Accepted: 28 July 2020

Investigation of genomic DNA methylation by ultraviolet resonant Raman spectroscopy

Francesco D'Amico[1*] | **Paolo Zucchiatti**[1,2,3] | **Katia Latella**[1,4] | **Maria Pachetti**[1,2] | **Alessandro Gessini**[1] | **Claudio Masciovecchio**[1] | **Lisa Vaccari**[1] | **Lorella Pascolo**[5]

[1]Elettra-Sincrotrone Trieste, Trieste, Italy

[2]Department of Physics, University of Trieste, Trieste, Italy

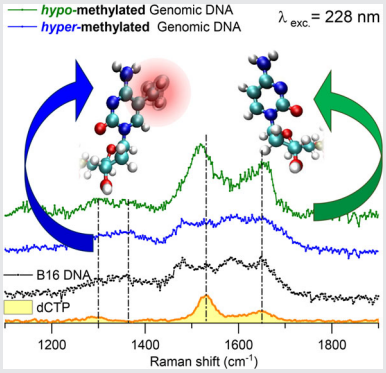
[3]Plasmon Nanotechnologies line, IIT, Genoa, Italy

[4]Department of Chemistry and Industrial Chemistry, University of Genova, Genoa, Italy

[5]Institute for Maternal and Child Health, IRCCS Burlo Garofolo, Trieste, Italy

***Correspondence**
Francesco D'Amico, Elettra-Sincrotrone Trieste S.C.p.A. Strada Statale 14 - km 163,5 in AREA Science Park, 34149 Basovizza, Trieste, ITALY.
Email: francesco.damico@elettra.eu

Abstract



Cytosine plays a preeminent role in DNA methylation, an epigenetic mechanism that regulates gene expression, the misregulation of which can lead to severe diseases. Several methods are nowadays employed for assessing the global DNA methylation levels, but none of them combines simplicity, high sensitivity, and low operating costs to be translated into clinical applications. Ultraviolet (UV) resonant Raman measurements at excitation wavelengths of 272 nm, 260 nm, 250 nm, and 228 nm have been carried out on isolated deoxynucleoside triphosphates (dNTPs), on a dNTP mixture as well as on genomic DNA (gDNA) samples, commercial from salmon sperm and non-commercial from B16 murine melanoma cell line. The 228 nm excitation wavelength was identified as the most suitable energy for enhancing cytosine signals over the other DNA bases. The UV Raman measurements performed at this excitation wavelength on hyper-methylated and hypo-methylated DNA from Jurkat leukemic T-cell line have revealed significant spectral differences with respect to gDNA isolated from salmon sperm and mouse melanoma B16 cells. This demonstrates how the proper choice of the excitation wavelength, combined with optimized extraction protocols, makes UV Raman spectroscopy a suitable technique for highlighting the chemical modifications undergone by cytosine nucleotides in gDNA upon hyper- and hypo-methylation events.

KEYWORDS

cytosine nucleotides, DNA methylation, epigenetics, isolated DNA, UV resonant Raman

Abbreviations: 5-Aza-dc, 5-aza-2-deoxycytidine; DNMT, DNA-methyl transferases; dNTP, deoxynucleoside triphosphate; gDNA, isolated genomic DNA; PCR, polymerase chain reaction; UVRR, UV Resonant Raman.

1 | INTRODUCTION

In the last decades, a new research field has emerged to explain mechanisms involved in a large variety of

phenomena, from embryogenesis to environmental adaptation, as well as regulation of metabolic functions in various disease states including cancer: this is epigenetics, which comprises the reversible and hereditable modifications of the DNA, without irreversible variations of the genome.[1, 2]

The epigenetic modification of the genomic sequences is a lifelong process that begins at the level of the developing germ cells in the embryo, influenced by maternal nutrition during pregnancy, and which continues after that influenced by environmental factors: errors in the epigenetic processes of cells, such as those produced by exposure to environmental chemicals, may cause pathologies and cancer as well.[2–4]

DNA cytosine methylation is one of the most relevant epigenetic mechanisms, which consists of the addition of a methyl group from S-adenosyl-methionine to the fifth position of the cytosine ring (5meC) in the CpG islands by DNA-methyl transferases (DNMT) with the consequent gene silencing.[5] Many clinical studies evidenced that there are substantial alterations in the global level of cytosine methylation in several cancer-cell-models.[2–4] In other contests, DNA is suggested to be highly methylated in functional gametes,[6] and thus global methylation could reflect the fertility potential of a patient. Therefore, where obtained with a fast, economic and simple protocol, the information on global methylation of DNA could be a useful diagnostic aid in clinical medicine as well as in basic research.[7] Many techniques, such as liquid chromatography, immune-quantifications, bisulfite conversion, polymerase chain reaction (PCR), and optical-based methods are currently adopted to quantify DNA methylation, mainly for research purposes.[3, 8]

However, there is still an urgent need for new protocols that may combine simplicity, high sensitivity, and low operating costs to be translated into clinical applications.[3, 7]

Recently, many vibrational spectroscopic techniques have been used to infer DNA methylation.[9–11] In this regard, techniques based on surface-enhanced Raman scattering (SERS) have been among the most used to investigate DNA in biological systems,[12, 13] and to infer DNA damage[14] and cytosine methylations mechanisms.[13, 15, 16] The goal is achieved thanks to the enhanced sensitivity of SERS, which allows detecting Raman signal even from a minute amount of sample. Although SERS provides excellent results in terms of sensitivity, it is frequently affected by poor reproducibility, thus foreseeing lower diagnostic reliability.[17, 18]

Another vibrational spectroscopy technique that is potentially able to infer DNA methylation is ultraviolet resonant Raman (UVRR) spectroscopy, which has been extensively used for the investigation of nitrogenous bases,[19–23] nucleosides,[24–27] nucleotides,[28] calf thymus, salmon sperm DNA and nucleic acid-based systems.[29–34] UVRR gives the possibility of working in an aqueous solution, thus keeping the DNA in its physiological conditions and it is label-free, preventing DNA chemical modifications.[35–37]

Despite the clear-cut proof of principle regarding the potential of UVRR in getting essential insights on the chemical modifications of nitrogenous bases, the possibility of transition to clinical applications is still faint. The first limit resides in the assessment of proper sample preparation protocols. DNA isolation is usually performed using commercially available kits, or by automatized extractors. The DNA, obtained following standard approaches, is adequate for molecular applications, but the presence of contaminants, such as proteins and residuals of the extraction chemicals, limits its exploitation for any vibrational spectroscopy analysis.[35] However, we previously demonstrated that introducing minimal modifications to the standard extraction protocols reduces the occurrence of the most frequent contaminants of the isolated DNA, as verified by FTIR and UVRR analysis at 266 nm.[35]

Interestingly, it has been widely reported that it is possible to selectively emphasize the vibrational bands associated to any single nitrogenous base by appropriately changing the UVRR excitation energy.[28, 31–34] Stated that it is possible to distinguish the spectral response of the cytosine from the signals coming from the other nitrogenous bases, in the present paper we verify the hypothesis that, properly exciting the sample, it is possible to reveal the spectral variations induced by the methylation on cytosine islands of DNA, thus opening to the applicability for epigenetic evaluations. Specifically, in the present work we found the right UVRR excitation wavelengths to investigate the relative abundance of single nitrogenous bases in isolated genomic DNA (gDNA) and, more importantly, to reveal and compare the spectral variations induced by cytosine methylation in different types of DNA samples.

We initially performed UVRR measurements on nucleotides aqueous solutions, commercial gDNA from salmon sperm (example of germ cell) and gDNA extracted from murine melanoma B16 cells (example of cancer cell), comparing the spectra at 272 nm, 260 nm, 250 nm and 228 nm of excitation wavelength. Finally, the excitation wavelength of 228 nm was selected to enhance cytosine signals in gDNA from human Jurkat cells (acute T-cell leukemia) treated with 5-aza-2-deoxycytidine (5-Aza-dc), as a negative control for CpG methylation genome, and enzymatically methylated by CpG Methyltransferase, as cytidine *hyper*-methylated standard.

2 | EXPERIMENTAL SECTION

For the Raman measurements, commercial PCR grade standards (100 mM in aqueous solution) of deoxynucleoside triphosphates (dNTPs) dATP, dCTP, dGTP, and dTTP (Sigma Aldrich) were diluted in Na_2SO_4 1.0 M up to 2 mM concentrations. Additionally, by mixing all the four solutions in equal parts, we prepared a further nucleotide mixture, at a concentration of 0.5 mM for each nucleotide.

Salmon sperm gDNA (Sigma Aldrich) is supplied highly purified and its guanine/cytosine percent content (%CG) is 41.2% accordingly to the technical note of the supplier.[38] Salmon sperm gDNA was diluted in 50 μL of MilliQ water with a final concentration of 1.0 mg/mL (namely 3.3 mM of nucleotides concentration) and 0.1 mg/mL (namely 0.33 mM of nucleotides concentration).

For DNA extraction from the B16 cells, we applied the protocol described in[35] and [36] in order to remove ethanol residues and to minimize the presence of the other contaminants. The isolated gDNA was diluted in MilliQ water to a final concentration of 0.33 mM (0.1 mg/mL).

gDNA from human male Jurkat cells treated with 5-aza-2 deoxycytidine was obtained commercially (Biolabs) and was diluted in MilliQ water up to a final concentration of 0.33 mM (0.1 mg/mL). A sample with identical concentration was prepared from gDNA isolated from human male Jurkat cells enzymatically treated with CpG methylase (Biolabs).

The purity of Jurkat and B16 gDNA was routinely verified by UV–visible absorption spectroscopy: the DNA samples used had a $A_{260nm/280nm}$ absorption intensity ratio larger than 1.8, confirming the absence of protein and DNA contamination. In addition, the samples were tested by attenuated total reflection -Fourier transformed infrared spectroscopy, as in[35] to verify the absence of contaminants from extraction protocol reagents. More details can be found in the Section 1 of the Supporting Information (SI).

UV–visible absorption measurements on nucleotides and gDNA have been carried out using a Perkin Elmer Lambda25 spectrophotometer. The final concentrations of the solutions in MilliQ water was 0.1 mM.

UVRR measurements have been performed at the IUVS beamline at Elettra Sincrotrone Trieste. A complete description of the experimental apparatus and measure protocol is reported elsewhere.[39] Briefly, synchrotron-based wavelengths at 272 nm, 260 nm, 250 nm, and 228 nm have been employed to excite 100 μL of the sample placed in a Suprasil quartz cell. Raman measurements were performed in a backscattering configuration, and the beam-power measured on the sample was approximately 2 μW. A Czerny-Turner spectrometer with a focal length of 750 mm equipped with a holographic dispersive element of 1800 lines/mm, coupled with a Peltier-cooled back-thinned CCD has been employed to get the Raman signal. The spectral bandwidth was set to 25 cm^{-1} for all the employed excitation wavelengths in order to get a sufficient count rate. The obtained resolving power (λ /$\Delta\lambda$) was 1460, 1527, 1588, and 1741 at 272 nm, 260 nm, 250 nm and 228 nm respectively. The samples were continuously oscillated during the measurement in order to avoid photo-damaging. Raman frequencies were calibrated to $\pm$1 cm^{-1} of accuracy by using cyclohexane spectra. For the correct wavenumbers of the cyclohexane we refer to peak positions of [40]

The nucleotides spectral intensity has been normalized to the SO_4^{2-} stretching peak area at 981 cm^{-1} as well as to the solvent spectral intensity, as done by Wen and coworkers.[33] The solvent contribution has been then removed from nucleotides solutions spectra (see Section 2 of SI for further details).

Water spectral contribution has been subtracted from the UVRR DNA spectra, after normalizing them to the OH-stretching band of water at 3300 cm^{-1}. This procedure has been applied to the excitation wavelengths 272 nm, 260 nm, and 266 nm. Differently, the water contribution in the DNA spectra collected with 228 nm of excitation wavelength has been subtracted by normalizing the spectral intensities to the water overtone band at 2200 cm^{-1}. Details can be found in SI.

3 | RESULTS AND DISCUSSION

3.1 | UV–vis absorption measurements

UV-absorption spectroscopy is a crucial preliminary step for UVRR spectroscopy. Indeed, with this technique, it is possible to tentatively define the excitation wavelengths more suitable for enhancing specific resonances over other signals and, therefore, for fully exploiting the potential of UVRR spectroscopy. To this aim, the molar extinction coefficients of dATP, dCTP, dGTP, and dTTP, as derived by UV–visible absorption measurements, are plotted in Figure 1. Figure 1 also shows the UV–visible absorption spectrum of the salmon sperm gDNA (pink curve). In order to estimate to which extent DNA bases contribute to the total absorption of gDNA, the A_{sum} curve is depicted, which represent the estimated DNA absorption curve obtained summing the contributions of the four nucleobases. Since it is known that the %GC content in salmon gDNA is 41.2%, A_{sum} curve has been calculated as follows:

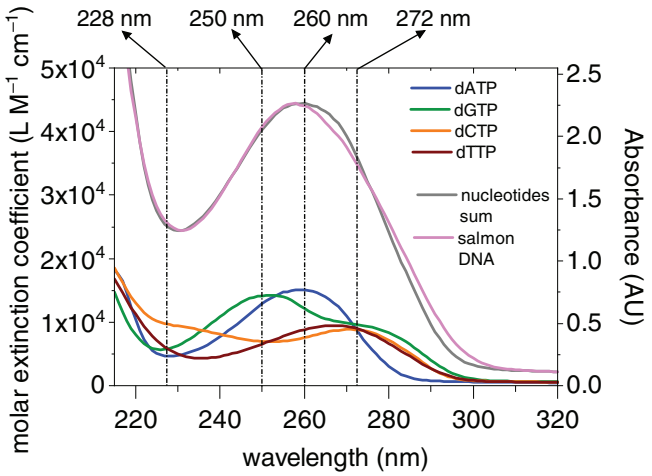


FIGURE 1 UV–vis absorption spectra of salmon sperm DNA (0.10 mg/mL,) dATP (blue), dGTP (green), dCTP (yellow) and dTTP (red) aqueous solutions. UV, ultraviolet

$$A_{\text{sum}} = 2*[0.412*(A_{\text{dCTP}} + A_{\text{dGTP}}) + (1-0.412)*(A_{\text{dATP}} + A_{\text{dTTP}})].$$

where A_{dCTP}, A_{dGTP}, A_{dATP} e A_{dTTP} are the dCTP, dGTP, dATP e dTTP molar extinction coefficients[41], while 0.412 is the CG molar ratio.

A_{sum} and salmon sperm gDNA curves match pretty well below 260 nm while they are slightly deviating above 260 nm. These minimal transpositions can be addressed to the different structural conformation of the nitrogenous bases in double-stranded DNA compared to dNTPs. Indeed, in the case of salmon sperm gDNA, the nitrogenous bases are involved in hydrogen bonds not present in the isolated nucleotides solutions. The H-bonding network can induce small modifications in the nucleotide absorption cross-sections, leading to changes in the absorption curve profile.[41] Nevertheless, despite the minimal discrepancy, it can be safely assumed that the gDNA UV-absorption spectrum is well represented by the linear combination of all the single nitrogenous bases UV-absorption spectra.

3.2 | UVRR at 272 nm, 260 nm, 250 nm, and 228 nm

By analyzing the individual contribution of nucleotides, it is possible to state that in the 220 nm – 300 nm spectral region, dATP and dGTP absorptions are characterized by a maximum at 260 nm and 250 nm, respectively, while dTTP and dCTP absorption maxima are located near 265 nm and 272 nm, respectively. However, the best condition to study dCTP is preferably 228 nm, where the dCTP absorption contribution is more intense than the ones of the other nucleotides.

The choice of the excitation wavelength should also be guided by additional considerations concerning the spectral fluorescence background. At excitation wavelengths higher than 250 nm, a fluorescence background moderately affects the UVRR spectrum of commercial salmon sperm gDNA (see Section 3 in SI), due to the intrinsic fluorescence of the nitrogenous basis. Besides, the UVRR spectrum of gDNA extracted from B16 cells acquired at 272 nm is strongly influenced by fluorescence (see Figure S3 in SI), due to retention in the analyzed solution of residuals of the DNA isolation chemicals despite the optimized extraction protocol.[35]

Nevertheless, even for this extreme condition, we could efficiently remove fluorescence background by subtracting a proper polynomial curve, without inducing modifications of the Raman spectral profile (see Figure S4 in SI). Indeed, the authors have already proven that the fluorescence background, induced by the residual presence of extraction chemical, does not modify the vibrational Raman spectral line-shape.[35]

Figure 2 shows the UVRR spectra at 272 nm of the gDNA isolated from B16 cells (spectrum a, blue dotted line) and the UVRR spectra of salmon sperm gDNA (spectrum c, red dotted line), both in aqueous solution (0.33 mM). The spectra are characterized by specific vibrational bands at 1655 cm^{-1}, 1578 cm^{-1}, 1484 cm^{-1} and 1337 cm^{-1}.[28, 31–34] The lower part of the same Figure (spectra d) reports the UVRR spectra derived from the aqueous solutions of dATP, dCTP, dGTP, and dTTP, as well as the spectrum obtained from the dNTP mixture in aqueous solution (spectrum d, black dotted line).

In order to carry out an accurate peak attribution and, in particular, to estimate the contribution of each nitrogenous base to the total intensity of the Raman spectrum, we calculated the sum_{N} curve as:

$$sum_{\text{N}} = \text{dATP} + \text{dTTP} + \text{dCTP} + \text{dGTP}.$$

and the sum_S curve as:

$$sum_{\text{S}} = (1-0.412)*(\text{dATP} + \text{dTTP}) + 0.412*(\text{dCTP} + \text{dGTP}).$$

where 0.412 is the %CG content in the salmon sperm DNA. sum_N and sum_S spectral intensities were then normalized to the 1484 cm^{-1} peak maximum of dNTPs and of salmon sperm gDNA spectra, respectively.

It is interesting to note how the sum_N curve strongly matches the dNTPs spectrum, within 10% in intensity in the worst case, and that the sum_S curve satisfactorily matches with the salmon sperm gDNA spectrum.

Therefore, with a good degree of confidence, it is reasonable to assume that the relative intensity contribution

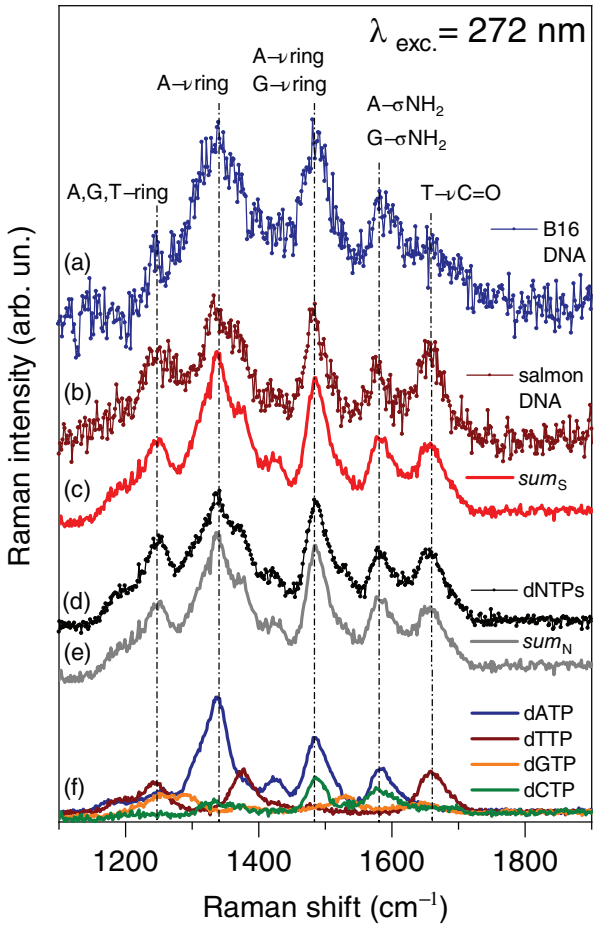


FIGURE 2 UVRR spectra at 272 nm of: (a) gDNA extracted from B16 cells 0.33 mM, (b) salmon sperm gDNA 0.33 mM, (d) dNTP mix aqueous solution, dATP (blue), dGTP (green), dCTP (yellow) and dTTP (red) aqueous solutions. The curves (c) and (e) represent the sum_N and sum_S curves respectively, calculated as explained in the text. Spectral intensities of (a), (b), (c), (d) and (e) are normalized to the 1350 cm^{-1} intensity. Spectral intensities of (c) have been normalized to the SO_4^{2-} peak at 981 cm^{-1}. Normalizations have been carried out before the solvent contribution subtractions. Vertical dotted lines are guide for the eyes. dNTP, deoxynucleoside triphosphates; gDNA, genomic DNA; UVRR, ultraviolet resonant Raman

of each dNTP to the total DNA spectrum, at a specific wavenumber ν^*, is given by

$$I_{\mathrm{Rel-base}}(\nu^*) = \frac{I_{\mathrm{base}}(\upsilon^*)}{\sum_{j=1}^{4} I_j(\upsilon^*)}.$$

$I_J(\upsilon^*)$ is the spectral intensity at the wavenumber υ^* of each dNTP spectrum. The relative intensities obtained in this way are affected by an error of about 5%, calculated on the base of the S/N ratio at each excitation wavelength.

Applying this formula to the salmon gDNA, we found that the 65% of the 1655 cm^{-1} peak intensity is due to the thymidine C=O stretching, in total accordance with the current scientific UVRR literature on nucleotides (dTMP-dCMP), calf thymus and salmon sperm gDNA.[28, 31–34] The peak at 1578 cm^{-1} is the convolution of adenosine $6NH_2$ scissoring vibrational mode at 1580 cm^{-1} (56% of the peak intensity) and guanosine $2NH_2$ scissoring and ring vibrations at 1575 cm^{-1} (33% of the peak intensity). The peak at 1484 cm^{-1} arises from adenosine (60% of the peak intensity) and guanosine (30% of the peak intensity) internal ring vibrations. Interestingly, the peak at 1337 cm^{-1} can be almost exclusively assigned to adenosine internal rings vibrations (83% on relative intensity), thus becoming a valid fingerprint to investigate the adenosine behavior in DNA complex systems. Lastly, the small component at ~1250 cm^{-1} is assigned to adenine, guanine, and thymidine internal ring vibrations combined with N-H bending.

The comparison between B16 and salmon sperm gDNA evidences a decrease of the 1655 cm^{-1} peak compared to the 1484 cm^{-1} one (see a dotted vertical line on Figure 2), that can be explained by a lower thymine contribution or to a different hydrogen-bonding environment behavior of the thymine C=O group between salmon sperm and B16 gDNA.

The same spectral features found at 272 nm in salmon sperm and B16 gDNA UVRR spectra are present also at 260 nm (see Figure 3). The main features of the gDNA spectra are still at 1655 cm^{-1}, 1578 cm^{-1}, 1484 cm^{-1} and 1337 cm^{-1}, but the relative intensity contribution of each dNTP differs among the two experimental conditions. In Figure 3, the UVRR spectra of dATP, dCTP, dGTP, and dTTP, as well as the spectrum obtained from the dNTP mixture at 260 nm are also reported (c).

Applying the same formula employed at 272 nm, we calculated the curves sum_N and sum_S, relative to the dNTP mixture solution and the salmon sperm gDNA, respectively. Thanks to the agreement between acquired and calculated spectra, we estimated the contribution of the single dNTP to the gDNA spectra.

The peak at 1578 cm^{-1} derives from adenosine (61%) and guanosine (35%), while the remaining 4% comes from the cytidine and thymidine. The peak at 1484 cm^{-1} is still addressed to guanosine (48%) and adenine (45%) internal ring vibrations. The peak at 1337 cm^{-1} is mainly assigned to adenosine internal ring vibrations. Nevertheless, lowering the excitation wavelength from 272 to 260 nm, the relative contribution to the peak intensity of adenosine decreases from 83% to 70%, in favor of the guanosine internal ring vibration, which rises from 10% to 23%.

Although the adenosine contribution at 1337 cm^{-1} and the one of thymidine at 1655 cm^{-1} are less dominant

for spectra collected at 272 nm, DNA spectra gathered at 260 nm suffer less from the fluorescence background, and should be preferred when fluorescence background subtraction cannot be accurately performed.

As at 260 nm excitation wavelength, there is a decrease in the intensity of the 1655 cm^{-1} peak of B16 gDNA, not detected for the salmon sperm. In this case, the decreasing effect is less intense than at 272 nm, and can be explained by the minor contribution of the thymine C=O stretching band to the overall peak intensity.

The UVRR spectra of B16 cells gDNA (a) and salmon sperm gDNA (b), together with dNTP spectra (0.1 mg/mL), collected at the excitation wavelength of 250 nm are shown in Figure 4. The peak at 1655 cm^{-1} is significantly suppressed compared to the same peak detected at 272 nm and 260 nm, both in dNTPs and gDNA spectra. Indeed, it becomes a small bump in the gDNA vibrational spectra, reflecting the relative intensities of the specific nitrogenous base vibrations for salmon sperm gDNA: 40% for guanine, 35% for thymine, 15% for cytosine, and 10% for adenine.

The peak at 1578 cm^{-1} evidences a clear predominance (75%) of the guanosine contribution. The

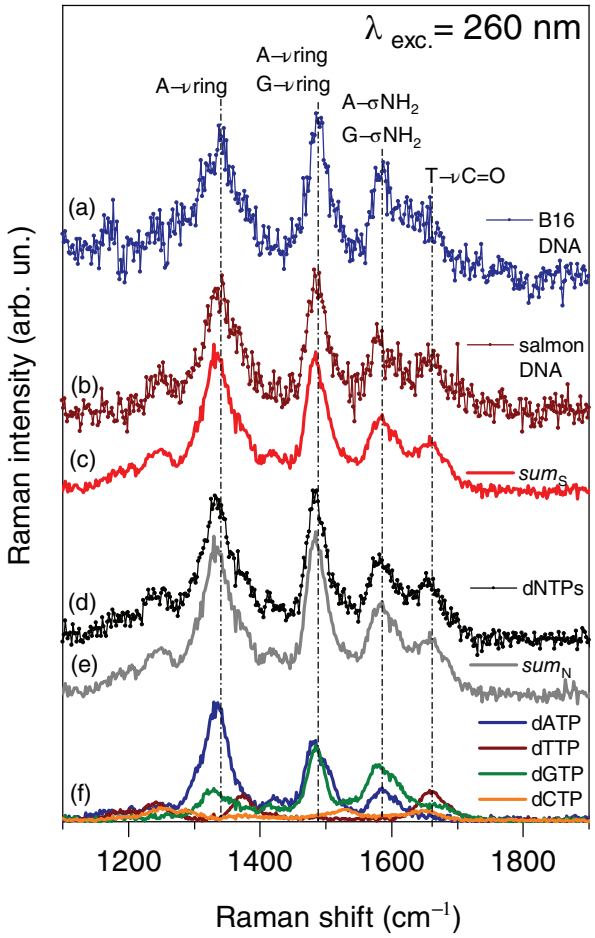


FIGURE 3 UVRR spectra at 260 nm of: (a) gDNA extracted from B16 cells 0.10 mg/mL, (b) salmon sperm gDNA 0.10 mg/mL, (d) dNTP mix aqueous solution, (f) dATP (blue), dGTP (green), dCTP (yellow) and dTTP (red) aqueous solutions. The curves (c), and (e), represent the sum_N and sum_S curves, calculated as explained in the text. Spectral intensities of (a), (b), (c), (d), and (e), are normalized to the 1350 cm^{-1} intensity. Spectral intensities of (c), have been normalized to the SO_4^{2-} peak at 981 cm^{-1}. Normalizations have been carried out before the solvent contribution subtractions. Vertical dotted lines are guide for the eyes. dNTP, deoxynucleoside triphosphates; gDNA, genomic DNA; UVRR, ultraviolet resonant Raman

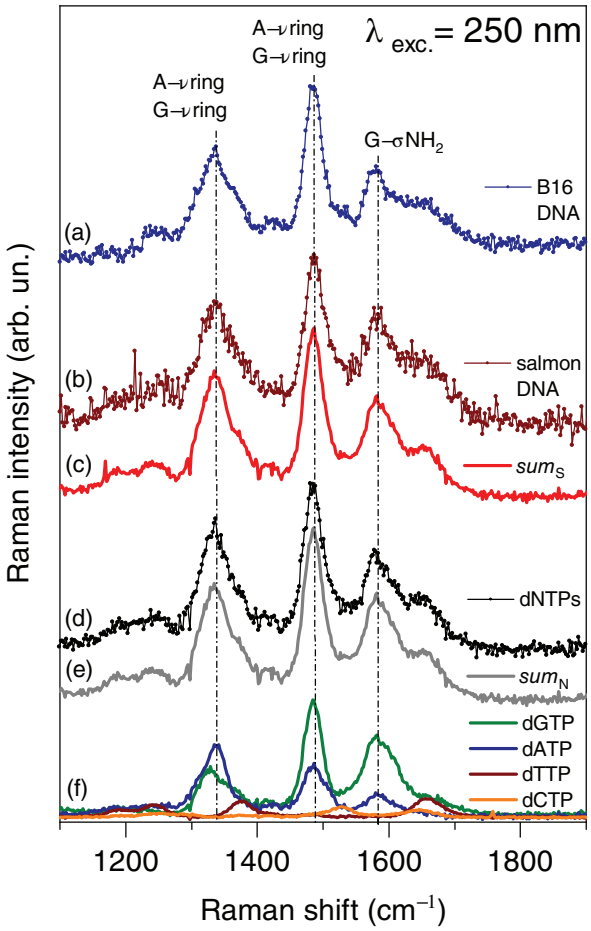


FIGURE 4 UVRR spectra at 250 nm of: (a) gDNA extracted from B16 cells 0.10 mg/mL, (b) salmon sperm gDNA 0.10 mg/mL, (d) dNTP mix aqueous solution, (f) dATP (blue), dGTP (green), dCTP (yellow) and dTTP (red) aqueous solutions. The curves (c), and (e) represent the sum_N and sum_S curves, calculated as explained in the text. Spectral intensities of (a), (b), (c), (d), and (e), are normalized to the 1350 cm^{-1} intensity. Spectral intensities of (c), have been normalized to the SO_4^{2-} peak at 981 cm^{-1}. Normalizations have been carried out before the solvent contribution subtractions. Vertical dotted lines are guide for the eyes. dNTP, deoxynucleoside triphosphates; gDNA, genomic DNA; UVRR, ultraviolet resonant Raman

adenosine abundance is slightly above 20%, while that of cytosine and thymine are less than 5%.

The peak at 1484 cm^{-1} derives from the overlap of guanine (66% relative intensity), adenine (25% relative intensity), cytosine and thymine (less than 3% relative intensity each) contributions. The 1337 cm^{-1} peak is dominated by adenine contribution (63% relative intensity), while guanine (33% relative intensity) and cytosine and thymine (2% relative intensity each) contribute less.

In accordance with the UV–Vis absorption spectra of the dNTPs shown in Figure 1, the 250 nm excitation wavelength is the most favorable to investigate the guanosine behavior, at the specific vibrations of the 1578 cm^{-1} and 1484 cm^{-1}.

Indeed, the excitation at 250 nm represents the best compromise between the spectral intensity of the DNA and the fluorescence background, the latter being negligible, thus favoring the accuracy required for a correct spectral analysis and interpretation.

Both in salmon sperm and B16 gDNA, guanosine and adenosine nitrogenous bases signals are the dominant contributions to the overall spectral intensity.

A common characteristic of the DNA spectra collected at 272 nm, 260 nm, and 250 nm is the absence of clear distinctive vibrational features for cytosine. This event is somewhat expected: as seen in Figure 1, the excitation wavelengths mentioned above favor the adenosine and guanosine resonance conditions. Consequently, cytosine bands in the UVRR gDNA spectra at 272 nm, 260 nm, and 250 nm are present, but always with a relative intensity negligible compared to the other nucleotides.

The situation completely reverses at 228 nm. As previously noticed, the UV-absorption spectra of Figure 1 shows a relative increase of the dCTP absorption concerning the other nitrogenous bases at this wavelength. This reaction implies that the UV resonant conditions also favor the enhancements of the cytosine vibrations in the UVRR spectrum.

Notably, by selecting excitation wavelengths lower than 228 nm, the spectral contribution from proteins amide bands residues could affect the UVRR spectra.[31] This observation is particularly relevant for non-commercial DNA isolated from cells, where the standard protocols are not able to completely eradicate the nuclear proteins. Therefore 228 nm can be considered the lowest excitation wavelength to adopt for accurately investigating extracted DNA by UVRR.

Figure 5 reports the UVRR spectrum of salmon sperm gDNA (a) (0.1 mg/mL), at an excitation source of 228 nm, as well as the mixture of those of dATP, dCTP, dGTP, dTTP (b) and the dNTPs.

Unlike what was done at 272 nm, 260 nm, and 250 nm, the subtraction of the spectral contribution of water has been performed by normalizing the spectra to the water overtone band located at 2200 cm^{-1} (for more details, see Section 5 in SI). As shown in the top panel of Figure 5, this choice permits a more accurate acquisition

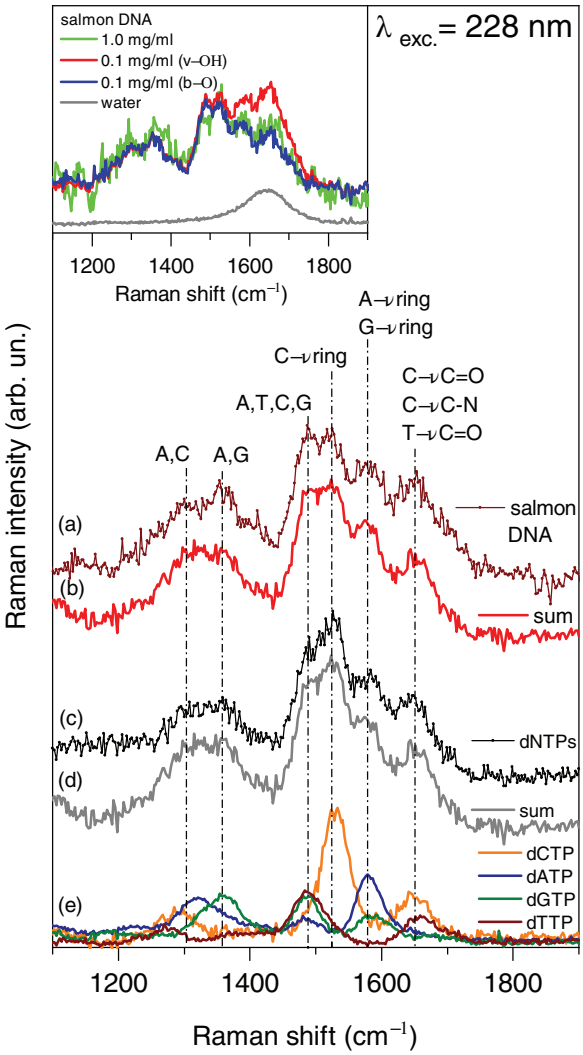


FIGURE 5 UVRR spectra at 228 nm of: (a) salmon sperm gDNA 0.10 mg/mL, (c) dNTP mix aqueous solution, (e) dATP (blue), dGTP (green), dCTP (yellow) and dTTP (red) aqueous solutions. The curves (b) and (d) represent the sum_N and sum_S curves, calculated as explained in the text. Spectral intensities of (a), (b), (c), and (d), are normalized to the 1350 cm^{-1} intensity. Spectral intensities of e, have been normalized to the SO_4^{2-} peak at 981 cm^{-1}. Normalizations have been carried out before the solvent contribution subtractions. The upper inset reports the spectra of the 1.0 mg/mL salmon sperm gDNA water solution after subtraction of the solvent contribution (green curve) as well as the 0.1 mg/mL salmon sperm gDNA aqueous solutions where two different solvent subtraction methods have been employed (red and blue curves). Water spectra (gray) have been reported for comparison. See the text for more details. dNTP, deoxynucleoside triphosphates; gDNA, genomic DNA; UVRR, ultraviolet resonant Raman

and a less noisy spectrum, also in the diluted sample. As expected, the DNA spectra at this energy are greatly different from those at higher wavelengths. The salmon sperm gDNA spectrum shows four important features centred at 1650 cm^{-1}, 1583 cm^{-1}, 1527 cm^{-1} and 1483 cm^{-1}. A broad vibrational feature is also visible between 1200 cm^{-1} and 1450 cm^{-1}.

The feature at 1650 cm^{-1} is mainly an overlap between the cytosine peak at 1650 cm^{-1} (C2 = O and C2-N3 stretching, according to [28, 31–34]), which contributes to 50% of the total intensity, and the thymidine peak at 1659 cm^{-1} (C4 = O and C4-C5 stretching, [23–25]), contributing to 30% of total intensity.

The peak at 1583 cm^{-1} in the gDNA spectra derives mainly from adenosine (58% of total intensity) and guanosine (24% of total intensity) internal ring vibrations. It is the analogue of the DNA peak found at 1578 cm^{-1}, employing 250 nm of excitation (Figure 4).

The spectral feature at 1527 cm^{-1} is constituted by more than 70% in the intensity of cytidine internal ring vibrations (more specifically, the N3-C4 and N1-C2 stretching) centred at 1527 cm^{-1}[128, 31–34] (229 nm). This feature is the most important we found since it represents the unique vibrational feature which can be addressed almost entirely to cytidine vibrational normal mode. The small tail at 1483 cm^{-1} is the equivalent of the one centred at 1484 cm^{-1} at 250 nm wavelength. Its intensity can be addressed to all the nucleotides with a small predominance of thymidine (35%). The broad feature between 1200 and 1450 cm^{-1} does not contain well-defined peaks. The only two features that we identified are two discontinuities at 1358 cm^{-1} and 1300 cm^{-1}: the first one characterized by the contributions of guanosine (50%), adenosine (30%) and thymidine (15%), while the second is mainly due to adenosine (40%) and cytidine (30%).

3.3 | UVRR of gDNA methylation standards

Stated that 228 nm is the most appropriate excitation wavelength to investigate the cytidine nucleotide behavior, UVRR at this wavelength can potentially reveal the chemical modifications occurring in cytidine, such as it *hypo*- and *hyper*- methylation due to epigenetic processes.

In order to prove such sensitivity, Figure 6 shows the UVRR spectra of CpG Jurkat *hyper*-methylated gDNA (blue curve) and the 5-azdact treated Jurkat *hypo*-methylated gDNA, both in aqueous solutions. The spectra of gDNA isolated from B16 cells (black curve) and that of dCTP (yellow) are also reported. At first glance, Figure 6 reveals the significant difference between *hyper*-methylated and *hypo*-methylated gDNA.

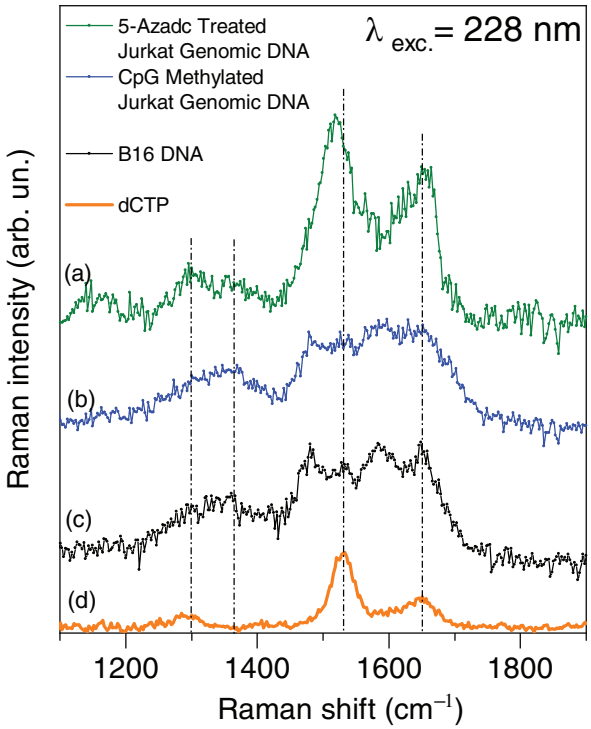


FIGURE 6 UVRR spectra of 5-azdact treated Jurkat genomic *hypo*-methylated gDNA, 0.10 mg/mL (green), CpG Jurkat genomic *hyper*-methylated gDNA 0.10 mg/mL (blue), B16 cell extracted DNA 0.10 mg/mL (black), dCTP (yellow) aqueous solutions; collected by using an excitation wavelength of 228 nm. Vertical dotted lines are guide for the eyes. gDNA, genomic DNA; UVRR, ultraviolet resonant Raman

Specifically, in the *hypo*-methylated DNA spectrum, it is possible to appreciate an evident rise in the peak centred at 1527 cm^{-1}. It strongly resembles the cytidine peak, slightly shifted to lower wavenumber. Such a strong peak is damped in the *hyper*-methylated DNA spectrum and shifted to higher wavenumber. At the same time, a shift of the 1650 cm^{-1} cytidine peak to higher wavenumbers occurs. These spectral changes are addressed to the different chemical conformation of the cytidine nucleotides, which are constituted to 5-aza-2-deoxycytidine in the *hypo*-methylated standard DNA and 5-methylcytidine in the case of *hyper*-methylated standard DNA.[42]

It is interesting that the B16 gDNA spectrum mainly resembles the CpG methylated Jurkat gDNA, while the salmon sperm gDNA spectrum appears to be an intermediate condition. Since this different spectral behavior between B16 and salmon gDNA sperm occurs only at 228 nm, it is conceivable that the modifications are induced by changes in the chemical and structural conformation of cytidine, and in particular that B16 cells

have a higher level of methylation compared to salmon sperm. This issue is in agreement with the fact that the development of germ cells, such as spermatozoa, has a lower level of methylation than somatic cells.[43]

This result opens up to the use of an analytical UVRR method of measurements at 228 nm on properly isolated gDNA to identify the changes in cytidine chemical conformation, related to epigenetic methylation processes.

4 | CONCLUSIONS

The present work aims at investigating the most advantageous experimental conditions for employing UVRR to reveal nitrogenous bases chemical changes in DNA isolated from cellular samples and in particular we showed how UVRR can provide distinct information on DNA cytidine modifications. To this purpose, we measured UVRR of standard *hyper-* and *hypo*-methylated DNA samples, also comparing them with other gDNAs. In accord with and extending previous reports,[19–34] our measurements show that any nucleotide has an optimal excitation wavelength: thymidine is well investigated at 272 nm, while 260 nm and 250 nm are more suitable for adenosine and guanosine. The precise investigation of cytidine can be performed only at 228 nm, where cytidine vibrations are predominant in the overall UVRR spectrum, and the methylation signal is sufficiently high to be monitored in extracted gDNA samples. Our measurements carried out at 228 nm on *hyper-* and *hypo*-methylated standard DNA samples evidence the high sensitivity of UVRR in detecting cytidine chemical modifications induced by *hyper-* and *hypo*-methylation of gDNA, also comparing gDNA from different species. It is relevant, in fact, that higher methylation level of B16 cells with respect to salmon sperm ones was revealed, as expected from somatic compared to germ cells.

Although further settings will be required, these observations open to the developments of UVRR protocols for detecting and quantifying DNA cytidine chemical modifications to be employed both in basic epigenetic studies and clinical diagnosis as well. It is well recognized, for instance, that DNA methylation plays a role in tumor development and therapy resistance: some demethylating agents, in fact, can be effective chemotherapeutics for solid tumors.[44] Analogously, evaluating methylation can be used to height male fertility potential, which becomes particularly relevant in assisted reproductive technology laboratories.[5, 45] Consequently, we could expect that the described UV Raman protocol could find a role in such clinical settings. It is clear that for clinical analysis the samples should consist in cell or tissue extracted DNA of high purity, as already described.[35]

ACKNOWLEDGMENTS

The authors are grateful to the CERIC-ERIC for granting the beamtime 20172056. L.P. acknowledges the support from the Institute from Maternal and Child Health, IRCCS Burlo Garofolo (5mille15D1). The authors are grateful to Dr. Martina Bradaschia (IRCSS Burlo Garofolo) for assistance with the manuscript editing.

CONFLICTS OF INTEREST

The authors declare no financial or commercial conflict of interest.

AUTHOR CONTRIBUTIONS

Francesco D'Amico conceptualize the experiment, Paolo Zucchiatti and Katia Latella prepared the samples and performed the DNA extractions, Francesco D'Amico, Katia Latella and Maria Pachetti perform the UVRR experiments, Alessandro Gessini contributed to prepare the experimental Raman set-up, Francesco D'Amico, Lisa Vaccari and Lorella Pascolo wrote the original draft, Francesco D'Amico, Claudio Masciovecchio, Lisa Vaccari and Lorella Pascolo supervised the entire work.

DATA AVAILABILITY STATEMENT

Data available on request from the authors

ORCID

Francesco D'Amico https://orcid.org/0000-0001-7995-826X
Paolo Zucchiatti https://orcid.org/0000-0002-9491-4668
Maria Pachetti https://orcid.org/0000-0002-1307-7475
Lisa Vaccari https://orcid.org/0000-0003-2355-114X
Lorella Pascolo https://orcid.org/0000-0003-1044-7263

REFERENCES

[1] S. Sharma, T. K. Kelly, P. A. Jones, *Carcinogenesis* **2010**, *31*, 27.
[2] G. Felsenfeld, *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a018200.
[3] B. Chowdhury, I. H. Cho, J. Irudayaraj, *J. Biol. Eng.* **2017**, *11*, 10.
[4] M. A. Dawson, T. Kouzarides, *Cell* **2012**, *150*, 12.
[5] S. Gunes, M. A. Arslan, G. Neslihan, T. Hekim, R. Asci, *J. Assist. Reprod. Gen.* **2016**, *33*, 553.
[6] D. Montjean, A. Zini, C. Ravel, S. Belloc, A. Dalleac, H. Copin, P. Boyer, K. McElreavey, M. Benkhalifa, *Andrology* **2015**, *3*, 235.
[7] S. Kurdyukov, M. Bullock, *Biology* **2016**, *5*, 3.
[8] N. Islam, S. Yadav, H. Haque, A. Munaz, F. Islam, S. Al Hossain, V. Gopalan, A. K. Lam, N. T. Nguyen, M. J. A. Shiddiky, *Biosens. Bioelectron.* **2017**, *92*, 668.
[9] J. G. Kelly, G. M. Najand, F. L. Martin, *Biophotonics* **2011**, *4*, 345.
[10] M. Banyay, A. Gräslund, *J. Mol. Biol.* **2002**, *24*, 667.
[11] L. Li, S. F. Lim, A. Puretzky, R. Riehn, H. D. Hallen, *Biophys. J.* **2018**, *114*, 2498.

[12] K. C. Bantz, A. F. Meyer, N. J. Wittenberg, H. Im, Ö. Kurtuluş, S. H. Lee, N. C. Lindquist, S. H. Oh, C. L. Haynes, *Phys. Chem. Chem. Phys.* **2011**, *13*, 11551.
[13] J. Morla-Folch, R. A. Alvarez-Puebla, L. Guerrini, *J. Phys. Chem. Lett.* **2016**, *7*, 3037.
[14] S. R. Panikkanvalappil, M. A. Mahmoud, M. A. Mackey, M. A. El-Sayed, *ACS Nano.* **2013**, *7*(7), 7524.
[15] S. Ganesh, K. Venkatakrishnan, B. Tan, *Nat. Comm.* **2020**, *11*, 1135.
[16] S. M. A. Hasan, Y. He, T.-W. Chang, J. Wang, M. R. Gartia, *J. Phys. Chem. C* **2019**, *123*, 698.
[17] N. E. Marotta, K. R. Beavers, L. A. Bottomley, *Anal. Chem.* **2013**, *85*, 1440.
[18] X. Liu, Y. Shao, Y. Tang, K. F. Yao, *Sci. Rep.* **2014**, *4*, 5835.
[19] T. G. Burova, V. V. Ermolenkov, G. N. Ten, D. M. Kadrov, M. N. Nurlygaianova, V. I. Baranov, I. K. Lednev, *J. Phys. Chem. A* **2013**, *117*, 12734.
[20] T. G. Burova, V. V. Ermolenkov, G. N. Ten, R. S. Shcherbakov, V. I. Baranov, I. K. Lednev, *J. Phys. Chem. A* **2011**, *115*, 10600.
[21] B. E. Billinghurst, S. A. Oladepo, G. R. Loppnow, *J. Phys. Chem. B* **2009**, *113*, 7392.
[22] T. Bourova, G. Ten, S. Andreeva, V. Berezin, *J. Raman Spectrosc.* **2000**, *31*, 827.
[23] M. Majoube, P. Milliè, L. Chinsky, P. Y. Turpin, G. Vergoten, *J. Mol. Struct.* **1995**, *355*, 147.
[24] S. Sasidharanpillai, G. R. Loppnow, *J. Phys. Chem. B* **2019**, *123*, 3898.
[25] B. E. Billinghurst, S. A. Oladepo, G. R. Loppnow, *J. Phys. Chem. B* **2012**, *116*, 10496.
[26] W. Suen, T. G. Spiro, L. C. Sowers, J. R. Fresco, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4500.
[27] A. Toyama, N. Hanada, J. Ono, E. Yoshimitsu, H. Takeuchi, *J. Raman Spectrosc.* **1999**, *30*, 623.
[28] S. P. A. Fodor, R. P. Rava, T. R. Hays, T. G. Spiro, *J. Am. Chem. Soc.* **1985**, *107*, 1520.
[29] J. Geng, M. Aioub, M. A. El-Sayed, B. A. Barry, *ChemPhysChem* **2018**, *19*, 1428.
[30] J. Geng, M. Aioub, M. A. El-Sayed, B. A. Barry, *J. Phys. Chem. B* **2017**, *121*, 8975.
[31] J. M. Benevides, S. A. Overman, G. J. Thomas Jr., *J. Raman Spectrosc.* **2005**, *36*, 279.
[32] S. P. A. Fodor, T. G. Spiro, *J. Am. Chem. Soc.* **1986**, *108*, 3198.
[33] Z. Q. Wen, G. J. Thomas Jr., *Biopolymers* **1998**, *45*, 247.
[34] L. Chinsky, P. Y. Turpin, *Nucleic Acids Res.* **1978**, *5*, 2969.
[35] P. Zucchiatti, K. Latella, G. Birarda, L. Vaccari, B. Rossi, A. Gessini, C. Masciovecchio, F. D'Amico, *J. Raman Spectr.* **2018**, *49*, 1056.
[36] F. Cammisuli, L. Pascolo, M. Morgutti, A. Gessini, C. Masciovecchio, F. D'Amico, *Appl. Spectr.* **2017**, *71*, 152.
[37] F. D'Amico, F. Cammisuli, R. Addobbati, C. Rizzardi, A. Gessini, C. Masciovecchio, B. Rossi, L. Pascolo, *Analyst* **2015**, *140*, 1477.
[38] J. Marmur, P. Doty, *J. Mol. Biol.* **1962**, *5*, 109.
[39] F. D'Amico, M. Saito, F. Bencivenga, M. Marsi, A. Gessini, G. Camisasca, E. Principi, R. Cucini, S. Di Fonzo, A. Battistoni, E. Giangrisostomi, C. Masciovecchio, *Nucl. Instrum. Methods Phys. Res., Sect. A* **2013**, *703*, 33.
[40] R. L. McCreery, Photometric Standards for Raman Spectroscopy. in *Handbook of Vibrational Spectroscopy*, **2006**, Vol. 1 (Eds: J. M. Chalmers, R. Peter), John Wiley & Sons Ltd, Chichester (England).
[41] M. J. Cavaluzzi, P. N. Borer, *Nucleic Acids Res.* **2004**, *31*, e13.
[42] S. K. Patra, S. Bettuzzi, *Biochem.* **2009**, *74*, 613.
[43] C. C. Oakes, S. La Salle, D. J. Smiraglia, B. Robaire, J. M. Trasler, *Dev. Biol.* **2007**, *307*, 368.
[44] J. F. Linnekamp, R. Butter, R. Spijker, J. P. Medema, H. W. M. van Laarhoven, *Cancer Treat. Rev.* **2017**, *54*, 10.
[45] L. Pascolo, D. E. Bedolla, L. Vaccari, I. Venturin, F. Cammisuli, A. Gianoncelli, E. Mitri, E. Giolo, S. Luppi, M. Martinelli, M. Zweyer, G. Ricci, *Reprod. Toxicol.* **2016**, *61*, 39.