Product Biomass Samples IR Batch Spectroscopy

Tags

REPLACER

Goals

Twelve samples of biomass were provided which were harvested from a test reactor and subsequently freeze-dried. The goal was to identify chemical similarities or disparities between the samples which would hint at the reactor output having a chemical drift over time. A secondary goal was to develop the ability to establish spatio-spectral correlations. This would allow us to distinguish chemical classes in these biomass samples at high spatial resolution.

Sample Production

The provided biomass consisted of freeze-dried ***A. platensis***, which takes the form a green powder with grain size estimated between 0.1 and 1 mm.

From each sample of dried biomass, I first produced a **re-hydrated sample**. To that end, I placed one grain of dried biomass into a test tube on a scale. Recording the weight of the freeze-dried biomass, I added 10x that mass of MilliQ water, by pipette, directly onto the biomass grain. The rehydrated samples were *not* stirred in order to not destroy the structure of the biomass.

To prepare consistent samples for PTIR microscopy / spectroscopy, I started with an optically polished CaF_2 window. This was cleaned by

- rinsing it with a sequence of solvents (MilliQ water → Ethanol → Isopropanol)
- mechanically cleaning it with a microfibre cloth soaked in Isopropanol
- rinsing it with the same solvents as above, but in reverse order
- blowing off excess water with pressurized nitrogen gas

A stainless steel spacer ring $(R_{\text{inner}} = 5\text{mm},\ d = 10\mu\text{m})$ was placed atop the CaF_2 window. A volume of $2\mu\text{l}$ of the rehydrated sample was pipetted onto the middle of the CaF_2 window, in the centre of the spacer ring.

A second CaF_2 window, cleaned in the same manner as described above, was placed on top to close the sandwich. Gentle pressure was applied to compress the sample to the thickness of the spacer ring.

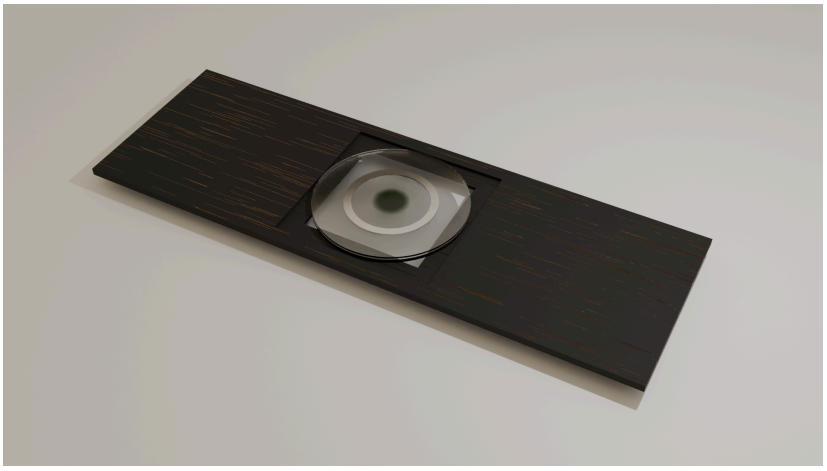


Illustration: Sample fixed to the custom sample holder, ready to be placed into the microscope.

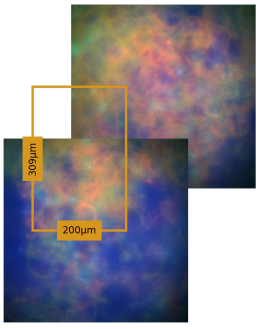
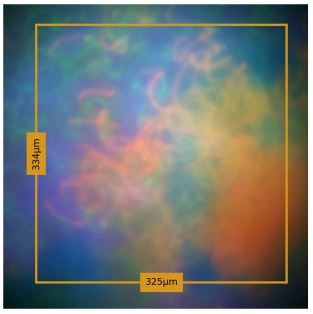
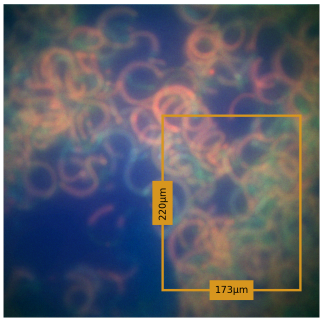
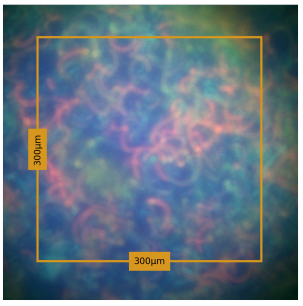
Sample	Labelled	Focus [µm]	Detector Gain
1	**0,1476**	2660.7	10x
2	illegible (cap blue, paper sticker: **0.01.25**)	2484.5	50x

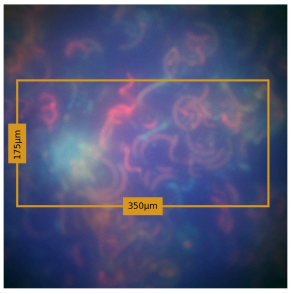
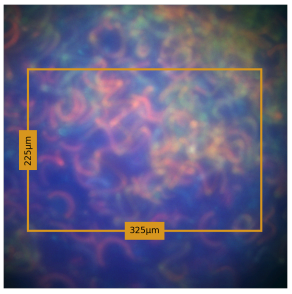
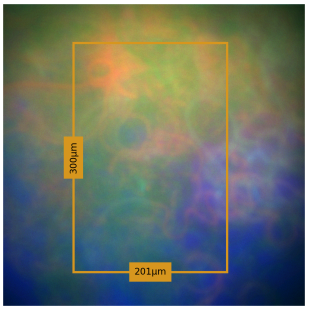
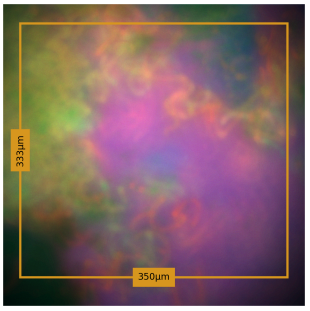
Sample	Labelled	Focus [µm]	Detector Gain
3	**30.01.25** A. platensis 0,15 g	2649.9	50x
4	**0,1653 g**	2592.0	50x
5	**98,6 mg**	2657.2	50x
6	03.02.25 A. platensis **0,19 g**	2569.8	50x
7	**A. platensis**	2641.9	50x
8	A. platensis 0,20 g **26.01.25**	2578.8	50x
9	~~A. platensis **0,23 g**~~		
10			
11			
12			

After production of the ninth sample, a crash of the setup was noticed. The crash was unfortunately determined to be irrecoverable. No further measurements could be made.

Measuring

Before spectral measurements, autofluorescence micrographs were captured to select a suitable region for spectroscopy:

Sample 1: **"0,1476"**	**Sample 2:** **"0.01.25"**	**Sample 3:** **"30.01.25"**	**Sample 4:** **"0,1653 g"**
			

Sample 5: **"98,6 mg"**	**Sample 6:** **"0,19 g"**	**Sample 7:** **"A. platensis"**	**Sample 8:** **"26.01.25"**
			

The Auto-Background procedure was executed once, before the first measurement. The PTIR setup was used in **co-propagation** mode, i.e. the IR pump and visible probe laser beams were both directed into the sample through the top **40x Cassegrain** objective. For each sampled point, the **average of 3 spectra** recorded at the same position was recorded.

Recorded Channels were

- `[0]` **OPTIR:** The amplitude of the PT signal, corrected for the background determined by the mIRage.

- `[1]` **Phase:** Phase of the PT signal
- `[2]` **X:** The real part (or in-phase component) of the PT signal
- `[3]` **Y:** The imaginary part (or quarter-phase delayed component) of the PT signal

The probe laser was modulated at a frequency of 100 kHz with a duty cycle of 1%.

Results

Spectra

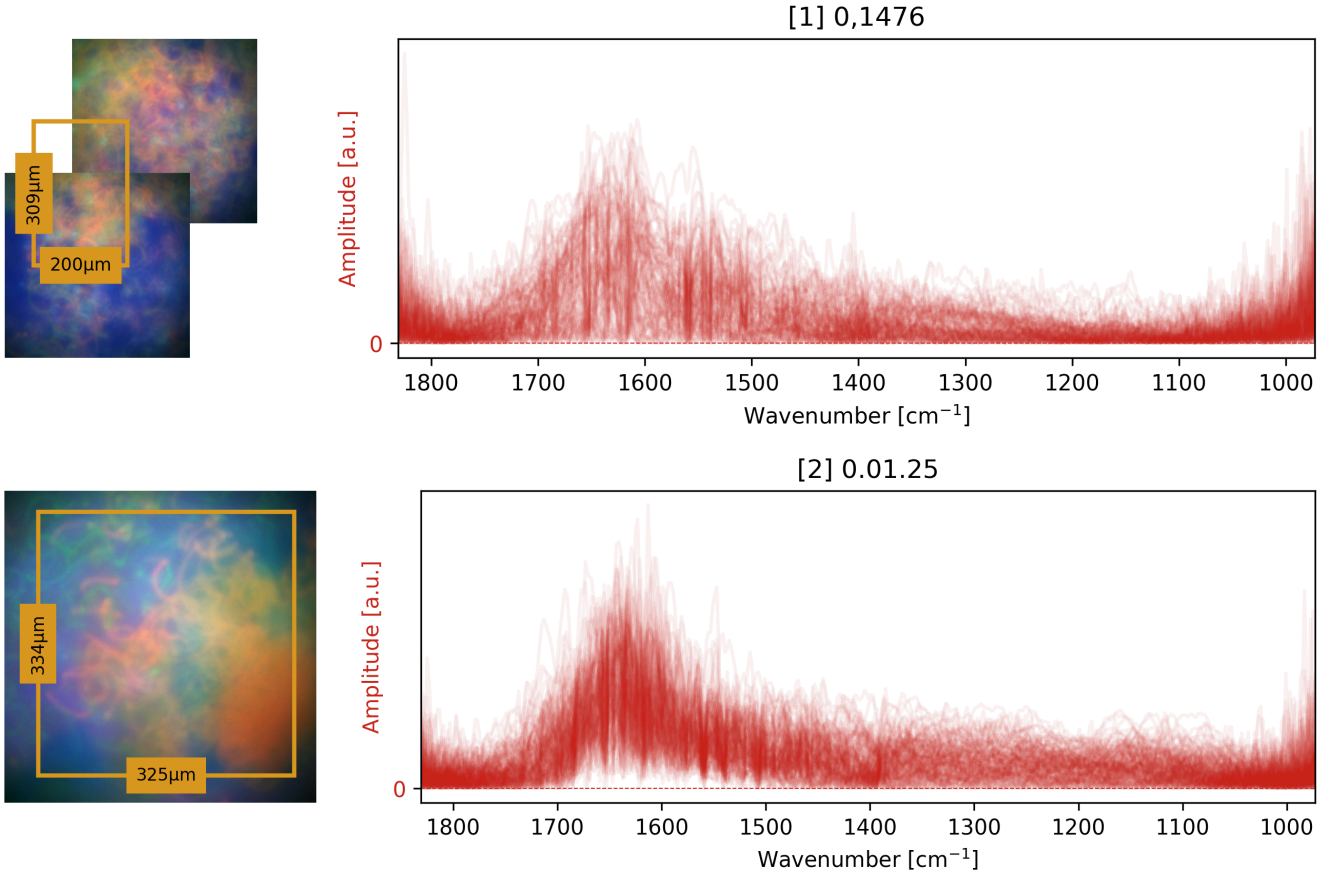
Definitions

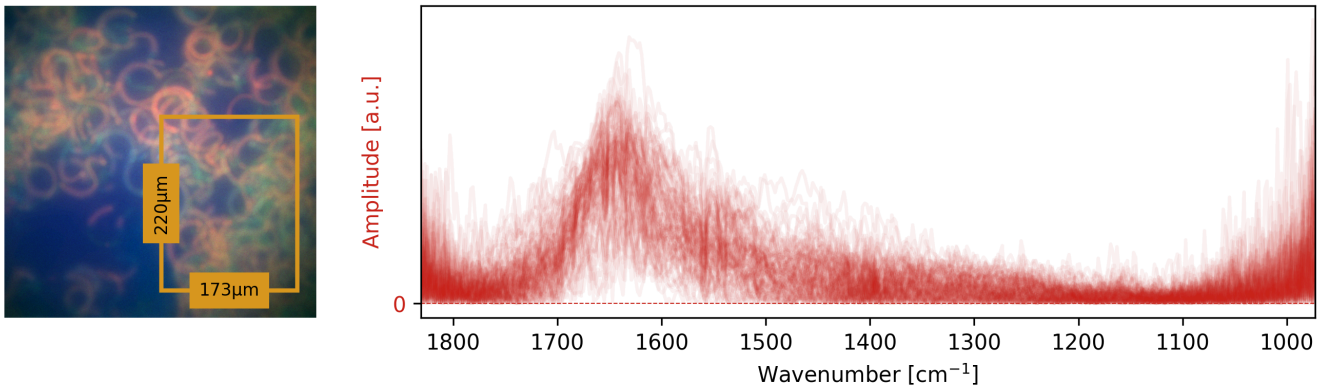
We assume that the "true" photothermal signal is encoded by $X + \mathrm{i}Y$. We thus calculate the **amplitude** of the signal as $\sqrt{X^2 + Y^2}$ and the **phase** as $\mathrm{atan2}(Y, X)$, instead of taking the phase as recorded by the software. We do this, because the phase of the average signal is not the same as the average phase of the signal:

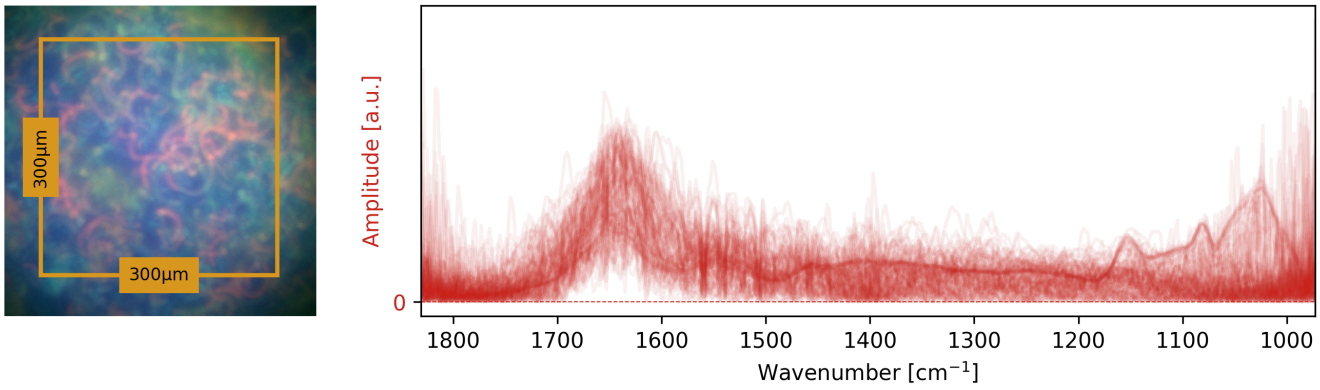
$$\mathrm{atan2}\big(\langle Y_i \rangle, \langle X_i \rangle\big) \neq \big\langle \mathrm{atan2}(Y_i, X_i) \big\rangle$$

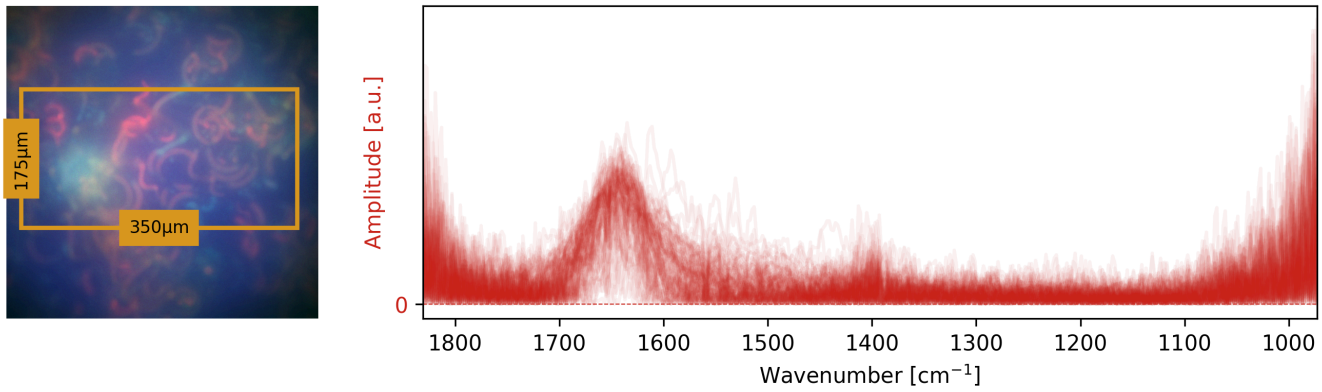
Individual Spectra

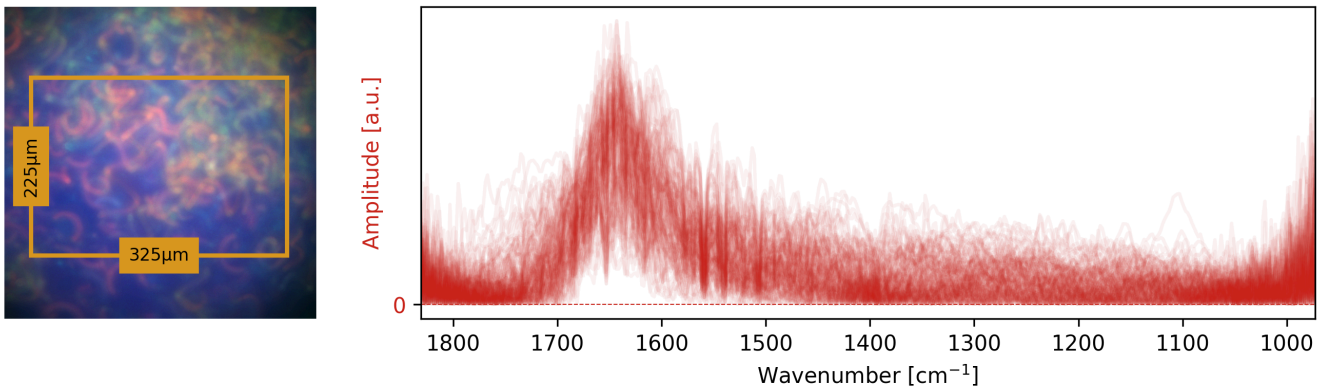
…were very noisy. Shown below are the L^2-normalized amplitude spectra from each sample. For each sample, an autofluorescence composite micrograph is given. The indicated areas are those, wherein spectra were recorded.

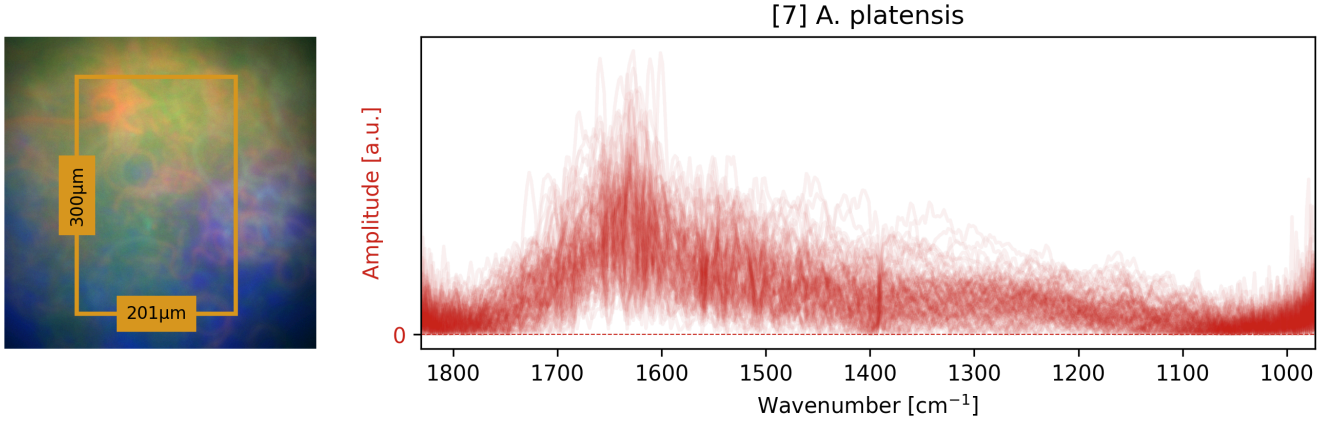


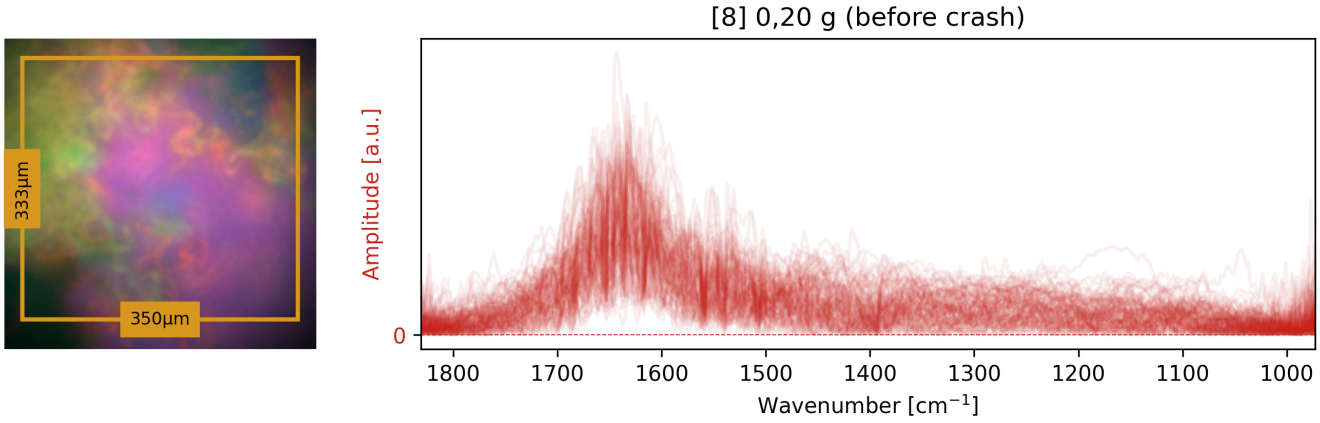

[3] 30.01.25
220µm
173µm
Amplitude [a.u.]
0
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]


[4] 0,1653g
300µm
300µm
Amplitude [a.u.]
0
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]


[5] 98,6mg
175µm
350µm
Amplitude [a.u.]
0
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]

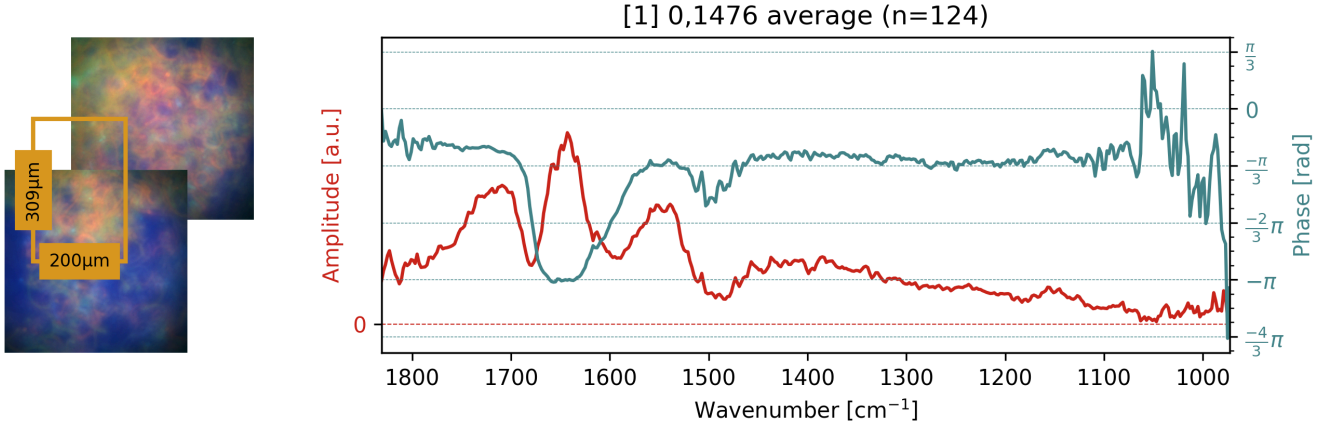

[6] 0,19 g
225µm
325µm
Amplitude [a.u.]
0
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]



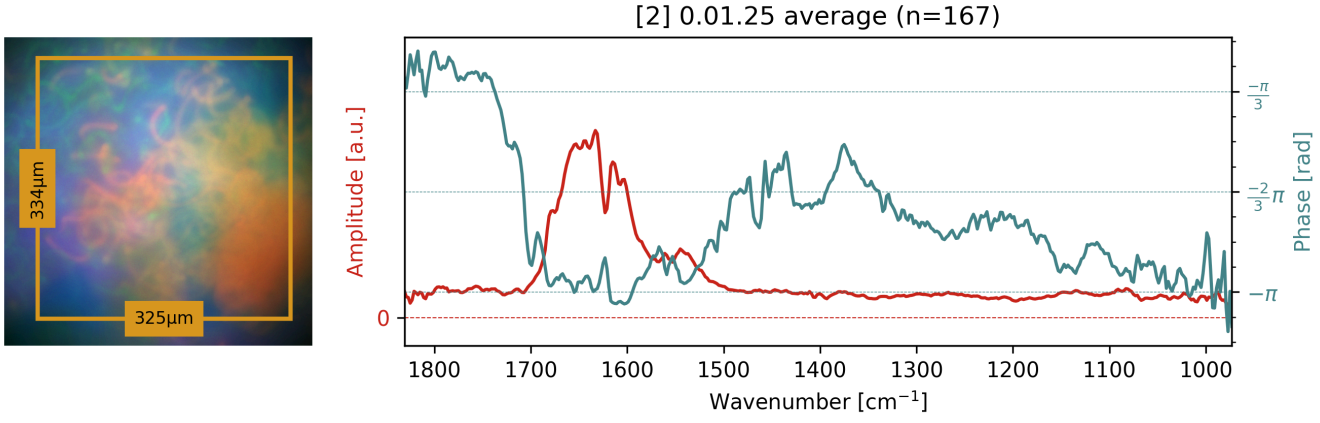


Average Spectra

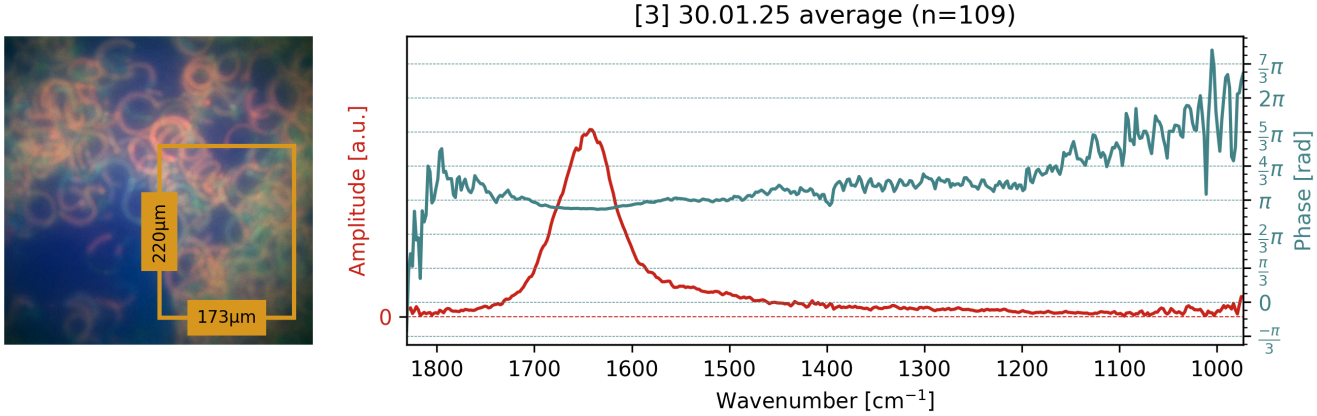
Here, average spectra are shown with amplitude and phase.



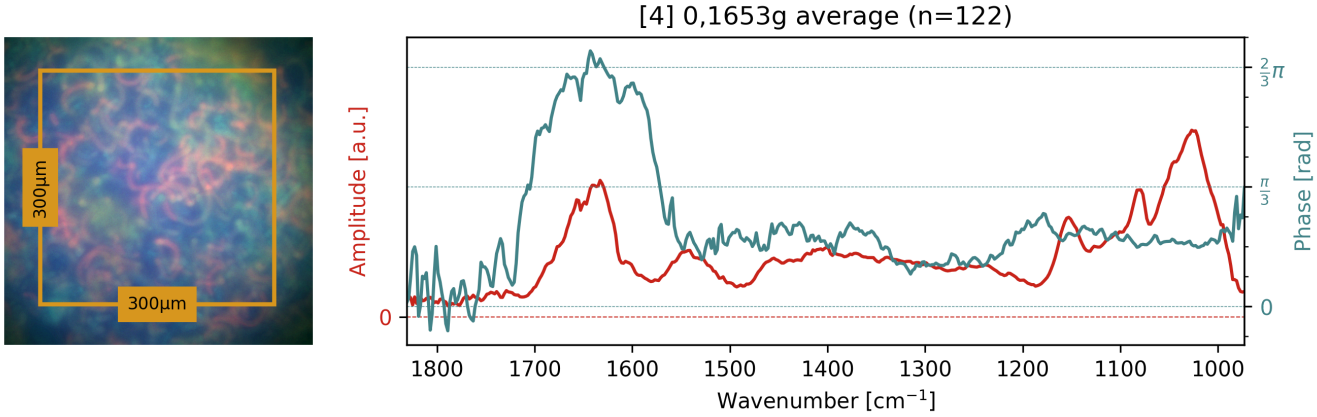
[1] Note how the main peak has a different phase from the secondary peaks. This might hint at the main peak corresponding to heating of a large volume (water background), taking a longer time for the temperature profile to relax and the secondary peaks corresponding to actual biomolecules, making up much less volume and therefore relaxing more quickly.



[2] Here, we similarly see a lagging phase in the region around 1650cm^{-1}.

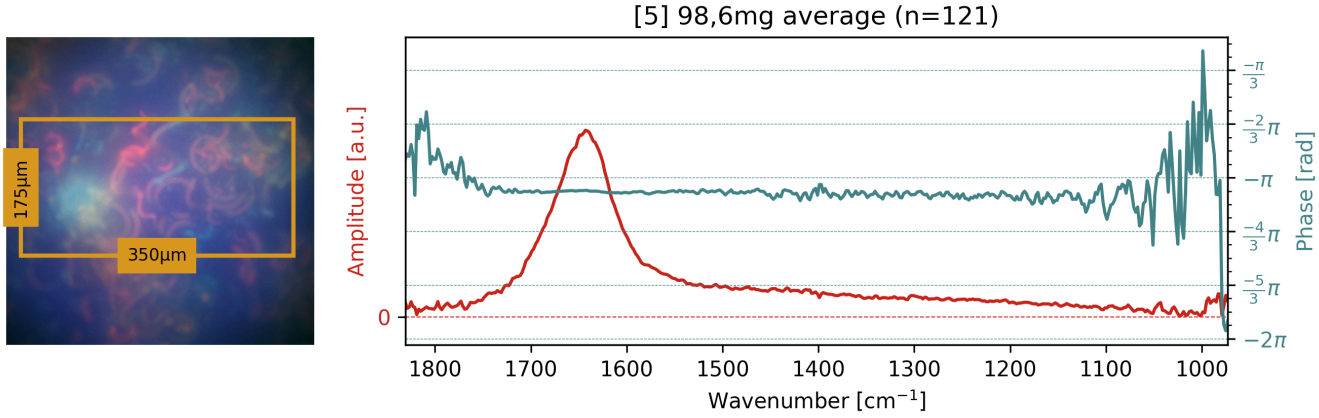


[3] Here, the phase seems to be invariant. The amplitude looks very smooth with only one visible peak and no secondaries. It doesn't look like any biomass is detected here, for some reason.

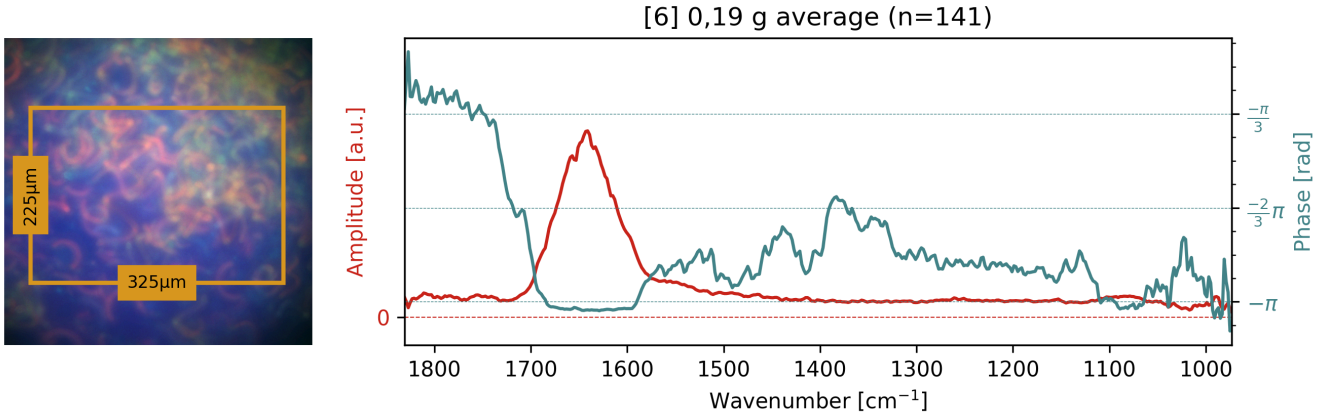


[4] Here, the phase doesn't lag around the water absorption but instead seems to advance. Reason unknown. Many additional peaks are visible in the amplitude of the signal. Their locations are consistent with usual bio-signatures, e.g.

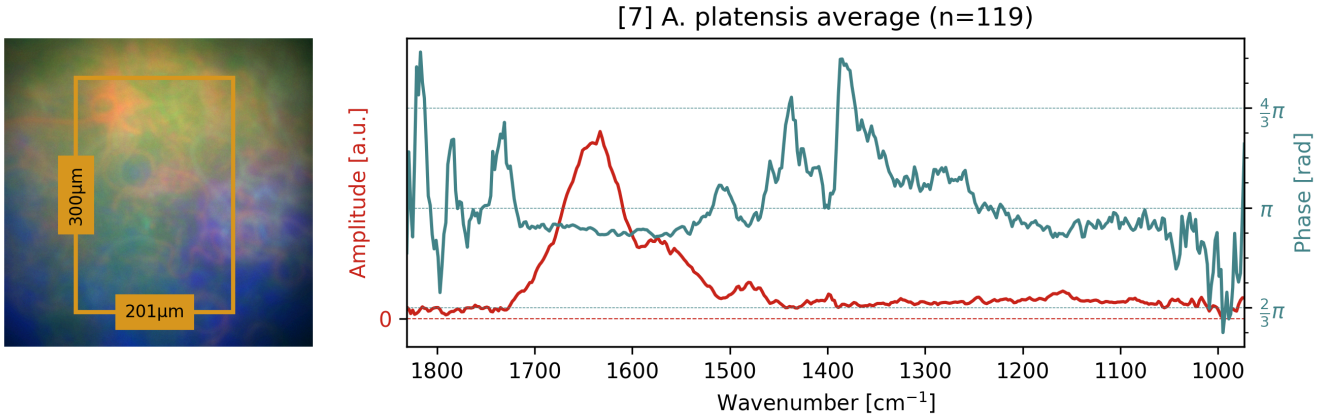
~1550 cm^{-1}	**~1460 cm^{-1}**	**~1150 cm^{-1}**	**~1080 cm^{-1}**	**~1020 cm^{-1}**
Amide II band	CH_2 scissor	C–O–C stretch	PO_4 stretch	carbohydrate C–O stretch



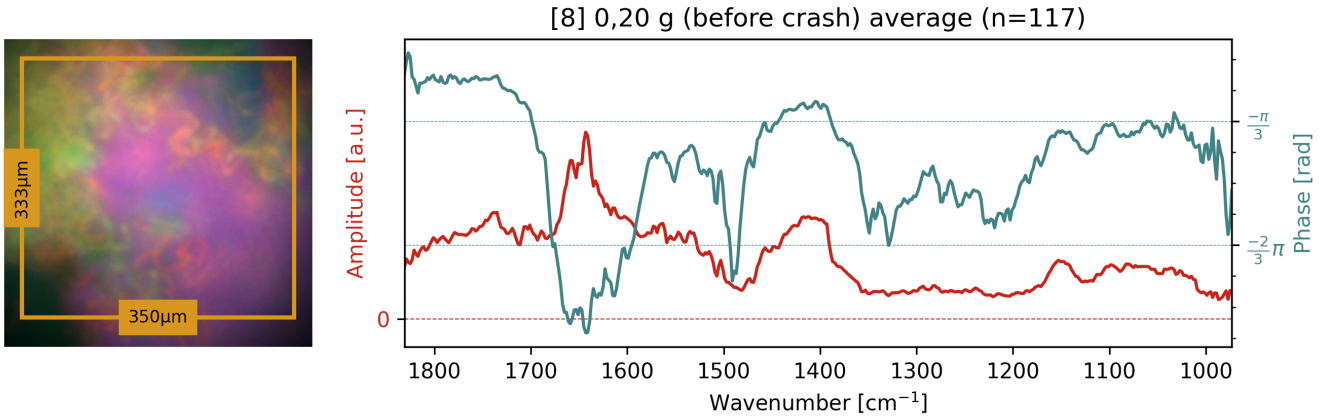
[5] as in [3], the phase is remarkably constant and no secondary peaks are visible.



[6] Secondary peaks are almost invisible. However, the phase clearly lags around the water absorption range.



[7] A peak in the high 1500s appears more like a shoulder to the side of the main peak at 1650 cm^{-1}. Smaller peaks are discernible in the amplitude at lower wavenumbers, and coincide with clear local phase lags.



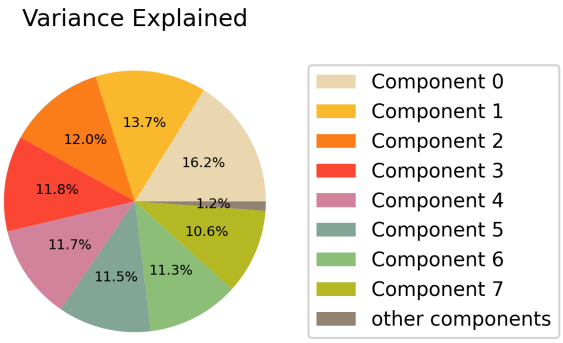
[8] Multiple peaks with irregular shapes are discernible in the amplitude. Many features (moth peaks and valleys) of the amplitude spectrum coincide with features in the phase spectrum.

Decomposition of Spectra

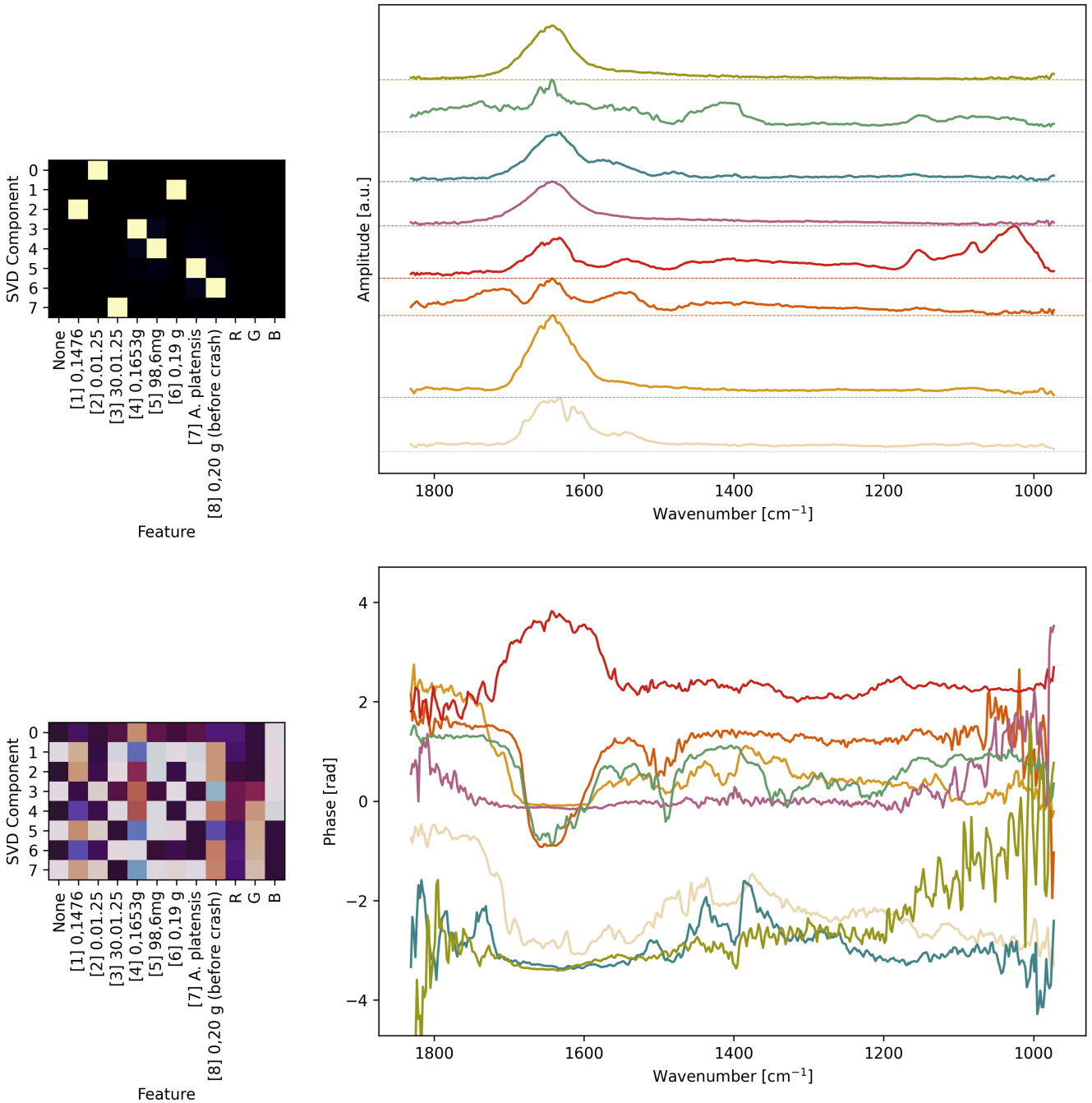
As usual, we cannot extract comprehensible information directly from the complete hyperspectral datasets and therefore must resort to some manner of dimension reduction. In this, it appears prudent to analyse the complex-valued spectra and to not disregard the phase.

Singular Value Decomposition

I performed PCA (or SVD, because `sklearn.decompose.PCA` doesn't support complex-valued input) The explained variance per SVD component looked like so:



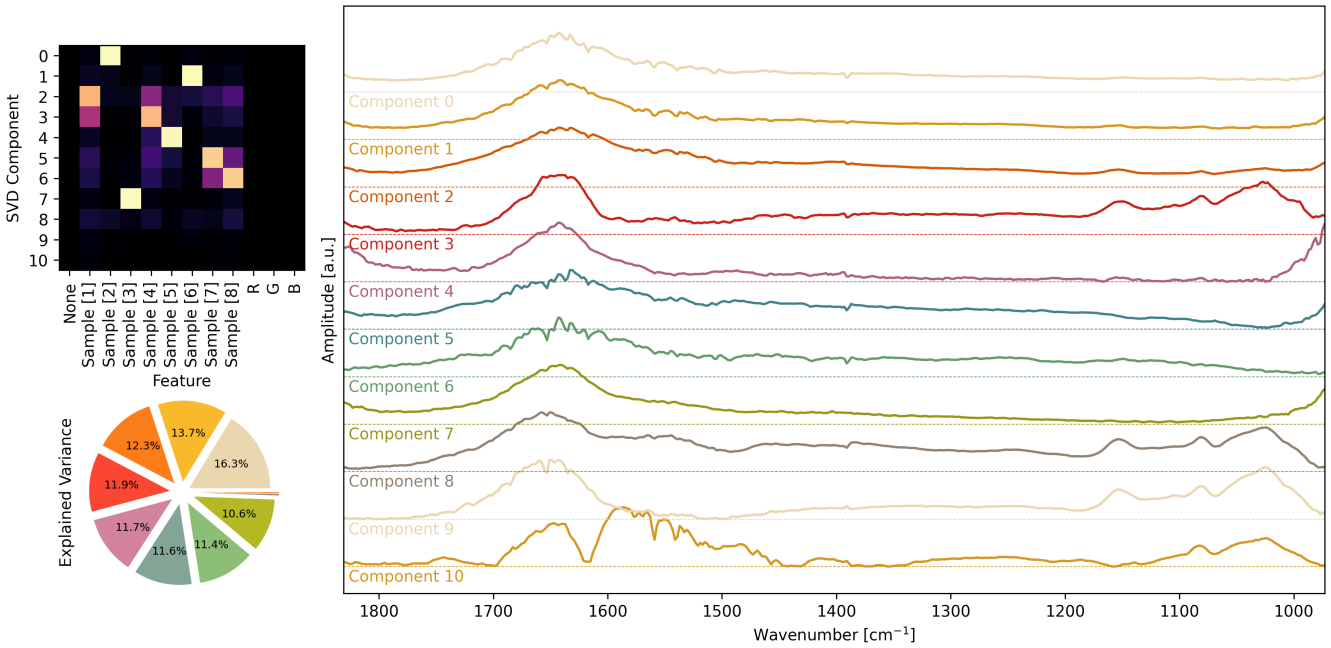
This means that the first eight components were significant and the rest could be disregarded, essentially. Inspecting the components gave the following:



- **top left:** magnitude of the mapping between SVD components and features (sample indices)
- **top right:** magnitude of the SVD components vs wavenumbers
- **bottom left:** phase of the mapping between SVD components and features (sample indices)
- **bottom right:** phase of the SVD components vs wavenumbers

The interpretation is as straight-forward as it is disappointing:
The eight relevant SVD components were nothing but the average spectra of the eight samples.

So, counter to the earlier idea, I attempted the decomposition disregarding the phase. When running the **SVD only on the amplitude** of the signal, we unfortunately get similar results:



- **top left:** magnitude of the mapping between SVD components and features (sample indices)
- **bottom left:** explained variance per SVD component. only 0.6% of the total variance of the dataset are explained by components beyond 7.
- **right:** SVD components vs wavenumbers

However, there is some overlap between measurements now. Still, it ought to be very clear that we should attempt to model the individual peaks and try decomposing into those.

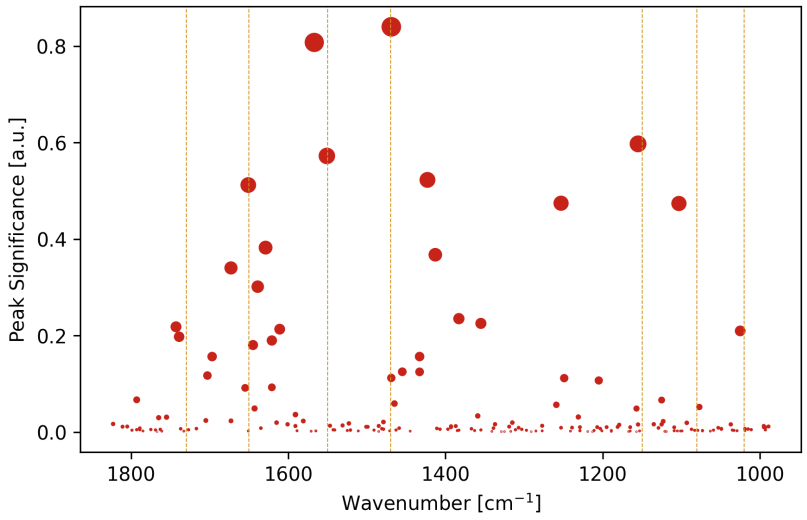
Attempts to Isolate Spectral Peaks

Let's accumulate information about all spectral peaks using `scipy.signal.find_peaks`. Obviously important is the position (in terms of wavenumbers) of each peak, but also, we require a measure for total absorption that a peak is associated with. Given the values of "prominence" and "width" that the `find_peaks` algorithm computes, we define a peak's

$$\text{significance} := \text{prominence} \cdot \text{width} \, .$$

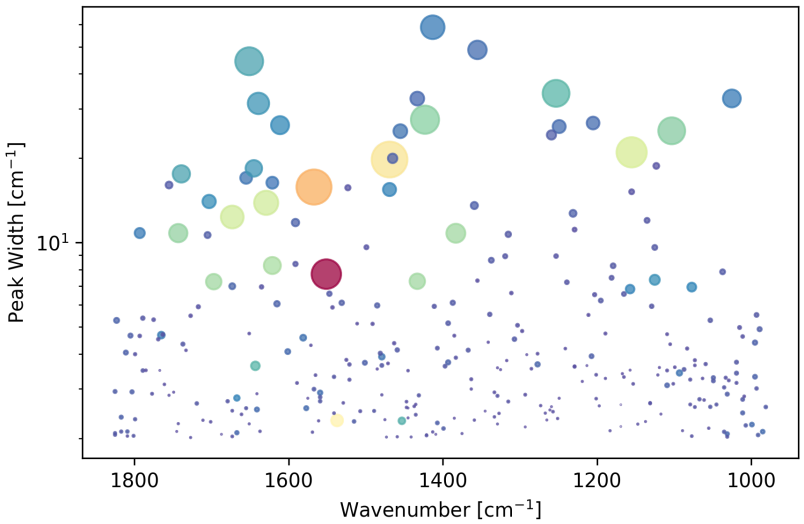
Then, we can visualize all peaks like so:

A



Significance of peaks vs. position. Prominence was normalized to $\max(\text{spectrum})$. The vertical lines highlight a selection of wavenumbers where we expect biosignatures.

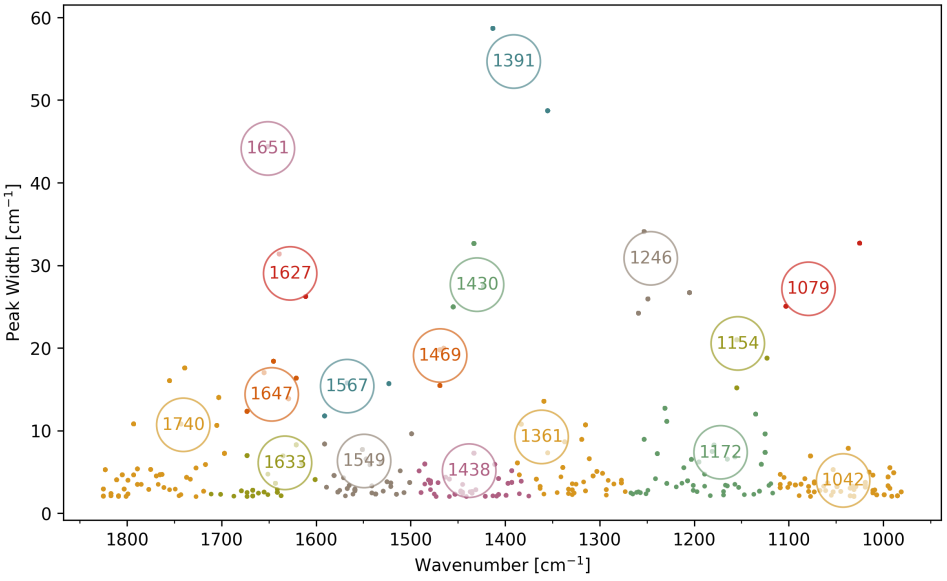
B



Here, peak width (y coordinate) and prominence (colour) are separated. (The spot size continues to signify significance). When searching for clusters, we should be aware that two fundamental peaks may lie at the same position but have different widths (like Amide-I and water absorption)

Next, we require a way to group these peaks into clusters.

I implemented weighted k-means clustering, tagging peaks by wavenumber and width, normalising width to 10x w.r.t. the wavelength dimension and duplicating datapoints by weight (width x prominence).



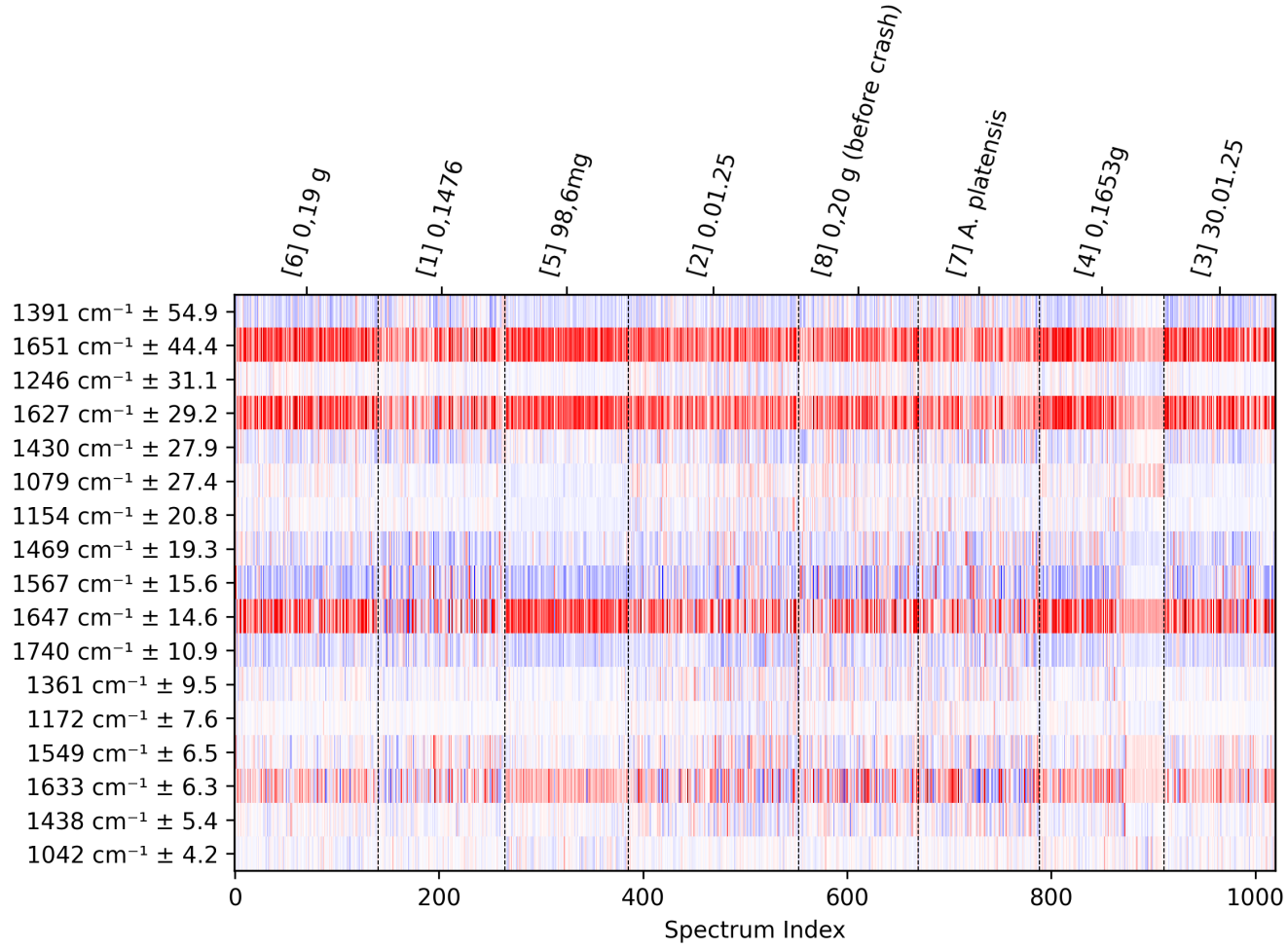
The following clusters of peaks are identified:

- 1391 cm^{-1} (width: 54.9) : **carboxylate symmetric stretch**
- 1651 cm^{-1} (width: 44.4) : **Amide I / Water** (tend more towards water)
- 1246 cm^{-1} (width: 31.1) : **symmetric PO4 stretch** or **C–N stretch**
- 1627 cm^{-1} (width: 29.2) : **Amide I / Water**
- 1430 cm^{-1} (width: 27.9) : **$CH_{2/3}$ scissoring/bending**
- 1079 cm^{-1} (width: 27.4) : **symmetric PO_4 stretch**
- 1154 cm-1 (width: 20.8) : **C–O–C stretch**
- 1469 cm^{-1} (width: 19.3) : **$CH_{2/3}$ scissoring/bending**
- 1567 cm^{-1} (width: 15.6) : **Amide II**
- 1647 cm^{-1} (width: 14.6) : **Amide I / Water** (tend more towards Amide I)
- 1740 cm^{-1} (width: 10.9) : **C=O stretch**
- 1361 cm^{-1} (width: 9.5)
- 1172 cm^{-1} (width: 7.6)
- 1549 cm^{-1} (width: 6.5) : **Amide II**

- 1633 cm^{-1} (width: 6.3) : **Amide I / Water**
- 1438 cm^{-1} (width: 5.4)
- 1042 cm^{-1} (width: 4.2)

Note: Position and width were given to the k-means algorithm without any transformation applied. Significance was used as a pseudo-weight, i.e. multiplicity of points. To achieve more "relevant" clusters, one could construct a transformation that virtually amplifies the distance between wider peaks. Then again, it should actually be more likely that a pair of peaks with a given distance in position should belong to the same cluster, if they are themselves wider.

As a measure of how well any of these peaks coincides with an actual spectrum, I calculated the scalar product of the spectrum with certain basis functions corresponding to each peak cluster's center. We get…



Red means positive, blue means negative.

Unfortunatetly, decomposing spectra into a basis of these peaks did not uncover any correlation between samples either.

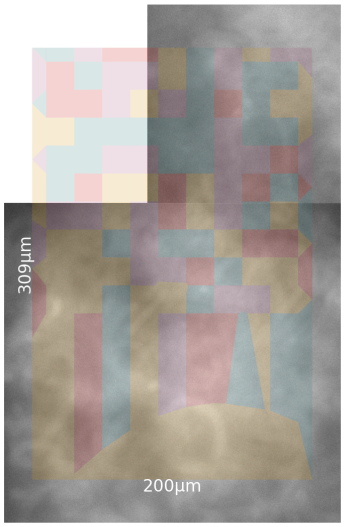
Hyperspectral Image Segmentation

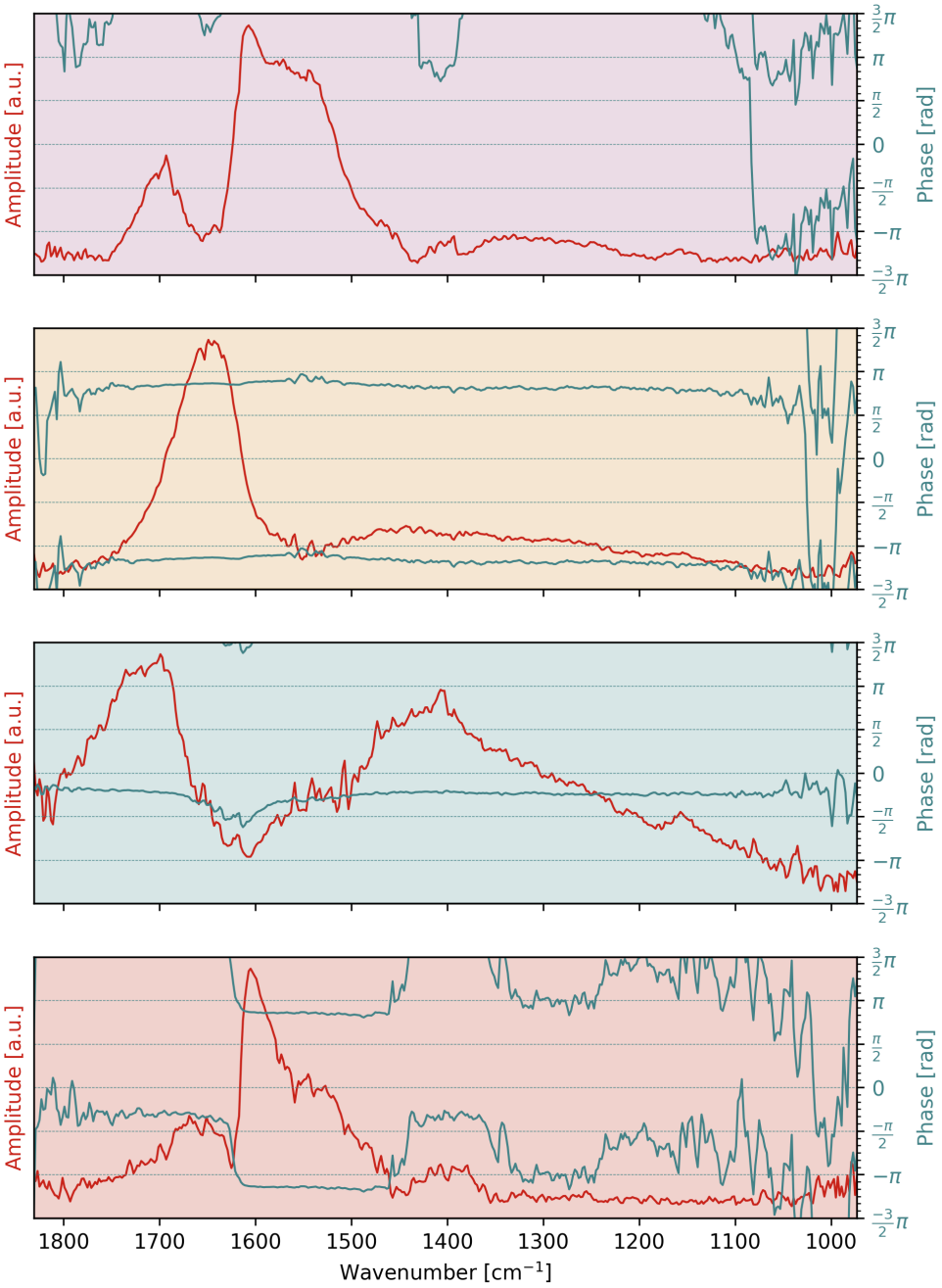
Instead of trying to model and track hypothetical spectral features through all measurements, let's attempt to classify spectra and see if these classifications are spatially correlated.

I calculated k-means clusters of the spectra in each sample (separately). I prescribed the number of clusters to be 4. We consider the X and Y channels.

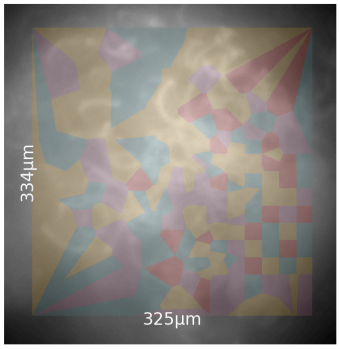
Each region of the image is then labelled with the cluster that its spatially closest spectrum belongs to. In the following, labelled images are presented on the left and representative spectra of the 4 clusters (the cluster centers) are shown on the right.

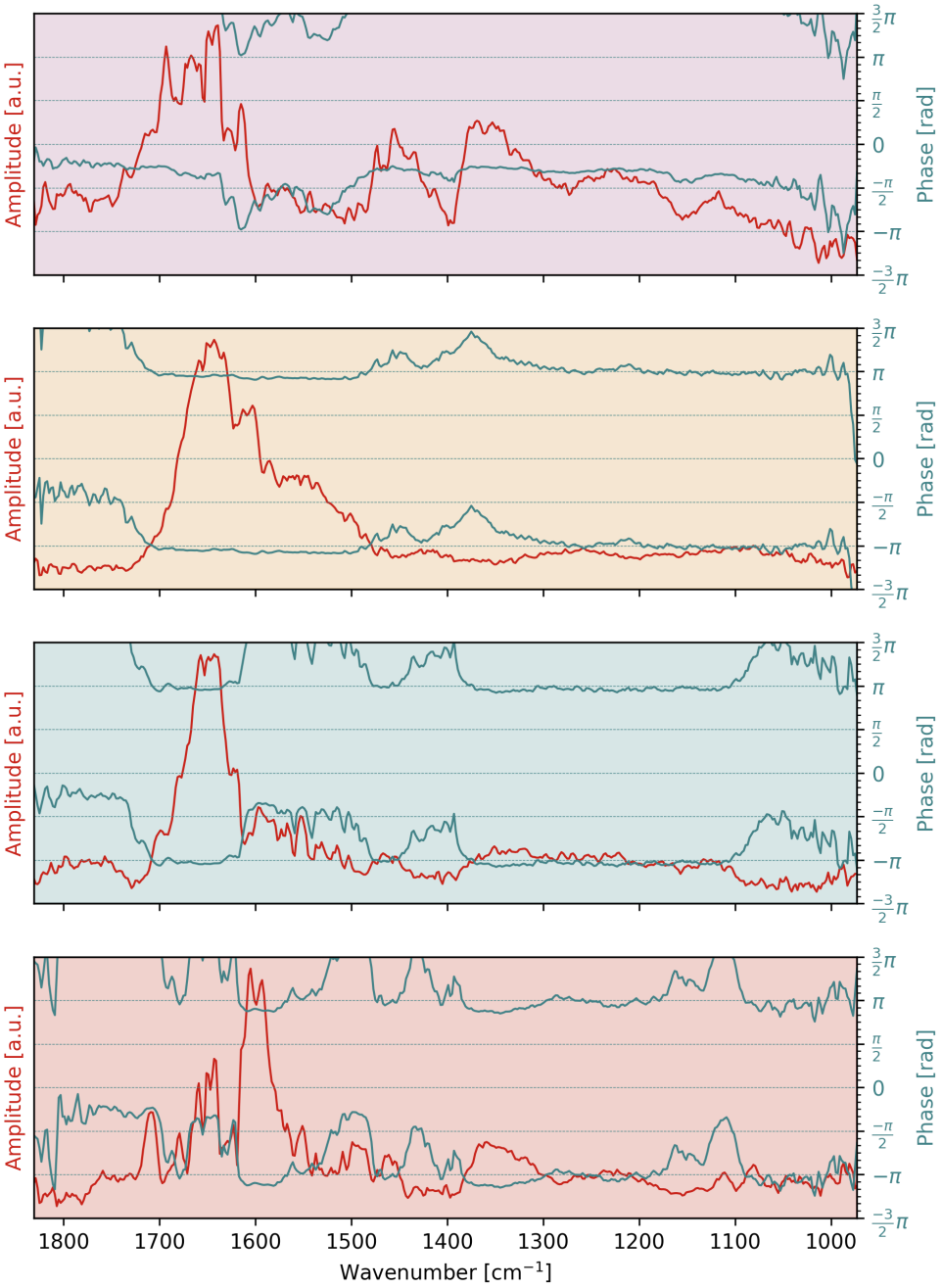
[1] 0,1476


309µm
200µm

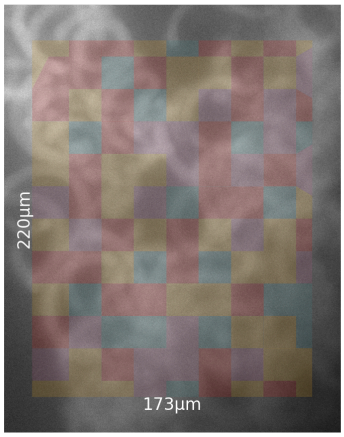

Amplitude [a.u.]
Phase [rad]
$\frac{3}{2}\pi$
π
$\frac{\pi}{2}$
0
$\frac{-\pi}{2}$
$-\pi$
$\frac{-3}{2}\pi$
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]

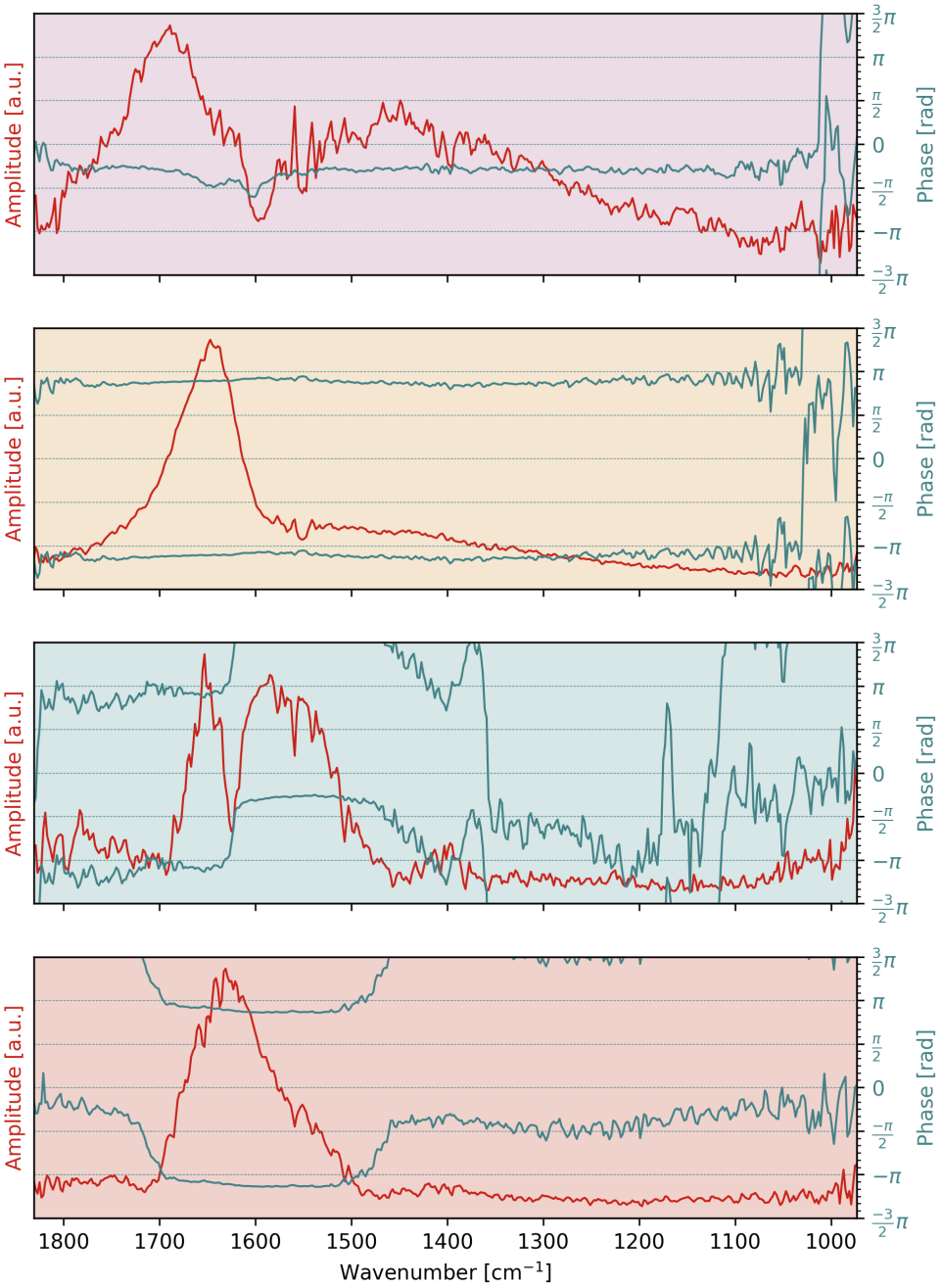
[2] 0.01.25


334µm
325µm

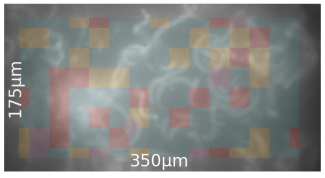

Amplitude [a.u.]
Phase [rad]
$\frac{3}{2}\pi$
π
$\frac{\pi}{2}$
0
$\frac{-\pi}{2}$
$-\pi$
$\frac{-3}{2}\pi$
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]

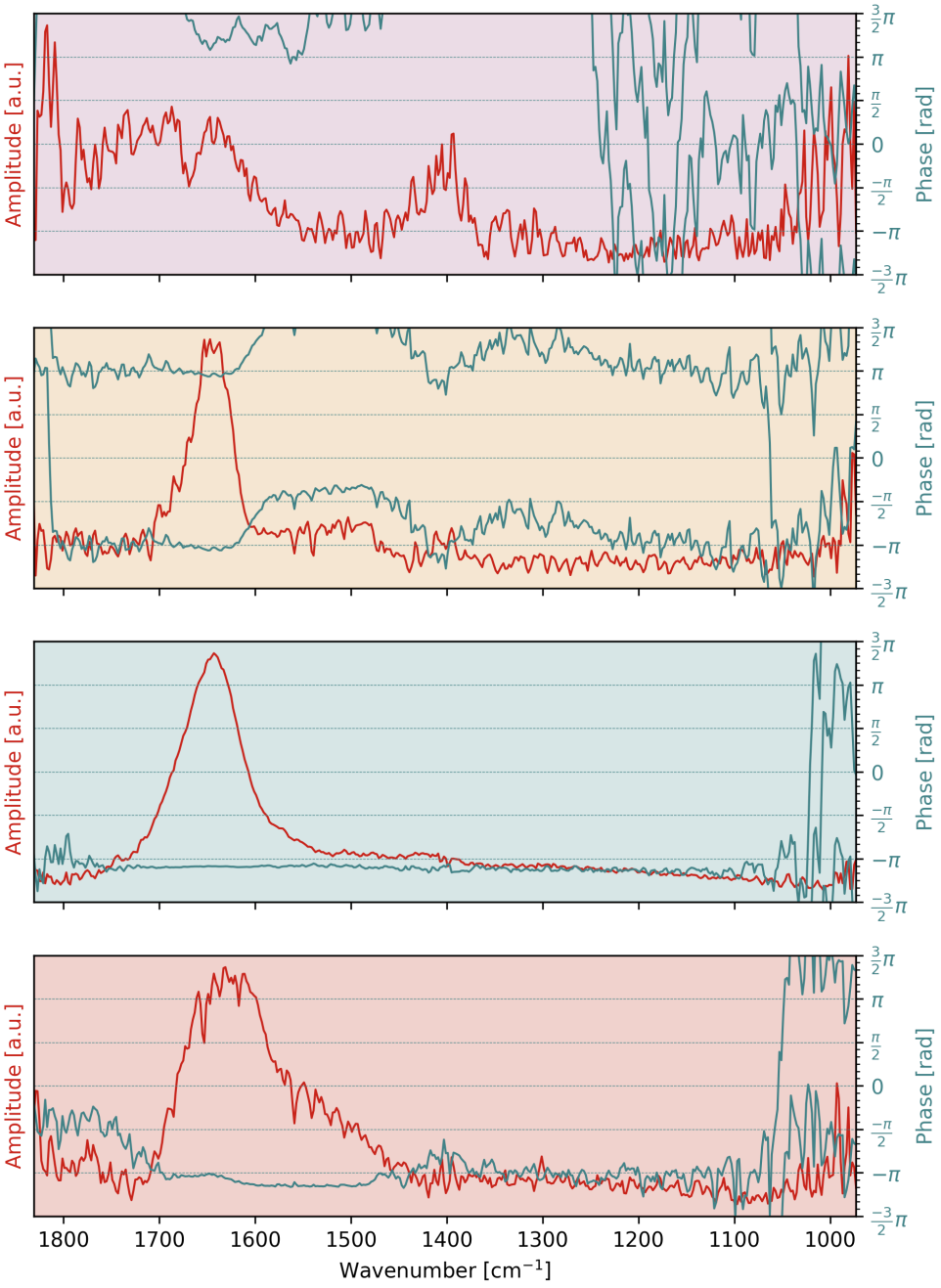
[3] 30.01.25


220µm
173µm

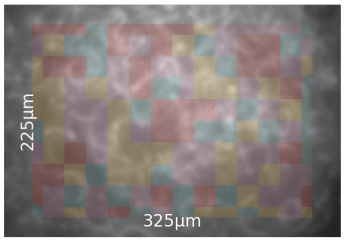

Amplitude [a.u.]
Phase [rad]
$\frac{3}{2}\pi$
π
$\frac{\pi}{2}$
0
$\frac{-\pi}{2}$
$-\pi$
$\frac{-3}{2}\pi$
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]

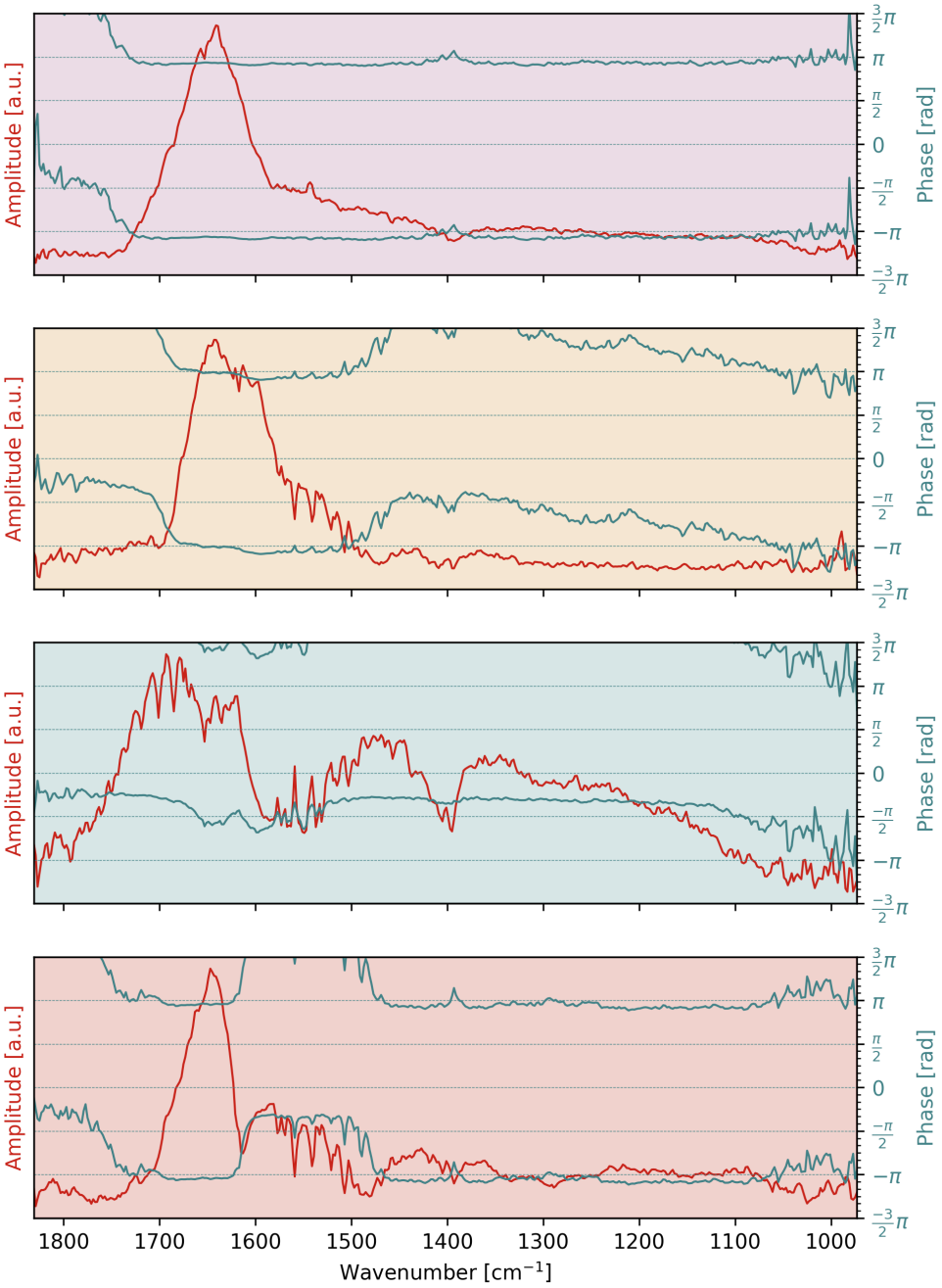
[5] 98,6mg


175µm
350µm

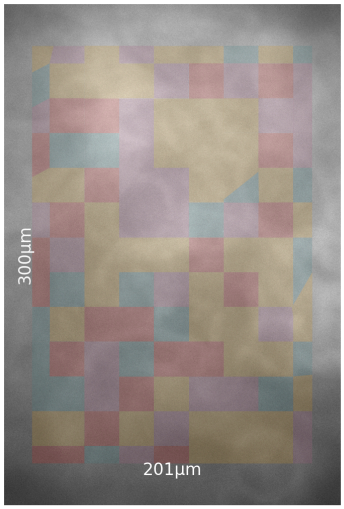

Amplitude [a.u.]
Phase [rad]
$\frac{3}{2}\pi$
π
$\frac{\pi}{2}$
0
$\frac{-\pi}{2}$
$-\pi$
$\frac{-3}{2}\pi$
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]

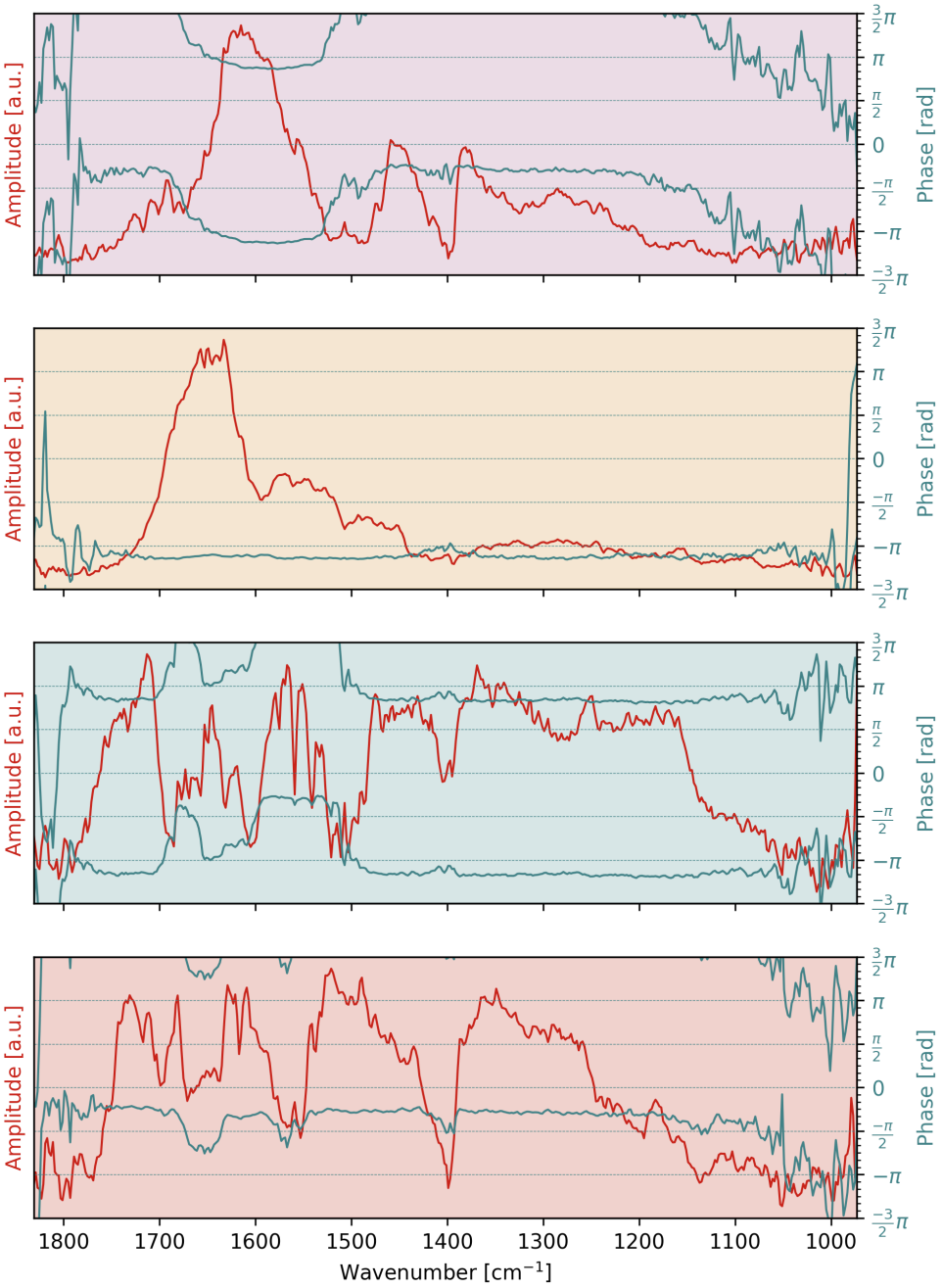
[6] 0,19 g


225µm
325µm

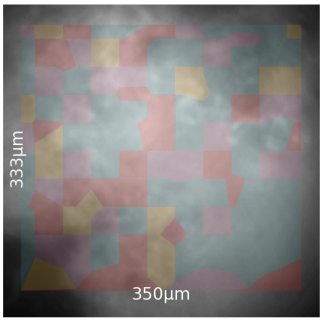

Amplitude [a.u.]
Phase [rad]
$\frac{3}{2}\pi$
π
$\frac{\pi}{2}$
0
$\frac{-\pi}{2}$
$-\pi$
$\frac{-3}{2}\pi$
1800
1700
1600
1500
1400
1300
1200
1100
1000
Wavenumber [cm^{-1}]

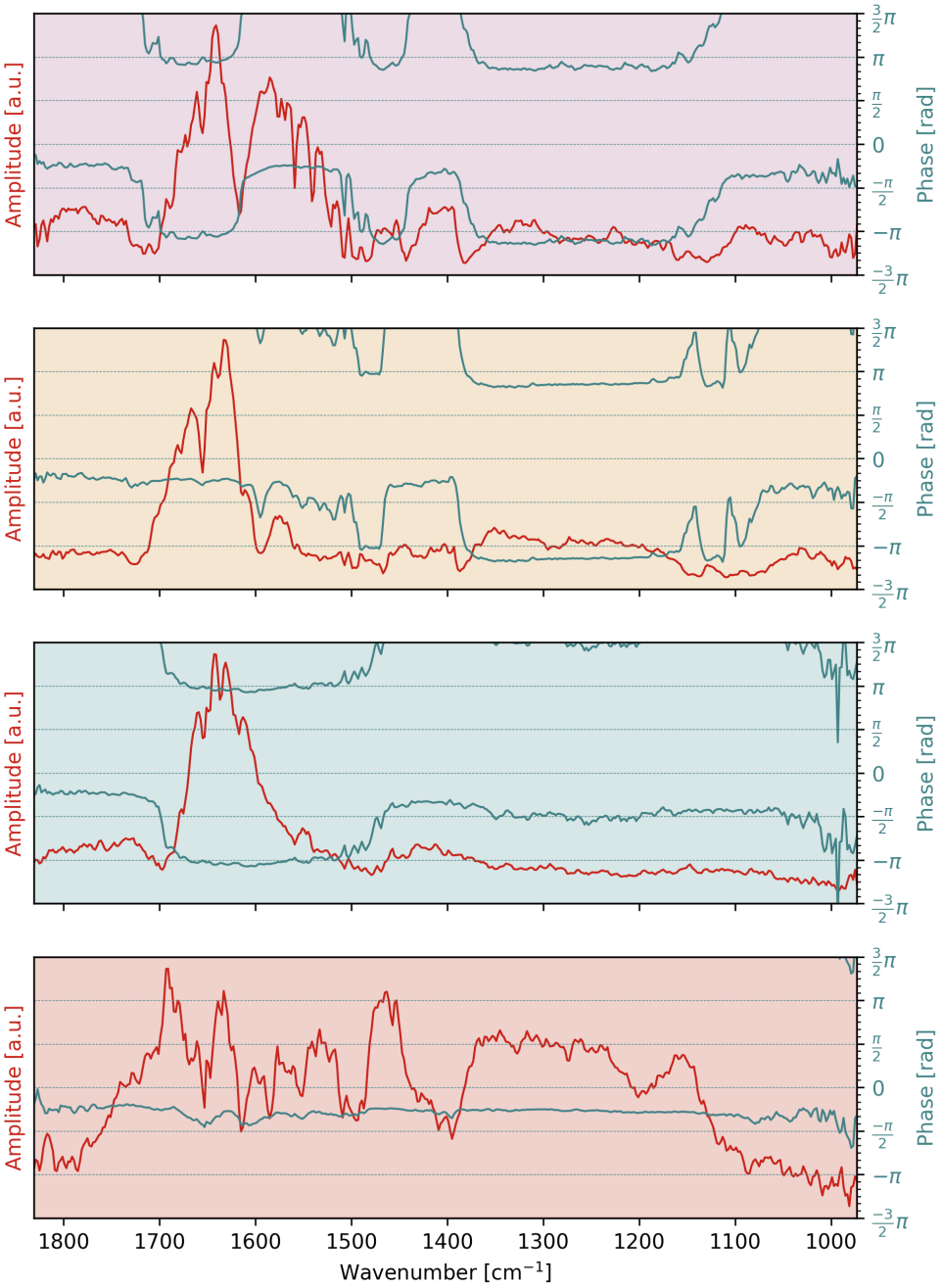
[7] A. platensis



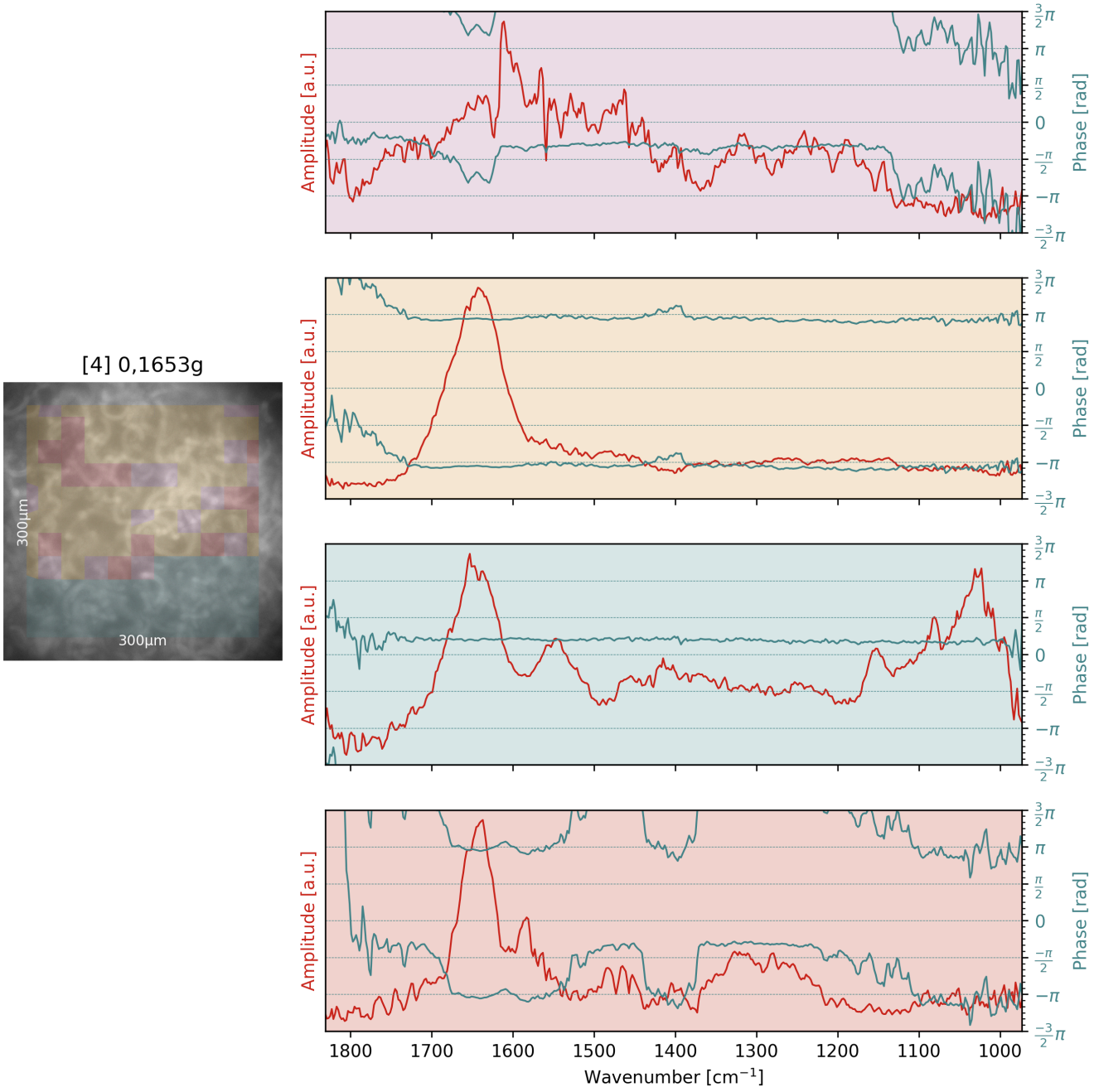


[8] 0,20 g (before crash)





In most of these, there seems to be no apparent spatio-spectral correlation.



Here, however, a clear spatial correlation is apparent. The upper two thirds of the image is labelled mostly as cluster 3 (yellow) with some clusters 1 (red) and 4 (purple) in between. Meanwhile, (something like) the lower third is exclusively labelled as cluster 2 (blue).

Both clusters 2 and 3 have, at their centres, comparatively low-noise spectra. Note that only cluster 2 shows the secondary peaks in the region between 1000 and 1200 cm^{-1} and that its Amide-II peak is the most prominent one and the one placed most closely to 1550 cm^{-1}.

One might surmise that in the blue-labelled region, and, to a lesser extent, in the red-labelled region, both proteins and lipids are being detected, while in the yellow-coloured region, only water is being detected. This conclusion is supported by the phase of the blue and yellow cluster centres: The phase of the yellow spectrum is close to constant at a value of $\pm\pi$. In the blue spectrum, the phase is close to constant as well, but around a value of approximately $+\frac{\pi}{6}$.

This can be assumed to be caused by different relaxation times for the IR-induced temperature profile, which, in turn, suggest that, in one case, a large volume of the solvent (MilliQ water) is being heated and in the other, a smaller volume of a biological structure being heated by the IR.