SCIENTIFIC REP⚙RTS

Sorting of Single Biomolecules based on Fourier Polar Representation of Surface Enhanced Raman Spectra

Aymeric Leray, Thibault Brulé[†], Mickael Buret, Gérard Colas des Francs, Alexandre Bouhelier, Alain Dereux & Eric Finot

Surface enhanced Raman scattering (SERS) spectroscopy becomes increasingly used in biosensors for its capacity to detect and identify single molecules. In practice, a large number of SERS spectra are acquired and reliable ranking methods are thus essential for analysing all these data. Supervised classification strategies, which are the most effective methods, are usually applied but they require pre-determined models or classes. In this work, we propose to sort SERS spectra in unknown groups with an alternative strategy called Fourier polar representation. This non-fitting method based on simple Fourier sine and cosine transforms produces a fast and graphical representation for sorting SERS spectra with quantitative information. The reliability of this method was first investigated theoretically and numerically. Then, its performances were tested on two concrete biological examples: first with single amino-acid molecule (cysteine) and then with a mixture of three distinct odorous molecules. The benefits of this Fourier polar representation were highlighted and compared to the well-established statistical principal component analysis method.

Biosensors are analytical devices measuring the concentration of specific analytes. They are ubiquitous in many fields (e.g. for detecting pesticides[1,2], for analysing food quality[3,4] and environment[5] or for biomedical diagnostics[6–8]). Optical biosensors receive increasing interest due to their high sensitivity and selectivity (for a review, see[9]). This unique selectivity is enabled by spectroscopically analysing inelastic optical responses such as fluorescent emission or Raman scattering. However, due to the extremely small scattering cross section, Raman spectroscopy was for a long time limited to steady state or slow dynamic process[10] in bulk materials or highly concentrated solutions[11,12]. For the last twenty years, the interest in Raman spectroscopy has been renewed when surface enhanced Raman scattering (SERS) has been demonstrated as an effective label-free spectroscopy for identifying and classifying single molecules[13–18]. The SERS effect is characterized by a significant enhancement of the Raman signal emitted by molecules present in close proximity of metallic nanoparticles. When excited by an appropriate laser frequency, localized surface plasmon resonances sustained by the metallic nanoparticles provide a high enhancement of the electric field intensity that can reach several orders of magnitude (at least 10^6).

In addition to their single molecule sensitivity, SERS sensors, by nature, provide a highly specific spectroscopic response also at the individual molecules level[13,19]. In practice, the combination of a high sensitivity with a high specificity result in the acquisition of a large number of distinct spectra. These numerous spectra may reflect the fact that either a large number of distinct molecules are present in the medium or that the analyte can be described by numerous states such as distinct conformations or positions in the SERS active region. In all cases, a reliable classification of these numerous SERS spectra is essential for accurately interpreting the output of the sensor.

The most sophisticated and effective methods for classifying SERS spectra are based on a predetermined library of spectra. They are called supervised methods and include linear discriminant analysis (LDA)[20,21] and partial least squares discriminant analysis (PLS-DA)[22,23]. However, building a valid classification model requires

Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS, Université Bourgogne Franche-Comté, 21000 Dijon (France). [†]Present address: Département de Chimie, Université de Montréal, Montréal, QC, H3C 3J7 (Canada). Correspondence and requests for materials should be addressed to A.L. (email: aymeric.leray@u-bourgogne.fr)

the acquisition of a training set of samples. Furthermore, these classification models are only valid for retrieving a known molecule with a given SERS based sensor and might not be adaptable to other platforms.

An alternative sorting tool exists to group SERS spectra by similarities into unknown classes without the need of prerequisite models. This multivariate analysis called principal component analysis (PCA) is used to reduce the dimensionality of measured SERS spectra into few principal components. The advantage of this strategy is that no *a priori* knowledge is required for data analysis. PCA has thus been largely used to regroup similar SERS spectra into clusters[24–28]. However, this method is limited to disentangle relatively simple mixture because it is graphically bound to three principal components.

In this work, we propose an original method that overcomes the limitations of PCA. Our representation sorts molecular spectra in a fast and visual way and provides mixture proportion without the need of precalibration step or reference spectra. Briefly, our approach is to reduce SERS spectra into a Fourier polar representation. Each spectrum is then transformed to a pole whose coordinates (modulus m and phase φ) correspond to the first component of its Fourier transform. Therefore, all SERS spectra are visualized into a scatter diagram. The geometry of the pool of poles gives an indication on the classification (e.g. binary or ternary mixture).

This strategy was initially applied in fluorescence lifetime imaging microscopy[29–31] and was recently used for unmixing fluorescence spectral images[32] and for spatially segmenting hyperspectral stimulated Raman scattering images[33]. However, to the best of our knowledge, it has never been applied to SERS experiments. Due to the large enhancement factor allowing single molecule measurements, SERS spectra are usually composed of multiple peaks with highly temporally fluctuating intensity. In this work, we have considered these specific issues of SERS experiments by investigating for instance the influence of the enhancement factor on the Fourier polar representation of multiple peaks spectra.

The paper is organized as followed. We first describe the principle of the Fourier polar representation and discuss the meaning of poles for single and multiple peaks spectra. We then apply it to small biomolecules, which are more difficult to be detected. We first consider one amino acid, the cysteine, whose levels are raised in Alzheimer's and Parkinson's diseases[34]. Experimental SERS spectra of single cysteine molecule are analysed with Fourier polar representation. We also apply it to a ternary mixture of distinct small odorous molecules: acetoin (butter odour), eugenol (clove odour) and IBMP (green pepper odour) which have been largely used for studying biological mechanisms involved in the perception of smell or taste and flavour[35]. We finally discuss the advantages and limitations of this representation compared to the principal component analysis.

Theoretical Framework

Case of single peak SERS spectrum. A phasor (contraction of phase vector) is a complex number representation written as a complex exponential $me^{i\varphi}$ where m is the vector modulus and φ is the phase. Each SERS spectrum $I(\lambda)$ is converted to a phasor by using a Fourier transform. The $(u_1;v_1)$ coordinates corresponding to the real and imaginary parts of the first component of the Fourier transform are expressed as

$$u_1 = \frac{\int_{\lambda_s}^{\lambda_f} I(\lambda) \times \cos(k\lambda)\, d\lambda}{\int_{\lambda_s}^{\lambda_f} I(\lambda)\, d\lambda} \tag{1}$$

$$v_1 = \frac{\int_{\lambda_s}^{\lambda_f} I(\lambda) \times \sin(k\lambda)\, d\lambda}{\int_{\lambda_s}^{\lambda_f} I(\lambda)\, d\lambda} \tag{2}$$

where λ_s and λ_f are respectively the start and final wavelengths of the spectral range and k is the angular wave number defined as $k = 2\pi/(\lambda_f - \lambda_s)$.

The polar coordinates corresponding to modulus m and phase φ are obtained from the $(u_1;v_1)$ coordinates with

$$\varphi = \tan^{-1}(v_1/u_1) \tag{3}$$

$$m = \sqrt{u_1^2 + v_1^2} \tag{4}$$

Take for instance a SERS spectrum composed of a single peak at wavelength λ_0 approximated by a Dirac delta function $I(\lambda) = I_0 \times \delta(\lambda - \lambda_0)$. The coordinates are simply given by

$$u_1 = \cos(k\lambda_0) \text{ and } m = 1 \tag{5}$$

$$v_1 = \sin(k\lambda_0) \text{ and } \varphi = k\lambda_0 \tag{6}$$

The equality $u_1^2 + v_1^2 = 1$ corresponds to the equation of a circle of radius 1 centred at the origin [0, 0]. In other words, an ideal single band SERS spectrum is represented by a pole localized on a circle centred at the origin with a radius of 1 (cf. Fig. 1A). Only φ provides information about the resonant wavelength of the band.

If we consider more realistic SERS spectra where Raman bands are modelled with Lorentzian functions (with full-width at half-maximum FWHM of Γ) given by: $I(\lambda) = \frac{I_0}{2\pi} \frac{\Gamma}{(\lambda - \lambda_0)^2 + (\Gamma/2)^2}$, the coordinates for an infinite spectral measurement width are given by
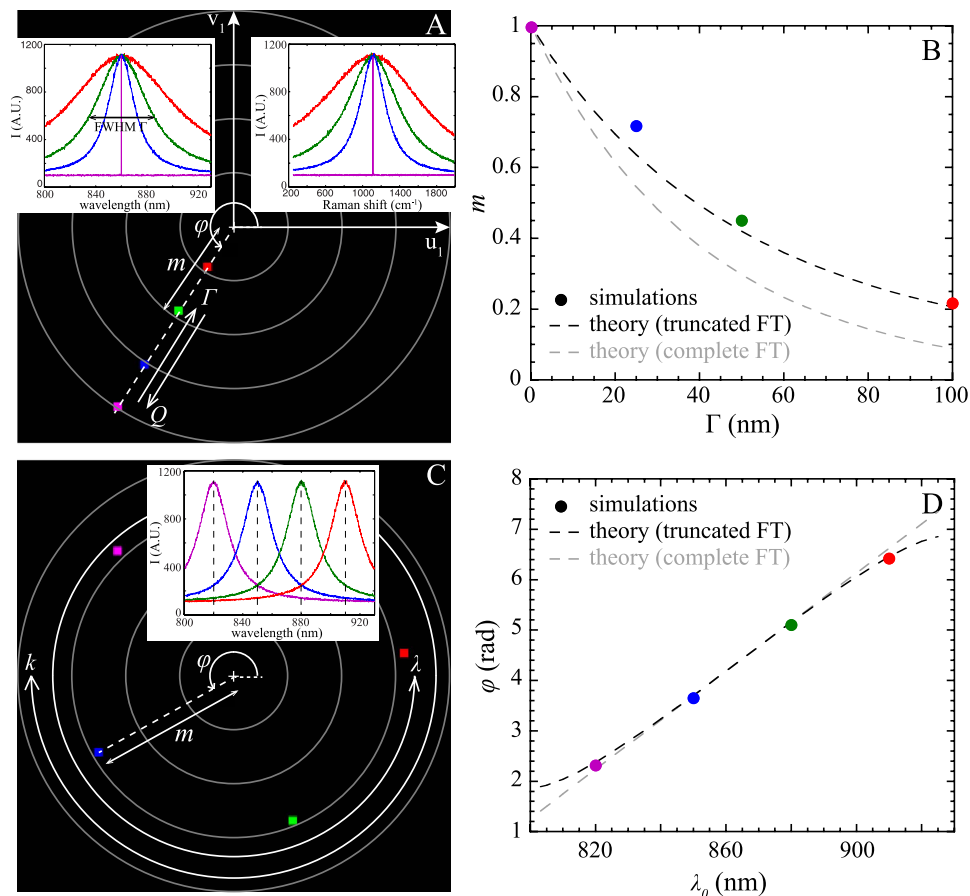
Figure 1. (**A**) Fourier polar representation of 4 single Raman band centred at 860 nm with increasing the full-width at half-maximum Γ of the Raman bands. (**B**) Corresponding modulus m as a function of Γ. When increasing Γ, poles come closer to origin. Theoretical modulus m is plotted for an infinite (in grey dotted line) and finite wavelength range (in black dotted line). (**C**) Fourier polar representations for 4 single Raman band of the same Γ but shifted with different λ_0. Poles turn anti-clockwise when increasing λ and clockwise when increasing k. Corresponding phase shifts for these peaks are represented in (**D**). Theoretical phase φ is plotted for an infinite (in grey dotted line) and finite wavelength range (in black dotted line).

$$u_1 = \exp(-\Gamma|k|/2)\cos(k\lambda_0) \text{ and } m = \exp(-\Gamma|k|/2) \qquad (7)$$

$$v_1 = \exp(-\Gamma|k|/2)\sin(k\lambda_0) \text{ and } \varphi = k\lambda_0 \qquad (8)$$

A Raman band (with a finite quality factor) is then represented by a pole located inside the unit circle and its modulus ($m < 1$) decreases exponentially when Γ increases (as indicated in Fig. 1A). Realistic Raman peaks with $\Gamma = 1$ nm leads to a modulus $m > 0.97$.

Equations 7 and 8 (which are represented in grey in Fig. 1B,C) are only fully valid for an infinite spectral measurement width. Small differences between the theoretical expressions and our Monte Carlo simulations (black dots) are explained by the finite wavelength range $[\lambda_s; \lambda_f]$ used to simulate the SERS spectra. In this case, the analytical calculations of the polar coordinates are not obvious. Results shown by the black dotted lines in Fig. 1B,D are obtained from computational study for SERS spectra modelled by Lorentzian functions with a finite spectral measurement width. The numerical results, performed with Mathematica 10 (Wolfram Research, Inc), are now in excellent agreement with Monte Carlo simulations.

Case of multiple peaks SERS spectrum. A full Raman spectrum composed of N peaks is ideally modelled by $I(\lambda) = I_0\sum_{i=1}^{N}\delta(\lambda - \lambda_i)$. The $[u_1; v_1]$ coordinates are
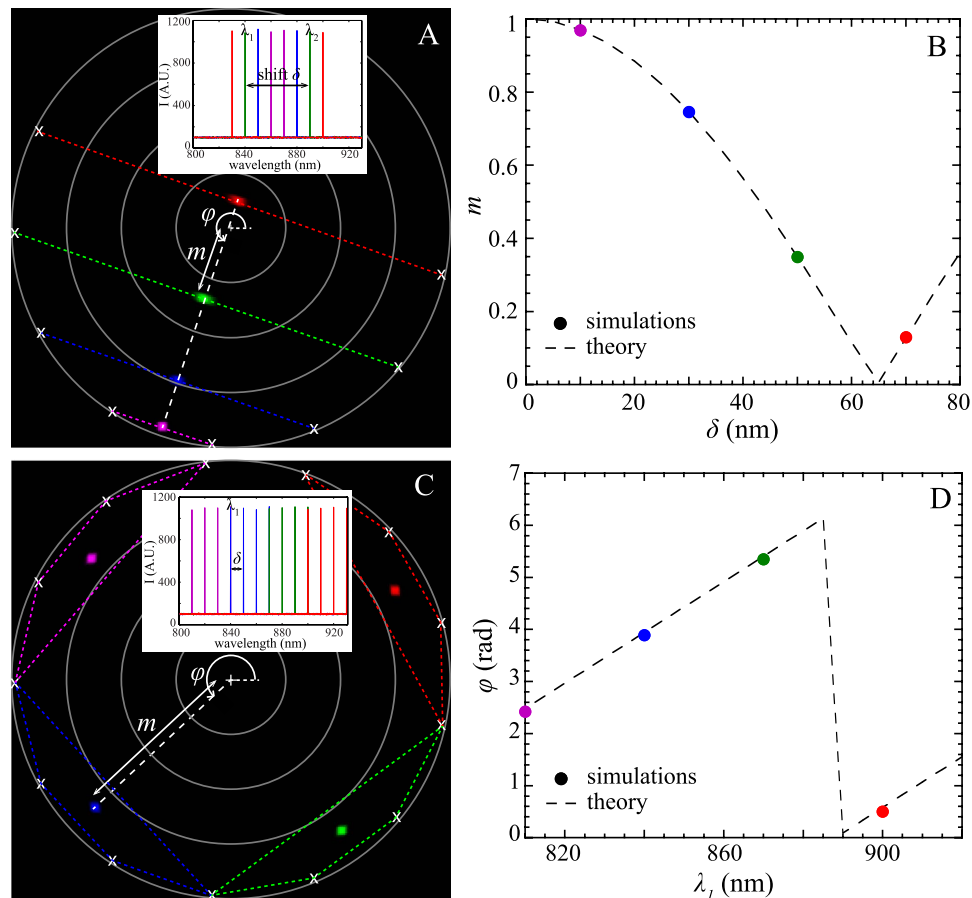
$$u_1 = \sum_{i=1}^{N}\frac{\cos(k\lambda_i)}{N} \qquad (9)$$

Figure 2. (**A**) Fourier polar representation for 4 double peaks spectra with variable spectral gap δ. (**B**) Simulated (coloured points) and theoretical (dotted lines) modulus m are plotted as a function of δ. (**C**) Fourier polar representation of four-peaks combs spectra separated with the same δ but shifted in wavelength. Corresponding φ as a function of the starting wavelength of the peak comb are compared in (**D**) with theoretical expectations.

$$v_1 = \sum_{i=1}^{N} \frac{\sin(k\lambda_i)}{N} \tag{10}$$

These coordinates correspond thus simply to the linear combination of the $[u_1;v_1]$ values of all individual SERS peaks. Calculations of the modulus m and the phase φ are easily feasible analytically for two peaks of identical height with wavelengths λ_1 and λ_2 ($\lambda_2 > \lambda_1$)

$$m = \left| \cos(k(\lambda_2 - \lambda_1)/2) \right| \text{ and } \varphi = k(\lambda_1 + \lambda_2)/2 \tag{11}$$

In this case, the Fourier polar representation is localized on the middle of the segment formed by the coordinates of each individual peaks (Fig. 2A). These theoretical expressions are in excellent agreement with simulations (cf. Fig. 2B). When $N \geq 3$, an analytical calculation of m and φ is not trivial and numerical computation is required.

However, some simple cases can still be detemined analytically. For instance, if we consider a comb of N Raman peaks of identical height that are uniformly separated by a fixed spectral shift δ and with first wavelength λ_1, the modulus and the phase shift are given by

$$m = \left| \frac{\sin(Nk\delta/2)}{N \ \sin(k\delta/2)} \right| \text{ and } \varphi = k\left(\lambda_1 + \frac{(N-1)\delta}{2}\right) \tag{12}$$

φ is governed by the central wavelength of the comb whereas the modulus is related to the number of peaks in the comb (at fixed δ, m decreases when N increases).

For $N = 4$, the polar representation is localized inside the unit circle ($m = \left| \sin(2k\delta)/(4 \ \sin(k\delta/2)) \right| < 1$) and more precisely inside the polygon formed by $[u_1;v_1]$ values of ideal individual single peak SERS spectra (Fig. 2C). The phase shift $\varphi = k(\lambda_1 + 3\delta/2)$ matches satisfactorily the simulated results (Fig. 2D).
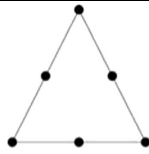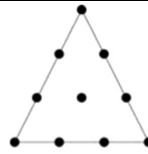
Proportions (p_1; p_2) in binary mixture	Proportions (p_1; p_2; p_3) in ternary mixture				
1 molecule	2 molecules	3 molecules	1 molecule	2 molecules	3 molecules
0; 1	0; 1	0; 1	0; 0; 1	0; 0; 1	0; 0; 1
		⅓; ⅔			0; ⅓; ⅔
	½; ½			0; ½; ½	0; ⅔; ⅓
		⅔; ⅓			⅓; ⅓; ⅓
1; 0	1; 0	1; 0	0; 1; 0	0; 1; 0	0; 1; 0
				½; 0; ½	⅔; 0; ⅓
					⅓; 0; ⅔
				½; ½; 0	⅓; ⅔; 0
					⅔; ⅓; 0
			1; 0; 0	1; 0; 0	1; 0; 0

Table 1. List of poles in a binary and ternary mixture of molecular states. The proportions p_1, p_2 and p_3 of each vibrational state (1, 2, 3) are reported according to the number of molecules simultaneously present in the SERS active volume that will participate to the Raman spectrum (by assuming similar SERS intensity spectrum for each state).

Case of a mixture.

A SERS spectrum does not necessarily stem from only one single molecule or even one molecular state. For instance, several molecular states of the same molecule may be activated simultaneously in the hotspot. The corresponding acquired spectrum will be a combinatory assembly of the SERS fingerprints of each molecular state.

To illustrate this situation, we first consider a binary mixture of two molecular states with proportions p_1 and p_2. If only one molecule is present in the enhancement site, the SERS measurements will only give two distinct spectra ($p_1 = 1$ or $p_2 = 1$) and the corresponding polar representation is a segment with two poles (Table 1). When two (and respectively three) states are mixed in SERS measurements, the corresponding polar representation is now a segment with 3 (and respectively 4) poles (whose proportions are listed in Table 1).

A ternary mixture of three states leads to a more complex combinatory analysis (Table 1). The 3 principal poles, corresponding to the isolated pure state, form a triangle. The number of poles can be 3, 6 and 10 according to the number of possible molecules (1, 2 and 3) detectable in the SERS active volume at the same time. The pool of poles in a ternary mixture has the shape of a triangle. The advantage of the Fourier polar representation becomes obvious. The polar plot graphically identifies the number of components of the mixture as well as their proportions.

In the next section, these theoretical considerations will be validated by simulations and experiments using both a binary and a ternary mixture.

Results

Binary mixture: Monte Carlo simulations.

For a deeper analysis of a binary mixture, we performed Monte Carlo simulations (see Material and Methods). We consider two molecular states with characteristic Raman spectra reported in blue and red in Fig. 3A. We first study high signal to noise ratio SERS spectra ($SNR > 30$). As previously mentioned, when two or three molecules are passing through the SERS active volume, the Fourier polar representation depicts a line with 3 or 4 poles (Fig. 3A). The poles at both ends of the segment correspond to the SERS spectra of individual molecular state (not mixed) and the internal poles (in green) result from the linear combination of each principal state. Note that the length of the segments corresponds directly to the mix ratio.

The Fourier polar representation was also compared with the well-established PCA method. As a reminder, PCA reduces the dimensionality of the measured SERS spectra by converting a set of correlated data into linearly orthogonal variables called principal components (PCs). PCA applied to the previous Monte Carlo simulations shows that one PC is sufficient to describe the SERS spectra, and three color-coded families (blue, red and green) are successfully retrieved. However, there is no longer a direct relation between the proportion and the position of the poles.

Let us then consider more realistic SERS spectra with lower signal to noise ratio ($SNR = 10$). The Fourier polar representation still exhibits 3 aligned poles (Fig. 3C) but these poles are widened with an oblong shape. The green pole is no longer exactly positioned in the midpoint.

With the PCA, two PCs are now necessary for interpreting the SERS spectra that are still classified into 3 families (see Fig. 3D1,D2). The blue and red families correspond to the pure spectra ($p_2 = 0$ and $p_2 = 1$). They form an angle of almost 90° indicating that they are independent. Again, there is no direct relation between the proportion and the position of the poles.
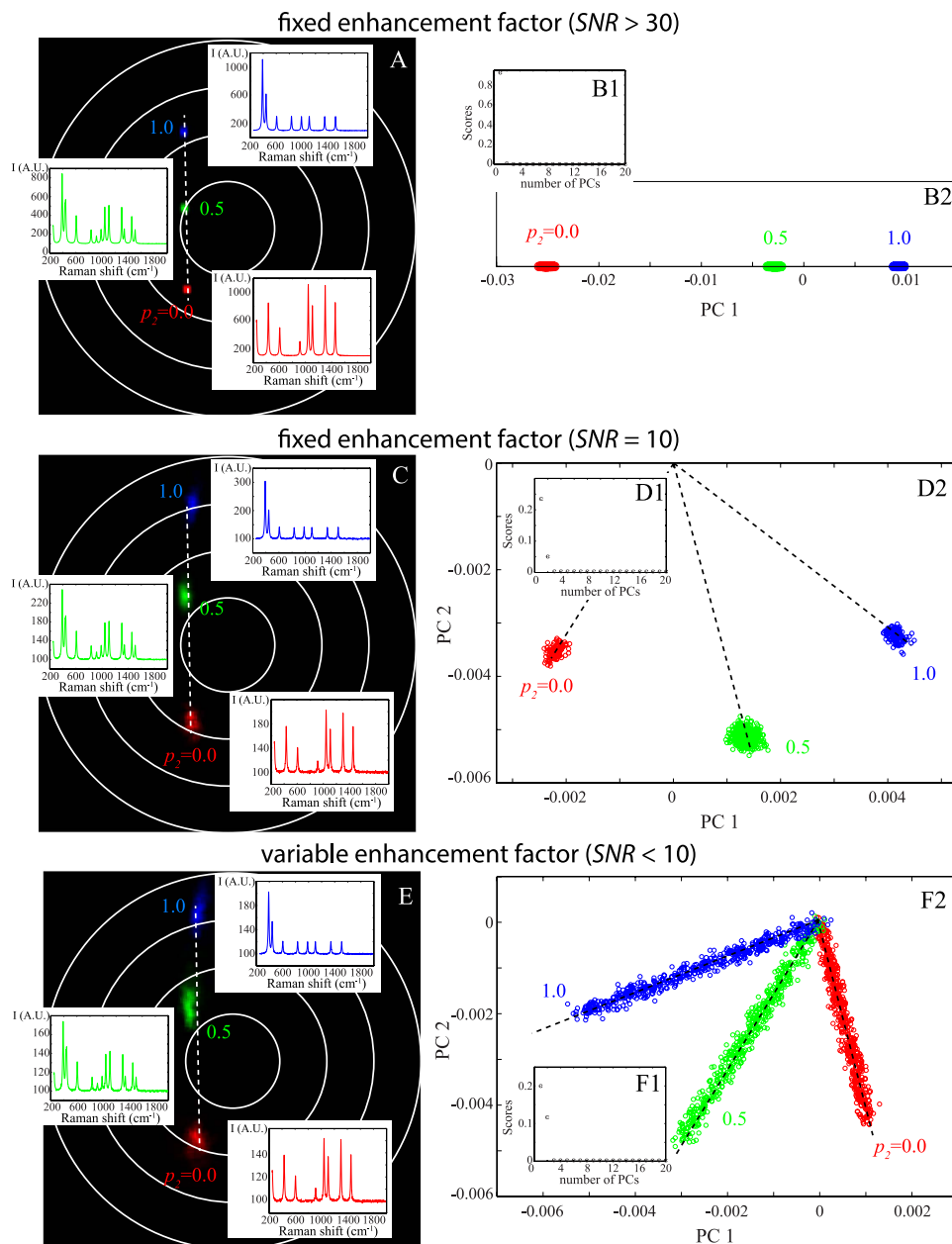
Figure 3. Comparison between the Fourier polar representation (left) and the PCA (right) applied to simulated spectra constituted of a binary mixture with varying mix ratio ($p_2 = 0$; 1/2 and 1). The influence of the signal to noise ratio (*SNR*) (at the top) and the enhancement factor *G* (at the bottom) are illustrated. At high *SNR* and fixed *G*, the corresponding polar representation and PCA are indicated respectively in (**A,B2**). For low SNR and fixed *G*, polar representation and PCA are indicated in (**C,D2**). When the enhancement factor is computationally modified, the Fourier polar representation (**E**) is almost not affected whereas PCA (**F2**) exhibits branches. The scores of PCs are indicated for each case in B1, D1 and F1.

In practice, molecules are extremely mobile inside the enhanced site implying that the total intensity of the SERS spectra is temporally fluctuating[36] and is extremely dependent on the position of the molecule in the SERS active volume[37,13]. In order to take these fluctuations into account, we add a variable enhancement factor in our Monte Carlo simulations. In presence of fluctuations, the polar representation remains unaffected while the results of the PCA are modified (Fig. 3E,F2). The previous 3 poles are now declined into branches.

Binary mixture: experimental application on cysteine. To validate our previous results deduced from Monte Carlo simulations, we apply the same methodology to experimental SERS spectra acquired when cysteine is circulated in a microfluidic sensor. We demonstrated in a previous work that cysteine is likely to form a dimer (i.e. cystine) and thus can be treated as a binary mixture[13]. The corresponding polar representation reported in Fig. 4A emphasizes three main families colorized in blue, red and green corresponding to three distinct spectra.
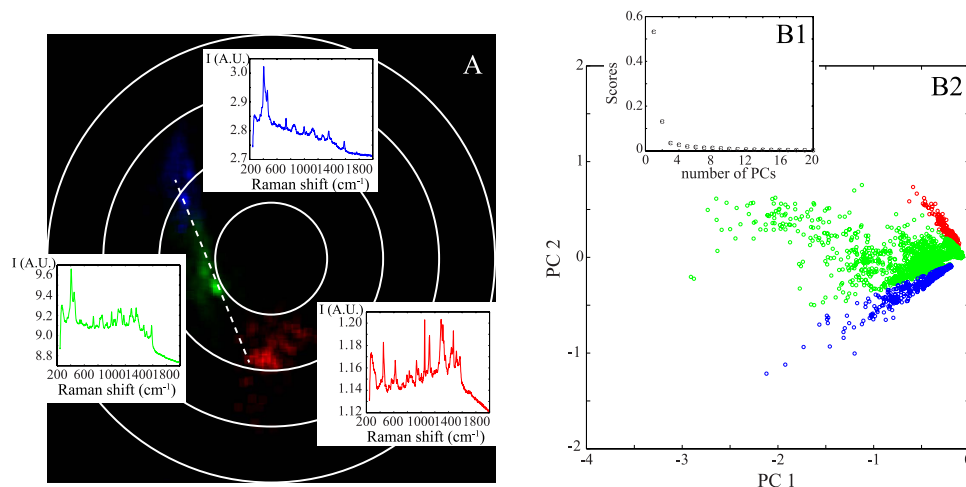
Figure 4. Comparison between Fourier polar representation (A) and PCA (B2) for analysing the experimental spectra of cysteine. Both PCA and polar representation are effective at sorting the spectra in three families coloured in blue, red and green. The corresponding Raman spectra are also indicated. As shown in (**B1**), two PCs are used for characterizing the experimental data.
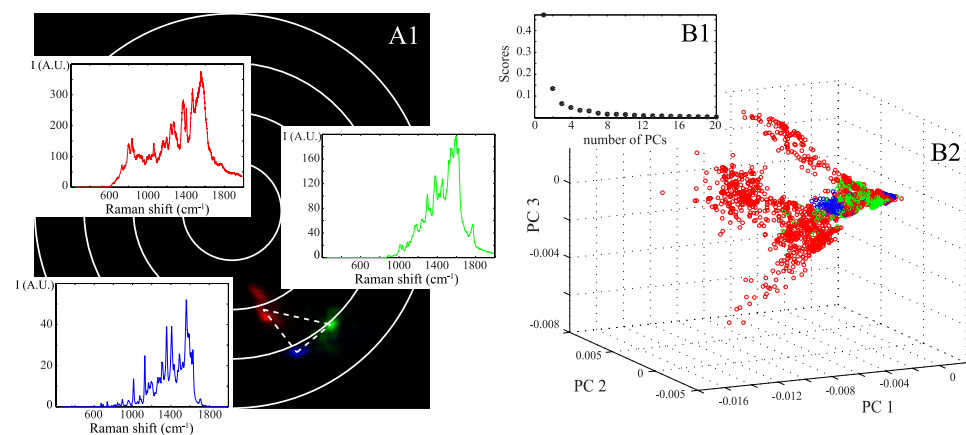

Figure 5. Comparison between the Fourier polar representation (A1) and the PCA (B2) applied on the experimental series of a mixture of three distinct aroma molecules: acetoin (indicated in blue), eugenol (in green) and IBMP (in red). The corresponding spectra are also shown in (**A1**). With PCA, 3 PCs are necessary for characterizing the experimental SERS spectra (see **B1**) but it is not possible to sort the three distinct molecules.

These families are almost localized on a line, and the extremities are formed from the blue and red families. This representation indicates that the red and blue families are nearly independent and that the green family is composed of a linear combination of the red and blue families (as suggested from Monte Carlo simulations).

The PCA applied to the experimental SERS data leads to the same conclusion. As indicated in Fig. 4B1, two principal components are necessary for describing the data. The results of PCA are then represented in two dimensions and three families are isolated (Fig. 4B2). We identify two nearly orthogonal distinct families colorized in red and blue. The third green family is obtained from a linear mixture of these two principal red and blue spectra. The blue and red families have already been attributed to the SERS spectra of the cysteine and of its dimer, respectively[13].

To conclude this section, when PCA can be reduced to a limited set of PCs, as in the case of single cysteine molecule, Fourier polar representation and PCA lead to similar sorting accuracy. We show in the following section the advantages of the polar representation when complex systems are investigated.

Ternary mixture: experimental application on three different odorous molecules.

We finally experimentally study a ternary mixture composed of three distinct aroma molecules: IBMP (green pepper odour), acetoin (butter odour) and eugenol (clove door). We successively acquire 512 SERS spectra of each individual molecules circulating into a microfluidic SERS sensor. As indicated in Fig. 5B1, at least three PCs are necessary to perform principal component analysis. The corresponding three-dimensional PCA are represented in Fig. 5B2.

Here, PCA is not able to distinguish the three odorous molecules. It seems that three principal components are not enough for sorting the ternary molecular mixture.

With the polar representation, it is not true anymore. As shown in Fig. 5A1, the Fourier polar representation forms a triangle with three poles, as expected theoretically for a ternary mixture with one molecule in the enhancement site (see Table 1). Because the polar coordinates are sensitive to both the number and the quality factor of the Raman bands, each type of molecule is well separated in the polar representation and can now be easily identified. The polar representation can thus be successfully used for sorting a ternary mixture of odorous molecules.

Discussion

To determine the relevance of PCA and polar representation for sorting SERS spectra, we first clarify the advantages and drawbacks of each method.

PCA and polar representation are two powerful sorting methods, which do not require any *a priori* knowledge about the measured samples.

It is well known that the performance of any classification methods largely depends on the mathematical pre-treatment of spectra[20]. In this work, we have voluntarily restricted this pre-treatment to simple offset subtraction to be as close as possible to raw data. We could of course envision more complicated corrections for denoising and/or flattening SERS spectra that would surely modify our results. However, these corrections will equivalently modify the PCA and the polar representation and would not alter the comparison between these two methods.

We have shown that PCA is a robust approach for separating SERS spectra even if *SNR* is low. We previously applied it for discerning the different forms of cysteine[13]. However, the drawbacks of PCA include: (i) the lack of quantitative determination of the proportion in a binary mixture, (ii) the lack of reproducibility in the orientation of the branches preventing a comparison between distinct sets of experimental spectra and (iii) the failure to sort a ternary mixture. Finally, the PCA method is probably not well adapted for classifying SERS spectra of complex samples. When the number of principal components exceeds 3, the graphical representation of the PCA outcomes is difficult.

These drawbacks are largely overcome with the suggested polar representation. The phasor correctly classifies binary and ternary mixture even at low *SNR*. The polar representation provides quantitative information on the unknown mixture ratio. It is a 2D histogram displaying the real and the imaginary parts of the first component of the Fourier transform and it is thus applicable to any molecular SERS signatures (i.e. large and small proteins). Additionally, we have demonstrated that the complex modulus and the phase shift (m and φ) are relatively constant regardless of *SNR* for a given spectrum. The polar representation can thus be used to compare distinct sets of experimental data. We have shown that the phasor depends on both the number and the quality factor of peaks in the SERS spectra. Another advantage of the phasor is that the representation can be directly implemented on a standard computer enabling thus a classification of SERS spectra with an automatic identification of identical spectra during an experimental acquisition. We emphasize that this automation would not have been possible with other analysis methods including unsupervised grouping techniques such as clustering[38].

Nevertheless, the Fourier polar representation is not free from weaknesses. Its major drawback is its extreme dependence on the offset subtraction that affects the first component of the Fourier sine and cosine transforms. It is therefore important to suppress fundamental frequency corresponding to the DC component (or offset). Otherwise, the calculated polar coordinates will be dependent on the level of this offset and the comparison between different sets of data will be impossible.

Here, we only calculated the polar coordinates of SERS spectra using the first component of the Fourier transform. In practice, our experimental SERS spectra are rich, and could thus be defined by a larger number of wavenumbers. We could then calculate the polar coordinates of SERS spectra with more components and plot several polar representations for each wavenumber[39]. It is worth noting that each polar representation will be different; in other words, the localization of the poles will be modified. The results indicated in Figs 1 and 2 would not be valid anymore as they are specific to the first component. An optimal wavenumber probably exists for maximizing the separation distance between spots. However, the precision of the polar representation is dependent on the amplitude of the Fourier components. This study was restricted to the first component because its amplitude is maximal, implying that the standard deviation is minimal. Furthermore, we experimentally showed that using the first component of Fourier transform was sufficient to correctly sort our data.

Conclusions

In this work, we have introduced an alternative method for accurately and graphically sorting temporally fluctuating single molecule SERS spectra. This non-fitting method called Fourier polar representation does not require any prerequisite models or classes. We have explained its general principle and described theoretically the position of the poles according to both the number and the quality factors of the Raman bands. Graphical patterns obtained with binary and ternary molecular mixtures have been discussed. We have then demonstrated that the Fourier polar representation was equivalent to the well-established PCA method for sorting binary mixture. In addition, our method provides the proportions of the different components in the mixture, which is not possible with PCA. We have finally experimentally demonstrated the added value of our method on a binary mixture and a ternary mixture of different biomolecules. We stress that this Fourier polar representation was here applied for the first time to sort single molecule SERS spectra. In particular, using simple base-line correction of spectra, the Fourier polar representation was able to discriminate a ternary mixture of distinct odorous molecules, whereas this sorting was not obvious with PCA. Because of its ease of implementation, this non-fitting analysis method should be of global concern to the spectroscopic sensor community.

Methods

Monte Carlo simulations. SERS spectra with controlled parameters are simulated on a standard computer using a Monte Carlo approach. In brief, the number of Raman bands and their corresponding wavelengths are fixed. Then we used a random number generator to set the spectral intensity. The density probability function of the generator is a Poisson distribution with a mean corresponding to spectral intensity. In order to mimic as closely as possible to standard experimental conditions, we consider a spectral band of 130 nm starting at 800.65 nm and divided into 1024 pixels leading to a spectral resolution of 0.127 nm. The total intensity of SERS spectra I_0 is Poisson distributed and thus signal to noise ratio (SNR) is shot-noise limited. We also added an offset (which is also Poisson distributed) and we have applied a variable enhancement factor (uniformly distributed between 0 and G) to each spectrum to take into account the mobility of the molecules into the active SERS area.

Acquisition of SERS spectra. In order to measure exploitable SERS spectra, we use a raspberry-like gold nanoparticles template synthetized from a colloidal solution. The nanoparticles are deposited on a glass coverslip and assembled in a microfluidic cell (fabrication steps are described in[13]). Cysteine was diluted into PBS to obtain a concentration of 50 μg/mL. The aroma molecules: 3-isobutyl-2-methoxypyrazine (IBMP), 3-hydroxybutanone (acetoin) and eugenol were diluted into deionized water to a concentration of 10 μM. These different solutions are circulated into a microfluidic channel using a peristaltic pump generating a flow rate of 2 μL/min.

The microfluidic device is placed on a custom-built inverted confocal Raman microscope (scanned with a piezoelectric translation stage) coupled with a 785 nm laser diode. Back scattered spectra are acquired with a 60× water immersion objective (NA = 1.2, Nikon) at a frequency of 1 Hz by using a spectrometer composed of a 600 lines per millimeter grating associated with a cooled CCD camera (1024 × 1024 pixels).

Data Analysis. SERS spectra are analysed by using two distinct methods: the PCA and the polar representation.

PCA is a well-established multivariate analysis technique converting the high dimensionality of the experimental data into a new variance weighted coordinate system with lower dimensions. The orthogonal linear transformation consists in determining the eigenvectors of the covariance matrix; the so-called principal components (PCs). In practice, each spectrum is first centred on its mean, and then the covariance matrix of the complete data set is calculated. The PCs and scores are resolved by using a homemade program designed with MATLAB.

The Fourier polar representation is programmed using MATLAB software. Fourier sine and cosine transforms are calculated for each spectrum and the intensity-weighted polar representation is displayed. Prior to each analysis (PCA and polar representation), SERS spectra are background corrected by estimating an average background from a part of the spectrum where the Raman signal is low and subtracting it from the full spectrum.

References

1. Li, P. *et al.* Polystyrene/Ag nanoparticles as dynamic surface-enhanced Raman spectroscopy substrates for sensitive detection of organophosphorus pesticides. *Talanta* **127,** 269–275 (2014).
2. Zheng, J., Pang, S., Labuza, T. P. & He, L. Semi-quantification of surface-enhanced Raman scattering using a handheld Raman spectrometer: a feasibility study. *Analyst* **138,** 7075–7078 (2013).
3. Craig, A. P., Franca, A. S. & Irudayaraj, J. Surface-Enhanced Raman Spectroscopy Applied to Food Safety. *Annu. Rev. Food Sci. Technol.* **4,** 369–380 (2013).
4. Sivashanmugan, K., Liao, J.-D., Liu, B. H. & Yao, C.-K. Focused-ion-beam-fabricated Au nanorods coupled with Ag nanoparticles used as surface-enhanced Raman scattering-active substrate for analyzing trace melamine constituents in solution. *Anal. Chim. Acta* **800,** 56–64 (2013).
5. Schmidt, H. *et al.* Detection of PAHs in seawater using surface-enhanced Raman scattering (SERS). *Mar. Pollut. Bull.* **49,** 229–234 (2004).
6. Lin, D. *et al.* Label-free blood plasma test based on surface-enhanced Raman scattering for tumor stages detection in nasopharyngeal cancer. *Sci. Rep.* **4,** 4751 (2014).
7. Lee, S. *et al.* Rapid and sensitive phenotypic marker detection on breast cancer cells using surface-enhanced Raman scattering (SERS) imaging. *Biosens. Bioelectron.* **51,** 238–243 (2014).
8. Mohs, A. M. *et al.* Hand-held spectroscopic device for *in vivo* and intraoperative tumor detection: contrast enhancement, detection sensitivity, and tissue penetration. *Anal. Chem.* **82,** 9058–9065 (2010).
9. Fan, X. *et al.* Sensitive optical biosensors for unlabeled targets: A review. *Anal. Chim. Acta* **620,** 8–26 (2008).
10. Wachsmann-Hogiu, S., Weeks, T. & Huser, T. Chemical analysis *in vivo* and *in vitro* by Raman spectroscopy—from single cells to humans. *Curr. Opin. Biotechnol.* **20,** 63–73 (2009).
11. Raman, C. V. & Krishnan, K. S. A New Type of Secondary Radiation. *Nature* **121,** 501–502 (1928).
12. Matthäus, C. *et al.* In *Methods in Cell Biology* (ed. Dr. John J. Correia & Dr. H. William Detrich, I.) 89, 275–308 (Academic Press, 2008).
13. Brulé, T. *et al.* Sorting of Enhanced Reference Raman Spectra of a Single Amino Acid Molecule. *J. Phys. Chem. C* **118,** 17975–17982 (2014).
14. Etchegoin, P. G., Meyer, M., Blackie, E. & Le Ru, E. C. Statistics of single-molecule surface enhanced Raman scattering signals: fluctuation analysis with multiple analyte techniques. *Anal. Chem.* **79,** 8411–8415 (2007).
15. Kneipp, K. *et al.* Single molecule detection using surface-enhanced raman scattering (SERS). *Phys Rev Lett* **78,** 1667–1670 (1997).
16. Qian, X.-M. & Nie, S. M. Single-molecule and single-nanoparticle SERS: from fundamental mechanisms to biomedical applications. *Chem Soc Rev* **37,** 912–920 (2008).
17. Nie, S. & Emory, S. R. Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering. *Science* **275,** 1102–6 (1997).
18. Michaels, A. M., Nirmal, M. & Brus, L. E. Surface enhanced Raman spectroscopy of individual rhodamine 6G molecules on large Ag nanocrystals. *J. Am. Chem. Soc.* **121,** 9932–9939 (1999).
19. Kneipp, K. *et al.* Detection and identification of a single DNA base molecule using surface-enhanced Raman scattering (SERS). *Phys. Rev. E* **57,** R6281–R6284 (1998).
20. Lee, K.-M., Herrman, T. J., Bisrat, Y. & Murray, S. C. Feasibility of Surface-Enhanced Raman Spectroscopy for Rapid Detection of Aflatoxins in Maize. *J. Agric. Food Chem.* **62,** 4466–4474 (2014).
21. Mert, S. & Çulha, M. Surface-enhanced Raman scattering-based detection of cancerous renal cells. *Appl. Spectrosc.* **68,** 617–624 (2014).

22. Campos, N. da S., Oliveira, K. S., Almeida, M. R., Stephani, R. & de Oliveira, L. F. C. Classification of Frankfurters by FT-Raman Spectroscopy and Chemometric Methods. *Molecules* **19,** 18980–18992 (2014).
23. Driskell, J. D. *et al.* Rapid microRNA (miRNA) detection and classification via surface-enhanced Raman spectroscopy (SERS). *Biosens. Bioelectron.* **24,** 917–922 (2008).
24. Chen, J., Huang, Y.-W. & Zhao, Y. Characterization of polycyclic aromatic hydrocarbons using Raman and surface-enhanced Raman spectroscopy. *J. Raman Spectrosc.* **46,** 64–69 (2015).
25. Avci, E. & Culha, M. Influence of Protein Size on Surface-Enhanced Raman Scattering (SERS) Spectra in Binary Protein Mixtures. *Appl. Spectrosc.* **68,** 890–899 (2014).
26. Jung, G. B., Nam, S. W., Choi, S., Lee, G.-J. & Park, H.-K. Evaluation of antibiotic effects on Pseudomonas aeruginosa biofilm using Raman spectroscopy and multivariate analysis. *Biomed. Opt. Express* **5,** 3238–3251 (2014).
27. Das, G. *et al.* Principal component analysis based methodology to distinguish protein SERS spectra. *J. Mol. Struct.* **993,** 500–505 (2011).
28. Guo, L. *et al.* Multivariate Statistical Analysis of Serum from Breast Cancer Patients Using Surface Enhanced Raman Spectrum. *Spectrosc. Spectr. Anal.* **33,** 1553–1556 (2013).
29. Jameson, D. M., Gratton, E. & Hall, R. D. The Measurement and Analysis of Heterogeneous Emissions by Multifrequency Phase and Modulation Fluorometry. *Appl Spectrosc Rev* **20,** 55–106 (1984).
30. Digman, M. A., Caiolfa, V. R., Zamai, M. & Gratton, E. The phasor approach to fluorescence lifetime imaging analysis. *Biophys J* **94,** L14–6 (2008).
31. Leray, A., Spriet, C., Trinel, D., Usson, Y. & Héliot, L. Generalization of the Polar Representation in Time Domain Fluorescence Lifetime Imaging Microscopy for Biological Applications: practical implementation. *J Microsc* **248,** 66–76 (2012).
32. Fereidouni, F., Bader, A. N. & Gerritsen, H. C. Spectral phasor analysis allows rapid and reliable unmixing of fluorescence microscopy spectral images. *Opt Express* **20,** 12729–12741 (2012).
33. Fu, D. & Xie, X. S. Reliable Cell Segmentation Based on Spectral Phasor Analysis of Hyperspectral Stimulated Raman Scattering Imaging Data. *Anal. Chem.* **86,** 4115–4119 (2014).
34. Heafield, M. T. *et al.* Plasma cysteine and sulphate levels in patients with motor neurone, Parkinson's and Alzheimer's disease. *Neurosci. Lett.* **110,** 216–220 (1990).
35. Golebiowski, J., Topin, J., Charlier, L. & Briand, L. Interaction between odorants and proteins involved in the perception of smell: the case of odorant-binding proteins probed by molecular modelling and biophysical data. *Flavour Fragr. J.* **27,** 445–453 (2012).
36. Margueritat, J. *et al.* Discerning the Origins of the Amplitude Fluctuations in Dynamic Raman Nanospectroscopy. *J. Phys. Chem. C* **116,** 26919–26923 (2012).
37. Pavel, I. *et al.* Label-Free SERS Detection of Small Proteins Modified to Act as Bifunctional Linkers. *J. Phys. Chem. C* **112,** 4880–4883 (2008).
38. Bonifacio, A., Beleites, C. & Sergo, V. Application of R-mode analysis to Raman maps: a different way of looking at vibrational hyperspectral data. *Anal. Bioanal. Chem.* 1–7 (2014). doi: 10.1007/s00216-014-8321-7
39. Cutrale, F., Salih, A. & Gratton, E. Spectral Phasor approach for fingerprinting of photo-activatable fluorescent proteins Dronpa, Kaede and KikGR. *Methods Appl. Fluoresc.* **1,** 35001 (2013).

Acknowledgements

Author Contributions
A.B., A.D., A.L., E.F. and T.B. conceived and designed the experiments. A.L., M.B. and T.B. performed the experiments. A.L., G.C.d.F. and T.B. analysed the data. A.L. and E.F. wrote the manuscript. All authors reviewed the manuscript.

Additional Information
Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Leray, A. *et al.* Sorting of Single Biomolecules based on Fourier Polar Representation of Surface Enhanced Raman Spectra. *Sci. Rep.* **6**, 20383; doi: 10.1038/srep20383 (2016).