Discrimination of rectal cancer through human serum using surface-enhanced Raman spectroscopy

Xiaozhou Li[1,2] · Tianyue Yang[2] · Siqi Li[3] · Su Zhang[4] · Lili Jin[5]

Abstract In this paper, surface-enhanced Raman spectroscopy (SERS) was used to detect the changes in blood serum components that accompany rectal cancer. The differences in serum SERS data between rectal cancer patients and healthy controls were examined. Postoperative rectal cancer patients also participated in the comparison to monitor the effects of cancer treatments. The results show that there are significant variations at certain wavenumbers which indicates alteration of corresponding biological substances. Principal component analysis (PCA) and parameters of intensity ratios were used on the original SERS spectra for the extraction of featured variables. These featured variables then underwent linear discriminant analysis (LDA) and classification and regression tree (CART) for the discrimination analysis. Accuracies of 93.5 and 92.4 % were obtained for PCA–LDA and parameter-CART, respectively.

✉ Xiaozhou Li
biophy@163.com

1   School of Science, Shenyang Ligong University, 110159 Shenyang, China

2   School of Physics, Liaoning University, 110036 Shenyang, China

3   Integrated Life Sciences, Kent State University, Kent, OH 44242, USA

4   Department of Physiology and Biochemistry, Shenyang Sport University, 110102 Shenyang, China

5   School of Life Science, Liaoning University, 110036 Shenyang, China

1 Introduction

Colorectal cancer is the third most common cancer in the USA, and its incidence increases with age [1]. The survival rate for rectal cancer is over 90 % when diagnosed at a local stage, but only 40 % of patients are diagnosed that early. Survival rates decline to 70–12 % when the cancer develops to regional and distant-stage disease [2]. Early screening and intervention of colorectal cancer is an effective way to reduce its mortality [3]. Routine diagnosis can be made by CT, MRI, colonoscopy, etc., but examination methods are expensive, inconvenient and invasive [4].

Raman spectroscopy is a nondestructive method that allows the secondary conformation of biological molecules to be observed. It has shown to be a promising diagnostic technique in the identification of various organs [5]. For instance, the combination of three Raman methods (near-infrared Fourier transform Raman, NIR–visible Raman and Vis Raman) has been used on colon tissue samples, and evident spectral differences were detected in comparison with normal tissues [6]. FT-Raman spectroscopy has shown its ability to distinguish different types of colorectal tissues such as smooth muscle tissue, connective tissue and fatty and slack tissue [7]. Research using micro-Raman spectroscopy revealed that Raman spectroscopy and optical microscopic images of breast tumor cells and colorectal tumor cells have significant differences [8]. Blood as a detection target has been used by Raman spectroscopy for many kinds of cancers, and promising results have been achieved [9, 10]. Surface-enhanced Raman spectroscopy (SERS) can enhance ordinary Raman signals by 9–10 orders of magnitude due to the enhancement effect caused by its physical and chemical mechanism [11]. It can also greatly quench the interference of fluorescence which is common in biosamples. SERS has been used for serum discrimination of

colorectal cancer patients and the healthy [12]. Combined with certain immunoassays and other extraction techniques, certain proteins or nucleic acids in serum can be detected by SERS [13, 14]. Our research explores the feasibility and effectiveness of using surface-enhanced Raman spectroscopy to differentiate serum from healthy volunteers, rectal cancer patients and from the same patients postoperation.

Due to the complex structures of biological molecules, chemometric methods have often been applied to reveal the inner differences of Raman spectral features. Principal component analysis (PCA) is able to extract the main factors out of the whole SERS spectra range and can simplify the development of classification algorithms [15]. Linear discriminant analysis (LDA) is usually used on the principal components (PCs) retrieved from PCA for discrimination. Classification and regression tree (CART) technique is a classification method based on recursive partitioning. It can divide samples by judging certain variables selectively at each step and finally divides all samples after a certain number of steps. All of the above-stated statistical techniques have been utilized successfully on the analysis of Raman spectroscopy of biosamples [16, 17].

In this paper, we investigated SERS spectra of serum taken from 22 healthy volunteers and 35 rectal cancer patients both before and after operation. Peak variations between groups were analyzed, and tentative biomedical causes were explained. The two analytical methods of PCA–LDA and parameter-CART were each used to form classifiers of the obtained Raman spectra. Finally, tenfold cross-validation was used to verify the diagnostic effectiveness of the two algorithms.

2 Materials and methods

2.1 Sampling

In this paper, blood serum taken from 22 healthy volunteers and 35 rectal cancer patients (both before and after their cancer operation) was evaluated using surface-enhanced Raman spectroscopy. Table 1 is the demographic distribution of the sample donors. All samples were obtained from the Tumor Hospital of Liaoning Province. Samples were collected before breakfast in the morning. For the postoperation samples, serum was taken 7–8 days after operation (this is before chemotherapy, if any). Obtained venous blood was separated in a centrifuge at 3000 rot/min for 10 min. The upper serum was extracted and refrigerated hermetically at -80 °C prior to data collection.

2.2 Instrumentation

Silver hydrosol was synthesized via a deoxidizing method using trisodium citrate and silver nitrate heated in a

Table 1 Demographics of study population

Groups	Controls ($n = 22$)	Cancer ($n = 35$)
Age (years, SD)	36 (12)	41 (16)
Gender (% F)	55 %	49 %

microwave oven as described by Liu et al. [18]. First, 20 mg of silver nitrate was added to 100 mL of water and heated to boiling temperature in a microwave oven. Then, 2 mL of 1 % trisodium citrate was added into the solution. The whole mixture was boiled for 2 min. The resulting colloid was cooled in ambient conditions and showed a yellow gray color. This method can produce uniform silver nanoparticles due to the heating mode of the microwave oven. The sample was then prepared by ultrasonically mixing 2 μL serum with 2 μL Ag colloid in a centrifuge. The SERS spectrum was measured by $180°$ back-scattering. Data were collected from an inverted microscope (British Renishaw) in the range of 400–1800 cm^{-1}. Radiation of 632.8 nm created by a He–Ne laser maintained at 3.5 mW was used for excitation. The exposure time was 10 s.

2.3 Data analysis

2.3.1 Pretreatment

Prior to PCA–LDA and parameter-CART, all spectra were smoothed using the pspline package in R. Baseline correction and area normalization were conducted on each Raman spectrum using the hyperSpec package in R.

2.3.2 Peak analysis

SERS peaks between groups were compared using one-way analysis of variance (ANOVA). Unlike the t test which can only be used in two-group cases, one-way ANOVA is used to test for differences among at least three groups.

2.3.3 PCA–LDA

Principal component analysis (PCA) can orthogonally transform a large set of data into a much smaller set of uncorrelated variables that contain most of the original information. Because of the data reduction effect of PCA, further discriminant analysis can be simplified and more robust. Linear discriminant analysis (LDA) is a method that helps to discriminate different groups based on the distances between each samples. LDA was utilized on the principal components (PCs) in this paper. PCA and LDA were performed using the hyperSpec and MASS packages in R.
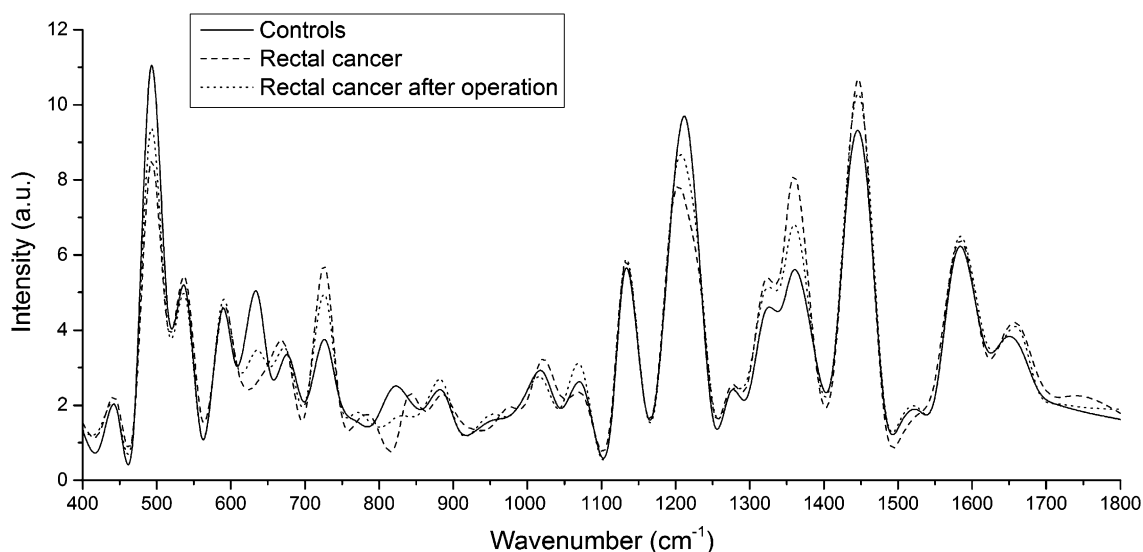
Fig. 1 Average SERS spectra of three groups—controls, rectal cancer patients before operation and after operation

2.3.4 Parameter-CART

For additional analysis, intensities of main peaks were selected as parameters and classification and regression tree (CART) was employed on those parameters. CART is a classification method which divides samples to smaller subsets in each level through the input parameters. The original group is called the root node, and the subsequent nodes are split from the preceding nodes. This splitting process continues until a set of terminal nodes which have no statistically significant splits or are too small to be split is reached. CART was performed using the rpart package in R.

3 Results and discussion

3.1 SERS

SERS data of serum taken from 22 normal people and 35 rectal cancer patients were measured. Three measurements were taken for each sample, and the averaged spectrum was recorded. Spectra were preprocessed by smoothing, baseline correction and area normalization. Figure 1 shows the average Raman spectra of controls ($n = 22$) and rectal cancer patients before operation and after operation ($n = 35$). Peaks at wavenumbers of 495, 590, 636, 727, 820, 886, 1021, 1073, 1132, 1211, 1323, 1355, 1446 and 1582 cm^{-1} can be consistently observed in all three groups.

Peaks at 495, 636, 727, 820, 1073, 1211, 1355 and 1446 cm^{-1} have the biggest difference between groups. In comparison with the controls, the intensities at 495 (S–S stretch of L-arginine), 636 (C–S stretch of tyrosine), 820
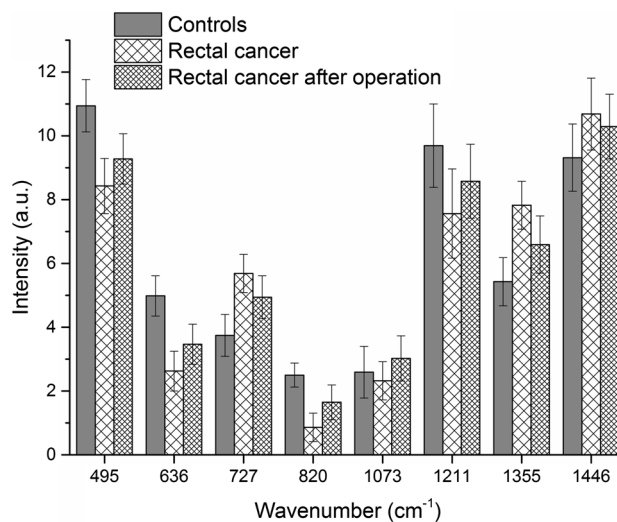
Fig. 2 Mean intensities and standard deviations of peaks with the most distinguishable differences between groups

(ring breathing of tyrosine), 1073 (C–C stretch of phospholipids) and 1211 cm^{-1} (C–C$_6$H$_5$ of phenylalanine and tryptophan) are lower for diseased samples, while bands at 727 (C–H bending of adenine and coenzyme), 1355 (CH$_3$CH$_2$ wagging of tryptophan, adenine and guanine) and 1446 cm^{-1} (CH$_2$ bending of collagen/lipids) are more intense (Fig. 2). All the peaks have changes with statistical significance (with p values calculated from one-way ANOVA <0.001, Table 2).

The decrease in peak intensities at 495, 636, 1073 and 1211 cm^{-1} indicates decreases in corresponding amino acids—arginine, tyrosine, phospholipids, phenylalanine and tryptophan—relative to the total SERS-active

Table 2 Assignments of Raman peaks, intensity with standard deviation, and *p* value

Wavenumber (cm^{-1})	Vibrational mode	Assignments [10, 12, 19, 20]	*p* value
495	S–S stretch	L-arginine	<0.001
636	C–S stretch	Tyrosine	<0.001
727	C–H bending	Adenine, coenzyme	<0.001
820	Ring breathing	Tyrosine	<0.001
1073	C–C stretch	Phospholipids	<0.001
1211	C–C$_6$H$_5$	Phenylalanine, tryptophan	<0.001
1355	CH$_3$CH$_2$ wagging	Tryptophan, adenine, guanine	<0.001
1446	CH$_2$ bending	Collagen, lipids	<0.001

Fig. 3 Scatter plot of the first two PCs for the three groups (*A* controls; *B* rectal cancer patients before operation; *C* rectal cancer patients after operation. Characters in *red* represent misclassified samples)

constituents in rectum cancer serum. This may be caused by the vigorous metabolism of tumor cells. The Raman peak at 820 cm^{-1} is also found to decrease, suggesting a reduction in collagen in rectum cancer group. The collagen decline is due to cytoplasmic mucin depletion and the elevated concentration of metalloproteinase [21]. This result is in agreement with colorectal cancer research [22]. The increase at 727 and 1355 cm^{-1} which are partly caused by adenine and guanine may be due to the abnormal metabolism of DNA or RNA in blood due to apoptosis and necrosis of neoplastic cells [12]. The peak at 1446 cm^{-1} corresponds to the CH$_2$ bending of collagen and lipids, which is higher in cancer serum samples. This same peak was also found to be higher in other cancerous tissues such as stomach cancers and was recognized as being of diagnostic significance in existing research [17].

Tumor markers in blood have some relationship with the occurrence of cancer. However, as the correlations are too complex, few biomarkers have the sensitivity and specificity required for the diagnosis of cancer and act as supplemental evidence instead. Besides, there are many factors that can influence blood components and cause false-positive or false-negative errors of biomarkers. Those factors include but are not limited to other diseases the patients have, special physiological periods (such as pregnancy or menstruation), medical treatment (such as operation or chemotherapy), drugs, the sealing of cancer cells, the combination of biomarkers to other antibodies, and the poor circulation of blood. These factors may cause fluctuation of hormones. For instance, the estradiol level in blood will vitiate with the development of follicular and will increase about 100 times with pregnancy [23], drugs or other secretions in blood. In our experiment, all postoperative samples were taken before patients received chemotherapy to avoid interference. Simultaneous detection of several biomarkers can greatly improve the predication accuracy. Microchip is an ideal tool for multi-biomarker detection, but its expense
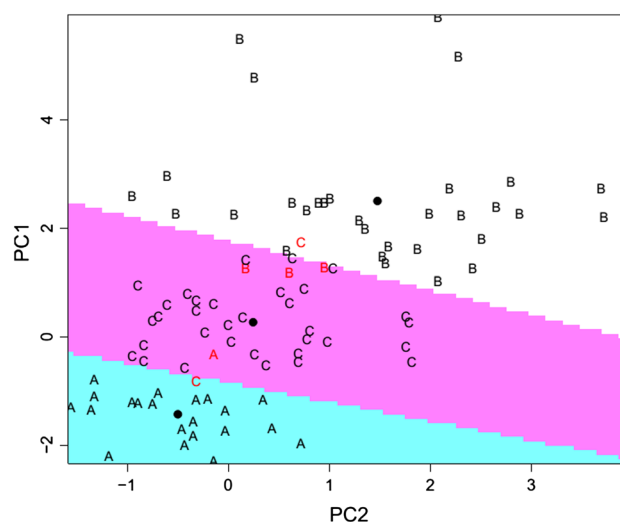
is very high and can hardly be used for large-scale screening. Optical spectroscopy techniques provide us with promising alternatives for multi-biomarker noninvasive diagnostic tests for cancers. PCA–LDA and parameter-CART were employed on the SERS spectra to form algorithms for classification.

3.2 PCA–LDA

Principal component analysis (PCA) was utilized on the original spectra to extract the key information between groups. The principal component model was used on the spectra. The first two principal components (PCs) contributed to more than 97 % of the total variation, so only PC1 and PC2 were maintained for further discriminant analysis. Linear discriminant analysis (LDA) was performed on the first two PCs for the discrimination of the three groups. Fisher's discriminant lines were retrieved from LDA and were drawn on the PC scatter plot. Figure 3 demonstrates the results of the predictions for each spectrum based on the first two principal components of PCA. In this analysis, most spots belong to different groups located at the opposite side of the line and scattered separately. These results indicate that samples can be discriminated well between groups.

3.3 Parameter-CART

The mechanism of CART allows us to visualize the parameters and their ranges for the partition of samples and to see which parameters are important in the discrimination

(a)

(b)
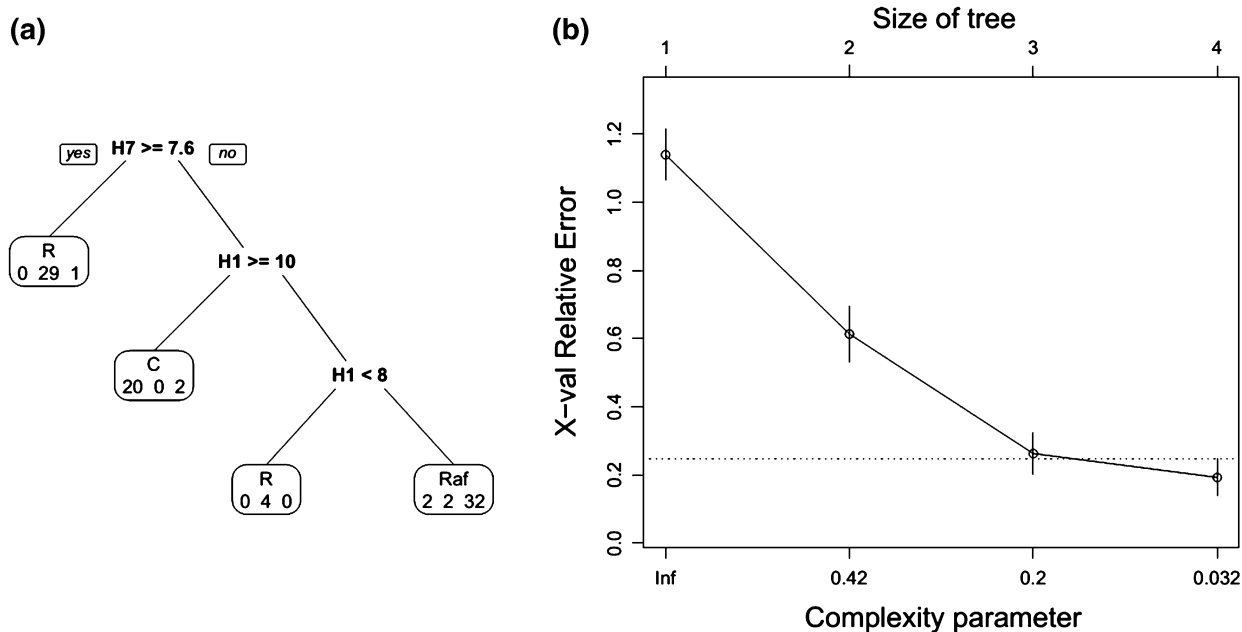
Fig. 4 Pruned CART (**a**) and relative error for different tree sizes (**b**) (*C* controls; *R* rectal cancer patients before operation; *Raf* rectal cancer patients after operation)

of spectral data. Thus, we selected peak intensities of the eight main peaks (Table 2) as parameters. CART was subsequently used on the parameters for the classification between groups. It can be seen that by judging the values of the parameters, samples were separated after several steps. The leaf nodes are the separated parts. Node splits were based upon the regression scores of each parameter. In each level (node), samples were split by a binary decision on whether the inequality was satisfied. In this experiment, the PC values were judged. The variance error reached a minimum when the size of tree reached 4, so there are 4 levels in our classification tree. CART analysis revealed that peak intensities can be used to discriminate groups efficiently, and the classification results are shown under each leaf node.

Figure 4 shows the relationship of number of nodes with respect to the cost (misclassification error) after tenfold cross-validation. The misclassification cost for both cross-validated and resubstitution errors decreased monotonically as the number of terminal nodes increased. The misclassification cost reached an optimal value at the point of 4 nodes. Figure 4 demonstrates the optimized CART. From the figure, we can see that after three steps, groups were separated. And each of the three parameters was used once, which means all three parameters were useful in discriminating different groups. From the CART, we can see that only the intensity of peak 1 and peak 8 was used as the criterion for decision. Through this analysis, we can see which parameters of spectroscopy have higher weights

for classification. This may give us another method to decide which peaks are more important for differentiating samples.

For the evaluation of the two algorithms, tenfold cross-validation was used on both PCA–LDA and CART. Table 3 summarizes the classification results of the two methods. The prediction accuracy, sensitivity and specificity vary from 90.9 to 95.5 %. CART performed better than PCA–LDA in our paper (with higher accuracy and sensitivity), indicating the effectiveness of our three parameters.

$$Accuracy = \frac{number\ of\ correctly\ identified\ cases}{total\ cases} \times 100\ \%$$

Sensitivity

$$= \frac{number\ of\ correctly\ identified\ cases\ as\ rectal\ cancer}{total\ cases\ of\ rectal\ cancer} \times 100\ \%$$

Specificity

$$= \frac{number\ of\ correctly\ identified\ cases\ as\ healthy}{total\ cases\ of\ healthy} \times 100\ \%$$

4 Conclusion

In this study, serum taken from rectal cancer patients before and after operation was tested using surface-enhanced Raman spectroscopy and results have been compared with those of controls. Raman peaks assigned to nucleic acids, proteins and lipids were found to have significant

Table 3 Diagnostic results for PCA–LDA and CART using tenfold cross-validation

Method	Group	Prediction			Accuracy (%)	Sensitivity (%)	Specificity (%)
		Controls	Cancer before operation	Cancer after operation			
PCA–LDA	Controls	21	0	1	93.5	Null	95.5
	Cancer before operation	0	32	3		91.4	
	Cancer after operation	1	1	33		94.3	
Parameter-CART	Controls	20	0	2	92.4	Null	90.9
	Cancer before operation	0	33	2		94.3	
	Cancer after operation	2	1	32		91.4	

differences between the three groups. Changes in the chemical components in serum are major factors in spectral changes, so the results indicate that corresponding component changes have occurred with the development of colon cancer. The multivariate statistical techniques of PCA–LDA and parameter-CART were utilized to develop diagnostic algorithms for the classification. Diagnostic accuracies of 93.5 and 92.4 % and specificities of 95.5 and 90.9 % were achieved for PCA–LDA and parameter-CART, respectively. This paper is an exploratory study for using SERS spectra of serum for the diagnosis of rectal cancer, and the techniques discussed will hopefully be perfected in future research for clinical applications.

References

1. R. Siegel, C. DeSantis, A. Jemal, CA Cancer J. Clin. **64**, 104 (2014)
2. N. Howlader, A.M. Noone, M. Krapcho, G. Johannesdottir, N. Neyman, *SEER Cancer Statistics Review, 1975–2010*, 6th edn. (National Cancer Institute, Bethesda, 2013)
3. J.D. Hardcastle, J.O. Chamberlain, M.H. Robinson, S.M. Moss, S.S. Amar, T.W. Balfour, P.D. James, C.M. Mangham, Lancet **348**, 1472 (1996)
4. J. Weitz, M. Koch, J. Debus, T. Höhler, P.R. Galle, M.W. Büchler, Lancet **365**, 153 (2005)
5. L.M. Moreira, L. Silveira Jr, F.V. Santos, J.P. Lyon, R. Rocha, R.A. Zângaro, A.B. Villaverde, M.T.T. Pacheco, Spectroscopy **22**, 1 (2008)
6. A. Synytsya, M. Judexova, D. Hoskovec, M. Miskovicova, L. Petruzelka, J. Raman Spectrosc. **45**, 903 (2014)
7. P.O. Andrade, R.A. Bitar, K. Yassoyama, H. Martinho, A.M.E. Santo, P.M. Bruno, A.A. Martin, Anal. Bioanal. Chem. **387**, 1643 (2007)
8. V. Ranc, The Analyst **138**, 5983 (2013)
9. A.T. Harris, A. Lungari, C.J. Needham, S.L. Smith, M.A. Lones, S.E. Fisher, X.B. Yang, N. Cooper, J. Kirkham, D.A. Smith, D.P. Martin-Hirsch, A.S. High, Head Neck Oncol. **1**, 1 (2009)
10. S. Feng, D. Lin, J. Lin, B. Li, Z. Huang, G. Chen, W. Zhang, L. Wang, J. Pan, R. Chen, H. Zeng, Analyst **138**, 3967 (2013)
11. P.L. Stiles, J.A. Dieringer, N.C. Shah, R.P. Van Duyne, Annu. Rev. Anal. Chem. **1**, 601 (2008)
12. D. Lin, S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen, S. Xie, H. Zeng, R. Chen, Opt. Express **19**, 13565 (2011)
13. G. Wang, R.J. Lipert, M. Jain, S. Kaur, S. Chakraboty, M.P. Torres, S.K. Batra, R.E. Brand, M.D. Porter, Anal. Chem. **83**, 2554 (2011)
14. Y. Chen, G. Chen, S. Feng, J. Pan, X. Zheng, Y. Su, Y. Chen, Z. Huang, X. Lin, F. Lan, R. Chen, H. Zeng, J. Biomed. Opt. **17**, 0670031 (2012)
15. S.K. Teh, W. Zheng, K.Y. Ho, M. Teh, K.G. Yeoh, Z. Huang, J. Biomed. Opt. **13**, 034013 (2008)
16. F.S. Aguiar, L.L. Almeida, A. Ruffino-Netto, A.L. Kritski, F.C. Mello, G.L. Werneck, BMC Pulm. Med. **12**, 1 (2012)
17. S. Feng, J. Pan, Y. Wu, D. Lin, Y. Chen, G. Xi, J. Lin, R. Chen, Sci. China Life Sci. **54**, 828 (2011)
18. K. Liu, S. Wu, M. Chen, S. Pan, Spectrosc. Spectr. Anal. **28**, 339 (2008)
19. S.-X. Li, Q.-Y. Zeng, L.-F. Li, Y.-J. Zhang, M.-M. Wan, Z.-M. Liu, H.-L. Xiong, Z.-Y. Guo, S.-H. Liu, J. Biomed. Opt. **18**, 027008 (2013)
20. S. Li, Y. Zhang, J. Xu, L. Li, Q. Zeng, L. Lin, Z. Guo, Z. Liu, H. Xiong, S. Liu, Appl. Phys. Lett. **105**, 091104 (2014)
21. D. Lin, J. Pan, H. Huang, G. Chen, S. Qiu, H. Shi, W. Chen, Y. Yu, S. Feng, R. Chen, Sci. Rep. **4**, 4751 (2014)
22. S. Li, G. Chen, Y. Zhang, Z. Guo, Z. Liu, J. Xu, X. Li, L. Lin, Opt. Express **22**, 25895 (2014)
23. S. Silbernagl, A. Despopoulos, *Color Atlas of Physiology*, 6th edn. (Thieme, Stuttgart, 2008)