2021-05-05

Raman Spectroscopic Characterisation of Non Stimulated and Stimulated Human Whole Saliva

Genecy Calado
Technological University Dublin

Isha Behl
Technological University Dublin

Hugh Byrne
Technological University Dublin, hugh.byrne@tudublin.ie

See next page for additional authors

Authors

Genecy Calado, Isha Behl, Hugh Byrne, and Fiona Lyng

Raman spectroscopic characterisation of non stimulated and stimulated human whole saliva.

Genecy Calado[1,2], Isha Behl[1,2], Hugh J. Byrne[3], Fiona M. Lyng[1,2].

[1]Radiation and Environmental Science Centre, FOCAS Research Institute, Technological University Dublin, City Centre Campus, Dublin 8, Ireland

[2]School of Physics and Clinical and Optometric Sciences, Technological University Dublin, City Centre Campus, Dublin 8, Ireland

[3]FOCAS Research Institute, Technological University Dublin, City Centre Campus, Dublin 8, Ireland

Corresponding Author: Genecy Calado <caladodentist@gmail.com>

Running title: Raman Analysis of Saliva.

Abstract

Human saliva is a unique biofluid which can reflect the physiopathological state of an individual. The wide spectrum of molecules present in saliva, compounded by the close association of salivary composition to serum metabolites, can provide valuable information for clinical diagnostic applications through highly sensitive vibrational spectroscopic techniques such as Raman spectroscopy. However, the nature of saliva, in terms of collection and patient-related characteristics, can be considered factors which may strongly affect the Raman spectral profile of salivary samples and disrupt the search for specific salivary biomarkers in the detection of diseases. The main objective of this study was to highlight spectral features associated with the type of collection in an intra- and inter-patient approach. Saliva was collected using both stimulated and non-stimulated approaches from 20 donors, concentrated by centrifugal filtration and further analysed using Raman spectroscopy. The methodology adopted for liquid saliva showed consistency in the qualitative analysis of the groups, confirming the reproducibility of this Raman spectroscopic approach. Using principal component analysis (PCA) and partial least squares – discriminant analysis (PLSDA), non stimulated saliva could be differentiated from stimulated saliva in both intra- and inter-patient analysis, with a classification efficiency of 77 and 87%, respectively. The bicinchoninic acid (BCA) assay showed a similar trend in terms of total protein concentration, showing a slight increase in stimulated saliva samples. These results are valuable in the process of developing and establishing Raman spectroscopy as a novel diagnostic tool in the future as well as controlling variability, in order to determine specific spectroscopic markers related to a multifactorial disease for diagnostic or follow-up purposes.

1. Introduction

Saliva is considered a dynamic biological fluid that has a large range of constituents, including proteins, polypeptides, nucleic acids, electrolytes, and hormones [1]. It is categorised as an exocrine secretion of the salivary glands, which is hypotonic in nature, with a pH of 7.2–7.4, or in some conditions slightly acidic [1]. Whole saliva is unique and complex, both in source and composition. It consists not only of secretions from the three major salivary glands (parotid, submandibular and sublingual) and the minor glands, but also gingival crevicular fluid, oral mucosa transudate, secretions from nasal and pharyngeal mucosa, keratin debris and blood cells [2,3].

Interest in human saliva as a potential diagnostic and prognostic fluid is steadily increasing because it provides access not only to relevant oral but also systemic disease information [4,5]. Saliva has been identified as functionally equivalent to blood serum, reflecting the physiological state of the body, including hormonal, nutritional, and metabolic variations, for example [6], and its composition can be linked with traditional biochemical parameters which appear in the serum [7]. Salivary collection is usually considered as one of the easiest methods of collection of bodily fluids, due to its noninvasive nature, which also does not require specialised equipment or supervision. Further, saliva collection is usually a well accepted procedure, from the patients' point of view, and is a cost-effective approach. Oral fluid sampling is also safe for both the operator and the patient and is amenable to repeated and voluminous sampling in short intervals of time. Considering these advantages, saliva is identified as a potential source of biological sample to be employed for specific patient screening as well as routine diagnostic screening [8,9].

It is very important, however, to highlight the fact that the composition of each saliva sample tends to vary and depend on the type of gland of origin and, consequently, the type of collection employed to obtain these samples [10]. Its composition differs according to the contribution of

each gland in order to obtain the total unstimulated saliva secretion, and can vary from 65%, 23%, and 8% to 4% for the submandibular, parotid, and sublingual glands, for example [11].

When resting, without exogenous or chemical stimulation, non stimulated saliva is characterised by its low and continuous salivary flow, denoted basal unstimulated secretion, present in the form of a film that covers, moisturises, and lubricates the oral tissues. In contrast, stimulated saliva is produced by a range of mechanical, gustatory, olfactory, or pharmacological stimuli, and contributes to around 80% to 90% of daily salivary production [12].

Clinically, the most practical way to differentiate between non stimulated and stimulated saliva is usually by its salivary flow rate. In adults, normal total stimulated salivary flow ranges from 1 to 3 mL/min, low ranges from 0.7 to 1.0 mL/min, while hyposalivation is characterised by less than 0.7 mL/min [13,14]. In normal, non stimulated salivary function, it ranges from 0.25 to 0.35 mL/min, low ranges from 0.1 to 0.25 mL/min, while hyposalivation is characterised by a salivary flow of less than 0.1 mL/min [14]. However, although widely adopted in clinics, the values denoted "normal" for stimulated and non stimulated salivary flow exhibit large biological variation [14].

Despite the complexity of the salivary milieu and the anatomy of salivary glands, the analysis of saliva and its different collection forms represent a promising approach to establishing potential biomarkers for several pathological conditions [15]. In this context, the scientific development of new technologies and associated "omics" approaches provide opportunities for the determination of new biomarkers for the diagnosis, staging, or prognosis of diseases [16].

Vibrational spectroscopy is one such evolving set of techniques which allows analysis of a multitude of biological samples, including saliva [17-20]. Recently, vibrational spectroscopic analysis of saliva has proven efficient to differentiate chronic periodontitis from aggressive periodontitis [21], diagnose type 2 diabetes and psoriasis [22], detect drugs [23], and discriminate smoking from non-smoking subjects [24]. Amongst the vibrational spectroscopic

methods, Raman spectroscopy (RS) can be considered a unique non-invasive laser-based analytical technique that aids biochemical component analysis [25]. RS is based on the molecular vibrations that are specific to certain types of biomolecules, including proteins, nucleic acids and lipids [25]. The Raman effect is based on vibration transitions under inelastic scattering of monochromatic light in visible, near ultraviolet or near infrared ranges. RS can thus provide a characteristic fingerprint of the molecular vibrations that are specific to certain types of biomolecules, including proteins, nucleic acids and lipids [25]. RS, in contrast to conventional biochemical methods, is a label-free and rapid technique, which usually requires only a small quantity of a sample without any preparation [26].

It has been reported that RS of saliva can be used for narcotic usage detection [27], for cancer diagnosis [28,29], and in forensic medicine [30]. It is notable, however, that the current literature lacks technical information regarding the methodologies employed [31]. Thus, there is an urgent need for a systematic optimisation of analysis protocols governing RS analysis of saliva samples. Most studies reported to date have focused on individual proteins under specific conditions, with the type of stimulation varying greatly [32]. Studies looking at protein changes in human saliva have typically analysed samples from individual glands, not whole saliva [33]. Notably, the literature is particularly scant on details regarding the sample collection protocols, and differentiation of stimulated and unstimulated production of saliva and on human whole saliva composition.

In an attempt to establish a standard RS protocol for analysis of saliva samples, as well as to better clarify factors correlated to the sampling procedure, such as type of collection, the aim of the present study was to develop, based on Raman spectra of saliva samples, a pre-analytical workflow to highlight spectral features associated with intra- and inter-patient characteristics which could further help to extract specific salivary diagnostic signatures of systemic or local pathological conditions.

2. Methodology

2.1 Subjects

Ethical approval to collect saliva samples from healthy donors was granted by the Technological University Dublin Research Ethics Committee (REC ref: 15/104). Written informed consent was obtained from each donor and the study was conducted in accordance with ethical principles founded in the Declaration of Helsinki.

2.2 Collection of saliva samples

Saliva samples were collected by both non-stimulated and stimulated techniques. In both techniques of saliva collection, all subjects were instructed to refrain from smoking, eating, drinking and tooth brushing for 1h prior to saliva collection. In each case, saliva was collected between 9:00 a.m. and 12:00 a.m., to minimise any interference of food. The participants rinsed their mouth with distilled water prior to collection for one minute, and waited five minutes before the collection commenced.

Resting drooling (minimal oral movements), known as the non-stimulated collection method, was used to collect about 2 mL of whole saliva from the oral cavity of healthy volunteers from Technological University Dublin. The saliva providers were asked to sit comfortably in an upright position and tilt their heads down slightly to pool saliva in the floor of the mouth. The first expectoration was discarded to eliminate food debris and unwanted substances which may contaminate the sample and cause analytical inaccuracy. Subsequently, the samples were expectorated into a pre-labelled, sterile, 15 mL plastic container (Nalgene, Eppendorf).

Stimulated whole saliva was also collected, by asking the volunteers to chew on a tasteless piece of parafilm (5x5cm, 0.30 g; Parafilm 'M'; American National CAL, Chicago, IL, USA). The first expectoration was discarded and the chewing-stimulated saliva was also expectorated

into test tubes, every 30s for two minutes. During the saliva collection period, the subjects chewed at their natural pace.

A total number of 30 saliva samples were collected from 20 volunteers. Non stimulated saliva and stimulated saliva were both collected from each of 10 of the donors, such that stimulated and non stimulated samples from the same donors could be compared. The remaining 10 samples were collected according to the stimulated protocol from 10 different donors, such that stimulated and non stimulated samples from different donors could be compared. From each sample, 10 spectra were acquired.

All salivary samples were aliquoted directly into 1 mL cryotubes and stored at -80°C. Although the samples could be centrifuged and analysed immediately, freezing was carried out for transportation from the clinic to the analytical laboratory, notably to minimise any enzymatic sample degradation [34]. They were further subjected to a freeze–thaw cycle to break down mucopolysaccharides [35], consequently reducing viscosity and minimising pipetting errors. Before spectroscopic measurement, the saliva sample was allowed to defrost for approximately 10 minutes at 4 C°.

2.3 Centrifugal filtration of saliva samples

Commercially available centrifugal filtration devices, Amicon Ultra- 0.5 mL (Millipore – Merck, Germany), with cut-off points at 3K, were employed in this study. Reported by Bonnier et al., the centrifugal filtration methodology was then adapted to concentre the saliva samples, as they retain constituent components only above a size of 3 kDa, allowing much of the aqueous sample pass to the filtrate [36].

As indicated by the manufacturer, the ultrafiltration membranes in Amicon® Ultra-0.5 devices "contain trace amounts of glycerine", which, as demonstrated by Bonnier et al. [37], can contaminate spectral analysis. Washing of the centrifugal devices prior to saliva analysis was therefore carried out by spinning the Amicon Ultra-0.5 mL once with a solution of NaOH (0.1

M) followed by two rinses with Milli-Q water (Millipore Elix S). For both washing and rinsing, 0.5 mL of the respective liquid was added to the filters and the centrifugation was applied for 30 minutes at 14000g followed by a spinning with the devices upside down at 1000g for two minutes in order to remove any residual solution contained in the filter.

For the sample concentration, 0.5 mL of saliva was placed in a 3K Centrifugal filtration device and centrifuged at $14000 \times g$ for 30 minutes. The filter devices were then placed upside down in a new Eppendorf and spun down at 1000g for 2 min in order to collect the remainder of the saliva (concentrate) retained in the filter devices. The concentrating factor is of the order of 10, with a resultant concentrate volume of ~70 µL. As a result, one fraction was obtained, representing proteins/components with a molecular weight higher than 3 kDa.

2.4 Instrumentation

A Raman Horiba Jobin Yvon LabRam HR 800, inverted, confocal Raman spectroscopic microscope was used to record the spectra from the concentrated saliva samples. The microscope has an automated xy stage and is coupled to a Peltier cooled CCD detector. A 50 mW diode laser with 532 nm wavelength was used, with a grating of 600 grooves/mm, while the confocal hole was set at 100 µm. A 96 well-plate with glass bottom (Thermo Fisher number 1, 0.17 mm thickness) was used as substrate. For the acquisition, 10 different regions were selected randomly using a 60X objective (MPLAN N Olympus, Japan), which also collected the backscattered light. The spectra were acquired over 3 accumulations, totalling 2 minutes per spectrum. A spectral fingerprint range from 400 to 1800 cm^{-1} was recorded for further analysis.

2.5 BCA assay

Total protein concentration of 9 randomly selected saliva samples (3 non stimulated samples and 3 stimulated samples from the same donors, and 3 stimulated samples from different

donors) was estimated via the bicinchoninic acid (BCA) protein assay (Micro BCA Protein Assay Kit - Thermo Scientific) by following the instructions of the manufacturer. The BCA assay is colorimetric based, giving a dark purple colour when two molecules of BCA chelate with protein and form a compound of the cuprous ion. The absorbance of the complex was measured at 562 nm using a microplate reader (Beckman Coulter Co.). BCA standard reagents A, B and C were freshly mixed in the ratio of 25:24:1. Bovine serum albumin (2 mg/ml) was used as a standard, with 13 working standards 0.5–2000 µg/mL. All the tubes (standards, test samples, and blank) were incubated at 37°C for 2 hours. After incubation, absorbance was measured at 562 nm against a reagent blank. The concentration of test samples was measured with reference to standards for further analysis.

2.6 Data Analysis

Pre-processing procedures

The spectral data processing was carried out using Matlab (MATHWORKS, US) with the PLS-Toolbox (Eigenvector Research Inc.) and in-house algorithms. Saliva, even when concentrated through centrifugal filtration, is a very dilute sample, and consequently the spectra acquired tend to be rather noisy. To improve the spectral quality, the raw Raman spectra were first smoothed using a Savitzky-Golay filter. The Savitzky-Golay parameters were incrementally increased to 13 pts and 9th order, which appeared to give satisfactory improvement without introducing distortion of the spectral features.. The baseline correction was applied using the rubberband method [38], and the spectra were vector normalised, aiming to reduce any variability caused by the fluctuation of measurement conditions or instrumental parameters.

Spectral correction method by non-negative least squares (NNLS)

To deal with possible interferences from the background that may mask important biological features, the non-negative least squares method was used to remove glass and/or water residuals in the saliva spectra. This in-house model considers the spectral data obtained as linear functions resulting from the underlying saliva components and the water background and glass substrate [39]. It aims to reconstitute a vector x that explains the observed spectra as well as possible, based on known observations. So, given the spectra obtained and a set of known observations, such as a matrix of (1) 60 glass spectra and 60 water spectra recorded from the model set samples considered in the study (see Supplemental Figure S1) and (2) a selection of 9 saliva components (from a pool of 11 components used to prepare artificial saliva according to the formula of Klimek *et al.* [40]), which were recorded at their maximum concentration in water (see Supplemental Figure S2, Figure S3 and Table S1) following the same parameters of instrumentation used for saliva samples; it is possible to find a nonnegative vector that estimates the contribution of these known observations to the spectra. The known observations are then multiplied by the nonnegative vector before being subtracted from the initial spectral matrix, correcting for both the glass and water contributions in saliva samples.

This method of correction was also successfully applied for wax and glass removal in formalin fixed paraffin preserved oral tissues by Ibrahim *et al.* [39]. Also, a recent study has demonstrated the same versatility of the NNLS method for glass correction in oral cytological samples [41].

The formula of Klimek et al. [40] was designed mainly to study dental erosion in *in-vitro* models. Only 9 out of 11 saliva components were recorded due to their suitable chemical properties allowing a Raman signal to be acquired. Sodium chloride and monopotassium chloride were rapidly dissociated in water affecting the Raman spectra and essentially providing a spectrum corresponding to water. As a result, the individual spectra of these components were not considered. Also, due to the inability of the mucin component to adequately represent the glycoprotein/protein content in the saliva spectrum (see Supplemental Figure S2), the spectrum

of an extra component (IgG – solubility 50 mg/L), was also included in the unsupervised analysis (Supplemental Figure S2). This component was used to better understand the protein content of saliva through the analysis of the weight of each component used in NNLS, and was also used in the unsupervised analysis. Concomitantly, the spectral information of the constituents of artificial saliva was used for peak assignment in analysis the real saliva samples, where appropriate.

Statistical Analysis

The pre-processed and corrected spectra of saliva samples were initially subjected to Principal Components Analysis (PCA) to allow an unsupervised evaluation of the variability existing in the data sets itself, as well as among non-stimulated saliva and stimulated saliva.

Furthermore, partial least squares discriminant analysis (PLSDA) was also used for further classification. Similar to PCA, PLSDA is a form of multivariate analysis which works as a linear classifier that aims to maximise the variance between groups and minimise the variance within groups, albeit in a supervised way. It is based on partial least squares regression (PLSR) [42]. Leave one patient out cross validation (LOPOCV) was applied, by leaving out all the spectra from each patient (in this case, donor) in turn during the cross validation of the classifier. Saliva spectral datasets of both groups were mean centered to exclude any common variances.

To further evaluate the performance of the PLSDA algorithm for differentiating between the three saliva groups, receiver operating characteristic (ROC) curves were also generated. Sensitivity was calculated from the fraction of "in class" spectra, while the specificity was calculated from the fraction of "not in class" spectra for a given threshold (0,1) based on the cofusion matrices generated by the PLSDA analysis of each classifier. The cross validated ROC curves follow the same method, except the class predicted when the spectra are left out during cross validation is used. The ROC curves are plots of the true positive rate (sensitivity) against the false positive rate (1-specificity) over a continuous range (from 0 to 1) of cut-points of a classifier. Each

point on the ROC curve represents a sensitivity/specificity pair, corresponding to a particular decision threshold. The statistical analysis tool, PLS-Toolbox, automatically provided a ROC curve of the threshold (from 0 to 1), based on the PLSDA confusion matrix analysis for each classifier. The estimated ROC curves are therefore based on predicted class for each group/observation. The area under the curve (AUC) was used to measure the quality of the model's predictions.

Data obtained from the BCA assay was subjected to statistical analysis (2 paired t-test) and $p < 0.05$ was deemed to be significant.

3. Results

The inverted sample measurement protocol has previously been described by Bonnier et al. [37] for the case of analysis of human serum, and was further adapted by Parachalil et al. [43] to glass coverslip bottomed chambers and by Medipally et al. [44] to glass coverslip bottomed 96 well plates. Using this methodology, as little as 30 μL can be measured reproducibly (**Figure 1a**).

The Raman spectrum of pure, as collected saliva is, however, dominated by that of water, as shown in **Figure 1b**, and features of biochemical constituents are not discernible. The constituents can be concentrated by centrifugal filtration, according to the protocols described by Bonnier et al. [36], ensuring the filtration devices have been thoroughly rinsed before use [37]. Centrifugal filtration using 3kDa centrifugal filtering devices seemed to provide a relevant Raman signal revealing biomolecular information in the spectra valuable for future analysis for oral cancer detection (**Figure 1b**).
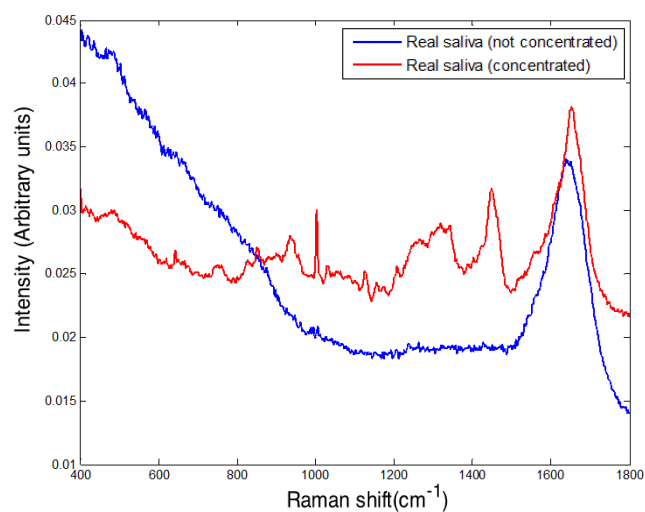
a)
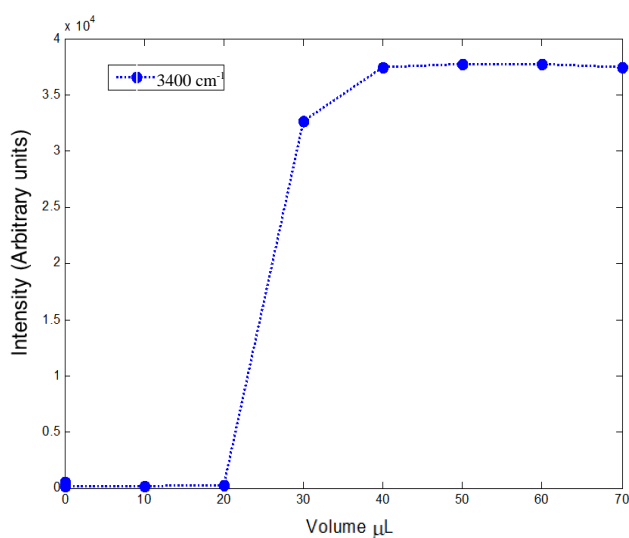
b)

Figure 1: (**a**) Raman intensity variation of different volume of water between 0-70 µL at 3400 cm^{-1}; and (**b**) 532 nm raw Raman spectrum focused by a 60X (water immersion) objective onto a sample of real saliva (around 75% more concentrated by centrifugal filtration) (blue) and a samples of real saliva not concentrated (before centrifugal filtration) (red) using a 96 well-plate with glass bottom no. 1 as substrate.

As the intrinsic contribution of water from each sample and possible residual contribution from the glass in the finger print region (400-1800 cm^{-1}) could interfere with the acquisition of the overall Raman spectra (**Figure 2**), those bands were removed using the NNLS method.
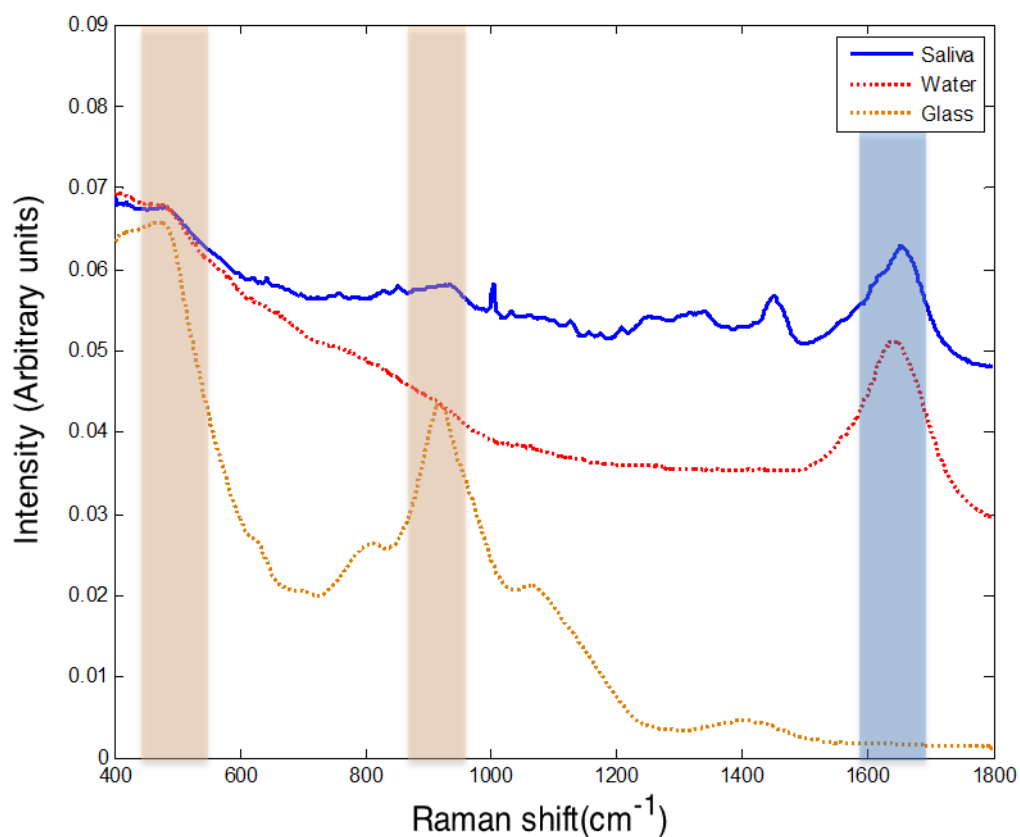
Figure 2: Raw Raman spectrum of a sample of filtered saliva (blue spectrum), water spectrum (red spectrum) and glass coverslip no. 1 (orange spectrum). The blue rectangular region denotes the strong influence of water on the important protein finger-print range. The orange rectangular region denotes the possible finger-print range where there could be spectral glass contamination.

Artificial saliva of the formula of Klimek et al. [40] was employed to help the biochemical assignment of the features of the filtered serum spectra. For comparison with the concentrated, filtered saliva, a concentration of the recommended formulation of 1:1.75 was used (Supplemental S5). Due to the inability of mucin to account for the spectral profile in the protein regions in saliva (e.g. the Amide I region ~1500- 1800cm^{-1}) (see Supplemental Figure S6), IgG was included along with 9 saliva components to analyse the weight of the components in different saliva groups (see Supplemental Figure S7). The weighted sum of all components was

acquired by the combination of components associated with least residual error and it was compared to each corresponding group mean spectrum.

Although the "fit" is still far from perfect, the inclusion of IgG in the spectral analysis significantly improves the correspondence, particularly in the region of the Amide I. In the non stimulated saliva samples, the least residual error showed that the component that best fit was IgG (glycoprotein) followed by mucin. In contrast, both groups of stimulated saliva samples (from same donors and different donors) had more contribution of mucin followed by IgG. These results support the use of IgG also for correction along with the other nine saliva components.

After the required pre-processing and corrections, the mean Raman spectra of the groups analysed; (i) non stimulated saliva, (ii) stimulated saliva (from the same donors as the non stimulated samples) and (iii) the second group of stimulated saliva (different donors), can be seen in **Figure 3**. The major vibrational assignments for all three groups, based on literature data [26, 30, 45, 46], can be seen in Table 1. The NNLS correction model seems to confer a significant improvement, in this case, on the water and glass subtraction.
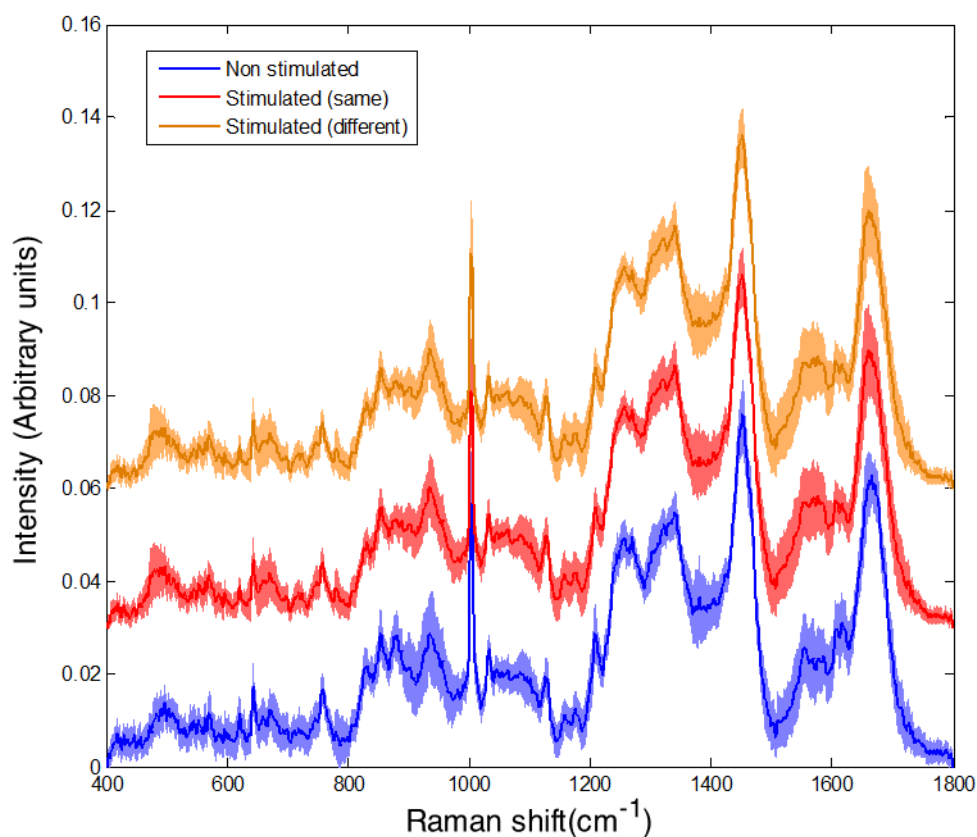
Figure 3: Mean Raman spectra of non stimulated saliva (blue), stimulated saliva from the same donors as non stimulated saliva (red), and stimulated saliva from different donors (orange). Spectra have been offset for clarity and the shading denotes standard deviation.

Table 1: Assignment of the main saliva proteins in the Raman bands to biomolecules [26, 30, 45, 46].

Raman shift/cm^{-1}	Major assignments
502	S-S disulphide stretching band (collagen)
654	Phenylalanine
760	Tryptophan

852	Proline/tyrosine
878	Hydroxyproline
938	Proline
1003	Phenylalanine
1032	Phenylalanine
1127	C-N stretching (proteins)
1208	Tryptophan
1268	Amide III
1340	Collagen
1450	Proteins
1552	Tryptophan
1658	Amide I

Qualitative analysis

In terms of composition, it is clear that the saliva mean fingerprint from each sample set is dominated by the polypeptide backbone of protein, represented by the amide I, C-H deformation bands and aromatic ring breathing peaks at 1658 cm^{-1}, 1450 cm^{-1} and 1003 cm^{-1}, respectively. Based on the current literature, these bands can be related to various glycoproteins that are known to be constituents of saliva, especially mucin matrices [46]; when correlated to the panel of saliva components recorded and available (see Supplemental Figure S2), these

peaks show compatibility with spectral features of IgG and mucin, reaffirming the high presence of glycoproteins/proteins.

There are also some peaks, including those in the low wavenumber range of 400-550 cm^{-1}, that indicate the presence of saccharide. The salivary mucus is rich in mucopolysaccharides, also known as glycosaminoglycans, which can explain this possible assignment [30]. When linked to the components present in Supplemental Figure S2, glucose, for example, shows similar spectral features, such as peaks at 422, 448 and 520 cm^{-1}, as does mucin (as it is also a protein with agglutination properties). Furthermore, considering again the saliva components used for NNLS correction (see Supplement Figure S3), some peaks in this range could also potentially represent some salivary electrolytes (in other words the buffer content of saliva), such as sodium phosphate (540 cm^{-1}), potassium phosphate (516 cm^{-1}) and calcium chloride (480 cm^{-1}).

Saliva is also known for its high concentration of proline-rich proteins. These type of proteins are one of the major components of the saliva from the parotid and submandibular gland in humans but mainly secreted by the submandibular gland [47]. The Raman peaks at 852 cm^{-1}, 878 cm^{-1} and 938 cm^{-1} are known to be correlated to proline presence and can also be easily assigned in the mean spectra of non-stimulated saliva [30].

The bands at 760 cm^{-1}, 1032 cm^{-1}, 1208 cm^{-1} and 1340 cm^{-1} are bands related to a wide range of proteins and lipids. When compared to the panel of artificial saliva components used for correction, these peaks also show some correlation with spectral bands of IgG (glycoproteins). Furthermore, the 1127 cm^{-1} peak seems to be related to carbohydrate, very commonly found in the oral environment [30].

Unsupervised analysis of non-stimulated saliva versus stimulated saliva

PCA was applied to gain more information on the differences between the non stimulated and stimulated saliva samples in general (**Figure 4a** and **Figure 4b**).
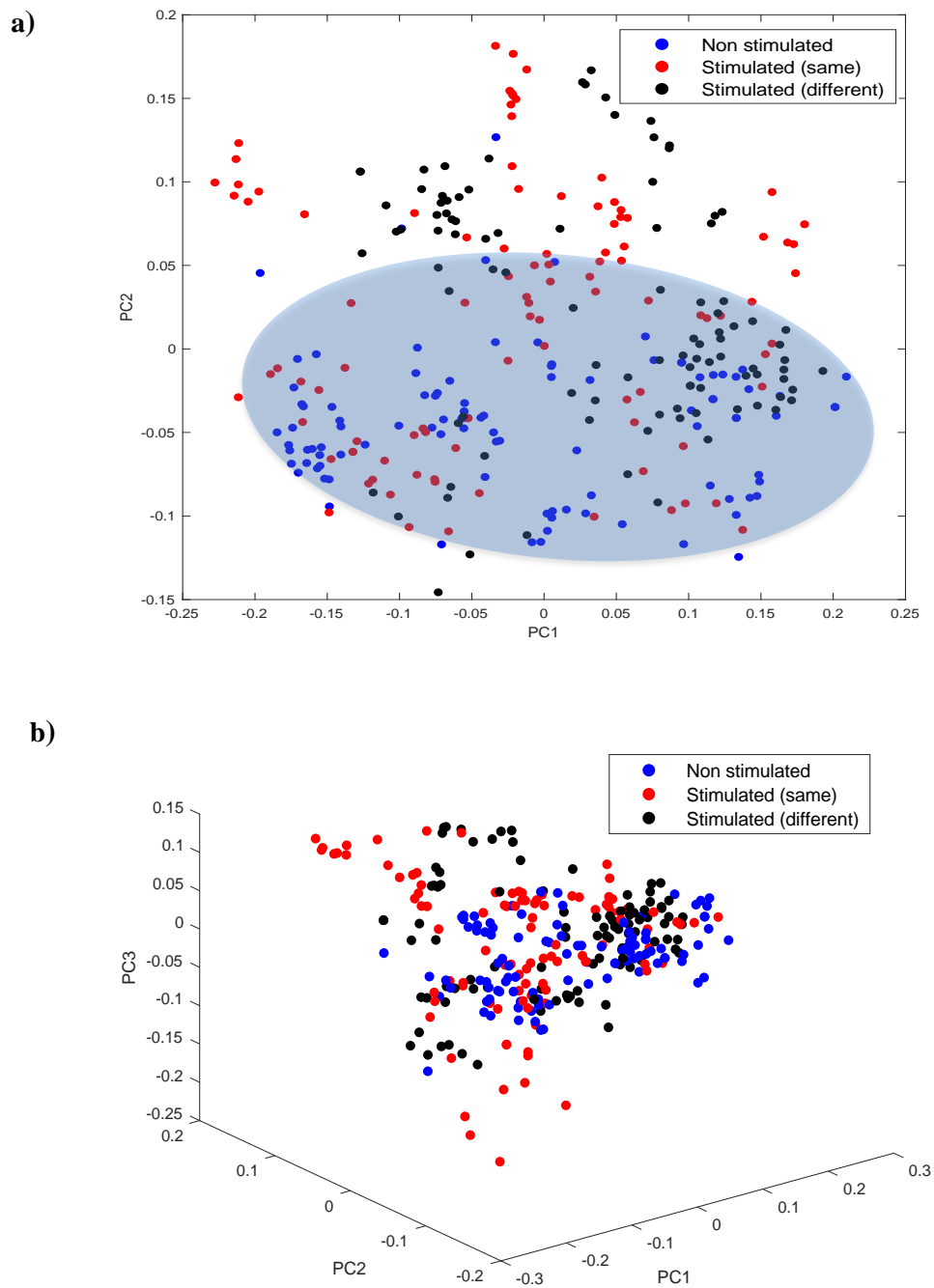

Figure 4: PCA of non stimulated and stimulated saliva from the same and different donors showing overlap between the groups according to PC1 in (**a**) 2D scatterplot and (**b**) 3D scatterplot. The PC2 axis shows, however, some degree of differentiation of the non stimulated saliva samples, as denoted by the blue ellipse.

PCA revealed a significant variance amongst the groups, indicating that the biochemical features of the saliva samples, either stimulated or non-stimulated, from the same or different donors, leads to little or no differentiation with respect to the PC1 axis, (**Figure 4**). Similar behaviour can be seen in a PCA of the dataset, following NNLS correction without IgG (Supplemental Figure S8). However, with respect to the PC2 axis, the non stimulated saliva group seems to be more tightly aggregated, showing more biochemical homogeneity (**Figure 4a**).

Although no clear differentiation of the sample types was achieved, the PC1 loading indicates that the variability across the samples could be correlated to protein (**Figure 5a**), featuring prominent protein bands such as 938 cm^{-1}, 1004 cm^{-1}, 1128 cm^{-1}, 1384 cm^{-1}, 1450 cm^{-1}, 1565 cm^{-1} and 1652 cm^{-1}, mentioned previously in the qualitative analysis. The same behaviour could be seen in the PC1 loading where the NNLS correction did not include IgG (Supplemental Figure S9).

The PC2 loading (**Figure 5b**), for which a more clear differentiation of the non stimulated saliva samples was achieved, suggests that the higher intensity of the major negative peaks (1004, 1450, 1450 and 1669 cm^{-1}) could be mostly assigned to higher concentration of glycoprotein/proteins in the non stimulated saliva, directly correlated to the Raman bands of IgG from the panel of components used in the NNLS correction (see Supplemental Figure S3).
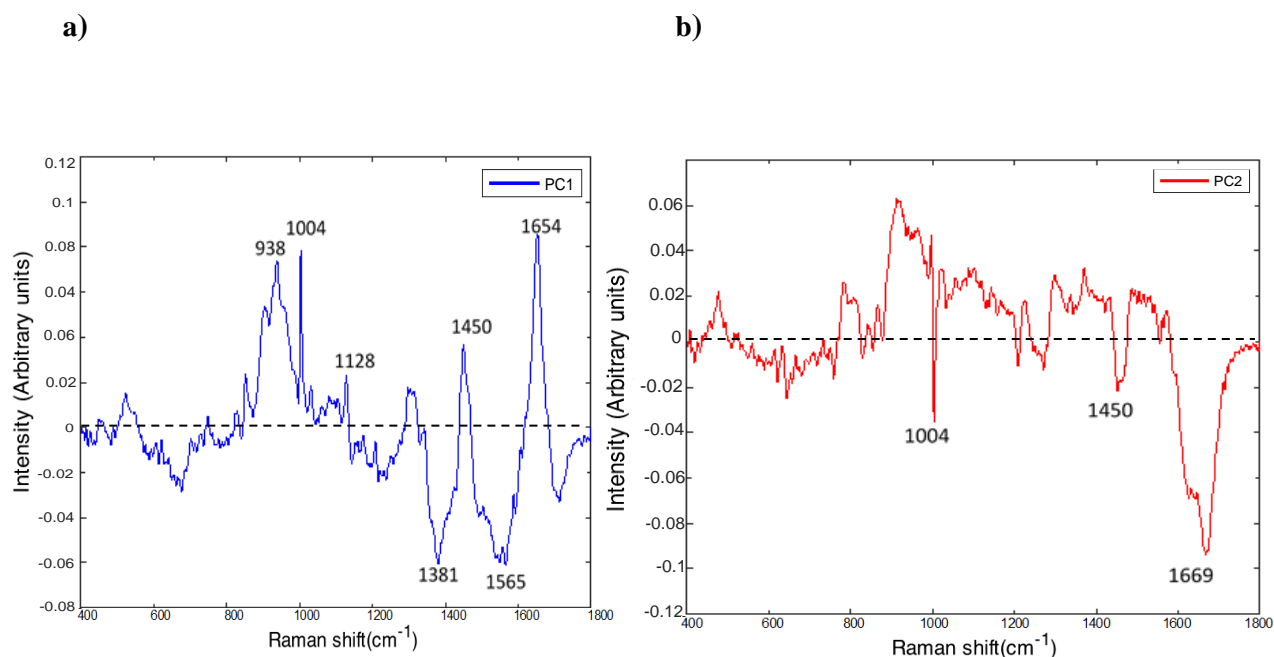
a)   b)

Figure 5: (**a**) PC1 loading from PCA analysis and (**b**) PC2 loading from PCA analysis.

The inclusion of IgG to represent the glycoprotein content of saliva samples, although it does not significantly improve the discrimination power of PCA for the stimulated saliva samples, was considered important for the classification of the non stimulated samples due not only to the improved "fit" with this component but also due to a clearer discrimination according to PC2 (Supplemental Figure S8), and was consequently used for in the supervised analysis.

The distribution of the different groups of saliva samples in PCA (Figure 4) suggests that non stimulated saliva seems to be, in an overall analysis, biochemically distinct (negative side of PC2), while stimulated saliva might represent a "mixture" of stimulated and non stimulated, composition-wise (spread across the negative and the positive side of PC2). Notably, the mean of the distributions of stimulated and non stimulated samples in PCA are offset, with respect to PC2, and therefore it is more appropriate to compare the mean spectra of the different groups to further analyse its composition.

Supervised analysis of non-stimulated saliva versus stimulated saliva – same donors

Stimulated and non stimulated saliva from the same donors showed very similar Raman spectral features when compared (**Figure 6**), apart from the spectral range around 1575 cm^{-1} (glycoprotein related). Visually comparing the mean spectrum from each group (**Figure 6a**), all the same major peaks are displayed in both, and in general the spectra appear very similar to each other. However, the intensity of some peaks has changed, notable those in the range of ~1500 - 1700 cm^{-1}.

These differences in intensity are further confirmed when the mean spectra of the non stimulated saliva group was subtracted from the mean spectra of the stimulated saliva group (same donors) (**Figure 6b**), showing again a higher presence of glycoproteins (1575 cm^{-1}) associated with the stimulated saliva. Following the approach of Ryzhikova et al. [48], the difference spectrum is displayed along with ± (x0.004) the standard deviation of the individual groups, indicating that the differences fall within 2 standard deviations of the mean, and that multivariate statistical methods are required to extract the subtle differences between the groups.
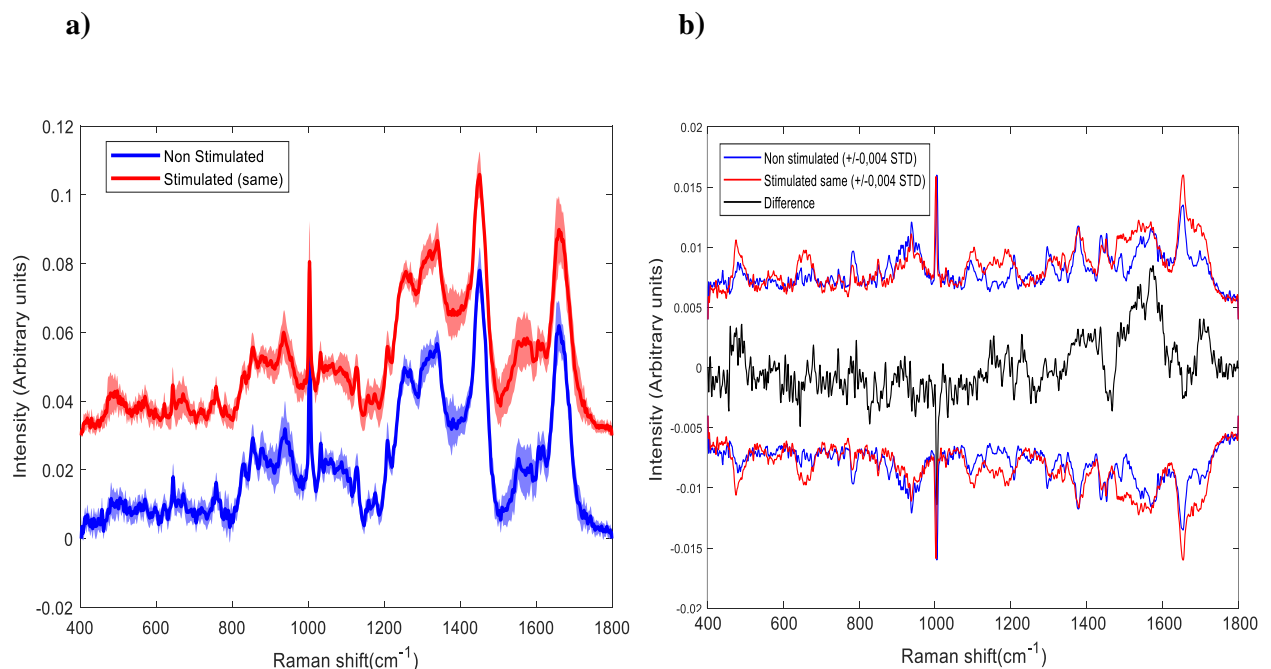
a)　　　　　　　　　　　　　　　　　　　　　b)

Figure 6: (**a**) Offset mean spectra of non stimulated saliva and stimulated saliva from the same donors with shading indicating the standard deviation and (**b**) difference spectrum of non stimulated saliva mean spectrum subtracted from stimulated (same donors) mean spectrum compared to $\pm$ (x0.004) the standard deviation of the individual groups.

PLSDA, along with the LOPOCV method, was subsequently utilised as a classification algorithm employed to quantitatively differentiate stimulated from non-stimulated saliva from the same donors. The resultant, cross validated, probability prediction plot indicates that it is possible to classify and separate the non-stimulated saliva group from the stimulated saliva group, although they come from the same donors (**Figure 7**). In the same way, the confusion matrix obtained (in a balanced analysis of 100 spectra collected) indicates sensitivity and specificity of 77% and 78%, respectively.
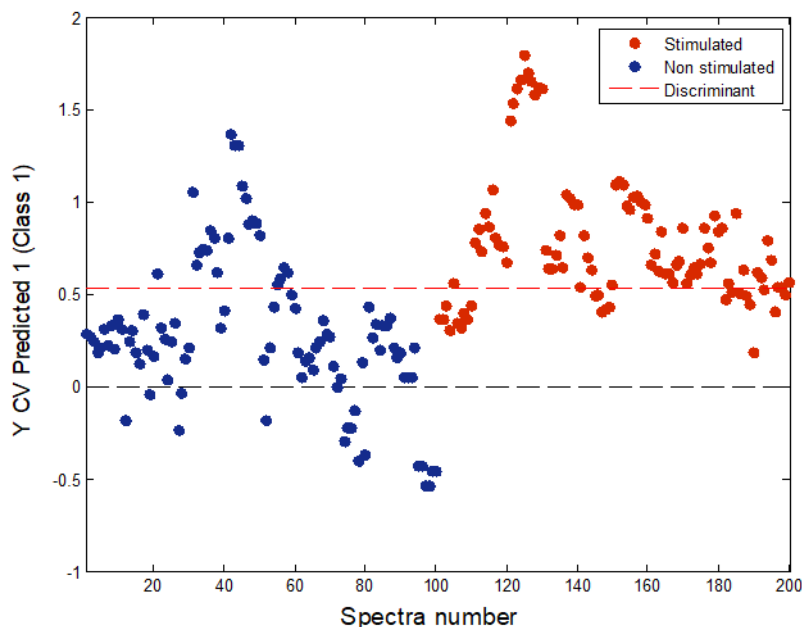

Figure 7: Cross validated probability prediction plot showing the discrimination between non stimulated saliva and stimulated saliva from the same donors. The discriminant (red line) is considered as the latent variable where the data best classifies.

Based on the ROC curves (**Figure 8a** and **Figure 8b**), the classifier was able to obtain excellent discrimination between non stimulated saliva (AUC=0.844) and stimulated saliva (AUC=0.844).
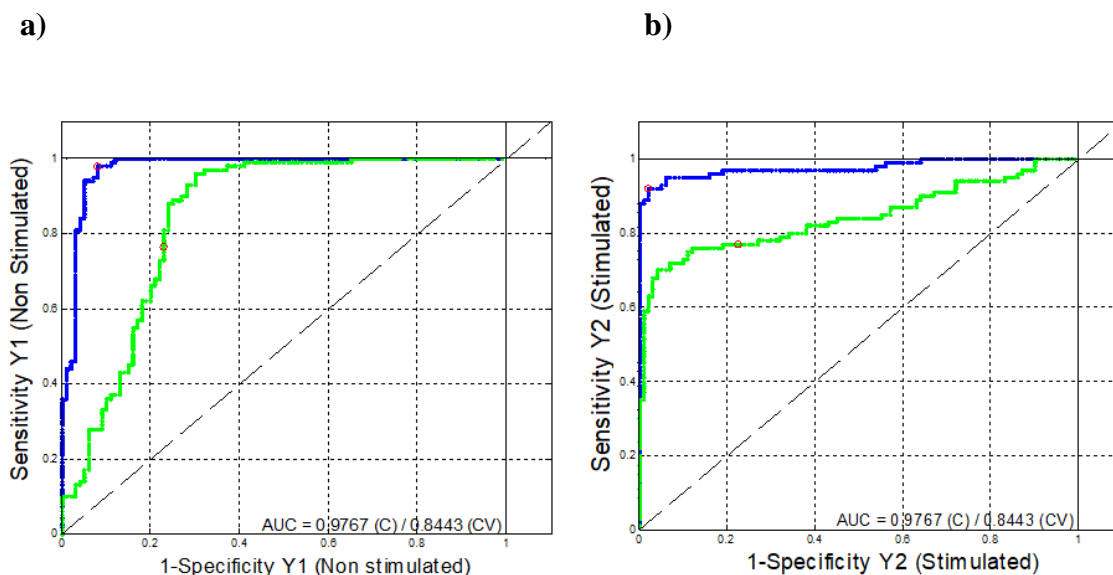
a)

b)

Figure 8: ROC curves for (**a**) non stimulated saliva samples and (**b**) stimulated saliva samples from the same donors. AUC is a measure of accuracy of the classifier, (C) is calibrated and (CV) is the cross validated AUC. The red dots represent the calculated sensitivity and 1-specificity on the y and x axis, respectively.

Supervised analysis of non-stimulated saliva versus stimulated saliva – different donors

As a second approach aiming to further classify these samples, the saliva samples from different donors were compared, non stimulated saliva samples from 10 donors (used in the first analysis), and stimulated from a new group of 10 different donors. Consistent with the analysis of samples from the same donors, only the intensity of some peaks seemed to change while the

general spectral appearance of the salivary profile of the non stimulated and stimulated groups continued to be very similar (**Figure 9a**).

Following the analysis with the same donors, similar differences in intensity could be further confirmed when the mean spectra of the non stimulated saliva group was subtracted from the mean spectra of the stimulated saliva group (different donors) (**Figure 9b**), showing this time a reduced presence of proteins (1575 cm^{-1}) associated with the stimulated saliva from different donors. It, however, could also show a higher intensity on proline related peak (938 cm^{-1}) associated with the non stimulated saliva. Again, the comparison with $\pm$ (x0.004) the standard deviation of the individual groups indicates that multivariate statistical methods are required to extract the subtle differences between the groups.

a)   b)
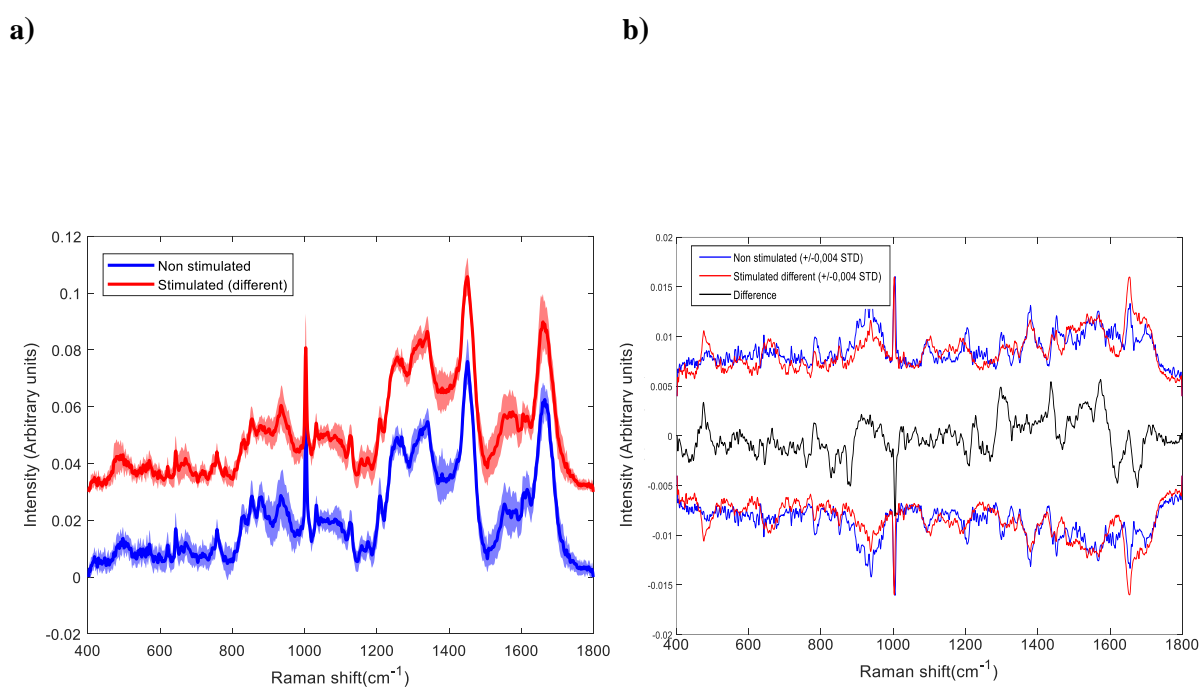

Figure 11: (**a**) Offset mean spectra of non stimulated saliva and stimulated saliva from the different donors with shading indicating the standard deviation and (**b**) difference spectrum of non stimulated saliva mean spectrum subtracted from stimulated (different donors) mean spectrum compared to $\pm$ (x0.004) the standard deviation of the individual groups.

To further assess the accuracy of saliva spectra, spectral differences of non stimulated saliva and stimulated saliva from different donors were mean centered and also explored in detail by the PLSDA multivariate algorithm and LOPOCV. The cross validated probability prediction plot from the different donors (**Figure 10**) could show an efficient classification between the groups which was confirmed with sensitivity and specificity (in a balanced analysis of 100 spectra) of 88% and 86%, respectively.
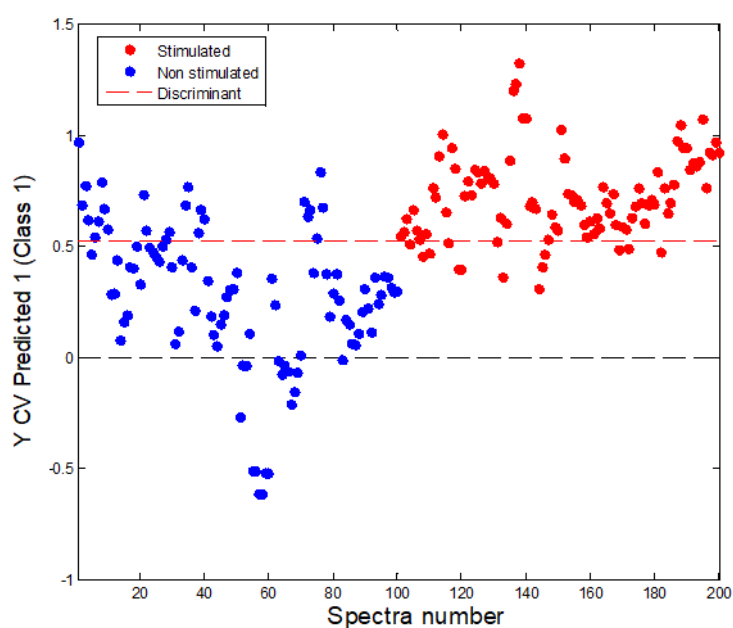

Figure 10: Mean spectra of non stimulated saliva and stimulated saliva from different donors.

The ROC curve plot for the second analysis showed that the classifier had an even better accuracy (AUC=0.8550) for both classes (non stimulated and stimulated) of samples from different donors (**Figure 11a** and **Figure 11b**).
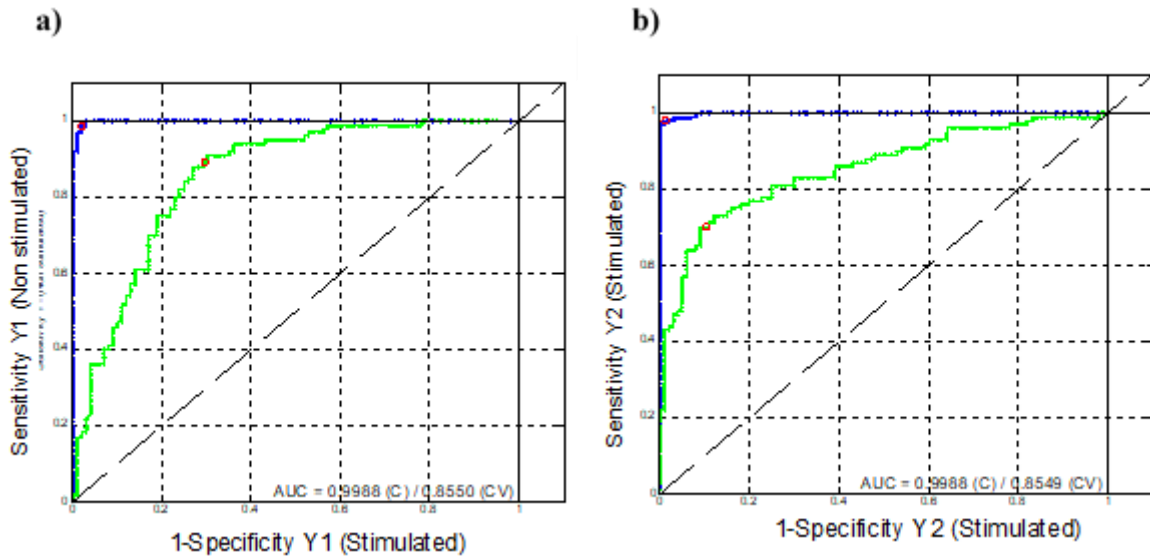
Figure 11: ROC curves for (**a**) non stimulated saliva samples and (**b**) stimulated saliva samples from distinct donors.

In terms of estimation of the total concentration of proteins, the BCA assay results did not show a significant difference between the non stimulated saliva samples and the stimulated saliva samples from the same donors ($p = 0.584$) or different donors ($p = 0.370$). However, the mean concentration of total proteins was slightly increased in stimulated saliva from the same donors (**Figure 12a**) and different donors (**Figure 12b**).

Figure 12: BCA assay showing the mean total protein concentration (µg/mL) and standard deviation (error bars = 95% confidence intervals) according to the different types of saliva in the same donors (**a**) and different donors (**b**).

4. Discussion

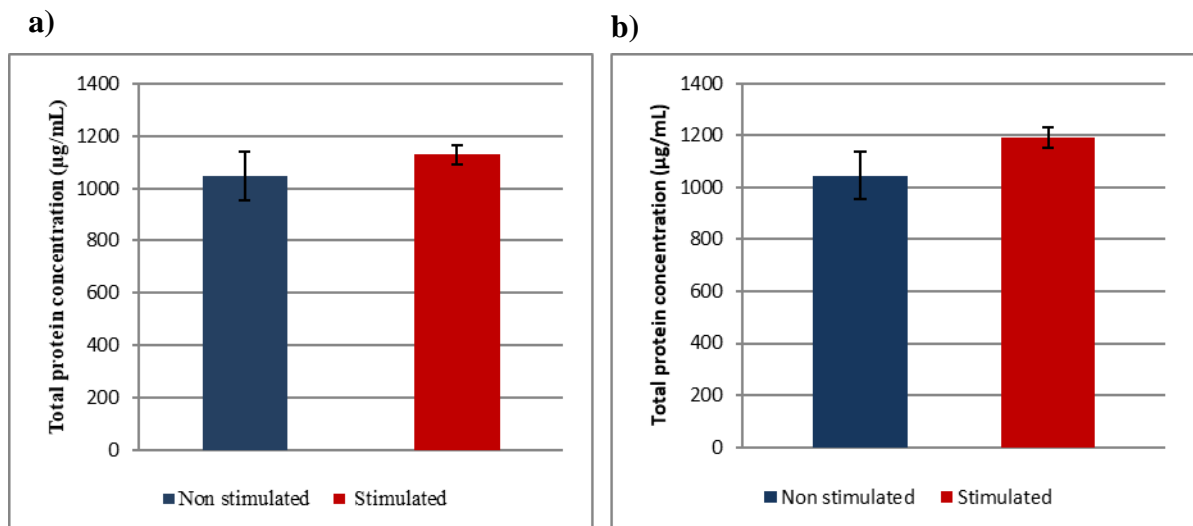In the development of label free spectroscopic methodologies for potential screening of biofluids for biomedical applications, standardisation of sample collection and analysis protocols is critical. In the case of saliva, sample collection can be performed under two conditions: stimulated and non stimulated. Non stimulated saliva is collected by drooling the saliva in the mouth and draining it in a wide bore sterile vessel or by swabbing or suction methods. Stimulated whole saliva is collected by masticatory action, that is, chewing paraffin wax or by gustatory stimulation by applying acetic acid in the mouth followed by collection of saliva [1,49].

Few reports have explored the effect of stimulation on human whole saliva composition. Most studies have focused on individual proteins under specific conditions, with the type of stimulation varying greatly. In an attempt to use non stimulated saliva to detect juvenile idiopathic arthritis, despite small divergences, no differences in the protein salivary status between patients and the control group were found [50]. Studies looking at protein changes in human saliva have typically analysed samples from individual glands, not whole saliva and many of these studies did not account for important variables, such as intra- and inter-patient differences [51]. Nevertheless, protein concentration is known to be influenced by type of stimulation, glandular source, etc. [52].

Discriminating chemometrically between non-stimulated whole saliva and stimulated whole saliva is an important step to confirm and standardise the applicability of Raman spectroscopy

for diagnostic purposes. To the best of our knowledge, this is the first study of its kind to compare the composition of stimulated and non-stimulated saliva based on the Raman spectral features as well as the individual spectral biochemical fingerprint based on an intra- and inter-patient approach.

The spectral components of saliva are complex, and have contributions from multiple constituent chemical species. For both saliva types, the spectra are dominated by water, proteins and electrolytes. These results consequently correlate with several literature sources, where the cited majority chemical components of saliva in the highest concentrations are electrolytes, mucus, antibacterial compounds, and various enzymes, and water [53-55].

As part of the preprocessing procedure, water can be removed using NNLS fitting, using spectra of constituents of a known artificial saliva formulation. It was noted however, that the combined spectra of the formulation did not account well for the protein content evident in the saliva spectra, and thus IgG was added to the mix of constituents. Also, it is important to highlight the fact that the artificial saliva formulation of Klimek et al., although representing a comprehensive mixture of saliva components, might not be representative enough for the complex spectroscopic band assignment, as this formula was primarily created to see the *in vitro* effects of dental erosion [40]. This explains the need for IgG as the salivary glycoprotein representative for improvement of the spectral fit (see Supplemental Figure S6 and S7).

When comparing the two different saliva groups, the qualitative analysis (spectral composition) of non stimulated saliva and stimulated saliva initially showed that, in both the intra-patient and inter-patient approach, the mean spectra of the different saliva groups had similar features. From the intra-patient approach, these results would be expected due to the fact that each saliva sample had the same donor source, despite being collected in a different way, which would mean a minimal variation in terms of biological composition of these samples. Thus, the same behaviour would also be expected from the inter-patient analysis due to the essential

components of saliva being still quite similar even when collected from different individuals [54].

Generally, the intensity of some bands, as seen in the mean spectra of each group, changed, and this is understandable since the relative contribution of the chemical species in saliva will likely change with each donor and can even change within the same donor throughout the day [30]. Besides, it is important to highlight the fact that, for example, the proline band at 938 cm^{-1} had a higher intensity in the non-stimulated group which could be explained by the major contribution of the submandibular gland in non stimulated samples as compared to stimulated samples [47]. However, higher intensity of glycoprotein related peaks, those at 1575 cm^{-1}, in the difference spectra of the groups can suggest that stimulated saliva also contains valuable protein contributions in its composition for possible clinical analysis.

It is already known that the average daily flow of whole saliva varies in healthy individuals between 1 and 1.5 L and the type of collection/stimulation. Percentage contributions of the different salivary glands during unstimulated flow are 20% from parotid, 65% from submandibular, 7% to 8% from sublingual, and less than 10% from numerous minor glands. Stimulated high flow rates drastically change the percentage contributions from each gland, with the parotid contributing more than 50% of total salivary secretions [56]. These different contributions might be responsible for the overlapping distribution of the stimulated saliva samples with the non stimulated saliva samples in PCA.

Since the structure and composition of saliva is complex, and given that the spectral profile from different groups based on stimulation and non stimulation are very similar, it was necessary to develop a more sophisticated and robust diagnostic model based on PCA and PLSDA by utilising the entire spectrum to determine the most diagnostically significant spectral features for classification of the saliva. However, there is of course a potential for a large variety

of contaminants in any saliva sample due to the eating habits of a particular donor [52], but these interferences do not appear to affect the spectroscopic signature.

In an unsupervised analysis, PCA did not provide a clear differentiation between the different saliva samples regardless of NNLS correction with IgG. However, PC1 of both analyses has shown that the slight differences in spectral profiles could be attributed to bands at 1004, 1128, 1450, 1655 cm^{-1}. As expected, these bands seem to correspond to the bond stretching of vs (C-C) of phenylalanine and Amide I which could be used to identify salivary proteins [57]. These results are also consistent with contributions from mucin and IgG (glycoproteins) spectral features from the recorded components, providing confirmation of such assignments.

Furthermore, the overall spectral profile of stimulated saliva seemed to have more influence from peaks at 938, 1442-50 cm^{-1} which could be correlated to proline and lipids in comparison to non stimulated samples from different donors. These results indicate that proline-rich proteins which are found in abundance in non stimulated saliva, can be subject to inter-patient variation in salivary composition. However, a possible association with calcium chloride and urea, according to our own panel of components, may open a possibility that Raman spectroscopy was able to detect possible differences in the microbiome of stimulated saliva, as urea is a resultant component of bacterial proliferation in the oral environment; and calcium chloride (present in the acquired pellicle) could be a result of the mechanical action during collection [58].

PCA showed that stimulated saliva was quite diverse (as mixture of both stimulated and non stimulated components), whereas non stimulated saliva was more homogeneous (or pure). According to the PC2 loading, non stimulated saliva also seemed to be strongly influenced by a peak at 1669 cm^{-1}, to which amino acids could be correlated [59]. On the other hand, the difference spectra of both groups could highlight the higher concentration in other glycoproteins in stimulated saliva [56].

To further complement the Raman spectral analysis, the BCA assay results seemed to be consistent with the fact that there are no major differences in the protein concentration of the two types of saliva samples. Although not statistically significant, stimulated samples had a slightly increased protein concentration in comparison to the non stimulated samples. This result is in accordance with some variability in intensity of some protein peaks, such as 1575 cm^{-1} seen in the difference spectrum. Ultimately, BCA results were able to confirm the rich general content of protein in saliva samples but could not differentiate the two groups based on protein concentration specifically. This could be attributed not only to the inter-patient issues but perhaps also attributed to the reduced number of samples used in this biochemical test.

Several factors may modify the salivary concentration. Thus, the composition of unstimulated saliva is different from stimulated saliva (which is more similar in composition to plasma) [60]. For example, an increase in the salivary flow rate, obtained by stimulation with acidic food, increases the concentrations of sodium, chloride and bicarbonate and decreases the concentration of salivary potassium and phosphate, compared with unstimulated saliva [61].

Stimulated saliva has certain drawbacks as the foreign substances which stimulate the saliva tend to modify the pH and the water phase of salivary secretion. However, for practical reasons, stimulated saliva samples may be preferred over non-stimulated saliva samples, as these can be collected in higher volumes and considerably faster than non-stimulated saliva samples or when in a clinical environment [1]. Not surprisingly, the proteomic profile of stimulated saliva samples has been reported to be diluted when compared with non-stimulated saliva samples, which is why non-stimulated saliva samples are preferred for proteomic analysis of saliva [62,63]. Other studies, however, have reported that the salivary protein content increased in stimulated saliva samples [64].

5. Conclusion

The field of salivary diagnostics has undeniable translational and clinical potential. Continuing advancements in vibrational spectroscopy technologies have revealed unprecedented insights toward understanding salivary composition as part of the body's overall health. Correct interpretation and utilization of this information may be useful not only for identifying local and systemic disorders but also, perhaps, to aid in the treatment strategies.

Raman spectroscopy can be of real interest for diagnostic purposes in case of complex diseases with multiple confounding factors. This is even more the case for salivary extracts whose biochemical composition may be affected by several conditions, not only the type of collection but also comorbidities related to complex diseases.

The qualitative results show that the specificity of the Raman signature of liquid saliva samples and its potential ability to be used as an identification technique for diagnostic purposes in the future due to its reproducibility even in different conditions of collection and considering inter-patient (donor) variability.

In an unsupervised analysis, PCA was not able to differentiate the different saliva samples, showing minimal changes inherent to individual saliva composition. However, it did indicate that non stimulated saliva was significantly more biochemically homogeneous, compared to stimulated saliva.

The good sensitivity and specificity obtained by PLSDA revealed that, even with high spectral similarities correlated to the salivary composition, the classifiers could provide differentiation between the groups, mainly between non stimulated samples and stimulated samples from different donors.

With a standardised collection procedure and protocol, the use of salivary samples for Raman spectroscopy can be a promising diagnostic method that can allow a novel non invasive and

cost-effective approach. Also, having set guidelines standardising the procedure, as the method proposed by this study, could resolve any confounding issues between studies and alleviate some of the inherent variability among individuals and populations when using saliva samples for Raman spectroscopic clinical analysis.

6. References

1. S. Sindhu, N. Jagannathan, Saliva: A Cutting Edge in Diagnostic Procedures, Journal of Oral Diseases, (2014), pp. 1-8. doi: https://doi.org/10.1155/2014/168584.

2. M. Castagnola, P.M. Picciotti, I. Messana, C. Fanali, A. Fiorita, T. Cabras, L. Calo, E. Pisano, G.C. Passali, F. Iavarone, G. Paludetti, E. Scarano, Potential applications of human saliva as diagnostic fluid. Acta Otorhinolaryngol Ital., 31 (2011), pp. 347-357.

3. Y.S. Cheng, T. Rees, J. Wright, A review of research on salivary biomarkers for oral cancer detection, Clin Transl Med., 24 (2014), p. 3. https://doi.org/10.1186/2001-1326-3-3

4. N.M. Ralbovsky, L. Halámková, K. Wall, C. Anderson-Hanley, I.K. Ledne, Screening for Alzheimer's Disease Using Saliva: A New Approach Based on Machine Learning and Raman Hyperspectroscopy, J Alzheimers Dis, 71 (2019):1351-1359. 10.3233/JAD-190675. PMID: 31524171

5. N. Buduneli, D.F. Kinane, Host-derived diagnostic markers related to soft tissue destruction and bone degradation in periodontitis, J Clin Periodontol, 38 (2011), pp. 85-105. DOI: 10.1111/j.1600-051X.2010.01670.x

6. Y.H. Lee, D.T. Wong, Saliva: an emerging biofluid for early detection of diseases, American Journal of Dentistry, 22 (2009), pp. 241-248.

7. C. Silvia, A. Giorgia, G. Rosalba, F. Elio, P. De, Saliva specimen: a new laboratory tool for diagnostic and basic investigation, Clinica Chimica Acta, vol. 383 (2007), pp. 30-40.

8. D. Malamud, Saliva as a diagnostic fluid, Malamud D. Saliva as a diagnostic fluid. Dent Clin North Am., 55 (2011), pp. 159-178. https://doi.org/10.1016/j.cden.2010.08.004

9. C.F. Streckfus, L.R. Bigler, Saliva as a diagnostic fluid, Oral Diseases, 8 (2002), pp. 69-76. https://doi.org/10.1034/j.1601-0825.2002.1o834.x

10. S. Hu, P. Denny, P. Denny, Y. Xie, J.A. Loo, L.E. Wolinsky, Y. Li, J. McBride, R.R.O. Loo, M. Navazesh, D.T. Wong Differentially expressed protein markers in human submandibular and sublingual secretions, Int J of Oncol., 25 (2004), pp. 1423-1430.

11. S. Chiappin, G. Antonelli, R. Gatti, E.F. De Palo, Saliva specimen: a new laboratory tool for diagnostic and basic investigation, Clin Chim Acta. 383 (2007), pp. 30-40, 2007 https://doi.org/10.1016/j.cca.2007.04.011

12. V. de Almeida Pdel, A.M. Grégio, M.A. Machado, A.A. de Lima, L.R. Azevedo, Saliva composition and functions: a comprehensive review, Contemp Dent Pract., 9 (2008), pp. 72-80.

13. J. Tenovuo, F. Lagerlöf. Saliva, in: A. Thylstrup, O. Fejerskov, Textbook of clinical cariology., second ed., Copenhagen, 1994.

14. M. Edgar, C. Dawes, D. O'Mullane, Saliva and oral health. third ed., London, 2004.

15. M.J. Baker, S.R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R.A. Lukaszewski, G. Thiéfin, G.D. Sockalingum, Developing and understanding biofluid vibrational spectroscopy: a critical review, Chem Soc Rev., 45 (2016), pp. 1803-1818.

16. G. Kuku, M. Saricam, F. Akhatova, A. Danilushkina, R. Fakhrullin, M. Culha, Surface-enhanced Raman scattering to evaluate nanomaterial cytotoxicity on living cells, Anal Chem., 88 (2016), pp. 9813–9820.

17. M.J. Baker, S.R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R.A. Lukaszewski, G. Thiéfinbg, G.D. Sockalingum, Developing and understanding biofluid vibrational spectroscopy: a critical review, Chem Soc Rev. 45 (2016), pp. 1803-1818.

18. S. Fendel, B. Schrader, Investigation of skin and skin lesions by NIR-FT-Raman spectroscopy, Fresenius J Anal Chem, 360 (1998), pp. 609–613. https://doi.org/10.1007/s002160050767

19. J.J. Mikkonen, J. Raittila, L. Rieppo, R. Lappalainen, A.M. Kullaa, S. Myllymaa, Fourier Transform Infrared Spectroscopy and Photoacoustic Spectroscopy for Saliva, Analysis. Appl Spectrosc., 70 (2016), pp. 1502-1210.

20. S. Derruau, C. Gobinet, A. Mateu, V. Untereiner, S. Lorimier, O. Piot, Shedding light on confounding factors likely to affect salivary infrared biosignatures, Anal Bioanal Chem., 411 (2019), pp. 2283-2290. doi: 10.1007/s00216-019-01669-6.

21. N.S. Ozek, I. Zeller, D.E. Renaud, P. Gümüş, N. Nizam, F. Severcan, N. Buduneli, D.A. Scott, Differentiation of chronic and aggressive periodontitis by FTIR spectroscop, J Dent Res., 95 (2016), pp. 1472-1478.

22. U. Bottoni, R. Tiriolo, S.A. Pullano, S. Dastoli, G.F. Amoruso, S.P. Nistico, S.P. Nisticò, A.S. Fiorillo, Infrared saliva analysis of psoriatic and diabetic patients: similarities in protein components, IEEE Trans Biomed Eng, 63 (2016), pp. 379-384. doi: 10.1109/TBME.2015.2458967.

23. S. Armenta, S. Garrigues, M. de la Guardia, J. Brassier, M. Alcalà, M. Blanco, Analysis of ecstasy in oral fluid by ion mobility spectrometry and infrared spectroscopy after liquid-liquid extraction, J Chromatogr A., 1384 (2015), pp. 1-8.

24. E. Al-Hetlani, L. Halámková, M.O. Amin, I.K. Lednev, Differentiating smokers and nonsmokers based on Raman spectroscopy of oral fluid and advanced statistics for forensic applications, J Biophotonics, 13 (2020), e201960123. doi: 10.1002/jbio.201960123.

25. S.A. Gonchukov, T.V. Lonkina, S.A. Minaeva, A.V. Sundukov, T.E. Migmanov, J. Lademann, M.E. Darvin, V.N. Bagratashvili, Confocal Raman microscopy of pathologic cells in cerebrospinal fluid, Laser Phys Lett, 11 (2014), pp. 1-4.

26. S.A. Gonchukov, A. Sukhinina, D. Bakhmutov, S. Minaeva, Raman spectroscopy of saliva as a perspective method for periodontitis diagnostics, 9 (2012), 73-77.

27. M.J. West, M.J. Went. Detection of drugs of abuse by Raman spectroscopy, Drug Test Anal., 3 (2011), pp. 532-538. doi: 10.1002/dta.217.

28. J.C.Y. Kah, K.W. Kho, C.G.L. Lee, C.J.R. Sheppard, Z.X. Shen, K.C. Soo, M.C. Olivo, Early diagnosis of oral cancer based on the surface plasmon resonance of gold nanoparticles, Int J Nanomedicine, 2 (2007), pp. 785-798.

29. S. Farquharson, C. Shende, F.E. Inscore, P. Maksymiuk, A. Gift, Analysis of 5-fluorouracil in saliva using surface-enhanced Raman spectroscopy, J.Raman Spectrosc., 36 (2005), 208-212.

30. K. Virkler, I.K. Lednev, Forensic body fluid identification: the Raman spectroscopic signature of saliva, Analyst, 135 (2010), pp. 512-517. doi: 10.1039/b919393f.

31. G. Calado, I. Behl, A. Daniel, H.J. Byrne, F.M. Lyng, Raman spectroscopic analysis of saliva for the diagnosis of oral cancer: a systematic review. Translational Biophotonics. (2019), e201900001. doi:i10.1002/ tbio.201900001

32. X. Fang, L. Yang, W. Wang, T. Song, C.S. Lee, DL DeVoe, B.M. Balglel, Comparison of electrokinetics-based multidimensional separations coupled with electrospray ionizationtandem mass spectrometry for characterization of human salivary proteins. Anal Chern., 79 (2007), pp. 5785-5792.

33. A. Walz, K. Stuhler, A. Wattenberg, E. Hawranke, H.E. Meyer, G. Schmalz, M. Blüggel, S. Ruhl, Proteome analysis of glandular parotid and submandibular-sublingual saliva in comparison to whole human saliva by two-dimensional gel electrophoresis, Proteomics, 6 (2006): pp. 1631-1639.

34. D. Esser, G. Alvarez-Llamas, M.P. de Vries, D. Weening, R.J. Vonk, H. Roelofsen, Sample Stability and Protein Composition of Saliva: Implications for Its Use as a Diagnostic Fluid, Biomark Insights, 3 (2008), pp. 25-27. doi:10.4137/bmi.s607

35. R. Mohamed, J.L. Campbell, J. Cooper-White, G. Dimeski, C. Punyadeera, The impact of saliva collection and processing methods on CRP, IgE, and Myoglobin immunoassays, Clin Transl Med., 1 (2012), p. 19. doi:10.1186/2001-1326-1-19

36. F. Bonnier, F. Petitjean, M.J. Baker, H.J. Byrne, Improved protocols for vibrational spectroscopic analysis of body fluids, J Biophotonics, 7 (2014), pp. 167-179. doi: 10.1002/jbio.201300130.

37. F. Bonnier, M.J. Baker, H.J. Byrne, Vibrational spectroscopic analysis of body fluids: avoiding molecular contamination using centrifugal filtration, Analytical Methods, 6 2014), pp. 5155-5160.

38. J. Trevisan, P.P. Angelov, P.L. Carmichael, A.D. Scott, F.L. Martin, Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. Analyst, 137 (2012), pp. 3202-3215. doi: 10.1039/c2an16300d.

39. O. Ibrahim, A. Maguire, A.D. Meade, S. Flint, M. Toner, H.J. Byrne, F.M. Lyng, Improved protocols for pre-processing Raman spectra of formalin fixed paraffin preserved tissue sections, Anal. Methods, (2017), 4709-4717. DOI: 10.1039/c6ay03308c

40. J. Klimek, E. Hellwig, G. Ahrens, Effect of plaque on fluoride stability in the enamel after amine fluoride application in the artificial mouth, Dtsch Zahnarztl Z., 10 (1982), pp. 836-840.

41. I. Behl, G. Calado, O. Ibrahim, A. Malkin, S. Flint, H.J. Byrne, F.M. Lyng, Development of methodology for Raman microspectroscopic analysis of oral exfoliated cells, Anal. Methods, 9 (2017), pp. 937-948.

42. R.D. Tobias, An introduction to partial least squares regression, SAS Conf Proc SAS Users Gr Int 20 (SUGI 20), (1995), pp. 2-5. doi:http://support.sas.com/techsup/technote/ts509.pdf.

43. D.R. Parachalil, J. McIntyre, H.J. Byrne, Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: recent advances, Anal Bioanal Chem., 412 (2020), pp. 1993-2007.

44. D.K. Medipally, A. Maguire, J. Bryant, J. Armstrong, M. Dunne, M. Finn, F.M. Lyng, A.D. Meade, Development of a high throuput (HT) Raman spectroscopy method for rapid screening of liquid blood plasma from prostate cancer patients, Analyst, 142 (2017), pp. 1216-1226. doi: 10.1039/c6an02100j.

45. Z. Movasaghi, S. Rehman, I.U. Rehman, Raman Spectroscopy of Biological Tissues, Applied Spectroscopy Reviews, 42 (2007), pp. 493-541.

46. S. Feng, S. Huang, D. Lin, G. Chen, Y. Xu, Y. Li, Z. Huang, J. Pan, R. Chen, H. Zeng, Surface-enhanced Raman spectroscopy of saliva proteins for the noninvasive differentiation of benign and malignant breast tumors, Int J Nanomedicine, 12 (2015), pp. 537-547. doi: 10.2147/IJN.S71811.

47. A. Bennick, Salivary proline-rich proteins, Mol Cell Biochem., 45 (1982), pp. 83-99.

48. E. Ryzhikova, O. Kazakov, L. Halamkova, D. Celmins, P. Malone, E. Molho, E.A. Zimmerman, I.K. Lednev, Raman spectroscopy of blood serum for Alzheimer's disease diagnostics: specificity relative to other types of dementia, J Biophotonics, 7 (2015), pp. 584-596. doi: 10.1002/jbio.201400060.

49. M. Navazesh, Methods for collecting saliva, Annals of the New York Academy of Sciences, 694 (1993), pp. 72–77.

50. A. Kobus, A. Kierklo, A. Zalewska, A. Kuźmiuk, S.D. Szajda, S. Ławicki, J. Bagińska, Unstimulated salivary flow, pH, proteins and oral health in patients with Juvenile Idiopathic Arthritis, BMC Oral Health., 17 (2017), p. 94. doi: 10.1186/s12903-017-0386-1.

51. E. Benedek-Spät, The composition of stimulated human parotid saliva, Arch Oral Biol., 18 (1973), 1091-1097.

52. S.P. Humphrey, R.T. Williamson, A review of saliva: normal composition, flow, and function, J Prosthet Dent., 85 (2001), pp. 162-169.

53. C. Dawes, G.N. Jenkins, Stimulus effects on protein and electrolyte concentrations in parotid saliva, J. Physiol., 170 (1964), pp. 86-100.

54. C.F. Streckfus, W.P. Dubinsky. A Review on Salivary Genomics and Proteomics Biomarkers in Oral Cancer, Expert Rev. Proteomics, 4 (2007), pp. 329-332.

55. C. Dawes, Circadian rhythms in human salivary flow rate and composition, J Physiol., 220 (1972), pp. 529-545.

56. W.M. Edgar, Saliva and dental health. Clinical implications of saliva: report of a consensus meeting, Br Dent J., 169 (1990), pp. 96-98.

57. M.C. Walsh, L. Brennan, J.P. Malthouse, H.M. Roche, M.J. Gibney, Effect of acute dietary standardisation on the urinary, plasma, and salivary metabolomic profiles of healthy humans. Am. J. Clin. Nutr., 84 (2006), pp. 531-539.

58. S. Gomar-Vercher, A. Simón-Soro, J.M. Montiel-Company, J.M. Almerich-Silla, A. Mira, Stimulated and unstimulated saliva samples have significantly different bacterial profiles, PLoS One., 13 (2018), e0198021. doi:10.1371/journal.pone.0198021

59. H. Byrne, G. Sockalingum, N. Stone, Raman Spectroscopy: Complement or Competitor?, In Moss, D. ed. Biomedical Applications of Synchrotron Infrared Microspectroscopy. Royal Society of Chemistry, RCS Analytical Spectroscopy Monographs, 11 (2011), pp.105-142. doi:10.1039/9781849731997-00105

60. J.A. Loo, W. Yan, P. Ramachandran, D.T. Wong, Comparative human salivary and plasma proteomes, J Dent Res., 89 (2010), pp. 1016-1023. doi:10.1177/0022034510380414

61. T. Jensdottir, B. Nauntofte, C. Buchwald, A. Bardow, Effects of sucking acidic candy on whole-mouth saliva composition, Caries Res, 39 (2005), pp. 468-474.

62. M. Yakob, L. Fuentes, M.B. Wang, E. Abemayor, D.T. Wong, Salivary biomarkers for detection of oral squamous cell carcinoma current state and recent advances, Curr Oral Health Rep, 1 (2014), pp. 133-141.

63. C.A. Schafer, J.J. Schafer, M. Yakob, P. Lima, P. Camargo, D.T. Wong, Saliva diagnostics: utilizing oral fluids to determine health status, Monogr Oral Sci, 24 (2014), pp. 88-98.

64. A. Nayak, G.H. Carpenter, A physiological model of tea-induced astringency, Physiol Behav,.95 (2008), pp. 290-294. doi: 10.1016/j.physbeh.2008.05.023.