Tracing active bugs in microbial communities by BONCAT and click chemistry-based enrichment of newly synthesised proteins

Patrick Hellwig[1,2], Daniel Kautzner[3], Robert Heyer[3,4], Anna Dittrich[5], Daniel Wibberg[6,7], Tobias Busche[8,9], Anika Winkler[8,9], Udo Reichl[1,2], Dirk Benndorf[1,2,10*]*

*: corresponding author

1. Otto-von-Guericke University Magdeburg, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany;
2. Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany
3. Multidimensional Omics Analyses group, Faculty of Technology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld,
4. Multidimensional Omics Analyses group, Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V., Bunsen-Kirchhoff-Straße 11, 44139 Dortmund
5. Department of Systems Biology, Institute of Biology, Otto-von-Guericke University Magdeburg, , Universitätsplatz 2, 39106 Magdeburg, Germany
6. Institute for Genome Research and Systems Biology, CeBiTec, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany
7. Institute of Bio- and Geosciences IBG-5, Computational Metagenomics, Forschungszentrum Jülich GmbH, 52425 Juelich, Germany
8. Center for Biotechnology - CeBiTec, Bielefeld University, Universitätsstraße 27, D-33615 Bielefeld, Germany.
9. Medical School East Westphalia-Lippe, Bielefeld University, Universitätsstraße 27, D-33615 Bielefeld, Germany.
10. Microbiology, Anhalt University of Applied Sciences, Bernburger Straße 55, 06354 Köthen, Germany

ORCID ID:

- P. Hellwig: 0000-0003-3280-9042
- A. Dittrich: 0000-0002-1056-4219
- D. Wibberg: 0000-0002-1331-4311
- T. Busche: 0000-0001-9211-8927
- U. Reichl: 0000-0001-6538-1332

Emails:

hellwig@mpi-magdeburg.mpg.de

daniel.kautzner@uni-bielefeld.de

anna.dittrich@ovgu.de

robert.heyer@isas.de

dwibberg@cebitec.uni-bielefeld.de

tbusche@cebitec.uni-bielefeld.de

awinkler@cebitec.uni-bielefeld.de

dirk.benndorf@hs-anhalt.de

ureichl@mpi-magdeburg.mpg.de

Keywords: Metaproteomics, mass spectrometry, BONCAT, click chemistry, microbiome microbial activity, anaerobic digestion

Running title: Microbial Protein Synthesis Tracking

Abstract (208 words):

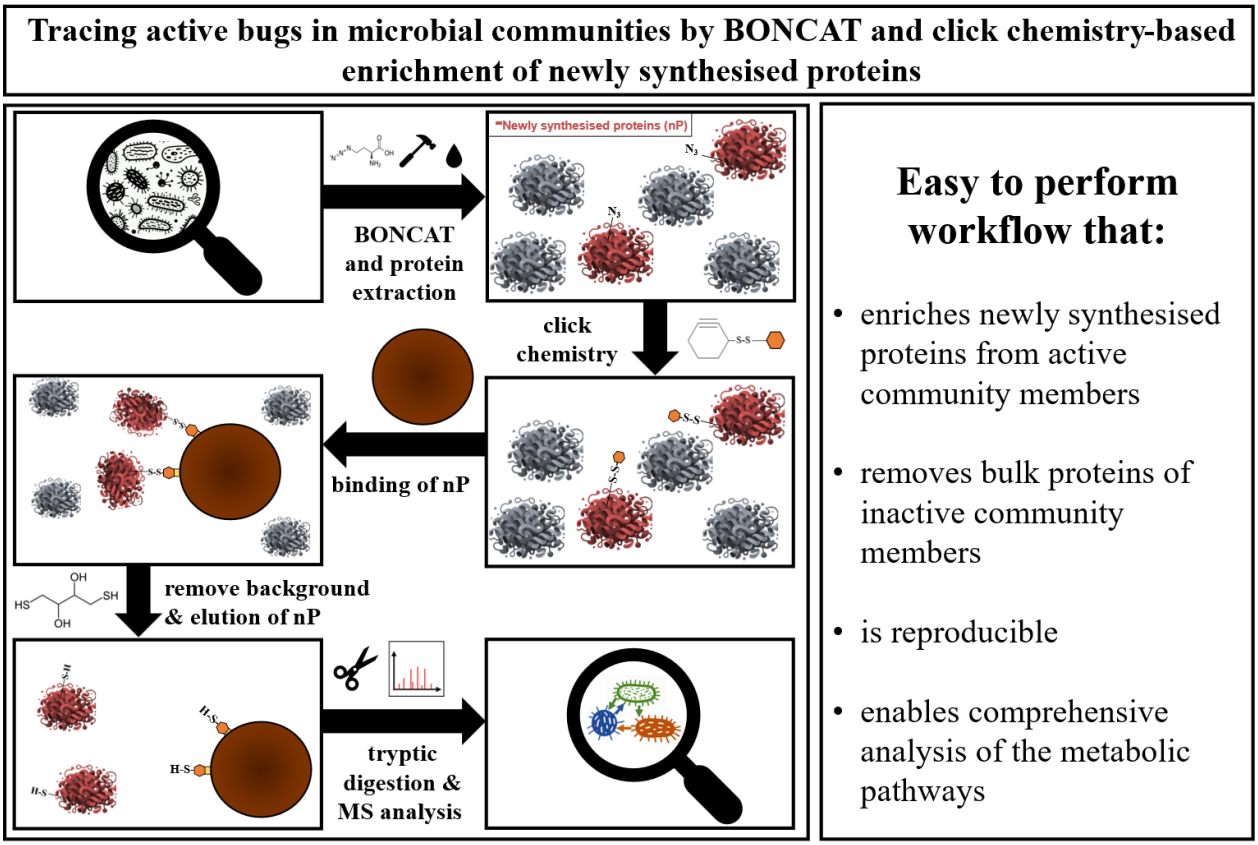
A comprehensive understanding of microbial community dynamics is fundamental to the advancement of environmental microbiology, human health, and biotechnology. Metaproteomics, i.e. the analysis of all proteins in a microbial community, provides insights into these complex systems. Microbial adaptation and activity depend to a significant extent on newly synthesized proteins (nP), however, the distinction between nP and bulk proteins is challenging. The application of bioorthogonal non-canonical amino acid tagging (BONCAT) with click chemistry has demonstrated efficacy in the enrichment of nP in pure cultures. However, the transfer of this technique to microbial communities has proven challenging due to the inefficient BONCAT process, which results in a low nP yield and a high contamination rate from non-labelled bulk proteins. To address this, a new workflow to improve the efficiency and specificity of nP enrichment was developed using a laboratory-scale mixture of labelled *E. coli* and unlabelled yeast. This workflow was successfully applied to an anaerobic microbial community with initially low BONCAT efficiency. A substrate shift from glucose to ethanol selectively enriched nP with minimal background. The identification of bifunctional alcohol dehydrogenase and a syntrophic interaction between an ethanol-utilizing bacterium and two methanogens (hydrogenotrophic and acetoclastic) demonstrates the potential of metaproteomics targeting nP to trace microbial activity in complex microbial communities.

54 Abbreviations:

AA	amino acids
AD	anaerobic digestion
ADH	alcohol dehydrogenase
AHA	4-azido-l-homoalanine
BONCAT	bioorthogonal non-canonical amino acid tagging
BSA	bovine serum albumin
CC	click chemistry
DBCO	dibenzocyclooctyne
DTT	dithiothreitol
FASP	filter-aided sample preparation
HPG	homopropargylglycine
HPLC	high-pressure liquid chromatography
IAA	iodoacetamide
KEGG	Kyoto Encyclopedia of Genes and Genomes
LBR	laboratory biogas reactor
LC	liquid chromatography
MAG	metagenome assembled genome
MC	microbial community
MS	mass spectrometry
Mgf	mascot generic file
ncAA	non-canonical amino acids

nP	newly synthesized proteins
PASEF	parallel accumulation–serial fragmentation
PBS	phosphate-buffered saline
PEG	polyethylene glycol
rpm	rounds per minute
RT	room temperature
SDS	sodium dodecyl sulfate
SOP	standard operation procedure
SS	disulfide

Graphical Abstract:



Introduction

Microbial communities (MC) include single-celled eukaryotes, archaea, bacteria, and viruses [1–3]. These communities are found in various environments, including soil and water, and within living organisms such as animals and humans. Additionally, MC are used for industrial processes such as wastewater treatment or energy production e.g. in anaerobic digestion (AD) [4]. The understanding of MC is still very limited due to their inherent complexity. Enhancing our understanding will help to cure diseases, optimize biotechnological processes, and even act against climate change [5–7]. Therefore, the analysis of MC gained prominence in the last decades. The analysis of MC is summarized under the term “Meta-omics”.

Metagenomics is the most common method currently used to characterize MC. It involves the genome analysis of each member of an MC and gives insights into the taxonomical composition and metabolic potential of an MC [8]. However, metagenomics requires amendment with expression analysis of proteins in MC (metaproteomics) to fully decipher the functions of the MC [9].

Microbial adaptation to changing environments or process parameters and the resulting changed microbial activity occurs primarily through the expression of newly synthesized proteins (nP), which can be hardly distinguished from already present bulk proteins. Consequently, the analysis of the nP would provide a novel dimension of metaproteomics, thereby facilitating a deeper comprehension of MC activities.

Advances in bioorthogonal chemistry have enabled the identification of nP. One technique is bioorthogonal non-canonical amino acid tagging (BONCAT), which allows the selective labelling of nP [10]. In BONCAT, homologous non-canonical amino acids (ncAA) are added to growing cells. During protein synthesis, the tRNA of an amino acid mistakenly selects the ncAA instead of the amino acid with a likelihood of 1:400 to 1:500 [11]. The incorporation rate of ncAA into the amino acid chain of a protein correlates negatively with the concentration of the

targeted amino acid in the medium [12]. 4-Azido-L-homoalanine (AHA) and L-Homopropargylglycine (HPG) are common ncAAs that replace methionine, but new ncAAs are being developed, such as for threonine [13]. Incorporation of ncAAs into nP results in a mass shift that can be detected by mass spectrometry (MS). Nevertheless, the low proportion of labelled nP in relation to the bulk proteins in MC is insufficient to permit the reliable identification of nP through metaproteomics.

As a remedy, BONCAT can be combined with click chemistry (CC) in the form of azide-alkyne cycloaddition [14; 15]. Thus, a fluorescence or biotin marker can be attached to the AHA (azide group) or HPG (alkyne group) [16]. BONCAT and CC have previously been applied to visualize and enrich nP from eukaryotic cells, such as neurons [12], HeLa cells [17], or HEK cells [18] and from viruses and their host cells [19]. However, BONCAT has been used mainly in cell culture experiments with only one cell type. Hatzenpichler et al. [20] showed that the application of BONCAT in MC is possible. Additionally, Reichart et al. [21] showed that sorting active ncAA-labelled microbes from non-active microbes with flow cytometry is possible. BONCAT and CC also show promise for bacteriophages in MC [22; 23].

In the case of metaproteomics, the direct detection of nP is of paramount importance for the understanding of MC. However, the currently available enrichment methods for nP are designed for clean, pure cultures with high ncAA-labeling. In contrast, low enrichment efficiencies of nP and an enormous, unspecific binding of bulk proteins are achieved when using more complex and significantly less ncAA-labelled MC samples. Therefore, there is a clear need for the development of reliable, specific, and reproducible methods for the detection of nP in MC *via* MS. The aim of this work was thus to develop a workflow that allows the direct enrichment of nP from growing MC for the detection *via* MS. The primary focus was the enrichment of already AHA-labelled nP since conditions of BONCAT have to be adapted individually for each experimental condition and MC in particular the dynamic of protein translation [20; 24].

Therefore, a test system for nP enrichment was developed. It should demonstrate the specificity of nP enrichment and be easy to reproduce in any laboratory. Based on these restrictions, a test system consisting of *E. coli* and *Saccharomyces cerevisiae* (yeast) was chosen. *E. coli* was labelled with AHA after a substrate shift from glucose to lactose that induces metabolic changes and thus expression of nP involved in lactose metabolism. Next, unlabelled yeast cells were added. The established workflow allowed the enrichment of *E. coli* nPs involved in lactose metabolism without contamination by yeast proteins.

As a proof of concept, the workflow was applied to enrich nP from an anaerobic MC derived from a laboratory-scale biogas reactor (LBR). Given the slow growth of this anaerobic MC, the efficiency of ncAA-labeling was expected to be suboptimal. However, if nP can be enriched from this MC and reliably detected in MS, the developed workflow should be applicable to other slowly growing MC. The LBR used for this study was operated with glucose as the primary substrate. The objective was to shift the substrate to ethanol so that the nP associated with the consumption of ethanol as a secondary substrate would be labelled with AHA and could subsequently be enriched with the established workflow. It was anticipated that successful nP enrichment would allow to trace the degradation pathway of ethanol under anaerobic conditions. It may even be possible to gain new insights into this poorly described metabolic pathway [25].

Material and Methods

The following chapter briefly describes the workflow established (see Figure 1). A detailed description of the methods, including a step-by-step standard operation procedure (SOP), can be found in Supplementary Note 1.

AHA-labeling of *E. coli* nP

The substrate of an at 37°C growing *E. coli* (DSM 5911) culture grown aerobically in M9 media was shifted from glucose to lactose while cells were incubated with 0.1 mM AHA. The same

was done for control cells without adding AHA. After the harvest, the two *E. coli* cultures were separately mixed with an overnight yeast culture. The mixture of *E. coli* and yeast is named "test system" in the following. For full details, refer to Supplementary Note 1: SOP 1.

AHA-labeling of nP from MC of LBR

In total, 50 mL of cell suspensions from an MC of LBR (1 L total volume, 40°C, 75 rpm, pH 7.4 – 7.5, see Supplementary Table 1) was sampled using a 50 mL syringe. The cultures were then transferred to nitrogen-flushed 100 mL serum bottles, diluted 1:1 with 50 mL anaerobic medium without glucose (see Supplementary Table 2), and incubated for one hour. Following this, 1 mL of 10% (vol./vol.) ethanol solution was added. Additionally, for BONCAT, 1 mL of 20 mM AHA was introduced, while controls received 1 mL of water. An extra control was prepared with 3.8 mM glucose instead of ethanol. The cultures were sealed and incubated for 24 hours at 40°C with gentle stirring until harvest. For full details, refer to Supplementary Note 1.

Cell disruption with a ball mill

Samples from the test system and the LBR MC were suspended in 5 mL of 20 mM Tris/HCl-MgCl buffer (pH 7.5) and treated with 1 g/mL silica beads and 0.5 µL/mL Cyanase™ Nuclease solution. The cells were disrupted in a ball mill, incubated at 37°C for 15 minutes, and then centrifuged. The resulting supernatant was transferred to a new reaction tube. For full details, refer to Supplementary Note 1: SOP 1.

Protein extraction with acetone (for clean samples)

The proteins of the test system were precipitated with the five-fold volume of ice-cold acetone at -20°C for at least 1 h and centrifugation afterwards. The supernatant was discarded, and the protein pellets were dried under a fume hood at room temperature (RT). The dry protein pellets were resuspended in incubation buffer I. The protein samples were centrifuged, and the

supernatants were transferred to new reaction tubes. For full details, refer to Supplementary Note 1: SOP 1.

Protein extraction with methanol and chloroform (for impure samples)

The MC proteins were extracted using a mixture of methanol, chloroform, and water, followed by centrifugation. After carefully removing the top phase, methanol was added, and the mixture was vortexed and centrifuged again. The resulting protein pellets were dried and resuspended in incubation buffer II. After centrifugation, the supernatant was transferred to new tubes. For full details, refer to Supplementary Note 1 and Wessel and Flügge (1984)[26].

Protein quantification

Amido black assay was used to quantify the protein concentration of each sample, as described earlier [27].

Click chemistry with fluorophores and SDS-PAGE with fluorescent scan

AHA incorporation was assessed using 10 µg protein from each sample. For this purpose, the dye Dibenzocyclooctyne (DBCO) Cyanin 5.5 was attached using CC following the SOP 2 in Supplementary Note 1. The fluorophore-tagged proteins were separated in a 1.5 mm 10 % SDS-PAGE and scanned twice: once with a LI-COR Odyssey Classic fluorescence scanner at 700 nm and a second time after Coomassie staining [27; 28]. For full details, refer to Supplementary Note 1: SOP 5.

Click chemistry with biotin linker

Based on the fluorescence signal (see previous paragraph), 100 µg protein of the test system and 200 µg protein of the MC were used for the attachment of DBCO-SS-Biotin via CC following the SOP 2 in Supplementary Note 1.

Enrichment of AHA-labelled proteins

The nP of the test system and the MC were enriched using Dynabeads™ MyOne™ Streptavidin C1 beads and the protocol described in SOP 3 in Supplementary Note 1. The beads were blocked with bovine serum albumin (BSA), amino acids (AA), or left blocked.

FASP digestion

The tryptic digestion was performed using FASP [27] with few modifications (Supplementary Note 1: SOP 4). The proteins contained in the dithiothreitol (DTT) elution phase and the elution of the SDS boiled beads were precipitated with acetone. 25 µg protein of each sample before CC ("not enriched samples") was precipitated with acetone (control without enrichment). The resulting protein pellets were dissolved in 200 µL 8 M urea buffer and transferred to a filter unit (Centrifugal Filter Unit, 10 kDa). In FASP digestion, a ratio of 1:100 (trypsin quantity to protein quantity) was used. 10 µg protein was assumed to calculate the amount of trypsin needed for the enriched nP fractions. After extraction and concentration via vacuum centrifuge, the eluted peptides were transferred to HPLC vials for MS measurement.

LC-MS/MS measurements

The peptides of each sample were measured using a timsTOF™ pro mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany) coupled online to an UltiMate® 3000 nano splitless reversed-phase nanoHPLC (Thermo Fisher Scientific, Dreieich, Germany) in PASEF® mode (see Supplementary Note 1).

Illumina library preparation, MiSeq sequencing, and metagenome assembly for protein database

The metagenome of the LBR was generated from three replicates using PCR-free libraries prepared with the Illumina TruSeq® DNA PCR-free kit [29]. The libraries were subjected to quality control and sequenced on the MiSeq platform (2 × 300 bp paired-end, v3 chemistry). After sequencing, data were stripped of adapters and low-quality reads [30], followed by assembly with Megahit (v1.1.1) [31] and gene prediction with Prodigal v.2.6.0 [32].

Metagenomic binning [33; 34] was performed with Bowtie 2 [35] and MetaBAT2 (76 Metagenome assembled genome (MAGs)) [36], while completeness and contamination were assessed with BUSCO (v5.7.0) [37]). The translated amino acid sequences of the predicted genes were then used for further analysis. See Supplementary Note 1 for all details. For the most abundant MAGs, the taxonomy was additionally evaluated with the Type (Strain) Genome Server (TYGS) [38; 39].

Protein identification using Mascot and the MetaProteomeAnalyzer

With Compass Data Analysis software (version 5.1.0.177, Bruker Corporation, Bremen, Germany) the raw files were converted into Mascot Generic Files (mgf) and mgf were searched with Mascot (version 2.6) for the peptide spectrum matches. For the test system, a defined UniProtKB/SwissProt database (08/05/2022) containing *E. coli* K12 (taxonomy_id:83333) and yeast (taxonomy_id:4932) proteins was used. A metagenome from the LBR was used for the MC samples (see Illumina library preparation, MiSeq sequencing, and metagenome assembly). Metaproteins were generated from the MC samples using MetaProteomeAnalyzer software (version 3.0 [40]). See Supplementary Note 1 for all details.

Further analysis of the metaprotein data

The metaproteins were uploaded to Prophane (version 6.2.6) [41] for basic local alignment search tool analysis of the functional and taxonomic annotation. For all unknown KO numbers after Functional Ontology Assignments for Metagenomes annotation [42], KofamKOALA [43] was used to identify possible missing KO numbers (see Supplementary Table 3 and Supplementary Note 1).

Results and Discussion

Development of a nP enrichment workflow with the test system

Before the development of the enrichment protocol, the *E. coli* (AHA-labelled) and yeast (unlabelled) test system was optimized by optimizing the AHA labelling of *E. coli* (see

Supplementary Figure 1). Attempts were made to enrich nP of *E. coli* from the test system with known enrichment protocols (from manufacturers or published) but these were unsuccessful (data not shown). The major issues were that MS of tryptic digests was either unable to detect any proteins, or detected similar proportions of yeast proteins (unspecific background) and *E. coli* proteins. The results demonstrated the necessity for the development of a test system for the monitoring of nP enrichment, and the requirement for the implementation of negative and positive controls to regulate non-specific binding in each experimental procedure.

A first successful enrichment of nP after click chemistry was obtained with DBCO-SS-PEG3-biotin (see Supplementary Figure 3). The nP were eluted with ease and precision under mild conditions by the addition of DTT through the reductive cleavage of the disulfide bridge in the DBCO-SS-PEG3-biotin. Therefore, DBCO-SS-PEG3-biotin appeared to be an optimal choice for our nP enrichment. However, many unspecific yeast proteins (background) were co-eluted with the nP (see Supplementary Figure 3).

Therefore, washing was optimized. An adapted washing protocol [18] and the use of phosphate-buffered saline (PBS) buffer with increased NaCl concentration (237 mM) reduced the amount of background yeast proteins. However, it lacked the reproducibility of nP elution (Supplementary Figure 4). For higher reproducibility of enrichment, the loss of nP during elution should be minimized.

Blocking the beads with BSA (e.g. [44; 45]), before loading the samples, reduced the amount of yeast protein background further but also increased the variability in nP elution (Supplementary Figure 4). Furthermore, the elution was found to be contaminated with BSA proteins. Blocking the beads seemed to be important for the removal of non-specific protein, but BSA was not optimal.

Inspired by Nicora et al. (2013), the beads were blocked with the AA: leucine (aliphatic hydrophobic), tryptophan (aromatic hydrophobic), histidine (polar positive), glutamine (polar

neutral), glutamic acid (polar negative) and glycine (nonpolar neutral). They should not cause steric effects and should prevent any unspecific interactions with the beads. The method ("AA method") maintained a low background and exhibited minimal variability and loss in nP elution (Figure 2 A and Supplementary Figure 4), making it the most promising approach.

Examining the AA method using the test system

In total, 1,558 ± 53 proteins (885 ± 25 *E. coli*, 590 ± 29 yeast, 83 ± 2 both) were identified in the AHA-labelled but non-enriched test system ("not enriched sample") (Figure 2). Over a third of these proteins were yeast proteins. After nP enrichment, 607 ± 16 proteins (555 ± 17 *E. coli*, 21 ± 2 yeast, 31 ± 1 both) were identified, reducing the amount of identified yeast proteins by over 96% while retaining 63% of the *E. coli* proteins (Figure 2). Additionally, the AA method resulted in a minimal background in the unlabelled control test system (without AHA-labelled proteins) with 10 ± 3 proteins (6 ± 3 *E. coli*, 1 ± 1 yeast, 3 ± 1 both).

Furthermore, the AA method allowed the detection of 48 proteins that were absent in the not enriched samples (see Supplementary Table 2). Of these, 29 were *E. coli* proteins related to cell metabolism, movement, transport, and protein biosynthesis. Additionally, 19 very low-abundant yeast proteins were detected, likely co-eluting with AHA-labelled *E. coli* nP. This co-elution did not occur in the AA method control, which confirmed the specificity of the elution of labelled nP by DTT. Boiling of the beads with SDS-PAGE sample buffer released no additional proteins showing that DTT was very efficient in elution (Supplementary Figure 5). Consequently, the selected elution with DTT is not only specific but also highly efficient.

The AA method allowed the enrichment of *E. coli* nP very efficiently with a low yeast protein background. Due to the substrate shift from glucose to lactose, the enriched *E. coli* nP should include proteins responsible for lactose degradation encoded by the *lac operon,* namely β-galactosidase (*lacZ*), galactoside O-acetyltransferase (l*acA*), and β-galactoside permease (*lacY*) [46; 47]. The AA method successfully enriched β-galactosidase (217 ± 6 spectra) and

galactoside O-acetyltransferase (5 ± 2 spectra). However, β-galactoside permease was not detected, neither in the enriched sample nor in the not enriched sample. As it is an integral membrane it was probably lost during protein extraction [48].

Lactose consists of galactose and glucose. Therefore, the pathway for galactose metabolism proteins should also be enriched with the AA method. The assignment of the detected nP to the KEGG pathway for galactose metabolism showed that all enzymes for galactose degradation were identified, including β-galactosidase (EC 3.2.1.23; 217 ± 6 spectra), aldose 1-epimerase (EC 5.1.3.3; 14 ± 1 spectra), galactokinase (EC 2.7.1.6; 50 ± 5 spectra), galactose-1-phosphate uridylyltransferase (EC 2. 7.7.12; 13 ± 3 spectra), UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9; 3 ± 1 spectra) and UDP-glucose 4-epimerase (EC 5.1.3.2; 34 ± 2 spectra) (see Figure 3). These proteins were all increased at least 2-fold in the enriched nP compared to the not enriched proteins. Glycolytic enzymes were also identified; however, they did not exhibit a twofold increase in the enriched nP compared to the not enriched sample. This might be explained by the fact, that cells were grown on glucose as substrate before the substrate shift (Figure 3). Furthermore, all dehydrogenases involved in the mixed acid fermentation of *E. coli* were enriched after the substrate shift by the AA method [49]. The substrate switch was carried out by centrifugation, which presumably led to anaerobic conditions for a short time and thus facilitated the formation of these enzymes. This further emphasizes the significance of nP as a conduit of knowledge regarding the activities of microorganisms.

In conclusion, the AA method is well-suited for monitoring the specific accumulation of nP in response to changing conditions in the test system. It is particularly noteworthy that the developed method successfully enriched nP while significantly reducing the unspecific, unregulated protein background. Therefore, this method holds promise as a tool for analyzing nP in complex MC.

<u>Enrichment of nP in anaerobic MC of an LBR using the AA method</u>

The next objective was to demonstrate that the established workflow was also effective for less strong AHA-labelled real MC. Therefore, the nP of an anaerobic MC from an LBR were labelled with AHA while the substrate was switched from glucose to ethanol.

Initial quality control of successful AHA incorporation by SDS-PAGE revealed a weak fluorescent signal that represents nP tagged with DBCO-Cyanin 5.5 by CC (see Supplementary Figure 6). The lower incorporation of AHA in the MC compared to a pure test sample reinforces the necessity for specific and effective enrichment of nP to reliably detect nP in MCs with MS. Unfortunately, in one out of three biological replicates of the experiment, no fluorescence was detectable (see Supplementary Figure 6). Therefore, this replicate was excluded from the further analysis. The other two replicates were processed with the AA method and subsequent metaproteome analysis.

To achieve more accurate identification of metaproteins within the LBR MC, a metagenome was sequenced from the LBR, and a corresponding protein database was generated. Additionally, 76 MAGs were binned. The taxonomic assignment of bins was utilized for subsequent analysis of active microbes.

In total, 1,960 ± 180 metaproteins were identified in the not enriched sample of the AHA-labelled MC (“Not enriched sample AHA”), while 441 ± 120 metaproteins were identified after the enrichment of nP (“Elution AHA”). Application of the AA method to the non-labelled control (“Elution no AHA”) showed a low non-specific protein background of 12 ± 11 metaproteins (Figure 4). Notably, with the AA method, 24 metaproteins were identified, which were absent in the not enriched sample. Consequently, the specific enrichment of nP should provide a better insight into the potential conversion of ethanol to methane in this LBR than the metaproteome analysed without enrichment.

<u>Ethanol-associated taxonomies based on nP enrichment</u>

Initially, the enrichment of the nP was considered for the 20 most abundant MAGs of the AA method and without the AA method (29 MAGs in total, see Supplementary Table 6). Bin_27 (family *Propionibacteriaceae*), Bin_24 (class *Methanomicrobia*), Bin_41 (class *Clostridia*), and Bin 53 (species *Candidatus Methanoculleus thermohydrogenitrophicus*) are of particular interest due to their high abundance. Bin_27 has been depleted, whereas the other three MAGs are similarly abundant in both samples. While certain MAGs, such as Bin_74 (family *Desulfobacteraceae*), Bin_62 (unknown *Bacteria*), Bin_46 (unknown *Bacteria*), Bin_11 (family *Synergistaceae*), and Bin_20 (family *Alcaligenaceae*), do not stand out due to their high abundance, they do exhibit strong enrichment with the AA method. In contrast, certain MAGs, including Bin_7 (family *Pseudomonadaceae*), Bin_23 (family *Methanobacteriaceae*), Bin_69 (family *Flavobacteriaceae*), Bin_71 (class *Clostridia*), Bin_75 (class *Actinomycetia*), and Bin_67 (family *Pseudomonadaceae*), exhibited a depletion pattern after enrichment (see Figure 5).

Upon closer examination of the metaprotein functions from the enriched MAGs, a notable abundance of stress proteins (up to 100% in some MAGs) was observed, likely induced by the addition of ethanol [50]. Consequently, stress proteins were excluded from consideration to ensure a metabolic enrichment of MAGs (see Figure 5). In the case of the high-abundant MAGs, the abundance after the removal of stress proteins has only decreased slightly from Bin_41. However, in the case of the low-abundance MAGs, there have been some significant changes. MAGs such as Bin_74 (family *Desulfobacteraceae*), Bin_46 (unknown *Bacteria*) and Bin_20 (Genus *Castellaniella*) exhibited a greater than 10-fold increase, while MAGs such as Bin_13 (family: Solibacteraceae), Bin_34 (class *Betaproteobacteria*), Bin_25 (Species *Candidatus Brevifilum fermentans*) and Bin_62 (unknown *Bacteria*) could no longer be identified after removal of stress-induced proteins from analysis. Thus, Bin_74, Bin_46, and Bin_20 mostly benefited from the addition of ethanol, while ethanol induced stress in other MAGs such as Bin_13, Bin_34, Bin_25, and Bin_62.

Ethanol-associated functions based on nP enrichment

Upon closer examination of Bin_46, Bin_74, and Bin_20, it was discovered that Bin_74 and Bin_20 exhibited gluconeogenesis activities. Bin_46 is however too low-abundant to make statements about the metabolism. A review of the genes revealed that each of the three MAGs possesses at least two genes that encode for a potential alcohol dehydrogenase (ADH). Additionally, *Castellaniella* (Bin_20) is known to utilize acetyl-CoA from acetate, ethanol, or pyruvate *via* the glyoxylate cycle [51–55]. The growth of these three strains is presumably slow. Therefore, the amount of accumulated nP was too low to reliably determine their complete metabolism. Without enrichment with the AA method, they would thus not have been the focus of analyses. In follow-up studies, their metabolism could be further analysed with longer incubation times, as such species growing on ethanol were then enriched [56].

The analysis of the Functional Ontology Assignments for Metagenomes in comparison to a sample with glucose (as in the LBR) showed that there were clear differences due to the substrate shift (see Supplementary Figure 13). A clustered heatmap of the Functional Ontology Assignments for Metagenomes and most abundant MAGs showed that Bin_27, Bin_24, Bin_41, and Bin_53 in particular account for most of the functions of the microbial community (Figure 6 A).

A detailed analysis of metaproteomes of all samples identified 29 different ADHs, with 7 ADHs (Bin_41, 2x Bin_27, Bin_66, Bin_24, Bin_70, and Bin_53) as nP. Out of these 7 ADHs, 3 poorly classified ADHs (FHCFJDLK_121438 (Bin_41), FHCFJDLK_86872 (Bin_53), and FHCFJDLK_294919 (Bin_70)) were increased (≥ 2.5-fold) after enrichment (see Supplementary Table 3 and 4). An analysis with Alphafold [57; 58] and KofamKOALA revealed that all three poorly classified ADHs are responsible for ethanol degradation (see Supplementary Table 4). However, Bin_70 was a very low abundant MAG (0.13%). This MAG belongs to the genus *Tepidiphilus*, which is known as a secondary fermenter and syntrophic

acetate oxidizer [59; 60]. The abundance of this MAG is however insufficient to definitively ascertain its metabolic processes in this MC. It appears that it degrades ethanol and generates ATP via membrane-bound proton pumps.

Based on the proteins, Bin_41 appears to be an acetate producer. All proteins for acetate formation from ethanol were found in the proteome of Bin_41 [61], except an aldehyde dehydrogenase, which is necessary for the further degradation of acetaldehyde [62–64]. It is therefore assumed that the ADH of Bin_41 is bifunctional, converting ethanol into acetaldehyde (EC: 1.1.1.1) and then directly to acetyl-CoA (EC: 1.2.1.10, Supplementary Figure 10) [65–68]. Bifunctionality poses a challenge in the annotation of KO numbers, as normally only one KO number is assigned to one protein. It was only through enrichment that we became aware of the protein and were able to consider the possibility of bifunctionality. Furthermore, NADH dehydrogenases, which are crucial for replenishing depleted NAD+ levels during anaerobic ethanol oxidation, were enriched from Bin_41 [69]. The removed electrons are probably transferred to a membrane-bound cytochrome-coupled formate dehydrogenase, which utilizes free CO_2 and protons for formate formation at iron-containing 2FE-2S clusters [70]. Additionally, ATP appeared to be generated through substrate chain phosphorylation [70; 71] (see Supplementary Table 4). However, a major portion of stored energy is probably consumed for the reduction of formate due to coupled proton influx [72].

Bin_53 (species *Candidatus Methanoculleus thermohydrogenitrophicus*) is closely related to *Methanoculleus*, which is known to use ethanol and secondary alcohols as electron donors for methanogenesis [73]. Ethanol is converted into acetate by a reaction that is coupled with NADP reduction, which is then recovered by NADPH oxidoreductase (EC 1.5.1.40) [74; 75] delivering reduction equivalents to hydrogenotrophic methanogenesis. Since ethanol is not completely oxidized to CO_2, an alternative source of CO_2 is required for hydrogenotrophic methanogenesis [76; 77]. A formate dehydrogenase (EC 1.17.1.9) was identified that converts formate into CO_2,

which can then be used for hydrogenotrophic methanogenesis and feeds released reduction equivalents into hydrogenotrophic methanogenesis. *Methanoculleus* is also known to use formate [77]. A syntropic interaction between Bin_53 and Bin_41 is plausible with Bin_41 delivering formate to Bin_53 [78–80]. Furthermore, Bin_53 also should be able to consume H_2 which could result in the inhibition of Bin_41 when accumulating in the medium [81–83]. But syntrophic transfer of H_2 probably did not play a major role in the LBR MC since no hydrogenase capable of releasing H_2 was detected in Bin_41, (see Figure 6 B). The phenomenon of syntrophic ethanol oxidation between fermenting bacteria and methanogenic archaea observed here has been repeatedly observed in anaerobic digesters [25; 84–86].

In Bin_24, all proteins necessary for acetoclastic methanogenesis were identified, indicating its potential to convert the acetate produced in Bin_41 and Bin_53 into methane and CO_2 (see Supplementary Figure 12 and Figure 6 B). Therefore, based on the enriched nP, a syntrophic relationship between these three MAGs is highly probable. In this scenario, ethanol is converted to acetate by Bin_53 and Bin_41, which transfer formate to each other, and Bin_24 consumes the acetate. However, other sources of H_2, formate, acetate, and CO_2, e.g. originating from fermentative degradation of microbial biomass [87], and additional partners involved in ethanol oxidation (e.g. by minor abundant bins) cannot be excluded.

<u>Further Improvements and important considerations for implementing the workflow</u>

A successful BONCAT is the precondition for this workflow. Our experiments demonstrated the sufficient efficacy of BONCAT in slow-growing anaerobic MC. We advise reading Hatzenpichler et al [20] and Landor et al. [24] as they address pivotal aspects for successful BONCAT and the potential cytotoxicity of BONCAT. Following a successful BONCAT, our workflow can be readily implemented in any laboratory (see SOPs in Supplementary Note 1). In the case of suboptimal nP yields despite a successful BONCAT, increasing the bead volume is recommended. The sensitivity and reproducibility of quantitative measurements could be

further improved by data-independent acquisition mass spectrometry measurements, such as DIA-PASEF [88]. In addition, fluorescence-activated cell sorting of ncAA-labelled cells from MC [21] could generate corresponding metagenomes for improved protein identification. Alternatively, the workflow could be adapted to HPG or β-ethynylserine and copper-based click chemistry, although this has not yet been tested, as copper-free click chemistry with AHA is demonstrably more efficient [13; 89–91].

Conclusion

The combination of BONCAT and the direct enrichment of nP is an effective approach for investigating MC using MS. Enriched nP reflects metabolic or adaptive processes, making it easier to get new insights into the MC by highlighting relevant proteins. The optimized nP enrichment workflow was demonstrated to be effective by a test system of *E. coli* and yeast and may serve as a reference for future nP enrichments. The application of the workflow on a slow-growing anaerobic MC enabled the description of syntrophic interaction in an MC consuming ethanol as substrate.

Acknowledgement

We want to thank Dr.-Ing. Ute Thron from GETEC Green Energy GmbH for providing the fermentation substrate for the inoculum of the LBR.

Funding

The bioinformatics support of the BMBF-funded project “Bielefeld-Gießen Center for Microbial Bioinformatics” (BiGi) (BMBF grant FKZ 031A533) within the German Network for Bioinformatics Infrastructure (de.NBI) is gratefully acknowledged. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 446327964).

Figures

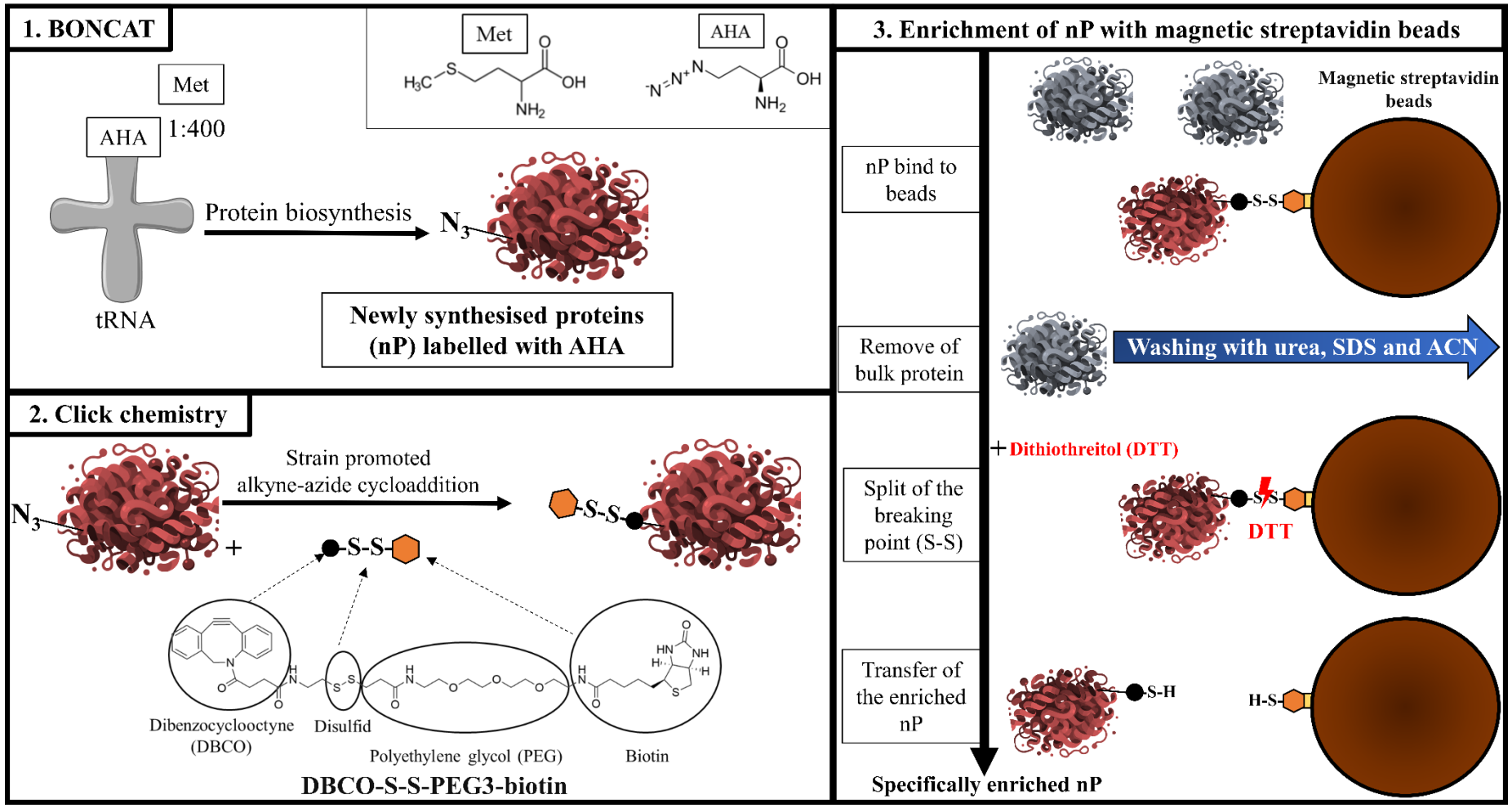


Figure 1: Workflow of the enrichment of newly synthesized proteins. **1.:** nP were labelled with AHA. **2.:** Free cysteines of the proteins were blocked with IAA (not shown). The labelled nP were tagged with DBCO-S-S-PEG$_3$-biotin within 1 h. **3.:** The biotin-tagged nP were bound to magnetic

streptavidin beads (blocked with amino acids in advance). All proteins not bound to the beads were removed by harsh washing with urea, sodium dodecyl sulfate (SDS) and acetonitrile (ACN). The bound proteins were eluted from the streptavidin beads with DTT. The specifically enriched nP were transferred to a new tube, tryptic digested and further analysed with LC-MS/MS.

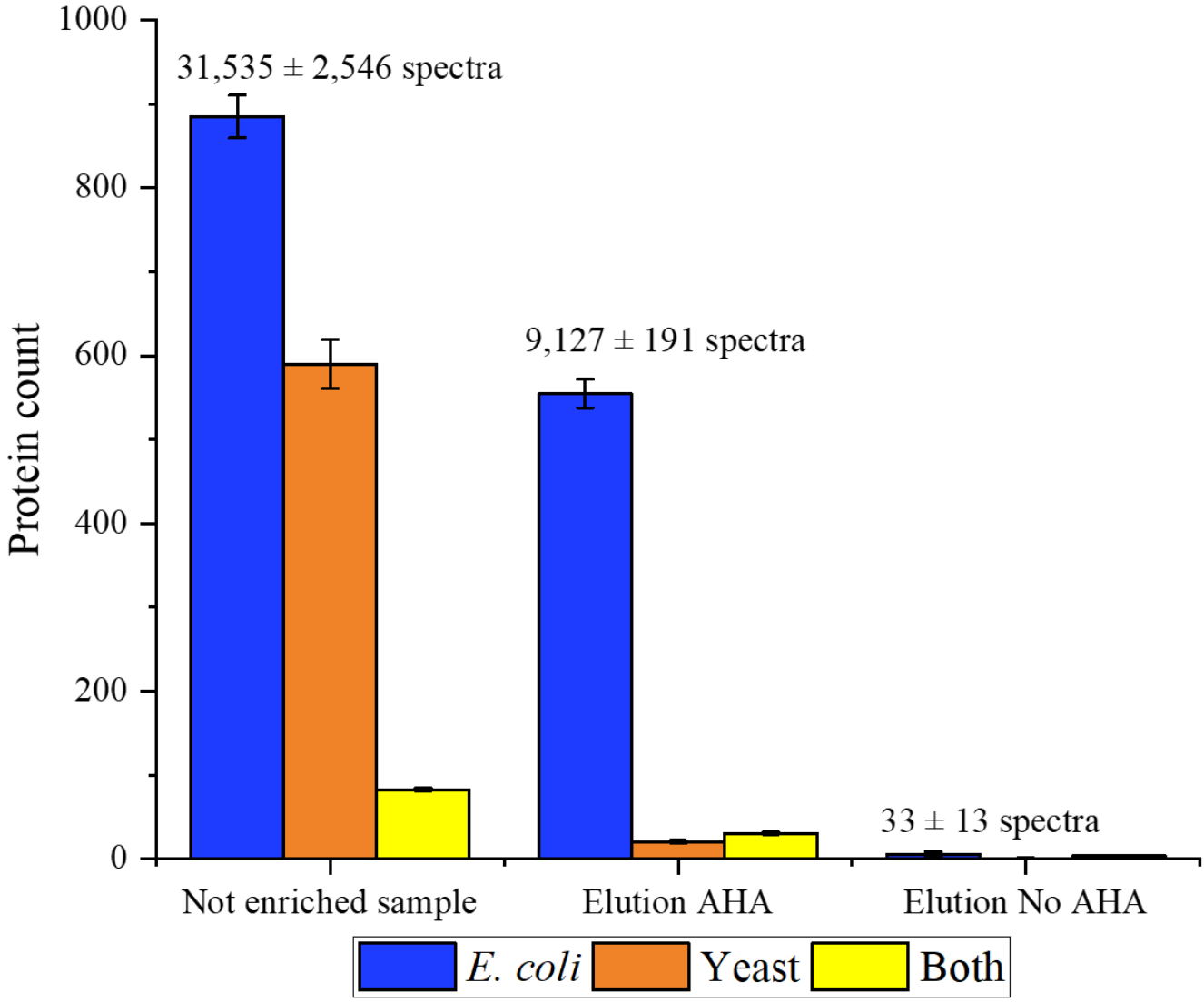


Figure 2: Number of identified proteins of the test system using the new enrichment method for nP. The number of proteins identified in the not enriched and AHA labelled sample ("Not enriched sample"), in the elution fraction of AHA labelled sample ("Elution AHA"), and the elution fraction without AHA labelling ("Elution No AHA"). All proteins were counted with at least two spectra. The proteins were grouped into the taxonomies *E. coli* and yeast. Proteins with shared peptides were grouped as "both". The bars represent the average of each group, and the error bars represent the standard derivation. The raw data can be found in Supplementary Table 2.

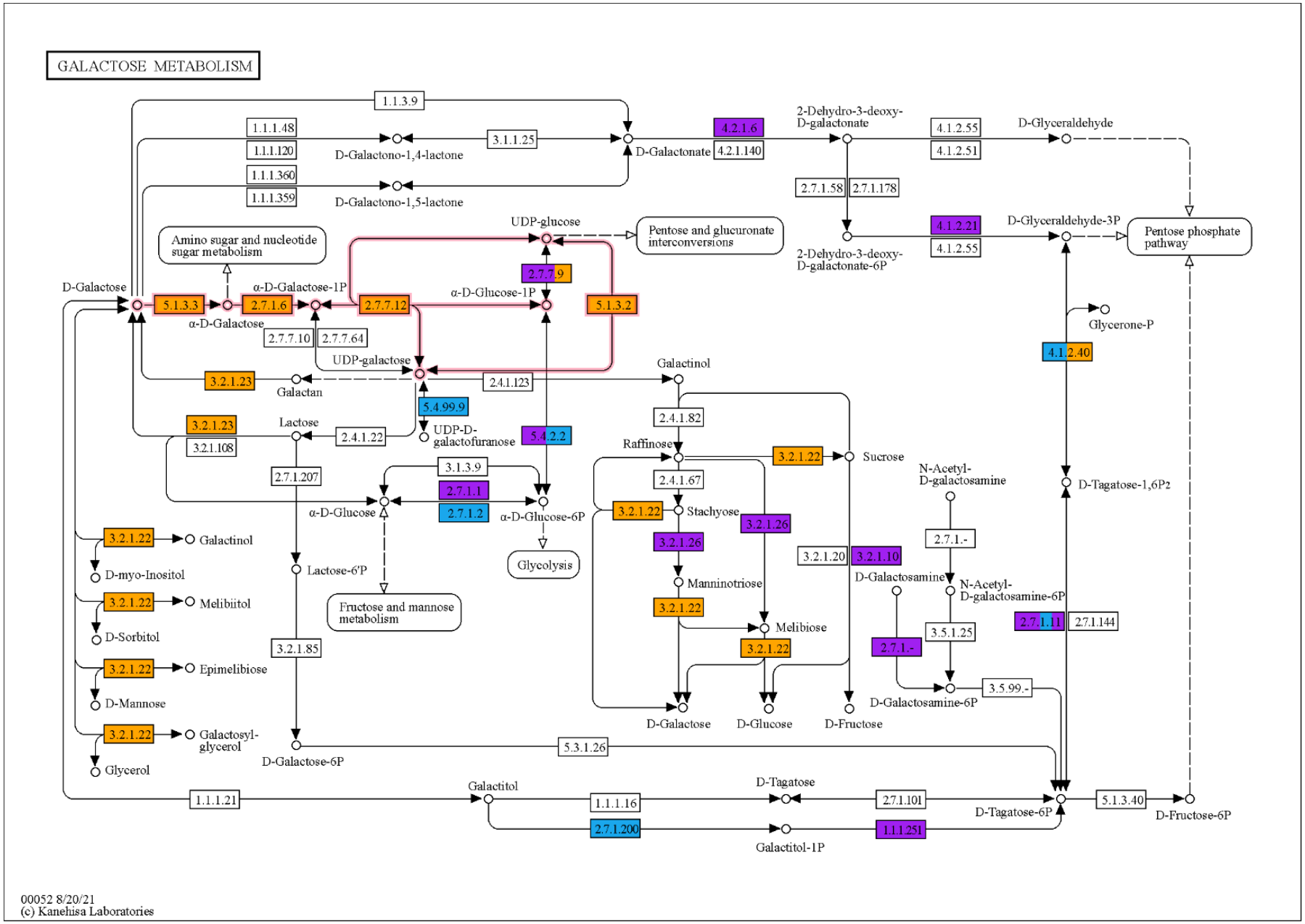

GALACTOSE METABOLISM
1.1.3.9
1.1.1.48
1.1.1.120
1.1.1.360
1.1.1.359
3.1.1.25
D-Galactono-1,4-lactone
D-Galactono-1,5-lactone
D-Galactonate
4.2.1.6
4.2.1.140
2-Dehydro-3-deoxy-D-galactonate
4.1.2.55
4.1.2.51
D-Glyceraldehyde
2.7.1.58
2.7.1.178
2-Dehydro-3-deoxy-D-galactonate-6P
4.1.2.21
4.1.2.55
D-Glyceraldehyde-3P
Pentose phosphate pathway
UDP-glucose
Pentose and glucuronate interconversions
Amino sugar and nucleotide sugar metabolism
2.7.7.9
D-Galactose
5.1.3.3
α-D-Galactose-1P
2.7.1.6
2.7.7.12
α-D-Glucose-1P
5.1.3.2
α-D-Galactose
2.7.7.10
2.7.7.64
UDP-galactose
Glycerone-P
4.1.2.40
3.2.1.23
Galactan
2.4.1.123
Galactinol
5.4.99.9
3.2.1.23
Lactose
2.4.1.22
UDP-D-galactofuranose
5.4.2.2
3.2.1.108
2.4.1.82
Raffinose
3.2.1.22
Sucrose
N-Acetyl-D-galactosamine
D-Tagatose-1,6P2
2.7.1.207
3.1.3.9
2.7.1.1
2.4.1.67
α-D-Glucose
2.7.1.2
α-D-Glucose-6P
3.2.1.22
Stachyose
3.2.1.26
3.2.1.22
Galactinol
3.2.1.26
2.7.1.-
D-myo-Inositol
Lactose-6'P
Glycolysis
3.2.1.20
3.2.1.10
D-Galactosamine
N-Acetyl-D-galactosamine-6P
Manninotriose
3.2.1.22
Melibiitol
Fructose and mannose metabolism
Melibiose
2.7.1.11
2.7.1.144
D-Sorbitol
3.2.1.85
3.2.1.22
3.5.1.25
2.7.1.-
3.2.1.22
Epimelibiose
3.5.99.-
D-Mannose
D-Galactosamine-6P
D-Galactose
D-Glucose
D-Fructose
3.2.1.22
Galactosyl-glycerol
5.3.1.26
Glycerol
D-Galactose-6P
1.1.1.21
Galactitol
1.1.1.16
D-Tagatose
2.7.1.101
D-Tagatose-6P
5.1.3.40
D-Fructose-6P
2.7.1.200
1.1.1.251
Galactitol-1P
00052 8/20/21
(c) Kanehisa Laboratories

Figure 3: KEGG map of all identified proteins from galactose metabolism in *E. coli* (map00052). **Orange** indicates proteins enriched at least 2-fold with the developed AA method. **Blue** shows the proteins with no significant difference in abundance between the AA method and not enriched sample. **Purple** represents all proteins unique identified in the not enriched sample. Boxes with several colours show similar proteins with different abundances for this step. **Pink arrows** highlight the conversion pathway of galactose to α-D-glucose-1-phosphate. Blank boxes were not identified in any sample. Sample data were normalized by dividing the number of spectra of each protein by the sum of the spectra for each sample. The abundance of each protein was averaged over three replicates, and the relative abundance ratio between the not enriched sample and elution was calculated for comparison [92–94].

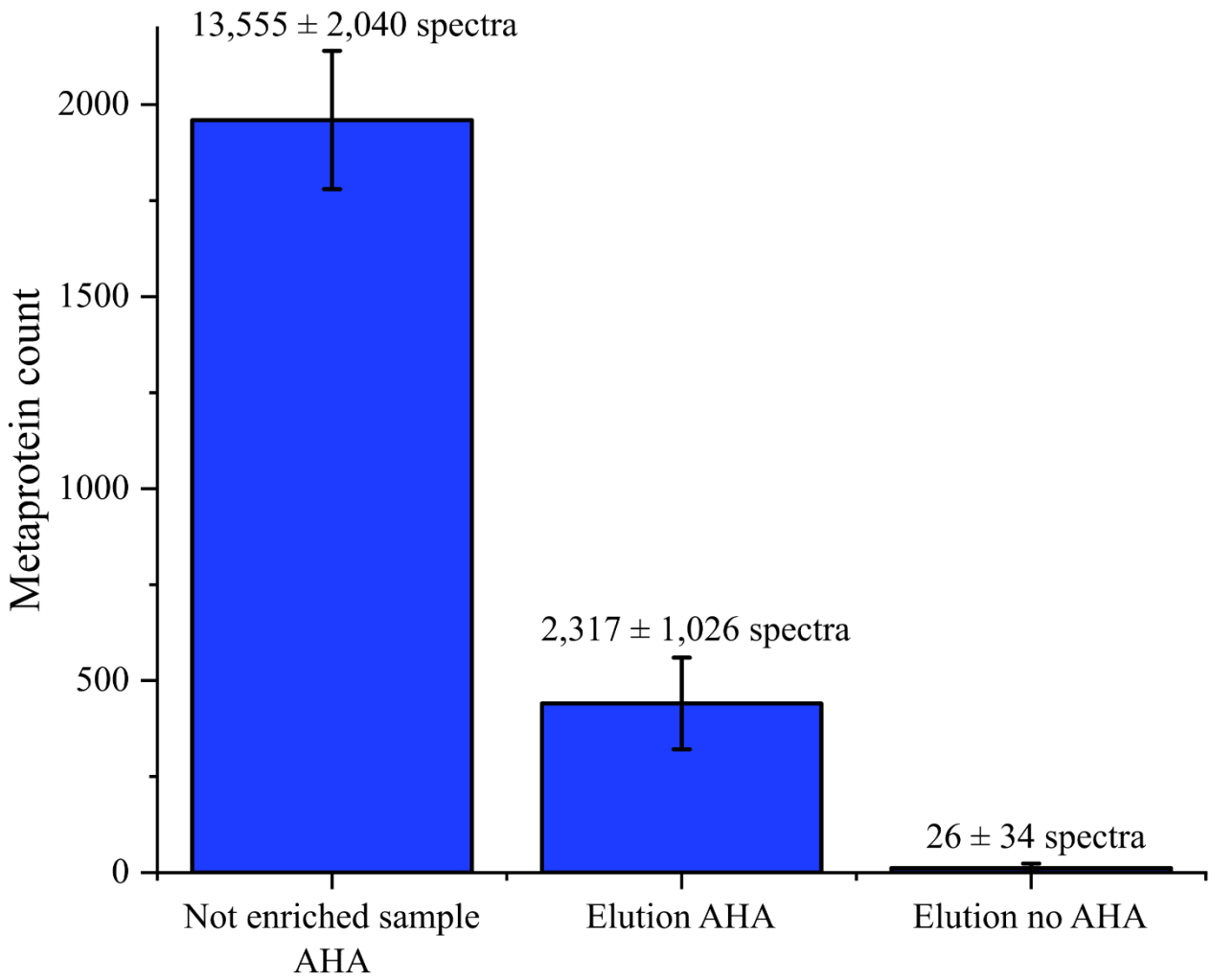


Figure 4: Number of identified metaproteins from AD using the AA method. The number of metaproteins identified in the not enriched and AHA labelled sample ("Not enriched sample"), in the elution fraction of AHA labelled sample ("Elution AHA"), and in the elution fraction without AHA labelling ("Elution No AHA"). The bars represent the average of each group, and the error bars represent the standard derivation. The raw data can be found in Supplementary Table 3.

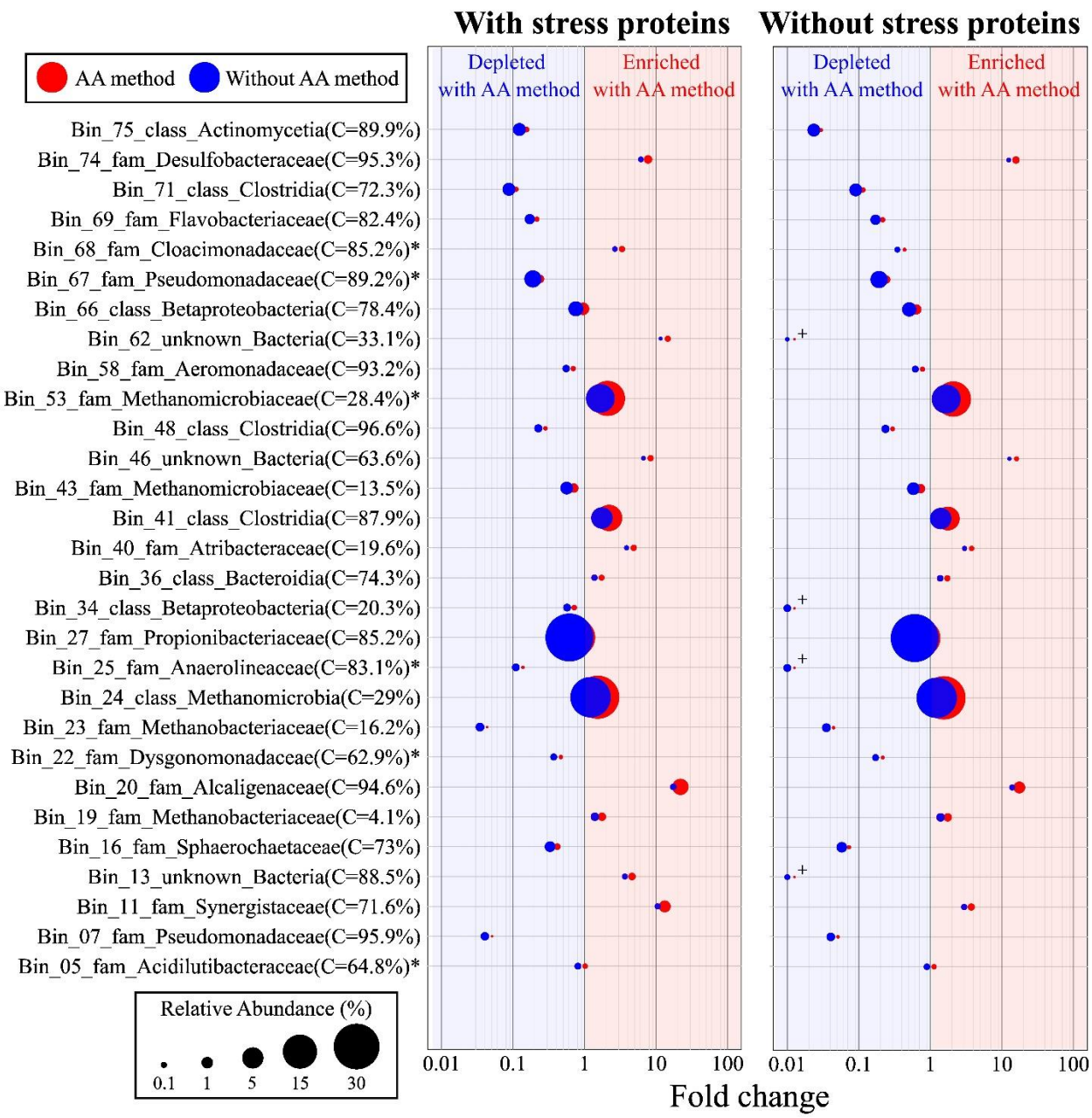
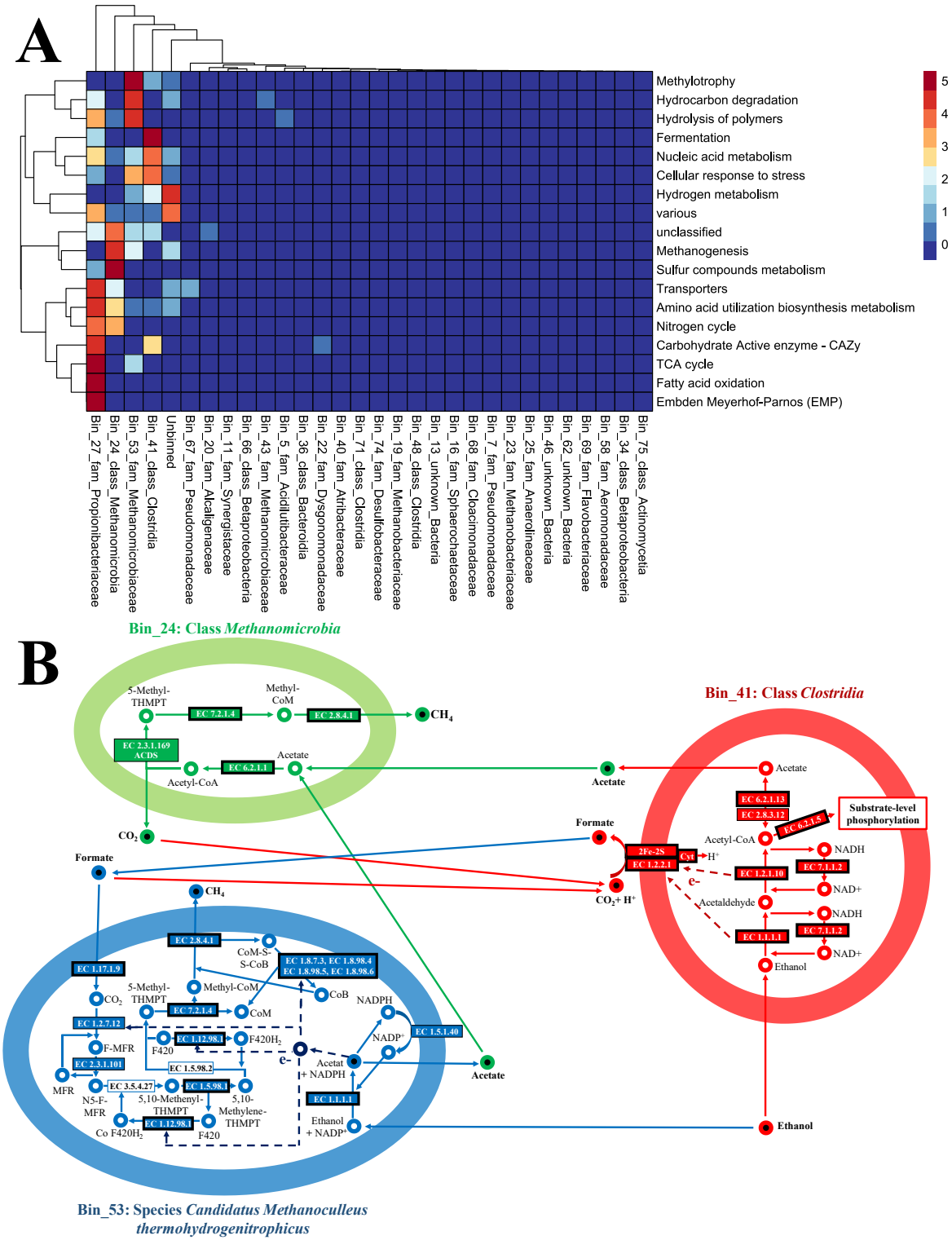


Figure 5: Bubble plot of the most abundant MAGs identified using the AA method. The 20 most abundant MAGs identified with the AA method and the 20 most abundant MAGs identified in the not-enriched sample were compared (a total of 29 different MAGs). The fold changes were calculated based on the ratio of the relative spectral abundance of the MAG in the two samples. The taxonomies were identified using the Type (Strain) Genome Server (see Supplementary Taxonomy report). The left bubble plot shows the MAG abundance based on all identified proteins and the right without stress proteins. C in brackets shows the completeness of the bins.

+ in the right bubble plot indicates that the MAG could no longer be identified. * are MAGs where the species are known: Bin_68 (*Candidatus Syntrophosphaera thermopropionivorans*), Bin_67 (*Halopseudomonas salegens*), Bin_53 (*Candidatus Methanoculleus thermohydrogenitrophicus*), Bin_5 (*Acidilutibacter cellobiosedens*), Bin_25 (*Candidatus Brevifilum fermentans*) and Bin_22 (*Fermentimonas caenicola*) (See also Supplementary Table 6).



497

498 Figure 6: Alternated biological processes in the MC due to ethanol addition (A) and proposed

499 syntrophic ethanol oxidation (B). **A:** The Functional Ontology Assignments for Metagenomes

500 based biological processes of all enriched nP and MAGs with the AA method were analysed

501 using a heatmap created with R Studio (2023.06.0 Build 421) and pheatmap (1.0.12). The data

were normalized using a z-score per row, based on the average relative spectral count. See also Supplementary Table 5 **B:** Intracellular metabolites are represented by unfilled circles, while extracellular metabolites are represented by filled circles. Metabolic pathways are designated by the colours of the arrows, which indicate the direction of metabolite reactions. Enzymes identified with the AA method are denoted by boxes on the arrows. The boxes with thicker lines indicate that the enzymes have been enriched at least twofold by the AA method. Exceptions include enzymes with EC numbers 1.5.98.2 and 3.5.4.27 that were not identified using the AA method or in the not enriched sample. The inspiration for this illustration is from Zhao et al. (2019) [95] Fig.7.

Supplementary

Supplementary note 1 material and methods

Supplementary note 2 additional figures

Supplementary table 1 lab scale biogas reactor

Supplementary table 2 protein data testsystem

Supplementary table 3 metaprotein data ethanol

Supplementary table 4 alcohol dehydrogenases

Supplementary table 5 biolog processes

Supplementary table 6 bubble plot

Supplementary table 7 protein quantification

Declarations

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement:

Proteome data were stored on PRIDE with the accession number: PXD047252, Username: reviewer_pxd047252@ebi.ac.uk, Password: 0kWF7k9n

The metagenome database for protein identification is stored on ENA under the accession number: PRJEB70937

Conflicts of Interest: Not applicable.

Author contributions (CRediT):

- Conceptualization: P.H., D.B.
- Project administration: P.H., D.B.
- Formal analysis: P.H., D.K., D.W.
- Funding acquisition: U.R., R.H.
- Investigation: P.H., D.K., D.W.
- Methodology: P.H., D.K., D.W., T.B., A.W.
- Project administration: D.B., U.R.
- Software: -
- Resources: D.B., U.R., R.H.
- Supervision: D.B., U.R.
- Validation: PH., D.K., D.W., T.B., A.W.
- Visualization: P.H., D.K.

- writing—original draft: P.H.
- writing—review, and editing: D.K., R.H., A.D., D.W., U.R., D.B., T.B., A.W.

All authors have read and agreed to the published version of the manuscript.

References

1 Finlay BJ, Clarke KJ. Ubiquitous dispersal of microbial species. *Nature* 1999;**400**:828.

2 Burman E, Bengtsson-Palme J. Microbial Community Interactions Are Sensitive to Small Changes in Temperature. *Front Microbiol* 2021;**12**:672910.

3 Suttle CA. The significance of viruses to mortality in aquatic microbial communities. *Microb Ecol* 1994;**28**:237–43.

4 Heyer R, Hellwig P, Maus I *et al.* Breakdown of hardly degradable carbohydrates (lignocellulose) in a two-stage anaerobic digestion plant is favored in the main fermenter. *Water Res* 2024;**250**:121020.

5 Waldrop MP, Firestone MK. Response of microbial community composition and function to soil climate change. *Microb Ecol* 2006;**52**:716–24.

6 Reid G, Younes JA, van der Mei HC *et al.* Microbiota restoration: natural and supplemented recovery of human microbial communities. *Nat Rev Microbiol* 2011;**9**:27–38.

7 Vrieze J de, Verstraete W. Perspectives for microbial community composition in anaerobic digestion: from abundance and activity to connectivity. *Environ Microbiol* 2016;**18**:2797–809.

8 Tringe SG, Mering C von, Kobayashi A *et al.* Comparative metagenomics of microbial communities. *Science* 2005;**308**:554–57.

9 Wilmes P, Bond PL. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* 2006;**14**:92–97.

10 Dieterich DC, Lee JJ, Link AJ *et al.* Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat Protoc* 2007;**2**:532–40.

11 Wouter van Bergen, Albert J.R. Heck, Marc P. Baggelaar. Recent advancements in mass spectrometry–based tools to investigate newly synthesized proteins. *Curr Opin Chem Biol* 2022;**66**:102074.

12 Dieterich DC, Link AJ, Graumann J *et al.* Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* 2006;**103**:9482–87.

13 Ignacio BJ, Dijkstra J, Mora N *et al.* THRONCAT: metabolic labeling of newly synthesized proteins using a bioorthogonal threonine analog. *Nat Commun* 2023;**14**:3367.

14 Kolb HC, Finn MG, Sharpless KB. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* 2001;**40**:2004–21.

15 Johnson JA, Lu YY, van Deventer JA *et al.* Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr Opin Chem Biol* 2010;**14**:774–80.

16 van Kasteren S, Rozen DE. Using click chemistry to study microbial ecology and evolution. *ISME Commun* 2023;**3**:9.

17 Bagert JD, Xie YJ, Sweredoski MJ *et al.* Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture. *Mol Cell Proteomics* 2014;**13**:1352–58.

18 Zhang G, Bowling H, Hom N *et al.* In-depth quantitative proteomic analysis of de novo protein synthesis induced by brain-derived neurotrophic factor. *J Proteome Res* 2014;**13**:5707–14.

19 Franco M, D'haeseleer PM, Branda SS *et al.* Proteomic Profiling of Burkholderia thailandensis During Host Infection Using Bio-Orthogonal Noncanonical Amino Acid Tagging (BONCAT). *Front Cell Infect Microbiol* 2018;**8**:370.

20 Hatzenpichler R, Scheller S, Tavormina PL *et al.* In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. *Environ Microbiol* 2014;**16**:2568–90.

21 Reichart NJ, Jay ZJ, Krukenberg V *et al.* Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. *ISME J* 2020;**14**:2851–61.

22 Hellwig P, Dittrich A, Heyer R *et al. Detection, isolation and characterisation of phage-host complexes using BONCAT and click chemistry*, 2024.

23 Pasulka AL, Thamatrakoln K, Kopf SH *et al.* Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environ Microbiol* 2018;**20**:671–92.

24 Landor LAI, Bratbak G, Larsen A *et al.* Differential toxicity of bioorthogonal non-canonical amino acids (BONCAT) in Escherichia coli. *J Microbiol Methods* 2023;**206**:106679.

25 Feng D, Guo X, Lin R *et al.* How can ethanol enhance direct interspecies electron transfer in anaerobic digestion? *Biotechnol Adv* 2021;**52**:107812.

26 Wessel D, Flügge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 1984;**138**:141–43.

27 Heyer R, Schallert K, Büdel A *et al.* A Robust and Universal Metaproteomics Workflow for Research Studies and Routine Diagnostics Within 24 h Using Phenol Extraction, FASP Digest, and the MetaProteomeAnalyzer. *Front Microbiol* 2019;**10**:1883.

28 Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970;**227**:680–85.

29 Maus I, Tubbesing T, Wibberg D *et al.* The Role of Petrimonas mucosa ING2-E5AT in Mesophilic Biogas Reactor Systems as Deduced from Multiomics Analyses. *Microorganisms* 2020;**8**.

30 Wibberg D, Andersson L, Tzelepis G *et al.* Genome analysis of the sugar beet pathogen Rhizoctonia solani AG2-2IIIB revealed high numbers in secreted proteins and cell wall degrading enzymes. *BMC Genomics* 2016;**17**:245.

31 Li D, Liu C-M, Luo R *et al.* MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;**31**:1674–76.

32 Hyatt D, Chen G-L, Locascio PF *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119.

33 Nguyen TH, Zucker J-D. Enhancing Metagenome-based Disease Prediction by Unsupervised Binning Approaches. In: . *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*. . IEEE, 1–5.

34 Berckx F, van Nguyen T, Bandong CM *et al.* A tale of two lineages: how the strains of the earliest divergent symbiotic Frankia clade spread over the world. *BMC Genomics* 2022;**23**:602.

35 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–59.

36 Kang DD, Li F, Kirton E *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019;**7**:e7359.

37 Manni M, Berkeley MR, Seppey M *et al.* BUSCO: Assessing Genomic Data Quality and Beyond. *Curr Protoc* 2021;**1**:e323.

38 Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 2019;**10**:2182.

39 Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL *et al.* TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 2022;**50**:D801-D807.

40 Muth T, Behne A, Heyer R *et al.* The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation. *J Proteome Res* 2015;**14**:1557–65.

41 Schiebenhoefer H, Schallert K, Renard BY *et al.* A complete and flexible workflow for metaproteomics data analysis based on MetaProteomeAnalyzer and Prophane. *Nat Protoc* 2020;**15**:3212–39.

42 Prestat E, David MM, Hultman J *et al.* FOAM (Functional Ontology Assignments for Metagenomes): a Hidden Markov Model (HMM) database with environmental focus. *Nucleic Acids Res* 2014;**42**:e145-e145.

43 Aramaki T, Blanc-Mathieu R, Endo H *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 2020;**36**:2251–52.

44 Kovalchuk SI, Anikanov NA, Ivanova OM *et al.* Bovine serum albumin as a universal suppressor of non-specific peptide binding in vials prior to nano-chromatography coupled mass-spectrometry analysis. *Anal Chim Acta* 2015;**893**:57–64.

45 Suelter CH, DeLuca M. How to prevent losses of protein by adsorption to glass and plastic. *Anal Biochem* 1983;**135**:112–19.

46 Kennell D, Riezman H. Transcription and translation initiation frequencies of the Escherichia coli lac operon. *J Mol Biol* 1977;**114**:1–21.

47 Roderick SL. The lac operon galactoside acetyltransferase. *C R Biol* 2005;**328**:568–75.

48 Zhao Y, Zhang W, Kho Y *et al.* Proteomic analysis of integral plasma membrane proteins. *Anal Chem* 2004;**76**:1817–23.

49 Clark DP. The fermentation pathways of Escherichia coli. *FEMS Microbiology Letters* 1989;**63**:223–34.

50 Cao H, Du Wei, Yang Y *et al.* Systems-level understanding of ethanol-induced stresses and adaptation in E. coli. *Sci Rep* 2017;**7**:44150.

51 Kornberg HL, Krebs HA. Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. *Nature* 1957;**179**:988–91.

52 Kornberg HL, Madsen NB. Synthesis of C4-dicarboxylic acids from acetate by a glyoxylate bypass of the tricarboxylic acid cycle. *Biochim Biophys Acta* 1957;**24**:651–53.

53 Dunn MF, Ramírez-Trujillo JA, Hernández-Lucas I. Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. *Microbiology (Reading)* 2009;**155**:3166–75.

54 Foss S, Heyen U, Harder J. Alcaligenes defragrans sp. nov., description of four strains isolated on alkenoic monoterpenes ((+)-menthene, alpha-pinene, 2-carene, and alpha-phellandrene) and nitrate. *Syst Appl Microbiol* 1998;**21**:237–44.

55 Kämpfer P, Denger K, Cook AM *et al.* Castellaniella gen. nov., to accommodate the phylogenetic lineage of Alcaligenes defragrans, and proposal of Castellaniella defragrans gen. nov., comb. nov. and Castellaniella denitrificans sp. nov. *Int J Syst Evol Microbiol* 2006;**56**:815–19.

56 Koch S, Kohrs F, Lahmann P *et al.* RedCom: A strategy for reduced metabolic modeling of complex microbial communities and its application for analyzing experimental datasets from anaerobic digestion. *PLoS Comput Biol* 2019;**15**:e1006759.

57 Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–89.

58 Varadi M, Anyango S, Deshpande M *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022;**50**:D439-D444.

59 Tang Y-Q, Matsui T, Morimura S *et al.* Effect of temperature on microbial community of a glucose-degrading methanogenic consortium under hyperthermophilic chemostat cultivation. *J Biosci Bioeng* 2008;**106**:180–87.

60 Poddar A, Lepcha RT, Das SK. Taxonomic study of the genus Tepidiphilus: transfer of Petrobacter succinatimandens to the genus Tepidiphilus as Tepidiphilus succinatimandens comb. nov., emended description of the genus Tepidiphilus and description of Tepidiphilus thermophilus sp. nov., isolated from a terrestrial hot spring. *Int J Syst Evol Microbiol* 2014;**64**:228–35.

61 Walter KA, Bennett GN, Papoutsakis ET. Molecular characterization of two Clostridium acetobutylicum ATCC 824 butanol dehydrogenase isozyme genes. *J Bacteriol* 1992;**174**:7149–58.

62 Kessler D, Leibrecht I, Knappe J. Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of Escherichia coli reside on a polymeric protein particle encoded by adhE. *FEBS Lett* 1991;**281**:59–63.

63 Paysan-Lafosse T, Blum M, Chuguransky S *et al.* InterPro in 2022. *Nucleic Acids Res* 2023;**51**:D418-D427.

64 Jones P, Binns D, Chang H-Y *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;**30**:1236–40.

65 Burton RM, Stadtman ER. The oxidation of acetaldehyde to acetyl coenzyme A. *J Biol Chem* 1953;**202**:873–90.

66 Fleischmann A, Darsow M, Degtyarenko K *et al.* IntEnz, the integrated relational enzyme database. *Nucleic Acids Res* 2004;**32**:D434-7.

67 Gómez-Manzo S, Escamilla JE, González-Valdez A *et al.* The oxidative fermentation of ethanol in Gluconacetobacter diazotrophicus is a two-step pathway catalyzed by a single enzyme: alcohol-aldehyde Dehydrogenase (ADHa). *Int J Mol Sci* 2015;**16**:1293–311.

68 Hitschler L, Nissen LS, Kuntz M *et al.* Alcohol dehydrogenases AdhE and AdhB with broad substrate ranges are important enzymes for organic acid reduction in Thermoanaerobacter sp. strain X514. *Biotechnol Biofuels* 2021;**14**:187.

69 Ramos AR, Grein F, Oliveira GP *et al.* The FlxABCD-HdrABC proteins correspond to a novel NADH dehydrogenase/heterodisulfide reductase widespread in anaerobic bacteria and involved in ethanol metabolism in Desulfovibrio vulgaris Hildenborough. *Environ Microbiol* 2015;**17**:2288–305.

70 Wawrik B, Marks CR, Davidova IA *et al.* Methanogenic paraffin degradation proceeds via alkane addition to fumarate by 'Smithella' spp. mediated by a syntrophic coupling with hydrogenotrophic methanogens. *Environ Microbiol* 2016;**18**:2604–19.

71 Zhang B, Lingga C, Bowman C *et al.* A New Pathway for Forming Acetate and Synthesizing ATP during Fermentation in Bacteria. *Appl Environ Microbiol* 2021;**87**:e0295920.

72 Schuchmann K, Chowdhury NP, Müller V. Complex Multimeric FeFe Hydrogenases: Biochemistry, Physiology and New Opportunities for the Hydrogen Economy. *Front Microbiol* 2018;**9**:2911.

73 Frimmer U, Widdel F. Oxidation of ethanol by methanogenic bacteria. *Arch. Microbiol.* 1989;**152**:479–83.

74 Song J, Wang J, Wang X *et al.* Improving the Acetic Acid Fermentation of Acetobacter pasteurianus by Enhancing the Energy Metabolism. *Front Bioeng Biotechnol* 2022;**10**:815614.

75 Berk H, Thauer RK. Function of coenzyme F420-dependent NADP reductase in methanogenic archaea containing an NADP-dependent alcohol dehydrogenase. *Arch. Microbiol.* 1997;**168**:396–402.

76 Maus I, Wibberg D, Stantscheff R *et al.* Complete genome sequence of the hydrogenotrophic, methanogenic archaeon Methanoculleus bourgensis strain MS2(T), Isolated from a sewage sludge digester. *J Bacteriol* 2012;**194**:5487–88.

77 Kougias PG, Campanaro S, Treu L *et al.* A novel archaeal species belonging to Methanoculleus genus identified via de-novo assembly and metagenomic binning process in biogas reactors. *Anaerobe* 2017;**46**:23–32.

78 Thiele JH, Zeikus JG. Control of Interspecies Electron Flow during Anaerobic Digestion: Significance of Formate Transfer versus Hydrogen Transfer during Syntrophic Methanogenesis in Flocs. *Appl Environ Microbiol* 1988;**54**:20–29.

79 Manzoor S, Schnürer A, Bongcam-Rudloff E *et al.* Complete genome sequence of Methanoculleus bourgensis strain MAB1, the syntrophic partner of mesophilic acetate-oxidising bacteria (SAOB). *Stand Genomic Sci* 2016;**11**:80.

80 Singh A, Schnürer A, Westerholm M. Enrichment and description of novel bacteria performing syntrophic propionate oxidation at high ammonia level. *Environ Microbiol* 2021;**23**:1620–37.

81 Reddy CA, Bryant MP, Wolin MJ. Characteristics of S organism isolated from Methanobacillus omelianskii. *J Bacteriol* 1972;**109**:539–45.

82 Bryant MP, Wolin EA, Wolin MJ *et al.* Methanobacillus omelianskii, a symbiotic association of two species of bacteria. *Arch Mikrobiol* 1967;**59**:20–31.

83 Morris BEL, Henneberger R, Huber H *et al.* Microbial syntrophy: interaction for the common good. *FEMS Microbiol Rev* 2013;**37**:384–406.

84 Seitz H-J, Schink B, Pfennig N *et al.* Energetics of syntrophic ethanol oxidation in defined chemostat cocultures. *Arch. Microbiol.* 1990;**155**:82–88.

85 Summers ZM, Fogarty HE, Leang C *et al.* Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. *Science* 2010;**330**:1413–15.

86 Keller A, Schink B, Müller N. Alternative Pathways of Acetogenic Ethanol and Methanol Degradation in the Thermophilic Anaerobe Thermacetogenium phaeum. *Front Microbiol* 2019;**10**:423.

87 Khesali Aghtaei H, Püttker S, Maus I *et al.* Adaptation of a microbial community to demand-oriented biological methanation. *Biotechnol Biofuels Bioprod* 2022;**15**:125.

88 Gómez-Varela D, Xian F, Grundtner S *et al.* Increasing taxonomic and functional characterization of host-microbiome interactions by DIA-PASEF metaproteomics. *Front Microbiol* 2023;**14**:1258703.

89 Debets MF, van Berkel SS, Schoffelen S *et al.* Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem Commun (Camb)* 2010;**46**:97–99.

90 Kuzmin A, Poloukhtine A, Wolfert MA *et al.* Surface functionalization using catalyst-free azide-alkyne cycloaddition. *Bioconjug Chem* 2010;**21**:2076–85.

91 Yao JZ, Uttamapinant C, Poloukhtine A *et al.* Fluorophore targeting to cellular proteins via enzyme-mediated azide ligation and strain-promoted cycloaddition. *J. Am. Chem. Soc.* 2012;**134**:3720–28.

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

93 Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 2019;**28**:1947–51.

94 Kanehisa M, Furumichi M, Sato Y *et al.* KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 2023;**51**:D587-D592.

95 Zhao Z, Wang J, Li Y *et al.* Why do DIETers like drinking: Metagenomic analysis for methane and energy metabolism during anaerobic digestion with ethanol. *Water Res* 2020;**171**:115425.