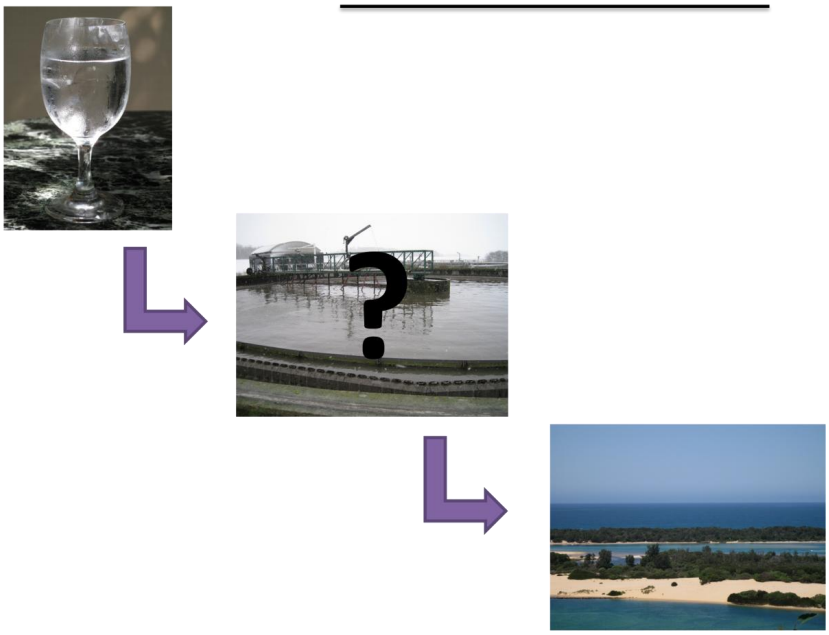
Metaproteomics

Biological origin of extracellular polymeric substance proteins in wastewater treatment plants Aalborg West, and Aalborg East



Master thesis by
Michael Kruse Olesen
Section of Biotechnology
Aalborg University

January 2012

Waste water treatmeant is an important part of the modern society and sustainability if we want to keep having pure drinking water, and preserve beautiful places like Lakes Entrace, Australia

Section of Biotechnology
Aalborg University
Biotechnology
http://bio.aau.dk/

Title:
Biological origin of extracellular polymeric substances in wastewater treatment plants Aalborg West, and Aalborg East

Subject:
Metaproteomics

Project period:
1 semester
K10, 10th semester 2011
Septemper 1st - January 18th 2012

Project group:
Master Student

Group members:
Michael Kruse Olesen

Supervisor(s):
Associate Professor Allan Stensballe
Professor Per Halkjær Nielsen

Report page count: 42

Appendix page count: 31

Total page count: 94

Preface

This project has been composed during the K10 project period from the 1st of September to the 18th of January 2012, by student Michael Kruse Olesen at the 10th semester of Biotechnology Engineering at the Technical-Scientific Faculty at Aalborg University. The main subject of the 10th semester was to produce the master thesis under which the project ”Metaproteomics” - Biological origin of extracellular polymeric substances in wastewater treatment plants Aalborg West, and Aalborg East" was undertaken.

The level of knowledge necessary to understand this report is therefore based on having successfully completed equivalent work in mass spectrometry, and waste water treatment. It is assumed that the reader is familiar with existing mass spectrometers, such as quadrupoles, time-of-flight, ion traps, and hybrid instruments that combines these mass analyzers into tandem mass spectrometry, as these topics will be discussed only in relation to the applied field of study. In addition, a common knowledge of wastewater treatment is recommended for comprehensive understanding of this project.

Books, websites, and articles are referred to in the text by the main author/s and the year of publication in parenthesis.
Figures made by the student have no references unless re-created from the literature.

All laboratory experiments carried out during the project period have been conducted by the student himself, under appropriate supervision.

Aalborg University, January 18th 2012

Michael Kruse Olesen

Abstract

The operation of waste water treatment plants (WWTP) was discovered decades ago by Srinath et al. (1959), and is now a well characterized, and established method in civil engineering. However, the biological processes in wastewater treatment is not fully understood, and increased knowledge is required to further develop new, and existing WWTP's. Improved control of the WWTP's can be established by learning more about the community ecology of the different WWTP's, which is the main goal for the EcoDesign project. In recent years, the application of genomics, transcriptomics, and proteomics have undergone continuing development, and the fields are maturing to encompass more complex systems, among others WWTP's. Since proteins expression is a direct reflection of microbial activity, they have a major importance in the function and stability of WWTP's. In this research, the application of protein extraction is investigated, and proteins associated with the biofilm formation is sought to be identified. These protein exists in the biofilm matrix, which consists of extracellular polymeric substances (EPS), in which proteins are included. Crude extracts were run on sodium dodecyl sulfate polyacrylamide (SDS-PAGE) gels, and analyzed by tandem mass spectrometry (LC-MS/MS). It was found that the extraction methods did not seem to be sufficient, which resulted in only few identified proteins. However, human Elastase-3A where identified in all applied extraction methods providing more evidence of the existence of this resistant protein in WWTP's. This report provides the current knowledge, and results obtained in this exciting field of study, and has already motivated further investigations.

Synopsis (Danish)

Driften af rensningsanlæg blev opdaget årtier siden af (Srinath et al., 1959), og er nu en velkarakteriseret og etableret metode inden for ingeniørfaget. Imidlertid er de biologiske processer i spildevandsresning ikke helt forstået og øget viden er nødvendig for at videreudkvikle nye og eksisterende renseanlæg. Forbedret styring af renseanlæg kan etableres ved at lære mere om det mikrobiologiske samfund i de forskellige renseanlæg, som er målet for EcoDesign projektet. I de senere år har anvendelsen af genomforskning, transkriptomforskning, og proteinforskning gennemgået en løbende udvikling, og disse metoder er ved at modnes til at omfatte mere komplekse systemer, blandt andet renseanlæg. Da proteiner direkte afspejler den mikrobielle aktivitet, har de en stor betydning i funktionen og stabiliteten i renseanlæg. I dette studie er anvendelsen af opresningmetoder forsøgt undersøgt, og proteiner associeret med biofilmdannelse søges at identificeres. Disse proteiner findes i biofilm-matricen, der består af ekstracellulære polymeriske stoffer, hvor i proteinerne er inkluderet. Råekstrakter blev kørt på natrium dodecyl sulfat polyakrylamid geler, og analyseret med hybrid masse spectrometri. Det blev konstateret at ekstraktionsmetoderne ikke synes at være tilstrækkelige, hvilket resulterede i kun få identificerede proteiner. Dog blev menneske Elastase-3A identificeret, og dermed gav mere bevis for dens eksistens i renseanlæg. Denne rapport viser vores nuværende status i dette spændende felt, og har allerede motiveret yderligere undersøgelser.

Contents

Preface **I**

Abstract **III**

Synopsis (Danish) **V**

Contents **VII**

1 Introduction **1**
1.1 Wastewater Treatment Principle 1
1.2 Application of High-Throughput Methods in Environmental Research 2
1.3 Mass Spectrometry Based Proteomics 3
1.4 Metaproteomics 5
1.5 Microorganisms Lives in Matrices of Extracellular Polymeric Substances 7
1.6 Current Environmental Issues Resolved by Biomembrane Bioreactors 10

2 Aim and Strategy **13**

3 Materials & Methods **17**
3.1 Methods 17

4 Results and Discussion **23**
4.1 Background of the two Investigated WWTP's 23
4.2 SDS-PAGE Analysis of the Extracted EPS Proteins - Strategy 1 23
4.3 Activated Sludge Binding Assay 27
4.4 Total Protein Extraction 32
4.5 Mass Spectrometry Data 34
4.6 Significance of Study 38

5 Conclusion **39**

6 Future Perspectives **41**

A Selected Metaprotemic Studies of Waste Water Treatment Plants **43**
A.1 Extraction by Wilmes and Bond (2004, 2006b); Wilmes et al. (2008) 44
A.2 Extractions by Park and Novak (2007); Park et al. (2008) 47
A.3 Extractions by Kuhn et al. (2011) 48

B Proteomics **51**
B.1 Quantitative Mass Spectrometry 53
B.2 Validation of High-throughput Proteomics 58
B.3 Metaproteomics 60

C Measured Protein Quantities from Selected Stuides **63**

D Quadrupole and Quadrupole Ion Trap Theory **65**

Bibliography **73**

1 Introduction

Waste water treatment is an important part of the modern society and sustainability, as it is necessary to purify the wastewater from municipal and industry sewages before releasing it into the lakes, rivers, and oceans of the world to avoid eutrophication (Oehmen et al., 2007). The collective effects of increased amounts of phosphorous, which is a key nutrient in aquatic water systems, have been observed to encourage growth of higher plants species, and blue-green algae, which ultimates distorts the natural competition between species in the water system, and this detrimental effect is referred to as eutrophication (Mainstone and Parr, 2002) (Oehmen et al., 2007). The adverse impacts of eutrophication includes extinction or reduction of various species, reduced penetration of sun light into the water system, and periods of very limited available oxygen (Mainstone and Parr, 2002), which is a must to rectify this issue.

1.1 Wastewater Treatment Principle

The operation of waste water treatment plants (WWTP) was discovered decades ago by Srinath et al. (1959), and is now a well characterized, and established method in civil engineering (Tchobanoglous et al., 2004). However, the biological processes in wastewater treatment is not fully understood (Oehmen et al., 2007). One of the main goals in wastewater treatment is the removal of phosphorous (P), and a widely used method is enhanced biological phosphorous removal (EBPR), which is simplified in Figure 1.1. Sewage water enters the WWTP, where it is mechanically filtered to remove large particles, and subsequently sand, and grease is separated in the primary settler (not shown on the figure). The bulk of the resulting waste water in Danish WWTP's is protein (25-35%), lipids (25-35%), carbohydrates (15-25%), and some complex macromolecules such as nucleic acids (Nielsen et al., 2010). These are transported into the WWTP, where they are degraded through various pathways. Though Figure 1.1 shows the anoxic, and aerobic tanks as separate this is often achieved in full scale plants by turning the aeration on/off. Thus, the resulting waste water is channelled between anoxic and aerobic conditions, which facilitates the dominant processes of nitrification/denetrification (Nielsen et al., 2010). After a defined number of cycling the activated sludge is transferred to the clarifier, where the sludge is allowed to settle to the buttom of the tanks leaving freshwater in the top that can be discharged into the aquatic water systems (Tchobanoglous et al., 2004). The activated sludge is then used to reinoculate new anoxic/aerobic tanks, and the excess sludge is either dewatered and used as fertilizer or transferred into thermofile digesters, which produces biogas for electricity production. An interesting note is that in this setup the return sludge is exposed to a prolonged anaeroibc phase, which favors the microorganisms that contribute to the EBPR process (Nielsen et al., 2010).

When the EBPR process is succesfully applied it is an efficient, and cheap method to remove phosphorus (P), but it has been observed in several studies that the performance may deteriorate even when the optimal conditions are sought to be applied (Oehmen et al., 2007). These issues is often a result of the changes in the community of microorganisms and the ability of activated sludge flocs to floculate (Park et al., 2008). Causes of deteriorating WWTP's can be caused by domination of unwanted microorganisms that does not contribute to EBPR, and compete for substrates (Bond et al., 1999; Blackall et al., 2002), or poor settling ability of the activated sludge flocs (Wilen et al., 2003) (Wilen et al., 2004). These unforseen, and unpredictable issues ultimately result in a poor performing WWTP, and therefore there is a need to improve the performance by

learning more about the system.

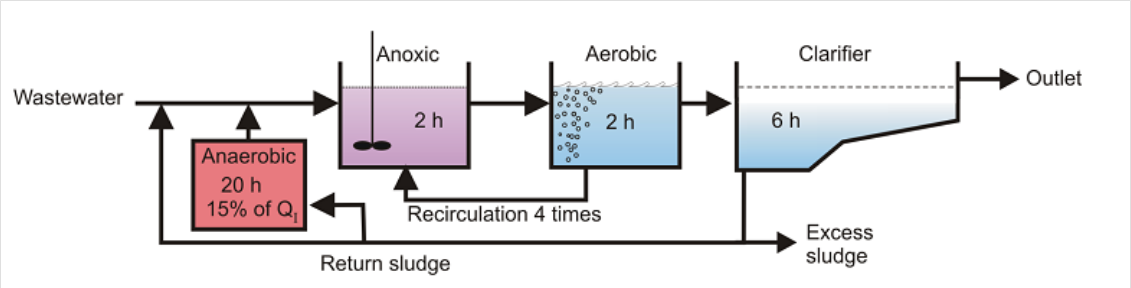


Figure 1.1: Simple schematic of a side-stream wastewater treatment plant: the wastewater is lead into an anaerobic digester, then into a cycling anoxic/aerobic reactor. Subsequently, the wastewater is transferred to the clarifier in which the activated sludge is allowed to settle to the bottom, which produces pure water in the top. Excess sludge is returned to the anaerobic digester. (Picture by Poul Larsen)

Improved control of the WWTP's can be established by learning more about the community ecology of the different WWTP's, which is the main goal for the EcoDesign project. High throughput strategies are great tools in ecodesigning, which is shown in Figure 1.2. For example, research projects include better control of the community composition, and the degradation of micropollutants. Years of research have been compiled into conceptual models, which is now being used as a basis to implement more advanced molecular methods such as genomics, transcriptomics, and proteomics.

1.2 Application of High-Throughput Methods in Environmental Research

In recent years, the application of genomics, transcriptomics, and proteomics have undergone continuing development, and the fields are maturing to encompass more complex systems, among others WWTP's (Wilmes and Bond, 2006b). The introduction of high throughput methods in environmental studies have made it possible to identify, classify, and quantify micro-communities at the molecular level (VerBerkmoes et al., 2009). Originally, the methods were only used for pure culture samples, but the methods have since been applied to community ecology studies. Applying these methods to the more complex systems has spawned the use of meta- as a prefix to emphasize that mixed species are involved (Wilmes and Bond, 2004; Martin et al., 2006). The workflow of these methods often involves sampling, purification, and then the subsequent use of one or several of the high-throughput methods, shown in Figure 1.2, which produces massive amounts of data for subsequent analysis (Wilmes and Bond, 2004). In this research the interest is proteins, which are the chemical active part of the microorganisms in the activated sludge. The proteins are responsible for the metabolic reactions that enables, e.g. substrate uptake, energy production, degradation of biotoxins, and mobility (Lodish et al., 2007).

Due to the vast number of abilities the protens are a very important part of the microorganisms, and the study of proteins, referred to as proteomics, enables valuable insights into the abilities of the community in activated sludge systems (Wilkins et al., 1996; Wilmes and Bond, 2006b). Following the sequencing of the human genome (Venter and et. al., 2001) the molecular methods have improved, and mixed communities, such as acid mine drainage have been genomically sequenced (Tyson et al., 2004). Having provided the genome enabled the subsequent detailed analysis of

the community proteome, which was performed on the same community, which is now used as a model system for meta-omics studies (Ram et al., 2005).

Lack of genomic information complicates detailed protein analysis, though it does not preclude it. It was shown by Wilmes and Bond (2004) that by combining 2D-electrophoresis with *de novo* peptide sequencing, it is possible to obtain novel proteins from unculturable organisms in laboratory-scale EBPR sequencing batch reactors (SBR). This provided valuable insight into active metabolic pathways, and highlighted the feasibility of functional 'omics' approaches in EBPR wastewater treatment. With increased availability of genomic sequences from mixed communities it was shown that with the combined analysis of genomics, and proteomics, which is referred to as proteogenomics, even more detailed insight into complex system can be achieved (Wilmes and Bond, 2006b,a), including the reconstruction of metabolic pathways. Though huge metagenomic studies have been undertaken, e.g. sequencing the Sargosso Sea (Venter et al., 2004), it is important to realize the benefits, and limitations of these studies. The metagenomes are the enabling tool that provide valuable data information of the abundances of the microorganisms, and also the entire potential for expressed proteins in the form of the genome sequences (Martin et al., 2006). It is highly stressed that metagenomics provides potential only, and further validation is necessary to obtain true identification of proteins in an environmental sample. Mass spectrometry based proteomics enables the identification of proteins on a large-scale, and samples the proteins directly, which provides true identification of proteins (Aebersold and Mann, 2003). The proteins can then be mapped back to the genome providing a more complete picture of the actual expression of proteins in a sample, and the methods for this process is explained in the following sections.

1.3 Mass Spectrometry Based Proteomics

Mass spectrometry (MS) based proteomics is currently the favored method for protein identification, because it enables the sequencing of proteins on a large high-throughput scale, and makes it possible to determine which the genes are actively being transcribed (Aebersold and Mann, 2003). MS allows the inquiry into cellular states, by sampling the entire protein complement of a cell under a given set of physiological conditions at a given time, referred to as the proteome of the cell (Wilkins et al., 1996). Proteomics enables evaluation of differences in protein expression under different sets of defined conditions, as well as the evaluation of differences in post modifications, such as disulfide bonding positions (Gorman et al., 2002; Wu, 2008; Xiang et al., 2009), glycosylation patterns (Hägglund et al., 2007; Medzihradszky, 2008), and phosphorylation site determination (Stensballe et al., 2001a; Mann and Jensen, 2003; Palumbo et al., 2010).

Sample preparation is a crucial aspect prior to MS, and may constitute the largest part of the work of the MS strategy. Traditional strategies include either a gel-free strategy known as multidimensional protein identification technology (MudPIT), or a gel-based strategy employing either 1D SDS-PAGE, or 2D SDS-PAGE as prefractionation steps (Steen and Mann, 2004; O'Farrell, 1975). In addition, when combined with protein dyes, such as coomassie, or silver-staining, gel-based techniques allow for a direct visualization of proteins of interest on the gel (Bradford, 1976; Krystal, 1987). The 2D strategy provides the highest resolution - ≈ 5000 proteins down to less than $1\,ng$ protein per spot - and robust methods have been invented for most types of proteins (Görg et al., 2004). Though the 2D strategy provides a good resolution it is not convenient for rapid large-scale identification, and it often fails to detect very hydrophobic proteins, or proteins with unusual pI values (O'Farrell, 1975). The experimental work is often laborious, and cutting out bands from the gel can be very tedious when a typical proteome consists of an average of 2000 proteins. More recent methods, such as filter-aided sample preparation (FASP) have been suggested in order to

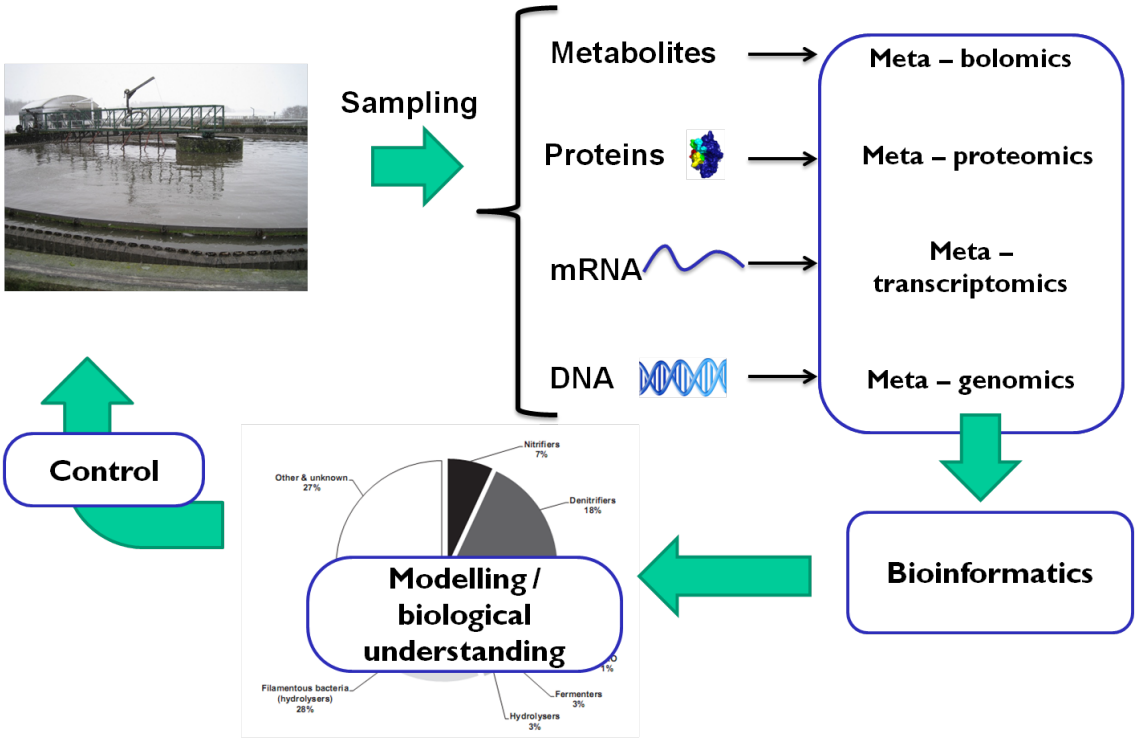


Figure 1.2: The workflow in the meta-omic strategies in community ecology. Wastewater treatments plants (WWTP's) are sampled for subsequent analysis of DNA, and mRNA by genomic sequencers, proteins by mass spectrometry, and metabolites by nuclear magnetic resonance. The high-throughput methods produce huge amounts of data, which is processed by bioinformatics. The resulting data is analyzed and inferences are made to create biologigal meaning. These approaches are thought to provide the valuable information that can be used to optimize the process operation of the WWTP's.

combine the advantages of in-gel, and in-solution digestion on filters, which provides a more rapid approach to obtain large-scale proteomes in single-run analyses (Wisniewski et al., 2009). Several studies, not only large-scales, apply the bottom-up approach. This method is commonly referred to as the shotgun proteomics approach, and it is favored for proteomics on a large-scale, because it provides rapid identification of proteins, and biological function (Steen and Mann, 2004; Marcotte, 2007). The general shotgun strategy involves three stages, shown in Figure 1.3: (1) sample preparation, and digestion, (2) LC-MS/MS, and (3) database searching, and bionformatics. The sample preparation is succeeded by digestion with proteases resulting in a peptide mixture (Link et al., 1999; Peng and Gygi, 2001). The peptides are the actual molecules, which are investigated, and the sequence of each peptide are sought to be traced back to its corresponding protein.

The peptide weight can then be analyzed through MS, and the peptides can be sequenced using tandem MS, or MS^n, where several mass analyzers are applied. In the latter technique, peptides are fragmented into product ions by applying internal energy to the precursor peptide, which provides evidence of the structure of the precursor ion (Sleno and Volmer, 2004). Tandem MS produces a vast number of spectra, and is therefore tedious to interpret manually. Therefore, database search algorithms are applied to analyze MS data. The computational analysis compare the acquired MS data with theoretical spectra, as shown in Figure 1.4. Several search algorithms exists, and each algorithm utilize different sets of strategies, e.g. spectral matching, theoretical

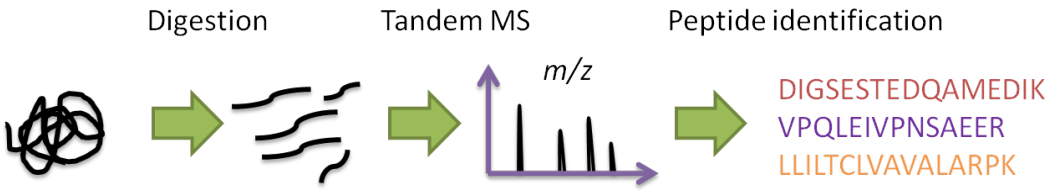


Figure 1.3: Shotgun proteomics approach: protein samples are digested with proteases, resulting in a peptide mixture. The peptides are analyzed by tandem MS, and peptide sequences are identified.

weight matching. Several excellent search engines exist, including SEQUEST (Eng. et al., 1994), MASCOT (Perkins et al., 1999), X-Tandem (Craig and Beavis, 2004), and MaxQuant (Cox and Mann, 2008). In recent years the field of proteomics have evolved, and is being applied to more complex systems including WWTP's (Wilmes and Bond, 2004), though several issues are still scrutinized. These are discussed in the following.

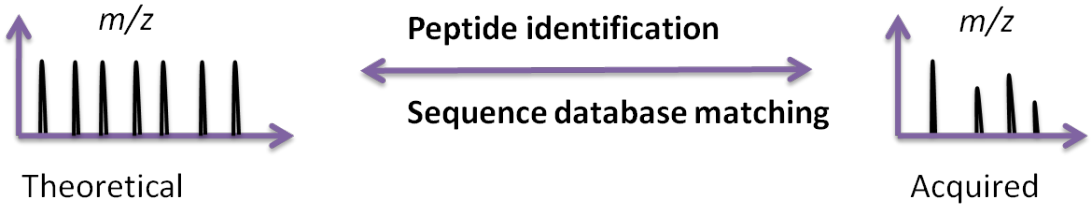


Figure 1.4: An example of spectral matching. A genomic database is used to create theoretical peptides, which are then fragmented *in silico* into theoretical MS/MS specta, which can be compared to acquired MS/MS spectra.

1.4 Metaproteomics

The feasibility of a large-scale proteomics study was shown already in 2001 on yeast (Washburn et al., 2001), and has since been applied to various other samples and organisms, including the EBPR system (Wilmes and Bond, 2004). With the parallel advances in next generation DNA sequencing the number of possible protein identifications expands rapidly each year (Mardis, 2008). Connecting metaproteomics with metagenomics, referred to as metaproteogenomics, it is possible to obtain high resolution information about the make-up and abundance of microorganisms, the functional potential, protein expressions, and lastly biochemical activity (Wilmes and Bond, 2006b; VerBerkmoes et al., 2009; Wilmes and Bond, 2009).

The application of proteogenomics was shown for acid mine drainage (AMD), which became the first mixed community to be analyzed by community genomics, and community proteomics (Tyson et al., 2004; Ram et al., 2005). It is now being used as a model system, because the community consists of only a handful of taxa, and because the different strains are genomically distinct, which makes the data processing more manageable (Tyson et al., 2004). It was possible to detect 2033 proteins from the five most abundant organisms in the growing acid mine drainage microbial biofilm, thus inferring gene expression, and also making it possible to map out active metabolic pathways *in situ* (Ram et al., 2005). In order to avoid randomly identified proteins, several studies have applied the restriction that at least two peptides must be mapped to one protein to trigger a

positive protein hit. In this way 49% of all the predicted proteins from the genomic data were identified using metaproteomics (Wilmes and Bond, 2006b), while others have applied a three peptides per protein restriction (Park et al., 2008). In addition to predicted identified proteins Ram et al. (2005) were also able to identify proteins from putative operons, thus confirming putative proteins. Thus, the detection of novel proteins is possible from metaproteomic data, which extends biological information. Several interesting observations have been made using the AMD model system. For examples, when a microorganism constitutes more than 30-40% of the entire poulation the number of protein identifications is saturated. When the abundance of the microorganism decreases, so does the number of proteins identified as well as the protein coverage. However, it was still shown that more than 100 proteins can be identified from an microorganism constituring around 1%.

The transition from proteomic analysis of single strains to several strains is not an easy task, and introduces several analytical challenges. Applying proteomics to more complex systems such as EBPR systems have been performed for lab-scale reactors (Wilmes and Bond, 2006b), but several issues still exist. The community in EBPR is not static over space and time, and each strain will change in abundance or potentially adapt to new niches in the system. This introduces a problem, because proteomics is based on the comparison of theoretical peptides obtained from genomic sequencing, with MS/MS data, which requires exact matches. Though the genome is more static compared to the proteome even single amino acid substitutions will lead to a mass difference that will prevent peptide identification (Peng and Gygi, 2001) (VerBerkmoes et al., 2009), as shown in Figure 1.5. Thus when a metagenome is constructed it only provides a snapshot of the genomically makeup at the time of sampling, and subsequent proteomic studies needs to be performed within the same timescale, to avoid the genome and proteome to drift apart. Logically, the affect of this issue is less on large, and abundant proteins, but becomes more severe for more diverged, less abundant, and smaller proteins. Though this constitutes an identification bias, it has also been used to obtain strain-variation information. Because MS is sensitive to amino-acid substitutions it is possible to perform strain-typing peptide by peptide, which makes it possible to analyze functional differences between closely related organism. This analysis is referred to as proteomics inferred genome typing (PIGT) (Denef et al., 2008), and was originally applied to the AMD system, but has also been applied for EBPR systems (Lo et al., 2007).

An issue in metaproteomics is the complexity of the system. The complexity of different environments is considered to follow the trend $soil > ocean > wastewater > AMD$ (Curtis et al., 2002). This shows that the EBPR system is not the most complex system to analyze, yet it is not the most simple either. However, the dynamic range in this system is still enormous. Each unique taxa in the systems produces unique proteins, and each protein is differentially expressed, which makes the dynamic range between low abundant proteins, and high abundant proteins huge. In addition, the intrinsic properties of electrospray, and the low coverage of the chromatography further complicates comprehensive protein identification (Michalski et al., 2011). However, these issues are continuously being reduced with improved liquid chromatography capabilities, and rapid scanning mass spectrometers, but also the post processing of data still improves with new observations and studies. Some of the issues of polymorphism, and lacking genomic data can be partially circumvented by *de novo* sequencing. This method extract peptide sequences directly from MS/MS, thus circumventing databased searches. For environmental samples where the selective pressure is high, and micro-diversity is abundant, it may be feasible to first apply database search, and then apply *de novo* sequencing for the remaining high quality spectra to increase protein identification (Nesvizhskii, 2007; Bern et al., 2007). Furthermore, protein extraction methods for environmental samples are not optimal, and it may therefore be rewarding to further analyze different constituents

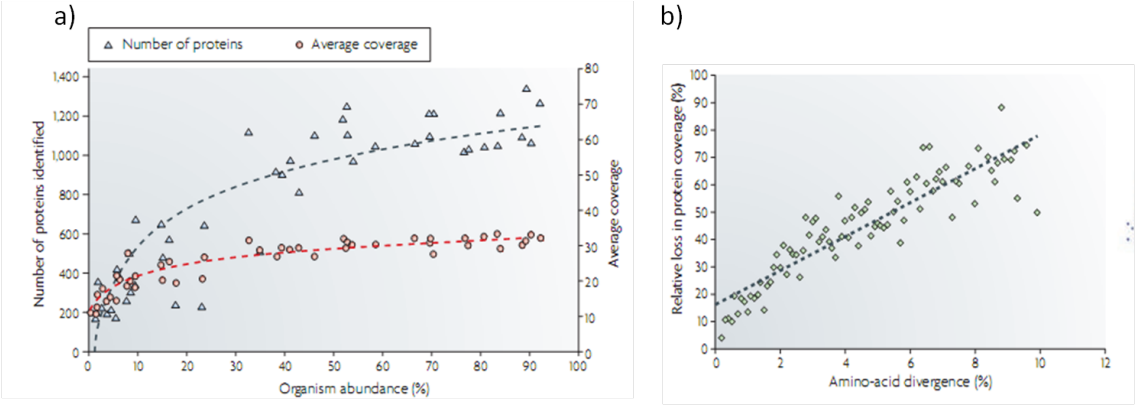


Figure 1.5: Interesting knowledge have come from the acid mine drainage, which is used as a model system for more complex systems. Two important observations are that a) the number of proteins becomes saturated when a miccroorganism becomes more abundant. Conversely, when a microorganism is less abudandant fewer proteins are identified, though it is in fact possible to observe proteins from them. b) Depicts the issues caused the by continuous adaptations, and mutations in the microorganism community. Mutations in genes causes the proteins to differ from the reference database. This phenomenon was predicted using *in silico* predictions, and shows that 50% of the protein coverage is lost, when the proteins differ in amino acids by 5-6%. Reprinted from (VerBerkmoes et al., 2009), obtained from (Denef et al., 2008)

in the system. Increasing the knowledge about the system then enables more qualified application of extraction steps for to create optimal samples for LC-MS/MS. Several studies have sought to extract EPS proteins from activated sludge systems, and these will be discussed in the following (Wilmes and Bond, 2004; Park and Novak, 2007; Kuhn et al., 2011).

1.5 Microorganisms Lives in Matrices of Extracellular Polymeric Substances

Microorganisms in WWTP's does not live in pure culture suspensions, but clog together in complex mixed communities in films, mats, flocs, etc, which is referred to as biofilm (Flemming and Wingender, 2010), as shown in Figure 1.6. The formation of biofilm creates a convenient environment for the microorganisms and allows them to gain abilities the microorganisms cannot perform in a planktonic state. For example, microorganisms are able to colonize the lungs of cystic fibrosis patients, because the form biofilms, that ensures their survival even under heavy antibiotic treatment (Yang et al., 2011). The major components of the biofilm are the extracellular materials that are secreted by the microorganisms themselves. The materials consists of several different polymeric compounds, such as carbohydrates, extracellular DNA, lipids, and proteins, which in general is referred to as extracelluar polymeric substances (EPS) (Raunkjær et al., 1994). The life inside the biofilm is convenient, because the constituents in the biofilm are retained and continuously recycled (Flemming and Wingender, 2010). In addition, the biofilm enables the retention of water in periods of starvation, and resistance to various toxins as well as preying protozoans, which means that the biofilm is both a food source, and a shielding layer (Flemming and Wingender, 2010; Yang et al., 2011).

The make up of EPS therefore has a profound impact on the stability, and function of the biofilm

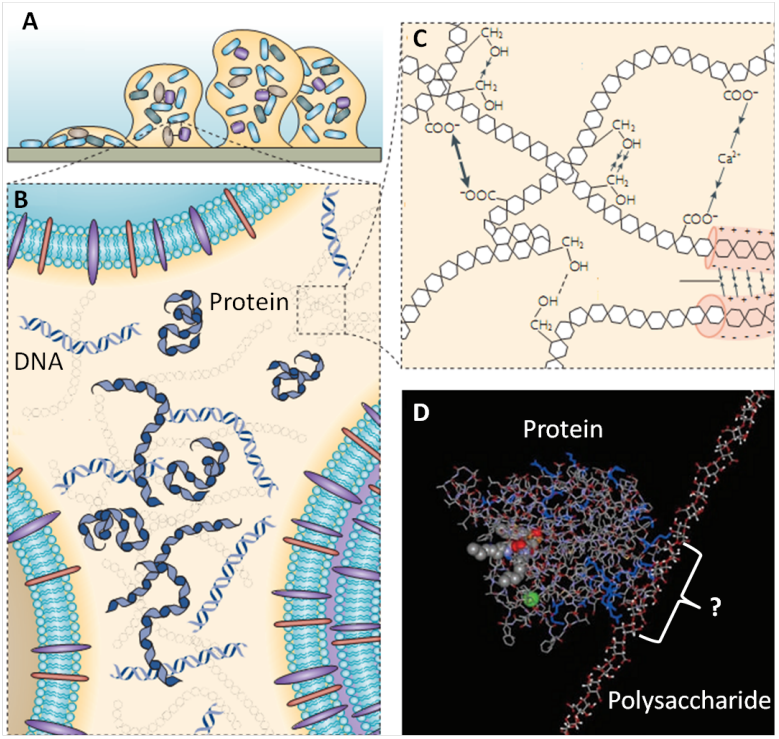


Figure 1.6: A) Biofilm initiation, and development: the microorganisms secrete exopolymers, which are used to adhere to surfaces, and other microorganisms. B) The likely structure of microorganisms in a biofilm, which shows the different extracellular polymeric substances interacting with each other, thus stabilizing the biofilm. C) Weak forces including charge-charge interactions, hydrogen bonding, Van der Waals interactions, ionic interactions, and others stabilizes the biofilm. For example high molecular weight polysaccharides are modified with carboxylic, and anionic groups that can interact with e.g. proteins as shown in D). The exact binding mechanism is unclear, but seems to be non-covalent. (Flemming and Wingender, 2010)

(Wilen et al., 2003), because the EPS is responsible for the cohesive forces that maintain the integrity of the biofilm, and binding to surfaces. Though the constituents seem to be simple polymers another picture forms when they are analyzed. For example the carbohydrates secreted in granular sludge have been shown in NMR studies to be heavily modified with e.g. carboxylic groups, acetyl groups, and other charged residues that enables intermolecular cross-linking (Seviour et al., 2010). Thus, the polysaccharides have both cationic, and anionic properties, and these properties have been shown to be important for various functions, such as adhesions, aggregation, retention of water, and protective barriers (Flemming and Wingender, 2010). Surprisingly, it has been observed that the extracellular DNA (eDNA) in biofilm not only originates from lysed cells, but is actively excreted by cells into the biofilm, and is thought to be important in the early formation of the biofilm (Whitchurch et al., 2002). The biofilm retains the DNA from cells thus enabling horizontal gene transfer (Flemming and Wingender, 2010), and have been shown to be abundant in activated sludge (Frølund et al., 1996). The eDNA seems to have three major functions in the biofilm. Firstly, it is a major structural component in biofilm, which is shown by deflocculation induced by DNase treatment (Dominiak et al., 2011). Secondly, it is used as an adhesin for attachment to various surfaces (Vilain et al., 2009), and thirdly, it possesses anti-microbial properties (Mulcahy et al., 2008). Lipids are also found in the biofilm, and they serve various

purposes, for example creation of hydrophobic interactions, and are also thought to be important for attachment and adhesion (Flemming and Wingender, 2010). They may be present in the form of lipopolysaccharides, and have been shown to have surface active properties (Matsuyama and Nakagawa, 1996), which have gained attention in the oil recovery business (Ron and Rosenberg, 2001), and as anti-microbial agents (Ron and Rosenberg, 2001). In addition, the level of saturated/unsaturated fatty acids in the matrix have been shown to vary in activated sludge according to season, which may imply adaptation to varying temperatures (Conrad et al., 2003). Extracellular proteins are present in sludge, where they create a second digestion system enabled by various EPS degrading enzymes (Frølund et al., 1995; Yu et al., 2007). These are extremely important in the biofilm, where the provide scaffolding to increase biofilm stability, and due to the vast amount of abilities, these proteins have a very interest. Consequently, these proteins have long been sought to be identified and exploited for various purposes, which is discussed in the following.

1.5.1 Extracelluar Polymeric Substance Proteins

Extracellular proteins are known to be present in large amount in biofilms, and these proteins contribute to the metabolic active community in activated sludge by performing various tasks (Frølund et al., 1996). Proteins in general can be considered to be the most important component, due to the vastness of chemical reactions carried out by proteins. Various enzymes are necessary to create the other complex polymers described above, while different conformations of residues can trap available ions, and provide structural integrity (Frølund et al., 1995). In addition to the the common residues, proteins can be post translational modified with various modifications, where the most commonly known are disulfide bonds (Gorman et al., 2002; Xiang et al., 2009; Wu, 2008), glycosylation (Hägglund et al., 2007; Medzihradszky, 2008), and phosphorylation (Stensballe et al., 2001b; Mann and Jensen, 2003), though more than hundreds exist. In addition to these covalent modifications, proteins in biofilms are able to assemble with polysaccharides to create lectins (Higgins and Novak, 1997), which are carbohydrate binding proteins, or create stable long β-sheet fibres known as amyloids (Otzen and Nielsen, 2008). Lectins are thought to be important extracellular proteins, because they posses a very high stability, i.e. they are less heat sensitive, and are more resistant to protease degradation. An example of a polysaccharide binding protein is shown in Figure 1.6 D, and though the exact binding mechanism is unknown, it is thought to be non-covalent - in contrast to glycoproteins (Broekman, 2007; Hägglund et al., 2007). Amyloids are associated with various properties, and functions, such as biofilm formation and stability, floc properties, anti-microbial activity, cell invasion, as well as neurodegenerative diseases (Larsen et al., 2008).

Some etracellular proteins are secreted to create an external digestive system, and because of the gel-like behavior of the biofilm, these enzymes need not to be anchored to the cells, because they are not able to readily diffuse (VerBerkmoes et al., 2009). The external digestive system functions to digest large polymers, that is not readily transportable through the cell membrane, into smaller fragments that can be internalized (Flemming and Wingender, 2010).

The EPS makes up most of the biofilm, but also includes water, and inorganic ions, that further contributes to the stability and integrity of the biofilm (Wilen et al., 2003). The EPS can be used as a food source in times of starvation or changes in the environment, but due to the complexity of the different EPS materials, degradation requires specialized enzymes, which is costly in terms of energy (Flemming and Wingender, 2010). The microorganisms reside in the biofilm, and make up between 10-15% of the total organic matter within the biofilm (Frølund et al., 1996). These microorganisms have an advantage as they protected against desiccation, anti-bacterial biocides,

preying protozoans, and changes in the environement (Flemming and Wingender, 2010). For this reason biofilm is an attractive position, which creates a competitive environment that forces each inhabitant in the biofilm to adapt or coorporate with other species in order to survive, thus promoting biodiversity and niche adaptation (Flemming and Wingender, 2010; Yang et al., 2011). In addition, the position of the microorgnisms within in the biofilm is also important, because optimization of position creates the best possible living conditions. Gradients from surface to the center of the biofilm exists, which include oxygen, nutrients, etc, which enable, for example anaerobic life in the center even though the biofilm is suspended in aerated conditions.

1.5.2 Extracting EPS Proteins

Because the EPS is a very important part of the properties in activated wastewater treatment a lot of effort has been put into the characterization, and identification of single EPS components, and in particular EPS proteins. Several of the methods was established in 1990's, and these methods were modified to be compatible with activated sludge with great success, and provided insights into the bulk quantities of the EPS components and functions (Raunkjær et al., 1994; Frølund et al., 1995, 1996). However, more detailed analyse were limited until the advancement of high-throughput methods. The method used for EPS protein extraction is the focus of this study, because no standardized extraction method currently exists. Several extraction methods have been reported in the literature, and the field of metaproteomics applied to wastewater is slowly growing, with the most important contributes made by (Wilmes and Bond, 2004; Park and Novak, 2007; Abram et al., 2009; Kuhn et al., 2011; Barr et al., 2011). The extraction methods are often extensive, time consuming, and have common features, which are also shared with common proteomic strategies. Initially, the sample obtained is treated in several washing steps to remove contaminants (Wilmes and Bond, 2004). Entire metaproteomes are then extracted by applying a strong denaturing extraction buffer, which contains denaturing compounds, such as urea, thiourea, SDS, CHAPS, EDTA, phenol, which induces cell lysis, and destabilizes the cells. Alternatively, the extraction of EPS proteins is often based on the application of EDTA or cation extraction resin, which extracts ions from the biofilm leading to the liberation of EPS components (Frølund et al., 1996; Park and Novak, 2007; Yu et al., 2007; Barr et al., 2011). The subsequent sample processing often includes the application of brute force methods such as sonication, french press, or flash-freezing by liquid nitrogen (Park and Novak, 2007; Wilmes and Bond, 2006b; Abram et al., 2009). Occasionally, filtration steps are applied in the treatment of samples, though this is not common, due to clocking, and/or sample loss (Park and Novak, 2007; Kuhn, 2010). The result of the preceding steps is often referred to as crude extracts, which must be purified to remove contaminants. This is achieved using precipitation strategies. Several strategies have previously been tested, including trichloroacetic acid (TCA), acetone, ammonium sulfate, and ethanol.

1.6 Current Environmental Issues Resolved by Biomembrane Bioreactors

Recently it was investigated by 'Ingeniøren' that the two main WWTP's of the danish capitol Copenhagen are discharging chemicals, heavy metals, and medical waste products into the Østersø (Marfelt and Wittrup, 2011; Djursing, 2008). However, these plants are not designed to retain or purify these highly toxic compounds. It is obvious that these compounds cannot be allowed to continuously contaminate the immediate surroundings, and therefore the plants must be equipped with new tools to accommodate these new objectives. One such tool is the application of biomembranes, which filters the effluent stream and discharges only clear water. Retrofitting membrane

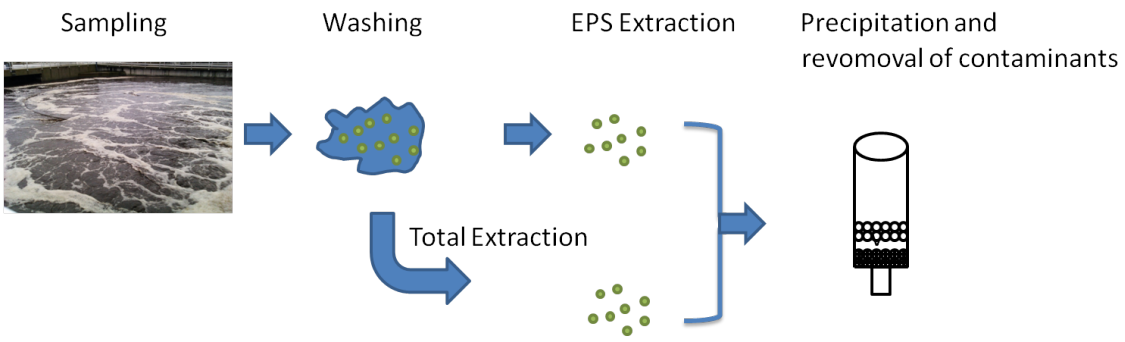


Figure 1.7: A generalized overview of the extraction method commonly applied to wastewater metaproteomic studies. Samples are washed and post-processed to obtain either EPS proteins, or the entire metaproteome. The resulting crude extracts are subsequently further purified using different precipiation methods.

bioreactors (MBR) to existing WWTP's is a promosing solution that not only retains the microcommunity, but also makes it possible create smaller plants, because secondary clarifiers are unnecessary, as shown in Figure 1.8. However, the application of MBR's is impeded by build up of organic and inorganic materials on the membrane. The membrane will eventually start to clog if this phenomenom is allowed to continue, which is referred to as membrane fouling. These issues are sought to be resolved, and in an effort to investigate the processes leading to fouling the identification of EPS proteins may provide valuable information. A fully operational MBR is not available for this study, and as a consequence the sludge flocs from operating WWTP's will be used as a model system. It is hypothesized that the sludge floc will constitute a good model system, because the MBR also functions as a EBPR WWTP, and the membrane is submerged in the anoxic/aerobic tanks from which the sludge flocs are obtained. The observations, and results will then subsequently be transferred to the MBR studies, when it becomes available. The strategy that will be followed to obtain the EPS protein identification will be detailed in the following chapter.

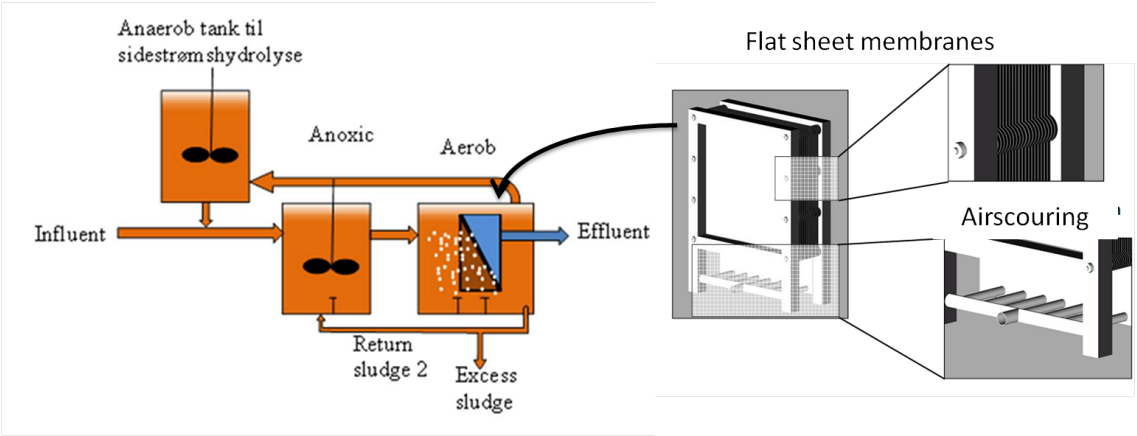


Figure 1.8: To the left: schematic of a wastewater treatment plant with an in-build membrane, which retains the microorganisms. The addition of the membrane into the anoxic/aerobic reactor allows the aeration to be used as a scouring meachanism, which removes particles from the membrane. To the right: an example of a membrane is the flat sheet membrane several membranes filters the water in parallel, thus providing a large surface area.

2 Aim and Strategy

The aim of this research is to implement extracelluar protein, and total protein extraction methods for subsequent analysis with liquid chromatography nano-electrospray tandem mass spectrometry (LC-MS/MS). Several methods have been described in the Introduction, and these were sought to be applied to two Danish full-scale WWTP's, Aalborg West, and Aalborg East (subsequently referred to as WWTP West, and WWTP East). The extraction development, and optimization is an important part of this research, and consequently the literature was well studied prior, and during the experimental work. These methods was carefully scrutinized and compiled into a reference work, which is located in Appendix A, and C. It was sought to adapt these methods to make them compatible with samples acquired from the two full-scale WWTP's. The purification and subsequent analysis on LC-MS/MS was performed in parallel with another research project in the genomic content of these WWTP's. The combination of genomic, and proteomic studies are referred to as proteogenomics, and the workflow is shown in Figure 2.3. Both studies are a part of the EcoDesign project, which has the goal to gain better controlling abilities and optimization of wastewater treatment operation, e.g. enhanced biological phosphorous removal.

The experiments will be focused on the extraction of extracellular proteins, i.e. EPS protein. In addition, the entire protein complement, i.e. the metaproteome of the community, will be sought to be extracted. The proteins within the metaproteome will then be searched for possible EPS proteins. The samples will be obtained from WWTP West, and WWTP East, because both of these plants have been genomically sequenced using shotgun strategies, thus predicted peptide sequences are available for mass spectrometry based database searching. The samples will be treated on ice at all times possible, and kept as close to freezing point as possible to avoid any chemical or enzymatic activity. The obtained samples will then be processed through four different downstream analyses, as shown in Figure 2.1. The protein extraction strategies are applied to break the sludge flocs apart, and to solubilize proteins for further purification using SDS-PAGE. Trypsin will then be used to generate peptide samples, which are preferred in LC-MS/MS, due to their inherent properties.

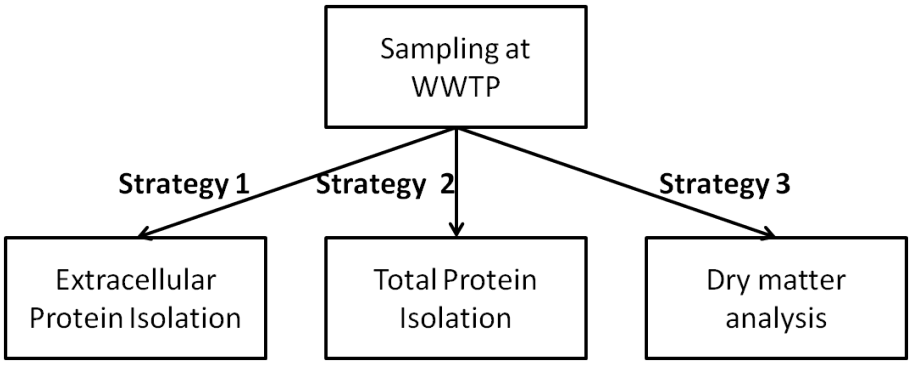


Figure 2.1: Schematic overview the strategies to process the obtained activated sludge samples. The extracellular proteins are isolated to determine the identity of proteins present in the sludge flocs, and the function of these proteins. The total protein complement, i.e the metaproteome is isolated to obtain as many proteins as possibly. These include cytosolic proteins, and likely also EPS proteins. The dry matter analysis is measured to determine the amount of non-volatile materials.

The extracellular proteins have a primary focus in research, because they may provide a detailed insight into the function of the biofilm layer, i.e. the EPS layer, and therefore these proteins are sought to be purified separately. Several strategies will be evaluated and tested for activated sludge, including the ones shown in Figure 2.2. It is important for this experiment that the activated sludge cells are treated as gentle as possible to avoid cell lysis. If the extraction method causes the cells to lyse, the subsequent protein extraction will be contaminated with cytosolic proteins, which might lead to false determination of EPS proteins. On the other hands, the extraction method is required to be sufficient in extracting the EPS proteins, which might require harsh condition for the activated sludge. This will require optimization to obtain the right equilibrium between cell lysis, and extraction yield.

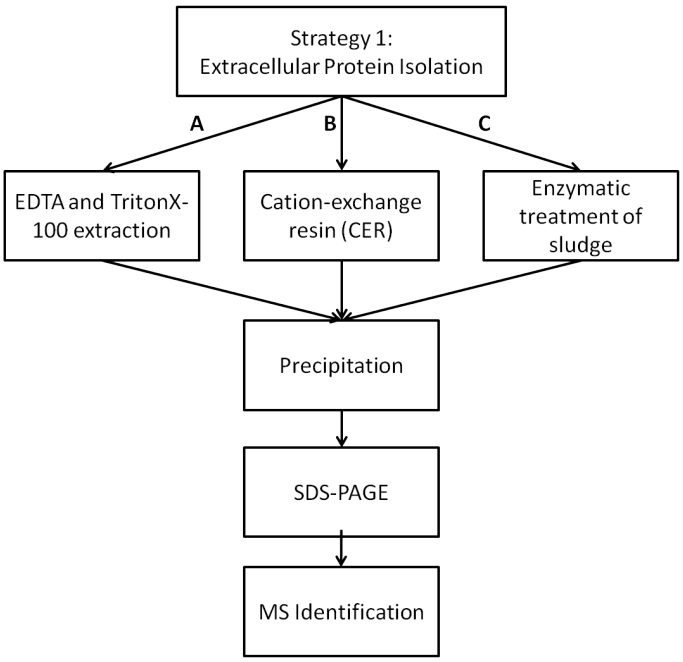


Figure 2.2: An overview of the different crude extraction methods used in Strategy 1. This strategy is divided into Strategy 1A, which is based on the application of EDTA, Strategy 1B, which is based on cation-exchange resin, and Strategy 1C, which is based on enzymatic treatment. Each method produces crude extracts, which will be further purified by precipitation methods, and subsequently proteins will be separated by SDS-PAGE. Mass spectrometry will then be used to identify proteins

One novel strategy will be tested to extract EPS proteins from activated sludge samples. This will be to apply a mixture, called Viscozyme, consisting of polysaccharide degrading enzymes. It is thought that short incubation times with the enzyme mixture will result in the cleavage of the polysaccharides in the samples, which breaks the sludge flocs apart, thus releasing EPS for extraction.

Furthermore, the use of arginine have previously been shown to create competitive binding, which aids the release of EPS proteins from *Pseudonomas aeruginosa* pure cultures (Broekman, 2007). This approach will be tested in all samples in an effort to optimize extraction yield.

Protein samples will be analyzed through LC-MS/MS, and in more detail by the use of a LC-QToF, and the analysis will be performed as shown in Figure 2.3. The resulting mass spectra will be searched against a sample specific peptide database created from activated sludge from the two

WWTP's (Albertsen et al., 2011), and against the SWISSPROT database, using the MASCOT search algorithm (Perkins et al., 1999).

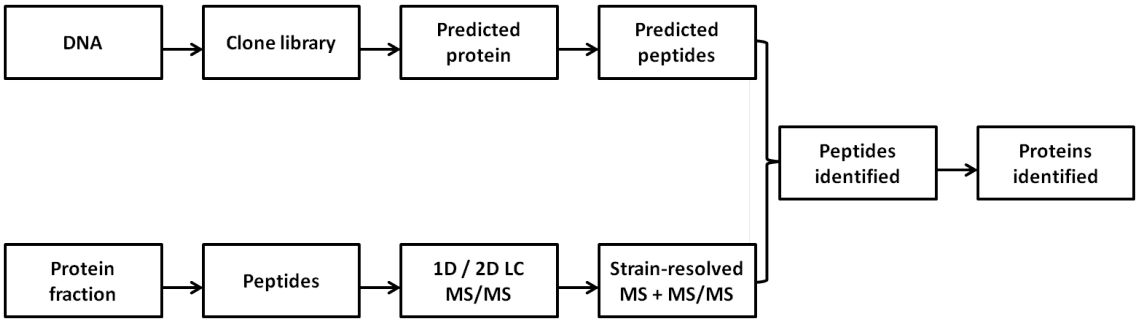
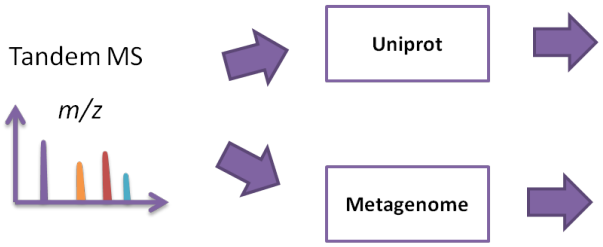


Figure 2.3: Proteogenomics: the combination of genomic, and proteomic analysis. Initially the two strategies are analyzed in parallel, and subsequently the results are used to provide mutual evaluation, and validation. In this study the focus is on the bottom part, i.e. the proteomic strategy.

In a preliminary study conducted prior to this research it was shown that using a sample specific database constructed from the metagenome of WWTP West, and WWTP East, resulted in an increased identification rate, as shown in Figure 2.4. Samples were generated with a CER-based method used originally in (Frølund et al., 1996). The sample were run on a SDS-PAGE gel, in-gel digested, and then analysed with LC-MS/MS. Subsequent, searches with MASCOT, showed that the protein identification rate increased by applying the metagenome database. The protein scores were not very high - in comparison a BSA standard have been observed to produce a protein score of 2565 - though the priliminary results motivates the application of the metagenome database.



Protein identification	Score	Taxonomic binning
Aspartyl-tRNA synthetase	59	Cyanobacteria
Oxidoreductase	55	Gammaproteobacteria

Protein identification	Score	Taxonomic binning
Hypothetical protein	98	Pseudomonas
Outer membrane porin	73	Dechloromonas
Iojap-like protein	62	Bacteroidetes
Transposase	59	Bacteria
ABC transporter	58	Betaproteobacteria
Predicted ORF	56	N/A
NTPase	53	Bacetroidetes
Glycosyltransferase	52	Chloroflexaceae

Figure 2.4: Preliminary proof of concept of the strategy. When LC-MS/MS data was searched against the UniProt database, only 2 proteins were identified. However, when the same data was searched against the metagenome database 7 proteins were identified.

The hypothesis of this study is that proteins are abundant in the activated sludge flocs. It is thought that the EPS layer contains several excreted proteins, that are microorganism specific, and there-

fore that the identification of these proteins may provide further insight into the biological activity of individual microorganisms in the activated sludge, and the community as a whole. Due to the high amounts of large polymers, e.g. polysaccharides, it is thought that the microorganisms excrete specialized enzymes to degrade the polymers into smaller fragments that can be assimilated by the microorganism. Thus, the identification of specialized enzymes might provide new discoveries of industrial interest. This study therefore has the potential to increase the current knowledge of EBPR activated sludge systems, and to provide more insight into this enigma.

3 Materials & Methods

The materials in used in this research is condensed into the consensus: chemical compound; (Manufacture: product number) solution make-up/concentration - next chemical compound etc. All solutions used for experiments were freshly made prior to the experiments. All chemicals was from Sigma unless otherwise stated.

Materials: ethanol; 60% - acryl-bisacrylamide mix; 30% 37.5:1 - Tris - sodium dodecyl sulfate (SDS) - ammonium persulfate - tetramethylethylenediamine (TEMED) - glycin, SDS sample buffer; containing 1.5 M Tris (pH 6.8) 20% SDS, 30% glycerol, 1.8 mg bromophenol blue - methanol; (Merck) 100% - acetic acid; 1% - ElgaWater; (Elga) - sodium thiosulfate; - formalin; - Na2CO3; - ammonium bicarbonate; - acetonitrile; (Merck) mass spec grade, dithiothretol (DTT); - iodoacetic acid; - Trypsin; (Promega) 10 $ng/\mu L$, 400 mL centrifuge tubes, NaCl; 0.9% - 50 mL greiner tubes, ethylenediaminetetraacetic acid (EDTA); - TritonX-100; - bovine serum albumin (BSA); - Ovalbumin; - ultracentrifuge tubes - Kimberly glass tubes - cloroform; - 2D sample buffer; 7 M urea, 2 M thiourea, 2% CHAPS, and 100 mM DTT - acetone - formic acid - Poros R1 beads - 1.2 μm filter, 15 mL greiner tube - tin foil cups - NaCl; - glass homegnizer - Na_3PO_4; - NaH_2PO_4; - NaCl; - KCl; - Dowex-Marathorn CER; - Complete Mini Protease inhibitor; - urea; - thiourea; - CHAPS; - rod sonicator, Viscozyme (Novo Nordic); - Tryple; (Invitrogen) 10X - ammonium bicarbonate; - SpeppacC18; (Oasis HlB 1cc cartridge) - phosphoric acid -

3.1 Methods

All the methods used in this report is listed in the following.

3.1.1 Collection of Activated Sludge

Activated sludge samples were collected from either of two WWTP's Aalborg West WWTP or Aalborg East. Samples were collected from the aerated nitrification/denitrification tanks. All samples were collected into containers that were already on ice, and the samples were kept on ice at all time while handling the samples. The samples were collected, and transported to the laboratory within 30 min or less.

3.1.2 Total Solids Analysis

The total solid analysis was performed in triplicates. Three times tin foil cups were filled with $\approx$100 mL activated sludge. The tin foil cups were then heated at 100^oC over night. The tin foil cups were measured before filling them, after filling them, and after heating, and the dry content was calculated.

3.1.3 In Solution Protein Extraction / Analysis of proteins in liquid phase matter

400 mL of activated sludge was centrifugated at 5000 g for 30 min. The supernatant was then collected, and acidified to 1% formic acid. The sample was then incubated with Poros R1 beads at 4^oC for 1 $hour$. The beads was then recovered by filtering the sample through a 1.2 μm filter into a 15 mL greiner tube. The peptides were then eluted by adding 1% formic acid, and then centrifugated in a benchtop centrifuge. THe supernatant was then transferred into new tubes, and analysed by LC-MS/MS.

3.1.4 Extraction Protocol Adapted from (Yu et al., 2007)

400 mL activated sludge was centrifugated at 5000 g for 30 min at 4^oC, and subsequently the supernatant was discarded. 200 mL 0.9% NaCl was added, and the sample was centrifugated at 5000 g for 30 min at 4^oC. The supernatant was discarded, then 200 mL 50 mM Tris-HCl pH 7.0 was added, and the sample was centrifugated at 5000 g for 30 min at 4^oC. The recovered pellet was then dissolved into 400 mL 50 mM Tris-HCl pH 7.0, and aliquoted into 50 mL greiner tubes. The preceeding steps is referred to as the washing steps. Before addition of reagent each sample was homogenized using a glass homogenizer for 1 min to disaggregate the microflocs, and the homogenized sample was transferred from the glass homogenizer into new 50$mLgreinertubes$. Reagents were then added to the 50 mL greiner tubes, and four different extractions buffers were used. The first extraction buffer contained 50 mM EDTA, the second contained 0.1% TritonX-100, the third contained 50 mM EDTA, 0.1% TritonX-100, and 50 μg BSA, and the fourth contained 50 mM EDTA, 0.1% TritionX-100, and 6 M guanidinium chloride. The three different samples were then centrifugated at 5000 g for 15 min at 4^oC, and the supernatant was collected into 70 mL ultracentrifuge tubes. The samples were then centrifugated at 30,000 g for 30 min, and the supernatant were collected. The resulting sample then constituted the "crude" extract.

In addition to the above method some alterations were made in subsequent purifications. After incubation with the extraction buffer containing 50 mM EDTA, 0.1% TritonX-100, and 50 μg BSA the sample was centrifugated at 12,5000 g for 30 min at 4^oC. The supernatant was collected into new 50 mL greiner tubes, and 6 M guanidinium chloride was added. The mixture was rotated, and incubated on ice for 30 min, while mixed every 15 min, i.e. twice. The samples were then centrifugated at 12,5000 g for 30 min at 4^oC, and the resulting supernatant was collected. The collected sample then constituted at two-step "crude extract".

Furthermore, all the above samples were repeated, and to each sample an additional 3 mg/mL arginine was added to the extraction buffer.

3.1.5 EPS Extraction using Cation Exchange Resin

The cation exchange method for extraction of extracellular proteins is based on (Frølund et al., 1996). However, in this study the modified version from (Barr et al., 2011) was used. The activated sludge was washed as described previously, and before addition of reagent each sample was homogenized using a glass homogenizer for 1 min to disaggregate the microflocs, and the homogenized sample was transferred from the glass homogenizer into new 50$mLgreinertubes$. The samples were centrifugated at 5000 g for 5 min at 4^oC, and subsequently 30 mL extraction buffer, consisting of 2 mM Na_3PO_4, 4 mM NaH_2PO_4, 9 mM NaCl, 1 mM KCl, pH. 7, 4^oC, were added. In addition, 21 g Dowex-Marathorn CER, 231 mg DTT, and three tables of Complete Mini Protease inhibitor was added. The mixture was rotated on ice for 5 $hours$, with mixing every 15 min. The samples were subsequently centrifugated at 15,000 g for 15 min at 4^oC. The supernatant was collected, and constituted "crude" extracts.

3.1.6 Activated Sludge Binding Capacity

The activated sludge was washed as described previously, and before addition of reagent each sample was homogenized using a glass homogenizer for 1 min to disaggregate the microflocs, and the homogenized sample was transferred from the glass homogenizer into new 50$mLgreinertubes$. 50 mM EDTA, 0.1% TritonX-100, were added to each greiner tube. Furthermore, BSA was added to the greiner tubes in varying concentrations: 50 μg, 100 μg, 250 μg, 500 μg, 1 mg, 5 mg, and 10 mg,

respectively. The different samples were then centrifugated at 5000 g for 15 min at 4^oC, and the supernatant was collected into 70 mL ultracentrifuge tubes. The samples were then centrifugated at 30,000 g for 30 min, and the supernatant were collected. The resulting samples then constituted the ”crude” extracts.

3.1.7 Broad Spectrum Enzymatic Treatment of Sludge

The activated sludge was washed as described previously. The 50 mL samples were centrifugated at 5000 g for 5 min at 4^oC, the supernatant was discarded, and the pellet was reconstituted in 50 mL 100 mM ammonium bicarbonate. The sample was then homogenized in a glass homogenizer for 1 min on ice to disaggregate the microflocs, and the homogenized sample was transferred from the glass homogenizer into new 50 $mLgreinertubes$. 50 μL Viscozyme wad added to the sample and incubated at 37^oC for 1 $hour$. 50 mM EDTA, and 0.1% TritonX-100 was added to the sample, and incubated on ice for 30 min. The sample were then centrifugated at 12.500 g for 30 min at 4^oC. The supernatant was recovered, and constituted ”crude” extracts.

3.1.8 Total Protein Extraction - Wilmes et. al 2004

The method was adapated from (Wilmes and Bond, 2004). The activated sludge was washed as described previously, and before addition of reagent each sample was homogenized using a glass homogenizer for 1 min to disaggregate the microflocs, and the homogenized sample was transferred from the glass homogenizer into new 50 $mLgreinertubes$. 10 mL urea-thiourea-CHAPS (UTCHAPS) extraction buffer, consisting of 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 10 mM Tris-HCl, 10 mM EDTA, 50 mM DTT, 1 tablet of complete protease inhibitor. In addition, 50 μg BSA was added. The sample was then incubated at 4^oC for 2 $hours$ with hand-mixing every 15 min. Cells were then lysed by a rod sonicator by three times 30 sec using a 3 mm probe with an amplitude of 180 μm. The amplitude of 180 μm was the 100% setting, and separate samples were sonicated using an amplitude of 144 μm, 108 μm, 72 μm, respectively. The four samples were then centrifugated at 12,500 g for 30 min at 4^oC. The supernatant was discarded, and the remaining pellet was reconstituted in a 2D sample buffer consisting of 7 M urea, 2 M thiourea, 2% CHAPS, and 100 mM DTT.

3.1.9 Chloroform and Methanol Precipitation for Protein Purification

Crude extracts was further purified using the chloroform and methanol precipitation protocol published by (Wessel and Flügge, 1984). 1 $volume$ of sample was transferred to a glass tube, then 3 $volumes$ of methanol, and the mixture was mixed thoroughly. Subsequently, 1 $volume$ of chloroform was added, and the mixture was mixed thoroughly. Finally 4 $volumes$ of ElgaWater was added, and mixed thoroughly before the samples was centrifugated at 12,500 g for 15 min at 4^oC. Afterwards, the upper phase was discarded, and an additional 4 $volumes$ of methanol was added, and the mixture was thoroughly mixed. The mixture was then centrifugated at 12,500 g for 15 min at 4^oC. The supernatant was discarded, and the remaining pellet was reconstituted in a 2D sample buffer consisting of 7 M urea, 2 M thiourea, 2% CHAPS, and 100 mM DTT.

3.1.10 Acetone Precipitation for Protein Purification

Crude extracts was further purified using an in-house acetone precipitation protocol. 1 $volume$ of sample was mixed with 8 $volumes$ of acetone, mixed thoroughly, and then stored over night at -20^oC. The following day the liquid phase was discarded, and the resulting pellet was reconstituted

in a 2D sample buffer consisting of 7 M urea, 2 M thiourea, 2% CHAPS, and 100 mM DTT.

3.1.11 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) was performed using in-lab created 7.5%, and 15%, and pre-made 4-20% SDS-PAGE gels.

The 7.5% resolving gels were cast using ElgaWater (48% v/v), 30% acryl-bisacrylamide mix (25% v/v), 1.5 M Tris (pH 8.8)(25% v/v), 10% SDS (1% v/v), 10% ammonium persulfate (1% v/v), and 2 μL TEMED. Similarly the resolving gels were cast using ElgaWater water 68% v/v), 17% acryl-bisacrylamide mix (50% v/v), 1.5 M Tris (pH 8.8)(13% v/v), 10% sodium dodecyl sulfate (SDS) (1% v/v), 10% ammonium persulfate (1% v/v), and 2 μL TEMED. The 15% resolving gels were cast using milliQ water (22% v/v), 30% acryl-bisacrylamide mix (50% v/v), 1.5 M Tris (pH 8.8)(26% v/v), 10% SDS (1% v/v), 10% ammonium persulfate (1% v/v), and 2 μL TEMED in the necessary. Similarly the resolving gels were cast using ElgaWater water 68% v/v), 17% acryl-bisacrylamide mix (50% v/v), 1.5 M Tris (pH 8.8)(13% v/v), 10% sodium dodecyl sulfate (SDS) (1% v/v), 10% ammonium persulfate (1% v/v), and 2 μL TEMED in the necessary. The gel were then stored in 1X running buffer at least one day to make sure that the polymerization was finished.

Both the in-lab created gels, and the pre-made gels were run in 1·X-running buffer constituting of 10 g SDS, 30.3 g Tris, and 144.1 g glycin in 10 L. Protein samples were diluted 1:1 with 2X SDS sample buffer consisting of 1.5 M Tris, 20% (w/v) SDS, 30% (v/v) glycerol, and 1.8 mg bromophenol blue. 20 μL of mixed protein, and sample buffer were loaded into each lane, and the gel were run at 90 V for 20 min, and then at 120 V until finished. All gels were run in a cold room at $\approx 4^oC$ to compensate for gel-heating during the electrophoresis.

3.1.12 Silver Staining

The method used was based on the initial research (Krystal, 1987), and the adapted method for SDS-PAGE (Shevchenko et al., 1996) SDS-PAGE gels were fixed in destaining solution conataining methanol, acetic acid, and ElgaWater in amounts (45:10:45), for 20-30 min. The gels were then rinsed in ElgaWater for at least 60 min or overnight. Subsequently, the gels were incubated for 10 min in sensitizing solution consisting of 0.02% sodium thiosulfate. Following, the gels were rinsed twice in ElgaWater. 4^oC 0.1% (w/v) silver solution were added to the gels, and incubated at room temperature for 30 min. Following, the gels were rinsed twice in ElgaWater. The gels were developed in 0.04% formalin, and 2% Na_2CO_3 until the gels were sufficiently silverstained, and subsequently the reaction were stopped using 1% acetic acid. The gels were then stored for maximum one month.

3.1.13 In Gel Tryptic Digestion

The method used was the in-house method used at the Mass Spectrometry Facility of University of Western Sydney. The volume of each reagent used is dependent on the amount of gel pieces, and has therefore been omitted. The reagents should at least cover the gel pieces, and must be equal volumes so that the concentration of reagent in the following is appropriate. Before excision the SDS-PAGE gels were rinsed with ElgaWater for a few hours. Single bands from SDS-PAGE gels were excised and then transferred into 1.5 mL eppendorf tubes. The gel pieces were then cutted into 1×1 mm sized pieces inside the eppendorf tube to reduce the amount of airborne contaminants. The gel pieces were washed twice with 50 mM ammonium bicarbonate, then destained with 25 mM ammonium bicarbonate 1 : 1 with aceteonitrile until clear, and dried under vacuum in a vacuum

centrifuge. The gel pieces were resuspended in 10 mM DTT, and incubated at 37^oC for 1 $hour$. 25 mM iodoacetic acid was added to the samples, and the samples were incubated at 37^oC for 1 $hour$. Acetonitrile was added, then all liquid was discarded. Following alkylation the gel pieces were washed three times with ElgaWater, then washed with 10 mM ammonium bicarbonate, and then washed once with acetonitrile. Following the last washing step acetonitrile was added, and the gel pieces were dried under vacuum in a vacuum centrifuge. 20 ng trypsin in 50 mM ammonium bicarbonate was added to gel pieces on ice. After 5 min each tube was observed to see if the gel pieces were completely covered. If not the pieces were covered with 50 mM ammonium bicarbonate, kept on ice for 5 min, and observed again. The samples were then digested for at least 12 $hours$. The peptides was extracted by transferring the supernatant into new 1.5 mL eppendorf tubes. Further extraction was performed by adding 1% formic acid at room temperature for 15 min. Then equal volume acetonitrile was added, and all liquid were transferred to the new eppendorf tube. The new tube were dried under vacuum in a vacuum centrifuge, and reconstituted in 0.1% formic acid. The samples were then analyzed by liquid chromatography electrospray tandem mass spectrometry.

3.1.14 LC-ESI MS/MS Analysis

Automated LC-ESI MS/MS was performed using a hybrid micro Q-TOF mass spectrometer in data-dependent mode with a Agilent1200 nanoflowHPLC system. Reversed phase columns (pre-column 2 cm, 75 μm id; separation column 12 cm, 50 μm internal diameter) were packed in-house with ReproSil-Pur C18-AQ 3 μm resin (Dr. Maisch GmbH, Ammer-Buch-Entringen, Germany) using a high-pressure vessel.

Aliquots of the tryptic peptides were injected onto the pre-column with a flow rate of 5 $\mu L/min$ and subsequently eluted at 10 nL/min using a 20 min gradient. The gradient was achieved by using a Solvent A consisting of 0.1% formic acid, 0.005% heptafluorobutyric acid, and a Solvent B consisting of 90% acetonitrile, 0.1% formic acid, and 0.005% heptafluorobutyric acid. The gradient was initialized with Solvent A, and Solvent B in the ratio 90:10, which lineary progressed in the 20 min to the ratio 55:45. The micro Q-TOF mass spectrometer was operated in data dependent mode to automatically switch between MS and MS/MS.

3.1.15 Protein Identification by Automatic Peptide Assignment using MASCOT

The resulting spectra from the DDA mode LC-MS/MS were searched toward a SWISS-PROT database using MASCOT search algorithm. MASCOT were set for MS/MS ion search with the following settings. Dataformat: mascotgeneric, peptide tolerance: $+/-0.10\,Da$, MS/MS tolerance: $+/-0.1\,Da$, Enzyme: semi-trypsin, allowed missed cleavage: 1-3, charge states: +2, +3, +4, variable modifications: oxidation methionine (mM), cysteine methylcarboxylation (cC) (for reduced samples), decoy: on, instrument type: ESI-QUAD-TOF.

4 Results and Discussion

The methods used in this large-scale protein purification research, described in the Materials and Methods Section 3, were tested and developed as the experimental work proceeded. The extraction methods were divided into two major strategies; the extracellular proteins extraction (Strategy 1A-C), and total protein purification (Strategy 2), and were adapted from various published sources, including (Wilmes and Bond, 2004) (Wilmes and Bond, 2006b) (Yu et al., 2007) (Park et al., 2008). In addition, a dry matter analysis was performed to measure non-volatile solids in Strategy 3. All experiments was performed, and repated at least once, except the cation-excange experiment.

4.1 Background of the two Investigated WWTP's

Samples were obtained from the two full-scale plants, WWTP West, and WWTP East. These two plants have been operated for EBPR, and the community of each plant have been monitored for more than a decade, using 35 different FISH probes (Nielsen et al., 2010). Prior to this research a metagenome was constructed for both plants, and the metagenome have been utilized to create a combined database for spectral search using the MASCOT search engine (Albertsen et al., 2011). Both WWTP's are of the general design of the one shown in Figure 1.1, and in more detail WWTP West features a sidestream configuration, while WWTP East is a conventional EBPR WWTP.

4.2 SDS-PAGE Analysis of the Extracted EPS Proteins - Strategy 1

The extraction methods was applied to WWTP West, and WWTP East, in an effort to evaluate the extraction efficiency, by comparing the two WWTP's by using SDS-PAGE. This has been done for full-scale plants, and laboratory scale SBR reactors previously (Park and Novak, 2007), and the experiments was grouped into Strategy 1, as described in the Aim & Strategy section. These strategies were therefore applied because the methods are rapid, and causes less loss of sample compared to 2D-PAGE. It seems that the trend is to use SDS-PAGE gels as a proof of concept, and when this methods reaches the limits of resolution, 2D-PAGE is applied (Kuhn, 2010).

The initial samples from both WWTP's were analyzed to determine if proteins were present in the bulk liquid. These extractions showed no trace of proteins, and subsequent gels were empty. Though this was expected it shows that the proteins are localized in the activated sludge (Frølund et al., 1996).

4.2.1 EPS Protein Extraction

The first extractions were a part of Strategy 1A (cf. Figure 2.2), and the detailed methods used are located in Section 3.1.4. The EDTA based method was selected because previous studies have reported a high extraction yield, and relative high purity, see Appendix C Table C.3. The method was divided into steps that included sludge washing, homogenization, EPS protein extraction, EPS protein separation using SDS-PAGE, LC-MS/MS analysis, and MASCOT database searching. Throughout these steps the samples were handled on ice to minimize any activity, that could cause loss of sample, because the enzymatic activity of activated sludge, and extracted EPS have

been observed to be high (Frølund et al., 1995; Yu et al., 2007). The initial steps were performed to introduce a stabilizing buffer to the live sample, in addition to removing excess polysaccharides (Wilmes and Bond, 2004), and the homogenization was included to break the sludge flocs into similar sized particles, and consequently to increase the surface area. The washing and homogenization of the sludge was performed for all samples, after which the samples branched into separate methods. The washing steps produced a clear liquid supernatant which was discarded, and the remaining sample consisted of big yet compact pellets. The pellets were not soluble, but they were made more soluble and dispersed by the homogenization. Al extractions resulted in a brown colored extract, which were run on a SDS gels for further separation of the extracted proteins. Purified protein samples are often transparent, but the brown color was expected, as it has been reported previously (Kuhn et al., 2011), and could originate from humic compounds (Frølund et al., 1996).

Different washing buffers were used in the initial phase including 100 mM ammonium acetate, 50 mM ammonium bicarbonate 0.9% (w/w) NaCl, 1X phosphate buffered saline (PBS), and 50 mM Tris-HCl. The use of any of these buffer excluding the 1X PBS buffer provided the same results. However, the 1X PBS buffer seemed to introduce and unnecessary high amounts of salts, that complicated the precipitation methods.

4.2.2 Testing Selected Extraction Buffers

The initial experiments was performed to evaluate the performance of different extraction buffers, where the use of the chelating agent ethylenediaminetetraacetic acid (EDTA), and the detergent TritonX-100, was tested separately, and furthermore the two reagents were tested when mixed. The different extraction buffers were utilized and the resulting sample were run on a premade 4-20% SDS-PAGE gel, shown in Figure 4.1, which was subsequently silverstained according to Section A.1. From these preliminary experiments it was clear that the combined use EDTA, and TritonX-100 provided the highest amount of EPS proteins, due to the significantly more intense band. Floc strength have previously been shown to correlate with di- and poly-valent ions (Higgins and Novak, 1997). Supposedly, the addition of EDTA in the extraction buffers preferable chelates monovalent ions, such as calcium ions, Ca^{2+}, and magnesium ions, Mg^{2+}, which extraction the EPS fraction associated with these ions (Park and Novak, 2007). Consequently, this may provide the answer why EDTA extraction provides the highest amount of protein, and why it exceeds extractions methods using strategies, such as centrifugation, ultrasonication, cation exchange resin, and formaldehyde (Yu et al., 2007). Though the EPS is associated with numerous ionic species (Raunkjær et al., 1994), it is very likely that proteins with hydrophobic residues are present in the EPS, and that the addition of detergent is able to make these protein more soluble. The addition of TritonX-100 is also widely used in 2D-PAGE buffers in addition to other denaturants (Wilmes and Bond, 2004). The results obtained therefore indicates that the combined use of EDTA, and TritonX-100 constitutes an efficient extraction buffer.

4.2.3 Testing Selected Precipitation Methods

Different precipitation methods, were tested, and evaluated in addition to the evaluation of the extraction buffers used. The two precipitations methods used were acetone precipitation, and chloroform/methanol precipitation. Both precipitation methods provided samples that were run on the 4-20% SDS-PAGE, which were shown previously in Figure 4.1. The chloroform/methanol precipitation could be completed on the same day, but the acetone precipitation was performed at -20^{o}C overnight. It seemed that the acetone precipitation was not very robust due to high amounts

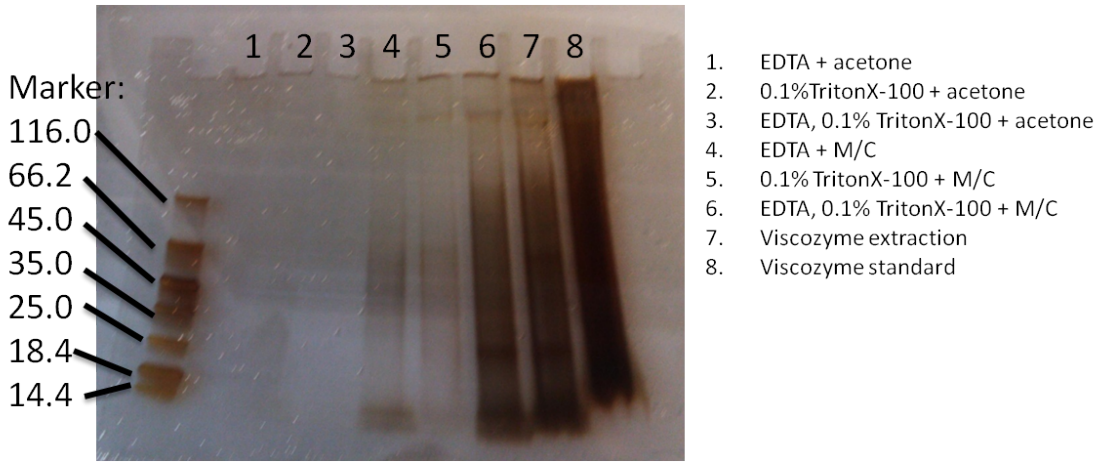


Figure 4.1: The obtained 4-20% SDS gel from the evaluation of extraction buffer, and precipitation method. The samples are indicated to the right, where the extraction method is given plus the precipitation method used. It is seen that the acetone precipitation method did not result in significant silverstaining, while the precipitation with chloroform and methanol (M/C) did result in silverstained bands. In addition, the extraction method using Viscozyme did also produce silverstained bands, and so did the loaded Viscozyme standard - the discussion of the latter samples is covered in subequent section.

of salts, and it was difficult to extract the resulting protein pellet. In previous studies, crude extracts were acid precipitated using trichloroacetic acid (TCA) or precipitated using a ammonium acetate in methanol, and then washed several times with ice cold acetone, with several subsequent washing steps (Denecke, 2006; Barr et al., 2011; Kuhn et al., 2011). However, in the context of the aforementioned studies, the use of acetone was to ensure that residual TCA was removed. After the poor result using the acetone precipitation method it might therefore produce better results by utilizing a preceding TCA precipitation to decrease the amount of salts, which is to be performed in follow up studies.

The chloroform/methanol precipitated pellet was easily extracted, and the resulting pellet was dissolved in 2D-buffer. The chloroform/methanol method is a robust solution to remove detergents, and lipids (Wessel and Flügge, 1984), which are present in the activated sludge. Furthermore, the additionally added TritonX-100 needs to be removed, because TritonX-100 suppresses ionization in MALDI-TOF, reduces the resolution in liquid chromatography (Zhang and Li, 2004) (Yeung et al., 2008). It is seen from the gel (Figure 4.1) that each of the three acetone precipitation did not provide any distinct bands. However, the chloroform/methanol precipitation resulted in significantly more silverstained lanes. The separate use of either EDTA or TritonX-100 did provide a low protein yield, which was observed as a light smear. However, when added simultaneously the EDTA and TritonX-100 provided a significantly more dense smeared silverstained lane for both precipitation methods. Curiously, the added BSA protein sample was not observed at the expected molecular weight, and could not be observed on the gel at all.

The silverstained proteins were sought to be identified by in-gel digestion, as described in Section 3.1.13, and the excision of samples is shown in Figure 4.2

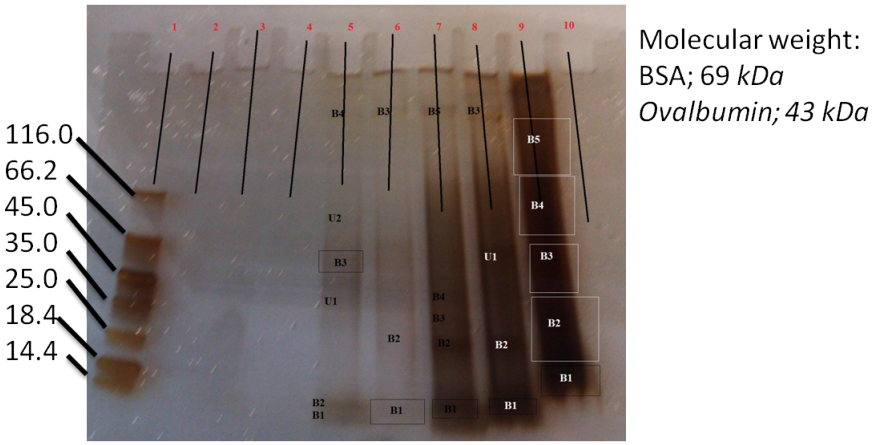
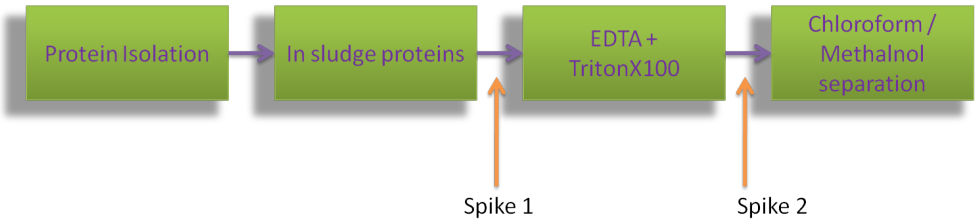


Figure 4.2: The same gel as shown in Figure 4.1, which highlights the excised samples. Samples were named with B for Band, where it seemed that a band was present, and numbers werea added from bottom to top. If no band was visible the excised locations were referred to as U for unknown.

4.2.4 Troubleshooting the Selected Extraction Buffer and Precipitation Method

The loss of the BSA standard necessitated an evaluation of the precipitation protocol. The extraction buffer was selected to the one containing EDTA, and TritonX-100, because it provided the best results, as described above. In addition, the precipitation protocol selected was the chloroform/methanol method for the same reason. It was then investigated if BSA was lost in the precipitation step or if was lost in a prior step using the spiking strategy shown in Figure 4.2.4. The spiking in of protein was performed in three different experiments: the first experiment was



performed using BSA as spike 1, and Ovalbumin as spike 2, and the second experiments was performed by reversing the proteins. The third experiments was performed by spiking BSA as spike 1, and 2, and the resulting SDS-PAGE gel from this experiment is shown in Figure 4.3. All three experiments confirmed that the chloroform/methanol precipitation method sufficiently extracted the spiked in protein. It was observed that if the $50\,\mu g$ BSA or $50\,\mu g$ Ovalbumin was spiked into a $50\,mL$ sludge sample then these protein did not appear on the resulting gel. However, if the BSA or Ovalbumin was spiked in just before the precipitation it was clear visible on the gel. Furthermore, if the BSA was spiked into the sludge, and just before the precipitation it was even more visible on the gel. The results therefore indicated that it was necessary to spike in more than $50\,\mu g$ per $50\,mL$ to obtain positive results. Originally the $50\,\mu g$ was chosen, because it was known through experience that this amount of protein would provide a clear band on the SDS-PAGE gels for other samples. This observation motivated the further investigation of what happened to the spiked in BSA, which investigated in a binding assay discussed in the following.

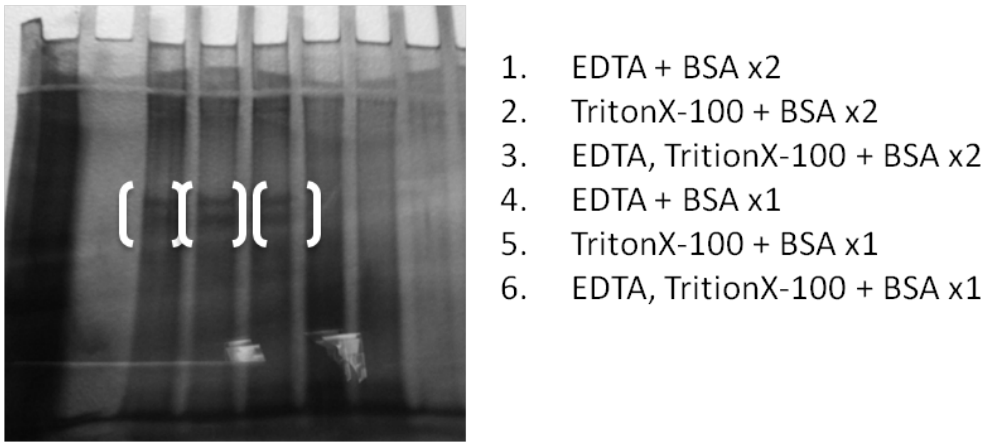


Figure 4.3: The obtained 4%/7.5% SDS PAGE Gel. Samples were spiked with BSA, as shown in Figure 4.2.4, and samples where BSA was spiked in twice resulted in silverstained bands, indicated in brackets. However, the samples, where BSA was spiked in as 'Spike 1', did not produce silverstained bands.

4.3 Activated Sludge Binding Assay

The loss of the spiked in BSA in the initial extractions of EPS proteins motivated the investigation of how the BSA was lost. Two possible scenarios was thought up. Either the BSA was lost during the purification, or the BSA was bound to the activated sludge. The latter was investigated by creating a saturation experiments designed to determine how much BSA was to be added so that the it was visible on the SDS-PAGE gels. The experiments was performed on obtained samples from both WWTP West, and WWTP East. These were performed using the EDTA, and TritonX-100 extraction buffer, where 7 samples were created by spiking in increasing amounts of BSA.

The samples from WWTP West indicated that the activated sludge did in fact have a binding capacity for BSA, which is shown in Figure 4.4. The BSA band was somewhat visible in the sample containing $100\,\mu g$ BSA, and was clearly visible in the sample containing $250\,\mu$. The total solid content, or suspended solids (SS) of the sample was measured to be $4.80\,mg/mL$, which means that according to the volume loaded on the gel $240\,mg$ solids was loaded. This lead to the estimation that the binding capacity of the activated sludge from WWTP West is between:

$$0.416\,mgBSA/gSS < Binding\ capacity < 1.042\,mgBSA/gSS \tag{4.1}$$

where *SS* is the amount of solids.

Similarly the samples from WWTP East, shown in Figure 4.5, also showed that the sample containing $50\,\mu g$ BSA was not visible on the SDS-PAGE gel, whereas the sample containing $250\,\mu g$ was clearly visible. The SS content of the sample was measured to be $4.84\,mg/mL$, which means that according to the volume loaded on the gel $242\,mg$ solids was loaded. This lead to the estimation that the binding capacity of the activated sludge from WWTP West is between:

$$0.207\,mgBSA/gSS < Binding\ capacity < 1.033\,mgBSA/gSS \tag{4.2}$$

where *S* is the amount of solids.

At first the observation that the activated sludge at both WWTP's was able to bind the spiked in protein was surprising. Assuming that the protein content measured by Frølund et al. (1996) can be used as a rough estimate of the current protein concentration then the the $50\,\mu g$ BSA only

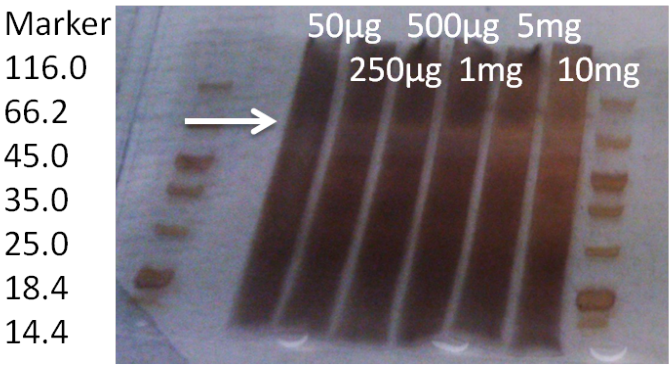


Figure 4.4: The results from the activated sludge binding assay from WWTP West. The amount of 'Spike 1' (confer Figure 4.2.4) spiked in BSA is given on the figure. The silverstained band of BSA, indciated with the arrow, was slightly visible at 100 μg BSA, and was clearly visible at 250 μg BSA. It is further seen that increasing the amount of BSA results in a more dense band.

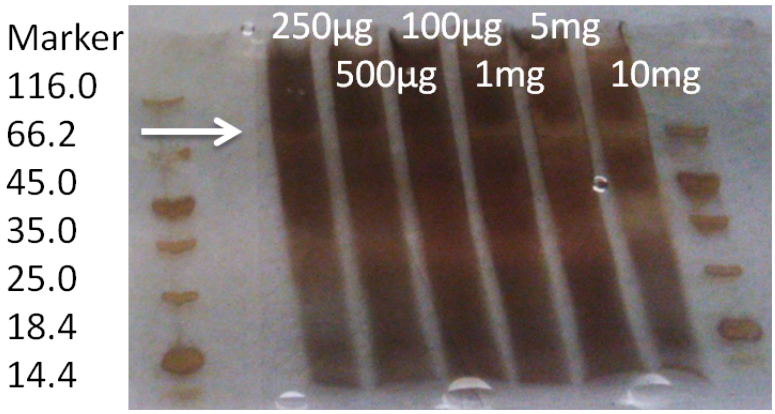


Figure 4.5: The results from the activated sludge binding assay from WWTP East. The amount of 'Spike 1' (confer Figure 4.2.4) spiked in BSA is given on the figure. The silverstained band of BSA, indciated with the arrow, was not visible at 50 μg BSA, and was clearly visible at 250 μg BSA. Again, it is seen that increasing the amount of BSA results in a more dense band.

constitutes around 0.44‰of the total protein content in a sample. It therefore was surprising that even though the spiked in BSA was added at this relative low amount it was extractable from the samples. Though the experiment did provide the threshold for required amount of spiked in protein, it did not provide any indications to where the BSA was lost. Due to lack of time this was not completely elucidated. However, a strategy was formulated to assess the loss of spiked in sample, as shown in Figure 4.6. The proposed mass balance would be able to determine not only the loss of spiked in sample, but would also enable to determine any sample loss. This would be performed by taking out a sample from each step in the extraction protocol (while subtracting the amount of sample each time), and measure the protein concentration using the modified Lowry (Frølund et al., 1996). This experiment was partially performed, where the samples from each step was obtained, and stored at -20^oC. Unfortunately, the setup of the modified Lowry, was too time consuming within the time frame of this master project, and would require numerous samples, standards, etc. Current effort is being put into downscaling this experiment into 96 well plates to

make this method more feasible.

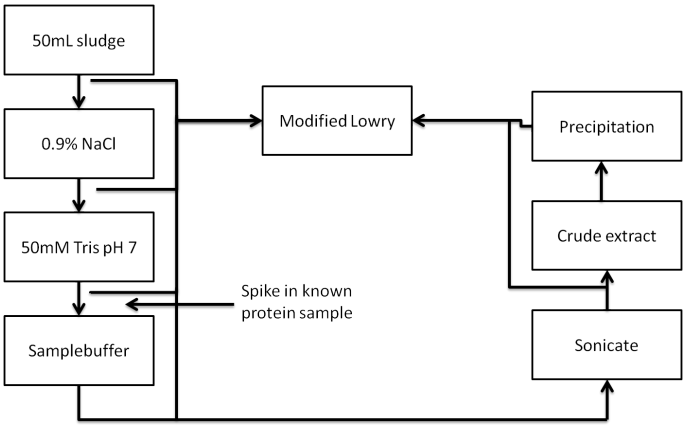


Figure 4.6: Proposed mass balance assessment to ascertain the location of loss of spiked in sample. A sample could be taken from each step, and the protein quantity could be examined using modified Lowry (Frølund et al., 1996).

Summing up on the spike in experiments it was shown that BSA was lost, when added in less than $100\,\mu g$. The actual mechanism for the loss is uncertain, though it is very likely that the BSA sample is adsorbed to the sludge flocs. It is already known that sludge flocs possess numerous polar sites, including the ones created by polysaccharides, proteins, and eDNA (Seviour et al., 2010; Flemming and Wingender, 2010). It is therefore not surprising that BSA is adsorbed to the surface of the biofilm, where it is either stored or assimilated by the microorganisms. An interesting experiment could be to spike in a radioactive labelled protein to locate the loss of sample in a traditional pulse-chase experiment. This experiments could be applied in a similar way as the strategy proposed in Figure 4.6 to provide a rapid overview of the location of the spiked in sample.

4.3.1 Addition of Guanidinium Chloride and Sonication in the EPS Protein Extraction

The initial extraction resulted in very few identified proteins. Investigation of the MS, and MS/MS chromatograms[1] indicated that only few peptides were fragmented, which could be the reason for the low identification rate. It was therefore sought to further optimize the extraction protocol. The addition of excess[2] guanidinium chloride (GCl) was included to solubilize hydrophobic proteins, and is a widely used method for this purpose (Denef et al., 2008). The use of GCl was tested in two separate experiments. The GCl was added to the selected extraction buffer, thus including EDTA, TritonX-100, and GCl. In addition, a two-step extraction method was created, where the GCl was added after the preceding EDTA, and TritonX-100 extraction, was performed. The samples were then split into two sets of samples, and one set was exposed to sonication, though the sonication was only tested on samples from WWTP West. Sonication is a widely used method used to solubilize proteins, and pellets of proteins, and it was applied in an attempt to separate lectins, i.e. polysaccharide binding proteins, and very hydrophobic proteins (Otzen and Nielsen, 2008; Denef et al., 2008). The resulting samples were then further purified using chloroform/methanol

[1] Mass spectrometry data analysis is covered separately in Section 4.5

[2] $6\,M$ GCl is close to saturation.

precipitation, and run on SDS-PAGE gel, which is shown in Figure 4.7 for WWTP West. Though the picture is of poor quality, the resulting gel was still of good quality.

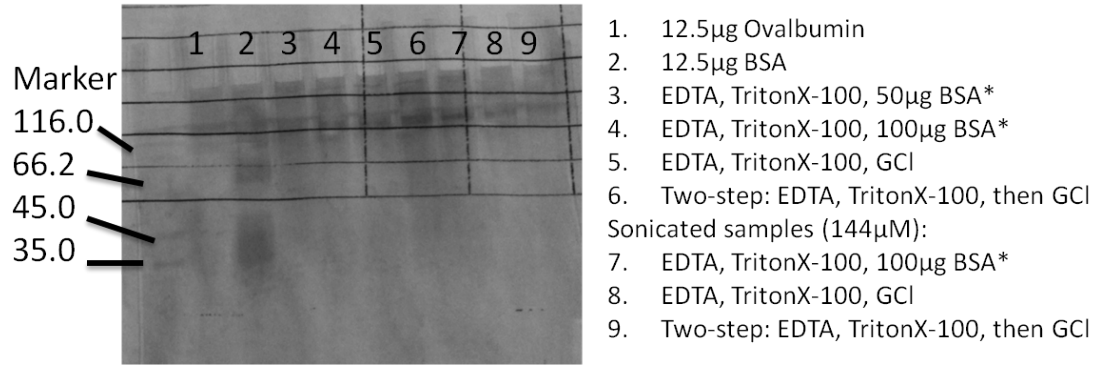


Figure 4.7: The obtained 4%/7.5% SDS-PAGE gel from GCl experiments from WWTP West. The picture is of poor quality, though it was very similar to the experiment performed on WWTP East samples. In general it seemed that the use of GCl did not have a significant effect.

The same experiment was subsequently performed on WWTP East, shown in Figure 4.8, and because the sonication had no apparent effect neither of the samples were exposed to sonication. Both the gels from WWTP West, and WWTP East, indicated that the addition of GCl did not have any apparent effect, though it might be generous to slightly attribute the GCl with more dense silverstaining. As a side experiment the two standards used for BSA, and Ovalbumin were run on the two gels, and it seemed that these standards was of low quality, due to the several degradation bands. Though the standards was old, this have been observed in previous work using the same manufacture.

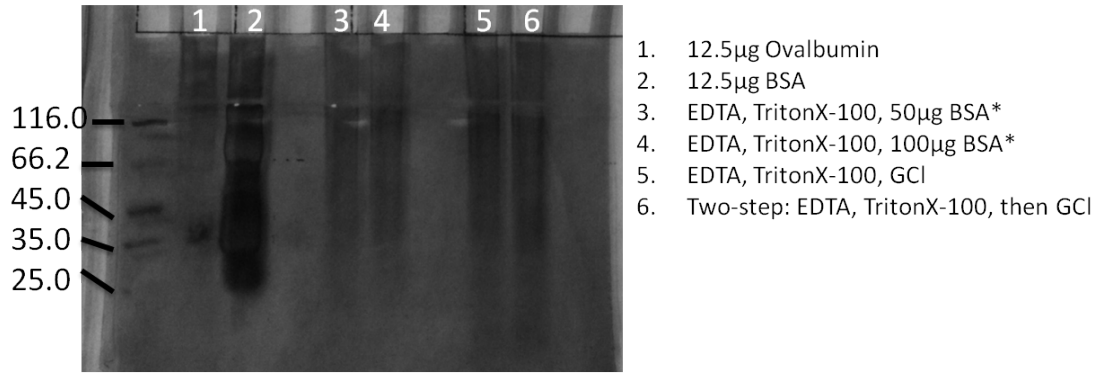


Figure 4.8: The obtained 4%/7.5% SDS-PAGE gel from GCl experiments from WWTP East. In general it seemed that the use of GCl did not have a significant effect, though it seemed that lane 5, and 6 was somewhat more stained.

The GCl experiment was the last experiment based on the chelation of EDTA, and thus Strategy 1A. Several methods, and approaches were applied and tested using this chelating agent. None of the methods seemed to be optimal for subsequent LC-MS/MS, and it seemed to be due to contaminants. It have been proposed previously that the high amounts of staining and smearing on the gels is caused by contaminants, and co-precipitating compounds, that impedes the downstream in-gel processing (Wilmes and Bond, 2004; Park and Novak, 2007; Abram et al., 2009; Kuhn,

2010; Kuhn et al., 2011). In an ongoing effort to determine the origin of these contaminants, samples have been made ready for nuclear magnetic resonance (NMR). NMR have previously been used on laboratory scale SBR's containing granular sludge flocs to determine the exact structure of one polysaccharide (Seviour et al., 2010). Therefore it might to possible to provide evidence of contaminating polysaccharides in the protein extracts. Though the authors of the previous study commented that an exact structure could not be obtained for regular sludge flocs, NMR might provide a useful fingerprint of the sample make-up. Possible ways to exclude these contaminants have been the application of sequential ammonium sulfate, in which protein pellets were washed with increased amounts of ammonium sulfate (Park et al., 2008). This method have shown to produce purer samples of proteins, though excluding some proteins, which was observed on SDS-PAGE gels. This have motivated the usage of ammonium sulfate, and this method is being investigating in ongoing experiments. In addition, several studies have applied an acid precipitation on the crude extract prior to the sequential ammonium precipitation, and though this was not tested in this research it seems prudent to test this in further studies (Wilmes and Bond, 2004; Broekman, 2007; Kuhn et al., 2011; Barr et al., 2011).

It was sought to find an alternative chelating agent than EDTA, and it was observed that previous methods have applied the use of cation-exchange resin (CER) (Frølund et al., 1996). Several studies have used this method to extract EPS proteins (Park and Novak, 2007; Yu et al., 2007; Barr et al., 2011), which motivated the application method. The results from the CER experiments are discussed in the following.

4.3.2 Extraction using Cation Exchange Resin

The next group of experiments was based on Strategy 1B, and was based on the application of cation-exchange resin, which is a widely used method (Frølund et al., 1996; Park and Novak, 2007; Barr et al., 2011). The extraction using cation exchange resin was applied as an alternative to the use of EDTA, and is thought to be more gentle to activated sludge. The extraction method is based on the adapted Frølund et al. (1996) method proposed by Barr et al. (2011). The two hour incubation on ice was proposed by Frølund et al. (1996), and following the chloroform/methanol precipitation, the samples were run on a SDS-PAGE gel, as shown in Figure 4.9. Because no clear bands was observed on the SDS-PAGE gel each lane of WWTP West, and WWTP East was excised into four equal sized corresponding to 25% of the entire lane, and processed for LC-MS/MS. Due to lack of time this experiment was not repeated, and it was therefore not possible to compare with other gels. In addition, the results from subsequent LC-MS/MS is currently pending. From the knowledge obtained from the literature it is known that extraction yield from CER extraction varies greatly between different WWTP's, see Appendix C, and it would therefore be interesting to evaluate the extraction yield, e.g. by modified Lowry.

Though Strategy 1B was not thoroughly studied, due to lack of time, it still seems to be an adequate alternative to the EDTA based methods, and the use of CER is used in ongoing work.

4.3.3 Extraction following Broad Spectrum Enzymatic Treatment

Strategy 1C was applied as a completely different apporach to the commonly used methods. The subsequent extraction of EPS proteins following Viscozyme treatment was a novel approach based on inducing defloculation by degrading the various polysaccharides in the EPS layer. Due to the diversity of the polysaccharides it was thought that by applying a broad spectrum of polysaccharide degrading enzymes at least some of the enzymes would be able to catalyze efficient degradation. The Viscozyme mixture is a commercial product, and the concentration was only given

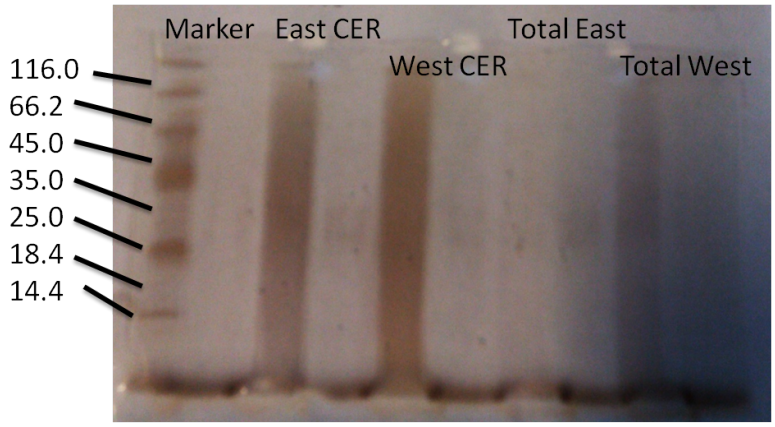


Figure 4.9: The obtained 4%/7.5% SDS-PAGE gel. It is seen that the CER extraction from WWTP West resulted in a more dense silverstaining than WWTP East. In a separate repeat experiment total extracts were run for both WWTP's.

in activity units. A dilution series on a SDS-PAGE gel was made, and it was finally concluded that the 3 μL of the mixture resulted in a adequate silverstained lanes. The extracted sample from the Viscozyme experiment was shown in Figure 4.1, and resulted in a densely silverstained lane (Lane 7). The extracted sample were precipitated using chloroform/methanol, and it seemed that it the lane contained at least one distinct band, which motivated further in-gel processing. The subsequent results from the LC-MS/MS is discussed in the Section 4.5.

The Viscozyme extraction method tested following the EDTA-based extraction, and a few concerns were hypothesized. The first was that the EDTA-based extraction method were only able to produce a few identified proteins. The addition of Viscozyme enzymes introduces numerous proteins, which would be sampled during subsequent LC-MS/MS analysis. A potential issue would then arise if these few extracted proteins were diluted with Viscozyme enzymes. From the analysis of the SDS-PAGE gel of the Viscozyme standard it was clear that the mixture contained numerous enzymes in the range 50-100. A trypsin digestion of the Viscozyme mixture would then result in an amount of peptides that would effectively dilute the EPS proteins of interest. However, it is very likely that this method can be optimized to produce interesting results. In addition, the dilution issue, becomes less of a problem if the sample is injected into a rapid-scanning mass analyzer, such as a LTQ-Orbitrap, which is also commonly used for activated sludge samples (Wilmes and Bond, 2006b).

4.4 Total Protein Extraction

The entire metaproteome was sought to be obtained using the total protein extraction method described in Section 3.1.8. This method was the only method applied in Strategy 2, and though more methods exists, these was not tested, because they were based on base extraction. It was thought that the addition of base to the activated sludge would create a massive stress response, and that the subsequent protein analysis would provide a distorted image of the protein complement. In addition, in a preliminary study performed prior to this research the use of base, and sulphide produced poor results.

Samples were treated using the same washing procedure, with a subsequent extraction buffer, which included protease inhibitors to avoid protein degradation. Sonication were utilized as an alternative to the french press used by Wilmes and Bond (2004), to lyse the activated sludge cells.

Samples were split into four separate samples, which were subsequently exposed to a three times 30 *sec* sonication at an amplitude of 180 μm, 144 μm, 108 μm, and 72 μm, respectively, which is shown in Figure 4.10 for WWTP West, and Figure 4.11.

The metaproteome extraction from WWTP West was observed to produce highly smeared lanes on the SDS-PAGE gel indicating that the protein content was very high. This was expected, and the reason why each sample were separated by an empty lane. It seemed that the lane exposed to the least intense sonication, i.e. the 40% sample, resulted in the most intensively silverstained lane. The samples 60%, 80%, and 100% samples were relatively less intense in decreasing order. It did however seem that each of lanes contained the same silverstaining pattern, and consequently only one lane - Lane 1 - was processed by in-gel digestion. Because no distinct bands were visible the lane was excised into four separate samples, as described in the CER extraction method.

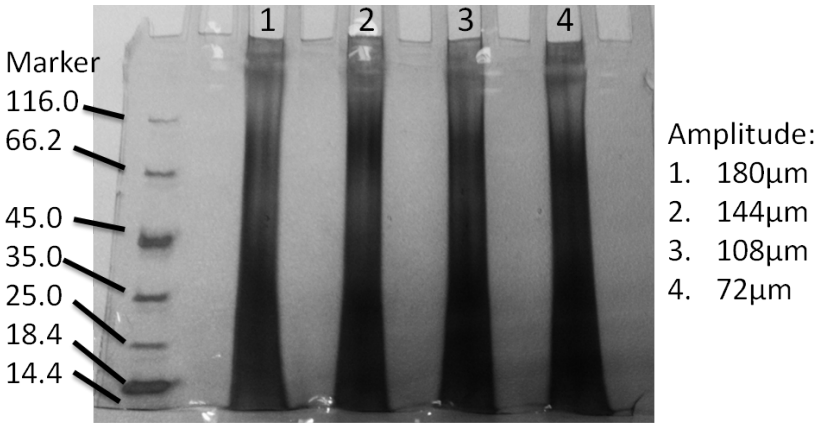


Figure 4.10: The obtained 4%/7.5% SDS-PAGE gel from the metaproteome extraction from WWTP West. One sample was extracted and split into four separate that samples that were subsequenlty exposed to varying sonication amplitude, as shown on the figure.

The metaproteome extraction from WWTP East resulted in a similar SDS-PAGE gel. However, each lane seemed to equally silverstained in this sample.

Several studies have commented that the intense smearing/streaking on SDS-PAGE, and 2D-PAGE gels, are caused by contaminating high-molecular weight polymers (Kuhn, 2010; Kuhn et al., 2011). Streaking is a common phenomenon on 2D-PAGE gels, which causes increased spot sizes, and consequently decreases the resolution (O'Farrell, 1975). For example, Kuhn (2010) observed intense smearing on SDS-PAGE gels, which motivated them to include washing steps. Though the washing removed some of the smearing, they observed that proteins were lost in the washing steps.

The extraction of metaproteome using SDS-PAGE have previously been discussed, and it have been commented that SDS-PAGE gels does not provide an adequate resolution for large-scale identification (Wilmes and Bond, 2004). SDS-PAGE gels have been utilized previously to study proteins in activated sludge by Ogunseitan (1993), and was subsequently adapted by Ehlers and Cloete (1999a,b) for the study of actively protein transcription in EBPR plants. The identification of proteins these studies were limited to protein profiles, which was thought to likely be attributed to the complexity of full-scale WWTP's, and the limitations of 1D SDS-PAGE (Wilmes and Bond, 2004). It seems this is also the case in this study. Both of the gels presented above, including the repeated metaproteome extraction shown in Figure 4.9 resulted in smeared lanes, where no distinct

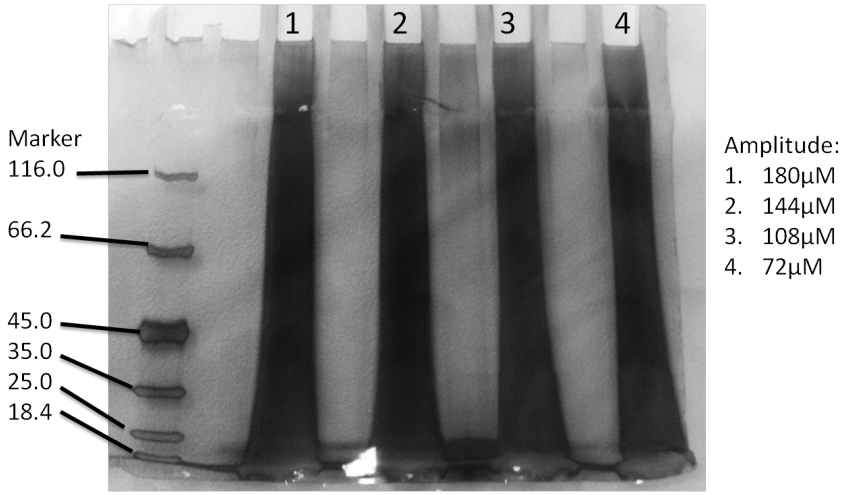


Figure 4.11: The obtained 4%/7.5% SDS-PAGE gel from the metaproteome extraction from WWTP East. One sample was extracted and split into four separate that samples that were subsequenlty exposed to varying sonication amplitude, as shown on the figure.

bands were visible. This might not be surprising as the previous 2D-studies were able to identify between 200, and 210 for full scale WWTP's, and between 590, and 630 proteins for laboratory scale SBR's (Kuhn et al., 2011; Wilmes and Bond, 2006b), which would be difficult or impossible to resolve on a 1D SDS-PAGE gel.

Though the metaproteome extractions only provided preliminary result it seemed from the subsequent MS data, that only few peptide ions were observed throughout each run. It therefore seems that the samples were also impeded by contaminating compounds, which suppress the identification rate in LC-MS/MS analysis. The original study by Wilmes and Bond (2004) utilized a Q-ToF, instrument, which was also used in this research, but they switched to use LTQ, and LTQ-Orbitraps (Wilmes et al., 2008). These latter instruments are able to sequence faster, and provide more sample depth, but it does not seem that the this is an issue in this research. Therefore, the extraction method, and subsequent protein separations needs to be optimized to provide a higher purity of proteins, as it currently seems to be the biggest issue.

4.5 Mass Spectrometry Data

The resulting SDS-PAGE gels from EPS protein extractions, and metaproteome extractions, were processed by in-gel digestion, and subsequently analyzed using LC-MS/MS. In general, the resulting mass spectra did only produce few protein identification, which will be discussed in the following. The protein marker of L-dedydrogenase was arbitrarily selected to provide a positive hit, and to evaluate the in-gel digestion. Hence, the resulting LC-MS/MS data was only processed if the marker protein was identified with a high coverage. This was the case for all in-gel digested samples, which demonstrates that the in-gel digestion method was efficient - at least for a standard protein.

4.5.1 Extraceullar Protein Identification

Samples from the extracellular extraction using EDTA, and TritonX-100 were processed for in-gel digestion, as shown in Figure 4.2. Only few proteins were identified from these samples. Most importantly, several samples were discovered to contain an identifiable protein. The human protein CEL3A was identified with a MASCOT score between 54, and 121 an accumulated count of six unique peptides, as shown in Table 4.1. At first this protein was neglected as a contaminant, but others have discovered this protein, which emphasized that this was in fact a positive identification (Park et al., 2008; Kuhn et al., 2011). Human keratin proteins was also identified, and have been reported before, but these proteins do not possess enzymatic activity, and the origin of these protein was too uncertain to include in this research - they are common contaminants in the in-gel digestion process. In general, the peptide scores, and in extension the protein score, was low. In comparison a BSA standard is easily able to produce a protein score of 349, with 30 peptides ranging between score 29-93. The peptides identified in this study is therefore in the low end of peptide score and number and, protein score.

Lane 7 Band 2	Protein score 54					
Observed m/z	Moluclear weight *Da*	Molecular weight calculated *Da*	Δmass *ppm*	Score	Precursor charges	Sequence
857.45577	1712.8968	1712.9209	14	25	2	LYTNGPLPDKLQQAR
912.9857	1823.9568	1823.9669	6	17	2	DLTYQVVLGEYNLAVK
873.0904	2616.2493	2616.2649	6	13	3	VVHGEDAVPYSWPWQVSLQYEK
Lane 7 Band 3	Protein score 56					
Observed m/z	Moluclear weight *Da*	Molecular weight calculated *Da*	Δmass *ppm*	Score	Precursor charges	Sequence
857.4570	1712.8995	1712.9209	13	30	2	LYTNGPLPDKLQQAR
912.9822	1823.9499	1823.9669	9	27	2	DLTYQVVLGEYNLAVK
Lane 8 Band 2	Protein score 68					
Observed m/z	Moluclear weight *Da*	Molecular weight calculated *Da*	Δmass *ppm*	Score	Precursor charges	Sequence
582.2821	1162.5496	1162.5560	5	29	2	WNWWGSTVK
912.9840	1823.9534	1823.9669	7	26	2	DLTYQVVLGEYNLAVK
873.0903	2616.2492	2616.2649	12	13	3	VVHGEDAVPYSWPWQVSLQYEK
Lane 8 U1	Protein score 121					
Observed m/z	Moluclear weight *Da*	Molecular weight calculated *Da*	Δmass *ppm*	Score	Precursor charges	Sequence
582.2844	1162.5543	1162.5560	1	13	2	WNWWGSTVK
646.3265	1290.6385	1290.6509	10	28	2	WNWWGSTVKK
857.4569	1712.8992	1712.9209	13	44	2	LYTNGPLPDKLQQAR
859.4181	1716.8216	1716.8359	8	36	2	VSAFIDWIEETIASH

Table 4.1: A detailed overview of the results obtained from samples in Strategy 1A,C. The observed m/z was the measured values of the precursor ions, which was calculated into a calculated weight. These values were compared to the theoretical weights, and the difference was reported in ppm. In addition, the MASCOT score, precursor ion charge, and the sequence of the identified peptide is given.

The CEL3A is a chymotrypsin-like elastase, which belongs to the chymotrypsin-like elastase family member 3A, and is better known as Elastase-3A. The protein contains a signal peptide, and after cleavage the protein molecular weight is about 29 *kDa*. This protein have also been identified in wastewater previously by base extraction, where it was found relatively close to the 29 *kDa* (Park et al., 2008), or close to 25 *Da* (Kuhn, 2010). The identified Elastase-3A in this research was located around 20 *kDA*, which might indicate a partial degradation, which was also noted by the two aforementioned studies.

The number of identified peptides, which have been detected previously for the Elastase-3A are 8 (Park et al., 2008), and 9 (Kuhn, 2010). In comparison, the three peptides identified in this research was also identified in the latter study. The Elastase-3A is an isoform of the human pancreatic elastases, which are serine proteases. It is secreted to the intestines, where it hydrolyzes

e.g. elastasin, as well as fibrous, and insoluble proteins, thus aiding with digestion (Park et al., 2008). The protein is possibly very resistant to degradation, because it is able to pass unharmed through the intestines, and it has been detected in human stool samples (Weiss et al., 2006; Kuhn, 2010). It seems that the protein is mostly originating from human, and possibly cows, where it is transferred to stool to the sewers, where it ultimately ends up in the activated sludge. Hence, it have been suggested as a biomarker for faecal contamination, due to its surprisingly high stability (Kuhn, 2010).

Summing up on the EPS extraction method it is clear that further optimization is required to obtain a higher identification rate. The current trend in the community seems to be focused on the metaproteome extraction. These extraction does not lead to potential loss of sample in the same the EPS protein extraction method does. The potential loss EPS proteins seems to be after the application of extraction buffer, when the activated sludge is discarded from the downstream processing. The metaproteome extractions might prove not to loose sample in the same, and the identification of EPS proteins may be possible by applying brute-force instead of refined methods. The addition of rapid scanning mass analyzers to environmental studies have shown to provide hundreds of proteins (Wilmes and Bond, 2006b). Though these studies was focused on metabolic pathways, an in-depth data-mining of the resulting MS data may provide indications of EPS proteins.

4.5.2 Metaproteome Protein Identification

As hinted in the SDS-PAGE gel evaluation of the extracted metaproteomes, it seemed that protein identifications were limited. The metaproteomes extracted from WWTP West, and WWTP East, was only processed one time for in-gel digestion, i.e. Figure 4.10, and Figure 4.11. As already mentioned, one lane from each gel was excised into four samples, illustrated in Figure 4.12.

The complete list of identified proteins by two peptides or more peptides, shown in Table 4.2, included 9 proteins. Most important, Elastase-3A was identified in both WWTP's, thus showing that it was possible to identify this protein in the metaproteome extractions, as well as the EPS protein extractions. It was striking that the molecular weight of the identified proteins were reversed compared to where they ought to appear on the SDS-PAGE gels. Supposedly, the identified proteins ought to appear at high molecular weights in Band 1, and then decrease through the other bands until Band 4. The only protein that appeared in the expected band was the spiked in BSA. This observation may provide more evidence to the limited resolution of SDS-PAGE gels, which have been commented previously (Ogunseitan, 1993; Ehlers and Cloete, 1999a,b; Wilmes and Bond, 2004).

The restriction of two peptides per protein was maintained for both samples. If this restriction was dropped it was noted that a total of 20 additional proteins could be included. These proteins were mostly proteins, which are known to be abundant inside the cells, and consequently these protein identifications were discarded with the exception of Elastase-2A, which was included in Table 4.2. Furthermore, human keratin proteins were excluded for the same reasons as explained previously. If these proteins were included an additional 13 proteins identified by a total of 28 peptides. Kuhn (2010) have argued that protein identification are lost when single-hit proteins are excluded, though it seems that this argue stands alone, as others have not included these proteins (Wilmes and Bond, 2004, 2006b; Wilmes et al., 2008; Park et al., 2008). It is debatable whether or not to include these proteins, though more peptides provide more reliable results, which was the reason to exclude single-hits in this research.

It was noted from Table 4.2, that only Elastase-3A, and the spiked in BSA were identified by more than 2 peptides. It is a likely assumption that Elastase-3A is only present in the EPS layer, based

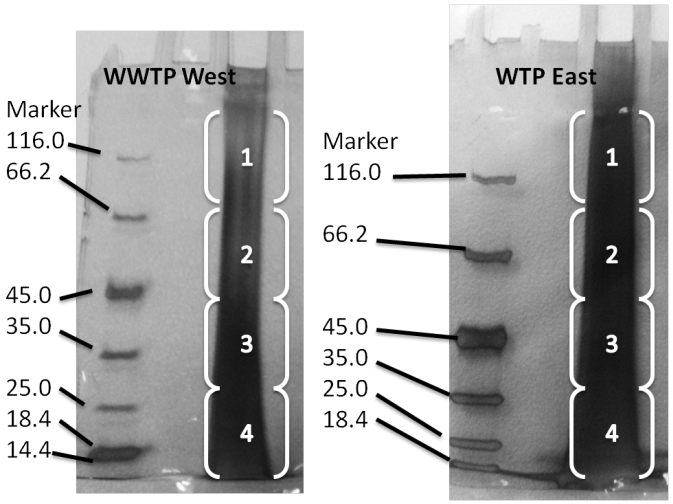


Figure 4.12: The SDS-PAGE gels, Figure 4.10, and Figure 4.11, was excised as shown in this figure. The excised bands was then processed through in-gel digestion, and analysed by LC-MS/MS. Note that this picture is included only as an illustration of the excision concept, thus the brackets are only rough indications of where the gels were actually excised.

Protein	Best matching species	MASCOT Score	Number of peptides	Molecular weight	Identified in sample
BSA*	Bos Taurus	113	4	69,293 *Da*	WWTP West Band 1
Lysozyme C	Phasianus versicolor	55	2	14,305 *Da*	WWTP West Band 1
Elastase-3A	Homo sapiens	148	5	29,470 *Da*	WWTP West Band 3
Elastase-2A**	Homo sapiens	90	1	28,869 *Da*	WWTP West Band 3
Chymotrypsin-C	Homo sapiens	38	2	34,362 *Da*	WWTP West Band 3
Elongation factor Tu	Herminiimonas arsenicoxydans	70	2	42,970 *Da*	WWTP West Band 4
ATP synthase subunit-β	Acidovorax avenae subsp.	40	2	51,345 *Da*	WWTP West Band 4
ATP synthase subunit-α	Aromatoleum aromaticum	56	2	55,502 *Da*	WWTP East Band 4
Elastase-3A	Homo sapiens	47	4	29,470 *Da*	WWTP East Band 4

Table 4.2: The complete list of protein identified with at least two peptides. *This sample was a spike in. **This protein identification was included, though it was only identified with 1 peptides, because it may prove relevant compared to the Elastase-3A.

on the EPS protein extractions, and similar results in the literature (Kuhn, 2010). The reason for the low identification rate compared to the hundreds of proteins identified by e.g. (Wilmes and Bond, 2006b) could originate from two issues. The first issue could be the recurrent indication of contaminants in the sample due to low purity samples. The other could be that the sonication step in the extraction needs to be further scrutinized in order to provide a more extensive lysis of cells. Additional steps could also be added to the extraction. For example, it was shown recently that a pre-treatment of activated sludge in liquid nitrogen followed by grinding in a mortar have provided optimization of 2D-PAGE gels, which could likely be extended to SDS-PAGE gels as well (Abram et al., 2009). The application of a mortar step was used for granular sludge, and may not be compatible to sludge flocs. The optimized extraction then utilized the UTCHAPS extraction buffer, followed by two subsequent cycles of DNase treatment. It has recently been shown that DNase treatment enables defloculation of activated sludge (Dominiak et al., 2011), and it therefore seems likely that this treatment also makes the cells less resistant to the applied extraction methods.

Based on the results from this research, and the reported optimization methods from the literature it is therefore recommended that optimization ought to be focused on the application of sufficient

sonication, pre-treatment of liquid nitrogen, and the addition of DNase treatment.

4.6 Significance of Study

In an effort to create a better understanding of activated sludge, and the EPS layer that constitutes the significant part of the this entity, several extraction methods were tested, and evaluated. As mentioned in the Introduction, this research is on-going, and the experience gained from this study have motivated the further investigation in order to obtain an adequate strategy for extraction. The EPS make-up is currently seen as 'dark matter of biofilms', because it is notoriously difficult to sample, and analyze, due the vast complexity of the various EPS components (Flemming et al., 2007). This has indeed also been observed in this research. Consequently, due to the lack of detailed knowledge the further investigation of this dark matter is of high priority, and relevance.

It is probably no coincidence that the very group that presented the first sufficient extraction strategy for activated wastewater (Wilmes and Bond, 2004) is still currently working with laboratory scale reactors. Using this strategy they have provided the largest, and most comprehensive datasets of metaproteomes in EBPR, which have provided valuable insights into the metabolic pathways (Wilmes and Bond, 2006b; Wilmes et al., 2008; Barr et al., 2011). It is very likely that the reduction of complexity of the community, and available substrates, makes the protein extraction more manageable.

This study have shown that the extraction method development, and evaluation is a tedious exercise. It has been commented by others, that the development of method is a long process, and the addition of steps to remove contaminant makes the extraction time consuming (Kuhn, 2010). However, this study has provided valuable practical experience, which is being applied as a basis for optimization and further development. The results indicate that more knowledge about the biological system must be included, or simply that new innovative methods should be invented.

5 Conclusion

The aim of this research was to implement extracellular protein, and total protein extraction extraction methods on activated sludge from two wastewater treatment plants, Aalborg West, and Aalborg East. This aim was achieved, but it was discussed that the obtained SDS-PAGE gels were heavily silverstained, and thus supposedly contained large amounts of proteins. Subsequent, analysis by LC-MS/MS revealed that only few proteins were identifiable. This was suggested to be due to contaminants that was co-purified in the extraction process, and that these contaminants reduced the compatibility with LC-MS/MS. It was suggested to utilize nuclear magnetic resonance to obtain a fingerprint of the sample make-up.

The results from this study indicated that the extraction methods used were insufficient in extracting the proteins. Several optimization steps were suggested for the extraction methods, including sonication, flash-freezing samples in liquid nitrogen, and DNase treatment. In addition, it was discussed to use prolonged precipitation methods, including acid precipitation, and sequential ammonium sulfate precipitation. In order to further optimize the extraction process a targeted strategy was suggested to determine if protein were lost during the extraction process by using the Modified Lowry method.

Protein were hypothesized to be located only in the activated sludge, and in-solution extractions confirmed that this was the case. It was also hypothesized that the database specific database would produce more protein identification, but this was not the case. This is mainly attributed to the impurity of the extracted samples, so this hypothesis was neither confirmed nor rejected.

Protein standards of bovine serum albumin, and ovalbumin were spiked in samples to provide positive hits, and it was found that these proteins were, in fact, extractable from the sample using all strategies. A more detailed analysis revealed that the activated had a low binding capacity, which is presented in Section 4.3.

The human protein Elastase-3A were identified, thus providing more evidence of the existence of this highly resistant protein in wastewater treatment plants. The results indicated that this protein ends up in the WWTP, and is captured by the activated sludge flocs in the extracellular polymeric substance (EPS) layer.

This research has provided the necessary knowledge required for further optimization, and ongoing experiments are being applied to further investigate, and ultimately to identify novel EPS proteins.

6 Future Perspectives

Waste water treatment is an important part of the modern society and sustainability. The continuous effort to optimize and develop new methods for the purification of wastewater is important, because fresh water is becoming increasingly depleted. The issue of water shortage must be taken seriously, and it is no coincidence that the field of water purification has a tremendous interest in the industry, e.g. Alfa Laval, Krüger, Novozymes, and Veolia Environment.

The wastewater treatment plants (WWTP) investigated in this research, and 4 additional plants, purifies water from 225,000 citizens annually, which corresponds to 15,000 liters of water per citizen, (Kommune, Kommune). WWTP West, and WWTP East, are two of the largest WWTP's in Denmark, and these plants are high technological plants that are continuously renewed and updated with modern methods. These modern plants enables purification of 1,000 liters per second, and the quality of the effluent water is much lower than the EU regulations (Kommune, Kommune). The removed phosphorous are recovered by dewatering excess sludge, and for example WWTP West generates 25-30 ton/day, or about 9,600 tons of sludge annually. The recovered sludge is then used as fertilizer for the agriculture, thus recycling phosphorous, which decreases the need for new generation of fertilizer (Kommune, Kommune). Due to new regulations the excess sludge were not allowed to be used as fertilizer, which motivated the use of excess sludge as starting material for the production of concrete. The recent application of excess sludge in biogas plants, and in this way the WWTP's generates environmentally friendly electricity, which have reduced the carbon dioxide emission between 1997-2003 by by 33%.

It is unquestioned that wastewater must be purified to avoid eutrophication (Oehmen et al., 2007). The WWTP's in Denmark provide robust removal of phosphorous from the applied enhanced biological phosphorus removal strategy (Srinath et al., 1959), and in this way relieves the aquatic systems for the potential dangers of pollution. The next step in wastewater treatment is to remove anthropogenic waste products, e.g. medical waste (Wilmes and Bond, 2004). For example it have been investigated that two WWTP's at the Danish capital Copenhagen emits medical waste products into the immediate aquatic systems (Djursing, 2008). In addition, emissions of heavy metals above the regulated quota, such as mercury, have also recently been observed (Marfelt and Wittrup, 2011). The WWTP's are not yet equipped to deal efficiently with these toxic compounds, which calls for new the innovative methods to resolve this issues.

This research, and the other studies included in the EcoDesign portfolio was funded by the Danish Council for Strategic Research in 2010 in an effort to bring wastewater treatment into a new boost of evolution. The scientific effort is joined with most of the other universities in Denmark, and publications, and solutions are increasingly generated. By knowing more about the community ecology, and by applying new innovative methods we believe that we can wastewater treatment an even more environmental and economically sustainable solution.

http://www.ecodesign.aau.dk/

A Selected Metaprotemic Studies of Waste Water Treatment Plants

Metaproteomic studies of full-scale wastewater treatment plants (WWTP) have proved to be a difficult tasks, and as a likely result the number of studies are limited (Park and Novak, 2007; Park et al., 2008; Park and Novak, 2009). In this chapter an extensive literature study have been made in an effort to compile methods previously used, and the important results that have been published. It is initiated by introducing the important metaproteomic studies of activated sludge systems, and the methods used for extracting proteins from these systems. It then covers an extensive review of the various quantitative protein assays, which results in the preferred method for protein quantitation for activated sludge. This chapter provided the reference work for the extraction method development, and evaluation used in this study.

The operation of WWTP's is a well characterized and established method in mechanical civil engineering. However, the biological processes in WWTP is yet to be fully understood. The application of these methods are rapidly developing post the human genome project (Venter and et. al., 2001). In recent years the application of genomics, transcriptomics, and proteomics have undergone continuing development, and the fields have matured enough to encompass highly complex bioligcal systems, among others WWTP's. The elucidation of biological processes have thus moved toward understanding of microbial communities, and the corresponding components of the central dogma is now described as meta-genomics, meta-transcriptomics, and meta-proteomics (VerBerkmoes et al., 2009) (Wilmes and Bond, 2006b) (Wilmes and Bond, 2009). The application of these fields requires extensive knowledge, and it is beyond the scope of this section provide details of each field. In this section the characterization of EPS protein through mass spectrometry based proteomics is extensively covered, with focus on extraction strategies. Currently, no generic purification strategies exists for full-scale WWTP. Research have so far been focused on less complex laboratory scale sequencing batch reactors (SBR), which have provided valuable insights into metabolic pathways in the activated sludge system (Wilmes and Bond, 2004). However, it has been attempted to purify proteins from full-scale plants, and these methods have been focused on SDS-PAGE (Park et al., 2008) (Kuhn et al., 2011), and 2D-PAGE (Kuhn et al., 2011). Though these methods did provide protein identifications it still seems that a full proteome, and extracellular metaproteome from activated sludge has yet to come.

Originally, EPS protein characterisation in activated sludge systems were focused on measuring the quantities of protein (Raunkjær et al., 1994). These results determined that proteins were abundant in the systems, in addition to the other EPS components, and that the biomass only constituted between 10-15% of the total organic fraction (Frølund et al., 1996). It was observed that extracted EPS proteins possessed a high enzyme activity, e.g. esterase, glucosidase, and lipase activity (Frølund et al., 1995). Though these methods did not have the resolution to determine the proteins responsible for the different activities, they did provide a motivation for further research. The first research to use mass spectrometry (MS) based proteomics was performed by Wilmes and Bond (2004), and showed the feasibility of this method. The general workflow of MS based proteomics is splitted into sample generation, sample preparation, LC-MS/MS, and subsequent validation experiments, as shown in Figure A.1. In full-scale WWTP's the sample generation is conviniently performed by the plant operator. However, this makes it more difficult to control the experiments, and samples may vary significantly according to seasons, or shock events such as

heavy rainfall. Consequently, focus is on the sample preparation, which is still under development. This part is covered in the following Section A.1. The subsequent LC-MS/MS analysis is performed by standard data-dependent acquisition, which is the preferred method for large-scale identifications (Aebersold and Mann, 2003). This part is covered in Section B, while validation is covered in Section B.2.

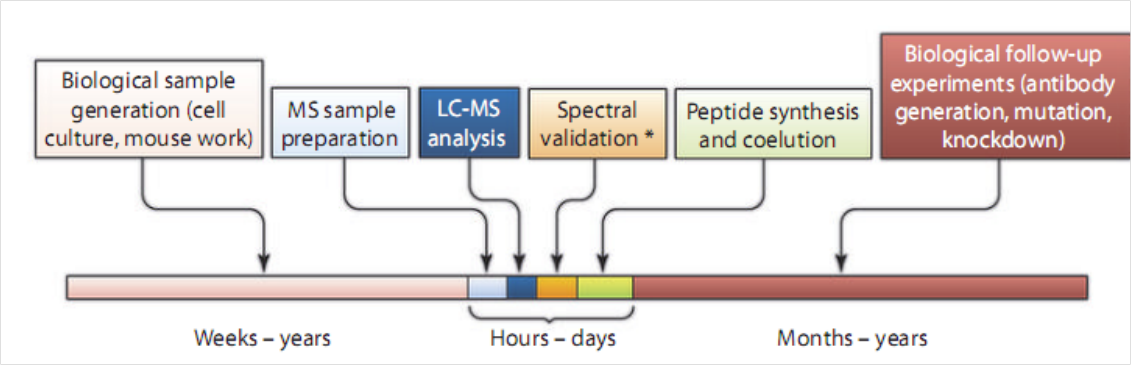


Figure A.1: The workflow of MS based proteomics is splitted into sample generation, sample preparation, LC-MS/MS, and subsequent validation experiments. (White, 2011)

A.1 Extraction by Wilmes and Bond (2004, 2006b); Wilmes et al. (2008)

The identification of proteins via. LC-MS/MS begins with sample preparation. It is necessary to remove contaminants that interfere, and contaminates the LC-MS/MS, such as detergents and salts (de Hoffman and Stroobant, 2009). In large-scale protein identification studies the sample is often comes from live cells - in this study the cells are the activated sludge - and thes sample contains all the different kinds of proteins: cytosolic proteins, membrane- or membrane bound proteins, extracellular proteins, etc. Depending on the analytical interest the cells can be exposed to changes or stresses, while a control sample in homeostasis can analyzed in parallel to determine changes in protein expression (Aebersold and Mann, 2003). This has for example also has been done for activated sludge to investigate the difference between a laboratory reactor peforming EBPR, and a laboratory reactor that did not perform sufficient EBPR (Wilmes and Bond, 2006b).

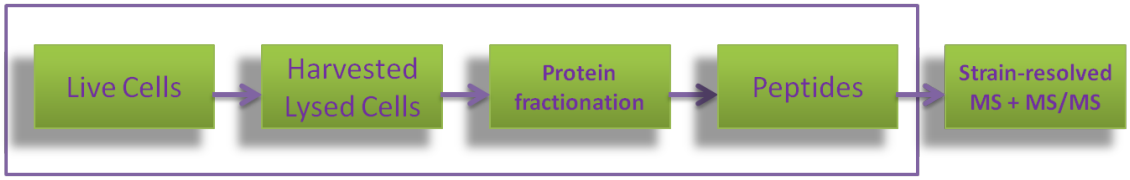


Figure A.2: fads

The characterization of EPS proteins have been done using several different strategies. SDS-PAGE gel has been utilized previously to study protein expression of waste water by Ogunseitan (1993), and was subsequently adapted by Ehlers and Cloete (1999a,b) for the study of actively protein transcription in EBPR plants. The identification of proteins these studies were limited to protein profiles, which was thought to likely be attributed to the complexity of full-scale WWTP's, and the limitations of 1D SDS-PAGE (Wilmes and Bond, 2004). A sample workflow was introduced

by Wilmes and Bond (2004), and included sample extraction, release of EPS components, extraction of EPS proteins, separation of EPS proteins, LC-MS/MS analysis, and subsequent database searching, as shown in Figure A.2. This method was applied to laboratory sequencing batch reactors (SBR) to reduce the complexity by feeding the reactor with synthetic waste water. This method have since been used as a more controllable model system to construct metabolic pathways (Wilmes and Bond, 2006b). 2D SDS-PAGE was utilized to improve the resolution of the protein separation (Wilmes and Bond, 2004), and to enable quantitation on the 2D gel (Wilmes and Bond, 2006b). Wilmes and Bond (2004) developed an extraction method, shown in Figure A.3, for the extraction of the entire community proteome, i.e. the metaproteome. In more detail, the sampled sludge is washed with 0.9% sodium chlorid, which is thought to remove exopolysaccharides. The lysis buffer, containing $7\,M$ urea, $2\,M$ thiourea, 4% (w/v) CHAPS, $10\,mMTris$, $1\,mMEDTA$, $50\,mM$ dithiothreitol (DTT), and protease inhibitor, is designed to be highly denaturing, and to extract bound di-valent ions, the latter accomblished by EDTA as shown in Figure A.4. It is build upon two concepts. Firstly, by removing bound ions in the biofilm it breakes apart, and becomes more soluable (Frølund et al., 1996). Secondly, the extensive use of denaturants solubilised the hydrophobic proteins, and the cell membranes are destabilised so much that the cells lyse and releases the cytosolic content. The lysis potentially introduces proteases, which are inhibited by the protease inhibitors (Kielberg and Rasmussen, 2005). Following the lysis step, the samples are precipitated by acidifying, and then extensively washed to remove excess reagents from the lysis buffer. The resulting extract is then resuspended in low stringency buffer, centrifugated, and the subsequent supernatant is transferred and referred to as the low stringency fraction. The remaining pellet is then resuspended in high stringency buffer, centrifugated, and the subsequent supernatant is then transferred and referred to as the high stringency fraction. Both samples are then transferred to isoelectric focusing strips for isolectric focusing (IEF), and subsequently furhter separated on a SDS-PAGE gel. Subsequently, protein spots were extracted and processed for MALDI-TOF MS, and Q-TOF MS.

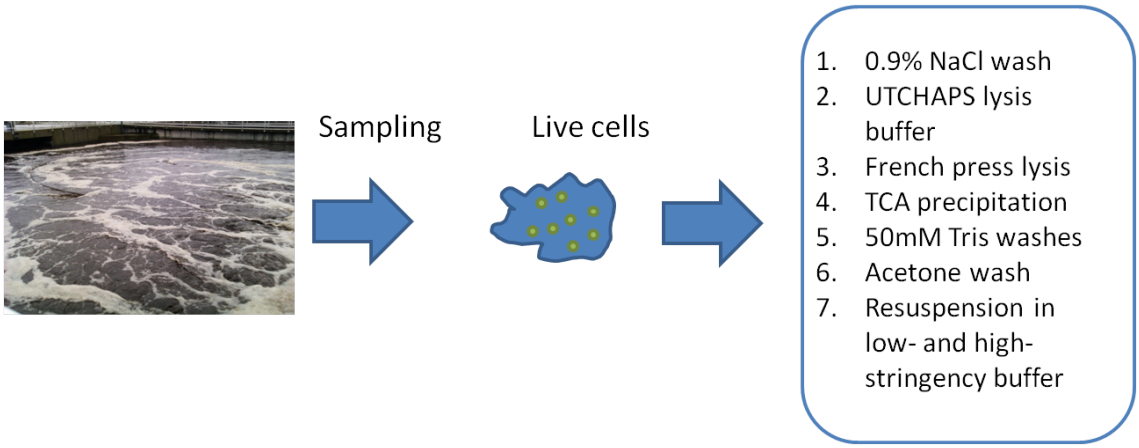


Figure A.3: The first sample preparation work flow for subsequent 2D SDS-PAGE proposed by (Wilmes and Bond, 2004). This work flow was then succeeded by LC-MS/MS for protein identification.

In a follow up study Wilmes and Bond (2006b) used the same purification strategy (Wilmes and Bond, 2004) in a quantitative manner by staining the gels with a fluorescent dye, SyproRuby, to investigate the difference between two enriched SBR. One reactor was performing EBPR, while the other was not performing sufficient EBPR, which resulted in two distinct 2D gels. Interestingly, they investigated the different phases of waste water treatment by sampling from the anaerobic

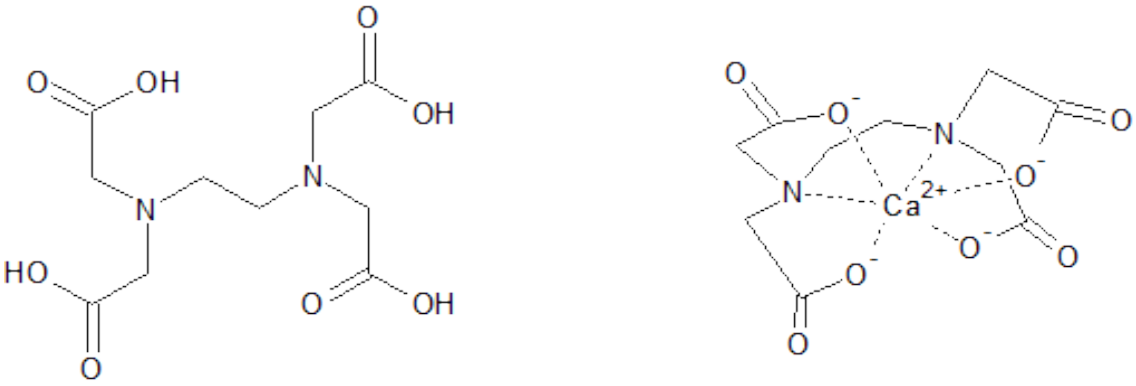


Figure A.4: The structure of etylenediaminetetraacetic acid (EDTA) shown in the left. The binding mechanism of EDTA is shown to the right, where a calcium ion is shown to be chelated in the center of the molecule. (Hart and Grace, 2005)

and aerobic phase, which resulted in 630 protein spots for the EBPR SBR and 590 protein spots for the non-performing EBPR SBR. It was noted that the result indicated that the protein expression changed in response to the cycling between anaerobic and aerobic phase. This was quantified further revealing that 9.4% for the EBPR SBR, and 14.7% for non-performing EBPR SBR of all spots were statistically different between the two phases, thus providing further evidence for the changes in protein expression patterns. Additional observation showed that the metaproteomes were very different, which was attributed to the fact that the microbial community structures were dissimilar. As mentioned earlier, SyproRuby was used, because it provides a good basis for quantitation. However, silverstaining is much more sensitive for metaproteomic mapping, due to higher sensitivity, but it was observed by Wilmes and Bond (2006b) that silverstained gels were less reproducible. It is therefore noted that if the 2D SDS-page gels are to be measured quantitatively on-gel SyproRuby is the choice, while silverstaining potentially leads to a higher discovery number due to higher sensitivity. With a large focus on quantitative proteomics it is therefore expected that SyproRuby will become unnecessary unless rapid information is needed (Ong and Mann, 2005).

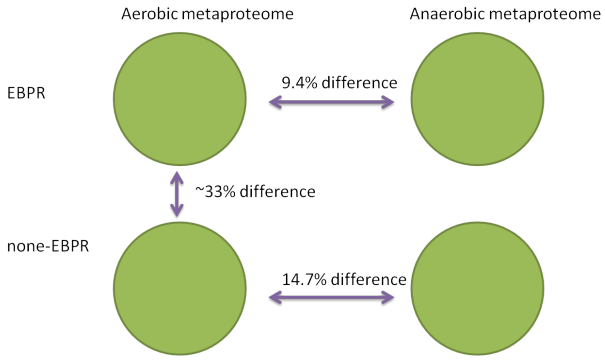


Figure A.5: Illustration of the four different SBR's. (Wilmes and Bond, 2006b)

More recently, community proteogenomics was used not only to determine strain-variant protein expression, but also to evaluate major metabolic pathways in EBPR SBR's (Wilmes et al., 2008). Four years after publishing the 2D SDS-gel approach the field of mass spectrometry had evolved rapidly. Instead of using a QToF two instruments were used - a LTQ, and a hybrid LTQ

Orbitrap both connected to a 2D chromatography setup consisting of SCX-RP for high resolution separation. Though the focus of this study was *Accumulibactor Phosphatis*, and therefore an investigation to identify proteins belonging to this microorganism within the metaproteome, the feasilibty of combining DNA sequencing and proteomics was clearly shown. About 10% of all predicted proteins were found for this microorganism, which was used to validate the presence of active metabolic pathways.

A.2 Extractions by Park and Novak (2007); Park et al. (2008)

Another purification strategy employs the findings of (Frølund et al., 1995), (Frølund et al., 1996) and, (Nielsen and Keiding, 1998). Park and Novak (2007) utilized the findings, and proposed the use of three different purification methods to obtain three distinct subproteomes for full scale WWTP's, as shown in Figure A.6. It was found by that bulk sludge had a considerable enzymatic activity, and that the activity seemed to be localized in the sludge floc matrix Frølund et al. (1995, 1996). Furthermore, it was shown that applying a cation exchange resin (CER) broke the flocs apart, and that exoenzymes was released into the bulk liquid, with little or no cell lysis. It was noted though that previous results from Wilen et al. (2003) indicated that the CER method mainly extracted divalent ions, i.e. calciums ions, Ca^{2+}, and magnesium ions, Mg^{2+}, but not polyvalent ions like aluminium, Al, and iron, Fe Park and Novak (2007). One suggestion for this process was that substitution of sodium ions with calcium, and magnesium ions occurred more readily than substitution of sodium ions with the much more tightly bound polyvalent ions (Wilen et al., 2003). Thus, (Park and Novak, 2007) concluded that a subproteome 1 would contain EPS protein extracted from the fraction of EPS bound by calcium, and magnesium ions.

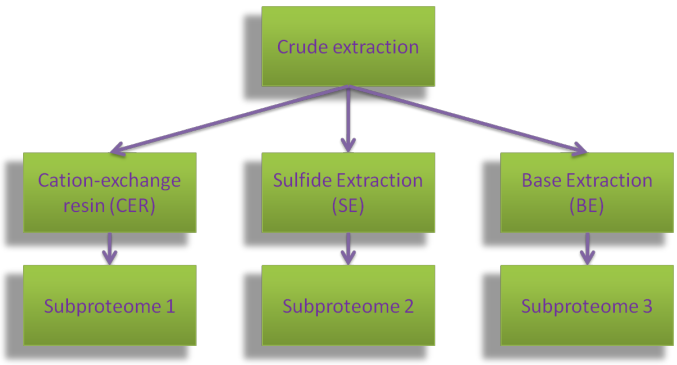


Figure A.6: The proposed strategy by (Park and Novak, 2007). The three different purifications lead to 3 distinct subset of the activated sludge proteome.

The second strategy by Park and Novak (2007) was based on the observations by Nielsen and Keiding (1998), where the relationship between sludge strength and sludge dewatering properties were investigated. It was shown that by removing Fe(III) via. addition of sulfide, the floc strength decreased making the sludge vulnerable to shear stress. By applying shear stress particles into the sample EPS was released into the bulk liquid. Thus Park and Novak (2007) used these observation to suggest a way to obtain a subproteome 2.

The third strategy of Park and Novak (2007) employs base extraction, though it seems that this method is poorly investigated. The strategy was based on the the water solubility chemistry of Al. By applying either low pH, or high pH it would be possible to solubilize Al, and as a conse-

quence release Al-bound EPS. The latter option was choosen to avoid refloculation of protonated biopolymers.

Following the proof of concepts of each of the tree strategies the three purification strategies were applied, with subsequent sequential ammonium sulfate precipitation, before SDS-PAGE (Park et al., 2008). The resulting gels were then processed for LC-MS/MS analysis on a 4000QTrap.

A.3 Extractions by Kuhn et al. (2011)

Recently, a study by Kuhn et al. (2011) used base extraction as a prefractionation method for subsequent purification using phenol base. As shown in Figure A.7, the first part (A) is a crude extraction method, followed by an extensive washing procedure, (B). Using phenol as an extraction method was originally described by Benndorf et al. (2007), and Benndorf et al. (2009), who used it extract proteins from soil and anaerobic ground water. Interestingly, fraction 1 was further fractionated using SDS, followed by miliQ water, and then 10% ammonium sulfate providing a fraction 1. This fraction was then exposed to the same procedure but with increased amounts of ammonium sulfate; 20%, and 30%, respectively. This method provided distinct bands, which indicated that further purification enabled descrimination of proteins, which was in contrast to what they had previously observed for other bacteria (Kuhn et al., 2011). In addition, results showed that contaminating material was captured in fraction 2, and the subsequent fraction 3, and 4 contained significantly reduced smear on the 2D gels (Kuhn et al., 2011).

A.3.1 Evaluation of Protein Quantitation Methods

Protein quantitation have been an important of protein research for decades. Before mass spectrometry evolved into the method of choice numerous methods were developed for various samples. However, several of these methods are not compatible with wastewater, because the samples contain all the various meterials, including the extracellular polymeric substances (EPS). For example the rapid method of 280 nm absorption is not applicable (Harris, 1987), because several other materials absorb at the same wavelength.

The quantitation of proteins in activated sludge was initiated in the 1990's with the evalutation, and adaptation of the standard protein assays to provide robust results for this complex system (Raunkjær et al., 1994). The standard protein assays included the Kjeldahl-N method (Bradstreet, 1954), the biuret method (Gornall et al., 1949), the folin-ciocalteau phenol reagent method - also known as the Lowry method (Lowry et al., 1951), bicinchoninic acid (BCA) (Smith et al., 1985), Bradford (Bradford, 1976), and silverstaining (Krystal, 1987). In short, the Kjeldal-N method involves acid digestion of the sample, which releases ammonia. The contained nitrogen in the ammonia is then measured, and the protein amount is then calculated by assuming that the proteins contain 16.5% (w/w) nitrogen. Obviously, the amount of measured protein changes according to amino acid present, but the method also includes contaminating nitrogen from other compounds within the EPS that contains nitrogen. The biuret method involves the binding of copper to the protein backbone in alkaline solution. The cross-binding results in a coloured complex, which can be calculated and compared to a standard curve. However, this method is not very sensitive and required protein content in the range of 1-20 mg for confident results, and have been observed to fail for measurement on wastewater (Kuhn, 2010). In addition, the copper-reaction was observed not to be strictly proportional for different amounts, and different proteins, which called for a more refined method (Lowry et al., 1951). The Lowry method is, in principle, a further developement of the biuret method, and is probably the most studied of the standard protein assays (Peterson,

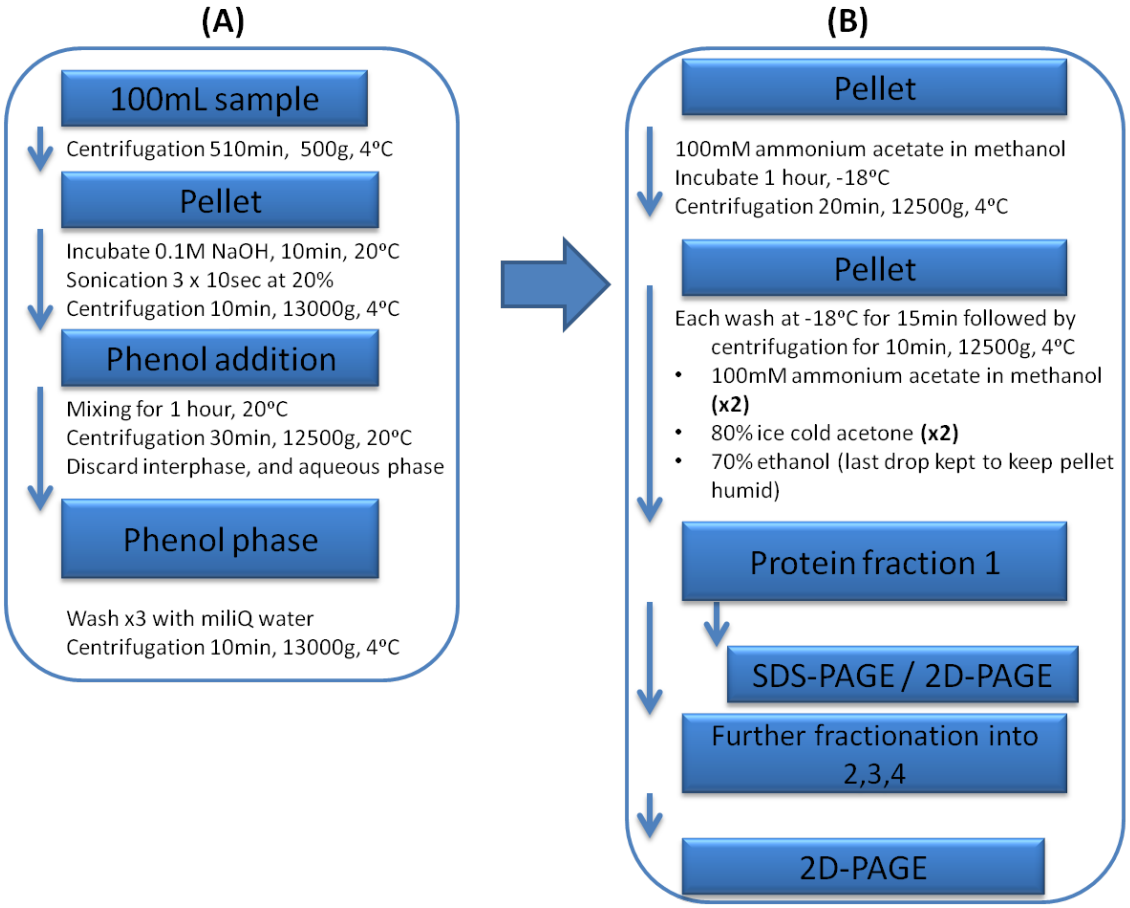


Figure A.7: (A) Extraction of proteins from sample. (B) Removal of phenol and further protein purification. Figure adapted from (Benndorf et al., 2009), method based on (Kuhn et al., 2011)

1979). It involves a two step procedure; firstly, the method also utilizes the reaction of copper in alkali, i.e. biuret method. Secondly, it employs the reduction phosphomolybdic-phosphotungstic (folin-ciocalteau phenol (FCP)) reagent by the copper-treated protein, which produces are measureable colour (Lowry et al., 1951). This reaction was shown to create proportional results, and the results indicated to be 100 times more sensitive than the biuret method, i.e. sensitivity is 0.1 mg protein/mL . A limitation of the Lowry method is that it reacts with phenol (Folin and Dennis, 1912). In addition a long list of interfering compounds has been reported including: S^{2-}, Mn(II), $(SO_3)^{2-}$, Fe(II), Fe(III), insoluble iron(III) hydroxide (Box, 1983), glycine, sodium citrate, $HgCl^2$, NaCl, EDTA, Triton X-100 (Peterson, 1979), humic acids () just to name a few. The simplest way for robust measurements in the presence of interfering compounds has been to incorporating the same amount of interfering compound into the reagent blank, which often also permits an increased tolerance for high amount of interfering compound (Peterson, 1979). Alternatively, the protein can be selectively isolated without the interfering compounds. These two method, which are widely used, and several other strategies was provided by (Peterson, 1979). The bicinchoninic acid (BCA) method is also based on the Lowry principle using BCA instead of FCP reagent. Originally suggested by Smith et al. (1985) the method is more sensitive (5 μg protein/mL), and more generally tolerant to interfering substances (Smith et al., 1985). However,

the method is sensitive to reducing sugars, which are present in wastewater making it unusable for wastewater (Raunkjær et al., 1994). The binding of Coomassie Brilliant Blue G-250 (CBBG) dye to proteins is used in the Bradford method (Bradford, 1976). This method was used to quantify proteins in larger scales (Simpson and Sonne, 1982), and is a good choice for pure protein solutions. However, the method has been showed to vary between different proteins, which is the biggest weakness of the method (Kielberg and Rasmussen, 2005).

Summing up, the different protein assays have been studied in great detail, and have been investigated for the use in wastewater (Raunkjær et al., 1994). The method of choice should have high specificity and sensitivity, and should be tolerant to the compounds in the wastewater. If the method is to be used with great confidence the different compounds in the wastewater must be monitored. It was suggested that no method were ideal, but that the most consistent method was the Lowry method (Raunkjær et al., 1994).

B Proteomics

Mass spectrometry (MS) based proteomics is currently the favored method for protein identification, post translational modification determination, because it enables the sequencing of proteins on a large high-throughput scale, and makes it possible to determine which the genes are actively being transcribed (Aebersold and Mann, 2003). MS allows the inquiry into cellular states, by sampling the entire protein complement of a cell under a given set of physiological conditions at a given time, referred to as the proteome of the cell (Wilkins et al., 1996). Proteomics enables evaluation of differences in protein expression under different sets of defined conditions, as well as the evaluation of differences in post modifications, such as disulfide bonding position (Gorman et al., 2002; Wu, 2008; Xiang et al., 2009), glycosylation patterns (Hägglund et al., 2007; Medzihradszky, 2008), and phosphorylation site determination (Stensballe et al., 2001a; Mann and Jensen, 2003; Palumbo et al., 2010). Each of the different modifications affects the properties of proteins including stability, solubility, function, activity, among others. The post translational modifications are not governed by the static genome, and must therefore be analyzed on a protein level (Mann and Jensen, 2003).

Sample preparation is a crucial aspect prior to MS, and may constitute the largest part of the work of the MS strategy. Traditional strategies includes either a gel-fre strategy known as multidimensional protein identification technology (MudPIT), or a gel-based strategy, shown in Figure B.1, employing either 1D SDS-PAGE, or 2D SDS-PAGE a prefractionation steps (Steen and Mann, 2004; O'Farrell, 1975). In addition, when combined with protein dyes, such as coomassie, or silver-staining, gel-based techniques allows for a direct visualization of proteins of interest on the gel (Bradford, 1976; Krystal, 1987). The 2D strategy provides the highest resolution - $\approx$ 5000 proteins down to less than 1 ng protein per spot - and robust methods have been invented for most types of proteins (Görg et al., 2004). Though the 2D strategy provides a good resolution it is not convenient for rapid large-scale identification, and it often fails to detect very hydrophobic proteins, or proteins with unusual pI values (). The experimental work is often laborious, and cutting out bands from the gel can be very tedious when a typical proteome consists of an average of 2000 proteins. More recently method, such as filter-aided sample preparation (FASP) have been suggested in order to combine the advantages of in-gel, and in-solution digestion on filters, which provides a more rapid approach to obtain large-scale proteomes in single-run analyses (Wisniewski et al., 2009).

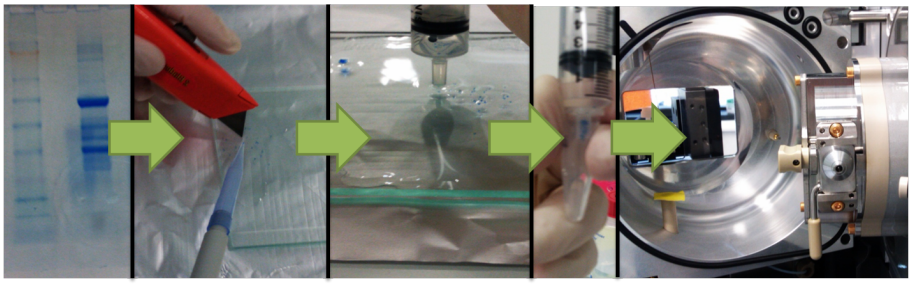


Figure B.1: A common gel-based proteomic apporach. Samples are run on a SDS-PAGE gel, and subsequently excised from the gel. The gel pieces are then transferred into tubes, and exposed to downstream processing, where the resulting samples ultimately are injected into the mass spectrometer.

Most mass spectrometers are able to analyze both intact, and digested proteins. The former strategy is referred to as top-down, and provides information of *in vivo* post translational modifications, charge states, and protein-protein interactions (Siuti and Kelleher, 2007). However, the top-down approach only provides a birds eye view, which can provide interesting results, but is uncommon in large-scale studies. For example protein-protein interactions, and protein comformation can be investigated by examining deuterium exchange in a an emergin field referred to as hydrogen/deuterium exchange MS, or HX-MS (Engen, 2009). Several studies, not only large-scales, applies the bottom-up approach. This method is commonly referred to as the shotgun proteomics approach, and it is favored for proteomics on a large-scale, because it provides rapid identification of proteins, and biological function (Steen and Mann, 2004; Marcotte, 2007). The general shotgun strategy involves three stages, shown in Figure B.2: (1) sample preparation, and digestion, (2) LC-MS/MS, and (3) database searching, and bionformatics. The sample preparation is succeeded by digestion with proteases resulting in a peptide mixture (Link et al., 1999; Peng and Gygi, 2001). The peptides are the actual molecules, which are investigated, and the sequence of each peptide are sought to be traced back to its corresponding protein.

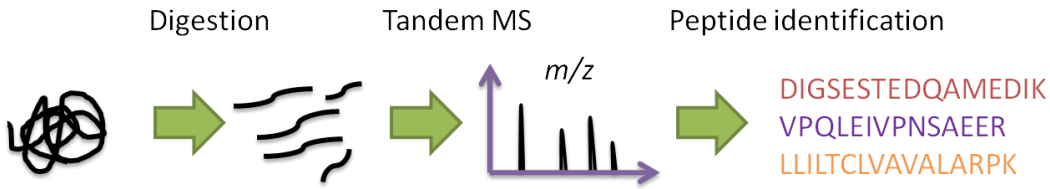


Figure B.2: Shotgun proteomics approach: protein samples are digested with proteases, resulting in a peptide mixture. The peptides are analyzed by tandem MS, and peptide sequences are identified. Though the workload is indicated on the figure each step may prove to be longer or shorter depending on the sample at hand.

The digestion of sample introduces increased complexity even if the sample contains only one protein. Larger proteins generally produces more peptides, and the mass spectrometer sequence one peptide at the time. The increased complexity is managed by introducing a separation column prior to injecting the sample into the mass spectrometer. Thus the interphase of liquid chromatography prior to MS analysis is applied in most instrument used today (Whitehouse et al., 1985).

The peptides are transferred from solution into the gas-phase, a process referred to as the ionization process, e.g electrospray ionization (ESI) (Fenn et al., 1989), and subsequently into the mass spectrometer. Often trypsin digested peptides are preferred, because of at least two inherent properties of these peptides, which makes them ideal for mass spectrometry. Firstly, the configuration with a N-terminal amino group, and a lysine/arginine C-terminal produces doubly charged ions due to gas-phase basicity, and the excess ions in the electrospray process (Dole et al., 1968) (Kebarle and Verkerk, 2009). Secondly, due tryptic peptides have two charges, they are measured at $m/z = m/2$, which makes it possible to measure as large as $8000\,kDA$ peptides using quadrupoles. Furthermore, the peptides will experience a factor two difference in applied voltage as described by the equation:

$$v = z^2 \sqrt{\frac{V_{acc} \cdot 2}{m}} \tag{B.1}$$

where v is velocity, z is charge, V_{acc} is applied voltage, and m is mass.

The peptide weight can then be analyzed through MS, and the peptides can be sequenced using tandem MS, or MS^n, where several mass analyzers are applied. In the latter technique, peptides are

fragmented into product ions by applying internal energy to the precursor peptide, which provides evidence of the structure of the precursor ion (Sleno and Volmer, 2004). The fragmentation is induced by increasing the internal energy of the peptides by e.g. collision induced dissociation (CID) with inert gas molecules (Hunt et al., 1986) (Shukla and Futrell, 2000). The schematic setup of a mass spectrometer involves an ion source, a collision cell, and one or more mass analyzers, as shown in Figure B.3. The ion source is used to transfer ions into the first mass analyzer, which is often used to select specific ions for collision induced dissociation in the collision cell, and fragmented ions are subsequently focused in the second mass analyzer, and detected according to their mass-over-charge ratio m/z in the detector (Aebersold and Mann, 2003). The m/z is then used to create a spectrum, where the x-axis is m/z, and the y-axis is ion counts; often the y-axis is scaled by setting the most intense ion to 100%. The product ion spectrum can then be used to assemble the precursor peptide, referred to as peptide sequencing (Link et al., 1999). The sequenced peptide can then be searched with bioinformatics algorithms, which are able to trace the peptide back to the protein, and subsequently the gene it originated from. In addition, this also provides information about which organism the protein originated from.

Figure B.3: The general setup of a tandem mass spectrometer. The ion source is used to transfer ions into the first mass analyzer, then into a collision cell. The resulting fragments are then transferred into the second mass analyzers, detected, and reported according to mass- over- charge, m/z.

Tandem MS produces a vast number of spectra, and is therefore tedious to interpret manually. Therefore, database search algorithms are applied to analyze MS data. The computational analysis compare the acquired MS data with theoretical spectra, as shown in Figure B.4. Several search algorithms exists, and each algorithm utilize different set of strategies, e.g. spectral matching, theoretical weight matching. Several excellent search engines exists including SEQUEST Eng. et al. (1994), MASCOT (Perkins et al., 1999), X-Tandem (Craig and Beavis, 2004), and MaxQuant (Cox and Mann, 2008). The search engine concept is to use genomic sequences, translate genes into proteins, fragment the proteins *in silico* into peptides, and create predicted MS/MS spectra from the theoretical peptides. The acquired data is then searched to fit the theoretical spectra, and positive hits are then backtraced to peptides, protein, and ultimately genes. A general observation is that the different search engines produces slightly different results, and validation of search results are often performed by using multiple search engines (Deutsch et al., 2010). The search engine strategy introduces potential false-positives, due to various reasons, which are further discussed below in the validation section.

B.1 Quantitative Mass Spectrometry

Protein identification is an important part of proteomics, but the mass analyzers are able to provide even more information. The turn toward quantiative analysis of proteins is also very important, because it enables comparison between samples, and makes it possible to infer even more biological information (Ong and Mann, 2005). Several challenges exists in this subfield of proteomics, including non-uniform ionization, e.g. in electrospray (Rappsilber et al., 2002), where it is commonly observed that hydrophobic peptides ionize more readily than more polar peptides, e.g. phosphorylated peptides (Zhao et al., 2008). However, with the rapidly evolving fast scanning

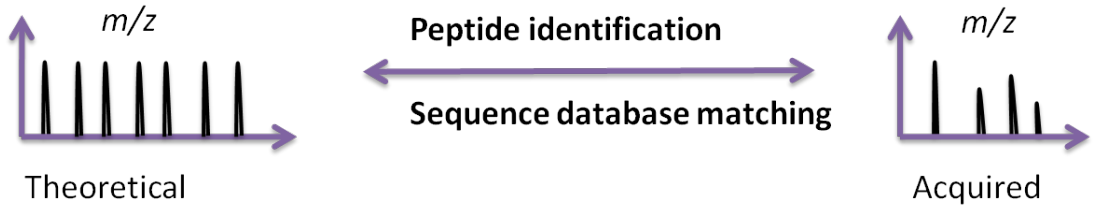


Figure B.4: An example of spectral matching. A genomic database is used to create theoretical peptides, which are then fragmented *in silico* into theoretical MS/MS specta, which can be compared to acquired MS/MS spectra.

high sensitivity mass spectrometers (Schwartz et al., 2002) (Hu et al., 2005) purely qualitative proteomes does not provide enough information to infer biological meaning (Ong and Mann, 2005). The abundance of a protein often provides an indication of the importance of the protein, and it is therefore important to know if a protein is abundant or close to the detection limit. As a consequence, the field of quantitative MS is of major concern and various strategies can be use to quantitate protein in complex mixtures. Two major strategies exists: label-free, and labelled techniques. Previoulsy the labelled methods were mostly applied, because the provided the most reliable and accurate data, but the field of label free techniques is continuously being improved to be comparable (Neilson et al., 2011). The quantitation of proteins using MS is often divided into absolute quantitation and relative quantitation. The former is the conversion of the MS output into a concentration or amount, and the latter is the determination of changes in protein abundance between two cell states (Ong and Mann, 2005).

B.1.1 Label-Free Techniques

The obvious advantages of label-free quantitative proteomics is that is does not require expensive isotopes labels, but the data analysis and validation can be challeging, and often replicates are demanded. The label-free strategies features two different methods: spectral counting (SC), and area under the curve (ACU), as shown in Figure B.5. The SC method is applied by counting the number of peptides assigned in a tandem MS analysis in data-dependent acquisition (DDA) mode. The DDA mode is used to select precursor peptides for sequencing. For example, in a QToF the quadrupole is set to survey scan mode to analyze the incoming sample. If it detects a high intensity ion it switches to MS/MS mode, thus fragmenting the incoming. This happens in live-time, and therefore a normal DDA method is set to fragment the most abundant ions. Thus, if a peptide is abundant it is more likely to be selected for fragmentation and will therefore produce a higher abundance of MS/MS spectra, which has been observed to be proportional to the amount of corresponding protein (Liu et al., 2004). To estimate the amount of protein a concept known as protein abundance index (PAI), was introduced by Rappsilber et al. (2002), to incorporate observable parameters (Ong and Mann, 2005; Neilson et al., 2011). PAI was defined as the number of peptides identified divided by the number of theoretically observable tryptic peptides, thus effectively normalizing the data. Subsequently it was found that the number of observed peptides and the protein abundance was logarithmic, which lead to the exponentially modified PAI, i.e. emPAI$= 10^{PAI} - 1$ (Ishihama et al., 2005). Though the emPAI value is only estimate of protein abundance ($\approx r = 0.89$) it is observed to be as accurate as protein staining or better, which means that the gels can be conveniently bypassed. However, the emPAI quantitation method has some shortcomings, and though it is beyond the scope of this introduction to provide full details, and

progress of SC, a few points are highlighted. Firstly, in the method of absolute protein expression (APEX), the differences in ionization efficiency, as a consequence of different physiochemical properties, introduces biases and variability, that have been sought to be removed by introducing a correction factor, O_i (Mallick et al., 2007) (Lu et al., 2007) (Braisted et al., 2008). Secondly, it is recognized that larger proteins will generate more peptides, thus affecting the number of spectral counts (SpC), which has been sought to be incorporated by correlating the SpC's to the length of the protein, L. The normalised spectral abundance factor (NSAF), thus improves the relative abundances by dividing the sum of all SpC/L [1], which have been shown to include changes down to 1.4 fold in relative abundance (Florens et al., 2006). Though these methods seem to provide reliable quantitation the methods are still being critically investigated, and sought to be improved. In closing, spectral counting methods are being combined to provide more robust and enhanced results by combining unique peptide count, SpC, and fragment ion intensity. In other words, for a given peptide the number of unique peptides, the number of MS/MS spectra, and the intensities for each MS/MS spectra is merged to give a normalized spectral index, SI_N (Griffin et al., 2010). This method seems to provide more accurate quantitation of protein, which is consistent with conventional Western Blots, and seems to be a promising tool for quantitating differentially expressed proteins in replicate runs (Neilson et al., 2011).

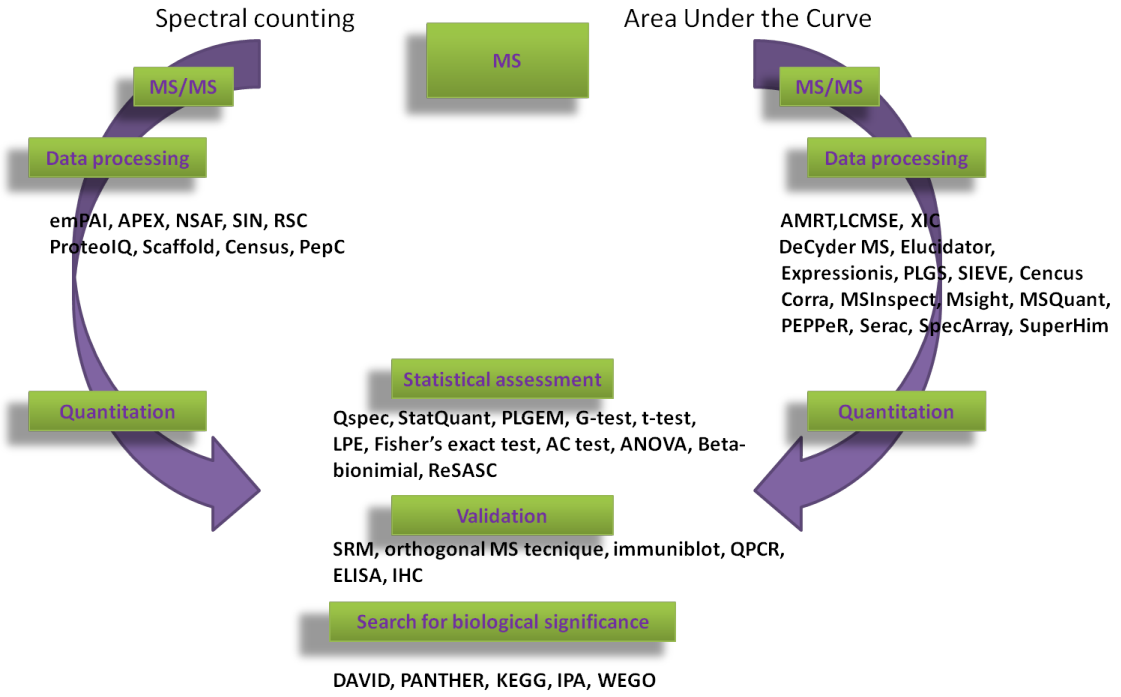


Figure B.5: The two strategies of label-free quantitative mass spectrometry: spectral counting, and area under the curve. The dataprocessing can be performed by using the numerous available software packages, though it is beyond the scope to discus each program. (Neilson et al., 2011)

Protein quantitation can alternatively be performed by integrating the peaks from the chromatogram from the liquid chromatography. The area under the curve (AUC) is an accurate method to measure the amount of protein providing good linearity for less complex samples ($R^2 = 0.9978$) in

[1] Strictly speaking this equation is spectral abundance factor (SAF), thus the true NSAF is $NSAF_k = \frac{SAF_k}{\sum_{i=1}^{N}(SpC/L)_i}$, where k refers to a specific protein

the range 10 $fmol$-10000 $pmol$ (Bondarenko et al., 2002), but also for more complex myoglobin samples with $R^2 = 0.991$ for the range 10 $fmol$-100 $pmol$ (Chelius and Bondarenko, 2002). The AUC is extracted from the MS-chromatogram, resulting in an extracted ion chromatogram (XIC), with time on the x-axis, and intensity (counts per second) on the y-axis (Ong and Mann, 2005). Though this process is straightforward the challenge occurs, when two samples are to be compared. Biological variation, e.g. differences in post translational modifications affects retention time, and intensity, and other factors such as differences in chemical background and interferences may further complicate comparison (Neilson et al., 2011). The data analytical task is then to clean, and normalize data (Listgarten and Emili, 2005). Retention times are normalized so that the values fit for identical precursor ions in separate runs (Wang et al., 2003). Though this method has successfully for identification of biomarkers, extensive reviewed by (Vandenbogaert et al., 2008), it is still vulnerable to co-eluting peptides, chromatographic resolution, and the switching between survey scan, and MS/MS, as shown in Figure B.6.

Figure B.6: Normal data dependant acquisition. The ESI is continuously spraying, and the instruments shuttles between survey scan, and MS/MS mode. When the survey scan is on, MS/MS if off, and vice versa.

The switching between survey scan and MS/MS compromises one or the other. When the mass analyzer is in survey scan, accurate chromatography is obtained, but protein identification is sacrificed. Conversely, when the mass analyzer is operating in MS/MS mode, protein identifications are increased, but the chromatography is switched off, and no other precursor ions are allowed. A way to counter this is to use data independent methods. As shown in Figure B.7, simply switching between low- and high-energy ($4\,V$ - 15-$130\,V$), thus transmitting everything simultaneously, it is possible to obtain very accurate quantitation (Silva et al., 2006).

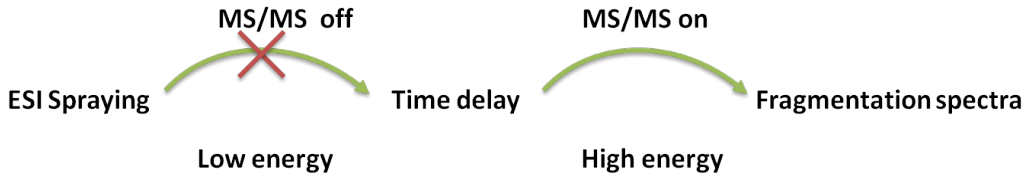


Figure B.7: MS^E is data-independent, because this mode is independent of the sample. The instrument shuttles between survey scan, and MS/MS following specified intervals, thus data is acquired without any selection process.

B.1.2 Labelled Techniques

The labelled methods can be roughly divided into four separate categories, as shown in Figure B.8. The ingenious part of the labelling techniques is that nonradioactive isotopes such as ^{13}C, ^{15}N, and ^{2}H can be incorporated into a peptide, which will share the same chemical properties as the same label-free peptide, but with a mass offset. The simple, but widely used, spiking in method can thus be applied by spiking in a known amount of labelled peptide, which cis used

to correlate absolute amount to the ion current output of the MS. The difference between the XIC of the labelled peptide and the label-free peptide can thus be correlated to the MS output, and provides an accurate quantitation (Kusmierz et al., 1990). Though this method is simple and effective, one drawback is that peptides needs to be synthesized for each protein to be quantitated. In addition, the synthetic peptide is spiked in late in the purification, and therefore does not take biological or methodological variations into consideration. Other more sophisticated strategies incorporate the stable isotopes by enzymatic reactions. As an example, trypsin digestion in the pressence of $H_2^{18}O$ result in substitution of carboxylic oxygen (Stewart et al., 2001). However, this method leads to exchange with one or both carboxylic oxygens, which compromises accurate quantitation.

Labelling techniques also includes chemical modification of proteins, and peptides, and are a widely used for several reasons. Proteins and peptides can be modified with uniquely identifiable tags, that can be targeted in affinity purifications, or be observed as unique fragments in MS/MS. For example, the isotope-coded affinity tag (ICAT) reacts with free cysteines, and consists of a linker region, and a biotin tag, that can be used for avidin chromatography (Gygi et al., 1999). Other chemical derivitisation methods are, permythlation to prevent unspecific binding in phospoprotein and peptide affinity chromatography (Ficarro et al., 2002), β-elimination (McLachlin and Chait, 2003), several amino terminal modifications, and the more recent but widely used isobaric tag for relative and absolute quantitation (iTRAQ) (Ross et al., 2004). In the latter method diagnostic tags are incorporated, which provides specific reporter ions that can be searched for in the MS/MS spectra. Furthermore, the analytical throughput was increased from binary (2-plex) samples to 4-plex, thus introducing a convenient method for multiplexing samples, and enabling the comparison of more than two simultaneous samples (Ross et al., 2004). A general requirement of labelling techiques is that the reaction must be specific, and the specific modifications must be applied to all available substrates. If the modification is only partially applied the modified proteins are divided into separate fraction, that compromises accurate quantitation, and reduces sensitivity (Ong and Mann, 2005). In addition, it is necessary for the labels to create a mass offset of 3-4 Da to avoid isotopic overlap, which results in uncertain quantiation (Ong and Mann, 2005). These two criteria is of the utmost importance in chemical labelling, but also the parameters that makes these methods prone to errors, loss of analyte, or complex spectra. As an example, some chemical modifications produce 1:1 diastereomers, that cannot be digested by highly specific proteases used in MS, thus halving the amount of analyte (McLachlin and Chait, 2003). If the protein is of low abundance the sensitivity of the MS might not be sufficient to detect this protein.

Metabolic incorporation of tracer compounds is a major class of labelling, that has been widely used for decades. Originally, these methods applied radioactive isotopes, that could be traced via. a geiger-tube effectively creating a ”handle” used to trace proteins. In this way the labelled fraction could be rapidly traced in every downstream step in a purification protocol, which have been referred to as pulse-chase experiment. More recently, the incorporation of stable isotopes have been utilized to label cell under different growth conditions, with light and heavy labels, respectively, to investigate relative difference, and to elucidate metabolic pathways (Doherty et al., 2009). The stable isotope labels are incorporated into metabolic active cells only in the presence of labelled growth media. About 5-6 cell doublings yields 99% labelling of the protein of interest, and the label-free proteins are continuously being diluted to insignificant amounts (Ong and Mann, 2005). Instead of labelling the entire proteome with expensive heavy isotopes a more targeted approach has been applied. The widely used stable isotope labelling by amino acids in cell culture (SILAC) is utilized by applying essential isotopically labelled amino acids, e.g. arginine, to the

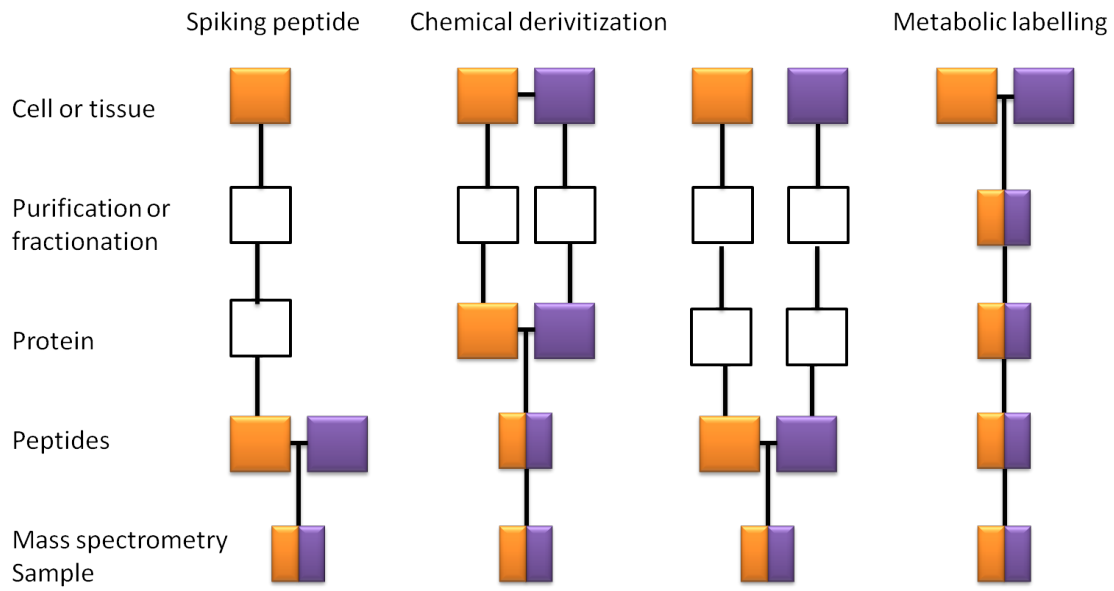


Figure B.8: Labelled techniques. The most simple is the spiking in method, which is often spiked in just before MS analysis. Another widely method is the chimical derivitisation methods, which include mixing samples in the beginning, or just before MS analysis. The last common method is metabolic labelling in which the samples are mixed at the beginning to ensure that the sample are treated equally. (Ong and Mann, 2005)

growth medium (Ong et al., 2002). In addition, it is possible to include three different labels; light, medium, and heavy. Though SILAC is expensive it is widely used, because of the simplicity of simply adding the labelled amino acid in the cell culture, and because it enables accurate measurements. By mixing cell cultures 1:1 early in the purification process, equal biases and purification variations are applied (except for unvoidable biological variations), which makes the relative quantitation more accurate. Several evolutions of the method have been applied including protein production variations in pulsed SILAC (Selbach et al., 2008), and protein turnover in dynamic SILAC (Doherty et al., 2009). In the former strategy a label is transiently added in hight amount, which enables the investigation of portein production. In the latter, the reverse strategy was used. Completely labelled cells were fed with non-labelled substrat, and the loss of labelled protein was measured relatively to newly produced proteins, thus inferring protein turnover rates for the entire organism.

B.2 Validation of High-throughput Proteomics

In recent years the proteomics community have recognized the necessity to improve the validity and the quality of MS data. This is being done in an effort to create better hypotheses, and to infer biological meaning. The application of search engines is based on the best match of MS/MS data to theoretical spectra. The validation process is then assess weather this best match is assigned correctly or not (Nesvizhskii and Aebersold, 2004). It has been suggested that the vast amount of data produced by the high throughput MS instruments, and the growing databases containing identification, localization, and post translational modifications, need to be scrutinized (White, 2011). The reasoning is that the poor quality databases, e.g. containing wrongly assigned site determination of modifications, could potentially lead to the pursuit of dead-end experiments, and

years of wasted research. The contamination of false-positives is more severe, than the exclusion of false negatives, and the filtering to reach a tolerable equilibrium is important in post processing of MS data. One method is to apply a target-decoy filtering to provide an estimate of false positives (Elias and Gygi, 2007), which can be used for further validation.

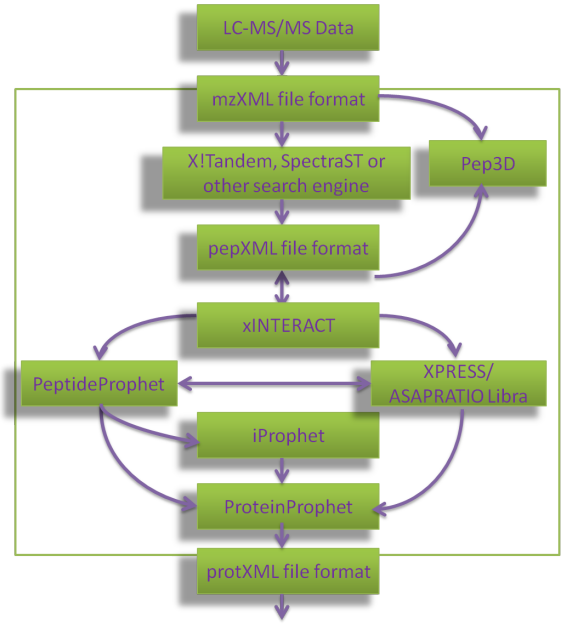


Figure B.9: Trans-Proteomic Pipeline (TTP): after LC-MS/MS data are submitted to the TTP dataprocessing, as shown in the flowdiagram. The different processing steps are described in the text. (Deutsch et al., 2010)

The Trans-Proteomic Pipeline is a compiled software package introduced in 2005 by Keller et al. (2005), which provides post data processing to improve information obtained from MS, shown in Figure B.9. In short, the pipeline is convenient because it converts proprietary LC-MS/MS Data from any instruments into the open-source formats, e.g. mzXML (Pedrioli et al., 2004). The resulting file can then be examined visually using Pep3D (Li et al., 2004) to assess the quality of the chromatogram, or searched using the common search algorithms. The resulting search files are then post-processed to improve the identification quality, and to quantitate the samples. Peptide identification is validated through PeptideProphet (Keller et al., 2002), which provides a model for false-positives, and false-negatives. As the name suggests this is initially performed on the peptide level, and this level is further validated using the recent iProphet[2]. The last validation step is performed on the protein level using ProteinProphet (Nesvizhskii and Aebersold, 2004), and results in a protein identification list summing up probabilities of proteins in the sample (Pedrioli et al., 2004). In summary the pipeline can be used for complex samples where the transition from MS/MS data, to peptides, and further to proteins can be very difficult. The pipeline enables filtering poor quality data, combines replicate MS-runs from the same or different instruments, assesses of peptide assignment, and infers the data into protein lists.

[2]In preparation by Shteynberg et. al

B.3 Metaproteomics

The feasibility of a large-scale proteomics study was shown already in 2001 on yeast (Washburn et al., 2001), and has since been applied to various other samples and organisms, including the EBPR system (Wilmes and Bond, 2004). With the parallel advances in next generation DNA sequencing the number of possible protein identifications expands rapidly each year (Mardis, 2008). The interconnected field of metagenomics provides valuable insights about the repertoire of genes, but if used alone, represents merely the blueprint of the potential of the microbial community. By connecting metaproteomics with metaproteomics, referred to as metaproteogenomics, it is possible to obtain high resolution information about the make-up and abundance of microorganisms, the functional potential, protein expressions, and lastly biochemical activity (Wilmes and Bond, 2006b) (VerBerkmoes et al., 2009) (Wilmes and Bond, 2009).

The application of proteogenomics was shown for acid mine drainage (AMD), which became the first mixed community to be analyzed by community genomics, and community proteomics (Tyson et al., 2004; Ram et al., 2005). It is now being used as a model system, because the community consists of only a handful of taxa, and because the different strains are genomically distinct, which makes the data processing more manageable (Tyson et al., 2004). It was possible to detect 2033 proteins from the five most abundant organisms in the growing acid mine drainage microbial biofilm, thus inferring gene expression, and also making it possible to map out active metabolic pathways *in situ* (Ram et al., 2005). In order to avoid randomly identified proteins, several studies have applied the restriction that at least to peptides must be mapped to one protein to trigger a positive protein hit. In this way 49% of all the predicted proteins from the genomic data were identified using metaproteomics (Wilmes and Bond, 2006b), while others have applied a three peptides per protein restriction (Park et al., 2008). In addition to predicted identified proteins Ram et al. (2005) were also able to identify proteins from putative operons, thus confirming putative proteins. Thus, the detection of novel proteins is possible from metaproteomic data, which extends biological information. Several interesting observations have been made using the AMD model system. For examples, when an microorganism constitutes more than 30-40% of the entire poulation the number of protein identifications is saturated. When the abundance of the microorganism decreases, so does the number of proteins identified as well as the protein coverage. However, it was stil shown that more than 100 proteins can be identified from an microorganism constituring around 1%.

The transition from proteomic analysis of single strains to several strains is not an easy task, and introduces several analytical challenges. Applying proteomics to more complex systems such as enhanced biological phosphorous removal (EBPR) systems have been performed for lab-scale reactors (Wilmes and Bond, 2006b), but several issues still exist. The community in EBPR is not static over space and time, and each strain will change in abundance or potentially adapt to new niches in the system. This introduces a problem, because proteomics is based on the comparison of theoretical peptides obtained from genomic sequencing, with MS/MS data, which requires exact matches. Though the genome is more static compared to the proteome even single amino acid substitutions will lead to a mass difference that will prevent peptide identification (Peng and Gygi, 2001) (VerBerkmoes et al., 2009), as shown in Figure B.10. Thus when a metagenome is constructed it only provides a snapshot of the genomically makeup at the time of sampling, and subsequent proteomic studies needs to be performed within the same timescale, to avoid the genome and proteome to drift apart. Logically, the affect of this issue is less on large, and abundant proteins, but becomes more severe for more diverged, less abundant, and smaller proteins. Though this constitutes an identification bias, it has also been used to obtain strain-variation information.

Because MS is sensitive to amino-acid substitutions it is possible to perform strain-typing peptide by peptide, which makes it possible to analyze functional differences between closely related organism. This analysis is referred to as proteomics inferred genome typing (PIGT) (Denef et al., 2008), and was orignally applied to the AMD system, but has also been applied for EBPR system (Lo et al., 2007).

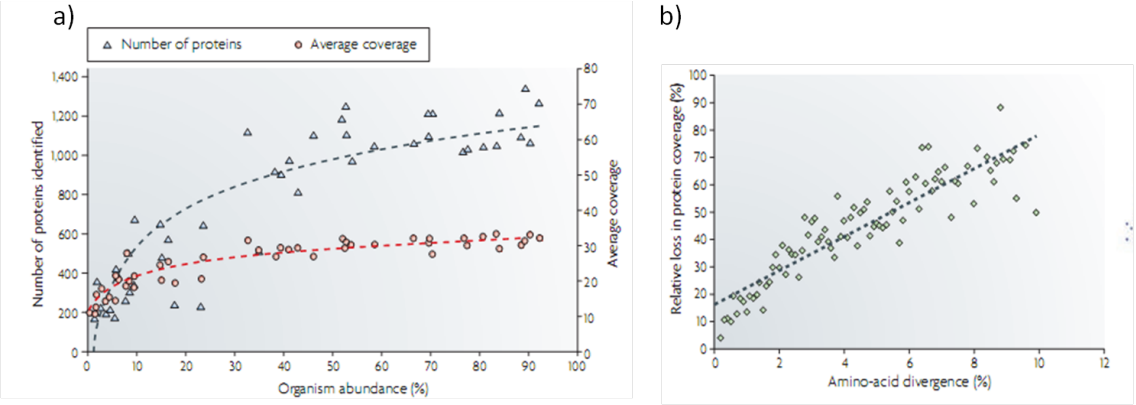


Figure B.10: Microorgnisms constantly evolves to adapt to the environment. A) shows that the species abundance has a direct affect on the number of proteins identified. When the community becomes increasingly abundant, the number of proteins increases dramatically. B) As the microorganisms is exposed to mutations the amino acid sequences drift away from previously made genomes. The figure shows that as the amino acids diverges, the protein coverage is decreases dramatically. Reprinted from (VerBerkmoes et al., 2009), obtained from (Denef et al., 2007, 2008)

Another issue in metaproteomics is the complexity of the system. The complexity of different environments is considered to follow the trend $soil > ocean > wastewater > AMD$ (Curtis et al., 2002). This shows that the EBPR system is not the most complex system to analyze, yet it is not the most simple either. However, the dynamic range in this system is still enormous. Each unique taxa in the systems produces unique proteins, and each protein is differentially expressed, which makes the dynamic range between low abundant proteins, and high abundant proteins huge. In addition, the intrinsic properties of electrospray, and the low coverage of the chromatography further complicates comprehensive protein identification (Michalski et al., 2011). However, these issues are continuously being reduced with improved liquid chromatography capabilities, and rapid scanning mass spectrometers, but also the post processing of data still improves with new observations and studies. Some of the issues of polymorphism, and lacking genomic data can be partially circumvented by *de novo* sequencing. This method extract peptide sequences directly from MS/MS, thus circumventing databased searches. For environmental samples where the selective pressure is high, and micro-diversity is abundant, it may be feasible to first apply database search, and then apply *de novo* sequencing for the remaining high quality spectra to increase protein identification (Nesvizhskii, 2007) (Bern et al., 2007).

C Measured Protein Quantities from Selected Stuides

The quantitation of protein and the other extracellular polymeric substances (EPS) have been performed previously, and have provided valuable information exploited in this report. In the following selected study is highlighted, because these studies were the reference work to compare and analyze the data obtained in this research.

As already mentioned, proteins in wastewater only constitutes a part of the total EPS. In addition, various ions, carbohydrates, humic acids, extracellular DNA, and other compounds are present, and these have previously been quantified in several studies for several different wastewater treatment plants (WWTP's) (Raunkjær et al., 1994) (Frølund et al., 1996). Currently it seems that the most used methods are the modified Lowry method (Frølund et al., 1995, 1996) for the determination of proteins, and humic acids, the anthrone method for the determination of carbohydrate (Gaudy, 1962), and the m-hydroxydiphenyl method for determination of uronic acids (Kintner III and van Buren, 1982). These method are widely used and have been utilized to quantitate the concentration of these compounds in wastewater (Frølund et al., 1996) (Yu et al., 2007) (Park and Novak, 2007) (Park et al., 2008).

Table C provides the results obtained by Frølund et al. (1996) from activated sludge obtained from two different WWTP's. It is seen that the amount of protein, humic compouns, carbohydrates are substantial, while uronic acids constitutes a relative smaller part. An interesting parameter is the protein-polysaccharide ratio. It is seen that the amount protein are approximately three times larger than polysaccharide, and that the amount of protein is bigger than humic compounds and carbohydrates combined, thus constituting the largest fraction of EPS.

Sludge from	Protein (mg/g VS)	Humic compounds (mg/g VS)	Carbohydrate (mg/g VS)	Uronic acids (mg/g VS)	Protein/Polysaccharide
Asaa	523 (+/- 8.0)	230 (+/- 20.0)	181 (+/-11.0)	21 (+/- 3.0)	2,9
Aalborg East	462 (+/- 25.0)	176 (+/- 26.0)	179 (+/- 13.0)	20 (+/- 4.0)	2,6

Table C.1: Results from CER by (Frølund et al., 1996)

While Frølund et al. (1996) utilized only the one extraction method per compound Park and Novak (2007) examined the three different extraction methods, shown in Table C, as described in Section A.1. Interestingly, the CER seems to provide the highest extraction yield for both proteins, and polysaccharides which merits the use of the CER method. However, the authors did show that the different methods provides distinct proteins in which case it would be feasible to use all of them to increase the proteome coverage.

Another study sought to inquire for the protein activities using several different purification strategies, which is summed up in Table C.3. From these results it seems that the method using formaldehyde and sodium chloride provides the highest protein yield, while the method employing EDTA is the second highest. Comparing these two with the CER methods, EDTA provides double, while formaldehyde and sodium chloride provides quadruple amounts of protein. Unfortunately the specific methods are only vaguely described, but highlights that other methods than CER can be utilized to provide high amounts of protein. The centrifugation method is unclear,

Sludge from	Protein (mg/g VS)			Polysaccharide (mg/gVS)			Protein/Polysaccharide		
	CER	Base	Sulfide	CER	Base	Sulfide	CER	Base	Sulfide
A	49.1	40.7	11.6	23.2	15.9	2.7	2.1	2.6	4.2
B1	51.3	-	18.9	20.8	-	5.0	2.5	-	3.8
B2	64.5	40.4	13.6	25.4	9.6	2.8	2.5	4.2	4.9
B3*	73.4	47.4	15.5	26.1	11.4	3.7	2.8	4.2	4.2
B4*	41.0	42.4	12.4	21.3	16.6	2.9	1.9	2.6	4.4
C1	36.7	31.1	16.0	13.0	11.4	4.6	2.8	2.7	3.5
D1*	41.0	49.7	34.0	21.3	15.3	9.7	1.9	3.2	4.5
D2*	58.7	42.9	37.1	29.2	17.9	12.8	2.0	2.4	2.9

Table C.2: Results obtained by (Park and Novak, 2007). In the plants marked with "*" base extraction was performed in the presence of N_2.

but provides the highest protein-polysaccharide ratio. Furthermore, the CER provides the second largest, while EDTA provides a considerable high ratio. It therefore seems that the CER, though not providing the highest protein yield, results in a relatively higher purity of protein.

Components	Extraction method	Protein (mg/g VSS)	Humic acids (mg/g VSS)	Polysaccharides (mg/g VSS)	Protein/Polysaccharide
Low bound EPS	Centrifugation	47.0	0.11	5.30	8.9
	Control	50.2	0.11	9.41	5.3
	Ultrasonication	42.7	0.87	16.2	2.6
	EDTA	63.0	18.1	12.2	5.2
Tight bound EPS	Formaldehyde	11.3	0.72	5.65	2.0
	Formaldehyde + ultrasonication	52.7	1.62	8.43	6.3
	Formaldehyde + NaOH	114.0	13.0	29.8	3.9
	CER	34.6	0.24	4.61	7.5

Table C.3: Results from (Yu et al., 2007)

D Quadrupole and Quadrupole Ion Trap Theory

In effort to obtain more specialized knowledge about the physics, physiochemical properties of mass spectrometry, this chapter was written as a succession of previous written chapter of ”electrospray ionization”, ”ion activation in mass spectrometry”, and ”collision induced dissociation”. The chapter is mainly focused on physical equations of quadrupoles, and ion traps, and how they are manipulated to work as ion filters.

The quadrupole ion mass analyzer is often used in sequence with other quadropoles or other mass anlayzers, because it is able to operate as a filter. The quadrupole makes it possible to select for specific ions or a window of ions, as shown in Figure D.1. Alternatively the quadrupole can be used to trap ions, which is referred to as an ion trap.

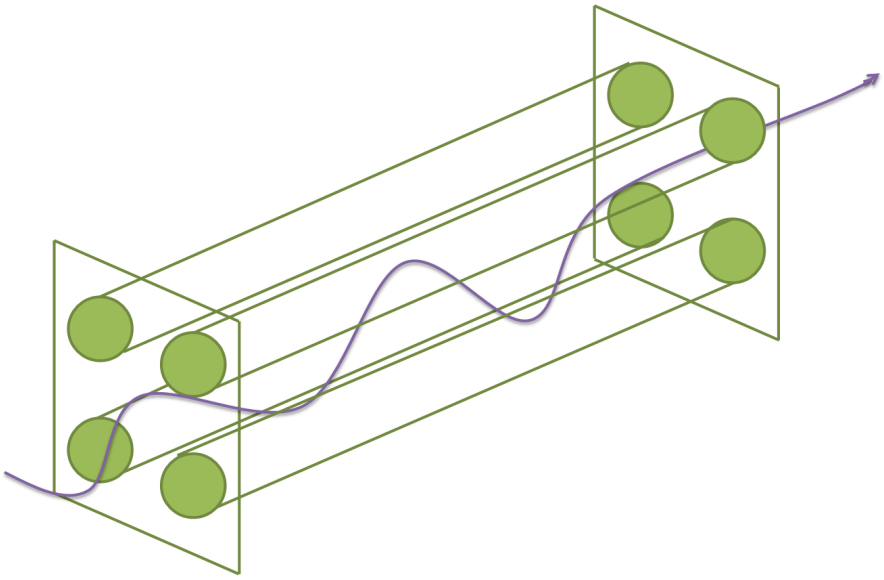


Figure D.1: Single quadropole

When the ions enters the quadrople they are subjected to an electric field posed by the four rods of the quadropole, termed quadrupolar alternative field. In positive mode positive ions enters the quadrupolar field and is attracted to the negatively charged rods. The quadrupole mass analyzer is operated so that the rods changes charges before the ions hits the rods, which makes them change direction. The fields are superimposed on a constant electric field, which creates opposing forces that acts on the ions (de Hoffmann and Stroobant, 2007, p. 91). The quadrupolar alternative field is a radio frequency (RF) field, and the combination of the DC-field, and the RF-field is responsible for the selection of ions. The DC-field is used to select for specific ions or a window of ions, and the RF-field is used to transfer the ions through the quadrupole. If only the latter is active all ions would be transmitted, with no selection. The ions are thus subjected to the forces (de Hoffmann and Stroobant, 2007, p. 91):

$$\phi_0 = +\left(U - V \cdot cos(\theta \cdot t)\right) \quad \text{and} \quad -\phi_0 = -\left(U - V \cdot cos(\theta \cdot t)\right) \tag{D.1}$$

where ϕ_0 is the potential applied to the rods, θ is the angular frequency, U is the direct potential and V is applitude of the RF voltage. The angular momentum is expressed as $\theta = 2 \cdot \pi \cdot v$, where v is the frequency, $v = \frac{1}{f}$ of the RF field given in s^{-1}.

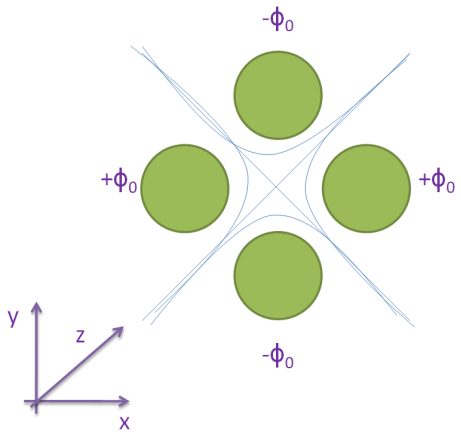


Figure D.2: The field in the cross section of the quadrupole

The ions is send along the z-axis shown in Figure D.2 and is affected by the electric fields from Equation D.1. The velocity of the ions is unchanged, but they will move in x and y plane according to mass to charge ratio (de Hoffmann and Stroobant, 2007, p. 91). The force on ion is given as mass multiplied by acceleration, $F = m \cdot a$, and additionally it affected by an electric field. This can be summed up in the formula:

$$F_x = m \cdot \frac{d^2x}{dt^2} = -z \cdot e \cdot \frac{\partial \phi}{\partial x} \tag{D.2}$$

$$F_y = m \cdot \frac{d^2y}{dt^2} = -z \cdot e \cdot \frac{\partial \phi}{\partial y} \tag{D.3}$$

where ϕ is a function of ϕ_0 given as (de Hoffmann and Stroobant, 2007, p. 91):

$$\phi_{(x,y)} = \frac{\phi_0 \cdot (x^2 - y^2)}{r_0^2} = \frac{(x^2 - y^2) \cdot (U - V \cdot cos(\theta \cdot t))}{r_0^2} \tag{D.4}$$

The quadropole is able to select for specific ions by stabilizing the trajectory of the specific ions in both the x and y direction. By differentiating Equation D.2, Equation D.3 with respect to time it is possible to get an expression that determine, which ions are able to obtain stable trajectories given by (de Hoffmann and Stroobant, 2007, p. 92):

$$\frac{d^2x}{dt^2} + \frac{2 \cdot z \cdot e}{m \cdot r_0^2} \cdot (U - V \cdot cos(\theta \cdot t)) \cdot x = 0 \tag{D.5}$$

$$\frac{d^2y}{dt^2} + \frac{2 \cdot z \cdot e}{m \cdot r_0^2} \cdot (U - V \cdot cos(\theta \cdot t)) \cdot y = 0 \tag{D.6}$$

The Equations D.5, and D.6 thus describes that stable trajectories if the ions never hit the rods, i.e. if x and y never reaches r_0. If the ions are unstable they will fluctuate increasingly until they hit one rod, discharge, and thus is rendered uncharged.

To determine the position of the particle in the x, and y direction it is necessary to integrate Equation D.5, and Equation D.6, respectively. Prior to the integration it is convenient to better define the motion of ions under the influence of an electric field. This was done by (Mathieu, 1868), who found that the motion could be described as the solution to the second-order linear differential equation given as:

$$\frac{d^2u}{d\varepsilon^2} + (a_u - 2 \cdot q_u \cdot cos(2\varepsilon)) \cdot u = 0 \tag{D.7}$$

where u is represents the coordinate axis x,y, or z. ε is introduced in the Mathieu equation as a dimensionless parameter, and is not present in the Paul equation. It is defined as:

$$\begin{aligned} \varepsilon &= \frac{\omega \cdot t}{2} \\ &\Updownarrow \\ \varepsilon^2 &= \frac{\omega^2 \cdot t^2}{4} \end{aligned} \tag{D.8}$$

where ω is frequency, and t is time. ε can therefore be introduced into the Paul equation by replacing t^2 with ε^2, while compensating with the multiplication of $4/\omega^2$. The introduction of this variable will later be shown to be important, because it wil reappear as the radial frequency, s^{-1}, of the RF potential applied to the ring electrodes. The change of variables is convenient, because $2\varepsilon = \omega t$.

By substituting $\varepsilon = \omega \cdot t/2$ it can be shown that (March, 1997):

$$\frac{d^2u}{dt^2} = \frac{\omega^2}{4} \cdot \frac{d^2u}{d\varepsilon^2} \tag{D.9}$$

and the mathieu equation can be substituted into this equation, while multiplying with m on both sides:

$$m \cdot \frac{d^2u}{dt^2} = \frac{-m \cdot \omega^2}{4} \cdot (a_u \cdot 2q_u \cdot cos(\omega \cdot t)) \cdot u \tag{D.10}$$

The left term inn Equation D.10 can be recognised as $F = m \cdot a$ in 3D space.

In addition to ε, the dimensionless parameters a_u, and q_u are introduced in the Mathieu equation. These are referred to as the trapping parameters, and successful tuning of these parameters allows ion selection (March, 1997). The trapping parameters are calculating by Equation D.7, and Equation D.5/D.6. In the following we will calculate the trapping parameters for the x-dimensions, i.e. a_x, and q_x. The calculations are tedious, but will be processed in steps. Firstly, Equation D.7, and Equation D.5 are set to equal:

$$-\frac{m \cdot \Omega^2}{4} \cdot a_x - \frac{m \cdot \Omega^2}{4} \cdot 2 \cdot q_x \cdot cos(\Omega \cdot t) = -\frac{2 \cdot e}{r_0^2} \cdot U + \frac{2 \cdot e}{r_0^2} \cdot cos(\Omega \cdot t) \cdot V \tag{D.11}$$

Thus we now have the complete picture, and it is possible to isolate a_x, and q_x. First, we isolate q_x and find that:

$$\frac{m \cdot \Omega^2}{4} \cdot (a_x - 2q_x \cdot cos(\Omega \cdot t)) \cdot \vec{x} = \frac{2 \cdot e}{r_0^2} \cdot (U + V \cdot cos(\Omega \cdot t) \cdot \vec{x} \tag{D.12}$$

$$\Updownarrow$$

$$2q_x = -\frac{8 \cdot e}{m \cdot \Omega^2 \cdot r_0^2} \cdot (U + V \cdot cos(\Omega \cdot t) + a_x \tag{D.13}$$

$$\Updownarrow \left(Y = \frac{8 \cdot e}{m \cdot \Omega^2 \cdot r_0^2} \right)$$

$$2q_x = -Y \cdot U + Y \cdot V \cdot cos(\Omega \cdot t) + a_x \tag{D.14}$$

$$\Updownarrow$$

$$q_x = \frac{\frac{Y}{2} \cdot U}{cos(\Omega \cdot t)} - \frac{Y}{2} \cdot V + \frac{a_x}{2} \cdot \frac{1}{cos(\Omega \cdot t)} \tag{D.15}$$

$$\Updownarrow$$

$$q_x = \frac{a_x}{2 \cdot cos(\Omega \cdot t)} - \frac{Y \cdot U}{2 \cdot cos(\omega \cdot t)} - \frac{Y}{2} \cdot V \tag{D.16}$$

The two first terms in Equation D.16 is reducible, thus the first two terms can be considered as a variable part, while the thrid terms is a constant part. For now, the interesting term in Equation D.16 is $-\frac{Y}{2}V = -\frac{4 \cdot e \cdot V}{m \cdot r_o^2 \cdot \Omega^2}$. The term for a_x can be isolated in a similar way, which generates the constant term $\frac{8 \cdot e \cdot U}{m \cdot \Omega^2 \cdot r_0^2}$. This is very convenient, because the latter term can now be substituted for a_x in Equation D.16:

$$q_x = \frac{8 \cdot e \cdot U}{m \cdot r_0^2 \cdot \Omega^2 \cdot 2cos(\Omega \cdot t)} - \frac{8 \cdot e \cdot U}{m \cdot r_0^2 \cdot \Omega^2 \cdot 2 \cdot cos(\Omega \cdot t)} - \frac{4 \cdot e \cdot V}{m \cdot r_0^2 \cdot \Omega^2} \tag{D.17}$$

$$\Updownarrow$$

$$q_x = -\frac{4 \cdot e \cdot V}{m \cdot r_0^2 \cdot \Omega^2} \tag{D.18}$$

Thus we find that the term for q_x, and the term for a_x:

$$a_x = \frac{8 \cdot e \cdot U}{m \cdot \Omega^2 \cdot r_0^2} \tag{D.19}$$

$$q_x = -\frac{4 \cdot e \cdot V}{m \cdot r_0^2 \cdot \Omega^2} \tag{D.20}$$

The trapping parameters can be exntended to the y-direction simply be applying a negative sign to the x-direction parameters. In a quadrupole r_0 is constant, $\omega = 2 \cdot \pi \cdot v$ is maintained constant, while U, and VU are variables. If an ion of any mass enters the quadrupole the coordinates x, and y can be determined through a time span as a function of U and V. The ion trajectory will either be stable or unstable. In the latter case the ion will fluctuate increasingly and x or y will eventually become equal to r_0. In practice, this means that the ion hits the rods, discharges, and is lost. A stable ion trajectory can be reached by the right combination of UV, and V, as described in the two following equations.

$$a_u = \frac{8 \cdot z \cdot e \cdot U}{m \cdot \omega^2 \cdot r_0^2}$$

$$\Updownarrow$$

$$a_u \cdot m \cdot \omega^2 \cdot r_0^2 = 8 \cdot z \cdot e \cdot U \tag{D.21}$$

$$\Updownarrow$$

$$U = \frac{m}{z} u \cdot \frac{\omega^2 \cdot r_0^2}{8 \cdot e} \cdot a_u \tag{D.22}$$

Similarly V can be isolated to give:

$$V = \frac{m}{z} \cdot \frac{\omega^2 \cdot r_0^2}{4 \cdot e} \cdot q_u \tag{D.23}$$

These variables can be predicted by creating a stability diagram, represented by a a_u, and q_u diagram, which is the solution to the Mathieu Equation. Thus, the solution to the Mathieu Equation

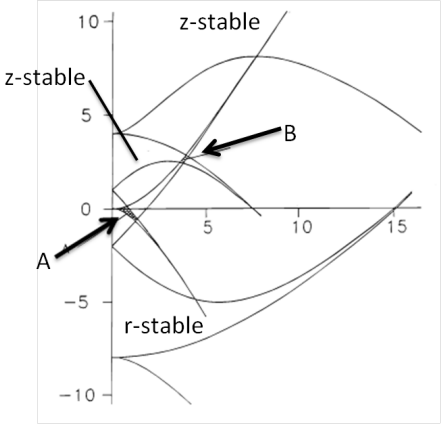


Figure D.3: Mathieu stability diagram, showing a a_x-, and q_z-space, where the stable regions are shown for r, and z is depicted. The ions must be stable in both the the $\vec{r}$-, and the $\vec{z}$ direction to obtain a stable trajectory through the quadrupole. The stable areas are where the $r-stable$, and $z-stable$ overlap, which is marked by A and B. (March, 1997)

D.7 can be describes as stability/instability diagrams, as shown in Figure D.3. Mostly all instrument focus on the A region on the figure, where the stability region is given by the the Mathieu parameters a_u, and q_u.

An important parameter to know in MS is the low-mass cutoff, that is the lowest ion weight, which can be stabilized by the mass analyzer. This can be investigation using the stability diagrams as follows. The subscript u of a, and q is given as $u = r, z$, and by converting to the subscripts a_z, and q_z it is possible to create a more detailed stability diagram as shown in Figure D.4. The low-mass cuttoff is seen on the picture where $\beta_z = 1$ intersects the q_z-axis, at $q_z = 0.908$. This point is also known as the working point, which can be stored in the mass analyzer.

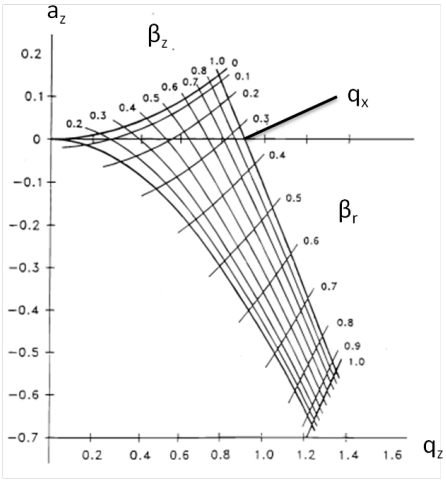


Figure D.4: Mathieu stability diagram, showing a a_z-, and q_z-space. This diagram shows the stability in both directions, i.e. $\vec{r}$, and $\vec{z}$, in the center of the quadrupole ion trap. β_z, and β_r values are given as stability boundaries. $\beta_z = 1$ boundary intersects the $q_z = 0.908$, and this location is known as q_max. (March, 1997)

The quadrupole ion trap is able to store ions in stable trajectories. These trajectories takes the

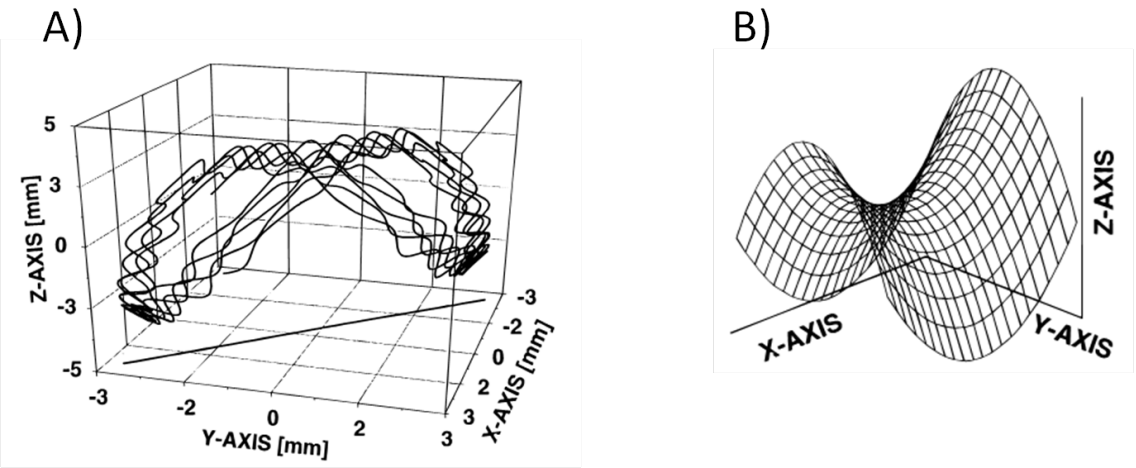


Figure D.5: A) An ion in an quadrupole ion trap moves in a Lissajous Curve, i.e. figure of eight. B) The trapped ion effectively moves on the poteintial surface, and will always seek to moves towards lower energy. When the ion moves on the potential-surface, the surface is altered, which results in a trajectory that is similar to a rollercoster. (March, 1997)

form as figures of eight in three dimensions, which is referred to as a Lissajous curve, which is govern by the frequency component $\omega_{r,0}$, and $\omega_{z,0}$, which results in secular motion, as shown in Figure D.5. The segular motion given by $\omega_{z,0}$ is given in rad/sec, but is often referred to as Hz. Multiple ions will follow different trajectories, but always in figure of eights, and the weights can be determined as m/z given by Equation D.22, and Equation D.23.

Bibliography

Abram, F., E. Gunnigle, and V. O'Flaherty (2009). Optimisation of protein extraction and 2-de for metaproteomic of microbial communites from anaerobic wastewater treatment biofilms. *Electrophoresis 30*, 4149–4151.

Aebersold, R. and M. Mann (2003). Mass spectrometry-based proteomics. *Nature 422*, 198–207.

Albertsen, M., L. B. S. Hansen, A. M. Saunders, P. H. Nielsen, and K. L. Nielsen (2011). The metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *The ISME Journal*, 1–13.

Barr, J. J., M. L. Hastie, T. Fukushima, M. R. Plan, G. Tyson, J. J. Gorman, and P. L. Bond (2011). Metaproteomic analysis of laboratory scale phosphorus removal reactors reveals functional insights of aerobic granular sludge. In *C0141*.

Benndorf, D., G. U. Balcke, H. Harms, and M. von Bergen (2007). Functional metaproteme analysis of protein extracts from contaminated soil and ground water. *ISME 1*, 224–234.

Benndorf, D., C. Vogt, N. Jehmlich, Y. Schmidt, H. Thomas, G. Woffendin, A. Shevchenko, H.-H. Richnow, and M. von Bergen (2009). Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediminets. *Biodegradation 20*, 737–750.

Bern, M., Yuhan, and D. Goldberg (2007). Lookup peaks: A hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Analytical Chemistry 79*, 1393–1400.

Blackall, L. L., G. R. Crocetti, A. M. Saunders, and P. L. Bond (2002). A review and update of the microbiology of enhanced biological phosphorus removal in wastewater treatment plants. *Biomedical and Life Sciences 81*, 681–691.

Bond, P. L., R. Erhart, M. Wagner, J. Keller, and L. L. Blackall (1999). Identification of some of the major groups of bacteria in efficient and nonefficient biological phosphorus removal activated sludge systems. *Applied and Environmental Microbiology 65*, 4077–4084.

Bondarenko, P. V., D. Chellius, and T. A. Saler (2002). Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Analytical Chemistry 74*, 4741–4749.

Box, J. D. (1983). Investigation of the folin-ciocalteau phenol reagent for the determination of polyphenolic substances in natural waters. *Water Research 5*, 511–525.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry 72*, 248–254.

Bradstreet, R. B. (1954). The kjeldahl method for organic nitrogen. *Analytical Chemistry 26*, 185–187.

Braisted, J. C., S. Kuntumalla, C. Vogel, E. M. Marcotte, A. R. Rodrigues, R. Wang, S.-T. Huang, E. S. Ferlanti, A. I. Saeed, R. D. Fleischmann, S. N. Peterson, and R. Pieper (2008). The apex quantitative proteomics tool: Generating protein quantitation estimates from lc-ms/ms proteomics results. *BMC Bioinformatics 9*, 1–11.

Broekman, S. (2007). *Etablierung der Methodik zur Darstellung des extrazellulären Proteoms in Biofilmen eines mucoiden Stammes von Pseudomonas aeruginosa*. Ph. D. thesis, Facchbereich Chemie der Universität Duisburg-Essen.

Chelius, D. and P. V. Bondarenko (2002). Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research 1*, 317–323.

Conrad, A., M. K. Suutari, M. M. Keinänen, A. Cadoret, P. Faure, L. Mansuy-Huault, and J.-C. Block (2003). Fatty acids of lipid fractions in extracellular polymeric substances of activated sludge flocs. *Lipids 38*, 1093–1105.

Cox, J. and M. Mann (2008). Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology 26*, 1367–1372.

Craig, R. and R. C. Beavis (2004). Tandem: matching proteins with mass spectra. *Bioinformatics 20*, 1466–1467.

Curtis, T. P., W. T. Sloan, and J. W. Scannel (2002). Estimating prokaryotic diversity and its limits. *Proceeding of the National Academy of Sciences, PNAS 99*, 10949–10499.

de Hoffman, E. and V. Stroobant (2009). *Mass Spectrometry - Principles and Applications* (3rd ed.). Wiley Inc.

de Hoffmann, E. and V. Stroobant (2007). *Mass Spectrometry - Principles and Applications*. Wiley.

Denecke, M. (2006). Protein extraction from activated sludge. *Water Science and Technology 54*, 175–181.

Denef, V. J., M. B. Shah, N. C. VerBerkmoes, R. L. Hettich, and J. F. Banfield (2007). Implications of strain- and species-level sequence divergence for community and isolate shotgun proteomic analysis. *Journal of Proteome Research 6*, 3152–3161.

Denef, V. J., N. C. VerBerkmoes, M. B. Shah, P. Abraham, M. Lefsrud, R. L. Hettich, and J. F. Banfield (2008). Proteomics-inferred genome typing (pigt) demonstrates inter-population recombination as a strategy for environmental adaptation. *Environmental Microbiology 11*, 313–325.

Deutsch, E. W., L. Mendoza, D. Shteynberg, T. Farrah, H. Lam, N. Tasman, Z. Sun, E. Nilsson, B. Pratt, B. Prazen, J. K. Eng, D. B. Martin, A. I. Nesvizhskii, and R. Aebersold (2010). A guided tour of the trans-proteomic pipeline. *Proteomics 10*, 1150–1159.

Djursing, T. (2008, January). Hovedpinepiller drøner direkte gennem rensningsanlæg. http://ing.dk/artikel/84889-hovedpinepiller-droener-direkte-gennem-rensningsanlaeg. Last visited: 30th december 2011.

Doherty, M. K., D. E. Hammond, M. J. Clague, S. J. Gaskell, and R. J. Beynon (2009). Turnover of the human proteome: determination of protein intracellular stability by dynamic silac. *Journal of Proteome Research 8*, 104–112.

Dole, M., L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson, and M. B. Alice (1968). Molecular beams of macroions. *The Journal of Chemical Physics 49*, 2240–2249.

Dominiak, D. M., J. L. Nielsen, and P. H. Nielsen (2011). Extracellular dna is abundant and important for microcolony strength in mixed microbial biofilms. *Environmental Microbiology 13*, 710–721.

Ehlers, M. M. and T. E. Cloete (1999a). Comparing the protein profiles of 21 different activated sludge systems after sds-page. *Water Research 33*, 1181–1186.

Ehlers, M. M. and T. E. Cloete (1999b). Protein profiles of phosphorus- and nitrate-removing activated sludge systems. *Water SA (South Africa) 25*, 351–356.

Elias, J. E. and S. P. Gygi (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods 4*, 207–214.

Eng., J. K., A. L. McCormack, and J. R. Yates III (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry 5*, 976–989.

Engen, J. R. (2009). Analysis of protein conformation and dynamics by hydrogen/deuterium exchange ms. *81 81*, 7870–7875.

Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse (1989). Electrospray ionization for the mass spectrometry of large biomolecules. *Science 64-71*, 260.

Ficarro, S., M. McCleland, P. Stukenberg, D. Burke, M. Ross, J. Shabonowitz, D. Hunt, and F. White (2002). Phosphoproteome analysis by mass spectrometry and its application to saccharomyces cerevisiae. *Nature Biotechnology 20*, 301–305.

Flemming, H.-C., T. R. Neu, and D. J. Wozniak (2007). The eps matrix: The "house of biofilm cells". *Journal of Bacteriology 189*, 7945–7947.

Flemming, H.-C. and J. Wingender (2010). The biofilm matrix. *Nature Review - Microbiology 8*, 623–633.

Florens, L., M. J. Carozza, S. K. Swanson, M. Fournier, M. K. Coleman, J. L. Workman, and M. P. Washburn (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *NIH Public Access - Methods 40*, 303–311.

Folin, O. and W. Dennis (1912). On phosphotungstic-phosphomolybdic compounds as color reagents. *Biological Chemistry 12*, 239–243.

Frølund, B., T. Griebe, and P. H. Nielsen (1995). Enzymatic activity in the activated-sludge floc matrix. *Applied Microbiol Biotechnology 43*, 755–761.

Frølund, B., R. Palmgren, K. Keiding, and P. H. Nielsen (1996). Extraction of extracellular polymers from activated sludge using a cation exchange resin. *Water Research 30*, 1749–1758.

Gaudy, A. F. (1962). Colorimetric determination of protein and carbohydrate. *Indust. Wat. Wastes 7*, 17–22.

Gorman, J. J., T. P. Wallis, and J. J. Pitt (2002). Protein disulfide bond determination by mass spectrometry. *Mass Spectrometry Reviews 21*, 183–216.

Gornall, A. G., C. J. Bardawill, and M. M. David (1949). Determination of serum proteins by means of the biuret reaction. *Biological Chemistry 177*, 751–766.

Görg, A., W. Weiss, and M. J. Dunn (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics 4*, 3665–3685.

Griffin, N. M., J. Yu, F. Long, P. Oh, S. Shore, Y. Li, J. A. Koziol, and J. E. Schnitzer (2010). Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nature Biotechnology 28*, 83–91.

Gygi, S. P., B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology 17*, 994–999.

Harris, D. A. (1987). *Spectrophotometry and spectrofluorimetry: A Practical Approach*. Oxford Press.

Hart, J. R. and W. R. Grace (2005). Ethylenediaminetetraacetic acid and related chelating agents. *Wiley Online Library*.

Hägglund, P., R. Matthiesen, F. Elortza, P. H. jrup, P. Roepstorff, O. N. rregaard Jensen, and J. Bunkenborg (2007). An enzymatic deglycosylation scheme enabling identification of core fucosylated n-glycans and o-glycosylation site mapping of human plasma proteins. *Journal of Proteome Research 6*, 3021–3031.

Higgins, M. J. and J. T. Novak (1997). Characterization of exocellular protein and its role in bioflocculation. *Journal of Environmental Engineering 123*, 479–485.

Hu, Q., R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. G. Cooks (2005). The orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry 40*, 430–443.

Hunt, D. F., J. R. Yates III, J. Shabanowitz, S. Winston, and C. R. Hauer (1986). Protein sequencing by tandem mass spectrometry. *Proceeding of the National Academy of Sciences 83*, 6233–6237.

Ishihama, Y., Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann (2005). Exponentially modified protein abundance index (empai) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular and Cellular Proteomics 4*, 1265–1272.

Kebarle, P. and U. H. Verkerk (2009). Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews by Wiley Periodicals, INC 28*, 898–917.

Keller, A., I. Eng, N. Zhang, X. jun Li, and R. Aebersold (2005). A uniform proteomics ms/ms analysis platform utilizing open xml file formats. *Molecular Systems Biology 1*, 1–8.

Keller, A., A. I. Nesvizhskii, E. Kolker, and R. Aebersold (2002). Empirical statistical model to estimate the accuracy of peptide identification made by ms/ms and database search. *Analytical Chemistry 74*, 5383–5392.

Kielberg, V. and L. Rasmussen (2005). *Proteiner - oprensning og karakterisering*. Nordisk Forlag. ISBN: 87-02-03301-1.

Kintner III, P. K. and J. P. van Buren (1982). Carbohydrate interference and its correction in pectin analysis using the m-hydroxydiphenyl method. *Journal of Food Science 3*, 756–759.

Kommune, A. Om Renseanlæg. http://www.forsyning.dk/aalborg-forsyning/kloak/spildevandsrensning/om-renseanlaeg.aspx.

Krystal, G. (1987). A silver-binding assay for measuring nanogram amounts of protein in solution. *Analytical Biochemistry 15*, 86–96.

Kuhn, R. (2010). *Gel-based metaproteomic approach for functional MBR biomass characterisation*. Ph. D. thesis, DOTTORATO DI RICERCA IN BIOCHIMICA E BIOLOGIA MEDICA CICLO XXII. Pages: 1-189.

Kuhn, R., D. Benndorf, E. Rapp, U. Reichl, L. L. Palese, and A. Pollice (2011). Metaproteome analysis of sewage sludge from membrane bioreactors. *Proteomics 11*, 1–7.

Kusmierz, J. J., R. Sumrada, and D. M. Desiderio (1990). Fast atom bombardment mass spectrometric quantitative analysis of methionine-enkephalin in human pituitary tissues. *Analytical Chemistry 62*, 2395–2400.

Larsen, P., J. L. Nielsen, D. Otzen, and P. H. Nielsen (2008). Amyloid-like adhesins produced by floc-forming and filamentous bacteria in activated sludge. *Applied Environmental Microbiology 74*, 1517–1526.

Li, X., P. G. A. Pedrioli, J. Eng, D. Martin, E. C. Yi, H. Lee, and R. Aebersold (2004). A tool to visualize and evaluate data obtained by liquid chromatography-electrospray ionization-mass spectrometry. *Analytical Chemistry 76*, 3856–3860.

Link, A. J., J. Eng, D. M. Schieltz, E. Carmack, G. J. Mize, D. R. Morris, B. M. Garvik, and J. R. Yates (1999). Direct analysis of protein complexes using mass spectrometry. *Nature 17*, 676–682.

Listgarten, J. and A. Emili (2005). Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Molecular and Cellular Proteomics 4*, 419–434.

Liu, H., R. G. Sadygov, and J. R. Yates III (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry 76*, 4193–4201.

Lo, I., V. J. Denef, N. C. VerBerkmoes, M. B. Shah, D. Goltsman, G. DiBartolo, G. W. Tyson, E. E. Allen, R. J. Ram, J. C. Detter, P. Richardson, M. P. Thelen, R. L. Hettich, and J. F. Banfield (2007). Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature 446*, 537–541.

Lodish, H., A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, and A. B. H. Ploegh (2007). *Molecular Cell Biology*. Freeman.

Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall (1951). Protein measurement with the folin phenol reagent. *Biological Chemistry 193*, 265–275.

Lu, P., C. Vogel, R. Wang, X. Yao, and E. M. Marcotte (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology Genetic Engeneering Reviews 25*, 117–124.

Mainstone, C. P. and W. Parr (2002). Phosphorus in rivers - ecology and management. *The Science of the Total Environment 282-283*, 25–47.

Mallick, P., M. Schirle, S. S. Chen, M. R. Flory, H. Lee, D. Martin, J. Ranish, B. Raught, R. Schmitt, T. Werner, B. Kuster, and R. Aebersold (2007). Computational prediction of proteotypic peptides for quantitative proteomics. *Nature Biotechnology 25*, 125–131.

Mann, M. and O. N. Jensen (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology 21*, 255–261.

March, R. E. (1997). An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry 32*, 351–369.

Marcotte, E. M. (2007). How do shotgun proteomics algorithms identify proteins? *Nautre Biotechnology 25*, 755–757.

Mardis, E. R. (2008). Next-generation dna sequencing methods. *Annual Review Genomics Human Genetics 9*, 387–402.

Marfelt, B. and S. Wittrup (2011, October). Alt for store mængder kviksølv og medicinrester ryger ud i øresund. http://ing.dk/artikel/122886-alt-for-store-maengder-kviksoelv-og-medicinrester-ryger-ud-i-oeresund. Last visited: 30th december 2011.

Martin, H. G., N. Ivanova, V. Kunin, F. Warnecke, K. W. Barry, A. C. McHardy, C. Yeates, S. He, A. A. Salamov, E. Szeto, E. Dalin, N. H. Putnam, H. J. Shapiro, J. L. Pangilinan, I. Rigoutsos, N. C. Kyrpides, L. L. Blackall, K. D. McMahon, and P. Hugenholtz (2006). Metagenomic analysis of two enhanced biological phosphorus removal (ebpr) sludge communities. *Nature Biotechnology 24*, 1263–1269.

Matsuyama, T. and Y. Nakagawa (1996). Surface-active exolipids: analysis of absolute chemical structures and biological functions. *Journal of Microbiological Methods 25*, 165–175.

McLachlin, D. T. and B. T. Chait (2003). Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Analytical Chemistry 75*, 6826–6836.

Medzihradszky, K. F. (2008). Characterization of site-specific n-glycosylation. *Methods in Molecular Biology 446*, 293–316.

Michalski, A., J. Cox, and M. Mann (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccesible to data-dependent lc-ms/ms. *Journal of Proteome Research 10*, 1785–1793.

Mulcahy, H., L. Charron-Mazenod, and S. Lewenza (2008). Extracellular dna chelates cations and induces antibiotic resistance in pseudomonas aeruginosa biofilms. *PLoS Pathogens 4*, 1–12.

Neilson, K. A., N. A. Ali, S. Muralidharan, M. Mirzaei, M. Mariani, G. Assadourian, A. Lee, S. C. van Sluyter, and P. A. Haynes (2011). Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics 11*, 535–553.

Nesvizhskii, A. I. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods 4*, 787–797.

Nesvizhskii, A. I. and R. Aebersold (2004). Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem ms. *Drug Discovery Today 9*, 173–181.

Nielsen, P. H. and K. Keiding (1998). Disintegration of activated sludge flocs in presence of sulfide. *Water Research 32*, 313–320.

Nielsen, P. H., A. T. Mielczarek, C. Kragelund, J. L. Nielsen, A. M. Saunders, Y. Kong, A. A. Hansen, and J. Volltertsen (2010). A conceptual ecosystem model of microbias communites in enhanced biological phosphorous removal plants. *Water Research*, 1–19.

Oehmen, A., P. C. Lemos, G. Carvalho, Z. Yuan, J. Keller, L. L. Backall, and M. A. M. Reis (2007). Advances in enchanced biological phosphorus removal: From micro to macro scale. *Water Research 41*, 2271–2300.

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *Biological Chemistry 25*, 4007–4021.

Ogunseitan, O. A. (1993). Direct extraction of proteins from environmental samples. *Journal of Microbial Methods 17*, 273–281.

Ong, S.-E., B. Blagoev, I. Krachmarova, D. B. Kristensen, H. Steen, A. Pandey, and M. Mann (2002). Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular and Cellular Proteomics 1*, 376–386.

Ong, S.-E. and M. Mann (2005). Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology 1*, 252–262.

Otzen, D. and P. H. Nielsen (2008). We find them here, we find them there: functional bacterial amyloid. *Cellular and Moldecular Life Sciences 65*, 910–927.

Palumbo, A. M., S. A. Smith, C. L. Kalcic, M. Dantus, P. M. Stemmer, and G. E. Reid (2010). Tandem mass spectrometry strategies for phosphoproteome analysis. *Mass Spectrometry Reviews 30*, 600–625. wileyonlinelibrary.com.

Park, C. and J. T. Novak (2007). Characterization of activated sludge exocellular polymers using several cation-associated extraction methods. *Water Research 41*, 1679–1688.

Park, C. and J. T. Novak (2009). Characterization of lectins and bacterial adhesins in activated sludge flocs. *Water Environment Research 81*, 755–764.

Park, C., J. T. Novak, R. F. Helm, Y.-O. Ahn, and A. Esen (2008). Evaluation of the extracellular proteins in full-scale activated sludges. *Water Research 42*, 3879–3889.

Pedrioli, P. G. A., J. K. Eng, R. Hubley, M. Vogelzang, E. W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R. H. Angeletti, R. Apweiler, K. Cheung, C. E. Costello, H. Hermajakob, S. Huan, R. K. Julian, E. Kapp jr, M. E. McComb, S. G. Oliver, G. Omenn, N. W. Paton, R. Simpson, R. Smith, C. F. Taylor, W. Zhu, and R. Aebersold (2004). A common opren reprensentation of mass spectrometry data and its application to proteomics reasearch. *Nature Biotechnology 22*, 1459–1466.

Peng, J. and S. P. Gygi (2001). Proteomics: the move to mixtures. *Journal of Mass Spectrometry 36*, 1083–1091.

Perkins, D. N., D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis 20*, 3551–3567.

Peterson, G. L. (1979). Review of the folin phenol protein quantification method of lowry rosebrough, farr and randall. *Analytical Biochemistry 100*, 201–220.

Ram, R. J., N. C. V. VerBerkmoes, M. P. Thelen, G. W. Tyson, B. J. Baker, R. C. Blake II, M. Shah, R. L. Hettich, and J. F. Banfield (2005). Community proteomics of a natural microbial biofilm. *Science 308*, 1915–1920.

Rappsilber, J., U. Ryder, A. I. Lamond, and M. Mann (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Research 12*, 1231–1245.

Raunkjær, K., T. Hvitved-Jacobsen, and P. H. r Nielsen (1994). Measurement of pools of protein, carbohydrate and lipid in domestic. *Water Research 2*, 251–262.

Ron, E. Z. and E. Rosenberg (2001). Natural roles of biosurfactants. *Environmental Microbiology 3*, 229–236.

Ross, P. L., Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin (2004). Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular and Cellular Proteomics 3*, 1154–1169.

Schwartz, J. C., M. W. Senko, and J. E. P. Syka (2002). A two-dimensional quadrupole ion trap mass spectrometer. *Journal of Americal Society for Mass Spectrometry 13*, 659–669.

Selbach, M., B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky (2008). Widespread changes in protein synthesis induced by microrna. *Nature Biotechnology 455*, 58–63.

Seviour, T., L. K. Lambert, M. Pijuan, and Z. Yuan (2010). Structural determination of a key exopolysaccharide in mixed culture aerobic sludge granules using nmr spectroscopy. *Environmental Science Technology 44*, 8964–8970.

Shevchenko, A., M. Wilm, O. Vorm, and M. Mann (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical Chemistry 68*, 850–858.

Shukla, A. K. and J. H. Futrell (2000). Tandem mass spectrometry: dissociation of ions by collisional activation. *Jorunal of Mass Spectrometry 35*, 1069–1090.

Silva, J. C., M. V. Gorenstein, Guo-Zhong, J. P. C. Vissers, and S. J. Gerommanos (2006). Absolute quantification of proteins by lcmse a virtue of parallel ms acquisition. *Molecular and Cellular Proteomics 5*, 144–156.

Simpson, I. M. and O. Sonne (1982). A simple, rapid, and sensitive method for measuring protein concentration in subcellular membrane fraction prepared by sucrose densitiy ultracentrifugation. *Analytical Biochemistry 119*, 424–427.

Siuti, N. and N. L. Kelleher (2007). Decoding protein modifications using top-down mass spectrometry. *Nature Methods 4*, 817–821.

Sleno, L. and D. A. Volmer (2004). Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry 39*, 1091–1112. www.interscience.wiley.com.

Smith, P. K., R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, and D. C. Klenk (1985). Measurement of protein using bicinchoninic acid. *Analytical Biochemistry 150*, 76–85.

Srinath, E. G., C. A. Sastry, and S. C. Pillai (1959). Rapid removal of phosphorus from sewage by activated sludge. *Cellular and Moldecular Life Sciences 15*, 339–340.

Steen, H. and M. Mann (2004). The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology 5*, 699–711.

Stensballe, A., S. eren Andersen, and O. N. Jensen (2001a). Characterization of phosphoproteins from electrophoretic gels by nanoscale fe(iii) affinity chromatography with off-line mass spectrometry. *Proteomics 1*, 207–222.

Stensballe, A., S. eren Andersen, and O. N. Jensen (2001b). Characterization of phosphoproteins from electrophoretic gels by nanoscale fe(iii) affinity chromatography with off-line mass spectrometry. *Proteomics 1*, 207–222.

Stewart, I. I., T. Thomson, and D. Figeys (2001). 18o labeling: a tool for proteomics. *Rapid Communications in Mass Spectrometry 15*, 2456–2465.

Tchobanoglous, G., F. L. Burton, and H. D. Stensel (2004). *Wastewater Engineering - Treatment and Reuse*. Mc Graw-Hill.

Tyson, G. W., J. Chapmann, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature 428*, 37–43.

Vandenbogaert, M., S. Li-Thiao-Te, H.-M. Kaltenbach, R. Zhang, T. Aittokallio, and B. Schwikowski (2008). Alignment of lc-ms images, with applications to biomarker discovery and protein identification. *Proteomics 8*, 650–672.

Venter, J. C. and et. al. (2001). The sequence of the human genome. *Science 291*, 1304.

Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. Wthie, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.-H. Rogers, and H. O. Smith (2004). Environmental genome shotgun sequencing of the sargasso sea. *Science 304*, 66–74.

VerBerkmoes, N. C., V. J. Denef, R. L. Hettich, and J. F. Banfield (2009). Systems biology functional analysis of natural microbial consortia using community proteomics. *Systems Biology 7*, 196–205.

Vilain, S., J. M. Pretorius, J. Theron, and V. S. Brözel (2009). Dna as an adhesin: Bacillus cereus requires extracellular dna to form biofilms. *Applied Environmental Microbiology 75*, 2861–2868.

Wang, W., H. Zhou, H. Lin, S. Roy, T. A. Shaler, L. R. Hill, S. Norton, P. Kumar, M. Anderle, and C. H. Becker (2003). Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Analytical Chemistry 75*, 4818–4826.

Washburn, M. P., D. Wolters, and J. R. Yates III (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology 19*, 242–247.

Weiss, F. U., M. Ruthenbürger, E. Hammer, U. Völker, and M. Lerch (2006). Assessment of isoform specificity of a polyclonal elastase elisa. *Pancreas 33*, 911–920.

Wessel, D. and U. I. Flügge (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical Chemistry 138*, 141–143.

Whitchurch, C. B., T. Tolker-Nielsen, P. C. Ragas, and J. S. Mattick (2002). Extracellular dna required for bacterial biofilm formation. *Science 295*, 1487.

White, F. M. (2011). The potential cost of high-throughput proteomics. *Science Signaling 4*, 1–3.

Whitehouse, C. M., R. N. Dreyer, M. Yamashita, and J. B. Fenn (1985). Electrospray interface for liquid chromatographs and mass spectrometers. *Analytical Chemistry 57*, 675–679.

Wilen, B.-M., B. Jin, and P. Lant (2003). The influence of key chemical constituents in activated sludge on surface and flocculating properties. *Water Research 37*, 2127–2139.

Wilen, B.-M., K. Keiding, and P. H. Nielsen (2004). Flocculation of activated sludge flocs by stimulation of the aerobic biological activity. *Water Research 38*, 3909–3919.

Wilkins, M. R., J. C. Sanchez, A. A. Gooley, R. D. Appel, I. Humphery-Smith, D. F. Hochstrasser, and K. L. Wilkins (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology Genetic Engeneering Reviews 13*, 19–50.

Wilmes, P., A. F. Anderson, M. G. Lefsrud, M. Wexler, M. Shah, B. Zhang, R. L. Hettich, P. L. Bond, N. C. VerBerkmoes, and J. F. Banfield (2008). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enchanced biological phosphorus removal. *The ISME Journal 2*, 853–864.

Wilmes, P. and P. L. Bond (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environmental Microbiology 6*, 911–920.

Wilmes, P. and P. L. Bond (2006a). Metaproteomics studying functional gene expression in microbial ecosystems. *TRENDS in Microbiology 14*, 92–97.

Wilmes, P. and P. L. Bond (2006b). Towards exposure of elusive metabolic mixed-culture processes: the application of metaproteomic analyses to activated sludge. *Water Science & Technology 54*, 217–226.

Wilmes, P. and P. L. Bond (2009). Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the earth's biogeochemical cycles. *Current Opinion in Microbiology 12*, 310–317.

Wisniewski, J. R., A. Zougman, N. Nagaraj, and M. Mann (2009). Universal sample preparation method for proteome analysis. *Nature Methods 6*, 359–362.

Wu, J. (2008). Disulfide bond mapping by cyanylatin-induced cleavage and mass spectrometry. *Methods in Molecular Biology 446*, 1–20.

Xiang, T., C. Chumsae, and H. Liu (2009). Localization and quantitation of free sulhydryl in recombinant monoclonal antibodies by differential labelling with 12c and 12c iodoacetic acid and lc-ms analysis. *Analytical Chemistry 81*, 8101–8108.

Yang, L., Y. Liu, T. Markussen, N. Høiby, T. Tolker-Nielsen, and S. Molin (2011). Pattern differentiation in co-culture biofilms formed by staphylococcus aurerus and pseudomonas aeruginosa. *FEMS Immunology and Medical Microbiology 62*, 339–347.

Yeung, Y.-G., E. Nieves, R. Angeletti, and E. R. Stanley (2008). Removal of detergent from protein digests for mass spectrometry analysis. *Analytical Biochemistry 15*, 135–137.

Yu, G.-H., P.-J. He, L.-M. Shao, and D.-J. Lee (2007). Enzyme activities in activated sludge flocs. *Applied Microbiol Biotechnology 77*, 605–612.

Zhang, N. and L. Li (2004). Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. *Rapid Communications in Mass Spectrometry 18*, 889–896.

Zhao, L., R. Wu, G. Han, H. Zhou, L. Ren, R. Tian, and H. Zou (2008). The highly selective capture of phosphopeptides by zirconium phosphonate-modified magnetic nanoparticles for phosphoproteome analysis. *Journal of the American Society for Mass Spectrometry 8*, 1176–1186. http://www.springerlink.com/content/q8h8831765481241/.