PhD thesis

Detection of microorganisms in clinical samples *via* surface-enhanced Raman spectroscopy

Sylwia Magdalena Berus MSc

Supervisor: prof. Agnieszka Kamińska

Doctoral dissertation prepared within International Doctoral Studies
of the Institute of Physical Chemistry of the Polish Academy of Sciences

Kasprzaka 44/52

01-224 Warsaw, Poland

Warsaw, 2025

Dedicated to …

… my dearest Mom

Acknowledgement

I would like to sincerely thank my supervisor, prof. Agnieszka Kamińska, for her invaluable support in every aspect of my work. Thank you for the wonderful work atmosphere, patience, and profound knowledge.

I would like to express my gratitude to prof. Jacek Waluk for the collaboration, knowledge, and professional assistance in editing my scientific articles.

My warmest thanks to everyone in the group I:

Tomasz Szymborski, Aneta Kowalska, Ariadna Nowicka, Marta Czaplicka, Evelin Witkowska, and Krzysztof Niciński. I appreciate every moment spent with you and the inspiring conversations that filled me with new energy.

I would like to express my special thankfulness to my coworkers with whom I had the pleasure of working

Monika Adamczyk Popławska, Dorota Korsak, Beata Młynarczyk-Bonikowska, Grażyna Przedpełska, Izabela Dróżdż, and Maciej Borowiec

Thank you for your kindness, valuable time, and commitment to research. I am very pleased that by combining our knowledge of chemistry, biology, and medicine, we were able to make a joint contribution to science.

Especially, I would like to thank my Mother for her understanding, patience, endless support, and words of encouragement. I am incredibly grateful for the opportunity to study and grow that I have been given.

And all my beloved ones for having faith in me and my abilities.

This work was financed by

Foundation for Polish Science under grant TEAM TECH/2017-4/23
(POIR.04.04.00-00-4210/17-00)

National Science Centre, Poland, under grant PRELUDIUM-20
2021/41/N/ST4/02164

Abstract

The doctoral dissertation presents a series of studies that aim to find a spectral image of selected disease entities using surface-enhanced Raman spectroscopy (SERS) and chemometric methods. SERS analyses performed in a label-free variant provided information about spectral changes caused by a given type of infection directly from the recorded SERS signal of a clinical sample. The reagentless procedure that does not require additional modification of the analyte can significantly reduce the risk of errors and the work and time involved.

As part of the research, one established the procedure for preparing clinical materials and the SERS measurement conditions using pre-optimised plasmonic nanostructures. An important element of the work was the comprehensive spectral characterisation of the control and experimental groups in relation to the considered disease entity. The assistance of chemometric methods provided insight into the correlation between data and identification of marker bands having diagnostic significance. Calibration models were created based on the properly defined classes, and their performance was tested at the stage of external validation simulating actual diagnostic conditions.

The first part of the research was devoted to the SERS analysis of bacteria from *Neisseria* spp., and urethral discharge in the context of gonorrhea recognition. The obtained data constituted a catalog of SERS spectra of microorganisms (reference data). Then, based on it, the unknown strain cultured from the clinical sample was correctly identified as *Neisseria gonorrhoeae* using its unique spectral image (the so-called 'fingerprint'). In the analysis of the urethral discharge, the experimental group related to the control one, exhibited different spectral character in the entire range. The appearance of new bands (especially the one at 724 cm^{-1}) indicated the certain contribution of pathogens in SERS signal formation. These differences are significant for identifying infections directly from SERS spectra of clinical samples. The chemometric analysis also indicated high accuracy in classifying external samples, *i.e.* 89% for SIMCA and 100% for PLS-DA.

The next section concerns the spectral characterisation of *Candida* spp. and bacteria of particular importance for the development of vaginal infections. The SERS spectra of microorganisms present the contribution of individual components, *e.g.* chitin, glucagon, and mannoproteins for *Candida* spp. and lipids, phospholipids, and proteins for bacteria.

The spectral variability was particularly observed for the clinical samples taken from women with abnormalities within microflora and biochemistry of the swab. The comprehensive analysis resulted in establishing the control group that provides an excellent basis for the diagnosis of VVC infection from the SERS spectra of the analysed clinical sample. The PLSR method visualized the clear differences between the two considered classes (control and VVC). It also revealed marker bands identified as those belonging to microorganisms (727, 1003, 1125, 1258, 1331, 1455, and 1670 cm^{-1}) and components of the swab, *i.e.* amino acids, proteins, saccharides, lactic acid (669, 891, and 1047 cm^{-1}).

The third section of the study concerns the analysis of saliva and nasopharyngeal swabs in terms of COVID-19 diagnosis. The chemometric methods extracted differential spectral information such as: for CoV(+) saliva – the elevated levels of methionine, nucleic acids, proteins (*e.g.*, ferritin) and immunoglobulins and for CoV(+) nasopharyngeal swabs – neopterin. Among all the tested methods, SVMC provided the best performance in the classification of external samples, achieving 90 % accuracy and 100 % sensitivity for saliva and 75 % accuracy and 88 % sensitivity for nasopharyngeal swabs.

The results presented in the doctoral dissertation indicated the competitiveness of the SERS technique compared to those currently accepted methods. It is also worth noting that the research was performed using a portable Raman spectrometer, which is a great advantage for point-of-care diagnostics.

Abstrakt (polish version)

Niniejsza rozprawa doktorska przedstawia serię badań przeprowadzonych w celu znalezienia spektralnego obrazu wybranych jednostek chorobowych z wykorzystaniem powierzchniowo wzmocnionej spektroskopii Ramana (*ang.* Surface-enhanced Raman spectroscopy – SERS) oraz metod chemometrycznych. Analiza SERS została wykonana w wariancie bezznacznikowym (tzw. label-free) w którym informacje o zmianach spektralnych wywoływanych danym typem infekcji pochodzą wprost z zarejestrowanego sygnału SERS próbki klinicznej. Procedura oparta na braku konieczności użycia odczynników oraz dodatkowej modyfikacji analitu pozwala zmniejszyć ryzyko popełnianych błędów oraz ograniczyć nakład pracy i czasu.

W ramach przeprowadzonych badań opracowano procedurę przygotowania próbek klinicznych oraz ustalono warunki pomiarowe analiz SERS z wykorzystaniem uprzednio zoptymalizowanych nanostruktur plazmonicznych. Istotnym elementem pracy była również kompleksowa charakterystyka spektralna grupy kontrolnej i eksperymentalnej pod kątem rozważanej jednostki chorobowej. Analiza otrzymanych widm z wykorzystaniem metod chemometrycznych zapewniła wgląd w korelację między danymi oraz identyfikację pasm markerowych mających znaczenie diagnostyczne. Na podstawie odpowiednio zdefiniowanych klas stworzono również modele kalibracyjne, których wydajność przetestowano na etapie walidacji zewnętrznej symulującej rzeczywiste warunki diagnostyki.

Pierwszą część badań poświęcono analizie SERS bakterii z rodzaju *Neisseria* spp., oraz wymazom z cewki moczowej w kierunku rozpoznawania chorób wenerycznych. Uzyskane dane stanowiły katalog widm SERS mikroorganizmów (dane referencyjne). Następnie na tej podstawie wyhodowany z próbki klinicznej szczep został poprawnie zidentyfikowany jako *Neisseria gonorrhoeae* wykorzystując jego unikatowy obraz spektralny (tzw. „odcisk palca”). W analizie wymazów z cewki moczowej grupa badana wykazała odmienny charakter spektralny względem grupy kontrolnej w całym analizowanym zakresie. Pojawienie się nowych pasm (szczególnie 724 cm^{-1}) wskazało na zdecydowany udział patogenów w tworzenie sygnału SERS. Stanowi to podstawę do rozpoznawania infekcji wprost z widm SERS próbek klinicznych. Dla utworzonych modeli kalibracyjnych otrzymano wysoką dokładność w klasyfikacji próbek zewnętrznych tj.: 89 % dla SIMCA oraz 100 % dla PLS-DA.

Kolejny rozdział pracy dotyczy spektralnej charakterystyki grzybów z rodzaju *Candida* spp., a także bakterii mających szczególne znaczenie dla rozwoju infekcji intymnych. Zarejestrowane sygnały SERS mikroorganizmów prezentują udział poszczególnych składników obecnych w ich strukturze np.: chityna, glukagon oraz mannoproteiny dla *Candida* spp. oraz lipidy, fosfolipidy, proteiny dla bakterii. Analizie SERS zostały również poddane próbki kliniczne dla których zaobserwowano zmienność spektralną powstałą na skutek zaburzeń w obrębie mikroflory i biochemii wymazu. Na tej podstawie ustalono grupę kontrolną stanowiącą doskonałą podstawę dla rozpoznawania infekcji VVC z widm SERS analizowanej próbki. Metoda PLSR wskazała na istnienie wyraźnych różnić między dwiema analizowanymi klasami. Pozwoliła również ustalić pasma markerowe, które zidentyfikowano jako te należące do mikroorganizmów (727, 1003, 1125, 1258, 1331, 1455 oraz 1670 cm^{-1}) oraz składników wymazów tj.: aminokwasy, białka, sacharydy, kwas mlekowy (669, 891 oraz 1047 cm^{-1}).

Trzecia części badań dotyczy analizy śliny i wymazów z nosogardzieli w aspekcie rozpoznawania infekcji COVID-19. Zastosowane metody chemometryczne pozwoliły na wydobycie różnicującej informacji spektralnej – dla śliny CoV(+) odnotowano podwyższony poziom metioniny, kwasów nukleinowych DNA/RNA, białek (np.: ferrytyna) oraz immunoglobulin, a dla wymazów z nosogardzieli CoV(+) zaobserwowano istotny wpływ neopteryny. Spośród testowanych metod z nadzorem, SVMC zapewniła najlepszą klasyfikację próbek zewnętrznych osiągając dokładność 90 % oraz czułość 100 % dla analiz śliny, a także dokładność 75 % oraz czułość 88 % dla wymazów z nosogardzieli.

Wyniki przedstawione w rozprawie doktorskiej wskazują, że technika SERS w połączeniu z metodami chemometrycznymi jest konkurencyjnym rozwiązaniem diagnostycznym dla metod obecnie przyjętych. Warto również zaznaczyć, że badania zostały wykonane przy użyciu przenośnego spektrometru Ramana, co stanowi ogromną zaletą dla diagnostyki wykonywanej w punkcie opieki.

Glossary

BV	bacterial vaginosis
Coinf.	coinfection
CoV(−)	COVID-19 negative patient samples
CoV(+)	COVID-19 positive patient samples
CT	charge transfer
EF	enhancement factor
FN	false negative
FOM	figures of merit
FP	fertile phase, false positive
Gon(−)	gonorrhea negative patient samples
Gon(+)	gonorrhea positive patient samples
IC	infected cohort
IFP	infertile phase
LDA	linear discriminant analysis
LSPR	localized surface plasmon resonance
LVs	latent variables
MRS	de Man-Rogosa-Sharpe agar
NAATs	nucleic acid amplification tests
NIPALS	non-linear iterative partial least squares
NPs	nanoparticles
PCA	principal component analysis
PCs	principal components
PCA-LDA	principal component analysis-linear discriminant analysis
PCR	principal component regression
PLS-DA	partial least-squares discriminant analysis
PLSR	partial least-squares regression
PNPs	plasmonic nanoparticles
POC	point-of-care
R^2_{cal}	R-squared (calibration)
R^2_{cv}	R-squared (cross-validation)
RMSEC	root mean square error (calibration)
RMSECV	root mean square error (cross-validation)
RT-PCR	real-time polymerase chain reaction
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SERS	surface-enhanced Raman spectroscopy
SIMCA	soft independent modelling of class analogy
STDs	sexually transmitted diseases
SVM	support vector machine
SVMC	support vector machine classification
S/N	signal-to-noise ratio
TN	true negative
TP	true positive
VF	vaginal fluid
VVC	vulvovaginal candidiasis
YPD	yeast peptone dextrose agar

Table of Contents

Chapter 1 Introduction

1.1 Discovery of the Raman effect

The Raman effect was first observed in 1928 by Indian physicist Sir Chandrasekhara Venkata Raman [1]. The motivation for conducting light scattering research was to understand the blue colour of the Mediterranean Sea. In 1930, he was awarded the Nobel Prize in Physics for his work on light scattering and the discovery of the phenomenon named after him – the Raman effect. Over the next 30 years, Raman experiments were based on large-volume and specially prepared samples to avoid any undesired effects, *e.g.* from impurities or fluorescence. The long integration time meant that spectrum acquisition took several days. The experiments were conducted using photographic plates, Hg lamps, and prism spectrographs [2,3].

Initially, physicists were deeply involved in understanding the Raman effect, resulting in hundreds of articles within the first seven years of discovery. Then, chemists began to take the interest, seeing in this technique new analytical possibilities. As James Hibben rightly noted, 'The Raman effect became the adopted child of chemistry'. By the late 1930s, the Raman effect had evolved into a leading method for the non-destructive analysis of inorganic and organic compounds with the unusual advantage of implementation to aqueous solutions [4].

1.2 The effect of light scattering

The primary forms of interaction between electromagnetic light and matter are absorption, emission, and scattering. In absorption, a molecule undergoes a transition process as a result of energy transfer between an initial and a posterior state. Depending on the energy of the incident light, the absorption process can occur at the rotational, vibrational, and electronic levels. Absorption occurs only when the energy of the incident photon is matched to the difference in energy levels. In the event of an energy mismatch, the light is scattered, and the molecule enters virtual state. Figure 1A presents the distribution of energy

levels, including electronic (ground and excited), vibrational states, and virtual ones created at the time of laser incidence [3].

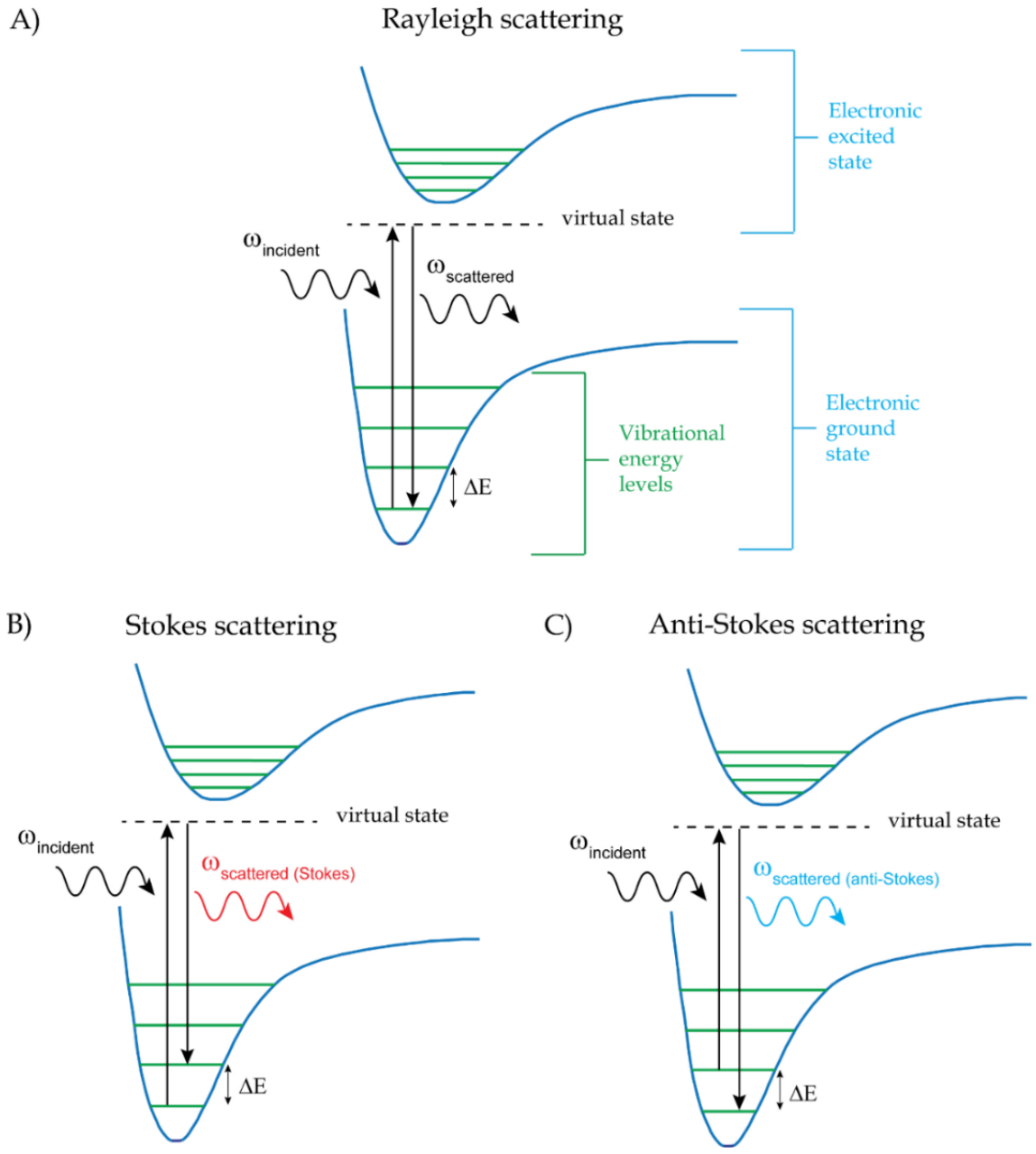

Figure 1 The energy transfer process in A) Rayleigh scattering but also B) Stokes and C) Anti-Stokes Raman scattering. The original drawing inspired by [3].

As a consequence of different energy expenditures, the effect of scattering can happen in two ways. When only the electron cloud is distorted, the Rayleigh scattering is formed. In this elastic process of light scattering, there is no change in photon energy ($E=E_0$). The energy of incident and scattered light remains the same (Figure 1A) [5]. When the vibrational

states of a molecule experience change (excitation or deactivation), the energy is transferred from the molecule to the photon or vice versa. This inelastic process of light scattering, known as Raman scattering, is the basis of Raman spectroscopy (Figure 1B and Figure 1C). The direction of energy transfer between the incident photon and the molecule means that Raman scattering can take two forms: Stokes and anti-Stokes scattering.

In Stokes scattering, the interaction with the incident photon increases the energy of the molecule. Consequently, the scattered photon has lower energy than the incident one by the value of the energy difference between vibrational levels ($\Delta E=E_0-E_{osc}$). The molecule is elevated from the ground state to the excited vibrational level (Figure 1B). In anti-Stokes scattering, the molecule, being at the excited vibrational state releases its energy and is transferred to the lower energy state. The scattered photon has higher energy than the incident one ($\Delta E =E_0+E_{osc}$) (Figure 1C) [3,6]. Thus, the scattering phenomenon is a two-photon effect that follows from quantum rules. First, the photon is absorbed into its virtual state and then emitted, resulting in a photon of the same, lower, or higher energy [7].

1.3 Theoretical fundamentals of Raman spectroscopy

Under the influence of incident electromagnetic light, the atoms of a molecule become polarised, *i.e.* the negative charge is shifted from the atomic nuclei [8]. The induced electric dipole moment is proportional to the intensity of the electric field with the proportionality constant as the polarizability, according to the equation [9]:

$$\vec{\mu} = \hat{\alpha} \cdot \vec{E} \tag{1.1}$$

where: $\vec{\mu}$ – the induced electric dipole moment,

$\hat{\alpha}$ – polarizability tensor,

$\vec{E}$ – electric field strength.

Since the polarizability tensor is additionally affected by the normal oscillation, the induced dipole moment is also modulated by the oscillation in question. For elastic scattering (Rayleigh), the polarizability can be seen as static, which does not change with time, so the frequency of the scattered electromagnetic wave is the same as the frequency of the external electromagnetic wave. For inelastic scattering, the polarizability tensor is changed during

the vibration, *i.e.* electron density is adjusted to the nuclear motions [7,8]. Hence, the time-dependent induced dipole moment can be expressed as:

$$\vec{\mu} = \underbrace{\hat{\alpha}_0 \vec{E}_0 cos\omega_0 t}_{term\ 1} + \underbrace{\frac{1}{2}\left(\frac{\partial \alpha}{\partial q}\right)\Big|_{q=0} \cdot q_0 \vec{E}_0 \cos(\omega_0 - \omega_R) t}_{term\ 2} + \underbrace{\frac{1}{2}\left(\frac{\partial \alpha}{\partial q}\right)\Big|_{q=0} \cdot q_0 \vec{E}_0 \cos(\omega_0 + \omega_R)\, t}_{term\ 3} \quad (1.2)$$

This equation presents the dependency of the polarizability α from the nuclear motion q according to Taylor expansion series around the nuclear equilibrium position q=0. The induced electron dipole moment $\vec{\mu}$ oscillates at three different frequencies: ω_0, $\omega_0 - \omega_R$, $\omega_0 + \omega_R$.

The first term of eq. (1.2) is related to the Rayleigh scattering, when the induced dipole oscillates with the same frequency as the driving electric field. The second and third terms of eq. (1.2) describe the oscillations of modulated frequencies (the Raman scattering), *i.e.* the difference or sum frequency between the external electromagnetic field (ω_0) and the frequency of the vibrational mode (ω_R). The second term of eq. (1.2) refers to the Stokes scattering resulting in the formation of the photon with smaller energy than the incident one and the excitation of vibration modes of the molecule. The photon is red-shifted in frequency in comparison to the light source. The third term of eq. (1.2) refers to the anti-Stokes scattering resulting in the formation of the photon with higher energy than the incident one and the deexcitation of vibration modes of the molecule. The photon is blue-shifted in frequency compared to the light source [8].

The processes involving induced dipole moment (expressed in equation 1.2) lead to the formation of Rayleigh band and sidebands (Stokes and anti-Stokes) associated with inelastic scattering that represent the vibrational frequencies of a molecule. Figure 2 presents the mentioned bands based on the example of the Raman spectrum of aspirin [8].

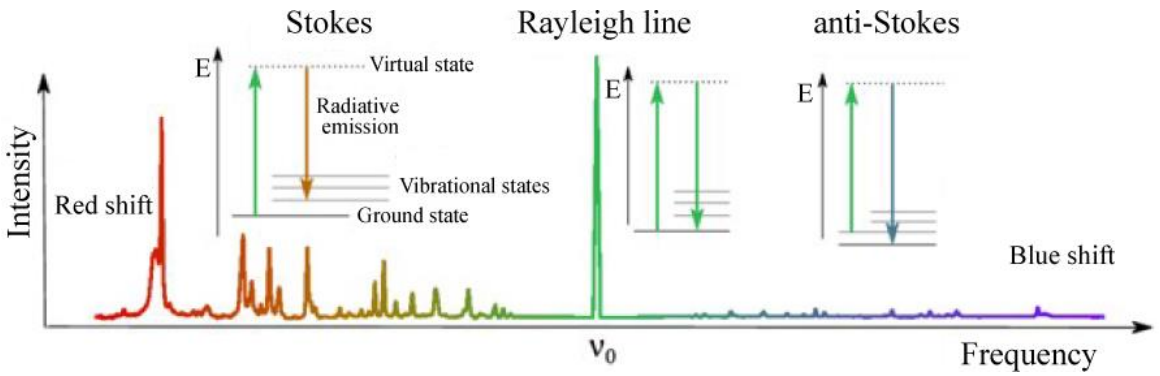

Figure 2 Band diagrams of the Raman effect representing Rayleigh line and sidebands such as Stokes (red-shifted) and anti-Stokes (blue-shifted) ones. The Raman spectrum of aspirin [10].

Raman spectroscopy plots the intensity of the Raman signal versus the frequency shift of the Raman signal in relation to the excitation source (Raman shifts / cm^{-1}) [5]. The integral intensity of Raman scattering depends on the fourth power of the frequency of scattered light and the magnitude of the change in molecular polarization, which is expressed as follows:

$$I_{RS} = I_0 \cdot (\omega_0 \pm \omega_R)^4 \cdot \left[\left(\frac{\partial \alpha}{\partial q}\right)\Big|_{q=0}\right]^2 \tag{1.3}$$

The (1.3) equation expresses directly the condition where the given vibration is Raman-active:

$$\left(\frac{\partial \alpha}{\partial q}\right)\Big|_{q=0} \neq 0 \tag{1.4}$$

Thus, Raman scattering occurs only when at least one component of the polarizability tensor is changed during the normal oscillation [11]. As each particle has a non-zero value of polarisability, it always scatters light in a Rayleigh-like manner and not necessarily in a Raman manner [7].

Although the two terms of equation 1.2 related to Stokes and anti-Stokes frequency are symmetric, their relative intensities are not equal. This is associated with the Boltzmann statistic, in which the probability of a photon encountering an already excited molecule is lower. Consequently, Raman scattering associated with Stokes processes is favored, resulting in more intense bands in the Stokes region [3,5].

1.4 Raman-based techniques – towards enhancing the Raman signal

Raman spectroscopy, as a branch of vibrational methods, probes specific oscillations in the material under analysis [12]. Spectra are composed of an appropriate sequence of specified bands and constitute the molecular representations of the analyzed individuum. The specific composition of the materials determines the creation of equally unique spectral signals, which provide insight into their molecular composition. These Raman signals acting as 'fingerprints' are the basis of qualitative analysis [13].

A number of significant advantages (*i.e.*, the analysis of a whole range of materials unrestricted to their phase or the non-contact nature of testing) make Raman spectroscopy appreciated in many research areas. Naturally, Raman spectroscopy also has several flaws. Beyond the fluorescence and possible material destruction, the limiting feature is the efficiency of the Raman effect itself. As one photon per million experiences inelastic scattering, the method's sensitivity can be insufficient, especially when testing materials at low concentrations [14].

To meet the ever-increasing demands for more advanced analyses, scientists began to streamline and improve Raman method. The contribution has resulted in 25 related techniques that solve problems encountered in Raman spectroscopy. One of them is tip-enhanced Raman spectroscopy. This method, which arises as a combination of Raman spectroscopy and scanning probe microscopy, delivers topographical and spectral information. The coherent anti-stokes Raman scattering improves the sensitivity by five orders of magnitude compared to spontaneous Raman. The adopted lower laser line makes this technique act non-invasively for sensitive biological systems. Femtosecond stimulated Raman spectroscopy enables tracking vibrational structural information with unprecedented temporal precision (below 50 fs) and spectral resolution (10 cm^{-1}). The confocal Raman microscopy provides high spatial resolution and clear image quality. Hence, it can be utilized to study the molecular composition of layered systems. The enhancement of the Raman signal can also be achieved by implementing plasmonic nanostructures. Such a modified technique is called surface-enhanced Raman spectroscopy (SERS) [13,15,16].

1.5 Surface-enhanced Raman spectroscopy (SERS)

1.5.1 SERS phenomenon

The origin of SERS dates back to the early 1970s when scientists attempted to measure vibrational spectra from monolayers deposited on substrates. At the time, such experiments were unimaginable due to the weak Raman scattering. However, in 1974, Martin Fleischmann published an article reporting the formation of a strong signal of pyridine molecules adsorbed on a rough silver surface [17]. In this experiment, the authors subjected the silver electrode to sequential cycles of electrochemical oxidation and reduction so as to increase the surface area and, thus, the number of illuminated molecules. Indeed, surface expansion was the first accepted explanation. The subsequent studies of van Duyne indicated that additional effects must have accompanied the signal enhancement, and roughness is not the only issue. Consequently, in 1977, the independent research of two groups led to the formulation of two mechanisms of Raman signal amplification – the increase of electromagnetic field on metallic nanostructures (by Jeanmarie and van Duyne) [18] and charge-transfer complex (by Albrecht and Creighton) [19]. Over the years, these two mechanisms have been extensively investigated, becoming known as electromagnetic (EM) and chemical (CT) enhancement [20]. Figure 3 illustrates the SERS phenomenon in which the incident light strikes the molecule adsorbed onto nanostructured particles, resulting in the emission of scattered light (Stokes and anti-Stokes).

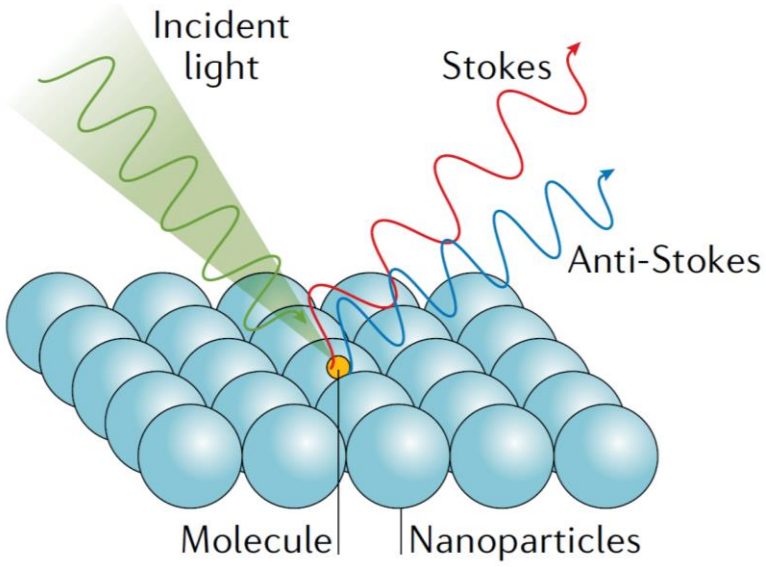

Figure 3 SERS phenomenon where the incident light strikes the molecule being adsorbed onto nanostructured particles, resulting in emitting radiation with a lower (red) and a higher (blue) frequency than the incident light, known as Stokes and anti-Stokes, respectively [21].

The electromagnetic theory involves the presence of two mechanisms of signal amplification: the initiating – local field (or near field) enhancement and the latter – re-radiation enhancement [22]. When the incident light strikes the metal surface, the electromagnetic field drives the delocalized conduction electrons to collective oscillations. The incidence of a laser beam with a resonance frequency corresponding to that of the electrons leads to the LSPR (localized surface plasmon resonance) process (Figure 4A). The term 'localised' refers to the fact that in metallic nanostructures, the electron oscillations are spatially localised in three dimensions of finite size and do not propagate. The nanoparticles that provide a strong LSPR effect are called plasmonic nanoparticles (PNPs). In the case of the visible and near-infrared region, the resonance frequencies fall into the range of Ag, Au, and Cu, making them appropriate materials [23]. The electromagnetic field experienced by molecules located on such plasmonic nanostructures is stronger compared to a field where the particle is not embedded. Also, Raman scattering is more effective under this condition. Since, at both stages, the enhancement is proportional to the square of the electromagnetic field, the total gain is proportional to the fourth power of the field [22].

The strength of the local field is highly dependent on the distance between the molecule and the metal surface. The inverse correlation implies that the SERS signal decreases when the distance increases. However, molecules not being in direct contact with the surface are still able to experience the enhanced electromagnetic field from the PNPs. The highest efficiency can be achieved for the interaction range of 1 – 10 nm between the molecule and the surface. In relation to the CT mechanism, the EM mechanism is called long-term and provides the strongest enhancement around 10^{10} [22,23].

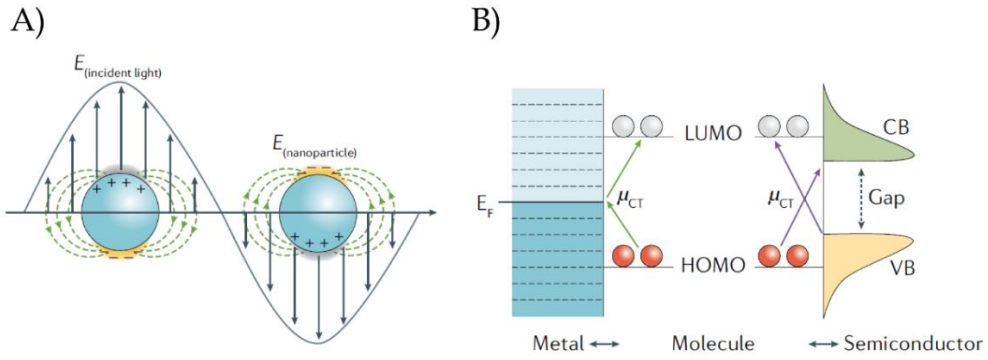

Figure 4 The mechanisms of signal enhancement in SERS spectroscopy A) electromagnetic effect – the localized surface plasmon resonance contribution and B) chemical effect – charge-transfer contribution; μ_{CT} – the CT transition, red and white circles represent molecular orbitals, HOMO – highest occupied molecular orbital, LUMO – lowest unoccupied molecular orbital, E_F – Fermi level, CB – conduction band, VB – valence band [21].

According to chemical theory, signal enhancement is achieved through a charge transfer (CT) process involving the Fermi level of the metal and the HOMO/LUMO orbitals of the adsorbed molecule (Figure 4B). The direction of the CT transition (μ_{CT}) depends on the laser energy, molecule, and material and can have a dual form, *i.e.* molecule-to-metal or metal-to-molecule. As a result of mentioned interaction, the photoinduced electron can be excited from the HOMO orbital of the molecule and transferred to the Fermi level of metal or excited from the Fermi level of the metal and transferred to the LUMO orbital of the molecule. For semiconducting materials with a full valence band and an empty conduction band, the energy gap between them serves as the Fermi level of plasmonic NPs in the CT process [21]. The consequence of these processes is the modified electronic and geometric structure of the molecule and its polarizability as well. The Raman cross-section of the vibrational mode is also altered compared to the free molecule. The chemical enhancement is at the level of $10^2 – 10^4$ and depends on the molecule type. Although the amplification so generated is smaller than the one that occurs according to the electromagnetic mechanism, it is also vital as it determines the spectral pattern (Raman shifts, ratios). Chemical amplification operates at the level of several Angstroms and is called short-term [21,22].

A particularly large enhancement effect is observed for small spatial regions of the plasmonic nanostructures, called ‘hot-spots’. They are located at the tips of elongated particles (nanorods) but also in gaps between nanoparticles, with the latter being favored for higher enhancement. The phenomenon of maximum enhancement in nanogaps can be explained based on Figure 5, which illustrates a molecule localized between two metallic spheres (dimer). The electric field polarises the nanoparticles, generating an excess of negative and positive charge distributed on opposite sides. In the case of polarization along the main axis (Figure 5A), the space between the induced surface charges is reduced, and the electric field between them increases as the nanoparticles approach each other. Hence, the reciprocal interactions of nanoparticles increase their polarizations. Consequently, the nanoparticle experiences the effect of the electric field, which is additionally amplified by the polarizing effect of the charges induced in the vicinity of the nanoparticle. This effect cannot be achieved when the electric field is polarised perpendicular to the principal axis (Figure 5B). In this configuration, the long-distance separation of such oriented, induced dipoles is not conducive to their reciprocal amplification under gap reduction [22]. In this aspect, theoretical studies comparing the enhancement factor (EF) for free and dimer forms of AgNPs were interesting. The results showed that the single Ag dimer with ~ 2 nm gap is

able to generate the enhancement equal to that of tens of thousands of single PNPs, and even the subsequent decreasing size gap results in increasing enhancement factors. These studies proved the importance of nanogaps in creating signal amplification [24].

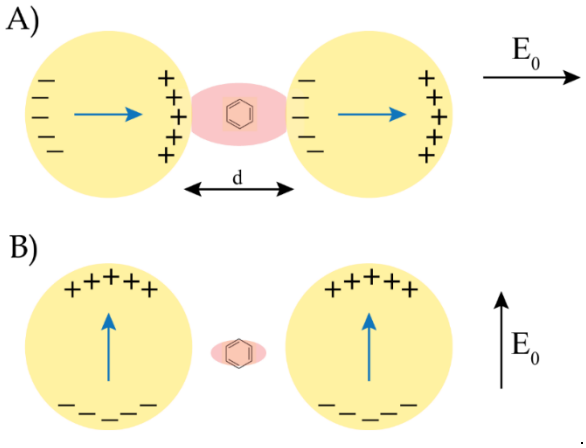

Figure 5 Two nanoparticles (dimer) polarized by the external electric field E_0 and separated by a distance *d* with a molecule located inside the formed gap. The polarization can be oriented A) along the main axis or B) perpendicularly to the main axis. The original drawing inspired by [22].

1.5.2 SERS substrates

The nanomaterials exhibiting LSPR properties that provide the enhancement of Raman signal are termed SERS substrates – mainly silver and gold but also copper and aluminum [22]. SERS substrates condition the SERS performance and play a key role in achieving high sensitivity and reproducibility in SERS detection. The ideal SERS substrates should meet certain characteristics. In addition to the magnitude of signal enhancement, high stability, adequate reproducibility, and affordable and simple production are also desirable features [20].

One of the categories of SERS substrates constitutes nanoparticles in suspension. These can be prepared *via* wet chemical synthesis (the reduction of Ag or Au ions in solution with reducing agents, *e.g.* citrate or sodium borohydride). This method enables the synthesis of nanoparticles with controlled size and morphology, *e.g.* nanospheres, nanorods, nanostars, nanocubes, or nanotriangles.

The nanoparticles can also be immobilized on a solid structure. According to one of the strategies, the solid substrate (*e.g.*, silicon, quartz, or glass) is functionalized with bifunctional molecules that contain a pendant functional group, typically $–HN_2$, –SH, –CN

(very often used is 3-(trimethoxysilyl)propylamine). Then, such a substrate is immersed in a solution with colloidal nanoparticles which are immobilized on the surface *via* the pendant group, making a very strong connection. The bifunctional molecule has a small Raman cross-section and is placed beneath the nanoparticles rather than between them, effectively preventing interference with the signal of analytes. This method offers the creation of substrates with a larger surface area (~ cm^2) and enhancement factor of ~10^4 [20].

Recent advances in nanofabrication techniques led to a broad range of SERS substrates. Nanolithography and all the lithography-based patterning techniques, which are part of top-down methods, provide the fabrication of highly ordered metallic nanostructure arrays. For example, in e-beam lithography, the narrow beam of electrons is focused and accelerated towards a thin layer of exposure-sensitive resist material coated on the substrate. This process leads to the modification of chemical properties of the resist layer (*e.g.*, solubility). Then, the patterned resist is used as a mask in the etching or deposition processes to create nanostructured metallic pattern. Although top-down methods enable the controlled formation of well-ordered periodic nanostructures with the pattern of interest, they require an additional step. As most of the fabricated materials are SERS-inactive, they have to be further decorated with SERS-active metal to create the plasmonic structures responsible for signal enhancement. There are several methods of metal deposition, such as electroless plating, electrodeposition, electrophoresis or colloidal coating technique, and physical vapour deposition. It is worth noting that the deposition process can also be needed for structural amendments associated with the nanogaps formation [25].

An important parameter for evaluating SERS substrate performance is the enhancement factor (EF). This parameter describes how much more signal is expected under SERS experiments compared to equivalent Raman experiments. The EF is a valuable parameter for specifying the performance of SERS substrates. Depending on the intended use, there are two conceptually different approaches for EF representation. In the first case, the EF is specified as the highest enhancement for a molecule localized in a 'hot-spot'. The second case refers to the average amplification from the whole surface area [22,26].

The EF, which may be of practical significance, is expressed as:

$$EF = \frac{I_{SERS}/N_{SERS}}{I_{Raman}/N_{Raman}} \tag{1.5}$$

where: I_{SERS} *and* I_{Raman} – the SERS and Raman signal intensities of the same analyte measured under the same conditions,

N_{SERS} *and* N_{Raman} – number of molecules illuminated by laser in SERS and Raman experiments.

Experiments have to be performed under the same conditions to obtain reliable results. Typically, for experiments with single molecules, EF takes values $10^8 – 10^{10}$ [26].

1.5.3 Application of SERS

Compared to normal Raman spectroscopy, the main characteristic of SERS is an increase in Raman intensity at the level of several orders of magnitude. Such a discovery of overcoming the low Raman scattering placed SERS at the center of attention in many areas of sensing, *e.g.* physics, chemistry, materials, and biomedicine. Signal enhancement, which is inherently built into SERS spectroscopy, enables the examination of substances at low concentrations or even single molecules. The signal-to-noise (S/N) ratio is also improved. Despite the increased intensity, SERS maintains valuable Raman spectroscopic features such as fingerprint-like information of the structure and a non-destructive and label-free nature. SERS utilizes the same equipment as Raman spectroscopy does and benefits from its advances in technology and instrumentation [27,28].

The mentioned features expand the horizon of SERS as an analytical tool for various applications. Nowadays, quality control and food safety are highly prioritized [29]. This area includes monitoring the freshness and internal components [30] as well as testing for the presence of metal ions [31], microorganisms [32], pesticides [33], and antibiotics [34]. In each of these aspects, SERS proved to be an appropriate tool. For example, Radu *et al.* demonstrated the detection of vitamin B2 and B12 in fortified cereals [35]. Dugandžić *et al.* developed a molecular dipicolylamine-based SERS sensor for the selective detection and quantification of copper (II) ions in water and white wine [36]. Camerlingo *et al.* designed glass slide substrates decorated with AuNPs to analyze apple/pear pulp and apple juices for fructose and pectin content, components of great importance for quality assessment [37].

SERS can also be used to identify natural dyes and glazes, often found in art materials, *e.g.* oil paintings [38], pastel paintings [39], and even textiles [40]. Generally, the analysed dye is incubated with Ag colloids and then is recognized by comparing with the reference,

so-called pattern-matching identification. Hence, in the art domain, SERS offers to render the initial colours of a faded painting, to trace the origin of the dyes and thus the history of the artwork, and to authenticate the work by comparing it with other works by the same artist. SERS also enables the recognition of counterfeit goods and currencies [41].

The unprecedented features have made SERS conquer new research fields, *e.g.* the identification of explosives and chemical warfare agents. The studies of Yang *et al.* proved that the specially designed sensor based on *p*-aminothiophenol-functionalised silver nanoparticles coated on silver molybdate nanowires could detect trace amounts of 2,4,6-trinitrotoluene with very high sensitivity and selectivity. With the continuously evolving technology, SERS can be a convenient tool for *in situ* hazard identification, being an essential technique for increasing safety [42,43]. SERS is highly advantageous for monitoring environmental pollution. With its ultrahigh sensitivity and ability to perform analyses in the field, SERS is well-placed to be regarded as a leading tool [44].

Throughout the years of research in manifold aspects, SERS evolved into an interdisciplinary technique. Of particular interest is the application of SERS for biomedical and biological purposes.

1.5.4 Biomedical application of SERS

There are two methodologies of SERS-based measurements performed in the aspect of biological and biomedical applications, *i.e.* direct and indirect (Figure 6). Direct strategy is oriented towards obtaining the vibrational information about the analysed system through the direct interaction between molecules and the SERS substrate. This label-free approach benefits from the original SERS feature, where information about the molecule is carried in the form of a unique spectral fingerprint (Figure 6A). In the indirect strategy, the SERS substrate is functionalized with a targeting moiety that can selectively recognize and bind to the molecule of interest. The capturing molecule is usually an antibody, an aptamer, or a nucleic acid sequence. The information about the target molecule is carried *via* Raman reporter, most often a dye or a sulfo-derivatives of the aromatic series. These chemical molecules with rather large scattering cross-sections result in a strong and distinctive signal enabling ultra-sensitive detection (Figure 6B).

Although both approaches can bring undesired effects from the matrix, well-designed can lead to high selectivity and sensitivity. With the indirect approach, obtaining molecular information about the target biomolecule itself is impossible. In turn, the direct one is susceptible to any changes in the environment of the analysed biomolecule. Hence, the choice of strategies should be considered and tailored to the purpose of the research, *e.g.* in the case of an indirect approach, whether it is justified to add additional complexity to the experiment and associated costs to capture the target in a mixture that may not necessarily complicate the SERS signal [23,45].

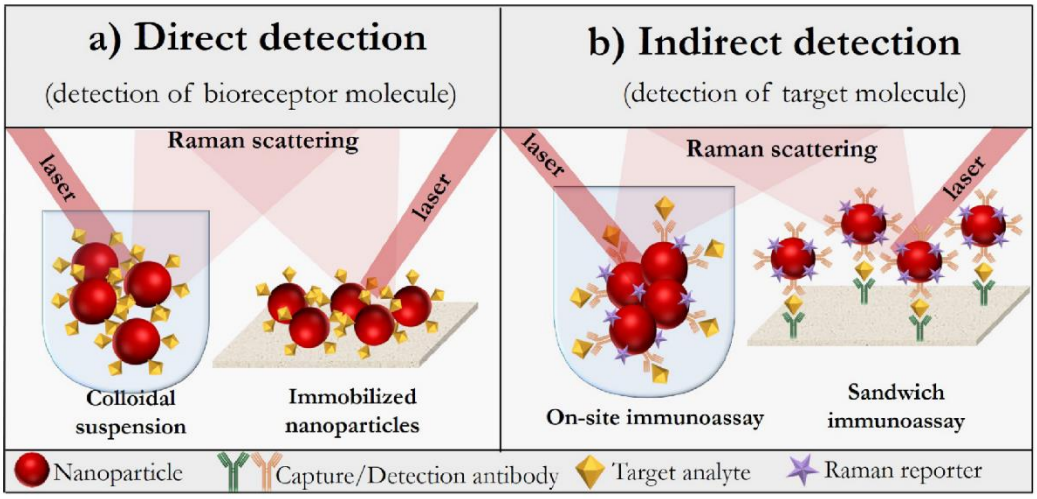

Figure 6 Two strategies of SERS detection: A) direct, the signal comes directly from the analysed molecules, and B) indirect, the nanoparticles are bound to the Raman reporters and then functionalized with targeting moiety, which specifically binds with the target [45].

The label-free strategy was utilized to study (macro-) biomolecules ranging from amino acids and proteins, through bases and nucleic acids, to metabolites in living cells. Of interest was the research of Han *et al.*, who designed the 'Western SERS' analytical procedure comprising protein electrophoresis, Western blot, colloidal silver staining, and SERS detection. This method successfully detected the myoglobin and bovine serum albumin on a nitrocellulose membrane [46]. Then, Ren's group developed the iodide-modified Ag nanoparticles method (Ag IMNPs) to examine proteins in their native state. In this approach, silver nanoparticles significantly enhanced the Raman signal. The coated iodide layers served as a barrier to prevent the direct interaction between the metal surface and proteins, maintaining their native structure. This method was tested on typical proteins (avidin, lysozyme, bovine serum albumin, cytochrome *c*, and hemoglobin) for which highly reproducible spectra were obtained [47].

Another class of investigated biomolecules is microRNAs (miRNA). Their detection has significant diagnostic value as they are recognized as regulators of gene expression that can serve as potential disease biomarkers. Driskell *et al.* fabricated silver nanorod arrays using oblique angle vapor deposition method to analyze miRNAs efficiently. This highly sensitive approach with sequence-dependent detection opens the possibility of classifying miRNA patterns with high accuracy [48].

The application of SERS for nucleic acids is usually based on the indirect approach for detecting and quantifying target strands. In this approach, the thiolated DNA forms are covalently attached to the plasmonic surface, and the intensity of the Raman label is changed upon the hybridization event. The direct analysis of DNA encounters difficulties due to the poor spectra reproducibility that hinders DNA detection. Scientists, driven by the desire to improve DNA detection in a label-free format, have put much effort into developing this issue [49]. One of the interesting studies in this aspect was investigating the role of the metal/liquid interface composition of silver colloids. The group of Guerrini analysed the commonly used plasmonic substrates (positively-charged spermine-coated nanoparticles and negatively-charged hydroxylamine-reduced and citrate-reduced silver colloids) and DNA probes of increasing complexity (from homopolymeric strands to genomic DNA duplexes). They documented that even the apparent minor changes in the chemistry of colloidal surfaces influence the affinity and the final SERS responses of single- and double-stranded DNAs [50].

SERS has also been tested to monitor diabetes in a rapid and minimally invasive manner. Douglas *et al.* developed the SERS-based sensor for the quantitative analysis of glucose in a model rat. The silver film over the nanosphere (AgFON) surface was functionalized with self-assembled monolayers (SAMs) that improve the signal from the analyte by partitioning glucose and localizing it within the first few nanometers of the SERS-active surface. Consequently, they presented the first *in vivo* application of SERS for glucose detection [51].

In medicine, SERS is also highly regarded for its ability to perform oncological imaging to detect cancerous tissue and distinguish it from normal ones. The imaging allows for monitoring treatment response, improving cancer staging, and screening for cancer recurrence [52]. The multiplexed detection of cancer-related molecular markers and their distribution in tumor microenvironment has been successfully demonstrated numerous times

[53–55]. Liu's group presented the fast detection of gastrointestinal cancers based on the visualization of the molecular phenotype of tissues with molecularly targeted SERS NPs contrast agents. In this strategy, antibody-conjugated SERS NPs were applied topically to the surface of the rat esophageal lumen to target EGFR and HER2, and a miniature spectral endoscope was used for the imaging of NPs bound in the esophageal lumen [56]. Recently, Zhao's group proposed a 3D se-SERS platform for multiplexed detection of exosomal proteins. The densely located 'hot-spots' effectively surround the analyte, providing sensitive and comprehensive SERS signals. By integration with deep learning models, seven proteins on the surface of patients' plasma exosomes were profiled, enabling the determination of cancer incidence and progression [57].

SERS has also been extensively exploited to analyse biomarkers (CTCs [58,59], circulating nucleic acids [60,61], proteins [62,63], and circulating tumor vesicles (CTVs) [64,65]) in liquid biopsies. A significant study in this area was done by Wei's group, who designed a microfluidic chip to monitor the level of carcinoembryonic antigen (CEA) and vascular endothelial growth factor (VEGF) expression in serum samples. The specific interaction of biomarkers with 4-mercaptobenzoic acid (4-MBA)-conjugated antibody functionalized gold nano-sheet substrate led to shifts in Raman spectra, thus allowing quantitative analysis due to linear tendency. The method's reliability was verified by comparison with the gold standard method making the SERS microfluidic chip promising in clinical settings for early diagnosis and prognosis of gastric cancer [66]. Also, many articles have demonstrated the successful differentiation between cancer and non-cancer groups based on the analysis of biomarkers in samples collected non-invasively [67], such as saliva [68–71], urine[72], and sputum [73].

SERS has gained an interest in the area of therapeutic drug monitoring (TDM), which is highly regarded due to its significance in providing patients with effective treatment and minimizing drug toxicity as well as the risk of adverse drug reactions [74]. Chen's group was involved in the fabrication of the substrate MIL-101(Cr)@Ag to determine the levels of chlorpromazine hydrochloride (CPZ) and aminophylline (AMP) in human serum [75]. Another study demonstrated the methodology that was meant to advance automating label-free SERS assays for TDM in clinical settings. In this strategy, a miniaturized solid-phase extraction method was integrated with a centrifugal microfluidic disk with incorporated SERS substrates. The analyses supported by machine learning methods yielded satisfactory values of the limit of detection for chosen drugs in human serum [76].

1.5.5 SERS for the analysis of microorganisms

Over the past few decades, tremendous progress has been made in SERS analysis of microorganisms, with particular attention to bacteria that, due to their pathogenicity, are a major clinical concern. To improve SERS performance, many aspects have been considered, *e.g.* SERS substrates and the associated size, shape, morphology, and type of active metal [77].

The detection of bacteria can be performed with AuNPs but also with AgNPs which are favored due to significantly larger enhancement. On the other hand, AuNPs are more biocompatible, highly stable, and not easily oxidized. The interaction of metallic nanoparticles with analysed bacteria is also an important issue. Beyond the simple mixing of bacteria with substrates, the nanoparticles can be synthesized directly in the presence of bacteria cells. The *in situ* synthesis increases the reproducibility and efficiency of the SERS signal [78].

Recently, it has been reported that the method of AgNPs production directs their specific interactions with bacterial cells. AgNPs fabricated in the presence of citrate (capping agent) can effectively partition through the cell envelope and bind to the components in the cell membrane, periplasmic space, and purine metabolites as well. This is opposite to the borohydride AgNPs (uncapped ones), which tend to be uniformly dispersed in the sample and do not aggregate on the surface of bacteria. Depending on the method chosen, nanoparticles selectively enhance the components, determining the formation of a specific bacterial SERS signal. Microorganisms are structurally complex biological systems for which signals from carbohydrates, proteins, and lipids overlap. Hence, the capped and uncapped AgNPs provide a means to differentiate between the signals of purine degradation and those of structural components of the cell membrane and periplasm space [79].

Silver nanomaterials (in addition to nanoparticle morphology) also take the forms of nanorods and nanowires and, when used in conjunction, can synergistically enhance the SERS signal, *e.g.* nanoplate-bacteria-nanorod spherocrystals [80]. Moreover, Ag- and Au-based substrates can be extended to form the whole range of multicomponent structures, *e.g.* core-shell bimetallic substrates, metal-semiconductor nanocomposite, metal-nonmetallic nanocomposite, and metal-polymer composite [81]. Recently, the green fabrication of plasmonic nanoparticles has become an interesting direction. In the innovative approach, AgNPs are synthesized using leaf extracts of *Nelomarckia cadamba* [82]. Chicken eggshells

can also serve as biotemplates for the subsequent sputtering of active metal [83]. Both alternatives demonstrated superior performance in bacteria detection.

The variability of spectral data can be related to the power and laser line, time of spectra acquisition, and calibration with standards. According to the literature, the commonly used laser lines for the analysis of biological samples are 633 and 785 nm, with the latter being preferred. This line enables acquiring signals rich in relevant structural information. The power of the laser line is also a crucial aspect as it determines the intensity of the scattered light. The optimal solution is the use of a laser line with 1.5 mW power, which provides an accurate SERS image of bacteria while avoiding the effects associated with thermal degradation of samples [84].

Microbiological aspects are another source of spectral variability. The nutrients of culture media exhibit spectral contribution that may hinder the complete manifestation of pure microorganisms. In particular, chromogenic media alter the SERS signal of bacteria due to the dyes they contain. The non-specific media, *e.g.* LB agar or BHI agar, are the best media selected for further comparison between different bacterial species [84]. To minimise potential interferences from the media, bacteria are prepared in washing/centrifugation cycles. Usually, three cycles are sufficient to obtain high-purity bacteria cells. The length of culture should also be balanced. It is not advisable to examine culture lasting longer than two days because of the decrease in signal intensity. As some strains are sensitive to changes in osmolarity, microorganisms are prepared in a buffer [84–86].

Upon the specific interactions with the environment, bacterial cells may exhibit physiological changes [86]. The study on bacteria culture indicated that treatments such as suspension in EtOH solution, UV exposure, freezing at −80 °C, or heating to 100 °C disrupt the structure of microorganisms, causing marked changes in their spectral image. A minimal centrifugal force is recommended to maintain bacterial viability [84]. On the other hand, an appropriate dose of stressor can act beneficially in magnifying the spectral differences among bacteria strains, which are owned by multiplex regulatory mechanisms that counteract stress and repair damage. It should be noted that the dose of stress must be balanced to trigger a sufficient biological effect and thus not exceed the tolerance limit of bacteria [87].

To effectively prevent unwanted spectral variability, the establishment of a standarised protocol is the appropriate solution. The analysis based on data obtained according to the

uniform procedure would facilitate the efficient identification and discrimination between bacteria species. Later, one can create the spectral library of microorganisms with special attention to measurement parameters [88].

The analysis of microorganisms is an important line of research in the context of reliable clinical diagnosis and the selection of customized treatments. To optimise the identification of pathogens, one can adapt various routes. Incorporating microfluidic systems reduces the analysis time and enables automated sample handling. Detection is performed with minimal sample volume and with reproducible measurement conditions. In turn, combining SERS with lateral flow assay tests increases the sensitivity compared to the standard approach. The detection is based on a color change on the test line and can be realized quantitatively by measuring the SERS signal intensity [89]. Choo's group created the SERS-LFA kits for rapid and accurate detection of *Orientia tsutsugamushi*. The analysis performed on 40 clinical samples demonstrated the consistency of results with a standard method (indirect immunofluorescence assay). In addition, the easy integration of the strips into a portable spectrometer ensures field measurements [90]. The created three-dimensional plasmonic nanostructures can be a convenient way to capture and identify pathogens effectively. Yeh *et al.* fabricated portable microfluidic platforms with carbon nanotube arrays of varying filtration porosity. Such a specially designed assay enabled the rapid enrichment and identification of viruses from clinical samples with 90 % accuracy. An additional benefit is that the captured viruses remain viable, permitting subsequent in-depth characterisation using various conventional methods [91].

In the case of fungal detection, a particularly interesting concept was the creation of an organic-inorganic nano-catcher based on magnetic nanoparticles and polymer nanoshells used with caspofungin drug. In this strategy, invasive fungi were isolated from complex samples and identified within minutes, which can be a significant advance in rapid clinical trials [92]. The SERS label-free analysis was also used to discriminate five species of apple spoilage fungi, achieving high accuracy for the tested algorithms [93]. The satisfactory results of SERS analysis were also obtained for the discrimination between *Trichophyton rubrum*, *Candida krusei*, *Scopulariopsis brumptii*, and *Aspergilus flavus*. The findings also demonstrated the possibilities of SERS and principal component analysis (PCA) methods in identifying *Trichophyton rubrum* in homogenized skin samples [94].

1.6 Spectral data analysis

Data obtained from spectroscopic measurements are highly susceptible to changes in settings or conditions. Hence, the genuine signal can often be distorted by noise and, more precisely, non-discriminatory sample-specific signal ranging from broadband background to high-frequency jitter [95]. The answer is data preprocessing, which eliminates the effects of unwanted signals without losing the genuine one. Preprocessing applied prior to chemometric methods prepares the data matrix and enhances its quality so as to increase the performance of subsequent analysis. Preprocessing usually consists of several steps following one another sequentially, and their choice should be tailored to the type of spectroscopic method. In Raman spectroscopy, several methods are used to remove the variability from different sources – fluorescence, cosmic rays, calibration errors, detector noise, laser power fluctuations, or signals from substrates [96,97]. These can be, *e.g.* baseline correction, normalization, or smoothing.

Figure 7 presents the workflow of Raman spectra analysis. The obtained raw spectra are initially preprocessed so that they can be deeply analysed with chemometric methods.

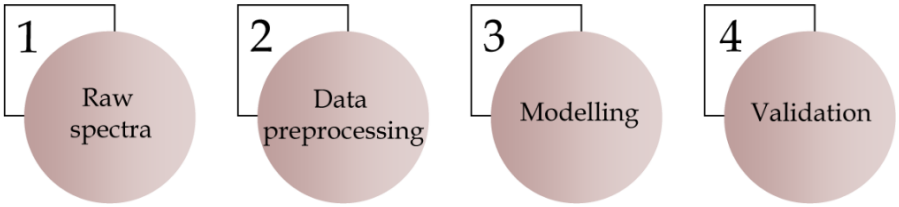

Figure 7 The workflow of Raman spectra analysis. The obtained raw spectra are initially preprocessed and then subjected to chemometric analysis, which consists of modelling and validation steps.

1.6.1 Chemometric analysis

The development of analytical methods brings with it the generation of an enormous amount of information (features, variables) by the number of samples (objects, observations) to be analysed in a short time [98]. Advances in information technology initiated the emergence of multidimensional data matrices, which use various mathematical and statistical algorithms to facilitate data analysis. Chemometric (multivariate data analysis) as a chemical discipline can be adopted to massive data to efficiently extract maximally relevant information and gain knowledge about the chemical system under analysis [99].

Prior to multivariate data analysis, spectroscopic data are displayed in the form of a matrix where columns represent variables (*e.g.*, Raman shifts) and rows represent observations (spectra), *i.e.* the spectroscopic profile of every analysed sample is represented by each row. Every spectroscopic measurement/observation consists of relevant signal and noise. The relevant signal is an actual representation of the underlying information associated with the property of interest. The noise is all the remaining information irrelevant to the interested property (*e.g.*, spectral noise). Hence, the primary idea of multivariate data analysis is to separate the relevant signal from the noise [96].

Depending on the property under consideration, chemometric methods are divided in various ways, and in the first distinction, there are classification/discrimination and regression/prediction methods. Figure 8 presents the classification of chemometric methods along with examples.

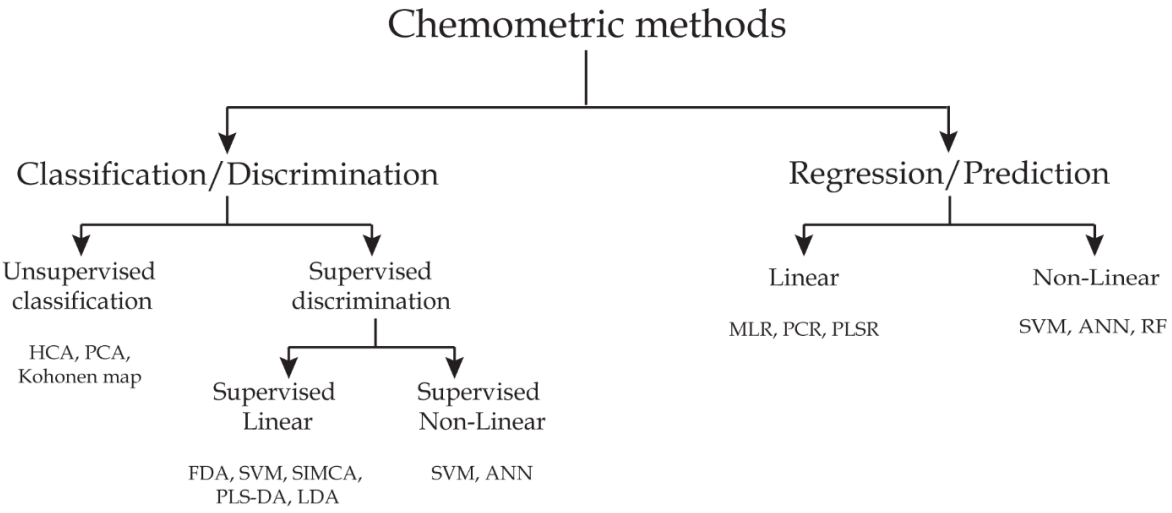

Figure 8 The classification of chemometric methods. The original drawing inspired by [97,100,101].

1. Regression and prediction

The regression methods are based on the concept of finding a correlation between a set of explanatory variables (*X*) and one (*y*) or several properties of interest (*Y*), *i.e.* relating properties of chemical samples (*Y*) to their chemical composition (*X*) or biological activity of studied compounds (*Y*) to their chemical structure (*X*) [101,102]. The typical examples of regression include, *e.g.* modelling the content of protein in wheat based on their NIR signals or modelling the antioxidant properties of tea extract based on chromatographic fingerprints [103].

In chemometrics, models are developed and optimized based on the selected sample set (well-known *X* and *y*), a so-called calibration or training set. The ultimate goal is to introduce an optimized model to gain information about unknown objects, *i.e.* to predict *y* values for them [101]. The prediction includes both quantitative (continuous-valued, *e.g.* protein content) and qualitative (discrete, *e.g.* healthy/ill, compliant/non-compliant, class membership) properties of the analysed system [103].

2. Linear modelling principle

The simplest regression model (univariate, linear model) relates one quantity of interest *y* (response) to one explicative variable (x_1) through the equation:

$$\hat{Y}_{(n,1)} = b_0 + b_1 x_{1(n,1)} \tag{1.6}$$

where: $\hat{Y}$ – the predicted value of the real *y*,

b – the regression vector, which is composed of b_0 (the offset) and b_1 (the slope).

When *X* is multivariate (*i.e.* each sample is described by *p* several variables), the multivariate linear model relates *X* to *y* as follows:

$$\hat{Y}_{(n,1)} = b_0 + b_1 x_{1(n,1)} + b_2 x_{2(n,1)} + \cdots + b_p x_{p(n,1)} = b_0 + X_{(n,p)} b_{(p,1)} \tag{1.7}$$

where: $b_1 \ldots b_p$ – the regression coefficients for each variable and refer to the regression vector *b* (in the right part of the equation),

b_0 – the offset,

X – the matrix of independent variables. Matrix dimensions are noted in parenthesis.

Creating the regression model means finding the optimal values of *b*, *i.e.* values that lead to the lowest error in predicting the response. Regression-based methods are multiple linear regression (MLR), principal component regression (PCR), and partial least squares regression (PLSR) [101,104,105].

3. Classification and discrimination

Another task of multidimensional data analysis is to classify a sample based on its individual and characteristic features. From a chemometric point of view, this notion is embodied in pattern recognition methods, which are defined primarily as a classification to assign samples to groups. Due to the considered feature, classification/discrimination methods can be divided into unsupervised and supervised [100].

Supervised methods use the patterns established for all analysed classes (predetermined classes) to identify and classify an unknown sample to one of the known classes based on its measurement pattern. The procedure of supervised methods comprises a few strategies. The selection of variables should be guided by information content, *i.e.* variables that encode relevant information are retained, and those that contain noise or have nonsignificant discriminatory power are discarded. The building of a model is based on training samples – objects have to be categorized in an exhaustive (*i.e.*, each observation belongs to a category) and mutually exclusive (*i.e.*, there are no observations belonging to different categories) manner. A model is validated on a new set of external samples to assess the reliability of the classification. Unsupervised methods do not rely on predefined classes, and objects are distributed along the dataspace, revealing the correlation between them. There is no information about class membership [97,98,100].

Supervised pattern recognition methods can be classified as those focused on the discrimination among classes (*e.g.*, linear discriminant analysis – LDA, partial least squares discriminant analysis – PLS-DA), and those oriented towards modelling classes (*e.g.*, soft independent modelling of class analogies – SIMCA). Although both types of methods have the same objectives, they differ in the way they work. In discrimination methods, also known as ‘hard-modelling’, models are created based on the differences between objects of various classes. This is in contrast to class modelling analysis or ‘soft modelling’ methods, which try to identify features that make individuals from the same class similar to each other. An important feature of discriminant methods is that the analysed object is always assigned to one of the given categories. In model-based methods, an object is likely to be rejected when it does not fit into a category model [97,98,100].

Figure 9 illustrates the graphical representation of the issue at hand for the case where the two classes span a two-dimensional feature space, and the hollow shapes represent samples to be classified (not included in the model). In class modelling, each target class is

modelled separately and positioned in the feature space (bounded by the contour as demonstrated in Figure 9A). This results in four possibilities for object classification; the classified object can fall into first class, second class, both classes, or outside of the modelled classes. Hence, in this particular case, the objects represented as a hollow star and a circle are correctly assigned to classes. A hollow square is classified as a member of both classes, while a hollow triangle is rejected from the modelled categories. In discrimination, data dimensionality is subdivided into regions, thus yielding the decision surface that delimits the boundaries between classes in the feature space. Hence, in the analysed case (Figure 9B), the black line divides the entire data space into two areas, thus designating two classes. The objects will always be classified into one of the classes. The objects represented as a hollow star and a circle are classified correctly, while the one marked as a hollow triangle is incorrectly assigned to the 'blue dots' class. Contrary to class modelling, the object represented as a hollow square is assigned to one class [106].

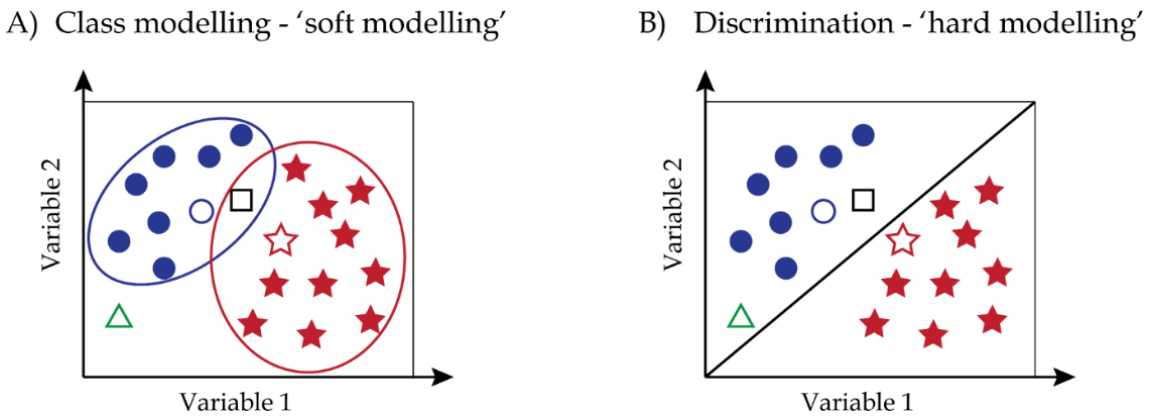

Figure 9 The illustration of two model concepts A) Class modelling 'soft modelling' and B) discrimination 'hard modelling' based on two-dimensional future space with two classes – blue dots and red stars (training objects). Hollow shapes represent the objects to be classified (test objects) (not taken for model construction). The original drawing inspired by [106].

The mathematical form of the functional relationship associated with the geometric shape of the decision boundary in the multidimensional (feature) space leads to another division. According to this property, classification can be performed in linear or non-linear strategy, where the latter can be subdivided into quadratic or polynomial. In linear methods, the decision boundary takes the form of a linear function of the original variables; a line, a plane, and a hyperplane in case of two, three, or more than three dimensions, respectively. In non-linear methods, more complex hyperspaces partition feature space according to existing classes, *e.g.* quadratic approach involves hyperelliposoids, hyperparaboloids,

hyperspheres. Hence, according to this classification, linear methods are PCA, PLSR, LDA, and PLS-DA. Examples of non-linear methods are artificial neural networks (ANN) and self-organizing maps (SOMs) [107]. Support vector machine (SVM) can handle simple linear tasks and more complex ones, *i.e.* non-linear [108].

4. Model optimization and validation

Chemometrics involves empirical models that, based on given measurements, provide information about the structure of a system and allow the prediction of one or more properties. However, for a specific data set, it is possible to create many models characterized by different performances (due to the representativeness of samples, the algorithm, or the method itself) [105]. Incorrect selection of latent variables (LVs) (*e.g.*, in PLSR, PCA-LDA) can also promote undesirable effects such as overfitting and underfitting. Over-fitting gives the illusion that the model is able to predict more than it actually does; the model is unstable due to noise and irrelevant information. Conversely, underfitting leads to a model with inadequate performance; the model lacks the relevant part of the signal [101]. Thus, it is important to evaluate the quality of the model. The chemometric strategies oriented towards it are collectively referred to as validation [105,109].

The predictive power of a model can be evaluated with the figures of merit (FOM): the coefficient of determination R^2 – a parameter that describes the fraction of the original data explained by the model (*e.g.*, $R^2 = 1$ means that 100 % of the data is explained), RMSE (root mean square error) – describes the standard deviation between reference and predicted values. To compare the performance of different models, the criteria are calculated at calibration and validation sets. The minimum error corresponds to the optimal complexity of the model, also known as the optimal model rank [101,110].

The most used validation technique to optimize the model is cross-validation. This approach utilizes the iterative resampling strategy, where the original data set is split into a training set (for learning the parameters) and a test set (for evaluating the model performance). In *K*-folds cross-validation, the data are divided into *K*-sub folds (subsets); one (validation set) is left out at a time to determine the predictive performance of the model created from the remaining *K-1* folds. This process is repeated *K*-times till all subsets are validated. [96]. Cross-validation involves an internal validation set and is the method of choice in case of insufficient samples. Whenever possible, it is recommended to introduce

the strategy of an external validation set. The brand-new samples, completely independent from the calibration data (*e.g.*, samples from the new batch, or measured later), can ideally serve to recreate the conditions of routine use of the model [101,105].

The following subsections describe the chemometric methods used in the research presented in this doctoral dissertation.

I. Principal component analysis (PCA)

PCA belongs to the class of data reduction methods. In PCA, the main systematic variation in a dataset is extracted by breaking down the information of original data (matrix *X*) into a series of variance-scaled orthogonal vectors called principal components (PCs) [111–113].

The data to be analysed are collected in a matrix *X* that comprises *n* rows (objects, *e.g.* spectra) and *p* columns (variables, *e.g.* Raman shifts). PCA decomposes the *X* matrix into the structured and the noise part, *i.e.* two smaller matrices, one of the scores (*T*) and the second of loadings (*P*) and the error (*E*) [101,114].

$$X_{(n,p)} = T_{(n,k)}P^{T}_{(k,p)} + E_{(n,p)} \tag{1.8}$$

where: *X* – the original data matrix with *n* (rows) *x p* (columns),

P – the loading matrix for *k* components,

T – the corresponding score matrix,

E – the residual variance not explained by the model.

The (1.8) equation can be expanded to:

$$X = t_1p_1^T + t_2p_2^T + t_3p_3^T + \cdots + t_kp_k^T + E \tag{1.9}$$

where: $t_ip_i^T$ – the orthogonal principal components creating the model and $t_1p_1^T$ is the first principal component and $t_2p_2^T$ is the second principal component, and so on.

The loading matrix $P_{(k\ x\ p)}$ identifies the k directions, *i.e.* principal components (PCs) along which the data are projected. The results of such projections, *i.e.* the coordinates of every object onto this reduced subspace, are collected in the score matrix $T_{(n\ x\ k)}$ [105].

The loading matrix P consists of the retained k PCs and acts as the transformation matrix between the original variable coordinate system and the new PC system. Each column in P matrix represents a PC, *i.e.* loading. In the PC-coordinate system, the particular object i is projected onto the PC1, PC2, PC3 and so forth, giving the corresponding y_{i1}, y_{i2}, $y_{i3,}$ and so forth. This means that object i (as well as every analysed one) is presented as a point in the PC-coordinate system with scores (coordinates) y_{i1}, y_{i2}, $y_{i3,}$ and so forth. The number of scores is the same as the number of PCs, *i.e.* k. The score matrix T contains scores for every object, with each row indicating a particular object. In turn, columns indicate the orthogonal score vectors, each representing the scores for one specific PC [96].

The first component (PC-1) explains the greatest possible variability as well as the second one (PC-2), which additionally comprises the information not being captured by the first one. Hence, each subsequent principal component explains the highest variance of the information under the condition that it is orthogonal to the previous one until all the data matrix is decomposed. The total number of principal components reflects the number of dimensions in the original data – the fewer PCs needed, the simpler the model is. The information about the explained variance by each PC is included along each axis [115,116].

PCA is a perfect tool that provides global insight into the analysed data set. This method identifies similarities, differences, and general patterns hidden in the data of interest. It can also be used to recognize artifacts and outliers. In light of ever-increasing data, PCA is a proper method to obtain condensed data that could be subsequently used instead of original ones [117].

II. Partial least squares regression (PLSR)

PLSR belongs to the partial least squares (PLS) methods (also called *projection to latent structures*). PLSR is used to predict a set of variables (Y) from a set of predictors (X). PLSR is able to predict many variables and handle the case of multicolinear predictors (*i.e.* the predictors are not linearly independent). It is a versatile tool for a large data set.

PLSR is aimed to find latent variables (denoted by T in matrix notation) that model X and simultaneously predict Y. This is expressed as a double decomposition of X and the predicted $\hat{Y}$.

$$X = TP^T \quad \text{and} \quad \hat{Y} = TBC^T \tag{1.10}$$

where: P – X loadings,

C – Y loadings,

B – the diagonal matrix.

Latent variables are ordered according to the amount of variance of $\hat{Y}$ that they explain. PLSR is focused on reducing the dimensionality of X and Y in latent subspaces with the use of score projections [101,118].

III. Partial least squares discriminant analysis (PLS-DA)

PLS-DA is a supervised method that evolved from PLSR algorithm aimed at resolving regression problems. PLS-DA is designed to handle the predictive and descriptive modelling tasks and the selection of discriminatory variables. Hence, PLS-DA links the reduction of the dimensionality of complex data with discriminant analysis (sample classification based on the constructed model). To provide pattern recognition information, PLS-DA involves explanatory and response variables. The analysed samples are recorded in a matrix X described by $n \times D$ dimension, where n objects and D features create explanatory variables. The corresponding matrix Y of $n \times q$ size refers to the same n objects with q properties and is described in a category system (response variables). Initially, such categorical variables are encoded into continuous variables data, and their number corresponds to the number of considered categories (*e.g.*, '0' and '1' dummy variables for describing two categories). The number of rows of matrix X is the same as that of matrix Y and refers to the number of objects. In turn, the number of columns of matrix Y corresponds to the number of established categories. Thus, PLS regression is used to build a model based on matrix X of the predictors and a dummy matrix Y that expresses the class membership, consequently leading to classification based on the generated prediction values (real values) [119–124].

IV. Soft independent modelling of class analogies (SIMCA)

SIMCA belongs to the supervised methods where a created model consisting of objects in well-defined classes (training set) is used to identify unknown objects (test set) [98]. SIMCA involves PCA method to model each class separately in a PCs dimension. A certain number of PCs is related to the spatial representation of the class (*e.g.*, line, plane, hyper-plane) and is significant in determining critical distance for classification. The number of PCs is retained to account for most of the variation within classes and includes a high signal-to-noise ratio by excluding so-called secondary or noise-laden principal components in the modelled class. The decision of class membership is made based on the similarities of the unknown sample (test sample) to its actual representative class in the lower dimensional space (PC space) [98,113].

The basis of the classification is to determine a confidence region for each modelled class. The membership criterion limit is designated with the squared Mahalanobis (T^2) and Euclidian (Q) distance. The Mahalanobis distance is the distance of an object to the center of the score space; it indicates how far the individuum is from the distribution of 'normal' objects in the space spanned by significant PCs. Euclidean distance is the distance of an object to its projection onto the PC space; it describes how well the individuum is fitted by the PCA model.

The new object to be classified is projected onto the model subspace, and the calculated T^2 and Q are used to designate its overall distance $d_{i,g}$ to the model:

$$d_{i,g} = \sqrt{\left(T_{i,g}^2\right)^2 + \left(Q_{i,g}\right)^2} \tag{1.11}$$

where: T^2 – the squared Mahalanobis,

Q – the Euclidean distance.

The subscript indicates that the i^{th} object is tested against g class. When the object falls within class limits, it is classified as a member of the class [105].

Figure 10A presents the graphical representations of two classes modelled by PCA prior to SIMCA analysis – Class 1 described by two PCs, and Class 2 described by a single PC. An object to be classified is compared to each class by projection on these two models

– a plane for Class 1 and a line for Class 2 (Figure 10B). This gives appropriate distances between this object and two classes [125].

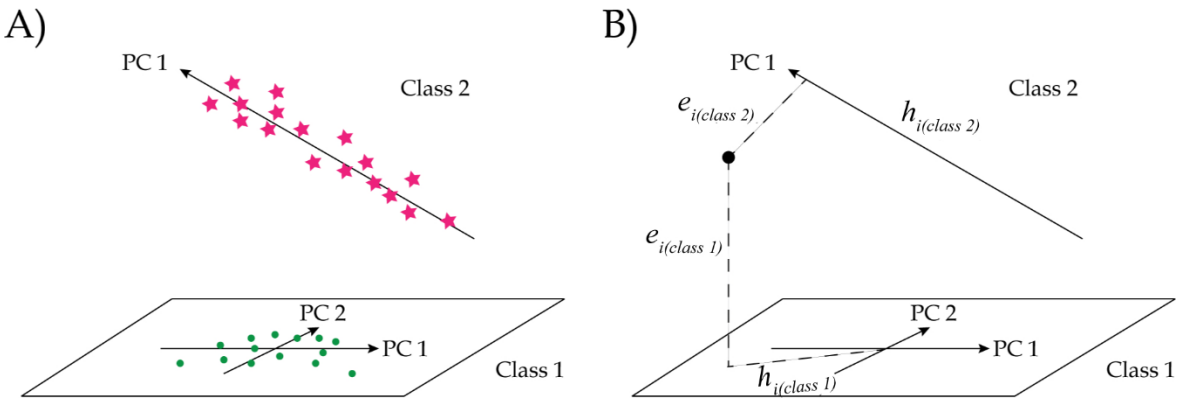

Figure 10 SIMCA analysis performed for two classes A) Class 1 is modelled by two PCs and Class 2 is modelled by a single PC, B) A new sample (marked as a black dot) is compared to each group by projecting it on models, a plane for Class 1 and a line for Class 2. This results in the distance $e_{i(\text{class 1})}$ and $h_{i(\text{class1})}$ for Class 1 and $e_{i(\text{class2})}$ and $h_{i(\text{class2})}$ for Class 2, where e_i is the Euclidean distance of the object to the model and h_i is the Mahalanobis distance within the PC space. The original drawing inspired by [125].

SIMCA determines the model distance and discrimination power. The model distance is calculated as the geometric distance from the principal components models; if it equals more than three, the class discrimination is successful, but if it is lower than one, the classification cannot be made. Discrimination power describes how specific variables contribute to class differentiation, and those with values higher than three are considered remarkable in total discrimination [98,115].

V. Linear discriminant analysis (LDA)

LDA aims at finding the linear discriminant that maximizes the between-class variability and minimizes the ratio within-class variability so that the separation between classes is the highest [126]. LDA may be viewed as a one-dimensional projection of objects onto 'LDA-axis' separating classes. LDA is not limited to the straight-line separation case. Depending on the complexity of the problem in question, the classification can be expressed with a quadratic separator or Mahalanobis distance. The quadratic DA can be an ideal solution for the separation of classes characterized by the main variability in different directions (but only when the training set is large and representative). LDA cannot handle the analysis of high-dimensional data, *e.g.* spectral data, when the number of variables

exceeds the number of samples. To overcome this limitation, it is recommended to use PCA as a pre-classification step (PCA-LDA). The use of only the diagnostically significant components with compact information about the system is a proper starting point for the subsequent LDA analysis. In addition, the requirement for fewer variables than samples within a class is much more easily fulfilled [110].

VI. Support vector machine classification (SVMC)

SVM is designed for analyses of linear data and non-linear ones using Kernel function. SVM is also intended for resolving the classification task – support vector machine classification (SVMC). The idea is to search for a hyperplane that would effectively separate the analysed classes by maximizing the margin of separation. This is achieved through the extraction of a subset of a training set that serves as support vectors. Support vectors are objects that lie on the borderlines between classes; their number determines the dimensionality of the feature space. In turn, a margin is defined as the geometrical distance of a blank space between the two classes. The margin is determined around the line of separation; the maximum margin is defined as a double distance from support vectors to the hyperplane that separates classes. Ideally, the margin is free of any object, and the separation is described as 'high quality'. Using the designated margin and support vectors, SVM establishes the hyperplane of separation [107,110,127,128]. The general idea of SVM is illustrated in Figure 11 – two classes (Class 1 and Class 2) to be separated by a hyperplane.

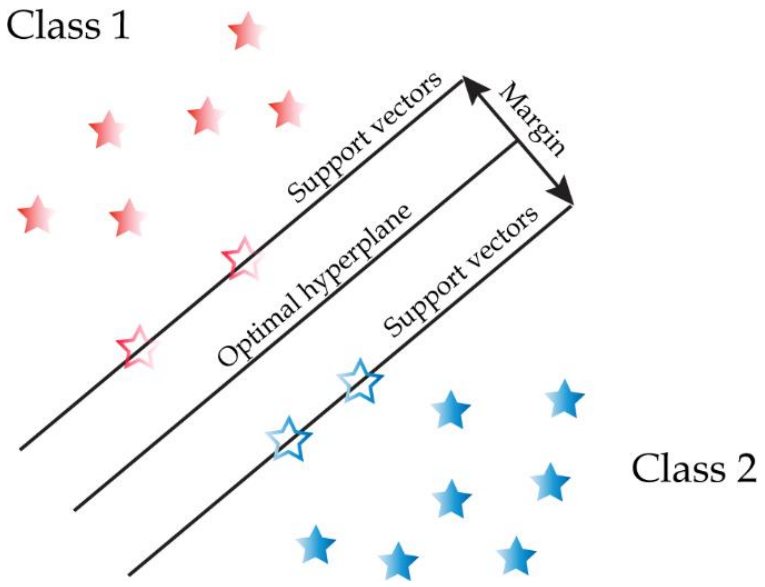

Figure 11 Representation of general classification hyperplane in SVM method that maximizes the margin of the training data. The original drawing inspired by [107,110].

To effectively separate classes, it is necessary to select a proper Kernel function (linear, polynomial, radial basis function, and sigmoid) and adjust C and gamma parameters [110]. The regularization parameter (C) describes the dependency between the correct classification of the training set and the size of the margin (model complexity). The Gamma parameter defines the level of influence of individual training points and affects the shape of the hyperplane [129].

1.7 The selected disease entities

1.7.1 Sexually transmitted diseases – gonorrhea

Gonorrhea is considered one of the most prevalent sexually transmitted diseases (STDs) caused by bacteria *Neisseria gonorrhoeae.* According to a 2020 report by the World Health Organization (WHO), there were 82.4 million new cases among people aged 15 – 49, and gonorrhea was recognized as a significant public health problem [130].

N. gonorrhoeae has an affinity mainly to the mucosal surface of the urogenital tract and can attach to the transitional and columnar epithelial and the stratified squamous epithelium of the ectocervix. This results in urethritis in men and cervicitis and urethritis in women. *N. gonorrhoeae* is able to enter anal, oral, or conjunctival mucosal and invade the bloodstream leading to the disseminated gonococcal affliction of the joints or skin. If gonococcal urethritis is left untreated or under-treated, it can evolve into an ascending infection of the upper genital tract. In consequence, the developed epididymitis or pelvic inflammatory diseases (PID) can lead to severe sequels, including infertility and ectopic pregnancy [131,132].

The incubation period of gonococcal urethritis is between two and eight days, and the clinical representation in men and women is substantially different. In men, acute gonococcal urethritis is mostly symptomatic: abundant and purulent discharge, macerated mucosa of the external orifice, and dysuria. In contrast to men, the asymptomatic course in women is less common, and if this disease does occur, the symptoms are non-specific [131,132].

One of the traditional diagnostic methods of gonorrhea is Gram staining of samples and microscopic evaluation for the presence of *N. gonorrhoeae* diplococci. The method's

sensitivity and specificity vary between studies and highly depend on the analysed specimen. In the case of urethral discharge of symptomatic men, the highest noted sensitivity ranges from 89 % to 98 %, and specificity is greater than 95 %. For asymptomatic courses the sensitivity can be lowered to 40 – 50 %. Such discrepancies in parameter values may be related to the quantity of bacterial load. Gram staining is not an appropriate method for the analysis of pharyngeal specimens (because of the prevalent *Neisseria* spp. with similar morphology) and rectal (due to the sensitivity lower than 40 %) [131].

The sensitivity of the culture method for urogenital specimens ranges between 72 % and 95 % and is much lower for rectal and pharyngeal specimens [131]. The obtained bacterial colonies should undergo presumptive identification by Gram staining and oxidase or catalase test. In the case of positive results, confirmatory identification tests should be performed (*e.g.*, fluorescent antibody test, coagglutination test, colourimetric test, or API biochemical test, including carbohydrate degradation test) [133,134]. Nowadays, nucleic acid amplification tests (NAATs) are recommended in high-income countries. The pathogen's viability condition does not have to be met, making storage and transport less restrictive. A good alternative is in-house polymerase chain reaction (PCR) tests, which can be used in resource-limited settings [131].

As was mentioned in the above subsections, SERS technique is a promising diagnostic tool. However, despite many scientific reports on the medical use of SERS, none concern the analysis of urethral discharge in terms of sexually transmitted diseases. The research presented in this dissertation is the first of its kind to address STDs problems.

1.7.2 Vaginal infections

Vaginal infections, medically termed vaginitis, refer to a series of symptoms of the reproductive tract resulting from disorders of the vaginal microflora. Vaginitis reduces the quality of life, causes discomfort, and can lead to serious complications. Depending on the etiological agent (microorganism), infections can be divided mainly into bacterial vaginosis (BV) caused by bacteria and vulvovaginal candidiasis (VVC) caused by fungi. Coinfections intensify the whole health problem. It is estimated that approximately 20 – 30 % of women with bacterial vaginosis (BV) are coinfected with *Candida* species [135–138].

Vulvovaginal candidiasis (VVC) affects one in every two women during their lifetime. Although most of them experience VVC infrequently, there is a group of women with recurrent VVC infection defined as three or more episodes within 12 months. VVC results from an overgrowth of *Candida* spp. (*e.g.*, *Candida albicans, Candida glabrata, Candida krusei*) relative to the natural microflora [139]. Symptoms of VVC, such as burning during micturition, irritation, soreness, and thick-curd-like vaginal discharge, are not pathognomonic and cannot be considered a major determinant of this particular infection. The pH of vaginal milieu is between 4.4 and 4.7 and does not increase under the influence of *Candida* spp. proliferation [140].

In clinical practice, VVC is diagnosed based on a microscopic examination of wet preparation (saline, 10 % KOH) of vaginal discharge that demonstrates budding yeast, hyphae, and pseudohyphae. Adding KOH upgrades the analysis by destroying cellular material (bacteria and cells) that might hinder the visualization of yeast and mycelia. In the case of negative results and persistent symptoms, a culture method is recommended [141].

BV occurs due to the replacement of *Lactobacillus* spp. to diverse facultative and strict anaerobes such as *Gardnerella vaginalis, Dialister* spp., *Mobiluncus* spp., *Prevotella* spp., *Atopobium* spp. and many others [142]. As the BV-related pathogens can also be found in non-infected individuals, they cannot be specific markers of the BV infection [143]. Hence, bacterial cultivation is not approved for the BV diagnosis. Instead, it is recommended to evaluate Gram-stained secretions and consider Amsel criteria [144].

According to the four Amsel criteria, meeting three of them means the presence of the infection: a thin and homogenous swab with milky consistency, an elevated pH (the cut-off value 4.5), a positive whiff test (the 'fishy' odor after the addition of 10 % KOH), clue cells (squamous epithelial cells covered with 75 – 100 % of Gram-variable short rods, curved rods, and cocci) visible under the microscope in quantity above 20 %. The presence of rare clue cells does not mean BV. However, these criteria cannot indicate the current infection unequivocally, and the interpretation of signs can be problematic, *e.g.* the pH reading errors may be caused by improper sampling (cervical mucous), menstruation, or recent douching [145].

The diagnosis of vaginal infections is challenging due to similar symptoms or even their absence and the shortcoming of the conventional diagnostic approach. The evaluation of Gram-stained smears is a subjective matter and requires good practice in preparing and

reading slides. The results depend primarily on the clinician's competence and experience, as well as the maintenance and settings of the microscope. The culture method is expensive and requires several days to obtain results. Current methods significantly hamper the diagnosis and monitoring of large-scale vaginal infections [142,146].

Several rapid tests were introduced to the market to facilitate the screening of infection, *e.g.* FemExam card or Candida Antigen test (Hydrex). The former is a colorimetric test that allows simultaneous determination of the elevated pH (4.7 or above) and detection of volatile amines. The latter is an immunochromatographic test for the direct and qualitative detection of *C. albicans*. BVBlue test (Gryphus Diagnostics, L.L.C.) defines the level of sialidase activity produced by BV-related pathogens. The undoubted advantage of these tests is that the final results can be obtained within half an hour. Similarly to the previously described methods, they are also based on visual assessment [142,147].

Recently, in an era of widespread exploitation of SERS in the medical field, scientists have demonstrated its great potential in the study of body fluids, cells, and tissues from the reproductive system of women. All these biological materials have already been analysed many times in various aspects, *e.g.* cancer diagnosis or forensic studies. The exfoliated cells of cervical tissues were examined to test the new way of cytology (spectro-cytology) for analysis of the healthy, pre-cancerous, and invasive stage of cervical cancer with a diagnostic accuracy of up to 94 % [148]. In turn, cervical fluids were analysed in terms of detecting human papillomavirus (HPV), which leads to an increased incidence of cervical cancer [149]. The possibility of tissue differentiation (HPV-positive and cervical cancer) was also attempted by searching their spectral fingerprinted response [150]. In forensic analysis, it is essential to effectively distinguish menstrual from peripheral blood, as they share a similar appearance due to the content of haemoglobin [151,152]. Equally important is the detection and identification of other bodily fluids, *e.g.* vaginal fluids [153–155]. However, as far as the state of science is concerned, there are no articles demonstrating the applicability of SERS in the diagnosis of vaginal infections. This dissertation tackles this problem for the first time, offering several studies that shed new light on the use of this technique as a diagnostic tool for vaginal infections.

1.7.3 COVID-19

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which leads to COVID-19, emerged in 2019 in China. It immediately spread worldwide, causing a pandemic in March 2020. Typical symptoms of COVID-19 are cough, fever, and other respiratory issues. The virus's transmission occurs upon close contact with an infected person *via* the droplet route [156].

In clinical practice, the diagnosis of COVID-19 is made based on the patient's symptoms. Results are confirmed with real-time polymerase chain reaction (RT-PCR) performed on nasopharyngeal or oropharyngeal swabs. RT-PCR is a sensitive method that is routinely used to detect most coronaviruses. Unfortunately, the analysis requires specialised and expensive equipment and highly skilled analysts. PCR-based methods are time-consuming (upwards of 4 – 8 hours, especially with the increasing number of potentially infected patients). Sometimes, it becomes necessary to send samples for testing hundreds of miles away to a properly equipped facility. This involves additional financial and time investment [157].

Whereas RT-PCR focuses on determining the presence of the virus, serological tests examine the immune response, *e.g.* immunoglobulins produced by the body when encountered with the virus. Tests based on IgM/IgG can play a key role in the fight against COVID-19 because they allow the classification of people who have developed an immune response due to SARS-CoV-2 infection. Serological tests are suitable for indirect diagnosis, broad scanning of human immunity, and keeping mortality statistics. They are also perfect options when access to molecular tests is limited or non-existent, allowing rapid identification of suspected cases. It is important to note that the stage and severity of the disease have a major impact on the final test result. Asymptomatic patients or those with few symptoms or in the early stage of the disease may have low levels of antibodies, which can lead to false-negative results [158,159].

In search of a new, more accurate method of COVID-19 diagnosis, scientists performed a series of studies on clinical samples with an integrated approach of SERS and chemometric methods or machine learning algorithms. Karunakaran *et al*. used the portable Raman device and AuNPs as SERS substrates to demonstrate different salivary responses of COVID-positive, COVID-recovered and healthy subjects. Studies performed over time proved that the composition of the saliva of patients who had COVID-19 normalized only

after a period of time [160]. Yang *et al.* distinguished between infected and non-infected nasopharyngeal swabs with SiO_2-coated silver nanorod arrays as a SERS substrate. The adopted SFNet algorithm was tailored to the specific spectral features inherent to SERS data, achieving test and blind test accuracy of 98.5 % and 99.04 %, respectively [161].

There are also other strategies that can be undertaken to recognize the disease. If the complex biological environment of analysed clinical samples suppresses the desired SERS signal of the target, magnetic biosensors can be used. An example of such a biosensor is a Fe_3O_4-Au nanocomposite modified with ACE2 to capture, enrich, and partially amplify Raman signals for SARS-CoV-2. The specially designed gold nanoneedle array enabled the virus detection. For the simulated throat and nasal swab, the SERS magnetic array provided the limit of detection of 100 copies per ml [162]. In other studies, the strategy was based on designing a silver nanorod array (AgNR) substrate with immobilized DNA probes to specifically capture RNA sequence of SARS-CoV-2 present in nasopharyngeal swab specimens. The most effective recurrent neural network models classified samples with an overall accuracy of 98.9 % [163]. An interesting idea was to design a mask with Au-TiO_2 system for the SERS label-free detection of the aerosolized SARS-COV-2 virus. This highly adsorptive chip with dense electromagnetic 'hot-spots' preconcentrated breath aerosols and effectively detected viruses [164].

The rapid and cost-effective method of COVID-19 diagnosis is fundamental to mitigate the current COVID-19 pandemic. Considering the recent performance of SERS, this technique is promising in achieving this goal. In this dissertation, the reader can become acquainted with another set of studies concerning SERS in the diagnostics of COVID-19 in a purely applied context.

Chapter 2
The objective of the doctoral dissertation

In medicine, making the proper diagnosis and thus implementing the right medication for effective treatment is of primary importance. The challenge of today's medicine is to develop a diagnostic approach that incorporates every crucial aspect, including the speed of analysis, reliability of the results, simplicity of performance, sensitivity, and accuracy. These methods should also be implementable in centers with fewer resources and difficult working conditions.

Over the past decades, surface-enhance Raman spectroscopy (SERS) has proved to be highly advantageous in biomedical aspects. This technique can evolve into a high-standard diagnostic method due to its fast analysis and simple reagent-free procedure, which reduces the risk of errors. However, what remains a challenge is undoubtedly the repeatability of the spectral signal closely related to the reproducibility of SERS substrates. Equally important are the optimisation and standardisation of the measurement conditions (power, laser line), the establishment of requirements for the method and storage of the tested material, and the preparation of the sample itself. The personal differences (patient-patient variability), which can result in additional spectral variances, should also be considered. This is especially important for the label-free approach, where the signal is the response of the sample as a whole. For the reliability of the results, it is beneficial to perform research on a large scale to cover the maximal spectral variance.

The research presented in this dissertation concerns the SERS analysis of microorganisms and clinical samples in the context of recognizing selected disease entities. The results of the SERS analysis performed in a label-free manner will be supported by chemometric methods. Although difficult due to the complexity of the signal, this strategy is preferable from a diagnostic point of view. It enables more frequent screening and monitoring of the state of infections because of the low cost, non-invasiveness, and lack of side effects. The direct approach would also facilitate diagnosis since the visual and sensor

examination is no longer needed. SERS substrates, which had previously been thoroughly optimised for the analysis of biological materials [165], have now been integrated with a handheld and portable Raman device. This adds value to the study, as it demonstrates the actual application of the method in point-of-care (POC) diagnostics.

The thesis of the doctoral dissertation is: the SERS analysis of clinical samples with the assistance of chemometric methods can be effectively used for the diagnosis of selected disease entities (*i.e.* gonorrhea, vaginal infections, and COVID-19). Within this thesis, there are several objectives:

1) establishing the procedure of clinical sample preparation and SERS measurement conditions;
2) determination of SERS signals of microorganisms (the chosen strains of *Neisseria* spp., and *Candida* spp., and also BV-related pathogens (*e.g.*, *Prevotella bivia*, *Gardnerella vaginalis*, *Mobiluncus mulieris*)) to create the catalog of reference SERS spectra;
3) determination and understanding of SERS signals of clinical samples considered as control (*i.e.* urethral discharge, vaginal fluid, saliva, nasopharyngeal swab);
4) determination and understanding of spectral differences caused by biochemical changes due to the ongoing disease (analysis within the disease entity, *i.e.* gonorrhea, vaginal infections, COVID-19). Within this aim, SERS is assisted by chemometric methods for the analysis of spectral data distribution (observation of correlations, similarities, differences) and determination of marker bands;
5) determination of the performance of chemometric methods in the spectral data analysis of clinical samples in terms of disease diagnosis. This step comprises establishing and optimising calibration models (for selected chemometric methods) – determining their robustness, and also predictive and classification abilities based on internal and external validation;
6) establishing the SERS response of a control group in relation to vaginal infections – considering the factors that alter the microflora and biochemical composition of secretion, which may result in spectral differences and deviation from the control group;
7) determination of the importance of the fertile/infertile phase of the menstrual cycle for the changes manifested on the SERS spectra of vaginal fluid;
8) comparison of the spectral effect associated with the fertile/infertile phase and vaginal infections.

Chapter 3
Materials and methods

3.1 Sexually transmitted diseases

3.1.1 The procedure of urethral discharge collection

Male, middle-aged patients were admitted to the Department of Dermatology, Infant Jesus Clinical Hospital (Warsaw, Poland) for medical appointments. The doctor initially interviewed patients regarding their health conditions and current symptoms. The analysed material (urethral discharge) was taken twice from each patient. The only indication for a proper test was that the patient should not urinate within four hours before the test. A sterile 10 μL inoculation loop was inserted into the man's urethra at 2 – 3 cm depth and gently rotated for 5 to 10 s. The collected samples were transferred into Eppendorf tubes®. Then, one of them was subjected to diagnostics in the medical office. At the same time, the other was immediately transported to the Institute of Physical Chemistry of the Polish Academy of Sciences (IPC PAS) for SERS analysis.

Depending on the diagnosis, two groups of clinical materials were distinguished:

a) Experimental group – 10 patients diagnosed with gonorrhea (referred to as Gon(+))
b) Control group – 10 patients tested negative for gonorrhea (referred to as Gon(−))

3.1.2 Conventional method for gonorrhea diagnosis

The diagnosis of gonorrhea was made based on microscopic images of Gram-stained urethral discharge. The analysed sample was placed onto a glass slide and, after drying, subjected to Gram staining in the following manner: *i)* staining with crystal violet (two minutes); *ii)* fixation with Lugol's fluid (two minutes); *iii)* decolorization with ethanol (30 seconds); *iv)* decolorization with fuchsin (two minutes). Then, the material was analysed under a microscope. The presence of Gram-negative diplococci inside leucocytes indicates gonorrhea.

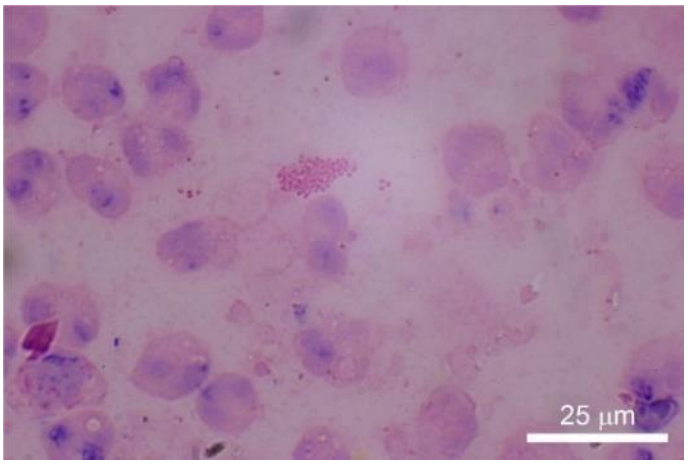

Figure 12 The microscopic image of Gram-stained urethral discharge taken from a patient diagnosed with gonorrhea.

Figure 12 presents the example of a photomicrograph of Gram-stained urethral discharge taken from a patient diagnosed with gonorrhea. The pink-stained cocci are recognized as Gram-negative *Neisseria gonorrhoeae* cells. Numerous pink-stained leucocytes can also be observed.

Microscopic analysis was performed with the upright microscope Delta Optical Evolution 100 Trino Plan LED (Delta Optical, Poland) equipped with ×10 eyepiece and ×100 objective lenses working on immersion (cedar immersion oil, Merck/Sigma Aldrich). The microscope was equipped with a camera port and Canon-compatible adapter. Canon 40D DSLR camera (Canon, Japan) was connected to a laptop to acquire images of analysed samples. Canon EOS Utility software recorded the images. RAW files were developed using ON1 Photo RAW software (ON1, Portland, OR, USA).

Figure 13 The colonies of *Neisseria gonorrhoeae* cultured from male urethral discharge [166].

The culture method was used to support the diagnosis. For this purpose, the collected material was applied onto a specific medium for *N. gonorrhoeae* (Chocolate agar + PolyviteX™ vcAT3 agar (bioMérieux)) and cultivated at 37 °C in the 5 % CO_2 condition for 48 h. The colonies were identified using oxidase and API tests (bioMérieux). Figure 13 presents the colonies of *N. gonorrhoeae* cultured from male urethral discharge.

3.1.3 SERS analysis of urethral discharge

To recognize the infection, the SERS analysis was performed in two variants:

a) Based on the identified cultured microorganisms – the recognition of infection is based on unique spectral patterns that microorganisms exhibit. Hence, the colonies cultured from the analysed sample are identified by comparison with the reference spectra of selected *Neisseria* spp. (the catalog of SERS spectra).

b) Based on the clinical samples – the infection alters the microflora and biochemical composition of secretion, leading to spectral changes in the analysed sample. In this strategy, the urethral discharge is mixed with two μL of saline solution (0.9 % NaCl solution), and then one μL is placed onto the prepared previously SERS substrate.

3.2 Vaginal infections

3.2.1 The procedure of vaginal fluid collection

a) The analysis of vaginal fluids – dysbiosis of natural microflora

The study was conducted on a cohort of 60 patients aged 18 – 69 years. The collection of clinical samples and the patient interview were performed in the Department of Dermatology, Infant Jesus Clinical Hospital (Warsaw, Poland). Samples were collected twice from each patient using sterile 10 μL inoculation loops and then transferred into Eppendorf tubes®. One was subjected to diagnostics according to standard procedure, and the other was immediately passed to the IPC PAS for SERS testing. The detailed characterisation of samples can be found in Table 1A, Table 2A, and Table 3A (appendix).

b) The analysis of the fertile/infertile phase of the menstrual cycle

The study was performed on two menstrual cycles of one woman belonging to the control class. The first cycle lasted 27 days, and the second 24 days. Every morning, samples were taken two, three, or four times, resulting in a total of 103 vaginal fluids to be analysed. Samples were collected using sterile 10 μL inoculation loops, transferred into Eppendorf tubes®, and then immediately measured. Each material was analysed sensorically for quantity, color, and consistency to determine the ovulation and phases of the menstrual cycle. A standard calendar method was also used to determine the fertile and infertile phase.

3.2.2 Conventional diagnostic methods of vaginal infections

The standard diagnostic method for vaginal infections included microscopic analysis of Gram-stained vaginal discharge and pH determination. In the case of the presence of blastospores and pseudohyphae during the microscopic analysis or by special order of the doctor, a culture of *Candida* spp. was performed.

The analysed vaginal fluid was placed onto a glass slide and, after drying, subjected to Gram staining in the following manner: *i)* staining with crystal violet (two minutes); *ii)* fixation with Lugol's fluid (two minutes); *iii)* decolorization with ethanol (30 seconds); *iv)* decolorization with fuchsin (two minutes). Then, the material underwent microscopic observations to assess the presence of clue cells and different microorganisms such as Gram-negative and Gram-positive bacteria, pseudohyphae, and blastospores of *Candida* spp. The samples were observed at ×100 objective lenses working on immersion using the upright microscope Delta Optical Evolution 100 Trino Plan LED (Delta Optical, Poland)). The Gram-stained vaginal fluid was evaluated using the I – IV purity scale described in Table 1.

Table 1 The evaluation scale for assessing vaginal fluid purity.

Purity Scale	Characterisation
I	pH=4.4; lactobacilli (many); 0 – 4 leucocytes
II	pH=4.7; 4 – 10 leucocytes with possibilities of slight inflammation, not eligible for treatment
III	pH=5.0 – 6.0 and higher; lactobacilli (lack) and 10 – 20 leucocytes, bacilli Gram-positive/Gram-negative or cocci Gram-positive (many) or 10 – 20 leucocytes or 0 – 4 leucocytes, but the presence of clues cells (indicative of *Gardnerella vaginalis*)
IV	pH=4.7; lactobacilli (many) and blastopores and pseudohyphae or pH=5 and higher; lactobacilli (lack), cocci Gram-negative inside leucocytes (indicative of *Neisseria gonorrhoeae)*

Figure 14 presents the selected photomicrographs of Gram-stained vaginal fluids taken from patients with different vaginal infections: vulvovaginal candidiasis (VVC), bacterial vaginosis (BV), Coinfection (BV and VVC), and gonorrhea. VVC (Figure 14A) is characterized by darkly stained yeast cells (reproduced in a process known as budding). Throughout the field of view, numerous Gram-positive bacilli could also be seen. Gonorrhea (Figure 14B) is recognized by stained pink *N. gonorrhoeae* cells located at epithelial cells. In the microscopic image of BV (Figure 14C), one can observe squamous epithelial cells

covered with Gram-variable small rods forming clue cells. The photomicrograph of coinfection (Figure 14D) reveals the BV pathogens that coexist with *Candida* spp. blastospores. For comparison, (Figure 14E) presents the vaginal discharge of a woman with balanced microflora. One can notice a pinkish, squamous epithelial cell with clearly defined and smooth edges covered with *Lactobacillus* spp. bacteria.

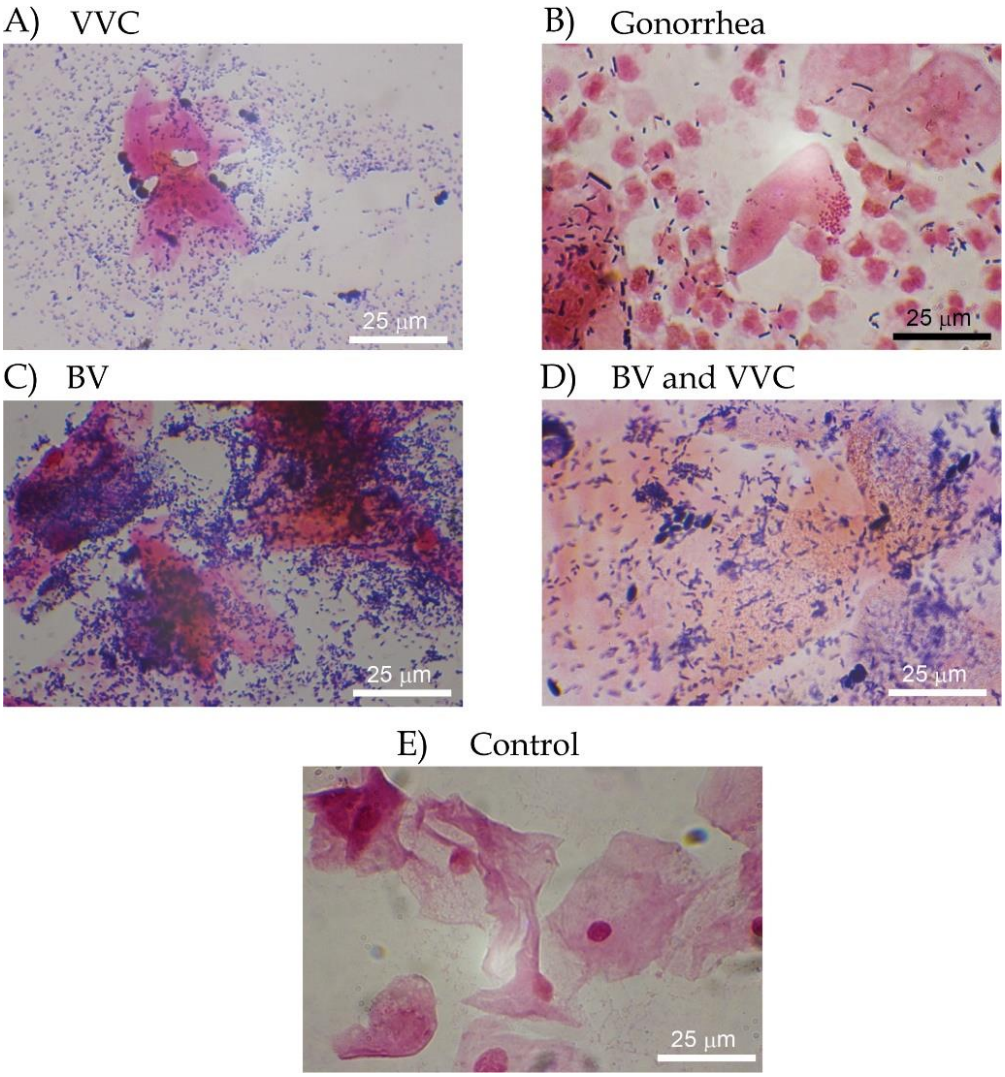

Figure 14 The microscopic images of Gram-stained vaginal fluids taken from patients with A) vulvovaginal candidiasis (VVC); B) gonorrhea; C) bacterial vaginosis (BV); D) coinfection (BV and VVC). For comparison, the microscopic image E) is a control one.

The pH of vaginal fluids was determined with pH litmus paper (Duotest pH 3.5 – 6.8 Macherey – Nagel GmbH&Co. KG.). The culture method for the presence of *Candida* spp. was performed with ChromID Candida (bioMérieux), which provides the direct identification of *Candida albicans*.

3.3 COVID-19

3.3.1 The collection of saliva and nasopharyngeal swabs

Patients with suspected COVID-19 were admitted to the Department of Clinical Genetics, Medical University of Łódź (Łódź, Poland) for medical appointments. Saliva samples were collected and transported to deep-well plates. Nasopharyngeal swabs were taken with sterile swabs and placed in vials with one ml of saline solution (0.9 % NaCl solution). After completion of the diagnostic tests, samples were immediately transferred to a freezer and stored at −80 °C. All samples, well-labeled and packaged in large batches, were transported with all safety precautions to the IPC PAS for SERS analysis.

3.3.2 Standard method for COVID-19 diagnosis

The standard method used for the COVID-19 diagnosis was quantitative real-time polymerase chain reaction (qRT-PCR). The procedure was as follows. In the first step, the chemagic™ 360 automated extraction platform (PerkinElmer, Waltham, MA, USA) was used to extract SARS-CoV-2 RNAs from 300 μL of a sample. The extraction was performed according to the manufacturer's instructions (the chemagic Viral DNA/RNA 300 Kit H96 (PerkinElmer, USA)). Viral RNA was eluted with 80 μL elution buffer and used for RT-PCR assay. The presence of SARS-CoV-2 was detected by qRT-PCR amplification of SARS-CoV-2 open reading frame 1ab (ORF1ab), nucleocapsid protein (NP) gene fragments, and a positive reference gene using DiaPlesQ^{TM} Novel Coronavirus (2019-nCoV) Detection Kit (SolGent CO, Ltd., Daejeon, Republic of Korea). The conditions for amplifications were 50 °C for 15 minutes (reverse transcription), 95 °C for 15 minutes (initial PCR activation), followed by 45 cycles of 95 °C for 20 seconds (denaturation) and 60 °C for 40 seconds (annealing/extension). The result was considered valid only when the reference gene's cycle threshold (Ct) value was lower or equal to 38. The result was considered positive when Ct values of both target genes were lower or equal to 38 and negative when they were both larger than 38. If only one of the target genes had Ct value lower or equal to 38 and the other larger than 38, it was interpreted as inconclusive.

3.4 The sample preparation and SERS measurements

3.4.1 Clinical samples

a) Urethral discharge and vaginal fluids

To perform SERS measurements, one μL of the analysed body fluid (diluted in two μL of saline solution) was applied onto the SERS substrate and allowed to dry slowly in a laminar flow chamber. For every sample, *ca.* 30 single spectra were recorded (every spectrum – three accumulations with time 6000 ms of a single accumulation).

b) Saliva and nasopharyngeal swabs

Prior to SERS measurements, the analysed samples were removed from the freezer in batches (a given deep-well plate with saliva or vials with nasopharyngeal swabs). After thawing at room temperature, 1.5 μL of every sample was applied on the SERS platform and left to dry slowly in a laminar flow chamber. For every sample, *ca.* 15 single spectra were recorded (every spectrum – three accumulations with time 6000 ms of a single accumulation).

3.4.2 Cultured microorganisms

Research presented in this dissertation includes the spectral analysis of different strains of microorganisms. Table 2 compiles a list of them with their origin and growth conditions.

Table 2 The set of analysed microorganisms.

Type of microorganism	Source	Medium and culture conditions
Neisseria gonorrhoeae ATCC 70825/FA 1090 *Neisseria sicca* ATCC 9913 *Neisseria meningitidis* ATCC 13102 *Neisseria lactamica* ATCC 23790	American Type Culture Collection (ATCC), National Medicines Institute, (Warsaw, Poland)	Medium: BD chocolate agar (GC II Agar with IsoVitaleX) Conditions: 48 hours at 37 °C in 5 % CO_2
Prevotella bivia 9557T *Fanyhessea vaginae* 38953T *Mobiluncus mulieris* 20071T *Mobiluncus curtisii* 21018T *Gardnerella vaginalis* 3717T *Finegoldia magna* 17636T *Aerococcus tetradius* 46590T *Anaerococcus christensenii* 28831T	Culture Collection of the University of Gothenburg (CCUG)	Medium: De Man-Rogosa-Sharpe (MRS) Agar Conditions: 48 hours at 37 °C in anaerobic conditions

Type of microorganism	Source	Medium and culture conditions
Candida glabrata *Candida albicans* SN148 *Candida albicans* dHP17	Institute of Biochemistry and Biophysics (IBB) (Warsaw, Poland)	Medium: Yeast extract Peptone Dextrose (YPD) Conditions: 24 hours at 37 °C in aerobic conditions

After the cultivation, the selected colonies were collected *via* inoculation loop and transferred to separate Eppendorf tubes® containing 100 μL of 0.9 % NaCl solution. Next, the analysed sample was vortexed for 10 seconds and centrifuged for three minutes at 1070×*g*. The supernatant was removed, and a new portion of saline solution was added. The bacteria purification was repeated three times, and the final pellet was resuspended in 10 μL of saline solution and appropriately mixed. For SERS analysis, one μL of such solution was placed onto the SERS substrate and left to dry for a few minutes. SERS measurements were performed with the following parameters: three accumulations with a single accumulation time of 6000 ms and 30 single measurements for each sample.

3.4.3 Chemical compounds

The following compounds were subjected to SERS analysis: D-glucose 99 % (Cambridge Isotope Laboratories, Inc.); Citric acid, BioUltra, anhydrous, >99.5 % (Sigma Aldrich); Lactic acid, lithium salt, 99 % extra pure (Thermo scientific); Acetic Acid Optima LC/MS (Fisher Chemical); Ferritin from equine spleen (Type I saline solution, Sigma Aldrich); L-tryptophan ≥99.5 % (NT) BioUltra (Sigma Aldrich); L-tyrosin ≥99 % (Merck); L-phenylalanine ≥98 % (Sigma Aldrich); L-methionine ≥99.5 % (NT) BioUltra (Sigma Aldrich); α-amylase from human saliva (Type XIII-A, lyophilized powder, Sigma Aldrich); Lysozyme from chicken egg white (lyophilized powder, Sigma Aldrich); Albumin from chicken egg white, lyophilized powder ≥98 % (Sigma Aldrich); Urea BioReagent ≥98 % (Sigma Aldrich).

SERS spectra were recorded for the liquid form of the chemical compounds. In the case of powder form, *ca.* 0.6 M aqueous solutions were prepared. For each tested chemical compound, 2.5 μL of the solution was applied on the SERS substrate, left to dry, and then analysed (three accumulations with time of a single accumulation 6000 ms, *ca.* 15 single measurements for each sample).

3.5 SERS setup

For the SERS investigations presented in this thesis, Bruker's BRAVO spectrometer equipped with Duo LASER™ (700 – 1100 nm) and a CCD camera was used. This is the handheld, portable Raman device with a built-in background cutoff function, 2 – 4 cm^{-1} spectral resolution, and 100 mW power. The experimental setup (Figure 15A) for SERS measurements consists of the Raman spectrometer, holder for SERS substrates, and laptop with OPUS software (Bruker Optic GmbH, 2012 version, Leipzig, Germany). Figure 15B presents a batch of SERS substrates after the process of laser ablation. Figure 15C shows a SERS substrate (with a layer of SERS-active metal) attached to a glass slide, ready for a sample to be applied and analysed.

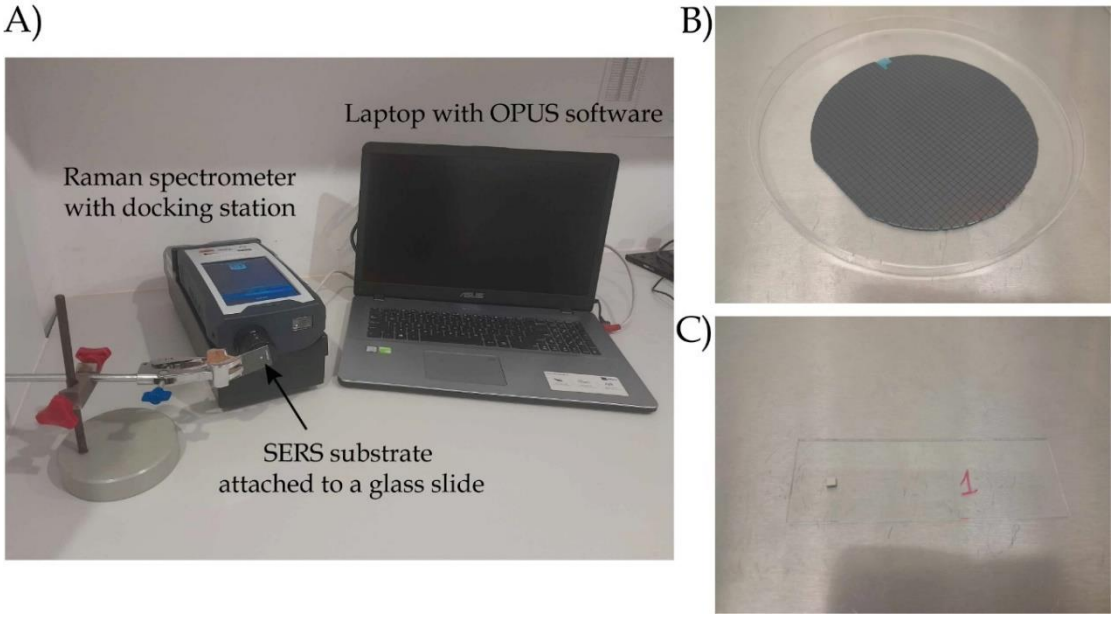

Figure 15 A) The experimental setup for SERS measurements, B) one batch of SERS substrates, and C) SERS substrate attached to a glass slide ready for the sample to be analysed.

The post-processing of spectra (with the OPUS software) was done in the following order: *i)* baseline correction (concave rubber band correction, number of iterations 6, number of baseline points 6); *ii)* smoothing (number of smoothing points 5); *iii)* normalization (min-max normalization, the whole range). Then, all the processed data were analysed using various chemometric methods implemented in the software Unscrambler (CAMO Software AS, version 10.3, Oslo, Norway).

SERS platforms

SERS platforms were silicon wafers roughened in the process of laser ablation. This process consists of three main steps *i)* cutting the silicon wafer into 3 × 3 mm squares corresponding to the SERS platforms' desired dimensions; *ii)* modification with the laser beam to roughen it at the nano- and micro-levels; *iii)* deposition of SERS-active metal – the optimal layer was 100 nm Ag sputtered in the process of physical vapour deposition (PVD magnetron sputtering device (Quorum, Q150T ES, Laughton, UK)).

Figure 16A presents the results of scanning electron microscopy (SEM) analysis for the Si material after laser ablation. Figure 16B demonstrates the base material with a 100 nm Ag layer (Si/Ag SERS platform) exposing the formed nanoaggregates of 80 nm ± 20 nm, which fit perfectly into the range of resonance amplification [165]. Figure 16C presents the *Prevotella bivia* cells located on the SERS platform [167].

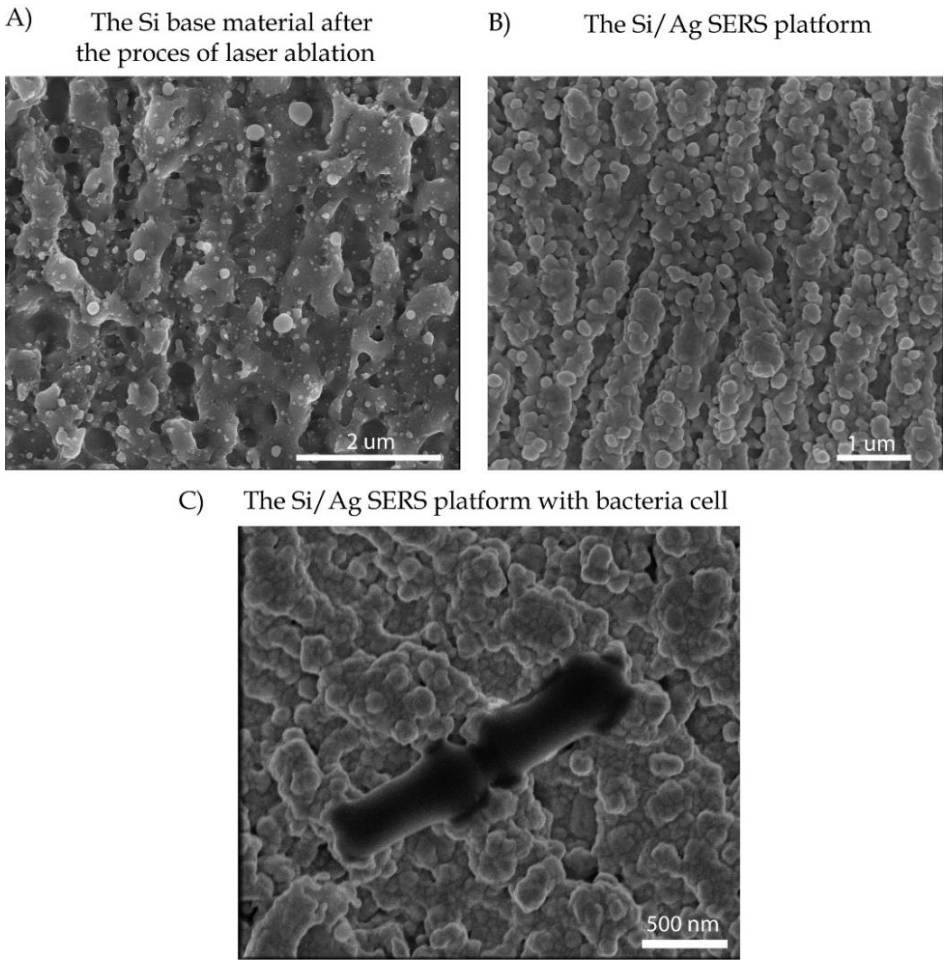

Figure 16 The SEM images of A) Si base material after the process of laser ablation and B) with a layer of 100 nm Ag forming Si/Ag SERS platforms ready to use, and C) *Prevotella bivia* cells on the SERS platform [165,167].

Chapter 4
Results and discussion

4.1 Sexually transmitted diseases

This section is devoted to the spectral analysis of selected microorganisms and urethral discharge for the detection and identification of *N. gonorrhoeae* in clinical samples. SERS measurements will be assisted with chemometric methods operated in supervised and unsupervised mode to extract the proper spectral information.

Firstly, selected strains of *Neisseria* spp. will be spectrally investigated to create the catalog of reference spectra. The empirical analysis of obtained data allows the initial insight into the spectral patterns and recognition of general differences between them. The subsequent application of chemometric methods significantly facilitates the analysis, revealing their subtle differences and specific correlations within the data analysed. Moreover, this catalog will be used to identify microorganisms cultured from a clinical sample, taking advantage of the unique spectral image they exhibit.

In the next step, the urethral discharge will be characterized to find the spectroscopic image of gonorrhea. In this strategy, information about the presence of *N. gonorrhoeae* cells is trapped in a specific SERS signal of the analysed clinical sample. The PCA analysis performed for the considered classes, namely gonorrhea-positive (Gon(+)) and gonorrhea-negative (Gon(−)), will help to extract the desired spectral information and present their dependencies in the PCs dimension. The supervised methods (SIMCA, PLD-DA) will be used to create and optimize the calibration models. Then, their performance will be tested at the stage of external validation – the recreation of actual diagnostic conditions.

4.1.1 SERS identification of bacteria in clinical samples

The etiological agent of gonorrhea – *Neisseria gonorrhoeae* is genomically, morphologically, and phenotypically closely related to other pathogenic and commensal *Neisseria* species, *e.g. Neisseria meningitidis*, *Neisseria lactamica*, and *Neisseria sicca*. *Neisseria meningitidis* is naturally carried as a commensal in the (naso)pharynx (10 to 15 % of the general population) but occasionally can cause fatal septicaemia and meningitis [131]. Although these bacteria mainly reside in the pharynx, their presence in the reproductive tract has also been reported [168–171].

The first step in identifying bacteria from a clinical sample is to create a catalog of reference SERS spectra ('spectral fingerprints') for selected microorganisms. In the considered case, the catalog comprises the above-mentioned *Neisseria* spp. (Figure 17).

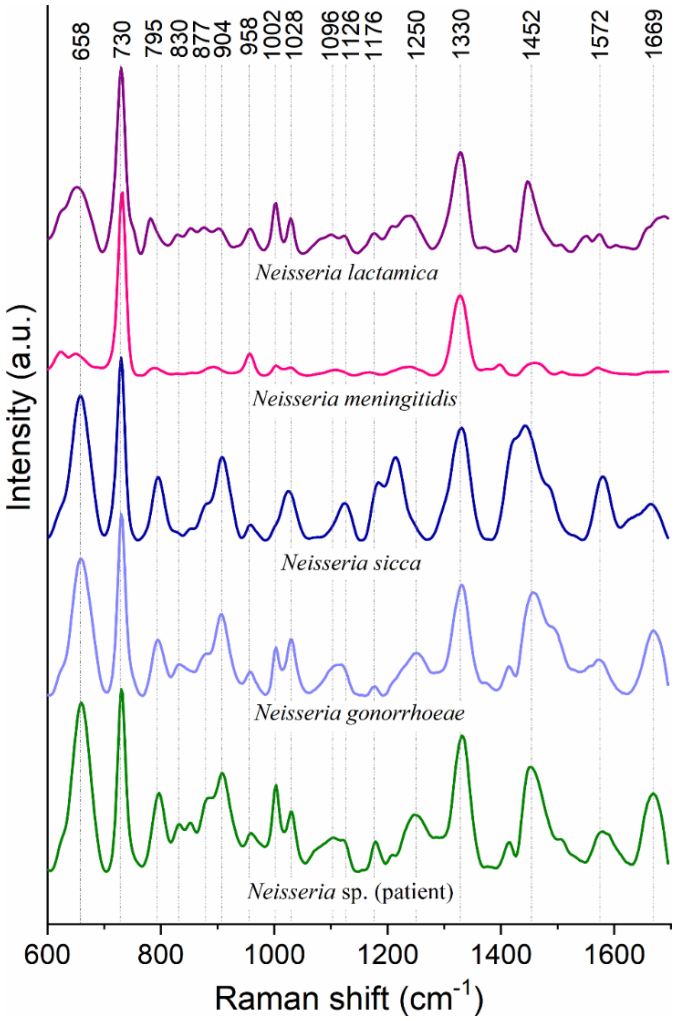

Figure 17 The reference SERS spectra of *Neisseria* spp. strains (*Neisseria lactamica, Neisseria meningitidis, Neisseria sicca, Neisseria gonorrhoeae*) and unknown strain isolated from the patient's urethra. The spectra were averaged from *ca.* 30 single spectra.

The SERS signal of microorganisms is formulated on the basis of molecular oscillations in the cell wall and cell membrane, more specifically, the peptidoglycan layer components, *N*-acetylmuramic acid, *N*-acetyl-D-glucosamine, proteins, lipids, lipopolysaccharides, polysaccharides, phospholipids. It was proved that the SERS spectrum of bacteria can also be conditioned by metabolic secretions of cells, *e.g.* metabolites of purine degradation (adenine, hypoxanthine, xanthine, uric acid, guanine, adenosine monophosphate) that occur due to the stress that bacteria experience when suspended in a nutrient-free environment. These secreted compounds appear mostly at the outer layer of bacterial cells and in the extracellular metabolome as a result of nucleotide degradation [172–175].

The SERS spectra of all *Neisseria* spp. are characterized by the intense band at 730 cm^{-1} arising from vibrations in adenine-related compounds (FAD, NAD, ATP, DNA) [176]. Additional features of great importance are 1330 cm^{-1} (CH_3CH_2 wagging in purine bases and/or NH_2 stretching in adenine, polyadenine) and 1452 cm^{-1} (CH_2 vibrations in lipids, proteins, and CH_2 scissoring of phospholipids). The presence of proteins is manifested by amide I and amide III signals located at 1250 and 1669 cm^{-1} [177–179]. The detailed band assignments are set in Table 3.

All tested strains exhibit a vast majority of bands at the same frequency (*e.g.*, 795, 904, 958, 1002, 1028, 1126, and 1572 cm^{-1}). Despite many spectral commonalities, these microorganisms show differences in the shape and intensity of the bands. Thus, the formulated unique spectral images reserved only for them provide an excellent reflection of their molecular external structure. *N. sicca* can be recognized with the band at 1002 cm^{-1}, appearing as a shoulder of 1028 cm^{-1}. In turn, *N. meningitidis* and *N. lactamica* reveal weak bands at 904 and 1572 cm^{-1} compared to *N. gonorrhoeae* and *N. sicca*. The SERS spectrum of *N. meningitidis* contains low-intensity bands in the entire range, which can be a distinctive feature. In conclusion, the spectral fingerprints of each *Neisseria* spp. display certain characteristic features that are fundamental for taxonomic affiliation.

Table 3 The tentative assignments of bands on the SERS spectra of *Neisseria* spp. bacteria [177–179]

Raman Shifts (cm^{-1})	**Tentative assignment**
658	C-S stretching, C-C twisting of proteins (tyrosine), COO^- deformation in amino acids, guanine and thymine (ring breathing modes)
730	Adenine-related compounds (FAD, NAD, ATP, DNA)
795	Ring breathing mode of cytosine, uracil, thymine, (O-P-O) symmetric stretching of nucleic acid
830	Stretching (O-P-O) of DNA, deformative vibrations of amine groups, tyrosine
877	C-C-N^+ symmetric stretching (lipids), C-O-C ring (carbohydrate)
904	C-C skeletal stretching, tyrosine
958	C=C deformation, C-N stretching, C-O stretching, CH_3 symmetric stretching of proteins (α-helix)
1002	Phenylalanine, C-C aromatic ring stretching
1028	C-N stretching, C-C stretching (phospholipids, carbohydrates), C-H in-plane bending mode of phenylalanine
1096	PO_2^- symmetric stretching, C-O-C stretching modes in polysaccharides, adenine, polyadenine, ν(C-C), ν(C-O), phospholipids
1126	=C-O-C= (unsaturated fatty acids in lipids), C-O-C stretching modes in nucleic acids, PO_2^- stretching in nucleic acid, C-O and C-C stretching in carbohydrates, C-N stretching in proteins
1176	C-H in-plane bending mode of tyrosine, ν(C=C), δ(COH) (lipid)
1250	Amide III (arising from the coupling of C-N stretching and N-H bending)
1330	CH_3CH_2 wagging in purine bases, NH_2 stretching in adenine and polyadenine, phospholipids
1452	CH_2 deformation (protein, lipids), CH_2 bending mode of protein and lipids, C-H vibrations proteins and lipids, CH_2 wagging, CH_2/CH_3 deformation, CH_2 bending and scissoring of phospholipids
1572	A, G (ring breathing modes of DNA and RNA bases) C=C bending mode (phenylalanine), guanine, adenine, tryptophan
1669	Amide I of proteins (β-sheet) ν(C=C) *cis*, lipids, fatty acids Carbonyl stretch (C=O)

The SERS measurements were complemented by PCA analysis to visualize the dependencies between analysed data and reveal the prominent variables. The results of PCA analysis performed for the considered association (reference strains of *Neisseria* spp. *vs.* unknown strain) are presented in Figure 18.

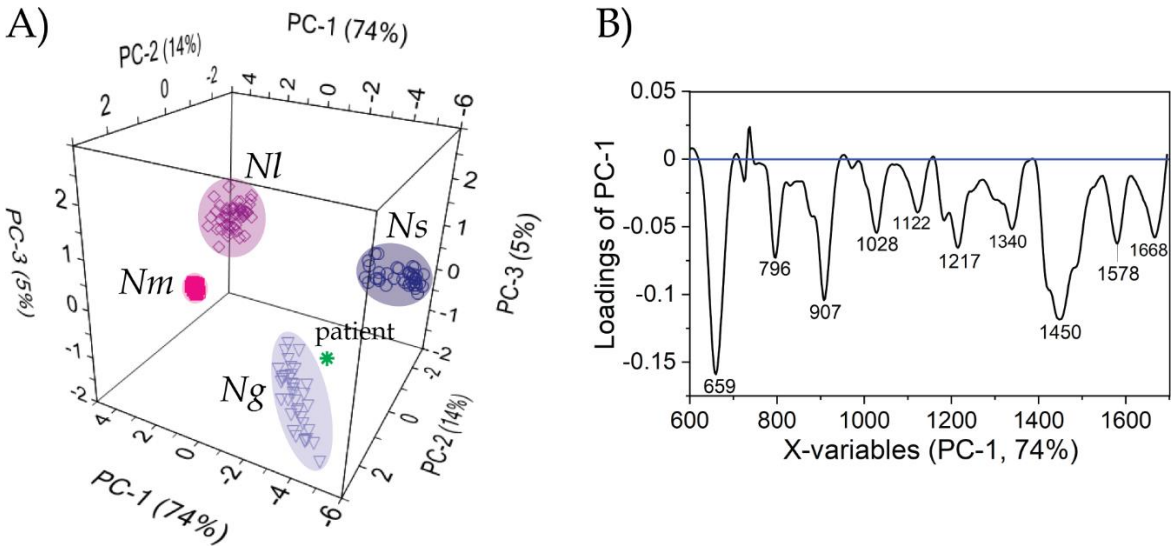

Figure 18 The results of PCA analysis calculated for the association (reference bacteria *vs.* unknown strain) in the form of A) 3D score plot (where *Ng* (*Neisseria gonorrhoeae*), *Nm* (*Neisseria meningitidis*), *Nl* (*Neisseria lactamica*), and *Ns* (*Neisseria sicca*)) and B) loading plot of PC-1.

In a new dimension (score plot), all analysed data are presented as points where every single point refers to one spectrum (Figure 18A). The spectra of the bacteria serving as reference are arranged in groups. The spectrum of the strain to be recognized is marked with a green asterisk. The plot score indicates that the first principal component (PC-1) explains 74 % of the spectral information and, together with the two subsequent components (PC-2 and PC-3) – 93 %. Moreover, PC-1 designates the boundary between *N. meningitidis* and *N. lactamica* against *N. gonorrhoeae* and *N. sicca*. The green asterisk representing the unknown strain is in close proximity to *N. gonorrhoeae*, directly indicating its affiliation to this group and thus recognition as *N. gonorrhoeae*. To confirm the veracity of the classification obtained, this strain was identified by genetic analysis – sequencing of the 16S rRNA gene. The consistency of results proved the reliability of the SERS-PCA approach for the microorganism's identification.

The loading plot is defined along a specific principal component and reveals the variables that affect the group separation with varying degrees of intensity. The significance of variables follows the principle – the greater the weight, the greater the influence.

The corresponding loading plot for this association (Figure 18B) reveals a significant impact of vibrations in proteins, lipids, and phospholipids with 659, 907, and 1450 cm^{-1} signals. Other variables at 796, 1028, 1217, 1578, and 1668 cm^{-1} indicate a prominent contribution of nucleic acids, phenylalanine, lipids, fatty acids, and also proteins due to the vibrations of amide I [178].

4.1.2 Comparison of methods for the identification of *Neisseria gonorrhoeae*

Most currently used tests for bacteria identification (Gram staining, oxidase test, catalase test, carbohydrate dehydrogenation) are based on visual assessments of observed effects or cultivated colonies. A number of chemical reagents are also required. Gram staining is a multistep method where the thickness of the swab and the time of culture significantly affect the result of the analysis [180,181]. The analysis can also be affected by the condition of reagents and the staining technique itself. In turn, the carbohydrate degradation test is time-consuming (4 hours), and false negative results may appear due to autolyzed organisms from a culture older than 24 hours [133]. Similarly to the SERS method, MALDI-TOF MS (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) requires a small sample volume for analysis. Both methods identify the tested strain according to a simple procedure and in a relatively short time. Unfortunately, the use of different matrices and their effect on the quality and reproducibility of results are the weakest points of MALDI-TOF MS [182,183]. The evident benefit of the SERS technique is the zero-reagent procedure, which significantly lowers expenses and minimises the probability of mistakes (Table 4) [184].

Table 4 The comparison of SERS and commonly used methods for the identification of *Neisseria gonorrhoeae* [133,180–183,185–188].

	SERS	**Presumptive identification methods**			**Confirmatory methods**	
		Gram staining	**Oxidase test**	**Catalase test**	**Carbohydrate degradation**	**MALDI TOF MS**
Idea	Identification at species and strain level	Cell wall type (Gram-positive, Gram-negative)	Biochemical reaction for the presence of cytochrome	Biochemical reaction for the presence of catalase	Biochemical reaction for identification of *Neisseria* spp.	Identification of bacteria at species and strain level
Time of analysis	15 min	15 min	2 min	few seconds	4 h	several dozen minutes
Materials and reagents	SERS platforms, saline solution (0.9 % NaCl solution)	crystal violet, iodine solution, 95 % ethanol, safranin	tetramethyl-*p*-phenylene-diamine hydrochloride	3 % or 30 % hydrogen peroxide	20 % glucose solution, maltose, lactose, sucrose, buffered balance salt	ethanol, acetonitrile, trifluoroacetic acid, different matrices: α-cyano-4-hydroksycinnamic acid, sinapinic acid, ferulic acid
Visual assessment	No	Change in cell color to pinkish-red	Change in cell color to dark-purple	Positive reaction with bubbling	For glucose: change in the colour of the solution from red to yellow	No
Advantages	Small amount of biological material required; simple procedure	Provides valuable information on the shape and size of the cells	Simple procedure with a single reagent; does not require expensive equipment	Simple procedure with a single reagent; does not require expensive equipment	Does not require expensive equipment	Small amount of biological material required; fast and easy measurement protocol
Limitations	Reproducibility of SERS measurements; the extended spectral library required	Colonies can be difficult to interpret; large number of steps and reagents	Can be performed only on fresh colonies (18 – 24 h); only platinum or inert transfer loops can be used	False positive results in the case of colony that was grown on blood or chocolate medium	Can be performed only on fresh colonies (18 – 24 h); Both false positive and false negative reactions are possible due to sample contamination and the presence of autolyzed organisms from culture older than 24h	Effects of matrix on quality and reproducibility; the extended spectral library required

4.1.3 Spectral pattern of Gon(+) and Gon(−) urethral discharge

Urethral discharge is a complex biochemical matrix abundant in various components. In the case of a control subject, such secretion contains leucocytes (neutrophilic granulocytes) and different epithelial cells (pseudostratified columnar epithelium, stratified squamous epithelium). The epithelial cells are mainly built of glycolipids, phospholipids, cholesterol, and various surface and transmembrane proteins. There are also urea, uric acid, creatinine, ammonia, and mineral salts (chlorides, phosphates, carbonates). Due to the frequent contact with semen, this secretion can be enriched with compounds such as albumins, putrescine, spermine, spermidine, cadaverine, lipids, prostaglandins, steroid

hormones, globulins, fructose, vitamins B_{12} and C, citric acid, acid phosphatase, fibrinolysis, hyaluronidase, mucus [189–191]. In the case of ongoing gonorrhea, the urethral discharge also contains *N. gonorrhoeae* cells. A recent study has shown that for asymptomatic disease, the gonococcal load is estimated at 2×10^5 copies per swab, and for symptomatic at 3.7×10^6 copies per swab [192].

The analysis presented in this section was performed on a total of 20 samples of urethral discharge taken from 10 patients diagnosed with gonorrhea (referred to as Gon(+)) and control ones (referred to as Gon(−)). The spectra averaged within a single group (Figure 19) provide the general biochemical description.

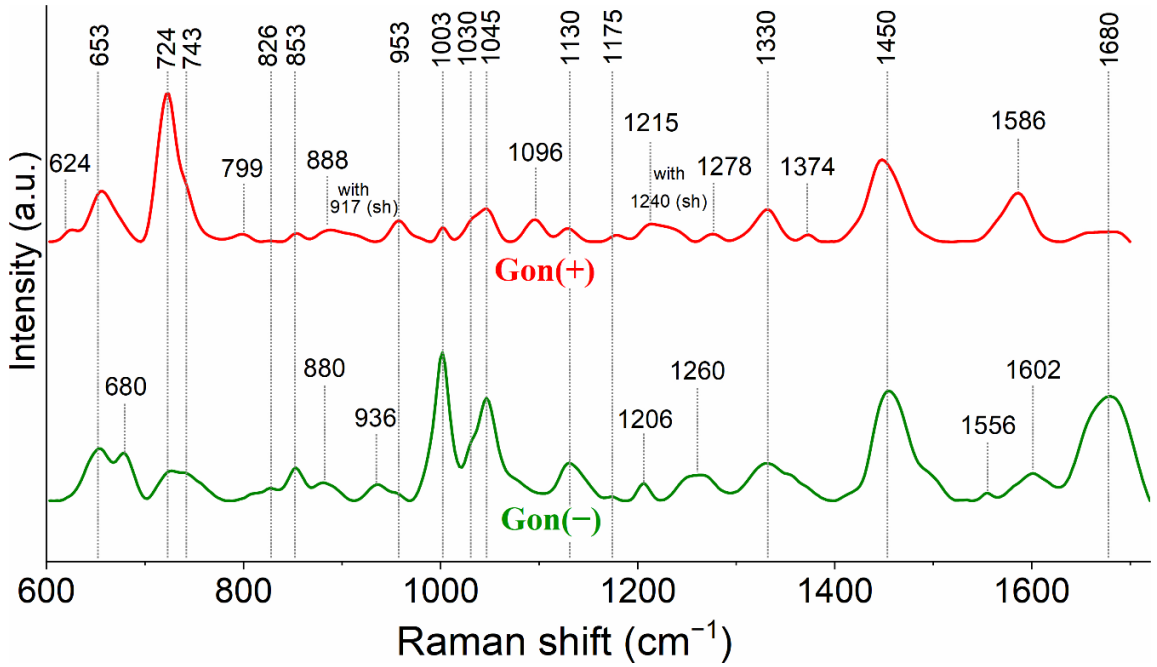

Figure 19 The averaged SERS spectra of Gon(+) and Gon(−) urethral discharge. Spectra were averaged from 10 Gon(+) and 10 Gon(−) samples *i.e.* 358 Gon(+) and 311 Gon(−) single spectra. The abbreviation (sh) stands for shoulder.

The SERS profiles of Gon(−) and Gon(+) specimens are distinct across the whole analysed range (Figure 19). The spectral differences are associated with the presence of bands (*e.g.*, 680 cm^{-1} for Gon(−) and 1096 cm^{-1} for Gon(+)) as well as their relative intensities (*e.g.*, 1003, 1045 cm^{-1}). Moreover, shifting and changes in the shape of the bands can also be observed. For Gon(−), 953 cm^{-1} is a shoulder to 936 cm^{-1}, while for Gon(+), the same band is narrow and separated. In turn, the low intensity 880 cm^{-1} band (for Gon(−)) is shifted to 888 cm^{-1} with 917 cm^{-1} shoulder for Gon(+). All of the alterations result from the changes at the molecular level of secretion due to the ongoing disease.

The Gon(−) spectrum is completely dominated by 1003, 1045, 1450, and 1680 cm^{-1} features. According to the specific biochemistry of the analysed secretions, these bands arise due to vibrations in phenylalanine and/or urea, spermidine trihydrochloride and/or proline, albumin and/or creatinine, and cholesterol and/or amide I, respectively. The spectral manifestation of tyrosine, guanine, and tryptophan is the band located at 1330 cm^{-1} [193,194]. For more information regarding band assignments, please see Table 5.

The Gon(+) spectrum reveals unique bands located at 724, 799, 953, 1096, 1215 with 1240, and 1374 cm^{-1}. These bands highly correspond to the SERS fingerprints of bacteria (Figure 17). Hence, they can be used as an excellent indicator of the disease. Notably, the most intense 724 cm^{-1} has already been proven to be a marker band of microorganisms [176]. Considering the overall SERS spectra and the biochemistry they represent, it can be concluded that the spectral contribution of urethral components is reduced in favor of bacteria cells in the case of Gon(+) samples.

Table 5 The tentative assignment of the most prominent bands observed on the SERS spectra of control urethral discharge [72,193–201].

Raman Shifts (cm^{-1})	**Tentative assignments**
653	Guanine, D-glucose, lactose, O=C-N deformation in uric acid, C-C twisting of tyrosine
680	Cysteine (C-S) stretching, G (ring breathing modes in DNA bases)
724	C-H stretching vibration in uric acid and/or hypoxanthine, C-N head group choline $(H_3C)_3N^+$ in lipids, ring breathing mode of Adenine (RNA/DNA), O-O stretching vibration in oxygenated proteins, ring breathing mode of tryptophan (protein assignment), coenzyme A
743	(CNC)s stretching, CH_2 rocking, NH_2 wagging, thymine
826	ν_{as}(O-P-O) of nucleic acids, phosphodiester, glucose, ring breathing mode of tyrosine (proteins), proline, hydroxyproline
853	ν(COC) of glycogen, ring breathing mode of tyrosine (proteins)
880	Proline, valine, glycine, tryptophan, glutamate, ν(C-C) hydroxyproline, $\nu_s P(OH)_2$ of phosphate
936	Spermine tetrahydrochloride, C-C stretching in α-helix, proline, and valine in proteins, glycogen
953 (sh)	Tyrosine in proteins, PO_4^{3-} symmetric stretching of spermine phosphate hexahydrate, ν_{sym} (CH_3) of proteins (α-helix), amino acids, polysaccharides
1003	Aromatic ring breathing in phenylalanine, C-N stretching in urea

1030 (sh)	CH in plane bending mode of phenylalanine in proteins, C-N stretching in proteins, CH_2CH_3 bending modes of collagen and phospholipids
1045	spermidine trihydrochloride, proline, C-O and C-N symmetric stretching in proteins, ν(C-O) & ν(C-C) of Glycogen
1130	C–C stretching in lipids (cholesterol, triglycerides) and fatty acids, C–N stretching mode in proteins, ν(C-O-C) in nucleic acids, $\nu(PO_2^-)$ in nucleic acid, C-O and C-C stretching in carbohydrates (*e.g.*, glycogen), CNC asymmetric stretching in spermine and/or spermidine, uric acid, ascorbic acid, L-serine,
1175	Cytosine, guanine, CH in plane bending mode of tyrosine in proteins
1206	A, T (vibration of DNA/RNA), tryptophan & phenylalanine $\nu(C\text{-}C_6H_5)$ mode, phenylalanine (protein assignment), C-C stretching in tyrosine
1260	Amide III in proteins (α-helices), collagen =C-H deformation in phospholipids, glucose
1330	Ring stretching in tyrosine, CH_3CH_2 deformation of collagen in proteins and lipids, Ring mode of Adenine and Guanine of nucleic acids, NH_2 stretching in adenine and polyadenine, (COH) & (CH) bending of glycogen, tryptophan, proteins (*e.g.*, albumin)
1450	$\delta(CH_2)$ in proteins, CH bending of lipids, Glucose, (CH_2) scissoring and (COH) bending of glycogen, CH_2, CH_3 bending of tryptophan, C-H stretching of glycoproteins including mucins, hydrocarbon chain of lipids, bending CH_2/CH_3 of triglycerides and fatty acids, albumin, creatinine,
1556	amide II in proteins, ν(C=C) tryptophan
1602	Spermine and/or spermidine, C=C in-plane bending of phenylalanine & tyrosine, proteins (*e.g.*, albumin)
1680	Amide I (β turn, random structure), cholesterol in lipids, spermidine phosphate hexahydrate

For the studied group of patients (Gon(−) and Gon(+)), the amount of obtained single spectra (*ca.* 30 spectra per one sample) was 311 and 358, respectively. Data presented in this configuration (Figure 20A and B) accurately illustrate spectral diversity between patients within the same class; the higher is for Gon(−) class. The superposition of unnormalized averaged spectra (Figure 20C) allows for the comparison of absolute intensities. The Gon(+) samples are characterised by higher spectral enhancement due to the presence of bacterial cells multiplied as a result of the ongoing infection.

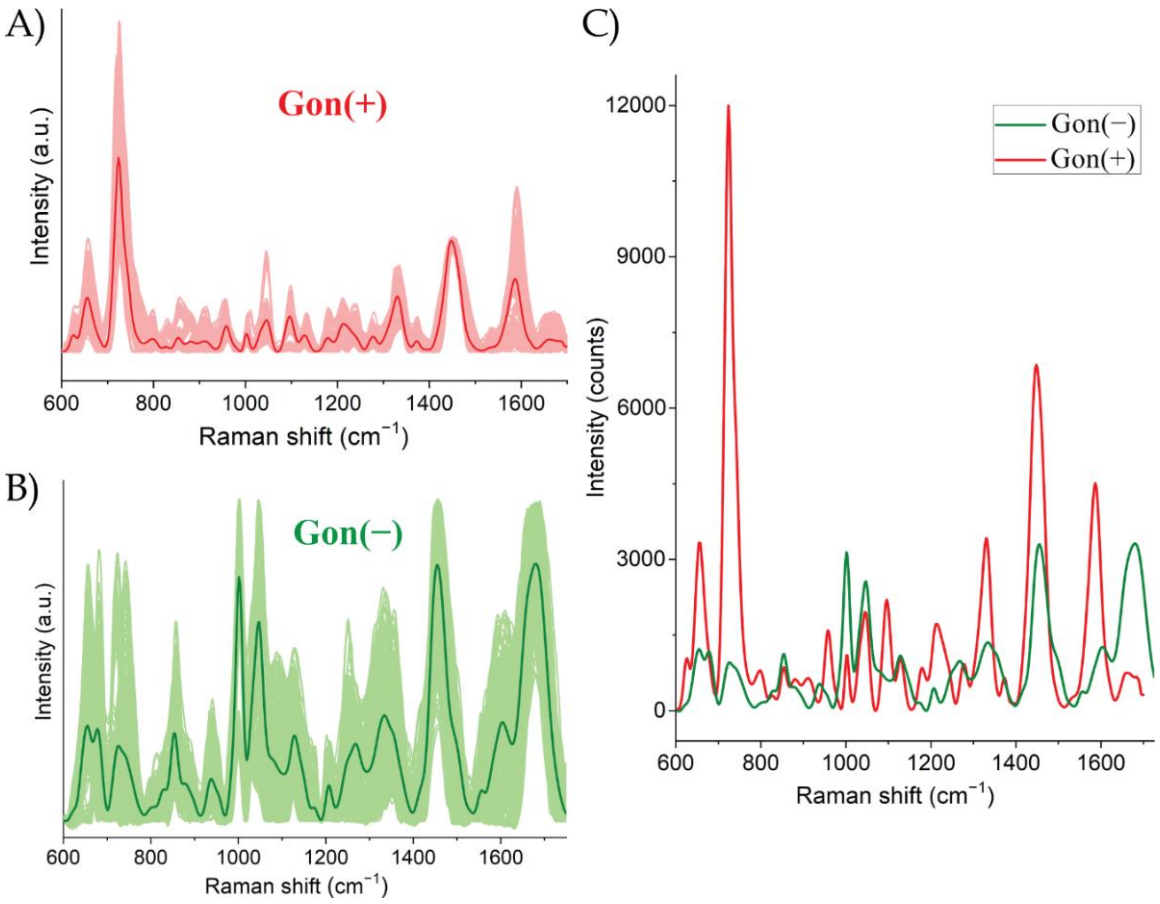

Figure 20 The SERS spectra of Gon(+) and Gon(−) urethral discharge presented as a superposition of all gathered spectra of A) 358 Gon(+) and B) 311 Gon(−) and also C) superimposed and unnormalized averaged SERS spectra.

4.1.4 Chemometric analysis of the spectra of Gon(+) and Gon(−) urethral discharge

1. The distinction between Gon(+) and Gon(−) classes

PCA analysis was performed for the association of 10 Gon(+) samples (358 single spectra) *vs.* 10 Gon(−) samples (311 single spectra) to make the distinction between classes (Figure 21). The two analysed groups are well separated, with the boundary defined by the PC-1 axis providing 49 % of the explained information. The three subsequent components (PC-1, PC-2, and PC-3) are responsible for 69 % of the spectral information. The Gon(−) class is characterized by a significantly higher distribution of spectral data than the Gon(+) class (Figure 21A and Figure 21B). This is probably related to inter-patient variability, where the composition of secretions is strongly influenced by diet, age, health condition, and other individual factors. Since pathogenic bacteria are not present in the Gon(−) secretion, the personal effects can be more visible. This is in contrast to the Gon(+) class, where the high

spectral contribution of proliferated bacterial cells makes the SERS signal more uniform across the group.

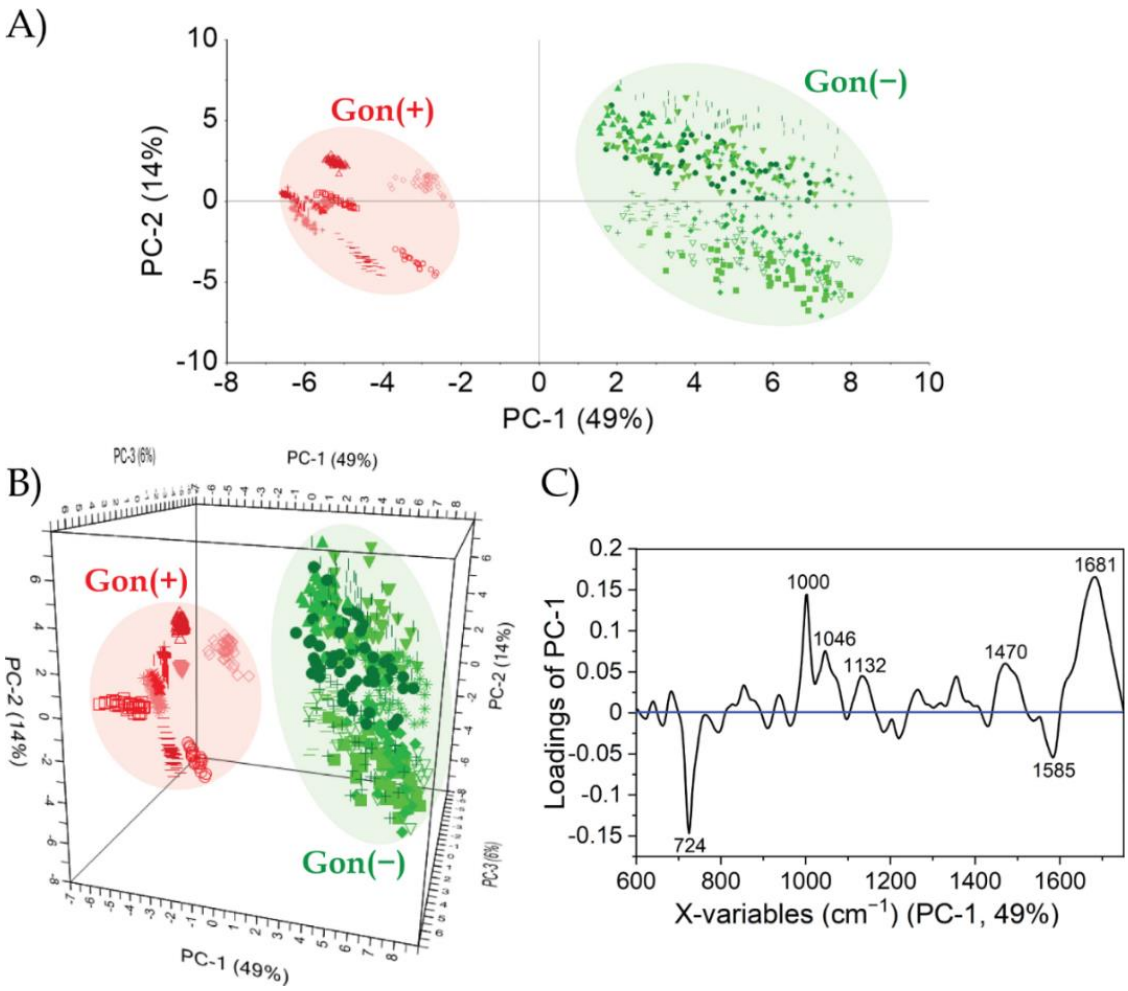

Figure 21 The results of PCA analysis of urethral discharge in the form of A) 2D score plot, (B) 3D score plot, and (C) corresponding loading plot for the PC-1. The calculation was performed for the association – 10 Gon(+) samples (358 single spectra) *vs.* 10 Gon(−) samples (311 single spectra).

The differences between classes reflected by the loading plot unveil variables with a significant impact for diagnostic purposes (Figure 21C). Those located at 1000, 1046, 1132, 1470, and 1681 cm^{-1} indicate the presence of biomolecules in Gon(−) secretion, *e.g.* proteins and lipids. Those at 724 and 1585 cm^{-1} indicate the pathological origin.

2. The recognition of gonorrhea based on SERS spectra of urethral discharge

Supervised chemometric methods (PLS-DA and SIMCA) were employed to improve the gonorrhea diagnostic. This two-step analysis consisted of *i)* the creation of a calibration model involving 549 spectra (8 Gon(−) and 8 Gon(+) samples) known as training set data;

ii) validation on 120 spectra (2 Gon(−) and 2 Gon(+) samples)) (test set data). To compare the classification abilities of the two considered methods, the analysis was performed for the same association involving the same data set.

In PLS-DA analysis, the calibration model was prepared with Kernel algorithm. The method of validation was cross-validation (random with 20 segments). The model performance was described by figures of merit (FOM) and reached high values of R^2_{cal} (0.96) and R^2_{cv} (0.96). The values of root mean square error of calibration (RMSEC) and cross-validation (RMSECV) were 0.09 and 0.1, respectively. The first latent variable (LV1), explaining 94 % of the variance in block Y with 49 % of the spectral data (X matrix), presents a distinct class separation (Figure 22).

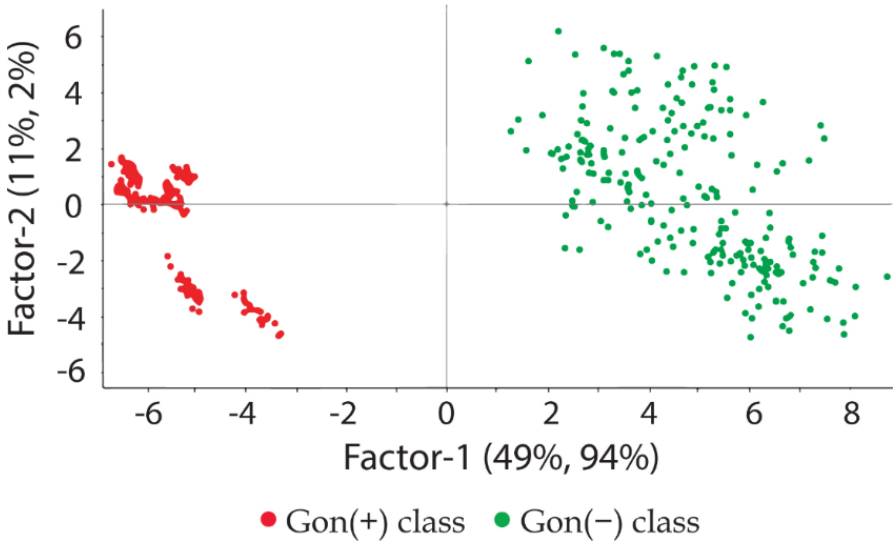

Figure 22 The 2D score plot of PLS-DA analysis performed for 8 Gon(+) (red dots) and 8 Gon(−) (green dots) samples created a calibration model (training set). The amount of spectral data was 549.

In the second step, the evaluated PLS-DA model was used to classify unknown clinical samples (four external samples, each consisting of 30 single spectra). The model exhibits impressive predictive abilities because the predictive values were close to 1 or 0, depending on the sample type. Standard deviations in the range of 0.04 – 0.31 demonstrate the model's robustness, which provides correct classification. The 100 % accuracy results from the proper recognition of all spectral data.

In SIMCA analysis, the analysed classes were independently evaluated with PCA prior to the subsequent validation (test set). In the analysed association, the Gon(+) and Gon(−) classes are described by two and three PCs, respectively. The significance level defining

classification limit equals 5 %. The 9.3 distance between these two models indicates an excellent distinction between classes. The Cooman's plot presents the distance of each tested spectrum to Gon(+) and Gon(−) PCA models (Figure 23A). In this analysis, 9 of 60 external spectra representing Gon(−) samples were rejected (upper right quadrant), and 4 of 60 Gon(+) spectra have membership of both classes (lower left quadrant). As a result of the correct classification of 107 samples, the accuracy reaches 89 %. The variable mainly responsible for the separation is located at 1681 cm^{-1} with a discrimination power of 7.04. Those located at 724, 907, 996, 1045, 1183, and 1220 cm^{-1}, reaching even 5.7, are also remarkable in the total distinction (Figure 23B).

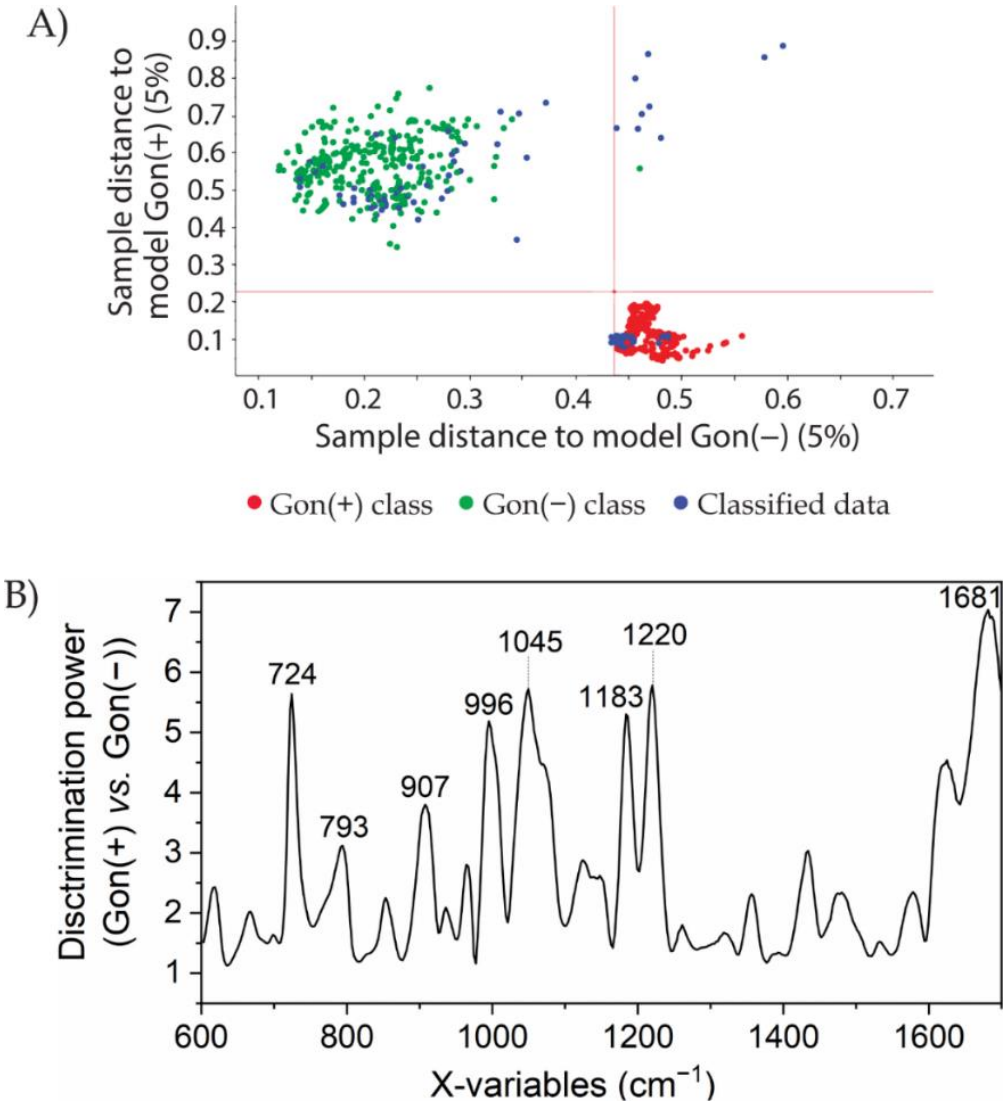

Figure 23 The results of SIMCA analysis in the form of A) Cooman's plot and B) discrimination power for 8 Gon(+) (red dots) and 8 Gon(−) (green dots) samples and test set data to be classified (blue dots). The amount of spectral data used for calibration (training set) was 549, and for validation (test set) 120.

4.1.5 Comparison of diagnostic methods for gonorrhea

The results indicated that SERS, supported by an appropriate chemometric method, can rapidly recognise the presence of *N. gonorrhoeae* in clinical samples. Due to the prominent advantages, this approach can facilitate the diagnosis of gonorrhea. Table 6 compares SERS and other techniques used in clinical practice under selected aspects.

Table 6 The comparison of SERS, Gram staining, NAATs, and culture method for STDs diagnosis [166,202–204].

	SERS	**Gram staining**	**NAATs**	**Culture method**
Idea	Spectral fingerprinting of clinical sample	The microscopic observation of Gram-stained clinical sample	Amplification of nucleic acids	The cultivation of particular bacteria, *e.g. Neisseria gonorrhoeae*, in appropriate conditions
Time of analysis	15 min	15 min	3 h	1 – 3 days
Visual assessments	No	Yes	No	Yes
Chemical reagents	No	Yes	Yes	Yes
Steps of analysis	Sample preparation; SERS measurements; Chemometric analysis	Distribution of secretion on a glass slide and fixation by heat; Staining with reagents; Microscopic analysis	Nucleic acid extraction; Amplification; Detection	The preparation of a suitable medium and the provision of suitable growth conditions
Advantages	Fast and label-free sample preparation; A small amount of analyte is required for the analysis	Valid diagnostic tool, especially in settings with more modest resources	Do not require the organism to be viable for detection; Screening for gonorrhea in asymptomatic patients	The ability to preserve microorganisms for additional studies, *e.g.* antimicrobial susceptibility
Limitations	Reproducibility of SERS platforms; Instrument failure; The library of SERS spectra that cover the high spectral variability is required for building a robust calibration model	Very low sensitivity in the case of asymptomatic patients and in the case of throat and rectal swabs (method not recommended); Other bacteria, especially *Neisseria* spp., which have similar morphology, affect the microscopic result in extragenital samples	Cross-reactivity with others *Neisseria* spp.; Inhibitory substances may lead to false negative results; Carryover contamination; Reagent and instrument failure; Not suitable for POC testing; Cost-prohibitive in developing countries	Sensitivity may be low in the case of long distance between the clinic and the laboratory; The methodology is intolerant of delays or inadequacies; The medium and growth conditions have to be tailored to the specific microorganisms; Gonococci are very demanding pathogens and do not tolerate dehydration, and they should be inoculated immediately after swabbing; Other methods should be used for the identification of the obtained colonies

The culture method is time-consuming, and the medium and growth conditions have to be tailored to the specific microorganism under analysis. As gonococci are very demanding pathogens, they should be immediately inoculated on a culture medium after swabbing. Gram staining requires access to chemical reagents and an experienced investigator to reliably determine the presence of pathogenic *N. gonorrhoeae* diplococci. Despite these limitations, microscopic imaging of Gram-stained samples is still a suitable

diagnostic tool in facilities with more modest resources [166]. NAATs belong to the most widely used and reliable tests for STDs diagnosis because of their relatively high sensitivity [202]. SERS and NAATS require three major steps. However, the reagent-based NAAT procedure can lead to significant errors related to carryover contamination or cross-reactivity with other *Neisseria* spp. [203]. As SERS spectrally images the actual composition of the tested material without reagent support, the possible contamination-related errors are considerably lower. Moreover, the reduced time of analysis (from 3 hours to 15 minutes) and simple procedure make SERS highly competitive with NAATs.

Point-of-care (POC) tests are beneficial in circumventing the timeliness of traditional diagnostics. They also allow patients to receive results and treatment before leaving the doctor's office. The diagnosis on the first day of presentation is a key aspect of managing treatment appropriately, thus limiting the widespread administration of broad-spectrum antibiotics. These are the reasons why POC testing is vital in medicine [205]. The WHO actively supports the development of new POC methods for STDs and, to this end, has established guidelines for developers and users, known as ASSURED, that have an actual utility. The first letters of ASSURED denote the characteristics of the ideal POC method (Table 7) [206].

Table 7 ASSURED criteria for an ideal point-of-care test for STD diagnosis.

ASSURED CRITERIA	
***A**ffordable*	by those at risk for infection
***S**ensitive*	very few false negatives
***S**pecific*	very few false positives
***U**ser-friendly*	very simple to perform (minimal steps required with minimal training)
***R**apid and Robust*	to enable treatment at first visit (rapid) and does not require refrigerated storage (robust)
***E**quipment-free*	easily collected noninvasive specimens and not require complex equipment
***D**eliverable*	to those who need them

The priority was to create simple, affordable, and sensitive POC tests for STDs diagnosis. According to these guidelines, diagnosis of gonorrhea using SERS and chemometric methods implemented in a label-free manner fits perfectly with POC testing standards. The entire SERS-based diagnostic set-up would consist of Raman spectrometer, a laptop with software, and SERS platforms. As the research presented, SERS analysis can be successfully performed with a compact, handheld spectrometer that can be easily located

in the doctor's office. No storage requirements exist because diagnostics can be performed immediately after sampling. Due to the low cost of analysis, this method is suitable for screening tests. It is these POC tests that are preferred by end-users (patients) who wish to receive a definitive diagnosis and implement treatment immediately [207].

4.1.6 Conclusions

This study indicated that selected strains of *Neisseria* spp. revealed distinct SERS signals that correlate closely with their external structure. These results provide a good foundation for bacteria identification based on the SERS spectra operating on a 'fingerprint principle'. In the trial test, the bacteria strain was cultivated from a urethral discharge of a patient with suspected gonorrhea. The correct results confirmed the reliability of the SERS-PCA method in the identification of microorganisms based on the previously created spectral catalog of *Neisseria* spp.

It was also demonstrated that the SERS signal of the urethral discharge has different characteristics depending on the presence of *N. gonorrhoeae* cells. The Gon(+) signal contains unique bands located at 724, 799, 953, 1096, 1215 with 1240 and 1374 cm^{-1} that highly indicate the spectral contribution of pathogenic microorganisms. An important aspect is the amplification of the SERS signal itself; the greater was noted for Gon(+). In PCA analysis, the PC-1, which explains the highest spectral variability, differs between the two analysed groups, allocating Gon(+) and Gon(−) on opposite sides in the score plot dimension.

Supervised chemometric methods were used to recognize 120 external spectra based on the established calibration models. In the case of PLS-DA, the accuracy was 100 %. SIMCA provides two more classification options – the analysed sample can get a dual membership, or it can be rejected. For the analysed association, SIMCA recognized correctly 107 out of 120 spectra, resulting in 89 % accuracy.

4.2 Vaginal infections

This chapter is devoted to the spectral investigation of selected *Candida* spp., bacteria related to BV infection, and vaginal secretions. Analysis of such body fluids is vital from the point of view of widespread vaginal infections and diseases that affect women's health and well-being. The vagina represents a moist biotope that contains a diverse microbiome changing through the stage of life. It is estimated that 1 – 4 ml of vaginal fluid contains up to 10^9 bacterial cells protecting the mucous membrane from infections and maintaining the physiological balance of the genital tract. The vaginal microbiome is composed of mycobiome (*Ascomycota*, *Saccharomycetales*), virome (*Papillomaviridae*, *Herpesviridae*), archeome (*Methanobrevibacter smithii),* and bacteriome with *Lactobacillus* spp. as dominate ones [208]. During the course of life, the microflora can change for various reasons, such as age, menstrual cycle, pregnancy, antibiotic treatment, or hygiene. The surrounding microbiome (*e.g.*, intestinal) can be the cause of the flora changes as well.

The main objective of this section is the spectral analysis of vaginal fluids to extract the spectral manifestation of pathogens that cause vaginal infections. Due to the complexity of the topic, the study has been enriched with additional issues, creating a comprehensive analysis of spectral patterns that reflect distinct biochemical compositions of analyzed samples. As was mentioned before, the body fluid is highly susceptible to change under the influence of different factors, *e.g.* contraception, vaginal irrigation, or dysbiosis. In the first instance, it is essential to establish the SERS spectrum of a control (vaginal fluid) sample and identify the parameters that influence the spectral response of an analysed sample. Hence, each sample was precisely described in several aspects, and any comments and observations during the diagnosis were noted. A separate subsection will be devoted to the analysis of the fertile/infertile phase based on two menstrual cycles. Such an established spectral signal with defined parameters served as a reference and will be used for further considerations – band assignments and the vaginal infections analysis. The bands will be identified experimentally and also based on the literature review. In the final step, SERS will be supported with chemometric methods for the analysis of VVC infection – the group differentiation and identification of diagnostic marker bands.

4.2.1 Spectral characterisation of microorganisms

Figure 24 presents the averaged SERS spectra of the common pathogens that are associated with BV infection. All these spectral fingerprints reveal many similar bands in the entire range, with the most intense at 730, 1325, and 1459 cm^{-1}. The main spectral variations are particularly concerned with mutual differences in intensity ratios of the individual bands (*e.g.*, 649, 923, and 1459 cm^{-1}). One can empirically notice that *M. curtisii* is recognized by the intense bands in the region 850 – 955 cm^{-1} and the one located at 649 cm^{-1}. In turn, *F. magna* and *G. vaginalis* can be distinguished from the rest of the studied microorganisms by weak bands over the entire spectral range.

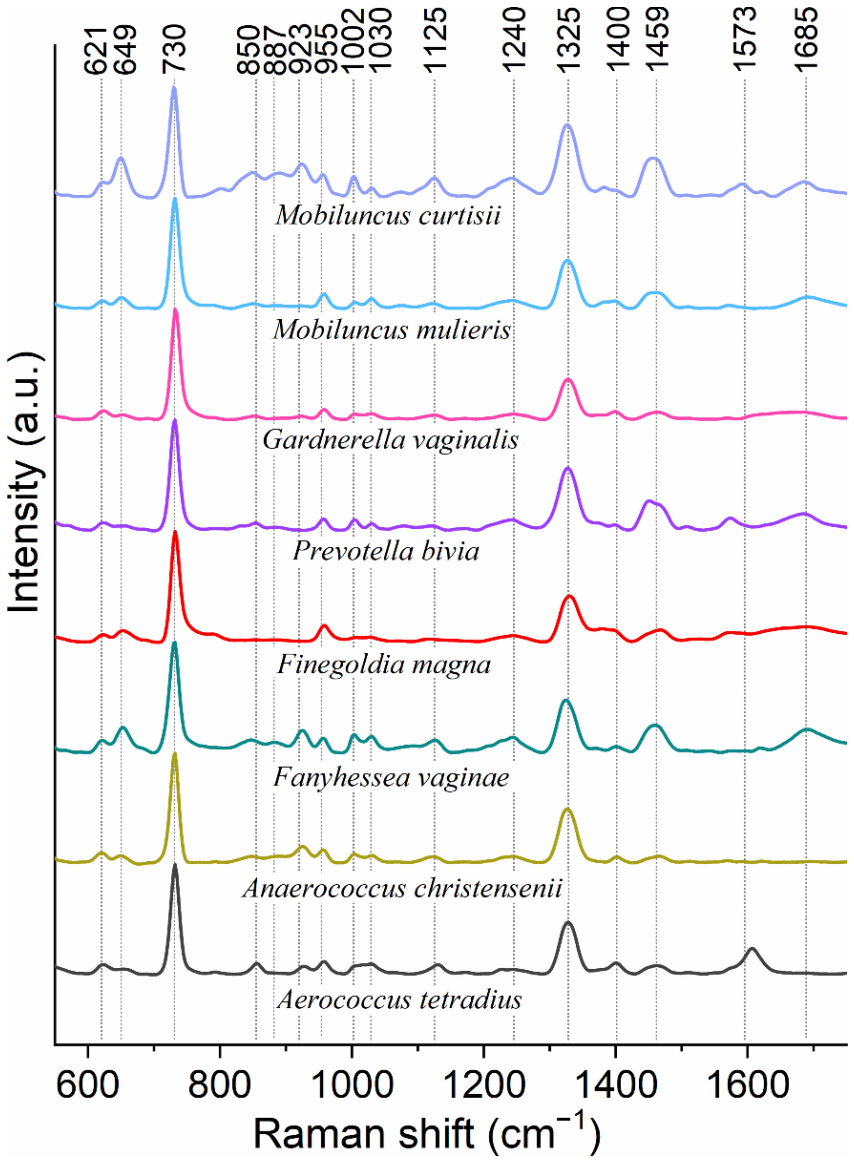

Figure 24 The averaged SERS spectra of common pathogenic bacteria present in the vagina. Bacteria were cultivated on MRS agar in anaerobic conditions at 37 °C and for 48 h. For every sample, *ca.* 30 single spectra were recorded.

Figure 25 presents the averaged SERS spectra of chosen *Candida* spp. Their spectral signals contain 725 cm^{-1} and also 1096 cm^{-1}, and 1126 cm^{-1}, which can be assigned to branched chains of 1,3-β-glucan and 1,3-β-D-glucan and/or *N*-acetyloglucosamine, respectively. These compounds constitute building blocks of chitin and glucagon that create most of the cell wall [209,210]. Other bands indicating the presence of glucagon and chitin in the fungal structure are located at 905, 957, 1209, 1244, 1325, 1370, 1459 cm^{-1}. Mannoproteins can be manifested by those at 1244, 1325, 1459, 1592, and 1688 cm^{-1}. The band at 652 cm^{-1} occurs due to C-S stretching and C-C twisting of proteins and/or COO^- deformation in amino acids [211].

A key aspect is that every *Candida* spp. reveals bands at 836, 905, 1209, 1370 cm^{-1} with a significant intensity that allows for their distinction from the analyzed bacteria. Bacteria are characterized by a lack of these bands or their negligible intensity.

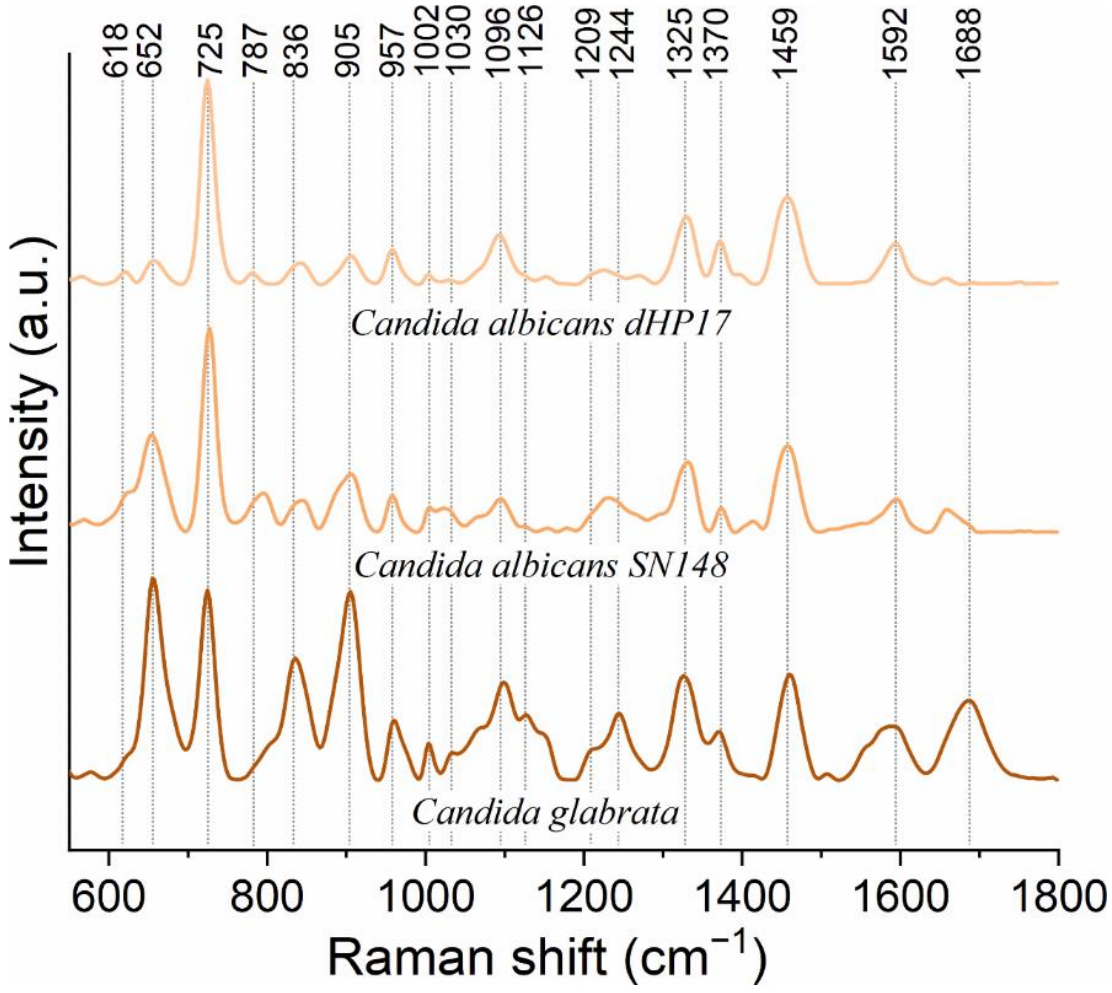

Figure 25 The averaged SERS spectra of chosen *Candida* spp. that can colonise vagina. These strains were cultivated on YPD medium in aerobic conditions at 37 °C and for 24 hours. For every sample, *ca.* 30 single spectra were recorded.

4.2.2 Variability of SERS signal of vaginal fluid

Based on the diagnosis and the interview with patients, the collected samples of vaginal fluids (characteristics in Table 1A, Table 2A, and Table 3A – appendix section) were divided into the categories (Table 8).

Table 8 The categories of analysed vaginal fluids.

Category	Characteristics
I	4.2 – 4.7 pH; I or II purity; many lactobacilli and epithelial cells
II	4.7 – 5.6 pH; II/III or III purity; Gram-positive and Gram-negative bacilli or Gram-positive cocci and few or lack lactobacilli; usually 10 – 20 leucocytes
III	vaginal irrigation (1 sample) loss of normal microflora (1 sample) contraception (vaginal ring with hormones) (1 sample)
IV	fertile/infertile phase of the menstrual cycle (a woman belonging to control class – 103 vagina fluid samples)
V	vaginal infections – vulvovaginal candidiasis (VVC), bacterial vaginosis (BV), gonorrhea, and coinfection (VVC and BV)

The I category comprises samples characterized by normal acidic pH and I purity. The microflora consists of many lactobacilli. Epithelial cells are also prominent. Therefore, this group is considered a control one in the conventional diagnostic methods, such as Gram staining. The II category of samples is characterized by disturbed microflora and elevated pH resulting from the proliferation of pathogenic bacteria. On the purity scale, samples are rated as II or II/III and do not classify for treatment (this is not vaginal infection). The III category contains samples taken from women with loss of normal microflora associated with age-related processes and those who performed vaginal irrigation and used contraception before medical appointments. The IV category of samples refers to changes in the fertile and infertile phase of the menstrual cycle (one woman belonging to the control class, 103 samples of vaginal fluids). The V category includes samples taken from women diagnosed with vaginal infections such as vulvovaginal candidiasis (VVC), bacterial vaginosis (BV), gonorrhea, and coinfection (VVC and BV).

The analysis of the mentioned categories is essential to determine whether the spectral changes between categories are significant to treat category I as a control class in terms of the SERS analysis. The spectral analysis of all aspects is provided in the sections below.

a) The spectral analysis of vaginal fluids of the I category

The I category comprises of 21 vaginal fluid samples. The averaged SERS spectra are set in Figure 26. The consistent characteristic within this category (acidic pH of 4.2 – 4.7 and I/II purity) leads to the fact that they all share similar SERS signals with the dominance of 665, 727, 891, 1336, and 1455 cm^{-1} bands. The only observed variation is the difference in their relative intensities, which arise from personal issues being unreported. Since all the specimens were taken disregarding the time of ovulation, this could be one of the right reasons for the sample-sample heterogeneity. The mutual dependence between circulating hormone levels and certain components of vaginal fluid was confirmed by Huggins [212]. Also, in my research on SERS time-dependent analysis of the menstrual cycle, I demonstrated the existence of spectral effect associated with biochemical variations of fertile and infertile phases (details in subsection d).

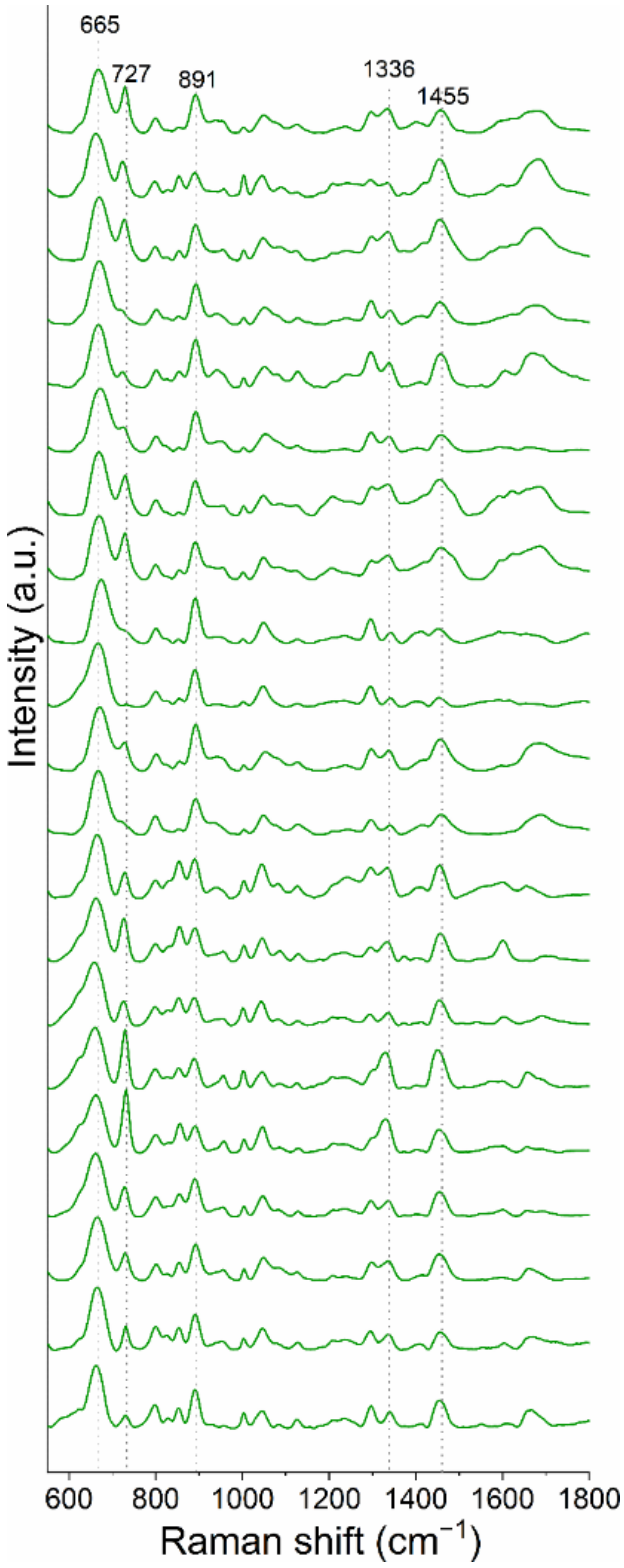

Figure 26 The averaged SERS spectra of vaginal fluids belonging to I category (I/II purity, 4.2 – 4.7 pH, many lactobacilli).

b) The spectral analysis of vaginal fluids of the II category

The SERS spectra of vaginal fluids belonging to the II category (Figure 27) have common bands at 727, 857, 894, 1125, 1336, 1455, and 1685 cm^{-1} identified as typical for bacteria, as I already demonstrated [167] (for comparison, please see Figure 24). The samples are characterized by II/III or III purity, which is a consequence of an overgrowth of non-lactobacilli bacteria. This is also directly linked with increased pH and leucocyte levels (usually 10 – 20 leucocytes). During the microscopic examination, the mentioned bacteria were recognized as pink-stained bacilli (Gram-negative bacteria) and blue-stained bacilli and cocci (Gram-positive bacteria). These bands with meaningful intensities only confirm the spectral impact of the number of bacterial cells on the total spectral image.

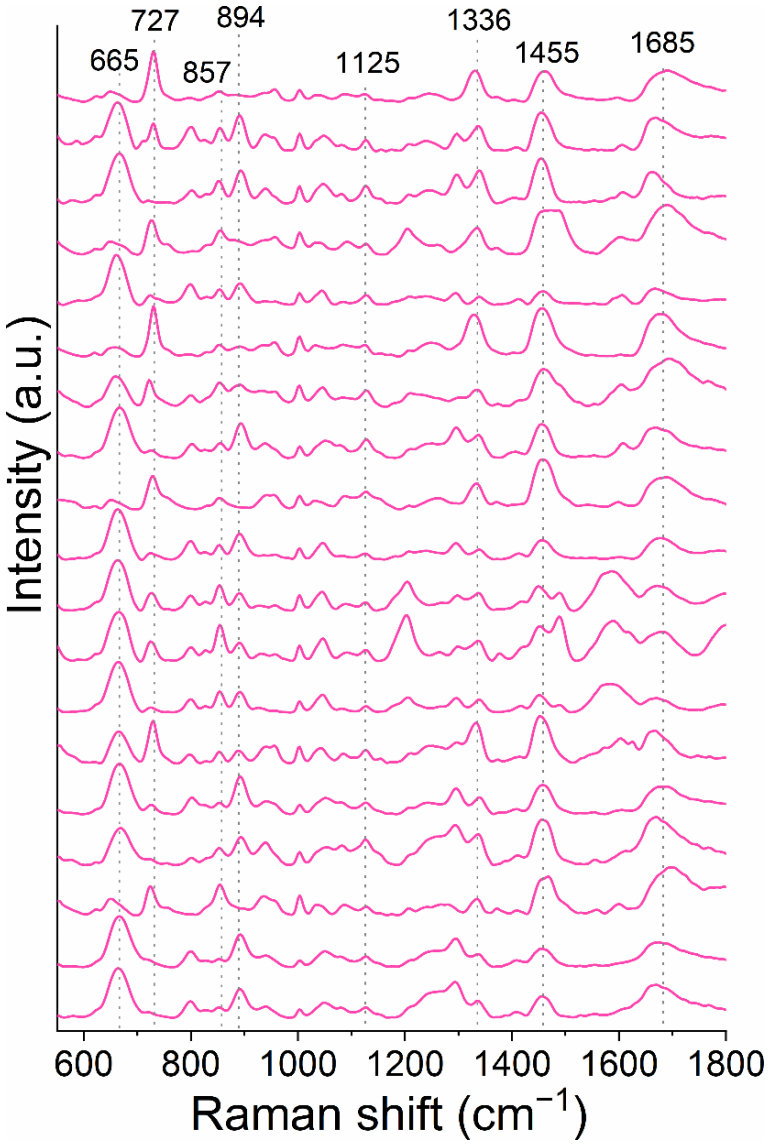

Figure 27 The averaged SERS spectra of vaginal fluids of the II category (pH=4.7 – 5.6, II/III or III purity, Gram-positive and Gram-negative bacilli or Gram-positive cocci, and few or lack lactobacilli).

c) The spectral analysis of vaginal fluids of the III category

Vaginal irrigation is a treatment that involves mechanically rimming the vagina with water, thus violating the normal microflora (gray spectrum in Figure 28). Indeed, the dominance of bands at 1003 cm^{-1} (ring breathing mode of phenylalanine (urea, albumin)) and 681, 853, 893, 1049 cm^{-1} (*e.g.*, lactic acid, acetic acid, citric acid, glycogen, glucose, proteins) exposes the significant contribution of biochemical compounds to the overall spectral response of such a sample. The absence of the 730 cm^{-1} band even more strengthens the thesis of its origin from bacteria. Such a completely different spectral profile indicates that the irrigation process alters the vaginal milieu significantly.

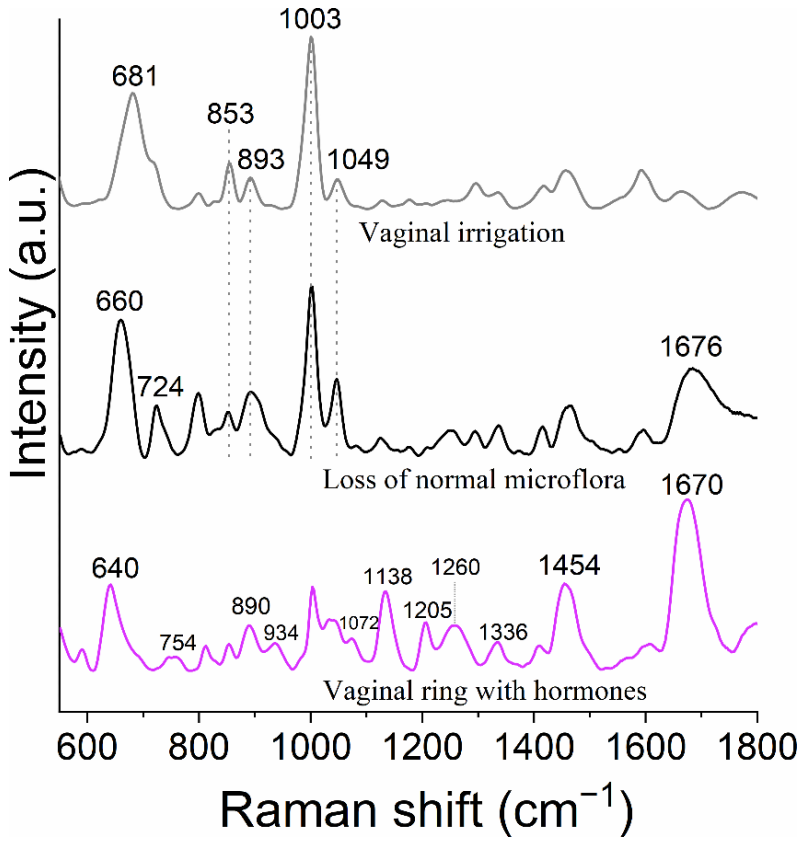

Figure 28 The averaged SERS spectra of vaginal fluids of the III category (samples influenced by vaginal irrigation practice, loss of normal microflora, and contraception usage).

A similar spectral response is observed for vaginal discharge from a woman whose age-related processes led to the loss of normal microflora, leaving only single epithelial cells and single Gram-positive cocci (black spectrum in Figure 28). The two analysed spectra (vaginal irrigation and loss of normal microflora) share similar bands as the composition of analysed materials is similar – microflora lost through the natural or intentional way.

Hence, the common bands come from biochemical molecules. Single epithelial cells are documented by the bands at 660 cm^{-1} and 1676 cm^{-1}, while the one at 724 cm^{-1} is assigned to bacteria [167].

The biochemical changes induced by the use of a hormone-loaded vaginal ring lead to the new spectral image with different relative band intensities (purple spectrum in Figure 28). Since the main components of this contraceptive are ethinylestradiol and etonogestrel (3-keto Desogestrel), bands located at 640, 754, 890, 934, 1072, 1138, 1205, 1260, 1336, 1454 cm^{-1} could partially arise due to mentioned hormones [213].

The PLSR analysis for vaginal fluids of the I, II, and III category

PLSR analysis was used to illustrate the correlation between the spectral responses of vaginal fluids belonging to the three considered categories. The model was created using the Kernel algorithm and cross-validation (random, 20-segments). The 2D score plot (Figure 29) discloses five main groups of spectra successfully separated according to the aspects analysed above. Such spatial distribution reveals that all samples belonging to the I category tend to be located next to each other while maintaining a certain distance from the rest. Factor-1 explains 43 % of the variance in the Y block, with 39 % of data within the X matrix. The first ten factors suggested by the model cover 91 % (X matrix) and 84 % (Y matrix) of total information. The model's performance is characterized by the low values of RMSECV and RMSEC (0.18 for both) and high values of R^2_{cal} and R^2_{CV} (0.86 for both).

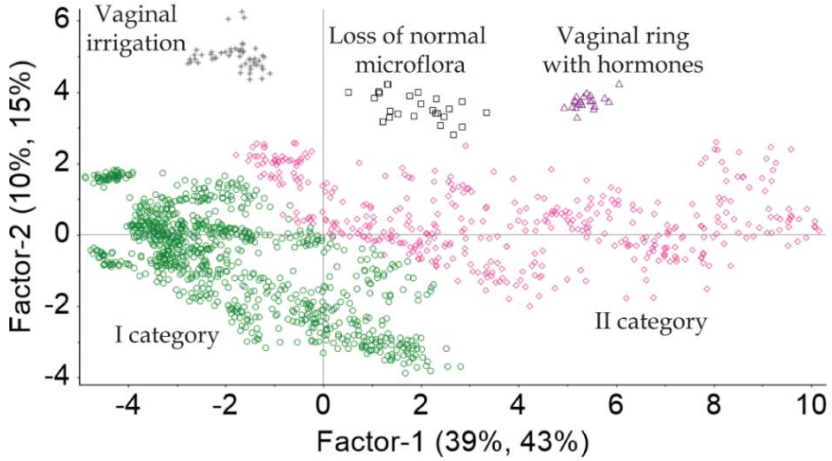

Figure 29 The 2D score plot of PLSR analysis performed for vaginal fluids of the I category (21 samples), II category (19 samples), and III category (vaginal irrigation practice (1 sample), a ring with hormones (1 sample), loss of normal microflora (1 sample)). For every sample, *ca.* 25 single spectra were recorded.

The above analysis demonstrated that all considered abnormalities in the microflora and the biochemistry of analysed vaginal fluids (II and III category) influence the SERS signal. Therefore, vaginal fluids grouped under the I category will be treated as a control class and used for further analysis in the aspect of SERS analysis of vaginal infections.

d) The spectral analysis of the fertile/infertile phase of the menstrual cycle (IV category)

To specify the hormone-dependent spectral changes within the menstrual cycle, the analysis was performed for all vaginal fluid samples (103 samples) collected within the first and the second menstrual cycles of one woman belonging to the control group. Based on the calendar and the organoleptic method, the fertility window was determined, resulting in the analysis of two classes of fertile and infertile days.

The 3D score plot of PLSR analysis provides information about the tendency of grouping spectra for two classes: fertile phase (FP) or infertile (IFP) phase (Figure 30). Partial overlap between groups can result from improper sampling or potential errors in phase determination, as biochemical changes can be unpredictable and out of control. The three subsequent factors (F-1, F-2, and F-3) gather the information at the level of 82 % (X matrix) and 59 % (Y matrix). The values of R^2_C and R^2_{CV} that characterize the model equal 0.70 and 0.54, respectively. RMSEC and RMSECV equal 0.28 and 0.34, respectively.

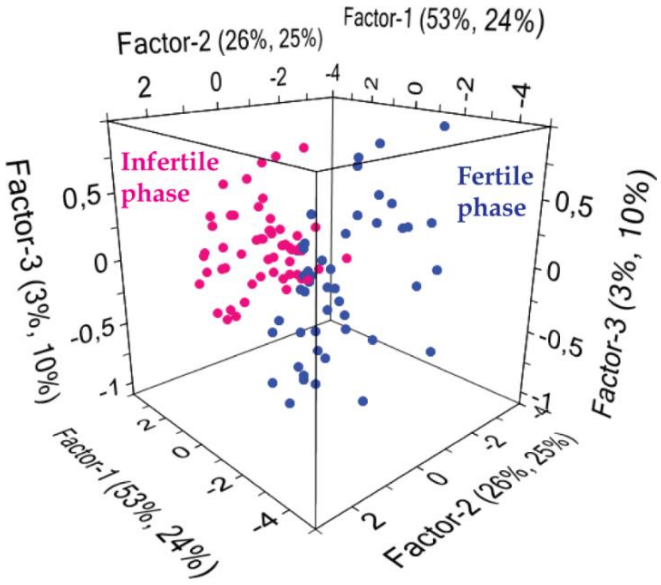

Figure 30 The 3D score plot of PLSR analysis performed for the vaginal fluid samples of two menstrual cycles. The number of samples represented as single points equals 103.

A detailed analysis of both menstrual cycles in terms of the fertile/infertile phase and also the differences in spectral images of the rest of the discharges (*i.e.*, menstrual blood and menstrual spotting) is presented in the article [214]. For a general view, all spectra collected over the entire two cycles are presented in the appendix section (Figure 1A).

The performed investigations demonstrated that biochemical processes of the menstrual cycle contribute to the spectral image of analysed vaginal fluid. At this stage of consideration, such hormone-dependent changes constitute crucial aspects during the diagnosis of vaginal infections. The interference of two spectral effects (one related to the biochemical processes within the menstrual cycle and the second to the ongoing infection) can hinder the reliable diagnosis. Thus, considering hormone-dependent changes as nonessential can result in the formation of an incorrect classification model, leading to diagnostic errors. However, creating separate models for at least two phases (as shown here, fertile/infertile) is labor-intensive and complicates the entire diagnostic procedure. Consequently, in the following paragraph, cycle- and infection-dependent spectral effects will be collated to estimate the magnitude of their impact.

The comparison of spectral effects: the phase of the menstrual cycle *vs.* vaginal infections

The comparative analysis of spectral effects related to the menstrual cycle phase and infections was performed on a series of vaginal fluid samples (subsection d) and patients diagnosed with vaginal infections of different etiological agents. Thus, the cohort of positively diagnosed patients (Infected cohort (IC)) included the following types of infection: bacterial vaginosis (BV) (3 patients – BV1, BV2, BV3), vulvovaginal candidiasis (VVC) (3 patients – VVC1, VVC2, VVC3), coinfection (1 patient – Coinf) and gonorrhea (1 patient).

The microscopic analysis revealed that the microflora of most IC patients is devoid of *Lactobacillus* spp.. They are replaced by numerous pathogenic bacteria, *i.e.* Gram-positive and Gram-negative bacilli and/or cocci, indicating an abnormal microbiome. The hallmark of VVC patients was the presence of blastospores and pseudohyphae. In turn, gonorrhea was recognized by Gram-negative cocci inside leucocytes. For Coinf. sample, numerous pathogenic bacteria, including *Gardnerella vaginalis*, were noted. The culture method also indicated coinfection with *C. albicans*. In every analyzed case, leucocyte level is in the range

of 4 – 10 or even higher, indicating inflammation. The pH level is elevated, reaching the value of 5 or higher. Thus, all samples reached III or IV purity. The detailed characteristics are presented in Table 2A in the appendix section.

Figure 31 presents the SERS spectra of all analysed samples – the infected cohort (VVC1, VVC2, VVC3, BV1, BV2, BV3, Coinf., Gonorrhea) along with the reference spectrum (VF), which is an average of all vaginal fluid samples of the first and second menstrual cycle. The apparent spectral differences between the two groups indicate a direct effect of infection. Both groups are characterized by the same bands at, *e.g.* 665, 723, 1003, 1047, 1330, 1455, and 1670 cm^{-1}. However, the differences that refer to their relative intensities alter the spectral response, making it crucial to achieve a complete differentiation. The significant observation is the change in the intensity of 723 cm^{-1}. This band increases for the IC samples, which is in line with the previously stated thesis regarding its origin (as the marker band for microorganisms) and with the biochemistry of the clinical samples related to the proliferation of pathogen cells. The increased intensity of bands at 1330, 1455, and 1670 cm^{-1} for IC can also indicate ongoing vaginal infections.

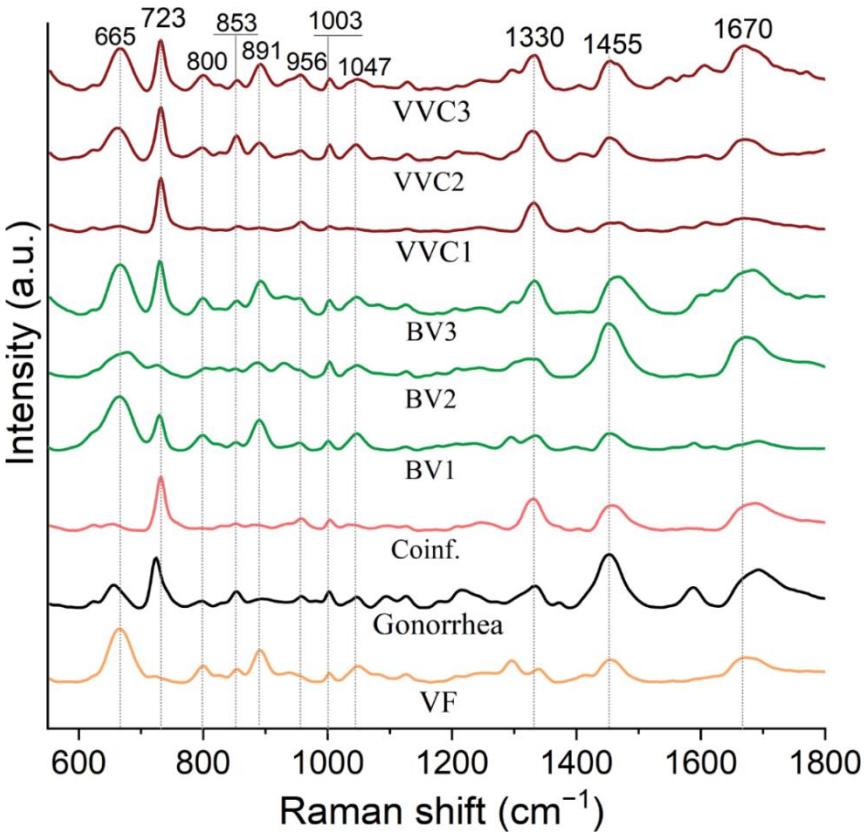

Figure 31 The averaged SERS spectra of vaginal fluids taken from control woman (VF) and those with gonorrhea, bacterial vaginosis (BV1, BV2, BV3), vulvovaginal candidiasis (VVC1, VVC2, VVC3) and coinfection (*Candida albicans* and *Gardnerella vaginalis*).

The dependencies between analysed data of this association (the first and the second menstrual cycle *vs.* IC class) were visualized with PLSR method in two distinct ways, such as VF with IC samples (Figure 32A) and FP, IFP of VF with IC samples (Figure 32B). These specific configurations deliver information about the magnitude of spectral effects resulting from the day of the menstrual cycle and infections as well. Both analysed models reveal the high-level distinction of spectral data described by subsequent factors. Hence, the F-1, F-2, and F-3 encompass 78 % (X matrix) with 60 % (Y matrix) for the VF-IC model and 79 % (X matrix) with 44 % (Y matrix) for the model with phase specification (FP-IFP-IC model).

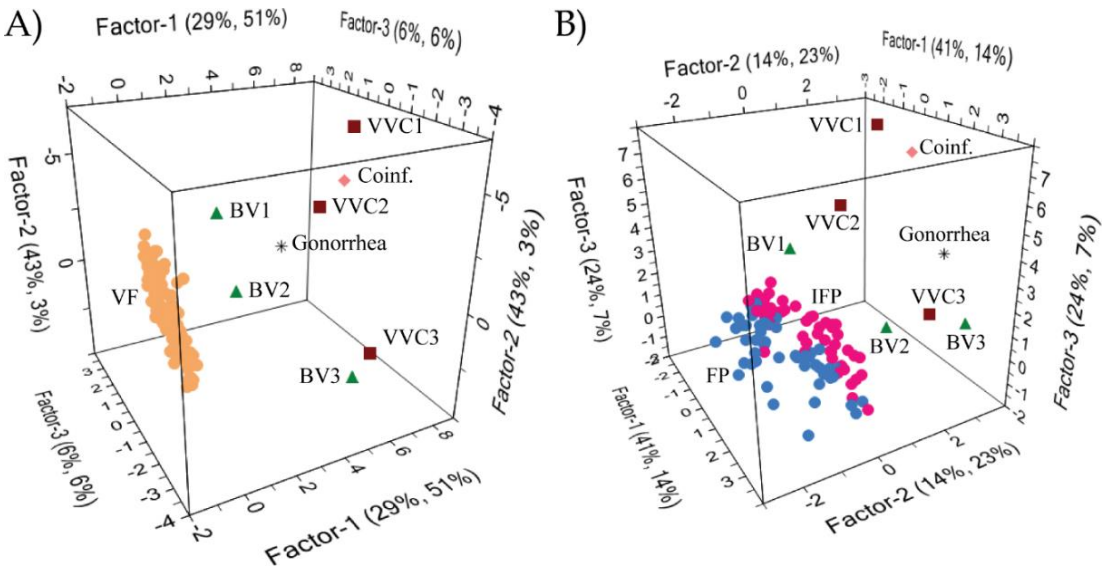

Figure 32 The 3D score plots of PLSR analysis for the association VF samples of the first and the second menstrual cycle and IC (VVC1, VVC2, VVC3, BV1, BV2, BV3, gonorrhea and Coinf.) divided into two ways A) VF and IC class (VF-IC model) and B) FP and IFP of VF and IC class (FP-IFP-IC model).

Such spatial data distribution reveals a prominent impact of the etiological agent and any other changes that infection brings. Considering the equally good separation for the FP-IFP-IC model (Figure 32B), where only samples among VF (FP and IFP) tend to overlap, the effect associated with infection is more significant than that of hormonal changes. In this configuration, Factor-1 allocates the vast majority of VF on the negative side of the score plot against IC samples on the opposite side with a relatively large distance along F-1. The high values of R^2 (R^2_C=0.85 and R^2_{CV}=0.80) and low values of RMSE (RMSEC=0.10 and RMSECV=0.12) indicate the excellent prediction abilities by the VF-IC model where the recognition of vaginal infections can be performed.

The weighted regression coefficient for the VF-IC model (Figure 33) reveals 665, 800, 891, 1052, and 1297 cm^{-1} variables which are typical for VF and come from biomolecules of secretion (lactic acid, citric acid, saccharides, glycogen, proteins, lipids, fatty acids, and nucleic acids) [154,199,215,216]. In turn, those at 730, 958, 1330, 1456, and 1600 cm^{-1} indicate specific oscillations in adenine (FAD, NAD, ATP, and DNA), proteins, and lipids (CH_2 deformation, CH_2 bending mode, C-H vibrations), phospholipids (NH_2 stretching, CH_2 bending and scissoring) that create the microorganism's cells [178,179]. As they arise due to the high accumulation of pathogenic microorganisms, they are suitable biomarkers of infection.

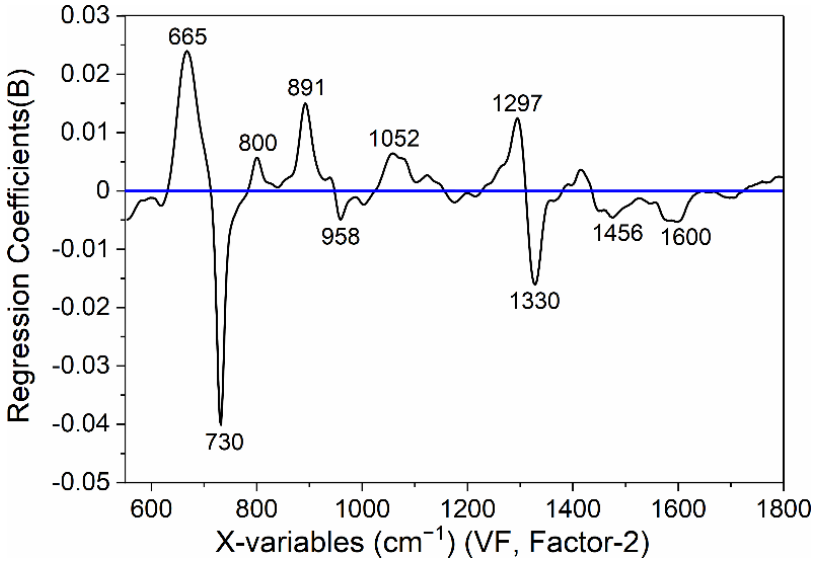

Figure 33 The weighted regression coefficient (VF response, Factor-2) for the VF-IC model.

The above analysis demonstrated that although there are spectral differences resulting from the phase of the menstrual cycle, these are considerably weaker compared to infection. This information is of great significance for diagnostics as it allows testing regardless of the day of the cycle.

4.2.3 Spectral response of a control vaginal fluid

The vaginal fluid is a body secretion composed of exfoliated cells, vulvar secretion, and secretion of Bartholin's and Skene's glands and exudate through the vaginal wall. It is enriched by the fluids from an upper reproductive tract, *e.g.* endometrial fluid, oviductal fluid, and cervical mucus, whose character and amount highly depend on the ovulatory cycle [212]. Water accounts for 92 – 98 % of the total composition of the cervical mucus.

The accompanying ingredients include electrolytes (*e.g.*, sodium chloride and potassium chloride), organic constituents (*e.g.*, ascorbic acid, citric acid, and squalene), lipids (*e.g.*, cholesterol and glycerides), phospholipids, amino acids, and fatty acids. Among the high molecular weight components, the most distinctive are glycoproteins (mucins) structured from long polypeptide chains with oligosaccharides as side chains. Other are neutral sugars (*e.g.*, fucose and galactose), amino sugars (*e.g.*, glucosamine, galactosamine), enzymes and enzyme inhibitors (*e.g.*, trypsin, antitrypsin, and amylase), and sialic acid. Proteins (*e.g.*, transferrin and albumin) with immunoglobulins are other constituents of a vaginal secretion arising from the direct synthesis by the vaginal mucosa. It is not excluded that they can appear by transduction process or as a product of the upper genital tract, *e.g.*, cervix (heptaglobin, β-lipoprotein, and α-macroglobulin). Carbohydrates (*e.g.*, glucose, maltotriose, maltose, and glycogen) are essential substrates in terms of enzymatic or microbial processes leading to lactic and acetic acid production [212,217]. Other compounds are urea, glycerol, vaginal peptidase, acid phosphatase, and pyridine [218].

To specify the SERS spectrum of a control vaginal fluid, all 21 reference signals (Figure 26) were averaged and presented with standard deviation in Figure 34A. This representative spectral image captures the entire range of components of the analysed vaginal fluids. The band assignments were based on the existing literature and SERS spectra of the most typical chemical compounds, *e.g.* lactic acid, citric acid, acetic acid, urea, glucose, and albumin (Figure 34B and Figure 34C). Moreover, a specially designed experiment aimed at the progressive separation of epithelial cells, microorganisms, and biochemical molecules contributed to a better understanding of signal enhancement. However, its detailed description is included in the article [219].

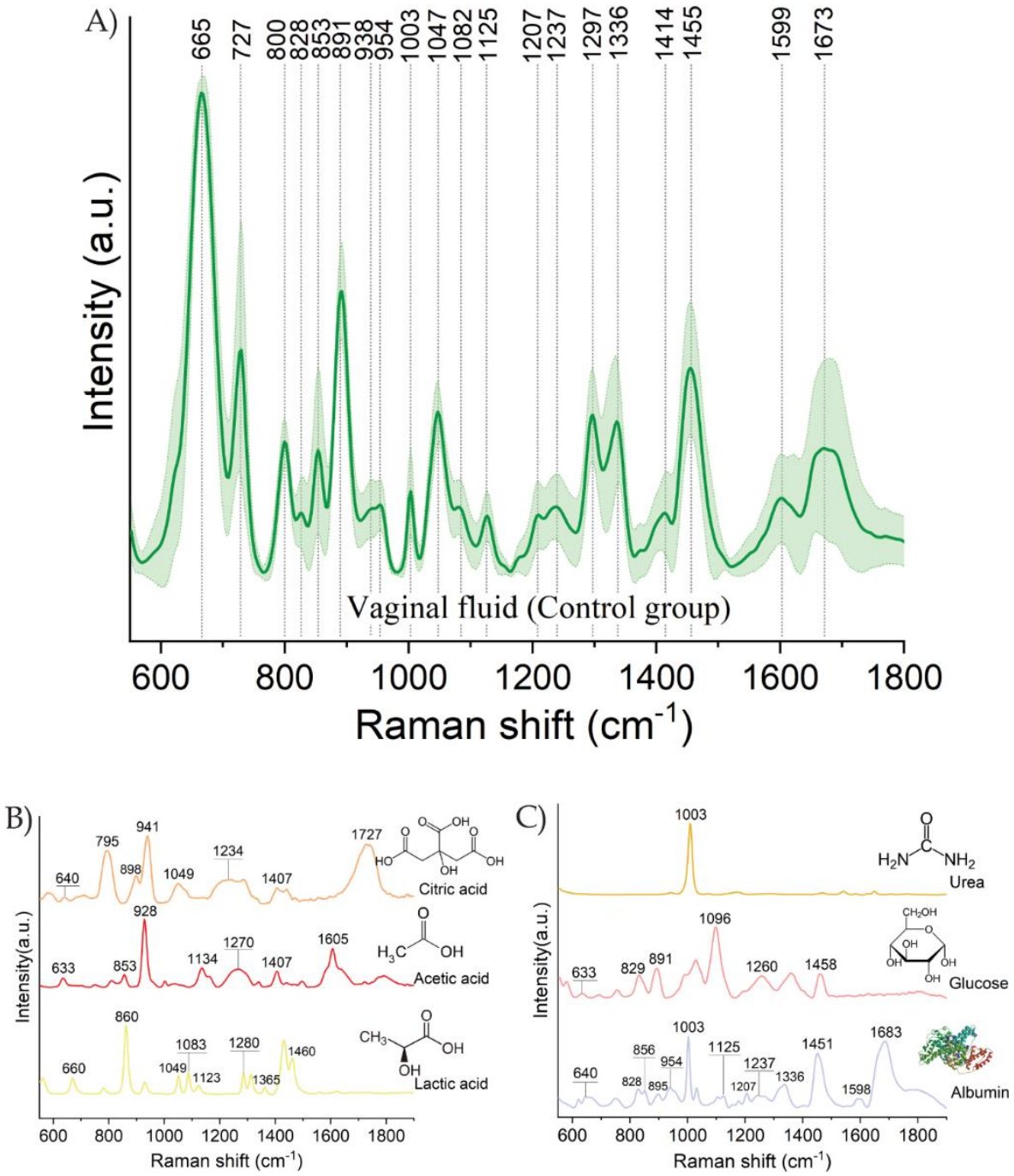

Figure 34 The averaged SERS spectrum of A) all vaginal fluids (21 samples) representing the control group and SERS spectra of B) and C) the common biochemical compounds present in a vagina.

The averaged SERS spectrum of control VF is dominated by the band at 665 cm^{-1}, which arises due to specific oscillations in thymine, guanine, tyrosine, cytosine, and other amino acids, while the one at 891 cm^{-1} indicates the presence of saccharides. The nucleic acids can be manifested by the band at 727 cm^{-1} (ring breathing mode of adenine) but also 800, 828 cm^{-1} (ν_{as}(O-P-O) and ν_s(O-P-O)), and 1082, 1125 and 1336 cm^{-1}. Various compounds such as lipids, proteins, glycoproteins, triglycerides, and fatty acids contribute

to the band at 1455 cm^{-1}. The spectral influence of proteins is evident in different bands, *e.g.* 828, 853, 891, 954, 1003, 1207, 1455, and 1599 cm^{-1}. Those located at 1237 and 1297 cm^{-1} are the signals of amide III of proteins, while the one at 1673 cm^{-1} arises due to amide I of proteins [151,154,155,215,216,220–223]. For more information, please see Table 9.

Table 9 The tentative assignments of the most intense bands present on the SERS spectra of the control vaginal fluid [151,154,155,215,216,220–223].

Raman shift (cm^{-1})	**Tentative assignment for control vaginal fluid**
665	Lactic acid, ring breathing mode of Thymine, Guanine (DNA/RNA), C-C twisting of proteins (tyrosine), C-S stretching of cytosine, COO^- def in amino acids, C-S stretching mode of cystine (collagen type I)
727	Ring breathing mode of Adenine (RNA/DNA), O-O stretching vibration in oxygenated proteins, ring breathing mode of tryptophan (protein assignment), C-N head group choline $(H_3C)_3N^+$ (lipid assignment), glycoproteins like mucins
800	ν_{as}(O-P-O) & ν_s(O-P-O) of nucleic acids, citric acid
828	ν_{as}(O-P-O) of nucleic acids, phosphodiester, glucose, ring breathing mode of tyrosine (proteins)
853	Acetic acid, lactic acid, ν(COC) of glycogen, ring breathing mode of tyrosine (proteins)
891	Citric acid, saccharides (glucose), proteins (*e.g.*, albumin)
938	C-C stretching in α-helix, proline, and valine in proteins, glycogen, citric acid
954	ν_{sym} (CH_3) of proteins (α-helix), amino acids, polysaccharides
1003	Aromatic ring breathing in phenylalanine (proteins), C-N stretching in urea
1047	C-O and C-N symmetric stretching in proteins, ν(C-O) & ν(C-C) of Glycogen, citric acid, ν(C-O), ν(C-CO_2), ν(C-CH_3) of lactic acid
1082	C-N stretching mode of proteins, $\nu_s(PO_2^-)$ in nucleic acids, carbohydrates (*e.g.*, glycogen), C–C stretching in lipids (cholesterol, triglycerides) and fatty acids, ν(C-O), δ(C-COH) of lactic acid
1125	C–C stretching in lipids (cholesterol, triglycerides) and fatty acids, C–N stretching mode in proteins, ν(C-O-C) in nucleic acids, $\nu(PO_2^-)$ in nucleic acid, C-O and C-C stretching in carbohydrates (*e.g.*, glycogen), ν(C-CH_3) of lactic acid, acetic acid
1207	A, T (vibration of DNA/RNA), tryptophan & phenylalanine *ν*(C-C_6H_5) mode, phenylalanine (protein assignment), C-C stretching in tyrosine
1237	Amide III of proteins arises from ν(C-N) coupled with δ_{bend}(N-H) (β-sheet), citric acid
1297	Amide III of proteins arises from ν(C-N) coupled with δ_{bend}(N-H) (α-helices), δ (CH_2) $_{twisting}$ of lipids (cholesterol and its ester, triglycerides) and fatty acids

1336	Ring mode of Adenine and Guanine of nucleic acids, NH_2 stretching in adenine and polyadenine, (COH) & (CH) bend of glycogen, tryptophan, proteins (*e.g.*, albumin)
1414	$\delta(CH_2)$ of nucleic acids, $\delta\ (CH_2)_{\text{bending}}$ of fatty acids, acetic acid, citric acid
1455	$\delta_{as}\ CH_3$ of lactic acid, $\delta(CH_2)$ in proteins, CH bending of lipids, Glucose, (CH_2) scissoring and (COH) bending of glycogen, CH_2, CH_3 bending of tryptophan, C-H stretching of glycoproteins including mucins, hydrocarbon chain of lipids, bending CH_2/CH_3 of triglycerides and fatty acids
1599	Ring C-C stretching of phenylalanine, C=C in-plane bending of phenylalanine & tyrosine, acetic acid, proteins (*e.g.*, albumin)
1673	Amide I of proteins arises from ν(C=O) (β-turn, random structure), ν(C=C) of the ring of cholesterol and its esters

4.2.4 Spectral recognition of vulvovaginal candidiasis (VVC)

Given the significant prevalence, this section is devoted to the spectral analysis of vaginal fluids samples taken from women diagnosed with VVC (the detailed characteristic in Table 3A). The averaged VVC signal was collated with the control one in separation (Figure 35A) and superposition (Figure 35B). Both spectral images reveal bands with similar shapes located at the same position with the dominance of 665, 727, 891, 1336, 1455, and 1673 cm^{-1}. Compared to the control class, bands at 727, 954, 1003, and 1125 cm^{-1} and in the region 1200 – 1700 cm^{-1} of VVC spectra are intensified as a consequence of the ongoing infection. This spectral trait is caused by the presence of *Candida* spp. and the pathogenic bacteria accompanying infection.

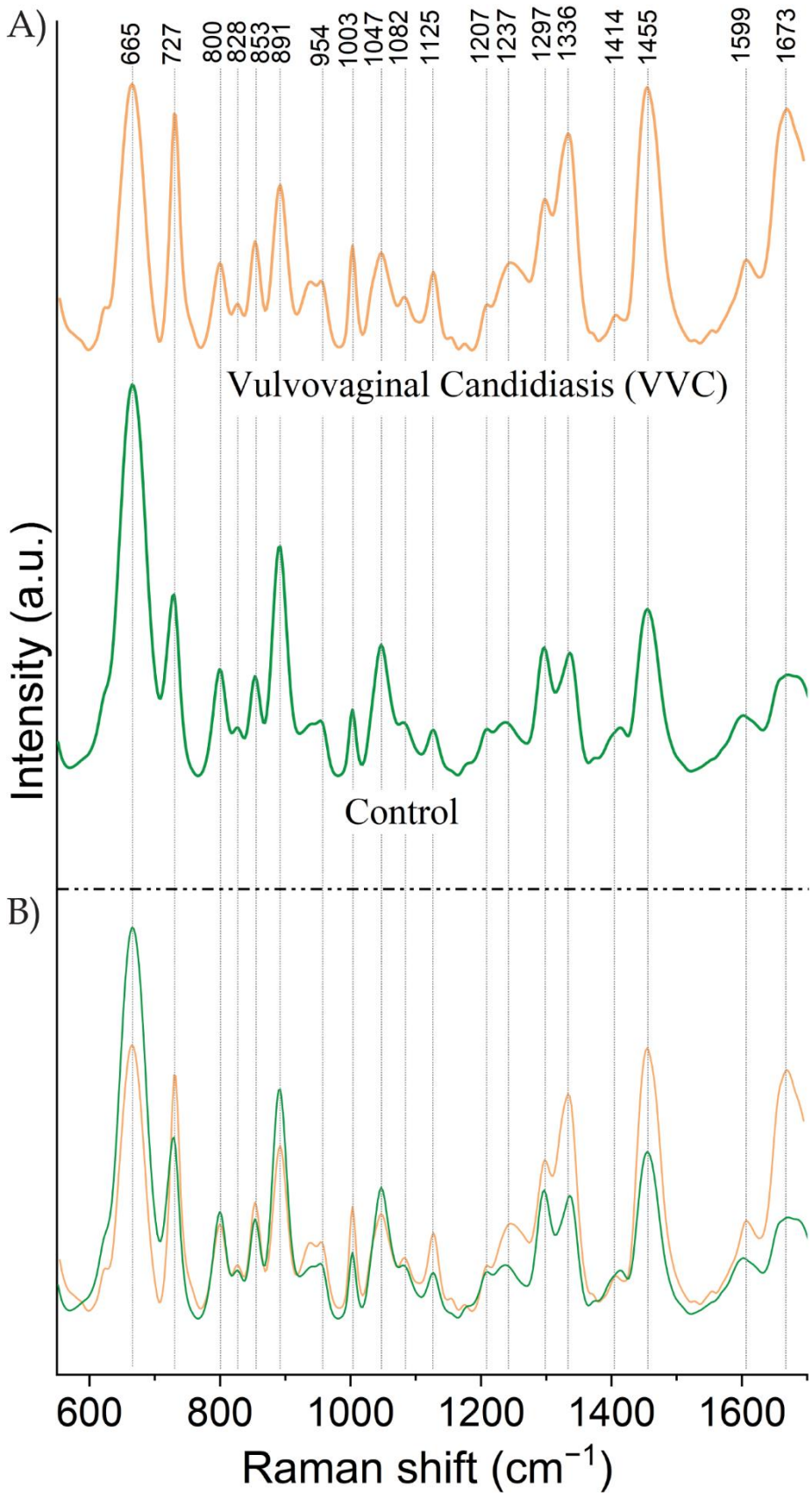

Figure 35 The averaged and normalized SERS spectra of vaginal fluids taken from all 21 patients belonging to the control group and all 12 patients diagnosed with VVC presented in A) separation and B) superposition.

The results of PLSR analysis for the analysed association – 21 control samples *vs*. 12 VVC samples are presented in Figure 36. The calibration model was created according to Kernel algorithm and cross-validation (random, 20-segments). The distribution of spectral data reveals well-separated classes with three factors gathering the information at the level of 61 % (X matrix) and 84 % (Y matrix). The high values of R^2_{cal} and R^2_{CV} (both equal 0.86) and low values of RMSEC and RSMECV (both equal 0.18) indicated an excellent predictive ability of the created model where the VVC infection can be recognized.

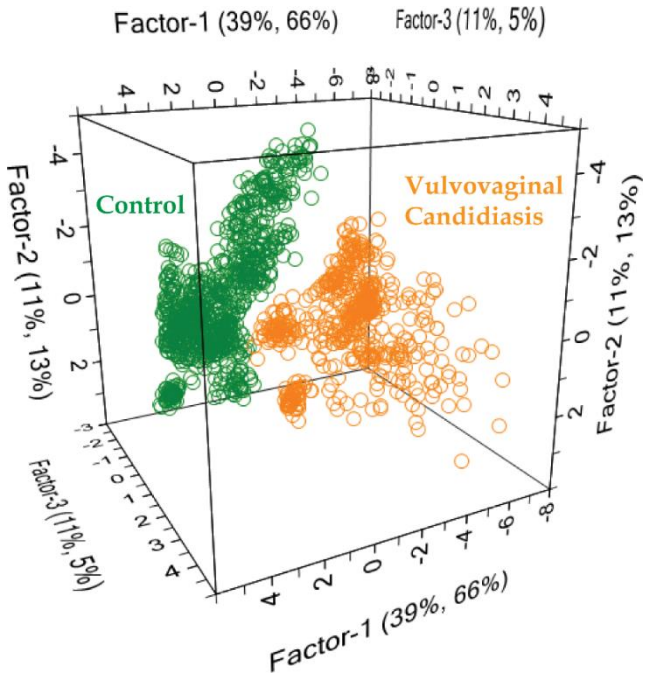

Figure 36 The 3D score plot of PLSR analysis performed for the association (21 control *vs.* 12 VVC samples). For every sample, *ca.* 25 single spectra were recorded and presented as single points.

The VVC class covers more space on the plot score dimension than the control one. This spread of the VVC class is directly related to the more diverse spectral information carried by an ongoing infection. The analysed VVC samples are highly variable in terms of pH, leucocytes, and whole microflora (in terms of quantity and quality of microorganisms). According to the biochemical description of samples (Table 3A), the majority of patients had microflora consisting of mycelial fragments. The pH was increased in the company of pathogenic bacteria (Gram-positive cocci and Gram-positive bacilli). For four patients (VVC1, VVC3, VVC5, VVC10), the presence of lactobacilli maintaining pH of 4.7 and accompanying blastospores or pathogens were noted. Also, for four patients (VVC1, VVC7, VVC8, VVC10), the infection was diagnosed solely based on the culture method.

The analysed weighted regression coefficient (Figure 37) referring to the F-1 and Control response reveals 669, 891, and 1047 cm^{-1} being significant for the control class. These bands represent the oscillations in amino acids, proteins, saccharides, and other components such as lactic acid and/or citric acid. While those characterizing VVC class – 727, 1003, 1125, 1258, 1331, 1455, and 1670 cm^{-1} highly correlate with the spectral 'fingerprints' of microorganisms.

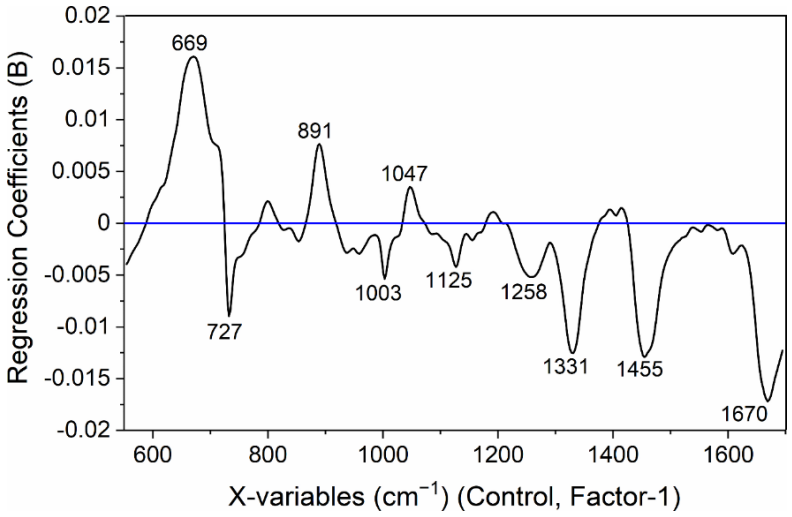

Figure 37 The weighted regression coefficients of PLSR analysis performed for the association (21 control samples *vs.* 12 VVC samples). This plot refers to the Factor-1 and Control response.

4.2.5 Conclusions

In this section, SERS and PLSR chemometric method were used to track the spectral variations of vaginal fluids encountered from biochemical and microflora disorders. It has been demonstrated that decreased purity, intentional activities (contraception, irrigation), or natural age-related processes alter the biochemistry and the whole vaginal milieu. Consequently, the SERS spectra of those vaginal fluids reveal distinct characteristics compared to the control class.

The time-dependent analysis of the menstrual cycle revealed the existence of spectral differences between fertile and infertile phases as a result of the changing levels of various substances involved in biochemical processes. Moreover, the comparative analysis provided significant information about the magnitude of spectral effects – those associated with changes in ongoing infection are greater than those within the menstrual cycle. This is a relatively comfortable situation, as the diagnosis can be made regardless of the day of the menstrual cycle.

In the last step, the established control class was related to the VVC class to detect the spectral characteristic induced by the ongoing infection. The analysis of 12 VVC patients revealed the presence of spectral features indicative of bacterial and *Candida* spp. proliferation tightly associated with infection. The assistance of PLSR allows for the profound analysis of the data set in the association Control *vs.* VVC class. The first three factors explaining the majority of spectral information designated a clear boundary between the two considered classes. The performance of the created model was described by satisfying FOM values, indicating its high ability to recognize VVC infection. PLSR revealed the variables of particular relevance for the differential analysis of control and VVC class. Hence, 727, 1003, 1125, 1258, 1331, 1455, and 1670 cm^{-1} constitute the spectral representation of microorganisms. The 669, 891, and 1047 cm^{-1} bands are indicative of the biochemical composition of vaginal fluid, such as amino acids, proteins, saccharides, and lactic acid.

4.3 COVID-19

This chapter is focused on the analysis of saliva and nasopharyngeal swabs to search for the spectroscopic image of SARS-CoV-2 infection. SERS measurements will be assisted with various chemometric methods operating in supervised mode. The aim is to find the best-performing algorithm for the accurate extraction of spectral information and the associated distinction between CoV(−) and CoV(+) classes. The first step encompasses the spectral analysis of the two materials collected from patients who tested positive and negative for COVID-19. Such characterisation includes the identification of spectral differences being significant for diagnostic purposes as well as their recognition at the molecular level. Then, calibration models consisting of well-defined classes will be created and optimized for the selected chemometric methods (PLS-DA, PCA-LDA, and SVMC). Their efficiency in COVID-19 recognition will be tested based on the classification involving external samples not belonging to the calibration models. The biochemical information associated with infection is specifically encoded in SERS spectra and can be revealed differently depending on the type of clinical material (saliva, nasopharyngeal swab). Therefore, the diagnostic parameters (accuracy, specificity, sensitivity) will enable us to compare chemometric methods adopted to SERS technique and also to specify the usefulness of the analysed clinical material for COVID-19 diagnosis.

4.3.1 Spectral characterisation of CoV(+) and CoV(−) saliva

Saliva is a colourless fluid consisting of 94 – 99 % water. The pH is usually 6.64 and is highly dependent on the level of CO_2 in the blood. The inorganic constituents (0.2 %) include electrolytes (*e.g.*, sodium, calcium, potassium, chloride, magnesium), whereas the organic ones (0.5 %) are oligopeptides, polypeptides, glycoproteins, glucose, α-amylase, maltase, serum albumin, urea, uric acid, ascorbic acid, creatinine, lactate, amino acids. The salivary compositions also contain hormones (testosterone, cortisol), enzymes, and immunoglobulins IgA, IgG, and IgM [224,225]. The presence of proteins such as lactoferrin, lysozyme, peroxidase, cystatins, mucins, cathelicidin (LL-37), salivary agglutinin (gp340, DMBT1), alpha-defensins, beta-defensins make saliva to have antiviral functions. Saliva also contains exfoliated epithelial cells, white blood cells, serum elements, oral microorganisms, and their metabolites. In the context of COVID-19 infection, salivary glands can be a significant reservoir for SARS-CoV-2 virus. The positive rate of COVID-19 in the saliva can exceed 92 %. This may explain the incidence of asymptomatic cases, where saliva is the main source of infection transmission. Saliva, as a clinical material, holds several vital traits that facilitate testing and screening for COVID-19. Therefore, it should be spectrally considered as a body fluid with potential use in diagnostics [226].

The complete analysis of saliva samples was performed on 169 patients, including 82 patients with a positive SARS-CoV-2 test result (hereafter CoV(+)) and 87 patients with a negative SARS-CoV-2 test result (hereafter CoV(−)). The averaged spectral responses of CoV(+) and CoV(−) saliva with standard deviation (SD) (the ribbon-like plot) are depicted in Figure 38A and as superimposed in Figure 38B. The calculated SD, which ranges from 15 to 39 %, unveils the spectral differences manifested as variations in the intensities of bands. They may correspond to individual patient traits, the stage of the disease, and the specific body response to infection.

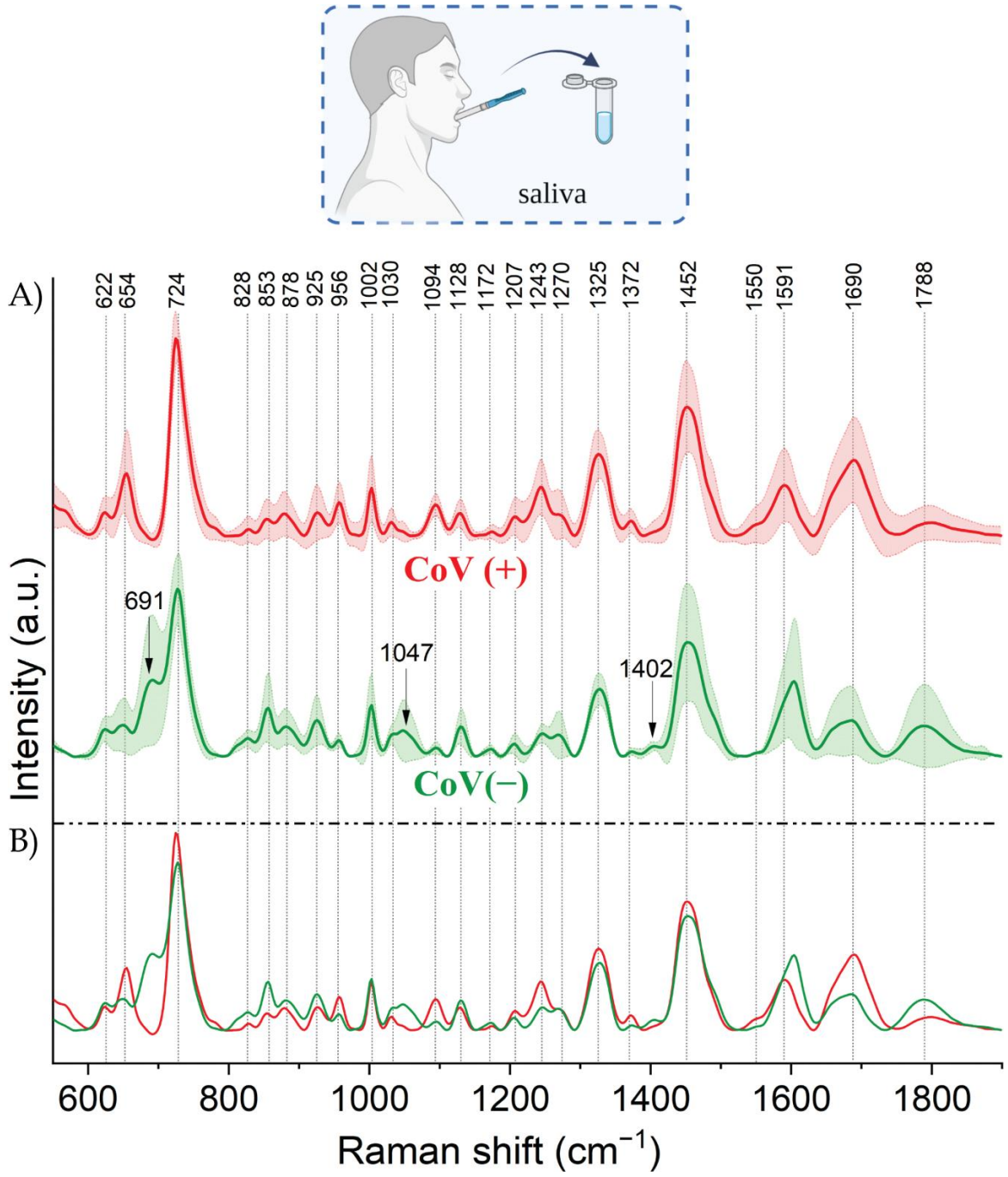

Figure 38 The averaged and normalized SERS spectra of CoV(+) and CoV(−) saliva samples presented A) with the standard deviation (SD) and B) as superimposed. The spectra were averaged from 72 CoV(+) and 77 CoV(−) samples. For every sample, *ca.* 15 single spectra were recorded.

The SERS spectrum of CoV(−) saliva (Figure 38A) reveals the presence of proteins in the form of signals of amide III (1243 and 1270 cm^{-1}), amide II (1550 cm^{-1}), and amide I (1690 cm^{-1}). The 1452 cm^{-1} can be assigned to, *e.g.* lipids, glycoproteins, triglycerides, and fatty acids, while the one at 1604 cm^{-1} to hydroxyproline, hypoxanthine, and C=C in-plane bending mode of phenylalanine and tyrosine. The unique signals 691, 1047, and 1402 cm^{-1}

for CoV(−) constitute a key differentiating element. According to the literature, these bands represent specific oscillations in cytosine, C-O, C-N stretching in proteins, and also C-CH_3 of glycogen and bending of the methyl group in proteins, respectively [68,199].

The intensity ratio of 853/828 cm^{-1} unveils the specific nature of tyrosine residues [227]. For CoV(−) saliva, this ratio equals 4.8. As a result of SARS-CoV-2 infection, the band at 853 cm^{-1} loses intensity, decreasing the 853/828 cm^{-1} ratio to 0.8 value. This can be explained by some specific interaction between tyrosine residues that now are hydrogen-bonded acceptors with viral proteins (spike glycoproteins) or other expressed molecules (ACE2 receptor) [226] and immunity proteins (IgA, IgM, IgG) [228].

Saliva, which is spectrally characterized by 654, 724, 1243, 1325, and 1452 cm^{-1}, exposes specific oscillations in methionine (*e.g.*, C-S stretching and CH_3, CH_2 deformation) or methionine adenosyl transferase (the enzyme that converts methionine to S-adenosyl methionine) [229]. It was observed that the intensity ratio of the individual bands to the band at 1002 cm^{-1} (I_{654}/I_{1002}, I_{724}/I_{1002}, I_{1325}/I_{1002}, I_{1452}/I_{1002}) is higher for CoV(+) samples. Two phenomena can support this conclusion. Firstly, in the presence of nonstructural proteins (*e.g.*, nsp10, nsp14, and nsp16) of SARS-CoV-2, S-adenosyl methionine is involved as a methyl donor in the process of viral RNA cap methylation (transferring a methyl group to the viral genome). This process is essential for virus replication, translation, and continued survival in host cells. The efficiency of S-adenosyl methionine synthesis is correlated with the level of methionine. If the methionine concentration is too low, the viral methyltransferase reaction will be blocked. The reduced ability of the virus to properly replicate ultimately leads to the degradation of viral genomes. Secondly, during SARS-CoV-2 infection, T-cells and macrophages are over-activated, resulting in a massive increase in the level of cytokines (IL-1β, IL-6, TNF-α) and proteins associated with the macrophage. Upon this activation, T-cells also have an increased requirement for methionine [230,231]. The spectral manifestations of T cells are 1094, 1325, 1372, 1452, and 1690 cm^{-1}. Their increased intensity (in relation to 1002 cm^{-1}) for CoV(+) saliva only confirms the mentioned overactivation [232].

Ferritin is a protein that stores iron and releases it during cell proliferation or metabolic renewal. Testing ferritin levels is important from the point of view of its deficiency as well as excess, which is associated with various diseases (*e.g.*, anemia) [233]. Recent studies have shown that salivary ferritin levels can rise significantly during COVID-19 infection [234].

Moreover, immunoglobulins (*e.g.*, IgG) are produced by immune system cells in response to some external agent (*e.g.*, viruses). These proteins, having the ability to recognize and specifically bind to their antigens, lead to the inactivation of microorganisms [235]. In these studies, the observed intensified bands in the region 1200 –1300 cm^{-1} but also 1452 and 1690 cm^{-1} may reflect the increased level of ferritin [221] and specific immunoglobulins in CoV(+) saliva [236].

Due to the diversity of bond oscillations that contribute to the SERS signal formation, the aforementioned bands at 1094 and 1325 cm^{-1} can also have different origins. Hence, they can arise due to specific oscillations – $\nu_s(PO_2^-)$ in nucleic acids and CH_3CH_2 wagging mode in purine bases of nucleic acids [199]. Their increased intensity for CoV(+) saliva can be explained by the multiplication of genetic material during infection.

The SERS measurements for chemical compounds were performed in a standardized manner – conditions and parameters (SERS substrates, device, laser line, laser power) are consistent with those for clinical samples. The defined signals help to reinforce the analysis and understand the formation of spectral responses of clinical samples. The averaged SERS spectra with the identified bands [237–240] are presented in Figure 39 and Figure 40.

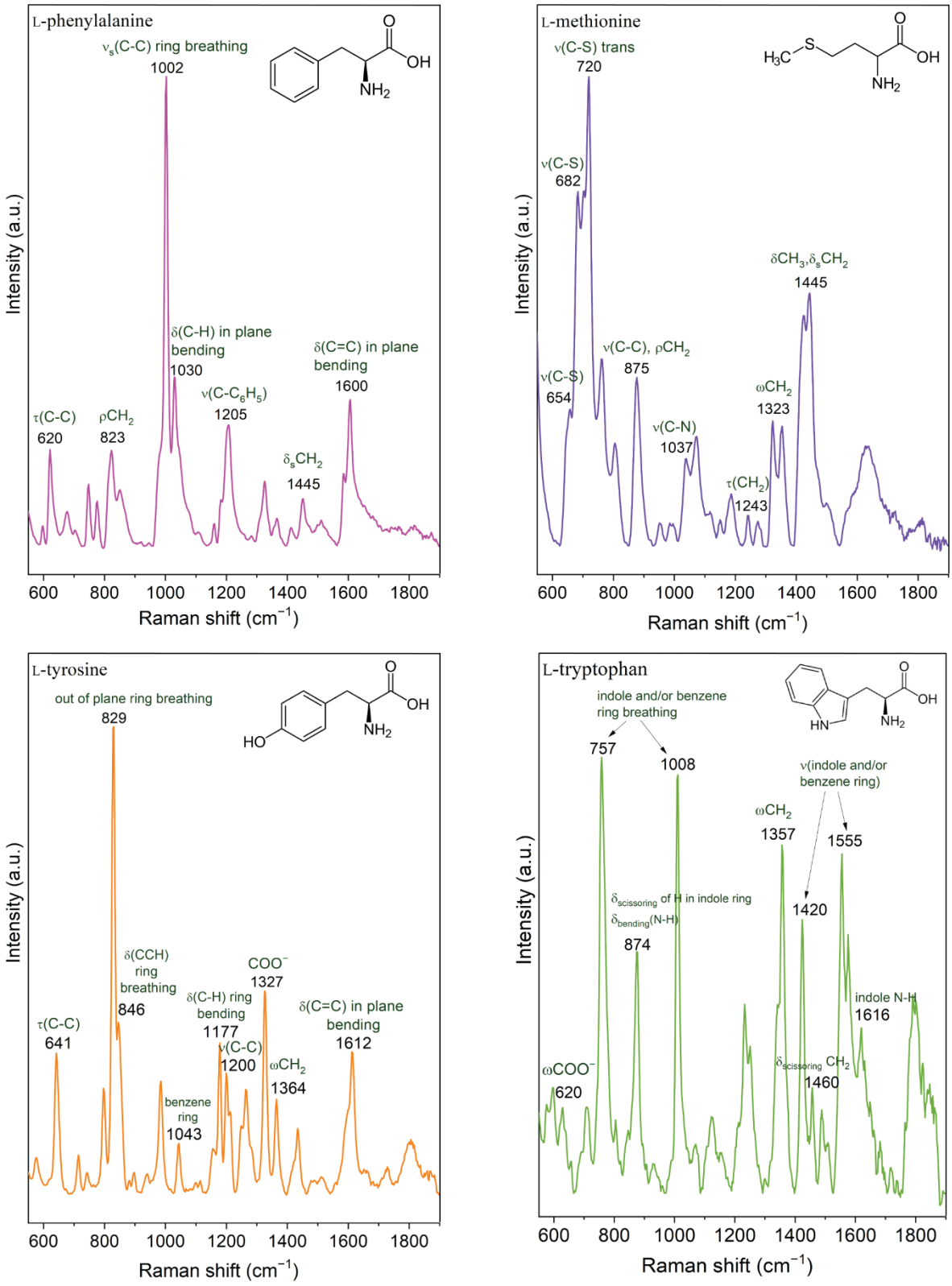

Figure 39 The averaged SERS spectra of selected amino acids (liquid form). For every sample, *ca.* 15 single spectra were recorded.

Legend: ν – stretching; s – symmetric; as – asymmetric; δ – deformation; ρ – rocking; ω – wagging; τ – twisting/torsion.

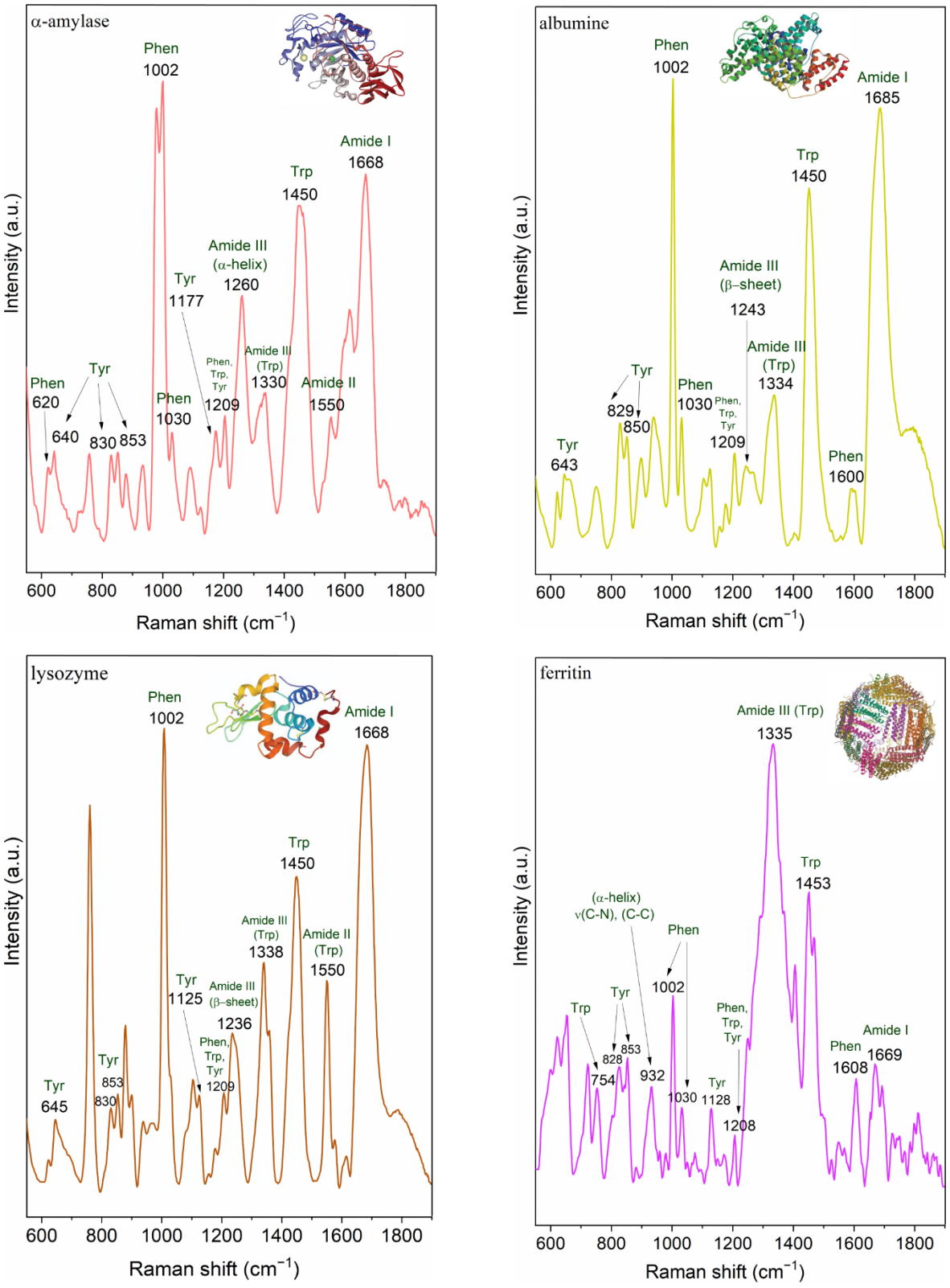

Figure 40 The averaged SERS spectra of selected proteins (liquid form). For every sample, *ca.* 15 single spectra were recorded. Phen – phenylalanine; Tyr – tyrosine; Trp – tryptophan.

Considering the literature review and the defined SERS spectra of the mentioned chemical compounds, the bands on the SERS spectra of saliva and nasopharyngeal swabs were identified. The entire spectral characteristic of saliva and nasopharyngeal swabs with the listed bands and the proposed assignments is presented in Table 10.

Table 10 The tentative assignments for the prominent bands observed on the SERS spectra of CoV(+) and CoV(−) saliva samples and nasopharyngeal swabs [68,155,245–247,199,220,221,227,241–244].

Raman shift (cm^{-1})				**Tentative assignment**
Saliva		**Nasopharyngeal swabs**		
CoV(−)	**CoV(+)**	**CoV(−)**	**CoV(+)**	
622	622	622	622	adenine, C-C twisting mode of phenylalanine (protein)
649	654	654	654	C-S stretching vibration in methionine, C-C twisting mode of tyrosine
–	–	–	688	G (ring breathing modes in the DNA bases)
691	–	–	–	δ(O–C=O) creatinine, cytosine
724	724	724	724	Ring breathing mode of adenine (RNA/DNA), O-O stretching vibration in oxygenated proteins, glycoproteins like mucins, ring breathing mode of tryptophan (protein assignment), C-N head group choline $(H_3C)_3N^+$ (lipid assignment)
828	828	828	828	ν_{as}(O-P-O) of nucleic acids, phosphodiester, glucose, ring breathing mode of tyrosine, transferrin (tyrosine, H-bonding)
853	853	853	853	ν(COC) of glycogen, ring breathing mode of tyrosine, transferrin (tyrosine, H-bonding)
878	878	878	878	Proline, valine, glycine, tryptophan, glutamate or ν(C-C) hydroxyproline, transferrin (tryptophan, H-bonding) or $\nu_sP(OH)_2$ of phosphate
925	925	925	925	C-C stretching proline ring, carboxylates including glucose and glycogen
956	956	956	956	hydroxyapatite, xanthine, proline, valine
1002	1002	1002	1002	aromatic ring breathing of phenylalanine, phenylalanine in lysozyme, lactoferrin, albumin, transferrin
1030	1030	–	–	C-H in-plane bending mode of phenylalanine, phenylalanine in lysozyme, lactoferrin, albumin
1047	–	1045	1045	C-O and C-N stretching in proteins, ν(C-O) & ν(C-C) of glycogen

1094	1094	1094	1094	$\nu_s(PO_2^-)$ in nucleic acids
1128	1128	1128	1128	C–C stretching in lipids (cholesterol, triglycerides) and fatty acids, C–N stretching mode in proteins, ν(C-O-C) in nucleic acids, $\nu(PO_2^-)$ in nucleic acid, C-O and C-C stretching in carbohydrates (*e.g.*, glycogen)
1172	1172	1172	1172	bending C-H tyrosine, transferrin (tyrosine, CH_3)
1207	1207	1207	1207	A, T (vibration of DNA/RNA), tryptophan & phenylalanine $\nu(C\text{-}C_6H_5)$ mode, phenylalanine (protein assignment), C-C stretching in tyrosine, hydroxyproline, tryptophan in lysozyme, lactoferrin, albumin
1243	1243	1243	1243	Amide III in proteins (random structure), *e.g.* albumin, lysozyme
1270	1270	1270	1270	Amide III in proteins (α-helix), *e.g.* transferrin, α-amylase
1325	1325	1325	1325	CH_3CH_2 wagging mode in purine bases of nucleic acids
1372	1372	1372	1372	Lipids, proteins (tryptophan), T, A, G (ring breathing modes of the DNA/RNA bases)
1402	–	–	–	Bending of methyl groups in proteins
1452	1452	1452	1452	$\delta(CH_2)$ in proteins, CH bending of lipids, Glucose, (CH_2) scissoring and (COH) bending of glycogen, CH_2, CH_3 bending of tryptophan, tryptophan in lysozyme, lactoferrin, albumin, C-H stretching of glycoproteins including mucins, hydrocarbon chain of lipids, bending CH_2/CH_3 of triglycerides and fatty acids
1550	1550	1553	1553	Amide II in proteins, ν(C=C) tryptophan
1604	1591	1586	1586	tryptophan, hydroxyproline, hypoxanthine, C=C in-plane bending mode of phenylalanine and tyrosine, cytosine (NH_2)
1690	1690	1685	1685	Amide I in proteins (*e.g.*, lysozyme, lactoferrin, albumin)
1788	1788	1780	1780	Albumie, lysozyme, tryptophan, tyrosine

4.3.2 Chemometric analysis of the spectra of saliva

The supervised chemometric methods PLS-DA, PCA-LDA, and SVMC were used to analyse the SERS spectra of saliva to recognize the spectral image of COVID-19. The same set of spectral data was used separately for each method, comprising two main parts (calibration and validation) (Table 11).

Table 11 The participation of individual saliva samples at the calibration and validation stage for all adopted methods (PLS-DA, PCA-LDA, and SVMC).

Operation / Material	Calibration			Validation		
	No. of samples	No. of		No. of samples	No. of	
		samples	spectra		samples	spectra
CoV(+) saliva	72	149	2220	10	20	300
CoV(−) saliva	77			10		

The calibration models for saliva were constructed using 149 samples (72 CoV(+) and 77 CoV(−)). The classification abilities of created models were tested on 20 external samples (10 CoV(+) and 10 CoV(−)) that were not involved in the calibration step. Validation was intended to imitate the actual diagnostic conditions. Therefore, to deviate from suggestions and provide conditions that were as close to the real ones as possible, all information about the sample origin was delivered after measurements. Approximately 15 single spectra were recorded for each sample, with a distribution for the calibration and validation stage of 2220 and 300 spectra, respectively.

To compare the classification performance of chemometric methods, the diagnostic parameters (accuracy, sensitivity, and specificity) were determined according to the following formulae:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \cdot \mathbf{100\%} \tag{4.1}$$

$$\boldsymbol{Sensitivity} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \cdot \mathbf{100\%} \tag{4.2}$$

$$\boldsymbol{Specificity} = \frac{\boldsymbol{TN}}{\boldsymbol{TN+FP}} \cdot \mathbf{100\%} \tag{4.3}$$

where: TP – true positive,
FP – false positive,
TN – true negative,
FN – false negative.

1. Partial least squares discriminant analysis (PLS-DA)

The PLSR calibration model was established using the Non-linear Iterative Partial Least Square (NIPALS) algorithm and random cross-validation (10-segments). Given the enormous biochemical variability in the analysed materials related to personal issues, the data fit the model correctly, resulting in relatively high R^2 and low RMSE values. Hence, R^2_{cal} and R^2_{CV} equals 0.70 and 0.69, respectively. RMSEC and RMSECV are 0.27 and 0.28, respectively.

The score plot (Figure 41A) represents that CoV(−) and CoV(+) saliva samples are largely separated from each other and tend to create two clusters depending on the origin – red one (CoV(+) samples) and green one (CoV(−) samples). The partial overlapping of samples can be elucidated by PCR method inaccuracies (false negative or false positive results) directly related to the misdiagnosis in the hospital. The first latent variable (LV1) explains 40 % of the variance in block Y with 22 % of the spectral data (X matrix), while LV2 explains 5 % with 21 % of the data within the X matrix. The eleven factors encompass 80 % (X data) and 70 % (Y data) of all spectral information.

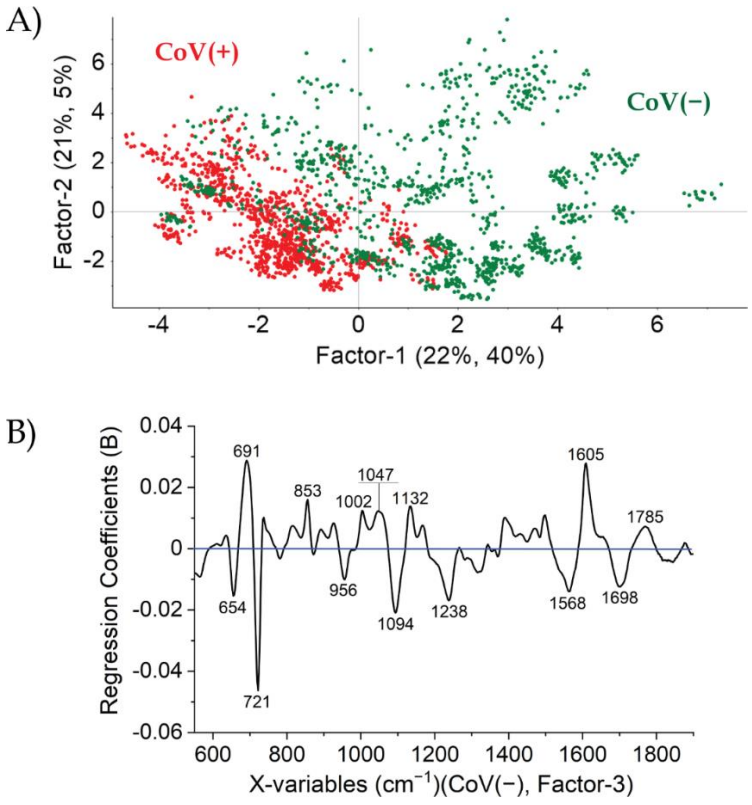

Figure 41 The results of PLS-DA analysis performed for 72 CoV(+) and 77 CoV(−) saliva samples in the form of A) 2D score plot and B) weighted regression coefficients (CoV(−) response, Factor-3). For each sample, *ca.* 15 single spectra were recorded and presented as dots on the score plot.

The plot of weighted regression coefficients calculated for the three components and CoV(−) response is presented in Figure 41B. The identified variables have positive coefficients (691, 853, 1002, 1047, 1132, 1605, 1785 cm^{-1}) and negative coefficients (654, 721, 956, 1094, 1238, 1568, 1698, cm^{-1}). The variables that have the largest value of regression coefficient play a significant role in a regression model, and for saliva, it is 721 cm^{-1} with 0.05 weight and also 691, 1605 cm^{-1} (0.03 weight) and 1094 cm^{-1} (0.02 weight).

The PLS-DA calibration model of saliva is characterized by 97 % sensitivity, 90 % specificity, and 93 % accuracy. In the validation of 20 external samples, three of them stated as CoV(−) and one stated as CoV(+) had incorrect predictive values, resulting in 90 % sensitivity, 70 % specificity, and 80 % accuracy (Table 12). For most analysed samples, the values of standard deviation (STD) are between 0.15 and 0.34, demonstrating the robustness of the model created.

Table 12 Diagnostic parameters (sensitivity, specificity, accuracy) calculated at calibration (C) and validation (V) stage of PLS-DA, PCA-LDA, and SVMC analysis for saliva.

Type of sample	Analysis	Sensitivity (%)	Specificity (%)	Accuracy (%)	Number of samples
Saliva	PLS-DA	**C: 97.0**	**90.0**	**93.0**	**149**
		V: 90.0	70.0	80.0	20
	PCA-LDA	**C: 97.0**	**79.0**	**88.0**	**149**
		V: 100.0	60.0	80.0	20
	SVMC	**C: 100.0**	**100.0**	**100.0**	**149**
		V: 100.0	80.0	90.0	20

2. Principal component analysis – linear discriminant analysis (PCA-LDA)

To reduce the dimensionality of spectral data, LDA calculations were preceded by PCA analysis. The first 12 principal components explaining 98 % of spectral variance were used as input data for the subsequent LDA investigations. The values of diagnostic parameters were determined considering the recognition of each spectrum of the given map. The final decision to assign a sample was made on the basis of the majority of the spectra of that sample classified into a particular class. Hence, in accordance with the formula X=n+1 (where n is the number of spectra of a particular sample). Consequently, the created PCA-LDA calibration model provides 97 % sensitivity, 79 % specificity, and 88 % accuracy with respect to the total of 149 saliva samples (Table 12). At the stage of external validation, PCA-LDA misclassified four of ten CoV(−) saliva samples, resulting in 100 % sensitivity, 60 % specificity, and 80 % accuracy.

3. Support vector machine classification (SVMC)

SVMC calculations were performed with the use of polynomial Kernel function (degree three) and cross-validation with 10 segments. The parameters were selected in such a way as to ensure the best accuracy of the calibrated model and thus prevent overfitting. Under the conditions of C=0.1 and γ=0.1, the model provides 100 % of the training and validation accuracy. At the stage of external validation, only two of ten CoV(−) samples were misclassified, resulting in the highest values of diagnostic parameters among all tested methods – sensitivity 100 %, specificity 80 %, and accuracy 90 % (Table 12). Hence, SVMC is able to capture the highest relevant spectral information, ensuring good sample recognition, which makes it the most suitable method.

4.3.3 Spectral characterisation of CoV(+) and CoV(−) nasopharyngeal swabs

Nasopharyngeal secretion consists of sodium, potassium, calcium electrolytes, polysaccharides, and immunoglobulins (IgA, IgG, IgM) [248]. Typically, the composition includes also proteins, mainly albumin, and γA globulin but also γG globulin, siderophilin, α_1 glycoprotein, β_1C-β_1A globulin, haptoglobin, α_2-macroglobulin and β-lipoprotein [249,250]. The vital component is mucus with mucins [251].

Due to the high positive rate, nasopharyngeal swabs are recommended as a first choice for SARS-COV-2 testing in clinical practice. If this is not possible, an oropharyngeal swab, an anterior nares specimen, or a nasal aspirate becomes an alternative [252]. Therefore, the spectral analysis of nasopharyngeal swabs as a potential material for COVID-19 diagnosis should be considered.

The complete analysis involving nasopharyngeal swabs was performed on 120 patients, including 59 patients tested positive for SARS-CoV-2 (hereafter CoV(+)) and 61 patients tested negative for SARS-CoV-2 (hereafter CoV(−)). The SERS spectra of CoV(+) and CoV(−) nasopharyngeal swabs, along with SD and in superposition, are presented in Figure 42A and Figure 42B, respectively.

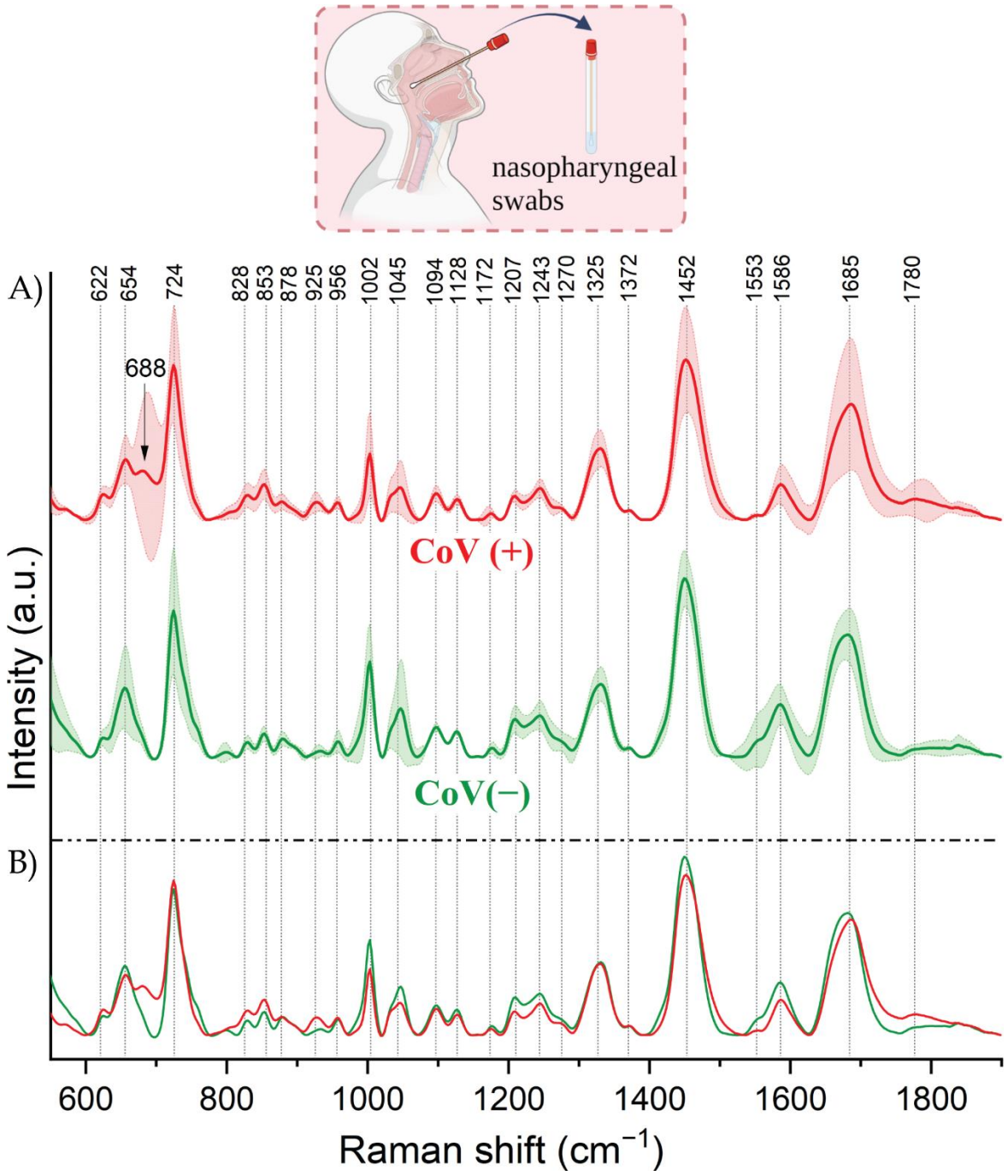

Figure 42 The averaged and normalized SERS spectra of CoV(+) and CoV(−) nasopharyngeal swabs presented A) with the standard deviation (SD) and B) as superimposed. The spectra were averaged from 51 CoV(+) and 53 CoV(−) samples. For every sample, *ca.* 15 single spectra were recorded.

Both analysed CoV(−) and CoV(+) nasopharyngeal swabs share intense bands located at 724, 1002, 1045, 1325, 1452, 1586, and 1685 cm^{-1}. These bands are the spectral representations of proteins, glycoproteins, purine bases of nucleic acids, lipids, and triglycerides [199,221]. The spectroscopic variations refer mainly to the relative band intensities in the entire region. The important trait is the 688 cm^{-1} band, which being unique

only to CoV(+) response, can hold a differentiating function. This band can be assigned to neopterin, 2-amino-4-oxo-6-(D-erythro-1, 2, 3-trihydroxypropyl)-pteridine [253]. Neopterin is an aromatic compound with nitrogen ring pyrazine-(2,3-d)-pyrimidine in its structure. In the human body, it is produced by monocytes and macrophages upon stimulation with interferon (IFN)-γ. Indeed, neopterin can be used as an inflammatory marker of the cell-mediated immune response, and its elevated level indicates the severity and progression of infection [254–256]. The current literature delivers information about the importance of neopterin in the COVID-19 course as a vital indicator in the immunopathological response to SARS-CoV-2 [257–259]. Hence, the aforementioned band that appears on the CoV(+) spectra may indicate the neopterin formation and accumulation as a result of this disease.

4.3.4 Chemometric analysis of the spectra of nasopharyngeal swabs

The chemometric analysis of nasopharyngeal swabs was performed using PLS-DA, PCA-LDA, and SVMC methods. In every case, calibration models were constructed using 104 samples (51 CoV(+) and 53 CoV(−)), and 16 samples (8 CoV(+) and 8 CoV(−)) were involved in external validation. For every sample *ca.* 15 single spectra were recorded, so the models were created from 1560 and tested on 240 spectra (Table 13).

Table 13 The participation of individual nasopharyngeal swabs on the calibration and validation step for all adopted methods (PLS-DA, PCA-LDA, SVMC).

Operation / **Material**	**Calibration**			**Validation**		
	No. of samples	No. of		No. of samples	No. of	
		samples	spectra		samples	spectra
CoV(+) nasopharyngeal swabs	51	104	1560	8	16	240
CoV(−) nasopharyngeal swabs	53			8		

1. Partial least squares discriminant analysis (PLS-DA)

The PLS-DA calibration model was created using Non-linear Iterative Partial Least Square (NIPALS) algorithm (cross-validation, random, 10-segments). The trained data fit the model with R^2_{cal} and R^2_{CV} equaling 0.66 and 0.63, respectively. The values of RMSEC and RMSECV were 0.29 and 0.30, respectively.

The 3D score plot (Figure 43A) discloses that all spectral data tend to create separated classes. The first latent variable (LV1) encompasses 27 % of the variance in block Y with 24 % of the spectral data (X matrix), and the second latent variable (LV2) covers 10 % with 15 % of the original data. The first thirteen variables, contributing 83 % (X matrix) with 67 % (Y matrix), fully describe the variability between the analysed data.

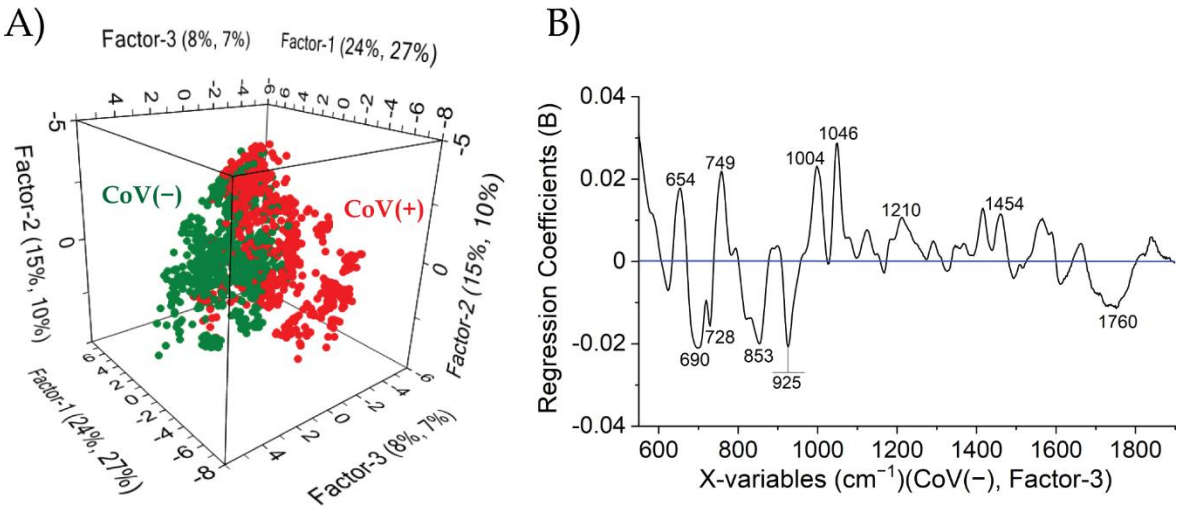

Figure 43 The results of PLSR analysis performed for 51 CoV(+) and 53 CoV(−) nasopharyngeal swabs in the form of A) 3D score plot and C) weighted regression coefficients (CoV(−) response, Factor-3). For each sample, *ca.* 15 single spectra were recorded and presented as dots on the score plot.

The plot of the weighted regression coefficient for Factor-3 with respect to CoV(−) response (Figure 43B) demonstrates variables with positive (654, 749, 1004, 1046, 1210, 1454 cm^{-1}) and negative (690, 728, 853, 925, 1760 cm^{-1}) coefficients. Among those mentioned, the ones at 1046, 1004, 690, 853, and 925 cm^{-1} are especially significant. According to the assignment, these bands indicate the high impact of macromolecules and proline ring-based compounds but also phenylalanine, tyrosine, cytosine, valine, proline, and glycogen [199,227,247].

The PLS-DA calibration model provided 100 % sensitivity, 96 % specificity, and 98 % accuracy. At the stage of external validation, three CoV(+) and two CoV(−) samples have incorrect predictive values, resulting in 63 % sensitivity, 75 % specificity, and 69 % accuracy (Table 14). The standard deviation usually ranges between 0.17 and 0.33.

Table 14 Diagnostic parameters (sensitivity, specificity, accuracy) calculated at calibration (C) and validation (V) stage of PLS-DA, PCA-LDA, and SVMC analysis for nasopharyngeal swabs.

Type of sample	Analysis	Sensitivity (%)	Specificity (%)	Accuracy (%)	Number of samples
Nasopharyngeal swabs	PLS-DA	**C: 100.0**	**96.0**	**98.0**	**104**
		V: 63.0	75.0	69.0	16
	PCA-LDA	**C: 96.0**	**83.0**	**89.0**	**104**
		V: 63.0	75.0	69.0	16
	SVMC	**C: 100.0**	**100.0**	**100.0**	**104**
		V: 88.0	63.0	75.0	16

2. Principal component analysis – linear discriminant analysis (PCA-LDA)

The original spectral data compressed to 15 principal components (by PCA) were used as input data for LDA calculations. The obtained results indicate that two of 51 CoV(+) and ten of 53 CoV(−) samples do not entirely match the typical spectral pattern of the majority of samples. Hence, the values of diagnostic parameters are 96 % (sensitivity), 83 % (specificity) and 89 % (accuracy). Since three of eight external samples stated as CoV(+) and two of eight external samples stated as CoV(−) were incorrectly classified, the sensitivity, specificity, and accuracy equal 63 %, 75 %, and 69 %, respectively (Table 14).

3. Support vector machine classification (SVMC)

In SVMC calculations, the optimal model was established with polynomial Kernel function (four degree), 10-segments cross-validation, and C=0.01, γ=0.1. With these optimised parameters, all 1560 spectra were correctly recognized (training and validation accuracy 100 %). Simulating actual diagnostic conditions, SVMC revealed an excellent sensitivity of 88 % with respect to eight CoV(+) patients. In turn, 63 % of specificity results from three of eight misclassified CoV(−) samples and is lower than for the previous two methods. However, the accuracy of 75 % is the highest among all tested methods (Table 14).

4.3.5 Conclusions

In this study, saliva and nasopharyngeal swabs were spectroscopically tested for their potential application as a clinical material for the recognition of SARS-CoV-2 infection. SERS data were meticulously analysed to extract differential spectral information. The indicated variability was related to personal issues only to a small extent. Significantly, they had a pathological basis, revealing all the processes involved during SARS-CoV-2 infection.

Thus, SERS spectra, being the molecular representation of material under analysis, provided valuable information about sample composition. Hence, the spectral response of CoV(+) saliva indicated an increased level of methionine, nucleic acids of DNA/RNA, proteins (*e.g.*, ferritin), and specific immunoglobulins. The CoV(+) nasopharyngeal swab unveils the significant role of neopterin in SERS signal formation.

The assistance of chemometric methods operated in supervised mode allowed for the development of a diagnostic approach working in a rapid and automated way. SVMC algorithm demonstrated the best sample recognition capabilities during model creation, reaching 100 % accuracy. The classification under real diagnostic conditions (external validation) resulted in satisfied diagnostic parameters – for saliva: sensitivity 100 %, specificity 80 %, and accuracy 90 %, and for nasopharyngeal swabs: sensitivity 88 %, specificity 63 %, and accuracy 75 %. Hence, both clinical materials are promising in the spectral analysis of CoV(−) and CoV(+) clinical samples.

Diagnosis with nasopharyngeal swabs requires close contact with patients and trained healthcare workers with the ability to collect samples precisely. The fundamental problems during testing are nasal obstruction, sneezing, and even bleeding. Saliva is more convenient material than nasopharyngeal swabs due to the non-invasive and cost-effective way of sampling. The patient is not exposed to discomfort, pain, or injury during intake. The possibility of sample self-collection reduces staff involvement and infection transmission. Saliva, which can be tested repeatedly and at any time, increases the chances of screening on a large scale [226,260].

The aforementioned aspects clearly show that saliva gains an advantage over nasopharyngeal swabs in diagnostic applications. It is also more convenient for spectroscopic analysis as, due to its homogeneous structure, it can be easily and evenly distributed on the SERS platform. SERS-based diagnostic tests with saliva as biomaterial can be a promising solution for COVID-19 diagnosis performed in a point-of-care variant.

Chapter 5
Summary and future prospects

The research presented in this dissertation constitutes a significant contribution to the practical application of SERS in diagnostics. The selected disease entities occur on a large scale, and their recognition using recommended methods does not always bring the expected results. An accurate diagnosis is crucial for implementing appropriate treatment and halting the disease progression, complications, and side effects. The SERS method, possessing many unprecedented features, has been repeatedly appreciated in biomedical research. This is evidenced by numerous scientific articles that reveal the ever-increasing practical application of SERS, implemented according to the label-free strategy or more complex assay as well. The first chapter of the dissertation provides only a basic overview of selected areas in which SERS demonstrates exceptional performance. However, it perfectly illustrates the interdisciplinary nature of the SERS technique.

The first section of the research was devoted to the analysis of men's urethral discharge to determine the spectral representation of gonorrhea. It was demonstrated that the spectral response of a clinical specimen taken from a patient with gonorrhea differs markedly from the control one. The Gon(+) spectrum is characterized by a significantly higher intensity and is richer in signals arising from bacterial cells that multiply due to the ongoing infection. The reduction of data dimensionality obtained with PCA method exposed correlations between classes and marker bands exhibiting diagnostic relevance. In turn, the supervised methods (PLS-DA, SIMCA) were used to predict outcomes and recognize spectral patterns. The established and evaluated models were tested in the external validation, demonstrating high reliability in sample recognition (up to 100 % of correctly classified spectra). The culture method with strain identification is frequently used in clinical practice for gonorrhea diagnosis. To overcome the limitations of existing techniques, the SERS-PCA methodology has been established to efficiently identify bacteria strains cultivated from urethral discharge.

Based on the unique spectral response, the strain was accurately identified as *N. gonorrhoeae*, proving the reliability of the new emerging method.

The second section concerned the comprehensive analysis of microorganisms (selected *Candida* spp. and bacterial pathogens) and vaginal fluids in the context of effective recognition of vaginal infections. The analysed vaginal fluids were categorised according to the aspects analysed, which included changes related to the fertile/infertile phase, microflora disorders, and the use of vaginal contraception or irrigation practice. The meticulous analysis provided insight into the correlation between biochemical changes and the signals, confirming the incredible sensitivity of the SERS method. The study also allowed to establish a control class, which is of purely practical significance and an elementary step for SERS as a future diagnostic method for vaginal infections. Research performed on a series of samples taken from women during infections (BV, VVC, gonorrhea, coinfection) demonstrated a significant impact of microbial cells on the SERS signal formation. The comparative analysis of spectral effects (menstrual cycle phase *vs.* vaginal infections) provided an estimate of their magnitude, thus indicating that the diagnosis can be made regardless of the day of the menstrual cycle. The information is of high value for subsequent analysis and the development of a reliable diagnostic method. In the last step, the spectral characterisation was performed for one of the most prevalent vaginal infections – VVC caused by the excessive growth of *Candida* spp. The previously defined control group served as an excellent reference for 12 samples taken from patients with VVC. The PLSR algorithm distinguished the analysed data, and the two classes thus created indicate the spectral manifestation of biochemical changes due to infection.

Section third presents the spectral analysis of two types of samples (saliva and nasopharyngeal swabs) in the context of SARS-CoV-2 infection. The research conducted on a large scale allowed for the inclusion of significant spectral variability related to the differences among patients, *e.g.* different course and responses of the organism to the disease. Thus, the averaged spectral signals represent the fingerprint of the entire composition of secretions altered and unaltered by the ongoing COVID-19. The use of chemometric methods allows for the visualization and extraction of any spectral differences and supports the understanding of SERS signal formation. In the case of saliva, a significant effect of methionine, nucleic acids, proteins, and specific immunoglobulins has been demonstrated. For nasopharyngeal swabs, the involvement of neopterin is important. Then, the external validation was conducted (20 saliva samples, 16 nasopharyngeal swabs) with

the previously optimised calibration models. The high sensitivity of SERS combined with chemometric methods (PLS-DA, PCA-LDA, and SVMC) delivered high-quality information about analysed system. Indeed, external validation is crucial in verifying the method as it demonstrates its performance in actual diagnostic conditions. The research indicated that the best-performing algorithm for the recognition of COVID-19 is SVMC, which provides up to 100% sensitivity.

The research presented in this dissertation addresses two aspects of special importance when optimizing SERS as a diagnostic method used in clinical settings. Namely, all studies were performed according to the label-free strategy that minimises the financial expenditure, time, and effort of qualified personnel. This approach enables the diagnosis of more patients in a short period of time while the non-invasiveness and low costs of analysis facilitate screening tests. Moreover, the research was conducted using a portable Raman spectrometer, which can be easily transported over various distances due to its size and weight. This solution is convenient and perfectly satisfies the requirements of point-of-care diagnostics, as well as the measurements taken in the field or at specially established points (*e.g.*, at entry points or border locations). All mentioned features combined with the speed of analysis (*ca.* 15 minutes) make SERS performed in the presented form a promising analytical tool with immense potential in disease diagnostics.

The presented research provides a comprehensive understanding of the formation of the spectral responses of clinical samples and addresses the issue of insufficient sensitivity of current diagnostic methods. Due to the complexity of the research topic and the large number of variables that potentially impact SERS measurements, several concepts need to be undertaken in future studies. For more extensive analyses, research can be performed on a larger scale involving hundreds or thousands of patients. Moreover, they can be conducted on the samples collected by several researchers from independent healthcare institutions to guarantee the highest confidence level in the results and investigate the potential sources of spectral variability. To maximize this effect, it would be advisable to consider the interlaboratory comparisons. This can be accomplished by sending samples from one patient to several laboratories for independent SERS analyses. Next, the obtained results are compared among these laboratories. In this aspect, it would certainly be crucial to standardise the entire measurement procedure, measurement conditions, equipment, and SERS platforms to eliminate factors that do not exist in the diagnostic context *per se*.

Chapter 6
Appendix

1. *The biochemical characteristics of the analysed vaginal fluids.*

The tables below provide detailed information about analysed vaginal fluids that were arranged in considered associations. All the information was collected during the diagnosis of patients.

Table 1A The characteristics of analysed vaginal fluids taken from patients of the control group and those with imbalanced microflora.

Type and number of samples	Age	pH	Characteristics	Purity	Remarks
	Control group				
Control group (21 samples)	24 – 34	4.2 – 4.7	lactobacilli (many), epithelial cells	I, II	–
	Samples with altered biochemistry/microflora				
Decreased purity (19 samples)	18 – 64	4.7 – 5.6 (elevated level)	usually 10 – 20 leucocytes, bacilli Gram-positive/Gram-negative or cocci Gram-positive (many), lactobacilli (usually few or lack)	II/III or III	–
Vaginal irrigation (1 sample)	18	–	Non-diagnostic sample	–	absence of typical discharge
Loss of normal microflora (1 sample)	62	–	Leucocytes (lack), lactobacilli (lack), cocci Gram-positive (single), epithelial cells (single)	–	no additional assessment
Vaginal ring with hormones (1 sample)	25	5	10 – 20 leucocytes, lactobacilli (lack), cocci Gram-positive (many)	–	large purulent discharge

Table 2A The characteristics of analysed vaginal fluids belonging to IC (Infected cohort).

Type of sample	Age	pH	Characteristics	Purity	Remarks
VVC1	–	5.2	4 – 10 leucocytes, lactobacilli (lack), Gram-positive bacilli, Gram-positive cocci (quite numerous), blastospores (very numerous)	IV	Culture method – abundant growth of *Candida albicans*, Non-painful abrasions
VVC2	30	5.0	10 – 20 leucocytes, lactobacilli (lack), Gram-positive bacilli (very numerous), pseudohyphae of *Candida* spp.	IV	–
VVC3	18	5.0	10 – 20 leucocytes, Gram-positive cocci, blastospores	IV	After 2 weeks of using globules, ectopic pregnancy
BV1	–	5.0	10 – 20 leucocytes, lactobacilli (lack), Gram-negative bacilli and Gram-positive cocci (very numerous)	III	Condylomas
BV2	–	5.6	4 – 10 leucocytes, lactobacilli (lack), Gram-negative bacilli and Gram-positive cocci (very numerous)	III	Condylomas
BV3	38	6.0	10 – 20 leucocytes, lactobacilli (lack), Gram-positive bacilli, Gram-negative bacilli and Gram-positive cocci (very numerous), *Gardnerella vaginalis* (very numerous)	III	–
Coinf.	54	5.0	4 – 10 leucocytes, lactobacilli (very numerous), Gram-positive bacilli, Gram-negative bacilli and Gram-positive cocci, Gram-negative cocci (very numerous), *Gardnerella vaginalis*	IV	*Candida albicans* (culture method)
Gonorrhea	–	–	Gram-negative cocci inside leucocytes (indicative of *Neisseria gonorrhoeae)*	IV	–

Table 3A The characteristics of analysed vaginal fluids taken from patients of the control group and those diagnosed with vulvovaginal candidiasis (VVC).

Type of sample	Age	pH	Characteristics	Purity	Remarks
Control group					
Control group (21 samples)	24 – 34	4.2 – 4.7	lactobacilli (many), epithelial cells	I-II	–
Vulvovaginal Candidiasis group					
VVC1	–	4.8	10-20 leucocytes, lactobacilli (quite numerous), bacilli Gram-negative (very numerous)	IV	Culture method – abundant growth of *Candida albicans*
VVC2	24	5.0	4-10 leucocytes, lactobacilli (numerous), bacilli Gram-positive (quite numerous), blastospores (numerous)	IV	Culture method – very abundant growth of *Candida tropicalis*
VVC3	49	4.7	lactobacilli (numerous), blastospores (very numerous)	IV	Culture method – very abundant growth of *Candida albicans*
VVC4	–	5.2	4-10 leucocytes, lactobacilli (lack), bacilli Gram-positive and cocci Gram-positive (quite numerous), blastospores (very numerous)	IV	Culture method – abundant growth of *Candida albicans*, Non-painful abrasions
VVC5	–	4.7	4-10 leucocytes, lactobacilli (numerous), pseudohyphae	IV	–
VVC6	18	5.0	10-20 leucocytes, cocci Gram-positive, blastospores	IV	After 2 weeks of using globules, ectopic pregnancy
VVC7	69	5.0	0 leucocytes, lactobacilli (singles), cocci Gram-positive (singles), epithelial cells (singles)	IV	Culture method – abundant growth of *Candida tropicalis* and *Candida glabrata*, IUD applied to support the prolapsed uterus, dry vaginal environment
VVC8	61	5.0	0-4 leucocytes, lactobacilli (lack), bacilli Gram-positive (very numerous), bacilli Gram-negative (quite numerous)	IV	Culture method – growth of *Candida* spp., atypical smear – no inflammation, flora is mediocre
VVC9	30	5.0	10-20 leucocytes, lactobacilli (lack),	IV	–

			bacilli Gram-positive (very numerous), pseudohyphae		
VVC10	63	4.7	4-10 leucocytes, lactobacilli (quite numerous), bacilli Gram-positive (quite numerous), bacilli Gram-negative (few), cocci Gram-positive (singles)	IV	Culture method – growth of *Candida* spp.
VVC11	39	5.0	4-10 leucocytes, lactobacilli (lack), bacilli Gram-positive (numerous), blastospores	IV	Culture method – growth of *Candida tropicalis*
VVC12	53	5.0	4-10 leucocytes, lactobacilli (lack), bacilli Gram-positive (numerous), bacilli Gram-negative (quite numerous), blastospores	IV	After treatment of vaginal infections with clindamycin

2. *The SERS spectra of vaginal secretions (i.e. menstrual blood, menstrual spotting, vaginal fluid) collected over two menstrual cycle.*

Figure 1A compiles the SERS spectra averaged within the single sample (including vaginal fluid – orange, menstrual blood – red, menstrual spotting – brown spectra) of the first (27 days) and second (24 days) menstrual cycle. Since SERS measurements were performed on two, three, or four independent samples daily, the amount of data is multiplicated – 83 (first cycle) and 62 (second cycle) samples in total analysis.

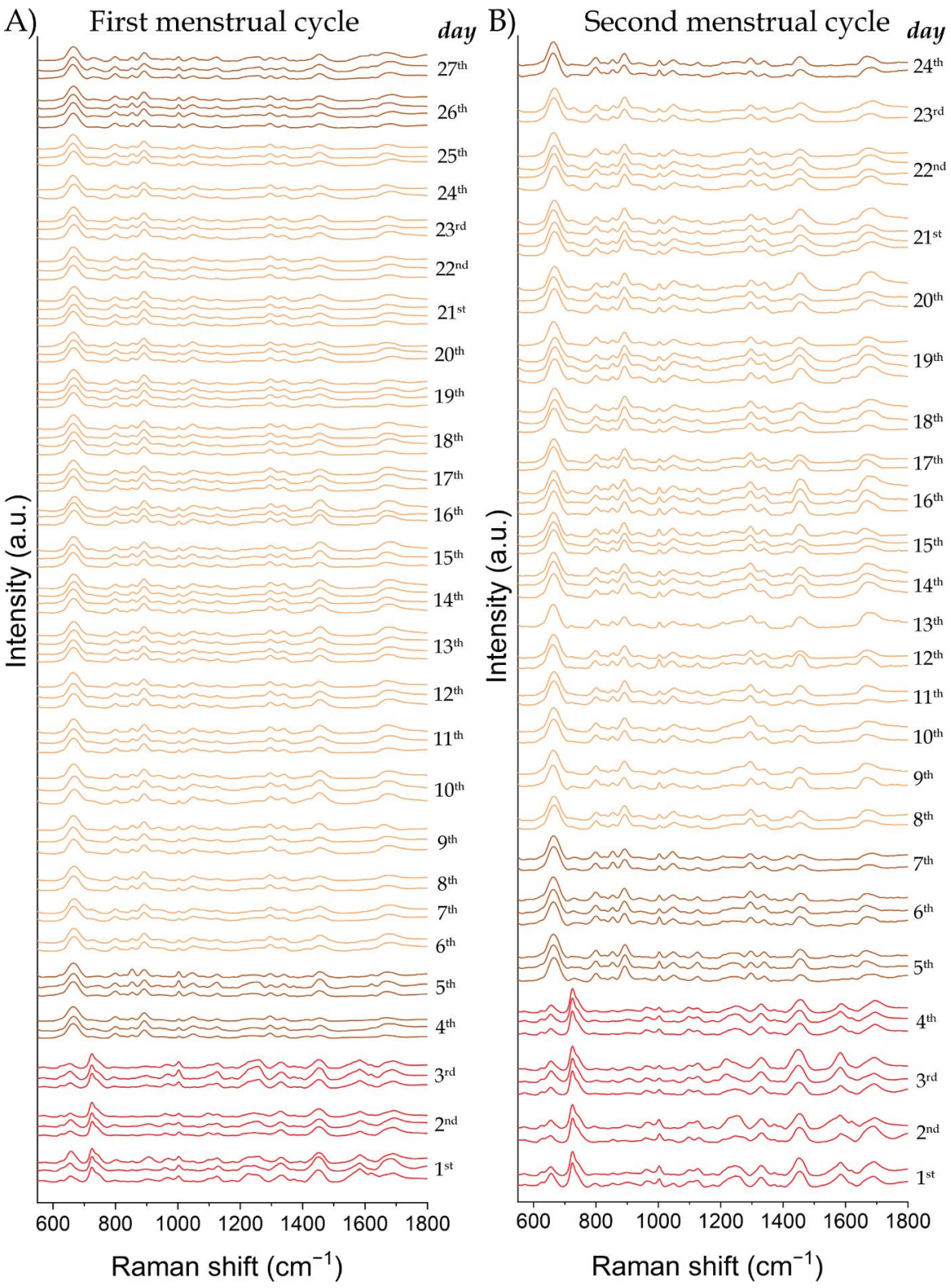

Figure 1A The averaged SERS spectra of body fluids taken from the patient's vagina within the entire A) first (83 samples of 27 days) and B) second (62 samples of 24 days) menstrual cycle. The colour of spectra indicates the type of analysed material: red – menstrual blood, brown – menstrual spotting, orange – vaginal fluid. The samples were collected two, three, or four times every morning. Presented spectra were averaged within a single sample (for every sample, *ca.* 15 spectra were recorded).

Chapter 7
Scientific achievements

7.1 Articles

9. S. M. Berus*, T. Szymborski, B. Młynarczyk-Bonikowska, G. Przedpełska, M. Adamczyk-Popławska, A. Kamińska, **Identifying changes in vaginal fluid using SERS: Advancing diagnosis of Vulvovaginal Candidiasis**, Sensing and Biosensing Research (2024) 100702. doi: 10.1016/j.sbsr.2024.100702
* corresponding author

8. S. M. Berus*, B. Młynarczyk-Bonikowska, G. Przedpełska, T. Szymborski, A. Kamińska, **Surface-enhanced Raman spectroscopy analysis of menstrual cycle: from biochemical changes to diagnostics of vaginal infections**, Sensors and Actuators B: chemical 407 (2024) 135571. doi:10.1016/j.snb.2024.135571
* corresponding author

7. T. Szymborski#, S. M Berus#, A. Nowicka, G. Słowiński, A. Kamińska, **Machine Learning for COVID-19 Determination Using Surface-Enhance Raman Spectroscopy**, Biomedicines 12 (2024) 167. doi:10.3390/biomedicines12010167
# equal contribution

6. S. M. Berus, A. Nowicka, J. Wieruszewska, K. Niciński, A.A. Kowalska, T. Szymborski, I. Dróżdż, M. Borowiec, J. Waluk, A. Kamińska, **SERS Signature of SARS-CoV-2 in Saliva and Nasopharyngeal Swabs: Towards Perspective COVID-19 Point-of-Care Diagnostics**, International Journal of Molecular Sciences 24 (2023) 9706. doi:10.3390/ijms24119706

5. S. M. Berus*#, M. Adamczyk-Popławska#, K. Goździk, G. Przedpełska, T. Szymborski, Y. Stepanenko, A. Kamińska, **SERS-PLSR Analysis of Vaginal Microflora: Towards the Spectral Library of Microorganisms**, International Journal of Molecular Sciences 23 (2022) 12576. doi:10.3390/ijms232012576.
* corresponding author, # equal contribution

4. S. M. Berus, M. Adamczyk-Popławska, B. Młynarczyk-Bonikowska, E. Witkowska, T. Szymborski, J. Waluk, A. Kamińska, **SERS-based sensor for the detection of sexually transmitted pathogens in the male swab specimens: a new approach for clinical diagnosis**, Biosensors and Bioelectronics 189 (2021) 113358. doi:10.1016/j.bios.2021.113358.

3. T. Szymborski, Y. Stepanenko, K. Niciński, P. Piecyk, S. M. Berus, M. Adamczyk-Popławska, A. Kamińska**, Ultrasensitive SERS platform made *via* femtosecond laser micromachining for biomedical applications**, Journal of Material Research and Technology 12 (2021) 1496–1507. doi:10.1016/j.jmrt.2021.03.083.

2. S. M. Berus, E. Witkowska, K. Niciński, E. Sadowy, W. Puzia, P. Ronkiewicz, A. Kamińska, **Surface-enhanced Raman scattering as a discrimination method of Streptococcus spp. and alternative approach for identifying capsular types of *S. pneumoniae* isolates**, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 233 (2020) 118088. doi:10.1016/j.saa.2020.118088.

1. A. A. Kowalska, S. M. Berus, Ł. Szleszkowski, A. Kamińska, A. Kmiecik, K. Ratajczak-Wielgomas, T. Jurek, Ł. Zadka, **Brain tumour homogenates analysed by surface-enhanced Raman spectroscopy: Discrimination among healthy and cancer cells**, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 231 (2020) 117769. doi:10.1016/j.saa.2019.117769.

7.2 Patent application

P.436251, 2020, **The method for determining the presence of venereal diseases pathogens,** S. M. Berus, B. Młynarczyk-Bonikowska, M. Adamczyk-Popławska, T. Szymborski, E. Witkowska-Iwan, A. Kamińska

7.3 Oral presentations

13. Invited talk at Doctoral seminar 2024 of Leibnitz Institute of Photonic Technology (IPHT)
18-21 March 2024, Dornburg, Germany, stationary
Surface-enhanced Raman spectroscopy in medical applications
S. M. Berus, B. Młynarczyk-Bonikowska, G. Przedpełska, M. Adamczyk-Popławska, A. Kowalska, T. Szymborski, E. Witkowska, A. Nowicka, J. Waluk, A. Kamińska

12. Invited talk at Virtual PhD seminar IPHT-IChF Leibnitz Institute of Photonic Technology (IPHT) in Jena and Institute of Physical Chemistry of the Polish Academy of Sciences
15 November 2023, virtual
Surface-enhanced Raman spectroscopy in analysis of clinical samples
S. M. Berus, B. Młynarczyk-Bonikowska, G. Przedpełska, M. Adamczyk-Popławska, T. Szymborski, J. Waluk, A. Kamińska

11. 8th Asian Spectroscopy Conference
03-07 September 2023, Tokamachi in Niigata, Japan, stationary
Spectral recognition of vaginal fluids alteration: towards SERS-based differential diagnosis of vaginal infections
S. M. Berus, B. Młynarczyk-Bonikowska, G. Przedpełska, T. Szymborski, A. Kamińska

10. ICACAS 2022: 16. International Conference on Analytical Chemistry and Applied Spectroscopy
06-07 October 2022, Tokyo, Japan, virtual
Spectral Recognition of Vaginal Fluids Alterations: The Integration of a Handheld Raman Device and a SERS-Based Sensor for Diagnosis of Vaginal Infections
S. M. Berus, B. Młynarczyk-Bonikowska, M. Adamczyk-Popławska, T. Szymborski, A. Kamińska

9. ICAVS 2022: 16. International Conference on Advanced Vibrational Spectroscopy
27-28 September 2022, Dubai, United Arab Emirates, virtual
SERS-based analysis assisted by chemometric methods towards differential diagnosis of vaginal infections
S. M. Berus, B. Młynarczyk-Bonikowska, M. Adamczyk-Popławska, A. Kamińska

8. RamanFest
15-16 September 2022, Paris, France, stationary
SERS-based study supported by chemometric methods for the differential diagnosis of vaginal infections
S. M. Berus, B. Młynarczyk-Bonikowska, M. Adamczyk-Popławska, T. Szymborski, A. Kamińska

7. European Symposium on Analytical Spectrometry Czech-Slovak Spectroscopic Conference ESAS CSSC
4-10 September 2022, Brno, Czech Republic, stationary
The integration of a handheld Raman device and a SERS-based sensor for spectral recognition of vaginal fluids alterations
S. M. Berus, B. Młynarczyk-Bonikowska, M. Adamczyk-Popławska, T. Szymborski, A. Kamińska

6. ICLSA 2022: 16. International Conference on Laser Spectroscopy and Applications,
16-17 August 2022, Bangkok, Thailand, virtual
SERS-based studies supported by PLSR analysis to determine spectral marker bands of microorganisms present in the human vagina
S. M. Berus, M. Adamczyk-Popławska, T. Szymborski, A. Kamińska

5. ICSSA 2022: 16. International Conference on Spectroscopy and Spectral Analysis
7-8 August 2022, Vancouver, Canada, virtual
SERS-based analysis of vaginal inhabiting microorganisms: towards the first database of spectral fingerprints
S. M. Berus, M. Adamczyk-Popławska, T. Szymborski, A. Kamińska

4. **Invited talk** at IPC-PAS Microsymposium 2022
25-26 January 2022, virtual
SERS-based sensor for the detection of sexually transmitted pathogens in the male swab specimens: a new approach for clinical diagnosis
S. M. Berus, M. Adamczyk-Popławska, B. Młynarczyk-Bonikowska, E. Witkowska-Iwan, T. Szymborski, J. Waluk, A. Kamińska

3. ICMS 2022: XVI International Conference on Molecular Spectroscopy

7-8 January 2022, Tokyo, Japan, virtual
SERS-based study supported by chemometric methods for the diagnosis of COVID-19 in clinical samples
S. M. Berus, A. Nowicka, J. Wieruszewska, K. Niciński, A. Kowalska, I. Dróżdż, M. Borowiec, A. Kamińska

2. SSPTChem 2021 Spring e-Conference
27-29 May 2021, virtual
SERS analysis of clinical materials for the diagnosis of sexually transmitted diseases
S. M. Berus, M. Adamczyk-Popławska, B. Młynarczyk-Bonikowska, E. Witkowska-Iwan, T. Szymborski, J. Waluk, A. Kamińska

1. National Student scientific Conference 'Closer to Chemistry' UJ
9-10 January 2021, virtual
Surface-enhanced Raman spectroscopy (SERS) as a tool for detection and identification of bacteria present in saliva
S. M. Berus, E. Witkowska-Iwan, D. Korsak, P. Ronkiewicz, A. Kamińska

7.4 Poster presentations

12. IPC PAS microsymposium 2025
13-15 January 2025, Warsaw, Poland, stationary
SERS biosensors for medical analysis
S. M. Berus, B. Młynarczyk-Bonikowska, G. Przedpełska, T. Szymborski, A. Kamińska

11. ICHF-IPHT Networking Meeting Agenda
4-5 November 2024, Warsaw, Poland, stationary
Spectral recognition of vaginal fluids alteration: towards SERS-based differential diagnosis of vaginal infections
S. M. Berus, B. Młynarczyk-Bonikowska, G. Przedpełska, T. Szymborski, A. Kamińska

10. IPC PAS Microsymposium 2024
9-10 January 2024, Warsaw, Poland, stationary
Spectral recognition of vaginal fluids alteration: towards SERS-based differential diagnosis of vaginal infections
S. M. Berus, B. Młynarczyk-Bonikowska, G. Przedpełska, T. Szymborski, A. Kamińska

9. 8th Asian Spectroscopy Conference
03-07 September 2023, Tokamachi in Niigata, Japan
Spectral recognition of vaginal fluids alteration: towards SERS-based differential diagnosis of vaginal infections
S. M. Berus, B. Młynarczyk-Bonikowska, G. Przedpełska, T. Szymborski, A. Kamińska
and
SERS-based sensor for the detection of sexually transmitted pathogens in the male swab specimens: A new approach for clinical diagnosis

S. M. Berus, M. Adamczyk-Popławska, B. Młynarczyk-Bonikowska, E. Witkowska-Iwan, T. Szymborski, J. Waluk, A. Kamińska

8. IPC PAS Microsymposium 2023
17-18 January 2023, Warsaw, Poland, virtual
SERS signature of SARS-CoV-2 infected saliva and nasopharyngeal swabs: towards perspective COVID-19 point-of-care diagnostics
S. M. Berus, A. Nowicka, J. Wieruszewska, K. Niciński, A. A. Kowalska, I. Dróżdż, M. Borowiec, J. Waluk, A. Kamińska

7. 4th Interdisciplinary FNP conference
6-7 October 2022, Warsaw, Poland, stationary
Future diagnOstic RaMan devIce for detection of medical and environmental pathogenic bacteria
S. M. Berus, K. Niciński, A. A. Kowalska, T. Szymborski, E. Witkowska, A. Nowicka, J. Wieruszewska, P. Piecyk, B. Dominiak, J. Waluk, A. Kamińska

6. AMLCI Materials Day 2021 „Materials for Energy and Sensing" Kent State University
9-10 April, 2021, virtual
SERS-based sensor for the detection of sexually transmitted pathogens in the male swab specimens: a new approach for gonorrhea diagnosis
S. M. Berus, M. Adamczyk-Popławska, B. Młynarczyk-Bonikowska, E. Witkowska-Iwan, T. Szymborski, J. Waluk, A. Kamińska

5. IPC PAS Microsymposium 2021
19-20 January 2021, Warsaw, Poland, virtual
Diagnosis of sexually transmitted diseases by means of SERS along with supervised and unsupervised pattern recognition methods
S. M. Berus, E. Witkowska, K. Niciński, A. Skoczyńska, P. Ronkiewicz, A. Kamińska

4. IPC PAS Microsymposium 2020
15-16 January 2020, Warsaw, Poland, stationary
Surface-enhanced Raman scattering as a discrimination method of Streptococcus spp. and alternative approach for identifying capsular types of S. pneumoniae isolates
S. M. Berus, E. Witkowska, K. Niciński, A. Skoczyńska, P. Ronkiewicz, A. Kamińska

3. Polish Photoscience Seminar, Świętokrzyska Polana
16-19 June 2019, Chrusty, Poland, stationary
Brain tumors analysed by Surface-Enhanced Raman Spectroscopy (SERS): discrimination among healthy and cancer cells
S. M. Berus, A. A. Kowalska, T. Szymborski, Ł. Zadka, Ł. Szlaszkowski, R. Jankowski, A. Kamińska

2. Chemsession'19 – XVI Warsaw Chemistry Doctoral Seminar
7 June 2019, Warsaw, Poland, stationary
SERS-based differentiation of pneumococci at species and strain level
S. M. Berus, E. Witkowska, K. Niciński, A. Skoczyńska, P. Ronkiewicz, A. Kamińska

1. IPC-PAS Microsymposium 2019
15-17 January, 2019, Warsaw, Poland, stationary
Brain tumors analysed by Surface-Enhanced Raman Spectroscopy (SERS): discrimination among healthy and cancer cells
S. M. Berus, A. A. Kowalska, T. Szymborski, Ł. Zadka, Ł. Szleszkowski, R. Jankowski, A. Kamińska

7.5 Scientific grants

1. PRELUDIUM-20, National Science Center, 01/2022 – 01/2024, position: **grant holder**

   2021/41/N/ST4/02164, **SERS-based study combined with chemometric methods for recognition of the molecular pattern of vaginal fluids: towards a differential diagnosis of vaginal infections**

2. Team-Tech/2017-4/23, Foundation for Polish Science, 10/2018 – 04/2021, position: grant researcher

   POIR.04.04.00-00-4210/17-00, **Future diagnostic Raman device for detection of medical and environmental pathogenic bacteria**

7.6 Awards and prizes

- The Best poster presenter at Materials Day 2021, Kent State University, 9-10 April, 2021
- The Best poster presenter at Polish Photoscience Seminar, Świętokrzyska Polana, 16-19 June 2019, Chrusty, Poland
- The best presentation at ICMS 2022: XVI International Conference on Molecular Spectroscopy, 7-8 January 2022, Tokyo, Japan
- Young Researches of IPC PAS, September 2022
- Young Researches of IPC PAS, September 2023
- Young Researches of IPC PAS, September 2024

List of Figures

List of Figures in the appendix

List of Tables

List of Tables in the appendix

List of Equations

Literature

[1] C. V Raman, F.R.S. Plate, A new radiation, Indian J. Phys. 2 (1928) 387–398.

[2] R. Singh, C. V. Raman and the Discovery of the Raman Effect, Phys. Perspect. 4 (2002) 399–420. doi:10.1007/S000160200002/METRICS.

[3] C. Lima, H. Muhamadali, R. Goodacre, The Role of Raman Spectroscopy within Quantitative Metabolomics, Annu. Rev. Anal. Chem. 14 (2021) 323–345. doi:10.1146/annurev-anchem-091420-092323.

[4] J. Bohning, The Raman Effect, Am. Chem. Soc. (1998). https://www.acs.org/education/whatischemistry/landmarks/ramaneffect.html (accessed August 10, 2024).

[5] Q. Tu, C. Chang, Diagnostic applications of Raman spectroscopy, Nanomedicine Nanotechnology, Biol. Med. 8 (2012) 545–558. doi:10.1016/J.NANO.2011.09.013.

[6] R.S. Das, Y.K. Agrawal, Raman spectroscopy: Recent advancements, techniques and applications, Vib. Spectrosc. 57 (2011) 163–176. doi:10.1016/J.VIBSPEC.2011.08.003.

[7] A. Adamski, M. Barańska, A. Bodzoń-Kułakowska, K. Chruszcz-Lipska, Ł. Hetmańczyk, A. Kaczor, A. Kraj, K. Małek, A. Migdał-Mikuli, E. Mikuli, M. Noga, P. Pietrzyk, E. Podstawka, L.M. Proniewicz, J. Silberring, Z. Sojka, P. Suder, J. Szklarzewicz, A. Wesełucha-Birczyńska, Wybrane metody spektroskopii i spektrometrii molekularnej w analizie strukturalnej, I, Jagiellonian University, 2005.

[8] D. Cialla-May, M. Schmitt, J. Popp, Theoretical principles of Raman spectroscopy, Phys. Sci. Rev. 4 (2019) 1–14. doi:10.1515/9783110515312-001.

[9] B. Schrader, Chemical Applications of Raman Spectroscopy, Angew. Chemie Int. Ed. 12 (1973) 884–908. doi:10.1002/ANIE.197308841.

[10] R. Kiselev, Towards clinical translation of Raman spectroscopy for tumor cell identification, 2019. doi:10.22032/DBT.38649.

[11] A. Orlando, F. Franceschini, C. Muscas, S. Pidkova, M. Bartoli, M. Rovere, A. Tagliaferro, A Comprehensive Review on Raman Spectroscopy Applications, Chemosensors. 9 (2021) 262. doi:10.3390/chemosensors9090262.

[12] K. Eberhardt, C. Stiebing, C. Matthaüs, M. Schmitt, J. Popp, Advantages and limitations of Raman spectroscopy for molecular diagnostics: an update, Expert Rev. Mol. Diagn. 15 (2015) 773–787. doi:10.1586/14737159.2015.1036744.

[13] P. Rostron, S. Gaber, D. Gaber, Raman Spectroscopy , Review, Int. J. Eng. Tech. Res. 6 (2016) 50–64.

[14] Hana Vašková, A powerful tool for material identification: Raman spectroscopy, Int. J. Math. Model. Methods Appl. Sci. 5 (2011) 1205–1212.

[15] A. Saletnik, B. Saletnik, C. Puchalski, Overview of Popular Techniques of Raman Spectroscopy and Their Potential in the Study of Plant Tissues, Molecules. 26 (2021)

1537. doi:10.3390/molecules26061537.

[16] D.R. Dietze, R.A. Mathies, Femtosecond Stimulated Raman Spectroscopy, ChemPhysChem. 17 (2016) 1224–1251. doi:10.1002/CPHC.201600104.

[17] M. Fleischmann, P.J. Hendra, A.J. McQuillan, Raman spectra of pyridine adsorbed at a silver electrode, Chem. Phys. Lett. 26 (1974) 163–166. doi:10.1016/0009-2614(74)85388-1.

[18] D.L. Jeanmaire, R.P. Van Duyne, Surface raman spectroelectrochemistry: Part I. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode, J. Electroanal. Chem. Interfacial Electrochem. 84 (1977) 1–20. doi:10.1016/S0022-0728(77)80224-6.

[19] M.G. Albrecht, J.A. Creighton, Anomalously Intense Raman Spectra of Pyridine at a Silver Electrode, J. Am. Chem. Soc. 99 (1977) 5215–5217. doi:10.1021/ja00457a071.

[20] Deepak F. L, Metal Nanoparticles and Clusters, Springer, 2018. doi:10.1007/978-3-319-68053-8.

[21] X.X. Han, R.S. Rodriguez, C.L. Haynes, Y. Ozaki, B. Zhao, Surface-enhanced Raman spectroscopy, Nat. Rev. Methods Prim. 1 (2022) 1–17. doi:10.1038/s43586-021-00083-6.

[22] R. Pilot, R. Signorini, C. Durante, L. Orian, M. Bhamidipati, L. Fabris, A Review on Surface-Enhanced Raman Scattering, Biosensors. 9 (2019) 57. doi:10.3390/BIOS9020057.

[23] C. Zong, M. Xu, L.-J. Xu, T. Wei, X. Ma, X.-S. Zheng, R. Hu, B. Ren, Surface-Enhanced Raman Spectroscopy for Bioanalysis: Reliability and Challenges, Chem. Rev. 118 (2018) 4946–4980. doi:10.1021/acs.chemrev.7b00668.

[24] W. Zhu, R. Esteban, A.G. Borisov, J.J. Baumberg, P. Nordlander, H.J. Lezec, J. Aizpurua, K.B. Crozier, Quantum mechanical effects in plasmonic structures with subnanometre gaps, Nat. Commun. 7 (2016) 1–14. doi:10.1038/ncomms11495.

[25] Y. Liu, M. Kim, S.H. Cho, Y.S. Jung, Vertically aligned nanostructures for a reliable and ultrasensitive SERS-active platform: Fabrication and engineering strategies, Nano Today. 37 (2021) 101063. doi:10.1016/J.NANTOD.2020.101063.

[26] E.C. Le Ru, B. Auguié, Enhancement Factors: A Central Concept during 50 Years of Surface-Enhanced Raman Spectroscopy, ACS Nano. 18 (2024) 9773–9783. doi:10.1021/acsnano.4C01474.

[27] S. Efrima, L. Zeiri, Understanding SERS of bacteria, J. Raman Spectrosc. 40 (2009) 277–288. doi:10.1002/jrs.2121.

[28] A.I. Pérez-Jiménez, D. Lyu, Z. Lu, G. Liu, B. Ren, Surface-enhanced Raman spectroscopy: benefits, trade-offs and future developments, Chem. Sci. 11 (2020) 4563–4577. doi:10.1039/D0SC00809E.

[29] L. Jiang, M.M. Hassan, S. Ali, H. Li, R. Sheng, Q. Chen, Evolving trends in SERS-based techniques for food quality and safety: A review, Trends Food Sci. Technol. 112 (2021) 225–240. doi:10.1016/J.TIFS.2021.04.006.

[30] B. Zhou, W. Ou, C. Zhao, J. Shen, G. Zhang, X. Tang, Z. Deng, G. Zhu, Y.Y. Li, J. Lu, Insertable and reusable SERS sensors for rapid on-site quality control of fish and meat products, Chem. Eng. J. 426 (2021) 130733. doi:10.1016/J.CEJ.2021.130733.

[31] H. Li, Q. Chen, M.M. Hassan, Q. Ouyang, T. Jiao, Y. Xu, M. Chen, AuNS@Ag core-shell nanocubes grafted with rhodamine for concurrent metal-enhanced fluorescence and surfaced enhanced Raman determination of mercury ions, Anal. Chim. Acta. 1018 (2018) 94–103. doi:10.1016/J.ACA.2018.01.050.

[32] C. Wang, J. Wang, M. Li, X. Qu, K. Zhang, Z. Rong, R. Xiao, S. Wang, A rapid SERS method for label-free bacteria detection using polyethylenimine-modified Au-coated magnetic microspheres and Au@Ag nanoparticles, Analyst. 141 (2016) 6226–6238. doi:10.1039/c6an01105e.

[33] B. Zhao, S. Feng, Y. Hu, S. Wang, X. Lu, Rapid determination of atrazine in apple juice using molecularly imprinted polymers coupled with gold nanoparticles-colorimetric/SERS dual chemosensor, Food Chem. 276 (2019) 366–375. doi:10.1016/J.FOODCHEM.2018.10.036.

[34] V. Moreno, A. Adnane, R. Salghi, M. Zougagh, Á. Ríos, Nanostructured hybrid surface enhancement Raman scattering substrate for the rapid determination of sulfapyridine in milk samples, Talanta. 194 (2019) 357–362. doi:10.1016/J.TALANTA.2018.10.047.

[35] A.I. Radu, M. Kuellmer, B. Giese, U. Huebner, K. Weber, D. Cialla-May, J. Popp, Surface-enhanced Raman spectroscopy (SERS) in food analytics: Detection of vitamins B2 and B12 in cereals, Talanta. 160 (2016) 289–297. doi:10.1016/J.TALANTA.2016.07.027.

[36] V. Dugandžić, S. Kupfer, M. Jahn, T. Henkel, K. Weber, D. Cialla-May, J. Popp, A SERS-based molecular sensor for selective detection and quantification of copper(II) ions, Sensors Actuators B Chem. 279 (2019) 230–237. doi:10.1016/J.SNB.2018.09.098.

[37] C. Camerlingo, M. Portaccio, R. Tatè, M. Lepore, I. Delfino, Fructose and pectin detection in fruit-based food products by surface-enhanced Raman spectroscopy, Sensors. 17 (2017) 1–12. doi:10.3390/s17040839.

[38] L.H. Oakley, S.A. Dinehart, S.A. Svoboda, K.L. Wustholz, Identification of organic materials in historic oil paintings using correlated extractionless surface-enhanced Raman scattering and fluorescence microscopy, Anal. Chem. 83 (2011) 3986–3989. doi:10.1021/ac200698Q.

[39] C.L. Brosseau, K.S. Rayner, F. Casadio, C.M. Grzywacz, R.P. Van Duyne, Surface-enhanced Raman spectroscopy: A direct method to identify colorants in various artist media, Anal. Chem. 81 (2009) 7443–7447. doi:10.1021/ac901219m.

[40] S. Bruni, V. Guglielmi, F. Pozzi, Surface-enhanced Raman spectroscopy (SERS) on silver colloids for the identification of ancient textile dyes: Tyrian purple and madder, J. Raman Spectrosc. 41 (2010) 175–180. doi:10.1002/JRS.2456.

[41] B. Sharma, R.R. Frontiera, A.I. Henry, E. Ringe, R.P. Van Duyne, SERS: Materials, applications, and the future, Mater. Today. 15 (2012) 16–25. doi:10.1016/S1369-

7021(12)70017-2.

[42] A. Hakonen, P.O. Andersson, M. Stenbæk Schmidt, T. Rindzevicius, M. Käll, Explosive and chemical threat detection by surface-enhanced Raman scattering: A review, Anal. Chim. Acta. 893 (2015) 1–13. doi:10.1016/J.ACA.2015.04.010.

[43] L. Yang, L. Ma, G. Chen, J. Liu, Z.Q. Tian, Ultrasensitive SERS Detection of TNT by Imprinting Molecular Recognition Using a New Type of Stable Substrate, Chem. – A Eur. J. 16 (2010) 12683–12693. doi:10.1002/CHEM.201001053.

[44] T.T.X. Ong, E.W. Blanch, O.A.H. Jones, Surface Enhanced Raman Spectroscopy in environmental analysis, monitoring and assessment, Sci. Total Environ. 720 (2020) 137601. doi:10.1016/J.SCITOTENV.2020.137601.

[45] K. V. Serebrennikova, A.N. Berlina, D. V. Sotnikov, B.B. Dzantiev, A. V. Zherdev, Raman Scattering-Based Biosensing: New Prospects and Opportunities, Biosensors. 11 (2021) 512. doi:10.3390/bios11120512.

[46] X.X. Han, H.Y. Jia, Y.F. Wang, Z.C. Lu, C.X. Wang, W.Q. Xu, B. Zhao, Y. Ozaki, Analytical technique for label-free multi-protein detection based on Western blot and surface-enhanced Raman scattering, Anal. Chem. 80 (2008) 2799–2804. doi:10.1021/ac702390U.

[47] L.J. Xu, C. Zong, X.S. Zheng, P. Hu, J.M. Feng, B. Ren, Label-free detection of native proteins by surface-enhanced Raman spectroscopy using iodide-modified nanoparticles, Anal. Chem. 86 (2014) 2238–2245. doi:10.1021/ac403974n.

[48] J.D. Driskell, A.G. Seto, L.P. Jones, S. Jokela, R.A. Dluhy, Y.P. Zhao, R.A. Tripp, Rapid microRNA (miRNA) detection and classification via surface-enhanced Raman spectroscopy (SERS), Biosens. Bioelectron. 24 (2008) 917–922. doi:10.1016/J.BIOS.2008.07.060.

[49] E. Garcia-Rico, R.A. Alvarez-Puebla, L. Guerrini, Direct surface-enhanced Raman scattering (SERS) spectroscopy of nucleic acids: from fundamental studies to real-life applications, Chem. Soc. Rev. 47 (2018) 4909–4923. doi:10.1039/C7CS00809K.

[50] A. Torres-Nuñez, K. Faulds, D. Graham, R.A. Alvarez-Puebla, L. Guerrini, Silver colloids as plasmonic substrates for direct label-free surface-enhanced Raman scattering analysis of DNA, Analyst. 141 (2016) 5170–5180. doi:10.1039/C6AN00911E.

[51] D.A. Stuart, J.M. Yuen, N. Shah, O. Lyandres, C.R. Yonzon, M.R. Glucksberg, J.T. Walsh, R.P. Van Duyne, In vivo glucose measurement by surface-enhanced Raman spectroscopy, Anal. Chem. 78 (2006) 7211–7215. doi:10.1021/ac061238u.

[52] Kenry, F. Nicolson, L. Clark, S.R. Panikkanvalappil, B. Andreiuk, C. Andreou, Advances in Surface Enhanced Raman Spectroscopy for *in Vivo* Imaging in Oncology, Nanotheranostics. 6 (2022) 31–49. doi:10.7150/NTNO.62970.

[53] C. Andreou, K. Plakas, N. Berisha, M. Gigoux, S. Pal, T. Merghoub, M. Detty, M. Kircher, Detection and imaging of multiple immunotherapy markers in the tumor microenvironment with SERS nanoprobes for automated treatment response monitoring, Front. Biol. Detect. From Nanosensors to Syst. XVI. 12861 (2024)

1286102. doi:10.1117/12.3000956.

[54] M. Li, J. Wu, M. Ma, Z. Feng, Z. Mi, P. Rong, D. Liu, Alkyne- and Nitrile-Anchored Gold Nanoparticles for Multiplex SERS Imaging of Biomarkers in Cancer Cells and Tissues, Nanotheranostics. 3 (2019) 113. doi:10.7150/NTNO.30924.

[55] J. Wu, X. Zhou, P. Li, X. Lin, J. Wang, Z. Hu, P. Zhang, D. Chen, H. Cai, R. Niessner, C. Haisch, P. Sun, Y. Zheng, Z. Jiang, H. Zhou, Ultrasensitive and Simultaneous SERS Detection of Multiplex MicroRNA Using Fractal Gold Nanotags for Early Diagnosis and Prognosis of Hepatocellular Carcinoma, Anal. Chem. 93 (2021) 8799–8809. doi:10.1021/acs.analchem.1c00478.

[56] A. Khan, S. Kang, P.Q. Bao, J.T.C. Liu, Y.W. Wang, In vivo multiplexed molecular imaging of esophageal cancer via spectral endoscopy of topically applied SERS nanoparticles, Biomed. Opt. Express. 6 (2015) 3714–3723. doi:10.1364/BOE.6.003714.

[57] M. Chen, H. Wang, Y. Zhang, H. Jiang, T. Li, L. Liu, Y. Zhao, Label-Free Multiplex Profiling of Exosomal Proteins with a Deep Learning-Driven 3D Surround-Enhancing SERS Platform for Early Cancer Diagnosis, Anal. Chem. 96 (2024) 6794–6801. doi:10.1021/acs.analchem.4c00669.

[58] K. Jibin, R. Babu V, R.S. Jayasree, Graphene-Gold Nanohybrid-Based Surface-Enhanced Raman Scattering Platform on a Portable Easy-to-Use Centrifugal Prototype for Liquid Biopsy Detection of Circulating Breast Cancer Cells, ACS Sustain. Chem. Eng. 9 (2021) 15496–15505. doi:10.1021/acssuschemeng.1c05084.

[59] H.Y. Cho, M.K. Hossain, J.H. Lee, J. Han, H.J. Lee, K.J. Kim, J.H. Kim, K.B. Lee, J.W. Choi, Selective isolation and noninvasive analysis of circulating cancer stem cells through Raman imaging, Biosens. Bioelectron. 102 (2018) 372–382. doi:10.1016/J.BIOS.2017.11.049.

[60] Y. Liu, S.H. Wu, X.Y. Du, J.J. Sun, Plasmonic Ag nanocube enhanced SERS biosensor for sensitive detection of oral cancer DNA based on nicking endonuclease signal amplification and heated electrode, Sensors Actuators B Chem. 338 (2021) 129854. doi:10.1016/J.SNB.2021.129854.

[61] Y. Si, L. Xu, N. Wang, J. Zheng, R. Yang, J. Li, Target MicroRNA-Responsive DNA Hydrogel-Based Surface-Enhanced Raman Scattering Sensor Arrays for MicroRNA-Marked Cancer Screening, Anal. Chem. 92 (2020) 2649–2655. doi:10.1021/acs.analchem.9b04606.

[62] J. Li, J. Wang, Y.S. Grewal, C.B. Howard, L.J. Raftery, S. Mahler, Y. Wang, M. Trau, Multiplexed SERS Detection of Soluble Cancer Protein Biomarkers with Gold-Silver Alloy Nanoboxes and Nanoyeast Single-Chain Variable Fragments, Anal. Chem. 90 (2018) 10377–10384. doi:10.1021/acs.analchem.8b02216.

[63] M. Su, S. Yang, M. Xu, S. Du, L. Zheng, X. Wang, C. Qu, H. Liu, Intrinsic SERS Fingerprints of Aptamer-Peptide Conjugates for Direct High-Specific Profiling Abnormal Protein Levels in Cancer Patients, Anal. Chem. 95 (2023) 12398–12405. doi:10.1021/acs.analchem.3c01988.

[64] W. Zhang, L. Jiang, R.J. Diefenbach, D.H. Campbell, B.J. Walsh, N.H. Packer, Y.

Wang, Enabling Sensitive Phenotypic Profiling of Cancer-Derived Small Extracellular Vesicles Using Surface-Enhanced Raman Spectroscopy Nanotags, ACS Sensors. 5 (2020) 764–771. doi:10.1021/acssensors.9b02377.

[65] U. Parlatan, M.O. Ozen, I. Kecoglu, B. Koyuncu, H. Torun, D. Khalafkhany, I. Loc, M.G. Ogut, F. Inci, D. Akin, I. Solaroglu, N. Ozoren, M.B. Unlu, U. Demirci, Label-Free Identification of Exosomes using Raman Spectroscopy and Machine Learning, Small. 19 (2023) 2205519. doi:10.1002/SMLL.202205519.

[66] Y. Huang, Z. Liu, X. Qin, J. Liu, Y. Yang, W. Wei, Ultrasensitive detection of gastric cancer biomarkers via a frequency shift-based SERS microfluidic chip, Analyst. 148 (2023) 3295–3305. doi:10.1039/D3AN00535F.

[67] L. Vázquez-Iglesias, G.M. Stanfoca Casagrande, D. García-Lojo, L. Ferro Leal, T.A. Ngo, J. Pérez-Juste, R.M. Reis, K. Kant, I. Pastoriza-Santos, SERS sensing for cancer biomarker: Approaches and directions, Bioact. Mater. 34 (2024) 248–268. doi:10.1016/J.BIOACTMAT.2023.12.018.

[68] X. Li, T. Yang, J. Lin, Spectral analysis of human saliva for detection of lung cancer using surface-enhanced Raman spectroscopy, J. Biomed. Opt. 17 (2012) 037003. doi:10.1117/1.JBO.17.3.037003.

[69] R.M. Borșa, V. Toma, A. Onaciu, C.S. Moldovan, R. Mărginean, D. Cenariu, G.F. Știufiuc, C.M. Dinu, S. Bran, H.O. Opriș, S. Văcăraș, F. Onișor-Gligor, D. Sentea, M.F. Băciuț, C.A. Iuga, R.I. Știufiuc, Developing New Diagnostic Tools Based on SERS Analysis of Filtered Salivary Samples for Oral Cancer Detection, Int. J. Mol. Sci. 24 (2023) 12125. doi:10.3390/IJMS241512125/S1.

[70] A. Fălămaș, H. Rotaru, M. Hedeșiu, Surface-enhanced Raman spectroscopy (SERS) investigations of saliva for oral cancer diagnosis, Lasers Med. Sci. 35 (2020) 1393–1401. doi:10.1007/S10103-020-02988-2.

[71] A. Hernández-Arteaga, J. de Jesús Zermeño Nava, E.S. Kolosovas-Machuca, J.J. Velázquez-Salazar, E. Vinogradova, M. José-Yacamán, H.R. Navarro-Contreras, Diagnosis of breast cancer by analysis of sialic acid concentrations in human saliva by surface-enhanced Raman spectroscopy of silver nanoparticles, Nano Res. 10 (2017) 3662–3670. doi:10.1007/S12274-017-1576-5.

[72] S. Huang, L. Wang, W. Chen, S. Feng, J. Lin, Z. Huang, G. Chen, B. Li, R. Chen, Potential of non-invasive esophagus cancer detection based on urine surface-enhanced Raman spectroscopy, Laser Phys. Lett. 11 (2014) 115604. doi:10.1088/1612-2011/11/11/115604.

[73] G. Saranya, M.M. Joseph, V. Karunakaran, J.B. Nair, V.N. Saritha, V.S. Veena, K. Sujathan, A. Ajayaghosh, K.K. Maiti, Enzyme-Driven Switchable Fluorescence-SERS Diagnostic Nanococktail for the Multiplex Detection of Lung Cancer Biomarkers, ACS Appl. Mater. Interfaces. 10 (2018) 38807–38818. doi:10.1021/acsami.8b15583.

[74] A. Jaworska, S. Fornasaro, V. Sergo, A. Bonifacio, Potential of Surface Enhanced Raman Spectroscopy (SERS) in Therapeutic Drug Monitoring (TDM). A Critical Review, Biosensors. 6 (2016) 47. doi:10.3390/BIOS6030047.

[75] R. Chen, Q. Chen, Y. Wang, Z. Feng, Z.W. Xu, P. Zhou, W. Huang, H. Cheng, L. Li, J. Feng, Ultrasensitive SERS substrate for label-free therapeutic drug monitoring of chlorpromazine hydrochloride and aminophylline in human serum, Anal. Bioanal. Chem. 415 (2023) 1803–1815. doi:10.1007/s00216-023-04621-x.

[76] G. Soufi, I. Badillo-Ramírez, L. Serioli, R. Altaf Raja, K. Schmiegelow, K. Zor, A. Boisen, Solid-phase extraction coupled to automated centrifugal microfluidics SERS: Improving quantification of therapeutic drugs in human serum, Biosens. Bioelectron. 266 (2024) 116725. doi:10.1016/J.BIOS.2024.116725.

[77] L. Liu, W. Ma, X. Wang, S. Li, Recent Progress of Surface-Enhanced Raman Spectroscopy for Bacteria Detection, Biosensors. 13 (2023) 350. doi:10.3390/BIOS13030350.

[78] X. Zhou, Z. Hu, D. Yang, S. Xie, Z. Jiang, R. Niessner, C. Haisch, H. Zhou, P. Sun, Bacteria Detection: From Powerful SERS to Its Advanced Compatible Techniques, Adv. Sci. 7 (2020) 2001739. doi:10.1002/ADVS.202001739.

[79] P.A. Mosier-Boss, K.C. Sorensen, R.D. George, P.C. Sims, A. Obraztsova, Surface enhanced Raman scattering of bacteria using capped and uncapped silver nanoparticles, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 242 (2020) 118742. doi:10.1016/J.SAA.2020.118742.

[80] W.Q. Wang, V. Hynninen, L. Qiu, A.W. Zhang, T. Lemma, N.N. Zhang, H.H. Ge, J.J. Toppari, V.P. Hytönen, J. Wang, Synergistic enhancement via plasmonic nanoplate-bacteria-nanorod supercrystals for highly efficient SERS sensing of food-borne bacteria, Sensors Actuators B Chem. 239 (2017) 515–525. doi:10.1016/J.SNB.2016.08.040.

[81] X. Zhou, Z. Hu, D. Yang, S. Xie, Z. Jiang, R. Niessner, C. Haisch, H. Zhou, P. Sun, Bacteria Detection: From Powerful SERS to Its Advanced Compatible Techniques, Adv. Sci. 7 (2020) 2001739. doi:10.1002/ADVS.202001739.

[82] B. Ankamwar, U.K. Sur, P. Das, SERS study of bacteria using biosynthesized silver nanoparticles as the SERS substrate, Anal. Methods. 8 (2016) 2335–2340. doi:10.1039/C5AY03014E.

[83] M.M.P. Arnob, W.C. Shih, 3-Dimensional Plasmonic Substrates Based on Chicken Eggshell Bio-Templates for SERS-Based Bio-Sensing, Micromachines. 8 (2017) 196. doi:10.3390/MI8060196.

[84] E. Witkowska, K. Niciński, D. Korsak, T. Szymborski, A. Kamińska, Sources of variability in SERS spectra of bacteria: comprehensive analysis of interactions between selected bacteria and plasmonic nanostructures, Anal. Bioanal. Chem. 411 (2019) 2001–2017. doi:10.1007/S00216-019-01609-4.

[85] E. Akanny, A. Bonhommé, F. Bessueille, S. Bourgeois, C. Bordes, Surface enhanced Raman spectroscopy for bacteria analysis: a review, Appl. Spectrosc. Rev. 56 (2021) 380–422. doi:10.1080/05704928.2020.1796698.

[86] P.A. Mosier-Boss, Review on SERS of bacteria, Biosensors. 7 (2017). doi:10.3390/bios7040051.

[87] W. Liu, L. Wei, D. Wang, C. Zhu, Y. Huang, Z. Gong, C. Tang, M. Fan, Phenotyping Bacteria through a Black-Box Approach: Amplifying Surface-Enhanced Raman Spectroscopy Spectral Differences among Bacteria by Inputting Appropriate Environmental Stress, Anal. Chem. 94 (2022) 6791–6798. doi:10.1021/acs.analchem.2c00502.

[88] W. Zhang, S. He, W. Hong, P. Wang, A Review of Raman-Based Technologies for Bacterial Identification and Antimicrobial Susceptibility Testing, Photonics. 9 (2022) 133. doi:10.3390/photonics9030133.

[89] H. Chen, A. Das, L. Bi, N. Choi, J. Il Moon, Y. Wu, S. Park, J. Choo, Recent advances in surface-enhanced Raman scattering-based microdevices for point-of-care diagnosis of viruses and bacteria, Nanoscale. 12 (2020) 21560–21570. doi:10.1039/D0NR06340A.

[90] S.H. Lee, J. Hwang, K. Kim, J. Jeon, S. Lee, J. Ko, J. Lee, M. Kang, D.R. Chung, J. Choo, Quantitative Serodiagnosis of Scrub Typhus Using Surface-Enhanced Raman Scattering-Based Lateral Flow Assay Platforms, Anal. Chem. 91 (2019) 12275–12282. doi:10.1021/acs.analchem.9b02363.

[91] Y.T. Yeh, K. Gulino, Y.H. Zhang, A. Sabestien, T.W. Chou, B. Zhou, Z. Lin, I. Albert, H. Lu, V. Swaminathan, E. Ghedin, M. Terrones, A rapid and label-free platform for virus capture and identification from clinical samples, Proc. Natl. Acad. Sci. 117 (2020) 895–901. doi:10.1073/pnas.1910113117.

[92] Y. Xu, F. Gu, S. Hu, Y. Wu, C. Wu, Y. Deng, B. Gu, Z. Chen, Y. Yang, A cell wall-targeted organic-inorganic hybrid nano-catcher for ultrafast capture and SERS detection of invasive fungi, Biosens. Bioelectron. 228 (2023) 115173. doi:10.1016/J.BIOS.2023.115173.

[93] Z. Guo, M. Wang, A.O. Barimah, Q. Chen, H. Li, J. Shi, H.R. El-Seedi, X. Zou, Label-free surface enhanced Raman scattering spectroscopy for discrimination and detection of dominant apple spoilage fungus, Int. J. Food Microbiol. 338 (2021) 108990. doi:10.1016/J.IJFOODMICRO.2020.108990.

[94] E. Witkowska, T. Jagielski, A. Kamińska, A. Kowalska, A. Hryncewicz-Gwóźdź, J. Waluk, Detection and identification of human fungal pathogens using surface-enhanced Raman spectroscopy and principal component analysis, Anal. Methods. 8 (2016) 8427–8434. doi:10.1039/C6AY02957D.

[95] T.W. Randolph, Scale-based normalization of spectral data, Cancer Biomarkers. 2 (2006) 135–144. doi:10.3233/CBM-2006-23-405.

[96] R. Gautam, S. Vanga, F. Ariese, S. Umapathy, Review of multidimensional data processing approaches for Raman and infrared spectroscopy, EPJ Tech. Instrum. 2 (2015). doi:10.1140/epjti/s40485-015-0018-6.

[97] A. González-Casado, A. María, J. Carvelo, R. González-Domínguez, A. Sayago, Á. Fernández-Recamales, An Overview on the Application of Chemometrics Tools in Food Authenticity and Traceability, Foods. 11 (2022) 3940. doi:10.3390/FOODS11233940.

[98] L.A. Berrueta, R.M. Alonso-Salces, K. Héberger, Supervised pattern recognition in

food analysis, J. Chromatogr. A. 1158 (2007) 196–214. doi:10.1016/j.chroma.2007.05.024.

[99] S.M. Sarsam, Reinforcing the decision-making process in chemometrics: Feature selection and algorithm optimization, ACM Int. Conf. Proceeding Ser. Part F147956 (2019) 11–16. doi:10.1145/3316615.3316644.

[100] A. Amirvaresi, H. Parastar, Miniaturized NIR spectroscopy and chemometrics: A smart combination to solve food authentication challenges, Front. Anal. Sci. 3 (2023) 1118590. doi:10.3389/FRANS.2023.1118590.

[101] O'Donnell Colm P., Fagan Colette, Cullen P.J, Process Analytical Technology for the Food Industry, Springer, 2014. doi:10.1007/978-1-4939-0311-5.

[102] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130. doi:10.1016/S0169-7439(01)00155-1.

[103] R.G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pomerantsev, O. Rodionova, J.M. Roger, B. Walczak, R. Tauler, Chemometrics in analytical chemistry—part II: modeling, validation, and applications, Anal. Bioanal. Chem. 410 (2018) 6691–6704. doi:10.1007/S00216-018-1283-4.

[104] A. Rácz, D. Bajusz, K. Héberger, Chemometrics in Analytical Chemistry, in: Appl. Chemoinformatics Achiev. Futur. Oppor., 2018: pp. 471–499. doi:10.1002/9783527806539.ch9.

[105] A. Biancolillo, F. Marini, Chemometric methods for spectroscopy-based pharmaceutical analysis, Front. Chem. 6 (2018) 412780. doi:10.3389/fchem.2018.00576.

[106] Z. Małyjurek, R. Vitale, B. Walczak, Different strategies for class model optimization. A comparative study, Talanta. 215 (2020) 120912. doi:10.1016/J.TALANTA.2020.120912.

[107] W.F. De Carvalho Rocha, C.B. Do Prado, N. Blonder, Comparison of chemometric problems in food analysis using non-linear methods, Molecules. 25 (2020) 3025. doi:10.3390/molecules25133025.

[108] E. Bona, P.H. Março, P. Valderrama, Chemometrics Applied to Food Control, in: Food Control Biosecurity, Academic Press, 2018: pp. 105–133. doi:10.1016/B978-0-12-811445-2.00004-0.

[109] M.D. Peris-Díaz, A. Krężel, A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry, TrAC Trends Anal. Chem. 135 (2021) 116157. doi:10.1016/J.TRAC.2020.116157.

[110] Esbensen Kim. H, B. Swarbrick, Multivariate Data Analysis, 6th Editio, CAMO Software AS, 2018.

[111] M. De Luca, F. Oliverio, D. Ioele, G.P. Husson, G. Ragno, Monitoring of water quality in South Paris district by clustering and SIMCA classification, Int. J. Environ. Anal. Chem. 88 (2008) 1087–1105. doi:10.1080/03067310802428232.

[112] D. Suhandy, M. Yulia, Discrimination of several Indonesian specialty coffees using Fluorescence Spectroscopy combined with SIMCA method, IOP Conf. Ser. Mater. Sci. Eng. 334 (2018) 012059. doi:10.1088/1757-899X/334/1/012059.

[113] H. Bächle, B. Zimmer, G. Wegener, Classification of thermally modified wood by FT-NIR spectroscopy and SIMCA, Wood Sci. Technol. 46 (2012) 1181–1192. doi:10.1007/s00226-012-0481-z.

[114] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods. 6 (2014) 2812–2831. doi:10.1039/c3ay41907j.

[115] G. Mooiweer, Multivariate Analysis of Spectroscopic Data, (2019) 1–161.

[116] H. Abdi, L.J. Williams, Principal component analysis, Wiley Interdiscip. Rev. Comput. Stat. 2 (2010) 433–459. doi:10.1002/wics.101.

[117] F. Kherif, A. Latypova, Principal component analysis, in: Mach. Learn. Methods Appl. to Brain Disord., Academic Press, 2020: pp. 209–225. doi:10.1016/B978-0-12-815739-8.00012-2.

[118] H. Abdi, L.J. Williams, Partial Least Squares Methods: Partial Least Squares Correlation and Partial Least Square Regression, Methods Mol. Biol. 930 (2013) 549–579. doi:10.1007/978-1-62703-059-5_23.

[119] C.P. Sharma, S. Sharma, V. Sharma, R. Singh, Rapid and non-destructive identification of claws using ATR-FTIR spectroscopy–A novel approach in wildlife forensics, Sci. Justice. 59 (2019) 622–629. doi:10.1016/j.scijus.2019.08.002.

[120] M. Bevilacqua, F. Marini, Local classification: Locally weighted–partial least squares-discriminant analysis (LW–PLS-DA), Anal. Chim. Acta. 838 (2014) 20–30. doi:10.1016/j.aca.2014.05.057.

[121] A. Höskuldsson, PLS regression methods, J. Chemom. 2 (1988) 211–228. doi:10.1002/cem.1180020306.

[122] L.C. Lee, C.Y. Liong, A.A. Jemain, Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps, Analyst. 143 (2018) 3526–3539. doi:10.1039/C8AN00599K.

[123] K.Y. Peerbhay, O. Mutanga, R. Ismail, Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa, ISPRS J. Photogramm. Remote Sens. 79 (2013) 19–28. doi:10.1016/j.isprsjprs.2013.01.013.

[124] A. Kalivodová, K. Hron, P. Filzmoser, L. Najdekr, H. Janečková, T. Adam, PLS-DA for compositional data with application to metabolomics, J. Chemom. 29 (2015) 21–28. doi:10.1002/cem.2657.

[125] A.M.C. Davies, T. Fearn, Back to basics: multivariate qualitative analysis, SIMCA, Spectrosc. Eur. 20 (2008) 15–19. doi:10.1255/SEW.2008.A1.

[126] N. Kumar, A. Bansal, G.S. Sarma, R.K. Rawal, Chemometrics tools used in analytical chemistry: An overview, Talanta. 123 (2014) 186–199.

doi:10.1016/J.TALANTA.2014.02.003.

[127] Y. Yao, Y. Liu, Y. Yu, H. Xu, W. Lv, Z. Li, X. Chen, K-SVM: An Effective SVM Algorithm Based on K-means Clustering, J. Comput. 8 (2013) 2632–2639. doi:10.4304/jcp.8.10.2632-2639.

[128] V. Mohan, Liver Disease Prediction using SVM and Naïve Bayes Algorithms Privacy Preserving Data Mining View project, Int. J. Sci. Eng. Technol. Res. 4 (2015) 816–820.

[129] G. Battineni, N. Chintalapudi, F. Amenta, Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM), Informatics Med. Unlocked. 16 (2019) 100200. doi:10.1016/J.IMU.2019.100200.

[130] Gonorrhoea (Neisseria gonorrhoeae infection), World Heal. Organ. (n.d.). https://www.who.int/news-room/fact-sheets/detail/gonorrhoea-(neisseria-gonorrhoeae-infection) (accessed June 7, 2024).

[131] M. Unemo, H.S. Seifert, E.W. Hook, S. Hawkes, F. Ndowa, J.A.R. Dillon, Gonorrhoea, Nat. Rev. Dis. Prim. 5 (2019). doi:10.1038/s41572-019-0128-6.

[132] M. Skerlev, I. Čulav-Košćak, Gonorrhea: New challenges, Clin. Dermatol. 32 (2014) 275–281. doi:10.1016/j.clindermatol.2013.08.010.

[133] E. van Dyck, A. Meheus, P. Piot, W.H. Organization, Laboratory Diagnosis of Sexually Transmitted Diseases, World Health Organization, 1999.

[134] L.K. Ng, I.E. Martin, The laboratory diagnosis of Neisseria gonorrhoeae, Can. J. Infect. Dis. Med. Microbiol. 16 (2005) 15–25. doi:10.1155/2005/323082.

[135] H. Atef Osman, H. Abd El Monem Mohamed, A Descriptive Comparative Study of Vaginal Infection Control Practices between Rural and Urban Women, Orig. Artic. Egypt. J. Heal. Care. 14 (2023) 888–904.

[136] S.M. Shawaky, M. Majed, A. Al Shammari, M. Shafik Sewelliam, A.A. El, R. Ghazal, A.N. Amer, A study on vaginitis among pregnant and non-pregnant females in Alexandria, Egypt: An unexpected high rate of mixed vaginal infection, AIMS Microbiol. 8 (2022) 167–177. doi:10.3934/microbiol.2022014.

[137] J.D. Sobel, C. Subramanian, B. Foxman, M. Fairfax, S.E. Gygax, Mixed vaginitis - More than coinfection and with therapeutic implications, Curr. Infect. Dis. Rep. 15 (2013) 104–108. doi:10.1007/s11908-013-0325-5.

[138] R. Palmeira-de-Oliveira, A. Palmeira-de-Oliveira, J. Martinez-de-Oliveira, New strategies for local treatment of vaginal infections, Adv. Drug Deliv. Rev. 92 (2015) 105–122. doi:10.1016/j.addr.2015.06.008.

[139] F. Blostein, E. Levin-Sparenberg, J. Wagner, B. Foxman, Recurrent vulvovaginal candidiasis, Ann. Epidemiol. 27 (2017) 575-582.e3. doi:10.1016/J.ANNEPIDEM.2017.08.010.

[140] J.D. Sobel, S. Faro, R.W. Force, B. Foxman, W.J. Ledger, P.R. Nyirjesy, B.D. Reed, P.R. Summers, Vulvovaginal candidiasis: Epidemiologic, diagnostic, and therapeutic considerations, Am. J. Obstet. Gynecol. 178 (1998) 203–211. doi:10.1016/S0002-

9378(98)80001-X.

[141] Vulvovaginal Candidiasis - STI Treatment Guidelines, (n.d.). https://www.cdc.gov/std/treatment-guidelines/candidiasis.htm (accessed March 10, 2024).

[142] J. Mashburn, Vaginal Infections Update, J. Midwifery Women's Heal. 57 (2012) 629–634. doi:10.1111/j.1542-2011.2012.00246.x.

[143] C.A. Spiegelt, Bacterial vaginosis, Clin. Microbiol. Rev. 4 (1991) 485–502. doi:10.1128/CMR.4.4.485.

[144] D.N. Fredricks, T.L. Fiedler, J.M. Marrazzo, Molecular Identification of Bacteria Associated with Bacterial Vaginosis, N. Engl. J. Med. 353 (2005) 1899–911. doi:10.1056/NEJMoa043802.

[145] R. Amsel, P.A. Totten, C.A. Spiegel, K.C.S. Chen, D. Eschenbach, K.K. Holmes, Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations, Am. J. Med. 74 (1983) 14–22. doi:10.1016/0002-9343(83)91112-9.

[146] D. Money, The laboratory diagnosis of bacterial vaginosis, Can. J. Infect. Dis. Med. Microbiol. 16 (2005) 77. doi:10.1155/2005/230319.

[147] R.E. Gutman, J.F. Peipert, S. Weitzen, J. Blume, Evaluation of clinical methods for diagnosing bacterial vaginosis, Obstet. Gynecol. 105 (2005) 551–556. doi:10.1097/01.AOG.0000145752.97999.67.

[148] V. Karunakaran, V.N. Saritha, M.M. Joseph, J.B. Nair, G. Saranya, K.G. Raghu, K. Sujathan, K.S. Kumar, K.K. Maiti, Diagnostic spectro-cytology revealing differential recognition of cervical cancer lesions by label-free surface enhanced Raman fingerprints and chemometrics, Nanomedicine Nanotechnology, Biol. Med. 29 (2020) 102276. doi:10.1016/j.nano.2020.102276.

[149] S. Choi, H.K. Park, G.E. Min, Y.H. Kim, Biochemical investigations of human papillomavirus-infected cervical fluids, Microsc. Res. Tech. 78 (2015) 200–206. doi:10.1002/jemt.22461.

[150] A. Ceja-Fdez, R. Carriles, A.L. González-Yebra, J. Vivero-Escoto, E. De La Rosa, T. López-Luke, Imaging and SERS Study of the Au Nanoparticles Interaction with HPV and Carcinogenic Cervical Tissues, Molecules. 26 (2021) 3758. doi:10.3390/molecules26123758.

[151] A. Sikirzhytskaya, V. Sikirzhytski, I.K. Lednev, Raman spectroscopy coupled with advanced statistics for differentiating menstrual and peripheral blood, J. Biophotonics. 7 (2014) 59–67. doi:10.1002/jbio.201200191.

[152] T. Reese, C. Suarez, W.R. Premasiri, M.L. Shaine, H. Ingraham, A.N. Brodeur, L.D. Ziegler, Surface enhanced Raman scattering specificity for detection and identification of dried bloodstains, Forensic Sci. Int. 328 (2021) 111000. doi:10.1016/j.forsciint.2021.111000.

[153] B. Vyas, L. Halámková, I.K. Lednev, A universal test for the forensic identification of all main body fluids including urine, Forensic Chem. 20 (2020) 100247. doi:10.1016/j.forc.2020.100247.

[154] A. Sikirzhytskaya, V. Sikirzhytski, I.K. Lednev, Raman spectroscopic signature of vaginal fluid and its potential application in forensic body fluid identification, Forensic Sci. Int. 216 (2012) 44–48. doi:10.1016/J.FORSCIINT.2011.08.015.

[155] C.K. Muro, K.C. Doty, L. de Souza Fernandes, I.K. Lednev, Forensic body fluid identification and differentiation by Raman spectroscopy, Forensic Chem. 1 (2016) 31–38. doi:10.1016/j.forc.2016.06.003.

[156] R. Maia, V. Carvalho, B. Faria, I. Miranda, S. Catarino, S. Teixeira, R. Lima, G. Minas, J. Ribeiro, Diagnosis Methods for COVID-19: A Systematic Review, Micromachines. 13 (2022) 1349. doi:10.3390/MI13081349.

[157] J. Kashir, A. Yaqinuddin, Loop mediated isothermal amplification (LAMP) assays as a rapid diagnostic for COVID-19, Med. Hypotheses. 141 (2020) 109786. doi:10.1016/J.MEHY.2020.109786.

[158] M. Yüce, E. Filiztekin, K.G. Özkaya, COVID-19 diagnosis —A review of current methods, Biosens. Bioelectron. 172 (2021) 112752. doi:10.1016/J.BIOS.2020.112752.

[159] R.W. Peeling, C.J. Wedderburn, P.J. Garcia, D. Boeras, N. Fongwen, J. Nkengasong, A. Sall, A. Tanuri, D.L. Heymann, Serology testing in the COVID-19 pandemic response, Lancet Infect. Dis. 20 (2020) e245–e249. doi:10.1016/S1473-3099(20)30517-X.

[160] V. Karunakaran, M.M. Joseph, I. Yadev, H. Sharma, K. Shamna, S. Saurav, R.P. Sreejith, V. Anand, R. Beegum, S. Regi David, T. Iype, K.L. Sarada Devi, A. Nizarudheen, M.S. Sharmad, R. Sharma, R. Mukhiya, E. Thouti, K. Yoosaf, J. Joseph, P. Sujatha Devi, S. Savithri, A. Agarwal, S. Singh, K.K. Maiti, A non-invasive ultrasensitive diagnostic approach for COVID-19 infection using salivary label-free SERS fingerprinting and artificial intelligence, J. Photochem. Photobiol. B Biol. 234 (2022) 112545. doi:10.1016/J.JPHOTOBIOL.2022.112545.

[161] Y. Yang, H. Li, L. Jones, J. Murray, H. Naikare, Y.Y.C. Mosley, T. Spikes, S. Hülck, R.A. Tripp, B. Ai, Y. Zhao, Advancing SERS Diagnostics in COVID-19 with Rapid, Accurate, and Label-Free Viral Load Monitoring in Clinical Specimens via SFNet Enhancement, Adv. Mater. Interfaces. 11 (2024) 2400013. doi:10.1002/admi.202400013.

[162] Y. Li, C. Lin, Y. Peng, J. He, Y. Yang, High-sensitivity and point-of-care detection of SARS-CoV-2 from nasal and throat swabs by magnetic SERS biosensor, Sensors Actuators B Chem. 365 (2022) 131974. doi:10.1016/J.SNB.2022.131974.

[163] Y. Yang, H. Li, L. Jones, J. Murray, J. Haverstick, H.K. Naikare, Y.Y.C. Mosley, R.A. Tripp, B. Ai, Y. Zhao, Rapid Detection of SARS-CoV-2 RNA in Human Nasopharyngeal Specimens Using Surface-Enhanced Raman Spectroscopy and Deep Learning Algorithms, ACS Sensors. 8 (2023) 297–307. doi:10.1021/acssensors.2c02194.

[164] C.S.H. Hwang, S. Lee, S. Lee, H. Kim, T. Kang, D. Lee, K.H. Jeong, Highly Adsorptive Au-TiO2Nanocomposites for the SERS Face Mask Allow the Machine-Learning-Based Quantitative Assay of SARS-CoV-2 in Artificial Breath Aerosols, ACS Appl. Mater. Interfaces. 14 (2022) 54550–54557. doi:10.1021/acsami.2c16446.

[165] T. Szymborski, Y. Stepanenko, K. Niciński, P. Piecyk, S.M. Berus, M. Adamczyk-Popławska, A. Kamińska, Ultrasensitive SERS platform made via femtosecond laser micromachining for biomedical applications, J. Mater. Res. Technol. 12 (2021) 1496–1507. doi:10.1016/j.jmrt.2021.03.083.

[166] T. Meyer, S. Buder, The Laboratory Diagnosis of Neisseria gonorrhoeae: Current Testing and Future Demands, Pathogens. 9 (2020) 91. doi:10.3390/pathogens9020091.

[167] S.M. Berus, M. Adamczyk-Popławska, K. Goździk, G. Przedpełska, T.R. Szymborski, Y. Stepanenko, A. Kamińska, SERS-PLSR Analysis of Vaginal Microflora: Towards the Spectral Library of Microorganisms, Int. J. Mol. Sci. 23 (2022) 12576. doi:10.3390/ijms232012576.

[168] W.A.T. Brunton, H. Young, D.R.K. Fraser, Isolation of Neisseria lactamica from the female genital tract. A case report, Sex. Transm. Infect. 56 (1980) 325–326. doi:10.1136/sti.56.5.325.

[169] Y.C. Faur, M.H. Weisburd, M.E. Wilson, Isolation of Neisseria meningitidis from the Genito-Urinary Tract and Anal Canal, J. Clin. Microbiol. 2 (1975) 178–182. doi:10.1128/jcm.2.3.178-182.1975.

[170] A.P. Johnsoon, The pathogenic potential of commensal species of Neisseria Gram-negative cocci were allocated a variety of generic and specific names including Micrococcus, J. Clin. Pathol. 36 (1983) 213–223. doi:10.1136/jcp.36.2.213.

[171] A.E. Wilkinson, Occurrence of neisseria other than the gonococcus in the genital tract, Br. J. Vener. Dis. 28 (1952) 24–27. doi:10.1136/sti.28.1.24.

[172] W.R. Premasiri, J.C. Lee, A. Sauer-Budge, R. Théberge, C.E. Costello, L.D. Ziegler, The biochemical origins of the surface-enhanced Raman spectra of bacteria: a metabolomics profiling by SERS, Anal. Bioanal. Chem. 408 (2016) 4631–4647. doi:10.1007/s00216-016-9540-x.

[173] W.R. Premasiri, D.T. Moir, M.S. Klempner, N. Krieger, G. Jones, L.D. Ziegler, Characterization of the Surface Enhanced Raman Scattering (SERS) of bacteria, J. Phys. Chem. B. 109 (2005) 312–320. doi:10.1021/jp040442n.

[174] W.R. Premasiri, D.T. Moir, L.D. Ziegler, Vibrational fingerprinting of bacterial pathogens by surface enhanced Raman scattering (SERS), Chem. Biol. Sens. VI. 5795 (2005) 19–29. doi:10.1117/12.605742.

[175] M. Kahraman, M.M. Yazici, F. Şahin, Ö.F. Bayrak, M. Çulha, Reproducible Surface-Enhanced Raman Scattering Spectra of Bacteria on Aggregated Silver Nanoparticles, Appl. Spectrosc. 61 (2007) 479–485.

[176] P. Kubryk, R. Niessner, N.P. Ivleva, The origin of the band at around 730 cm-1 in the SERS spectra of bacteria: A stable isotope approach, Analyst. 141 (2016) 2874–2878. doi:10.1039/c6an00306k.

[177] T. Lemma, A. Saliniemi, V. Hynninen, V.P. Hytönen, J.J. Toppari, SERS detection of cell surface and intracellular components of microorganisms using nano-aggregated Ag substrate, Vib. Spectrosc. 83 (2016) 36–45.

doi:10.1016/j.vibspec.2016.01.006.

[178] N. Paccotti, F. Boschetto, S. Horiguchi, E. Marin, A. Chiadò, C. Novara, F. Geobaldo, F. Giorgis, G. Pezzotti, Label-Free SERS Discrimination and In Situ Analysis of Life Cycle in Escherichia coli and Staphylococcus epidermidis, Biosensors. 8 (2018) 131. doi:10.3390/bios8040131.

[179] O. Prakash, S. Sil, T. Verma, S. Umapathy, Direct Detection of Bacteria Using Positively Charged Ag/Au Bimetallic Nanoparticles: A Label-free Surface-Enhanced Raman Scattering Study Coupled with Multivariate Analysis, J. Phys. Chem. C. 124 (2020) 861–869. doi:10.1021/acs.jpcc.9b09311.

[180] R.B. Moyes, J. Reynolds, D.P. Breakwell, Differential Staining of Bacteria: Gram Stain, Curr. Protoc. Microbiol. 15 (2009) A.3C.1-A.3C.8. doi:10.1002/9780471729259.mca03cs15.

[181] J.M. Spence, L. Wright, V.L. Clark, Laboratory Maintenance of *Neisseria gonorrhoeae*, Curr. Protoc. Microbiol. 8 (2008) 4A.1.1-4A.1.26. doi:10.1002/9780471729259.mc04a01s8.

[182] E.N. Ilina, A.D. Borovskaya, M.M. Malakhova, V.A. Vereshchagin, A.A. Kubanova, A.N. Kruglov, T.S. Svistunova, A.O. Gazarian, T. Maier, M. Kostrzewa, V.M. Govorun, Direct bacterial profiling by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry for identification of pathogenic neisseria, J. Mol. Diagnostics. 11 (2009) 75–86. doi:10.2353/jmoldx.2009.080079.

[183] S.C. Wunschel, K.H. Jarman, C.E. Petersen, N.B. Valentine, K.L. Wahl, D. Schauki, J. Jackman, C.P. Nelson, E. White, Bacterial Analysis by MALDI-TOF Mass Spectroscopy: An Inter-Laboratory Comparison, J. Am. Soc. Mass Spectrosc. 16 (2020) 27. doi:10.1016/j.jasms.2004.12.004.

[184] G. Kaiser, Isolation and Identification of Neisseriae, Mycobacteria, and Obligate Anaerobes - Biology LibreTexts, (2020). https://bio.libretexts.org/Bookshelves/Ancillary_Materials/Laboratory_Experiments/Microbiology_Labs/Microbiology_Labs_II/Lab_16%3A_Isolation_and_Identification_of_Neisseriae%2C_Mycobacteria%2C_and_Obligate_Anaerobes (accessed November 22, 2020).

[185] T. Gregersen, Rapid method for distinction of gram-negative from gram-positive bacteria, Eur. J. Appl. Microbiol. Biotechnol. 5 (1978) 123–127. doi:10.1007/BF00498806.

[186] K. Moaledj, Comparison of Gram-staining and alternate methods, KOH test and aminopeptidase activity in aquatic bacteria: their application to numerical taxonomy, J. Microbiol. Methods. 5 (1986) 303–310. doi:10.1016/0167-7012(86)90056-4.

[187] P. Shields, L. Cathcart, Oxidase Test Protocol, (2010) 1–9. https://www.asmscience.org/content/education/protocol/protocol.3229 (accessed November 23, 2020).

[188] K. Reiner, Catalase Test Protocol, 2010. www.asmscience.org (accessed November 23, 2020).

[189] R. Eliasson, Cholesterol in human semen., Biochem. J. 98 (1966) 242–243. doi:10.1042/bj0980242.

[190] P.J. Horner, K. Blee, L. Falk, W. Van Der Meijden, H. Moi, 2016 European guideline on the management of non-gonococcal urethritis, Int. J. STD AIDS. 27 (2016) 928–937. doi:10.1177/0956462416648585.

[191] D.H. Owen, D.F. Katz, A review of the physical and chemical properties of human semen and the formulation of a semen simulant, J. Androl. 26 (2005) 459–469. doi:10.2164/jandrol.04104.

[192] D. Priest, J.J. Ong, E.P.F. Chow, S. Tabrizi, S. Phillips, M. Bissessor, C.K. Fairley, C.S. Bradshaw, T.R.H. Read, S. Garland, M. Chen, Neisseria gonorrhoeae DNA bacterial load in men with symptomatic and asymptomatic gonococcal urethritis, Sex. Transm. Infect. 93 (2017) 478–481. doi:10.1136/sextrans-2016-052950.

[193] R. Dong, S. Weng, L. Yang, J. Liu, Detection and direct readout of drugs in human urine using dynamic surface-enhanced Raman spectroscopy and support vector machines, Anal. Chem. 87 (2015) 2937–2944. doi:10.1021/acs.analchem.5b00137.

[194] U. Neugebauer, J.H. Clement, T. Bocklitz, C. Krafft, J. Popp, Identification and differentiation of single cells from peripheral blood by Raman spectroscopic imaging, J. Biophotonics. 3 (2010) 579–587. doi:10.1002/jbio.201000020.

[195] G. Del Mistro, S. Cervo, E. Mansutti, R. Spizzo, A. Colombatti, P. Belmonte, R. Zucconelli, A. Steffan, V. Sergo, A. Bonifacio, Surface–enhanced raman spectroscopy of urine for prostate cancer detection: A preliminary study, Anal. Bioanal. Chem. 407 (2015) 3271–3275. doi:10.1007/s00216-015-8610-9.

[196] V. Sikirzhytski, K. Virkler, I.K. Lednev, Discriminant Analysis of Raman Spectra for Body Fluid Identification for Forensic Purposes, Sensors. 10 (2010) 2869–2884. doi:10.3390/s100402869.

[197] A. Bertoluzza, C. Fagnano, P. Finelli, M.A. Morelli, R. Simoni, R. Tosi, Raman and infrared spectra of spermidine and spermine and their hydrochlorides and phosphates as a basis for the study of the interactions between polyamines and nucleic acids, J. Raman Spectrosc. 14 (1983) 386–394. doi:10.1002/jrs.1250140606.

[198] Y. Zou, P. Xia, F. Yang, F. Cao, K. Ma, Z. Mi, X. Huang, N. Cai, B. Jiang, X. Zhao, W. Liu, X. Chen, Whole blood and semen identification using mid-infrared and Raman spectrum analysis for forensic applications, Anal. Methods. 8 (2016) 3763–3767. doi:10.1039/c5ay03337c.

[199] A.C.S. Talari, Z. Movasaghi, S. Rehman, I.U. Rehman, Raman spectroscopy of biological tissues, Appl. Spectrosc. Rev. 50 (2015) 46–111. doi:10.1080/05704928.2014.923902.

[200] K. Virkler, I.K. Lednev, Raman spectroscopic signature of semen and its potential application to forensic body fluid identification, Forensic Sci. Int. 193 (2009) 56–62. doi:10.1016/j.forsciint.2009.09.005.

[201] X. Chen, Z. Huang, S. Feng, J. Chen, L. Wang, P. Lu, H. Zeng, R. Chen, Analysis and differentiation of seminal plasma via polarized SERS spectroscopy, Int. J.

Nanomedicine. 7 (2012) 6115–6121. doi:10.2147/IJN.S37782.

[202] E. Olshen, L.A. Shrier, Diagnostic tests for chlamydial and gonorrheal infections, Semin. Pediatr. Infect. Dis. 16 (2005) 192–198. doi:10.1053/j.spid.2005.04.008.

[203] D.M. Whiley, J.W. Tapsall, T.P. Sloots, Nucleic acid amplification testing for Neisseria gonorrhoeae: An ongoing challenge, J. Mol. Diagnostics. 8 (2006) 3–15. doi:10.2353/jmoldx.2006.050045.

[204] G.J. Locksmith, New diagnostic tests for gonorrhea and chlamydia, Prim. Care Update Ob. Gyns. 4 (1997) 161–167. doi:10.1016/S1068-607X(97)00044-9.

[205] C.A. Gaydos, J.H. Melendez, Point-by-Point Progress: Gonorrhea Point of Care Tests, Expert Rev. Mol. Diagn. 20 (2020) 803–813. doi:10.1080/14737159.2020.1778467.

[206] R.W. Peeling, K.K. Holmes, D. Mabey, A. Ronald, Rapid tests for sexually transmitted infections (STIs): the way forward, Sex. Transm. Infect. 82 (2006) v1–v6. doi:10.1136/STI.2006.024265.

[207] A.M. Rompalo, Y.H. Hsieh, T. Hogan, M. Barnes, M. Jett-Goheen, J.S. Huppert, C.A. Gaydos, Point-of-care tests for sexually transmissible infections: what do 'end users' want?, Sex. Health. 10 (2013) 541–545. doi:10.1071/SH13047.

[208] L. Abou Chacra, F. Fenollar, Exploring the global vaginal microbiome and its impact on human health, Microb. Pathog. 160 (2021) 105172. doi:10.1016/J.MICPATH.2021.105172.

[209] B. Gieroba, A. Sroka-Bartnicka, P. Kazimierczak, G. Kalisz, I.S. Pieta, R. Nowakowski, M. Pisarek, A. Przekora, Effect of Gelation Temperature on the Molecular Structure and Physicochemical Properties of the Curdlan Matrix: Spectroscopic and Microscopic Analyses, Int. J. Mol. Sci. 21 (2020) 6154. doi:10.3390/ijms21176154.

[210] A.M.R. Gherman, N.E. Dina, V. Chiș, A. Wieser, C. Haisch, Yeast cell wall – Silver nanoparticles interaction: A synergistic approach between surface-enhanced Raman scattering and computational spectroscopy tools, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 222 (2019) 117223. doi:10.1016/J.SAA.2019.117223.

[211] A. Sujith, T. Itoh, H. Abe, A.A. Anas, K. Yoshida, V. Biju, M. Ishikawa, Surface enhanced Raman scattering analyses of individual silver nanoaggregates on living single yeast cell wall, Appl. Phys. Lett. 92 (2008) 103901. doi:10.1063/1.2891086.

[212] G.R. Huggins, G. Preti, Vaginal odors and secretions, Clin. Obstet. Gynecol. 24 (1981) 355–377. doi:10.1097/00003081-198106000-00005.

[213] I. Durickovic, M. Marchetti, Raman spectroscopy as polyvalent alternative for water pollution detection, IET Sci. Meas. Technol. 8 (2014) 122–128. doi:10.1049/IET-SMT.2013.0143.

[214] S.M. Berus, B. Młynarczyk-Bonikowska, G. Przedpełska, T. Szymborski, A. Kamińska, Surface-enhanced Raman spectroscopy analysis of menstrual cycle: from biochemical changes to diagnostics of vaginal infections, Sensors Actuators B Chem. (2024) 135571. doi:10.1016/J.SNB.2024.135571.

[215] N. Kuhar, S. Sil, T. Verma, S. Umapathy, Challenges in application of Raman spectroscopy to biology and materials, RSC Adv. 8 (2018) 25888–25908. doi:10.1039/c8ra04491k.

[216] M. Pecul, A. Rizzo, J. Leszczynski, Vibrational Raman and Raman optical activity spectra of D-lactic acid, D-lactate, and D-glyceraldehyde: Ab initio calculations, J. Phys. Chem. A. 106 (2002) 11008–11016. doi:10.1021/jp021030m.

[217] G.R. Huggins, G. Preti, Volatile constituents of human vaginal secretions, Am. J. Obstet. Gynecol. 126 (1976) 129–136. doi:10.1016/0002-9378(76)90477-4.

[218] D.H. Owen, D.F. Katz, A vaginal fluid simulant, Contraception. 59 (1999) 91–95. doi:10.1016/S0010-7824(99)00010-4.

[219] S.M. Berus, T. Szymborski, B. Młynarczyk-Bonikowska, G. Przedpełska, M. Adamczyk-Popławska, A. Kamińska, Identifying changes in vaginal fluid using SERS: Advancing diagnosis of vulvovaginal candidiasis, Sens. Bio-Sensing Res. 46 (2024) 100702. doi:10.1016/J.SBSR.2024.100702.

[220] K. Virkler, I.K. Lednev, Raman spectroscopy offers great potential for the nondestructive confirmatory identification of body fluids, Forensic Sci. Int. 181 (2008) e1–e5. doi:10.1016/J.FORSCIINT.2008.08.004.

[221] L. Ashton, V.L. Brewster, E. Correa, R. Goodacre, Detection of glycosylation and iron-binding protein modifications using Raman spectroscopy, Analyst. 142 (2017) 808–814. doi:10.1039/C6AN02516A.

[222] K. Czamara, K. Majzner, M.Z. Pacia, K. Kochan, A. Kaczor, M. Baranska, Raman spectroscopy of lipids: a review, J. Raman Spectrosc. 46 (2015) 4–20. doi:10.1002/JRS.4607.

[223] L.E. Kamemoto, S.K. Sharma, T. Acosta, H. Luk, A.K. Misra, A.C. Dykes, M.T. Goodman, Near-Infrared Micro-Raman Spectroscopy for in Vitro Detection of Cervical Cancer, Appl. Spectrosc. 64 (2010) 255–261. doi:10.1366/000370210790918364.

[224] J.L. Chicharro, A. Lucía, M. Pérez, A.F. Vaquero, R. Ureña, Saliva composition and exercise, Sport. Med. 26 (1998) 17–27. doi:10.2165/00007256-199826010-00002.

[225] P. Graphics Inc, Saliva Composition and Functions: A Comprehensive Review, J. Contemp. Dent. Pract. 9 (2008) 72–80.

[226] M. Baghizadeh Fini, Oral saliva and COVID-19, Oral Oncol. 108 (2020) 104821. doi:10.1016/J.ORALONCOLOGY.2020.104821.

[227] G. Cao, M. Chen, Y. Chen, Z. Huang, J. Lin, J. Lin, Z. Xu, S. Wu, W. Huang, G. Weng, G. Chen, A potential method for non-invasive acute myocardial infarction detection based on saliva Raman spectroscopy and multivariate analysis, Laser Phys. Lett. 12 (2015) 125702. doi:10.1088/1612-2011/12/12/125702.

[228] B. Isho, K.T. Abe, M. Zuo, A.J. Jamal, B. Rathod, J.H. Wang, Z. Li, G. Chao, O.L. Rojas, Y.M. Bang, A. Pu, N. Christie-Holmes, C. Gervais, D. Ceccarelli, P. Samavarchi-Tehrani, F. Guvenc, P. Budylowski, A. Li, A. Paterson, Y.F. Yun, L.M. Marin, L. Caldwell, J.L. Wrana, K. Colwill, F. Sicheri, S. Mubareka, S.D. Gray-

Owen, S.J. Drews, W.L. Siqueira, M. Barrios-Rodiles, M. Ostrowski, J.M. Rini, Y. Durocher, A.J. McGeer, J.L. Gommerman, A.C. Gingras, Persistence of serum and saliva antibody responses to SARS-CoV-2 spike antigens in COVID-19 patients, Sci. Immunol. 5 (2020). doi:10.1126/sciimmunol.abe5511.

[229] A. Torreggiani, S. Barata-Vallejo, C. Chatgilialoglu, Combined Raman and IR spectroscopic study on the radical-based modifications of methionine, Anal. Bioanal. Chem. 401 (2011) 1231–1239. doi:10.1007/s00216-011-5203-0.

[230] E. Decroly, I. Imbert, B. Coutard, M. Bouvet, B. Selisko, K. Alvarez, A.E. Gorbalenya, E.J. Snijder, B. Canard, Coronavirus Nonstructural Protein 16 Is a Cap-0 Binding Enzyme Possessing (Nucleoside-2′ O)-Methyltransferase Activity, J. Virol. 82 (2008) 8071–8084. doi:10.1128/JVI.00407-08.

[231] R.M. Hoffman, Q. Han, Oral Methioninase for Covid-19 Methionine-restriction Therapy, In Vivo (Brooklyn). 34 (2020) 1593–1596. doi:10.21873/INVIVO.11948.

[232] T. Ichimura, L. Da Chiu, K. Fujita, H. MacHiyama, T. Yamaguchi, T.M. Watanabe, H. Fujita, Non-label immune cell state prediction using Raman spectroscopy, Sci. Rep. 6 (2016) 37562. doi:10.1038/srep37562.

[233] N. Jagannathan, C. Thiruvengadam, P. Ramani, P. Premkumar, A. Natesan, H.J. Sherlin, Salivary Ferritin as a Predictive Marker of Iron Deficiency Anemia in Children, J. Clin. Pediatr. Dent. 37 (2012) 25–30. doi:10.17796/jcpd.37.1.ap20543762015370.

[234] L. Franco-Martínez, J.J. Cerón, M.R. Vicente-Romero, E. Bernal, A.T. Cantero, F. Tecles, C.S. Resalt, M. Martínez, A. Tvarijonaviciute, S. Martínez-Subiela, Salivary Ferritin Changes in Patients with COVID-19, Int. J. Environ. Res. Public Health. 19 (2021) 41. doi:10.3390/ijerph19010041.

[235] M.R. Giuca, M. Pasini, S. Tecco, G. Giuca, G. Marzo, Levels of salivary immunoglobulins and periodontal evaluation in smoking patients, BMC Immunol. 15 (2014) 1–5. doi:10.1186/1471-2172-15-5.

[236] I. Ettah, L. Ashton, Engaging with Raman Spectroscopy to Investigate Antibody Aggregation, Antibodies. 7 (2018) 24. doi:10.3390/antib7030024.

[237] L.M. Joke De Gelder, Kris De Gussem, Peter Vandenabeele, Reference database of Raman spectra of biological molecules, J. Pure Appl. Microbiol. 38 (2007) 1133–1147. doi:10.1002/jrs.

[238] G. Zhu, X. Zhu, Q. Fan, X. Wan, Raman spectra of amino acids and their aqueous solutions, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 78 (2011) 1187–1195. doi:10.1016/J.SAA.2010.12.079.

[239] H. Lee, M.S. Kim, S.W. Suh, Raman spectroscopy of sulphur-containing amino acids and their derivatives adsorbed on silver, J. Raman Spectrosc. 22 (1991) 91–96. doi:10.1002/JRS.1250220208.

[240] B. Fazio, C. D'Andrea, A. Foti, E. Messina, A. Irrera, M.G. Donato, V. Villari, N. Micali, O.M. Maragò, P.G. Gucciardi, SERS detection of Biomolecules at Physiological pH via aggregation of Gold Nanorods mediated by Optical Forces and

Plasmonic Heating, Sci. Rep. 6 (2016) 26952. doi:10.1038/srep26952.

[241] X. Lin, D. Lin, X. Ge, S. Qiu, S. Feng, R. Chen, Noninvasive detection of nasopharyngeal carcinoma based on saliva proteins using surface-enhanced Raman spectroscopy, J. Biomed. Opt. 22 (2017) 105004. doi:10.1117/1.JBO.22.10.105004.

[242] L.A. Austin, S. Osseiran, C.L. Evans, Raman technologies in cancer diagnostics, Analyst. 141 (2016) 476–503. doi:10.1039/C5AN01786F.

[243] E.M. Oliveira, M. Rogero, E.C. Ferreira, J.A. Gomes Neto, Simultaneous determination of phosphite and phosphate in fertilizers by Raman spectroscopy, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 246 (2021) 119025. doi:10.1016/J.SAA.2020.119025.

[244] Y. Wang, Y. Sun, M. Li, L. Xiong, X. Xu, N. Ji, L. Dai, Q. Sun, The formation of a protein corona and the interaction with α-amylase by chitin nanowhiskers in simulated saliva fluid, Food Hydrocoll. 102 (2020) 105615. doi:10.1016/j.foodhyd.2019.105615.

[245] B.N. Zamora-Mendoza, R. Espinosa-Tanguma, M.G. Ramírez-Elías, R. Cabrera-Alonso, G. Montero-Moran, D. Portales-Pérez, J.A. Rosales-Romo, J.F. Gonzalez, C. Gonzalez, Surface-enhanced raman spectroscopy: A non invasive alternative procedure for early detection in childhood asthma biomarkers in saliva, Photodiagnosis Photodyn. Ther. 27 (2019) 85–91. doi:10.1016/J.PDPDT.2019.05.009.

[246] P. Hu, X.S. Zheng, C. Zong, M.H. Li, L.Y. Zhang, W. Li, B. Ren, Drop-coating deposition and surface-enhanced Raman spectroscopies (DCDRS and SERS) provide complementary information of whole human tears, J. Raman Spectrosc. 45 (2014) 565–573. doi:10.1002/JRS.4499.

[247] K. Virkler, I.K. Lednev, Forensic body fluid identification: The Raman spectroscopic signature of saliva, Analyst. 135 (2010) 512–517. doi:10.1039/B919393F.

[248] Y. Sakakura, Y. Majima, S. Yoshii, T. Taniguchi, Y. Miyoshi, M. Ohyama, Nasal Secretion from Normal Subjects, Auris Nasus Larynx. 6 (1979) 71–78. doi:10.1016/S0385-8146(79)80010-3.

[249] R.D. Rossen, W.T. Butler, T.R. Cate, C.F. Szwed, R.B. Couch, Protein Composition of Nasal Secretion During Respiratory Virus Infection, Proc. Soc. Exp. Biol. Med. 119 (1965) 1169–1176. doi:10.3181/00379727-119-30406.

[250] A.C.C. Goulart, R.A. Zângaro, H.C. Carvalho, I.K. Lednev, L. Silveira, Diagnosing COVID-19 in nasopharyngeal secretion through Raman spectroscopy: a feasibility study, Lasers Med. Sci. 38 (2023) 1–8. doi:10.1007/s10103-023-03871-6.

[251] R. Bansil, B.S. Turner, The biology of mucus: Composition, synthesis and organization, Adv. Drug Deliv. Rev. 124 (2018) 3–15. doi:10.1016/J.ADDR.2017.09.023.

[252] R. Ahirwar, S. Gandhi, K. Komal, G. Dhaniya, P.P. Tripathi, V.M. Shingatgeri, K. Kumar, J.G. Sharma, S. Kumar, Biochemical composition, transmission and diagnosis of SARS-CoV-2, Biosci. Rep. 41 (2021).

doi:10.1042/BSR20211238/229295.

[253] A. Kamińska, E. Witkowska, A. Kowalska, A. Skoczyńska, I. Gawryszewska, E. Guziewicz, D. Snigurenko, J. Waluk, Highly efficient SERS-based detection of cerebrospinal fluid neopterin as a diagnostic marker of bacterial infection, Anal. Bioanal. Chem. 408 (2016) 4319–4327. doi:10.1007/s00216-016-9535-7.

[254] C.P.Y. Chan, J.W.Y. Choi, K.Y. Cao, M. Wang, Y. Gao, D.H. Zhou, B. Di, H.F. Xu, M.F. Leung, A. Bergmann, M. Lehmann, Y.M. Nie, G.W.H. Cautherley, D. Fuchs, R. Renneberg, B.J. Zheng, Detection of serum neopterin for early assessment of dengue virus infection, J. Infect. 53 (2006) 152–158. doi:10.1016/j.jinf.2005.11.008.

[255] G. Reibnegger, D. Fuchs, L.C. Fuith, A. Hausen, E.R. Werner, G. Werner-Felmayer, H. Wachter, Neopterin as a marker for activated cell-mediated immunity: application in malignant disease., Cancer Detect. Prev. 15 (1991) 483–490.

[256] B. Widner, C. Murr, B. Wirleitner, C. Mayr, N. Spöttl, G. Baier-Bitterlich, D. Fuchs, The importance of neopterin as a laboratory diagnostic marker of immune activation, Pteridines. 10 (1999) 101–111. doi:10.1515/pteridines.1999.10.3.101.

[257] M. Chauvin, M. Larsen, B. Quirant, P. Quentric, K. Dorgham, L. Royer, H. Vallet, A. Guihot, B. Combadière, C. Combadière, J. Barallat, J. Mayaux, C.E. Luyt, A. Mathian, Z. Amoura, J. Boddaert, F. Armestar, G. Gorochov, E. Martinez-Caceres, D. Sauce, Elevated Neopterin Levels Predict Fatal Outcome in SARS-CoV-2-Infected Patients, Front. Cell. Infect. Microbiol. 11 (2021) 764. doi:10.3389/fcimb.2021.709893.

[258] R. Bellmann-Weiler, L. Lanser, F. Burkert, S. Seiwald, G. Fritsche, S. Wildner, A. Schroll, S. Koppelstätter, K. Kurz, A. Griesmacher, G. Weiss, Neopterin Predicts Disease Severity in Hospitalized Patients With COVID-19, Open Forum Infect. Dis. 8 (2021) ofaa521. doi:10.1093/OFID/OFAA521.

[259] W. Hailemichael, M. Kiros, Y. Akelew, S. Getu, H. Andualem, Neopterin: A Promising Candidate Biomarker for Severe COVID-19, J. Inflamm. Res. 14 (2021) 245. doi:10.2147/JIR.S290264.

[260] P. Harikrishnan, Saliva as a Potential Diagnostic Specimen for COVID-19 Testing, J. Craniofac. Surg. 31 (2020) e653–e655. doi:10.1097/SCS.0000000000006724.