microbial biotechnology

Minireview

Metaproteomics of complex microbial communities in biogas plants

Robert Heyer,[1,2] Fabian Kohrs,[1,2] Udo Reichl[1,2] and Dirk Benndorf[1,2]*

[1]*Bioprocess Engineering*, Otto von Guericke University Magdeburg, Universitätsplatz 2, Magdeburg 39106, Germany.
[2]*Max Planck Institute for Dynamics of Complex Technical Systems*, Sandtorstr. 1, Magdeburg 39106, Germany.

Summary

Production of biogas from agricultural biomass or organic wastes is an important source of renewable energy. Although thousands of biogas plants (BGPs) are operating in Germany, there is still a significant potential to improve yields, e.g. from fibrous substrates. In addition, process stability should be optimized. Besides evaluating technical measures, improving our understanding of microbial communities involved into the biogas process is considered as key issue to achieve both goals. Microscopic and genetic approaches to analyse community composition provide valuable experimental data, but fail to detect presence of enzymes and overall metabolic activity of microbial communities. Therefore, metaproteomics can significantly contribute to elucidate critical steps in the conversion of biomass to methane as it delivers combined functional and phylogenetic data. Although metaproteomics analyses are challenged by sample impurities, sample complexity and redundant protein identification, and are still limited by the availability of genome sequences, recent studies have shown promising results. In the following, the workflow and potential pitfalls for metaproteomics of samples from full-scale BGP are discussed. In addition, the value of metaproteomics to contribute to the further advancement of microbial ecology is evaluated. Finally, synergistic effects expected when metaproteomics is combined with advanced imaging techniques, metagenomics, metatranscriptomics and metabolomics are addressed.

Introduction

Over the past 10 years, conversion of biomass to methane in biogas plants (BGPs) has become a reliable source of renewable energy. In 2013, about 7500 BGPs produced 3.5% of the annual electricity demand in Germany (Fachagentur Nachwachsende Rohstoffe e.V. (FNR, 2013). In contrast to burning the biomass, the main advantage of biogas production is the possibility to utilize substrates with high water content.

In the future, the importance of biogas process might even grow, because it could be used for energy storage by biological methanation (Luo *et al.*, 2012; Bensmann *et al.*, 2014) or for anaerobic treatment of wastewater (Angelidaki *et al.*, 2011). During the biogas process, a complex microbial community degrades biomass or organic waste including crop silage, dung, manure, sludge from wastewater treatment plants, household garbage or waste from food industry to methane. In the first step, polymeric substrates are hydrolysed to monomers by extracellular enzymes released by primary fermenters, i.e. *Clostridium thermocellum* or *Caldicellulosiruptor saccharolyticus*. Afterwards, primary fermenters such as *Clostridium acetobutylicum* convert monomers to hydrogen, carbon dioxide, short-chain fatty acids and primary alcohols. During the subsequent acetogenesis, secondary fermenters including *Syntrophomonas wolfei* metabolize primary alcohols and short-chain fatty acids to hydrogen, carbon dioxide and acetate. The released hydrogen is captured by homoacetogenic Bacteria and hydrogenotrophic methanogens such as *Acetobacterium woodii* resp. *Methanothermobacter*

thermoautotrophicus. This syntrophic interaction enables the secondary fermenters to gain energy under thermodynamically unfavourable conditions. Finally, acetoclastic methanogens, i.e. *Methanosarcina barkeri*, consume acetate and convert it to methane and carbon dioxide. For a continuous high-yield biogas production, all of the metabolic pathways of these four main steps of the biogas process have to be finely tuned.

Previous attempts to optimize biogas production focused on the impact of physicochemical and technical process parameters on performance of BGPs (Appels *et al.*, 2008; Ward *et al.*, 2008; Holm-Nielsen *et al.*, 2009; Weiland, 2010; Angelidaki *et al.*, 2011). However, several problems still impair the optimal conversion of biomass to methane (Ward *et al.*, 2008).

First, microbial communities degrade only 30–60% of the fed biomass (Angelidaki *et al.*, 2011), because lignin and cellulose are resistant to hydrolysis. In contrast, the microbial communities in the gut of sheep (Toyoda *et al.*, 2009) or termites (Burnum *et al.*, 2011) are able to utilize both lignin- and cellulose-rich grass and wood with high efficiency.

Second, in order to avoid process disturbances, a lot of energy and effort is spent to adjust optimal process conditions for the microbial community, e.g. optimal ammonia concentrations (Appels *et al.*, 2008) or stable process temperatures. However, microbial communities are able to adapt to challenging process conditions, partially questioning these efforts (Chen *et al.*, 2008).

Third, dynamic operation of BGPs would be favourable to produce electricity on demand and to stabilize the electric grid, but is rarely applied due to risk of acidification (Munk *et al.*, 2010) and missing control strategies. Dynamic operation could be applied by different feeding amounts easily, but more detailed knowledge about the metabolic limits of the microbial communities is required.

In summary, a lack of understanding concerning the composition and performance of the microbial community hinders further optimization of the biogas process (Weiland, 2010). Therefore, improvement of space–time yield of BPGs requires the clarification of the following three key questions of applied microbial ecology: (i) who is there, (ii) who is doing what with whom and (iii) how can we adjust initial conditions and control the composition of the microbial communities as previously suggested by Verstraete and colleagues (2007).

In order to answer these questions, different approaches namely microscopy, metagenomics, metatranscriptomics, metaproteomics and metabolomics are available. In particular, metaproteomics, targeting the identification of proteins/enzymes from the individual species of the microbial community, represent a promising approach. The main advantage of metaproteomics is the possibility to link the function of proteins with a certain taxonomy and to correlated their presence with metabolic activity (Wilmes and Bond, 2006). However, metaproteomics is challenged by four major problems (Muth *et al.*, 2013): (i) contamination by products of biomass degradation, (ii) sample complexity, (iii) redundant protein identifications and (iv) lack of detailed databases. In this paper, the value of metaproteomics for analyses of BGPs is discussed, and an optimized workflow comprising sampling, protein purification, separation, mass spectrometry (MS), bioinformatics and result evaluation is described.

Tools for the characterization of microbial communities

For the analysis of complex microbial communities, a wide spectrum of elaborated methods is available as shown in Table 1. Besides the characterization of genes, mRNAs, proteins and metabolites using dedicated assays, microscopic analysis of microorganisms is also a valuable option. Due to different targets, the methods provide different levels of information concerning the spatial organization, the taxonomic composition as well as the function and metabolic activity of the individual microbial species.

Microscopy is a well-known technology used to investigate the organization of microbial communities regarding abundance and spatial distribution (Grotenhuis *et al.*, 1991). However, most microorganisms cannot be classified by morphology alone. Nevertheless, in BGPs, the F420 cofactor (Heine-Dobbernack *et al.*, 1988) is involved in methanogenesis and shows an intrinsic fluorescence allowing specific detection of hydrogenotrophic methanogens. For further differentiation, specific staining methods such as fluorescence in situ hybridization (FISH) can be used (Sekiguchi *et al.*, 1999; Nettmann *et al.*, 2010). Nevertheless, strong background fluorescence from sample impurities, i.e. humic and fulvic acids (Senesi *et al.*, 1989), often interferes with staining procedures (Hofman-Bang *et al.*, 2003; Bastida *et al.*, 2009). In addition to microscopy, flow cytometry can be applied to discriminate between individual strains and to follow dynamics of microbial communities (Müller *et al.*, 2012). Molecular biological analysis of genes or mRNA is a more robust and precise method for the phylogenetic or functional characterization (Hofman-Bang *et al.*, 2003; Klocke *et al.*, 2008; Nelson *et al.*, 2011; Ziganshin *et al.*, 2013). In particular, the presence of 16S rRNA genes is frequently used for phylogenetic studies (Amann *et al.*, 1995). The presence of functional genes or corresponding mRNA is utilized for measuring the functional diversity. Due to its low stability, the presence of mRNA is a good indicator of gene expression. While the analysis of RNA requires a

Table 1. Tools for the characterization of microbial communities.

Method	Target	Spatial organization	Taxonomy	Function	Metabolic activity	Analysed parameters per run[a]	Supplements metaproteomics
microscopy	microorganisms	+	±	–	–	1 sample	indicates successful cell lysis
flow cytometry	microorganisms	–	±	–	–	1 sample/1–3 stainings	
FISH microscopy	microorganisms	+	+	±	±	1 sample/1–3 stainings	
TRFLP/DGGE	genes/mRNA	–	±	±	–	1 sample/1 gene	
TRFLP/DGGE + clone library	genes/mRNA	–	+	±	–	1 sample/1 gene	
metagenome sequencing	genes	–	+	+	–	≈10,000 contigs	database for metaproteomics
metatranscriptome sequencing	mRNA	–	+	+	±	≈10,000 contigs	database for metaproteomics
metaproteomics	proteins	–	+	+	±	≈1,000 proteins	Re-annotation of genes by proteogenomics
metabolomics	intermediates	–	–	±	+	≈10–20 intermediates	proves activity of proteins
enzyme activity assays	enzymes	–	–	±	+	1 enzyme	activity values for genes/proteins

a. Numbers of analysed parameters per run are estimated. Actual numbers depend on the experimental setup.
Comparison of standard methods for the investigation of microbial communities, concerning its target, effort, price as well as the type and amount of information obtained (–, no information; ±, qualitative information; +, quantitative information). The evaluation of these methods was done to the best of our knowledge and refers to the number of analysed parameters per run. However, only a broad overview about available methods can be given within the scope of this review.
DGGE, denaturing gradient gel electrophoresis; TRFLP, terminal restriction fragment length polymorphism.

previous reversed transcription to cDNA, DNA is directly amplified by polymerase chain reaction (PCR) with primers specific to phylogenetic groups or selected functional genes. Afterwards, the equally sized PCR products are separated by denaturing gradient gel electrophoresis (DGGE) or terminal restriction fragment length polymorphism (TRFL) revealing the fingerprint. In the case of 16S rRNA based community analysis, the individual microorganisms can be identified by a clone library and the actual community profile can be generated by normalization with the species-specific abundance of the 16S rRNA gene (Klappenbach *et al.*, 2001). In BGPs, the functional analysis was successfully applied for quantification of the methyl CoM reductase gene and its mRNA (Munk *et al.*, 2012).

The development of 454 pyrosequencing (Margulies *et al.*, 2005) and illumina sequencing (Bentley *et al.*, 2008) enabled the investigation of the whole metagenome/ metatranscriptome of microbial communities instead of analysing single genes or individual mRNA (Schlüter *et al.*, 2008; Wirth *et al.*, 2012; Zakrzewski *et al.*, 2012). In contrast to a metagenome, representing the genetic potential of a community, the metatranscriptome is a snapshot of the actual gene expression.

However, final metabolic activity is determined, among other factors, by the concentrations of proteins, which are strongly influenced by their half-life periods. Thus, a better description of the metabolic function of microbial communities is expected from the abundance of the

microbial enzymes and proteins [metaproteome (Wilmes and Bond, 2006)]. Most proteomic approaches, however, are performed under denaturing conditions and therefore provide only information regarding the abundance of proteins instead of enzyme activities. In addition, the latter are influenced by temperature, pH value as well as on the concentration of substrates and products. Therefore, enzyme activity assays, e.g. enzymes of hydrolysis (Gasch *et al.*, 2013) or enzymes of methanogenesis (Refai *et al.*, 2014), were also established for analysis of microbial communities from BGPs and could be applied to confirm metaproteome data. Alternatively, concentrations of intracellular and extracellular metabolites (metabolome) could be determined as they also represent the microbial activity. The currently performed routine sampling of full-scale BGPs clearly provides only a minimum of information regarding the metabolic activity of microbial communities, e.g. the composition and amount of the substrates, the volume and composition of the gas, and the concentration of the short-chain fatty acids in the digestate (Hill and Holmberg, 1988). However, without expensive labelling of substrates with stable isotopes, metabolites cannot be assigned to phylogenetic groups.

Obviously, each method for microbial community characterization has its own advantages. Only a combination of different methods will allow to draw a more realistic picture of the microbial conversion of biomass to methane, and to derive successful measures for process

design and further optimization. The problems and limits of the individual methods should be carefully considered, as done in the review of Hofman-Bang and colleagues (2003) for molecular biological methods.

Metaproteomic workflow

The implementation of metaproteomics approaches covering a wide range of BGP samples requires considering several key challenges: (i) experimental design, (ii) sampling, (iii) protein purification, (iv) protein separation, (v) liquid chromatography (LC)-tandem mass spectrometry MS/MS), (vi) bioinformatics and (vii) examining the protein identification (Figure 1).

Reproducible scientific studies require carefully planned and documented experiments. In order to correlate metaproteome data from different BGPs with process data, at least a minimum set of meta-information has to be provided, comprising BGP design, process temperature, pH value, nitrogen content, inocula, gas composition and volume as well as feed composition. In addition, technical and biological replicates are required. However, for most industrial scale BGPs, no true replicates exist, because each BGP is individual concerning its construction and operational parameters. An acceptable workaround would be to sample at least two independent technical replicates at close time points. Otherwise, investigations using lab-scale equipment are required to complement studies. Here, critical process conditions can be applied without risking the crash of a full-scale fermenter. Depending on the scientific question, simplifying the complex microbial community by feeding a synthetic medium (Wilmes *et al.*, 2008; Abram *et al.*, 2009) or the prior use of synthetic communities could be useful (Laube and Martin, 1981; Tatton *et al.*, 1989; Scholten and Conrad, 2000; Plugge *et al.*, 2010)

A key issue for metaproteome studies is sample complexity. Deeper insight into the metaproteome can be gained through a combination of orthogonal separation steps as shown by Kohrs and colleagues (2014). However, higher resolution requires significantly higher experimental effort. Researchers should consider this before initiating a comprehensive study in which insight into the metaproteome is often indispensable to validate research hypotheses. As long as a sufficient number of representative samples are retained, a more comprehensive metaproteomics analysis or the sequencing of the corresponding metagenome can be carried out.

Sampling is quite straightforward and representative as long as the BGP is well mixed, the dead volume of the sampling tube is discarded and the samples are frozen immediately. When sampling full-scale BGPs, the following issues have to be considered: (i) sampling before feeding and at same time of the day, (ii) mixing the BGP

before sampling and (iii) discarding sufficient material before sampling in order to flush the sample port.

Sample preparation includes cell lysis, protein extraction, protein quantification and separation. Main problems during cell lysis and protein extraction are the high amount of sample impurities and the different levels of microbial community organization, such as scattered microorganisms, biofilms on the substrates or granules (Hofman-Bang *et al.*, 2003). Consequently, robust lysis of all cells and removal of as many contaminants as possible is required. Phenol extraction followed by ammonium acetate in methanol precipitation sometimes combined with cell lysis in a ball mill has already been successfully applied to characterize samples from activated sludge (Kuhn *et al.*, 2011), soil (Benndorf *et al.*, 2007) and BGPs (Heyer *et al.*, 2013). Phenol extraction separates proteins and humic substances and is essential when extracting proteins from full-scale BGPs. For lab-scale fermenters fed with synthetic media, cell lysis with ultrasonic sound and separation of debris from proteins by centrifugation is sufficient (Abram *et al.*, 2009). Subsequent dissolution of precipitated proteins in buffers, especially after phenol extraction, is difficult but high molar urea buffers delivered good results (Keiblinger *et al.*, 2012; Heyer *et al.*, 2013). However, the use of high molar urea buffer or the presence of remaining humic substances may interfere with standard protein assays (Kuhn *et al.*, 2011), namely Bradford (Bradford, 1976), Lowry (Lowry *et al.*, 1951) and BCA (Smith *et al.*, 1985) assays. In contrast, acceptable protein quantification can be achieved by using the amido black assay (Popov *et al.*, 1975; Schweikl *et al.*, 1989; Hanreich *et al.*, 2013) or by quantification of protein intensities in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Due to the high sample complexity, the proteins have to be separated prior to MS analysis. Common approaches are protein separation according to molecular weight or isoelectric point, e.g. SDS-PAGE (Laemmli, 1970) or two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (Klose, 1975; O'Farrell, 1975). Subsequent protein fractions are tryptically digested in-gel into peptides (Shevchenko *et al.*, 1996). A complete gel-free approach involving separation of tryptic peptides by one or higher dimensional LC appears feasible (Link *et al.*, 1999; Washburn *et al.*, 2001; Wolters *et al.*, 2001). However, running the proteins through a SDS-gel without separation and subsequent in-gel digest is useful because sample impurities remain in the gels. Furthermore, such a step prevents clogging of the columns and capillaries of the chromatographic system (Kohrs *et al.*, 2014). For routinely metaproteomics, SDS-PAGE procedures achieved best results since remaining impurities from environmental samples seem to hinder reproducible separation of
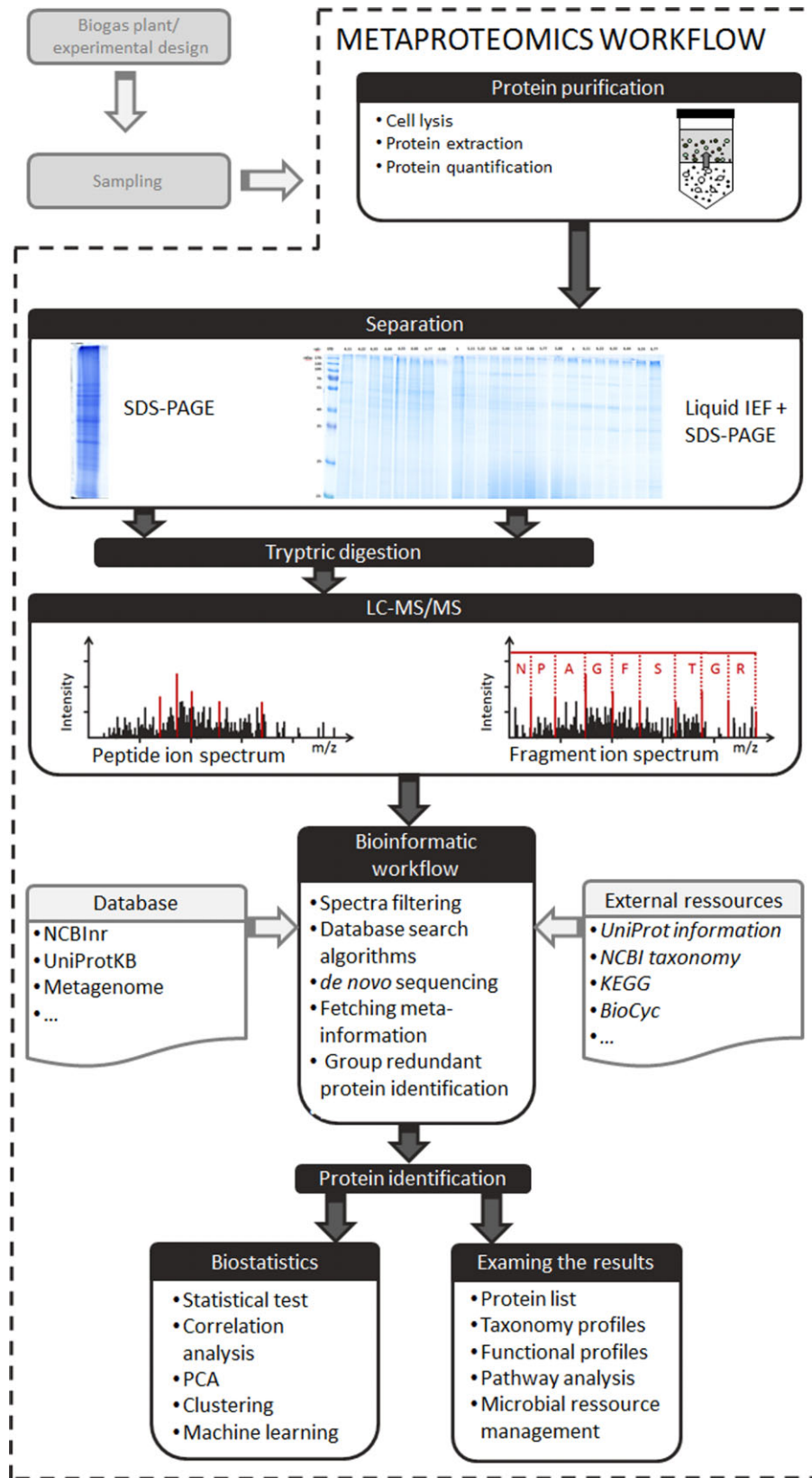
Fig. 1. Metaproteomics workflow comprising sampling, protein purification, separation, mass spectrometry, bioinformatic workflow and result evaluation.

proteins by 2D-PAGE. For improving resolution of metaproteomics, liquid-isoelectric focusing (IEF) can be carried out prior to SDS-PAGE (Kohrs *et al.*, 2014). Alternatively, ultracentrifugation can be used to separate the cellular and the extracellular fractions of proteins (Binner *et al.*, 2011).

MS, often combined with LC, is the standard approach for protein identification. In MS, peptides are ionized and separated according to their mass-to-charge ratio. In order to distinguish between peptides with identical amino acid composition but different amino acid order, peptide ions are further fragmented. For protein identification, these fragment spectra are compared against a protein database. A complete overview about MS techniques can be found in Wöhlbrand and colleagues (2013).

For samples with high complexity containing a large number of peptides, separation of peptides by LC and high-resolution MS are essential. Nevertheless, the probability that peptides with similar mass-to-charge ratio coelute from the LC systems increases drastically. Finally, the common fragmentation results in low-quality spectra that fail in a database search. Worse, the number of peptides with different mass-to-charge ratios is so high that, due to the limited scan and separation speed of the MS, only 5–30 of the most abundant peptide ions can be analysed in one cycle. Although certain rules are applied to carefully select peptides for fragmentation, this selection is still often random due to the high number of peptides. Therefore, the reproducibility between such LC-MS/MS experiments is low (Tabb *et al.*, 2010). Running technical replicates in LC-MS/MS and extending the LC gradients are appropriate strategies to manage this problem. Besides protein identification, quantification of at least key proteins is often important for the characterization of microbial communities. Common quantification strategies include chemical labelling, isotopic labelling, label-free quantification as well as quantification of protein intensity in gels. Signals from fluorescence labels, often used in gel-based approaches, can be disturbed by intrinsic fluorescence of humic-like sample impurities (Li *et al.*, 2004). In addition, the remaining humic compounds can also react with established chemical labels (Gygi *et al.*, 1999; Lottspeich and Kellermann, 2011). Due to this uncertainty, label-free quantification by peptide respective spectra abundance remains the last option. However, normalization of abundances is beneficial and the exponential modified protein abundance index (Ishihama *et al.*, 2005) or the normalized spectra abundance factor (Zybailov *et al.*, 2007) are frequently applied. Another promising quantification approach is the metabolic labelling with isotopically labelled substrates (Jehmlich *et al.*, 2008). Incorporation of stable isotopes into proteins can be monitored by MS and allows to draw conclusions about metabolic activity in the microbial community. However, the application is restricted to microcosm experiments due to the high costs for fully labelled substrates.

Routinely, peptide and protein identification are carried out by comparison of fragment spectra against theoretical spectra from a database by algorithms, e.g. Mascot (Perkins *et al.*, 1999) or X!Tandem (Craig and Beavis, 2004). Standard databases for protein identification are NCBInr (Acland *et al.*, 2014), UniProtKB/Swiss-Prot or UniProtKB/TrEMBL (Consortium, 2012). With respect to metaproteomics, more specific databases or searches against metagenomes from the same or similar samples [e.g. for BGP samples (Schlüter *et al.*, 2008; Rademacher *et al.*, 2012; Wirth *et al.*, 2012; Zakrzewski *et al.*, 2012)] resulted in the identification of more proteins and are strongly recommended.

Raw data contains many low-quality spectra (Muth *et al.*, 2013). During preprocessing, low-quality spectra can be removed without any significant loss of information (Ma *et al.*, 2011). Recently, preprocessing and data handling was embedded into complete bioinformatic platforms, e.g. OpenMS (Sturm *et al.*, 2008), Proteome Discoverer or ProteinScape (Thiele *et al.*, 2010).

Besides a probability-based score as a measure for correctness of peptide identification, the false discovery rate (FDR) evolved in the proteomic community as a standard (Elias *et al.*, 2005). FDR is mainly influenced by database size, e.g. a doubling of the database size doubles more or less the probability of false positive hits and thus doubles the FDR. Therefore, searching against large databases can cause the removal of valuable hits in order to reach a low FDR (e.g. less than 5%).

The problem of metaproteomics in contrast to other proteomics with pure or defined mixed cultures is that the taxonomic composition of complex microbial communities is not known and the database cannot be reduced to keep the number of false positive hits low. In this context, the idea of Jagtap and colleagues (2013) to repeat the search in a qualified database with reduced size containing only sequences from species identified in a first search round seems to be an option. The number of false positive hits decreases and consequently more spectra are regarded as correctly identified. However, this strategy may lead to an underestimation of the FDR and mask the lack of suitable database entries for many microorganisms due to cultivation problems (Amann *et al.*, 1995) .

In the next step, protein identification is achieved based on identified peptides (Bradshaw *et al.*, 2006). Although identifications based on two peptide per protein are favoured in peer-reviewed journals, the so-called 'single hit wonders' are not necessarily worse. In fact, identifications based on single peptides are also accepted when using high-resolution MS because the quality of the peptide identification is also considered important (Gupta

and Pevzner, 2009). Nevertheless, reliability and number of correctly identified peptides and proteins can be increased by combining multiple algorithms (Ma *et al.*, 2011; Vaudel *et al.*, 2011). Even the best algorithm can only identify proteins whose sequence is covered in a database. An approach to overcome this pitfall is *de novo* sequencing of peptides using acquired spectra (Frank and Pevzner, 2005), and to search for homologue proteins using the MS-driven basic local alignment search tool (MS-BLAST) (Shevchenko *et al.*, 2001). However, the evaluation of *de novo* results requires manual inspection. Therefore, a more straightforward strategy is sequencing the metagenome of the analysed sample.

After successful protein identification, the importance of a single identification can be improved by acquiring meta-information concerning taxonomy and function from repositories, e.g. UniProt (Consortium, 2012). Moreover, redundant protein identifications due to similar peptides from homologue proteins can be grouped based on a similar peptide sets (Schneider *et al.*, 2012), one shared peptide (Kohrs *et al.*, 2014; Lü *et al.*, 2014) or by sequence similarity [e.g. UniRef-Cluster (Suzek *et al.*, 2007)] to so called metaproteins (Muth *et al.*, 2015). Finally, protein taxonomy can be redefined by the common ancestor taxonomy of all proteins in a group (Huson *et al.*, 2007). It allows a reliable phylogenetic assignment of metaproteins avoiding risky assignments on species or strains.

For a better survey, taxonomic composition can be visualized in a Krona plot (e.g. Krona plot for a mesophilic/thermophilic BGP: Figure 2; Fig. S1) (Ondov *et al.*, 2011) that is based on identified peptides or spectra, and National Center for Biotechnology Information taxonomy (Acland *et al.*, 2014). For comparison of taxonomic profiles from different samples or time points, the richness of species, their community organization and their dynamics can be calculated, as extensively discussed in the concept of Microbial Resource Management (Verstraete *et al.*, 2007; Marzorati *et al.*, 2008; Wittebolle *et al.*, 2009).

Shifting to protein functions, overview plots, such as a Voronoi Treemap (Bernhardt *et al.*, 2013) or a common pie chart, based on gene ontologies (Ashburner *et al.*, 2000) or UniProt Keywords (Consortium, 2012) are beneficial. Even more important is the assignment of identified proteins to biochemical pathways. A straightforward mapping to MetaCyc pathways (Caspi *et al.*, 2014) or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Figure 3, Fig. S2) (Kanehisa and Goto, 2000) can be achieved using KEGG ontologies or enzyme commission numbers (Bairoch, 2000). Often, proteome studies result in long lists of upregulated and downregulated proteins confirmed by statistical tests (Karp and Lilley, 2007). For better exploitation of data,

correlation analysis between taxa, functions or process parameters can reveal unexpected functional relationships improving the knowledge about the microbial community. Moreover, differences between BGPs can be monitored by principal component analysis or cluster analysis of protein or taxonomic profiles, e.g. cluster of different BGPs based on SDS-PAGE profiles (Heyer *et al.*, 2013). In the future, the use of machine learning algorithms (Kelchtermans *et al.*, 2014) might result in further improvements.

In microbial ecology, a wide mixture of different methods are commonly applied to investigate a specific problem. Hence, knowledge of how to combine these methods is important. First of all, a standardization of sample preparation is essential. With regard to the use of multi-omic approaches, the sample preparation workflow should comprise most analytes namely DNA, RNA, proteins and metabolites in an adequate manner (Roume *et al.*, 2013). As previously discussed in *Tools for the characterization of microbial communities* (Table 1), metaproteomics delivers thorough information about taxonomy and function while conclusions regarding community organization and metabolic activity can only be obtained to a limited degree. Further insight in microbial communities could be gained by combining data from advanced microscopy approaches, e.g. FISH and metaproteomics. For instance, hypotheses regarding syntrophic interaction of *Coprothermobacter* and *Methanothermobacter* in a thermophilic reactor treating thermally pretreated sludge (Gagliano *et al.*, 2014) might have benefited from an additional proteomic study. Moreover, protein identification in metaproteomics profits from high-quality genome databases, making metagenomics/metatranscriptomics and metaproteomics partners rather than competitors. The fact that proteomics can also be used to improve the quality of gene annotation in genome studies (Gupta *et al.*, 2007) indicates that this interaction might not be an 'one way road'. In particular, 'proteogenomics' approaches might be applied to assist the annotation of metagenome data using the recently published proteogenomic software Peppy (Risk *et al.*, 2013). Another option might be the combination of flow cytometry and metaproteomics (Jehmlich *et al.*, 2010) as cell sorting enriches microbial subpopulations and therefore reduces complexity of samples prior to metaproteome analysis.

Advances in the field enabled by metaproteomics

So far, only a few metaproteome studies of BGPs were carried out (Table 2) (Abram *et al.*, 2009; 2011; Hanreich *et al.*, 2012; 2013; Yan *et al.*, 2012; Heyer *et al.*, 2013; Kohrs *et al.*, 2014; Lü *et al.*, 2014). Of those, only the work

Fig. 2. Krona plot of a mesophilic BGP, based on the data of Kohrs and colleagues (2014). The abundance of the taxonomic groups corresponds to the percentage of spectra based on a total number of 9485 spectra.

of Hanreich and colleagues (2012), Heyer and colleagues (2013) and Kohrs and colleagues (2014) analysed full-scale BGPs. Most studies reported on the massive problems related to sample impurities, especially humic-like substances, requiring extensive sample preparation with phenol. As a consequence, the separation of proteins by 2D-PAGE, even with improvement by paper bridge loading (Hanreich *et al.*, 2012), sometimes failed (data not shown). In order to reduce the high sample complexity, proteins can be separated by SDS-PAGE as discussed in Heyer and colleagues (2013) and Kohrs and colleagues

(2014). In addition to SDS-PAGE, Lü and colleagues (2014) used IEF to separate the proteins.

While in early metaproteome studies, only a few proteins were detected (Abram *et al.*, 2009; 2011; Hanreich *et al.*, 2012; 2013; Yan *et al.*, 2012), recent high-resolution separations using liquid IEF and SDS-PAGE (Kohrs *et al.*, 2014; Lü *et al.*, 2014) enabled the identification of up to 1000 proteins (Table 2). Assignment of these 1000 protein identifications to the biogas process enabled to cover the main steps of hydrolysis, fermentation, acetogenesis and methanogenesis. In addition, the most important
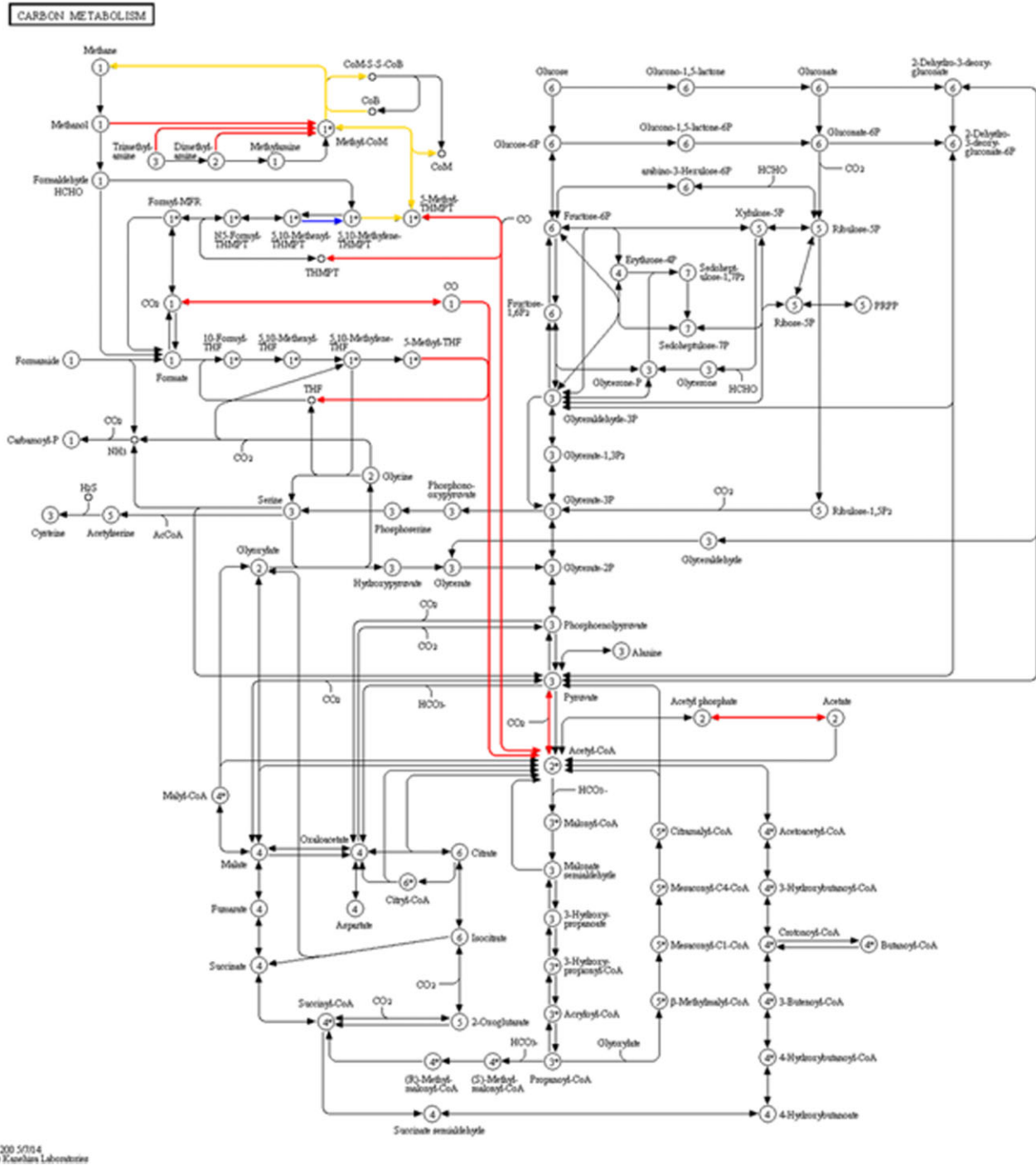
Fig. 3. Carbon metabolism of a mesophilic biogas plant, based on the data of Kohrs and colleagues (2014). KEGG pathway map of the carbon metabolism with the identified proteins for methanogenesis from different Archaea (red: *Methanosarcinales*, blue: *Methanomicrobiales*, gold: both groups).

phylogenetic groups known to be involved in biomass conversion to methane were identified.

The majority of the identified bacterial proteins belonged to the orders *Actinobacter*, *Bacteriodia*, *Bacilli*, *Chlostridiales*, *Thermotogae* and different *Proteobacter*

groups. Archaeal proteins were dominated by the orders *Methanobacteria* and *Methanomicrobia*. A comparison of phylogenetic profiles derived from metaproteomics and molecular biological studies revealed significant differences in the relative abundance of methanogens [about

Table 2. Overview about previous metaproteome studies.

Author	Fermenter	Substrate	Process temperature	Separation method	Identified proteins
Abram *et al.* (2011)	3–5 L lab scale	synthetic glucose-based wastewater	15°C	2D-PAGE (388 spots)	33 proteins
Yan *et al.* (2012)	2 L lab scale	blue algae, sludge	35°C	2D-PAGE (200–300 spots)	3 proteins
Hanreich *et al.* (2012)	8 L lab scale	beet silage, chopped rye	55°C	2D-PAGE (350 spots)	7 enzymes of methanogenesis + several housekeeping proteins
Hanreich *et al.* (2013)	500 ml batch test	straw, hay, digestate from maize fermentation	38°C	2D-PAGE (300 spots)	80 proteins
Heyer *et al.* (2013)	6 agricultural biogas plants	mainly grain silage, slurry and manure	5× mesophilic 1× thermophilic	SDS-PAGE	100–150 proteins
Kohrs *et al.* (2014)	mesophilic agricultural biogas plants	maize silage, forage rye, cattle manure and slurry	43°C	LC-MS/MS, SDS-PAGE, Liquid-IEF + SDS-PAGE	757–1,639 proteins
	thermophilic agricultural biogas plants	maize whole crop silage and poultry manure	52°C	LC-MS/MS, SDS-PAGE, Liquid-IEF + SDS-PAGE	1,663–2,091 proteins
Lü *et al.* (2014)	1 L bottle	office paper + sludge + buffer	55°C	LC-MS/MS, SDS-PAGE, Liquid-IEF	500 proteins

20–30% in metaproteome data compared with the 4% derived by metagenomics (Hanreich *et al.*, 2013)]. Both approaches may be subject to bias resulting from differences in cell lysis and extraction of proteins respective genes. When comparing both results with predicted community structures based on modelling of the biogas process, e.g. the Anaerobic Digestion Model number 1 (Batstone *et al.*, 2002), abundance of methanogens based on metaproteome data seems to be more correct. Moreover, the difference between abundances based on metaproteome and genomic data is not restricted to methanogens. For example, Lü and colleagues (2014) were astonished about only a few proteins from the genus *Gelria* by metaproteomics, although it was highly abundant in the pyrosequencing data. In this case, the bias might have been introduced by the lack of protein entries for *Gelria* in the UniProt database.

Besides proteins from Archaea and Bacteria, several proteins from plants and animals are frequently identified in samples from BGPs. They are originated from plant feedstock or manure and represent the incomplete usage of substrate. In addition, a few proteins were identified to belong to Fungi (Kohrs *et al.*, 2014) and to Bacteriophages (Lü *et al.*, 2014). Most likely, the identified proteins were not correctly assigned phylogenetically due to homologous protein sequences. At present, however, it cannot be ruled out completely that Fungi (Trinci *et al.*, 1994) or Bacteriophages (Suttle, 2007) have any relevance in BGPs.

As already discussed, the main advantage of metaproteomics is the functional characterization of microbial communities together with the phylogenetic assignment. Lü and colleagues (2014) showed, for example, that hemicellulose was hydrolysed by the genus *Caldicellulosiruptor* and that celluloses were degraded by the cellulosome of *Clostridium thermocellum*. Surprisingly, Lü and colleagues (2014) also observed a high proteolytic activity from *Clostridium proteolyticus*, indicating its function as predator or scavenger of dead biomass. The observed proteolytic activity nicely confirmed a study (Binner *et al.*, 2011) demonstrating the fast degradation of externally added cellulolytic enzymes.

For the subsequent fermentation step, mainly proteins from sugar uptake, glycolysis and to some extent from the pentose phosphate pathway were identified. This is in accordance with carbohydrates as the major substrate of biogas production. Whether the Entner–Doudoroff pathway is of relevance for fermentation (Abram *et al.*, 2011) or not (Abram *et al.*, 2011; Lü *et al.*, 2014) depends on process conditions. Pyruvate, which is produced during the glycolysis, is further converted to ethanol, acetate, lactate (Kohrs *et al.*, 2014) respectively to propionate, butyrate and butanoate (Lü *et al.*, 2014).

Amino acids derived from feedstock proteins are fermentable substrates and precursors for synthesis of microbial biomass. Based on metaproteome data, Heyer and colleagues (2013) reported about an imbalance of amino acids available for microbial anabolism. On the one hand, the enzyme glutamate dehydrogenase degrades glutamate and represents catabolism; on the other hand, enzymes like aspartokinase and dihydrodipicolinate reductase represent anabolism and are involved in *de novo* synthesis of methionine, lysine and threonine surpassing their low proportion in maize protein (Ridley *et al.*, 2002).

In contrast to the previous steps of the biogas process, the investigation of acetogenesis is even more challenging. In fact, Lü and colleagues (2014) were able

to identify the majority of key bacterial enzymes for the acetyl–CoA pathway. However, this pathway not only enables the production of acetate from hydrogen and carbon dioxide but can also be used to oxidize acetate to carbon dioxide. The impact of these two possibilities on the biogas process is further discussed in the review of Müller and colleagues (2013). Consequently, the direction of this pathway can only be determined based on the presence of species identified by metaproteomics or the absence of proteins from acetoclastic methanogenesis.

Proteins involved in methanogenesis are highly abundant in metaproteomes. Nearly all enzymes of the hydrogenotrophic and the acetoclastic pathway were identified. Under psychrophilic and mesophilic conditions, the acetoclastic pathway is favoured (Abram *et al.*, 2011; Hanreich *et al.*, 2013; Heyer *et al.*, 2013; Kohrs *et al.*, 2014), whereas, under thermophilic conditions, the hydrogenotrophic pathway is preferred (Kohrs *et al.*, 2014; Lü *et al.*, 2014). The presence of the acetoclastic pathway under thermophilic conditions was only identified once (Hanreich *et al.*, 2012), and seems to be an individual case. Furthermore, enzymes for methanogenesis from single carbon atom compounds were detected (Heyer *et al.*, 2013; Kohrs *et al.*, 2014) demonstrating the usage of methanol and methylamines released from biomass in BGPs.

Although MS-based metaproteomes of different samples show some similarities due to high abundance of methanogenic enzymes and also the presence of similar dominating phylogenetic groups, each BGP seems to have its own protein signature. Surprisingly, separation of proteins by SDS-PAGE is sufficient to produce individual protein patterns (Heyer *et al.*, 2013). These protein patterns were stable for time periods of several months and changes were correlated to process disturbance, namely an acidification of the BGP. Subsequent protein identification revealed a drastic decrease of the concentration of the enzyme methyl CoM reductase in advance of acidification. Accordingly, this key enzyme of methanogenesis could be used as a predictive biomarker. A low level of the corresponding mRNA (from mcrA gene) was previously reported to be correlated to disturbed methanogenesis (Munk *et al.*, 2012).

Conclusion and outlook

In depth, analysis of microbial communities in BGPs is required to use their full potential for biogas production. The comparison of methods for the characterization of microbial communities and recent results regarding the functional and taxonomic composition of these microbial communities obtained by various research groups showed that metaproteomics is developing into a powerful tool for the exploration of the biogas process. Besides identifying major pathways of biomass degradation, it links single metabolic pathways with microbial taxa. Each BGP shows its own, time stable protein pattern. Strong alterations in this pattern can be linked with process disturbances, and some enzymes were identified as potential biomarkers for process monitoring and fault detection.

Metaproteome analysis of BGPs is still hampered, however, by sample impurities, sample complexity, redundancy of protein identifications and a lack of genome sequences required for protein identifications. Nevertheless, the presented workflow overcomes at least parts of these problems. In the future key issues to be addressed include comprehensive sample preparation, a suitable protein separation, grouping of redundant proteins and the incorporation of meta-information from online repositories. In particular, more efficient protein extraction, improved MS and new algorithms for the verification of protein identification are urgently required to further improve this workflow and to exploit the full potential of metaproteomics. In addition, it has to be taken into account that metaproteomics is no stand-alone approach. For comprehensive analysis of microbial communities, metaproteomics should be applied in concert with microscopy, cytometry, metagenomics, metatranscriptomics and metabolomics.

References

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

Abram, F., Enright, A.M., O'Reilly, J., Botting, C.H., Collins, G., and O'Flaherty, V. (2011) A metaproteomic approach gives functional insights into anaerobic digestion. *J Appl Microbiol* **110:** 1550–1560.

Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., *et al.* (2014) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **42:** D7–D17.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. *Microbiol Rev* **59:** 143–169.

Angelidaki, I., Karakashev, D., Batstone, D.J., Plugge, C.M., and Stams, A.J. (2011) Biomethanation and its potential. *Methods Enzymol* **494:** 327–351.

Appels, L., Baeyens, J., Degreve, J., and Dewil, R. (2008) Principles and potential of the anaerobic digestion of waste-activated sludge. *Prog Energ Combust* **34:** 755–781.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* **25:** 25–29.

Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res* **28:** 304–305.

Bastida, F., Moreno, J.L., Nicolas, C., Hernandez, T., and Garcia, C. (2009) Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. *Eur J Soil Sci* **60:** 845–859.

Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A., *et al.* (2002) The IWA Anaerobic Digestion Model No 1 (ADM1). *Water Sci Technol* **45:** 65–73.

Benndorf, D., Balcke, G.U., Harms, H., and von Bergen, M. (2007) Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *Isme J* **1:** 224–234.

Bensmann, A., Hanke-Rauschenbach, R., Heyer, R., Kohrs, F., Benndorf, D., Reichl, U., and Sundmacher, K. (2014) Biological methanation of hydrogen within biogas plants: a model-based feasibility study. *Appl Energ* **134:** 413–425.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456:** 53–59.

Bernhardt, J., Michalik, S., Wollscheid, B., Volker, U., and Schmidt, F. (2013) Proteomics approaches for the analysis of enriched microbial subpopulations and visualization of complex functional information. *Curr Opin Biotechnol* **24:** 112–119.

Binner, R., Menath, V., Huber, H., Thomm, M., Bischof, F., Schmack, D., and Reuter, M. (2011) Comparative study of stability and half-life of enzymes and enzyme aggregates implemented in anaerobic biogas processes. *Biomass Conversion Biorefinery* **1:** 1–8.

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

Bradshaw, R.A., Burlingame, A.L., Carr, S., and Aebersold, R. (2006) Reporting protein identification data: the next generation of guidelines. *Mol Cell Proteomics* **5:** 787–788.

Burnum, K.E., Callister, S.J., Nicora, C.D., Purvine, S.O., Hugenholtz, P., Warnecke, F., *et al.* (2011) Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. *Isme J* **5:** 161–164.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **42:** D459–D471.

Chen, Y., Cheng, J.J., and Creamer, K.S. (2008) Inhibition of anaerobic digestion process: a review. *Bioresour Technol* **99:** 4044–4064.

Consortium, U. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **40:** D71–D75.

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

Elias, J.E., Haas, W., Faherty, B.K., and Gygi, S.P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* **2:** 667–675.

Fachagentur Nachwachsende Rohstoffe e.V. (FNR) (2013) Basisdaten Bioenergie Deutschland.

Frank, A., and Pevzner, P. (2005) PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **77:** 964–973.

Gagliano, M.C., Braguglia, C.M., Gianico, A., Mininni, G., Nakamura, K., and Rossetti, S. (2014) Thermophilic anaerobic digestion of thermal pretreated sludge: role of microbial community structure and correlation with process performances. *Water Res* **68C:** 498–509.

Gasch, C., Hildebrandt, I., Rebbe, F., and Röske, I. (2013) Enzymatic monitoring and control of a two-phase batch digester leaching system with integrated anaerobic filter. *Energy, Sustainability Soc* **3:** 1–11.

Grotenhuis, J., Smit, M., Plugge, C., Xu, Y., Van Lammeren, A., Stams, A., and Zehnder, A. (1991) Bacteriological composition and structure of granular sludge adapted to different substrates. *Appl Environ Microbiol* **57:** 1942–1949.

Gupta, N., and Pevzner, P.A. (2009) False discovery rates of protein identifications: a strike against the two-peptide rule. *J Proteome Res* **8:** 4173–4181.

Gupta, N., Tanner, S., Jaitly, N., Adkins, J.N., Lipton, M., Edwards, R., *et al.* (2007) Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res* **17:** 1362–1377.

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

Hanreich, A., Heyer, R., Benndorf, D., Rapp, E., Pioch, M., Reichl, U., and Klocke, M. (2012) Metaproteome analysis to determine the metabolically active part of a thermophilic microbial community producing biogas from agricultural biomass. *Can J Microbiol* **58:** 917–922.

Hanreich, A., Schimpf, U., Zakrzewski, M., Schluter, A., Benndorf, D., Heyer, R., *et al.* (2013) Metagenome and metaproteome analyses of microbial communities in mesophilic biogas-producing anaerobic batch fermentations indicate concerted plant carbohydrate degradation. *Syst Appl Microbiol* **36:** 330–338.

Heine-Dobbernack, E., Schoberth, S.M., and Sahm, H. (1988) Relationship of intracellular coenzyme F(420) content to growth and metabolic activity of methanobacterium bryantii and methanosarcina barkeri. *Appl Environ Microbiol* **54:** 454–459.

Heyer, R., Kohrs, F., Benndorf, D., Rapp, E., Kausmann, R., Heiermann, M., *et al.* (2013) Metaproteome analysis of the microbial communities in agricultural biogas plants. *New Biotechnol* **30:** 614–622.

Hill, D., and Holmberg, R. (1988) Long chain volatile fatty acid relationships in anaerobic digestion of swine waste. *Biological Wastes* **23:** 195–214.

Hofman-Bang, J., Zheng, D., Westermann, P., Ahring, B.K., and Raskin, L. (2003) Molecular ecology of anaerobic reactor systems. *Adv Biochem Eng Biotechnol* **81:** 151–203.

Holm-Nielsen, J.B., Al Seadi, T., and Oleskowicz-Popiel, P. (2009) The future of anaerobic digestion and biogas utilization. *Bioresour Technol* **100:** 5478–5484.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* **17:** 377–386.

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

Jagtap, P., Goslinga, J., Kooren, J.A., McGowan, T., Wroblewski, M.S., Seymour, S.L., and Griffin, T.J. (2013) A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics* **13:** 1352–1357.

Jehmlich, N., Schmidt, F., Hartwich, M., von Bergen, M., Richnow, H.H., and Vogt, C. (2008) Incorporation of carbon and nitrogen atoms into proteins measured by protein-based stable isotope probing (Protein-SIP). *Rapid Commun Mass Spectrom* **22:** 2889–2897.

Jehmlich, N., Hubschmann, T., Salazar, M.G., Volker, U., Benndorf, D., Muller, S., *et al.* (2010) Advanced tool for characterization of microbial cultures by combining cytomics and proteomics. *Appl Microbiol Biotechnol* **88:** 575–584.

Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28:** 27–30.

Karp, N.A., and Lilley, K.S. (2007) Design and analysis issues in quantitative proteomics studies. *Proteomics* **7** (Suppl. 1): 42–50.

Keiblinger, K.M., Wilhartitz, I.C., Schneider, T., Roschitzki, B., Schmid, E., Eberl, L., *et al.* (2012) Soil metaproteomics – Comparative evaluation of protein extraction protocols. *Soil Biol Biochem* **54:** 14–24.

Kelchtermans, P., Bittremieux, W., De Grave, K., Degroeve, S., Ramon, J., Laukens, K., *et al.* (2014) Machine learning applications in proteomics research: how the past can boost the future. *Proteomics* **14:** 353–366.

Klappenbach, J.A., Saxman, P.R., Cole, J.R., and Schmidt, T.M. (2001) rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* **29:** 181–184.

Klocke, M., Nettmann, E., Bergmann, I., Mundt, K., Souidi, K., Mumme, J., and Linke, B. (2008) Characterization of the methanogenic Archaea within two-phase biogas reactor systems operated with plant biomass. *Syst Appl Microbiol* **31:** 190–205.

Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26:** 231–243.

Kohrs, F., Heyer, R., Magnussen, A., Benndorf, D., Muth, T., Behne, A., *et al.* (2014) Sample prefractionation with liquid isoelectric focusing enables in depth microbial metaproteome analysis of mesophilic and thermophilic biogas plants. *Anaerobe* **29:** 59–67.

Kuhn, R., Benndorf, D., Rapp, E., Reichl, U., Palese, L.L., and Pollice, A. (2011) Metaproteome analysis of sewage sludge from membrane bioreactors. *Proteomics* **11:** 2738–2744.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680–685.

Laube, V.M., and Martin, S.M. (1981) Conversion of cellulose to methane and carbon dioxide by triculture of *Acetivibrio cellulolyticus*, *Desulfovibrio* sp., and *Methanosarcina barkeri*. *Appl Environ Microbiol* **42:** 413–420.

Li, Y., Dick, W.A., and Tuovinen, O.H. (2004) Fluorescence microscopy for visualization of soil microorganisms – a review. *Biol Fertil Soils* **39:** 301–311.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., *et al.* (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17:** 676–682.

Lottspeich, F., and Kellermann, J. (2011) ICPL labeling strategies for proteome research. *Methods Mol Biol* **753:** 55–64.

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

Luo, G., Johansson, S., Boe, K., Xie, L., Zhou, Q., and Angelidaki, I. (2012) Simultaneous hydrogen utilization and in situ biogas upgrading in an anaerobic reactor. *Biotechnol Bioeng* **109:** 1088–1094.

Lü, F., Bize, A., Guillot, A., Monnet, V., Madigou, C., Chapleur, O., *et al.* (2014) Metaproteomics of cellulose methanisation under thermophilic conditions reveals a surprisingly high proteolytic activity. *Isme J* **8:** 88–102.

Ma, Z.Q., Chambers, M.C., Ham, A.J.L., Cheek, K.L., Whitwell, C.W., Aerni, H.R., *et al.* (2011) ScanRanker: quality assessment of tandem mass spectra via sequence tagging. *J Proteome Res* **10:** 2896–2904.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

Marzorati, M., Wittebolle, L., Boon, N., Daffonchio, D., and Verstraete, W. (2008) How to get more out of molecular fingerprints: practical tools for microbial ecology. *Environ Microbiol* **10:** 1571–1581.

Munk, B., Bauer, C., Gronauer, A., and Lebuhn, M. (2010) Population dynamics of methanogens during acidification of biogas fermenters fed with maize silage. *Eng Life Sci* **10:** 496–508.

Munk, B., Bauer, C., Gronauer, A., and Lebuhn, M. (2012) A metabolic quotient for methanogenic Archaea. *Water Sci Technol* **66:** 2311–2317.

Muth, T., Behne, A., Heyer, R., Kohrs, F., Benndorf, D., Hoffmann, M., Lehtevä, M., Reichl, U., Martens, L., Rapp, E. (2015) The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation. *Journal of Proteome Research* **14:**1557–1565.

Muth, T., Benndorf, D., Reichl, U., Rapp, E., and Martens, L. (2013) Searching for a needle in a stack of needles: challenges in metaproteomics data analysis. *Mol Biosyst* **9:** 578–585.

Müller, B., Sun, L., and Schnürer, A. (2013) First insights into the syntrophic acetate-oxidizing bacteria – a genetic study. *Microbiologyopen* **2:** 35–53.

Müller, S., Hubschmann, T., Kleinsteuber, S., and Vogt, C. (2012) High resolution single cell analytics to follow microbial community dynamics in anaerobic ecosystems. *Methods* **57:** 338–349.

Nelson, M.C., Morrison, M., and Yu, Z.T. (2011) A meta-analysis of the microbial diversity observed in anaerobic digesters. *Bioresour Technol* **102:** 3730–3739.

Nettmann, E., Bergmann, I., Pramschufer, S., Mundt, K., Plogsties, V., Herrmann, C., and Klocke, M. (2010) Polyphasic analyses of methanogenic archaeal communities in agricultural biogas plants. *Appl Environ Microbiol* **76:** 2540–2548.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250:** 4007–4021.

Ondov, B.D., Bergman, N.H., and Phillippy, A.M. (2011) Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12:** 385.

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

Plugge, C.M., Scholten, J.C.M., Culley, D.E., Nie, L., Brockman, F.J., and Zhang, W.W. (2010) Global transcriptomics analysis of the Desulfovibrio vulgaris change from syntrophic growth with Methanosarcina barkeri to sulfidogenic metabolism. *Microbiol-Sgm* **156:** 2746–2756.

Popov, N., Schmitt, M., Schulzeck, S., and Matthies, H. (1975) Eine storungsfreie mikromethode zur bestimmung des proteingehaltes in gewebehomogenaten. *Acta Biol Med Ger* **34:** 1441–1446.

Rademacher, A., Zakrzewski, M., Schluter, A., Schonberg, M., Szczepanowski, R., Goesmann, A., *et al.* (2012) Characterization of microbial biofilms in a thermophilic biogas system by high-throughput metagenome sequencing. *FEMS Microbiol Ecol* **79:** 785–799.

Refai, S., Berger, S., Wassmann, K., and Deppenmeier, U. (2014) Quantification of methanogenic heterodisulfide reductase activity in biogas sludge. *J Biotechnol* **180:** 66–69.

Ridley, W.P., Sidhu, R.S., Pyla, P.D., Nemeth, M.A., Breeze, M.L., and Astwood, J.D. (2002) Comparison of the nutritional profile of glyphosate-tolerant corn event NK603 with that of conventional corn (Zea mays L. *J Agr Food Chem* **50:** 7235–7243.

Risk, B.A., Spitzer, W.J., and Giddings, M.C. (2013) Peppy: proteogenomic search software. *J Proteome Res* **12:** 3019–3025.

Roume, H., Muller, E.E.L., Cordes, T., Renaut, J., Hiller, K., and Wilmes, P. (2013) A biomolecular isolation framework for eco-systems biology. *Isme J* **7:** 110–121.

Schlüter, A., Bekel, T., Diaz, N.N., Dondrup, M., Eichenlaub, R., Gartemann, K.H., *et al.* (2008) The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. *J Biotechnol* **136:** 77–90.

Schneider, T., Keiblinger, K.M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G., Roschitzki, B., *et al.* (2012) Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *Isme J* **6:** 1749–1762.

Scholten, J.C.M., and Conrad, R. (2000) Energetics of syntrophic propionate oxidation in defined batch and chemostat cocultures. *Appl Environ Microbiol* **66:** 2934–2942.

Schweikl, H., Klein, U., Schindlbeck, M., and Wieczorek, H. (1989) A vacuolar-type ATPase, partially purified from potassium transporting plasma membranes of tobacco hornworm midgut. *J Biol Chem* **264:** 11136–11142.

Sekiguchi, Y., Kamagata, Y., Nakamura, K., Ohashi, A., and Harada, H. (1999) Fluorescence in situ hybridization using 16S rRNA-targeted oligonucleotides reveals localization of methanogens and selected uncultured bacteria in mesophilic and thermophilic sludge granules. *Appl Environ Microbiol* **65:** 1280–1288.

Senesi, N., Miano, T., Provenzano, M., and Brunetti, G. (1989) Spectroscopic and compositional comparative characterization of IHSS reference and standard fulvic and humic acids of various origin. *Sci Total Environ* **81:** 143–156.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal Chem* **68:** 850–858.

Shevchenko, A., Sunyaev, S., Loboda, A., Shevehenko, A., Bork, P., Ens, W., and Standing, K.G. (2001) Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and BLAST homology searching. *Anal Chem* **73:** 1917–1926.

Smith, P., Krohn, R.I., Hermanson, G., Mallia, A., Gartner, F., Provenzano, M., *et al.* (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* **150:** 76–85.

Sturm, M., Bertsch, A., Gropl, C., Hildebrandt, A., Hussong, R., Lange, E., *et al.* (2008) OpenMS-An open-source software framework for mass spectrometry. *BMC Bioinformatics* **9:** 163.

Suttle, C.A. (2007) Marine viruses – major players in the global ecosystem. *Nat Rev Microbiol* **5:** 801–812.

Suzek, B.E., Huang, H.Z., McGarvey, P., Mazumder, R., and Wu, C.H. (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23:** 1282–1288.

Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.J.L., Bunk, D.M., *et al.* (2010) Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res* **9:** 761–776.

Tatton, M.J., Archer, D.B., Powell, G.E., and Parker, M.L. (1989) Methanogenesis from ethanol by defined mixed

continuous cultures. *Appl Environ Microbiol* **55:** 440–445.

Thiele, H., Glandorf, J., and Hufnagel, P. (2010) Bioinformatics strategies in life sciences: from data processing and data warehousing to biological knowledge extraction. *J Integr Bioinform* **7:** 141.

Toyoda, A., Iio, W., Mitsumori, M., and Minato, H. (2009) Isolation and identification of cellulose-binding proteins from sheep rumen contents. *Appl Environ Microbiol* **75:** 1667–1673.

Trinci, A.P.J., Davies, D.R., Gull, K., Lawrence, M.I., Nielsen, B.B., Rickers, A., and Theodorou, M.K. (1994) Anaerobic Fungi in Herbivorous Animals. *Mycol Res* **98:** 129–152.

Vaudel, M., Barsnes, H., Berven, F.S., Sickmann, A., and Martens, L. (2011) SearchGUI: an open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* **11:** 996–999.

Verstraete, W., Wittebolle, L., Heylen, K., Vanparys, B., de Vos, P., van de Wiele, T., and Boon, N. (2007) Microbial resource management: the road to go for environmental biotechnology. *Eng Life Sci* **7:** 117–126.

Ward, A.J., Hobbs, P.J., Holliman, P.J., and Jones, D.L. (2008) Optimisation of the anaerobic digestion of agricultural resources. *Bioresour Technol* **99:** 7928–7940.

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

Weiland, P. (2010) Biogas production: current state and perspectives. *Appl Microbiol Biotechnol* **85:** 849–860.

Wilmes, P., and Bond, P.L. (2006) Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* **14:** 92–97.

Wilmes, P., Andersson, A.F., Lefsrud, M.G., Wexler, M., Shah, M., Zhang, B., *et al.* (2008) Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *Isme J* **2:** 853–864.

Wirth, R., Kovacs, E., Maroti, G., Bagi, Z., Rakhely, G., and Kovacs, K.L. (2012) Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. *Biotechnol Biofuels* **5:** 41.

Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., *et al.* (2009) Initial community evenness favours functionality under selective stress. *Nature* **458:** 623–626.

Wolters, D.A., Washburn, M.P., and Yates, J.R. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73:** 5683–5690.

Wöhlbrand, L., Trautwein, K., and Rabus, R. (2013) Proteomic tools for environmental microbiology-A roadmap from sample preparation to protein identification and quantification. *Proteomics* **13:** 2700–2730.

Yan, Q., Li, Y.C., Huang, B., Wang, A.J., Zou, H., Miao, H.F., and Li, R.Q. (2012) Proteomic profiling of the acid tolerance response (ATR) during the enhanced biomethanation process from Taihu Blue Algae with butyrate stress on anaerobic sludge. *J Hazard Mater* **235:** 286–290.

Zakrzewski, M., Goesmann, A., Jaenicke, S., Junemann, S., Eikmeyer, F., Szczepanowski, R., *et al.* (2012) Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing. *J Biotechnol* **158:** 248–258.

Ziganshin, A.M., Liebetrau, J., Proter, J., and Kleinsteuber, S. (2013) Microbial community structure and dynamics during anaerobic digestion of various agricultural waste materials. *Appl Microbiol Biotechnol* **97:** 5161–5174.

Zybailov, B.L., Florens, L., and Washburn, M.P. (2007) Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol Biosyst* **3:** 354–360.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Krona plot of a thermophilic BGP, based on the data of Kohrs and colleagues (2014) for a thermophilic BGP. The abundance of the taxonomic groups corresponds to the percentage of spectra based on a total number of 18139 spectra.
Fig. S2. Carbon metabolism of a thermophilic BGP, based on the data of Kohrs and colleagues (2014). KEGG pathway map of the carbon metabolism with the identified proteins for methanogenesis from different Archaea (red: *Methanosarcinales*, blue: *Methanomicrobiales*, green: proteins of anabolism).