Review

In situ identification of environmental microorganisms with Raman spectroscopy

Dongyu Cui a, b, Lingchao Kong c, Yi Wang a, b, *, Yuanqing Zhu b, c, Chuanlun Zhang a, b, d, e, **

a Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, 511458, China
b Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China
c State Environmental Protection Key Laboratory of Integrated Surface Water-Groundwater Pollution Control, School of Environmental Science & Engineering, Southern University of Science and Technology, Shenzhen, 518055, China
d Shenzhen Key Laboratory of Marine Archaea Geo-Omics, University of Southern University of Science and Technology, Shenzhen, 518055, China
e Shanghai Sheshan National Geophysical Observatory, Shanghai Earthquake Agency, Shanghai, 200062, China

A R T I C L E   I N F O

Article history:
Received 7 January 2022
Received in revised form 13 May 2022
Accepted 15 May 2022

Keywords:
Raman spectroscopy
Environmental microorganisms
Single cells
Metabolic activities

A B S T R A C T

Microorganisms in natural environments are crucial in maintaining the material and energy cycle and the ecological balance of the environment. However, it is challenging to delineate environmental microorganisms' actual metabolic pathways and intraspecific heterogeneity because most microorganisms cannot be cultivated. Raman spectroscopy is a culture-independent technique that can collect molecular vibration profiles from cells. It can reveal the physiological and biochemical information at the single-cell level rapidly and non-destructively in situ. The first part of this review introduces the principles, advantages, progress, and analytical methods of Raman spectroscopy applied in environmental microbiology. The second part summarizes the applications of Raman spectroscopy combined with stable isotope probing (SIP), fluorescence in situ hybridization (FISH), Raman-activated cell sorting and genomic sequencing, and machine learning in microbiological studies. Finally, this review discusses expectations of Raman spectroscopy and future advances to be made in identifying microorganisms, especially for uncultured microorganisms.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Chinese Society for Environmental Sciences, Harbin Institute of Technology, Chinese Research Academy of Environmental Sciences. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Microorganisms are indispensable for maintaining the balance of the ecosystem [1]. Moreover, microorganisms are essential for human health [2], disease treatment [3], food processing [4], and treatment of environmental pollution [5,6]. While advances have been made in the species diversity and compositions of natural microbial populations, the physiological and biochemical properties of environmental microorganisms are still poorly known [7]. The studies of environmental microorganisms mainly include culture-dependent and culture-independent approaches.

Traditional methods can reveal the physiological and biochemical characteristics of isolated microorganisms, which have significantly promoted the development of microbiology. However, more than 99% of microorganisms in the environment are difficult to isolate and grow in pure cultures under current technical conditions [8,9]. Some microorganisms, especially anaerobic microorganisms, grow slowly, increasing the time cost of microbial identification and the difficulty of isolation and cultivation [10]. In addition, ex situ culture changes the physiological status, metabolic characteristics, and interactions with other microbes in the community [11]. Therefore, it is necessary to apply more practical techniques to explore environmental microorganisms, especially uncultivated microorganisms.

The current omics technologies as culture-independent methods can explore the phylogenetic diversity, the metabolic pathways, and the potential physiological and biochemical functions of microorganisms in situ. Metagenomics can directly catalog and analyze all microorganisms in various environments, including...
unknown and uncultured ones [12]. But metagenomics can only infer the potential metabolic pathways of microorganisms because there are many factors affecting gene expression. The results of metagenomic sequencing cannot reveal dynamic properties of spatiotemporal activities and responses to environmental changes [13]. Transcriptomics can snapshot all transcripts present in a cell. Thus, this method can complement metagenomics in verifying the actual gene expression [14,15]. But currently, it is challenging to obtain full-length and low-abundance transcripts due to the high loss rate of mRNA [13]. Metabolomics is closely linked to phenotyping of an organism and intuitive in reflecting the response of living systems to the external environment [16,17]. However, metabolomics usually relies on mass spectrometry to analyze metabolites [18], so it inevitably faces the problems of large sample volume and poor reproducibility of test results. Qualitative analysis of large-scale products based on metabolome data has always been a key issue hindering the development of metabolomics [19].

In short, the current omics technologies brought many dazzling achievements to the field of environmental microbiology [12,13,20]. Metagenomics, transcriptomics, and metabolomics can complement each other in phenotypic and genotypic analyses of microbial populations. However, these technologies are generally biased towards features that represent the overall community. These technologies cannot analyze the proportion of active members and species attribution of metabolites in environmental samples. It is worth mentioning that single-cell multi-omics is on the rise. Although it still faces some challenges, such as low genome coverage in the current single-cell sequencing, the future is expected to witness profound new findings [21].

Raman spectroscopy is an effective method to complement the current omics technologies [22]. Raman spectroscopy is unsuitable for the non-targeted identification of microorganisms or for a comprehensive overview of metabolites, which are the advantages of the current omics technologies. But Raman spectroscopy can detect the metabolic activities and the physiological and biochemical characteristics of environmental microorganisms at the single-cell level [23]. Heterogeneous cells may contain special substances, or different cells may have different proportions of the same substances, which can be analyzed by Raman spectroscopy [24]. Moreover, omics technologies are unable to assist downstream experiments using the same batch of samples because they are destructive. Raman spectra can be interpreted by combining with other approaches or detecting standard samples of the presumed substance.

In addition to omics technologies, Raman spectroscopy can be combined with other biotechnologies to better understand microbial functions [25]. These complementary technologies include but are not limited to stable isotope probing (SIP), fluorescence in situ hybridization (FISH), Raman-activated cell sorting, and genomic sequencing. The combination of Raman spectroscopy and SIP can explore the metabolites and semi-quantify the metabolic activities of microorganisms [26]. Raman spectroscopy combined with FISH can quickly identify target microbes in complex samples by marking the DNA of a specific species [27]. Raman-activated cell sorting can collect target microbes in complex environmental samples according to the biomarker Raman bands [28]. The sorted cells still intact in cellular functions can be sequenced or used for cultivation. Thus, Raman spectroscopy is an effective targeted method to identify uncultivated microorganisms in situ. Combined with different techniques, Raman spectroscopy provides the possibility to obtain Raman fingerprints and whole-genome sequence of uncultured microorganisms, analyze the metabolic preferences of uncultivated species from the environment, and explore the heterogeneity of individuals in a community.

2. Raman spectroscopy

2.1. Principles of Raman scattering

Raman scattering was discovered by the Indian physicist Chandrasekhar Venkata Raman (C. V. Raman) in 1928 [29]. A sample is excited by a monochromatic laser in a Raman spectrometer. The molecules in the sample interact with the excited light to emit scattered light. Most scattered photons have the same frequency as the incident laser, called Elastic scattering or Rayleigh scattering. A few scattered photons (1/10^7) have different frequencies from the incident laser. This scattering is named Inelastic scattering or Raman scattering. The frequency variation of the scattered light depends on the vibrational characteristics of the chemical bonds of molecules in the sample.

The Raman scattering can also be explained by the quantum mechanical model [30] (Fig. 1). The scattering phenomenon can be described as a two-photon process in a quantum mechanical model. First, an incident photon interacts with a molecule. The molecule absorbs the energy and undergoes a transition from the ground state to a higher energy state (excited virtual state). Second, the molecule in the excited virtual state releases a scattered photon after a short time and returns to the ground state. Rayleigh scattering means that the incident photon and the scattered photon have the same energy. In contrast, Raman scattering means the two photons have different energy.

Raman scattering can be divided into Stokes scattering and anti-Stokes scattering [31]. In Stokes scattering, the excited molecule does not completely return to the ground state. The energy of the scattered photon (v0 - v) is less than the incident photon (v0). In anti-Stokes scattering, the molecule has higher energy before being excited. When the excited molecule returns to the ground state, the energy of the scattered photon is v0 + v. Compared with Stokes scattering, anti-Stokes scattering can only be caused by molecules with higher energy. The number of such molecules is limited, resulting in very low signal intensity of anti-Stokes scattering. Therefore, Stokes scattering is usually used as the signal source of Raman spectroscopy to reflect the material composition of the sample.

Raman spectroscopy can be regarded as a collection of molecular vibrations in the sample. Therefore, it can reflect ingredient and chemical information of the sample. Raman spectroscopy has high spatial resolution [32]. The laser diameter is around 1 μm with an objective of 100 × . It can obtain target microbes' physiological and biochemical information at the single-cell level [33,34]. The single-cell Raman spectra can illustrate a “fingerprint” and reveal the composition (such as proteins, nucleic acids, lipids) of each cell [35].

2.2. Common methods of Raman spectroscopy used in microbiology

Up to now, more than 25 different types of Raman spectroscopy have been developed [36]. Confocal Raman spectroscopy is
commonly used for studying microorganisms because of its micron-level resolution and clear image quality. It has excellent potential for detecting single microbial cells in situ due to its non-destructive and water-free performance. Berg et al. [37] used confocal Raman spectroscopy to detect a *Beggiopta* species under different growth conditions and plotted an image of the sulfide distribution in a living cell. They proved that the sulfur in *Beggiopta* under culture conditions is stored in cyclooctasulfur (S₈) and inorganic polysulfides (S₅³⁻). Serrano et al. [38] compared two strains of methanogenic archaea from permafrost and non-permafrost, respectively, by confocal Raman spectroscopy. Although the two strains have a high homology, their chemical compositions are significantly different.

Traditional Raman spectroscopy is relatively insensitive and requires long acquisition times due to the low Raman scattering cross section [39]. Molecular structures that are less abundant in cells are difficult to be detected. When the excitation wavelength is within the range of electron excitation of molecules, the resonant Raman scattering significantly improves the electric field and selectively enhances the Raman signal intensity of pigment-containing substances in microbial cells. Due to the high specificity and selectivity, the detection time and detection limit of intracellular substances are significantly reduced. It helps to detect molecular structures with suitable chromophores, including those less abundant in cells. Therefore, the vibration information of the target substance can be obtained by changing the wavelength of the exciter and combining it with spontaneous Raman spectroscopy. Mukherjee et al. [40] detected a new characteristic peak of *Escherichia coli* at 740 cm⁻¹ under laser excitation at 633 nm using resonant Raman spectroscopy and verified that this peak may belong to the second subunit of cytochrome bd. This is the first study to use a red laser to achieve resonance enhancement in bacteria in situ, laying a foundation for real-time detection of intracellular cytochrome subunit changes and even intracellular respiration without labeling. Miyaoaka et al. [41] detected in situ antibiotic Ampicillin B (Amb) produced by the marine bacteria *Streptomyces nodosus*. The Raman intensity of Amb correlates with the abundance of Amb in cells, which proves that Raman spectroscopy is an effective method for identifying antibiotic producers and (semi-) quantifying intracellular biological antibiotics [41].

Surface-enhanced Raman spectroscopy can improve the electric field and molecular polarizability through the interaction between electromagnetic waves and metal, thus effectively enhancing the Raman signal [42]. This technique was first invented by Fleischmann et al. [43] in 1973. Metals (such as gold, silver, or copper) that are stable in the air are usually used as signal amplification materials. The strength of output signals is related to the type, particle size, and roughness of metals. Bodelein et al. [44] used surface-enhanced Raman spectroscopy to explore the metabolic changes of two pure culture strains, *Chromobacterium violaceum* (C. violaceum) and *Pseudomonas aeruginosa* (P. aeruginosa), in coculture. Co-cultivation induces *C. violaceum* to produce violacein and inhibits the secretion of pyocyanin produced by *P. aeruginosa* due to the adverse effects of pyocyanin on the growth of *C. violaceum*. Liu et al. [45] designed a silver nanorod as a base of surface-enhanced Raman spectroscopy, detected 22 kinds of common pathogens, and obtained Raman spectra of these bacteria. Then the Raman spectra were effectively distinguished by various mathematical methods, which built a foundation for rapidly identifying pathogens from a complex environment. Other studies using surface-enhanced Raman spectroscopy are highlighted by Laucks et al. [46] who detected five polar psychrophilic marine bacteria and Xu et al. [47] who detected and identified seven kinds of marine pathogen *Vibrio parahealmolgyticus*.

### 2.3. Advantages of Raman spectroscopy in studying environmental microorganisms

Raman spectroscopy can measure all molecular vibrations of a single microbial cell [48]. A single-cell Raman spectrum contains more than 1000 Raman bands reflecting the information of macromolecules, such as proteins, nucleic acids, lipids, and carbohydrates. The width of the Raman band is related to the structure and crystal form of the compound. Raman bands are much broader when detecting objects in solution. But the vast majority of Raman bands overlap each other on Raman spectra because the bands are concentrated in the fingerprint region. Only the most intense Raman peaks could be explained. Nonetheless, Raman spectra can still be used as effective microbial fingerprints, especially when ignoring individual Raman peaks and looking at a Raman spectrum as a whole.

The advantages of Raman spectroscopy in environmental microbiology are as follows:

1. **Detection at the single-cell level**

   The magnification of Raman microscopes can reach 100 times, the spot diameter of the laser can be less than 1 μm, and the resolution can be up to 0.4 μm. The diameter of most microbes is about 1 μm. Therefore, the Raman spectra of microorganisms can be obtained at the unicellular level and reflect physiological and biochemical information of microbes.

2. **Non-destructive to sample material**

   Raman spectroscopy can be a non-destructive method if the parameters are suitable. A laser with sufficient energy and adequate detection time is required to obtain obvious Raman signals. But excessive energy or detection time can easily lead to the destruction of chemical structures in cells, failing to collect Raman signals. During the first detection, the user needs to gradually increase the energy and detection time from low to high to obtain high-quality Raman signals without destroying the sample.

3. **Rapid detection with micro samples**

   In order to exceed the detection limit of compounds, most detection methods require the collection of sufficient target microbes in the environment. In contrast, Raman spectroscopy requires only a small volume of samples to detect compounds within a single cell.

4. **No interference by water**

   Microbial samples usually need to be processed in liquid. The strong infrared absorption peak of water will affect the spectral analysis for other substances in the microbe, so liquid samples are not suitable for infrared spectroscopy. Raman spectroscopy is an ideal technique for studying aqueous materials in that the Raman signal strength of water is very weak due to the asymmetry of chemical bonds of water molecules.

### 2.4. Analyses of single-cell Raman spectra (SCRS)

Based on the above advantages, Raman spectroscopy, combined with other technologies, can detect target single cells in environmental samples in situ, analyze the phenotypes [2], and perform subsequent genomic sequencing. Cells from pure cultures can be revived and reproduced after Raman detection and cell sorting [49]. The SCRS can be divided into three regions: fingerprint region
The Raman bands of most biological macromolecules are displayed in the fingerprint region, reflecting important physiological and biochemical information of a single cell (Table 1). For example, three types of amide bonds could reflect the secondary structure of the protein: Amide I (1660 cm⁻¹), Amide II (1574 cm⁻¹), and Amide III (1235 cm⁻¹) [50]. Five bases of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) have strong Raman bands in the fingerprint region: adenine (735 cm⁻¹), cytosine (787 cm⁻¹), guanine (651 cm⁻¹), and uracil (771 cm⁻¹). The characteristic Raman bands of exocyclic and endocyclic ring deformation of a monosaccharide are located at 540 cm⁻¹ and 410 cm⁻¹, and the characteristic Raman bands of the C=O–C4 glycosidic bond in glycogen are located at 480 cm⁻¹. In addition, some biological macromolecules have special Raman bands, which can be used as biomarkers of the Raman spectrum, such as poly-β-hydroxybutyrate (PHB) (C=O stretching of the ester functional group at 1735 cm⁻¹) [51], carotenoids (C–CH₃ deformation at 1003 cm⁻¹, C=C stretching at 1155 cm⁻¹, C=C stretching at 1512 cm⁻¹) [52], rhodopin (the C=C stretching at 1187 cm⁻¹, the C=C–H in-plane rocks at 1237 cm⁻¹) [53], cytochrome C (pyrrole breathing at 750 cm⁻¹) [54], and green fluorescent protein (GFP) (the C=C–N=–C=N= portion of the imidazolone ring at 1552 cm⁻¹) [55]. Thus, these notable Raman bands can be used to rapidly identify microorganisms from the community and for semi-quantitative enumeration of important intracellular components.

Raman bands in the high wavenumber region usually represent −CH₂, −CH₃, −NH, and OH groups, which are very obvious and show up in Raman spectra of all microbes, such as the −CH₂ symmetric stretch vibration near 2850 cm⁻¹ and the Fermi resonance −CH₂ stretch vibration near 2886 cm⁻¹. Raman bands of these groups overlap strongly in the high wavenumber region, so there is generally little effective information that can identify biological macromolecules [87]. Proteins and lipids are mainly assigned at 2800–3000 cm⁻¹. The relative contribution of DNA is small [88]. The Raman band of water overlaps with that of the OH group of carbohydrates at 3300–3500 cm⁻¹ and with that of the N–H group of proteins at 3329 cm⁻¹ [89]. In addition, the Raman band of the −CH vibration (3060 cm⁻¹) from the unsaturated fatty acids is different to that of the −CH bond vibration (2880 cm⁻¹) from saturated fatty acids. The ratio of the integral area of the two bands can represent the ratio of the unsaturated fatty acid to the total fatty acids in the cell, which is commonly used to characterize intracellular lipids.

### 3. Raman spectroscopy in combination with other technologies

#### 3.1. Raman–SIP combination for analyzing the phenotypes of environmental microorganisms

Microorganisms gain energy by coupling electron donors and electron acceptors [90]. The principle of SIP is that microorganisms can incorporate the heavier isotope of an atom (¹³C vs. ¹²C, for example) into cell components when adding substrates labeled with the ¹³C isotope to environmental samples [91]. SIP is a powerful method to explore the physiological information and functional properties of individual microbes from complex environments [92,93]. As described above in the principle of Raman spectroscopy, the absorbance of Raman spectra is related to molecular vibration. The molecular mass increases when an atom is enriched by a heavier isotope, whereas the molecular vibration frequency decreases and the absorbance of the Raman peak shifts to the left. Raman–SIP can detect whether a given substrate is degraded and which microbes incorporate the heavier isotope. Redshifts caused by these changes can be found in Raman spectra of microorganisms labeled by the heavy isotope (Fig. 3). Therefore, this method can verify potential metabolic pathways and metabolic preferences of individuals in communities at the single-cell level, which contributes to understanding the ecological role of individual microbes from natural habitats [94–101].

Heavy water (D₂O) is an important substance to measure the metabolic activity of microbes. Moreover, compounds labeled by stable isotope deuterium (D) are commonly used to explore the ability of microbes to metabolize specific substances [103]. In general, the silent region of SCRIS has no Raman bands. The C–D band (2040–2300 cm⁻¹) in the “silent region” is easy to observe. Microbes with biological activity could incorporate D of D₂O into the active cells through a H/D exchange reaction mediated by NADPH electron transfer chain. The C–D band can be observed in Raman spectra, which could semi-quantify the metabolic activity of microorganisms [94]. For fast-growing E. coli cells, the C–D band is detectable after 20 min of labeling [26]. D-labeled substrates have been used to analyze the products of several complex carbohydrates degraded by cecal microbes in mice [94]. D-glucose has been used as the sole carbon source to monitor lipid production in living cells [104]; D–amino acids have been used as substrates to detect the newly synthesized proteins and their proportion in total proteins [105,106]. In addition, if bacteria are still metabolically active in an environment with antibiotics, the bacteria can absorb heavy water, and the C–D band can be observed on their Raman spectra. Therefore, antibiotic-resistant bacteria in environmental samples can be detected by adding D₂O and antibiotics at the lowest inhibitory concentration. This method has been applied to study antibiotic-resistant bacteria in the human gut and surface water samples [2,103].

¹⁴N– and ¹³C–labeled substrates are also commonly used to explore the metabolic characteristics of microorganisms. Studies have found that using ¹³C–glucose with different proportions as the sole carbon source of Pseudomonas fluorescens, redshifts of the Raman peaks are found in the fingerprint region of their Raman spectra. These redshifts are mainly manifested at the peak of phenylalanine and increase with the amount of ¹³C absorbed in the cell biomass [107]. ¹³C–naphthalene was used as the sole carbon source to cultivate Pseudomonas. It was found that Pseudomonas has a degrading effect on naphthalene in groundwater biofilms [27].
Cui et al. [102] used surface-enhanced Raman spectroscopy in combination with the stable isotope $^{15}$N to monitor different bacteria from multiple angles to assess their ability to assimilate various nitrogen sources. Raman signal intensity after redshift was linearly correlated with the content of $^{15}$N assimilated by single cells and was not affected by the type and concentration of nanoparticles used in this method. Cui et al. [95] cultivated soil samples by $^{15}$N$_2$ and identified nitrogen-fixing bacteria in the soil. Milucka et al. [108] used confocal Raman spectroscopy combined with FISH, SIP, and other methods to verify that anaerobic oxidation of

### Table 1

| Wavenumber (cm$^{-1}$) | Biological Assignment | Vibrational Assignment | Reference |
|------------------------|-----------------------|------------------------|-----------|
| **Fingerprint Region (400–1800 cm$^{-1}$)** | | | |
| 642 | Tyrosine | C–C twisting | [56] |
| 650 | Amide IV | OCN bending | [50] |
| 651 | Guanine | Ring stretching | [57] |
| 700 | Amide V | Out-of-plane NH bending | [50] |
| 728 | Tryptophan | Ring breathing | [58] |
| 735 | Adenine | Ring breathing | [57] |
| 748 | Tryptophan | Indole ring bending | [59] |
| 750 | Cytochrome C | Pyrrole breathing | [54] |
| 771 | Uracil | Ring breathing | [60] |
| 776 | Thymine | Ring breathing | [57] |
| 787 | Cytosine | Ring breathing | [57] |
| 813 | A-type RNA of the viroid | C–O–P–O–C symmetric stretching | [61] |
| 840–860 | Polysaccharide structure | 1,4-glycosidic link stretching | [62,63] |
| 875 | Tryptophan | out-of-plane bending of indole ring and indole CH | [59] |
| 951 | Proteins (α-helix) | CH$_2$ symmetric stretching | [64] |
| 1000 | Deoxyribose | C–O stretching | [65] |
| 1003 | Carotenoid | C–CH$_3$ deformation | [52] |
| 1004 | Phenylalanine | Benzene ring breathing | [50] |
| 1043 | HCO$_3^-$ in sodium bicarbonate | symmetric stretching | [66] |
| 1064 | Lipids | Skeletal C–C stretching | [67] |
| 1078 | Glycogen | C–C stretching | [68] |
| 1093 | DNA | PO$_2^-$ symmetric stretching | [69] |
| 1123 | Lipids and proteins | C–C stretching | [70] |
| 1155 | Carotenoid | C–C stretching | [52] |
| 1157 | Carotenoid | In-phase vibrations of the conjugated –C–C– | [71] |
| 1172 | Tyrosine | C–H in-plane bending | [56] |
| 1187 | Bacteriorhodopsin | C–C stretching | [72] |
| 1199 | Tyrosine | Ring breathing | [58] |
| 1228 | DNA | PO$_2^-$ antisymmetric stretching | [56] |
| 1235 | Amide III | In-phase combination of the NH bending and CN stretching | [50] |
| 1237 | Bacteriorhodopsin | C–C–H in-plane rocks | [72] |
| 1260 | Lipids | CH$_2$ in-plane deformation | [73] |
| 1267 | Lipids | HC– in–plane deformation | [74] |
| 1300 | Lipids | In-plane twisting | [62] |
| 1304 | Lipids | CH$_2$ deformation | [67] |
| 1321 | Lipids | CH$_2$ deformation | [75] |
| 1358 | Tryptophan | Indole ring bending | [76] |
| 1420–1450 | Lipids | CH$_3$ scissoring | [77] |
| 1420–1470 | Proteins and lipids | CH$_3$ bending | [78] |
| 1440 | Lipids | CH$_2$ and CH$_3$ deformation | [79] |
| 1446 | Proteins and lipids | CH$_2$ bending | [56] |
| 1512 | Carotenoid | C–C stretching | [52] |
| 1517 | β-Carotene | C–C stretching | [80] |
| 1542 | Tryptophan | Indole ring stretching | [59] |
| 1574 | Amide II | Out-of-phase combination of the NH bending and the CN stretching | [50] |
| 1625 | Chlorophyll | Stretching of conjugated vinyl groups | [81] |
| 1660 | Amide I | C–O stretching vibrations out-of-plane C–N stretching | [50] |
| 1660 | Chlorophyll b | Stretching of formyl carbonyl group | [81] |
| 1735 | PHB | C–O stretching of the ester functional group | [82] |
| **Silent Region (1800–2700 cm$^{-1}$)** | | | |
| 2040–2300 | | C–D vibrations | [83] |
| **High Wavenumber Region (2700–3600 cm$^{-1}$)** | | | |
| 2700–3100 | Lipids and amino acid side chains of proteins and carbohydrates | C–H stretching | [84] |
| 2879 | Lipids and proteins | CH$_2$ and CH stretching | [85] |
| 2910–2965 | Proteins | CH$_2$ stretching | [86] |
| 3350–3550 | Water | OH stretching | [86] |
methane coupled with sulfate reduction in marine sediments was completed alone by methanotrophic archaea. Meanwhile, sulfate-reducing Deltaproteobacteria were first observed to dismutate zero-valent sulfur into sulfur dioxide. These results have important implications for understanding the global carbon and sulfur cycle.

Raman–SIP can identify photosynthetic bacteria in environmental samples and explore their carbon fixation ability and the heterogeneity of single cells in the population. Li et al. [109] detected single cells of $^{13}$C–labeled photosynthetic algae and photosynthetic bacteria by resonance Raman spectroscopy; the results showed that carotenoids are good biomarkers for Raman spectroscopy. The redshift of the Raman band can be detected when the $^{13}$C content of a single cell is only 10%. It is a non-invasive method to quickly and quantitatively detect the carbon fixation ability of single cells. On this basis, Jing et al. [110] isolated two typical bacteria, *Synechococcus* spp. and *Pelagibacter* spp. with redshift bands of $^{13}$C–carotenoid in seawater by Raman-activated cell ejection and linked phenotypes to genotypes of the carbon sequestrating capacity of uncultured marine microbes by reconstructing their genomes.

### 3.2. Raman–FISH combination for identifying the target microorganisms from natural environments

FISH can use fluorophore-bearing nucleotide fragments as probes to hybridize to 16S DNA or rRNA of target microorganisms, making it fluoresce under a fluorescence microscope, thus allowing us to identify a specific type of microorganisms from complex environmental samples [111,112]. Raman spectroscopy can be combined with FISH to obtain Raman spectra of target microbes in environmental samples and analyze their physiological and biochemical characteristics based on these Raman spectra. The traditional FISH-microautoradiography relies on radioisotopes such as $^{14}$C as substrates to label intracellular biological macromolecules, which has been used to detect the *in situ* absorption of the target microorganisms in complex environmental samples [113]. By contrast, as shown in Fig. 4, Raman–FISH does not reply on radioisotopes and can be a powerful method for non-destructive exploration of the physiological and biochemical characteristics of uncultivated microorganisms [27].

Microorganisms have different metabolic pathways for the same substrate under different environmental conditions. Raman–FISH can quantify the metabolism of target microorganisms *in situ*. Studies have shown that *Pseudomonas* can degrade naphthalene in groundwater [27]. Gordon et al. [114] used this technique to study the differences in the growth rate of photosynthetic autotrophic bacteria within and between species under natural conditions and different culture conditions. Escudero et al. [115] verified the strong correlation between the abundance of *Acidovorax* and the distribution of pyrite by Raman–FISH to explore the interaction between deep underground microbes and minerals, indicating that ore will have an important impact on the distribution of underground microbes.

Some artificial environments, such as sewage treatment plants and bioreactors, use microorganisms to degrade or enrich pollutants, including enhanced biological phosphorus removal and microbial nitrogen fixation. According to the Raman signal intensity, Raman spectroscopy combined with quantitative FISH can quantify the proportion of species in the environment and detect the metabolic activity of species against specific elements or compounds. Fernando et al. [116] detected the phosphorus-accumulating microorganisms in the wastewater treatment systems of eight factories. In addition to *Candidatus Accumulibacter, Tetrasphaera*, which does not show the classic phenotypic characteristics of phosphorus-accumulating microorganisms, is also an important participant in enhanced biological phosphorus removal. In this study, Fernando and colleagues developed a method based on Raman–FISH to quantify intracellular polyphosphate (poly–P). Since the Raman signal intensity of poly–P at 1170 cm$^{-1}$ is linearly related to the poly–P density after drying standard solutions of different concentrations, the intracellular poly–P content could be calculated by detecting the Raman signal intensity of target cells at 1170 cm$^{-1}$. This method can explore the individual contribution of a variety of phosphorus-accumulating microorganisms during the comprehensive action and has reference significance for quantifying the material cycle of individual species in various ecosystems [116].

### 3.3. Raman activated cell sorting in combination with genomic sequencing for enhancing research in environmental microbiology

SCRS as microbial Raman fingerprints can reflect biomolecular vibrations inside specific single cells. By comparing with Raman spectra of putative substances, it is possible to explore the phenotypic characteristics of microorganisms to establish the association between phenotypes and genotypes. Meanwhile, based on the fingerprint region of SCRS, the targeted microbes can be sorted and collected by Raman-activated cell sorting. The collected cells can be genome amplified and sequenced for higher quality. High-quality genomes are constructed for analyzing potential metabolic pathways. The advantage of Raman-activated cell sorting is that the single cells of target uncultivated microbes can be isolated and collected from environmental samples without external labeling and then be sequenced or cultured [28]. Compared to traditional
methods, single-cell genomic sequencing can get a complete genome of a single uncultivated species, and single-cell culture can avoid a complex screening process.

Up to now, different target microorganisms, such as antimicrobial-resistant bacteria and pathogenic bacteria, have been isolated from mixed flora, surface water, seawater, mouse intestinal tract, human intestinal tract, and human oral cavity. Their species identification and functional gene analysis have been carried out [103,117,118]. As shown in Fig. 5, there are three main types of Raman activated sorting technologies: Raman-activated cell ejection, Raman tweezers, and microfluidic cell sorting [28].

Raman-activated cell ejection is a type of Raman-activated cell sorting coupled with a Raman spectroscopy to the cell isolation system (Fig. 5a), Wang et al. [117] used two independent approaches to achieve single-cell sorting: A 337 nm pulsed laser was applied to the thin water layer around a cell. The pulsed laser energy can evaporate the thin water layer and give the cell energy to move forward. The cell is then pushed from the slide onto the collection tube cover by a laser-induced forward transport device to achieve single-cell separation. Song et al. [118] used an all-in-one system to sort and collect single cells. They sequenced the obtained cells, and new functional genes were discovered in single-cell genomics. It is the first time that Raman-activated cell ejection coupled with single-cell genomics to explore uncultivated microorganisms.

Optical tweezers are optical traps integrated by laser focusing. Optical tweezers can capture micron-sized particles in fluids. When the microbial cells flow through the optical trap, they are fixed in the optical trap with non-contact. The cells can move with the position of the optical trap controlled by the laser source. Raman tweezers is the combination of optical tweezers and Raman spectroscopy (Fig. 5b). It can capture single target cells for Raman detection and monitor their dynamic changes between bacteriophages and the host. Traditional phage detection methods require at least 45 min, while Pilat et al. [119] detected phage replicated in bacterial cells within 5 min after introducing the phage into the host cell. This provides a fast, non-destructive, and real-time detection method for understanding the mechanism of phage infection. Huang et al. [49] used Raman tweezers to identify and separate E. coli, Saccharomyces, and Pseudomonas in the mixture, then transferred the sorted target bacteria into the capillary tube. The accuracy of the Raman tweezer was verified by genome amplification, and the activity of separated cells was verified by isotope labeling and cell culturing.

The microfluidic cell sorting technique uses micron-sized pipes to create a microfluidic environment (Fig. 5c). It integrates microbial cell detection, separation, cultivation, and other independent operation processes into one chip. The chip can carry out many tasks in liquid, such as accurate cell counting, Raman detection, and single-cell sorting. The biomarker Raman bands of target microbes can be used as screening conditions to achieve rapid and high-throughput sorting of target species in environmental samples [120]. Mcilvenna et al. [120] developed a high-throughput sorting system that does not require fixed cells for Raman detection, simultaneously performing Raman signal acquisition, real-time identification, and cell sorting. They used photosynthetic bacteria as a type strain, and the sorting accuracy was 96.3%. Liao et al. [121] simplified the operation of the surface-enhanced Raman spectroscopy antimicrobial drug sensitivity testing by using a microfluidic system. They could separate the drug-sensitive and drug-resistant strains within 3.5 h and conduct the experiment in situ by changing the reagent on the chip to change the liquid environment and analyzing the bacterial response.

3.4. Single-cell Raman spectra of environmental microorganisms analyzed by machine learning

Machine learning refers to an algorithm model that completes certain functions through data training and is currently the mainstream artificial intelligence implementation method [122]. Machine learning has been successfully applied to the classification, clustering, and prediction of large data sets in many fields [123–125]. The composition and physiological and biochemical characteristics of environmental microorganisms are relatively complex, and traditional data processing cannot meet the needs of data analysis [126]. Many data visualization methods have been
applied to the study of Raman spectroscopy. Due to the high sensitivity of Raman spectroscopy, when detecting two very similar cells, the Raman bands cannot be directly analyzed by simple data processing. To get accurate results, data visualization methods such as two-dimensional correlation analysis and linear discriminant analysis help us discover minor differences in Raman spectra. Raman spectroscopy can also connect with artificial neural networks [127], support vector machines [128], random forests [129], AdaBoost [130], and other algorithm models [131]. Herein, the Raman spectroscopy can classify and predict different microorganisms, extract their features, and analyze their “Raman fingerprints” in depth. These methods exhibit high efficiency and precision, possessing great potential in rapidly identifying uncultured pathogens and uncultured marine microorganisms [132–136].

Raman spectroscopy combined with machine learning has identified pathogens from the environment [134–136]. Ho et al. [134] used a convolutional neural network to identify 30 pathogens in low signal-to-noise ratio Raman spectra with an accuracy rate of over 82%, and identification accuracies of corresponding antibiotic treatment reached 97%. Yang et al. [135] compared the clustering and identification capabilities of various machine learning algorithms for nine species of Staphylococcus. Among these methods, density-based spatial clustering of applications with noise has the best clustering ability (Rand index 0.9733), and the convolutional neural network has the highest accuracy rate of identification (accuracy at 98.21%, area under the curve at 99.93%). Wang et al. [136] used confocal Raman spectroscopy (785 nm) combined with neural network technology to classify and identify Arcobacter. Arcobacter was selected from Campylobacter and Helicobacter by cluster analysis based on Raman spectra, and 18 species of Arcobacter were distinguished at the species level using the convolutional neural network, with an accuracy rate of 97.2%. They also constructed a fully connected artificial neural network based on SCRS to determine the actual proportion of specific Arcobacter in the bacterial mixture. These studies provided references for determining the proportion of target pathogens in environmental samples.

The combination of Raman spectroscopy and machine learning has also made progress in marine microbial research. Yu et al. [136] combined Raman spectroscopy with long and short-term memory neural networks. After optimizing the model parameters, they trained the model based on eight strains isolated from the marine organism Urechis unicinctus, and the accuracy of classification prediction exceeded 94%. Yu et al. [137] also fitted the Raman spectra of three marine strains (Staphylococcus hominis, Vibrio alginolyticus, and Bacillus licheniformis) to generative adversarial networks. The model only needed 100 SCRS of each strain. After being iteratively trained, the classification accuracy of these strains reached 100%. Liu et al. [139] collected ten species of marine actinomycetes and two species of non-marine actinomycetes, and detected their Raman spectra. They used one-dimensional convolutional neural networks and several traditional classification algorithms (support vector machines, principal component analysis, etc.) to classify the original data with an accuracy rate of 95%, and one-dimensional convolutional neural networks is preferred for processing the raw data. Heraud et al. [140] measured DunalieIIe Tertoiocta cells with a 780 nm laser. The enhanced Raman bands of Chlorophyll a and beta-carotene in Raman spectra are indicative of whether nitrogen limitation exists. They successfully predicted more than 90% of the cellular nutritional status by partial least squares discriminant analysis and other multivariate classification methods. Our laboratory obtained the Raman spectra of nine types of archaea (seven halophilic archaea, one thermophilic archaeon, and one marine archaeon) by confocal Raman spectroscopy. We predicted the target microorganisms in the mixture of these organisms and verified the results by Raman-activated cell ejection and genomic sequencing. The accuracy rate of support vector machines was greater than 88%. It also was the first time that Raman–FISH is applied to obtain the Raman spectra of uncultured marine group II archaea [133].

4. Conclusion and outlook

Microorganisms are ubiquitous in the environment. They are the main driving force of the global biogeochemical cycles and are associated with almost all multicellular life forms [141]. Most microorganisms from nature cannot be studied in pure culture. These microorganisms are called microbial “dark matter”. Although microbial “dark matter” has been brought to light by omics technologies, the physiological and biochemical research on microorganisms cannot be completely achieved only by metagenomic data analyses or cultivated strains. Moreover, it is essential to analyze the physiological and biochemical characteristics of individual microbial cells in situ because of the interaction among microorganisms in nature and the microbial host-specific metabolism. Detection at the single-cell level enables Raman spectroscopy as a method to complement the current omics technologies. Raman spectroscopy performs near-in-situ analysis of microbial phenotypes in the environment. Combined with downstream single-cell separation and sequencing, Raman spectroscopy provides new insights for the research of environmental microbiology. It enhances the possibility of deciphering environmental microbial “dark matter”.

Raman spectroscopy in combination with other biotechnological methods is further expanding its range of application. Raman spectroscopy combined with SIP and FISH avoids the deficiency of simply detecting the material composition of cells. These integrated methods can accurately identify target microbes from the environment and quantify the metabolic activity of single cells on specific elements or nutrients according to the Raman bands of isotope-labeled substances.

Furthermore, with the increasing development of artificial intelligence, the combination of Raman spectroscopy and machine learning can improve the speed and accuracy of identifying target microbes. This combination promises rapid screening and classification of uncultivated bacteria and archaea and helps to visualize and fingerprint them from the natural environment.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was financially supported by the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (No. K19313901), the National Natural Science Foundation of China (91851210, 42141003), the State Key R&D project of China grant (No. 2018YFA0605800), the Stable Support Plan Program of Shenzhen Natural Science Foundation (202009252173954005), the Guangdong-Shenzhen Joint Fund (201915121010080), the Shenzhen Key Laboratory of Marine Archaea Geo-omics, Southern University of Science and Technology (ZDSYS201802081843490), and the financial support from China Postdoctoral Science Foundation (2020M682769).
spectroscopy combined with generative adversarial networks, Sci. Total Environ. (2020) 726, https://doi.org/10.1016/j.scitotenv.2020.138477.

[138] K.D. Wang, L. Chen, X.Y. Ma, L.N. Ma, K.C. Chou, Y.K. Cao, L.H. Khan, G. Golz, X.N. Lu, Arcobacter identification and species determination using Raman spectroscopy combined with neural networks, Appl. Environ. Microbiol. 86 (20) (2020), https://doi.org/10.1128/AEM.00924-20.

[139] Y.Y. Liu, J.J. Xu, Y. Tao, T. Fang, W.B. Du, A.P. Ye, Rapid and accurate identification of marine microbes with single-cell Raman spectroscopy, Analyst 145 (9) (2020) 3297–3305, https://doi.org/10.1039/c9an02069a.

[140] P. Heraud, J. Beardall, D. McNaughton, B.R. Wood, In vivo prediction of the nutrient status of individual microalgal cells using Raman microspectroscopy, FEMS Microbiol. Lett. 275 (1) (2007) 24–30, https://doi.org/10.1111/j.1574-6968.2007.00861.x.

[141] C.L.L. Zhang, W. Xie, A.B. Martin-Cuadrado, F. Rodriguez-Valera, Marine Group II Archaea, potentially important players in the global ocean carbon cycle, Front. Microbiol. 6 (2015), https://doi.org/10.3389/fmicb.2015.01108.