Metaproteomics of activated sludge from a wastewater treatment plant - a pilot study

“Dataset Brief”

Sebastian Püttker\textsuperscript{1,*}, Fabian Kohrs\textsuperscript{1,*}, Dirk Benndorf\textsuperscript{1}, Robert Heyer\textsuperscript{1}, Erdmann Rapp\textsuperscript{2} and Udo Reichl\textsuperscript{1,2}

\textsuperscript{1} Bioprocess Engineering, Otto von Guericke University Magdeburg, Magdeburg, Germany
\textsuperscript{2} Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
* Equal authorship

Correspondence: Dr. rer. nat. Dirk Benndorf, Bioprocess Engineering, Otto von Guericke University Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany. E-Mail: benndorf@mpi-magdeburg.mpg.de. Fax: ++49 / 391 / 67-11209.

Abbreviations: M, molar; WWTP, wastewater treatment plant; S, summer; W, winter; 3D, three dimensional; IAA, iodoacetamide; FA, formic acid; MGF, Mascot generic format; MPA, MetaproteomeAnalyzer; KEGG, Kyoto Encyclopedia of Genes and Genomes; KO, KEGG Orthology; EC, Enzyme Commission

Keywords: Metaproteomics / Wastewater treatment plant / Activated sludge / MetaProteomeAnalyzer / Offgel isoelectric focusing

3,080 Words
Abstract

In this study the impact of protein fractionation techniques prior to LC/MS analysis was investigated on activated sludge samples derived at winter and summer condition from a full-scale wastewater treatment plant (WWTP). For reduction of the sample complexity, different fractionation techniques including RP-LC (1D-approach), SDS-PAGE and RP-LC (2D-approach) as well as RP-LC, SDS-PAGE and liquid IEF (3D-approach) were carried out before subsequent ion trap MS analysis. The derived spectra were identified by MASCOT search using a combination of the public UniProtKB/Swiss-Prot protein database and metagenome data from a WWTP (data are available via ProteomeXchange with identifier PXD001547). The results showed a significant increase of identified spectra, enabled by applying IEF and SDS-PAGE to the proteomic workflow. Based on meta-proteins, a core metaproteome and a corresponding taxonomic profile of the wastewater activated sludge were described. Functional aspects were analyzed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway library by plotting KEGG Orthology identifiers (KO numbers) of protein hits into pathway maps of the central carbon (map01200) and nitrogen metabolism (map00910). Using the 3D-approach, most proteins involved in glycolysis and citrate cycle and nearly all proteins of the nitrogen removal were identified, qualifying this approach as most promising for future studies.
Due to anthropogenic processes, a high amount of organic substances like sewage, fertilizers and detergents from municipal and industrial sources are released into wastewater. Discharging of polluted water into the environment without further treatment endangers public health and can cause eutrophication or poisoning of aquatic ecosystems. Therefore, the separation of solid material and the removal of organic and inorganic contaminants in wastewater treatment plants (WWTP) are essential to guarantee water purity. The biological treatment is performed by a complex community of microorganisms commonly referred to as activated sludge and comprises the degradation of organic material, the removal of nitrogen by nitrification and denitrification as well as the elimination of phosphorus by phosphate accumulating bacteria. However, the microbial population in activated sludge is sensitive to external influences. In particular, contamination with toxic compounds from industry, dilution of wastewater by strong rainfall or formation of layers of scum due to seasonal changes in temperature can lead to unstable operational conditions.

Molecular biological approaches (i.e. community fingerprinting, metagenome sequencing) revealed that only a small portion of the microbial community typically found in WWTPs is cultivable in the lab. In order to understand its function without isolating single strains, metaproteomic analysis has been applied to activated sludge samples [1, 2]. Major challenges for this approach are the presence of humic compounds interfering with protein separation, the high complexity of the community, and the lack of sufficient genomic sequences for protein identification [1, 3].

In this work a proteomic workflow previously established for digestates of biogas plants [4] was applied to characterize composition of microbial communities of activated sludge from WWTP. Different gel-free and gel-based fractionation techniques were performed, to achieve a high resolution of low abundant proteins and to enable detailed description of the complex metaproteome. Activated sludge was sampled in winter (W, 21st of January 2014) and summer (S, 2nd of June 2014) period to cover the core
metaproteome as well as to represent major seasonal changes. Samples were collected from three
different spots of the aerobic basin of the WWTP Gerwisch (Saxony-Anhalt, Germany) and handled as
triplicates (see Supplementary Table S1 for general process data). A detailed description of the methods
used is given in Supplementary Note 1. Briefly, cell disruption and protein extraction were performed
simultaneously using a ball mill with liquid phenol (10 g phenol plus 1 ml water). Proteins were
precipitated afterwards with ammonium acetate in methanol (0.1 M) and washed with ice-cold 80%
acetone and 70% ethanol. Finally, extracts were suspended in 7 M urea/2 M thiourea buffer containing
0.01 g mL\(^{-1}\) DTT and protein contents were determined using the amido black dye assay. For
fractionation of the protein samples, SDS-PAGE, gel-free liquid IEF (allowing the recovery of proteins
after isoelectric focusing) and RP-LC were used and combined (1D: RP-LC; 2D: SDS-PAGE + RP-LC; 3D: IEF
+ SDS-PAGE + RP-LC) as published previously [4]. SDS-PAGE gels were cut into equal pieces for tryptic
digestion, resulting in ten fractions for the 2D and 120 fractions for 3D-approach, respectively. After LC
the samples were directly transferred to an ESI-amazon™ iontrap MS (Bruker Daltonics, Bremen,
Germany) for generation of MS/MS spectra. Identification of the MS spectra was performed using
MASCOT 2.4 (Matrix Science, London, England). In a first round of data evaluation the public
UniProtKB/Swiss-Prot protein database (state 2014/10/23, [5]) has been used for protein identification,
but were insufficient. In previous studies the usage of a metagenome database showed at least a two-
fold increase in protein identifications compared to public protein databases such as UniProt/Swiss-Prot
[2]. Obviously, a database derived by metagenome studies from the same plant should result in the best
identification rate and should therefore be preferred to third-party metagenome. Nevertheless, the
functional core microbiome is expected to be very similar at different time points and between
comparably operated plants, enabling the usage of third-party metagenomes. However, metagenome
data from the WWTP sampled in this study were not available, for which reason data, comprising
1,687,643 protein sequences, from the similar operated WWTP in Aalborg (Denmark, [6]) was used as an
alternative source. With this additional database about twice as much identifications could be achieved.
For data analysis the in-house software MetaProteomeAnalyzer 1.0.4 (MPA) was used for parsing MASCOT .dat-files and claiming a FDR <1% [7]. Identified peptide sequences based on the metagenome data were assigned from UniProt/Swiss-Prot database, using the protein basic local alignment search tool algorithm (BLAST [8]) and an e-value cutoff of 0.0001.

The protein profiles of activated sludge samples from W and S period showed strong background smearing of the Coomassie Brilliant Blue dye for all lanes after SDS-PAGE (2D-approach) (Figure 1A). Humic compounds, which share similar physico-chemical characteristics as proteins, apparently could not be completely eliminated during the sample preparation procedure. However, a few weak protein bands at 25 kDa, 33 kDa and 43 kDa were observed for all three cascade sample points. Winter and summer period samples showed no significant differences in the protein profiles but rather the intensities of S period samples were slightly higher as judged by visual assessment. The combination of a liquid isoelectric focusing prior to a SDS gel electrophoresis (3D-approach) led to an increased number of protein bands over the pH range 3–10 (Figure 1B). Also, the background staining with Coomassie dye was slightly reduced due to the distribution of humic compounds over a wide pH range. In addition, part of humic compounds precipitated in the IPG-Strip during the focusing step. Unfortunately, some regions of the gel lanes still remained smeared. Some protein bands were only observed in single pH fractions whereas others occurred in up to five fractions, indicating the presence of protein isoforms with different pH values as previously shown in 2D-gels by Kuhn et al. (2011) [1]. The visual comparison of gels from W and S period samples revealed unique bands (Figure 1B, red boxes) indicating that seasonal changes were already detectable by combined liquid isoelectric focusing and SDS-PAGE.

In a next step, tryptic digests were analyzed by LC-MS/MS and all separation techniques quantitatively compared regarding the number of identified peptides and proteins. Furthermore, proteins were assigned using the MPA software. The number of peptides was characterized as total identified spectra and furthermore the number of non-redundant peptides was calculated (Figure 2A). Hereby, the 1D-
approach revealed a maximum of 50 identified spectra and 27 non-redundant peptides for sample W_3 and 68 total spectra and 44 non-redundant peptides in sample S_3, respectively. The peptides were assigned to 327 (W_3) and 1,395 proteins (S_3, Figure 2B). Performing SDS-PAGE prior to LC-MS/MS (2D-approach) increased the maximum number of identified spectra to 782 in sample W_1 (442 non-redundant) and 722 in sample S_1 (312 non-redundant). The best 2D-approach yielded 4,366 and 2,799 protein identifications in sample W_1 and S_2. Obviously the 1D-approach showed unevenly distributed identifications resulting in insufficient reproducibility, whereas the 2D-approach revealed more consistent results.

For the 3D-approach liquid IEF was carried out before gel electrophoresis and LC. In comparison to the previous approaches, 8,120 peptides (including 2,037 non-redundant hits) in sample W_1 and 4,613 peptides (1,512 non-redundant hits) in sample S_1. Peptides of both samples were matched to 10,095 proteins in sample W_1 and 9,483 in sample S_1 (Figure 2B), respectively. The high number of identified proteins compared to the peptides was mainly attributed to redundant assignments of peptides to highly conserved proteins from different species, e.g. chaperones, elongation factors or keratins as discussed by Benndorf and Reichl (2014) [3]. In order to remove redundancy, all proteins sharing at least one peptide were combined to meta-proteins (Figure 2B) for the following analyses. Using the MPA software, taxonomies of meta-proteins were assigned based on the taxonomy of the most specific peptide. After this processing step significantly fewer meta-proteins remained for the winter sample: a top of 13 meta-protein for the 1D-approach (W_3), 206 meta-proteins for the 2D-approach (W_1) and 936 meta-proteins for the 3D-approach (W_1; similar trends were obtained for sample S_1 in Figure 2B).

Further taxonomic and functional analyses were exclusively carried out with the datasets obtained from the 2D-approach. Therefore, hierarchical taxonomic profiles were produced with the Krona visualization tool [9] using the spectral count of the meta-proteins along with the corresponding taxonomic annotations. A Krona plot exemplarily displaying the complete results of the identified taxonomic
hierarchy from superkingdom to species was created for the sample W_1 (Downloadable File D1, supporting material). Surprisingly, the chart is not only dominated by bacterial (61% of all spectra) but also by eukaryotic meta-proteins (37%), which were mainly represented by the phylum chordata (*Mus musculus, Homo sapiens*). Major proteins were keratins and the pancreatic enzyme elastase. The latter one is released into the colon and was also detected previously in activated sludge samples [1, 3]. Small fractions of archaeal (2%) and viral (0.4%) meta-proteins were also identified. The presence of virus and phages in activated sludge was shown before by metagenomics and electron microscopy [10]. Numerous fractions with unknown taxonomy, in particular for lower taxonomic levels, were observed showing that not all meta-proteins could be assigned to specific organisms. This is due to peptides sharing the common ancestor on a higher taxonomic level.

Focusing on the microbial activity during wastewater treatment, a detailed comparison of the W_1 and S_1 samples (2D-approach) was carried out by considering only results of bacterial and archaeal meta-proteins in a next step. Applicable taxonomic levels for comparison were superkingdom, phylum, class and order (Downloadable Files D2 and D3, supporting material). The Krona plots of both samples revealed a different number of spectral counts for bacteria (462 spectra in W_1; 225 in S_1). Nevertheless, similarities could be observed regarding the ratios of bacteria (95-96%) and archaea (4–5%). At the phylum level both samples were dominated by Proteobacteria (69% in sample W_1; 74% in sample S_1), followed by Firmicutes (similar percentages in both samples and replicates; data not shown). The class level revealed major fractions of Betaproteobacteria (W_1: 53%; S_1: 63%) and Gammaproteobacteria (W_1: 15%; S_1: 13%) and minor fractions of Alpha- and Deltaproteobacteria, showing no significant differences in activated sludge samples from W and S period. Slight variations between S and W conditions at the order level were observed regarding the members of the Gamma- and Betaproteobacteria, particularly Pseudomonadales, Enterobacteriales, Burkholderiales and Nitrosomonadales. However, this variation could not be validated as a general trend neither by
evaluating the 2D replicates nor the 3D analyses due to lack of sensitivity or missing statistical proofs.

Subsequently, this observation has to be validated with further data. The presence of the Archaea is mainly represented by Euryarchaeota (92% in sample W_1, 100% in sample S_1), including the order Archaeoglobales and Methanosarcinales. Archaea of the order Thermococcales, Methanococcales and Halobacteriales were only identified in sample S_1.

A functional analysis of the identified meta-proteins was performed using the Kyoto Encyclopedia of Genes and Genomes pathway database (KEGG, [11, 12]) and the KEGG Orthology (KO) identifier. The KEGG pathway maps of the carbon metabolism (map01200) were used exemplarily for a general overview as well as the nitrogen metabolism (map00910) as a WWTP specific metabolism, respectively. Unfortunately the data of the 2D-approach showed only few identified KOs and allowed no discrimination between samples of winter and summer (data not shown), possibly due to a lack of sensitivity to detect low abundant proteins. However, using the data from the 3D-approach, revealed a much more detailed identification of proteins involved in the carbon metabolism (map01200, Figure 3).

Several enzymes from glycolysis and citrate cycle were identified in W and S samples, mainly assigned to members of the Proteobacteria (Alpha-, Beta- and Gammaproteobacteria). Single enzymes representing the reductive pentose phosphate cycle (Calvin cycle, Enzyme Commission number (EC): 2.7.1.19; 4.1.1.39) were found uniquely in sample W_1, all assigned to Bacilli of the phylum Firmicutes.

The process of nitrification, the conversion of ammonia in two steps to nitrite by Nitrosomonadales and to nitrate by Nitrospirales, plays an important role in the wastewater treatment process. For monitoring nitrification, data of the 3D-approach were mapped in the KEGG pathway map for nitrogen (map00910), equivalent to the carbon metabolism (Supplementary Figure S1). In sample W_1 ten proteins were identified, covering not only all the enzymes involved in nitrification (ammonia monooxygenase EC 1.14.99.39, hydroxylamine oxidoreductase EC 1.7.2.6, nitrite oxidoreductase EC 1.7.99.4) but also most enzymes of denitrification such as nitrate, nitrite and nitrous oxide reductases (EC 1.7.99.4; 1.7.2.1;
1.7.2.4). The sample from summer period S_1 showed similar results but revealed only ammonia
monooxygenase as enzyme for nitrification.

In conclusion, metaproteomic data of WWTP activated sludge samples from winter and summer period
were fractionated with different techniques prior to LC-MS/MS analysis. The results revealed an
increased protein identification enabled by application of liquid IEF and SDS gel electrophoresis (2D and
3D approach, Figure 2). Additionally, the availability of metagenomic data showed a significant gain of
about two-fold peptide identifications thereby widen the potential of metaproteomics in analysis of
activated sludge. On 2D-level, taxonomic and functional assignment of meta-proteins revealed a core
metaproteome and characterized its most abundant community members, which presence is in
accordance with previous studies [6]. Significant seasonal changes between W and S could not be
observed, probably due to weak differences in the process conditions (Supplementary Table 1) or due to
the lack of resolving power of the 2D-technique. However, using the 3D-approach comprehensive
identification of the protein set involved in nitrogen removal was enabled (Supplementary Figure S1),
qualifying this technique as most promising for metabolic pathway analysis of complex microbial
communities. For defining a core metaproteome of WWTPs and discovering potential seasonal or
process related changes, more comprehensive studies using the 3D-approach are required.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium
(http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [13] with the dataset
identifier PXD001547.

The Authors would like to thank Dipl.-Ing. (FH) Günter Beckermann, Dipl.-Ing. Matthias Regner and
Caroline Chrzan (B.Sc.) (Städtische Werke Magdeburg GmbH und Co. KG, WWTP Gerwisch, Germany) for
providing samples and related process data. Robert Heyer was supported by the German Environmental
Foundation (DBU), grant number 20011/136. The authors also thank the PRIDE team for technical
assistance in uploading the data.
The authors have declared no conflict of interest.
References


**Figure Legends**

**Figure 1.** SDS-PAGE gels of the 2D and 3D-approach applied for protein extracts of activated sludge of winter (W) and summer (S) period. STD: proteins standard; (A): SDS-PAGE gels of the 2D-approach; (B): SDS-PAGE of the 3D-approach. Red boxes indicate differences in the protein patterns of W and S.

**Figure 2.** Peptides, proteins and meta-proteins identified by LC-MS/MS approach for all replicates of winter (W) and summer (S) period using respective separation techniques. (A): peptides identified only once (single peptide hits) and repeatedly (multiple peptide hits); (B): proteins and meta-proteins identified.

**Figure 3.** KEGG pathway map showing central carbon metabolism (map01200), KEGG Ontologies (KO) and Enzyme Commission numbers (E.C.) identified by 3D-approach of activated sludge of winter (W) and summer (S) period. Green-colored edges indicate identified KOs from sample W_1, red colored edges sample S_1, and blue colored edges KOs present in both samples.

**Supplementary Figure S1.** KEGG pathway map showing nitrogen metabolism (map00910), KEGG Ontologies (KO), and Enzyme Commission numbers (E.C.) identified by 3D-approach of activated sludge of winter (W) and summer (S) period. All enzymes identified in sample W_1 were marked with green color, all enzymes in sample S_1 with red, respectively.
Supplementary Table S1. Process data acquired from afflux and drain of biological treatment step.

<table>
<thead>
<tr>
<th>Process parameter</th>
<th>Winter</th>
<th>Summer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Afflux (01/21/2014)</td>
<td>Drain (06/02/2014)</td>
</tr>
<tr>
<td>Afflux volume [m³]</td>
<td>100,347</td>
<td>94,469</td>
</tr>
<tr>
<td>Temperature [°C]</td>
<td>12.1</td>
<td>15.1</td>
</tr>
<tr>
<td>pH value [-]</td>
<td>7.89</td>
<td>7.62</td>
</tr>
<tr>
<td>Oxygen concentration [mg/L]</td>
<td>-</td>
<td>2.6</td>
</tr>
<tr>
<td>Chemical oxygen demand [mg/L]</td>
<td>685.0</td>
<td>269.0</td>
</tr>
<tr>
<td>Biological oxygen demand [mg/(L*5d)]</td>
<td>241</td>
<td>208</td>
</tr>
<tr>
<td>Phosphorus concentration [mg/L]</td>
<td>13.10</td>
<td>7.52</td>
</tr>
<tr>
<td>Total nitrogen concentration [mg/L]</td>
<td>96.30</td>
<td>34.90</td>
</tr>
<tr>
<td>Ammonium concentration [mg/L]</td>
<td>68.80</td>
<td>30.00</td>
</tr>
<tr>
<td>Nitrate concentration [mg/L]</td>
<td>0.45</td>
<td>0.23</td>
</tr>
<tr>
<td>Nitrite concentration [mg/L]</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Dry matter [g/L]</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Sludge activity [mg (Oxygen/ Dry matter)]</td>
<td>4.40</td>
<td>2.26</td>
</tr>
<tr>
<td>Sludge content [mL Sludge/L]</td>
<td>103</td>
<td>100</td>
</tr>
<tr>
<td>Sludge volume index [mL Sludge/ Dry matter]</td>
<td>9.86</td>
<td>4.22</td>
</tr>
<tr>
<td>Total acid content [mmol/L]</td>
<td>4.22</td>
<td>4.74</td>
</tr>
</tbody>
</table>

1: Afflux data recording is carried out two-daily, therefore process parameters shown in this column are taken from January 20th.