universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Dynamic exo- and endoproteome analysis of *Deinococcus radiodurans*"

verfasst von / submitted by

Kaan Georg Kutlucinar BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 863
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Biologische Chemie
Betreut von / Supervisor:	Univ.-Prof. Dr. Wolfram Weckwerth
Mitbetreut von / Co-Supervisor:	Dr. rer. nat. Tetyana Milojevic

Acknowledgment

First of all, I would like to thank my advisor Prof. Wolfram Weckwerth for accepting me in his laboratory and for the opportunity to work in proteomics field.

I also thank Prof. Christa Schleper for allowing me to work in her laboratory and to culture *Deinococcus radiodurans*.

Dr. Palak Chaturvedi, Ella Katarina Nukarinen, Nima Ranjbar, Dr. Bernhard Wurziger and Dr. Melina Kerou are thankfully acknowledged for their advice on experimental techniques.

Special thanks go to Romana Bittner, Lena Fragner and Sonja Tischler for their patient answers to all of my questions, and for their assistance with the mass spectrometers.

I would like to express my gratitude to all members of the department of ecogenomics and systems biology. Their friendship and support created a perfect working environment and helped me to overcome difficulties.

I like to thank my close friends Saren, Mate and Enis for their encouragement and support in my study. I also like to thank Katja, Marie, Öykü, Walpurga, Klaudia, Simone, Timo and Gregor for all the good times we spent together during our studies.

Lastly I would like to express my deepest gratitude to my family.

Abstract (in English)

Deinococcus radiodurans is a gram positive polyextremophylic bacteria belonging to the Deinococcus genus. Its polyextremophylic features like UV-resistance, ionising radiation resistance, oxidising agent resistance and desiccation resistance makes *D. radiodurance* an interesting model organism for investigation of the possible transfer of microbes through space. For this investigation the bacteria is exposed to space conditions in the Japan experimental module (JEM) of the international space station (ISS)[1–3].

Sensitive and quantitative analysis of proteins using high throughput mass spectrometric analysis methods can help us to understand the role of proteins in biological systems[4]. To elucidate important proteins of *Deinococcus radiodurans*, which play an important role under different environmental or artificial stresses, we have tested different sample preparation methods and found that the MCW-Phenol method resulted in the best proteome coverage of all tested methods. MCW-Phenol method did not only result in the best proteome coverage it also made an integrative metabolomic analysis possible.

Abstract (in German)

Deinococcus radiodurans ist eine polyextremophiles gram positives Bakterium, das zu Deinococcus Genus gehört. Die polyextremophilen Eigenschaften, wie UV-Resistenz, ionisierende Strahlung Resistenz, oxidationsmitteln Resistenz und Trocknungsresistenz macht *Deinococcus radiodurans* ein interessanter Modellorganismus für die Untersuchung von möglichen Übertragungen von Mikroben durch den Weltraum. Für diese Untersuchung wurde dieses Bakterium unter Weltraumbedingungen in „Japan experimental module" (JEM) an der Internationale Raumstation (ISS) ausgesetzt[1–3].

Sensitive und quantitative Analysen von Proteinen mit Hilfe von massenspektrometrischen Methoden kann uns helfen die Rolle von Proteinen in biologischen Systemen zu verstehen[4]. Um die wichtige Proteine von *Deinococcus radiodurans*, die unter verschiedenen umweltbedingten oder künstlichen Stress eine Rolle spielen aufzuklären, haben wir verschiedene Probenvorbereitungsmethoden getestet und wir haben herausgefunden, dass unter getesteten Methoden die MCW-Phenol Methode die größte Deckung von Proteome ergeben hat. MCW-Phenol Methode hat nicht nur mit der größten Deckung vom Proteome ergeben, sondern hat auch die integrative metabolische Analyse ermöglicht.

Table of Contents

1 Introduction

1.1 Extremophiles

Scientists have found organisms in physically and/or geochemically extreme conditions. It has been found that some of these organisms not only tolerate these extreme conditions, but they also require them[5]. Organisms that can thrive in physically or geochemically extreme conditions such as high or low temperatures, high pressure, desiccation, radiation, high or low pH etc., are called extremophiles[5,6]. If they can thrive more than one extreme condition they are called polyextremophiles.[6] Extremophiles can be found in all domains of life. They can be found in ecosystems such as hot springs, geysers, deep sea, hypersaline environments, deserts, atmosphere, ice, permafrost and snow[6] (Figure 1).

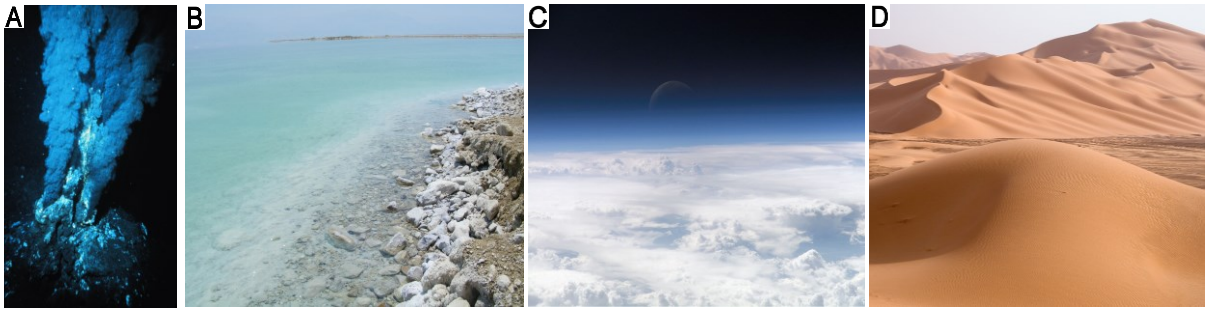

Figure 1: Extreme environments: A) Hydrothermal vent, B) High salinity of Dead Sea, C) Atmosphere and Universe, D) Desert (pictures taken from Wikipedia)

To survive in extreme environments extremophiles have developed different mechanisms. Depending on the extreme condition extremophiles protect themselves by keeping the extreme condition out of the cell, by extruding the harmful substance out of the cell, by changing their physiology or by improving repair mechanisms[6].

For example, thermophiles and hyperthermophiles acquired their ability to resist high temperatures by altering their physiology. To maintain a favorable membrane fluidity their membrane lipids contain more saturated fatty acids compared to mesophiles[7]. To prevent protein denaturation, extremophile proteins posses an increased number of large hydrophobic residues, disulfide bonds, and ionic interactions[8].

Another important biomolecule, which should be protected from high temperatures is the DNA. DNA is protected from denaturation and chemical modification by increasing its stability through the presence of monovalent and divalent salts, which screen the negative charges of the phosphate groups[6,7].

The obtained data from the studies of extremophiles contributed to basic molecular biology and evolutionary biology[6]. Additionally, it also raised the possibility for finding life on other planets and moons in our solar system. Findings on extremophiles supported the panspermia hypothesis, which states that life can migrate through space from one location to another[7,9].
Extremophiles and enzymes from extremophiles (extremozymes)[10] are tools which have a wide use in biotechnology[6,11]. One of the best known extremozymes is a thermostable DNA polymerase isolated from thermophilic bacterium *Thermus aquaticus* by Alice Chien[12], often abbreviated as Taq Polymerase. Taq Polymerase had big impact on biotechnology by enabling polymerase chain reaction (PCR) at higher temperatures[6,13]. Other important fields where extremophiles and/or extremozymes are used are biofuel production, biomining, laundry detergents, food industry, and medical applications. The wide spectrum of applicability in different industrial fields was a commercial success[6,11].

1.2 Deinococcus radiodurans

Deinococcus radiodurans (hereafter referred as *D. radiodurans*) was first isolated in 1956 from a gamma-irradiated canned meat[14]. It was first named *Micrococcus radiodurans,* but 16S rRNA analysis proved that it is a unique phylogenetic group of bacteria. It was then placed in its own genus *Deinococcus and* renamed *Deinococcus radiodurans.* The word Deinococcus derives from Ancient Greek. Deinos means strange or unusual and coccus means grain or berry (strange berry). Radiodurans derives from Latin where radius and durare stands together for radiation surviving[1].
D. radiodurans is a Gram positive, red pigmented, nonsporulating, nonpathogenic bacterium, with a cell envelope, with a multilayer structure and lipid composition similar to Gram negative bacteria[1]. It is obligatory aerobic chemoorganotroph[15] and has a resistance to desiccation, UV radiation, ionizing radiation and oxidizing agents[1,15–17]. *D.* radiodurans can tolerate up to 5000 Gy without loss of viability[18]. The multi-resistance against radiation desiccation toxic chemicals of *D. radiodurans* seems to be evolved as a polyvalent strategy, instead of parallel strategies for every single resistance[19].

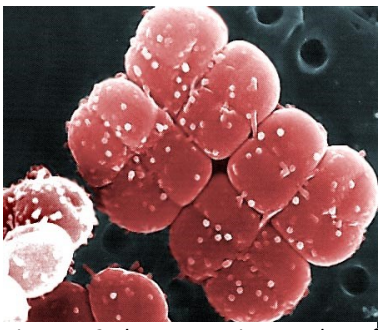

Figure 2:Electron micrograph of Deinococcus radiodurans. Image has been downloaded from webpage: http://www.scifun.ed.ac.uk/card/facts.html

D. radiodurans cells with a radius between 0.5 and 3.5 µm form diads and tetrads (Figure 2) by dividing alternatively in two planes[1,15]. Cells can grow between 4°C and 45°C and have an optimum growth temperature of 30°C[15]. The genome

of *D. radiodurans* consists of two chromosomes and two plasmids containing 3,187 open reading frames (ORFs). Depending on their culture medium and growth stage cells, have different copy number of their genome. In the exponential phase cells can contain up to ten copies of their genome[1,20].

Since the Earth does not have natural sources of radiation which produce doses comparable to the resistance capabilities of *D. radiodurans*, it was an open question how this feature evolved in *D. radiodurans*[19]. One hypothesis is that radiation resistance is a consequence of desiccation resistance[21]. This theory was supported by finding that other desiccation resistant organisms are also resistant to radiation[19].

There are multiple factors contributing to radiation resistance of an organism. One of them is the efficiency of the DNA repair mechanism. The mechanism which *D.radiodurans* uses is called extended synthesis-dependent strand annealing[1] (ESDSA). In comparison to synthesis-dependent strand annealing (SDSA) ESDSA is able to repair hundred of times more DNA double-strand breaks (DSBs)[19]. *D. radiodurans* uses also enzmatic and non enzmatic mechanisms to keep the reactive oxygen species (ROS) at low concentrations. It has an increased catalase and superoxide dismutase activity and different antioxidant molecules, manganese complexes to maintain a low ROS-level[1,19,22]. *D. radiodurans* has different metabolic properties, which contributes also keeping the ROS level low. Some of these mechanisms involve the conversion of glucose into precursors for dNTPs, proteolysis and import of exogenous peptides and amino acids using 90 ABC transporters[1,19,23].

Ionizing radiation and UV radiation cause proteome oxidation, which is measured by analyzing the protein carbonylation (PC) and alters the function of the biosynthetic machinery. Results of Krisko and Radman 2010 are indicating a relationship between PC and cell death[24]. These finding underline the importance of the antioxidant protection of the proteins beside the protection of the DNA. The importance of proteome protection is more vital than genome protection, because only an active proteome can repair damaged cell functions including the genome[19].

Escherichia coli and *Saccharomyces cerevisiae* are most popular microorganisms engineered for biotechnological applications. Interestingly, studies with several species of *Deinococcus* have shown that they could be an attractive alternative. The most important feature, what makes several *Deinococcus* species interesting for biotechnology is their chromosomal modifiability, which allows gene expression without maintaining a plasmid. This feature is than followed by the ability of growing on minimal media and the ability of metabolizing cheaper carbon sources (e.g. lignocellulosic biomass)[16,25]. Beside their capability of being a

host bacteria for industrial biotechnology, their polyextremophylic features especially UV radiation and ionizing radiation resistance favors them for application in bioremediation of organic molecules or heavy metals in radiation polluted areas[16,26].

1.3 Tanpopo Mission

Tanpopo (aka dandelion) mission is an astrobiology project by Japanese scientist to investigate the possible transfer of microbes and organic compounds by capturing microbes and micrometeoroids at the altitude of International Space Station (ISS) using the Japan Experimental Module (JEM) of the ISS. Another focus of Tanpopo mission are exposure experiments, where dehydrated microbes are exposed to space conditions. Following to the long term exposure experiment on ISS viability of the cells will be investigated[2,3].

1.4 Mass spectrometry, comparative proteomics and functional data interpretation

All the proteins expressed by a cell are known as the proteome of a cell[27]. The proteome of a cell changes when the cell or organism undergoes different stresses. Proteomics is the discipline, which uses high-throughput methods for elucidation of quantitative changes in protein expression between stressed and unstressed biological samples at a given time[28]. Beside comparative expression analysis of the expressed proteins, proteomics investigates also posttranslational modifications and protein-protein interactions[29].

Protein microarray and mass spectrometry are two relevant high-throughput technologies which are crucial to proteomics studies. Protein microarrays are used for high-throughput analysis of protein-protein, protein-nucleic acid, and protein molecule interactions[29].

Technological advances in the past decades made mass spectrometers a preferred tool for proteomics studies[30]. To identify a sample mass, spectrometers ionize probes in the gas phase and measure the mass charge ratio (m/z) of single constituent components. Every mass spectrometer has minimum three basic components: an ion source, a mass analyzer and a detector. The application of mass spectrometers in proteomics were possible by the development of the soft ionization techniques[31]. The development of soft ionization techniques Matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) enabled ionization of peptides and intact proteins without fragmentation[32].

Mass spectrometry based proteomics uses two different strategies: Top down and bottom up. Each strategy has its own advantages and disadvantages. In top down strategy intact proteins are analyzed. However, due to the chemical complexity of intact proteins, the separation of a

complex mixture is difficult. This property narrowed the top down proteomics research mostly in single or small number of proteins. Nevertheless, top down proteomics has an advantage in identification and localization of posttranslational modifications (PTMs). In the bottom up strategy, extracted proteins are digested into peptides and are separated by using one or two dimensional liquid chromatography before analyzing with mass spectrometer. The bottom up strategy is a high-throughput method for protein identification and quantification, while the limited sequence coverage decreases the identification of PTMs[31,33].

After high-throughput measurement with mass spectrometry, the obtained large amount of raw data requires computational and statistical analysis for correct interpretation. In bottom up proteomics, different methods can be used to interpret fragment ion spectra. One of them is the database searching technique, which compares data from an in silico digested database with the acquired fragment ion spectra. There are different computational tools which involves different algorithms to produce a list of best matching peptides and to score them. According to the score of the matching peptides the potential candidates can be interpreted as true identification[31].

Proteomics answers questions in basic biological research, clinical studies, and toxicology, either by quantitatively monitoring the protein expression in a cell or tissue or by structural proteomics which investigates the physical interaction of proteins[34]. These studies analyzes signaling pathways and develop protein biomarker assay systems[35].

2 Aims of this Study

The aim of this study was to compare different sample preparation methods for proteomics analysis of polyextremophile microorganism *Deinococcus radiodurans* and find the strengths and weaknesses of these different methods. The optimization of sample preparation and reproducibility of proteome analysis is necessary to reveal proteomic changes in *Deinococcus sp.* in response to environmental stresses and to cope with small sample amounts from astrobiology experiments such as the Tanpopo mission. Furthermore, an integrative extraction method was tested to reveal the possibility for the simultaneous analysis of metabolites and proteins from the same sample. The reasoning for this was to cope with small sample amounts. Accordingly, primary metabolites were analyzed using an integrative metabolite and protein extraction method using GC-MS.

3 Materials and Methods

3.1 Strain and Cell Condition

The strain was obtained from Tanpopo space mission/Tokyo University. *Deinococcus radiodurans R1* was grown at 30°C in mTGE medium (1 % bacto trypton, 0.6 % beef extract, 0.2 % glucose) in an incubator with shaking at 150 rpm until an OD600 value between 2.150 and 2.350 was achieved (approx... 16 h).

3.2 Sample Preparation

D. radiodurans cells were collected from their liquid cultures by centrifugation at 1520 g for 15 min at 4°C. These cells were washed three times with 10 mM potassium phosphate buffer pH 7.0.

3.2.1 Metabolite extraction and analysis

Dry weight of the washed samples was established after lyophilisation. Dried cells were resuspended in cold 1 mL MCW-mix (Methanol/Chloroform/Water mixture 2,5:1:0,5 (v:v:v) and transferred into bead tubes with 0.5 g FastPrep™ Lysing Matrix B (MP Biomedicals, Santa Ana, USA). Cells were lysed at 4,5 m/s for 30 sec by using a FastPrep®-24 Instrument (MP Biomedicals, Santa Ana, USA) with a 5 min cool down on ice between cycles.

After tubes were centrifuged with 21000 g for 15 min at 4°C the supernatant was transferred into a clean 2 mL plastic tube.

Phase separation was induced by adding 200 µL H_2O. The polar phase was transferred into a new plastic tube and dried in vacuum.

Polar metabolites were methoximated by resolubilization in 20 µL of a 40 mg mL-1 solution methoxyamine hydrochloride in pyridine through shaking at 30° C for 90 min. After methoximation 80 µL silylation mix (30 µL alkane mixture in 1 mL N- methyl-N-trimethylsilyltrifluoroacetamid) was added and incubated for 30 min at 37° C. Following the silylation plastic tubes were centrifuged for 2 min at 14000 g and the supernatant was transferred into GC-microvials.

3.2.2 Protein Extraction

3.2.2.1 Endoproteins

Cells were resuspended in protein extraction buffer (100 mM NaCl, 100 mM tris-HCl pH 7.5, 10 % (v/v) Glycerol, 3 % SDS (m/v)) directly after washing and transferred into bead tubes with 0.5 g *FastPrep™ Lysing Matrix* B (MP Biomedicals, Santa Ana, USA). Cells were lysed

at 4,5 m/s for 30 sec by using a *FastPrep®-24 Instrument* (MP Biomedicals, Santa Ana, USA) with a 5 min cool down on ice between cycles.

After tubes were centrifuged with 21000 g for 15 min at 4°C, the supernatant was transferred into a clean 2 mL plastic tube.

a) TCA – Precipitation

For precipitation of proteins with TCA, 20 % TCA concentration was achieved by adding 100 % ice cold in water solubilized TCA to the collected supernatant. After vortexing, the sample was incubated 30 min on ice. After precipitation plastic tubes were centrifuged with 21000 g for 15 min at 4°C and the protein pellet was washed three times with ice cold acetone.

b) Phenol Extraction

An equal volume of phenol saturated with Tris-HCl, pH 7.0 (Carl Roth Nr: 0038,3) was added to the collected supernatant and vortexed for 5 min. To separate phenol phase from water phase, plastic tubes were centrifuged with 20000 g for 2 min at 4° C. The lower phenolic phase was transferred into a 15 mL falcon tube covered with five volumes of cold 0.1 M ammonium acetate in methanol, vortexed and stored at -20°C overnight in order to precipitate the proteins. Falcons were centrifuged with 5346 g for 30 min at 4°C and the protein pellet was washed two times with methanol and one time with acetone.

3.2.2.2 Exoproteins

The growth medium was filtered using a 0,2 µm syringe filter and proteins were precipitated by adding the right amount of 100 % ice cold TCA to achieve 20 % (w/v) end concentration. Samples were wortexed and incubated over night. After incubation samples were centrifuged with 38759 g for 30 min at 4° C and the protein pellet was washed 3 times with ice cold acetone.

3.2.3 Protein Quantification

For protein quantification bicinchoninic acid assay (BCA assay) was used. For this assay protein pellet was resuspended in 200 µL of 6 M urea and 5 % SDS. In a 96 well flat bottom plate 2 µL of solubilized protein was mixed with 200µL BCA standard working reagent which consists of 100 vol. of reagent A [sodium bicinchoninate (0.1 g), $Na_2CO_3 \cdot H_2O$ (2.0 g), sodium tartrate (dihydrate) (0.16 g), NaOH (0.4 g), $NaHCO_3$ (0.95 g), made up to 100 mL. If necessary, adjust the pH to 11.25 with $NaHCO_3$ or NaOH] and 2 vol of reagent B [$CuSO_4$ $5 \cdot H_2O$ (0.4 g) in 10 mL of water]. After incubating at 60 °C for 20 min the absorption was measured at 562 nm. The measurement of bovine serum albumin (BSA) for the standard

curve between 0.5 µg/µL and 10 µg/µL and of the samples were done in three technical replicates.

3.2.4 In Gel Digestion

100 µg of protein was mixed with 5x Laemmli buffer and heated for 5 min at 95° C and after placed into the pocket of a 6 cm long 12 % Polyacrylamide gel with a 5 % stacking part and run with 100 V. The Gel was stained with Coomassie Brillant Blue R-250 solution (0,1 % Coomassie Brilliant Blue R250, 40 % methanol, 10 % acetic acid) and destained with destaining solution (40 % methanol 2 % acetic acid). The protein lane is than divided in four parts (Figure 3). Every part was cut into approx.. 3 x 3 square pieces and put in a 1.5 mL plastic tube. For further destaining the gel pieces, 1 mL of 200 mM ammonium bicarbonate (Ambic) and 50 % Acetonitrile (ACN) solution was put in to the plastic tubes and incubated at 37 °C for 30 min with shaking and time to time vortexing. This process was repeated until sufficient destaining was achieved. After discharging the solution from the plastic tubes 500 µL of 50 mM Ambic and 5 % ACN solution was added and incubated at 37 °C for 15 min with shaking and regular shaking. Gel pieces were incubated in 500 µL 100 % ACN at 37 °C for 10 min. This step was repeated until gels were sufficiently dried. After discharging ACN gel pieces were dried in vacuum for 5 min. Protease solutions were prepared with protease buffer (25 mM Ambic, 10 % ACN and 5 mM CaCl2). For trypsin digestion gels were rehydrated with 12,5 ng/µL trypsin (Roche product number: 000000011418025001) solution and for trypsin/chymotrypsin digestion gels were rehydrated with 6,25 ng/µL trypsin and 6,25 ng/µL chymotrypsin (Roche product number: 000000011418467001) mixture. After rehydration gels were incubated at 37° C in an oven for 16 h.

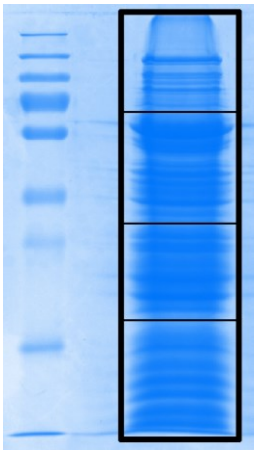

Figure 3: SDS.PAGE gel cutting template.

3.2.5 Peptide Extraction

In first step 150 µL of 50 % ACN, 1 % formic acid was added to the plastic tubes, which were than incubated 5 min on room temperature and sonicated for 3 min in a low intensity ultrasound bath. After transferring the supernatant in a 1.5 mL low binding tube first step was repeated and transferred to the same low binding tube. In the last step of peptide extraction 100 µL of 90 % ACN, 1% formic acid was added to the gel pieces and incubated 5 min on

room temperature and transferred in to the same low binding tube. This step is also repeated if the gel pieces were not dry and plastic like. Extracted peptides were dried in vacuum.

3.2.6 Desalting

3.2.6.1 C_{18} (Bond Elute 96 plate)

Peptides were resuspended in 75 µL peptide solubilization buffer (4 % ACN, 0.25 % formic acid) incubated on ice for 10 min, resuspended by pipetting up and down and sonicated for 3 min in a low intensity ultrasound bath. After incubating 5 min at room temperature, they were kept on ice.

Each C_{18} well was activated by sucking two times 700 µL methanol through the C_{18} material. Following activation step, wells were washed two times with 700 µL water. Samples were than put in the center of the well and incubated for 3 min at room temperature, and than slowly aspirated. Bound samples were desalted by washing the plates with 5 times 400 µL water. Sample and the first two flow through of the desalting step was collected for graphite desalting. Desalted bound peptides on C_{18}-material were eluted with 200 µL methanol. The last step was repeated. Desalted peptides in methanol were transferred in a clean low binding tube.

3.2.6.2 Graphite

In empty spin columns 10 mg graphite powder was weighed. The Graphite was washed with 500 µL of 20 % ethanol. The in C_{18} step collected flow through was acidified with 100 % trifluoroacetic acid (TFA) to the final concentration of 1 %. The ethanol is then spun down by centrifugation at 2000 g for one min. For the column preparation first 100 µL of 1 M ammonium hydroxide was added, vortexed and spun down by centrifugation at 2000 g for one minute. This step was repeated once. The graphite was activated by adding 100 µL 100 % ACN and spun down at 2000 g for one min. Finally, 100 µL of 1 % TFA was added and spun down at 2000 g for one min. The last step was also repeated once. The acidified sample is than incubated for 10 min with graphite and mixed periodically. The column was than centrifuged at 1000 g for three min. The column was washed two times with 200 µL 1 % TFA and centrifuged at 2000 g for one min. A new collection tube was placed and peptides were eluted by four times adding and centrifuging of 100 µL 0,1 % formic acid in 50 % ACN at 2000 g.

3.3 Sample Preparation MS/MS-Shotgun Analysis

C_{18} and graphite desalted peptides were combined together and dried under vacuum. Before MS/MS shotgun analysis peptides were solubilized with peptide solubilization buffer and subsequently incubated for 10 min on ice. Finally, samples were sonicated for 3 min in a low intensity ultrasound bath and incubated for 5 min at room temperature. In solubilities were spun down by centrifugation at 21000 x g for 10 min at 4° C. 20 µL of the supernatant was transfered to LC microvials.

3.3.1 Measurement Parameters

3.3.1.1 GC-MS

After derivatization, samples were analyzed on an Agilent® 6890 gas chromatograph coupled to a LECO Pegasus® 4D GC×GC-TOF spectrometer. An Agilent HP-5MS column (30 m length, 0,25 mm diameter, and 0,25 µm film) was used to separate the components. The injection temperature was set 230° C with a back inlet split ratio of 2 using helium. The starting temperature of the oven was set to 70° C for 1 min and the temperature was increased 9° C per minute until the temperature of 330° C was reached. This temperature was kept for 8 min. The mass range of the mass spectrometer was set 40-600 m/z with a 20 spectra/second data acquisition rate. The detector voltage was 1550 V.

3.3.1.2 LC-MS

Depending to the sample 1 µg or 2,5 µg of each biological replicate were applied to a reverse phase C18 column (Thermo scientific, EASY-Spray 500 mm, 2 µm particle size) and separated during a 180 min gradient until 40 % solution B (see below) with a flow rate of 300 nL min−1. Liquid chromatography was performed with two mobile phases (A: 0,1 % formic acid, B: 90 % ACN and 0,1 % formic acid) and their mixtures. MS measurement was performed on an Orbitrap Elite (Thermo Fisher Scientific, Bremen, Germany) with the following settings: Full scan range 350–1800 m/z resolution 120000, max. 10 MS2 scans (activation type CID), repeat count 1, repeat duration 30 sec, exclusion list size 500, exclusion duration 30 sec, charge state screening enabled with rejection of unassigned and +1 charge states, minimum signal threshold 500.

3.3.2 Bioinformatic Parameters

3.3.2.1 Proteomics

Peptides were identified from the raw data by searching with the SEQUEST algorithm included in the commercial software called proteome Discoverer (Thermo, Germany - version 1.4). Four raw data files belonging to a biological replicate were merged. Min and max precursor mass was set at 350 Da and 10000 Da. The variable modifications were set as acetylation of the N terminus, oxidation of methionine and carbamylation of cysteine, with mass tolerances of 10 ppm for the parent ion and 0.8 Da for the fragment ion. Up to two missed cleavage sites were permitted. A Uniprot database with 3085 entries was used. Identification confidence was set to 5% false discovery rate based on q-value.

To minimize false identifications, peptide identifications were accepted if they exceeded the filter parameter Xcorr score vs charge state with SequestNode Probability Score (+1=1.5, +2=2.0, +3=3, +4=4, +5=5, +6=6, +7=7, +8=8) and if they were high confident peptides. Protein identifications were only accepted if minimum 2 peptides were identified.

3.3.2.2 Metabolomics

The raw data was processed with LECO ChromaTOF® software. Compounds were annotated by the comparison of the retention index and spectra library (minimum match factor of 800) with reference compounds.

3.3.3 Exoproteomics Pilot Experiment

The pilot experiment was performed in the same way to the exoproteomic experiment described above. However, in this experiment, 40 µg of proteins were mixed with 5x Laemmli buffer and heated for 5 min at 95° C and than placed in to the pocket of a 12 % Polyacrylamide gel with a 5 % stacking part and run with 100 V just 1 cm in the 12 % part of the gel. The graphite desalting step was not performed. Furthermore, some LC-MS measurement parameters were different: A 120 min gradient until 40 % solution B was used, max MS2 scans were set to 20, repeat duration to 50 sec and the minimum threshold was set to 1000.

4 Results

4.1 Method Optimization for Endoproteomics

To investigate proteome coverage, different protein extraction and digestion methods were compared. For each method 3 biological replicates were used. The compared parameters of the sample preparation methods are listed in Table 1.

The amount of peptides injected to the mass spectrometer is different because of a calculation error. This mistake introduced a bias towards the method optimization. Due to the differences in the amount of the injected peptide, comparison between the methods with TCA precipitation and methods with phenol extraction was not possible. However, we experimentally show that with the performed in gel digestion method the injection of a 2,5 µg theoretical amount of peptides is possible and did not cause an overloading of the column.

Extraction Method	Digestion Method	Amount of protein loaded to the gel	Amount of peptides injected in to MS
Phenol	Trypsin	100 µg	2,5 µg
MCW-Phenol	Trypsin	100 µg	2,5 µg
TCA	Trypsin	200 µg	1 µg
TCA	Trypsin	100 µg	1 µg
TCA	Trypsin + Chymotrypsin	100 µg	1 µg

Table 1: Sample preparation parameters

4.1.1 SDS-PAGE Gels of Intracellular Proteins

Figure 4 - Figure 8 shows SDS-PAGE gels from different protein extraction methods. On each gel, 3 biological replicates were loaded. Each biological replicate looks almost identical. TCA precipitation demonstrated sharper protein bands than phenol extraction. Comparing the protein bands of normal phenol extraction with the phenol extraction with a previous metabolite extraction using MCW (Methanol/Chloroform/Water), intensity of the protein bands larger than 15 kDa increases in metabolite extracted samples and also the bands smaller than 15 kDa became sharper. The smear at small sized proteins could indicate a degradation of larger proteins.

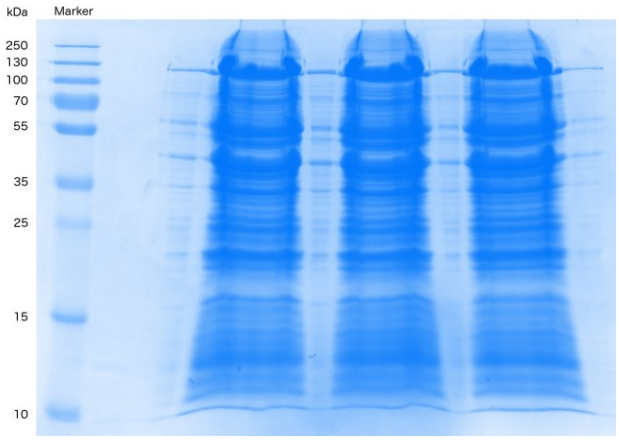

Figure 8: SDS-PAGE gel of 100 µg of phenol extracted intracellular proteins after metabolite extraction with MCW for trypsin digestion.

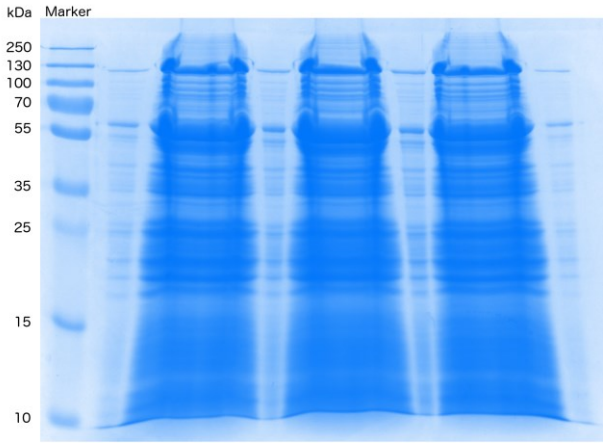

Figure 8: SDS-PAGE gel of 100 µg of phenol extracted intracellular proteins for trypsin digestion.

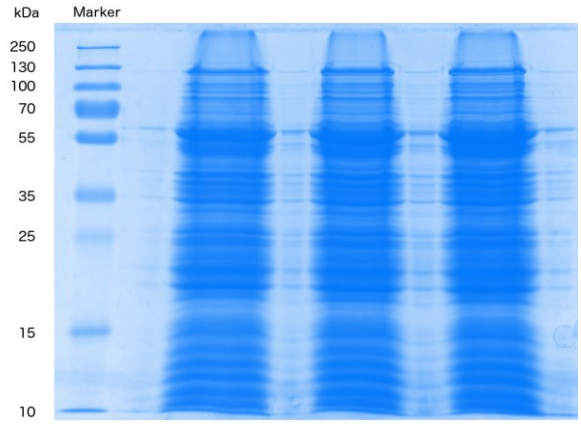

Figure 8: SDS-PAGE gel of 100 µg of TCA-precipitated intracellular proteins for chymotrypsin/ trypsin digestion.

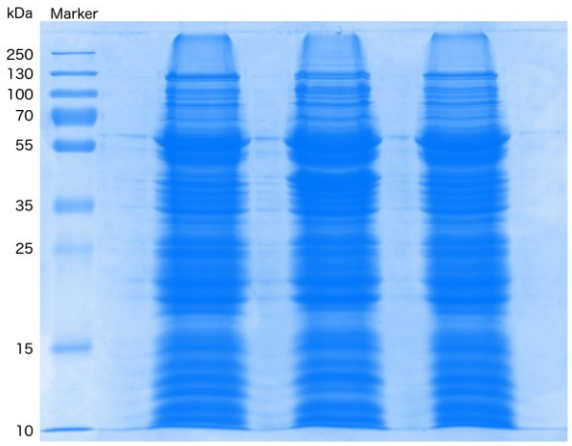

Figure 8: SDS-PAGE gel of 100 µg of TCA-precipitated intracellular proteins for trypsin digestion.

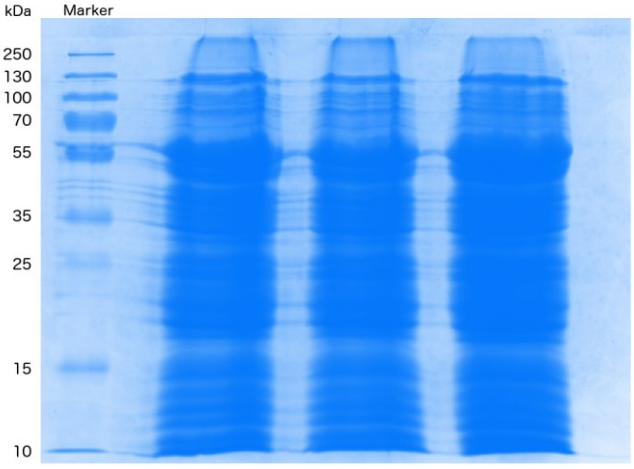

Figure 8: SDS-PAGE gel of 200 µg of TCA-precipitated intracellular proteins for trypsin digestion.

4.1.2 Chromatograms

As mentioned in the methods part, every line of the SDS-PAGE was divided in four fractions. Figure 9 shows base peak chromatograms of the digested proteins from the second gel part of the bottom of every method. Every chromatogram has a different pattern with some similarities. MCW + Phenol and Phenol method has the broadest spectrum.

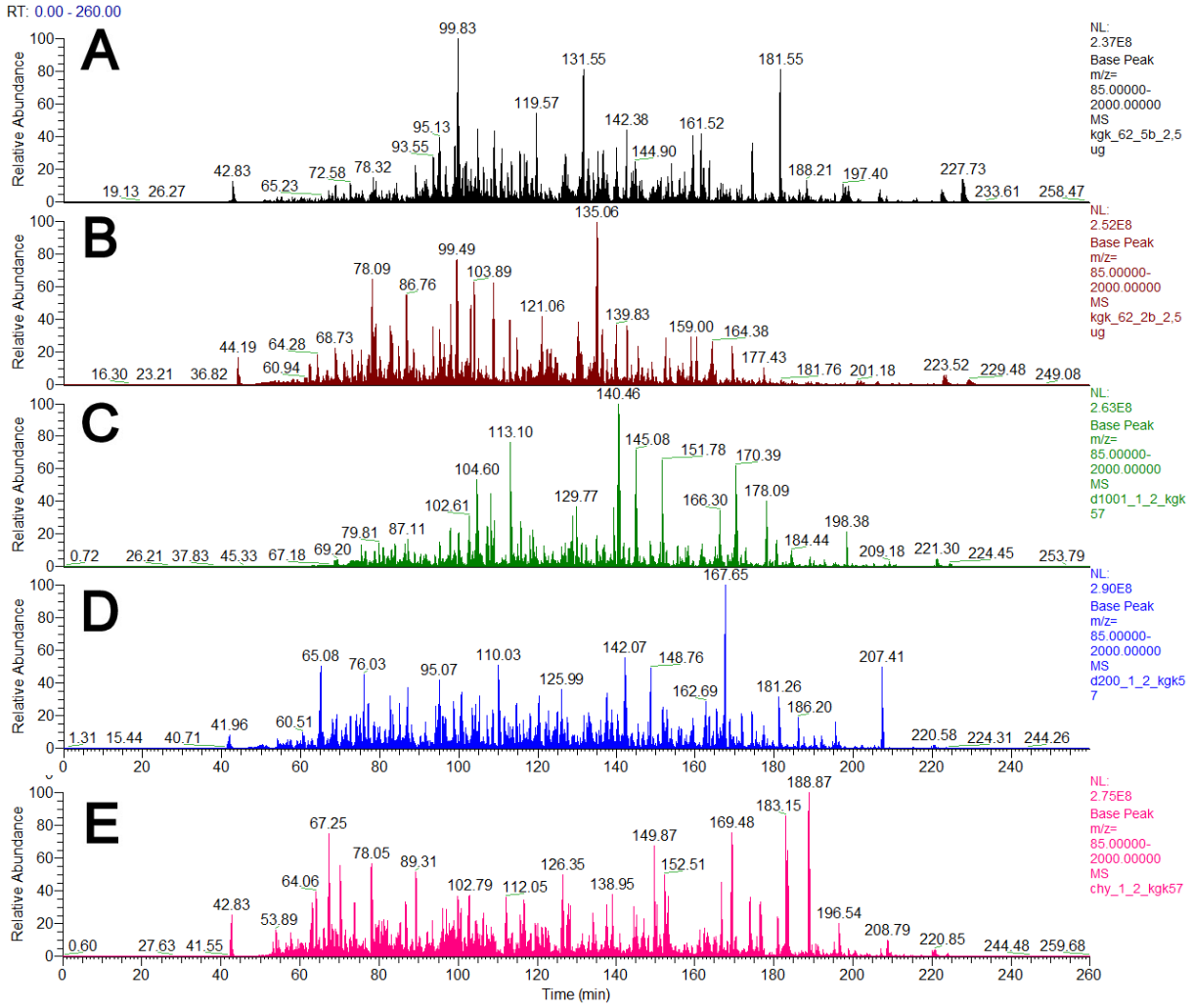

Figure 9: Base peak chromatogram of every second measurement from different sample preparations. A) MCW + Phenol method, B) Phenol method, C) TCA 100 µg method, D) TCA 200 µg method, E) TCA – Chy/Try method.

4.1.3 Mass spectrometry results

The measured spectra are examined and filtered with the software "Proteome Discoverer 1.4".

Methode	Number of proteins	Number of peptides	Number of unique peptides	Coverage
MCW + Phenol	1458	9107	8648	47%
Phenol	1255	7749	7314	41%
TCA 100 µg	1040	6690	6310	34%
TCA 100 µg Chymotrypsin/ Trypsin	865	6178	5814	28%
TCA 200 µg	818	4585	4190	27%

Table 2: Average mass spectrometry results of biological replicates

Table 2 lists the average mass spectrometry results of the biological replicates for each method. According to this data, in phenol extracted samples the best coverage was achieved by phenol extraction with preliminary metabolite extraction using MCW, and in TCA precipitated samples by TCA 100 µg. MCW + Phenol method resulted an average protein number of 1458 which corresponds to 47 percent of the proteome. TCA 100 µg resulted in an average protein number of 1040 which corresponds to 34 percent of the proteome.

Figure 10 show the venn diagram of identified proteins from every biological replicate of all methods. Out of 1929 unique proteins, 832 were at least identified once in every method. The MCW + Phenol method has the most individual identifications. 266 proteins were only identified in the MCW + Phenol method. This can only be compared to the Phenol method because only there the same amount of peptides was injected.

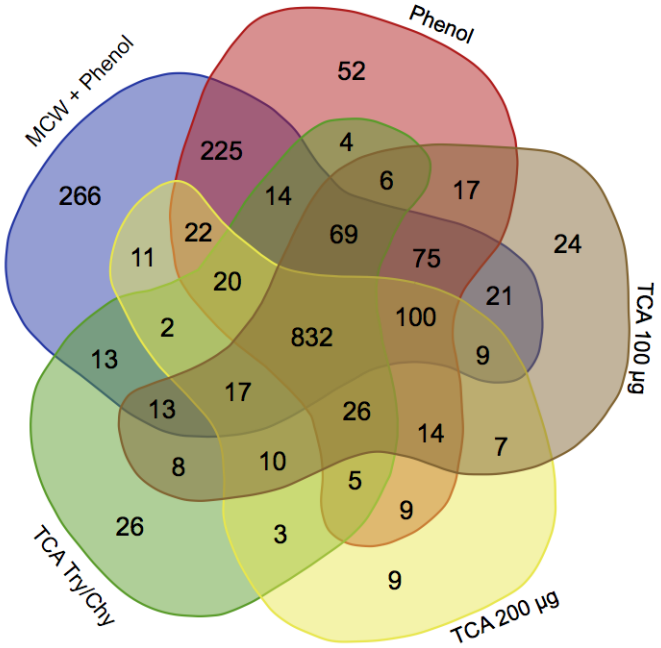

Figure 10: Venn diagram of identified proteins from every sample preparation method.

4.1.4 Reproducibility of the results

Reproducibility was tested by three biological replicates for each method. The identified proteins of each replicate is visualized with a circle in a venn diagram in Figure 11.

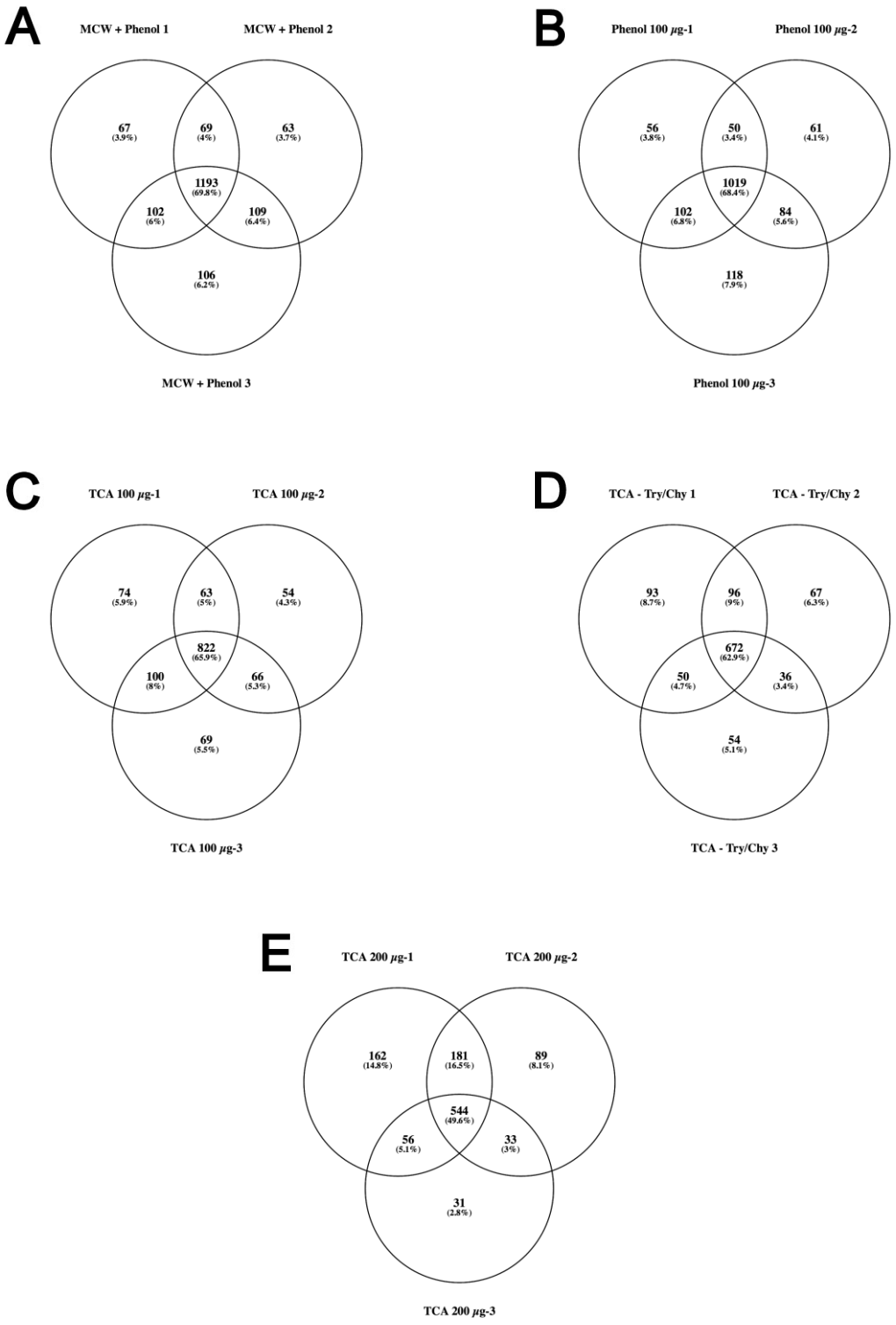

Figure 11: Reproducibility of the protein coverage in endoproteomics methods. Each circle represents a biological replicate. A) Phenol extraction with preliminary metabolomics extraction B) Phenol extraction C) TCA 100 µg D) TCA 100 µg Trypsin/Chymotrypsin digestion E) TCA 200 µg

In case of the MCW + Phenol method 69,8 % of all proteins were identified by each replicate and 16,4 % of all proteins were identified in two of three replicates. In the Phenol extraction 68,4 of all proteins were identified by each replicate and 15,8 % of all proteins were identified in two of three replicates. The reproducibility of these methods are high enough to select one of them as a scientific method to compare control and treated samples.

In the method called TCA 100 µg, 65,9 % of all proteins occurred in all replicates and 18,3 % were identified in two of three replicates. In case of Trypsin/Chymotrypsin digestion the reproducibility decreased. 62,9 % of all proteins were found in all replicates and 17,1 % were identified in two of three replicates. TCA 200 µg had the worst reproducibility. Only 49,6 % of the proteins were identified by each replicate and 24,6 % were identified in two of three replicates. In TCA 200 the decrease of the reproducibility occurred mainly because of the decrease of identified proteins in the third replicate. Which could have been caused by an experimental mistake.

4.1.5 Normalized Spectral Abundance Factors (NSAF)

NSAF is a quantitative profiling method for shotgun proteomics, which uses the number of spectral counts (SpC) and protein length (L). NSAF value for a protein x is calculated as described in the Equation 1[36].

$$(NSAF)_x = \frac{(SpC/L)_x}{\sum_{i=1}^{N} (SpC/L)_i}$$

Equation 1: NSAF for protein x

TOP10	MCW + Phenol	Phenol	TCA 100 µg	TCA 200 µg	TCA 100 µg Chymotrypsin/ Trypsin
1	Q9RRB6	Q9RVQ0	Q9RVQ0	Q9RVQ0	Q9RY32
2	Q9RVQ0	Q9RRB6	Q9R342	Q9RY32	Q9RU24
3	Q9RV45	Q9RV45	Q9RSA5	Q9RU24	Q9RSA5
4	Q9RUR8	Q9RU24	Q9RU24	Q9RUR8	Q9RVQ0
5	Q9RTJ4	Q9RUP1	Q9RXT9	Q9RXT9	P56867
6	Q9R342	Q9RXT9	Q9RV45	Q9RV45	Q9RXR9
7	Q9RSS9	Q9RY32	Q9RUR8	P56867	Q9RV45
8	Q9RT27	Q9RUY5	P56867	Q9RXR9	Q9RXJ0
9	Q9RWQ9	Q9RU54	Q9RSQ4	Q9RVJ0	Q9RT27
10	Q9RZ12	Q9RSS9	Q9RY32	Q9R342	Q9RRB6

Table 3: Top 10 most abundant proteins of different sample preparation methods.

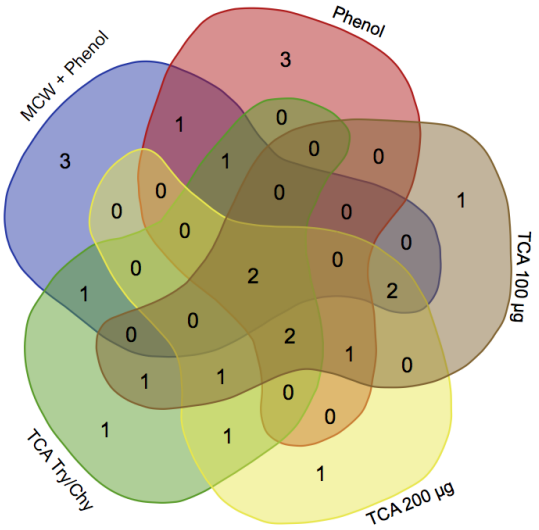

Figure 12: Distribution of 10 most abundant proteins between different sample preparation methods in a venn diagram.

Table 3 lists the 10 most abundant proteins in descending order, Figure 12 shows the distribution of the proteins in Table 3. In Table 3 only two proteins showed up in all methods: Q9RVQ0 and Q9RV45. Beside MCW + Phenol method Q9RVQ0 is in each method with only trypsin digestion the most abundant protein. In MCW + Phenol method Q9RVQ0 comes in second line. In case of the Trypsin/Chymotrypsin method it is in the fourth line. TCA 100 µg and TCA 200 µg are the methods which have 7 identical proteins in their 10 most abundant proteins. According to Figure 12 and Table 3 the calculated protein abundance changes in each method. This makes it impossible to determine the most abundant protein by only using one method.

The quantitative reproducibility of the methods is roughly examined by comparing the 10 most abundant proteins in each biological replicate. Venn diagrams in Figure 13 -Figure 14 show the quantitative reproducibility. MCW + Phenol and Phenol method are the most reproducible methods 8 of 10 proteins were identified in all biological replicates.

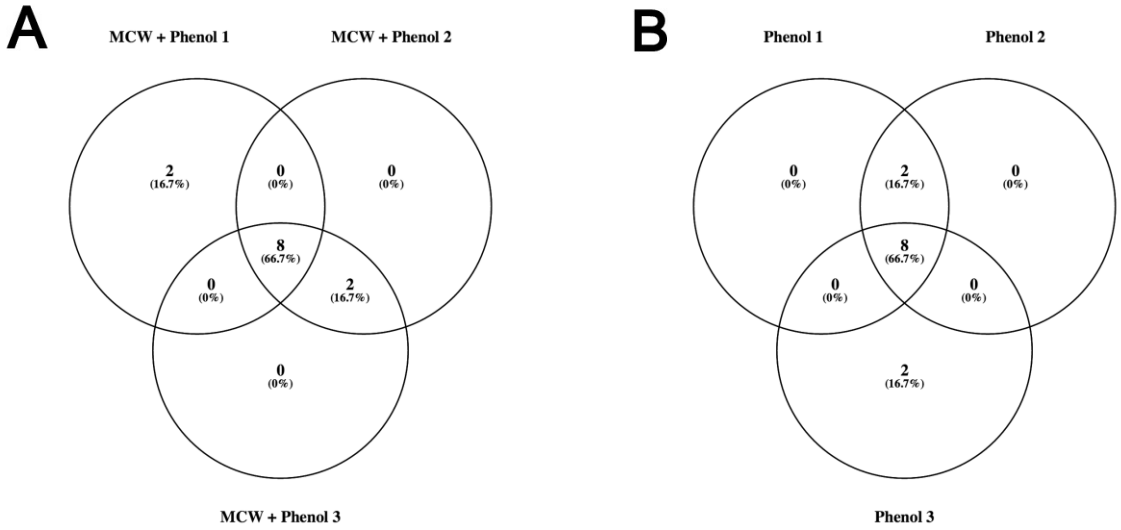

Figure 13: Venn diagram of 10 most abundant proteins of biological replicates. A) MCW + Phenol Method B) Phenol Method

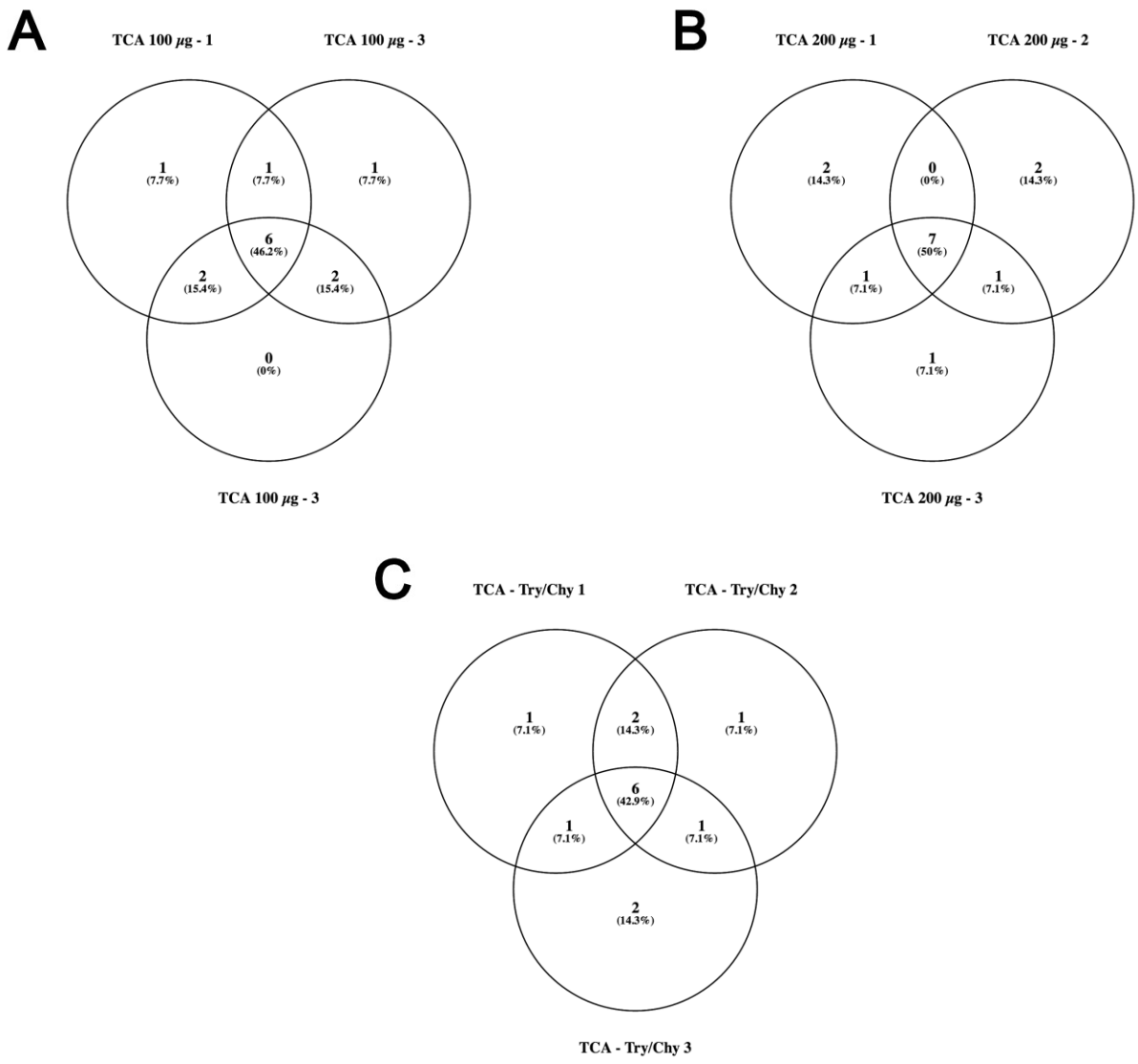

Figure 14: Venn diagram of 10 most abundant proteins of biological replicates. A) TCA 100 µg Method B) TCA 200 µg Method C) TCA – Try/Chy Method

4.1.6 Comparison of membrane protein identification performance

Membrane proteins are difficult to identify through shotgun proteomics because of solubilisation problems and low protein abundance[37].

The Uniprot database for *Deinococcus radiodurans* has 582 proteins (19% of its proteome), which are annotated as membrane proteins. Figure 15 shows the number of membrane protein identification for each method. A comparison between phenol extracted and TCA precipitated methods can not be made, because the theoretical injected amount of peptides is different between these groups. The different parameters between TCA precipitated methods do not show a significant increase in membrane proteins. In case of phenol extracted methods, phenol extraction with a preliminary metabolite extraction identified 57 more membrane proteins than phenol extraction. The MCW + Phenol method has increased the identified proteins by 203 proteins.

19 % of the proteome of *D. radiodurans* is annotated as membrane proteins and 57 membrane proteins correspond to 28 % of total increase. These ratios can be interpreted as a membrane protein enrichment.

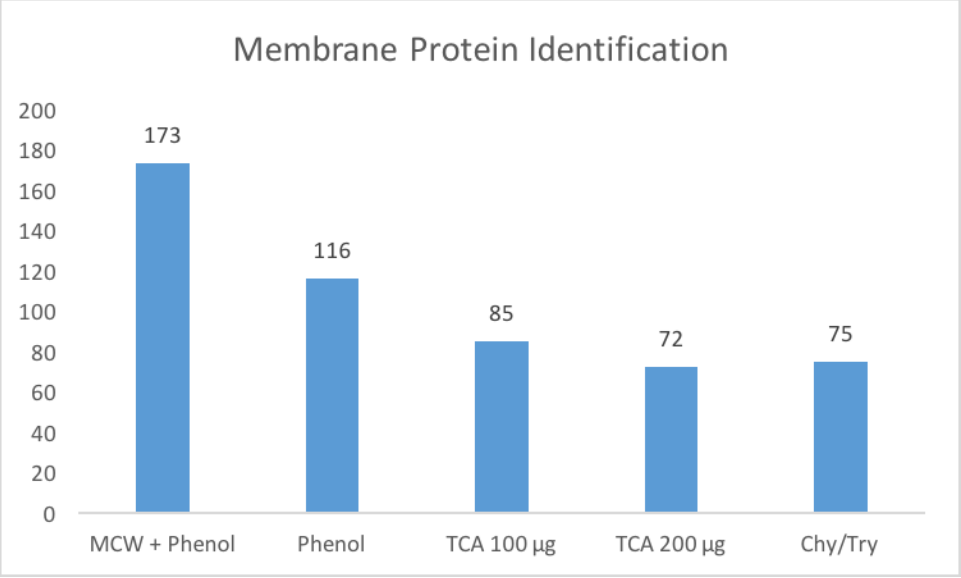

Figure 15: Comparison of the identified membrane protein amounts for each sample preparation method.

4.1.7 Genome Annotation

Genome annotation adds additional information to a raw DNA sequence. Raw DNA sequence is analyzed and interpreted to understand its biological meaning. Genome annotation has three levels: nucleotide level, protein level and process level[38].

In protein level annotation a protein is named and a putative function is assigned[38]. This annotation can be made manually by experimental biologist or computationally. Computational annotation is performed by comparison of unannotated protein with manually annotated proteins. Manual annotation and computational annotation differ in performance and quality. The manual annotation produces high quality annotation but it requires time. In contrast to manual annotation computational annotation is fast but produces low quality annotations [39].

The Gene Ontology (GO) project provides a database, with structured, controlled vocabularies and classifications freely for the scientific community for annotation of genes, gene products and sequences (http://www.geneontology.org/). In GO database gene products are categorized in three non-overlapping domains (molecular function, biological process and cellular component). Every annotation in GO database has to be attributed to a source such as literature reference, another database or a computational analysis[40].

The identified proteins are annotated using the Gene Ontology (GO) database[41]. Table 4 shows the percentage of identified proteins by different sample preparation methods for selected protein functions.

		% Proteins identified by each sample preparation method				
Functional protein classification	Protein number	MCW + Phenol	Phenol	TCA 100 µg	TCA 200 µg	TCA - Chy/Try
Transcription factor activity, sequence-specific DNA binding	35	40%	37%	26%	23%	14%
Catalytic activity	1282	68%	60%	47%	42%	40%
Structural molecule activity	56	80%	63%	55%	52%	52%
Transporter activity	169	37%	24%	20%	18%	18%
Binding	916	71%	61%	48%	42%	38%
Electron carrier activity	33	76%	73%	58%	55%	42%
Antioxidant activity	20	80%	85%	75%	55%	70%
Sigma factor activity	3	67%	67%	33%	33%	67%
Enzyme regulator activity	4	100%	0%	100%	75%	100%
Molecular transducer activity	29	59%	28%	17%	7%	7%

Table 4: Distribution of identified proteins in protein functions

MCW + Phenol method did not only identify the most proteins of the total proteome, it also identified the most proteins of every functional protein classification, besides antioxidant activity, in which phenol method scored the highest percentage. Non of the methods changed the distribution of the proteins into the functional classifications sufficiently enough to mention a gene enrichment.

Figure 16 shows the number of identified annotated proteins from the method MCW + Phenol in a bar graph.

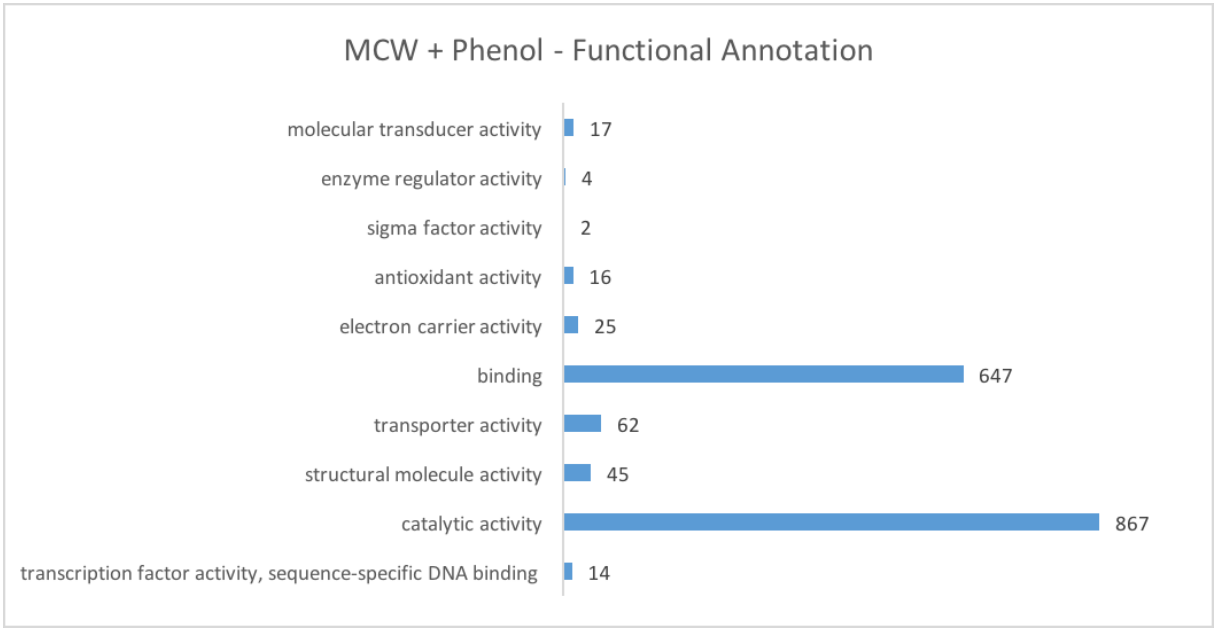

Figure 16: Distribution of proteins in functional annotations

The distribution of identified proteins in the subcategory catalytic activity was also investigated. Table 5 shows the percentage of identified proteins by different sample preparation methods for subcategories of catalytic activity.

		% Proteins identified by each sample preparation method				
Subcategory Catalytic activity	Protein number	MCW + Phenol	Phenol	TCA 100 µg	TCA 200 µg	TCA - Chy/Try
Recombinase activity	2	50%	100%	50%	50%	50%
Glycogen debranching enzyme activity	2	100%	50%	50%	50%	50%
Transposase activity	17	0%	0%	0%	0%	0%
Oxidoreductase activity	254	69%	63%	52%	44%	44%
Transferase activity	385	64%	56%	40%	37%	34%
Hydrolase activity	434	68%	58%	45%	40%	38%
Lyase activity	91	74%	64%	58%	54%	43%
Isomerase activity	58	79%	74%	64%	64%	55%
Ligase activity	80	91%	83%	60%	55%	46%
Lipoate synthase activity	1	100%	100%	100%	100%	100%
Deaminase activity	7	86%	86%	57%	43%	43%

Table 5: Distribution of identified proteins in subcategory catalytic activity

In Table 5 MCW + Phenol method has the most coverage in every subcategory of catalytic activity.

Figure 16 and Figure 17 show the distribution of the identified proteins in cellular compartments and biological processes.

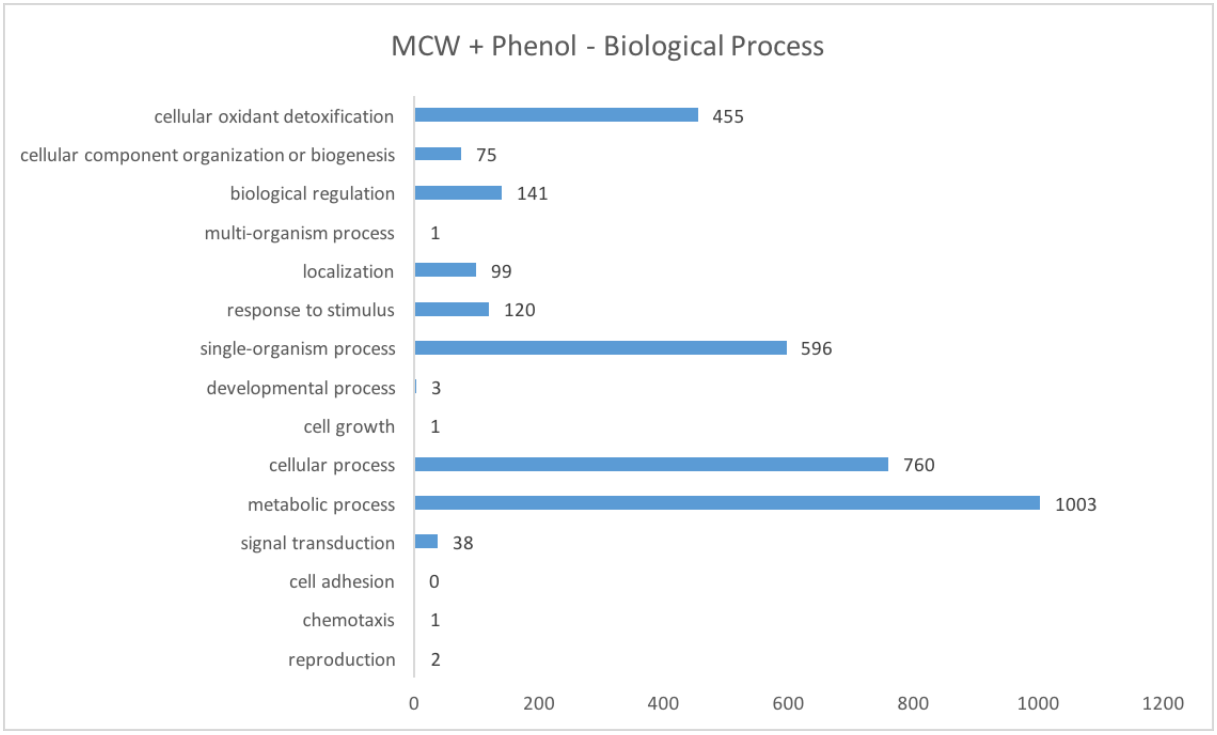

Figure 17: Distribution of identified proteins in biological functions

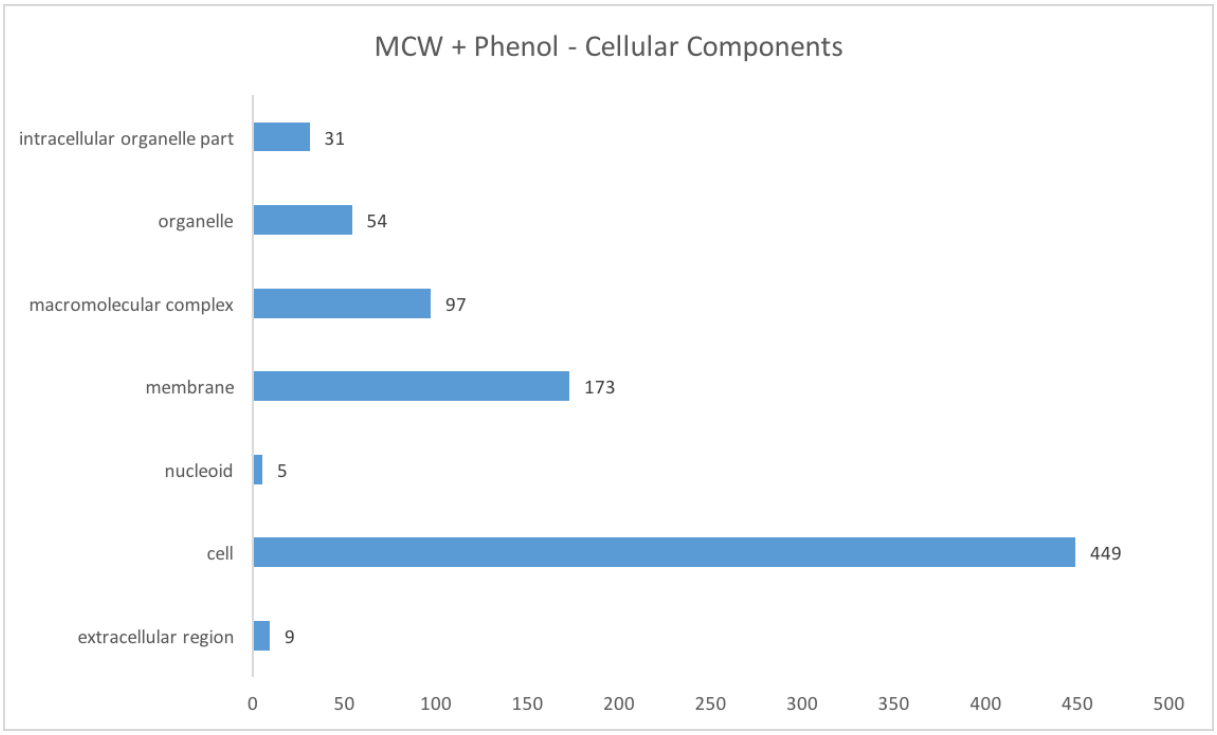

Figure 18: Distribution of identified proteins in cellular compartments

4.2 Method optimization for exoproteomics

Exoproteome comprises all the proteins outside of the cells, which can either be actively secreted and nonsecreted proteins. By investigating exoproteins, environmental interactions and physiological state of living systems can be elucidated[42].

In our experiments we precipitated proteins in the medium with TCA-precipitation as described in the methods part. 100 µg protein were placed into the pockets of a SDS-PAGE gel. In Figure 19 we can see the resulting SDS-PAGE gel of 5 biological replicates.

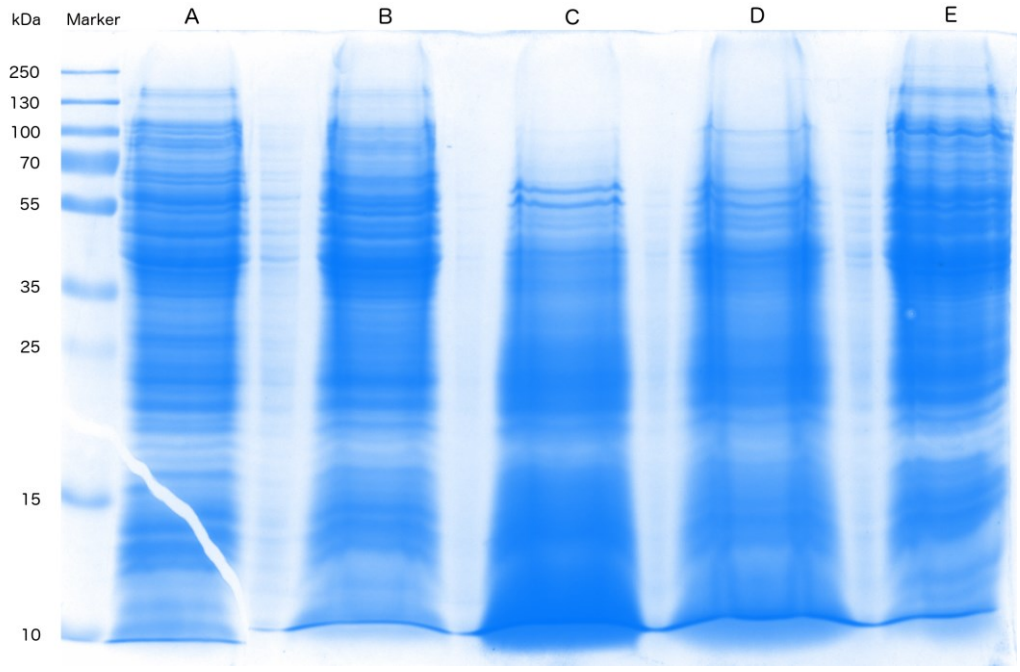

Figure 19: SDS-PAGE gel of 100 µg TCA-precipitated exoproteins

The SDS-PAGE runs of each biological replicate do not look identical. In lane C and D there are obvious degradation of larger proteins. We excluded them and worked with biological replicates, which are represented in lane A, B, and E. Protein bands in lane A, B and E show that the precipitation of extracellular proteins was successful.

In Figure 20 we can see the base peak chromatogram of all fractions from the first biological replicate. The peak intensities in the chromatograms are low. This indicates a low amount of peptides, which could be caused by an insufficient digestion. Proteome Discoverer 1.4 analysis of the raw file found 9 proteins from the first biological replicate, and 2 proteins from each of the biological replicates two and three. These proteins are listed in Table 6.

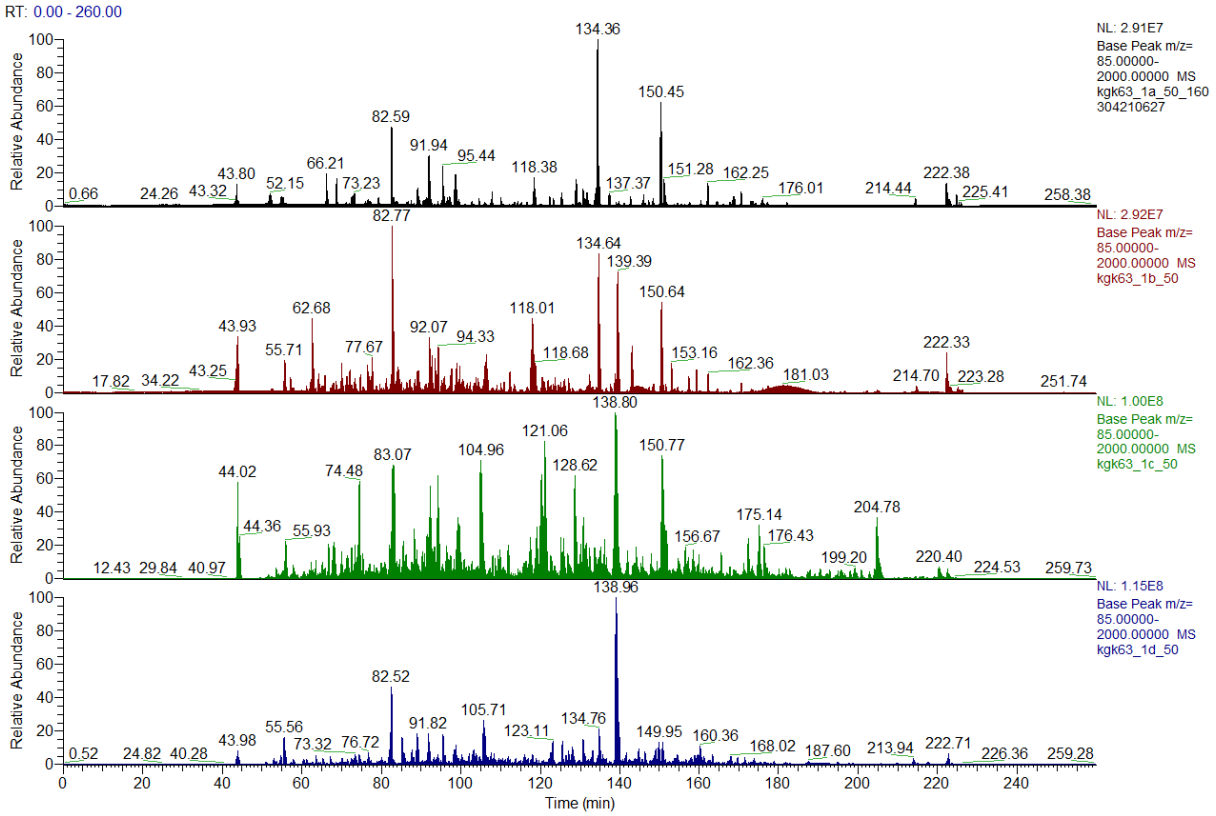

Figure 20: Base peak chromatogram of exoproteins

Biological replicate 1			
Accession	Description	Score	Coverage
P56867	Hexagonally packed intermediate-layer surface protein	50.34	13.08 %
Q9RVW0	DNA-directed RNA polymerase subunit beta	41.14	1.62 %
Q9RRB6	S-layer protein	29.44	8.48 %
Q9RSZ6	ATP-dependent Clp protease ATP-binding subunit ClpX	10.92	5.71 %
Q9RYM8	Probable subtilase-type serine protease	8.00	4.12 %
Q9RWW7	Pilin, type IV	6.48	7.01 %
Q9RV58	Protein DR_1172 OS=Deinococcus radiodurans	5.83	7.72 %
Q9RS05	Serine protease, subtilase family	5.13	7.99 %
Q9RUR8	Uncharacterized protein DR_1314 OS=Deinococcus radiodurans	4.57	10.22 %
Biological replicate 2			
Q9RVW0	DNA-directed RNA polymerase subunit beta	27.96	1.62 %
P56867	Hexagonally packed intermediate-layer surface protein	20.65	7.38 %
Biological replicate 3			
P56867	Hexagonally packed intermediate-layer surface protein	10.87	5.80 %
Q9RVW0	DNA-directed RNA polymerase subunit beta	6.07	1.29 %

Table 6: Identified proteins from 3 biological replicates

4.2.1 Exoproteins Pilot Experiment

1 µg of peptides were applied on a reverse phase C18 column and measured with Orbitrap Elite. The raw data was processed with Proteome Discoverer 1.4 as described in the methods part. The measurement resulted with 122 proteins. The list of the proteins can be found in supplementary materials Table 9.

Figure 21 show the number of the identified annotated exoproteins. 55 of the identified exoproteins were annotated with catalytic activity, 6 with transporter activity, 2 with binding, 1 with electron activity, and 1 with antioxidant activity.

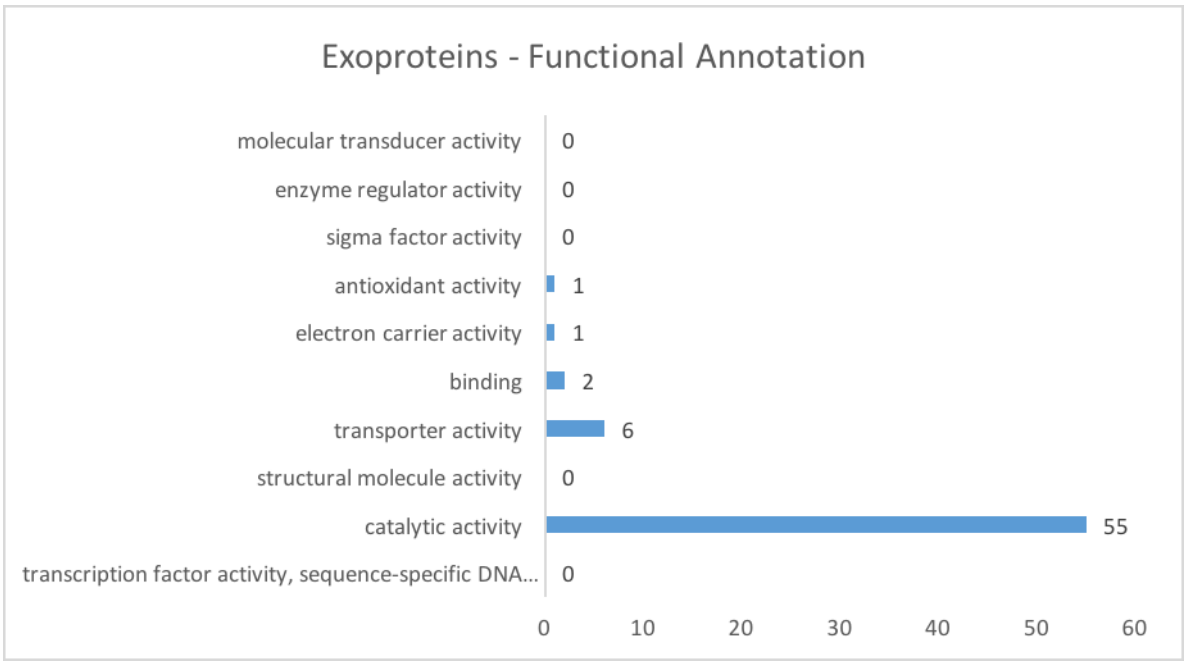

Figure 21: Distribution of exoproteins in functional annotations

4.3 Metabolomics

High throughput analytical techniques of metabolomics investigate all metabolites in a sample, enabling the quantitative analysis of molecular phenotypes of a cell, tissue or organism[43].

Due to the physicochemical diversity of metabolites there is no single protocol for the analysis of all small biological molecules. To increase the coverage, a mixture of techniques must be applied and the results must be combined. At the present time the combination of GC-MS and LC-MS gives promising results by detecting amino acids, sugars, organic acids and free fatty acids using GC-MS and secondary metabolites using LC-MS[43].

As a result of the development of comprehensive and accurate standard techniques, metabolomics enabled the identification of biomarkers and the analysis of environmental perturbations. Furthermore, the integration of transcriptomics, proteomics and metabolomics data will give a deeper understanding of metabolomic networks[43].

By using only GC-MS we have identified 71 metabolites. The identified metabolites are listed in Table 7. Figure 22 shows the GC-MS-spectrum of *Deinococcus radiodurans.*

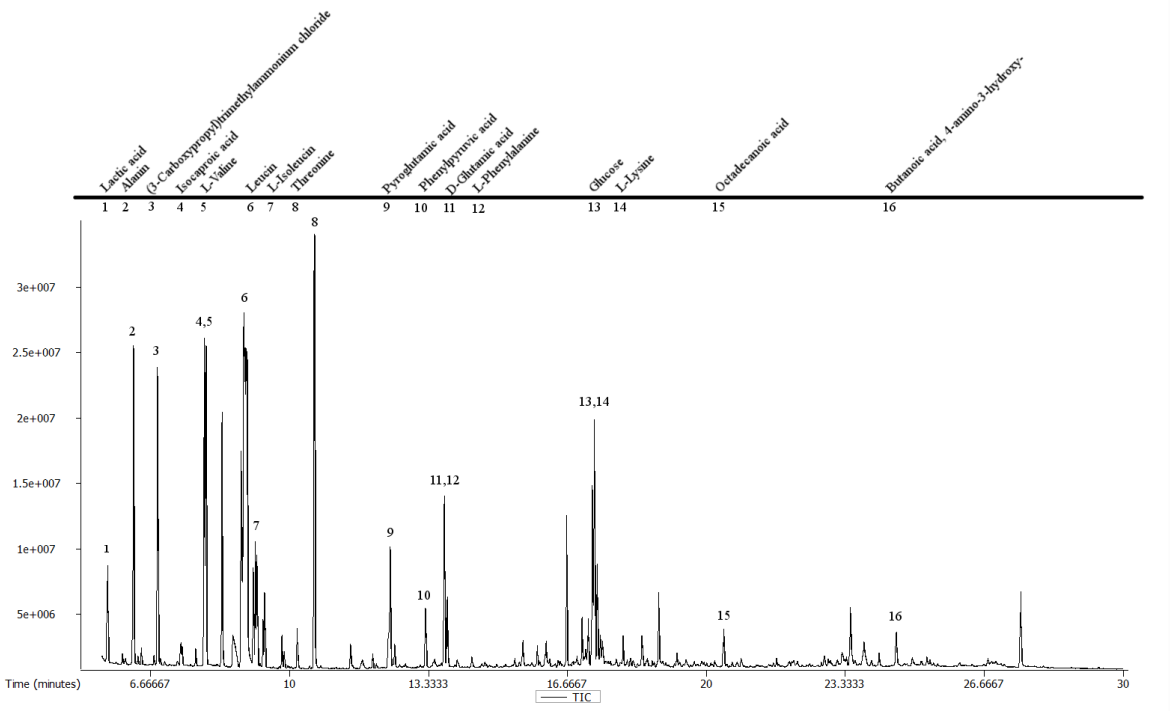

Figure 22: GC-MS-spectrum of Deinococcus radiodurans

Peak	Name	Retention Index	Quant Masses
30	(3-Carboxypropyl)Trimethylammonium chloride	1147	147
932	5-Methylthioadenosine	2824,4	236
893	Adenosine	2670,1	230
990	Adenosine-5-monophosphate	3122,5	169
902	Sucrose	2707,5	361
520	Hexonic acid	1988,7	187
317	Pyrophosphoric acid	1689,9	451
23	2-Hydroxybutyric acid sodium salt	1133,1	131
275	Alanyl-alanine	1615,4	116
462	Allantoic acid	1915,1	331
212	Argininosuccinic acid	1504,5	73
49	Butanoic acid, 2-amino- (2TMS)	1178,8	130
188	Butanoic acid, 4-methylthio-2-oxo-	1465,1	61
38	D-(-)-beta-Hydroxybutyric acid sodium salt	1165,7	117
227	DL-Pyroglutamic acid	1535,1	156
98	Ethanolamine	1276,1	174
763	Glucose-6-phosphate methoxyamine	2373,8	73
372	Glutamine	1788,6	156
403	Glyceric acid-3-phosphate	1834,4	211
21	Glycine	1123,7	102
20	Hydroxylamine	1120,6	146
55	Isocaproic acid, 2-oxo- (1MEOX) (1TMS)	1184,5	99
18	Isovaleric acid, 2-oxo- (1MEOX) (1TMS)	1115	89
6	Lactic acid	1065,9	117
16	L-Alanine	1107,8	190
36	Leucine	1158,5	86
351	L-Glutamine	1755,1	227
224	L-Methionine	1532	176
687	L-Tryptophan	2225,1	202
72	L-Valine	1225,3	218
112	Maleic acid	1313	147

324	Methionine, N-acetyl-	1703	99
666	Normetadrenaline	2194,5	174
697	Octadecanoic acid	2245,8	117
231	Phenylpyruvic acid	1540,1	84
277	Phospho(enol)pyruvic acid monosodium salt hydrate	1617,9	211
681	Spermidine	2215,1	174
169	Tartronic acid, 2-(methylaminomethyl)	1418,5	116
165	Thymine	1411,9	255
926	Trehalose	2813,9	361
126	Uracil	1348,4	99
92	Urea	1268,7	189
12	Valine	1089,7	72
142	Alanine	1369,9	188
228	Aspartic acid	1535,8	232
410	Citric acid	1844,7	273
288	D-Glutamic acid	1634,9	246
457	Fructose	1909,8	103
128	Fumaric acid	1352,2	245
477	Glucose	1937	319
262	Glutaric acid, 2-oxo-	1588,7	198
407	L-Arginine	1840,9	256
251	L-Cysteine	1570	220
109	L-Isoleucine	1305,3	158
483	L-Lysine	1942,1	156
290	L-Phenylalanine	1640,6	218
143	L-Serine	1373,2	204
405	Ornithine	1837,7	142
110	Proline	1308,1	142
114	Succinic acid	1320,5	75
158	Threonine, allo-	1401,4	57
314	L-Asparagine	1688,2	116
494	Tyrosine, 3,5-diiodo-	1958,6	218
360	3-Hydroxytyramine hydrochloride	1772,4	174

1015	3-Methylamino-1,2-propanediol	3276,7	116
994	5-Aminoimidazole-4-Carboxamide-1-beta-D-Ribofuranosyl	3136,1	169
28	alpha-ketoisovaleric acid sodium salt	1141,3	73
307	D-Glucose-3-sulfate sodium salt	1674,6	217
105	Pyridoxamine-5'-phosphate	1292,2	98
1011	Tryptamine, 5-hydroxy-	3231,6	174
560	Xanthine	2042,8	353

Table 7: List of detected metabolites of Deinococcus radiodurans

5 Discussion

5.1 Proteomics

5.1.1 Endoproteomics

The proteomic method with the largest coverage gives a scientist a wider spectrum of proteins which can be used to interpret the reactions of microorganism to different stresses and environments. One important strategy is to reduce the complexity of the biological samples by on-line and off-line fractionation of proteins and peptides before the mass spectrometry analysis[44].

In our experiments we reduced the complexity of the biological samples by off-line protein fractionations through a one dimensional SDS-Page gel and by an on-line peptide fractionation using a C_{18}-column prior the mass spectrometry analysis and tried different sample preparation methods to find the one with the best coverage for the polyextremophile microorganism *Deinococcus radiodurans*.

Unfortunately, not all the methods could be compared with each other, because of a calculation error of the injected peptide amount. Methods where the proteins were extracted using phenol and methods where the proteins were precipitated using TCA must be interpreted separately.

Mass spectrometry results of the methods, where proteins were precipitated using TCA showed that digestion of 200 µg protein instead 100 µg, and using the chymotrypsin/trypsin mixture instead trypsin did not increase the proteome coverage. The method where 100 µg protein was digested with trypsin provided the best proteome coverage of the methods where the proteins are precipitated using TCA. Enrichment of membrane proteins or a functional group was also not detected by any of these methods.

In case of phenol extracted proteins, the method which has a preliminary metabolite extraction using MCW-mix showed the best coverage. The preliminary metabolite extraction increased the

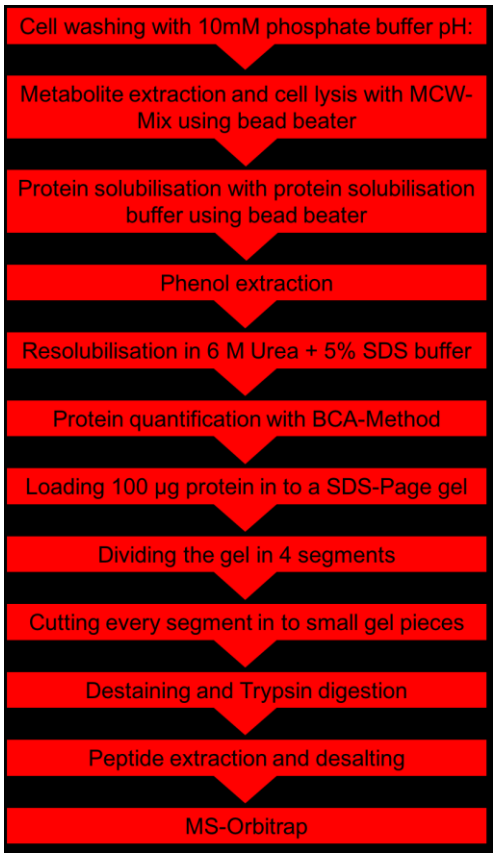

Figure 23: MCW – Phenol Method summary

average number of identified proteins by 203, which is 6% of the proteome. 57 proteins of the increase are membrane proteins which corresponds to 28% of the increase. This results could indicate that MCW-mix leads to an enrichment of membrane proteins. To verify this hypothesis other cell lines or other organisms should be compared similarly as in this work.
MCW-Phenol method was not only the method with the best proteome coverage, it also gave the most identified proteins in every category of molecular functions and allowed the integrative metabolomic studies. The summary of this method is shown in Figure 22.

5.1.2 Functional analysis of most abundant endoproteins

The sample preparation techniques we have used for the endoproteome analysis are shown in Table 1. From these analyses it was possible to estimate protein abundance using NSAF method (described in the section 4.1.5). To judge the efficiency and reproducibility we selected 10 most abundant proteins and compared them. As expected the different extraction methods resulted in heterogeneous efficiency with respect to the 10 most abundant proteins and to the total proteome coverage (see Figure 10 and Figure 12). This phenomena is well described in the the proteomics field[45].
A semiquantitative functional description of 10 most abundant proteins of the MCW-Phenol method was done using GO-database (Table 8).

Protein ID	Name	Protein Function
Q9RRB6	S-layer protein SlpA	Major constituent of the S-layer. Plays an important role in the structural organization and integrity of the S-layer (PubMed:16946272, PubMed:26074883). Binds the carotenoid deinoxanthin, a strong protective antioxidant specific of this bacterium, and could be part of the first lane of defense against UV radiation, especially under desiccation (PubMed:26909071).
Q9RVQ0	Uncharacterized protein	
Q9RV45	S-layer-like array-related protein	carbohydrate binding
Q9RUR8	Uncharacterized protein DR_1314	electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity

Q9RTJ4	Uncharacterized protein	
Q9R342	Elongation factor Tu	This protein promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis. - GTPase activity - GTP binding - translation elongation factor activity
Q9RSS9	50S ribosomal protein L10	structural constituent of ribosome - large ribosomal subunit rRNA binding
Q9RT27	Acyl carrier protein	Carrier of the growing fatty acid chain in fatty acid biosynthesis. - ACP phosphopantetheine attachment site binding involved in fatty acid biosynthetic process
Q9RWQ9	60 kDa chaperonin	Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides; generated under stress conditions. - ATP binding
Q9RZ12	Uncharacterized protein	

Table 8: Functional annotation of 10 most abundant proteins from the MCW-Phenol method.

In Table 8 the GO annotations of the 10 most abundant proteins from the MCW-Phenol method are shown. Here we can see that in the unstressed environment *D. radiodurans* expresses S-layer proteins (Q9RRB6, Q9RV45), proteins necessary for protein translation (Q9R342, Q9RSS9), lipid synthesis proteins (Q9RT27) and chaperones to prevent protein misfolding (Q9RWQ9) in high concentration.

S-layers have many functions such as cell rigidity, cell shape, cell adhesion and cell protection. In *D. radiodurans* S-layer plays also a role in resisting high doses of ionising and UV radiation. SlpA is a primary component of the S-layer and it is important in the organization and integrity of the S-layer[46].

60-kDa chaperonin belongs to the Type I chaperonins which are in bacterial cytoplasm (GroEL). They are formed like a ring and help in assembly of unfolded proteins, which are induced under stress conditions, presumably to repair misfolded proteins[47]. It has been shown that after heat shock the expression of 60.k-Da chaperonin has been induced in *D. radiodurans*[48]. Our results show that 60-kDa chaperonin is relative to all other detected proteins highly abundant in unstressed cells, which could be a precautious protection to heat shock.

Two proteins from 10 most abundant proteins in *D. radiodurans* contribute to protein translation, Elongation factor-Tu (EF-Tu) and 50S ribosomal protein L10. EF-Tu is a important GTPase contributing to the elongation of a polypeptide chain by bringing aminoacyl-tRNA to ribosomal A-site[49].

Another interesting protein at the 10 most abundant proteins is Q9RUR8. Q9RUR8 is an uncharacterized protein with the gene name DR1314 and is homologous to the PRC-H (photosynthetic reaction center–H) barrel domain of *Rhodopseudomonas viridis,* where it regulates electron transfer between quinones[50,51]. In literature it is predicted that in non-photosynthetic energy metabolism systems PRC-H barrel functions as electron-transfer chain regulator[51]. Additionally, it has been proven that DR1314 is a heat shock protective gene[52].

5.1.3 Exoproteomics

Unfortunately, even though we could observe proteins in the SDS-PAGE gel and have loaded a theoretical peptide amount of 2,5 µg, we could not identify high amounts of proteins. The spectra in Figure 20 show low intensity. This could be caused by an unnoticed peptide sample loss in the drying process using speed vac. Another reason of the low amount of protein identification could be an insufficient digestion. In the pilot experiment described in the methods, we could identify 122 proteins. Most of the identified exoproteins have catalytic activity (Figure 21).

5.2 Metabolomics

As mentioned in the endoproteomics section above we have used an integrative extraction protocol (described in literature[53,54]) for the extraction of proteins from *D. radiodurans*. Consequently, we had to test whether we are able to extract and analyze metabolites based on the same protocol using an initial MeOH/Chloroform/Water extraction for the soluble metabolite fraction. Based on this protocol it was possible to identify 71 metabolites using GC-MS. The amount of identified metabolites could be increased if unpolar metabolites the secondary metabolites are also measured.

The integrative extraction protocol makes the extraction of metabolites and proteins from the same sample for a convenient subsequent statistical data mining possible, as described in literature [53,54]. This is especially important if only very little amounts of material is available, e.g. space samples.

Therefor, we propose to use the integrative extraction protocol for future analyses, especially in astrobiological experiments such as the Tanpopo mission. The protocol can also be used to

the analysis of RNA and DNA[53,54]. With the RNA extraction it would be possible to do a RNA sequencing and correlate transcript dynamics to proteins and even metabolite dynamics.

6 Literature

1. Slade, D. & Radman, M. Oxidative Stress Resistance in Deinococcus radiodurans. *Microbiol. Mol. Biol. Rev.* **75,** 133–191 (2011).
2. Yamagishi, A. *et al.* TANPOPO: Astrobiology Exposure and Micrometeoroid Capture Experiments. (2008).
3. Kawaguchi, Y. *et al.* The Possible Interplanetary Transfer of Microbes: Assessing the Viability of Deinococcus spp. Under the ISS Environmental Conditions for Performing Exposure Experiments of Microbes in the Tanpopo Mission. *Orig. Life Evol. Biospheres* **43,** 411–428 (2013).
4. Lipton, M. S. *et al.* Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. *Proc. Natl. Acad. Sci.* **99,** 11049–11054 (2002).
5. Pikuta, E. V., Hoover, R. B. & Tang, J. Microbial Extremophiles at the Limits of Life. *Crit. Rev. Microbiol.* **33,** 183–209 (2007).
6. Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* **409,** 1092–1101 (2001).
7. Rampelotto, P. H. Resistance of Microorganisms to Extreme Environmental Conditions and Its Contribution to Astrobiology. *Sustainability* **2,** 1602–1623 (2010).
8. Reed, C. J., Lewis, H., Trejo, E., Winston, V. & Evilia, C. Protein Adaptations in Archaeal Extremophiles. *Archaea* **2013,** 1–14 (2013).

9. Wickramasinghe, C. Bacterial morphologies supporting cometary panspermia: a reappraisal. *Int. J. Astrobiol.* **10,** 25–30 (2011).

10. Hough, D. W. & Danson, M. J. Extremozymes. *Curr. Opin. Chem. Biol.* **3,** 39–46 (1999).

11. Coker, J. A. Extremophiles and biotechnology: current uses and prospects [version 1; referees: 2 approved]. *F1000Research* **5,** 396 (2016).

12. Chien, A., Edgar, D. B. & Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *J. Bacteriol.* **127,** 1550–1557 (1976).

13. Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–491 (1988).

14. Anderson, A. W., Nordan, H. C., Cain, R. F., Parrish, G. & Duggan, D. Studies on a radioresistant Micrococcus. I. Isolation, morphology, cultural characteristics, and resistance to gamma radiation. *Food Technol* **10,** 575–577 (1956).

15. Battista, J. R. AGAINST ALL ODDS: The Survival Strategies of Deinococcus radiodurans. *Annu. Rev. Microbiol.* **51,** 203–224 (1997).

16. Gerber, E. *et al. Deinococcus* as new chassis for industrial biotechnology: biology, physiology and tools. *J. Appl. Microbiol.* **119,** 1–10 (2015).

17. Makarova, K. S. *et al.* Genome of the Extremely Radiation-Resistant Bacterium Deinococcus radiodurans Viewed from the Perspective of Comparative Genomics. *Microbiol. Mol. Biol. Rev.* **65,** 44–79 (2001).

18. Moseley, B. E. B. & Mattingly, A. Repair of irradiated transforming deoxyribonucleic acid in wild type and a radiation-sensitive mutant of Micrococcus radiodurans. *J. Bacteriol.* **105,** 976–983 (1971).

19. Krisko, A. & Radman, M. Biology of Extreme Radiation Resistance: The Way of Deinococcus radiodurans. *Cold Spring Harb. Perspect. Biol.* **5,** a012765–a012765 (2013).

20. Hansen, M. T. Multiplicity of genome equivalents in the radiation-resistant bacterium Micrococcus radiodurans. *J. Bacteriol.* **134,** 71–75 (1978).

21. Mattimore, V. & Battista, J. R. Radioresistance of Deinococcus radiodurans: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *J. Bacteriol.* **178,** 633–637 (1996).

22. Daly, M. J. *et al.* Small-Molecule Antioxidant Proteome-Shields in Deinococcus radiodurans. *PLoS ONE* **5,** e12570 (2010).

23. de Groot, A. *et al.* Alliance of Proteomics and Genomics to Unravel the Specificities of Sahara Bacterium Deinococcus deserti. *PLoS Genet.* **5,** e1000434 (2009).

24. Krisko, A. & Radman, M. Protein damage and death by radiation in Escherichia coli and Deinococcus radiodurans. *Proc. Natl. Acad. Sci.* **107,** 14373–14377 (2010).

25. Leonetti, J.-P. & Matic, I. Use of bacteria for the production of bioenergy. (2009).

26. Brim, H. *et al.* Engineering Deinococcus radiodurans for metal remediation in radioactive mixed waste environments. *Nat Biotech* **18,** 85–90 (2000).

27. Wilkins, M. R. *et al.* From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. *Nat Biotech* **14,** 61–65 (1996).

28. Anderson, N. L. & Anderson, N. G. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19,** 1853–1861 (1998).

29. Chandramouli, K. & Qian, P.-Y. Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Hum. Genomics Proteomics* **2009,** 1–22 (2009).

30. Cravatt, B. F., Simon, G. M. & Yates III, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **450,** 991–1000 (2007).

31. Graham, C., McMullan, G. & Graham, R. L. J. Proteomics in the microbial sciences. *Bioeng. Bugs* **2,** 17–30 (2011).

32. Yates, J. R. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **33,** 1–19 (1998).

33. Gregorich, Z. R., Chang, Y.-H. & Ge, Y. Proteomics in heart failure: top-down or bottom-up? *Pflüg. Arch. - Eur. J. Physiol.* **466,** 1199–1209 (2014).

34. Anderson, N. L., Matheson, A. D. & Steiner, S. Proteomics: applications in basic and applied biology. *Curr. Opin. Biotechnol.* **11,** 408–412 (2000).

35. Mallick, P. & Kuster, B. Proteomics: a pragmatic perspective. *Nat. Biotechnol.* **28,** 695–709 (2010).

36. Zybailov, B. *et al.* Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces c erevisiae*. *J. Proteome Res.* **5,** 2339–2347 (2006).

37. Fischer, F. & Poetsch, A. Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome Sci.* **4,** 1 (2006).

38. Stein, L. Genome annotation: from sequence to biology. *Nat. Rev. Genet.* **2,** 493–503 (2001).

39. Clark, J. I. It's All GO for Plant Scientists. *PLANT Physiol.* **138,** 1268–1279 (2005).

40. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32,** 258D–261 (2004).

41. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43,** D1049–D1056 (2015).

42. Armengaud, J., Christie-Oleza, J. A., Clair, G., Malard, V. & Duport, C. Exoproteomics: exploring the world around biological systems. *Expert Rev. Proteomics* **9,** 561–575 (2012).

43. Weckwerth, W. Metabolomics: an integral technique in systems biology. *Bioanalysis* **2,** 829–836 (2010).

44. Camerini, S. & Mauri, P. The role of protein and peptide separation before mass spectrometry analysis in clinical proteomics. *J. Chromatogr. A* **1381,** 1–12 (2015).

45. Saravanan, R. S. & Rose, J. K. C. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *PROTEOMICS* **4,** 2522–2532 (2004).

46. Farci, D. *et al.* Purification and characterization of DR_2577 (SlpA) a major S-layer protein from Deinococcus radiodurans. *Front. Microbiol.* **6,** (2015).

47. Horwich, A. L., Fenton, W. A., Chapman, E. & Farr, G. W. Two Families of Chaperonin: Physiology and Mechanism. *Annu. Rev. Cell Dev. Biol.* **23,** 115–145 (2007).

48. Airo, A., Chan, S. L., Martinez, Z., Platt, M. O. & Trent, J. D. Heat shock and cold shock in Deinococcus radiodurans. *Cell Biochem. Biophys.* **40,** 277–288 (2004).

49. Sergiev, P. V., Bogdanov, A. A. & Dontsova, O. A. How can elongation factors EF-G and EF-Tu discriminate the functional state of the ribosome using the same binding site? *FEBS Lett.* **579,** 5439–5442 (2005).

50. Das, A. D. & Misra, H. S. Hypothetical Proteins Present During Recovery Phase of Radiation Resistant Bacterium Deinococcus radiodurans are Under Purifying Selection. *J. Mol. Evol.* **77,** 31–42 (2013).

51. Anantharaman, V. & Aravind, L. The PRC-barrel: a widespread, conserved domain shared by photosynthetic reaction center subunits and proteins of RNA metabolism. *Genome Biol.* **3,** 1 (2002).

52. Schmid, A. K., Howell, H. A., Battista, J. R., Peterson, S. N. & Lidstrom, M. E. Global Transcriptional and Proteomic Analysis of the Sig1 Heat Shock Regulon of Deinococcus radiodurans. *J. Bacteriol.* **187,** 3339–3351 (2005).

53. Weckwerth, W., Wenzel, K. & Fiehn, O. Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *PROTEOMICS* **4,** 78–83 (2004).

54. Valledor, L. *et al.* A universal protocol for the combined isolation of metabolites, DNA, long RNAs, small RNAs, and proteins from plants and microorganisms. *Plant J.* **79,** 173–180 (2014).

7 Supplementary Materials

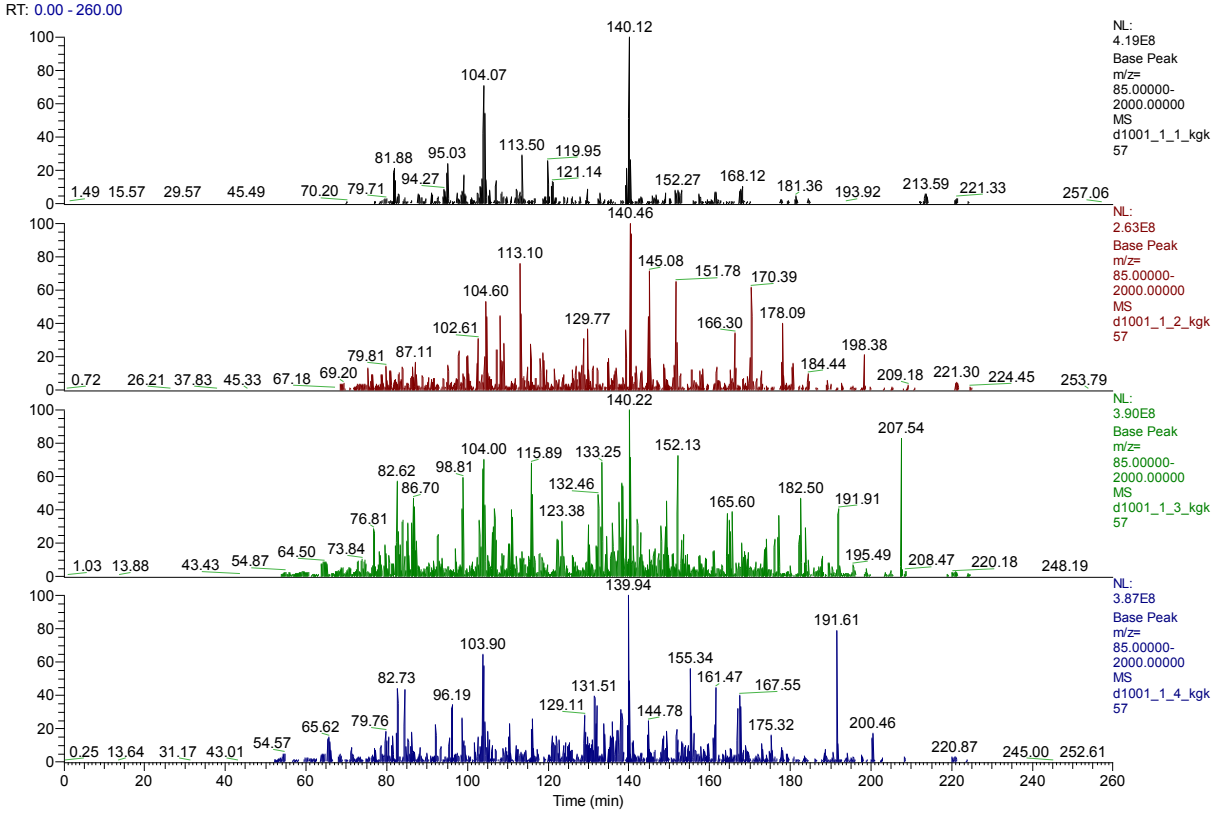

Figure 24: Base peak chromatograms of TCA 100 µg method

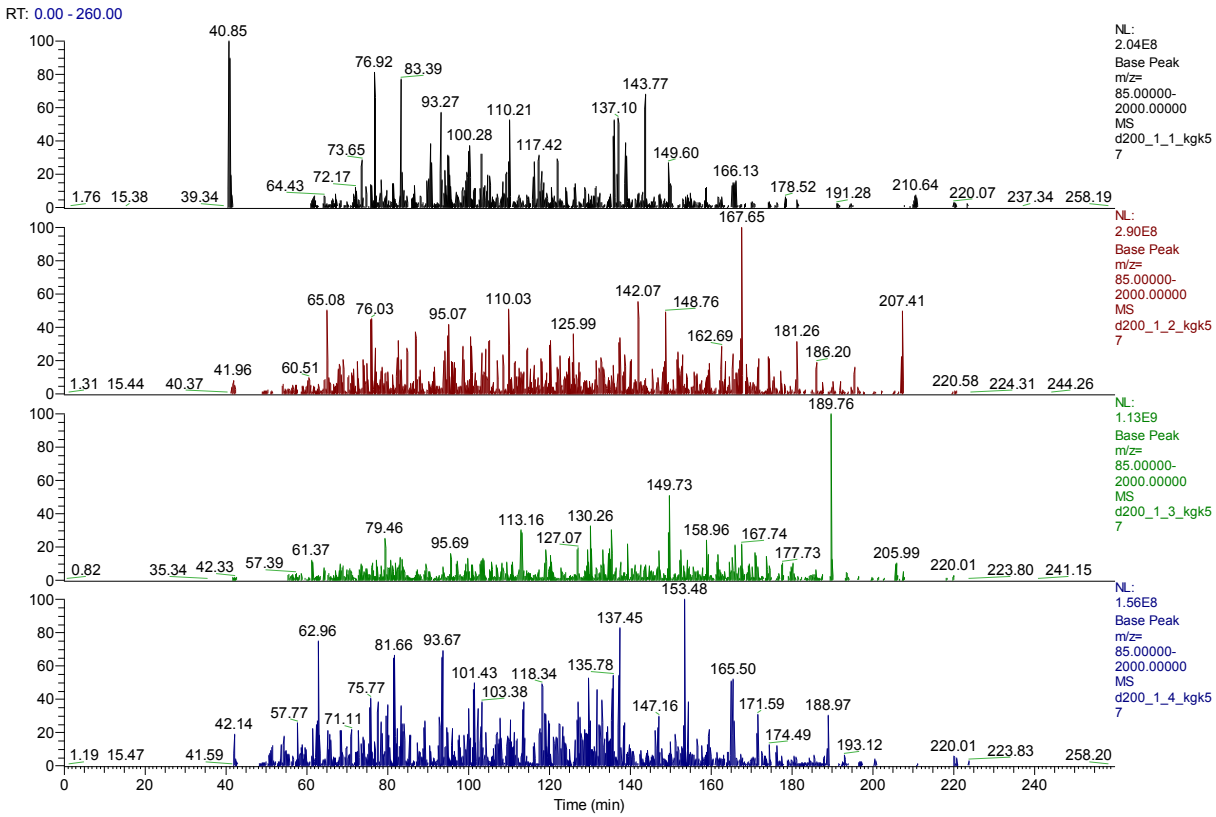

Figure 25: Base peak chromatograms of TCA 200 µg method

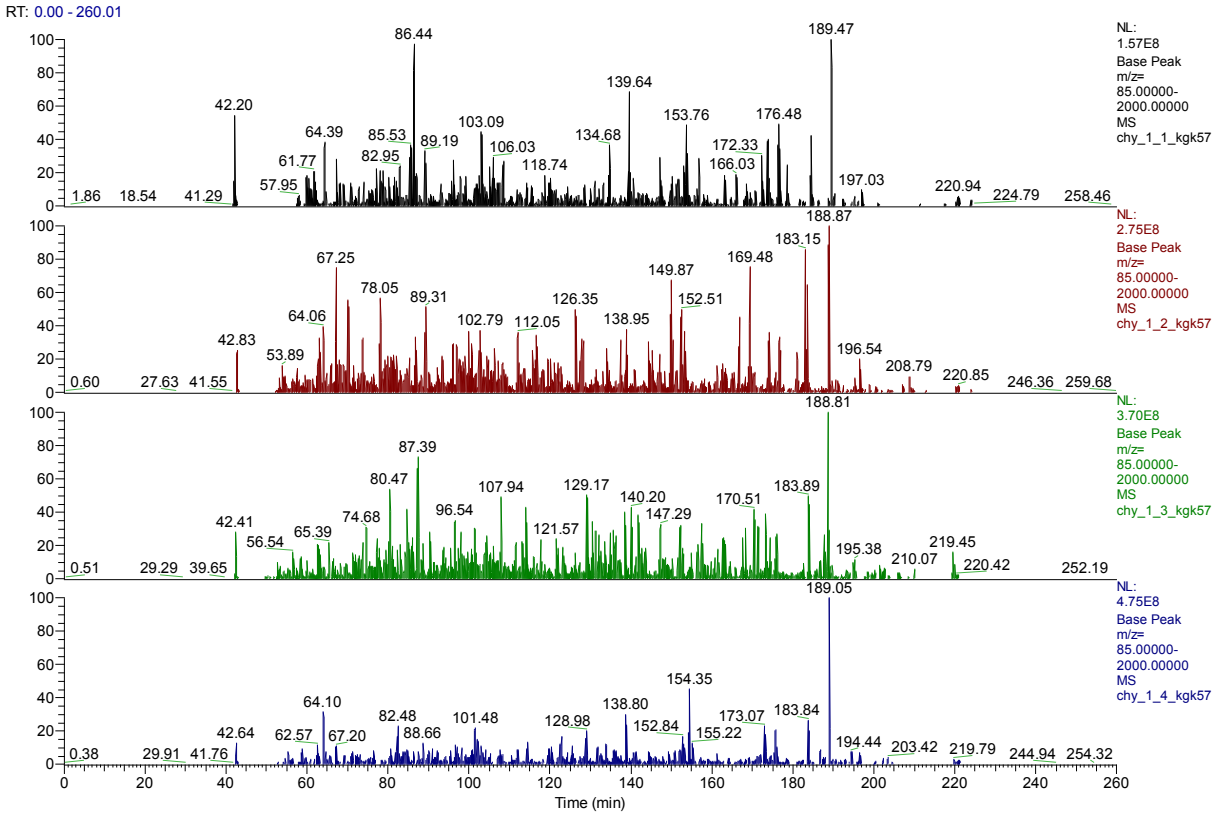

Figure 26: Base peak chromatograms of Chy/Try method

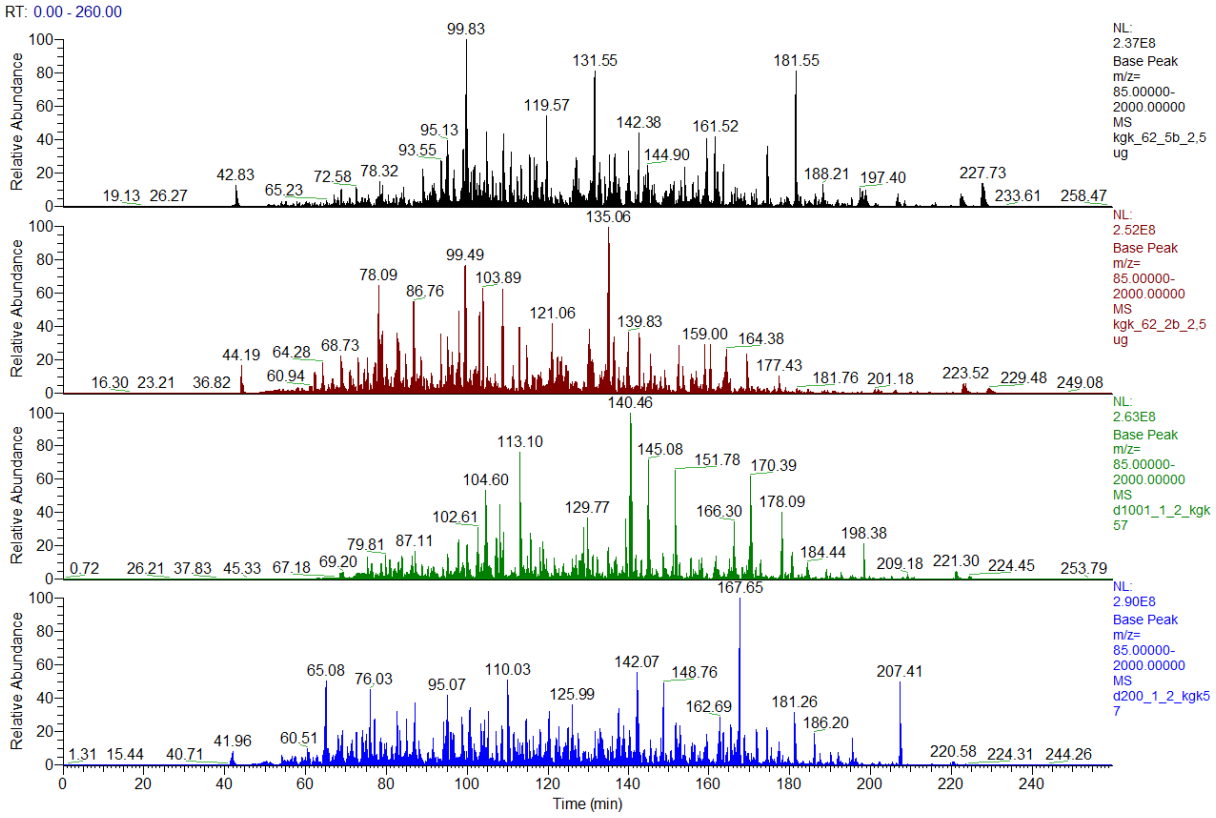

Figure 27: Base peak chromatograms of MCW - Phenol method

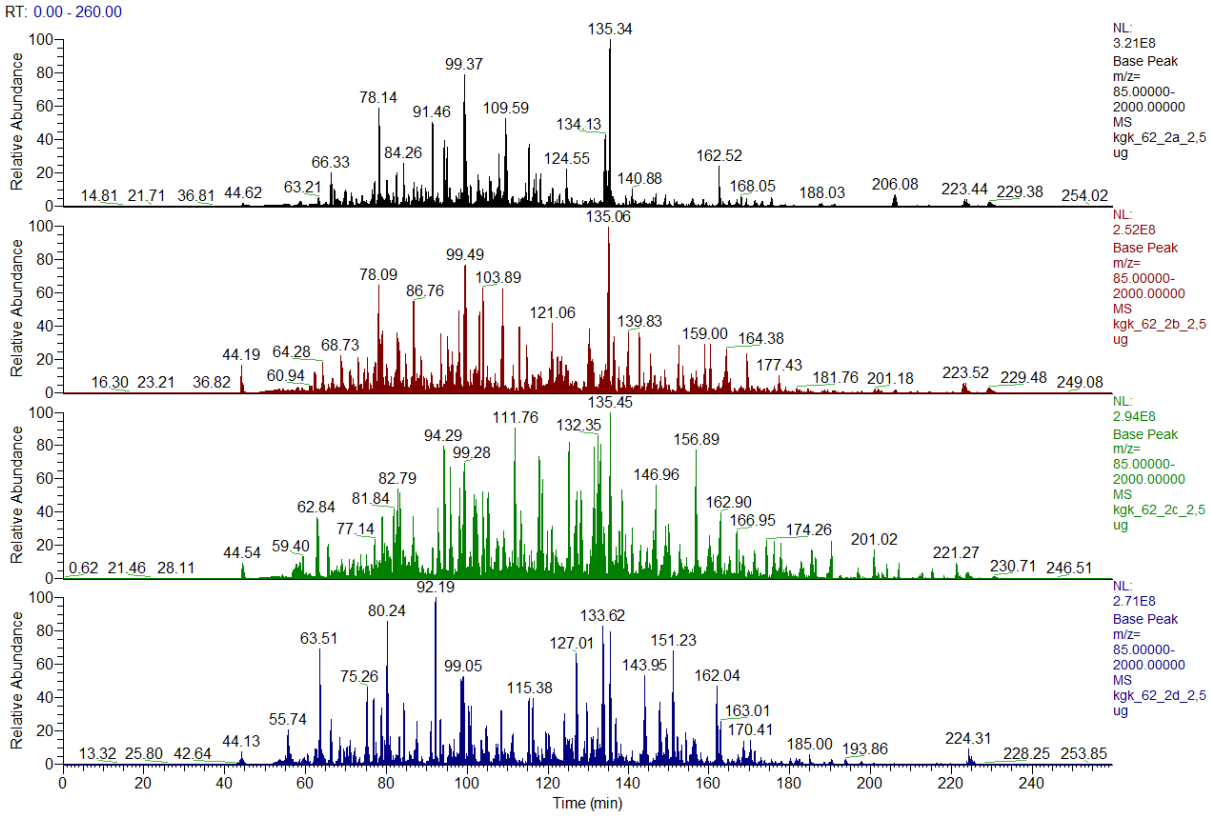

Figure 28: Base peak chromatograms of phenol method

Accession	Score	Coverage	
P56867	668.56	47.78 %	Hexagonally packed intermediate-layer surface protein
Q9RRB6	489.85	49.44 %	S-layer protein SlpA
Q9RS05	254.71	42.71 %	Serine protease, subtilase family, C-terminal
Q9RXD5	180.20	44.56 %	Outer membrane protein (Uncharacterized protein)
Q9RVA5	154.20	28.61 %	S-layer protein (SLH family protein)
Q9RWG0	148.88	32.47 %	Uncharacterized protein
Q9RZQ3	119.63	18.84 %	Extracellular nuclease, putative (Nuclease)
Q9RVP1	90.50	23.07 %	Extracellular solute binding protein, family 5
Q9RWW7	84.84	10.19 %	Pilin, type IV, putative
Q9RS04	73.77	11.65 %	Serine protease, subtilase family, N-terminal
Q9RV45	69.12	55.04 %	S-layer protein (S-layer-like array-related protein)
Q9RVQ0	68.98	37.66 %	Uncharacterized protein
Q9RW57	55.84	15.06 %	Serine protease (Serine protease, subtilase family)
Q9RYM9	55.06	27.78 %	Uncharacterized protein DR_A0282
Q9RWG1	53.37	14.21 %	Uncharacterized protein

Q9RTV1	52.12	24.63 %	ABC transporter, periplasmic substrate-binding protein, putative
Q9RWI3	51.16	17.25 %	Uncharacterized protein
Q9RUE0	50.70	31.40 %	Uncharacterized protein
Q9RX10	49.12	18.66 %	5`-nucleotidase family protein (Bifunctional metallophosphatase/5'-nucleotidase)
Q9RYL0	46.97	20.52 %	Uncharacterized protein
Q9RRG5	46.55	43.09 %	Uncharacterized protein
Q9RYQ9	46.33	15.69 %	Zinc metalloendopeptidase, leishmanolysin family
Q9RR60	44.59	15.64 %	Enolase (EC 4.2.1.11) (2-phospho-D-glycerate hydro-lyase) (2-phosphoglycerate dehydratase)
Q9RRF0	41.11	31.82 %	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)
Q9RU24	38.32	17.17 %	Probable ABC transporter-binding protein DR_1571
Q9RTR8	38.15	15.76 %	Uncharacterized protein
Q9RUY5	36.91	13.94 %	Uncharacterized protein
Q9RXD2	35.71	34.64 %	S-layer protein (S-layer-like array-related protein)
Q9RYN0	35.15	27.43 %	Uncharacterized protein
Q9RWU1	34.94	10.54 %	Uncharacterized protein
Q9RXF0	34.85	15.78 %	Peptide ABC transporter, periplasmic peptide-binding protein, putative
Q9RTV5	34.60	12.23 %	Immunogenic protein (TRAP transporter)
Q9RX56	34.23	25.85 %	Uncharacterized protein
Q9RS17	33.91	7.97 %	Uncharacterized protein
Q9R342	32.38	7.41 %	Elongation factor Tu (EF-Tu)
Q9RSA5	30.50	56.02 %	Chemical-damaging agent resistance protein C (Tellurium resistance protein TerD)
Q9RT97	30.38	21.28 %	Penicillin-binding protein 2
Q9RTK5	30.20	8.33 %	Uncharacterized protein
Q9RY36	29.21	9.46 %	Uncharacterized protein
Q9RWB6	29.11	16.03 %	Uncharacterized protein
Q9RTX7	28.27	15.89 %	Uncharacterized protein

Q9RSD6	27.84	6.47 %	Aminopeptidase (Peptidase M29)
Q9RZS3	27.00	10.90 %	Uncharacterized protein
Q9RTC3	26.96	23.19 %	Uncharacterized protein
Q9RUA2	26.70	25.23 %	Uncharacterized protein
Q9RTQ7	26.37	22.52 %	Uncharacterized protein
Q9RTQ0	25.99	26.02 %	Uncharacterized protein
Q9RT72	25.81	10.29 %	Peptidyl-prolyl cis-trans isomerase, cyclophilin-type (Peptidylprolyl isomerase)
Q9RSP6	25.81	9.70 %	Lipase (Lipase, putative)
Q9RTN7	24.95	5.63 %	Aconitate hydratase A
Q9RUR2	24.90	28.41 %	Uncharacterized protein
Q9RUP1	24.79	15.45 %	Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.-)
Q9RXI8	24.67	28.18 %	Malate dehydrogenase (EC 1.1.1.37)
Q9RR98	24.49	12.14 %	Uncharacterized protein
Q9RUR8	22.65	20.80 %	Uncharacterized protein DR_1314
Q9RU46	22.62	10.40 %	Uncharacterized protein
Q9RZQ1	22.47	10.89 %	Serine protease, subtilase family (Uncharacterized protein)
Q9RYM8	20.35	13.19 %	Probable subtilase-type serine protease
Q9RY23	20.08	8.28 %	Chaperone protein DnaK (HSP70) (Heat shock 70 kDa protein) (Heat shock protein 70)
Q9RZD6	19.90	11.01 %	Ferredoxin-nitrite reductase
Q9RZQ2	19.16	13.09 %	Uncharacterized protein
Q9RWV1	18.86	21.07 %	Amino acid ABC transporter, periplasmic amino acid-binding protein
Q9RUD9	18.57	14.32 %	Uncharacterized protein
Q9RVL8	18.38	7.60 %	Uncharacterized protein
Q9RUZ9	18.34	10.10 %	Uncharacterized protein
Q9RXK5	18.03	9.31 %	Elongation factor G (EF-G)
Q9RWI0	17.87	4.52 %	Uncharacterized protein
Q9RWI2	17.67	7.31 %	Uncharacterized protein
Q9RSA0	17.51	9.51 %	Probable 2,3-bisphosphoglycerate-independent

			phosphoglycerate mutase
Q9RWB2	17.26	14.32 %	Citrate synthase (EC 2.3.3.16)
Q9RUY1	16.75	22.92 %	Uncharacterized protein
Q9RUS6	16.32	13.62 %	Uncharacterized protein
Q9RTE5	15.00	8.47 %	Uncharacterized protein
Q9RXT9	14.90	22.38 %	Thiosulfate sulfurtransferase (EC 2.8.1.1) (Rhodanese-like protein)
Q9RU17	14.82	10.84 %	Uncharacterized protein
Q9RWB7	14.65	15.58 %	Site-determining protein
Q9RT25	13.31	37.68 %	Uncharacterized protein
Q9RZ90	13.19	5.79 %	Serine protease (Serine protease, subtilase family)
Q9RVS8	12.96	31.21 %	Thioredoxin
Q9RRH4	12.57	21.41 %	Uncharacterized protein
Q9RZ02	12.45	9.14 %	Urocanate hydratase (Urocanase) (EC 4.2.1.49) (Imidazolonepropionate hydrolase)
Q9RRU1	12.40	7.97 %	N-acetylmuramoyl-L-alanine amidase
Q9RYN6	12.24	5.32 %	Flavin monoamine oxidase (Flavin monoamine oxidase-related protein)
Q9RVK6	12.14	18.56 %	Uncharacterized protein
Q9RU58	11.99	5.05 %	Serine protease, subtilase family
Q9RTU7	11.51	8.18 %	Oligopeptidase A
Q9RX29	11.46	23.12 %	Uncharacterized protein
Q9RUF8	10.70	9.95 %	Acetyl-CoA acetyltransferase
Q9RRX1	10.43	6.37 %	Acyl-CoA dehydrogenase (Acyl-CoA dehydrogenase, putative)
Q9RSR9	10.43	4.71 %	Oligoendopeptidase F, putative
Q9RUP2	10.08	6.81 %	Phosphoglycerate kinase (EC 2.7.2.3)
Q9RR70	9.74	4.53 %	Fumarate hydratase class II (Fumarase C) (EC 4.2.1.2)
Q9RT29	9.50	5.71 %	META domain-containing protein (Uncharacterized protein)
Q9RSU7	9.13	10.02 %	ArgE/DapE/Acy1 family protein (Peptidase M20)
Q9RU28	8.59	9.83 %	Peptide ABC transporter ATP-binding protein (Peptide ABC

			transporter, ATP-binding protein)
Q9RY79	8.30	3.99 %	Peptidylprolyl isomerase (EC 5.2.1.8)
Q9RRA6	8.26	11.04 %	Iron ABC transporter, periplasmic substrate-binding protein, putative
Q9RU54	8.22	7.21 %	Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)
Q9RRI3	8.04	3.76 %	Medium-chain fatty acid--CoA ligase
Q9RY35	7.96	17.83 %	Uncharacterized protein
Q9RY39	7.90	10.88 %	Glutamine cyclotransferase
Q59337	7.24	4.48 %	Catalase (EC 1.11.1.6)
Q9RYU9	7.14	5.86 %	Uncharacterized protein
Q9RT74	7.03	5.40 %	Tetratricopeptide repeat family protein
Q9RTP9	6.92	11.36 %	Uncharacterized protein
Q9RRG3	6.79	17.93 %	Uncharacterized protein
Q9RVQ3	6.57	4.91 %	Uncharacterized protein
Q9RS64	5.83	14.01 %	DNA protection during starvation protein 1 (EC 1.16.-.-)
Q9RUR5	5.72	48.72 %	Uncharacterized protein
Q9RUD0	5.66	8.77 %	Serine protease, subtilase family
Q9RX57	5.43	5.13 %	Uncharacterized protein
Q9RW56	5.40	5.74 %	1-pyrroline-5-carboxylate dehydrogenase
Q9RRI4	5.39	6.05 %	Ferric enterobactin esterase-related protein
Q9RWG7	4.87	6.37 %	V-type ATP synthase beta chain (V-ATPase subunit B)
Q9RYZ6	4.84	8.19 %	Phosphate-binding protein
Q9RYX4	4.72	4.10 %	Probable guanine deaminase
Q9RXP5	4.68	2.50 %	Glycogen operon protein GlgX homolog (EC 3.2.1.-)
Q9RU43	4.60	5.37 %	Carboxyl-terminal protease, putative
Q9RXR9	4.60	15.54 %	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)
Q9RWP9	4.47	2.51 %	Metalloprotease, putative
Q9RTL8	4.46	3.70 %	Glucose-6-phosphate isomerase (GPI)
Q9RUH9	4.45	8.22 %	Cell ssuface protein containing Ig-like domain protein (Uncharacterized protein)

Table 9: List of identified exoproteins with pilot experiment