



HAL
open science

Coherent anti-Stokes Raman scattering cell imaging and segmentation with unsupervised data analysis

Damien Boildieu, Tiffany Guerenne-del Ben, Ludovic Duponchel, Vincent Sol, Jean-Michel Petit, Eric Champion, Hideaki Kano, David Helbert, Amandine Magnaudeix, Philippe Leproux, et al.

► To cite this version:

Damien Boildieu, Tiffany Guerenne-del Ben, Ludovic Duponchel, Vincent Sol, Jean-Michel Petit, et al.. Coherent anti-Stokes Raman scattering cell imaging and segmentation with unsupervised data analysis. *Frontiers in Cell and Developmental Biology*, 2022, 10, pp.933897. 10.3389/fcell.2022.933897. hal-03726275

HAL Id: hal-03726275

<https://hal.science/hal-03726275>

Submitted on 16 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN ACCESS

EDITED BY
Zhichao Fan,
UCONN Health, United States

REVIEWED BY
Rongrong Liu,
Northwestern University, United States
Ilhan Candan,
Dicle University, Turkey

*CORRESPONDENCE
Philippe Leproux,
philippe.leproux@xlim.fr

SPECIALTY SECTION
This article was submitted to Cellular
Biochemistry,
a section of the journal
Frontiers in Cell and Developmental
Biology

RECEIVED 01 May 2022
ACCEPTED 11 July 2022
PUBLISHED 16 August 2022

CITATION
Boildieu D, Guerenne-Del Ben T,
Duponchel L, Sol V, Petit J-M,
Champion É, Kano H, Helbert D,
Magnaudeix A, Leproux P and Carré P
(2022), Coherent anti-Stokes Raman
scattering cell imaging and
segmentation with unsupervised
data analysis.
Front. Cell Dev. Biol. 10:933897.
doi: 10.3389/fcell.2022.933897

COPYRIGHT
© 2022 Boildieu, Guerenne-Del Ben,
Duponchel, Sol, Petit, Champion, Kano,
Helbert, Magnaudeix, Leproux and
Carré. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Coherent anti-Stokes Raman scattering cell imaging and segmentation with unsupervised data analysis

Damien Boildieu^{1,2}, Tiffany Guerenne-Del Ben³,
Ludovic Duponchel⁴, Vincent Sol³, Jean-Michel Petit³,
Éric Champion⁵, Hideaki Kano⁶, David Helbert²,
Amandine Magnaudeix⁵, Philippe Leproux^{1*} and Philippe Carré²

¹University of Limoges, CNRS, XLIM, UMR 7252, Limoges, France, ²University of Poitiers, CNRS, XLIM, UMR 7252, Poitiers, France, ³University of Limoges, PEIRENE, UR 22722, Limoges, France, ⁴University of Lille, CNRS, UMR 8516, LASIRE - Laboratoire de Spectroscopie Pour Les Interactions, La Réactivité et L'Environnement, Lille, France, ⁵University of Limoges, CNRS, Institut de Recherche sur Les Céramiques, UMR 7315, Limoges, France, ⁶Department of Chemistry, Faculty of Science, Kyushu University, Fukuoka, Japan

Coherent Raman imaging has been extensively applied to live-cell imaging in the last 2 decades, allowing to probe the intracellular lipid, protein, nucleic acid, and water content with a high-acquisition rate and sensitivity. In this context, multiplex coherent anti-Stokes Raman scattering (MCARS) microspectroscopy using sub-nanosecond laser pulses is now recognized as a mature and straightforward technology for label-free bioimaging, offering the high spectral resolution of conventional Raman spectroscopy with reduced acquisition time. Here, we introduce the combination of the MCARS imaging technique with unsupervised data analysis based on multivariate curve resolution (MCR). The MCR process is implemented under the classical signal non-negativity constraint and, even more originally, under a new spatial constraint based on cell segmentation. We thus introduce a new methodology for hyperspectral cell imaging and segmentation, based on a simple, unsupervised workflow without any spectrum-to-spectrum phase retrieval computation. We first assess the robustness of our approach by considering cells of different types, namely, from the human HEK293 and murine C2C12 lines. To evaluate its applicability over a broader range, we then study HEK293 cells in different physiological states and experimental situations. Specifically, we compare an interphasic cell with a mitotic (prophase) one. We also present a comparison between a fixed cell and a living cell, in order to visualize the potential changes induced by the fixation protocol in cellular architecture. Next, with the aim of assessing more precisely the sensitivity of our approach, we study HEK293 living cells overexpressing tropomyosin-related kinase B (TrkB), a cancer-related membrane receptor, depending on the presence of its ligand, brain-derived neurotrophic factor (BDNF). Finally, the segmentation capability of the approach is evaluated in the case of a single cell and also by considering cell clusters of various sizes.

KEYWORDS

cell imaging, cell segmentation, coherent anti-Stokes Raman scattering, unsupervised data analysis, coherent Raman imaging, label-free imaging, supercontinuum, multivariate curve resolution

1 Introduction

In the context of label-free cell imaging, vibrational spectroscopic approaches have proven their effectiveness to visualize the cellular content and processes (Matthäus et al., 2008; Klein et al., 2012; Palonpon et al., 2013; Smith et al., 2016). As chemically selective techniques, they can provide both structural and functional information from the sample. Among these techniques, coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS) have been extensively applied to live-cell imaging in the last 2 decades (Zumbusch et al., 1999; Freudiger et al., 2008; Cheng and Xie, 2015). More specifically, both methods have been successfully used to probe the intracellular lipid, protein, nucleic acid, and water content with a high-acquisition rate and sensitivity.

In this context, multiplex CARS (MCARS) microspectroscopy (Kee and Cicerone, 2004; Kano and Hamaguchi, 2005) is a powerful technology for extracting rich vibrational information from biological samples. It is based on the excitation of the sample by two laser beams, namely, the monochromatic pump (ω_p frequency) and the broadband Stokes (ω_s frequency) beam. When the difference $\omega_p - \omega_s$ matches the vibration frequency of a vibrational mode, a resonant CARS signal is generated at the frequency $2\omega_p - \omega_s$, that is, at a wavelength lower than that of the pump laser, preventing any fluorescence background in the measured MCARS spectrum. This is a considerable advantage for this technique because conventional Raman often presents a fluorescence that completely masks the vibrational signal, especially in biology. However, it is well-known that MCARS is affected by the presence of a so-called non-resonant background (NRB), which is inherent to its physical mechanism. In this regard, phase retrieval approaches like the maximum entropy method (MEM) and time-domain Kramers–Kronig (TDKK) are generally used to extract the pure vibrational signal and to recover conventional Raman-like spectra (Cicerone et al., 2012). These spectra are then more easily interpretable and allow generating high-contrast images from the identified Raman bands. More recently, phase extraction from MCARS spectra was investigated by using a deep learning approach, based on the concept of supervised methods and using a long short-term memory network (LSTM) architecture (Houhou et al., 2020). Another study introduced a workflow for fast Raman signal extraction, denoising, error correction, and the applicability of this workflow to machine learning (Camp et al., 2020).

On the other hand, chemometric methods as principal component analysis (PCA) or multivariate curve resolution

(MCR) are widely used in spectroscopy-based data analysis (Pisapia et al., 2018; Ghaffari et al., 2019; Amigo, 2020). They consist of an unsupervised, statistical multivariate exploration of the collected data, being considered as established approaches, especially in the field of Raman microspectroscopy. PCA's purpose is to project data into a subspace, allowing to decorrelate the measurements and to maximize the variance of projections. Thus, PCA facilitates the observation of the relevant information of the dataset. Following this concept of projection of a multicomponent signal into subspaces, MCR is an iterative matrix decomposition method. It decomposes the considered dataset by means of its projection into a subspace guided by different constraints (for instance, non-negativity constraint). This is a definite advantage because this signal unmixing approach allows us to find contributions in the sample that are far more easily interpreted than those obtained from PCA. Specifically, the MCR approach decomposes the initial dataset into a matrix of “pure concentration profiles” and a matrix of “pure spectral profiles,” providing a concentration map (quantitative information) and the associated spectrum (qualitative information) of each contribution within the investigated sample. For example, the algorithm allows, in a context of chemical analysis, to extract the “pure” chemical species of the sample (De Juan and Tauler, 2021).

To date, MCR has never been applied to the analysis of MCARS hyperspectral datasets. The only work that could come close to such an approach was published by Zhang et al., 2013, with idea to extract images of major components in breast cancer cells using SRS. In this case, the multivariate exploration involved a spectral stack of only ~80 SRS images, and representative spectra of lipid droplets, nuclei, and culture medium were used at the initial estimation step (supervised method). Regarding CARS data processing in the context of cell imaging, few studies report on the use of chemometric methods within the proposed workflow, including PCA (Parekh et al., 2010; Pohling et al., 2011; Lee et al., 2014), singular value decomposition (SVD) (Masia et al., 2013), or hierarchical cluster analysis (HCA) (El Mashtoly et al., 2014). These studies aim at performing an unsupervised quantitative chemical analysis of an unknown biological sample and/or a classification of cells or subcellular organelles. Overall, they rely on relatively complex data analysis pipelines, including a heavy spectrum-to-spectrum phase retrieval step, which limit their dissemination into the biomedical field. A more recent work focused on using PCA and k-means clustering to perform a comparison between CARS and SRS in the context of tissue imaging (Bocklitz et al., 2018). Based on all these works, it is

evident that an MCR-based approach should be able to bring a new look on MCARS data by generating less biased cell images while facilitating the interpretation of the extracted chemical and biological information. Moreover, such approach would be all the more powerful as it could take both the spectral and spatial dimensions of the dataset into account, and the sole spectra being usually exploited in the field.

Today, MCARS microspectroscopy using a dual-output sub-nanosecond laser source is recognized as a mature and straightforward technology for bioimaging, offering the high spectral resolution of conventional Raman spectroscopy with reduced acquisition time (Kano et al., 2019; Kaneta et al., 2021). We have recently demonstrated its efficiency for cell cycle studies (Guerenne-Del Ben et al., 2019) or cancer cells characterization (Guerenne-Del Ben et al., 2020). In the proposed study, we naturally combine the potential of both this imaging technique and the MCR source separation method under the classical signal non-negativity constraint and, even more originally, under a new spatial constraint based on cell segmentation. We thus introduce a new approach of MCARS hyperspectral cell imaging and segmentation, based on a simple workflow, without any phase retrieval computation, for data acquisition, chemical analysis, and visualization.

The imaging capability of this approach is first described using a MCARS dataset obtained from a fixed human embryonic kidney 293 (HEK293) cell in the interphase of the cell cycle (G_1/S). This cell is subsequently used as the “reference” cell. Second, we compare the results obtained for the reference HEK293 cell with those obtained for a premyoblastic cell from the C2C12 murine line under similar conditions (*i.e.*, fixed during the interphase of the cell cycle). This comparison allows us to assess the robustness of our approach by considering two very different cell types. To evaluate its applicability over a broader range, we then study HEK293 cells in different physiological states and experimental situations (Guerenne-Del Ben et al., 2019; Guerenne-Del Ben et al., 2020). For this purpose, we compare the results of the reference (interphasic) cell to those obtained from another cell in mitosis (prophase), during which important nuclear and cytoplasmic rearrangements occur. We also present a comparison between the reference (fixed) cell and a living cell, in order to visualize the potential changes induced by the fixation protocol in cellular architecture. Next, with the aim of assessing more specifically the sensitivity of our approach, we study HEK293 living cells overexpressing tropomyosin-related kinase B (TrkB), a cancer-related membrane receptor, depending on the presence of its ligand, brain-derived neurotrophic factor (BDNF). Finally, the application of the segmentation constraint in the MCR framework is introduced and discussed in the case of the single reference cell. Then, the segmentation capability of the approach is evaluated by considering cell clusters of various sizes.

2 Materials and methods

2.1 Cell samples preparation

The datasets of HEK293 cells used in this study were obtained from cell samples prepared by T. Guerenne-Del Ben, according to the protocols described in Guerenne-Del Ben et al., 2019 and Guerenne-Del Ben et al., 2020. HEK293 cell line was provided by Pr. F. Lalloue, CAPTuR, UMR INSERM 1308, Faculty of Medicine, University of Limoges, France. HEK293 cells were non-modified or overexpressing the TrkB receptor.

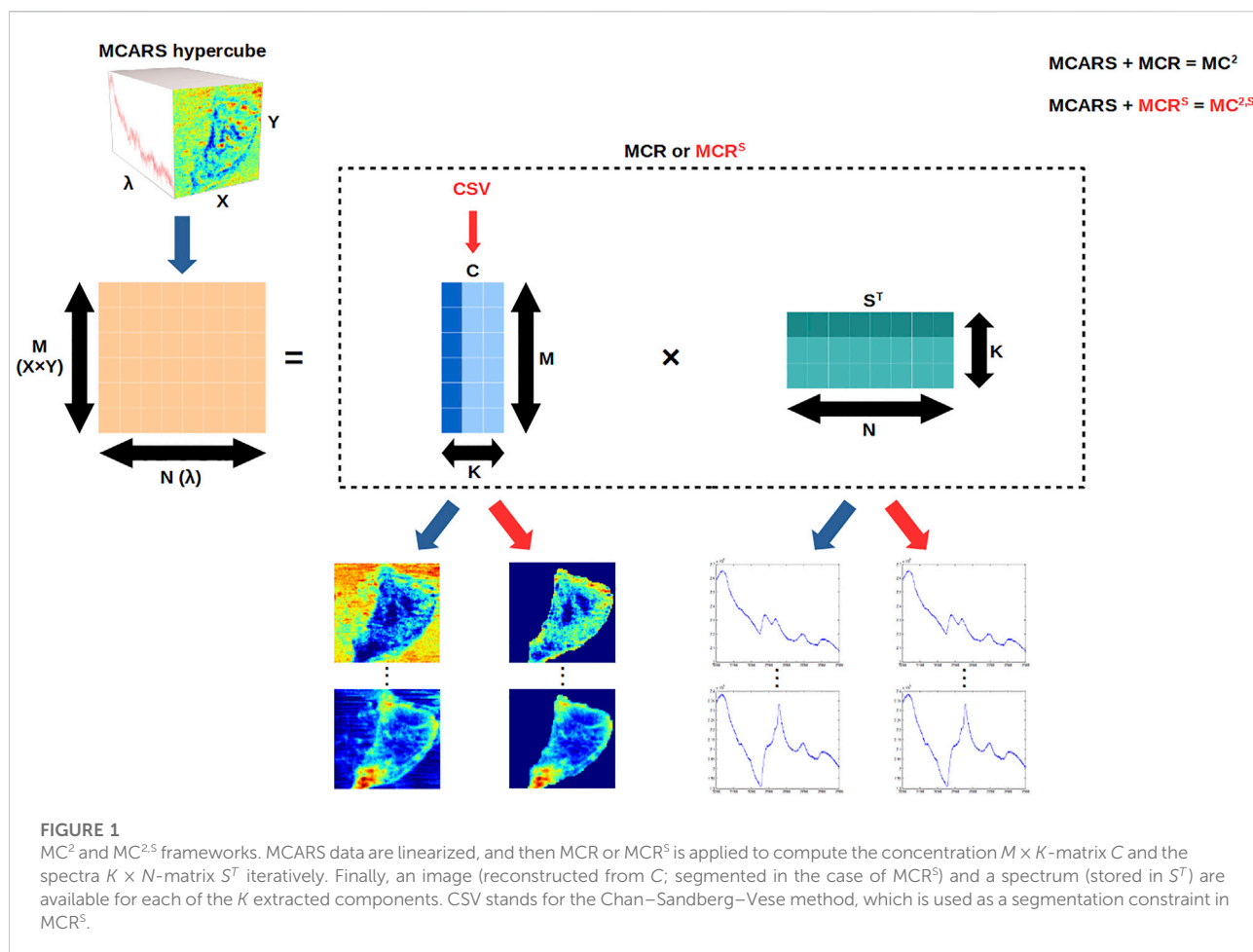
Non-modified HEK293 cells were routinely cultured in a complete medium consisting in high-glucose (4.5 g/L) DMEM (Gibco), supplemented with 10% (v/v) fetal calf serum (Eurobio), 1 μ g/ml amphotericin B (Gibco), 100 units/ml penicillin, and 100 μ g/ml streptomycin (Gibco) in an incubator heated at 37°C under 5% CO₂ humidified atmosphere. HEK293 cells overexpressing TrkB were cultured in a complete medium supplemented with 750 μ g/ml geneticin (G418, Roth). C2C12 cell line (ATCC, Manassas, VA, United States) was cultured in DMEM supplemented with L-glutamine, 10% (v/v) fetal calf serum, 50 units/mL penicillin, and 50 μ g/ml streptomycin.

For MCARS analysis, the cells were seeded at a density of 8,750 cells/cm² on 18-mm glass coverslips in 12-well plates for 48 h.

For the synchronization of non-modified HEK293 cells (study of the cell cycle), a first blocking with thymidine was performed for 18 h. Then for obtaining G_1/S or mitotic cells, after a release phase in complete medium of, respectively, 9 and 4 h, the cells were exposed to either 2-mM thymidine (double block procedure) for 17 h or 100 ng/ml nocodazole (Sigma-Aldrich) during 12 h before being replaced in fresh complete medium (for more details, refer to Guerenne-Del Ben et al., 2019). When a fixation was performed, the cells were washed three times in DPBS (Gibco), immersed in 4% paraformaldehyde (PFA) for 10 min at room temperature, and washed again three times in DPBS. A subset of living cells was labeled with the nuclear dye Hoechst 33342 (ThermoFisher) at 10 μ g/ml for 15 min and room temperature before being rinsed in DPBS. In addition, a subset of fixed cells was labeled with the nuclear dye DAPI (4',6-diamidino-2-phenylindole, Sigma-Aldrich) at 1 μ g/ml for 5 min.

HEK293 cells overexpressing TrkB were cultured in the presence of recombinant human BDNF (Peprotech) at a final concentration of 75 ng/ml, or its solvent (DPBS), for 72 h. Cell nuclei were then stained with Hoechst 33342 as described earlier.

Glass coverslips with living or fixed cells were mounted on a microscope slide in DPBS and sealed with nail polish.



2.2 MCARS microspectroscopy and fluorescence imaging

The MCARS microspectroscopy and associated experimental conditions are described in [Guerrene-Del Ben et al., 2019](#). In brief, the MCARS setup is based on a passively Q-switched microchip laser (Horus Laser, 1,064 nm, 1 ns, 20 kHz), a photonic crystal fiber for generating the broadband infrared Stokes wave, focusing (Olympus, UPlanSApo 60x, N.A. = 1.2, water immersion) and collecting (Nikon, S Plan Fluor ELWD 60x, N.A. = 0.7) microscope objectives, and a spectrometer (Horiba, LabRam HR Evolution). CARS spectra were acquired from 2,500 to 3,200 cm^{-1} with $\sim 0.8 \text{ cm}^{-1}$ spectral resolution, 50 ms pixel dwell time, and 300 nm lateral step for cross-sectional mapping. The lateral and axial resolutions were $\sim 300 \text{ nm}$ and $\sim 2 \mu\text{m}$, respectively. The laser power of the pump and Stokes radiations at the sample position was 55 and 9 mW, respectively, for which no morphological change of cells was observed during the experiments.

Fluorescence imaging was realized on the same system by using a halogen light source, appropriate excitation and emission filters, and a dedicated CCD camera (Thorlabs, 1500M-GE).

2.3 Maximum entropy method

As we have seen in the introduction, CARS spectra suffer from the presence of NRB that greatly distorts the acquired spectra and therefore the images that are extracted afterward. For the purpose of spectroscopic validation of our approach, we used the MEM algorithm from [Vartiainen et al. \(2006\)](#) as a reference method to extract the imaginary part of the third order non-linear susceptibility ($\text{Im}\{\chi^{(3)}\}$) and thus obtain conventional Raman-like spectra. The implementation of MEM was performed using MATLAB software (R2018b, MathWorks).

2.4 Multivariate curve resolution

MCARS datasets were analyzed using MCR ([De Juan et al., 2014](#)). MCR is a signal unmixing method, which aims at finding the K components of a multicomponent signal. The data matrix D , which contains M measures for N Raman shifts, is decomposed into two matrices, as depicted in [Figure 1](#): the concentration $M \times K$ -matrix C , including the K component

concentration maps along its columns (corresponding to the projection of the initial data), and the spectra $K \times N$ -matrix S^T , including the K component spectra along its rows (corresponding to the projection basis). More formally, the model is defined as follows:

$$D = CS^T + E, \quad (1)$$

where E being the error matrix, containing data not represented by the basis and expected to be noise or irrelevant information. In the case of chemical mixture analysis, $c_{i,j}$ is the concentration of compound j for measuring (i.e. pixel) i and row j of S^T , which is the characteristic spectrum of this compound.

In order to evaluate the quality of the MCR decomposition according to the value of K , the so-called lack-of-fit (LOF) was used:

$$LOF = \frac{\sum_{i,j} e_{i,j}^2}{\sum_{i,j} d_{i,j}^2}, \quad (2)$$

where $e_{i,j}$ and $d_{i,j}$ being elements of row i and column j of matrices E and D , respectively. The LOF corresponds to the ratio between the data contained in E (i.e., not expressed by the model) and those contained in the experimental dataset. Based on the variation of the LOF as a function of K for the datasets used in our study, the number of components was set to $K = 5$. This choice will be investigated later in the Discussion.

Even if MCR allows us to explore complex signals, this approach nevertheless presents what is called ambiguities, meaning that different pairs of C and S^T matrices can be acceptable solutions. In this context, applying constraints to C and S^T during MCR calculation allows reducing these ambiguities and the number of possible solutions, in order to potentially converge toward a unique decomposition. Two of the most common constraints are non-negativity and normalization. The non-negativity constraint ensures to have only positive values for all elements of C and/or S^T matrices, while the normalization constraint implies that the sum of all values of each row or column of C and/or S^T is equal to one. Obviously, constraints have to be selected according to the acquisition method, the study subject, and the intrinsic nature of the data. In this study, we applied the non-negativity constraint to the concentration matrix C and to the spectra matrix S^T , while the normalization constraint was only applied to each row (i.e., measure) of the concentration matrix.

One way to solve the MCR problem is to use an alternating regression algorithm to compute C and S^T . Here, we used the common “alternating least squares” (ALS) algorithm, the method being then called MCR-ALS. Moreover, we applied the non-negativity constraint within the decomposition process by using non-negative least squares (NNLS) (Lawson and Hanson, 1974). Practically, these results are a three-step process: first, the C matrix is computed by NNLS, then it is normalized and, in turn, the S matrix is computed by NNLS. This process is then repeated

until convergence. Furthermore, we used the SIMPLISMA algorithm (Windig and Guilment, 1991) for computing the initial S matrix of the iterative process.

We implemented the whole MCR-ALS workflow in Python language, based on the freely available “pyMCR” package (Camp, 2019). In the following sections, the combination of MCARS acquisition and MCR data analysis (with the two classical constraints) will be denoted as MC^2 (Figure 1).

2.5 New cell segmentation constraint (MCR^S)

To extract cells from their environment and refine the intracellular analysis, we have developed in this work a new spatial constraint to be applied to the concentration matrix C , additionally to the previous constraints in the MCR framework. Specifically, we applied a segmentation constraint based on the Chan–Sandberg–Vese (CSV) method (Chan et al., 2000), which has already shown its ability to segment biological cells (Dufour et al., 2005; Maška et al., 2013). With the use of this additional spatial constraint, the framework will be denoted as MCR^S and the whole approach as MC^{2-S} in this work (Figure 1).

The CSV method is an iterative segmentation method based on active contours. It consists of splitting the given image into two regions, using an energy-dependent model. The goal of the algorithm is to minimize the fitting energy between the model and the input image I :

$$\arg \min_{c_1, c_2, \mathcal{C}} F(c_1, c_2, \mathcal{C}) \quad (3)$$

$$\text{with} \quad F(c_1, c_2, \mathcal{C}) = \mu \cdot \text{Length}(\mathcal{C}) + \nu \cdot \text{Area}(\text{inside}(\mathcal{C})) + \lambda_1 \int_{x,y \in \text{inside}(\mathcal{C})} \|I(x, y) - c_1\|^2 dx dy + \lambda_2 \int_{x,y \in \text{outside}(\mathcal{C})} \|I(x, y) - c_2\|^2 dx dy,$$

where \mathcal{C} is the border between the two regions, and c_1 and c_2 are the region averages. This equation is solved using a partial differential equation. The method requires the following parameters to be defined:

- 1) “Length penalty” μ (to be set between 0 and 1) regularizes the length of \mathcal{C} . It is the most important parameter to tune (Getreuer, 2012). A small value leads to define the border of the cells with potentially a high level of detail, while a large one generates a smoother boundary. Here, we use $\mu = 0.35$. The rationale behind this choice is stated in Section 3 (more specifically, in the sub-section entitled “Comparison between MC^2 and MC^{2-S} methods applied to the reference cell”;
- 2) “Area penalty” ν tunes the penalty ($\nu > 0$) or reward ($\nu < 0$) to be applied to the area inside \mathcal{C} during minimization. As we do not have any *a priori* on the size of cells in the acquisition window, ν is set to zero;
- 3) “Fit weights” λ_1 and λ_2 allow to control deviation from the original image in terms of pixel intensity for areas inside and

outside \mathcal{C} , respectively. We use $\lambda_1 = \lambda_2 = 1$ in order to not to penalize one area against the other;

- 4) “Time step” Δt is the discretization step used to solve the partial differential equation. A small value gives better results with increased computation time, while a large value can lead to convergence problems. We use the default value $\Delta t = 0.5$ (Getreuer, 2012).

The new algorithm corresponding to one iteration of MCR^s is the following:

- 1) compute C using NNLS;
- 2) refold C into K images of size $X \times Y$ (consistently with the initial dataset);
- 3) apply the CSV method to these images in order to compute a segmentation mask;
- 4) use the mask to set concentration values outside \mathcal{C} to zero;
- 5) unfold segmented images in order to retrieve a (spatially) constrained C matrix;
- 6) apply the normalization constraint to C ;
- 7) compute S using NNLS.

The addition of the segmentation constraint to the MCR-ALS framework was also implemented in Python.

2.6 Projection onto reference spectra

In the case of living HEK293 cells overexpressing TrkB cancer-related protein, a particular approach was used to compare cells that were incubated with or without BDNF. First, the MC² approach was applied to a cell that was incubated without BDNF, considering the spectra of the extracted components as reference spectra. Then, the MCARS dataset of a second cell, incubated with BDNF, was projected onto the basis of reference spectra by using the NNLS regression algorithm. The resulting concentration matrix was then normalized, as previously. With this approach, we could visualize the spatial distribution of the same components in both cells, comparatively.

2.7 Statistical analysis of cell segmentation

In order to generate the ground truth masks, bright-field images of the cells were opened within NIH ImageJ software (<https://imagej.nih.gov/ij/>) and manually segmented using the “Polygons selection” tool. The comparison between ground truth and MC^{2.5} masks, and the calculation of Dice similarity coefficients were realized by using “imshowpair” and “dice” functions in MATLAB software (R2018b, MathWorks).

The statistical analysis of Dice coefficients was performed using PAST software (O. Hammer, D.A.T. Harper, P.D. Ryan,

PAST: paleontological statistics software package for education and data analysis, *Palaeontol. Electron.* 4 (2001) 1–9). After having tested the normality of the values using a Shapiro–Wilk test, a one-way analysis of variance (ANOVA) was performed and followed by a Tukey’s multiple comparison test. Differences were considered significant for $p < 0.05$.

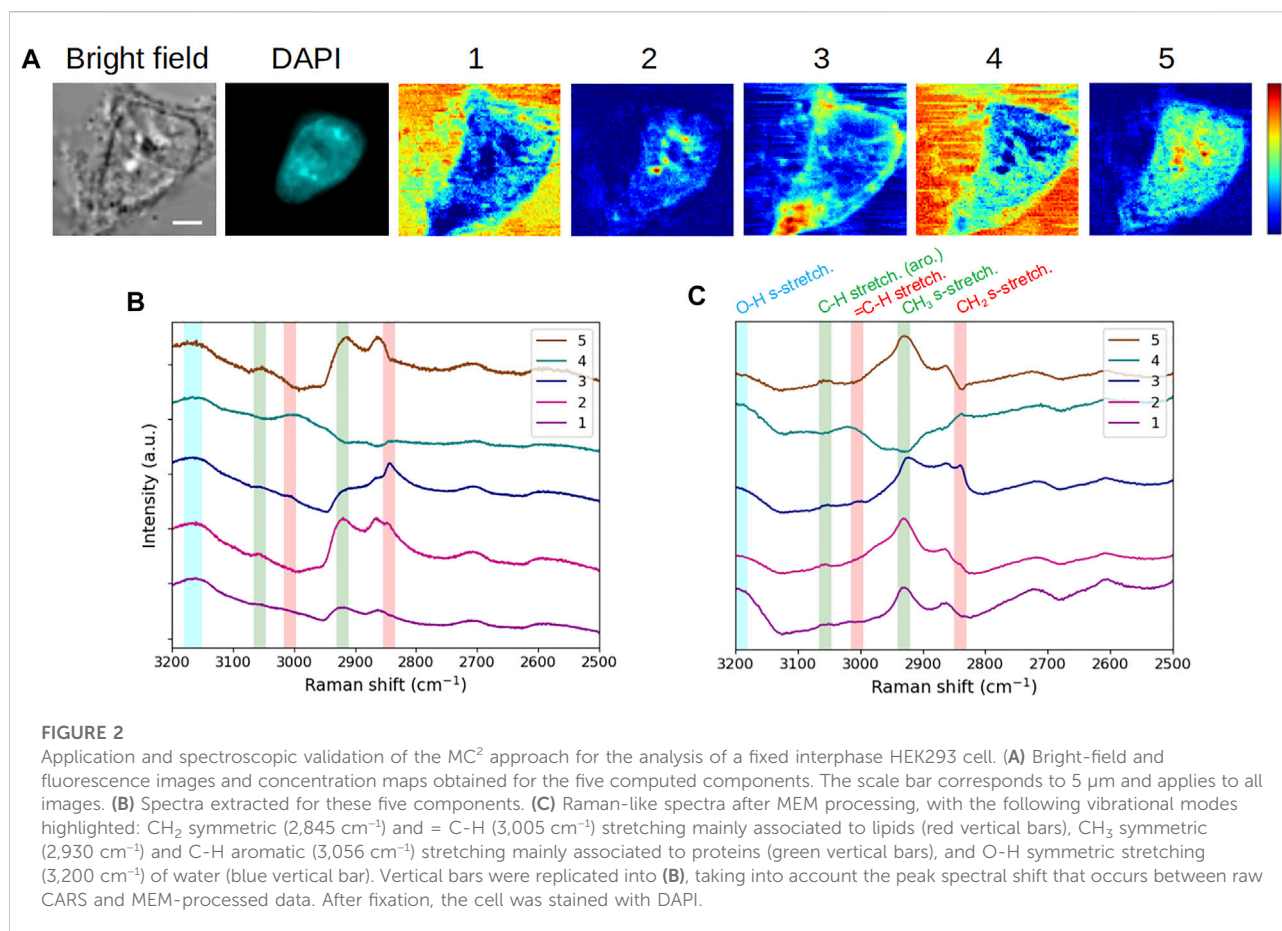
3 Results

3.1 Spectroscopic validation of the approach

We first applied the MC² approach to the analysis of a fixed, DAPI-stained, interphase HEK293 cell. Figure 2A displays the bright-field and fluorescence images, as well as the five concentration maps, while the corresponding component spectra are plotted in Figure 2B. In order to validate the spectroscopic features observed on these extracted spectra (as expected for conventional raw MCARS spectra), they were processed by the MEM algorithm so as to extract the corresponding conventional Raman-like spectra (Figure 2C). We have highlighted in Figure 2C the vibrational modes that are most relevant for cell analysis in this spectral range, namely, CH₂ symmetric (2,845 cm⁻¹) and = C-H (3,005 cm⁻¹) stretching, mainly associated to lipids (red vertical bars), CH₃ symmetric (2,930 cm⁻¹) and C-H aromatic (3,056 cm⁻¹) stretching, mainly associated to proteins (green vertical bars), and O-H symmetric stretching (3,200 cm⁻¹) of water (blue vertical bar). It is reminded that no vibrational signature of DAPI is expected in the high wavenumber (C-H stretching) region (Krause et al., 2007).

It appears from Figure 2C that components #1 and #4 are characterized by a high water content, concomitantly with the presence of proteins (for component #1) and lipids (for component #4). Components #2 and #5 are clearly dominated by the protein content but differ from each other by the lipid content (weak in component #2 and absent in component #5). Last, component #3 stands out by a strong lipid content, together with proteins. Therefore, from the spectroscopic viewpoint, the chemical species behind these components exhibit five different proportions of lipids, proteins, and water. This spectroscopic and chemical information of a given spectral contribution can then be related to the associated concentration map in order to observe the spatial distributions and biochemically investigate the intra- and extracellular contents.

In the following, we perform such biochemical study by simply using the spectral contributions extracted by the MC² approach, that is, without phase retrieval computation by the MEM. To this aim, we have replicated and slightly shifted the vertical bars of Figure 2C onto Figure 2B, taking into account the spectral shift of peaks that occurs between raw CARS spectra and MEM-processed ones (Freudiger et al., 2008; Capitaine et al., 2018). In that respect and in a first approach, the spectroscopic



analysis can be made from the MC² spectra by appreciating the presence of CARS peaks in the different spectral channels indicated by the vertical bars. A more thorough interpretation is to consider the significance (mathematically, the slope and the amplitude) of the dispersive lines associated to these peaks. This is particularly useful when the CARS peak is not apparent, due to an insufficient ratio between vibrationally resonant and non-resonant contributions, as in the case of CH₃ symmetric stretching for spectral contribution #3 (Figure 2B).

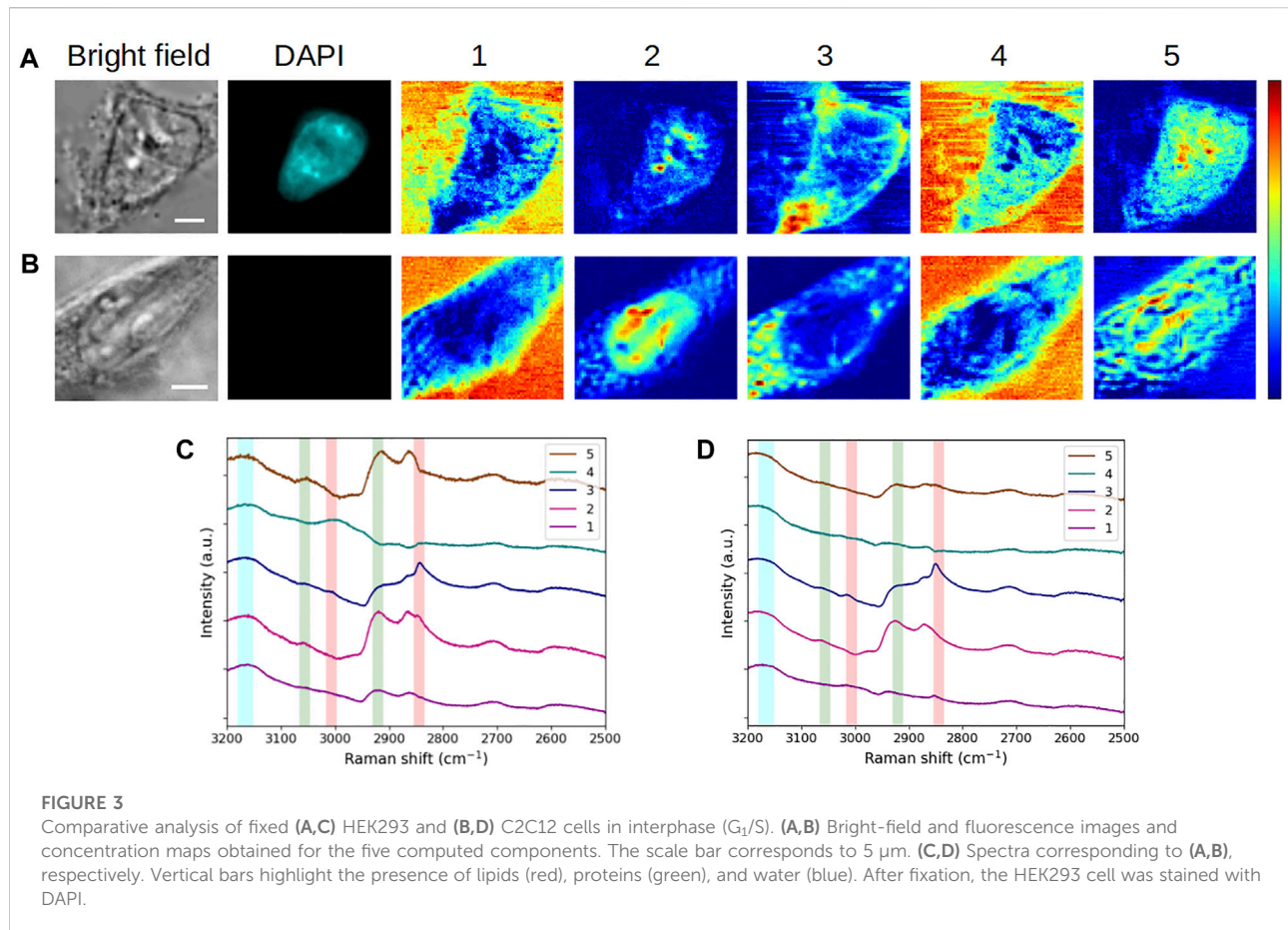
Finally, the fixed interphasic HEK293 cell observed in Figure 2 will be considered as the reference cell throughout the rest of this work. As a supplemental validation of our approach, we also applied MCR to the reference cell dataset after a prior step of MEM processing, taking into account the impact of the NRB spectrum source (here, the raw MCARS spectrum of the solvent) and of the baseline detrending (corresponding to the dark noise/background related to the spectrometer). Supplementary Figure S1 shows the evolution of the final outputs, according to these different processing steps. In this figure, the results obtained in the last case (NRB normalization combined with background subtraction) can thus be compared with MC² results (first row). From the spectral point of view, we stress the obvious

quality difference of $\text{Im}\{\chi^{(3)}\}$ spectra plotted in Figure 2C (MEM applied to the five spectra extracted by MCR) and Supplementary Figure S1 (MCR applied to all spectra first processed by MEM). Moreover, the $\text{Im}\{\chi^{(3)}\}$ spectra of Figure 2C demonstrate that MCR performs well with non-linear CARS intensity, since both the spectral shape and the exhibited vibrational information are consistent with Raman-like spectra and with the images of Figure 2A, respectively.

3.2 Comparative analysis of fixed cells in interphase (G₁/S)

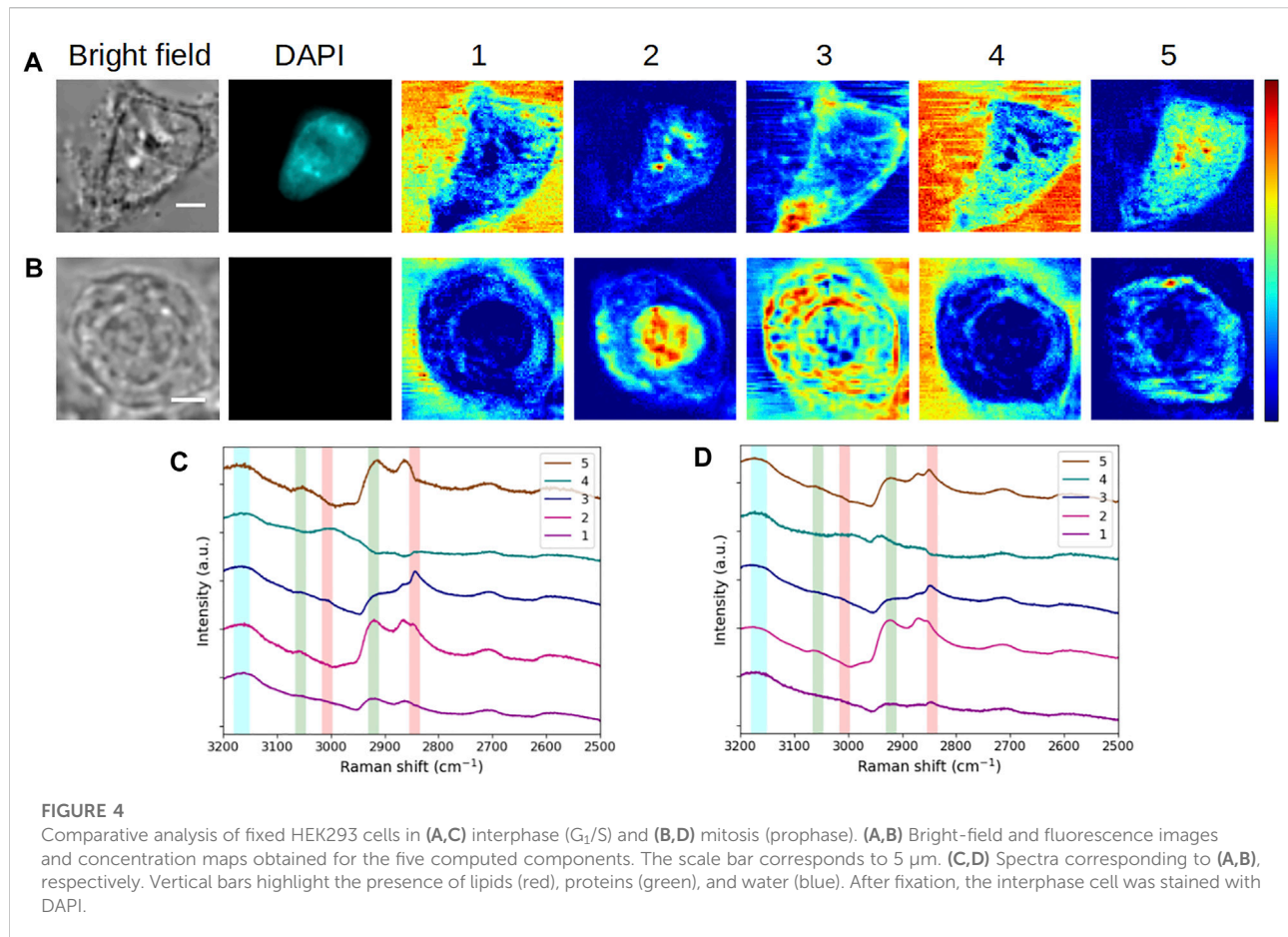
In this part of the work, we assessed the robustness of the MC² method by comparatively analyzing the HEK293 reference cell and a C2C12 interphasic unstained cell. Figure 3 shows the extractions obtained on these 2 cells. It should be noted that two separate MCR analyses were conducted on these two datasets.

When considering the concentration maps of the reference cell (Figure 3A), it is clear that components #1 and #4 (dominated by a high-water content, as already explained before) are mainly



related to the aqueous extracellular environment. From a spatial point of view, component #2 is localized at the nucleus as depicted by its round shape, overlapping with DAPI staining. The high-concentration spots highlight specific areas that can reasonably be attributed to nucleoli in relation to DAPI staining together with bright-field images. Nucleoli are active nuclear subcompartments whose one of the main functions is to be the site for the initial steps of ribosome biogenesis. This implies the presence of proteins linked to the transcription machinery and post-transcriptional modifications by small nucleolar ribonucleoproteins in the dense fibrillary component. The rRNA assembly with the ribosomal proteins then occurs in the granular component of the nucleolus (Pederson, 2011). This is thus in agreement with the high protein content emerging in spectral profile #2 (Figure 3C). Regarding component #3, it is concentrated in an area corresponding to the plasma membrane (or its very close periphery) and it is found in the cytoplasm as well. This is in line with the chemical species highlighted by the associated spectrum, that is, a lipid-rich content with the presence of proteins. Finally, component #5, which corresponds to a high protein content and an absence of lipids, is located throughout cell nucleus and cytoplasm, with a higher concentration in the nucleus and near nucleoli.

Overall, the extracted spectra of HEK293 and C2C12 cells (Figures 3C,D, respectively) exhibit a reasonable stability when they are observed pairwise for each component (except for component #5, which will be examined later). Accordingly, the concentration maps (Figures 3A,B) reveal a comparable spatial distribution of the signal between both cells for the first four components. Hence, with a high-water content, components #1 and #4 highlight the extracellular milieu. The observations made for component #2 in the case of the reference cell are applicable to the second cell, that is, a high protein content (with more or less lipids) emphasizes the nucleus and nucleoli (visible in bright-field for both cells). Regarding component #3 (lipid-rich content with the presence of proteins), the distribution of the signal in the cytoplasm of the C2C12 cell is more or less punctiform, suggesting that the related cytoplasmic structures correspond to lipid droplets. Lipid droplets are organelles consisting in a phospholipid monolayer surrounding a core composed of neutral lipids. They are well described in C2C12 premyoblastic cells (Billecke et al., 2015; Ichimura et al., 2015), and their abundance in myoblast cytoplasm was recently associated with a facilitated (induced) differentiation into myotubes (Tan et al., 2021). Moreover, in both cells, a portion of component #3 appears to surround the



nuclear structure and could be related to the nuclear envelope. For all these reasons, component #3 seems to relate to cellular membrane constituents such as phospholipids.

In the case of component #5, a variability is observed between both cells. For the reference HEK293 cell, this component represents a pure protein content in the cytoplasm and, more abundantly, in the nucleus, which is difficult to assign to a particular cellular compartment. For the C2C12 cell, spectrum #5 (Figure 3D) reveals a mixed content of proteins, lipids, and water. In the corresponding concentration map (Figure 3B), the signal is less diffuse and more reticular than for the HEK293 cell. This would suggest that endoplasmic reticulum (ER) is highlighted, assuming that lateral resolution is not sufficient to resolve individual ER layers and that the seeming presence of signal in the nucleus is due to positioning the focal plane at the nucleus periphery. However, further experiments would be necessary to support this assertion, including the specific labeling and fluorescence imaging of the ER. Anyhow, in both cases, the concentration map of component #5 is complementary to the other four maps. Regarding the variability of the spatial distribution between HEK293 and C2C12 cells, it can be hypothesized that component #5 is specific to the cell line,

given that the cells studied here belong to distinct types with very different physiological functions.

3.3 Comparative analysis of fixed cells in interphase (G_1/S) and mitosis (prophase)

Next, we compared the reference interphasic cell with an HEK293 unstained cell that was fixed in mitosis (prophase), as displayed in Figure 4. As previously, MCR analyses were performed separately.

Again, it is interesting to see that a certain stability is maintained at the spectral level between both cells for components #1–#4. Accordingly, in the concentration maps of the mitotic cell (Figure 4B), the extracellular environment is disclosed by components #1 and #4. Component #2 (high protein content) shows a strong and rather homogeneous signal at the place of nucleus, in agreement with the nuclear rearrangements occurring during prophase. The observed intensity variations can be correlated with the level of chromatin condensation. The concentration map of component #3 (lipid-rich content with proteins) highlights concentric structures looking like

membranes, the outermost layer corresponding to the plasma membrane. During mitosis, a major rearrangement of membrane structures takes place and is essential for the completion of a proper cell division leading to fully functional daughter cells (recently reviewed in [Carlton et al., 2020](#)). The fractionation of some of the organelles such as the ER and the Golgi apparatus, disassembled during late prophase (or prometaphase), leads to a new membrane organization with a modification of the protein/lipid ratio in the resulting membranes. Notably, an important part of the membrane components is bound to the plasma membrane. It is interesting to note that the concentric lamellar signal observed in concentration map #3 suggests such an organization, which would be consistent with the division state of the studied cell, visually estimated in prophase to late prophase.

Regarding component #5, we notice a divergence in the spectra and concentration maps corresponding to the interphasic and mitotic cells, respectively. We can assign this divergence to the different physiological states of the cells. In the case of the mitotic cell, the extracted spectrum displays a mixed protein-lipid content and the resulting concentration map exhibits a specific, polarized, and intracellular structure in the cytoplasm. As seen in [Supplementary Figure S2A](#), this structure (plotted in green here) is entangled with component #3 (plotted in red). The same trend is visualized for another (late) prophase cell ([Supplementary Figure S2B](#)), with an even more distinct polarization of component #5. These observations would suggest that this component might be linked to the mitotic spindle of the cells. The results obtained for an early-metaphase cell ([Supplementary Figure S2C](#)) consolidate this hypothesis, as well.

3.4 Comparative analysis of interphase (G_1/S) fixed and living cells

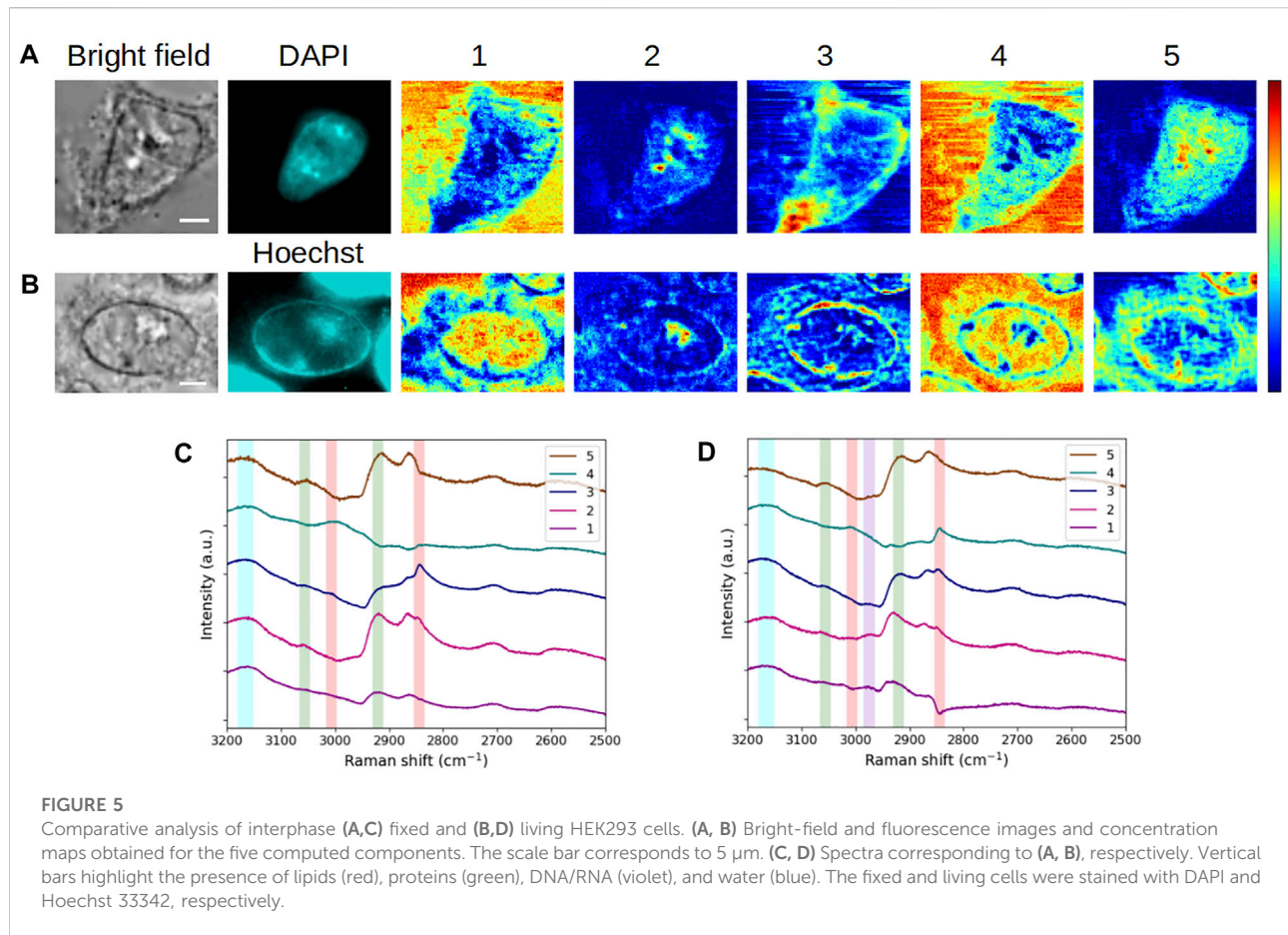
Cell fixation prior to *in situ* labeling and/or the use of several imaging methods have non-negligible effects on cell structure. These effects differ according to the fixation reagent. For example, formaldehyde, routinely used, can react with numerous functional groups of macromolecules, especially inducing protein or protein-DNA cross-linking ([Hoffman et al., 2015](#)). The consequences of fixation from a chemical standpoint are accompanied by the alteration of cell mechanical properties, which were investigated in [Kim et al. \(2017\)](#) by means of atomic force microscopy and scanning ion conductance microscopy. For their part, [Hobro and Smith \(2017\)](#) could use Raman spectroscopic imaging and PCA to evaluate several fixation methods. In this context, we sought to assess our approach for the study of cell alterations induced by fixation. In that way, [Figure 5](#) presents the concentration maps and the associated spectra, resulting from the application of the MC^2 method to the reference cell ([Figures 5A,C](#)), fixed with PFA, and a living HEK293 cell ([Figures 5B,D](#)).

First, some common features persist when comparing the concentration maps of the 2 cells. Components #1 and #4, dominated by the water content, again points out here the extracellular environment. Nucleoli visualization is still found in component #2, while component #3 discloses membrane structures. However, we can notice some obvious differences between both cells, which are described later.

In the living cell, component #1 highlights not only the extracellular milieu but also the nucleus, which is visible for neighbor cells in the field of view, as well. In addition to the nucleus, the cytoplasm is shown by component #4, which is spectrally characterized by the combined presence of water and lipids. Interestingly, this component spotlights the border of the nucleus and the nucleoli as low-signal areas, indicating that water and lipids are not found together in these areas (the sole lipids being expected at the nuclear envelope and non-expected in the nucleoli). Overall, the presence of the intracellular aqueous content is not observed for the fixed cell. This would be consistent with a loss of cellular content due to the fixation that may induce shrinkage, especially when performed at room temperature.

In the case of component #2, besides nucleoli, the border of the nucleus (the nuclear envelope) is highlighted in the living cell, in agreement with bright-field and Hoechst 33342 fluorescence images. A further investigation of spectrum #2 brings out some vibrational information in the channel plotted in violet ([Figure 5D](#)). This contribution is assigned to CH_3 antisymmetric stretching in the $2,960\text{--}2,980\text{ cm}^{-1}$ range ([Matthews et al., 2010](#)). In [Supplementary Figure S3](#), we have compared the spectra of component #2 obtained for all fixed and living cells considered so far, including both raw and MEM-processed spectra. This figure displays a stronger contribution of the interphase living HEK293 cell in the highlighted $2,960\text{--}2,980\text{ cm}^{-1}$ channel. Following [Lu et al., 2015](#), this contribution can be attributed to DNA in live cells and can thus be associated with a structure linked to the nuclear envelope in the present case ([Figure 5B](#), map #2). Indeed, during interphase, heterochromatin binds indirectly to the nuclear envelope by means of proteins involved in tethering chromatin. The absence of such a perinuclear signal in the fixed cell ([Figure 5A](#), map #2) would be consistent with the chemical mechanism of action related to cell fixation with PFA.

Regarding component #3, in the living cell, the cytoplasm and the nuclear border clearly appear. Taking into account the combined presence of lipids/proteins and the reticular aspect of the signal in the cytoplasm, one can suggest that the revealed structures correspond to the nuclear envelope in contiguity with the rough ER. This assumption is supported by the weak contribution of CH_3 antisymmetric stretching to component #3 ([Figure 5D](#), spectrum #3, violet channel), which is assigned to ribosomal RNA in the cytoplasm ([Lu et al., 2015](#)). Finally, component #5, dominated by proteins and exempt from lipids, discloses the inner border of the nuclear envelope, likely



corresponding to the nuclear lamina and the cytoplasm, likely due to the presence of resident soluble proteins in the latter. The absence of lipids in this component results in a “double ring” signal, localized on both sides of the nuclear envelope.

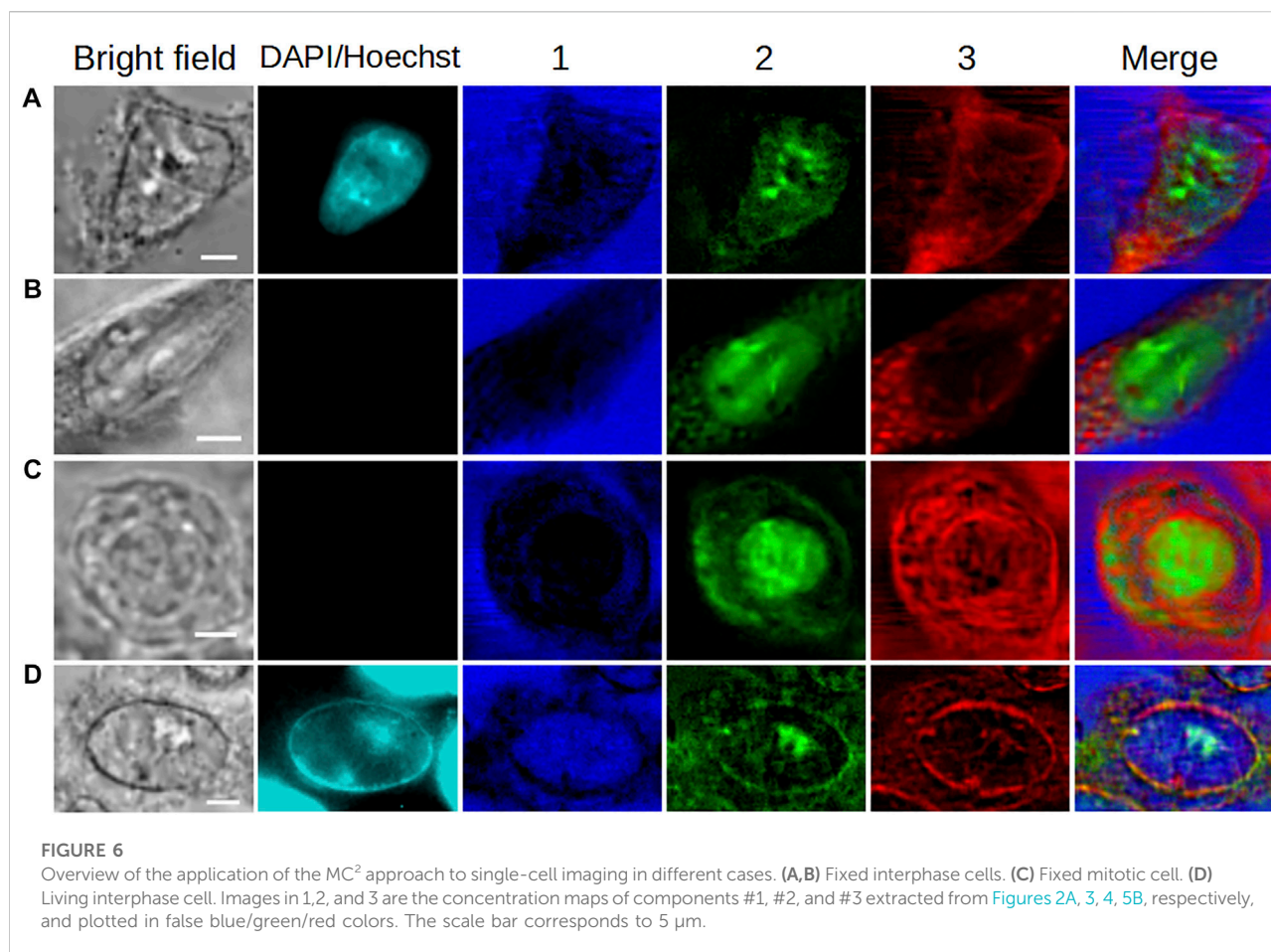
In order to illustrate the potential of the MC^2 approach for single-cell imaging, Figure 6 summarizes the results obtained with HEK293 and C2C12 cells in the different physiological states (fixed interphase, fixed mitotic, and living interphase cells). In this figure, we plotted the concentration maps of components #1, #2, and #3 from Figures 2A, 3, 4, 5B in false colors, specifically in blue, green, and red, respectively. The merge column shows the overlay of these three maps, illustrating the complementarity of the extracted components under each biological condition. Moreover, an easier comparative analysis of all results can be made from Figure 6.

For a further validation of our approach, the workflow of Supplementary Figure S1 (NRB normalization + background subtraction + MEM + MCR) was applied to the four cells of Figure 6. Output maps and spectra are available in Supplementary Figures S4 and S5. Generally speaking, combining MEM and MCR leads to more noisy images and spectra, due to additional processing steps and thus potential loss of information. In Supplementary Figure S4, the parallel representation of “MCR” and “MEM + MCR” results allows

to validate our new approach, especially by comparing the high-protein and high-lipid content channels (second and third columns, in each case). However, even if our approach is workable and concentration maps are quite similar in this figure, we draw attention to the fact that, rigorously, the method is not expected to be quantitative since the raw CARS signal is not proportional to the concentration.

3.5 Analysis of living cells overexpressing TrkB and treated or not with brain-derived neurotrophic factor

Last, we evaluated the robustness and the applicability of the MC^2 method for detecting changes in cells where a variable was introduced. For this purpose, we analyzed the datasets of two living HEK293 cells overexpressing the TrkB receptor at a basal level or activated by the introduction of its ligand, BDNF, in the culture medium for 72 h (Figure 7). In this case, we used the spectra calculated for the non-treated cell (Figure 7C) as a reference, and then we projected both datasets onto this base to construct the concentration maps of the non-treated (Figure 7A) and TrkB-activated (Figure 7B) cells. In addition,



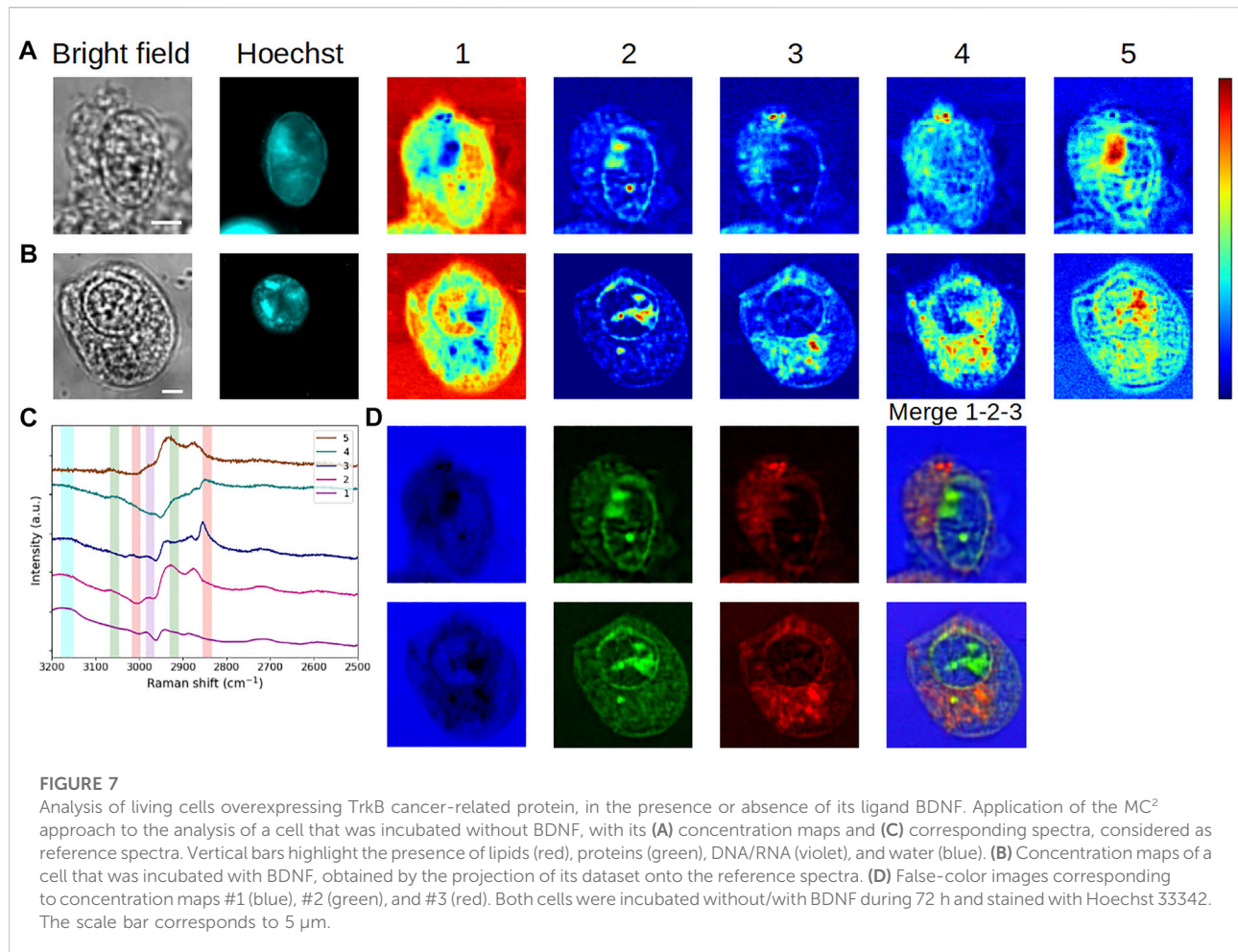
both cells can be visualized in false colors for components #1–3, as displayed in Figure 7D.

As an initial investigation, the information contained in components #1–3 is rather similar to that of the previously studied living HEK293 cell. Specifically, in both cells, some water appears in the nucleus (except the nucleoli), the nucleoli, and the nuclear border are highlighted, as well as the lipid content in the cytoplasm. As previously, the DNA contribution is found in spectrum #2, as confirmed in Supplementary Figure S2 (green spectra). Concerning component #3, given the punctiform distribution of the corresponding intracytoplasmic signal and considering the shape of spectrum #3, it is stressed that this component exhibits lipid droplets in both cells. Thereupon, the apparent larger accumulation of lipid droplets in the TrkB-activated cell would be related to a modification of the lipid metabolism, as suggested by Guerrenne-Del Ben et al., 2020. Further than this work, here the comparative study of the non-treated and treated cells is substantially facilitated by the great complementarity between components #1–3. In the case of component #4, the obvious increase of high-intensity spots in the second cell may also correlate with the alteration of the lipid metabolism.

Finally, for component #5, the signal is distributed throughout the nucleus and the cytoplasm of both cells, but a more reticular structure is observed in the non-treated one. In both cases, an intense spot is visible in the nucleus, partially overlapping with nucleoli and low-water content areas. Another feature is the presence of the signal, in the TrkB-activated cell, at the place of the plasma membrane. Yet, proteins constitute the main vibrational signature in spectrum #5, which furthermore differs from other spectra by the complete lack of contribution in the water region. Then, this high protein content near the plasma membrane may reflect the consequences of cell exposure to the growth factor, and particularly the activation of TrkB by BDNF at the cell surface. At this step, it is more difficult to interpret the intracellular content revealed by component #5.

3.6 Comparison between MC² and MC^{2,S} methods applied to the reference cell

We have seen in the previous sections that the MC² approach allows to extract interesting biological information from the MCARS spectra. Nevertheless, we should not forget that only



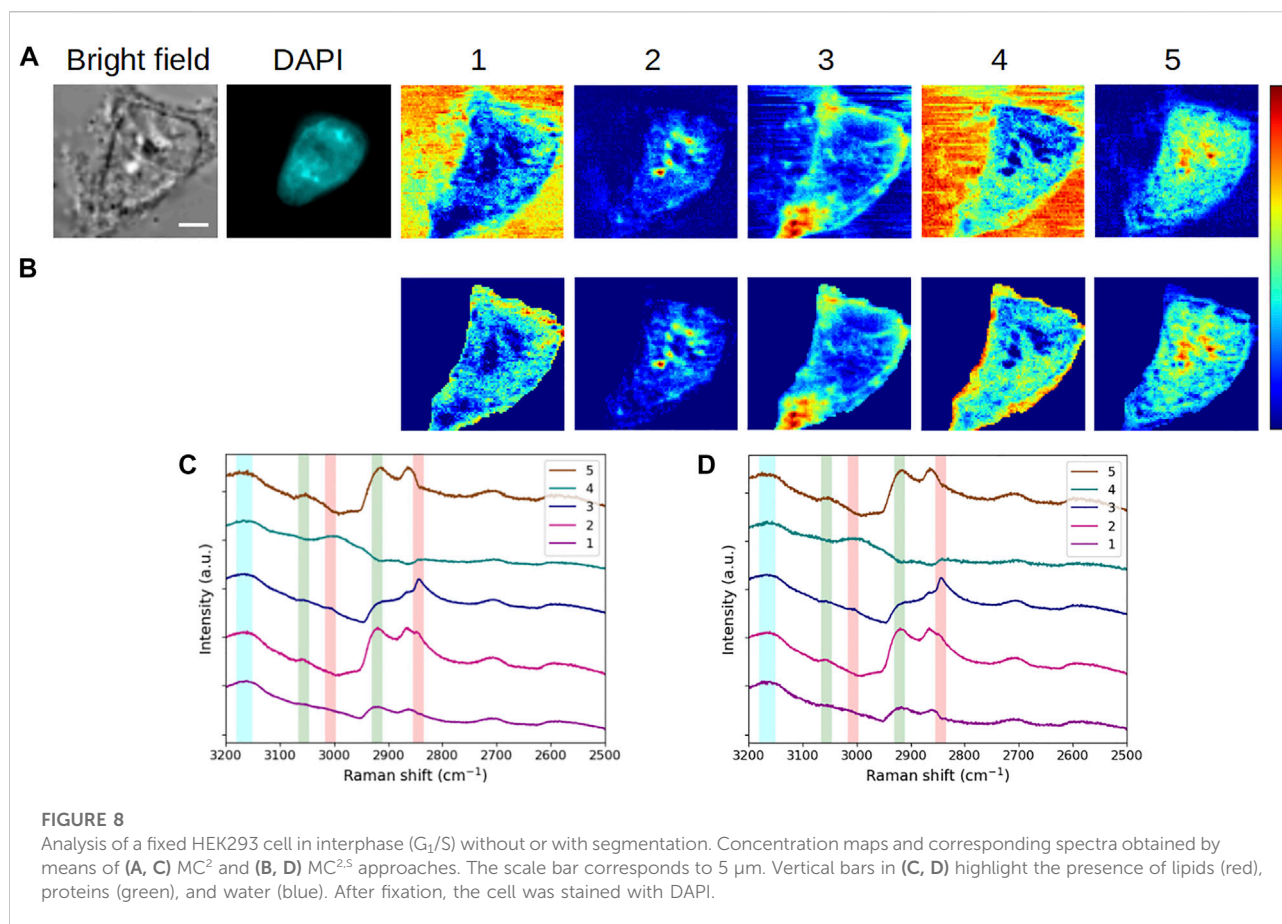
the spectral data were exploited to generate these maps and that the spatial information was not exploited in this approach. We thus developed the MC^{2,S} method in order to take into account simultaneously the spectral and spatial dimensions, to refine the extraction of components and to perform cell segmentation.

As mentioned in Section 2, the length penalty μ is an important parameter to set (between 0 and 1) in the MC^{2,S} approach. Several values of μ were tested for the reference cell, as illustrated in Supplementary Figure S6. We could then study the evolution of the cell contour according to μ variation, for each of the five components extracted by MCR^S. Small values of μ lead to cropping the bottom-left part of the cell, particularly in the case of components #1 and #5 (arrow heads, Supplementary Figure S6). From $\mu = 0.35$, this bottom-left curve becomes comparable to what is observed without using the segmentation constraint. Higher values of μ induce an oversegmentation, visible for components #1, #3, and #4 on the upper side of the cell (full arrows, Supplementary Figure S6). For these reasons, we set the value of μ to 0.35.

Then, to appreciate the impact of the addition of the segmentation constraint to the MCR framework, both MC² and

MC^{2,S} methods were applied to the reference cell. The results are shown in Figures 8A,C and in Figures 8B,D, respectively.

The vibrational modes of each extracted spectrum were not modified by the introduction of the segmentation constraint, though some noise appeared on the spectral profiles. Regarding concentration maps, the value added by the new constraint is undeniable since the cell is clearly delimited from the background under these new conditions. In addition, the improved contrast within the cell allows a better rendering of intracellular structures. The most obvious changes are observed in concentration maps #1 and #4. In these maps, the extracellular milieu was initially the main contributing element. By using the segmentation constraint in the MCR process, the resolution is tightened on the intracellular area, resulting in a more accurate extraction of the aqueous content inside the cell. As a consequence, the main contributions in concentration maps #1 and #4 now appear in the inner periphery of the plasma membrane (this is particularly true for component #4) and, secondarily, in the nucleus and cytoplasm. Beyond the investigation of intracellular water, MC^{2,S} would be of interest for the study of elements that are prominent in the cell environment.



3.7 Application of MC² and MC^{2,S} methods to the analysis of cell clusters

Three different unlabeled, fixed, HEK293 cell clusters were analyzed to assess the robustness of MC² and MC^{2,S} methods and their applicability to more complex situations (Figure 9). The corresponding spectra are given in Supplementary Figures S7–S9.

When looking at the concentration maps obtained by the MC² approach, we can notice a good reproducibility in the extraction of components for each cell cluster, with respect to previous results. Specifically, components #1 and #4 reveal the extracellular environment. Component #2 is associated with nuclei/nucleoli. Component #3 is mostly cytoplasmic and highlights a perinuclear structure (obvious in 2-cell and 5-cell clusters), probably made up of intracellular membranes belonging to the ER. Component #5 allows to visualize other cytoplasmic elements distributed asymmetrically within the cells.

The application of the MC^{2,S} method provides the suitable segmentation of each cell cluster and a better visualization of intracluster structures. This benefit is noticeable in the false color representation of concentration maps #1, #2, and #3. In

particular, it is interesting to see the presence of water in the intercellular space (5-cell and 8-cell clusters) and on the periphery of the clusters. Of course, the study of aqueous content would be supplemented by taking component #4 into account and by investigating the contribution of vibrational modes reflecting the presence of other chemical species. The MC^{2,S} approach would thus be of interest to determine the state of water, namely, free water or water hydrating macromolecules.

3.8 Statistical analysis of cell segmentation

The accuracy of the segmentation included in the MC^{2,S} analysis was evaluated by calculating the Dice similarity coefficients that are well recognized as a valuable metrics (Hermesen et al., 2019; Eelbode et al., 2020). Dice coefficients estimate the spatial overlap between the ground truth, corresponding in our case to a manual segmentation made by a cell biologist, and the segmentation mask generated by the MC^{2,S} method. They are defined by $DC(\text{Ground truth}, \text{MC}^{2,S} \text{ segmentation}) = 2(|\text{Ground truth} \cap \text{MC}^{2,S} \text{ segmentation}|) / (|\text{Ground truth}| + |\text{MC}^{2,S} \text{ segmentation}|)$, where \cap is the

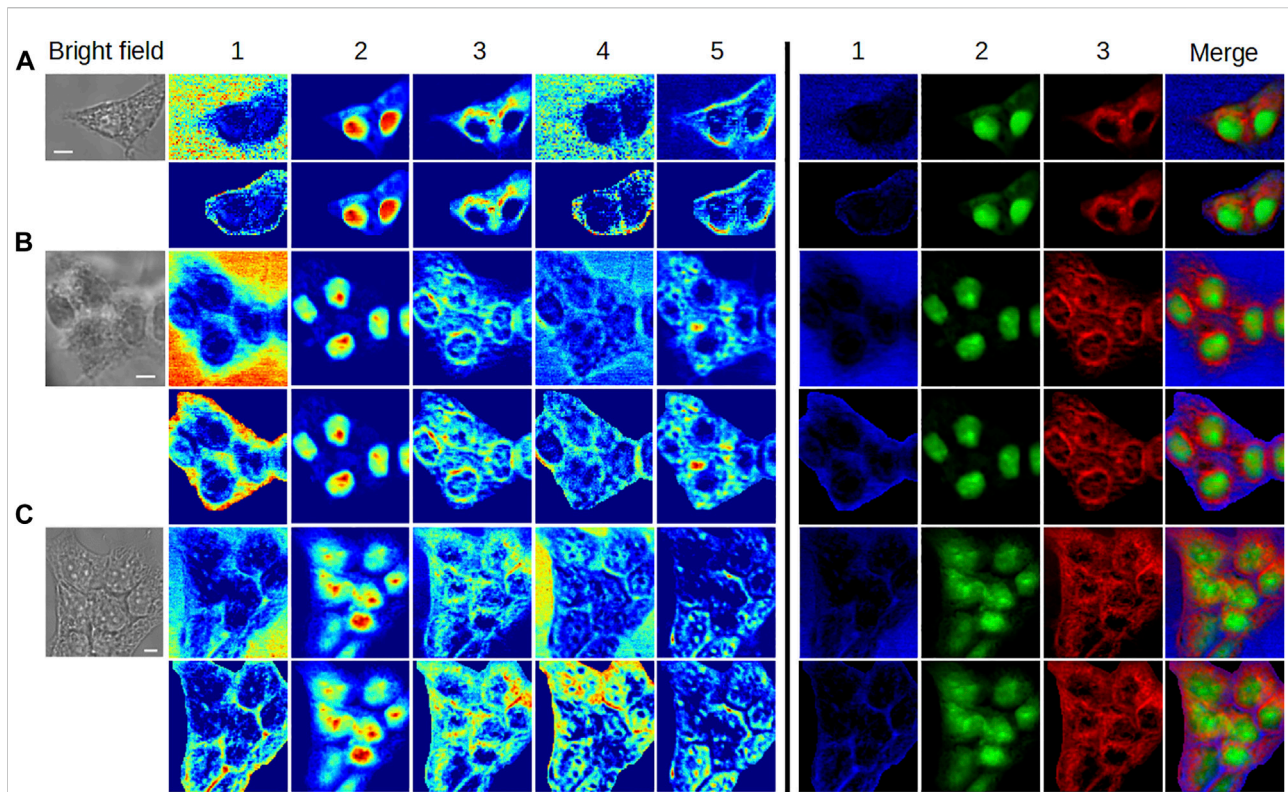


FIGURE 9

Application of MC^2 and $MC^{2.5}$ approaches to the analysis of cell clusters. (A) 2-cell cluster, (B) 5-cell cluster, and (C) 8-cell cluster. For each cluster, the concentration maps computed by MC^2 and $MC^{2.5}$ are plotted in first and second rows, respectively, using two different colormaps. The scale bar corresponds to 5 μm . The corresponding spectra are available in [Supplementary Material](#).

intersection. Dice coefficients are comprised between 0 and 1, 0 being a total absence of overlapping, and 1 a total overlap between manual and automatic segmentations. We calculated the Dice coefficients for a total of 18 comparisons that correspond to cells or clusters included in [Guerenne-Del Ben et al., 2019](#) and to unpublished results obtained in the context of the cited work. These comparisons were classified according to four categories, as shown in [Supplementary Figure S10](#) (interphase fixed cells), [Supplementary Figure S11](#) (mitosis fixed cells), [Supplementary Figure S12](#) (interphase living cells), and [Supplementary Figure S13](#) (fixed cell clusters).

Overall mean of Dice coefficients was 0.86 ± 0.07 . In light of this result, $MC^{2.5}$ segmentation can be considered of good quality, all the more so as manual segmentation from bright-field images is a source of error and CARS measurement differs from bright-field by its optical sectioning capability (causing lateral shift and/or size modification between both modalities, as seen in several cases). Finally, we plotted in [Figure 10](#) the calculated Dice coefficients, according to the physiological state of the cells. Following the p -value analysis of these data ($p > 0.05$), no significant difference was observed between the different cell categories, highlighting the robustness of the proposed method toward the cell morphology.

4 Discussion

4.1 Determination of the number of components used in multivariate curve resolution

Here, we discuss the strategy that we adopted to set the number of components to $K = 5$.

From the signal analysis point of view, we consider that the analyzed MCARS hypercube is created by various sources of information, each of them having an internal structure that can be exploited. There are many methods in the literature for solving this source separation problem. A first family of methods relies on stochastic assumptions, the main approach in this field being PCA. The latter consists in diagonalizing the covariance matrix to build a projection basis and thus decorrelate the data (specifically, the second-order moments). A second family of methods aims at providing the sparsest approximation, allowing to reconstruct the initial dataset as best as possible. A typical example is the matching pursuit ([Shi et al., 2013](#)). We position our approach in this second family of methods. As previously mentioned, in the case of MCR, the C matrix can be interpreted as the projection of original data onto a new vector basis defined by S^T , which contains K rows (i.e., K

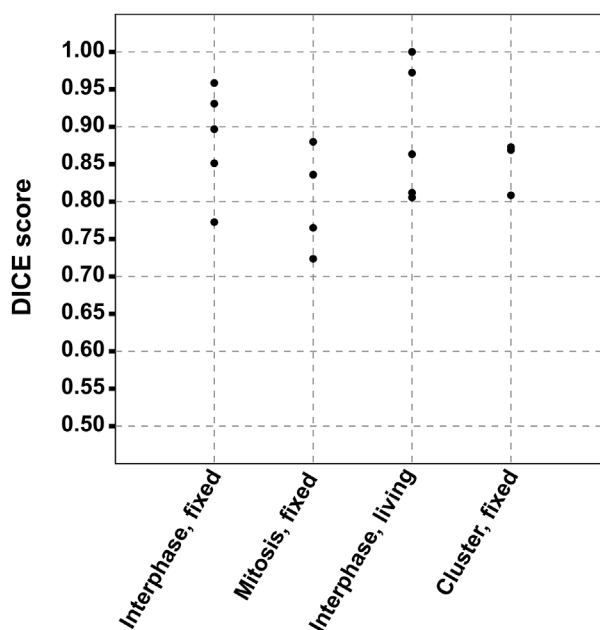


FIGURE 10

Dice coefficients according to the physiological state of the cells. The similarity was calculated between the ground truth, corresponding to the manual segmentation from bright-field images and the segmentation mask generated by the MC^{2S} method.

TABLE 1 LOF evolution according to the number of components. First row is the number of components K extracted by MCR, second row is the LOF mean μ_{LOF} (in percent), and third row is the LOF standard deviation σ_{LOF} (in percent). The corresponding curve is given in [Supplementary Figure S7](#).

K	1	2	3	<u>4</u>	5	<u>6</u>	7	8	9	10	11	12	13	14	15
μ_{LOF} (%)	0.481	0.369	0.262	<u>0.243</u>	0.214	<u>0.202</u>	0.20	0.192	0.193	0.189	0.186	0.186	0.186	0.175	0.175
σ_{LOF} (%)	0.771	0.592	0.463	<u>0.443</u>	0.373	<u>0.361</u>	0.365	0.356	0.363	0.362	0.355	0.354	0.355	0.327	0.327

The “elbow” points are underlined, and the selected value is in bold.

components). Whichever method is chosen, a difficult question is to estimate the size of the projection space ensuring the selection of the informative components from the analyzed signal.

For the PCA approach, most of the estimators work exclusively from the eigenvalues of the covariance matrix. It is usual to use a screen test, which consists in detecting the existence of a significant decay in the eigenvalues diagram for determining the rank (the number of eigenvectors to retain). In practice, this method is not always reliable because of its subjectivity. Another method is to use the approximation error to determine the best choice of the number of components: during the iterative construction of the projection basis, the Euclidean norm of the approximation error is used in the stopping criterion. In the case of a linear transformation like conventional PCA, these two methods of rank selection are actually equivalent.

For the MCR approach, it is common practice to determine the dimension of the projection space by studying the eigenvalues of the covariance matrix—by means of PCA—before the MCR computation. However, in our case, we focus on building the best

approximation of the initial dataset. Therefore, we introduce the use of the LOF to evaluate the quality of the MCR decomposition according to the value of K . In this context, the LOF mean and standard deviation, μ_{LOF} and σ_{LOF} , respectively, were calculated for the whole group of datasets considered in the present work. The elbow method (Thorndike, 1953) was then applied to determine the best value of K . Table 1 shows μ_{LOF} and σ_{LOF} with K varying from 1 to 15. The corresponding curve is plotted in [Supplementary Figure S14](#). According to the elbow method, we derive $4 \leq K \leq 6$ with $0.243 \geq \mu_{LOF} \geq 0.202$ and $0.443 \geq \sigma_{LOF} \geq 0.361$. Given that $K = 5$ provides new information compared to $K = 4$, whereas $K = 6$ only brings redundancy, and the number of components is set to $K = 5$. As a comparison, by applying the criterion of decay of the eigenvalues, we obtain $K = 2$, which is clearly insufficient in view of the information effectively contained in the datasets. Thus, we suggest that our strategy is appropriate to realize the unsupervised exploration of the different cell samples.

4.2 Robustness and relevance of the MC² method for label-free cell imaging

For all the cells considered in this study, except those overexpressing TrkB, the first four MCR components are remarkably stable, highlighting cell common constituents (in both spectra and maps) and allowing a consistent comparison between the samples. Component #5 appears to be more specific to each studied individual condition and contains information related to this specificity, namely, cell type, step of the cell cycle, and (patho) physiological or metabolic status. In this sense, component #5 is highly relevant and confirms the capability of the MC² method to extract peculiar information. However, as this component is specific, a direct comparison of different cells (regarding component #5) is not sufficient for a reliable analysis. Hence, the use of MC² on a reference cell/condition and the subsequent projection of data obtained with different cells/conditions onto the reference spectra is particularly useful. This kind of data relativization is commonly used in biology, for example in metabolic activity assays, or for representing qRT-PCR data ($\Delta\Delta\text{CT}$ method). The choice of the reference cell/condition is then crucial, and modifying this choice allows to carry out a systematic and deeper analysis, considering the specificity of component #5. Of course, this approach is applicable to all components extracted by MCR. Furthermore, it is stressed that it is valid when only a unique variable is introduced in the experimental context (e.g. pharmacological treatment, ligand binding, *etc.*), as in the case of living cells overexpressing TrkB and treated or not with BDNF.

In view of the overall results presented in this work, the MC² method shows its great potential in label-free cell imaging. First, the observations made here are consistent with the previous ones using largely the same datasets (Guerenne-Del Ben et al., 2019; Guerenne-Del Ben et al., 2020). Beyond that, we could make more accurate analyses and comparisons of experimental situations, thanks to the superior unmixing of cell constituents and the extraction of additional contributions. For instance, when comparing cells in different stages of the cell cycle, it was possible to propose a better estimation of the mitotic subphase (from late prophase/prometaphase to metaphase). New information came also from the comparison of the living cell with the fixed one regarding the loss of cellular content due to the fixation. Yet, the reader should note that these findings are based on the study of individual cells to establish a proof-of-concept, and that they cannot, at this step, be generalized to cell populations.

4.3 Refinement of intracellular/intracluster analysis by the addition of a segmentation constraint (MC^{2,S} method)

As displayed in Figure 8 for the reference cell, the MC^{2,S} method performs a good extraction of the cell from its

environment and improves the image contrast within the cell, allowing to refine the intracellular analysis. The use of the segmentation constraint makes the LOF increase from 0.165 to 55.6% in this case, expressing that a major part of data (the extracellular milieu) was removed. Therefore, MC^{2,S} can be of particular interest for studying intracellular events linked to a component that is in abundance in the extracellular milieu. The aqueous content, associated with components #1 and #4, is a perfect illustration of this perspective, as seen in the corresponding concentration maps (Figure 8B). In that respect, it is worth noting that the study of “biological water” can raise different view angles and interests, depending on the involved disciplinary field (interestingly outlined in Jungwirth, 2015). Anyhow, we stress that the detailed knowledge of the water state and distribution in cellular compartments would have a high (patho) physiological relevance. Moreover, several optical approaches were recently introduced for intra/extracellular water imaging, based on MCARS (Nuriya et al., 2019), stimulated Raman excited fluorescence (Shi et al., 2019), or fluorescence lifetime measurement (Rao et al., 2019). In this context, we suggest that the MC^{2,S} approach could significantly contribute to the field by taking into account both spatial and spectral dimensions from data acquisition to numerical processing. Obviously, our approach can be extended to the O-H stretching and fingerprint regions, as we already demonstrated ultra-multiplex CARS with 500–4,000 cm⁻¹ coverage in previous works (Kano et al., 2019; Nuriya et al., 2019; Kaneta et al., 2021).

Finally, the potentialities offered by the MC^{2,S} method are confirmed in the case of cell clusters, whatever number of cells is considered. Here, the extraction of information is refined inside the cluster and, in particular, the water content can be clearly visualized between the cells forming the cluster. These results show that the MC^{2,S} method is also applicable to the study of biological tissues.

To conclude, this work establishes MC^{2,S} as a new methodology for MCARS hyperspectral cell imaging and segmentation, based on a simple, unsupervised workflow without any spectrum-to-spectrum phase retrieval computation. Such easy-to-use methodology, combined with the constant simplification of MCARS instrumentation, should substantially participate in disseminating coherent Raman technologies into the biomedical field.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://gitlab.xlim.fr/datacart/mc2s>.

Author contributions

TG-DB prepared the samples. PL developed the custom-made MCARS system. TG-DB and PL performed the experiments. LD performed a preliminary chemometric exploration of experimental data. DB designed and implemented the numerical methods, processed all experimental data, and performed the analysis, under the supervision of DH, PC, PL, and AM. PL and AM interpreted the results. DB, PL, AM, DH, and PC wrote the manuscript; LD, J-MP, VS, ÉC, and HK reviewed it. All authors provided critical input and helped shape the research.

Funding

We acknowledge French government support managed by the National Research Agency under the Investments for the Future program, with the reference ANR-10-LABX-0074-01 Sigma-LIM.

References

- Amigo, J. M. (2020). *Hyperspectral imaging*. Elsevier Vol. 32. Available at: <https://www.sciencedirect.com/bookseries/data-handling-in-science-and-technology/vol/32>.
- Billecke, N., Bosma, M., Rock, W., Fleissner, F., Best, G., Schrauwen, P., et al. (2015). Perilipin 5 mediated lipid droplet remodeling revealed by coherent Raman imaging. *Integr. Biol.* 7, 467–476. doi:10.1039/c4ib00271g
- Bocklitz, T., Meyer, T., Schmitt, M., Rimke, I., Hoffmann, F., von Eggeling, F., et al. (2018). Invited Article: Comparison of hyperspectral coherent Raman scattering microscopies for biomedical applications. *APL Photonics* 3, 092404. doi:10.1063/1.5030159
- Camp, C. H., Jr., Bender, J. S., and Lee, Y. J. (2020). Real-time and high-throughput Raman signal extraction and processing in CARS hyperspectral imaging. *Opt. Express* 28, 20422–20437. doi:10.1364/OE.397606
- Camp, C. H., Jr. (2019). pyMCR: A Python library for multivariate curve resolution analysis with alternating regression (MCR-AR). *J. Res. Natl. Inst. Stand. Technol.* 124, 1–10. doi:10.6028/jres.124.018
- Capitaine, E., Ould Moussa, N., Louot, C., Bardet, S. M., Kano, H., Duponchel, L., et al. (2018). Fast epi-detected broadband multiplex CARS and SHG imaging of mouse skull cells. *Biomed. Opt. Express* 9, 245–253. doi:10.1364/BOE.9.000245
- Carlton, J. G., Jones, H., and Eggert, U. S. (2020). Membrane and organelle dynamics during cell division. *Nat. Rev. Mol. Cell Biol.* 21, 151–166. doi:10.1038/s41580-019-0208-1
- Chan, T. F., Sandberg, B. Y., and Vese, L. A. (2000). Active contours without edges for vector-valued images. *J. Vis. Commun. Image Represent.* 11, 130–141. doi:10.1006/jvci.1999.0442
- Cheng, J. X., and Xie, X. S. (2015). Vibrational spectroscopic imaging of living systems: An emerging platform for biology and medicine. *Science* 350, aaa8870. doi:10.1126/science.aaa8870
- Cicerone, M. T., Aamer, K. A., Lee, Y. J., and Vartiainen, E. (2012). Maximum entropy and time-domain Kramers-Kronig phase retrieval approaches are functionally equivalent for CARS microspectroscopy. *J. Raman Spectrosc.* 43, 637–643. doi:10.1002/jrs.3169
- De Juan, A., Jaumot, J., and Tauler, R. (2014). Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Anal. Methods* 6, 4964–4976. doi:10.1039/c4ay00571f
- De Juan, A., and Tauler, R. (2021). Multivariate curve resolution: 50 years addressing the mixture analysis problem - a review. *Anal. Chim. Acta* 1145, 59–78. doi:10.1016/j.aca.2020.10.051
- Dufour, A., Shinin, V., Tajbakhsh, S., Guillén-Aghion, N., Olivo-Marin, J. C., Zimmer, C., et al. (2005). Segmenting and tracking fluorescent cells in dynamic 3-D

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2022.933897/full#supplementary-material>

microscopy with coupled active surfaces. *IEEE Trans. Image Process.* 14, 1396–1410. doi:10.1109/tip.2005.852790

Eelbode, T., Bertels, J., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., et al. (2020). Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index. *IEEE Trans. Med. Imaging* 39, 3679–3690. doi:10.1109/TMI.2020.3002417

El-Mashtoly, S. F., Niedieker, D., Petersen, D., Krauss, S. D., Freier, E., Maghnouj, A., et al. (2014). Automated identification of subcellular organelles by coherent anti-Stokes Raman scattering. *Biophys. J.* 106, 1910–1920. doi:10.1016/j.bpj.2014.03.025

Freudiger, C. W., Min, W., Saar, B. G., Lu, S., Holtom, G. R., He, C., et al. (2008). Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science* 322, 1857–1861. doi:10.1126/science.1165758

Getreuer, P. (2012). Chan-Vese segmentation. *Image Process. Line* 2, 214–224. doi:10.5201/ipl.2012.g-cv

Ghaffari, M., Chateigner-Boutin, A. L., Guillon, F., Devaux, M. F., Abdollahi, H., Duponchel, L., et al. (2019). Multi-excitation hyperspectral autofluorescence imaging for the exploration of biological samples. *Anal. Chim. Acta* 1062, 47–59. doi:10.1016/j.aca.2019.03.003

Guerenne-Del Ben, T., Couderc, V., Duponchel, L., Sol, V., Leproux, P., and Petit, J. M. (2020). Multiplex coherent anti-Stokes Raman scattering microspectroscopy detection of lipid droplets in cancer cells expressing TrkB. *Sci. Rep.* 10, 16749. doi:10.1038/s41598-020-74021-z

Guerenne-Del Ben, T., Rajaofara, Z., Couderc, V., Sol, V., Kano, H., Leproux, P., et al. (2019). Multiplex coherent anti-Stokes Raman scattering highlights state of chromatin condensation in CH region. *Sci. Rep.* 9, 13862. doi:10.1038/s41598-019-50453-0

Hermesen, M., de Bel, T., den Boer, M., Steenbergen, E. J., Kers, J., Florquin, S., et al. (2019). Deep learning-based histopathologic assessment of kidney tissue. *J. Am. Soc. Nephrol.* 30, 1968–1979. doi:10.1681/ASN.2019020144

Hobro, A. J., and Smith, N. I. (2017). An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. *Vib. Spectrosc.* 91, 31–45. doi:10.1016/j.vibspec.2016.10.012

Hoffman, E. A., Frey, B. L., Smith, L. M., and Auble, D. T. (2015). Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J. Biol. Chem.* 290, 26404–26411. doi:10.1074/jbc.R115.651679

Houhou, R., Barman, P., Schmitt, M., Meyer, T., Popp, J., Bocklitz, T., et al. (2020). Deep learning as phase retrieval tool for CARS spectra. *Opt. Express* 28, 21002–21024. doi:10.1364/OE.390413

- Ichimura, T., Chiu, L., Fujita, K., Machiyama, H., Kawata, S., Watanabe, T. M., et al. (2015). Visualizing the appearance and disappearance of the attractor of differentiation using Raman spectral imaging. *Sci. Rep.* 5, 11358. doi:10.1038/srep11358
- Jungwirth, P. (2015). Biological water or rather water in biology? *J. Phys. Chem. Lett.* 6, 2449–2451. doi:10.1021/acs.jpcclett.5b01143
- Kaneta, D., Goto, M., Hagihara, M., Leproux, P., Couderc, V., Egawa, M., et al. (2021). Visualizing intra-medulla lipids in human hair using ultra-multiplex CARS, SHG, and THG microscopy. *Analyst* 146, 1163–1168. doi:10.1039/d0an01880e
- Kano, H., and Hamaguchi, H. (2005). Ultrabroadband ($> 2500 \text{ cm}^{-1}$) multiplex coherent anti-Stokes Raman scattering microspectroscopy using a supercontinuum generated from a photonic crystal fiber. *Appl. Phys. Lett.* 86, 121113. doi:10.1063/1.1883714
- Kano, H., Maruyama, T., Kano, J., Oka, Y., Kaneta, D., Guerenne, T., et al. (2019). Ultra-multiplex CARS spectroscopic imaging with 1-millisecond pixel dwell time. *OSA Contin.* 2, 1693. doi:10.1364/OSAC.2.001693
- Kee, T. W., and Cicerone, M. T. (2004). Simple approach to one-laser, broadband coherent anti-Stokes Raman scattering microscopy. *Opt. Lett.* 29, 2701–2703. doi:10.1364/OL.29.002701
- Kim, S. O., Kim, J., Okajima, T., and Cho, N. J. (2017). Mechanical properties of paraformaldehyde-treated individual cells investigated by atomic force microscopy and scanning ion conductance microscopy. *Nano Conver.* 4, 5. doi:10.1186/s40580-017-0099-9
- Klein, K., Gigler, A. M., Aschenbrenner, T., Monetti, R., Bunk, W., Jamitzky, F., et al. (2012). Label-free live-cell imaging with confocal Raman microscopy. *Biophys. J.* 102, 360–368. doi:10.1016/j.bpj.2011.12.027
- Krause, M., Radt, B., Rösch, P., and Popp, J. (2007). The investigation of single bacteria by means of fluorescence staining and Raman spectroscopy. *J. Raman Spectrosc.* 38, 369–372. doi:10.1002/jrs.1721
- Lawson, C. L., and Hanson, R. J. (1974). *Solving least squares problems*. New York: Prentice-Hall.
- Lee, Y. J., Vega, S. L., Patel, P. J., Aamer, K. A., Moghe, P. V., Cicerone, M. T., et al. (2014). Quantitative, label-free characterization of stem cell differentiation at the single-cell level by broadband coherent anti-Stokes Raman scattering microscopy. *Tissue Eng. Part C Methods* 20, 562–569. doi:10.1089/ten.TEC.2013.0472
- Lu, F. K., Basu, S., Igras, V., Hoang, M. P., Ji, M., Fu, D., et al. (2015). Label-free DNA imaging *in vivo* with stimulated Raman scattering microscopy. *Proc. Natl. Acad. Sci. U. S. A.* 112, 11624–11629. doi:10.1073/pnas.1515121112
- Masia, F., Glen, A., Stephens, P., Borri, P., and Langbein, W. (2013). Quantitative chemical imaging and unsupervised analysis using hyperspectral coherent anti-Stokes Raman scattering microscopy. *Anal. Chem.* 85, 10820–10828. doi:10.1021/ac402303g
- Maška, M., Daněk, O., Garasa, S., Rouzaut, A., Muñoz-Barrutia, A., Ortiz-de-Solorzano, C., et al. (2013). Segmentation and shape tracking of whole fluorescent cells based on the Chan-Vese model. *IEEE Trans. Med. Imaging* 32, 995–1006. doi:10.1109/TMI.2013.2243463
- Matthäus, C., Bird, B., Miljković, M., Chernenko, T., Romeo, M., Diem, M., et al. (2008). Chapter 10: Infrared and Raman microscopy in cell biology. *Methods Cell Biol.* 89, 275–308. doi:10.1016/S0091-679X(08)00610-9
- Matthews, Q., Jirasek, A., Lum, J., Duan, X., and Brolo, A. G. (2010). Variability in Raman spectra of single human tumor cells cultured *in vitro*: correlation with cell cycle and culture confluency. *Appl. Spectrosc.* 64, 871–887. doi:10.1366/000370210792080966
- Nuriya, M., Yoneyama, H., Takahashi, K., Leproux, P., Couderc, V., Yasui, M., et al. (2019). Characterization of intra/extracellular water states probed by ultrabroadband multiplex coherent anti-Stokes Raman scattering (CARS) spectroscopic imaging. *J. Phys. Chem. A* 123, 3928–3934. doi:10.1021/acs.jpca.9b03018
- Palonpon, A., Ando, J., Yamakoshi, H., Dodo, K., Sodeoka, M., Kawata, S., et al. (2013). Raman and SERS microscopy for molecular imaging of live cells. *Nat. Protoc.* 8, 677–692. doi:10.1038/nprot.2013.030
- Parekh, S. H., Lee, Y. J., Aamer, K. A., and Cicerone, M. T. (2010). Label-free cellular imaging by broadband coherent anti-Stokes Raman scattering microscopy. *Biophys. J.* 99, 2695–2704. doi:10.1016/j.bpj.2010.08.009
- Pederson, T. (2011). The nucleolus. *Cold Spring Harb. Perspect. Biol.* 3, a000638. doi:10.1101/cshperspect.a000638
- Pisapia, C., Jamme, F., Duponchel, L., and Ménez, B. (2018). Tracking hidden organic carbon in rocks using chemometrics and hyperspectral imaging. *Sci. Rep.* 8, 2396. doi:10.1038/s41598-018-20890-4
- Pohling, C., Buckup, T., and Motzkus, M. (2011). Hyperspectral data processing for chemoselective multiplex coherent anti-Stokes Raman scattering microscopy of unknown samples. *J. Biomed. Opt.* 16, 021105. doi:10.1117/1.3533309
- Rao, C., Verma, N. C., and Nandi, C. K. (2019). Unveiling the hydrogen bonding network of intracellular water by fluorescence lifetime imaging microscopy. *J. Phys. Chem. C* 123, 2673–2677. doi:10.1021/acs.jpcc.8b12439
- Shi, L., Hu, F., and Min, W. (2019). Optical mapping of biological water in single live cells by stimulated Raman excited fluorescence microscopy. *Nat. Commun.* 10, 4764. doi:10.1038/s41467-019-12708-2
- Shi, Z., Tang, W., Duren, Z., and Jiang, Z. (2013). Subspace matching pursuit for sparse unmixing of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* 52, 3256–3274. doi:10.1109/TGRS.2013.2272076
- Smith, R., Wright, K. L., and Ashton, L. (2016). Raman spectroscopy: an evolving technique for live cell studies. *Analyst* 141, 3590–3600. doi:10.1039/C6AN00152A
- Tan, Y., Jin, Y., Zhao, P., Wu, J., and Ren, Z. (2021). Lipid droplets contribute myogenic differentiation in C2C12 by promoting the remodeling of the actin filament. *Cell Death Dis.* 12, 1102. doi:10.1038/s41419-021-04273-8
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika* 18, 267–276. doi:10.1007/BF02289263
- Vartiainen, E. M., Rinia, H. A., Müller, M., and Bonn, M. (2006). Direct extraction of Raman line-shapes from congested CARS spectra. *Opt. Express* 14, 3622–3630. doi:10.1364/OE.14.003622
- Windig, W., and Guilment, J. (1991). Interactive self-modeling mixture analysis. *Anal. Chem.* 63, 1425–1432. doi:10.1021/ac00014a016
- Zhang, D., Wang, P., Slipchenko, M. N., Ben-Amotz, D., Weiner, A. M., Cheng, J. X., et al. (2013). Quantitative vibrational imaging by hyperspectral stimulated Raman scattering microscopy and multivariate curve resolution analysis. *Anal. Chem.* 85, 98–106. doi:10.1021/ac3019119
- Zumbusch, A., Holtom, G. R., and Xie, X. S. (1999). Three-dimensional vibrational imaging by coherent anti-Stokes Raman scattering. *Phys. Rev. Lett.* 82, 4142–4145. doi:10.1103/PhysRevLett.82.4142