universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Characterization of ammonia-oxidizing archaea by Raman microspectroscopy

Verfasser

Christoph Böhm

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Wien, 2012

Studienkennzahl lt. Studienblatt:	A441
Studienrichtung lt. Studienblatt:	Genetik - Mikrobiologie
Betreuerin / Betreuer:	Univ.-Prof. Dr. Michael Wagner

"... knowledge must continually be renewed by ceaseless effort, if it is not to be lost. It resembles a statue of marble which stands in the desert and is continually threatened with burial by the shifting sand. The hands of service must ever be at work, in order that the marble continue to lastingly shine in the sun. To these serving hands mine shall also belong." **Albert Einstein** (1950)

Table of contents

1 Introduction

1.1 Bacterial ammonia-oxidation

The oxidation of NH_3 to NO_2^-, the first step of nitrification, is one of the key processes in the global nitrogen cycle and also very important for human built systems like wastewater treatment plants (WWTPs). For more than a hundred years, this process is known to be performed by certain chemolithoautotrophic bacteria (Winogradsky, 1890) and today several ammonia-oxidizing genera within the *Beta-* and *Gammaproteobacteria* have been recognized (Teske et al., 1994; Purkhold et al., 2000). Biochemically, bacterial ammonia-oxidation consists of two-steps catalyzed by two enzymes, ammonia monooxygenase (Amo) and hydroxylamine oxidoreductase (Hao). In the first step NH_3 is oxidized to NH_2OH and in the second step it is then further oxidized to NO_2^- (Fig. 1.1).

Amo: $NH_3 + O_2 + H^+ + 2\ e^- \rightarrow NH_2OH + H_2O$

Hao: $NH_2OH + H_2O \rightarrow NO_2^- + 5\ H^+ + 4\ e^-$

Figure 1.1. Ammonia-oxidation of AOB catalyzed by two enzymes (Amo and Hao).

The bacterial Amo is a membrane-associated enzyme and it consists of three subunits (AmoA, AmoB and AmoC) encoded by the genes *amoA*, *amoB* and *amoC* (Klotz and Norton, 1998). The subunit AmoA is used as a phylogenetic marker of ammonia-oxidizing bacteria (AOB) (Purkhold et al., 2000) and contains the active site of the enzyme (Ensign et al., 1993) .

1.2 Ammonia-oxidizing archaea

Initially, microbial ammonia-oxidation was thought to be exclusively performed by certain bacterial species. However, this hypothesis was challenged a couple of years ago when unique *amo* genes were detected on an archaeal-associated metagenomic scaffolds from the Sargasso Sea (Venter, 2004) and on terrestrial metagenome fragments derived from unusual mesophilic *Crenarchaeota* (Treusch et al., 2005). The final proof for the existence of ammonia-oxidizing archaea (AOA) was delivered in the same year, when *Nitrosopumilus maritimus*, an autotrophic member of the so-called marine Crenarchaea capable of growing by the oxidation of ammonia, was isolated (Könneke et al., 2005).
In the meantime, cultivation-independent methods revealed that AOA occur in almost every nitrifying environment (Wuchter, 2004; Treusch et al., 2005; Francis et al., 2005; Beman and Francis, 2006; Leininger et al., 2006; Park et al., 2006; Lam et al., 2007; Mincer et al., 2007; Nakagawa et al., 2007;

Coolen et al., 2007; Weidler et al., 2007; Herfort et al., 2007; Chen et al., 2008; Reigstad et al., 2008; Santoro et al., 2008; Shen et al., 2008; Urakawa et al., 2008; de la Torre et al., 2008; Hansel et al., 2008; Hatzenpichler et al., 2008; Herrmann et al., 2008; de Vet et al., 2009; Sauder et al., 2011) and even outnumber AOB in certain habitats (Wuchter, 2004; Beman and Francis, 2006; Leininger et al., 2006; Park et al., 2006; Nakagawa et al., 2007; de la Torre et al., 2008; Shen et al., 2008; Hatzenpichler et al., 2008; Reigstad et al., 2008; Erguder et al., 2009; Martens-Habbena et al., 2009; Zhang et al., 2010; Xia et al., 2011; Pratscher et al., 2011), which raises interesting questions regarding their ecophysiology and importance for the global nitrogen cycle.

Amo is also an essential enzyme for AOA and thus, it is now used as a phylogenetic marker for both AOA and AOB (Treusch et al., 2005). Recently, AOA were re-assigned to a newly proposed phylum called the *Thaumarchaeota* (Brochier-Armanet et al., 2008; Spang et al., 2010). A widely accepted assumption is that all *Thaumarchaeota* which carry *amoA* are able of autotrophic nitrification (Mussmann et al., 2011). However, it was postulated that maybe not all *Thaumarchaeota* possess *amoA* genes (Agogué et al, 2008, Muller et al., 2010). This hypothesis was then disproven and furthermore explained to be probably caused by a failure of quantitative PCR primers of certain amoA sequences (Konstantinos et al., 2009). Furthermore, recent findings suggest that not all *Thaumarchaetoa* perform autotrophic ammonia-oxidation. For example, they occur in amounts in several industrial WWTPs that cannot be explained by ammonia oxidation alone (Mussmann et al., 2011).

Currently, only a handful of AOA are well described, cultivated or even isolated in pure culture. In contrast, it was recently shown by 454 amplicon sequencing of archaeal *amoA* genes that in less than 1 g of certain soils up to 83 different AOA species can be found (Pester et al., 2012). Apparently, we have just started to discover their diversity and understand their role in nature.

1.3 Lipids of AOA

Mainly (hyper-)thermophilic archaea synthesize a characteristic core membrane lipid called glycerol dibiphytanyl glycerol tetraether (GDGT) (Schouten et al., 2007). About 20 different types of GDGTs have been described, encompassing crenarchaeol (Fig. 1.2), which so far has been exclusively been detected in all AOA (Damsté, 2002; Schouten et al., 2008; de la Torre et al., 2008; Pitcher et al., 2009) and has thus been used as biomarker in environmental studies of these organisms (Leininger et al., 2006; Pitcher et al., 2011a; Pitcher et al., 2011b)

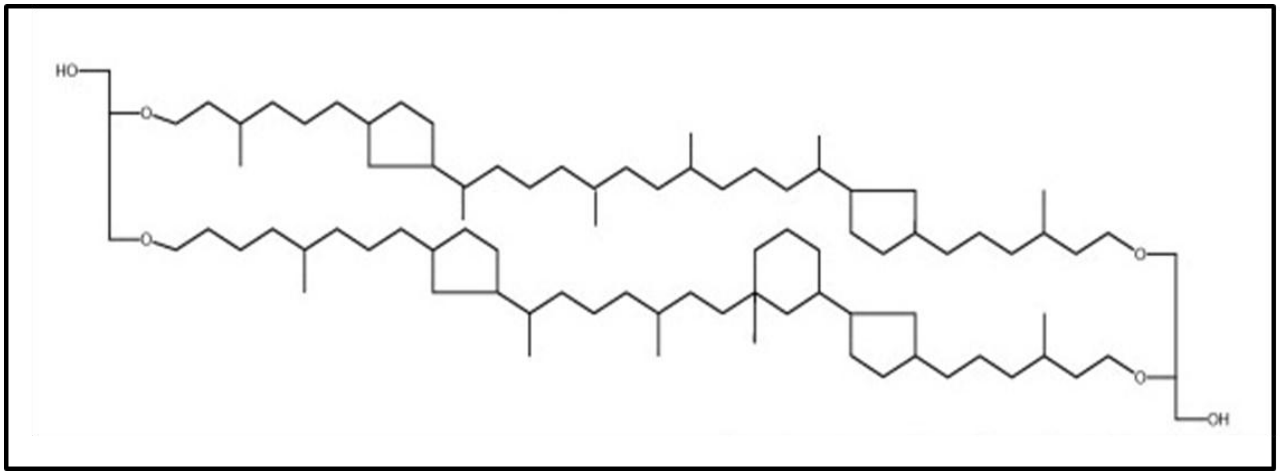

Figure 1.2. Chemical structure of crenarchaeol (Smith et al., 2012).

Crenarchaeol, which would be more accurately termed thaumarchaeol, contains four cyclopentane and one cyclohexane ring, long alkyl chains, 11 methyl - side groups and two biphytanyl glycerol diethers. Many archaea have been found in habitats with extreme conditions and the ester linkages in the membrane lipids are less stable than ether linkages, which could be an explanation for their occurrence in (hyper-)thermophilic archaea (Damsté et al., 2002).

The chemical structure of crenarchaeol is highly similar to the structure of other GDGT lipids found in many archaeal species, besides the cyclohexane ring. It was proposed that pelagic *Crenarchaeota* inherited the ability to build cyclopentanes-containing GDGT membrane lipids from hyperthermophilic *Crenarchaeota*. Nevertheless, the pelagic *Crenarchaeota* have a much cooler tempered environment. Hence, it was suggested that they modified their GDGT lipids so that they also contain a cyclohexane ring. This evolutionary step was thought to prevent a highly densed packing characteristic of the membrane lipids, which is needed for higher temperatures (Damsté et al., 2011). On the contrary, crenarchaeol was recently also found in hydrothermal marine sediments (Schouten et al., 2003) and thermal hot springs (Pearson et al., 2004). It was further postulated that besides to the water temperature, the water chemistry (e.g. salinity, pH, etc.) plays a major role in the distribution of GDGTs (Pearson et al., 2004).

During the diploma thesis I also had contact with other lipids, which do not occur in currently known AOA but do have some structural resemblances (Fig. 4.2). This membrane spanning lipid is called 13,16-Dimethyl-octacosanedioic acid or Iso-diabolic acid. It was previously thought to be restricted to certain thermophilic *Thermoanaerobacter* species (Jung et al., 1994; Balk et al., 2009). However, it was recently also found in other bacteria, especially in the subclasses 1 and 3 of *Acidobacteria*, where this lipid accounts for up to 43% of the total fatty acids (Damsté et al., 2011).

AOA are very difficult to cultivate and in many cases only an enrichment culture is available. This makes further genomic analyses very challenging and thus, new ways for the identification of AOAs are needed. Since so far all analyzed AOA contain the special membrane lipid crenarchaeol (Fig. 1.2)

this might be a putative factor to distinguish them from other microorganisms. Thus, Raman spectroscopy was chosen for this diploma thesis because it is able to reveal the chemical composition on a single cell level. This technique can also be applied to living cells which makes further analyses possible like single cell genomics or cultivation.

1.4 Raman microspectroscopy

Raman microspectroscopy is based on molecular vibrations which derive from an inelastic light scattering process ("Raman effect"). The existence of inelastic light scattering was first postulated by the Austrian physicist A. Smekal in 1923 and five years later observed from two indian physicists (Raman and Krishnan, 1928). In 1930, C.V. Raman proved the inelastic light scattering process, which was then named after him (Fechner, 2005). Raman spectroscopy measures the intensities of wavelengths of inelastically scattered light.

During a modern Raman measurement a laser beam is focused onto a sample and the photons interact with the atoms/molecules in different ways. Photons can be either absorbed or scattered. Absorption occurs most likely if the wavelength of the radiation is in the infrared (IR) or in the ultraviolet (UV). The IR absorption leads to an excitation of vibrational modes of the molecules, whereas the UV absorption leads to an excitation of an electronic transition, which is often followed by fluorescence. However, scattering is a little bit more complex. When monochromatic light is directed on a microbiological sample, radiation will pass through the obstacle (transmission). Nevertheless, a small amount of radiation will be scattered from the molecules. First of all, there is the elastic scattering process, which means that there is no measureable loss of energy (Rayleigh scattering). Second of all, approximately 1 out of 10^6 to 10^8 photons will be scattered inelastically, which means that the radiation is scattered at optical frequencies different from the frequency of the incident photons (Raman scattering) (Schrader, 1995; Petry et al., 2003).

This process is called the "Raman effect". Raman scattering can be differentiated between the so called Stokes and Anti-Stokes scattering. The characteristic wavelengths of a Raman spectrum describe the wavelength/frequency shift (Raman shift) of the Stokes and Anti-Stokes scattering in relation to the Rayleigh scattering (Bugay and Findlay, 1999; Schittkowski and Brüggemann, 2002). These shifts are characteristic for every molecule in a sample.

The incident photons of the Stokes scattering lose energy on the vibrational level (excited state). On the contrary, the incident photons of the Anti-Stokes scattering gain energy on the vibrational level (Fig. 1.3) (Popp and Kiefer, 2006). When working with biological samples, often fluorescence can be observed (Fig. 1.3 E). The excitation of fluorescence is sometimes magnitudes bigger than the actual Raman cell spectrum, which can make reasonable Raman spectrum acquisition very challenging or sometimes nearly impossible. Additionally, biological samples often contain fluorophores which

fluoresce when the wavelength of the used Raman laser is in the UV range (Petry et al., 2003). This drawback can be avoided by using a laser with a different wavelength or by photo-bleaching of the sample (Ivleva et al., 2004).

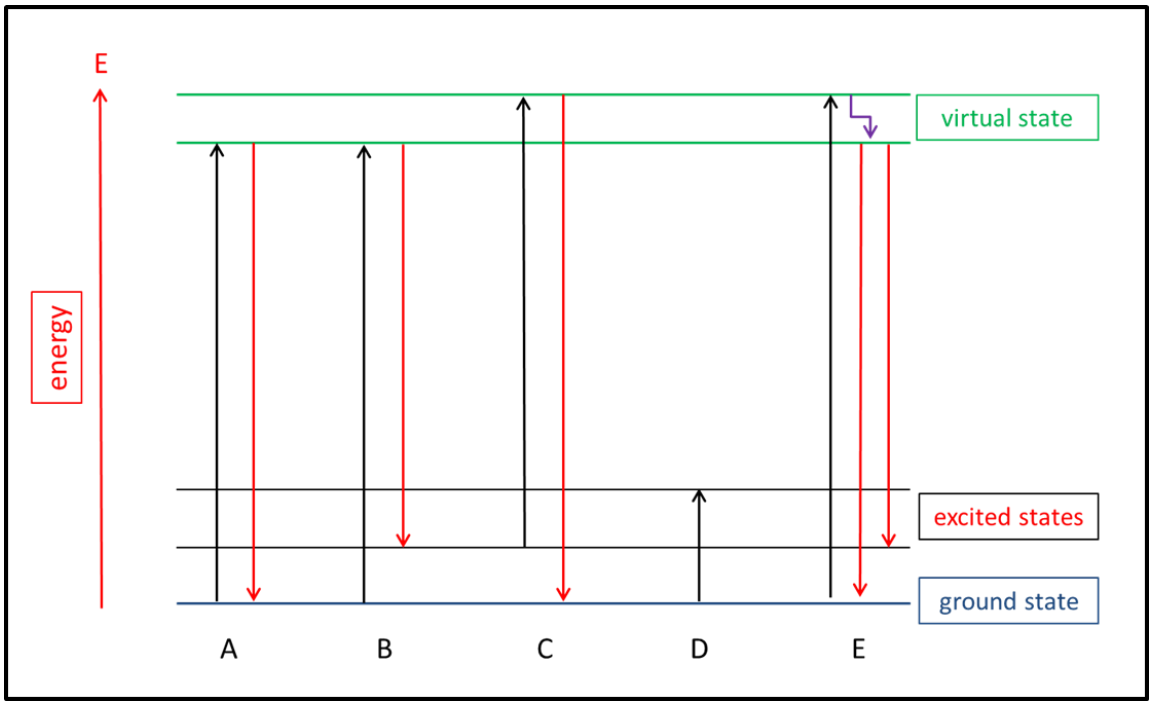

Figure 1.3. Schematic illustration of the differences between Rayleigh (elastic) scattering (A), Raman (Stokes) scattering (B), Raman (anti-Stokes) scattering (C), absorption (D) and fluorescence (E).

Raman spectra of prokaryotic microorganisms provide information about major cellular compounds (e.g. storage compounds, carbohydrates, nucleic acids, proteins and lipids (Fig. 1.4)) and the intra- and intermolecular interactions at a single cell resolution (Petry et al., 2003). Raman microspectroscopy only requires a very small amount of sample, hardly any preparations and the technique is in general non-destructive. In addition, spectra can be acquired from living or fixed samples for qualitative and quantitative analysis of their chemical composition and Raman microspectroscopy can be combined with fluorescence *in situ* hybridization (FISH) (Huang et al., 2007, Wagner et al., 2009) and single cell stable isotope probing (Huang et al., 2007; Haider et al., 2010).

During the last decade, different kinds of lasers have been used for Raman microspectroscopy. Argon and krypton lasers were applied in many laboratories, whereas currently also helium-neon (He-Ne) and neodymium-doped yttrium aluminum garnet (Nd:YAG) lasers are in use, which have a wavenumber of up to 1064 nm (Petry et al., 2003). Those near-infrared lasers pushed the application of Raman microspectroscopy in the biological sciences since they often avoid the excitation of fluorescence or even open new opportunities for optical trapping of cells (Barbarossa et al., 1991; Xie et al., 2002; Creely et al., 2005; Min et al., 2005; Huang et al., 2009).

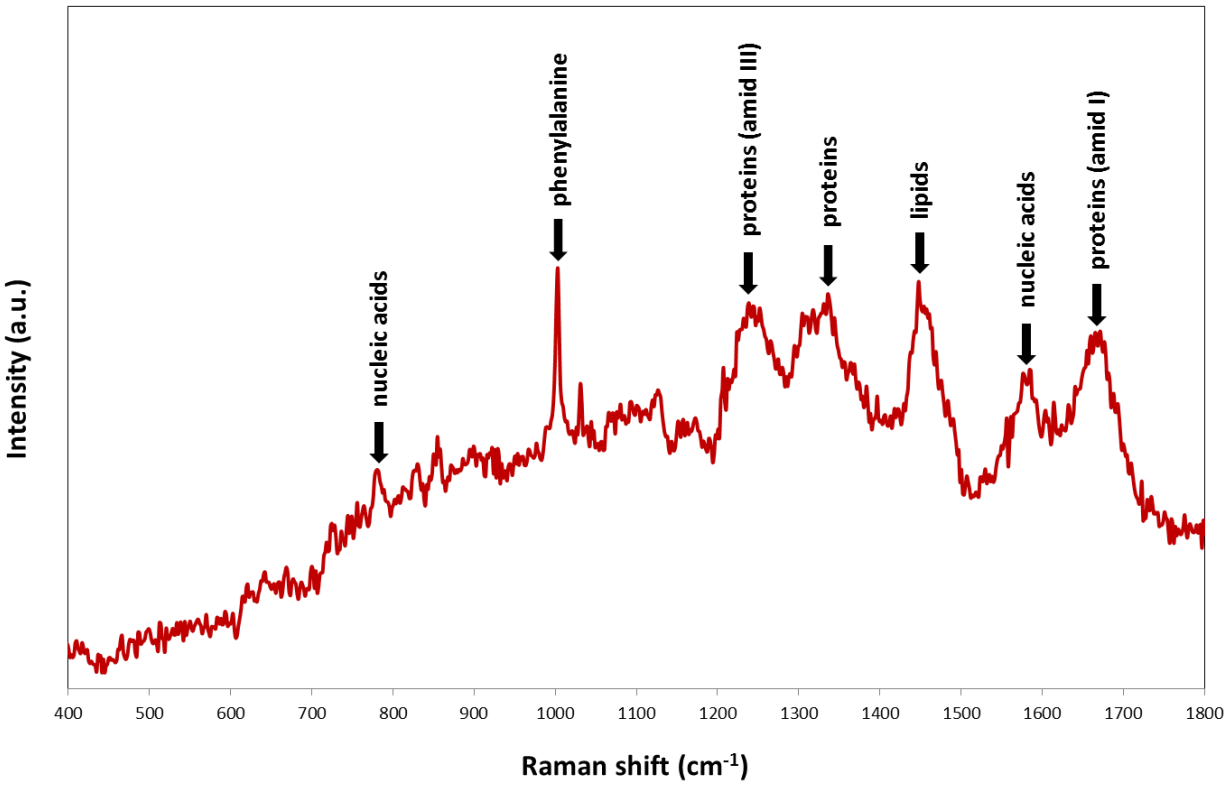

Figure 1.4. Illustration of the complexity of a Raman cell spectrum (*E. coli*). Indicated are certain peaks and bands which originate from major cell macromolecules. The spectrum was not processed in any way.

A Raman cell spectrum is a very complex fingerprint, which can consist of thousands of single peaks that merge into broad peaks, shoulders and bands (Fig. 1.4) and in this diploma thesis, a huge amount of cell spectra had to be compared under each other. Furthermore, peaks of specific compounds like lipids of AOA can potentially be very small and thus, be overseen by the human eye. Therefore, a more sophisticated high-throughput approach was mandatory – a cluster analysis in the case of this study.

1.5 Cluster analysis

Cluster analysis is a method applied to large data sets to find meaningful groups and similarities and it is a very important process in data mining. These similar groups are called clusters and they are based on specific features, so that the given data points of one cluster are more similar to each other than to data points of other clusters (Jiang et al., 2004). There are many different algorithms to achieve a clustering and they vary in how they join objects together into groups, using the measurement of similarity or distance.

In this study a hierarchical clustering was performed on all acquired Raman cell spectra of different prokaryotic species. They were assigned into clusters and sub clusters, which derived from so called distances to each other. In general, clusters contain objects with a lower distance (higher similarity) to

each other than to other objects. Basically, there are two main types of cluster methods that rely either on divisive or agglomerative clustering (Fig. 1.4).

Divisive or top-down clustering is a variant of hierarchical clustering where it starts at the top with all objects in one cluster. This cluster is then split into groups using a clustering algorithm. In contrast, agglomerative or bottom-up clustering starts at the bottom with all objects being a cluster by itself. They merge into one single cluster that contains all the objects by the use of cluster algorithms (Fraley, 1998). In this study an agglomerative clustering algorithm was used.

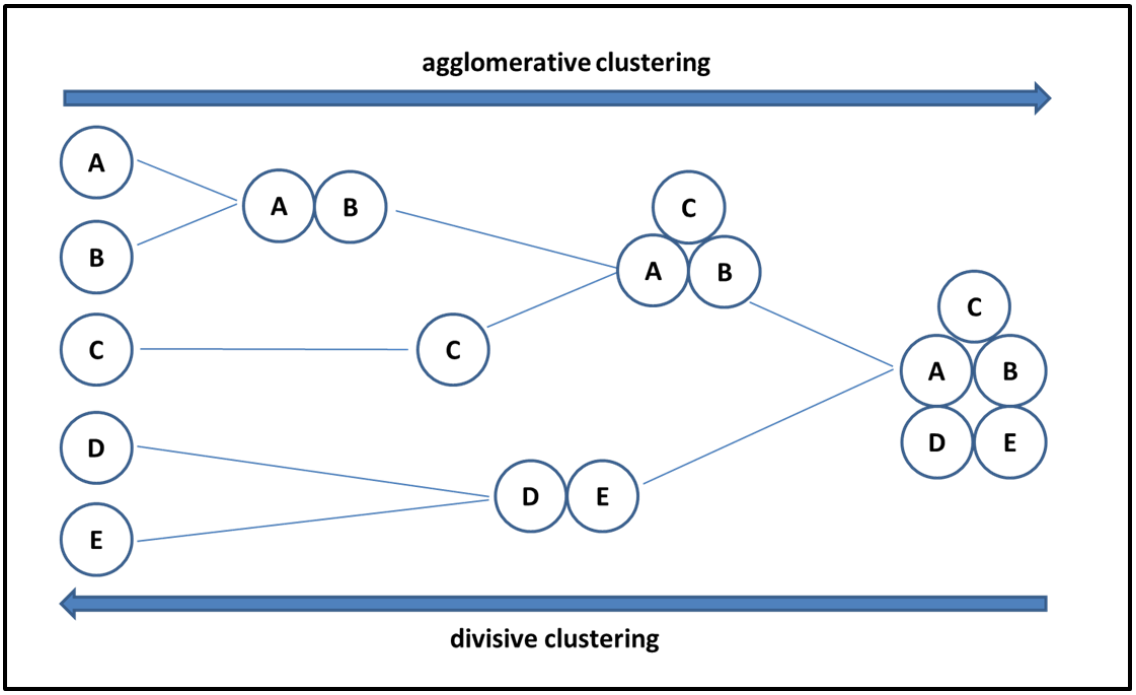

Figure 1.4. Schematic illustration of the differences between agglomerative (bottom-up) and divisive (top-down) clustering.

1.6 Aims of this study

It was the major aim of this study to explore if AOA can be reliably identified via Raman microspectroscopy. For this purpose, a comprehensive Raman reference spectra library of several AOA as well as of many other prokaryotic organisms, representing various phyla, was established and analyzed. In addition, also the spectra of various chemical compounds including lipids were recorded in order to make an attempt to understand which cellular compounds contributed in which manner to the obtained AOA spectra.

Ultimately, in collaboration with Dr. David Berry, a Post-Doc at the Department of Microbial Ecology (Vienna), a statistical software application should be established that allows to calculate the probabilities of a Raman spectrum of unknown origin to be an AOA species or not. This approach should prove itself by testing it first on AOA enrichment cultures and then on a more complex level

on environmental samples which contain putative AOA.

Ultimately, my work should provide a basis for combining Raman based single cell identification of AOA with an optical tweezer system (König, 2000; Creely et al., 2005; Huang et al., 2009), which optically traps living cells in a capillary for a quick Raman spectrum acquisition. If a trapped single cell would then have a highly calculated probability to be an AOA, this (still living) cell could then be specifically separated from the sample and further analysed by either multiple displacement amplification (MDA) and genome sequencing or cultivation.

2 Materials and Methods

2.1 Software

Table 2.1. Software used.

Name of software	Reference/Manufacturer
Adobe Photoshop CS4	Adobe Systems, San Jose, California, USA
irAnalyze	LabCognition, Analytical Software GmbH & Co. KG, Köln, Germany
LabSpec 5	Horiba, Kyoto, Japan
LSM 510 Meta V.3.2. sp2	Carl Zeiss MicroImaging GmbH, Jena, Germany
LSM Image Browser	Carl Zeiss MicroImaging GmbH, Jena, Germany
Microsoft Office 2010	Microsoft Corporation, Redmond, WA, USA
probeBase	Loy et al., 2003
R	R Development Core Team (2012). R Foundation for Statistical Computing, Vienna, Austria
Zotero	Center for History and New Media, George Mason University, USA

2.2 Technical equipment

Table 2.2. Technical equipment used.

Equipment	Company
CCD camera BX41	Olympus Corporation, Tokio, Japan
Centrifuge Mikro 22 R	Andreas Hettich GmbH & Co KG, Tuttlingen, Germany
Centrifuge Rotina 35 S	Andreas Hettich GmbH & Co KG, Tuttlingen, Germany
Hybridization oven UE-500	Memmert GmbH, Schwabach, Germany
Microscope LSM 510 Meta (CLSM)	Carl Zeiss MicroImaging GmbH, Jena, Germany

pH-meter inoLab pH Level 1	Wissenschaftlich-Technische Werkstätten GmbH & Co KG, Weilheim, Germany
Raman Spectrometer HR800	Horiba, Kyoto, Japan
Vortex Genie	Vortex Genie 2, Scientific Industries, New York, USA
Water bath DC10	Thermo Haake, Karlsruhe, Germany
Water purification system Milli-Q	Biocel, Millipore GmbH, Vienna, Austria

2.3 Expendable items

Table 2.3. Expendable items used.

Expandable item	Company
CaF_2 carrier slide („Raman slide“)	Crystran, Poole, UK
Cover slips (24×50 mm)	Paul Marienfeld, Bad Mergentheim, Germany
Erlenmeyer-Kolben DURAN®, various sizes	Schott Glas, Mainz, Germany
Eppendorf Reaktionsgefäße, various sizes	Eppendorf AG, Hamburg, Germany
Pipette tips, various volumes	Carl Roth GmbH & Co, Karlsruhe, Germany
Sampling vessels (50 ml)	Greiner Bio-One GmbH, Frickenhausen, Germany
Slides (10 wells)	Paul Marienfeld, Bad Mergentheim, Germany
Syringes, various sizes	B. Braun Melsungen AG, Melsungen, Germany
Syringe filters, various sizes	Satorious AG, Goettingen, Germany

2.4 Chemicals

Table 2.4. Chemicals used.

Chemical	Company
Blocking reagent	Roche Diagnostics Vienna GmbH, Vienna, Austria
Casein-peptone	Fluka Chemie AG, Buchs, Switzerland
Citifluor AF1	Agar Scientific Ltd., Stansted, UK
DAPI	Lactan Chemikalien und Laborgeräte GmbH, Graz, Austria
Dextran-sulfate	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
EDTA	Carl Roth GmbH & Co KG, Karlsruhe, Germany
Ethanol absolute	Fluka Chemie AG, Buchs, Switzerland
Formaldehyde (37 % (w/w))	Carl Roth GmbH & Co KG, Karlsruhe, Germany
Formamide	Fluka Chemie AG, Buchs, Switzerland
Glucose	Fluka Chemie AG, Buchs, Switzerland
Hydrochloric acid (37% (w/w)	Carl Roth GmbH & Co KG, Karlsruhe, Germany
Hydrogen peroxide (30%)	Carl Roth GmbH & Co KG, Karlsruhe, Germany
Paraformaldehyde	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Potassium acetate	J. T. Baker, Deventer, Holland
Sodium chloride	Carl Roth GmbH & Co KG, Karlsruhe, Germany
Sodium dodecyl sulfate	Carl Roth GmbH & Co KG, Karlsruhe, Germany
Sodium hydroxide	Carl Roth GmbH & Co KG, Karlsruhe, Germany
Tris	Carl Roth GmbH & Co KG, Karlsruhe, Germany
Yeast extract	Oxoid Ltd., Hampshire, England

2.5 List of all 16S rRNA probes used in this study

For the selection of the appropriate probe (Tab. 2.5), the online database probeBase (Loy et al., 2003) was used.

Table 2.5. Characteristics and specifications of all 16S rRNA probes used.

primer name	Sequence 5'-3'	specific to	FA conc. (%)	reference
Arch915	GTG CTC CCC CGC CAA TTC CT	Archaea	not determined	Stahl and Amann., 1991
CREN512	CGG CGG CTG ACA CCA G	most *Crenarchaeota*	0	Jurgens et al., 2000
EUB338	GCT GCC TCC CGT AGG AGT	most Bacteria	0 – 50	Amann et al., 1990
EUB338 II	GCA GCC ACC CGT AGG TGT	Planctomycetales	0 – 50	Daims et al., 1999
EUB338 III	GCT GCC ACC CGT AGG TGT	Verrucomicrobiales	0 - 50	Daims et al., 1999

2.6 Buffers, media and solutions

2.6.1 General buffers

PBS stock solution

solution 1 NaH_2PO_4: 200mM (35.6 g/l)

solution 2 Na_2HPO_4: 200 mM (27.6 g/l)

(pH of solution 1 adjusted the with solution 2 to 7,2 - 7,4)

1xPBS solution

NaCl 130 mM (7.6 g/l)

PBS stock solution 10 mM (50 ml/l)

MQ ad 1000 ml

(pH adjusted to 7.2 – 7.4)

3xPBS solution

NaCl	390 mM (22.8 g/l)
PBS stock solution	30 mM (150 ml/l)
MQ	ad 1000 ml

(pH adjusted to 7.2 – 7.4)

TE buffer

Tris	10 mM
EDTA	5 mM

(pH adjusted to 8.0 using HCl)

2.6.2 General solutions

0.5 M EDTA, pH (8.0)

186.1 g of Na_2-EDTA x $2H_2O$ were dissolved in 700 ml MQ. In addition, the pH was adjusted to 8.0 by 10 M NaOH. At last, the volume was filled up to 1000 ml with MQ.

1 M Tris/HCl, pH (8.0)

121 g of Tris were dissolved in 800 ml of MQ and the pH was adjusted to 8.0 with HCl (conc.).

Lysozyme solution (10 mg/mL)

Lysozyme	100 mg
0.5 M EDTA	1 ml
1 M Tris/HCl	1 ml
MQ	8 ml

2.6.3 Culture medium for *Edaphobacter modestus* (DSM 18101)

Casein peptone	0.50 g
Glucose	0.10 g
Yeast extract	0.25 g
MQ	ad 1000 ml

(pH adjusted to 5.0)

2.6.4 Paraformaldehyde solution

4 % Paraformaldehyde (PFA) solution

3xPBS solution

1 M NaOH

1 M HCl

Procedure: 33 ml MQ were heated at 65°C. 2 g PFA and 1 M NaOH were added until the solution was clear. Further, 16.6 ml 3xPBS were added. The solution was then cooled down to room temperature and the pH was adjusted to 7.2 - 7.4 by 1 M HCl. Last but not least, the solution was filter-sterilized (0.22 µm) and stored at -20°C.

2.6.5 Fluorescence *in situ* hybridization buffers

Table. 2.6 Hybridization buffer of FISH.

FA conc. (%)	0	5	10	20	25	30	35
5M NaCl (µl)	180	180	180	180	180	180	180
1M Tris/HCl (µl)	20	20	20	20	20	20	20
MQ (µl)	800	750	700	600	550	500	450
Formamide (µl)	0	50	100	200	250	300	350
10% SDS (µl)	1	1	1	1	1	1	1

Table. 2.7 Washing buffer of FISH.

FA conc. (%)	0	5	10	20	25	30	35
5M NaCl (ml)	9	6.3	4.5	2.15	1.49	1.02	0.7
1M Tris/HCl (ml)	1	1	1	1	1	1	1
MQ (ml)	ad 50ml	ad 50ml	ad 50ml	ad 50ml	ad 50ml	ad 50ml	ad 50ml
EDTA (ml)	0	0	0	0.5	0.5	0.5	0.5

2.6.6 Catalyzed reporter deposition fluorescence *in situ* hybridization buffers

Hybridization buffer (HB) – 20 ml

5M NaCl	3.6 ml
1M Tris-HCl	400 µl
Dextran sulfate	2 g
Formamide (X %)	x ml
Blocking reagent (10 %)	2 ml
SDS (10 %)	20 µl
MQ	ad 20 ml

Volume of formamide in 20 ml of HB

20 % FA in HB	4 ml FA
25 % FA in HB	5 ml FA
30 % FA in HB	6 ml FA

Washing buffer (WB) – 50 ml

5M NaCl	x µl
1M Tris-HCl	1 ml
0.5 M EDTA	500 µl
SDS (10 %)	50 µl
MQ	ad 50 ml

Volume of 5 M NaCl in 50 ml of WB

20 % FA	2150 µl
25 % FA	1490 µl
30 % FA	1020 µl

2.7 List of microorganisms used for Raman spectra acquisition

Table 2.8. Alphabetical list of microorganisms used in this diploma thesis and their Raman acquisition parameters. Filter 0 = 100 %; filter 0.3 = 58 %; filter 0.6 = 28 %, filter 1= 8 % laser intensity. The Raman reference spectra library consists of all listed microorganisms (exclusive of the two AOA enrichment cultures SV8-6 and SV9-19). The used objectives (Olympus) were Mplan-achromat with a numerical aperture of 0.90.

species	strain	acquisition time (s)	pinhole (µm)	filter	objective magnification
Acetonema longum	DSM 6540	45	600	0.3	100
Acidobacterium capsulatum	DSM 11244	50	600	0.6	100
AOA enrichment SV8-6 ***	-	25	600	0.6	100
AOA enrichment SV9-19 ***	-	30	600	0.6	100
Bacillus mycoides	DSM 309	25	600	1	100
Burkholderia cepacia	DSM 7288	40	600	0.3	100
Desulfoacinum infernum	DSM 9756	20	600	0.6	100
Desulfobacca acetoxidans	DSM 11109	30	600	1	100
Desulfobacterium niacini	DSM 2650	40	600	1	100
Desulfobacula phenolica	DSM 3384	30	600	0.6	100
Desulfobulbus propionicus	DSM 2032	18	600	0.6	100
Desulfocella halophila	DSM 11763	30	600	0.6	100
Desulfofustis glycolicus	DSM 9705	35	600	0.6	100
Desulfomicrobium apsheronum	DSM 5918	33	600	1	100
Desulfomusa hansenii	DSM 12642	25	600	1	100
Desulfovibrio halophilus	DSM 5663	45	600	0.6	100
Desulfovibrio longus	DSM 6739	35	600	0.6	100
Desulfovibrio piger	DSM 749	30	600	0.6	100
Desulovibrio oxyclinae	DSM 11498	60	600	0.6	100
Edaphobacter aggregans	DSM 19364	35	600	0.3	100
Edaphobacter modestus	DSM 18101	32	600	0.3	100
Escherichia coli	DSM 30083	20	500	0	50
Fervidobacterium pennivorans	DSM 9078	40	600	0.6	100

Gemmata obscuriglobus	DSM 5831	40	600	0.6	100
Candidatus *Kuenenia stuttgartiensis* *	-	40	600	0.6	100
Methanothermobacter margburgensis **	-	45	600	1	100
Methylobacter tundripaludum sp. nov.	DSM 17260	35	600	0.6	100
Methylocystis rosea sp. nov.	DSM 17261	30	600	0.6	100
Nitrolancetus hollandicus	-	27	600	1	100
Nitrosominus uzonensis	-	60	600	0	100
*Nitrosopumilus maritimus****	-	120	500	0	100
Nitrososphaera gargensis	-	60	600	0.3	100
*Nitrososphaera viennensis****	-	80	600	0.6	100
Nitrospira moscovienses	-	25	600	0.6	100
Rhodopirellula baltica	DSM 10527	40	600	0.6	100
Sarcina ventriculi	DSM 3758	35	600	0.6	100
Sphaerobacter thermophilus	DSM 20745	34	600	1	100
Sporotomaculum syntrophicum	DSM 14795	30	600	0.6	100
Streptococcus salivarius	DSM 20560	10	500	0	100
Sulfolobus acidocaldarius ***	DSM 639	25	600	0.6	100
Sulfolobus islandicus ***	Y.N.15.51	30	600	1	100
Sulfolobus tokodaii ***	strain 7	25	600	0.6	100
Thermosipho africanus	DSM 5309	50	600	0.6	100
Thermotoga maritima	DSM 3109	40	600	0.6	100

* donated from Dr. Markus Schmid, Lab-scale reactor, Nijmegen, The Nederlands, Radboud University

** donated from Dr. Rudolf Thauer (see description Schmid et al., 2000)

*** donated from Dr. Christa Schleper, Vienna, Austria, University Vienna

2.8 List of compounds/materials used for Raman spectra acquisition

Table 2.9. Alphabetical list of compounds/materials used in this diploma thesis and their Raman acquisition parameters. The used objectives (Olympus) were Mplan-achromat with a numerical aperture of 0.90.

compound/material	origin	acquisition time (sec)	pinhole (µm)	filter	objective magnification
1,2-di-O-phytanyl-*sn*-glycerol	INstruchemie BV, Delfzijl, The Nederlands	10	600	0.6	100
1,2-di-O-phytanyl-*sn*-glycero-3-phosphoethanolamine	INstruchemie BV, Delfzijl, The Nederlands	10	600	0.6	100
calciumdifluoride	Crystran, Poole, UK	various	600	0	100
crenarchaeol	provided by Jaap. S. Damsté	20	600	0.6	100
cylcohexane	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	3	300	0	10
cylcohexane – cyclopentane (1:1) mix	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	5	300	0	10
cyclopentane	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	2	500	0	10
glycogen	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	10	600	0.3	100
methylcyclohexane	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	3	500	0.3	50
methylcyclopentane	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	2	500	0	10

2.9 Arctic AOA enrichment cultures

The two arctic AOA enrichment soil samples (SV8-6 and SV9-19) were collected from Spitsbergen, in the Svalbard (an archipelago in the Artic, 78º north). They were grown in an autotrophic freshwater medium with the addition of streptomycin and ammonia. Archaeal but no bacterial *amoA* genes could be amplified by PCR. In addition, quantitative polymerase chain reaction (qPCR) based on the archaeal amoA showed a 17% (SV8-6) and a 26% (SV9-19) AOA-content in these enrichments. Furthermore, no sequences of fungi or other eukaryotes could be found by PCR. Finally, restriction fragment length polymorphism (RFLP) of the archaeal amoA gene featured a different pattern between the AOA enrichment cultures, indicating that the enriched AOA represent different species (Alves, 2011).

2.10 Cultivation of *Edaphobacter modestus*

An actively growing culture of *Edaphobacter modestus* (Koch et al., 2008), DSM No.: 18101, was ordered from DSMZ (Germany). The culture was cultivated and maintained in a special medium (chapter 2.6.3) at room temperature (RT) for two weeks before a sample was PFA-fixed (chapter 2.15.1.) for Raman spectrum acquisition.

2.11 Raman microspectroscopy

2.11.1 Raman spectrometer

Raman spectra were acquired using a Raman spectrometer (Horiba, HR 800). This spectrometer is coupled to a fluorescence microscope (Olympus, BX41) and equipped with a Nd:YAG laser emitting photons of a wavelength of 532.09 nm. To focus the laser beam onto the single cells a x100/x50 Mplan-achromat objective (Olympus) with a numerical aperture of 0.90 was used, which led to a laser spot size of approximately 800 nm. The spectral resolution was 1.5 cm^{-1}.

2.11.2 Calibration of the spectrometer

The Raman spectrometer was calibrated every day by the use of a calibration script by the Labspec 5 software (Horiba). Pure silica was used for this calibration measurement. In addition, also a laser alignment should be done at least every week to ensure that the positions where the photons are generated and where most photons are captured stay the same. Unfortunately, this was not executed during this diploma thesis which could have resulted in varying spectra intensities for specific acquisition parameters. The red diode of the CCD detector should be activated and the green dot of the Labspec 5 software must then be aligned with the center of the red diode. This green dot indicates where the optimal position of the laser should be. Finally, the laser has to be aligned to this position by the use of the internal mirrors.

2.11.3 Treatment of the samples

Approximately 4 to 10 µl of PFA-fixed sample (depending on the density of the culture) were pipetted on a CaF_2 carrier slide (Crystran). In order to immobilize the cells by drying, the slide was then put at 46°C for approximately 10 to 15 minutes, then shortly dipped into double distilled water (MQ) to remove most of the salts and last but not least air dried.

2.11.4 Raman spectra acquisition

Raman spectra with a high signal to noise ratio (SNR) were acquired using the LabSpec 5 software (Horiba) at 20 to 120 sec of acquisition time. Additionally, the LabSpec 5 controlled filter for the laser power was set from the range 0 to 1 to prevent cells from taking photodamage. Additional tests were run to evaluate the laser intensity for each filter. Filter setting 0 allowed a laser intensity of 100 %. Filter 0.3 let pass around 53 %, filter 0.6 around 28 % and filter 1 around 8 % of laser intensity. Furthermore, the pinhole size of the CCD detector was set to 600 µm because this resulted in Raman cell spectra with a higher intensity compared to a smaller pinhole adjustment. Moreover, cell spectra were recorded at the range from wavenumbers 400 cm^{-1} to 3200 cm^{-1} with the most relevant Raman peaks of microbial cells being located between 400 cm^{-1} and 1900 cm^{-1}. Hence, only spectra within this shorter range are displayed in this diploma thesis.

2.12 Processing of Raman raw data

2.12.1 Smoothing

Acquired Raman spectra were smoothed using the Labspec 5 software from Horiba. This step was done to denoise the spectra. This function is based on a so-called linear Savitsky-Golay smoothing.
"Savitsky-Golay smoothing fits a polynomial function of a specific "degree" through a range ("size") of adjacent pixels, and replaces those pixels with the polynomial curve. Typically, the smaller the "degree" and the larger the "size", the more significant the smoothing" (Labspec 5 user manual). In the case of this diploma thesis, the following parameters were executed two times: degree 2, size 3.

2.12.2 Baselining

In this study the Labspec 5 software (Horiba) was used to perform baselining. In general, this method is applied to remove fluorescence from a Raman spectrum. The baselining of Labspec 5 is based on a curve fitting approach where a polynomial curve that has the best fit to a series of data points is constructed. The chosen data points depend on the polynomial degree, 8^{th} degree in the case of this diploma thesis. The 8^{th} degree equation of a polynomial curve will exactly fit 9 data points. The Labspec 5 line-segmented baseline approach works as follows:
The program fits an 8^{th} degree polynomial regression to all spectrum data points. The points above the curve are then excluded from consideration and the program calculates the new best fitted polynomial 8^{th} degree curve for the data points that are left. This operation is repeated until there are

no more data points left to exclude. Furthermore, the program uses the non-excluded points and draws straight lines (line-segmented) between those points. The data points below these lines become subtracted from the total Raman spectrum. The general formula to calculate the baseline offsets is pictured in Figure 2.1.

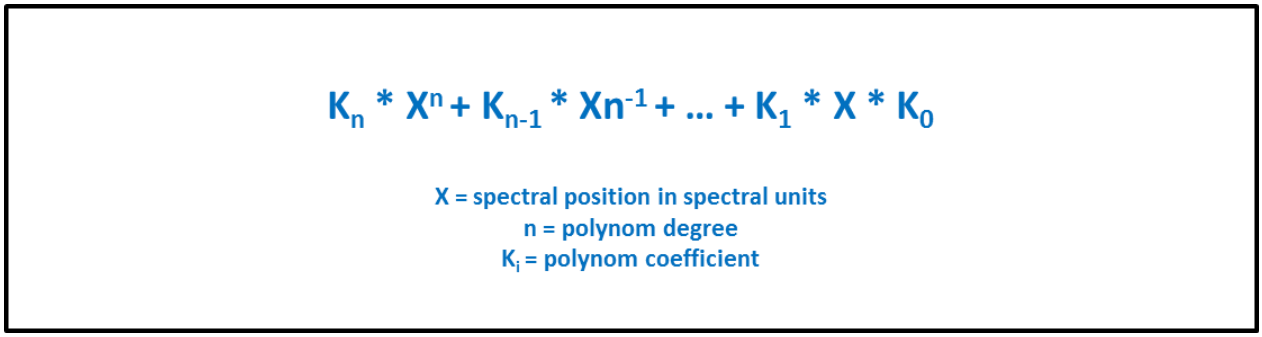

Figure 2.1. The general formula to calculate the baseline offsets.

2.12.3 Normalization

Raman spectra were normalized in order to compare them to each other. The intensities of a Raman spectrum are heavily influenced by the acquisition parameters (e.g. filter, acquisition time, objective, selected pinhole diameter). The higher the numbers of photons are (e.g. less or no filter), the more intense are the peaks of the Raman spectrum. Hence, three different methods of normalization were applied in this study:

a) Mean normalization was directly applied by the Labspec 5 software (Horiba). Every data point (of a wavenumber) was divided by the sum of all data points of each spectrum.

b) Median normalization was manually applied to the data set of the recorded microorganisms by taking the median intensity of all data points of a Raman spectrum and dividing the intensity value of every wavenumber of a spectrum by this median value.

c) Phenylalanine (Phe) normalization was also manually applied to the data set of the recorded microorganisms by taking the value of the Phe peak at the position 1004 cm^{-1} from a Raman cell spectrum and dividing every single data point on all spectra positions by this Phe value. This was executed for all microorganisms. Thus, at the end the Phe data point at the position 1004 cm^{-1} of all acquired Raman spectra had a normalized value of 1.

2.12.4 Spectra alignment based on phenylalanine

All acquired Raman cell spectra were aligned so that their most prominent Phe peak was located at the wavenumber position 1004 cm^{-1}. The proper position of the Phe peak has been confirmed in the past by many other studies (Maquelin et al., 2000; Buschman et al., 2001; Xie et al., 2002; Huang et

al., 2003; Krishna et al., 2004; Mannie et al., 2005; Schallreuter et al., 2005; Shao et al., 2005; De Gelder et al., 2007; Hu et al., 2008; Teh et al., 2009; Meyer and Smith, 2011). This step was crucial for this study because specific AOA peaks might be very small. Thus, minor shifts in the spectrum caused by an inaccurate calibration of the spectrometer could cause major problems in finding of those characteristic peaks.

2.12.5 Mean spectra

In order to work with a smaller and easier manageable number of spectra, the single cell spectra of a specific microbial strain were combined to a mean spectrum. In case that certain cell spectra of a species contained the characteristic Raman peaks of the storage compound polyhydroxybutyrate (PHB), an isolated mean spectrum was created out of them.

2.12.6 Polyhydroxybutyrate filter script

In order to improve the search for AOA specific peaks, it was necessary to subtract storage compounds from certain mean Raman cell spectra. It might have been possible to discover AOA specific peaks even though the cell spectrum was covered by storage compound peaks, but the goal was to prevent a clustering of microorganisms which was mainly based on storage compounds even though the organisms themselves where not closely related. In this diploma thesis only PHB was considered because it is a very abundant storage compound, no other ones could be clearly identified within our library and PHB does have a very complex spectrum, which took a lot of time to analyze and fully understand. A polyhydroxybutyrate (PHB) filter script was built based on a difference Raman spectrum of *Sarcina ventriculi* with and without the storage of PHB, which derived from a difference in the growth stage (Fig 3.8). The script was built and executed in R (R Development Core Team, 2011) by Dr. David Berry. Further, the processed (chapter 2.12) data set of all mean cell spectra (=Raman reference spectra library) was read in and the PHB filter script evaluated the Raman wavenumbers between 1724 and 1741 cm^{-1}, the area where PHB has its most characteristic peak. This peak was chosen for monitoring the storage of PHB because no other significantly strong cell-derived Raman peaks could be detected in this region. This was not true for the other peaks of the PHB spectrum. Furthermore, if there was a data point in this area that was as high as or even higher than the phenylalanine peak in the same spectrum, it was handled as a PHB containing cell spectrum. The script then chose the maximum value out of the data points between 1724 and 1741 cm^{-1} and aligned the difference spectrum (Fig. 3.9) with reference to this position to the height of this putative PHB peak. In addition, every Raman spectrum was manually inspected to prove the storage of PHB. The cluster dendrograms (Fig. 3.1 – 3.6) also worked as a positive control of this method because

two spectra of the same species, with subtracted and without PHB storage, should theoretically cluster together, at least for *S. ventriculi* this must be true.

2.13 Random Forest

Random Forest (RF) (Breiman, 2001) is a powerful statistical classifier which performs classification and regression based on a forest of decision trees (chapter 2.13.1) that were grown using randomly selected subsets of the input data. In this diploma thesis the RF package (Liaw and Wiener, 2002) in the open software R (R Development Core Team, 2011) was used. RF is one of the most accurate learning algorithms today (Caruana et al., 2008), and it can handle very large data sets (Breiman, 2001). Further, RF has proved to be an accurate method for various prediction questions (Chen and Liu, 2005; Díaz-Uriarte and Alvarez de Andrés, 2006; Prasad et al., 2006; Dutilh et al., 2011). Furthermore, it has its roots in two methods called classification and regression trees (CART) (Breiman, 1984) and bootstrap aggregating (bagging) (Breiman, 1996). CART is "*… a recursive partitioning method*" and it is used to build classification and regression trees (Statsoft(a)). To achieve this, raw data is used and decision trees (chapter 2.13.1) are generated by growing them to their maximum size. In addition, a certain number of potential optimal trees are produced. At the end, the best tree is evaluated by the use of the Gini diversity index (chapter 2.13.3) (Wu et al., 2007). Bagging is a very effective method to improve the predictive power (stability and classification accuracy) of a classifier. Furthermore, bagging reduces the variance of a predictor, but its success relies on the instability of the used learning algorithm (Dietterrich, 2000). Especially, for large, high dimensional data sets this approach is used quite often (Bühlmann and Yu, 2002).

2.13.1 Decision trees

A decision tree is a predictive model where distinct rules are applied to calculate a target value (Horning). Furthermore, a specific algorithm is used to determine where the best split at a node is. RF creates many of those decision trees using a randomly selected subset of spectral data at each node.
Each tree is grown using a binary partitioning that means each parent node is split into two children nodes. Furthermore, each tree is grown, at least partially, at random. Once a node is split, the process is then repeated for every following child node (Steinberg et al, 2004), which means that all the trees are grown to their absolute maximum extent possible and left unpruned (Breiman, 2001). Pruning involves editing of a tree to simplify its structure (e.g. by removing nodes) (Zhang and Shasha, 1989). After a "forest" of N decision trees is generated, all of these trees "vote" (classify) for a distinct target attribute (class) and Random Forest then chooses the classification with the most out of N votes as it works with the principle: "*The winner takes it all*" (Lüthy, 2009). In addition, RF can

inject randomness (e.g. candidate predictors at nodes are chosen partly at random; growing of trees is achieved by the selection of a random subsamples of the training data), so that each tree is different (Breiman, 2001; Steinberg et al., 2004).

2.13.2 Error rates

RF does not require the need of a special test data set to calculate its accuracy. For every tree grown, approximately one third of all cases (training data) are not in the bootstrap sample, they are simply left out (out of bag (OOB)). "*Each tree is constructed using a different bootstrap sample from the original data*" (Breiman – statistics Berkeley). Each OOB case of the individual tree construction is put down to this tree to obtain a test set classification for each case in about one third of the trees. At the end, one class achieves the most votes every time a specific case was OOB (Breiman – statistics Berkeley). The final OOB error estimate for the whole forest is calculated simply by cumulating the individual OOB results. These values are then averaged over the trees in the forest and the OOB error curve flattens out when enough trees are added to the RF. Furthermore, RF uses these OOB samples to compare the error rates of the original input data with the error rates of the variable permutated data and calculates the importance of variables (Breiman, 2001).

2.13.3 Weighting of variables

RF gives an output of the most important variables in the classification and produces an unbiased estimate of the test set prediction error (Breiman - statistics Berkeley). Those relevant predictor variables have been used quite a lot for many different data like DNA sequencing and microarrays (Lunetta et al., 2004; Arun and Langmead, 2005; Bureau et al., 2005; Huang et al., 2005; Díaz-Uriarte and Alvarez de Andrés, 2006; Qi et al., 2006; Ward et al., 2006; Statnikov et al., 2008; Wu et al., 2008; Moorthy and Mohamad, 2011). Furthermore, RF generated precise results even when the important predictor variables were correlated (Strobl et al., 2007). In addition, RF does not overfit, so it is possible to run any number of trees in a short amount of time (Breiman, 2001). Moreover, you gain information about the importance of all variables for a desired group clustering because they are weighted by the RF algorithm.

Two types of variable importance measures are offered by the program: the Gini importance and the raw importance score, also known as the permutation accuracy importance. The Gini importance is calculated by the principle of impurity reduction (Breiman – statistics Berkeley). This is based on the Gini coefficient, which is *"…a measure of inequality in a population* (frequency distribution)" (Damgaard). A low Gini coefficient value implies that values are similar, whereas a high value expresses a higher inequality (Damgaard). Taken together, the Gini importance is the sum of the

averaged impurity decreases of all nodes in a Random Forest where a specific predictor variable was chosen for splitting a node (Schwarz et al., 2010). Additionally, a split at a node is made when the Gini impurity criterion for a parent is higher than a daughter node (Izenman, 2008). When all the Gini decreases for each variable in the forest are added up, the outcome is an importance score which, often but not always, is consistent with the raw importance score (Breiman – statistics Berkeley). The idea behind the raw importance is not to measure the gain of information, but the comparison of the measurement of prediction accuracy from the model with the original data and the prediction accuracy of the permuted data. When a variable is permuted, it loses its influence to the response. If the prediction accuracy of the permuted data decreases substantially, then it is said, that this variable seems to be important. Further, this is done averaged over all trees in the forest to measure the variable importance values. The mathematics behind it is that every tree from the forest is taken, the OOB cases are put down it, and the numbers of votes for the correct class are calculated. Then the data in the OOB cases are randomly permuted and then these cases are then put down the tree. The number of votes for the correct class in the permuted OOB data set is then subtracted from the number of votes for the correct class in the original OOB data set. The average of this number over all trees in the forest is the raw importance score for a variable (Breiman, 2001, Breiman - Berkeley)). In this study only the raw importance score was used, since it is known to produce statistically more precise results (Díaz-Uriarte and Alvarez de Andrés, 2006; Strobl et al., 2007).

2.14 Clustering – R function

The data was hierarchically clustered by using the R software (R Development Core Team, 2011) and its function hclust, which is available in the standard R software without any additional packages. Furthermore, this function needs two arguments to work, a distance matrix and an algorithm method.

2.14.1 Euclidean distances

In this study, Random Forest (chapter 2.13.) calculated the weight of each peak of a Raman spectrum based on the whole data set of the microorganisms mean spectra to be important for AOA. In addition, RF produces data which have to be processed into a final graphical output file, the cluster dendrogram. To achieve this, a hierarchical clustering analysis was applied which assigned a given set of objects to groups on the basis of dissimilarity (Mimmack et al., 2001). A distance matrix was generated which derived from distances between these clustered groups. For this study, Euclidean distances (Fig. 2.2) were used. Geometrically, an Euclidean distance is the shortest distance (a straight line) between any two given points. Moreover, it is one of the most commonly used types of distances and it is simply calculated by the Phytagorean formula (Fig. 2.3).

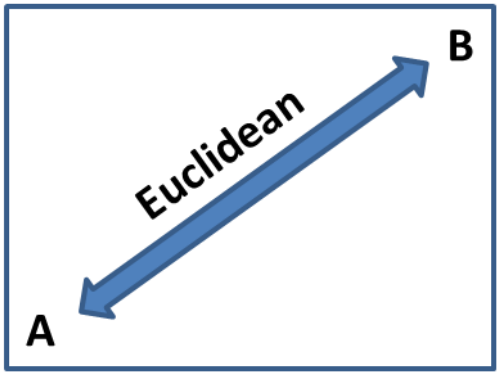

Figure 2.2. Graphical illustration of an the Euclidean distance.

$$d(x,y) = \sqrt{\sum_{i}^{n} (x_i - y_i)^2}$$

Figure 2.3. Formula of Pythagoras.

2.14.2 Ward's method

The Ward's method (Joe and Ward, 1963) was used in this study as the clustering algorithm of choice to build the cluster dendrogram. This method "*...uses an analysis of variance approach to evaluate the distances between clusters"* (Ender). Its objective is to find at each level those two distinct clusters whose merger gives the minimum sum of squares of any two clusters that can be formed at this stage (Hervada-Sala and Jarauta-Bragulat, 2004).

2.14.3 Classification weighting

How the weighting works is described in chapter 2.13.3. Two weighting patterns were used in this study, the AOA weighting (a) and the AOA+ weighting (b).

a) The AOA weighting comprises all four AOA mean spectra (*N. gargensis, N. maritimus, N. uzonensis and N. viennensis*) and evaluates which similarities they have in common against all other acquired cell spectra of the reference library. Those similar peaks receive a higher weighting compared to other peaks. The cluster dendrogram is then calculated based on these weightings.

b) The AOA+ weighting is a modified weighting pattern for Random Forest. It contains not only all four AOA species (see a), but also certain other microorganism. In short, all species from the Raman reference library which contain iso-diabolic acid (*Acidobacterium capsulatum, Edaphobacter*

aggregans, Edaphobacter modestus, Fervidobacterium pennivorans, Thermosipho africanus and *Thermotoga maritima*) and the bacterium *Desulfovibrio oxyclinae*.

2.14.4 Arctic AOA enrichment cultures

The prediction of a Raman spectrum of unknown origin to belong to the group of AOA was accomplished by using the RF package (Liaw and Wiener, 2002) of R (R Development Core Team, 2011). Every Raman spectrum of the two arctic AOA enrichments SV8-6 (Fig. 8.41 – 8.50) and SV9-19 (Fig. 8.51 – 8.70) was individually evaluated. The calculation is based on a bootstrap algorithm and it shows the probability of a spectrum to be assigned to the AOA cluster of the dendrogram (Fig. 3.1 – 3.6).

2.15 Fluorescence *in situ* hybridization

Fluorescence *in situ* hybridization (FISH) is a culture-independent method, which is used to detect specific prokaryotic microorganisms by fluorescently labeled rRNA targeted oligonucleotide probes. It allows to differentiate between different taxonomic levels and to perform a quantitative analysis. For the best results, cells have to be chemically fixed before the probes can hybridize under stringent conditions. Furthermore, FISH can be performed on a slide or in a liquid. After the hybridization, a final washing step is conducted and the cells can then be detected by an epifluorescence microscope (Daims et al., 2005). In this diploma thesis only the liquid FISH technique was used. Hence, the following steps (besides PFA fixation) will only refer to this type of hybridization.

2.15.1 Cell fixation

A tube with 1 ml of a culture was centrifuged at 14,000 rpm for 6 min. After the disposal of the supernatant the cell pellet was resuspended in 1 ml 1xPBS and again centrifuged with the same parameters. The supernatant was discarded and 250 ml of 1xPBS were added. After that, 750 ml of 4% PFA were added and the content of the tube was resuspended and put at 4°C for 3 hours. The tube was then centrifuged (14,000 rpm, 6 min) and the supernatant was disposed. 1 ml of 1xPBS solution was added, followed by the disposal of the supernatant and centrifugation (14,000 rpm, 6 min). The final step was performed by adding some PBS/EtOH (1:1 solution). The amount of solution varied due to the density of the culture. Usually the amount was between 20 and 100 µl. All PFA-fixed microbial samples were stored at – 20°C.

2.15.2 Dehydration of the fixed sample

About 6 µl of a PFA-fixed sample were pipetted in a 2 ml tube with the addition of 150 µl EtOH (96%) to dehydrate the cells. After 5 min on room temperature (RT), the tube was centrifuged (14,000 rpm, 5 min). Finally, the supernatant was discarded and the sample dried within seconds.

2.15.3 *In situ* hybridization

50 µl of the hybridization buffer (HB) (Tab. 2.6) were pipetted onto the cells and 1 µl of the specific probe was added. The tube was closed and put at 46°C for 2 hours where it was used as a hybridization chamber.

2.15.4 Washing of the hybridized sample

The tube was centrifuged at 46°C (14,000 rpm, 5 min) and subsequently, the supernatant was discarded. The pellet was then resuspended in 500 µl of washing buffer (WB) (Tab. 2.7) and put back at 46°C for 5 min. After this step the tube was centrifuged again at 46°C (14,000 rpm, 10 min) and the supernatant was discarded. Subsequently, the sample was resuspended in 50 µl ice-cold MQ and again centrifuged (14,000 rpm, 10 min, 4°C). Finally, the sample was resuspended in approximately 30 µl of PBS:EtOH (1:1) and stored at -20°C for further usage (e.g. analysis under the fluorescence microscope, Raman spectrum acquisition).

2.16 Catalyzed reporter deposition fluorescence *in situ* hybridization

Most known AOA cannot be successfully detected with conventional FISH using mono-labeled oligonucleotide probes (chapter 2.15). In addition, if a cell has a low ribosomal content or a high autofluorescence, there is a need for a technique which circumvents these drawbacks. The catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) works with a tyramide signal amplification. The sample is hybridized with a horseradish peroxidase conjugated probe, and these probes are then detected by the addition of fluorophore labeled tyramides. The sample can again be analyzed by epifluorescence microscopy and the signal intensity is much stronger compared to the standard FISH (Hoshino et al., 2008). The CARD-FISH protocol (Pernthaler et al., 2002) of this study was adapted and performed on slides.

2.16.1 Cell fixation

The cell fixation for the CARD FISH protocol was performed similar to the standard FISH approach (chapter 2.15.1.).

2.16.2 Embedding

10 µl of the PFA-fixed sample were pipetted onto the wells of the glass slide and dried in an oven at 46°C for about 10 min. After this step, an increasing EtOH series was performed. The slide was put into 50%, 80% and 96% EtOH for 3 min each. After this dehydration step, the cells on the slide wells were covered by 10 µl of 0.1% agarose. The slide was air dried at 30°C for about 10 min.

2.16.3 Permeabilization of the cell wall

15 µl proteinase K (15 µl/ml) were pipetted on each well of the slide and incubated for 10 min at RT. Further, the slide was put into a tube with MQ and incubated for 1 min at RT. The slide was then incubated in a tube with 0.01 M HCl for 10 min at RT to bleach endogenous peroxidases and to inactivate proteinase K, followed by 30 min incubation in methanol + 0.15 % H_2O_2. Finally, the slide was washed in MQ, dipped in 96 % EtOH and air dried.

2.16.4 *In situ* hybridization and washing

The HB (chapter 2.6.6) and the probe working solution (50 ng/µl) were mixed in a 300:1 ratio. Further, 10 µl of this mixture were pipetted onto the wells of the slide with the samples. The slide was then put into a tube (hybridization chamber) with 50 ml volume in which a tissue was soaked with the remaining HB. The tube was closed and put in an oven at 46°C for 2 hours. During this step, the WB (chapter 2.6.6) was pre-warmed at 48°C in a water bath. After 2 hours, the slide was removed of the hybridization chamber und put into a 50 ml tube with washing buffer for 10 min at 48°C. Finally, the slide was briefly dipped into ice-cold MQ.

2.16.5 Tyramide signal amplification

The slide was incubated in a tube with 1xPBS at RT for 10 min. After this step, the liquid was removed by blotting the slide on a piece of paper. The slide was then incubated in a humid chamber in a substrate mix (1 part (1.8 µl) of dye-labeled tyramide (1:10) and 100 parts (standard 180 µl) of amplification buffer + 0.0015 % H_2O_2 end concentration (1.8 µl; freshly prepared; 1 ml MQ + 5 µl

H_2O_2 (30 %)) for 45 min at 45°C in the dark. Subsequently, the slide was washed in a tube with 1xPBS at RT for 10 min in the dark and then in MQ for 1 min in the dark. After a final air drying step in the dark the slide was ready for storage at -20°C or for the analysis under the fluorescence microscope. For this purpose the wells of the slide were mounted with Citifluor and a cover slip was put on it.

3 Results

3.1 Cluster dendrogram of the Raman library reference microorganisms

The following cluster dendrograms are based on mean spectra of the Raman reference library microorganisms (Tab. 2.8 – exclusive the two artic AOA enrichment cultures). The Raman cell spectra were processed as follows: They were aligned to the phenylalanine peak at wavenumber 1004 cm^{-1}. Furthermore, they were smoothed, baselined (line-segmented 8^{th} degree) and normalized (chapter 2.12). Three different methods of normalization were applied: Normalization based on the phenylalanine peak at position 1004 cm^{-1} (chapter 3.1.1). Moreover, median (chapter 3.1.2) and mean normalization (chapter 3.1.3) approaches were applied.

Random Forest (chapter 2.13) was used to weight the variables of the data set. Two different kinds of weighting types were performed. Weighting towards only the AOA species (chapter 2.14.3a) and weighting towards the AOA+ group (chapter 2.14.3b). The RF weighting calculation was based on 10,000 trees and 31 variables tried at each split. The OOB error estimate was calculated for all cluster dendrograms individually.

3.1.1 Phenylalanine normalized data set

The two cluster dendrograms based on the phenylalanine normalized (chapter 2.12.2.c) Raman reference data set are shown in Figure 3.1 (AOA weighted) and Figure 3.2 (AOA+ weighted). Figure 3.1 featured two main clusters. Cluster 1 contained all AOA species (red), all iso-diabolic acid containing species (green) and five mean spectra of microorganisms, which were not expected to cluster within the AOA group (indicated by a black arrow). Cluster 1 also consisted of two main sub-clusters, A and B. All four AOA mean spectra were assigned within the sub cluster B. Cluster 2 (purple) contained all *Sulfolobus* species (blue). The calculated OOB error estimate was 12.5 %. This value was the highest error rate of all dendrograms.

Figure 3.2 was based on the AOA+ RF weighting and showed a similar, but slightly improved result. Again, all four AOA species (red) were assigned to the same sub cluster (B), but much closer. In addition, only two iso-diabolic acid containing organisms were left inside the sub cluster B, plus three unexpected species (*D. oxyclinae, M. tundripaludum and M. rosea*). *D. oxyclinae* clustered together with *N. viennensis* in both RF weighting methods (Fig. 3.1 and 3.2). The OOB error estimate of Figure 3.2 was again 12.5 %.

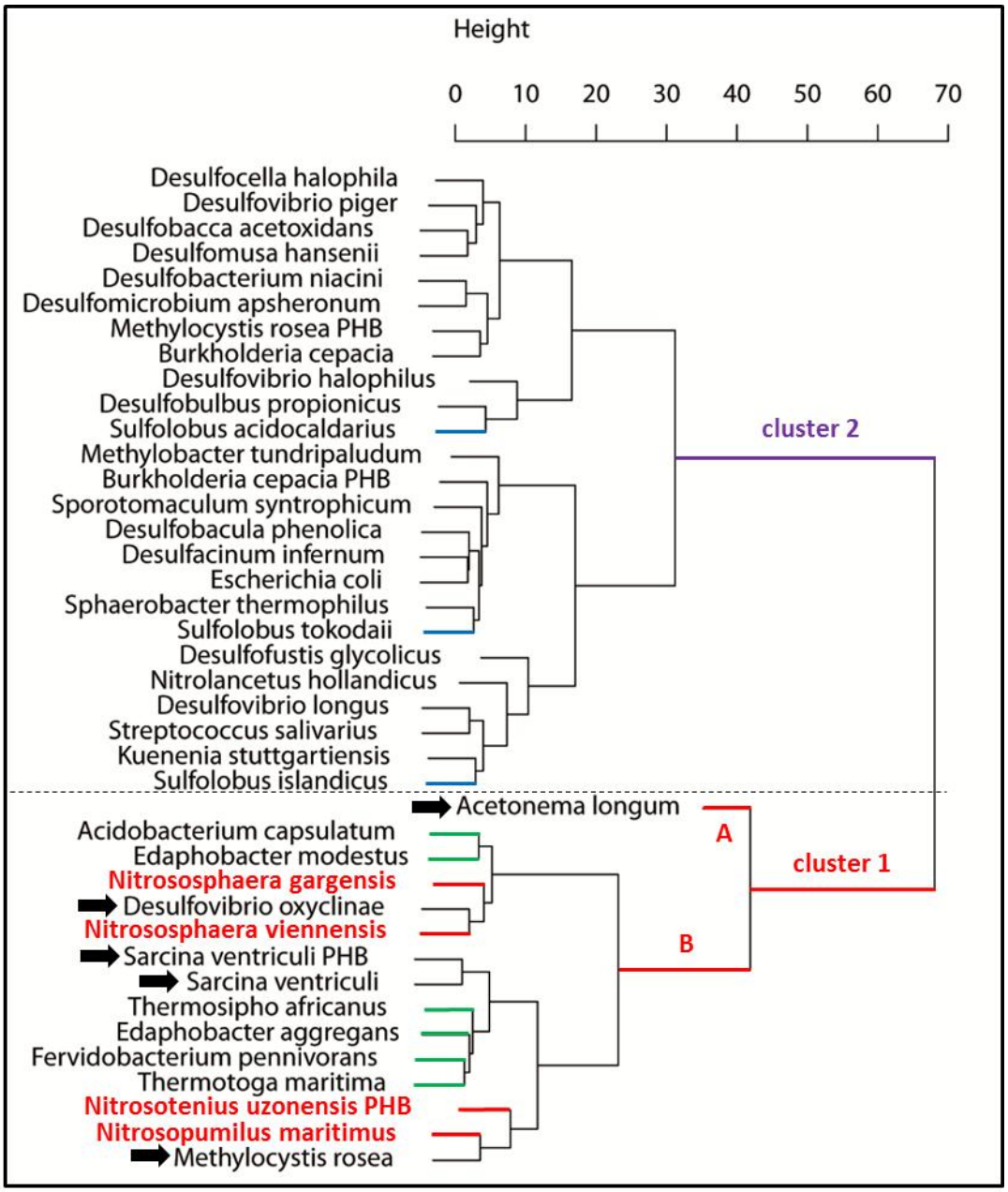

Figure 3.1. Cluster dendrogram based on the Raman reference library. Cluster 1 contains the AOA (red). In addition, the sub clusters A and B of cluster 1 are indicated in the figure. Cluster 2 (violet) does not contain AOA species. *Sulfolobus* species are indicated by a blue labeling. The classification was based on a Random Forest weighting towards only the AOA mean spectra. PHB next to a species name indicates that these Raman mean spectra contained signals from the storage compound polyhydroxybutyrate, which was subtracted by the PHB filter script (chapter 2.12.6). The data were mean spectra and processed as follows: phenylalanine-shifted to position 1004 cm^{-1}, line-segmented 8^{th} degree baselining and phenylalanine normalized (chapter 2.12).

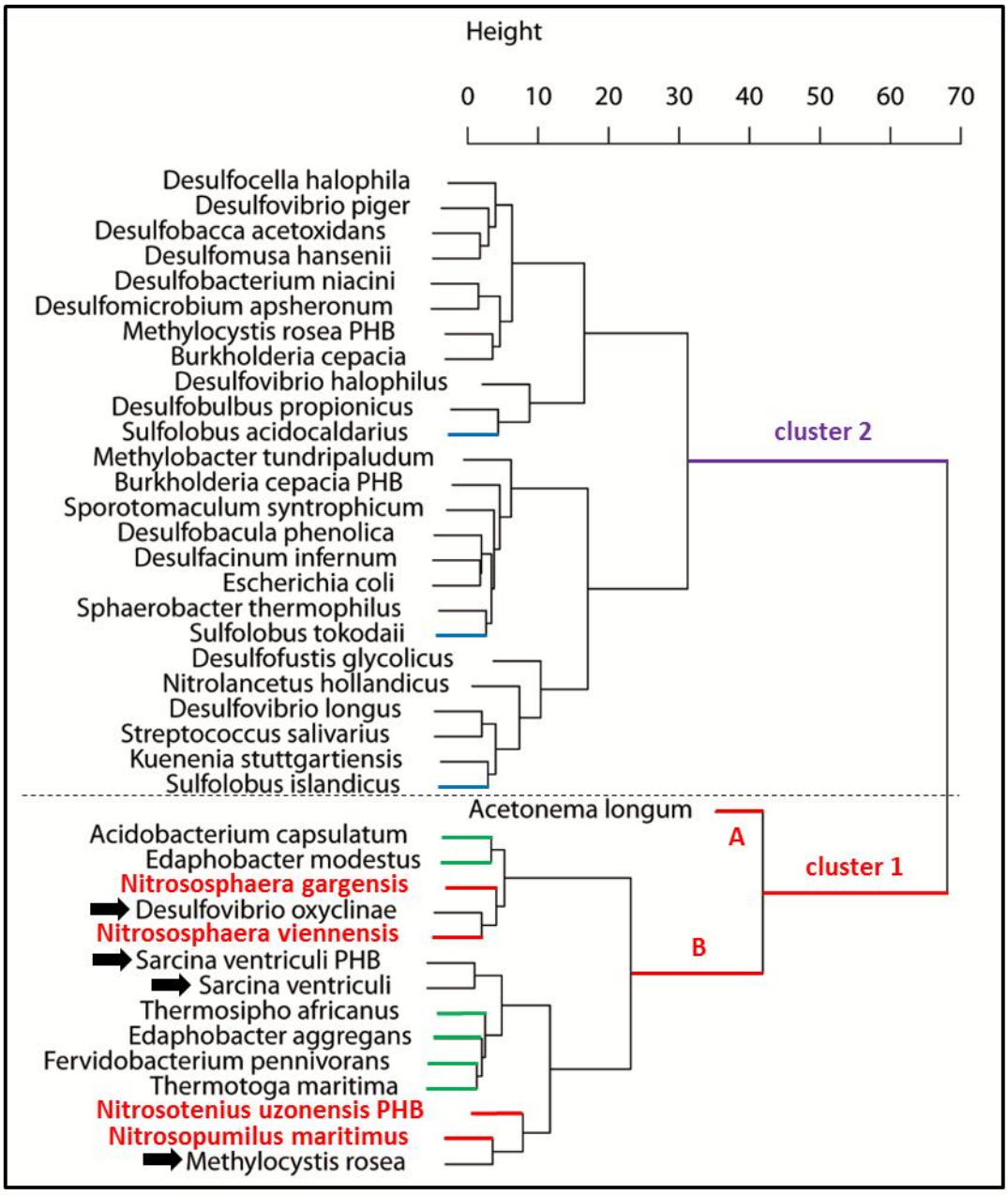

Figure 3.2. Cluster dendrogram based on the Raman reference library. Cluster 1 contains the AOA (red). In addition, the sub clusters A and B of cluster 1 are indicated in the figure. Cluster 2 (violet) does not contain AOA. *Sulfolobus* species are indicated by a blue label. The classification was based on a Random Forest weighting towards the AOA+ mean spectra (chapter 2.14.3). PHB next to a species name indicates that these Raman mean spectra contained signals from the storage compound polyhydroxybutyrate, which was subtracted by the PHB filter script (chapter 2.12.6). The data were mean spectra and processed as follows: phenylalanine-shifted to position 1004 cm^{-1}, line-segmented 8^{th} degree baselining and phenylalanine normalized (chapter 2.12).

3.1.2 Median normalized data set

The two cluster dendrograms based on the median normalized (chapter 2.12.2.b) Raman reference data set are shown in Figure 3.3 (AOA weighted) and Figure 3.4 (AOA+ weighted). Figure 3.3 showed two main clusters. Cluster 1 contained all AOA species (red), all iso-diabolic containing species (green) and three mean spectra of microorganisms, which were not expected to cluster within the AOA group (indicated by a black arrow). Cluster 1 also consisted of two main sub-clusters, A and B. All four AOA mean spectra were assigned to the sub cluster B. Cluster 2 (purple) contained *S. acidocaldarius*. However, the other two *Sulfolobus* species (blue) were found in cluster 1, sub cluster A. The calculated OOB error estimate was 12.5 %.

Figure 3.4 was based on the AOA+ RF weighting and showed an improved result compared to the AOA only weighting (Fig. 3.3). The AOA *N. uzonensis* (red) was now located in sub cluster A, and the other three AOA (red) were in the sub cluster B. In addition, all iso-diabolic acid containing organisms (green) were inside the sub cluster B, plus one unexpected species (*D. oxyclinae*), indicated by a black arrow. In comparison to the AOA-only weighting, two organisms (*A. longum, M. rosea*) which were not expected to be assigned to the cluster 1 fell out of it. *D. oxyclinae* clustered together with *N. viennensis* in both RF weighting methods (Fig. 3.3 and 3.4). The OOB error estimate of Figure 3.4 improved to 5 % in comparison to the AOA only weighting (12.5 %).

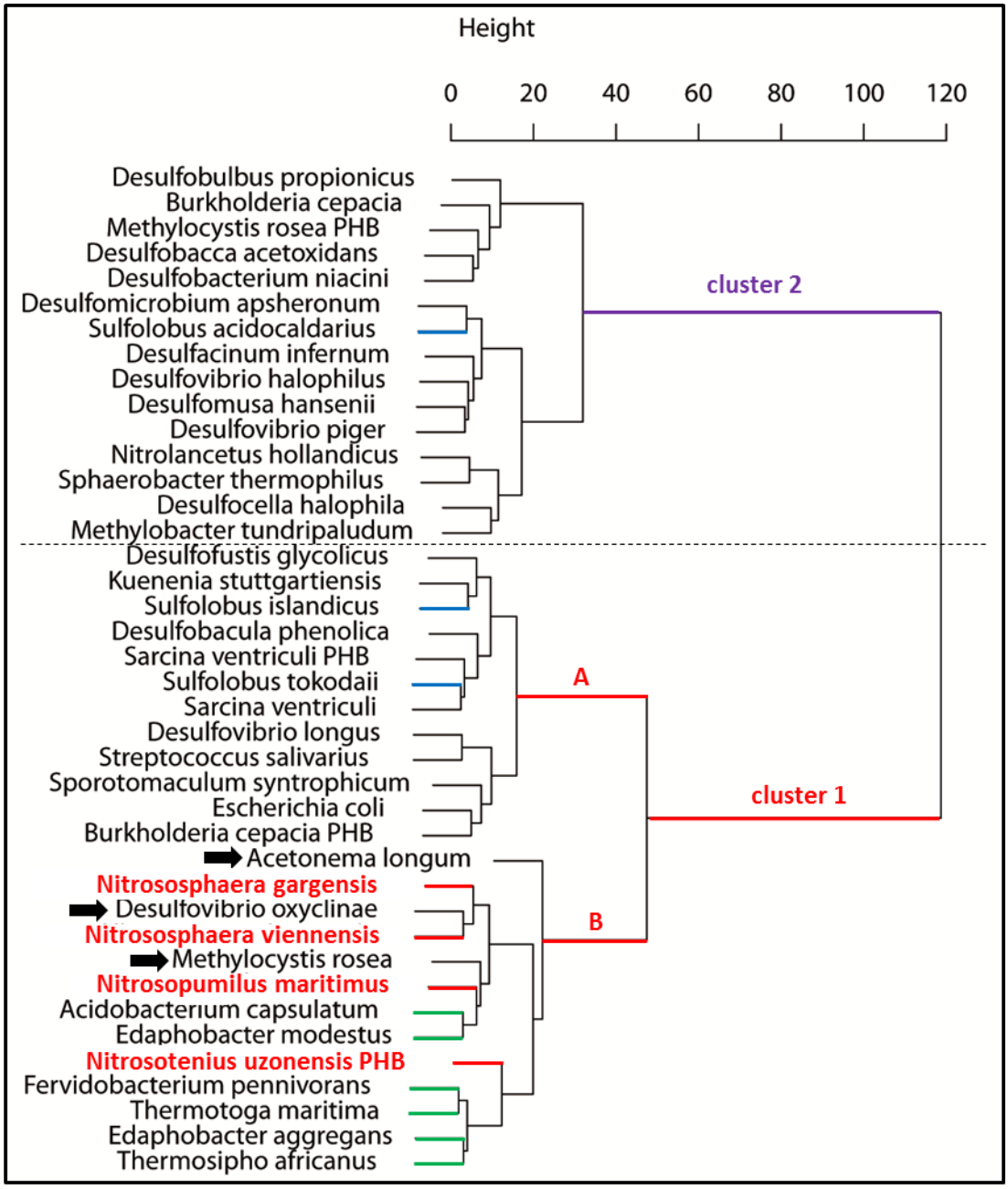

Figure 3.3. Cluster dendrogram based on the Raman reference library. Cluster 1 contains the AOA (red). In addition, the sub clusters A and B of cluster 1 are indicated in the figure. Cluster 2 (violet) does not contain AOA species. *Sulfolobus* species are indicated by a blue label. The classification was based on a Random Forest weighting towards only the AOA mean spectra. PHB next to a species name indicates that these Raman mean spectra contained signals from the storage compound polyhydroxybutyrate, which was subtracted by the PHB filter script (chapter 2.12.6). The data were mean spectra and processed as follows: phenylalanine-shifted to position 1004 cm^{-1}, line-segmented 8^{th} degree baselining and phenylalanine normalized (chapter 2.12).The data were mean spectra and processed as follows: phenylalanine-shifted to position 1004 cm^{-1}, line-segmented 8^{th} degree baselining and median normalized (chapter 2.12).

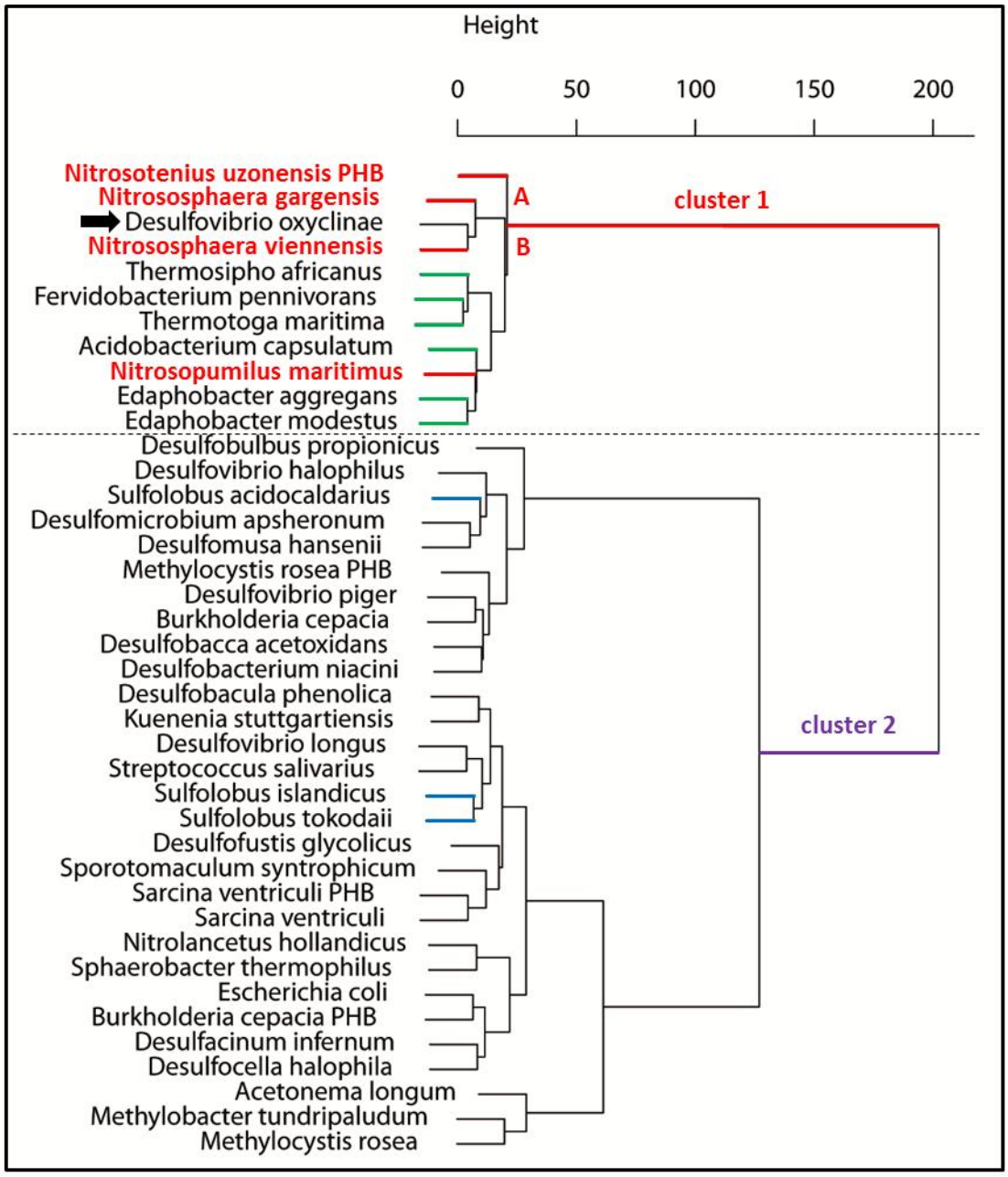

Figure 3.4. Cluster dendrogram based on the Raman reference library. Cluster 1 contains the AOA (red). In addition, the sub clusters A and B of cluster 1 are indicated in the figure. Cluster 2 (violet) does not contain AOA. *Sulfolobus* species are indicated by a blue label. The classification was based on a Random Forest weighting towards the AOA+ mean spectra (chapter 2.14.3). PHB next to a species name indicates that these Raman mean spectra contained signals from the storage compound polyhydroxybutyrate, which was subtracted by the PHB filter script (chapter 2.12.6). The data were mean spectra and processed as follows: phenylalanine-shifted to position 1004 cm^{-1}, line-segmented 8^{th} degree baselining and phenylalanine normalized (chapter 2.12).

3.1.3 Mean normalized data set

The two cluster dendrograms based on the mean normalized (chapter 2.12.2.a) Raman reference data set are shown in Figure 3.5. (AOA weighted) and Figure 3.6. (AOA+ weighted). Figure 3.5 featured two main clusters. Cluster 1 (red) contained all AOA species, all iso-diabolic containing species (green) and two mean spectra of microorganisms which were not expected to cluster within the AOA group (indicated by a black arrow). Cluster 1 also consisted of two main sub-clusters, A and B. *N. viennensis* and *N. gargensis* clustered in sub cluster A and *N. uzonensis* and *N. maritimus* were assigned to sub cluster B. Cluster 2 (purple) contained all three *Sulfolobus* species (blue). The calculated OOB error estimate was 12.5 %.

Figure 3.6 was based on the AOA+ RF weighting and showed a highly similar result. Differences in the assignment could only be seen in cluster 2 and the scale of the overall distances. In both figures, *M. rosea* clustered together with *N. maritimus* and *D. oxyclinae* was closely assigned to *N. viennensis*. The OOB error estimate of Figure 3.6 improved to 5 % compared to the AOA-only weighting (12.5 %).

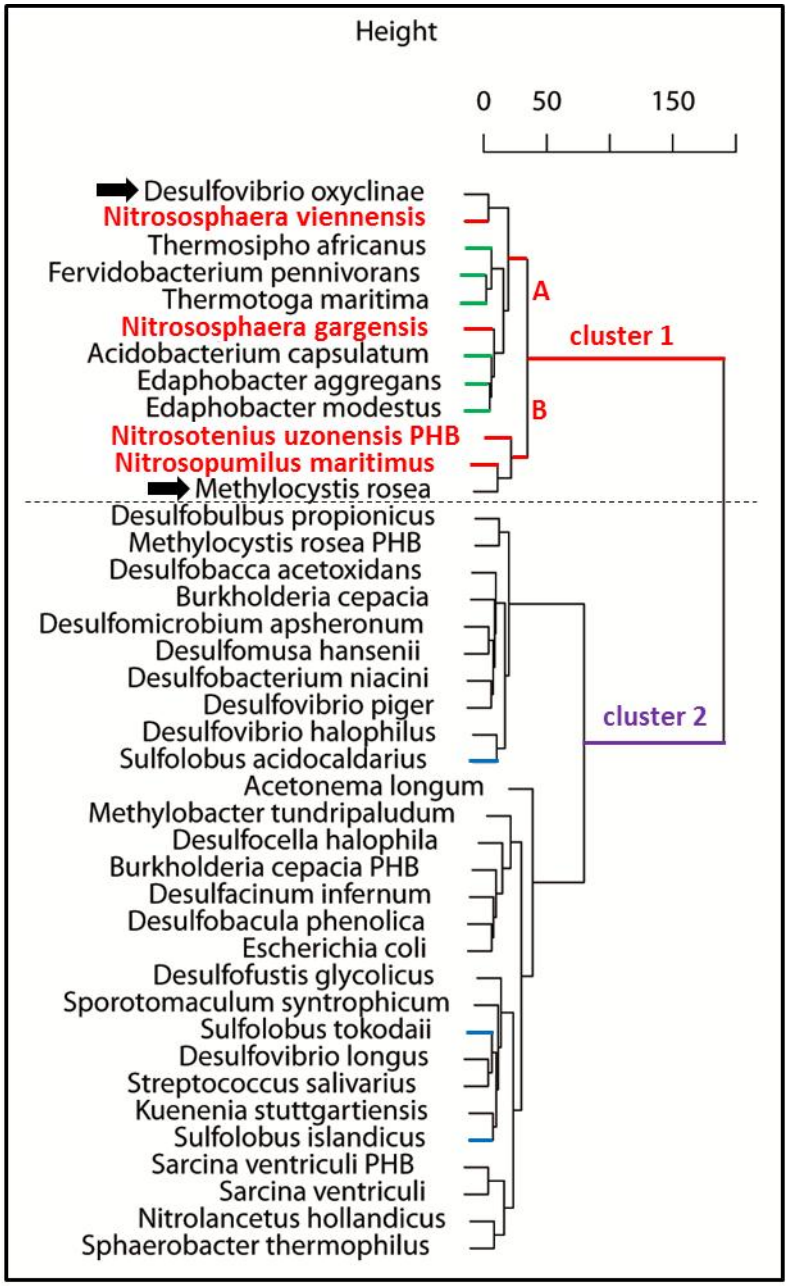

Figure 3.5. Cluster dendrogram based on the Raman reference library. Cluster 1 contains all AOA species (red). In addition, the sub clusters A and B of cluster 1 are indicated in the figure. Cluster 2 (violet) does not contain AOA. *Sulfolobus* species are indicated by a blue label. The classification was based on a Random Forest weighting towards only the AOA mean spectra. PHB next to a species name indicates that these Raman mean spectra contained signals from the storage compound polyhydroxybutyrate, which was subtracted by the PHB filter script (chapter 2.12.6). The data were mean spectra and processed as follows: phenylalanine-shifted to position 1004 cm^{-1}, line-segmented 8^{th} degree baselining and phenylalanine normalized (chapter 2.12).

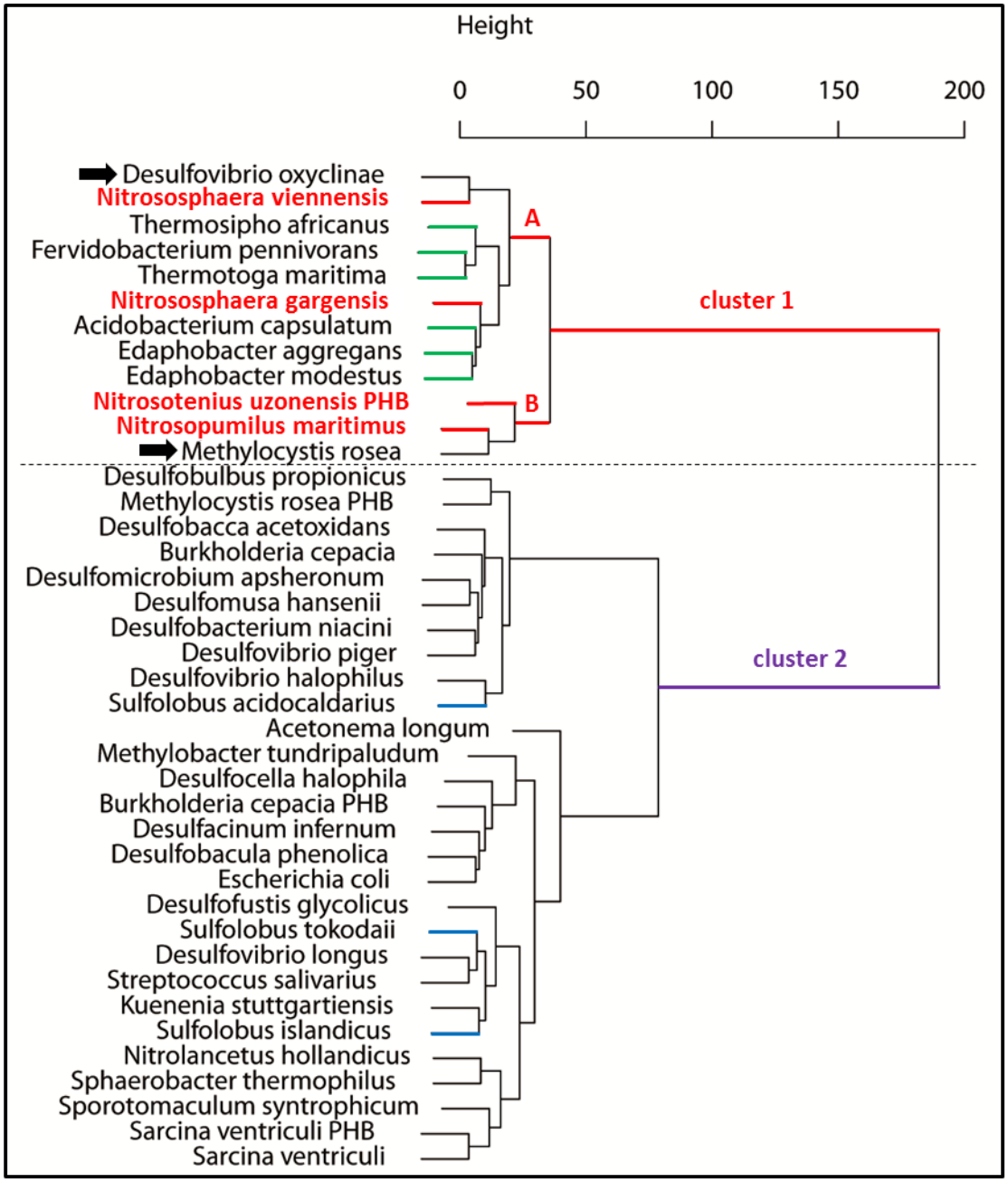

Figure 3.6. Cluster dendrogram based on the Raman reference library. Cluster 1 contains all AOA species (red). In addition, the sub clusters A and B of cluster 1 are indicated in the figure. Cluster 2 does not contain AOA (violet). *Sulfolobus* species are indicated by a blue labeling. The classification was based on a Random Forest weighting towards the AOA+ mean spectra (chapter 2.14.3). PHB next to a species name indicates that these Raman mean spectra contained signals from the storage compound polyhydroxybutyrate, which was subtracted by the PHB filter script (chapter 2.12.6). The data were mean spectra and processed as follows: phenylalanine-shifted to position 1004 cm^{-1}, line-segmented 8^{th} degree baselining and phenylalanine normalized (chapter 2.12).

Taken together, the best results (5% OOB error estimate) were achieved by the AOA + weighting and the use of the mean and median normalization. *D. oxyclinae* clustered together with *N. viennensis* in all approaches. The *Sulfolobus* species were never assigned to the same sub cluster as the AOA species. It was not possible to exclude all iso-diabolic acid containing species from the AOA cluster.

3.2 Artic AOA enrichments: SV8-6 and SV9-19

3.2.1 FISH/CARD-FISH images

Dr. Markus Schmid performed a FISH (chapter 2.15) and CARD-FISH (chapter 2.16) of the arctic AOA enrichment samples. In both enrichments, SV8-6 (Fig. 3.7 A) and SV9-19 (Fig. 3.7 B and D) signals were observed after performing a FISH with the general bacterial probe EUB338 (green). In addition, in SV9-19 some cells could be successfully labeled by performing a CARD-FISH with the general archaeal probe ARCH915 (red) (Fig. 3.7 C). Furthermore, in SV9-19 also a FISH was performed with the general bacterial probe EUB338 after a successful CARD-FISH (probe: ARCH915) was done (Fig. 3.7 D). Unfortunately, it was not possible to achieve an archaeal hybridization signal for the cells from the SV8-6 enrichment (data not shown).

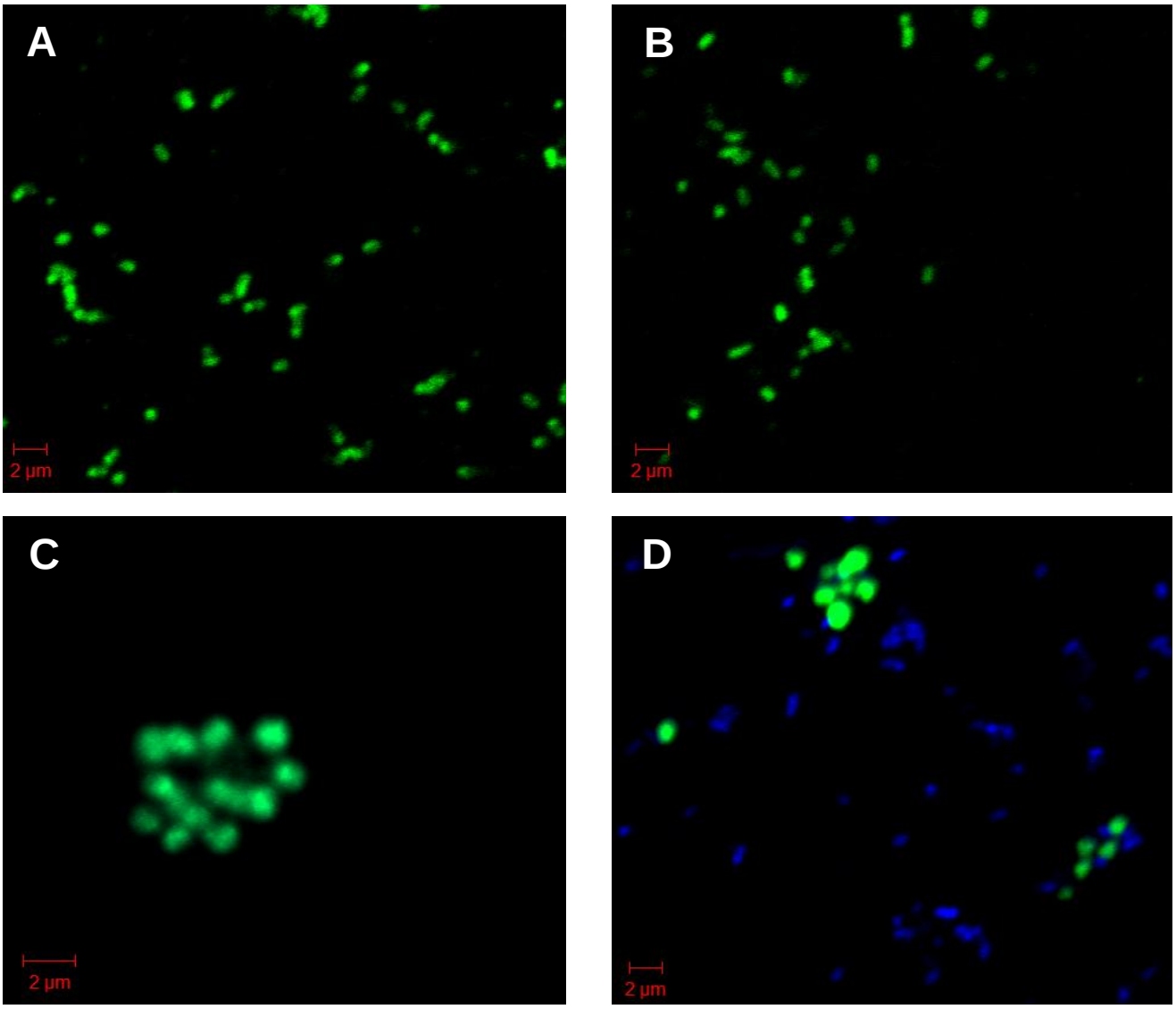

Figure 3.7. Images showing probe-hybridized cells of the two arctic AOA enrichments. A = SV8-6, FISH, probe: EUB338 mix, FA 10%; B = FISH, SV9-19, probe: EUB338 mix, FA 10%; C = SV9-19, CARD-FISH, probe: ARCH915, FA 20%; D = SV9-19, FISH, probe: EUBmix338 (blue), FA 10 % and CARD-FISH, probe: ARCH915 (green), FA 20%.

3.2.2 AOA cluster probabilities of arctic AOA enrichment cells

Figure 3.8 shows the probabilities of Raman spectra from randomly chosen individual cells of the arctic AOA enrichments SV8-6 and SV9-19 to cluster together with the AOA reference spectra (mean normalization of the reference spectra library data set and use of an AOA+ weighting) (Fig. 3.6) based on the AOA prediction script (chapter 2.14.3).

The minimum percentage of probability to be positively assigned to the AOA cluster was achieved from cell 13 of the AOA enrichment culture SV9-19 with 35.21 % (Fig. 3.8). All cells with a lower percentage were not assigned to the AOA cluster of the dendrogram (Fig. 3.6). Hence, they were not

considered to be an AOA species. Based on this prediction approach SV8-6 featured a total AOA content of 30%, whereas 45% of all measured SV9-19 cells were assigned to the AOA cluster.

SV8-6	
cell 1	91.42
cell 2	25.51
cell 3	19.96
cell 4	28.32
cell 5	23.95
cell 6	26.6
cell 7	20.29
cell 8	21.94
cell 9	88.25
cell 10	67.24

SV9-19	
cell 1	88.82
cell 2	95.55
cell 3	91.42
cell 4	82.3
cell 5	67.84
cell 6	53.45
cell 7	80.75
cell 8	17.27
cell 9	21.9
cell 10	13.18
cell 11	11.19
cell 12	14.65
cell 13	35.21
cell 14	13.25
cell 15	11.04
cell 16	16.15
cell 17	59.56
cell 18	15.15
cell 19	15.61
cell 20	17.26

Figure 3.8. Cluster probabilities (in %) of the arctic AOA enrichments SV8-6 and SV9-19. Cells which were assigned to the AOA cluster (red) of Fig. 3.6 using the AOA prediction script (chapter 2.14.3). The cell spectra were processed as follows: phenylalanine shifted, baselined (8^{th} degree, line segmented) and mean normalized. Furthermore, an AOA+ weighting was used for calculating the RF weightings.

3.2.3 Images of predicted AOA cells from arctic AOA enrichments

Raman spectra of the arctic AOA enrichments SV8-6 (n = 10) and SV9-19 (n = 20) were acquired. These 30 cells were chosen at random and an image of every cell was recorded directly by the Labspec 5 software. The following images (Fig. 3.9 A - L) show the red indicated cells from Figure 3.8, which were predicted to be AOA species and thus were assigned to the AOA cluster of Figure 3.6. The quality of the shown images is rather low. This is a result of an abdication of an embedding solution, which would have caused problems for the Raman acquisition. Theoretically, also the use of

water- instead of air-objectives would improve the quality of the images.

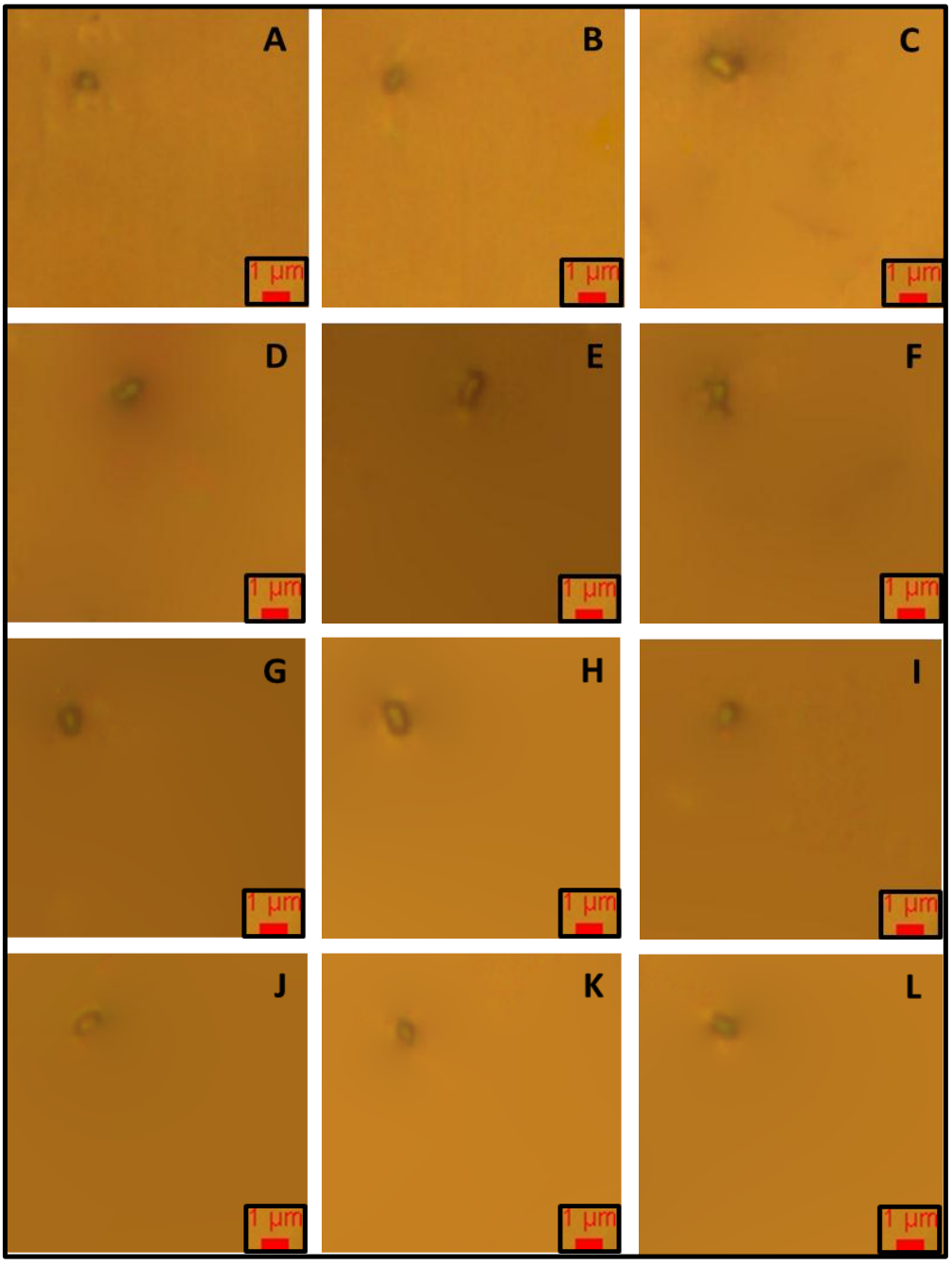

Figure 3.9. Brightfield images of cells from two AOA enrichment cultures SV8-6 and SV9.19, which were assigned to the AOA cluster of Figure 3.4. The following cell numbers are referring to the Figure. 3.8: SV8-6 cell 1 (A), cell 9 (B) and cell 10 (C), and SV9-19 cell 1 (D), cell 2 (E), cell 3 (F), cell 4 (G), cell 5 (H), cell 6 (I), cell 7 (J), cell 13 (K) and cell 17 (L) were assigned to the AOA cluster of Figure 3.6 by the application of the AOA prediction script (chapter 2.14.3).

3.3 Raman spectra of storage compounds

Storage compounds are very abundant in both, bacteria and archaea. During the diploma thesis I observed that a storage compound like PHB can overlay the whole cell spectrum of a microorganism (Fig. 3.11). The main goal of this study was to find AOA specific peaks and such an immense

contribution made this very challenging. As a result, the subtraction of storage compounds was necessary to reveal even small peaks which may be covered by storage compounds. Thus, Raman spectra of certain abundant ones were acquired during this diploma thesis in order to create a filter for them, because different cell types can have similar storage compounds even if they are not closely related.

3.3.1 Glycogen

The branched polymer glycogen (Sigma-Aldrich) was analyzed by Raman microspectroscopy on a CaF_2 carrier slide (Crystran). The most pronounced peaks of the mean Raman spectrum (Fig. 3.10) can be seen at the Raman shift positions: 440, 479, 575, 710, 756, 835, 938, 1054, 1082, 1124, 1259, 1335, 1377 and 1458 cm^{-1}. This spectrum is in accordance with other Raman spectroscopy analyses of glycogen (Majed and Gu., 2010). The spectra were acquired by using the following parameters in the Labspec 5 (Horiba) software: 10 seconds of acquisition time, pinhole size 600 µm and filter 0.3 (58 % laser intensity).

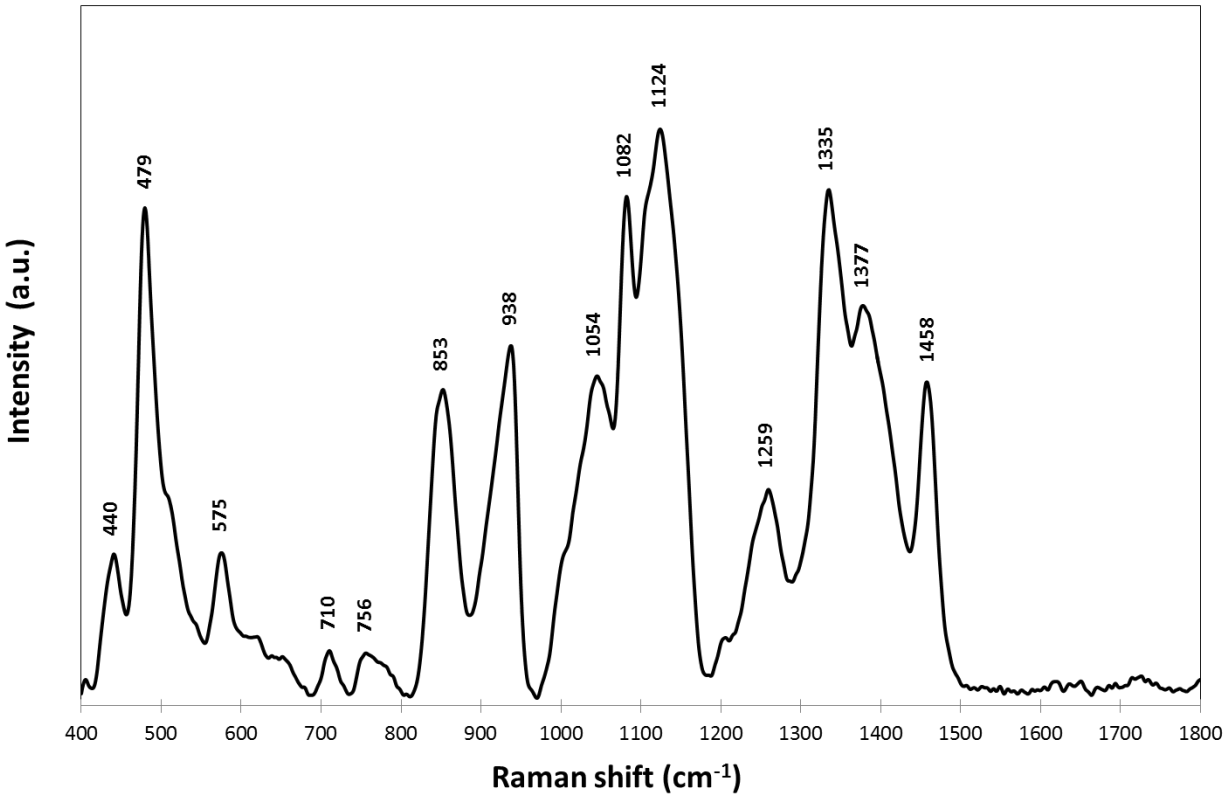

Figure 3.10. Mean Raman spectrum of glycogen (n = 13). The data were smoothed, line-segmented baselined (8^{th} degree) and mean normalized (chapter 2.12).

No indicative glycogen peaks could be found in our library of microorganisms. Hence, I assumed it did not have a significant influence to the clustering process.

3.3.2 Polyhydroxybutyrate

During this diploma thesis numerous Raman spectra were acquired that contained the storage compound polyhydroxybutyrate (PHB). It was found in the following microorganisms: *Burkholderia cepacia* (Fig. 8.7), *Sarcina ventriculi* (Fig. 8.27), *Nitrosotenius uzonensis* (Fig. 8.34) and *Methylocystis rosea* (Fig. 8.5). Presence of PHB affects large parts of a cellular Raman spectrum (Fig. 3.11). The Raman difference spectrum of *Sarcina ventriculi*, with and without PHB storage (Fig. 3.12), showed characteristic peaks at the Raman shift positions 619 cm^{-1}, 833 cm^{-1}, 860 cm^{-1}, 901 cm^{-1}, 958 cm^{-1}, 1060 cm^{-1}, 1106 cm^{-1}, 1145 cm^{-1}, 1209 cm^{-1}, 1236 cm^{-1}, 1299 cm^{-1}, 1355 cm^{-1}, 1425 cm^{-1}, 1457 cm^{-1} and 1737 cm^{-1}.These results were quite similar to other Raman measurements of PHB with the exception of minor peak shifts (e.g. 1737 instead of 1735 cm^{-1}) (Furukawa et al., 2006; De Gelder et al., 2008; Ciobotă et al., 2010; Majed and Gu, 2010).

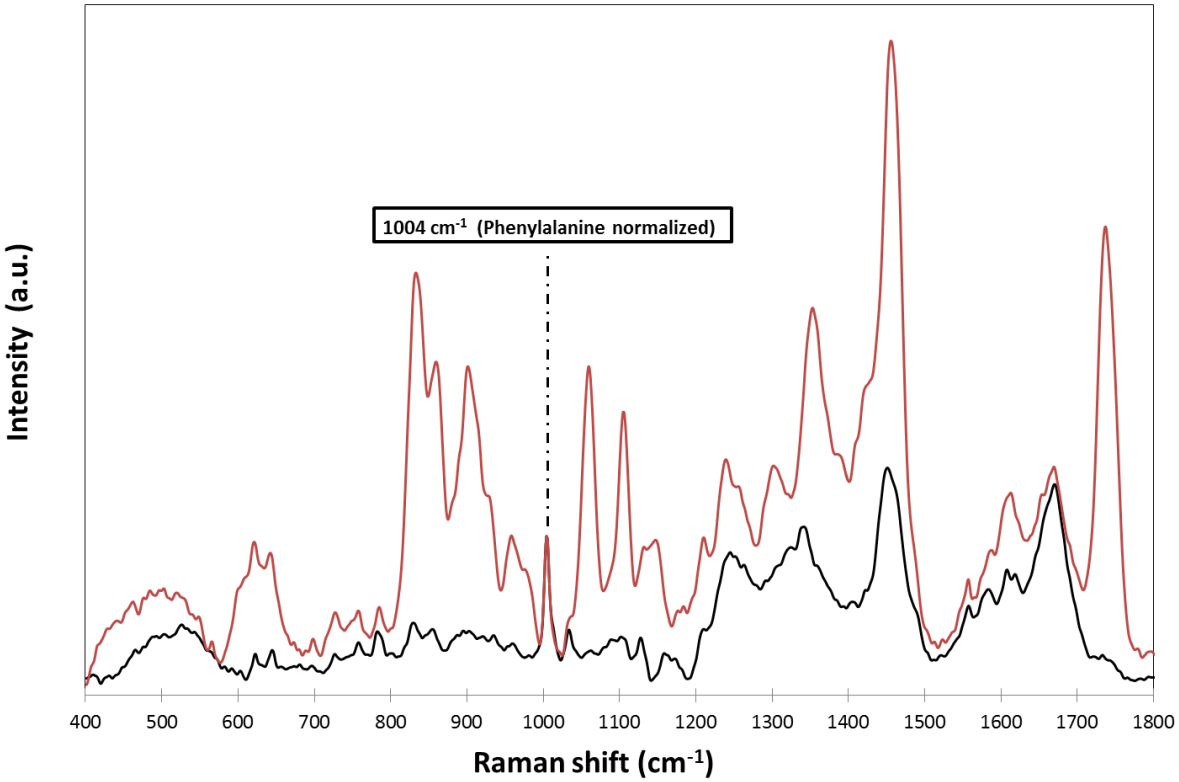

Figure 3.11. Overlay spectra of two *Sarcina ventriculi* Raman spectra with (red) and without (black) the storage of polyhydroxybutyrate. The spectra were acquired from two cells of the same culture. The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8^{th} degree) and normalized to the height of the Phe peak at position 1004 cm^{-1} (chapter 2.12).

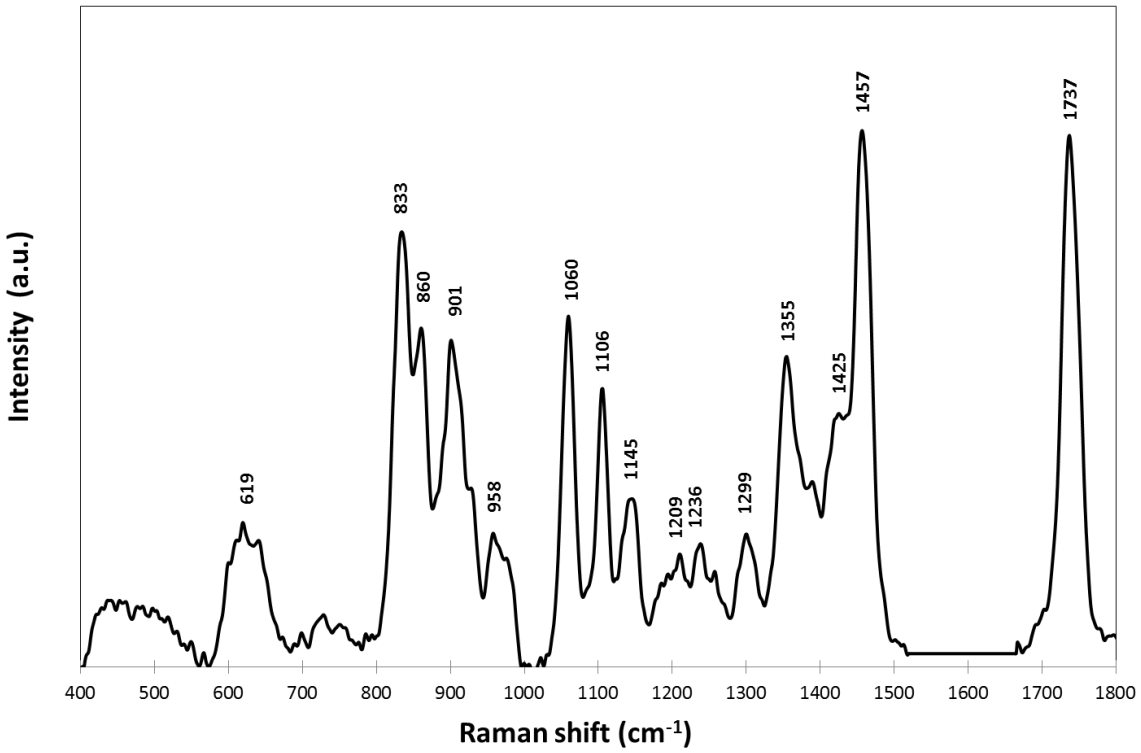

Fig. 3.12. Difference spectrum of two *Sarcina ventriculi* Raman spectra - with and without the storage compound polyhydroxybutyrate (Fig. 3.11). The remaining peaks showed the characteristic spectrum of PHB.

No other storage compounds were analyzed by Raman spectroscopy during this study because PHA and glycogen belong to the most abundant ones in prokaryotic cells. In addition, Raman spectra of other storage compounds like sulfur (Ward, 1968) and polyphosphates (Majed and Gu, 2010) were checked in the literature and compared to the library of microorganism of this study. No presence of these could be observed.

3.4 CaF_2 Raman spectrum

Figure 3.13 features the mean Raman spectrum of the used CaF_2 carrier slide (Crystran). CaF_2 showed a highly characteristic peak at position 321 cm^{-1}. Furthermore, there are some less pronounced peaks in the area between 400 and 1100 cm^{-1}, which were shown in more detail in Figure 3.14. The marked peaks indicated the most prominent peaks at around wavenumber 517, 809, 909 and 1556 cm^{-1}. Two different baselining approaches were performed on the Raman spectrum CaF_2. The ratio between the peaks at position 517 and 809 cm^{-1} differed significantly based on the degree of the polynomial equation used for baselining (Fig. 3.15).

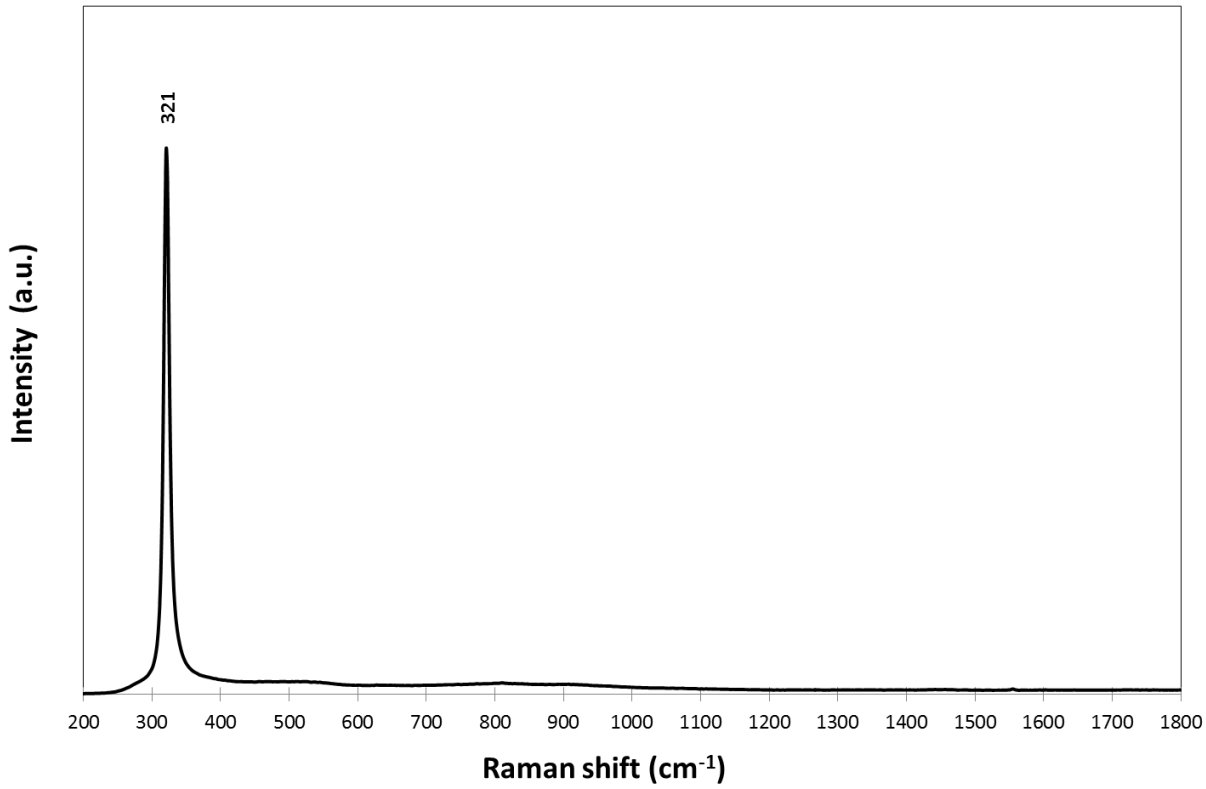

Figure 3.13. Mean Raman spectrum of the CaF_2 Raman carrier slide (Crystran) used in this study (n = 17). The data were not processed.

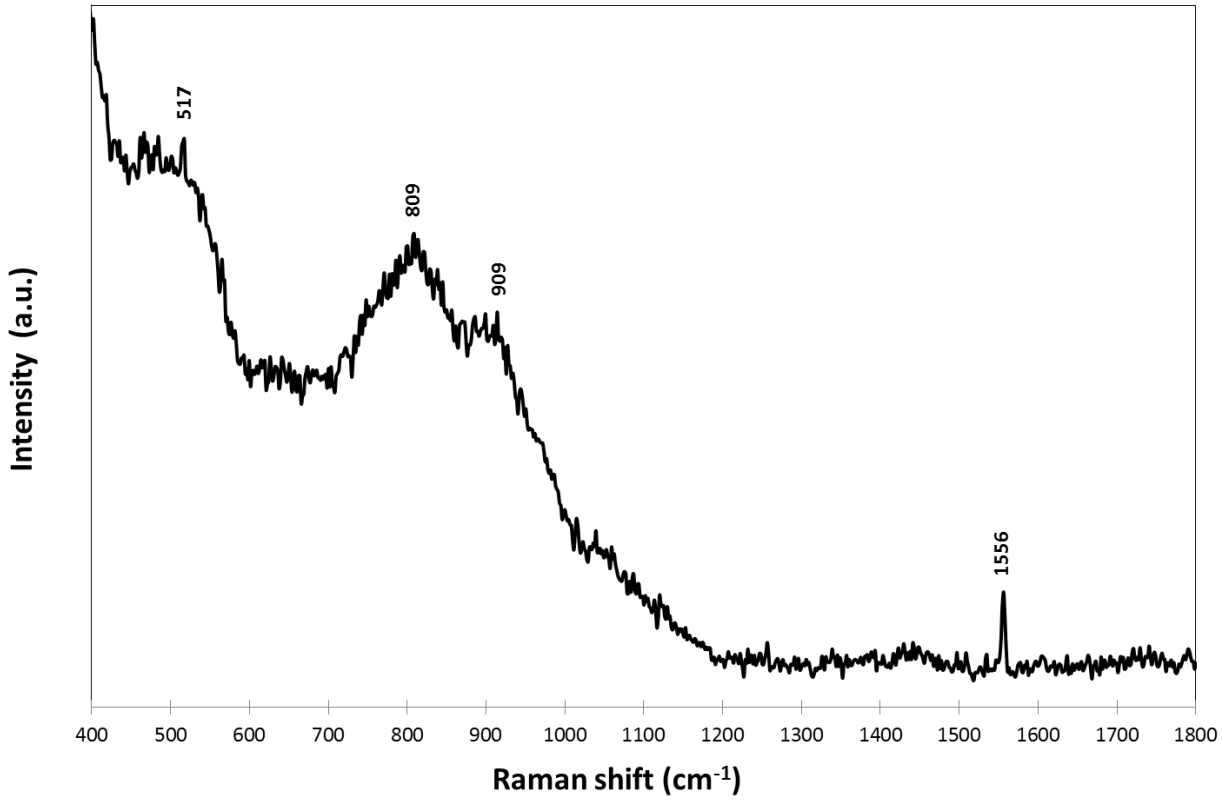

Figure 3.14. Raman spectrum of the CaF_2 Raman carrier slide (Crystran) used in this study. The data was not processed. Unlike the Fig. 3.14, the spectrum of this figure starts at 400 and not at 300 cm^{-1}. This shows the influence of CaF_2 to the acquired Raman spectra during this diploma thesis.

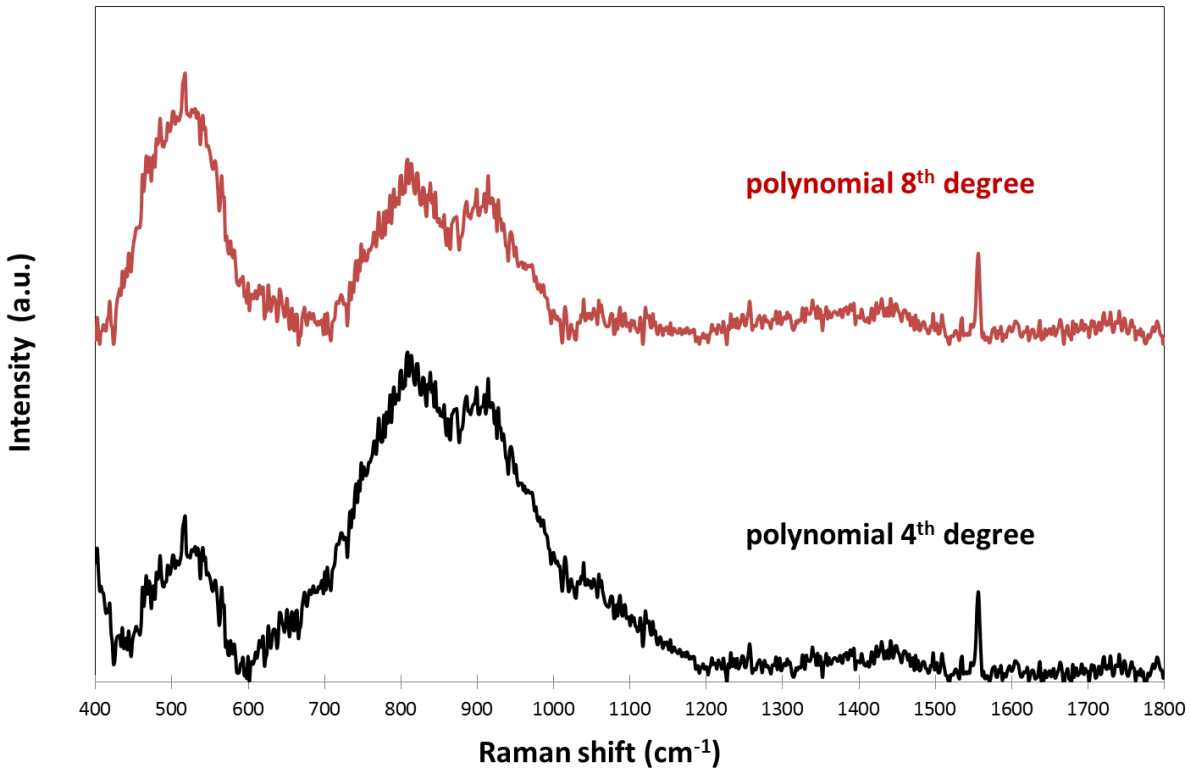

Figure 3.15. Comparison of the effect of different baselining parameters to the same Raman spectrum (Fig. 3.14) of the used CaF_2 carrier slide (Crystran) used in this study. The data were line-segmented baselined 8th degree (red) and 4th degree (black).

3.4.1 CaF_2 intensity based on Raman acquisition time

At a rather low acquisition time of 20 seconds (without the use of a laser intensity filter, pinhole size: 600 µm, the CaF_2 background signal does only significantly influence the spectrum of a cell at wavenumber 321 cm^{-1}. However, this situation changed dramatically the longer the acquisition time became (Fig. 3.16). Subsequently, also the use of laser intensity filters had a major impact to the height of Raman peaks of CaF_2. Characteristic CaF_2 peaks after 120 seconds of acquisition time (no filter, pinhole size: 600 µm) could be observed at the wavenumbers 520, 804, 898 and 1554 cm^{-1}. To put this in perspective, cell spectra in this study where acquired using different parameters. From low intensity (*D. infernum*; 20 seconds, pinhole size: 600 µm, filter 0.6 for 28% laser intensity) to high intensity settings (*N. maritimus*; 120 seconds, pinhole size: 500 µm, no filter).

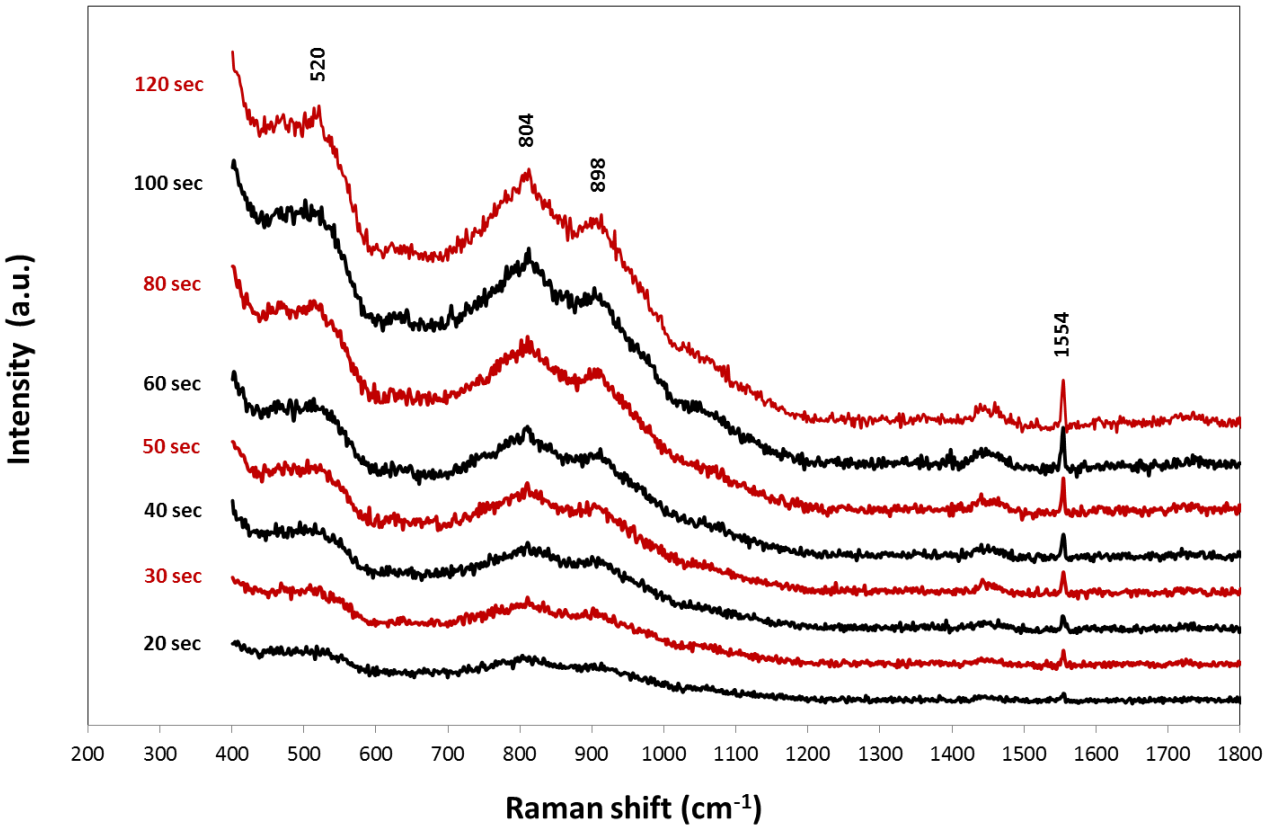

Figure 3.16. Influence of the acquisition time to the Raman spectrum of the CaF_2 carrier slide (Crystran) used in this study. The data were unprocessed raw spectra and no filter was used during the spectra acquisition. The wavenumbers of four peaks that were most pronounced after 120 sec of acquisition time with the use of no laser intensity filter (pinhole size 600 µm) were indicated in the image.

3.4.2 Influence of CaF_2 to cell spectra

I discovered that the increase of intensity of the CaF_2 peaks was stronger than the increase of the cell spectra peaks when the laser acquisition time was increased (data not shown). Subsequently, the cell spectra that were recorded with more intense laser settings contained a stronger CaF_2 background signal contribution. Furthermore, also smaller cells had a stronger CaF_2 contribution because less cell material and more slide support was measured with the relatively large pinhole diameter applied in this study. The two cell spectra in Figure 3.17 were acquired with the same acquisition time (60 seconds), but the spectrum of *N. gargensis* (red) was recorded with the filter 0.3 (58 % laser intensity), whereas the cell spectrum of *D. oxyclinae* (black) was acquired with the filter 0.6 (28 % laser intensity). Hence, the laser intensity for *D. oxyclinae* was only around half as much and in addition, the cells of *N. gargensis* were even smaller than the ones of *D. oxyclinae*. As a consequence, the spectrum of *N. gargensis* contained a much stronger CaF_2 background compared to the cell signal. The effect of the used baselining parameters (line-segmented, 8^{th} degree) on these two cell spectra were shown in Figure 3.18. The formerly quite different cell spectra were changed to very

similar looking spectra due to the chosen baselining parameters and the resulting subtraction.

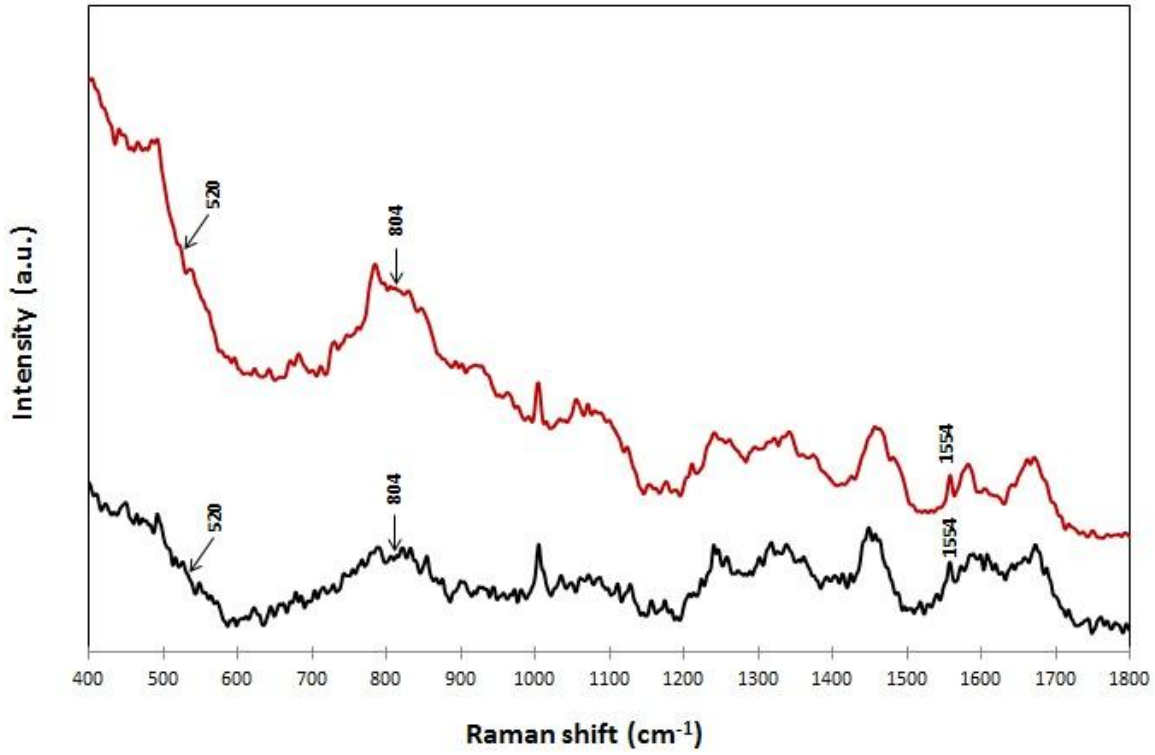

Figure 3.17. Raman spectra comparison between *N. gargensis* (red) and *D. oxyclinae* (black). The data were aligned to the phenylalanine peak at position 1004 cm^{-1}, smoothed, (chapter 2.12). The most characteristic wavenumbers of the CaF_2 peaks (Fig. 3.16) were indicated to show their influence to the whole cell spectrum.

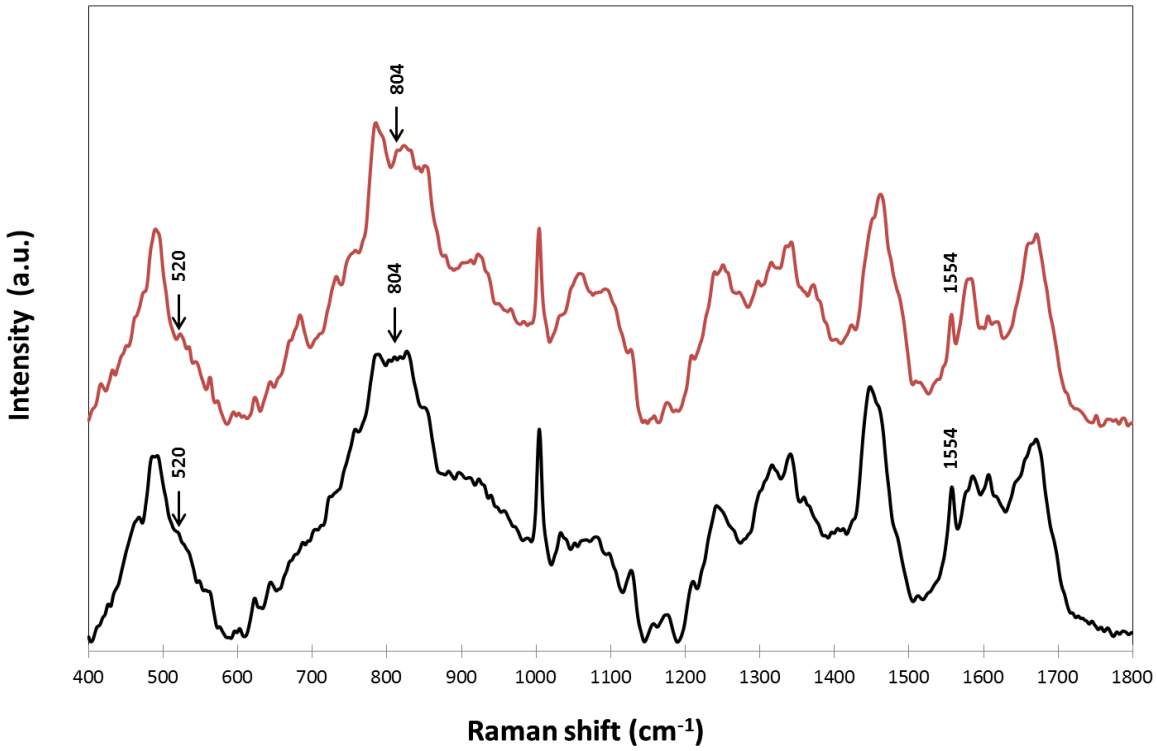

Figure 3.18. Impact of baselining to the Raman spectra comparison (Fig. 3.17) between *N. gargensis* (red) and *D. oxyclinae* (black). The data were aligned to the phenylalanine peak at position 1004 cm^{-1}, smoothed, line-segmented baselined (8^{th} degree) and mean normalized (chapter 2.12). The most characteristic wavenumbers of the CaF_2 peaks (Fig. 3.16) were indicated to show their influence to the whole-cell spectrum.

3.5 Raman spectrum of crenarchaeol

3.5.1 Raman spectrum of crenarchaeol with CaF_2 background signal

The lipid crenarchaeol was a gift from Jaap. S. Damsté and measured on a CaF_2 carrier slide. Hence, the acquired Raman spectrum of crenarchaeol (Fig. 3.19) contained the characteristic CaF_2 background signal, which was mainly indicated by the very strong peak at wavenumber 321 cm^{-1} (see Fig. 3.13 for a comparison with CaF_2).

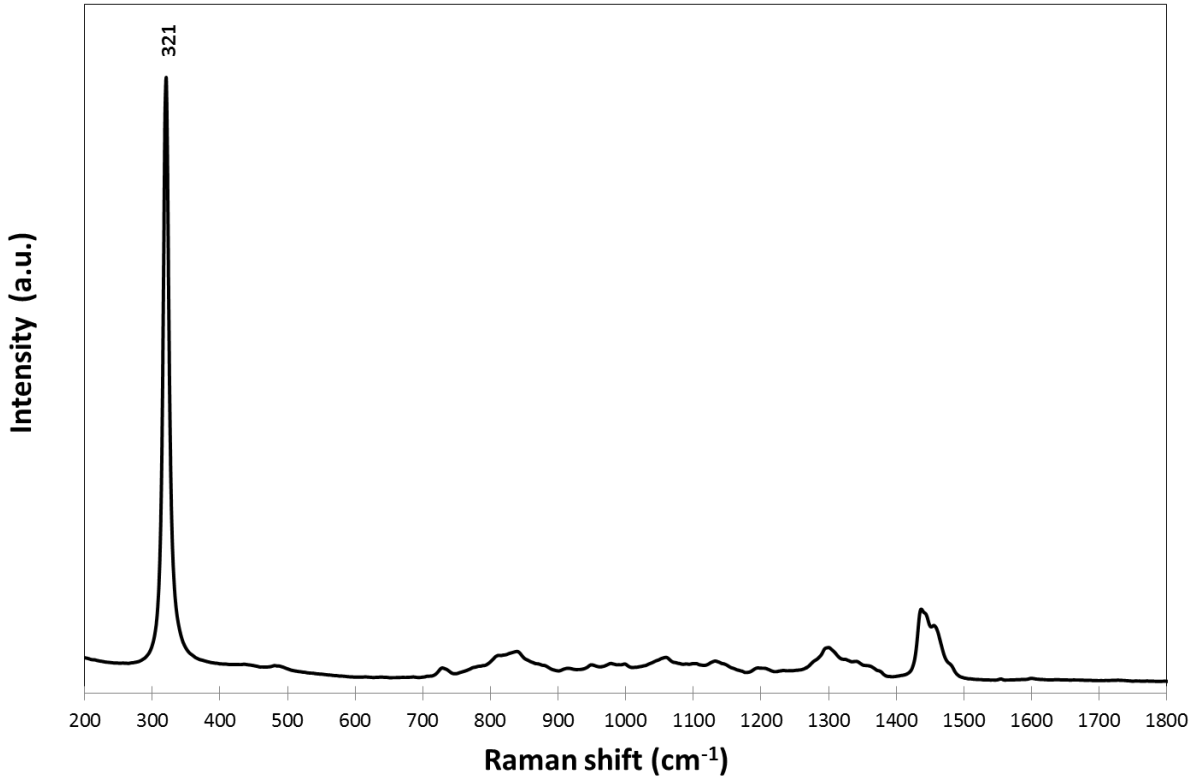

Figure 3.19. Mean Raman spectrum of crenarchaeol (n = 7). The data were smoothed and the most pronounced peak of the spectrum (321 cm^{-1}) indicated a CaF_2 background signal (see Fig. 3.13). The Raman spectra were acquired by Dr. Markus Schmid.

3.5.2 Raman spectrum of crenarchaeol without CaF_2 background signal

Crenarchaeol was the only spectrum obtained in this study (besides CaF_2 itself) which was recorded between the wavenumbers 200 and 3200 cm^{-1}, thus the influence of CaF_2 could be investigated and subtracted because it had its indicator peak at the position 321 cm^{-1} (Fig. 3.13). Based on the height of this peak, the CaF_2 spectrum was aligned to the spectrum of crenarchaeol. However, the overlay was not precise concerning the wavenumbers starting at 1100 cm^{-1} of the CaF_2 spectrum. Numerous low values of the data points of the CaF_2 were below zero compared to the crenarchaeol spectrum data points after the alignment to the height of the peak (321 cm^{-1}). Hence, CaF_2 was subtracted from crenarchaeol only between 200 and 1100 cm^{-1}, a region in which all pronounced peaks of the spectrum of CaF_2 are located (Fig. 3.13). The CaF_2 subtracted spectrum of crenarchaeol (red) was relatively consistent with the non-subtracted spectrum (black) (Fig. 3.20). Moreover, the peak positions stayed the same, but the intensities of the peaks changed slightly. Taken together, the short acquisition time (20 sec) required for the acquisition of the crenarchaeol spectrum resulted in a rather weak contribution of CaF_2 to the whole spectrum.

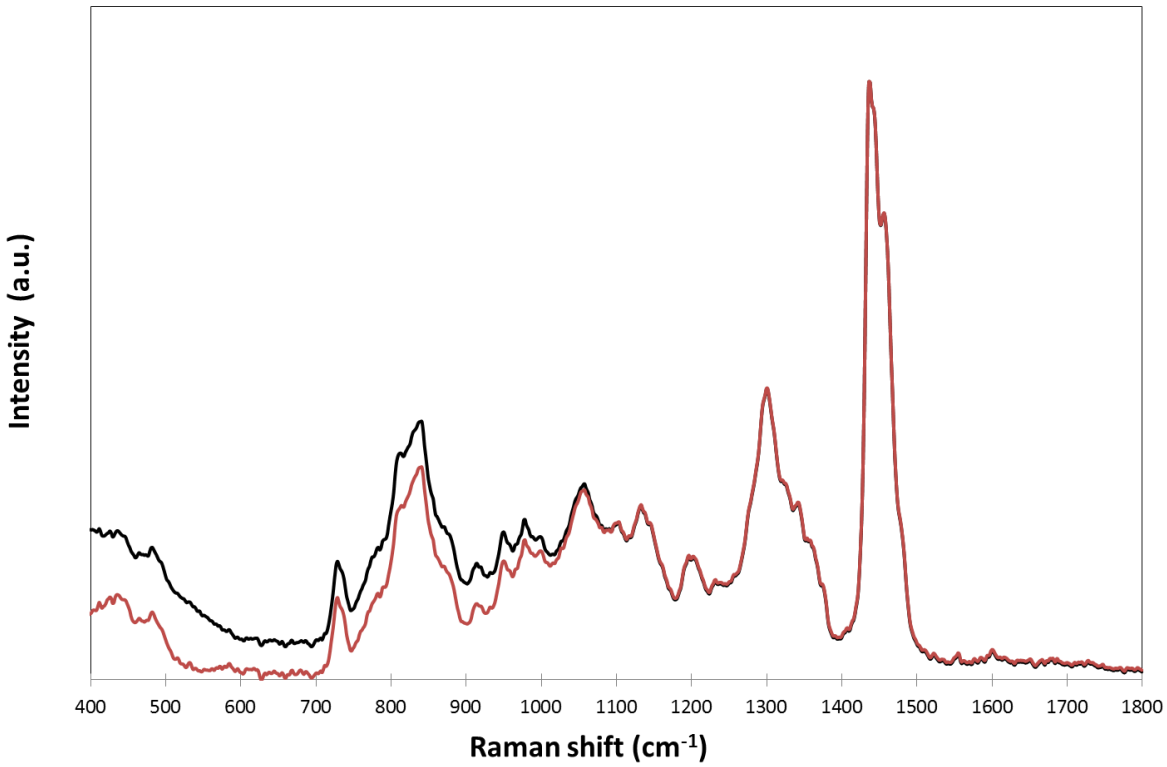

Figure 3.20. Overlay of Raman spectra of crenarchaeol with CaF_2 background (black) and after CaF_2 subtraction in the region from 400 to 1100 cm^{-1} (red) (n = 7). The data were smoothed and the Raman spectra were acquired by Dr. Markus Schmid.

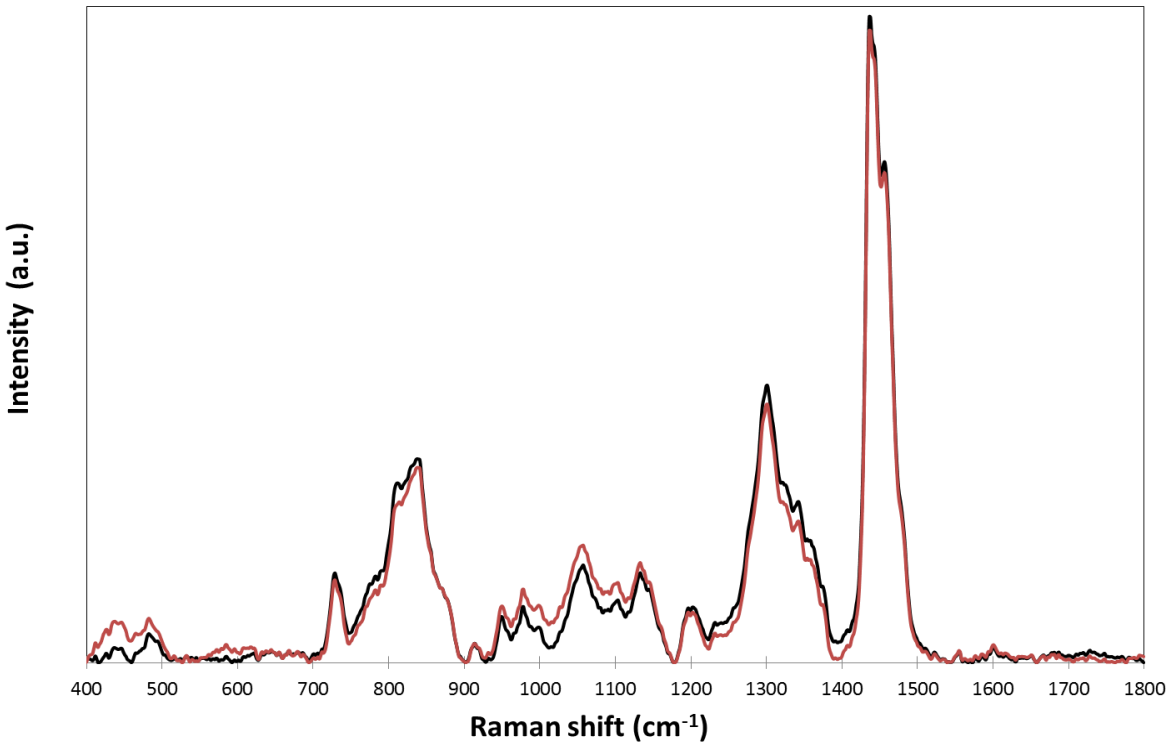

Figure 3.21. Influence of baselining to the Raman spectra of crenarchaeol with (black) and without (red) CaF_2 background signal from Fig. 3.20. The data were polynomial baselined (8^{th} degree) and the Raman spectra were acquired by Dr. Markus Schmid.

3.6 Peak assignment of crenarchaeol

The most pronounced Raman peaks of crenarchaeol (Fig. 3.22) were analyzed using the irAnalyze software (LabCognition). For all peaks a tentative assignment could be made with the exception of the small peak at position 1600 cm^{-1} (Tab. 3.1).

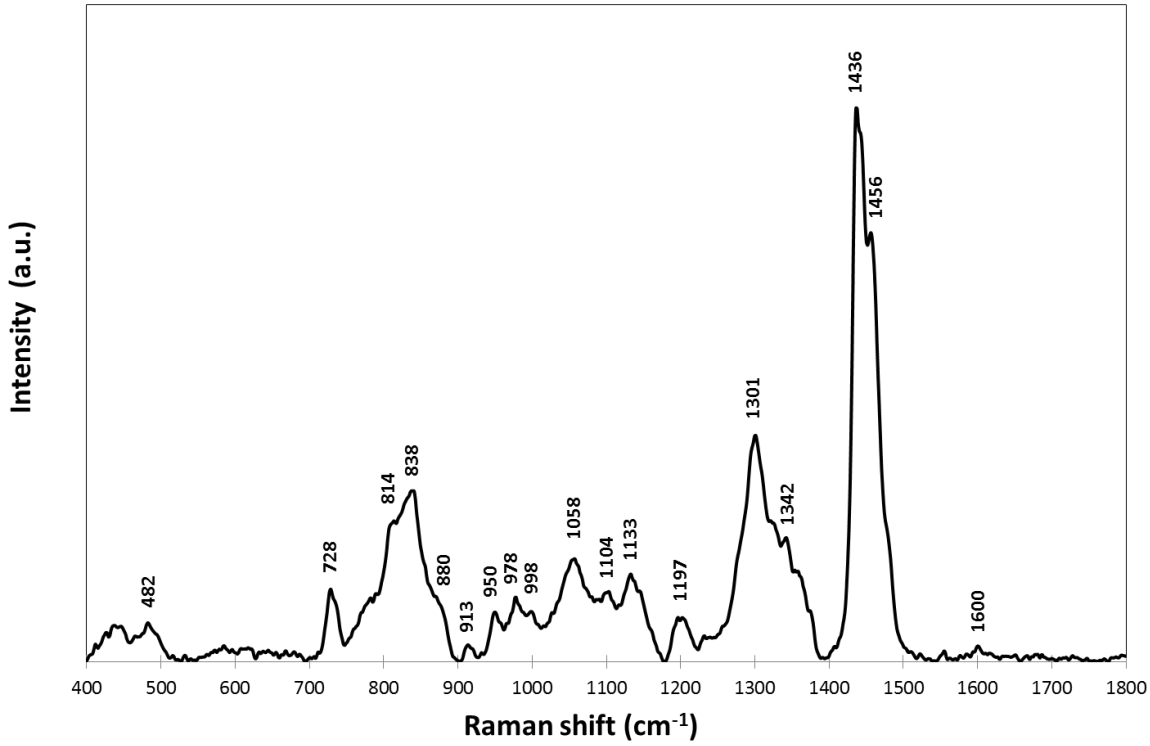

Figure 3.22. Mean Raman spectrum of crenarchaeol (after subtraction of the CaF_2 background), n=7. The data were smoothed and polynomial baselined (8^{th} degree). The wavenumbers of the most pronounced peaks were indicated. The Raman spectra were acquired by Dr. Markus Schmid.

Table 3.1 Crenarchaeol (Fig. 3.22) peak assignment based on the irAnalyze software. Relative intensities were denoted by: s = strong, m = medium, w = weak, v = very, sh = shoulder.

Raman band (cm^{-1})	functional group	vibration
479 (w)	aliphatic hydroxy compound	C-O in plane, C-O-C Def
728 (m)	long chain hydrocarbon	C-H, rock CH_2
814 (sh)	branched alkyl, methyl/long chain alkyl/aliphatic ether	C-C stretch (weak)/CCC stretch/C-O stretch (sym)
838 (s)	branched alkyl, methyl/long chain alkyl/aliphatic ether	C-C stretch (weak)/CCC stretch/C-O stretch (sym)
880 (sh)	long chain hydrocarbon/aliphatic ether	C-C-C stretch (weak)/C-O stretch (sym)
913 (w)	branched alkyl, methyl	C-C stretch

950 (w)	unsaturated hydrocarbon (ether conjugated)/branched alkyl, methyl	C-H bend out of plane/C-C stretch
978 (w)	large ring or long chain alkyl	C-C stretch ring
998 (sh)	large ring or long chain alkyl	C-C stretch ring
1058 (m)	large ring or long chain alkyl	C-C stretch ring
1104 (sh)	aliphatic ether	C-O stretch (asym)
1133 (m)	aliphatic ether	C-O stretch (asym)
1197 (m)	unsaturated hydrocarbon (ether conjugated)	C-O stretch
1301 (s)	aliphatic ether (polyethoxy)	CH_2 bend
1342 (sh)	branched alkyl, methyl	C-H bend, CH3
1436 (vs)	branched alkyl, methyl	C-H, bendCH2/CH3
1456 (sh)	branched alkyl, methyl	C-H, bendCH2/CH3
1600 (vw)	unkown	unknown

3.7 Raman spectra of diphytanoyl lipids

Raman spectra of two lipids, 1,2-di-O-phytanyl-*sn*-glycerol and 1,2-di-O-phytanyl-*sn*-glycero-3-phosphoethanolamine (Fig. 3.23), which have some structural similarities with crenarchaeol (chapter 1.3), were acquired using the following parameters: 10 sec of acquisition time, pinhole size 600 µm and filter 0.6 (28 % laser intensity) by Labspec 5 (Horiba). Furthermore, the chosen acquisition parameters (low intensity due to a short acquisition time and the used laser filter) resulted in a very minor contribution of CaF_2 to these spectra. The two lipids differed between each other just by one phosphoethanolamine side group. The Raman spectra showed a very high similarity with each other except for two remarkably different peaks at positions 761 and 1093 cm^{-1}, indicating the influence of the additional side group. Compared to crenarchaeol (Fig. 3.24), they lacked peaks at the Raman shift positions 479, 706, 998, 1133, 1197 and 1600 cm^{-1}. On the other hand, the majority of the peaks were rather comparable between crenarchaeol and the two diphytanoyl lipids

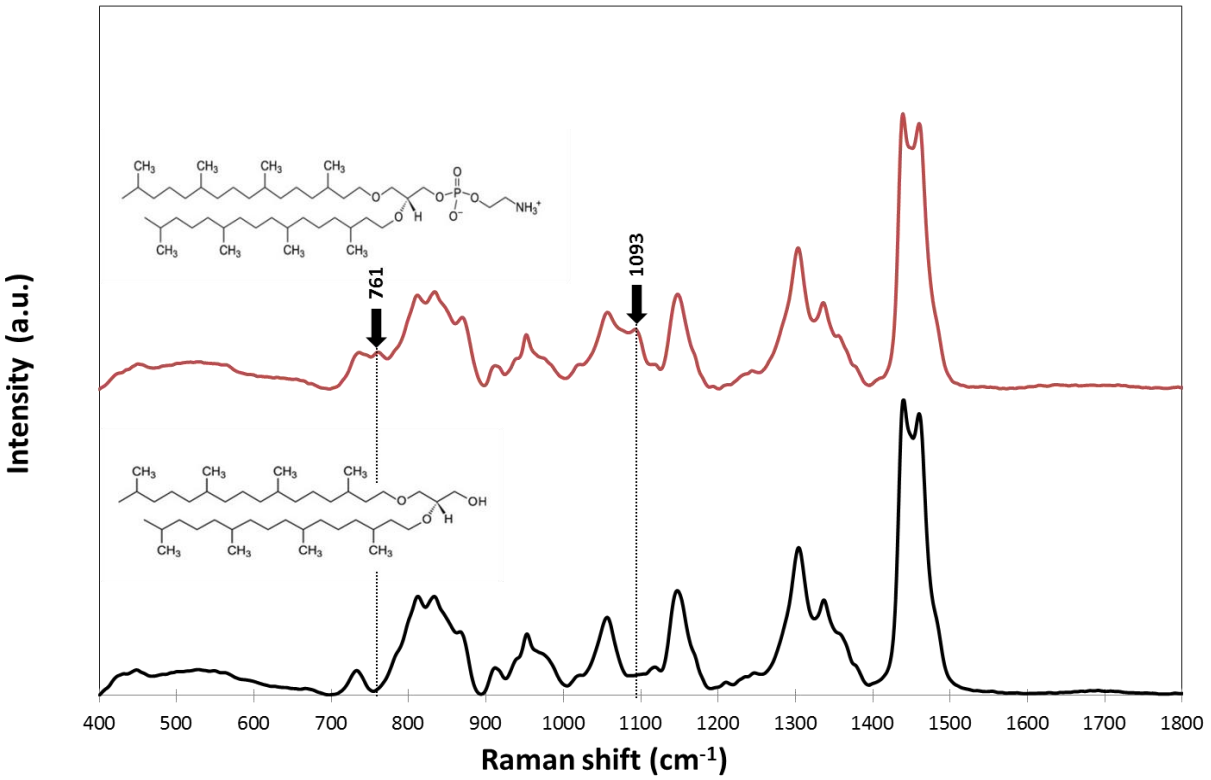

Figure 3.23. Comparison between the mean Raman spectra of the lipids 1,2-di-O-phytanyl-*sn*-glycerol (black, n = 6) and 1,2-di-O-phytanyl-*sn*-glycero-3-phosphoethanolamine (red, n = 6). The data were smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12). The wavenumbers of the most pronounced differences were indicated in the figure. These Raman spectra were acquired by Dr. Markus Schmid.

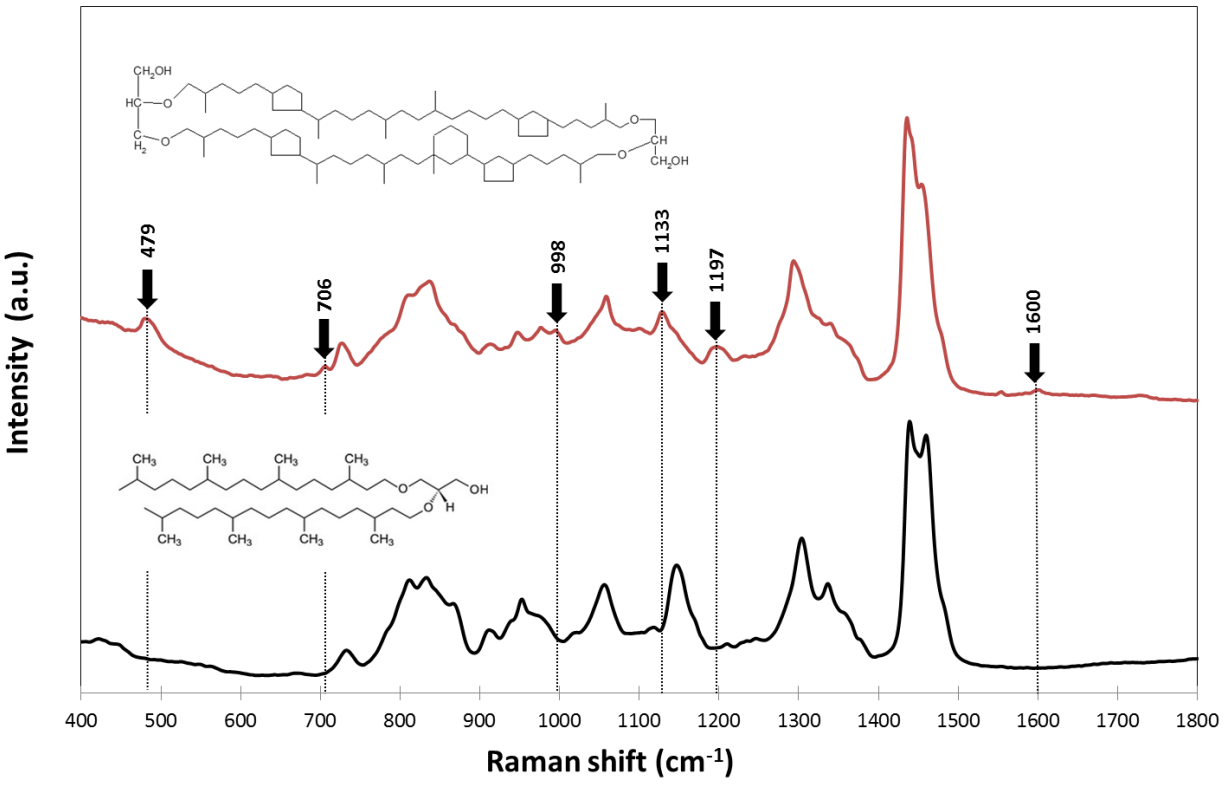

Figure 3.24. Comparison between the mean Raman spectra of the lipids 1,2-di-O-phytanyl-*sn*-glycerol (black) and crenarchaeol (red). The data were smoothed but not baselined.The wavenumbers of the most pronounced differences were indicated in the figure and the Raman spectra were acquired by Dr. Markus Schmid.

3.8 Raman spectra of cycloalkanes

Raman spectra of various cycloalkanes: cyclohexane (Fig. 3.25 A), cyclopentane (Fig. 3.25 B), an equimolar mix of cyclohexane and cyclopentane (Fig. 3.25 C), methylcyclohexane (Fig. 3.25 D) and methylcyclopentane (Fig. 3.25 E) were acquired (acquisition parameters can be seen in Table. 2.8). Cyclohexane showed certain characteristic peaks at the positions 425, 801, 1028, 1158, 1266, 1347, 1445 and 1466 cm^{-1}. On the other hand, cyclopentane had distinct Raman peaks at the positions 889, 1025, 1278, 1449 and 1481 cm^{-1}. Subsequently, the equimolar mix of both compounds (Fig. 3.25 C) showed all of those peaks. In addition, the peaks of cyclohexane were stronger in this equimolar mix, indicating that cyclohexane gives stronger Raman signals than cyclopentane. Furthermore, the methyl-group in cyclohexane (Fig. 3.25 D) and cyclopentane (Fig. 3.25 E) caused numerous additional weak peaks. Additionally, the most characteristic peak of cyclohexane at wavenumber 801 cm^{-1} was significantly shifted to the left (position 770 cm^{-1}) in the Raman spectrum of methylcyclohexane, whereas the methyl-group of cyclopentane did not cause a noteworthy shift. Taken together, these findings indicated that side groups can cause strong shifts in the Raman spectra of cycloalkanes, which makes an accurate theoretical prediction of the exact position of the cycloalkanes in the crenarchaeol spectrum very difficult.

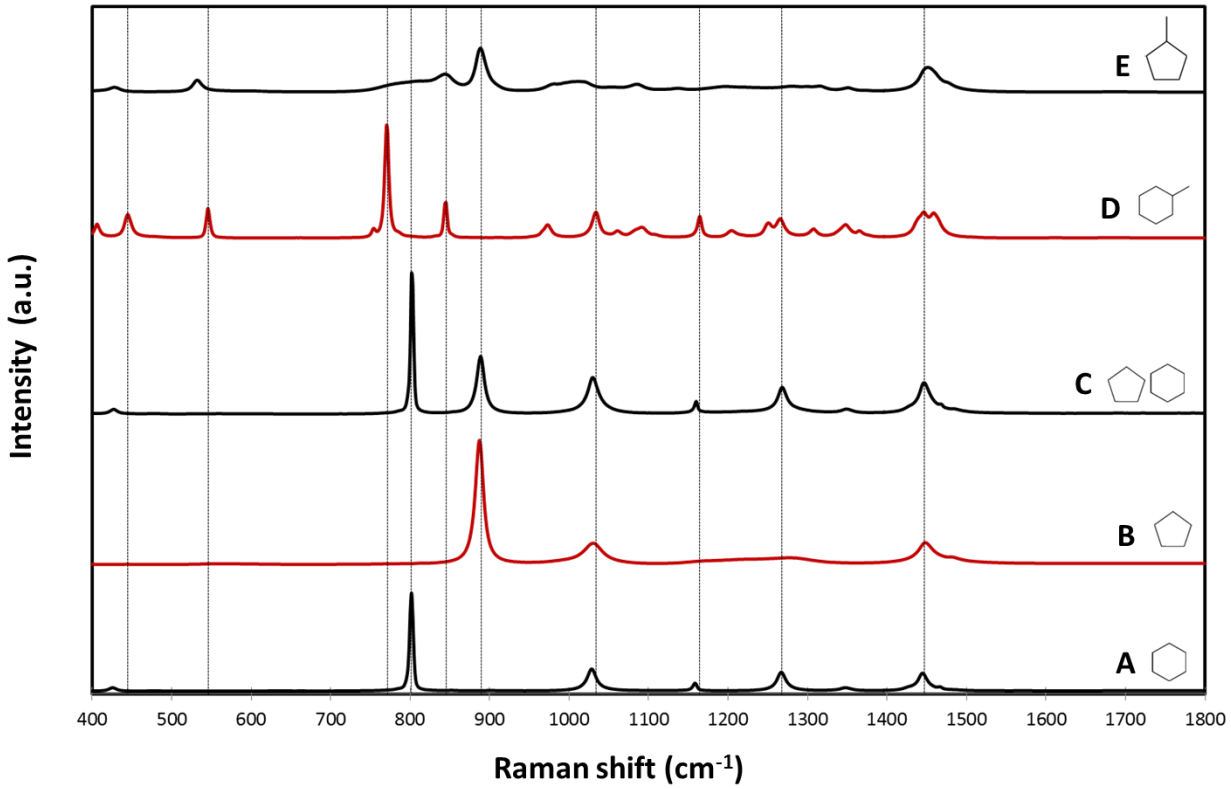

Figure 3.25. Comparison between the mean Raman spectra of cyclohexane ((A), n = 10), cyclopentane ((B), n = 5), equimolar mix of cyclopentane and cyclohexane ((C), n = 10), methylcyclohexane ((D), n = 3) and methylcyclopentane ((E), n = 5). For an easier comparison certain peaks were indicated in the figure. The data were smoothed and acquired by Dr. Markus Schmid.

4 Discussion

4.1 Macromolecules – lipids as discriminating factor

The major macromolecules of a cell mainly consist of nucleic acids, polysaccharides, proteins and lipids. Nucleic acids are nucleotide polymers and store the genetic information of a cell. There are four different nucleotides (plus uracil in RNAs) that are used for the biosynthesis of nucleic acids, which are of course the same for all prokaryotes. Polysaccharides consist of polymers of a huge variety of simple sugars and can occur as functional and structural components of a cell (e.g. glycoproteins) or as storage compounds (e.g. glycogen). Most proteins are synthesized from a series of up to 20 different amino acids (Poeggel, 2005; Engelking, 2010). A typical cell contains thousands of different proteins - each of them has a different structure and function. Proteins are usually not stored in prokaryotes like polysaccharides and lipids. Lipids serve for example as membrane components or storage form. Bloor (1943) suggested the following main subclasses: simple lipids (e.g. triglycerides, waxes), compound lipids (e.g. phospholipids, glycolipids), and derived lipids (e.g. various forms of fatty acids).

The reason why lipids and polysaccharides should be generally well suited as discrimination factors between different microbial strains via Raman microspectroscopy is because nucleic acids and proteins of all organisms consist of the same basic components, whereas lipids and polysaccharides (i) can have a rather unique chemical composition in certain prokaryotic strains and (ii) can occur in large amounts in a cell, which is why they can cause strong Raman peaks. This is also the reason why the membrane spanning lipid crenarchaeol is at least in theory a good Raman indicator for AOA.

4.2 Random Forest

In this study, RF was used to calculate similarities and differences between Raman spectra of different organisms. To this end, these data were used to build a clustering tree and to find peaks which were most characteristic for AOA spectra. Additionally, these data built the basis for the AOA prediction tool in order to cluster spectra of unknown origin in- or outside the AOA group. To calculate differences (distances) between numerous spectra, the Euclidean distances were chosen. Furthermore, the Ward's method was used to create clusters based on these distance values. These two methods will be discussed below in more detail and also some alternatives will be shown.

4.2.1 Euclidean Distance

Euclidean distances are the most popular distances for computing distances between objects of multi-

dimensional clusters. However, this type of distances is known to have one drawback. They do not take into consideration that some variables might be correlated (Mimmack et al., 2001). There are certain other distance measures, which are better suited for the analysis of correlated data (e.g. Mahalanobis distance (Mahalanobis, 1936)). The Mahalanobis distance is basically the Euclidean distance but normalized by the variance of each variable. This way, the covariance among variables is considered when distances are calculated (De Maesschalck et al., 2000; Imai et al., 2001; Cunderlik and Burn, 2006; Abril et al., 2011). Unfortunately, the Mahalanobis distance was not a selectable distance for RF. Nevertheless, Euclidean distances showed the best results over numerous other distance methods in many studies (Golub, 1999; Ramoni, 2002). However, other distance methods for clustering will be tested in the future once the CaF_2 filter (chapter 2.12.6) is working for the whole Raman reference spectra library.

4.2.2 Ward's method

The Ward's method is known to show a good discriminant efficiency (Karydis, 2009). However, it is also known that it tends to create small clusters with roughly the same small size (Statsoft(b)). In addition, this method is different from other algorithms, as it does not compute distances between clusters. It is forming clusters by maximizing the homogeneity within clusters (Sharma, 1996). In order to obtain more appropriate cluster groups it is suggested to first use hierarchical clustering to determine the number of clusters and then use an iterative partitioning (Punji and Stewart, 1983). For this diploma thesis only the hierarchical clustering method was applied. In order to improve the results of this study, aiming for the pipeline proposed by Punji and Stewart (1983) could be considered.

4.2.3 Alternative: Proximities for scaling

There is a built-in clustering method in the R package of Random Forest that is based on proximities between pairs of observations in order to visualize dis/-similarities (distances) in a given data set. The graphical output file is a multi-dimensional plot, called the multi-dimensional scaling (MDS) plot. The proximities (range between 1 and 0, where 1 are identical spectra) are a measure of similarity between two spectra. Proximity values are calculated as the number of trees for which any two spectra show a terminal node, normalized by dividing by the number of trees. Further, a matrix of these proximity values is created. This similarity matrix is used to calculate Euclidean distances between spectra and they are then projected/visualized into a lower-dimensional space by the use of a metric scaling algorithm (Izenman, 2008).
In summary, MDS attempts to put spectra in space with a distinct number of dimensions based on the

calculated distances (Statsoft(c)). This kind of clustering has been successfully performed on various data in the past (Svetnik et al., 2003; Shi et al., 2004). However, this clustering method was not performed in this diploma thesis because the highly popular Ward's method was already well-established in the scientific field. In addition, there was the possibility to change the distance measure from Euclidean distance to different ones to try to improve the results.

4.3 Challenges and issues during this study

4.3.1 Raman background spectrum of CaF_2 carrier slide

At the end of my diploma thesis an issue concerning the CaF_2 carrier slides, which were used for the acquisition of the Raman spectra, was discovered. In the beginning, this type of slide was chosen because it was generally assumed in the group that it only had a single Raman band at the position 321 cm^{-1} that was strong enough to affect Raman spectra of microbial cells (Fig. 4.1). Hence, this carrier slides would have been perfect for the acquisition of cell spectra, as I aimed for background-free cell spectra in the region between 400 and 3200 cm^{-1}.

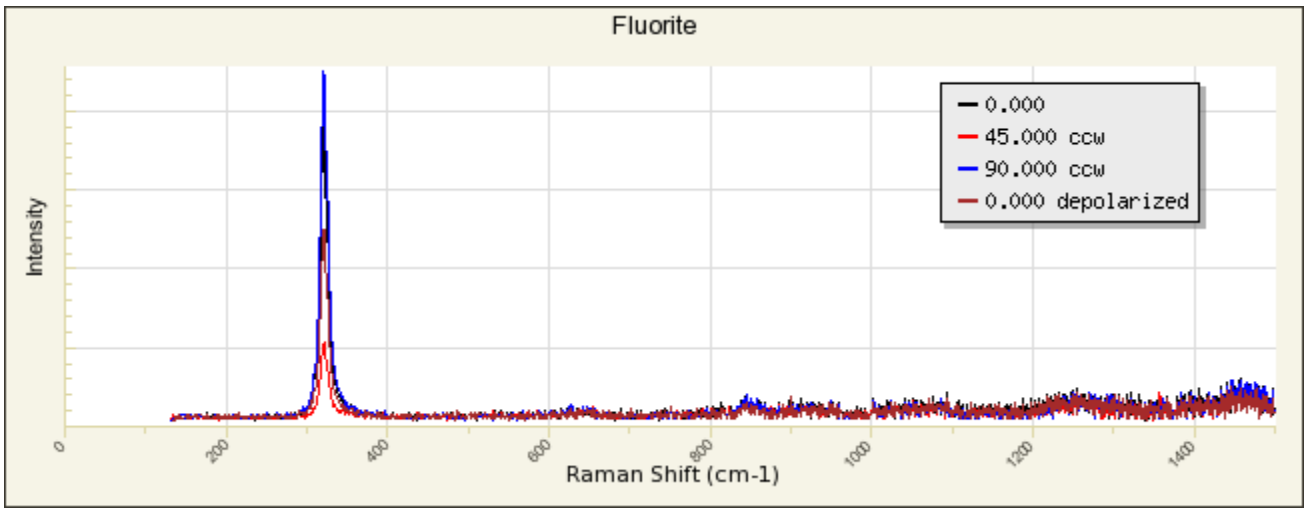

Figure 4.1. Raman spectrum of CaF_2 from the RRUF database; Locality: Hunan Province, China, Source: Eugene Schlepp; RRUF ID: R050045.2.

When carefully inspecting the Raman spectra of various reference organisms, I encountered a characteristic peak pattern in numerous spectra in the region between 400 and 1100 cm^{-1}. Interestingly, these spectra also showed a strong signal increase towards 400 cm^{-1}. This observation raised the hypothesis that CaF_2 might have a stronger influence on the spectra than previously assumed. Hence, I had a closer look at the spectrum of the CaF_2 slide. At the start of this study I recorded a Raman spectrum of this slide with an acquisition time of 10 sec (without a filter). Doing this, I achieved similar results (data not shown) compared to a CaF_2 spectrum that I found in the

literature (Fig. 4.1). Unfortunately, this short acquisition time was not in accordance to the ones which were used to record the cell spectra. Numerous spectra were recorded for more than 60 seconds acquisition time, some of them even without a laser intensity filter. Thus, these extended acquisition settings were applied to the CaF_2 slide (without any cells) and an unexpected more complex spectrum (Fig. 3.16) was obtained. Especially in the region between wavenumber 400 and 1100 cm^{-1} strong peaks, which resembled the peak pattern of cells that were obtained using long acquisition times (Fig. 3.17) were found. Therefore, I assumed that these spectra did not only contain information about the cell, but also background signal of the CaF_2 carrier slide. Unfortunately, all cell spectra were recorded at starting wavenumber 400 cm^{-1}. This is why it was very challenging to conclude how strong the CaF_2 impact to each individual spectrum was, because the carrier slide had its unique indicator peak at position 321 cm^{-1} (in the region around position 400 cm^{-1}, not only peaks of CaF_2, but also peaks caused by cellular material can be found). In addition, I discovered that the bands of CaF_2 became more prominent in relation to the bands of a cell spectrum the longer the acquisition time was chosen (data not shown). Therefore, I am quite certain that cells, which were exposed to a longer laser radiation and in addition, also small sized cells (the laser hits less cell material and less Raman photons of the CaF_2 material underneath the cell get absorbed by it), contained a stronger CaF_2 background signal. In addition, as mentioned in chapter 2.11,2, the misalignment of our laser position also caused an artificial error which could not be corrected by the choice of a smaller pinhole. In theory, when the laser is perfectly aligned which the position where the most photons are caught, and the laser hits a cell, a very small pinhole should exclude any background signal of a carrier slide. These issues led to an imprecise Random Forest calculation because the AOA were not only the smallest cells of this library, but they were also obtained using the most intense Raman acquisition parameters. In addition, a significantly stronger CaF_2 background also affected the subtraction by baselining because the algorithm cannot clearly distinguish between peaks caused by the cell, fluorescence or background. Thus, at this stage it can only be speculated if Random Forest was really selecting for AOA peaks or rather for small cells with a CaF_2 background signal. In summary, these findings have to be kept in mind while reading the discussions and conclusions about the cluster dendrograms (Fig. 3.1 – 3.6). Moreover, I can exclude that the Raman spectrum of CaF_2 derived from any contamination on the surface of the slide, since it was not only washed with EtOH and $CHCl_3$, but also various carrier slides were tested in parallel. On the other hand, I am confident that the CaF_2 spectrum maybe covered AOA specific, maybe even crenarchaeol peaks. After a proper CaF_2 subtraction, the results may become even better. Unfortunately, the time window of this diploma thesis was too small to analyze this issue any further. Hence, it will be addressed in the near future.

Nevertheless, at the end of this diploma thesis our group started to experiment with other materials as carrier slides and aluminum looked quite promising. There still was a background signal, but it

seemed rather low compared to the intensity of the cell spectrum. More tests with aluminum will be necessary to confirm its suitability as a useful Raman carrier slide.

4.3.2 Baseline parameters

In this diploma thesis a baselining approach (line segmented, 8th degree) was performed on all Raman cell spectra. This approach was necessary because fluorescence can be one of the major contributors to a cell spectrum. Raman raw spectra of the same pure culture do very often not look exactly the same. The peaks are at the same position, but the intensity values and the form of the spectrum curve sometimes varies considerably because background signals like the fluorescence underlie the spectrum. Unfortunately, this effect is rather unpredictable which results in a very challenging comparison of peak heights between different cell spectra, even more between different phyla. A baselining approach basically tries to remove fluorescence by subtracting underlying data by the calculation of a (in most cases) polynomial curve. This application is commonly used for processing of Raman spectra (Lieber and Mahadevan-Jansen, 2003). Furthermore, there are also methods to reject the development of a fluorescence background (e.g. by the use of FT-Raman spectroscopy (Chase, 1986), by the use of "*Kerr gated temporal rejection with shifted excitation Raman difference spectroscopy*" (Matousek et al., 2002), or by the use of the shifted-excitation Raman difference spectroscopy (SERDS) technique (da Silva Martins et al., 2010).

For this diploma thesis, a line-segmented, 8th degree baselining approach was chosen because it generated spectra which were aligned to the baseline throughout the whole range. In addition, this method was used before for various Raman approaches. Nevertheless, an "optimal" baselining approach is absolutely vital for the weighting process of RF, which means the baselining should be reasonable for all recorded spectra from different phyla. Nevertheless, it has to be considered that every baselining method is changing the data. This was shown in this study for CaF_2 (Fig. 3.15), certain microorganisms (Fig. 3.18) and crenarchaeol (Fig. 3.21). In order to choose the best suited baselining type and degree parameter it is not just necessary to understand how each method works (chapter 2.12.3), but also to run various empirical tests. Unfortunately, the time window in this diploma thesis was again too small, so it was not possible to dig much deeper into this topic. In conclusion, the chosen baselining type affected the outcome of this study and it should be mentioned that there is still room for optimization here. Especially after a proper CaF_2 subtraction of the Raman reference spectra library (chapter 4.3.1), which will significantly influence the subtraction process of the used baselining type.

4.3.3 Different normalization methods

Normalization is a crucial step when different Raman spectra are compared. The absolute values of the intensities will most likely vary because of different acquisition parameters (e.g. filter, acquisition time, pinhole size). However, if certain compounds of various cells need to be compared, spectra of varying intensity have to be processed to make them comparable. Accordingly, numerous normalization methods are performed in the scientific field, for instance: the mean normalization (Stone et al., 2002; Teh et al., 2008), the normalization by acquisition time (Orendorff et al., 2005), the normalization to the intensity of a specific peak that is shared by all compared spectra (Stone et al., 2002; Gniadecka et al., 2004; Ferraro et al., 2002), the normalization to the value 1 at the maximum intensity of a spectrum (Caspers et al., 2003) or the normalization by using the standard deviation to the values of the intensities (Baeten et al., 1998).

Three different normalization methods were used for this diploma thesis based on the height of most prominent phenylalanine peak, the median and mean value of the data set (see chapter 2.12.2). In addition, the effect of these normalization methods on the reproducibility of the Raman spectra of multiple cells from the same culture had been evaluated. The obtained results indicated a consistent outcome. Most single Raman spectra of a pure culture were comparable under each other after the application of one of these three normalization methods (data not shown). Furthermore, the effect of the subtraction by baselining plays a major role for all three normalization methods because putative peaks derived by cell compounds could be lost during a baselining approach. In this study, the baselining was performed before the normalization step. As a result, fluorescence, which can have a big impact on a spectrum of a biological sample, gets subtracted first. However, small signals can be lost due to this process. A second possibility would have been to normalize the spectra first and then to the baselining. In this case the fluorescence would have also been normalized which generates an artificial error because the amount of fluorescence is not the same for every spectrum of the library.

The idea behind the Phe normalization is that basically all microorganisms should have a comparable ratio between the amount of the basic amino acid Phe and other cell compounds. This approach seemed plausible but this method also has a noteworthy drawback. If the height of the phenylalanine peak is influenced by any other cell/storage or pigment peaks (e.g. carotenoids), a source of error is created that will affect the RF-weighting. However, in this diploma thesis reasonable results could still be achieved with the Phe normalization approach, but it did not generate the best ones out of the three methods. The OOB error estimate of this method was 12.5 %, which was the highest percentage of all three used methods. The median normalization was a reasonable method because it is relatively resistant to outlier peaks when a mean out of numerous single spectra is created. However, outlier peaks can originate for example from a different growth stage and must not be artifacts or storage compounds which were not of interest for goal of the clustering. In the case of a different growth

stage, spectra information would be lost if the majority of the single spectra were in another level of their life cycle. Nevertheless, the OOB error estimate of the mean normalization method (with AOA+ weighting) was 5 %. The same percentage was achieved by using the mean normalization. This method is more prone to outlier peaks compared to the median normalization. On the other hand, therefore different growth stages are potentially better considered when mean spectra are used.

In conclusion, the median and mean normalization generated the best results, when the AOA+ weighting was used for the classification. The mean normalization is more prone to outlier peaks than the median normalization (e.g. storage compounds in only some of the spectra) when mean spectra are used. The Phe normalization can be influenced by other chemical compounds which generate bands at the wavenumber position 1004 cm^{-1}. Furthermore, also outlier peaks contribute to the mean spectrum of Phe normalized data sets, just like discussed for the mean normalization. To improve the results even further, it might be worth considering individual and not mean spectra for each reference species. This would probably improve the precision of the clustering tree and the discrimination between AOA and non-AOA because it cannot be totally excluded that some spectra might have been from a contaminant. Especially for the AOA, this would cause a tremendous influence, since the RF looks for peaks which are specific for these type of archaea. Furthermore, the useage of single rather than mean spectra would create a much easier possibility in finding those contaminant spectra in the clustering tree. Those could then be eliminated from the data set.

4.3.4 Storage compounds

Many different storage compounds can be found in different prokaryotic microorganisms, for example: polyhydroxyalkanoates (PHA), polyglucans, extracellular polysaccharides, lipids, polyphosphates, sulphur granules or even triglycerols (TAG) and wax esters (WEs) (Shively, 1974; Dawes, 1992; Alvarez et al., 1997; Lee, 2000; Steinbüchel, 2001; Alvarez and Steinbüchel, 2002; Hezayen et al., 2002; Bredemeier et al., 2003; Wältermann and Steinbüchel, 2005; De Gelder et al., 2008). In addition, also pigments (e.g. carotenoids) do occur in a lot of prokaryotes and can cause an enormous influence on the Raman spectrum of the cell containing them (Hayashi et al., 1989; Marshall et al., 2007; Maquelin et al., 2009; Willemse-Erix et al., 2009). PHB for example can accumulate to approximately 80% of dry mass if the bacterial cell lives under highly stressed conditions or is grown in a rich medium (Luzier, 1992; Kim et al., 1994; Wong and Lee, 1998; York et al., 2003; Thuoc, 2009). In some cases, storage compounds can also be beneficial for the analysis by Raman, for example if only specific lineages of a phylogenetic group are able to build them, Raman spectroscopy could easily distinguish them from each other.

This fact shows that the knowledge about these compounds and their influence on the Raman spectra of cells is crucial for the goals of this diploma thesis because these storage compounds and pigments

often lead to very strong Raman peaks and they can overlay the complete cell spectrum (Fig. 3.11). Hence, this makes a meaningful RF weighted cluster dendrogram nearly impossible because all the prokaryotes with the same storage compounds or pigments would cluster together. The application of storage compound/pigment filters is therefore absolutely necessary. This study mainly concentrated on the removal of PHB signals from microbial Raman spectra (chapter 2.12.6) as this storage compound was frequently found in the analyzed reference strains reflecting its widespread distribution. Hence, there is still some work to do in the future in that area to subtract also the peaks from the remaining compounds and thus, further improve the RF weighting and clustering.

4.3.4.1 Polyhydroxybutyrate

Polyhydroxybutyrate (PHB) belongs to the class of polyhydroxyalkanoates (PHAs), and is one of the most common intracellular polymers, which is also involved in the enhanced biological phosphorus removal process in waste water treatment plants (Majed and Gu, 2010). This compound is built and stored in big granules when the environmental conditions become unfavorable for the growth requirements of the cell (e.g. high carbon/nitrogen ratio) (Misra et al., 2004; De Gelder et al., 2008).
PHB consists of just one type of monomer called 3-hydroxybutyrate (3HB). Nevertheless, more than a hundred other types of monomers have been shown to be present in microbial polyesters (Steinbüchel and Valentin, 1995; Steinbüchel and Doi, 2001; Jendrossek and Handrick, 2002).
In addition, PHA exists in two forms – amorphous and crystalline. PHA that accumulates inside the cell, also called the native form, is in the amorphous state. The proteins and phospholipids of the surface layer are sensitive to chemical and physical stress, which leads to a changing of the polymer structure into the crystalline state when it is stressed and damaged. Crystalline PHA also occurs extracellular, when the polymer is released by a lysing cell. The ratio of crystalline to amorphous PHA in such polymers is about 50:50 to 60:40 (Jendrossek and Handrick, 2002).
The suggestion that not-damaged intracellular PHB granules are only amorphous and not crystalline was proven by certain studies (Horowitz and Sanders, 1995; Merrick et al., 1999; Jarute et al., 2004). Additionally, the strong peak at wavenumber 1740 cm^{-1} of certain PHB spectra was assigned to a C=O stretch and further, suggested to be shifted to position 1726 cm^{-1} when PHB is crystalline instead of amorphous (Murakami et al., 2007).
The position of this shifted peak was assigned to 1725 cm^{-1} in certain other publications (Jarute et al., 2004; Izumi and Temperini, 2010). Jarute et al. (2004) postulated that this shift in the crystalline form of PHB is a result of an intensified hydrogen bonding.
Furthermore, it was postulated that besides this peak, also a band at position 1731 cm^{-1} can be observed in the Raman spectra of crystalline PHB. In addition, the broadness of the band at wavenumber 1725 cm^{-1} was shown to be altered by the degree of crystallinity of the sample (Izumi

and Temperini, 2010).

Murakami et al. (2007) also revealed additional bands which are located at the positions 434, 1259, 1402 and 1444 cm^{-1} that are characteristic for the crystalline state of PHB.

Furthermore, PHB also exists in two types of molecular conformations, the so called helical alpha- and planar beta-form (Murakami et al., 2007; Chaturvedi et al., 2009).

Murakami et al. (2007) revealed that there are specific peaks which only occur in the beta-form of PHB: 966, 935, 908, 858 and 1735 cm^{-1}. There does not seem to be peaks specific for the alpha conformation.

In the Raman spectra library of this diploma thesis, several different PHA contributions could be observed. Amorphous PHB (indicator peak: 1740 cm^{-1}) could be discovered in the Raman spectra of *Methylocystis rosea* (1734 cm^{-1}; Fig. 8.5) and *Sarcina ventr*iculi (1734 cm^{-1}; Fig. 8.27), whereas crystalline PHB (main indicator peaks: 1725 and 434 cm^{-1}) could be observed in *Burkholeria cepacia* (1727 and 433 cm^{-1}; Fig. 8.7) and *Nitrosotenius uzonensis* (1731 and 433 cm^{-1}; Fig. 8.34). I assume that spectra with strong crystalline PHB contribution (*B. cepacia*, *N. uzonensis*) could be observed because of the fixation of the culture, which damaged the polymer. Minor amounts of crystallinity and hence, small modifications in the spectra compared to the natural PHB can probably be explained by additional monomers in the polyesters of the cells storage compound as described by Izumi and Temperini (2010). Last but not least, no Raman peaks specific for the beta-conformation of PHB could be found within the spectra library of this study.

4.3.4.2 Polyhydroxybutyrate filter

Most prokaryotes contain either the PHA sub-type called PHB or the co-complex PHB-PHV. Hence, a filter for the subtraction of PHB was created (chapter 2.12.6). For this filter, a difference spectrum of two spectra of *S. ventriculi* (with and without the storage of PHB) (Fig. 3.12) was used. At the time the PHB filter was created and used, neither did the group know how complex the Raman spectrum of this storage compound is, nor what is causing the shifts of certain peaks in different bacteria (chapter 4.3.4.1).

The reason why a difference spectrum of *S. ventriculi* was used for the PHB filter in order to subtract any PHB influence in numerous species was that the group initially thought that Raman spectra of PHB, originating from different phylogenetic groups, do not vary much and in addition, that a spectrum of bacterial PHB is always different than synthetic PHB. However, bacterial PHAs are often not only consisting just out of one type of monomer and in addition, there are other factors like the conformation or different states (amorphous, crystalline).

The PHB filter used for this study is actually a filter for amorphous α-PHB. Indeed, it is not known if *S. ventriculi* accumulates pure α-PHB. Since it should have its indicator peak at 1740 cm^{-1} (*S.*

ventriculi: 1734 cm^{-1}), it is very likely that this is not the case. Hence, this filter should better not be used as a general PHB filter, especially not for Raman cell spectra that contain crystalline PHB (*B. cepacia, N. uzonensis*). Nevertheless, the filter was created as follows: The two Raman cell spectra of *S. ventriculi* (with and without PHB) were aligned to the intensity of the prominent phenylalanine peak at position 1004 cm^{-1}. This emanated from the assumption that different cells from a pure culture should contain a comparable amount of this amino acid. Moreover, these two aligned spectra were subtracted from each other and the resulting spectrum was compared with various PHB spectra from the literature (De Gelder et al., 2008; Ciobotă et al., 2010). It became obvious that the difference spectrum did not just consist of PHB but also of an unknown contribution in the region between wavenumbers 1520 and 1670 cm^{-1} (data not shown). Hence, this contribution from unknown origin was cut out, resulting in a flat line in the spectrum (Fig. 3.12). The PHB filter was then applied to all cell spectra of the reference library that seemed to contain some sort of PHA. The wavenumber region 1725 – 1740 cm^{-1} appeared to be a good indicator peak for monitoring the storage of PHA because no contribution from other cell compounds could be observed in that region in the Raman reference library. After the end of the diploma thesis the group realized the complexity of the PHB Raman spectrum and further realized that the used filter did not work properly for all microorganisms, especially not the ones with crystalline PHB, because the alignment of the amorphous PHB filter was based on the region 1725 – 1740 cm^{-1}. Consequently, this resulted in the subtraction of some false bands in certain cell spectra because the whole filter spectrum was aligned to this one shifting peak, whereas not all peaks of PHB do shift the same way as the band of this one specific region.

Self-evidently, the filter performed well when applied to the PHB-containing Raman spectrum of *S. ventriculi* (spectra with subtracted and without PHB storage clustered together in all normalization and weighting approaches (Fig. 3.1 – 3.6)) because the filter was based on a *S. ventriculi* difference spectrum. On the contrary, the filter did not operate perfectly for the other microorganisms which contained PHB, e.g. *B. cepacia* (Fig. 3.1 – 3.6). The mean spectra of *B. cepacia* (with subtracted and without PHB contribution) were assigned to different sub clusters after all normalization and weighting methods were applied (Fig. 3.1 – 3.6). It is very likely that *S. ventriculi* and B. *cepacia* stored a different sub type of PHA. Furthermore, I assume that PHA is not the only distinguishing factor between cells with and without this storage compound. Prokaryotic cells accumulate polymeric materials for a reason, so it could be that these cells were in a different growth stage or encountered stress, which could for example have resulted in an alteration in their membrane lipid composition, which is known for certain microorganisms (Ray et al., 1971; Darveau et al., 1980; Hazel and Williams, 1990; Nichols et al., 2000). In the difference spectrum of *S. ventriculi*, small additional peaks could be observed, which did not show up in various PHB spectra of the literature (De Gelder et al., 2008; Ciobotă et al., 2010). These peaks were excluded from the PHB filter because they were

most probably not similar for other microorganisms of different phyla since they did not originate from PHB. This supports the hypothesis of the formation of other Raman active compounds next to the PHB signals under certain conditions. In conclusion, the filter was not working as intended, which will also have affected the clustering of *N. uzonensis* and *M. rosea*. While the latter one also stored amorphous PHB, the effect will be less severe, but for *N. uzonensis* which had a rather strong contribution of crystalline PHB the subtraction generated an artificial error. These mistakes also affected the quality of the RF weighting of AOA spectra in general and hence, also the results of the AOA cluster prediction tool.

4.3.4.3 Validity of filter application and spectra processing

Although the used PHB filter (chapter 2.12.6) did not seem to work perfectly (chapter 4.3.4.2), the application of a storage compound filter appears to be both valid and necessary. Perfect results over a variety of microorganisms will not be accomplished with this kind of application because before the storage compound can be subtracted, the spectra have to be baselined due to fluorescence and normalized due to different reasons (e.g. acquisition parameters, cell size). Baselining has shown to be a putative source of error (chapter 4.3.2.). However, the results after baseline-subtraction should still be reasonable enough for further RF weighting. One way to improve the filter-approach would be to acquire and compare the Raman spectra of certain PHAs and then create various filters out of it instead of just one, which should produce a more robust data. In addition, also the possibility of generating two data sets for clustering should be mentioned. One data set including only cell spectra with and one without storage compounds like PHB would deny the need for a filter and yet, specific AOA bands could be discovered. Maybe there are characteristic sub-types of storage compounds for AOA which can be overseen when using a general PHA filter.
Furthermore, the type of baselining plays a major role, not just for a reasonable comparison of different spectra, but also for an effective and valid subtraction of storage compounds. The varying and rather unpredictable contribution of fluorescence to the full spectrum makes baselining very challenging. Accordingly, it even becomes more complicated if not only fluorescence but also cell derived data was subtracted by baselining.

4.4 AOA cluster enigmas

4.4.1 Iso-diabolic acid

As seen in the cluster dendrograms (Fig. 3.1 – 3.6), several bacteria did cluster together with the AOA. I discovered the influence of the CaF_2 background signal on this clustering (chapter 4.3.1), but

there might also be other contributors. In most cases, the bacteria *A. capsulatum, E. aggregans, E. modestus, F. pennivorans, T. africanus* and *T. maritima* were assigned to the AOA cluster and they all contain a very characteristic membrane spanning lipid called 13,16-dimethyl octacosanedioic acid (iso-diabolic acid). It shows a structural similarity with crenarchaeol (Fig. 4.2), as both have long methylated alkyl chains. Recently, even a biosynthetic relationship between these two lipids was suggested (Damsté et al., 2011). In addition, iso-diabolic acid was previously thought to be restricted to certain thermophilic *Thermoanaerobacter* species (Jung et al., 1994; Balk et al., 2009). However, it was recently also found in other bacteria, especially in the subclasses 1 and 3 of *Acidobacteria*, where this lipid accounts for up to 43% of the total fatty acids (Damsté et al., 2011).

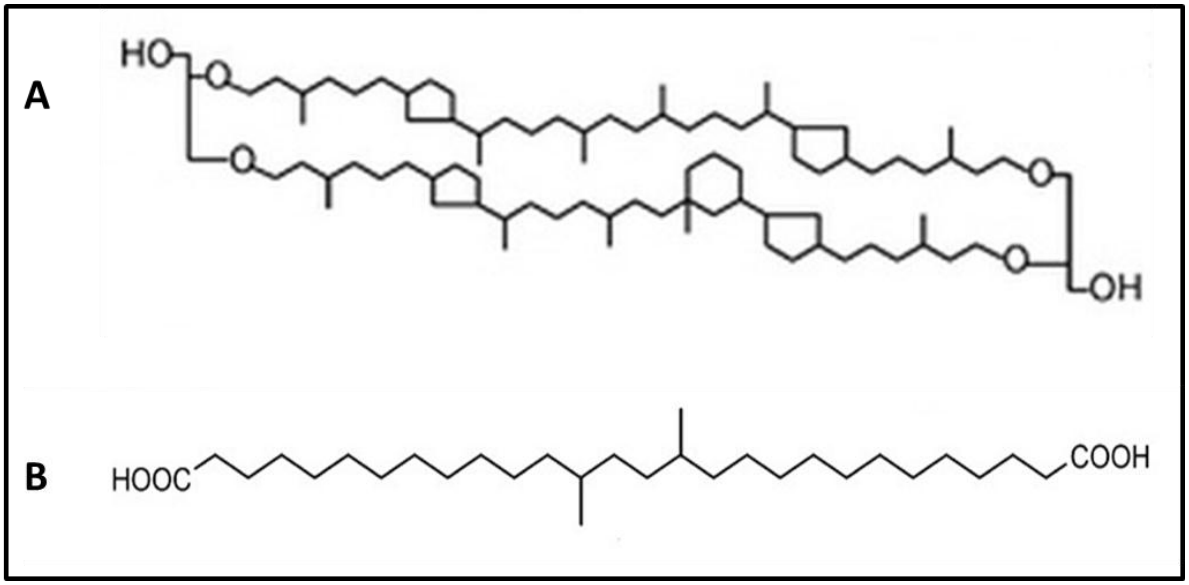

Figure 4.2. Difference between crenarchaeol (A) and iso-diabolic acid (B)

All acquired Raman cell spectra, which contain iso-diabolic acid were assigned to the AOA in one or another cluster dendrogram depending on the type of normalization or RF weighting (Fig. 3.1 – 3.6), so it is possible that the Raman bands of long methylated alkyl chains resulted in this clustering.

4.4.2 *Sulfolobus* species

Three *Crenarchaeota* (*S. acidocaldarius, S. islandicus and S. tokodaii*) were analyzed during this study and none of them clustered together with one of the AOA (Fig. 3.1 – 3.6). This was rather unexpected since it is known that these *Crenarchaeota* also contain GDGT (Damsté, 2002; Ellen et al., 2008) and this lipid was thought to be a discriminative factor for the cluster dendrogram because of its structural similarity with crenarchaeol. Moreover, all *Acidobacteria* (*A. capsulatum, E. aggregans and E. modestus*) of the Raman reference library were assigned to the AOA cluster depending on the cluster parameters (Fig. 3.1 – 3.6). Hence, I hypothesized that iso-diabolic acid

could have a major influence to this clustering (chapter 4.6), since also the other iso-diabolic acid containing microorganisms were assigned to the AOA cluster after application of certain normalization and weighting methods (Fig. 3.1 – 3.6) Therefore, long methylated alkyl chains were thought to be a major discriminating factor for the RF weighting. However, also the *Sulfolobus* species of the Raman reference library contain GDGT. Consequently, it seemed that were also other factors which affected the clustering of AOA and iso-diabolic acid containing prokaryotes. One possibility was the influence of the CaF_2 background signal from the carrier slide, which was used for the acquisition of the cell spectra. The longer the acquisition time and the weaker the used laser intensity, the stronger became the influence of the CaF_2 background signal in relation to the cell spectra peaks (data not shown). The cell spectra of *S. acidocaldarius* and *S. tokodaii* were recorded with 28 % laser intensity (filter 0.6), an acquisition time of 25 sec and a pinhole size of 600 µm, whereas the cells of *S. islandicus* were recorded with the same settings but 30 sec of acquisition time. This resulted in a weaker CaF_2 background signal compared to the spectra of iso-diabolic acid containing organisms (e.g. *A. capsulatum*: 45 sec, filter 0.6, pinhole 600 µm; *E. aggregans*: 35 sec, filter 0.6, pinhole 600 µm) and the AOA (e.g. *N. uzonensis*: 60 sec, no filter, pinhole 600 µm; *N. maritimus*: 120 sec, no filter, pinhole 500 µm) (Tab. 2.8). In short, the influence of GDGT to the RF weighting can only be validated by the removal of the CaF_2 background signal from the whole data set. I still think that the slide contribution is not the only relevant contributor to this clustering, because there were also other bacteria with an intense laser acquisition time and they were not assigned to the AOA cluster (e.g. *Acetonema longum* (45 sec / laser intensity 58 % (filter 0.3)), *Burkholderia cepacia* (40 sec / 58 % laser intensity (filter 0.3)).

4.4.3 *Desulfovibrio oxyclinae*

Surprisingly, the sulfate-reducing bacterium *Desulfovibrio oxyclinae* (DSM 11498) (Krekeler et al., 1997) was assigned to the ammonia oxidizing archaeum *N. viennensis* in all cluster dendrograms (Fig. 3.1 – 3.6). However, the comparison of their raw (Fig. 3.14) with the baselined spectra (Fig. 3.15) showed the significant influence of the applied baselining method. There are several aspects which can be improved. First of all, the CaF_2 background spectrum needs to be subtracted from all spectra of the reference library. Second of all, different kinds of baselining methods have to be evaluated to achieve the best result possible. And last but not least, a Raman spectrum of *D. oxyclinae* should be acquired using a different carrier slide material, to validate the CaF_2 subtraction success. Taken together, there were certain similarities between the spectra of *N. viennensis* and *D. oxyclinae*, but without the knowledge about the content of fluorescence and CaF_2 background, it was hard to predict whether *D. oxyclinae* had in fact a very similar spectrum which derived e.g. from similar lipids or not.

4.4.4 *Methylocystis rosea*

Besides *D. oxyclinae* (chapter 4.4.3), also the mean spectrum of *Methylocystis rosea* (without the storage of PHB) (Fig. 8.4) was assigned to the AOA cluster using different normalization and weighting methods (Fig. 3.1, 3.2, 3.3, 3.5 and 3.6). In contrast, the spectrum of *Methylocystis rosea* with PHB storage (Fig. 8.5) was not (Fig. 3.1 – 3.6). This was surprising because the PHB filter script (chapter 2.12.6) was performed on this mean spectrum. In addition, *M. rosea* and *S. ventriculi* had their most characteristic PHB bands on the same Raman shift positions. Therefore, it can be assumed that they contained a similar type of PHA. Consequently, the subtraction was valid and a wrongly applied PHB filter can be excluded as the reason for the assignment of *M. rosea* to the AOA cluster. Furthermore, I discovered that the Raman spectrum of *M. rosea* (without PHB storage) featured some very strong Raman bands (1154 and 1511 cm^{-1}), which could be assigned to a pigment subclass called carotenoids.

The most important carotenoids for many prokaryotes are beta-carotene (Fig. 4.3) and bacterioruberin. It was shown that carotenoids have their most pronounced Raman peaks at the positions around 1000, 1152 and 1505 cm^{-1} (Marshall et al., 2007; Fendrihan et al., 2009). The mean spectrum of *M. rosea* with PHB storage, which was assigned outside the AOA cluster did not contain the peaks of this pigment. Consequently, this could explain why this spectrum was not assigned to the AOA cluster after PHB subtraction. The overlaying Raman spectrum of carotenoids probably led to the assignment of *M. rosea* to the AOA cluster. Another possibility would be that cells which contain PHB are less penetrable for Raman photons from the CaF_2 material underneath the measured cells, thus less CaF_2 background signals are in those spectra and hence, they do not cluster. Furthermore, the Raman peak of carotenoids which is centered at wavenumber 1000 cm^{-1} is a rather broad peak and thus, it added up to the one of phenylalanine (1004 cm^{-1}). Therefore, these pigments have a strong influence to any clustering approach when the Phe-normalization method (chapter 2.12.2.b) is performed. Consequently, in order to perform this type of normalization, a carotenoid filter has to be applied first to the Raman reference library.

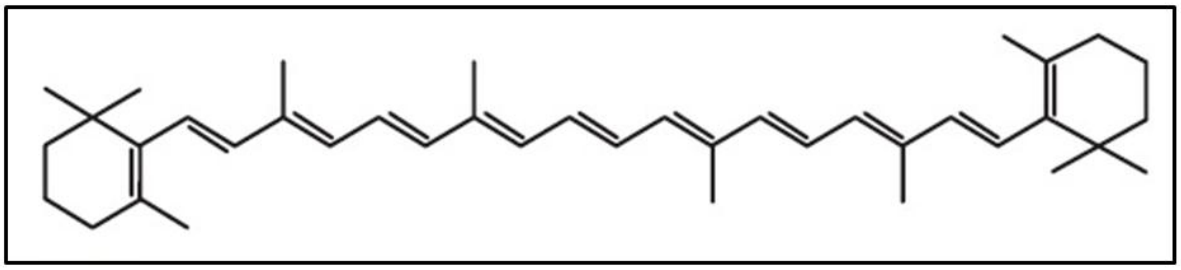

Figure 4.3. Chemical structure of beta-carotene (Marshall et al., 2007)

4.5 Peak assignment of crenarchaeol

Certain noteworthy differences and similarities between the Raman spectra of crenarchaeol and 1,2-di-O-phytanyl-sn-glycerol could be observed (Fig. 3.24). The differences were indicated by the following peaks: 479 (w), 706 (vs), 998 (sh), 1133 (m), 1197 (m) and 1600 cm^{-1} (vw). Consequently, those bands are putative indicators for either cyclohexane or –pentane of crenarchaeol because besides the ether linkages and the two types of cycloalkanes, these two lipids are very much alike. The following analysis is based on the baselined spectrum of crenarchaeol (Fig. 3.19) and the results of the irAnalyze software (Tab. 3.1).

For the Raman peak of crenarchaeol at wavenumber 479 cm^{-1} a hydroxyl compound was suggested, which can be found in the pure crenarchaeol. The very small peak at position 706 cm^{-1} was subtracted by the baselining approach and therefore not included in this analysis which is based on a baselined spectrum. However, the medium peak at wavenumber 998 cm^{-1} (a shoulder peak) was suggested to derive from a large ring or long alkyl chain. Subsequently, this band probably originated from the cyclohexane or –pentane rings, as this peak was not detectable in 1,2-di-O-phytanyl-sn-glycerol, which also has large alkyl chains (Fig. 3.21). Furthermore, the medium peak at position 1133 cm^{-1} was proposed to derive from an aliphatic ether. Crenarchaeol has four aliphatic ether linkages in its chemical structure (Fig. 1.2) whereas 1,2-di-O-phytanyl-sn-glycerol has two of them in its structure. Therefore, this could be a wrong peak assignment for crenarchaeol. Moreover, the software suggests an unsaturated hydrocarbon to be responsible for the band at wavenumber 1197 cm^{-1}, which was disproven by the chemical structure of crenarchaeol. Finally, there was no suggestion for the very small peak at position 1600 cm^{-1}. It might be possible that this small peak derived from a minor contamination of the sample. The software exploits a comprehensive database of many different chemical molecules and linkages, but a peak can derive from many different origins or even be generated by numerous contributors. This was also true for putative cyclohexane and –pentane bands, because the crenarchaeol peaks at the wavenumbers 978 and 1058 cm^{-1} were suggested by the irAnalyze program to originate from large rings and/or long chain alkyls. Theoretically, the peaks in these regions could be either from one of these bondings or from both of

them, which makes a peak assignment rather difficult.

The similarities between crenarchaeol and 1,2-di-O-phytanyl-sn-glycerol (Fig. 3.21) were indicated by the following peaks: 728 (m), 814 (s), 838 (vs), 880 (sh), 913 (m), 950 (sh), 978 (sh), 1058 (sh), 1104 (m), 1301 (vs) and 1342 cm^{-1}. Several bands which might be the result of long (branched) alkyl groups like: 728, 814, 838, 880, 913, 950, 1342, 1436 and 1456 cm^{-1} (Tab. 3.1) could be detected. In addition, the crenarchaeol peaks at wavenumber 1104 and 1301 cm^{-1} were suggested to be promising additional candidates for aliphatic ether bands. Pure cyclohexane and –pentane had strong Raman peaks at the wavenumber positions 801, 889, 1025, 1266 and 1449 cm^{-1} (Fig. 3.22). These bands could not be seen in the Raman spectrum of crenarchaeol. However, it was discovered that even simple methyl side groups can result in a significant Raman shift of strong peaks in cycloalkanes (Fig. 3.19). Furthermore, there is the slight possibility that the Raman spectrum of crenarchaeol might be shifted inside the cell compared to the extracted pure lipid form which the spectra were taken from in this diploma thesis.

4.6 Detection of crenarchaeol in whole-cell Raman spectra

Based on the preliminary results (Fig. 3.1 – 3.6) it is not possible to discriminate the AOA spectra from those of all other prokaryotic organisms just based on crenarchaeol (= AOA-only weighting in RF). The findings of this study (chapter 4.3.1) indicated that the CaF_2 carrier slide caused a strong influence to the acquired Raman cell spectra that can currently not be removed without a re-recording of all reference spectra. Furthermore, the impact of the CaF_2 background signal increased compared to the intensity of a cell spectrum if longer acquisition times and/or a lower filter were chosen (data not shown). In addition, the Raman peaks of cyclohexane and cyclopentane from crenarchaeol could be much weaker compared to the other bands of cell compounds than expected. Nevertheless, keeping in mind the location of the CaF_2 peaks, it might very well be possible that the CaF_2 background spectrum covered characteristic crenarchaeol peaks. Furthermore, the most pronounced Raman peaks of crenarchaeol seemed to derive from long methylated alkyl groups, which can also be found in other lipids (Fig. 3.21). In the case of crenarchaeol it was already possible to subtract the CaF_2 background signal (Fig. 3.17) since the spectrum was recorded starting at the Raman shift position 200 cm^{-1} and CaF_2 had its indicator peak at position 321 cm^{-1}. On the contrary, the Raman reference spectra library was recorded from beginning at wavenumber 400 cm^{-1}, which made a subtraction much more difficult. Either all Raman spectra have to be re-recorded to include the major CaF_2 peak or a different carrier slide has to be chosen. Ultimately, I will try to subtract the CaF_2 background from all acquired Raman cell spectra to discover if crenarchaeol is in fact the discriminating factor of AOA against other archaea and bacteria. The usage of the AOA+ weighting led to a tighter clustering of the AOA and a lower OOB error estimate (Fig. 3.4 and 3.6). Hence, at

the moment it seems that the AOA can be discriminated against most other prokaryotes, with the exception of organisms that contain iso-diabolic acid or certain pigments.

4.7 Arctic AOA enrichments

In order to test the reliability of the AOA prediction tool and the peak areas which were chosen by RF to be specific for AOA, AOA spectra from outside the library have to be analyzed. Our group had access to two arctic AOA enrichments from Dr. Christa Schleper. The presence of AOA was confirmed by qPCR and CARD-FISH. Raman spectra of randomly chosen cells should then feature a similar AOA-content like the qPCR suggested.

4.7.1 Quality of acquired Raman spectra

The signal to noise ratio of the acquired spectra of the arctic AOA enrichments, SV8-6 (Fig. 8.41 – 8.50) and SV9-19 (Fig. 8.51 – 8.70), was relatively low compared to the spectra of the reference library (Fig. 8.1 – 8.40). The Raman spectra of these AOA enrichment cultures were difficult to measure, because many cells showed unexpected photo damage during spectrum acquisition. It would have been possible to extend the acquisition time for all recorded cells, but this would have resulted in loss of many more cell spectra. Since I chose cells at random to acquire Raman spectra from, I had to use low intensity laser settings to avoid losing phylogenetic groups of cells which were not able to withstand stronger parameters. I would have lost the natural diversity of cells inside the enrichment cultures which would have made a comparison to the qPCR data very difficult, so I decided to choose low laser intensity settings in order to acquire a Raman spectrum from every selected cell even though the resulting SNR was not favorable for the AOA prediction tool. Hence, the final cell spectra of SV8-6 and SV9-19 lacked a good signal to noise ratio. It is necessary to re-record the cell spectra of these enrichment cultures in liquid inside a capillary to avoid the issue of photo damage. Unfortunately, this approach was not yet fully established by the end of this thesis. This approach could provide spectra of higher quality which would have a positive effect on the quality of the AOA cluster prediction of these cells.

4.7.2 Morphology

The morphology of the AOA used to generate the reference spectra library showed that they were in general roundish shaped and approximately 0.7 – 1.0 µm in size (data not shown). CARD-FISH results of the arctic AOA enrichment SV8-6 (Fig. 3.D C) suggested a similar morphology. Bright field images of the AOA enrichment cells of SV8-6 and SV9-19 (Fig. 3.F A – Fig. 3.F L) which were

positively assigned to the AOA cluster (mean normalization, AOA+ weighting) using the AOA prediction script also featured cells which were small in size, about 0.7 – 1.0 µm, but not uniform in morphology. Some were roundish shaped (Fig. 3.F A, -B, -F, -G, -I, -K and -L), whereas others had a rod-shaped structure (Fig. 3.F C, -D, -E, -H and –J). Because of the low quality of their Raman spectra, the currently unpredictable CaF_2 influence, the issues concerning the choice of baselining type and issues like storage compounds, the observed variety of morphologies might suggest that the results were adversely affected by an inaccurate RF weighted data set.

4.7.3 Significance of cluster assignment

The Raman reference library data set was artificially influenced by the CaF_2 background signal (chapter 3.4.2). It is hard to predict if the AOA clustering will become worse or even better after the subtraction of these background signals. Nevertheless, the cluster percentages of randomly picked arctic AOA enrichment cells (Fig. 3.6) looked promising compared to the qPCR results (chapter 2.9), but also the low signal to noise ratio of the spectra of the arctic AOA enrichment cultures (chapter 4.7.1) played a major role concerning the probabilities of clustering because the spectra of the Raman reference library were of higher quality. First, a proper CaF_2 filter has to be implemented to the Raman reference library data set and then the arctic AOA have to be re-recorded using less-damaging methods. Finally, the AOA cluster prediction tool has to calculate new probabilities for the assignment of the AOA enrichment cells.

5 Summary

Nitrification is the oxidation of ammonia to nitrite and further to nitrate. It is one of the key steps in the global nitrogen cycle and also a very important reaction for industrial agriculture and in wastewater treatment plants. Hence, it is fundamental to understand the key players of ammonia oxidation, which were thought for a century to be exclusively ammonia oxidizing bacteria (AOB). Recently, ammonia oxidizing archaea (AOA) were discovered to occupy a major role in the global nitrification, as they outnumber AOB in numerous habitats. The ammonia monooxygenase (Amo) has been used as a phylogenetic marker for the AOA for a couple of years. In addition, AOA can be detected in environmental samples by CARD-FISH. However, this technique requires fixation of the target cells and thus, makes further single cell genomic analyses very difficult.

In this diploma thesis Raman microspectroscopy was used because it is a non-destructive technique and it allows the identification of a plethora of macromolecules of a cell. The long term goal is to use Raman microspectroscopy to identify living AOA by a signature spectrum while they are trapped in an optical laser tweezer system. With tweezer sorted cells downstream applications like cultivation or MDA combined with genome sequencing (single cell genomics) will be performed. In this study a comprehensive Raman reference spectra library of various phyla was created. In addition, the spectra were statistically analyzed by an accurate classifier (Random Forest) in order to evaluate, which Raman bands are characteristic or even unique for AOA. Certain challenges had to be dealt with (e.g. carrier slide background signal, storage compounds, pigments, normalization and baselining of the Raman spectra). Preliminary results showed that AOA can be assigned to one cluster. Nevertheless, it was not possible to exclude certain other species from this cluster. There was strong evidence that long methylated alkyl groups of lipids (e.g. crenarchaeol, iso-diabolic acid) gave rise to Raman signature bands which resulted in a clustering of these specific organisms. However, I discovered that the used carrier slide material – calciumdifluoride – caused a more intense background signal than expected at the beginning of the study. Additionally, an AOA prediction script based on the Random Forest algorithm, which calculated significance values for all peaks was created in collaboration with Dr. David Berry and applied on two arctic AOA enrichment cultures. Cells were chosen at random and automatically assigned to either the AOA or non-AOA cluster. The results of the AOA cluster assignments based on the prediction script (predicted AOA content of the arctic AOA enrichments SV8-6: 30%; SV9-19: 45%) were plausible compared to the qPCR data (AOA content of the arctic AOA enrichments SV8-6: 17%; SV9-19: 26%) since just 30 cells were measured in total.

In order to prove the preliminary results, certain experiments have to be performed (e.g. subtraction of carrier slide and storage compound spectra). Taken together, the obtained results show the potential of Raman microspectroscopy for rapid and non-destructive characterization of the chemical composition of microbial cells – a feature that is particularly attractive if combined with subsequent

cell sorting.

6 Zusammenfassung

Nitrifikation ist die Oxidation von Ammonium zu Nitrit und weiter zu Nitrat. Es ist einer der wichtigsten Schritte im globalen Stickstoffkreislauf und außerdem ist die Reaktion auch sehr wichtig für die industrielle Landwirtschaft und das Funktionieren von Kläranlagen. Deshalb ist es unerlässlich die wichtigsten verantwortlichen Mikroorganismen in diesem System zu verstehen. Für lange Zeit glaubte man, dass für den ersten Schritt der Nitrifikation ausschließlich Ammoniak-oxidierende Bakterien (AOB) verantwortlich sind, jedoch wurden vor einigen Jahren die Ammonium/Ammoniak-oxidierenden Archaeen (AOA) entdeckt. Sie sind den AOB in bestimmten Umweltproben zahlenmäßig weit überlegen und darum besonders interessant für Nitrifikationsforscher. Bisher wurde die Ammoniummonooxygenase (Amo) als phylogenetischer Marker für die Identifizierung von AOA verwendet und zusätzlich CARD-FISH eingesetzt um diese Organismen in Umweltproben nachzuweisen. Durch die Anwendung dieser Techniken werden allerdings die Zellen zerstört bzw. chemisch stark verändert, was eine weitere genomische Analyse auf Einzelzellniveau sehr schwierig gestaltet.

In dieser Diplomarbeit hatte ich die Raman Mikrospektroskopie verwendet, da sie mehr oder weniger zerstörungsfrei arbeitet und in kurzer Zeit eine Vielzahl an Makromolekülen einer Zelle identifizieren kann. Das Ziel auf lange Sicht hin ist es mittels Raman Mikrospektroskopie lebende AOA Zellen anhand von spezifischen Ramanspektren zu identifizieren während sie in einem Laserpinzetten-System gefangen sind. In der Folge könnte man dann diese, immer noch lebenden, Zellen aussortieren und kultivieren bzw. genomische Einzelzell-Analysen basierend auf MDA damit durchführen. Um erste Schritte in diese Richtung zu entwickeln, wurde im Verlauf dieser Diplomarbeit eine umfassende Referenzspektrenbibliothek angelegt welche aus Organismen diverser Phyla besteht. Die Spektren dieser Organismen wurden dann durch einen maschinellen Lernalgorithmus (Random Forest) analysiert und für AOA wichtige Bereiche wurden identifiziert. Im Verlauf der Studie gab es verschiedene Schwierigkeiten mit denen ich konfrontiert wurde (z.B. Hintergrundsignal der Trägerfläche, Speicherkomponenten, Pigmente, Normalisierung und „Baselinen“ der Ramanspektren). Die Ergebnisse dieser Diplomarbeit haben jedoch gezeigt, dass es möglich ist die AOA basierend auf ihren Ramanspektren in einem Cluster zu vereinen, jedoch war es nicht möglich einen reinen AOA-Cluster zu erzeugen. Es gab starke Hinweise, dass lange methylierte Alkylgruppen von Lipiden (z.B. Crenarchaeol, „iso-diaolic acid“) zu charakteristischen Ramansignalen führten welche eine gemeinsame Cluster-Zuweisung von AOA und bestimmten Bakterien bedingten. Im Verlauf der Studie wurde des Weiteren herausgefunden, dass die verwendete Trägerfläche für Zellen ein weitaus stärkeres Hintergrundsignal generierte als zu Beginn der Studie vermutet. Zusätzlich wurde in Zusammenarbeit mit Dr. David Berry ein AOA-Vorhersageskript entworfen, welches aufgrund der vom Random Forest Algorithmus gewichteten AOA-spezifische

Spektrumbereiche berechnen konnte, mit welcher Wahrscheinlichkeit ein Spektrum unbekannter Herkunft von einem AOA stammt oder nicht. Ich hatte dieses Skript an Ramanspektren von zufällig ausgewählten Zellen von zwei arktischen AOA-Anreicherungskulturen ausprobiert. Die Resultate der Cluster-Zuweisung, basierend auf dem AOA Vorhersage Skript (AOA-Inhalt der arktischen AOA Anreicherungskulturen: SV8-6: 30%; SV9-19: 45%), wirkten plausibel im Vergleich zu den qPCR Daten (AOA-Inhalt der arktischen AOA Anreicherungskulturen: SV8-6: 17%; SV9-19: 26%). Um diese Resultate zu bestätigen müssen in Zukunft noch einige darauf aufbauende Experimente ausgeführt werden, wie zum Beispiel die Subtraktion vom Trägermaterial-Hintergrundspektrum und diverser Speichersubstanzen. Zusammenfassend kann man sagen, dass die erzielten Resultate eindeutig das Potential der Raman Mikrospektroskopie aufzeigen, wenn es um eine schnelle, nicht destruktive Charakterisierung der chemischen Zusammensetzung von mikrobiellen Zellen geht – eine Eigenschaft die diese Technik vor allem in Kombination mit nachfolgender Zellsortierung attraktiv macht.

7 List of abbreviations

(a.u.)	arbitrary units
16S rRNA	small subunit of rRNA
λ	wavelength
µ	mikro (10^{-6})
°C	degree Celsius
%	percent
abs	absolut
Amo	ammonium monooxygenase
amoA	gene coding for subunit A of Amo
amoB	gene coding for subunit B of Amo
amoC	gene coding for subunit C of Amo
AOA	ammonia-oxidizing archaea
AOB	ammonia-oxidizing bacteria
AOP	ammonia-oxidizing prokaryotes
bagging	bootstrap aggregating
CARD	catalyzed reporter deposition
CART	classification and regression trees
$CaCl_2$	calciumchloride
CaF_2	calciumdifluoride
CLSM	confocal laser scanning microscope
cm	centimeter(s)
cm^{-1}	reciprocal centimeter/wavenumber
conc.	concentration
Cy3	5,5'-di-sulfo-1,1'-di-(X-carbopentynyl)-3,3,3',3'-tetra-methylindol-Cy3.18-derivative N-hydroxysuccimidester
Cy5	5,5'-di-sulfo-1,1'-di-(X-carbopentynyl)-3,3,3',3'-tetra-methylindol-Cy5.18-derivative N-hydroxysuccimidester
DAPI	4’-6’-di-amidino-2-phenylindole
DNA	desoxyribonucleic acid
e^-	electron
EDTA	Ethylenedinitrilotetraacetic acid
et al.	*et alteri*
EtOH	ethanol
FA	formamide

Fig.	Figure
FISH	fluorescence in situ hybridization
fluos	5,(6)-carboxyfluorescein-N-hydroxysuccimidester
g	gram(s)
GDGT	glycerol dibiphytanyl glycerol tetraether
h	hour(s)
H^+	proton(s)
H_2O	water
H_2O_2	hydrogen peroxide
Hao	hydroxylamine oxidoreductase
HB	hybridization buffer
He-Ne	helium-neon
m	milli 10^{-3}
M	molar
MDA	multiple displacement amplification
MDS	multi-dimensional scaling
MO	microorganism
MQ	Milli-Q (double distilled water)
min	minute(s)
NaCl	sodium chloride
NaOH	sodium hydroxide
Nd:YAG	neodymium-doped yttrium aluminum garnet
NH_3	ammonia
NO	nitrous oxide
NO_2^-	nitrite
NO_3^-	nitrate
O_2	oxygen
OOB	out of box
PBS	phosphate buffered saline
PHA	polyhydroxyalkanoates
Phe	phenylalanine
PFA	paraformaldehyde
qPCR	quantitative polymerase chain reaction
RNA	ribonucleic acid
rpm	rotations per minute
rRNA	ribosomal ribonucleic acid

RF	random forest
RT	room temperature
SDS	sodium dodecyl sulfate
sec	second(s)
SERDS	shifted-excitation Raman difference spectroscopy
SNR	signal to noise ratio
sp.	species
Tab.	Table
TAG	triglycerols
temp.	temperature
vol.	volume
v/v	volume/volume
WB	washing buffer
WEs	wax esters
WWTP	wastewater treatment plant
w/v	weight/volume

8 Appendix

8.1 *Acidobacteria*

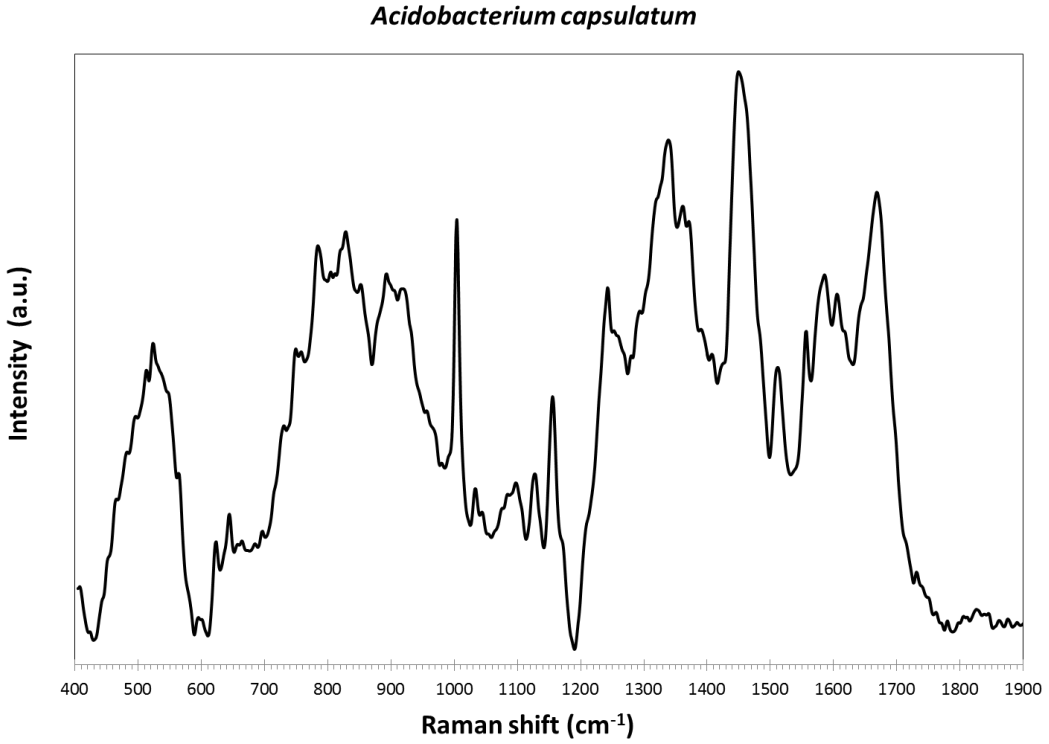

Figure 8.1. Mean Raman spectrum of *Acidobacterium capsulatum* (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

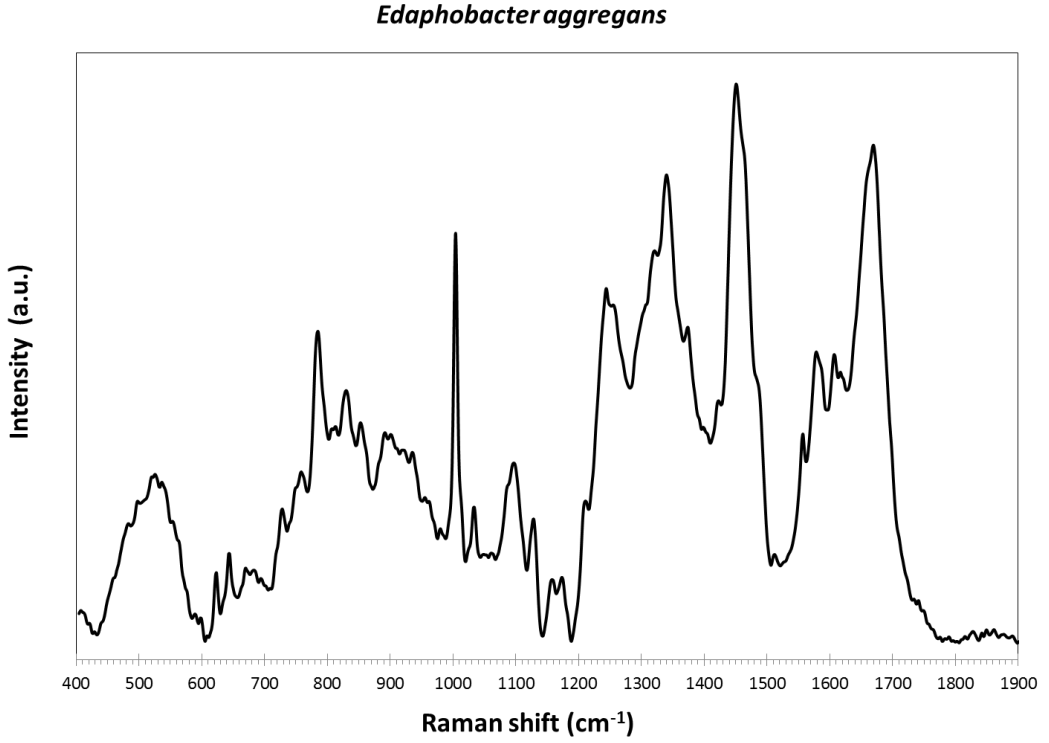

Figure 8.2. Mean Raman spectrum of *Edaphobacter aggregans* (n = 14). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

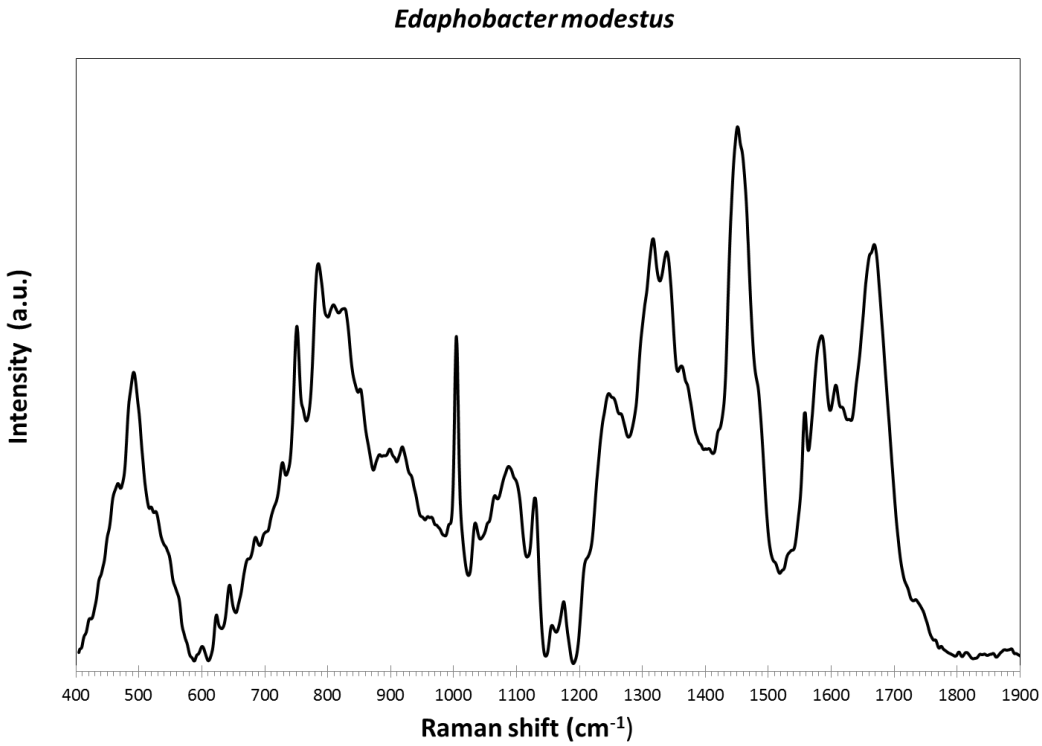

Figure 8.3. Mean Raman spectrum of *Edphobacter modestus* (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.2 *Alphaproteobacteria*

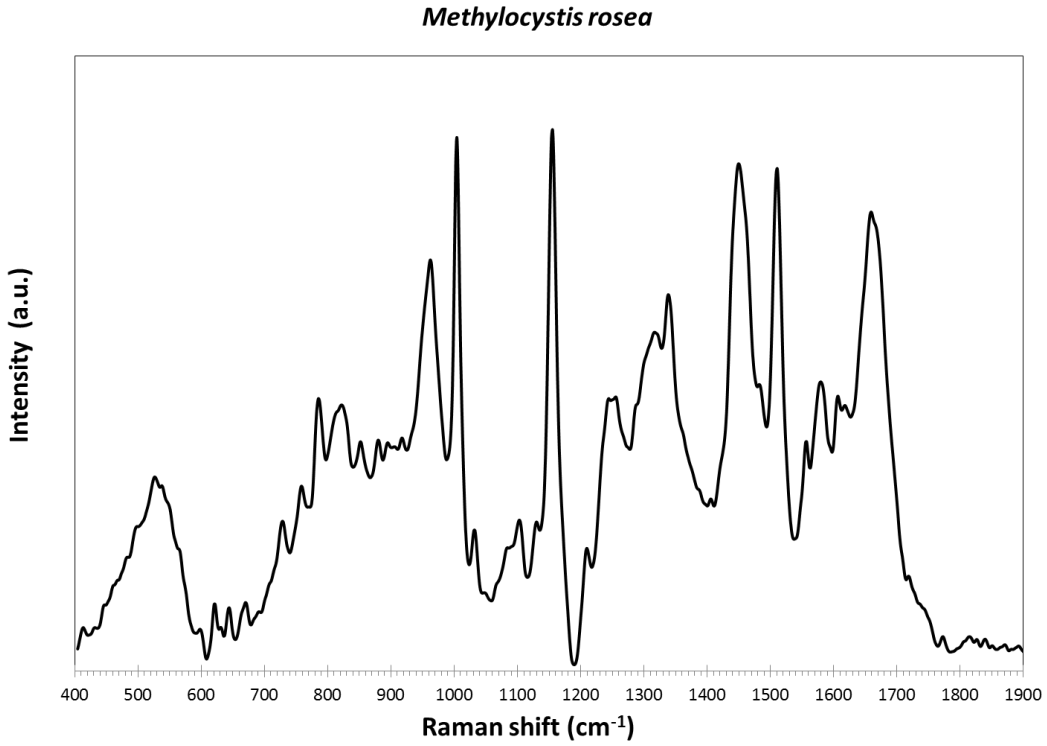

Figure 8.4. Mean Raman spectrum of *Methylocystis rosea* (n = 9). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

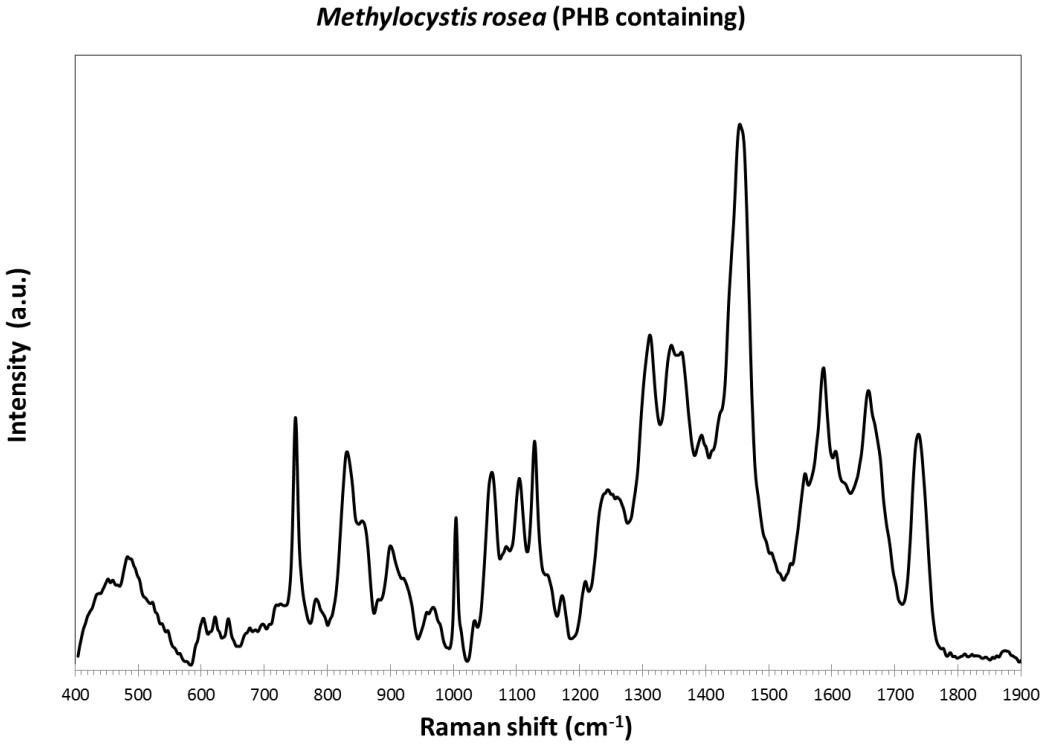

Figure 8.5. Mean Raman spectrum of *Methylocystis rosea* containing PHB, (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.3 *Betaproteobacteria*

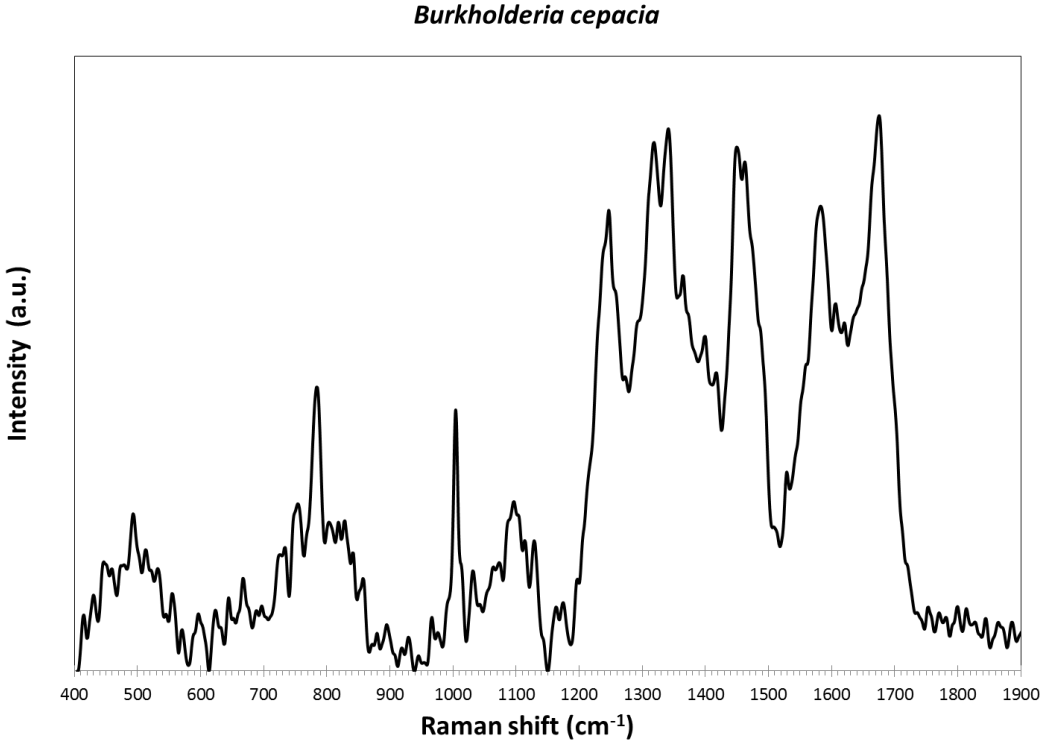

Figure 8.6. Mean Raman spectrum of *Burkholderia cepacia*, (n = 9). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

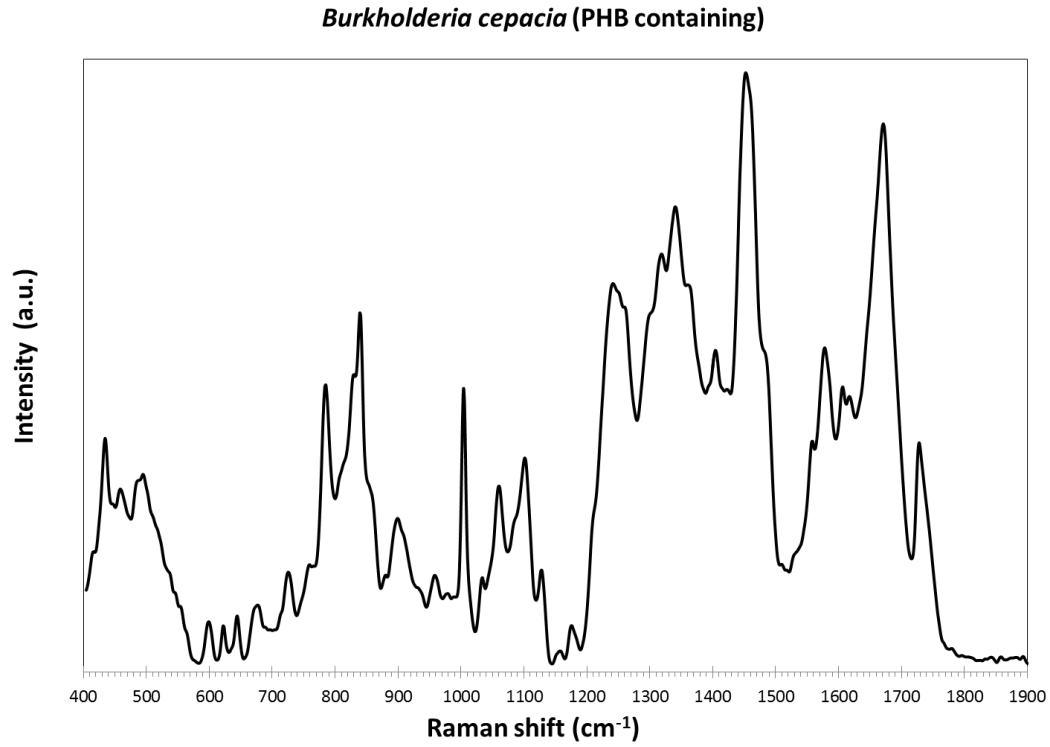

Figure 8.7. Mean Raman spectrum of *Burkholderia cepacia* containing PHB, (n = 6). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.4 *Chloroflexi*

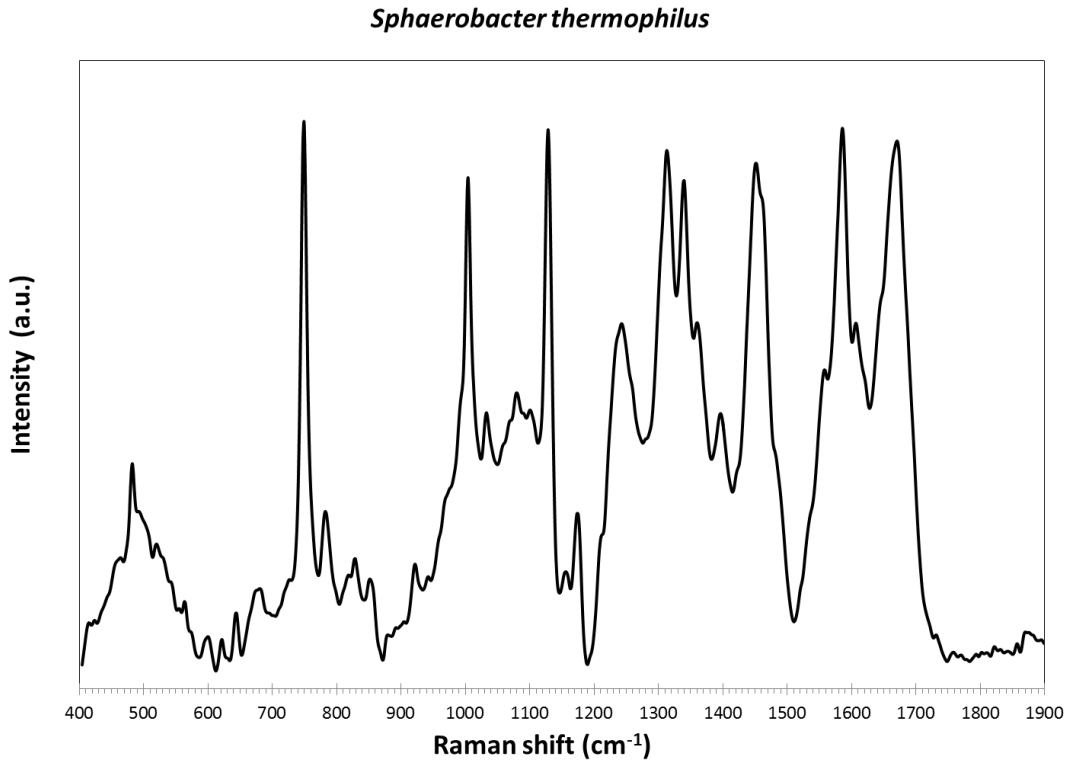

Figure 8.8. Mean Raman spectrum of *Sphaerobacter thermophilus*, (n = 12). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.5 *Crenarchaeota*

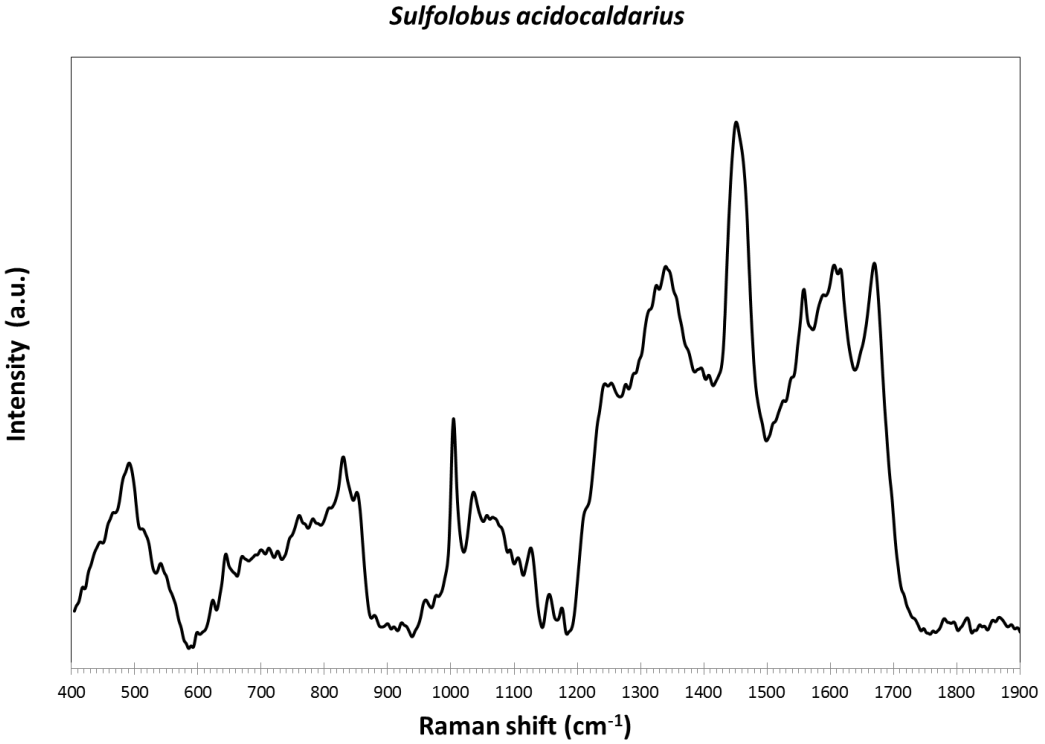

Figure 8.9. Mean Raman spectrum of *Sulfolobus acidocaldarius*, (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

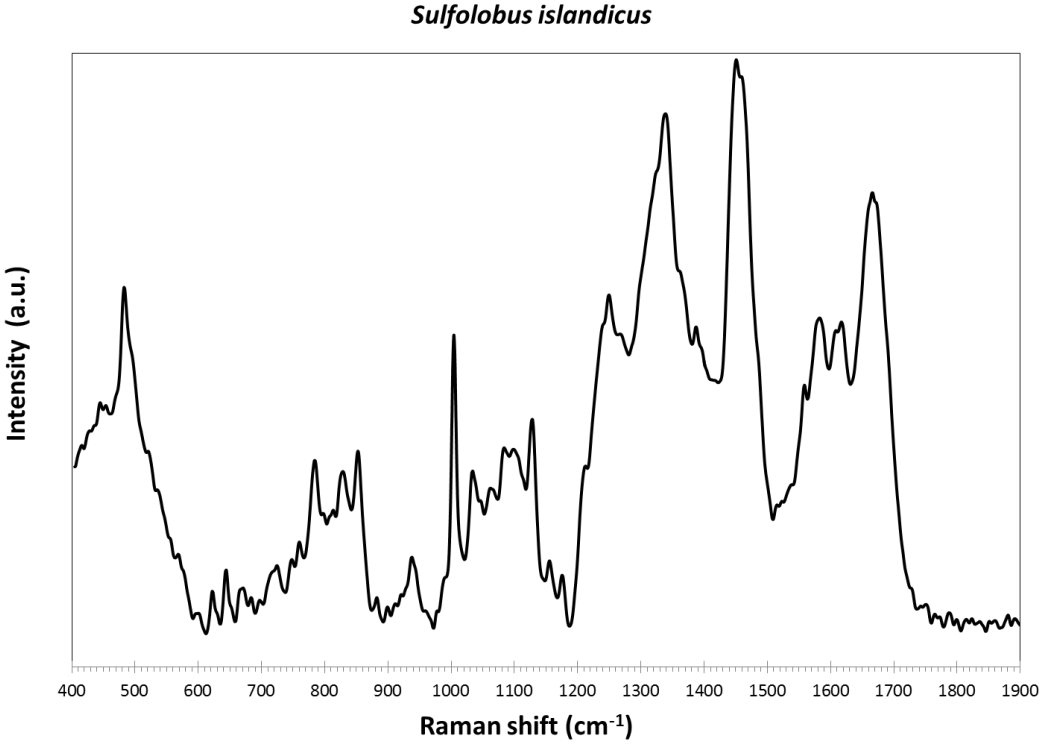

Figure 8.10. Mean Raman spectrum of *Sulfolobus islandicus*, (n = 10). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

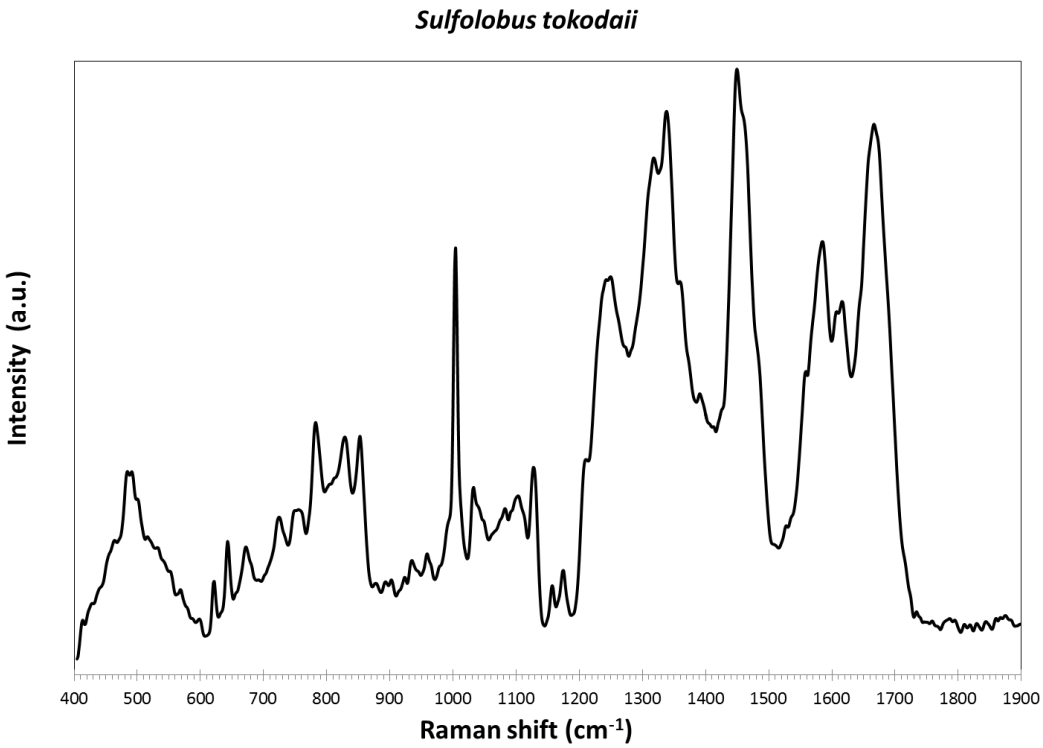

Figure 8.11. Mean Raman spectrum of *Sulfolobus tokodaii*, (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.6 *Deltaproteobacteria*

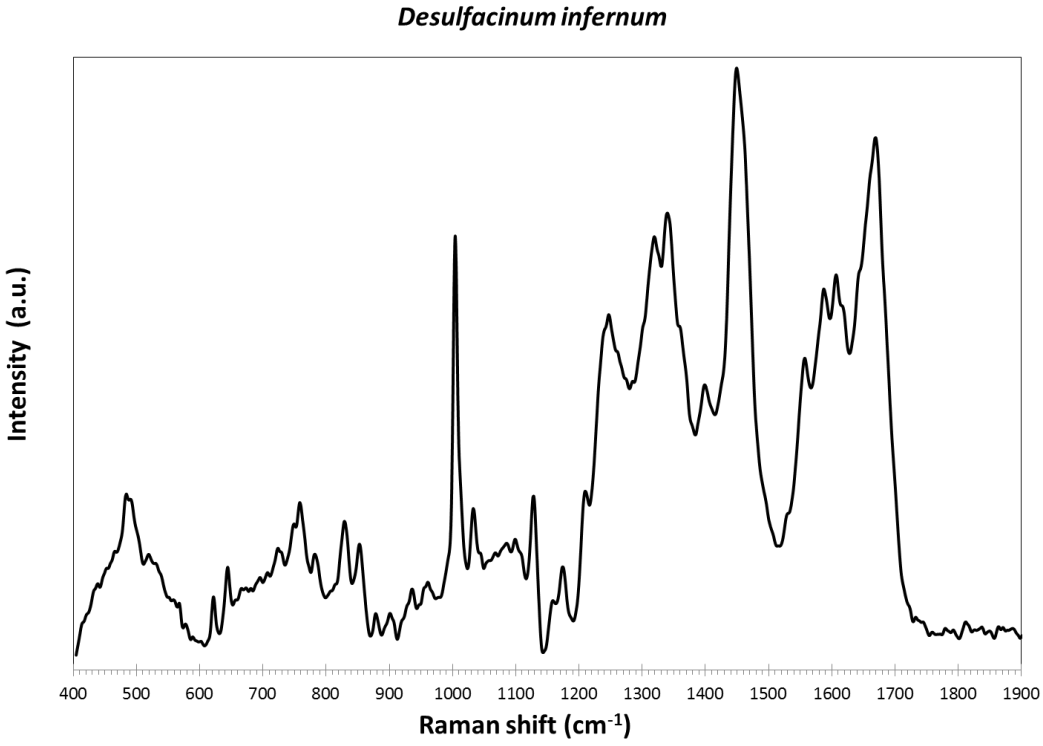

Figure 8.12. Mean Raman spectrum of *Desulfacinum infernum*, (n = 24). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

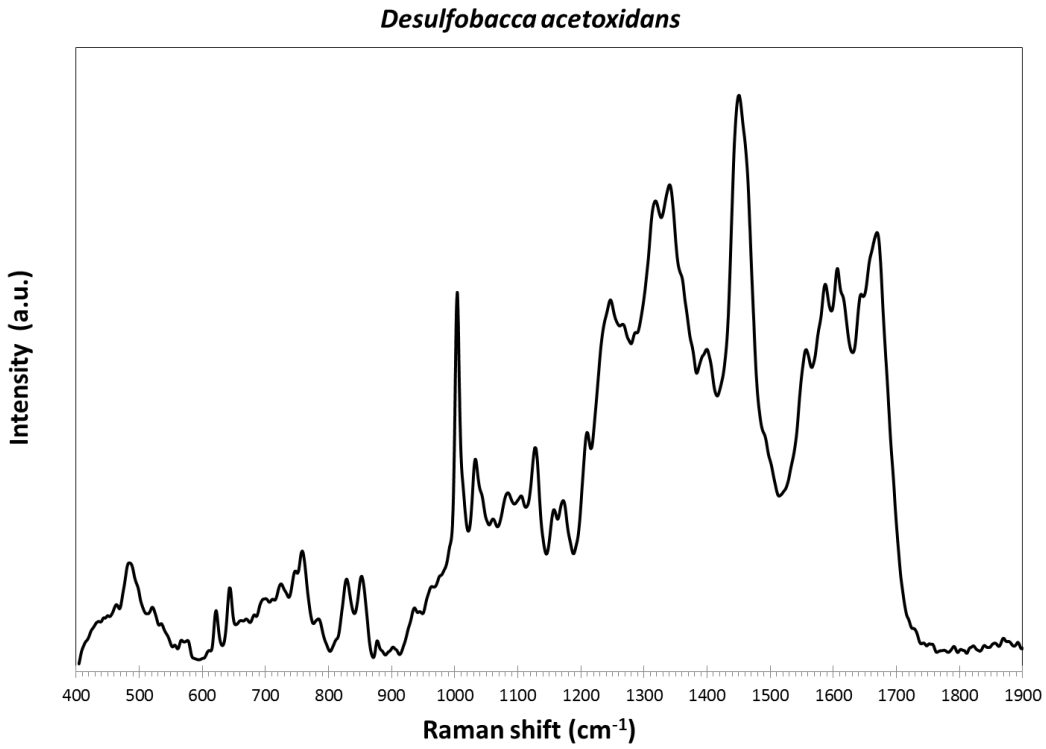

Figure 8.13. Mean Raman spectrum of *Desulfobacca acetoxidans*, (n = 16). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

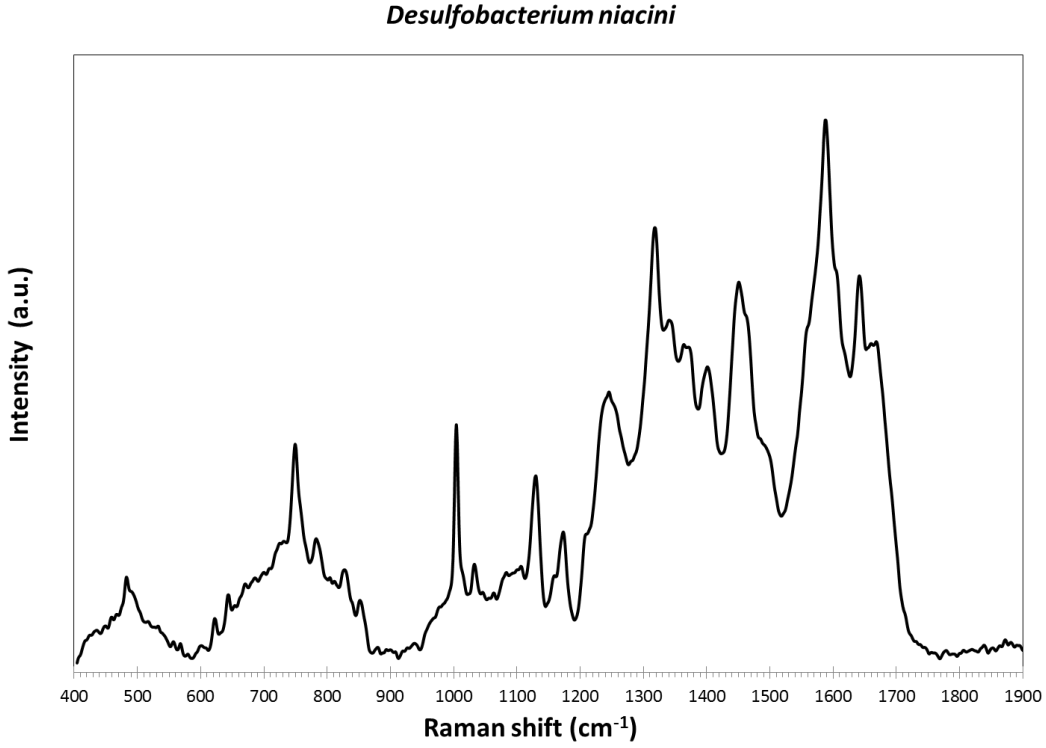

Figure 8.14. Mean Raman spectrum of *Desulfobacterium niacini*, (n = 21). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

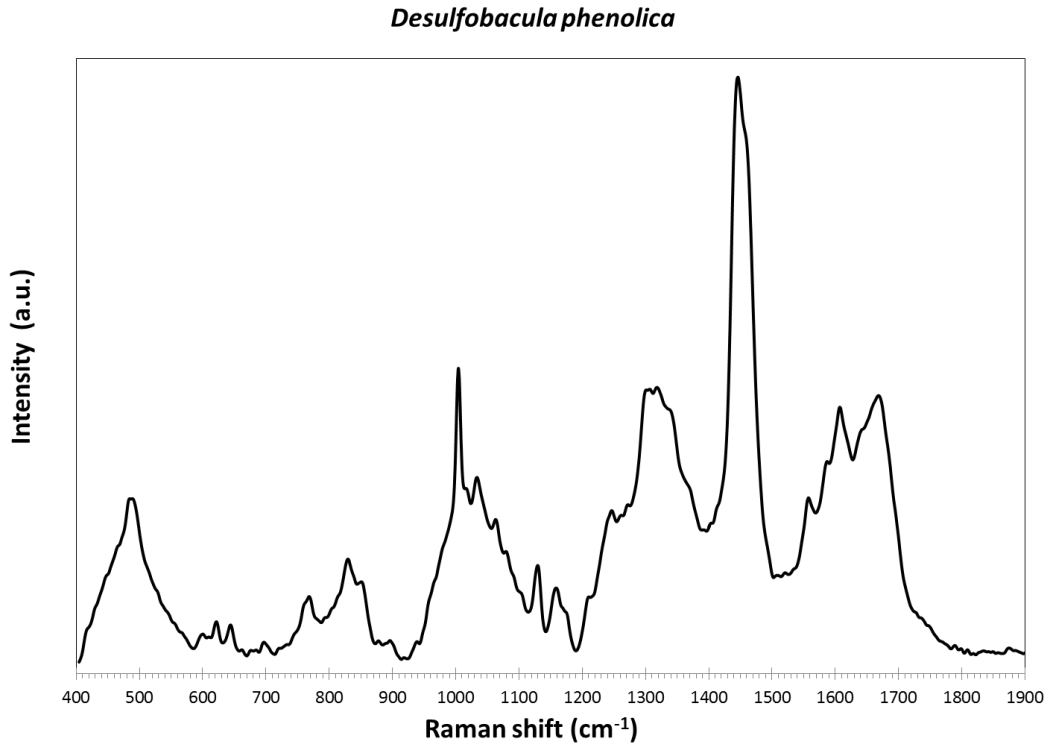

Figure 8.15. Mean Raman spectrum of *Desulfobacula phenolica*, (n = 18). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

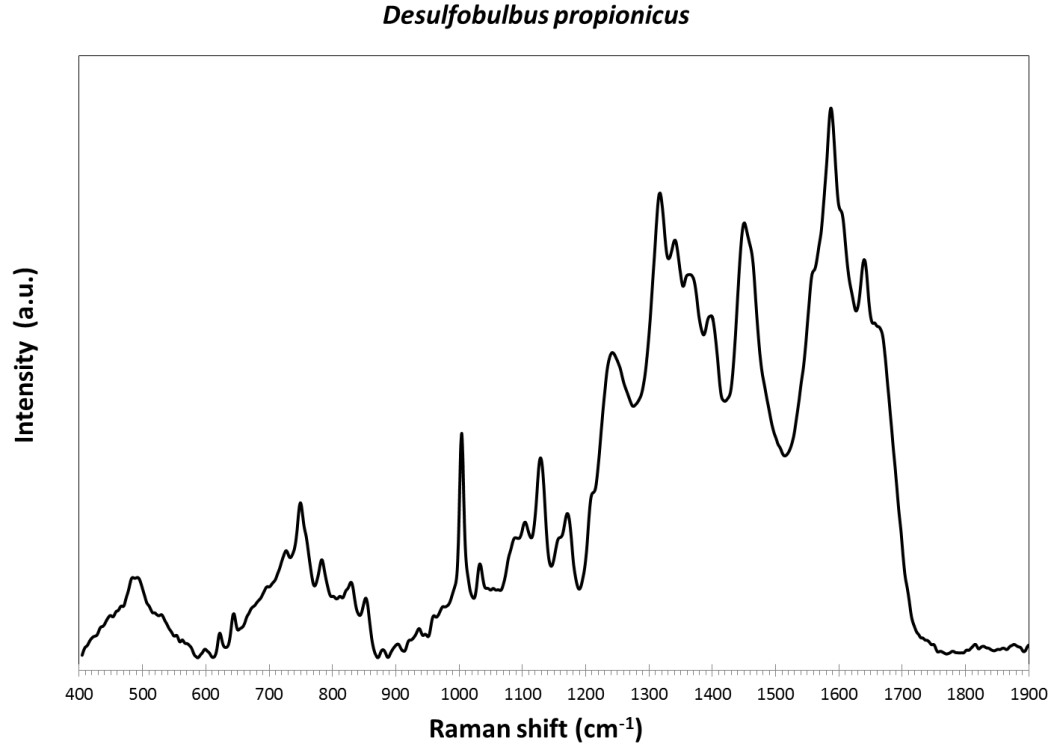

Figure 8.16. Mean Raman spectrum of *Desulfobulbus propionicus*, (n = 21). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

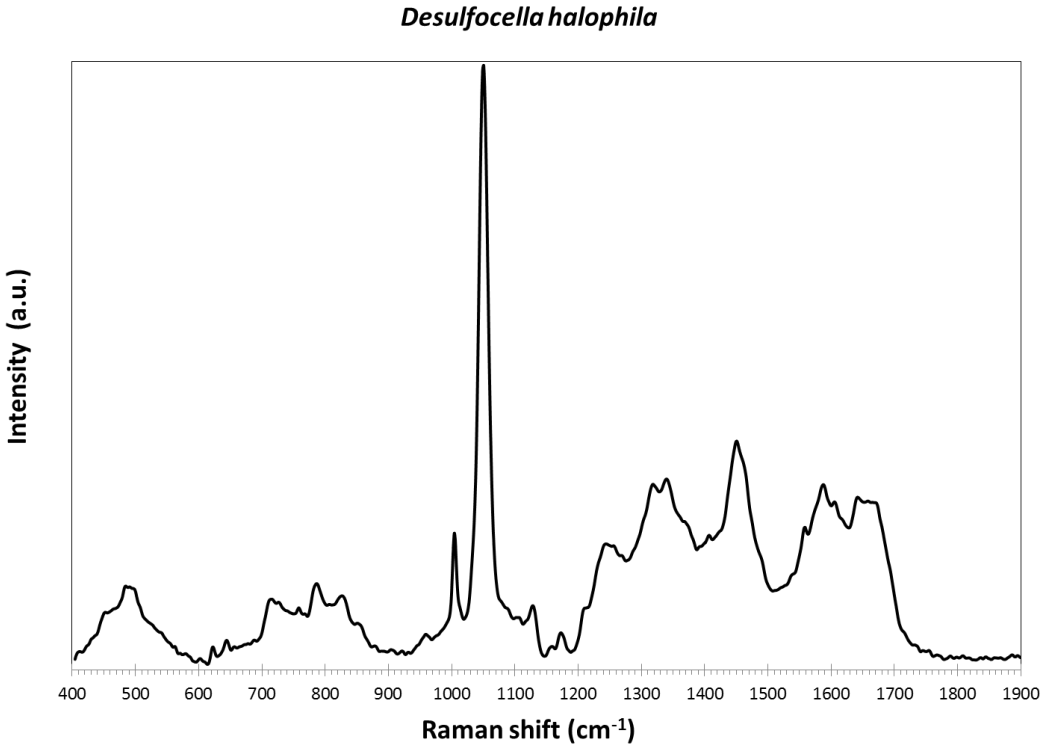

Figure 8.17. Mean Raman spectrum of *Desulfocella halophila*, (n = 16). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

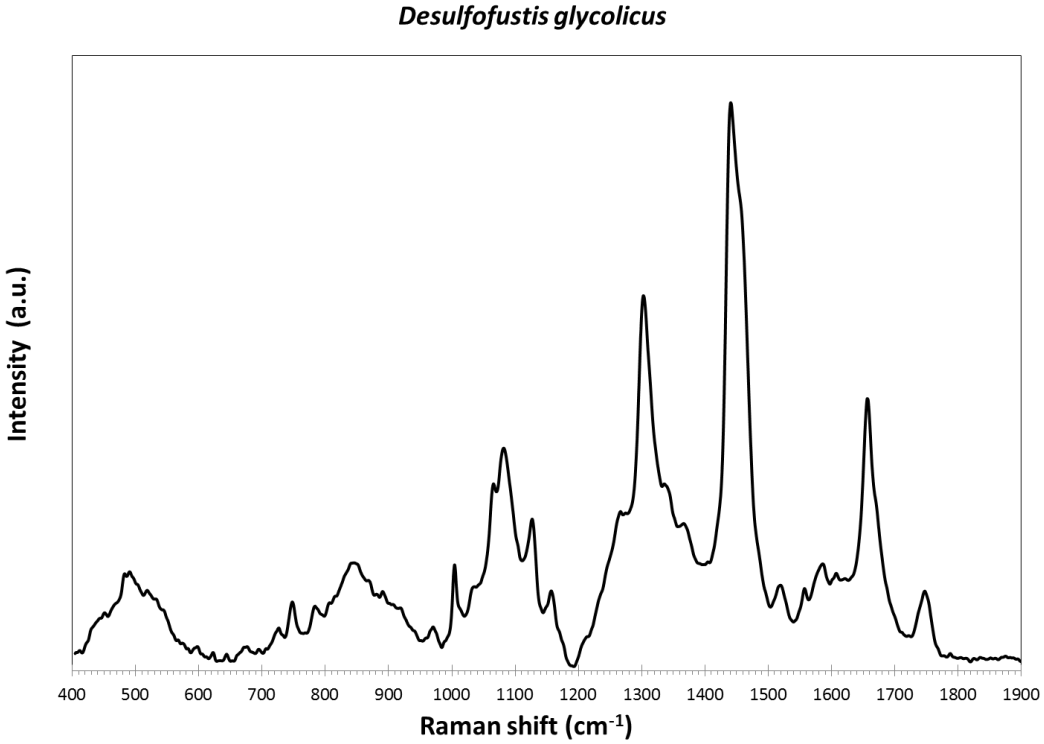

Figure 8.18. Mean Raman spectrum of *Desulfofustis glycolicus*, (n = 13). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

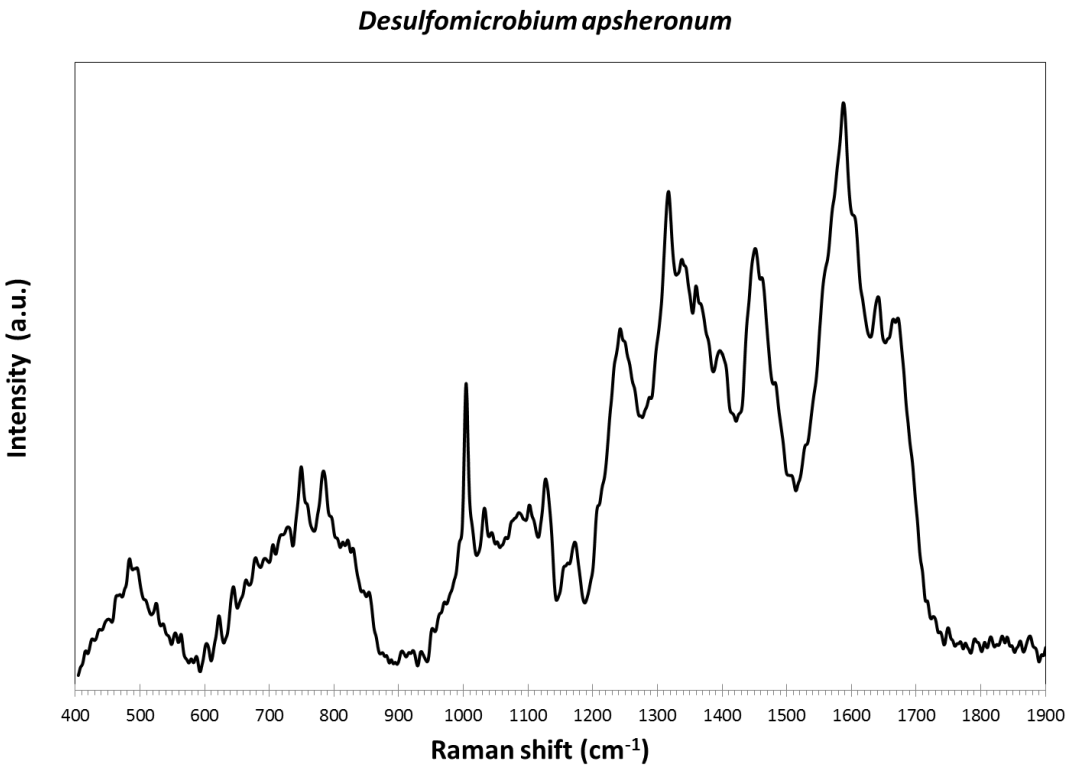

Figure 8.19. Mean Raman spectrum of *Desulfomicrobium apsheronum*, (n = 11). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

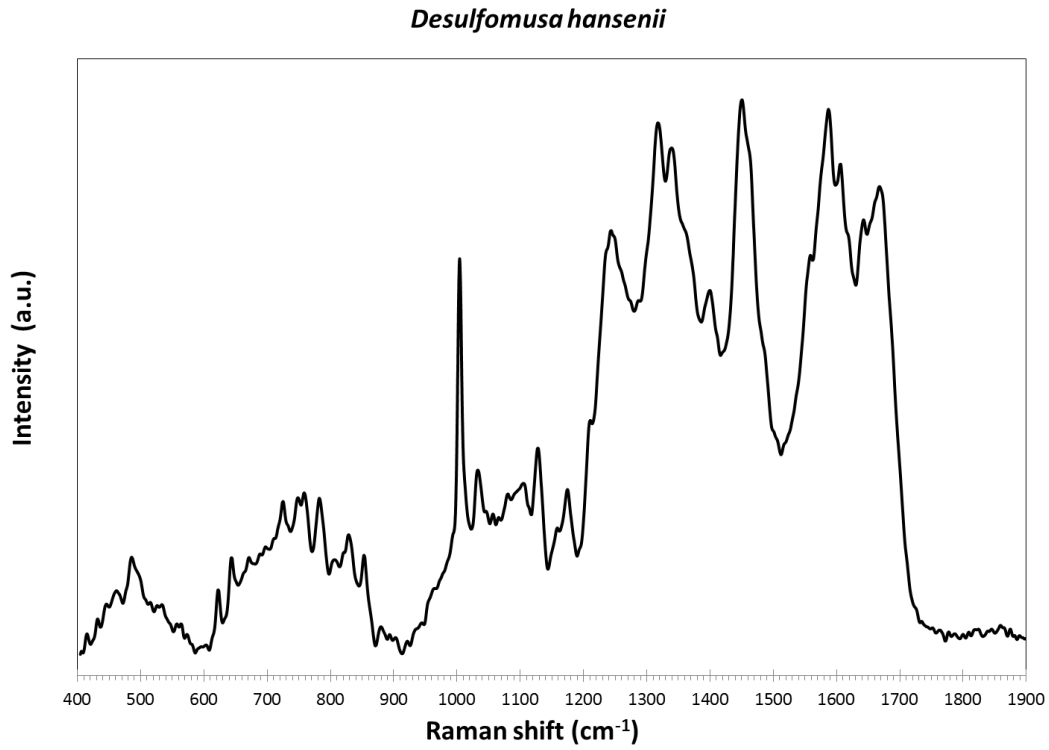

Figure 8.20. Mean Raman spectrum of *Desulfomusa hansenii*, (n = 18). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

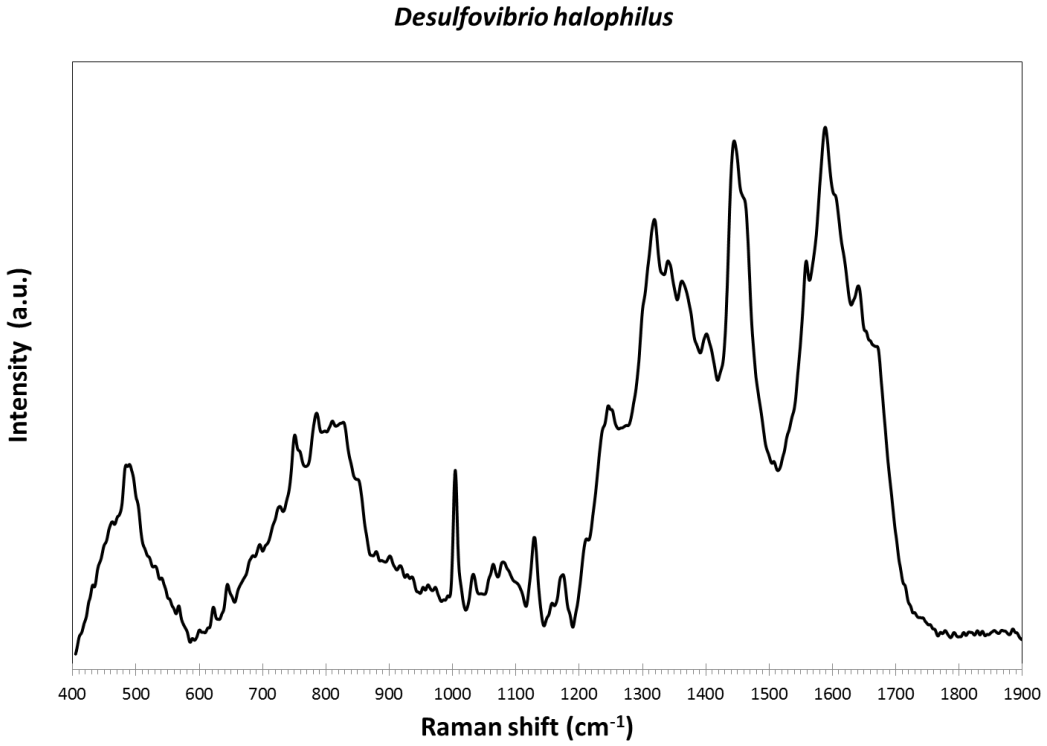

Figure 8.21. Mean Raman spectrum of *Desulfovibrio halophilus*, (n = 13). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

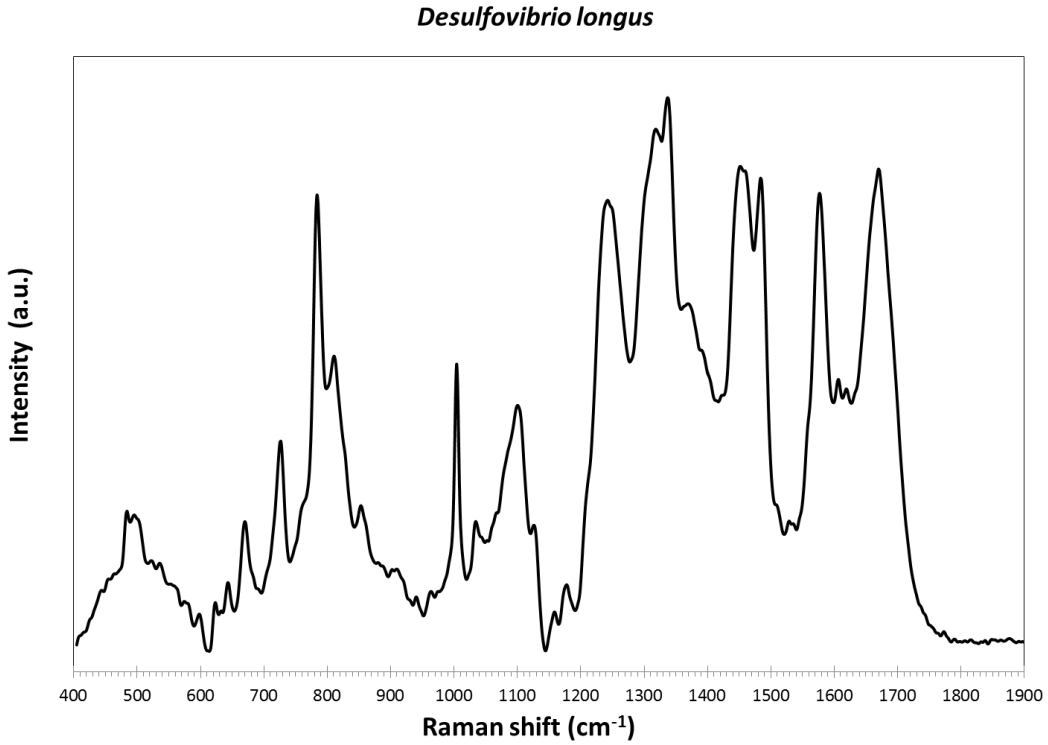

Figure 8.22. Mean Raman spectrum of *Desulfovibrio longus*, (n = 20). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

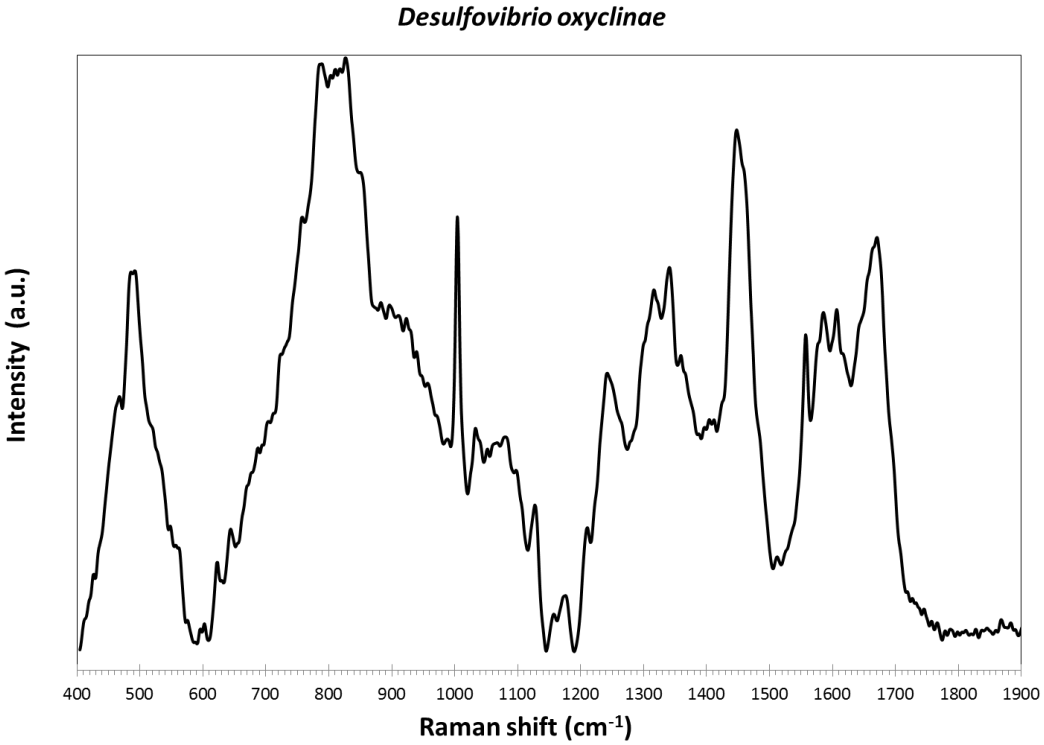

Figure 8.23. Mean Raman spectrum of *Desulfovibrio oxyclinae*, (n = 14). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

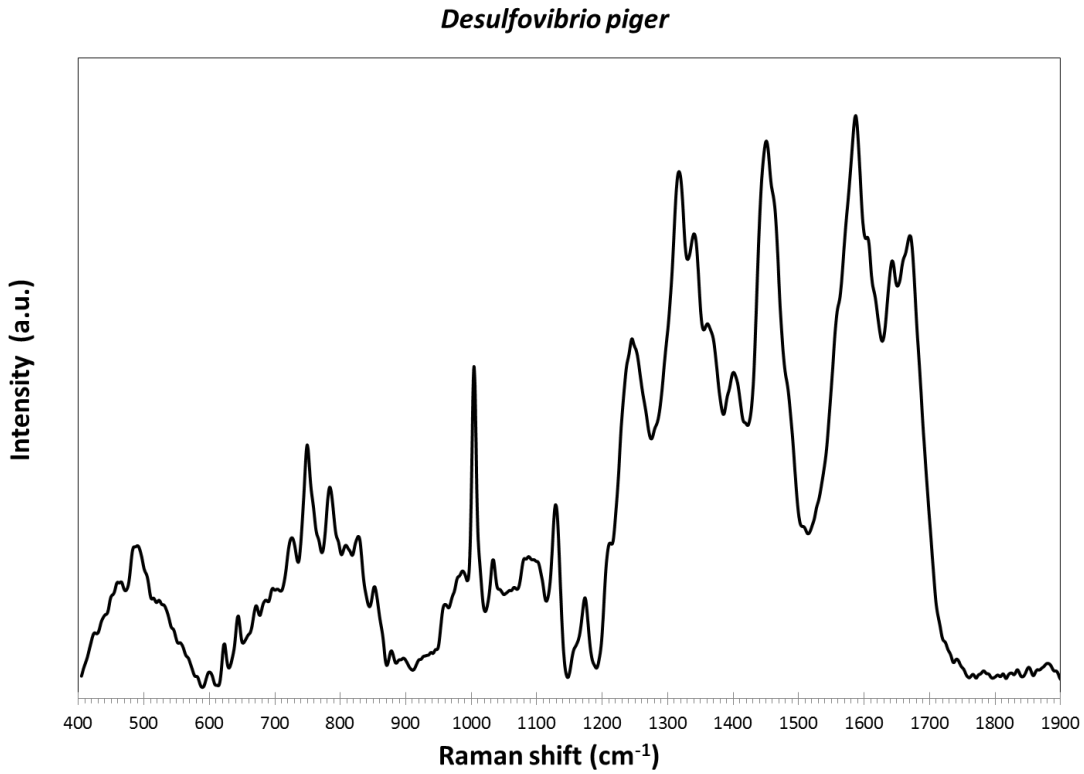

Figure 8.24. Mean Raman spectrum of *Desulfovibrio piger,* (n = 17). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.7 *Firmicutes*

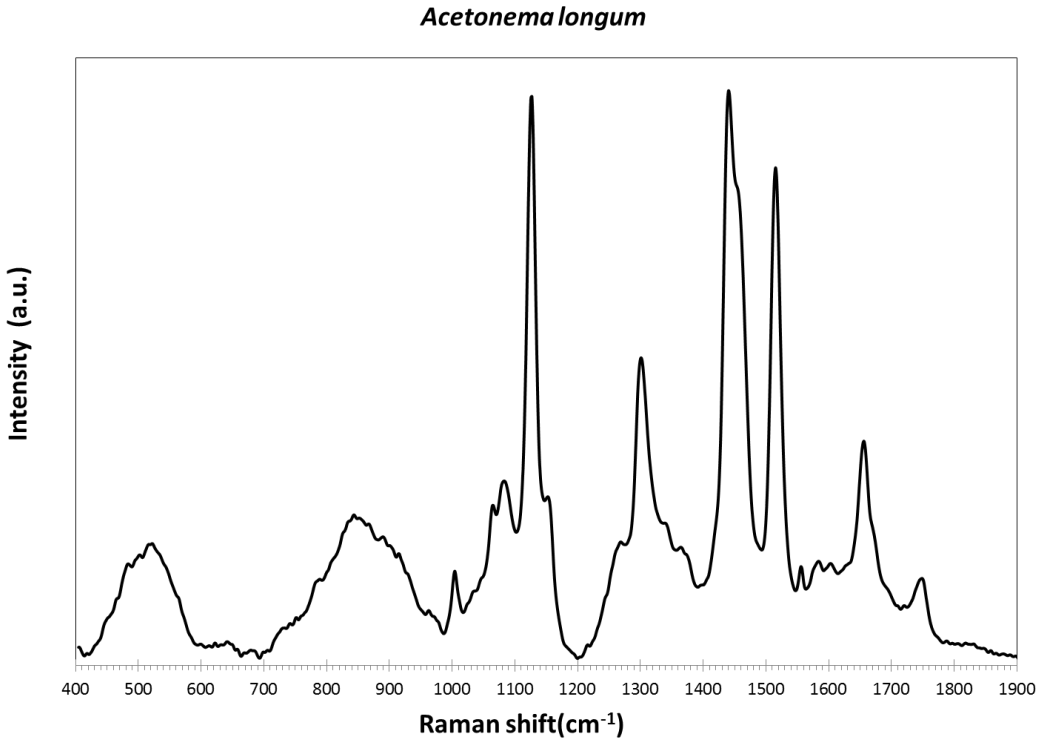

Figure 8.25. Mean Raman spectrum of *Acetonema longum*, (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

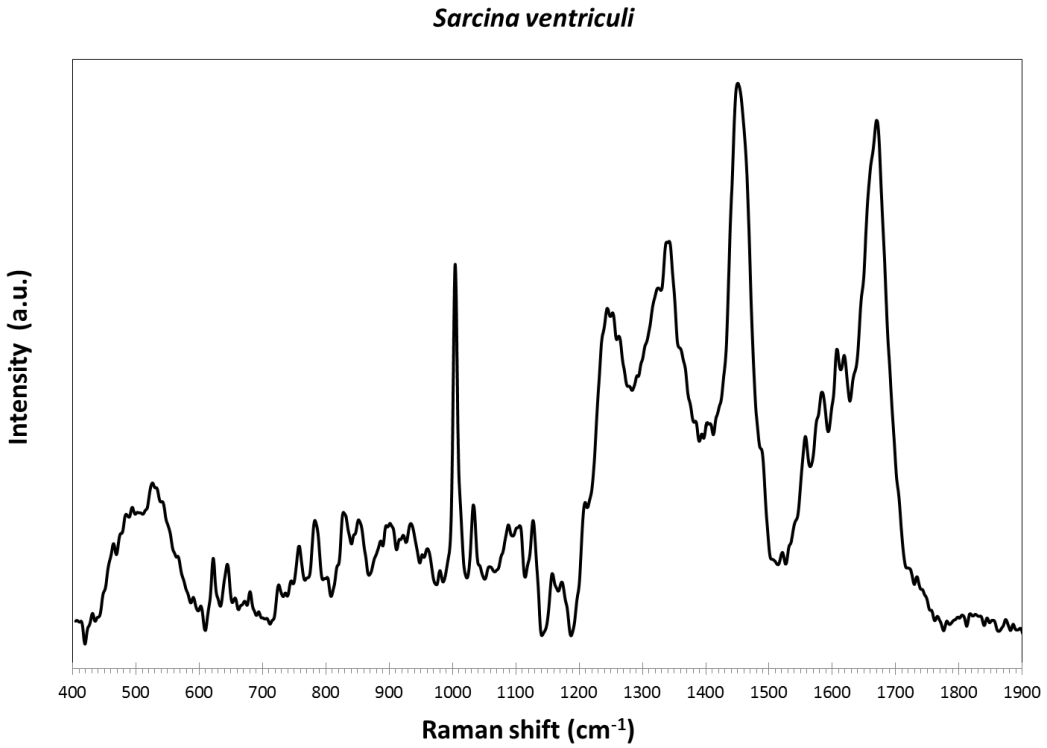

Figure 8.26. Mean Raman spectrum of *Sacrina ventriculi*, (n = 6). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

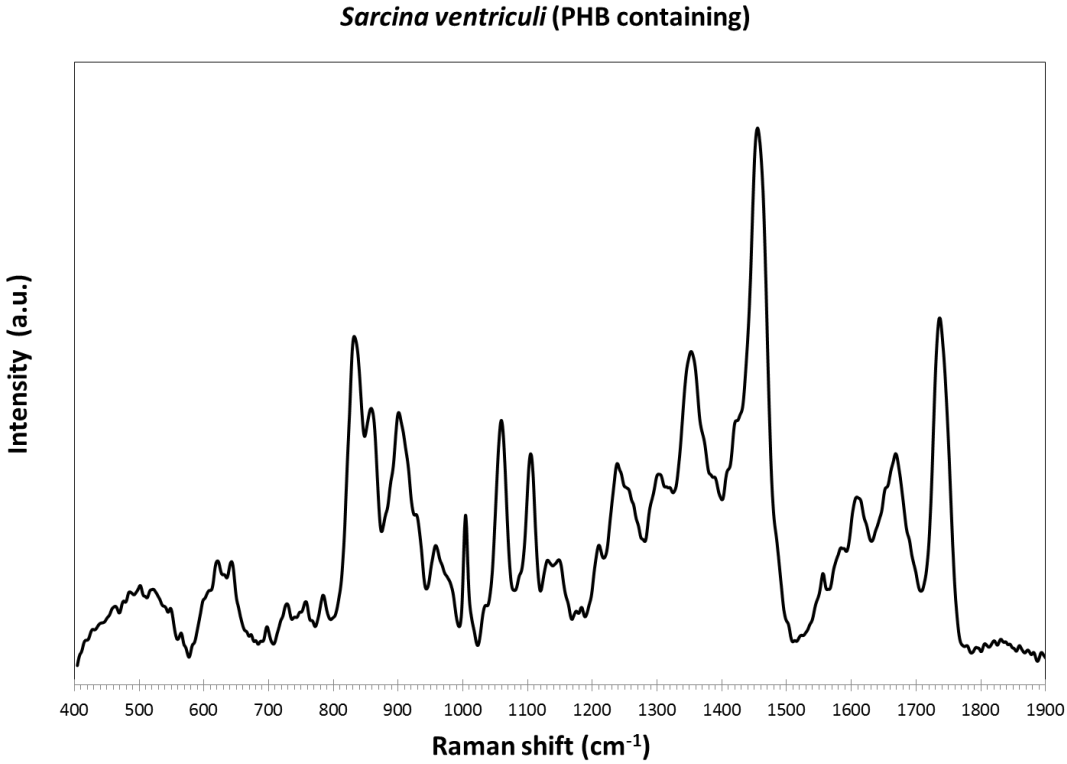

Figure 8.27. Mean Raman spectrum of *Sacrina ventriculi* containing PHB, (n = 14). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

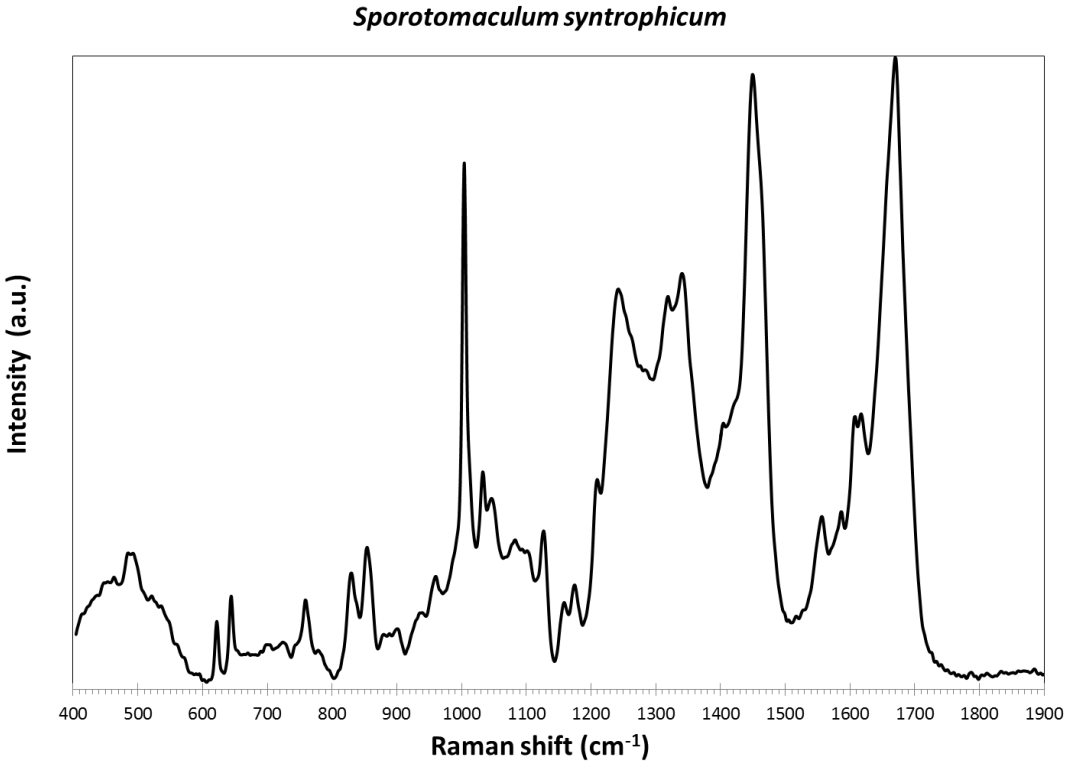

Figure 8.28. Mean Raman spectrum of *Sporomaculum syntrophicum*, (n = 17). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

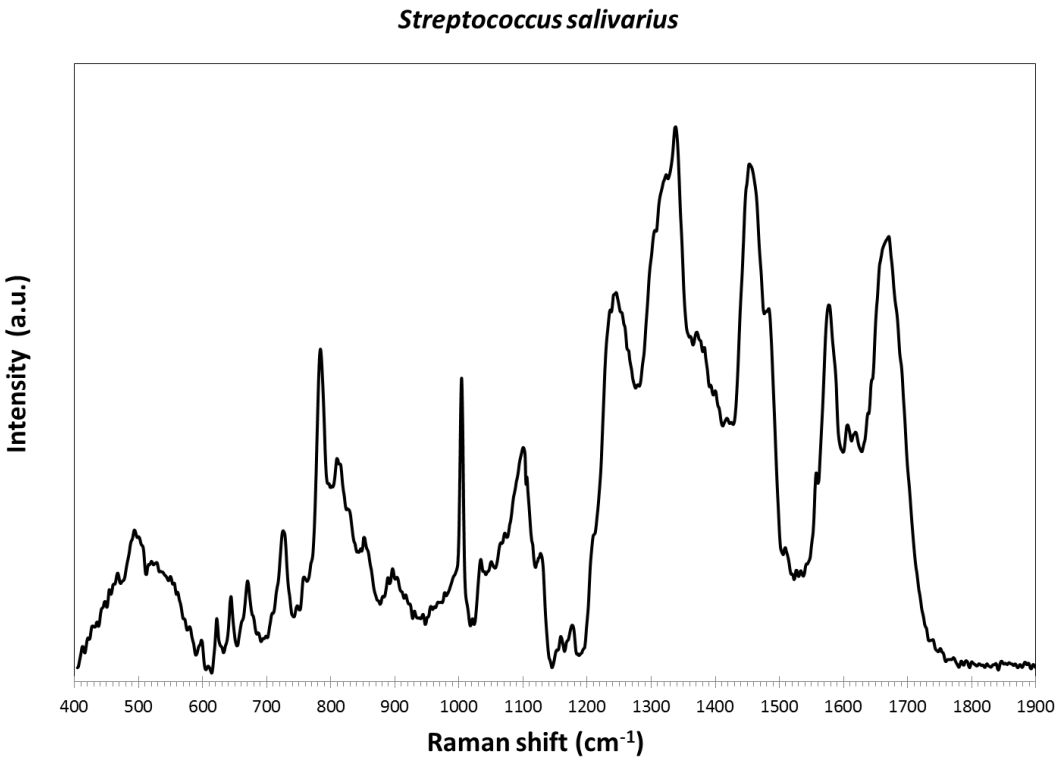

Figure 8.29. Mean Raman spectrum of *Streptococcus salivarius*, (n = 13). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.8 *Gammaproteobacteria*

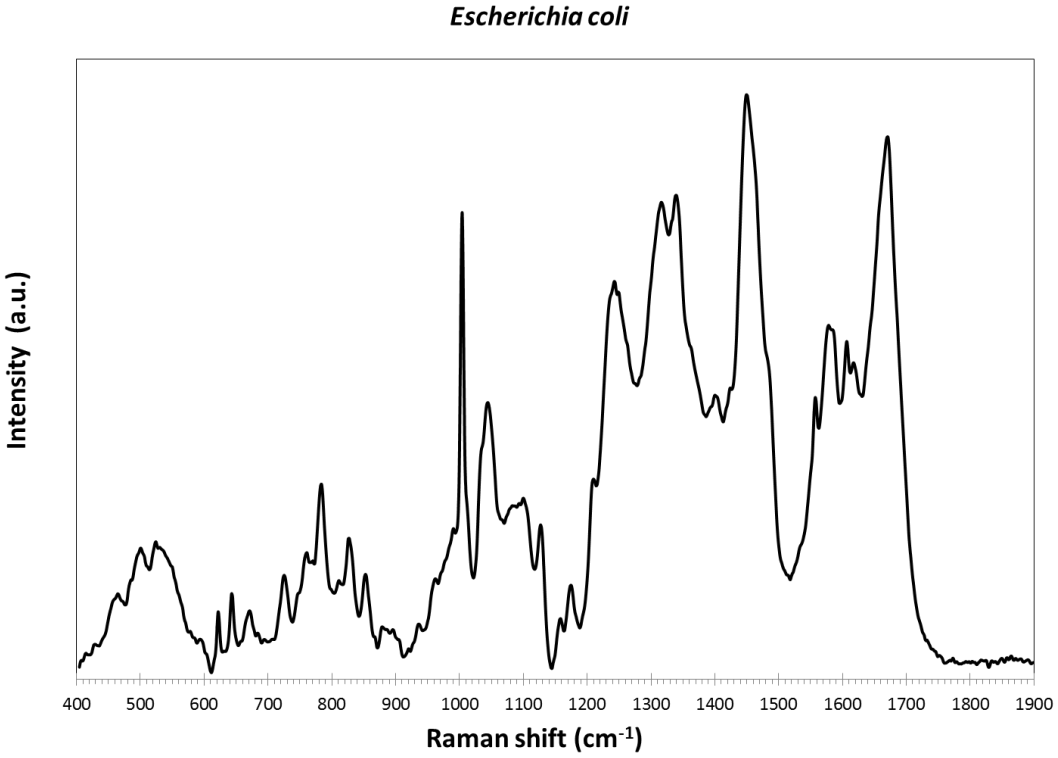

Figure 8.30. Mean Raman spectrum of *Escherichia coli*, (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

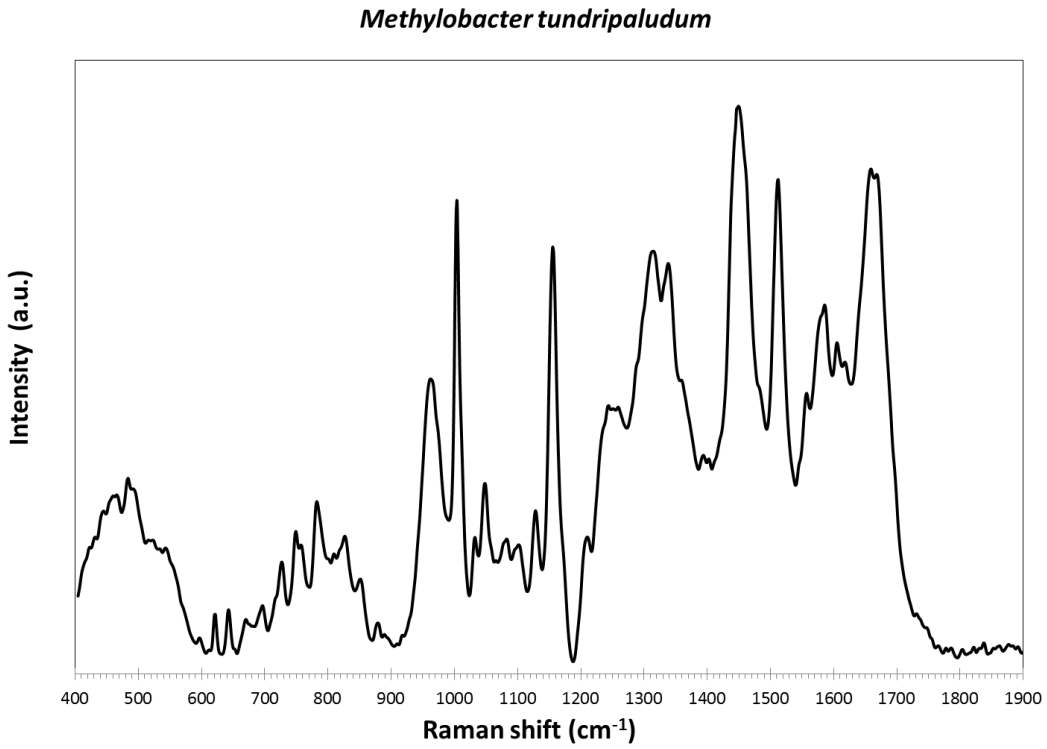

Figure 8.31. Mean Raman spectrum of *Methlobacter tundripaladum*, (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.9 *Nitrospirae*

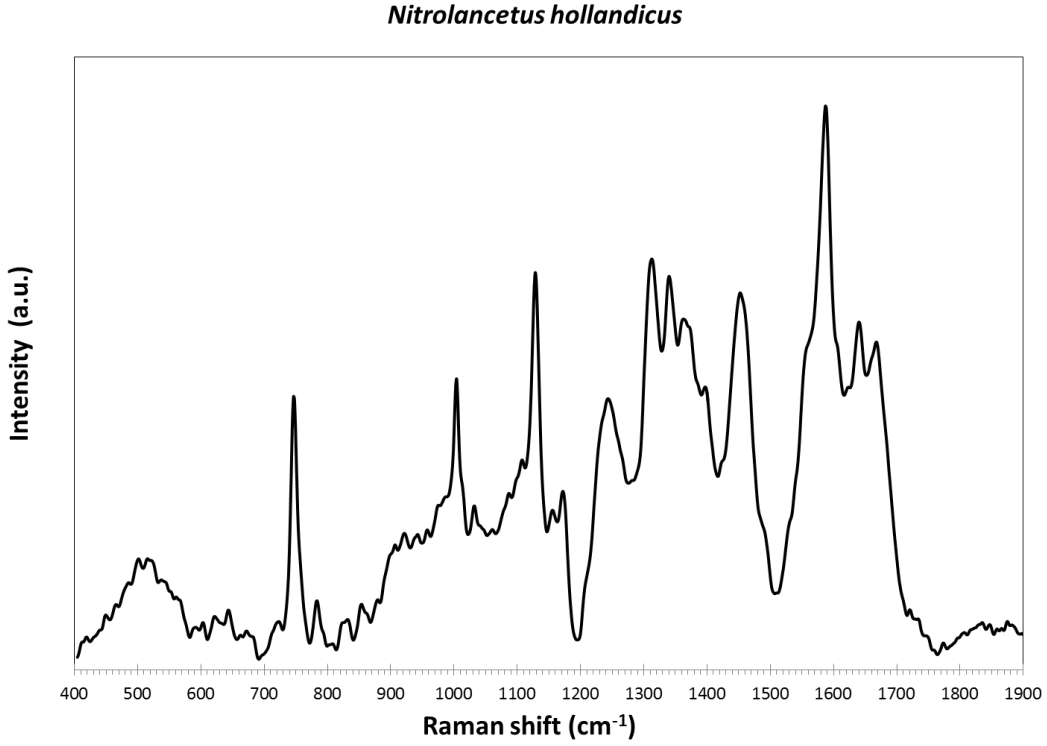

Figure 8.32. Mean Raman spectrum of *Nitrolancetus hollandicus*, (n = 16). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.10 *Planctomycetes*

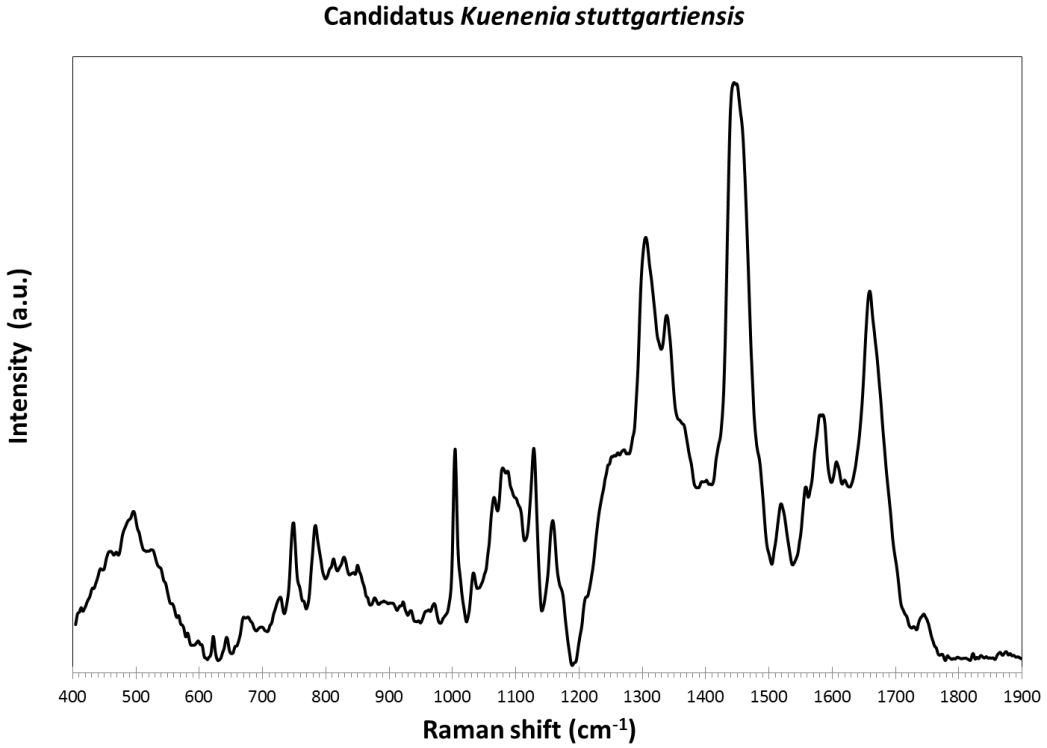

Figure 8.33. Mean Raman spectrum of Candidatus *Kuenenia stuttgartiensis*, (n = 14). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.11 *Thaumarchaeota*

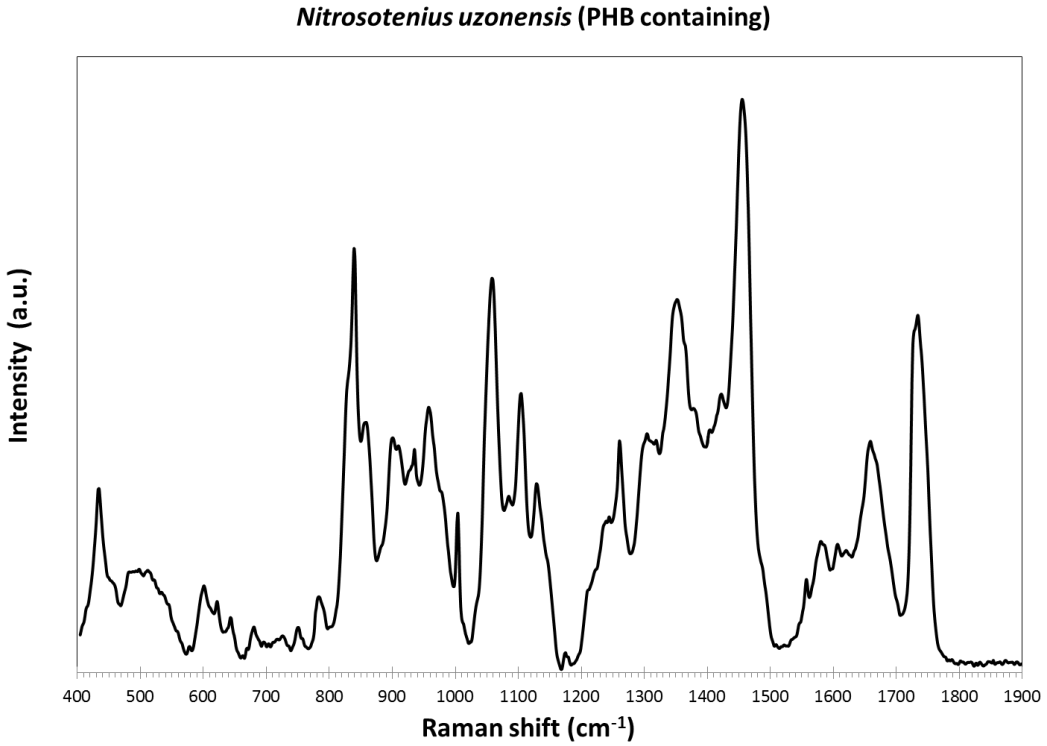

Figure 8.34. Mean Raman spectrum of *Nitrosotenius uzonensis* containing PHB, (n = 8). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

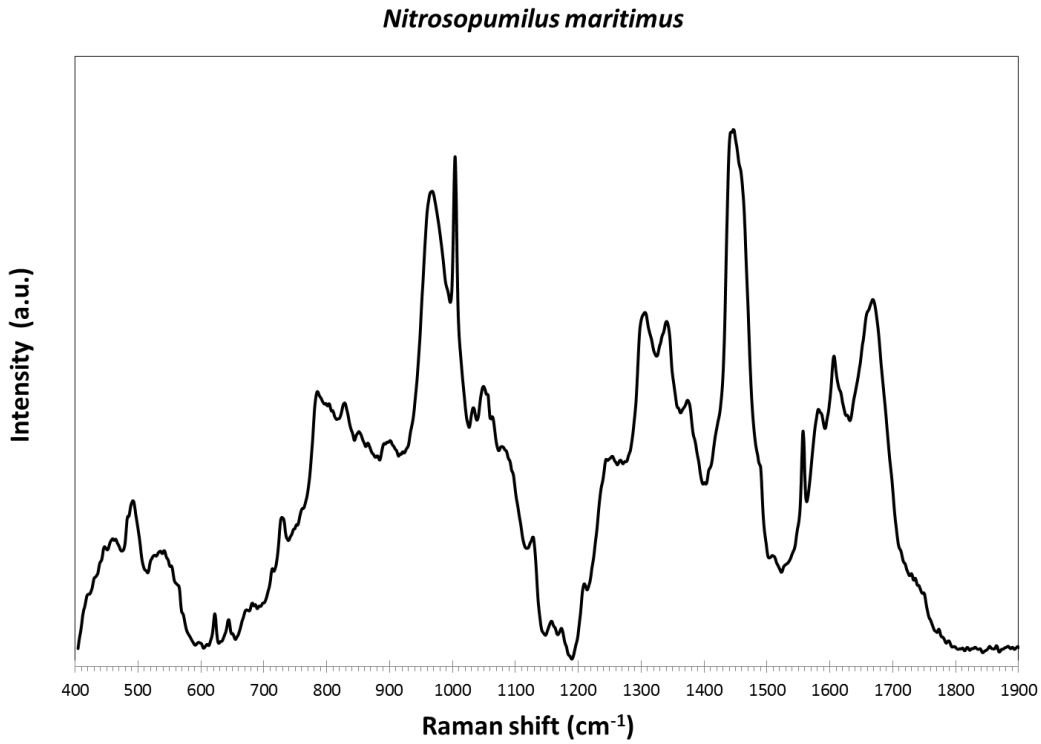

Figure 8.35. Mean Raman spectrum of *Nitrosopumilus maritimus*, (n = 11). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

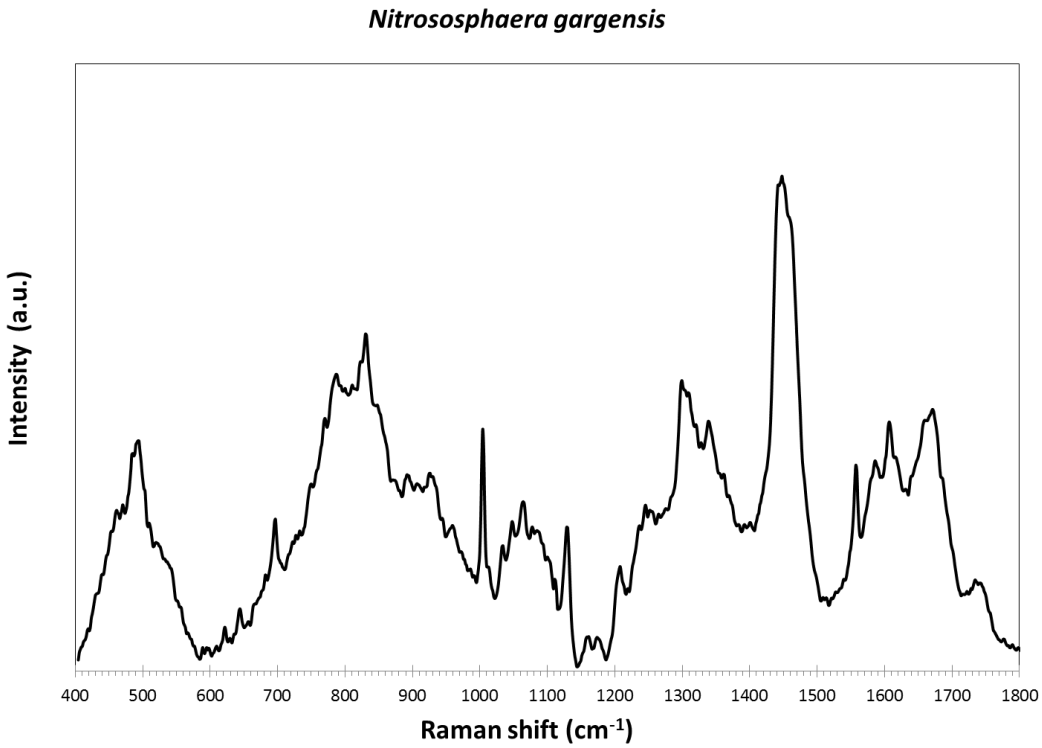

Figure 8.36. Mean Raman spectrum of *Nitrososphaera gargensis*, (n = 8). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

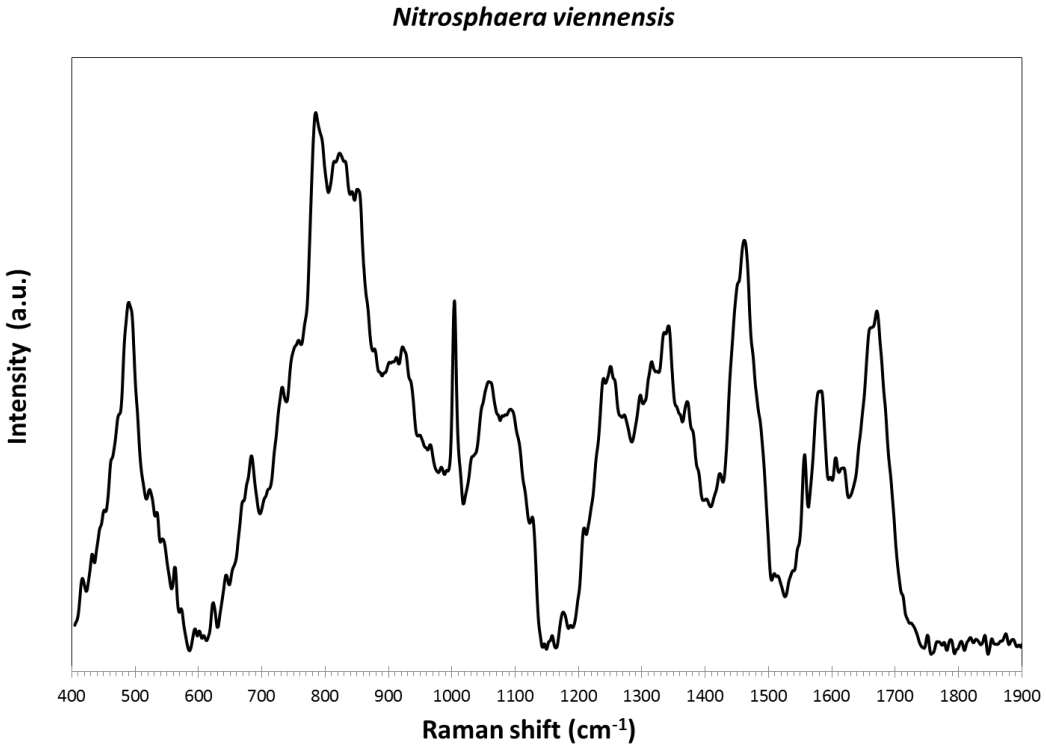

Figure 8.37. Mean Raman spectrum of *Nitrososphaera viennensis*, (n = 8). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.12 *Thermotogae*

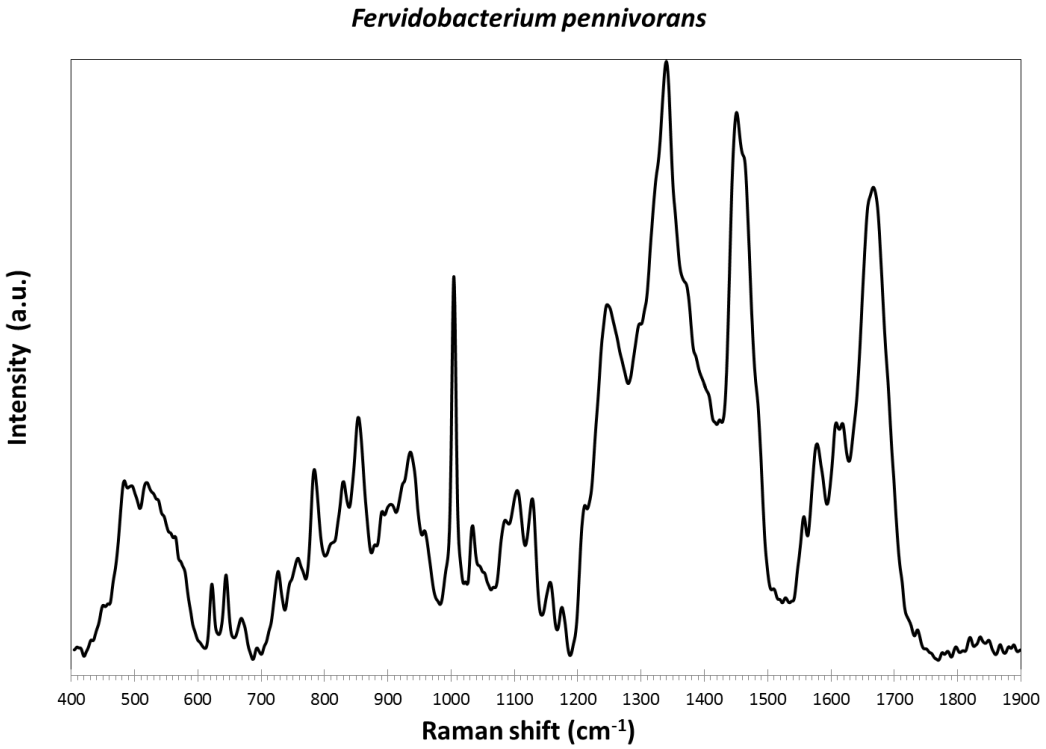

Figure 8.38. Mean Raman spectrum of *Fervidobacterium pennivorans*, (n = 15). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

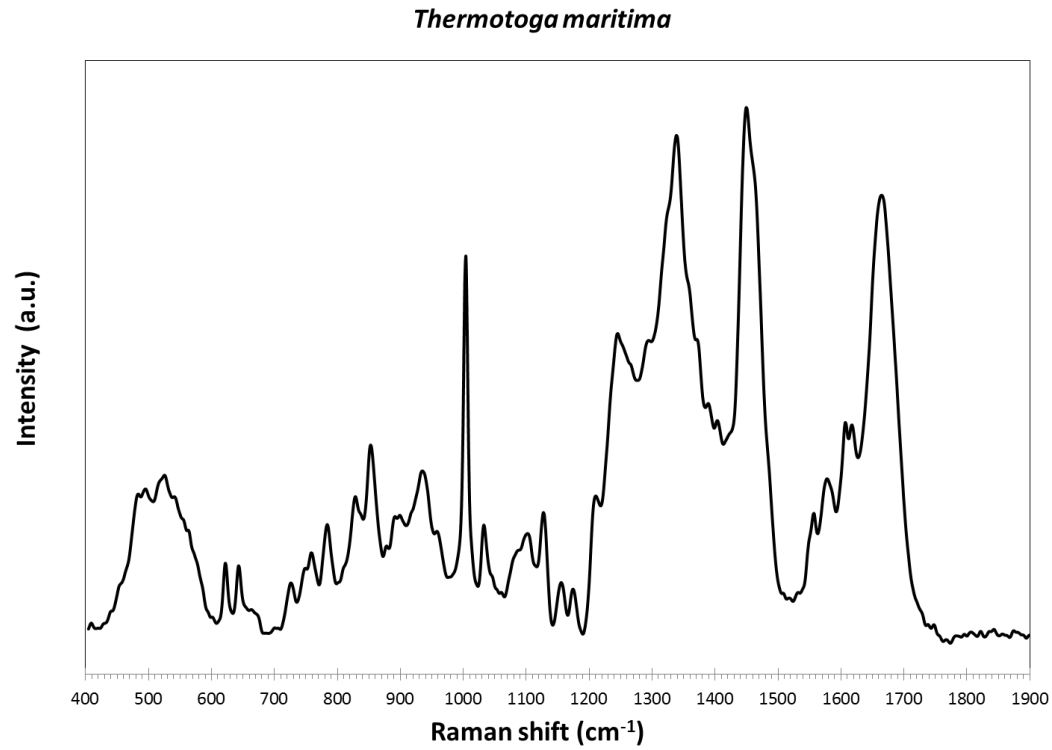

Figure 8.39. Mean Raman spectrum of *Thermotoga maritima*, (n = 16). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

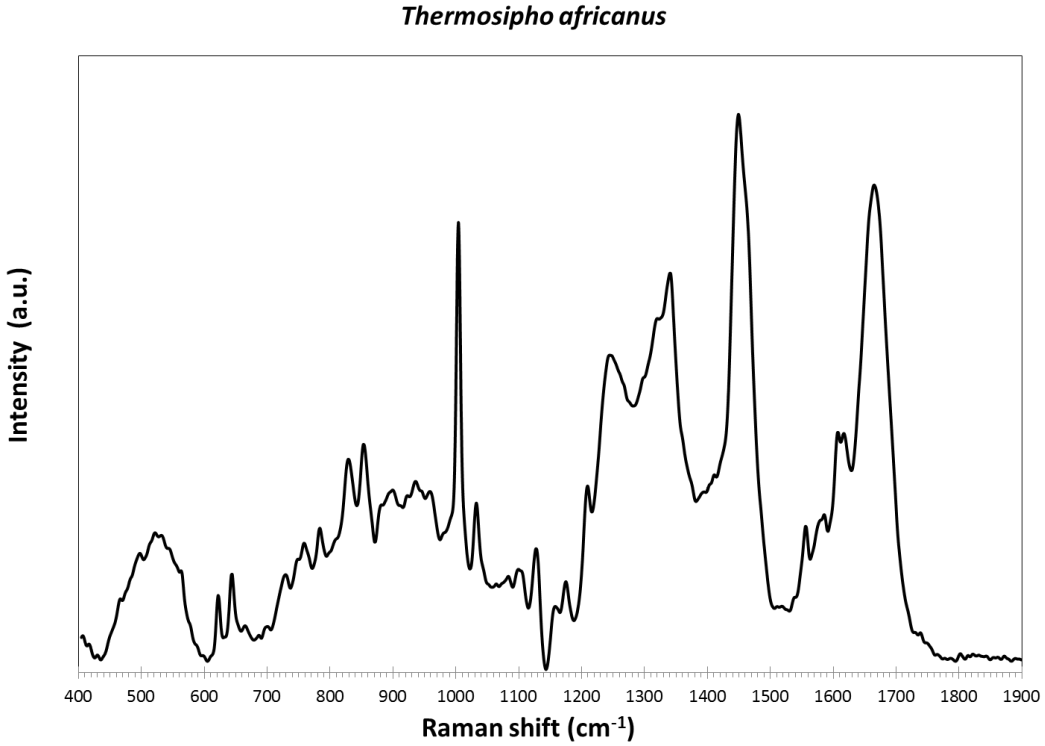

Figure 8.40. Mean Raman spectrum of *Thermosipho africanus*, (n = 14). The data were phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.13 arctic AOA enrichment SV8-6

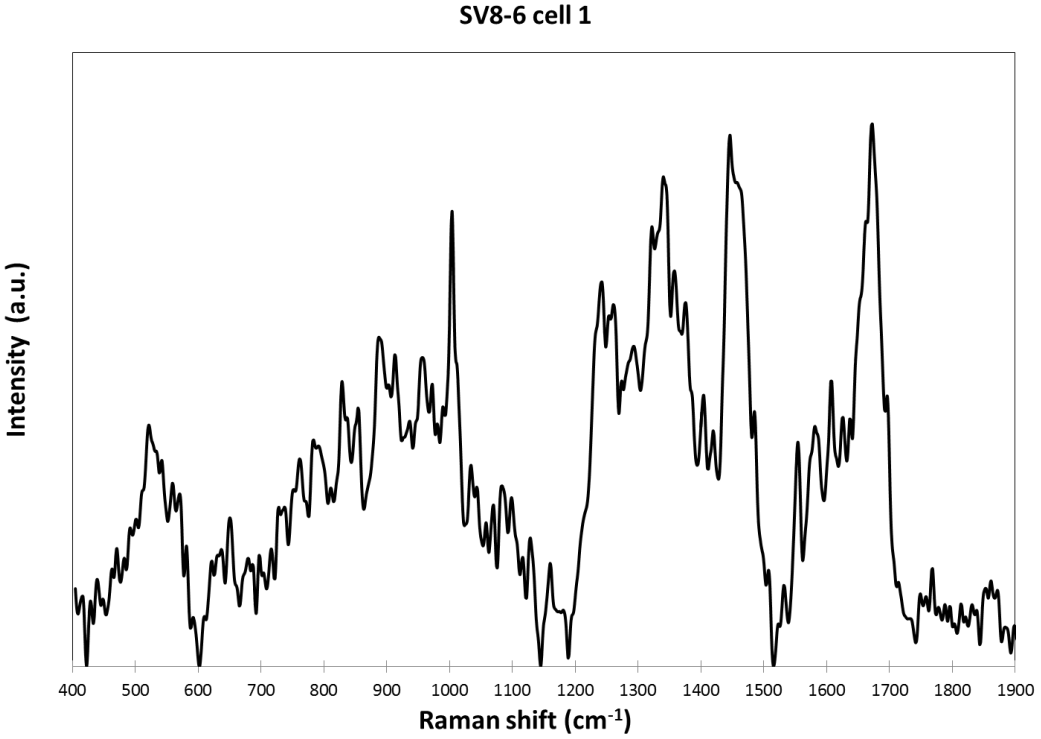

Figure 8.41. Raman spectrum of SV8-6 cell 1. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

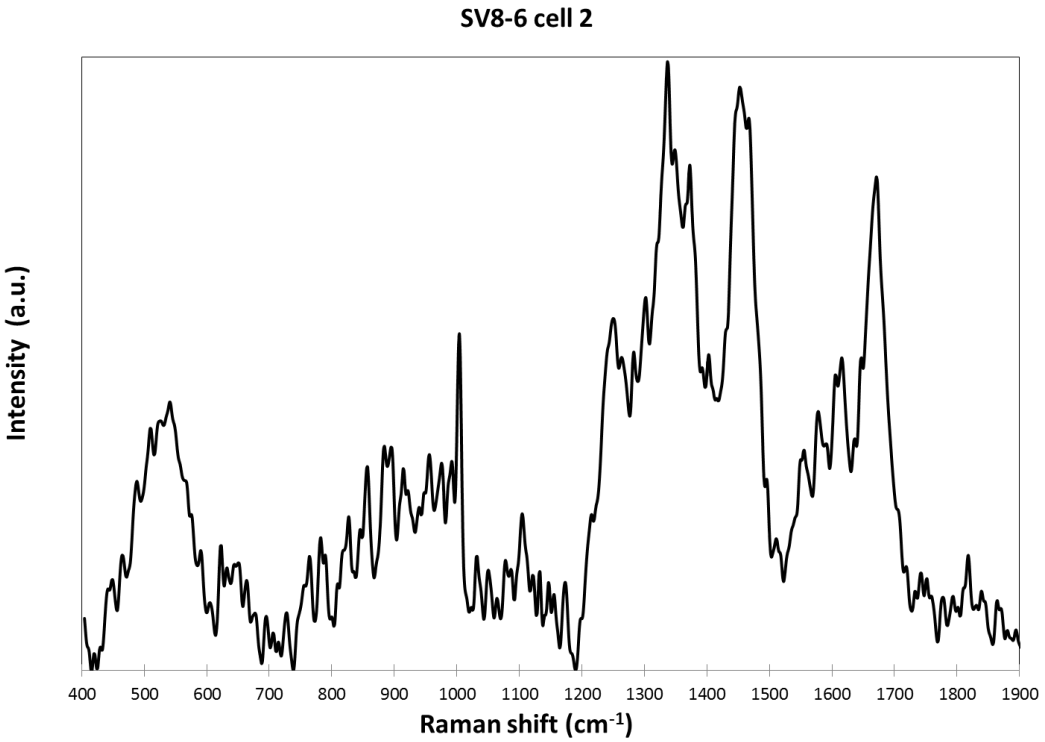

Figure 8.42. Raman spectrum of SV8-6 cell 2. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

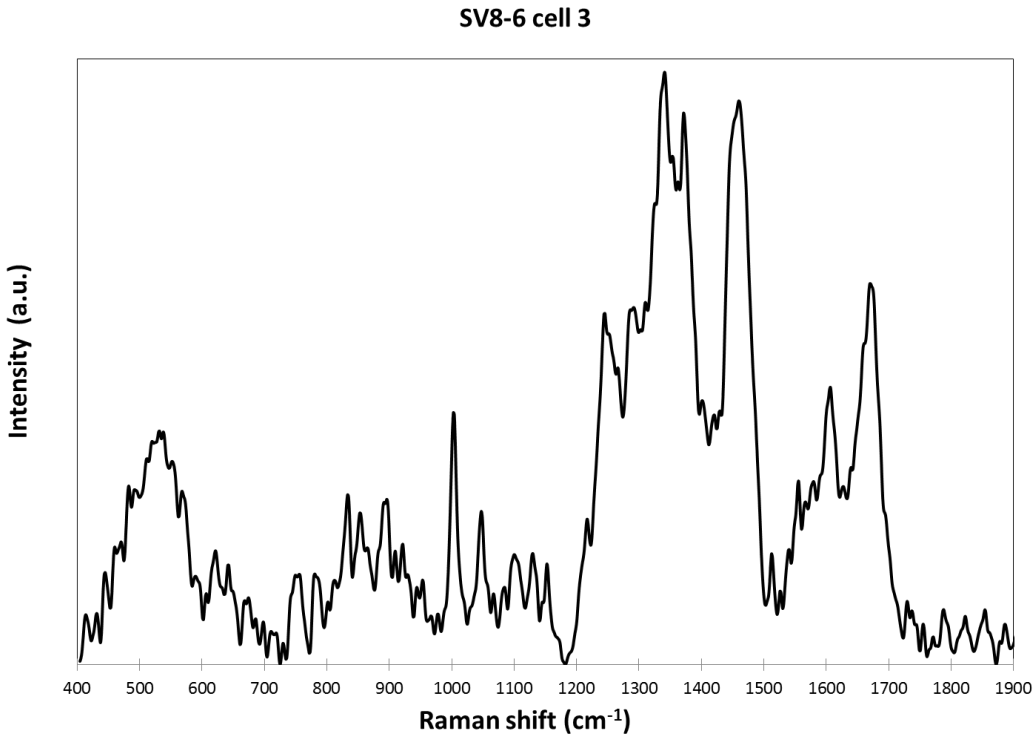

Figure 8.43. Raman spectrum of SV8-6 cell 3. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

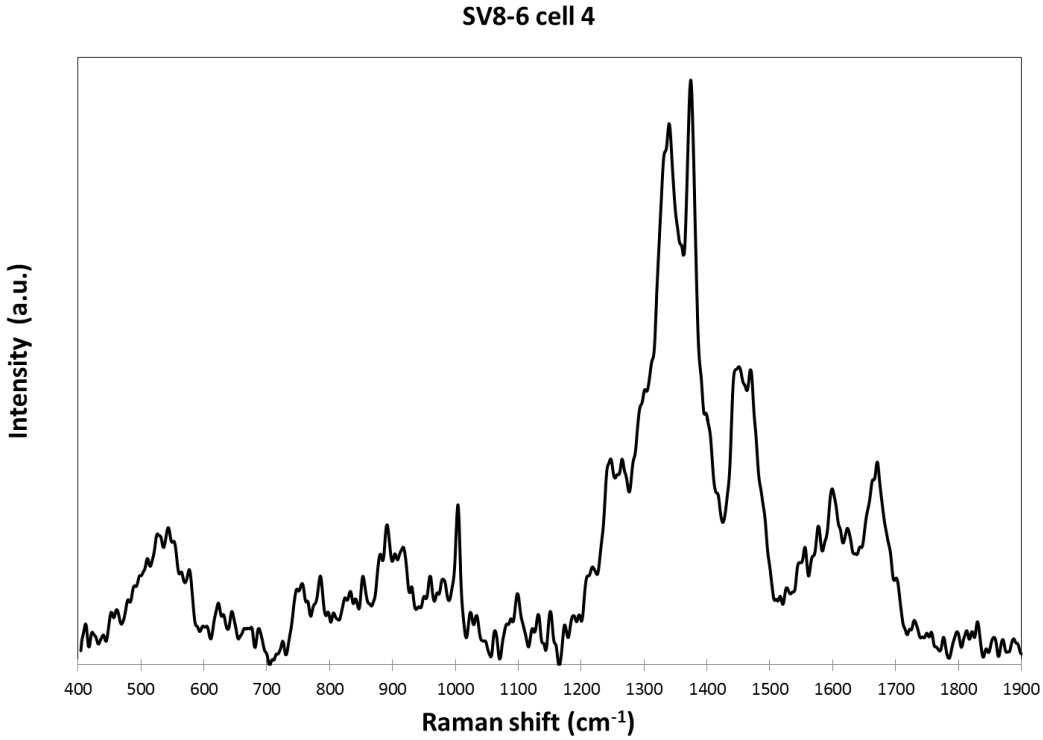

Figure 8.44. Raman spectrum of SV8-6 cell 4. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

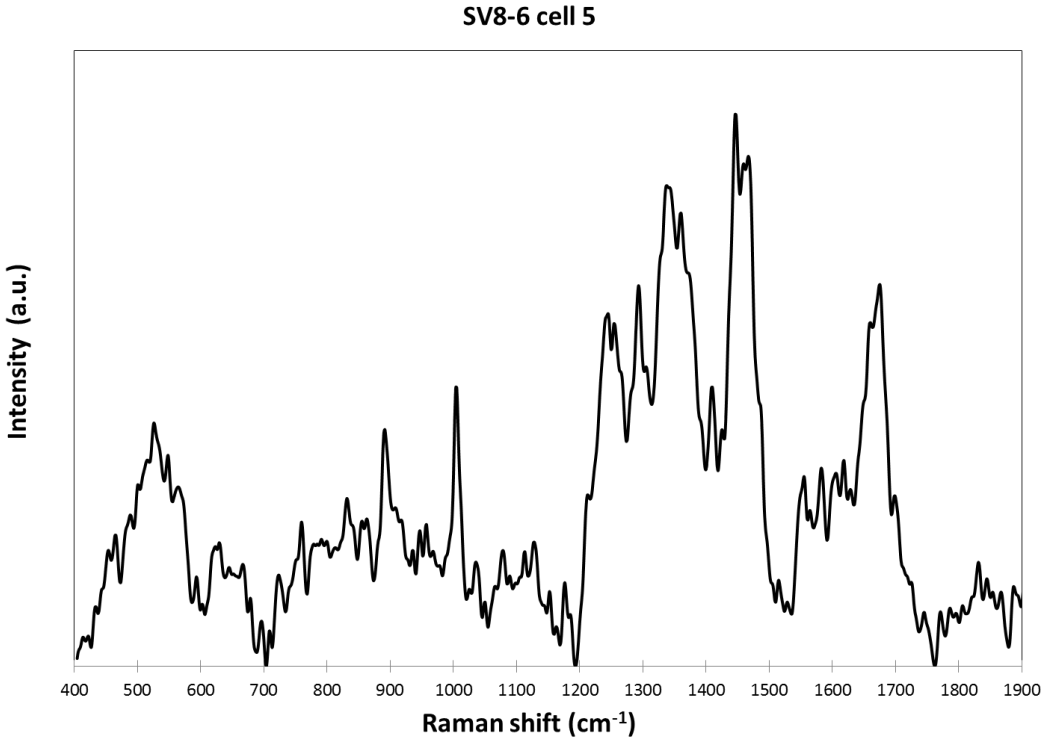

Figure 8.45. Raman spectrum of SV8-6 cell 5. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

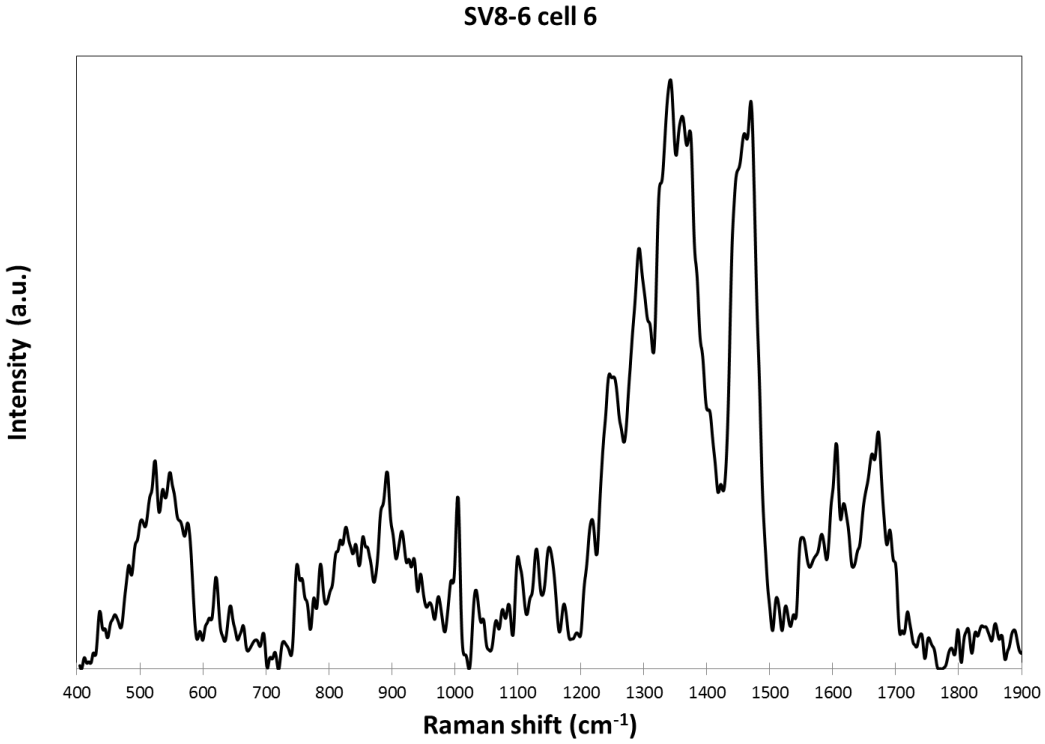

Figure 8.46. Raman spectrum of SV8-6 cell 6. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

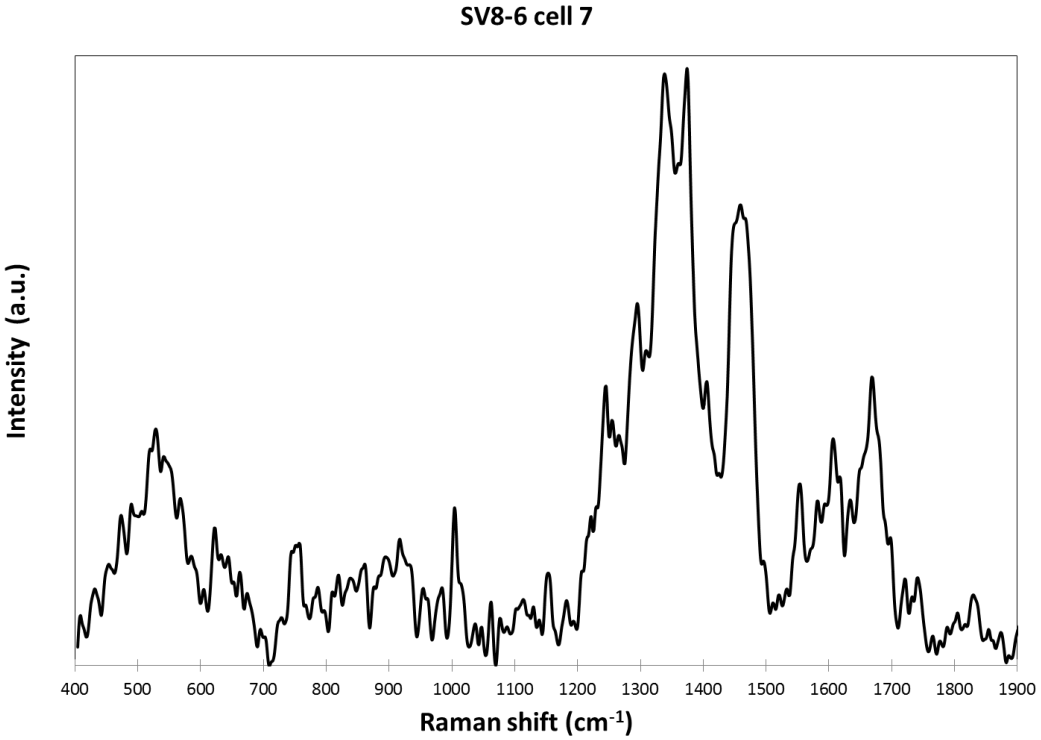

Figure 8.47. Raman spectrum of SV8-6 cell 7. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

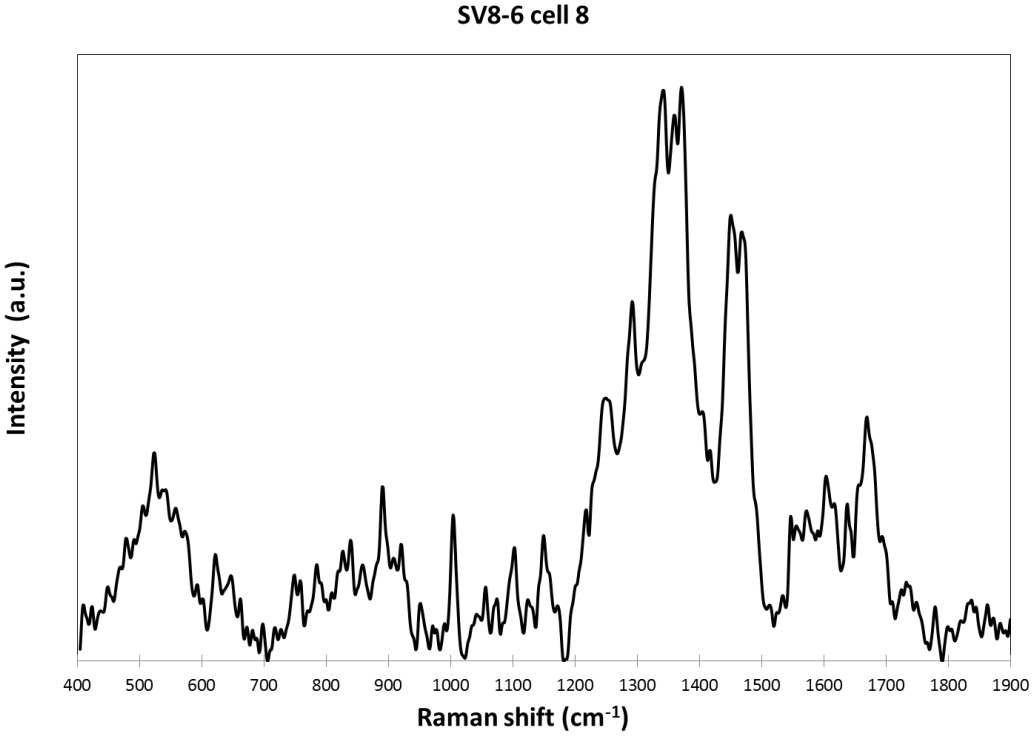

Figure 8.48. Raman spectrum of SV8-6 cell 8. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

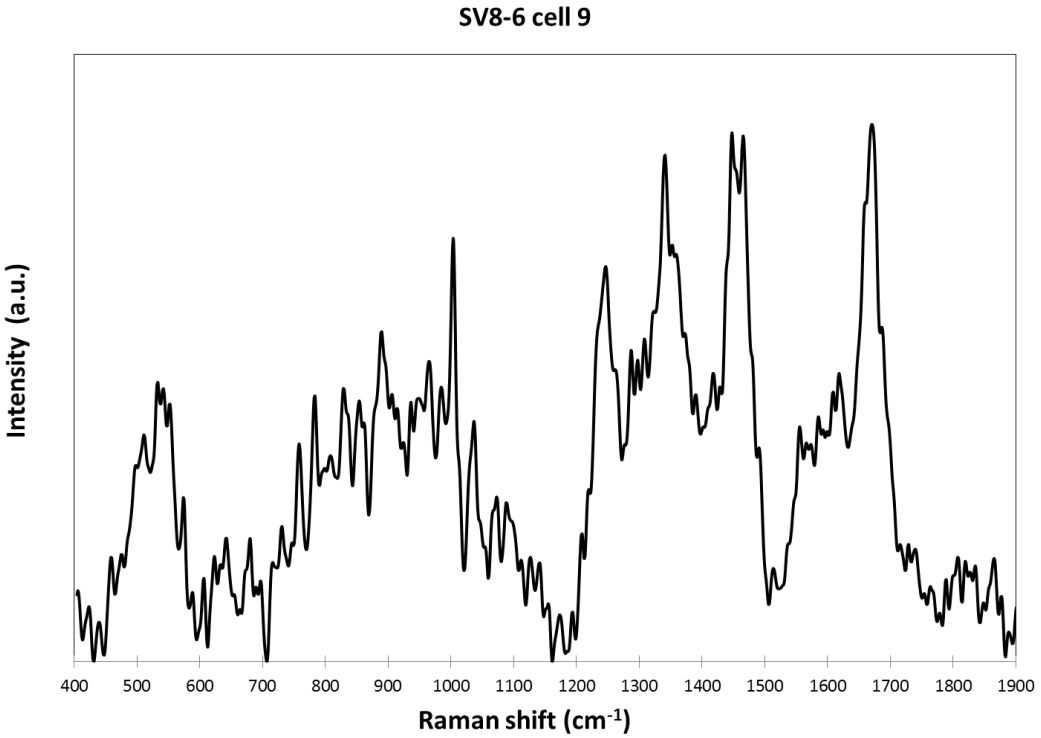

Figure 8.49. Raman spectrum of SV8-6 cell 9. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

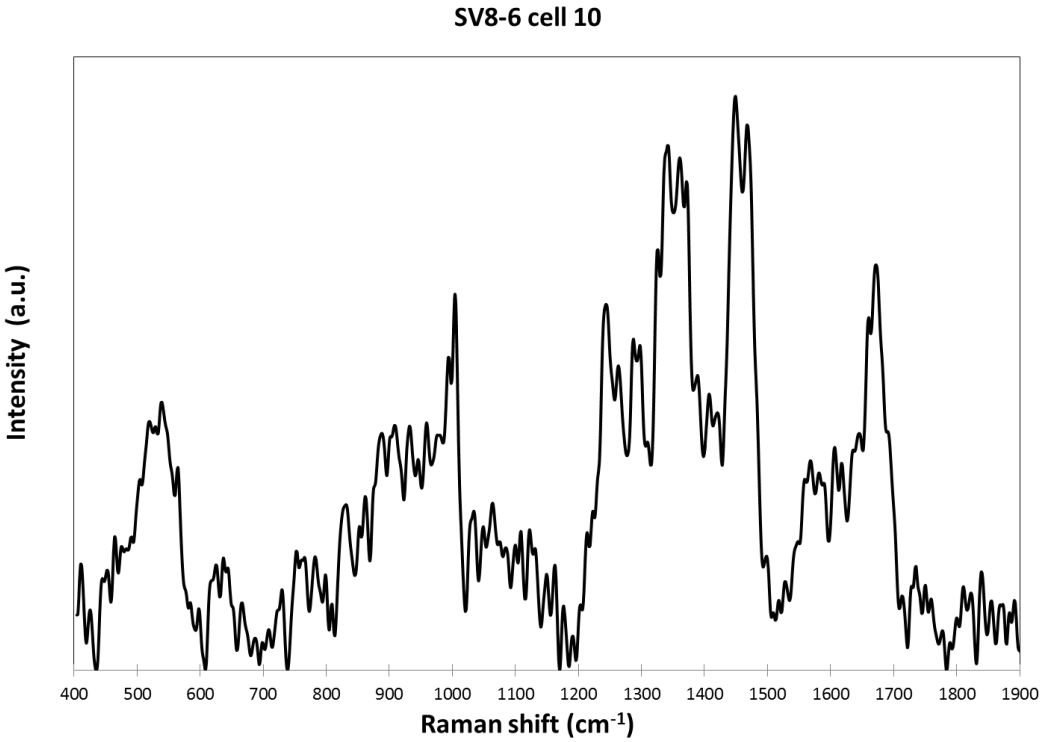

Figure 8.50. Raman spectrum of SV8-6 cell 10. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

8.14 arctic AOA enrichment SV9-19

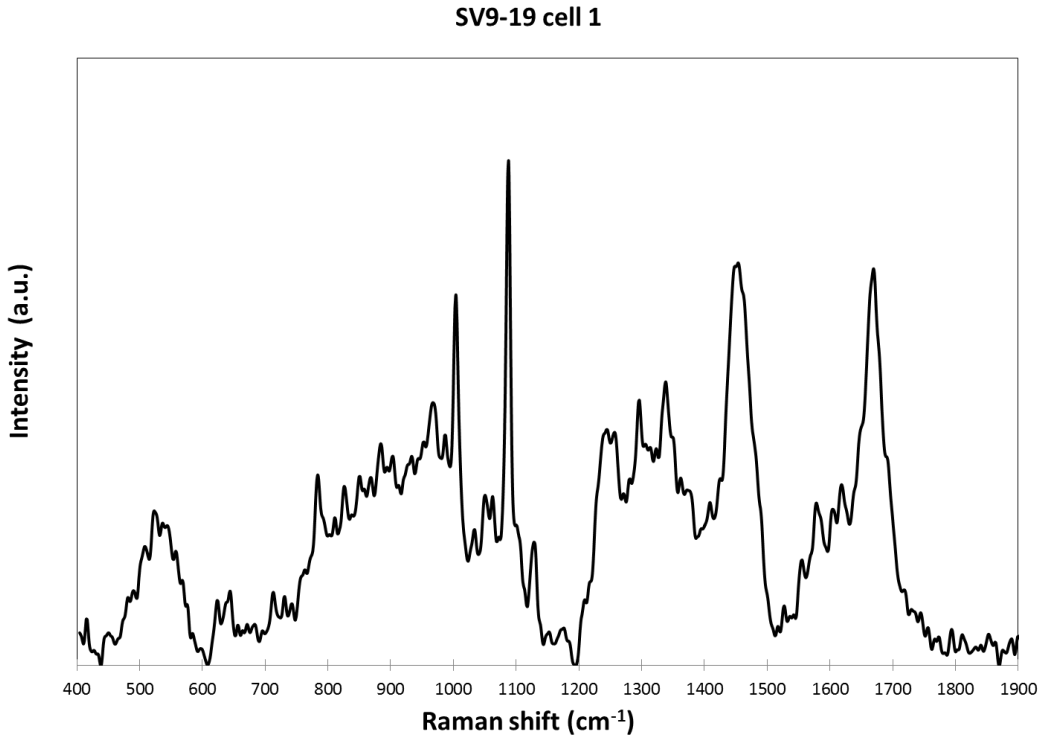

Figure 8.51. Raman spectrum of SV9-19 cell 1. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

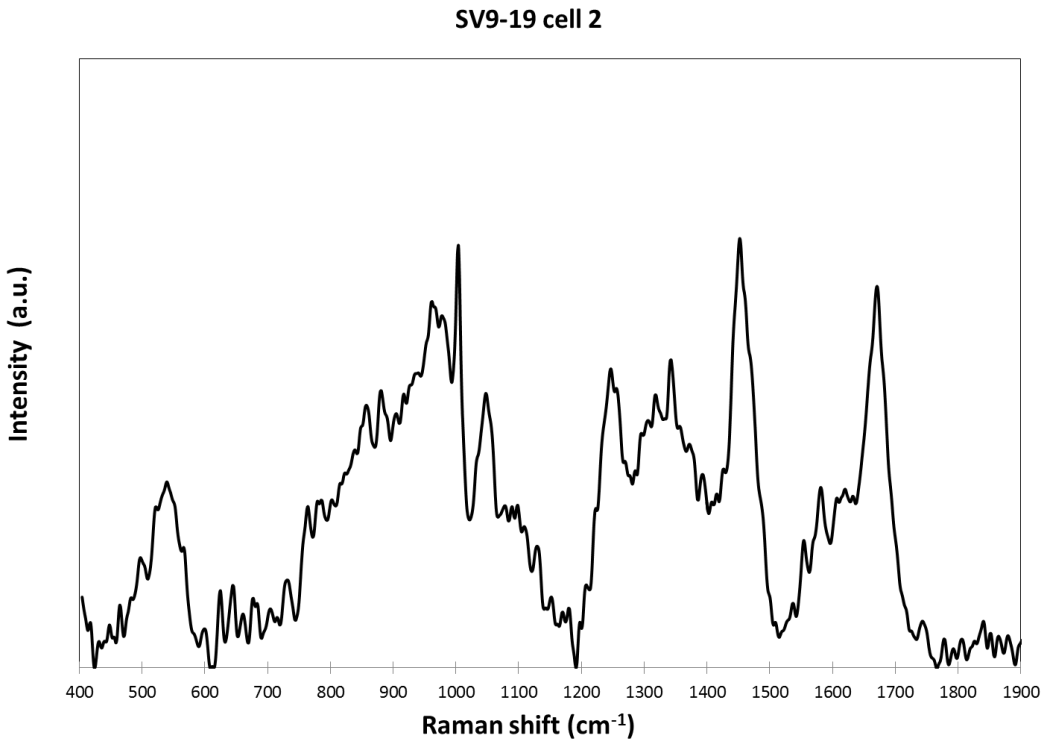

Figure 8.52. Raman spectrum of SV9-19 cell 2. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

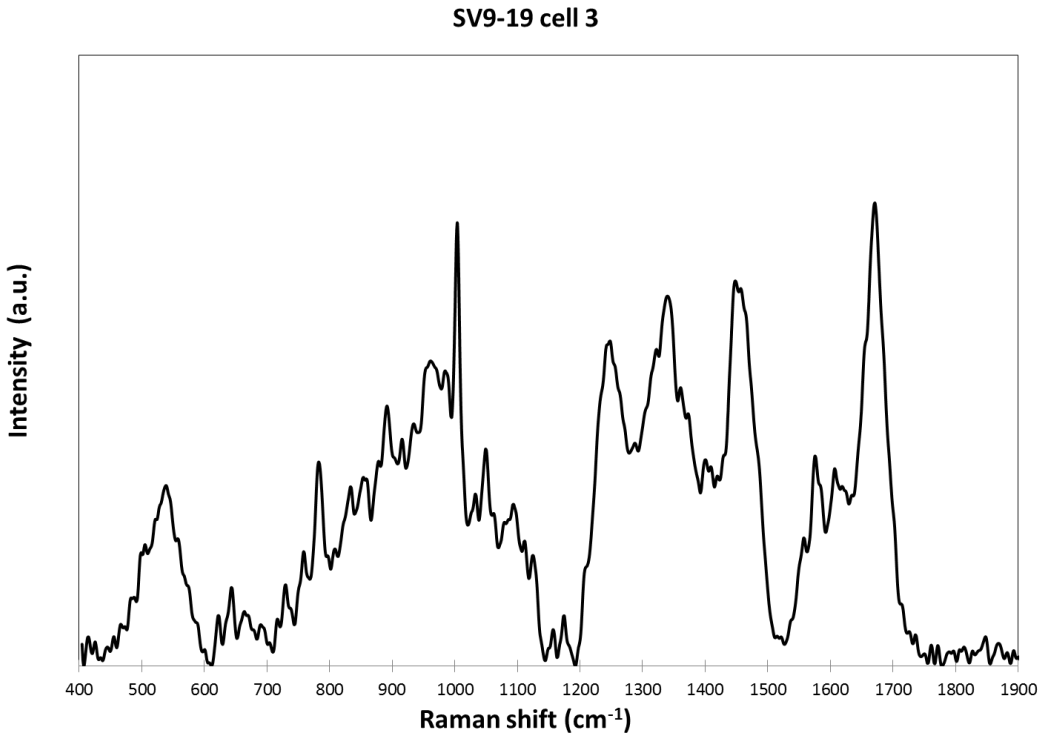

Figure 8.53. Raman spectrum of SV9-19 cell 3. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

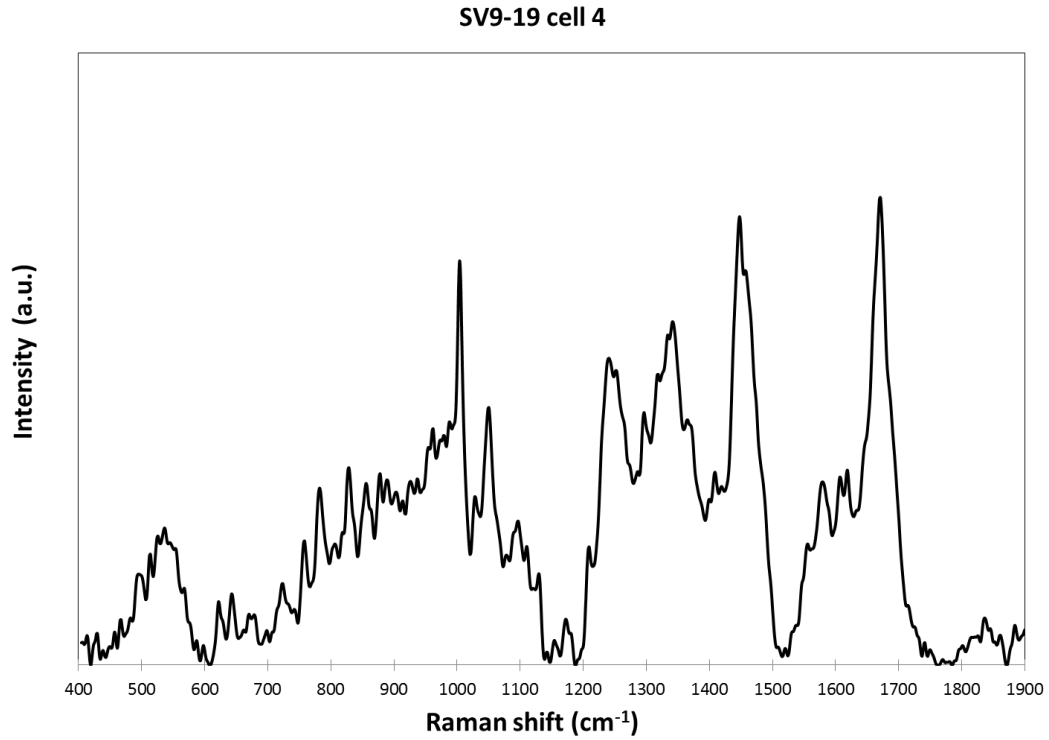

Figure 8.54. Raman spectrum of SV9-19 cell 4. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

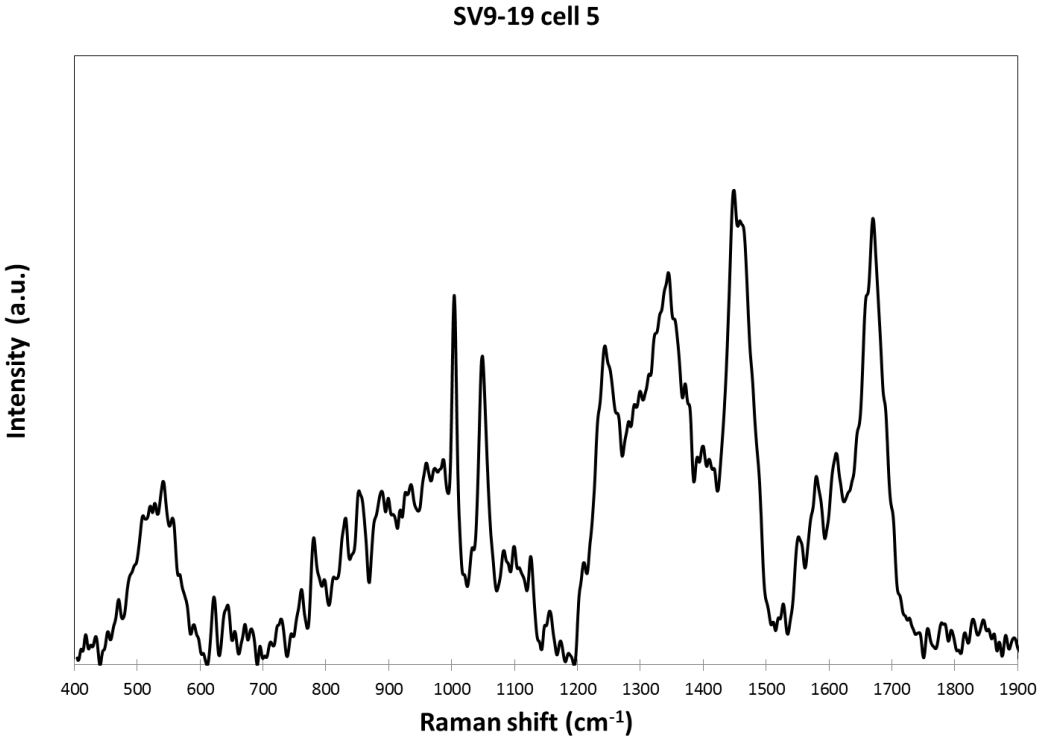

Figure 8.55. Raman spectrum of SV9-19 cell 5. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

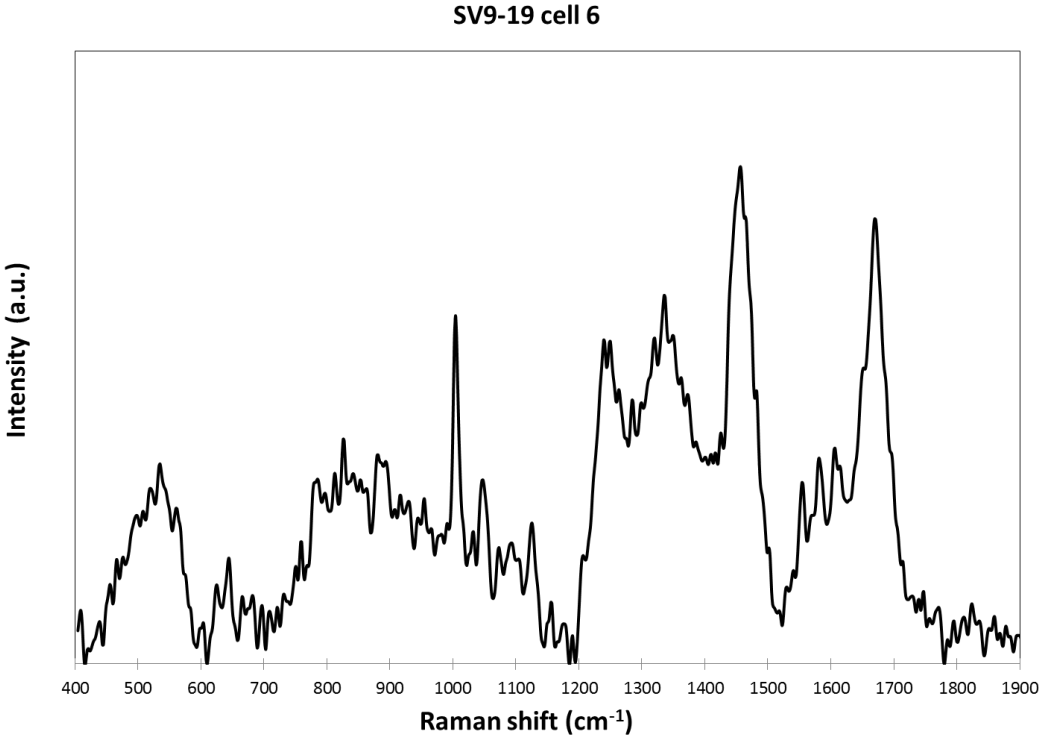

Figure 8.56. Raman spectrum of SV9-19 cell 6. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

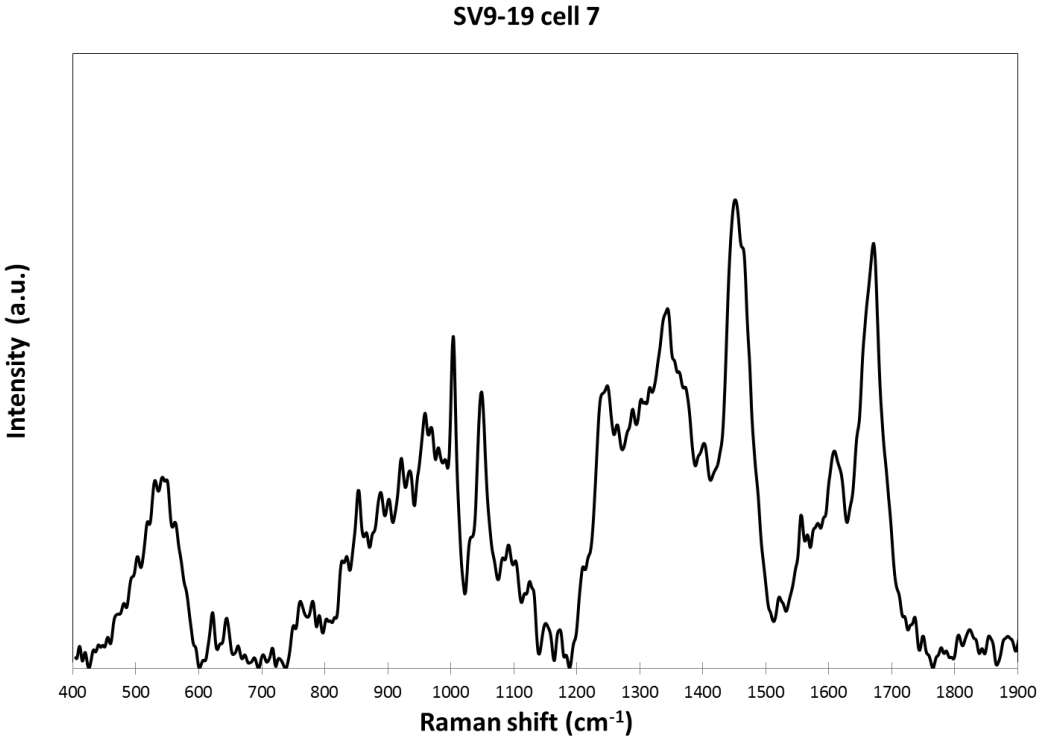

Figure 8.57. Raman spectrum of SV9-19 cell 7. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

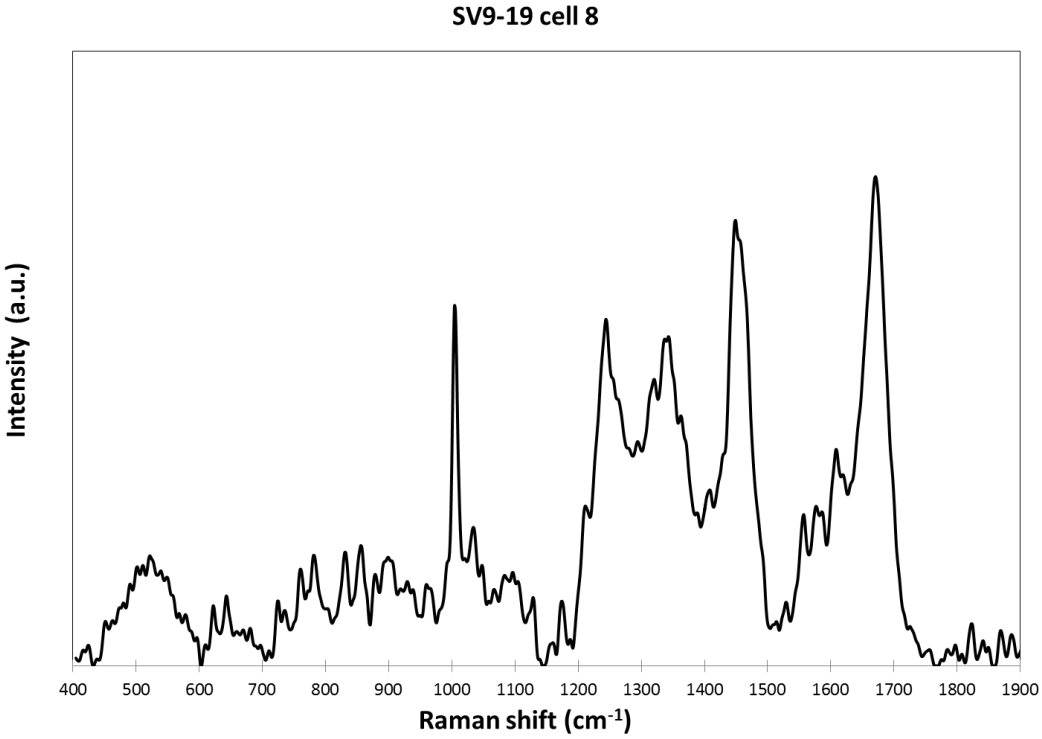

Figure 8.58. Raman spectrum of SV9-19 cell 8. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

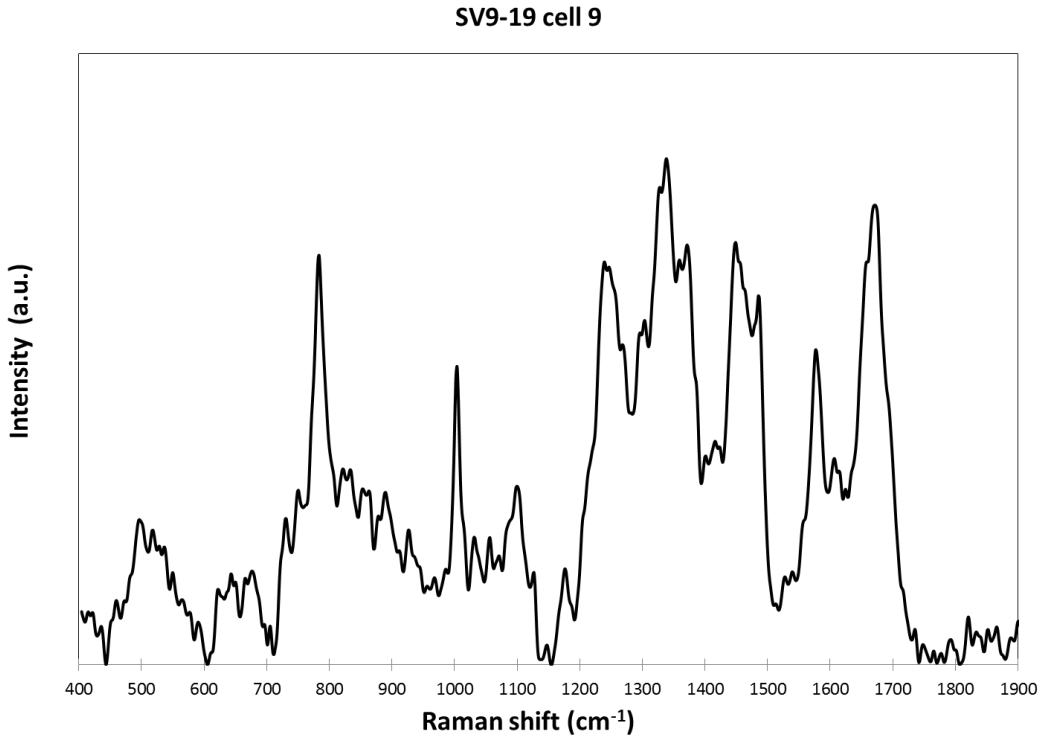

Figure 8.59. Raman spectrum of SV9-19 cell 9. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

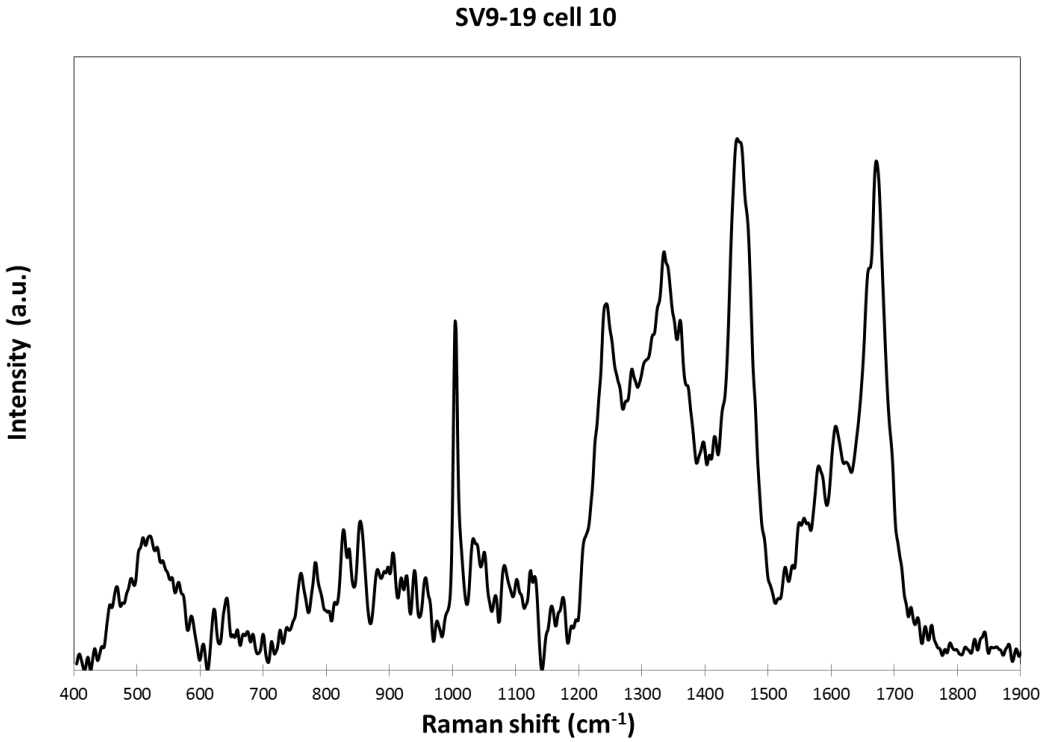

Figure 8.60. Raman spectrum of SV9-19 cell 10. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

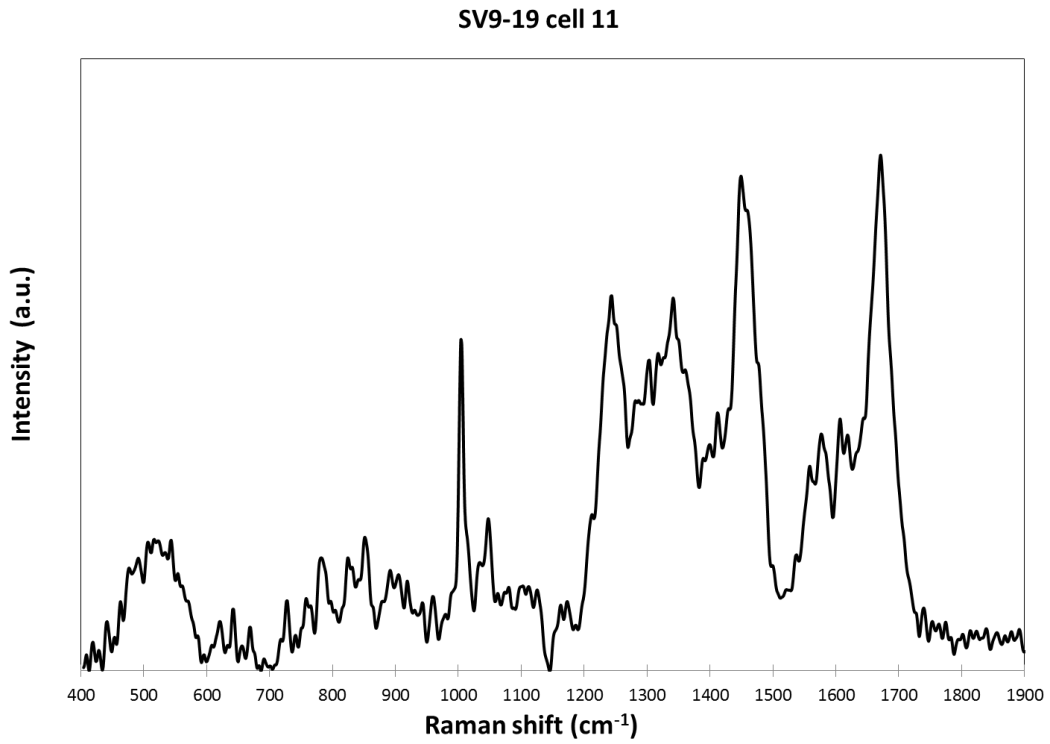

Figure 8.61. Raman spectrum of SV9-19 cell 11. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

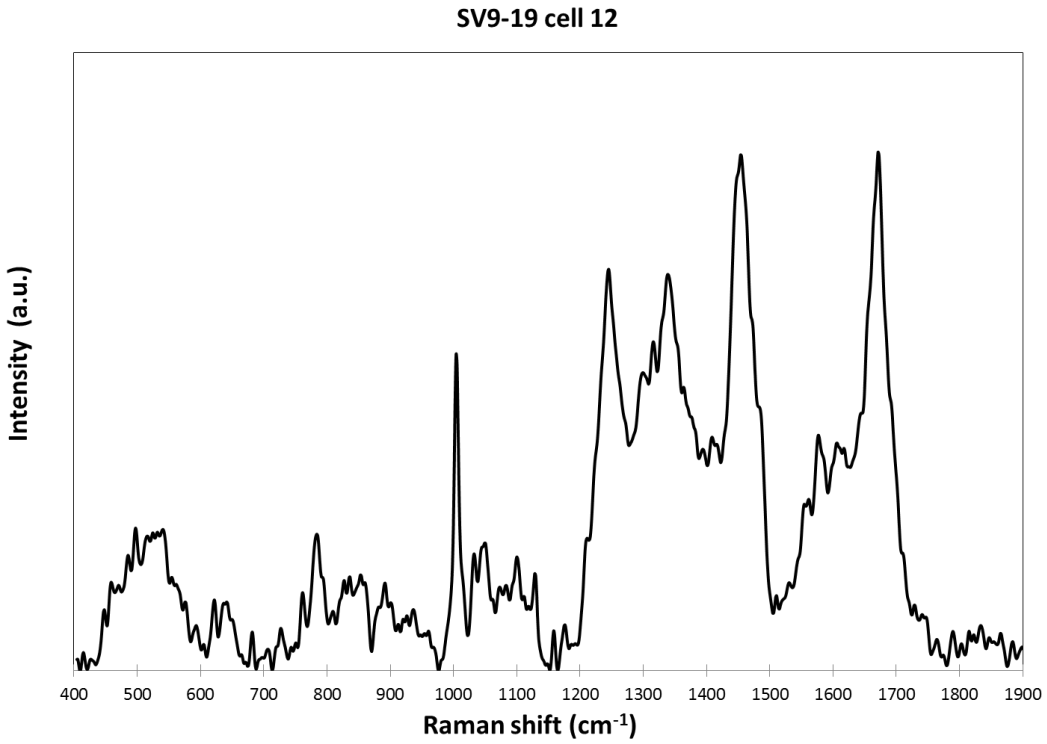

Figure 8.62. Raman spectrum of SV9-19 cell 12. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

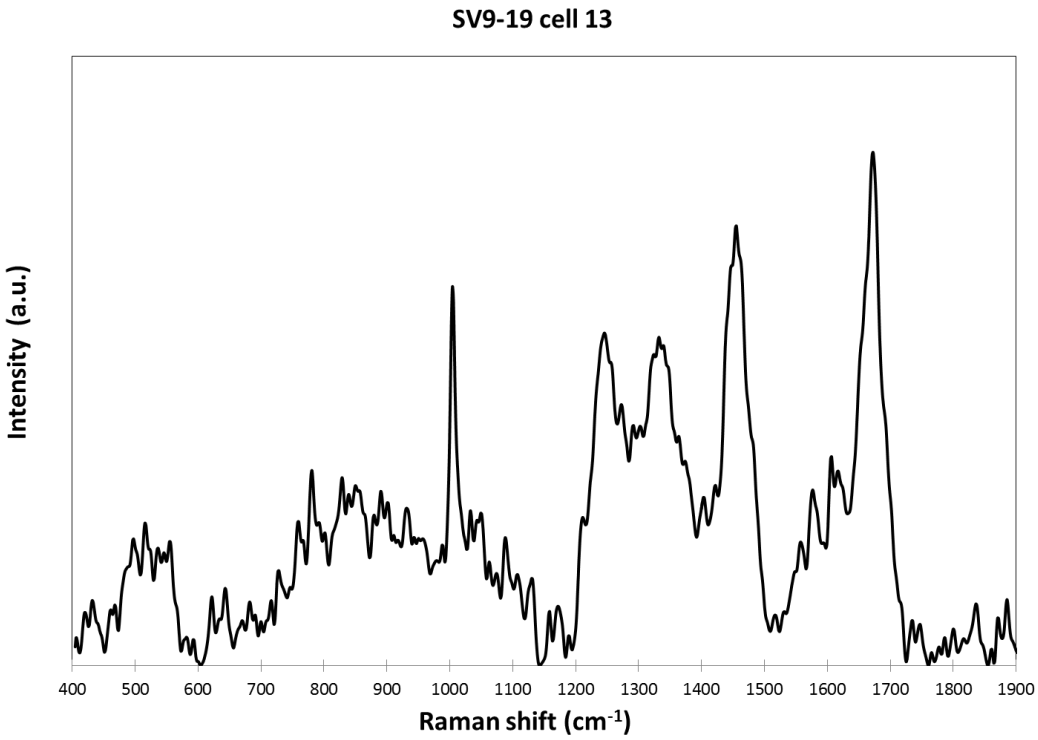

Figure 8.63. Raman spectrum of SV9-19 cell 13. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

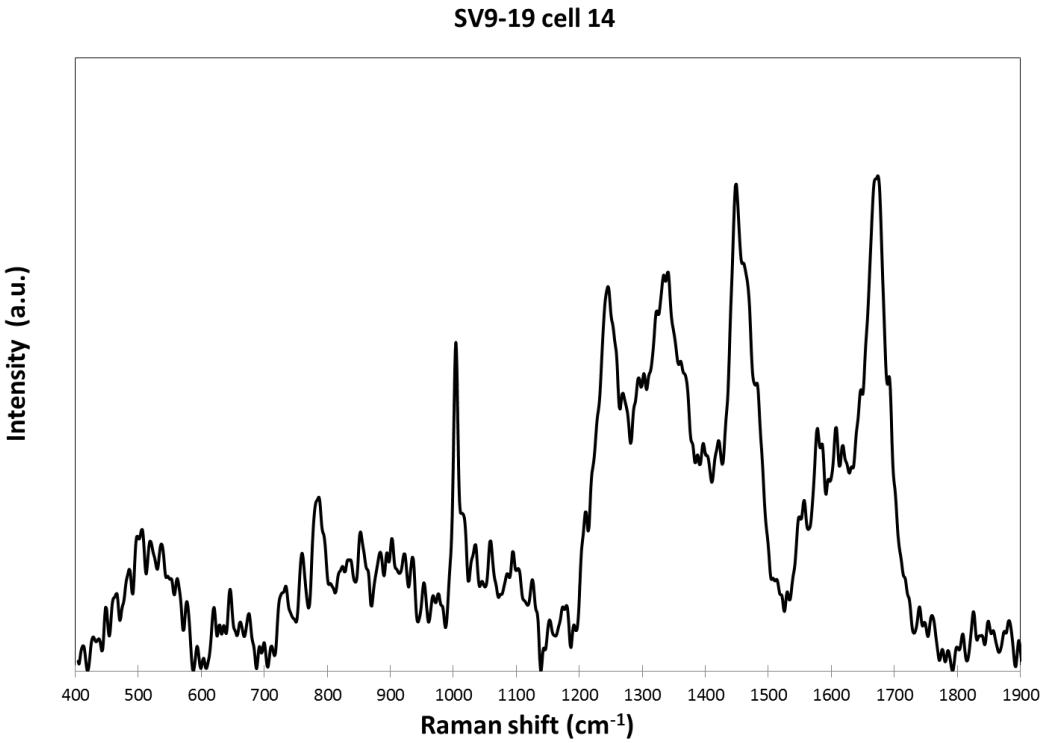

Figure 8.64. Raman spectrum of SV9-19 cell 14. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

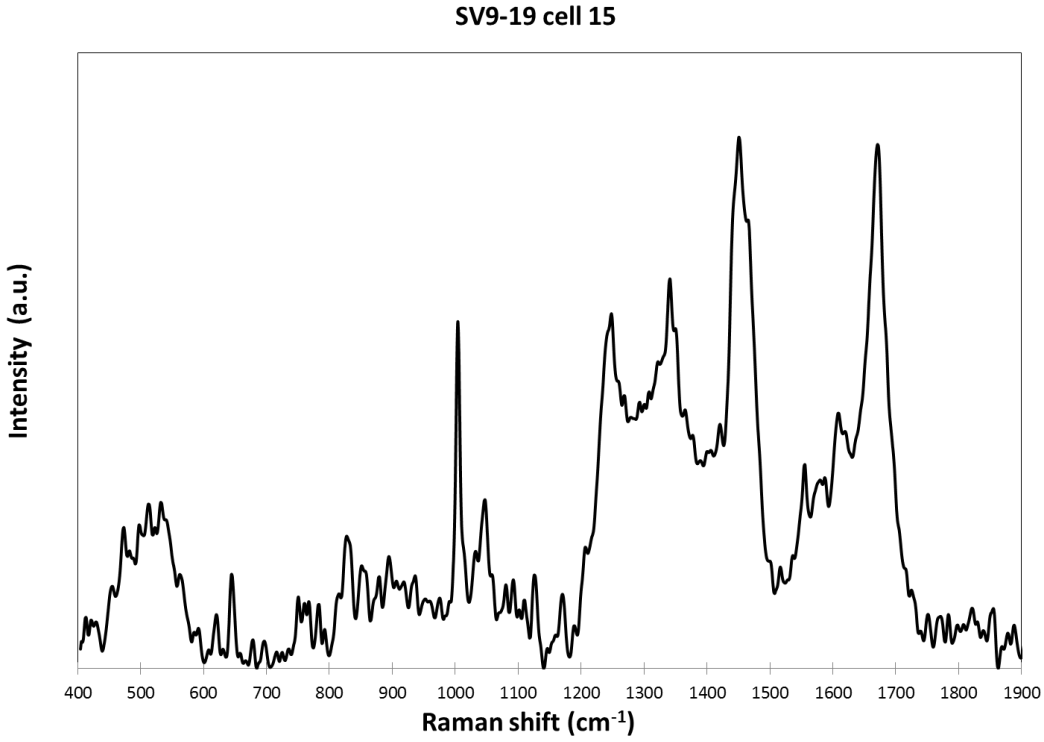

Figure 8.65. Raman spectrum of SV9-19 cell 15. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

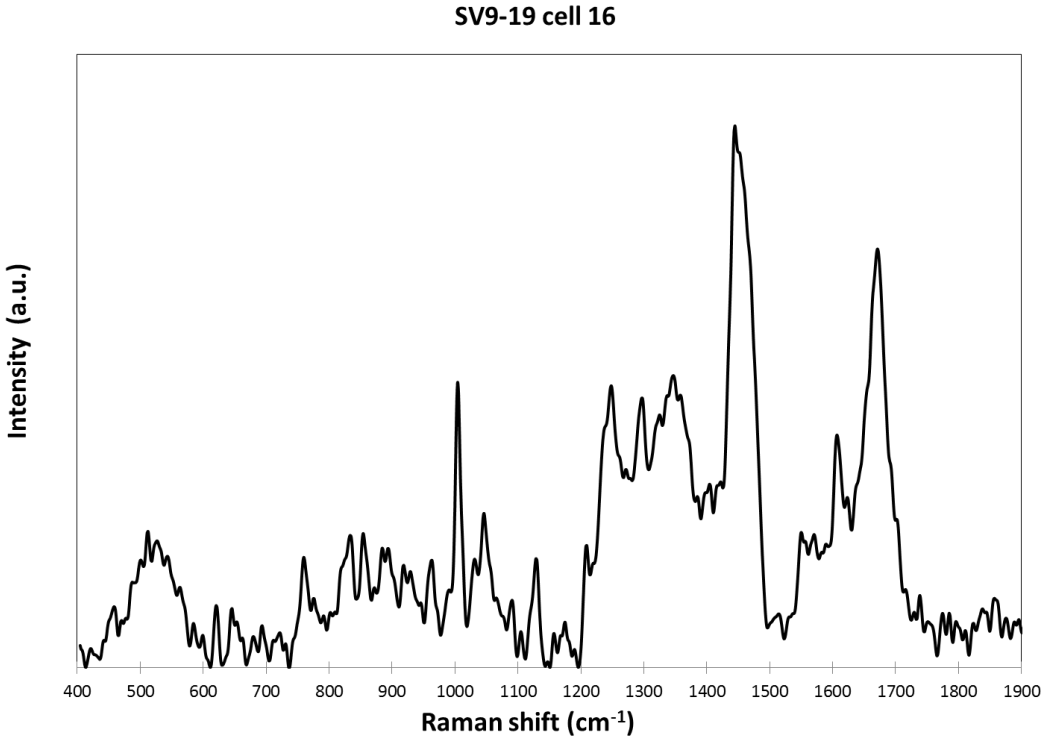

Figure 8.66. Raman spectrum of SV9-19 cell 16. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

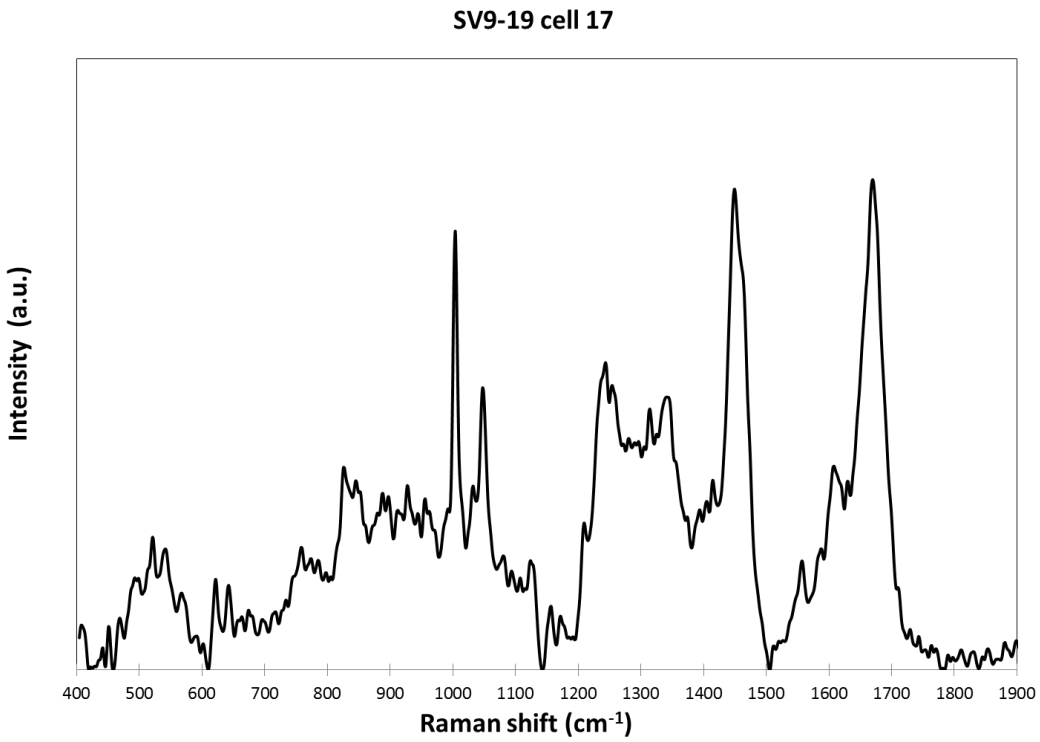

Figure 8.67. Raman spectrum of SV9-19 cell 17. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

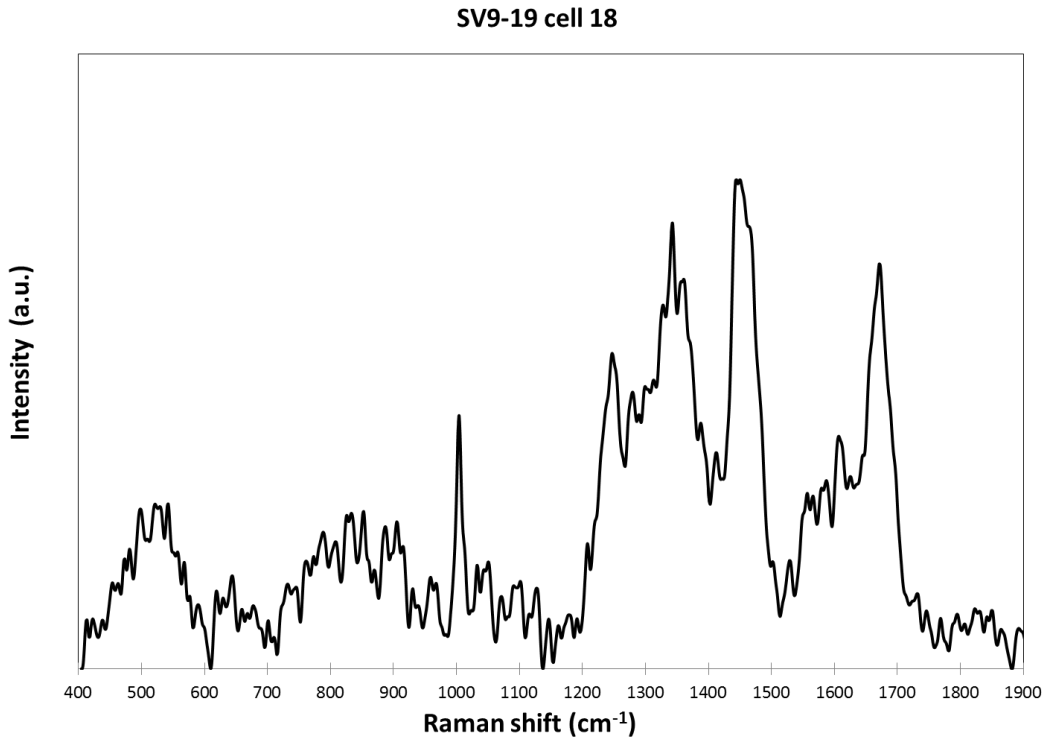

Figure 8.68. Raman spectrum of SV9-19 cell 18. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

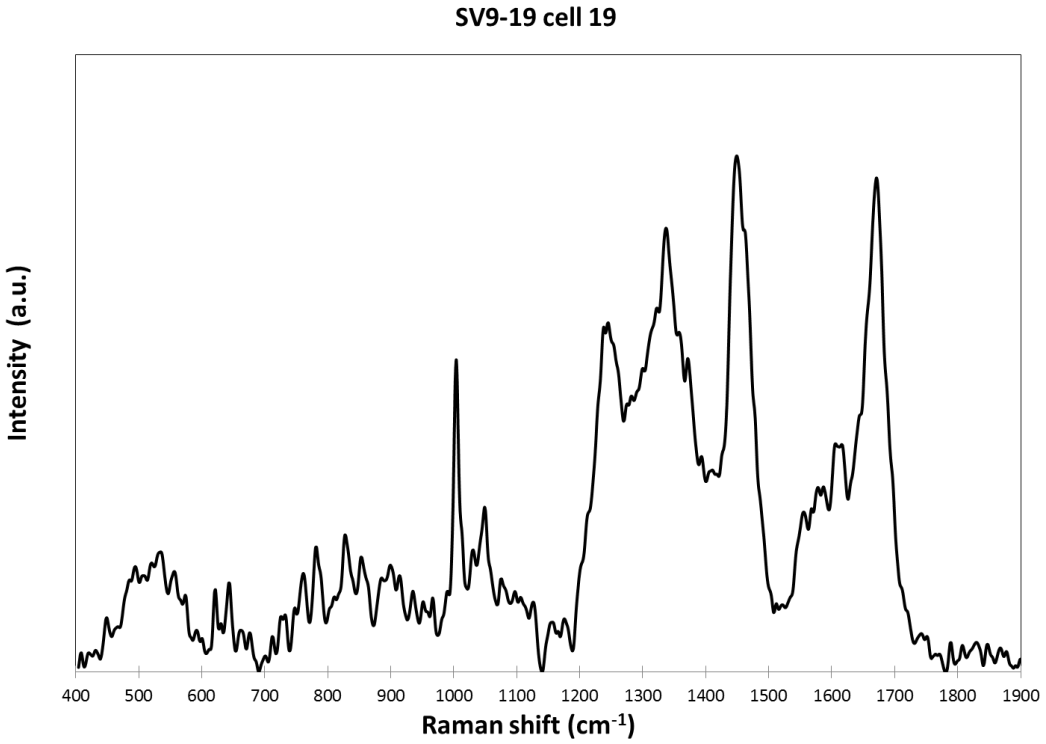

Figure 8.69. Raman spectrum of SV9-19 cell 19. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

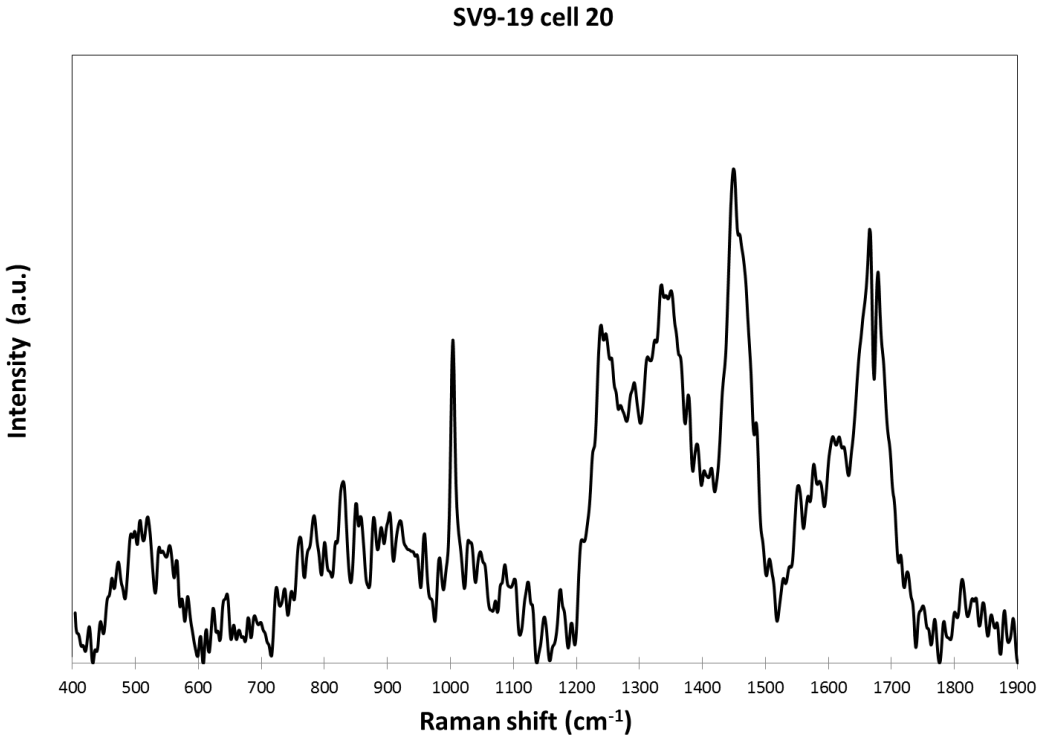

Figure 8.70. Raman spectrum of SV9-19 cell 20. The data was phenylalanine peak aligned, smoothed, line-segmented baselined (8th degree) and mean normalized (chapter 2.12).

9 References

Abril, D., Torra, V. and Navarro-Arribas. (2011). Supervised learning using Mahalanobis distance for record linkage. 6th international summer school on aggregation operators – AGOP 2011

Alvarez, H., and Steinbüchel, A. (2002). Triacylglycerols in prokaryotic microorganisms. Applied Microbiology and Biotechnology 60, 367–376.

Alvarez, H.M., Pucci, O.H., and Steinbüchel, A. (1997). Lipid storage compounds in marine bacteria. Applied Microbiology and Biotechnology 47, 132–139.

Alves R. (2011). Ammonia-oxidizing archaea from high Arctic soils. Master Thesis. Faculty of sciences, University of Lisbon. Department of genetics in ecology, university of Vienna.

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. 56, 1919–1925.

Agogué, H., Brink M., Dinasquet J. and Herndl G.J. (2008) Major gradients in putatively nitrifying and non-nitrifying Archaea in the deep North Atlantic. Nature 456, 788-791.

Arun, K., and Langmead, C. (2005). Structure based chemical shift prediction using Random Forests non-linear regression (Computer Science Department).

Balk, M., Heilig, H.G.H.J., van Eekert, M.H.A., Stams, A.J.M., Rijpstra, I.C., Sinninghe-Damsté, J.S., de Vos, W.M., and Kengen, S.W.M. (2009). Isolation and characterization of a new CO-utilizing strain, *Thermoanaerobacter thermohydrosulfuricus* subsp. carboxydovorans, isolated from a geothermal spring in Turkey. Extremophiles 13, 885–894.

Baeten, V., Hourant, P., Morales, M.T., and Aparicio, R. (1998). Oil and Fat Classification by FT-Raman Spectroscopy. Journal of Agricultural and Food Chemistry 46, 2638–2646.

Barbarossa, V., Galluzzi, F., Tomaciello, R., and Zanobi, A. (1991). Raman spectra of microcrystalline graphite and a-C:H films excited at 1064 nm. Chemical Physics Letters 185, 53–55.

Beman, J.M., and Francis, C.A. (2006). Diversity of Ammonia-Oxidizing Archaea and Bacteria in the Sediments of a Hypernutrified Subtropical Estuary: Bahia del Tobari, Mexico. Applied and Environmental Microbiology 72, 7767–7777.

Bloor, W.R. (1943). The biochemistry of the fatty acids. Reinhold Publishing Corp., New York.

Bredemeier, R., Hulsch, R., Metzger, J.O., and Berthe-Corti, L. (2003). Submersed culture production of extracellular wax esters by the marine bacterium Fundibacter jadensis. Mar. Biotechnol. 5, 579–583.

Breiman, L., Friedman, J., Stone., J.S., Olshen, R.A. (1984), Classification and Regression Trees, Wadsworth Internatioanl Group, Bellmont, California, 358 p.

Breiman L. – statistics Berkeley. University of California, Department of Statistics http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Breiman, L. (1996). Bagging Predictors. Machine Learning, 123–140.

Breiman, L. (2001). Random Forests. Machine Learning, 5–32.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). Mesophilic crenarchaeota: proposal for a third archaeal phylum, the *Thaumarchaeota*. Nature Reviews Microbiology 6, 245–252.

Bühlmann, P., and Yu, B. (2002). Analyzing bagging. The Annals of Statistics 30, 927–961.
Bugay, D.E., and Findlay, W.P. (1999). Pharmaceutical Excipients: Characterization by IR, Raman, and NMR Spectroscopy (Informa Healthcare).

Bureau, A., Dupuis, J., Falls, K., Lunetta, K.L., Hayward, B., Keith, T.P., and Van Eerdewegh, P. (2005). Identifying SNPs predictive of phenotype using random forests. Genetic Epidemiology 28, 171–182.

Buschman, H.P., Deinum, G., Motz, J.T., Fitzmaurice, M., Kramer, J.R., van der Laarse, A., Bruschke, A.V., and Feld, M.S. (2001). Raman microspectroscopy of human coronary atherosclerosis: Biochemical assessment of cellular and extracellular morphologic structures in situ. Cardiovascular Pathology 10, 69–82.

Caruana, R., Karampatziakis, N., and Yessenalina, A. (2008). An empirical evaluation of supervised learning in high dimensions. (ACM Press), pp. 96–103.

Caspers, P.J., Lucassen, G.W., and Puppels, G.J. (2003). Combined In Vivo Confocal Raman Spectroscopy and Confocal Microscopy of Human Skin. Biophysical Journal 85, 572–580.

Chase, D.B. (1986). Fourier transform Raman spectroscopy. Journal of the American Chemical Society 108, 7485–7488.

Chaturvedi, D., Mishra, S., Tandon, P., Dayal Gupta, V. and Siesler, H.W. (2009). Vibrational dynamics of poly(β-hydroxybutyrate)-α form. Polymer engineering and science. Volume 49, issue 5, pages 850-861.

Chen, X.-P., Zhu, Y.-G., Xia, Y., Shen, J.-P., and He, J.-Z. (2008). Ammonia-oxidizing archaea: important players in paddy rhizosphere soil? Environmental Microbiology 10, 1978–1987.

Chen, X.-W., and Liu, M. (2005). Prediction of protein-protein interactions using random decision forest framework. Bioinformatics 21, 4394–4400.

Ciobotă, V., Burkhardt, E.-M., Schumacher, W., Rösch, P., Küsel, K., and Popp, J. (2010). The influence of intracellular storage material on bacterial identification by means of Raman spectroscopy. Analytical and Bioanalytical Chemistry 397, 2929–2937.

Coolen, M.J.L., Abbas, B., van Bleijswijk, J., Hopmans, E.C., Kuypers, M.M.M., Wakeham, S.G., and Sinninghe Damsté, J.S. (2007). Putative ammonia-oxidizing Crenarchaeota in suboxic waters of the Black Sea: a basin-wide ecological study using 16S ribosomal and functional genes and membrane lipids. Environ. Microbiol. 9, 1001–1016.

Creely, C.M., Singh, G.P., and Petrov, D. (2005). Dual wavelength optical tweezers for confocal Raman spectroscopy. Optics Communications 245, 465–470.

Daims, H., Brühl, A., Amann, R., Schleifer, K.H., and Wagner, M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. Syst. Appl. Microbiol. 22, 434–444.

Daims, H., Stoecker, K, Wagner M. (2005). Fluorescence in situ hybridization for the detection of prokaryotes. In Advanced Methods in Molecular Microbial Ecology, pp. 213-239. (Osborn AM, Smith CJ, ed.). Bios-Garland, Abingdon, UK.

Damgaard, C. Gini Coefficient. From MathWorld-A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/GiniCoefficient.html

Damsté, J.S., Rijpstra, W.I.C., Hopmans, E.C., Weijers, J.W.H., Foesel, B.U., Overmann, J., and Dedysh, S.N. (2011). 13,16-Dimethyl Octacosanedioic Acid (iso-Diabolic Acid), a Common Membrane-Spanning Lipid of Acidobacteria Subdivisions 1 and 3. Applied and Environmental Microbiology 77, 4147–4154.

Damsté, J.S., Schouten, S., Hopmans, E.C., van Duin, A.C., Geenevasen, J.A. (2002). Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. The Journal of Lipid Research 43, 1641–1651.

Darveau, R.P., Charnetzky, W.T., and Hurlbert, R.E. (1980). Outer membrane protein composition of Yersinia pestis at different growth stages and incubation temperatures. J. Bacteriol. 143, 942–949.

Dawes, E.A. (1992) Storage polymers in prokaryotes. In: Prokaryotic Structure and Function: A New Perspective (Mohan, S., Dowand, C. and Cole, J.A., Eds.), pp. 81-122. Cambridge University Press, Cambridge.

Díaz-Uriarte, R., and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7, 3.

Dietterich, T. (2000) An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization, Machine Learning, vol. 40, no. 2, pp. 139–157.

Dutilh, B.E., Jurgelenaite, R., Szklarczyk, R., van Hijum, S.A.F.T., Harhangi, H.R., Schmid, M., de Wild, B., Francoijs, K.-J., Stunnenberg, H.G., Strous, M. (2011). FACIL: Fast and Accurate Genetic Code Inference and Logo. Bioinformatics 27, 1929–1933.

Ellen, A.F., Albers, S.-V., Huibers, W., Pitcher, A., Hobel, C.F.V., Schwarz, H., Folea, M., Schouten, S., Boekema, E.J., Poolman, B. (2008). Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components. Extremophiles 13, 67–79.

Enders, P.B., Multivariate Analysis - Hierarchical Cluster Analysis, http://www.philender.com/courses/multivariate/notes2/cluster.html

Engelking, L.R. (2010). Textbook of Veterinary Physiological Chemistry (Academic Press).

Ensign, S.A., Hyman, M.R., and Arp, D.J. (1993). In vitro activation of ammonia monooxygenase from *Nitrosomonas europaea* by copper. J. Bacteriol. 175, 1971–1980.

Erguder, T.H., Boon, N., Wittebolle, L., Marzorati, M., and Verstraete, W. (2009). Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. FEMS Microbiology Reviews 33, 855–869.

Fechner, P. (2005). Raman-Spektroskopie und atmosphärische Rasterelektronenmikroskopie - Charakterisierung pharmazeutischer Hilfsstoffe. Martin-Luther-Universität Halle-Wittenberg.

graduate thesis.

Fendrihan, S., Musso, M., and Stan-Lotter, H. (2009). Raman spectroscopy as a potential method for the detection of extremely halophilic archaea embedded in halite in terrestrial and possibly extraterrestrial samples. Journal of Raman Spectroscopy 40, 1996–2003.

Ferraro, J.R., Nakamoto, K., and Brown, C.W. (2002). Introductory Raman spectroscopy (Amsterdam: Academic Press).

Fraley, C. (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. The Computer Journal 41, 578–588.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005). Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. Proc. Natl. Acad. Sci. U.S.A. 102, 14683–14688.

Furukawa, T., Sato, H., Murakami, R., Zhang, J., Noda, I., Ochiai, S., and Ozaki, Y. (2006). Raman microspectroscopy study of structure, dispersibility, and crystallinity of poly(hydroxybutyrate)/poly(l-lactic acid) blends. Polymer 47, 3132–3140.

De Gelder, J., Willemse-Erix, D., Scholtes, M.J., Sanchez, J.I., Maquelin, K., Vandenabeele, P., De Boever, P., Puppels, G.J., Moens, L., and De Vos, P. (2008). Monitoring Poly(3-hydroxybutyrate) Production in Cupriavidus necator DSM 428 (H16) with Raman Spectroscopy. Analytical Chemistry 80, 2155–2160.

De Gelder, J., Scheldeman, P., Leus, K., Heyndrickx, M., Vandenabeele, P., Moens, L., and Vos, P. (2007). Raman spectroscopic study of bacterial endospores. Analytical and Bioanalytical Chemistry 389, 2143–2151.

Gniadecka, M., Philipsen, P.A., Sigurdsson, S., Wessel, S., Nielsen, O.F., Christensen, D.H., Hercogova, J., Rossen, K., Thomsen, H.K., Gniadecki, R. (2004). Melanoma Diagnosis by Raman Spectroscopy and Neural Networks: Structure Alterations in Proteins and Lipids in Intact Cancer Tissue. Journal of Investigative Dermatology 122, 443–449.

Golub, T.R. (1999). Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science 286, 531–537.

De Gussem, K. (2007). Optimisation of Raman spectroscopy for the analysis of Basidomycota: spores, latex and mycelium. Universiteit Gent faculteit Wetenschappen. diploma thesis.

Hansel, C.M., Fendorf, S., Jardine, P.M., and Francis, C.A. (2008). Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Appl. Environ. Microbiol. 74, 1620–1633.

Hatzenpichler, R., Lebedeva, E.V., Spieck, E., Stoecker, K., Richter, A., Daims, H., and Wagner, M. (2008). A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. Proc. Natl. Acad. Sci. U.S.A. 105, 2134–2139.

Hayashi, H., Noguchi, T., and Tasumi, M. (1989). Studies on the interrelationship among the intensity of a Raman marker band of carotenoids, polyene chain structure, and efficiency of the energy transfer from carotenoids to bacteriochlorophyll in photosynthetic bacteria. Photochemistry and Photobiology 49, 337–343.

Hazel, J.R., and Williams, E.E. (1990). The role of alterations in membrane lipid composition in enabling physiological adaptation of organisms to their physical environment. Prog. Lipid Res. 29, 167–227.

Herfort, L., Schouten, S., Abbas, B., Veldhuis, M.J.W., Coolen, M.J.L., Wuchter, C., Boon, J.P., Herndl, G.J., and Sinninghe Damsté, J.S. (2007). Variations in spatial and temporal distribution of Archaea in the North Sea in relation to environmental variables. FEMS Microbiol. Ecol. 62, 242–257.

Herrmann, M., Saunders, A.M., and Schramm, A. (2008). Archaea Dominate the Ammonia-Oxidizing Community in the Rhizosphere of the Freshwater Macrophyte Littorella uniflora. Applied and Environmental Microbiology 74, 3279–3283.

Hervada-Sala, C., and Jarauta-Bragulat, E. (2004). A program to perform Ward's clustering method on several regionalized variables. Computers & Geosciences 30, 881–886.

Hezayen, F.F., Steinbüchel, A., and Rehm, B.H.A. (2002). Biochemical and enzymological properties of the polyhydroxybutyrate synthase from the extremely halophilic archaeon strain 56. Arch. Biochem. Biophys. 403, 284–291.

Hoefs, M., Schouten, S., De Leeuw, J.W., King, L.L., Wakeham, S.G., and Damste, J. (1997). Ether lipids of planktonic archaea in the marine water column. Appl. Environ. Microbiol. 63, 3090–3095.

Horning, N. (American museum of natural history's center, center for biodiversity and conservation), http://www.whrc.org/education/indonesia/pdf/DecisionTrees_RandomForest_v2.pdf

Horowitz, D.M. and Sanders, J.K.M. (1995) Biomimetic, amorphous granules of polyhydroxyalkanoates: composition, mobility, and stabilization in vitro by proteins. Canadian journal of Microbiology, 41(13), 115-123.

Hoshino, T., Safak Yilmaz, L., Noguera, D.R., Daims, H. and Wagner, M. (2008), Quantification of target molecules needed to detect microorganisms by fluorescence in situ hybridization (FISH) and catalyzed reporter deposition-FISH. Appl. Environ. Microbiol. 74(16):5068.

Hu, Y., Shen, A., Jiang, T., Ai, Y., and Hu, J. (2008). Classification of normal and malignant human gastric mucosa tissue with confocal Raman microspectroscopy and wavelet analysis. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 69, 378–382.

Huang, W.E., Ward, A.D., and Whiteley, A.S. (2009). Raman tweezers sorting of single microbial cells. Environmental Microbiology Reports 1, 44–49.

Huang, X., Pan, W., Grindle, S., Han, X., Chen, Y., Park, S.J., Miller, L.W., and Hall, J. (2005). A comparative study of discriminating human heart failure etiology using gene expression profiles. BMC Bioinformatics 6, 205.

Huang, Z., McWilliams, A., Lui, H., McLean, D.I., Lam, S., and Zeng, H. (2003). Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. International Journal of Cancer 107, 1047–1052.

Ivleva, N.P., Niessner, R., and Panne, U. (2004). Characterization and discrimination of pollen by Raman microscopy. Analytical and Bioanalytical Chemistry 381, 261–267.

Izenman, A.J. (2008). Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning (Springer).

Izumi, C.M.S. and Temperini, M.L.A. (2010). FT-Raman investigation of biodegradable polymers: Poly(3-hydroxybutyrate) and poly (3-hydroxybutyrate-co-3-hydroxyvalerate). Vibrational spectroscopy 54, 127-132.

Jarute, G., Kainz, A., Schroll, G., Baena, J.R. and Lendl, B. (2004). On-line determination of the intracellular poly(β-hydroxybutyric acid) content in transformed *Escherichia coli* and glucose during PHB production using stopped-flow attenuated total reflection FT-IR spectrometry. Anal. Chem. Vol 76, 6353-6358.

Jendrossek, D., and Handrick R. (2002). Microbial degradation of polyhydroxyalkanoates. Annu. Rev. Microbial. 56:403-32

Jiang, D., Tang, C., and Zhang, A. (2004). Cluster analysis for gene expression data: a survey. IEEE Transactions on Knowledge and Data Engineering 16, 1370–1386.

Joe, H. and Ward, Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association. Volume 58, No. 301, page 236-244, Published by: American Statistical Association, Stable URL: http://www.jstor.org/stable/2282967

Jung, S., Zeikus, J.G., and Hollingsworth, R.I. (1994). A new family of very long chain alpha,omega-dicarboxylic acids is a major structural fatty acyl component of the membrane lipids of *Thermoanaerobacter ethanolicus* 39E. J. Lipid Res. 35, 1057–1065.

Jurgens, G., Glöckner, F., Amann, R.,Saano, A., Montonen, L., Likolammi, M., and Münster, U. (2000). Identification of novel Archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent in situ hybridization(1). FEMS Microbiol. Ecol. 34, 45–56.

Karydis, M. (2009). Eutrophication assessment of coastal waters based on indicators: a literature review. Global NEST Journal Vol 11, pp 373–390.

Kim, B.S., Lee, S.C., Lee, S.Y., Chang, H.N., Chang, Y.K., and Woo, S.I. (1994). Production of poly(3-hydroxybutyric acid) by fed-batch culture ofAlcaligenes eutrophus with glucose concentration control. Biotechnology and Bioengineering 43, 892–898.

Klotz, M.G., and Norton, J.M. (1998). Multiple copies of ammonia monooxygenase (amo) operons have evolved under biased AT/GC mutational pressure in ammonia-oxidizing autotrophic bacteria. FEMS Microbiol. Lett. 168, 303–311.

Kneipp, K., Wang, Y., Kneipp, H., Perelman, L., Itzkan, I., Dasari, R., and Feld, M. (1997). Single Molecule Detection Using Surface-Enhanced Raman Scattering (SERS). Physical Review Letters 78, 1667–1670.

Koch, I.H., Gich, F., Dunfield, P.F., and Overmann, J. (2008). *Edaphobacter modestus* gen. nov., sp. nov., and *Edaphobacter aggregans* sp. nov., acidobacteria isolated from alpine and forest soils. International Journal of Systematic and Evolutionary Microbiology 58, 1114–1122.

König, K. (2000). Robert Feulgen Prize Lecture. Laser tweezers and multiphoton microscopes in life sciences. Histochem. Cell Biol. 114, 79–92.

Könneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A.

(2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437, 543–546.

Konstantinidis, K.T., Braff, J., Karl, D.M., and DeLong, E.F. (2009). Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. Appl. Environ. Microbiol. 75, 5345–5355.

Krekeler, D., Sigalevich, P., Teske, A., Cypionka, H., and Cohen, Y. (1997). A sulfate-reducing bacterium from the oxic layer of a microbial mat from Solar Lake (Sinai), Desulfovibrio oxyclinae sp. nov. Archives of Microbiology 167, 369–375.

Krishna, C.M., Sockalingum, G.D., Kurien, J., Rao, L., Venteo, L., Pluot, M., Manfait, M., and Kartha, V.B. (2004). Micro-Raman spectroscopy for optical pathology of oral squamous cell carcinoma. Appl Spectrosc 58, 1128–1135.

Kuypers, M.M., Blokker, P., Erbacher, J., Kinkel, H., Pancost, R.D., Schouten, S., and Sinninghe Damste, J.S. (2001). Massive expansion of marine archaea during a mid-Cretaceous oceanic anoxic event. Science 293, 92–95.

Labspec 5 user manual: http://sindhu.ece.iisc.ernet.in/nanofab/twiki/pub/Main/MicroRamanMicroPL/HR800_Softw-Hardw_Manual_v02.pdf

Lam, P., Jensen, M.M., Lavik, G., McGinnis, D.F., Muller, B., Schubert, C.J., Amann, R., Thamdrup, B., and Kuypers, M.M.M. (2007). From the Cover: Linking crenarchaeal and bacterial nitrification to anammox in the Black Sea. Proceedings of the National Academy of Sciences 104, 7104–7109.

Lee, S.Y. (2000). Bacterial polyhydroxyalkanoates. Biotechnology and Bioengineering 49, 1–14.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C., and Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442, 806–809.

Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

Lieber, C.A., and Mahadevan-Jansen, A. (2003). Automated method for subtraction of fluorescence from biological Raman spectra. Appl Spectrosc 57, 1363–1367.

Loy, A., Horn, M., and Wagner, M. (2003). probeBase: an online resource for rRNA-targeted oligonucleotide probes. Nucleic Acids Res. 31, 514–516.

Lüthy, S. (2009). Merkmalswichtigkeit im Random Forest (Masterarbeit). Eidgenössische Technische Hochschule Zürich, Department für Mathematik.

Lunetta, K.L., Hayward, L.B., Segal, J., and Van Eerdewegh, P. (2004). Screening large-scale association study data: exploiting interactions using random forests. BMC Genet. 5, 32.

Luzier, W.D. (1992). Materials derived from biomass/biodegradable materials. Proc. Natl. Acad. Sci. U.S.A. 89, 839–842.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In Proceedings National Institute of Science, India, Vol. 2, No. 1. (16 April 1936), pp. 49-55

Majed, N., and Gu, A.Z. (2010). Application of Raman Microscopy for Simultaneous and Quantitative Evaluation of Multiple Intracellular Polymers Dynamics Functionally Relevant to Enhanced Biological Phosphorus Removal Processes. Environmental Science & Technology 44, 8601–8608.

Mannie, M.D., McConnell, T.J., Xie, C., and Li, Y. (2005). Activation-dependent phases of T cells distinguished by use of optical tweezers and near infrared Raman spectroscopy. Journal of Immunological Methods 297, 53–60.

Maquelin, K., Choo-Smith, L.-P., van Vreeswijk, T., Endtz, H.P., Smith, B., Bennett, R., Bruining, H.A., and Puppels, G.J. (2000). Raman Spectroscopic Method for Identification of Clinically Relevant Microorganisms Growing on Solid Culture Medium. Analytical Chemistry 72, 12–19.

Maquelin, K., Hoogenboezem, T., Jachtenberg, J.-W., Dumke, R., Jacobs, E., Puppels, G.J., Hartwig, N.G., and Vink, C. (2009). Raman spectroscopic typing reveals the presence of carotenoids in Mycoplasma pneumoniae. Microbiology 155, 2068–2077.

Marshall, C.P., Leuko, S., Coyle, C.M., Walter, M.R., Burns, B.P., and Neilan, B.A. (2007). Carotenoid Analysis of Halophilic Archaea by Resonance Raman Spectroscopy. Astrobiology 7, 631–643.

Martens-Habbena, W., Berube, P.M., Urakawa, H., de la Torre, J.R., and Stahl, D.A. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. Nature 461, 976–979.

Matousek, P., Towrie, M., and Parker, A.W. (2002). Fluorescence background suppression in Raman spectroscopy using combined Kerr gated and shifted excitation Raman difference techniques. Journal of Raman Spectroscopy 33, 238–242.

Meyer, M.W., and Smith, E.A. (2011). Optimization of silver nanoparticles for surface enhanced Raman spectroscopy of structurally diverse analytes using visible and near-infrared excitation. The Analyst 136, 3542.

Mimmack, G.M., Mason, S.J., and Galpin, J.S. (2001). Choice of Distance Matrices in Cluster Analysis: Defining Regions. Journal of Climate 14, 2790–2797.

Min, Y.-K., Yamamoto, T., Kohda, E., Ito, T., and Hamaguchi, H. (2005). 1064 nm near-infrared multichannel Raman spectroscopy of fresh human lung tissues. Journal of Raman Spectroscopy 36, 73–76.

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., and DeLong, E.F. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. Environ. Microbiol. 9, 1162–1175.

Misra, A.K., Thakur, M.S., Srinivas, P., and Karanth, N.G. (2004). Screening of poly-β-hydroxybutyrate-producing microorganisms using Fourier transform infrared spectroscopy (Springer).

Merrick, J.M., Steger, R. and Dombroski, D. (1999) Hydrolysis of native poly(hydroxyl-butyrate) granules (PHB), crystalline PHB, and articial amorphous PHB granules by intracellular and extracellular depolymerases. Int. J, Biol. Macromol. 25: 129-134.

Moorthy, K., and Mohamad, M.S. (2011). Random forest for gene selection and microarray data classification. Bioinformation 7, 142–146.

Muller, F., Brissac, T., Le Bris, N., Felbeck, H., and Gros, O. (2010). First description of giant Archaea (*Thaumarchaeota*) associated with putative bacterial ectosymbionts in a sulfidic marine habitat. Environ. Microbiol. 12, 2371–2383.

Murakami, R., Sato, H., Dybal, J., Iwata, T. and Ozaki, Y. (2007). Formation and stability of β-structure in biodegradable ultra-high-molecular-weight poly(3-hydroxybutyrate) by infrared, Raman, and quantum chemical calculation studies. Polymer 48, 2672-2680.

Mussmann, M., Brito, I., Pitcher, A., Sinninghe Damste, J.S., Hatzenpichler, R., Richter, A., Nielsen, J.L., Nielsen, P.H., Muller, A., Daims, H. (2011). Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. Proceedings of the National Academy of Sciences 108, 16771–16776.

Nakagawa, T., Mori, K., Kato, C., Takahashi, R., and Tokuyama, T. (2007). Distribution of Cold-Adapted Ammonia-Oxidizing Microorganisms in the Deep-Ocean of the Northeastern Japan Sea. Microbes and Environments 22, 365–372.

Nichols, D.S., Olley, J., Garda, H., Brenner, R.R., and McMeekin, T.A. (2000). Effect of temperature and salinity stress on growth and lipid composition of Shewanella gelidimarina. Appl. Environ. Microbiol. 66, 2422–2429.

Orendorff, C.J., Gole, A., Sau, T.K., and Murphy, C.J. (2005). Surface-Enhanced Raman Spectroscopy of Self-Assembled Monolayers: Sandwich Architecture and Nanoparticle Shape Dependence. Analytical Chemistry 77, 3261–3266.

Park, H.-D., Wells, G.F., Bae, H., Criddle, C.S., and Francis, C.A. (2006). Occurrence of Ammonia-Oxidizing Archaea in Wastewater Treatment Plant Bioreactors. Applied and Environmental Microbiology 72, 5643–5647.

Pearson, A.; Huang, Z., Ingalls, A.E., Romanek, C.S., Wiegel, J., Freeman, K.H., Smittenberg, R.H. and Zhang, C.L. (2004) Nonmarine crenarchaeol in Nevada hot springs. Appl. Environ. Microbiol. 70(9):5229.

Pernthaler, A., Pernthaler, J., and Amann, R. (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. Appl. Environ. Microbiol. 68, 3094–3101.

Pester, M., Rattei, T., Flechl, S., Gröngröft, A., Richter, A., Overmann, J., Reinhold-Hurek, B., Loy, A., and Wagner, M. (2012). amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions. Environmental Microbiology 14, 525–539.

Petry, R., Schmitt, M., and Popp, J. (2003). Raman spectroscopy--a prospective tool in the life sciences. Chemphyschem 4, 14–30.

Pitcher, A., Hopmans, E.C., Mosier, A.C., Park, S.-J., Rhee, S.-K., Francis, C.A., Schouten, S., and Damsté, J.S.S. (2011a). Core and intact polar glycerol dibiphytanyl glycerol tetraether lipids of ammonia-oxidizing archaea enriched from marine and estuarine sediments. Appl. Environ. Microbiol. 77, 3468–3477.

Pitcher, A., Rychlik, N., Hopmans, E.C., Spieck, E., Rijpstra, W.I.C., Ossebaar, J., Schouten, S., Wagner, M., and Sinninghe Damsté, J.S. (2009). Crenarchaeol dominates the membrane lipids of Candidatus *Nitrososphaera gargensis*, a thermophilic Group I.1b Archaeon. The ISME Journal 4, 542–552.

Pitcher, A., Wuchter, C., Siedenberg, K., Schouten, S., and Sinninghe Damsté, J.S. (2011b). Crenarchaeol tracks winter blooms of ammonia-oxidizing Thaumarchaeota in the coastal North Sea. Limnology and Oceanography 56, 2308–2318.

Poeggel, G. (2005). Kurzlehrbuch Biologie, Stuttgart, Thieme Verlag.

Popp, J., and Kiefer, W. (2006). Raman Scattering, Fundamentals. In Encyclopedia of Analytical Chemistry, R.A. Meyers, ed. (Chichester, UK: John Wiley & Sons, Ltd).

Prasad, A.M., Iverson, L.R., and Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. Ecosystems 9, 181–199.

Pratscher, J., Dumont, M.G., and Conrad, R. (2011). Ammonia oxidation coupled to CO2 fixation by archaea and bacteria in an agricultural soil. Proceedings of the National Academy of Sciences 108, 4170–4175.

Punji, G., and Stewart, D.W. (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. Journal of Marketing Research 20, 134.

Purkhold, U., Pommerening-Röser, A., Juretschko, S., Schmid, M.C., Koops, H.P., and Wagner, M. (2000). Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and amoA sequence analysis: implications for molecular diversity surveys. Appl. Environ. Microbiol. 66, 5368–5382.

Qi, Y., Bar-Joseph, Z., and Klein-Seetharaman, J. (2006). Evaluation of different biological data and computational classification methods for use in protein interaction prediction. Proteins 63, 490–500.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Raman, C.V., and Krishnan, K.S. (1928). A New Type of Secondary Radiation. Nature 121, 501–502.

Ramoni, M.F. (2002). From the Cover: Cluster analysis of gene expression dynamics. Proceedings of the National Academy of Sciences 99, 9121–9126.

Ray, P.H., White, D.C., and Brock, T.D. (1971). Effect of growth temperature on the lipid composition of *Thermus aquaticus*. J. Bacteriol. 108, 227–235.

Reigstad, L.J., Richter, A., Daims, H., Urich, T., Schwark, L., and Schleper, C. (2008). Nitrification in terrestrial hot springs of Iceland and Kamchatka. FEMS Microbiol. Ecol. 64, 167–174.

Rodin, A.S., Litvinenko, A., Klos, K., Morrison, A.C., Woodage, T., Coresh, J., and Boerwinkle, E. (2009). Use of wrapper algorithms coupled with a random forests classifier for variable selection in large-scale genomic association studies. J. Comput. Biol. 16, 1705–1718.

Santoro, A.E., Francis, C.A., de Sieyes, N.R., and Boehm, A.B. (2008). Shifts in the relative abundance of ammonia-oxidizing bacteria and archaea across physicochemical gradients in a subterranean estuary. Environ. Microbiol. 10, 1068–1079.

Sauder, L.A., Engel, K., Stearns, J.C., Masella, A.P., Pawliszyn, R., and Neufeld, J.D. (2011). Aquarium Nitrification Revisited: *Thaumarchaeota* Are the Dominant Ammonia Oxidizers in Freshwater Aquarium Biofilters. PLoS ONE 6, e23281.

Schallreuter, K.U., Chavan, B., Rokos, H., Hibberts, N., Panske, A., and Wood, J.M. (2005). Decreased phenylalanine uptake and turnover in patients with vitiligo. Molecular Genetics and Metabolism 86, 27–33.

Schittkowski, T., and Brüggemann, D. (2002). Raman-Spektroskopie und ihr Einsatz zur ortsaufgelösten Bestimmung von Gaskonzentrationen in rußenden Flammen. Chemie Ingenieur Technik 74, 1012–1016.

Schmid, M., Twachtmann, U., Klein, M., Strous, M., Juretschko, S., Jetten, M., Metzger, J.W., Schleifer, K.-H., and Wagner, M. (2000). Molecular Evidence for Genus Level Diversity of Bacteria Capable of Catalyzing Anaerobic Ammonium Oxidation. Systematic and Applied Microbiology 23, 93–106.

Schouten, S., Hopmans, E.C., Baas, M., Boumann, H., Standfest, S., Konneke, M., Stahl, D.A., and Sinninghe Damste, J.S. (2008). Intact Membrane Lipids of "Candidatus *Nitrosopumilus maritimus*," a Cultivated Representative of the Cosmopolitan Mesophilic Group I Crenarchaeota. Applied and Environmental Microbiology 74, 2433–2440.

Schouten, S., Hopmans, E. C., Forster, A. van Breugel, Y., Kuypers, M. M. M. and Sinninghe Damste´, J. S. (2003). Extremely high sea-surface temperatures at low latitudes during the middle Cretaceous as revealed by archaeal membrane lipids. Geology 31:1069–1072.

Schouten, S., van der Meer, M.T.J., Hopmans, E.C., Rijpstra, W.I.C., Reysenbach, A.-L., Ward, D.M., and Sinninghe Damsté, J.S. (2007). Archaeal and Bacterial Glycerol Dialkyl Glycerol Tetraether Lipids in Hot Springs of Yellowstone National Park. Applied and Environmental Microbiology 73, 6181–6191.

Schouten, S., Wakeham, S.G., Hopmans, E.C. and Sinninghe Damsté, J.S. (2003). Biogeochemical evidence that thermophilic archaea mediate the anaerobic oxidation of methane. Appl. Environ. Microbiol. 69:1680-1686.

Schrader, B. (1995). Infrared and raman spectroscopy: methods and applications. Berichte der Bunsengesellschaft für physikalische Chemie, Volume 100, Issue 7, page 1268

Schwarz, D.F., Konig, I.R., and Ziegler, A. (2010). On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data. Bioinformatics 26, 1752–1758.

Shao, J., Zheng, J., Liu, J., and Carr, C.M. (2005). Fourier transform Raman and Fourier transform infrared spectroscopy studies of silk fibroin. Journal of Applied Polymer Science 96, 1999–2004.

Sharma, S. (1996) Applied multivariate techniques. John Wiley and sons, inc., New York, 493 Pages.

Shen, J., Zhang, L., Zhu, Y., Zhang, J., and He, J. (2008). Abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea communities of an alkaline sandy loam. Environmental Microbiology 10, 1601–1611.

Shi, T., Seligson, D., Belldegrun, A.S., Palotie, A., and Horvath, S. (2005). Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. Mod. Pathol. 18, 547–557.

Shively, J.M. (1974). Inclusion Bodies of Prokaryotes. Annual Review of Microbiology 28, 167–188.

da Silva Martins, M.A., Ribeiro, D.G., Pereira Dos Santos, E.A., Martin, A.A., Fontes, A., and da Silva Martinho, H. (2010). Shifted-excitation Raman difference spectroscopy for in vitro and in vivo biological samples analysis. Biomed Opt Express 1, 617–626.

Smith, R.W., Bianchi, T.S., and Li, X. (2012). A re-evaluation of the use of branched GDGTs as terrestrial biomarkers: Implications for the BIT Index. Geochimica Et Cosmochimica Acta 80, 14–29.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D.A., Wagner, M., and Schleper, C. (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. Trends in Microbiology 18, 331–340.

Stahl, D. A., and Amann, R. (1991). Development and application of nucleic acid probes. In Nucleic Acid Techniques in Bacterial Systematics, E. Stackebrandt and M. Goodfellow, eds. (John Wiley & Sons), pp. 205-248.

Statnikov, A., Wang, L., and Aliferis, C.F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics 9, 319.

Statsoft(a), electronic statistics textbook, http://www.statsoft.com/textbook/classification-and-regression-trees/

Statsoft(b), electronic statistics textbook, http://www.statsoft.com/textbook/cluster-analysis/

Statsoft(c), electronic statistics textbook, http://www.statsoft.com/textbook/multidimensional-scaling/

Steinberg D., Golovnya, M. and Cardell, S. (2004) Data mining with random forests. Salford systems. http://nymetro.chapter.informs.org/prac_cor_pubs/RandomForest_SteinbergD.pdf

Steinbüchel, A. (2001). Perspectives for biotechnological production and utilization of biopolymers: Metabolic engineering of polyhydroxyalkanoate biosynthesis pathways as a successful example. Macromolecular Bioscience 1, 1–24.

Steinbuchel, A., Doi Y, eds. (2001). Biopolymers, Polyesters II, Properties and Chemical Synthesis. Weinheim: Wiley-VCH.468 pp.

Steinbuchel, A., Valentin H.E. (1995). Diversity of bacterial polyhydroxyalkanoic acids. FEMS Microbiol. Lett. 128:219–28

Stone, N., Kendall, C., Shepherd, N., Crow, P., and Barr, H. (2002). Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. Journal of Raman Spectroscopy 33, 564–573.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8, 25.

Teh, S.K., Zheng, W., Ho, K.Y., Teh, M., Yeoh, K.G., and Huang, Z. (2008). Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue. British Journal of Cancer 98, 457–465.

Teh, S.K., Zheng, W., Ho, K.Y., Teh, M., Yeoh, K.G., and Huang, Z. (2009). Near-infrared Raman spectroscopy for optical diagnosis in the stomach: Identification of Helicobacter-pylori infection and intestinal metaplasia. International Journal of Cancer n/a-n/a.

Teske, A., Alm, E., Regan, J.M., Toze, S., Rittmann, B.E., and Stahl, D.A. (1994). Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria. J. Bacteriol. 176, 6623–6630.

Thuoc, D.V. (2009). Production of poly(3-hydroxybutyrate) and ectoines using a halophilic bacterium. Department of Biotechnology, Lund University.

Tornabene, T., and Langworthy, T. (1979). Diphytanyl and dibiphytanyl glycerol ether lipids of methanogenic archaebacteria. Science 203, 51–53.

de la Torre, J.R., Walker, C.B., Ingalls, A.E., Könneke, M., and Stahl, D.A. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. Environmental Microbiology 10, 810–818.

Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.-P., and Schleper, C. (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. Environmental Microbiology 7, 1985–1995.

Urakawa, H., Tajima, Y., Numata, Y., and Tsuneda, S. (2008). Low temperature decreases the phylogenetic diversity of ammonia-oxidizing archaea and bacteria in aquarium biofiltration systems. Appl. Environ. Microbiol. 74, 894–900.

Venter, J.C. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science 304, 66–74.

de Vet, W.W.J.M., Dinkla, I.J.T., Muyzer, G., Rietveld, L.C., and van Loosdrecht, M.C.M. (2009). Molecular characterization of microbial populations in groundwater sources and sand filters for drinking water production. Water Res. 43, 182–194.

Wältermann, M., and Steinbüchel, A. (2005). Neutral lipid bodies in prokaryotes: recent insights into structure, formation, and relationship to eukaryotic lipid depots. J. Bacteriol. 187, 3607–3619.

Ward, A.T. (1968). Raman spectroscopy of sulfur, sulfur-selenium, and sulfur-arsenic mixtures. J. Phys. Chem., 72 (12), pp 4133-4139.

Ward, M. M., Pajevic, S., Dreyfuss, J. and Malley, J. D. (2006). Short-term prediction of mortality in patients with systemic lupus erythematosus: Classification of outcomes using random forests. Arthritis and Rheumatism 55 (1), 74–80.

Weidler, G.W., Dornmayr-Pfaffenhuemer, M., Gerbl, F.W., Heinen, W., and Stan-Lotter, H. (2007). Communities of archaea and bacteria in a subsurface radioactive thermal spring in the Austrian Central Alps, and evidence of ammonia-oxidizing Crenarchaeota. Appl. Environ. Microbiol. 73, 259–270.

Willemse-Erix, D.F.M., Scholtes-Timmerman, M.J., Jachtenberg, J.-W., van Leeuwen, W.B., Horst-Kreft, D., Bakker Schut, T.C., Deurenberg, R.H., Puppels, G.J., van Belkum, A., Vos,

M.C. (2009). Optical fingerprinting in bacterial epidemiology: Raman spectroscopy as a real-time typing method. J. Clin. Microbiol. 47, 652–659.

WINOGRADSKY, S. 1890 Recherches sur les organismes de la Nitrification. Ann. Inst. Pasteur (Paris), 4, 213-257. 760-771.

Wong, H.H., and Lee, S.Y. (1998). Poly-(3-hydroxybutyrate) production from whey by high-density cultivation of recombinant Escherichia coli. Applied Microbiology and Biotechnology 50, 30–33.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S. (2007). Top 10 algorithms in data mining. Knowledge and Information Systems 14, 1–37.

Wu, X., Wu, Z., and Li, K. (2008). Identification of differential gene expression for microarray data using recursive random forest. Chin. Med. J. 121, 2492–2496.

Wuchter, C. (2004). Temperature-dependent variation in the distribution of tetraether membrane lipids of marine Crenarchaeota: Implications for TEX 86 paleothermometry. Paleoceanography 19.

Xia, W., Zhang, C., Zeng, X., Feng, Y., Weng, J., Lin, X., Zhu, J., Xiong, Z., Xu, J., Cai, Z. (2011). Autotrophic growth of nitrifying community in an agricultural soil. The ISME Journal 5, 1226–1236.

Xie, C., Dinno, M.A., and Li, Y. (2002). Near-infrared Raman spectroscopy of single optically trapped biological cells. Optics Letters 27, 249.

York, G.M., Lupberger, J., Tian, J., Lawrence, A.G., Stubbe, J., and Sinskey, A.J. (2003). Ralstonia eutropha H16 encodes two and possibly three intracellular Poly[D-(-)-3-hydroxybutyrate] depolymerase genes. J. Bacteriol. 185, 3788–3794.

Zhang, L.-M., Offre, P.R., He, J.-Z., Verhamme, D.T., Nicol, G.W., and Prosser, J.I. (2010). Autotrophic ammonia oxidation by soil thaumarchaea. Proceedings of the National Academy of Sciences 107, 17240–17245.

Zhang, K., and Shasha, D. (1989). Simple fast algorithms for editing distance between trees and related problems. SIAM J. on Computing, volume 18, issue 6, pp 1245-1262.

10 Acknowledgements

This diploma thesis was performed from 01.02.2011 to 31.11.2011 at the Department of Microbial Ecology (DOME) of the Vienna Ecology Centre (University of Vienna) under the leadership of Professor Dr. Michael Wagner.

I kindly want to thank the following people:

- Markus Schmid, who helped me during the whole time here at DOME. He was not just a great teacher but also a good friend. Thank you for the great support and teaching me the techniques of Raman microspectroscopy.
- Michael Wagner, who is responsible for teaching me a lot of things, not only related to science. As the leader of the Raman group he inspired me at every meeting. It was such a pleasure to be a part of his team. You can really see that he loves his job.
- David Berry, who was there for me all the time when I needed him, even though he had so much other stuff to do. Thank you so much David, the whole Raman topic would have been only half as interesting without your constant work on the statistics.
- Alexander Galushko, who is not only a good and respectful teacher, but also one of the funniest people I know. Somehow he managed to stay in a good mood all the time, no matter what happened. Thank you very much Sascha, I really appreciated your never ending advice.
- Holger Daims, who is not just a really nice human being, but also a very good group leader. Thanks for the nice time in your Nitri-group.
- Jochen Reichert, who was my best laboratory buddy and personal friend during my diploma thesis. We knew each other before, but this adventure really built a good friendship. I can always count on your support Jochen, thank you so much for that.
- Andreas Anderluh, for giving the best high fives in the department and all the other fun during work.

Thanks also to all other Domies, which I did not mention above. I thank very much for your accompany me during my time here.

Most of all I want to thank my parents, who always encouraged me to do what I am interested in. Without your support there is no way I could have achieved my goals. Thanks for everything.

11 Curriculum vitae (CV)

Christoph Heinrich Böhm

Personal info

Date of birth	01.02.1984
Place of birth	Vienna
Nationality	Austria

Education

1990 – 1994	Elementary school (Sommerein)
1994 – 1998	Secondary school (Bruck/Leitha)
1998 – 2003	Commercial academy (Bruck/Leitha)
2004 – 2004	Assistant of the management board of Toyota Toyfl (Hennersdorf)
2004 – 2012	Studies at the University of Vienna
	Genetics and Microbiology
	Specialized in molecular microbiology