



UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE CHIMICHE

DIPARTIMENTO DI CHIMICA

CORSO DI LAUREA IN CHIMICA

**Studio della correlazione tra i segnali SERS di campioni di DNA e
l'età del donatore**

Relatore: Dr. Lucio Litti

Laureanda: Giada Dall'Amico

2056366

Anno Accademico 2022/2023

INDICE

Abstract	3
1 INTRODUZIONE	4
1.1 <i>Metilazione del DNA</i>	4
1.2 <i>La tecnica Raman</i>	7
1.3 <i>L'amplificazione SERS</i>	10
2 MATERIALI E METODI	14
2.1 <i>Materiali utilizzati</i>	14
2.2 <i>PARTE SPERIMENTALE</i>	15
2.2.1 <i>Substrati solidi con stampa inkjet</i>	15
2.2.2 <i>Substrati solidi con drop-casting</i>	15
2.2.3 <i>Substrati colloidali</i>	16
3 RISULTATI E DISCUSSIONI	17
3.1 <i>Substrati</i>	17
3.2 <i>Concentrazione campioni</i>	17
3.3 <i>Aggregazione delle AuNPs</i>	17
3.4 <i>ANALISI DEI DATI</i>	19
3.4.1 <i>Descrizione generale del modello PLS</i>	19
3.4.2 <i>Risultati ottenuti</i>	19
3.5 <i>Implementazioni per il proseguo del progetto</i>	23
4 CONCLUSIONE	25
5 BIBLIOGRAFIA	26

Abstract

Uno degli effetti più noti del naturale invecchiamento di un organismo sul suo DNA è quello di un aumento del numero dei gruppi metile (-CH₃) attaccati alle sue basi azotate. L'età di metilazione (methylage) è infatti un parametro che serve per stimare l'età biologica del campione ed un grande scostamento dal valore previsto può indicare il precoce insorgere di malattie come l'Alzheimer o il cancro.

Lo scopo di questa tesi è quello di sviluppare un metodo basato sulla tecnica Raman e SERS, grazie all'utilizzo di nanoparticelle di oro (AuNPs), per analizzare campioni di DNA di donatori con età compresa tra i 2 e i 91 anni. Il diverso grado di metilazione del DNA dei campioni è stato rilevato sulla base della variazione di specifici marcatori spettrali, grazie ad un'analisi multivariata mediante Partial Least Squares Regression. La metodologia qui presentata si propone come un semplice e rapido strumento applicato all'analisi di campioni forensi.

“I actually think it will be possible one day to be immortal. [...] Only 20 percent of our longevity is genetically determined. The rest is what we do, how we live our lives and increasingly the molecules that we take. It's not the loss of our DNA that causes aging, it's the problems in reading the information, the epigenetic noise”.

David Andrew Sinclair

1- INTRODUZIONE

1.1 Metilazione del DNA

Le sequenze genetiche di un essere umano, che regolano e determinano le caratteristiche, le funzioni e lo sviluppo dell'organismo, sono contenute, come scoperto nel 1869 da Friedrich Miescher, in 46 cromosomi suddivisi in 23 coppie omologhe. Ciascuna di queste è costituita da una lunga catena composta da due filamenti complementari di acido desossiribonucleico (DNA) ed è proprio la sequenza delle sue quattro basi azotate, che si ibridano attraverso i legami a idrogeno (A=T e C≡G), a definire e custodire l'informazione genetica.^[1] Tuttavia, tale sequenza da sola non è sufficiente per spiegare appieno i caratteri fenotipici che ci contraddistinguono e ruolo importante nei cambiamenti di quest'ultimi è oggi attribuito anche all'epigenetica. Con questo termine si indica l'insieme delle modifiche che influenzano l'espressione dei geni, ed è proprio a questa classe che appartiene anche la metilazione del DNA (DNAm).^[2,3,4] Su questo importante processo biochimico si sono recentemente concentrati numerosi studi, perché è stata notata e confermata una relazione tra la variazione del grado di metilazione e l'invecchiamento biologico, come anche l'insorgere di patologie come i tumori (in particolare al seno e alla prostata), l'Alzheimer, il Parkinson, l'aterosclerosi e le malattie infiammatorie (e per tutte queste l'età è uno dei principali fattori di rischio).^[2,3]

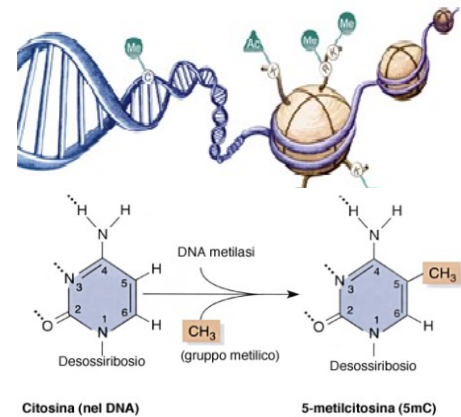


Figura 1: Metilazione del DNA (sopra) e base coinvolta (sotto).

In modo generale si può descrivere l'invecchiamento come un “declino lento e dipendente dal tempo di un insieme di molteplici funzioni biologiche”^[2] e molti dei suoi segnali, come il silenziamento dei geni di riparazione del DNA o di quelli antinfiammatori, comuni in tutti i tipi di cellule umane, possono essere ricondotti ad un'espressione genica alterata. Gli studi effettuati si sono concentrati sul DNA metilato, sia perché è risultato chimicamente più stabile rispetto ad altri biomarcatori come l'RNA messaggero o le proteine^[3], sia perché, essendo un'alterazione epigenetica e non una mutazione, rappresenta un cambiamento più facilmente reversibile e quindi risulta un promettente bersaglio per ideare futuri approcci terapeutici^{1, [2]}

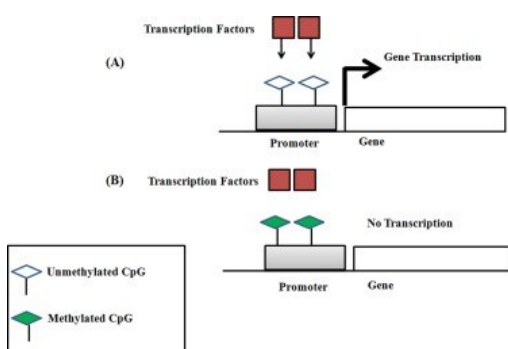


Figura 2: Metilazione e silenziamento genico.

La **metilazione del DNA** comporta il trasferimento covalente di un gruppo metilico dal donatore SAM (S-adenosilmetionina)², o dal folato attivo (5-MTHF), alla posizione C-5 dell'anello di citosina del DNA per formare la 5 metilcitosina (5mC) o altre forme meno diffuse come la 5hmC o la T (fig. 1).^[3] Nei mammiferi, attraverso questo processo, che è mediato dagli enzimi DNA metil-transferasi (DNMT)³, vengono principalmente attaccate le citosine della sequenza 5'-CG-3' (CpG). Tuttavia, esistono alcune di queste regioni genomiche, lunghe fino a 3 kb, in cui,

¹ “As we get better at reversing aging it will be possible to take one medicine and within weeks feel and even look younger. Imagine going to a doctor to get a pill for diabetes, and this same medicine will prevent heart disease, Alzheimer's, cancer, and will give you more vitality too”. David Andrew Sinclair

² che si trasforma quindi in SAH, S-adenosilomocisteina, che è un inibitore delle DNMT.^[3]

³ DNMT1 preserva la metilazione del DNA durante la replicazione, DNMT3A e DNMT3B metilano *de novo* il DNA in nuove regioni genomiche, DNMT2 è una t-RNA metiltransferasi, mentre DNMT3L regola l'attività delle *de novo*.^[3]

almeno fino all'età avanzata, generalmente non si verifica la metilazione. Queste zone particolari, che contengono un alto numero di dinucleotidi CpG, sono comunemente denominate "isole CpG"⁴. Sono molto importanti soprattutto perché lì si trovano circa il 60-70% dei promotori dei geni (soprattutto di quelli "housekeeping", cioè quelli che regolano le funzioni vitali di base in tutte le cellule), con i rispettivi fattori e complessi proteici coinvolti nella trascrizione e regolazione (e.g. il complesso Polycomb).^[2,3,4] Questi, allo stesso tempo, svolgono anche un'azione protettiva contro la metilazione (fig. 3, sopra). La presenza di gruppi promotori metilati nelle isole CpG, infatti, impedisce il legame con i fattori di trascrizione e questo, di conseguenza, determina il silenziamento genico o trasformazioni maligne responsabili dell'insorgere dei tumori, che sono appunto una delle malattie più strettamente collegate con l'età (fig. 2). Si ipotizza che molte delle isole non metilate siano normalmente riconosciute e legate al complesso Polycomb (proteina KDM2B che recluta PRC1 e quindi PRC2), che però con l'età si degrada, permettendo gradualmente alla DNA metilasi di accedere (fig. 3, sotto).^[2] L'ipermetilazione del DNA dei geni bersaglio della proteina del gruppo Polycomb (PcG), che in medicina è stata chiamata "commutazione Polycomb", è stata riscontrata anche in molti studi sul cancro ed ha come risultato una perdita di plasticità dal momento che la metilazione è considerata una modifica molto stabile.^[2]

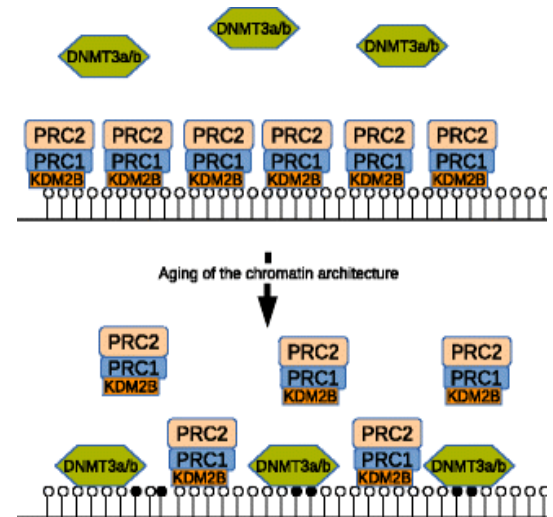


Figura 3: Meccanismo di protezione dalla metilazione svolto dal complesso Polycomb.

I cambiamenti nei modelli di metilazione del DNA si verificano per tutta la vita a partire dalla nascita e devono essere distinti in due diverse categorie: una ipometilazione a livello generale e una ipermetilazione di siti⁵ e regioni specifiche.^[4] Per quanto riguarda l'andamento complessivo, gli studi sul sangue hanno evidenziato che, dopo l'aumento nel primo anno di vita, i livelli mediani globali di metilazione del DNA sono relativamente stabili e restano sostanzialmente costanti durante l'infanzia e l'adolescenza. Sembra poi, da ricerche effettuate sia nel cervello che nel sangue, che la quantità di DNAm aumenti rapidamente fino all'età adulta, per poi stabilizzarsi e rimanere costante fino all'età avanzata. Allo stesso tempo però, alcuni studi evidenziano come la variabilità tra gli individui, denominata "deriva epigenetica", sembri aumentare maggiormente a partire dalla fase adulta della vita.^[4] Nella fase matura, molte ricerche condotte sul sangue concordano nel ritenere che vi sia una diminuzione media della metilazione del DNA, cosa che solitamente si verifica anche con il cancro.^[2,4]

I motivi più rilevanti per spiegare questa riduzione in età avanzata sono il rallentamento del metabolismo di base (necessario per la biosintesi di SAM) e la riduzione dell'attività delle DNMT legate alla diminuzione di quella dell'ormone della crescita.^[2] C'è però da sottolineare che questi studi sono stati effettuati principalmente sul sangue e, dal momento che il grado di metilazione è specifica per ogni tessuto, non risulta ancora possibile generalizzare l'andamento individuato. L'ipometilazione complessiva potrebbe essere spiegata dal fatto che con l'età aumenta la metilazione solamente nelle isole CpG, mentre le regioni ripetitive e i siti CpG "non insulari", che sono la maggioranza, tendono a subire una diminuzione. Inoltre, i cambiamenti riscontrati (aumento nella prima fase della vita e diminuzione nella terza età) non sono simmetrici poiché il tasso di variazione è molto più alto all'inizio rispetto all'età avanzata.^[4] Anche le posizioni genomiche dei cambiamenti

⁴ Si stima che, in media, dopo la fase di sviluppo, il genoma delle cellule somatiche sia costituito per circa l'1,5% da 5mC.^[2,3]

⁵ Come quelli, in alcune isole CpG, dei promotori dei geni soppressori tumorali, che con la metilazione si inattivano.^[3]

sono piuttosto diverse, dal momento che in età giovanile l'aumento di DNAm avviene a livello generale e soprattutto nei siti CpG non insulari, mentre con il progredire dell'età viene perso a livello globale, ma permane nelle isole.^[4] Tuttavia, nel tentativo di creare un modello predittivo per l'età dei campioni di DNA, la “deriva epigenetica”, che comprende i cambiamenti epigenetici causati da fattori ambientali o da eventi stocastici durante le divisioni cellulari, rappresenta una variabile difficile da prevedere e controllare. Generalmente la si può osservare negli studi trasversali (gruppi di individui di età diversa), ma per identificare le sue regioni bersaglio all'interno di un individuo sono necessari anche studi longitudinali (stessi individui in più punti temporali).^[4]

A causa quindi del complesso andamento generale, della variabilità tra i tessuti e della **deriva epigenetica**, molti modelli di previsione dell'età proposti finora hanno preferito concentrarsi solo su siti specifici del genoma che ormai si ritengono strettamente correlati all'età (AR-CpG) e/o a determinate malattie per tutti gli individui. Alcuni di questi sembrano essere altresì comuni anche a tutti i tessuti.^[3,4]

Questi loci, ovvero le posizioni specifiche all'interno dei cromosomi dove si trova un determinato gene, costituiscono la base dell’**“orologio epigenetico”** e per questi sembra esserci una relazione lineare tra l'età epigenetica e quella cronologica (fig. 4 e 5). Gli studi sull'orologio epigenetico fin ora si sono concentrati principalmente sul sangue, trovando 11 siti tessuto-specifici che sono considerati buoni predittori dell'età (8 si trovano nelle isole CpG e mostrano un aumento, mentre 3 si trovano al di fuori, nelle coste delle isole, e presentano una diminuzione) e altri 5 pan-tessutali.^[4] Ad oggi, i loci genici più validi e valutati (con un R pari al 96% e 3.9 anni di Errore Standard) sembrano essere quelli di ELOVL2 (intertestuale), che ha il vantaggio di poter essere valutato anche in vecchi campioni di sangue.^[3] Il fatto stesso, però, che sembra effettivamente esistere questo “orologio epigenetico” comune tra tutti gli individui, dà fondamento all'utilizzo del DNA metilato come biomarcatore. Si capisce, quindi, perché ci sia molto interesse da parte della comunità scientifica per trovare un valido modello in grado di predire l'età epigenetica di un individuo, con applicazioni che sarebbero utili in campo forense e soprattutto medico. In questo modo, infatti, sarebbe possibile prevedere gli esiti di salute per coloro che appartengono ai gruppi epigeneticamente anziani (sopra la retta in fig. 5) o epigeneticamente giovani (sotto).^[4]

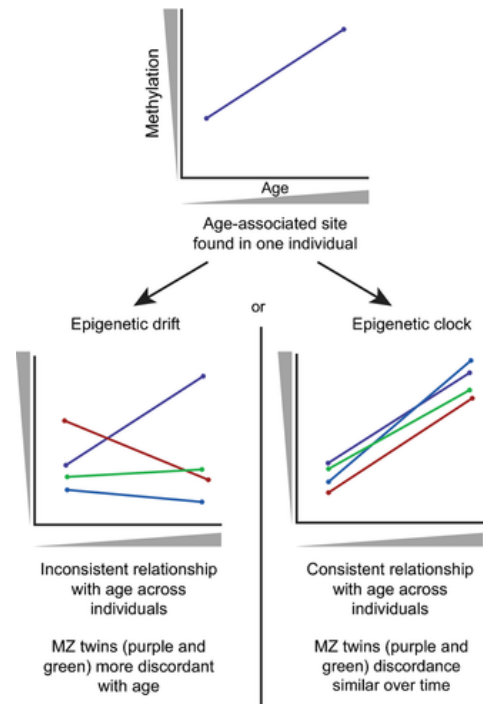


Figura 4: Differenza tra deriva epigenetica ed orologio epigenetico.

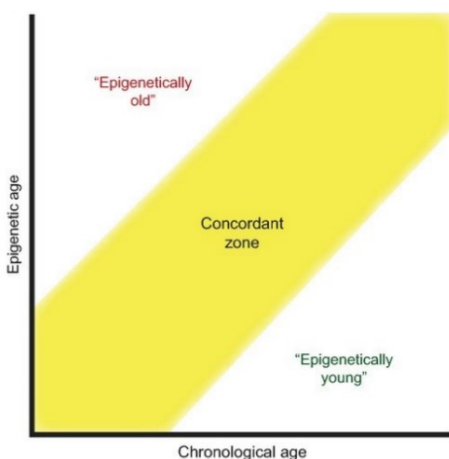


Figura 5: Relazione tra età cronologica ed epigenetica.

Tutti gli studi analizzati concordano sul fatto che la metilazione sia associata in qualche modo all'invecchiamento e che nel corso della vita i modelli epigenetici subiscano delle variazioni, ma resta ancora da chiarire se questo ne sia una causa o una conseguenza. Attualmente, la tecnica d'indagine principale in questo campo sembra essere il sequenziamento con l'array Illumina 450K^[4], che si concentra su siti locali e specifici dei geni che vengono metilati (e quindi inattivati). In questa tesi, invece, è stata utilizzata una tecnica diversa, il Raman potenziato mediante SERS, per cercare di costruire un modello dell'età dai segnali del DNA “complessivo”.

1.2 La tecnica Raman

La spettroscopia (di scattering) Raman è, come l'IR, una spettroscopia vibrazionale, ma l'origine dei segnali associati è nettamente diversa, poiché in questa tecnica si irradia il campione con intensi raggi laser monocromatici nel UV-visibile-NIR e viene misurata la luce diffusa.^[5]

In questa spettroscopia, quando la radiazione elettromagnetica interagisce con gli elettroni della molecola, induce su di essi un dipolo elettrico oscillante, che è il responsabile della diffusione della radiazione incidente, formata da tre componenti con diverse energie: Rayleigh, Stokes ed Anti-Stokes. In tutte e tre un fotone incidente di energia $h\nu_0$ porta la molecola da uno stato vibrazionale a uno degli infiniti stati virtuali situati tra lo stato fondamentale e il primo stato eccitato. Quello che le distingue è quindi il diverso modo di rilassamento.

- **Lo scattering Rayleigh:** rappresenta la quasi totalità della radiazione diffusa e deriva da un processo di diffusione elastico che quindi non scambia energia con il sistema. Il fotone diffuso, perciò, ha la stessa energia della radiazione incidente.
- **Lo scattering Stokes:** avviene se la molecola, che partiva dallo stato fondamentale $v = 0$, si rilassa anelasticamente tornando ad uno stato vibrazionale più elevato $v \geq 1$. Questo corrisponde a un fotone con energia minore, pari a $h\nu_0 - \Delta E_{0-v}$.
- **Lo scattering Anti-Stokes:** si verifica quando la molecola, partendo da uno stato vibrazionale eccitato $v \geq 1$, torna allo stato vibrazionale fondamentale $v = 0$ rilassandosi anelasticamente. Il fotone ha quindi un'energia maggiore, pari a $h\nu_0 + \Delta E_{0-v}$.

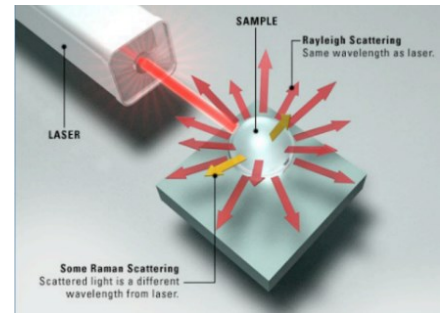


Figura 6: Luce diffusa elasticamente e anelasticamente.

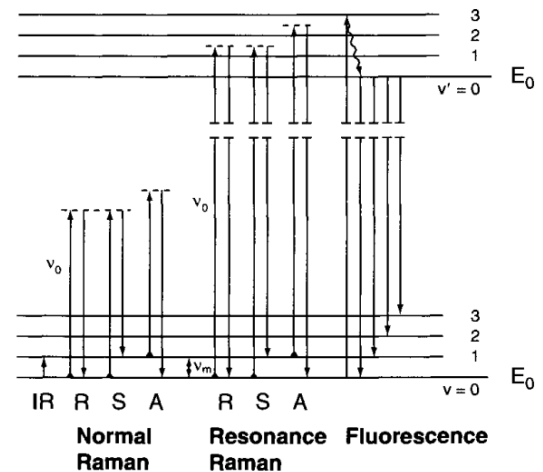


Figura 7: Rappresentazione schematica delle transizioni energetiche coinvolte nel Raman.

Gli spettri Raman di solito riportano sull'asse x il Raman Shift mentre sull'asse y sono riportate le intensità, che sono proporzionali al numero di fotoni diffusi rilevati. Il Raman Shift misura la differenza tra la luce laser utilizzata e quella diffusa di Stokes ($\Delta\bar{\nu}$ in fig. 8). Solitamente non viene utilizzata quella di Anti-Stokes perché, secondo la distribuzione di Boltzmann, risulta essere la meno probabile tra le due⁶.

È possibile spiegare il fenomeno dello scattering Raman attraverso una **derivazione classica per la polarizzazione** indotta in un sistema molecolare biatomico da un campo elettrico monocromatico.^[5]

Il campo elettrico dell'onda elettromagnetica in funzione del tempo può essere scritto secondo:

$$\vec{E} = \vec{E}_0 \cos(\omega_0 t) \quad (1)$$

Dove E_0 è l'ampiezza massima, ω_0 la pulsazione dell'onda elettromagnetica pari a $\omega_0 = 2\pi\nu_0$ e ν_0 la frequenza del laser usato.

⁶ Questo non viene spiegato con la trattazione classica, ma con la teoria quantistica delle perturbazioni dipendenti dal tempo. All'equilibrio termico, alla temperatura T, vale: $\frac{I_{\text{Stokes}}}{I_{\text{Anti-Stokes}}} = \frac{(\nu_0 - \nu_1)^4}{(\nu_0 + \nu_1)^4} e^{(h\nu_1/kT)}$

Quando una molecola biatomica viene irradiata da questa luce, viene indotto un momento di dipolo elettrico \vec{p} pari a:

$$\vec{p} = \alpha \cdot \vec{E} = \alpha \vec{E}_0 \cos(\omega_0 t) \quad (2)$$

Dove α è una grandezza tensoriale chiamata polarizzabilità e descrive il modo e l'entità con la quale la distribuzione elettronica molecolare viene perturbata sotto l'influenza di un campo elettrico (fig. 9).

Se la molecola sta vibrando con una frequenza ν_m (fig. 7), lo spostamento nucleare q può essere descritto, introducendo l'ampiezza vibrazionale q_0 , come:

$$q = q_0 \cos(2\pi\nu_m t) \quad (3)$$

Quando la variazione è piccola, è possibile scrivere α come una funzione lineare di q con uno sviluppo di Taylor troncato al primo ordine:

$$\alpha = \alpha_0 + \left(\frac{\partial\alpha}{\partial q}\right)_0 q + \dots \quad (4)$$

Dove α_0 rappresenta la polarizzabilità nella posizione di equilibrio mentre $\left(\frac{\partial\alpha}{\partial q}\right)_0$ rappresenta la variazione di α rispetto alla variazione di q , calcolato nella posizione di equilibrio.

Mettendo ora assieme le equazioni (2), (3) e (4) si ottiene:

$$\begin{aligned} \vec{p} &= \alpha \cdot \vec{E} = \alpha \vec{E}_0 \cos(\omega_0 t) = \alpha_0 \vec{E}_0 \cos(\omega_0 t) + \left(\frac{\partial\alpha}{\partial q}\right)_0 q \vec{E}_0 \cos(\omega_0 t) + \dots = \\ &= \alpha_0 \vec{E}_0 \cos(2\pi\nu_0 t) + \left(\frac{\partial\alpha}{\partial q}\right)_0 q_0 \cos(2\pi\nu_m t) \vec{E}_0 \cos(\omega_0 t) + \dots \end{aligned} \quad (5)$$

Utilizzando ora in (5) la seconda formula trigonometrica di Werner si ottiene:

$$\vec{p} = \underbrace{\alpha_0 \vec{E}_0 \cos(2\pi\nu_0 t)}_{1: \text{Rayleigh}} + \left(\frac{\partial\alpha}{\partial q}\right)_0 \vec{E}_0 q_0 \cdot \left\{ \underbrace{\cos[2\pi(\nu_0 + \nu_m) t]}_{2: \text{Anti-Stokes}} + \underbrace{\cos[2\pi(\nu_0 - \nu_m) t]}_{3: \text{Stokes}} \right\} \quad (6)$$

Da (6) si vede che le componenti 2 e 3 sono linearmente dipendenti dall'intensità del campo elettrico incidente e dalla variazione della polarizzabilità $\left(\frac{\partial\alpha}{\partial q}\right)_0$, che deve essere diversa da zero perché la vibrazione sia Raman attiva (fig. 9).

La regola di selezione, dunque, è diversa da quella dell'IR, che si basa sulla variazione del momento di dipolo: $\left(\frac{\partial\mu_{gg}}{\partial q}\right)_0 \neq 0$. Le due tecniche, quindi, forniscono informazioni differenti o addirittura complementari se la molecola è centro-simmetrica (regola di mutua esclusione).

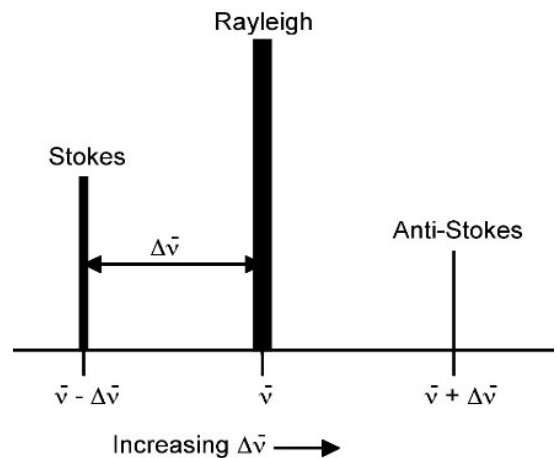


Figura 8. Rappresentazione del Raman Shift e della posizione e intensità relativa delle tre componenti.

In generale, si può affermare che le vibrazioni forti al Raman sono quelle attribuibili ad un legame covalente (molecola polarizzabile con elettroni debolmente legati), mentre sono forti all'IR i legami ionici e polari, che sono invece molto deboli al Raman dal momento che gli atomi tendono a trattenere vicino a sé gli elettroni stabilmente.

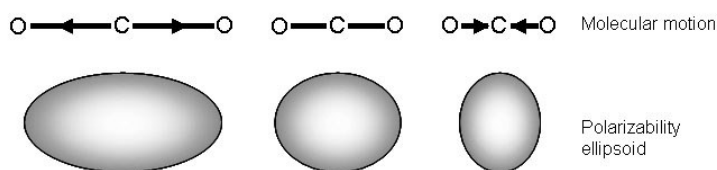


Figura 9: Rappresentazione grafica della variazione del moto/momento di dipolo (sopra) e della polarizzabilità (sotto) di CO_2 .

Il Raman inoltre, a differenza dell'IR, è spesso usato con campioni biologici e può quindi essere utilizzato per lo scopo perseguito con questa tesi. Il motivo di ciò risiede nel fatto che la tecnica consente di ottenere spettri anche da campioni in soluzione acquosa (l'acqua, infatti, è un debole diffusore Raman). Tuttavia,

mostra anche due importanti limitazioni per le quali sono state trovate soluzioni:

1. Alcuni composti emettono fluorescenza e questo potrebbe coprire i loro segnali Raman molto più deboli. Dal momento che questo si verifica principalmente a lunghezze d'onda del visibile e vicino UV, è sufficiente lavorare a energie minori per ridurre il fenomeno (e.g. 785 nm).
2. Il segnale Raman è intrinsecamente molto debole per la bassa sezione d'urto (circa 1 fotone su circa 10 milioni incidenti subisce una diffusione Raman). Questo è stato migliorato grazie all'uso di Surface Enhanced Raman Scattering, che amplifica il segnale con l'utilizzo di nanoparticelle metalliche di oro o argento, che presentano inoltre una bassa citotossicità ed un'eccellente biocompatibilità (specialmente Au)^{7, [6]}

⁷ Nei farmaci/biosensori le interazioni DNA/AuNP agiscono inoltre da agenti protettivi contro la degradazione delle nucleasi, aumentando il tempo di circolazione, l'efficacia e la biodisponibilità dei trattamenti.^[6]

1.3 L'amplificazione SERS

La tecnica del Surface Enhanced Raman Scattering, o SERS, è la tecnica di rilevamento in cui i segnali Stokes della luce diffusa vengono notevolmente potenziati grazie all'utilizzo di superfici metalliche composte da nanoparticelle (NPs) di oro o d'argento, su cui vengono adsorbite le molecole di interesse.^[7] I fattori di amplificazione (EF) medi sono generalmente compresi tra $10^4/10^5$ ^[8], ma in alcuni punti specifici, chiamati hot spots (fig. 11), si può arrivare fino ad un'amplificazione di 10^8 o più, che consente di ottenere informazioni spettroscopiche molto dettagliate fino al rilevamento della singola molecola.

SERS, in realtà, non fornisce esattamente le stesse informazioni della tecnica Raman amplificata, dal momento che, con l'adsorbimento sulla superficie, la simmetria del sistema cambia e le regole di selezione risultano meno rigide.^[9] È stato per esempio osservato che le intensità delle bande in SERS tendono a diminuire con l'aumento della frequenza vibrazionale e, infatti, gli stretching C-H, che sono vibrazioni interessanti per questo lavoro di tesi, tendono ad essere relativamente deboli.

La tecnica è stata scoperta casualmente nel 1974 da Fleischmann e dai suoi colleghi mentre stavano studiando i segnali Raman prodotti dalla piridina adsorbita su una superficie rugosa di elettrodi d'argento e da allora molti meccanismi sono stati proposti per spiegare l'amplificazione del segnale, ma, ad oggi, le due teorie più accreditate si dividono in **teoria elettromagnetica**, che si riferisce all'aumento del campo elettrico, e in **teoria chimica**, che si focalizza sull'aumento della polarizzabilità.^[1,9]

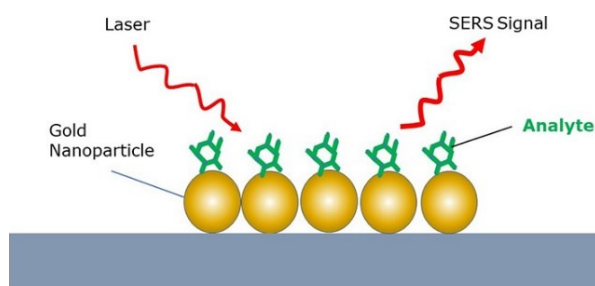


Figura 10: Schema del funzionamento di SERS.

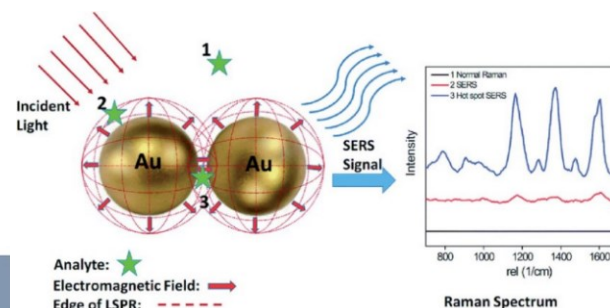


Figura 11: Rappresentazione della risonanza plasmonica e degli hot spots.

1) Quando la luce colpisce la superficie del metallo, il suo campo elettromagnetico interagisce con gli elettroni presenti nella “nuvola” di carica elettrica negativa delle NPs, che si muove liberamente attorno agli atomi del metallo (una semplificazione della configurazione delle cariche nel plasma bulk, che corrisponde a un dipolo elettrico^[10])⁸. Questo provoca un movimento longitudinale collettivo degli elettroni (che si può immaginare come un’onda che si propaga sulla superficie), chiamato “plasmon”, che genera un campo elettrico locale (LSP) molto intenso nei pressi della superficie metallica interagente con le molecole lì adsorbite. L’amplificazione massima del segnale,

⁸ I plasmi bulk, infatti, possono sostenere oscillazioni longitudinali del plasma (plasmoni), la cui frequenza di risonanza nasce dalla forza di ripristino che la distribuzione di carica alterata esercita sulle cariche mobili quando vengono spostate dall’equilibrio, ad esempio dal passaggio di un fotone. Per un plasma bulk la frequenza plasmonica è data da: $\omega_p^2 = \frac{ne^2}{\epsilon_0 m}$

Dove n, e, m indicano rispettivamente la densità numerica, la carica e la massa degli elettroni liberi e ϵ_0 la permittività relativa del vuoto.^[10]

conosciuta anche come risonanza plasmonica, si verifica quando la frequenza della luce incidente corrisponde alla frequenza di risonanza ω_p degli e^- della banda di conduzione del metallo e dipende dalla forma, dalle dimensioni e dalla composizione del materiale metallico. Il grande EF che si riesce ad ottenere dipende dal fatto che il campo viene amplificato due volte, sia “in entrata”, ingrandendo l’intensità della luce incidente, sia “in uscita” amplificando il segnale Raman prodotto. Semplificando si può considerare che in entrambe i plasmoni locali di superficie amplifichino il campo al quadrato, anche se in realtà la luce dispersa ha una frequenza diversa da quella incidente. I campi, perciò, possono essere entrambi quasi in risonanza con il plasmon di superficie solo se il Raman Shift è piccolo e l’effetto risulta tanto più grande quanto più la molecola si trova vicino alla nanoparticella (fig. 11).

L’esempio classico di una sfera metallica in un campo elettromagnetico esterno illustra bene la fisica essenziale alla base del fenomeno. Se il raggio della particella è molto più piccolo della lunghezza d’onda della luce, il campo elettrico risulta uniforme in tutta la particella e l’approssimazione elettrostatica (Rayleigh) è valida. In questo caso, il campo indotto sulla superficie della sfera e il campo esterno (laser) sono legati dall’equazione^[9]:

$$E_{indotto} = \{[\varepsilon_1(\omega) - \varepsilon_2]/[\varepsilon_1(\omega) + 2\varepsilon_2]\} \cdot E_{laser} \quad (7)$$

con $\varepsilon_1(\omega)$ funzione dielettrica complessa dipendente dalla frequenza del metallo e ε_2 è la permittività relativa della fase ambiente, che nel lavoro svolto è il solvente che circonda le nanostrutture. Questa funzione è risonante alla frequenza per cui la parte reale di $\varepsilon_1(\omega)$ è pari a $Re(\varepsilon_1(\omega_p)) = -2\varepsilon_2$.^[9] La scelta di utilizzare Au o Ag deriva proprio dal fatto che la condizione di risonanza è soddisfatta alle frequenze del visibile comunemente utilizzate in spettroscopia Raman e dal fatto che la parte immaginaria della funzione dielettrica, che misura le perdite di energia (riferite comunemente ad assorbimenti), a tale frequenza è molto piccola e questo porta a risonanze più nitide e intense.^[9] Il fenomeno SERS può avvenire sulla singola nanoparticella metallica, ma è ancora più elevato in presenza di aggregati di NPs (e.g. nei colloidali) perché si formano numerosi hot spots associati ai “gap plasmonici”, cioè ai piccoli spazi tra le nanoparticelle metalliche in cui, grazie alle interferenze costruttive, la risonanza plasmonica è particolarmente intensa (fig. 11).^[8]

2) Il meccanismo chimico, molto più debole per quanto concomitante con il primo, può spiegare come le molecole CO e N₂, pur con la stessa polarità, differiscano di un fattore di 200 nelle loro intensità SERS. Questo è stato giustificato mediante trasferimenti di carica tra le specie chemiosorbite e la superficie metallica.^[9] Questo meccanismo si verifica quando i livelli elettronici HOMO e LUMO della molecola cadono simmetricamente intorno al livello di Fermi (fig. 12) della superficie metallica, così che sono sufficienti anche fotoni con metà dell’energia (è appropriata la luce visibile o infrarossa) normalmente richiesta per l’assorbimento (che si colloca generalmente nello ultravioletto), in quanto il metallo funge da intermedio durante il trasferimento di carica.

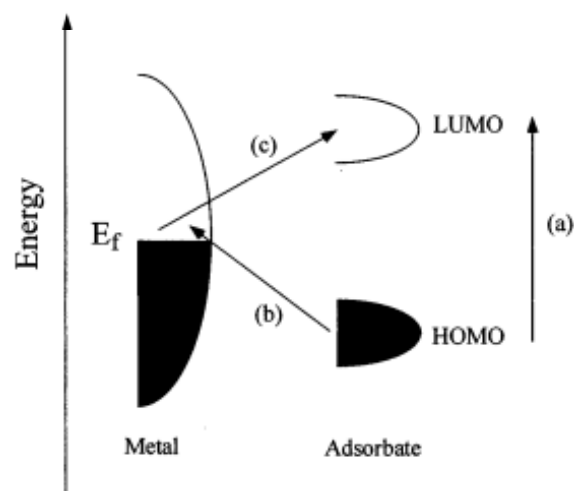


Figura 12: Rappresentazione dei livelli e dei possibili trasferimenti energetici tra molecola e metallo.

⁹ Per le altre forme cambia il fattore geometrico 2 di risonanza.

Un importante ruolo e una delle maggiori sfide attuali in questo ambito è lo **sviluppo di substrati SERS** altamente riproducibili, affidabili e con proprietà ottiche e di campo quasi uniformi e prevedibili. Soprattutto a livello accademico, l'interesse è rivolto verso la produzione di substrati realizzati in modo estremamente preciso, con un perfetto controllo della dimensione/geometria dello hot spot per ottenere gli EF più alti e riproducibili possibili.^[7]

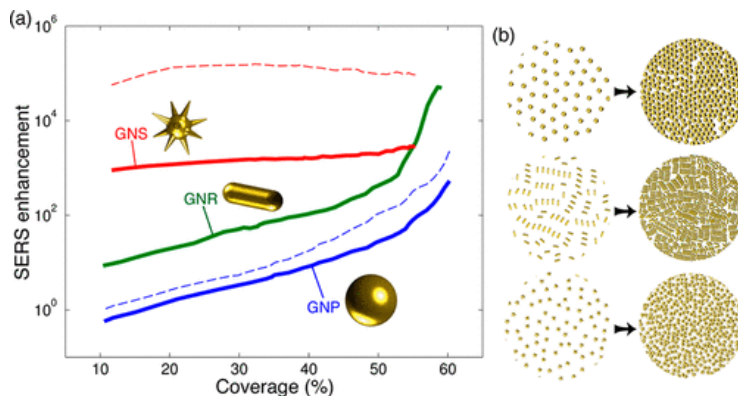


Figura 13: Confronto tra gli EF a seconda della % di copertura per le diverse forme nei substrati solidi.

I **substrati solidi** sono generalmente più stabili e convenienti da maneggiare, sia durante l'uso che durante lo stoccaggio. Possono essere ottenuti mediante l'utilizzo di una **stampante inkjet** che utilizza un inchiostro composto da nanoparticelle di AuNPs e altri additivi come l'etilene glicole per controllarne la viscosità e ottenere uno spot (film metallici) di dimensione, spessore e compattezza controllata (fig. 13). Inkjet è molto interessante a livello industriale soprattutto perché è facilmente scalabile e riduce gli sprechi di inchiostro. Una tecnica meno onerosa in termini di tecnologia e costi, ma anche meno precisa (minor controllo della geometria e della copertura), consiste nell'ottenere gli spots attraverso la tecnica di **drop-casting**, cioè di deposizione (manuale) in situ delle gocce di NPs su un substrato solido come il vetro, scaldando ad alta temperatura per favorire la crescita e l'autoassemblaggio delle NPs, oltre che per evaporare il solvente. La fase più critica di questa tecnica è la diffusione delle gocce a causa dei flussi verso l'esterno¹⁰.^[11]

Quando però si vuole utilizzare SERS per indagini chimiche comuni, sono più importanti l'omogeneità su larga scala, la riproducibilità da lotto a lotto del substrato, una fabbricazione veloce, semplice ed economica, una robusta intensità del segnale SERS, un'elevata affinità verso l'analita di interesse e una facile funzionalizzazione (e.g. con PEG-SH). La maggior parte di questi requisiti sono agevolmente ottenibili utilizzando **substrati colloidali** di NPs in soluzione.

Le NPs nude possono essere prodotte (come avviene in questa tesi) per ablazione laser pulsata in solventi (LASiS o PLAL) che, a differenza della sintesi chimica, non contamina le nanoparticelle con stabilizzanti. Le NPs più utilizzate mostrano una semplice morfologia sferica che, pur presentando un EF minore rispetto a quello ottenibile mediante altre forme anisotrope (e.g. nanostelle), sono spesso utilizzate per la loro tendenza ad aggregare in soluzione¹¹ a causa delle cariche superficiali presenti e dell'effetto della carica ionica del solvente. Un vantaggio, infatti, dell'utilizzo dei colloidali risiede nel fatto che i

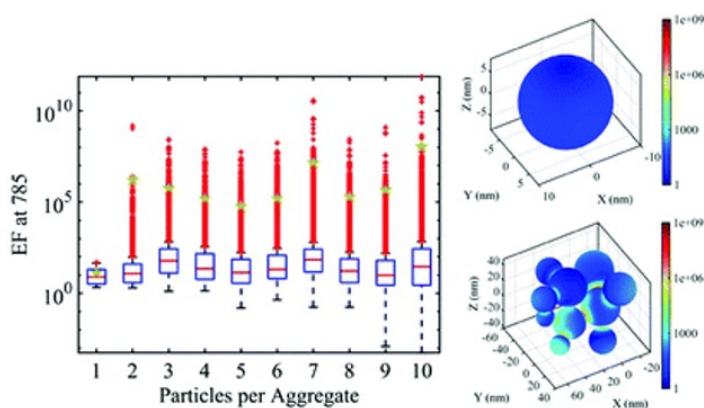


Figura 14: Rappresentazione degli EF per NPs aggregati, con gli hot spots rappresentati dalle stelle (sx) e illustrazione con i colori degli hot spots in un aggregato (dx).

¹⁰ Si può formare l'indesiderato effetto "coffee ring".

¹¹ Questi processi possono essere modellizzati con la teoria cinetica che stima che vi siano circa 10^4 collisioni di particelle al secondo in una soluzione da 1 mL contenente ~ 5 nM nanostrutture.^[7]

vari contributi del potenziale, come quello attrattivo di Van der Waals, quello di superficie elettrostatico, quello dato dalla composizione del solvente e dai ligandi di superficie, possono essere bilanciati per ottenere nanostrutture con gap ideali (0,8 -1,5 nm) per l'eccitazione SERS nel visibile.^[7] Grazie alla simulazione B.E.M (Boundary Element Method) si è potuto verificare che gli hot spots negli aggregati colloidali sono quasi indipendenti l'uno dall'altro (fig. 14, sx) e l'aumento che si registra per un aggregato più grande rispetto al dimero risiede nel maggior numero di punti caldi disponibili.^[8] In generale, si stima che in un colloide solo lo 0,5% degli elementi di superficie aggregati contribuiscano a un'amplificazione maggiore di 10^4 , che poi si riduce drasticamente a meno dello 0,05% per più dello 10^7 . Il valore di amplificazione stimabile o calcolabile è un valore medio e a 785 nm per AuNPs aggregati è dell'ordine di 10^5 .

Indipendentemente dal tipo di substrato SERS utilizzato, bisogna sempre tenere in considerazione che è necessario un certo intervallo temporale perché le molecole riescano ad attaccarsi abbastanza stabilmente alla superficie delle nanoparticelle. Proprio per questo, e dato che l'intensità del segnale cresce con l'aumentare dell'incubazione^[1], gli spettri SERS devono essere registrati entro un certo tempo ben preciso e nelle stesse condizioni sperimentali per poter essere riproducibili. L'adsorbimento sulla superficie delle AuNPs non funzionalizzate dei nucleotidi, i monomeri del DNA (fig. 15), è fortemente influenzato dalle proprietà chimiche degli stessi e perciò l'esatta natura e la forza dell'interazione del doppio filamento dipendono in modo significativo dalla sequenza delle basi (l'ordine di affinità in presenza di PBS sembra essere $dA > dG > dC >> dT$) che si legano mediante legami a idrogeno. Il double stranded DNA, a causa della maggiore rigidità, delle repulsioni elettrostatiche tra i fosfati e le NPs e della minore disponibilità delle sue basi, presenta un'interazione più complessa e meno favorevole rispetto ai suoi monomeri.^[6]

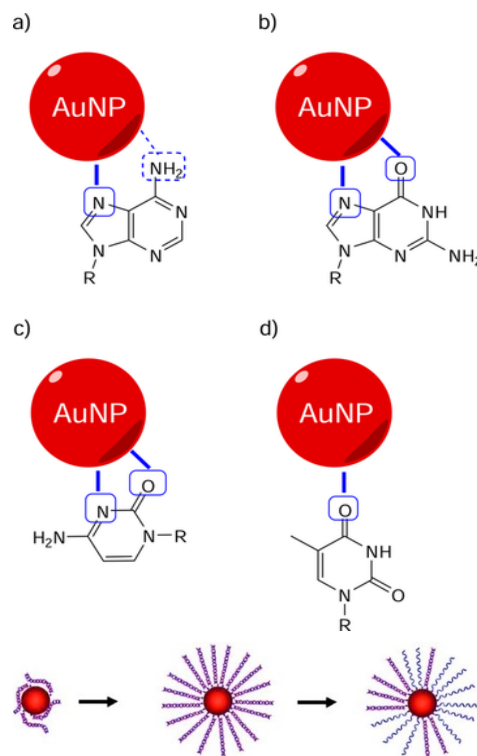


Figura 15: Sopra interazione dell'AuNP con: a) A, b) G, c) C, d) T; sotto con il dsDNA. a C crescente.

2- MATERIALI E METODI

2.1 Materiali utilizzati

- Campioni di DNA:** I 28 campioni utilizzati sono quelli riportati in tabella 1 e hanno un'età compresa tra i 2 e i 91 anni, divisa in 10 fasce di età. Sono stati forniti e preparati in collaborazione con il Dipartimento di Chimica dall'Università degli Studi di Torino ed il Centro Regionale Antidoping e di Tossicologia "A. Bertinara" di Orbassano (TO), dopo essere stati estratti dalla saliva di volontari. I campioni si trovavano in circa 230 μL di buffer TE (10 mM Tris, 0,1 mM EDTA, pH =8,0). Sono sempre stati conservati in frigo a 4°C, tranne per circa 10/15 minuti necessari per prelevare il volume previsto in tabella sotto. I campioni presentano una concentrazione di DNA varia (con * se $C < 1 \text{ ng}/\mu\text{L}$) e per le analisi sono stati tutti diluiti secondo lo schema sotto riportato, al fine di omogeneizzarli in concentrazione.
- Nanoparticelle d'oro:** Le AuNPs sferiche utilizzate sono state prodotte per ablazione laser in una soluzione $5 \cdot 10^{-5} \text{ M}$ di NaCl e acqua distillata.
- Strumenti:** gli **spettri Uv-Vis** sono stati registrati utilizzando uno spettrometro Agilent Instrument Cary 5000 UV-Vis-NIR. Per gli **spettri Raman** invece è stato utilizzato un Renishaw InVia μRaman (con microscopio) con laser a diodi che emette a 785 nm. Tutti gli spettri dell'UV-Vis e della soluzione colloidale sono stati registrati con una cuvetta in quarzo da 2 mm (fig. 21).

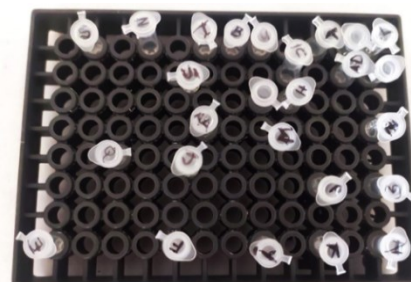


Figura 16: Provette con campioni di DNA e acqua milliQ.

Tabella 1: Volume di campione prelevato per omogenizzarli tutti a C pari a 1 $\text{ng}/\mu\text{L}$, fascia d'età e ordine di misura.

Campione	Età	Fascia d'età	C ini ($\text{ng}/\mu\text{L}$)	V prele (μL)	V acqua (μL)	Numero
A	2	0-10	4.19	23.87	76.13	12
B	4	0-10	12.63	7.92	92.08	2
C	5	0-10	12.79	7.82	92.18	8
D	11	10-20	15.51	6.45	93.55	14
E	17	10-20	5.19	19.27	80.73	20
F	19	10-20	5.75	17.39	82.61	22
G	22	20-30	6.07	16.47	83.53	16
H	24	20-30	12.68	7.89	92.11	4
I	26	20-30	12.58	7.95	92.05	18
K	31	30-40	6.29	15.90	84.10	24
J	33	30-40	25.01	4.00	96.00	6
L	37	30-40	19.34	5.17	94.83	10
M	46	40-50	4.13	24.21	75.79	26
N	48	40-50	5.72	17.48	85.52	17
O	48	40-50	6.4	15.63	84.38	28
P	51	50-60	2.54	39.37	60.63	11
Q	55	50-60	1.78	56.18	43.82	27
R	56	50-60	27.14	3.68	96.32	1
S	60	60-70	1.59	62.89	37.11	19
T	61	60-70	2.65	37.74	62.26	13
U	67	60-70	4.27	23.42	76.58	7
V*	74	70-80	0.79	*	-	25
Y	80	70-80	2.02	49.50	50.50	23
Z	83	80-90	4.83	20.70	79.30	5
AA	85	80-90	2.42	41.32	58.68	9
AB*	89	80-90	0.71	*	-	15
AC*	91	90-100	0.03	*	-	21
AD	91	90-100	4.49	22.27	77.73	3

2.2 PARTE SPERIMENTALE

Il lavoro sperimentale può essere diviso in due parti: la prima consiste nell'ottenere un buon substrato SERS in termini di praticità e segnale, mentre la seconda nella registrazione degli spettri con la tecnica scelta.

2.2.1 Substrati solidi con stampa inkjet

Inchiostro: È stato utilizzato il surnattante ottenuto da una precedente centrifuga con NPs (ablate in data 14/03/23) funzionalizzate con PEG-SH. Per concentrare le particelle rimanenti, è stata eseguita per tre volte una centrifuga a 3000 RCF per mezz'ora. Una volta prodotto l'inchiostro (circa 1 mL totale recuperato), è stato eseguito lo spettro UV-visibile (fig. 23) per controllare che le NPs non fossero aggregate (per non intasare gli ugelli della stampante con cartuccia da 1 pL) ed è stato quindi aggiunto etilen glicole al 10% per migliorarne la viscosità e 2/3 mL ogni 0,5 mL di NPs di tween 20, un tensioattivo per evitare la formazione di aggregati.

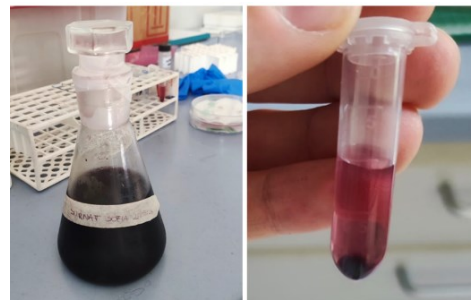


Figura 17: A sx il surnattante di partenza, a dx l'inchiostro concentrato sul fondo dopo la centrifuga.

Substrato: Lo spot circolare da 1 mm^2 , composto da 6 strati, è stato stampato sul vetrino ricoperto di alluminio. Sopra si è attaccato, con la colla verde, un pozzetto in plastica per contenere la soluzione di campione. Non sono stati registrati gli spettri Raman.

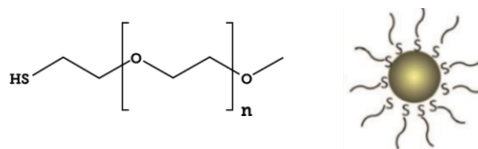


Figura 18: A sx la formula del metossi-polietilen glicole-tiolato (mPEG-SH), a dx AuNP con PEG-SH.

2.2.2 Substrati solidi con drop-casting

Substrato: Lo spot ($\sim 1 \text{ cm}^2$) è creato depositando l'inchiostro precedentemente preparato sul supporto solido costituito dal vetrino coperto di alluminio. L'uso della piastra riscaldante a circa $100 \text{ }^\circ\text{C}$ ha permesso al solvente di evaporare più velocemente, creando così il film (fig. 20). Si è dovuto però prestare attenzione a non creare l' indesiderato effetto "coffee ring"¹², dato che si utilizzavano AuNPs sferiche.^[11] I vetrini sono stati poi cotti in forno per 3 ore a $250 \text{ }^\circ\text{C}$. Sopra il film è stato quindi attaccato, con la colla verde, il pozzetto di plastica per contenere la soluzione di campione (fig. 19).

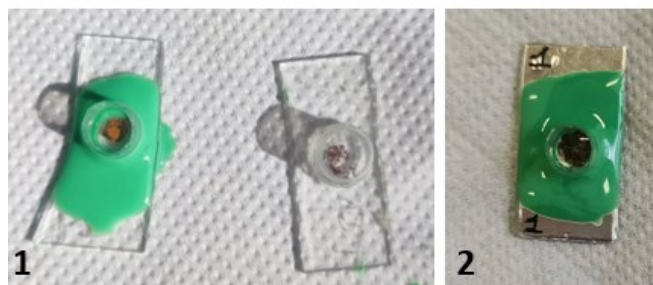


Figura 19: In 1 vetrini senza alluminio prima e dopo aver attaccato il pozzetto con la colla verde. In 2 substrato con soluzione di DNA in incubazione.

¹² Sospensioni colloidali o soluzioni di specie non volatili evaporando lasciano residui solidi simili ad anelli lungo la linea di contatto. Le ragioni fisiche per cui avviene sono il "pinning", l'inchiodamento, della linea di contatto appuntata e l'intensa evaporazione.

Campioni: I campioni sono stati omogenizzati portandoli tutti alla C di 1 ng/ μ L in 100 μ L utilizzando acqua milliQ come solvente (tabella 1). 50 μ L di questa soluzione sono stati prelevati e lasciati a sedimentare per un'ora (tempo di incubazione) nel pozzetto.

Spettri: Con l'utilizzo del microscopio, è stata scelta una zona in cui il film di AuNPs era compatto ed è stata eseguita una mappatura intorno all'origine (con 9 punti) con il laser a 785 nm, 10 s, 10x, \approx 10 mW. A lunghezze d'onda minori si è vista fluorescenza.

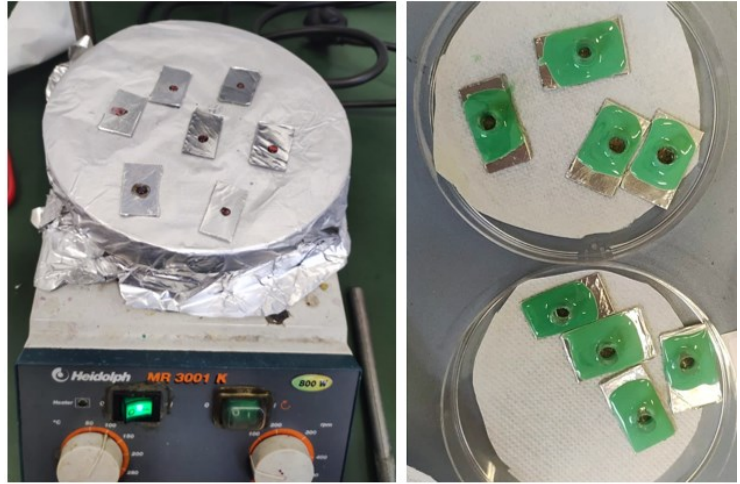


Figura 20: A sinistra vetrini ricoperti di alluminio dopo il drop-casting, a destra vetrini finiti con pozzetto.

2.2.3 Substrati colloidali

Campioni: I campioni sono stati portati tutti a C= 1 ng/ μ L in 100 μ L utilizzando acqua milliQ come solvente. A 50 μ L di questa soluzione sono stati aggiunti 150 μ L di AuNPs non funzionalizzate ablate in data 06/06/23. Si sono poi aggiunti 250 μ L di PBS a pH=7,4 (0,01M di fosfato, 0,138 M di NaCl e 0,0027 M di KCl) per aggregarle e il colore della soluzione è passato subito dal rosso al violaceo, per tendere al grigio dopo circa 10 minuti. Dopo aver lasciato i campioni in incubazione per mezz'ora, quando la soluzione era diventata grigio chiaro con riflessi viola-blu, è stata eseguita la misura.

Spettri: Sono stati registrati tra 190 e 3500 cm^{-1} con il laser a 785 nm, 60s, 10x, \approx 10 mW. Il bianco e il campione R sono stati registrati il primo giorno, i campioni da 2/11 il successivo, poi dal 12/19 il terzo e dal 20/28 il quarto giorno (tabella 1).

Analisi: I risultati sono stati analizzati con il metodo di regressione multivariata PLSR, utilizzando la k-fold-validation come tecnica di convalida incrociata per valutare le prestazioni del modello e scegliere graficamente il numero di variabili latenti ottimali.



Figura 21: Cuvetta da 2 mm usata per lo spettro UV-Vis a sx e per lo spettro Raman a dx.

3- RISULTATI E DISCUSSIONI

3.1 Substrati

La prima parte del lavoro si è concentrata sulla ricerca del miglior substrato in termini di precisione/tempo a disposizione per questa analisi. I substrati ottenuti con la stampante inkjet non sono stati utilizzati per le misure perché troppo dispendiosi in termini di tempo, anche se, come si vede da figura 22, permettevano un miglior controllo della dimensione (1 mm^2 contro 1 cm^2) dello spot e, di conseguenza, un minor spreco di inchiostro. I substrati solidi prodotti con il drop-casting sembravano essere un buon compromesso, ma gli spettri ottenuti con questi non erano molto chiari ed utilizzabili per l'analisi successiva.

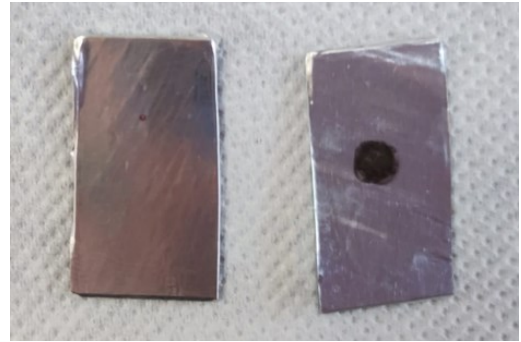


Figura 22: Confronto tra le dimensioni dello spot stampato con stampante inkjet (dx) e con drop-casting (sx).

L'utilizzo di substrati colloidali si è rivelato molto più facile e veloce, sia per quanto riguarda la preparazione del colloide che per il tempo di incubazione necessario, che si è potuto dimezzare dal momento che il substrato non doveva aderire a una superficie solida.

3.2 Concentrazione campioni

È stata scelta pari a $1\text{ ng}/\mu\text{L}$ in modo che con il campione R, utilizzato come riferimento in quanto aveva la concentrazione più elevata, fosse possibile eseguire diverse prove per trovare il substrato e il metodo migliori. Era inoltre importante avere una concentrazione omogenea tra tutti i campioni, che fosse però sufficientemente alta da poter essere rivelata. In aggiunta a questo, il valore così selezionato era anche abbastanza basso da essere sicuri dell'orientazione dei campioni di DNA sulla superficie metallica. È stato infatti studiato che, all'aumentare della concentrazione, i filamenti di DNA passano da paralleli alla superficie (concentrazione bassa) a perpendicolari (concentrazione più alta), con il potenziale rischio di denaturarsi (fig. 15, sotto).^[1,6] Questo influenza l'esito degli spettri rendendo l'analisi meno riproducibile.^[1]

3.3 Aggregazione delle AuNPs

1- NPs per il film dei substrati solidi

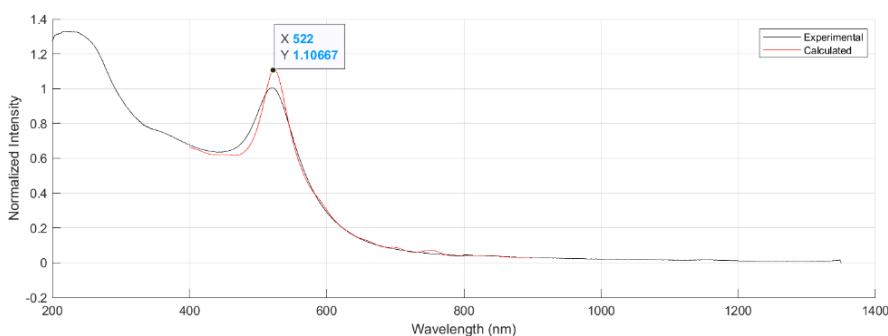


Figura 23: Spettro UV-Vis dell'inchiostro prodotto dopo la terza centrifuga. La C calcolata da Fitspettri è pari a $3.23 \cdot 10^{-10}\text{ mol/L}$.

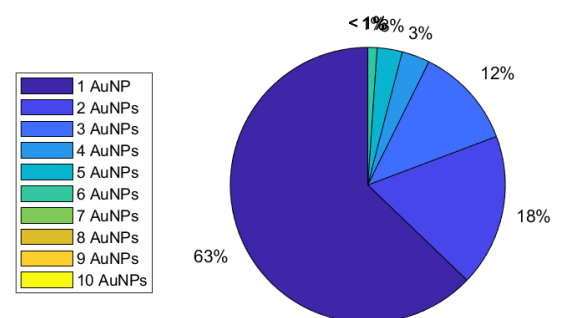


Figura 24: Aggregazione % in nanoclusters dell'inchiostro ottenuta con Fitspettri.

2- NPs per il substrato colloidale

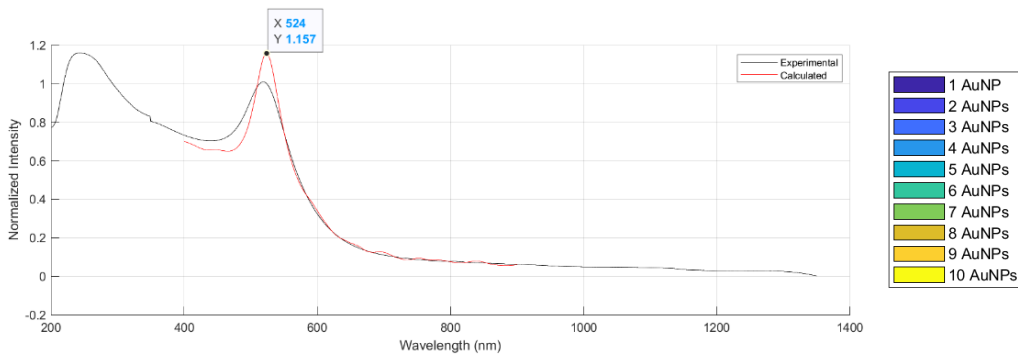


Figura 25: Spettro UV-Vis delle nanoparticelle utilizzate come colloide. La C calcolata da Fitspettri è pari a $1.30 \cdot 10^{-09}$ mol/L.

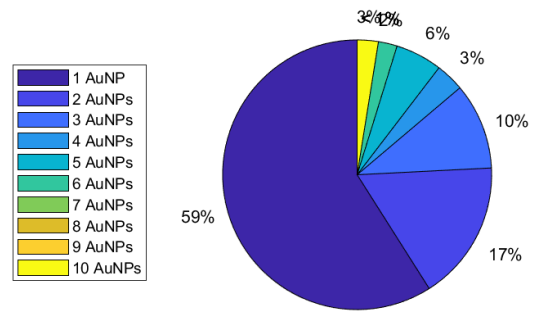


Figura 26: Aggregazione % in nanoclusters delle NPs utilizzate come colloide ottenuta con Fitspettri.

Lo spettro UV-Vis, in entrambi i casi (fig. 23 e 25), si presenta semplice, caratterizzato da un unico picco non allargato intorno al tipico valore di 520 nm, che concorda e giustifica la colorazione rossa della AuNPs sferiche con d tra i 5-20 nm. Questo significa che, come desiderato, la maggioranza delle nanoparticelle ($\sim 60\%$) non era aggregata, cosa che viene confermata anche dai grafici a torta. L'aggregazione, infatti, avrebbe avuto l'effetto di complicare lo spettro sviluppando gradualmente nuove bande, che avrebbero coperto la regione del visibile e del vicino infrarosso, e questo avrebbe quindi causato il passaggio a una colorazione marrone/grigio scuro. La concentrazione calcolata con il programma Matlab Fitspettri per la soluzione recuperata e concentrata dal surnattante è, come prevedibile, un ordine di grandezza inferiore rispetto all'altra ($3.23 \cdot 10^{-10}$ mol/L contro $1.30 \cdot 10^{-09}$ mol/L).^[8]

Le particelle utilizzate per l'inchiostro nella prima parte, per sfavorire l'aggregazione, erano prima state funzionalizzate con il PEG-SH.

Nel secondo caso sono state usate AuNPs senza alcun agente cappante e si è poi impiegato il PBS per aggregarle. Aggiungendo il PBS a una soluzione, con o senza il DNA, si ottenevano soluzioni di colore leggermente diverso, rispettivamente grigetta con riflessi blu una e più viola l'altra (fig. 27). Questo era una prova del fatto che il DNA si fosse attaccato alla superficie delle AuNPs, rendendone più difficoltosa l'aggregazione. Il colore più scuro indicava comunque che le NPs si erano aggregate.



Figura 27: Colorazione dopo l'aggiunta del PBS a una soluzione senza DNA (sx) e una con (dx).

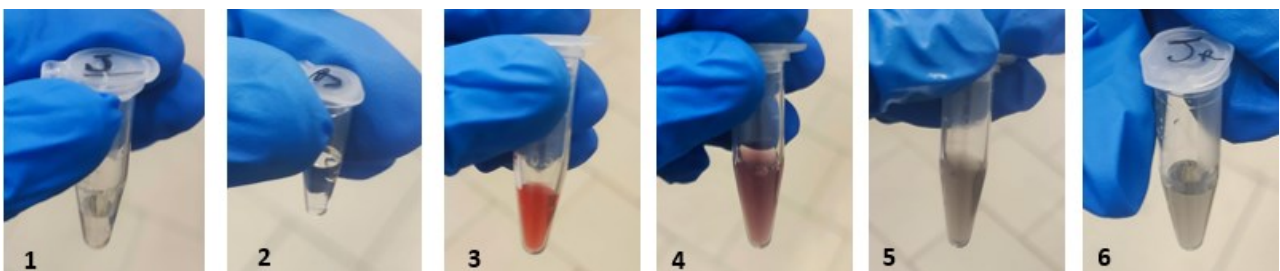


Figura 28: Evoluzione dei colori del campione J. In 1 la soluzione trasparente di campione appena prelevato dal frigo, in 2 campione diluito con acqua, in 3 la soluzione rossa-arancio dopo l'aggiunta dei 150 μ L di AuNPs alla soluzione con il campione, in 4 la soluzione violetta dopo l'aggiunta del PBS, in 5 la soluzione grigetta dopo circa 30 minuti di sedimentazione, in 6 la soluzione quasi trasparente dopo la misura dello spettro (dopo circa 1 ora).

3.4 ANALISI DEI DATI

3.4.1 Descrizione generale del modello PLS

Per mettere in relazione gli spettri SERS ottenuti e creare un modello di previsione dell'età, è stata utilizzata la regressione parziale dei minimi quadrati (PLS). Questo metodo è sempre più spesso utilizzato in chemometria per estrarre informazioni da una serie di spettri che includono molte variabili (le intensità ad ogni Raman Shift per il caso specifico, dati omogenei), spesso correlate tra di loro.^[12,13] PLS è infatti appropriato quando le variabili sono numerose, rumorose, incomplete e dove è molto probabile che vi sia multicollinearità (cioè, che non siano indipendenti tra loro come accade per gli spettri) o quando è necessario modellare simultaneamente diverse variabili di risposta Y (quando $Y=1$, cioè l'età del donatore di DNA in questo caso, il modello si chiama PLS1).^[14,15]

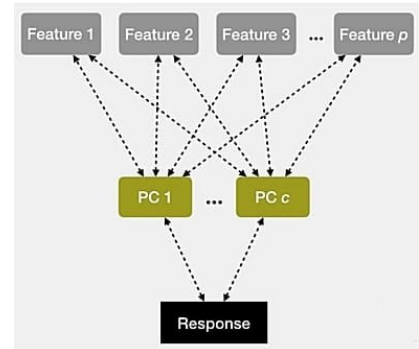


Figura 29: Schema di PCA/PLS

Analogamente al metodo PCA (Principal Component Analysis), anche in questa regressione vengono costruite nuove variabili non correlate, chiamate latenti (LV), attraverso una combinazione lineare delle variabili iniziali con dei coefficienti chiamati “pesi”, in modo tale che siano poi considerate solo quelle significative e, quindi, il loro numero risulti inferiore rispetto a quello delle variabili originali (fig. 29).^[14,15] Anche in questo modello la prima LV creata è quella più importante, ma, se nella PCA le componenti sono costruite per descrivere e contenere la maggior varianza dell'insieme X (insieme delle variabili di partenza), nel metodo PLS invece vengono scomposte simultaneamente sia la X che la Y e le componenti di X vengono scelte in modo da massimizzare la covarianza con le variabili risposta per spiegarne le variazioni il meglio possibile. Una variabile latente in questa regressione è rilevante solo se migliora la previsione di Y per nuovi dati. Questo passaggio generalizza la PCA, ma il modello è una generalizzazione anche della regressione lineare multipla (MLR). Infatti, i vettori latenti ottenuti da X (predittori) vengono poi utilizzati per prevedere Y, come in MLR.^[15] Generalmente, per scegliere il numero di variabili latenti ottimali da usare (per non andare in overfitting ed essere troppo aderenti ai dati utilizzati o spiegare troppo poco del modello) e per valutare le prestazioni della regressione si utilizza la convalida incrociata. La variante qui impiegata si chiama k-fold cross validation con $k=10$. È utilizzata per selezionare graficamente il numero di LV ottimali, che sono 4 in questa analisi, e poi con 4 LV viene stimata la prestazione del modello (qui, il coefficiente di determinazione R^2 , che indica quanto della varianza totale della variabile di risposta è spiegato dalla regressione). Con questa procedura, il set di dati viene suddiviso in k parti di dimensioni simili e in ogni iterazione k-1 gruppi sono utilizzati per l'addestramento, mentre il rimanente è utilizzato come set di convalida e sui suoi n elementi esclusi spesso si calcola l'indice di prestazione (e.g. R^2 o PRESS, cioè $\sum_{i=1}^n (Y_i - \tilde{Y}_{i,VL})^2$). Il processo è quindi ripetuto k volte (serve perciò quando, come qui, si hanno pochi dati a disposizione) e la prestazione può essere calcolata, per esempio, rispettivamente come la loro media o la somma.^[15]

3.4.2 Risultati ottenuti

Gli spettri ottenuti allo spettrofotometro Raman sono stati filtrati per l'eventuale presenza di raggi cosmici con il software proprio dello strumento. Il resto delle elaborazioni, ovvero correzione della linea di base e PLSR è stato eseguito in Matlab utilizzando funzioni built-in. Nei grafici che seguono vengono evidenziate in rosso le vibrazioni^[1,12,16] su cui si è focalizzata l'indagine vista la provata relazione tra metilazione ed età (vedi intro. 1.1 e tabella 2). Queste sono quelle che riguardano il gruppo $-CH_3$ che con l'età si dovrebbe legare a C nelle isole CpG. Fortunatamente, secondo alcune ricerche, dovrebbe essere valido il seguente ordine per l'intensità dei segnali SERS: $IA \approx IC \gg IG > IT$ e perciò, si dovrebbero riuscire a vedere abbastanza bene i segnali della base C di interesse.^[1]

Tabella 2: Picchi SERS del DNA legati alla metilazione.

Picco (cm ⁻¹)	1212	1247-1267	1298	1320-1330	1378	1407	1446	1528	1628
Assegn.	T: bending CH ₃	C,(G): stretch. anello	Def. dei CH ₃	C,G,A: def. CH-CH ₃ / Vibr. anello	T(+A): bending CH ₃	T: def. C5-CH ₃	C,G,A: def. CH-CH ₃ / Respiro dell'anello	C: Vibr./respiro anello	C: stretch. C=O (vicino bend. H ₂ O)

Da studi precedenti è noto che la maggior parte delle differenze spettrali sono osservate negli intervalli 600-900 e 1200-1400 cm⁻¹, poiché lì si concentrano le zone delle impronte digitali dei nucleosidi.^[1] Inoltre, sotto i 300 cm⁻¹ e tra i 3200 e i 3400 cm⁻¹ (stretching O-H), i segnali dell'acqua sono rilevanti^[17] perciò si è scelto di lavorare con il secondo range, analizzando i dati solo da 900 a 2000 cm⁻¹.

Come già specificato, gli spettri raccolti ed in seguito presentati sono stati ottenuti utilizzando come substrato SERS una soluzione colloidale di AuNPs. Il grafico 30 mostra li campione R (volontario di 56 anni) di tabella 1. Si notano alcuni dei picchi più caratteristici riportati in letteratura (tabella 2 e 3) per il DNA ed in particolare, si vedono bene i picchi dei fosfati e alcuni dei picchi correlati alla metilazione. Lo spettro del bianco (solo acqua milliQ, con AuNPs e PBS), rappresentato in fig. 30, dovrebbe idealmente presentare solo i deboli segnali dell'acqua nel range di interesse, ovvero il bending intorno a 1650 cm⁻¹.^[17] Nello spettro di fig. 31 non è in effetti visibile il picco caratteristico del fosfato intorno a 1100 cm⁻¹, ma purtroppo sono presenti segnali in corrispondenza dei punti di interesse a 1378 e 1528 cm⁻¹.

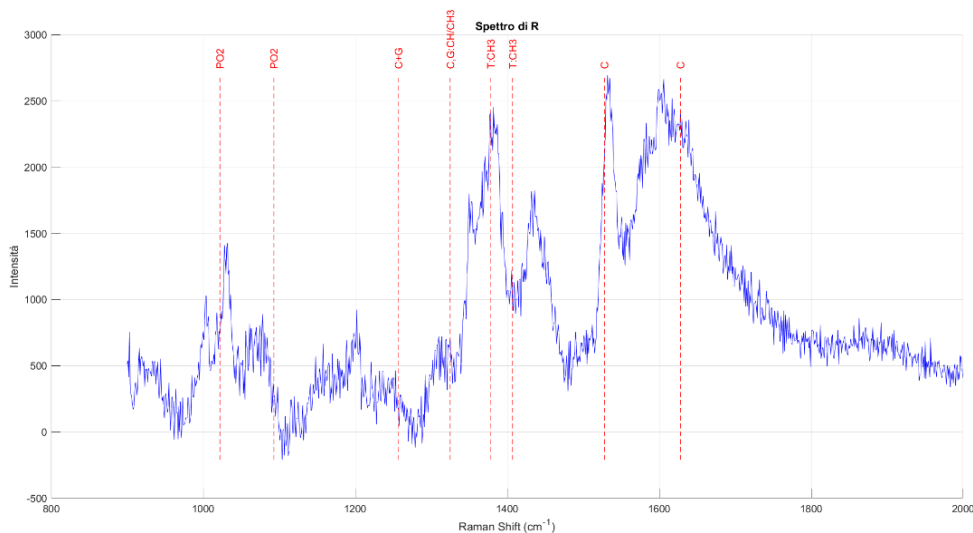


Figura 30: Spettro SERS del campione R.

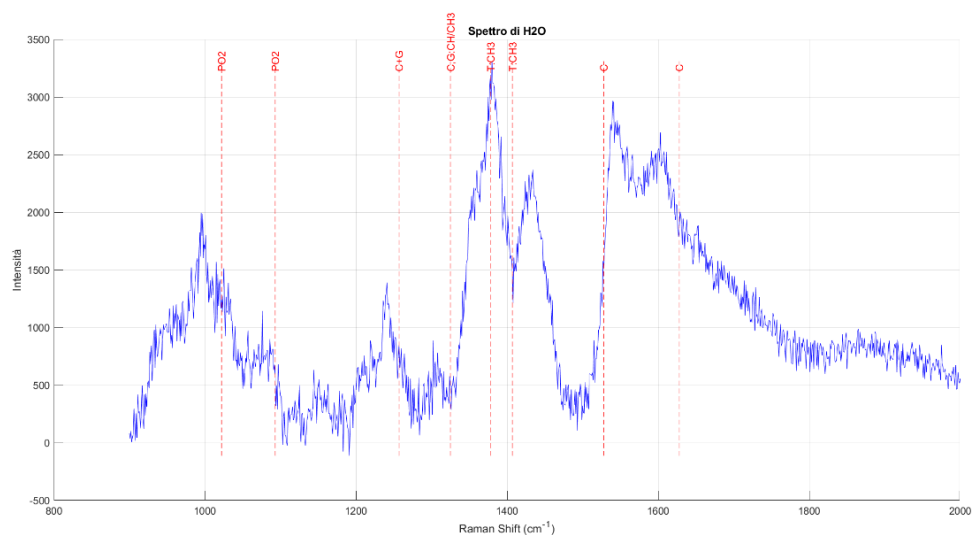


Figura 31: Spettro SERS del bianco (acqua milliQ).

Tabella 3: Altri picchi SERS del DNA .

Picco (cm ⁻¹)	Assegnazione [1,12,16]
919	Stretching C-C del ribosio
958	Stretching ribosio-fosfato
1010-1035	Backbone/ Str. simm. PO ₂ ⁻
1090-1096	Stretch. simm PO ₂ ⁻
1178	T
1215	Stretching assim. PO ₂ ⁻
1354	G
1415-1425	Backbone
1485-1490	G(+A): vibr. dell'anello, stretch. C=N
1506	G(+A): vibrazione dell'anello
1509	A
1575	A, G: vibrazione dell'anello
1650-1655	T, C: stretching C=O. (vicino al bending di H ₂ O)
1684	T: stretching C=O
1721	G: stretching C=O

L'intero dataset viene presentato in fig 32, ponendo i Raman Shift e l'età del donatore in asse x e y, mentre le intensità di ciascuno spettro sono espresse con una scala cromatica secondo cui i colori freddi come il blu indicano i valori più bassi, mentre quelli caldi come il giallo i più intensi.

Si nota che le regioni dello spettro in cui si hanno i picchi più intensi sono comprese tra 1340 e 1390 cm^{-1} , tra 1420 e 1460 cm^{-1} e tra 1530 e 1640 cm^{-1} per tutte le età e soprattutto nei primi due range spettrali sono contenuti alcuni dei segnali più rilevanti per l'indagine. I campioni che li hanno riportato maggiori intensità SERS sono quelli corrispondenti alla metà della prima fascia d'età (~5 anni), della terza (~25 anni) e la sesta (~60 anni).

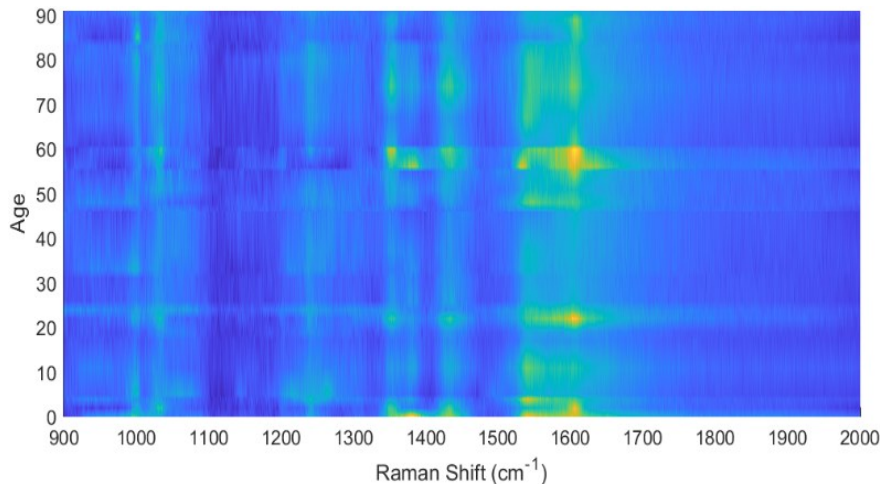


Figura 32: Grafico 3D degli spettri in funzione del Raman Shift (asse x), dell'età (asse y) e delle intensità spettrali (asse z).

L'intero dataset viene quindi sottoposto al modello PLSR. Figura 33 presenta le prime 4 variabili latenti che il modello ha individuato come le più significative e correlate all'età del donatore. Questi profili sono utili per visualizzare i segnali, e quindi i relativi Raman Shift, più informativi per ogni LV. I segnali significativi sono quelli con pesi maggiori (positivi o negativi) perché sono quelli che hanno una maggiore relazione con la variazione della risposta età (cioè, sembrano esserne le maggiori responsabili).^[14] Il profilo per la prima componente LV1 mostra che i segnali attribuibili al backbone e ai fosfati non sembrano apparentemente legati ai cambiamenti nella risposta età. Questo concorda anche con il grafico 32 in cui si osserva la stessa intensità (colore) nei loro rispettivi range caratteristici lungo tutta la serie. Questa osservazione può essere altresì spiegata con il fatto che il backbone del DNA non dovrebbe subire modifiche sostanziali con l'invecchiamento. Le variazioni più rilevanti attribuibili al CH_3 sembrano essere quelle che cadono a circa 1257, 1378, 1407 e 1528 cm^{-1} (peso maggiore), mentre gli altri riferimenti di tabella 2 hanno pesi trascurabili (perciò non riportate in rosso). I picchi a 1378 e a 1528 cm^{-1} , però, perdono un po' di importanza essendo presenti anche nello spettro del bianco. Alcuni segnali presenti nella tabella 3, in particolare quelli intorno a 950, 1180 e 1350 cm^{-1} , hanno pesi abbastanza significativi anche se non sono strettamente correlati alla metilazione¹³. Non si riscontrano evidenti picchi significativi non presenti nelle tabelle, perciò sembra effettivamente che quelli collegati alla metilazione (soprattutto quello a 1257 e a 1407 cm^{-1}) siano quelli più rilevanti da considerare. LV2 presenta un andamento pressoché analogo a LV1, mentre LV3 e LV4 paiono gradualmente meno significative.

¹³ La T contiene un CH_3 e la G si trova nelle isole CpG metilate perciò è plausibile che i loro picchi siano rilevanti. Il picco a circa 1620 cm^{-1} non è particolarmente significativo perché prossimo al bending dell'acqua.

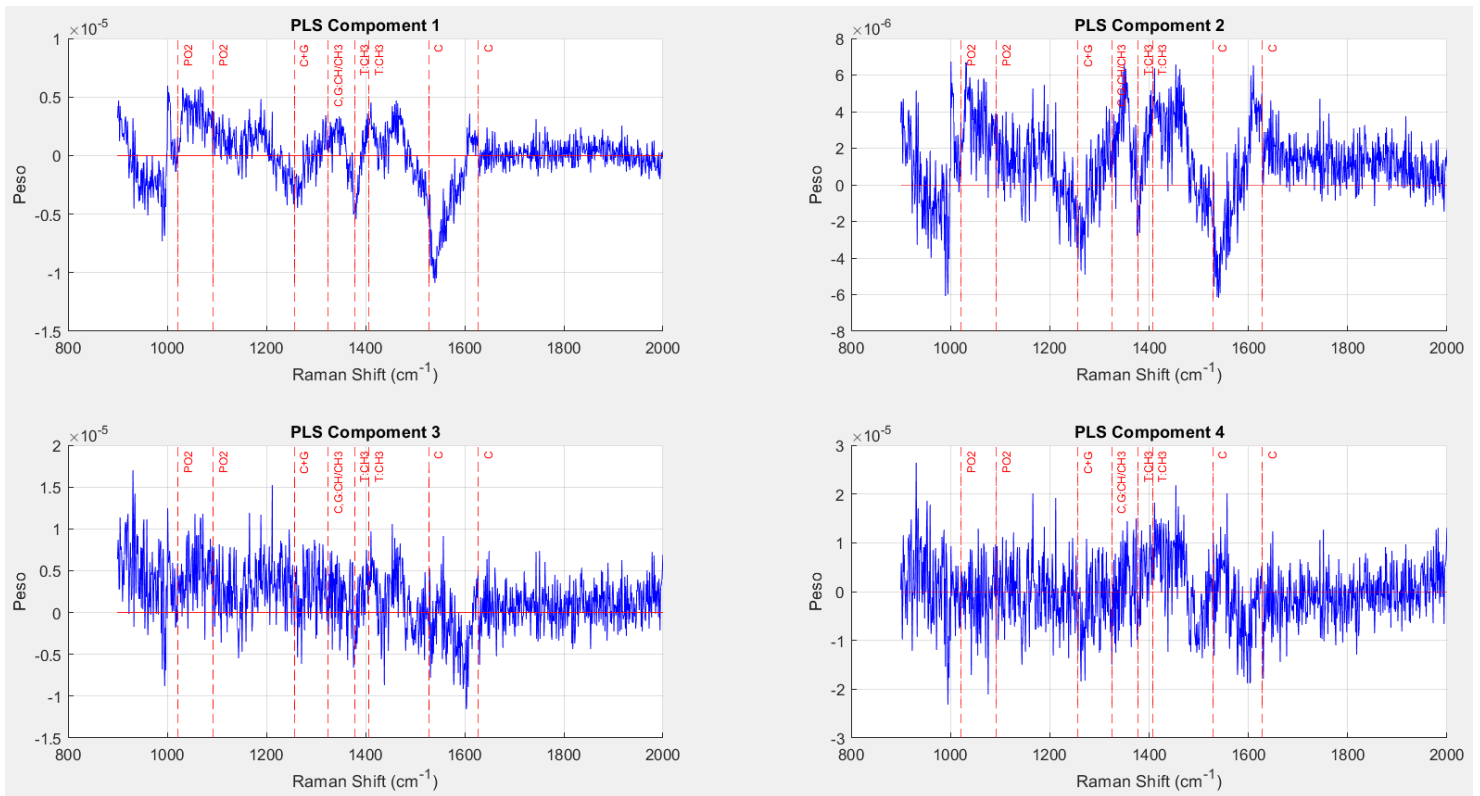


Figura 33: Grafico dei pesi delle componenti PLSR.

La scelta del numero di variabili latenti da usare per la costruzione di un modello predittivo tramite PLSR è di fondamentale importanza, come già detto. In questa regressione, la quantità di varianza spiegata da una variabile latente indica la sua importanza nella previsione di Y.^[14] Figura 34 illustra la varianza cumulativa espressa al crescere del numero di LV considerate, ovvero quanto l'insieme di questi è adeguato al descrivere la maggior parte delle informazioni contenute in tutto il dataset. In questo caso si evince che l'utilizzo di 4 LV comporta l'espressione cumulativa dell'83%.

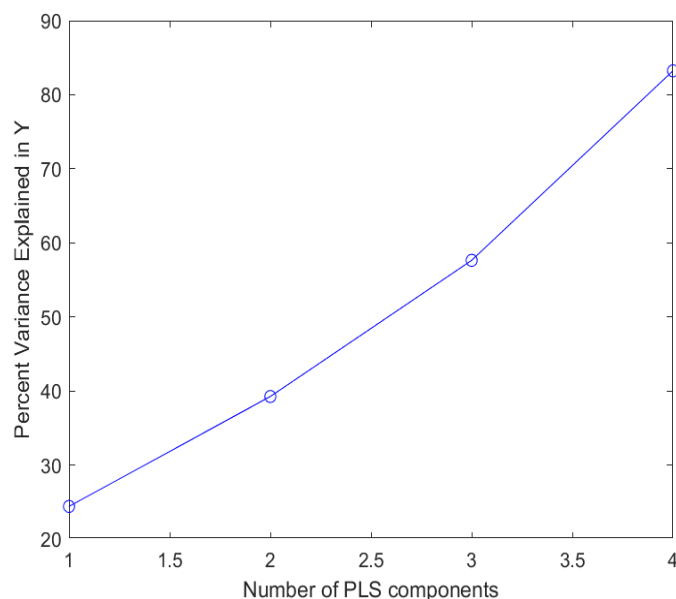


Figura 34: Grafico della varianza cumulativa delle componenti.

Infine, in figura 35 si riportano i risultati del modello predittivo, ovvero il cuore dell'analisi. Qui si possono confrontare i valori previsti dal modello PLSR1 con i valori noti dell'età dei donatori per i singoli campioni. Si può osservare che i punti si distribuiscono abbastanza bene attorno alla retta della correlazione ideale. R^2 è stato scelto come indice di prestazione in quanto è risultato il più informativo, accurato ed efficace nella valutazione dei modelli statistici.^[18] Questo è stato qui calcolato utilizzando quattro variabili latenti per la regressione, ed essendo superiore allo 0.8, può essere utilizzato per fare previsioni.^[3,18] Si deduce quindi che il modello, per ora, è in grado di prevedere solo la fascia d'età, cioè se il campione di DNA appartiene ad un bambino, un adolescente, un adulto o un anziano. Ad esempio, un ragazzo di 17 anni viene confuso per uno di 28, un uomo di 55 anni lo considera di 40, mentre uno di 85 anni lo stima di 77. Il modello sembra adattarsi abbastanza bene nella zona centrale e questo potrebbe essere comunque utile per future ricerche perché questa fascia (i primi e gli ultimi anni di vita sono stati ampiamente studiati) risulta attualmente quella più complessa da studiare dato che è quella in cui si ha la maggiore variabilità nei modelli epigenetici e in cui iniziano ad insorgere numerose patologie.^[4]

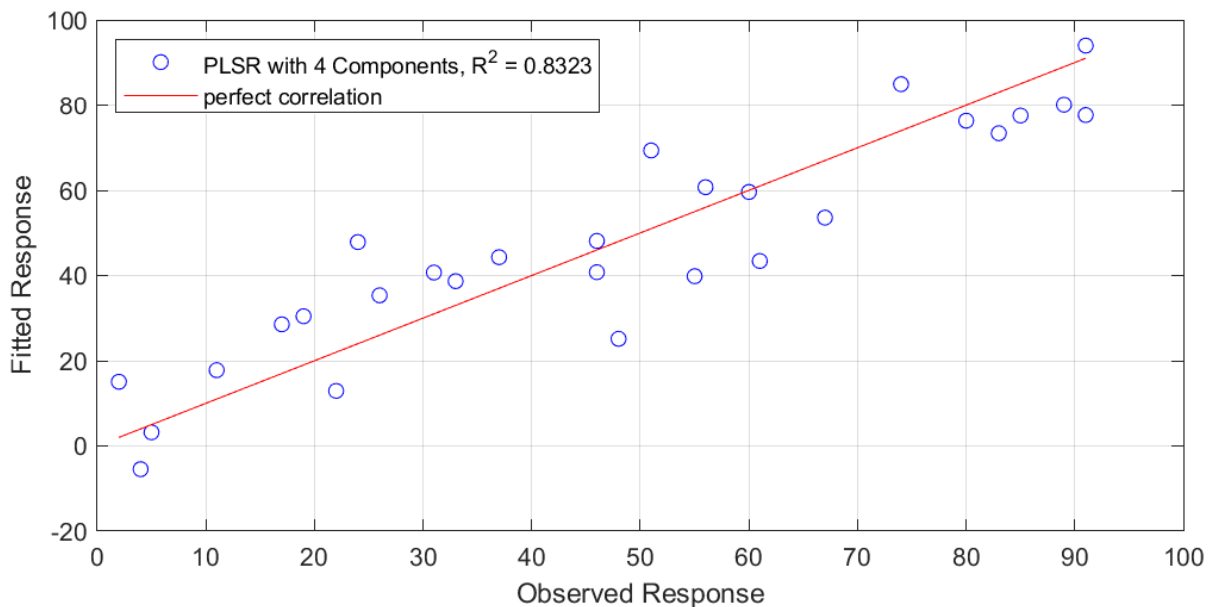


Figura 35: Grafico dei risultati yfitPLS.

3.5 Implementazioni per il proseguo del progetto

Per quanto i risultati ottenuti e presentati nelle sezioni precedenti siano interessanti e promettenti, si individuano alcune traiettorie di sviluppo per ulteriori implementazioni e miglioramenti.

1) Indagine: I soggetti volontari coinvolti per l'estrazione dei campioni di DNA sono discriminati riguardo la sola età. Da quanto riportato in letteratura e discusso nei capitoli precedenti, esistono numerosi altri fattori che influenzano il grado di metilazione e sarebbe quindi auspicabile implementare una scheda di anamnesi maggiormente approfondita. Il rischio altrimenti è di non essere in grado di distinguere effetti intrinseci e specifici dovuti all'invecchiamento. Ad esempio, sono importanti le informazioni sulle malattie diagnostiche (e.g. il diabete di tipo 2 o il tumore), sullo stile di vita (e.g. se svolge attività fisica), sulle abitudini alimentari (che possono incidere nel metabolismo del folato), sull'indice di massa corporea (e.g. l'obesità è associata alla metilazione), sulla qualità dell'aria respirata e sulle condizioni e abitudini di vita.^[2] Tra questi ultimi, quello da

tenere più in considerazione, perché strettamente correlato a livelli ridotti di metilazione, è l'esposizione al fumo attivo (fig. 36). Infatti, tra adulti fumatori e non fumatori sono state identificate differenze statisticamente significative in 187 siti CpG (dal DNA nel sangue intero), che possono fungere da biomarcatori dell'esposizione (e.g. in AHRR soppressore tumorale e in F2RL3 correlato al rischio di malattie cardiovascolari, al cancro ai polmoni e alla mortalità).^[19] In modo analogo a quanto ottenuto per gli adulti, anche gli EWAS (Epigenome-Wide Association Study) che analizzavano il ruolo del fumo materno nei neonati, hanno identificato siti CpG meno metilati, anche se con un effetto meno pronunciato, in diversi geni correlati al fumo (in AHRR ma non F2RL3). Guardando i livelli intermedi degli ex-fumatori, si è osservato che il processo è reversibile, ma per un recupero completo possono essere necessari fino a 20 anni.^[19] Per ottenere modelli affidabili però, oltre ad indagini trasversali, sarebbe necessario svolgere anche le più costose analisi longitudinali^[4] ed eseguire più repliche dello stesso spettro.

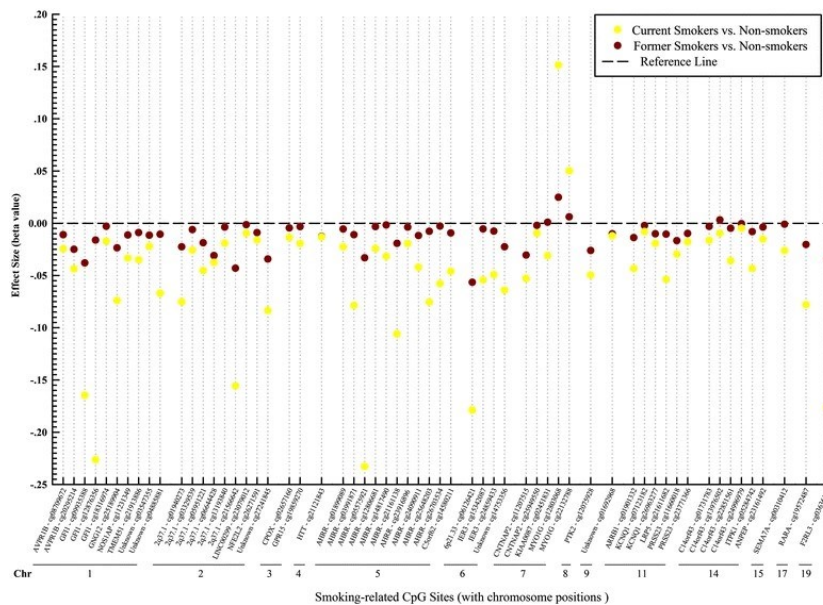


Figura 36: Relazione inversa per gli adulti tra il fumo e la metilazione nei siti CpG specifici ("beta values" negativi).

2) Prelievo: I campioni sono stati prelevati dalla saliva. La maggior parte delle informazioni che abbiamo da letteratura a proposito della metilazione dei mammiferi deriva, però, da studi effettuati sul sangue ed è inoltre ben noto che i profili di metilazione del DNA sono tessuto-specifici. Tra i vantaggi nell'utilizzare il DNA proveniente dal sangue si trovano sia la quantità del DNA, che la sua qualità, dal momento che questo contiene meno interferenti (e.g. DNA batterico o inibitori della PCR nella saliva) di quello salivare.^[20] Infatti, la quantità media di DNA raccolto con una provetta per prelievo venoso da 15 mL è di $277 \pm 214 \mu\text{g}$ contro i $92 \pm 74 \mu\text{g}$ ottenuti con il tampone salivare (4 mL). Alcuni studi sul DNA salivare, inoltre, hanno anche dimostrato che solo circa il 50% del DNA totale prelevato è di origine umana, mentre il restante è batterico. Lo svantaggio nell'utilizzare il sangue intero, però, risiede nel metodo di prelievo, la flebotomia, che è più dolorosa per i volontari e più costosa del tampone.^[20]

3) Substrati SERS: L'interazione tra il doppio filamento di DNA e la superficie metallica è, teoricamente, repulsiva perché entrambe sono cariche negativamente. Aggiungendo alla soluzione specie con cariche positive (e.g. spermina) si favorirebbe sia l'aggregazione delle nanoparticelle^[1,6] che l'adsorbimento del DNA, facilitato eventualmente anche dall'uso di filamenti tiolati (se disponibili).^[6] Per produrre spettri altamente riproducibili e controllare l'aggregazione dei colloidi,

mantenendo una distanza fissa ottimale di 0.9 nm, alcuni studi suggeriscono di aggiungere alla soluzione anche il CB[n] (cucurbit), con n preferenzialmente pari a 5 o 7, anche se questo potrebbe interferire con l'adesione del DNA.^[21] Un'altra accortezza proposta per essere più sicuri dei picchi, consisterebbe nel costruire i pozzetti con polistirene nero e non con quello reticolato, che potrebbe perdere contaminanti e produrre picchi a circa 1000, 1030 e 1600 cm⁻¹.^[21] Inoltre, per eliminare l'effetto "coffee ring" con il drop casting, senza cambiare tipo di metallo o solvente, si possono utilizzare (almeno in maggior parte) AuNPs anisotrope ellissoidali ($\alpha > 1.2$). È stato infatti riportato che queste favoriscano una deposizione quasi uniforme durante l'evaporazione grazie alle forti interazioni interparticellari a lungo raggio presenti all'interfaccia.^[22]

4- CONCLUSIONE

Il presente progetto di tesi ha avuto l'obiettivo di individuare una metodologia per estrarre l'informazione dell'età del donatore da spettri Raman ottenuti mediante la tecnica SERS. Per raggiungere lo scopo è stata usata una soluzione di nanoparticelle di Au come substrato colloidale e regressione tramite il modello del PLS. Gli spettri ottenuti hanno effettivamente riscontrato l'importanza delle vibrazioni della base C e dei CH₃, coerenti con numerosi studi collegano questi segnali con l'invecchiamento e con numerose patologie. Analizzando i dati attraverso il modello PLSR con quattro componenti principali e con la k-fold-cross validation, si è ottenuto un modello per prevedere l'età di un campione di DNA che però, per ora, si è dimostrato efficace unicamente per stimare a che range d'età (bambini, ragazzi, adulti o anziani) il donatore appartenga. Per migliorare il modello e renderlo più attendibile, bisognerebbe includere nell'indagine una serie di informazioni importanti sugli altri fattori che possono influenzare il grado di metilazione, oltre che sviluppare materiali e metodologie più sofisticate per i substrati SERS.

RINGRAZIAMENTI

Vorrei ringraziare il mio relatore, il Dr. Lucio Litti, per avermi proposto questo interessante progetto, per la dedizione, la pazienza e il tempo dedicatomi. Un grazie anche a tutte le persone del gruppo NOL.

Un ringraziamento al Prof. Eugenio Palladio dell'Università di Torino e alla Prof. Giorgia Sciutto dell'Università di Bologna per la collaborazione al progetto.

Un grazie speciale a mia sorella e mia mamma, senza le quali non sarei riuscita ad arrivare alla fin qui, che mi hanno sostenuto in tutti i momenti difficili di questo lungo percorso personale e universitario.

5- BIBLIOGRAFIA

- [1]- Safar W.; Azziz A.; Edely M. and Lamy de la Chapell M., *Review Conventional Raman, SERS and TERS Studies of DNA Compounds, Chemosensors*, **2023**, 11(7), 399, 1-49.
- [2]- Jung M.; P. Pfeifer G., *Aging and DNA methylation, BMC Biology*, **2015**,1-8.
- [3]- Goela N.; Karira P.; Garg V. K.; *Role of DNA methylation in human age prediction, Elsevier*, **2017**, 166, 33-41.
- [4]- J. Jones M; J. Goodman S.; S. Kobor M., *Aging Cell, Anatomical Society and John Wiley & Sons Ltd*, **2015**, 924-932.
- [5]- R. Ferraro J.; Nakamoto K.; W. Brown C., *Introduction to Raman Spectroscopy*, Elsevier Science, San Diego, 1994.
- [6]- Carnerero J. M.; Jimenez-Ruiz A.; Castillo P. M.; Prado-Gotor R., *Covalent and Non-Covalent DNA–Gold-Nanoparticle Interactions: New Avenues of Research, ChemPhysChem*, **2017**, 18, 17-33
- [7]- *Review: Present and Future of Surface-Enhanced Raman Scattering, ACS Nano*, **2020**, 14, 1, 28–117.
- [8]- Litti L; Meneghetti M., *Predictions on the SERS enhancement factor of gold nanosphere aggregate samples, Phys. Chem. Chem. Phys.*, **2019**, 21, 15515-15522.
- [9]- Campion A; Kambhampati P., *Surface-enhanced Raman scattering, Chem. Society Reviews*, **1998**, 27, 241-250.
- [10]- Murray W. A; L. Barnes W, *Plasmonic Materials, Adv. Mater.* **2007**, 19, 3771–3782.
- [11]- Rodríguez-Hernández J.; Bormashenko E, *Breath Figures*, Springer Nature Switzerland AG, Gewerbestrasse, 2020
- [12]- *Plasmonic hot spots reveal local conformational transitions induced by DNA double-strand breaks, Scientific Reports*, **2022**, 12.
- [13]- N. Ur Rehman; A. Al-Harrasi; R. Boqué; F. Mabood; M. Al-Broumi; J. Hussain; S. Alameri, *FT-NIRS Coupled with PLS Regression as a Complement to HPLC Routine Analysis of Caffeine in Tea Samples, Foods* **2020**, 9, 827.
- [14]- Abdi H, *Partial least squares regression and projection on latent structure regression (PLS Regression), Focus*, **2010**, 2, 97-106.
- [15]- Wold S.; Sjostrom M.; Eriksson L; *PLS-regression: a basic tool of chemometrics, Elsevier*, **2001**, 58, 109-130.
- [16]- K. A. Okotrub, N. V. Surovtsev, V. F. Semeshin, L. V. Omelyanchuk, *Raman Spectroscopy for DNA Quantification in Cell Nucleus, Cytometry Part A*, **2015**, 87A, 68-73.
- [17]- Pattenaude S. R.; Streacker L. M.; Ben-Amoz D, *Temperature and polarization dependent Raman spectra of liquid H₂O and D₂O, J Raman Spectrosc.*, **2018**, 49, 1860-1866.
- [18]- Chicco D., Warrens M. J.; Jurman G., *The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Comput*, **2021**.
- [19]- Gao X.; Jia M.; Zhang Y.; Breitling L. P; Brenner H, *DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clinical Epigenetics*, **2015**, 7, 1-10.
- [20]- A. Philibert R.; Zadorozhnyaya O.; R. H. Beach S.; H. Brody G., *A Comparison of the Genotyping Results Using DNA Obtained from Blood and Saliva, Psychiatr Genet.*, **2008**; 18(6), 275-281.
- [21]- Grys D. B.; Chikkaraddy R.; Kamp M.; Scherman O. A.; Baumberg J. J., de Nijs B., *Eliminating irreproducibility in SERS substrates, J Raman Spectrosc.* **2021**, 52, 412-419.
- [22]- J. Yunker P.; Still T.; Lohr M. A.; A. G. Yodh, *Suppression of the coffee-ring effect by shape-dependent capillary interactions, Nature*, **2011**, 476, 1308.