



MASTERARBEIT

Titel der Masterarbeit

„Spectral characterisation of different glycerol diphytanyl ethers as pure substances and part of the cell membrane of archaea using time-of-flight secondary ion mass spectrometry (ToF-SIMS) and matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-MS)“

Verfasser

Marcus Wieder, B.rer.nat MSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2014

Studienkennzahl lt. Studienblatt:

A 066 863

Studienrichtung lt. Studienblatt:

Masterstudium Biologische Chemie

Betreuerin / Betreuer:

Univ.-Prof. Mag. Dr. Michael Wagner

Abstract

Introduction

Glycerol diphytanyl diethers and glycerol diphytanyl glycerol tetraethers (GDGT) are major components of archaeal lipids and could be utilized as a biomarker to differentiate between archaea and bacteria given an instrument that is able to resolve mass spectra in the high mass range with high lateral resolution and mass accuracy. An instrument possessing such features is the ToF (Time-of-Flight) - SIMS (Secondary Ion Mass Spectroscopy). Spectra of pure substances of Glycerol diphytanyl diethers and GDGTs were analyzed with MALDI-MS/MS and ToF-SIMS to find significant fragments for these lipids followed by ToF-SIMS measurements of prokaryotes to test the utility of these fragments in a real life application. One specific GDGT - crenarchaeol - is unique for the phylum thaumarchaeota (chemolithoautotrophic ammonia-oxidizers) and the aim of this thesis was to find crenarchaeol ions and fragments with sufficient intensity to identify thaumarchaeota in a mixture of different prokaryotes.

Methods

Spectra were obtained from two glycerol diphytanyl diethers (1,2-Di-O-Phytanyl-sn-Glycerol and 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamine) and one glycerol diphytanyl glycerol tetraether (crenarchaeol) using ToF-SIMS and MALDI-MS. Spectra were analyzed for the molecular ion peak ($[M]^+$ and $[M]^-$), fragments and adducts. Two prokaryotes (the bacteria *E. coli* and the thaumarchaeota *N. viennensis*) were analyzed using ToF-SIMS. The spectra of the prokaryotes were investigated using known biomarkers for organic substances and subsequently with multivariate statistical tools (PCA and correlation coefficients). Furthermore in a bioinformatic subproject of this thesis an analysis for two missing enzymes in the biosynthesis pathway of crenarchaeol was performed using cluster of orthologous genes, phylogenetic and biochemical considerations.

Results and Summary

It was possible to identify the molecule ion peak for the purified glycerol diphytanyl diethers with MALDI-MS but it was not possible to find the molecule ion peak for crenarchaeol in the MALDI-MS spectra. We were able to identify the molecule ion and the Na^+ adduct peak of all analyzed molecules using ToF-SIMS but for crenarchaeol the intensity was very low. In the subsequent analysis of crenarchaeol in the membrane of entire prokaryotic cells this finding was not reproducible - only one fragment was found that might originate from crenarchaeol.

ToF-SIMS analysis was often not able to reconstruct the contour of clusters of prokaryotes as seen in the microscope using high to medium molecular masses. Furthermore it was not possible to differ between single cells of archaea and also hardly possible to distinguish single *E. coli* cells - the lateral resolution and secondary ion yield was too little for organic medium to high molecular mass molecules.

In addition it was not always possible to distinguish between biomass and non-biomass and the intensity of high-molecular mass peaks (above 600) was very low. The destruction of organic molecules on the surface and beneath by the erosion gun using Cs^+ or O_2^+ in combination with the analysis gun using Bi_3^+ was most likely too severe.

The statistical analysis (PCA) of mean-centered and normalized spectra showed that it was possible to distinguish between *E. coli* and *N. viennensis*.

For the two missing enzymes in the biosynthesis pathway of crenarchaeol it was possible to give a ranked list of in silico predicted candidate genes, which should be used as a guideline for further research.

Abstract

Einleitung

Glycerol-Diphytanyl-Diether und Glycerol-Diphytanyl-Glycerol-Tetraether (GDGT) sind Hauptbestandteile von archaealen Lipiden und könnten als Biomarker zur Einzelzellidentifizierung und -differenzierung von Archaea und Bakterien verwendet werden - eine Voraussetzung hierzu ist ein analytisches Instrument mit der benötigten räumlichen Auflösung sowie ausreichender Massenauflösung. Ein analytisches Instrument, welches dieses Anforderungsprofil erfüllt, ist das ToF (Time-of-Flight) - SIMS (Secondary Ion Mass Spectroscopy). Glycerol-Diphytanyl-Diether und Glycerol-Diphytanyl-Glycerol-Tetraether wurden als Reinsubstanzen auf ihr Fragmentierungsverhalten mit MALDI (Matrix-unterstützte Laser- Desorption/Ionisation) -MS/MS und ToF-SIMS untersucht gefolgt von ToF-SIMS Messungen von Prokaryonten. Von primärem Interesse war ein spezielles GDGT - Crenarchaeol - welches einzigartig für das Phylum Thaumarchaeota (chemolithoautotrophe Ammoniak oxidierende Archaea) ist. Das Ziel dieser Masterarbeit war die Identifizierung von Crenarchaeol-Ionen, -Addukten und -Fragmenten welche zur Identifizierung von Thaumarchaeota im Gemisch mit anderen Prokaryoten genutzt werden können.

Methoden

Analysiert wurden zwei Glycerol-Diphytanyl-Diether (1,2-Di-O-Phytanyl-sn-Glycerol und 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamin) und ein Glycerol-Diphytanyl-Glycerol-Tetraether (Crenarchaeol) unter Verwendung von ToF-SIMS und MALDI-MS. Spektren wurden mit Hilfe gängiger Molekülsignale $[M]^+$, $[M]^-$, Fragmenten und Addukten analysiert. Weiters wurden zwei Prokaryonten (Archaeon *N. viennensis* und Bakterium *E. coli*) mit ToF-SIMS analysiert. Die Spektren dieser Organismen wurden mit Hilfe von bekannten Biomarkern für Biomasse untersucht. Weiters wurden Methoden der multivariaten Statistik (PCA und Korrelationsanalysen) zur Auswertung angewendet. Auch wurde im Bioinformatik-Teil dieser Arbeit nach zwei noch nicht charakterisierten Enzymen im Biosynthesepfad von Crenarchaeol mit Hilfe von orthologen Genen, phylogenetischen und biochemischen Überlegungen gesucht.

Resultate und Zusammenfassung

MALDI-MS zeigte positive Molekülsignale für beide Glycerol-Diphytanyl-Diether, jedoch nicht für Crenarchaeol. ToF-SIMS konnte bei allen drei Reinsubstanzen positive Molekülsignale zeigen. Die folgende Analyse der Prokaryonten mit ToF-SIMS konnte Crenarchaeol in Archaea nicht nachweisen - jedoch wurde ein Fragment gefunden welches von Crenarchaeol stammen könnte. Die ToF-SIMS Analyse war häufig nicht in der Lage die Topologie der Zellen mit mittleren bis hohen Massen abzubilden. Weiters war es oft nicht möglich zwischen Bereichen mit Biomasse und Bereichen ohne Biomasse mit Hilfe der Spektren zu unterscheiden und die Intensität für Moleküle über 600 Dalton war sehr gering. Die Zerstörung der Oberfläche durch den primären Ionenstrahl war zu vollständig und der Ioneneinschlag fragmentierte die organischen Moleküle an und unter der Oberfläche. Die statistische Analyse ergab, dass es bei gemittelten und normierten Spektren möglich ist mittels PCA zwischen *E. coli* und *N. viennensis* zu unterscheiden. Die Loadings, welche hierfür verantwortlich waren, sind jedoch wahrscheinlicher ein Resultat unterschiedlicher Präparationsmethoden. Die Suche nach den zwei noch nicht charakterisierten Enzymen im Biosynthesepfad von Crenarchaeol ergab eine Liste von möglichen Kandidaten welche als Startpunkt für weitere biochemische Untersuchungen verwendet werden kann.

Contents

1	Introduction	1
1.1	Introduction to archaea	2
1.1.1	Archaea in general	2
1.1.2	The phylum Thaumarchaeota	3
1.2	Biochemistry of archaeal lipids	4
1.2.1	Introduction to lipids	4
1.2.2	Archaeal lipids	6
1.2.3	Crenarchaeol	8
2	Analytical strategy and methods	13
2.1	Starting point	13
2.2	ToF-SIMS	13
2.2.1	Principles	14
2.2.2	Measurement settings	22
2.3	MALDI-ToF	30
2.3.1	Principles	30
2.3.2	Measurement settings	32
3	Bioinformatic strategies, results and discussion	33
3.1	Introduction in the method	33
3.1.1	Cluster of orthologues genes - Concept	33
3.1.2	Cluster of orthologues genes - Data source	34
3.1.3	Workflow	35
3.2	Results and discussion	36
4	Characterization of Lipids	41
4.1	Spectral characterization of 1,2-Di-O-Phytanyl-sn-Glycerol	41
4.1.1	Sample preparation	41
4.1.2	MALDI-ToF measurements results and discussion	43
4.1.3	TOF-SIMS measurements results and discussion	45
4.2	Spectral characterization of 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamine	51
4.2.1	Sample preparation	51
4.2.2	MALDI-ToF measurements results and discussion	51
4.2.3	TOF-SIMS measurements results and discussion	54

4.3	Spectral characterization of Crenarchaeol	59
4.3.1	Sample preparation	59
4.3.2	MALDI-ToF measurements results and discussion	59
4.3.3	ToF-SIMS measurements results and discussion	60
5	Characterization of procaryotes	64
5.1	Sample preparation	64
5.2	TOF-SIMS measurements results and discussion	66
5.2.1	Data set 0103m*	69
5.2.2	Data set 0208m*	85
5.2.3	Data set 0416m*	93
5.2.4	Data set 0525m*	101
5.2.5	Chemometrics	105
6	Conclusion	123
	Literatur	132

List of Figures

1.1	Phylogeny of Archaea	3
1.2	Hydrolyzability of lipids	5
1.3	Core structure of alkylglycerols	6
1.4	The structure of crenarchaeol	8
1.5	Head groups of GDGTs	9
1.6	Head-to-head condensation of two diphytanyl glycerol diether	9
1.7	Simple biosynthesis pathway to archaeol	11
2.1	Molecular dynamics model of poly-vs. monoatomic sputtering	17
2.2	Model of sputtering process	19
2.3	Ion Formation Processes in SIMS	20
3.1	Workflow	35
3.2	Function vs abundance	38
4.1	Chemical structure of the glycerolipids	42
4.2	MALDI-ToF spectrum of PG	44
4.3	MALDI-CID spectrum of PG from precursor ion	46
4.4	MALDI-CID fragmentation scheme for PG	47
4.5	ToF-SIMS: proposed fragmentation of PG in positive mode	47

4.6	ToF-SIMS: spectrum for PG in positive mode	49
4.7	ToF-SIMS: proposed fragmentation of PG in negative mode	49
4.8	ToF-SIMS: spectrum for PG in negative mode	50
4.9	Full MALDI-ToF spectrum of PGE	52
4.10	MALDI-CID: spectrum of PGE from precursor ion	54
4.11	MALDI-ToF: fragmentation scheme for PGE	55
4.12	ToF-SIMS: spectrum for PGE in positive mode	57
4.13	ToF-SIMS: proposed fragmentation of PGE in negative mode	57
4.14	ToF-SIMS: spectrum for PGE in negative mode	58
4.15	ToF-SIMS: proposed fragmentation of crenarchaeol in positive mode	61
4.16	ToF-SIMS: spectrum for crenarchaeol in positive mode	61
4.17	ToF-SIMS: proposed fragmentation of crenarchaeol in negative mode	63
4.18	ToF-SIMS: spectrum for crenarchaeol in negative mode	63
5.1	p0103m002 - Microscope (20x and 40x) and ToF-SIMS camera image of E. coli	70
5.2	Cell Imaging with AFM and SEM	78
5.3	Visualization of p0208m102 - a selection of the most characteristic ions for biomass	86
5.4	Spectrum of p0208m102	87
5.5	Visualization of n0208m102 - a selection of the most characteristic ions for biomass	91
5.6	Spectrum of sample n0208m102	92
5.7	Visualization of p0416m101 - a selection of the most characteristic ions for biomass	94
5.8	Spectrum between of the measurement p0416m001	97
5.9	Spectrum of measurement p0416m001 between 1010 u and 1140 u	98
5.10	Visualization of p0416m001 - a selection of the most characteristic ions for biomass and microscope image	99
5.11	Depth Profile of p0416m001 for $C_{18}H_{38}O_2$	100
5.12	Proposed fragmentation pattern for $C_{44}H_{45}P_2O_8^+$	100
5.13	Spectrum of p0525m112	103
5.14	Spectrum of measurement p0525m112 between 700 u and 960 u	104
5.15	Raw Data of p0208m* - Scores of PC 1 vs scores of PC 2	113
5.16	Raw Data of p0208m* - Loadings of PC 1 and PC 2	114
5.17	Preprocessed Data of p0208m* - Shown are the scores of PC 1 and PC 2	114
5.18	Preprocessed Data of p0208m* - Loadings of PC 1 and PC 2	115
5.19	An example for biomass and non-biomass on the same sample	117
5.20	PCA scores for p0208m* - Biomass vs. Non-biomass	118
5.21	PCA scores for p0208m* - Biomass vs. Non-biomass - first 50 masses excluded from analysis	119
5.22	PC1 scores vs time - Biomass vs. Non-biomass - first 50 masses excluded from analysis	121

5.23	PC2 scores vs time - Biomass vs. Non-biomass - first 50 masses excluded from analysis	122
------	---	-----

List of Tables

1.1	Phylogenetic pattern for the head-to-head condensation	12
1.2	Phylogenetic pattern for the enzyme that catalyzes the cyclohexane ring formation	12
2.1	ToF-SIMS measurement parameters for lipids	24
2.2	ToF-SIMS measurement parameters for experiments 0103m*	25
2.3	ToF-SIMS measurement parameters for 0208m*	26
2.4	ToF-SIMS measurement parameters for experiments 0208m*	27
2.5	ToF-SIMS measurement parameters for 0416m*	28
2.6	ToF-SIMS measurement parameters for experiments 0525m*	29
3.1	Results for the head-to-head condensation based on phylogenetic patterns	39
3.2	Results for the internal cyclisation to form a cyclohexane	40
4.1	ToF-SIMS: lists of fragments for PG obtained in positive mode	48
4.2	ToF-SIMS: List of fragments for PG obtained in negative mode	48
4.3	ToF-SIMS: list of fragments for PGE obtained in positive mode	56
4.4	ToF-SIMS: list of fragments for PGE obtained in negative mode	56
4.5	ToF-SIMS: fragment list for crenarchaeol obtained in positive mode	62
4.6	ToF-SIMS: list of fragments for crenarchaeol obtained in negative mode	62
5.1	p0208m102 - Peak list used to visualize biomass	88
5.2	n0208m102 - Peak list used to visualize biomass	90
5.3	p0416m001 - Peak list obtained in high molecular mass range	96
5.4	p0525m112 - Peak list obtained in high molecular mass range	102
5.5	p0525m10* - Two selected masses of the correlation matrix	108
5.6	73 masses for the correlation matrix	109
5.7	Correlation matrix	110
5.8	PCA scores - without data preprocessing for data series p0208m*	111
5.9	PCA scores - with data preprocessing for data series p0208m*	112

Chapter 1

Introduction

Many questions in the life sciences require understanding and characterization of surface properties of cells and cellular components. A complete characterization of such targets include shape and morphology, but also information about the chemical abundance of organic and inorganic molecules and their three dimensional distribution in complex biological samples. Therefore it is desirable to be able to analyze intact organic molecules with the highest possible spatial resolution and with high mass resolution. There is a wide range of tools available to characterize surface properties (some limited to atoms, some extendable to molecules) which includes electron spectroscopy for chemical analysis (ESCA), static/dynamic time-of-flight secondary ion mass spectrometry (TOF-SIMS), scanning force microscopy, Auger electron spectroscopy (AES), contact angle methods, reflection Fourier transform infrared spectroscopy (FTIR) and near edge X-ray absorption fine structure (NEXAFS). Many of these techniques have also powerful imaging options and can detect and localize biological compounds with a tens of micrometer resolution [Belu *et al.*, 2003]. In this master thesis two mass spectrometric techniques based on the time-of-flight analyzer were used to characterize and image surfaces properties: matrix-assisted laser desorption/ionisation (MALDI-MS)¹ and secondary ion mass spectrometry (SIMS)². The aim of the thesis was to identify a glycerol diphytanyl glycerol tetraether

¹MALDI-MS/MS measurements were done in co-operation with Mag.pharm. Dr. Ernst Pittenauer, research group of (Bio)Polymer Analysis, Institute of Chemical Technologies and Analytics, Vienna University of Technology

²ToF-SIMS measurements were done in co-operation with DI Dr. Markus Holzweber and Sabrina Koch, MSc., Institute of Chemical Technologies and Analytics, Vienna University of Technology

compound (Crenarchaeol) present in archaeal cell membranes unique for the phylum Thaumarchaeota in the domain Archaea. Subsequently this lipid should be used as a biomarker for fast identification of single cells of Thaumarchaeota in environmental samples [Damsté *et al.*, 2002].

The second part of the master thesis dealt with the biosynthesis of crenarchaeol which has a well-characterized core pathway but the transition from the core pathway to the final product is not well understood [Eguchi *et al.*, 2000, Boucher *et al.*, 2004]. Starting from the known core pathway, two reactions that lead to the final biosynthesis of crenarchaeol were further investigated. Candidates for this reactions - one of them with a unique reaction mechanism - were analyzed and ranked based on phylogenetic considerations, phylogenetic patterns in orthologous cluster of proteins and biochemical plausibility.

1.1 Introduction to archaea

1.1.1 Archaea in general

Archaea are prokaryotic microorganisms and only recently (1990) recognized as a third, independent domain of life, separated from bacteria and eucarya. Traditionally, archaea were found in extreme environments (high temperature, high salt concentration, extreme pH) in which they dwelled as the dominant life form, but over the last decade it became more and more clear that they are ubiquitous and detectable in nearly all environments from marine to terrestrial ecosystems using culture-independent techniques [Schleper *et al.*, 2005].

Archaea were first distinguished from bacteria and eukarya by C.R. Woese based on the phylogenetic analysis of small-subunit rRNA sequences [Woese *et al.*, 1990]. There are also unique ecological and biochemical properties supporting the division along this lines. Among these differences most remarkably are the distinct features in polar membrane lipids (including their structures, biosynthesis, enzymes and their genes). Also their metabolic diversity and the ability of the majority of archaeal organisms to adapt to extreme environments is outstanding since many of them are thermophilic (heat-loving),

methanogenic (methane-producing) and halophilic (salt-loving) organisms.

The domain archaea is divided into five phyla since 2008, four of them shown in Figure 1.1 [Brochier-Armanet *et al.*, 2008, DasSarma *et al.*, 2009, McLain, 2005].

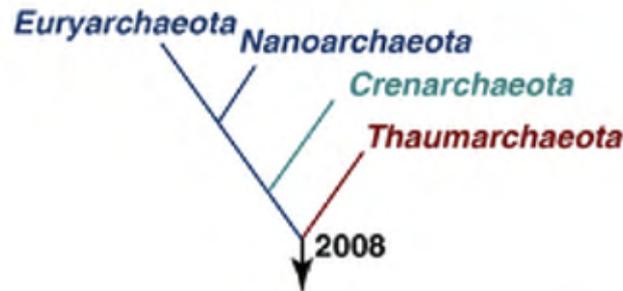


Figure 1.1: Phylogeny of Archaea. The tree is based on the analysis of 53 ribosomal proteins. The phylum Korarchaeota is not shown [Spang *et al.*, 2010].

1.1.2 The phylum Thaumarchaeota

In 2008 Thaumarchaeota were proposed as a new phylum within the archaea. This was done after it became clear that the division within the Crenarchaeota into (1) mesophilic Crenarchaeota and (2) hyperthermophilic Crenarchaeota was not supported by newly available sequence data and eventually that mesophilic Crenarchaeota have to be considered as a new phylum [Brochier-Armanet *et al.*, 2008].

All Thaumarchaeota described so far are ammonium oxidizing prokaryotes. Ammonium oxidation is a process which has been long studied in autotrophic Beta-proteobacteria and Gamma-proteobacteria and was also thought to be limited to bacteria. Thaumarchaeota are in contrast to their bacterial competitor for ammonium adapted to low ammonium concentration and their membrane-bound copper monooxygenases distinguishes itself through its wide substrate range [Pester *et al.*, 2011].

What is further of great importance for this thesis is the unique structure of membrane lipids in Thaumarchaeota with glycerol diphytanyl glycerol tetraethers (GDGT) as the most dominant class of lipids. There are many different types of GDGTs, a few of them are most promising to become unique surface biomarkers for this phylum. One of them

and unique for the phylum Thaumarchaeota is crenarchaeol, a GDGT with four cyclopentane rings and one cyclohexane moiety [Pitcher *et al.*, 2009].

1.2 Biochemistry of archaeal lipids

1.2.1 Introduction to lipids

Lipids in general are not easy to define because they are a large and heterogeneous group of substances. One common characteristic is their solubility in organic solvents such as methanol, chloroform, acetone and benzene because of their most important property: hydrophobicity.

Lipids play many different roles in organisms: fats and oils are the main form of stored energy in many eukaryots, phospholipids and sterols are important structural elements of biological membranes, lipids play crucial roles as enzyme cofactors, light-absorbing pigments (retinal), enzyme cofactors, hydrophobic anchors for proteins, chaperones, emulsifying agents, vitamins, hormones and intracellular messengers.

Lipids can be classified based on their biogenesis. Lipids are either biosynthesized from fatty acids and, as the two most basic subunits acetyl-CoA and malonyl-CoA or synthesized from isoprenoid units and acetyl-CoA and mevalonic acid as the basic subunits. Using these differences further subclasses can be defined based on the polarity of the molecules.

- Neutral lipids
 - fatty acids
 - waxes
 - carbonyl and alcohol hydrocarbons
 - mono-, di-, and triglycerids

 - Polar lipids
 - Phospholipids
-

– Glycolipids

Another characteristic which divides the known lipids is that they are either hydrolyzable or non-hydrolyzable as seen in Figure 1.2.

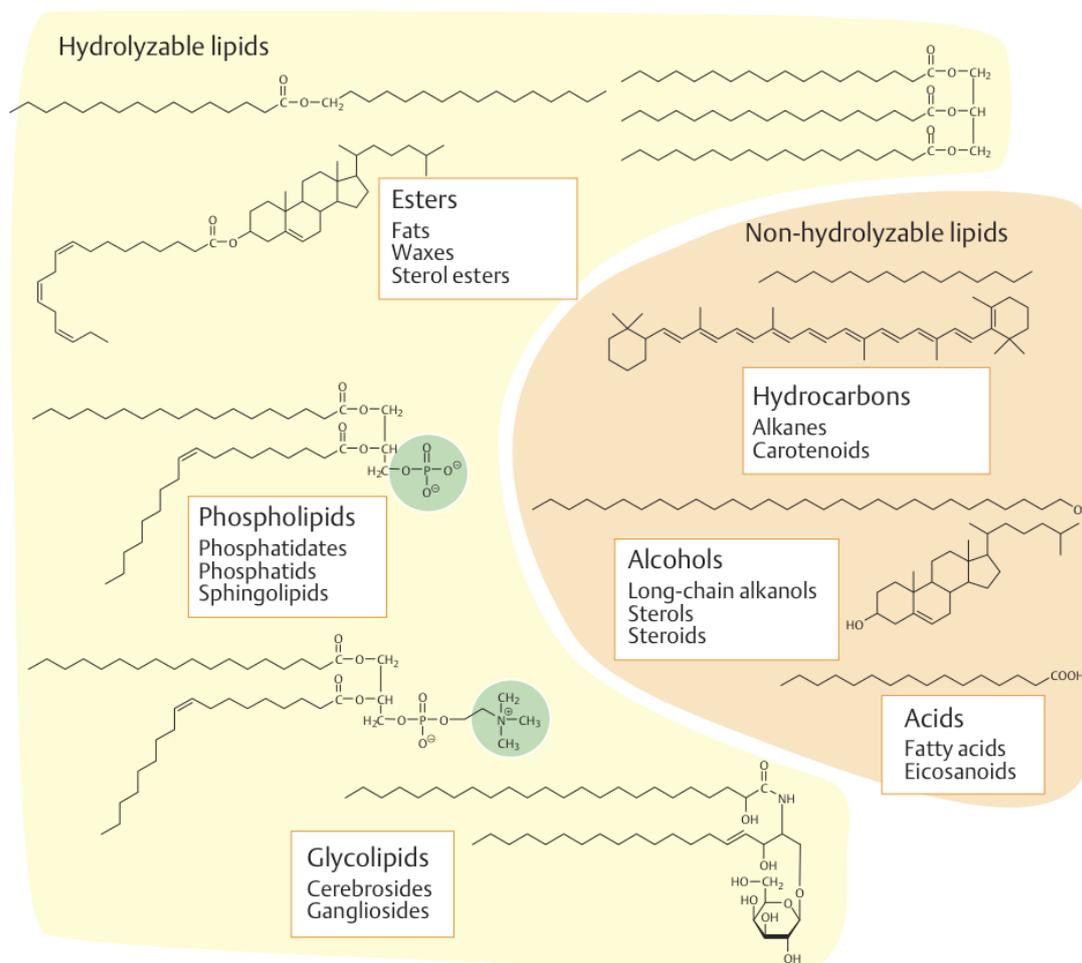


Figure 1.2: Classification of lipids based on the possibility to hydrolyze part of the molecule - usually the ester [Koolman & Rohm, 2004].

For this study lipids which contain a glycerol (dialkylglycerols and trialkylglycerols) - so called glycerolipids (not to be confused with glycolipids - lipids with a sugar moiety as polar head group) - are of further interest so this introduction will focus on them. These lipids can be divided in complex lipids which yield more than 3 fragments when hydrolyzed and simple lipids which give only two or less fragments. The basic structures

of diacylglycerols is shown in Figure 1.3. The prefix *sn* (for stereospecifically numbered) is used for lipids with a stereocenter (shown in Red on C2). Using the *sn* nomenclature for lipids the numbers 1,2 and 3 are assigned to the three carbon atoms of the glycerol starting with the stereocenter which is always *sn*-2. Then the molecule must be rotated so that the stereocenter is in the back and the oxygen bound to the stereocenter faces to the left. From top to bottom the numbers are assigned starting with 1 at the top. It should be noted that the glycerol is not planar. The alkylglycerids found in nature have a L-configuration in the Fischer projection [Koolman & Rohm, 2004, Berg *et al.*, 2008].

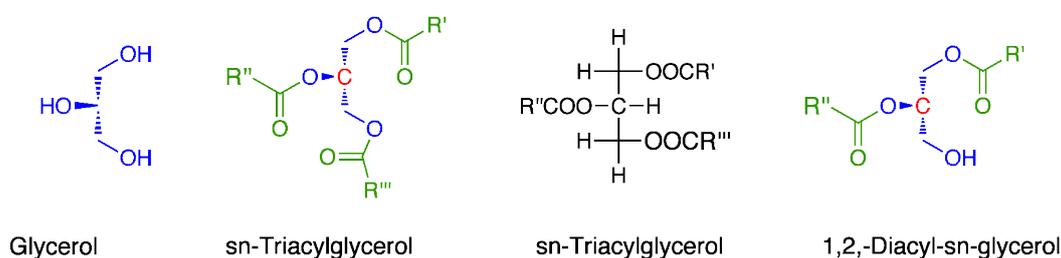


Figure 1.3: Core structure of alkylglycerols. The glycerol-subunit is shown in blue, the ester-linkage and alkyl moiety is shown in green and the stereocenter - if available - of the glycerol is shown in red.

1.2.2 Archaeal lipids

Archaeal lipids are distinguished from their eukaryotic and bacterial counterparts through four key differences: (1) the stereochemistry at the glycerol moiety, (2) the ether bonds, (3) the possibility to form monolayers and (4) the biosynthesis based on isoprenoid sub-units. What they have in common are the polar species which are linked to the glycerol. What is not shown as separated point is the fact that many archaeal tetraether lipids contain cyclopentyl moieties - something that was only recently discovered in anaerobic bacteria as well [Weijers *et al.*, 2006]. But what remains exclusively for Thaumarchaeota and has some importance for this thesis is the fact that - despite the cyclopentane group in the lipids - cyclohexane groups are a unique feature of this phylum.

The most commonly observed membrane lipid in archaea is a diphytanyl glycerol diether species (archaeol) containing two phytanyl chains each consisting of 20 carbon atoms.

To be more specific on the differences:

Chirality of the glycerol

In eukaryotic and bacterial lipids the glycerol (shown in Figure 1.3) group is always connected with the polar head group at the sn-3 position of the glycerol whereas the archaeal counterpart shows a very typical sn-glycerol-1-phosphate - therefore the bacterial/eukaryotic and archaeal glycerolipids are enantiomers. This is an exclusive feature for archaea in contrast to bacteria and eukaryotes [Koga & Morii, 2007, McIntyre *et al.*, 2008].

Ether bonds

The hydrocarbon chains are linked exclusively by ether bonds to the glycerol moiety in archaeal lipids. This fact is true for all archaea but it is not unique for this domain. Whereas most bacteria and eukaryotes normally link at least one of their hydrocarbon chains with an ester bond, in combination with ester also ether bonds are possible and not so unusual - for example at least one-fifth of the human phospholipid pool are ether lipids. Also macrocyclic GDGTs are possible in bacteria, in *Thermoanaerobacter ethanolicus* tetraether lipids were identified. For the biosynthesis of such lipids in bacteria normally alkyl alcohols are used rather than fatty acids [Matsumi *et al.*, 2011]. But the reverse is not true: ester bonds are not found in archaea and therefore a unique feature for bacteria and eukaryotes [Koga & Morii, 2007, McIntyre *et al.*, 2008].

Monolayer with bipolar lipids

Archaeal lipids can form monolayers in the cell membrane, spanning from one polar head-group to the other with the thickness of the cell membrane determined by the length of the two alkyl groups. But, despite the fact that this was also shown for thermophilic bacteria, it must also be considered that archaea incorporate more than one type of lipids in their membrane and not all of them are bipolar lipids [Weijers *et al.*, 2006].

Biosynthesis

Archaeal lipids are biosynthesized from isoprenoid subunits thus their branched alkyl chains in contrast to the mostly unbranched alkyl chains of bacteria and eukaryots in which the building blocks for the biosynthesis of the alkyl chains are fatty acids. The core pathway has some features which are conserved components of the mevalonate pathway and some new features - a very good and comprehensive summary is given in [Boucher *et al.*, 2004]. The mevalonate pathway is used in bacteria and eukaryots to synthesise biomolecules utilizing isoprenoids including steroids, protecting pigments, light-responsive co-factors and respiratory electron carriers [Matsumi *et al.*, 2011].

1.2.3 Crenarchaeol

Crenarchaeol is the major core lipid present in the cell membrane of Thaumarchaeota.

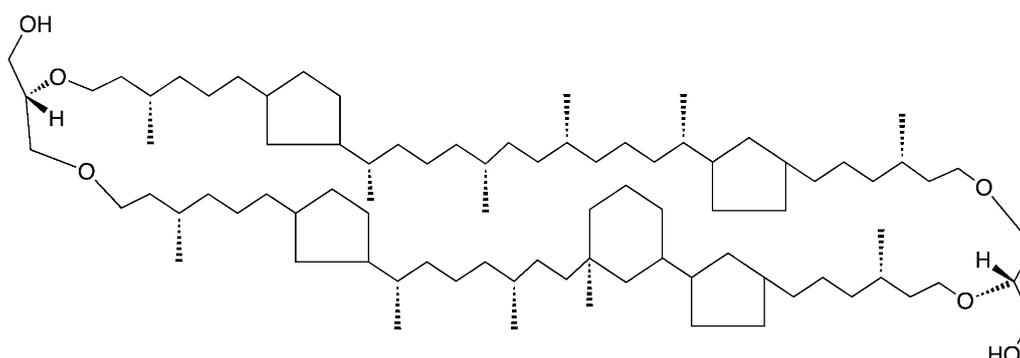


Figure 1.4: The structure of crenarchaeol. What should be noted are the four cyclopentyl and the one cyclohexyl moieties, the chirality at the glycerol and also the stereo-center at the cyclohexane.

Crenarchaeol is a macrocyclic **g**lycerol **d**iphytanyl **g**lycerol **t**etraether (GDGT) with four cyclopentyl moieties and one cyclohexyl moiety. As a bipolar membrane lipid it forms a monolayer. The cyclopentane and also the cyclohexane are formed through internal cyclization. The hydroxyl moiety has two possibilities to bind polar head groups: (1) it can form a phosphoester and connect to mono- or dihexoses, or (2) the lactol group of

the hexose can bind directly without a phosphate to the lipid and form an acetal with the hydroxyl group of the glycerol [Pitcher *et al.*, 2009, Damsté *et al.*, 2002].

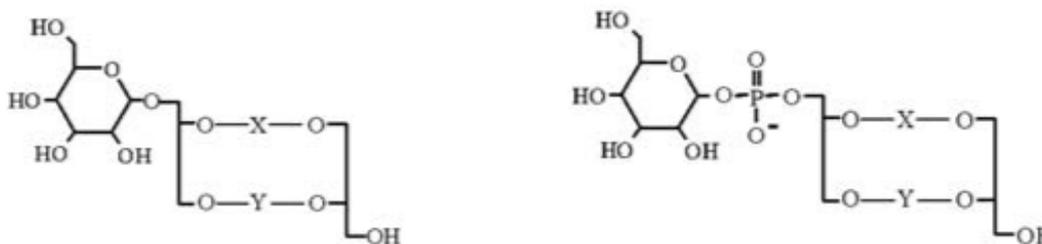


Figure 1.5: Head groups of glycerol diphytanyl glycerol tetraethers. The left structure shows a GDGT with a monohexose moiety as polar head group binding directly on the glycerol forming an acetal. The right structure shows a monohexose binding to the glycerol phosphate forming a phosphoester. X and Y stand for the alkyl chains associated with GDGTs [Pitcher *et al.*, 2009]

Biosynthesis of Crenarchaeol

For the majority of the biosynthesis pathway of crenarchaeol the key enzymes and structures are well known and also part of the known biosynthesis of archaeal lipids up to the formation of diphytanyl glycerol diether (archaeol) - as shown in Figure 1.7).

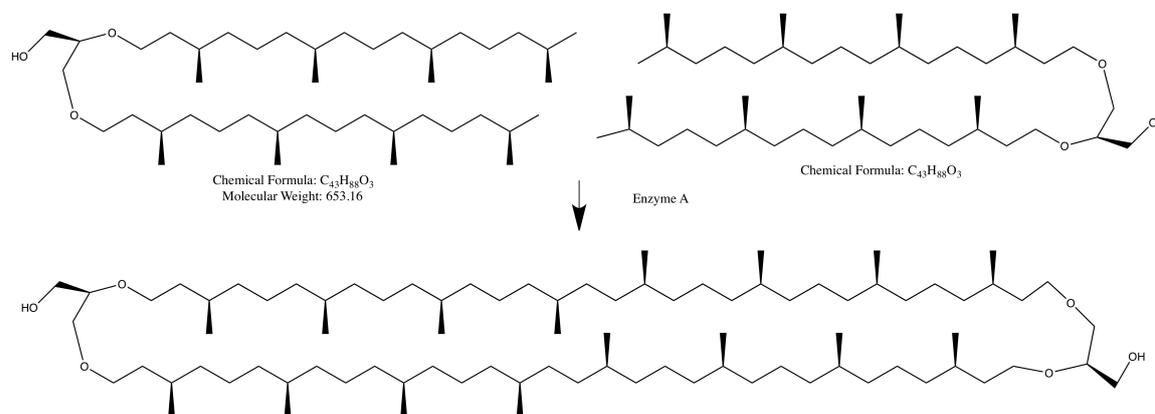


Figure 1.6: Head-to-head condensation of two diphytanyl glycerol diether to form a glycerol diphytanyl glycerol tetraether. The possible double bounds were omitted but it should be noted that it is not clear if the condensation takes place via the saturated phytanyl or unsaturated phytyl chains.

Two problems arise subsequent to the synthesis of the diphytanyl glycerol diether:

- (1) How are two units of glycerol diphytanyl (C20) diether connected to give rise to a single glycerol diphytanyl (C40) glycerol tetraether unit (shown in Figure 1.6) and
- (2) how is the cyclohexane ring introduced in a supposedly saturated phytanyl chain.

Crenarchaeol is built from two diphytanyl or - also possible - two diphytyl glycerol diether subunits through a head-to-head condensation of the phytanyl or phytyl tails and subsequently an internal ring formation to four cyclopentanes and one cyclohexane. Until today there are only speculations about the mechanism of the head-to-head condensation (unusual double-bond migration [Eguchi *et al.*, 2000]) and the cyclisation on the saturated chain (radical mechanism with metallocofactors [Galliker, 1990]) and the enzymes involved. Furthermore the structure of the actual substrate for the reaction that give rise to the cyclohexane is not known - two possibilities are discussed in the literature: a unsaturated diphytyl or the reduced, saturated form of diphytyl: diphytanyl. A phytane substrate would make a direct condensation as well as a subsequent internal condensation unlikely but there are some evidences pointing in this direction [Murakami *et al.*, 2007, Sato *et al.*, 2008, Nishimura & Eguchi, 2006, Galliker, 1990].

Phylogenetic pattern

In the bioinformatic part of this thesis, enzymes for the head-to-head condensation and for the cyclohexane ring formation were looked for and analyzed based on phylogenetic considerations. These considerations are quite obvious, but they are summarized at this point into two different tables for the two reactions.

Phylogenetic pattern for the head-to-head condensation is shown in Table 1.1 and the phylogenetic pattern for the ring formation is shown in Ta:RingFormation.

The head-to-head condensation is an observed process over all phyla of the archaea and it is not know if the enzyme/s which catalyze this reaction are also present in eukarya and bacteria (at least in bacteria this seems very likely since they are able to synthesis

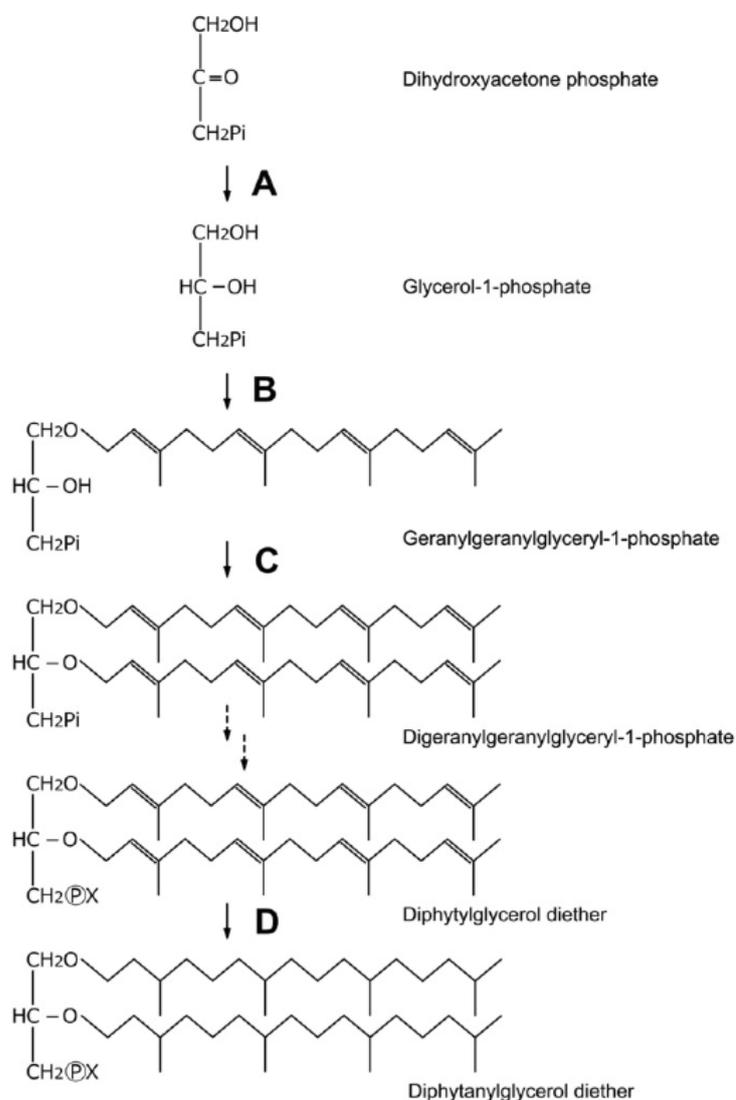


Figure 1.7: Simple biosynthesis pathway to archaeol. The first two steps show the reduction of dihydroxyacetone to glycerol. What follows is a very schematical description of the incorporation of the first and second phytanyl chain, the binding of the polar head-group (shown as PX - P for phosphate and X a polar rest) and the reduction from phytanyl to phytanyl. The following enzymes are responsible for the indicated reactions: A, glycerol-1-phosphate dehydrogenase; B, (S)-3-O-geranylgeranylgeranyl phosphate synthase; C, (S)-2,3-di-O-geranylgeranylgeranyl phosphate synthase; D, geranylgeranyl reductase (GGR) [Chong, 2010].

Archaea					Eukarya	Bacteria
Crenarch.	Thaumarch.	Nanoarch.	Korarch.	Euryarch.		
Y	Y	Y	Y	Y	?	?

Table 1.1: Phylogenetic pattern for the head-to-head condensation. What is known is that the enzyme/s for the head-to-head condensation has/have to be present in at least all archaea. Y means that they must be present there, ? stands for unknown if present or not.

long bipolar lipids).

Phylogenetic pattern Cyclohexane ring formation:

Archaea					Eukarya	Bacteria
Crenarch.	Thaumarch.	Nanoarch.	Korarch.	Euryarch.		
?	Y	?	?	?	?	?

Table 1.2: Phylogenetic pattern for the enzyme that catalyzes the cyclohexane ring formation. What is known is that this enzyme/s has/have to be present in all Thaumarchaeota

Chapter 2

Analytical strategy and methods

2.1 Starting point

It was known that it is possible to detect cholesterol and glycerolipids in freeze-dried thin sections of tissues, eukaryotic cells and bacteria with ToF-SIMS. Also achieved was imaging the distribution of these lipids with a resolution of about 100 nm [Johansson, 2006, McMahon *et al.*, 1996, Seedorf *et al.*, 1995]. The spectral characterization of archaeol as pure lipid - done by Christine Heim *et al* - extended the field of known ToF-SIMS spectra to archaeal lipids. In 2007 Volker Thiel *et al* published a paper in which the spectral characterization of cryosections of microbial mats is shown which contained archaea. The conclusion in this paper (and supported by the results of other studies albeit not to that extent) was that it is possible to analyze intact archaeal membrane lipids (including glycerol diphytanyl glycerol tetraethers) with ToF-SIMS in the molecular mass range of up to 1700 units with sufficient sensitivities. Worth to note are the sample preparation methods (cryofixation) in this papers [Heim *et al.*, 2009, Thiel *et al.*, 2007].

2.2 ToF-SIMS

Time-of-Flight (ToF) secondary ion mass spectrometry (SIMS) is a highly sensitive surface technique for characterizing atoms and molecules with high mass resolution (m/Δ

5000-10000) [Touboul *et al.*, 2005]. It yields positive or negative mass spectra from the outer atomic layers of the sample. Mass imaging with a spatial resolution of theoretically 4-8 nm and practically achieved 20 nm (for organic molecules about 150 nm are feasible) with high resolution raster (up to 1024 x 1024 pixel resolution) are recent achievements. Due to the capability to remove time-dependent defined amount of surface atom layers it enables the analyst to high-resolution depth profiling into the sub-micrometer range. In combination with computerized data acquisition the lateral as well as the depth resolution makes ToF-SIMS a powerful tool for three-dimensional chemical imaging for organic and inorganic molecules - therefore its wide field of applications, spanning from the analysis of semiconductors to thin sections of brain samples [Barshick *et al.*, 2000].

2.2.1 Principles

Mass spectrometry in its various forms always involves three steps: (1) evaporation and ionization of the sample, (2) separation of the formed ions based on the mass-to-charge (m/z) ratio and (3) detection of the separated ions.

The first step - the formation of gaseous ions from a solid or liquid sample - is the central process of the SIMS technique. In SIMS, secondary ions (sample ions) are produced via bombardment of the sample with primary ions setting off a collision cascade during which energy and momentum is transferred to the molecules in the sample surface and - exceeding a kinetic energy threshold - emitting molecules in the gas phase. In ToF-SIMS the separation of the masses is done with a time-of-flight mass analyzer and the detection is done with a micro channel plate connected to a scintillator.

There are two different approaches to SIMS analysis: (1) static SIMS and (2) dynamic SIMS. This differentiation depends only on the ion dose of the primary ion beam. The characteristic use for static SIMS is surface spectroscopy. This is achieved with low ion current density and little destruction of the surface area. This is granted by the static SIMS limit which is $< 10^{13}$ primary ions/cm² as a rule of thumb (10^{13} primary ions/cm² corresponds to the density of silicon in silicon wafer - amount of atoms in about 1 % of a

silicon monolayer) but greatly depending on the density of the sample. This limit is applied because when the primary ion hits the surface it produces secondary ions (to a very small percentage) but to a much greater extent get implanted in the sample surface. The energy transferred to the surface molecules damage the surface at this particular area and result in the production of artifacts and fragments which do not reflect the actual chemical properties of the surface. The diameter of the area which is affected by an ion impact is normally restricted to 5-10 nm and from this zone no further molecular information can be obtained.

Static SIMS is usually performed with a pulsed ion beam and a time of flight mass analyzer.

In contrast dynamic SIMS is characterized by higher primary ion dose density exceeding the static SIMS limit. This results in multiple sputtering events at the same site on the surface. For organic samples this results in high fragmentation. This mode is not used in this study for analysis of the surface but only for depth profiling, utilizing the oxygen and the cesium co-sputtering gun [Barshick *et al.*, 2000].

Primary Ion Sputtering

The bombardment of the primary ion beam results at the surface of the sample in the production of ions and various molecules of sample material and combinations of the primary ion species with sample material as well as electrons and photons. These emitted secondary particles carry negative, positive and neutral (about 95%) charges and have kinetic energies ranging from zero to several hundred eV.

The primary ions for the sputtering process are generated in ion guns. In modern bioanalytics polyatomic ion beams such as SF_5^+ , Au_n^+ , C_n^+ and Bi_n^+ have widely been used with an increase in the number of high-mass secondary ions generated during the sputtering process and an increase in secondary ion yield (measured secondary ions per primary ion) in comparison to monoatomic ion beams like Ar^+ , Cs^+ or Ga^+ . The model concept of

how a cluster primary ion interacts with the surface is as following: a cluster ion hits the surface of a sample and breaks apart. Each single ion of the prior polyatomic primary ion retains only a fraction of the initial kinetic energy. As a result the penetration depth of the ion as well as surface-localized damage decreases and therefore the chemical structure of the sample is better conserved than with monoatomic ion beams.

Primary ions are generated in an ion source and subsequently directed through ion optics on the surface of the probe as a focused ion beam [Benninghoven *et al.*, 1987, Mahoney, 2010, Kollmer, 2004].

Three methods are nowadays mostly used for ion formation: Liquid metal ion guns (LMIG), electron impact gas ion guns (and as recent development gas cluster ion guns) and surface ionization guns.

There are different atoms associated with each ionization method - metals like gallium, indium, gold and bismuth are produced in a liquid metal ion gun; oxygen, xenon, pentafluorosulfanyl and fullarene like C_{60}^+ are ionized in a electron impact gas ion gun and ionized cesium is produced through thermal surface ionization [Stapel & Benninghoven, 2001].

The two erosion ion sources used in this thesis and part of the ToF-SIMS V apparatus of ION-TOF (Münster) are used in form of a dual source column (DSC). Three guns were used:

- (1) Bi-LMIG emitter which produces Bi_n^+ primary ions with $n = 1 \dots 7$,
- (2) a Cs-thermal ionization ion source (this source is used for dual beam depth profiling in negative ion mode because the ionization probability of negative secondary ion species increases significantly through implantation of electropositive Cs atoms) and
- (3) an electron impact ion gun producing oxygen ions in form of O_2^+ (this source is used for dual beam depth profiling in positive ion mode because the ionization probability of positive secondary ion species increases significantly due to the implantation of O_2^+).

The implanting (sputtering) of atoms like Cs and O has as result a chemical yield en-

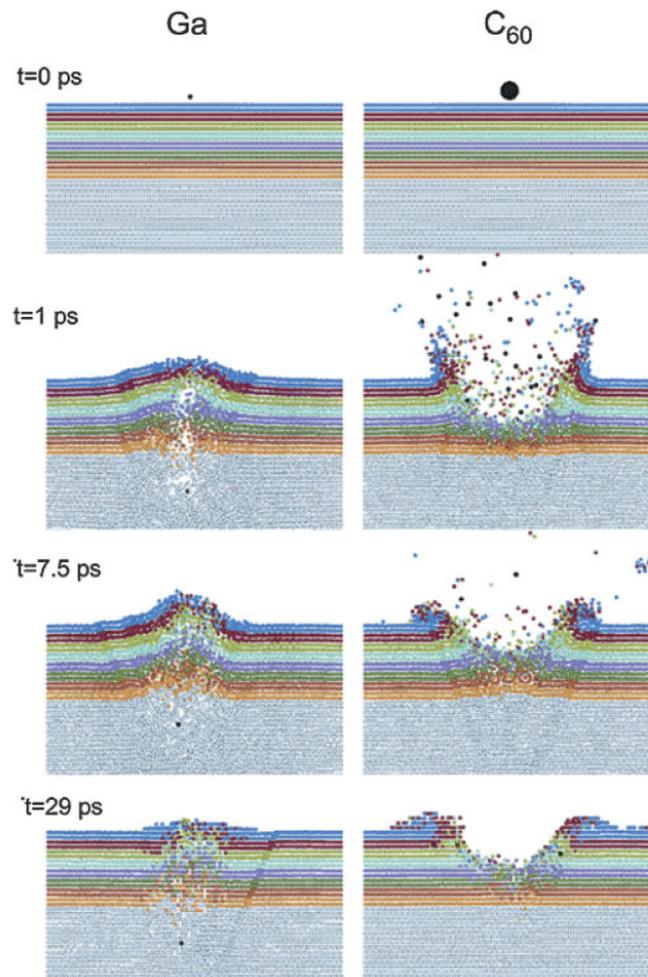


Figure 2.1: Molecular dynamics model of poly- vs. monoatomic sputtering. Shown is the evolution over time (0-29 ps) of a sputtering event leading to the emission of atoms due to 15keV gallium in comparison to a C₆₀ bombardment of a silver surface [Postawa *et al.*, 2004].

hancement due to their role as electron donator/acceptor.

Liquid metal ion guns consist of a metal covered tungsten tip which forms - when heated and a high voltage field is applied - a so called taylor-cone. The bismuth atoms are ionized through field emission at the apex of the taylor cone. The use of a Bi-LMIG as the primary ion source for acquiring spectra has many reasons: liquid metal ion guns in general have various very positive effects like long lifetimes ($> 500 \mu\text{Ah}$), very high brightness and can be focused - depending ion-optical set up - from several μm to about 50 nm [Nagy & Walker, 2007, Herzog & Viehböck, 1949, Benninghoven *et al.*, 1987].

Ions from the primary ion source collide with the sample and set atoms in motion during direct collision. This can be described as elastic collisions between point particles (Sigmund's linear cascade theory) as well as collision of atoms already in motion with other atoms in the sample (knock-on effects). The particles are governed by a linear Boltzmann transport equation. [Sigmund, 1969]. There are also non-linear cascade effects and, at least in metals, thermal spikes which should be considered when describing the energy transport and the ejection from primary to secondary ions. The energy of the primary ion beam is typical in the keV range - so compared to bond energies of molecules (2-7 eV) the primary ions are highly energetic and break down molecules essentially to atoms on the collision site, as the collision cascade moves further away from the point of impact the energy of the cascade and the fragmentation rate declines - ultimately resulting in directing part of the collision energy to the surface and the emission of molecular fragments. [Belu *et al.*, 2003, Montaudo & Lattimer, 2001, Barshick *et al.*, 2000, Jones *et al.*, 2006].

The sputtering yield determines the rate at which material is removed and therefore defines the depth profiling rate. The ion yield $S_{atoms/ion}$ is a function of the primary ion species, the sample and the kinetic energy and angle of the bombarding ion beam with respect to the sample. The sputtering process, especially for organic molecules, is not fully understood. A simple model is illustrated in Figure 2.2 and is described with the

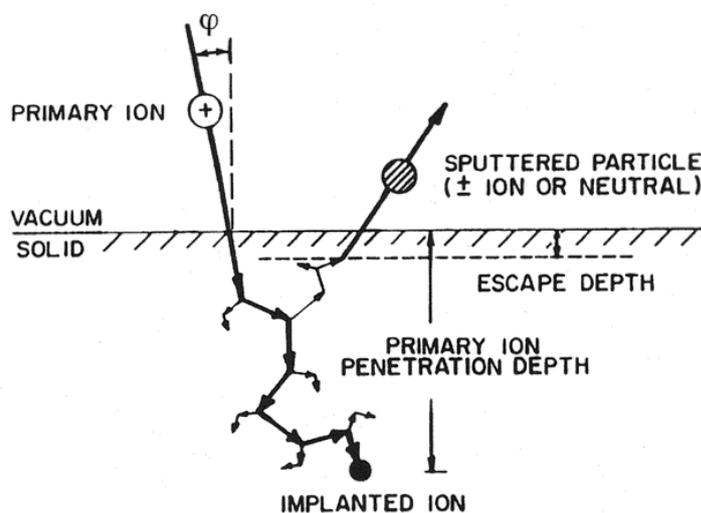


Figure 2.2: Model of sputtering process of the primary ions beam. A primary ion enters with an angle ψ and loses its energy through a series of collisions and gets implanted in the solid [Barshick *et al.*, 2000].

following parameters:

$$S_{\text{atoms/ion}} = ckE_0 \frac{\pi R^2 n_0}{\cos \varphi} \frac{M_1 M_2}{(M_1 + M_2)^2} \quad (2.1)$$

where S is the ion yield, c is a proportionality constant, k is a negative exponential function of the binding energy of the surface atoms, R is the collision radius calculated by using a rigid sphere model, n_0 is the number of lattice atoms per unit volume, φ is the angle between an imaginary line standing at an angle of 90 degree on the sample's surface and the direction of impact, M_1 and M_2 are the masses of the bombarding and target atoms and E_0 is the energy of the bombarding ion.

This model best describes the behavior of inorganic, monoatomic solid materials and must be considered with caution for organic or biological samples [Barshick *et al.*, 2000].

Ion Formation Processes

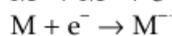
Although some sources describe the ion formation process utilized in SIMS techniques as soft ionization methods (like MALDI and ESI) and [Mikkelsen & Cortón, 2004, Francis

et al., 2008, Stephan, 2001] it should be kept in mind that it is ultimately better described as a hard ionization method because of the direct charge transmission between the primary ion beam and the formed ions. In SIMS, both positive and negative ions are formed. Ionization is strongly influenced by electron exchange processes between the emitted particle and the surface. Furthermore it is important to note and to take into consideration that the secondary ion yield can vary over several orders of magnitude across the periodic table and that it is very dependent on the chemical state of the surface. The dependence of secondary ion yield on the state of the surface also holds true for organic molecules but they are generally not so critical as for inorganic materials.

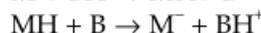
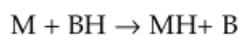
Ion Formation Processes in SIMS

(M = Polymer Molecule)

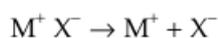
Gain or Loss of Electrons



Gain or Loss of Protons



Ionization of Salts



Ion Attachment

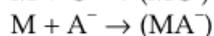
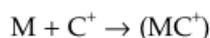


Figure 2.3: Ion Formation Processes in SIMS [Montaudo & Lattimer, 2001].

The mechanisms for the ion formation are [Vickerman & Briggs, 2001]:

- (1) ionization through gain or loss of an electron forming M^+ and M^- ,
- (2) protonation $[M+H]^+$,
- (3) de-protonation $[M-H]^-$ and

(4) ion attachment resulting in $[\text{MK}]^+$ or $[\text{MA}]^-$ adduct ions (K for Kation, A for Anion).

Separation of secondary ions

The formed secondary ions have to be separated. The majority of SIMS instruments use either a time-of-flight or a sector field mass analyzer. Both analysis methods separate the secondary ions based on their mass-to-charge ratio (m/z).

The sector field mass analyzer uses electrostatic and magnetic fields to obtain a separation of the different ions. The physics of the sector field mass analyzer can be ultimately described with the Lorentz force law:

$$F = q(E + v \times B) \quad (2.2)$$

where E is the electric field strength, B is the magnetic field induction, q is the charge of the ion and v is the velocity and \times the cross product. The force on one ion in a linear homogenous electric field in the direction of the electric field with positive ions is $F = qE$ and with negative ions is $-qE$ [Barshick *et al.*, 2000].

The time-of-flight mass analyzer uses a different approach to separate the ions. The mass (m) is determined according to the time it takes for the charged particles to move through a field-free flight tube with length (L) after the ions have been accelerated in an extraction field to a common energy (E). This can be described with the following equation:

$$E = \frac{mv^2}{2} = \frac{mL^2}{2zt^2} \quad (2.3)$$

where v is the velocity and t is the flight time. This equation written in terms of t shows that the less mass the ion has the higher its velocity and the earlier it arrives at the detector:

$$t = L \left(\frac{m}{2zE} \right)^{1/2} \quad (2.4)$$

The assignment of the masses of the ions is done by matching of the detection event and the associated time it took the ions to travel through the field free tube [Belu *et al.*, 2003]. The separation of the masses is done with a two meter reflectron type time-of-flight mass analyzer with first order energy focusing and a beam deflector for alignment and dynamic emittance matching. Before detection the ions enter a post acceleration unit. The detection of the ions is done with a combined micro channel plate (MCP) connected to a scintillator allowing reliable detection down to single ions. A multi-stop time-to-digital converter (TDC) with 50 ps time resolution is used to reconstruct the initial channel arrival times.

2.2.2 Measurement settings

The ToF-SIMS experiments were done on a ToF-SIMS⁵ instrument (Ion TOF, Münster, Germany) equipped with a dual-source column (DSC) for presputtering and depth profiling and a Bi-LMIG as primary ion source for spatially mass analysis. The DSC consists of an Oxygen-EI source (Electron Impact Gas Ion Source) and a Cs-Source (Thermal ionization Cesium source). The Bi emission current was set to 0.95 μA if not specified otherwise and all measurements were done in the high current bunched mode (HCBU) for high mass resolution ($M/\Delta M = 5000$). The settings for the individual measurements are displayed in Table 2.1 to Table 2.6. The scanning mode of the primary ion beam was random, charge compensation was not required. The samples were measured on different spots in negative and positive polarity.

If not specified otherwise mass calibration was done with at least four of the following ions when analyzing positive ions: $^{12}\text{C}^+$ (m/z 12.01), $^{12}\text{CH}_3^+$ (m/z 15.03), $^{16}\text{O}^+$ (m/z 16.00), $^{24}\text{Mg}^+$ (24.30 m/z), $^{12}\text{C}_2\text{H}_5^+$ (m/z 29.06), Ca^+ (24.30 m/z), $^{12}\text{C}_3\text{H}_7^+$ (43.09 m/z), $^{12}\text{C}_4\text{H}_9^+$ (57.11 m/z).

Mass calibration with negative ions was done with at least four of the following ions if not specified otherwise: $^{12}\text{C}_2^-$, $^{12}\text{C}_3^-$, $^{12}\text{C}_4^-$, $^{12}\text{C}_5^-$, PO_2^- , PO_3^- , I^- .

Labeling of Samples

The designation of experiments and samples comply with the following rules : the first 4 numbers refer to the month and the day of the measurement, the optional letter n and p before the first four numbers (e.g. p0103m*, n0103m*) limits the series to either only the positive or only the negative measurements of the day. What follows the first four numbers is of the following format (D stands for digit): mDDD (e.g. p0103m001). Surface spectroscopy has the format m00D, short co-sputtering with oxygen or cesium is indicated m10D and long co-sputtering is indicated by m11D (D increases starting from 1 to the number of the actual analysis). The asterisk is used if a series is addressed which was taken on a specific day. What is optional and only used in the results section is to a suffix to the sample name. The two options are either an E (for E. coli) or a N (for N. viennensis) to indicate the organism which was analyzed.

1207m*	p1207m001	p1207m002	n1207m003	p1207m004	n1207m005	p1207m006	n1207m007
Sample							
Substance	Crenarchaeol	Crenarchaeol	Crenarchaeol	1,2-Di-O-Phytanyl -sn-Glycerol	SAB	1,2-Di-O-Phytanyl -sn-Glycero-3- Phosphoethanolamin	SAB
Substrate	Si	Si	Si	Si	Si	Si	Si
Dissolved in	MeOH/CHCl ₃	MeOH/CHCl ₃	MeOH/CHCl ₃	CHCl ₃	CHCl ₃	CHCl ₃	CHCl ₃
Drying	1h@46°C	1h@46°C	1h@46°C	1h@46°C	1h@46°C	1h@46°C	1h@46°C
Analysis							
Primary Ion Species	Bi ₃ ⁺	Bi ₃ ⁺	Bi ₃ ⁺	Bi ₃ ⁺	Bi ₃ ⁺	Bi ₃ ⁺	Bi ₃ ⁺
Primary Ion K.E. [keV]	25	25	25	25	25	25	25
Polarity	positive	positive	negative	positive	negative	positive	negative
Ion current [pA]	0.354	0.354	0.354	0.354	0.354	0.354	0.354
Ion dose density [cm ⁻²]	5.05*10 ¹¹	5.05*10 ¹¹	5.05*10 ¹¹	5.07*10 ¹¹	5.07*10 ¹¹	5.07*10 ¹¹	5.07*10 ¹¹
Operation mode	HCBU	HCBU	HCBU	HCBU	HCBU	HCBU	HCBU
Raster Size (μm)	300	300	300	300	300	300	300
Raster Resolution (Pixel)	128	128	128	128	128	128	128
Number of scans	93	93	93	97	97	97	97
Total Ion Counts	28119970	26439987	30901743	33107113	11562616	25767564	9293507
DSC							
DSC species	-	-	-	-	-	-	-
Raster Size [μm]	-	-	-	-	-	-	-
Beam energy [keV]	-	-	-	-	-	-	-

Table 2.1: ToF-SIMS measurement parameters for lipids. Three pure substances in solution were measured with TOF-SIMS and their spectra were interpreted. For more information on the spectra see the results section. The abbreviation SAB stands for 'same as before', K.E. stands for 'Kinetic Energy', @stands for 'incubation at'.

0103m*	p0103m001	p0103m002	n0103m001	n0103m002
Sample				
Organism	E. coli	E. coli	E. coli	E. coli
Stain	KI	KI	KI	KI
Substrate	Si	Si	Si	Si
Fixation	EtOH/PFA	EtOH/PFA	EtOH/PFA	EtOH/PFA
Buffer	PBS	PBS	PBS	PBS
Drying	12h@46°C	12h@46°C	12h@46°C	12h@46°C
Analysis				
Primary Ion Species	Bi ₃ ²⁺	Bi ₃ ²⁺	Bi ₃ ²⁺	Bi ₃ ²⁺
Primary Ion K.E. [keV]	50	50	50	50
Polarity	pos	pos	neg	neg
Ion dose density[cm ⁻²]	2.64*10 ⁸	2.64*10 ⁸	2.64*10 ⁸	2.54*10 ⁸
Total dose	0.2*10 ¹¹	1.06*10 ¹¹	0.2*10 ¹¹	1.06*10 ¹¹
Ion current [pA]	0.063	0.063	0.063	0.063
Operation mode	HCBU	HCBU	HCBU	HCBU
Raster Size (μm)	200	500	200	500
Raster Resolution (Pixel)	1024	512	1024	512
Number of scans	10	52	10	50
Total ion dose	24895392	71970094	16812217	46566694
DSC				
DSC	-	-	-	-
Raster Size [μm]	-	-	-	-
Beam energy [keV]	-	-	-	-
Biomass	N	Y	Y	N
Crenarchaeol	N	N	N	N

Table 2.2: ToF-SIMS measurement parameters for 0103m*. N.c. = N. viennensis; E.c.= E. coli; K.E. = Kinetic Energy.

0208m*	p0208m101	p0208m102	n0208m101	n0208m102	p0208m001	p0208m002	p0208m005	n0208m001	n0208m002
Sample									
Organism	N.v.	E.c.	E.c.	N.v.	N.v.	N.v.	E.c.	N.v.	E.c.
Stain	KI	KI	KI	-	-	-	KI	KI	KI
Substrate	Si								
Fixation	EtOH/PFA								
Buffer	PBS								
Drying	12h@46°C								
Analysis									
Primary Ion Species	Bi ₃ ⁺								
Primary Ion K.E. [keV]	25	25	25	25	25	25	25	25	25
Polarity	pos	pos	neg	neg	pos	pos	pos	neg	neg
Primary ion current [pA]	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
Ion dose density [cm ⁻²]	1.01*10 ¹³	5.16*10 ¹²	1.01*10 ¹³						
Total Dose	6.53*10 ⁸	1.16*10 ⁹	6.53*10 ⁸						
Operation mode	HCBU								
Raster Size (μm)	80	150	80	80	80	80	80	80	80
Raster Resolution (Pixel)	256	256	256	256	256	256	256	256	256
Number of scans	124	220	124	124	124	124	124	123	123
Total Ion Counts	28980732	74299443	38570572	11046224	19458858	38061288	43140103	23211421	29548321
DSC									
DSC species	O ₂ ⁺	O ₂ ⁺	Cs ⁺	Cs ⁺	-	-	-	-	-
Raster Size [μm]	500	500	500	500	-	-	-	-	-
Kinetic energy [keV]	0.2	0.2	0.25	0.25	-	-	-	-	-
Ion current [nA]	71	71	10	10	-	-	-	-	-
Sputter time [s]	5	5	5	5	-	-	-	-	-
Ion dose density [cm ⁻²]	8.8*10 ¹⁶	8.8*10 ¹⁶	1.2*10 ¹⁶	1.2*10 ¹⁶	-	-	-	-	-
Biomass									
Biomass	Y	Y	N	N	Y	Y	Y	N	N
Crenarchaeol	N	N	N	N	N	N	N	N	N

Table 2.3: ToF-SIMS measurement parameters for experiments 0208m*. N.c. = N. viennensis; E.c.= E. coli; K.E. = Kinetic Energy

0208m*	p0208m111	p0208m112	n0208m111	n0208m111
Sample				
Organism	N.v.	E.c.	N.v.	E.c.
Stain	-	KI	-	KI
Substrate	Si	Si	Si	Si
Fixation	EtOH/PFA	EtOH/PFA	EtOH/PFA	EtOH/PFA
Buffer	PBS	PBS	PBS	PBS
Drying	12h@46°C	12h@46°C	12h@46°C	12h@46°C
Analysis				
Primary Ion Species	Bi ₃ ⁺	Bi ₃ ⁺	Bi ₃ ⁺	Bi ₃ ⁺
Primary Ion K.E. [keV]	25	25	25	25
Polarity	pos	pos	neg	neg
Ion current [pA]	0.92	0.92	0.92	0.92
Ion dose [cm ⁻²]	1*10 ¹³	1*10 ¹³	1*10 ¹³	1*10 ¹³
Total dose density	6.53*10 ⁸	6.53*10 ⁸	6.53*10 ⁸	6.53*10 ⁸
Operation mode	HCBU	HCBU	HCBU	HCBU
Raster Size (μm)	80	80	80	80
Raster Resolution (Pixel)	256	256	256	256
Number of scans	123	123	123	123
Total Ion Count	9402321	1542983	14567947	19822898
DSC				
DSC species	O ₂ ⁺	O ₂ ⁺	Cs ⁺	Cs ⁺
Raster Size [μm]	500	500	500	500
Kinetic energy [keV]	250	250	250	250
Ion current [nA]	71	71	10	10
Sputter time [s]	62	62	62	62
Ion dose density [cm ⁻²]	6.3*10 ¹⁸	6.3*10 ¹⁸	7.7*10 ¹⁷	7.7*10 ¹⁷
Biomass	Maybe	Maybe	1	1
Crenarchaeol	N	N	N	N

Table 2.4: ToF-SIMS measurement parameters for experiments 0208m*. N.c. = N. viennensis; E.c. = E. coli; K.E. = Kinetic Energy

0416*	p0416m101	p0416m102	p0416m001	p0416m111
Sample				
Organism	N.v.	E.c.	N.v.	N.v.
Stain	KI	-	KI	KI
Substrate	Si	Si	Si	Si
Fixation	EtOH/PFA	EtOH/PFA	EtOH/PFA	EtOH/PFA
Buffer	PBS	PBS	PBS	PBS
Method for drying	12h@46°C	12h@46°C	12h@46°C	12h@46°C
Analysis				
Primary Ion Species	Bi ₃ ⁺	Bi ₁ ⁺	Bi ₃ ⁺	Bi ₃ ⁺
Primary Ion K.E. [keV]	25	25	25	25
Polarity	pos	pos	pos	pos
Ion current [pA]	0.31	0.09	0.31	0.31
Ion dose density [cm ⁻²]	0.3*10 ¹³	0.1*10 ¹³	1.12*10 ¹⁷	0.3*10 ¹³
Total dose	6.53*10 ⁸	6.53*10 ⁸	1.12*10 ¹¹	6.53*10 ⁸
Operation mode	HCBU	HCBU	HCBU	HCBU
Raster Size (μm)	70	50	100	100
Raster Resolution (Pixel)	256	512	64	256
Number of scans	159	12	87071	159
Total Ion Count	32128882	3191964	1657560118	38928332
DSC				
DSC species	O ₂ ⁺	O ₂ ⁺	-	Cs ⁺
Raster Size [μm]	300	300	-	300
Kinetic energy [keV]	2	2	-	-
Ion current [nA]	153	153	-	11
Sputter time [s]	5	5	-	75
Ion dose density [cm ⁻²]	9.5*10 ¹⁷	9.5*10 ¹⁷	-	1.7*10 ¹⁸
Biomass	Y	Y	Y	Y
Crenarchaeol	N	N	N	N

Table 2.5: ToF-SIMS measurement parameters for experiments 0416m*. N.c. = N. viennensis; E.c.= E. coli; K.E. = Kinetic Energy

0525m*	p0525m101	p0525m102	p0525m103	p0525m104	p0525m001	p0525m002	p0525m111	p0525m112
Sample								
Organism	N.v.							
Stain	-	-	-	-	-	-	-	-
Substrate	ITO							
Fixation	-	-	-	-	-	-	-	-
Buffer	(NH ₄) ₂ CO ₃							
Drying	12h@80°C							
Analysis								
Primary Ion Species	Bi ₃ ⁺	Bi ₁ ⁺	Bi ₃ ⁺	Bi ₁ ⁺	Bi ₁ ⁺	Bi ₁ ⁺	Bi ₃ ⁺	Bi ₃ ⁺
Primary Ion K.E. [keV]	25	25	25	25	25	25	25	25
Polarity	pos							
Ion current [pA]	N/A							
Ion dose density [cm ⁻²]	N/A							
Total dose	N/A							
Operation mode	HCBU							
Raster Size (μm)	200	200	100	200	200	200	100	100
Raster Resolution (Pixel)	1024	1024	512	1024	1024	1024	512	256
Number of scans	2	6	7	1	1	2	18	103
Total Ion Count	43462584	31895412	5760531	1846567	10234189	13371390	7214763	8051319
DSC								
Depth profile species	O ₂ ⁺	O ₂ ⁺	O ₂ ⁺	O ₂ ⁺	-	-	O ₂ ⁺	O ₂ ⁺
Raster Size [μm]	500	500	500	500	-	-	500	500
Kinetic energy [keV]	1	1	1	1	-	-	1	1
Ion current [nA]	132	132	132	132	-	-	132	132
Sputter time [s]	5	5	5	5	-	-	9	52
Total dose density [cm ⁻²]	8.1*10 ¹⁷	8.1*10 ¹⁷	8.1*10 ¹⁷	8.1*10 ¹⁷	-	-	1.4*10 ¹⁸	8.5*10 ¹⁸
Biomass								
Biomass	N	Y	Y	Y	N	N	N	Y
Crenarchaeol	N	N	N	N	N	N	N	N

Table 2.6: ToF-SIMS measurement parameters for experiments 0525m*. The ion current was not recorded during this analysis so it was not possible to fill in the corresponding entries - N/A stands for entries not available. N.c. = *N. viennensis*; E.c. = *E. coli*; K.E. = Kinetic Energy

2.3 MALDI-ToF

2.3.1 Principles

Historically, gas chromatography coupled with mass spectrometry (GC-MS) was the technique for structural and quantitative analysis of lipids (alongside with silver ion and reversed-phase high-performance liquid chromatography). GC was used as method for separation, coupled with electron or chemical ionization and triple quadrupole analyzer (QMS) to further filter sample ions based on their mass-to-charge ratio (another possibility would be Fourier-transform infrared spectroscopy (FTIR) for detection). But complex lipids needed to be chemically modified when using GC-MS systems prior to analysis (the carboxyl group is usually derivatized with a reagent containing a nitrogen atom (e.g. either picolinyl (3-hydroxymethylpyridinyl) ester or 4,4-di-methyloxazoline (DMOX) derivatives) to introduce a functional group that can be easily ionized - minimizing the chance of double-bond ionization and migration which would make structural interpretation nearly impossible), a drawback that led to the investigation and invention of other methods [Christie, 1998].

The development of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was guided by the need to rapidly analyze macromolecular samples with accurate determination of their molecular mass with high mass resolution [Griesser *et al.*, 2004]. This was achieved and first reported by Karas and Hillenkamp in 1988 [Karas & Hillenkamp, 1988].

MALDI as well as ESI have the advantage of typically soft ionization when applied to biological molecules (proteins, lipids, metabolites, etc.) and therefore are able to yield high molecular mass spectra of biomolecules and yield the possibility to detect the intact molecule ion of interest without labeling. Combination of the MS techniques with separation methods such as liquid chromatography (LC) or gas chromatography and selective fragmentation by tandem mass spectrometry (MS/MS) and collision induced

decay (CID) further extended the range of samples this two ionization methods can be applied to. Which one of the two methods is used is usually determined by the aggregate state the sample and the analytical question addressed: ESI is used to obtain mass spectra from liquids whereas MALDI is applied for solid samples and can also be used for imaging (MALDI-MSI) [Sparvero *et al.*, 2012].

Usually for MALDI analysis, the sample is mixed with an organic matrix (usually 1:10⁴ in excess) that absorbs laser radiation and the mixture mounted on a stainless steel target. Analyte ions are generated by irradiation of the mixture of sample and matrix with a pulsed UV or IR laser (normally N₂ with $\lambda = 337$ nm or Nd/YAG with $\lambda = 355$) with a beam diameter of about 50 μm , thus limiting the resolution of the image to the beam diameter. The laser radiation is absorbed by the matrix and the matrix subsequently vaporized and, as part of this transformation this mechanism also ejects the analyte and ionizes it alongside. MALDI is nowadays mostly used with the ToF mass analyzer. The masses are separated based on their mass to charge ratio which is determined by the time they need to pass through a field free path. The advantage of a Time-of-Flight analyzer is the ability to detect simultaneously numerous compounds with excellent sensitivity [Carrol & Beavis, 1998, Benabdellah *et al.*, 2010].

The reason to analyze the pure lipids with MALDI-ToF lies in one advantage of the MALDI-ToF: the possibility to use ion gates. This enables the selection of an intact molecule with the ion gate and let the precursor molecule undergo post source decay and collision induced decomposition. This leads to characteristic fragments which can be linked directly to the precursor molecule and since the fragments from MALDI and SIMS are comparable it hints at fragments to look out in the ToF-SIMS spectra.

For the CID setting the precursor-of-interest passes a gas collision cell filled with helium (as collision gas) and fragments are formed while colliding with the inert noble gas [Pittenauer & Allmaier, 2009].

2.3.2 Measurement settings

All MALDI-ToF MS/MS measurements were performed on a AXIMA TOF² instrument [Kratos, UK]. This MALDI instrument is equipped with a N₂ laser which operates at $\lambda=337$ nm. The acceleration voltage was set to 20 kV. Only CID MS/MS was performed, using Helium as collision gas. Mass calibration was done with Na⁺, [THAP+H]⁺ and [THAP+Na]⁺. For mass calibration of the CID MS/MS measurement the internal master file was used which uses the peptide P14R1 as a reference for calibration. The assigned names of the fragments are in accordance to the suggested international nomenclature [Cheng *et al.*, 1998].

Chapter 3

Bioinformatic strategies, results and discussion

3.1 Introduction in the method

Candidates which could catalyze the head-to-head condensation and the cyclohexane ring formation were analyzed and ranked through phylogenetic considerations, phylogenetic patterns in orthologues cluster of proteins and biochemical plausibility.

3.1.1 Cluster of orthologues genes - Concept

Orthology is one of the two major forms of homology, the other one is paralogy. Orthology describes the concept of genes being of vertical descent from a single ancestral gene in the last common ancestor of the compared genomes in contrast to paralogy which defines genes that derive from a single gene through duplication without a speciation event [Koonin, 2005]. Therefore cluster of orthologues genes are the direct descendants of such a gene in the last common ancestor (LCA) of the species that are being compared. The basic idea is to obtain data from the eggNog-Database www.eggnog.embl.de for cluster of orthologues genes and to screen for enzymes which are either present in all thaumarchaeota or present in all archaea. Subsequently the resulting enzymes are scored

due to their assigned functions based on a homology search [Powell *et al.*, 2012].

To make up for the limited number of thaumarchaeota present in EggNog (only *Nitrosopumilus maritimus* SCM1) inParanoid (Stand-alone InParanoid Version 4.1) was used to obtain orthologues groups of genes considering five thaumarchaeota for which closed genomes were available [Remm *et al.*, 2001].

3.1.2 Cluster of orthologous genes - Data source

The eggNog-Database (evolutionary genealogy of genes: Non-supervised Orthologous Groups) contains orthologous groups of genes (COGs) with functional description and functional categories based on the original COG/KOG categories. The database spans over the three domains of life, containing 721,801 orthologous groups in 1133 species. OCG are calculated for 943 Bacteria, 69 Archaea and 121 Eukaryotes Clusters of orthologous groups of genes consists of individual proteins or groups of paralogs from at least 3 species and thus corresponds to a conserved domain. The COG assignments are the results of a homology search based on the SIMAP database.

The eggNog Database has at the moment only one thaumarchaeota - *Nitrosopumilus maritimus* SCM1 - integrated. Since there are more closed genomes of thaumarchaeota available we used inParanoid to screen for pairwise orthologous between *Nitrosopumilus maritimus* SCM1 and one of the following thaumarchaeota (obtained from the MicroScope Microbial Genome Annotation & Analysis Platform): (1) *Candidatus Nitrososphaera gargensis*, (2) *Candidatus Nitrosoarchaeum limnia* SFB1, (3) *Candidatus Nitrosoarchaeum koreensis* MY1, (4) *Cenarchaeum symbiosum* A. and a list was obtained with all original COGs of *Nitrosopumilus maritimus* SCM1 and all homologs from the other thaumarchaeota. This also means that with this analysis no new COGs were build - COGs were only appended if *Nitrosopumilus maritimus* SCM1 was present.

This data was integrated in the final analysis of the phylogenetic pattern.

3.1.3 Workflow

The workflow is schematically shown in Figure 3.1. A script (S) was written that reads in a list obtained from the EggNog-database and appends all the COGs for which inParanoid has found bidirectional best hits in Thaumarchaeota. This list of COGs is then analyzed based on the organisms which are present in the different COGs (this was done with two text files - the one contained for every TaxID a column with either the letter A (Archaea), B (Bacteria), E (Eukarya) and a second text file with a different letter for the five phyla of Archaea) and assignment based on the COG/KOG database was done for every COG in the form of a functional category (as described in [Tatusov *et al.*, 1997]) and a text description.

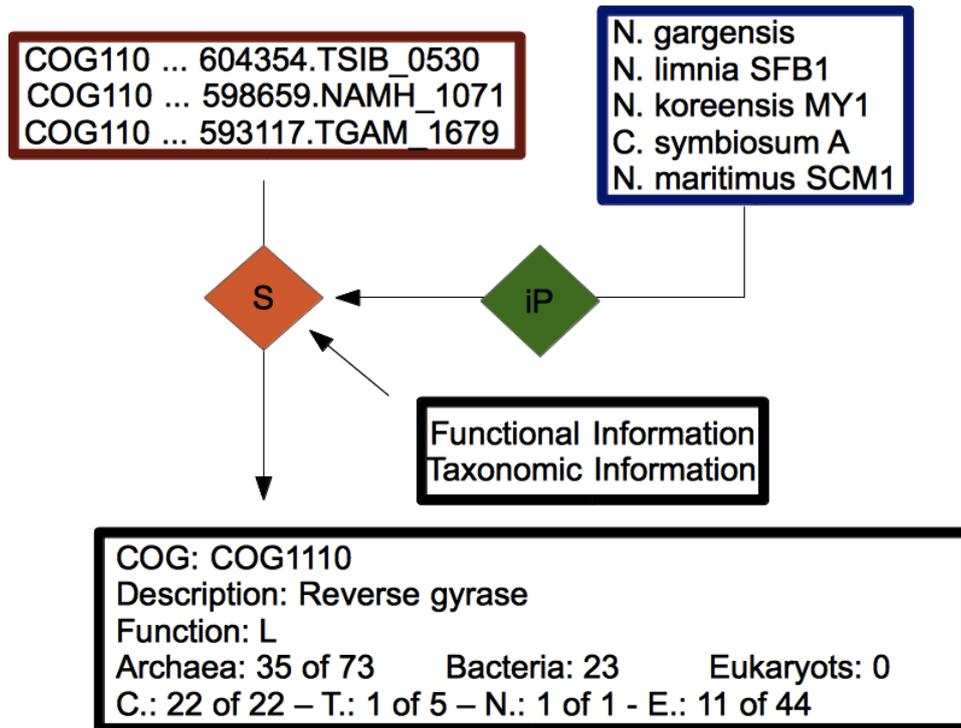


Figure 3.1: Workflow for the analysis of COGs. The red rectangle contains the original data from the EggNog-Database, the blue rectangle the genomic information from the MaGe database. iP stands for inParanoid, S is the actual script which deals with all the data and its output is a list of COGs with taxonomic and functional information - both in short form (18 letter code for different cellular functions) and a long, descriptive form.

3.2 Results and discussion

It should be noted that the functional description in the EggNog database for the COGs obtained from the original COG/KOG database (<http://www.ncbi.nlm.nih.gov/COG/new/>) is from the very same database and was last updated 2003. So the description of the COGs was further analyzed through a manually cross-platform analyses with the entry in the UniProt database for these COGs.

The one-letter-code of the functional categories is divided into four categories:

INFORMATION STORAGE AND PROCESSING: [J] Translation, ribosomal structure and biogenesis; [A] RNA processing and modification; [K] Transcription; [L] Replication, recombination and repair; [B] Chromatin structure and dynamics

CELLULAR PROCESSES AND SIGNALING: [D] Cell cycle control, cell division, chromosome partitioning; [Y] Nuclear structure; [V] Defense mechanisms; [T] Signal transduction mechanisms; [M] Cell wall/membrane/envelope biogenesis; [N] Cell motility; [Z] Cytoskeleton; [W] Extracellular structures; [U] Intracellular trafficking, secretion, and vesicular transport; [O] Posttranslational modification, protein turnover, chaperones

METABOLISM: [C] Energy production and conversion; [G] Carbohydrate transport and metabolism; [E] Amino acid transport and metabolism; [F] Nucleotide transport and metabolism; [H] Coenzyme transport and metabolism; [I] Lipid transport and metabolism; [P] Inorganic ion transport and metabolism; [Q] Secondary metabolites biosynthesis, transport and catabolism

POORLY CHARACTERIZED: [R] General function prediction only; [S] Function unknown;

Head-to-head condensation

The enzymes are first ranked based on the phylogenetic pattern: present in the maximum amount of Archaea (73) without regard of presence in other orders. After the ranking deletion of all enzymes which are - based on their functional annotation from the COG/KOG database and in accordance with the UniProt entry - not suitable to

catalyze a reaction in the lipid biosynthesis. In Table 3.1 the remaining enzymes which are not unsuitable based on their phylogenetic pattern or obvious biochemical considerations (e.g. RNA Synthetase) are shown. 4 of this 8 enzymes are available in all archaea - except Nanoarchaea - but have already a functional assignment. That this four enzymes are not present in Nanoarchaea is understandable since the Nanoarchaea known today do not synthesis their own lipids.

All four enzymes play a crucial role in the lipid biosynthesis, they are shown here because of their phylogenetic pattern and the possibility that if they catalyze so closely related reactions that they might as well have other substrates and products.

The other five - COG1646, COG1608, COG1371, COG2078, COG1602 - are present in most archaea (including all five thaumarchaeota). COG1646, COG1608, COG1371 and COG2078 are present in prokaryotes and eukaryots, COG1602 is the only COG in this list unique for archaea. COG1608 is the most unlikely candidate - a kinase is not very likely for this reactions. The other four are uncharacterized and fit the constrains - further investigations should be conducted to functionally characterize this COGs.

Cyclohexane formation

For the enzyme which catalyses the internal cyclohexane formation the screening method resulted in many more plausible candidates - shown in Table 3.2 - than for the head-to-head condensation. It is not possible to discriminate against further for this enzymes because there are no functional constrains available for the resulting COGs and the phylogenetic pattern is not very strictly defined, so the information of their abundance in other species is redundant and not helpful. In Figure 3.2 a part of the raw data (with the restriction that the COGs must present in all Thaumarchaeota and at least in 69 of 73 archaea) is shown - the x-axis correspond to the functional categories whereas the y-axis shows the abundance. Here should be noted that the two functional categories which are represented by the highest abundance are [J] and [R] - [J] being of no further interest since it was neither surprising that enzymes corresponding to translation, ribosomal structure

and biogenesis are very conserved within one domain nor was it likely that such an enzyme has also substrates being part of the lipid biosynthesis. Of further interest are mainly [L], [S] and [R] - [S] and [R] because since these are uncharacterized reactions with no known enzyme/s it is also plausible that the function is not predicted. Table 3.2 shows an sorted list (based on the number of archaea) of possible enzymes which should be functionally characterized in further biochemical experiments. The sorting of this list should not be considered as ranking - the cyclohexane forming enzyme/s can exclusively be present in Thaumarchaeota, but it is also possible that they are available with other substrates in bacteria and/or eukarya.

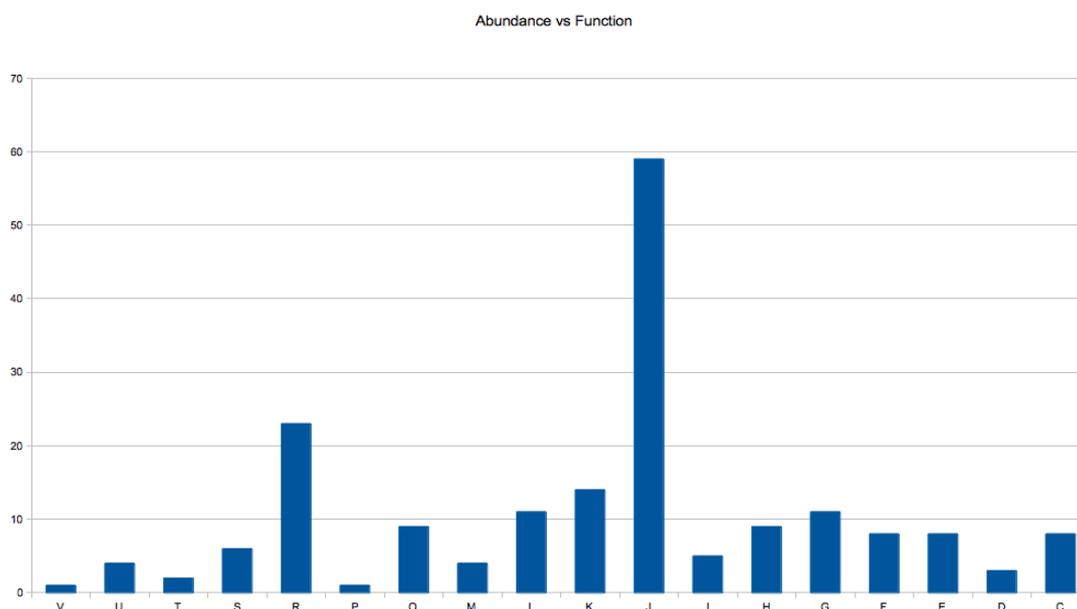


Figure 3.2: Function vs abundance. This is a histogram made from all COGs which are present in all five Thaumarchaeota and in at least 69 from 73 archaea (inclusive the Thaumarchaeota). The data was not further filtered for this representation - further filtering for this data was done based on functional annotation which gave as result the list in Table 3.2. With a few exceptions only proteins with functional category [L] - Lipid biosynthesis, [S] - Function unknown and [R] - General function prediction only were considered.

COG	Description	Function	Archaea	Bacteria	Eukarya	C.	T.	N.	K.	E.
COG3425	3-hydroxy-3-methylglutaryl CoA synthase	I	72	168	107	22	5	0	1	44
COG0020	Undecaprenyl pyrophosphate synthase	I	72	901	116	22	5	0	1	44
COG1257	Hydroxymethylglutaryl-CoA reductase	I	72	173	104	22	5	0	1	44
COG1646	Predicted phosphate-binding enzymes	R	72	72	0	22	5	0	1	44
COG1608	Predicted archaeal kinase	R	70	8	15	20	5	0	1	44
COG1371	Uncharacterized conserved protein	S	70	63	59	22	5	1	1	41
COG2078	Uncharacterized conserved protein	S	70	119	111	22	5	1	1	41
COG1602	Uncharacterized conserved protein	S	69	0	0	22	5	0	1	41

Table 3.1: Results for the head-to-head condensation based on phylogenetic patterns. C. = Crenarchaeota, T. = Thaumarchaeota, N. = Nanoarchaeota, K. = Korarchaeota, E.= Euryarchaeota

COG	Description	Function	Archaea	Bacteria	Eukarya	C.	T.	N.	K.	E.
COG2078	Uncharacterized conserved protein	S	70	119	111	22	5	1	1	41
COG1602	Uncharacterized conserved protein	S	69	0	0	22	5	0	1	41
COG1287	Uncharacterized membrane protein	R	69	25	116	20	5	0	1	43
COG1634	Uncharacterized Rossmann fold enzyme	R	68	0	0	19	5	0	1	43
COG2237	Predicted membrane protein	S	66	0	0	21	5	0	1	39
COG1931	Uncharacterized protein conserved in archaea	S	61	0	0	22	5	0	1	33
COG2090	Uncharacterized protein conserved in archaea	S	60	0	0	19	5	0	1	35
COG1698	Uncharacterized protein conserved in archaea	S	58	0	0	17	5	0	1	35
COG1679	Uncharacterized conserved protein	S	52	32	13	12	5	0	1	34
COG2412	Uncharacterized conserved protein	S	51	0	0	15	5	0	0	31
COG2835	Uncharacterized conserved protein	S	48	402	7	22	5	0	1	20
COG3356	Predicted membrane protein	S	45	0	0	19	5	0	0	21
COG1817	Uncharacterized protein conserved in archaea	S	41	14	0	7	5	0	1	28
COG1430	Uncharacterized conserved protein	S	41	260	2	5	5	0	0	31
COG2512	Uncharacterized membrane-associated protein	S	39	0	0	18	5	0	1	15
COG1641	Uncharacterized conserved protein	S	37	142	4	0	5	0	1	31
COG1711	Uncharacterized protein conserved in archaea	S	36	0	0	0	5	0	0	31
COG1915	Uncharacterized conserved protein	S	30	45	0	0	5	0	0	25
COG4046	Uncharacterized protein conserved in archaea	S	28	0	0	20	5	0	1	2
COG2107	Predicted periplasmic solute-binding protein	R	27	109	1	18	5	0	0	4
COG0586	Uncharacterized membrane-associated protein	S	27	720	4	11	5	0	1	10
NOG114159	Uncharacterized	S	25	0	0	13	5	0	1	6
COG4911	Uncharacterized conserved protein	S	17	18	0	12	5	0	0	0
COG3875	Uncharacterized conserved protein	S	15	76	4	0	5	0	1	9
NOG124503	Uncharacterized	S	11	0	0	6	5	0	0	0
NOG131523	Uncharacterized	S	9	0	0	4	5	0	0	0
NOG87543	Uncharacterized	S	9	0	0	0	5	0	0	4
NOG129618	Uncharacterized	S	8	0	0	0	5	0	0	3
COG1615	Uncharacterized conserved protein	S	7	150	0	0	5	0	0	2
NOG81378	Uncharacterized	S	7	0	0	0	5	0	0	2
NOG248731	Uncharacterized	S	6	8	0	0	5	0	0	1
NOG76061	Uncharacterized	S	6	4	0	1	5	0	0	0
NOG47610	Uncharacterized protein	S	5	12	0	0	5	0	0	0

Table 3.2: Results for the internal cyclisation to form a cyclohexane based on phylogenetic patterns. C. = Crenarchaeota, T. = Thaumarchaeota, N. = Nanoarchaeota, K. = Korarchaeota, E.= Euryarchaeota

Chapter 4

Characterization of Lipids

Three pure substances were analyzed with MALDI-ToF in positive ion mode and ToF-SIMS in positive and negative ion mode. The obtained peaks were further interpreted and if possible their molecular formula was calculated. For ToF-SIMS spectra this was done based on the accurate mass of the peak in comparison with the isotope abundance in addition to basic chemical interpretation. Isotope abundances were calculated either with the provided ION-TOF software (Surfacelab6) or with the web tool ChemCalc ¹ (contact: Luc Patiny Director of Chemical Information, Ecole Polytechnique Fédérale de Lausanne (EPFL)).

4.1 Spectral characterization of 1,2-Di-O-Phytanyl-sn-Glycerol

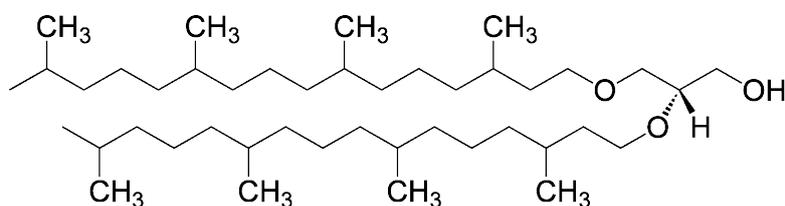
4.1.1 Sample preparation

The polar 1,2-Di-O-Phytanyl-sn-Glycerol (abbreviation PG) is commercially available (Avanti Polar Lipids, 9938 AB, Netherlands). It is dissolved in 1 mL CH₃Cl at a concentration of 5 mg/mL. Prior to analysis the lipid was stored in a freezer at -18°C.

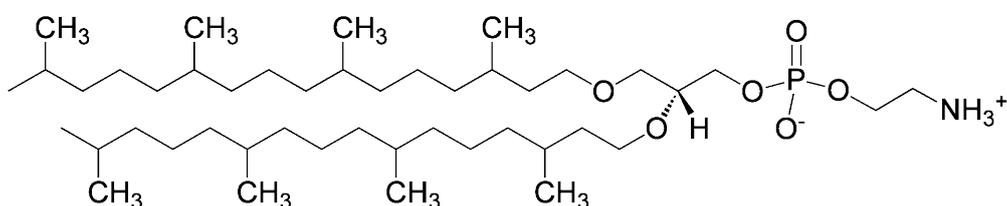
For the ToF-SIMS measurements 20 μL of the solution were pipetted on a Si 111 wafer (Infineon, Austria).

For the MALDI-ToF analysis 2',4',6'-Hydroxyacetophenone (THAP) (Sigma-Aldrich,

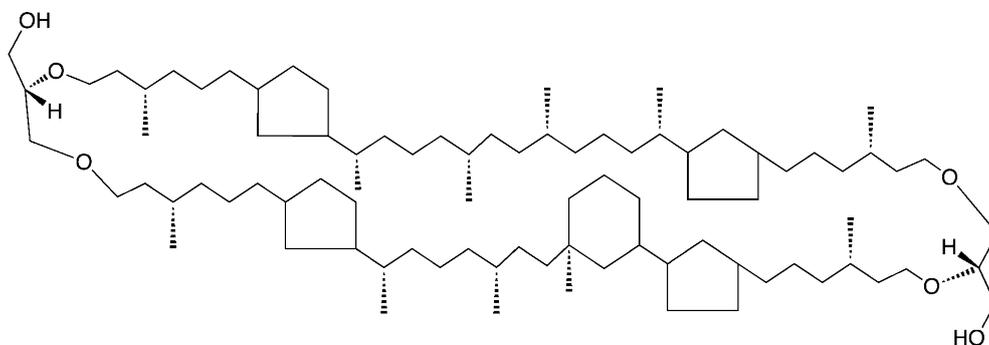
¹<http://www.chemcalc.org/>

1,2-Di-O-Phytanyl-sn-Glycerol

Chemical Formula: $C_{43}H_{88}O_3$
 Exact Mass: 652.67
 Molecular Weight: 653.16
 Elemental Analysis: C, 79.07; H, 13.58; O, 7.35

1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamin

Chemical Formula: $C_{45}H_{94}NO_6P$
 Exact Mass: 775.68
 Molecular Weight: 776.20
 Elemental Analysis: C, 69.63; H, 12.21; N, 1.80; O, 12.37; P, 3.99

Crenarchaeol

Chemical Formula: $C_{86}H_{162}O_6$
 Exact Mass: 1291.24
 Molecular Weight: 1292.20
 Elemental Analysis: C, 79.93; H, 12.64; O, 7.43

Figure 4.1: Chemical structure of the glycerolipids. The structures were obtained from Avanti, Polar Lipids, The Netherlands for the first two and for crenarchaeol it is based on the structure from [Damsté *et al.*, 2002]. It should be noted here that the structure for crenarchaeol in Wikipedia, the NCBI PubChem and the Lipid MAPS are missing four CH_2 units and therefore wrong.

Austria) was used as matrix. A stock solution of 20 mg THAP and 1 mg NaCl dissolved in 1 mL methanol (Sigma-Aldrich, Austria) was prepared and ultra-sonicated for 5 minutes. Three different samples were spotted on the MALDI target, which had been cleaned with acetone (Sigma-Aldrich, Austria) and 2-propanol (Sigma-Aldrich, Austria) prior to analysis. The first sample contained 0.8 μL matrix, the second sample contained 0.8 μL matrix and the analyte PG (1:100 diluted in the matrix) and the third sample contained the matrix plus a standard (ricinoleic acid) to calibrate the mass spectrum.

The sample spots on the MALDI target were prepared using the dried droplet technique, which means that the samples were pipetted on the sample spots of a stainless steel target and the organic solvents evaporated. Four spots were mounted for each of the three different solutions.

For the full MALDI-ToF spectra about 200 profiles (corresponding to the number of laser pulses (200) and for every laser pulse a spectrum is acquired) and for the CID the number of laser pulses and acquired spectra were about 5000.

4.1.2 MALDI-ToF measurements results and discussion

The structure of 1,2-Di-O-Phytanyl-sn-Glycerol is shown in Figure 4.1. Spectra were only obtained in positive mode because it has been observed that the sodium adducts of the intact lipid molecules yield the highest sensitivity [Pittenauer & Allmaier, 2009].

Positive mode

The full MALDI-MS spectrum is shown in Figure 4.2, the intact molecule peak as sodium adduct $[M+\text{Na}]^+$ is located at m/z 676. The spectrum obtained from collision induced decay (CID) of the intact sodium adduct is shown in Figure 4.3. The full MALDI-ToF spectrum shows two different precursor ions at m/z 676 and 677 (others are shown as well, e.g. 677.1 but mass differences of a factor of lower than 1 u are not useful to consider at this point because the mass window of the ion gate for selecting precursor ions was adjusted to ± 1 u). The first ion at m/z 676 is the Na adduct of the

intact 1,2-Di-O-Phytanyl-sn-Glycerol and the ion molecule at 677 m/z represents the protonated molecular ion Na adduct $[M+H+Na]^+$. The molecule peak with the highest intensity ($[M+Na]^+$) was selected as a precursor ion and the ion gate was set symmetrically at $m/z 676 \pm 2.5$ u and collision induced decay was performed to investigate the fragmentation of the intact molecule and to analyse which fragments are obtained from 1,2-Di-O-Phytanyl-sn-Glycerol.

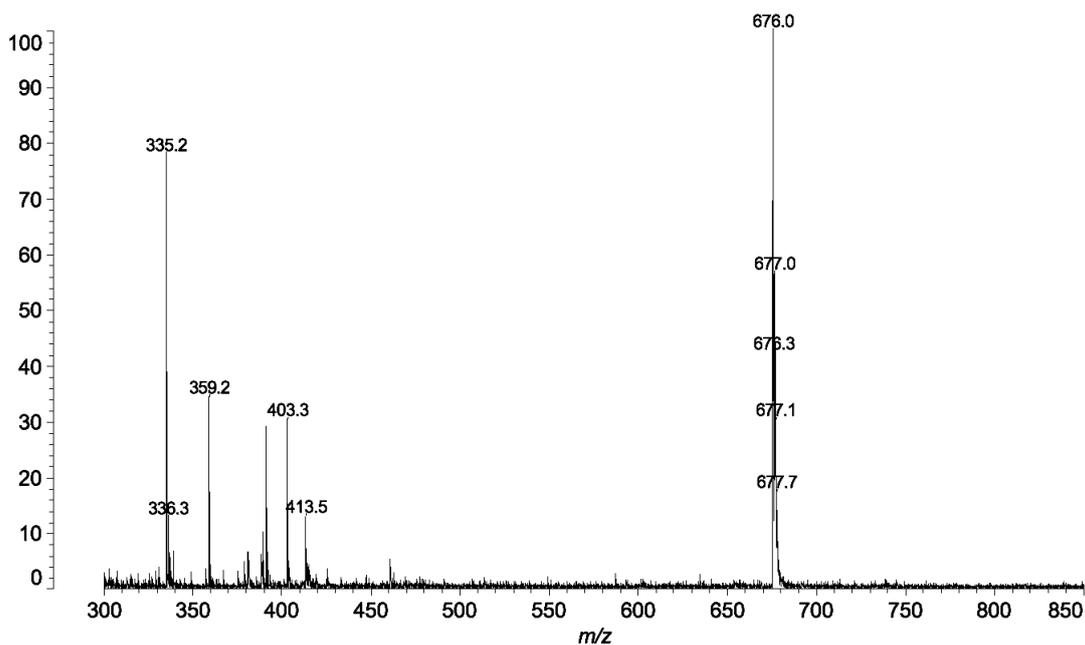


Figure 4.2: MALDI-ToF spectrum of 1,2-Di-O-Phytanyl-sn-Glycerol in positive mode. The most prominent peak is the Na adduct at $m/z [653+23]^+$. Shown is the spectrum between 300 and 850 mass units, no further smoothing was performed. Normalization was done relative to the peak with highest intensity.

The CID spectrum is shown in Figure 4.3. The masses below 81 units can not be explained by fragmentation. What is most striking about the other fragments is the fact that they are all observed as sodium-adducts and no other positive charge was introduced (like protonation of the free hydroxy group). The fragmentation scheme is shown in Figure 4.4. All introduced double bonds should be considered as delocalized rather than fixed. For easier interpretation, the positions were assigned randomly.

The fragment "A"+Na⁺ with m/z 506 can be explained through introduction of olefins and elimination of a C₁₂ group.

The fragment "B"+Na⁺ with m/z 81 represents iso-butane and can be regarded as result of continuous elimination of ethyl and iso-butane moieties of the phytanyl chain. The iso-butane most likely originates from the tertiary carbons of the phytanyl chain and appears as the sodium adduct in the spectrum.

The fragment "C"+Na⁺ is formed as glycerol ion with one intact (or partially reduced) phytanyl chain. The glycerol backbone stays intact but the hydroxy moiety can also be eliminated. Multiple combinations are possible, thus the two unspecific generic labels (R₁ and R₂).

The fragment "D"+Na⁺ with m/z 153 consists of an intact glycerol moiety and additionally an ethyl bound via an ether linkage. This ion is detectable as Na adduct.

A comparison of the spectrum with the following positive ToF-SIMS spectrum of the same substance (shown in Figure 4.6) shows no match of the besides the molecular ion peak and the Na adducts. This might be due to the fact that SIMS is to a much lesser extent a soft ionization method than MALDI and it can very well be the case that all longer fragments are destructed in this process. In MALDI the energy is normally absorbed by the matrix and not the analyte itself - therefore the amount of energy transferred to the analyte is much smaller. The difference between the MALDI and the ToF-SIMS spectrum might also correspond to the differences in ionization/fragmentation mechanisms.

4.1.3 TOF-SIMS measurements results and discussion

Positive mode

Mass calibration was done with C⁺ (m/z 12.01), CH₃⁺ (m/z 15.03), C₂H₅⁺ (m/z 29.06), C₃H₇⁺ (43.09 m/z), C₄H₉⁺ (57.11 m/z).

The structure of 1,2-Di-O-Phytanyl-sn-Glycerol is given in Figure 4.1. The obtained results were compared to [Heim *et al.*, 2009] peak list for archaeol. A peak for the molecule

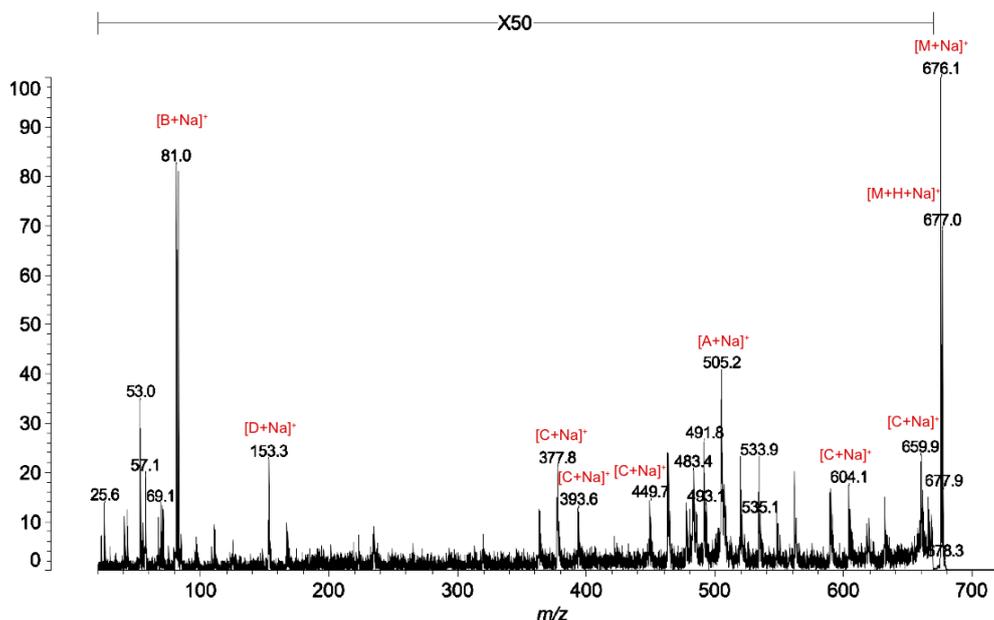


Figure 4.3: MALDI-CID Spectrum of 1,2-Di-O-Phytanyl-*sn*-Glycerol from precursor ion at m/z $[653+23]^+$ in positive mode. Shown is the spectrum (profiles 1-5,000) between 20 and 700 mass units, no further smoothing was performed. Normalization was done relative to the peak with highest intensity. The 50x correspond to the whole spectrum except the mass at $[653+23]^+$.

in protonated form $[M+H]^+$ (at 653 m/z) and as Na adduct $[M+Na]^+$ (at 675 m/z), as well as the dehydrated form $[M-H_2O]^+$ (at 634 m/z) were detected and are in accordance to [C. Heim *et al*]. What was further detected was a fragment at m/z 620 - but this fragment had such low intensity in contrast to the other peaks that it was not further considered for this analysis. Furthermore, three fragments were detected at m/z 211, 281, 371/373 (371 corresponds to the 373 molecule - 2H). The fragment at 281 is a polysiloxane contamination (also in accordance to C. Heim *et al*). The fragment at m/z 91, specific for glycerol, is not shown because the mass range was set from 150-1400 u.

Negative mode

The obtained results were compared to the peak list for archaeol in C. Heim *et al*. A peak for the deprotonated molecule as $[M-H]^-$ (at 651 m/z) is recognizable but less pronounced than in the positive spectrum - comparing the intensity of the two

No	Center Mass	Assignment	Peak Area [a.u.]
1	212	$C_{15}H_{33}^+$	29621
2	373	$C_{22}H_{45}O_4^+$	1544
3	635	$[M-H_2O]^+$	842
4	654	$[M+H]^+$	295
5	654	$[M+H]^+$	11711
6	676	$[M+Na]^+$	3200

Table 4.1: List of fragments for 1,2-Di-O-Phytanyl-sn-Glycerol obtained in positive mode. [a.u.] stands for arbitrary unit.

different ions obtained with comparable settings. Archaeol clearly favors the formation of positively charged secondary ions, in particular through formation of sodium adducts.

No	Center Mass	Assignment	Area [a.u.]
1	91	$C_3H_7O_3^-$	37757
2	295	$C_{20}H_{39}O^-$	50078
3	297	$C_{20}H_{41}O^-$	18053
4	371	$C_{22}H_{43}O_4^-$	4455
5	652	$[M-H]^-$	576
6	668	$[M-OH]^-$	141

Table 4.2: List of fragment for 1,2-Di-O-Phytanyl-sn-Glycerol obtained in negative mode

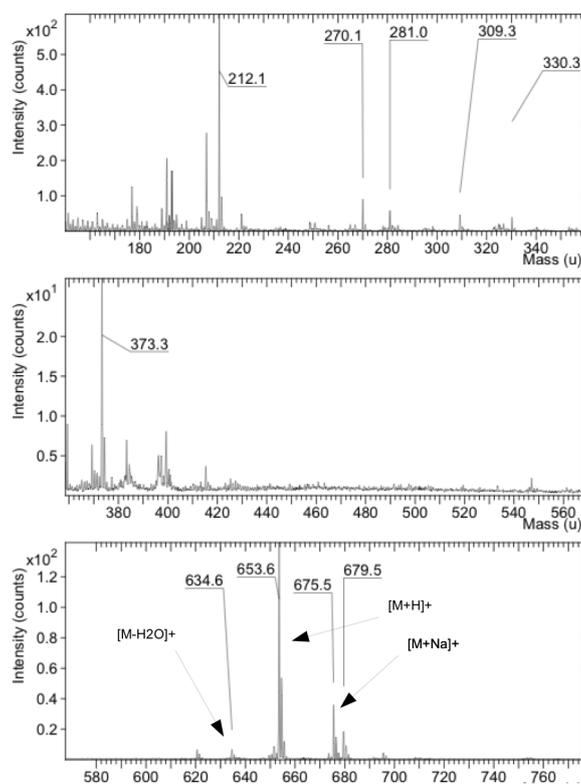


Figure 4.6: ToF-SIMS: Spectrum for 1,2-Di-O-Phytanyl-*sn*-Glycerol in positive mode. Mass range from 150 to 1400 u. Information about the peaks is provided in the text.

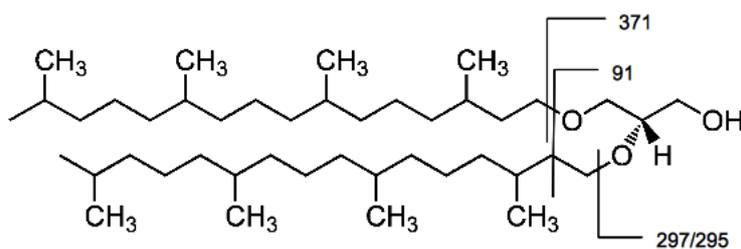


Figure 4.7: ToF-SIMS: proposed fragmentation of 1,2-Di-O-Phytanyl-*sn*-Glycerol in negative mode.

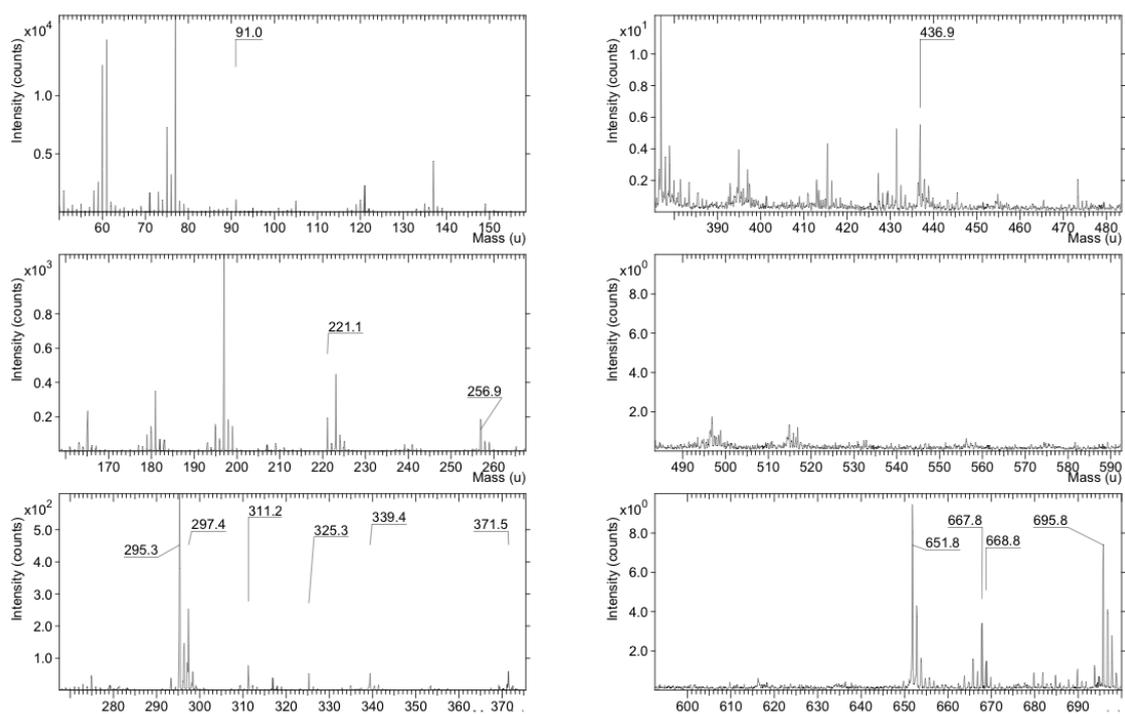


Figure 4.8: ToF-SIMS: spectrum for 1,2-Di-O-Phytanyl-sn-Glycerol in negative mode. Mass range from 150 to 1400 u. Information about the peaks is provided in the text

4.2 Spectral characterization of 1,2-Di-O-Phytanyl-sn-Glycerol-3-Phosphoethanolamine

4.2.1 Sample preparation

The polar 1,2-Di-O-Phytanyl-sn-Glycerol-Ethanolamine (abbreviation PGE) is commercially available (Avanti Polar Lipids, 9938 AB, Netherlands). It is delivered as a solution in 1 mL CH₃Cl with a concentration of 5 mg/mL. Prior to analysis the lipid was stored in a freezer at -18°C.

For the ToF-SIMS measurements the dried droplet technique was applied. 5 μ L of the solution were pipetted on a Si 111 wafer (Infineon, Austria) and dried in a heating unit. For the MALDI-ToF analysis 2',4',6'-Hydroxyacetophenone (THAP) (Sigma-Aldrich, Austria) was used as matrix. A stock solution of 20 mg THAP and 1 mg NaCl (Sigma-Aldrich, Austria) dissolved in 1 mL methanol (Sigma-Aldrich, Austria) was prepared and ultra-sonicated for 5 minutes. Three different samples were spotted on the MALDI target, which had been cleaned with 2-propanol (Sigma-Aldrich, Austria) and acetone (Sigma-Aldrich, Austria) prior to analysis. Three different samples were prepared for this analysis: (1) 0.8 μ L matrix, (2) 0.8 μ L matrix and the analyte PGE (1:200 diluted) and (3) the matrix plus a standard to calibrate the spectrometer (Ricinoleic acid).

The sample spots on the MALDI target were prepared by the dried droplet technique. This means that aliquots of the sample solution were pipetted on the sample spots of a stainless steel target and the organic solvent were evaporated.

4.2.2 MALDI-ToF measurements results and discussion

The structure of 1,2-Di-O-Phytanyl-sn-Glycerol-Ethanolamine is given in Figure 4.1. The spectrum was obtained in positive mode.

Positive mode

The full MALDI-MS spectrum is displayed in Figure 4.9. No further data processing was performed. Three peaks were selected for further CID analysis : (1) m/z 777 [M+H]⁺

(2) m/z 799 $[M+Na]^+$ (3) m/z 820 $[M-H+2Na]^+$ (shown in Figure 4.10). The peak at m/z 207 $[THAP+K]^+$ and 335 (THAP dimer) are originate from the THAP matrix. The ToF-SIMS spectrum (shown in Figure 4.14) shows peaks at m/z 820, 798, 776, 754, 732, 400, 398, 355, 324, 281, 221, 207, 147, 142, 124 and 91. Unexplained peaks are at m/z 732, 400, 398, 355, 324, 281, 221 and 207. Except the intact lipid molecule, the Na adducts and the protonated phosphoethanolamine, none of the MALDI and SIMS fragments have the same or arguable similar m/z ratio.

For further characterization of the diphytanyl glycerol ethanol amine MALDI-ToF MS/MS CID experiments were performed. The measurements were performed for the protonated molecular ion at m/z 777 ($[M+H]^+$) and the two Na adducts at m/z 799 and m/z 821 as a precursor ion and the ion gate was set symmetrically at m/z 776.6, 798.81 and 820.86 ± 2 units. The peaks of all three spectra were analyzed.

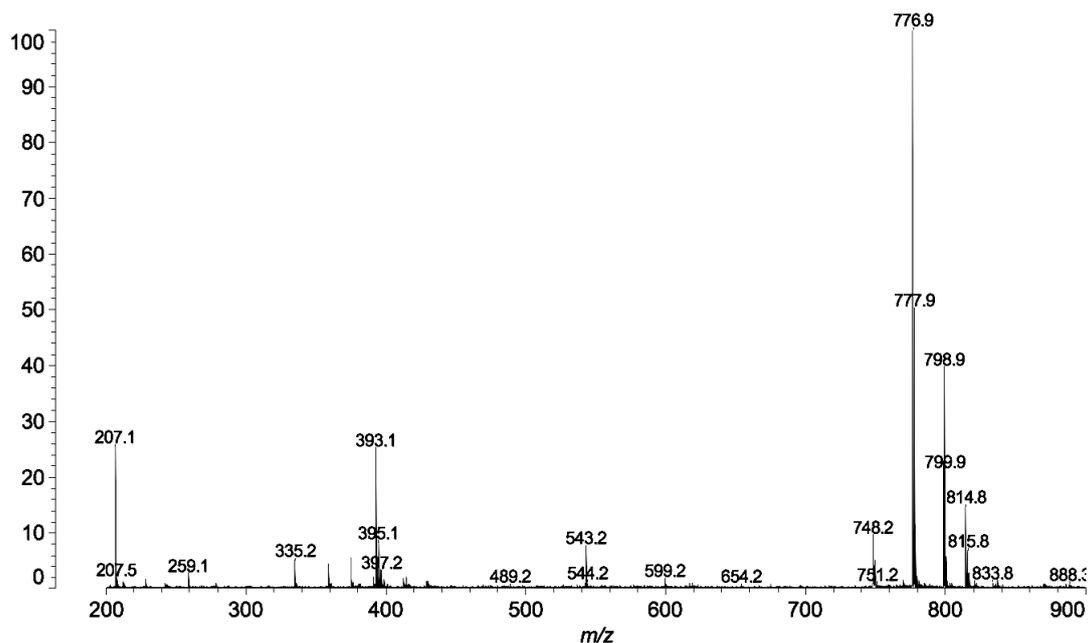


Figure 4.9: Full MALDI-ToF spectrum of 1,2-Di-O-Phytanyl-sn-Glycerol-Ethanolamine in positive mode. The most prominent peak represents the protonated molecule at m/z $[776+H]^+$, followed by the Na adduct $[M+Na]^+$ at 799 and $[M-H+2Na]^+$ at m/z 820. Shown is the spectrum between 200 and 900 mass units, no further smoothing was performed. Normalization was done relative to the peak with the highest intensity.

The peaks in the range from m/z 493 to 776 correspond to the stepwise cleavage of carbon atoms and the introduction of olefine structures in one of the phytanyl chains (sometimes as methane - m/z 16 , methyl - m/z 15 but also as carbene (methylene) - m/z 14). But there are also other cleavages: the mass difference between the peak at m/z 537 and 594 (Δ 57) corresponds most likely to a butyl substituent group, the difference between m/z 635 and 666 (Δ 30) to a formaldehyde, the difference between m/z 706 and 734 (Δ 30) is most likely due to the elimination of a ethylene and the mass difference between the peak at 776 and 735 (m/z 41) corresponds to a protonated propin. This fragments can be added to the fragment ion named "A+R" to give the corresponding mass, partial deprotonation and ionization assumed.

The following mentioned fragments are shown in Figure 4.11. The generic labels (R_1 and R_2) refer to moieties explained in more detail in Figure 4.11. The fragment "A+R" has one intact phytanyl chain bound to a phosphoglycerol and two unspecific generic labels (R_1 and R_2), one at the sn-1 glycerol hydroxy group and the other on the phosphoester originally connecting to the ethanolamine.

The fragment $[D-27]^+$ indicates a partial elimination of the polar head group as phosphoethanolamine. The phosphoethanolamine forms as separate fragment (but only in the CID spectrum of the protonated intact molecule which is not shown here). The fragment $[D]^+$ contains an ethylen moiety attached to a hydroxy group to form a phosphoester.

Fragment B can be explained by internal rearrangement of part of the branched phytanyl chain and the sn-2 carbon atom of the glycerol moiety. The molecule forms a tetrahydropyran structure with two corresponding unspecific generic labels (R_1 and R_2). This fragment determines the diether and phosphoglycerol structure of the lipid. In phosphoglycerol diester the remaining carboxyl group forms the tetrahydropyran structure resulting in a lactol group.

The protonated fragment $[C]^+$ is a rearrangement of the phosphoric acid group form-

ing a heterocycle. This fragment is unique for phospholipids.

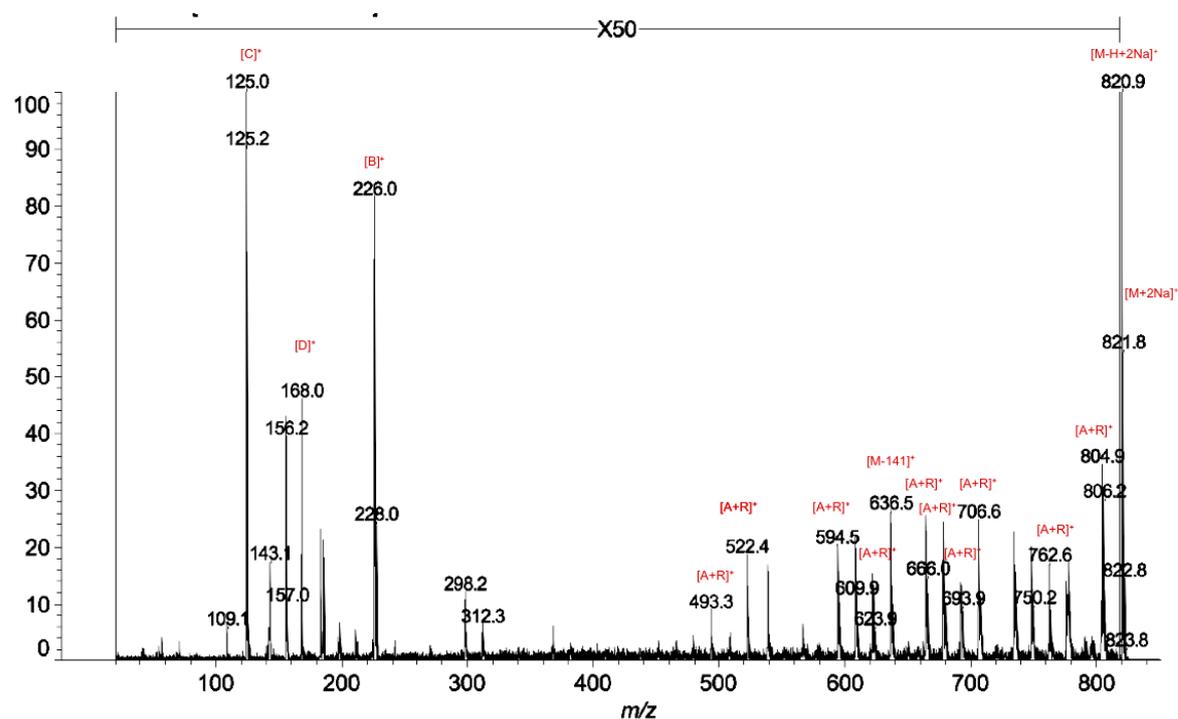


Figure 4.10: MALDI-CID: Spectrum of 1,2-Di-O-Phytanyl-sn-Glycerol-Ethanolamine from precursor ion $[M-H+2Na]^+$ at m/z 820 in positive mode. Shown is the spectrum (profiles 1-5,000) between 0 and 850 units, no further smoothing but baseline subtraction was performed. Normalization was done relative to the peak with the highest intensity.

4.2.3 TOF-SIMS measurements results and discussion

Positive mode

The structure of 1,2-Di-O-Phytanyl-sn-Glycerol-Ethanolamine is given in Figure 4.1, the ToF-SIMS spectrum is shown in Figure 4.14. There is no reported spectrum for PGE prior to this analysis and so it is not possible to compare the results to known fragmentation.

The most prominent and conceivable peaks are listed in Table 4.3. The first two peaks at m/z 91 and m/z 124 are most likely the glycerol and the phosphoethanolamine moieties, followed by a signal at 142 which is very likely the hydrated form of phosphoethanolamine.

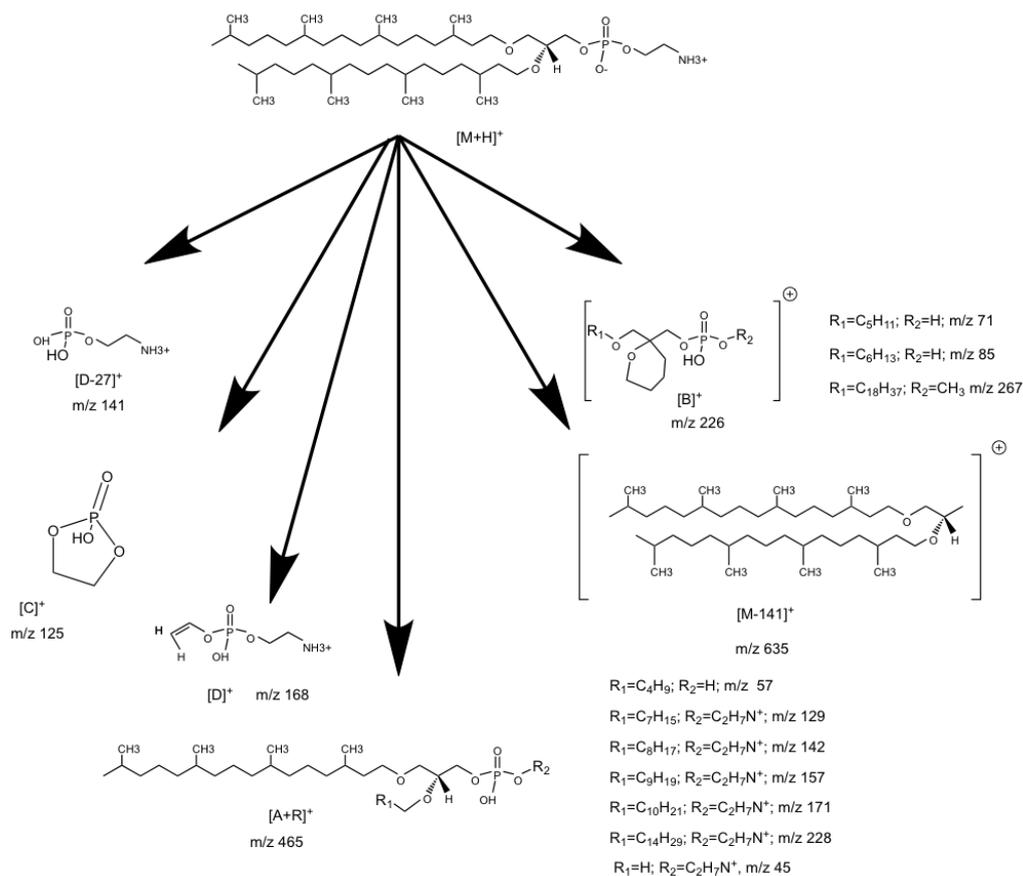


Figure 4.11: MALDI-ToF: proposed fragmentation scheme for 1,2-Di-O-Phytanyl-sn-Glycerol-Ethanolamine in positive mode.

Three sodium adducts are likely represented in the spectrum. A phosphoethanolamine with Na, $[PGE+Na]$ and $[PGE+2Na]$. The peak for the protonated form of the molecule is at m/z 777. This peak and the Na adducts of the peak are characteristic (and also visible in the MALDI spectrum), the others are very likely independent of the lipid structure but unique for the polar head group. The peak at 281 most likely originates from a siloxane contamination (derived from polydimethylsiloxane (PDMS)).

No	Center Mass	Assignment	Area [a.u.]
1	91	$\text{C}_3\text{H}_7\text{O}_3^+$	97696
2	124	$\text{C}_2\text{H}_7\text{NPO}_3^+$	148713
3	142	$\text{C}_2\text{H}_9\text{NPO}_4^+$	141621
4	147	$[\text{C}_2\text{H}_7\text{NPO}_3+\text{Na}]^+$	541104
5	777	$[\text{M}+\text{H}]^+$	3493
6	799	$[\text{M}+\text{Na}]^+$	1228
7	821	$[\text{M}-\text{H}+2\text{Na}]^+$	1787

Table 4.3: ToF-SIMS: list of fragments for 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamine obtained in positive mode

Negative mode

The peaks at m/z 137, 163, 223 are siloxane contaminations. The polar head group is responsible for the first signals at m/z 109 for CH_3PO_4^- and m/z 140 for $\text{C}_2\text{H}_7\text{NPO}_4^-$ as well as m/z 158 as Diphosphate $\text{P}_2\text{O}_6\text{H}^-$. The peak at m/z 476 likely corresponds to a fragment with - fragmentation shown in Figure 4.13 - as $\text{C}_{25}\text{H}_{49}\text{PO}_6^-$. The deprotonated molecule peak at m/z 775 $[\text{M}-\text{H}]^-$ is the peak with most intensity in the high mass range above 500 u. Also shown is the dehydrated and deprotonated peak at m/z 757 $[\text{M}-\text{H}_2\text{O}]^-$ and $[\text{M}]^-$. The peak at m/z 772 may also result from deprotonation.

No	Center Mass	Assignment	Area [a.u.]
1	110	CH_3PO_4^-	134695
2	140	$\text{C}_2\text{H}_7\text{NPO}_4^-$	122104
3	159	$\text{P}_2\text{O}_6\text{H}^-$	77531
4	476	$\text{C}_{25}\text{H}_{49}\text{PO}_6^-$	17787
5	758	$[\text{M}-\text{H}_2\text{O}]^-$	3748
6	773	$[\text{M}-3\text{H}]^-$	10787
7	775	$[\text{M}-\text{H}]^-$	97653
8	776	$[\text{M}]^-$	1657

Table 4.4: ToF-SIMS: list of fragments for 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamine obtained in negative mode.

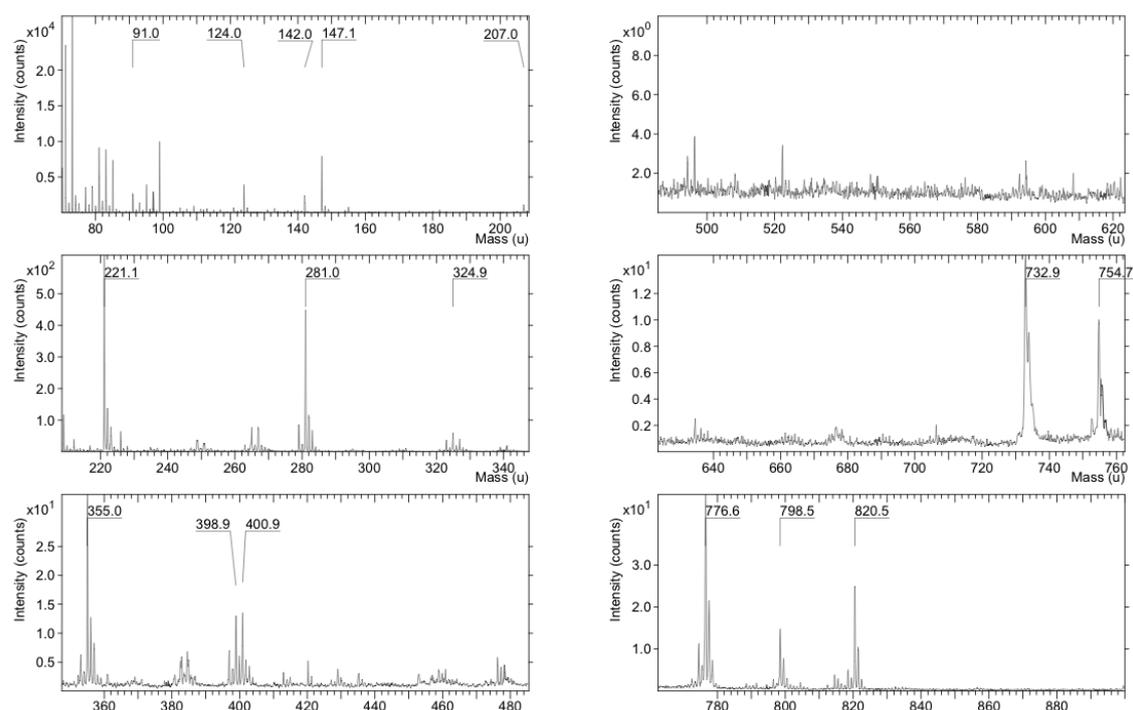


Figure 4.12: ToF-SIMS: spectrum for 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamine in positive mode. Mass range from 150 to 1400 u. Information about the peaks is provided in the text.

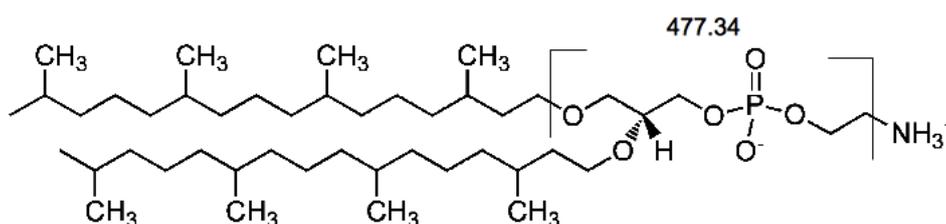


Figure 4.13: ToF-SIMS: proposed fragmentation of 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamine in negative mode for the peak at m/z 476.34. Shown is the protonated form.

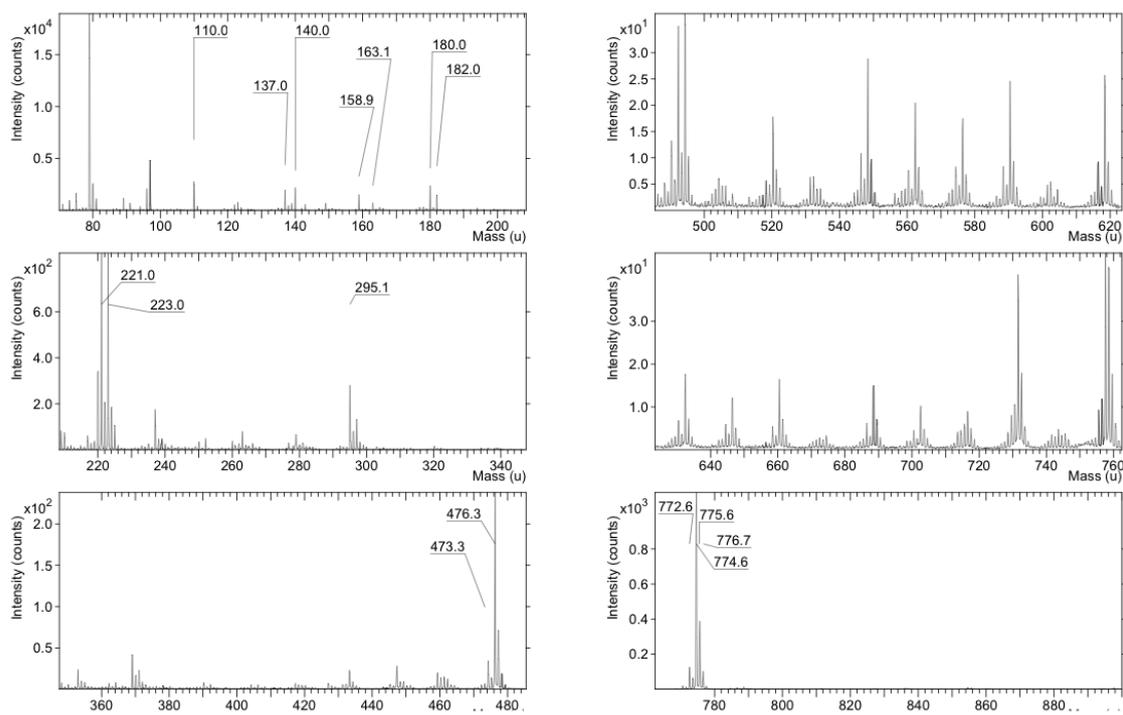


Figure 4.14: ToF-SIMS: spectrum of 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamine in negative mode. Mass range from 50 to 1400 u. Information about the peaks is provided in the text

4.3 Spectral characterization of Crenarchaeol

4.3.1 Sample preparation

Crenarchaeol (in the following abbreviated as 'CAO') is not commercially available, it was kindly provided by Jaap Damsté (Royal Netherlands Institute for sea research). The component was received without polar head-groups. It was diluted in 1:1 (v/v) MeOH/CH₃Cl (unknown manufacturer) at an unknown concentration. Prior to analysis, the lipid was stored in a freezer at -18°C.

For the ToF-SIMS measurements the dried droplet technique was used. 5 μ L of the solution were pipetted on a Si 111 wafer (Infineon, Austria) and dried in a heating unit. For the MALDI-ToF analysis 2',4',6'-Hydroxyacetophenone (THAP) (Sigma-Aldrich, Austria) was used as matrix. A stock solution of 20 mg THAP and 1 mg NaCl dissolved in 1 mL methanol (Sigma-Aldrich, Austria) was prepared and ultra-sonicated for 5 minutes. Three different samples were spotted on the MALDI target, which was cleaned prior to analysis with acetone (Sigma-Aldrich, Austria) and 2-propanol (Sigma-Aldrich, Austria). The first sample was 0.8 μ L matrix, the second sample contained 0.8 μ L matrix and the analyte CAO (1:100 diluted in the matrix) and the third sample was the matrix plus a standard to calibrate the spectrum (Ricinoleic acid). The sample spots on the MALDI target were prepared using the dried droplet technique, which means that the samples were transferred on the sample spots of a stainless steel target and the organic solvents evaporated. Four spots were prepared for each of the three different solutions.

4.3.2 MALDI-ToF measurements results and discussion

The spectrum of crenarchaeol was acquired using the same settings as for the other two glycerol diphytanyl ethers. The protonated molecular ion $[M+H]^+$ and the sodium adduct $[M+Na]^+$ was visible in the spectrum, but at low intensity, that CID was not possible. Most likely the CAO molecules in the ToF are present as neutrals, since hydroxy groups are not very efficient in the formation of positive charged ions. Competitive discrimination against crenarchaeol with the matrix is another explanation for the low

intensity observed with MALDI.

4.3.3 ToF-SIMS measurements results and discussion

Positive mode

The structure of CAO is shown in Figure 4.1. Since MALDI-ToF did not reveal any insight into fragmentation the ToF-SIMS spectra were only analyzed based on the peak for crenarchaeol at 1291 and some irdgendwas to conceive adducts and fragments - but their assignment was speculative.

The peak of the protonated molecule at m/z 1292 is visible even though for this position only very little counts are seen in contrast to the lower mass fragments. The Na-adduct at m/z 1314 is also present and exceeds the intact protonated molecule peak in intensity. The peak for m/z 273 is most likely $C_{20}H_{17}O^+$ and the fragment this ion might originate from has a molecular mass of 295 and is shown in blue in Figure 4.15 - it has to undergo excessive dehydrogenation to become the observed mass and should be considered as completely speculative. What is more reasonable is the fragmentation shown in green in Figure 4.15 with the molecular formula $C_{21}H_{39}O_3^+$ obtained by deprotonation of $[340-H]^+$. There are many other peaks in this spectrum, many of them were assigned with organic fragments made up of C, O, and H but none of them could be explained through fragmentation of crenarchaeol. The peaks at 360 ($C_{19}H_{40}N_2O_4^+$), 401 ($C_{18}H_{34}O_7K^+$), 413 ($C_{24}H_{30}PO_4^+$), 429 ($C_{22}H_{31}PO_4K^+$), 443 ($C_{23}H_{39}O_8^+$), 487 ($C_{24}H_{39}O_{10}^+$), 531 ($C_{26}H_{52}O_9Na^+$), 686 ($C_{25}H_{34}N_2^+$) and 741 ($C_{38}H_{61}N_2^+$) are all organic molecules which can be derived from various combinations of parts of crenarchaeol, but none of them can be explained through fragmentation of crenarchaeol. The peaks at m/z 685 and 281 are polysiloxan contaminations (PDMS) from the wafer.

Negative mode

It is possible to identify a peak at m/z 1290.4 which can be assigned to the deprotonated form of crenarchaeol. The peak at 1290 shows a very weak intensity though, even

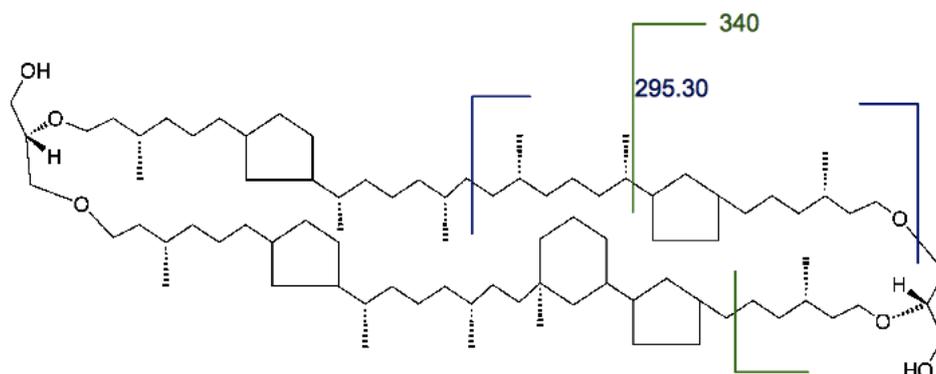


Figure 4.15: ToF-SIMS: proposed fragmentation of crenarchaeol in positive mode. Shown are the protonated molecules - m/z 295 and m/z 340.

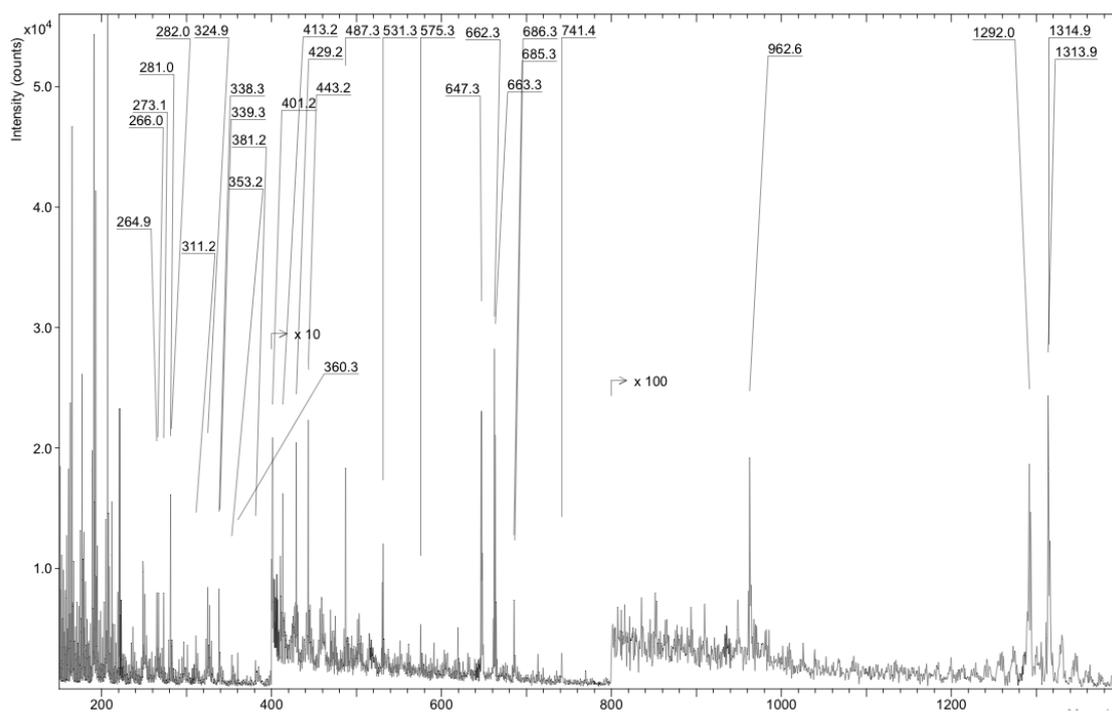


Figure 4.16: ToF-SIMS: spectrum for crenarchaeol in positive mode. Mass range from 150 to 1400 u. Information about the peaks is provided in the text.

No	Center Mass	Assignment	Area [a.u.]
1	273	[295-22H] ⁺	7276
2	339	[340-H] ⁺	1790
3	1292	[M+H] ⁺	198
4	1315	[M+Na] ⁺	230

Table 4.5: Fragment list for crenarchaeol obtained in positive mode.

weaker intensity than seen in the positive spectrum for the protonated ions ([M+H]⁺, [M+Na]⁺).

The peak at 947 was with reasonable mass deviation (10.404 ppm) assigned to the molecular formula C₆₅H₂₅Na₂O₆, a possible ion fragmentation is shown in Figure 4.17. As for peak at m/z 562 and m/z 590, both of them have very little mass deviation for their proposed molecular formula (ppm of 0.923 for C₃₂H₈₆NaO₆ and ppm of 1.477 for C₃₃H₇₈NaO₄) but both structures can not be explained solely from the fragmentation of crenarchaeol.

No	Center Mass	Assignment	Area [a.u.]
1	562	C ₃₃ H ₇₈ NaO ₄	4792
2	590	C ₃₂ H ₈₆ NaO ₆	1186
3	947	C ₆₅ H ₂₅ Na ₂ O ₆	829
4	1290	[M-H] ⁻	164

Table 4.6: ToF-SIMS: list of fragments for crenarchaeol obtained in negative mode

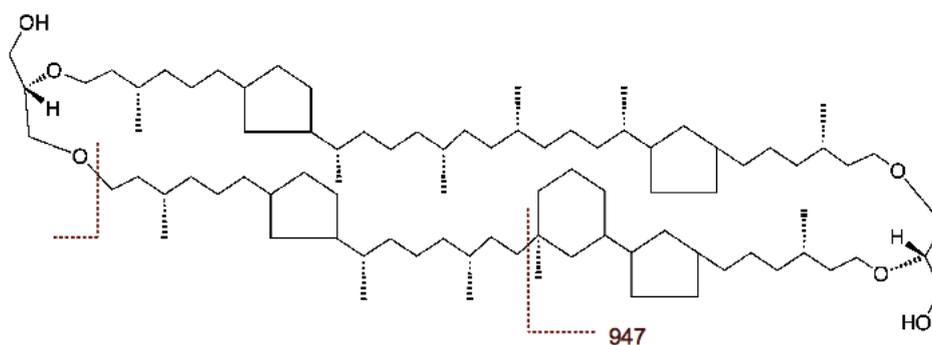


Figure 4.17: ToF-SIMS: proposed fragmentation of crenarchaeol in negative mode for the peak at m/z 947.

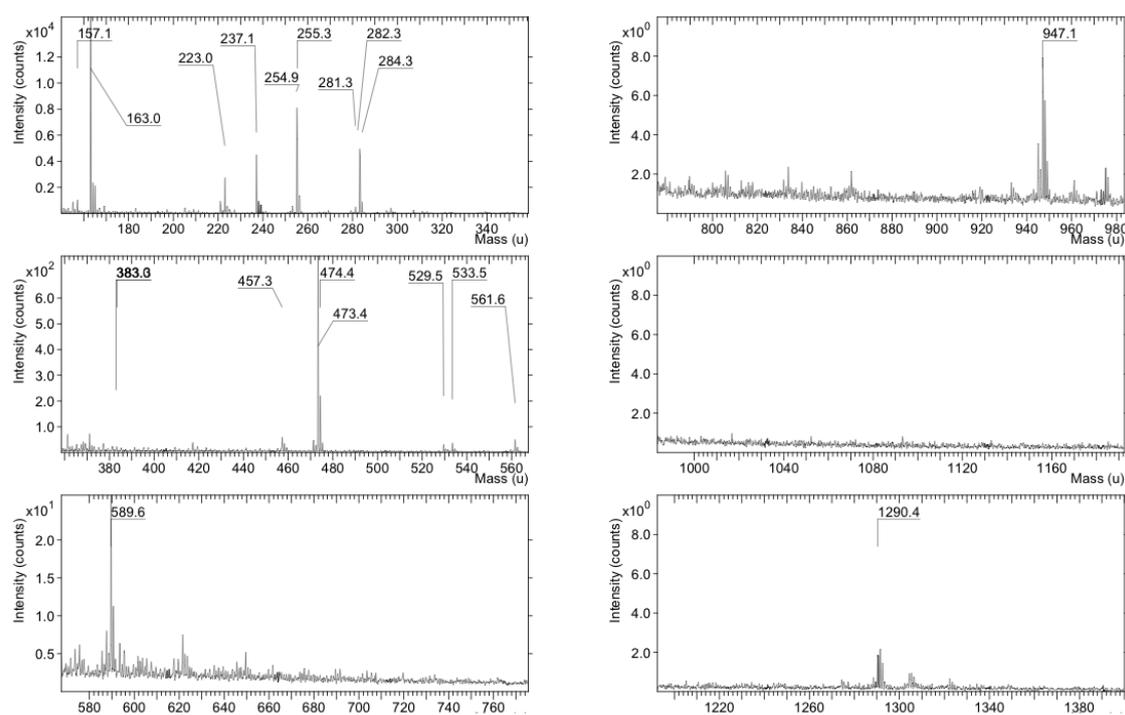


Figure 4.18: ToF-SIMS: spectrum for crenarchaeol in negative mode. Mass range from 150 to 1400 u. Information about the peaks is provided in the text

Chapter 5

Characterization of procaryotes

5.1 Sample preparation

All liquids and media were prepared using UV-light treated and filtered water obtained from a water purification system (MQ Biocel, Millipore, Billerica, USA). Media were furthermore sterilized before usage in a water vapor high-pressure autoclave (H+P, München, Germany) at 121 °C and pressure of $1 \cdot 10^5$ Pa for 20 min. All objects which came in contact with the cells and the wafer were furthermore treated with 97% EtOH p.a. (Merck, Darmstadt, Germany).

For this thesis two organisms were investigated : the bacterium *E. coli* and the archaeon *N. viennensis* [Tourna *et al.*, 2011] - the latter one was obtained as a pure culture from the lab of the Department of Genetics in Ecology, A-1090 Vienna, Austria.

E. coli grew on a medium containing 8g/L yeast extract (AppliChem, Darmstadt, Germany), 8 g/L bacto-peptone (AppliChem, Darmstadt, Germany) and 8 g/L NaCl (AppliChem, Darmstadt, Germany) in dH₂O for 48h at 37°C. Subsequently aliquots were taken (usually 4-8 times 1mL) and stored at -18 °C. *N. viennensis* grew on a very complex medium for ammonia oxidizing archaea in soil which ingredients are not given here but the complete composition can be obtained from the Department of Genetics in Ecology, A-1090 Vienna, Austria. Aliquots of *N. viennensis* were taken and stored prior to anal-

ysis at $-18\text{ }^{\circ}\text{C}$.

After this there were optional steps from which at least one and maximal three were performed: fixation in PFA/EtOH, staining with KI, sample washing with 4% NaCl (AppliChem, Austria), 8% PBS (10 mM of PBS stock solution (Stock solution containing 200 mM Na_2HPO_4 and 200 mM NaH_2PO_2) and 130 mM NaCl) or 8% NH_5CO_3 (AppliChem, Austria). Which of this steps was performed for an individual sample is given in the sample settings (Table 2.2 to Table 2.4).

Fixation

Fixation was done by Markus Schmid from the Department of Microbial Ecology, University of Vienna as explained here: 1 mL 37% Formaldehyde solution (Carl ROTH, Karlsruhe, Germany) was mixed with 3.3 mL 3xPBS and 5.7 mL dH_2O) were added to 1 mL of sample and incubated for 3 h at 4°C . Fixation was stopped by centrifugation of the sample (21250g for 15 min) and the supernatant was discarded and the pellet resuspended in 1:1 PBS and EtOH. Fixed cells were stored at $-20\text{ }^{\circ}\text{C}$.

Washing step

The washing steps for the three different solutions (8% NH_5CO_3 , 4% NaCl and 8% PBS) were the same. Only for the NH_5CO_3 the procedure after the washing step did slightly differ from the other solutions - drying was done for 12 hours at $80\text{ }^{\circ}\text{C}$ to induce thermal decomposition of the buffer into HN_3 , CO_2 , H_2O which ideally leads to a buffer free environment.

The first step was to spin the 1 mL aliquots at 7.000 rpm for 5 minutes, the supernatant was discarded and the droplet was diluted in 1 mL of the washing solution. This step was repeated three times. Some other washing techniques were also tested: Centrifugal Filter Devices, Dialysis (SnakeSkin Dialysis Tubing, 10K MWCO, Thermo Scientific, Rockford, IL, USA) and Slida-A-Lyzer Dialysis Cassettes (5K MWCO, Thermo Scientific, Rockford, IL, USA) - all used as described in their manual.

In practice the different washing techniques did not show any visible improvement under

the microscope and in the spectra of the ToF-SIMS and were discarded due to the fact that they all take longer than the washing step with the centrifuge.

It should be noted that even though the washing step with 4% NaCl is described here it does not show up in the actual experimental settings because NaCl formed salt crystals during drying and was therefore discarded as an alternative to PBS.

Staining

A 5% KI solution (Solution Lugoli) was prepared for the KI stain. This was done because *N.viennensis* is - like all archaea - very small and not easy to spot in a light microscope for the untrained eye. The stain with the Solution Lugoli increases the contrast of cells. For the stain 5 g/L Iodine and 10 g/L potassium iodide (both obtained from Thermo Scientific, Rockford, IL, USA) were dissolved in 10 mL dH₂O and rinsed three times on the mounted and dried cells on the silicon wafer. After each staining step the wafer was rinsed with dH₂O. KI staining showed no visible improvement of the contrast of the archaea in the light microscope so in later analysis settings it was as possible contamination source discarded.

5.2 TOF-SIMS measurements results and discussion

Sample labeling

The sample and series naming followed a few simple rules: the first 4 numbers are the month and the day of the measurement (mm/DD), the optional letter n and p before the first four numbers (e.g. p0103m*, n0103m*) limits the series to either only the positive or only the negative measurements of the day. What follows the first four numbers is of the following format (D stands for digit): mDDD (e.g. p0103m001). This follows a simple systematic: surface analysis has the format m00D, short co-sputtering with oxygen or cesium uses m10D and long co-sputtering uses m11D (D increases starting from 1). The asterisk is used if a series is addressed which was taken on a specific day. The letter m in the sample stands for measurement.

Sputtering modes

All measurements were done in the high-current bunched (HCBU) operation mode of the Bi-gun for high molecular mass resolution.

There were three fundamentally different analysis modes as indicated by the label of the samples.

- (1) Surface spectroscopy (00D) done with Bi_3^+ clusters
- (2) Prior to the surface measurement a short sputtering cycle - 'pre-sputtering' - from the oxygen or the cesium source (operated between 0.2 and 2 keV kinetic energy) was performed. Afterwards analysis with the bismuth cluster for data acquisition - the samples are named (10D)
- (3) Continuous co-sputtering with O_2^+ or Cs^+ ions of low kinetic energy (usually between 0.2 and 2 keV) and measurement with the Bi-cluster gun (11D).

Short co-sputtering was mainly done to remove contaminants from the surface of the sample, long co-sputtering was done to penetrate the outer surface layer of the gram-negative *E. coli* and the lipopolysaccharide layer of *N. viennensis*. Furthermore all samples were carefully characterized with a reflected light microscope (Olympus Bx60) for artifacts that could be confused with cells in the ToF-SIMS camera. Concerning ion imaging it should be noted that a mass binning of 4 pixels was performed in order to improve the signal to noise ratio.

Analytical question

When the analysis first turned to prokaryotes and cells two questions arose:

- (1) Is it possible to identify biomass?
 - (2) Is it possible to use crenarchaeol to locate and distinguish between against different cells?
-

After the first few measurements and finally with the analysis series 0416m* and 0525m* it became evident that the performed experiments can not yield satisfying answers on these questions. It was possible to identify biomass, but this was mostly succeeded using low molecular mass ions. No high mass molecular ions characteristic for lipids, sugars or proteins could be located at intensities high enough for lateral imaging or depth-profiling. Therefore the data series p0208m* (only positive) was used and tested with respect to the following questions:

- (1) Is it possible to see common features (correlations) in the spectrum of *E. coli* and *N. viennensis*? What does correlate? What doesn't?
- (2) Is it possible to distinguish between *N. viennensis* and *E. coli* based on their spectra with principal component analysis (PCA)? What are the loadings? What changes during depth profile?

These questions are addressed in the chemometrics section. As it was not the aim of this thesis to analyze the ToF-SIMS spectra of prokaryotes besides the binary decision of crenarchaeol yes/no, only a few examples of the acquired data are fully discussed. This and a short summary of every data set is given in the sections below.

It should be noted that molecular formulas are automatically assigned to the masses obtained in the spectrum, using isotope abundances. This is in most cases very effective for low molecular masses but the higher the molecular weight gets the more possible structures can be assigned to a mass, even if the mass is confined to a window of ± 15 mDa ($\pm 0.015u$). For that reason, the molecular mass instead of the molecular formula is given when the molecular formula is not plausible.

5.2.1 Data set 0103m*

The organism in this analysis was *E. coli*.

scale measurement settings for the data set is given in Table 2.2. The mass It should be noted that for all spectra ranged from 0.1 u to 900 u obtained by operating the Bi gun at a cycle time of 100 μ s. These measurements were done to identify biomass - the detection of high-mass molecular ions was not the primary objective. The negative spectra were calibrated with C_2^- , C_3^- , C_4^- , PO_2^- , PO_3^- and I^- . The positive spectra were calibrated using CH_3^+ , $C_2H_5^+$, Na^+ and K^+ . What should be noted here is that this analysis was performed using a 50 keV Bi_3^{2+} cluster instead of 25 keV Bi_3^+ used in all of the following measurements. Two examples will be fully discussed (p0103m002 and n0103m002):

p0103m002

The sample consists of fixated *E. coli* cells. The area of interest was marked prior to analysis with a laser from a laser micro-dissection system (LEICA LMD 7000). The image obtained from the transmitted-light microscope (since Si wafers are not transparent, sample areas had to be located at the fringe of the wafer) is shown in Figure 5.1 along with the light microscope image and the ToF-SIMS CCD camera image.

A peak list was automatically generated using the Auto Peak Search Mode (Single Range) of the SurfaceLAB6 software applying the following parameters: Start Mass (u): 0, End Mass (u): 900 and Intensity Threshold (cts): 150. After the automatic peak search a manual search was performed to supervise the indicated peaks and identify relevant peaks not found by the auto peak search.

On the following pages the ion images (500x500 μ m) of all measured ions above the threshold are displayed (p. 70-73). These images display the lateral intensity distributions of the respective secondary ion species, the relative intensities are displayed on a color scale. A four pixel binning was applied to improve the signal to noise ration (for every pixel a spectrum was obtained, the spectra of four pixels were accumulated to one spectrum and the four pixels merged into one) The color code was as following: starting from black to blue, orange, red, purple and white the intensity of the secondary ions

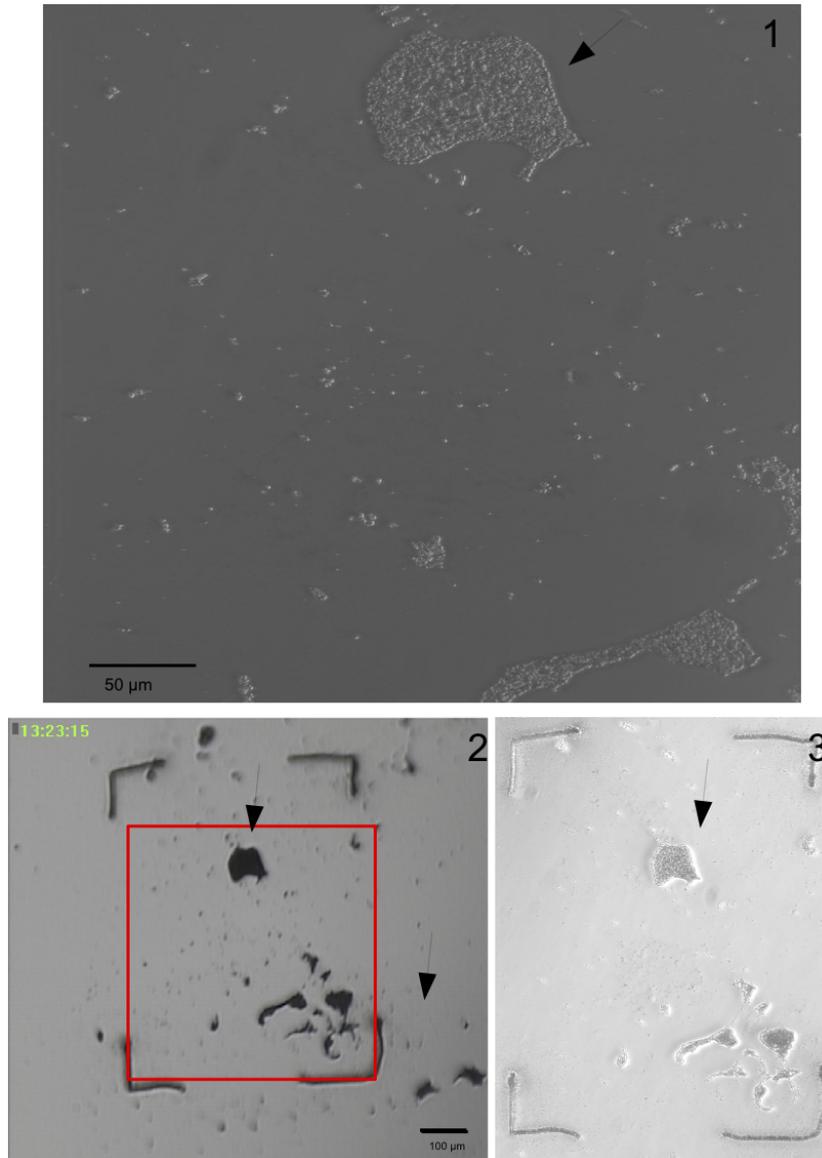


Figure 5.1: p0103m002 - Light Microscope (20x), LMD Microscope (40x) and ToF-SIMS CCD camera image of *E. coli*. The measurement area is indicated by the red rectangle and has an area of $500 \times 500 \mu\text{m}$. The light microscope image (Nr.3) can be seen on the right and the ToF-SIMS camera image on the left (Nr.2), above is the LMD microscope image (Nr. 1). The black arrow indicates the same sample spot in all three images, *E. coli* cells can be seen in the LMD microscope image (40x - Nr.1). The marked edges (seen in image Nr. 1 and Nr. 2) were obtained by sample irradiation with the laser beam of the LMD instrument.

increases.

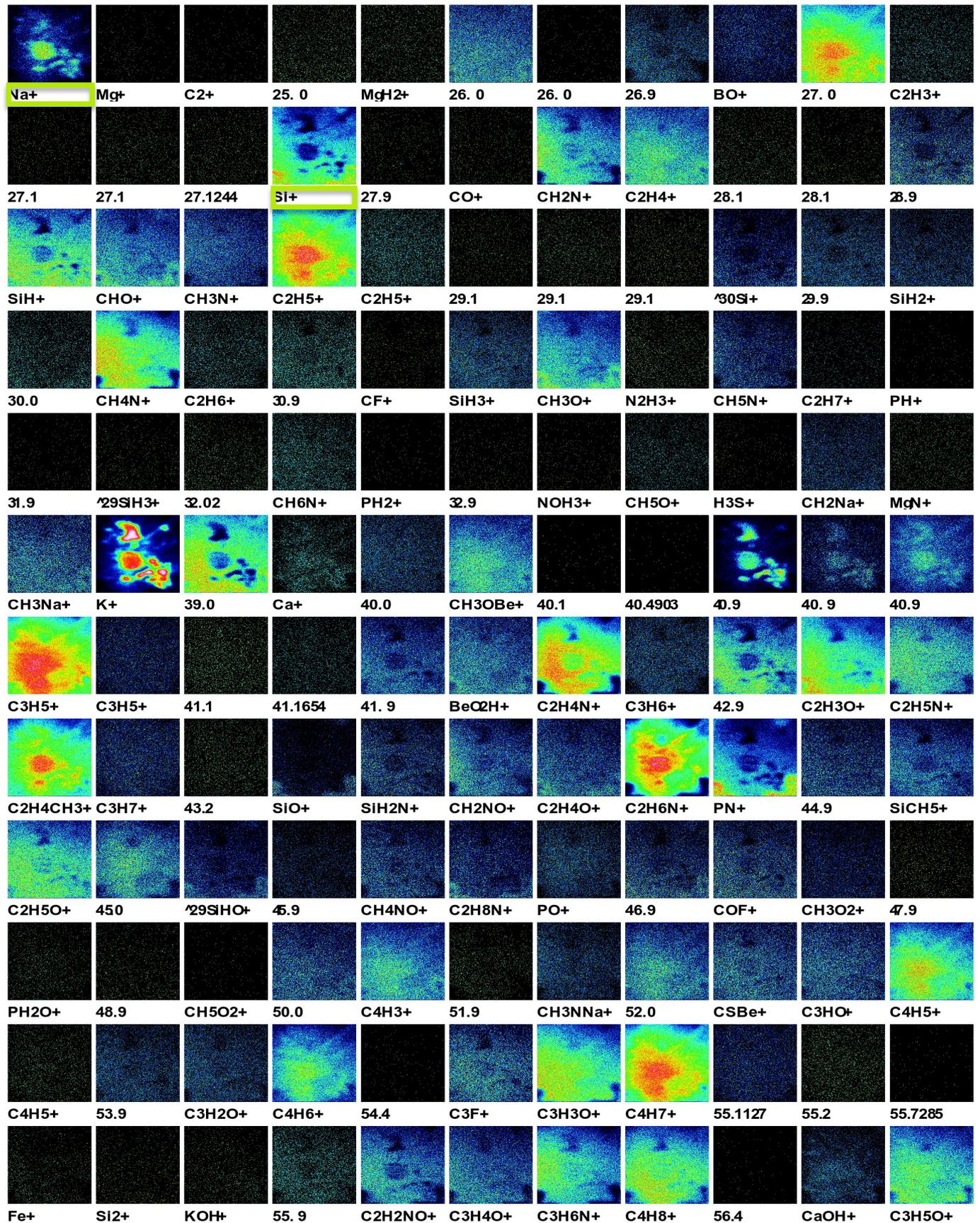
Comparison of the secondary ion intensity distribution images with the optical images from the CCD camera, the light microscope and the LMD microscope suggests that:

(1) All important masses for this analysis are framed in light green. Looking at the Na^+ and K^+ intensity distributions in contrast to Si^+ reveals where most of the biomass is located. This is in accordance with the cells based on the images from the optical microscopes with the exception of the round shaped feature in the center which is visible in many ion images but not in the optical images. The distribution of K_2PO_3^+ and, with less intensity, K_2PO_2^+ might show the actual position of the cells since the mass in the center is not shown here and should be regarded as organic contamination.

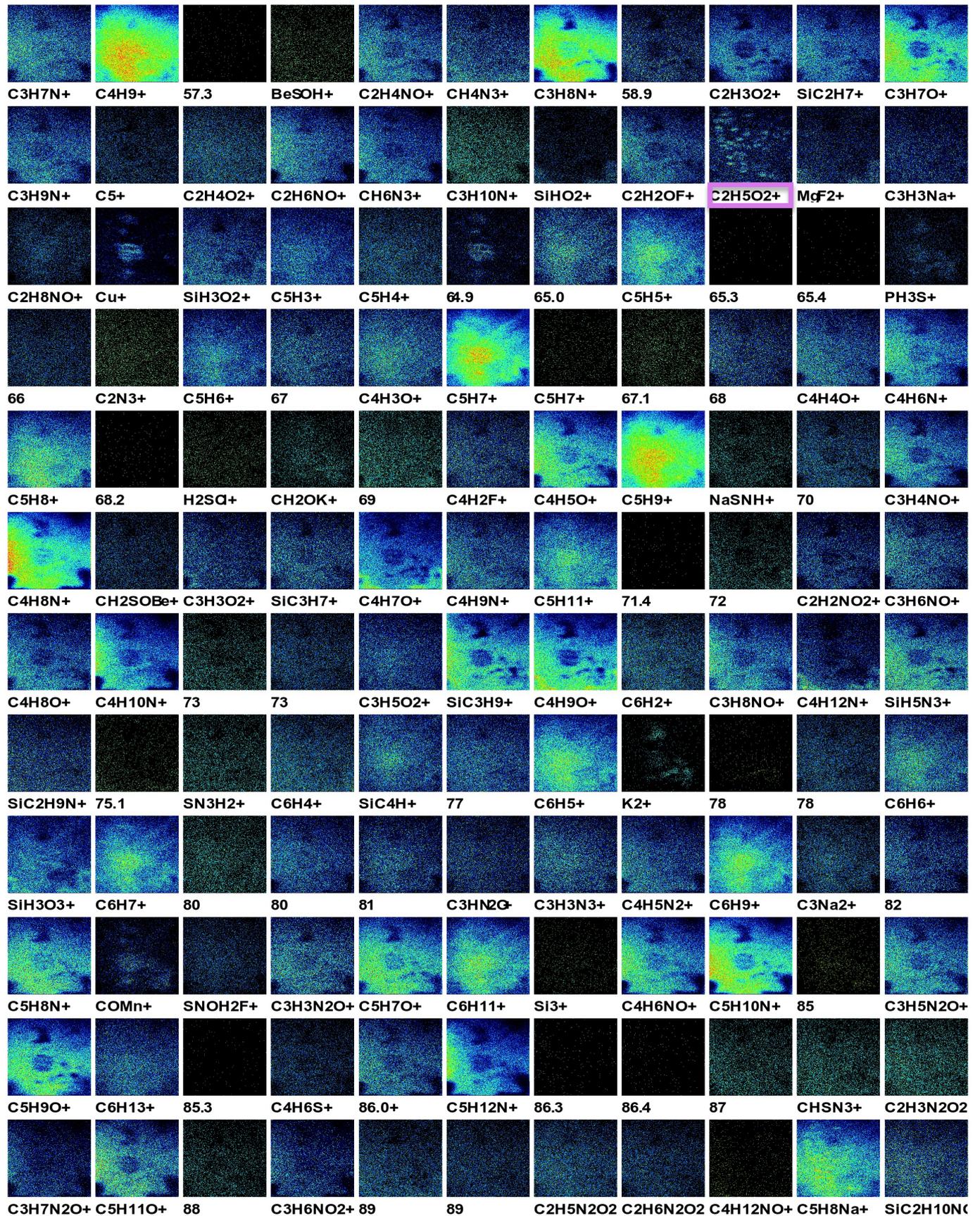
(2) The lateral intensity distribution of long alkane molecules and in general high molecular mass ions (a selection framed in red) are surprisingly not in accordance with the distribution of cells shown in the microscope and consequently render useless in this analysis to indicate cells. This finding is not unique to this analysis. The intensity of ions, even high molecular weight ions decreases when the ion beam hits the surface of biomass and increases when hitting spots without cells. For small organic molecules like C_2H_4 that pattern is not observed but the lateral distribution does not resemble any topography seen on the wafer in the microscope or the ToF-SIMS camera.

(3) The pattern seen in the ion image of e.g. C_4H_{11} and $\text{C}_3\text{H}_9\text{N}$ (a selection is framed in magenta) is very interesting. All molecules indicating this pattern are medium molecular mass organic molecules. The pattern does not correspond to single cells as they are much smaller.

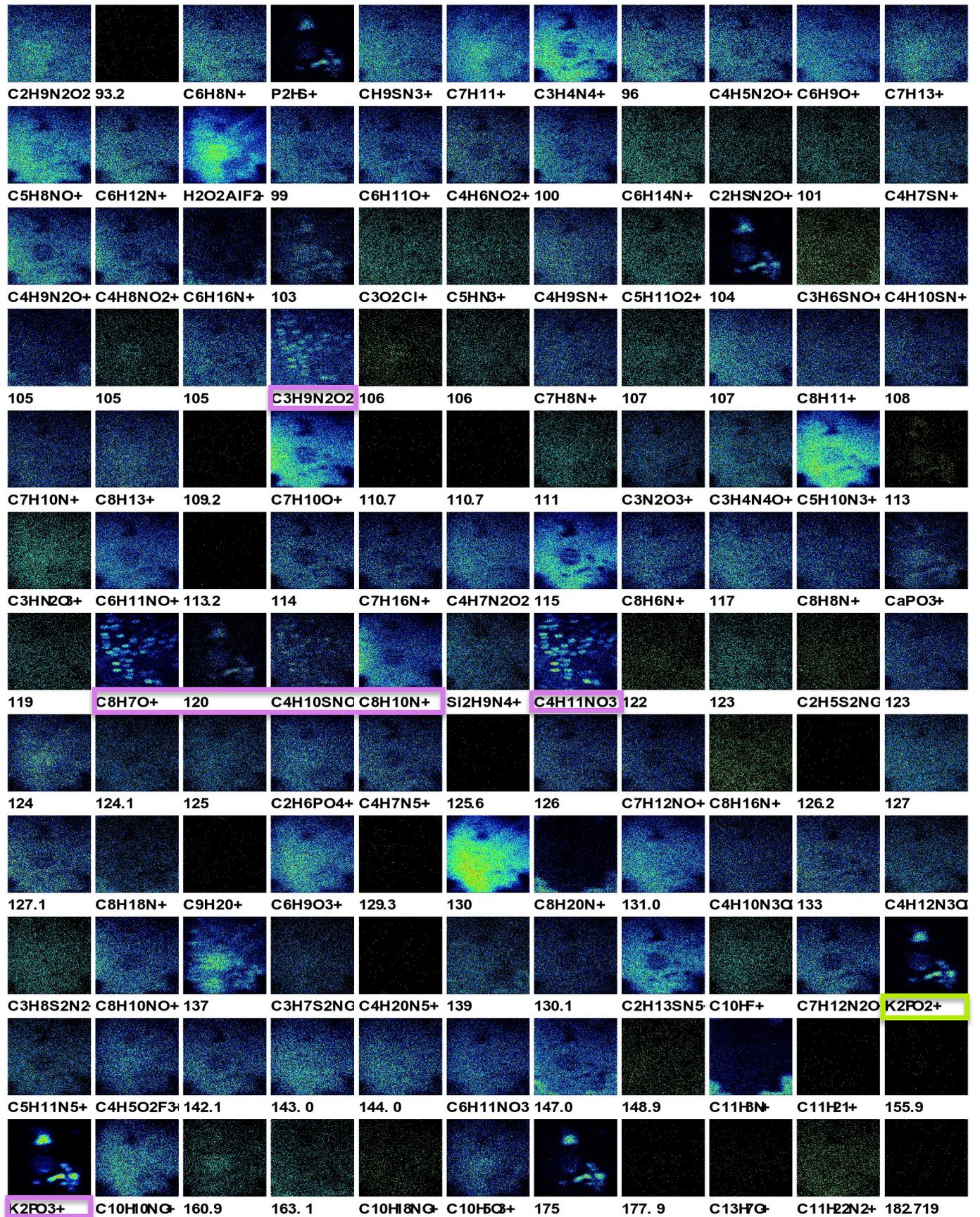
p0103n002 Spectra 1/4



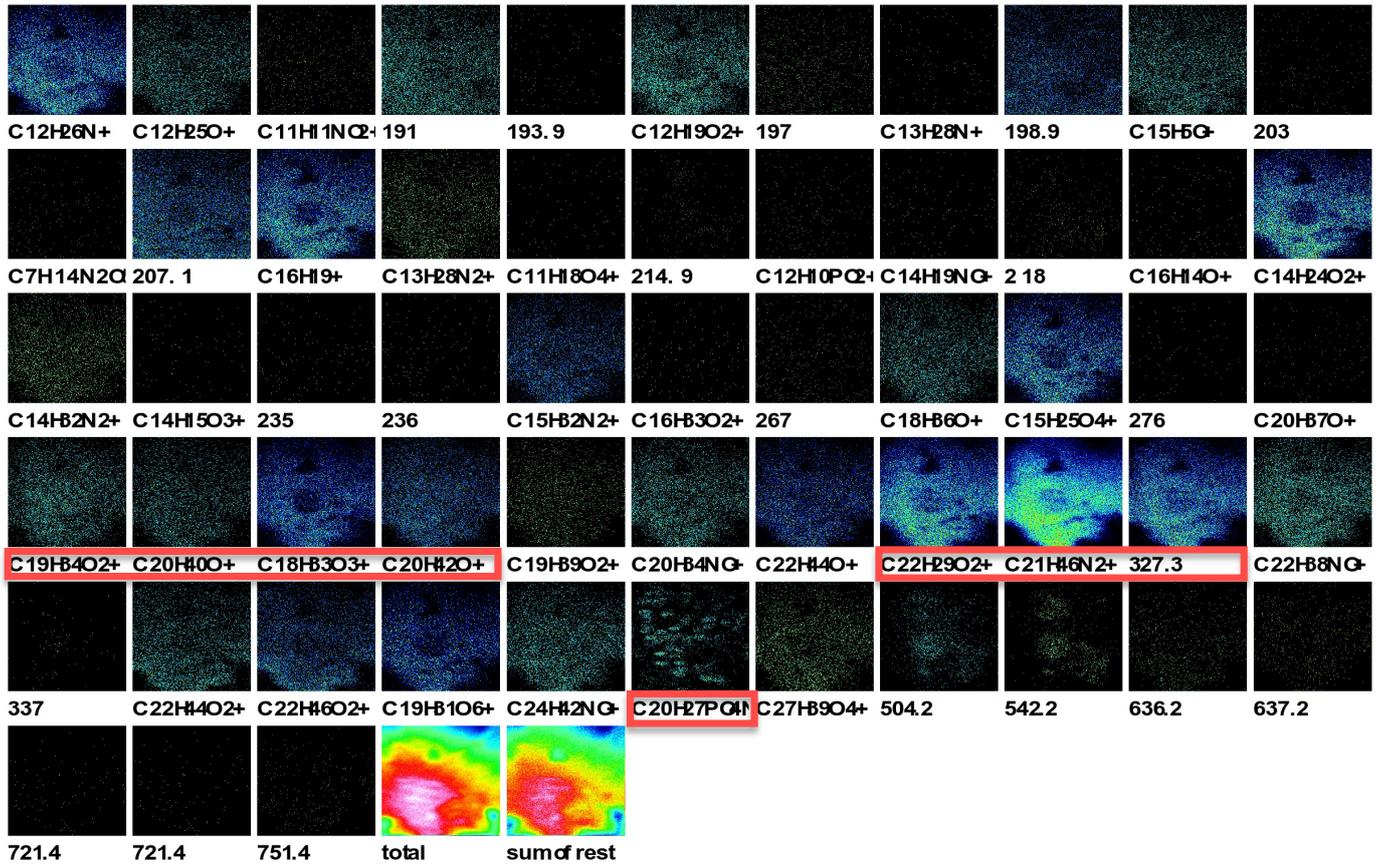
p0103n02 Spectra 2/4



p0103n002 Spectra 3/4



p0103n002 Spectra 4/4



n0103m002

The very same sample spot was used to analyze the spectrum in negative mode (the static SIMS limit was not exceeded - both measurements had an total ion dose density in the range of 10^{11} primary ions per cm^2). A peak list was automatically generated using the Auto Peak Search Mode (Single Range) of the SurfaceLAB6 software with the following constraints: Start Mass (u): 0, End Mass (u): 900 and Intensity Threshold (cts): 150. Subsequent to the automatic peak search a manual search was performed to supervise the indicated peaks and identify relevant peaks not found by the auto peak search.

The intensity distribution of all secondary ions that fulfill the requirements of the peak search are displayed on p. 79 to 83.

A closer look at the ion images of the ion distributions reveal some interesting features: In principal in negative ion mode there are some particular secondary ion species that are indicative of organic compounds : CN^- (which shows highest intensity at locations without biomass in the microscope image), CNO^- , $\text{C}_2\text{H}_3\text{O}^-$, CHO_2^- , PO_2^- and PH_2O_4^- and an ion at molecular mass 71 and 79 u which all have enough intensity to be used as marker for the E. coli cells. High molecular ions (starting at about 200 u) have low intensity (e.g. $\text{C}_{16}\text{H}_{31}\text{O}_2^-$, $\text{C}_{36}\text{H}_{53}\text{O}_4^-$) but still show similar patterns than e.g. the ions at m/z 71 and 79. In comparison to the images of the optical microscope some secondary ions reveal the same pattern like the secondary ions in positive mode and show the round shape in the middle of the wafer which is not in accordance with the obtained optical microscope image (e.g. in positive mode m/z 40.9 and in negative mode m/z 78.94). Since this pattern is not in accordance with the image obtained from the optical microscope one explanation is that the sample was contaminated and that it is not possible to draw a clear boarder between biomass and non-biomass.

This very inconclusive finding - that it was possible to detect secondary ions associated with organic compounds but also that these ions give rise to patterns which are not seen in the optical microscope led to a variety of hypothesis:

- (1) the cells undergo lysis when dehydrated - that is the reason why the wafer surface is contaminated with cell contents
- (2) the washing steps are not sufficient to clean the cells from medium and/or buffer
- (3) contaminants cover the cells and wafer surface, thus higher ion dose (or co-sputtering) must be applied to remove contaminants.

The question if cells undergo lysis when they are dried was tackled with electron microscopy and atomic force microscopy (examples seen in Figure 5.2). The morphology of the cell wall was preserved after dehydration which suggests that at least the cell wall must have been intact to be measured.

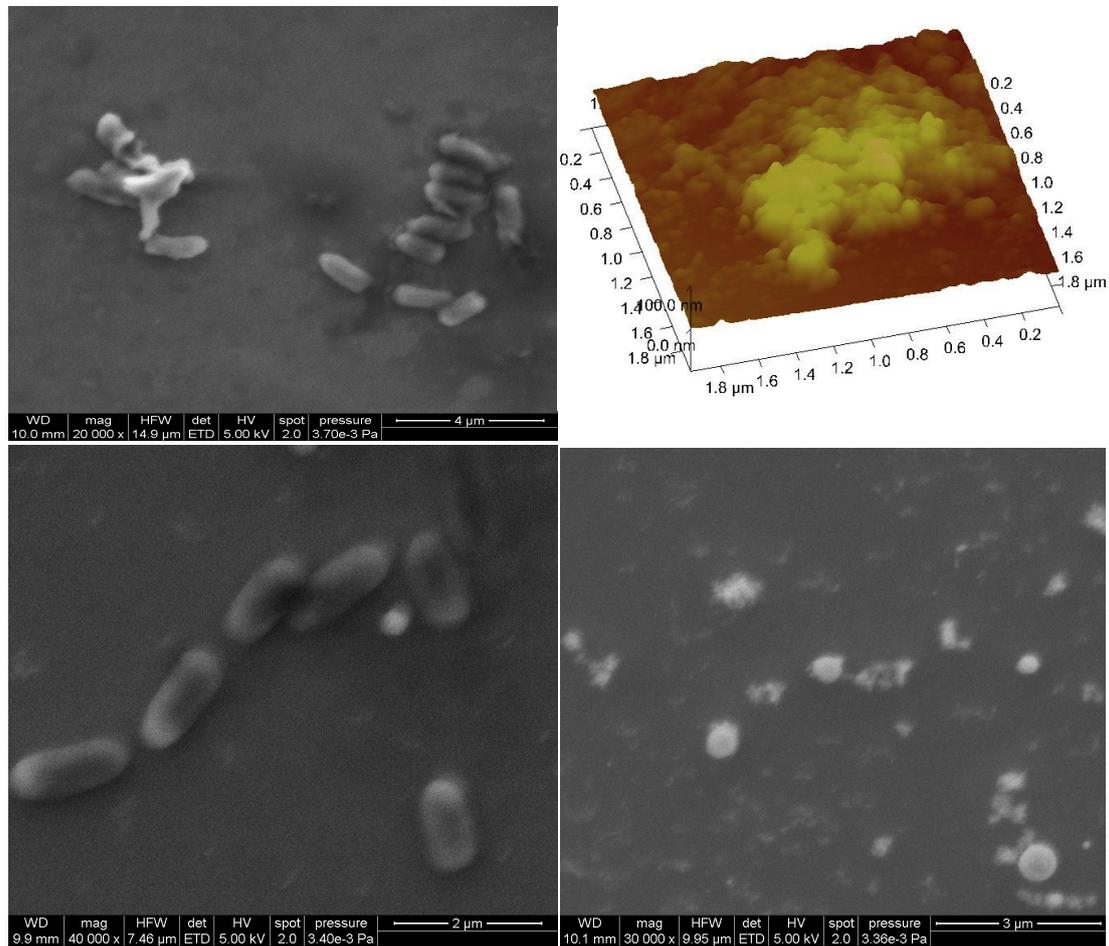
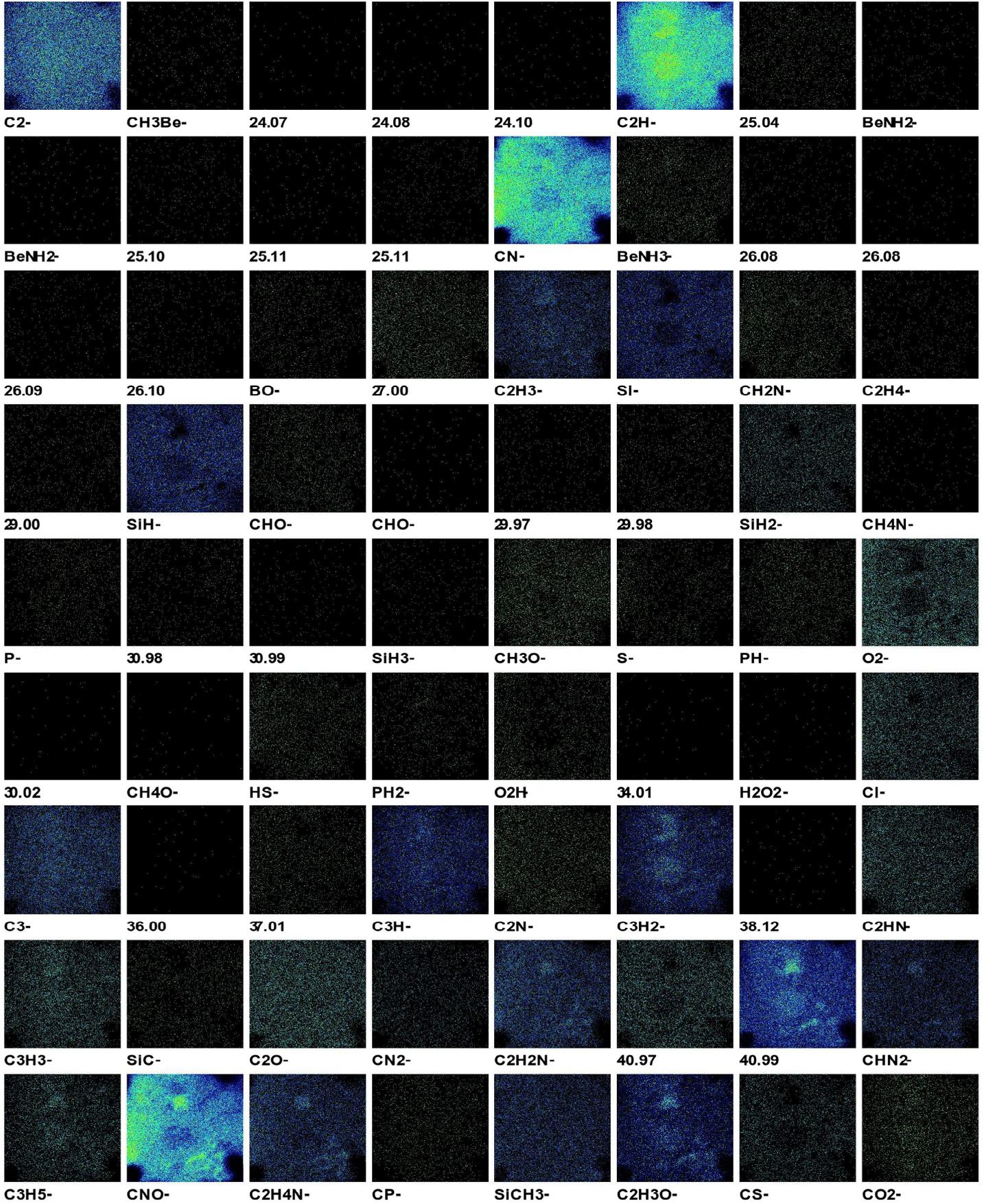
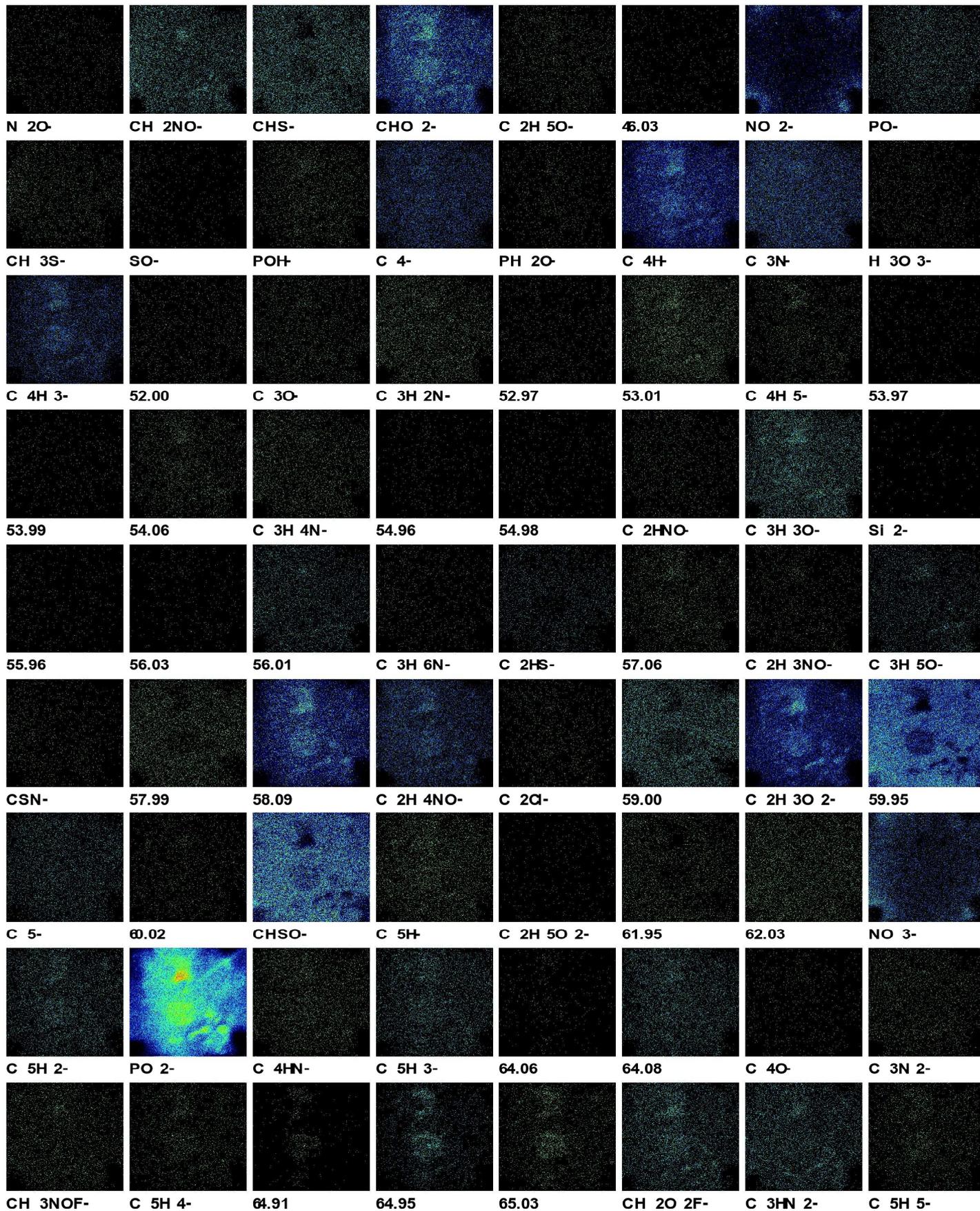
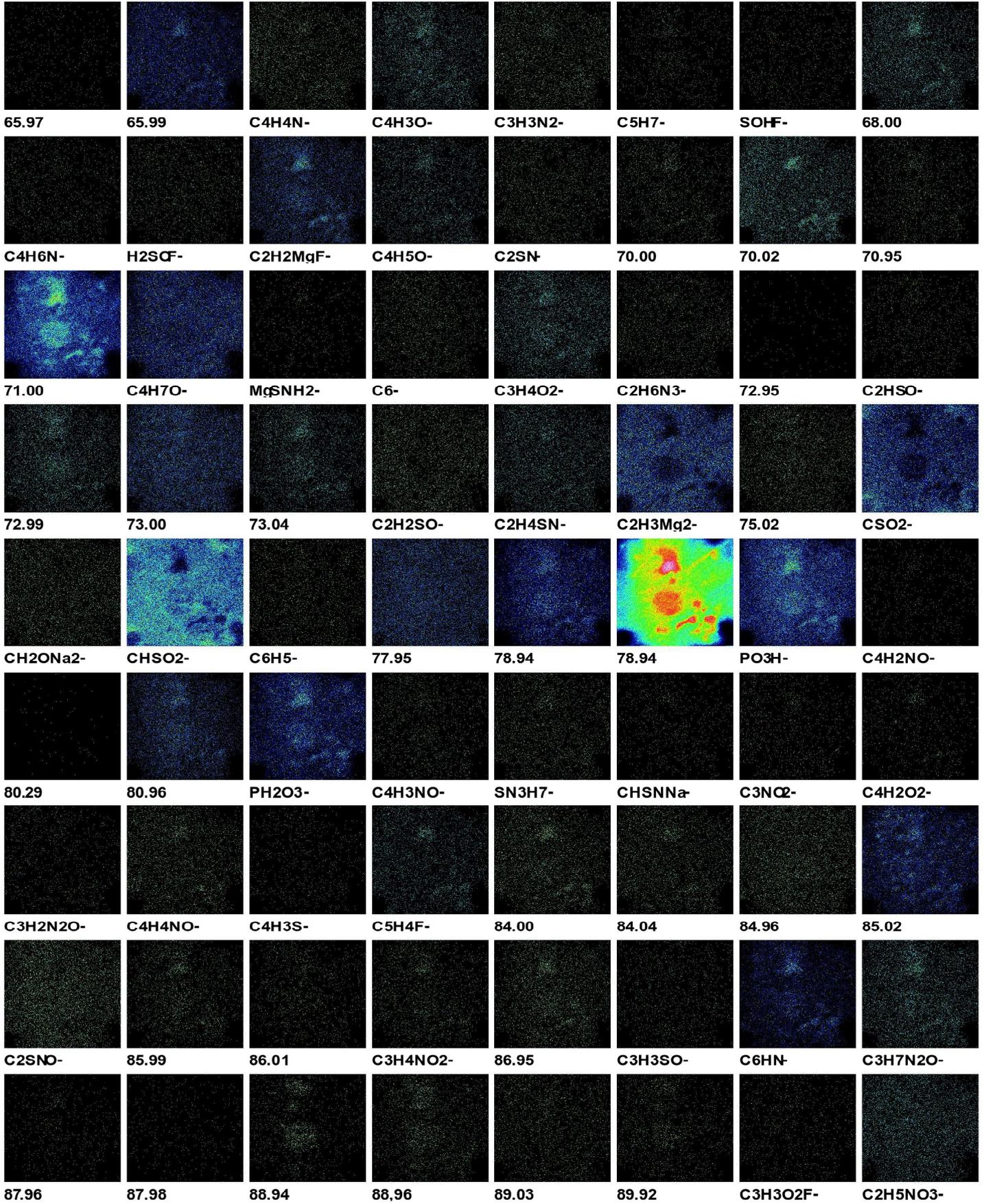
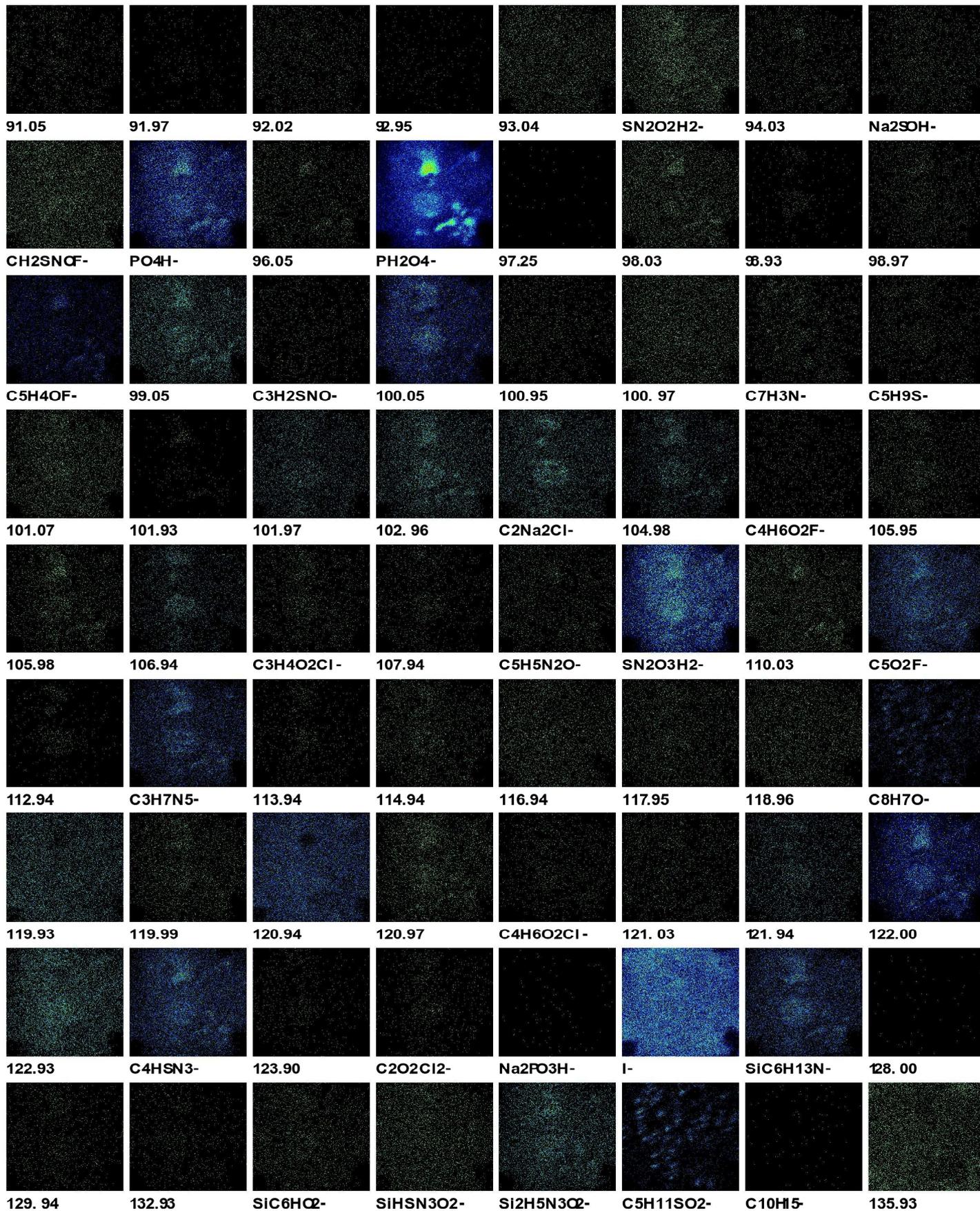


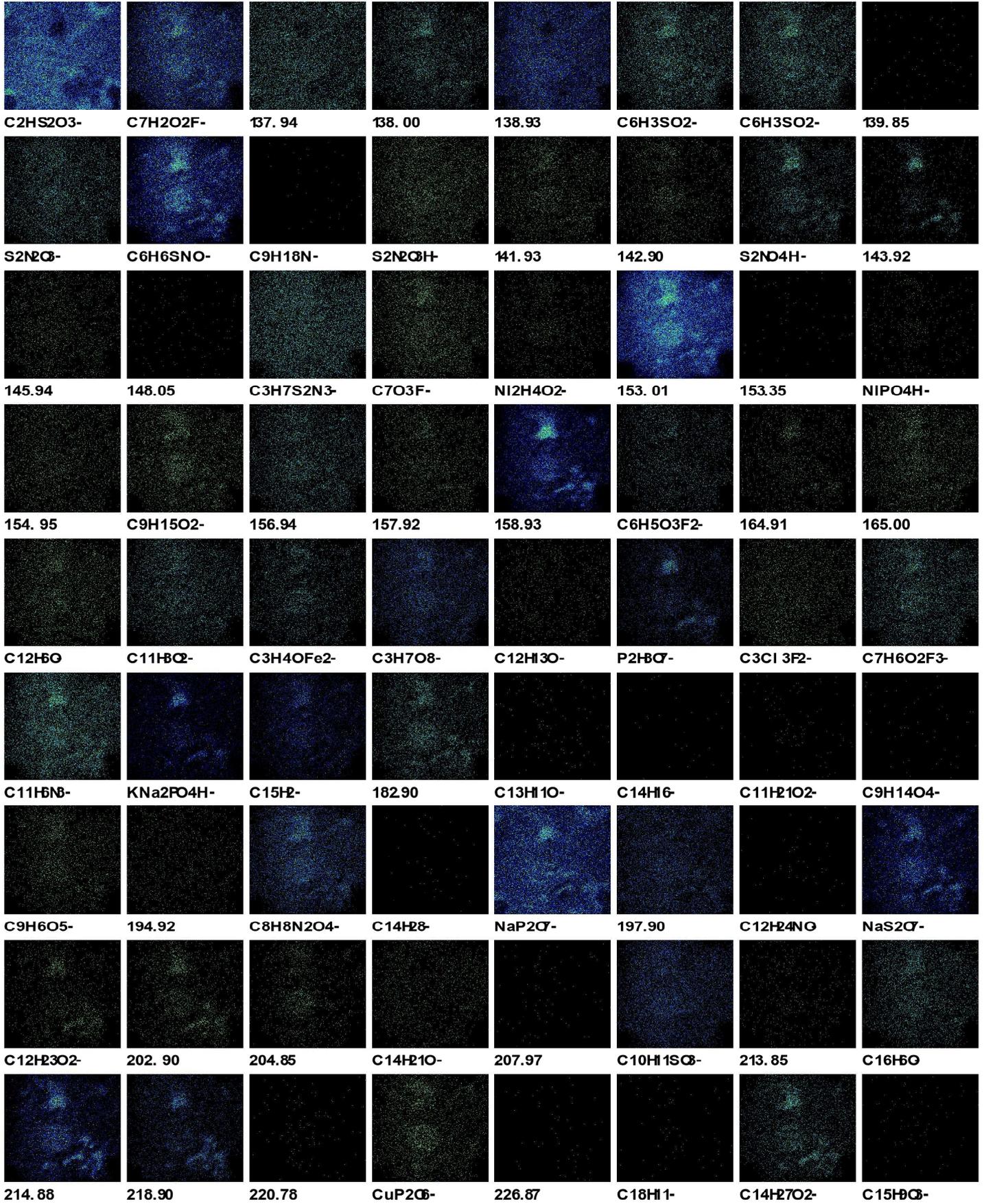
Figure 5.2: Cell Imaging of *E. coli* and *N. viennensis* with AFM and SEM. The images on the left are SEM images of *E. coli*. At top *E. coli* is KI stained and fixated and on bottom *E. coli* is only fixated - both have the characteristic rod shape which is associated with *E. coli*. On the right side *N. viennensis* cells are shown (fixated and with KI stain). On bottom a REM picture of single *N. viennensis* cells is shown with the typical spherical like appearance. On top a AFM picture obtained from a cluster of biomass on a wafer with *N. viennensis*. Since *N. viennensis* is very small and not very well seen in the ToF-SIMS camera it was critical to make sure that the structures regarded as *N. viennensis* cells are actually not residues from the buffer of other contamination forming crystal like structures. It was not possible to confirm this using the AFM - the round structures in this image are far too small to be of prokaryotic origin.



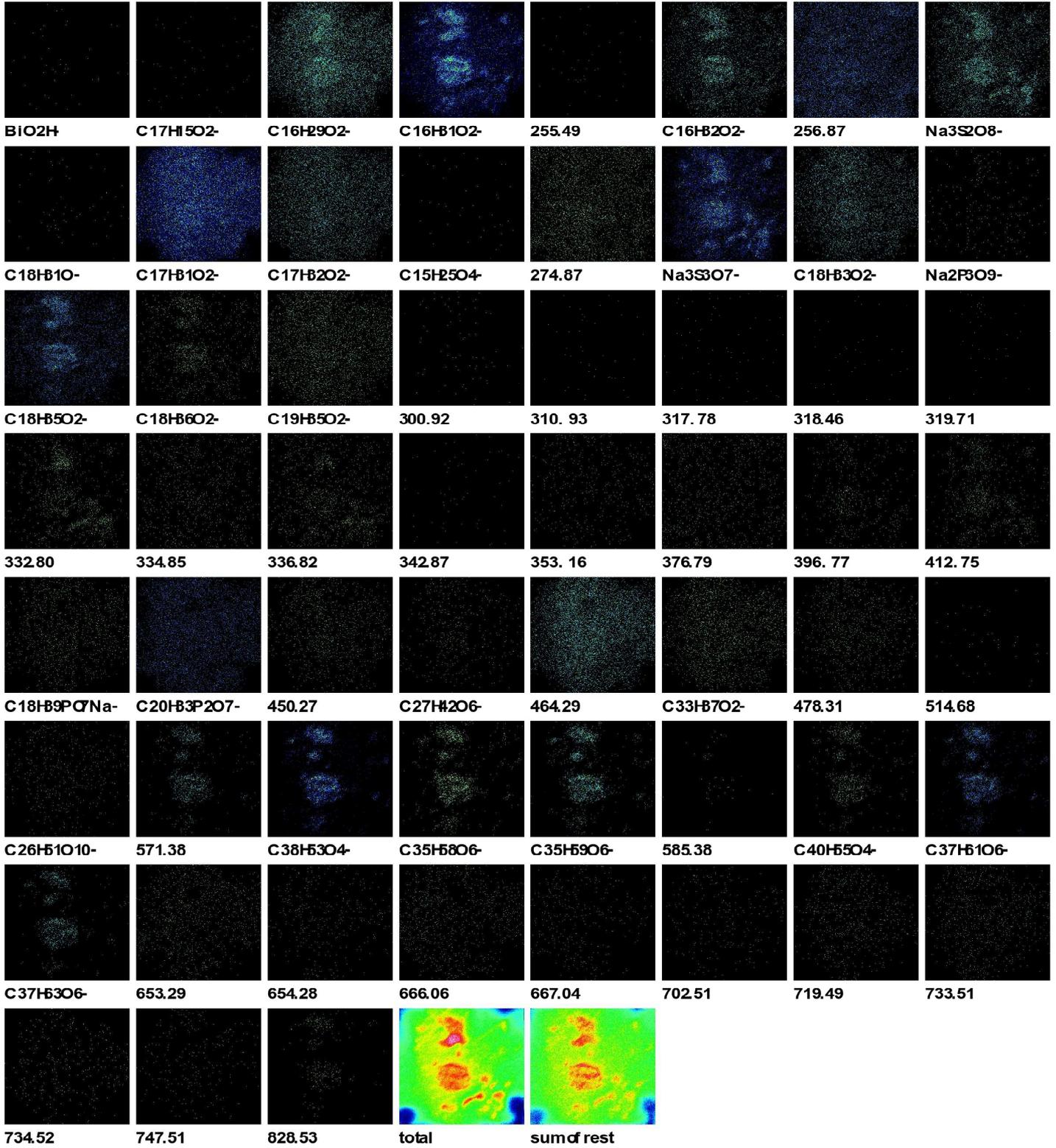








n0103n002 Spectrum 6/6



5.2.2 Data set 0208m*

Since crenarchaeol was not seen in the spectra before and it was also not possible to distinguish biomass from contamination (although possible with e.g. K_2PO_2^2 and K_2PO_2^3 without the contamination - but this ions are the exception) another analysis was done.

The analysis settings were changed:

- the bismuth cluster changed from Bi_3^{2+} to Bi_3^+
- the number of scans and the primary ion beam density was increased
- short sputtering cycles were done at the beginning of the analysis (5 seconds) with an ion dose density of $8.8 \cdot 10^{16} \text{ions} \cdot \text{cm}^{-2}$ for Cs^+ and $1.2 \cdot 10^{16} \text{ions} \cdot \text{cm}^{-2}$ for O_2^{2+}
- co-sputtering analyzes to obtain depth profiles were conducted

A peak list was automatically generated using the Auto Peak Search Mode (Single Range) of the SurfaceLAB6 software with the following constraints: Start Mass (u): 0, End Mass (u): 1700 and Intensity Threshold (cts): 150. Subsequent to the automatic peak search a manual search was performed to supervise the indicated peaks and identify relevant peaks not found by the auto peak search.

The intensity distribution of all secondary ions that fulfill the requirements of the peak search are displayed on p. 79 to 83.

The settings for the analysis can be seen in Table 2.3 and Table 2.4. Two measurements are discussed in more detail: p0208m102 and n0208m102.

p0208m102

The sample consists of E. coli cells. Mass calibration was performed utilizing C_2H_5^+ , C_3H_7^+ , C_4H_9^+ and C_5H_9^+ . A short sputtering cycle (5 s - non-interlaced sputtering) was performed prior to analysis with the bismuth gun using the oxygen O_2^+ gun (0.2 kV) on a $500 \times 500 \mu\text{m}$ area centered in the area of the subsequent bismuth bombardment. Secondary ion species are selected due to their ability to reflect the structures seen in the ToF-SIMS CCD camera image. A selection of the masses best suited for this is given in Figure 5.3 and the spectrum with the labeled masses is displayed in Figure 5.6 and

the peaks with the center masses and corrected peak areas are shown in Table 5.1. The conclusion based on this spectrum and ion images is the same as in all spectra made from prokaryotes in this study: ions which seem to resemble the structure of the cells are inorganic (K^+ , Na^+) and phosphonate or phosphodioxid cluster ions. Small organic secondary ion species seem to preferentially emerge from areas containing biomass, but also at lower intensities at the surrounding areas. The larger the organic molecule ions, the more they appear as not emerging from areas containing cells but the surrounding areas - which are indicated by the high intensities of silica containing secondary ions. Biomass was characterized as the negative image of the silica distribution and further indicated by the distribution of the phosphor molecules and small organic molecules - for example the $C_2H_6N^+$ distribution. This image was then compared with the ToF-SIMS CCD image and a sample investigation with an optical microscope was also conducted - but it was not possible to find the sample area.

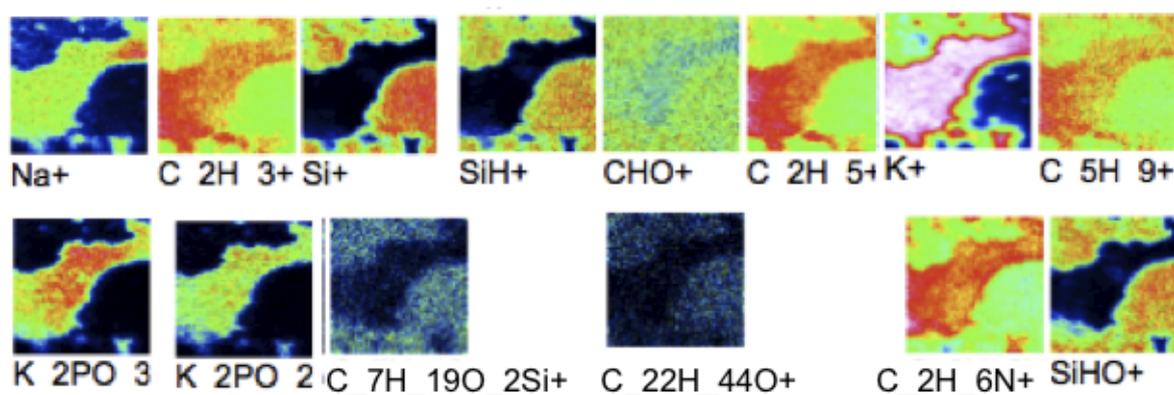


Figure 5.3: Visualization of p0208m102 - a selection of the most characteristic ions for biomass

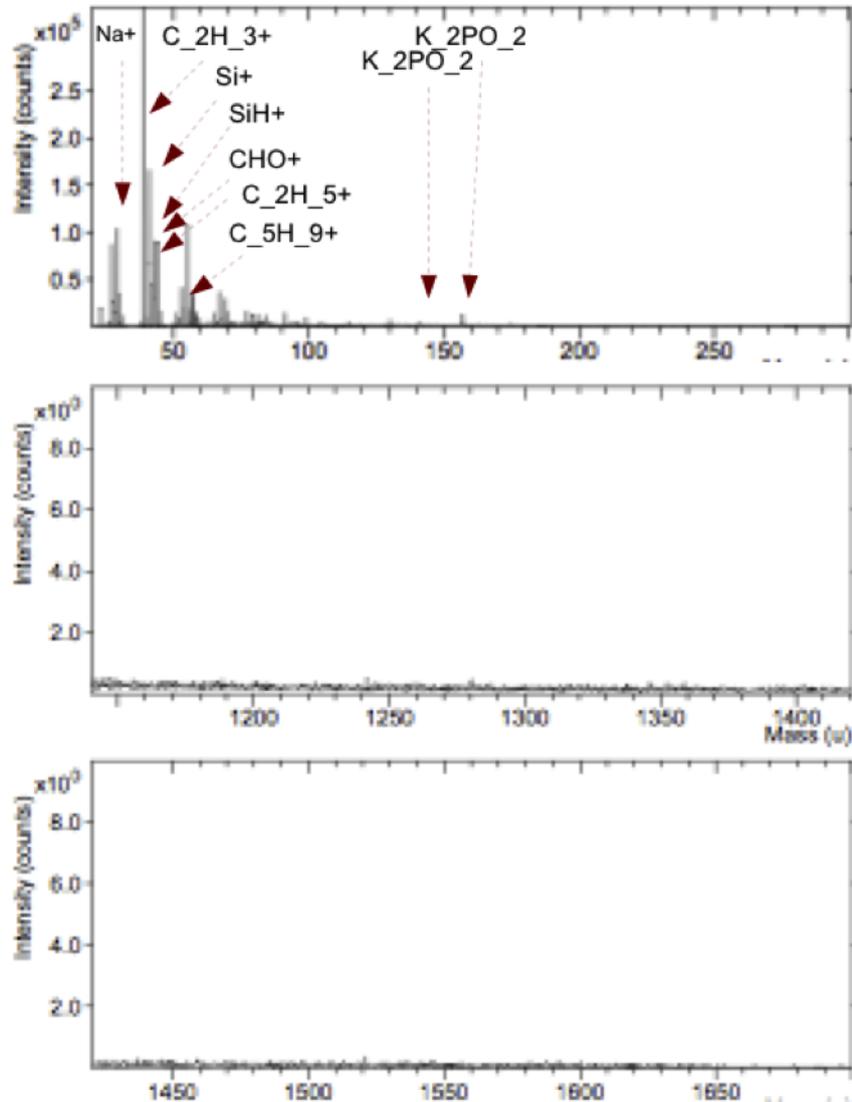


Figure 5.4: Spectrum of p0208m102 - the most characteristic peaks are labeled. Not shown are $C_{22}H_{44}O^+$ and $C_7H_{19}O_2Si^+$, $C_2H_6N^+$ and $SiHO^+$ because they were either in different mass ranges or their intensity was too small to be displayed without a multiplication factor. Also a section of the spectrum between 1100 to 1700 u is shown. There are no usable peaks with enough intensity.

No	Center Mass	Assignment	Area [a.u.]
1	23	Na ⁺	423,767
2	27	C ₂ H ₃ ⁺	1,863,687
3	28	Si ⁺	511,246
4	28.9	SiH ⁺	279,981
5	29.0	CHO ⁺	162,021
6	29,04	C ₂ H ₅ ⁺	10,192,577
7	39	K ⁺	164
8	44	C ₂ H ₆ N ⁺	2,301,241
9	45	SiHO ⁺	371,339
10	69	C ₅ H ₉ ⁺	959,003
11	141	K ₂ PO ₂ ⁺	141,840
12	157	K ₂ PO ₃ ⁺	444,586
13	163	C ₇ H ₁₉ O ₂ Si ⁺	30,373
14	324	C ₂₂ H ₄₄ ⁺	10,206

Table 5.1: p0208m102 - Peak list used to visualize biomass

n0208m102

Mass calibration was performed utilizing C^- , C_2^- , C_3^- , C_4^- and C_5^- . A short sputtering cycle (5 s - non-interlaced sputtering) was done prior to analysis with the bismuth gun using low kinetic energy Cs^+ ions (0.25 keV) on a 500x500 area overlaying the area of the later bismuth bombardment. Peaks and masses for imaging were selected due to their ability to reflect the structure seen in the ToF-SIMS CCD camera image. The spectrum in the high mass range is shown in Figure 5.6 as an example of the most striking problem of very low signal intensities in the high molecular mass regions. The selected ions and their lateral intensity distribution are shown in Figure 5.5.

In this analysis it is not clear if the seen structure represents bacterial cells - although there are some organic secondary ion signals ($C_2H_2^-$, PO_2^- , $C_4H_{11}N_2O^-$) origination from the structure seen in the ToF-SIMS DCC which is regarded as cells there is no further proof of such a thesis (a peak list with selected ions and ions used to visualize the structure is shown in Table 5.2). Although this sample was analyzed using light microscope it was not possible to find the sample area. Normally this structures (as seen in the red square) represent cells and normally there are no contaminations that mimic such cell bulks - but this can not be proofed for this designated structure.

What should be noted are the high molecular mass fragments labeled in the spectrum which can all be derived from fatty acids, especially $C_{39}H_{79}O_{14}$ (m/z 771), $C_{28}H_{46}O_4Bi^-$ (m/z 654), $C_{30}H_{47}O_5^-$ (m/z 487), $C_{20}H_{41}O_{12}^-$ (m/z 473), $C_{20}H_{33}P_2O_7^-$ (m/z 447), $C_{18}H_{29}PO_4^-$ (m/z 340) and $C_{17}H_{27}PO_4^-$ (m/z 326). Unfortunately they are only seen in low secondary ion concentration hence visualization is not convincing.

Of interest in this analysis is the first layer the ion beam must penetrate when hitting *E. coli* - the lipopolysaccharide layer. The idea was that it might be possible to find saccharides in the spectra and - as a known compound present in *E. coli* - hyaluronans (monomers: made from D-glucuronic acid (m/z 194.14) and D-N-acetylglucosamine (m/z 221.21)). Unfortunately neither $[M]^+$, $[M]^-$ nor $[M+Na]^+$ were found [Shard *et al.*, 1997].

No	Center Mass	Assignment	Area [a.u.]
1	24	NaH ⁻	255
2	25	C ₂ H ⁻	16,711,680
3	26	CN ⁻	65,280
4	63	PO ₂ ⁻	33,023
5	76	CHPS ⁻	12,732,818
6	77	CH ₂ PS ⁻	32,896
7	78	CH ₂ S ₂ ⁻	40,960
8	79	C ₂ H ₁₁ N ₂ O ¹	40,960
9	28	Si ⁻	16,744,576
10	27	C ₂ H ₃ ⁻	2,147,552
11	49	C ₄ H ⁻	16,744,576
12	119	C ₃ H ₇ N ₂ O ₃ ⁻	444,586
13	69	C ₂ H ₃ N ₃ ⁻	65,280
14	71	CHN ₃ O ⁻	8,421,504
15	85	C ₂ H ₃ N ₃ ⁻	16,744,576
16	85	OHN ₃ ⁻	40,960
17	59.9	-	16,728,128
18	60	N ₂ O ₂ ⁻	8,421,504
19	61	NOP ⁻	2,147,552

Table 5.2: n0208m102 - Peak list used to visualize biomass

To summarize the result: it is often possible to identify biomass based on specific secondary ions in comparison with the ToF-SIMS CCD camera image of the sample with this setting and this sample selection but crenarchaeol was not detected in any of the spectra.

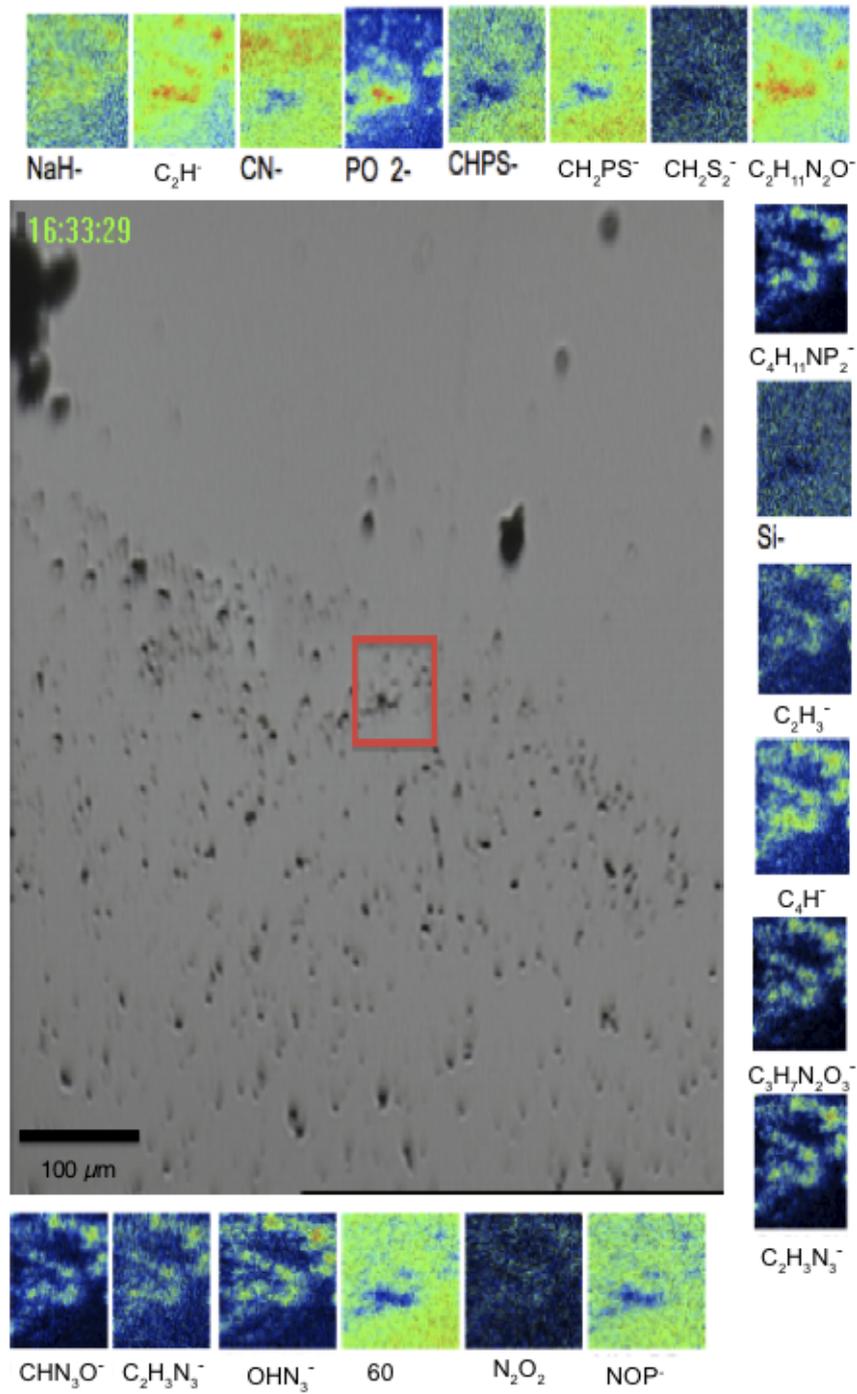


Figure 5.5: Visualization of n0208m102 - a selection of the most characteristic ions for biomass. The measurement area is indicated by the red rectangle and has an area of $80 \times 80 \mu\text{m}$. A discussion is given in the text.

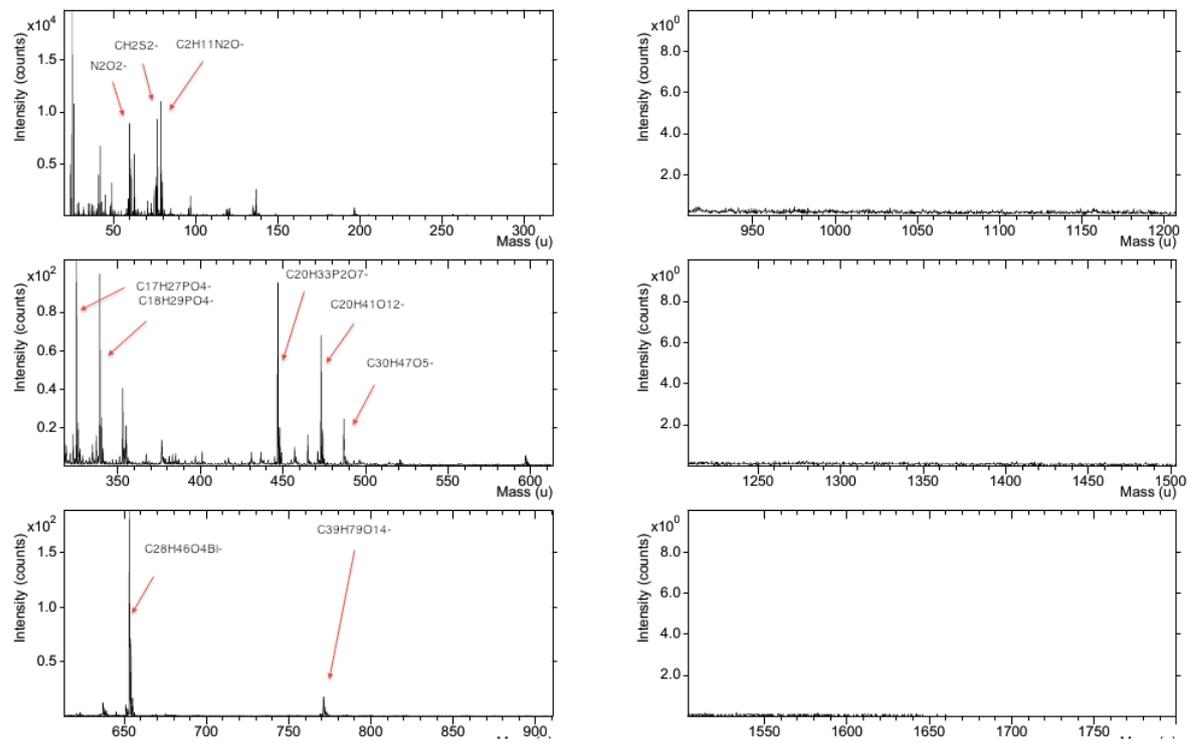


Figure 5.6: Spectrum of sample n0208m102

5.2.3 Data set 0416m*

This data series was done because it was not possible to detect crenarchaeol in all previous measurements. The main focus of this series was on the measurement p0416m001 which exceeded the static SIMS level using only the bismuth gun over the course of nearly 12 h primary ion bombardment. This was done because bombardment with the DSC damages the sample and crenarchaeol - supposedly resulting in a complete fragmentation of the molecule. It is presumed that the bombardment with bismuth ions should avoid this limitation. From the measurement series 0416m* only two data sets will be discussed: p0416m101 and p0416m001. All other measurements were done to choose the right place for the long-term analysis p0416m001. A peak list was automatically generated using the Auto Peak Search Mode (Single Range) of the SurfaceLAB6 software with the following constrains: Start Mass (u): 0, End Mass (u): 1500 and Intensity Threshold (cts): 200. After the automatic peak search a manual search was performed based on the spectra and masses were assigned to relevant peaks. The settings for the analysis can be seen in Table 2.5.

p0416m101

Mass calibration was performed utilizing $C_2H_5^+$, $C_3H_7^+$, $C_4H_9^+$ and $C_5H_9^+$. A short sputtering cycle (5 s - non-interlaced sputtering) was performed prior to analysis with the bismuth cluster gun using the oxygen ion gun (2 keV) on a 300x300 area overlaying the area of the later bismuth bombardment. Peaks and masses are selected due to their ability to reflect the topology seen in the ToF-SIMS CCD camera image (see Figure 5.7). This measurement reveals a trend that has been seen before but probably through the higher energy of the co-sputtering gun gains strength in this and the following analyzed data sets. Biomass can be associated with low intensity signals of oxoacids of phosphorus and associated salts but with high intensity signals with Na^+ , Ca^+ , CaH^+ , CaO^+ , iron, copper and nickel. All of these elements are present in prokaryotes. The silica secondary ion distribution reveals a negative image of the biomass distribution as well as the silicon hydride.

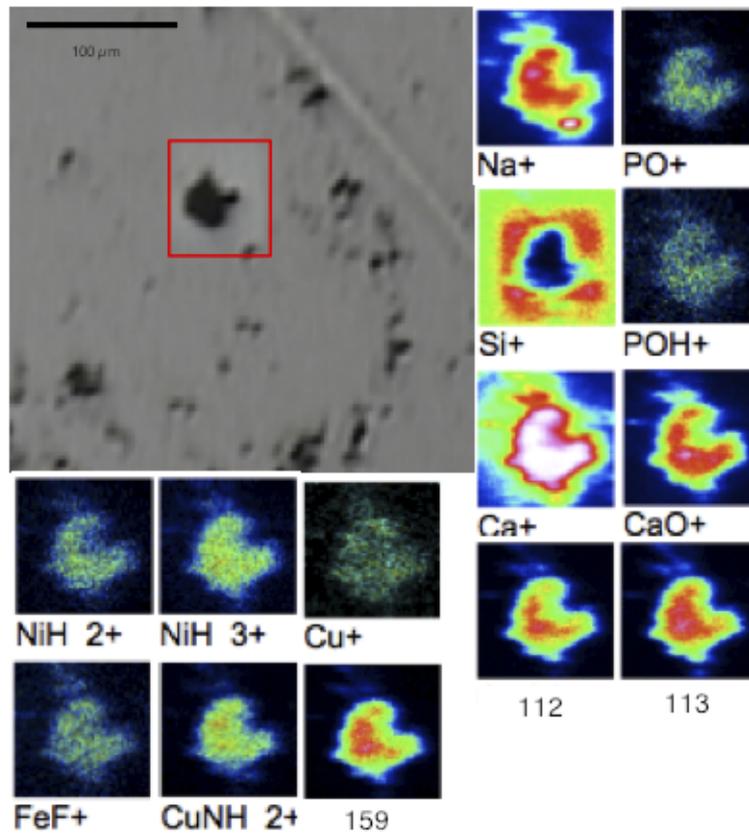


Figure 5.7: Visualization of p0416m101 - a selection of the most characteristic ions for biomass. The measurement area is indicated by the red rectangle and has an area of $70 \times 70 \mu\text{m}$. A discussion is given in the text.

p0416m001

The sample consists of *N. viennensis* cells. This analysis was unique: 87071 scanning cycles were obtained, exceeding the static SIMS limit (Total Ion Dose of 1.12×10^{17} ions/cm²) with a pulsed Bi_3^+ (25 keV) beam over the course of nearly 12 h bombardment (data acquisition time of 58119 seconds). It was possible to determine biomass and reconstruct the topology of the ToF-SIMS camera image. Also high mass fragments were seen with higher intensity but crenarchaeol was not detected. The lateral distribution of a few selected secondary ions which show the topology of the biomass are shown in Figure 5.10. What is striking is the presence of many inorganic compounds - iron, chromium, ce-

sium, cobalt, tin, aluminum, niobium - this might of course be due to the automated molecule selection based on the exact mass. This might be the result of calibration with molecules from the lower end of the molecular weight range of measured molecules. But in the case of this analysis different ions for calibration were used than in the other measurements containing some medium weight molecular masses: C^+ ($m/z = 12.01$), CH_2^+ ($m/z = 14.03$), Na^+ ($m/z = 22.99$), SiH^+ ($m/z = 29.09$), $C_3H_3^+$ ($m/z = 39.06$), CaO^+ ($m/z = 56.08$), $CaOH^+$ ($m/z = 57.09$), $FeNH_2^+$ ($m/z = 71.87$), Ca_2O^+ ($m/z = 96.16$), $K_2SO_4H^+$ ($m/z = 175.27$).

Although the intensity of high molecular mass secondary ions decreases with higher mass (763 about 100 counts, 850 about 75 counts, 960 about 60 counts and 1017 about 45 counts) the intensity is much higher than in all previous measurements and only in part matched by the measurement p0525m112 (peak list shown in Table 5.4) for which the wafer material was changed from silica to ITO and ammonium hydrogen carbonate was used as buffer instead of PBS.

Also high molecular mass secondary ions can be used to visualize cells. Good examples are $C_{44}H_{45}P_2O_8^+$ ($m/z = 764$) and $C_{31}H_{60}O_{16}Na^+$ ($m/z = 712$) - this was not possible in all previous analysis most likely due to the low ion density of the bismuth beam and fewer destruction of the surface layer.

The recurring peaks at high molecular masses are noticeable. Due to the long measurement time secondary ions with low intensity can be detected. Although the counts for these signals (ranging from about 100 to 20 counts per binned channel) are low in number they stand out against the noise background of the baseline. One blocks (indicated by the blue rectangles) was further analyzed (displayed in Figure 5.9) to get an insight in the fragmentation pattern. The block spans from $m/z = 1010$ to $m/z = 1140$.

The recurring peaks at high molecular masses (starting at about 700 u) may have many sources. Normally such patterns hint at degrading biopolymers but the small mass

differences between many peaks (mass difference of 4 units (e.g. 1017 and 1021, 1073 and 1077)) make it more likely that this are the results of deprotonation.

What should also be noted is the molecule with the assigned molecular formula of $C_{44}H_{45}P_2O_8^+$ and a molecular mass of 764. Not only can it be used to map the biomass it is also possible to derive this molecule from crenarchaeol through fragmentation (shown in Figure 5.12). Unfortunately, this remains speculative since we weren't able to analyze crenarchaeol with MALDI and with polar head groups (phosphate moieties and hexoses) with MALDI and ToF-SIMS. It remains unclear which fragments the analysis of crenarchaeol with polar head groups would yield.

A depth profile for all ions above 280 u was also obtained - no characteristic change in intensities was seen. The secondary ion intensity of these ions had no obvious gradient. Only one secondary ion showed a slight increase in the course of the measurement: $C_{18}H_{38}O_2$. The depth profile for this ion is shown in Figure 5.11.

No	Center Mass	Assignment	Area [a.u.]
1	286	$C_{18}H_{38}O_2^+$	2129984
2	295	-	16728128
3	347	-	8421504
4	513	$C_{32}H_{65}O_4^+$	12632256
5	541	$C_{28}H_{80}Na_2OP_2^+$	8388863
6	652	$C_{46}H_{60}NaO^+$	65280
7	712	$C_{31}H_{60}O_{16}Na^+$	255
8	764	$C_{44}H_{45}P_2O_8^+$	2147552
9	850	$C_{55}H_{49}Na_2O_2P_2$	16711680
10	962	$C_{42}H_{82}O_{22}Na^+$	128

Table 5.3: p0416m001 - Peak list obtained in high molecular mass range

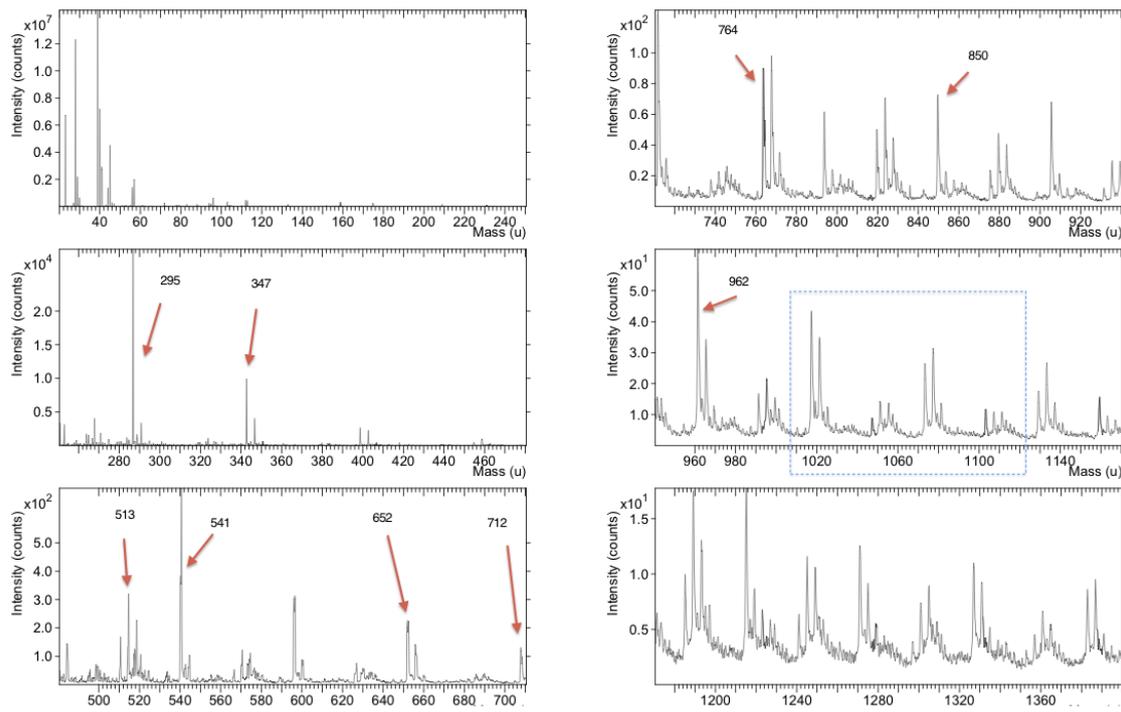


Figure 5.8: Spectrum of p0416m001 - the most characteristic ions in high molecular mass range are marked.

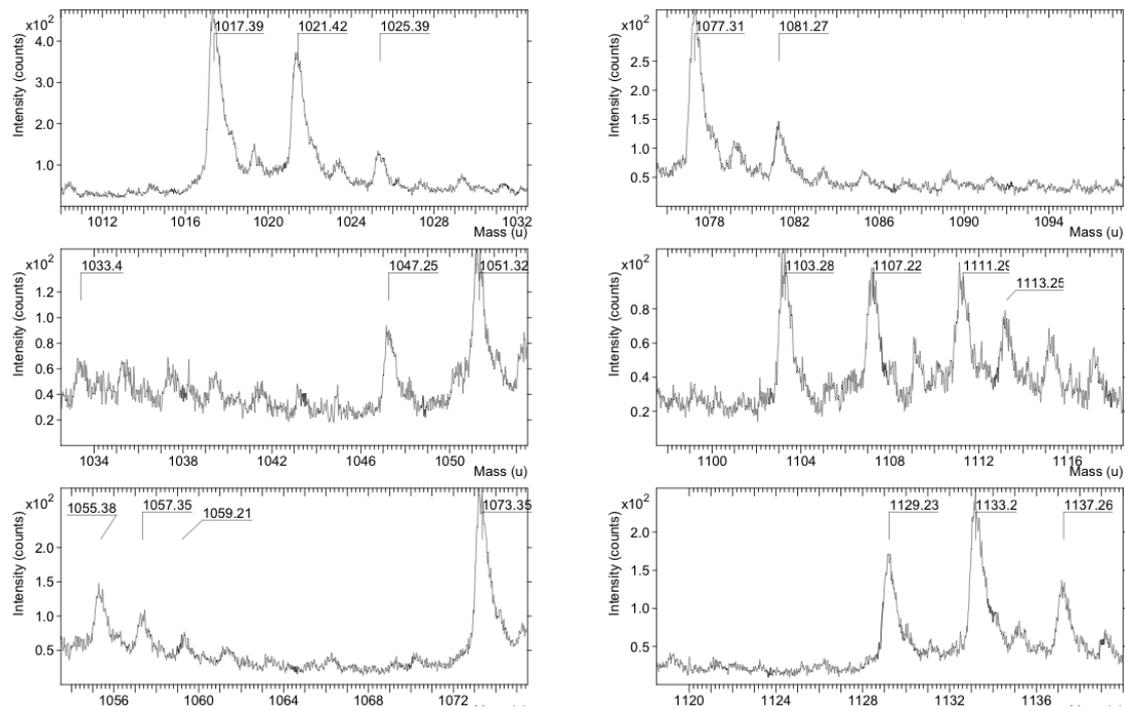


Figure 5.9: Spectrum of measurement p0416m001 between 1010 u and 1140 u

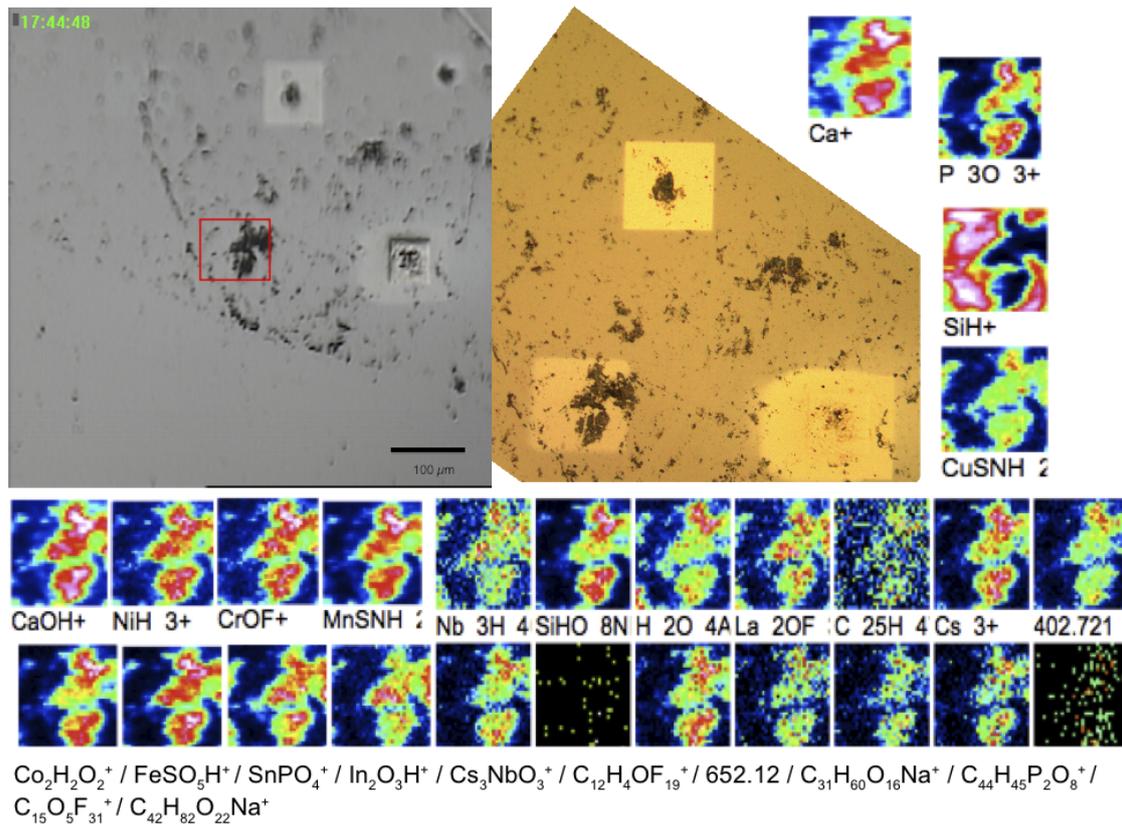


Figure 5.10: Visualization of p0416m001 - ToF-SIMS camera image and microscope image with a selection of the most characteristic ions for biomass. The left picture at top shows the ToF-SIMS camera image (measurement area is indicated by the red rectangle and has an area of $100 \times 100 \mu\text{m}$), the right picture at top the image taken with a reflected-light microscope (80 x).

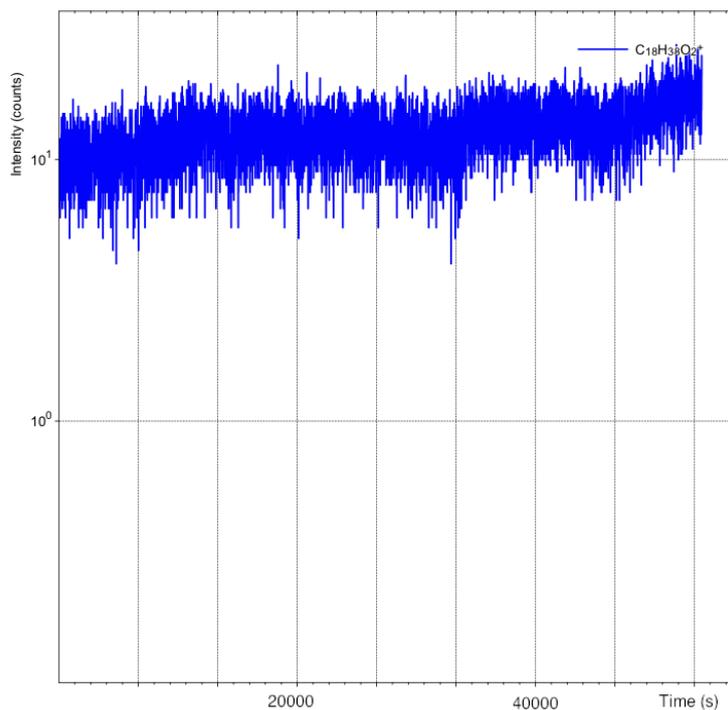


Figure 5.11: Depth Profile of p0416m001 for $C_{18}H_{38}O_2$ - the y-axis corresponds to the sputter time. Cycles were not binned.

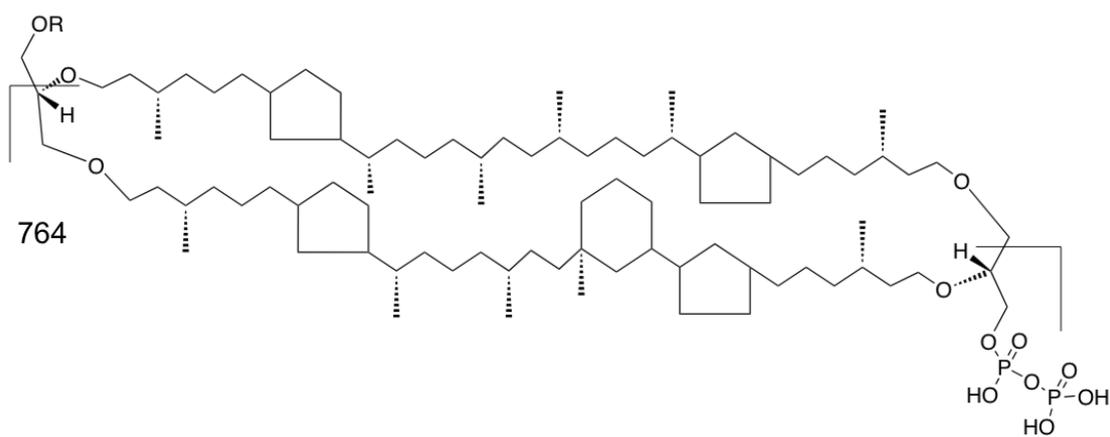


Figure 5.12: Proposed fragmentation pattern for $C_{44}H_{45}P_2O_8^+$.

5.2.4 Data set 0525m*

For the data set 0525m* only one visualization is shown because this run was simply to obtain many spectra with little total ion dose from different locations on the sample slide to look for crenarchaeol. The location of the analysis was changed during the analysis in this data set to get surface spectra of as many as possible cells. For this analysis the Si-wafers were exchanged for ITO (Indium-Tin oxide) coated glass-wafers. Also the washing buffer was not PBS but $\text{NH}_4\text{HCO}_3^+$ and no ethanol or PFA was used to fix the cells. The settings for the analysis can be seen in Table 2.6.

What should be noted is that in the spectrum there are higher molecular mass secondary ions present (e.g. $\text{C}_{34}\text{H}_{74}\text{NO}^+$ (m/z 513) and $\text{C}_{31}\text{H}_{71}\text{KNNa}_2\text{O}_6^+$ (m/z 639)), even though with low intensity compared to medium or low molecular mass ions. But in this regard a significant improvement to all previous measurements (with the exception of the long time analysis) is visible.

The spectrum of measurement p0525m112 between m/z 700 u and 960 u is displayed in Figure 5.14. These peaks are also present in the peak list in Table 5.4 - peak number 10 to 18. Each of them can be seen as a fragment of lipids (e.g. $\text{C}_{39}\text{H}_{79}\text{O}_{14}^+$) but molecular formula assignment must be handled carefully - low secondary ion counts and high molecular masses tend to broaden the peak therefore making assignment based on the center of the peak inaccurate. What can be said definitely is that the change in experiment settings (ITO-slide, change of buffer from PBS to ammonium hydrogen carbonate) resulted in an increase of high molecular mass fragments.

To summarize the result: crenarchaeol was not observed in any of the obtained spectra. The two different settings (PBS in contrast to Ammoniumhydrogencarbonate) are not directly compared here because for such an analysis the biomass must be completely excluded from the analysis to make sure that matrix effects due to organic substances are at a minimum. Such a condition was not met with this data set.

No	Center Mass	Assignment	Area [au]
1	359	$C_{23}H_{82}^+$	16,744,576
2	361	$C_{22}H_{82}N^+$	8,454,016
3	375	$C_{32}H_{82}N_2^+$	2,129,984
4	377	$In_3O_2^+$	16,711,808
5	382	$C_{25}H_{81}^+$	32896
6	490	$C_{23}H_{81}NNaO_4P^+$	1,263,2256
7	492	$C_{26}H_{78}NNaO_4^+$	40,960
8	513	$C_{34}H_{74}NO^+$	16,728,128
9	639	$C_{31}H_{71}KNNa_2O_6^+$	8,388,863
10	752	-	270
11	753	-	439
12	767	-	250
13	769	$C_{39}H_{79}O_{14}^+$	86
14	769	-	273
15	883	-	572
16	898	-	270
17	900	-	901
18	916	-	348

Table 5.4: p0525m112 - Peak list obtained in high molecular mass range

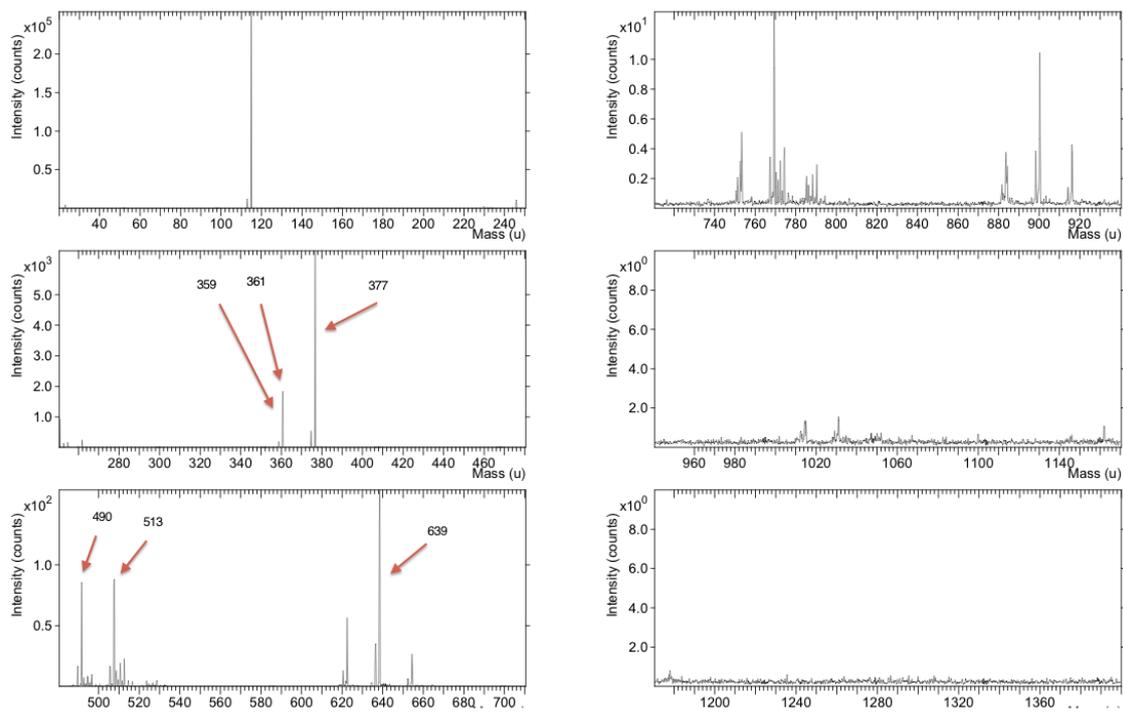


Figure 5.13: Spectrum of p0525m112 - the most characteristic ions in high molecular mass range are marked. The masses for the peaks above m/z 700 u can be seen in Figure 5.14

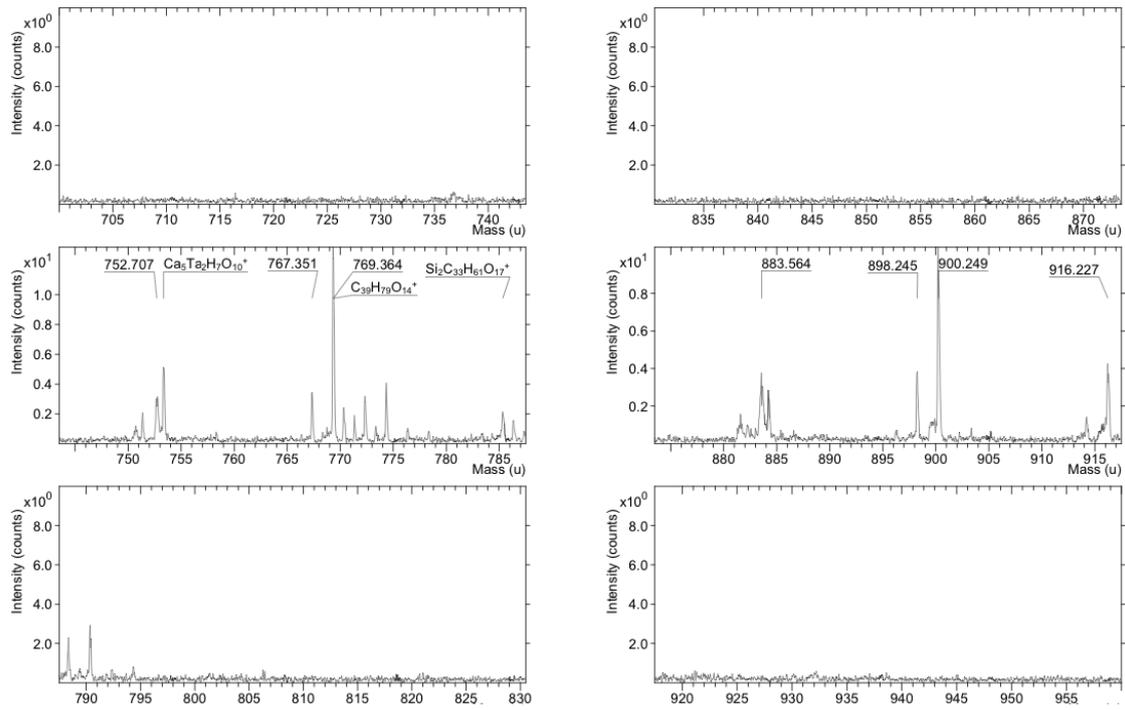


Figure 5.14: Spectrum of measurement p0525m112 between 700 u and 960 u

5.2.5 Chemometrics

All statistical tests performed in this section were done using the R software package (www.r-project.org) and local scripts written in Perl.

Data acquisition was done as following: Export of the whole spectra in ASCII format from the SurfaceLab6 software. In the end a mass binning on every unit ($\pm 0.5u$) was the best option (only on mass level - channel binning was not done) - this was a very critical step because many approaches failed because of too low mass binning (100 to 1000 channels) and the problems that arise from the fact that channel number to associated masses can differ by very little from measurement to measurement. Sometimes this is enough to have important masses binned to different units. What was also done but discarded later was exporting the spectra without mass binning. This failed because of the differences in calibration and the problems with zero intensity at certain masses in the PCA calculations.

Another crucial step was the question how the data should be preprocessed. What was done for all spectra was a Poisson correction of the peak intensities in order to minimize channel time effects at the detection system. Optional mean centering (for better description of variation from mean) and normalization by the accumulated signal intensities (to emphasise relative changes in ion intensities) were performed and evaluated. Scaling (by the sum or the standard deviation of the particular peak intensities (secondary ion)) was not done.

Correlation of data sets

First a correlation test was performed with two objectives:

- (1) Are the relative peak intensities in the spectra correlate with the different microbial species ?
- (2) Which secondary ion species have a negative or positive correlation in signal intensity?

The first question was tackled by a correlation matrix of the data series p0520m10*

in which all measurements in positive mode which had comparable cycle time (in order to analyze masses above 1200 units) were included. The intensity for every unit is considered. The Pearson correlation coefficient is computed and a covariance matrix is obtained Table 5.7. Of the 13 correlation coefficients over the value 0.89 nine are correlated to the same microbial species and 4 show high correlation to the other species. A closer look at the four cross-species pairs reveals that of the 8 samples 4 were only surface spectra, 3 were short co-sputtered spectra and only 1 was a long co-sputtered spectrum. But without regard of the different spectra this shows that correlation coefficients are not sufficient to cluster the different bacteria. The reason for this might very well be the fact that the spectra contain (with exception of all long co-sputtered spectra) mainly surface material and this might be responsible for this inconclusive findings.

Question (2) was asked because it was simply not clear if the short co-sputtering does a good job in cleaning the surface of the sample from organic and inorganic contaminants. The idea was that if differences in intensity of the observed spectra are due to differences in the ratio of wafer to biomass than biomass should correlate with organic ions in contrast to all wafer borne masses which should not or inversely correlate with the intensity of the biomass. Question (2) was answered with a setting different to question (1) : a manually obtained peak list was used in combination with the test set p0525m10*. This set was used because it was the biggest set of the same species under the same conditions with short co-sputtering. The manually obtained peak list contained 73 peaks (see Table 5.6). This peak list is used to obtain all masses with a correlation factor above 0.9 and below -0.5. The peaks for sodium (Na) and indium (In) were excluded from the analysis.

A list of 715 correlations satisfying the requirements were obtained. Interestingly only 58 negative (below -0.5) in contrast to 657 positive correlations were found. With very little exceptions negative correlation was found between organic low-mass molecules and organic high-mass molecules as well as inorganic molecules and high-mass organic molecules. Not all correlations and masses can be considered here so two generic molecules

have been selected: In_2^+ and $\text{C}_{64}\text{H}_{126}^+$.

As shown for these two examples there is no clear trend visible. In_2^+ negatively correlates with organic middle- to high-mass range molecules but also correlates positively with - besides inorganic ions as Cs^+ , indium-molecules and Ca^+ - high-mass organic molecules such as $\text{C}_{39}\text{H}_{67}\text{O}_6\text{Na}^+$ which are very likely derived from lipids. For $\text{C}_{64}\text{H}_{126}^+$ the picture is much more convincing. Only high molecular-mass molecules correlate with this ion, the only molecule with less than 22 carbon atoms correlates inversely. A reasonable explanation for this inverse correlation with $\text{C}_3\text{H}_5\text{N}_3\text{O}^+$ might be that since nitrogen is mostly found in proteins it would not be expected in the cell wall whereas all other fragments can be derived from lipids. But even though the correlation pattern for $\text{C}_{64}\text{H}_{126}^+$ appears promising, this is the exception and not the rule and therefore it is not possible to determine with certainty if a short co-sputter process is sufficient to get rid of all contaminations based on these findings. It is not clear if the short co-sputtering cycle at the beginning allows to analyze the sample surface and not the contaminations.

Pearson Correlation Coefficient	Molecular Formula of the correlated species	
(-0.54694442)	In_2^+	$\text{C}_4\text{H}_5\text{N}_3\text{O}^+$
0.97962498	In_2^+	Cs^+
0.99994966	In_2^+	$\text{C}_3\text{H}_3\text{Ni}_2^+$
(-0.53749007)	In_2^+	$\text{C}_{11}\text{H}_9\text{N}^+$
0.982777005	In_2^+	LaO_2^+
(-0.50995977)	In_2^+	$\text{C}_3\text{H}_7\text{NO}^+$
(-0.522270048)	In_2^+	$\text{C}_8\text{H}_{18}\text{O}_5\text{Na}^+$
0.99979515	In_2^+	In_2O^+
0.99740632	In_2^+	In_2OH^+
0.9625962608	In_2^+	$\text{C}_{15}\text{H}_{20}\text{O}_3^+$
0.98903096	In_2^+	InSnO^+
(-0.519423570)	In_2^+	$\text{C}_{14}\text{H}_{26}\text{O}_8\text{Na}^+$
0.963402168	In_2^+	$\text{C}_{22}\text{H}_{45}\text{O}_4^+$
0.9640922263	In_2^+	$\text{C}_{22}\text{H}_{41}\text{O}_2\text{K}^+$
0.99700358	In_2^+	Ca^+
0.99981490	In_2^+	$\text{In}_4\text{O}_4\text{H}^+$
0.99960041	In_2^+	C_4H_8^+
0.99990672	In_2^+	CaOH^+
0.99486455	In_2^+	$\text{C}_3\text{H}_7\text{O}^+$
0.97392451	In_2^+	$\text{C}_{39}\text{H}_{67}\text{O}_6\text{Na}^+$
0.996290285	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{22}\text{H}_{48}\text{NO}_2^+$
0.997610432	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{22}\text{H}_{41}\text{O}_2\text{Na}^+$
0.95539269	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{25}\text{H}_{52}\text{O}_5^+$
0.997774225	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{28}\text{H}_{48}\text{O}_4^+$
0.99626272	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{34}\text{H}_{51}\text{O}_2^+$
0.97804450	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{34}\text{H}_{67}\text{O}_2^+$
0.99939111	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{32}\text{H}_{62}\text{O}_4^+$
0.99993613	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{36}\text{H}_{80}$
0.99896460	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{43}\text{H}_{74}\text{O}_2^+$
0.9974135	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{43}\text{H}_{72}\text{O}_3^+$
0.99130956	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{43}\text{H}_{72}\text{O}_3^+$
0.99937387	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{35}\text{H}_{70}\text{O}_{16}\text{Na}^+$
0.998179503	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_{63}\text{H}_{124}\text{O}_2^+$
(-0.56798815)	$\text{C}_{64}\text{H}_{126}^+$	$\text{C}_3\text{H}_5\text{N}_3\text{O}^+$

Table 5.5: p0525m10* - Two selected masses of the correlation matrix obtained from 73 selected masses of the data set p0525m10*.

Nr.	Mass	Assigned formular	Area	Nr.	Mass	Assigned formular	Area
1	18.036472	NH_4^+	255	27	132.903195	Cs^+	16744576
2	22.991205	Na^+	16711680	28	145.078354	$\text{C}_8\text{H}_{11}\text{F}_2^+$	16711808
3	27.023644	C_2H_3^+	65280	29	146.068086	$\text{C}_7\text{H}_{14}\text{SO}^+$	8454016
4	29.039987	$\text{C}_2\text{H}_4\text{H}^+$	8421504	30	147.073545	$\text{C}_8\text{H}_9\text{N}_3^+$	8454016
5	30.035888	CH_4N^+	33023	31	148.071528	$\text{C}_8\text{H}_8\text{N}_{20}^+$	12632256
6	39.023133	C_3H_3^+	16744576	32	154.907672	$\text{C}_3\text{H}_3\text{Ni}_2^+$	2129984
7	39.962815	Ca^+	8454016	33	155.078224	$\text{C}_{11}\text{H}_9\text{N}^+$	40960
8	41.039529	C_3H_5^+	2129984	34	170.862932	LaO_2^+	16711808
9	44.051375	$\text{C}_2\text{H}_6\text{N}^+$	16711808	35	173.087813	$\text{C}_{12}\text{H}_{13}\text{O}^+$	16728128
10	55.056642	C_4H_7^+	32896	36	207.028487	$\text{C}_{14}\text{H}_7\text{O}_2^+$	32896
11	55.970603	CaO^+	12632256	37	217.112950	$\text{C}_8\text{H}_{18}\text{O}_5\text{Na}^+$	8388863
12	56.056039	C_4H_8^+	40960	38	229.798262	In_2^+	12632256
13	56.966107	CaOH^+	16728128	39	245.798629	In_2O^+	40960
14	58.977261	NaHCl^+	8388863	40	246.797421	In_2OH^+	16728128
15	59.051242	$\text{C}_3\text{H}_7\text{O}^+$	2147552	41	248.816590	$\text{C}_{15}\text{H}_{20}\text{O}_3^+$	16744576
16	60.054521	CH_6N_3^+	128	42	250.792372	InSnO^+	8388863
17	67.055724	C_5H_7^+	255	43	281.061741	$\text{C}_{17}\text{H}_{13}\text{O}_4^+$	2147552
18	70.070861	$\text{C}_5\text{H}_{10}^+$	16711680	44	284.320809	$\text{C}_{19}\text{H}_{42}\text{N}^+$	128
19	73.055576	$\text{C}_3\text{H}_7\text{NO}^+$	65280	45	298.333105	$\text{C}_{20}\text{H}_{44}\text{N}^+$	255
20	83.050550	$\text{C}_3\text{H}_5\text{N}_3^+$	8454016	46	324.963290	$\text{C}_{18}\text{H}_{14}\text{PO}_4^+$	16711680
21	84.086178	$\text{C}_6\text{H}_{12}^+$	2129984	47	326.363902	$\text{C}_{22}\text{H}_{48}\text{N}^+$	65280
22	86.102391	$\text{C}_6\text{H}_{14}^+$	8421504	48	326.941729	$\text{K}_3\text{N}_7\text{O}_7^+$	8421504
23	99.045072	$\text{C}_3\text{H}_5\text{N}_3\text{O}^+$	2129984	49	345.132658	$\text{C}_{14}\text{H}_{26}\text{O}_8\text{Na}^+$	2147552
24	111.048356	$\text{C}_4\text{H}_5\text{N}_3\text{O}^+$	16711808	50	358.700621	$\text{C}_{22}\text{H}_{48}\text{NO}_2^+$	33023
25	114.902866	In^+	33023	51	360.686356	$\text{C}_{22}\text{H}_{41}\text{O}_2\text{Na}^+$	16744576
26	127.085904	$\text{C}_7\text{H}_{11}\text{O}_2^+$	32896	52	374.715875	$\text{C}_{22}\text{H}_{45}\text{O}_4^+$	8454016
53	376.681234	$\text{C}_{22}\text{H}_{41}\text{O}_2\text{K}^+$	2129984	63	510.578897	$\text{C}_{32}\text{H}_{62}\text{O}_4^+$	255
54	420.690179	$\text{C}_{25}\text{H}_{26}\text{PO}_4^+$	16711808	64	512.625562	$\text{C}_{36}\text{H}_{80}^+$	128
55	432.682636	$\text{C}_{25}\text{H}_{52}\text{O}_5^+$	32896	65	524.564324	$\text{In}_4\text{O}_4\text{H}^+$	8388863
56	448.667183	$\text{C}_{28}\text{H}_{48}\text{O}_4^+$	33023	66	622.521250	$\text{C}_{43}\text{H}_{74}\text{O}_2^+$	2147552
57	476.655289	$\text{C}_{34}\text{H}_{52}\text{O}^+$	8421504	67	636.522309	$\text{C}_{43}\text{H}_{72}\text{O}_3^+$	128
58	491.567165	$\text{C}_{34}\text{H}_{51}\text{O}_2^+$	12632256	68	638.501564	$\text{C}_{43}\text{H}_{72}\text{O}_3^+$	255
59	492,622428	$\text{C}_3\text{1H}_5\text{6O}_4^+$	65280	69	654.508414	$\text{C}_{39}\text{H}_{67}\text{O}_6\text{Na}^+$	16711680
60	494.599520	$\text{C}_{34}\text{H}_{72}\text{N}^+$	16711680	70	769.431520	$\text{C}_{35}\text{H}_{70}\text{O}_{16}\text{Na}^+$	65280
61	496.585527	$\text{C}_{30}\text{H}_{42}\text{PO}_4^+$	40960	71	900.283216	$\text{C}_{64}\text{H}_{126}^+$	8421504
62	507.604743	$\text{C}_{34}\text{H}_{67}\text{O}_2^+$	16728128	72	910.524580	$\text{C}_{63}\text{H}_{122}\text{O}_2^+$	33023
73	916.138048	$\text{C}_{63}\text{H}_{124}\text{O}_2^+$	16744576				

Table 5.6: List of the 73 masses for the correlation matrix

	A001E	B002E	C001N	D002N	E005E	F101N	G102E	H111N	I112E	J001N	K101N	L101N	M102N	N103N	O104N	P121N	Q122N
A001E	1	0.97	0.63	0.92	0.82	0.85	0.96	0.21	0.81	0.67	0.67	0.49	0.03	0.65	0.25	0.02	0.02
B002E	0.97	1	0.67	0.88	0.65	0.85	0.87	0.23	0.64	0.52	0.48	0.55	0.03	0.72	0.25	0.02	0.02
C001N	0.63	0.67	1	0.72	0.34	0.81	0.55	0.24	0.34	0.42	0.29	0.85	0.16	0.73	0.59	0.03	0.03
D002N	0.92	0.88	0.72	1	0.74	0.93	0.89	0.36	0.74	0.71	0.64	0.53	0.06	0.67	0.34	0.03	0.03
E005E	0.82	0.65	0.34	0.74	1	0.63	0.93	0.07	1	0.82	0.93	0.24	0.02	0.31	0.18	0	0
F101N	0.85	0.85	0.81	0.93	0.63	1	0.82	0.25	0.64	0.64	0.6	0.62	0.1	0.68	0.43	0.03	0.03
G102E	0.96	0.87	0.55	0.98	0.93	0.82	1	0.15	0.93	0.76	0.81	0.42	0.03	0.55	0.24	0.01	0.01
H111N	0.21	0.23	0.24	0.36	0.07	0.25	0.15	1	0.11	0.51	0.12	0.13	0	0.14	0.22	0	0
I112E	0.81	0.64	0.34	0.74	1	0.64	0.93	0.11	1	0.85	0.94	0.25	0.03	0.3	0.2	0.01	0.01
J001N	0.67	0.52	0.42	0.71	0.82	0.64	0.76	0.51	0.85	1	0.91	0.33	0.06	0.25	0.37	0	0
K101N	0.67	0.48	0.29	0.64	0.93	0.6	0.81	0.12	0.94	0.91	1	0.23	0.05	0.19	0.26	0	0
L101N	0.49	0.55	0.85	0.53	0.24	0.62	0.42	0.13	0.25	0.33	0.23	1	0.29	0.76	0.67	0.14	0.15
M102N	0.03	0.03	0.16	0.06	0.02	0.1	0.03	0	0.03	0.06	0.05	0.29	1	0.21	0.83	0.98	0.98
N103N	0.65	0.72	0.73	0.67	0.31	0.68	0.55	0.14	0.3	0.25	0.19	0.76	0.21	1	0.43	0.15	0.15
O104N	0.25	0.25	0.59	0.34	0.18	0.43	0.24	0.22	0.2	0.37	0.26	0.67	0.83	0.43	1	0.71	0.72
P121N	0.02	0.02	0.03	0.03	0	0.03	0.01	0	0.01	0	0	0.14	0.98	0.15	0.71	1	1
Q122N	0.02	0.02	0.03	0.03	0	0.03	0.01	0	0.01	0	0	0.15	0.98	0.15	0.72	1	1

Table 5.7: Correlation matrix for the spectra. The Pearson correlation coefficients are rounded to two decimal figures. Red marked are all correlation coefficients which are ≥ 0.89 . The sample names are abbreviated: A001E = p0103m001E, B002E = p0103m002E, C001N = p0208m001N, D002N = p0208m002N, E005E = p0208m005E, F101N = p0208m101N, G102E = p0208m102E, H111N = p0208m111N, I112E = p0208m112E, J001N = p0417m001N, K101N = p0417m101N, L101N = p0525m101N, M102N = p0525m102N, N103N = p0525m103N, O104N = p0525m104N, P121N = p0525m121N, Q122N = p0525m122N

Principal component analysis on test series p0208m*

All further statistical tests were performed with the test series p0208m*. This was the largest test series and the parameters were matchable in regards of experimental factors (Total ion count, ion dose density ($6.53 \cdot 10^{13}$), staining (KI), primary ion kinetic energy (25 keV), etc.). Two question were of interest:

- (1) Is it possible to differentiate between *E. coli* and *N. viennensis* based on their mass spectra and what are the most influential ion species if it is indeed possible?
- (2) Is it possible to distinguish between a region in an analysis sector with biomass and without biomass?

Only positive spectra were used, it was not differentiated between co-sputtering (m11D - D is a placeholder for an arbitrary digit) and pre-sputtering (m10D). Mass binning was performed on every unit ($\pm 0.5u$). The relevant principal component (PC) were manually defined based on their eigenvalues and manually plotting of the first 6 principal components.

The first PCA was performed without further data pre-processing and resulted in the following scores for the first four principal components:

Sample	score for PC1	score for PC2	score for PC3	score for PC4
p0208m001N	-0.0492314011	-0.2805663083	-0.1617951331	0.4391444255
p0208m002N	-0.1768248614	-0.4336154132	-0.0558753417	0.3286735201
p0208m005E	-0.7757410527	0.4880581207	0.2662746451	0.2972766249
p0208m101N	-0.1193360012	-0.360782611	-0.2028994357	0.4700399992
p0208m102E	-0.5911221636	-0.3963222679	-0.3140051709	-0.6160796393
p0208m111N	-0.0240365315	-0.4557752647	0.8717467405	-0.1008457798
p0208m112E	-0.0170531014	0.0096419153	0.0093960933	0.0095344083

Table 5.8: PCA scores - without data preprocessing for data series p0208m*

Clustering of the samples for *N. viennensis* can be seen, but *E. coli* does not cluster in the first two principal components and also the other four components don't change the picture. The PCA is shown in Figure 5.15. A closer look at the loadings responsible for the

separation (Figure 5.16) shows that the following masses had most influence (with loadings above 1,000,000) for the first PC: 23 (Na^+), 27 (C_2H_3^+), 28 (CH_2N^+), 29 (C_2H_5^+), 39 (Ca^+), 41 (C_3H_3^+), 42 ($\text{C}_2\text{H}_4\text{N}^+$), 43 ($\text{C}_2\text{H}_5\text{N}^+$), 44 ($\text{C}_2\text{H}_6\text{N}^+$), 45 (CH_3NO^+), 53 (C_4H_5^+), 55 (C_4H_7^+), 56 (CaO^+), 57 (CaOH^+), 67 (C_5H_7^+), 69 (C_2H_7^+) and as only medium-mass molecule: 157 (BaF^+). These are all very low-mass organic and inorganic molecules/fragments which are ubiquitous in all biological samples.

What must be noted here is that the assignment from mass to molecular formula is based on the peak with maximum intensity in the mass range of 1 u and that might not necessarily represent the influence of that mass on the loading - but that reduction in mass accuracy was the prize to pay for comparing spectra in this way.

The next step was to preprocess the data. Another PCA was performed with (1) normalized data, with (2) normalized and mean-centered and with (3) normalized, mean-centered and a molecular weight cut-off at mass 100 units ignoring peaks below this threshold. A separation of the two species was achieved with all three approaches, but the latter approaches (normalized, mean-centered and a molecular weight cut-off at mass 100 units ignoring peaks below this threshold) was the most recognizable and the PCA result is shown in plain numbers in Table 5.9 as well as graphically in Figure 5.17 with the corresponding loadings in Figure 5.18.

Sample	score for PC1	score for PC2	score for PC3	score for PC4
p0208m001N	-0.4271705979	-0.4314310561	0.1410690032	0.764625038
p0208m002N	-0.2885457887	-0.1990346119	0.3493353936	-0.1797852352
p0208m005E	-0.4339090623	0.6495281735	-0.1225443353	0.1204713614
p0208m101N	-0.3234923873	-0.1603215866	0.5700925085	-0.4925536215
p0208m102E	-0.3699342915	0.1568011221	0.0344839107	-0.0522854179
p0208m111N	-0.4117422808	-0.4241740386	-0.7184851348	-0.3507941959
p0208m112E	-0.367367831	0.3495030772	-0.0251337984	-0.0106333737

Table 5.9: PCA scores - with data preprocessing (normalized, mean-centered and a molecular weight cut-off at mass 100 units ignoring peaks below this threshold) for data series p0208m*

A closer look at the loadings reveals that the following ions have the highest impact on PC 1 (loadings above Δ of 0.04): 102 ($C_3H_6N_2O_2^+$), 103 ($C_3H_3S_2^+$), 104 ($CSNNa_2^+$), 114 ($C_5H_8NO_2^+$), 118 ($C_7H_2O_2^+$), 120 ($C_5H_9O_2F^+$), 140 ($C_6H_8SN_2^+$), 156 ($C_4H_7NO_4Na^+$). And for the PC 2 the following ions were mostly responsible for the discrimination against the samples (loadings above Δ of 0.01): 104 ($CSNNa_2^+$), 118 ($C_7H_2O_2^+$), 120 ($C_5H_9O_2F^+$), 129 ($SiC_3H_7N_3O^+$), 132 ($C_6H_7N_3O^+$), 136 ($Si_2CH_{12}N_4^+$), 140 ($C_6H_8SN_2^+$), 146 ($C_6H_2O_3^+$), 156 ($C_4H_7NO_4Na^+$), 158 ($C_6H_7PO_3^+$), 174 ($C_6H_7PO_4^+$).

The principal component 1 is not sufficient to separate the species, but in combination with PC 2 or with PC 2 alone there is a clustering of *E. coli* and *N. viennensis*. What should be noted separately are the two masses that contribute the most to the differentiation on PC 2: for *E. coli* that was 156 ($C_4H_7NO_4Na^+$) and for *N. viennensis* it is not such a clear picture but even though one mass stands out: 104 ($CSNNa_2^+$). It should also be noted that from a biological point of view there is no persuasive reason for that.

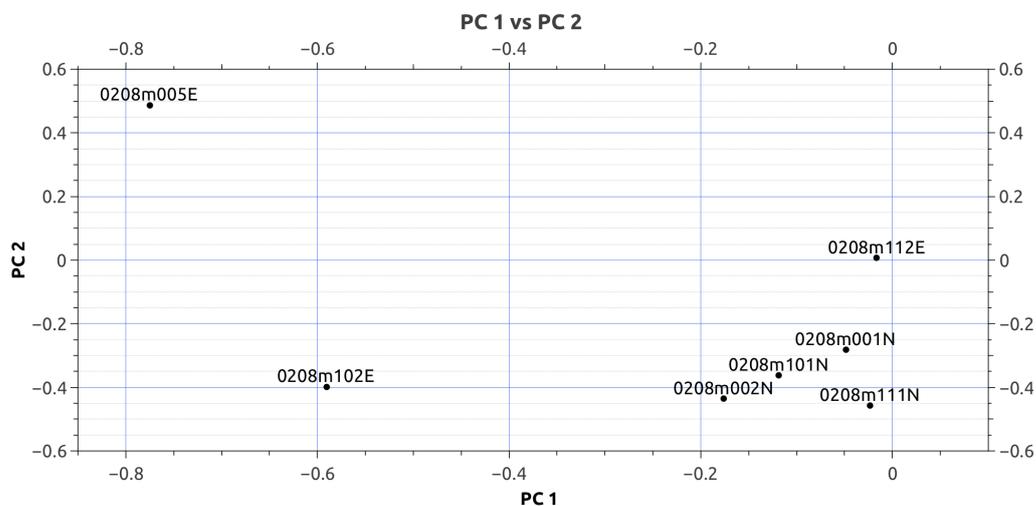


Figure 5.15: Raw Data of p0208m* - Scores of PC 1 vs scores of PC 2. *N. viennensis* clusters but *E. coli* is distributed over the plot without a clear clustering.

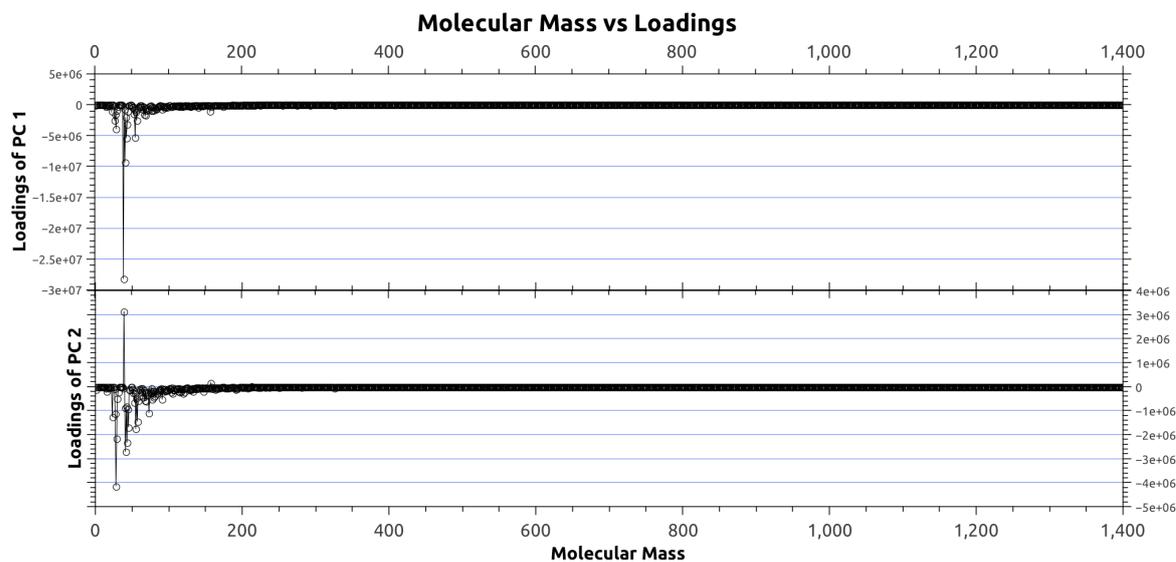


Figure 5.16: Raw Data of p0208m* - Loadings of PC 1 and PC 2. As can be seen the low molecular mass molecules are the factors contributing most to the distribution of the two species.

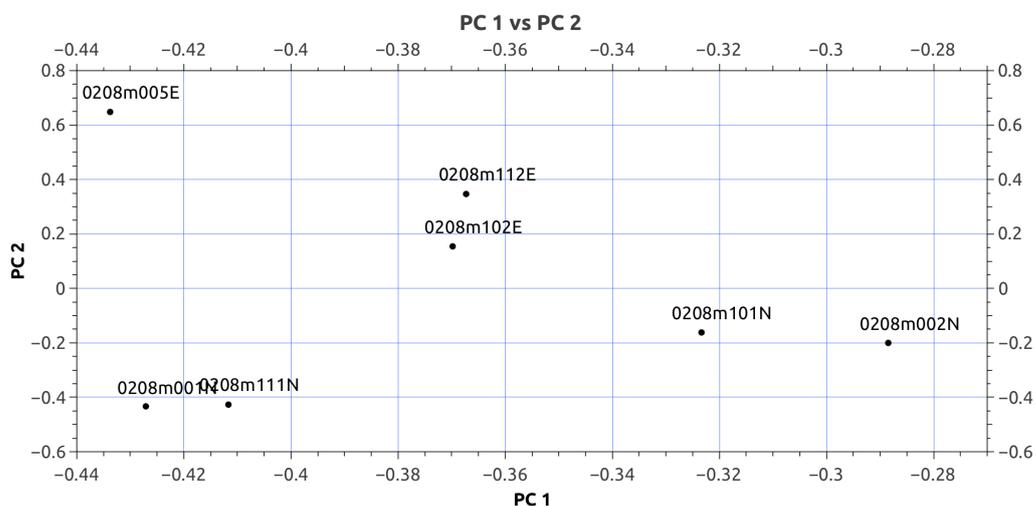


Figure 5.17: Preprocessed Data of p0208m* - Shown are the scores of PC 1 and PC 2 of the Principal compound analysis. Data preprocessing involved: mean-centering and normalization with the total ion current and a mass cut-off of 100 units.

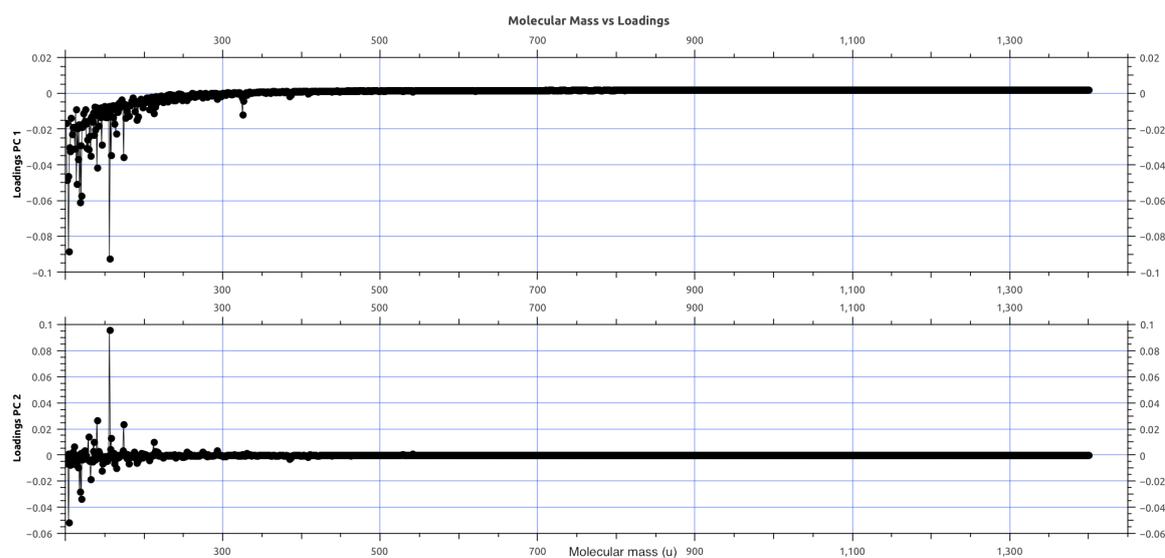


Figure 5.18: Shown are loadings from the Principal component analysis for PC 1 and PC 2 of p0208m*. Data preprocessing involved: mean-centering and normalizing with the total ion current and a cut-off of 100 units.

Principal component analysis of biomass vs non-biomass

A striking question was if it is possible to discrimination against biomass and non-biomass areas via PCA - this means if signals from biomass and non-biomass areas clusters in a PCA. To answer this question, six data sets (p0208m001, p0208m002, p0208m005, p0208m102, p0208m111, p0208m112) were analyzed and a data set was produced based on spectra from different regions on the wafer and split in two different groups: non-biomass region and biomass region. This was done with a mass filter. Every pixel which has been hit with the primary ion beam was analyzed. For every pixel it was decided to which group (biomass or non-biomass) it belongs based on the amount of ions which were markers for biomass/non-biomass and if they surpass or undercut a certain threshold on this pixel. Biomass was defined with very little silica counts and many CaOH^+ , the opposite was true for non-biomass. The marker for biomass changed sometimes to other alkali or earth alkali metal adducts or low mass organic molecules (OH, PO, SNO). This was done only in accordance with the silicon peak and with other medium to high-molecular

mass images. In Figure 5.19 the region for biomass and non-biomass are imaged based on the lateral abundance of two different ions and an image as seen from the ToF-SIMS camera is also shown.

The first PCA was calculated from mean-centered data that were additionally normalized by the TIC (Total Ion Count). It is shown in Figure 5.21. The first 5 principle component scores are not able to differentiate between the biomass (biomass is indicated by an appending B of the sample name in Figure 5.21) and non-biomass (appending NB). A closer look at the loadings (not shown here) revealed that all masses relevant in the loadings for the principal components can be found below 50 u. Therefore a mass cut-off of 50 u was performed, ignoring all masses below 50 u for further analysis. Furthermore two additional measurements were added to the analysis: m111N (*N. viennensis*) and m112E (*E. coli*) - both measurements were obtained using the oxygen gun with an ion dose density of $6.3 \cdot 10^{18}$ ions*cm². Also in this analysis there is no clustering of the two different samples.

Furthermore a PCA was performed with a data set of only biomass ROIs and an additional data set with only non-biomass ROIs (in case that the buffer and the media of the two species were detectable on non-biomass ROIs) but no clear clustering between the microbial species was observed.

What needs to be emphasized at this point : if the whole spectrum (i.e. the complete mass range) is taken into consideration it is not possible to distinguish between biomass and contamination on wafer. In other words, this may very well mean that the contaminants on the wafer surface and the biomass are more similar than biomass and wafer are distinguishable.

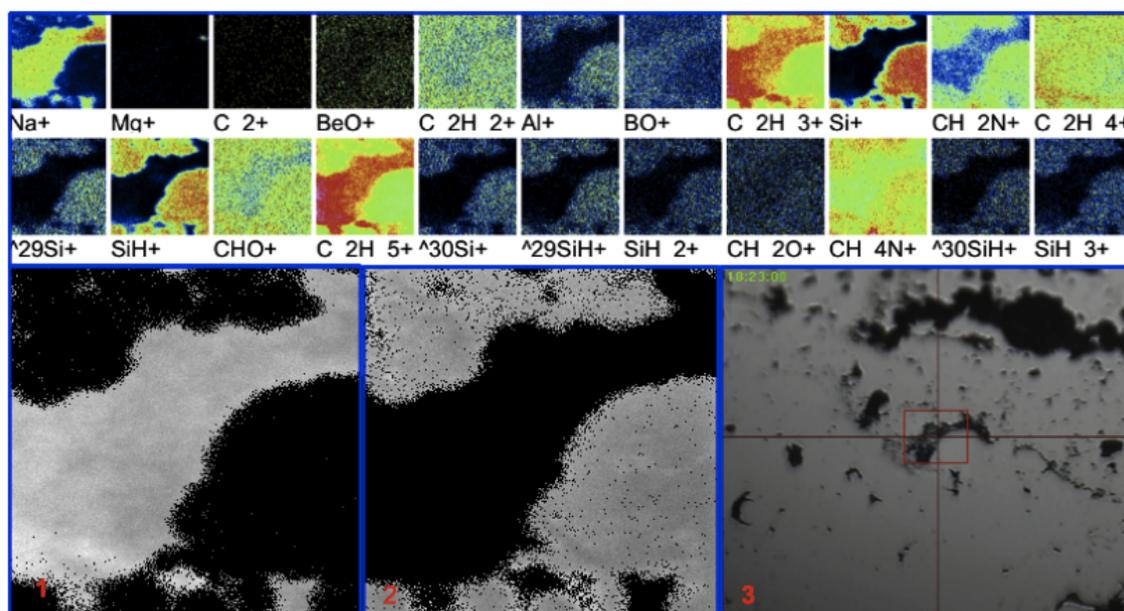


Figure 5.19: An example for biomass and non-biomass on the same sample - measurement p0208m102. Figure 3 shows the wafer as seen from the ToF-SIMS camera after the ion bombardment. The red cross and the red square mark the area of analysis. The small pictures above represent the lateral distribution for the ion on the wafer. In accordance with other, not shown ions biomass was defined with two constrains: maximum Na^+ and minimum Si^+ , the contrary was true for non-biomass. Figure 1 and 2 show the lateral distribution of pixel that were filtered. Figure 1 shows which pixel were considered for the spectra of biomass whereas Figure 2 shows which pixel were considered for the non-biomass spectra.

Analysis of the bismuth sputtering process

In the hope to find crenarchaeol, an unusual analysis was carried out: an overnight aquisition with the bismuth gun exceeding the static SIMS limit and steadily eroding the sample. This analysis was done with 87071 scans and a primary ion dose of $1.12 \cdot 10^{17}$ ions/cm² yielding secondary ion counts of 1,657,560,118 ($1.657 \cdot 10^9$) ions (acquisition time of 58119 seconds). In this process the static limit was exceeded and the surface was degraded by the bismuth gun. The reason for this analysis was the possibility that crenarchaeol is destructed when co-sputtering is applied and when no co-sputtering is used only contamination is seen in the spectrum. But a skimming of the surface with a Bi_3^+ cluster ion beam could be less destructive, possible resulting in longer fragments

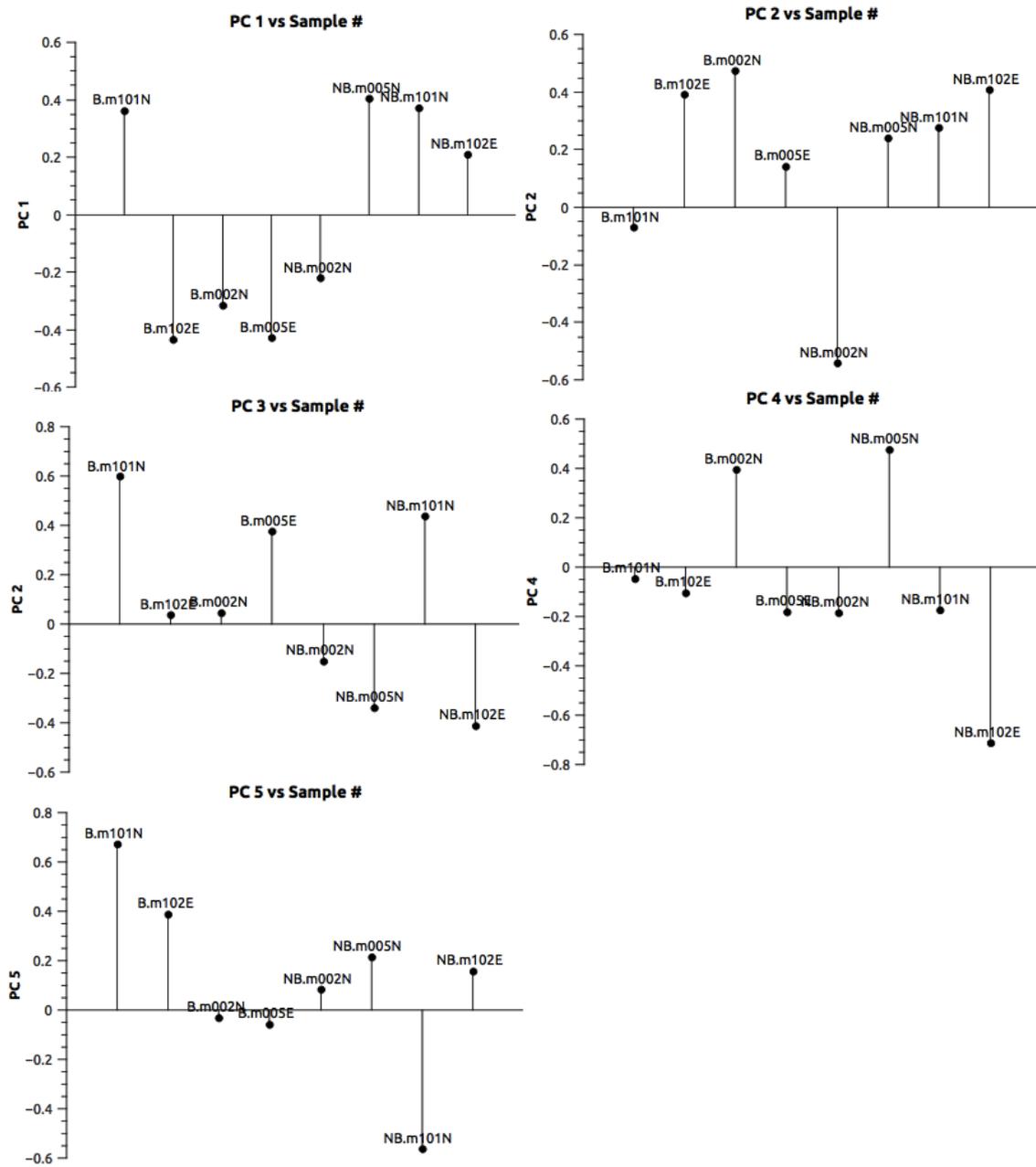


Figure 5.20: PCA scores for p0208m* - Biomass vs. Non-biomass.

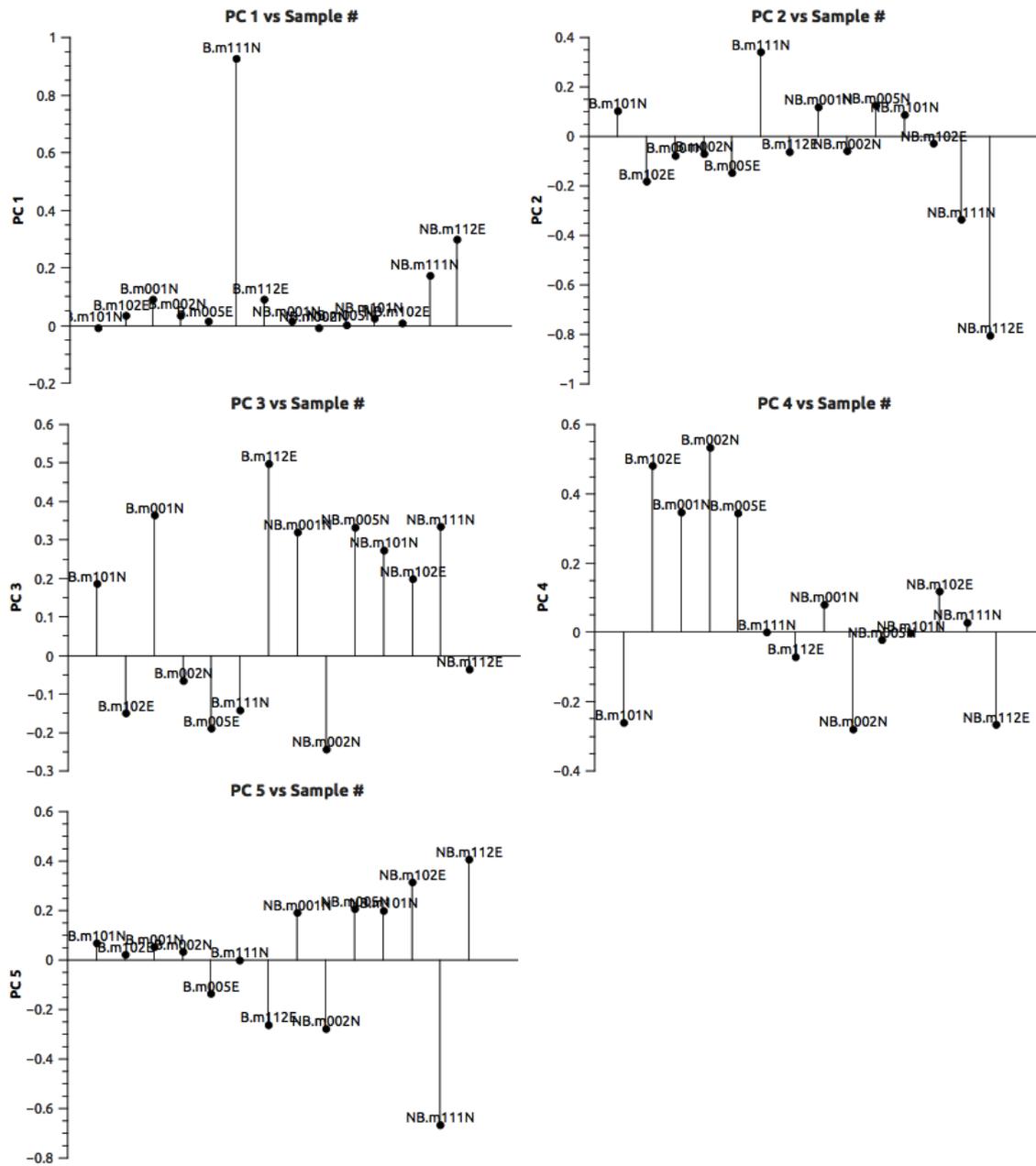


Figure 5.21: PCA scores for p0208m* - Biomass vs. Non-biomass. First 50 masses were excluded from analysis.

visible in the mass spectrum. This was true (secondary ions were detectable till 1280 units) but nevertheless crenarchaeol was not detected (as shown in Figure 5.8). But in the course of this analysis, a deep crater was formed on the wafer surface, thus resembling a perfect sample to analyze what happens during sample erosion.

For this analysis the data set was pre-processed. Mean-centering and normalization with TIC (total ion count) was done and also all secondary ions below a mass limit of 40 u were discarded. The data was partitioned into clusters of 100 single spectra (over all resulting in 10600 clusters) - representing the time scale of about 12 hours. Due to software problems it was only possible to obtain masses below an upper limit of 515 units, this lead to a serious reduction of processing time for the SurfaceLab6 software which was clearly not developed to handle such a big data volume (approximately 6 Gb).

The modified data set should nonetheless represent the whole data set - this opinion is based on two facts:

- (1) The loading plots of prior analyzed samples (shown are only Figure 5.18 and Figure 5.16 but loading plots were done for every data set and loadings above 200 units had seldom influence on the scores of the PC)
- (2) and the literature (e.g. [Thompson *et al.*, 2004]). The scores are mostly influenced by the ions with the highest abundance and change in abundance and therefore this mass cut-off is suitable to decrease the data volume with only a minimal loss on information. Also this subset should be enough to reveal statistical significant changes in the composition of the wafer surface.

First a correlation matrix was calculated between all mass peaks from all 10600 cluster against themselves. The result of that analysis is not shown here, but the conclusion is that the correlation declines as expected. Entry $m_{(1,1)}$ has a Pearson correlation coefficient of 1 whereas $n_{(1,10600)}$ has a coefficient of 0.737.

Next a PCA was performed with the particular clusters treated as individual measure-

ments. Scores for PC 1 (shown in Figure 5.22) and PC 2 (shown in Figure 5.23) change the longer the bismuth cluster beam is bombarding the surface and leads to a sigmoidal curve for the PC 1 and to a parabolic curve for the PC 2. A closer look at the loadings reveals the ions which have most impact on this behaviour in the PC versus time plot. For the PC 1 these are: CaOH^+ , $[64]^+$, $[78]^+$, CH_5SNO^+ , $[93]^+$, $[94]^+$, Ca_2O^+ , $[102]^+$, Ca_2O_2^+ , $\text{Ca}_2\text{O}_2\text{H}^+$, $[132]^+$. The loadings for the PC 2 reveal the following ions as the major player for the changes: C_3H_7^+ , C_4H_7^+ , $\text{C}_3\text{H}_6\text{N}^+$, $[64]^+$, C_5H_7^+ , Ca_2O^+ , Ca_2O_2^+ , $[158]^+$, Ca_2PO_4^+ , $[230]^+$, $[286]^+$. Most of the ions are small organic molecules and alkaline earth metal adducts.

Secondary ions with higher masses than 286 units don't play a role in the behaviour of either the PC1 or the PC2.

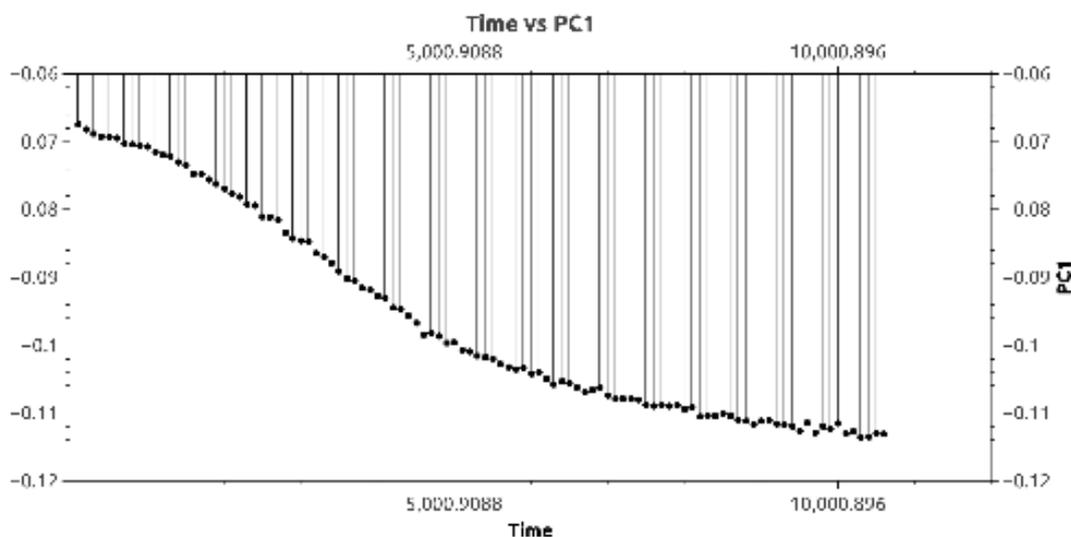


Figure 5.22: PC1 scores vs time - Biomass vs. Non-biomass. First 50 masses were excluded from analysis.

What should be noted is that the slope of the curve declines indicating that the spectra obtained from the last clusters will not change any further for the particular ions represented by this principal component.

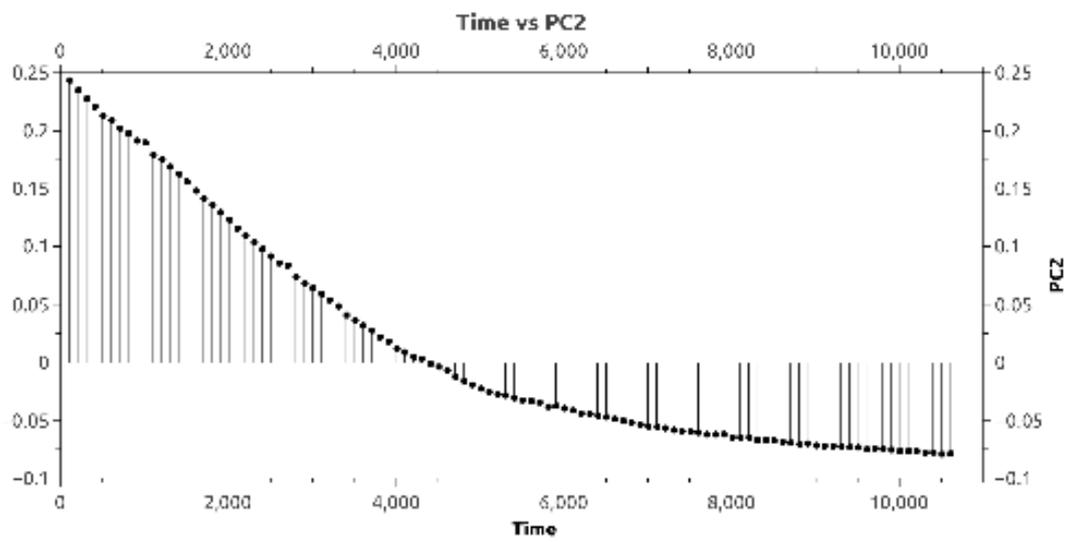


Figure 5.23: PC2 scores vs time - Biomass vs. Non-biomass. First 50 masses were excluded from analysis.

Chapter 6

Conclusion

Analysis of the pure substances

MALDI-MS yielded great success in analyzing 1,2-Di-O-Phytanyl-sn-Glycerol and 1,2-Di-O-Phytanyl-sn-Glycero-3-Phosphoethanolamine. The molecular peak and the Na adducts were identified and fragmentation patterns were analyzed by using the capabilities of the tandem MS. It was not possible to identify the intact mass peak for crenarchaeol or the Na adduct. This was quite unfortunate because therefore fragmentation could not be followed by MS/MS. It was possible to analyze the pure components of all three glycerol diphytanyl ethers with ToF-SIMS (in positive and negative mode) - crenarchaeol included. A comparison of the obtained spectra with the corresponding positive ToF-SIMS spectra of the same substance showed very little matching fragments beside the molecular ion peak for the lipid and the Na adducts. This might be due to the fact that SIMS, in contrast to MALDI, is not optimized to yield longer fragments and during the sputtering process many of them are destroyed. The difference between the MALDI and the ToF-SIMS spectrum might also correspond to the differences in ionization/fragmentation mechanisms. In the literature it was shown that the sodium adducts of lipids yield the most secondary ion intensity, thus favoring the positive ion mode but even in positive mode the intensities for crenarchaeol and fragments are very low seriously challenging their usefulness as biomarker with SIMS with this settings.

This can also indicate that the ionization and ion formation process in ToF-SIMS is not ideal for crenarchaeol. To further verify this different polyatomic cluster ions as primary ion species in combination with chemical yield enhancement (with e.g. Au) should be considered. Also it would be important to investigate the matrix effects and competitive ionization in the matrix for a final statement on this subject.

Moreover, it has to be noted that these findings just apply to the analysis of crenarchaeol without polar head groups, which is a very 'unnatural' situation. It might very well be possible that with the crenarchaeol as it is present in archaea (with polar head groups) analysis with MALDI and SIMS with high secondary ion intensity for crenarchaeol is possible.

Analysis of crenarchaeol in *N. viennensis*

The analysis of *E. coli* and *N. viennensis* showed the importance of sample preparation for ToF-SIMS analysis. It was very often not possible to identify biomass based solely on ToF-SIMS ion yield spectra of organic molecules despite the fact that cells were mounted on a Si-Wafer and the buffer was either carefully washed or vaporised. Without a reference picture of the position of cells identification of biomass with ToF-SIMS spectra is not possible and leads to wrong results. In the cases in which biomass was clearly distinguished from non-biomass with ToF-SIMS spectra this was possible by small organic or inorganic molecules - most likely neither specific to cells nor to biomass but most likely residues from the buffer, cell medium or contamination which had different affinity to inorganic and organic surfaces.

Higher total ion doses showed clearly an improvement in the visualization of biomass. From the test series 0208 onward it became more and more possible to identify biomass. The buffer was changed from a phosphate to an ammonium hydrogen carbonate buffer because this buffer was easier to remove (rinsing vs vaporisation) and therefore a way to decrease organic contamination on the surface and also the cell preparation was changed from fixated (treated with PFA/EtOH) and stained (KI) cells to unstained cells without fixation. This change in sample treatment in combination with higher total ion doses

showed an increase of high molecular mass secondary ions - but crenarchaeol was still not seen. Out of a lack of analysis time there was no comprehensive investigation in the effects of buffer/cell preparation on the secondary ion yield.

The measurement p0416m001 (long-time measurement only with the bismuth gun (Bi_3^+) showed promising results: visualization of cells with high molecular secondary ion molecules was possible.

In this thesis it was shown that in ToF-SIMS imaging mode it was not always possible to distinguish biomass from wafer surface based on specific ions. It stands to discussion in how far contamination render identification of biomass impossible in a ToF-SIMS analysis. Another explanation might be that when the measuring Bi-beam hits the surface it first detects only contamination and when co-sputtering with the DSC is applied to degrade the contaminants the high ion density of the DSC also damages the upcoming surface layers.

A look at other studies showed that most biological samples were either measured as flash-frozen thin sections [Thiel *et al.*, 2007] or freeze-fractured samples [Johansson, 2006] and the method used in this thesis (dried droplet technique) is normally used for low to medium molecular weight ions in combination with multivariate statistical analysis [Thompson *et al.*, 2004, Thompson *et al.*, 2006] making use of silver or carbon polyatomic primary ion beams and the BioToF-SIMS instrument (Ionoptika, UK). The sections and the fractured samples have the big advantage that it is possible to get a direct signal with the primary measurement beam of statistically nearly all levels of the cells.

But besides the technical problem of sample preparation and the problem of biomass/non-biomass differentiation another even more overwhelming challenge awaits: when it was finally possible to identify biomass (based on mostly inorganic molecules) a mass peak list with hundreds of organic molecules appears. Simply based on the deviation introduced through calibration, isotopic abundance and isobars the mass resolution of the TOF-SIMS instrument is not high enough to allow identification of molecules in the high

mass range by calculation from first principles of MS and this problem gets even more severe since the fragmentation can not be followed by a tandem MS/MS set up. This leads to the strange effect that even though molecules with high mass are more significant for a sample they are also much harder to identify simply because the chemical space for molecules with e.g. mass 40 in comparison to mass 400 is much smaller. To bring some light into these peaks the comprehensive use of reference samples is inevitable for unambiguous identification of molecular fragments and molecular ions, otherwise most of the assigned molecules are at risk of being not distinct.

It was possible to distinguish between *N. viennensis* and *E. coli* using a PCA. The loadings show that the differences are mainly small inorganic molecules - most likely a result of different cell cultivation and buffers or adsorption effects. These findings can not be interpreted based solely on the results of this thesis, systematic and well planned analysis has to follow to understand the problem we were facing here. The same goes for the fact that it was basically not possible to distinguish between biomass and non-biomass in some PCA analysis especially if the results were not preprocessed. A finding that contradicts the primary assumption made in every ToF-SIMS analysis of biomass: that biomass is verifiably different from non-biomass (in the case of this study the wafer it is mounted on). This findings are very hard to interpret. Maybe this was due to the fact that all spectra were obtained with settings best suited for finding crenarchaeol. Very often the aim was not to find many cells but the aim was to identify one single cell and to screen for the lipid. What also might have played a role was the non-supervised selection of peaks used to calculate the PCA. A mass threshold was usually introduced (only masses above a certain molecular weight) but compared to other studies and the PCAs performed (below 100 peaks selected) this was not very strict. But this are no conclusive explanations, further studies have to give an answer here.

Bibliography

- [Barshick *et al.*, 2000] BARSHICK, C., DUCKWORTH, D., & SMITH, D. 2000. *Inorganic Mass Spectrometry: Fundamentals and Applications*. Practical Spectroscopy. Taylor & Francis.
- [Belu *et al.*, 2003] BELU, ANNA M, GRAHAM, DANIEL J, & CASTNER, DAVID G. 2003. Time-of-flight secondary ion mass spectrometry: techniques and applications for the characterization of biomaterial surfaces. *Biomaterials*, **24**(21), 3635–53.
- [Benabdellah *et al.*, 2010] BENABDELLAH, FARIDA, SEYER, ALEXANDRE, QUINTON, LOÏC, TOUBOUL, DAVID, BRUNELLE, ALAIN, & LAPRÉVOTE, OLIVIER. 2010. Mass spectrometry imaging of rat brain sections: nanomolar sensitivity with MALDI versus nanometer resolution by TOF–SIMS. *Anal Bioanal Chem*, **396**(1), 151–162.
- [Benninghoven *et al.*, 1987] BENNINGHOVEN, A., RÜDENAUER, F. G., & WERNER, H.W. 1987. *Secondary Ion Mass Spectrometry: Basic Concepts, Instrumental Aspects, Applications, and Trends*. Wiley.
- [Berg *et al.*, 2008] BERG, J.M., TYMOCZKO, J.L., & STRYER, L. 2008. *Biochemistry*. W. H. Freeman.
- [Boucher *et al.*, 2004] BOUCHER, Y, KAMEKURA, M, & DOOLITTLE, W.F. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Molecular Microbiology*, **52**(2), 515–527.
- [Brochier-Armanet *et al.*, 2008] BROCHIER-ARMANET, CÉLINE, BOUSSAU, BASTIEN, GRIBALDO, SIMONETTA, & FORTERRE, PATRICK. 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Micro*, **6**(3), 245–52.
- [Carrol & Beavis, 1998] CARROL, J A, & BEAVIS, R C. 1998. 9. Matrix-Assisted Laser Desorption and Ionization. *Experimental Methods in the Physical Sciences*, **30**, 413–448.
- [Cheng *et al.*, 1998] CHENG, C, GROSS, M L, & PITTENAUER, E. 1998. Complete structural elucidation of triacylglycerols by tandem sector mass spectrometry. *Anal Chem*, **70**(20), 4417–26.

- [Chong, 2010] CHONG, PARKSON LEE-GAU. 2010. Archaeobacterial bipolar tetraether lipids: Physico-chemical and membrane properties. *Chemistry and Physics of Lipids*, **163**(3), 253–265.
- [Christie, 1998] CHRISTIE, W.W. 1998. Gas Chromatography–Mass Spectrometry Methods for Structural Analysis of Fatty Acids. *Lipids*, **33**(4), 343–353.
- [Damsté *et al.*, 2002] DAMSTÉ, JAAP S SINNINGHE, SCHOUTEN, STEFAN, HOPMANS, ELLEN C, VAN DUIN, ADRI C T, & GEENEVASEN, JAN A J. 2002. Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. *J Lipid Res*, **43**(10), 1641–51.
- [DasSarma *et al.*, 2009] DASSARMA, S., COKER, J.A., & DASSARMA, P. 2009. Archaea (overview). *Pages 1 – 23 of:* IN CHIEF: Å Å MOSELIO SCHAECHTER, EDITOR (ed), *Encyclopedia of Microbiology (Third Edition)*, third edition edn. Oxford: Academic Press.
- [Eguchi *et al.*, 2000] EGUCHI, T, TAKYO, H, MORITA, M, KAKINUMA, K, & KOGA, Y. 2000. Unusual double-bond migration as a plausible key reaction in the biosynthesis of the isoprenoidal membrane lipids of methanogenic archaea. Electronic supplementary information (ESI) available: ¹H NMR and ¹H–¹H COSY spectra. See <http://www.rsc.org/suppdata/cc/b0/b003948i>. *Chem. Commun.*, 1545–1546.
- [Francis *et al.*, 2008] FRANCIS, J.T, NIE, H.Y, TAYLOR, A.R, WALZAK, M.J, CHANG, W.H, MACFABE, D.F, & LAU, W.M. 2008. ToF-SIMS cluster ion imaging of hippocampal CA1 pyramidal rat neurons. *Applied surface science*, **255**, 1126–1130.
- [Galliker, 1990] GALLIKER, PETER KASPAR. 1990. *Zur Biosynthese der Etherlipide aus Methanobacterium thermoautotrophicum*. Ph.D. thesis, ETH Zürich.
- [Griesser *et al.*, 2004] GRIESSER, H J HANS J, KINGSHOTT, PETER, MCARTHUR, S L SALLY L, MCLEAN, K M KEITH M, KINSEL, G R GARY R, & TIMMONS, R B RICHARD B. 2004. Surface-MALDI mass spectrometry in biomaterials research. *Biomaterials*, **25**(20), 4861–75.
- [Heim *et al.*, 2009] HEIM, CHRISTINE, SJÖVALL, PETER, LAUSMAA, JUKKA, LEEFMANN, TIM, & THIEL, VOLKER. 2009. Spectral characterisation of eight glycerolipids and their detection in natural samples using time-of-flight secondary ion mass spectrometry. *Rapid Commun Mass Spectrom*, **23**(17), 2741–53.
- [Herzog & Viehböck, 1949] HERZOG, R. F. K., & VIEHBÖCK, F. P. 1949. Ion Source for Mass Spectrography. *Physical Review*, **6**(76), 855–856.
- [Johansson, 2006] JOHANSSON, B. 2006. ToF-²SIMS imaging of lipids in cell membranes. *Surf. Interface Anal.*, **38**(11), 1401–1412.
-

- [Jones *et al.*, 2006] JONES, EMRYS A., FLETCHER, JOHN S., THOMPSON, CHARLOTTE E., JACKSON, DEAN A., & LOCKYER, NICHOLAS P. 2006. ToF-SIMS analysis of bio-systems: Are polyatomic primary ions the solution? *Applied Surface Science*, 6844–6854.
- [Karas & Hillenkamp, 1988] KARAS, M, & HILLENKAMP, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, **60**(20), 2299–301.
- [Koga & Morii, 2007] KOGA, Y, & MORII, H. 2007. Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiology and Molecular Biology Reviews*, **71**(1), 97–120.
- [Kollmer, 2004] KOLLMER, F. 2004. Cluster primary ion bombardment of organic materials. *Applied surface science*, **231**(Jun), 153–158.
- [Koolman & Rohm, 2004] KOOLMAN, J., & ROHM, K. 2004. *Color Atlas of Biochemistry*. Thieme Flexibook. Thieme.
- [Koonin, 2005] KOONIN, EUGENE V. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.*, **39**(Jan), 309–38.
- [Mahoney, 2010] MAHONEY, CHRISTINE M. 2010. Cluster secondary ion mass spectrometry of polymers and related materials. *Mass Spectrom. Rev.*, **29**(2), 247–93.
- [Matsumi *et al.*, 2011] MATSUMI, R, ATOMI, H, DRIESSEN, A.J.M, & OOST, J VAN DER. 2011. Isoprenoid biosynthesis in Archaea—biochemical and evolutionary implications. *Research in Microbiology*, **162**(1), 39–52.
- [McIntyre *et al.*, 2008] MCINTYRE, THOMAS M., SNYDER, F., & MARATHE, G. K. 2008. *Biochemistry of Lipids, Lipoproteins and Membranes*. 5 edn. Elsevier. Chap. Ether-linked lipids and their bioactive species.
- [McLain, 2005] MCLAIN, J.E.T. 2005. ARCHAEA. *Pages 88 – 94 of: IN CHIEF: DANIEL HILLEL, EDITOR (ed), Encyclopedia of Soils in the Environment*. Oxford: Elsevier.
- [McMahon *et al.*, 1996] MCMAHON, J. M., SHOR, R. T., MCCANDLISH, C.A., & BRENNAN, P.J. 1996. Identification and Mapping of Phosphocholine in Animal Tissue by Static Secondary Ion Mass Spectrometry and Tandem Mass Spectrometry Identification and Mapping of Phosphocholine in Animal Tissue by Static Secondary Ion Mass Spectrometry and Tandem Mass Spectrometry Identification and Mapping of Phosphocholine in Animal Tissue by Static Secondary Ion Mass Spectrometry and Tandem Mass Spectrometry. *Rapid Commun Mass Spectrom*, **10**(3), 335–340.
- [Mikkelsen & Cortón, 2004] MIKKELSEN, S.R., & CORTÓN, E. 2004. *Bioanalytical Chemistry*. Wiley.
-

- [Montaudo & Lattimer, 2001] MONTAUDO, G., & LATTIMER, R.P. 2001. *Mass Spectrometry of Polymers*. Taylor & Francis.
- [Murakami *et al.*, 2007] MURAKAMI, M, SHIBUYA, K, NAKAYAMA, T, NISHINO, T, YOSHIMURA, T, & HEMMI, H. 2007. Geranylgeranyl reductase involved in the biosynthesis of archaeal membrane lipids in the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *FEBS Journal*, **274**(3), 805–814.
- [Nagy & Walker, 2007] NAGY, G, & WALKER, A.V. 2007. Enhanced secondary ion emission with a bismuth cluster ion source. *International Journal of Mass Spectrometry*, **262**(Feb), 144–153.
- [Nishimura & Eguchi, 2006] NISHIMURA, Y, & EGUCHI, T. 2006. Biosynthesis of archaeal membrane lipids: digeranylgeranyl glycerophospholipid reductase of the thermoacidophilic archaeon *Thermoplasma acidophilum*. *Journal of Biochemistry*, **139**(6), 1073–1081.
- [Pester *et al.*, 2011] PESTER, MICHAEL, SCHLEPER, CHRISTA, & WAGNER, MICHAEL. 2011. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol*, **14**(3), 300–6.
- [Pitcher *et al.*, 2009] PITCHER, A, RYCHLIK, N, HOPMANS, E.C, SPIECK, E, RIJPS-TRA, W.I.C, OSSEBAAR, J, SCHOUTEN, S, WAGNER, M, & DAMSTÉ, J.S.S. 2009. Crenarchaeol dominates the membrane lipids of *Candidatus Nitrososphaera gargensis*, a thermophilic Group I. 1b Archaeon. *The ISME Journal*, **4**(4), 542–552.
- [Pittenauer & Allmaier, 2009] PITTENAUER, ERNST, & ALLMAIER, GÜNTER. 2009. The renaissance of high-energy CID for structural elucidation of complex lipids: MALDI-TOF/RTOF-MS of alkali cationized triacylglycerols. *J Am Soc Mass Spectrom*, **20**(6), 1037–47.
- [Postawa *et al.*, 2004] POSTAWA, Z, CZERWINSKI, B, SZEWCZYK, M, SMILEY, E.J, WINOGRAD, N, & GARRISON, B.J. 2004. Microscopic Insights into the Sputtering of Ag111 Induced by C60 and Ga Bombardment. *J. Phys. Chem.*, **108**(Oct), 7831–7838.
- [Powell *et al.*, 2012] POWELL, SEAN, SZKLARCZYK, DAMIAN, TRACHANA, KALLIOPI, ROTH, ALEXANDER, KUHN, MICHAEL, MULLER, JEAN, ARNOLD, ROLAND, RATT-TEI, THOMAS, LETUNIC, IVICA, DOERKS, TOBIAS, JENSEN, LARS J, VON MERING, CHRISTIAN, & BORK, PEER. 2012. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Research*, **40**(Database issue), D284–9.
- [Remm *et al.*, 2001] REMM, M, STORM, C E, & SONNHAMMER, E L. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol*, **314**(5), 1041–52.
-

- [Sato *et al.*, 2008] SATO, S, MURAKAMI, M, YOSHIMURA, T, & HEMMI, H. 2008. Specific partial reduction of geranylgeranyl diphosphate by an enzyme from the thermoacidophilic archaeon *Sulfolobus acidocaldarius* yields a reactive prenyl donor, not a dead-end product. *Journal of bacteriology*, **190**(11), 3923.
- [Schleper *et al.*, 2005] SCHLEPER, CHRISTA, JURGENS, GERMAN, & JONUSCHEIT, MELANIE. 2005. Genomic studies of uncultivated archaea. *Nat Rev Micro*, **3**(6), 479–88.
- [Seedorf *et al.*, 1995] SEEDORF, U, FOBKER, M, VOSS, R, MEYER, K, KANNENBERG, F, MESCHEDE, D, ULLRICH, K, HORST, J, BENNINGHOVEN, A, & ASSMANN, G. 1995. Smith-Lemli-Opitz syndrome diagnosed by using time-of-flight secondary-ion mass spectrometry. *Clin Chem*, **41**(4), 548–52.
- [Shard *et al.*, 1997] SHARD, A.G, DAVIES, M.C, TENDER, S.J.B, BENNETTI, L, & PURBRICK, M.D. 1997. X-ray Photoelectron Spectroscopy and Time-of-Flight SIMS Investigations of Hyaluronic Acid Derivatives. *Langmuir*, Aug, 2808–2814.
- [Sigmund, 1969] SIGMUND, P. 1969. Theory of sputtering. I. Sputtering yield of amorphous and polycrystalline targets. *Physical Review*, **184**, 383.
- [Spang *et al.*, 2010] SPANG, ANJA, HATZENPICHLER, ROLAND, BROCHIER-ARMANET, CÉLINE, RATTEI, THOMAS, TISCHLER, PATRICK, SPIECK, EVA, STREIT, WOLFGANG, STAHL, DAVID A, WAGNER, MICHAEL, & SCHLEPER, CHRISTA. 2010. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends in Microbiology*, **18**(8), 331–40.
- [Sparvero *et al.*, 2012] SPARVERO, LOUIS J, AMOSCATO, ANDREW A, DIXON, C EDWARD, LONG, JOSEPH B, KOCHANEK, PATRICK M, PITT, BRUCE R, BAYŃR, HÜLYA, & KAGAN, VALERIAN E. 2012. Mapping of phospholipids by MALDI imaging (MALDI-MSI): realities and expectations. *Chemistry and Physics of Lipids*, **165**(5), 545–562.
- [Stapel & Benninghoven, 2001] STAPEL, D, & BENNINGHOVEN, A. 2001. Application of atomic and molecular primary ions for TOF-SIMS analysis of additive containing polymer surfaces. *Applied Surface Science*, **174**(Apr), 261–270.
- [Stephan, 2001] STEPHAN, T. 2001. TOF-SIMS in Cosmochemistry. *Planetary and Space Science*, **47**, 859–906.
- [Tatusov *et al.*, 1997] TATUSOV, R L, KOONIN, E V, & LIPMAN, D J. 1997. A genomic perspective on protein families. *Science*, **278**(5338), 631–7.
- [Thiel *et al.*, 2007] THIEL, V, HEIM, C, ARP, G, HAHMANN, U, SJÖVALL, P, & LAUSMAA, J. 2007. Biomarkers at the microscopic range: ToF-SIMS molecular imaging of Archaea-derived lipids in a microbial mat. *Geobiology*, **5**(4), 413–421.
-

- [Thompson *et al.*, 2006] THOMPSON, C. E, ELLIS, J, FLETCHER, J. S, GOODACRE, R, HENDERSON, A, LOCKYER, N. P, & VICKERMAN, J. C. 2006. ToF-SIMS studies of *Bacillus* using multivariate analysis with possible identification and taxonomic applications. *Applied surface science*, **252**(19), 6719–6722.
- [Thompson *et al.*, 2004] THOMPSON, CE, JUNGNICHEL, H, LOCKYER, NP, STEPHENS, GM, & VICKERMAN, JC. 2004. ToF-SIMS studies as a tool to discriminate between spores and vegetative cells of bacteria. *Applied surface science*, **231**(Jan), 420–423.
- [Touboul *et al.*, 2005] TOUBOUL, DAVID, KOLLMER, FELIX, & NIEHUIS, EDWALD. 2005. Improvement of biological time-of-flight-secondary ion mass spectrometry imaging with a bismuth cluster ion source Improvement of biological time-of-flight-secondary ion mass spectrometry imaging with a bismuth cluster ion source Improvement of biological time-of-flight-secondary ion mass spectrometry imaging with a bismuth cluster ion source. *Journal of the American Society for Mass Spectrometry*, **10**(16), 1608–1618.
- [Tourna *et al.*, 2011] TOURNA, MARIA, STIEGLMEIER, MICHAELA, SPANG, ANJA, KÖNNEKE, MARTIN, SCHINTLMEISTER, ARNO, URICH, TIM, ENGEL, MARION, SCHLOTTER, MICHAEL, WAGNER, MICHAEL, RICHTER, ANDREAS, & SCHLEPER, CHRISTA. 2011. *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *Proceedings of the National Academy of Sciences*, **108**(20), 8420–5.
- [Vickerman & Briggs, 2001] VICKERMAN, J.C., & BRIGGS, D. 2001. *ToF-SIMS: surface analysis by mass spectrometry*. IM.
- [Weijers *et al.*, 2006] WEIJERS, JOHAN W H, SCHOUTEN, STEFAN, HOPMANS, ELLEN C, GEENEVASEN, JAN A J, DAVID, OLIVIER R P, COLEMAN, JOANNA M, PANCOST, RICH D, & DAMSTÉ, JAAP S SINNINGHE. 2006. Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environ Microbiol*, **8**(4), 648–57.
- [Woese *et al.*, 1990] WOESE, C R, KANDLER, O, & WHEELIS, M L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA*, **87**(12), 4576–9.
-

EDUCATION AND TRAINING

Dates 1992-1996
Name and type of organisation Volksschule Henndorf
providing education and training

Dates 1996-2000
Name and type of organisation Hauptschule Henndorf
providing education and training

Dates 2000-2005
Name and type of organisation Bundesoberstufen Realgymnasium mit
providing education and training naturwissenschaftlichem Schwerpunkt
Level in national or international Matura
classification

Dates October 2005 - October 2006
Civilian service done and finished working for Arbeiter-
Samariter Bund Salzburg

Dates October 2006 - March 2010
Field of study Bachelorstudium Genetik
Name and type of organisation University Salzburg
providing education and training
Title gained Bachelor rerum naturalium

Dates	Since March 2010 to March 2014
Field of study	Masterstudium Biologische Chemie
Title of Master thesis	Spectral characterisation of different glycerol diphytanyl ethers as pure substances and part of the cell membrane of archaea using time-of-flight secondary ion mass spectrometry (ToF-SIMS) and matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-MS)
Supervised and grades by	o. Prof. Dr. Michael Wagner
Department/s the Master thesis was done in	Department of microbial ecology Fakultät für Lebenswissenschaften Universität Vienna Althanstr. 14 A-1090 Wien Institute of Chemical Technologies and Analytics (ICTA) Technical University of Vienna Getreidemarkt 9/164 A-1060 Vienna
Name and type of organisation providing education and training	University Vienna

Dates	From March 2010 to November 2013
Field of study	Masterstudium Molekulare Biologie (Schwerpunkt Molekulare Medizin)
Title of Master thesis	Bacterial effector proteins in the evolu- tion of pathogenic and symbiotic bac- teria
Submitted and graded	November 2013
Grade	Sehr gut (1)
Supervised and grades by	Prof. Dr. Thomas Rattei
Department/s the Master thesis was done in	Division of Computational Systems Biology Fakultät für Lebenswissenschaften Universität Vienna Althanstr. 14 A-1090 Vienna
Name and type of organisation providing education and training	University Vienna
