



# DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

## Reliable Cell Segmentation Based on Spectral Phasor Analysis of Hyperspectral Stimulated Raman Scattering Imaging Data

The Harvard community has made this article openly available.  
[Please share](#) how this access benefits you. Your story matters.

Citation	Anal. Chem. 2014, 86, 41154119
Accessed	January 29, 2016 1:44:01 PM EST
Citable Link	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:17948082">http://nrs.harvard.edu/urn-3:HUL.InstRepos:17948082</a>
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a>

*(Article begins on next page)*

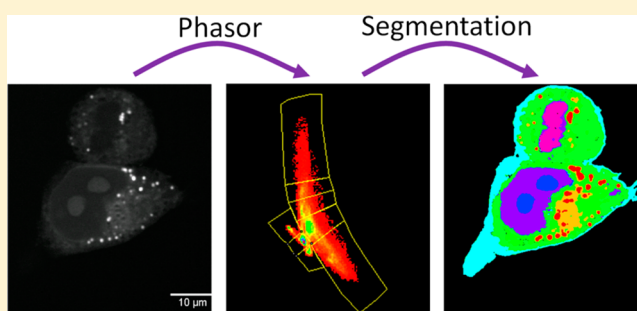
# Reliable Cell Segmentation Based on Spectral Phasor Analysis of Hyperspectral Stimulated Raman Scattering Imaging Data

Dan Fu and X. Sunney Xie\*

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

**S** Supporting Information

**ABSTRACT:** Hyperspectral stimulated Raman scattering (SRS) imaging has rapidly become an emerging tool for high content analyses of cell and tissue systems. The label-free nature of SRS imaging combined with its chemical specificity allows *in situ* and *in vivo* biochemical quantification at submicrometer resolution without sectioning and staining. Current hyperspectral SRS data analysis methods are based on either linear unmixing or multivariate analysis, which are not sensitive to small spectral variations and often provide obscure information on the cell composition. Here, we demonstrate a spectral phasor analysis method that allows fast and reliable cellular organelle segmentation of mammalian cells, without any *a priori* knowledge of their composition or basis spectra. We further show that, in combination with a branch-bound algorithm for optimal selection of a few wavenumbers, spectral phasor analysis provides a robust solution to label-free single cell analysis.



Raman spectroscopy is a powerful technique for chemical identification and quantification. Many biological applications of Raman microspectroscopic imaging have benefited greatly from its label-free nature and chemical specificity.<sup>1–5</sup> Recent developments in confocal Raman imaging have enabled visualization of many different subcellular organelles with increased contrast.<sup>6,7</sup> However, the imaging speed is fundamentally limited by the weak Raman scattering cross-section. Rapid developments in coherent Raman microscopy in recent years provide excellent alternatives to confocal Raman imaging with substantial improvement in imaging speed.<sup>8,9</sup> Two different coherent Raman imaging techniques are commonly used: coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS). While both techniques are capable of imaging cells and tissue at submicrometer resolution with high speed, SRS offers a distinctive advantage in preserving the spectral fidelity of Raman and provides easier quantification without any need for a complicated spectral retrieval procedure.<sup>10,11</sup>

Hyperspectral SRS (hsSRS) imaging requires either scanning the Raman shift sequentially or probe discrete spectral regions of the Raman spectrum simultaneously (also known as multiplex).<sup>12</sup> Most recently, a number of groups have demonstrated hsSRS imaging based on spectral scanning.<sup>13–16</sup> Different lasers and scanning mechanisms were used to achieve sequential Raman shift tuning to reconstruct Raman spectra at individual pixels. The spectral noise associated with sequential scanning may reduce the spectral fidelity. Nonetheless, all methods have been shown to be able to enhance subcellular organelle contrasts using either linear analysis or multivariate analysis in the C–H stretching region. Linear analysis allows

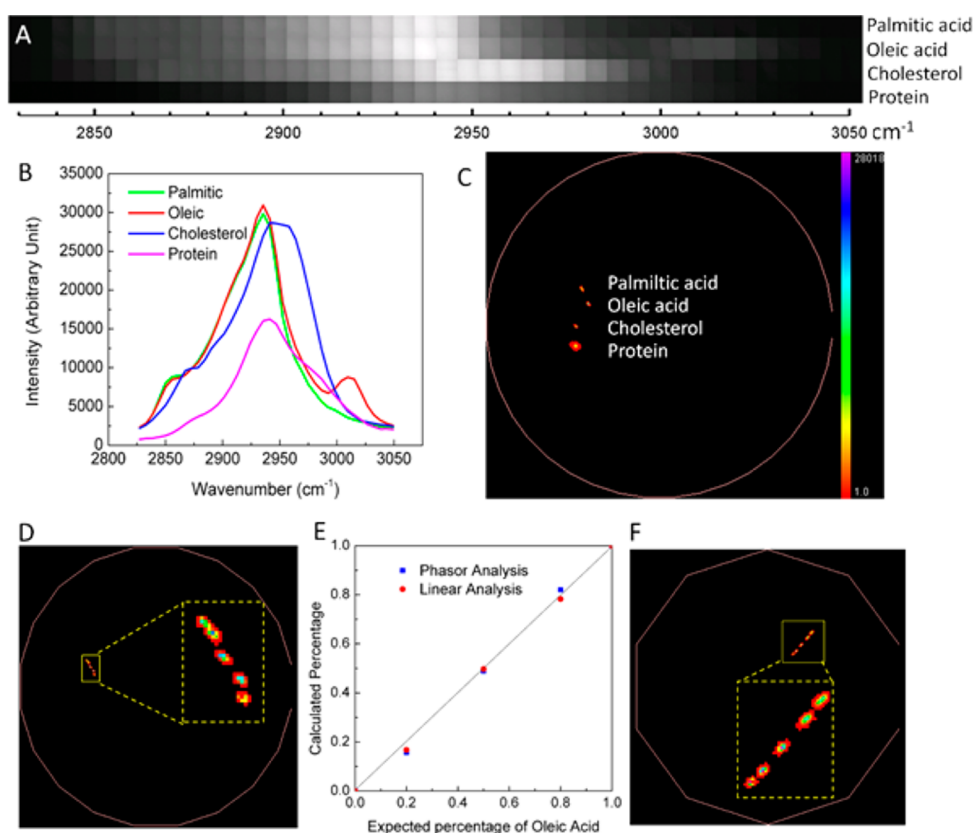
direct mapping of chemical composition provided that *a priori* knowledge of system composition and spectral calibration of individual chemical species are available. In comparison, multivariate analysis such as principle component analysis does not require *a priori* information, but it results in maps that may not directly correspond to individual chemical components or cellular components, which can complicate interpretation. Here, we report a spectral phasor analysis method that provides fast and reliable subcellular organelle segmentation. In addition, we demonstrate that effective data reduction without sacrifice in segmentation quality can be achieved with a branch/bound algorithm for optimal selection of wavenumbers, allowing for more than 10-fold speedup in imaging.

Phasor analysis in optical imaging was first introduced by Clayton for the analysis of fluorescence lifetime imaging (FLIM) data.<sup>17</sup> Later on, Gratton demonstrated its great potential in simplifying the analysis of FLIM and FRET images.<sup>18,19</sup> FLIM data collected in the time domain often exhibit complex nonexponential decay behavior. Simple exponential fitting is prone to error and requires expertise. The phasor analysis transforms the histogram of the time delays at each pixel to a single point on a polar plot, which significantly simplifies the analysis and presents a graphical global view of the processes affecting the fluorescence decay occurring at each pixel. Because the phasor analysis is a data representation method, it is not restricted to fluorescence lifetime data.

**Received:** December 20, 2013

**Accepted:** March 31, 2014

**Published:** March 31, 2014



**Figure 1.** (A) hsSRS images of four different chemical species in solutions: palmitic acid, oleic acid, cholesterol, and protein. Images span from 2830 to 3050  $\text{cm}^{-1}$  with a total of 40 frames; (B) Average SRS spectra of the four chemical species; (C) Phasor plot of the four hsSRS data sets; (D) Phasor plot of the 20-frame hsSRS images of five palmitic acid/oleic acid mixture solutions. The dotted yellow box shows a zoomed-in image of the phasor clusters; (E) Comparing the calculated percentage of oleic acid in the mixtures using phasor analysis with that from using linear analysis; (F) Phasor plot of the five solutions using only 10 frames from 2940–3050  $\text{cm}^{-1}$ .

Recently, it has been adapted successfully for fluorescence spectral imaging and pump–probe imaging.<sup>20,21</sup> Polar plot representation has been used previously for hyperspectral CARS imaging to show nonresonant background induced phase-variation.<sup>22</sup> However, in that case, each point on the polar plot represents a single Raman frequency, not the entire spectrum, which is different from the phasor analysis.

hsSRS imaging with its label-free nature and rich spectral information naturally lends itself to the phasor analysis. The three-dimensional (3D) hsSRS data stack can be transformed into a two-dimensional phasor plot based on simple Fourier Transform (for details, see Supporting Information). Each point on the phasor plot (also called the phasor) corresponds to a SRS spectrum at a particular pixel. The spatial closeness of any two phasors is determined by their spectral similarity. Therefore, a cluster of phasors can be mapped back to an area in the sample image with closely resembled Raman spectra. By dividing the phasor plot into a number of clusters, a sample image can be segmented into different areas grouped by their Raman spectra. We adapted an ImageJ macro developed by Fereidouni<sup>20</sup> to perform phasor analysis on our hsSRS data acquired by a spectral focusing SRS setup.<sup>16</sup>

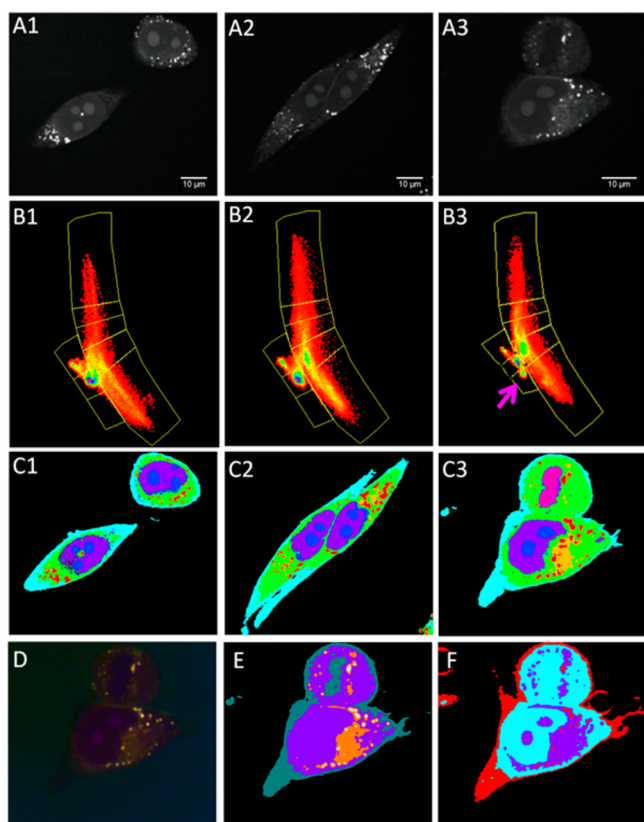
We first evaluate the capability of spectral phasor analysis in distinguishing different chemicals in solution. hsSRS imaging data spanning the C–H stretching region were acquired (see Supporting Information on data acquisition). Four chemical species purchased from Sigma-Aldrich were measured: oleic acid, palmitic acid, cholesterol, and protein (bovine serum

albumin). Figure 1A shows the hsSRS images of the four chemicals in  $\text{CDCl}_3$  solution at 10% w/v concentration (except for protein, which is dissolved in deuterated water). The average SRS spectra of each species are plotted out in Figure 1B. They all have broad Raman spectra within the C–H region but differ in spectral shape. Importantly, the four solutions can be easily distinguished as four spatially separated clusters in the phasor plot (Figure 1C). The protein cluster shows a larger spread due to its lower SRS intensity.

Next, we demonstrate that, in the phasor plot, the phasor of a mixture of two different species exhibits a linear relationship with the phasors of the two individual species.<sup>20</sup> We acquired 20-frame hsSRS images of oleic acid (10% w/v in  $\text{CDCl}_3$ ) and palmitic acid (10% w/v in  $\text{CDCl}_3$ ) mixed at five different ratios: 100%/0%, 80%/20%, 50%/50%, 20%/80%, and 0%/100%. Each frame was  $256 \times 256$  pixels. The SRS spectra of oleic acid and palmitic acid only differ slightly at the unsaturated C–H stretching peak 3010  $\text{cm}^{-1}$ . In order to reduce the phasor cluster size, the hsSRS images were averaged down to  $128 \times 128$  pixels. Figure 1D shows the phasor plot of the reduced hyperspectral data sets. Five clusters are observed. They mostly lie along a line with the top and bottom clusters representing palmitic acid and oleic acid, respectively. To quantify the concentration ratio, we located the central point of each cluster and measured the relative distances of each cluster to the clusters of oleic acid and palmitic acid. The result was compared to simple linear fitting (Figure 1E). We note that the quantification capability of a single phasor plot is marginally

weaker than that of simple linear analysis. This is because phasor analysis is a data reduction procedure and the phase plot contains much less information compared to the full hyperspectral data set. Interestingly, we found that phasor analysis of the hsSRS data containing only the last 10 frames from 2940 to 3050  $\text{cm}^{-1}$  (where the largest spectral differences between oleic and palmitic acids are found) can better quantify the concentration compared with that obtained from the full 20-frame hyperspectral data set (Figure 1F). This interesting finding suggests added flexibility in data acquisition because current hsSRS imaging systems perform discrete spectral scanning instead of parallel spectral acquisition.

To investigate the capability of spectral phasor analysis in cell biochemical and morphological characterization, we acquired hsSRS data of formaldehyde fixed mammalian cells from 2810 to 3090  $\text{cm}^{-1}$  for a total of 60 frames. The SRS images and corresponding phasor plots (zoomed-in images) of three different cell samples are shown in Figure 2A,B. Even though the cells show different morphology, their phasor plots exhibit remarkable similarity. The biochemical composition of a cell is



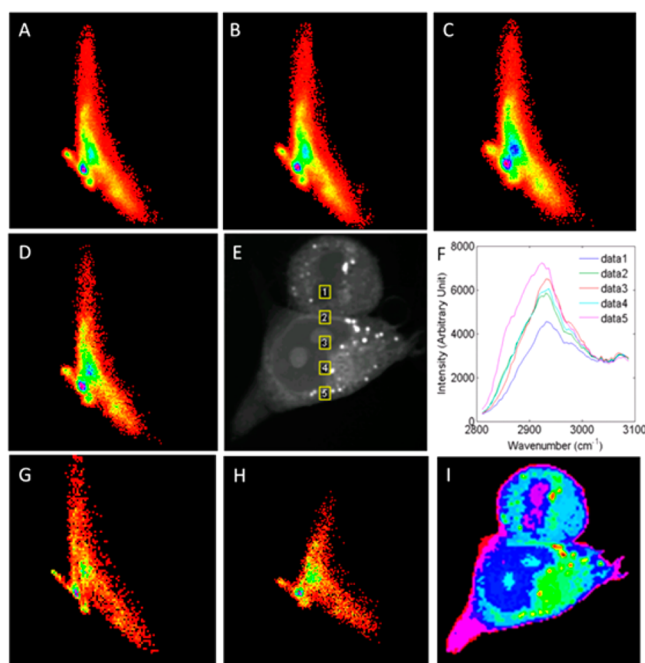
**Figure 2.** (A1–A3) Maximum intensity projection (projecting the pixel value with the highest intensity along the spectral axis onto a 2D image) images of 60-frame hsSRS images from 2810–3090  $\text{cm}^{-1}$  of three different cell samples; (B1–B3) Zoomed-in phasor plots of hsSRS data of the three samples. Yellow boxes show manual segmentation of the plots based on their clustering pattern; (C1–C3) Color-coded cell images based on segmented phasor plots; (D) Principle component analysis of the hsSRS data set of 2A3. Only the first three principle components are shown due to the excessive noise of latter components; (E) Cell segmentation using K-means cluster analysis with five clusters. More clusters result in background segmented into two or more regions; (F) Cell segmentation using Ward's agglomerative cluster analysis with five clusters.

mostly lipids, protein, nucleic acids, and water. Phasor analysis provides a graphical way of globally visualizing the spectral variations of a cell due to compositional change and allows users to interactively segment a cell based on the pattern of the phasor plot. For the purpose of illustrating the important features of a spectral phasor plot, we chose to use manual segmentation (details of our rationale for segmentation is shown in Supporting Information). The most notable features of the phasor plots in Figure 2B are the two long tails. These long tails are due to limited spatial resolution of SRS imaging (Figure S1, Supporting Information). The pixels near the surface of a lipid droplet or the surface of the cell show up with varying degree of contributions from contents of the two sides of the surface. As we have shown previously, phasors of mixed species lie along a line with the two ends of the line determined by the phasors of the two species. In the cell phasor plots, the top long tails arise from varying lipid/protein composition near the lipid droplet surface, and the bottom right long tails arise from varying protein/water composition near the cell surface. We segmented each long tail as one region of interest (ROI). The rest were divided into four or five ROIs based on their clustering patterns (the exact sizes of clusters are not critical for segmentation, see Supporting Information and Figures S1–S2). The segmented phasor plots of the three samples are shown in Figure 2B1–3. Each ROI is assigned a different color, which color-codes the cellular region with phasors located in that ROI. Figure 2C1–3 shows the color-coded cell images using this manual segmentation method. The color-coding clearly delineates many subcellular components of the cell based on their morphological features and established SRS imaging characteristics (Figure S3, Supporting Information). For example, red, violet, blue, and cyan colors show lipid droplets, nucleus, nucleoli, and cell membrane, respectively. Interestingly, we note that the phasor plot of the third sample has an additional cluster (pointed out with a pink arrow) that does not exist in the other two samples. That cluster arises from the condensed chromosomes of a cell in its mitotic phase, which changes the Raman spectrum when compared with that of an interphase cell nucleus. The spectral change may be due to a combination of increasing nucleic acid to protein ratio and changing protein density. Although there is no obvious spectral peak shift, the high sensitivity of phasor to spectral variation enables reliable cell segmentation, even for the difficult case of identifying mitotic cells.

To compare the performance of spectral phasor analysis with multivariate analysis,<sup>23,24</sup> we performed principle component analysis, K-means cluster analysis, and Ward's agglomerative cluster analysis of the hsSRS data set corresponding to Figure 2A3 (see details in Supporting Information). The results are shown in Figure 2D–F, respectively. In all three cases, the nucleus and mitotic chromosome features are not delineated as clearly as that of spectral phasor analysis (Figure 2C3). In addition, spectral phasor analysis has significant speed advantage: it is ten times faster than K-means clustering analysis and more than a thousand times faster than Ward's agglomerative cluster analysis.

It is important to note that large hsSRS data sets are not required for cell segmentation. We demonstrate this by down-sampling our previously acquired hsSRS data set. Figure 3A–C shows the phasor plot of the sample shown in Figure 2A3 down-sampled to 30, 20, and 10 frames, respectively. Clearly, the overall shape of the phasor plots does not change with down-sampling, but there is an increase of nucleus and





**Figure 3.** (A–C) The effect of wavenumber sampling frequency on phasor analysis. Phasor plots of the 60-frame hsSRS data set of the cell in Figure 2A3 is down-sampled to 30, 20, and 10 frames, respectively; (D) Phasor plot of the 10-frame hsSRS data with  $2 \times 2$  spatial averaging; (E) Sampling of SRS spectra at five different locations in the cell marked with yellow boxes; (F) Corresponding average spectra of the five ROIs; (G) Phasor plot of the 60-frame hsSRS data with  $4 \times 4$  spatial averaging; (H) Phasor plot of 4-frame SRS images at 2852, 2932, 2956, and 3088  $\text{cm}^{-1}$  with  $4 \times 4$  spatial averaging; (I) Color-coded cell image mapped by automatic segmentation of H into 64 regions.

nucleolus cluster size, which decreases the ability of subcellular segmentation. This can be partly compensated by spatial averaging (Figure 3D). It is possible to use only a few selected frames to segment the cell. However, optimal segmentation requires acquisition of SRS images that reflect the largest spectral variations among subcellular components. We selected five different ROIs and used their SRS spectra to determine optimal wavenumbers (Figure 3E,F). We adapted a branch/bound algorithm with determinant criterion that has been previously used for optimal wavelength selection for multi-component spectrophotometry.<sup>25</sup> The branch/bound algorithm for finding the optimal set of wavelengths involves partition of feasible wavelength combination sets into subsets and comparison of their upper limits of determinants. The partitioning of subsets effectively reduces the number of calculations needed. The determinant criterion is similar to the mean square error criteria commonly used for error assessment but with improved computational advantage. Using this algorithm, we obtained four optimal wavenumbers: 2852, 2932, 2956, and 3088  $\text{cm}^{-1}$ . The four SRS frames corresponding to these wavenumbers were then used for phasor analysis. To reduce noise,  $4 \times 4$  spatial averaging of the four frames was applied.

The phasor plot is shown in Figure 3H. When compared to that obtained from  $4 \times 4$  spatial averaging of the full 60 frames hyperspectral data (Figure 3G), we observe a change in phasor shape, but the major clustering features are well-preserved. An automatic segmentation of the 4-frame phasor plot into 64

ROIs results in a color-coded image (Figure 3I) similar to Figure 2C3, proving the validity of using only four selected SRS frames for cell segmentation. With our spectral focusing SRS approach, this can be easily implemented by moving the mechanical delay through a few predetermined positions.

In summary, to the best of our knowledge, we demonstrate the first use of spectral phasor analysis for hyperspectral Raman data and show that high wavenumber hsSRS imaging in the C–H stretching region (2800–3100  $\text{cm}^{-1}$ ) is useful in distinguishing many different subcellular components including lipid droplets, nucleus, nucleoli, and membrane structures. The method is more effective than other multivariate analysis methods in subcellular segmentation and provides faster speed. Live cell analysis requires either faster imaging at video-rate or line-by-line frequency tuning to reduce motion artifact.<sup>14,26</sup> Because SRS imaging provides intrinsic 3D optical sectioning, the spectral phasor analysis should be equally applicable to tissue as well. It is important to point out that high spectral resolution hyperspectral data is not required for segmentation purposes; a few Raman images at discrete wavenumbers that capture the largest spectral difference of chemical components are as effective as full high-resolution hyperspectral data in distinguishing major subcellular structures. This allows significant speedup of the imaging process, which can be readily implemented with spectral-scanning hsSRS, but not confocal Raman or hyperspectral CARS. Nonetheless, the spectral phasor analysis method presented here is applicable to vibrational data acquired by any other method including hyperspectral CARS, confocal Raman, and infrared imaging, independent of their spectral coverage and spectral resolution. However, a linear relationship between spectral intensity and molecular concentrations is a prerequisite in spectral phasor analysis. We believe that spectral phasor based cell segmentation will be useful in biological and biomedical imaging based on many different forms of vibrational microscopy. With the high sensitivity of phasor to small spectral variation, the spectral phasor analysis could also be used to monitor biochemical changes of cells or cell–drug interactions.

## ■ ASSOCIATED CONTENT

### Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [xie@chemistry.harvard.edu](mailto:xie@chemistry.harvard.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work is supported by the U.S. Department of Energy's Basic Energy Sciences Program (DE-FG02-09ER16104). We thank Professor Enrico Gratton for helpful discussions. We thank Dr. Fereidouni for kindly providing the ImageJ plugin - Spectral Phasor.

## ■ REFERENCES

- (1) Puppels, G. J.; de Mul, F. F. M.; Otto, C.; Greve, J.; Robert-Nicoud, M.; Arndt-Jovin, D. J.; Jovin, T. M. *Nature* **1990**, *347*, 301–303.

- (2) Uzunbajakava, N.; Lenferink, A.; Kraan, Y.; Volokhina, E.; Vrensen, G.; Greve, J.; Otto, C. *Biophys. J.* **2003**, *84*, 3968–3981.
- (3) Caspers, P. J.; Lucassen, G. W.; Carter, E. A.; Bruining, H. A.; Puppels, G. J. *J. Invest. Dermatol.* **2001**, *116*, 434–442.
- (4) Parekh, S. H.; Lee, Y. J.; Aamer, K. A.; Cicerone, M. T. *Biophys. J.* **2010**, *99*, 2695–2704.
- (5) Haka, A. S.; Shafer-Peltier, K. E.; Fitzmaurice, M.; Crowe, J.; Dasari, R. R.; Feld, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12371–12376.
- (6) Klein, K.; Gigler, A. M.; Aschenbrenner, T.; Monetti, R.; Bunk, W.; Jamitzky, F.; Morfill, G.; Stark, R. W.; Schlegel, J. *Biophys. J.* **2012**, *102*, 360–368.
- (7) Okada, M.; Smith, N. I.; Palonpon, A. F.; Endo, H.; Kawata, S.; Sodeoka, M.; Fujita, K. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 28–32.
- (8) Evans, C. L.; Potma, E. O.; Puoris'haag, M.; Cote, D.; Lin, C. P.; Xie, X. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16807–16812.
- (9) Saar, B. G.; Freudiger, C. W.; Reichman, J.; Stanley, C. M.; Holtom, G. R.; Xie, X. S. *Science* **2010**, *330*, 1368–1370.
- (10) Freudiger, C. W.; Min, W.; Saar, B. G.; Lu, S.; Holtom, G. R.; He, C.; Tsai, J. C.; Kang, J. X.; Xie, X. S. *Science* **2008**, *322*, 1857–1861.
- (11) Day, J. P. R.; Domke, K. F.; Rago, G.; Kano, H.; Hamaguchi, H.-o.; Vartiainen, E. M.; Bonn, M. *J. Phys. Chem. B* **2011**, *115*, 7713–7725.
- (12) Fu, D.; Lu, F. K.; Zhang, X.; Freudiger, C.; Pernik, D. R.; Holtom, G.; Xie, X. S. *J. Am. Chem. Soc.* **2012**, *134*, 3623–3626.
- (13) Suhailim, J. L.; Chung, C.-Y.; Lilledahl, M. B.; Lim, R. S.; Levi, M.; Tromberg, B. J.; Potma, E. O. *Biophys. J.* **2012**, *102*, 1988–1995.
- (14) Ozeki, Y.; Umemura, W.; Otsuka, Y.; Satoh, S.; Hashimoto, H.; Sumimura, K.; Nishizawa, N.; Fukui, K.; Itoh, K. *Nat. Photonics* **2012**, *6*, 845–851.
- (15) Zhang, D.; Wang, P.; Slipchenko, M. N.; Ben-Amotz, D.; Weiner, A. M.; Cheng, J.-X. *Anal. Chem.* **2013**, *85*, 98–106.
- (16) Fu, D.; Holtom, G.; Freudiger, C.; Zhang, X.; Xie, X. S. *J. Phys. Chem. B* **2013**, *117*, 4634–4640.
- (17) Clayton, A. H. A.; Hanley, Q. S.; Verveer, P. J. *J. Microsc.* **2004**, *213*, 1–5.
- (18) Digman, M. A.; Caiolfa, V. R.; Zama, M.; Gratton, E. *Biophys. J.* **2008**, *94*, L14–L16.
- (19) Stringari, C.; Cinquin, A.; Cinquin, O.; Digman, M. A.; Donovan, P. J.; Gratton, E. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13582–13587.
- (20) Fereidouni, F.; Bader, A. N.; Gerritsen, H. C. *Opt. Express* **2012**, *20*, 12729–12741.
- (21) Robles, F. E.; Wilson, J. W.; Fischer, M. C.; Warren, W. S. *Opt. Express* **2012**, *20*, 17082–17092.
- (22) Jurna, M.; Garbacik, E. T.; Korterik, J. P.; Herek, J. L.; Otto, C.; Offerhaus, H. L. *Anal. Chem.* **2010**, *82*, 7656–7659.
- (23) Miljkovic, M.; Chernenko, T.; Romeo, M. J.; Bird, B.; Matthaus, C.; Diem, M. *Analyst* **2010**, *135*, 2002–2013.
- (24) Shinzawa, H.; Awa, K.; Kanematsu, W.; Ozaki, Y. *J. Raman Spectrosc.* **2009**, *40*, 1720–1725.
- (25) Yi-Zeng, L.; Yu-Long, X.; Ru-Qin, Y. *Anal. Chim. Acta* **1989**, *222*, 347–357.
- (26) Kong, L.; Ji, M.; Holtom, G. R.; Fu, D.; Freudiger, C. W.; Xie, X. S. *Opt. Lett.* **2013**, *38*, 145–147.