COMBINATION OF LABEL-FREE SURFACE-ENHACED RAMAN SPECTROSCOPY WITH CONVOLUTIONAL NEURAL NETWORK FOR DNA RECOGNITION

[1]Anastasia SKVORTSOVA, [1]Andrei TRELIN, [1]Vaclav SVORCIK, [1]Olga GUSELNIKOVA, [1]Oleksyi LYUTAKOV

[1]*UCT - University of Chemistry and Technology, Prague, Czech Republic, EU,*
skvortss@vscht.cz

Abstract

Nowadays, the rapid identification of bacterial antibiotic resistance is one of the major biomedical challenges. Classical methods of detection (culture and sensitivity testing, microbial whole-genome sequencing) fail in the context of time requirements. In this work, we propose the express method for the detection of gene encoding enzyme responsible for bacterial antibiotic resistance. Proposed analytical approach is based on a combination of unique advantages provided by surface enhanced Raman spectroscopy (SERS) and artificially created convolutional neural network (CNN). SERS is known for the extremely high sensitivity and fast analysis, while CNN seems to be a promising alternative to find even ambiguous spectral properties produced by the Raman signal.

Keywords: SERS, DNA, neural networks, CNN, antibiotic resistance

1.   INTRODUCTION

The culture-based approach for diagnosis of antibiotic-resistive bacteria development is currently used. However, culturing of the bacteria takes more than 24 hours and together with further species identification and evaluation can prolong time of analysis up to several days [1], which is crucial in need of urgent treatment. On the other hand, such methods as multiplex PCR, [2] electrochemical detection [3], and mass spectrometry [4] can achieve high accuracy for recognition of specific DNA, responsible for antibiotic resistence. Nevertheless, these techniques require expensive instruments, sophisticated sample preparation, and trained staff for obtained data analysis [5].

Recently, Surface-enhanced Raman scattering (SERS) analysis became a promising analytical tool in the field of sensing of DNA oligomers, with especial focus on the presence of antibiotic-resistant genes [6;7;8]. The main advantages of this method are related to the possibility to provide rich spectral information obtained through direct analysis of biomedical samples [9]. However, poor reproducibility and possible interferences of SERS bands due to the complex structure of biological samples restrict the direct SERS application. As an overcoming solution the machine-learning spectral data analysis was recently proposed [10,11]. It is well known, that typical post-processing statistical tools, such as PCA, linear discriminant analysis or 2D correlation fail in the need to find non-linear spectral correlations [12], when machine learning algorithms are characterized by a capability to found complex relationships in data analysis, making the overall analytical procedure highly reliable.

In our previous work [10,11], combination of SERS and artificial neural networks for relevant DNA oligomers recognition and its promising applications were demonstrated for the first time, and now we have focused on the optimization of the biosensor to increase the reproducibility, accuracy and usefulness of the measurements.

2. EXPERIMENTAL

2.1. Reagents

Acetic acid (reagent grade, ≥99 %), diethyl ether, deionized water, 4-aminobenzoic acid (≥99.0 %), p-Toluenesulfonic acid monohydrate (ACS reagent, ≥98.5 %), high-purity water (EMD MILLIPORE) were purchased from Sigma-Aldrich and used without further purification. The oligonucleotides (ONDs) were obtained from Thermofisher. The OND sequences used in this study are presented in Table S1. DVD+R disks (Verbatim) were purchased from local shop (Alza).

2.2. Preparation of samples

For preparation of plasmon-active substrate the polymer grating was used. The Au thin film was deposited onto a patterned polymer surface by vacuum sputtering (discharge power of 7.5 W, sputtering time 350 s, resulted in thickness 25 nm).

The obtained gratings were spontaneously modified by soaking in 1 mM freshly prepared aqueous solution of 4-carboxybenzenediazonium tosylate (ADT-COOH) and rinsed with water and ethanol twice.

0.5 nmol of Au-C (NH2-terminated) was grafted to surface through EDC/sulfo-NHS activation. For DNA hybridization experiments the 10 mM Tris buffer was used. All samples were kept in a refrigerator at 5 ˚C

2.3. Characterization techniques

For characterization of the sample surface and nanomechanical properties the peak force AFM technique was used. AFM scratch tests were carried out on control sample on glass substrate by profiling across a scratch at an angle of 90° relative to the surface. SERS spectra were measured On ProRaman-L spectrometer with 785 nm excitation wavelengths. The spectra were recorded at a resolution of 2 cm^{-1} in the 3000–200 cm^{-1} wavenumber range. UV-Vis spectra were measured using Spectrometer Lambda 25 (Perkin-Elmer) in 300–1100 nm wavelength range.

2.4. CNN data processing

All SERS spectra were evaluated using home-made CNN design with implementation of TensorFlow 2 framework [13]. The collected data set consisted of individual spectra marked with corresponding categories. Spectra preprocessing involved their normalization and noise removing. The training was continued until manually observed convergence, which took 50 epochs in total, training and validation losses were converged to 0.1299 and 0.1291 respectively.

3. RESULT AND DISCUSSION

Recent advances in achieving low limits of detection and possibility to get massive spectral data in a short period of time make surface-enhanced Raman scattering a useful tool for the detection of DNA oligomers [10]. However, spectral differences are easily masked by background noise in complex biomolecules; thereby peak interference reduces universality of the method. On the other hand, artificially created convolutional neural network (CNN) is characterized by the ability to split the interfered peaks and reveal the most important regions of the spectrum [14] for the more accessible data interpretation.

To get rid of the long sample preparation, polymer grating as a cost-effective and commercially available alternative for the biosensor platform was used. After sputtering of the plasmonic metal (Au), periodical structure of the grating provided the SERS effectivity. The thickness of gold was optimized to get the highest enhancement factor (**Figure 1**). Optimization have demonstrated that the value of EF reach 2.3*10^6 in the case of 25 nm thick gold layer (350 s of sputtering).
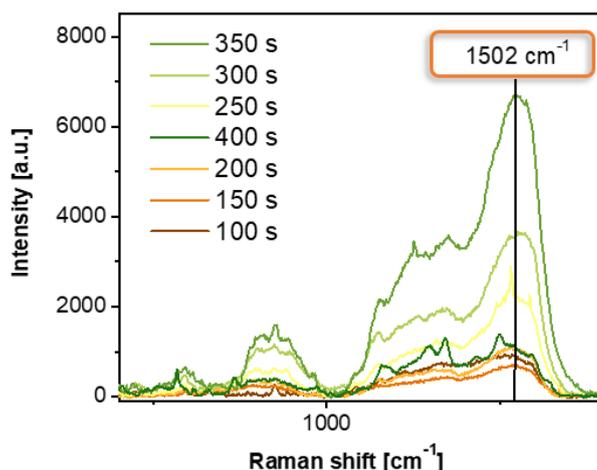
Figure 1 Calculation of EF of gold grating: SERS spectra of R6G on Au grating with various thicknesses

Activation of the low-reactive surface of the substrate was achieved by diazonium modification for 4-carboxyphenyl groups grafting, which are then able to covalently bond with the amino-terminated OND (C-OND) (**Figure 2**).
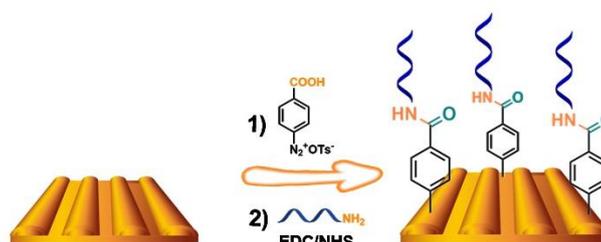
Figure 2 Scheme of preparation of gold grating with grafted capture oligonucleotide (Au-C)
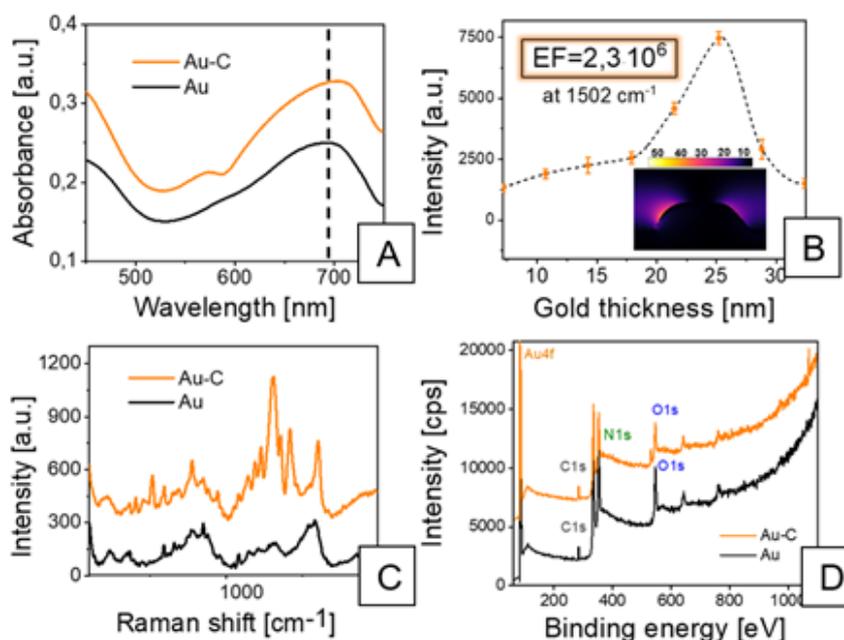
Figure 3 Characterization of Au and Au—C: (a) UV-Vis spectroscopy, (b) Calculation of enhancement factor of Au using R6G (insert is simulation of EM field distribution); (c) SERS spectra; (c) survey XPS spectra

SERS spectra of Au-C confirm the successfulness of C-OND grafting via the appearance of new characteristic bands typical for the oligonucleotide structure. (**Figure 3C**). The attachment of 4-carboxyphenyl groups leads to the appearance of carboxyl group related peaks (1596 cm^{-1}) and aromatic ring (786, 852, 1086, 1214 cm^{-1}). After the coupling with amino-terminated OND-C, there are new peaks in the region 780-1600 cm^{-1}, which are typical for the oligonucleotide structure (bands at 778 cm^{-1}, 1310 cm^{-1}, 1359 cm^{-1}, and 1422 cm^{-1}).

Also, the simulation of plasmon intensity across the grating was performed by finite-difference-time-domain method (FDTD). Numerical simulation of the electric field (EF) distribution on gold grating was performed for excitation with 785 nm light (**Figure 3 A**). Cross sections of one period of the structure yielded the distribution of the localized EF, suggesting that the energy is indeed concentrated in the slopes of the grating (**Figure 3B**).

According to SEM and AFM images, morphology of the resulting substrate consisted of parallel wave-like patterns, with a height of about 168±3 nm and a period of about 710±7 nm (as measured from a cross-section profile, **Figure 4**). After the diazonium modification and covalent grafting of amino-terminated C-OND the pristine grating structure was conserved.
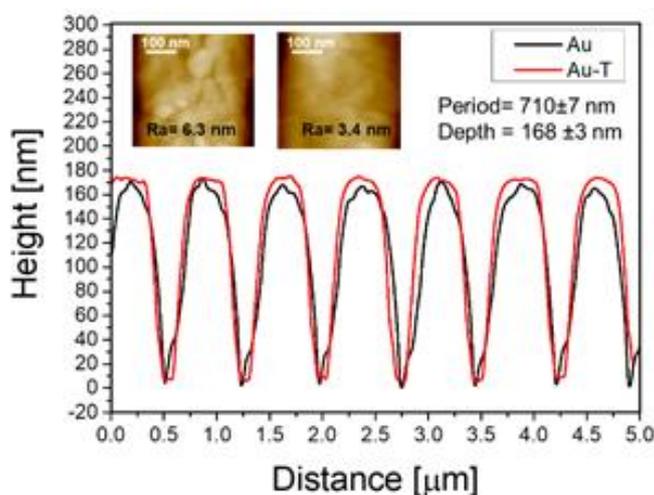
Figure 4 profiles of Au and Au-C, inserts are magnified scans (0.4×0.4 μm2)

To get spectral information about the presence of NDM-encoding gene (T-OND), samples already grafted with the first OND (Au-C) were immersed in a buffer solution to capture the second one. Massive SERS spectra collected prepared samples from were used as a NN input for the deep learning to identify the presence of NDM-encoding gene.

The validation data were utilized for convolutional NN training to automated multiclass classification of the corresponding ONDs using a trained neural network for differentiation between various DNA. The deep convolutional neural network implemented within Keras was used for classification [15]. The network was trained on 75% of randomly selected data and verified on 25% of data. After verifying the convergence, the accuracy reached 99.92% with a level of confidence exceeding 95%.

Also it should be noted that the total time from the beginning of the measurement to obtaining the results is 1-1.5 hours with a preset program of the neural network that is a huge advantage compared to gold-standard cultural-based methods taking more than 24 hours.

4.   CONCLUSION

Here we demonstrated the proof-of-concept performance of sensor by the analysis of ONDs related to specific genes responsible for antibiotic resistance. We utilized grafted SERS substrates for DNA capturing. Prepared sensor was involved in the hybridization with T-OND and other mutated ONDs. Further collected SERS spectra

measured by Raman spectrometer were utilized as a learning dataset for CNN. Several types of ONDs were detected with accuracy higher than 99,9% with a level of confidence exceeding 95%.

As developed sensor is operated by the ONDs in the similar way as PCR methods does, it is applicable not only for bacterial resistivity but also for tumors, diseases, food and toxicants. This entitles our method to be a universal easy-to-performed platform for the detection of DNA oligomers.

ACKNOWLEDGEMENTS

REFERENCES

[1]   LEE, A., MIRRET, S., RELLER, L.B., WEINSTEIN, M.P. Detection of bloodstream infections in adults: how many blood cultures are needed? *J Clin Microbiol*. 2007, vol. 45, pp. 3546-3548.

[2]   STROMMENGER, B., KETTLITZ, C., WERNER, G., WITTE, W. Staphylococcus aureus antibiotic resistance genes in detection of nine clinically relevant multiplex PCR assay for simultaneous. *J Clin Microbiol.* 2003, vol. 41, pp. 4089-4094.

[3]   TSUI-MING-KUO, MO-YUAN SHEN, SHIH-YING HUANG. Facile Fabrication of a Sensor with a Bifunctional Interface for Logic Analysis of the New Delhi Metallo-β-Lactamase (NDM)-Coding Gene. *ACS Sens.* 2016, vol. 1, no. 2, pp. 124–130.

[4]   CHARRETIER, Y., SCHRENZEL, J. Mass spectrometry methods for predicting antibiotic resistence. *Proteomics Clinical Applications.* 2016, vol. 10, pp. 964-981.

[5]   DANIEL, D., GALVAN, Y. Surface-Enhanced Raman Scattering for Rapid Detection and Characterization of Antibiotic-Resistant Bacteria. *Advanced Helthcare materials.* 2018, vol. 7.

[6]   WOOD, R., JENSEN, T., WADSWORTH, C., CLEMENT, M., NAGPAL, P., PITT, P. Analysis of Identification Method for Bacterial Species and Antibiotic Resistance Genes Using Optical Data From DNA Oligomers. *Front. Microbiol.* 2020.

[7]   GALVAN, D., YU, Q. Surface-Enhanced Raman Scattering for Rapid Detection and Characterization of Antibiotic-Resistant Bacteria. *Advanced healthcare materials.* 2018, vol. 7, no. 13.

[8]   GUSELNIKOVA, O., POSTNIKOV, P., KOLSKA, Z., ZARUBA, K., KOHOUT, K., ELASHNIKOV, R., SVORCIK, V., LYUTAKOV, O. Homochiral metal-organic frameworks functionalized SERS substrate for atto-molar enantio-selective detection. *Applied Materials Today.* 2020, vol. 20.

[9]   NOVELLI-ROUSSEAU, A., ESPAGNON, I., FILIPUTTI, D. Culture-free antibiotic-susceptibility determination from single-bacterium Raman spectra. *Scientific reports*. 2018.

[10]   GUSELNIKOVA, O., TRELIN, A., SKVORTSOVA, A., ULBRICH, P. Label-free surface-enhanced Raman spectroscopy with artificial neural network technique for recognition photoinduced DNA damage. *Biosensors and Bioelectronics.* 2019, vol.145.

[11]   ERZINA, M., TRELIN, A., GUSELNIKOVA, O. Precise cancer detection via the combination of functionalized SERS surfaces and convolutional neural network with independent inputs. *Sensors and Actuators B: Chemical.* 2020, vol. 308.

[12]   MADRY, A., MAKELOV, A., SCHMIDT, L., TSIPRAS, D., VLADU, A. Towards deep learning models resistant to adversarial attacks. 2017.

[13]   CHOLLET F. Keras. 2015.

[14]   VIRTANEN P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods.* 2020, vol. 17, pp. 261–272.

[15]   HO, C., JEAN, N., HOGAN C., BLACKMON, L. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nature.* 2019, vol. 10, no. 4927.