. Lander, J. L. Rinn, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11667-11672.

[17] H. F. Dong, J. P. Lei, L. Ding, Y. Q. Wen, H. X. Ju, X. J. Zhang, *Chem. Rev.* **2013**, *113*, 6207-6233.

[18] H. Rohrig, J. Schmidt, E. Miklashevichs, J. Schell, M. John, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1915-1920.

[19] M. M. C. D.L. Nelson, *Lehninger Principles of Biochemistry*, Barcelona, **2009**.

[20] T. Carell, C. Brandmayr, A. Hienzsch, M. Muller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem.-Int. Edit.* **2012**, *51*, 7110-7131.

[21] a) E. Chargaff, B. Magasanik, R. Doniger, E. Vischer, *Journal of the American Chemical Society* **1949**, *71*, 1513-1514; b) E. Chargaff, B. Magasanik, E. Vischer, C. Green, R. Doniger, D. Elson, *J.*

Biol. Chem. **1950**, *186*, 51-67; c) B. Magasanik, E. Vischer, R. Doniger, D. Elson, E. Chargaff, *J. Biol. Chem.* **1950**, *186*, 37-50.

[22] P. Yakovchuk, E. Protozanova, M. D. Frank-Kamenetskii, *Nucleic Acids Res.* **2006**, *34*, 564-574.

[23] a) A. Brunet, C. Tardin, L. Salome, P. Rousseau, N. Destainville, M. Manghi, *Macromolecules* **2015**, *48*, 3641-3652; b) S. Rao, S. Raj, B. Cossins, M. Marro, V. Guallar, D. Petrov, *Biophys. J.* **2013**, *104*, 156-162.

[24] A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. Vanboom, G. Vandermarel, A. Rich, *Nature* **1979**, *282*, 680-686.

[25] F. DiMaio, X. Yu, E. Rensen, M. Krupovic, D. Prangishvili, E. H. Egelman, *Science* **2015**, *348*, 914-917.

[26] A. Rich, *J. Biol. Chem.* **1996**, *271*, 11595-11598.

[27] S. C. Harvey, *J. Struct. Biol.* **2015**, *189*, 1-8.

[28] W. P. Wahls, L. J. Wallace, P. D. Moore, *Mol. Cell. Biol.* **1990**, *10*, 785-793.

[29] P. B. Moore, *Annu. Rev. Biochem.* **1999**, *68*, 287-300.

[30] G. Egger, G. N. Liang, A. Aparicio, P. A. Jones, *Nature* **2004**, *429*, 457-463.

[31] S. Balbo, R. J. Turesky, P. W. Villalta, *Chemical research in toxicology* **2014**, *27*, 356-366.

[32] A. Portela, M. Esteller, *Nat. Biotechnol.* **2010**, *28*, 1057-1068.

[33] T. Kouzarides, *Cell* **2007**, *128*, 693-705.

[34] O. J. Rando, H. Y. Chang, in *Annu. Rev. Biochem., Vol. 78*, **2009**, pp. 245-271.

[35] A. Zemach, I. E. McDaniel, P. Silva, D. Zilberman, *Science* **2010**, *328*, 916-919.

[36] F. Capuano, M. Mulleder, R. Kok, H. J. Blom, M. Ralser, *Anal. Chem.* **2014**, *86*, 3697-3702.

[37] X. M. He, S. H. Chang, J. J. Zhang, Q. Zhao, H. Z. Xiang, K. Kusonmano, L. Yang, Z. S. Sun, H. M. Yang, J. Wang, *Nucleic Acids Res.* **2008**, *36*, D836-D841.

[38] a) M. Tahiliani, K. P. Koh, Y. H. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930-935; b) M. Munzel, D. Globisch, T. Bruckl, M. Wagner, V. Welzmiller, S. Michalakis, M. Muller, M. Biel, T. Carell, *Angew. Chem.-Int. Edit.* **2010**, *49*, 5375-5377.

[39] M. A. Palafox, V. K. Rastogi, S. Kumar, H. Joe, *Spectroc. Acta Pt. A-Molec. Biomolec. Spectr.* **2013**, *111*, 104-122.

[40] a) M. Whiteman, A. Jenner, B. Halliwell, *Chemical research in toxicology* **1997**, *10*, 1240-1246; b) M. D. Evans, M. Dizdaroglu, M. S. Cooke, *Mutat. Res.-Rev. Mutat. Res.* **2004**, *567*, 1-61.

[41] J. Liu, G. F. Jia, *J. Genet. Genomics* **2014**, *41*, 21-33.

[42] a) M. Lee, B. Kim, V. N. Kim, *Cell* **2014**, *158*, 980-987; b) G. F. Jia, Y. Fu, X. Zhao, Q. Dai, G. Q. Zheng, Y. Yang, C. Q. Yi, T. Lindahl, T. Pan, Y. G. Yang, C. He, *Nat. Chem. Biol.* **2011**, *7*, 885-887.

[43] N. Tretyakova, P. W. Villalta, S. Kotapati, *Chem. Rev.* **2013**, *113*, 2395-2436.

[44] C. Badouard, M. Masuda, H. Nishino, J. Cadet, A. Favier, J. L. Ravanat, *J. Chromatogr. B* **2005**, *827*, 26-31.

[45] A. C. Beach, R. C. Gupta, *Carcinogenesis* **1992**, *13*, 1053-1074.

[46] M. L. Metzker, *Nat. Rev. Genet.* **2010**, *11*, 31-46.

[47] S. Griffiths-Jones, H. K. Saini, S. van Dongen, A. J. Enright, *Nucleic Acids Res.* **2008**, *36*, D154-D158.

[48] P. A. Sharp, *Cell* **2009**, *136*, 577-580.

[49] Y. Liang, D. Ridzon, L. Wong, C. F. Chen, *BMC Genomics* **2007**, *8*.

[50] A. K. L. Leung, P. A. Sharp, *Mol. Cell* **2010**, *40*, 205-215.

[51] S. M. Eacker, T. M. Dawson, V. L. Dawson, *Nat. Rev. Neurosci.* **2009**, *10*, 837-841.

[52] A. K. Pandey, P. Agarwal, K. Kaur, M. Datta, *Cell. Physiol. Biochem.* **2009**, *23*, 221-232.

[53] R. Garzon, G. A. Calin, C. M. Croce, in *Annual Review of Medicine, Vol. 60*, Annual Reviews, Palo Alto, **2009**, pp. 167-179.

[54] L. P. Lim, N. C. Lau, E. G. Weinstein, A. Abdelhakim, S. Yekta, M. W. Rhoades, C. B. Burge, D. P. Bartel, *Genes Dev.* **2003**, *17*, 991-1008.

[55] K. A. Cissell, S. Shrestha, S. K. Deo, *Anal. Chem.* **2007**, *79*, 4754-4761.

[56] A. J. Qavi, J. T. Kindt, R. C. Bailey, *Anal. Bioanal. Chem.* **2010**, *398*, 2535-2549.

[57] S. Streit, C. W. Michalski, M. Erkan, J. Kleeff, H. Friess, *Nat. Protoc.* **2009**, *4*, 37-43.

[58] *Essentials of Nucleic Acid Analysis: A Robust Approach*, Royal Soc Chemistry, Cambridge, **2008**.

[59] D. M. Duan, K. X. Zheng, Y. Shen, R. Cao, L. Jiang, Z. X. Lu, X. Y. Yan, J. Li, *Nucleic Acids Res.* **2011**, *39*.

[60] A. Gaur, D. A. Jewell, Y. Liang, D. Ridzon, J. H. Moore, C. F. Chen, V. R. Ambros, M. A. Israel, *Cancer Res.* **2007**, *67*, 2456-2468.

[61] M. Wanunu, T. Dadosh, V. Ray, J. M. Jin, L. McReynolds, M. Drndic, *Nat. Nanotechnol.* **2010**, *5*, 807-814.

[62] E. Crew, S. Rahman, A. Razzak-Jaffar, D. Mott, M. Kamundi, G. Yu, N. Tchah, J. Lee, M. Bellavia, C. J. Zhong, *Anal. Chem.* **2012**, *84*, 26-29.

[63] A. H. Alhasan, D. Y. Kim, W. L. Daniel, E. Watson, J. J. Meeks, C. S. Thaxton, C. A. Mirkin, *Anal. Chem.* **2012**, *84*, 4153-4160.

[64] a) J. D. Driskell, A. G. Seto, L. P. Jones, S. Jokela, R. A. Dluhy, Y. P. Zhao, R. A. Tripp, *Biosens. Bioelectron.* **2008**, *24*, 917-922; bJ. D. Driskell, R. A. Tripp, *Chemical Communications* **2010**, *46*, 3298-3300.

[65] a) E. C. Le Ru, P. G. Etchegoin, in *Principles of Surface-Enhanced Raman Spectroscopy*, Elsevier, Amsterdam, **2009**, pp. 29-120; b) E. C. Le Ru, P. G. Etchegoin, in *Principles of Surface-Enhanced Raman Spectroscopy*, Elsevier, Amsterdam, **2009**, pp. 121-183.

[66] A. Smekal, *Naturwissenschaften* **1923**, *11*, 873-875.

[67] C. V. Raman, K. S. Krishnan, *Nature* **1928**, *121*, 501-502.

[68] J. W. Strutt, *Philosophical Magazine Series 4* **1871**, *41*, 107-120.

[69] E. C. Le Ru, P. G. Etchegoin, in *Principles of Surface-Enhanced Raman Spectroscopy*, Elsevier, Amsterdam, **2009**, pp. 185-264.

[70] a) K. Kneipp, Y. Wang, H. Kneipp, L. T. Perelman, I. Itzkan, R. Dasari, M. S. Feld, *Phys. Rev. Lett.* **1997**, *78*, 1667-1670; b) S. M. Nie, *Abstr. Pap. Am. Chem. Soc.* **2001**, *221*, U244-U244.

[71] R. F. Aroca, R. A. Alvarez-Puebla, N. Pieczonka, S. Sanchez-Cortez, J. V. Garcia-Ramos, *Adv. Colloid Interface Sci.* **2005**, *116*, 45-61.

[72] R. A. Álvarez-Puebla, *The Journal of Physical Chemistry Letters* **2012**, *3*, 857-866.

[73] S. Schlücker, *Angewandte Chemie International Edition* **2014**, *53*, 4756-4795.

[74] E. C. Le Ru, P. G. Etchegoin, *Chem. Phys. Lett.* **2006**, *423*, 63-66.

[75] a) N. Pazos-Perez, S. Barbosa, L. Rodriguez-Lorenzo, P. Aldeanueva-Potel, J. Perez-Juste, I. Pastoriza-Santos, R. A. Alvarez-Puebla, L. M. Liz-Marzan, *J. Phys. Chem. Lett.* **2010**, *1*, 24-27; b) J. Morla-Folch, L. Guerrini, N. Pazos-Perez, R. Arenal, R. A. Alvarez-Puebla, *ACS Photonics* **2014**, *1*, 1237-1244.

[76] E. Hao, G. C. Schatz, *J. Chem. Phys.* **2004**, *120*, 357-366.

[77] R. Aroca, *Surface-enhanced Vibrational Spectroscopy*, Chichester, **2006**.

[78] a) E. C. Le Ru, P. G. Etchegoin, in *Principles of Surface-Enhanced Raman Spectroscopy*, Elsevier, Amsterdam, **2009**, pp. 367-413; b) M. J. Natan, *Faraday Discuss.* **2006**, *132*, 321-328.

[79] a) Q. Zhang, Y. H. Lee, I. Y. Phang, C. K. Lee, X. Y. Ling, *Small* **2014**, *10*, 2703-2711; b) A. Torres-Nunez, K. Faulds, D. Graham, R. A. Alvarez-Puebla, L. Guerrini, *Analyst* **2016**, *141*, 5170-5180.

[80] G. Rusciano, A. C. De Luca, G. Pesce, A. Sasso, G. Oliviero, J. Amato, N. Borbone, S. D'Errico, V. Piccialli, G. Piccialli, L. Mayol, *Anal. Chem.* **2011**, *83*, 6849-6855.

[81] H. Y. Jung, Y. K. Park, S. Park, S. K. Kim, *Analytica Chimica Acta* **2007**, *602*, 236-243.

[82] S. Abalde-Cela, P. Aldeanueva-Potel, C. Mateo-Mateo, L. Rodriguez-Lorenzo, R. A. Alvarez-Puebla, L. M. Liz-Marzan, *Journal of The Royal Society Interface* **2010**, *7*, S435-S450.

[83] R. A. Álvarez-Puebla, R. Contreras-Cáceres, I. Pastoriza-Santos, J. Pérez-Juste, L. M. Liz-Marzán, *Angewandte Chemie International Edition* **2009**, *48*, 138-143.

[84] R. A. Alvarez-Puebla, R. F. Aroca, *Anal. Chem.* **2009**, *81*, 2280-2285.

[85] a) L. Mikac, M. Ivanda, M. Gotic, T. Mihelj, L. Horvat, *Journal of Nanoparticle Research* **2014**, *16*; b) K. M. M. Abou El-Nour, A. a. Eftaiha, A. Al-Warthan, R. A. A. Ammar, *Arabian Journal of Chemistry* **2010**, *3*, 135-140.

[86] J. Kneipp, H. Kneipp, B. Wittig, K. Kneipp, *Nano Letters* **2007**, *7*, 2819-2823.

[87] E. C. Le Ru, P. G. Etchegoin, *Principles of Surface-Enhanced Raman Spectroscopy*, **2009**.

[88] E. C. Le Ru, P. G. Etchegoin, in *Principles of Surface-Enhanced Raman Spectroscopy*, Elsevier, Amsterdam, **2009**, pp. 299-365.

[89] L. Guerrini, L. Barrett, J. A. Dougan, K. Faulds, D. Graham, *Nanoscale* **2013**, *5*, 4166-4170.

[90] H. I. Peng, B. L. Miller, *Analyst* **2011**, *136*, 436-447.

[91] Y. W. C. Cao, R. C. Jin, C. A. Mirkin, *Science* **2002**, *297*, 1536-1540.

[92] D. Graham, B. J. Mallinder, W. E. Smith, *Angew. Chem.-Int. Edit.* **2000**, *39*, 1061-+.

[93] J. A. Dougan, K. Faulds, *Analyst* **2012**, *137*, 545-554.

[94] a) A. Barhoumi, D. Zhang, F. Tam, N. J. Halas, *Journal of the American Chemical Society* **2008**, *130*, 5523-5529; b) A. Barhoumi, N. J. Halas, *J. Phys. Chem. Lett.* **2011**, *2*, 3118-3123; c) S. R. Panikkanvalappil, M. A. Mackey, M. A. El-Sayed, *Journal of the American Chemical Society* **2013**, *135*, 4815-4821.

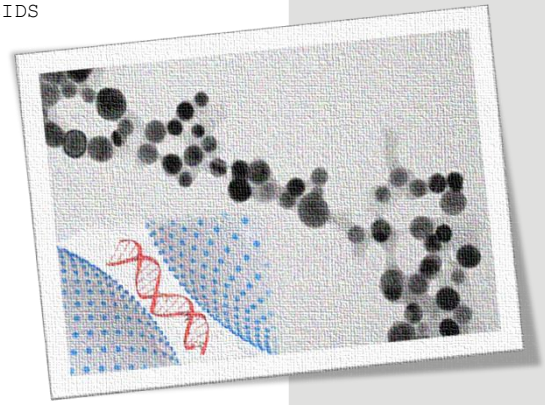
[95] E. Prado, A. Colin, L. Servant, S. Lecomte, *J. Phys. Chem. C* **2014**, *118*, 13965-13971.

[96] J. L. Abell, J. M. Garren, J. D. Driskell, R. A. Tripp, Y. P. Zhao, *Journal of the American Chemical Society* **2012**, *134*, 12889-12892.

[97] L. Guerrini, Z. Jurasekova, C. Domingo, M. Pérez-Méndez, P. Leyton, M. Campos-Vallette, J. V. Garcia-Ramos, S. Sanchez-Cortes, *Plasmonics* **2007**, *2*, 147-156.

[98] E. Papadopoulou, S. E. J. Bell, *Chem.-Eur. J.* **2012**, *18*, 5394-5400.

[99] L. Guerrini, Z. Krpetic, D. van Lierop, R. A. Alvarez-Puebla, D. Graham, *Angew. Chem.-Int. Edit.* **2015**, *54*, 1144-1148.



CHAPTER 2

POSITIVELY CHARGED NANOPARTICLES FOR DIRECT SERS ANALYSIS OF NA

ABSTRACT

Direct SERS analysis of NAs has been largely performed using negatively charged nanoparticles as SERS substrates. However, its applicability has traditionally suffered from reproducibility and/or sensitivity issues. Recently, it has been shown that the employment of positively charged silver colloids coated with spermine (AgNP@Sp) provides several key advantages. Spermine molecules anchored at the surface of the nanoparticles ensure the electrostatic attraction with the negatively charged phosphate backbone of the NA, leading to the generation of stable aggregates in suspension. Here, the NA sequences remain trapped at interparticle junctions (hot spots). In this chapter, it is described the synthesis of AgNP@Sp, their characterization and, finally, the optimization of the experimental conditions to maximize the performance of these colloids as a SERS platform for direct NA analysis.

OUTLINE

2.1 INTRODUCTION

Negatively charged silver colloids have been the most widely used substrates for the analysis of NAs in direct SERS.[1] Here, nucleotides preferably interact with the metallic surface via their exocyclic amine moiety or heterocyclic nitrogen,[2] adopting a flattened configuration.[3] Normally, the resulting SERS spectra appear to be dominated by adenine modes because of adenine has a stronger scattering cross-section than the other bases[3] and highest binding affinity for silver and gold (mainly via the N7 nitrogen atom).[2] This mechanism of interaction significantly restricted the SERS analysis to ssDNA since in the duplex form, the nucleotides are hidden with the double-helix, leaving exposed the negative charged phosphate groups which display low affinity to such colloids.[4] Further, aggregating agents are normally employed to promote nanoparticle aggregation, therefore generating intense SERS signal. Particularly, magnesium sulfate ($MgSO_4$) is extensively used[1a-c] as a “passive” aggregating agent. The addition of $MgSO_4$ produces an increment of the ionic concentration facilitating nanoparticle aggregation and thus, enhancing the SERS signal.[1b] In other studies, Mg^{2+} ions are also described as neutralizers of the DNA phosphate backbone that mediate the electrostatic interaction with iodide-modified silver nanoparticles.[1c]

Nonetheless, the use of external agents to induce the aggregation of the nanoparticles leads to poor or null control over both the overall nanoparticle assembly and the final position of the DNA within the aggregate. Moreover, as it was investigated by Torres-Nuñez and co-workers,[5] when negative NPs are

aggregated with $MgSO_4$, SERS results more dependent on nucleobase composition and nanoparticle surface chemistry than when no aggregating agent is used. Thus, removal of the external aggregating agent highly simplifies the sensing scheme and could allow higher reproducibility.

El-Sayed and co-workers[6] efficiently circumvented the issues associated with the use of an external aggregating agent by using highly concentrated colloids. Here, conventional citrate-capped silver nanospheres were used at a high concentration in order to physically retain the DNA within the interparticle junctions without the requirement of an aggregating agent.[6] However, this approach is only effective when the concentration of DNA is relatively high (i.e., ca. 1 mg/mL).

Furthermore, other strategies have been exploited to overcome the electrostatic repulsion between the phosphate groups of the NAs, and metal surface, such as the addition of small positively charged molecules. Graham *et al.*[7] described the employment of spermine as an effective agent for the controlled aggregation of the colloids. Spermine acts as an NA charge modifier, it interacts with the NA balancing the negatively charged phosphate groups.[7] However, spermine itself can induce agglomeration of the DNA and can lead to over aggregation of the system, compromising their stability.[8]

An alternative approach for the NA detection is the use of positively charged colloids capable of directly interacting with the phosphate backbone of DNA via electrostatic interactions. One example was published by Gill *et al.*[9] using

hexadecyltrimethylammonium bromide (CTAB) capped silver nanoparticles to ensure the effective absorption of dye-labeled DNA onto the surface and, thus, efficient SERRS signals can be produced.[9] Other positively charged polymers employed are for instance branched polyethyleneimine (PEI) chains[10] or poly-L-Lysine (PLL).[11] Although the presence of a polymer or surfactant improves colloidal stability and affinity towards DNA, a dramatic limitation for direct SERS analysis is imposed by the intrinsic thickness of the surface layers (i.e., DNA absorbs relatively far from the metal). In addition, high background signals are likely to be observed for polymer coated nanoparticles.

Novel synthesis of spermine-coated positively charged silver nanoparticles (AgNP@Sp) was reported by V. Lierop *et al.*[12] Spermine molecules (**Figure 2.1**) employed during the preparation of the colloids remained bound to the NP surface, providing an overall positive charge that stabilizes the NPs in suspension via electrostatic repulsions. Thus, upon addition of NA, the sequences rapidly adhere on the NPs through their phosphate backbone.[12]

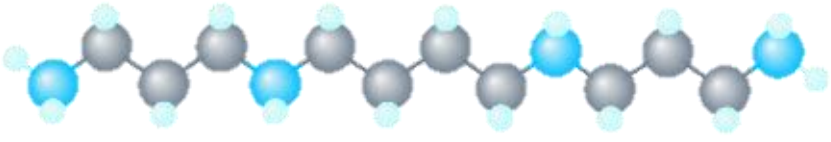


Figure 2.1 Three-dimensional arrangement of the atoms that constitute spermine molecule. Grey = Carbon atom, Light Blue = Hydrogen atom and Blue = Nitrogen atom. *Obtained from: https://pubchem.ncbi.nlm.nih.gov/compound/spermine*

As a result, a fast nanoparticle aggregation (**Figure 2.2**) takes place where NPs form DNA-mediated clusters[13] which entrap the NA sequences at the inter-particle junctions (where the SERS signal is maximized) leading to intense SERS spectra with high spectral reproducibility.

In this system, NAs has double key features; firstly acts as an aggregating agent of individual nanoparticles, secondly, develops the role as an effective stabilizing agent of the so-formed clusters. The advantages of the NAs analysis using AgNP@ can be summarized as follows;

- Absorption onto the nanoparticles surface is achieved due to the electrostatic interaction between the secondary amines of spermine ligands and the negatively charged phosphate groups of the NAs.
- The nanoparticle aggregation is induced by the NAs itself, removing the need for the addition of an external aggregating agent.
- High-quality SERS signals and extremely sensitive are obtained due to the location of the NAs sequences; near to the plasmonic surface and located at the inter-particle hotspots of the aggregates.

Because of NAs molecules act as an ‘electrostatic glue’ for AgNP@Sp, illumination of the colloidal suspension with a 532 nm laser yields highly intense and reproducible averaged SERS spectra of NA at nanograms per milliliter level[13-14], while integration with microfluidic devices could reduce the required amount of sample to picogram levels.[15]

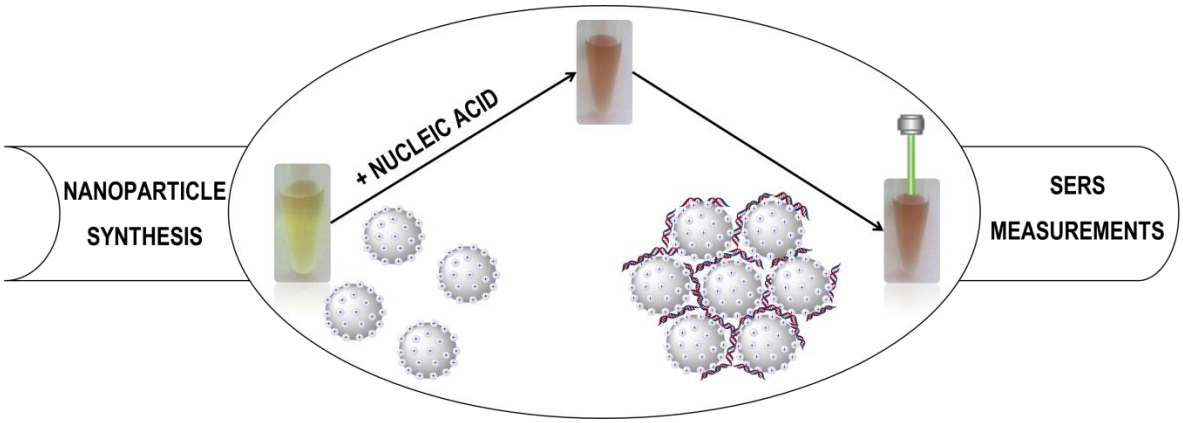


Figure 2.1 General scheme of NA detection using AgNP@Sp (the aggregation is driven by the NA itself). The change of colour is because of the transition from isolated colloids to stable clusters.

2.2 SYNTHESIS OF AgNP@Sp

Synthesis of positively-charged spermine coated silver nanoparticles (AgNP@Sp) was performed following the previously reported protocol.[12, 14] Briefly, 20 µL of a 0.5 M $AgNO_3$ solution were added to 10 mL of Milli-Q water, followed by addition of 7 µL of a 0.1 M aqueous spermine tetrahydrochloride. After 20 min of stirring, 250 µL of $NaBH_4$ (0.01 M) were rapidly injected into the mixture under vigorous stirring. The solution was gently stirred for another 10 min. Before being used, colloids were left to age overnight and the visible sediment at the bottom of the vial was eliminated from the sample. Ultraviolet-visible (UV-Vis) spectra before and after aging are shown in **Figure 2.3**.

Importantly, in order to prevent the electrostatic deposition of the positive nanoparticles onto the glassware walls (negatively charged), both glass vials and the magnetic stirring bars were pre-coated with a positively charged polymer

(polyethyleneimine (PEI)), as described elsewhere.[13] All material was dipped into a PEI solution 0.4% (v/v) for at least 2 h and, after that, extensively washed with Milli-Q water to remove any unbound PEI molecules.

The synthesis yields quasi-spherical silver nanoparticles of ca. 25nm diameter, (fully characterization is presented below). The bulk pH of the colloids is ~ 6, and the final nanoparticle concentration is ca. 0.3 nM, calculated according to the Lambert-Beer's law by using an extinction coefficient of 54.8×10^8 M^{-1} cm^{-1} obtained from the literature.[16]

2.3 CHARACTERIZATION

After the synthesis, the optical and physical properties of colloids were characterized using UV-Vis spectroscopy, Transmission Electron microscopy, SERS and ζ-potential.

UV-Vis spectroscopy

The measurement of the UV-Vis excitation spectra is one of the most common characterization optical techniques for SERS substrates. The main information extracted is associated with the response of the localized surface plasmon resonance (LSPR). From the characteristics of their surface resonance and in comparison with theoretical predictions, parameters such as polydispersity, uniformity or the concentration can be extracted.[17]

Figure 2.3A shows the direct optical absorption measurements of the AgNP@Sp recently synthesized and after 24 hours aging. Colloids show a LSPR centred at ~ 391 nm and the difference between both spectra is appreciable. After the synthesis, the spectrum is broader indicating a certain degree of polydispersity (different sized silver nanoparticles or aggregates could be found in solution).

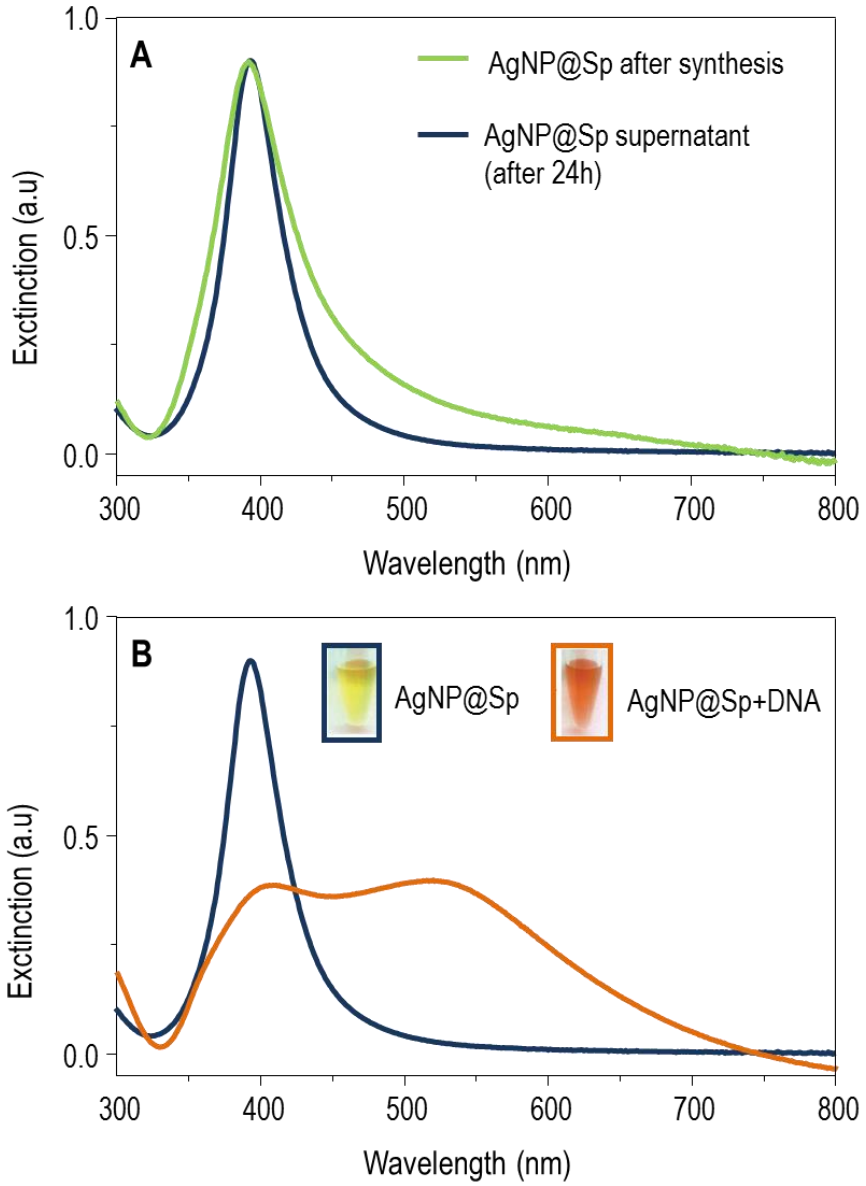


Figure 2.2 Extinction spectra of AgNP@Sp A) before and after aging and B) before and after the addition of a sequence of DNA (single-stranded) when they became aggregate. UV-Vis spectra were recorded using a Thermo Scientific Evolution 201 UV-visible spectrophotometer.

Upon overnight aging (**Figure 2.3A**), sediment appears at the bottom of the vial which is removed, leading to a narrower spectrum with a well-defined peak (i.e., higher monodispersity).

NA-induced aggregation of AgNP@Sp is well-reflected in the extinction spectra (**Figure 2.3B**). Upon NA addition, the LSPR band decreases in intensity and becomes broader, while a second large feature emerges at longer wavelength due of the formation of aggregates in solution.

Transmission Electron Microscopy (TEM)

Transmission electron microscopy (TEM) is the characterization technique that provides direct measurement of the nanoparticle morphology and size.[17-18] AgNP@Sp was characterized with TEM mainly because it can produce images with higher magnification and greater resolution than other microscopic techniques such as the Scanning Electron Microscopy (SEM).[19]

The colloidal solution was dried onto a copper grid coated with a thin layer of carbon. The corresponding TEM images (**Figure 2.4**) show quasi-spherical silver nanoparticles of ~ 25 nm diameter. In order to obtain a more accurate approximation in terms of size, a statistics on size was performed by measuring the diameter of N=150 nanoparticles from different TEM pictures, the results are represented in the form of a histogram in **Figure 2.4E**. The average size obtained is 24.7 ± 4.5 nm.

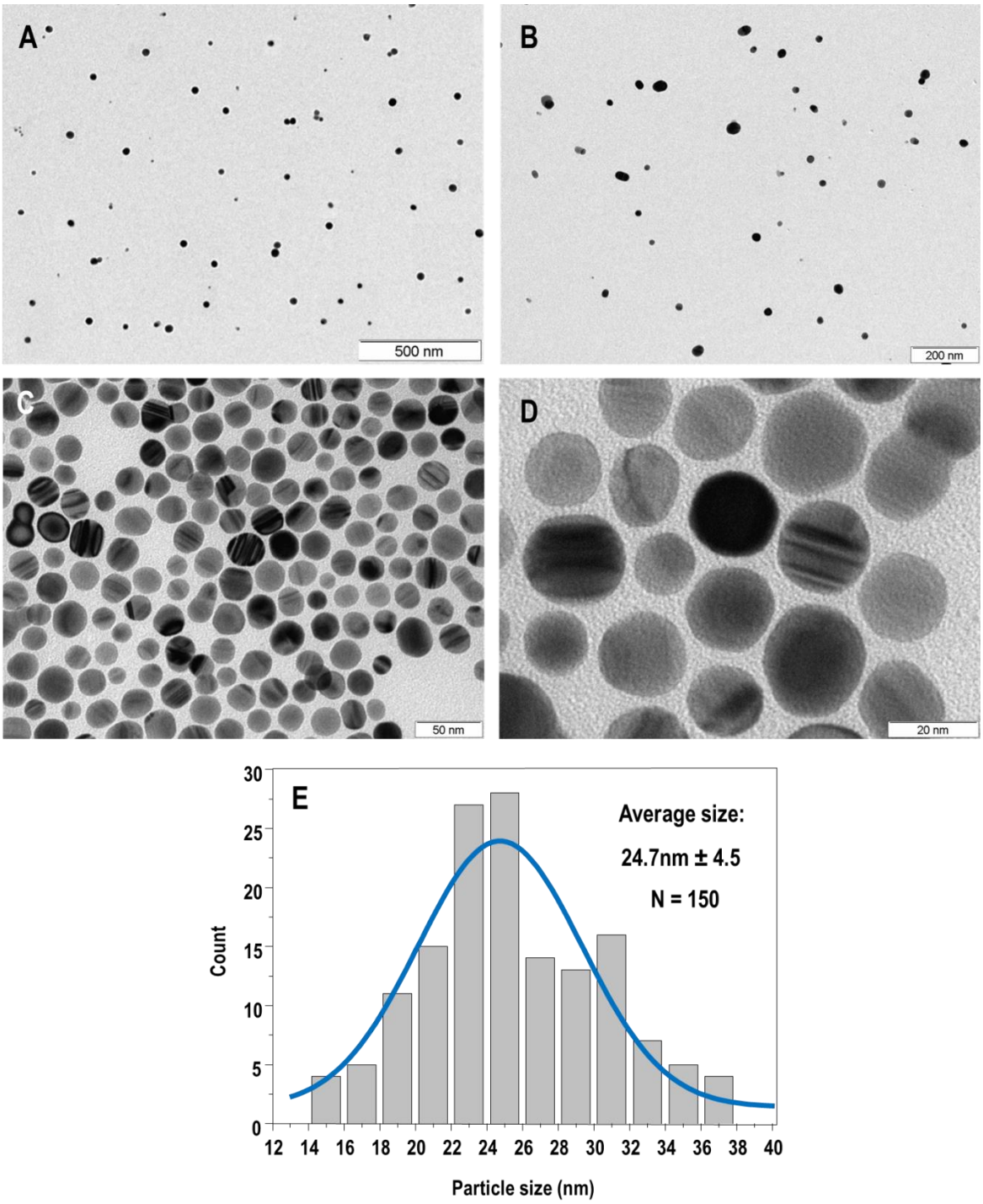


Figure 2.3 A-D) Representative TEM images of AgNP@Sp colloids from lower to higher magnification. TEM Samples were prepared by drop casting 5 or 10 µL of the colloidal suspension onto TEM grids. TEM analysis was performed on the dried sample. E) The corresponding histogram of nanoparticle diameters (24.7 ± 4.5 nm). TEM was performed with a JEOL JEM-1011 transmission electron microscope.

As previously shown, the mixture of NAs with AgNP@Sp induces a rapid aggregation that is clearly demonstrated by their UV-Vis extinction spectrum (**Figure 2.3B**). In order to visualize such phenomenon by TEM, images before and after the addition of a DNA sequence are compiled in **Figure 2.5**. In order to perform a fair comparison, AgNP@Sp was also aggregated with $MgSO_4$. In the corresponding TEM images of DNA-mediated clusters is visible the halo of DNA corona surrounding the nanoparticle aggregates.

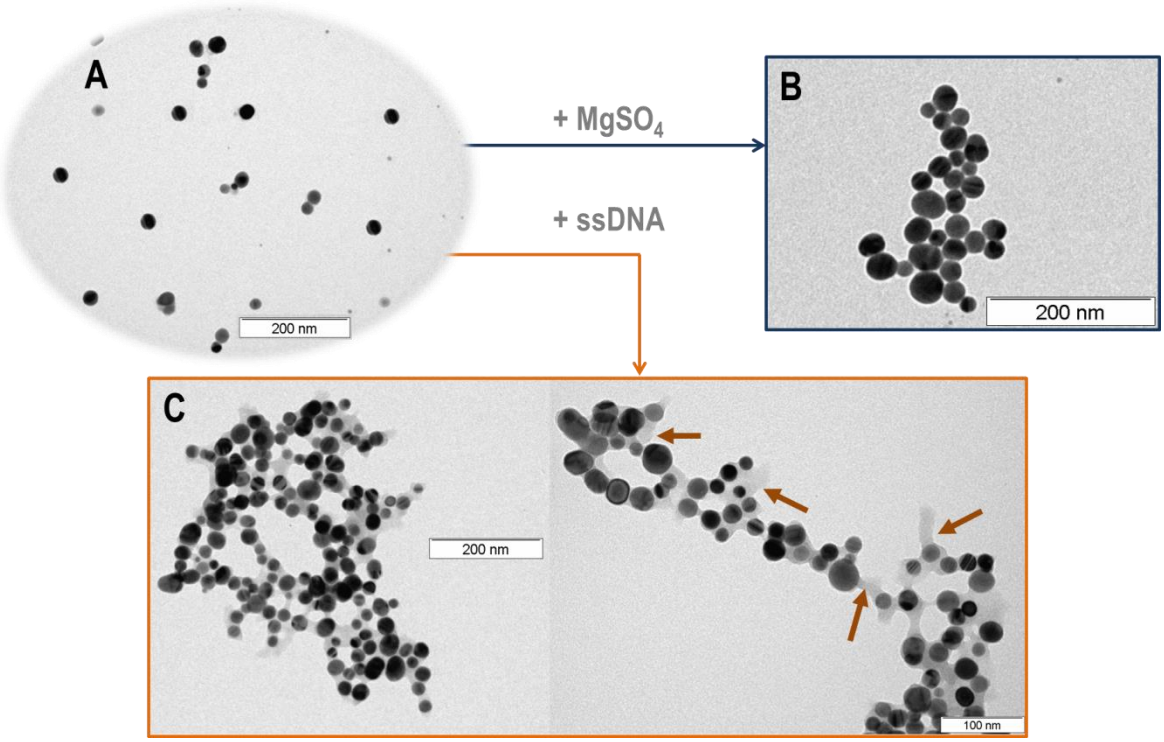


Figure 2.5 TEM images of AgNP@Sp colloids A) before addition of ssDNA, B) after mixed with $MgSO_4$ and, C) the corresponding TEM image of ssDNA-mediated clusters.

SERS spectroscopy

Surface-enhanced Raman scattering (SERS) spectroscopy is a powerful technique for providing information about the surface chemistry of the nanoparticles.[20] Thus, SERS was used as a characterization technique to acquire the spectral background signals of AgNP@Sp colloids.

To do so, a small volume of AgNP@Sp was mixed with an aqueous solution of $MgSO_4$ in order to induce nanoparticle aggregation and, thus, provide detectable SERS signals. $MgSO_4$ is a passive electrolyte that induces aggregation by increasing the ionic concentration without firmly absorbing onto the silver surface.[17] The cluster suspensions were illuminated using a green laser (532 nm) and a long-working distance objective (0.17 NA, working distance 30 mm) focusing the laser on the sample (setup is shown in **Figure 2.6A**).

The spectrum of the background SERS signals obtained from that suspension is illustrated in **Figure 2.6B**. Importantly, AgNP@Sp show weak SERS background in the spectral region of interest for NAs, thus removing the need of a digital subtraction from the SERS spectra of the analysed molecules, which can introduce spectral reproducibility issues.

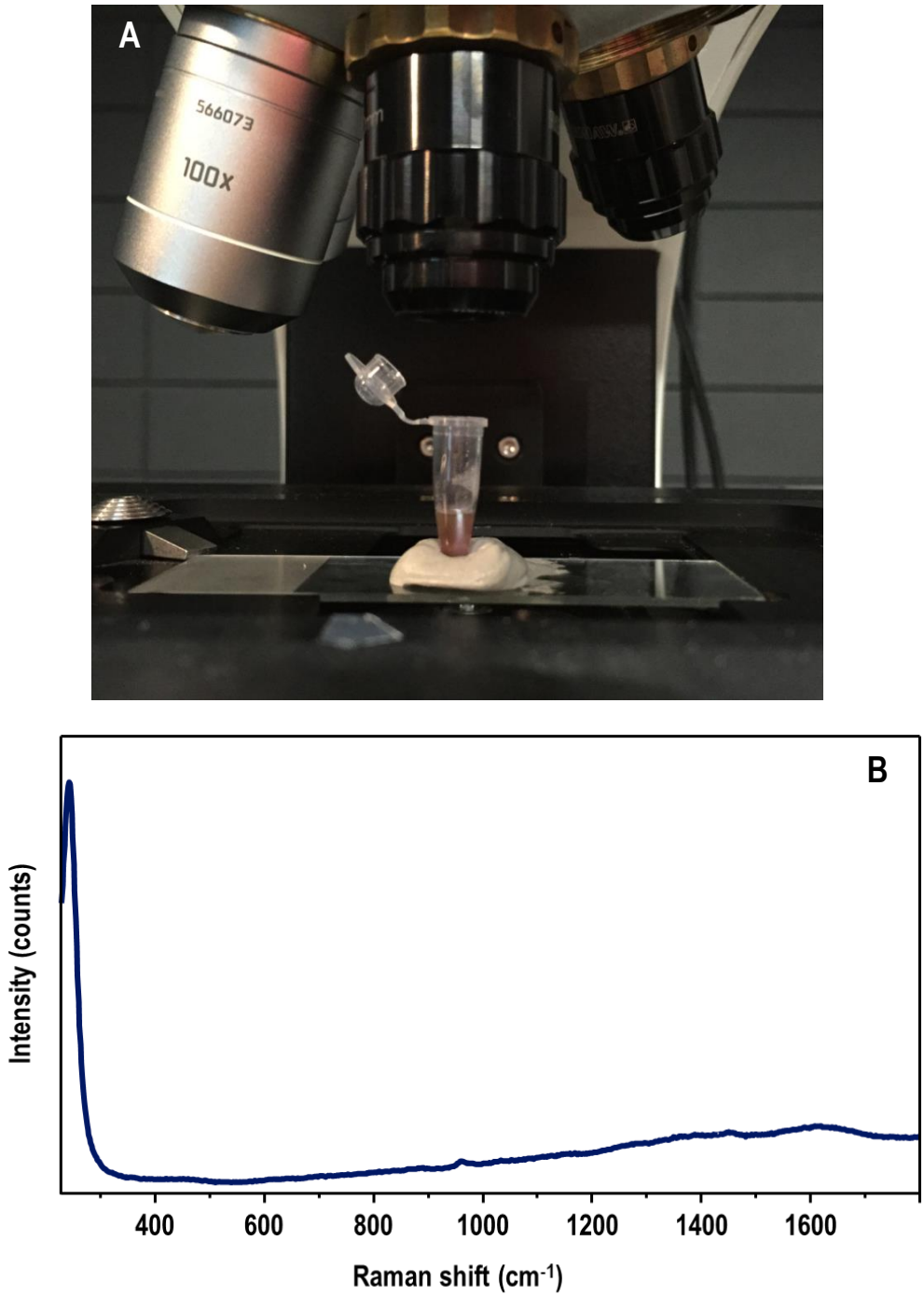


Figure 2.6 A) Setup for SERS experiments; conducted using a Renishaw InVia Reflex confocal microscope. B) Background SERS spectrum of AgNP@Sp irradiated with 532 nm laser. The sample investigated contained 80 µL of AgNP@Sp + 10 µL 0.1M $MgSO_4$.

ζ-Potential

In addition to UV-Vis spectroscopy, TEM and SERS spectroscopy a complementary technique to analyze the positively charged surface of the colloids was required. In this regard, measurements to determine the ζ-potential were carried out (Figure 2.7), demonstrating a positive value for ζ-potential at +41 mV.

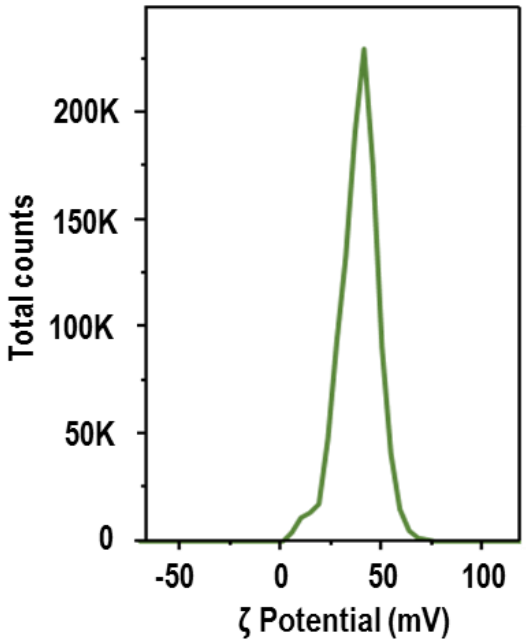


Figure 2.7 ζ potential distribution plot of AgNP@Sp (the ζ-potential corresponds to 41.0 ± 1.5 mV). ζ-potential measurements were performed on a Malvern Nano Zetasizer.

2.4 OPTIMIZATION FOR SERS MEASUREMENTS WITH NA

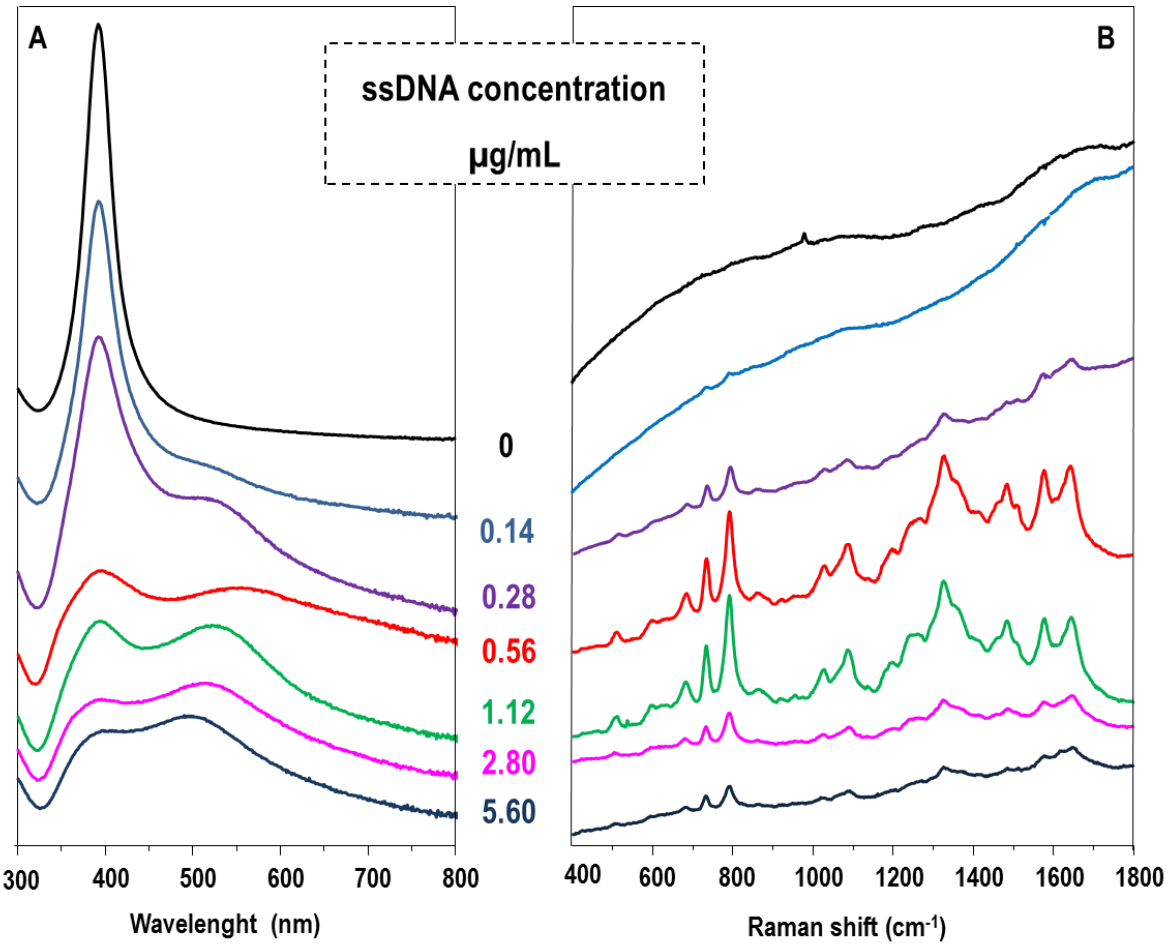
Since NAs act as aggregating agent and cluster-stabilizer in addition of being the target analyte, the optimization of the molar DNA/NP ratio is required prior to the SERS study. It is very important that the DNA concentration employed yields long-term stable clusters in order to maximize the SERS signal and to ensure the spectral reproducibility.

A range of ssDNA concentrations was mixed with AgNP@Sp colloids at fixed concentration (0.3 nM) and investigated by UV-Vis and SERS spectroscopy after 2 hours from the addition of the ssDNA. Stacked extinction and SERS spectra are illustrated in **Figure 2.8.** The addition of the genetic material into the colloids promotes an immediate change in the colour from the typical greenish-yellow to red/orange. Different colours are produced depending on the aggregation state of the nanoparticles, which are directly related to the amount of DNA added to the solution (**Figure 2.8C**).

The same change is also revealed by the decrement in the intensity of the LSPR band at 391 nm and the progressive red-shift to ~525 nm (**Figure 2.8A**). From 0 to 0.28 μg/mL it was revealed that the samples do no show long-term stability, and silver nanoparticles were progressively depositing on the Eppendorf tube wall (where the samples were prepared). When the ssDNA concentration is increased (from 0.56 to 1.12 μg/mL), the second plasmonic contribution at ~525 nm becomes more important, indicating the increment of the cluster formation as corroborated by the generation of intense SERS signals (**Figure 2.8B**).

When the concentration of ssDNA is further increased (from 2.80 to 5.60 μg/mL) the extent of the aggregation decreases as well as the SERS intensity. At the high amount of DNA, the nanoparticle surface is saturated preventing the formation of clusters and increasing the presence of single nanoparticles.

Additionally, the samples corresponding to 1.12 µg/mL and above have shown long-term stability.



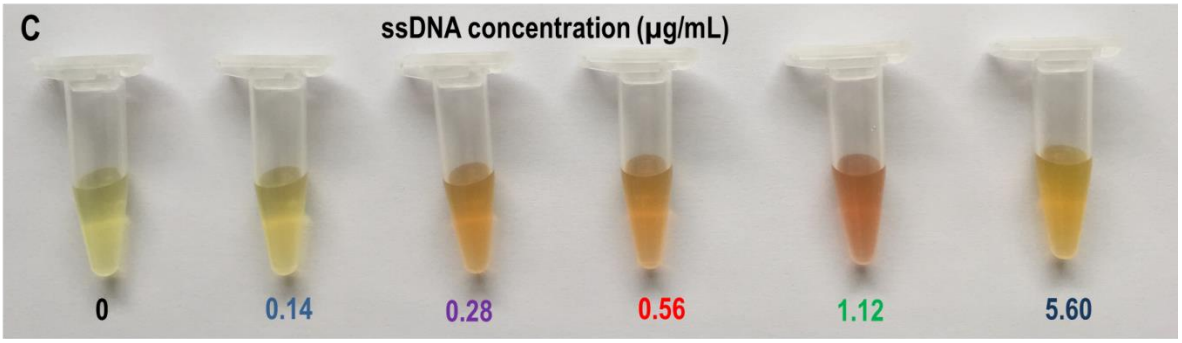


Figure 2.8 A) Extinction spectra and B) SERS spectra of DNA-driven aggregation of AgNP@Sp at different concentrations. The samples were left 2 hours after the addition of ssDNA and then were quickly sonicated before measurements. C) Representative picture of the Eppendorf tubes with the solutions of AgNP@Sp adding from 0 to 5.60 µg/mL ssDNA. The progressive change of the colour is directly related to the amount of DNA present.

From the original non-baselined SERS spectra illustrated in **Figure 2.8B**, it can be observed that the spectral profiles remain unaltered from 0.56 to 2.80 μg/mL, showing intense SERS features. However, outside this range of concentrations, SERS spectra undergo changes related to band intensities or even spectral shifts. It is worth noting that between colloidal batch-to-batch there are inherent dissimilarities associated with NP concentration and size distribution, thus the DNA/NP ratio could suffer subtle changes depending on the batch of NPs used. However, in the appropriate range of DNA/NP ratios such fluctuations do not impact the overall SERS response.

Based on these results, 1.12 μg/mL was selected as the optimal ssDNA concentration for the given NP concentration, yielding long-term stable clusters in solution.

This study is based on single-stranded DNA sequences composed of 21 bases. However, the behaviour of the AgNP@Sp can be extended to longer DNA sequences, double-stranded DNA, RNA or even another kind of structures. In order to determine the optimal concentration that should be employed in those particular cases, a screening with different concentrations of analytes was repeated in a similar way as presented in **Figure 2.8**.

SERS Signal stability

As the nanoparticle aggregation occurs fast and generates long-term stable clusters in suspension, it allows performing SERS measurements after several hours upon the analyte addition (that is, when the process of DNA adhesion and nanoparticle clustering have reached its equilibrium). In order to define the time required to achieve such equilibrium, the relative intensities of the ring breathing bands of G, A and C+T were monitored over time (**Figure 2.9**).

It can be concluded that within the first ~60 min upon the addition of DNA to the colloids, some degree of fluctuations is displayed, however, after that time signal stability is ensured. For this reason, samples were incubated for 2 hours prior to SERS analysis.

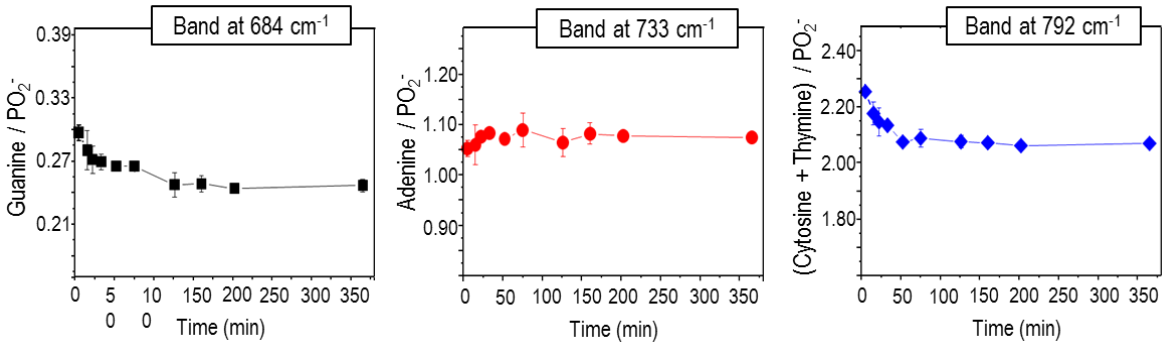


Figure 2.9 Ratiometric peak intensities of the ring-breathing modes (G, A and C+T at ~684, 733 and 792 cm^{-1}, respectively) *vs* the internal standard (υPO_2^- at ~1089 cm^{-1}) extracted from SERS spectra of ds1 acquired at a different time upon the initial addition of DNA to the colloids.

Spectral reproducibility

Generally, aggregated nanoparticles are characterized by having high SERS activity. However, the lack of control on their geometry and size severely affects the SERS reproducibility. The spectral reproducibility is known as an important limiting factor for the successful applicability of label-free SERS analysis in DNA.

For this reason, the spectral reproducibility of AgNP@Sp as SERS platform for the detection of DNA was studied. SERS spectra were acquired in solution using a long working distance lens (30 mm). This setup (already shown in **Figure 2.6A**) allows the simultaneous analysis of a large number of clusters during the acquisition time due to the continuous Brownian motion within the scattering volume of the objective. The obtained spectrum results from averaging numerous contributions, providing highly averaged SERS spectra with well-defined features. The averaging effect is a key factor to obtain stable and reproducible SERS signals.[21]

Figure 2.10 shows the results obtained under the abovementioned experimental conditions. In this case, instead of ssDNA, a double-stranded RNA was used as the molecular probe. 10 SERS spectra of 10 different repetitions of dsRNA (0.56 μg/mL) in AgNP@Sp (0.3 nM) were reported. The repetitions were prepared in 10 different tubes, mixing 200 μL of AgNP@Sp with the same volume of dsRNA. After 2 hours aging, SERS measurements were performed using a green laser (532 nm). As is demonstrated in **Figure 2.10**, the spectral profile

remains unperturbed for all 10 replicas, showing a consistent sample-to-sample reproducibility.

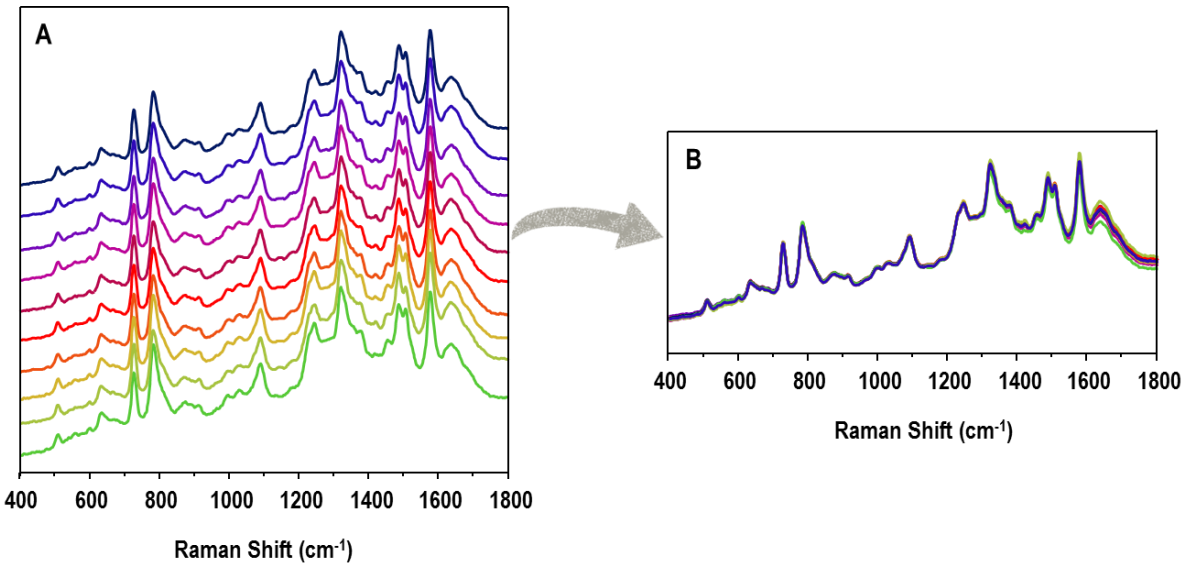


Figure 2.10 No-baselined SERS spectra of dsRNA, A) stacked and B) overlapped. The spectra were normalized at ~1090 cm^{-1}. The repetitions were performed as follows: 200 µL of AgNP@Sp from the same colloidal batch were placed in 10 different vials, then mixed with the same volume of a dsRNA buffered solution and finally investigated by SERS under an average regime.

Detection Limit

The lowest amount of DNA that can be detected under these experimental conditions can be deduced from **Figure 2.8** and complementary calculations. Firstly, by observing the SERS spectra at different ssDNA concentrations it is appreciable that the lowest concentration in which Raman features can be detected is 0.28 µg/mL. Considering that the total volume of the suspension is 200 µL, the corresponding amount of ssDNA in the sample is 56 ng. However, it has to be considered that the volume illuminated by the laser is much less than the 200 µL contained in the tube. In order to obtain an approximation of the real illuminated

volume, *Equation 2.1* was used. When a macrosampling objective is used, the illuminated volume (V) is defined by the local length (f) and the lens diameter (D).[22]

$$V = 3.21\lambda^3 \left[{}^{f}/_{D}\right]^4 \qquad 2.1$$

In our particular case, the laser employed was 532nm, f = 30 mm and D = 10 mm. By the theoretical calculation, the illuminated volume obtained is ~40 μm^3 ($4 \cdot 10^{-9}$ µL). However, it has to be considered that other experimental factors contribute to making the scattering volume much larger than the theoretical calculations.[23] For instance, the DNA-AgNP@Sp clusters are in continuous diffusion, in-and-out the scattering volume, increasing the actual number of investigated nanoparticles and NA sequences.[17]

In order to take into consideration of these factors, an additional and conservative calculation of the illuminated volume was carried out (**Figure 2.11**). In this case, the laser was assumed as a cylinder of 1 mm in diameter and 6 mm high (sample depth). The calculated volume is 4.7 µL, corresponding to less than 1.3 ng in the probed volume. Using this approximation unreliable claims regarding sensitivity are avoided, safely including experimental factors.

In conclusion, limit of detection for ssDNA analysis using AgNP@Sp can be re-estimated as less than 1.3 ng.

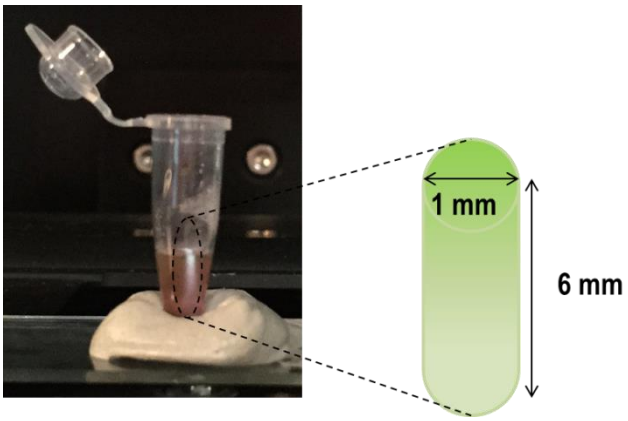


Figure 2.11 Schematic consideration for the approximate calculation of the detection limit of DNA analysis using AgNP@Sp. The illuminated volume by the green laser (532 nm) was considered as a cylinder within the sample volume.

2.5 VIBRATIONAL ASSIGNMENT OF THE MAIN BANDS OF THE SERS SPECTRA

Vibrational assignment of DNA sequences was based on both the data reported in the literature[24] and by measuring the SERS spectra of 21-*mer* homopolymeric sequences of the four bases (**Table 2.1**) illustrated in **Figure 2.12**. The spectral interval 600–800 cm^{-1} contains Raman features associated with vibrations involving concerted ring stretching motions (ring breathing) of purine or pyrimidine residues. In this region, the bands were ascribed to each DNA base; the band mainly ascribable to the guanine residues at ca. 684 cm^{-1}, and strong bands at 733 and 792 cm^{-1} are mainly due to ring breathing vibrations of adenine and the combination of cytosine + thymine residues, respectively.

Considering this data, SERS spectra of a double stranded DNA (ds1, **Table 2.1**) using AgNP@Sp as a plamonic substrate were acquired (**Figure 2.13**). Within the SERS spectra of DNA, we can identify four broad spectral regions of interest:

I. The ring breathing range, from 550 to 820 cm^{-1}; which mainly contains the relatively well-separated ring stretching features of purine and pyrimidine residues, often in combination with stretching of the glycosidic bond and possibly also stretching of bonds within the linked deoxyribose ring.[24e] Thus, these features are very informative about the helical geometry and as identifiers of each nucleobase.

II. The 820–1150 cm^{-1} spectral range is dominated by the intense symmetric stretching vibration of the phosphodioxy moiety at 1089 cm^{-1}. This data clearly indicates that the phosphate groups are in close proximity to the metal surface, providing a further evidence of the electrostatic-driven adsorption of the DNA sequences onto the spermine-coated silver nanoparticles.

III. The highly overlapped nucleobase bands located between 1150 and 1600 cm^{-1}. It contains a complicated pattern of features mainly ascribed to the in-plane vibrations of base residues.

IV. Finally, in the 1600-1800 cm^{-1} region overlap carbonyl group stretching vibrations of thymine (C2=O and C4=O), guanine (C6=O), and cytosine (C2=O). [24e]

On **Table 2.2** are exposed the Raman shifts and the tentative vibrational assignment of the main SERS bands of single-stranded DNA on AgNP@Sp colloids.

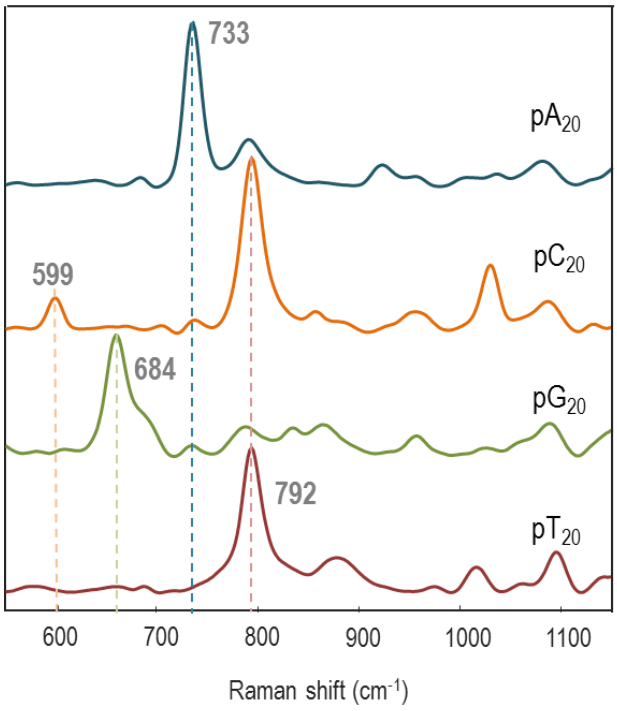


Figure 2.12 SERS spectra of 20-mer DNA homopolymers in the 550-1150 cm^{-1} spectral region. All spectra were normalized to the υPO_2^- band at ca. 1089 cm^{-1}. The illustrated spectra are from averaging 8 independent repeats per sample.

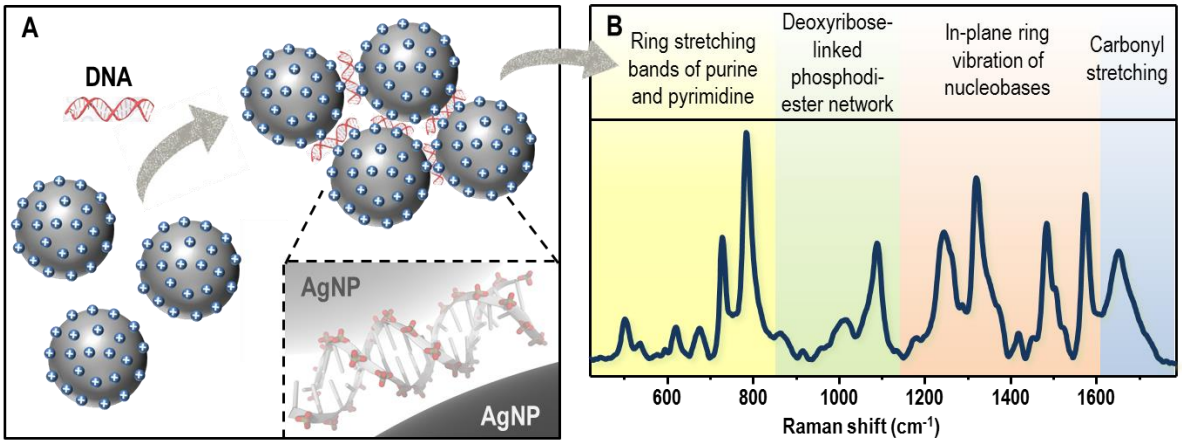


Figure 2.13 A) Schematic outline of the DNA-driven aggregation of cationic nanoparticles into stable clusters in suspension, yielding B) intense SERS spectra (21-bp double-stranded DNA, ds1). The four broad spectral regions of interest are highlighted in different colors.

NAME	SEQUENCES (5' ⟶ 3')
pA_{20}	AAA AAA AAA AAA AAA AAA AA
pC_{20}	CCC CCC CCC CCC CCC CCC CC
pG_{20}	GGG GGG GGG GGG GGG GGG GG
pT_{20}	TTT TTT TTT TTT TTT TTT TTT TT
ds1	ACG TTA GCG CAA TCT GAA TCC TGC AAT CGC GTT AGT CTT AGG

Table 2.1 Reference 20-mer homopolymers used as a reference for determining the spectral markers and ds1, 21 duplex DNA.

Raman shift (cm^{-1})	Vibrational assignment of ssDNA
510	G, ring deformation
599	C, ring bending mode
660	G, ring breathing (C3'-endo/anti) A-form
684	G, ring breathing (C2'-endo/anti) B-form
733	A, ring breathing
792	C+T, ring breathing (mainly C)
920	Deoxyribose, C-C stretching
959	Deoxyribose – Phosphate
1026	Deoxyribose + C
1089	PO_2^-, symmetric stretching
1183	G, C8-H bending
1246	C, ring stretching
1290	A + C, ring stretching
1327	A + G, ring modes
1390	A, ring mode
1415	A, G
1460	C + T, ring modes. (A+G minor contributions)
1485	G + A, ring mode (mainly G)
1509	A, ring mode
1574	A + G, ring modes
1640	C = O (mainly C)

Table 2.2 List of the main Raman shifts and the tentative vibrational assignment of the main SERS bands of ss1 on AgNP@Sp colloids.

2.6 CONCLUSIONS

Positively charged silver nanoparticles are reliable candidates for the analysis of negatively charged analytes. Specifically, by using AgNP@Sp the electrostatic interaction between NAs and the colloids is ensured due to the spermine molecules anchored at the surface. Moreover, their simple preparation, the stable aggregates obtained without the need of an external aggregating agent and the intense SERS signals obtained, make them an optimal SERS platform for NAs analysis.

The synthesis of AgNP@Sp yields quasi-spherical silver nanoparticles of ca. 25 nm diameter as demonstrated by TEM images. The localized surface plasmon resonance centered at ca. 391 nm becomes narrower after 24 hours, improving the monodispersity of the nanoparticle solution. The bulk pH of the colloids is ~ 6, whereas the zeta potential is equal to 41.0 ± 1.5 mV. Moreover, nanoparticle concentration can be calculated by Lambert-Beer Law leading to ca. 0.3nM and the detection limit can be estimated as less than 1.3 ng for ssDNA analysis.

In conclusion, SERS analysis of DNA or RNA can be efficiently and reproducibly achieved in solution by using AgNP@Sp, a large ensemble of stable clusters in solution can be investigated during the acquisition time, providing highly averaged SERS spectra with well-defined features.

REFERENCES

[1] a) E. Papadopoulou, S. E. J. Bell, *Chem.-Eur. J.* **2012**, *18*, 5394-5400; b) E. Papadopoulou, S. E. J. Bell, *Angew. Chem.-Int. Edit.* **2011**, *50*, 9058-9061; c) L.-J. Xu, Z.-C. Lei, J. Li, C. Zong, C. J. Yang, B. Ren, *Journal of the American Chemical Society* **2015**, *137*, 5149-5154; d) S. E. J. Bell, N. M. S. Sirimuthu, *Journal of the American Chemical Society* **2006**, *128*, 15580-15581.

[2] N. H. Jang, *Bulletin of the Korean Chemical Society* **2002**, *23*, 1790-1800.

[3] A. Barhoumi, D. Zhang, F. Tam, N. J. Halas, *Journal of the American Chemical Society* **2008**, *130*, 5523-5529.

[4] L. Guerrini, Z. Krpetic, D. van Lierop, R. A. Alvarez-Puebla, D. Graham, *Angew. Chem.-Int. Edit.* **2015**, *54*, 1144-1148.

[5] A. Torres-Nunez, K. Faulds, D. Graham, R. A. Alvarez-Puebla, L. Guerrini, *Analyst* **2016**, *141*, 5170-5180.

[6] S. R. Panikkanvalappil, M. A. Mackey, M. A. El-Sayed, *Journal of the American Chemical Society* **2013**, *135*, 4815-4821.

[7] a) D. Graham, W. E. Smith, A. M. T. Linacre, C. H. Munro, N. D. Watson, P. C. White, *Anal. Chem.* **1997**, *69*, 4703-4707; b) H. S. Basu, L. J. Marton, *Biochem. J.* **1987**, *244*, 243-246.

[8] B. C. Hoopes, W. R. McClure, *Nucleic Acids Res.* **1981**, *9*, 5493-5504.

[9] R. Gill, G. W. Lucassen, *Anal. Methods* **2010**, *2*, 445-447.

[10] M. Thomas, A. M. Klibanov, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9138-9143.

[11] a) B. L. Frey, R. M. Corn, *Anal. Chem.* **1996**, *68*, 3187-3193; b) Y. D. Ma, Y. Guo, J. Li, J. Guan, L. Xu, W. S. Yang, *Chem.-Eur. J.* **2009**, *15*, 13135-13140.

[12] D. van Lierop, Z. Krpetic, L. Guerrini, I. A. Larmour, J. A. Dougan, K. Faulds, D. Graham, *Chemical Communications* **2012**, *48*, 8192-8194.

[13] L. Guerrini, Ž. Krpetić, D. van Lierop, R. A. Alvarez-Puebla, D. Graham, *Angew. Chem.-Int. Edit.* **2015**, *54*, 1144-1148.

[14] M. Masetti, H. N. Xie, Z. Krpetic, M. Recanatini, R. A. Alvarez-Puebla, L. Guerrini, *Journal of the American Chemical Society* **2015**, *137*, 469-476.

[15] J. Morla-Folch, H.-n. Xie, P. Gisbert-Quilis, S. Gómez-de Pedro, N. Pazos-Perez, R. A. Alvarez-Puebla, L. Guerrini, *Angew. Chem.-Int. Edit.* **2015**, *54*, 13650-13654.

[16] J. Yguerabide, E. E. Yguerabide, *Anal. Biochem.* **1998**, *262*, 157-176.

[17] P. G. E. E. C. Le Ru *Principles of Surface-Enhanced Raman Spectroscopy*, Amsterdam, **2009**.

[18] J. T. Lue, *Journal of Physics and Chemistry of Solids* **2001**, *62*, 1599-1612.

[19] P. C. Lin, S. Lin, P. C. Wang, R. Sridhar, *Biotechnology Advances* **2014**, *32*, 711-726.

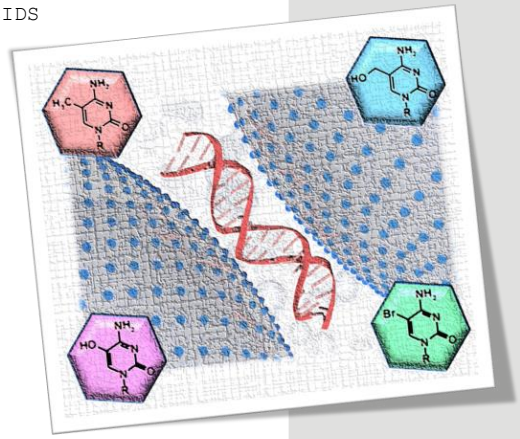
[20] S. Schlücker, *Angewandte Chemie International Edition* **2014**, *53*, 4756-4795.

[21] R. Aroca, *Surface-enhanced Vibrational Spectroscopy*, Chichester, **2006**.

[22] R. A. Álvarez-Puebla, *The Journal of Physical Chemistry Letters* **2012**, *3*, 857-866.

[23] N. J. Everall, *Analyst* **2010**, *135*, 2512-2522.

[24] a) E. Taillandier, J. Liquier, M. Ghomi, *J. Mol. Struct.* **1989**, *214*, 185-211; b) J. M. Benevides, S. A. Overman, G. J. Thomas, *Journal of Raman Spectroscopy* **2005**, *36*, 279-299; c) J. M. Benevides, A. H. J. Wang, G. A. Vandermarel, J. H. Vanboom, G. J. Thomas, *Biochemistry* **1989**, *28*, 304-310; d) D. Serban, J. M. Benevides, G. J. Thomas, *Biochemistry* **2002**, *41*, 847-853; e) H. Deng, V. A. Bloomfield, J. M. Benevides, G. J. Thomas, *Biopolymers* **1999**, *50*, 656-666; f) P. Y. Turpin, L. Chinsky, A. Laigle, B. Jolles, *J. Mol. Struct.* **1989**, *214*, 43-70.



CHAPTER 3

DIRECT DETECTION OF MODIFIED CYTOSINE NUCLEOBASE IN DNA

ABSTRACT

Recognition of chemical modifications in canonical nucleobases of nucleic acids is of key importance since such modified variants act as different genetic encoders, introducing variability in the biological information contained in DNA. In this chapter, it is demonstrated the feasibility of direct SERS, in combination with chemometrics and microfluidics, for the identification and relative quantification of 4 different cytosine modifications in both single- and double-stranded DNA. The minute amount of DNA required per measurement, in the sub-nanogram regime, removes the necessity of pre-amplification or enrichment steps (which are also potential sources of artificial DNA damages). These findings show great potentials for the development of fast, low-cost and high-throughput screening analytical devices capable of detecting known and unknown modifications in nucleic acids opening new windows of activity in several fields such as biology, medicine, and forensic sciences.

OUTLINE

3.1 INTRODUCTION

Chemical modifications in NA

Chemical modifications of the canonical nucleobases of nucleic acids (especially cytosine variants for DNA) play a major role in the increasing of the genetic diversity and, thus, the codification of biological information.[1] In fact, the vastly different functions and properties of cells inside higher organisms demand the specific silencing and activation of cell-type-specific genes which is achieved through an additional layer of epigenetic information beyond the simple genomic sequence.[2] Recently, this chemical diversity in DNA has been demonstrated to be greater than initially predicted.[3] These discoveries suggest that there may be additional underestimated modifications that contribute to the gene regulation.[4]

In addition to epigenetic modifications, an ample variety of nucleobase lesions (such as alkylation, oxidation, deamination, and cross-linking) can take place as a result of carcinogen attacks to DNA.[5] Accumulation of such mutations in genes controlling cell growth, proliferation, programmed cell death, and cell differentiation is likely to cause cancer.[6]

Detection methods for modified DNA nucleobases

Recent discoveries of new nucleotide variants with epigenetic functions have directed an intense research toward the development of novel methods to detect, profile, and sequence these base modifications in the genome and

transcriptome.[2] These strategies span from pure detection and quantification methods (such as immunological detection,[7] ^{32}P-postlabelling,[5a] and liquid chromatography–mass spectrometry (LC-MS)[8]) to genome-wide profiling methods and single-base-resolution sequencing methods.[9] A relatively new field in this exciting area is the screening of DNA samples for unknown or unanticipated lesions, which is referred to as “adductomics”.[10] Nowadays, this sort of analysis is primarily performed with high-resolution LC-MS, which relies on the fragmentation of protonated modified nucleobases, subsequently differentiated according to their molecular masses.[5a, 10] However, DNA adduct analysis by mass spectrometry is costly and time-consuming as it requires prior sample preparation which normally involves several standard steps: *a)* DNA denaturalization, *b)* hydrolysis into the corresponding monomers, *c)* enrichment of the adducts, *d)* removal of unmodified nucleobases and *e)* addition of the appropriate internal standard.[10] In addition, extreme caution must be paid to avoid artificial generation of DNA lesions during these processing steps.

Surface-enhanced Raman scattering (SERS) spectroscopy is an effective method for the identification of unknown species.[11] Moreover, as shown in chapter 2, AgNPs@Sp act as efficient SERS substrate for NA analysis. Herein, the direct SERS analysis of DNA based on AgNP@Sp was applied to the challenging task of quantifying structurally analogous nucleobase variants.

In this chapter, it is described the exploitation of positively charged silver nanoparticles as an effective plasmonic substrate for the ultrasensitive quantification of different modifications in DNA, specifically in the cytosine nucleobase (C). Different modifications in the cytosine base (**Figure 3.Figure 3.1**) were selected: 5-methylcytosine (^{m}C) and 5-hydroxymethylcytosine (^{hm}C) were chosen as the most important epigenetic modifications in mammalian DNA;[12] 5-bromocytosine (^{br}C) as the representative inflammation-induced DNA damage;[13] and 5-hydroxycytosine (^{h}C) as the example of oxidative DNA lesion.[14]

The combination of direct SERS detection and microfluidics (**Figure 3.2**) allows performing a reliable ultrasensitive analysis of DNA samples at the picogram regime (therefore removing the need of pre-amplification/enrichment steps) while offering the possibility to develop simple low-cost devices for real-time automated DNA screening analysis.

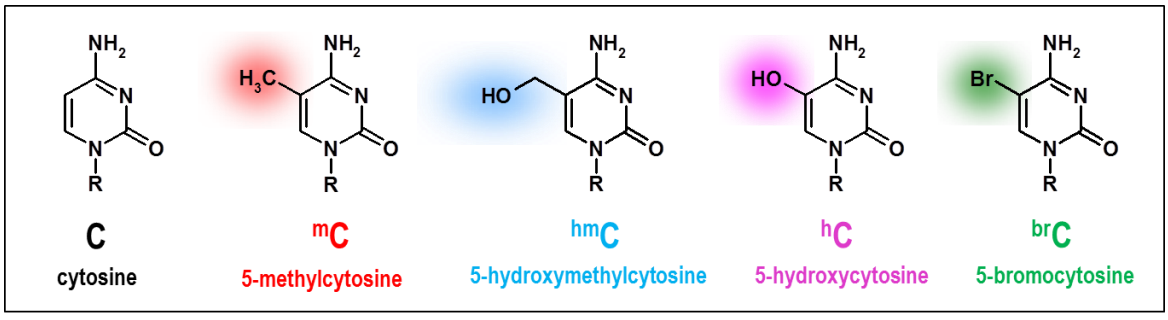


Figure 3.1 Molecular structures of the cytosine modifications.

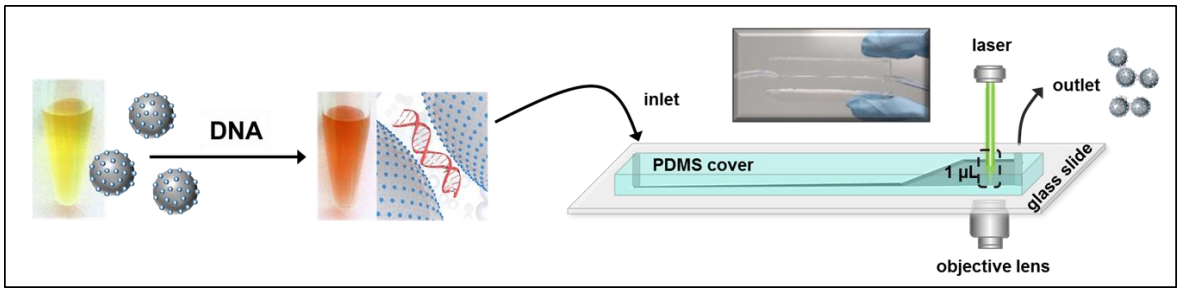


Figure 3.2 Outline of DNA-driven aggregation of AgNP@Sp into stable clusters in suspension (the yellowish colloids immediately turn to deep orange upon DNA addition) and schematic of the on-line detection of DNA via the SERS-microfluidic system.

3.2 RESULTS AND DISCUSSION

Single-stranded DNA with modified cytosines variants

In order to demonstrate the direct SERS recognition and relative quantification of the different cytosine variants in ssDNA, a variety of single-strands were selected for this study (listed in **Table 3.1**). ssC is an oligonucleotide composed by the four canonical nucleobases, whereas its modified counterparts $ss^{mod}Cn$ (where mod = *m*, *hm*, *h* and *br*) contain one, three or all five cytosine variants (n=1, 3, 5).

The averaged SERS spectra of all ssDNA sequences are reported in **Figure 3.3A-D**. Each cytosine variation imposes a characteristic alteration of the SERS spectral profile. In **Figure 3.3E** are illustrated the baseline corrected spectra of ssC, $ss^{m}C5$, $ss^{hm}C5$, $ss^{h}C5$, and $ss^{br}C5$ in the most informative spectral regions.

NAME	SEQUENCE (5' ⟶ 3')
ssC	CAT CGC AGG TAC CTG TAA GAG
ss$^{m}C_1$	^{m}CAT CGC AGG TAC CTG TAA GAG
ss$^{m}C_3$	^{m}CAT ^{m}CGC AGG TA^{m}C CTG TAA GAG
ss$^{m}C_5$	^{m}CAT ^{m}CG^{m}C AGG TA^{m}C ^{m}CTG TAA GAG
ss$^{hm}C_1$	^{hm}CAT CGC AGG TAC CTG TAA GAG
ss$^{hm}C_3$	^{hm}CAT ^{hm}CGC AGG TA^{hm}C CTG TAA GAG
ss$^{hm}C_5$	^{hm}CAT ^{hm}CG^{hm}C AGG TA^{hm}C ^{hm}CTG TAA GAG
ss$^{h}C_1$	^{h}CAT CGC AGG TAC CTG TAA GAG
ss$^{h}C_3$	^{h}CAT ^{h}CGC AGG TA^{h}C CTG TAA GAG
ss$^{h}C_5$	^{h}CAT ^{h}CG^{h}C AGG TA^{h}C ^{h}CTG TAA GAG
ss$^{br}C_1$	^{br}CAT CGC AGG TAC CTG TAA GAG
ss$^{br}C_3$	^{br}CAT ^{br}CGC AGG TA^{br}C CTG TAA GAG
ss$^{br}C_5$	^{br}CAT ^{br}CG^{br}C AGG TA^{br}C ^{br}CTG TAA GAG
ss-com	GTA GCG TCC ATG GAC ATT CTC

Table 3.1 Names and nucleobase sequence of single-stranded DNA. Nucleobase modifications are: 5-methylcytosine (^{m}C), 5-hydroxymethylcytosine (^{hm}C), 5-hydroxycytosine (^{h}C) and 5-bromocytosine (^{br}C). ss-com is the complementary DNA sequence.

The general vibrational assignment is reported in the same **Figure 3.3** in accordance with previous works.[15] Simple visual analysis of the superimposed SERS spectra reveals pronounced alteration of the vibrational patterns, which are mainly circumscribed to the C related bands. For instance, we observe a relative intensity decrease of the C ring breathing band at 790 cm^{-1} (with minor T contribution), and the vibration centered at 595 cm^{-1} which is ascribed to a C ring bending mode.[16]

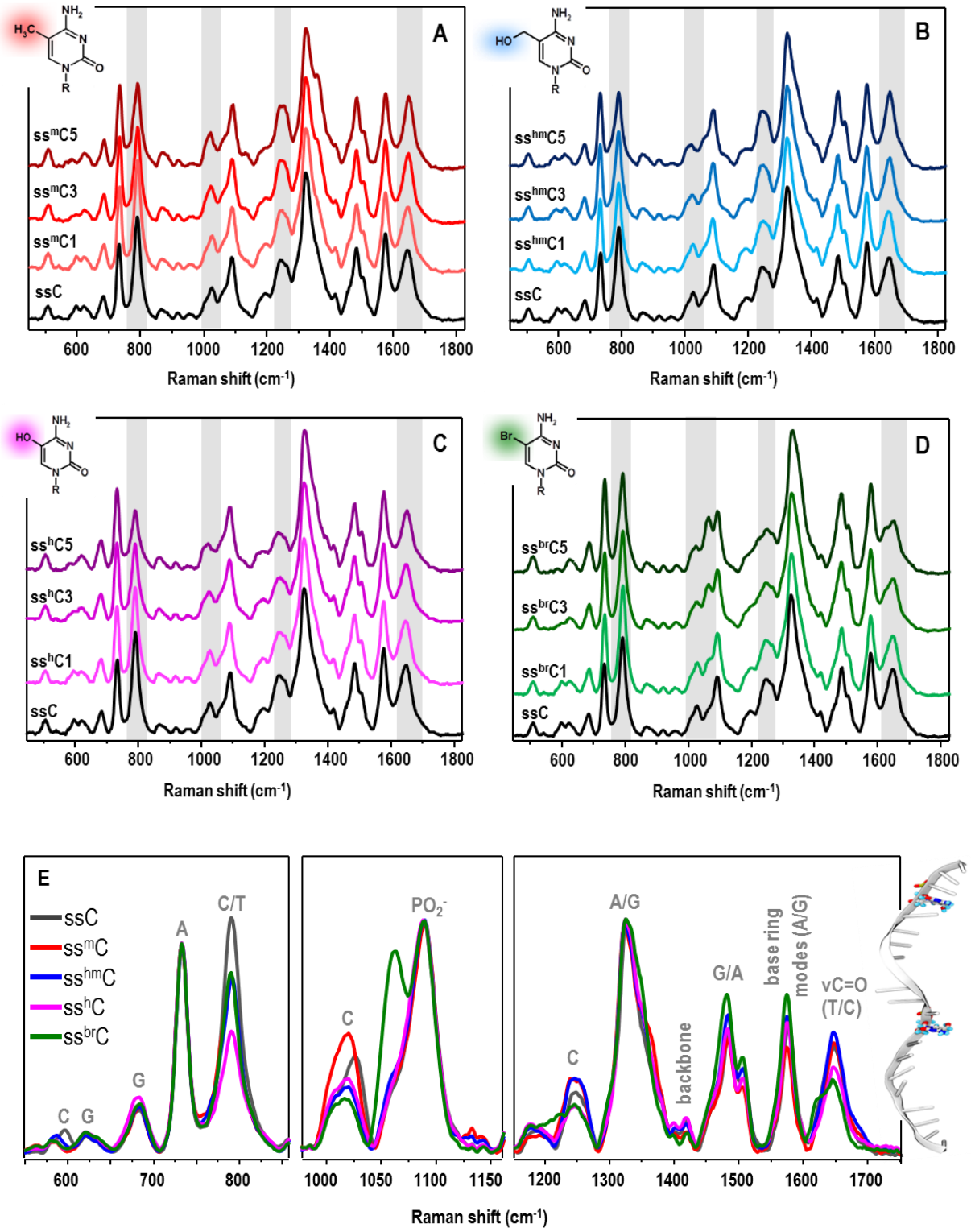


Figure 3.3 A, B, C, and D) Baseline corrected SERS spectra of single-stranded sequences; ssC, ss^{m}C, ss^{hm}C, ss^{h}C and ss^{br}C respectively. C related band are highlighted in gray. E) Overlapped baseline corrected SERS spectra of ssC, ss^{m}C5, ss^{hm}C5, ss^{h}C5 and ss^{br}C5. AgNP@Sp concentration was kept constant to ≈0.3 nM (final concentration in the sample 50 nM).

Similarly, the C ring bending band centered at 1026 cm^{-1} suffers a marked downshift of 7 cm^{-1} and a drastic variation of intensity, as does the peak height of both the C ring stretching feature at 1246 cm^{-1} [17] and the broad carbonyl stretching band at 1647 cm^{-1}.[17a] Importantly, the extension of these changes differs from nucleobase modification to modification.

Double-stranded DNA with modified cytosines variants

The ss oligonucleotides were also hybridized to their complementary sequence to yield the corresponding double-stranded structures dsC, ds^{m}C5, ds^{hm}C5, ds^{h}C5 and ds^{br}C5 (see Experimental Section). All investigated 5-modified cytosine nucleobases have the same base-pairing chemistry as cytosine and, thus, the corresponding double-stranded helixes are expected to structurally differ only for the specific cytosine modification. As for ss, the nucleobase modifications mostly alter the vibrational bands with larger C contributions (**Figure 3.4**). The C/T ring breathing mode undergoes a marked intensity decrease and a 3 nm blue-shift (for ds^{m}C5 and ds^{br}C5). Simultaneously, the C band at 598 cm^{-1} disappears from the spectra of the modified duplexes which also reveal marked perturbations of the C ring features at 1020 cm^{-1} and 1242 cm^{-1}.

In the case of bromocytosine, the SERS spectrum of the single strand ss^{br}C5 displays new intense features at 1060 cm^{-1} and, as a shoulder, at 1622 cm^{-1}. These bands were tentatively assigned to bending modes of the NH_2 group[18] which adopts a co-planar geometry as a result of the Br substitution.[18b] However,

these bands almost disappear from the SERS spectra of ds^{br}C5 with an intensity drop which is far below what one could expect by applying a mere dilution factor (in the investigated duplexes only 5 out 11 cytosines carry structural modifications). It has been reported that, in the isolated cytosine, the introduction of the Br atom at the C5 position promotes a deformation of the ring and the planarity of the NH_2 group (in the unmodified molecule exists in an out-of-plane geometry). Such co-planarity maximizes the interaction between the nitrogen and the ring leading to structural changes of the amino group.[18] On the other hand, it has been shown that the C5-bromination of cytosine has little effect on the atomic charge distribution of the nucleobase[18b] as well as on the overall duplex stability.[18b, 19] Therefore, it can be deduced that, in the duplex conformation, the NH_2 group geometry is largely determined by the strong Watson-Crick base pairing which significantly reduces the influence of the Br group on the amino conformation. This may provide a possible explanation for the drastic intensity decrease of the 1060 and 1622 cm^{-1} bands when ss^{br}C5 is hybridized into the corresponding ds^{br}C5 form.

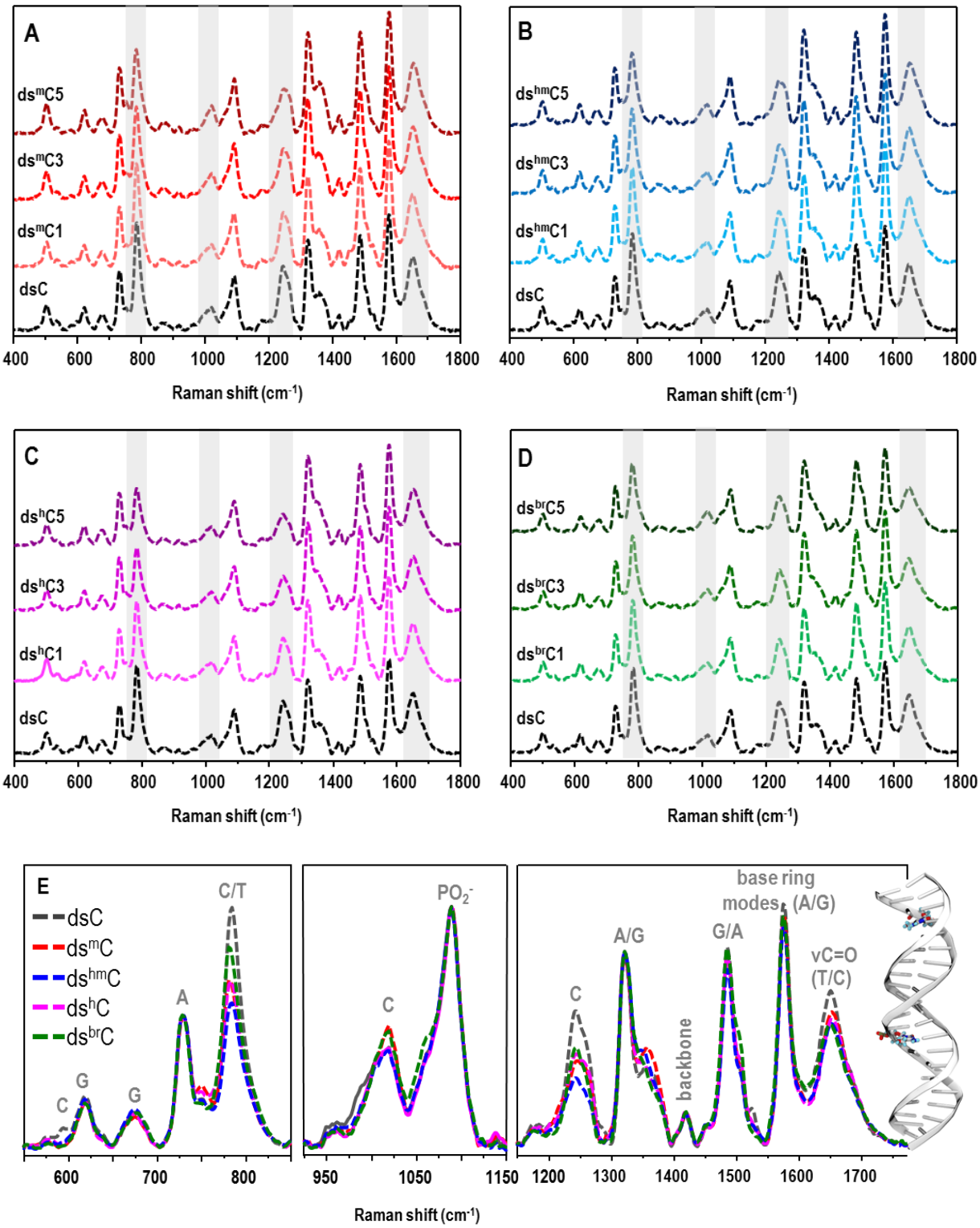


Figure 3.3 A, B, C, and D) Baseline corrected SERS spectra of single-stranded sequences; dsC, ds^{m}C, ds^{hm}C, ds^{h}C and ds^{br}C respectively. C related band are highlighted in gray. E) Overlapped baseline corrected SERS spectra of dsC, ds^{m}C5, ds^{hm}C5, ds^{h}C5 and ds^{br}C5. AgNP@Sp concentration was kept constant to ≈0.3 nM (final concentration in the sample 50 nM).

Statistical differentiation of DNA containing modified cytosine variants

Partial least squares discriminant analysis (PLS-DA) was applied to differentiate between sequences containing the different investigated variations of cytosine bases. The PLS-DA is a supervised multivariate statistical technique describing the variability among different groups, it allows the comparison of the entire SERS spectra instead of comparing each peak separately. From the established model, the 5 sample groups are composed by the SERS spectra from the DNA sequence with unmodified cytosine (ssC) and the other four with their respective modifications (ss^{m}C5, ss^{hm}C5, ss^{h}C5 and ss^{br}C5), all sequences having all 5 modified cytosines. The same analysis was repeated for the double-stranded DNA, in that case, only 5 out 11 cytosines suffer the modification. The results show that the 5 sample groups are distinguishable from each other with 100% sensitivity and specificity, in the cases of both single and double strands (**Figure 3.5A** and **Figure 3.5B**, respectively).

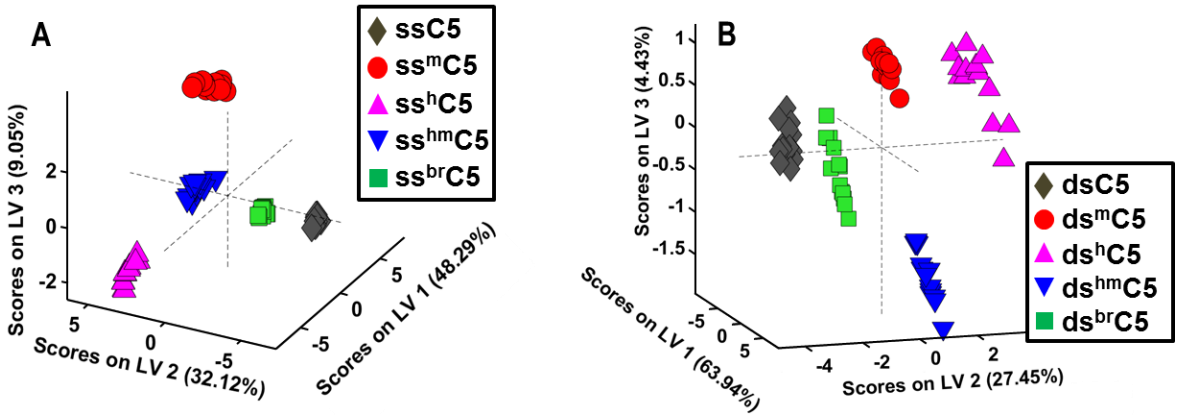


Figure 3.5 A, B) PLS-DA of the ss/ds SERS spectra, respectively. Score plot of loading vectors (LV) 1, 2, and 3 of the obtained SERS spectra, each spectrum is represented by a data point. The mean spectra demonstrated here are from averaging more than 30 measurements (6 independent repeats) per sample.

Quantification of relative content of modified cytosine variants

Next, we investigated the correlation between the extent of the spectral changes with the number of modified cytosine (1, 3 and 5) in ss and ds. The corresponding SERS spectra are illustrated in **Figure 3.3** and **Figure 3.4**, respectively. Most of the relevant spectral changes are localized on the vibrational bands dominated by C contributions. Thus, we monitored the SERS intensities of three of these C-related features (at 784/790, 1026/1020 and 1246/1242 cm^{-1}) as a function of the relative abundance of modified cytosines in each sequence. Additionally, we also included in such comparison the spectral intensity registered at 1060 cm^{-1}, where is centered the new bromo-related band appearing in the ss^{br}C5 SERS spectra (**Figure 3.3D** and **E**). Each peak height was normalized to the phosphodioxy band height at 1089 cm^{-1}, which was used as an internal standard since the content of PO_2^- groups is constant for each sequence. Furthermore, this specific vibrational feature has shown to be scarcely sensitive to structural rearrangements such as the DNA hybridization.[20]

The corresponding plots for ssDNA are represented in **Figure 3.6A**. The C ring breathing mode undergoes a linear drop in intensity as the variant population increases, even though with a different extent depending on the modification type (i.e., different slopes in **Figure 3.6A**).

On the other hand, the bromo-related feature at 1060 cm^{-1} only displays a relative increase when ^{br}C nucleotides are included in the strand. Moreover, bromo- and hydroxyl-modifications also promote a progressive decrease of both

1026 and 1246 cm^{-1} bands, whereas the introduction of the methyl and hydroxymethyl groups poorly affects the C ring vibration at 1246 cm^{-1}. In contrast, an increment of ^{hm}C content imposes a linear drop of the 1026 cm^{-1} band intensity while an opposite trend is observed for ^{m}C.

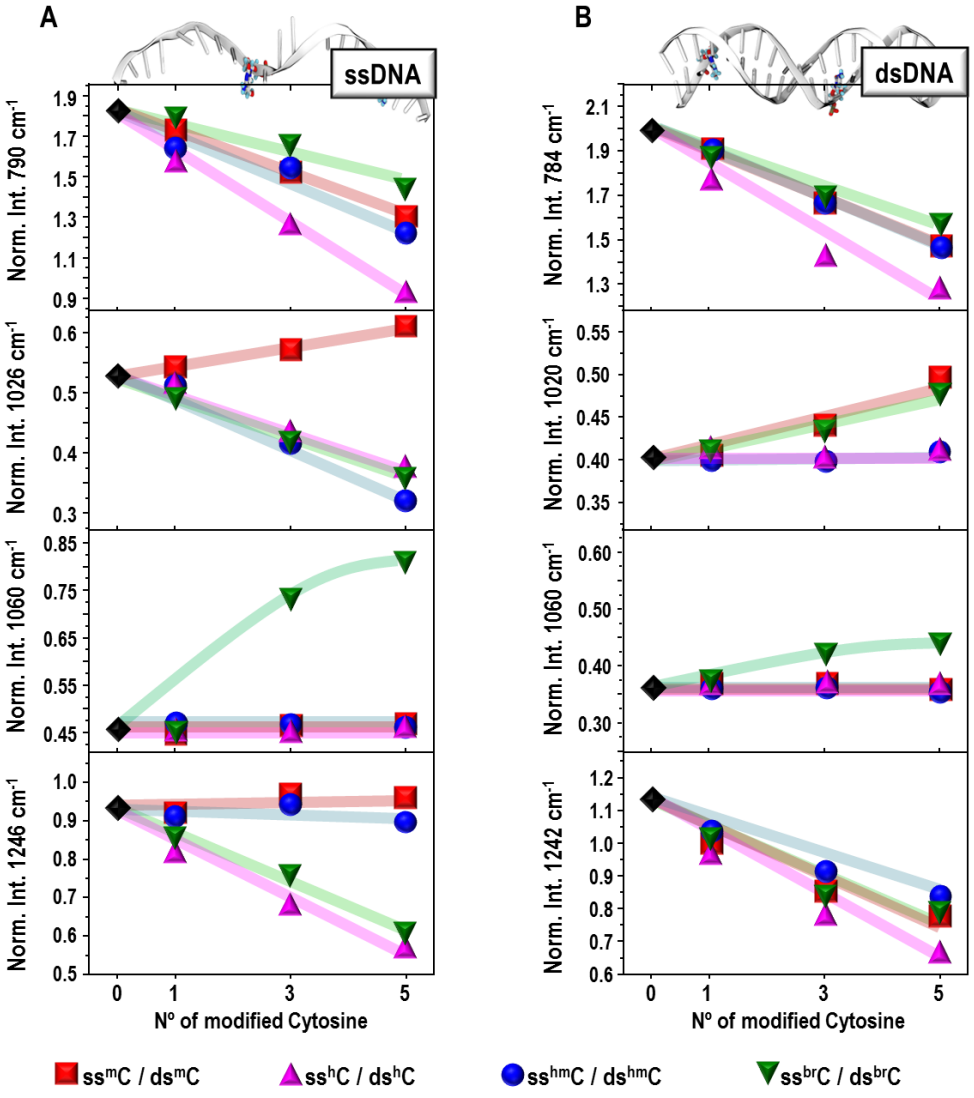


Figure 3.6 Comparison of the relative SERS intensity of the C spectral marker bands as a function of the nucleobase variant content. A) ssDNA: SERS intensities of the band at 790 cm^{-1}, 1026 cm^{-1}, 1060 cm^{-1} and 1246 cm^{-1} as a function of the content of modified cytosine bases. B) dsDNA: SERS intensities of the band at 784 cm^{-1}, 1020 cm^{-1}, 1060 cm^{-1} and 1242 cm^{-1} as a function of the content of modified cytosine bases. The PO_2^- band at 1089 cm^{-1} was used in both cases as the internal standard.

These results combined are extremely important since they do not only highlight the existence of a quantitative correlation between the detected spectral changes and the relative content of modified cytosine nucleobases but plainly show that, even with the simple analysis of just four of the many vibrational features forming the overall spectrum, each modification perturbs the SERS profile in a unique fashion.

Such conclusion was further corroborated by PLS-DA of each set of SERS spectra of ssDNA with different cytosine modification content. Statistical analysis was separately performed on ssC, ss^{m}C1, ss^{m}C3 and ss^{m}C5 data (**Figure 3.7**), then the classification was repeated in the same fashion for the remaining nucleobase variations. The plots show that each single strand with an increasing number of modified cytosine can be successfully differentiated into a separate group with 100% specificity.

Figure 3.6B illustrates the outcome of the same identical study performed on the corresponding dsC, ds^{m}C5, ds^{hm}C5, ds^{h}C5 and ds^{br}C5 duplexes. The general response of the intensity of the C ring breathing mode to the change in variant contents is fairly close to what observed for the single strands. Similarly, the relative intensity measured at 1060 cm^{-1} exclusively registers an increase in the case of brominated cytosines. Nevertheless, the monitoring of the two remaining C-associated bands at 1020 and 1242 cm^{-1} against the variant content provides the most intriguing results.

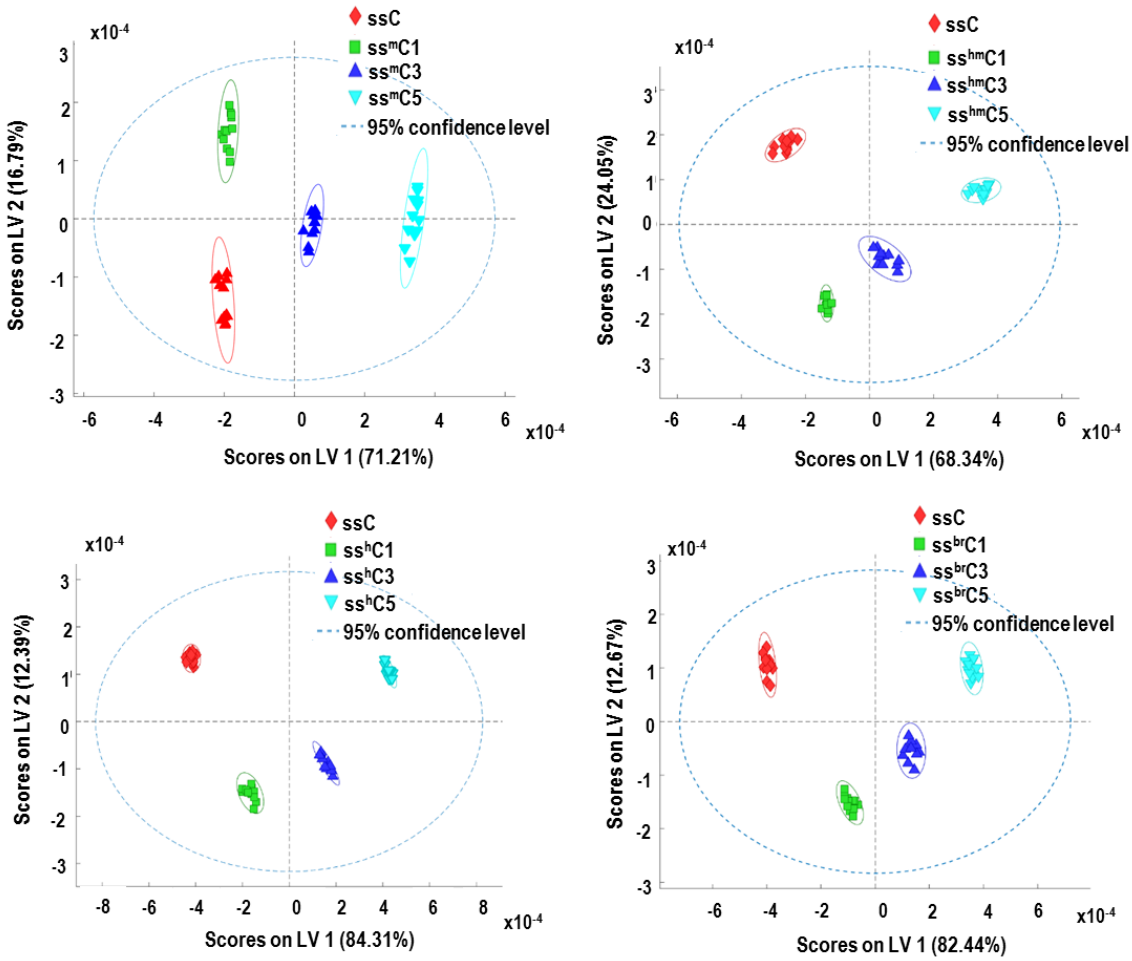


Figure 3.7 PLS-DA of ssDNA SERS spectra. Score plot of loading vectors (LV) 1, and 2, with 95% confidence ellipses, each spectrum is represented by a data point. The position of each point reflects characteristic features of the spectrum.

Once more, their corresponding plots in **Figure 3.6B** show a concentration-dependent correlation of their intensities but their overall patterns strikingly diverge from those observed for ss (for instance, the change in relative intensity trend of the ≈1020 cm^{-1} band for ^{h}C, ^{hm}C and ^{br}C and the ≈1242 cm^{-1} band for ^{m}C and ^{hm}C). This is consistent with the remarkable structural rearrangement and electronic redistribution imposed by the strong Watson–Crick hydrogen bonding and nucleobase stacking within the double helix.

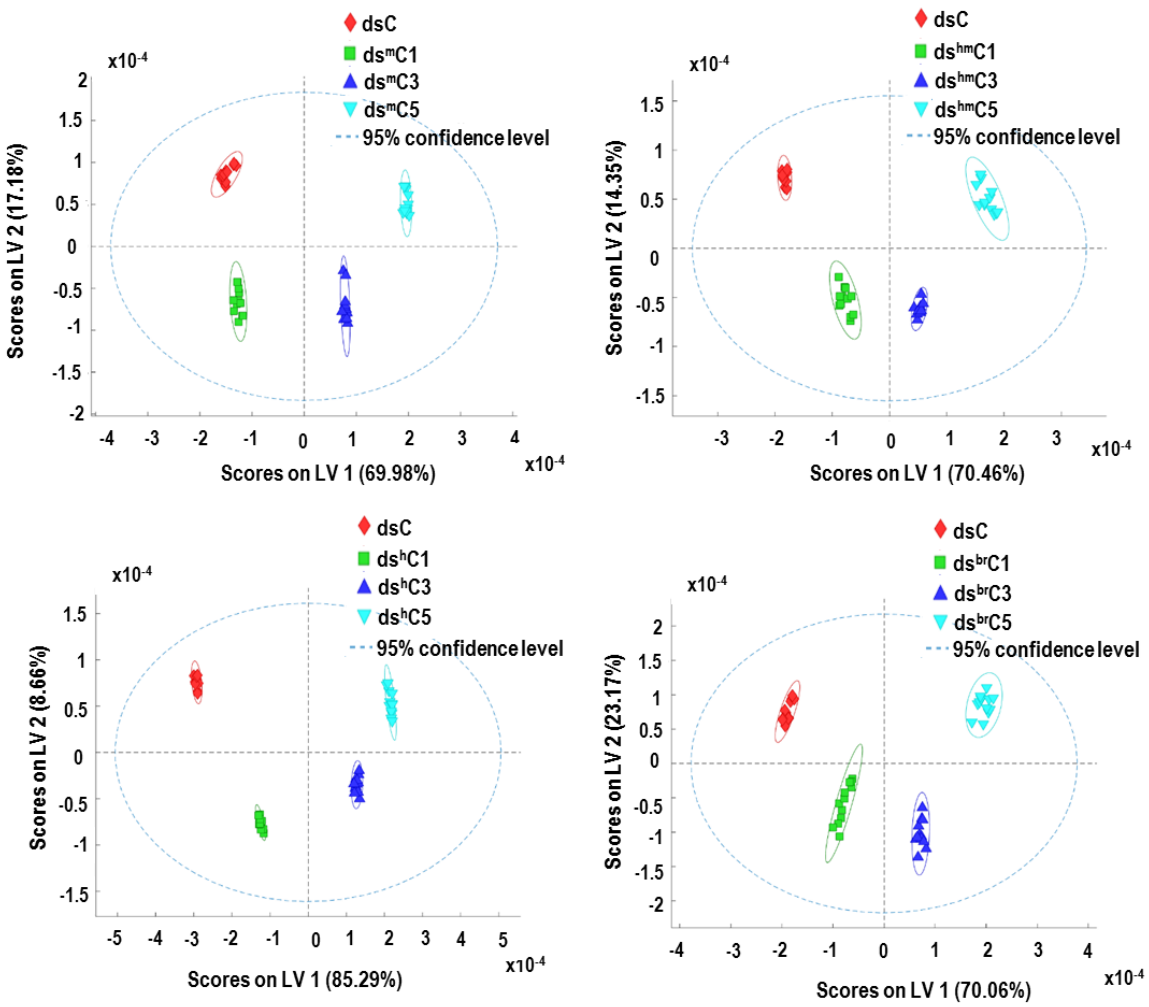


Figure 3.8 PLS-DA of dsDNA SERS spectra. Score plot of loading vectors (LV) 1, and 2 with 95% confidence ellipses, each spectrum is represented by a data point. The position of each point reflects characteristic features of the spectrum.

These results show that introduction of nucleobase variants leads to different sets of spectral perturbations depending if such modifications are included either in a single strand or in the duplex. Similarly, PLS-DA of each set of SERS spectra of dsDNA with different cytosine modification content (**Figure 3.8**) was capable of differentiating them into separate groups with 100% specificity. Beyond these considerations, the data illustrated in **Figure 3.6** demonstrates the capability of our direct method to identify and quantify down to one cytosine variant in ssDNA

oligonucleotides (21 bases, five of them cytosines), as well as in dsDNA (21 base pairs, 42 bases, 11 of them cytosines).

Combining SERS and microfluidics

The reported SERS measurements were performed in a static configuration on colloidal particles in suspension using a long working distance objective. However, under these conditions, the illuminated volume is many orders of magnitude smaller than the actually prepared sample.[15a] Thus, we combined our direct SERS method with a microfluidic device able to manipulate and analyze microscopic volumes of sample in a rapid way (**Figure 3.2**). Remarkably the use of microfluidic device also offers the possibility of automating the whole procedure, being feasible to perform real-time analyses in a continuous way at a relatively low cost.[21]

To demonstrate the reliability of such approach in differentiating and recognizing nucleobase modifications, we collected SERS spectra of the individual samples ssC, ss^{m}C5, ss^{hm}C5, and ss^{br}C5, together with a binary mixture of two of them, which was used as a blind experiment for our data classification. We set the acquisition time to 3 min for each measurement but well-defined SERS spectra can be obtained even for shorter exposures. Samples were analyzed under a continuous flow rate of 10 μL h^{-1}, which in the described conditions entailed a sample volume of less than 1 μL (corresponding to less than 300 picograms of DNA content). The obtained mixture spectra were compared with the pure spectra

of the individual samples and the prediction was made through the non-negative least squares algorithm.[22] The prediction result reveals the composition of the mixture with the precise percentages of each component (39.2% ssC and 60.8% ss^{hm}C), perfectly overlapping with the actual constitution (**Figure 3.9**).

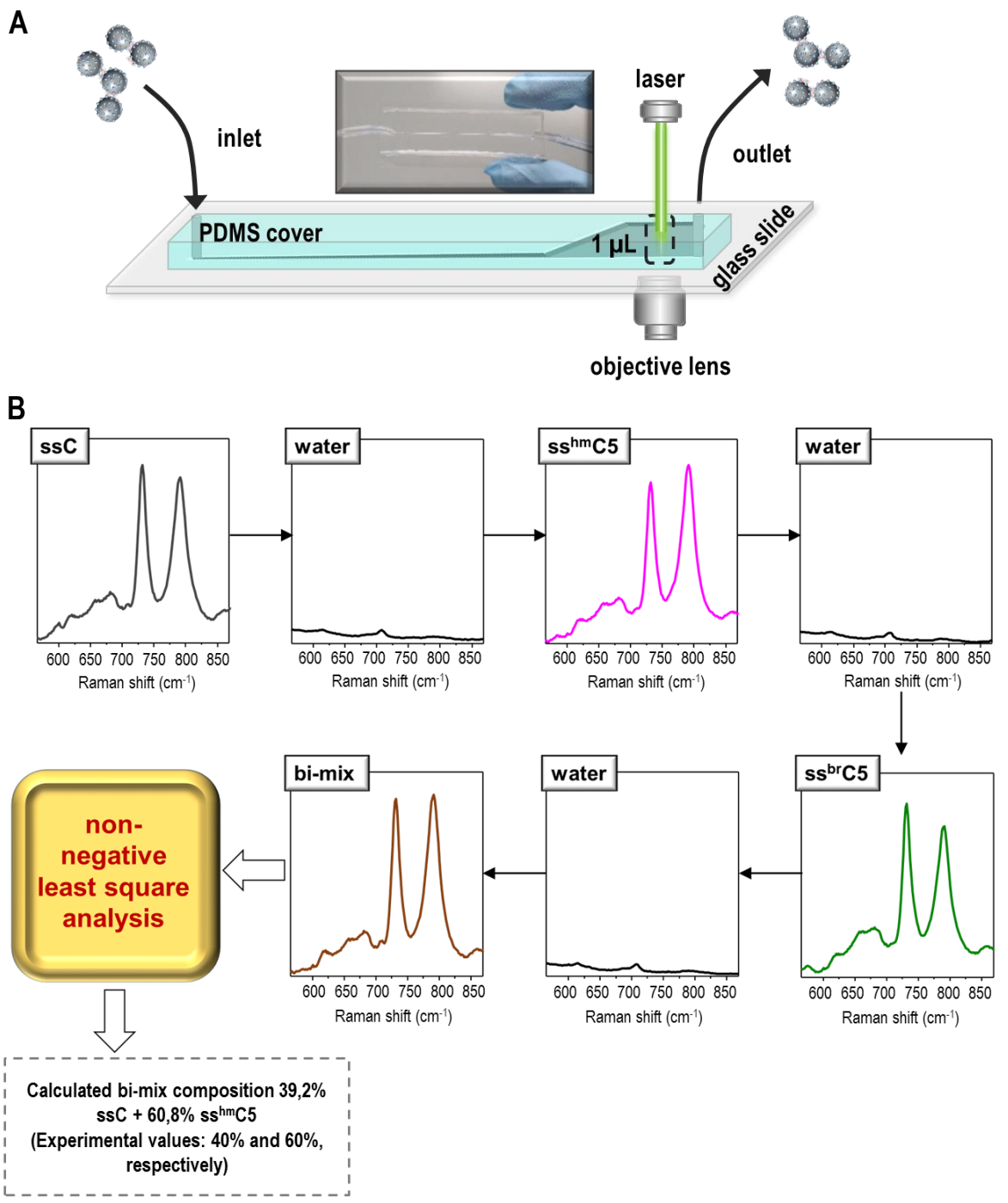


Figure 3.9 A) Schematic representation of the on-line detection of ssDNA via SERS microfluidic system. B) AgNP@Sp were premixed with ssDNA (under the same experimental conditions applied to static SERS measurements) and then pumped into the microfluidic device. SERS spectra were acquired under continuous flow and milli-Q water was used to remove any remaining particles between one measurement and another. The mixture spectra were compared with the pure spectra of the individual samples and the prediction was made through the combination of non-negative least squares algorithm.

3.3 CONCLUSIONS

In summary, we demonstrated the ability of direct SERS analysis of DNA to accurately identify and quantify structurally analogous nucleobase variants in both single-stranded and duplex structures. This study represents a significant step forward in the ultrasensitive detection, vibrational characterization and quantification of cytosine modifications in DNA by direct SERS, which was limited so far to the qualitative differentiation of methyl- and hydroxymethyl-cytosine variants in short single-stranded sequences at high DNA concentrations (0.1 mM).[23]

Classification of the SERS spectra was achieved via PLS-DA. Implementation of the sensing strategy into a microfluidic device further reduced the required amount of DNA per measurement to the picogram level while enabling fundamentally new applications, such as an increase in automation and parallelization, which paves the way to screening and systematic testing analysis.

These findings offer the opportunity for the future development of a fast, low-cost and a high-throughput analytical tool capable of detecting known and unknown modifications in nucleic acids (DNA and RNA) and, thus, opening new windows in biology and beyond. For instance, we may foresee a method based on, first, restricted digestion of long genomic DNA into shorter fragments that, secondly, will be analyzed by SERS combined with microfluidics for rapid and ultrasensitive analyte screening. Finally, the acquired spectra will be compared

with a previously constructed SERS library comprising the fingerprint vibrational spectra of an ample set of synthetically modified DNA.

3.4 EXPERIMENTAL SECTION

Materials

All materials were of highest purity available and obtained from Sigma Aldrich, unless stated otherwise. DNA oligonucleotides were purchased from ATDBio (Southampton, UK) and 0.4 mM stock solutions were prepared by dissolving into Milli-Q water. The single-strands (ss) selected for this study are listed in **Table 3.1**. Duplex structures were prepared by annealing at 95°C for 10 minutes equimolar solutions of ssC, ss^{m}Cn, ss^{hm}Cn, ss^{h}Cn and ss^{br}Cn (n = 1, 3 and 5), respectively, with their complementary strand (ss-com) in 0.3 M PBS (final concentration of ds in solution was 20 μM). The solutions were stored at -20 °C until required.

Microfluidic Device Fabrication

A polydimethylsiloxane (PDMS) microfluidic device, consisting of a single microchannel with variable dimensions, was fabricated by replica molding. The inlet channel has a width of 200 μm and a depth of 100 μm, and was enlarged in the outlet channel (detection section) to 800 μm-width and 3 mm-depth. The pattern was transferred to PDMS using an aluminum master mold. Briefly, PDMS

was mixed with curing agent (10:1) (Sylgard 184, Dow Corning) and poured onto the mold. Once degassed in vacuum, the mixture was cured at 80°C for an hour. Then, the PDMS layer was peeled off from the master, and inlet/outlet holes were punched with a 2.5 mm-sized biopsy punch. A 130-170 μm-thick cover glass was used to seal the bottom of the PDMS layer, since it provides a minimal SERS signal when compared to conventional 1 mm-thick glass slides. Thus, both PDMS and cover glass were exposed to oxygen plasma treatment (100 W, 3% O_2, 0.2 mbar, 40 s) (Plasma Flecto 10, Plasma technology GmbH) for increasing their adhesion and were immediately bonded. The final device has a volume of 25 μL.

SERS experiments

For single-stranded SERS studies, 10 μL of 1 μM ssDNA solutions were mixed with 200 μL of AgNP@Sp ([NP] ca. 0.3 nM). The final concentration in the samples was ca. 50 nM. For double-stranded SERS studies, 5 μL of 1 μM dsDNA solutions were mixed with 200 μL of AgNP@Sp ([NP] ca. 0.3 nM). The final concentration in the samples was 25 nM. After the addition of the DNA, the colloids were left to equilibrate for 3 h and redispersed by quick sonication before running the SERS measurements.

Equipment and Instrumental Settings

SERS experiments were conducted using a Renishaw InVia Reflex confocal microscope equipped with a high-resolution grating consisting of 1800

grooves/mm for visible wavelengths, additional band-pass filters, and a CCD camera. A 532 nm laser was focused onto the colloidal solution by a long-working distance objective (0.17 NA, working distance 30 mm) and the spectra were acquired at room temperature with 20 accumulations, 15 s exposure time and a laser power at the sample of 6.9 mW. All SERS spectra reported in the manuscript were obtained by averaging the SERS responses of 6 different replications per each sample. These replications were obtained by combining the same colloids (prepared the day before the measurements and left aging overnight) with the specific DNA solution. It is worth highlighting that, in the case of difference SERS spectra, the baseline correction was performed after the digital subtraction between the original no baseline-corrected spectra to avoid the possible generation of spectral artefacts.

PLS-DA is a supervised multivariate statistical technique describing the variability among spectra. Instead of comparing each peak separately, this technique allows comparison of the entire spectra. Acquired SERS spectra were imported to the PLS tool box (Eigenvector Research, Inc.) in Matlab for PLS-DA modelling. The spectra were normalized using Multiplicative Scatter Correction[24] and mean centered before being calculated through the model.

Microfluidic device set-up

For on-line monitoring of nucleobase modifications (on-chip measurements), the fabricated microfluidic device was placed onto the translation stage of the

Renishaw Invia Reflex system and appropriately aligned. Inlet and outlet were connected to a syringe and a waste container, respectively, via polytetrafluoroethylene (PTFE) (i.d. 0.8 mm) and tygon (i.d. 1.1 mm) tubes. Ethanol and later Milli-Q water were flushed into the microfluidic device to remove all air bubbles and condition the microchannel. Then, samples were pumped at a flow rate of 10 μL h^{-1} by using a standard syringe pump (NE-1000, New Era Pump Systems, Inc.). A 532 nm laser was focused halfway height of the outlet channel to minimize the PDMS and glass signal background to the SERS spectra by a long-working distance objective (0.17 NA, 30 mm working distance, 6.9 mW laser power). SERS spectra were acquired within 3 min measure per sample (18 accumulations, 10 s exposure time) at room temperature. The analyzed volume in the described conditions was less than 1 μL.

REFERENCES

[1] T. Carell, C. Brandmayr, A. Hienzsch, M. Muller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem.-Int. Edit.* **2012**, *51*, 7110-7131.

[2] C. X. Song, C. Q. Yi, C. He, *Nat. Biotechnol.* **2012**, *30*, 1107-1116.

[3] G. Egger, G. N. Liang, A. Aparicio, P. A. Jones, *Nature* **2004**, *429*, 457-463.

[4] M. Lee, B. Kim, V. N. Kim, *Cell* **2014**, *158*, 980-987.

[5] a) S. Balbo, R. J. Turesky, P. W. Villalta, *Chemical research in toxicology* **2014**, *27*, 356-366; b) J. K. Nicholson, E. Holmes, J. Kinross, R. Burcelin, G. Gibson, W. Jia, S. Pettersson, *Science* **2012**, *336*, 1262-1267.

[6] a) X. M. He, S. H. Chang, J. J. Zhang, Q. Zhao, H. Z. Xiang, K. Kusonmano, L. Yang, Z. S. Sun, H. M. Yang, J. Wang, *Nucleic Acids Res.* **2008**, *36*, D836-D841; b) A. Portela, M. Esteller, *Nat. Biotechnol.* **2010**, *28*, 1057-1068.

[7] G. F. Jia, Y. Fu, X. Zhao, Q. Dai, G. Q. Zheng, Y. Yang, C. Q. Yi, T. Lindahl, T. Pan, Y. G. Yang, C. He, *Nat. Chem. Biol.* **2011**, *7*, 885-887.

[8] M. Munzel, D. Globisch, T. Bruckl, M. Wagner, V. Welzmiller, S. Michalakis, M. Muller, M. Biel, T. Carell, *Angew. Chem.-Int. Edit.* **2010**, *49*, 5375-5377.

[9] M. L. Metzker, *Nat. Rev. Genet.* **2010**, *11*, 31-46.

[10] N. Tretyakova, P. W. Villalta, S. Kotapati, *Chem. Rev.* **2013**, *113*, 2395-2436.

[11] S. Schlücker, *Angewandte Chemie International Edition* **2014**, *53*, 4756-4795.

[12] a) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303; b) M. Tahiliani, K. P. Koh, Y. H. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930-935.

[13] a) M. Whiteman, A. Jenner, B. Halliwell, *Chemical research in toxicology* **1997**, *10*, 1240-1246; b) C. Badouard, M. Masuda, H. Nishino, J. Cadet, A. Favier, J. L. Ravanat, *J. Chromatogr. B* **2005**, *827*, 26-31.

[14] M. D. Evans, M. Dizdaroglu, M. S. Cooke, *Mutat. Res.-Rev. Mutat. Res.* **2004**, *567*, 1-61.

[15] a) L. Guerrini, Z. Krpetic, D. van Lierop, R. A. Alvarez-Puebla, D. Graham, *Angew. Chem.-Int. Edit.* **2015**, *54*, 1144-1148; b) J. M. Benevides, S. A. Overman, G. J. Thomas, *Journal of Raman Spectroscopy* **2005**, *36*, 279-299; c) P. Y. Turpin, L. Chinsky, A. Laigle, B. Jolles, *J. Mol. Struct.* **1989**, *214*, 43-70.

[16] M. Shanmugasundaram, M. Puranik, *Journal of Raman Spectroscopy* **2009**, *40*, 1726-1748.

[17] a) W. Suen, T. G. Spiro, L. C. Sowers, J. R. Fresco, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4500-4505; b) C. P. Beetz, G. Ascarelli, *Spectroc. Acta Pt. A-Molec. Biomolec. Spectr.* **1980**, *36*, 299-313.

[18] a) V. K. Rastogi, M. A. Palafox, K. Lang, S. K. Singhal, R. K. Soni, R. Sharma, *Indian J. Pure Appl. Phys.* **2006**, *44*, 653-660; bM. A. Palafox, V. K. Rastogi, S. Kumar, H. Joe, *Spectroc. Acta Pt. A-Molec. Biomolec. Spectr.* **2013**, *111*, 104-122.

[19] W. S. Ross, C. C. Hardin, I. Tinoco, S. N. Rao, D. A. Pearlman, P. A. Kollman, *Biopolymers* **1989**, *28*, 1939-1957.

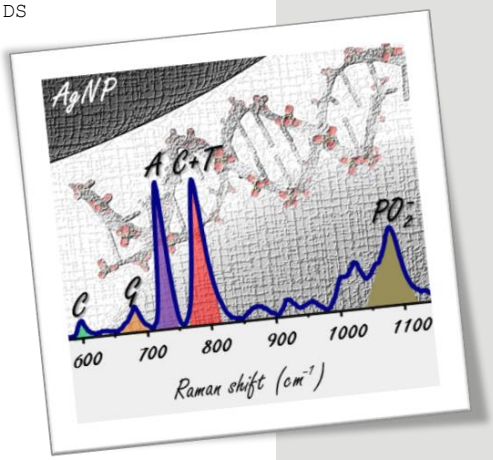
[20] J. G. Duguid, V. A. Bloomfield, J. M. Benevides, G. J. Thomas, *Biophys. J.* **1996**, *71*, 3350-3360.

[21] E. K. Sackmann, A. L. Fulton, D. J. Beebe, *Nature* **2014**, *507*, 181-189.

[22] R. Bro, S. DeJong, *J. Chemometr.* **1997**, *11*, 393-401.

[23] A. Barhoumi, N. J. Halas, *J. Phys. Chem. Lett.* **2011**, *2*, 3118-3123.

[24] P. Geladi, D. Macdougall, H. Martens, *Appl. Spectrosc.* **1985**, *39*, 491-500.



CHAPTER 4

DIRECT QUANTIFICATION OF DNA BASE COMPOSITION

ABSTRACT

Despite the vast number of current nanotechnology applications, the determination of the base composition in nucleic acids, a fundamental parameter in genomic analyses and taxonomic classification, is still restricted to time-consuming and poorly sensitive conventional methods. In this chapter, AgNP@Sp are exploited to demonstrate the possibility of determining the base composition in single and double stranded DNA by using the previously described label-free SERS method.

OUTLINE

4.1 INTRODUCTION

Direct SERS for DNA base quantification

Target DNA available is usually in low amount; therefore, the development of optical strategies that could detect the genetic material with extremely high sensitivity (i.e.; PCR-free methods) represents a key requirement for their effective implementation into rapid, easy-to-use, cost-effective and high accuracy DNA sensors.[1] Particularly, DNA base composition is a fundamental parameter in genomic analyses[2] and taxonomic classification.[3] Traditional methods for screening of nucleobase quantification, such as HPLC, melting point, and buoyant density, are time-consuming, labour-intensive and normally requires relatively large amounts of DNA.[4] Surprisingly, in the multiple SERS-based sensing approaches for DNA analysis reported in the literature, quantification of base composition in nucleic acids have been limited so far to synthetic short (12-mer) bi-polymeric Adenine/Cytosine sequences.[5] Herein, it is demonstrated the potential of SERS in the quantification of the base composition in complex single- and double-stranded DNA (short and genomic).

4.2 RESULTS AND DISCUSSION

Selection of the spectral markers

SERS spectra were acquired using AgNP@Sp as previously described in chapter 2. Within the SERS spectra of DNA, we can identify different spectral

regions of interest (see in detail section 2.5). While the nucleobase spectral features in the ca. 800-1800 cm^{-1} appear as a complex pattern of overlapping bands, intense and well-defined ring stretching bands of purine and pyrimidine bases are localized in the in the 500-820 cm^{-1} region. Thus, we selected these bands as spectral markers for extracting the relative nucleobase content. **Table 4.1** and **Table 4.2** show the spectral markers selected for each base in ssDNA and dsDNA, respectively.

While structural classification of large biomolecules requires multivariate analysis of the Raman spectra, an univariate approach is more suited when relative quantifications are required.[6] Identification of an internal standard is also necessary to remove the fluctuations of the absolute intensity measurements.[6b] In our case, the intense υPO_2^- band at ca. 1089 cm^{-1} was designated as the internal standard since it is poorly affected by DNA structural changes.[5b, 7] Additionally, the ratio between the overall content of phosphate groups and nucleobase units is constant and independent of the base composition.

21-mer single-stranded DNA		
NUCLEOBASE	**PEAK POSITION (cm^{-1})**	**VIBRATIONAL ASSIGNMENT**
Cytosine - C	599	C, ring bending mode
Guanine - G	684	G, ring breathing
Adenine - A	733	A, ring breathing
Thymine - T	790	T, ring breathing

Table 4.1 Characteristic bands selected as spectral markers for of the relative nucleobase content in ssDNA.

21-bp double-stranded DNA		
NUCLEOBASE	**PEAK POSITION (cm^{-1})**	**VIBRATIONAL ASSIGNMENT**
G (=C)	684	G, ring breathing
A (=T)	733	A, ring breathing

Table 4.2 Characteristic bands selected as spectral markers for of the relative nucleobase content in dsDNA.

Extraction of the quantification parameters

Eleven 21-mer single-stranded DNA (**Table 4.3**) and 21-base pair (bp) duplexes (**Table 4.4**) of different base sequence and composition were selected as probe samples to train the sensing system and extract the quantification parameters.

NAME	SEQUENCES (5' ⟶ 3')
ss1	CAT CGC AGG TAC CTG TAA GAG
ss2	GAT GCG TCC ATG GAC ATT CTC
ss3	CAT CGC AGA TAC CTG TAA TAG
ss4	CTC TCC TTG ACT CAT CCA TAG
ss5	ATC GTC AAG AAT TCG AAC CAG
ss6	CTC CCA GTG CGG ACC CTG CGT
ss7	GAT TAC TGA TGT GTC TCA GTA
ss8	CTA ATG ACT ACA CAG AGT CAT
ss9	TAA CAT AGA ATA CTA GTT TAA
ss10	ATT GTA TCT TAT GAT CAA ATT
ss11	GAT TGA ATT GAG AGA TAG TTG

Table 4.3 Reference 21-mer single-stranded sequences (ssDNA) used as training samples.

NAME	SEQUENCES (5' ⟶ 3')
ds1	CAT CGC AGG TAC CTG TAA GAG GTA GCG TCC ATG GAC ATT CTC
ds2	GAT GCG TCC ATG GAC ATT CTC CTA ACA TGA CTC TCT GTC CGA
ds3	GTC ATA CAG CAT CTG CAA GCA CAG TAT GTC GTA GAC GTT CGT
ds4	TCT GTC TTA CGA TAT ACT ATA AGA CAG AAT GCT ATA TGA TAT
ds5	GAT TAC TGA TGT GTC TCA GTA CTA ATG ACT ACA CAG AGT CAT
ds6	CCA GTC AGT CAC GCC TCA GAG GGT CAG TCA GTG CGG AGT CTC
ds7	CCG TGC CCG TAG CGA GCC GCC GGC ACG GGC ATC GCT CGG CGG
ds8	TAA CAT AGA ATA CTA GTT TAA ATT GTA TCT TAT GAT CAA ATT
ds9	CCG CGC CGC GCG CGC GGC GCGG GGC GCG GCG CGC GCG CCG CGCC
ds10	AAT ATA ATA TAT ATA TTA TATT TTA TAT TAT ATA TAT AAT ATAA
ds11	AAT CGC AGG TAC GTG TAA GAG GTA GCG TCC ATG GAC ATT CTC

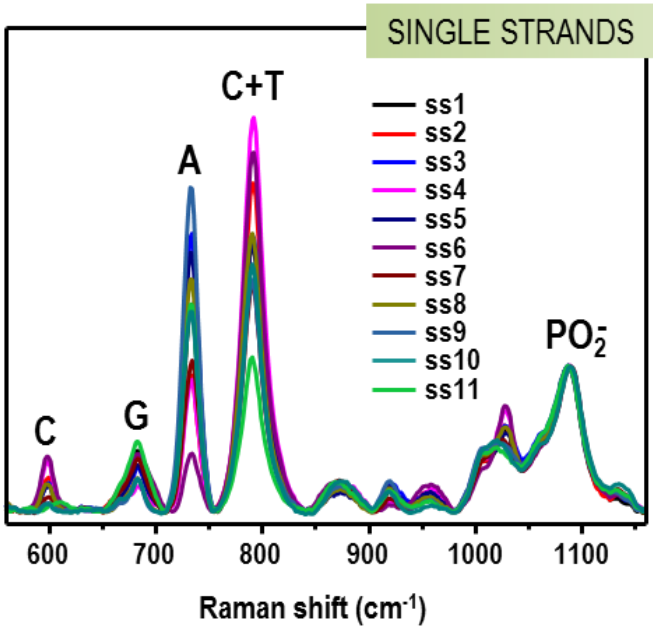
Table 4.4 Reference 21-bp double-stranded sequences (dsDNA) used as training samples.

Figure 4.1 and **Figure 4.2** show the SERS spectra of the probe samples in the sub-1150 cm^{-1} spectral range and their corresponding nucleobase content (in percentage). The spectra were normalized to the internal spectral standard at 1089 cm^{-1}. Baseline correction was identically applied to all spectra to remove fluctuations of the background signal. A general vibrational assignment is also included in the figures.

In the case of single strands (**Figure 4.1**), we can clearly distinguish the intense ring-breathing modes of guanine at ca. 684 cm^{-1} and adenine at ca. 733 cm^{-1}, while the intense feature at ca. 792 cm^{-1} contains two distinct components: the cytosine and the thymine ring breathing bands centred at ca. 793 and 790 cm^{-1}, respectively.[7a] An additional C-band at ca. 599 cm^{-1} is ascribed to a ring bending mode.[8]

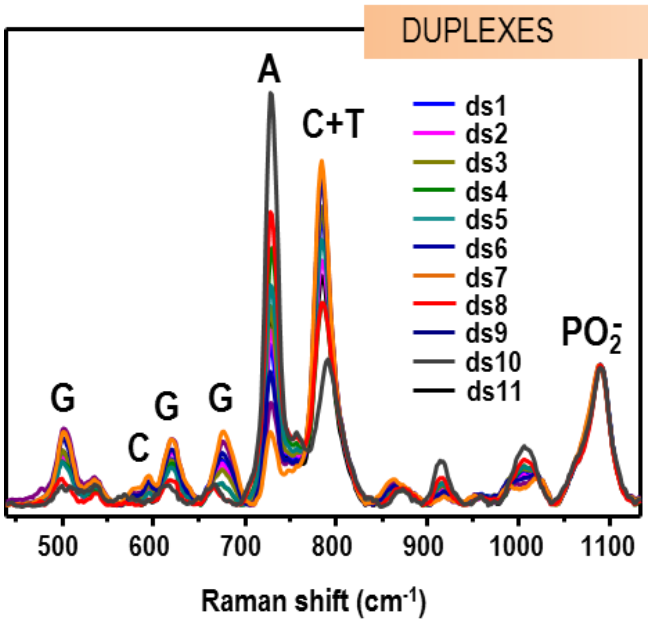
Differently, a significant reshaping of the overall spectral profile takes place when the individual strands hybridize in the double helix (**Figure 4.2**). Among others, we highlight the notable shift of the nucleobase ring-breathing modes of G, A, and C+T to ca. 676, 729 and 784 cm^{-1}, respectively (**Figure 4.2**), while the C ring bending vibration undergoes a significant intensity drop and a G ring deformation strongly emerges at ca. 620 cm^{-1}. [7a]

Based on the spectral analysis of single strands, it is possible to select the peak heights of the strong and well-separated features at ca. 599, 684 and 733 cm^{-1} as the quantitative spectral markers for C, G, and A relative content, respectively. On the other hand, no pure thymine features can be isolated in the SERS spectra. However, in principle it would be possible to extract the relative thymine content from the composed C+T ring breathing band at ca. 792 cm^{-1} once determined the initial C content via the analysis of the 599 cm^{-1} feature.



	%A	%T	%C	%G
ss1	28.6	19.0	23.8	28.6
ss2	19.0	28.6	28.6	23.8
ss3	33.3	23.8	23.8	19.1
ss4	19.1	33.3	38,1	9.5
ss5	38.1	19.0	23.8	19.0
ss6	9.5	19.0	42.9	28.6
ss7	23.8	38.1	14.3	23.8
ss8	38.1	23.8	23.8	14.3
ss9	47.6	33.4	9.5	9.5
ss10	33.4	47.6	9.5	9.5
ss11	33.3	33.3	0.0	33.3

Figure 4.1 SERS spectra of 21-mer single-stranded DNA sequences of different composition and base sequence, in the 550-1160 cm^{-1} spectral region. All spectra were normalized to the υPO_2^- band at ca. 1089 cm^{-1}. The illustrated spectra are from averaging 8 independent repeats per sample.



	%A	%T	%C	%G
ds1	23.8	23.8	26.2	26.2
ds2	26.2	26.2	23.8	23.8
ds3	26.2	26.2	23.8	23.8
ds4	35.7	35.7	14.3	14.3
ds5	30.9	30.9	19.1	19.1
ds6	19.1	19.1	30.9	30.9
ds7	19.1	19.1	30.9	30.9
ds8	40.5	40.5	9.5	9.5
ds9	0	0	50	50
ds10	50	50	0	0
ds11	26.2	23.8	21.4	28.6

Figure 4.2 SERS spectra of 21-bp double-stranded DNA chains of different composition and base sequence, in the 450-1150 cm^{-1} spectral region. All spectra were normalized to the υPO_2^- band at ca. 1089 cm^{-1}. The illustrated spectra are from averaging 8 independent repeats per sample.

Thus, in order to obtain the relative thymine content, an iterative process based on the outcome of the SERS measurements of the single strand probes in **Figure 4.1** was performed. The process applied for the relative quantification is summarized in **Figure 4.3**. The relative contributions of the C and T residues to the peak height of the 792 cm^{-1} band was progressively adjusted to an optimal ratio of (aC + 0.22 × bT), where a and b are the relative content of C and T in the strand, respectively. One cytosine contributes approximately 4 times more than one thymine to the peak height of the 792 cm^{-1} band, which is consistent with previous normal Raman studies reported in the literature.[9]

However, in the SERS spectra of the duplexes, the C-band at ca. 599 cm^{-1} can no longer be selected as a reliable spectral marker due to its intrinsic weakness and partial overlapping with the emerging G feature at ca. 620 cm^{-1}. Nonetheless, based on the Chargaff's first parity rule of nucleotide base composition in double-stranded DNA (i.e.; [A] = [T] and [G] = [C]),[10] it is still possible to estimate the relative GC and AT base content in DNA duplexes from the peak height ratios of G band at ca. 676 cm^{-1} and A band at ca. 729 cm^{-1} (normalized to the internal standard υPO_2^- band).

Thus, from the SERS analysis of the 21-mer strands and 21-bp duplexes, we calculated the normalized peak height values for each spectral marker corresponding to 1% of *(i)* individual nucleobase content in single-stranded DNA and *(ii)* combined AT or GC content in DNA duplexes (**Table 4.5**).

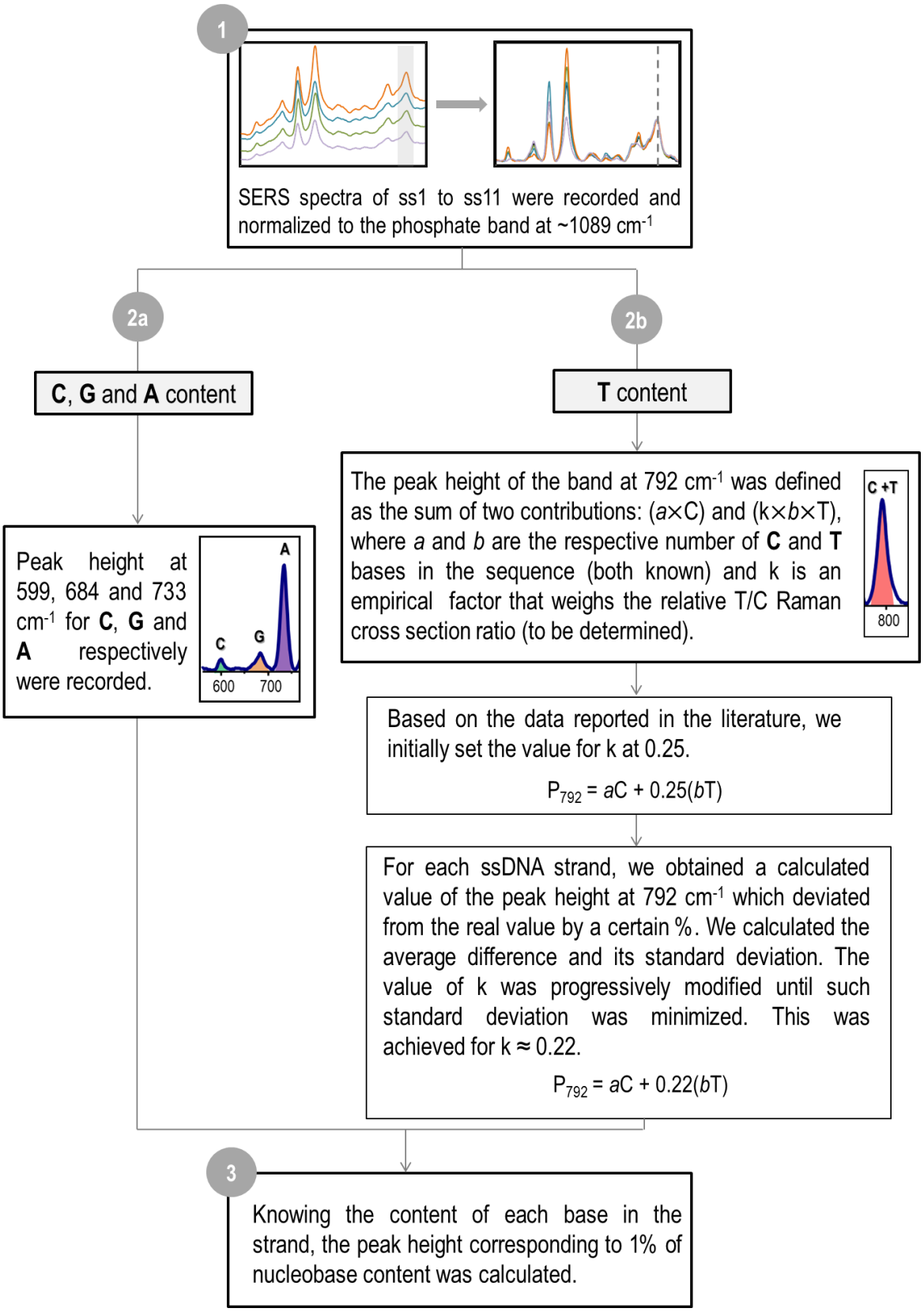


Figure 4.3 Steps followed for determining the peak height normalized to the internal standard and corresponding to 1% of nucleobase content in the DNA strand.

21-mer single-stranded DNA		Peak height normalized to the internal standard υPO_2^- and corresponding to 1% of nucleobase content in the strand
NUCLEOBASE	PEAK POSITION (cm^{-1})	
Cytosine - C	599	0.00766
Guanine - G	684	0.01716
Adenine - A	733	0.04655
Thymine - T	792	0.02410

21-bp double-stranded DNA		Peak height normalized to the internal standard υPO_2^- and corresponding to 1% of AT or GC content in the duplexes
NUCLEOBASE	PEAK POSITION (cm^{-1})	
G (= C)	676	0.01106
A (= T)	729	0.04550

Table 4.5 Estimated peak height values for individual nucleobase contents in the single-stranded DNA and combined AT/CG contents in double-stranded DNA. The values were firstly normalized to the internal standard (peak height values of the phosphate symmetric stretching at ~1089 cm^{-1}) and then referred to 1% content in the molecule.

Verification of the quantification parameters

The validity of such nucleobase quantification approach was verified by selecting three different DNA chains as unknown samples (**Table 4.6**). Specifically, SERS spectra were acquired of a 21-mer single-stranded sequence ($ssDNA_{21}$), a 21-bp synthetic duplex ($dsDNA_{21}$) and genomic DNA from calf thymus (ctDNA), as is shown in **Figure 4.4**.

The relative quantifications of each individual nucleobase for ssDNA and AT/GC contents in duplexes, based on the peak height ratios acquired from their direct SERS analysis and the values from **Table 4.5**, are represented in **Figure 4.5** together with the reference values.

NAME	BASE SEQUENCES (5' ⟶ 3')
ssDNA$_{21}$	CTA ACA TGA CTC TCT GTC CAG
dsDNA$_{21}$	CTC CCA GTG CGG ACC CTG CGT GAG GGT CAC GCC TGG GAC GCA
ctDNA	High-quality double-stranded DNA isolated from the thymus of calves

Table 4.6 Single-stranded and double-stranded DNA used as unknown samples.

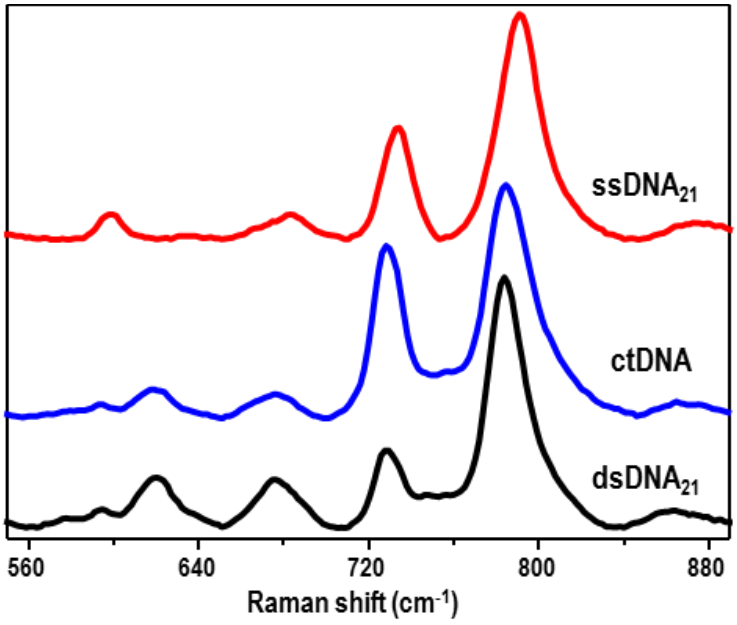


Figure 4.4 SERS spectra of ssDNA, dsDNA, and ctDNA in the 550-890cm^{-1} spectral region. The illustrated spectra are from averaging 8 independent repeats per sample.

The results show an outstanding accuracy of the direct SERS method at predicting the relative nucleobase contents in DNA chains of different composition and length (from ca. 7 nm long 21-bp dsDNA, to 10-15 kb DNA from calf thymus).

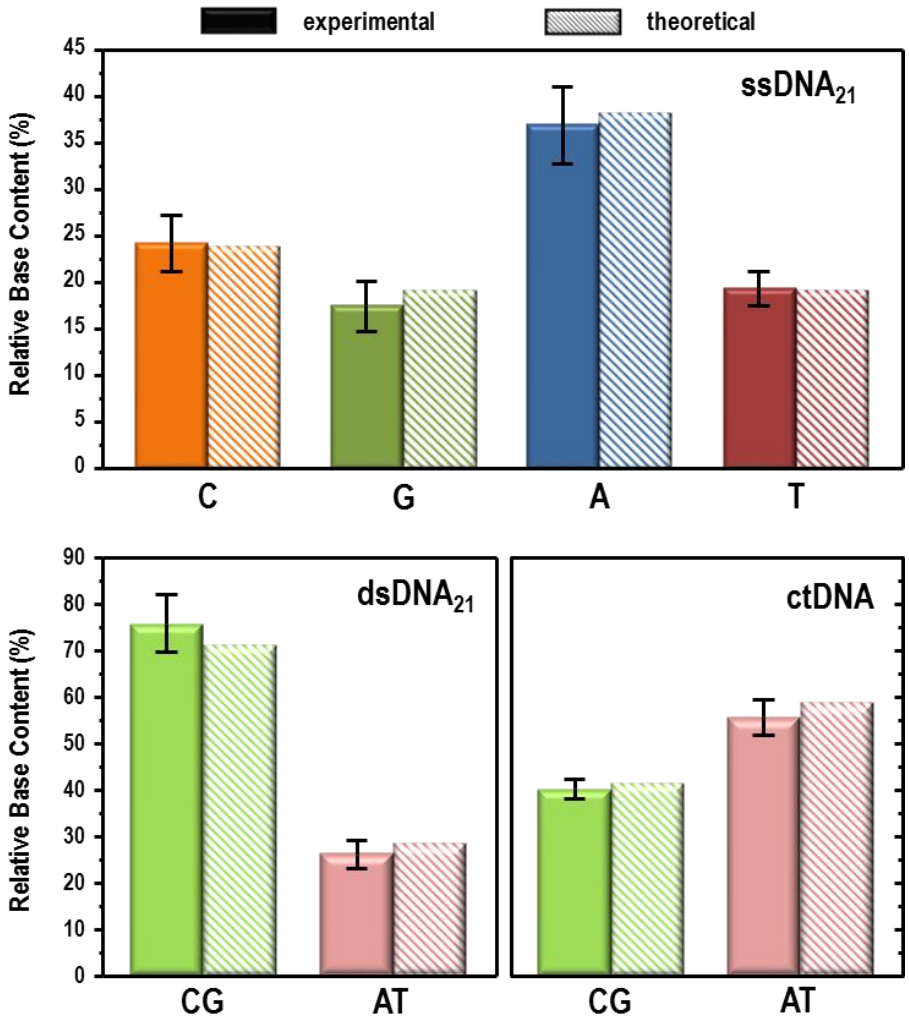


Figure 4.5 Quantification of nucleobase content (%) in the 21-mer single strand $ssDNA_{21}$ and quantification of CG and AT base content (%) in the 21-bp double-stranded DNA ($dsDNA_{21}$) and genomic DNA from calf thymus (ctDNA). Error bars equal to two standard deviations (N = 8).

4.3 CONCLUSIONS

In summary, this chapter describes a simple direct SERS method for the ultrasensitive, fast and label-free relative quantification of the nucleobase content in DNA. The peak heights of characteristic nucleobase markers were normalized to the internal standard (the phosphate symmetric stretching at ~1089 cm^{-1}) to yield spectral ratios that can be quantitatively correlated with the relative number of purine/pyrimidine bases in the chain. Different to indirect optical approaches based on DNA-nanoparticle conjugates, this simple, inexpensive and flexible method can straightforwardly be applied to DNA samples independently of their size (from short to long genomic DNA), conformation (single-stranded *vs* duplex), and base sequence.

4.4 EXPERIMENTAL SECTION

Materials

All materials were obtained from Sigma-Aldrich. 21-mer single-stranded DNA were all purchased from Eurofins Genomics (Ebersberg, Germany). Deoxyribonucleic acid from calf thymus (Type XV, Activated, lyophilized powder) was purchased from Sigma-Aldrich. Short dsDNA chains were prepared by annealing (90 °C for 10 min) equimolar mixtures of the corresponding complementary strands in PBS 0.3 M.

Sample preparation

Samples for SERS measurements were prepared by mixing 150 µL of AgNP@Sp (ca. 1.1 nM) with 12 µL of DNA solutions (10 µM for ssDNA; 5 µM for short dsDNA and 15 mg/mL for calf thymus DNA). Samples were left unperturbed for 2 hours, and then quickly sonicated before being investigated by SERS.

Instrumentation

A Renishaw InVia Reflex confocal microscope equipped with a 532 nm laser was used to carry out the SERS analysis. The laser was focused on the colloidal suspension with a macro lens providing an efficient spot of 0.6 mm (see in detail at section 2.4). This experimental set-up also avoids the uncontrolled formation of aggregates via light-induced trapping.[11] 8 different replications per each sample were acquired under the same experimental conditions (10 s exposure time, 20 accumulations and ca. 34.5 mW laser power at the sample). All SERS spectra were processed before the peak calculations; normalized at ~1090 cm^{-1} and baselined.

REFERENCES

[1] T. Vo-Dinh, A. M. Fales, G. D. Griffin, C. G. Khoury, Y. Liu, H. Ngo, S. J. Norton, J. K. Register, H. N. Wang, H. Yuan, *Nanoscale* **2013**, *5*, 10127-10140.

[2] X. R. Li, M. J. Scanlon, J. M. Yu, *Nucleic Acids Res.* **2015**, *43*, 3614-3625.

[3] P. E. Fournier, K. Suhre, G. Fournous, D. Raoult, *Int. J. Syst. Evol. Microbiol.* **2006**, *56*, 1025-1029.

[4] N.-P. Hua, T. Naganuma, *Electrophoresis* **2007**, *28*, 366-372.

[5] a) E. Papadopoulou, S. E. J. Bell, *Chem.-Eur. J.* **2012**, *18*, 5394-5400; b) L.-J. Xu, Z.-C. Lei, J. Li, C. Zong, C. J. Yang, B. Ren, *Journal of the American Chemical Society* **2015**, *137*, 5149-5154.

[6] a) S. Kumar, T. Verma, R. Mukherjee, F. Ariese, K. Somasundaram, S. Umapathy, *Chemical Society Reviews* **2016**; b) S. E. J. Bell, N. M. S. Sirimuthu, *Chemical Society Reviews* **2008**, *37*, 1012-1024.

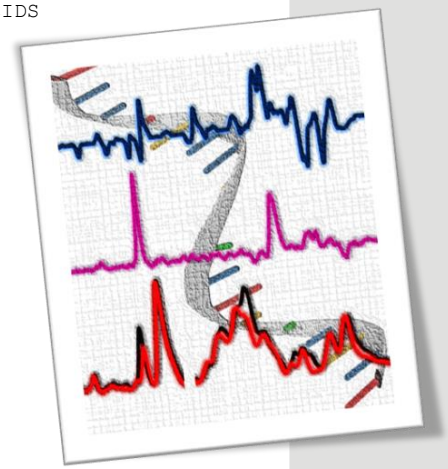
[7] a) L. Guerrini, Ž. Krpetić, D. van Lierop, R. A. Alvarez-Puebla, D. Graham, *Angew. Chem.-Int. Edit.* **2015**, *54*, 1144-1148; b) J. G. Duguid, V. A. Bloomfield, J. M. Benevides, G. J. Thomas, *Biophys. J.* **1996**, *71*, 3350-3360.

[8] M. Shanmugasundaram, M. Puranik, *Journal of Raman Spectroscopy* **2009**, *40*, 1726-1748.

[9] V. Reipa, G. Niaura, D. H. Atha, *RNA-Publ. RNA Soc.* **2007**, *13*, 108-115.

[10] E. Chargaff, *Experientia* **1950**, *6*, 201-209.

[11] A. Lehmuskero, P. Johansson, H. Rubinsztein-Dunlop, L. M. Tong, M. Kall, *ACS Nano* **2015**, *9*, 3453-3469.



CHAPTER 5

FAST OPTICAL CHEMICAL AND STRUCTURAL CLASSIFICATION OF RNA

ABSTRACT

As more biological activities of ribonucleic acids continue to emerge, the development of novel analytical tools for RNA identification and characterization is necessary to acquire an in-depth understanding of their functions and chemical properties. In this chapter, it is shown the capacity of label-free direct SERS analysis to access highly specific structural information of RNAs at the ultrasensitive level by using AgNP@Sp. This method represents a key advance in the ribonucleic acid analysis and will have a direct impact on a wide range of different fields, including medical diagnosis, drug design, and biotechnology, by enabling the rapid, high-throughput, simple and low-cost identification and classification of structurally similar RNAs.

OUTLINE

5.1 INTRODUCTION

Small RNAs

The role of ribonucleic acid (RNA) in cellular processes is one of the most dynamic and fast growing fields of biology.[1] New biological activities are continuously emerging besides RNA key functions in the gene regulation, protein synthesis or catalysis of biological reactions, all essential for the advance of molecular diagnostics. Among others, small RNAs are proved to play a major function in gene regulation, mainly in the form microRNAs (*miRNA*), and double stranded RNAs such as small interfering RNAs (siRNA) and short-hairpin RNAs (shRNA).[2]

miRNA are a class of non-protein-coding, small single-stranded (about 19−23 nucleotides) ribonucleic acids that play critical regulatory roles in almost every biological processes.[3] Mounting evidence associates aberrant expressions of *miRNA* with a number of diseases and genetic disorders as well as highlighting the importance of *miRNA* as both oncogenes and tumor suppressors.[3] Thus, *miRNA* can be used not only as biomarkers for cancer diagnostics and monitoring, but also exploited as new tools for cancer treatment *via* modulation of the *miRNA* pathways and activities[3a, 4] and in gene therapy.[2]

On the other hand, 21-23-mer siRNA and shRNA silence gene expression in the RNA interference (RNAi) process.[5] Growing understanding of genetic and the mechanisms of RNAi pathways fuelled the development of the RNAi-based

therapeutic strategy, where the highly specific mechanisms of sequence-specific gene silencing are exploited as efficient tools to design novel drugs that selectively interfere with disease-causing or disease-promoting genes.[6]

Additionally, aside from the four canonical bases, many more chemically modified nucleotides exist in multiple types of RNA increasing their chemical diversity (i.e., epigenetic modifications) to fulfil all the different RNA functions within the biological systems.[7] However, the lack of robust analytical methods has hindered, for decades, the in-depth understanding of their distribution and biochemical roles[8] and only very recent developments of new genomic technologies allowed to progress in this direction.[9]

Analytical methods for small RNAs

The analysis of small RNAs poses several challenges due to their intrinsic characteristics such as the small size of their sequences, which greatly complicates the application of PCR; their low abundance; and their sequence similarity.[3a] Thus, emerging technologies, such as the optical sensing through plasmons, have the potential of reshaping the detection and characterization scenario of nucleic acids[10] by offering simple, fast and highly sensitive platforms that overcome the limitations of PCR and Northern blot based microarrays.

While label-free direct SERS analysis of DNA has enjoyed in the last few years a renewed interest,[11] a surprisingly scarce number of studies are reported

for RNA.[12] Further, direct label-free SERS investigation of RNA has been limited to short single strands (microRNA), often in the form of simple homopolymeric or bi-polymeric sequences, and mostly performed by simple drying the sample directly onto plasmonic substrates with the consequent lack of spectral reproducibility. Thus, the development of fast, sensitive and efficient detection and characterization methods becomes more and more imperative.

In this chapter, it is demonstrated the potential of direct SERS analysis to screen composition and conformations of RNAs at the ultrasensitive level. Specifically, it is proven the capability of this technique to *i)* recognize the distinctive vibrational features of the structural characteristics of RNA including those of fully complementary double-stranded, small interfering and short hairpin RNA; and *ii)* classify highly similar RNA sequences with subtle differences such as a single-base variance (including that of uracil and thymine that defines RNA or DNA) or even a nucleobase modification (i.e., N^6-methyladenine and pseudouridine) in both single strands and duplexes.

5.2 RESULTS AND DISCUSSION

Acquisition of SERS spectra

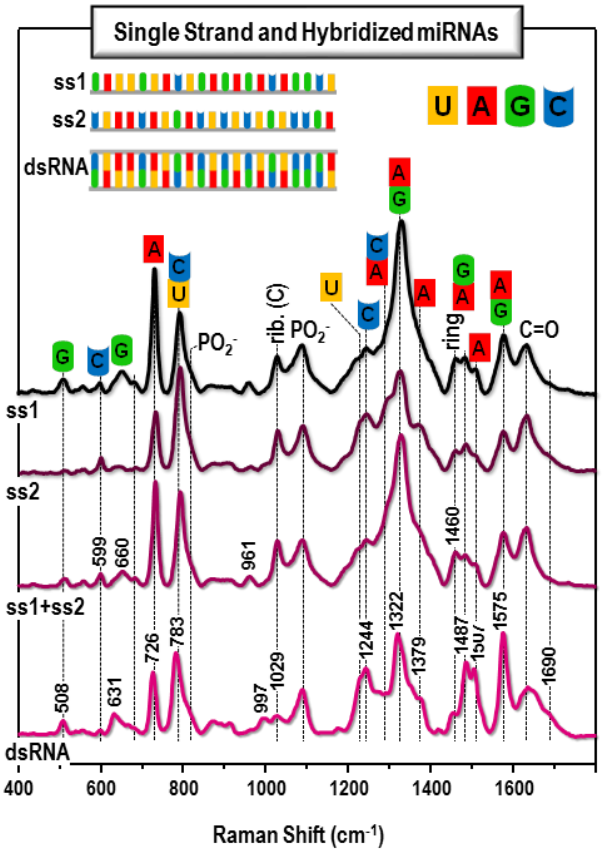
The microRNAs and single-stranded DNAs selected for this study are listed in **Table 5.1**. Duplex structures were prepared by annealing equimolar solutions of

ss1 and *ss2; ss1* and *ss5*; *ssDNA1* and *ssDNA2* yielding the corresponding *dsRNA*, *siRNA,* and *dsDNA,* respectively.

Vibrational assignment of the main bands appearing in the SERS spectrum of RNA is specified in **Figure 5.1.** In order to assign the vibrational features for single-stranded RNA, normal Raman studies reported in the literature,[13] as well as direct analysis of homopolymeric sequences (**Figure 5.2**), were compared. The study of that analysis is required for accessing and interpreting the rich structural information contained in the RNA SERS spectrum. For an appropriate comparison, all SERS spectra were normalized to the phosphate stretching band at 1089 cm^{-1}.

NAME	SEQUENCES (5' → 3')
ss1	GAU UGU ACU GAG AGA CAG GCU
ss2	AGC CUG UCU CUC AGU ACA AUC
ss3	AGC CUG UCU CUA AGU ACA AUC
ss4	AGC CUG ACU CUC AGU ACA AUC
ss5	CCU GUC UCU CAG UAC AAU CAG
ss^{m}A	^{m}AGC CUG UCU CUC ^{m}AGU ^{m}AC^{m}A ^{m}AUC
ssψ	AGC CψG ψCψ CψC AGψ ACA AψC
shRNA	GAU UGU ACU GAG AGA CAC CGG CUU GUC UCU CAG UAC AAU CAG
ssDNA1	GAT TGT ACT GAG AGA CAG GCT
ssDNA2	AGC CTG TCT CTC AGT ACA CTA
pA	AAA AAA AAA AAA AAA AAA AAA
pC	CCC CCC CCC CCC CCC CCC CCC
pG	GGG GGG GGG GGG GGG GGG GGG
pU	UUU UUU UUU UUU UUU UUU UUU
pT	TTT TTT TTT TTT TTT TTT TTT

Table 5.1 Names and nucleobase sequences of RNAs and single-stranded DNA. Nucleobase modifications are: N^6-methyladenine (^{m}A) and pseudouridine (ψ).



RAMAN SHIFT (cm^{-1})	VIBRATIONAL ASSIGNMENT ssRNA
512	G, ring deformation
599	C, ring vibration
631	U, ring vibration
660	G, ring breathing (C3?*endo/anti), A-form*
684	G, ring breathing (C2?*endo/anti), B-form*
730	A, ring breathing
792	C + U, ring breathing (mainly C)
814	O-P-O symm. stretching, *A-form*
913	Ribose, C-C stretching
960	Ribose-phosphate
1029	Ribose + C
1089	PO_2^-, symm. stretching
1191	G, C8-H bending
1231	U, ring stretching
1244	C, ring stretching
1292	A + C, ring stretching
1327	A + G, ring modes
1378	A, ring mode
1420	A,G
1460	C + U (A + G minor contributions), ring modes
1487	G + A, ring mode s (mainly G)
1507	A, ring mode
1575	A + G, ring mode s
1633	C=O (mainly U+C)
1720	C=O (G)

Figure 5.1 Hybridization of single stranded miRNAs sequences into full complementary duplexes. Averaged SERS spectra of miRNAs *ss1* and *ss2*, and double-stranded RNA (*dsRNA*). The (*ss1+ss2*) spectrum was digitally calculated by summing the two separated SERS spectra of *ss1* and *ss2*, both previously normalized to the 1089 cm^{-1} band. Raman shifts and the tentative vibrational assignment of the main SERS bands of single stranded RNA on AgNP@Sp colloids are listed in the left side table.

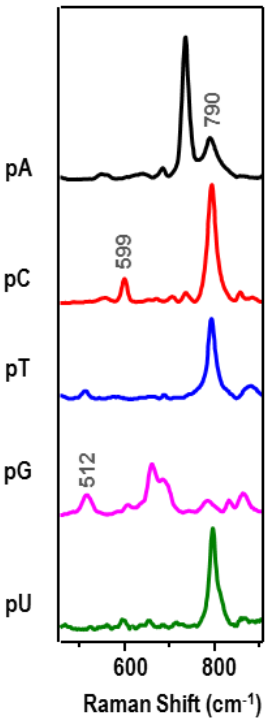


Figure 5.2 Averaged SERS spectra of 21-mer homopolymeric ssDNAs (*pA*, *pC*, *pT*, and *pG*) and 21-mer homopolymeric RNA (*pU*) in the ring breathing spectral region. The spectra were baseline corrected and normalized to the $\upsilon(PO_2^-)$ band at 1089 cm^{-1}.

Changes in RNA Conformation

To investigate the impact of the Watson-Crick hydrogen bonding and base stacking formation on the vibrational profiles of single-stranded RNAs, the full complementary miRNA *ss1,* and *ss2* were hybridized into the resulting duplex, *dsRNA,* by thermal annealing at 90ºC. The SERS spectra of these structures are illustrated in **Figure 5.1**, together with the digital sum of the individual single strands spectra (*ss1+ss2*). As a first approximation, *ss1+ss2* can be considered as the hypothetical equimolar mixture of non-hybridized *ss1* and *ss2* sequences and used as a reference to properly disclose the spectral reshaping associated with the structural rearrangement into the duplex conformation.

Phosphate bands at 814 cm^{-1} and 1089 cm^{-1}, sensitive to the ribose-phosphate chain backbone structure and helix conformation,[13a] do not undergo significant changes both in frequency position and relative intensity upon strand hybridization, suggesting that *ss1*, *ss2,* and *dsRNA* adopt a similar conformation when electrostatically adsorbed onto the plasmonic surface. On the contrary, a large set of spectral differences clearly emerges for the nucleotide bands, which can be summarized as follows:

I. As a result of the base stacking interactions, the ring-breathing modes of cytosine (792 cm^{-1}) and adenine (730 cm^{-1}) suffer a frequency-shift in *dsRNA* (-9 and -4 cm^{-1}, respectively). This is also consistent with the relative intensity increase of the U-ring vibration at 631 cm^{-1}.

II. The hydrogen bonding promotes an intensity decrease and a ~7 cm^{-1} red-shift of the carbonyl stretching (mainly U+C) and the rise of a new shoulder (~1690 cm^{-1}) ascribed to the C=O stretching of paired guanine (G).[13b, 14]

III. Changes in the molecular conformation are also evident in the relative intensity increase of both U and C ring stretching (~1231 and 1244 cm^{-1}), attributed to the RNA folding,[15] and the characteristic prominent hyperchromicity associated with H-bonding of A+G bands (1487 and 1575 cm^{-1}).[16]

It is worth noticing that single-stranded RNAs were thermally treated before the SERS analysis to enhance the spectral reproducibility. In fact, the lack of the methyl group in uracil (U) as compared with thymine (T) (**Figure 5.3**) decreases its hydrophobicity. As a result, although U preferably pairs with adenine (A), it can also weakly associate with almost any other nucleobases. Thus, one may expect that single-stranded RNAs in solutions would exist in a broad set of different configurations for loosely interacting strands. In fact, the SERS spectra of *miRNA* directly analyzed from stock solutions removed from the freezer show a significant sample-to-sample irreproducibility (**Figure 5.4A**), whereas *ds-, sh-* and *si-RNA*s do not suffer from such problem. In particular, such lack of homogeneous spectral response appears to be mostly circumscribed to the sharp ring breathing mode of adenine at ~ 732 cm^{-1} and, to a lesser extent, to the intense A+G ring mode at ~ 1327 cm^{-1}. Interestingly, SERS spectra of single-stranded DNAs do not reveal such sample-to-sample variability (**Figure 5.4B**). Such limitation was overcome by incubating *miRNA* solutions at 60ºC for 10 min before their addition to the silver colloids.

The gentle thermal treatment promotes the extension of the chain into a more linear conformation which has been previously demonstrated to drastically improve the spectral reproducibility in direct SERS analysis of single-stranded DNAs.[17] As shown in **Figure 5.4**, the sample-to-sample variability dramatically decreases when the above-mentioned thermal treatment is applied to *miRNA.*

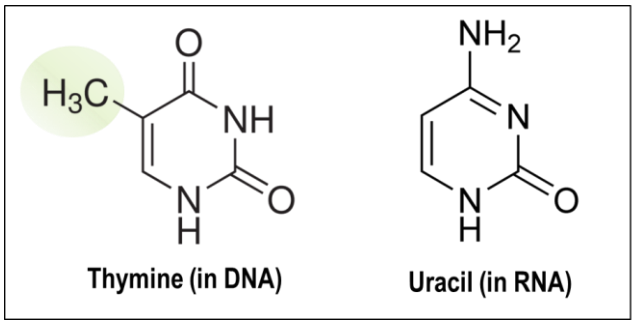


Figure 5.3 Molecular structures of Thymine and Uracil.

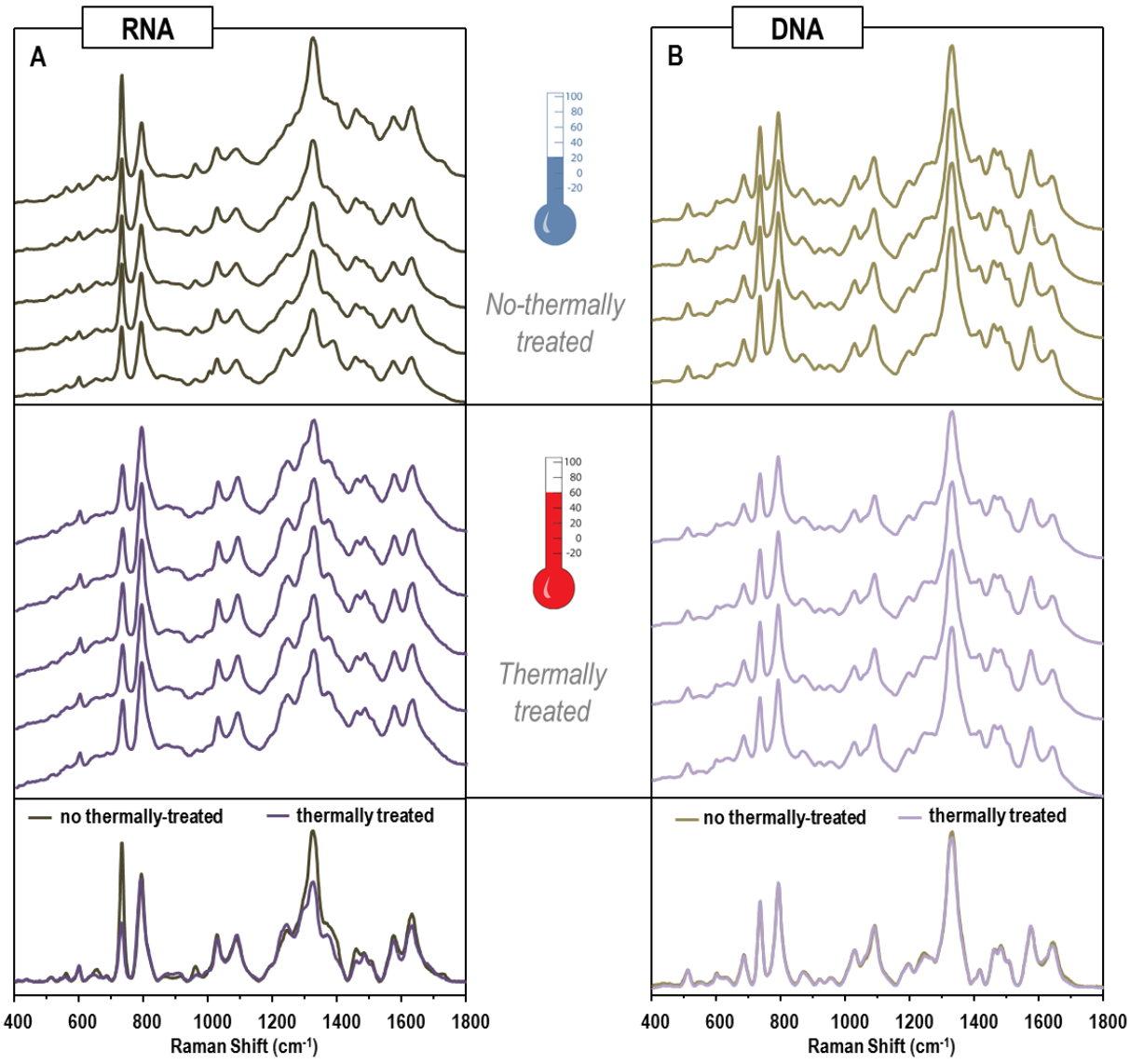


Figure 5.4 Original SERS spectra of nonthermally treated and thermally treated A) *ss2* and B) *ssDNA1* added to AgNP@Sp. The spectra are 5 different repetitions each of identically prepared samples. At the bottom, baseline-corrected averaged SERS spectra obtained by averaging the previous 5 spectra at the different conditions.

In addition to fully complementary duplexes, double-stranded RNAs exist in several other variants, such as small interfering (*siRNA*) and short hairpin (*shRNA*). *siRNA* are short double-stranded segments with two overhanging nucleotides, whereas *shRNA* also include in their structure an additional loop.[6a]

The investigated *siRNA* was formed by the hybridization between *ss1* and the partially complementary *ss5* strand (**Figure 5.5**) resulting in a duplex structure of 19 base pairs and two overhanging bases at both ends. The *ss5* sequence has an identical base composition of *ss2*, thus *siRNA* structurally differs from *dsRNA* only for the unpaired condition of the two extremities.

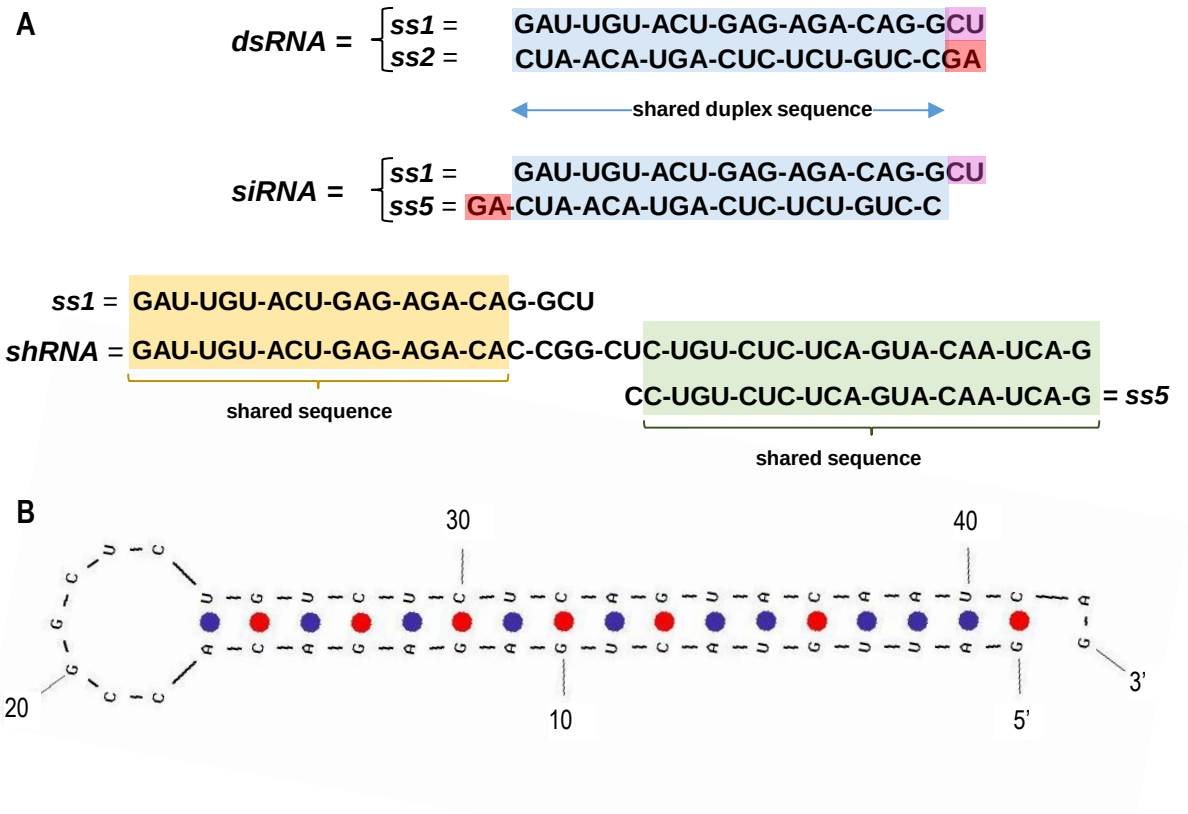


Figure 5.5 A) Base sequences and structural similarities *of dsRNA, siRNA, and shRNA.* B) Hairpin structure of the *shRNA,* obtained from *https://eu.idtdna.com/calc/analyzer.*

On the other hand, a partially self-complementary single-stranded RNA of 43 nucleobases was selected to yield a *shRNA* structure containing a stem of 17 base pairs, a loop of 7 nucleotides and a dinucleotide overhang at the 3' end (**Figure 5.5B**). Except for the central loop, the base sequences of the duplex stem and the overhang region are identical to those contained in the *siRNA,* while the base composition differs only for an additional extra cytosine (**Figure 5.5**).

Thus, it may be expected that the spectral reshaping of the SERS profile should be dominated by the different structural conformation of the RNA rather than its base composition.

This is corroborated by the analysis of the SERS spectra of *dsRNA* and *shRNA* (**Figure 5.6**). At first glance, these spectra show little deviations. Nevertheless, once subtracted, the corresponding SERS difference spectrum reveals a spectral profile which is remarkably similar to (*ss1+ss2)-dsRNA* throughout the whole spectral range (**Figure 5.6**) with few exceptions such as relatively less extended intensity changes of the composite broad C=O stretching and the ring-breathing modes at 726 and 783 cm^{-1}. In particular, in the *shRNA-dsRNA* difference spectrum, the A band at 726 cm^{-1} undergoes a minor spectral shift as compared to the 783 cm^{-1} ring breathing mode.

By assuming that the extension of such shifts is correlated with the relative number of mismatched nucleobases, we can then ascribe this observation to the larger % of unpaired C bases in *shRNA* (4 out of 10), with an additional

contribution of an extra unpaired U base, as compared to A, which has only 1 unpaired base out of the 11 contained in the whole hairpin sequence.

As previously described, the small interfering RNA is structurally analogous to *dsRNA*, except for the unpaired dinucleotide ends (GA and CU). Consistently, their SERS spectra only show subtle but still detectable differences, well-highlighted in the corresponding difference spectrum *siRNA-dsRNA* (multiplied by a factor of 4 in **Figure 5.6**). In this case, the spectral markers of H-bonding (such as the purine bands at 1244, 1487 and 1575 cm^{-1} and the broad carbonyl stretching feature) do not exhibit noticeable alterations and the detectable spectral changes appear to be mainly due to the base unpairing, as indicated by the red-shift and intensity decrease of the C+U ring breathing mode, the intensity increase at ~1330 cm^{-1} (ascribed to the disordering of the purine A-G bases[18]) and the weakening of the A feature at 1507 cm^{-1}.

A clear visual representation of the different degrees of similarities between *shRNA*, *siRNA,* and *dsRNA* is depicted in **Figure 5.7** using partial least squares-discriminate analysis (PLS-DA) to classify the SERS spectra of the different RNAs. The spectral profile of *shRNA* significantly diverges from those of *siRNA* and *dsRNA*, which are on the contrary very similar even though they can still be differentiated with 100% sensitivity and specificity.

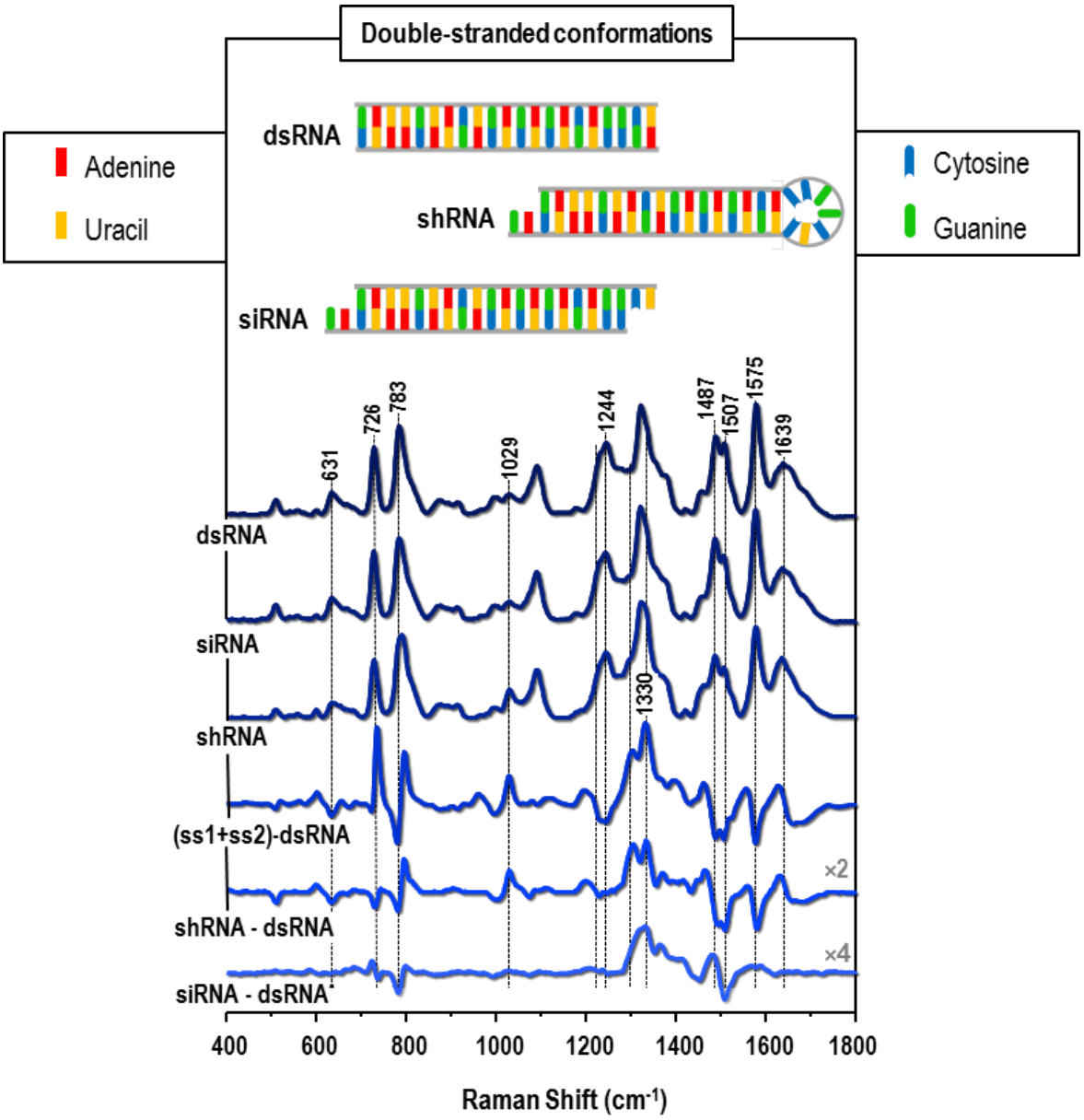


Figure 5.6 Averaged SERS spectra of double-stranded RNA (*dsRNA*), small interfering RNA (*siRNA*) and short hairpin RNA (*shRNA*). Difference spectra *(ss1+ss2)-dsRNA, shRNA-dsRNA* and *siRNA-dsRNA* obtained by digital subtraction of the SERS spectrum of *dsRNA* from the SERS spectra of *shRNA* and *siRNA;* and the pondered SERS spectrum *(ss1+ss2)* illustrated in Figure 5.1. For a better comparison, the difference spectra *shRNA-dsRNA* and *siRNA-dsRNA* were both multiplied by a factor of 2 and 4, respectively.

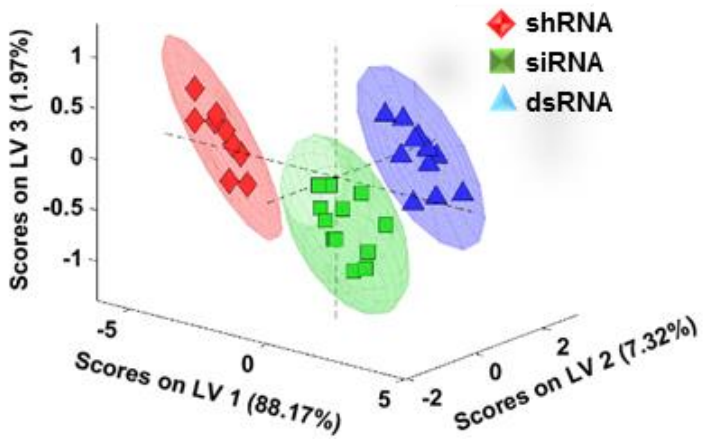


Figure 5.7 PLS-DA of SERS spectra from *dsRNA*, *shRNA* and *siRNA*. Score plot of loading vectors (LV) 1, 2 and 3 with 95% confidence ellipses, in which each spectrum is represented by a data point. The position of each point reflects characteristic features of the spectrum.

Changes in RNA Composition – Single-base mismatch

The diversity in nucleobase composition of *ss1* and *ss2*, complementary oligonucleotides with specular base content (6A, 5U, 3C and 7G for *ss1*; and 5A, 6U, 7C and 3G for *ss2*), is plainly reflected in their SERS spectra. The decrement of G and A bands and the intensity increase of the C and U features can be straightforwardly visualized in **Figure 5.1** when going from the SERS spectrum of *ss1* to that of *ss2*. On the other hand, more intriguing is the possibility to differentiate single-base mismatches in highly similar *miRNA* sequences. To test the ability of this method to achieve such result, SERS spectra of highly similar *miRNA*s (*ss2*, *ss3,* and *ss4*) with identical base sequence except for one single-base mismatch were acquired (see sequences at **Table 5.1**).

Specifically, if compared to *ss2*, one C was replaced with an A in the case of *ss3*, whereas for *ss4* the extra A was included to the detriment of one U. Subtraction of the SERS spectrum of *ss2* from those of *ss3* and *ss4* was performed to better show the spectral changes associated with the nucleobase substitution (**Figure 5.8**). In both difference spectra emerge intense positive bands associated with A motions (730, ~1327 and 1507 cm^{-1}) whereas no apparent changes occur in the bands with major G contribution, such as the broad ring breathing band at ~660 cm^{-1}.

On the other hand, a general intensity decrease of the C features, such as those at 599, 792 and 1029 cm^{-1}, is only observed in the case of *ss3* whereas the difference spectrum of *ss4* shows a decrease of the U-ring stretching at ~1231 cm^{-1} and a much milder intensity decrease of the ring breathing band at 792 cm^{-1}. The intense band at 792 cm^{-1} is indeed the result of different contributions mainly from ring-breathing modes of C (larger contribution)[15] and U,[18] with an additional overlapping with a weaker A feature.

Consistently, when one C is replaced with one A in *ss3*, the 792 cm^{-1} band appears in the difference SERS spectrum as an intense negative feature whereas when one U is removed in favour of A, as for *ss4*, the overall intensity is left largely unperturbed (i.e., the positive A contribution and the negative U contribution to the 792 cm^{-1} band approximately balance themselves out). Finally, both difference spectra reveal a relative intensity decrease of the C=O stretching, mainly ascribed

to U+C. Overall, these spectral changes are consistent with a C→A single-base mismatch in *ss3*, and U→A single-base mismatch in *ss4*.

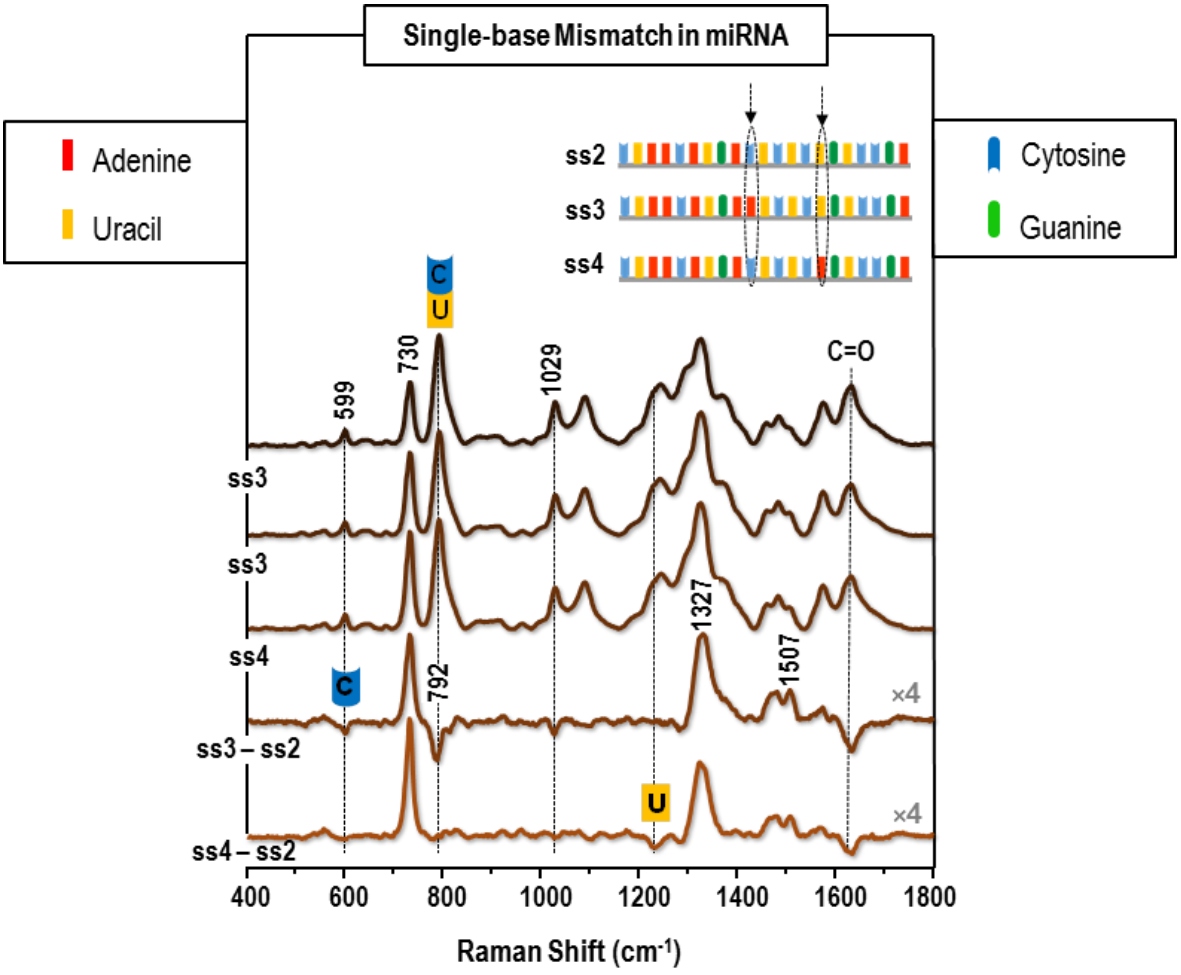


Figure 5.8 Averaged SERS spectra of single-stranded miRNAs *ss2*, *ss3*, and *ss4*. The spectra were baseline corrected and normalized to the υ(PO2-) band at ~1089 cm-1. Difference spectra (*ss3-ss2*) and (*ss4-ss2*) were obtained by digital subtraction of the SERS spectrum of *ss2* from the SERS spectra of *ss3* and *ss4*. For a better comparison, the difference spectra were both multiplied by a factor of 4. *miRNA* structures and sequences are outlined at the top of the figure.

Direct SERS single-base mismatch detection in *miRNAs* was limited so far to the simple differentiation *via* partial least squares discriminate analysis of *miRNA* members of the let-7 family achieved by investigating a relatively large amount of samples (1 μg) dried onto silver nanorods arrays.[12a]

Changes in RNA Composition – Chemical Modifications

In addition to the different content of the four canonical nucleobases, structural differences within RNAs can arise from chemical modifications, such as epigenetic alterations. Among all post-transcriptional modifications that have been identified in RNAs of all organisms, pseudouridine (ψ) is found to be the most abundant one and it is recognized as the "fifth" nucleoside in RNA.[19] The second most common RNA modification is nucleobase methylation.[20] In particular, N^6-methyladenine is the most abundant internal messenger RNA modification in eukaryotes, as well as in viruses.[8-9] To demonstrate the reliability of this direct method to recognize different types of nucleobase modifications in the SERS spectra of both *miRNA* and *dsRNA*, we acquired the characteristic vibrational changes imposed on the spectral profile of RNA by replacing A and U with their epigenetic variants N^6-methyladenine (^{m}A) and pseudouridine (ψ), respectively.

N^6-methylation of adenine (**Figure 5.9**) dictates characteristic spectral changes in the nucleotide spectral fingerprint[21] which are plainly resolved in the corresponding SERS profile of both single strand and duplex. **Figure 5.9B** illustrates the SERS spectra of *ss2* and *dsRNA* and their corresponding *ss^{m}A* and

ds^{m}A analogous, where the 5 adenine of the *ss2* were fully replaced by the *^{m}A* variant. It is important to note that the methyl group at the N6 position of adenosine does not prevent the Watson–Crick base-paring.[8, 22] The spectral changes imposed by the methylation on both *ss*- and *dsRNA* are striking.

First of all, a remarkable spectral redshift (> 10 cm^{-1}) of the A ring-breathing modes is observed and their relative intensity decrease. In the case of the duplex, this mode appears as a doublet at 726 cm^{-1}, ascribed to the unmodified A bases in the complementary *ss1* sequence, and at 740 cm^{-1}, assigned to the methylated adenines of the *ss^{m}A* strand.

Furthermore, 3-4 cm^{-1} redshifts are observed in several other A related bands, such as those at 1487 cm^{-1} and 1575 cm^{-1}, while the very intense A+G feature at ~1327 cm^{-1} undergoes a notable intensity drop and the 1507 cm^{-1} adenine band almost disappears from the spectra.

These spectral changes nicely match those described in the normal Raman studies of the N^{6}-methylation of the individual adenine nucleoside.[21] Interestingly, while the alterations of the vibrational profile are largely limited to A related bands in the *miRNA* spectra, for the corresponding duplex we also register additional differences such as the broadening and red-shift of the carbonyl stretching vibration (mainly U) from 1635 cm^{-1} to 1648 cm^{-1}, as well as the more subtle 2 cm^{-1} blue-shift of the C+U ring breathing mode

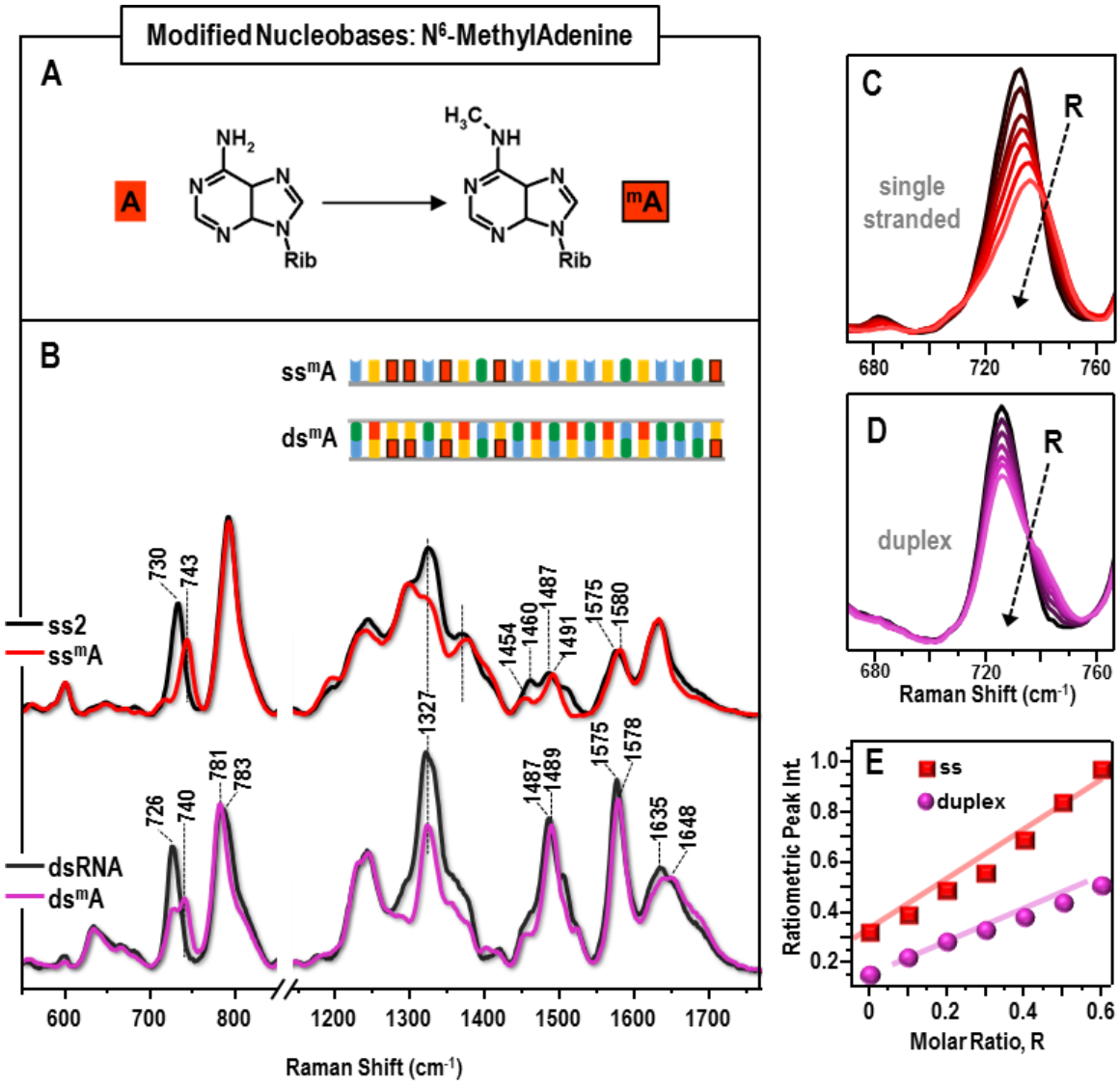


Figure 5.9 A) molecular structures of adenosine (A) and N6-methyladenosine (mA). B) Averaged SERS spectra of *ss2* and *dsRNA*, and the analogous *ss^{m}A* and *ds^{m}A* samples where the 5 adenines were replaced with the N6-methyladenine variant (thus, *ds^{m}A* contains 5 *^{m}A* from *ss^{m}A* and 6 A from *ss1)*. C) and D) Detail of the A ring breathing spectral region for the SERS spectra of *ss2*+*ss^{m}A* and *dsRNA*+*ds^{m}A* mixtures at different molar ratios, R ([*ss^{m}A*] *vs.* total [*ss^{m}A*]+[*ss2*] content, and [*ds^{m}A*] *vs.* total [*ds^{m}A*]+[*dsRNA*] content, respectively). Molar ratios decrease going from the darkest spectrum to clearest one as follows: R = 0; 0.1; 0.2; 0.3; 0.4; 0.5 and 0.6. E) Ratiometric peak intensities I_{743}/I_{729} and I_{743}/I_{725} for *miRNA* and *dsRNA* mixtures, respectively, against the corresponding molar ratios. RNA structures and sequences are outlined at the top of the figure (red shapes correspond to Adenine, yellows to Uracil, greens to Guanine and blues to Cytosine).

Once more, these results are consistent with the formation of the A-U base pairs in the double helix leading to different perturbations of the U electronic structure depending on the structural properties of the Watson-Crick counterpart (A *vs.* ^{m}A).

Pseudouridine (ψ) is derived from uridine *via* base-specific isomerization.[19a] This process introduces an additional hydrogen bond donor, N-H, in the nucleobase (**Figure 5.10A**) which lays the foundation for the special properties of ψ relative to U such as the enhanced rigidity conferred to the phosphodiester backbone and the structural stabilization of the neighboring nucleosides, *via* increased base stacking, for both single- and double-stranded RNAs.[19]

As for *ss^{m}A*, the analysis of the SERS spectra of pseudouridine-containing *miRNA*, *ssψ*, (**Figure 5.10**) reveals spectral changes to the uracil bands that qualitatively agree with normal Raman studies of the individual nucleosides,[23] these include:

- The drastic red-shift of the carbonyl stretching band.
- The intensity perturbation and 2 cm^{-1} shift of the C+U ring breathing mode.
- The large reshaping of the spectral contour of the broad band in the 1220-1250 cm^{-1} range, containing U-ring vibrations.[23]

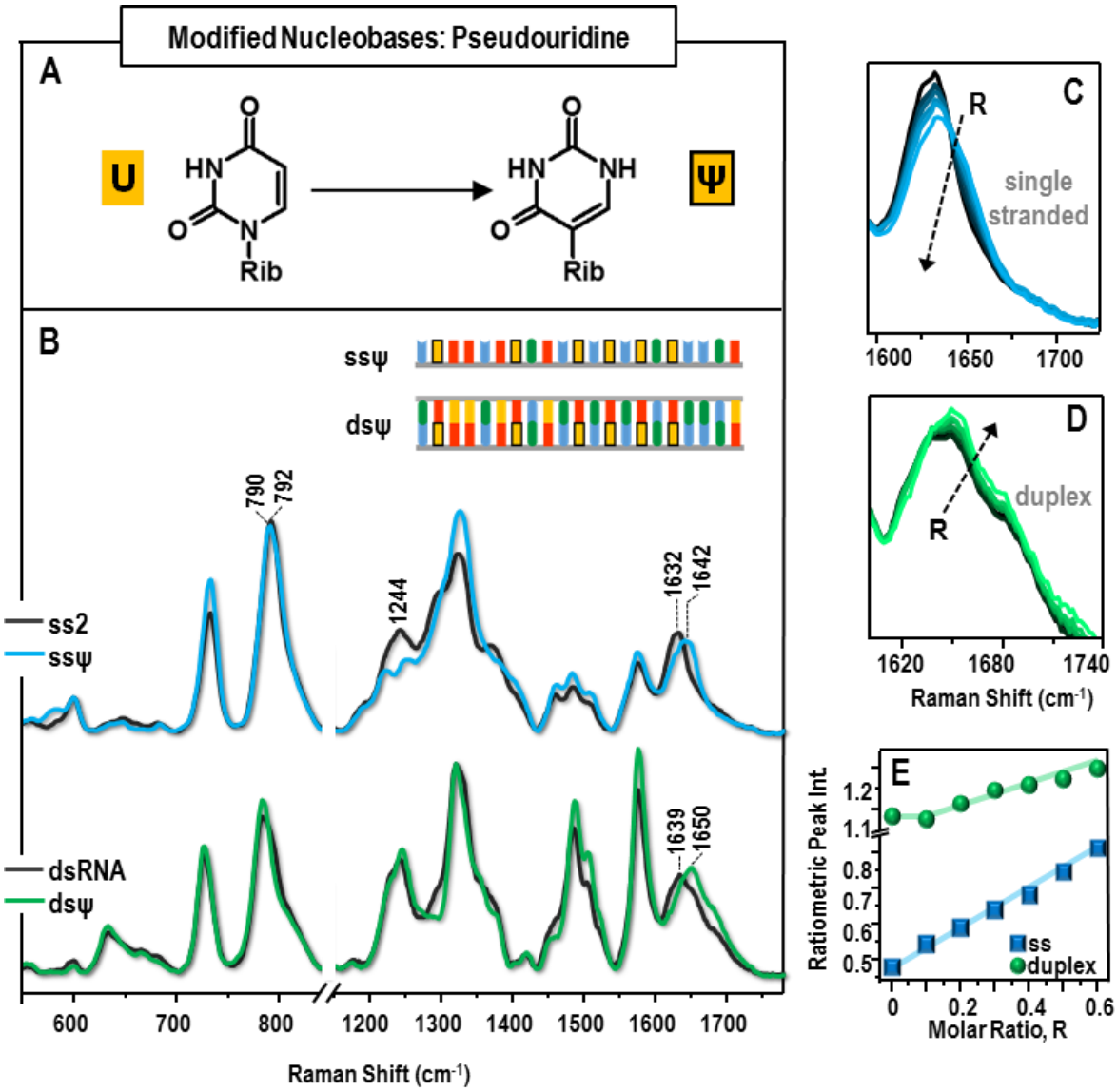


Figure 5.10 A) Molecular structures of uridine (U) and pseudouridine (ψ). B) Averaged SERS spectra of *ss2* and *dsRNA*, and the analogous *ssψ* and *dsψ* samples where the 6 uracils of *ss3* were replaced with the pseudouridine variant (thus, *dsψ* contains 6 ψ from *ssψ* and 5 U from *ss1)*. C) and D) Detail of the carbonyl spectral region for the SERS spectra of *ssψ*+*ss2* and *dsRNA*+*dsψ* mixtures at different molar ratios, R ([*ssψ*] *vs.* total [*ssψ*]+[*ss2*] content, and [*dsψ*] *vs.* total [*dsψ*]+[*dsRNA*] content, respectively). Molar ratios decrease going from the darkest spectrum to clearest one as follows: R = 0; 0.1; 0.2; 0.3; 0.4; 0.5 and 0.6. E) Ratiometric peak intensities I_{1650}/I_{1625} and I_{1655}/I_{1630} for *miRNAs* and *dsRNA* mixtures, respectively, against the corresponding molar ratios. RNA structures and sequences are outlined at the top of figure (red shapes correspond to Adenine, yellows to Uracil, greens to Guanine and blues to Cytosine).

In this case, however, it is also registered a general intensity increase of purine bands, mostly ascribed to A (i.e., 730, 1327 and 1575 cm^{-1}), even though with no significant spectral shifts. Noteworthy, this base-neighboring effect of pseudouridine is far less explicit in the SERS spectrum of the corresponding duplex, *ds*ψ, which however remains the main pseudouridine spectral marker (i.e., the large red-shift of the carbonyl stretching band).

Once identified the A ring breathing band and the carbonyl stretching as the major spectral markers of N^6-methyladenine and pseudouridine variants, respectively, the possibility to quantitatively correlate the spectral changes in the SERS profiles with the relative contents of modified nucleobases in the biomolecule were investigated.

To do so, mixtures of modified and unmodified *miRNAs* (*ss2*+*ss*^{m}*A* and *ss2*+*ss*ψ) as well as duplexes (*dsRNA*+*ds*^{m}*A* and *dsRNA*+*ds*ψ) at different molar ratios, R (where R is equal to the concentration of the modified sample *vs* the overall RNA content) were prepared.

The corresponding SERS spectra are illustrated in **Figure 5.11** and **5.12** whereas the characteristic spectral regions containing the A ring breathing and C=O stretching markers are detailed in **Figure 5.9C, D** and **Figure 5.10C, D**, respectively.

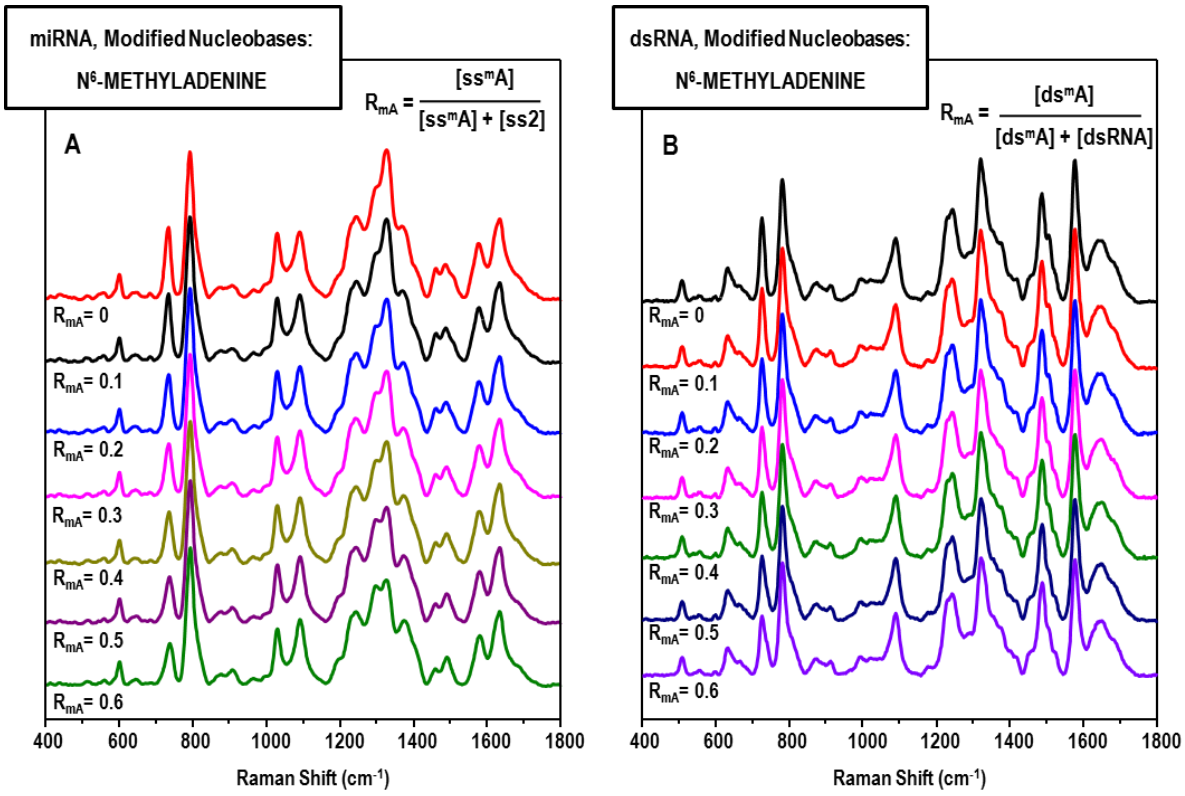


Figure 5.11 Averaged SERS spectra of A) *ss2* and *ss^{m}A* mixtures and B) *dsRNA* and *ds^{m}A* mixtures at different molar ratios, R. The overall concentration was kept constant through the whole set of measurements.

The relative SERS intensities I_{743}/I_{729} for *ss^{m}A,* and I_{743}/I_{725} for *ds^{m}A* are plotted against R in **Figure 5.9E**. Similarly, the values I_{1650}/I_{1625} for *ssψ* and I_{1655}/I_{1630} for *dsψ* were monitored at different molar ratios as illustrated in **Figure 5.10E**. The results show outstanding linear correlations between the content of modified bases and the extent of spectral perturbations introduced in the SERS spectra, with sensitivity below the single-base discrimination for N^6-methyladenine in both *ss*- and *dsRNA*, and for pseudouridine in single stranded sequences.

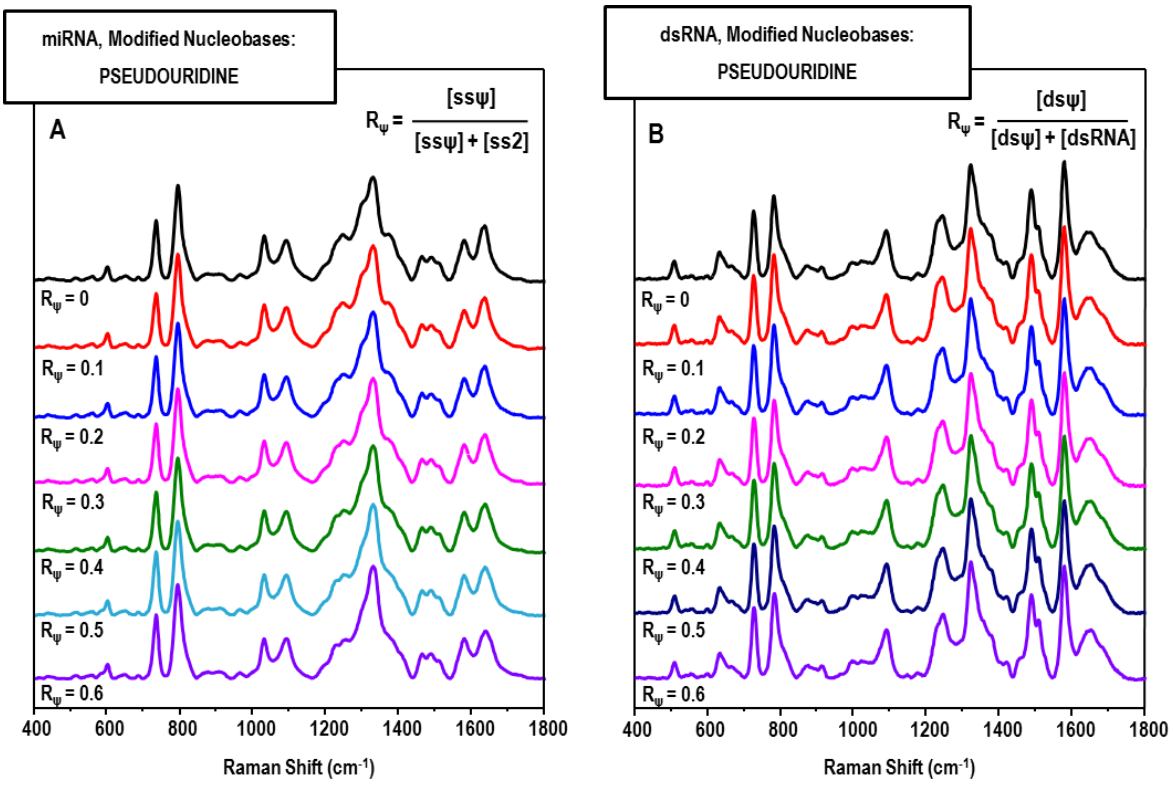


Figure 5.12 Averaged SERS spectra of A) *ss2* and *ssψ* mixtures and B) *dsRNA* and *dsψ* mixtures at different molar ratios, R. The overall concentration was kept constant through the whole set of measurements.

Comparison of DNA and RNA analogous

The reliability of direct SERS to fully extract the rich structural information contained in the Raman spectra of nucleic acids was also exploited to differentiate and structurally characterize RNA and DNA analogous. To do so, it was investigated single and double-stranded DNA and RNA with identical sequence, except of course for the thymine bases replaced by their uracil counterparts and the deoxyribose sugar moieties substituted by the ribose.

The direct comparison of SERS spectra is illustrated in **Figure 5.13** and unravels a set of spectral differences that can be broadly classified into the following classes: spectral changes associated with (i) different backbone conformations, (ii) nucleobase modifications (i.e., removal of the methyl group from the T structure yielding the U base) and (iii) indirect perturbations on the remaining A, G, and C nucleobases imposed by (i) and (ii).

Regarding the comparison of the unpaired strands, the analysis of backbone conformation markers indicates that *miRNA* exist mainly in the A-form secondary structure, as revealed by the intense spectral markers at ~660 cm^{-1} (G band diagnostic of the C3´-endo/anti ribofuranose puckering) and the presence of a shoulder at ~814 cm^{-1}, ascribed to the sugar-phosphate backbone vibration[14, 24] (**Figure 5.13**). However, although the A-form helicity is found to be predominant, the coexistence of the G ring breathing band at ~684 cm^{-1}, characteristic of the C2´-endo/anti conformation of the B-type backbone, suggests that some nucleotide residues can adopt additional conformations.[14, 24]

Differently, *ssDNA* SERS spectra only display a narrow G band at 684 cm^{-1} with no apparent phosphate contribution at 814 cm^{-1}, which is consistent with a B-like conformation.[13a] It is also worth noticing the different spectral contour of the weak bands in the 850-1050 cm^{-1} region, originating mainly from vibrations of the sugars, which are also sensitive to the backbone conformation.[13a]

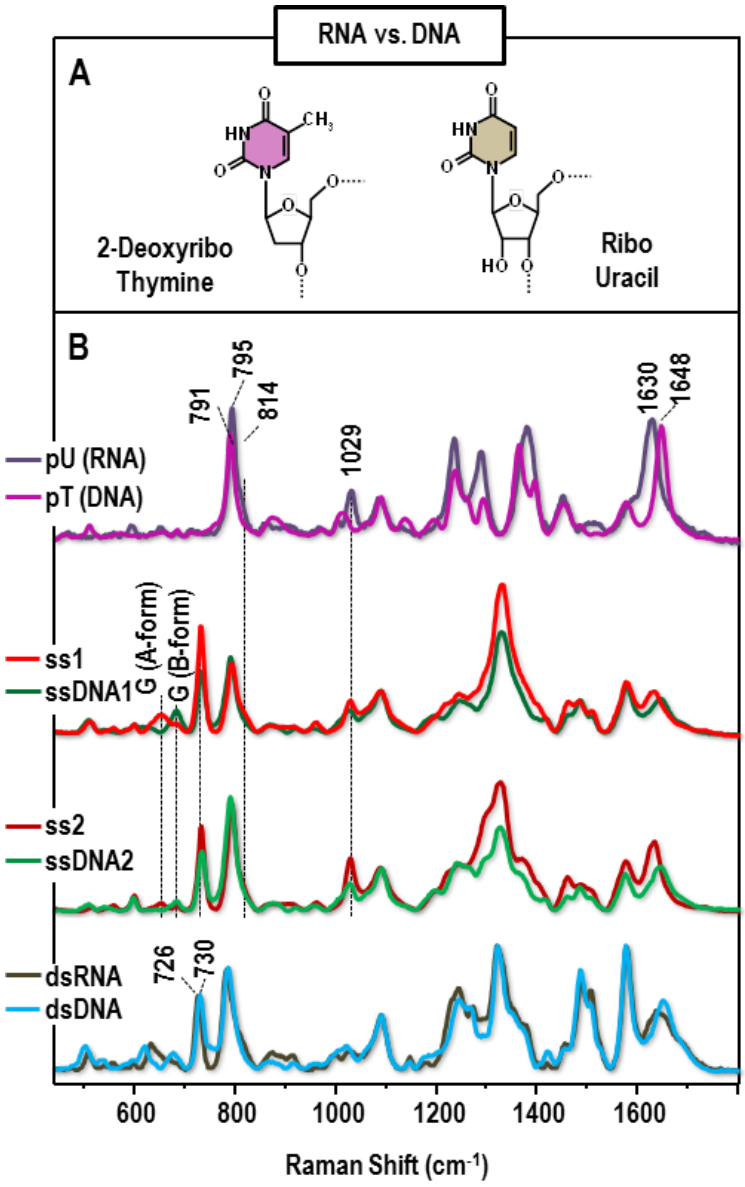


Figure 5.13 A) RNA and DNA structures. B) Averaged SERS spectra of uracil RNA and thymine DNA homopolymers (*pU* and *pT*, respectively); the two complementary micro-RNAs *ss1* and *ss2*, and the two complementary single-stranded *ssDNA1* and *ssDNA2*. The averaged SERS spectra of the corresponding *dsRNA* and *dsDNA* duplexes are also shown.

The replacement of T with U in the RNA structure is mainly revealed in the SERS spectra by the relative intensity increase and dramatic blue-shift of the broad C=O stretching as well as for the 2 cm^{-1} shift of the C+U ring breathing

mode, (see **Figure 5.14** for details of the 600-950 cm^{-1} spectral region). This agrees with what observed by directly comparing the SERS spectra of homopolymeric uracil RNA, *pU*, and thymine DNA, *pT* (**Figure 5.13**). Furthermore, the analysis of the homopolymers suggests that the relative intensity increase of the feature at ~1029 cm^{-1} observed for *miRNA* may also be due to the U presence in the structure (**Figure 5.13**).

Similar considerations in terms of helicity form and U/T nucleobase modification can be derived when comparing the SERS spectra of RNA and DNA duplexes (**Figure 5.13**). Additionally, we also observe a significant 4 cm^{-1} down-shift of the A ring breathing mode (~730 cm^{-1}) in *dsRNA* which does not occur for *miRNA* (better visualized in **Figure 5.14**). This experimental result is congruent with the different electronic perturbation that U and T impose on the A structure upon duplex formation *via* base stacking and H-bonding, as also reported in the normal Raman studies of bi-polymeric A-U and A-T nucleic acids.[25]

Interestingly, in the case of duplexes, the remaining features of the spectra show minor or null differences whereas for *miRNA* and *ssDNA* some remarkable changes are observed in several bands which are not strictly related to either the backbone conformation or T/U nucleobases. Specifically, these spectral differences can be mainly described as the variation of the intensity ratio between purine and pyrimidine bands, such as the A ring breathing band and the intense

A+G ring band centred at 1327 cm^{-1} (purine), and the C+U ring breathing feature (pyrimidine).

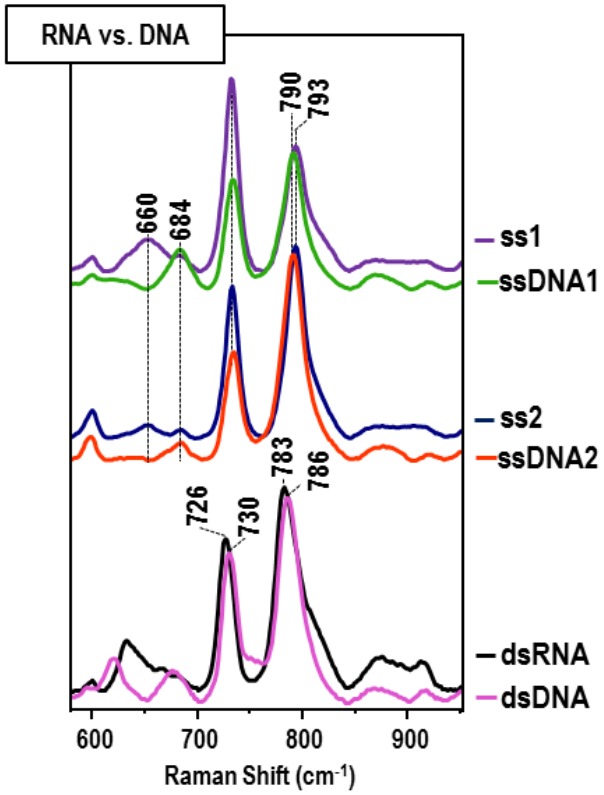
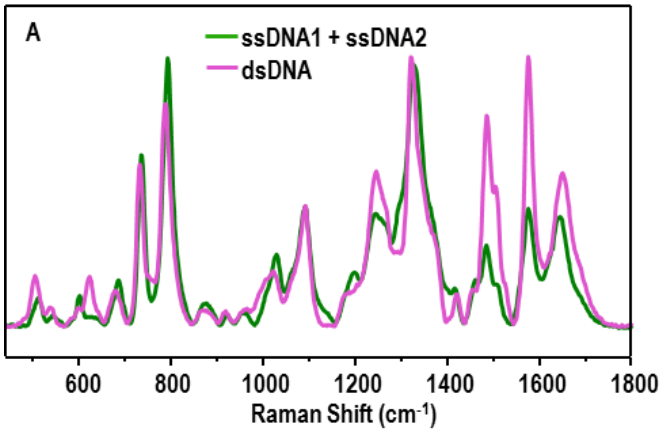


Figure 5.14 Details of the 580-950cm^{-1} spectral range for the averaged SERS spectra of the complementary microRNA *ss1* and *ss2*, the complementary single-stranded *ssDNA1* and *ssDNA2*, and their corresponding *dsRNA* and *dsDNA* duplexes.

Overall, the data showed that the relative intensity of these purine bands decreases in a very similar fashion when *miRNAs* are thermally treated (**Figure 5.4**) and uridine replaces pseudouridine in single stranded sequences (**Figure 5.10**). Moreover, as clearly visible in **Figure 5.13**, SERS spectra of *dsRNA* and *dsDNA* only differ from spectral alterations associated with nucleobase modification (T *vs.* U) and backbone conformation with minor deviations with

regard to the general purine *vs.* pyrimidine relative intensities. Contrary, for single stranded sequences the difference is remarkable (see for instance the peak height ratios between the A and C+U breathing modes, **Figure 5.15**). All these experimental evidence converge to indicate that *dsRNA* and *dsDNA* share similar adsorption conformations on silver surfaces (i.e., the distribution of SERS enhancements profited by the ensemble of nucleobases is similar for RNA/DNA duplexes of identical base sequence).

On the other hand, when individual *miRNA* strands are not forced to pair into the rigid double helix, they are likely free to adopt a different range of possible conformations which can be restricted by an appropriate thermal treatment or modified upon inclusion of a nucleobase variant conferring, such as pseudouridine, an enhanced rigidity to the phosphodiester backbone. Controlling the nucleic acid conformation onto the metallic surface is a key feature in determining the overall spectral reproducibility as well as affecting the final spectral profile, and must keep into careful consideration in the direct SERS analysis of single-stranded RNAs.



B

SAMPLE	I ring br. C+U/T / I ring br. A
ssDNA1	1.22
ssDNA1 (therm. treat.)	1.29
ss2	0.78
ss2 (therm. treat.)	1.75
ssψ (therm. treat.)	1.36
ssDNA1 + ssDNA2	1.19
ss1 + ss2	0.90
dsDNA	1.39
dsRNA	1.32
dsψ	1.37

Figure 5.15 A) Averaged SERS spectrum of *dsDNA* and digitally calculated SERS spectrum of *ssDNA1+ssDNA2*, both normalized to the phosphate band at 1089 cm^{-1}. The *ssDNA1+ssDNA2* pondered spectrum was obtained by summing the two separated SERS spectra of *ssDNA1* and *ssDNA2*. B) Peak height ratios between the C+U or C+T ring breathing mode (~730 cm^{-1}) and the A ring breathing mode (~792 cm^{-1}). These values were calculated from the SERS spectra illustrated in this figure as well from Fig. 5.4, Fig. 5.10 and Fig. 5.13.

5.3 CONCLUSIONS

In summary, in this chapter, we demonstrated the potential of label-free direct SERS analysis to fully disclose the unique vibrational information of a variety of different small RNAs with high sensitivity. As vibrational spectra and molecular features are intrinsically linked, it was possible to identify and classify highly similar RNA structures based on both their conformation and composition. Among others, it was reported the first example of SERS recognition of fully complementary duplexes, as well as a short hairpin and small interfering RNAs. We further diversified microRNAs sequences with single-base mismatches.

Nucleobase variants such as pseudouridine and N^6-methyladenine were recognized and quantified with single-base sensitivity within both single strands and duplexes. Finally, direct comparison of structurally analogous RNA and DNA molecules also permitted to characterize distinctive signatures of their corresponding sugar moieties as well as their secondary structure.

This work represents a step forward in the ribonucleic acid analysis by SERS, paving the way for the final translation of its analytical potential to the in-depth investigation of these intriguing biomolecules. Owing to the extremely rich structural information provided by this simple, highly sensitive and low-cost sensing approach, this method could have an important impact in basic research, therapeutic, drug design and diagnosis applications.

5.4 EXPERIMENTAL SECTION

Materials

All materials were of highest purity available and obtained from Sigma-Aldrich. RNA sequences were purchased from Eurofins Genomics (Germany) except for *shRNA*, purchased from IDT Integrated DNA Technologies (Denmark). 200 μM stock solutions were prepared by dissolving into Milli-Q water. The microRNAs and single-stranded DNAs selected for this study are listed in **Table 5.1**. Duplex structures were prepared in 0.3 M PBS by annealing at 90°C for 10 minutes of equimolar solutions of *ss1* and *ss2*; *ss1* and *ss5*; *ssDNA1* and *ssDNA2*

yielding the corresponding *dsRNA*, *siRNA* and *dsDNA*. Short hairpin RNA structure (*shRNA*) was also self-hybridized in PBS 0.3 M by annealing at 90°C for 10 minutes. The final concentration of the duplex structures in the stock solutions was 20 μM. The solutions were stored at -20 °C until required.

Sample preparation

AgNP@Sp was prepared as is detailed in chapter 2. Samples for SERS measurements were prepared as follows: 100 μL of AgNP@Sp ([NP] ~0.3 nM) were mixed with 4 μL of RNA or DNA aqueous solutions (10 μM for single-stranded sequences and 5 μM for duplexes, respectively). The colloidal suspension immediately turned its color from yellow to orange, revealing the nanoparticle aggregation. The samples were left to equilibrate for 3 h and redispersed by quick sonication before acquiring the SERS spectra. It is worthy of note that the sensitivity of the method can be easily improved down to ~30 nM of *miRNA* and ~10 nM of *dsRNA* (corresponding respectively to ~21 and ~14 ng per 100 μL of investigated colloidal sample) by diluting the nanoparticle suspension 4 times. Further, the overall amount of sample required to perform the SERS study can be drastically reduced, below the pg level of RNA, *via* straightforward implementation of the SERS method with microfluidics devices, as described in our previous study.[11g]

Instrumentation

SERS experiments were conducted using a Renishaw InVia Reflex confocal microscope and a 532 nm laser. A long-working distance objective (0.17 NA, working distance 30 mm) was used to focus the laser on the sample (laser power at the sample of 6.9 mW). All SERS spectra illustrated in the article were obtained by averaging the SERS responses of 5 different replications per each sample (15 accumulations, 10 s exposure time). These replications were obtained by combining the same colloids (prepared the day before the measurement and left aging overnight) with the specific RNA/DNA solution.

Partial least squares Discriminant Analysis (PLS-DA) is a linear classification method for highlighting differences between multivariate data sets of different classes. It encompasses the partial least squares regression properties to fine-tune a principal component analysis model. Acquired SERS spectra were imported to the PLS toolbox (Eigenvector Research, Inc.) in Matlab for PLS-DA modeling. The spectra were normalized using Multiplicative Scatter Correction[26] and mean centred before being imported for the construction of the model.

REFERENCES

[1] a) P. A. Sharp, *Cell* **2009**, *136*, 577-580; b) C. E. Dumelin, Y. Y. Chen, A. M. Leconte, Y. G. Chen, D. R. Liu, *Nat. Chem. Biol.* **2012**, *8*, 913-919.

[2] L. He, G. J. Hannon, *Nat. Rev. Genet.* **2004**, *5*, 522-531.

[3] a) H. F. Dong, J. P. Lei, L. Ding, Y. Q. Wen, H. X. Ju, X. J. Zhang, *Chem. Rev.* **2013**, *113*, 6207-6233; b) D. P. Bartel, *Cell* **2004**, *116*, 281-297.

[4] R. Garzon, G. A. Calin, C. M. Croce, in *Annual Review of Medicine, Vol. 60*, **2009**, pp. 167-179.

[5] a) D. D. Rao, J. S. Vorhies, N. Senzer, J. Nemunaitis, *Adv. Drug Deliv. Rev.* **2009**, *61*, 746-759; b) A. Z. Fire, *Cell Death Differ.* **2007**, *14*, 1998-2012.

[6] a) D. H. Kim, J. J. Rossi, *Nat. Rev. Genet.* **2007**, *8*, 173-184; b) A. de Fougerolles, H.-P. Vornlocher, J. Maraganore, J. Lieberman, *Nat Rev Drug Discov* **2007**, *6*, 443-453.

[7] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, D195-D201.

[8] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, *Nature* **2012**, *485*, 201-206.

[9] M. Lee, B. Kim, V. N. Kim, *Cell* **2014**, *158*, 980-987.

[10] a) C. H. Lu, B. Willner, I. Willner, *ACS Nano* **2013**, *7*, 8320-8332; b) H. I. Peng, B. L. Miller, *Analyst* **2011**, *136*, 436-447; c) D. Li, S. P. Song, C. H. Fan, *Acc. Chem. Res.* **2010**, *43*, 631-641.

[11] a) S. R. Panikkanvalappil, M. A. Mackey, M. A. El-Sayed, *J. Am. Chem. Soc.* **2013**, *135*, 4815-4821; b) S. R. Panikkanvalappil, M. A. Mahmoud, M. A. Mackey, M. A. El-Sayed, *ACS Nano* **2013**, *7*, 7524-7533; c) E. Papadopoulou, S. E. J. Bell, *Angew. Chem.-Int. Edit.* **2011**, *50*, 9058-9061; d) L.-J. Xu, Z.-C. Lei, J. Li, C. Zong, C. J. Yang, B. Ren, *J. Am. Chem. Soc.* **2015**, *137*, 5149-5154; e) L. Guerrini, Ž. Krpetić, D. van Lierop, R. A. Alvarez-Puebla, D. Graham, *Angew. Chem.-Int. Edit.* **2015**, *54*, 1144-1148; f) M. Masetti, H.-n. Xie, Ž. Krpetić, M. Recanatini, R. A. Alvarez-Puebla, L. Guerrini, *J. Am. Chem. Soc.* **2015**, *137*, 469-476; g) J. Morla-Folch, H.-n. Xie, P. Gisbert-Quilis, S. Gómez-de Pedro, N. Pazos-Perez, R. A. Alvarez-Puebla, L. Guerrini, *Angew. Chem.-Int. Edit.* **2015**, *54*, 13650-13654.

[12] a) J. D. Driskell, A. G. Seto, L. P. Jones, S. Jokela, R. A. Dluhy, Y. P. Zhao, R. A. Tripp, *Biosens. Bioelectron.* **2008**, *24*, 917-922; b) M. Muniz-Miranda, C. Gellini, M. Pagliai, M. Innocenti, P. R. Salvi, V. Schettino, *J. Phys. Chem. C* **2010**, *114*, 13730-13735; c) E. Prado, N. Daugey, S. Plumet, L. Servant, S. Lecomte, *Chem. Comm.* **2011**, *47*, 7425-7427; d) J. L. Abell, J. M. Garren, J. D. Driskell, R. A. Tripp, Y. P. Zhao, *Journal of the American Chemical Society* **2012**, *134*, 12889-12892; e) Y. P. Chen, G. Chen, X. W. Zheng, C. He, S. Y. Feng, Y. Chen, X. Q. Lin, R. Chen, H. S. Zeng, *Med. Phys.* **2012**, *39*, 5664-5668.

[13] a) A. J. Hobro, D. M. Standley, S. Ahmad, N. I. Smith, *Phys. Chem. Chem. Phys.* **2013**, *15*, 13199-13208; b) J. M. Benevides, M. Tsuboi, J. K. H. Bamford, G. J. Thomas, *Biophys. J.* **1997**, *72*, 2748-2762.

[14] P. Carmona, M. Molina, A. Rodriguez-Casado, *Journal of Raman Spectroscopy* **2009**, *40*, 893-897.

[15] V. Reipa, G. Niaura, D. H. Atha, *RNA* **2007**, *13*, 108-115.

[16] N. Fujimoto, A. Toyama, H. Takeuchi, *J. Mol. Struct.* **1998**, *447*, 61-69.

[17] A. Barhoumi, D. Zhang, F. Tam, N. J. Halas, *J. Am. Chem. Soc.* **2008**, *130*, 5523-5529.

[18] G. J. Thomas Jr, M. C. Chen, K. A. Hartman, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.* **1973**, *324*, 37-49.

[19] a) J. H. Ge, Y. T. Yu, *Trends Biochem. Sci.* **2013**, *38*, 210-218; b) M. Charette, M. W. Gray, *IUBMB Life* **2000**, *49*, 341-351.

[20] J. Liu, G. F. Jia, *J. Genet. Genomics* **2014**, *41*, 21-33.

[21] S. Mansy, W. L. Peticolas, R. S. Tobias, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **1979**, *35*, 315-329.

[22] Y. Niu, X. Zhao, Y.-S. Wu, M.-M. Li, X.-J. Wang, Y.-G. Yang, *Genomics Proteomics Bioinformatics* **2013**, *11*, 8-17.

[23] T. Ueda, K. Shinozaki, K. Ushizawa, M. Tsuboi, *J. Raman Spectrosc.* **1997**, *28*, 947-952.

[24] J. M. Benevides, S. A. Overman, G. J. Thomas, *J. Raman Spectrosc.* **2005**, *36*, 279-299.

[25] C. A. Grygon, T. G. Spiro, *Biopolymers* **1990**, *29*, 707-715.

[26] P. Geladi, D. Macdougall, H. Martens, *Appl. Spectrosc.* **1985**, *39*, 491-500.

GENERAL CONCLUSIONS

This thesis focused on exploring the analytical potential of a novel direct SERS method for the analysis based on the use of specifically designed positively charged nanoparticles. This highly sensitive label-free approach was successfully implemented in various applications of biological interest including identification and quantification of chemically-modified nucleobases in single and double stranded DNA, quantification of nucleobase composition in short and genomic DNA, and vibrational classification of structurally similar small RNA. Further, it was proven the efficient integration of the direct SERS method into microfluidic chips for precise manipulation of DNA samples at small length scales. Moreover, the combination of direct SERS analysis with statistical techniques such as PLS-DA provided extra discriminatory capacity. Thus, the main conclusions of the thesis can be summarized as follows:

1. It was shown that positively charged spermine-coated silver nanoparticles (AgNP@Sp) are efficient plasmonic substrates for the direct SERS analysis nucleic acids. Specifically, by using AgNP@Sp the electrostatic interaction between NAs and the colloids is ensured due to the spermine molecules anchored at the surface. This results in the entrapment of the biomolecules at the interparticle gaps of stable nanoparticle clusters in suspension, yielding intense and highly reproducible SERS spectra at the ultrasensitive level without the need of an external aggregating agent. This,

combined with the simple colloidal preparation, makes AgNP@Sp an optimal SERS platform for label-free NAs analysis.

2. AgNP@Sp colloids were successfully exploited for the identification and ultrasensitive quantification of multiple chemically-modified cytosine bases in both single and double stranded DNAs. Specifically, the investigated nucleobase variants included: 5-methylcytosine and 5-hydroxymethylcytosine (the most important epigenetic modifications in mammalian DNA); 5-bromocytosine (as a representative inflammation-induced DNA damage); and 5-hydroxycytosine as an example of oxidative lesion. Classification of the SERS spectra was achieved via PLS-DA.

3. The SERS method was efficiently integrated into a microfluidic device which allowed for the drastic reduction of the amount of DNA required per measurement down to the picogram level. Further, implementation with microfluidics enables fundamentally new applications, such as an increase in automation and parallelization, which paves the way to screening and systematic testing analysis.

4. Quantification capabilities of direct SERS approach were further validated in the determination of the relative nucleobase content in DNA. A large set of DNA probes were initially selected to train the sensing system. Once extracted from the SERS spectra the appropriate quantification parameters, this simple, inexpensive and flexible method for nucleobase

quantification was successfully applied to DNA samples independently of their size (from short to long genomic DNA), conformation (single-stranded *vs* duplex), and base sequence.

5. It was demonstrated the potential of label-free direct SERS analysis to fully disclose the unique vibrational information of a variety of different small RNAs with high sensitivity. As vibrational spectra and molecular features are intrinsically linked, it was possible to identify and classify highly similar RNA structures based on both their conformation and composition. Among others, it has been reported the first example of SERS recognition of fully complementary duplexes, as well as a short hairpin and small interfering RNAs. We further diversified microRNAs sequences with single-base mismatches. Nucleobase variants such as pseudouridine and N6-methyladenine were recognized and quantified with single-base sensitivity within both single strands and duplexes. Finally, direct comparison of structurally analogous RNA and DNA molecules also permitted to characterize distinctive signatures of their corresponding sugar moieties as well as their secondary structure.

In summary, this thesis demonstrates the tremendous analytical potential of this novel direct SERS method for nucleic acids analysis. Further, the combination with statistical techniques, such as PLS-DA, and the integration with microfluidics increments the discriminatory ability while enabling the online screening of

picogram amounts of DNA, respectively. This work opens new ways for the future development of a fast, low-cost, and high-throughput analytical technique for nucleic acids analysis.

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

AGRAÏMENTS

Resumir en poques línies totes les hores, persones, bons i no tan bons moments i tot el suport rebut durant aquest temps, no és tasca fàcil.

Primer de tot m'agradaria agrair-li al Prof. Ramón Álvarez-Puebla l'oportunitat que m'ha brindat de treballar en el seu grup. Gràcies per creure en mi des del primer moment, per ensenyar-me a desenvolupar un pensament científic crític i ser capaç de generar dubtes, de qüestionar-nos a nosaltres mateixos sent així la millor manera d'aprendre i millorar.

Sens dubte, agrair al Dr. Luca Guerrini el seu esforç i dedicació dia a dia, no només a aquesta tesi sinó també a la ciència. Per ensenyar-me tants i tants coneixements, per mostrar-me com treballar al detall i per transmetre'm la seva passió per aprendre d'allò desconegut. I al Dr. Nicolás Pazos, perquè sense les seves classes sobre plasmònica, nanopartícules i síntesis, instruïdes al laboratori, entre mesura i mesura, aquests tesis tampoc hagués sigut possible.

Agrair especialment a la Mian Rong i al Yang Zhe, de la Nanyany Technological University, per fer-me sentir com a casa, per ajudar-me en tot, acompanyar-me i fins i tot fer-me de traductors.

Fer una menció especial per totes les grans persones que gràcies a aquesta etapa he pogut conèixer. Companys de PhD, post-docs i gent del CTQ, gràcies per l'amistat, la festa i els debats científics. Especialment a tu Manu, per la teva facilitat

de treure'ns un somriure, perquè ja són molts anys al teu costat i molts bons moments compartits.

Finalment, a la meva família, per fer-me creure que sempre podem aconseguir allò que volem, si ho volem de veritat. A la teta, per transmetre'm les ganes de menjar-me al món. Al Dídac, per ser el més polifacètic de la família. No només informàtic, solucionador de problemes sinó també un bon amic. A la meva germana Teresa, per continuar cuidant-me encara que jo li digui que no és necessari. Gràcies per estar sempre disposada a ajudar-me i aconsellar-me.

Finalment, agrair amb tot l'amor al millor company de viatge, el Marc, per la seva paciència, comprensió, tendresa i per creure en mi en tot moment. I clarissimament, a la meva mare i al meu pare, pel seu suport incondicional, sigui on sigui ells em fan sentir que tot és possible.

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

APPENDIX I – List of Figures

Figure 3.1 99

Molecular structures of the cytosine modifications

Figure 3.2 100

Outline of DNA-driven aggregation of AgNP@Sp and schematic of the on-line detection of DNA via the SERS-microfluidic system.

APPENDIX II – List of Tables

h

APPENDIX III – List of Publications

J. Morla-Folch; P. Gisbert-Quilis; M. Masetti; E. Garcia-Rico; R.A. Alvarez-Puebla*; L. Guerrini*. *Conformational SERS classification of K-RAS point mutations for cancer diagnosis.* Angewandte Chemie International Edition, (2017).

J. Morla-Folch, R.A. Alvarez-Puebla*; L. Guerrini*. *Direct quantification of DNA base composition by Surface-enhanced Raman Scattering Spectroscopy.* The Journal of Physical Chemistry Letters, (2016).

J. Morla-Folch, H. Xie, R.A. Alvarez-Puebla*; L. Guerrini*. *Fast Optical Chemical and Structural classification of RNA.* ACS Nano, (2016).

J. Morla-Folch; H. Xie; P. Gisbert-Quilis; S. Gómez-de Pedro; N. Pazos-Perez; R.A. Alvarez-Puebla*; L. Guerrini*. *Ultrasensitive Direct Quantification of Nucleobase Modifications in DNA by Surface-Enhanced Raman Scattering: The Case of Cytosine.* Angewandte Chemie International Edition, (2015).

B. Mir-Simon; J. Morla-Folch; P. Gisbert-Quilis; N. Pazos-Perez; H. Xie; N. Bastus; V. Puntes; R.A. Alvarez-Puebla* and L. Guerrini*. *SERS Efficiencies of Micrometric Polystyrene Beads Coated with Gold and Silver Nanoparticles: The Effect of Nanoparticle Size*. Journal of Optics, (2015).

J. Morla-Folch; L. Guerrini; N. Pazos-Perez; R. Arenal; R.A. Alvarez-Puebla*. *Synthesis and optical properties of homogeneous nanoshurikens*. ACS Photonics (2014).

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT ROVIRA I VIRGILI
DIRECT LABEL-FREE SURFACE-ENHANCED RAMAN SCATTERING ANALYSIS OF NUCLEIC ACIDS
Judit Morlà Folch

UNIVERSITAT
ROVIRA i VIRGILI