1 **Fractionation of biogas plant sludge material improves metaproteomic characterization to**

2 **investigate metabolic activity of microbial communities**

3 "Technical Brief"

4 *Fabian Kohrs[1], Sophie Wolter[1], Dirk Benndorf[1,*], Robert Heyer[1], Marcus Hoffmann[2], Erdmann Rapp[2],*

5 *Andreas Bremges[3], Alexander Sczyrba[3], Andreas Schlüter[3], and Udo Reichl[1,2]*

6 [1] Bioprocess Engineering, Otto von Guericke University Magdeburg, Magdeburg, Germany

7 [2] Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems,

8 Magdeburg, Germany

9 [3] Center for Biotechnology, Bielefeld University, Bielefeld, Germany

10 * **Correspondence**: Dr. rer. nat. Dirk Benndorf, Bioprocess Engineering, Otto von Guericke University

11 Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany. E-Mail: benndorf@mpi-

12 magdeburg.mpg.de. Fax: ++49 / 391 / 67-11209.

13 **Abbreviations**: **BGP**, biogas plant; **E**, *Ebendorf*; **G**, *Getec*; **U**, untreated; **P**, pellet fraction; **S**,

14 supernatant; **M**, molar; **KEGG**, Kyoto Encyclopedia of Genes and Genomes; **KO**, KEGG Orthology

15 **Keywords**: Metaproteomics / Biogas plant / Microbial communities / Differential centrifugation /

16 Metabolism

17 **2766 Words**

Abstract

With the development of high resolving mass spectrometers, metaproteomics evolved as a powerful tool to elucidate metabolic activity of microbial communities derived from full-scale biogas plants. Due to the vast complexity of these microbiomes, application of suitable fractionation methods are indispensable, but often turn out to be time and cost intense, depending on the method used for protein separation. In this study, centrifugal fractionation has been applied for fractionation of two biogas sludge samples to analyze proteins extracted from (i) crude fibers, (ii) suspended microorganisms, and (iii) secreted proteins in the supernatant using a gel-based approach followed by LC-MS/MS identification (data are available via ProteomeXchange with identifier PXD001508). This fast and easy method turned out to be beneficial to both the quality of SDS-PAGE and the identification of peptides and proteins compared to untreated samples. Additionally, a high functional metabolic pathway coverage was achieved by combining protein hits found exclusively in distinct fractions. Sample preparation using centrifugal fractionation influenced significantly the number and the types of proteins identified in the microbial metaproteomes. Thereby, comparing results from different proteomic or genomic studies, the impact of sample preparation should be considered.

34 **Text Body**

35 Metaproteomics is as an orthologous method to molecular approaches such as metagenome
36 sequencing or community profiling, and has been shown to be a suitable tool for functional analyses
37 of environmental microbiomes [1]. Extraction of microbial communities from contaminated matrices
38 [2] or crude sludge samples [3] as well as the reduction of protein complexity is of crucial importance
39 for functional analysis of metabolic pathways using metaproteomics. Since many years,
40 centrifugation has been applied for enrichment of biomass from lab scale biogas plants (BGPs) [4],
41 but not yet for full scale plants. Furthermore, sample pre-fractionation [5] increases resolution and
42 identification of peptides and proteins from complex mixtures. In this paper, the efficiency of
43 centrifugal fractionation of samples for reduction of sample complexity and the impact on
44 subsequent analyses is demonstrated. For this purpose, proteins extracted from (i) crude fibers, (ii)
45 suspended microorganisms, and (iii) secreted proteins in the supernatant were investigated using a
46 gel-based approach followed by protein identification using LC-MS/MS experiments and database
47 search.

48 In detail, 50 mL of sludge sample was taken in two replicates from two different main fermenters of
49 BGPs located near Magdeburg (Germany, *Ebendorf* ("E1", "E2") and *Getec* ("G1", "G2"); see
50 Supplementary Table S1 for process parameters) and was transported directly to the laboratory in
51 airtight containers for subsequent analysis. Untreated (suffix "_U") sludge of each sample was
52 retained as a control ("E1_U", "E2_U"; "G1_U", "G2_U"). In a first step, fractionation centrifugation
53 was applied for 3 min at 3,000 x g at 20°C to pelletize crude fibers and any associated
54 microorganisms (suffix "_P3000"). Derived supernatants were submitted to a second centrifugation
55 for 10 min at 16,400 x g at 20°C to separate suspended microorganisms in the pellet fraction (suffix
56 "_P16400") from extracellular or secreted proteins in the supernatant ("_S16400"). The respective
57 notations of all fractions are summarized in Supplementary Figure S1. In a second step, cell
58 disruption in a ball mill with simultaneous phenol extraction of proteins, subsequent protein
59 precipitation using 0.1 M ammonium acetate in methanol and washing steps using ice-cold 80%

60 acetone and 70% ethanol were performed according to Heyer and Kohrs et al. [6]. The protein
61 content of the S16400 fractions was directly measured using an amido black assay, whereas the
62 precipitated proteins of the other fractions were suspended in 7 M urea/2 M thiourea buffer
63 containing 0.01 g mL^{-1} DTT before protein staining with amido black. For subsequent analysis by 1D
64 SDS-PAGE 100 µg of protein were precipitated by acetone. Dried pellets were rinsed in 40 µL SDS
65 sample buffer, and loaded on a 1.5 mm thick gel (see [6] for detailed description). Finally, gels were
66 stained overnight using colloidal Coomassie blue dye.

67 SDS-PAGE of both samples (Figure 1 and Supplementary Figure S2) showed similar protein profiles
68 for untreated and P3000 fractions. This suggests that the protein patterns of untreated fractions
69 were dominated by proteins derived from the crude fiber fraction. The protein patterns of
70 G1/G2_P16400 were slightly changed in comparison to the P3000 fractions (e.g. in the area of 55-
71 70 kDa or ~30 kDa). Surprisingly, the protein pattern of E1/E2_P16400 contained the same dominant
72 bands as in the untreated and the P3000 fraction, but with a significantly higher intensity. The
73 supernatant fractions S16400 of both samples showed only a single distinct protein band but
74 exhibited the most intense background smear.

75 To investigate, if an impact on fraction's protein purity and a higher functional pathway coverage can
76 be confirmed by subsequent LC-MS analysis, complete PAGE lanes were cut off and chopped
77 vertically into ten equal pieces. All chemicals listed below were of LC-MS grade. Devices and columns
78 for LC-MS/MS analysis were obtained from Thermo Fisher Scientific (Bremen, Germany). In-gel
79 reduction with DTT was followed by alkylation with iodoacetamide and final digestion was performed
80 using trypsin (SERVA Electrophoresis GmbH, Heidelberg, Germany) in an enzyme to substrate ratio of
81 1:50 overnight at 37°C [6]. Peptides were extracted, dried and suspended in 11 µL chromatographic
82 liquid phase A (97.95% water, 2% ACN, 0.05% TFA) of which 5 µL were injected into the LC system
83 (UltiMate 3000 RSLCnano splitless LC system). Isocratic loading of diluted peptides on a trap column
84 (Dionex Acclaim PepMap100 C18 µ-pre-column, 5 µm particle size, 100 Å pore size, 300 µm inner
85 diameter, 5 mm length) was performed with a flow rate of 7 µL min^{-1} chromatographic liquid phase

86 A. Afterwards chromatographic separation took place on a RP column (Dionex Acclaim PepMap C18
87 RSLCnano, 2 µm particle size, 100 Å pore size, 75 µm inner diameter, 150 mm length) at 40°C and a
88 flow rate of 300 nL min^{-1}. A binary A/B-solvent gradient was applied starting at 4% of phase B for
89 4 min, running on with a linear increase to 55% B within 120 min (A: 97.9% water, 2% ACN, 0.1%
90 formic acid; B: 80% ACN, 10% trifluorethanol, 9.9% water, 0.1% formic acid). Eluting peptides were
91 directly submitted to an online coupled OrbiTrap Elite hybrid ion trap-orbitrap MS equipped with a
92 Nanospray Flex Ion Source and distal coated silica tips (SilicaTip Emitters, FS360-20-10-D-20, New
93 Objective, MA, USA). Precursor ions were scanned in the orbital trap in positive mode at a resolution
94 of 30,000 and a m/z range of 350-2,000. The 20 most intense precursors were submitted for CID
95 fragmentation and scanned in the linear trap with a mass range and scan rate referred to as "normal"
96 parameter settings. Acquired data were exported as Mascot Generic Format using Proteome
97 Discoverer 1.4 (Thermo Fisher Scientific, Bremen, Germany) without any post processing. Spectra
98 were searched using the MASCOT 2.5 search engine (Matrix Science, London, England) against three
99 BGP metagenomes (Supplementary Table S2) jointly concatenated to UniProtKB/Swiss-Prot database
100 (state 2014/10/23) with the following parameters: all species, trypsin, one missed cleavage,
101 monoisotopic mass, carbamidomethylation (cysteine) and oxidation (methionine) as variable
102 modifications, ±0.03 Da precursor and ±0.4 Da MS/MS fragment tolerance, 1 ^{13}C and +2/+3 charged
103 peptide ions. Resulting MASCOT DAT files containing peptide and protein identifications were finally
104 uploaded to the *MetaProteomeAnalyzer* software [7] (see https://code.google.com/p/meta-
105 proteome-analyzer/ for download and supporting material) with a FDR of less than 1%. To prevent
106 redundant peptide to protein assignments, all proteins that share at least one peptide were fused to
107 meta-proteins. Taxonomies of meta-proteins were derived from its peptides using their common
108 ancestor in the phylogenetic tree.

109 While the number of identified spectra was in the same range for samples U and P3000, a
110 significantly higher (up to two-fold) number of spectra were identified in fractions P16400 and
111 S16400 for both technical and biological replicates (Supplementary Table S3). Obviously, the removal

of solid non-protein compounds resulted in an improved quality of the spectra. All fractions delivered exclusively identified peptides (Supplementary Table S3). Moreover, more than half of the peptides in fractions P16400 and S16400 were exclusive. Surprisingly, only fraction P16400 showed a significantly increased number of unique peptides and meta-proteins, whereas in fraction S16400 the lowest number of unique peptides and meta-proteins from all fractions was detected. This is in accordance with observations that proteins in the extracellular space have a limited half-life in the substrate of a biogas plant [8]. Overall, the improved number of identifications obtained from LC-MS/MS demonstrated that centrifugal fractionation is beneficial for analysis of BGPs samples.

The following evaluation of the results obtained by the different fractionation methods focusses on the functions assigned to meta-proteins in different fractions. As already shown in previous studies [5], eukaryotic proteins originating from forage plants and animal residues were found. Although fractionation removed crude fibers, the proportion of plant and animal proteins in the different fractions was not decreased significantly. Probably, these substrates were already well suspended or partly degraded. However, the main objective of investigating samples from BGPs is identification of metabolic activity performed by the microbial community. Therefore, the success of centrifugal fractionation was evaluated based on the number of biological functions, i.e. carbon fixation, methane formation or conversion of metabolites, identified by LC-MS/MS. For this purpose, Kyoto Encyclopedia of Genes and Genomes (KEGG [9]) Orthologies (KO) of bacterial and archaeal meta-proteins identified together in both of the technical replicates were assigned to the KEGG pathway map of central carbon metabolism (map01200, Figure 3), and depicted in a Venn diagram [10] (Figure 2). This way KOs found exclusively in single fractions as well as overlaps for the respective pathway map could easily be examined. For sample E the amount of KOs could be increased from 45 for the unfractionated case (E_U) to 60 when investigations of sub-fractions were included. Moreover, the set of KOs identified in sample G increased from 46 (G_U) to 67 considering hits identified in both pellets and supernatant. However, no exclusive KO has been identified in fraction P3000. Accordingly, this fraction was neglected for the following analysis.

138 To illustrate the pathway coverage, all KOs were directly plotted into the central carbon map. For
139 fraction E_U, KOs could be assigned to methanogenesis, glycolysis/gluconeogenesis, pentose
140 phosphate shunt and citric acid cycle (Figure 3). A comprehensive evaluation of these pathways
141 revealed that some enzymes for methanogenesis (coenzyme M methyltransferase, K14082) or
142 archaeal pentose phosphate shunt (3-hexulose-6-phosphate synthase, K08093) and citric acid cycle
143 (malate dehydrogenase, K00025) were only identified in the cellular fraction E-P16400, whereby only
144 fumarate hydratase (K 01676) was unique in E_U. For sample G, coenzyme M methyltransferase
145 (K14082) and formylmethanofuran dehydrogenase (K00200), both involved in the hydrogenotrophic
146 methanogenesis, and the enzyme malate dehydrogenase (K00025), involved in the citrate cycle, were
147 identified exclusively in fractions P16400 and S16400 (Supplementary Figure S3). As before, a minor
148 number of KOs was exclusively detected in the unfractionated sample (e.g. ribulose-bisphosphate
149 carboxylase, K01601 and acetate kinase, K00925, both involved in carbon fixation).

150 Centrifugation of primary environmental samples is frequently applied as a first step in molecular
151 biological analyses [11]. Both, gel separation (Figure 1) and subsequent LC-MS/MS analysis
152 (Supplementary Table S3) benefit from the removal of non-protein particles prior to sample
153 preparation. Although no major impact on the taxonomic distribution of organisms in the different
154 fractions was observed here (data not shown), fractionation could bias the composition of the
155 microbial metaproteome. Accordingly, the issue of sample preparation should be paid more
156 attention, if results from different proteomic or genomic studies are compared that were acquired
157 from different fractions derived by centrifugation techniques. For future metaproteomics studies on
158 BGPs, it is strongly recommended to focus rather on the P16400 fraction than on a 'quick and dirty'
159 analysis of the unfractionated sample, independently of the fractionation procedures following
160 downstream of protein extraction.

161 In summary, centrifugal fractionation is a straightforward sample preparation technique adding less
162 than one hour to the proteomic workflow for analysis of samples from BGPs. It can be used as an
163 alternative to or in combination with other advanced fractionation protocols published recently [5]

because it fractionates the sample before lysis and protein extraction. In particular, the resolution of explorative metaproteome studies in BGPs is increased by yielding more and exclusive identifications on spectra and peptide level. Finally, it allows higher pathway coverage (Figure 3) and thus facilitates interpretation of functional microbiomes.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD001508.

The Authors would like to thank Claudia Andres (GETEC Green Energy AG, Magdeburg, Germany) and Matthias Neuss (ABO Wind, Wiesbaden, Germany) for providing BGP samples and related process data. Robert Heyer was supported by the German Environmental Foundation (DBU), grant number 20011/136, and Marcus Hoffmann was supported by HighGlycan, grant number 278535. They also thank the PRIDE team for technical assistance in uploading the data.

The authors have declared no conflict of interest.

177 **References**

178 [1] Wilmes, P., Bond, P. L., The application of two-dimensional polyacrylamide gel electrophoresis
179 and downstream analyses to a mixed community of prokaryotic microorganisms. Environ Microbiol
180 2004, 6, 911–920.
181 [2] Benndorf, D., Balcke, G. U., Harms, H., Bergen, M. von, Functional metaproteome analysis of
182 protein extracts from contaminated soil and groundwater. ISME J 2007, 1, 224–234.
183 [3] Hanreich, A., Heyer, R., Benndorf, D., Rapp, E. et al., Metaproteome analysis to determine the
184 metabolically active part of a thermophilic microbial community producing biogas from agricultural
185 biomass. Can. J. Microbiol. 2012, 58, 917–922.
186 [4] Wöhlbrand, L., Trautwein, K., Rabus, R., Proteomic tools for environmental microbiology - a
187 roadmap from sample preparation to protein identification and quantification. Proteomics 2013, n/a.
188 [5] Kohrs, F., Heyer, R., Magnussen, A., Benndorf, D. et al., Sample prefractionation with liquid
189 isoelectric focusing enables in depth microbial metaproteome analysis of mesophilic and
190 thermophilic biogas plants. Anaerobe 2014, 29, 59–67.
191 [6] Heyer, R., Kohrs, F., Benndorf, D., Rapp, E. et al., Metaproteome analysis of the microbial
192 communities in agricultural biogas plants. New Biotechnology 2013, 30, 614–622.
193 [7] Muth, T., Behne, A., Heyer, R., Kohrs, F. et al., The MetaProteomeAnalyzer: A Powerful Open-
194 Source Software Suite for Metaproteomics Data Analysis and Interpretation. J Proteome Res. 2015,
195 14(3):1557-65.
196 [8] Binner, R., Menath, V., Huber, H., Thomm, M. et al., Comparative study of stability and half-life of
197 enzymes and enzyme aggregates implemented in anaerobic biogas processes. Biomass Conv. Bioref.
198 (2011) 1:1–8.
199 [9] Kanehisa, M., Goto, S., KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research
200 2000, 28, 27–30.
201 [10] Oliveros, J. C., VENNY. An interactive tool for comparing lists with Venn Diagrams, 2007.
202 http://bioinfogp.cnb.csic.es/tools/venny/index.html.

203 [11] Klocke, M., Mähnert, P., Mundt, K., Souidi, K. et al., Microbial community analysis of a biogas-
204 producing completely stirred tank reactor fed continuously with fodder beet silage as mono-
205 substrate. Systematic and Applied Microbiology 2007, 30, 139–151.

206 **Figure Legends**

207 **Figure 1.** *Coomassie blue stained 1D gels of fractionated BGP samples. 100 µg of respective fractions*
208 *were loaded next to a molecular weight standard. (A) BGP Ebendorf "E1". (B) BGP Getec "G1".*

209 **Figure 2.** *Venn diagrams showing KEGG Orthologies (KO) assigned to the central carbon metabolism*
210 *(map01200) for all sample fractions of the BGPs Ebendorf and Getec. Diagrams were created using*
211 *VENNY.*

212 **Figure 3.** *KEGG pathway map showing the central carbon metabolism (map01200) for samples of the*
213 *biogas plant Ebendorf. Blue: common KEGG Orthologies (KO) of the unfractionated sample (E_U),*
214 *pellet after centrifugation at 16,400 x g (E_P16400) and the supernatant (E_S16400); Red: KOs*
215 *identified exclusively in E_U; Green: KOs further identified in E_P16400; Yellow: KOs identified*
216 *exclusively in E_S16400.*

217 **Supplementary Figure S1.** *Schematic workflow for centrifugal fractionation and sample notation.*
218 *Untreated sample (U) was centrifuged at 3.000 x g to obtain a pellet fraction (P3000) containing*
219 *substrate fibers and attached biomass. The supernatant, containing suspended microorganisms or*
220 *secreted enzymes, was submitted to a second centrifugation step at 16.400 x g. The derived pellet*
221 *(P16400) containing microorganisms was separated from the supernatant (S16400) containing*
222 *extracellular proteins.*

223 **Supplementary Figure S2.** *Replicates of Coomassie blue stained 1D gels of fractionated BGP samples.*
224 *100 µg of respective fractions were loaded next to a molecular weight standard. (A) BGP Ebendorf*
225 *"E2". (B) BGP Getec "G2".*

226 **Supplementary Figure S3.** *KEGG pathway map showing central carbon metabolism (map01200) for*
227 *sample Getec. Blue: common KEGG Orthologies (KO) of the unfractionated sample (G_U), pellet after*
228 *centrifugation at 16,400 x g (G_P16400) and the supernatant (G_S16400); Red: KOs identified*
229 *exclusively in G_U; Green: KOs further identified in G_P16400; Yellow: KOs identified exclusively in*
230 *G_S16400.*

231 **Supplementary Material**

232 **Supplementary Table S1.** *General process data and substrate composition of BGPs investigated.*

Biogas plant	Ebendorf	Getec
Process temperature	40°C	40°C
Reactor fill level	3,000 m^3	3,728 m^3
Feed rate	80 tons d^{-1}	102 tons d^{-1}
Substrate composition	40% corn silage, 40% turkey dung, 20% sweet millet silage	67% corn silage 16% sugar beet puree 9% pressed sugar beet shred 8% turkey dunk
pH	8.9	7.75
Volatile fatty acids to total alkalinity	0.2	(Not analyzed)
Acidic equivalent content	(Not analyzed)	0.393 g L^{-1}
Totals gas production	950 m^3 h^{-1}	420 m^3 h^{-1}
Methane content	56%	53%
Ammonia concentration	6.0 g L^{-1}	2.2 g L^{-1}

233

234 **Supplementary Table S2.** *Assembly of public metagenomes used to identify LC-MS/MS spectra.*

Reference	Description
Rademacher et al. 2012 doi:10.1111/j.1574-6941.2011.01265.x	Metagenome derived from a two-stage leach-bed fermenter (100 L hydrolysis reactor and 30 L anaerobic filter), fed with rye silage and winter barley, operated at 55°C, sampled after 21 days.
Hanreich et al. 2013 doi:10.1016/j.syapm.2013.03.006	Metagenome derived from 500 mL batch reactors inoculated with sludge of a full-scale maize-fermenting BGP, fed with cut straw and hay, operated at 38°C, sampled after five and 30 days.
Bremges et al. 2015 (submitted) Study accession PRJEB8813	Metagenome derived from a full-scale BGP. The metagenomic dataset supporting the results of this article is provided by the Center for Biotechnology, University Bielefeld, Bielefeld, Germany, and available under sample accession numbers ERS697688- ERS697693 (www.ebi.ac.uk/ena/data/view/PRJEB8813).

235

236 **Supplementary Table S3.** *General statistics of samples after LC-MS/MS analysis and data processing[1].*

	Identified Spectra	Peptides[2]	Unique Peptides[3]	Meta-proteins
E1_U	11,671	3,257	1,942	1,396
E1_P3000	13,856	3,912 (34.9%)	2,408	1,598
E1_P16400	22,666	5,245 (60.1%)	3,187	1,808
E1_S16400	21,579	2,680 (54.9%)	1,647	917
E2_U	8,781	1,952	1,166	907
E2_P3000	9,019	1,913 (28.3%)	1,114	870
E2_P16400	23,453	3,903 (63.9%)	2,402	1,344
E2_S16400	24,524	2,499 (65.1%)	1,558	812
G1_U	14,860	2,756	1,627	1,240
G1_P3000	16,207	2,872 (33.7%)	1,743	1,255
G1_P16400	26,853	3,983 (55.0%)	2,418	1,431
G1_S16400	45,148	2,950 (60.9%)	1,885	957
G2_U	14,095	2,921	1,770	1,171
G2_P3000	18,026	3,003 (37.8%)	1,742	1,005
G2_P16400	25,754	4,732 (55.0%)	2,803	1,594
G2_S16400	21,508	2,557 (54.7%)	1,631	803

237 1: Using *MetaProteomeAnalyzer* software [7].

238 2: In brackets: percentage of peptides found exclusively in the respective fraction but not in the unfractionated sample.

239 3: Peptides non-redundantly assigned to only one protein hit. For details, please refer to [7].