Raman spectroscopic studies coupled with MCR-ALS applied on some biomedical systems

（MCR-ALS 法を組み合わせたラマン分光法による医・生物学的応用研究）

Keita Iwasaki

2021

The United Graduate School of Agricultural Sciences,

Tottori University

Raman spectroscopic studies coupled with MCR-ALS applied on some biomedical systems

（MCR-ALS 法を組み合わせたラマン分光法による

医・生物学的応用研究）

THESIS BY

Keita Iwasaki

As Partial Fulfillment of the Requirement

For the Award of the Degree of Doctor of Philosophy

At

Department of Bioresources Science

The United Graduate School of Agricultural Sciences,

Tottori University

September 2021

Contents

Chapter I.

General Introduction

I - i Theoretical Background of Raman spectroscopy

Raman effect is one of light scattering effects which was reported first by C. V. Raman on 1928 (1). When a light with a particular wavelength such as laser light is irradiated on to a substance, several processes can happen (absorption, scattering etc.). Raman scattering is an inelastic scattering process in which the irradiated light either loses or gains energy as a result of interaction with the molecule. Wavelengths of such Raman scattered photons will be shifted from the irradiated light. This shift is related to the energy of electronic, vibrational and rotational states. Among them, the study of Raman effect for vibrational modes is most popular for both ease to measure and interests to recognize structure of targeting molecule based on its molecular vibration. Thus vibrational Raman spectroscopy can be grouped with infrared absorbance spectroscopy (a complementary technique which also probes molecular vibrations!) as vibrational spectroscopy. Figure 1-1 illustrates Raman scattering with other interaction together.

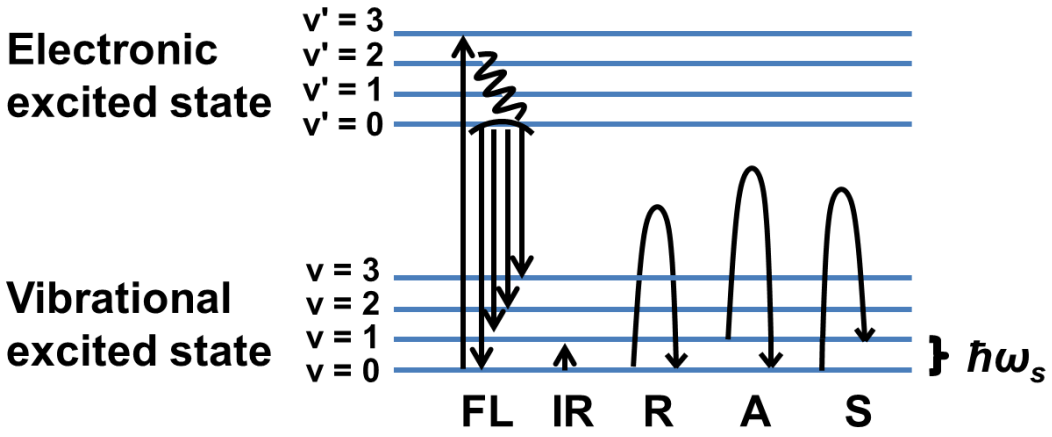


Figure 1-1. Energy level of several interactions of light between substance. Fluorescence (FL), Infrared absorption (IR), Rayleigh scattering (R), anti-Stokes Raman scattering (A) and Stokes Raman scattering (S).

Raman scattering can occur in two manners, the Stokes and anti-Stokes scattering. The Energy of light and vibrational excitation states in a molecule is related in the figure 1-1. Here, $\hbar$ (Dirac's constant, $h/2\pi$) is called as the reduced plank constant and the character ω is the anguler frequency of light. Then energy of the light is explained as $E = \hbar\omega$. Each subscript i and r with ω appearing in the figure indicates the irradiated light and Raman scattering light, respectively. In the Stokes site (Figure 1-1. S), the molecule at vibrationally ground state ($v = 0$) is excited by the incident radiation ($\hbar\omega_i$) virtually (middle in electronic between vibrational state) results in the scattered light with loss of energy with $\hbar$ ($\omega_i - \omega_r$), which corresponds to the energy difference from the excited to ground state ($\hbar\omega_S$). On the other hand, when the molecule at vibrationally excited state ($v = 1$) is

excited by the incident radiation ($\hbar\omega_i$), the scattered light gains energy as much as $\hbar\omega_r$ with transition of the originally vibrational excited state ($\nu = 1$) to the ground one ($\nu = 0$). Practically, Stokes and anti-Stokes scattering of the light take place simultaneously, because a lot of molecules in the system are distributed in different vibrational states at $\nu = 0, 1, 2..$ with the population ratio determined by the rule of Boltzmann distribution. However, depending on the fact that much larger population of ground state than excited state in the sample, the number of photons by Stokes scattering manner is much larger than that of anti-Stokes manner. Therefore, in most cases, the Stokes site is detected for molecular study purpose usually.

Generally speaking, the Stokes Raman scattering is called as Raman Spectrum, and the scattered light can be realized as a phenomenon in which several interactions with light and molecular vibrational modes results in generating photons with shifted longer wavelength (loss of frequency as well as energy) than the excitation light. In order to make sure how much shifted, a Laser light is used as an excitation light source. The photons are dispersed into certain range of wavelength by a spectrophotometer. Customary, the horizontal axis of a Raman spectrum is called “Raman shift” as an energy difference from that of the excited light. The unit for it is usually written as a reciprocal wavelength expressed in cm (wavenumber, cm^{-1}). The Wavenumber is common for infrared absorption spectroscopy (IR) as Figure 1-1 shows same energy transition to Raman scattering. Raman shift is comparable measure for Raman spectra frequently measured in different excitation laser sources.

Figure 1-2 shows an example of Raman spectrum of adenine in an aqueous solution measured by 632. 8 nm excitation. When the Raman spectrum is measured for even pure molecule, there might be some bands (called as Raman bands) each of them reflecting certain vibration modes in the molecule.

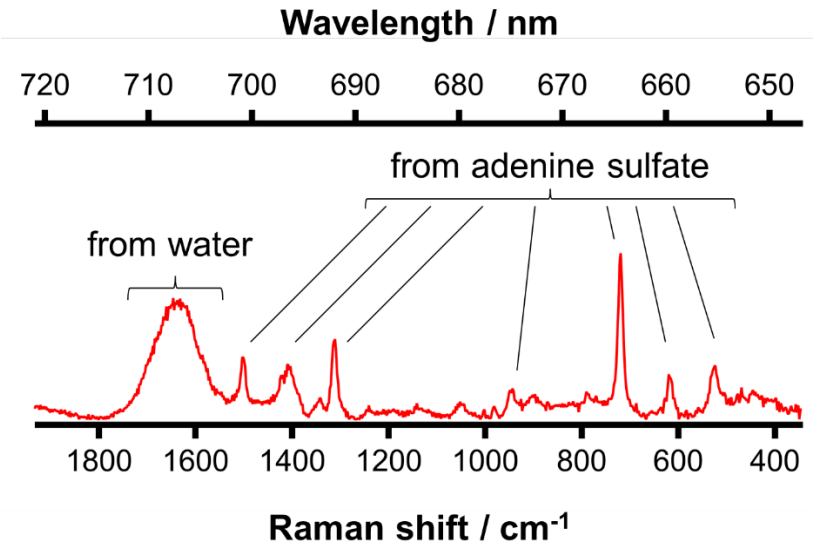


Figure 1-2. An example of Raman spectrum of adenine sulfate (5 mg/ml) in 0.5 M HCl aq. solution. The excitation wavelength was 632. 8 nm and Laser power at sample was 2 mW. Exposure time was 300 sec and 5 times average to acquire this Raman spectrum.

Raman cross section is an optical property unique to every molecular species. Simply speaking, a molecule with higher polarizability tends to give stronger Raman scattering light. When we apply Raman spectroscopy to some biological system, the relatively low polarizability of water is a strong benefit because the relatively weak Raman scattering causes little influence for measuring such samples, in solution, tissues, cells, etc. This aspect may be one of stronger points of Raman spectroscopy than FT-IR spectroscopy, in which the overwhelmingly strong water absorption bands frequently hinders the measurement of weaker sample bands. We have to think about the hinderance of strong auto-fluorescence light. It frequently occurs simultaneously to Raman scattering (Figure 1-1. FL) and causes a serious difficulty in measuring much weaker Raman bands. Theoretically speaking, the relative intensity of fluorescence light is 10^4 to 10^8 stronger than Raman scattered light.

I - ii Overview on Raman spectra of typical biomolecules

Raman spectrum is unique to each molecule and gives characteristic pattern, which is called as “molecular finger print”. This pattern is originated by the combination of molecular vibrational modes. The assignment of each band in a Raman spectrum assignable to certain vibrational mode is very important for understand the molecular structure and molecular environments. Representative biomolecules such as protein, nucleic acids, saccharides, lipids and pigments are deeply and repeatedly investigated by both experimental observations and theoretical calculations.

The relationship between the structure of proteins and Raman spectra are well studied. For example, a sharp and strong band observed in Raman spectra of proteins is due to the side chain of

Phenylalanine (Phe). This band is assigned to so called breathing mode of benzene ring (mode ν_1), often used as a protein marker band. This band is frequently used as standard Raman band for a protein by the clear visibility; it is observed very clearly in most protein samples and frequently is distinguished for the sample with strong auto-fluorescence. Phe gives relatively weak Raman bands also, at 1203, 1032 and 624 cm^{-1} due to different vibrational modes. Raman spectra due to the secondary structures of protein or peptide are well studied. Vibrational modes assignable to peptide bond are called as amide A, I, II, III, V, IV, VI and VII. They can be observed in Raman and infrared spectra and used as a measure to estimate the secondary structure of proteins and peptides. Amide I (C=O stretching and NH bending around 1660 cm^{-1}) and III (C=N stretching + HN bending around 1250 cm^{-1}) are Raman active, it is known that the wavenumber shift of these two bands would be found depending on secondary structure such as α-helix, β-sheets, random coils and turns (2, 3).

In the lipid, C−H stretching vibration from CH_2 and CH_3 appeared at around 2800 cm^{-1} called CH region, and CH_2 bending mode is observed at around 1444 cm^{-1} (4) that are often dealt as marker of total lipid abundance. Unsaturated fatty acids or their triglyceride gave C=C stretching mode around 1650 cm^{-1}. Ester bond around 1750 cm^{-1} is also picked up as triglyceride marker. In addition, CH_2 twisting around 1300 cm^{-1}, =C−H deformation around 1260 cm^{-1} and adjacent three C—C stretching mode around 1100 cm^{-1} are used for detail discussion related index of unsaturation.

These spectral features depend on the phase of the matter (liquid/solid state) and molecular surroundings (water solution/organic solvent etc.). Therefore, it is especially necessary to be aware of such effects during the experiment involving biological and medical applications and make comparisons with appropriate molecular standards. It is important to understand that most of the basic studies for the assignment of Raman bands of biomolecules were done in pure molecular systems, while the actual biological or medical samples are usually complex systems containing mixtures of biomolecules. Here, the problem is that superposition (or overlapping) of some Raman bands derived from different molecules. Picking up one case to explain about the problem with lipid and protein, the superposition of Amide I and C=C stretching mode around 1650 cm^{-1} make discussions difficult. To solve this problem, exploration of Raman marker band is important by eliminating possibility of such superposition in the interested system.

I - iii Application of Raman spectroscopy for biological and medical researches

No necessity of special sample preparation and low invasiveness are benefits to use Raman spectroscopy for biological and medical applications. Combination of microscopy with Raman spectroscopy (called Raman micro-spectroscopy) gives label-free molecular imaging. Since the report of Raman spectrum measurement of a single cell (5) and the report of Raman imaging of the division process of fission yeast (6), many studies have been conducted by Raman micro-spectroscopy to study biological and medical interests and problems. When it is desired to qualitatively compare

experimental conditions different samples, mean Raman spectrum is compared and the difference spectrum is taken to see the possibility of such as band shift and relative intensity ratio in many cases. Moreover, since the Raman scattering intensity is proportional to the sample concentration, a semi-quantitative analysis method can be performed by searching for a marker band of the target molecule. In addition, if the sample is measured by scanning in 2D plane or 3D space, semi-quantitative analysis and distribution analysis can be performed simultaneously by imaging (7). As a basic analysis method, the area of the target marker band is calculated by Gaussian fitting or the sum of the differences between the baseline and the peak intensity (8), and the area is treated as a variable, which is called univariate analysis.

As mentioned in the previous section, there are various vibration modes even in a single molecule, and there are numerous molecules in the biomolecular samples, so the contribution of unknown molecules cannot be denied. Therefore, quantitative interpretation of Raman spectrum measurement requires a deep understanding of the molecular species presumed in the sample and their standard Raman spectrum. In addition, many of the possible variables measured with great effort are wasted in univariate analysis when it is wanted to applied classification and regression. In order to solve these problems, multivariate analysis has been performed in the last 2 decades in the application of Raman spectroscopy for biological and medical samples.

I - iv Multivariate Analysis

In multivariate analysis, there are mainly two categories: multivariate classification (MC) and multivariate regression (MR). MC tries to finds patterns in the given data and helps in simplification by compressing or grouping the data. It can be further divided into either unsupervised or supervised learning. Unsupervised MC methods [e.g., singular value decomposition (SVD), principal components analysis (PCA), cluster analysis, independent component analysis, multivariate curve resolution-alternating least squares (MCR-ALS) analysis, etc.] are exploratory in nature and needs no *a priori* information (9). Especially, PCA is one of the most common technique to reduce dimension (number of variable) to understand data structure. In contrast, supervised MC methods (e.g., linear discriminant analysis (LDA), multiple linear regression, partial least squares discriminant analysis, support vector machines (SVM), neural networks (NN), etc.) requires *a priori* knowledge, such as class labels identifying the sample group. These methods make classification model and are applied to unknown samples for purposes such as diagnosis of disease and classification of certain type of cells etc. MR analyses such as linear regression, principal component regression, partial least squares regression, etc. model the data by training from given data sets with known variables such as concentration in order to estimate relationships among the variables. It will eventually be applied to predict the unknown once a model is developed with sufficient accuracy (9). Combination of these analysis is also common. For example, PCA (unsupervised) followed by LDA (supervised) is called

PC-LDA. Since LDA requires smaller variable than sample population, PCA is applied to reduce dimension while keep original data information as much as possible.

In this section, SVD, PCA and MCR-ALS have been explained in detail as they have been employed extensively in this thesis (Chapter II and III). To apply multivariate analysis, Raman spectral data is recognized as vector and matrix. A single spectrum is recognized as m-dimensional column vector $\boldsymbol{x}$.

$$\boldsymbol{x}^t = (x_1, x_2, \ldots, x_{m-1}, x_m)$$

Here, n is number of points per spectrum that is usually within the number of channels in a detector. Since we disperse Raman scattered light using a grating, each channel receives light of different wavelength. They convert incoming photons into electron charges to eventually obtain a Raman spectrum. Here, n channels with corresponding wavenumbers forms the horizontal axis (Raman shift). When there are m number of spectra arranged in a 2D data matrix $\boldsymbol{X}$ (n × m), the expression of the equation is the following:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

SVD

SVD is applied when one wants to know independent spectral components number and to remove noise by using the results of matrix decomposition. SVD decompose matrix $\boldsymbol{X}$ as follows (10):

$$X = U\Lambda V^t + E \quad (1)$$

Where U is matrix aligned eigenvector of XX^t. The Λ is a diagonal matrix of singular value, each element has square root λ, which is eigen value of AA^t. V is a matrix aligned eigenvector of X^tX. E is a residual matrix. When the singular values are plotted in descending order, number of significant components and boundary between spectral and noise components can be identified. The difference usually become small along to low order of singular values and eventually almost identical. We can regard the point as boundary of spectral and noise components. This plot is also useful to estimate number of components in MCR-ALS. Finally, de-noise is achieved by substituting singular value in Λ into 0 only after the boundary and calculating $\boldsymbol{X}$ again by multiplication of $\boldsymbol{U}$, substituted $\boldsymbol{\Lambda}$, and $\boldsymbol{V^t}$.

PCA

In the given data, PCA tries to find new 'axis', which make variance from the axis to original variable maximum. The axis is called first principal components or loading vector. Once the first principal component is determined, second principal components as well as third, fourth, … are determined to make variance maximum again under keeping a condition of orthogonality to previous loading vector. Mathematically, a result of PCA decomposition is related to SVD and described as

bilinear model by following (10):

$$X = TP^t + E \quad (2)$$

Where T is score, P and E are loading and residual matrix, respectively. The loading matrix ***P*** is derived from using Eq (1).

$$\Lambda V^t = P \quad (3)$$

Then, T is a score matrix corresponding U in Eq (1).

To get matrices P and T by using SVD results, original data matrix **X** should be standardized before the SVD. Instead of SVD, the major alternative way to introduce these metrices is nonlinear iterative partial least square (NIPALS) algorithm that can be used to find only first few principal components so as to save computational time.

Each principal component preserves spectral feature with corresponding score. The scores which is useful for unsupervised MC can also be used for supervised MC and MR such as PC-LDA and principal components regression. Although these methods achieve high classification accuracy, they have an inherent problem i.e., loadings which contain spectral information lack physical meaning because of the presence of both negative and positive values (Raman spectrum is never negative!). One idea to solve this problem and get physically meaningful spectra is application of non-negative constraints to MCR-ALS. This is also called non-negative matrix factorization (NMF).

MCR-ALS

In MCR-ALS analysis, matrix approximation sought by a linear combination of desired number of spectral components can be written as follows:

$$A \approx WH$$

In this low-rank approximation, A is original mapping data of dimension m × n (m denotes number of points per spectrum and n denotes the total number of spectra). Note the A is transposed X in Eq (1) and Eq (2). W (m × k matrix) represents spectral components and rows of H (k × n matrix) represent intensity profile of each spectral component. The parameter k, the number of components, can be flexibly decided by referring SVD analysis or *a priori* estimation. W and H were iteratively refined using alternating least squares, so that the Frobenius norm $||A-WH||^2$ is minimized with non-negative constraints W≥0 and H≥0 (11).

In this thesis, to obtain sparser solutions, L1 penalty term for H (lasso regression) of α is applied as follows:

$$(W^tW + \alpha^2 E)H = W^tA$$

where E is a k × k matrix whose elements are all unity. In addition, L2 penalty term for W (ridge regression) of β is also applicable as follows:

$$(HH^t + \beta^2 I)W = HA^t$$

where I is a k × k identity matrix. L1 penalty term for W and L2 penalty term for H are also applicable in homemade program developed by pylon (8). Non-negativity in W and H endows us to interpret the result with physical meaning. Only weak point of this technique is the result is not unique and depends on initial conditions, setting of k and use of penalty terms.

I - v Objective of this thesis

Even though MCR-ALS technique are demonstrated and getting attention, practical use for biological and medical studies are not so many (12, 13). Based on such backgrounds, I studied some biomedical systems with Raman spectroscopy coupled with MCR-ALS. This thesis consists of the results of two different such applications;

1) **Visualizing wax ester fermentation in single *Euglena gracilis* cells by Raman microspectroscopy and multivariate curve resolution analysis**

2) **Identification of Molecular Basis for Objective Discrimination of Breast Cancer Cells (MCF-7) from Normal Human Mammary Epithelial Cells by Raman Microspectroscopy and Multivariate Curve Resolution Analysis**

In Chapter II, Raman micro spectroscopy coupled with MCR-ALS was demonstrated as a screening method for organisms suitable for biomass production by using *Euglena*.

In Chapter III, As medical application, we searched for Raman spectral markers to classify cultured breast cancer and normal cells, and discovered specific lipids as marker.

Both Chapters II and III use classical univariate analysis methods and popular multivariate analysis and discuss the findings obtained by performing MCR-ALS analysis with together. The purpose of this thesis is to show the significant finding of each of these studies in the biological and medical field as well as usefulness of MCR-ALS in Raman spectroscopy.

Chapter II.

Visualizing wax ester fermentation in single *Euglena gracilis* cells by Raman microspectroscopy and multivariate curve resolution analysis

II - i Introduction

Fossil fuels contribute to two thirds of the global energy demand out of which oils contribute 33% (14, 15). In an age of increasing population growth, overconsumption and depleting oil supplies, continued use of petroleum sourced fuels is both unsustainable and damaging to environment with long-standing negative impacts on public health and global climate(16, 17). Therefore, there is an urgent need to find suitable renewable energy sources. Microalgal biofuels are currently the most favored substitute for liquid fossil fuels than other nontoxic, eco-friendly alternatives such as plant or animal biomass derived energy. Microalgae offers several advantages: (1) easy and quick growth under various conditions, (2) does not compete for arable land and water with edible crops, and (3) provides carbon neutral renewable energy by converting CO_2 to useful products such as fatty acids, alcohols, and neutral lipids. Many algae generally produce substantial amounts of triacylglycerol of medium-chain fatty acids such as palmitic (C16:0) and stearic (C18:0) acids, sometimes up to 70% of its dry weight(18, 19).

One such microalgae that has received considerable attention in the past few decades as a biotechnological tool to produce drop-in jet fuel is *Euglena gracilis*, a photosynthetic unicellular flagellate eukaryote. *Euglena* being a mixotroph, feeds as an autotroph in the presence of sunlight to produce sugars through photosynthesis while survives as a heterotroph taking in dissolved organic compounds as nutrition under dark conditions. One of the main reasons for its attraction is because of its ability to produce wax esters, chiefly myristyl myristate (MM). MM is made up of myristic (C14:0) acid and myristyl alcohol (C14:0), each of which can individually be utilized for jet fuel because of their low freezing point/high cetane number compared to other medium-chain fatty acids (20). Typically, *Euglena* cells accumulate storage polysaccharide called paramylon granules, a β-1,3-glucan under aerobic conditions. However, such stored paramylon is broken down to glucose and further converted to wax esters when put under anaerobic conditions. Since the anaerobic cells gain subtle levels of ATP during the process, the phenomenon is called “wax ester fermentation” (21).

Though *Euglena* cells have huge potential and can serve as tiny factories for biofuel production, inherent problem associated with large scale culturing is the slow growth rate of algal strains with high oil content (21, 22). It appears that the synthesis and storage of wax esters as cytosolic lipid particles is *Euglena*’s defense mechanism to cope with stress (18). Therefore, much effort has been put to genetically engineer or optimize culturing conditions of algae for enhanced biofuel production (21, 23-26). To evaluate any constructed algal strain or the choice of culture conditions, polysaccharide/lipid profiles must be characterized. Conventional quantification methods employ labor intensive, time consuming, destructive chemical extraction procedures followed by expensive mass spectrometric measurements thereby limiting scientific progress.

Therefore, we set out to develop a Raman spectroscopy (RS) based molecular imaging method to characterize various metabolites in *Euglena* in a simple and straightforward manner. Raman

spectrum, which is also called a molecular fingerprint, provides wealth of chemical information with high specificity. Combining RS with a microscope endows subcellular resolution. Moreover, it is a rapid, non-destructive, live cell compatible technique that requires no additional dye probes or extensive sample preparation for molecular imaging. Previously, metabolic heterogeneity of live *Euglena* was studied in real time by stimulated Raman scattering. However, only the heavily crowded C−H stretching region could be analyzed (27). Spontaneous Raman spectroscopy has also proved to be useful in studying enhanced lipid production in yeasts (28). In this work, we performed space- and time-resolved Raman imaging of single living *Euglena* cells under anaerobic conditions and analyzed fingerprint region rich in molecular and structural information to identify/visualize paramylon and products of wax ester fermentation.

We identified Raman spectral markers for β-1,3-glucan/esters and constructed their intracellular distribution images by simple univariate approach. In order to obtain carbon chain length specific information of lipids and further probe any other unknown components, we employed multivariate curve resolution analysis and succeeded in identifying MM (C28), a major product of wax ester fermentation, which is ideal for a drop-in bio jet fuel.

II - ii Material and methods

Sample preparation

Euglena gracilis SM-ZK, a non-photosynthetic mutant was used in this study. First, *Euglena* was pre-cultured aerobically in Koren–Hutner (KH) medium until stationary phase, diluted 20 times with fresh medium, and cultured aerobically for another 2 days. To perform anaerobic digestion, 1.5 ml of aerobically grown culture was taken in an eppendorf tube of the same volume and sealed with parafilm. All steps were done on a rotary shaker (120 rpm) at 26 ℃ under dark conditions (25, 29, 30). For Raman spectroscopic measurements, since *Euglena* are flagellates, 20 µl of culture at each time (0 h, 12 h, 24 h, and 48 h) was put on a concanavalin-A coated glass bottom dish. Then, after standing for about 5 min, a few ml of lukewarm (~ 35 ℃) 2% agarose solution was added to further restrict their motion. The glass bottom dish containing *Euglena* cells was then transferred to the microscope as it is for Raman imaging experiment and two cells were measured at each time. All chemical standards were bought either from Sigma-Aldrich or Wako, Japan, and measured using glass bottom dish.

Raman spectroscopy

Raman spectra were measured using a homemade confocal Raman microspectrometer equipped with a He–Ne Laser (632.8 nm) (12). The laser beam was introduced into an inverted microscope (Olympus, IX70) and tightly focused onto the sample on the microscope stage using oil immersion objective lens (100×, NA=1.3). Backscattered light including the inelastically scattered photons was collected by the same objective lens and passed through an edge filter to remove elastic

scattering light. In the Raman path, a 50 μm pinhole was set up to achieve confocality before light entered polychromator (Chromex, 250IS). A liquid nitrogen cooled CCD detector operating at −120 ℃ (Princeton Instruments, Spec-10) was used to record Raman spectra. The entrance slit width of the polychromator was set to 50 μm and measurements were done using a 600 g/mm grating, resulting in spectral resolution of ~ 4.5 cm^{-1}. Lateral and axial resolutions were 300 nm and 3 μm, respectively.

For imaging experiments, a step size of 0.6 μm in X- and Y-direction was used with the help of a piezo stage (Physik Instrumente). Each *Euglena* cell, being relatively large, took about ~40 min to scan the whole cell with an exposure of just 1 s/spectrum. Laser power of 4 mW at the sample point was used for all measurements. An exposure time of 30 s and 60 s was used for measuring several points of lipids and β-1,3-glucan standards, respectively, and averaged. CCD detector and piezo stage were controlled using the LabVIEW software (National Instruments). All measurements were done at room temperature (22 ℃).

Data analysis

Data pre-processing such as dark subtraction, intensity correction (using white light spectrum), and spectral de-noising by singular value decomposition analysis were all carried out in IGOR Pro (Wavemetrics). All standard spectra were an average of several points and the fluorescence background was removed by assuming a polynomial baseline.

Raman imaging data from *Euglena* were analyzed by multivariate curve resolution performed on homemade program written in Python which was used previously (12, 31).

A seven-component model (initialized with six random components and one fixed straight baseline) was constructed. To obtain sparser solutions, L1 penalty term (lasso regression) of $\alpha^2 = 0.008$ and L2 penalty term (ridge regression) of $\beta^2 = 0.008$ were applied.

II - iii Results and discussion

Raman microspectroscopy and conventional univariate imaging of single *Euglena gracilis* cells

To understand wax ester fermentation in *Euglena* at the molecular level, we measured space- and time-resolved Raman spectra and images of single cells grown under anaerobic conditions (Fig 2-1.).

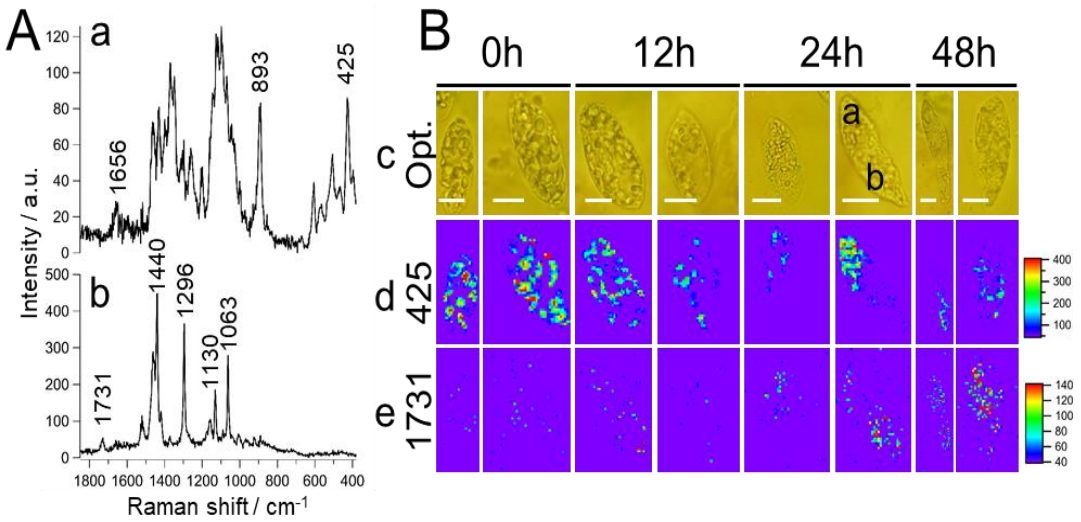


Fig 2-1. Raman microspectroscopy and imaging of single *Euglena gracilis* cells under anaerobic condition. **A** Space-resolved Raman spectra measured at polysaccharide-rich region (a), and ester-rich region (b) from a cell at 24 h. **B** Optical images of single *E. gracilis* cells (c), time-resolved univariate Raman images of polysaccharides (d), and esters (e). Scale bar in each optical image measures 10 μm and measured points are indicated using alphabets

As mentioned earlier, stored polysaccharides in *Euglena* are converted to wax esters. Therefore, to identify and discuss Raman spectral markers during wax ester fermentation, two most relevant space-resolved Raman spectra from a *Euglena* cell are presented in Fig. 2-1. A. Spectrum at point a (Fig. 2-1. A-a) had COO− asymmetric stretching at 1656 cm^{-1}, COO− symmetric stretch, and C–H deformation modes between 1500 and 1200 cm^{-1}, C–C and C–O stretch modes of pyranose rings between 1150 and 1050 cm^{-1}, and C–C–C ring deformation mode at 425 cm^{-1} indicating polysaccharide-rich region. In addition, we observed a band at 893 cm^{-1}, a region which is sensitive to glycosidic linkages. In fact, Raman spectroscopic studies on series of carbohydrate monomers have revealed C–H equatorial bending vibration of β-anomer between 905 and 885 cm^{-1} (12, 32, 33). We can safely assume that the observed polysaccharide spectrum may particularly be rich in paramylon, a β-glucan (Scheme 2-1 a).

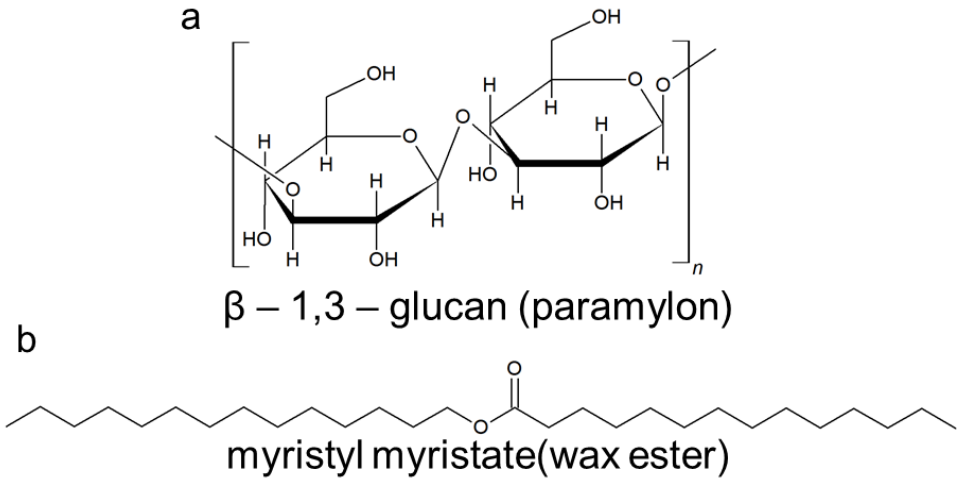
However, actual comparison with pure β-glucan is necessary. Major features in Raman spectrum measured at point b (Fig 2-1. A-b) include C=O stretch of ester linkage at 1731 cm^{-1}, C–H-

bending vibrations of the aliphatic chain at 1440 cm^{-1}, in-plane CH_2 twist at 1296 cm^{-1}, and C–C stretch between 1150 and 1050 cm^{-1}. It is important to note that the absence of any band in the C=C stretch region around 1650 cm^{-1} clearly indicates that this strain only accumulates esters containing saturated hydrocarbon chains (Scheme 2-1 b).

It is then straightforward to choose 425 cm^{-1} and 1731 cm^{-1} bands to be markers of paramylon and wax esters, respectively. To visualize dynamic intracellular distributions of these components, we performed time-resolved Raman imaging experiment of single *Euglena* cells at 0 h, 12 h, 24 h, and 48 h under anaerobic conditions. Two representative cells at each time are presented in Fig. 2-1 B. It is apparent from univariate Raman images that cells at 0 h (pre-grown under aerobic conditions) have accumulated polysaccharides, while ester content is negligible. As the culture time progresses, stored polysaccharide content decreases slowly, while wax esters start accumulating, especially from 24 h. This is a clear indication of wax ester fermentation in *Euglena*.

Identification of carbon chain lengths in wax esters

Though we were able to visualize the fatty acid biosynthetic machinery at work, there is no information on the nature of wax esters produced. Because, the C=O stretch of ester linkage (1731 cm^{-1}) used for molecular imaging does not indicate carbon chain lengths in compounds



Scheme 2-1. Molecular structures. **a** β-1,3-glucan (paramylon) and **b** myristyl myristate

containing > 12 carbons (34), which is quite important in the context of its application for biofuel production. Therefore, to characterize the chain length of wax esters in detail within single *Euglena* cells, we set out to identify Raman markers that are sensitive to carbon chains. To achieve this, we measured series of standard wax esters with different chain lengths together with myristic acid and myristyl alcohol, precursors of MM which is a promising candidate for drop-in jet fuel (Fig. 2-2).

It is known that the position of C=O-stretching band in fatty acid methyl ester depends on the chain

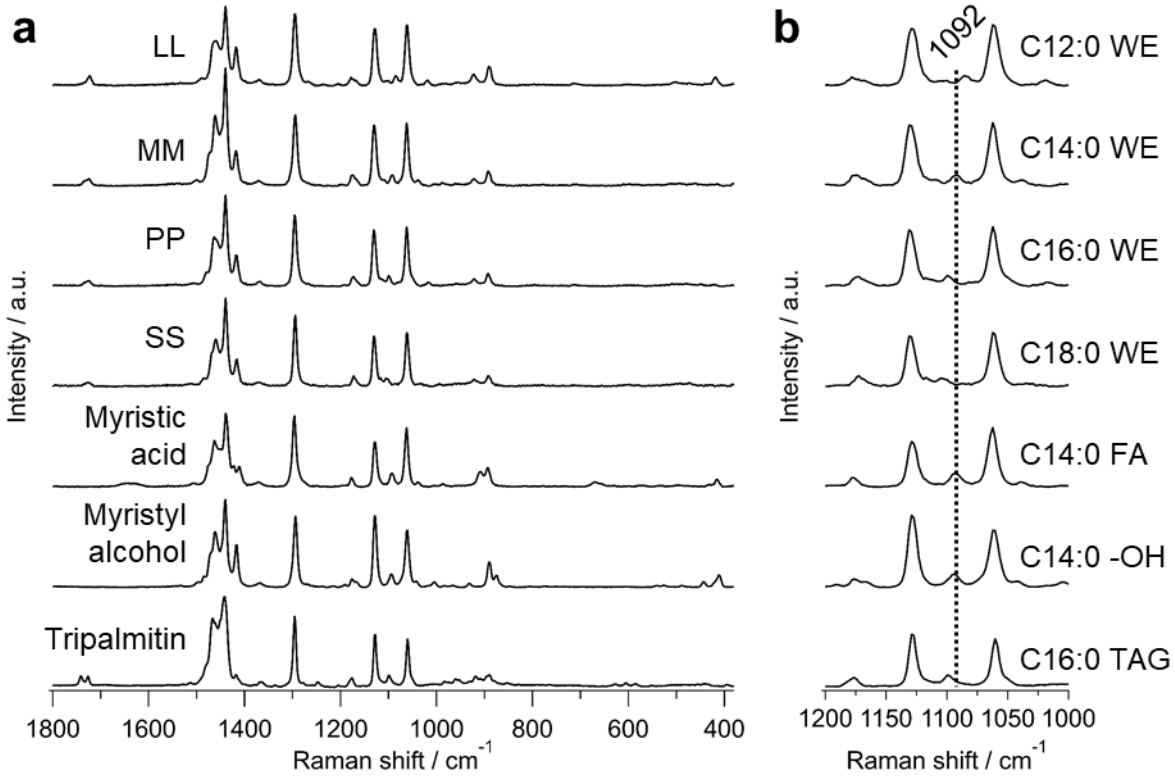


Fig. 2-2. Comparison of Raman spectra of standard wax esters, lipid, and alcohol. a Fingerprint region (1800–800 cm^{-1}) of lauryl laurate (LL), myristyl myristate (MM), palmityl palmitate (PP), stearyl stearate (SS), myristic acid, myristyl alcohol, and tripalmitin. b Enlarged view of the 1200–1000 cm^{-1} region containing C–C stretch information, which is useful for chain length analysis. Corresponding carbon chain lengths for each compound is indicated using common notation. Fluorescence background was subtracted using a polynomial baseline and all spectra were normalized to 1296 cm^{-1} band

lengths. However, it is useful only for oils containing < 12 carbon atoms and the change in band position is minimal for fatty acids > 12 carbons (34). Although the overall spectral pattern looked very similar (Fig. 2-2. a), careful screening of C–C-stretching region revealed significant difference that can be attributed to chain lengths (Fig. 2-2. b). Raman bands at 1130 cm^{-1} and 1063 cm^{-1} have been assigned to in-phase and out-of-phase skeletal C–C-stretching vibrations, respectively, for all-trans chain conformation. The band in between these two is a superposition of all-trans C–C stretch with a single gauche defect and C–C stretching of gauche conformation which is indicative of gauche isomer formation (35-38). Its position has been found to be sensitive to carbon chain lengths and we observed a systematic shift to higher wavenumber with increasing carbon number (Table 2-1).

Table 2-1. Carbon chain length dependence of gauche conformation sensitive C–C-stretching band in saturated chains

Compound	Chain length	Band position [a] (cm^{-1})
Lauryl laurate	12:0/12:0	1084.7 ± 0.9
Myristyl myristate	14:0/14:0	1092.2 ± 0.8
Palmityl palmitate	16:0/16:0	1099.1 ± 0.8
Stearyl stearate	18:0/18:0	1103.7 ± 0.9
Myristic acid	14:0	1092.6 ± 0.4
Myristyl alcohol	14:0	1094.1 ± 0.3
Tripalmitin	16:0/16:0/16:0	1098.6 ± 0.7

[a] Gaussian fitting was used to determine band positions and fitting errors are included

MM and both its precursors which contain C14:0 show Raman band close to 1092 cm^{-1} while others are shifted in either direction. Even though we succeeded in identifying chain length-specific Raman spectral markers, we must keep in mind that the difference in band position is quite small and that the measured samples were all pure compounds in solid state. This indicator has been shown to fail if the lipids are in liquid state (34).

Extracting pure biomolecular information using MCR analysis

In the present context, *Euglena* cells contain heterogeneous distributions of many different biomolecules with varying phases and Raman spectrum measured at any given point in the cell is a mixture of all components. For example, C–C-stretching region of Raman spectrum is quite crowded with overlapping contributions not only from lipids or esters but also from other intracellular biomolecules such as protein, nucleic acids, polysaccharides, etc. Therefore, simple univariate approach is not suitable for such complex biological samples, especially to predict chain lengths. In fact, if we take a closer look into the space-resolved spectrum from ester-rich region between 1150–1050 cm^{-1} (Fig. 2-1A-b), it is hard to find any C–C gauche band. However, a broad and an intense band can be observed in the same region from polysaccharide-rich Raman spectrum (Fig. 2-1A-a) indicating the complexity involved. Therefore, we applied MCR analysis to extract pure biomolecular information and to visualize intracellular abundance of each component in a straightforward manner.

Results of seven components MCR model is given in Fig. 2-3 in which a straight baseline was intentionally included to eliminate varying offset. Other six components were automatically extracted. Let us look into the assignment of each in detail. Figure 2-3b includes O–H-bending vibration of water around ~1600 cm^{-1} and an overall broad fluorescence background. Raman spectrum in Fig. 2-3c includes phenylalanine ring breathing mode at 1004 cm^{-1} and amide I band at 1660 cm^{-1}, indicating proteins. Next component (Fig. 2-3d) contains intense bands at 1522 cm^{-1} and

1158 cm^{-1} which represent stretching modes of C=C and C–C of polyene chain in carotenoids, respectively. It is important to note that protein and carotenoid spectra were obtained as a natural consequence of MCR analysis without any *a priori* knowledge of their presence and, thus, could be very useful in exploratory analysis. Spectrum in Fig. 2-3e can be assigned to polysaccharide. Unexpectedly, we extracted two lipid components, named lipid 1 and 2, as shown in Fig. 2-3f, g, respectively.

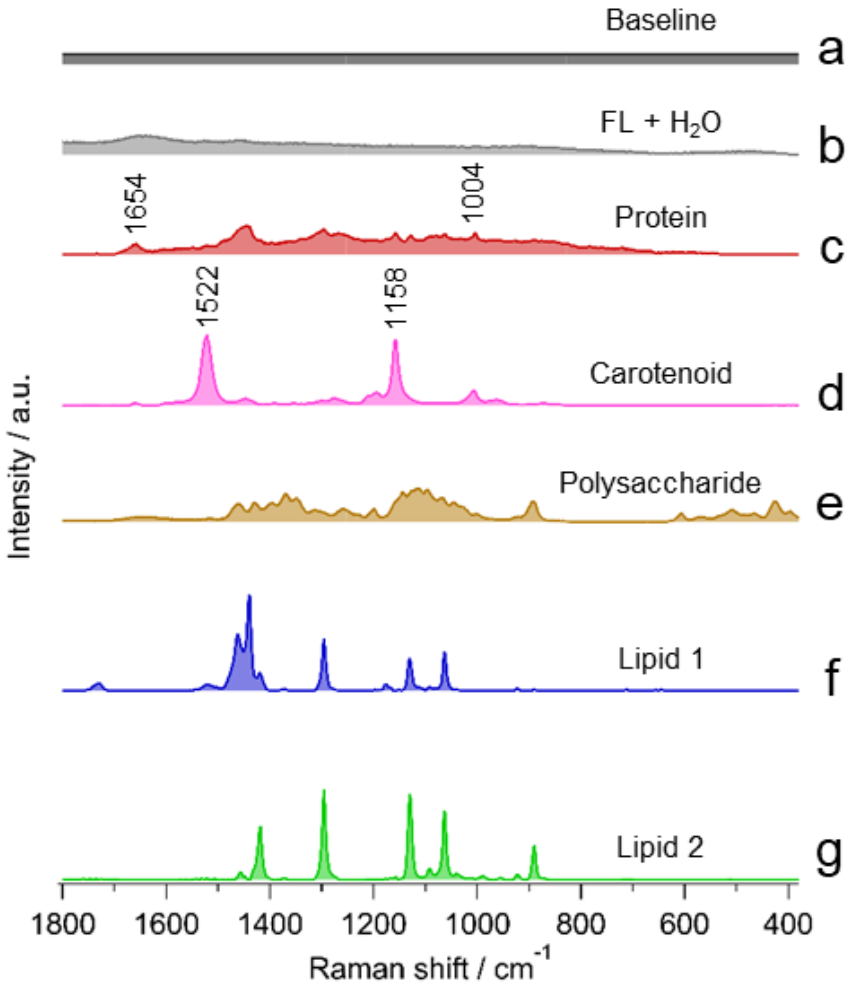


Fig. 2-3. Results of MCR analysis assuming seven components. (a) Baseline, (b) fluorescence background (FL) and water, (c) protein, (d) carotenoid, (e) polysaccharide, and (f, g) lipids 1 and 2,

Comparison of MCR-extracted components with pure standards

To understand the origin of polysaccharide and lipids from MCR analysis more specifically, we compared them with series of pure chemical standards of wax esters and their precursors. After screening, comparison with expected compounds such as β-1,3-glucan and MM is shown in Fig. 2-4.

This is mainly because *Euglena* is known to store appreciable amounts of paramylon (a β-1,3-glucan) as energy reserves under aerobic conditions which are almost converted exclusively to wax esters containing saturated carbon chains. Gas chromatographic analysis showed esters with C28 to be the major component along with minor contributions from other even numbered esters in C24–C32 range (39). Indeed, MCR-extracted polysaccharide component matches very well with β-1,3-glucan and can unambiguously be assigned to paramylon in *Euglena*. It is intriguing that two seemingly similar lipid components were extracted separately in MCR analysis (Fig. 2-4d, e). Lipid 1 with bands at 1732 cm^{-1}, 1440 cm^{-1}, 1296 cm^{-1}, 1130 cm^{-1}, 1092 cm^{-1}, and 1063 cm^{-1} matches quite well with MM and can be assigned to C28 ester containing two saturated C14 chains. A closer look into lipid 2 reveals the absence of 1732 cm^{-1} and C–H-bending vibrations at 1440 cm^{-1}, while 1417 cm^{-1} and 890 cm^{-1} are more pronounced. Absence of C=O stretch band of ester raises the question whether lipid 2 is really a lipid/ester. However, the presence of 1092 cm^{-1} along with other C–C-stretching vibrational bands indicates C14 carbon chain, indirectly suggesting that it could either be myristic acid or myristyl alcohol. However, it does not correspond well with neither as expected, especially in C=O stretch and C–H deformation region, essentially leaving the spectrum unassigned.

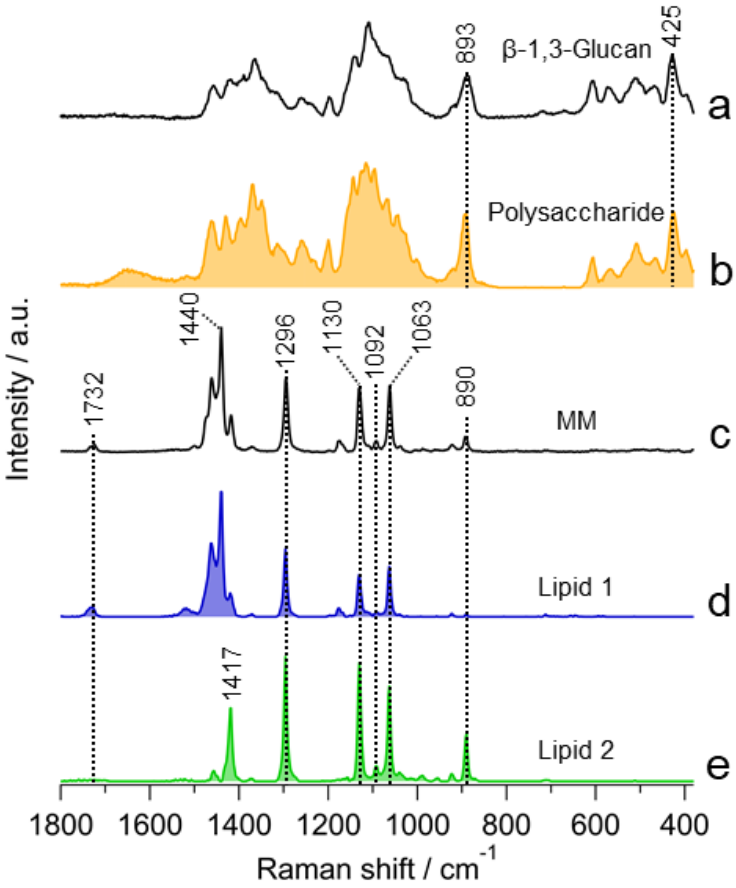


Fig. 2-4. Comparison of MCR spectral components with pure standards. (a) β-1,3-glucan, (b) MCR-extracted polysaccharide (same as Fig. 2-3e), (c) myristyl myristate (MM), and (d, e) MCR lipids 1 and 2, respectively (same as Fig. 2-3f, g)

MCR analysis of Raman images of standard myristyl myristate

Since lipid 2 extracted from *Euglena* with 1092 cm^{-1} band does not match either with wax ester or their precursors, we performed Raman imaging on pure MM solid film (obtained after drying 10 mg/ml MM in hexane) and carried out detailed MCR analysis (Fig. 2-5). A two-component MCR model constructed from data of pure MM showed surprising results. The two spectra were, indeed, identical to the two lipid components obtained from the MCR analysis of living *Euglena* cells, i.e., Fig. 2-5A-a (MCR_MM1) and Fig. 2-5A-c (lipid 1) were identical and both correspond well to averaged MM spectrum measured earlier (Fig. 2-4c). Spectral profile of second component (MCR_MM2), in which bands at 1732 cm^{-1} and 1440 cm^{-1} were missing, was identical to 'lipid 2' from *Euglena* cells, indicating its origin to MM. Only plausible explanation is the presence of crystal

polymorphs (several crystalline structures with the same chemical composition).

It is known that long-chain esters/triglycerides exist in three major polymorphic forms, namely α, β′, and β. Their stability varies in the order β > β′ > α. While the subcell structure of α form is hexagonal with no ordered arrangement of chain planes (H), β′ is orthorhombic with every second chain being perpendicular to the rest (O⊥) and β is triclinic with all chain planes parallel (T//). In a Raman spectrum, C–H deformation modes between 1500 and 1400 cm^{-1} are sensitive to crystal structure. First

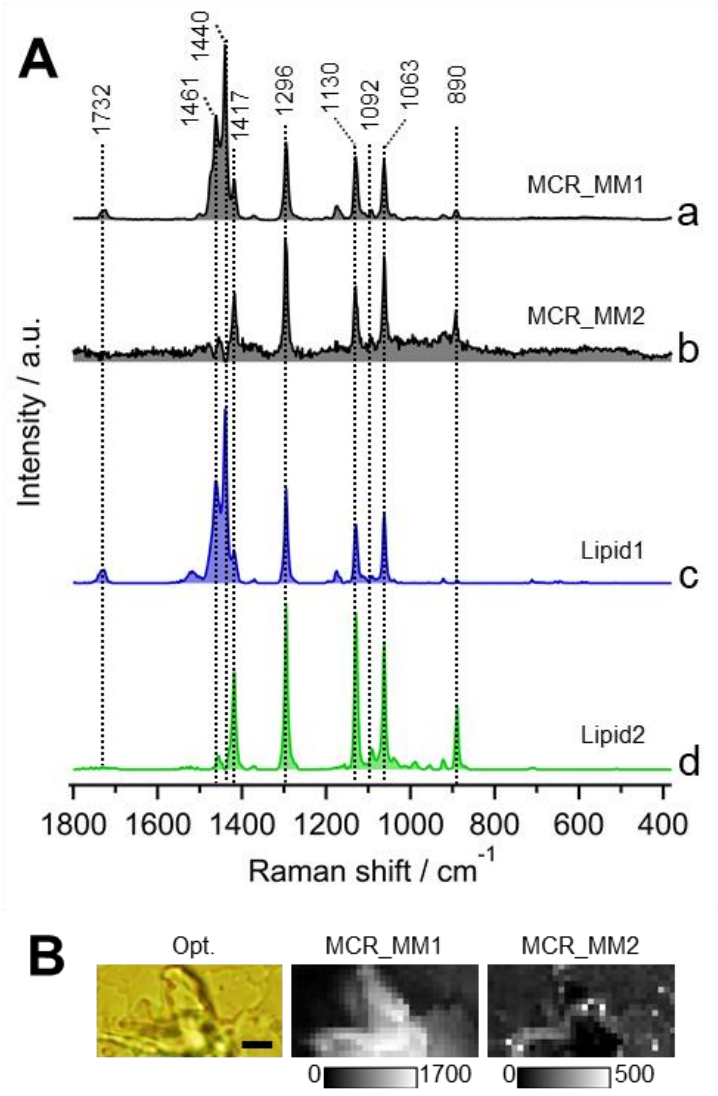


Fig. 2-5. Results of MCR analysis of pure myristyl myristate. **A** Comparison of two MCR-extracted spectra from (a, b) pure myristyl myristate solid (MCR_MM1 and 2), and (c, d) *Euglena* (lipid 1 and 2). **B** Molecular distribution images of extracted components in pure myristyl myristate after MCR

set of spectra (MCR_MM1 and lipid 1) in which three defined bands at 1461 cm^{-1}, 1440 cm^{-1}, and 1417 cm^{-1} were observed corresponding to β′ polymorph. In fact, 1417 cm^{-1} band is associated with splitting of the Raman active methylene scissoring mode in β′ form (40-42). In the second set,

MCR_MM2 and lipid 2, intense bands at 1417 cm^{-1}, 1296 cm^{-1}, and 1130 cm^{-1} that are characteristics of all-trans conformation of carbon chains in crystalline domains were observed (43). In addition, 890 cm^{-1} band corresponding to terminal C–C-stretching vibration was also prominent. However, it is interesting to note that C=O-stretching (1732 cm^{-1}) and C–H-bending (1440 cm^{-1}) vibrations in both spectra were absent. Since intensities of Raman bands depend on both crystal orientation and incident polarization, it is possible that these bands are weak in this particular sample due to crystal orientation. However, it may also be due to the presence of two polymorphs of MM in *Euglena* cells. This may have serious implications as physical properties like molecular packing and freezing point, which are crucial for MM's efficient storage and eventual application as a bio jet fuel, will be different for different polymorphs. Polarized-Raman spectroscopic measurements should be performed to obtain further insights to make clear distinction between polymorphs.

We then constructed molecular distribution images of MCR-extracted components which revealed heterogeneous pattern without much resemblance to each other (Fig. 2-5B). This result further confirms the presence of two different forms in the standard MM sample.

Time-resolved MCR component images of *Euglena* cells

Once the assignment of all MCR-extracted spectral components was accomplished, we constructed time-resolved Raman images to visualize intracellular biomolecular distribution (Fig. 2-6). First, let us look into baseline (Fig. 2-6a). Although there is no difference at early culture times, significant increase in localized areas was observed in cells from 24 h. On the other hand, varying degrees of fluorescence background could be observed in cells at any given time (Fig. 2-6b). Protein synthesis seems to be active as its intracellular abundance increases and gets more or less evenly distributed throughout the cells as culture time progresses (Fig. 2-6c). However, irrespective of time, carotenoids were randomly distributed indicating cellular individuality (Fig. 2-6d). Details on wax ester fermentation, which is our main target, can be visualized in Fig. 2-6e–g. Paramylon accumulated under aerobic condition during pre-culture seems to decrease with time under anaerobic condition (Fig.2- 6e). Complementarily, abundance of myristyl myristate (MM1 and MM2), which were not present to begin with at 0 h, slowly starts increasing with passing culture time (Fig. 2-6f, g). Strong accumulation of wax esters in a localized fashion can be observed starting from 24 h. Interestingly, MM1 distribution (from MCR analysis) is similar to the abundance images obtained using univariate method (Fig. 2-1e). However, it is also important to note that MCR analysis led to identification of MM2, whose intracellular distribution pattern is quite different from MM1, reiterating the existence of two forms of myristyl myristate. Further clarification of these two forms is left for future studies.

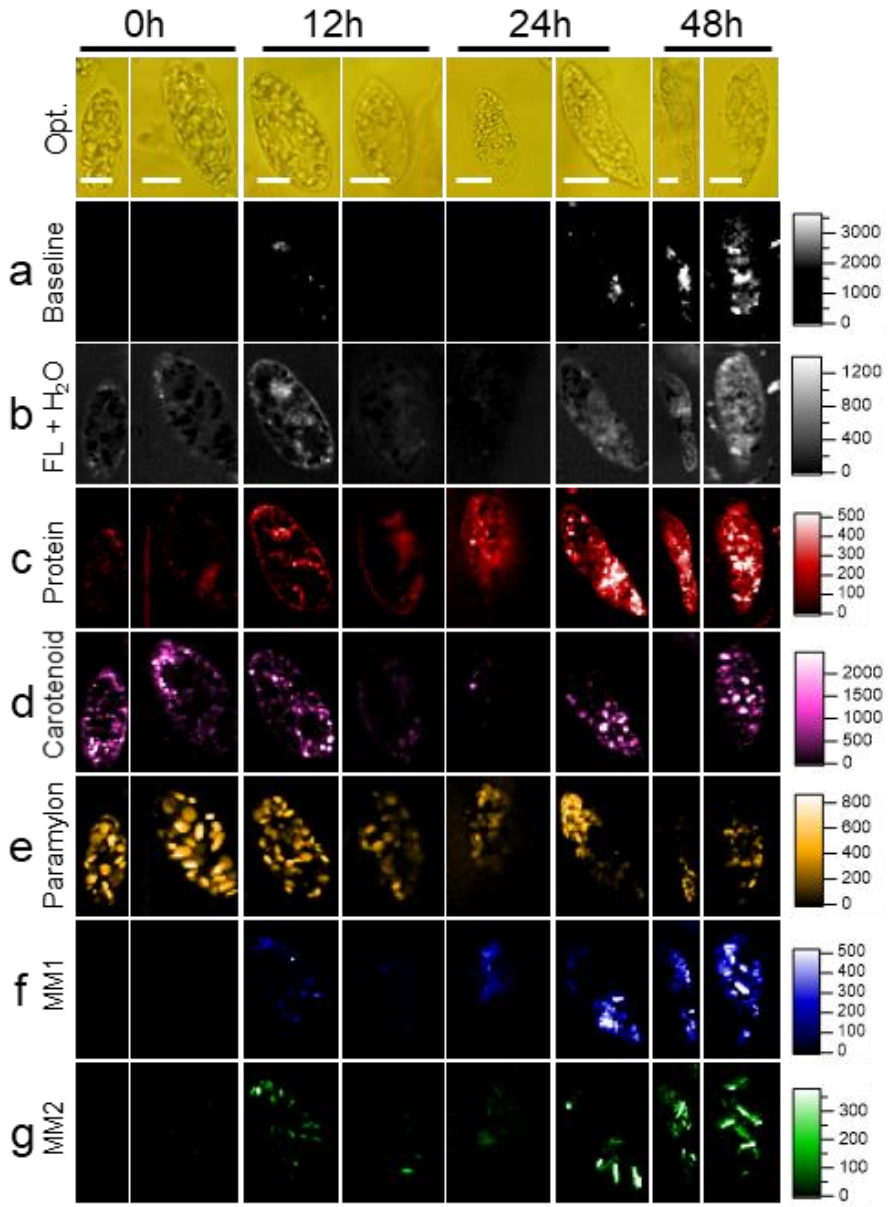


Fig. 2-6. Raman images constructed from MCR analysis. (a) Baseline, (b) fluorescence background and water, (c) protein, (d) carotenoid, (e) paramylon, and (f, g) myristyl myristate 1 and 2, respectively. Corresponding optical images are included (Opt.). Scale bar measures 10 μm. Color scale in Raman images indicates molecular abundance.

II - iv Summary

We have demonstrated the unique ability of Raman microscopy coupled with MCR analysis to investigate wax ester fermentation and obtain carbon chain length-specific information in single living *Euglena* cells. In the present study, conversion of aerobically accumulated paramylon to MM, a C28 wax ester (C14:0–C14:0), has been successfully visualized. Interestingly, two polymorphic forms of MM with different distribution patterns may have been separated during MCR analysis for the first time in *Euglena* cells. Even though this work focused on specifically identifying MM, we believe that this method can be applied to characterize other metabolites in many different cell types, including but not limited to humans, animals, plants, etc. Moreover, this approach is directly applicable to mutant strains or under other culture conditions. Therefore, our approach is expected to further our understanding of lipid metabolism in *Euglena* and its regulatory apparatus at the cellular level to realize microalgae as an economically viable biofuel feedstock. Moreover, it is clear from the present example that simple univariate analysis, though useful to some extent, is limited by overlapping contributions and that multivariate approach is absolutely necessary to study complex samples of biological origin.

Chapter III.

Identification of Molecular Basis for Objective Discrimination of Breast Cancer Cells (MCF-7) from Normal Human Mammary Epithelial Cells by Raman Microspectroscopy and Multivariate Curve Resolution Analysis

III - i Introduction

Despite advances in prognosis and treatment, cancer incidence and mortality are rapidly increasing around the world. According to the estimates by the International Agency for Research on Cancer in 2018, there were about 18.1 million new cases and 9.6 million cancer deaths. Among women, breast cancer with 2.1 million cases and over 0.6 million deaths tops the chart (44). Breast cancer can be diagnosed through multiple tests including an X-ray mammogram, ultrasound imaging, magnetic resonance imaging, fine needle aspiration cytology, and tissue biopsy etc. Presently, histopathology remains to be a gold standard in breast cancer diagnosis and treatment. However, identification of molecular signatures using this invasive procedure is expensive, involves tedious sample preparation, is time consuming and sometimes leads to ambiguous results due to human interpretations. Thus, it has severe limitations especially during surgeries. Therefore, it is necessary to develop alternative methods that are low- or non-invasive and economical while achieving rapid diagnosis with high accuracy.

Raman spectroscopy (RS), a powerful vibrational spectroscopic technique based on inelastic scattering of light, has been proposed to be a good alternative to overcome such difficulties. Advantages of RS are manifold: (1) non-invasive, i.e., suitability to *in vivo* applications, (2) no need for staining or genetic manipulation, (3) high sensitivity and specificity due to rich molecular information. Indeed, RS has been gaining much attention and has been successfully applied in disease prognosis and diagnosis (9, 45, 46), discriminate cells and tissues (47, 48), image living cells in a label-free manner (12, 49) and probe metabolic pathways (7). However, there are limitations to RS as well. First, traditional raster scanning methods employed in RS are extremely slow procedures, especially when considering the size of the tissues examined during histopathology. To solve this problem, researchers have proposed various methods, such as hand-held Raman probes for guided biopsy (50) and autofluorescence combined with selective Raman sampling (51), etc. Second, since RS measures molecular vibrations, different molecules containing similar chemical bonds show similar frequencies and, in most cases, it is not appropriate to simply use a single band for spectral interpretation. To make matters worse, Raman hyperspectral imaging results in a large volume of data with thousands of Raman spectra to handle. Therefore, we need to employ multivariate analyses (MA) for meaningful interpretation. To this end, a variety of multivariate analytical methods have been developed. Some of the most popular unsupervised multivariate classification methods applied to Raman spectroscopic data include singular value decomposition (SVD), principal components analysis (PCA), and cluster analysis, etc., which are suitable for exploratory analysis. On the other hand, if *a priori* information about the samples is available, supervised methods such as linear discriminant analysis (LDA), neural networks, and support vector machine (SVM) etc., are well suited to model the given Raman hyperspectral data and apply it to predict unknown samples.

It is surprising to note that some of the early studies demonstrating the potential application

of RS to cancers was done by Manfait and co-workers as early as 1982 (52, 53). This was followed by several studies especially focusing on RS-based breast cancer diagnosis in the early nineties (54, 55). Due to technological advancements and the development of chemometrics in the last two decades, the volume of RS-based literature has kept growing rapidly. To put things into perspective, a simple 'Topic' search with a keyword 'Raman AND Cancer' in the Web of Science database returned ~5000 documents. Even though application of RS has been proven to be successful under laboratory situations, it is important to understand that these MA results are subjective to many factors, including design of experiment and analysis, data pre-processing and overall quality of data. Therefore, experience of the person, instrument performance and acquisition parameters also play a crucial role. Owing to limitations in standardizing the whole procedure, universal adoption of RS in clinics has still not been achieved. Another major drawback is that none of the above-mentioned MA procedures discriminate/classify/predict based on inherent chemical information but strictly treat Raman spectroscopic data only mathematically. Therefore, to overcome these limitations, we employed an alternative approach called multivariate curve resolution-alternating least squares (MCR-ALS) in which pure chemical components and their abundances are extracted from Raman hyperspectral data to establish a molecular basis for reliable diagnosis. In this study, we identified for the first time that linoleate rich triglycerides serve as the marker for objective discrimination of MCF-7 and HMEpC cells in 632.8 nm excited chemometrics assisted Raman microspectroscopy.

III - ii Material and methods

Cell Culture

MCF-7 malignant breast cancer cell line was cultured in DMEM low glucose without phenol red (Thermo Fisher Scientific, Tokyo, Japan) with added supplements (0.1 mM sodium pyruvate, 2 mM L-Glutamine, 1% (v/v) antibiotics and 5% (v/v) fetal bovine serum). HMEpC primary cells obtained from normal mammary glands (Cell Applications, Inc., San Diego, CA, USA) as control were cultured in Human Mammary Epithelial Cell Media (TOYOBO Life Science, Osaka, Japan). Both MCF-7 and HMEpC cells were incubated at 37 °C and 5% CO_2. Cells were sub-cultured at ~80% of cell confluence and Raman spectra were obtained from cells incubated for 3 days after gently washing with PBS (-) on Poly-L-Lysine-coated glass bottom dish.

Raman Microspectroscopy

Raman spectra were measured using a homemade confocal Raman micro-spectrometer (38, 56). An excitation part consists of He-Ne laser (632.8 nm) coupled to an inverted microscope (IX70, Olympus) with an oil immersion objective lens (100×, NA = 1.3) to focus the excitation laser on specific points of cultured cells. Stokes Raman scattered light was collected using the same objective lens in back scattering geometry using a long pass filter. To improve axial resolution, a confocal pinhole of 50 μm was used in collection path. A polychromator (Chromex, 250IS) dispersed the

scattered light and was detected with a CCD device (Princeton instruments, Spec-10) cooled at −120 °C with liquid nitrogen. To achieve optimal throughput while measuring the whole finger print region, we used a 600 g/mm grating and set the slit width of polychromator to 50 μm. All Raman measurements were done at room temperature (22 °C) and the laser power was set to 4 mW at the sample position. Raman spectra were obtained from 5 random points in each cell with an exposure of 30 s/point. A total of 60 cells (30 cells for each kind) were measured and averaged. For lipid standards, several unsaturated fatty acids including docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), arachidonic acid (AA), γ-linolenic acid (GLA), α-linolenic acid (ALA), linoleic acid (LA), oleic acid (OA), and palmitoleic acid (PMA) purchased from Tokyo Chemical Industry Co., Ltd, Tokyo, Japan. were measured under the same conditions.

Data Analysis

Data pre-processing such as dark light subtraction, cosmic ray removal, and data de-noising by SVD were performed by IGOR Pro (Wavemetrics, Portland, OR, USA). Generally, no Raman bands are expected in the so-called silent region between ~2800 cm^{-1}–1800 cm^{-1}. Therefore, a preliminary analysis of Raman spectra in the whole fingerprint region between ~1800 cm^{-1}–370 cm^{-1} was carried out. Since no significant Raman band was observed except for strong contribution from background, the fingerprint region between 739 ~ 1800 cm^{-1} was chosen for multivariate analysis.

Discriminant Analysis

The first PCA was performed on mean-centered data using NIPALS algorithm with random cross validation to extract principal components (PC). Using prior knowledge of principal components, an LDA model for two classes was constructed by including the first 4 PC scores assuming equal prior possibilities. Furthermore, to construct an SVM model, nu-SVM with linear kernel type was employed with 10-fold cross validation. PCA, LDA and SVM were performed using Unscrambler (Camo Analytics, Oslo, Norway).

MCR-ALS

Parameter k represents the number of spectral components and was set to 7 in this study based on SVD analysis (31). First, 7 SVD components were used as initial points for further analysis. To obtain sparser solutions, additional L1 penalty term (lasso regression) of $\alpha^2 = 0.005$ and L2 penalty term (ridge regression) of $\beta^2 = 0.005$ were applied respectively. MCR-ALS was performed using a homemade program specifically developed for Raman spectroscopic applications using Python .

III - iii Results and discussion

Univariate Analysis of Normal and MCF-7 Cells Gives Little Information for Objective Discrimination

Average Raman spectra of 30 cells each of normal human mammary epithelial cells (HMEpC) and breast cancer cells (MCF-7) are presented in Figure 3-1A. Some of the prominent bands observed in both spectra such as 1657 cm^{-1} (amide I/-C=C- str), 1446 cm^{-1} (CH_2/CH_3), 1300 cm^{-1} (CH_2 twisting), 1263 cm^{-1} (=C-H), and 1003 cm^{-1} (Phenyl alanine) indicate the contribution of proteins and lipids. Raman bands at 879 cm^{-1} and 786 cm^{-1} observed in cancer cells can be assigned to C-C stretch (protein, amino acid hydroxyproline and lipids) and O-P-O symmetric stretch (nucleic acids), respectively (4, 57). Since simple comparison only suggests general variation in proteins and nucleic acids, we integrated intensities of important Raman bands, calculated ratios of various biomacromolecules for each cell, and their averages along with standard deviations (S.D.) were used to identify markers for discrimination as shown in Figure 3-1B. Some ratios such as nucleic acid/lipid (Figure 3-1B(c)), protein/lipid (Figure 3-1B(d)) and C-C str/lipid (Figure 3-1B(e)) show significant differences between normal and cancer cells. However, it is important to note that these are calculated by univariate approach (using one representative band/species) and it is impossible to avoid band overlaps from other components in the same region. For example, band around ~1440 cm^{-1} has been traditionally used as a lipid marker but it originally represents CH_2 and CH_3 vibrations, which inevitably contains contributions from most other biomolecules. Therefore, instead of single band analysis, there is a need for multivariate methods that consider the whole spectrum for reliable diagnosis.

Application of Multivariate Statistical Methods to Discriminate Cancer Cells

To develop Raman spectroscopy as a diagnostic tool, it is imperative to detect subtle biochemical changes in disease conditions by employing multivariate statistics. In order to identify spectral differences and discriminate normal/cancer cells, we averaged only those spectra obtained

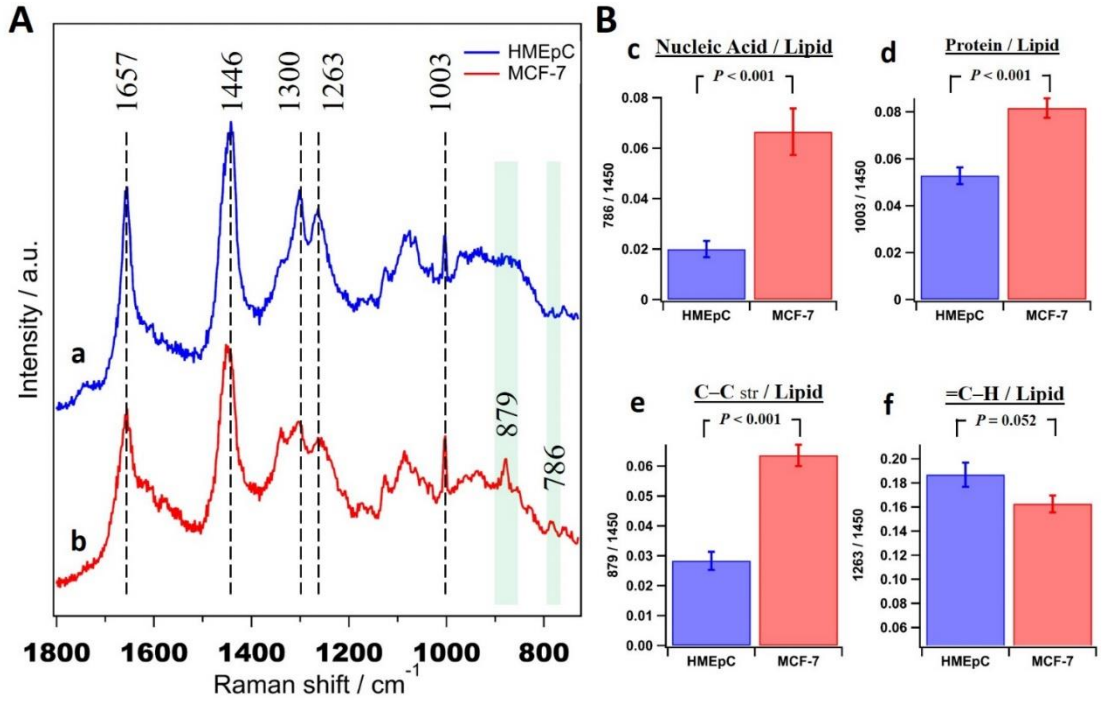


Figure 3-1. Comparison of average Raman spectra of MCF-7 and HMEpC. **(A)** Averaged Raman spectra (±S.D.) obtained from 30 cells of **(a)** HMEpC and **(b)** MCF-7, respectively. The consistent band positions were shown with broken lines and significant differences were highlighted by shaded bars. **(B)** Biomolecular ratios of **(c)** nucleic acid/lipid, **(d)** protein/lipid, **(e)** C–C str/lipid (chain lengths) and **(f)** =C–H/lipid (unsaturation).

from five different points in a cell and retained Raman spectrum representative of each individual cell for further analysis (60 spectra in total).

Principal Components Analysis

PCA essentially reduces the dimensionality of hyperspectral data to a few principal components (PC) without losing much information. Indeed, it is one of the oldest and widely used multivariate methods in data analysis and has previously been applied to Raman spectroscopic data from cancer cells and tissues. Results of PCA showed a good degree of classification of the two groups of cells. PCA identified 7 PCs. The first four components that contribute 88% are presented in Figure 3-2. PC scores indicate PC1 to be the main contributor (64%) as it essentially can classify efficiently when taken with any of the next three PCs (Figure 3-2B). A closer look into loadings (Figure 3-2A) reveals that PC1 spectrum is dominated by bands of lipid origin such as 1657 cm^{-1}, 1440 cm^{-1}, 1300 cm^{-1}, and 1263 cm^{-1}. In addition to these bands in PC1, PC2 showed markers of protein (1003 cm^{-1})

and nucleic acids (782 cm^{-1} and 1576 cm^{-1}). Nucleic acid marker band at 1576 cm^{-1}, which was not clearly observed in the average spectra, can be seen in PCA. Although PC loadings may provide molecular information to some degree, it is important to note that all of them show both positive and negative features. Moreover, most of the bands are mixed and are observed in multiple loadings, making it wrong to interpret the data in a physically meaningful way.

Linear Discriminant Analysis

In order to further the analysis, we used PC classifiers and constructed a discrimination model based on LDA. The discrimination plot of LDA presented in Figure 3-3 shows good separation of normal and cancer cells. Results are tabulated in a confusion matrix in Table 3-1. Constructed model achieved 98% discrimination accuracy with 96% sensitivity and 100% specificity.

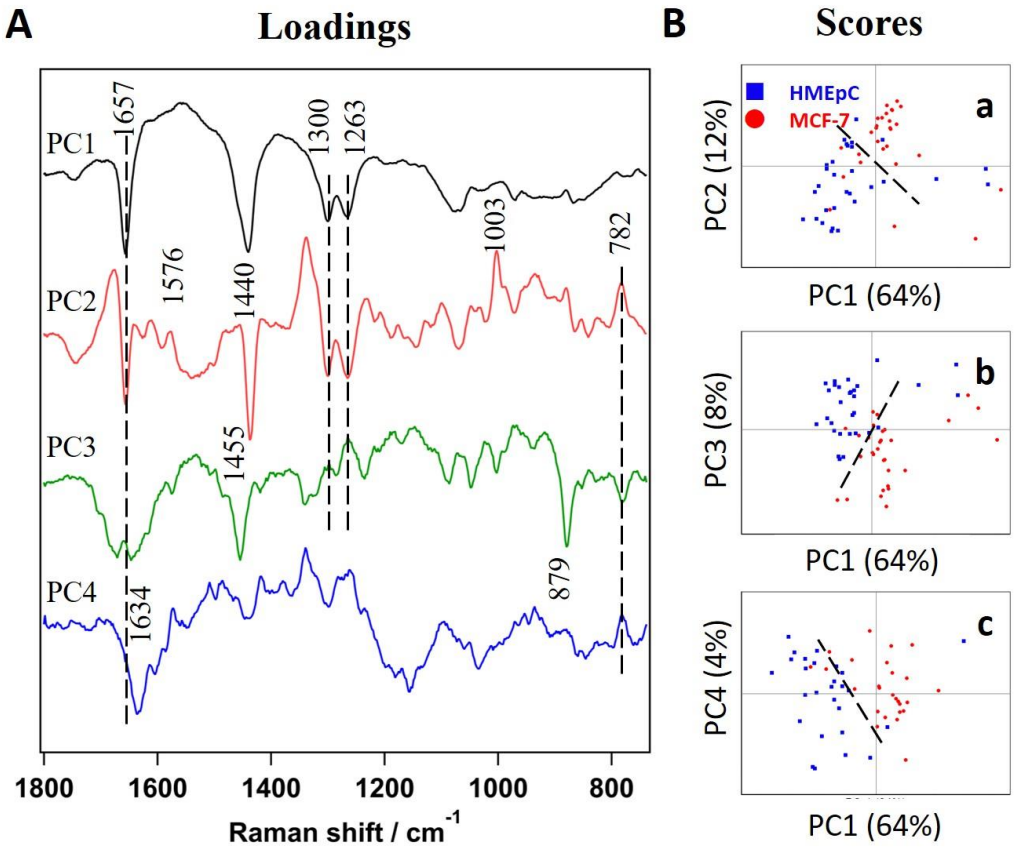


Figure 3-2. Results of principal components analysis (PCA) analysis. **(A)** First 4 principal components (PC) loadings, PC1 (64 %), PC2 (17%), PC3 (8%), and PC4 (4%). Broken lines show same band positions regardless of positive or negative tendency. **(B)** Scores plots of **(a)** PC2, **(b)** PC3 and **(c)** PC4 vs. PC1, respectively. Broken lines are drawn as visual guides to discriminate HMEpC and MCF-7.

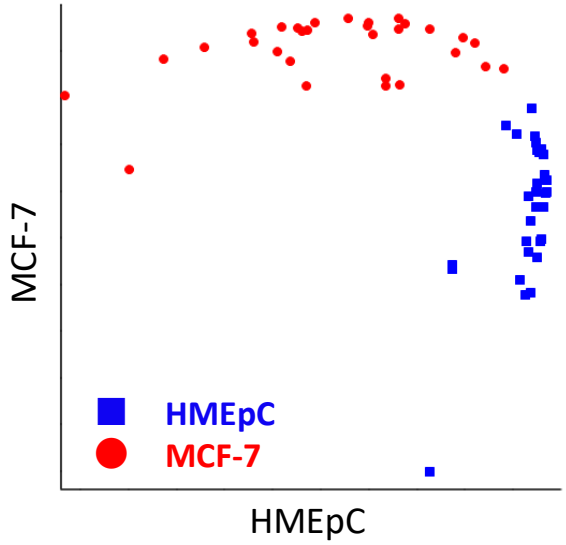


Figure 3-3. LDA Discrimination Plot. Linear discrimination score of normal HMEpC and breast cancer MCF-7 cells are plotted by blue boxes and red circles, respectively.

Table 3-1. Confusion matrix of linear discriminant analysis (LDA)

		Actual	
		HMEpC	**MCF-7**
Predicted	**HMEpC**	30	1
	MCF-7	0	29

Support Vector Machine Analysis

Unlike LDA in which data are expected to be normally distributed, SVM makes no assumptions to the data and has gained much popularity among machine learning methods. To further test the applicability of other supervised learning model, we employed the SVM algorithm and the resultant confusion matrix is given in Table 3-2. Indeed, the constructed SVM model with linear classification and 10-fold cross validation could achieve superior discrimination with training accuracy of 100% and validation accuracy of 98%, as shown in Table 3-2.

Table 3-2. Confusion matrix of SVM.

		Actual	
		HMEpC	**MCF-7**
Predicted	HMEpC	30	0
	MCF-7	0	30

Multivariate Curve Resolution Analysis

In order to understand molecular level differences and to establish a reasonable basis for successful discrimination by statistical methods such as LDA or SVM, we performed exploratory MCR-ALS analysis to obtain pure chemical components. Extracted spectral profiles of 7 components from the MCR-ALS model are presented in Figure 3-4A.

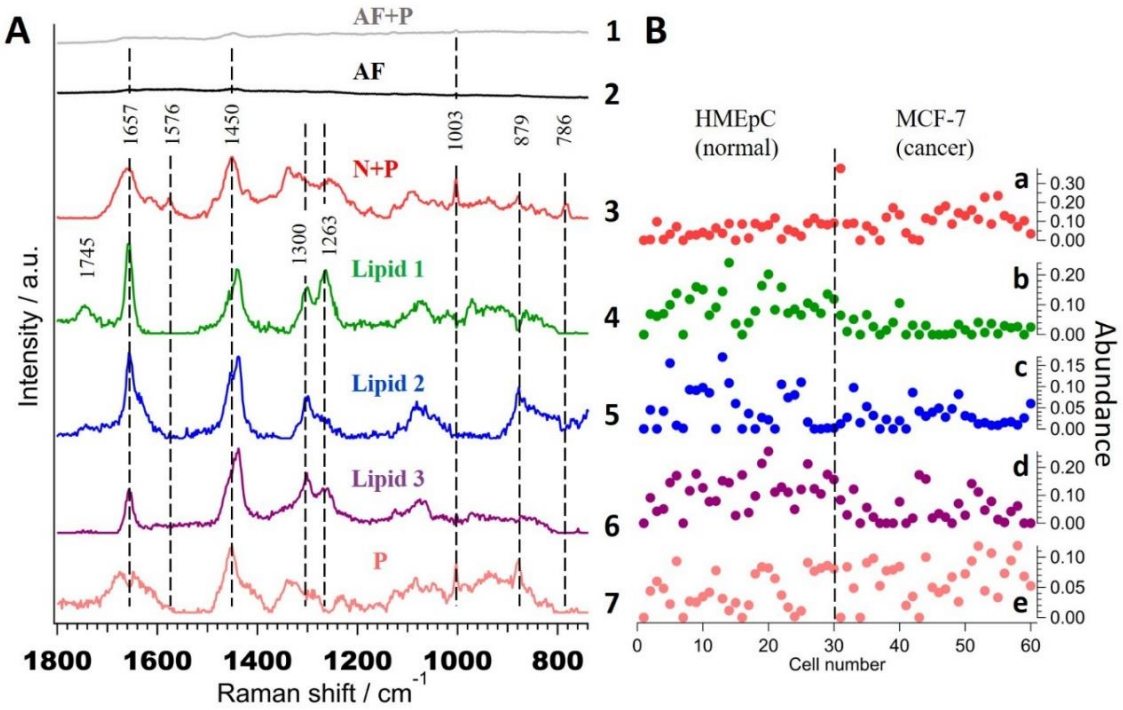


Figure 3-4. Seven components MCR-ALS analysis. **(A)**The extracted spectral components, **(1)** autofluorescence with protein [AF + P], **(2)** autofluorescence [AF], **(3)** nucleic acid with protein [N + P], **(4)** Lipid 1, **(5)** Lipid 2, **(6)** Lipid 3 and **(7)** Protein [P]. **(B)** Abundance profiles of **(a)** N + P, **(b-d)** lipid 1-3 and **(e)** protein, respectively. Broken line in B separates HMEpC and MCF-7 cells.

Unlike the results of PCA, these spectral profiles are meaningful as they correspond to pure molecular species or groups. Respective abundance profiles obtained from 'H' matrix of seven components MCR-ALS analysis, i.e., contribution of each component in single cells are shown in Figure 3-4B. Component 1 (Figure 3-4(A1)) with bands at 1003 cm^{-1}, 1450 cm^{-1}, and 1657 cm^{-1} with

broad background can be assigned to that part of autofluorescence which coexists with some proteins while component 2 (Figure 3-4(A2)) to com-monly observed autofluorescence background in Raman spectra of biological samples under this excitation conditions. Component 3 (Figure 3-4(A3)) containing bands typical to that of proteins at 879 cm^{-1}, 1003 cm^{-1}, 1657 cm^{-1} and nucleic acids at 786 cm^{-1} and 1576 cm^{-1} could be assigned to 'nucleic acid + protein' (denoted as 'N + P') that coexist together. Its abundance profile (Figure 3-4(Ba)) suggests slightly higher concentration in MCF-7 cells. Interestingly, components 4–6, which seem spectrally similar, were separated as independent components. Bands at 1263 cm^{-1}, 1300 cm^{-1}, 1440 cm^{-1}, and 1657 cm^{-1} indicate that these are lipids and hence named as 'Lipid 1′ (Figure 3-4(A4)), 'Lipid 2′ (Figure 3-4(A5)), and 'Lipid 3′ (Figure 3-4(A6)). Their abundance profiles indicate 'Lipid 1′ (Figure 3-4(Bb)) to be lower in MCF-7 compared to HMEpC cells whereas no significant difference can be observed in other two lipids (Figure 3-4(Bc,d)). Finally, component 7 (Figure 3-4(A7)) can be assigned to 'proteins' (denoted as 'P') alone, based on the spectral profile with no significant difference in their abundance.

Even though we get concentration information from MCR-ALS analysis, it should not be compared directly as it is not an absolute quantity. Therefore, it is safe to calculate average relative abundance of extracted components to understand meaningful trends. Figure 3-5A shows relative concentrations along with their standard error of three separated lipid components to 'N + P' (Figure 3-5(Aa–c)), to 'P' (Figure 3-5(Ad–f)) and to other lipids (Figure 3-5(Ag–i)). Of all nine ratios, four of them; 'Lipid 1′ to' N + P' or 'P' (Figure 3-5(Aa–d)) and 'Lipid 3′ to 'N + P' or 'P' (Figure 3-5(Ac–f)) seem to have statistically significant differences. Further to perform objective discrimination based on obtained pure molecular information, we constructed scatter plots to visualize all nine combinations in a similar fashion (Figure 3-5B). Although several of them seem to show a fair degree of separation (as indicated by broken lines in Figure 3-5(Bj,m,p and q)), considering statistical averages, we could conclude that ratios involving 'Lipid 1′ to other biomacromolecules such as nucleic acids and proteins serve as reliable "Raman spectral marker" for discriminating cancer from normal cells. Moreover, it is important to note that though scatter plots show lower discrimination than some of the other chemometric methods, this disadvantage is overcome by the advantage of the physically meaningful spectra.

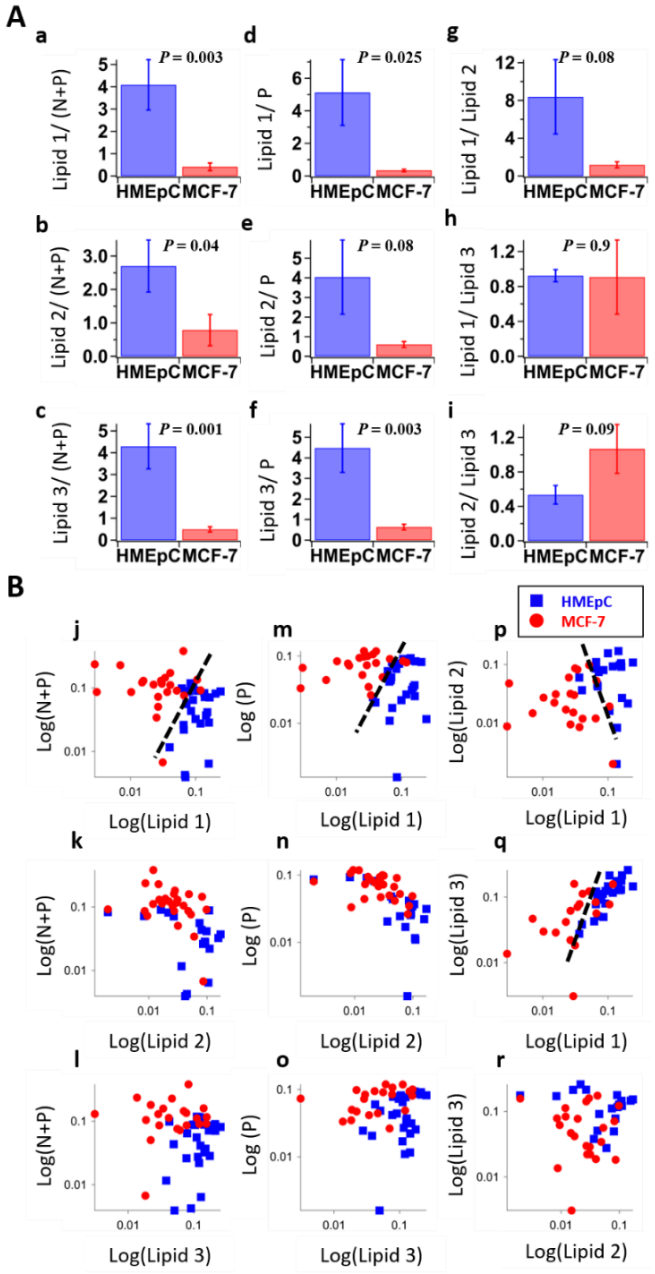


Figure 3-5. Discrimination analysis by MCR-ALS. **(A)** Relative abundance of MCR-ALS extracted components, **(a)** Lipid 1/ (N + P), **(b)** Lipid 2/ (N + P), **(c)** Lipid 3/ (N + P), **(d)** Lipid 1/P, **(e)** Lipid 2/P, **(f)** Lipid 3/P, **(g)** Lipid 1/Lipid 2, **(h)** Lipid 1/Lipid 3, **(i)** Lipid 2/Lipid 3. N + P: nucleic acid with protein, P: protein. Error bars are standard error of mean. p values obtained by t-test were denoted on top of histograms. **(B)** Scatter plots of each logarithmic abundance, **(j)** Lipid 1 vs. (N + P), **(k)** Lipid 2 vs. (N + P), **(l)** Lipid 3 vs. (N + P), **(m)** Lipid 1 vs. P, **(n)** Lipid 2 vs. P, **(o)** Lipid 3 vs. P, **(p)** Lipid 1 vs. Lipid 2, **(q)** Lipid 1 vs. Lipid 3, **(r)** Lipid 3 vs. Lipid 2. Some labels of measured cells were omitted in those plots since the values of abundance were calculated into zero by MCR-ALS. Broken lines serve as visual guides to separate two groups of cells.

Molecular Assignment of MCR-ALS Extracted Lipid Components

Now that we have identified sensitive lipid spectral markers, it is necessary to assign these components at the molecular level to develop an objective method to discriminate cancer cells from normal ones. To begin with, all three lipid components (Figure 3-4A(4,5,6)) show bands at 1657 cm^{-1} and 1263 cm^{-1} corresponding to -C=C- stretching and =C-H modes, respectively. Therefore, we can safely say that none of the three components are saturated lipids. In order to screen for potential candidates, we measured a series of standard fatty acids from palmitoleic acid with unsaturation index of 1 to docosahexaenoic acid with 6 double bonds to cover a wide range of polyunsaturated fatty acids (PUFA) as given in Figure 3-6A.

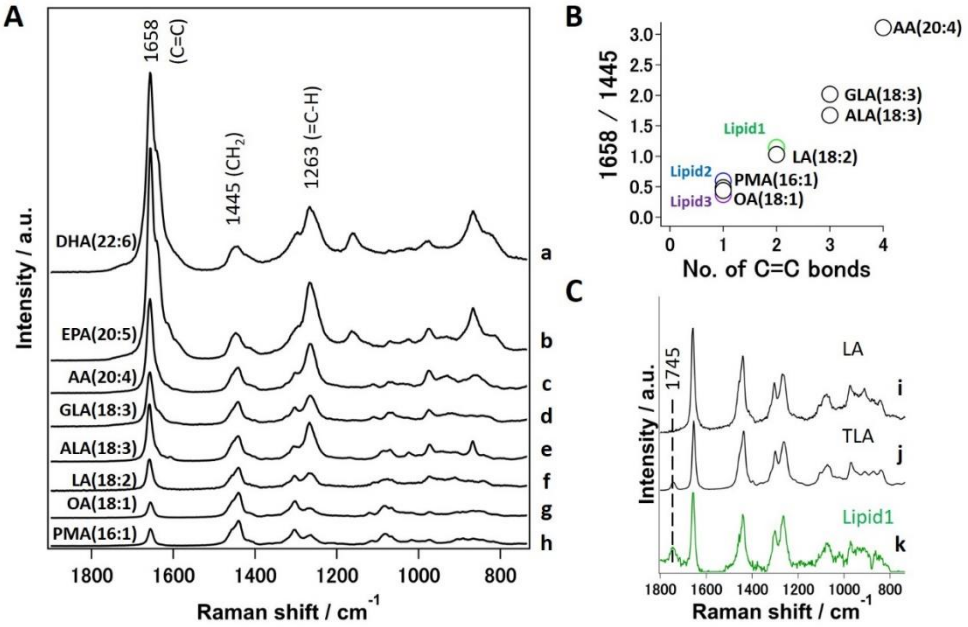


Figure 3-6. Molecular level assignment of Raman spectral marker. (A) Standard Raman spectra of various unsaturated fatty acids normalized using 1445 cm^{-1} band; (a) docosahexaenoic acid(DHA), (b) eicosapentaenoic acid (EPA), (c) arachidonic acid (AA), (d) γ-linolenic acid (GLA), (e) α-linolenic acid (ALA) (f) linoleic acid (LA), (g) oleic acid (OA) and (h) palmitoleic acid (PMA). (B) Unsaturation index plot. Relative intensity ratio of $C=C/CH_2$ vs. number of C=C bonds in standard unsaturated fatty acids. Relative intensities obtained from three lipids on MCR-ALS components are also plotted and denoted in color. (C) Comparison of standard (i) LA, (j) Trilinolein (TLA) and (k) Lipid 1 extracted by MCR-ALS.

For further comparison of fatty acid standards, we normalized these spectra with band area of 1445 cm^{-1}. We can observe that intensity of 1658 cm^{-1} greatly increases with an increasing number of double bonds. In fact, it is well known that Raman intensity of C=C stretching mode is directly proportional to the number of double bonds in the molecule. Therefore, it is rather straightforward to construct a calibration model to predict the unsaturation index from measured Raman spectra by calculating Raman intensity at 1658 cm^{-1} (C=C stretching vibration) to that at 1445

cm^{-1} (CH_2 deformation), i.e., 1658/1445 (4). Indeed, a linear relationship is observed when a ratio of 1658/1445 is plotted against number of double bonds in chemical structure of fatty acids as shown in Figure 3-6B. To predict the molecular structure of MCR-ALS extracted lipid components, we estimated the ratio of 1658/1440 in a similar fashion and compared with the constructed model as marked in Figure 3-6B. We could therefore assign 'Lipid 1' to di-unsaturated fatty acid (linoleic acid, LA) whereas 'Lipid 2' and 'Lipid 3' could be assigned to mono-unsaturated fats. It is important to note that 'Lipid 1' contains a Raman band at 1745 cm^{-1} corresponding to C=O stretch of esters. Since the focus is to mainly identify 'Lipid 1' in an unambiguous manner, we further measured trilinoleic acid (TLA), a triglyceride (TG) with three linoleic acid groups. A comparison reveals a perfect match between TLA (Figure 3-6C(j)) and 'Lipid 1' (Figure 3-6C(k)) as opposed to simple LA (Figure 3-6C(i)), in which 1745 cm^{-1} band is not observed as shown in Figure 3-6C. Therefore, we believe the relative content of TGs with high LA content is the main factor that helped to discriminate normal (HMEpC) and cancer (MCF-7) cells. Although 'Lipid 2' and 'Lipid 3' have been identified as mono-unsaturated fats, further unambiguous assignment to the likes of Oleic acid (18:1) or palmitoleic acid (16:1) could not be achieved in this study.

The choice of breast cancer cell line for this work (MCF-7) was established from invasive ductal carcinoma (IDC) of a Caucasian patient and the cells are known to be estrogen (ER) and progesterone receptor (PgR)-positive. It is important to note that ER and PgR-positive IDC is the most common subtype accounting for >70% of breast cancers (58). Therefore, analysis of such a cancer cell line adds meaningful value to understanding Raman spectral markers. Indeed, many researchers have used MA such as PCA, LDA, and SVM for RS data of cancers for a long time and reported marked differences in proteins and fat profiles in general, which corroborates well with this study (59-63). Although these methods discriminated cancers well, as can also be seen from our own data, they do not give insights into the chemical changes responsible for diagnosis, thereby making it difficult to be translated to clinics. To overcome this, Haka et al. developed a method to model tissue spectra as linear combinations of known components and succeeded in discriminating cancers with some chemical information. Indeed, they showed that relatively low abundance of fats could be used as an indicator to distinguish breast cancer tissues (50, 64). However, such analysis has several assumptions and may overlook underlying pathology. Other researchers also reported decreased overall lipid content in human breast cell/biopsy samples compared to normal breast cells/tissues using RS but without molecular level information (54, 55, 65-67). Our results specifically showed that relative abundance of linoleate-rich glyceride to other biomacromolecules, such as nucleic acids and proteins, to be the major difference and possibly the reason for successful discrimination of breast cancer cells from normal epithelial cells. Interestingly, a previous attempt by Sixian et al. could not find strong correlation with PUFA and protein by Raman spectroscopy (68). We believe it was because they calculated the ratio considering all fats as a single entity. It is important to note from this study

that, although there are three groups of unsaturated fatty acids, only a linoleate-rich component could serve a reliable discrimination index.

Alterations in lipid metabolism have been shown to play a critical role in development, promotion, and maintenance of cancers (69, 70). Therefore, reprogramming of lipid metabolism is being considered a hallmark of malignancy and can be used as a novel target for anti-cancer strategy (71, 72). In particular, the role of unsaturated fatty acids including LA is of great importance as it is used for synthesizing arachidonic acid (AA). For example, cyclooxygenase (COX) enzymes convert AA to bioactive lipids such as prostaglandins (PG), which play key roles in adhesive, migratory, and invasive behavior of cells during development and progression of breast and other cancers (73-75). Therefore, we suspect from our results that a certain amount of AA could have been used up for the synthesis of PG, thereby depleting LA-rich TG in MCF-7 cells.

From the nature of the analysis used in this study, one might expect that several protein and/or saccharide components should also have been extracted. However, it is important to understand the limitations involved. Since we use spontaneous Raman microspectroscopy, one of the main limitations in detecting several more biomolecular components is their local intracellular concentrations. Limitation to resolve multiple components arises from the inherent nature of MCR-ALS with applied penalties (L1- and L2-norms). It is not possible to unmix two spectral components if there is no difference in their intracellular distribution pattern. Essentially, such components are treated as a single component. Considering the above limitations, it is understandable as to why weak or minor molecular components such as saccharides could not be detected/separated in this study.

III - iv Summary

In this study, we tried to address the age-old problem of efficiently extracting hidden information from chemically rich Raman hyperspectral data. In addition to demonstrating the utility of discrimination analysis such as LDA and SVM, we developed and employed MCR-ALS with non-negative constraints to extract physically meaningful Raman spectra using mammary epithelial cells and breast cancer cells as a model case. In a truly exploratory fashion, without *a priori* information, we obtained various biomolecular spectra including three individual lipid groups and successfully identified relative ratios of linoleate-rich glyceride as the Raman spectral marker and molecular basis for objective diagnosis of breast cancer. We would like to emphasize that this is the first report that discusses cancer pathology in detail while discriminating breast cancer cells unambiguously using specific fatty acid content in chemometrics-assisted RS. However, further studies are necessary to determine whether the differences in linoleate-rich triglycerides can be directly related to cancer states. Although both cell lines used in this study are of epithelial source, it is important to understand that most tumors are like organs and have more than one type of cell. Therefore, while the model holds true to this breast cancer cell line with 633 nm excitation, it is imperative that we further test on large

numbers of other cell lines and with different excitations wavelengths as well to have general consensus. Once established, spectral markers identified in the present study being at the cellular level have the potential to be used as an adjunct or even an alternative to cytological diagnosis, especially because specimens for cytology have scattered cells in them that are appropriate for RS. Moreover, RS can be performed on any biological sample including cells, tissues and body fluids etc. We believe such an approach when further developed can be adopted to real clinical applications for rapid yet objective diagnosis of certain types of cancers.

Chapter IV.

General conclusion

In this thesis, Raman spectroscopy (RS) coupled with multivariate analysis, especially MCR-ALS was utilized to study in two biomedical systems.

In a study described in Chapter II, we succeeded in visualizing molecular-specific information in *Euglena* during wax ester fermentation by Raman micro-spectroscopy. It is obvious from our results that simple univariate approach is insufficient and that MCR-ALS is crucial to extract hidden information from Raman spectra. Even though we have not measured any mutants in this study, our approach is directly applicable to other systems and is expected to deepen the knowledge on lipid metabolism in microalgae, which eventually leads to new strategies that will help to enhance biofuel production efficiency in the future.

In a study described in Chapter III, we applied MCR-ALS in medical diagnosis to help understanding molecular basis. Human mammary epithelial cells (HMEpC) and breast cancer cells (MCF-7) were measured by RS and to estimate diagnostic accuracy by PC-LDA and SVM categorized supervised multivariate analysis (MA). While those MA gave fairly high accuracy, the molecular basis was still unclear. Furthermore, to elucidate the underlying molecular changes in cancer cells, MCR-ALS was applied to extract physically meaningful spectra from complex cellular data. Relative abundance of linoleate rich lipid component seems to be strictly regulated between the two groups of cells. This study successfully identified Raman spectral markers and demonstrated the potential of RS to become an excellent cytodiagnostic tool that can both accurately and objectively discriminates breast cancer from normal cells.

These results from two practical studies successfully show that Raman spectroscopy coupled with MCR-ALS is strong analytical tool to extract biomolecular information. Although this technique presently needs deep consideration of what is extracted, this problem will be solved combining such as standard Raman spectrum library and user-friendly interface in future. I believe Raman spectroscopy coupled with MCR-ALS technique probably endows researches opportunity for reading "molecular finger print" as it is.

References

1. Raman CV, Krishnan KS. A New Type of Secondary Radiation. Nature. 1928;121(3048):501-2.

2. Krimm S, Bandekar J. Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. Advances in protein chemistry. 1986;38:181-364.

3. Carey PR. Biochemical applications of Raman and resonance Raman spectroscopies. New York: Academic Press; 1982.

4. Czamara K, Majzner K, Pacia MZ, Kochan K, Kaczor A, Baranska M. Raman spectroscopy of lipids: a review. Journal of Raman Spectroscopy. 2015;46(1):4-20.

5. Puppels GJ, Mul FFMd, Otto C, Greve J, Robert-Nicoud M, Arndt-Jovin DJ, et al. Studying single living cells and chromosomes by confocal Raman microspectroscopy. Nature. 1990;347(6290):301-3.

6. Yu-San H, Takeshi K, Masayuki Y, Hiro-o H. Molecular-Level Investigation of the Structure, Transformation, and Bioactivity of Single Living Fission Yeast Cells by Time- and Space-Resolved Raman Spectroscopy. Biochemistry. 2005;44(30):10009-19.

7. Laor D, Sade D, Shaham-Niv S, Zaguri D, Gartner M, Basavalingappa V, et al. Fibril formation and therapeutic targeting of amyloid-like structures in a yeast model of adenine accumulation. Nature communications. 2019;10(1):62.

8. Chuan-Keng H, Masahiro A, Hiro-o H, Shinsuke S. Disentangling Dynamic Changes of Multiple Cellular Components during the Yeast Cell Cycle by in Vivo Multivariate Raman Imaging. Analytical Chemistry. 2012;84(13):5661-8.

9. Iwasaki K, Noothalapati H, Yamamoto T. Chapter 15 - Recent advances in Raman spectroscopy of proteins for disease diagnosis. In: Ozaki Y, Baranska M, Lednev IK, Wood BR, editors. Vibrational Spectroscopy in Protein Research: Academic Press; 2020. p. 435-59.

10. Hideyuki S, Kimie A, Wataru K, Yukihiro O. Multivariate data analysis for Raman spectroscopic imaging. Journal of Raman Spectroscopy. 2009;40(12):1720-5.

11. Ando M, Hamaguchi H-o. Molecular component distribution imaging of living cells by multivariate curve resolution analysis of space-resolved Raman spectra. Journal of biomedical optics. 2014;19(1).

12. Noothalapati H, Sasaki T, Kaino T, Kawamukai M, Ando M, Hamaguchi HO, et al. Label-free Chemical Imaging of Fungal Spore Walls by Raman Microscopy and Multivariate Curve Resolution Analysis. Sci Rep. 2016;6:27789.

13. Chen P-HH, Shimada R, Yabumoto S, Okajima H, Ando M, Chang C-TT, et al. Automatic and objective oral cancer diagnosis by Raman spectroscopic detection of keratin with multivariate curve resolution analysis. Scientific reports. 2016;6:20097.

14. BP Statistical Review of World Energy - 2015. BP p.l.c.; 2016 June 2016.
15. Global Energy & CO2 Status Report - 2017. International Energy Agency; 2018 march 2018.
16. Council NR. Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use. Washington, DC: The National Academies Press; 2010.
17. Allred BW, Smith WK, Twidwell D, Haggerty JH, Running SW, Naugle DE, et al. Sustainability. Ecosystem services lost to oil and gas in North America. Science. 2015;348(6233):401-2.
18. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, et al. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant J. 2008;54(4):621-39.
19. Malcata FX. Microalgae and biofuels: a promising partnership? Trends Biotechnol. 2011;29(11):542-9.
20. Klopfenstein WE. Effect of molecular weights of fatty acid esters on cetane numbers as diesel fuels. Journal of the American Oil Chemists Society. 1985;62(6):1029-31.
21. Inui H, Ishikawa T, Tamoi M. Wax Ester Fermentation and Its Application for Biofuel Production. Adv Exp Med Biol. 2017;979:269-83.
22. Alam F, Date A, Rasjidin R, Mobin S, Moria H, Baqui A. Biofuel from Algae- Is It a Viable Alternative? Procedia Engineering. 2012;49:221-7.
23. Ogawa T, Tamoi M, Kimura A, Mine A, Sakuyama H, Yoshida E, et al. Enhancement of photosynthetic capacity in Euglena gracilis by expression of cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphatase leads to increases in biomass and wax ester production. Biotechnol Biofuels. 2015;8:80.
24. Radakovits R, Jinkerson RE, Darzins A, Posewitz MC. Genetic engineering of algae for enhanced biofuel production. Eukaryot Cell. 2010;9(4):486-501.
25. Tomiyama T, Kurihara K, Ogawa T, Maruta T, Ogawa T, Ohta D, et al. Wax Ester Synthase/Diacylglycerol Acyltransferase Isoenzymes Play a Pivotal Role in Wax Ester Biosynthesis in Euglena gracilis. Sci Rep. 2017;7(1):13504.
26. Wang Y, Seppanen-Laakso T, Rischer H, Wiebe MG. Euglena gracilis growth and cell composition under different temperature, light and trophic conditions. PLoS One. 2018;13(4):e0195329.
27. Wakisaka Y, Suzuki Y, Iwata O, Nakashima A, Ito T, Hirose M, et al. Probing the metabolic heterogeneity of live Euglena gracilis with stimulated Raman scattering microscopy. Nat Microbiol. 2016;1(10):16124.
28. Kamila K, Huadong P, Bayden RW, Victoria SH. Single cell assessment of yeast metabolic engineering for enhanced lipid production using Raman and AFM-IR imaging.

Biotechnology for Biofuels. 2018;11(1):106.
29. Tomiyama T, Goto K, Tanaka Y, Maruta T, Ogawa T, Sawa Y, et al. A major isoform of mitochondrial trans-2-enoyl-CoA reductase is dispensable for wax ester production in Euglena gracilis under anaerobic conditions. PLoS One. 2019;14(1):e0210755.
30. Mitsuhiro K, Takahiro I. Suppression of DYRK ortholog expression affects wax ester fermentation in Euglena gracilis. Journal of Applied Phycology. 2018;30(1):367-73.
31. Noothalapati H, Iwasaki K, Yamamoto T. Biological and Medical Applications of Multivariate Curve Resolution Assisted Raman Spectroscopy. Anal Sci. 2017;33(1):15-22.
32. Corbett EC, Zichy V, Goral J, Passingham C. Fourier transform Raman studies of materials and compounds of biological importance—II. The effect of moisture on the molecular structure of the alpha and beta anomers of d-glucose. Spectrochimica Acta Part A: Molecular Spectroscopy. 1991;47(9–10):1399-411.
33. Cael JJ, Koenig JL, Blackwell J. Infrared and raman spectroscopy of carbohydrates : Part IV. Identification of configuration- and conformation-sensitive modes for D-glucose by normal coordinate analysis. Carbohydrate Research. 1974;32(1):79-91.
34. Beattie JR, Steven EJB, Bruce WM. A critical evaluation of Raman spectroscopy for the analysis of lipids: Fatty acid methyl esters. Lipids. 2004;39(5):407-19.
35. Snyder R, Hsu S, Krimm S. Vibrational spectra in the C-H stretching region and the structure of the polymethylene chain. Spectrochimica Acta Part A: Molecular Spectroscopy. 1978;34(4):395-406.
36. Robert CS, Ira WL. Raman spectra and vibrational assignments for dipalmitoyl phosphatidylcholine and structurally related molecules. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism. 1975;388(3):361-73.
37. Snyder RG, Schachtschneider JH. Vibrational analysis of the n-paraffins—I. Spectrochimica Acta. 1963;19(1):85-116.
38. Noothalapati H, Iwasaki K, Yoshimoto C, Yoshikiyo K, Nishikawa T, Ando M, et al. Imaging phospholipid conformational disorder and packing in giant multilamellar liposome by confocal Raman microspectroscopy. Spectrochimica acta Part A, Molecular and biomolecular spectroscopy. 2017;187:186-90.
39. Koritala S. Microbiological synthesis of wax esters by euglena gracilis. Journal of the American Oil Chemists Society. 1989;66(1):133-4.
40. Da Silva E, Bresson S, Rousseau D. Characterization of the three major polymorphic forms and liquid state of tristearin by Raman spectroscopy. Chem Phys Lipids. 2009;157(2):113-9.
41. Michiyo M, Masahiro A, Keisuke S, Ikuyo N, Koichi C, Katsuhiro A, et al. Simultaneous imaging of fat crystallinity and crystal polymorphic types by Raman

microspectroscopy. Food Chemistry. 2016;196(Food Chemistry 72 4 2001):411-7.
42. Michiyo M, Masahiro A, Keisuke S, Hiro OH. Differentiation of Animal Fats from Different Origins: Use of Polymorphic Features Detected by Raman Spectroscopy. Applied Spectroscopy. 2010;64(11):1244-50.
43. Ashok Zachariah S, Bo-Han L, Shih-Ting L, Masahiro A, Chien-Lung W, Hiro-o H. Estimating Percent Crystallinity of Polyethylene as a Function of Temperature by Raman Spectroscopy Multivariate Curve Resolution by Alternating Least Squares. Analytical Chemistry. 2017.
44. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
45. Sahu A, Gera P, Malik A, Nair S, Chaturvedi P, Murali Krishna C. Raman exfoliative cytology for prognosis prediction in oral cancers: A proof of concept study. Journal of biophotonics. 2019.
46. Hemanth N, Suguru U, Naoki O, Yoshikazu K, Masahiro A, Hiro-o H, et al. Towards the development of a non-bioptic diagnostic technique for eosinophilic esophagitis using Raman spectroscopy. Vibrational Spectroscopy. 2016;85:7-10.
47. Tolstik T, Marquardt C, Matthaus C, Bergner N, Bielecki C, Krafft C, et al. Discrimination and classification of liver cancer cells and proliferation states by Raman spectroscopic imaging. Analyst. 2014;139(22):6036-43.
48. Lloyd GR, Orr LE, Christie-Brown J, McCarthy K, Rose S, Thomas M, et al. Discrimination between benign, primary and secondary malignancies in lymph nodes from the head and neck utilising Raman spectroscopy and multivariate analysis. Analyst. 2013;138(14):3900-8.
49. Iwasaki K, Kaneko A, Tanaka Y, Ishikawa T, Noothalapati H, Yamamoto T. Visualizing wax ester fermentation in single Euglena gracilis cells by Raman microspectroscopy and multivariate curve resolution analysis. Biotechnology for Biofuels. 2019;12.
50. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Diagnosing breast cancer by using Raman spectroscopy. Proc Natl Acad Sci U S A. 2005;102(35).
51. Kong K, Rowlands CJ, Varma S, Perkins W, Leach IH, Koloydenko AA, et al. Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy. Proc Natl Acad Sci U S A. 2013;110(38).
52. Manfait M, Jeannesson P, Jardillier JC, Ginot L, Alix AJP. Raman-Spectroscopy of Cancer-Cells - a New Approach to the Study of the Drug-Cell Interactions. Annales De Biologie Clinique. 1982;40(4):394-.
53. Nabiev IR, Morjani H, Manfait M. Selective Analysis of Antitumor Drug-Interaction

with Living Cancer-Cells as Probed by Surface-Enhanced Raman-Spectroscopy. European Biophysics Journal. 1991;19(6):311-6.

54. Frank CJ, Redd DC, Gansler TS, McCreery RL. Characterization of human breast biopsy specimens with near-IR Raman spectroscopy. Anal Chem. 1994;66(3):319-26.

55. Frank CJ, McCreery RL, Redd DC. Raman spectroscopy of normal and diseased human breast tissues. Anal Chem. 1995;67(5):777-83.

56. Noothalapati H, Ikarashi R, Iwasaki K, Nishida T, Kaino T, Yoshikiyo K, et al. Studying anti-oxidative properties of inclusion complexes of alpha-lipoic acid with gamma-cyclodextrin in single living fission yeast by confocal Raman microspectroscopy. Spectrochim Acta A Mol Biomol Spectrosc. 2018;197:237-43.

57. Notingher I, Verrier S, Romanska H, Bishop AE, Polak JM, Hench LL. In situ Characterisation of Living Cells by Raman Spectroscopy. Journal of Spectroscopy. 2002;16(2):43-51.

58. Vici P, Pizzuti L, Natoli C, Gamucci T, Di Lauro L, Barba M, et al. Triple positive breast cancer: a distinct subtype? Cancer Treat Rev. 2015;41(2):69-76.

59. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. Cancer Res. 2002;62(18):5375-80.

60. Chowdary MV, Kumar KK, Kurien J, Mathew S, Krishna CM. Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy. Biopolymers. 2006;83(5):556-69.

61. Pichardo-Molina JL, Frausto-Reyes C, Barbosa-Garcia O, Huerta-Franco R, Gonzalez-Trujillo JL, Ramirez-Alvarado CA, et al. Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. Lasers Med Sci. 2007;22(4):229-36.

62. Lyng FM, Traynor D, Nguyen TNQNQ, Meade AD, Rakib F, Al-Saady R, et al. Discrimination of breast cancer from benign tumours using Raman spectroscopy. PloS one. 2019;14(2).

63. Hedegaard M, Krafft C, Ditzel HJ, Johansen LE, Hassing S, Popp J. Discriminating Isogenic Cancer Cells and Identifying Altered Unsaturated Fatty Acid Content as Associated with Metastasis Status, Using K-Means Clustering and Partial Least Squares-Discriminant Analysis of Raman Maps. Analytical Chemistry. 2010;82(7):2797-802.

64. Haka A, Volynskaya Z, Gardecki J, Nazemi J, Lyons J, Hicks D, et al. In vivo Margin Assessment during Partial Mastectomy Breast Surgery Using Raman Spectroscopy. Cancer Research. 2006;66(6).

65. Bitar RA, Martinho HaS, Tierra-Criollo CJ, Ramalho ZLN, Netto MM, Martin AA. Biochemical analysis of human breast tissues using Fourier-transform Raman spectroscopy. J

Biomed Opt. 2006;11(5).
66. Brozek-Pluska B, Musial J, Kordek R, Bailo E, Dieing T, Abramczyk H. Raman spectroscopy and imaging: applications in human breast cancer diagnosis. Analyst. 2012;137(16):3773-80.
67. Marro M, Nieva C, de Juan A, Sierra A. Unravelling the Metabolic Progression of Breast Cancer Cells to Bone Metastasis by Coupling Raman Spectroscopy and a Novel Use of Mcr-Als Algorithm. Anal Chem. 2018;90(9):5594-602.
68. Sixian Y, Haohua T, Youbo Z, Yuan L, Eric JC, Marina M, et al. Raman Spectroscopic Analysis Reveals Abnormal Fatty Acid Composition in Tumor Micro- and Macroenvironments in Human Breast and Rat Mammary Cancer. Scientific Reports. 2016;6(1).
69. Kuo CY, Ann DK. When fats commit crimes: fatty acid metabolism, cancer stemness and therapeutic resistance. Cancer Commun (Lond). 2018;38(1):47.
70. Yi M, Li J, Chen S, Cai J, Ban Y, Peng Q, et al. Emerging role of lipid metabolism alterations in Cancer stem cells. J Exp Clin Cancer Res. 2018;37(1):118.
71. Peck B, Schulze A. Lipid desaturation - the next step in targeting lipogenesis in cancer? FEBS J. 2016;283(15):2767-78.
72. Cheng C, Geng F, Cheng X, Guo D. Lipid metabolism reprogramming and its potential targets in cancer. Cancer Commun (Lond). 2018;38(1):27.
73. Wang D, Dubois RN. Prostaglandins and cancer. Gut. 2006;55(1):115-22.
74. Menter DG, Dubois RN. Prostaglandins in cancer cell adhesion, migration, and invasion. Int J Cell Biol. 2012;2012:723419.
75. Ma X, Yang Q, Wilson KT, Kundu N, Meltzer SJ, Fulton AM. Promoter methylation regulates cyclooxygenase expression in breast cancer. Breast Cancer Res. 2004;6(4):R316-21.

Acknowledgement

I would like to express my deep gratitude to Professor Tatsuyuki Yamamoto, the Faculty of Life and Environmental Science of Shimane University, for his advice and encouragement as a major supervisor. This thesis would have not completed without his kind help.

I am grateful to Assistant Professor Hemanth Noothalapati, the Faculty of Life and Environmental Science of Shimane University. I have grateful respect for his scientific prospection and motivation.

I would like to express my sincerest and profound gratitude to Associate Professor Keisuke Yoshikiyo, the Faculty of Life and Environmental Science of Shimane University, for his hearty help and encouragement throughout this study as a vice supervisor.

I am grateful to Professor Masaaki Azuma, the Faculty of Agriculture of Tottori University, for his relevant remark and guidance as a vice supervisor.

I'm also grateful to all members of our laboratory and my family for their help.

Keita Iwasaki

Raman spectroscopic studies coupled with MCR-ALS applied on some biomedical systems

Abstract

Raman spectroscopy has been gaining attention as a valuable molecular characterization tool especially in the context of biological and medical research. Label-free molecular imaging and low invasiveness are some of the major beneficial points. Since biological and medical sample are mostly complex systems, utilization of Raman spectral data is restricted by only using classical univariate analytical methods. Multivariate curve resolution – alternating least square (MCR-ALS) is one kind of multivariate analysis to fully utilize Raman spectral information. This thesis consists of two practical applications of MCR-ALS to describe its usefulness and how to process entire analysis in biological and medical field.

1) Visualizing wax ester fermentation in single *Euglena gracilis* cells by Raman microspectroscopy and multivariate curve resolution analysis

Global demand for energy is on the rise at a time when limited natural resources are fast depleting. To address this issue, microalgal biofuels are being recommended as a renewable and eco-friendly substitute for fossil fuels. *Euglena gracilis* is one such candidate that has received special interest due to their ability to synthesize wax esters that serve as precursors for production of drop-in jet fuel. However, to realize economic viability and achieve industrial-scale production, development of novel methods to characterize algal cells, evaluate its culture conditions, and construct appropriate genetically modified strains is necessary. Here, we report a Raman microspectroscopy-based method to visualize important metabolites such as paramylon and ester during wax ester fermentation in single *Euglena gracilis* cells in a label-free manner.

We measured Raman spectra to obtain intracellular biomolecular information in *Euglena* under anaerobic condition. First, by univariate approach, we identified Raman markers corresponding to paramylon/esters and constructed their time-lapse chemical images. However, univariate analysis is severely limited in its ability to obtain detailed information as several molecules can contribute to a Raman band. Therefore, we further employed multivariate curve resolution analysis to obtain chain length-specific information and their abundance images of the produced esters. Accumulated esters in *Euglena* were particularly identified to be myristyl myristate (C28), a wax ester candidate suitable to prepare drop-in jet fuel. Interestingly, we found accumulation of two different forms of myristyl myristate for the first time in *Euglena* through our exploratory multivariate analysis.

2) Identification of Molecular Basis for Objective Discrimination of Breast Cancer Cells (MCF-7) from Normal Human Mammary Epithelial Cells by Raman Microspectroscopy and Multivariate Curve Resolution Analysis

Raman spectroscopy (RS), a non-invasive and label-free method, has been suggested to improve accuracy of cytological and even histopathological diagnosis. To our knowledge, this novel technique tends to be employed without concrete knowledge of molecular changes in cells. Therefore, identification of Raman spectral markers for objective diagnosis is necessary for universal adoption of RS. As a model study, we investigated human mammary epithelial cells (HMEpC) and breast cancer cells (MCF-7) by RS and employed various multivariate analyses (MA) including principal components analysis (PCA), linear discriminant analysis (LDA), and support vector machine (SVM) to estimate diagnostic accuracy. Furthermore, to elucidate the underlying molecular changes in cancer cells, we utilized multivariate curve resolution analysis–alternating least squares (MCR-ALS) with non-negative constraints to extract physically meaningful spectra from complex cellular data. Unsupervised PCA and supervised MA, such as LDA and SVM, classified HMEpC and MCF-7 fairly well with high accuracy but without revealing molecular basis. Employing MCR-ALS analysis we identified five pure biomolecular spectra comprising DNA, proteins and three independent unsaturated lipid components. Relative abundance of one lipid component seems to be strictly regulated between the two groups of cells and could be the basis for excellent discrimination by chemometrics-assisted RS. It was unambiguously assigned to linoleate rich glyceride and therefore serves as a Raman spectral marker for reliable diagnosis. This study successfully identified Raman spectral markers and demonstrated the potential of RS to become an excellent cytodiagnostic tool that can both accurately and objectively discriminates breast cancer from normal cells.

These results from two practical studies successfully show that Raman spectroscopy coupled with MCR-ALS is strong analytical tool to extract biomolecular information. Although this technique presently needs deep consideration of what is extracted, this problem will be solved combining such as standard Raman spectrum library and user-friendly interface in future. I believe Raman spectroscopy coupled with MCR-ALS technique probably endows researches opportunity for reading "molecular finger print" as it is.

MCR-ALS 法を組み合わせたラマン分光法による医・生物学的応用研究

摘要

ラマン分光法は、生物学的および医学的研究を研究するためのツールとして注目をあつめている。ラベルフリー分子イメージングが可能なことと、低侵襲性で非破壊な分子分析が可能であることが主な理由である。生物試料には多数の分子が存在するため、従来の単変量分析のみを用いたラマンスペクトルの解析は分子情報のすべてを活かすことができない。交互最小二乗多変量曲線分解(MCR-ALS) は、ラマンスペクトルに含まれる分子の情報を純粋なラマンスペクトル成分に分解して、半定量的に評価できる多変量解析であるが、医・生物学分野における適用例はまだ多くない。本研究は、医・生物学分野において MCR-ALS を適用した 2 つの実践的な研究についてその研究成果を示すとともに、ラマン分光法と MCR-ALS を組み合わせた手法の有用性を示すことを目的とした。本論文は、次の 2 章から成る。

1）顕微ラマン分光法と多変量曲線分解を用いたユーグレナのワックスエステル発酵の可視化

世界のエネルギー需要が増加する一方で、化石燃料の急速な枯渇が問題となっている。微細藻類バイオ燃料は、化石燃料の代替品として注目を集めており、そのうち *Euglena gracilis* は、ワックスエステルの一種であるミリスチン酸ミリスチル(C28)を多く蓄積する。このワックスエステルはジェット燃料に利用可能なドロップイン燃料の原料として期待されている。産業規模の生産を実現するには、高効率でワックスエステルを生産するような遺伝子組換え株をスクリーニングしたり、培養条件を評価したりする手法が必要である。そこで、本研究では、顕微ラマン分光法を用いてワックスエステル発酵に関わる代謝物を視覚化する方法を示した。

ユーグレナは好気性の条件でワックスエステル発酵の前駆物資であるパラミロンを蓄積し、嫌気性の条件でミリスチン酸ミリスチル(C28)を多く蓄積することが知られている。好気性条件下で培養したユーグレナを嫌気的条件に移し、パラミロンがミリスチン酸ミリスチルに代謝されていく過程の分子イメージングを試みた。まず、パラミロン/ワックスエステルに対応するラマンマーカーバンドを決定し、ラマンイメージングを構築した。しかし、この単変量解析では、複数の分子がラマンバンドに寄与する可能性があるため、正確な可視化ができているとは言えなかった。そこで、多変量曲線分解(MCR-ALS)を用いたところ、ミリスチン酸ミリスチル (C28) の鎖長特異的なラマンスペクトル成分が得られ、正確な分子イメージングが可能となった。また、ミリスチン酸ミリスチルの標準ラマンスペクトルが異なる 2 種類存在し、同一のラマンスペクトル成分がユーグレナでも得られることを明らかにした。単変量解析ではこの 2 種類のラマンスペクトルを反映したイメージングを

得ることはできないため、MCR-ALS 解析を適用する重要性も同時に示すことができた。

2) 顕微ラマン分光法および多変量曲線分解による正常ヒト乳腺上皮細胞と乳癌細胞 (MCF-7) の客観的識別のための分子認識

非侵襲、ラベルフリーで分子認識が可能なラマン分光法は、細胞学的および組織病理学的診断の精度向上をもたらすことが期待されている。しかし、ラマン分光法を用いた医療診断に関する研究報告では、機械学習を適用することで高い診断精度を示す一方で、多くの例では細胞内の分子変化に関する具体的な理解を伴っていない傾向がある。ラマン分光法を用いた診断は、現在は臨床で用いられている段階ではなく、将来において普遍的に利用されるためには、分子情報に基づいた客観的な診断ができることを示す必要がある。そこで本研究では、ヒト乳上皮細胞 (HMEpC) および乳がん細胞 (MCF-7) を細胞診モデルとして測定し、ラマンスペクトルマーカーの探索をおこなった。診断精度の算出のために、主成分分析 (PCA)、線形判別分析 (LDA)、およびサポートベクトルマシン（SVM）などの、様々な多変量分析を行った。教師あり学習である LDA と SVM によって、HMEpC と MCF-7 は正確に分類されたが、PCA では両者の識別を可能にする分子に基づいた情報は不明瞭なままであった。

そこで、がん細胞の分子変化を解明するために、MCR-ALS を用いて、物理的に意味のあるラマンスペクトル成分の抽出を試みた。結果として、DNA、タンパク質、3 つの独立した不飽和脂質成分を含む 5 つの純粋な生体関連分子ラマンスペクトルが抽出された。また、これらの成分の半定量的な比較によって、3 つの不飽和脂質のうちリノール酸に帰属された脂質の相対的量が、SVM や LDA が示した高い診断精度の分子学的基礎である可能性が示唆された。この研究は、乳がん診断におけるラマンスペクトルマーカーの同定に成功し、ラマン分光法が診断技術としてがん細胞と正常細胞を正確かつ客観的に区別できる優れた手法となる可能性を示した。

これら 2 つの実践的研究の結果は、ラマン分光法と MCR-ALS による解析が、生体分子情報を抽出するための強力な分析手法であることを示した。現状では、抽出されたラマンスペクトルの特徴から特定分子への帰属は深く考察する必要がある。標準ラマンスペクトルライブラリの構築やそれを参照するユーザーフレンドリーなインターフェイスと組み合わせることで、様々な分野の研究者が「分子の指紋」をそのまま読み解く事が可能になる。

List of Publications

1. Visualizing wax ester fermentation in single *Euglena gracilis* cells by Raman microspectroscopy and multivariate curve resolution analysis.
   Biotechnol Biofuels **12**, 128 (2019).

   Chapter II

2. Identification of Molecular Basis for Objective Discrimination of Breast Cancer Cells (MCF-7) from Normal Human Mammary Epithelial Cells by Raman Microspectroscopy and Multivariate Curve Resolution Analysis.
   Int. J Mol. Sci. **22,** 800 (2021).

   Chapter III