Published on Feb 27, 2025 by Arcadia Science **DOI**: 10.57844/arcadia-cd7e-443b

DIY Raman spectroscopy for biological research

We optimized an open-source Raman spectrometer for solid and liquid biological samples, including microorganisms, organic solvents, and biochemicals. Here, we share a calibration protocol, data processing notebooks, sample prep methods, and a nascent spectral reference library.

Contributors (A-Z)

Audrey Bell, Ben Braverman, Keith Cheveralls, Megan L. Hochstrasser, Erin McGeever, David G. Mets, Sunanda Sharma

Version 4 · Mar 31, 2025

Purpose

In the field of biology, researchers have historically gained rich scientific insights by observing the interaction between light and matter. Optical microscopy and spectroscopy fundamentally require relatively few components — namely, a light source, a detector, and a sample. However, capturing the right photons to interrogate a biological sample meaningfully can be challenging, especially if it's dynamic or living. Here, we began with an open-source spontaneous Raman spectrometer (preliminarily used to study chili, beer, and algae in a hackathon **[1]** and optimized it for biological samples. Raman spectroscopy is a label-free vibrational optical spectroscopy method that can reveal molecular composition, structure, and environmental information. We tested sample preparation, calibration methods, and stage configurations to optimize

the Raman signal from various samples, including media, reagents, and cells in liquid and solid cultures. We're sharing resources for optimizing this inexpensive and easily fabricated Raman spectrometer for biology: a calibration protocol, Jupyter Notebooks with Python code for applying calibration and data processing, notes on troubleshooting the system and optimizing biological sample signal, and a preliminary spectral library. We hope biologists interested in exploring a rapid approach to collecting high-dimensional information about the chemical composition of a sample will find these materials helpful. Biologists and biochemists - from students to professional researchers — can build this system and apply our methods and code to analyze biological and living samples.

- All associated **code** for analyzing the spectral data is available in this GitHub repository.
- **Data** from this pub, including the raw and processed spectra, are available in the "data" folder on GitHub. The spectral library from this pub is available in a CSV file in the "spectral_library" folder.
- A **protocol** we created for "Calibration of the OpenRAMAN DIY Raman spectrometer" is available on protocols.io.

The strategy

Biologists have increasingly used Raman spectroscopy to collect spatially and temporally resolved information about life and its processes **[2][3]**. Given that little to no sample preparation is required, Raman applies to a wide range of dynamic systems. When monochromatic light is focused on a sample, the sample absorbs, reflects, or scatters the photons. A small percentage of these photons scatter inelastically, which means their energy and wavelength change through interaction with the sample. These slight energy shifts, or Raman shifts, indicate the vibration of specific chemical bonds in the sample (Figure 1). Researchers have used this technique to assess phenotypic heterogeneity in bacteria and yeast **[4]**, mammalian cells **[5]**, plants **[6]**, filamentous fungi **[7]**, and protists **[8]**. Furthermore, Raman spectroscopy is promising as a label-free method of tracking metabolic activity **[9][10]**, even at the scale of a single cell **[11]**, and can be used to probe specific mechanisms such as cell

inflammation **[12]**. The field has recently expanded to link Raman spectroscopy with bioinformatics tools to enable spatially-resolved, systems-level “spectromics” on cells **[13]**.

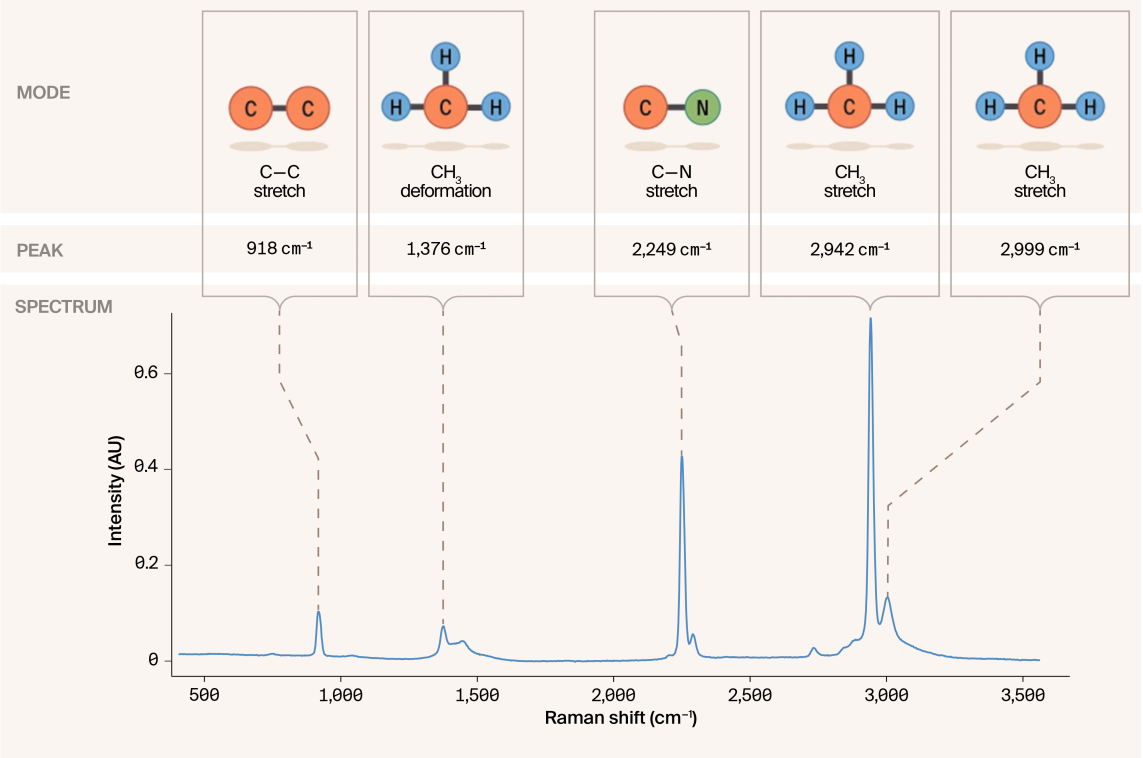

Figure 1

Introduction to Raman spectra through an overview of acetonitrile.

Raman is a spectroscopy technique in which each peak in a spectrum corresponds to vibrational modes of a specific molecular bond in the material. This overview figure shows how the peaks in acetonitrile, a common reference material, correspond to various vibrational modes. Data are from 2024-10-11.

Raman spectroscopy can also capture dynamic changes in samples across time points or in real time. For instance, researchers have used the technique to study the degradation of nanocarrier drug-delivery systems **[14]**, molecular changes in human lung carcinoma epithelial cells **[15]**, and to monitor enzyme-catalyzed reactions **[16]**. As labels are unnecessary and acquisition times can be short, this technique has special relevance in observing a changing living system with comparatively little risk of altering that system.

While many published works on Raman spectroscopy use expensive commercial or custom systems, there are a few examples of low-cost Raman systems. We previously built one of these, OpenRAMAN (“Starter Edition”), to explore rapid analysis of biological samples. This system has two configurations: the solid cuvette, which has a sample stage, and the liquid/standard cuvette, which has a tube holder (Figure 2). The system is < $3,500 (USD), has a detailed build guide, an active user community, and has accompanying open-source software available. However, it hasn't been used extensively for biological applications or to capture dynamic phenotypes. We sought to improve our implementation of this DIY Raman system and demonstrate its utility for biological research.

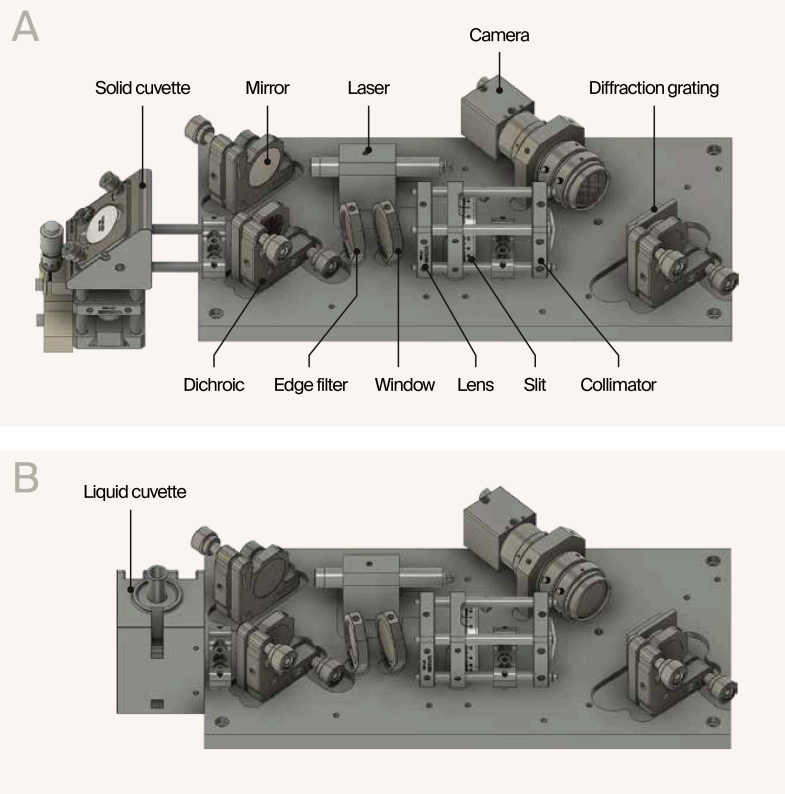

Figure 2

Solid and liquid configurations of the OpenRAMAN (Starter Edition).

(A) Schematic of the solid configuration used for capped liquids, powders, minerals, dried solutions, and solid cultures.

(B) Schematic of the liquid configuration used for samples in borosilicate tubes.

The problem

In our first implementation of the OpenRAMAN system, data were easy to acquire but didn't contain many Raman peaks that could be used for analysis. In addition, the system needs to be better calibrated to interrogate samples meaningfully, compare spectra across samples, and compare them to published literature. We reviewed data from that implementation, including 2D images from the CMOS camera and the 1D

spectra, to identify where we could improve the system. We observed significant background noise, likely from stray light, and broad, aberrant lines from a neon bulb. Such a source should generate clear lines in a 2D image, translating into sharp, high-amplitude peaks in a spectrum. Furthermore, the Raman spectra collected previously had broad peaks and possible fluorescence. Together, these observations suggest that the optical path wasn't optimized.

Our solution

In this follow-up work, we had several goals:

1. To improve the quality of the acquired Raman spectra
2. To create a calibration workflow using known reference materials
3. To assess the performance of the system regarding spectral resolution and positional accuracy of peaks
4. To develop methods for preparing biological samples
5. To collect an initial library of biologically relevant spectra
6. To observe dynamic phenomena in living cells.

To achieve these goals, we realigned the system and developed procedures to measure its calibration and performance. We collected reproducible data on samples relevant to biological research with sufficient spectral resolution to distinguish Raman features. Through this effort, we demonstrated that this low-cost system can successfully support biological investigations.

System optimization

In addition to Raman scattered photons, the spectrum of any given sample potentially contains signal and noise from many other sources. Sample fluorescence, emissions from the optical components, environmental light and cosmic rays, and noise sources such as shot noise, readout noise, fixed pattern noise, and dark noise can all be present to varying degrees **[17][18]** and decrease signal quality. An optimized system aims to *maximize* the number of Raman-scattered photons from your sample that reach the detector and *minimize* all other photons.

In our system, sample illumination generated by a 532 nm (green) laser is reflected by mirrors and focused through a lens onto the sample surface. A small percentage (up to one in 107) of photons are scattered back with different energy from the incident light (Raman-scattered) and return through the sample path along with light that's the same energy as the laser (Rayleigh-scattered). The returning light passes through a dichroic mirror and filters that reject most of the Rayleigh-scattered light. The light is focused on a 50 µm slit to limit the out-of-focus light and thus increase the spectral resolution, then collimated before hitting the diffraction grating. The grating spatially separates light with differing wavelengths and projects them onto the detector. We used both system configurations, the solid and liquid cuvette, with different sample paths (Figure 2).

We began this work by taking apart the system (except the laser, which we verified was working as expected) and assessed each component to ensure it was clean and placed correctly. Beginning with the camera placement, we worked step-by-step on the optical path using a fluorescent light bulb to align the lenses and slit. We then placed a neon light source in the light path and refined the position of each component to optimize the position and intensity of the resultant spectrum in the 2D image. We aligned the diffraction grating, optimizing the signal in our region of interest (ROI), which was 2048 pixels wide and 100 high. We limited the ROI height to avoid including noise from pixels that don't receive light. Finally, we turned on the laser and optimized the incident light path, ensuring maximum light (lux) reached the sample end of the optical path with a digital light meter (Urceri, MT-912).

We used the suggested spectrometer cover to reduce the noise caused by stray light and built an enclosure using corrugated black plastic, as in our previous work. We acquired all spectra using the Spectrum Analyzer suite (r123) and processed them with the code in the linked GitHub repository. To ensure we'd limited stray light sources, we acquired a “dark spectrum” with the laser light off (Figure 3, blue line). The dark spectrum had minimal signal compared to the intensity of a spectrum from the neon source (Figure 3, compare blue and orange lines). This neon bulb, following calibration, provided well-defined peaks as expected from an atomic light source. These sources have atoms in the gas phase, so they don’t exhibit vibrational or rotational states and, therefore, have narrow peaks.

After optimizing the light path, we measured the laser power at the sample surface during the alignment using a Thorlabs PM16-120 sensor. The final post-alignment measure was 2.9 + 0.08 mW. We then used 4 mL of HPLC-grade acetonitrile (VWR) in

a capped quartz cuvette (Starna Cells) in the solid configuration as a test standard. We tuned two parameters contributing to signal quality: exposure (1–10,000 ms) and number of averaged acquisitions (1–100). Increasing the exposure duration can increase the number of photons reaching the detector, improving the signal, but can also pick up cosmic rays or other noise events. Increasing the number of averaged acquisitions can mitigate cosmic rays, but increases read noise with each acquisition.

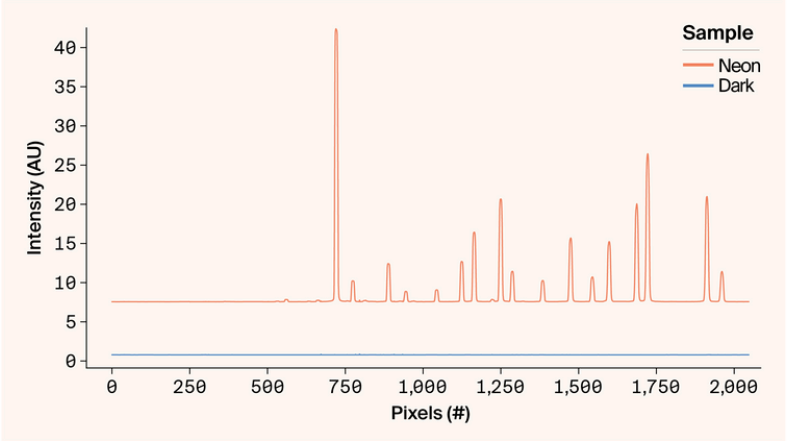

Figure 3

Comparison of the neon spectrum and the dark spectrum.

The neon spectrum is used as a calibrant, while the dark spectrum measures the system's background noise. We used 1,000 ms for neon exposure to avoid saturating the detector and 10,000 ms for the dark since this was the longest exposure we'd likely use for actual samples. Neither spectrum has been processed after acquisition, and both were acquired with five averaged acquisitions. Data are from 2024-10-18.

After verifying the presence of expected Raman peaks, we conducted a parameter sweep to identify the optimal acquisition time and number of averaged acquisitions (Figure 4). We could detect the most intense peaks of acetonitrile at very short acquisitions — 10–50 ms (Figure 4, A, right axis, 10–50). Minor peaks became evident at exposures of 100 ms and were resolved at exposures of 501 ms and above (Figure 4, A, right axis, 100–501). A spectrum from a single 1,000 ms exposure contained eight detectable peaks (Figure 4, B, right axis, 1), though slightly less noise was evident after averaging two similarly exposed spectra (Figure 4, B, right axis, compare 1 and 2) and increasing the number of averaged spectra increased the resolvability of minor peaks (Figure 4, B, right axis). Based on these results, we decided to use 1,000 ms and 10,000 ms as standard settings, and average between one and five acquisitions. In

most cases, we began with 1,000 ms exposure and increased to 10,000 ms if peaks weren't well resolved. The results showed us that neon and acetonitrile are useful as calibrants, with the parameters we tested, and could be the basis of our calibration protocol.

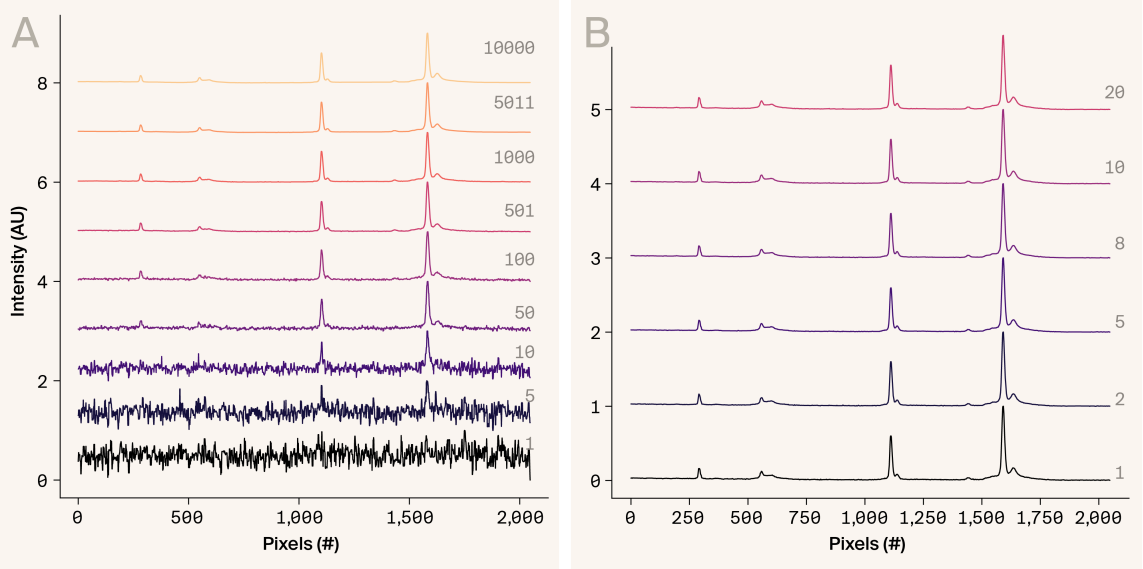

Figure 4

Comparison of acquisition parameters for samples of acetonitrile.

We analyzed acetonitrile at different exposures and the number of averaged acquisitions to determine suitable baseline acquisition parameters.

(A) Acetonitrile spectra collected with exposure times ranging from 1 ms to 10,000 ms; we averaged five spectra in all cases.

(B) Acetonitrile spectra collected by averaging one to 20 acquisitions; we used 1,000 ms exposure in all cases. Data are from 2024-10-11 and were baselined with airPLS and min-max scaled.

The resource

This resource has several components: a calibration protocol for the OpenRAMAN system (both configurations), a Jupyter Notebook for generating calibration equations, a Python script for applying this calibration to sample data, suggested acquisition

parameters for biological samples, and a small spectral library with raw and processed data as well as peak lists. Together, the components should allow any user to calibrate this DIY Raman system, acquire usable Raman spectra on biological samples, measure system performance, and compare results to our library.

> All associated **code**, the **spectral library**, and all **data** are available on GitHub (DOI: 10.5281/zenodo.14908269). If you run into issues, please comment on the protocol or this pub, and we'll be happy to discuss it.

Calibration overview

> **TRY IT**: Our full protocol for calibrating the OpenRAMAN system is available on protocols.io (DOI: 10.17504/protocols.io.yxmvmemj6g3p/v1).

We developed a standard calibration protocol to collect spectra from reference materials. We then used these spectra to generate equations to compare data from this instrument to other instruments. We used the equations to convert between the acquired units (pixels, #) and wavelength (nanometers, nm)/Raman shift (wavenumbers, cm^{-1}) based on the known peaks of the two reference materials.

Atomic emission sources are those where electrons from known atoms are excited and emit photons of specific energy when the electrons return to the ground state, resulting in spectra with sharp peaks at known fixed wavelengths that are robust to local environmental changes. For these reasons, they're typically used for calibration. We chose to use neon as an atomic emission source, consistent with the OpenRAMAN documentation, because neon bulbs are inexpensive, easy to acquire, and have well-known spectral peaks commonly used for calibration of 532 nm Raman instruments **[19]**.

We then turned on the laser and acquired a spectrum of acetonitrile as an additional reference material. We used acetonitrile as a standard for this test, as it's an organic liquid with multiple strong, narrow peaks across our range of interest. In contrast to neon, the acquired acetonitrile spectrum comprises Raman-scattered photons and can be used to verify the conversion.

We exported data from samples and dark and blank spectra in CSV format and imported them into the calibration notebook. We applied median filtering and baselining to both spectra to prevent peak finding and fitting issues. We selected 15 well-resolved peaks in the neon emission spectrum as reference points. These peaks have known wavelength positions and thus can be used to convert pixel numbers to nanometers. We used the SciPy signal software package (v1.13.1) to find the 15 corresponding peaks in the acquired neon spectrum and the lmfit package (v1.3.1) to fit Gaussians to each peak and calculate the center and width. We then plotted the measured peaks (in pixel #) against the known reference peaks (in nm) and fit a linear equation. We then used this equation to convert the neon spectrum from pixel to wavelength. We calculated the difference between the measured and reference peaks to calculate the error across the detector and average positional error. Both these metrics are related to the accuracy of acquired spectra with our system, based on the difference between the x-axis position of the peaks of a standard sample in our data versus literature references. As shown in Figure 5, this error forms a parabolic shape, related to how the diffraction grating disperses light across the detector, which is described by the equation:

$$dsin(\theta) = m\lambda$$

where θ = wavelength, λ = diffraction angle, and *d* = grating spacing.

We applied the conversion equation to the acetonitrile spectrum and then converted from wavelength (nm) to Raman shift (cm^{-1}) using the Raman shift equation for 532 nm excitation systems:

$$\text{Raman shift} = (10^7)\left(\frac{1}{532} - \frac{1}{\text{wavelength}}\right)$$

We then found and fit peaks in the spectrum, calculating the center and width. We plotted the measured peaks (in cm^{-1}) against the known reference peaks (cm^{-1}) and fit another linear equation. This is a minor adjustment to account for slight variations in laser behavior and environment that could affect Raman scattering. In this step, it's also easy to catch systematic errors in conversion or issues such as signal attenuation that could indicate a problem in the path from laser to sample to detector. We show the resulting calibrated spectra, peaks, and deviation from reference values across the detector in Figure 5.

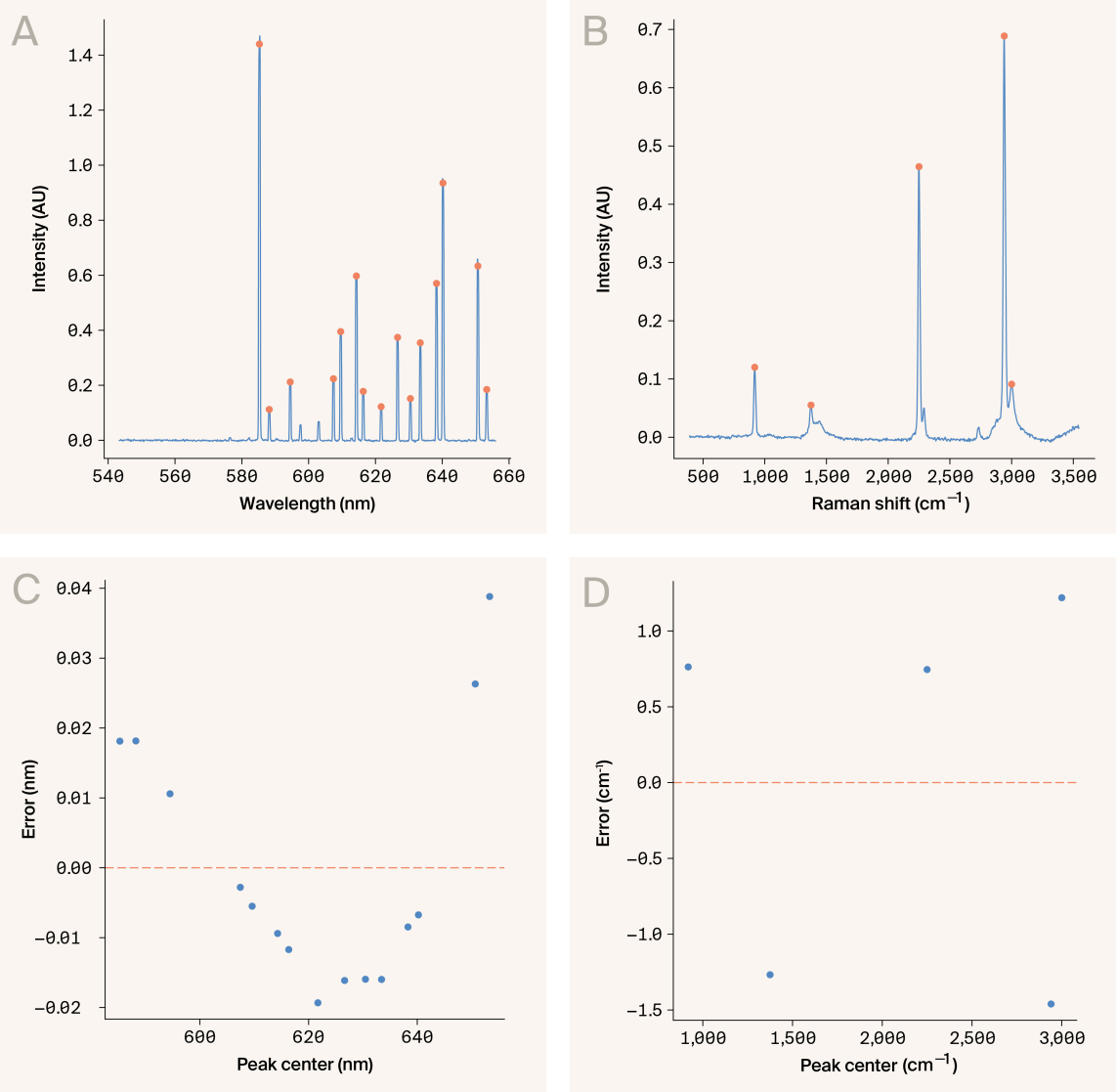

Figure 5

Calibrant data and error across the detector in solid configuration.

We used neon and acetonitrile data to generate the calibration equations.

(A) Neon spectrum in wavelength (nm) with fitted peaks collected at 1,000 ms exposure.

(B) Acetonitrile spectrum in Raman shift (cm^{-1}) with fitted peaks collected at 1,000 ms exposure.

(C) Difference between observed and reference peak values for neon.

(D) Difference between observed and reference peak values for acetonitrile. Data are from 2024-08-27.

System performance

We can measure the system's performance in several ways: how accurately the peaks of a sample are detected, the spectral resolving power of the instrument, and signal intensity. We used data from neon and acetonitrile in both configurations of the cuvette to generate performance metrics and characterize the system's behavior. The performance metrics included expected peak positional error, full width at peak half maximum (FWHM), and signal-to-noise ratio (SNR). Peak positional error is the deviation from expected, based on reference spectra, peak positions (± cm^{-1}). For Raman systems, FWHM is an indicator of the spectral resolution when measuring a reference material with narrow spectral peaks such as an atomic emission source **[18]**. For more complex samples, FWHM can change based on properties, such as crystallinity **[20]**, or environmental conditions, such as temperature **[21]**.

The stated performance of the OpenRAMAN (Starter Edition) is a resolution of 35 cm^{-1} on the 820 cm^{-1} peak of isopropanol, with a range of about 500 to 3,500 cm^{-1}. The specific range limits can change as the alignment of the detector to the grating is altered.

We calculated the metrics based on the 2,942 cm^{-1} peak of acetonitrile, which is expected to be very strong. We subtracted the dark spectrum and applied median filtering (kernel size = 5). We also applied baseline correction to the spectra, removing background signals from stray light, fluorescence, or other emissions and “flattening” the spectrum to more easily identify peaks. There are several approaches for baselining; here, we used the airPLS algorithm **[22]**. After finding the peaks, we calculated the SNR based on the following equation:

$$SNR = \frac{\text{peak intensity} - \text{background signal}}{\text{rms}(\text{background signal})}$$

We defined the background signal as the intensities between 1,900 and 2,000 cm^{-1}, part of the “quiet region” of a Raman spectrum **[23]**. This region typically doesn't have peaks from fundamental modes, especially for spectra from biological samples. Table

1 reports the system performance, measured on acetonitrile in the solid configuration on 2024-09-18 and liquid on 2024-08-09.

Metric	Explanation	Solid	Liquid
Error	Distance between measured value and reference value for peak position	1.354 cm^{-1}	1.755 cm^{-1}
FWHM	Full width at half maximum of a peak	19.394 cm^{-1}	21.225 cm^{-1}
SNR	Signal-to-noise ratio based on equation	78:1	48:1

Table 1

System performance.

We calculated key metrics for solid and liquid configurations using acetonitrile in a capped quartz cuvette or a borosilicate tube. For each calculation, we used the 2,942 cm^{-1} peak.

Spectral library

While the application of Raman spectroscopy to biological samples is increasing, there are still only a few accessible libraries. We collected and processed data for a spectral library focused on samples relevant to biological research (Table 2). For many samples, we acquired data in both solid and liquid configurations. All liquid configuration samples were in disposable borosilicate tubes (VWR, 47729-566) placed in the sample holder with no additional position adjustments for focus. We put powders, crystals, and solutions in the solid configuration on a mirrored grade 304 stainless steel substrate, which increases the Raman signal for biological samples **[24]**. We cleaned this substrate with 70% ethanol and dried the substrate between samples. We put solid biological cell cultures and media on matte black foil (single use) and targeted the colony surface using the visible beam to find the best focus position. All solutions listed below, other than those listed in the category “solvent,” are aqueous solutions.

Category	Sample	Source	Configuration	Notes
Mineral	Optical calcite	Ward's Science	Solid	Crystal
Salt	Magnesium sulfate heptahydrate, ≥ 99%	Sigma-Aldrich	Solid	Powder
			Liquid	1 M solution
	Calcium sulfate dihydrate	Ward's Science	Solid	Powder
	Sodium sulfate, ≥ 99%	Sigma-Aldrich	Solid	Powder
	Potassium phosphate monobasic, ≥ 99%	Sigma-Aldrich	Solid	Powder
	Sodium phosphate dibasic heptahydrate, 98–102%	Sigma-Aldrich	Solid	Powder
Solvent	Acetonitrile, ≥ 99.5%	VWR	Solid	In capped quartz cuvette
			Liquid	-
	Isopropanol, 200 proof	VWR	Liquid	-
	Ethanol, 200 proof	VWR	Liquid	-
Amino acid	Glycine, 99%	VWR	Solid	Powder
			Solid	0.001–1 M solution
			Liquid	0.001–1 M solution
	L-Methionine, ≥ 98%	Sigma-Aldrich	Solid	Powder
	L-Tyrosine, 99%	Beantown Chemical	Solid	Powder
Carboxylic acid	Citric acid, ≥ 99.5%	Sigma-Aldrich	Solid	Powder

Category	Sample	Source	Configuration	Notes
Fatty acid	Palmitic acid, 95%	AmBeed	Solid	Powder
Carbohydrate	D-(+)-glucose, 99.5%	Sigma-Aldrich	Solid	Powder
	Sucrose	Ward's Science	Solid	Powder
	Methylcellulose	Sigma-Aldrich	Solid	Powder
Biological	*Halobacterium* sp. NRC-1	Carolina Biological	Solid	Colony on agar
	Halobacterium sp. NRC-1	Carolina Biological	Liquid	Liquid culture
	Halobacterium agar	Carolina Biological	Solid	-
	Halobacterium medium	Carolina Biological	Liquid	-
	E. coli K-12	Carolina Biological	Solid	Colony on agar
	E. coli K-12	Carolina Biological	Liquid	Liquid culture
	LB agar	Sigma-Aldrich	Solid	-
	LB medium	Sigma-Aldrich	Liquid	-
	Chlamydomonas reinhardtii CC124	UTEX	Solid	Colony on agar
	Chlamydomonas reinhardtii CC124	UTEX	Liquid	Liquid culture
	TAP agar	UTEX	Solid	-
	TAP medium	UTEX	Liquid	-
Background	Matte black foil	Rosco	Solid	-
	Stainless steel	Yodaoke	Solid	Cleaned with 70% ethanol
	No sample	-	Liquid	

Category	Sample	Source	Configuration	Notes
	Borosilicate tube	VWR	Liquid	Cleaned with 70% ethanol

Table 2

List of samples in spectral library.

We applied standard acquisition and processing parameters for the spectral library presented in this pub. The parameters were median filtering (kernel size = 5), zero dB gain, five averaged acquisitions, and a 100-pixel ROI. We chose these based on the initial results from the acetonitrile parameter sweep (Figure 4) and other preliminary tests. We exported all data in CSV format and calibrated it using the neon and acetonitrile calibration data for that day and configuration, which was median-filtered (kernel size = 3) and baselined using the airPLS algorithm from the pybaselines module. We didn't usually apply background subtraction, which would remove the substrate (e.g., borosilicate tube or foil signal) but could increase noise. We note the exposure and any differences in acquisition or processing in the figure captions. We report peaks in a spreadsheet that's available with this pub in the "spectral_library" folder of our GitHub repo.

Background contributions

In addition to each sample measurement, we collected spectra of background materials to assess the spectral contributions of the substrates and the apparatus itself (Figure 6).

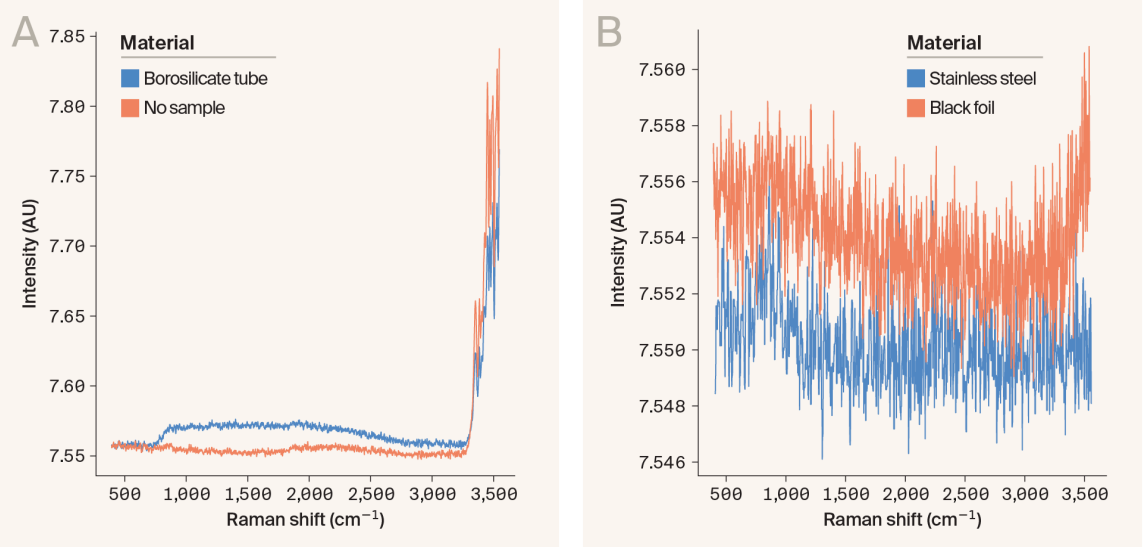

Figure 6

Comparison of background contributions for liquid and solid configurations.

We compared the spectra of the substrates for each configuration. All spectra are raw with no post-processing.

(A) For the liquid configuration, we compared the borosilicate tube, which holds samples, to the empty plastic tube holder. We used 1,000 ms exposure for these spectra that were collected on 2024-08-27.

(B) For the solid configuration, we compared two substrates used for different samples. We used 1,000 ms exposure and applied filtering but didn't baseline these spectra collected on 2024-08-27 and 2024-10-11.

These “dark” spectra typically showed no resolvable features and low background noise. The borosilicate tube spectrum (Figure 6, A) rose at ~800 cm^{-1}, while the liquid configuration with no sample or tube present had a broad feature at ~3,300 cm^{-1}. The broad feature was likely due to the plastic we used to make the tube holder. The stainless steel spectrum (Figure 6, B) rose around 800 cm^{-1}, and the black foil signal slightly rose at around 3,400 cm^{-1}. In some cases below, we used background subtraction to remove the contribution of these components from the spectra.

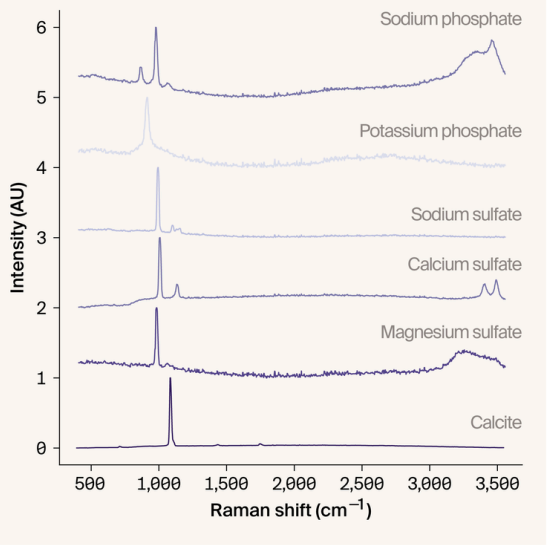

Figure 7

Spectra of minerals and salts.

We analyzed a set of common laboratory reagents between 2024-10-11 and 2024-10-18. We used 10,000 ms exposure, baselining using airPLS, and min-max scaling for all spectra.

Minerals and salts

We analyzed a set of minerals and salts: optical (crystalline) calcite, magnesium sulfate heptahydrate, calcium sulfate dihydrate, sodium sulfate, potassium phosphate monobasic, and sodium phosphate dibasic heptahydrate (Figure 7). We analyzed all of these samples with the instrument in the solid configuration, and all but the calcite (a crystal) were in powder form. We compared each of the spectra to peaks reported in reference literature and found suitable matches in nearly all cases, with most peaks within ± 5 cm^{-1}.

Solvents

We analyzed three common organic solvents in the liquid configuration: acetonitrile, isopropanol, and ethanol (Figure 8). We also analyzed acetonitrile in a quartz cuvette in the solid configuration. With regard to peak intensities and positions, the spectra from acetonitrile collected in both solid and liquid configurations were qualitatively very similar. However, peaks were slightly broader in the liquid configuration. There was more visible noise in all liquid sample spectra, and the expected broad background feature started at around ~3,300 cm^{-1}. The peaks in each spectrum matched published references well (± 5 cm^{-1}); we note the deviations in the linked spreadsheet.

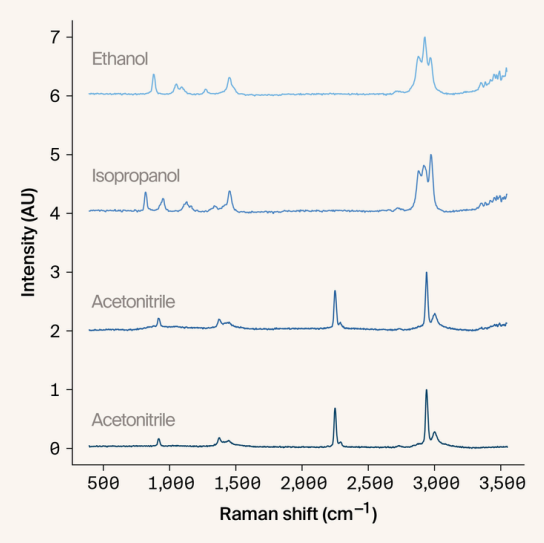

Figure 8

Comparison of organic solvents for liquid and solid configurations.

We analyzed a set of common organic solvents. From top to bottom, the spectra are ethanol (liquid configuration), isopropanol (liquid), acetonitrile (liquid), and acetonitrile (solid). We collected these spectra between 2024-08-27 and 2024-09-03. We used 1,000 ms exposure and applied min-max scaling for all spectra.

Glycine parameter sweep

Before analyzing many biomolecules, we did a parameter sweep with one sample — glycine — to determine parameters that may usefully serve as a baseline for spectrum acquisition from other molecules. Glycine is an organic molecule with peaks between 1,000 and 3,100 cm^{-1}. Using the solid configuration, we collected spectra sweeping through two parameters: the exposure time (100–10,000 ms, Figure 9, A) and number of averaged acquisitions (1–100, Figure 9, B. We found that the signal improves noticeably from one to five averaged acquisitions and only modestly with increasing

acquisitions. Across the sampled range, increasing exposure notably improves the signal with 10,000 ms, providing decreased noise. We established 10,000 ms and five averaged acquisitions as our typical parameters for solid biochemical powders, to balance the SNR and overall acquisition time needed.

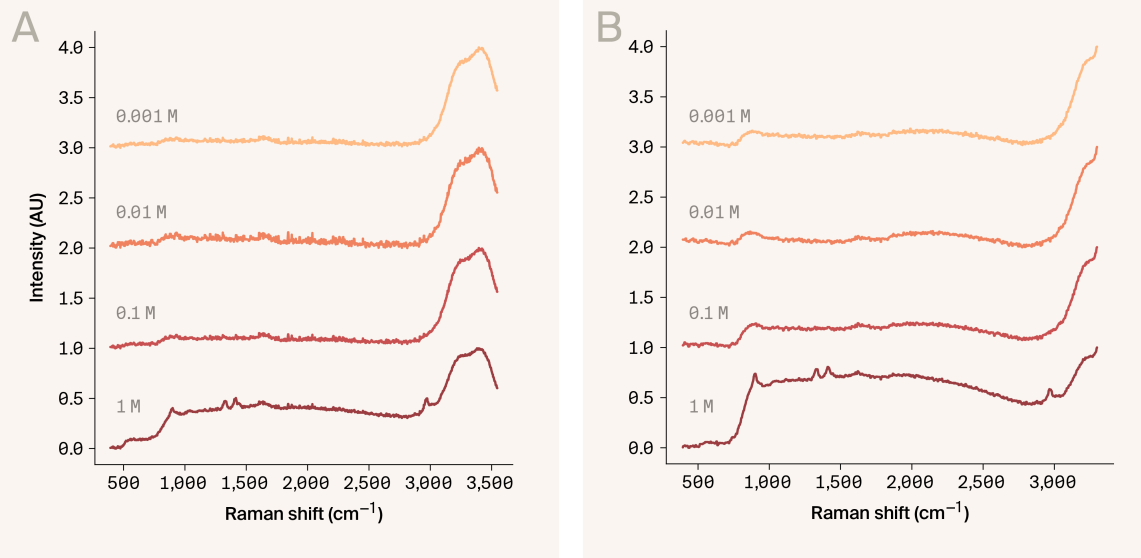

Figure 9

Glycine parameter sweep.

We analyzed glycine powder using different parameters. In all cases, we used the solid configuration and cropped the spectra from 1,000–3,500 cm^{-1} to show the major peaks better, baselined with airPLS, and min-max scaled. Data are from 2024-09-03.

(A) Glycine powder spectra collected with five averaged acquisitions and exposure ranging from 100 to 10,000 ms.

(B) Glycine powder spectra collected with 1,000 ms exposure and from 1 to 100 averaged acquisitions.

Biomolecules

We analyzed a panel of organic biomolecules in powder form with the spectrometer in the solid configuration (Figure 10). We chose three amino acids (glycine, L-methionine, and L-tyrosine), citric acid, palmitic acid, and three carbohydrates (D-glucose, sucrose, and methylcellulose). Consistent with expectation, the glycine, tyrosine, sucrose, and

methylcellulose spectra had strong background fluorescence (Figure 10, A). However, it was difficult to identify peaks below ~1,000 cm^{-1} in each case. We assessed several different baselining algorithms from the pybaselines module to remove the fluorescence and used a modified polynomial (Figure 10, B), though it still has artifacts due to fluorescence at < 1000 cm^{-1}. Regardless, we could resolve the major peaks of every compound, except for methylcellulose, due to its high fluorescence background. We compared each of the spectra to peaks reported in reference literature and found suitable matches in nearly all cases, with most peaks within ± 5 cm^{-1}.

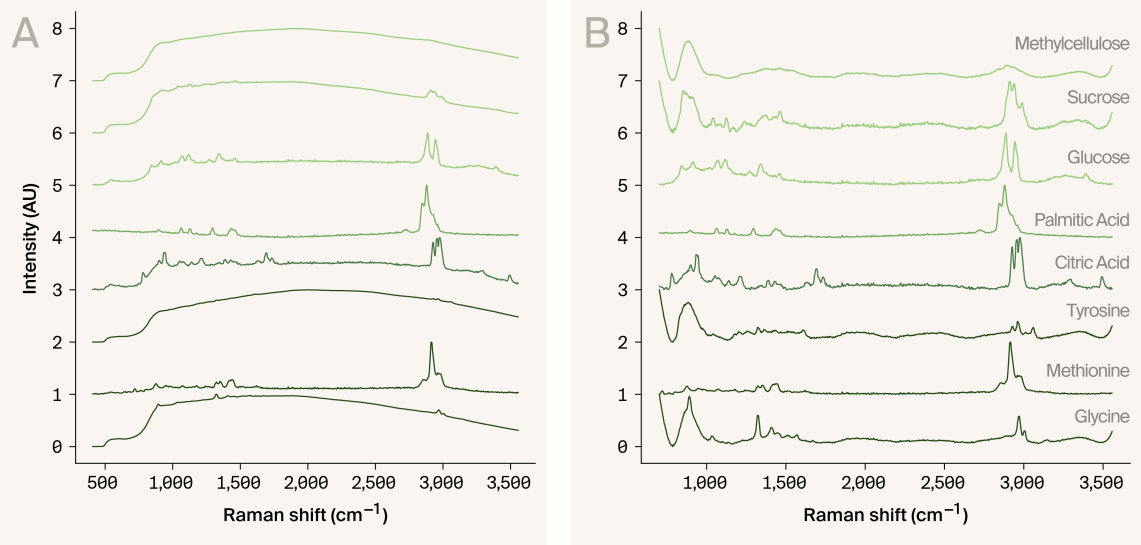

Figure 10

Spectra of biomolecules.

We analyzed biomolecules, including amino acids (glycine, L-methionine, L-tyrosine), citric acid, palmitic acid, and carbohydrates (D-glucose, sucrose, and methylcellulose). We acquired all spectra on 2024-10-11 with 10,000 ms exposure and min-max scaled.

(A) Unbaselined spectra.

(B) Spectra with polynomial fit baseline removed.

Glycine dilution series

To determine the detection limit of our system for a target biomolecule, we tested a dilution series of glycine powder in Millipore water ranging from 1 to 0.001 M (Figure 11). We used solid (Figure 11, A) and liquid (Figure 11, B) configurations for this test. We pipetted 200 µL of each solution onto cleaned stainless steel for the solid configuration and used 3 mL of solution in the borosilicate tube for the liquid. In both configurations, we could only distinguish glycine peaks from the 1 M solution, though we could see the water O-H stretching mode at all concentrations in the solid configuration. The background signal from the borosilicate vial and liquid sample holder obscured that region in the liquid configuration; therefore, we truncated it in the figure above.

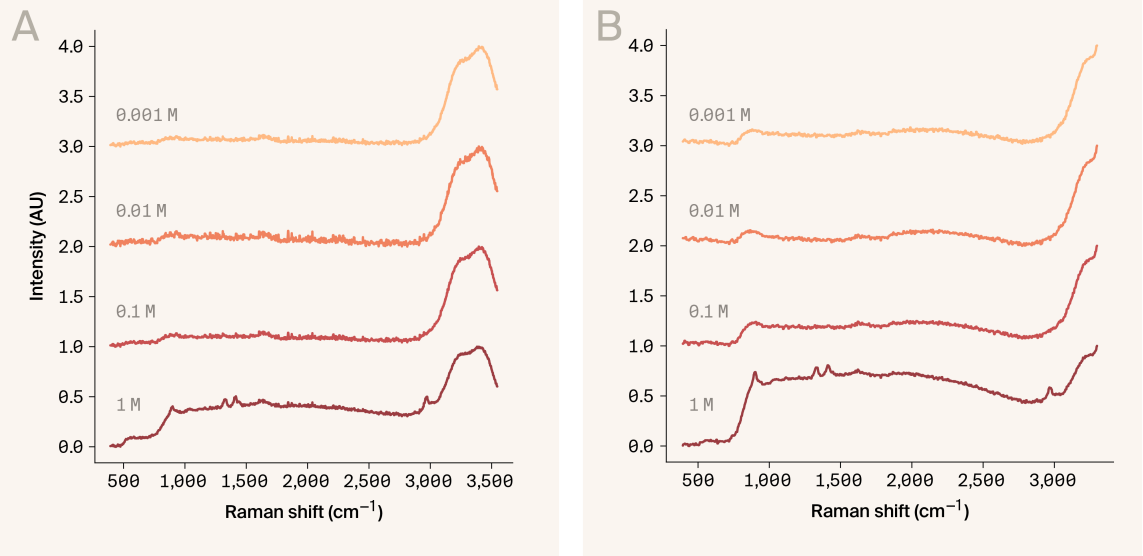

Figure 11

Glycine dilution series in solid and liquid configurations.

We analyzed glycine at different concentrations (0.001 to 1 M) in solid and liquid configurations. We acquired all spectra at 10,000 ms exposure, baselined using airPLS, and min-max scaled.

(A) Samples in solid configuration collected on 2024-10-18.

(B) Samples in liquid configuration collected on 2024-08-29. We truncated the liquid sample spectra at 3,300 cm^{-1} to remove the background feature.

Biological samples

Having established the effectiveness of this instrument in collecting spectra from biomolecules, we then evaluated its utility for collecting spectra from living biological samples. We first assessed different preparations for biological samples, focusing on lower-effort methods since one of our interests is rapid, scalable phenotyping. Using the two configurations of the system, we compared the spectra from solid and liquid samples of three different microorganisms: *Escherichia coli* K-12, *Halobacterium* sp. NRC-1, and *Chlamydomonas reinhardtii* 124 (Figure 12). *E. coli* is one of the most common model bacteria used in laboratory studies and has little or no pigmentation. The solid culture was grown for 24 h at 37 °C on LB agar, whereas the liquid culture was grown for 16 h at 37 °C in liquid LB medium shaking at 200 rpm. Before analysis, we pipetted the liquid culture up and down to more uniformly suspend the *E. coli* cells. *Halobacterium sp.* NRC-1 is a model extremophilic archaeon that produces multiple C40 and C50 carotenoids and survives low water and high salt conditions. We purchased the solid culture on *Halobacterium* agar from Ward's Science and stored it at room temperature before analysis. We grew the liquid culture for 24 h at 30 °C, 200 rpm, then allowed it to settle at room temperature for 48 h. The cells formed a denser film, which we then disrupted and suspended before analysis. *Chlamydomonas reinhardtii* 124 is a photosynthetic, single-celled alga that produces chlorophyll and carotenoid pigments and is motile. We grew the liquid culture in TAP medium in a rotating drum at room temperature under a 12 h light-dark cycle. We grew the solid culture on TAP agar at room

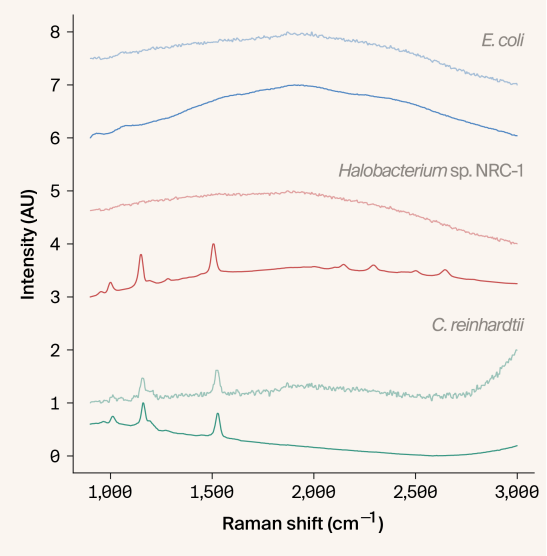

Figure 12

Biological samples in both configurations.

We analyzed three species in the solid and liquid configurations between 2024-08-27 and 2024-09-03. Each spectrum is n = 1; solid samples are darker colors, and liquid samples are lighter. We used 10,000 ms exposure, baselining with airPLS, and min-max-scaled all spectra.

temperature under continuous light and placed in the dark overnight before data acquisition.

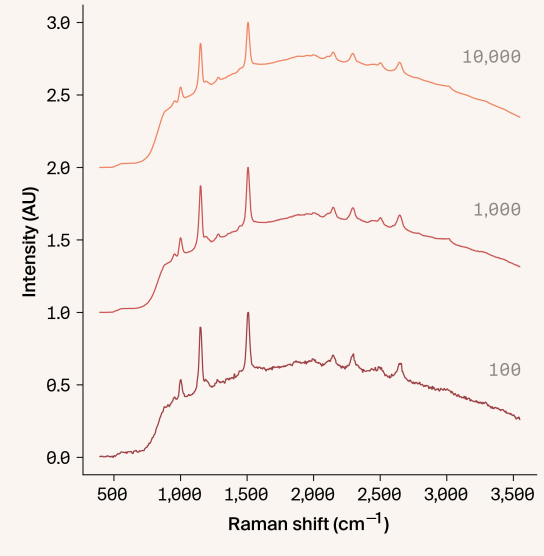

Figure 13

***Halobacterium* parameter sweep (solid configuration).**

We analyzed a solid culture of *Halobacterium* sp. NRC-1. Each spectrum is n = 1 and collected on 2024-09-03. We used baselining with airPLS and min-max scaled all spectra for all samples.

In both configurations and sample preps, we couldn't recover peaks from *E. coli*. However, we could recover peaks in both configurations for *C. reinhardtii*, though solid cultures had stronger signals. For *Halobacterium* sp., we could only recover peaks in the solid configuration. The liquid configuration likely had more background due to the sample holder, borosilicate, and media suspension, which made it harder to recover Raman peaks. Therefore, we used solid preparations for subsequent analyses.

We did a parameter sweep using a solid culture of *Halobacterium* sp. NRC-1 to understand how much signal we could recover at low exposures that would be more suitable for dynamic analysis (Figure 13). As with all solid biological samples, we placed a small piece of the colony with agar on black foil onto the sample stage. We tested three exposures: 100 ms, 1,000 ms, and 10,000 ms. Exposures of 10,000 ms provided only modest improvements in peak SNR over 1,000 s exposure, suggesting that shorter exposures could be used for assessing changes over time for this species and possibly those with similarly detectable pigments.

We then assessed variation between replicates of the same sample. We analyzed three biological replicates of each species in the solid configuration (Figure 14), placing samples on black foil and focusing the laser on the colony's surface. In all cases, we saw a fluorescence background from the sample, which is expected given the excitation wavelength we're using and the fact that these are biological samples **[2]**. As

before, with *E. coli*, this background was strong enough that we couldn't discern any Raman peaks. However, we could clearly distinguish several consistent peaks for the other two species across replicates.

We could see over ten peaks across 800 to 3,000 cm^{-1} for *Halobacterium*. These peaks were: 957 cm^{-1}, 1,002 cm^{-1}, 1,152 cm^{-1}, 1,196 cm^{-1}, 1,284 cm^{-1}, 1,446 cm^{-1}, 1,507 cm^{-1}, 2,107 cm^{-1}, 2,149 cm^{-1}, 2,296 cm^{-1}, 2,444 cm^{-1}, 2,501 cm^{-1}, and 2,647 cm^{-1}. The peaks below 2,000 cm^{-1} are likely due to the vibration of carotenoid pigments in *Halobacterium*, which usually yield strong signals under 532 nm excitation **[25]**. Those above 2,000 cm^{-1} may be combinations or overtones of the fundamental modes. There's also the possibility that some of the peaks — at 957, 1,284, and 1,444 — may be due to other biomolecules, such as phosphate groups from phospholipids or nucleic acids, amide groups in proteins, or CH_2 or CH_3 groups in lipids and proteins.

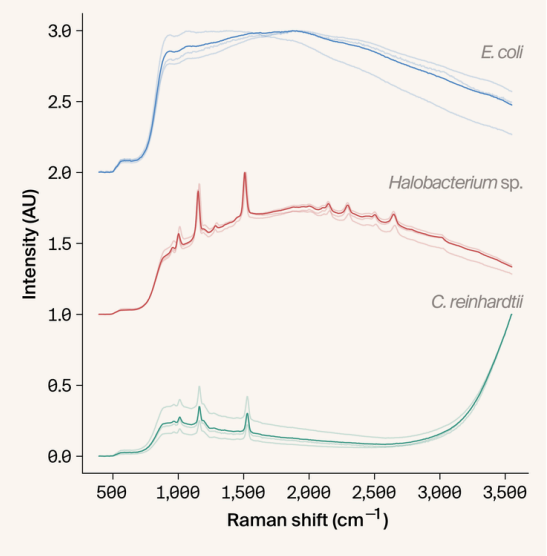

Figure 14

Three biological samples in solid configuration.

We collected spectra on multiple samples for three species on 2024-09-03. Each light line is n = 1, and the darker line is the average for the three. We used 10,000 ms exposure, baselining with airPLS, and min-max scaled all spectra.

For *C. reinhardtii*, we saw a fluorescence background that changed over time with increased light exposure. However, we could still distinguish several peaks at 966 cm^{-1}, 1,010 cm^{-1}, 1,160 cm^{-1}, 1,195 cm^{-1}, 1,275 cm^{-1}, and 1,527 cm^{-1}. These are similar to *Halobacterium*, suggesting that a carotenoid pigment is present and enhanced under this excitation wavelength. The 966 and 1,275 peaks could also be due to chlorophyll a. A combination of carotenoid and chlorophyll peaks is typically responsible for most of the peaks in this species **[26]**.

Time-series analysis of *Chlamydomonas reinhardtii* cc124

We noticed a visible change in the color of the laser spot on the surface of *C. reinhardtii* cultures over time and a change in the background fluorescence of spectra over time. We decided to investigate how the spectrum of this culture changes with continuous laser light exposure, capturing a 1-second exposure spectrum every minute for 20 minutes (Figure 15, A). During this time, we observed that the visible laser spot on the sample changed from red to orange, a change that's potentially consistent with the known phenomenon of chlorophyll fluorescence decay **[27]**. This occurs when dark-adapted photosynthetic organisms are exposed to light for an extended time, which leads to an increase in fluorescence emission intensity and subsequent decrease. Our previous work using the phenotype-o-mat observed this phenomenon over 20 minutes **[28]** with exposure to 460 nm light.

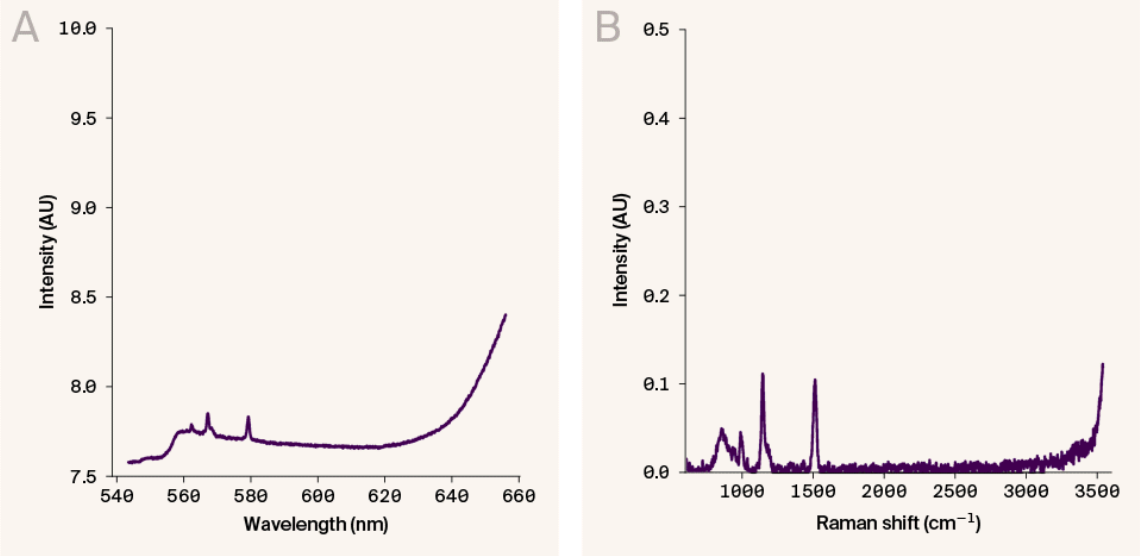

Figure 15

***C. reinhardtii* 124 time series**.

We analyzed solid cultures of *Chlamydomonas reinhardtii* 124 over 20 minutes, with spectra captured every minute under continuous laser light. We acquired all data at 1,000 ms exposure and a single spectrum acquisition on 2024-09-04 and didn't apply filtering.

(A) Unbaselined spectra.

(B) Baselined spectra using the function built into the OpenRAMAN Spectrum Analyzer software (version r123).

In the current work, we're using overnight dark-adapted cells exposed to continuous 532 nm light, a wavelength that *Chlamydomonas* cells don't absorb as well **[29]**. Research with *C. reinhardtii* grown under green light has shown enhanced energy transfer from light-harvesting chlorophyll protein complexes to photosystem I and II **[29]**. The green light is possibly absorbed by carotenoids, which are also present in this strain and have roles in light harvesting and preventing photooxidative damage **[30]**.

The fluorescence background in the collected spectra has two possible features, one with a peak at or below 560 nm and the other at > 660 nm (Figure 15, A). The overall intensity of the background in the spectra increased over time, with the fluorescence < 560 nm increasing more than that at 660+ nm. The peak > 660 nm may be the known ~680 nm peak observed in *C. reinhardtii* cells due to emissions from photosystem II **[31][32]**. Our detection range cuts off at 660 nm, so we can't define the true lambda max or peak behavior over time. Similarly, we can't fully define the lambda max of the shorter wavelength fluorescence, which could be the tail end of fluorescence emissions from pigment binding complexes observed in other green algae **[33]**, another chromophore that emits at this wavelength, or a photodegradation product that's being produced over time.

We then compared the Raman spectra, separated by baselining the original spectra, over time (Figure 15, B). We didn't see a notable change in the number of peaks or their positions, but the intensities decreased over time. This could be due to the increasing fluorescence background obscuring the Raman signal or possibly actual changes in the pigments responsible for the prominent Raman peaks. These findings indicate we can capture dynamic phenotypes with Raman and fluorescence analysis for this and similar photosynthetic organisms over time. Chlorophyll fluorescence decay in response to continuous light exposure is well studied. With the addition of Raman spectroscopy, we can capture changes to chemicals and pigments other than chlorophyll during this process.

Additional methods

We used ChatGPT to streamline and clarify the text we wrote and quickly test out different plot ideas by providing spectra and asking for various plot types. We used GitHub Copilot to help write and clean up code, with it suggesting code and comment ideas that we then selected from. GitHub Copilot also auto-suggested code for

repeating or modifying sections, especially for generating similar figures with different data. Additionally, we used Grammarly Business to suggest wording ideas, pick and choose bits to use, reformat text according to a style guide, and streamline and edit the text we wrote.

Key takeaways

The key takeaway from this effort is that DIY Raman, specifically this implementation of the OpenRAMAN (Starter Edition), can acquire high-dimensional compositional and time-varying data on biological samples, including biomolecules, salts, liquid and solid cultures of living cells. However, solutions analyzed in either configuration must be relatively concentrated (1 M) to distinguish multiple peaks. Biological samples give much more signal when in a solid state (i.e., colonies on a plate) than liquid cultures, likely because of lower background and higher density. The results correlate well with published references and appear to be reproducible. Our current hardware, protocol, and code implementation enables straightforward acquisition, calibration, and data processing. This low-cost system is helpful for biology and biochemistry laboratory research and has potential as an easy-to-build tool for rapid phenotyping.

Next steps

The OpenRAMAN system is flexible and can be modified to improve performance and utility for biological samples. We plan to change it to enable higher throughput acquisition. The most obvious next step would be to improve the sample end. For instance, we could include an objective, XYZ-automated stage, and a camera, allowing for better focusing on a sample and moving from point to point across acquisitions. In this way, we could map data on samples that are standard formats for biology, such as colonies on a Petri dish or wells of a multi-well plate. In addition, having automated metadata saving would help streamline the data collection process.

We're also interested in other upgrades to the system. Adding shutters to control the light path would be helpful for time-series acquisitions in which we don’t want the sample continuously exposed to light. Using a laser with more power or a different wavelength for this system would change how we interrogate the sample. A higher-powered laser would allow for potentially more signal, and we may be able to include

an objective to focus the beam and improve the spatial resolution further. A different wavelength, such as 785 nm, could decrease the background fluorescence expected in biological samples but may have trade-offs in the intensity of the Raman scattering **[34]**.

One aspect of this study we didn't fully explore was the behavior of *C. reinhardtii* cells over time, given that our detector range didn't fully capture the major fluorescence peaks. We're interested in further pursuing this research area and can modify the system to change our edge filters and alignment to capture a different range. We can also study cells exposed to different dark and light cycles, overall laser light exposure, and wavelengths of light. We think this will give us a better understanding of time-dependent phenotypes in this and related species through combined Raman and fluorescence spectroscopy.

We'll share updates to our Raman system and its associated protocols and code as we develop them. We'll also continue to build out our Raman spectral library, focusing on adding samples relevant to biological research. Please comment on the pub if you've questions, thoughts, or suggestions! We'd love to hear about your results and feedback if you use this system for biological research. In addition, we'd like to hear about what datasets and levels of data processing were helpful for you from this effort.

References

1. Avasthi P, Braverman B, Bigge BM, Essock-Burns T, Lane R, Mets DG, Patton AH, York R. (2024). Raman spectroscopy enables rapid and inexpensive exploration of biology. https://doi.org/10.57844/ARCADIA-085E-3ECF
2. Butler HJ, Ashton L, Bird B, Cinque G, Curtis K, Dorney J, Esmonde-White K, Fullwood NJ, Gardner B, Martin-Hirsch PL, Walsh MJ, McAinsh MR, Stone N, Martin FL. (2016). Using Raman spectroscopy to characterize biological materials. https://doi.org/10.1038/nprot.2016.036
3. Pezzotti G. (2021). Raman spectroscopy in cell biology and microbiology. https://doi.org/10.1002/jrs.6204

4 García-Timermans C, Props R, Zacchetti B, Sakarika M, Delvigne F, Boon N. (2020). Raman Spectroscopy-Based Measurements of Single-Cell Phenotypic Diversity in Microbial Populations. https://doi.org/10.1128/msphere.00806-20

5 Rocha RA, Fox JM, Genever PG, Hancock Y. (2021). Biomolecular phenotyping and heterogeneity assessment of mesenchymal stromal cells using label-free Raman spectroscopy. https://doi.org/10.1038/s41598-021-81991-1

6 Morel O, Gierlinger N. (2023). Chemical tissue heterogeneity of young Arabidopsis stems revealed by Raman imaging combined with multivariate data analysis. https://doi.org/10.1016/j.microc.2023.108900

7 Shigeto S, Takeshita N. (2022). Raman Microspectroscopy and Imaging of Filamentous Fungi. https://doi.org/10.1264/jsme2.me22006

8 Sasaki R, Toda S, Sakamoto T, Sakuradani E, Shigeto S. (2023). Simultaneous Imaging and Characterization of Polyunsaturated Fatty Acids, Carotenoids, and Microcrystalline Guanine in Single *Aurantiochytrium limacinum* Cells with Linear and Nonlinear Raman Microspectroscopy. https://doi.org/10.1021/acs.jpcb.3c00302

9 Wakisaka Y, Suzuki Y, Iwata O, Nakashima A, Ito T, Hirose M, Domon R, Sugawara M, Tsumura N, Watarai H, Shimobaba T, Suzuki K, Goda K, Ozeki Y. (2016). Probing the metabolic heterogeneity of live Euglena gracilis with stimulated Raman scattering microscopy. https://doi.org/10.1038/nmicrobiol.2016.124

10 Greig JC, Tipping WJ, Graham D, Faulds K, Gould GW. (2024). New insights into lipid and fatty acid metabolism from Raman spectroscopy. https://doi.org/10.1039/d4an00846d

11 Yue S, Cheng J-X. (2016). Deciphering single cell metabolism by coherent Raman scattering microscopy. https://doi.org/10.1016/j.cbpa.2016.05.016

12 Borek-Dorosz A, Pieczara A, Orleanska J, Brzozowski K, Tipping W, Graham D, Bik E, Kubrak A, Baranska M, Majzner K. (2024). Raman microscopy reveals how cell inflammation activates glucose and lipid metabolism. https://doi.org/10.1016/j.bbamcr.2023.119575

13 Sigle M, Rohlfing A-K, Kenny M, Scheuermann S, Sun N, Graeßner U, Haug V, Sudmann J, Seitz CM, Heinzmann D, Schenke-Layland K, Maguire PB, Walch A, Marzi J, Gawaz MP. (2023). Translating genomic tools to Raman spectroscopy analysis enables high-dimensional tissue characterization on molecular resolution. https://doi.org/10.1038/s41467-023-41417-0

14 Chernenko T, Matthäus C, Milane L, Quintero L, Amiji M, Diem M. (2009). Label-Free Raman Spectral Imaging of Intracellular Delivery and Degradation of

Polymeric Nanoparticle Systems. https://doi.org/10.1021/nn9010973

15 Verrier S, Notingher I, Polak JM, Hench LL. (2004). In situ monitoring of cell death using Raman microspectroscopy. https://doi.org/10.1002/bip.20063

16 Westley C, Fisk H, Xu Y, Hollywood KA, Carnell AJ, Micklefield J, Turner NJ, Goodacre R. (2017). Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy. https://doi.org/10.1002/chem.201701388

17 Hennelly B, Barton S. (2018). Signal to noise ratio of Raman spectra of biological samples. https://doi.org/10.1117/12.2307591

18 McCreery RL. (2000). Raman Spectroscopy for Chemical Analysis. https://doi.org/10.1002/0471721646

19 Carrabba MM. (2001). Wavenumber Standards for Raman Spectrometry. https://doi.org/10.1002/0470027320.s0705 (PDF: https://www.s-a-s.org/assets/docs/0470027320_Wavenumber_Standards_for_Raman_Spectrometry.p

20 Liu X, Luan H, Jinglin Y, Wang Shuo, Wang Shujun, Copeland L. (2020). A method for characterizing short-range molecular order in amorphous starch. https://doi.org/10.1016/j.carbpol.2020.116405

21 Liu H-N, Cong X, Lin M-L, Tan P-H. (2019). The intrinsic temperature-dependent Raman spectra of graphite in the temperature range from 4K to 1000K. https://doi.org/10.1016/j.carbon.2019.05.016

22 Zhang Z. (2016). airPLS. https://github.com/zmzhang/airPLS

23 Vardaki MZ, Gregoriou VG, Chochos CL. (2024). Biomedical applications, perspectives and tag design concepts in the cell – silent Raman window. https://doi.org/10.1039/d3cb00217a

24 Lewis AT, Gaifulina R, Isabelle M, Dorney J, Woods ML, Lloyd GR, Lau K, Rodriguez-Justo M, Kendall C, Stone N, Thomas GM. (2016). Mirrored stainless steel substrate provides improved signal for Raman spectroscopy of tissue and cells. https://doi.org/10.1002/jrs.4980

25 Jehlička J, Culka A, Mana L, Oren A. (2019). Comparison of Miniaturized Raman Spectrometers for Discrimination of Carotenoids of Halophilic Microorganisms. https://doi.org/10.3389/fmicb.2019.01155

26 Pandey S, Archana G, Bagchi D. (2022). Micro-Raman spectroscopy of the light-harvesting pigments in Chlamydomonas reinhardtii under salinity stress. https://doi.org/10.1016/j.saa.2022.121613

27 Kalaji HM, Schansker G, Brestic M, Bussotti F, Calatayud A, Ferroni L, Goltsev V, Guidi L, Jajoo A, Li P, Losciale P, Mishra VK, Misra AN, Nebauer SG, Pancaldi S, Penella C, Pollastrini M, Suresh K, Tambussi E, Yanniccari M, Zivcak M, Cetner MD, Samborska IA, Stirbet A, Olsovska K, Kunderlikova K, Shelonzek H, Rusinowski S, Bąba W. (2016). Frequently asked questions about chlorophyll fluorescence, the sequel. https://doi.org/10.1007/s11120-016-0318-y

28 Braverman B, Mets DG, York R. (2024). The phenotype-o-mat: A flexible tool for collecting visual phenotypes. https://doi.org/10.57844/ARCADIA-112F-5023

29 Ueno Y, Aikawa S, Kondo A, Akimoto S. (2018). Adaptation of light-harvesting functions of unicellular green algae to different light qualities. https://doi.org/10.1007/s11120-018-0523-y

30 Bohne F, Linden H. (2002). Regulation of carotenoid biosynthesis genes in response to light in Chlamydomonas reinhardtii. https://doi.org/10.1016/s0167-4781(02)00500-6

31 Dinc E, Tian L, Roy LM, Roth R, Goodenough U, Croce R. (2016). LHCSR1 induces a fast and reversible pH-dependent fluorescence quenching in LHCII in *Chlamydomonas reinhardtii* cells. https://doi.org/10.1073/pnas.1605380113

32 Bonente G, Pippa S, Castellano S, Bassi R, Ballottari M. (2012). Acclimation of Chlamydomonas reinhardtii to Different Growth Irradiances. https://doi.org/10.1074/jbc.m111.304279

33 Mascia F, Girolomoni L, Alcocer MJP, Bargigia I, Perozeni F, Cazzaniga S, Cerullo G, D'Andrea C, Ballottari M. (2017). Functional analysis of photosynthetic pigment binding complexes in the green alga Haematococcus pluvialis reveals distribution of astaxanthin in Photosystems. https://doi.org/10.1038/s41598-017-16641-6

34 Hawes RC, George KP, Nelson DCarl, Beckwith Richard. (1966). Laser excitation of Raman spectra. https://doi.org/10.1021/ac50155a041