NANO LETTERS

pubs.acs.org/NanoLett
Letter

Single-Molecule SERS Discrimination of Proline from Hydroxyproline Assisted by a Deep Learning Model

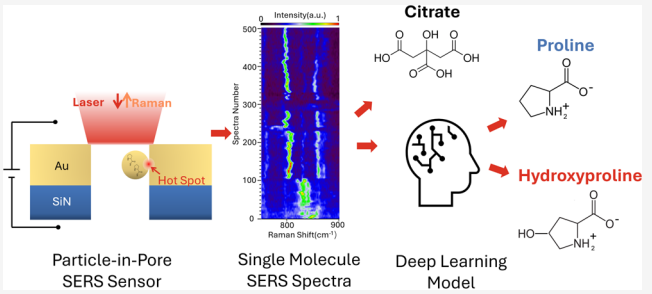
Yingqi Zhao, Kuo Zhan, Pei-Lin Xin, Zuyan Chen, Shuai Li, Francesco De Angelis, and Jian-An Huang*

Cite This: *Nano Lett.* 2025, 25, 7499−7506

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Discriminating low-abundance hydroxylation is a crucial and unmet need for early disease diagnostics and therapeutic development due to the small hydroxyl group with 17.01 Da. While single-molecule surface-enhanced Raman spectroscopy (SERS) sensors can detect hydroxylation, subsequent data analysis suffers from signal fluctuations and strong interference from citrates. Here, we used our plasmonic particle-in-pore sensor, occurrence frequency histogram of the single-molecule SERS spectra, and a one-dimensional convolutional neural network (1D-CNN) model to achieve single-molecule discrimination of hydroxylation. The histogram extracted spectral features of the whole data set to overcome the signal fluctuations and helped the citrate-replaced particle-in-pore sensor to generate clean signals of the hydroxylation for model training. As a result, the discrimination of single-molecule SERS signals of proline and hydroxyproline was successful by the 1D-CNN model with 96.6% accuracy for the first time. The histogram further validated that the features extracted by the 1D-CNN model corresponded to hydroxylation-induced spectral changes.



KEYWORDS: *Post-translational modification, SERS, CNN, Plasmonic nanopore, Hydroxyproline and proline, Single molecule*

Post-translational modification (PTM) plays a fundamental role in regulating cellular processes by significantly affecting the protein structure and dynamics. In particular, prolyl 4-hydroxylation catalyzes the addition of a hydroxyl group to proline residues in collagens, which regulates hypoxia-inducible transcription factor (HIF) and activates hundreds of genes in an oxygen-dependent manner to drive cancers.[1,2] This important discovery was awarded the Nobel Prize in Physiology or Medicine in 2019. As a critical regulator of cancer development, HIF hydroxylation suppresses HIF-1α and constrains tumor growth. However, if impaired, for example, by the loss of the von Hippel-Lindau protein, which is common in clear cell renal carcinoma, HIF accumulates and unleashes its oncogenic potential.[3,4] HIF hydroxylation can serve as a potential biomarker of tumor behavior assessment and therapy response,[5] as well as therapeutic targeting.[6] Yet, the current detection methods based on mass spectrometry and antibodies are facing challenges in detecting low-abundance and site-specific PTM.[7,8] Therefore, single-molecule analysis of trace amounts of PTM biomarkers in human biofluids is of great significance for early-stage disease diagnosis, low-abundance protein studies, and therapeutic development.[9]

Mass spectrometry is currently the dominant technology for hydroxylation analysis. However, its sensitivity is constrained by spontaneous nonspecific oxidations and sample loss, which usually requires 10^6−10^8 molecular copies, and therefore it is challenging to detect low-abundance hydroxylation. Nanopore resistive pulse sensing has emerged as a promising single-molecule detection method, which distinguishes biomolecules based on current changes as they pass through a nanopore. Using this technique, single-molecule PTM detection has been demonstrated for phosphorylation,[10] acetylation,[11] propionylation,[12] glycosylation,[13] nitration, and oxidation.[14] However, the difference in molecular structure (−OH group with 17.01 Da) between hydroxylated proline and proline is too small to induce sufficient current changes in nanopores for reliable detection, making resistive pulse analysis of single-molecule hydroxylation particularly challenging.

Surface-enhanced Raman spectroscopy (SERS) is another promising tool for label-free PTM detection due to its high sensitivity and ability to provide molecular structural information.[15] Similar to the nanopore resistive pulse sensors, significant research efforts have focused on using plasmonic colloids to detect PTMs with large Raman cross sections and distinct SERS spectral features, for example, phosphorylation,[16,17] nitration,[18] and oxidation.[19] Yet, few SERS-based detections have been reported for small PTMs with small

Received: February 21, 2025
Revised: April 9, 2025
Accepted: April 11, 2025
Published: April 17, 2025

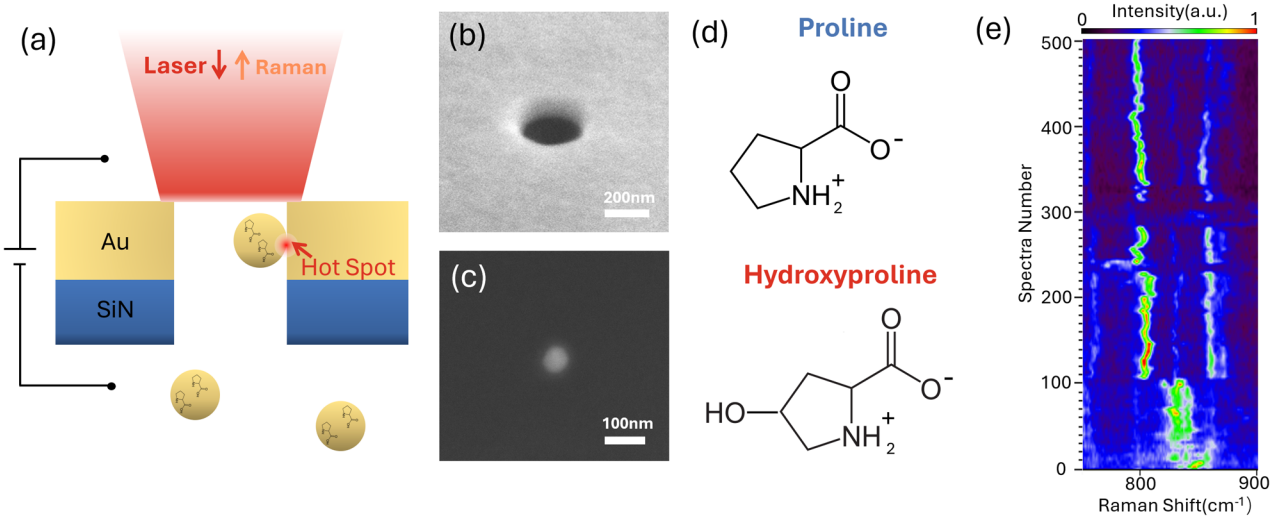


Figure 1. (a) Schematic illustration of the particle-in-pore sensor that allows single AuNPs loaded with analytes to be trapped in a gold nanopore to generate a plasmonic hot spot upon laser illumination. (b) SEM image of the nanopore. (c) SEM image of the AuNP. (d) The molecular structures of Pro and Hyp at pH 5.5 in solution. (e) Typical fluctuating single-molecule SERS spectra of Pro. The color bar is the normalized signal intensity in arbitrary units.

Raman cross sections, including hydroxylation, due to the difficulties in obtaining sufficient SERS spectra for analysis or the limitation of SERS system sensitivity.[20,21] Another problem caused by the small PTM Raman cross section is that the peptide backbone signal can screen or merge the PTM spectra.[1] As a result, there has been no report on the SERS detection of hydroxylation at the single-molecule level.

In contrast to the above SERS methods, our recent plasmonic particle-in-pore sensor demonstrated SERS spectroscopic detection of the 20 amino acids and peptides at the single-molecule level.[22,23] The sensor generated a single gap-mode plasmonic hot spot with single-molecule sensitivity between a gold nanoparticle and the nanopore side wall, as shown in Figure 1(a). When a molecule enters the hot spot, part of it is located in the small hot spot and excited,[24] thereby generating SERS signals from the corresponding molecule moiety and avoiding the backbone signal coverage. However, challenges in the single-molecule SERS detection of the hydroxylation lie in the spectra interpretation due to the fluctuation and blinking of the single-molecule spectra. What is worse, the citrate surfactants on the nanoparticle surface or solution may interfere with the SERS data analysis of single-molecule hydroxylation.

In this study, we report the single-molecule SERS discrimination of proline (Pro) and hydroxyproline (Hyp) using the particle-in-pore sensor, the occurrence frequency histogram of SERS peaks, and a deep learning model of a one-dimensional convolutional neural network (1D-CNN). To analyze the highly dynamic single-molecule data, we employed occurrence frequency histograms to filter out intensity fluctuations caused by nanoparticle morphology variation while preserving structural information for molecular discrimination. Using these histograms, we investigated the substitution of citrates on gold nanoparticles (AuNPs) with analytes to minimize citrate interference in spectral interpretation and benefit the training of the 1D-CNN model. As a result, the 1D-CNN model achieved more than 96% accuracy in both testing and post evaluation. Finally, we used 1D gradient-weighted positive feature visualization to validate discrimination based on Raman bands corresponding to hydroxylation-induced spectral changes in the pyrrolidine ring. To the best of our knowledge, this is the first successful discrimination of an amino acid with and without hydroxylation, one of the most challenging PTMs for SERS, at the single-molecule level, laying the solid foundation for the detection of low-abundance HIF hydroxylation.

SENSOR FABRICATION AND MEASUREMENT

Figure 1(a) presents a schematic illustration of the particle-in-pore sensor, which consists of a gold nanopore array with a 200 nm diameter that was drilled on a 100 nm gold film supported on a free-standing SiN membrane with a silicon frame by a focused ion beam. The detailed fabrication process is recorded in Supplementary Note 1. Figure 1(b) shows an SEM image of a typical nanopore. The SiN membrane chip is encapsulated into a microfluidic device with top and bottom chambers, across which an electric potential is applied. The analyte, Pro or Hyp, is physically adsorbed onto a 50 nm diameter AuNP (Figure 1(c)) before being injected into the bottom reservoir. The negatively charged AuNPs are then driven through the nanopore by the cross-membrane electric potential. A 785 nm laser illuminates the nanopore from above, generating an optical force that pushes the nanoparticle toward the nanopore sidewall to generate the gap-mode hot spot formed between the AuNP and the nanopore sidewall with single-molecule sensitivity.[22] During the particle trapping period, the analyte could diffuse into the hot spot, where it is excited to produce single-molecule SERS spectra.

A monolayer of Pro or Hyp was attached to AuNPs by incubation in a 5% PBS buffer containing Pro and Hyp at pH 5.5. The concentration of molecules required to achieve monolayers capping the AuNP surfaces was determined by empirical values of maximum solvent accessibilities of amino acids[25] and nanoparticle surface area; the details are discussed in the Supplementary Note 2. At pH 5.5, both molecules exist in their zwitterionic form, as shown in Figure 1(d). The purchased AuNPs were initially stabilized by a citrate layer, which acted as a surfactant. Upon incubation in a Pro or Hyp solution, the citrate was replaced by amino acids due to their higher binding affinity to AuNPs,[26,27] as discussed in

https://doi.org/10.1021/acs.nanolett.5c01177
Nano Lett. 2025, 25, 7499−7506

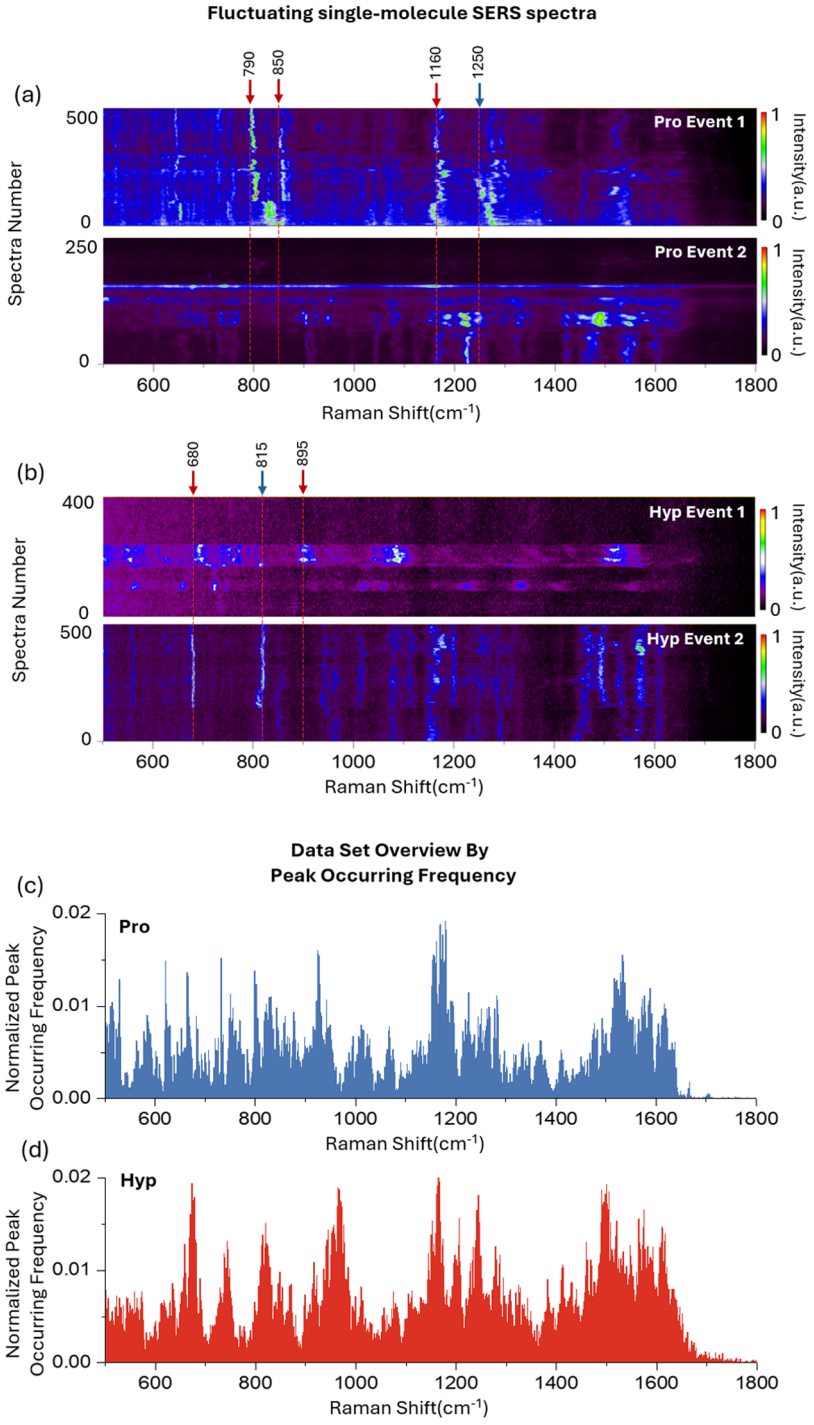


Figure 2. Waterfall plot of typical single-molecule SERS spectra of (a) Pro and (b) Hyp trapping events shows the vibrating feature of single-molecule SERS spectra. The color bars indicate the normalized signal intensity in arbitrary units. The red arrow indicates that the band shows in only one trapping event, while the blue arrow indicates that the band shows in both trapping events. Distribution histograms of normalized peak occurrence frequency of (c) Pro and (d) Hyp on AuNP with 48 h monolayer incubation. The number of spectral events contained in the histograms for Pro and Hyp is 11,002 and 9769, respectively.

Supplementary Note 3. Typical single-molecule SERS spectra of Pro show fluctuating SERS peaks, as plotted in Figure 1(e).

MOLECULAR PROFILING FROM SINGLE-MOLECULE SERS SPECTRA

The explanation of single-molecule spectra is challenging due to fluctuations and blinking features.[28] Figure 2(a) and (b) present typical SERS waterfall plots generated by the sensor, showing two distinct trapping events for each molecule. In a single trapping event, only a subset of vibrational modes is typically excited. For instance, in Pro event 1, three bands near 790 cm^{-1} (δ_{COO^-}),[21] 850 cm^{-1} (ρ_{CH_2}),[29] and 1160 cm^{-1} (t_{CH2})[29] are observed but absent in event 2, while the band near 1250 cm^{-1} (ω_{CH2})[29] appears in both. Similarly, for Hyp, the band at 815 cm^{-1} (t_{CH2})[21] appears in event 1, but the 680 cm^{-1} (δ,ω_{COO^-})[21] is present in event 2, while a band at 895

https://doi.org/10.1021/acs.nanolett.5c01177
Nano Lett. 2025, 25, 7499−7506

cm^{-1} (cit)[30] is shown in a few spectra in event 1, but exhibits long trapping in event 2. This variation arises because when a AuNP is trapped near the nanopore sidewall, the resulting hot spot is only a few angstroms in size—comparable to the molecular dimensions. As a result, only specific portions of the molecule reside within the hot spot, leading to selective excitation of vibrational bands based on molecular position and orientation.[31] Given that molecules diffuse on the nanoparticle surface and the nanoparticle still has partial Brownian motion while trapped in the nanopore, the particle-in-pore sensor captures different molecular moieties in each trapping event or spectrum. By collecting large data sets, the system effectively reconstructs a more complete molecular profile.

Single-molecule SERS bands exhibit fluctuations in both the position and intensity. The peak intensity varies due to factors such as the distance between the nanoparticle and the nanopore sidewall, hot spot size, molecular position, and orientation, making it less reliable for distinguishing molecular structural differences. However, despite peak position fluctuations, the central position can be statistically extracted from a sufficiently large data set, providing insights into structural changes. By analyzing peak positions and their frequency of occurrence across multiple trapping events, the general spectral characteristics can be determined.[32]

The peak frequency histogram is an efficient tool to obtain a general understanding of the large amount of single-molecule SERS data.[32] We quantified the peak occurring frequency at each Raman shift and plotted normalized histograms for Pro and Hyp, as shown in Figure 2(c) and (d). Raw data were preprocessed through cosmic-ray-removal, normalization, and baseline correction. The number of peak occurrences was then divided by the total number of effective spectra (details provided in the Supplementary Note 1). The overall spectral profiles of Pro and Hyp are similar due to their structural resemblance, although they differ in specific peak positions. By displaying peak occurrence frequency and position (Raman shift), the histograms highlight single-molecule SERS bands at high spectral resolution that would otherwise overlap in multimolecule SERS spectra or powder Raman spectra. For instance, in the Pro histogram in Figure 2(c), four CH_2 rocking bands appear near 825 cm^{-1}, 844 cm^{-1}, 855 cm^{-1}, and 878 cm^{-1},[29] whereas these bands may merge into two or three in powder Raman spectra (as shown in Figure S4 in the Supporting Information) or multimolecule SERS spectra.[20,21] Although not all vibrational modes are excited in a single trapping event, the particle-in-pore sensor captures detailed spectral information that might be lost in conventional multimolecule SERS spectra. Due to the lack of reference for single-molecule proline/hydroxyproline SERS spectra, precise peak assignment in a conventional way is not straightforward, besides the interference of citrate on AuNP, making it even difficult to discriminate analytes only based on the histogram. However, the histogram of peak occurrence frequency was also proven to be a useful tool for analyzing citrate substitution and determining the valid data set for 1D-CNN model training, which will be discussed below.

CITRATE SUBSTITUTION

The presence of surfactants on AuNPs complicates precise peak assignment in single-molecule SERS data obtained from the particle-in-pore system. Due to the system's extreme sensitivity, even residual citrate—used as a stabilizing surfactant for AuNPs—can no longer be ignored. To investigate citrate substitution, we analyzed AuNPs incubated in the analyte solution (Pro and Hyp in this work) for 0 and 48 h and then measured the peak occurrence frequency. Based on our previous protocol, a 48 h incubation time was chosen, as it provides a balance between sufficient substitution, colloidal stability, and efficiency.[23] The analyte concentration was adjusted to ensure monolayer coverage of the AuNPs. Additionally, citrate detection by the particle-in-pore sensor using AuNPs with 1/8 monolayer Hyp coverage and a 24 h incubation time was conducted as an example of insufficient citrate substitution.

In the absence of analytes, AuNPs remained capped with citrate with the corresponding peak occurrence frequencies shown in Figure 3(a) (black curve). Four high-intensity bands

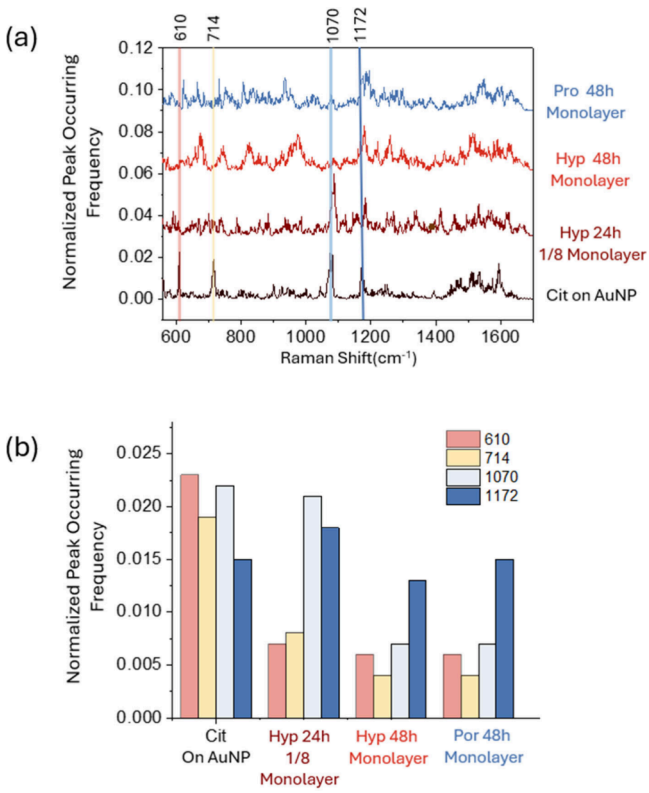


Figure 3. (a) The normalized peak occurring frequency generated in particle-in-pore sensors using AuNPs with citrate capping (black curve), 24 h incubation in solution with Hyp concentration for 1/8 monolayer (dark red curve), 48 h incubation in solution with Hyp concentration for a monolayer (red curve), and 48 h incubation in solution with Pro concentration for a monolayer (blue curve). Four characteristic peaks of citrate-capped AuNPs at 610 cm^{-1}, 714 cm^{-1}, 1070 cm^{-1}, and 1172 cm^{-1} are indicated by red, yellow, light blue, and blue guiding lines. (b) Histograms of the band height at 610 cm^{-1}, 714 cm^{-1}, 1070 cm^{-1}, and 1172 cm^{-1}, generated by AuNPs with citrate capping, 24 h incubation in solution with Hyp concentration for 1/8 monolayer, 48 h incubation in solution with Hyp concentration for a monolayer, and 48 h incubation in solution with Pro concentration for a monolayer.

at 610 cm^{-1} (δ_{COO}),[33] 714 cm^{-1} (δ_{OCO}),[30] 1070 cm^{-1} (ν_{CO}),[33] and 1172 cm^{-1} (δ_{COO})[33] were chosen to indicate citrate presence. The observed peak occurrence frequencies qualitatively correlate with the molecular population adsorbed on the AuNP surface, as shown in Figure 3(b). Under sufficient substitution conditions by the 48 h monolayer incubation, the

https://doi.org/10.1021/acs.nanolett.5c01177
Nano Lett. 2025, 25, 7499−7506

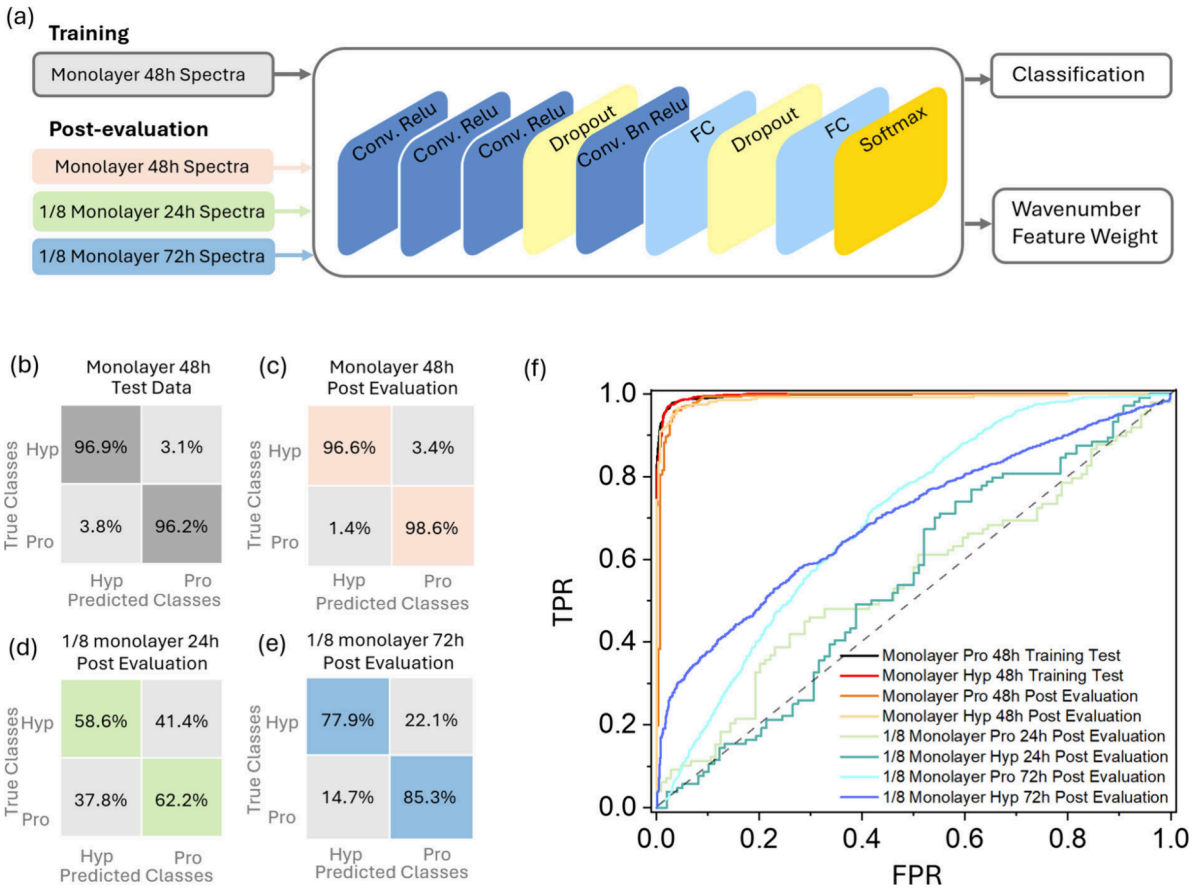


Figure 4. (a) Schematic representation of the CNN model architecture and the data classification framework. The confusion matrices illustrate the prediction accuracy for different data sets: (b) training data for monolayer Hyp and Pro after 48 h of incubation, (c) post evaluation of monolayer Hyp and Pro after 48 h of incubation, (d) post evaluation of 1/8 monolayer Hyp and Pro after 24 h of incubation, and (e) post evaluation of 1/8 monolayer Hyp and Pro after 72 h of incubation. (f) ROC curves for different data sets.

1070 cm^{-1} peak remains visible but its intensity significantly decreases. In the case of insufficient substitution, the 610 cm^{-1} and 714 cm^{-1} bands are reduced compared to those at the 0 h incubation case but remain higher than those in the 48 h incubation condition. Due to overlap with the Pro/Hyp vibrational modes, the 1172 cm^{-1} band does not exhibit a clear trend. After 48 h, the citrate bands become almost negligible, indicating minimal interference. However, this does not confirm complete citrate displacement, as analyte adsorption onto AuNPs occurs through a dynamic process. Even when the analyte concentration is sufficient to form a monolayer, a complete citrate substitution is unlikely. Additionally, citrate may form a secondary capping layer via hydrogen bonding,[34,35] making total elimination of citrate interference nearly impossible. Nevertheless, with the aid of a deep learning model, differences between Pro and Hyp can be extracted, while residual citrate signals are treated as common features. To minimize interference from citrates, the data set obtained from AuNPs with monolayer capping and 48 h incubation was used for initial 1D-CNN model training.

DEEP LEARNING ANALYSIS

Deep learning-based models have been widely used in spectroscopy data analysis due to their self-adaptation characterization and high accuracy. A 1D-CNN model has been used for the prediction of Pro and Hyp.[36,37] Here, we used MATLAB to run the codes of the 1D-CNN model for classification and post evaluation as shown in Figure 4(a). A total of 22,350 and 550 effective spectra with 1463 features were classified using 5-fold cross-validation for the CNN classification model and post evaluation. In the former, 80% of the spectra were used as the training set and 20% of the spectra were used as the test set, following these steps: (1) data set preparation, (2) training process, (3) evaluation metrics.

The model demonstrated accuracy higher than 96% in identifying single-molecule spectra of Pro and Hyp in the training set and post evaluation, as shown in Figure 4(b) and (c). It also showed tolerance to citrate interference by discriminating the data obtained under different incubation conditions. For the 72 h incubation and 1/8 monolayer coverage data set, the post evaluation accuracy is 77.9% and 85.3% for Hyp and Pro, respectively, in Figure 4(e). The lowering of the accuracy compared with the monolayer Pro or Hyp data set might be due to the interference of a larger amount of citrates remaining on AuNPs. However, the 24h 1/8 monolayer Pro and Hyp data set showed an accuracy of less than 70% in Figure 4(d), because the portion of citrate signals dominated the data set, which was also indicated by the peak occurrence frequency red curve in Figure 3(a) and histogram in Figure 3(b). Finally, the receiver operating characteristic (ROC) curves of the monolayer 48 h post evaluation of Pro and Hyp in Figure 4(f) show both high sensitivity and high specificity of the model in identifying new single-molecule spectra of Pro and Hyp.

https://doi.org/10.1021/acs.nanolett.5c01177
Nano Lett. 2025, 25, 7499−7506

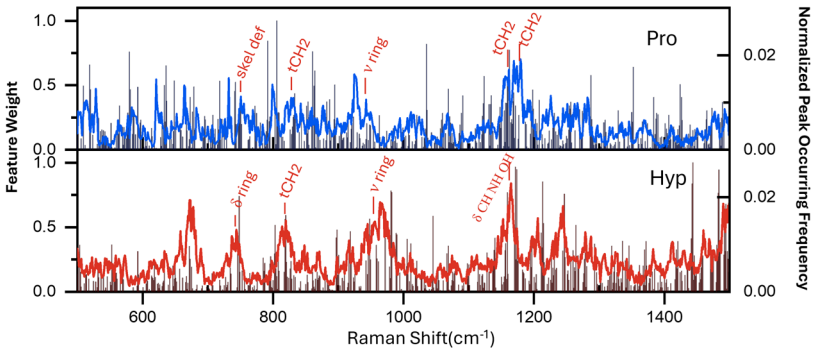


Figure 5. Feature weight extracted from the 1-CNN model. The blue and red curves show the normalized peak occurring frequency of a monolayer of Pro and Hyp with 48 h of incubation, respectively.

INTERPRETATION OF THE CNN OUTPUT

The 1D gradient-weighted feature visualization has become an emerging tool frequently used in recent works to extract spectral features from CNN models.[38,39] However, many previous studies have applied this method to analyze multimolecule data with subtle spectral changes, rather than single-molecule data, which is inherently inhomogeneous.[38,39] In this work, we extended its application to single-molecule spectral analysis and compared the extracted feature map of average spectra of the whole data set with the peak occurrence frequency to understand the performance of the 1D-CNN model. Figure 5 compares the positive feature weights extracted from the trained 1D-CNN model with the normalized peak occurrence frequency in the 600−1500 cm^{-1} region. The positive feature weights show narrow spikes per Raman shift (cm^{-1}), indicating the high importance of the Raman bands for discrimination. We observed that high positive feature weight spikes (≥0.4) occur in the regions 750−800 cm^{-1}, 800−825 cm^{-1}, 900−1000 cm^{-1}, and 1150−1200 cm^{-1}, which correspond to band differences induced by structural changes. Table 1 lists the SERS shift and vibration modes for Pro and Hyp. In the 750−800 cm^{-1} region, the Hyp ring vibration band appears at 740 cm^{-1}, while Pro shows a ring vibration-related band at 750 cm^{-1}.[20,29] In the region 800−825 cm^{-1}, Hyp and Pro show a peak at 817 cm^{-1} and 824 cm^{-1}, which corresponds to the ring fragment CH_2 movement.[21] The shift from 817 to 824 cm^{-1} is attributed to the structural change induced by the OH group on the ring. Similarly, at the region 900−1000 cm^{-1}, Hyp and Pro show peaks at 943 cm^{-1} and 955 cm^{-1}, respectively, corresponding to the ν_{Ring} mode.[21] In the 1150−1200 cm^{-1} region, Hyp shows peaks at 1165 cm^{-1}, respectively, corresponding to $\delta_{CH\ NH\ OH}$,[20] while Pro shows peaks at 1159 cm^{-1} and 1174 cm^{-1} corresponding to t_{CH2}.[29] Interestingly, the 1D-CNN model also gives high positive feature weight at positions with no peak occurrence, for example, the different valley at 1036 cm^{-1} for Pro and 1044 cm^{-1} and 1208 cm^{-1} for Hyp. Therefore, many highly weighted bands are aligned with the SERS bands with high occurring frequencies, indicating that the major contributions to the 1D-CNN performance originate from the difference in molecular structure between Hyp and Pro.

Table 1. SERS Bands Having a High Feature Weight in Figure 5 and the Corresponding Vibration Mode

Pro Band (cm^{-1})	Vibration Mode	Reference
750	skel deformation	29
824	t CH2	21
943	ν ring	21
1157, 1174	t CH2	29
Hyp Band (cm^{-1})	**Vibration Mode**	**Reference**
740	δ ring or ring torsion	20
817	t CH2	21
955	ν ring	21
1163	δ CH NH OH	20

In summary, we demonstrate, for the first time, the discrimination of proline and hydroxyproline at the single-molecule level using the particle-in-pore sensor assisted by the 1D-CNN model. The general characteristics of the single-molecule data set were obtained by calculating the peak occurrence frequency and plotting the frequency distribution. Using the peak occurrence frequency diagrams, we investigated the substitution of citrates by the analyte on AuNPs, revealing that sufficient incubation with the analyte solution reduced citrate bands in the spectra, indicating successful substitution. With the aid of a 1D-CNN model, we achieved more than 96% accuracy in discriminating proline from hydroxyproline. Even with 1/8 monolayer Pro and Hyp coverage, the discrimination accuracy was above 77%, demonstrating that the trained CNN model can tolerate some citrate interference. However, when excessive citrate remains on the AuNP surface, such as in the 1/8 monolayer, 24 h incubation data set, the 1D CNN model fails to correctly distinguish proline from hydroxyproline. The 1D gradient-weighted positive feature visualization was used to output normalized feature gradients, which were compared to peak occurrence frequencies. High positive feature weights were assigned to regions corresponding to vibration modes related to ring vibrations and OH bending. The successful discrimination of proline and hydroxyproline—one of the most challenging PTMs to detect at the single-molecule level—demonstrates the potential of our technology for various PTM discriminations. The single-molecule sensitivity of our method for detecting slight structural changes, including those with small Raman cross sections, also paves the way for site-specific PTM analysis in low-abundance peptides and proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.nanolett.5c01177.

Materials, fabrication, and CNN model; distribution of nanopore size; schematic illustration of the microfluidic

https://doi.org/10.1021/acs.nanolett.5c01177
Nano Lett. 2025, 25, 7499−7506

device; DLS size distribution of AuNP; calculation of amino acids and peptide concentration; citrate substitution on AuNP; Raman spectra of Pro and Hyp powder; AUC obtained from the CNN model (PDF)

AUTHOR INFORMATION

Corresponding Author

Jian-An Huang − *Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, 90220 Oulu, Finland; Research Unit of Disease Networks, Faculty of Biochemistry and Molecular Medicine and Biocenter Oulu, University of Oulu, 90220 Oulu, Finland;* orcid.org/0000-0002-6564-5972; Email: jianan.huang@oulu.fi

Authors

Yingqi Zhao − *Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, 90220 Oulu, Finland; Biocenter Oulu, University of Oulu, 90220 Oulu, Finland;* orcid.org/0000-0001-6054-8076

Kuo Zhan − *Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, 90220 Oulu, Finland; Biocenter Oulu, University of Oulu, 90220 Oulu, Finland*

Pei-Lin Xin − *Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, 90220 Oulu, Finland; Biocenter Oulu, University of Oulu, 90220 Oulu, Finland*

Zuyan Chen − *The Biomimetics and Intelligent Systems (BISG) research unit, Faculty of Information Technology and Electronic Engineering, University of Oulu, 90220 Oulu, Finland*

Shuai Li − *The Biomimetics and Intelligent Systems (BISG) research unit, Faculty of Information Technology and Electronic Engineering, University of Oulu, 90220 Oulu, Finland*

Francesco De Angelis − *Istituto Italiano di Tecnologia, 16163 Genoa, Italy;* orcid.org/0000-0001-6053-2488

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.nanolett.5c01177

Author Contributions

Yingqi Zhao fabricated the particle-in-pore devices, determined the Raman measurement protocol, collected Raman spectra, analyzed the Raman spectra, and wrote the manuscript. Kuo Zhan wrote the scripts for Raman data preprocessing, effective spectrum selection, and peak occurrence frequency calculation and developed a 1D-CNN model for PTM discrimination. Pei-Lin Xin contributed to writing the Raman spectra preprocessing and data selection scripts. Francesco De Angelis contributed to the device fabrication and revised the manuscript. Jianan Huang conceived the idea, acquired funding, supervised the work, and revised the manuscript. The manuscript was written with contributions from all authors. All authors have approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research receives support from Academy Research Fellow projects: TwoPoreProSeq (project number 347652), Biocenter Oulu emerging project (DigiRaman), spearhead project (LiveProSeq), and DigiHealth project (project number 326291), a strategic profiling project at the University of Oulu that is supported by the Academy of Finland and the University of Oulu. Jianan Huang acknowledges Prof. Jean-François Masson at the University of Montreal for valuable discussion on the CNN model and citrate interference.

REFERENCES

(1) Ma, H.; Han, X. X.; Zhao, B. Enhanced Raman spectroscopic analysis of protein post-translational modifications. *TrAC Trends in Analytical Chemistry* **2020**, *131*, 116019.

(2) Iommarini, L.; Porcelli, A.; Gasparre, G.; Kurelac, I. Non-Canonical Mechanisms Regulating Hypoxia-Inducible Factor 1 Alpha in Cancer. *Frontiers in Oncology* **2017**, *7*, 286.

(3) Semenza, G. Hypoxia-Inducible Factors in Physiology and Medicine. *Cell* **2012**, *148* (3), 399−408 Review.

(4) Kaelin, W. G. The von Hippel-Lindau tumor suppressor protein and clear cell renal carcinoma. *Clin. Cancer Res.* **2007**, *13* (2), 680s−684s.

(5) Zhou, T.; Erber, L.; Liu, B.; Gao, Y.; Ruan, H. B.; Chen, Y. Proteomic analysis reveals diverse proline hydroxylation-mediated oxygen-sensing cellular pathways in cancer cells. *Oncotarget* **2016**, *7* (48), 79154−79169.

(6) Jonasch, E.; Donskov, F.; Iliopoulos, O.; Rathmell, W. K.; Narayan, V. K.; Maughan, B. L.; Oudard, S.; Else, T.; Maranchie, J. K.; Welsh, S. J.; et al. Belzutifan for Renal Cell Carcinoma in von Hippel-Lindau Disease. *N Engl J. Med.* **2021**, *385* (22), 2036−2046.

(7) Smart, T. J.; Hamed, R. B.; Claridge, T. D. W.; Schofield, C. J. Studies on the selectivity of proline hydroxylases reveal new substrates including bicycles. *Bioorg Chem.* **2020**, *94*, 103386.

(8) Zurlo, G.; Guo, J.; Takada, M.; Wei, W.; Zhang, Q. New Insights into Protein Hydroxylation and Its Important Role in Human Diseases. *Biochim. Biophys. Acta* **2016**, *1866* (2), 208−220.

(9) Ponomarenko, E.; Poverennaya, E.; Ilgisonis, E.; Pyatnitskiy, M.; Kopylov, A.; Zgoda, V.; Lisitsa, A.; Archakov, A. The Size of the Human Proteome: The Width and Depth. *International Journal of Analytical Chemistry* **2016**, *2016* (1), 7436849.

(10) Rosen, C.; Rodriguez-Larrea, D.; Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat. Biotechnol.* **2014**, *32* (2), 179−181.

(11) Ensslen, T.; Sarthak, K.; Aksimentiev, A.; Behrends, J. C. Resolving Isomeric Posttranslational Modifications Using a Biological Nanopore as a Sensor of Molecular Shape. *J. Am. Chem. Soc.* **2022**, *144* (35), 16060−16068.

(12) Zhang, L.; Gardner, M. L.; Jayasinghe, L.; Jordan, M.; Aldana, J.; Burns, N.; Freitas, M. A.; Guo, P. Detection of single peptide with only one amino acid modification via electronic fingerprinting using reengineered durable channel of Phi29 DNA packaging motor. *Biomaterials* **2021**, *276*, 121022.

(13) Fahie, M. A.; Chen, M. Electrostatic Interactions between OmpG Nanopore and Analyte Protein Surface Can Distinguish between Glycosylated Isoforms. *J. Phys. Chem. B* **2015**, *119* (32), 10198−10206.

(14) Cao, C.; Magalhães, P.; Krapp, L. F.; Bada Juarez, J. F.; Mayer, S. F.; Rukes, V.; Chiki, A.; Lashuel, H. A.; Dal Peraro, M. Deep Learning-Assisted Single-Molecule Detection of Protein Post-translational Modifications with a Biological Nanopore. *ACS Nano* **2024**, *18* (2), 1504−1515.

(15) Iarossi, M.; Darvill, D.; Hubarevich, A.; Huang, J.; Zhao, Y.; De Fazio, A.; O'Neill, D.; Tantussi, F.; De Angelis, F. High-Density Plasmonic Nanopores for DNA Sensing at Ultra-Low Concentrations by Plasmon-Enhanced Raman Spectroscopy. *Adv. Funct. Mater.* **2023**, *33* (41), 2301934.

(16) Du, J.; Xu, G.; Liu, C.; Zhang, R. The role of phosphorylation and dephosphorylation of shell matrix proteins in shell formation: an in vivo and; in vitro study. *CrystEngComm* **2018**, *20* (27), 3905−3916.

(17) Geddis, A.; Mendive-Tapia, L.; Sujantho, A.; Liu, E.; McAughtrie, S.; Goodwin, R.; Vendrell, M.; Campbell, C. Label-

Free SERS Sensors for Real-Time Monitoring of Tyrosine Phosphorylation. *Anal. Chem.* **2024**, *95* (45), 17915−18310.
(18) Li, Y.; Li, D.; Cao, Y.; Long, Y. Label-free in-situ monitoring of protein tyrosine nitration in blood by surface-enhanced Raman spectroscopy. *Biosens. Bioelectron.* **2015**, *69*, 1−7.
(19) Sjöberg, B.; Foley, S.; Cardey, B.; Fromm, M.; Enescu, M. Methionine oxidation by hydrogen peroxide in peptides and proteins: A theoretical and Raman spectroscopy study. *Journal of Photochemistry and Photobiology B-Biology* **2018**, *188*, 95−99.
(20) Guerrero, A.; Aroca, R. Surface-enhanced Raman scattering of hydroxyproline. *J. Raman Spectrosc.* **2012**, *43* (4), 478−481.
(21) Cárcamo, J.; Aliaga, A.; Clavijo, E.; Garrido, C.; Gómez-Jeria, J.; Campos-Vallette, M. Proline and hydroxyproline deposited on silver nanoparticles. A Raman, SERS and theoretical study. *J. Raman Spectrosc.* **2012**, *43* (6), 750−755.
(22) Huang, J.-A.; Mousavi, M. Z.; Zhao, Y.; Hubarevich, A.; Omeis, F.; Giovannini, G.; Schütte, M.; Garoli, D.; De Angelis, F. SERS discrimination of single DNA bases in single oligonucleotides by electro-plasmonic trapping. *Nat. Commun.* **2019**, *10* (1), 5321.
(23) Huang, J.-A.; Mousavi, M. Z.; Giovannini, G.; Zhao, Y.; Hubarevich, A.; Soler, M. A.; Rocchia, W.; Garoli, D.; De Angelis, F. Multiplexed Discrimination of Single Amino Acid Residues in Polypeptides in a Single SERS Hot Spot. *Angew. Chem., Int. Ed.* **2020**, *59* (28), 11423−11431.
(24) Zhao, Y.; Iarossi, M.; De Fazio, A.; Huang, J.; De Angelis, F. Label-Free Optical Analysis of Biomolecules in Solid-State Nanopores: Toward Single-Molecule Protein Sequencing. *Acs Photonics* **2022**, *9* (3), 730−742.
(25) Tien, M.; Meyer, A.; Sydykova, D.; Spielman, S.; Wilke, C. Maximum Allowed Solvent Accessibilites of Residues in Proteins. *PLoS One* **2013**, *8* (11), No. e80635.
(26) Franco-Ulloa, S.; Tatulli, G.; Bore, S.; Moglianetti, M.; Pompa, P.; Cascella, M.; De Vivo, M. Dispersion state phase diagram of citrate-coated metallic nanoparticles in saline solutions. *Nat. Commun.* **2020**, *11* (1), 5422.
(27) Buglak, A.; Kononov, A. Comparative study of gold and silver interactions with amino acids and nucleobases. *Rsc Advances* **2020**, *10* (56), 34149−34160.
(28) Zrimsek, A.; Chiang, N.; Mattei, M.; Zaleski, S.; McAnally, M.; Chapman, C.; Henry, A.; Schatz, G.; Van Duyne, R. Single-Molecule Chemistry with Surface- and Tip-Enhanced Raman Spectroscopy. *Chem. Rev.* **2017**, *117* (11), 7583−7613.
(29) Herlinger, A. W.; Long, T. V. II. Laser-Raman and infrared spectra of amino acids and their metal complexes. III. Proline and bisprolinato complexes. *J. Am. Chem. Soc.* **1970**, *92* (22), 6481−6486.
(30) SIIMAN, O.; BUMM, L.; CALLAGHAN, R.; BLATCHFORD, C.; KERKER, M. SURFACE-ENHANCED RAMAN-SCATTERING BY CITRATE ON COLLOIDAL SILVER. *J. Phys. Chem.* **1983**, *87* (6), 1014−1023.
(31) Shin, H.; Yeon, G.; Choi, H.; Park, S.; Lee, K.; Kim, Z. Frequency-Domain Proof of the Existence of Atomic-Scale SERS Hot-Spots. *Nano Lett.* **2018**, *18* (1), 262−271.
(32) Li, W.; Guo, L.; Ding, X.; Ding, Y.; Ji, L.; Xia, X.; Wang, K. High-Throughput Single-Molecule Surface-Enhanced Raman Spectroscopic Profiling of Single-Amino Acid Substitutions in Peptides by a Gold Plasmonic Nanopore. *ACS Nano* **2024**, *18* (29), 19200−19207.
(33) MUNRO, C.; SMITH, W.; GARNER, M.; CLARKSON, J.; WHITE, P. CHARACTERIZATION OF THE SURFACE OF A CITRATE-REDUCED COLLOID OPTIMIZED FOR USE AS A SUBSTRATE FOR SURFACE-ENHANCED RESONANCE RAMAN-SCATTERING. *Langmuir* **1995**, *11* (10), 3712−3720.
(34) Park, J.; Shumaker-Parry, J. Structural Study of Citrate Layers on Gold Nanoparticles: Role of Intermolecular Interactions in Stabilizing Nanoparticles. *J. Am. Chem. Soc.* **2014**, *136* (5), 1907−1921.
(35) Majzik, A.; Fülöp, L.; Csapó, E.; Bogár, F.; Martinek, T.; Penke, B.; Bíró, G.; Dékány, I. Functionalization of gold nanoparticles with amino acid, β-amyloid peptides and fragment. *Colloids and Surfaces B-Biointerfaces* **2010**, *81* (1), 235−241.
(36) Li, J.; Dukes, P.; Lee, W.; Sarkis, M.; Vo-Dinh, T. Machine learning using convolutional neural networks for SERS analysis of biomarkers in medical diagnostics. *J. Raman Spectrosc.* **2022**, *53* (12), 2044−2057.
(37) Lu, P.; Lin, D.; Chen, N.; Wang, L.; Zhang, X.; Chen, H.; Ma, P. CNN-assisted SERS enables ultra-sensitive and simultaneous detection of Scr and BUN for rapid kidney function assessment. *Analytical Methods* **2023**, *15* (3), 322−332.
(38) Shi, G.; Wu, H.; Luo, S.; Lu, X.; Ren, B.; Zhang, Q.; Lin, W.; Chen, R.; Guo, P.; Chen, H.; et al. 1D Gradient-Weighted Class Activation Mapping, Visualizing Decision Process of Convolutional Neural Network-Based Models in Spectroscopy Analysis. *Anal. Chem.* **2023**, *95* (26), 9959−9966.
(39) Xiong, C.; Zhu, S.; Yan, D.; Yao, Y.; Zhang, Z.; Zhang, G.; Chen, S. Rapid and precise detection of cancers via label-free SERS and deep learning. *Anal. Bioanal. Chem.* **2023**, *415* (17), 3449−3462.

https://doi.org/10.1021/acs.nanolett.5c01177
Nano Lett. 2025, 25, 7499−7506