UNIVERSITY OF CALIFORNIA,
IRVINE

Decoding Bacterial Metabolism: surface enhanced Raman scattering and deep learning

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Materials Science and Engineering

by

Hong Wei

Dissertation Committee:
Professor Regina Ragan, Chair
Associate Professor Allon I. Hochbaum
Assistant Professor Joe Patterson

2023

DEDICATION

To

my parents

for letting me follow my heart and supporting all my choices

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to express my heartfelt gratitude to my advisor, Professor Regina Ragan, for her invaluable guidance, unwavering support, and encouragement throughout my PhD journey. Her dedication to my success has been exceptional, always being available for discussion and providing me with critical feedback on my research work. I am grateful to her for believing in me even when I did not believe in myself. Her mentorship has taught me important skills, such as critical thinking and effective scientific communication, which have helped me grow both as a researcher and as a person. I am incredibly lucky to have had her as my advisor.

I would like to extend my sincere thanks to my committee members, Professor Allon I. Hochbaum, for his insightful guidance on the project and invaluable feedback, and Professor Joe Patterson, for his co-advisory role and for creating a supportive and encouraging atmosphere during our monthly meetings, especially during the challenges of the COVID-19 pandemic.

I would like to thank Dr. Will Thrift for his mentoring when I first joined the group and for introducing me to the habit of reading the New York Times and the Medium. I would like to thank my collaborator, Dr. Serxho Selmani, for the enjoyable chats we had during confocal fluorescence experimental sessions, which made the long hours in the dark room more bearable. I would like to thank my collaborator, Yixin Huang, for his meticulous work in bacteria culturing, which was critical for the success of our experiments. I would like to thank my collaborator, Yen Hsiang Huang, for his help in preparing wastewater samples. I would like to express my gratitude to my labmates, Adrian E Garcia, Chloe Groome, Hector Pascual, Peter J Santiago, Judy Tran, for their camaraderie and support. Additionally, I would like to thank the

undergraduate students I mentored, Naeiri Shahmirzaeian Savara and Gabriel D Reginato for their valuable contributions to my research.

I would like to express my deep appreciation to my parents, for their unconditional love and support. Their encouragement and belief in me have been a constant source of inspiration and motivation throughout this journey. I would also like to thank my cat Romeo, for keeping me company and purring so loudly.

Thank you all for making this Ph.D. thesis possible.

Portions of Chapter 1 and 4 reprinted with permission from William John Thrift, Sasha Ronaghi, Muntaha Samad, Hong Wei, Dean Gia Nguyen, Antony Superio Cabuslay, Chloe E Groome, Peter Joseph Santiago, Pierre Baldi, Allon I Hochbaum, Regina Ragan. ACS nano 14 (11), 15336-15348. Portions of Chapter 5 reprinted with permission from Hong Wei, Yixin Huang, Peter J Santiago, Khachik E Labachyan, Sasha Ronaghi, Martin Paul Banda Magana, Yen-Hsiang Huang, Sunny C. Jiang, Allon I. Hochbaum and Regina Ragan, PNAS. 2023;120(7):e2210061120.

Finally, I would like to acknowledge the support of the National Science Foundation (CBET-1926612), the National Science Foundation Materials Research Science and Engineering Center program through the UC Irvine Center for Complex and Active Materials (DMR-2011967), PhD Bridge Fellowship, Inclusive Excellence Ambassador Fellowship, and UCI Graduate Division Completion Fellowship. I am also grateful to UCI's Laser Spectroscopy Facility and the Irvine Materials Research Institute for providing access to their facilities.

VITA

Hong Wei

Research Experience

Prof. Regina Ragan's Research Group, Graduate Student Researcher, UC Irvine 2017-2022

Machine Learning Analysis of Complex Biological Surface Enhanced Raman Scattering (SERS) Data

- Designed, implemented and scaled machine learning workflows via TensorFlow-Keras for 1D spectra data from Raman spectroscopy including data collection, data pre-processing, building datasets, model training and refinement, evaluation and deployment to production
- Developed support vector machine classification models and convolutional neuron network high-dimensional multivariate regression models to detect the rapid metabolite profile change in bacteria due to heavy metal stress respond. This work demonstrated the limit of detection for arsenic and chromium in water achieved 6 orders of magnitude lower than EPA regulatory limit
- Deployed and delivered an on-field water quality plasmonic optical sensing system via transfer learning achieved over 92% accuracy for identification wastewater samples, contains heavy metal ion above or below the EPA regulatory limit. This work demonstrated the selectivity ability of the sensing system over foreign ions from practical wastewater samples
- Developed support vector machine models to determine differences between wild-type and gentamicin resistant type *E.coli* at 10^2 CFU/mL. This work demonstrated rapid bacteria strain identification
- Collaborated in a multidisciplinary team of chemists, biologists, and ecologists to design, manufacture and validate biosensing platforms for the identification of heavy metal ion in wastewater by metabolite sensing

Applied Physics of Plasmonic Nanostructures

- Manufactured plasmonic nanostructures via oscillation field electrohydrodynamic flow, optimized fabrication method resulting in 100x reduction of required amount plasmonic nanoparticle solution, decreased fabrication time (from 30 to 4 minutes), and allowing for room temperature development
- Developed method to control for nanoparticle cluster size and improved uniformity and density of surface nanostructure
- Determined nanoparticle cluster size, density, and uniformity with scanning electron microscopy
- Investigated localized surface plasmon resonance (LSPR) and dissipative assembly of plasmonic nanostructures using Raman microscopy and *in situ* fluorescence microscopy

Plasmonic Optical Sensor Fabrication

- Fabricated templates to assist chemical fiber dissipative assembly using photolithography and electron beam lithography in cleanroom (class 100)

- Invented novel oscillation field electrohydrodynamic flow driven chemical self-assembly method utilizing gold nanoparticles and EDC/sNHS crosslinking for more efficient manufacturing of plasmonic optical sensors
- Engineered the combination of plasmonic optical sensor and whole cell sensor with state-of-the-art performance utilizing metabolite system and stress responds of cells for antimicrobial susceptibility testing, heavy metal ion detection in drinking water, bacteria strain identification, nutrient source identification, and diauxic shift of various sugar types

UC Irvine Material Research Institute, Lab Assistant, UC Irvine
2018-2021

- Trained over 30 users on FEI Magellan scanning electron microscopy/EDS/electron beam lithography and over 30 users on Anton Paar Tosca atomic force microscopy
- Wrote and updated SOP for Anton Paar Tosca atomic force microscopy
- Provided 24/7 support for facility users from industry and academia, helped identified root causes

Education

PhD, Material Science and Engineering, UC Irvine, Irvine, CA, USA 2018 – 2023

MS, Material Science and Engineering, UC Irvine, Irvine, CA, USA 2016 – 2018

BS, Electronic Science and Technology, Inner Mongolia University, Hohhot, China 2012 – 2016

Teaching, Mentoring, & Leadership

- Teaching assistant for material science undergraduate courses: Principles of materials science and engineering, material kinetics and phase transformation, and electronic and optical properties of material 2018 – 2021
- Inclusive Excellence Ambassador, mentored 6 freshman year undergraduate student from underrepresented community to prepare their arrival to UC Irvine 2020
- Graduate Resource Center International Peer Mentor, mentored 12 international students from different departments to prepare their arrival to a new country and graduate school 2017
- Mentored 4 undergraduate students and 2 junior PhD students on functionalization of plasmonic gold nanoparticle with lipoic acid ligand and ligand exchange to engineer moiety to have affinity property to a variety of metabolites

Fellowship/Awards

Research Highlight from UC Irvine 2023

UCI Graduate Division Completion Fellowship 2022

Inclusive Excellence Ambassador Fellowship, UC Irvine 2020

PhD Bridge Fellowship, UC Irvine 2018 – 2019

Conferences and workshops

- Materials Research Society, oral presentations 2022

Title: Machine Learning Analysis of Spectral Data using Bacterial Metabolic Networks for Signal Amplification

- The Minerals, Metals & Materials Society, oral presentations 2022

Title: Machine Learning Analysis of Spectral Data using Bacteria for Signal Amplification

- American Chemistry Society, poster presentations 2022

Title: Sensing platform for heavy metal ion detection with a bioinspired approach

- Scientific Exchange, poster presentations 2021

Title: Bacterial Stress Response Monitored by Surface Enhanced Raman Scattering and Machine Learning

- High-Performance Computing and Data Science Summer Institute, UC San Diego 2020

Publications

Peer-Reviewed Journal Articles

- Decoding the metabolic response of Escherichia coli for sensing trace heavy metals in water (first author, *PNAS* 2023, 120 (7), e2210061120)
- Sensing antibiotics in wastewater using surface enhanced Raman scattering (second author, *ACS ES&T* 2023, accepted)
- Electrically fueled active supramolecular materials (co-author, *J. Am. Chem. Soc.* 2022, 144, 17, 7844–7851)
- Deep learning analysis of vibrational spectra of bacterial lysate for rapid antimicrobial susceptibility testing (co-author, *ACS Nano* 2020, 14, 11, 15336–15348)

Work-in-Progress

- Nanoantennas as Reporters of Dissipative Assembly in Oscillatory Electric Fields (first author, in preparation)

ABSTRACT OF THE DISSERTATION

Decoding Bacterial Metabolism: surface enhanced Raman scattering and deep learning

by

Hong Wei

Doctor of Philosophy in Material Science and Engineering

University of California, Irvine, 2023

Professor Regina Ragan, Chair

Self-assembly utilizes thermodynamic driving forces in order to achieve molecular scale control with less waste and lower costs than lithographic methods. Thus the ability to understand reactivity between functional groups on molecular structures in confined geometries provides knowledge on how to organize building blocks such as molecules and nanoparticles (NPs) into supramolecular structures and metasurfaces using scalable methods. The overall objective of my work in understanding driving forces for forming plasmonic nanostructures on surfaces from colloidal solutions was to design plasmonic nanostructures with molecular scale control for sensor applications. I show that the resulting plasmonic nanostructures can produce unique light matter interactions, controlling localized interactions with plasmons and phonons for molecular sensors with detection limits at single molecule and single cell levels.

Colloidal NPs are typically stabilized via electrostatic repulsion, and thus, chemical crosslinking in solution can lead to uncontrolled aggregation. Here, I developed a robust fabrication method using long range electrokinetic forces, oscillation field electrohydrodynamic (AC-EHD) flow, to drive chemical reactions between suspended NPs on working electrode surfaces. This process forms "nanogaps" between the NPs and enables manufacturing of surface

enhanced Raman scattering (SERS) sensors with record performance. By varying field frequency, isolated NPs can be deposited onto the surface at 100 Hz. At higher frequencies, e.g., 500-1500 Hz, AC-EHD flow plays a more important role to form 2D Au nanostructures with controlled nanogap spacing. In order to fabricate plasmonic nanostructures with high density over large areas, a two step deposition was utilized to first deposit isolated NPs as seeds then apply higher frequencies to grow to clusters. The resulting SERS sensor is highly sensitive to bacterial metabolites, many which have aromatic rings. Thus the bacterial metabolic responses to environmental stimuli is robustly differentiable when using machine learning (ML) algorithms to analyze the spectral data. Studies on nutrient deprivation are performed to understand general stress response, which can be useful for process monitoring in biotechnology industries. Changing nutrient source of *E. coli* cultures from glucose to the less preferred xylose and sucrose, provides insight on metabolic network response. The metabolite profile of bacteria in response to antibiotic treatment differentiates resistance and susceptibility as soon as 5 min to produce a new platform for rapid antimicrobial susceptibility testing. Similarly, bacterial stress responses to toxic heavy metals were used to monitor water quality. We show that detection of arsenic ions (As^{3+}) and chromium ions (Cr^{6+}) is possible at the single cell level. ML analysis of the vibrational spectra of metabolites released in response to As^{3+} and Cr^{6+} exposure detects concentrations 10^8 times lower than those leading to cell death. Transfer learning of trained algorithms to test contaminants in tap water and wastewater were able to achieve 92% accuracy.

Chapter 1

Introduction

1.1 The history and development of colloidal self-assembly

The observation of colloidal suspensions dates back to the 19th Century. In 1857, Michael Faraday (1791-1867) delivered an exciting lecture to the Royal Society in London about the suspensions of "Ruby" Au and its properties, which are completely different from bulk material.[1] The term colloidal was introduced in 1861 by Graham.[2] Faraday's pioneer work marks the initiation of the modern colloidal science research field, which has been recognized as an essential part of nanotechnology.[3] Colloids are defined to have particles as solute in solution with diameter in the range of 1 to 1000 nm, where unique properties compared with bulk material may be observed. When the particle diameter is in the range of 1-100 nm, they are called nanoparticles (NPs). NPs with high valence electron density, including metals such as Au, Ag, and Al, semiconductors such as InAs,[4] and metal oxides such as ITO,[5] may exhibit plasmonic effects in result from the interaction between electromagnetic wave and excitation of valence electrons at metallic-dielectric interfaces. The excitation of plasmons can improve the performance of photovoltaics, optoelectronic devices, photocatalysis, and chemical and biological sensors;[6] My work involves the use of Au NPs in biosensors and environmental sensors. Both Au and Ag NPs are easily fabricated with control of size and shape, and surfaces can be readily functionalized with chemical and biomolecular ligands.[7] In addition, Au NPs are more stable and compatible with living systems for biomedical applications, including imaging agents and drug delivery.[8,9] Taken together, colloidal Au NP systems have been demonstrated to

have broad impact on physics, chemical, and biological applications and are of great interest for further exploration.

1.2 Colloidal self-assembly

Colloidal self-assembly is a process in which colloidal particles are assembled into ordered structures through various methodologies, including chemical approaches, interparticle forces, external forces, and their combinations. Chemical reactions between surface ligands of two NPs can form chemical bonds, linking Au NPs together.[10] Besides chemical ligands, biomolecules such as DNA,[5] protein linkers,[11,12] and antibodies[13] can also be used to conjugate with Au NPs. For example, the DNA origami technique has been used as templates for assembly of NPs into three dimensional structures.[14] Au NP's interaction with protein denaturation can lead to a compact-packed assembly and retention of protein structure can lead to an assembly with larger interparticle spacings.[15] Moreover, antibody–antigen recognition can also self-assemble Au NPs into macroscopic materials.[16] The combination of chemical reactions with interparticle forces, such as capillary forces and electrostatic forces, can achieve controllable assembly. For example, DNA functionalized Au NPs can assemble through base pair interaction or electrostatic attraction. Selective functionalization can be performed using templates fabricated from lithography methods to confine the deposition of colloidal building blocks onto selected regions and achieve large area ordering based on electrostatic or hydrophobic/hydrophilic interactions.[17,18]

External forces, such as optical, electric, and magnetic fields, offer a powerful means of driving colloidal self-assembly with increased control and diversity of assembled structures. For example, optical plasmonic tweezers can manipulate NPs by concentrating an incident laser into a nanoscale region to enhance the electric field, strengthening the optical force to trap NPs;[19]

magnetic fields can direct NP organization by magnetic dipole-dipole interactions, which can switch between attractive and repulsive interactions by controlling the angle between the magnetic field and the direction connecting the dipoles;[20] Furthermore, electrokinetic phenomena, such as electrophoresis, dielectrophoresis, and electrohydrodynamic flow, can be used to direct lateral assembly of NPs on electrode surfaces to form two dimensional NP clusters or thin films.[21] Section 2.2 will discuss electrohydrodynamic flow in more detail.

1.3 Colloidal dissipative assembly

Self-assembly processes can occur far from equilibrium, known as dissipative assembly, which generates dynamic and adaptive systems that consume energy to maintain their state. Dissipative assembly is crucial for life, as exemplified by the self-replication of DNA, the construction of cell walls, and immune function.[22] Synthetic self-assembly may utilize similar principles as organic systems, and the availability of finely tuned external controls and materials in synthetic systems may unlock additional functionalities and unique structures not commonly encountered in nature. For example, dissipative assembly in materials such as gels[23] can lead to self-healing materials. In addition, dissipative assembly in colloidal NPs solutions can produce systems with spatio-temporal control of tunable optical properties.[24]

In the case of NPs in colloid systems, various external stimuli such as temperature, magnetic fields, light, electric fields[25,26] and their combinations have been explored as driving forces for dissipative assembly.[27,28] For example, temperature can trigger the dissipative assembly of NPs with functionalized thermosensitive molecules or polymers;[29,30] a magnetic field can direct the anisotropic assembly of patchy NPs;[31] additionally, NPs functionalized with photoswitchable moieties can undergo aggregation and disassembly in response to light.[32] Chemically driven dissipative assembly is a process that utilizes the addition of chemical fuel to

introduce interactions between NPs, where the consumption of the fuel may bring the system back to equilibrium for disassembly.[33] Grötsch *et al.* have demonstrated the use of 1-Ethyl-3-(3-(dimethylamino)propyl)carbodiimide (EDC) as a fuel to drive the dissipative assembly of 5 nm Au NPs. Specifically, the addition of EDC fuel induced the assembly of the NPs after 30 min, and subsequently, the consumption of the fuel resulted in disassembly after 24 hours.[34] Electric stimuli have been reported for the dissipative assembly of microparticles; however, their potential in the dissipative assembly of NPs remains relatively unexplored.[27] Recent research has demonstrated the use of direct current bias to achieve dissipative assembly of NPs via electrostatic effects. Specifically, a bias of 0.7 V was employed for assembly and 0.1 V was used for disassembly.[35,36] Electrohydrodynamic (EHD) flow is an electrokinetic phenomenon that can achieve lateral assembly of NPs. The reversibility of the EHD flow,[37] which allows for the disassembly of assemblies if the bias is not maintained, is hindered by the kinetic trapping of resulting structures on surfaces due to electrostatic interactions, making it difficult to experimentally monitor and thereby understand the disassembly behavior of NPs at the electrode-liquid interface. Moreover, the dynamic optical responses of the assembled structures under oscillatory electric fields have not been extensively explored, limiting the further development of electrically tunable optical materials.

1.4 Optical properties of Au colloids

1.4.1 Localized surface plasmon resonance

When the incident laser impinges at a metallic-dielectric interface, it can drive the collective oscillation of delocalized conduction electrons at the dielectric-metal interfaces when the frequency of the incident light matches the resonance frequency of free electrons in the plasmonic materials. This phenomenon is called surface plasmon resonance (SPR).[38] For the

metallic-dielectric interface, the complex reflection coefficient r_p for p-polarized incident laser electric field can be described by the Maxwell equations,[39]

$$r_p = \frac{E_i}{E_r} = |r_p| e^{j\varphi} = \left|\frac{\tan(\alpha-\beta)}{\tan(\alpha+\beta)}\right| e^{j\varphi} \qquad \text{Equation 1.1}$$

In conductive nanostructures, the SPR can be highly localized to a specific position, which is referred to as localized SPR (LSPR). NPs of interest that can generate a strong LSPR effect are called plasmonic NPs, which are typically metals such as Au, Ag and Al, semiconductors such as InAs, and metal oxide such as ITO, because they show strong SPR in UV and near-infrared regions. LSPR will lead to the resonant absorption or scattering of the incident light.[40] The frequency depends on NP's size, shape, and surrounding dielectric environment.

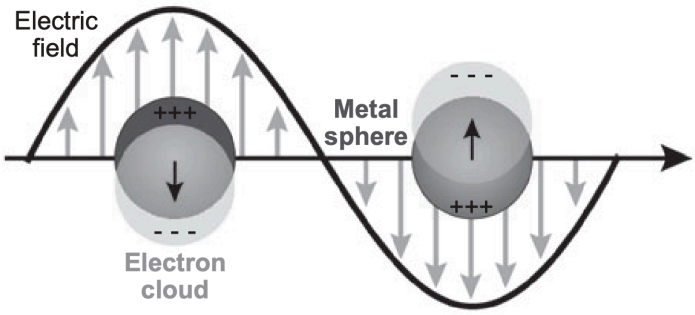

Figure 1.1 Illustration of LSPR effect.[41]

1.4.2 Surface enhanced Raman scattering

Raman scattering is an analytical technique that provides vibrational fingerprint of molecules via non-linear inelastic scattering events. However, Raman is an intrinsically very weak phenomenon due to the small Raman cross section.[42] In 1974, Fleischmann et al. observed an unusual enhancement of Raman signal from pyridine adsorbed on a roughened silver

surface.[43] In 1977, Richard P. Van Duyne et al. first termed this phenomenon as surface enhanced Raman scattering (SERS). Two enhancement mechanisms contribute to the amplified Raman signal, electromagnetic and chemical mechanisms. The electromagnetic mechanism originates from the interaction of incident light to the valence electrons at the metallic-dielectric interface, which can result in 10^6 fold intensity enhancement of the Raman scattering.[44] Chemical enhancement arises from a modification of the polarizability of a molecule. The contribution of Raman signal enhancement from the chemical enhancement is usually considered much smaller than the electromagnetic mechanism; its magnitude may reach 10^2-10^4.[45] In Chapter 2.1, I provide a detailed explanation of the electromagnetic mechanism in SERS.

SERS is a highly sensitive analytical technique that results from the LSPR of metal NPs when illuminated with laser light at the plasmon resonance frequency. The LSPR induces a strong electromagnetic field within the confined nanogap between the metal NPs, which interacts with molecules located in close proximity, resulting in an amplified Raman signal.[46] SERS surfaces can be used to investigate biochemical processes in cells[47] and tissues[48] and integrated into microfluidic devices to detect pathogens.[49] The high enhancement factor enabled by SERS has opened up the possibility of detecting very few molecules, even at a single molecule level.[50] However, a long standing challenge in SERS is the fabrication of sensors with highly reproducible responses. I developed a novel synthesis protocol for the formation of two-dimensional Au NP clusters using AC-EHD flow. The protocol involves the growth of Au NP clusters from isolated single Au NPs on the surface, serving as seeds, with controllable uniform gap spacing and high reproducibility. A comprehensive description of the experimental procedure can be found in Chapter 3.

1.5 Machine learning for analytical technologies

Machine learning (ML), especially deep learning, is emerging as a powerful tool to interpret data obtained from analytical technologies such as Raman scattering. The combination of Raman scattering and deep learning has shown promise for antimicrobial susceptibility testing, and recent reports suggest that analysis can benefit from enhancements in SERS.[51] Since the use of principal component analysis to demonstrate single-molecule detection by Le Ru *et al.*,[52] great progress has been made in applying sophisticated machine learning techniques to analyze SERS spectra. These include fully connected artificial neural networks for analyte concentration regression,[53] DNA classification,[54] cancer detection,[55] convolutional neural networks for the classification of metabolite signals,[56] support vector machines for the classification of drug use from urine,[57] and genetic algorithms for cancer diagnoses.[58]

1.6 Thesis outline

In this remainder thesis, Chapter 2 offers scientific explanations of key concepts that are essential for understanding the research question. Specifically, the chapter provides an overview of SERS and the underlying electromagnetic and chemical mechanisms that contribute to its ultimate sensitivity. The chapter delves deeper into the influence of nanogap size and geometry on electromagnetic mechanism of SERS, which is critical for optimizing the sensitivity of SERS-based sensing platforms. Additionally, the chapter discusses the mechanism of electrohydrodynamic (EHD) flow and its relevance to the assembly of plasmonic NPs for the fabrication of SERS sensors. Oscillatory electric field frequency can affect EHD flow, which is an important consideration for ensuring the stability and reproducibility of SERS signals. Furthermore, the chapter explores the application of machine learning in analytical chemistry, a field that has gained popularity in recent years. The chapter outlines the basics of machine learning algorithms and how they can be used to analyze large and complex datasets generated

by analytical techniques like SERS. Overall, by offering a comprehensive overview of SERS mechanisms, EHD flow, and ML in analytical chemistry, Chapter 2 highlights the importance of interdisciplinary approaches to developing more sensitive analytical techniques.

Chapter 3 describes a novel methodology for regulating the interaction between surface ligands on Au NPs assisted by EHD flow. This chapter studies the effect of frequency under an applied AC potential in influencing the lateral arrangement of clusters of Au NP formed as a result of EHD flow and time scales for assembly and disassembly. The resultant electromagnetic field enhancements associated with excitation of the LSPR of NP as they are assembled into clusters, provides a means of tracking assembly in situ using optical methods. Dissipative assembly is monitored using in situ fluorescence microscopy and SERS. Previous studies have reported that metal enhanced fluorescence is maximum when the emission spectra of the fluorophore is tuned to the LSPR frequency of metal nanostructures. Lateral assembly of Au NP can shift the LSPR from approximately 520 nm toward the emission line of Nile Red, 561 nm. We show that increases and decreases in fluorescence intensity are correlated with turning on and off external electric stimuli and thus able to report dissipative assembly of Au NP in real time. Furthermore, SERS spectra measured in situ during EHD assembly of Au NP functionalized with 4-mercaptobenzoic (4-MBA) as reporter molecules shows signal saturates as early as 20 s. To the best of our knowledge, this study presents the first demonstration of *in situ* monitoring of the dissipative assembly of NPs under AC-EHD flow using confocal fluorescence spectroscopy.

Chapter 4 demonstrates the SERS + deep learning approach is a promising candidate for use in clinical samples for rapid antimicrobial susceptibility testing (AST) and bacteria nutrient metabolism identification. The response of *P. aeruginosa* and *E. coli* bacterial communities to

antibiotics is rapidly detected in SERS spectral data when using sensor surfaces with controlled nanogap spacing and chemistry. Deep learning analysis of SERS data is able to differentiate the response of untreated cells from those exposed to antibiotics in 10 min post exposure with greater than 99% accuracy and temporally follow the evolution with 5 min resolution with greater than 99% accuracy. The bacterial response to varying antibiotic doses is differentiated with greater than 96% accuracy from untreated bacteria, even when treated with antibiotic dosages up to 10-fold lower than the minimum inhibitory concentration observed in conventional growth assays. We also found that the metabolism features captured by the unsupervised PCA model can correlate with energy nucleotides and metabolic pathways involved in nutrient uptake, suggesting that changes in nucleotide concentrations are consistent with the nutrient-dependent metabolic signatures observed. The SVM classification model achieved high accuracy on differentiating different nutrient metabolism profile with the false prediction rate of control spectra being classified as nutrient conditions was less than 0.6%. The phenomenon of carbon catabolite repression in bacteria refers to the preference of utilizing one nutrient source over another when multiple sources are available. Section 4.3 in this chapter investigated the impact of glucose and xylose as nutrient sources on the growth of *E.coli*. Using tSNE analysis, we observed that the control and xylose clusters were more similar since they didn't exhibit growth, whereas the glucose and the mixture containing glucose and xylose clusters were more alike since they both consumed glucose to support growth.

Chapter 5 describes a work that uses the sensitivity of the *Escherichia coli (E. coli)* stress response to transduce the signal of Cr^{6+} and As^{3+} ions into chemical signals that are detected with chemically assembled SERS surfaces. A support vector machine (SVM) model achieves higher than 97% classification accuracy for decoding *E. coli* stress response to different concentrations

of metal ions for concentrations as low as 68 pM for Cr^{6+} and 5 pM for As^{3+}. Due to their distinct mechanisms of toxicity in bacteria, this sensing platform also distinguishes the metabolic response of As^{3+} and Cr^{6+} with high accuracy when analyzed with SVM models. In addition, convolutional neural networks (CNN) show sensitive and quantitative determination of concentrations across a dynamic range of 0.68 pM - 68 μM for Cr^{6+} and 5 fM - 5 mM for As^{3+}. At the lowest concentrations investigated, the metabolic response is detectable when the ratio of metal ions to bacterium in solution is 0.6 for As^{3+} and 8.2 for Cr^{6+}. Finally, by using a pretrained model for analysis of previously unseen tap water and wastewater samples spiked with As^{3+}, SERS detection and ML analysis requires only 80 spectra per class (40 sec total acquisition time) to achieve greater than 96.5% accuracy for classifying concentrations above or below the WHO recommended limit.

References

(1) Faraday, M. X. The Bakerian Lecture. —Experimental Relations of Gold (and Other Metals) to Light. *Philos. Trans. R. Soc. Lond.* **1857**, *147*, 145–181. https://doi.org/10.1098/rstl.1857.0011.

(2) Graham, T. X. Liquid Diffusion Applied to Analysis. *Philos. Trans. R. Soc. Lond.* **1861**, *151*, 183–224. https://doi.org/10.1098/rstl.1861.0011.

(3) Whitesides, G. M.; Grzybowski, B. Self-Assembly at All Scales. *Science* **2002**, *295* (5564), 2418–2421. https://doi.org/10.1126/science.1070821.

(4) Arcangeli, A.; Rossella, F.; Tomadin, A.; Xu, J.; Ercolani, D.; Sorba, L.; Beltram, F.; Tredicucci, A.; Polini, M.; Roddaro, S. Gate-Tunable Spatial Modulation of Localized Plasmon Resonances. *Nano Lett.* **2016**, *16* (9), 5688–5693. https://doi.org/10.1021/acs.nanolett.6b02351.

(5) Xi, M.; Reinhard, B. M. Localized Surface Plasmon Coupling between Mid-IR-Resonant ITO Nanocrystals. *J. Phys. Chem. C* **2018**, *122* (10), 5698–5704. https://doi.org/10.1021/acs.jpcc.8b01283.

(6) Thrift, W. J.; Nguyen, C. Q.; Darvishzadeh-Varcheie, M.; Zare, S.; Sharac, N.; Sanderson, R. N.; Dupper, T. J.; Hochbaum, A. I.; Capolino, F.; Abdolhosseini Qomi, M. J.; Ragan, R. Driving Chemical Reactions in Plasmonic Nanogaps with Electrohydrodynamic Flow. *ACS Nano* **2017**, *11* (11), 11317–11329. https://doi.org/10.1021/acsnano.7b05815.

(7) Nejati, K.; Dadashpour, M.; Gharibi, T.; Mellatyar, H.; Akbarzadeh, A. Biomedical Applications of Functionalized Gold Nanoparticles: A Review. *J. Clust. Sci.* **2022**, *33* (1), 1–16. https://doi.org/10.1007/s10876-020-01955-9.

(8) Giljohann, D. A.; Seferos, D. S.; Daniel, W. L.; Massich, M. D.; Patel, P. C.; Mirkin, C. A. Gold Nanoparticles for Biology and Medicine. *Angew. Chem. Int. Ed Engl.* **2010**, *49* (19), 3280–3294. https://doi.org/10.1002/anie.200904359.
(9) Amina, S. J.; Guo, B. A Review on the Synthesis and Functionalization of Gold Nanoparticles as a Drug Delivery Vehicle. *Int. J. Nanomedicine* **2020**, *15*, 9823–9857. https://doi.org/10.2147/IJN.S279094.
(10) Akselrod, G. M.; Huang, J.; Hoang, T. B.; Bowen, P. T.; Su, L.; Smith, D. R.; Mikkelsen, M. H. Large-Area Metasurface Perfect Absorbers from Visible to Near-Infrared. *Adv. Mater.* **2015**, *27* (48), 8028–8034. https://doi.org/10.1002/adma.201503281.
(11) Kim, N. H.; Lee, S. J.; Moskovits, M. Reversible Tuning of SERS Hot Spots with Aptamers. *Adv. Mater.* **2011**, *23* (36), 4152–4156. https://doi.org/10.1002/adma.201101847.
(12) Zon, V.; Sachsenhauser, M.; Rant, U. Preparation of Gold Nanoparticle Dimers via Streptavidin-Induced Interlinking. *J. Nanoparticle Res.* **2013**, *15*. https://doi.org/10.1007/s11051-013-1974-x.
(13) Busch, R. T.; Karim, F.; Weis, J.; Sun, Y.; Zhao, C.; Vasquez, E. S. Optimization and Structural Stability of Gold Nanoparticle–Antibody Bioconjugates. *ACS Omega* **2019**, *4* (12), 15269–15279. https://doi.org/10.1021/acsomega.9b02276.
(14) Kuzyk, A.; Jungmann, R.; Acuna, G. P.; Liu, N. DNA Origami Route for Nanophotonics. *ACS Photonics* **2018**, *5* (4), 1151–1163. https://doi.org/10.1021/acsphotonics.7b01580.
(15) Ofir, Y.; Samanta, B.; M. Rotello, V. Polymer and Biopolymer Mediated Self-Assembly of Gold Nanoparticles. *Chem. Soc. Rev.* **2008**, *37* (9), 1814–1825. https://doi.org/10.1039/B712689C.
(16) Shenton, W.; Davis, S. A.; Mann, S. Directed Self-Assembly of Nanoparticles into Macroscopic Materials Using Antibody–Antigen Recognition. *Adv. Mater.* **1999**, *11* (6), 449–452. https://doi.org/10.1002/(SICI)1521-4095(199904)11:6<449::AID-ADMA449>3.0.CO;2-A.
(17) Gong, C.; Deng, W.; Zou, B.; Xing, Y.; Zhang, X.; Zhang, X.; Jie, J. Large-Scale Assembly of Organic Micro/Nanocrystals into Highly Ordered Patterns and Their Applications for Strain Sensors. *ACS Appl. Mater. Interfaces* **2014**, *6* (14), 11018–11024. https://doi.org/10.1021/am502060y.
(18) Vogel, N.; Retsch, M.; Fustin, C.-A.; del Campo, A.; Jonas, U. Advances in Colloidal Assembly: The Design of Structure and Hierarchy in Two and Three Dimensions. *Chem. Rev.* **2015**, *115* (13), 6265–6311. https://doi.org/10.1021/cr400081d.
(19) Zhang, Y.; Min, C.; Dou, X.; Wang, X.; Urbach, H. P.; Somekh, M. G.; Yuan, X. Plasmonic Tweezers: For Nanoscale Optical Trapping and Beyond. *Light Sci. Appl.* **2021**, *10* (1), 59. https://doi.org/10.1038/s41377-021-00474-0.
(20) Wang, M.; He, L.; Yin, Y. Magnetic Field Guided Colloidal Assembly. *Mater. Today* **2013**, *16* (4), 110–116. https://doi.org/10.1016/j.mattod.2013.04.008.
(21) Adams, S. M.; Campione, S.; Capolino, F.; Ragan, R. Directing Cluster Formation of Au Nanoparticles from Colloidal Solution. *Langmuir* **2013**, *29* (13), 4242–4251. https://doi.org/10.1021/la3051719.
(22) Grzybowski, B. A.; Wilmer, C. E.; Kim, J.; Browne, K. P.; Bishop, K. J. M. Self-Assembly: From Crystals to Cells. *Soft Matter* **2009**, *5* (6), 1110–1128. https://doi.org/10.1039/B819321P.

(23) Selmani, S.; Schwartz, E.; Mulvey, J. T.; Wei, H.; Grosvirt-Dramen, A.; Gibson, W.; Hochbaum, A. I.; Patterson, J. P.; Ragan, R.; Guan, Z. Electrically Fueled Active Supramolecular Materials. *J. Am. Chem. Soc.* **2022**, *144* (17), 7844–7851. https://doi.org/10.1021/jacs.2c01884.
(24) Young, K. L.; Ross, M. B.; Blaber, M. G.; Rycenga, M.; Jones, M. R.; Zhang, C.; Senesi, A. J.; Lee, B.; Schatz, G. C.; Mirkin, C. A. Using DNA to Design Plasmonic Metamaterials with Tunable Optical Properties. *Adv. Mater.* **2014**, *26* (4), 653–659. https://doi.org/10.1002/adma.201302938.
(25) Maestas, J. R.; Ma, F.; Wu, N.; Wu, D. T. Electric-Field-Driven Assembly of Dipolar Spheres Asymmetrically Confined between Two Electrodes. *ACS Nano* **2021**, *15* (2), 2399–2412. https://doi.org/10.1021/acsnano.0c04939.
(26) Nan, F.; Han, F.; Scherer, N. F.; Yan, Z. Dissipative Self-Assembly of Anisotropic Nanoparticle Chains with Combined Electrodynamic and Electrostatic Interactions. *Adv. Mater.* **2018**, *30* (45), 1803238. https://doi.org/10.1002/adma.201803238.
(27) Grzelczak, M.; Liz-Marzán, L. M.; Klajn, R. Stimuli-Responsive Self-Assembly of Nanoparticles. *Chem. Soc. Rev.* **2019**, *48* (5), 1342–1361. https://doi.org/10.1039/C8CS00787J.
(28) Harraq, A. A.; Choudhury, B. D.; Bharti, B. Field-Induced Assembly and Propulsion of Colloids. *Langmuir* **2022**, *38* (10), 3001–3016. https://doi.org/10.1021/acs.langmuir.1c02581.
(29) Balasubramaniam, S.; Pothayee, N.; Lin, Y.; House, M.; Woodward, R. C.; St. Pierre, T. G.; Davis, R. M.; Riffle, J. S. Poly(N-Isopropylacrylamide)-Coated Superparamagnetic Iron Oxide Nanoparticles: Relaxometric and Fluorescence Behavior Correlate to Temperature-Dependent Aggregation. *Chem. Mater.* **2011**, *23* (14), 3348–3356. https://doi.org/10.1021/cm2009048.
(30) Hamner, K. L.; Maye, M. M. Thermal Aggregation Properties of Nanoparticles Modified with Temperature Sensitive Copolymers. *Langmuir* **2013**, *29* (49), 15217–15223. https://doi.org/10.1021/la4037887.
(31) Mehraeen, S. *Self-Assembly of Nanostructures and Patchy Nanoparticles*; BoD – Books on Demand, 2020.
(32) Kundu, P. K.; Das, S.; Ahrens, J.; Klajn, R. Controlling the Lifetimes of Dynamic Nanoparticle Aggregates by Spiropyran Functionalization. *Nanoscale* **2016**, *8* (46), 19280–19286. https://doi.org/10.1039/C6NR05959G.
(33) Wang, D.; Kowalczyk, B.; Lagzi, I.; Grzybowski, B. A. Bistability and Hysteresis During Aggregation of Charged Nanoparticles. *J. Phys. Chem. Lett.* **2010**, *1* (9), 1459–1462. https://doi.org/10.1021/jz100406w.
(34) Grötsch, R. K.; Wanzke, C.; Speckbacher, M.; Angı, A.; Rieger, B.; Boekhoven, J. Pathway Dependence in the Fuel-Driven Dissipative Self-Assembly of Nanoparticles. *J. Am. Chem. Soc.* **2019**, *141* (25), 9872–9878. https://doi.org/10.1021/jacs.9b02004.
(35) Ma, Y.; Zagar, C.; Klemme, D. J.; Sikdar, D.; Velleman, L.; Montelongo, Y.; Oh, S.-H.; Kucernak, A. R.; Edel, J. B.; Kornyshev, A. A. A Tunable Nanoplasmonic Mirror at an Electrochemical Interface. *ACS Photonics* **2018**, *5* (11), 4604–4616. https://doi.org/10.1021/acsphotonics.8b01105.
(36) Ma, Y.; Sikdar, D.; Fedosyuk, A.; Velleman, L.; Klemme, D. J.; Oh, S.-H.; Kucernak, A. R. J.; Kornyshev, A. A.; Edel, J. B. Electrotunable Nanoplasmonics for Amplified Surface

Enhanced Raman Spectroscopy. *ACS Nano* **2020**, *14* (1), 328–336. https://doi.org/10.1021/acsnano.9b05257.
(37) Gong, C.; Deng, W.; Zou, B.; Xing, Y.; Zhang, X.; Zhang, X.; Jie, J. Large-Scale Assembly of Organic Micro/Nanocrystals into Highly Ordered Patterns and Their Applications for Strain Sensors. *ACS Appl. Mater. Interfaces* **2014**, *6* (14), 11018–11024. https://doi.org/10.1021/am502060y.
(38) Jung, L. S.; Campbell, C. T.; Chinowsky, T. M.; Mar, M. N.; Yee, S. S. Quantitative Interpretation of the Response of Surface Plasmon Resonance Sensors to Adsorbed Films. *Langmuir* **1998**, *14* (19), 5636–5648. https://doi.org/10.1021/la971228b.
(39) Kooyman, R. P. H. Physics of Surface Plasmon Resonance.
(40) Ding, S.-Y.; Yi, J.; Li, J.-F.; Ren, B.; Wu, D.-Y.; Panneerselvam, R.; Tian, Z.-Q. Nanostructure-Based Plasmon-Enhanced Raman Spectroscopy for Surface Analysis of Materials. *Nat. Rev. Mater.* **2016**, *1* (6), 1–16. https://doi.org/10.1038/natrevmats.2016.21.
(41) Stiles, P.; Dieringer, J.; Shah, N.; Duyne, R. Surface-Enhanced Raman Spectroscopy. *Annu. Rev. Anal. Chem. Palo Alto Calif* **2008**, *1*, 601–626. https://doi.org/10.1146/annurev.anchem.1.031207.112814.
(42) del Pino, J.; Feist, J.; Garcia-Vidal, F. J. Signatures of Vibrational Strong Coupling in Raman Scattering. *J. Phys. Chem. C* **2015**, *119* (52), 29132–29137. https://doi.org/10.1021/acs.jpcc.5b11654.
(43) Fleischmann, M.; Hendra, P. J.; McQuillan, A. J. Raman Spectra of Pyridine Adsorbed at a Silver Electrode. *Chem. Phys. Lett.* **1974**, *26* (2), 163–166. https://doi.org/10.1016/0009-2614(74)85388-1.
(44) Jeanmaire, D. L.; Van Duyne, R. P. Surface Raman Spectroelectrochemistry: Part I. Heterocyclic, Aromatic, and Aliphatic Amines Adsorbed on the Anodized Silver Electrode. *J. Electroanal. Chem. Interfacial Electrochem.* **1977**, *84* (1), 1–20. https://doi.org/10.1016/S0022-0728(77)80224-6.
(45) Persson, B. N. J.; Zhao, K.; Zhang, Z. Chemical Contribution to Surface-Enhanced Raman Scattering. *Phys. Rev. Lett.* **2006**, *96* (20), 207401. https://doi.org/10.1103/PhysRevLett.96.207401.
(46) Langer, J.; Jimenez de Aberasturi, D.; Aizpurua, J.; Alvarez-Puebla, R. A.; Auguié, B.; Baumberg, J. J.; Bazan, G. C.; Bell, S. E. J.; Boisen, A.; Brolo, A. G.; Choo, J.; Cialla-May, D.; Deckert, V.; Fabris, L.; Faulds, K.; García de Abajo, F. J.; Goodacre, R.; Graham, D.; Haes, A. J.; Haynes, C. L.; Huck, C.; Itoh, T.; Käll, M.; Kneipp, J.; Kotov, N. A.; Kuang, H.; Le Ru, E. C.; Lee, H. K.; Li, J.-F.; Ling, X. Y.; Maier, S. A.; Mayerhöfer, T.; Moskovits, M.; Murakoshi, K.; Nam, J.-M.; Nie, S.; Ozaki, Y.; Pastoriza-Santos, I.; Perez-Juste, J.; Popp, J.; Pucci, A.; Reich, S.; Ren, B.; Schatz, G. C.; Shegai, T.; Schlücker, S.; Tay, L.-L.; Thomas, K. G.; Tian, Z.-Q.; Van Duyne, R. P.; Vo-Dinh, T.; Wang, Y.; Willets, K. A.; Xu, C.; Xu, H.; Xu, Y.; Yamamoto, Y. S.; Zhao, B.; Liz-Marzán, L. M. Present and Future of Surface-Enhanced Raman Scattering. *ACS Nano* **2020**, *14* (1), 28–117. https://doi.org/10.1021/acsnano.9b04224.
(47) Hering, K.; Cialla, D.; Ackermann, K.; Dörfer, T.; Möller, R.; Schneidewind, H.; Mattheis, R.; Fritzsche, W.; Rösch, P.; Popp, J. SERS: A Versatile Tool in Chemical and Biochemical Diagnostics. *Anal. Bioanal. Chem.* **2008**, *390* (1), 113–124. https://doi.org/10.1007/s00216-007-1667-3.

(48) Zhou, Y.; Liu, J.; Zheng, T.; Tian, Y. Label-Free SERS Strategy for In Situ Monitoring and Real-Time Imaging of Aβ Aggregation Process in Live Neurons and Brain Tissues. *Anal. Chem.* **2020**, *92* (8), 5910–5920. https://doi.org/10.1021/acs.analchem.9b05837.
(49) Wang, Y.; Rauf, S.; Grewal, Y. S.; Spadafora, L. J.; Shiddiky, M. J. A.; Cangelosi, G. A.; Schlücker, S.; Trau, M. Duplex Microfluidic SERS Detection of Pathogen Antigens with Nanoyeast Single-Chain Variable Fragments. *Anal. Chem.* **2014**, *86* (19), 9930–9938. https://doi.org/10.1021/ac5027012.
(50) Le Ru, E. C.; Etchegoin, P. G. Single-Molecule Surface-Enhanced Raman Spectroscopy. *Annu. Rev. Phys. Chem.* **2012**, *63* (1), 65–87. https://doi.org/10.1146/annurev-physchem-032511-143757.
(51) Thrift, W. J.; Ronaghi, S.; Samad, M.; Wei, H.; Nguyen, D. G.; Cabuslay, A. S.; Groome, C. E.; Santiago, P. J.; Baldi, P.; Hochbaum, A. I.; Ragan, R. Deep Learning Analysis of Vibrational Spectra of Bacterial Lysate for Rapid Antimicrobial Susceptibility Testing. *ACS Nano* **2020**, *14* (11), 15336–15348. https://doi.org/10.1021/acsnano.0c05693.
(52) Le Ru, E. C.; Meyer, M.; Etchegoin, P. G. Proof of Single-Molecule Sensitivity in Surface Enhanced Raman Scattering (SERS) by Means of a Two-Analyte Technique. *J. Phys. Chem. B* **2006**, *110* (4), 1944–1948. https://doi.org/10.1021/jp054732v.
(53) Alharbi, O.; Xu, Y.; Goodacre, R. Simultaneous Multiplexed Quantification of Nicotine and Its Metabolites Using Surface Enhanced Raman Scattering. *Analyst* **2014**, *139* (19), 4820–4827. https://doi.org/10.1039/C4AN00879K.
(54) Shi, H.; Wang, H.; Meng, X.; Chen, R.; Zhang, Y.; Su, Y.; He, Y. Setting Up a Surface-Enhanced Raman Scattering Database for Artificial-Intelligence-Based Label-Free Discrimination of Tumor Suppressor Genes. *Anal. Chem.* **2018**, *90* (24), 14216–14221. https://doi.org/10.1021/acs.analchem.8b03080.
(55) Li, J.; Skeete, Z.; Shan, S.; Yan, S.; Kurzatkowska, K.; Zhao, W.; Ngo, Q. M.; Holubovska, P.; Luo, J.; Hepel, M.; Zhong, C.-J. Surface Enhanced Raman Scattering Detection of Cancer Biomarkers with Bifunctional Nanocomposite Probes. *Anal. Chem.* **2015**, *87* (21), 10698–10702. https://doi.org/10.1021/acs.analchem.5b03456.
(56) Lussier, F.; Missirlis, D.; Spatz, J. P.; Masson, J.-F. Machine-Learning-Driven Surface-Enhanced Raman Scattering Optophysiology Reveals Multiplexed Metabolite Gradients Near Cells. *ACS Nano* **2019**, *13* (2), 1403–1411. https://doi.org/10.1021/acsnano.8b07024.
(57) Dong, R.; Weng, S.; Yang, L.; Liu, J. Detection and Direct Readout of Drugs in Human Urine Using Dynamic Surface-Enhanced Raman Spectroscopy and Support Vector Machines. *Anal. Chem.* **2015**, *87* (5), 2937–2944. https://doi.org/10.1021/acs.analchem.5b00137.
(58) Li, X.; Yang, T.; Li, S.; Jin, L.; Wang, D.; Guan, D.; Ding, J. Noninvasive Liver Diseases Detection Based on Serum Surface Enhanced Raman Spectroscopy and Statistical Analysis. *Opt. Express* **2015**, *23* (14), 18361–18372. https://doi.org/10.1364/OE.23.018361.
(59) Cullen, W. R.; Reimer, K. J. Arsenic Speciation in the Environment. 52.
(60) Prasad, S.; Yadav, K. K.; Kumar, S.; Gupta, N.; Cabral-Pinto, M. M. S.; Rezania, S.; Radwan, N.; Alam, J. Chromium Contamination and Effect on Environmental Health and Its Remediation: A Sustainable Approaches. *J. Environ. Manage.* **2021**, *285*, 112174. https://doi.org/10.1016/j.jenvman.2021.112174.

Chapter 2

Background

2.1 Surface enhanced Raman scattering

Raman scattering is a phenomenon that involves the inelastic scattering of photons when they interact with matter, leading to the generation of two types of scattered photons: Stokes scattering and Anti-Stokes scattering. In Stokes scattering, photons lose energy when molecules are promoted from the ground state to the excited vibrational state and subsequently relax back to an energy state higher than their original state. In contrast, Anti-Stokes scattering involves the opposite process where photons gain energy from molecular phonon modes. Consequently, inelastically scattered photons contain information about the vibrational modes of molecules. However, Raman signal is usually low due to a small scattering cross section, which is quantified as the area of the incident beam over which incident photons are effectively converted into emitted Raman photons (10^{-11} – 10^{-15} nm^2).[1] Surface enhanced Raman scattering (SERS) provides highly sensitive vibrational fingerprinting information of analytes through the enhancement of electromagnetic fields generated by the excitation of localized surface plasmon resonance.[2] The chemical mechanism of SERS involves the transfer of electrons between adsorbed molecules and NPs in direct contact. Electrons from the Fermi level of the plasmonic NP transfer to the lowest unoccupied molecular orbital of the molecule, results in the formation of charge transfer intermediates with larger Raman cross section than that of the free molecule.[3] Section 2.2.1 discusses the electromagnetic enhancement mechanism in detail, as the enhancement resulting from this mechanism can be up to 10^4 – 10^6 times stronger than the chemical mechanism. Sections 2.1.2 and 2.1.3 focus on the influence of the gaps between NPs on

SERS enhancement. The nanogap spacing and the 2D arrangement of NPs are crucial factors that affect SERS enhancement.

2.1.1 Electromagnetic enhancement

The collective oscillations of conduction electrons in NP when excited optically at the material's plasmon resonance frequency results in surface plasmons, which exhibit an enhanced local electromagnetic field surrounding the plasmonic NP. The excitation of surface plasmons induces a strong spatial localization and hence amplification of the laser light in small spatial regions, called hotspots.[4] Therefore, the electromagnetic field experienced by the molecules residing in hotspots is much stronger than the field they would experience without the plasmonic NP, which is known as the local field enhancement. The presence of the plasmonic NP nearby the molecule can significantly enhance the Raman signal.

$$G = G_1(\omega_0)G_2(\omega_R) = \frac{|E_{loc}(\omega_0)|^2|E_{loc}(\omega_R)|^2}{|E_0(\omega_0)|^2|E_0(\omega_R)|^2} \approx \frac{|E_{loc}(\omega_0)|^4}{|E_0(\omega_0)|^4} \qquad \text{Equation 2.1}^4$$

Equation 2.1 elucidate a two step enhancement process of the electromagnetic mechanism. $G_1(\omega_0)$ and $G_2(\omega_R)$ represents the enhancement factor of the local electromagnetic field occurs at the metallic-dielectric interface at the incident laser frequency ω_0 and the Raman polarizability at resonance frequency ω_R, respectively. For the first step, when the LSPR is excited inelastic light scattering of the incident electric field $E_0(\omega_0)$ on the plasmonic NP leads to an increased local electric field $E_{loc}(\omega_0)$ in the vicinity of the plasmonic NP. For the second step, plasmonic NPs serve as optical antennae to transfer Raman signal from the near field to the

far field. Raman signal from molecules in hotspots NP can be enhanced when experiencing the local electric field $E_{loc}(\omega_R)$. When the wavelengths of the incident laser and Stokes scattered signal are close to each other, the SERS enhancement factor is approximately proportional to the fourth power of the enhancement of the local electric field.

2.1.2 SERS in nanoclusters

It is well known that the electromagnetic mechanism and resulting SERS enhancement of hotspots strongly depend on its size,[5] making it a critical aspect of SERS-based sensing platforms. For instance, reducing the gap size of an Au NP dimer from 10 to 2 nm increases the SERS enhancement factor from 10^5 to 10^9.[5] This is because the intensity of the electromagnetic field within the gap is inversely proportional to the distance between the NPs. Therefore, smaller nanogap spacing results in a stronger electromagnetic field and higher SERS enhancement. However, decreasing the gap size to the sub-nanometer scale introduces electron tunneling effects between the coupled NPs,[6] limiting the ability to further increase the local electromagnetic field and the SERS enhancement factor.

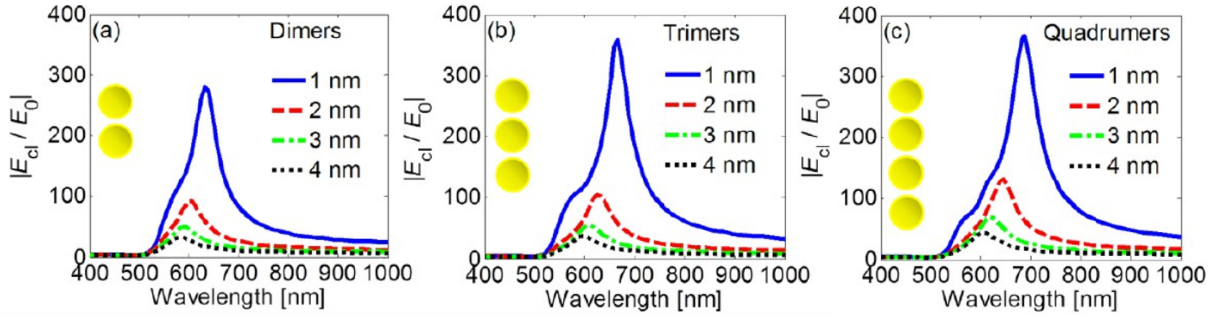

Figure 2.1 Theoretical electric field enhancement $|E_{cl}/E_0|$ computed in dimers (left), linear trimers (middle), and linear quadrumers (right) versus illumination wavelength. E_0 is the plane wave field without clusters and E_{cl} is the field with clusters.[7]

Finite element method simulations of electromagnetic coupling can provide guidance for advanced nanostructure fabrication to tune optical properties.[8–10] By using electromagnetic full-wave simulations, the Ragan Lab investigates the effect of changing interparticle gap spacing on SERS enhancement. Figure 2.1 shows calculated electric field enhancement as a function of gap spacings between NPs ranging from 1 - 4 nm in dimers, linear trimers and linear quadrumers. The result shows that a slight increase in the gap size can significantly decrease the local electric field enhancement. When the gap size is 1 nm for the linear quadrumer , the enhancement is approximately 365, 130 when the gap is 2 nm, and less than 45 when the gap is 4 nm. Also there is a resonance blue shift for increasing gap size for all three cases. Calculation of the local electric field enhancement can estimate the theoretical SERS enhancement by computing as the fourth power of the field enhancement, which would range from about 10^{10} for the linear quadrumer case with 1 nm gap to about 3 x 10^{8} when the gap is 2 nm and less than 4 x 10^{6} when the gap is 4 nm.

2.1.3 SERS geometry dependence

The SERS activity observed in metal NP clusters is often modeled as a dipole-dipole interaction between two neighboring NPs.[5] Clusters with more NPs in variant arrangements should also be considered to investigate the interaction of the LSPR between NPs. Understanding how SERS geometry governing field enhancement of electromagnetic fields confined in hotspots is a topic of great significance to designing SERS sensing platforms.[11] The Ragan Lab correlated UV-vis spectroscopy data from optical response of a surface with a variety of geometries (blue dotted line), and overlaid absorption spectra simulated with the full-wave finite element to exam how different geometries contribute to the measured attenuation spectrum. Four distinct peaks are observed. Simulations identify these peaks to be associated with monomers at 536 nm,

dimers at 686 nm (black dotted line), trimers at 740 nm (orange dotted line) and quadrumers at 782 nm (green dotted line) with 0.9 nm gap spacing. These results indicate that the optical response is dominated by clusters with 0.9 nm gap spacing. Consider that the absorption maximum of a 2 nm gap dimer geometry, calculated in simulations and plotted with a dashed line in Figure 2.2, the dimer adsorption peak is shifted by approximately 50 nm for the 0.9 nm dimer case. From this analysis, it is clear that the influence of gap spacing is greater than the cluster size when one considers both lead to spectra shifts but the SERS signal may vary by orders of magnitude when the gap spacing changes on the order of a nanometer.[12]

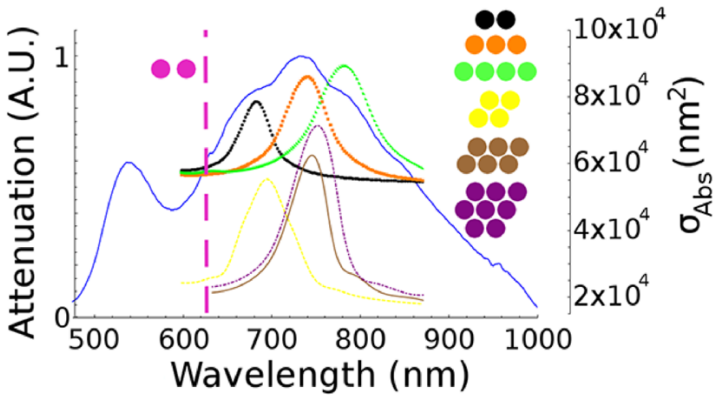

Figure 2.2 Normalized attenuation spectra. The spectrum is overlaid with the calculated absorption cross section of observed oligomers with various geometries with a color-coded schematic.[12]

2.2 Electrohydrodynamic flow

Electrohydrodynamic (EHD) flow is the motion of electrically charged fluids resulting from the application of electric fields.[13] One of the applications of EHD is the generation of flow around charged NPs near an electrode. The earliest modern report of electrically induced 2D crystallization of colloidal particles dates back to 1984, when Richetti *et al.* applied an oscillatory electric field between two parallel-plate electrodes and observed the formation of

hexagonal arrays of micro-sized particles on the electrode surface.[14] When the electric field was removed, the arrays disassembled due to Brownian motion. In 1997, Trau *et al.* attributed the ordered monolayer particle assembly to the electrodynamic (EHD) flow and proposed an EHD mechanism in which the distortions of the electric field alter the body force distribution in the electrode's charge polarization layer.[15] The action of the applied field on this charge produces flow, and the flow scales with the square of the field strength since the induced charge and the electrical body force are each proportional to the applied voltage.[16] This section focuses on studying EHD flow under oscillatory electric field.

2.2.1 Origin of EHD flow

AC-induced charge electroosmosis is a phenomenon that occurs when an oscillatory electric field is applied to a charged surface, causing fluid flow to be generated. This flow arises from the movement of charged ions in response to the electric field.

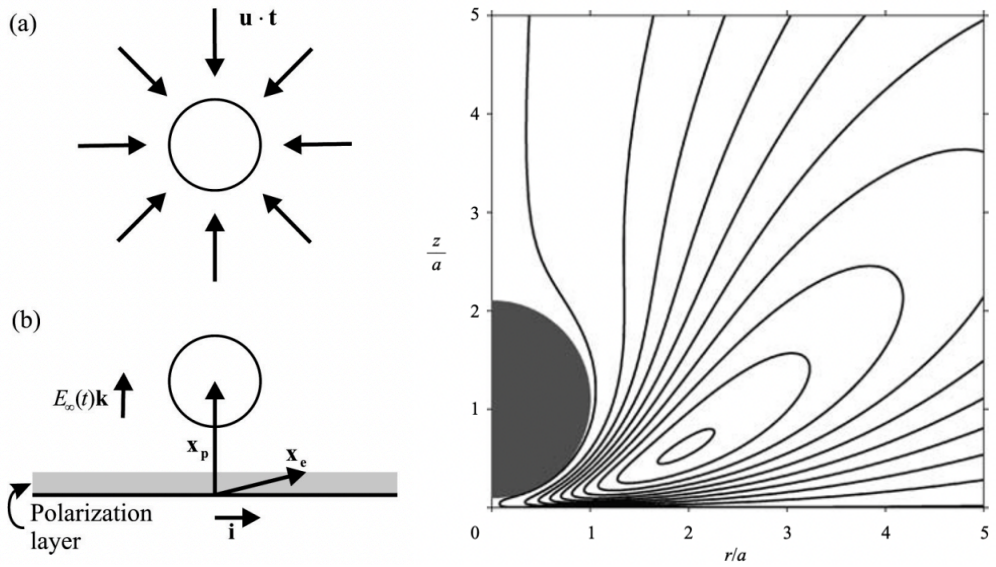

Figure 2.3 Schematic depicting a spherical particle near an electrode (left).[16] (a) represents the plan view depicting flow towards the test particle. (b) An elevation view showing a particle

located at X_p outside the polarization layer, and X_e is a location in the polarization layer near the electrode. Streamline representing EHD flow fields (right).[17]

The equations that govern AC-induced charge electroosmosis are the Navier-Stokes equations for fluid motion (Equation 2.2)[18] and the Poisson equation for electric potential (Equation 2.3).[19] The Navier-Stokes equations describe the conservation of momentum for a fluid. $\rho \frac{D\vec{v}}{Dt}$ is the force on each fluid particle. $-\nabla p$ represents the volumetric stress tensor, which prevents motion due to normal stresses; $\nabla \cdot T$ known as the stress deviator tensor, causing motion due to horizontal friction and shear stresses; $\vec{f}$ is the force term acting on every single fluid particle. The Poisson equation describes the relationship between the electric potential and the charge distribution in the system, where ρ is the electric charge, Φ is electric potential and ϵ_0 is permittivity.

$$\rho \frac{D\vec{v}}{Dt} = -\nabla p + \nabla \cdot T + \vec{f} \qquad \text{Equation 2.2}$$

$$\nabla^2 \Phi = -\rho/\epsilon_0 \qquad \text{Equation 2.3}$$

In addition to these fundamental equations governing fluid motion, there exist additional equations that are essential in characterizing the electrical properties of charged surfaces, including the surface charge density and surface conductivity. These equations are typically derived from models of the electrical double layer, which forms at the interface between the charged surface and the fluid. The specific equations used to describe AC-induced charge electroosmosis are dependent on the specific system being investigated, taking into consideration factors such as the geometry of the charged surface and the properties of the fluid. Nevertheless,

these equations are useful for predicting the behavior of the system and optimizing experimental parameters to achieve desired fluid flow patterns in Figure 2.3.

2.2.2 EHD flow frequency dependence

EHD flow can exist only at specific frequency range, which refers to the characteristic frequency of the EHD flow. At frequencies higher than the characteristic frequency, ions in the electrolyte do not have time to respond fast enough to establish the induced electronic double layer. In this way, the EHD flow cannot be generated.[20] At frequencies lower than the characteristic frequency, the induced electronic double layer screens the powered electrode and with it also the ability of the electric field to penetrate the microfluidic chamber, the ions in the suspension have enough time to form the electrical double layer. In this situation, the applied electric field drops almost entirely across the electrical double layer, resulting in a near-zero electric field in the suspension and again ceasing EHD flow.[20]

2.3 Machine learning

2.3.1 The development of machine learning

Machine learning (ML) has undergone significant advancements in the past two decades, evolving into a practical technology with extensive commercial applications.[21] ML can be categorized into three types, namely supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the training data comprises a set of (x,y) pairs, and the objective is to generate a prediction y' for a given query x' through a learned mapping $f(x)$. The input data x can take diverse forms, such as images, protein sequences, DNA sequences, protein 3D structure, molecular structure, mass spectrometry, and Raman spectrometry.[22] The output generated can either be a single value y for each input x or a

probability distribution over y given x. The mapping function f can take various forms, such as decision trees, logistic regression, support vector machines, and neural networks.

In recent years, supervised learning has witnessed a significant advancement in the field of deep learning, which utilizes multilayer networks of threshold units, with each unit computing a parameterized activation function of its inputs. These networks are commonly referred to as deep networks. To adjust the parameters of the network based on errors at the output, gradient-based optimization algorithms are employed. The internal layers of deep networks can be seen as providing learned representations of the input data, which facilitates the extraction of features that are essential in solving complex problems.

Unsupervised learning is a type of machine learning that does not require labeled training data. Dimension reduction methods, such as principal component analysis, factor analysis, and autoencoders, have been developed to reduce the dimensionality of the input data. Clustering is another unsupervised learning technique that involves finding a partition of the observed data and predicting future data, without explicit labels indicating a desired partition.

Unlike supervised learning, where training examples indicate the correct output for a given input, or unsupervised learning, where labeled data is not available, reinforcement learning relies on training data that only provides an indication of whether an action is correct or not. In reinforcement learning, an agent learns to interact with its environment by taking actions and receiving rewards or penalties based on its performance. Through trial-and-error, the agent seeks to maximize its cumulative rewards by learning a policy that maps states to actions.

2.3.2 Machine learning in SERS

SERS is a powerful analytical technique for identifying and characterizing materials based on their molecular vibrations. However, the interpretation of SERS spectra can be challenging, especially for chemically similar compounds that exhibit subtle spectral differences and mixtures with complex backgrounds.[23] The integration of machine learning algorithms with SERS can overcome this challenge and improve the accuracy of chemical identification and trace chemical detection.

Material identification is one of the most common applications of SERS spectroscopy. For example, machine learning algorithms can be utilized to classify Raman spectra of bacteria strains[24] and different types of cancer cells.[25] Besides material identification, Raman spectroscopy is also suitable for quantitative analysis such as the characterization of metabolisms.[26,27] In addition, the application of machine learning to SERS spectra of complex organic molecules has assisted researchers to gain a better understanding of their structural characteristics and potential reactivity.[28] Machine learning algorithms can predict the optoelectronic properties of plasmonic nanostructures, with the aim of guiding the design and modeling of the next generation of catalysts, sensors, and photothermal devices.[29]

References

(1) Langer, J.; Jimenez de Aberasturi, D.; Aizpurua, J.; Alvarez-Puebla, R. A.; Auguié, B.; Baumberg, J. J.; Bazan, G. C.; Bell, S. E. J.; Boisen, A.; Brolo, A. G.; Choo, J.; Cialla-May, D.; Deckert, V.; Fabris, L.; Faulds, K.; García de Abajo, F. J.; Goodacre, R.; Graham, D.; Haes, A. J.; Haynes, C. L.; Huck, C.; Itoh, T.; Käll, M.; Kneipp, J.; Kotov, N. A.; Kuang, H.; Le Ru, E. C.; Lee, H. K.; Li, J.-F.; Ling, X. Y.; Maier, S. A.; Mayerhöfer, T.; Moskovits, M.; Murakoshi, K.; Nam, J.-M.; Nie, S.; Ozaki, Y.; Pastoriza-Santos, I.; Perez-Juste, J.; Popp, J.; Pucci, A.; Reich, S.; Ren, B.; Schatz, G. C.; Shegai, T.; Schlücker, S.; Tay, L.-L.; Thomas, K. G.; Tian, Z.-Q.; Van Duyne, R. P.; Vo-Dinh, T.; Wang, Y.; Willets, K. A.; Xu, C.; Xu, H.; Xu, Y.; Yamamoto, Y. S.; Zhao, B.; Liz-Marzán, L. M. Present and Future of Surface-Enhanced Raman Scattering. *ACS Nano* **2020**, *14* (1), 28–117. https://doi.org/10.1021/acsnano.9b04224.

(2) Sharma, B.; Frontiera, R. R.; Henry, A.-I.; Ringe, E.; Van Duyne, R. P. SERS: Materials, Applications, and the Future. *Mater. Today* **2012**, *15* (1), 16–25. https://doi.org/10.1016/S1369-7021(12)70017-2.
(3) Schlücker, S. Surface-Enhanced Raman Spectroscopy: Concepts and Chemical Applications. *Angew. Chem. Int. Ed.* **2014**, *53* (19), 4756–4795. https://doi.org/10.1002/anie.201205748.
(4) Esteban, R.; Borisov, A. G.; Nordlander, P.; Aizpurua, J. Bridging Quantum and Classical Plasmonics with a Quantum-Corrected Model. *Nat. Commun.* **2012**, *3* (1), 825. https://doi.org/10.1038/ncomms1806.
(5) McMahon, J. M.; Li, S.; Ausman, L. K.; Schatz, G. C. Modeling the Effect of Small Gaps in Surface-Enhanced Raman Spectroscopy. *J. Phys. Chem. C* **2012**, *116* (2), 1627–1637. https://doi.org/10.1021/jp207661y.
(6) Itoh, T.; Procházka, M.; Dong, Z.-C.; Ji, W.; Yamamoto, Y. S.; Zhang, Y.; Ozaki, Y. Toward a New Era of SERS and TERS at the Nanometer Scale: From Fundamentals to Innovative Applications. *Chem. Rev.* **2023**, *123* (4), 1552–1634. https://doi.org/10.1021/acs.chemrev.2c00316.
(7) Adams, S. M.; Campione, S.; Capolino, F.; Ragan, R. Directing Cluster Formation of Au Nanoparticles from Colloidal Solution. *Langmuir* **2013**, *29* (13), 4242–4251. https://doi.org/10.1021/la3051719.
(8) Myroshnychenko, V.; Rodríguez-Fernández, J.; Pastoriza-Santos, I.; Funston, A. M.; Novo, C.; Mulvaney, P.; Liz-Marzán, L. M.; Abajo, F. J. G. de. Modelling the Optical Response of Gold Nanoparticles. *Chem. Soc. Rev.* **2008**, *37* (9), 1792–1805. https://doi.org/10.1039/B711486A.
(9) García-Martín, A.; Ward, D. R.; Natelson, D.; Cuevas, J. C. Field Enhancement in Subnanometer Metallic Gaps. *Phys. Rev. B* **2011**, *83* (19), 193404. https://doi.org/10.1103/PhysRevB.83.193404.
(10) Vallecchi, A.; Albani, M.; Capolino, F. Collective Electric and Magnetic Plasmonic Resonances in Spherical Nanoclusters. *Opt. Express* **2011**, *19* (3), 2754. https://doi.org/10.1364/OE.19.002754.
(11) Yap, F. L.; Thoniyot, P.; Krishnan, S.; Krishnamoorthy, S. Nanoparticle Cluster Arrays for High-Performance SERS through Directed Self-Assembly on Flat Substrates and on Optical Fibers. *ACS Nano* **2012**, *6* (3), 2056–2070. https://doi.org/10.1021/nn203661n.
(12) Thrift, W. J.; Nguyen, C. Q.; Darvishzadeh-Varcheie, M.; Zare, S.; Sharac, N.; Sanderson, R. N.; Dupper, T. J.; Hochbaum, A. I.; Capolino, F.; Abdolhosseini Qomi, M. J.; Ragan, R. Driving Chemical Reactions in Plasmonic Nanogaps with Electrohydrodynamic Flow. *ACS Nano* **2017**, *11* (11), 11317–11329. https://doi.org/10.1021/acsnano.7b05815.
(13) Selvakumar, R. D.; Qiang, L.; Kang, L.; Traoré, P.; Wu, J. Numerical Modeling of Solid-Liquid Phase Change under the Influence an External Electric Field. *Int. J. Multiph. Flow* **2021**, *136*, 103550. https://doi.org/10.1016/j.ijmultiphaseflow.2020.103550.
(14) Richetti, P.; Prost, J.; Barois, P. Two-Dimensional Aggregation and Crystallization of a Colloidal Suspension of Latex Spheres. *J. Phys. Lett.* **1984**, *45* (23), 1137–1143. https://doi.org/10.1051/jphyslet:0198400450230113700.
(15) Trau, M.; Saville, D. A.; Aksay, I. A. Assembly of Colloidal Crystals at Electrode Interfaces. *Langmuir* **1997**, *13* (24), 6375–6381. https://doi.org/10.1021/la970568u.

(16) Ristenpart, W. D.; Aksay, I. A.; Saville, D. A. Assembly of Colloidal Aggregates by Electrohydrodynamic Flow: Kinetic Experiments and Scaling Analysis. *Phys. Rev. E* **2004**, *69* (2), 021405. https://doi.org/10.1103/PhysRevE.69.021405.
(17) Ristenpart, W. D.; Aksay, I. A.; Saville, D. A. Electrohydrodynamic Flow around a Colloidal Particle near an Electrode with an Oscillating Potential. *J. Fluid Mech.* **2007**, *575*, 83–109. https://doi.org/10.1017/S0022112006004368.
(18) Chemin, J.-Y.; Desjardins, B.; Gallagher, I.; Grenier, E. *Mathematical Geophysics: An Introduction to Rotating Fluids and the Navier-Stokes Equations*; Clarendon Press, 2006.
(19) Srebrenik, S.; Weinstein, H.; Pauncz, R. Analytical Calculation of Atomic and Molecular Electrostatic Potentials from the Poisson Equation. *Chem. Phys. Lett.* **1973**, *20* (5), 419–423. https://doi.org/10.1016/0009-2614(73)85188-7.
(20) Bhatt, K. H.; Grego, S.; Velev, O. D. An AC Electrokinetic Technique for Collection and Concentration of Particles and Cells on Patterned Electrodes. *Langmuir* **2005**, *21* (14), 6603–6612. https://doi.org/10.1021/la050658w.
(21) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349* (6245), 255–260. https://doi.org/10.1126/science.aaa8415.
(22) Greener, J. G.; Kandathil, S. M.; Moffat, L.; Jones, D. T. A Guide to Machine Learning for Biologists. *Nat. Rev. Mol. Cell Biol.* **2022**, *23* (1), 40–55. https://doi.org/10.1038/s41580-021-00407-0.
(23) Bajomo, M. M.; Ju, Y.; Zhou, J.; Elefterescu, S.; Farr, C.; Zhao, Y.; Neumann, O.; Nordlander, P.; Patel, A.; Halas, N. J. Computational Chromatography: A Machine Learning Strategy for Demixing Individual Chemical Components in Complex Mixtures. *Proc. Natl. Acad. Sci.* **2022**, *119* (52), e2211406119. https://doi.org/10.1073/pnas.2211406119.
(24) Rahman, A.; Kang, S.; Wang, W.; Huang, Q.; Kim, I.; Vikesland, P. J. Lectin-Modified Bacterial Cellulose Nanocrystals Decorated with Au Nanoparticles for Selective Detection of Bacteria Using Surface-Enhanced Raman Scattering Coupled with Machine Learning. *ACS Appl. Nano Mater.* **2022**, *5* (1), 259–268. https://doi.org/10.1021/acsanm.1c02760.
(25) Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y.; Choi, B. H.; Kang, K.-W.; Jeong, H.; Park, Y.; Hong, S.; Kim, H. K.; Choi, Y. Early-Stage Lung Cancer Diagnosis by Deep Learning-Based Spectroscopic Analysis of Circulating Exosomes. *ACS Nano* **2020**, *14* (5), 5435–5444. https://doi.org/10.1021/acsnano.9b09119.
(26) Thrift, W. J.; Ronaghi, S.; Samad, M.; Wei, H.; Nguyen, D. G.; Cabuslay, A. S.; Groome, C. E.; Santiago, P. J.; Baldi, P.; Hochbaum, A. I.; Ragan, R. Deep Learning Analysis of Vibrational Spectra of Bacterial Lysate for Rapid Antimicrobial Susceptibility Testing. *ACS Nano* **2020**, *14* (11), 15336–15348. https://doi.org/10.1021/acsnano.0c05693.
(27) Yan, S.; Liu, C.; Fang, S.; Ma, J.; Qiu, J.; Xu, D.; Li, L.; Yu, J.; Li, D.; Liu, Q. SERS-Based Lateral Flow Assay Combined with Machine Learning for Highly Sensitive Quantitative Analysis of Escherichia Coli O157:H7. *Anal. Bioanal. Chem.* **2020**, *412* (28), 7881–7890. https://doi.org/10.1007/s00216-020-02921-0.
(28) Wu, Z.; Ramsundar, B.; N. Feinberg, E.; Gomes, J.; Geniesse, C.; S. Pappu, A.; Leswing, K.; Pande, V. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chem. Sci.* **2018**, *9* (2), 513–530. https://doi.org/10.1039/C7SC02664A.
(29) Masson, J.-F.; Biggins, J. S.; Ringe, E. Machine Learning for Nanoplasmonics. *Nat. Nanotechnol.* **2023**, *18* (2), 111–123. https://doi.org/10.1038/s41565-022-01284-0.

Chapter 3

Nanoantennas as Reporters of Dissipative Assembly in Oscillatory Electric Fields

3.1 Introduction

Self-assembly of optically active materials from nanoparticle building blocks in colloidal solution relies on the ability of directed interactions. For example, dynamic assembly of plasmonic nanoparticles (NP) with resonances at optical frequencies has been used to change absorbance and reflectance on surfaces using electrostatic forces.[1,2] Electrotunable systems using functionalized Au NP have used small voltage pulses to tune between window and mirror states.[3] Optical forces on colloidal solutions of Ag NP in combination of electrostatic interactions have initiated dissipative assembly of chains of NP along the polarization axis of an optical beam.[4] Rich 2D phase behavior with varying electric field strength is observed when NP are confined between electrodes separated by distances a few orders of magnitude larger than the nanoparticle's diameter.[5] Thus understanding electric stimuli as a driving forces for dissipative assembly of NP with resonances at optical frequencies can produce reconfigurable metasurfaces.

Electrohydrodynamic (EHD) flow resulting from an applied AC potential, also called AC-induced charge electroosmosis, is an electrokinetic phenomenon that can drive lateral assembly of nanoparticles in response to a perturbation that produces an electric field gradient on an electrode-liquid interface.[6–10] EHD flow is a dissipative driving force;[8,11,12] assemblies may disassemble if the electrical stimuli is not maintained. Dissipative assembly of micron scale particles using EHD flow has been studied extensively but electrical stimuli for dissipative assembly of NP remains relatively unexplored.[5,8,10,13,14] As colloidal NP are typically stabilized

via electrostatic repulsion, chemical crosslinking in solution will normally lead to aggregation. EHD flow is confined to the working electrode surface and thus mitigates uncontrolled aggregation in bulk solution. Though assemblies are transient in principle, NPs may be kinetically trapped on surfaces due to electrostatic interactions, making it difficult to experimentally observe and thereby monitor time scales of disassembly of NP at electrode-liquid interfaces.

Here we investigate the effect of frequency under an applied AC potential in influencing the lateral arrangement of clusters of Au NP formed as a result of EHD flow and time scales for assembly and disassembly. The resultant electromagnetic field enhancements, associated with excitation of the local surface plasmon resonance (LSPR) of NP as they are assembled into clusters, provides a means of tracking assembly in situ using optical methods. Dissipative assembly is monitored using in situ fluorescence microscopy and surface enhanced Raman scattering (SERS). Previous studies have reported that metal enhanced fluorescence is maximum when the emission spectra of the fluorophore is tuned to the LSPR frequency of metal nanostructures.[15] Lateral assembly of Au NP will shift the LSPR from approximately 520 nm toward the emission line of Nile red, 561 nm. We show that increases and decreases in fluorescence intensity are correlated with turning on and off external electric stimuli and thus able to report dissipative assembly of Au NP in real time. Furthermore, SERS spectra measured in situ during EHD assembly of Au NP functionalized with 4-mercaptobenzoic acid (4-MBA) as reporter molecules shows signal saturates as early as 20 s.

3.2 Materials and Methods

3.2.1 Electrode Materials

The working electrode is composed of P-type, boron doped, (100) silicon wafers (University Wafer, USA) with resistivity of 0.001-0.005 ohm-cm. Silicon wafers are cleaned for 5 min by 20% v/v hydrofluoric acid (HF, Fisher Scientific, USA) in deionized (DI) water with resistivity of 18.2 MΩ cm^{-1}, obtained from a Milli-Q Millipore System and then immersed in DI water to regrow a thin silicon dioxide layer. *The potential of HF to cause severe injury mandates extreme caution during usage.* After cleaning, silicon wafers are spin coated with 1 wt% random copolymer Poly(styrene-co-methyl methacrylate)-α-Hydroxyl-ω-Tempo moiety (PS-r-PMMA) (Mn =7400, Mw =11800, Mw /Mn =1.60, Polystyrene content: 59.6 mol%, Polymer Source, Inc., Canada) in toluene (Fisher Scientific, USA) at 3000 rpm for 45 s, annealed under vacuum at 170 °C for 48 hr, and then rinsed with toluene to leave a brush layer. Diblock copolymer poly(styrene-b-methyl methacrylate) (PS-b-PMMA) (Mn S-b-MMA 170000-b-145000 g mol^{-1}) (Polymer Source, Inc., Canada) is spin coated at 5000 rpm for 45 s and then annealed for 72 hr at 170 °C.

In the case of fluorescence imaging, Si wafers are first immersed in 3:1 H_2SO_4 : H_2O_2 (piranha solution) at room temperature for 30 s in order to hydroxylate the surface, and functionalized with amine group by soaking in a solution of 2% 3-(aminopropyl)triethoxysilane (APTES) in toluene and baked on a hotplate at 110 °C for 30 min. *The potential of piranha solution to cause severe injury mandates extreme caution during usage.* ITO coated glass slides, with sheet resistance of 70-100 ohm cm^{-1} (Delta Technologies, USA), are used as the counter electrode. ITO slides are cleaned by rinsing with ethanol, isopropyl alcohol (IPA), and DI water and then dried by N_2. Indium wire (Chip Quik, Canada) is used for electrical contact by soldering to surfaces. APTES, DMSO, ethanol, and IPA were purchased from Sigma Aldrich (USA).

3.2.2 Experimental Setup

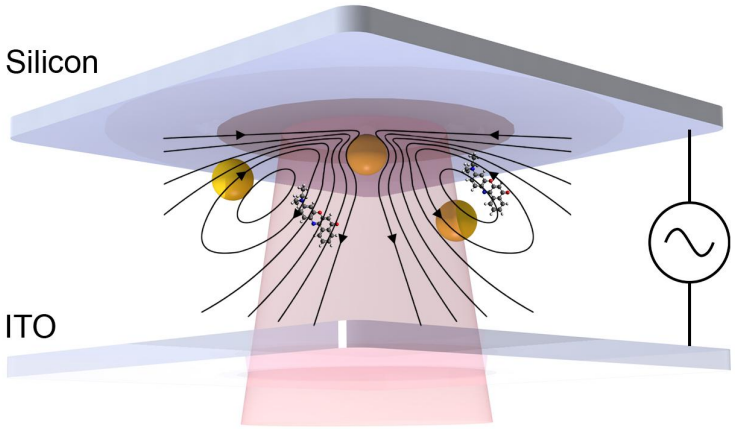

Figure 3.1 Schematic of microfluidic cell with oscillatory electric field between a silicon substrate (top) and ITO coated glass slide (bottom) for driving assembly of Au NP. The transparent ITO allows for *in situ* fluorescence imaging and SERS measurements.

A schematic of the experimental setup for in situ monitoring of Au NP EHD driven assembly is shown in Figure 1a. A Si substrate, serving as a working electrode, and an indium tin oxide (ITO) coated glass slide, serving as a counter electrode, are assembled in a capacitor architecture using a 90 µm spacer layer (9816L, 3M, USA). Before assembly, PMMA domains on PS-b-PMMA/Si are immersed in dimethyl sulfoxide (DMSO) for 5 min and then 5 vol % ethylenediamine in DMSO for 5 min in order to functionalize the PMMA regions with amine groups for coupling to Au NP. In the microfluidic cell, 20 µL of the following freshly prepared solution is used: equal parts of 20 mM N-hydroxysulfosuccinimide (s-NHS), and 8 mM 1-Ethyl-3-(3-dimethyl aminopropyl) carbodiimide (EDC) in a 0.1 M 2-[N-morpholino]ethanesulfonic acid (MES) buffer, 4 µL in total, is added to 0.25 mL of 2.6 nM lipoic acid functionalized Au NP

solution (Nanocomposix, USA). EDC, ethylenediamine, and s-NHS, MES buffer were purchased from Sigma Aldrich (USA).

After assembling the cell, EHD assembly is conducted using an oscillatory electric field as the assembly stimuli. A potential with an amplitude of 5 V is applied across electrodes for 2 min and the frequency is varied as follows: 100 Hz, 500 Hz, 1000 Hz and 1500 Hz. In the case of two-step deposition, an AC potential with amplitude of 5 V and frequency of 100 Hz is applied for 2 min. This is followed by a second deposition step that is conducted with the same AC potential and the frequency is varied between 500 and 1500 Hz. After deposition, the microfluidic cell is dismantled and the silicon surface is thoroughly rinsed with DI water and IPA and then dried with N_2 for further characterization. The samples were rinsed with water and IPA to mitigate clusters from forming due to capillary forces on the surface upon drying.

3.2.3 Characterization

SEM is performed on a Magellan 400 XHR SEM (FEI, USA) to examine the morphology of Au NP after EHD assembly. The SEM images have a surface area of approximately 30 μm^2 and a minimum of 5 images are acquired for each frequency condition used during deposition. In order to determine how frequency affects the number of Au NP in each cluster, Wolfram Mathematica(™) is utilized for image analysis to determine the number of clusters of a particular size on the surface, N. The distance differentiating clusters is defined as 100 nm. The surface coverage as a function of cluster size is determined by calculating the surface area covered of a particular cluster size, where n is the number of NP in the cluster, as a percentage of the total surface area (A) of the SEM images, $N \times n\pi r^2/A$.

For confocal fluorescence microscopy, Si was functionalized with APTES and Au NP functionalized with lipoic acid ligands were deposited on the surface with AC potential amplitude of 5 V and frequency of 100 Hz for 2 min with EDC and s-NHS, the same conditions as on the PS-b-PMMA surface. The Si electrode and Au NP seeds were coated with amorphous carbon with a thickness of 5 nm using a EM ACE 600 high vacuum sputter coater (Leica Microsystems, Germany). Then the Si and ITO electrodes were assembled into microfluidic cells in a capacitor architecture using a 20 μm spacer layer (3M 9816L). The confocal fluorescence microscopy is performed using a Zeiss LSM 780 confocal microscope (Zeiss, Germany) with LD C-Apochromat 63x/1.15 W Korr M27 lens, which has a free working distance of 600 μm. Nile red was used as the fluorophore. In the microfluidic channel, the solution was composed of 8 μL of an aqueous solution of 0.5 mM Nile red and 150 μL of 1.3 nM Au NP. An applied AC potential with amplitude of 5 V and a frequency of 500 Hz was held on for 20 s and off for 20 s for 12 cycles in total, while exciting the sample with a 561 nm laser throughout.

In situ SERS spectral imaging was performed by a i-Raman Plus Portable Raman Spectrometer (BWTEK, USA). The cell was mounted with the ITO electrode on the bottom and the Si electrode on top, allowing for *in situ* SERS monitoring in an inverted geometry.

3.3 Results and Discussion

3.3.1 EHD flow frequency dependence on cluster size

An oscillatory electric field is applied across the Si and ITO electrodes with Au NP in solution, having resistance of 8-9 MΩ, in the channel between electrodes. The interaction of NP with the electrodes with an applied AC potential can include electrophoretic deposition, EHD flow leading to lateral assembly, and assembly on defects.[16] The Si working electrode is coated with a PS-b-PMMA diblock copolymer thin film. Au NP are functionalized with lipoic acid for

carbodiimide crosslinking to amine terminated PMMA regions in order to lock-in assemblies for SEM imaging after the cell is disassembled.

In order to investigate how the frequency of the oscillatory field affects the assembly of NP, scanning electron microscopy (SEM) images were acquired on the Si working electrode after deposition with an AC potential having an amplitude of 5 V and with the frequency varied between 100 and 1500 Hz. Figure 3.2 a-d shows representative SEM images when the deposition is performed at this potential with frequencies of 100 Hz, 500 Hz, 1000 Hz and 1500 Hz, respectively. Examination of the images shows that the cluster size increases as the frequency of the AC potential increases. In order to obtain statistics on how cluster size varies with frequency, analysis of the surface coverage for different cluster sizes was performed over at least five SEM images, each with an area of $30\mu m^2$. The corresponding statistical analysis of observed cluster sizes is performed by first identifying the number of clusters with a specific number (n) of NP via image analysis by Wolfram Mathematica(™) and dividing the area covered by a particular cluster size with the total surface area. The percent area is shown in Figure 3.2 e., 3.2 f., 3.2 g. and 3.2 h., for deposition at frequency of 100 Hz, 500 Hz, 1000 Hz, and 1500 Hz, respectively.

From this analysis, deposition at a frequency of 100 Hz leads to primarily isolated NP as observed in Figure 3.2 e. This is consistent with prior observations showing that electrophoresis is the main driving force for assembly at low frequencies.[17] When deposition is performed at 500 Hz, there is a preference for clusters with three NP (Figure 3.2 f); at this frequency EHD flow forces influence assembly more than electrophoresis. As the deposition frequency increases to 1000 Hz and 1500 Hz, shown in Figure 3.2 g and 3.2 h, respectively, one can observe that larger

clusters, size greater than 10, begin to form at appreciable levels and there is no preference for a particular size when deposition occurs at 1500 Hz.

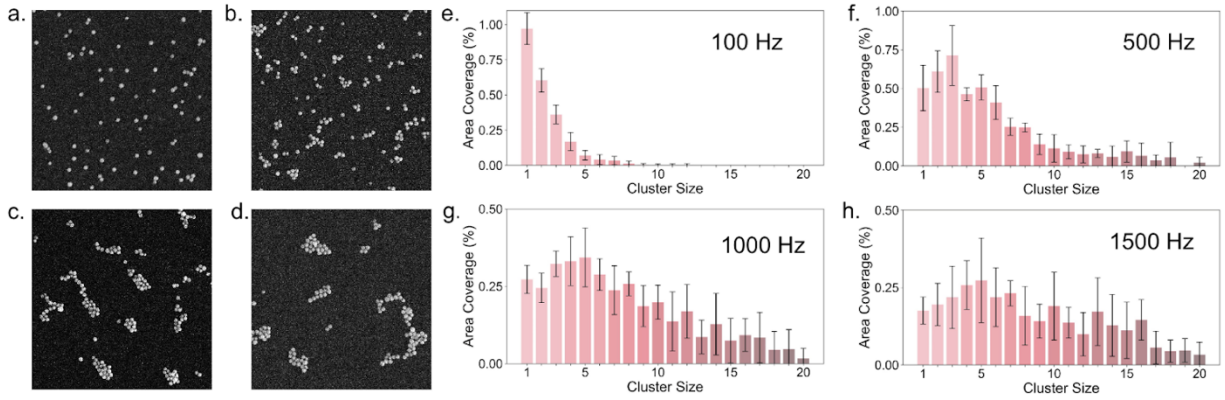

Figure 3.2 Representative SEM images of surfaces with a field of view of 2 µm x 2 µm after EHD assembly for 2 min with an AC potential having an amplitude of 5 V at frequency of a. 100 Hz, b. 500 Hz, c. 1000 Hz, and d. 1500 Hz. The occurrence of a particular cluster size is represented as a percent in terms of the fractional area covered with respect to the total surface area.

A 2-step deposition process was performed to investigate the ability to seed EHD flow by placing Au NP on the surface to serve as perturbations of the electrode potential. Based on the above analysis of Figure 3.3, a frequency of 100 Hz and potential amplitude of 5 V leads to primarily isolated NP. Thus these conditions were used for deposition for 2 min on a Si working electrode with a diblock copolymer thin film that has PMMA lamellar domains with approximate width of 80 nm and ITO as the counter electrode. The presence of the PS-b-PMMA diblock copolymer template aids in dispersing Au seeds on the surface as NP selectively attach to PMMA and not PS using EDC crosslinking chemistry.[18] After the first deposition, the microfluidic cell is dismantled and a freshly prepared solution of Au NP with EDC and s-NHS in MES buffer is added and the cell is reassembled. Then the potential of 5 V is applied with

variable frequency. Representative SEM images are shown in Figure 3.3 a, 3.3 b, and 3.3 c with frequency of 500 Hz, 1000 Hz, and 1500 Hz in the second step. The 2-step deposition process leads to increased surface coverage of NP in all cases. The seeded assembly also appears to nucleate the growth of more clusters and there is a higher number of clusters on the surface of smaller size. Even at frequencies of 1000 Hz and 1500 Hz few clusters are observed of size greater than 10.

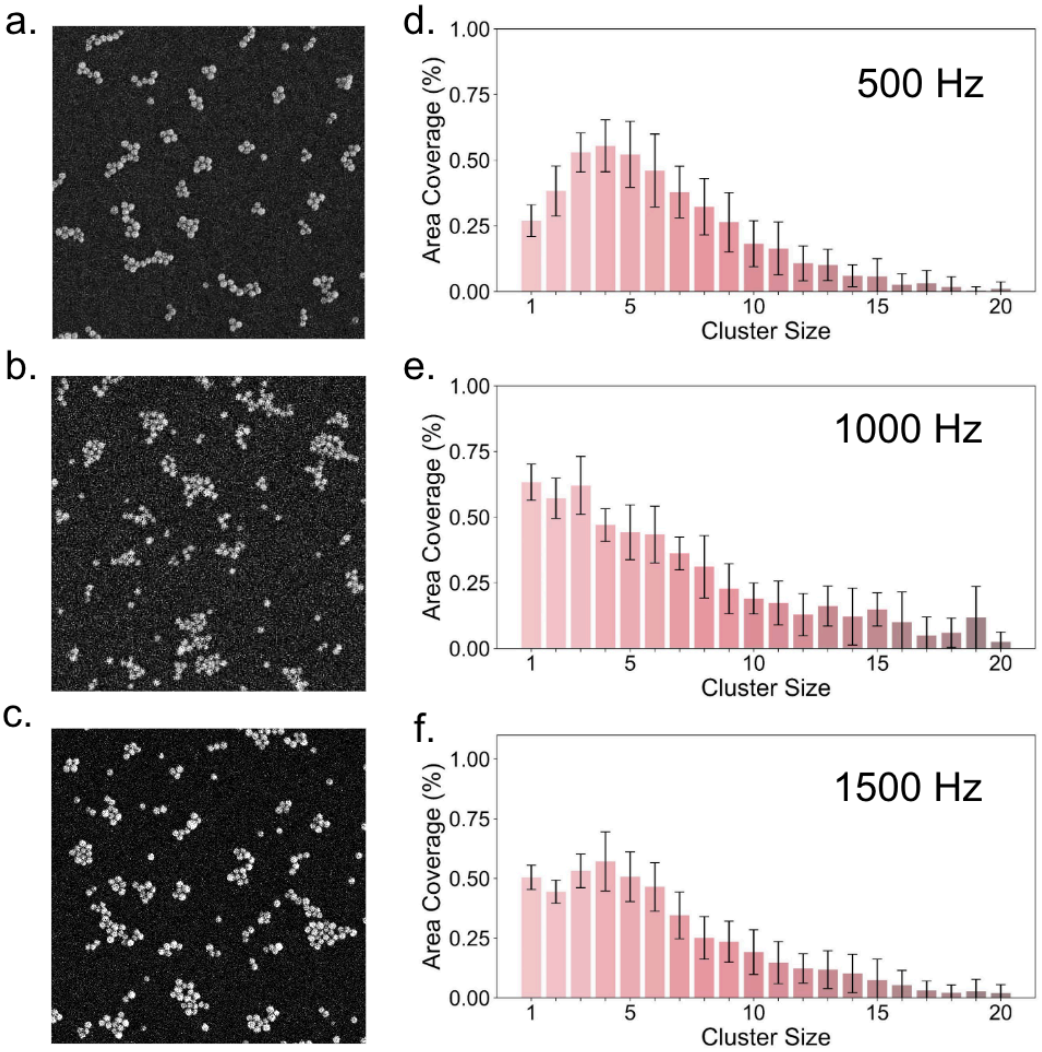

Figure 3.3 Representative SEM images of surfaces with a field of view of 2 µm x 2 µm after EHD assembly for 2 min with an AC potential having amplitude of 5 V and frequency of 100

Hz, followed by a second deposition for 2 min at **a.** 500 Hz, **b.** 1000 Hz, and **c.** 1500 Hz with corresponding statistical analysis of occurrence of cluster size shown in **d., e., f.**, respectively.

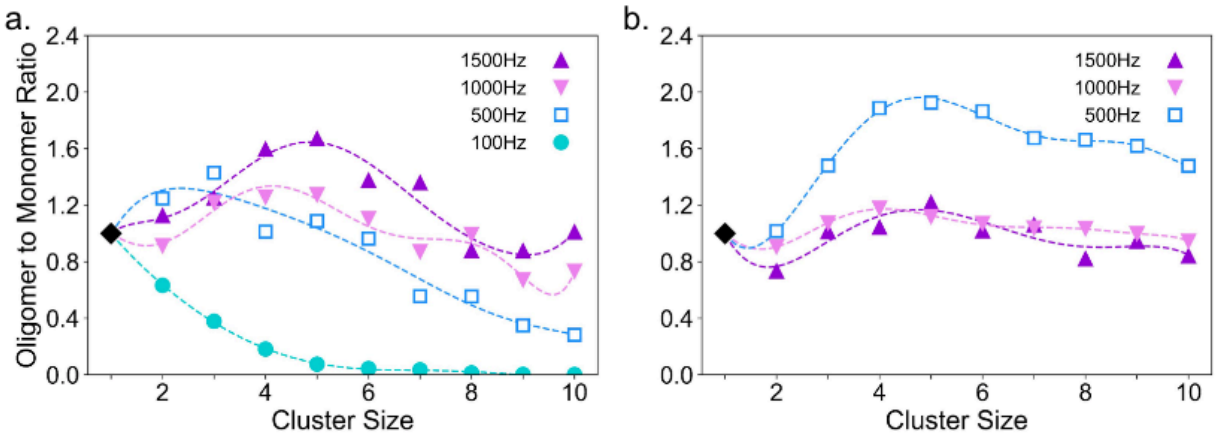

Figure 3.4 Oligomer to monomer ratio for **a.** single-step deposition and **b.** two step deposition. The teal closed circles represent deposition at 100 Hz, open blue squares represent deposition at 500 Hz, inverted triangles represent deposition at 1000 Hz, and triangles represent deposition at 1500 Hz. The AC potential amplitude was 5 V in all cases.

Figure 3.4 summarizes how the AC frequency affects oligomerization, i.e., the formation of NP clusters during single-step (Figure 3.2) and two-step deposition (Figure 3.3). In order to do so, the number of isolated NP observed on the Si working electrode after deposition is normalized to 1 and the occurrence of larger clusters, in terms of oligomer to monomer ratio, is plotted as a function of cluster size and frequency. Oligomer to monomer ratio for 100 Hz is represented by solid cyan circles, hollow blue squares represent 500 Hz, solid pink inverted triangle represents 1000 Hz and solid purple triangle represent 1500 Hz. Dashed lines are polynomial fits provided as a guide for the eye. Figure 3.4 a clearly conveys that at 100 Hz produces primarily monomers and at higher frequencies, EHD flow is the primary mechanism for deposition. Oligomerization shows different trends after the two-step deposition process. In Figure 3.4 b, deposition at 500 Hz has a higher degree of oligomerization than the single-stop

process due to the presence of the Au seeds on the surface tipping the balance toward deposition by EHD flow over electrophoresis. That is this observation is consistent that the Au NP are the source of the electrode perturbation driving EHD flow on these surfaces. Deposition at 1000 and 1500 Hz appears less influenced by the presence of Au seeds, thus we interpret that EHD flow is already a stronger driving force than electrophoresis at these frequencies. It appears Au seeds are deposited simultaneous with cluster growth. This would also explain the lack of preference for a particular cluster size when performing the two-step deposition process.

3.3.2 In situ optical imaging

Before fluorescence imaging, Au seeds are deposited on Si substrates using the same conditions of Figure 2a, V_{AC} = 5 V and f = 100 Hz, to generate EHD flow. Si was coated with APTES to adhere NP to the surface via electrostatic interactions. (The diblock copolymer is not used in this case due to a high fluorescence signal.) A thin carbon layer is then sputter coated on the Si electrode to mitigate further electrostatic interactions between the NP in solution and the surface. The microfluidic channel is filled with an aqueous solution of Au NP, 1.3 nM, and Nile red, 25 μM. EHD flow is introduced by applying an AC potential with an amplitude of 5 V and frequency of 500 Hz between the working and counter electrode. In order to measure the temporal response of dissipative assembly of Au NP in response to EHD flow, surfaces are imaged with confocal fluorescence spectroscopy *in situ*. The laser beam, with a wavelength of 561 nm, is focused with an objective through the ITO slide onto the Si electrode surface. Nile red has an absorption maximum at 552 nm and emission maximum at 633 nm. The emission wavelength is near the localized surface plasmon resonance (LSPR) of a Au NP dimer with inter-particle distance of approximately 2 nm in aqueous solution as measured previously for this system.[7,19] In the experimental setup we expect the inter-particle spacing will be larger as the

carbon layer is approximately 5 nm in thickness. This will redshift the LSPR such that larger clusters will be resonant with Nile red emission. Thus when Nile red fluorophores are in the vicinity of Au NP clusters, the fluorescence emission has been reported to be enhanced by more than an order of magnitude.[15] The observed enhancements to the fluorescence signal when the AC field is on are consistent with dye molecules in NP gaps of distance of approximately 5-10 nm.[20]

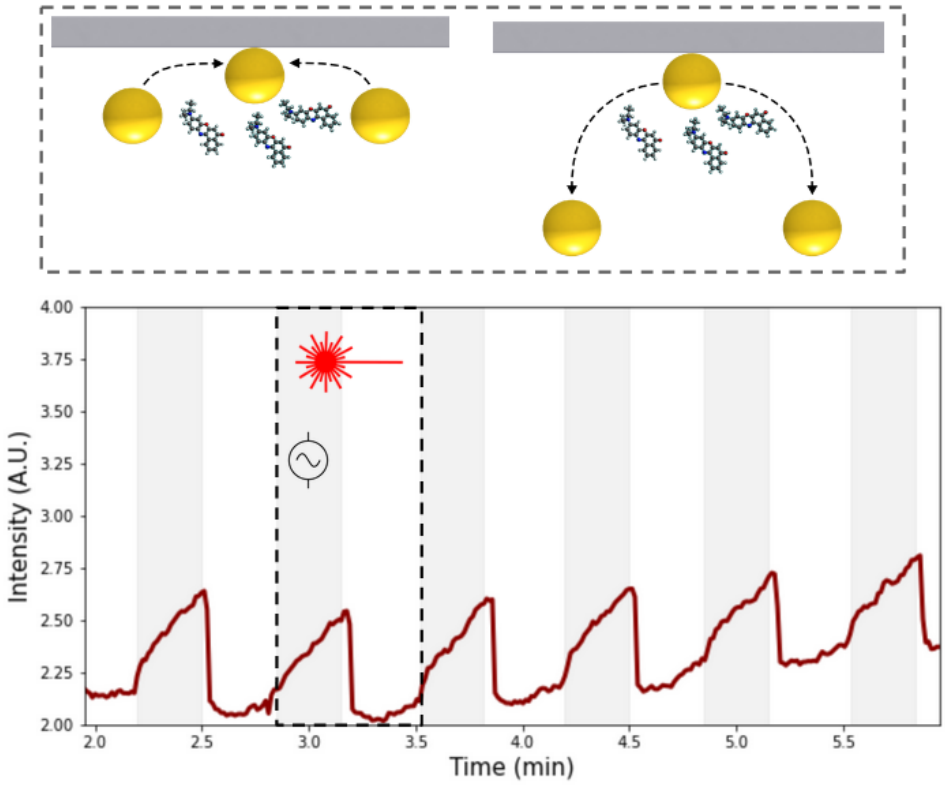

Figure 3.5 Fluorescence intensity variation of Nile red in Au NP solution when the oscillatory electric field is cycled on for 20 s and off for 20 s. The laser is on for the duration of the experiment. The inset above is a schematic illustrating assembly and disassembly in response to the external stimuli.

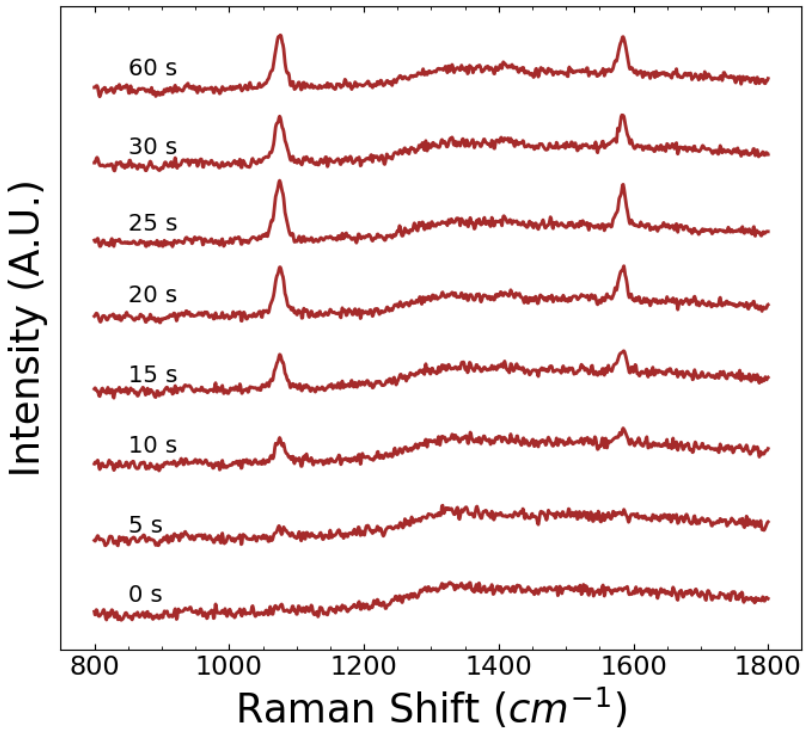

Figure 3.6 SERS *in situ* imaging of EHD driven deposition of 4-MBA functionalized Au NP on surfaces with lipoic acid functionalized Au seeds on a PS-b-PMMA coated Si electrode.

In addition to fluorescence imaging, SERS imaging is performed *in situ* to monitor the dissipative assembly of Au NPs in response to EHD flow. Au seeds functionalized with lipoic acid are assembled on a diblock copolymer Si electrode similarly as those for fluorescence imaging, except a diblock copolymer thin film is used instead of APTES for chemically crosslinking Au seeds on the surface as SERS is less sensitive to a fluorescence background. A liquid cell was formed by sandwiching a 1.3 nM solution of Au NP functionalized with 4-mercaptobenzoic acid (4MBA), a Raman reporter molecule, between the seeded Si electrode and an ITO-coated glass slide. EHD flow was introduced by applying an AC bias with an amplitude of 5 V at 500 Hz to grow NP clusters between the seeds and the 4MBA-AuNP. A laser beam with a wavelength of 785 nm is focused on the Si electrode and the SERS signal of 4MBA at 1080 cm^{-1} and 1590 cm^{-1} is used to measure the assembly dynamics in real time.

3.4 Conclusion

In conclusion, we provide direct evidence for the role of the EHD flow in the dissipative self-assembly process of 40 nm Au NPs. Owing to unique LSPR properties of metal NPs, profound organic dye molecules and Au NPs interactions result in significant modulation in fluorescence intensity of dye molecules, aggregated metal NPs can give rise to a giant local electric field at the gap region and thus expected to display much larger fluorescence enhancement compared to isolated Au NPs. We demonstrate the direct visualization and real-time monitoring of dissipative self-assembly by using NPs as building blocks. Fluorescence represents a refined approach to monitor the dissipative kinetics with high sensitivity and reliability. It offers excellent contrast ratio and is attractive for developing platforms for highly sensitive sensing and imaging applications. This simple method can be easily extend for detection of many other analytes which can induce the aggregation of metal NPs through various assembly methods. Further surface modification of metal NPs with proper recognition moiety and with targeting groups is critical for selective interactions and effect assembly based on aggregation of metal NP's strong plasmon coupling interactions and large fluorescence enhancement.

3.5 Supplemental Information

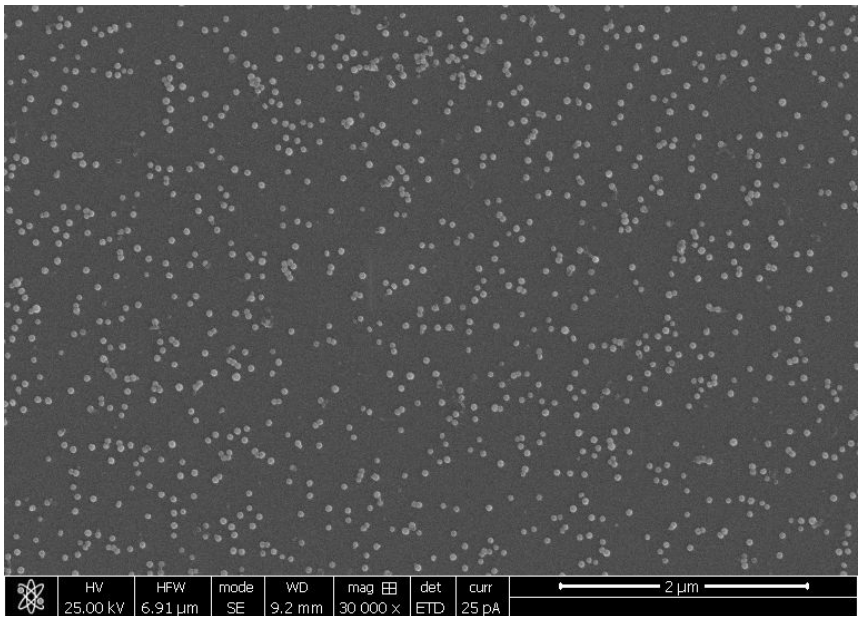

Figure 3.7 SEM image of the substrate for confocal measurement of plasmonic nanoparticle dissipative assembly.

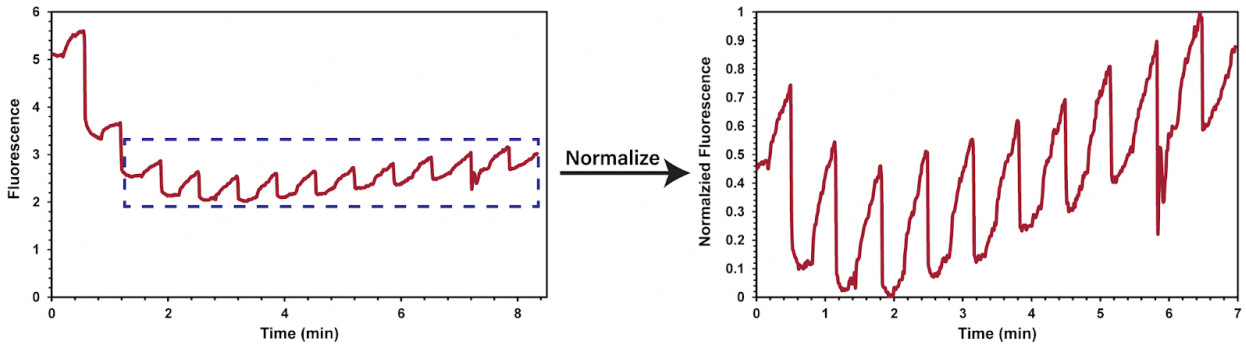

Figure 3.8 Reversible fluorescence intensity of Nile red at 561 nm for Au NP dissipative self-assembly. The dye molecules settle on the carbon surface and exhibit higher fluorescence intensity before the application of the electric field. The initial 3 cycles facilitate the purging of the dye molecules back into the solution. The fluorescence signal of Nile red is plotted against time, with the first 10 s capturing the initial fluorescence intensity using the 561 nm laser only. Subsequently, the laser was turned on continuously, and an oscillatory electric field was applied for 20 s to monitor the assembly of gold nanoparticles, followed by 20 s with the electric field turned off to monitor their disassembly, constituting one cycle. The experiment consisted of 12 cycles.

Table 3.1 Coverage, clusters/monomer ratio for single step deposition

Cluster distribution	5 V 100 Hz	5 V 500 Hz	5 V 1000 Hz	5 V 1500 Hz
NP coverage (#/μm^2)	28	37	31	27
Percentage coverage	3.52%	4.68%	5.82%	4.93%
2/1 ratio (cluster)	0.47 (0.24)	1.25 (0.62)	0.86 (0.43)	0.89 (0.44)
3/1 ratio	0.30 (0.10)	1.43 (0.48)	0.86 (0.29)	1.07 (0.36)
4/1 ratio	0.12 (0.03)	1.01 (0.25)	0.70 (0.17)	1.15 (0.29)
5/1 ratio	0.08 (0.02)	1.09 (0.22)	0.69 (0.14)	1.02 (0.20)
6/1 ratio	0.06 (0.01)	0.96 (0.16)	0.61 (0.10)	0.93 (0.15)
7/1 ratio	0.02 (0.00)	0.56 (0.08)	0.48 (0.07)	0.69 (0.10)
8/1 ratio	0.02 (0.00)	0.55 (0.07)	0.49 (0.06)	0.51 (0.06)
9/1 ratio	0.00 (0.00)	0.35 (0.04)	0.34 (0.04)	0.47 (0.05)
10/1 ratio	0.02 (0.00)	0.28 (0.03)	0.30 (0.03)	0.39 (0.04)

Table 3.2 Coverage, clusters/monomer ratio two step deposition

Cluster distribution	5 V 100 Hz	5 V 100 Hz + 5 V 500 Hz	5 V 100 Hz + 5 V 1000 Hz	5 V 100 Hz + 5 V 1000 Hz
NP coverage (#/μm^2)	18	36	46	39
Percentage coverage	2.27%	4.57%	5.82%	4.93%
2/1 ratio(cluster)	0.63 (0.32)	1.45 (0.73)	0.86 (0.43)	0.89 (0.44)
3/1 ratio	0.3789 (0.1263)	2.04 (0.68)	0.86 (0.29)	1.07 (0.36)
4/1 ratio	0.18 (0.0455)	2.17 (0.54)	0.70 (0.17)	1.15 (0.29)
5/1 ratio	0.08 (0.02)	2.06 (0.41)	0.69 (0.14)	1.02 (0.20)
6/1 ratio	0.04 (0.01)	1.82 (0.30)	0.61 (0.10)	0.93 (0.15)
7/1 ratio	0.04 (0.01)	1.49 (0.21)	0.48 (0.07)	0.69 (0.10)
8/1 ratio	0.01 (0.00)	1.29 (0.16)	0.49 (0.06)	0.51 (0.06)
9/1 ratio	0.00 (0.00)	1.05 (0.12)	0.34 (0.04)	0.47 (0.05)
10/1 ratio	0.00 (0.00)	0.75 (0.0751)	0.30 (0.03)	0.39 (0.04)

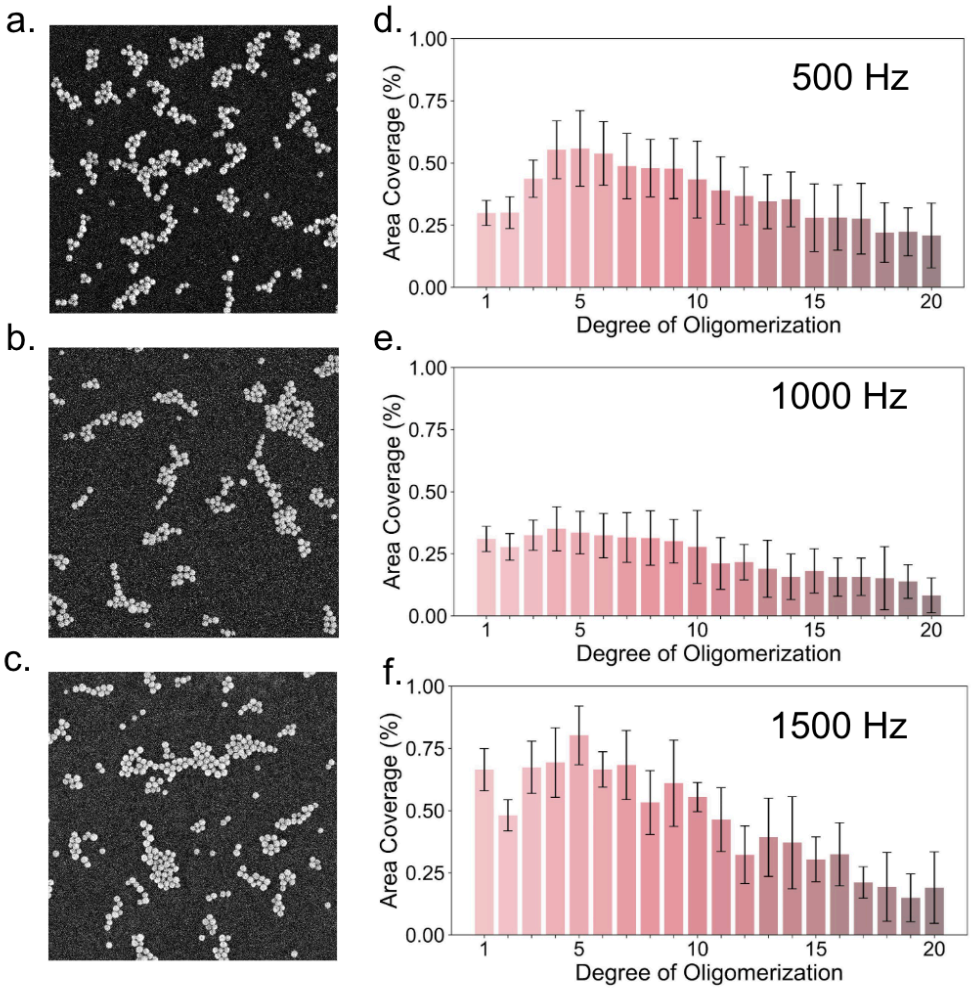

Figure 3.9 Representative SEM images and statistical analysis for structures assembled via a two-step deposition process. The first step is performed with a bias of 5 V and frequency of 100 Hz to have isolated NPs on the surface prior to second step depositions. The second step depositions are investigated at 5 V with different frequencies: **a.** 500 Hz, **b.** 1000 Hz, **c.** 1500 Hz and their corresponding statistical analysis are shown in **d.**, **e.**, **f.**, respectively. The deposition time for the second step is 4 mins.

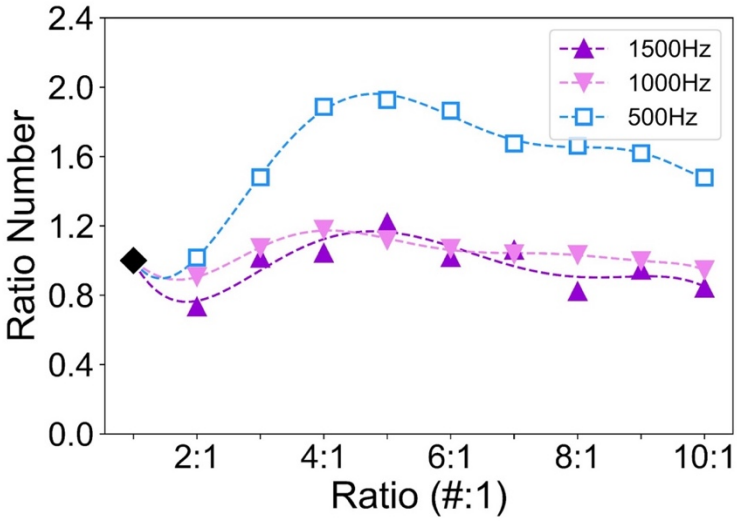

Figure 3.10 Cluster size ratio analysis plot. The x axis in ratio indicates the number of NPs inside a cluster, and y axis represents the ratio between different numbers of NPs inside a cluster to isolated NPs.

Table 3.3 Coverage, clusters/monomer ratio for two step deposition (4 mins deposition time for the second step)

Cluster distribution	5 V_p 100 Hz + 5 V_p 500 Hz	5 V_p 100 Hz + 5 V_p 1000 Hz	5 V_p 100 Hz + 5 V_p 1500 Hz
NP coverage (#/μm^2)	81	48	89
Percentage coverage	10.24%	6.00%	11.17%
2/1 ratio(cluster)	1.02 (0.51)	0.91 (0.45)	0.73 (0.37)
3/1 ratio	1.48 (0.49)	1.07 (0.36)	1.01 (0.34)
4/1 ratio	1.89 (0.47)	1.18 (0.29)	1.04 (0.26)
5/1 ratio	1.93 (0.39)	1.12 (0.22)	1.22 (0.24)
6/1 ratio	1.87 (0.31)	1.07 (0.18)	1.02 (0.17)
7/1 ratio	1.68 (0.24)	1.04 (0.15)	1.06 (0.15)
8/1 ratio	1.66 (0.21)	1.03 (0.13)	0.82 (0.10)
9/1 ratio	1.62 (0.18)	1.00 (0.11)	0.95 (0.11)
10/1 ratio	1.48 (0.15)	0.94 (0.09)	0.84 (0.08)

References

(1) Ma, Y.; Zagar, C.; Klemme, D. J.; Sikdar, D.; Velleman, L.; Montelongo, Y.; Oh, S.-H.; Kucernak, A. R.; Edel, J. B.; Kornyshev, A. A. A Tunable Nanoplasmonic Mirror at an Electrochemical Interface. *ACS Photonics* **2018**, *5* (11), 4604–4616. https://doi.org/10.1021/acsphotonics.8b01105.

(2) Ma, Y.; Sikdar, D.; Fedosyuk, A.; Velleman, L.; Klemme, D. J.; Oh, S.-H.; Kucernak, A. R. J.; Kornyshev, A. A.; Edel, J. B. Electrotunable Nanoplasmonics for Amplified Surface Enhanced Raman Spectroscopy. *ACS Nano* **2020**, *14* (1), 328–336. https://doi.org/10.1021/acsnano.9b05257.

(3) Montelongo, Y.; Sikdar, D.; Ma, Y.; McIntosh, A. J. S.; Velleman, L.; Kucernak, A. R.; Edel, J. B.; Kornyshev, A. A. Electrotunable Nanoplasmonic Liquid Mirror. *Nat. Mater.* **2017**, *16* (11), 1127–1135. https://doi.org/10.1038/nmat4969.

(4) Nan, F.; Han, F.; Scherer, N. F.; Yan, Z. Dissipative Self-Assembly of Anisotropic Nanoparticle Chains with Combined Electrodynamic and Electrostatic Interactions. *Adv. Mater.* **2018**, *30* (45), 1803238. https://doi.org/10.1002/adma.201803238.

(5) Maestas, J. R.; Ma, F.; Wu, N.; Wu, D. T. Electric-Field-Driven Assembly of Dipolar Spheres Asymmetrically Confined between Two Electrodes. *ACS Nano* **2021**, *15* (2), 2399–2412. https://doi.org/10.1021/acsnano.0c04939.

(6) Adams, S. M.; Campione, S.; Capolino, F.; Ragan, R. Directing Cluster Formation of Au Nanoparticles from Colloidal Solution. *Langmuir* **2013**, *29* (13), 4242–4251. https://doi.org/10.1021/la3051719.

(7) Thrift, W. J.; Nguyen, C. Q.; Darvishzadeh-Varcheie, M.; Zare, S.; Sharac, N.; Sanderson, R. N.; Dupper, T. J.; Hochbaum, A. I.; Capolino, F.; Abdolhosseini Qomi, M. J.; Ragan, R. Driving Chemical Reactions in Plasmonic Nanogaps with Electrohydrodynamic Flow. *ACS Nano* **2017**, *11* (11), 11317–11329. https://doi.org/10.1021/acsnano.7b05815.

(8) Trau, M.; Saville, D. A.; Aksay, I. A. Assembly of Colloidal Crystals at Electrode Interfaces. *Langmuir* **1997**, *13* (24), 6375–6381. https://doi.org/10.1021/la970568u.

(9) Goel, M.; Singh, A.; Bhola, A.; Gupta, S. Size-Tunable Assembly of Gold Nanoparticles Using Competitive AC Electrokinetics. *Langmuir* **2019**, *35* (24), 8015–8024. https://doi.org/10.1021/acs.langmuir.8b03963.

(10) Ristenpart, W. D.; Aksay, I. A.; Saville, D. A. Assembly of Colloidal Aggregates by Electrohydrodynamic Flow: Kinetic Experiments and Scaling Analysis. *Phys. Rev. E* **2004**, *69* (2), 021405. https://doi.org/10.1103/PhysRevE.69.021405.

(11) Harraq, A. A.; Choudhury, B. D.; Bharti, B. Field-Induced Assembly and Propulsion of Colloids. *Langmuir* **2022**, *38* (10), 3001–3016. https://doi.org/10.1021/acs.langmuir.1c02581.
(12) Dutcher, C. S.; Woehl, T. J.; Talken, N. H.; Ristenpart, W. D. Hexatic-to-Disorder Transition in Colloidal Crystals Near Electrodes: Rapid Annealing of Polycrystalline Domains. *Phys. Rev. Lett.* **2013**, *111* (12), 128302. https://doi.org/10.1103/PhysRevLett.111.128302.
(13) Ristenpart, W. D.; Aksay, I. A.; Saville, D. A. Electrohydrodynamic Flow around a Colloidal Particle near an Electrode with an Oscillating Potential. *J. Fluid Mech.* **2007**, *575*, 83–109. https://doi.org/10.1017/S0022112006004368.
(14) Grzelczak, M.; Liz-Marzán, L. M.; Klajn, R. Stimuli-Responsive Self-Assembly of Nanoparticles. *Chem. Soc. Rev.* **2019**, *48* (5), 1342–1361. https://doi.org/10.1039/C8CS00787J.
(15) Tam, F.; Goodrich, G. P.; Johnson, B. R.; Halas, N. J. Plasmonic Enhancement of Molecular Fluorescence. *Nano Lett.* **2007**, *7* (2), 496–501. https://doi.org/10.1021/nl062901x.
(16) Ristenpart, W. D.; Jiang, P.; Slowik, M. A.; Punckt, C.; Saville, D. A.; Aksay, I. A. Electrohydrodynamic Flow and Colloidal Patterning near Inhomogeneities on Electrodes. *Langmuir* **2008**, *24* (21), 12172–12180. https://doi.org/10.1021/la801419k.
(17) Ferrick, A.; Wang, M.; Woehl, T. J. Direct Visualization of Planar Assembly of Plasmonic Nanoparticles Adjacent to Electrodes in Oscillatory Electric Fields. *Langmuir* **2018**, *34* (21), 6237–6248. https://doi.org/10.1021/acs.langmuir.8b00992.
(18) Choi, J. H.; Adams, S. M.; Ragan, R. Design of a Versatile Chemical Assembly Method for Patterning Colloidal Nanoparticles. *Nanotechnology* **2009**, *20* (6), 065301. https://doi.org/10.1088/0957-4484/20/6/065301.
(19) Nguyen, C. Q.; Thrift, W. J.; Bhattacharjee, A.; Ranjbar, S.; Gallagher, T.; Darvishzadeh-Varcheie, M.; Sanderson, R. N.; Capolino, F.; Whiteson, K.; Baldi, P.; Hochbaum, A. I.; Ragan, R. Longitudinal Monitoring of Biofilm Formation via Robust Surface-Enhanced Raman Scattering Quantification of Pseudomonas Aeruginosa-Produced Metabolites. *ACS Appl. Mater. Interfaces* **2018**, *10* (15), 12364–12373. https://doi.org/10.1021/acsami.7b18592.
(20) Lu, D.; Hou, S.; Liu, S.; Xiong, Q.; Chen, Y.; Duan, H. Amphiphilic Janus Magnetoplasmonic Nanoparticles: PH-Triggered Self-Assembly and Fluorescence Modulation. *J. Phys. Chem. C* **2022**, *126* (35), 14967–14975. https://doi.org/10.1021/acs.jpcc.2c03753.

Chapter 4

Illuminating Bacterial Metabolism with Plasmonic Structures

4.1 Introduction

Like all living organisms, bacteria are equipped with biochemical machinery to survive and adapt in diverse and changing environments all over the world. These responses to dynamic conditions elicit changes in bacteria metabolic networks, and their metabolite profiles can shift on timescales as short as minutes.[1] Many of these environmental changes constitute stresses, which trigger physiological responses within the cell. Stresses, ranging from nutrient restriction[2] to exposure to antibiotics,[3] elicit profound metabolic consequences in bacteria. The resulting changes in metabolite profiles can be detected by conventional[3] and next-generation[4] metabolomic techniques. Consequently, we hypothesize and demonstrate that bacterial cultures can be used as whole-cell sensors of antibiotics and nutrient stressors by the detection and decoding of their metabolic responses to these stressors. Specifically, the bacterial metabolic response transduces stresses into chemical (metabolite) signals that are amplified with surface enhanced Raman scattering (SERS) surfaces. When decoding the spectral signals using machine learning (ML) algorithms, a sensitive and accurate sensing platform for rapid antimicrobial testing and nutrient source detection. In this Chapter, we employ SERS+ML to investigate the metabolism of glucose, sucrose, and the joint regulation of glucose and xylose metabolic pathways for *E.coli*.

4.2 Deep neural network models for antimicrobial susceptibility testing

Physicians often prescribe antibiotics to treat clinical infections based on insufficient information; an informed response is delayed since the culturing of a patient's sample requires

24-72 hours.[5] As a result, a full third of prescribed antibiotics would qualify as either overuse or misuse.[6] As shown in Figure 4.1, bacterial infections exhibiting antimicrobial resistance (AMR) cause 700,000 deaths per year globally,[7] and by 2050 it is expected to increase to 10 million deaths per year.[8] A rapid diagnostics method is required urgently to reduce the inappropriate use of antimicrobials contributing to this severe global AMR issue, and rapid antimicrobial susceptibility testing (AST) is a promising method to tackle this problem by assisting physicians to select proper antibiotic treatment in the time span of a typical doctor visit.[9]

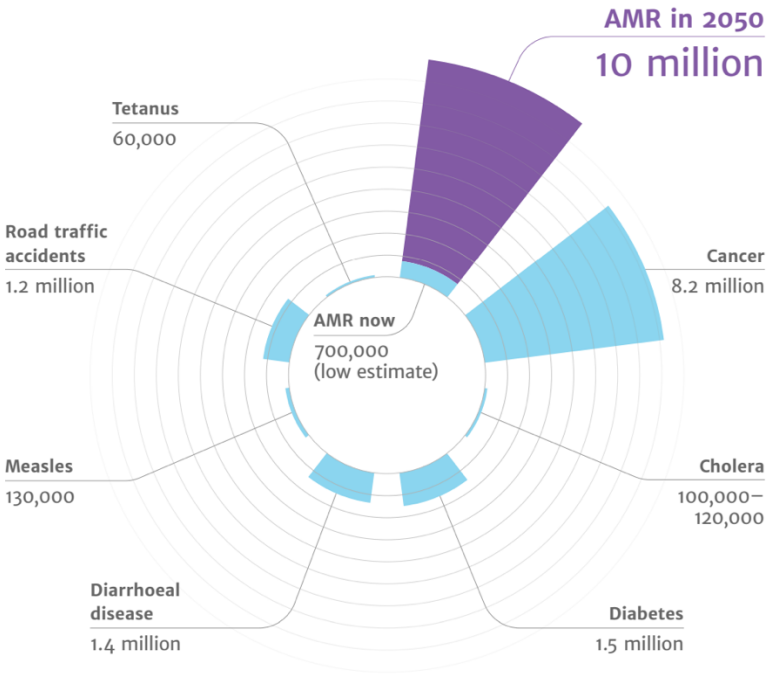

Figure 4.1 Deaths attributable to AMR every year by 2050

Phenotypic AST has been considered the gold standard[10] to provide direct metrics of antibiotic susceptibility from patient samples, yet the time to culture cells for phenotypic AST delays an informed diagnosis and treatment plan. On the other hand, examination of bacterial response to antibiotics using mass spectroscopy has shown the dysregulation of core metabolic

cellular functions is correlated to lethality.[11–14] The metabolite profiles exhibit changes as soon as 30 minutes after exposure when bacteria are exposed to effective antibiotics. However, there is no real time device solution to detect the dynamic change of metabolites as mass spectroscopy requires sophisticated sample preparation.[15] The focus of this section is sensors able to detect this rapid metabolic response and characterize their performance as a rapid phenotypic AST diagnostic device.

A combination of SERS spectra and machine learning data analysis is a promising new approach that has the potential to enable rapid AST. However, in order to build a robust ML model for rapid AST, SERS data reflecting AMR states need to be acquired for model training. Proper association between AMR states and SERS data sets requires validation by traditional AST method which takes 24-72 hours for growth assays.[25] Thus we need to minimize the amount of SERS data needed from growth assays which reflects phenotypic bacterial response to antibiotics, referred to as labeled data. SERS+ML will be a highly competitive method to achieve time efficient rapid AST.[26]

In order to more accurately determine the dosage and time point at which differentiating the SERS spectra becomes possible, we use a deep neural network (DNN) model, a supervised learning approach, to classify the spectra from *E. coli* and *P. aeruginosa* lysate with respect to their temporal and gentamicin dosage treatment conditions. The DNN is trained directly on the raw spectral data and we develop multiple two-class DNN models with groupings of consecutive treatment conditions. For example, one experiment for the temporal datasets would classify the 0 and 5 min treatment conditions (first class) against the 10, 20, and 40 min treatment conditions (second class). The resultant classification accuracy of these experiments are 99% ± 0.1% for both *E. coli* and *P. aeruginosa*. Thus, we are able to detect bacterial response to antibiotics after

10 min with greater than 99% accuracy. Examining all two-class model results, as shown in Table 4.1, we can see that the two-class feed-forward DNN models are able to distinguish between all possible groupings of temporal conditions with near-perfect mean 10-fold cross validation accuracy greater than 99% even when grouping the 0 min temporal response alone in a class. Thus, after only 5 min, bacterial response to antibiotics is clearly evident in the SERS spectra. The DNN performed equally well on all possible groupings in the two class analysis of dosage-variant datasets, also shown in Table 4.1. Because the two-class models performed so well, we explored five-class models in order to analyze the performance of the DNN model for differentiating each individual condition.

Table 4.1 2-Class DNN model performance metrics.

	***E. coli* Temporal**			***P. aeruginosa* Temporal**		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
(0min) | (5min, 10min,20min,40min)	99% ± 0.1%	100%	99%	99% ± 0.2%	100%	99%
(0min, 5min) | (10min,20min,40min)	99% ± 0.1%	99%	100%	99% ± 0.1%	99%	100%
(0min, 5min, 10min) | (20min,40min)	99% ± 0.2%	99%	100%	99% ± 0.2%	99%	99%
(0min, 5min, 10min, 20min) | (40min)	99% ± 0.1%	99%	100%	99% ± 0.1%	99%	99%

	***E. coli* Dosage**			***P. aeruginosa* Dosage**		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
(0 µg/mL) | (0.1 µg/mL, 0.5 µg/mL, 1 µg/mL, 10 µg/mL)	99% ± 1%	100%	99%	98% ± 1%	99%	95%
(0 µg/mL, 0.1 µg/mL) | (0.5 µg/mL, 1 µg/mL, 10 µg/mL)	97% ± 1%	98%	95%	98% ± 1%	98%	98%
(0 µg/mL, 0.1 µg/mL, 0.5 µg/mL) | (1 µg/mL, 10 µg/mL)	95% ± 1%	93%	95%	99% ± 1%	98%	99%
(0 µg/mL, 0.1 µg/mL, 0.5 µg/mL, 1 µg/mL) | (10 µg/mL)	99% ± 1%	90%	99%	98% ± 1%	94%	99%

The five-class models achieved comparable accuracy as can be seen from the confusion matrices for each of the models (Figure 4.2). For the *P. aeruginosa* time variant dataset, we were able to achieve a mean 10-fold cross validation classification accuracy of 99 ± 0.2%. For the dosage variant dataset, the model had a mean 10-fold cross validation classification accuracy of 98 ± 1.0%. The *E. coli* temporal and dosage datasets performed similarly well, with mean 10-fold cross validation classification accuracies of 99 ± 0.3% and 95 ± 1% respectively. Looking at the confusion matrices in Figure 4.2 b,d for both temporal datasets, we can clearly see early differentiation with the model showing a strong ability to differentiate classes as early as 5 min. For both models trained on the dosage datasets, we see clear differentiation with dosages as low as 0.1 μg/mL. It is important to note that the majority of misclassification results are for similar dosages or time points and thus SERS data is able to measure bacterial response below the MIC and track the temporal evolution of the bacterial response on a time scale of 5 min with high accuracy. Classification accuracy, equal to sensitivity, is considered the primary metric in this work since the goal of rapid AST is accurate antibiotic treatment.

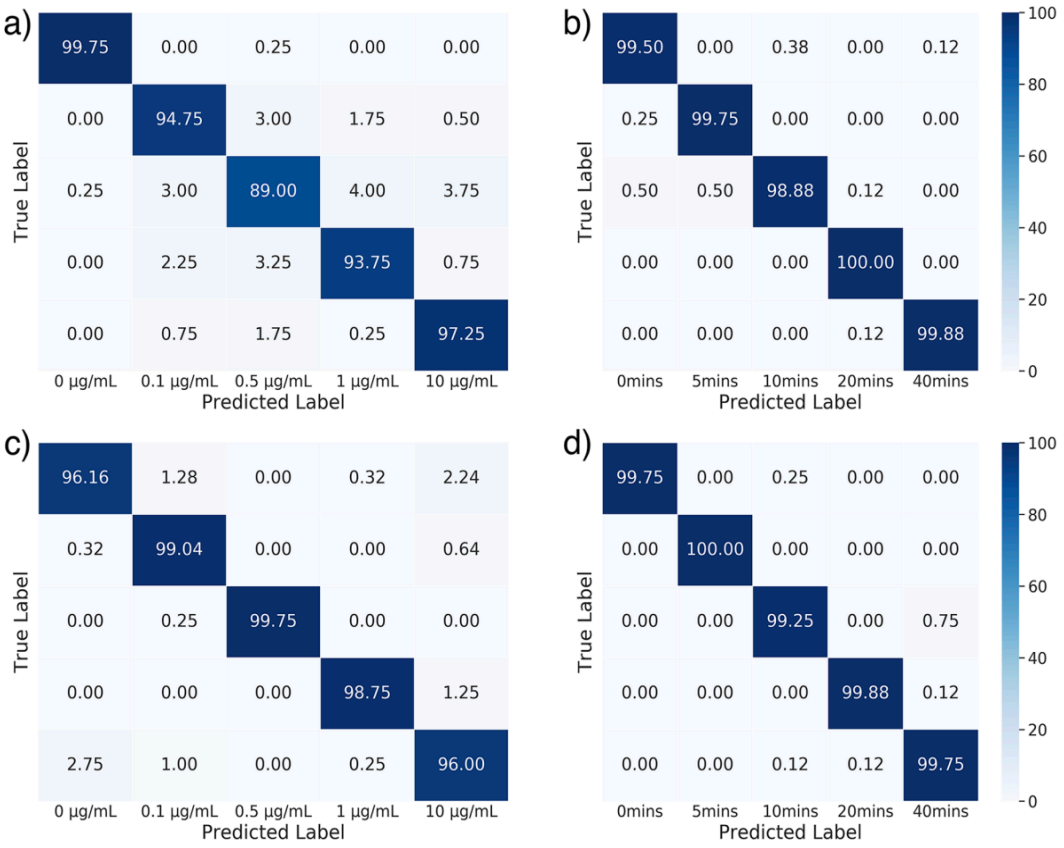

Figure 4.2 Five-class DNN model confusion matrices (values are listed as percentages). The datasets used are *E. coli* (a) gentamicin dose and (b) temporal response dataset, and *P. aeruginosa* (c) gentamicin dose and (d) temporal response dataset.

4.3 Classifying type of nutrient source by support vector machine

Nutrients are essential for bacteria survival, with carbon and nitrogen sources playing a crucial role in their metabolism. While some nutrients can be readily metabolized, others are not, and there may be a preference for one nutrient over others when multiple sources are available. The complex and multifaceted nature of nutrient utilization in bacteria underscores the importance of understanding their nutrient source metabolism. This knowledge is critical for predicting their competitive success in natural environments. In natural environments, bacteria

may need to compete for limited resources, including nutrients. Therefore, selection of preferred nutrient sources can greatly influence bacterial growth rates and competitive success with other microorganisms. *Escherichia coli* (*E.coli*), an industrial biotechnology workhorse,[18] can uptake a variety of carbon source and induce distinct metabolic pathways to survive.[19] Among all nutrient source, glucose is a crucial source of energy for *E.coli*, as it serves as the primary metabolic fuel for this organism, and is a major precursor for the biosynthesis of various carbohydrates including glycogen, ribose, deoxyribose, galactose, glycolipids, glycoproteins and proteoglycans. Sucrose is an attractive industrial carbon source due to its abundance. However, wild-type *E.coli* is unable to metabolize sucrose, which can lead to starvation stress when no other consumable sugar source is available.[20]

SERS spectra of *E. coli* lysate were obtained after growth in different nutrient sources, including glucose, sucrose, and a control without a carbon source. Representative SERS spectra in shown in Fig. 4.3 and the most significant peak intensity change is highlighted by the grey bands. Principal component analysis (PCA) was used to reduce the dimensionality of the SERS spectra highlight nutrient-dependent spectral features. The PC loadings as shown in Fig. 4.4 were then used for revealing differential metabolic signatures in response to nutrient exposure. The largest loading features at 734 cm^{-1} (adenosine[25]) and 1030 cm^{-1} (phenylalanine[24] or adenine[23]) were found to correlate with energy nucleotides and metabolic pathways involved in nutrient uptake, suggesting that changes in nucleotide concentrations are consistent with the nutrient-dependent metabolic signatures observed.

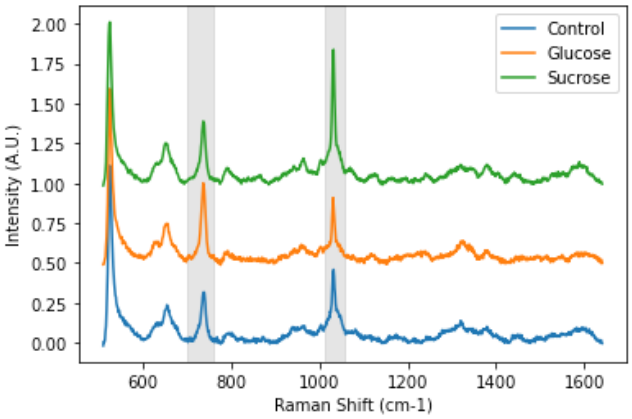

Figure 4.3 Representative SERS spectra (vertically offset) acquired from *E. coli* lysate after culturing in different nutrient sources, including glucose (orange line), sucrose (green line), and a control without a carbon source (blue line).

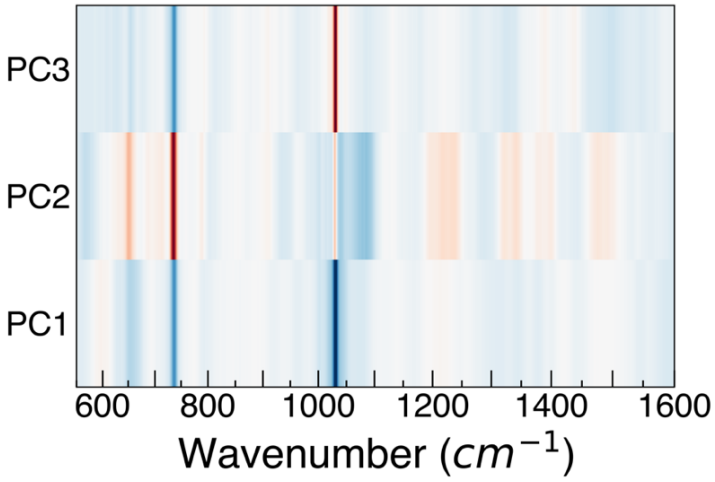

Figure 4.4 PC1, 2 and 3 heat map of the dataset containing SERS spectra of *E. coli* lysate after culturing in different nutrient sources, glucose, sucrose, and a control without a carbon source.

A support vector machine (SVM) discriminative ML model was trained to access the accuracy of discriminating between different nutrient sources and evaluate metabolism features. Each nutrient source was designated as a class in the SVM model, which was trained using 80% of the spectral data. The resulting 20% of the data, which were not seen during the SVM model

training, were used as a holdout set to determine the classification accuracy of the algorithm predictions. The resulting classification accuracy was depicted in the confusion matrices presented in Fig. 4.5. Notably, the false predictions rate of control spectra being classified as nutrient conditions was less than 0.6%.

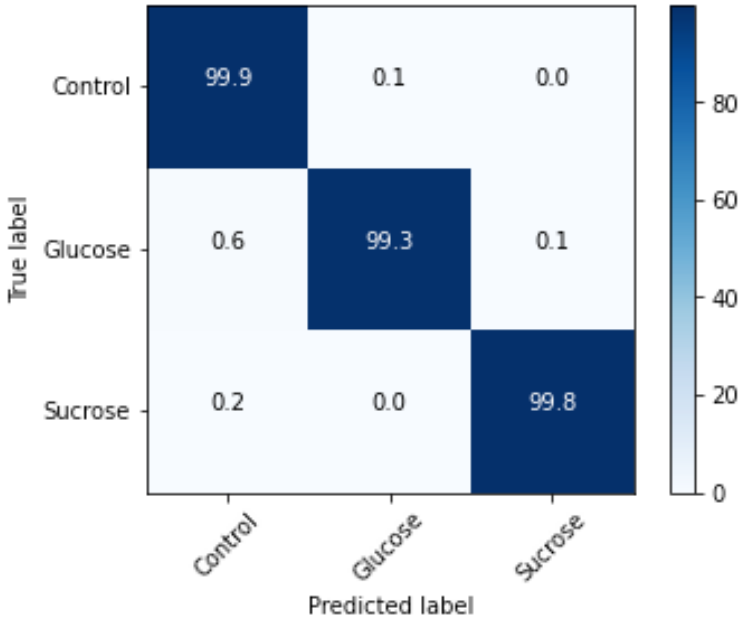

Figure 4.5 SVM confusion matrix for classification between SERS spectra of *E. coli* lysate after culturing in glucose, sucrose, and a control without a carbon source, respectively.

4.4 Visualizing carbon catabolite repression in *E.coli* lysate spectra using t-distributed stochastic neighbor embedding

When there are multiple sugar sources available, the simultaneous utilization of multiple sugar sources can be hindered in *E.coli*. For example, wild type *E.coli* consumes glucose during the first exponential growth phase, followed by a diauxic lag phase before consuming xylose.[21] This regulatory phenomenon by which the expression of functions for the use of secondary

carbon sources and the activities of the corresponding enzymes are reduced in the presence of a preferred carbon source is known as carbon catabolite repression (CCR).[22]

The unsupervised ML algorithm t-distributed stochastic neighbor embedding (tSNE) was employed to compare similar data points in lower dimensional space. The resulting tSNE plots demonstrated clear differences in spectral data that correlated with nutrient exposure concentration, providing preliminary validation of our hypothesis that the observed differences in metabolic responses in the cell lysate are reflected in spectral data, rather than a result of algorithm training. Specifically, at 0.5 h, *E.coli* cells grown in media containing 0.1% glucose consumed glucose to support growth, while cells cultured in media containing 0.05% glucose and 0.05% xylose also consumed glucose to grow. When the only nutrient source was xylose, cell didn't exhibit growth. The tSNE plot revealed that the control and xylose clusters were more similar since they didn't exhibit growth, whereas the glucose and the mixture containing glucose and xylose clusters were more alike since they both consumed glucose to support growth.

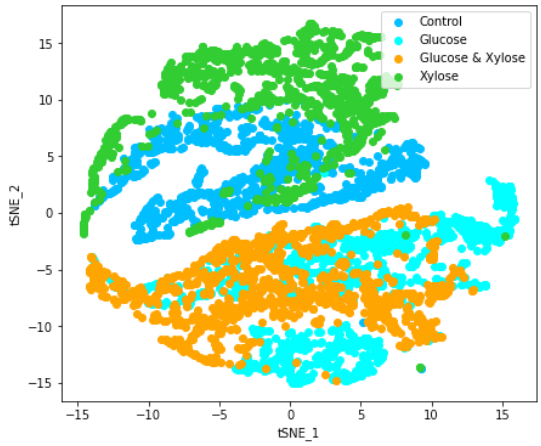

Figure 4.6 tSNE clustering analysis for SERS spectra of *E. coli* lysate after culturing in different nutrient sources, glucose, xylose, a mixture containing glucose and xylose, and a control without a carbon source, respectively.

4.5 Methods

4.5.1 SERS Sensor fabrication

SERS surfaces are fabricated in microfluidic channels with a capacitor architecture to apply an AC potential across electrodes. Fabrication is performed silicon substrates (NOVA Electronic Materials, P-type, boron doped <100> with resistivity of 0.001-0.005 ohm-cm) with dimensions of 15 mm × 15 mm that are spin coated with poly(styrene-b-methyl methacrylate) (PS-b-PMMA, Mn S-b-MMA 170000-b-145000 g mol^{-1}) thin films of approximate thickness of 25 nm; Si substrates serve as the working electrode. Indium tin oxide (ITO) coated glass slides (Delta Technologies) serve as the counter electrode.

Silicon substrates were cleaned by 20% v/v hydrofluoric acid (HF, Fisher Scientific, 48%) / deionized (DI) water (Milli-Q Millipore System, 18.2 MΩ cm^{-1}) for 5 minutes to remove the native oxide layer and then immersed in DI water to regrow a thin oxide layer. *The potential of HF to cause severe injury mandates extreme caution during usage*. Random copolymer poly(styrene-co-methyl-methacrylate)-α-hydroxyl- ω-Tempo moiety (PS-r-PMMA, Polymer Source, Mn =7400, Mw =11800, Mw /Mn =1.60, 59.6 mol% Polystyrene content) random copolymer dissolved in toluene (Fisher Scientific), 1 wt%, was spin-coated at 3000 rpm for 45 seconds on silicon substrates. PS-r-PMMA films were annealed under vacuum at 170 °C for 48 hours followed by a rinse with toluene to leave a brush layer. PS-b-PMMA is spin coated at 5000 rpm for 45 seconds and then annealed for 72 hours at 170 °C. In order to selectively functionalize PMMA domains on PS-b-PMMA diblock copolymer films with amine functional groups for crosslinking with Au NPs, PS-b-PMMA/Si were immersed in dimethyl sulfoxide (DMSO, Sigma Aldrich) for 5 min and then 5 % vol ethylenediamine (ED, Sigma Aldrich) in DMSO for another 5 min. ITO counter electrodes were cleaned by ethanol (Sigma Aldrich),

isopropyl alcohol, and DI water and then dried by N_2 before attaching a platinum wire and silver paste (Epoxy Technology) to make electrical contact.

A microfluidic cell was formed between electrodes using a 90 μm spacer layer composed of 3M 9816L. A solution of 2 μL N-hydroxysulfosuccinimide (s-NHS, Sigma Aldrich), 20mM, and 2 μL 1-Ethyl-3-(3-dimethyl aminopropyl) carbodiimide (EDC, Sigma Aldrich), 8 mM, in a 2-(Nmorpholino) ethane sulfonic acid buffer (MES, Sigma Aldrich, 0.1 M, pH=4.7) added to a 0.25 mL solution of 2.6 nM lipoic acid functionalized Au NP solution. 20 μL of the solution containing Au NP, s-NHS, and EDC is added to the microfluidic cell. An AC electrical stimuli with potential of 5 V_p and frequency of 100 Hz is applied for 2 min to deposit a seed layer to induce EHD flow. The second deposition step was conducted at a potential of 5 V_p and frequency of 1000 Hz for 2 min to grow Au NP clusters. After deposition, the electrode cell was dismantled and the sensor surface was thoroughly rinsed with DI water and isopropyl alcohol (IPA, Sigma Aldrich) and then dried with N_2.

4.5.2 SERS spectroscopy

All SERS spectroscopy measurements are conducted using a confocal Renishaw InVia micro Raman system with a 785 nm diode laser, a laser power of 14 μW, an exposure time of 0.5 s, and a 60× water immersion objective with a 1.2 numerical aperture. Bacteria cell lysate is used as the immersion media. After soaking the SERS substrate in the sample for 15 min, Raman maps are collected with a spacing of 4 μm spacing between points. For each sample one 20 × 20 pixel Raman map is acquired.

4.5.3 Bacterial culture preparation

Pseudomonas aeruginosa (strain PA14 wild type) and *Escherichia coli* (strain MC4100, K-12) cultures were revived by streaking from a frozen culture stock onto LB Lennox agar (IBI Scientific) plates and incubated at 37 °C for 24 h. Individual colonies from these plates were used to inoculate solutions of 100 mL of LB in triplicate, which were subsequently grown for 18 h at 37 °C and shaking at 230 rpm. The 18 h cultures were centrifuged at 5000 rpm for 5 min, then re-suspended in fresh LB at an optical density at 600 nm (OD600) of 0.50 as measured by a BioChrom Colourwave CO7500 colorimeter.

After subculturing the bacteria, they were washed twice with PBS. For carbon starvation experiments, the washed cells were resuspended in 5 mL of M63 medium supplemented with 1% glucose (w/v), 1% sucrose (w/v), or deionized water to reach an OD of 0.5. These cultures were then incubated in a shaking incubator at 300 rpm and 37℃ for 2 hours. For carbon catabolite repression experiments, the washed cells were resuspended in 5 mL of M63 medium supplemented with 0.1% glucose, 0.1% xylose, or 0.05% of glucose/xylose mixture to reach an OD of 0.2. Each carbon source profile included three test tubes per colony. These cultures were also incubated in a shaking incubator at 300 rpm and 37℃. Samples were collected at 0.5 h after the start of growth. All the samples collected came from their individual tubes, ensuring that there was no reduction in cell culture volume that could affect the cells. After 0.5 h, the cells were washed twice with PBS and centrifuged at 5000 rpm for 5 minutes. The resulting pellet was resuspended in 100 uL of deionized water and heated at 97℃ for 30 minutes to lyse the cells and release their contents into the water. After 30 minutes, the culture was centrifuged at 12000 rpm for 10 minutes, and 75 μL of supernatant were collected into three separate centrifuge tubes, each containing 25 μL of supernatant. The samples were then stored at -20 ℃ for further analysis.

4.5.4 Pre-processing of SERS spectra data

For data preprocessing, asymmetric least square correction is utilized for baseline correction, and a Savitzky–Golay filter is used for data smoothing. In order to normalize the data, the vibrational band of silicon at 520 cm^{-1} is used as an internal standard and set to 1. Principal component analysis (PCA) was performed for dimensional reduction. We determined that 17 PCA components captured 90% of variant of the dataset. t-distributed stochastic neighbor embedding (tSNE) was also performed to visualize the concentration data in lower dimensional space and show there are spectral differences in the data observed without labeling data for algorithms.

4.5.5 Antimicrobial susceptibility test feed-forward deep neural network

The four datasets used in these experiments (*E. coli* temporal, *E. coli* dose, *P. aeruginosa* temporal, and *P. aeruginosa* dose) were all pre-processed using the spectra pre-processing method. 800 spectra from each condition are used to train the model. The DNNs were trained using 10-fold cross validation. For each cross fold the data was divided into a training, validation, and test set. As a supervised machine learning method, training data is used to fit the parameters of the DNN; knowledge of the class that each point of training data belongs to is necessary. The validation set for each cross fold is unique and is a random 20% subset of the training set of that cross fold. The DNNs used in this study are feed-forward layered networks with fully connected layers and a logistic output for the two-class models or a softmax output for the five-class models. We used the python hyperparameter optimization library SHERPA in combination with 10-fold cross validation to select the best hyperparameters and architecture based on mean cross validation accuracy.[16] All of the models utilize the Adam optimizer with a learning rate of 0.001 and a batch size of 20. The DNN architectures used have 1−3 hidden

layers with 10−500 neurons per layer, and rectified linear unit (ReLU) activation functions. The loss function used for the two-class models was binary cross entropy and the loss function used for the five-class models was categorical cross entropy. Due to the relatively small number of samples compared to the very large number of features, we use Scikit-learn to perform principal component analysis (PCA) with an explained variance of 90% for dimensionality reduction. In order to avoid overfitting, we used early stopping in combination with data augmentation. Early stopping is a technique used to halt model training when the loss of the validation set starts to increase compared to the training loss, indicating overfitting. For our experiments, we used a patience of 10. Additionally, in some of the networks we utilized dropout and L2 regularization to combat overfitting. For data augmentation we used the Synthetic Minority Oversampling Technique (SMOTE) to combat the problem of class imbalance.[17] Although the original dataset does not have significant class imbalance between the five classes, when we split the data into two groups for our two-class models we introduced class imbalance. For example, one of the two-class models is 0 min, 5 min, 10 min, 20 min as one class (3200 data points) and 40 min as one class (800 data points). SMOTE augments the data by synthesizing training examples in the minority class. Specifically, SMOTE chooses a random data point from the minority class and a random neighbor from its five nearest neighbors, and then a synthetic example is created at a randomly selected point between the two examples in feature space. Data augmentation was only performed on the training set; the test set was not augmented. To measure the performance of each of the models, we used 10-fold mean cross validation accuracy.

4.5.6 Support vector machine classification model

A support vector machine (SVM) discriminative model was trained on SERS spectra of *E. coli* lysate after culturing in different nutrient sources, glucose, sucrose, and a control without

a carbon source. The SVM models are trained using 17 PCA components. A holdout set is composed of 20% of the data that is used for final validation and not seen at all during training. The model is trained with the remaining 80% of the spectral data labeled with their appropriate class to define a hyperplane separating data into the correct classes. SVM models are trained with Scikit-learn using default parameters, with radial basis function (RBF) kernel, Margin parameter (C) =1, and γ=scale.

4.6 Conclusion

We have demonstrated that the response of *P. aeruginosa* and *E. coli* bacterial communities to antibiotics is rapidly detected in SERS spectral data when using sensor surfaces with controlled nanogap spacing and chemistry. Deep learning analysis of SERS data is able to differentiate the response of untreated cells from those exposed to antibiotics in 10 min post exposure with greater than 99% accuracy and temporally follow the evolution with 5 min resolution with greater than 99% accuracy, significantly faster and more accurate than current SERS AST methods.[26–28] The bacterial response to varying antibiotic doses is differentiated with greater than 96% accuracy from untreated bacteria, even when treated with antibiotic dosages up to 10-fold lower than the minimum inhibitory concentration observed in conventional growth assays. The high classification accuracy/sensitivity and specificity in following temporal response of monocultures and differentiating susceptibility and resistance demonstrates the SERS + deep learning approach described here making this method a promising candidate for use in clinical samples for rapid AST. We also found that the metabolism features captured by the unsupervised PCA model can correlate with energy nucleotides and metabolic pathways involved in nutrient uptake, suggesting that changes in nucleotide concentrations are consistent with the nutrient-dependent metabolic signatures observed. The SVM classification model

achieved high accuracy on differentiating different nutrient metabolism profile with the false prediction rate of control spectra being classified as nutrient conditions was less than 0.6%. For CCR, the tSNE plot revealed that the control and xylose clusters were more similar since they didn't exhibit growth, whereas the glucose and the mixture containing glucose and xylose clusters were more alike since they both consumed glucose to support growth.

References

(1) Lambert, G.; Kussell, E. Memory and Fitness Optimization of Bacteria under Fluctuating Environments. *PLoS Genet.* **2014**, *10* (9), e1004556. https://doi.org/10.1371/journal.pgen.1004556.

(2) Cashel, M.; Gallant, J. Two Compounds Implicated in the Function of the RC Gene of Escherichia Coli. *Nature* **1969**, *221* (5183), 838–841. https://doi.org/10.1038/221838a0.

(3) Belenky, P.; Ye, J. D.; Porter, C. B. M.; Cohen, N. R.; Lobritz, M. A.; Ferrante, T.; Jain, S.; Korry, B. J.; Schwarz, E. G.; Walker, G. C.; Collins, J. J. Bactericidal Antibiotics Induce Toxic Metabolic Perturbations That Lead to Cellular Damage. *Cell Rep.* **2015**, *13* (5), 968–980. https://doi.org/10.1016/j.celrep.2015.09.059.

(4) Thrift, W. J.; Ronaghi, S.; Samad, M.; Wei, H.; Nguyen, D. G.; Cabuslay, A. S.; Groome, C. E.; Santiago, P. J.; Baldi, P.; Hochbaum, A. I.; Ragan, R. Deep Learning Analysis of Vibrational Spectra of Bacterial Lysate for Rapid Antimicrobial Susceptibility Testing. *ACS Nano* **2020**, *14* (11), 15336–15348. https://doi.org/10.1021/acsnano.0c05693.

(5) Pulido, M. R.; García-Quintanilla, M.; Martín-Peña, R.; Cisneros, J. M.; McConnell, M. J. Progress on the Development of Rapid Methods for Antimicrobial Susceptibility Testing. *J. Antimicrob. Chemother.* **2013**, *68* (12), 2710–2717. https://doi.org/10.1093/jac/dkt253.

(6) Fleming-Dutra, K. E.; Hersh, A. L.; Shapiro, D. J.; Bartoces, M.; Enns, E. A.; File, T. M., Jr; Finkelstein, J. A.; Gerber, J. S.; Hyun, D. Y.; Linder, J. A.; Lynfield, R.; Margolis, D. J.; May, L. S.; Merenstein, D.; Metlay, J. P.; Newland, J. G.; Piccirillo, J. F.; Roberts, R. M.; Sanchez, G. V.; Suda, K. J.; Thomas, A.; Woo, T. M.; Zetts, R. M.; Hicks, L. A. Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011. *JAMA* **2016**, *315* (17), 1864–1873. https://doi.org/10.1001/jama.2016.4151.

(7) Rochford, C.; Sridhar, D.; Woods, N.; Saleh, Z.; Hartenstein, L.; Ahlawat, H.; Whiting, E.; Dybul, M.; Cars, O.; Goosby, E.; Cassels, A.; Velasquez, G.; Hoffman, S.; Baris, E.; Wadsworth, J.; Gyansa-Lutterodt, M.; Davies, S. Global Governance of Antimicrobial Resistance. *Lancet Lond. Engl.* **2018**, *391* (10134), 1976–1978. https://doi.org/10.1016/S0140-6736(18)31117-6.

(8) Sugden, R.; Kelly, R.; Davies, S. Combatting Antimicrobial Resistance Globally. *Nat. Microbiol.* **2016**, *1* (10), 1–2. https://doi.org/10.1038/nmicrobiol.2016.187.

(9) *Tackling drug-resistant infections globally : final report and recommendations / the Review on Antimicrobial Resistance chaired by Jim O'Neill.* Wellcome Collection. https://wellcomecollection.org/works/thvwsuba (accessed 2022-12-08).

(10) van Belkum, A.; Bachmann, T. T.; Lüdke, G.; Lisby, J. G.; Kahlmeter, G.; Mohess, A.; Becker, K.; Hays, J. P.; Woodford, N.; Mitsakakis, K.; Moran-Gilad, J.; Vila, J.; Peter, H.; Rex, J. H.; Dunne, W. M.; JPIAMR AMR-RDT Working Group on Antimicrobial Resistance and Rapid Diagnostic Testing. Developmental Roadmap for Antimicrobial Susceptibility Testing Systems. *Nat. Rev. Microbiol.* **2019**, *17* (1), 51–62. https://doi.org/10.1038/s41579-018-0098-9.
(11) *Bactericidal Antibiotics Induce Toxic Metabolic Perturbations that Lead to Cellular Damage - PubMed*. https://pubmed.ncbi.nlm.nih.gov/26565910/ (accessed 2022-12-08).
(12) Rowan, A. D.; Cabral, D. J.; Belenky, P. Bactericidal Antibiotics Induce Programmed Metabolic Toxicity. *Microb. Cell Graz Austria* **2016**, *3* (4), 178–180. https://doi.org/10.15698/mic2016.04.493.
(13) Dwyer, D. J.; Collins, J. J.; Walker, G. C. Unraveling the Physiological Complexities of Antibiotic Lethality. *Annu. Rev. Pharmacol. Toxicol.* **2015**, *55*, 313–332. https://doi.org/10.1146/annurev-pharmtox-010814-124712.
(14) Brynildsen, M. P.; Winkler, J. A.; Spina, C. S.; MacDonald, I. C.; Collins, J. J. Potentiating Antibacterial Activity by Predictably Enhancing Endogenous Microbial ROS Production. *Nat. Biotechnol.* **2013**, *31* (2), 160–165. https://doi.org/10.1038/nbt.2458.
(15) Chiu, S. W.-Y.; Cheng, H.-W.; Chen, Z.-X.; Wang, H.-H.; Lai, M.-Y.; Wang, J.-K.; Wang, Y.-L. Quantification of Biomolecules Responsible for Biomarkers in the Surface-Enhanced Raman Spectra of Bacteria Using Liquid Chromatography-Mass Spectrometry. *Phys. Chem. Chem. Phys.* **2018**, *20* (12), 8032–8041. https://doi.org/10.1039/C7CP07103E.
(16) Hertel, L.; Collado, J.; Sadowski, P.; Ott, J.; Baldi, P. Sherpa: Robust Hyperparameter Optimization for Machine Learning. *SoftwareX* **2020**, *12*, 100591. https://doi.org/10.1016/j.softx.2020.100591.
(17) Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. https://doi.org/10.1613/jair.953.
(18) Monk, J. M.; Koza, A.; Campodonico, M. A.; Machado, D.; Seoane, J. M.; Palsson, B. O.; Herrgård, M. J.; Feist, A. M. Multi-Omics Quantification of Species Variation of Escherichia Coli Links Molecular Features with Strain Phenotypes. *Cell Syst.* **2016**, *3* (3), 238-251.e12. https://doi.org/10.1016/j.cels.2016.08.013.
(19) Desai, T. A.; Rao, C. V. Regulation of Arabinose and Xylose Metabolism in Escherichia Coli. *Appl. Environ. Microbiol.* **2010**, *76* (5), 1524–1532. https://doi.org/10.1128/AEM.01970-09.
(20) Mohamed, E. T.; Mundhada, H.; Landberg, J.; Cann, I.; Mackie, R. I.; Nielsen, A. T.; Herrgård, M. J.; Feist, A. M. Generation of an E. Coli Platform Strain for Improved Sucrose Utilization Using Adaptive Laboratory Evolution. *Microb. Cell Factories* **2019**, *18* (1), 116. https://doi.org/10.1186/s12934-019-1165-2.
(21) Solopova, A.; van Gestel, J.; Weissing, F. J.; Bachmann, H.; Teusink, B.; Kok, J.; Kuipers, O. P. Bet-Hedging during Bacterial Diauxic Shift. *Proc. Natl. Acad. Sci.* **2014**, *111* (20), 7427–7432. https://doi.org/10.1073/pnas.1320063111.
(22) Görke, B.; Stülke, J. Carbon Catabolite Repression in Bacteria: Many Ways to Make the Most out of Nutrients. *Nat. Rev. Microbiol.* **2008**, *6* (8), 613–624. https://doi.org/10.1038/nrmicro1932.
(23) Ludin, K. M.; Hilti, N.; Schweingruber, M. E. Schizosaccharomyces Pombe Rds1, an Adenine-Repressible Gene Regulated by Glucose, Ammonium, Phosphate, Carbon Dioxide

and Temperature. *Mol. Gen. Genet. MGG* **1995**, *248* (4), 439–445. https://doi.org/10.1007/BF02191644.
(24) Báez-Viveros, J. L.; Osuna, J.; Hernández-Chávez, G.; Soberón, X.; Bolívar, F.; Gosset, G. Metabolic Engineering and Protein Directed Evolution Increase the Yield of L-Phenylalanine Synthesized from Glucose in Escherichia Coli. *Biotechnol. Bioeng.* **2004**, *87* (4), 516–524. https://doi.org/10.1002/bit.20159.
(25) Peterkofsky, A.; Gazdar, C. Glucose and the Metabolism of Adenosine 3′:5′-Cyclic Monophosphate in Escherichia Coli. *Proc. Natl. Acad. Sci.* **1971**, *68* (11), 2794–2798. https://doi.org/10.1073/pnas.68.11.2794.
(26) Han, Y.-Y.; Lin, Y.-C.; Cheng, W.-C.; Lin, Y.-T.; Teng, L.-J.; Wang, J.-K.; Wang, Y.-L. Rapid Antibiotic Susceptibility Testing of Bacteria from Patients' Blood via Assaying Bacterial Metabolic Response with Surface-Enhanced Raman Spectroscopy. *Sci. Rep.* **2020**, *10*, 12538. https://doi.org/10.1038/s41598-020-68855-w.
(27) Kirchhoff, J.; Glaser, U.; Bohnert, J. A.; Pletz, M. W.; Popp, J.; Neugebauer, U. Simple Ciprofloxacin Resistance Test and Determination of Minimal Inhibitory Concentration within 2 h Using Raman Spectroscopy. *Anal. Chem.* **2018**, *90* (3), 1811–1818. https://doi.org/10.1021/acs.analchem.7b03800.
(28) Chang, K.-W.; Cheng, H.-W.; Shiue, J.; Wang, J.-K.; Wang, Y.-L.; Huang, N.-T. Antibiotic Susceptibility Test with Surface-Enhanced Raman Scattering in a Microfluidic System. *Anal. Chem.* **2019**, *91* (17), 10988–10995. https://doi.org/10.1021/acs.analchem.9b01027.

Chapter 5

Decoding the metabolic response of *Escherichia coli* for sensing trace heavy metals in water

5.1 Introduction

Heavy metal contamination from natural and anthropogenic sources is a serious threat to human and ecosystem health, and heavy metal use in a wide variety of industrial and agricultural processes is growing exponentially.[1,2] Contaminated water is a major source of exposure leading to toxic heavy metal accumulation in humans, plants, and livestock. The development of portable and low-cost sensors which can be broadly deployed to locally and frequently monitor the quality of drinking and irrigation water, agricultural and industrial runoff is needed to safeguard sensitive ecosystems and human health. Arsenic, cadmium, chromium, copper, lead, and mercury rank among the priority metals of public health significance.[1] Currently, monitoring water quality typically requires samples to be sent to specifically certified laboratories for inductively coupled plasma-mass spectrometry (ICP-MS) analysis for quantification[3] to determine if contaminants are below safety guidelines set by the World Health Organization (WHO)[4] or regulatory agencies. Other laboratory methods with the necessary limit of detection (LOD) and dynamic range rely on similarly sophisticated and centralized analytical instruments, such as atomic absorption, X-ray fluorescence, or atomic emission spectrometries.[3]

Alternatively, biosensors, using physicochemical signal transduction, such as optical, electrochemical, piezoelectric, and thermal signal outputs, represent low-cost solutions that are compatible for integration in portable systems to detect heavy metal ions. Molecular recognition

labels include enzymes,[5] antibodies,[6] whole cells,[7] aptamers,[8] molecularly imprinted polymers,[9,10] and DNA.[11] Encapsulation of enzymes in hydrogels yields sensors with a LOD needed for monitoring water quality, but they have limited shelf life.[5] Aptamers, on the other hand, exhibit high specificity and stability but are not easily engineered to detect a variety of analytes. Antibodies, relying on the formation of metal–chelated complexes, are versatile sensing elements, yet cross reactivity with other ions lead to lack of specificity.[12] Whole cell-based biosensors rely on mature cell culturing technology and can be incorporated in a range of physicochemical sensor platforms for multiple assays. Whole cell biosensors have received increasing attention as an ultra-sensitive means of detecting hazardous contaminants as they can be engineered to be responsive to different toxins.[13]

Many cellular metabolites have high Raman cross-sections,[14] which can be detected in surface enhanced Raman scattering (SERS) measurements.[15,16] SERS is a highly sensitive, and label free detection scheme,[17] which offers single molecule LOD when using carefully designed nanoarchitectures.[18–20] Indeed, SERS signals from Au decorated nanofiber probes inserted into breast cancer cells have been shown to detect toxic metal exposure at a LOD of 5 nM for mercury and 100 nM for silver.[21] Obtaining reproducible response in biosensors is a longstanding challenge.[22] In particular, the reproducibility of SERS surfaces depends on nanoparticle morphology, nanogap distance and surface chemistry.[23] Our previously demonstrated chemically assembled SERS surfaces composed of spherical nanoparticles with controlled nanogap spacing of 0.9 nm and chemistry exhibit reproducible billion-fold signal enhancements over areas of 1 cm^2; surfaces are able to detect metabolites from bacterial communities on a time scales of minutes[15,24] and accurately quantify analyte concentrations down to 10 fM when using machine learning analysis of spectral data.[18] In this work, the

sensitivity of the *Escherichia coli (E. coli)* stress response is used to transduce the signal of Cr^{6+} and As^{3+} ions into chemical signals that are detected with chemically assembled SERS surfaces. Arsenite is one of the most common toxic valence states (III) of As and high arsenite concentrations are indicators of phytoplankton bloom, high microbial populations, and pollution from mining activity.[25] Cr pollution is largely related to industrial applications in the field of energy production, manufacturing of metals and chemicals, and subsequent waste and wastewater management.[26] Cr^{6+} is much more toxic than Cr^{3+}.[4] A support vector machine (SVM) model achieves higher than 97% classification accuracy for decoding *E. coli* stress response to different concentrations of metal ions for concentrations as low as 68 pM for Cr^{6+} and 5 pM for As^{3+}. Due to their distinct mechanisms of toxicity in bacteria, this sensing platform also distinguishes the metabolic response of As^{3+} and Cr^{6+} with high accuracy when analyzed with SVM models. In addition, convolutional neural networks (CNN) show sensitive and quantitative determination of concentrations across a dynamic range of 0.68 pM - 68 μM for Cr^{6+} and 5 fM - 5 mM for As^{3+}, well below WHO recommended limits of 10 μg/L for As^{3+} and 50 μg/L for Cr^{6+}, respectively.[4] At the lowest concentrations investigated, the metabolic response is detectable when the ratio of metal ions to bacterium in solution is 0.6 for As^{3+} and 8.2 for Cr^{6+}. Finally, by using a pretrained model for analysis of previously unseen tap water and wastewater samples spiked with As^{3+}, SERS detection and ML analysis requires only 80 spectra per class (40 sec total acquisition time) to achieve greater than 96.5% accuracy for classifying concentrations above or below the WHO recommended limit.

5.2 Biochemical signal transduction of metal ions into vibrational spectra

The inherent metabolic stress response of *E. coli* cultures is used to transduce the presence of heavy metal ions in water into metabolites. We then fingerprint the metabolic

response with a combination of SERS detection and ML analysis (SERS +ML). *E. coli* cultures were exposed to Cr^{6+} or As^{3+} ions ($K_2Cr_2O_7$ or $NaAsO_2$) in minimal media for 2 hr (Fig. 1a). Metabolites from the cells were extracted by thermal lysis, and the lysate was deposited on SERS surfaces composed of Au nanoparticle (NP) clusters for spectral data acquisition (Fig. 1b-c). SERS surfaces were fabricated in microfluidic channels with electrodes in a capacitor architecture to achieve reproducible billion-fold signal enhancements (Fig. 5.1e-f).[27] SERS spectra of control samples prepared under the same conditions without Cr^{6+} or As^{3+} in the exposure medium were used to determine the limit of blank (LOB).[28] The full concentration range of samples was collected over the course of several experiments. Each subset of concentrations was collected with a control group included which was not exposed to any metal. To avoid training the algorithm to classify based on background fluctuations, inherent biological variation, or manufacturing variations of SERS surfaces, control samples were measured in biological duplicate and on multiple SERS surfaces (see Methods for more details).

The exposure of bacterial cultures to toxic metal ions is expected to result in significant changes in metabolite concentrations. Such metabolic shifts resulting from stress responses often involve differential regulation of nucleotides central to biosynthetic processes within the cell. Metabolic changes in response to antibiotic stress have been reported to be detectable within 30 minutes of exposure by mass spectrometry.[29] Some metabolic stress responses are general, for example those triggered by the sigma factor regulon, RpoS, which can be regulated by proteins dependent on concentrations of the nucleotide adenosine triphosphate (ATP).[30] ATP accumulates in *E. coli* as part of its stress response to antibiotics[31] and ATP-coupled pumps are associated with As^{3+} transport out of cells in response to toxic exposure.[32] Uracil, another nucleotide, is a building block of RNA and thus related to protein translation, and its concentration is closely

correlated with oxidative stress responses in bacteria.[29,33] Another nucleotide, adenine, regulates the cell cycle in bacteria, including cell division and DNA repair, processes modulated in stress conditions.[34] To verify that SERS surfaces are sensitive to these and similarly Raman active metabolites associated with bacterial stress response, SERS spectra of 1 mM aqueous solutions of key nucleotides ATP, uracil and adenine were acquired and representative spectra are shown in Fig. 5.1d.

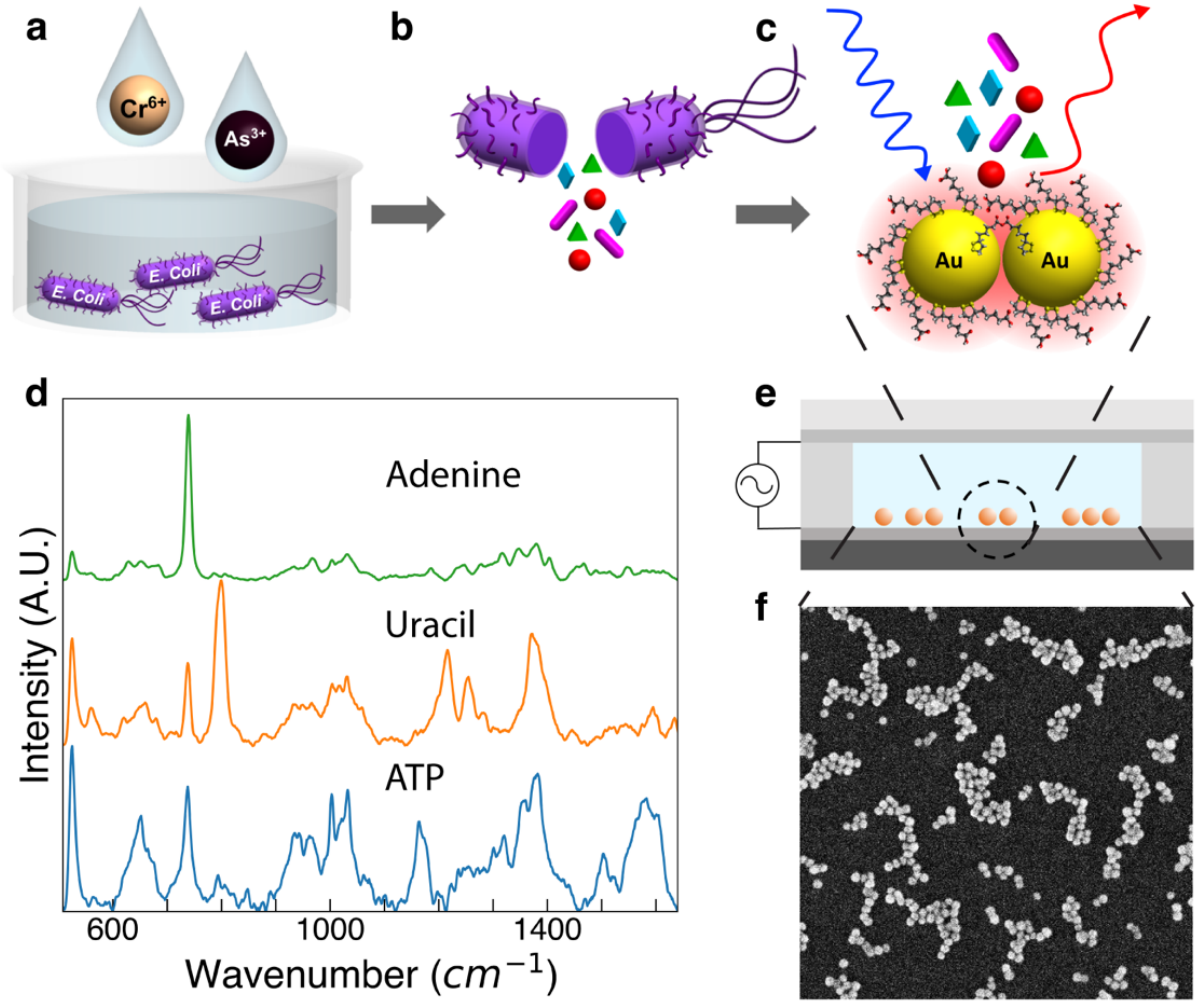

Figure 5.1 Heavy metal detection scheme and SERS spectra of key metabolites. **a** *E. coli* is cultured in growth media supplemented with Cr^{6+} or As^{3+} salts. **b** Cells are thermally lysed, and **c**

lysate supernatant is deposited on SERS surfaces. **d** Representative SERS spectra of key nucleotides involved in bacterial stress responses, ATP, uracil and adenine. **e** Schematic of fabrication of SERS surfaces: a microfluidic cell with AC electric fields across electrodes induces EHD flow to drive lateral assembly and subsequent cross linking reactions between Au NP. **f** Scanning electron microscopy image shows Au NP form close-packed clusters of various sizes. Field of view is 2 μm × 2 μm.

5.3 Training data acquisition for fingerprinting bacterial stress response

SERS spectra were acquired from lysate from *E. coli* cells exposed to heavy metal ion solutions at various concentrations untreated (control). The concentration range investigated with SERS + ML for $NaAsO_2$ was 0.65 pg/L to 650 mg/L (13 concentrations) and for $K_2Cr_2O_7$ was 0.1 ng/L to 10 mg/L (9 concentrations). The corresponding molarities are 5 fM to 5 mM for As^{3+} and 0.68 pM to 68 μM for Cr^{6+}. The concentration range was chosen to span the WHO recommended limit for these metals in drinking water, which are 10 μg/L (0.13 μM) and 50 μg/L (0.96 μM) for As^{3+} and Cr^{6+}, respectively. SERS spectra acquired from pure solutions of Cr^{6+} (6.8 pM) and As^{3+} (0.5 pM) without *E. coli* cells show that the vibrational peaks observed from lysate samples are due to the cellular metabolites instead of heavy metal ions itself (Fig. 5.2).

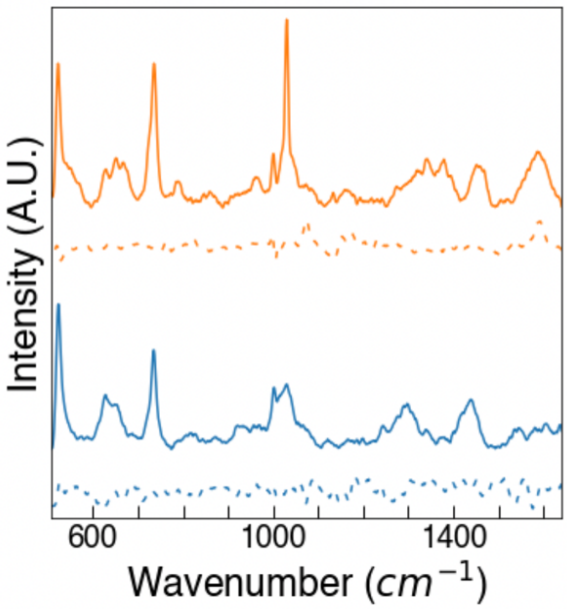

Figure 5.2 Surface enhanced Raman scattering spectra from Cr^{6+} (dotted blue) and As^{3+} (dotted orange) salts dissolved in deionized water at concentrations of 6.8 pM and 0.5 pM, respectively. Spectra from bacteria cultures exposed to metals at the same concentration in deionized water are plotted above for Cr^{6+} (blue) and As^{3+} (orange). All spectra are background subtracted.

Average SERS spectra of *E. coli* lysate after metal ion exposure show spectral feature differences to the eye (Fig. 5.3 a, b). Principal component analysis (PCA), used for dimensional reduction of SERS spectra, more clearly highlights some spectral feature changes unique to samples exposed to each metal or common to exposure to either metal. Before PCA, SERS spectra undergo baseline correction, data smoothing and normalization (see Methods). We found that 22 PCA components, shown in Fig. 5.4, capture 93.3% and 94.8% of variances for Cr^{6+} and As^{3+} concentration data, respectively. In Fig. 5.3 c, d, the first three principal component (PC) loadings, which account for greater than 75% of spectral variance used for sample classification, are shown in a heat map. The scores are plotted in Fig. 5.5. Metabolite vibrational mode

assignments are shown in Table 5.1. The stress response to metal toxins involves differential regulation of nucleotides related to biosynthetic processes within the cell. The largest loading features in PC1, PC2, and PC3 correlate with energy nucleotides, which are associated with energy metabolism pathways involved in toxic metal stress response in bacteria,[35–37] suggesting that changes in nucleotide concentrations in response to metal exposure are consistent with the features upon which the algorithm is classifying the different exposure conditions. Thus, this platform is promising to identify biochemical networks involved in toxin stress response when combined with network models as performed by Yang *et al.* to identify metabolic mechanisms of antibiotic lethality.[38]

Table 5.1 PC1, 2 and 3 loading peak relationship to metabolites for Cr^{6+} and As^{3+}.

	Cr^{6+} Features	Metabolite(s)	As^{3+} Features	Metabolite(s)
PC1	734 cm^{-1}	adenine[44–47]	734 cm^{-1}	adenine[44–47]
	1030 cm^{-1}	phenylalanine[48–50] adenosine[45]	724 cm^{-1}	hypoxanthine[51]
	1036 cm^{-1}	dAMP[45]	1030 cm^{-1}	phenylalanine[48,50] adenosine[45]
	1040 cm^{-1}	uracil[52], ATP[53], thymine[44]	1036 cm^{-1}	dAMP[45]
	1017 cm^{-1}	phenylalanine[54]	1000 cm^{-1}	phenylalanine[55,56]
PC2	1030 cm^{-1}	phenylalanine[48,50] adenosine[45]	734 cm^{-1}	adenine[44–47]
	724 cm^{-1}	hypoxanthine[51]	1269 cm^{-1}	lipid[57]

	1003 cm^{-1}	phenylalanine[58]	607 cm^{-1}	indole[59]
	1010 cm^{-1}	indole[60]	1000 cm^{-1}	phenylalanine[55]
	671 cm^{-1}	cysteine[57,61,62]	1506 cm^{-1}	indole[59]
PC3	734 cm^{-1}	adenine[44–47]	1030 cm^{-1}	phenylalanine[48,50] adenosine[45]
	1030 cm^{-1}	phenylalanine[48,50] adenosine[45]	734 cm^{-1}	adenine[44–47]
	1040 cm-1	uracil[52], ATP[53], thymine[44]	652 cm^{-1}	guanine[63,64]
	671 cm^{-1}	cysteine[57,61,62]	671 cm^{-1}	cysteine[57,61,62]
	641 cm^{-1}	tyrosine[58]	1269 cm^{-1}	lipid[57]

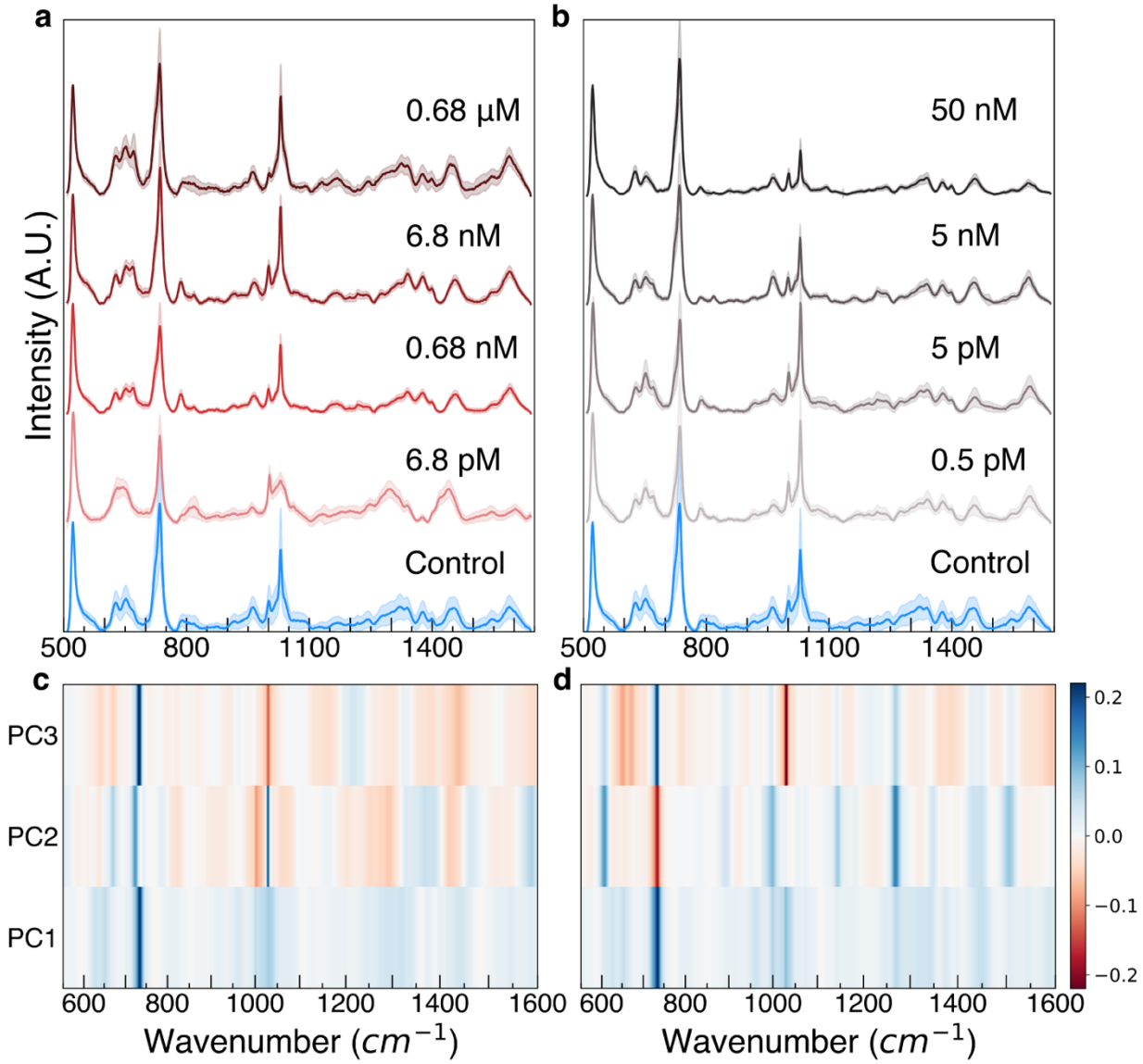

Figure 5.3 Concentration dependent averaged SERS spectra (vertically offset with standard deviation shaded above and below each spectrum) acquired from *E. coli* cultured in media with indicated **a** $K_2Cr_2O_7$ and **b** $NaAsO_2$ concentrations. PC1, 2 and 3 heat map of **c** the Cr^{6+} dataset and **d** the As^{3+} dataset containing spectra of lysate from control and the full range of metal concentration exposure.

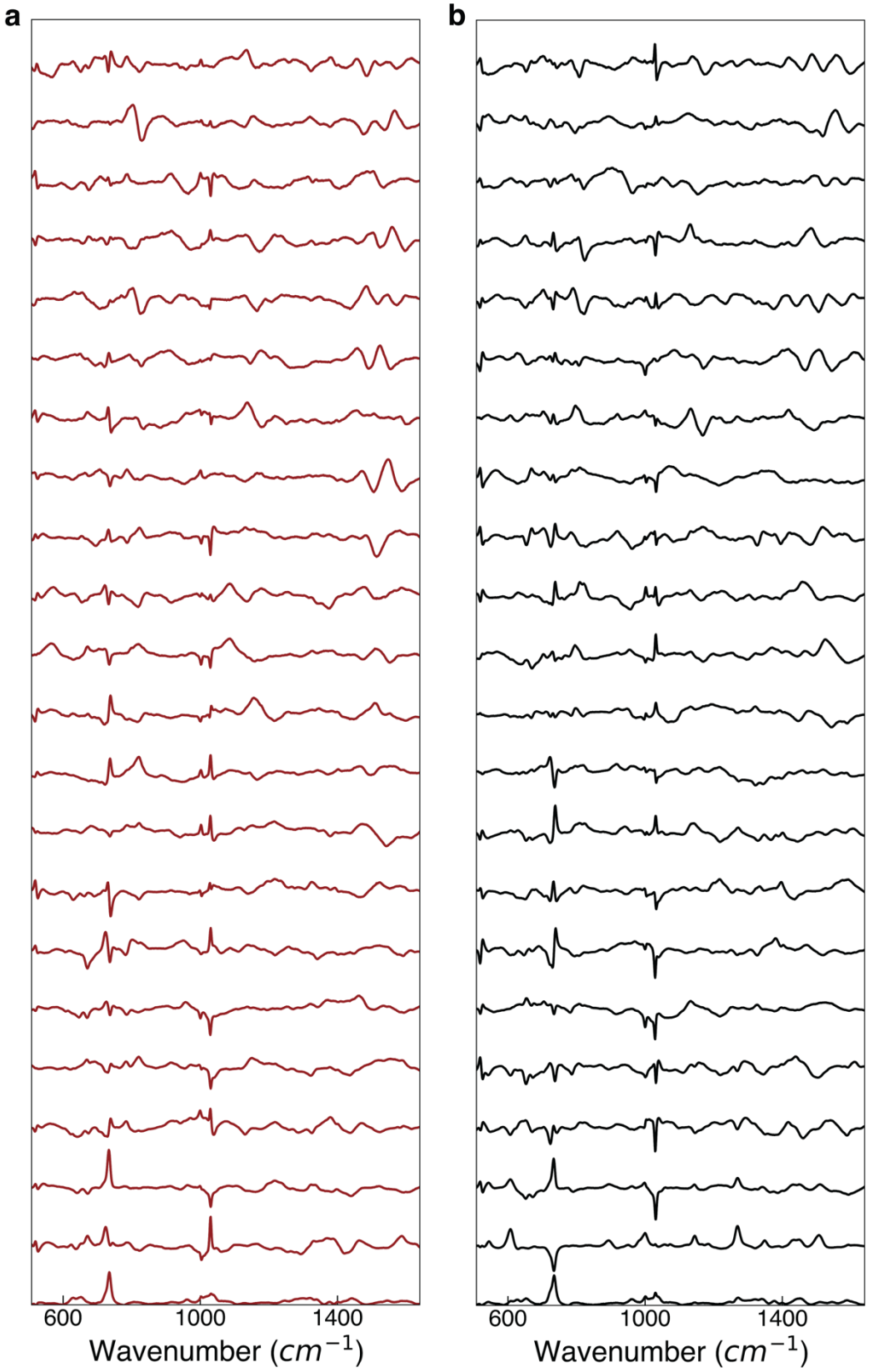

Figure 5.4 PC loadings from **a.** Cr^{6+} and **b.** As^{3+}; bottom to top is PC1 to PC 22.

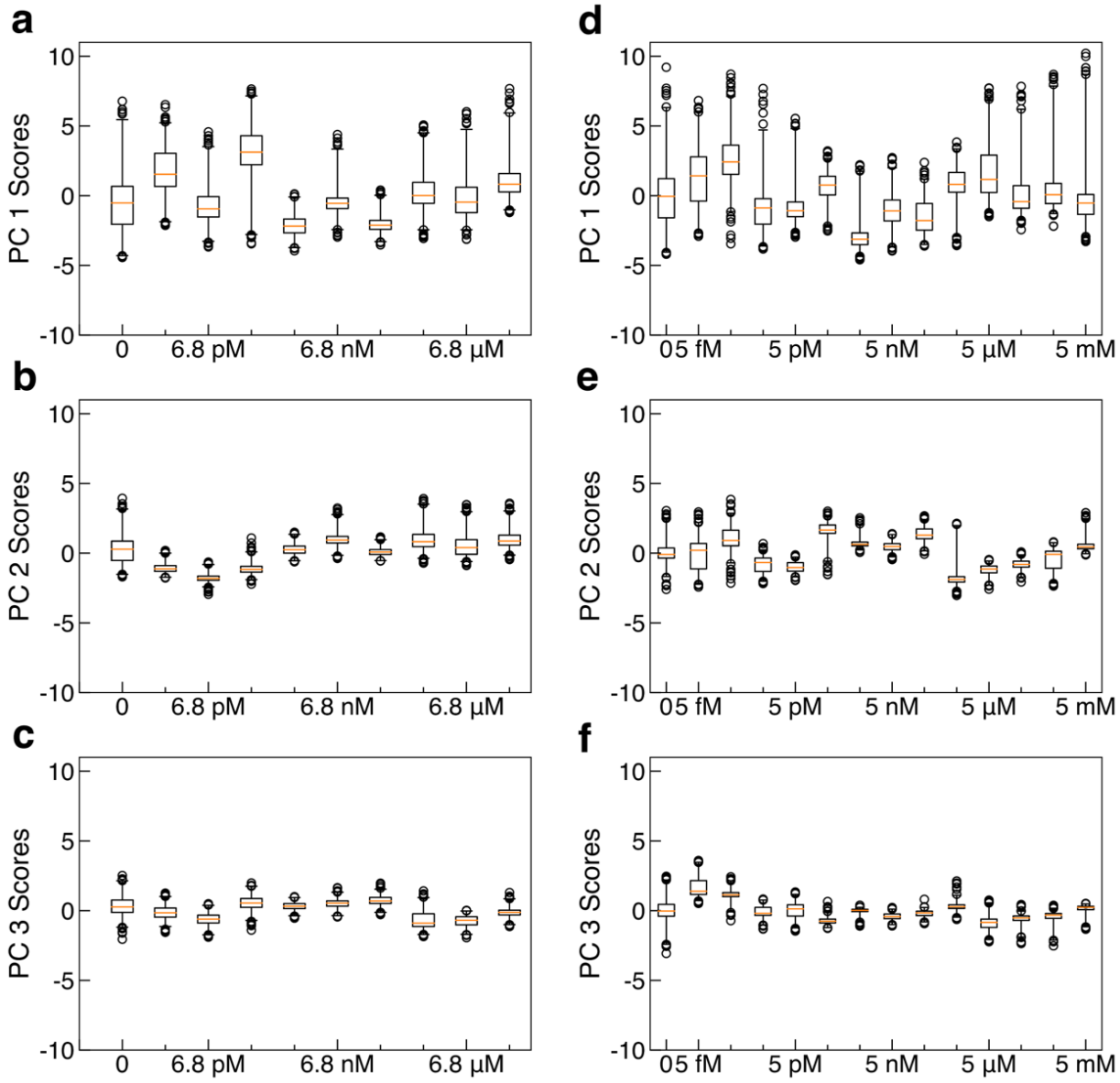

Figure 5.5 Boxplot showing principal component 1, 2 and 3 scores for Cr^{6+} (**a, b, c**) and As^{3+} (**d, e, f**), respectively. Isolation forest was used for 1% outlier removal before plotting the data.

5.4 Classifying Lysate Spectral concentrations by Support Vector Machine

We hypothesized that, while differences in lysate spectra associated with heavy metal exposure might be difficult to identify by eye, ML algorithms could accurately classify these

differences as a function of metal concentration. An unsupervised ML algorithm, t-distributed stochastic neighbor embedding (tSNE), is used for comparing similar data points in lower dimensional space. The tSNE plots show clear differences in the spectral data that correlate with exposure concentration (Fig. 5.6). These plots represent preliminary validation of our hypothesis that the differences in metabolic responses observed in the cell lysate are evident in spectral data and not a result of algorithm training. These components are used as inputs for training two independent support vector machine (SVM) discriminative ML models, one for Cr^{6+} and one for As^{3+}, in order to demonstrate the ability to accurately distinguish different heavy metal exposure concentrations as a means to evaluate water safety. The classes in each discriminative model are the concentrations of metal ions: the model for Cr^{6+} has 10 classes (for 9 metal concentrations + control) and for As^{3+} there are 14 classes (for 13 metal concentrations + control).

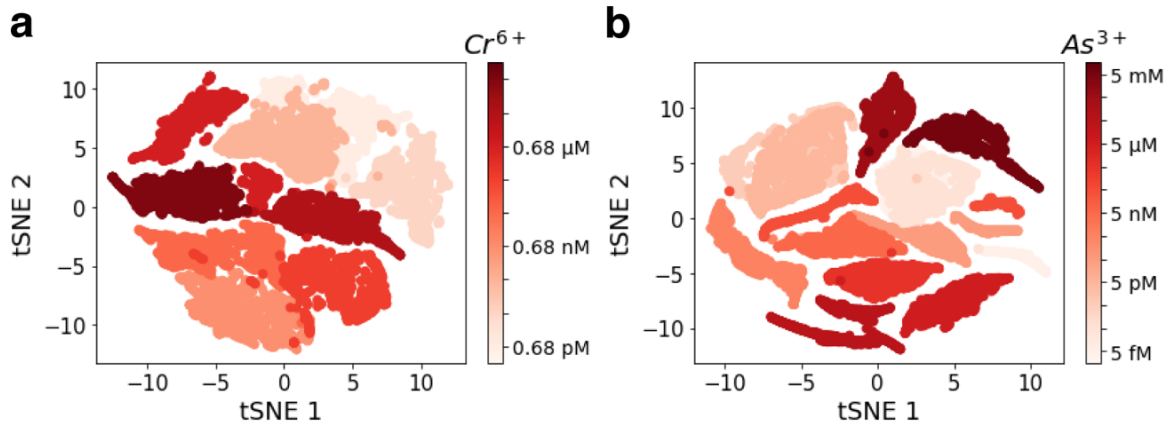

Figure 5.6 t-distributed stochastic neighbor embedding cluster plot of Cr^{6+} (left) and As^{3+} (right) showing different metal exposure concentrations exhibit differential features when using unsupervised algorithms.

The training datasets are imbalanced since the size of the control class (measured in biological duplicate) dataset (9600 spectra) is 8 times larger than the classes corresponding to a

single concentration (1200 spectra). The synthetic minority over-sampling technique (SMOTE) is a standard method to manage imbalanced data sets by performing data augmentation (see Methods).[39] SMOTE is performed after dataset division to prevent data leakage. The model is trained with 80% of the spectral data and the resulting classification accuracy is determined by algorithm predictions on a holdout set (not seen by the SVM model during training) composed of the remaining 20% of the data. The classification accuracy of the holdout set is plotted in the confusion matrices for Cr^{6+} (Fig. 5.7a) and As^{3+} (Fig. 5.7b). The concentration label of Cr^{6+} and As^{3+} datasets is transformed to logarithmic scale. The LOD was determined to be at the value when the prediction accuracy was higher than 98% in distinguishing from the control sample. At concentrations of 6.8 pM for Cr^{6+} and 0.5 pM for As^{3+}, there are less than 0.3% false predictions of control rather than the true concentration (Fig. 5.7a, b). Thus SERS + ML yields a LOD of 6.8 pM for $Cr^{6+}$and 0.5 pM for As^{3+}. The SVM classification model was also evaluated by traditional sensor performance metrics of sensitivity, specificity and accuracy (Table 5.2). Overall, above the LOD, the sensitivity, specificity and accuracy are all higher than 97% for both As^{3+} and Cr^{6+}. In order to put these metrics in perspective, we compare the analysis from SVM models to analysis of the culture optical density (OD) data (Fig. 5.7c, f) used for assessing cell growth and inhibition by stressors. There is no significant difference in culture OD 2 hr after exposure to Cr^{6+} even at concentrations of 340 μM, and there is a significant difference in OD for As^{3+} compared to control only at concentrations greater than 100 μM. At an OD of 0.5, the LOD determined from the SVM model corresponds to approximately 0.6 As^{3+} ions per bacterium in solution and 8.2 Cr^{6+} ions per bacterium in solution. This correlates well with the recommended safe concentration of Cr^{6+} being ten times higher than As^{3+}. Thus, SERS + ML achieves 6 orders of magnitude lower concentration detection versus methods based on growth inhibition.

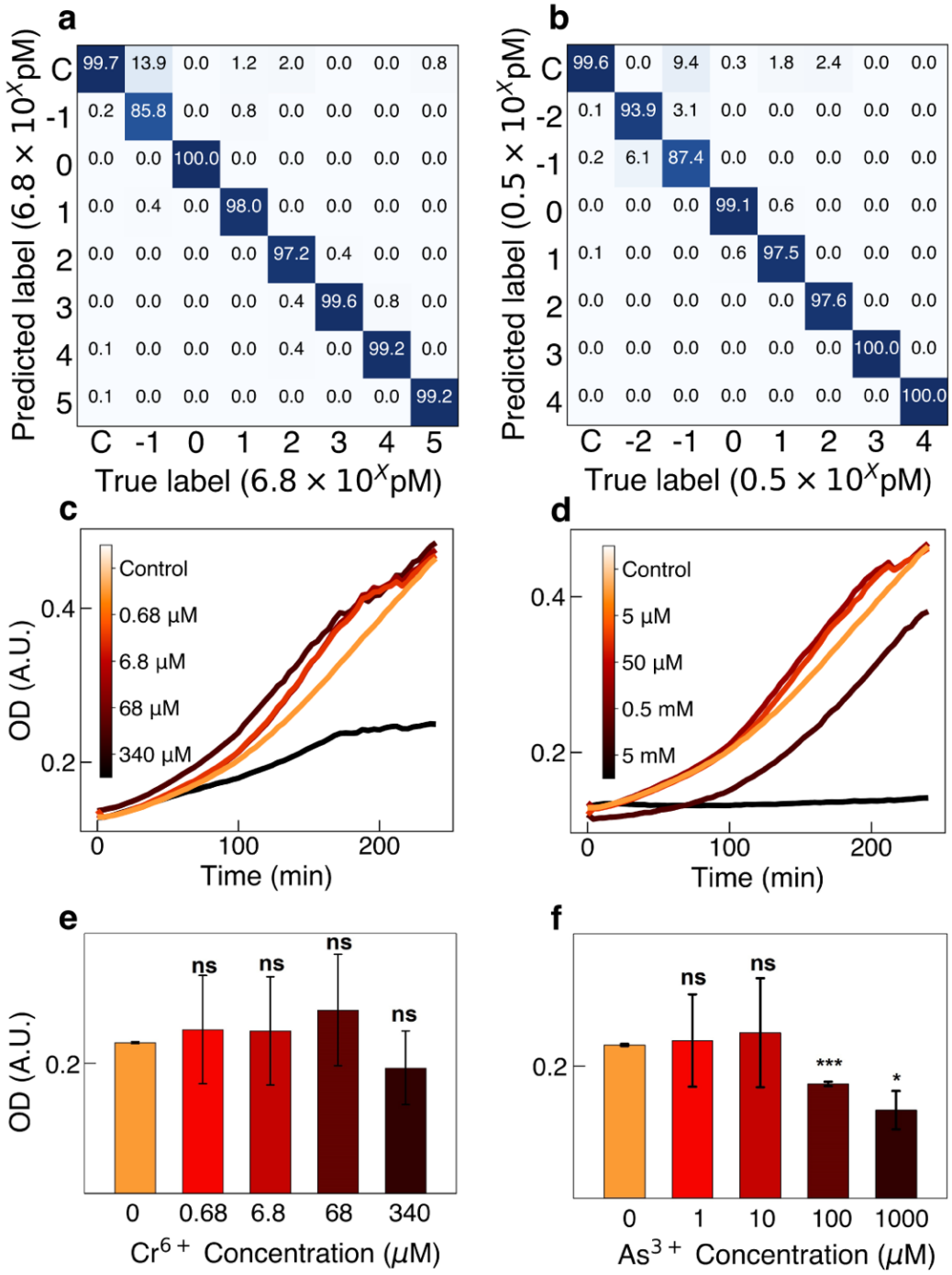

Figure 5.7 Classifying Lysate Spectral concentrations. **a** SVM confusion matrices showing accuracy of classifying of different concentrations of Cr^{6+} (label scale bar is on a log scale in units of 6.8 pM) and **b** As^{3+} (label scale bar is on a log scale in units of 0.5 pM) in the correct concentration class. Growth curve for **c** Cr^{6+} and **d** As^{3+} at different exposure concentrations. Corresponding OD from the growth curves at 2 hr for different concentrations of **e** Cr^{6+} and **f**

As^{3+}, where ns = no significant difference between the experimental groups and control, $* = p \leq 0.05$, $** = p \leq 0.01$ and $*** = p \leq 0.001$. Experiments were done in biological duplicate.

Table 5.2 Calculated metrics of sensor performance from SVM confusion matrices in Figure 5.7.

Cr^{6+}	Sensitivity (%)	Specificity (%)	Accuracy (%)
0.68 pM	85.8	99.86	98.1
6.8 pM	100	100	100
68 pM	98	99.9	99.7
0.68 nM	97.2	99.9	99.6
6.8 nM	99.6	99.8	99.8
68 nM	99.2	99.9	99.8
0.68 μM	99.2	99.99	99.9

As^{3+}	Sensitivity (%)	Specificity (%)	Accuracy (%)
5 fM	93.9	99.5	98.8
50 fM	87.4	99.1	97.6
0.5 pM	99.1	99.9	99.8
5 pM	97.5	99.9	99.6
50 pM	97.6	100	99.7
0.5 nM	100	100	100

5 nM	100	100	100

$$Sensitivity = \frac{True\ Positive}{True\ Positives + False\ Negatives} \quad (1)$$

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \quad (2)$$

$$Accuracy = \frac{True\ Positive + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives} \quad (3)$$

5.5 Classification of Type of Heavy Metal Ion Contaminants

We hypothesized that the metabolic consequences of As^{3+} and Cr^{6+} exposure should be differentiable by SERS + ML of cell lysate due to differences in the mechanism of toxicity of these two metals. A SVM binary classification model was trained on lysate from cells exposed to Cr^{6+} at concentrations in the range of 0.68 pM - 0.68 μM and As^{3+} at concentrations 0.5 pM - 0.5 μM, at 10-fold concentration increments. These ranges span the LOD achieved with SERS + ML for each of the two metals. The algorithm training process follows an analogous flow (baseline correction, smoothing, normalization, data reduction) as described for the classification of concentration in the prior section (see Methods). Using this approach, Cr^{6+} and As^{3+} contamination can be distinguished with a high classification accuracy of 98.8% (Fig. 5.8a). The ability to distinguish between different types of heavy metal ions in water is of great importance for determining pollution source and water treatment process. Analysis of the two metal data sets with tSNE show there are clear differences in spectral data even when the data is not labeled during training (Fig. 5.8b).

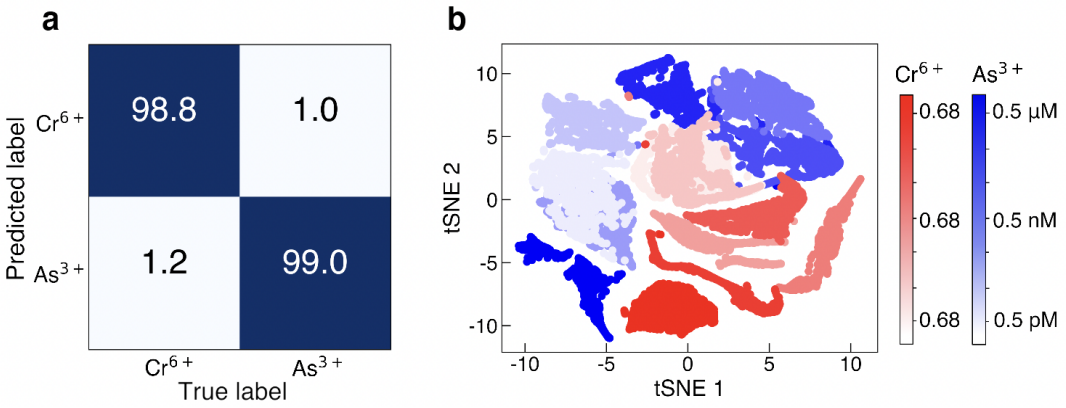

Figure 5.8 Investigation of different types of heavy metal ion contamination. **a** SVM confusion matrix for classification between Cr^{6+} and As^{3+} for concentration range 0.68 pM to 0.68 µM and 0.5 pM to 0.5 µM, respectively. **b** tSNE clustering analysis for different concentrations of Cr^{6+} and As^{3+} in red and blue, respectively.

5.6 Convolutional neural network regression for sensitive quantification heavy metal concentrations

In addition to evaluating how SERS + ML is able to assign a concentration to a particular class (Fig. 5.7), we also demonstrate that algorithms can predict the actual concentration of heavy metal ions in water. Monitoring concentration changes below EPA regulatory and WHO recommended limits is important for early detection of contaminants entering water supplies before adverse effects occur. CNN was used for regression analysis as it outperforms SVM in terms of throughput and regression error.[40] Two independent 1-dimensional (1D) CNN regression models are trained on Cr^{6+} and As^{3+} concentration-dependent cell lysate spectral data. The same 10 and 14 metal concentration classes for Cr^{6+} and As^{3+}, respectively, were used as before (Fig. 5.7). The CNN model architecture (Fig. 5.9a) contains four 1D convolutional layers with inputs of 22 PCA components representing the Cr^{6+} and As^{3+} concentration data. The first convolutional layer has the same padding and a stride of 1 to preserve the spatial dimensions of

the input data. Each convolutional layer uses a rectified linear (ReLU) activation function, and is followed with batch normalization and dropout with 20% random dropout rate to avoid overfitting (see Methods). As before, the spectral data is baseline corrected, smoothed, normalized, and dimensionally reduced using PCA before input into the model. The holdout set for validation is composed of 20% of the data and the remainder is used for training.

First, we use 10-fold cross validation for hyperparameter tuning and model performance evaluation. The number of epochs (training cycles) in the 1D CNN was determined by monitoring the convergence of the training and validation loss. The loss function is calculated to determine the mean square error (MSE) error between the predicted values and the true values. As one can see in the Fig. 5.10 the algorithm converges to a loss value of approximately 0.1 at an epoch of 35. In order to utilize SERS + ML for a variety of contaminants in practice, it is important to evaluate required data set size achieving accurate results. A randomly chosen subset of the data composed of 100 spectra per class is first analyzed. The coefficient of determination (R^2) of linear regression was also calculated as a complementary metric to MSE to evaluate model performance.[41] MSE and R^2 score were calculated as a function of training data size and plotted in Fig. 5.9b-c. As one can see the MSE (R^2 score) values are high (low) for this smaller dataset and exhibit high fluctuations. The training dataset includes 960 spectra per class per exposure condition, this requires 10 min for acquisition. The control dataset contains 7680 spectra. As before SMOTE is used for data augmentation for the concentration classes to balance with control data. When the training dataset has 1000 spectra per class, which contains only 40 generated spectra, the model achieved a MSE value of 0.17 (0.23) for As^{3+} (Cr^{6+}) and R^2 score of 0.98 (0.97) for As^{3+} (Cr^{6+}). If further augmentation is performed using SMOTE to produce 7680 spectra per class to balance with control, the MSE reduces to 0.09 (0.11) for As^{3+} (Cr^{6+}) and R^2

score increases to 0.99 for both As^{3+} and Cr^{6+}. Thus we can achieve robust model performance using SERS spectra which can be acquired rapidly.

The 1D CNN regression model performance on the balanced data set is plotted in Fig. 5.9d-e. The results are presented as box plots where the data in the boxes contains 50% of the predicted values of the holdout data, vertical lines extend to include up to 99% of predicted values, and the remaining outliers are represented individually by blue dots. The narrow height of the box plots show that SERS + ML provides concentration quantification with high precision. The gray shaded region at the bottom of figures highlights the LOB. The resulting LOD is highlighted with a vertical dashed line and is defined as having less than 0.5% overlap with control data. The values are in agreement with that determined by the SVM model (Fig. 5.7a-b) demonstrating robust performance of SERS + ML regardless of algorithm type. The 1D CNN regression model also allows for determining a limit of quantification (LOQ), highlighted with a vertical dashed line, where the overlap between neighboring concentrations is less than 0.5%. The values of LOQ are 68 pM for Cr^{6+} and 5 pM for As^{3+}. The dynamic range spans from the LOQ to 68 μM for Cr^{6+} and LOQ to 5 mM for As^{3+}. Chronic exposure at doses of 50 μg/L of arsenic in drinking water is correlated with disease, such as cancer.[42] In addition to regulatory limits, the U.S. Environmental Protection Agency (EPA) defines a maximum contaminant level goal in drinking water that is known to have no adverse effects on the health of people. For arsenic, this value is zero. The EPA regulatory limit (10 μg/L for As^{3+} and 100 μg/L for Cr^{6+}) is the value that is enforceable and provides a buffer for health safety. There is value, therefore, in detection at concentrations lower than the regulatory limit.

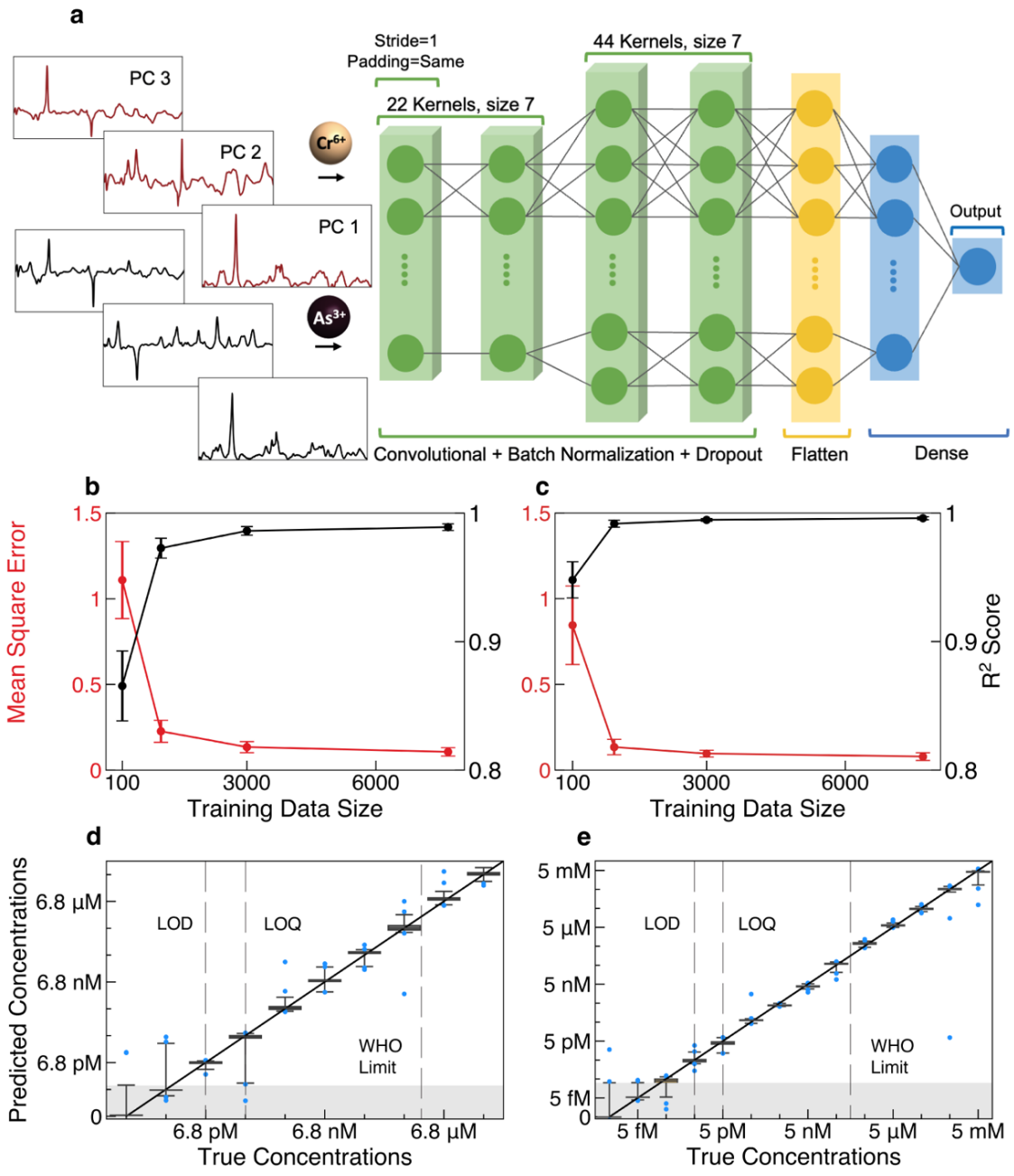

Figure 5.9 1D CNN regression model for quantitative concentration determination. **a** Schematic of process flow in training 1D CNN architectures using 22 PC from Cr^{6+} and As^{3+} concentration data. The 1D CNN model is 4 layers deep. The flatten layer is used to convert the data into a 1D array for inputting it to the fully connected dense layer. The output layer has one node with linear activation function to produce a predicted value. The MSE and R^2 variance as a function of training class size for **b** Cr^{6+}, and **c** As^{3+}. The training data size from each class is 100, 1000, 3000 and 7680. Each training algorithm runs 10 times to generate a mean value and standard

deviation for MSE and R^2. CNN regression boxplots for **d** Cr^{6+} and **e** As^{3+}. Boxes contain 50% of predicted concentration values, and vertical lines indicate the range containing 99% of predicted concentration values. Blue dots show the remaining 1% outliers.

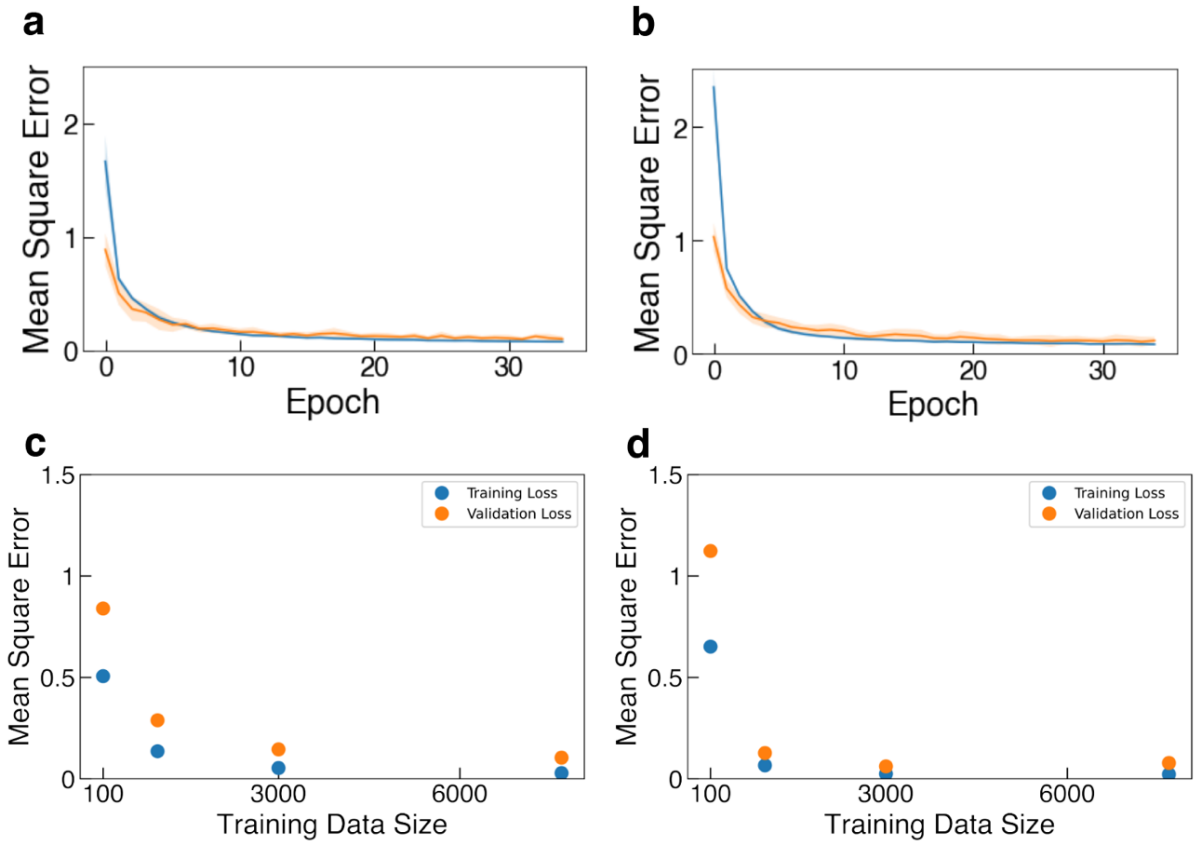

Figure 5.10 Overlaid learning curve from 1D CNN regression model for **a** Cr^{6+} and **b** As^{3+}. The blue (orange) line is training (validation) mean loss, and the blue (orange) shading is training (validation) mean loss with standard deviation. The epoch size used in 1D CNN algorithm is 35. The blue dots represent training loss and the orange dots represents validation loss for **c** Cr^{6+} and **d** As^{3+} as a function of the training data size.

5.7 Determination of Contaminant Levels in Tap Water and Wastewater Samples

Water samples from different sources unseen by the trained algorithm are analyzed to demonstrate that SERS+ML is generalizable. Drinking water, water used in agriculture, and wastewater will contain different types of impurities which may perturb the stress response of *E.*

coli. It is not feasible to fully train a new model for every different water sample. Transfer learning is an effective method to analyze similar systems with small datasets while still achieving high prediction accuracy. During transfer learning, the weights and bias of the first and second convolutional layers are adjusted and other layers are fixed. In practice, this method could be applied by spiking contaminants in water samples for fine tuning the model for the water sample of interest. In order to demonstrate this principle, a 1D CNN model was pretrained with spectra from DI water samples spiked with As^{3+} at 0.05, 0.5, and 5 nM (below WHO recommended level) and 5, 50, and 500 µM (above WHO recommended level). Then unseen tap water samples are spiked with As^{3+} at concentrations of 1.3 nM, 13 nM and 1.3 µM. A binary model is assembled to predict if tap samples contain As^{3+} above or below WHO recommended levels. The number of spectra per class needed to fine tune the model is 80, which takes only 2 min of acquisition time for the entire training dataset. The results are shown in Fig. 5.11b where the model was able to categorize tap water samples as above or below regulatory limits with 99% accuracy. It is worth noting that the different As^{3+} concentrations in the tap water samples is not the same as in the DI water samples. This is important to determining accuracy of evaluating unknown samples.

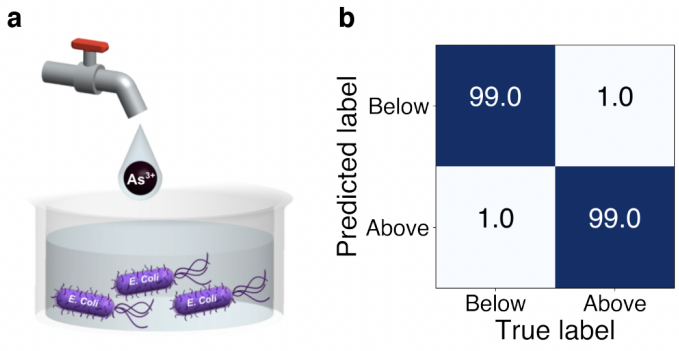

Figure 5.11 Performance of SERS + ML on unseen tap water samples. **a** *E. coli* is cultured in growth media and added to tap water supplemented with As^{3+} salts at concentrations of 1.3 nM,

13 nM and 1.3 µM for 2 hr. **b** CNN confusion matrix of binary classification of spectral lysate data exposed to tap water at concentrations above and below WHO standard for drinking water for As^{3+}.

In order to analyze more complex samples, As^{3+} was also spiked in secondary treated wastewater from a local wastewater treatment plant. These samples are more complex as they contain heavy metal contaminants in the background. Table 5.3 shows the primary pollutant analysis summary from the sanitation district where the As concentration in the background is approximately 19.4 nM. The process of determining if the concentration in the unspiked sample is above or below WHO level for As is shown in Fig. 5.12. Wastewater samples are spiked with concentrations of 1.3 nM, 13 nM, 1.3 µM and 13 µM. Again spanning above and below WHO recommended levels, 130 nM, for model fine tuning of the above pretrained DI model used for tap water. Fig. 5.12 shows classification accuracy of differentiating the different classes used for training. When applying the model to the unspiked sample the model predicts that the As^{3+} concentration is below WHO level with 92% accuracy. The total data acquisition time is 8 min; thus, acquiring samples in the field to fine tune a model in a short amount of time produces high accuracy.

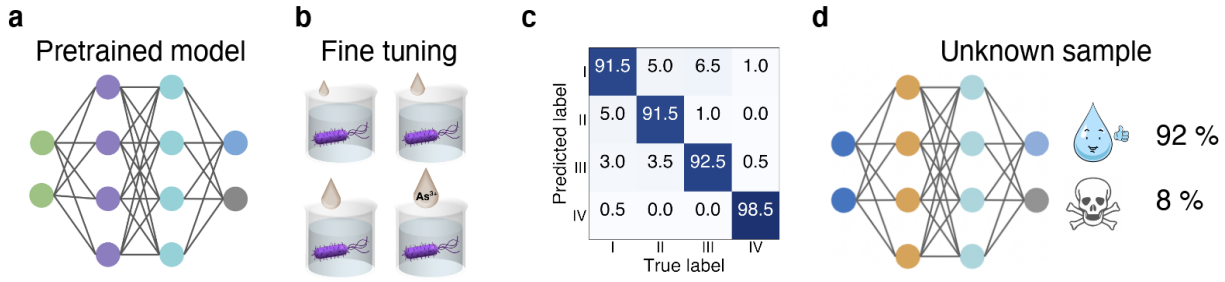

Figure 5.12 Performance of SERS + ML on unseen wastewater samples. **a** The model is pretrained on DI water **b** is fine tuned with waste water samples spiked with (I) 1.3 nM, (II) 13 nM, (III) 1.3 µM, and (IV) 13 µM As^{3+}. **c** The accuracy of differentiating the different As^{3+}

concentrations in spiked wastewater samples after pretraining. **d** The fine tuned model is able to determine that the concentration of As^{3+} in the original wastewater sample is below WHO recommended level with 92% accuracy.

Table 5.3 Priority pollutants analysis summary from 2019-20 Orange County Sanitation District resource protection division, pretreatment program annual report.[23] (As concentration of 2.52 µg/L is equivalent to 19.4 nM)

Monitoring location	Analysis	Total Average Concentration	UNIT	Flow (MGD)	Mass (lbs / day)
EFF-001	As	2.52	µg/L	101	2.12
EFF-001	Cd	0.02	µg/L	101	0.017
EFF-001	Cr	1.07	µg/L	101	0.898
EFF-001	Cu	4.92	µg/L	101	4.13
EFF-001	Hg	5.14	ng/L	101	0.004
EFF-001	Ni	7.75	µg/L	101	6.51
EFF-001	Pb	0.464	µg/L	101	0.39
EFF-001	Sb	1.32	µg/L	101	1.11
EFF-001	Se	5.81	µg/L	101	4.88
EFF-001	Zn	24.5	µg/L	101	20.6

5.8 Conclusion

The *E. coli* whole cell sensors are shown to transduce metal ions into chemical signals using the inherent metabolic stress response. Robust and sensitive SERS surfaces with high enhancement factors[18,24,27] are able to gather large, reproducible datasets needed for ML analysis. The dataset size per class for training and validation is composed of 1200 spectra, which requires 10 min when using the SERS surfaces developed by the authors. Thus we can achieve robust model performance using SERS spectra which can be acquired rapidly. Changes in the metabolite profile in *E. coli* cell lysate associated with a stress response to heavy metal toxins in water are observable in SERS spectra even when using unsupervised feature extraction methods such as t-SNE, which computes similarity of data in lower dimensional space. There are clear differences in the spectral response across the entire range of concentrations to which cells were exposed (Fig. 5.6). These plots represent validation of our hypothesis that the differences in metabolic responses observed in the cell lysate are evident in spectral data and not a result of algorithm training.

When using SVM, a supervised algorithm, for data analysis, the resulting changes in metabolite concentrations in *E. coli* cell lysate are observable in SERS spectra and differentiable across exposure concentrations with a dynamic range of 10^5 (Fig. 5.7). The spectral changes are distinct from control samples (unexposed) down to concentrations at which the number of As^{3+} in solution per cell is approximately 1. For Cr^{6+} exposure this number is approximately 10 ions per cell. These values correlate well with the fact that the EPA regulatory limit of Cr^{6+} is ten times higher than As^{3+}. Overall, the limit of detection of SERS + ML is 100,000 lower than the WHO recommended and US EPA regulatory levels (Fig. 5.7). Detection well below regulatory limits is beneficial because the EPA maximum contaminant level goal for As^{3+} is zero.

Consequently, this platform is promising for monitoring changes in water quality below regulatory limits to provide early warning of water contamination and accurate longitudinal tracking of contaminant concentrations. The metabolite changes detected by this system can also distinguish between Cr^{6+} and As^{3+} induced responses in water with a classification accuracy of 99% (Fig. 5.8). Identifying the type of metal contamination is critical to locating the source and determining necessary treatment.[43] When using 1D CNN regression algorithms, the LOQ is 68 pM for Cr^{6+} and 5 pM for As^{3+} with a dynamic range of 6 orders of magnitude (Fig. 5.9). The 1D CNN regression model yields the same LOD as SVM (Fig. 5.7 a-b) demonstrating robust performance of SERS + ML regardless of algorithm type.

Monitoring quality of tap water and water discharged from water treatment facilities will require analysis of samples with a distribution of impurities, which may perturb the stress response of *E. coli*. It is not feasible to fully train a new model for every type of water sample in the field. Transfer learning is shown to be an effective method to analyze similar systems with smaller training datasets while still achieving high prediction accuracy. By obtaining water samples and spiking with known concentrations of contaminants, a new model can be quickly fine-tuned with a smaller data set. Transfer learning using data obtained in several seconds is sufficient to determine if drinking water or wastewater is unsafe (Fig. 5.12), i.e., above or below WHO recommended limits with greater than 96% accuracy. For more complex samples, secondary treated wastewater, the fine-tuned models can determine if the unspiked waste water sample is above or below recommended safety limits with 92% accuracy. While here we demonstrated that transfer learning is an effective way to evaluate one type of metal contaminant in an 'unknown' samples with multiple background contaminants, we envision an assay approach could be used to examine water samples for the presence of other toxins. Overall, we

demonstrate that trained algorithms are rapidly generalizable across different water samples. The whole cell SERS + ML platform is promising for application to other water sources, such as recycled water, and to other metals of concern such as lead, mercury and cadmium.

5.9 Methods

5.9.1 Sensor Fabrication

SERS surfaces are fabricated in microfluidic channels with a capacitor architecture to apply an AC potential across electrodes (Fig. 5.1) to induce electrohydrodynamic (EHD) flow. Fabrication is performed silicon substrates (NOVA Electronic Materials, P-type, boron doped <100> with resistivity of 0.001-0.005 ohm-cm) with dimensions of 15 mm × 15 mm that are spin coated with poly(styrene-b-methyl methacrylate) (PS-b-PMMA, Mn S-b-MMA 170000-b-145000 g mol^{-1}) thin films of approximate thickness of 25 nm; Si substrates serve as the working electrode. Indium tin oxide (ITO) coated glass slides (Delta Technologies) serve as the counter electrode. EHD, which results as Au NPs attach to the working electrode and locally perturb the surface potential, is used as an external driving force for crosslinking reactions between 40 nm lipoic acid functionalized Au NPs (Nanocomposix, 0.13 nM) to form anhydride linking group, which define nanogap spacings. Chemical cross linking reactions between NP leads to Au NP clusters with reproducible SERS signal over a large area.[28]

Silicon substrates were cleaned by 20% v/v hydrofluoric acid (HF, Fisher Scientific, 48%) / deionized (DI) water (Milli-Q Millipore System, 18.2 MΩ cm^{-1}) for 5 minutes to remove the native oxide layer and then immersed in DI water to regrow a thin oxide layer. *The potential of HF to cause severe injury mandates extreme caution during usage*. Random copolymer poly(styrene-co-methyl-methacrylate)-α-hydroxyl- ω-Tempo moiety (PS-r-PMMA, Polymer

Source, Mn =7400, Mw =11800, Mw /Mn =1.60, 59.6 mol% Polystyrene content) random copolymer dissolved in toluene (Fisher Scientific), 1 wt%, was spin-coated at 3000 rpm for 45 seconds on silicon substrates. PS-r-PMMA films were annealed under vacuum at 170 °C for 48 hours followed by a rinse with toluene to leave a brush layer. PS-b-PMMA is spin coated at 5000 rpm for 45 seconds and then annealed for 72 hours at 170 °C. In order to selectively functionalize PMMA domains on PS-b-PMMA diblock copolymer films with amine functional groups for crosslinking with Au NPs, PS-b-PMMA/Si were immersed in dimethyl sulfoxide (DMSO, Sigma Aldrich) for 5 min and then 5 % vol ethylenediamine (ED, Sigma Aldrich) in DMSO for another 5 min. ITO counter electrodes were cleaned by ethanol (Sigma Aldrich), isopropyl alcohol, and DI water and then dried by N_2 before attaching a platinum wire and silver paste (Epoxy Technology) to make electrical contact.

A microfluidic cell was formed between electrodes using a 90 µm spacer layer composed of 3M 9816L. A solution of 2 µL N-hydroxysulfosuccinimide (s-NHS, Sigma Aldrich), 20mM, and 2 µL 1-Ethyl-3-(3-dimethyl aminopropyl) carbodiimide (EDC, Sigma Aldrich), 8 mM, in a 2-(Nmorpholino) ethane sulfonic acid buffer (MES, Sigma Aldrich, 0.1 M, pH=4.7) added to a 0.25 mL solution of 2.6 nM lipoic acid functionalized Au NP solution. 20 µL of the solution containing Au NP, s-NHS, and EDC is added to the microfluidic cell. An AC electrical stimuli with potential of 5 V_p and frequency of 100 Hz is applied for 2 min to deposit a seed layer to induce EHD flow. The second deposition step was conducted at a potential of 5 V_p and frequency of 1000 Hz for 2 min to grow Au NP clusters. After deposition, the electrode cell was dismantled and the sensor surface was thoroughly rinsed with DI water and isopropyl alcohol (IPA, Sigma Aldrich) and then dried with N_2. Supplementary Fig. S5.4 show reproducible intensity across the SERS surface and Supplementary Fig. S5.5 compares to intensity from a

benzenethiol monolayer obtained from samples fabricated using EHD and drop casting, where the latter has lower signal and highly variable intensity.

5.9.2 Media, Heavy Metal and Carbon Source Supplement

M63 media (VWR Life Science) solution was made by first diluting 1 liter of pre-sterilized M63 5x (BioWORLD, GeneLinx International Inc.) stock solution using autoclaved Millipore water. Filter-sterilized magnesium sulfate anhydrous ($MgSO_4$, Fisher Scientific) water solution, of volume 1 mL and molarity of 1 M, was added to the diluted media solution following standard protocol. Sodium arsenate stock solution (RICCA, 100 mM), was first filter-sterilized and then diluted with sterilized DI water to reach concentrations of 0.1 mM and 0.1 μM and stored under 4 ℃. Potassium dichromate (Fisher Scientific) solution was made by first dissolving sodium dichromate crystal into sterilized DI water to reach concentrations of 17 mM, and then the solution was filter-sterilized and diluted with sterilized DI water again to reach concentrations of 0.34 mM and 0.34 μM and stored at 4 °C. Prior to exposure to bacterial cultures, working solutions were placed at room temperature for 30 min to equilibrate to ambient temperature, and then titrated to the culture to target exposure concentration. Anhydrous Dextrose (glucose, Fisher Scientific), 1 g, was dissolved in 10 mL DI water and filter-sterilized to form 10% (w/v) glucose stock solution, which was added into the media solution later to provide energy source for bacteria.

5.9.3 Growth and Subculture Condition

A sterilized wooden applicator was used to streak *Escherichia coli* K12 strain MG1655 (Yale Stock Center via the Goulian Lab) frozen stock onto an lysogeny broth (LB, IBI scientific) agar plate. The plate was then placed into an incubator and incubated stationarily for 18 hr. A

single colony was picked from the plate after incubation and used to inoculate 5 mL sterile LB solution in a test tube. The inoculated culture tube was then placed in the shaking incubator (I series 24R, New Brunswick) set at 37 ℃ and speed of 250 rpm for 18 hours. After incubation, the final optical density (OD) was approximately 1.5 as measured with a colorimeter (WPA CO7500 colorimeter, Biochrom Ltd., Cambridge, UK). From the shaking culture, 3 mL was transferred to a 50 mL conical centrifuge tube and centrifuged at the speed of 5000 rpm for 5 min (Sorvall Legend X1R centrifuge, Fisher Scientific). Then the supernatant was disposed and the pellets were resuspended in 1 mL of 1× phosphate-buffered saline (PBS, Fisher Scientific, 10× solution) solution. The pellet-PBS mixture was transferred to 1 mL centrifuge tubes, centrifuged at 5000 g for 5 minutes (accuSpin Micro 17, Fisher Scientific) and the supernatant was disposed. The washing step was repeated. After, the pellet was resuspended in 1 mL M63 defined media, resulting in a milky M63-pellet mixture with very high OD. M63 media supplemented with 1% (w/v) glucose was pipetted into sterilized test tubes and the pellet-M63 mixture was titrated into the test tubes to reach the final OD of 0.5. The total volume of liquid in the tube was 5 mL. Three tubes, having a 15 mL culture, were prepared for a single colony. These tubes were then moved to the shaking incubator for subculturing with the shaking speed set at 250 rpm and temperature at 37℃ for 6 hr. Then, the 15 mL subculture was transferred to 50 mL centrifuge tubes, centrifuged twice at a speed of 5000 rpm for 5 min and washed with 1 mL of PBS twice. The subculture was resuspended in 1 mL M63 defined media before being exposed to heavy metals.

5.9.4 Bacterial Exposure to Heavy Metal and Growth Curve Measurement

E. coli (K12 MG1655 strain) is cultured in defined media M63 to achieve an optical density of 0.5 and supplemented with 1% (w/v) glucose to mitigate conflating stress from heavy metal stress ions with nutrition limitation. The subcultures prepared as described in the prior

section were washed with 1 mL PBS twice and resuspended in M63 defined media. M63 media supplemented with 1% glucose (w/v) was pipetted into wells of white-opaque 96-well microplates. Different concentrations of heavy metal ($NaAsO_2$ or $K_2Cr_2O_7$) was added to the wells. Specifically, 0, 1, 10, 100, 1000 µM of $NaAsO_2$ and 0, 0.34, 3.4, 34, 170 µM of $K_2Cr_2O_7$ were exposed to cultures for 2 hr. The resuspended culture was pipetted into the wells to make the OD of the culture 0.5. Each condition was done in biological duplicate. After pipetting, the microplate was placed in the SkanIt Microplate Reader (Thermo Scientific) at 37 ℃ and shaken at a speed 300 rpm and high force. The OD of the culture in each well was measured every 5 min for 6 hours to generate growth curves.

Preparation of cultures exposed to tap water and wastewater from Orange County Sanitation District (OCSD) involves similar steps as those exposed to DI water spiked with As^{3+}, except after washing with PBS, the subculture was resuspended in tap water or wastewater supplemented with 1% (w/v) glucose at a OD of 0.5, and the heavy metal salts were dissolved in tap water or wastewater instead of the defined media. The secondary treated wastewater was treated by primary sedimentation followed by activated sludge process with nitrification and denitrification at OCSD. Before spiking with As^{3+}, the secondary treated wastewater was filtered with 0.45 µm MCE Membrane (MF-Millipore).

5.9.5 Lysate Sample Preparation

After two hr exposure, the washing steps used to prepare subculture was performed again to eliminate residual media and heavy metals from the pellet and avoid mixing of these with metabolites released during lysing. The pellet was then resuspended in 100 µL Millipore water and heated to a temperature of 97ºC for 30 min to lyse the cells. Then the pellet-water mixture

was centrifuged at 12000 g for 10 min. Then 100 µL supernatant in each tube was evenly divided into 4 parts by pipetting into 4 different 1 mL sterile centrifuge tubes, 25 µL each transfer. These supernatant samples were placed in the -20 ℃ freezer to store for further analysis.

5.9.6 Data Acquisition

Spectral data is acquired by placing a droplet of with volume of 25 µL of lysate from *E.coli* cells untreated (control) or exposed to heavy metal ion solutions at various concentrations on SERS surfaces. The measured concentration range for $NaAsO_2$ was 0.65 pg/L to 650 mg/L (13 concentrations) and for $K_2Cr_2O_7$ was 0.1 ng/L to 10 mg/L (9 concentrations) spaced by one order of magnitude as shown in Table 1. The corresponding concentrations in molarity of As^{3+} and Cr^{6+} are shown in Table 5.4.

Table 5.4 Cr^{6+} (10 classes) and As^{3+} (14 classes) for machine learning models. C is the control class. Superscripts indicate SERS data acquired on the same SERS surface.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Cr^{6+}	C	0.68 pM*	6.8 pM*	68 pM*	680 pM†	6.8 nM†	68 nM†	680 nM‡	6.8 µM‡	68 µM‡	X	X	X	X
As^{3+}	C	5 fM§	50 fM§	500 fM¶	5 pM¶	50 pM¶	500 pM¶	5 nM¶	50 nM¶	500 nM#	5 µM‖	50 µM‖	500 µM‖	5 mM‖

For each exposure concentration, a dataset of 1200 SERS spectra is acquired using a Renishaw InVia micro Raman system with an integration time of 0.5 s, 146 µW laser power at 785 nm excitation wavelength and a 60X water immersion lens with 1.2 NA. Raman maps were acquired in an array of 20 x 20 with 3 µm steps between measurement points, resulting in 400 spectra per map. Three maps were acquired over different regions of the sample surface resulting in a total of 1200 spectra per concentration for each metal ion defining a class for initial training of machine learning algorithms. The dataset acquisition takes 10 min and the droplet does not

evaporate during this period of time. In order to ensure the algorithm is not being trained to detect batch-to-batch variations of SERS surfaces, between two and six concentration classes, including control samples, were acquired on different regions of the same SERS surface (droplets exposed to isolated regions), indicated by superscripts in Table 1. Furthermore, the control group, prepared under the same conditions in the absence of Cr^{6+} or As^{3+} exposure, were measured in from lysate samples prepared in biological duplicate on different days, from the eight different SERS surfaces, also fabricated on different days, used for the other metal concentrations exposure conditions to train algorithms to not identify differences based on normal variability of experimental conditions such as culture growth, device fabrication, and processing steps.

5.9.7 Pre-processing of SERS Spectra Data

For data preprocessing, asymmetric least square correction is utilized for baseline correction, and a Savitzky–Golay filter is used for data smoothing. In order to normalize the data, the vibrational band of silicon at 520 cm^{-1} is used as an internal standard and set to 1. The diblock copolymer layer, between Si and NP clusters, is 25 nm thick and thus Si surfaces is not affected by the signal enhancement of Au NP clusters. The metal ion concentration unit was labeled with a log scale since concentrations investigated span several orders of magnitude. Principal component analysis (PCA) was performed for dimensional reduction. We determined that 22 PCA components captured 93.3% and 94.9% of variances for Cr^{6+} and As^{3+} concentration data, respectively. t-distributed stochastic neighbor embedding (tSNE) was also performed to visualize the concentration data in lower dimensional space and show there are spectral differences in the data observed without labeling data for algorithms.

5.9.8 Support Vector Machine Classification Model

Two independent support vector machine (SVM) discriminative models are trained on Cr^{6+} and As^{3+} exposed lysate spectra data for the classes shown in Table 1. The training datasets are imbalanced since the size of the control class dataset (9600 spectra) is 8 times larger than the classes corresponding to a single concentration (1200 spectra). The synthetic minority over-sampling technique (SMOTE) is used to oversample skewed classes in the dataset and achieve a balanced dataset. SMOTE works by selecting a random example from the minority class, then k of the nearest neighbors for that example are found. A randomly selected neighbor is chosen and a synthetic example is created at a randomly selected point between the two examples in feature space. SMOTE can alleviate overfitting by increasing stability with respect to random fluctuations and thereby increase the generalization capability of the classifier.[35] SMOTE is performed after data split within each cross-validation fold to prevent data leakage.

The SVM models are trained using 22 PCA components. A holdout set is composed of 20% of the data that is used for final validation and not seen at all during training. The model is trained with the remaining 80% of the spectral data labeled with their appropriate class to define a hyperplane separating data into the correct classes. SVM models are trained with Scikit-learn using default parameters, with radial basis function (RBF) kernel, Margin parameter (C) =1, and γ=scale. In order to evaluate SVM model performance, sampling cross validation is performed using 10-fold stratified sampling on the training dataset for initial evaluation of model performance. Here, each fold is shuffled and used as validation data to estimate prediction accuracy. The cross-validation results are in the Supplementary Fig. S6 and Fig.S7. The final model is trained with 80% training data and tested with 20% holdout set.

5.9.9 Statistical Analysis

The statistical significance between the OD when exposed for 2 hours to different heavy metal concentrations (Fig. 3e and f) was calculated using a two-tailed Student's t-test. All growth experiments were done with biological duplicates (n=2) in 96-well plates. The OD after 2 hours of exposure was calculated as the average of three replicate wells and the error bars represent the standard deviation of the OD of the three wells. The degrees of freedom for all statistical calculations in the two plots are 2. The *t* values and *p* values are shown in Table 5.5.

Table 5.5 *t* value and *p* values of final OD after 2-hour exposure to heavy metals.

$[As^{3+}]$ (μM)	0	0.68	6.8	68	340
t value	N/A	0.1677	0.3912	30.5232	5.8938
p value	N/A	0.8822	0.7334	0.0011	0.0276
$[Cr^{6+}]$ (μM)	0	1	10	100	1000
t value	N/A	0.4158	0.3767	1.0135	1.1834
p value	N/A	0.7179	0.7426	0.4175	0.3583

5.9.10 Convolutional Neural Network Regression Model

The 1D CNN model architecture utilizes Keras framework with Tensorflow backend. 22 PCA components are used as input for both Cr^{6+} (0.68 pM - 68 μM) and As^{3+} (5 fM - 5 mM) exposed lysate spectra datasets, respectively. The first convolutional layer is the data input layer, which has 22 kernels with size 7 and 1 stride to preserve the spatial size with the same padding. The second convolutional layer also has 22 kernels with size 7. The third and fourth

convolutional layers are identical, with 44 kernels with size 7. Each convolutional layer followed by a batch normalization layer and a dropout layer with 20% random dropout rate. Batch normalization mitigates changes in the distribution of network activations due to the change in network parameters during training. Dropout layers are used to prevent overfitting. Followed with convolutional layers, a flatten layer is added to reshape the 2D extracted feature into a 1D vector followed by a dropout layer. Fully connected layers with 22 nodes with a L2 norm regularization (0.001) and ReLU activation function are applied to process the 1D vector. Finally, using the linear function, the weighted sum of the flatten layer is condensed into a one-unit neuron containing the prediction result between 0 and 9 (Cr^{6+}) or 13 (As^{3+}), where the continuous score supplies predicted concentrations.

Hyperparameters of the 1D CNN regression model including number of hidden layers and units, activation function, dropout rate, batch size, kernel size and number of epochs are optimized by monitoring training and validation loss during 10-fold cross validation. To be specific, EarlyStopping was used by monitoring the increase in validation loss to determine the number of epochs. Early termination was determined when the validation loss was increasing for 10 consecutive epochs, indicating that the 1D CNN had reached maximum convergence. During 10-fold cross validation, they all reach the convergence at approximately 35 epochs, which was thus chosen for the final model. During 10-fold cross validation, the loss function is calculated to determine the average of the squared differences between the predicted and true values. The overlaid learning curve from 10-fold cross validation shows no obvious gap between training loss and validation loss, which show the absence of overfitting.

Due to the large size of control dataset acquired to capture variability of experimental conditions, including biological culture conditions and device fabrication, the data classes are

imbalanced. Again SMOTE is used to balance the training dataset and here the training dataset size is varied to contain 100, 1000, 3000 and 7680 randomly selected spectra from each class to determine the size of needed training data for accurate predictions. As before 20% of the spectral data is set aside as a holdout set, i.e., not used in training. The performance of the 1D CNN regression model is evaluated by calculating mean square error (MSE) and coefficient of determination (R^2) scores for four different dataset sizes. The R^2 metric is the ratio of explained sum of squares and the total sum of squares and is sensitive in the order of predicted and actual targets. MSE and R^2 score mean values and standard deviation (std) are calculated by running the calculations ten times.

The final CNN model is trained with tuned hyperparameters on 80% of the spectral data (training set) and the model performance is evaluated on the remaining 20% of the spectral data (hold out set), with batch size 44, number of epoch 35, and using Adam for gradient descent optimization. The holdout set in the classes are unbalanced where the control class has 1920 spectra and other classes have 240 spectra. We thus use random downsampling of the control to include 240 spectra to balance the data and represent in the box plot on Fig. 5d-e.

5.9.11 Transfer Learning

The transferred convolutional neural network is built by Tensorflow 1.8 in Python 3.6. The 1D CNN binary classification model is pre-trained to identify the heavy metal concentration in DI water. The classes contain spectra from DI water samples spiked with concentrations of As^{3+} of 0.05 nM, 0.5 nM, and 5 nM (below WHO recommended level) and 5 μM, 50 μM, and 500 μM (above WHO recommended level). The pre-trained model is then transferred to identify if the As^{3+} concentration in tap water samples is above or below WHO recommended level. The

concentrations tested are 1.3 nM, 13 nM, (below) and 1.3 μM (above). For wastewater, the four classes tested contain 1.3 nM, 13 nM, (below) 1.3 μM, and 13 μM (above) concentration of As^{3+}. The fully connected layer and output layer of the pretrained model is replaced with an output layer which has 1 node with sigmoid activation function. The weights of the third and forth convolutional layers are frozen throughout fine tuning, and the weight of the first and second layer are set to be trainable. Before fine tuning, the model is compiled with binary cross-entropy as loss function, accuracy as metric, and Adam optimizer with a 0.001 learning rate is used. 80 examples from each class from the new water type are used to fine tune the compiled the transferred model. The performance of the transferred model is tested by 1040 tap water samples.

5.10 Supplemental Information

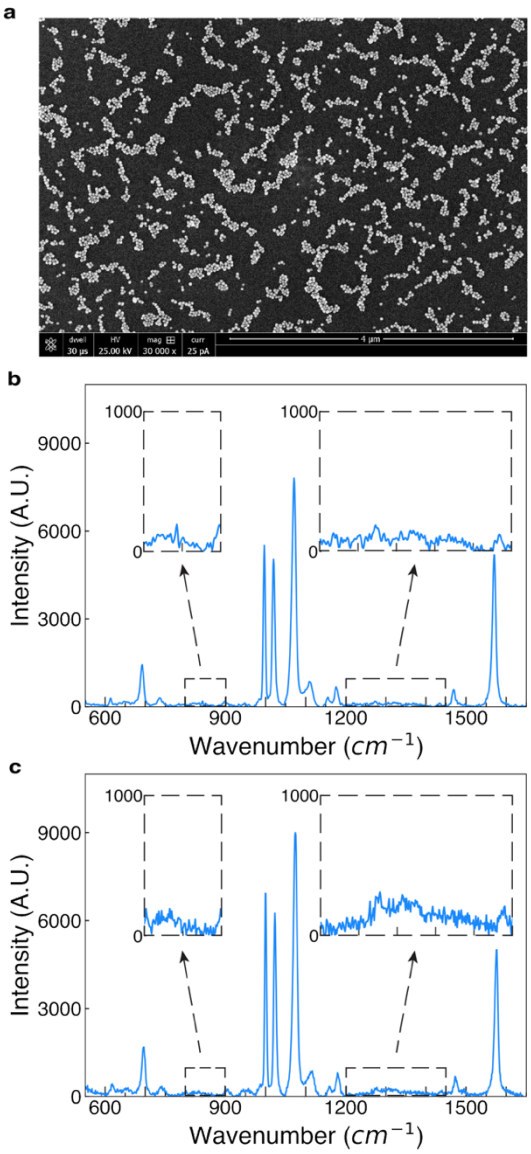

Figure S5.1 **a.** Scanning electron microscopy image of sensor surface with field of view of approximately 30 µm^2. The scale bar is 4 µm. Background subtracted Raman spectra from **b.** the BW Tek i-Raman Plus (Model: BWS465-785S) portable Raman system and **c.** the Renishaw Invia micro Raman system. No modes observed in the ranges of 800 - 900 cm^{-1} and 1200 - 1450 cm^{-1} from gas phase benzenethiol Raman spectra. These featureless regions are highlighted by dotted squares and used for signal to noise ratio calculation. The inset of b. and c. shows the intensity of the SERS signal from featureless regions for ease of visualization. The peak intensity used for signal to noise ratio calculation at 1075 cm^{-1} represent C−C symmetric stretching and C−S stretching vibrational modes.[1]

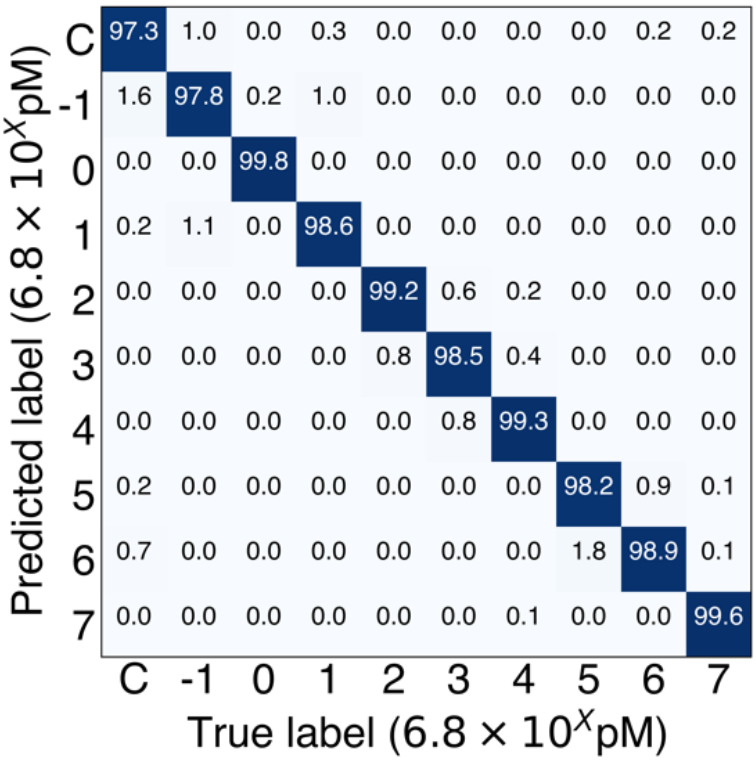

Figure S5.2 Confusion matrices from 10-fold cross validation of SVM classification models for Cr^{6+} obtained from the training dataset.

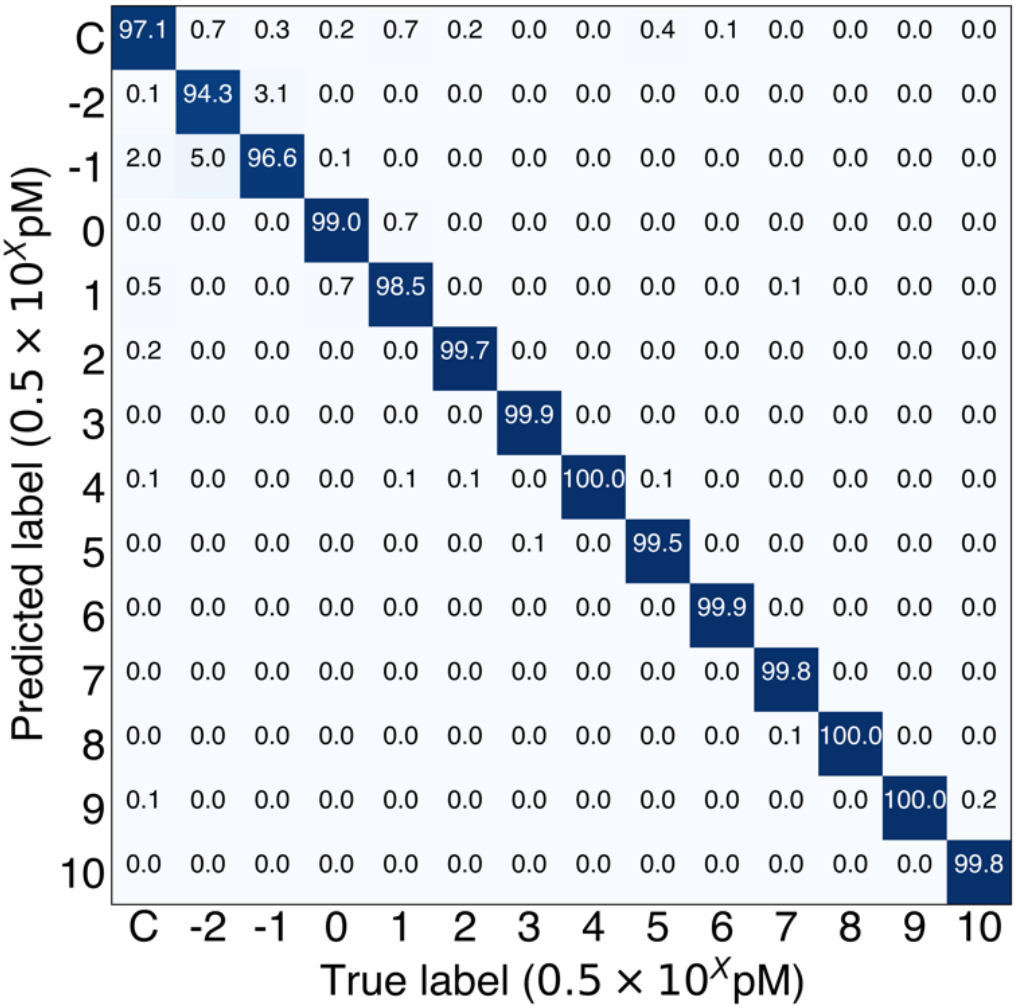

Figure S5.3 Confusion matrices from 10-fold cross validation of SVM classification models for As^{3+} obtained from the training dataset.

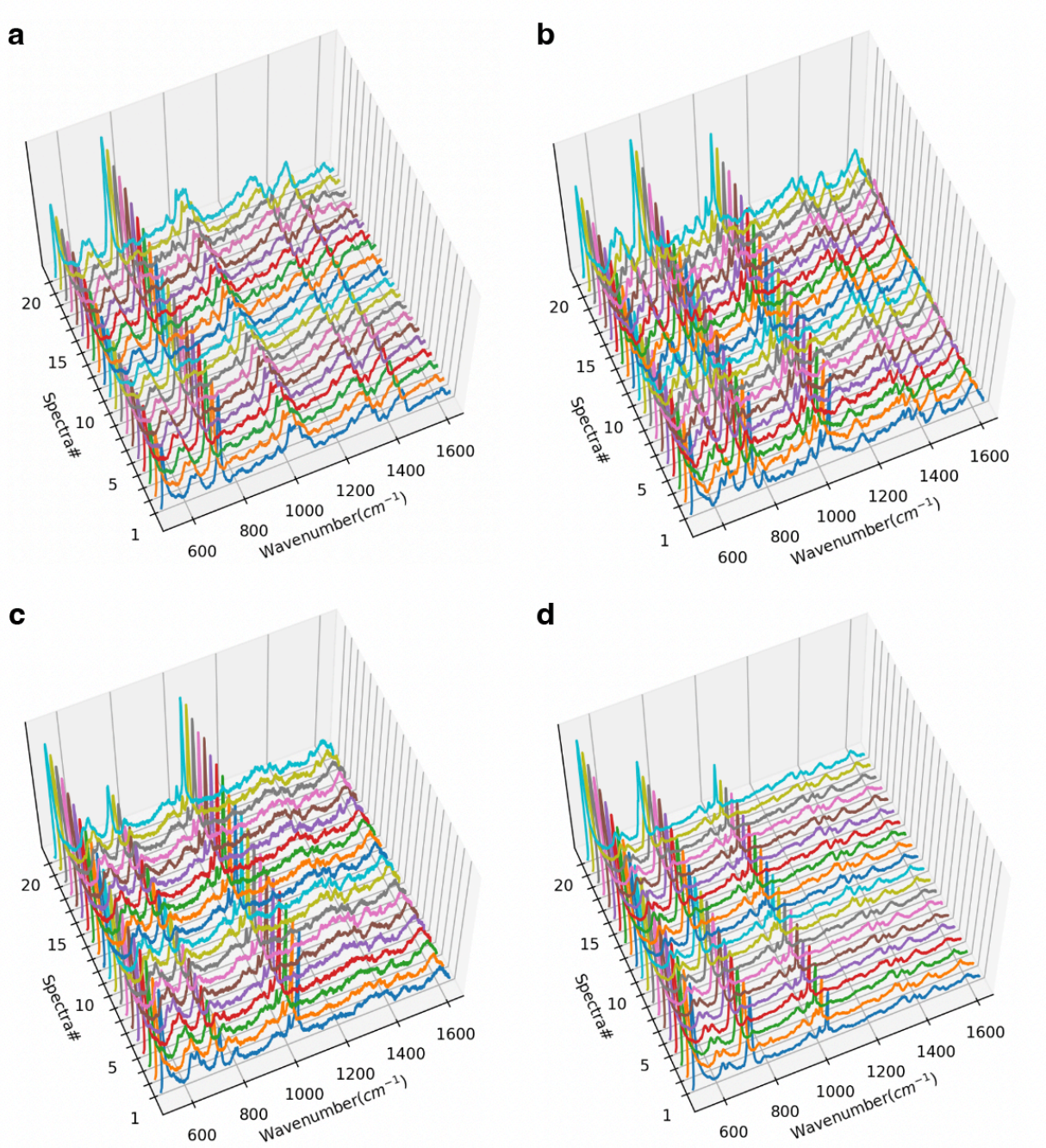

Figure S5.4 Waterfall plots of 20 randomly chosen spectra at concentrations of **a.** 10x and **b.** 1000x LOD from Cr^{6+} (6.8 pM) and **c.** 10x and **d.** 1000x LOD from As^{3+} (0.5 pM) showing uniform signals across sensor surfaces.

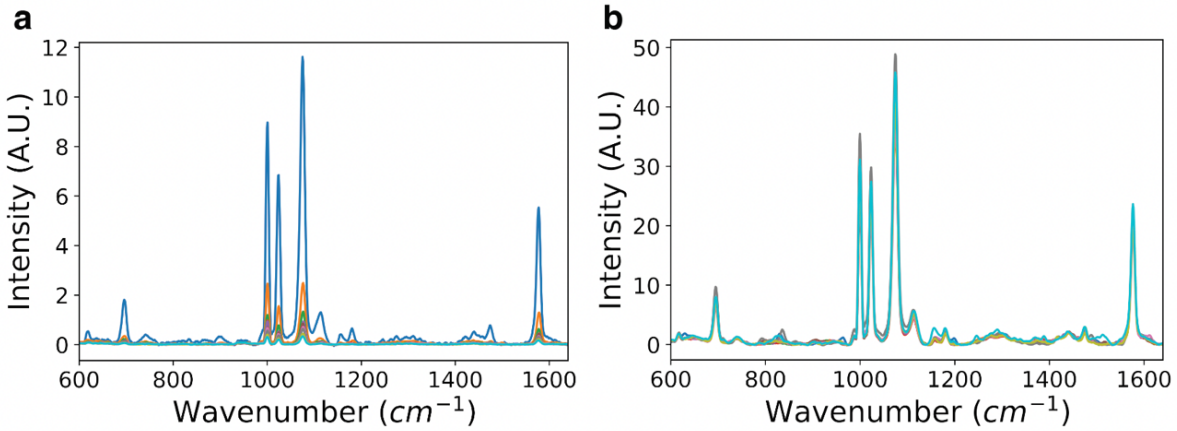

Figure S5.5 Randomly selected spectra from SERS surfaces fabricated by a. Drop-casting Au nanoparticles and b. with EHD flow. Per convention, the silicon peak at 520 cm^{-1} (not shown) is used as an internal standard and normalized to 1 for all spectra. The total SERS intensity is different in Fig. a and b due to different enhancement factors from the two types of surfaces.

5.11 References

(1) Tchounwou, P. B.; Yedjou, C. G.; Patlolla, A. K.; Sutton, D. J. Heavy Metals Toxicity and the Environment. *EXS* **2012**, *101*, 133–164. https://doi.org/10.1007/978-3-7643-8340-4_6.

(2) Bradl, H. *Heavy Metals in the Environment: Origin, Interaction and Remediation*; Elsevier, 2005.

(3) Turdean, G. L. Design and Development of Biosensors for the Detection of Heavy Metal Toxicity. *Int. J. Electrochem.* **2011**, *2011*, e343125. https://doi.org/10.4061/2011/343125.

(4) *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First and Second Addenda*; WHO Guidelines Approved by the Guidelines Review Committee; World Health Organization: Geneva, 2022.

(5) Zhang, Y.; Ren, T.; Tian, H.; Jin, B.; He, J. Hydrogel-Encapsulated Enzyme Facilitates Colorimetric Acute Toxicity Assessment of Heavy Metal Ions. *ACS Appl. Mater. Interfaces* **2018**, *10* (31), 26705–26712. https://doi.org/10.1021/acsami.8b08949.

(6) Xing, C.; Liu, L.; Zhang, X.; Kuang, H.; Xu, C. Colorimetric Detection of Mercury Based on a Strip Sensor. *Anal. Methods* **2014**, *6* (16), 6247–6253. https://doi.org/10.1039/C3AY42002G.

(7) Kim, H. J.; Jeong, H.; Lee, S. J. Synthetic Biology for Microbial Heavy Metal Biosensors. *Anal. Bioanal. Chem.* **2018**, *410* (4), 1191–1203. https://doi.org/10.1007/s00216-017-0751-6.

(8) Wu, Y.; Liu, L.; Zhan, S.; Wang, F.; Zhou, P. Ultrasensitive Aptamer Biosensor for Arsenic(III) Detection in Aqueous Solution Based on Surfactant-Induced Aggregation of Gold Nanoparticles. *Analyst* **2012**, *137* (18), 4171–4178. https://doi.org/10.1039/C2AN35711A.

(9) BelBruno, J. J. Molecularly Imprinted Polymers. *Chem. Rev.* **2019**, *119* (1), 94–119. https://doi.org/10.1021/acs.chemrev.8b00171.

(10) Sharma, G.; Kandasubramanian, B. Molecularly Imprinted Polymers for Selective Recognition and Extraction of Heavy Metal Ions and Toxic Dyes. *J. Chem. Eng. Data* **2020**, *65* (2), 396–418. https://doi.org/10.1021/acs.jced.9b00953.

(11) Zheng, J.; Wai, J. L.; Lake, R. J.; New, S. Y.; He, Z.; Lu, Y. DNAzyme Sensor Uses Chemiluminescence Resonance Energy Transfer for Rapid, Portable, and Ratiometric Detection of Metal Ions. *Anal. Chem.* **2021**, *93* (31), 10834–10840. https://doi.org/10.1021/acs.analchem.1c01077.

(12) Wang, Y.; Zhang, C.; Liu, F. Antibody Developments for Metal Ions and Their Applications. *Food Agric. Immunol.* **2020**, *31* (1), 1079–1103. https://doi.org/10.1080/09540105.2020.1828293.

(13) Kannappan, S.; Ramisetty, B. C. M. Engineered Whole-Cell-Based Biosensors: Sensing Environmental Heavy Metal Pollutants in Water—a Review. *Appl. Biochem. Biotechnol.* **2021**. https://doi.org/10.1007/s12010-021-03734-2.

(14) Sherman, L. M.; Petrov, A. P.; Karger, L. F. P.; Tetrick, M. G.; Dovichi, N. J.; Camden, J. P. A Surface-Enhanced Raman Spectroscopy Database of 63 Metabolites. *Talanta* **2020**, *210*, 120645. https://doi.org/10.1016/j.talanta.2019.120645.

(15) Thrift, W. J.; Ronaghi, S.; Samad, M.; Wei, H.; Nguyen, D. G.; Cabuslay, A. S.; Groome, C. E.; Santiago, P. J.; Baldi, P.; Hochbaum, A. I.; Ragan, R. Deep Learning Analysis of Vibrational Spectra of Bacterial Lysate for Rapid Antimicrobial Susceptibility Testing. *ACS Nano* **2020**, *14* (11), 15336–15348. https://doi.org/10.1021/acsnano.0c05693.

(16) Sun, D.; Cao, F.; Tian, Y.; Li, A.; Xu, W.; Chen, Q.; Shi, W.; Xu, S. Label-Free Detection of Multiplexed Metabolites at Single-Cell Level via a SERS-Microfluidic Droplet Platform. *Anal. Chem.* **2019**, *91* (24), 15484–15490. https://doi.org/10.1021/acs.analchem.9b03294.
(17) Sun, B.; Jiang, X.; Wang, H.; Song, B.; Zhu, Y.; Wang, H.; Su, Y.; He, Y. Surface-Enhancement Raman Scattering Sensing Strategy for Discriminating Trace Mercuric Ion (II) from Real Water Samples in Sensitive, Specific, Recyclable, and Reproducible Manners. *Anal. Chem.* **2015**, *87* (2), 1250–1256. https://doi.org/10.1021/ac503939d.
(18) Thrift, W. J.; Ragan, R. Quantification of Analyte Concentration in the Single Molecule Regime Using Convolutional Neural Networks. *Anal. Chem.* **2019**, *91* (21), 13337–13342. https://doi.org/10.1021/acs.analchem.9b03599.
(19) Graham, D.; Thompson, D. G.; Smith, W. E.; Faulds, K. Control of Enhanced Raman Scattering Using a DNA-Based Assembly Process of Dye-Coded Nanoparticles. *Nat. Nanotechnol.* **2008**, *3* (9), 548–551. https://doi.org/10.1038/nnano.2008.189.
(20) Lim, D.-K.; Jeon, K.-S.; Hwang, J.-H.; Kim, H.; Kwon, S.; Suh, Y. D.; Nam, J.-M. Highly Uniform and Reproducible Surface-Enhanced Raman Scattering from DNA-Tailorable Nanoparticles with 1-Nm Interior Gap. *Nat. Nanotechnol.* **2011**, *6* (7), 452–460. https://doi.org/10.1038/nnano.2011.79.
(21) Zhao, X.; Campbell, S.; El-Khoury, P. Z.; Jia, Y.; Wallace, G. Q.; Claing, A.; Bazuin, C. G.; Masson, J.-F. Surface-Enhanced Raman Scattering Optophysiology Nanofibers for the Detection of Heavy Metals in Single Breast Cancer Cells. *ACS Sens.* **2021**, *6* (4), 1649–1662. https://doi.org/10.1021/acssensors.1c00332.
(22) Sahu, S.; Roy, R.; Anand, R. Harnessing the Potential of Biological Recognition Elements for Water Pollution Monitoring. *ACS Sens.* **2022**, *7* (3), 704–715. https://doi.org/10.1021/acssensors.1c02579.
(23) Kim, J.-M.; Lee, C.; Lee, Y.; Lee, J.; Park, S.-J.; Park, S.; Nam, J.-M. Synthesis, Assembly, Optical Properties, and Sensing Applications of Plasmonic Gap Nanostructures. *Adv. Mater.* **2021**, *33* (46), 2006966. https://doi.org/10.1002/adma.202006966.
(24) Nguyen, C. Q.; Thrift, W. J.; Bhattacharjee, A.; Ranjbar, S.; Gallagher, T.; Darvishzadeh-Varcheie, M.; Sanderson, R. N.; Capolino, F.; Whiteson, K.; Baldi, P.; Hochbaum, A. I.; Ragan, R. Longitudinal Monitoring of Biofilm Formation via Robust Surface-Enhanced Raman Scattering Quantification of Pseudomonas Aeruginosa-Produced Metabolites. *ACS Appl. Mater. Interfaces* **2018**, *10* (15), 12364–12373. https://doi.org/10.1021/acsami.7b18592.
(25) Cullen, W. R.; Reimer, K. J. Arsenic Speciation in the Environment. 52.
(26) Prasad, S.; Yadav, K. K.; Kumar, S.; Gupta, N.; Cabral-Pinto, M. M. S.; Rezania, S.; Radwan, N.; Alam, J. Chromium Contamination and Effect on Environmental Health and Its Remediation: A Sustainable Approaches. *J. Environ. Manage.* **2021**, *285*, 112174. https://doi.org/10.1016/j.jenvman.2021.112174.
(27) Thrift, W. J.; Nguyen, C. Q.; Darvishzadeh-Varcheie, M.; Zare, S.; Sharac, N.; Sanderson, R. N.; Dupper, T. J.; Hochbaum, A. I.; Capolino, F.; Abdolhosseini Qomi, M. J.; Ragan, R. Driving Chemical Reactions in Plasmonic Nanogaps with Electrohydrodynamic Flow. *ACS Nano* **2017**, *11* (11), 11317–11329. https://doi.org/10.1021/acsnano.7b05815.
(28) Armbruster, D. A.; Pry, T. Limit of Blank, Limit of Detection and Limit of Quantitation. *Clin. Biochem. Rev.* **2008**, *29* (Suppl 1), S49–S52.
(29) Belenky, P.; Ye, J. D.; Porter, C. B. M.; Cohen, N. R.; Lobritz, M. A.; Ferrante, T.; Jain, S.; Korry, B. J.; Schwarz, E. G.; Walker, G. C.; Collins, J. J. Bactericidal Antibiotics Induce

Toxic Metabolic Perturbations That Lead to Cellular Damage. *Cell Rep.* **2015**, *13* (5), 968–980. https://doi.org/10.1016/j.celrep.2015.09.059.
(30) Battesti, A.; Majdalani, N.; Gottesman, S. The RpoS-Mediated General Stress Response in Escherichia Coli. *Annu. Rev. Microbiol.* **2011**, *65*, 189–213. https://doi.org/10.1146/annurev-micro-090110-102946.
(31) Akhova, A. V.; Tkachenko, A. G. ATP/ADP Alteration as a Sign of the Oxidative Stress Development in Escherichia Coli Cells under Antibiotic Treatment. *FEMS Microbiol. Lett.* **2014**, *353* (1), 69–76. https://doi.org/10.1111/1574-6968.12405.
(32) Yang, H.-C.; Fu, H.-L.; Lin, Y.-F.; Rosen, B. P. Pathways of Arsenic Uptake and Efflux. *Curr. Top. Membr.* **2012**, *69*, 325–358. https://doi.org/10.1016/B978-0-12-394390-3.00012-4.
(33) Cooke, M. S.; Evans, M. D.; Dizdaroglu, M.; Lunec, J. Oxidative DNA Damage: Mechanisms, Mutation, and Disease. *FASEB J.* **2003**, *17* (10), 1195–1214. https://doi.org/10.1096/fj.02-0752rev.
(34) Low, D. A.; Weyand, N. J.; Mahan, M. J. Roles of DNA Adenine Methylation in Regulating Bacterial Gene Expression and Virulence. *Infect. Immun.* **2001**, *69* (12), 7197–7204. https://doi.org/10.1128/IAI.69.12.7197-7204.2001.
(35) Zhai, Q.; Xiao, Y.; Zhao, J.; Tian, F.; Zhang, H.; Narbad, A.; Chen, W. Identification of Key Proteins and Pathways in Cadmium Tolerance of Lactobacillus Plantarum Strains by Proteomic Analysis. *Sci. Rep.* **2017**, *7* (1), 1182. https://doi.org/10.1038/s41598-017-01180-x.
(36) Tremaroli, V.; Workentine, M. L.; Weljie, A. M.; Vogel, H. J.; Ceri, H.; Viti, C.; Tatti, E.; Zhang, P.; Hynes, A. P.; Turner, R. J.; Zannoni, D. Metabolomic Investigation of the Bacterial Response to a Metal Challenge. *Appl. Environ. Microbiol.* **2009**, *75* (3), 719–728. https://doi.org/10.1128/AEM.01771-08.
(37) Booth, S. C.; Weljie, A. M.; Turner, R. J. Metabolomics Reveals Differences of Metal Toxicity in Cultures of Pseudomonas Pseudoalcaligenes KF707 Grown on Different Carbon Sources. *Front. Microbiol.* **2015**, *6*.
(38) Yang, J. H.; Wright, S. N.; Hamblin, M.; McCloskey, D.; Alcantar, M. A.; Schrübbers, L.; Lopatkin, A. J.; Satish, S.; Nili, A.; Palsson, B. O.; Walker, G. C.; Collins, J. J. A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell* **2019**, *177* (6), 1649-1661.e9. https://doi.org/10.1016/j.cell.2019.04.016.
(39) Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. https://doi.org/10.1613/jair.953.
(40) Ameri, A.; Akhaee, M. A.; Scheme, E.; Englehart, K. Regression Convolutional Neural Network for Improved Simultaneous EMG Control. *J. Neural Eng.* **2019**, *16* (3), 036015. https://doi.org/10.1088/1741-2552/ab0e2e.
(41) Abrol, A.; Fu, Z.; Salman, M.; Silva, R.; Du, Y.; Plis, S.; Calhoun, V. Deep Learning Encodes Robust Discriminative Neuroimaging Representations to Outperform Standard Machine Learning. *Nat. Commun.* **2021**, *12* (1), 353. https://doi.org/10.1038/s41467-020-20655-6.
(42) 2012 Edition of the Drinking Water Standards and Health Advisories (EPA 822-S-12-001). **2012**, 20.
(43) US EPA, O. *Metals*. https://www.epa.gov/caddis-vol2/metals (accessed 2022-02-16).

(44) Chan, T.-Y.; Liu, T.-Y.; Wang, K.-S.; Tsai, K.-T.; Chen, Z.-X.; Chang, Y.-C.; Tseng, Y.-Q.; Wang, C.-H.; Wang, J.-K.; Wang, Y.-L. SERS Detection of Biomolecules by Highly Sensitive and Reproducible Raman-Enhancing Nanoparticle Array. *Nanoscale Res. Lett.* **2017**, *12* (1), 344. https://doi.org/10.1186/s11671-017-2121-x.
(45) Bell, S. E. J.; Sirimuthu, N. M. S. Surface-Enhanced Raman Spectroscopy (SERS) for Sub-Micromolar Detection of DNA/RNA Mononucleotides. *J. Am. Chem. Soc.* **2006**, *128* (49), 15580–15581. https://doi.org/10.1021/ja066263w.
(46) Madzharova, F.; Heiner, Z.; Gühlke, M.; Kneipp, J. Surface-Enhanced Hyper-Raman Spectra of Adenine, Guanine, Cytosine, Thymine, and Uracil. *J. Phys. Chem. C* **2016**, *120* (28), 15415–15423. https://doi.org/10.1021/acs.jpcc.6b02753.
(47) Yao, G.; Zhai, Z.; Zhong, J.; Huang, Q. DFT and SERS Study of ^{15}N Full-Labeled Adenine Adsorption on Silver and Gold Surfaces. *J. Phys. Chem. C* **2017**, *121* (18), 9869–9878. https://doi.org/10.1021/acs.jpcc.7b00818.
(48) Chan, J. W.; Taylor, D. S.; Zwerdling, T.; Lane, S. M.; Ihara, K.; Huser, T. Micro-Raman Spectroscopy Detects Individual Neoplastic and Normal Hematopoietic Cells. *Biophys. J.* **2006**, *90* (2), 648–656. https://doi.org/10.1529/biophysj.105.066761.
(49) Cheng, W.-T.; Liu, M.-T.; Liu, H.-N.; Lin, S.-Y. Micro-Raman Spectroscopy Used to Identify and Grade Human Skin Pilomatrixoma. *Microsc. Res. Tech.* **2005**, *68* (2), 75–79. https://doi.org/10.1002/jemt.20229.
(50) Zhu, J.; Zhou, J.; Guo, J.; Cai, W.; Liu, B.; Wang, Z.; Sun, Z. Surface-Enhanced Raman Spectroscopy Investigation on Human Breast Cancer Cells. *Chem. Cent. J.* **2013**, *7*, 37. https://doi.org/10.1186/1752-153X-7-37.
(51) Cui, L.; Chen, P.; Chen, S.; Yuan, Z.; Yu, C.; Ren, B.; Zhang, K. In Situ Study of the Antibacterial Activity and Mechanism of Action of Silver Nanoparticles by Surface-Enhanced Raman Spectroscopy. *Anal. Chem.* **2013**, *85* (11), 5436–5443. https://doi.org/10.1021/ac400245j.
(52) Farquharson, S.; Smith, W. W.; Lee, V. Y.-H.; Elliott, S.; Sperry, J. F. Detection of Bioagent Signatures: A Comparison of Electrolytic and Metal-Doped Sol-Gel Surface-Enhanced Raman Media. In *Chemical and Biological Early Warning Monitoring for Water, Food, and Ground*; SPIE, 2002; Vol. 4575, pp 62–72. https://doi.org/10.1117/12.456923.
(53) Chen, T. T.; Kuo, C. S.; Chou, Y. C.; Liang, N. T. Surface-Enhanced Raman Scattering of Adenosine Triphosphate Molecules. *Langmuir* **1989**, *5* (4), 887–891. https://doi.org/10.1021/la00088a001.
(54) Pan, J.; Shao, X.; Zhu, Y.; Dong, B.; Wang, Y.; Kang, X.; Chen, N.; Chen, Z.; Liu, S.; Xue, W. Surface-Enhanced Raman Spectroscopy before Radical Prostatectomy Predicts Biochemical Recurrence Better than CAPRA-S. *Int. J. Nanomedicine* **2019**, *Volume 14*, 431–440. https://doi.org/10.2147/IJN.S186226.
(55) Ivleva, N. P.; Wagner, M.; Szkola, A.; Horn, H.; Niessner, R.; Haisch, C. Label-Free in Situ SERS Imaging of Biofilms. *J. Phys. Chem. B* **2010**, *114* (31), 10184–10194. https://doi.org/10.1021/jp102466c.
(56) Cui, L.; Zhang, Y.-J.; Huang, W. E.; Zhang, B.-F.; Martin, F. L.; Li, J.-Y.; Zhang, K.-S.; Zhu, Y.-G. Surface-Enhanced Raman Spectroscopy for Identification of Heavy Metal Arsenic(V)-Mediated Enhancing Effect on Antibiotic Resistance. *Anal. Chem.* **2016**, *88* (6), 3164–3170. https://doi.org/10.1021/acs.analchem.5b04490.

(57) Stone, N.; Kendall, C.; Smith, J.; Crow, P.; Barr, H. Raman Spectroscopy for Identification of Epithelial Cancers. *Faraday Discuss.* **2004**, *126*, 141–157; discussion 169-183. https://doi.org/10.1039/b304992b.
(58) Maquelin, K.; Kirschner, C.; Choo-Smith, L.-P.; van den Braak, N.; Endtz, H. P.; Naumann, D.; Puppels, G. J. Identification of Medically Relevant Microorganisms by Vibrational Spectroscopy. *J. Microbiol. Methods* **2002**, *51* (3), 255–271. https://doi.org/10.1016/S0167-7012(02)00127-6.
(59) De Marchi, S.; Bodelón, G.; Vázquez-Iglesias, L.; Liz-Marzán, L. M.; Pérez-Juste, J.; Pastoriza-Santos, I. Surface-Enhanced Raman Scattering (SERS) Imaging of Bioactive Metabolites in Mixed Bacterial Populations. *Appl. Mater. Today* **2019**, *14*, 207–215. https://doi.org/10.1016/j.apmt.2018.12.005.
(60) Jayan, H.; Pu, H.; Sun, D.-W. Detection of Bioactive Metabolites in Escherichia Coli Cultures Using Surface-Enhanced Raman Spectroscopy. *Appl. Spectrosc.* **2022**, *76* (7), 812–822. https://doi.org/10.1177/00037028221079661.
(61) Yao, G.; Huang, Q. DFT and SERS Study of L-Cysteine Adsorption on the Surface of Gold Nanoparticles. *J. Phys. Chem. C* **2018**, *122* (27), 15241–15251. https://doi.org/10.1021/acs.jpcc.8b00949.
(62) Jing, C.; Fang, Y. Experimental (SERS) and Theoretical (DFT) Studies on the Adsorption Behaviors of l-Cysteine on Gold/Silver Nanoparticles. *Chem. Phys.* **2007**, *332* (1), 27–32. https://doi.org/10.1016/j.chemphys.2006.11.019.
(63) Giese, B.; McNaughton, D. Density Functional Theoretical (DFT) and Surface-Enhanced Raman Spectroscopic Study of Guanine and Its Alkylated Derivatives. *Phys. Chem. Chem. Phys.* **2002**, *4* (20), 5171–5182. https://doi.org/10.1039/B203830G.
(64) Wang, W.; Hynninen, V.; Qiu, L.; Zhang, A.; Lemma, T.; Zhang, N.; Ge, H.; Toppari, J. J.; Hytönen, V. P.; Wang, J. Synergistic Enhancement via Plasmonic Nanoplate-Bacteria-Nanorod Supercrystals for Highly Efficient SERS Sensing of Food-Borne Bacteria. *Sens. Actuators B Chem.* **2017**, *239*, 515–525. https://doi.org/10.1016/j.snb.2016.08.040.