Hierarchical Cluster Analysis of hyperspectral Raman images:

a new point of view leads to 10 000× speedup

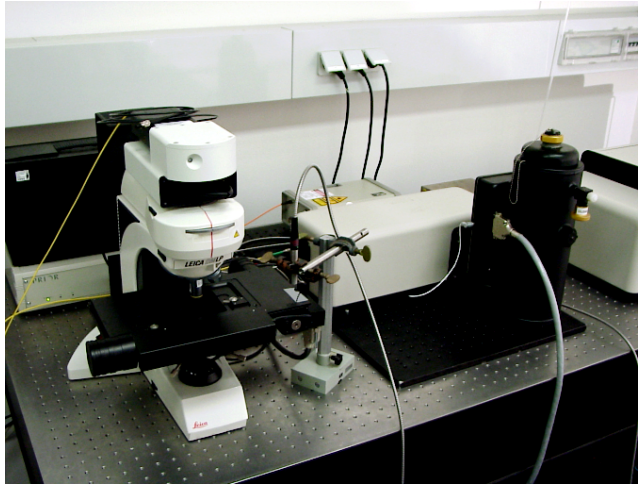
UseR!, Aalborg, 2015-07-01

C. Beleites[1], A. Bonifacio[2], C. Krafft[1], V. Sergo[2] and J. Popp[1,3]

[1] Dept. Spectroscopy and Imaging, IPHT Jena, Jena, Germany
[2] Dept. Arch. and Engineering, Univ. of Trieste, Trieste, Italy
[3] Inst. of Physical Chemistry and Abbe School of Photonics, University of Jena, Jena, Germany

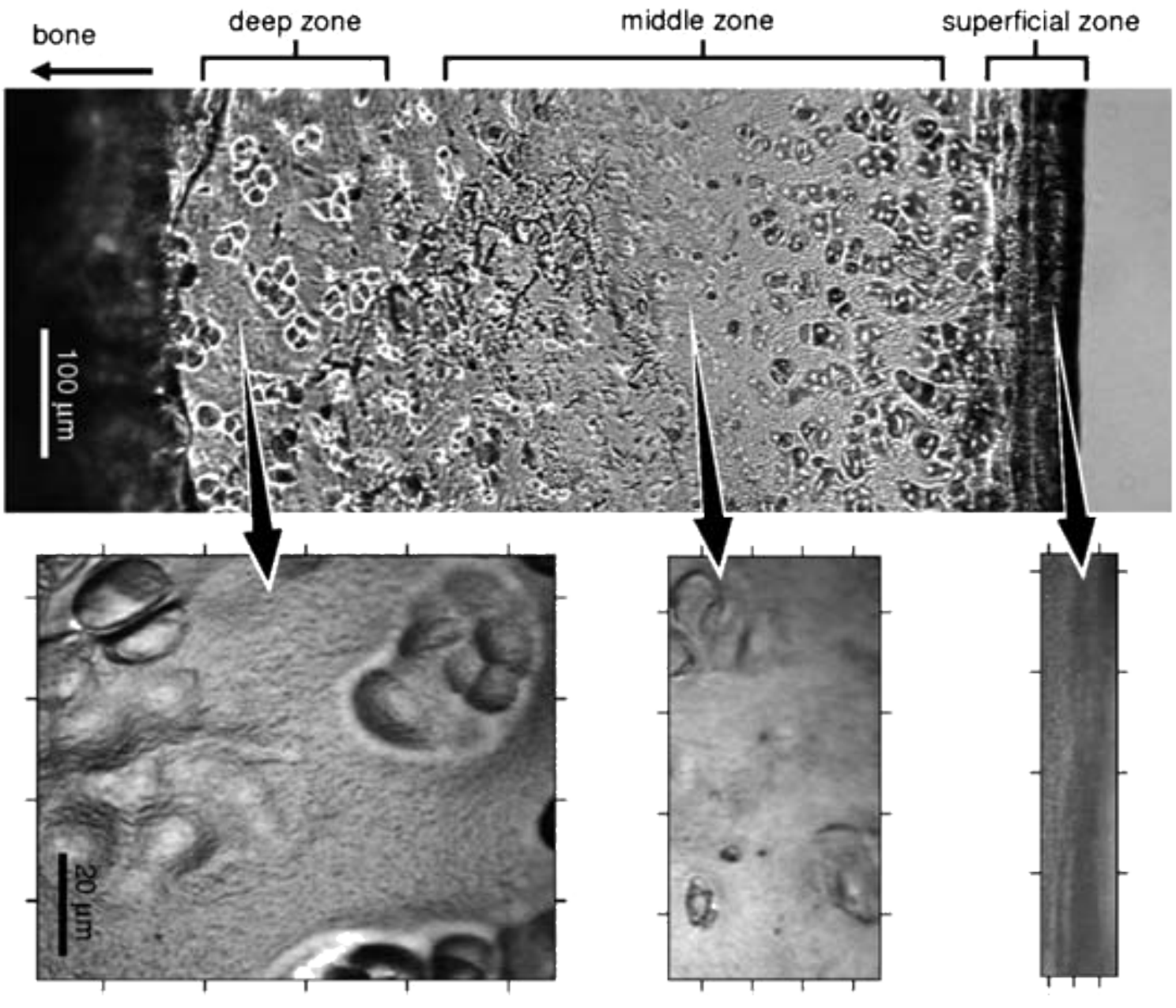
Mitglied der Leibniz-Gemeinschaft

Vibrational Spectroscopy

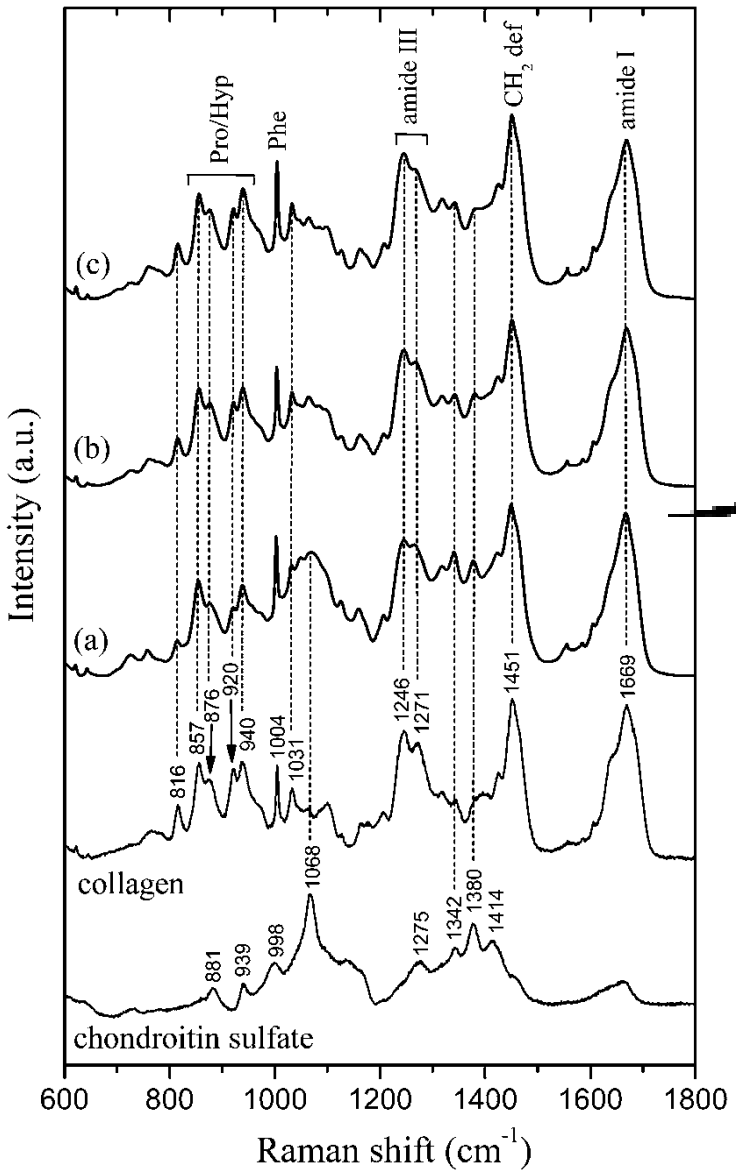


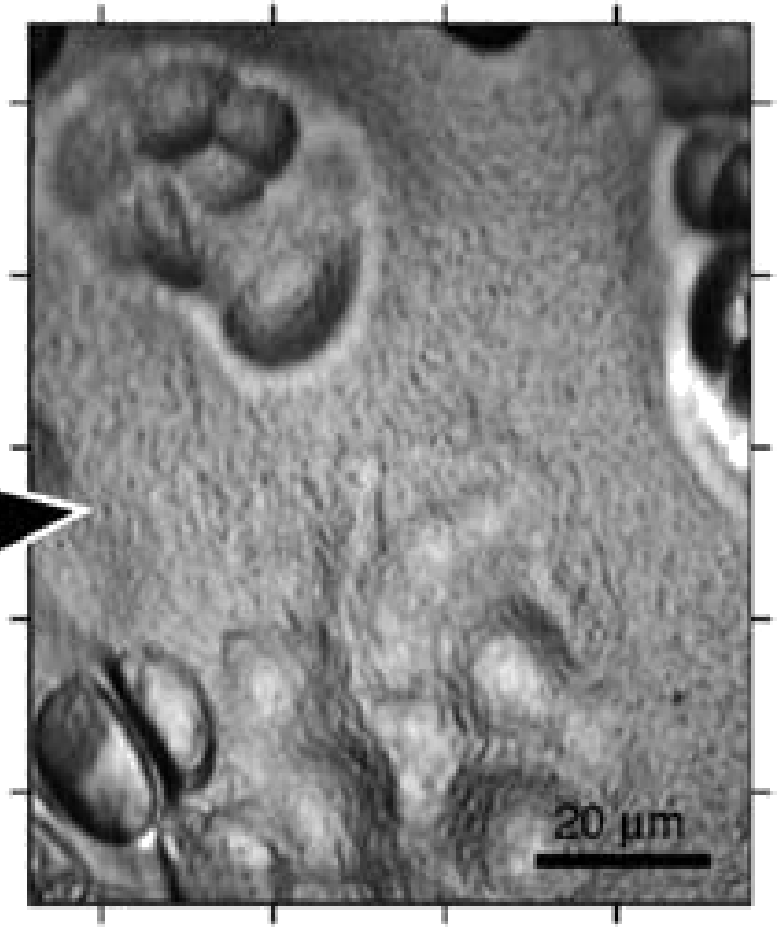
- Molecular vibrations:
  atoms vibrate against each other
- *n*-atomic molecule:
  $3n - 6$ different vibrations
- Excite vibration: discrete energy
- Characteristic bands
  ⇒ *biochemical composition*
  Lipids, proteins, carbohydrates, . . .
- Fingerprint region
  ⇒ *identify substance*
  ⇒ *identify cell or tissue type*

Chondrocytes in Articular Cartilage



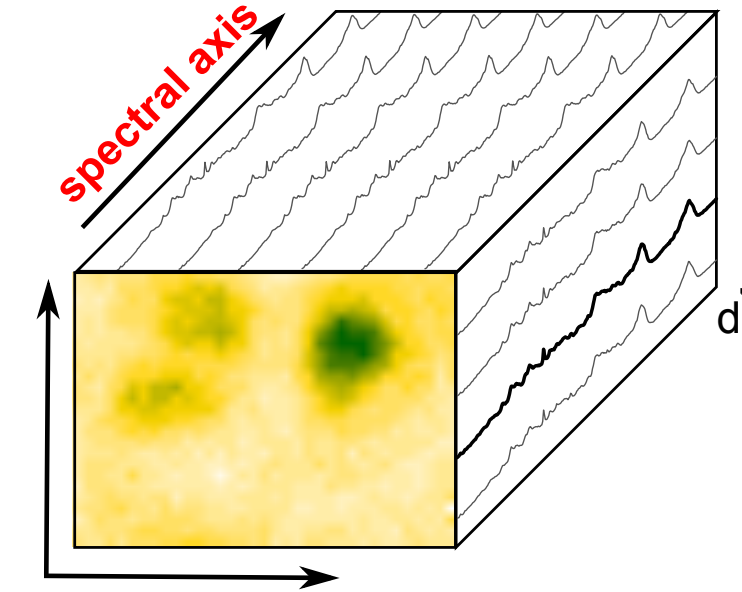
Raman spectra: Articular Cartilage



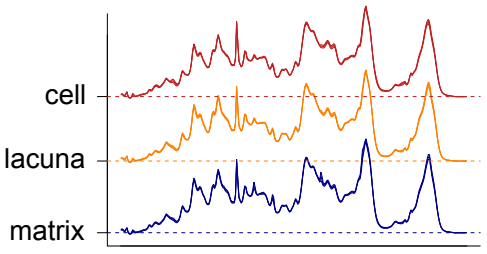


Bonifacio *et al.*, Analyst, 135 (2010) 3194 – 3204. DOI: 10.1039/c0an00459f

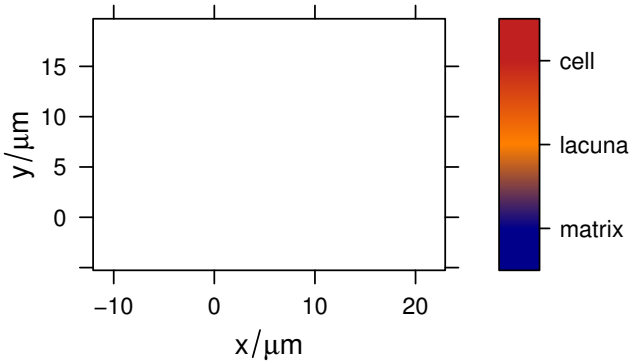
Hyperspectral Data



statistical
data analysis

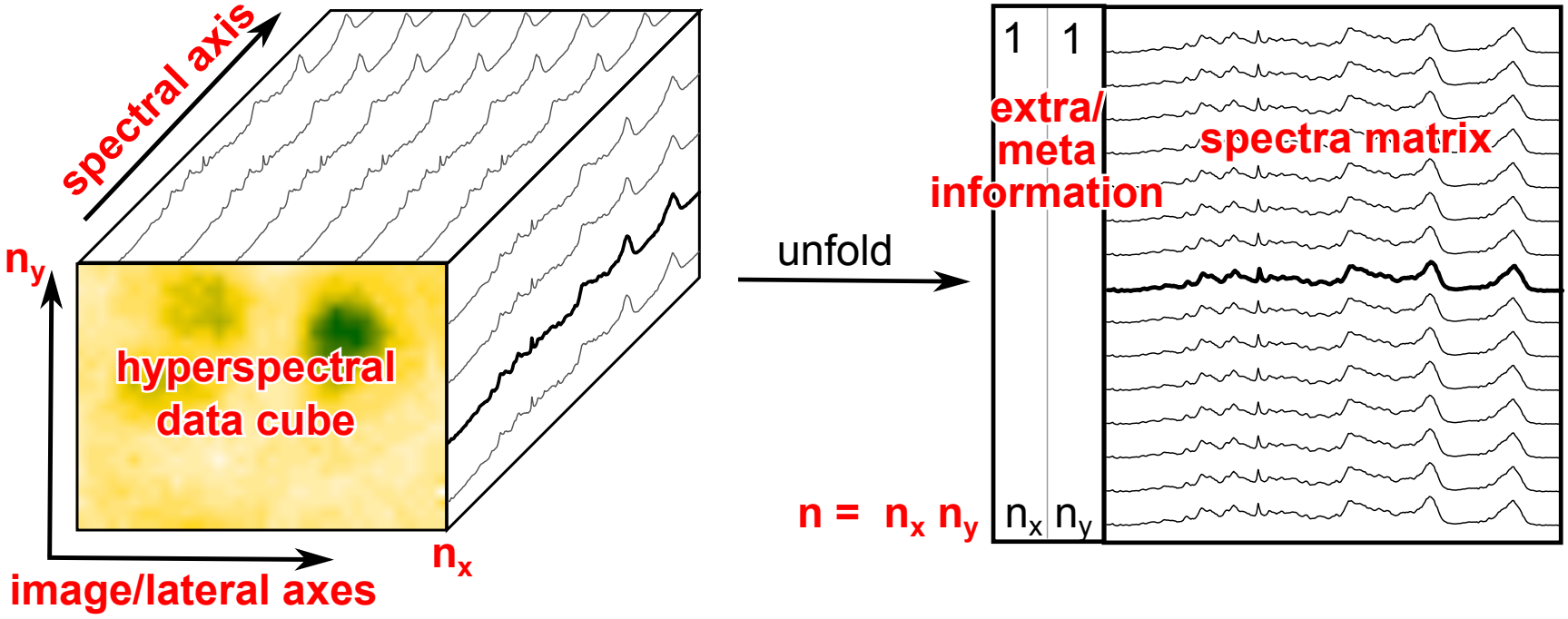


spectral information



spatial information

Hyperspectral Data



Cluster analysis

Hierarchical

- ✔ finds small clusters among many other spectra
- ✔ get number of clusters from dendrogram

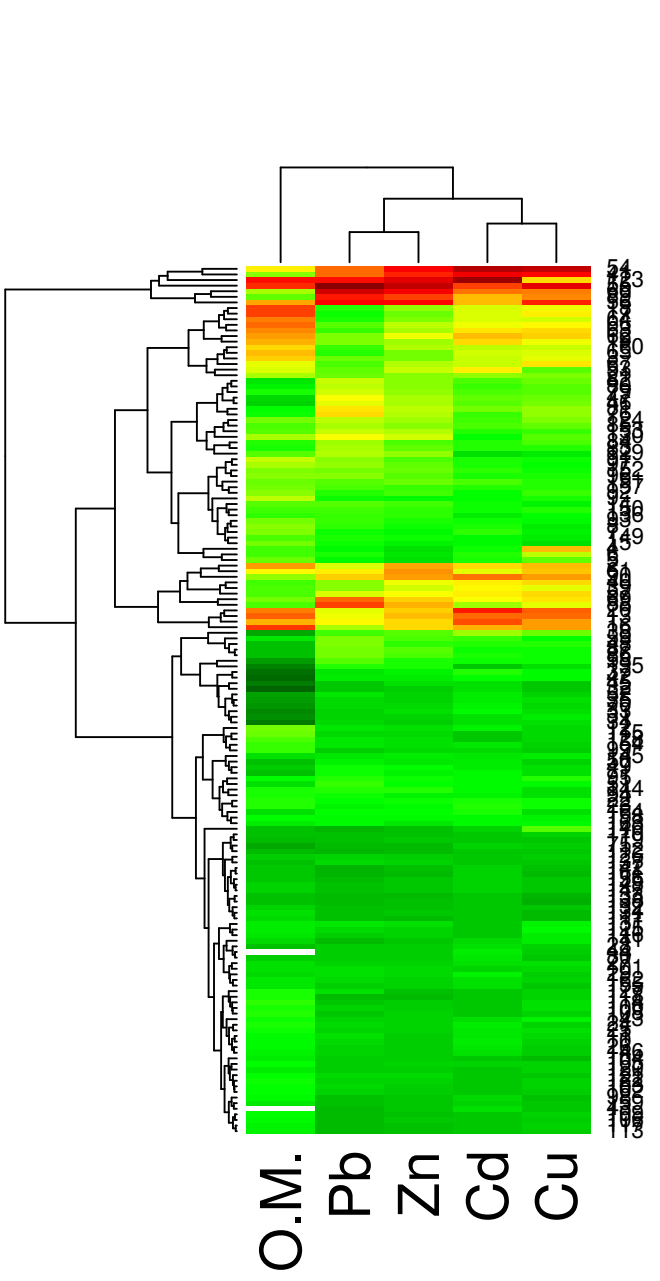
- ✘ memory use: $\approx n^2$
- ✘ run time:
  - $\sim n^2$
  - *out-of-memory* algorithms
  - z. B. `fastcluster`
    Müllner D, JStatSoft, 53, 1 – 18.

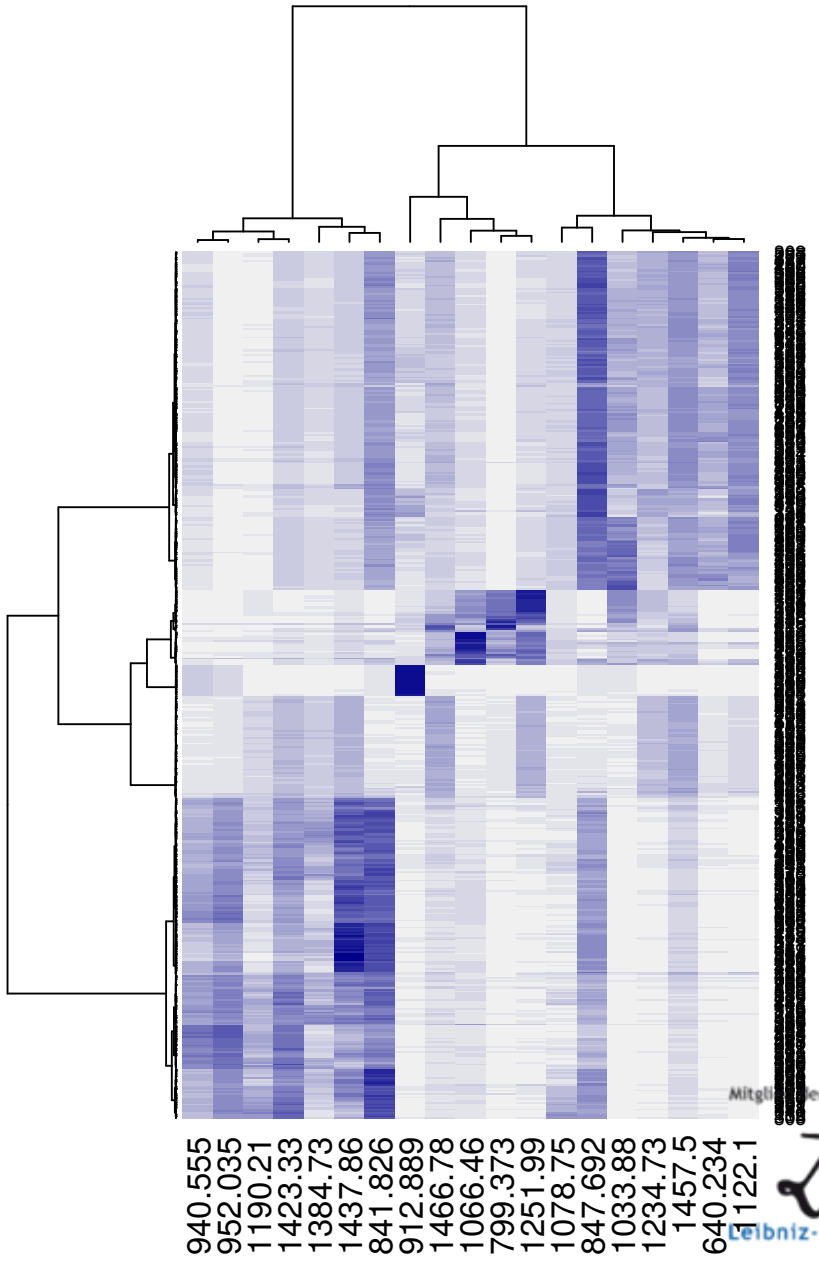
***k*-means**

- ✘ small clusters may not be found
- ✘ number of clusters must be known beforehand

- ✔ memory use: uncritical
- ✔ fast
- ✘ sometimes sensitive to initialization

Heatmaps


O.M.
Pb
Zn
Cd
Cu


940.555
952.035
1190.21
1423.33
1384.73
1437.86
841.826
912.889
1466.78
1066.46
799.373
1251.99
1078.75
847.692
1033.88
1234.73
1457.5
640.234
1122.1

Cluster analysis the other way round

What happens if *images* are clustered rather than *spectra*?

Bonifacio, Beleites & Sergo: Anal Bioanal Chem, 407 (2015), 1089 – 1095
DOI: 10.1007/s00216-014-8321-7

Cluster analysis the other way round

What happens if *images* are clustered rather than *spectra*?

⇝ Clusters of spectral bands that have similar (spatial) distribution.

Bonifacio, Beleites & Sergo: Anal Bioanal Chem, 407 (2015), 1089 – 1095
DOI: 10.1007/s00216-014-8321-7

Cluster analysis the other way round

What happens if *images* are clustered rather than *spectra*?

⇝ Clusters of spectral bands that have similar (spatial) distribution.

✔ Similar to spectra interpretation work flow

Bonifacio, Beleites & Sergo: Anal Bioanal Chem, 407 (2015), 1089 – 1095
DOI: 10.1007/s00216-014-8321-7

Cluster analysis the other way round

What happens if *images* are clustered rather than *spectra*?

⇝ Clusters of spectral bands that have similar (spatial) distribution.

✔ Similar to spectra interpretation work flow

⚠ Several substances may share the same spectral band position

- band may be assigned with only one substance, or
- neighbour channels are assigned to different substances
- substances form subclusters

Bonifacio, Beleites & Sergo: Anal Bioanal Chem, 407 (2015), 1089 – 1095
DOI: 10.1007/s00216-014-8321-7

Cluster analysis the other way round

What happens if *images* are clustered rather than *spectra*?

⇝ Clusters of spectral bands that have similar (spatial) distribution.

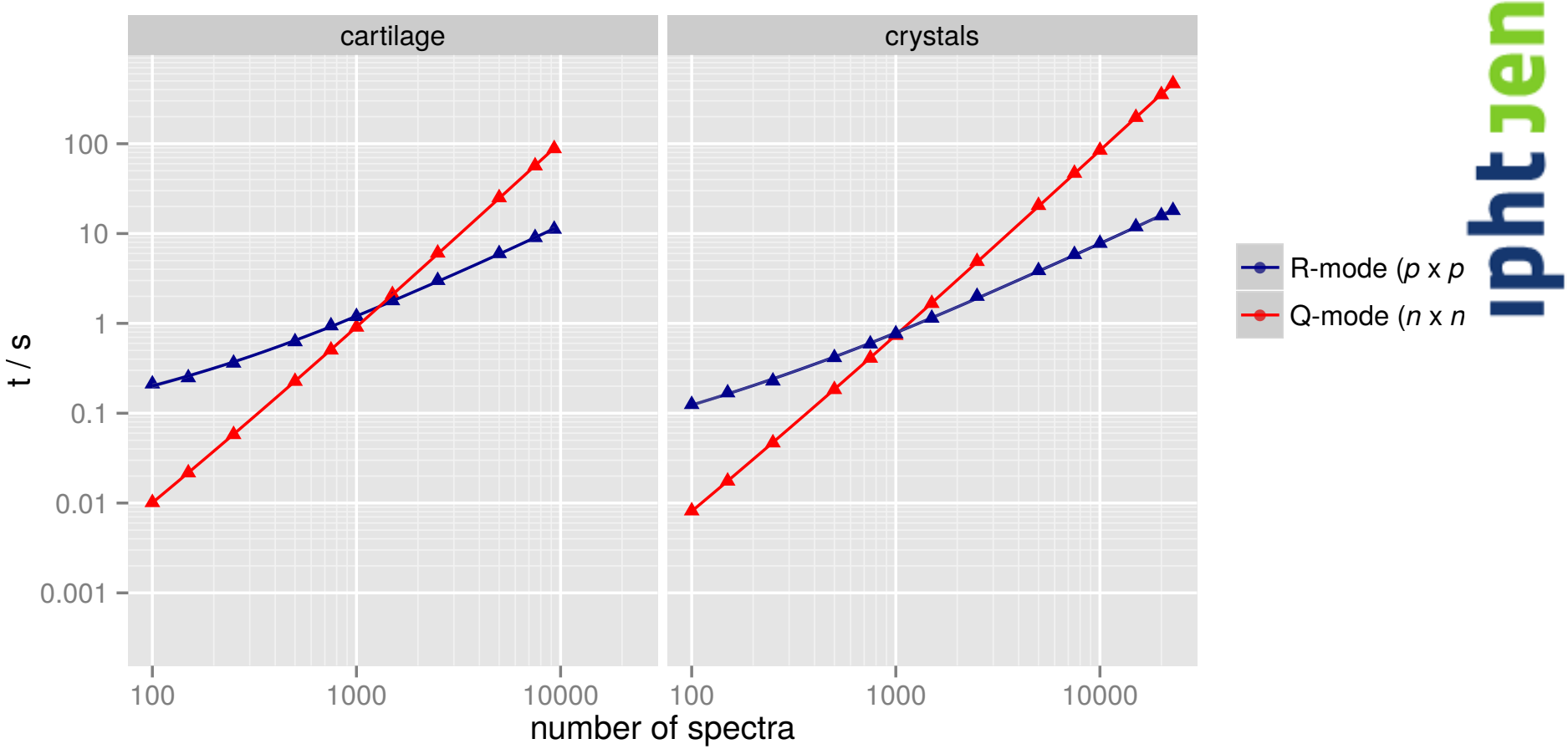
✔ Similar to spectra interpretation work flow

⚠ Several substances may share the same spectral band position
- band may be assigned with only one substance, or
- neighbour channels are assigned to different substances
- substances form subclusters

✔ (Spatial) distribution is not any more separated into hard clusters, agrees well with continuous concentration distributions.

Bonifacio, Beleites & Sergo: Anal Bioanal Chem, 407 (2015), 1089 – 1095
DOI: 10.1007/s00216-014-8321-7

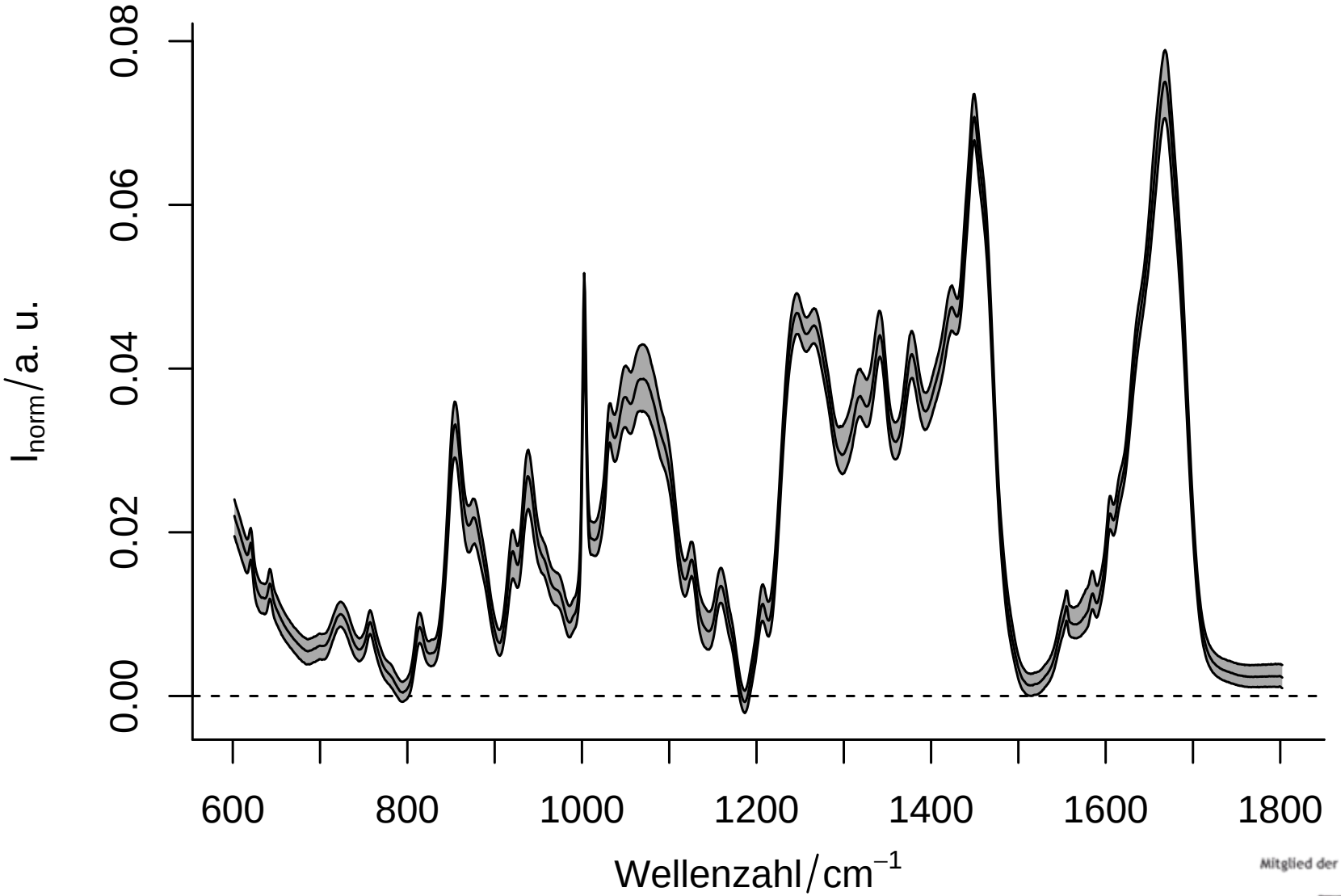
Hierarchical Cluster Analysis: Run time



cartilage from articular joint
$n = 9\,304$ Raman spectra $\times p = 1\,272$ wavenumber channels

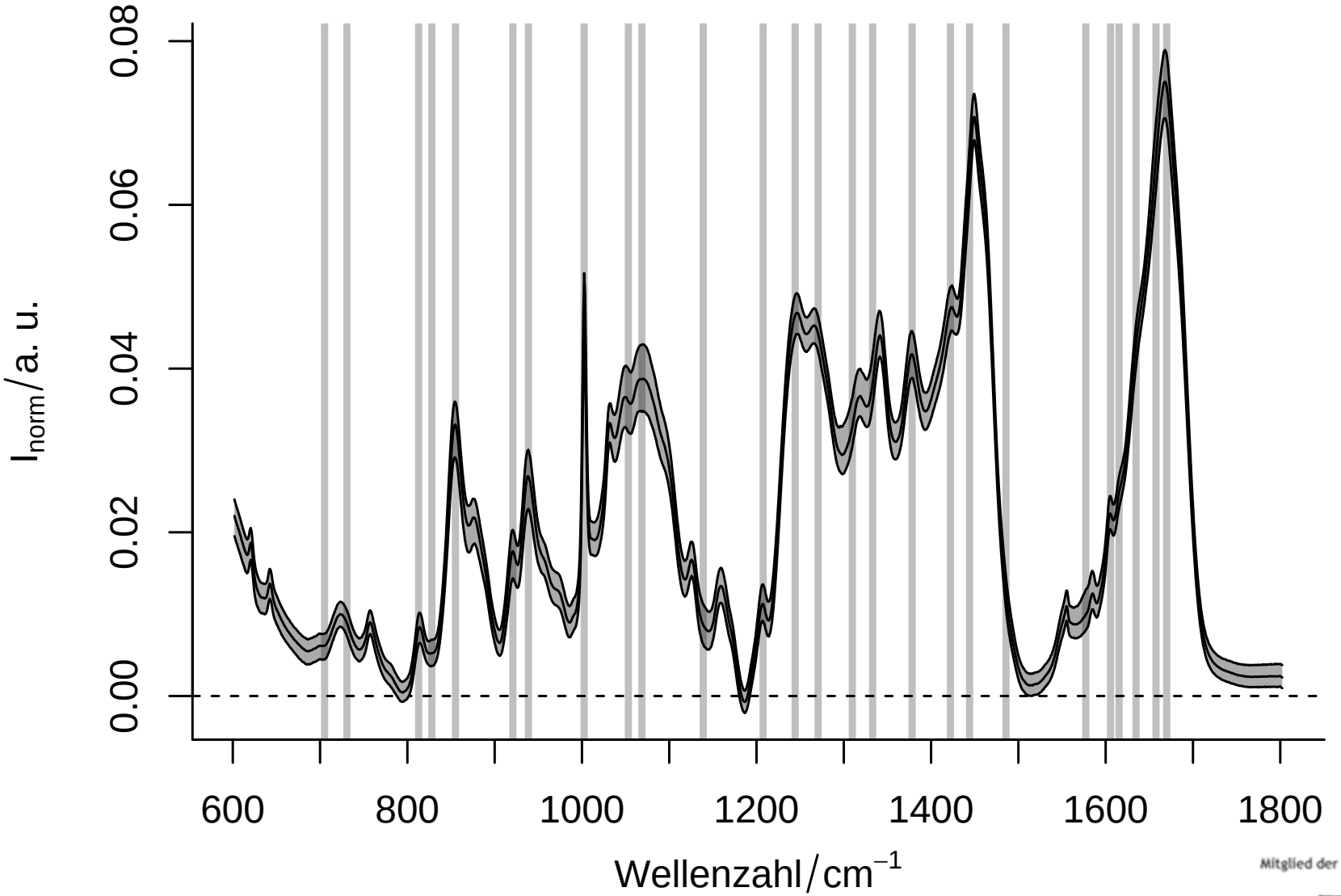
crystals heterogeneous mixture of 4 different substances
$n = 22\,801$ Raman spectra $\times p = 1\,019$ wavenumber channels

Raman spectra: cartilage



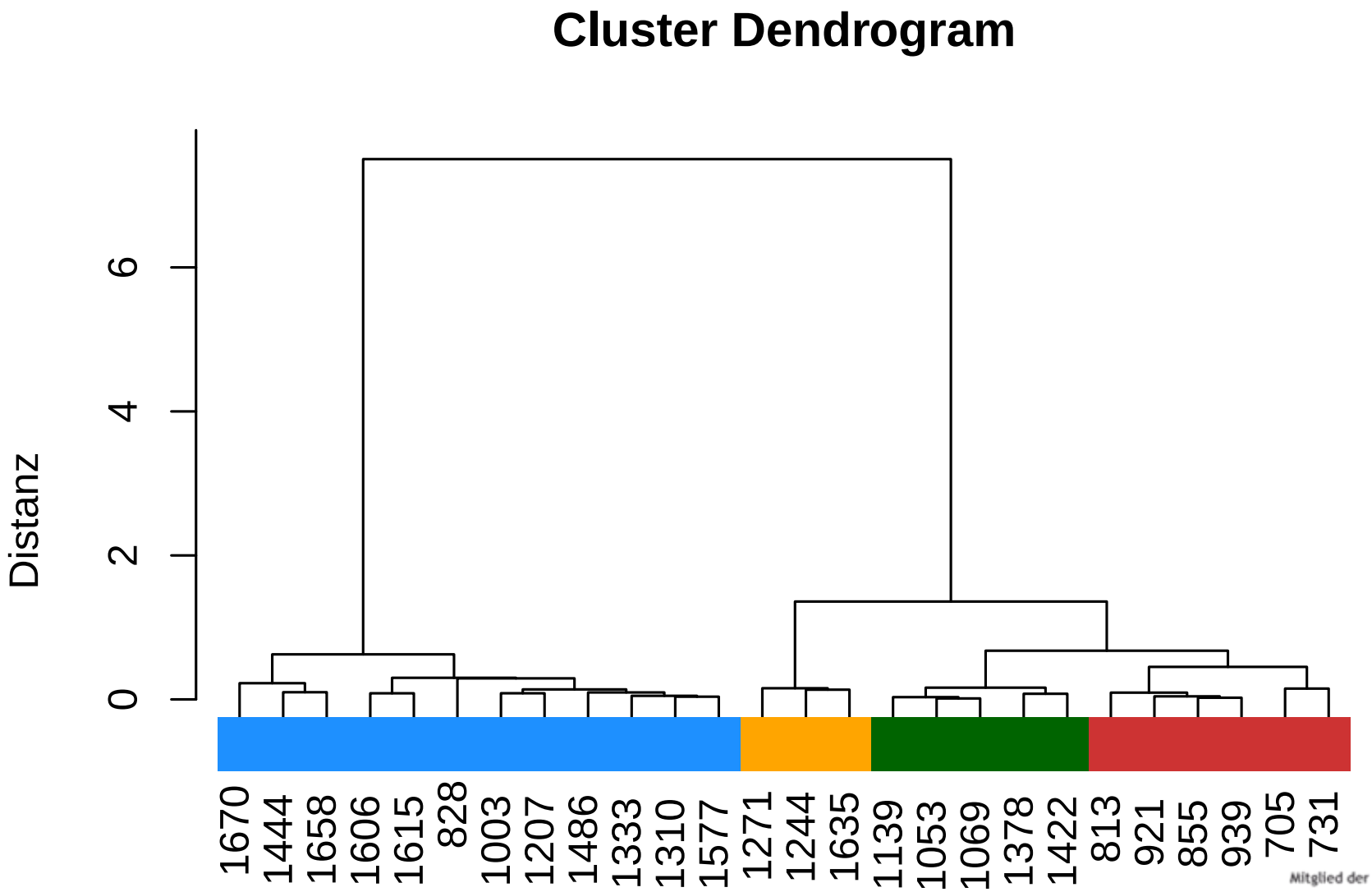
Mitglied der Leibniz-Gemeinschaft

Raman spectra: cartilage



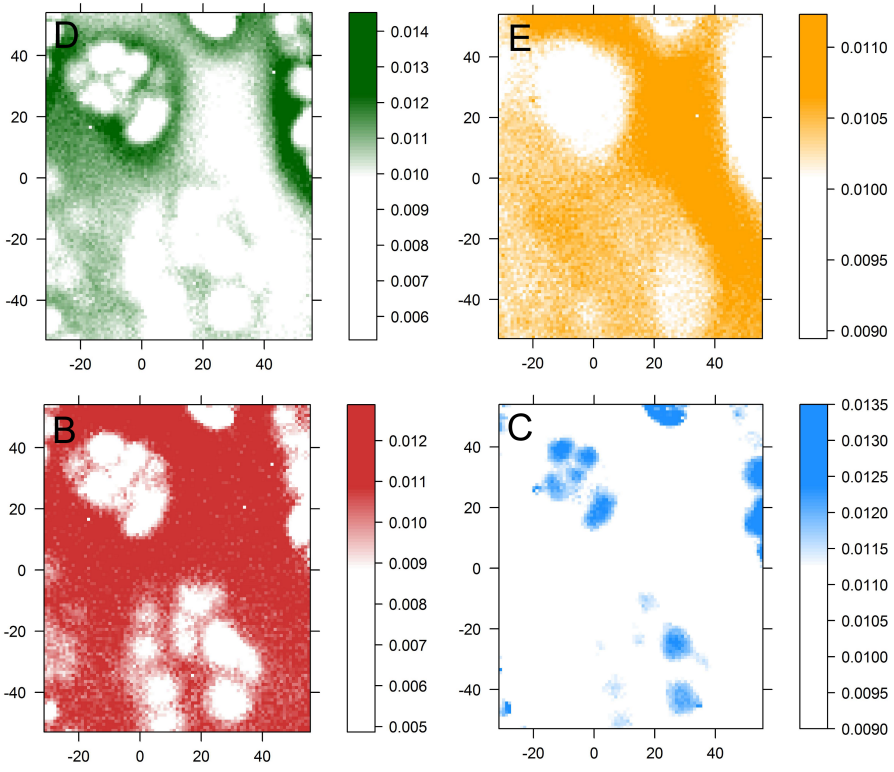
Mitglied der Leibniz-Gemeinschaft

Dendrogram selected bands



Mitglied der Leibniz-Gemeinschaft

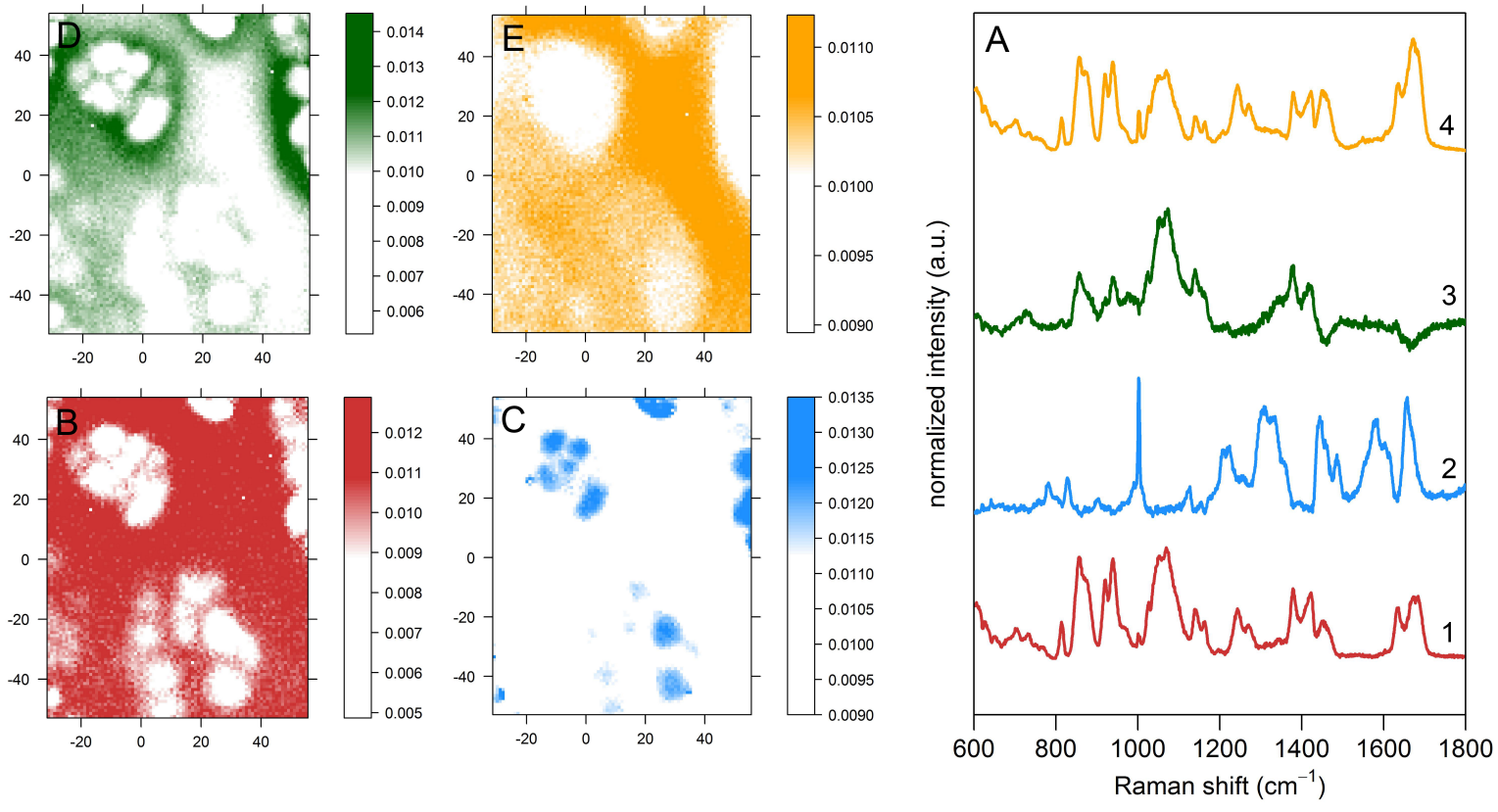
Results



Bonifacio *et al.*, Analyst, 135 (2010) 3194 – 3204. DOI: 10.1039/c0an00459f

Bonifacio, Beleites & Sergo: Anal Bioanal Chem, 407 (2015), 1089 – 1095
DOI: 10.1007/s00216-014-8321-7

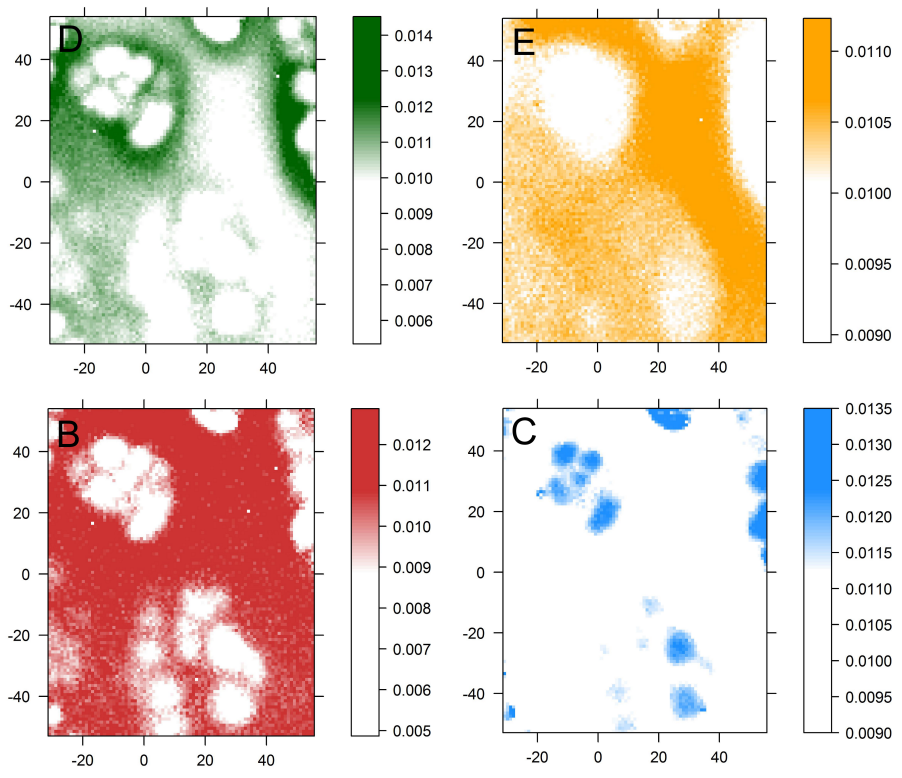
Results

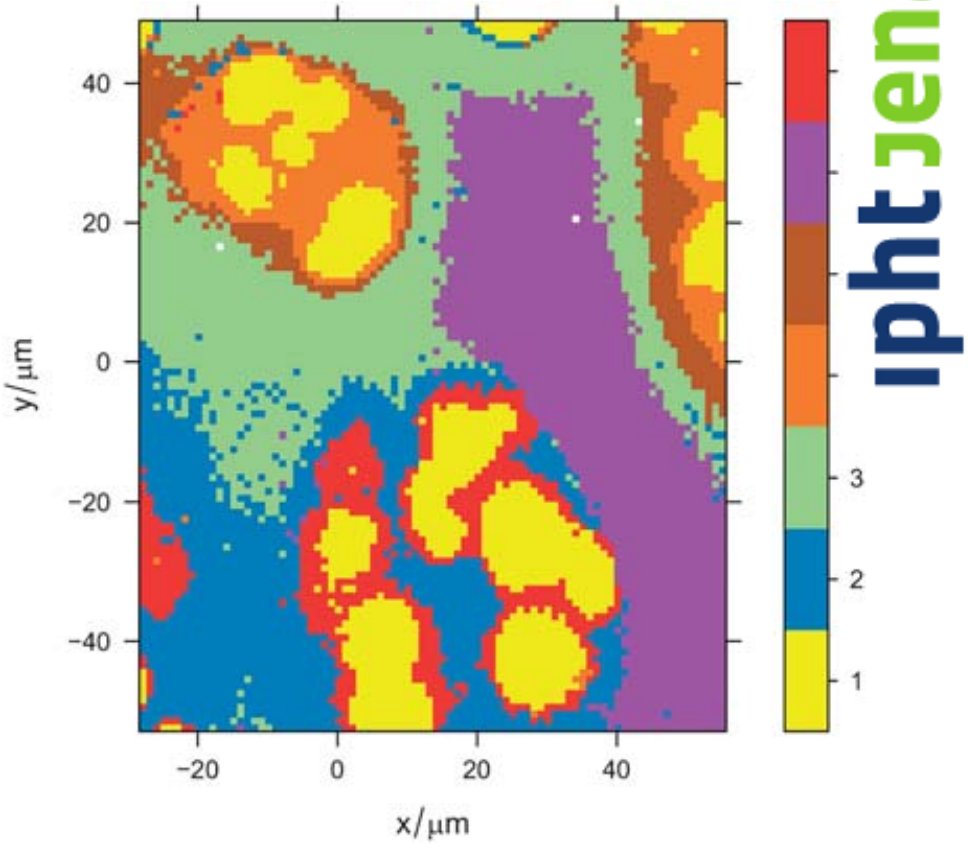


Bonifacio *et al.*, Analyst, 135 (2010) 3194 – 3204. DOI: 10.1039/c0an00459f

Bonifacio, Beleites & Sergo: Anal Bioanal Chem, 407 (2015), 1089 – 1095
DOI: 10.1007/s00216-014-8321-7

Results

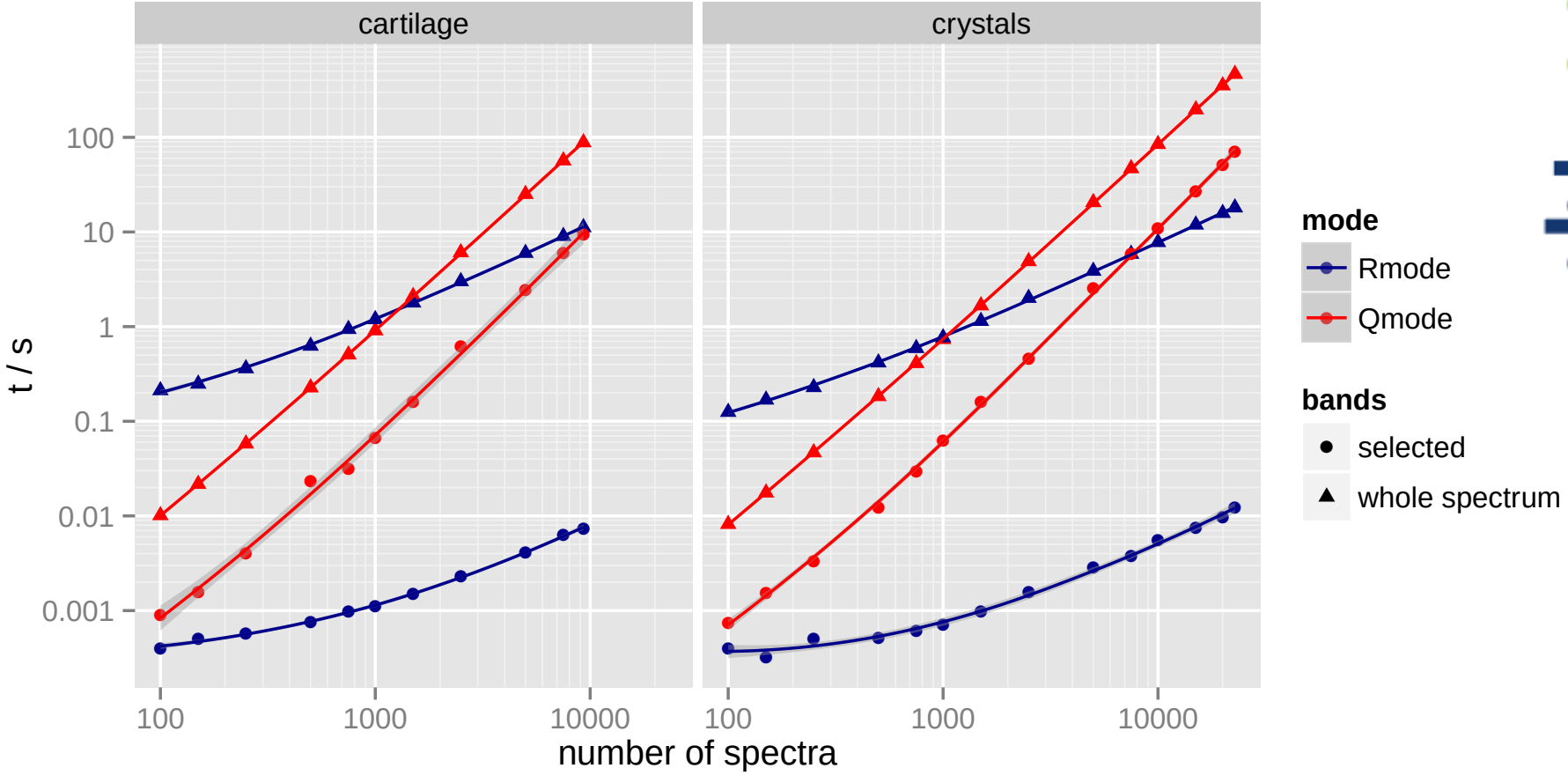




Bonifacio *et al.*, Analyst, 135 (2010) 3194 – 3204. DOI: 10.1039/c0an00459f

Bonifacio, Beleites & Sergo: Anal Bioanal Chem, 407 (2015), 1089 – 1095
DOI: 10.1007/s00216-014-8321-7

Hierarchical Cluster Analysis: Run time



Summary and Acknowledgements

- **Use domain-specific information**
- Vibrational spectroscopy: Clustering spectra vs. images
  ⇝ similar results
- Spectroscopic interpretation
- Reduce run-time by orders of magnitude

- Bonifacio *et al.*, Analyst, 135 (2010) 3194 – 3204.
  DOI: 10.1039/c0an00459f
- Bonifacio *et al.*: Anal Bioanal Chem, 407 (2015), 1089 – 1095.
  DOI: 10.1007/s00216-014-8321-7

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Mitglied der Leibniz-Gemeinschaft