Tracing active members 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

ureichl@mpi-magdeburg.mpg.de

benndorf@mpi-magdeburg.mpg.de

Keywords: Metaproteomics; mass spectrometry; BONCAT; click chemistry; microbiome; microbial activity; anaerobic digestion; biogas; microbial interactions; Metagenomics

Running title: Microbial Protein Synthesis Tracking

Abstract (196 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 an important 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 and has therefore not been used on microbial communities before. To address this, a new workflow with efficient and specific 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.

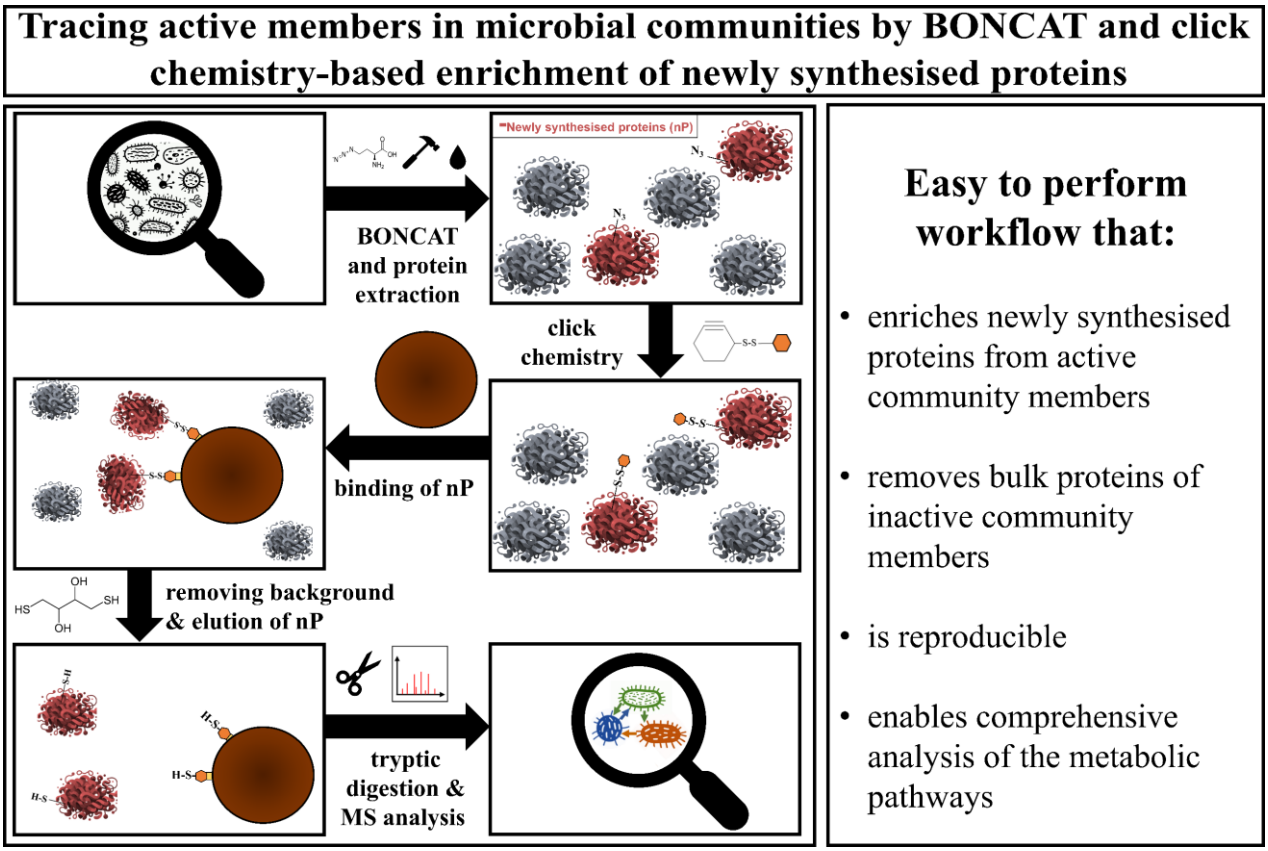
53 Abbreviations:

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
SIP	stable isotope labeling
SOP	standard operation procedure
SS	disulfide
TYGS	Type (Strain) Genome Server

54

55 **Graphical Abstract:**



56

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]. Due to its immense importance in the analysis and characterization of MC gained prominence in the last decades.

Metagenomics is currently the most widely used method to characterise MCs. It involves the simultaneous genomic analysis of all members of an MC and provides insight into the taxonomic composition and metabolic potential of an MC [8]. However, only in combination with other meta-omics techniques such as metaproteomics, which analyses all proteins of an MC, it is possible 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.

The standard method of tracking nP is stable isotope labeling (SIP). In protein-SIP stable isotopes, such as ^{13}C, ^{15}N, and ^{2}H (deuterium), are introduced into the cultivation media. These isotopes are incorporated into amino acids and the modified amino acids will be incorporated in nP [10, 11]. This allows a distinction between labeled nP from non-labeled proteins. Alternatively, advances in bioorthogonal chemistry have enabled another way of identification of nP. One technique is bioorthogonal non-canonical amino acid tagging (BONCAT), which

allows the selective labelling of nP [12]. In BONCAT, analog non-canonical amino acids (ncAA) are added to cells or MCs. During protein synthesis, the tRNA-amino acyl synthetase of an amino acid mistakenly selects the ncAA instead of the canonical amino acid. 4-Azido-L-homoalanine (AHA) and L-Homopropargylglycine (HPG). Both are the ncAAs replacing methionine [13]. Note, incorporation of AHA and HPG does not depend on genetic modification of tRNAs in the uptaking cells. Studies in *E. coli* showed that the likelihood of incorporation is 1:390 for AHA and 1:500 for HPG [13]. 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 [14]. This amino acid should therefore be avoided. However, Ignacio et al. (2023) [15] have recently developed a method called THRONCAT, in which ncAA incorporation is not affected by external canonical amino acid concentrations, which could significantly expand the range of applications for ncAA labeling in the future. The 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 so far 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 [16, 17]. Thus, a fluorescence or biotin marker can be attached to the AHA (azide group) or HPG (alkyne group) [18]. BONCAT and CC have previously been applied to visualize and enrich nP from eukaryotic cells, such as neurons [14], HeLa cells [19], or HEK cells [20] and from viruses and their host cells [21]. However, BONCAT combined with CC has been used mainly in cell culture experiments with only one cell type. Hatzenpichler et al. [22] showed that the application of BONCAT in MC is possible and discussed the advantages and disadvantages of BONCAT and its application to MC in great detail [23]. Additionally, Reichart et al. [24] 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 [25, 26].

In the case of metaproteomics, the detection of nP is of paramount importance for the understanding of MC. However, the currently available enrichment methods for nP are designed for 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 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 enrichment of nP from growing MC for the detection *via* MS. The main focus was on the enrichment of AHA-labelled nP since the conditions of BONCAT have to be individually adapted to each experimental condition and MC in particular to the dynamics of protein translation [22, 27]. Therefore, we developed a test system for nP enrichment. 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 [28].

Material and Methods

In the following we briefly describe the workflow established (Figure 1). A detailed description of the methods, including a step-by-step standard operation procedure (SOP), can be found in Supplementary Note 1. All cultivation experiments were performed in triplicates.

AHA-labeling of *E. coli* nP

The carbon source of an E. coli (DSM 5911) culture grown aerobically at 37°C in M9 media was switched from glucose to lactose while the cells were incubated with 100 µM AHA (sterile filtrated). The same substrate shift 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 (see SOP 1).

AHA-labeling of nP from MC of LBR

50 mL of cell suspensions from an LBR MC (1 L total volume, 40°C, 75 rpm, pH 7.4 – 7.5, Supplementary Table 1) was sampled using a 50 mL syringe. The cell suspensions were then transferred to nitrogen-flushed 100 mL serum bottles, diluted 1:1 with 50 mL anaerobic medium without glucose (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.

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 of silica beads and 0.5 µL/mL of Cyanase™

Nuclease solution. The cells were lysed in a ball mill, incubated at 37°C for 15 minutes, and then centrifuged. The resulting supernatant was transferred to a new reaction tube(see SOP 1).

Protein extraction with acetone (for pure cultures)

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 followed by centrifugation. 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 (see SOP 1).

Protein extraction with methanol and chloroform (for environmental samples)

The MC proteins were extracted using a mixture of methanol, chloroform, and water, followed by centrifugation [29]. 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.

Protein quantification

Amido black assay was used to quantify the protein concentration of each sample, as described earlier [30].

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 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 [30, 31]. (see SOP 5).

Click chemistry with biotin linker

Based on the fluorescence signal (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 (see SOP 2).

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. The beads were blocked with bovine serum albumin (BSA), amino acids (AA), or left blocked.

FASP digestion

The tryptic digestion was performed using FASP [30] with few modifications (see SOP 4). The proteins contained in the dithiothreitol (DTT) elution phase and in the elution phase 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. For the enriched nP fractions, 100 ng trypsin was used. 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.

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

Three samples of the LBR used in this study were collected in different weeks for metagenomic sequencing to generate a protein database for the metaproteomic analysis in this study. The

samples from the LBR were sequenced independently using PCR-free libraries prepared with the Illumina TruSeq® DNA PCR-free kit [32]. 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 [33], followed by assembly with Megahit (v1.1.1) [34] and gene prediction with Prodigal v.2.6.0 [35]. Metagenomic binning [36, 37] was performed with Bowtie 2 [38] and MetaBAT2 (76 Metagenome assembled genome (MAGs)) [39], while completeness and contamination were assessed with BUSCO (v5.7.0) [40]). The translated amino acid sequences of the predicted genes were generated and the replicates were merged for a protein database generation with Contig informations. For the most abundant MAGs, the MAG sequence was uploaded to the Type (Strain) Genome Server (TYGS) to evaluate the taxonomy [41, 42].

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 [43]).

Further analysis of the protein data

The protein data were uploaded to Prophane (version 6.2.6) [44] 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 [45], KofamKOALA [46]

was used to identify possible missing KO numbers (Supplementary Table 3 and Supplementary Note 1).

Results and Discussion

Development of a nP enrichment workflow with the test system

Prior to developing the enrichment protocol, the *E. coli* (AHA-labelled) and yeast (unlabelled) test system was refined by optimising the AHA labelling of *E. coli,* measured by detecting the proportion of fluorescent *E. coli* cells generated through click chemistry. (Supplementary Figure 1). Attempts were made to enrich nP of *E. coli* from the test system with known enrichment protocols (e.g., [20, 21, 47]) 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 highlighted the need for a test system to monitor nP enrichment, and the importance of implementing 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 and the use of Dynabeads™ MyOne™ Streptavidin C1 beads (Supplementary Figure 2). The nP were eluted under mild conditions from streptavidin beads 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 (Supplementary Figure 2).

Therefore, washing was optimized. An adapted washing strategy [20] and the use of phosphate-buffered saline (PBS) buffer with increased NaCl concentration (237 mM) reduced the amount of background yeast proteins. However, it strongly reduced the reproducibility of nP elution

(Figure 2 A). For higher reproducibility of enrichment, the loss of nP during elution should be minimized.

Blocking the beads with 1 mg/mL BSA (e.g. [48, 49]), before loading the samples, reduced the amount of yeast protein background further but unfortunately also increased the variability in nP elution (Figure 2 A). Additionally, the elution phase 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) [50], the beads were blocked with an AA blocking solution containing 16.67 µg/mL of leucine (aliphatic hydrophobic), tryptophan (aromatic hydrophobic), histidine (polar positive), glutamine (polar neutral), glutamic acid (polar negative) and glycine (nonpolar neutral) each. These AAs should not cause steric effects and should prevent any unspecific interactions with the beads. This method ("AA method") maintained a low background and exhibited minimal variability and loss in nP elution (Figure 2 A), 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) were identified in the AHA-labelled but non-enriched test system ("not enriched sample") (Figure 2 B). Over a third of these proteins were yeast proteins. After nP enrichment, 607 ± 16 proteins (555 ± 17 *E. coli*, 21 ± 2 yeast) were identified, reducing the amount of identified yeast proteins by over 96% while retaining 63% of the *E. coli* proteins (Figure 2 B). 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).

Furthermore, the AA method allowed the detection of 48 proteins that were absent in the not enriched samples (Supplementary Table 2). Of these, 29 were *E. coli* proteins that 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 Figures 3 and 4). Consequently, the selected elution with DTT is not only specific but also highly efficient.

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 (*lacA*), and β-galactoside permease (*lacY*) [51, 52]. Indeed, the AA method successfully enriched β-galactosidase and galactoside O-acetyltransferase by a relative foldchange of 3.7. However, β-galactoside permease was not detected, neither in the enriched sample nor in the not enriched sample. As it is an integral membrane protein it was probably lost during protein extraction [53].

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 and at least 2-fold enriched, 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) (Figure 3). Glycolytic enzymes were also identified; however, they did not exhibit a 2-fold increase in the enriched nP compared to the not enriched sample. This might be explained by the fact, that cells were already grown on glucose as substrate before the substrate shift (Figure 3). Furthermore, all dehydrogenases involved in the mixed acid fermentation of *E. coli* [54] were enriched after the substrate shift by the AA method(Supplementary Table 2). 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.

Enrichment of nP in anaerobic MC of an LBR using the AA method

The next objective was to demonstrate that the established workflow is also effective for weaker AHA-labelled and more complex 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 (Supplementary Figure 5). 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 (Supplementary Figure 5). Therefore, this replicate was excluded from further analysis. The other two replicates were processed with the AA method and subsequent metaproteome analysis.

To achieve more accurate identification of proteins within the LBR MC, a metagenome was sequenced from the LBR, and a corresponding protein database was generated. 76 MAGs were binned.

In total, 2063 (replicate 1) and 2065 (replicate 2) proteins were identified in the not enriched sample of the AHA-labelled MC (“Not enriched sample AHA”), while 525 (replicate 1) and 356 (replicate 2) proteins 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 7 (replicate 1) and 25 (replicate 2) proteins (Figure 4 A). Notably, with the AA method, 26 proteins were identified (both replicates), which were absent in the not enriched sample (Figure 4 B). Consequently, the specific enrichment of nP should

provide a better insight into the potential conversion of ethanol to methane in this LBR than the proteome analysed without enrichment.

Ethanol-associated taxa based on nP enrichment

Initially, the enrichment of the nP was considered for the 20 most abundant MAGs of the AA method and without the AA method each based on their protein abundance divided by the MAG length (29 MAGs in total, Supplementary Table 6). The ratio of the MAGs between the AA method and without the AA method was calculated based on their protein abundance. The results of the two replicates demonstrate a consistent pattern of variation in protein abundance among the MAGs under investigation. 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 protein abundance in all samples. The protein abundance of Bin_27 has been depleted by the AA method, whereas the proteins of the other three MAGs are similarly abundant in all 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 protein abundance, they do exhibit strong protein 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 in their protein abundance after using the AA method (Figure 5).

Upon closer examination of the protein functions from the enriched MAGs, a notable abundance of stress proteins (up to 100% in some MAGs) was observed. Although ethanol can have a positive effect on methanogenesis in AD [55], the addition of ethanol in combination with the transfer of the MC from continuous LBR to serum bottles and mixing with a medium may have induced the expression of these stress proteins [56, 57]. Consequently, stress proteins

were excluded from further analyses to ensure a metabolic enrichment of MAGs (Figure 5). In the case of the high-abundant MAGs, the protein 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*) showed more than 10-fold protein enrichment using the AA method. In contrast, proteins from 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 the stress-induced proteins were removed from the analysis. Therefore, the addition of ethanol resulted in increased expression of metabolic proteins in Bin_74, Bin_46, and Bin_20, whereas MAGs such as Bin_13, Bin_34, Bin_25, and Bin_62 primarily synthesized stress proteins.

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 expressed gluconeogenesis proteins. Too few proteins could be identified from Bin_46 to make conclusions about its metabolism. A review of the individual 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 [58–62]. The incubation period for these three strains in the presence of ethanol was likely insufficient for the strains to multiply sufficiently to allow for the detection of these proteins. Without enrichment with the AA method, they would thus not have been the focus of analyses. In follow-up studies, the incubation times with ethanol could be extended so that also the metabolisms can be analysed [63].

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 (Supplementary Figure 6). 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 of both replicates (Figure 6 and Supplementary Figure 7).

A detailed analysis of proteomes 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, only in replicate 1)) were increased (≥ 2.5-fold) after enrichment (Supplementary Table 3 and 4). An analysis with Alphafold [64, 65] 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 [66, 67]. The abundance of this MAG is however insufficient to definitively ascertain its metabolic processes in this MC. 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 [68], except an aldehyde dehydrogenase, which is necessary for the further degradation of acetaldehyde [69–71]. 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 8) [72–75]. Bifunctionality poses a challenge in the annotation of KO numbers, as normally only one KO number is assigned to one protein. Note, that it was only through enrichment of nP 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 [76]. 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 [77]. Additionally, ATP appeared to be generated

406 through substrate chain phosphorylation [77, 78] (Supplementary Table 4). However, a major
407 portion of stored energy is probably consumed for the reduction of formate due to coupled
408 proton influx [79].

409 Bin_53 (species *Candidatus Methanoculleus thermohydrogenitrophicus*) is closely related to
410 *Methanoculleus*, which is known to use ethanol and secondary alcohols as electron donors for
411 methanogenesis [80]. Ethanol is converted into acetate by a reaction that is coupled with
412 NADP+ reduction, which is then recovered by NADPH oxidoreductase (EC 1.5.1.40) [81, 82]
413 delivering reduction equivalents to hydrogenotrophic methanogenesis (Supplementary Figure
414 9). Since ethanol is not completely oxidized to CO_2, an alternative source of CO_2 is required
415 for hydrogenotrophic methanogenesis [83, 84]. A formate dehydrogenase (EC 1.17.1.9) was
416 identified that converts formate into CO_2, which can then be used for hydrogenotrophic
417 methanogenesis and feeds released reduction equivalents into hydrogenotrophic
418 methanogenesis. *Methanoculleus* is also known to use formate [84]. A syntropic interaction
419 between Bin_53 and Bin_41 is plausible with Bin_41 delivering formate to Bin_53 [85–87].
420 Furthermore, Bin_53 also should be able to consume H_2 which could result in the inhibition of
421 Bin_41 when accumulating in the medium [88–90]. But syntrophic transfer of H_2 probably did
422 not play a major role in the LBR MC since no hydrogenase capable of releasing H_2 was detected
423 in Bin_41, (Figure 7). The phenomenon of syntrophic ethanol oxidation between fermenting
424 bacteria and methanogenic archaea observed here has been repeatedly observed in anaerobic
425 digesters [28, 91–93].

426 In Bin_24, all proteins necessary for acetoclastic methanogenesis were identified, indicating its
427 potential to convert the acetate produced in Bin_41 and Bin_53 into methane and CO_2
428 (Supplementary Figure 10 and Figure 7). Therefore, based on the enriched nP, a syntrophic
429 relationship between these three MAGs is highly probable. In this scenario, ethanol is converted
430 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 [94], and additional partners involved in ethanol oxidation (e.g. by minor abundant bins) cannot be excluded (Supplementary Figures 11 and 12).

Further Improvements and important considerations for implementing the workflow

A successful BONCAT experimentis the precondition for this workflow. Our experiments demonstrated the sufficient efficacy of BONCAT in slow-growing anaerobic MC. We advise reading Hatzenpichler et al [22, 23, 95] and Landor et al. [27] as they address pivotal aspects of a successful BONCAT experiment and the potential cytotoxicity of BONCAT. Following a successful BONCAT, our workflow can be readily implemented in any laboratory (SOPs in Supplementary Note 1). In the case of suboptimal nP yields despite a successful BONCAT experiment, 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® [96]. In addition, fluorescence-activated cell sorting of ncAA-labelled cells from MC [24] could generate corresponding metagenomes for improved protein identification. Alternatively, the workflow could be adapted to HPG, pyrrolysine, or β-ethynylserine, although this has not yet been tested but the same biotin linker is also available for copper-based click chemistry [15, 97].

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.

Acknowledgment

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

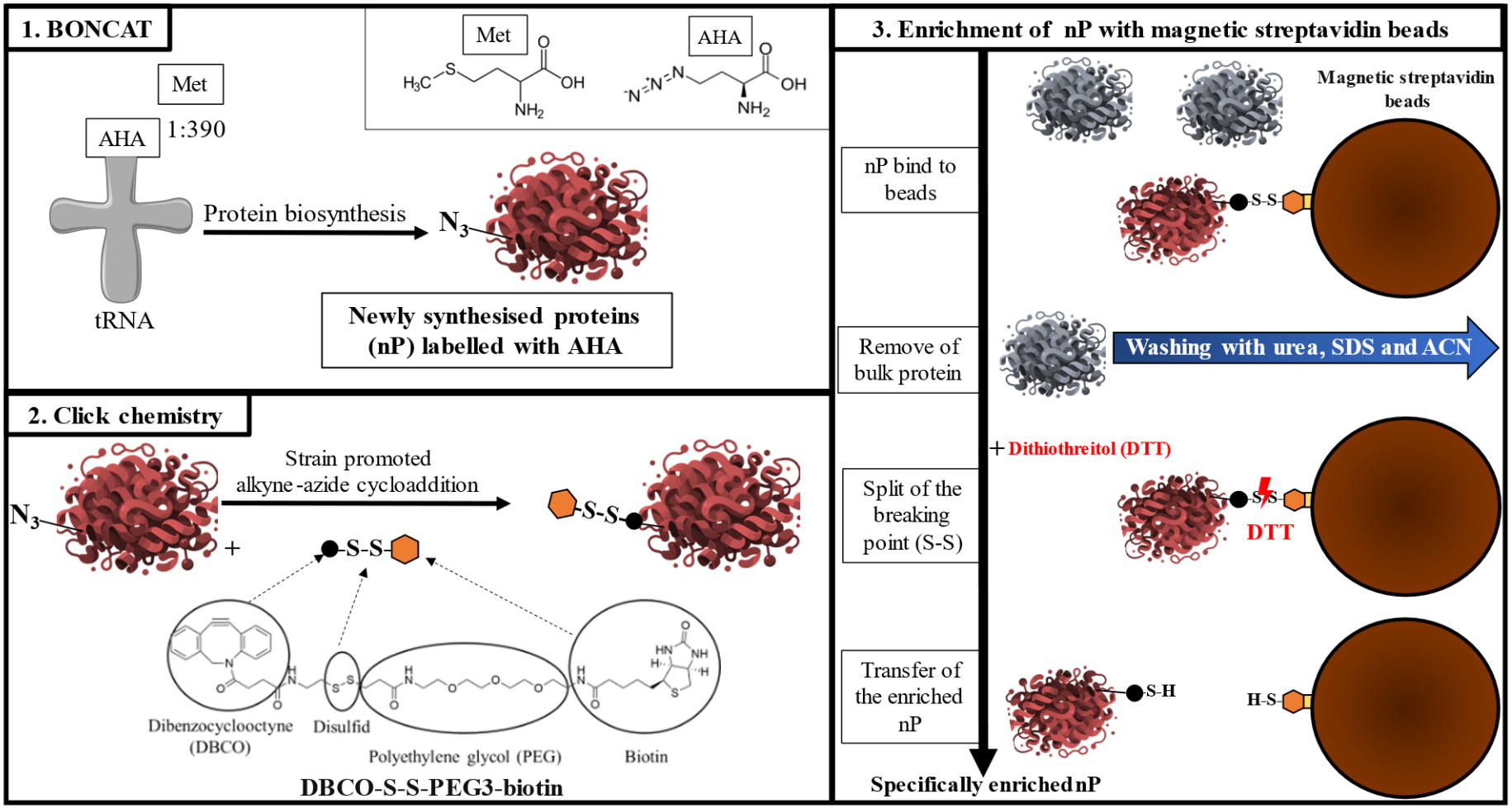
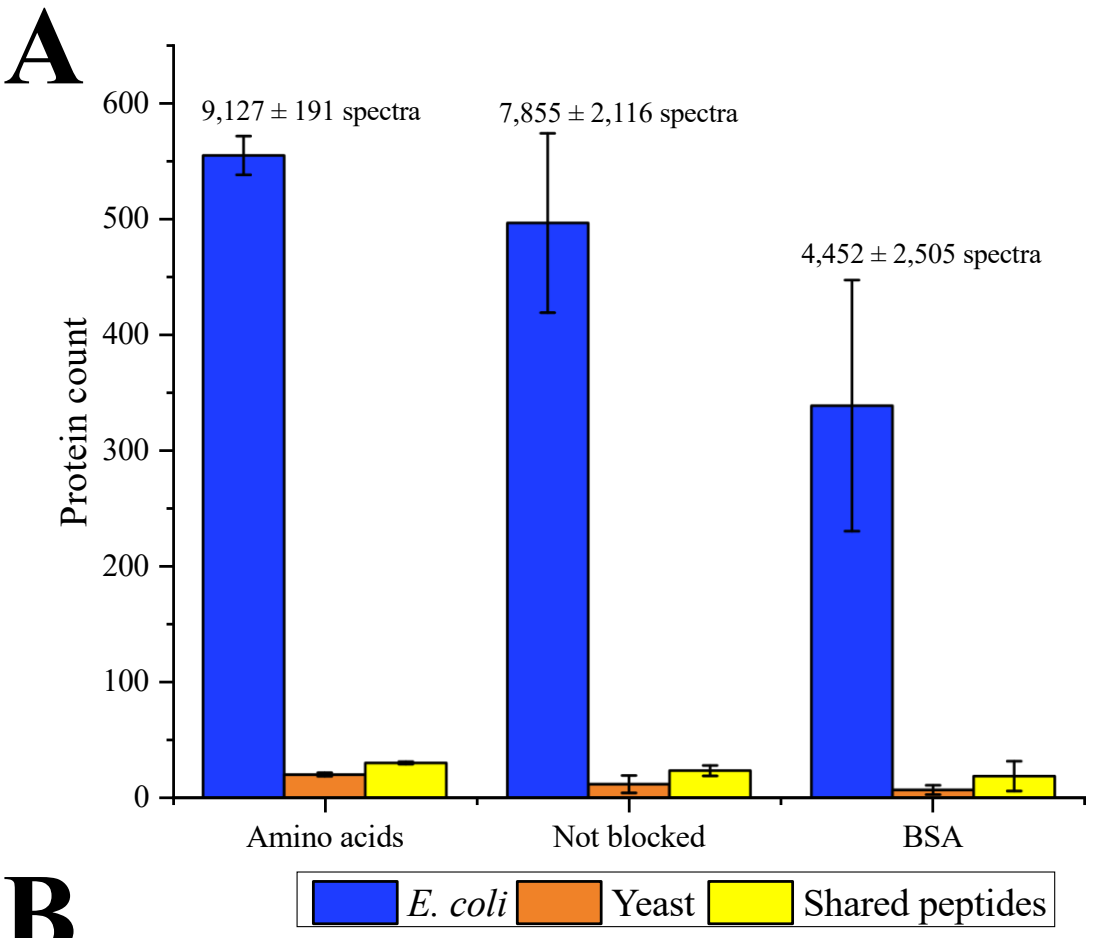


Figure 1: Workflow of the enrichment of newly synthesized proteins. **1.:** nP were labelled with AHA. **2.:** Free cysteines of the proteins were blocked with IAA (not shown). The labelled nP were tagged with DBCO-S-S-PEG$_3$-biotin within 1 h. **3.:** The biotin-tagged nP were bound to magnetic

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.


A
Protein count
600
500
400
300
200
100
0
9,127 ± 191 spectra
7,855 ± 2,116 spectra
4,452 ± 2,505 spectra
Amino acids
Not blocked
BSA
E. coli
Yeast
Shared peptides

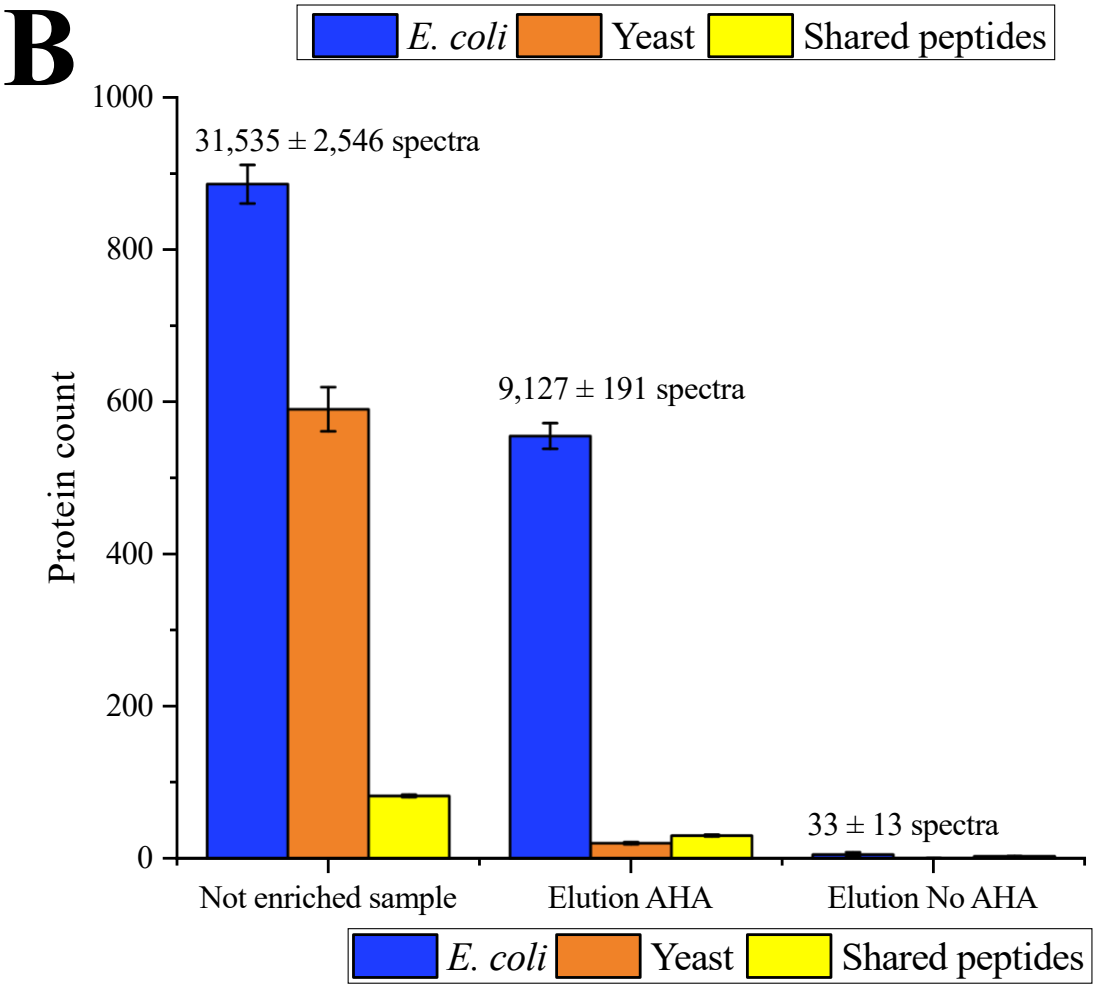

B
Protein count
1000
800
600
400
200
0
31,535 ± 2,546 spectra
9,127 ± 191 spectra
33 ± 13 spectra
Not enriched sample
Elution AHA
Elution No AHA
E. coli
Yeast
Shared peptides

Figure 2: Number of identified proteins of the test system using the new enrichment method for nP. **A:** Comparison of the number of identified proteins from the test system between the tested methods for nP enrichment. All proteins identified in the elution with AHA were compared between the three enrichment methods: blocking with amino acids, blocking with BSA, and no blocking. **B:** The different fractions of nP enrichment with the AA method were compared. 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 in the elution fraction without AHA labelling ("Elution No AHA") are shown. All proteins with at least two spectra were considered. The proteins were grouped into the taxonomies *E. coli* and yeast. Proteins with shared peptides were grouped as "Shared peptides". The bars represent the average of each group, and the error bars represent the standard derivation of three technical replicates. The raw data can be found in Supplementary Table 2.

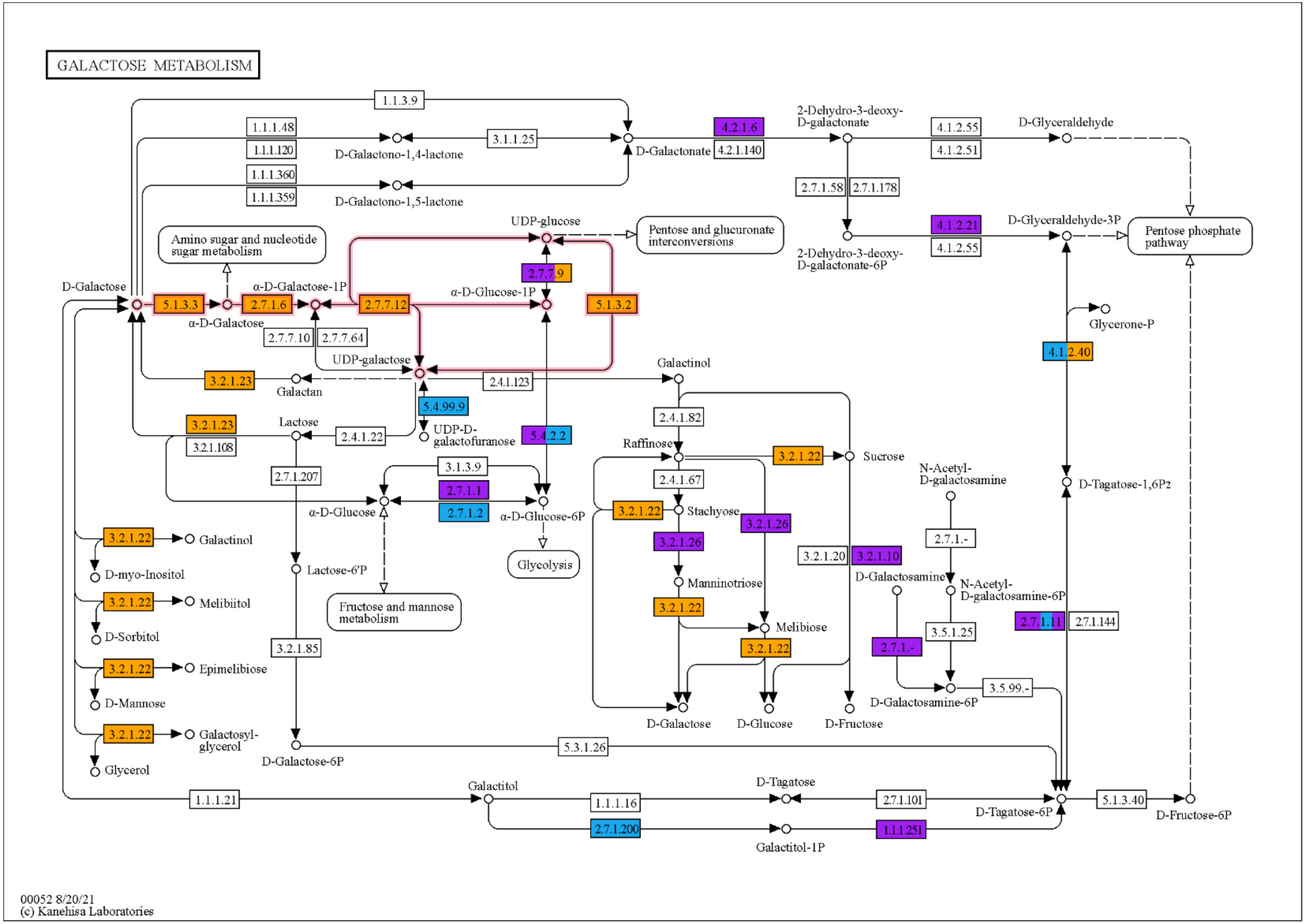
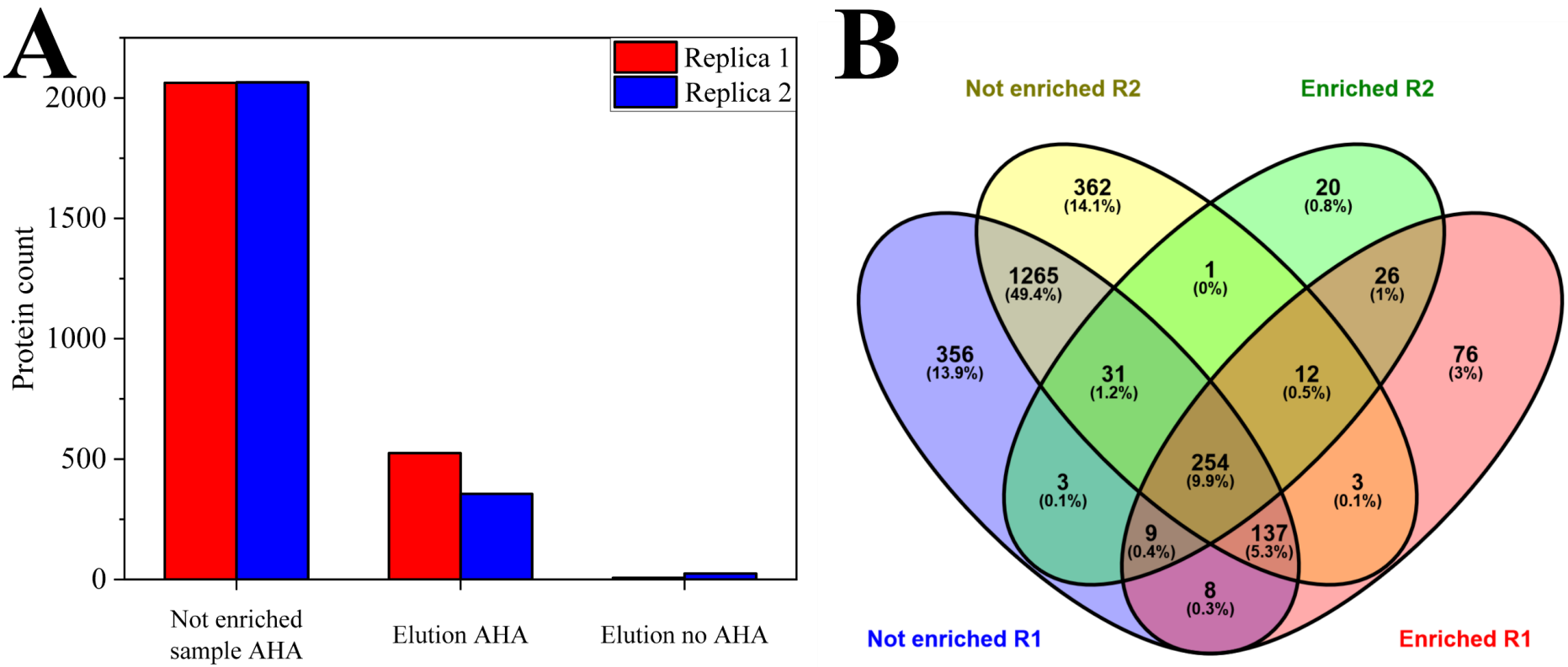

GALACTOSE METABOLISM
1.1.3.9
1.1.1.48
1.1.1.120
1.1.1.360
1.1.1.359
3.1.1.25
D-Galactono-1,4-lactone
D-Galactono-1,5-lactone
D-Galactonate
4.2.1.6
4.2.1.140
2-Dehydro-3-deoxy-D-galactonate
4.1.2.55
4.1.2.51
D-Glyceraldehyde
2.7.1.58
2.7.1.178
2-Dehydro-3-deoxy-D-galactonate-6P
4.1.2.21
4.1.2.55
D-Glyceraldehyde-3P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

Figure 3: KEGG map of all identified proteins from galactose metabolism in *E. coli* in the test system (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 uniquely 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 [98] .



493

494 Figure 4: Number of identified metaproteins from AD using the AA method. The proteins were grouped based on shared peptides and the number of

495 the resulting metaproteins was plotted. Each sample was prepared in biological duplicates, which are indicated in the plot by red and blue bars. The

496 samples are the not enriched but AHA labelled sample ("Not enriched sample"), the enriched nPs of the AHA labelled sample in the elution fraction

497 ("Elution AHA"), and in the elution fraction of the control without AHA labelling ("Elution No AHA"). **B:** The Venn diagram illustrates the complete

set of proteins identified in the respective replicates of the non-enriched (“Not enriched”) and nP-enriched (“Enriched”) samples [99]. The raw data can be found in Supplementary Table 3.

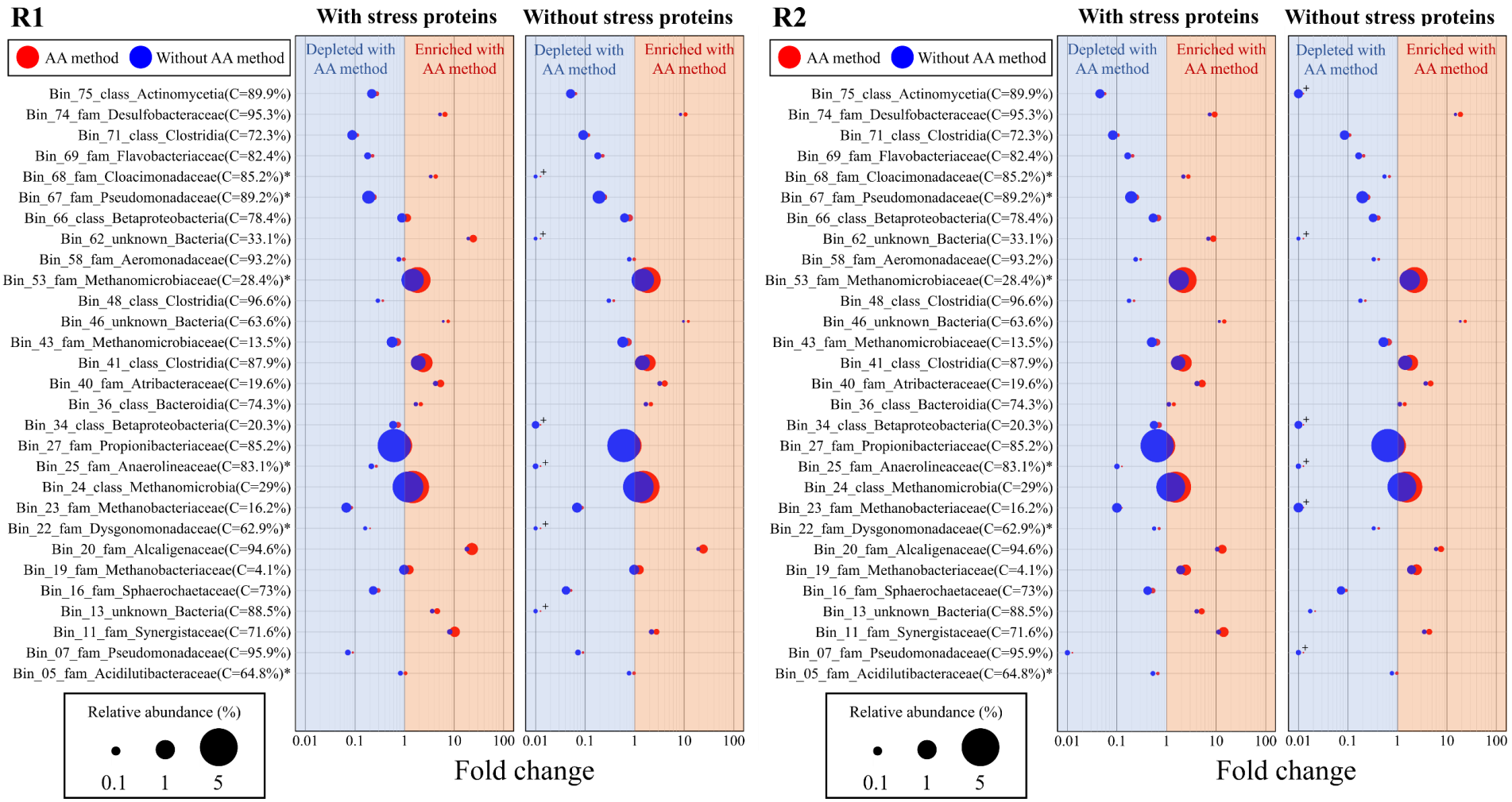
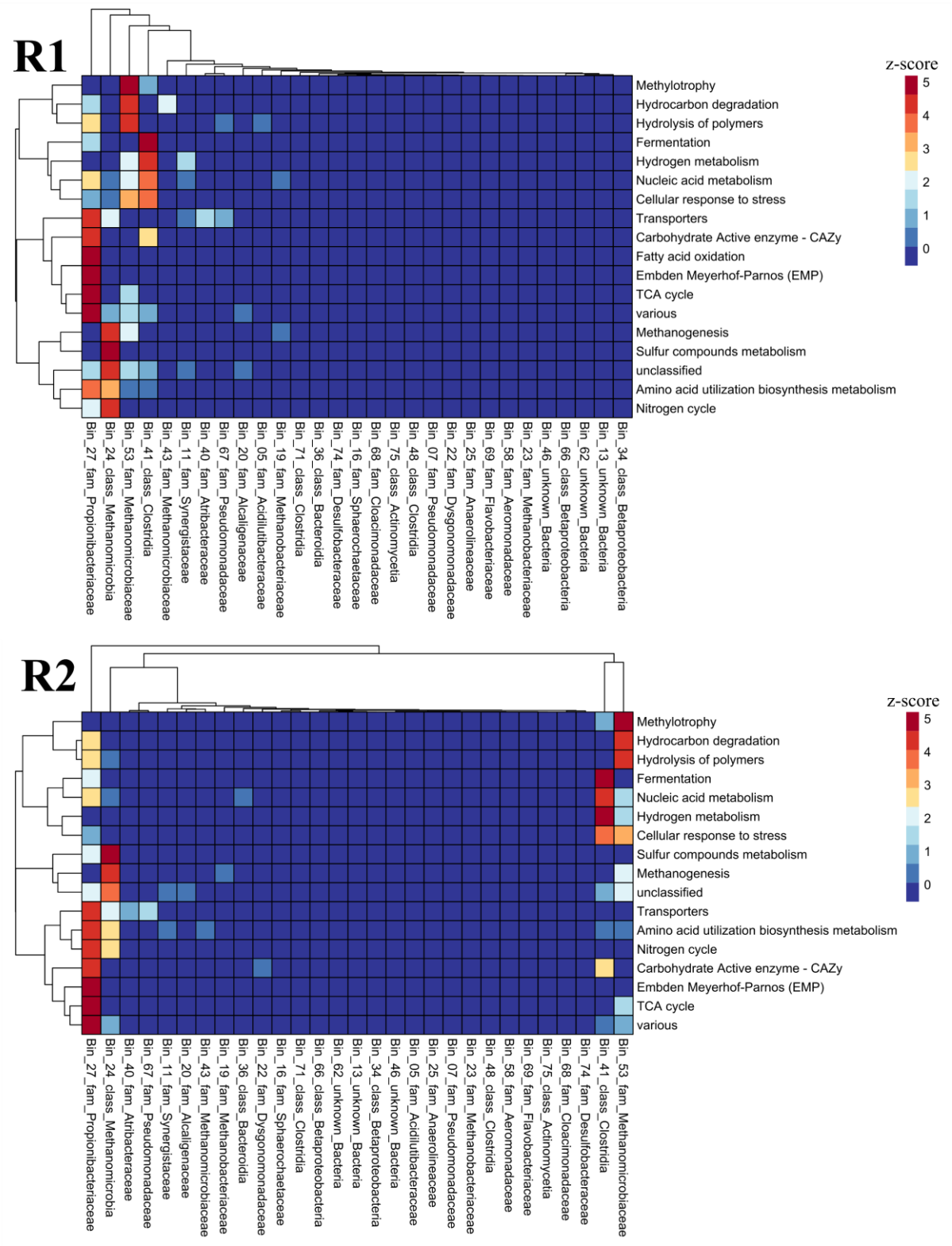


Figure 5: Bubble plot of the MAGs with the highest protein abundance identified using the AA method. The protein abundance of the 20 most abundant MAGs identified by the AA method and the 20 most abundant MAGs identified in the non-enriched sample were compared in 2 replicates (a total of

29 different MAGs). The fold changes were calculated based on the ratio of the relative spectral abundance of the MAG, which was normalised by the MAG length in the samples. The taxonomies were identified using the TYGS. The left bubble plot of each replicate 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*) (also Supplementary Table 6).



509

510 Figure 6: Alternated biological processes in the MC due to ethanol addition The functional

511 ontology assignments for metagenomes of all enriched nP and MAGs with the AA method were

analysed using a heatmap created with R Studio(2023.06.0 Build 421) and pheatmap (1.0.12). Previously, the protein abundances were normalised by dividing the spectral count of each protein by the total spectral abundance of the sample. Afterwards, a pivot table was generated, to sum up the relative abundances of the proteins based on the functional assignment and the MAG. The functional ontology assignments were normalized using a z-score per row (overall MAGs). also Supplementary Table 5.

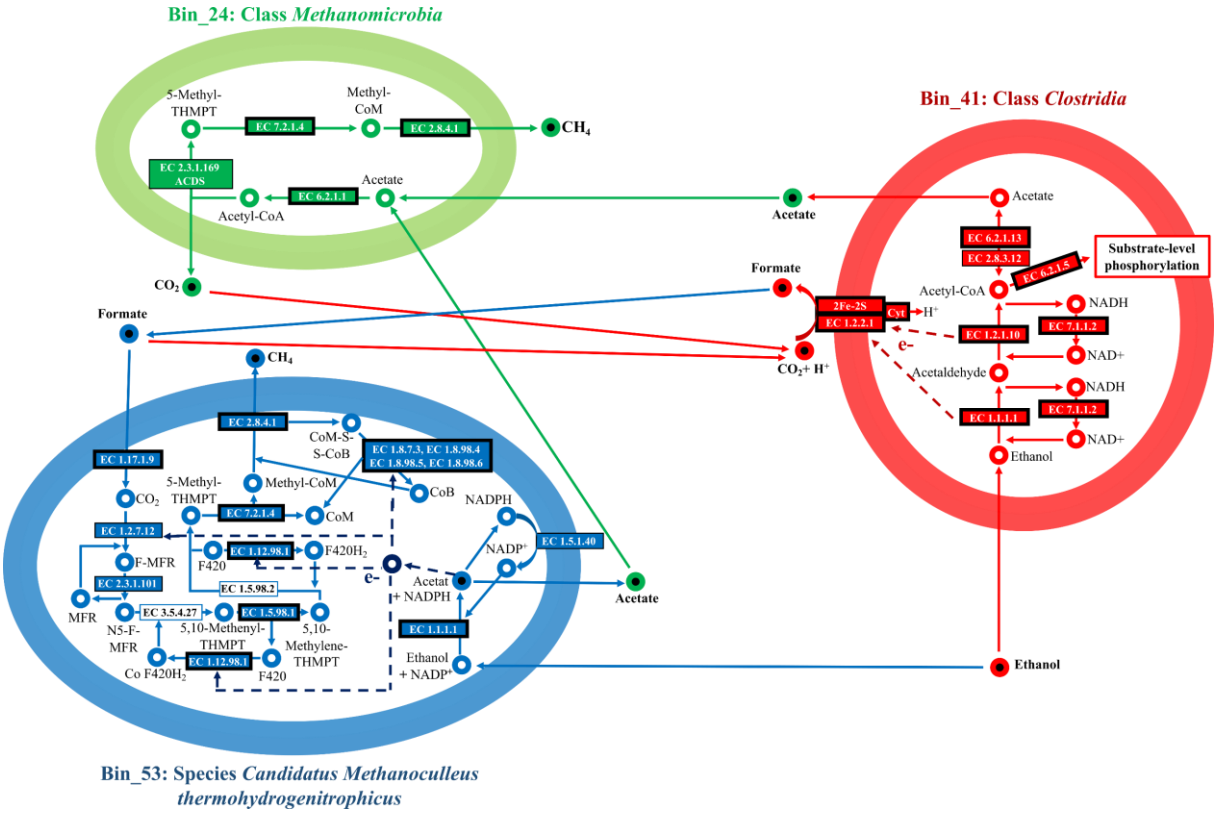


Figure 7: Suggestion for an anaerobic syntrophic ethanol oxidation based on nP enrichment. Intracellular metabolites are represented by unfilled circles, while extracellular metabolites are represented by filled circles. Metabolic pathways are designated by the colours of the arrows, which indicate the 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) [100] Fig.7.

527

Supplementary

Supplementary note 1: SOPs and detailed method description

Supplementary note 2: additional figures and tables

Supplementary table 1: lab scale biogas reactor information

Supplementary table 2: protein data testsystem

Supplementary table 3: metaprotein data ethanol

Supplementary table 4: identified proteins for syntrophic interactions

Supplementary table 5: biolog processes

Supplementary table 6: input bubble plot

Supplementary table 7: raw data 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. doi:10.1038/23616.
2. Burman E, Bengtsson-Palme J. Microbial Community Interactions Are Sensitive to Small Changes in Temperature. Front Microbiol. 2021;12:672910. doi:10.3389/fmicb.2021.672910.
3. Suttle CA. The significance of viruses to mortality in aquatic microbial communities. Microb Ecol. 1994;28:237–43. doi:10.1007/BF00166813.
4. Heyer R, Hellwig P, Maus I, Walke D, Schlüter A, Hassa J, 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. doi:10.1016/j.watres.2023.121020.
5. Waldrop MP, Firestone MK. Response of microbial community composition and function to soil climate change. Microb Ecol. 2006;52:716–24. doi:10.1007/s00248-006-9103-3.
6. Reid G, Younes JA, van der Mei HC, Gloor GB, Knight R, Busscher HJ. Microbiota restoration: natural and supplemented recovery of human microbial communities. Nat Rev Microbiol. 2011;9:27–38. doi:10.1038/nrmicro2473.
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. doi:10.1111/1462-2920.13437.
8. Tringe SG, Mering C von, Kobayashi A, Salamov AA, Chen K, Chang HW, et al. Comparative metagenomics of microbial communities. Science. 2005;308:554–7. doi:10.1126/science.1107851.
9. Wilmes P, Bond PL. Metaproteomics: studying functional gene expression in microbial ecosystems. Trends Microbiol. 2006;14:92–7. doi:10.1016/j.tim.2005.12.006.

10. Biévre PD, Taylor PDP. Table of the isotopic compositions of the elements. International Journal of Mass Spectrometry and Ion Processes. 1993;123:149–66. doi:10.1016/0168-1176(93)87009-H.
11. Becker GW. Stable isotopic labeling of proteins for quantitative proteomic applications. Briefings in Functional Genomics. 2008;7:371–82. doi:10.1093/bfgp/eln047.
12. Dieterich DC, Lee JJ, Link AJ, Graumann J, Tirrell DA, Schuman EM. Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. Nat Protoc. 2007;2:532–40. doi:10.1038/nprot.2007.52.
13. Kiick KL, Saxon E, Tirrell DA, Bertozzi CR. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. Proc Natl Acad Sci U S A. 2002;99:19–24. doi:10.1073/pnas.012583299.
14. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). Proc Natl Acad Sci U S A. 2006;103:9482–7. doi:10.1073/pnas.0601637103.
15. Ignacio BJ, Dijkstra J, Mora N, Slot EFJ, van Weijsten MJ, Storkebaum E, et al. THRONCAT: metabolic labeling of newly synthesized proteins using a bioorthogonal threonine analog. Nat Commun. 2023;14:3367. doi:10.1038/s41467-023-39063-7.
16. Kolb HC, Finn MG, Sharpless KB. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. Angew. Chem. Int. Ed. 2001;40:2004–21. doi:10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.
17. Johnson JA, Lu YY, van Deventer JA, Tirrell DA. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. Curr Opin Chem Biol. 2010;14:774–80. doi:10.1016/j.cbpa.2010.09.013.
18. van Kasteren S, Rozen DE. Using click chemistry to study microbial ecology and evolution. ISME Commun. 2023;3:9. doi:10.1038/s43705-022-00205-5.

19. Bagert JD, Xie YJ, Sweredoski MJ, Qi Y, Hess S, Schuman EM, Tirrell DA. 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;13:1352–8. doi:10.1074/mcp.M113.031914.
20. Zhang G, Bowling H, Hom N, Kirshenbaum K, Klann E, Chao MV, Neubert TA. In-depth quantitative proteomic analysis of de novo protein synthesis induced by brain-derived neurotrophic factor. J Proteome Res. 2014;13:5707–14. doi:10.1021/pr5006982.
21. Franco M, D'haeseleer PM, Branda SS, Liou MJ, Haider Y, Segelke BW, El-Etr SH. Proteomic Profiling of Burkholderia thailandensis During Host Infection Using Bio-Orthogonal Noncanonical Amino Acid Tagging (BONCAT). Front Cell Infect Microbiol. 2018;8:370. doi:10.3389/fcimb.2018.00370.
22. Hatzenpichler R, Scheller S, Tavormina PL, Babin BM, Tirrell DA, Orphan VJ. In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. Environ Microbiol. 2014;16:2568–90. doi:10.1111/1462-2920.12436.
23. Hatzenpichler R, Krukenberg V, Spietz RL, Jay ZJ. Next-generation physiology approaches to study microbiome function at single cell level. Nature Reviews Microbiology. 2020;18:241–56. doi:10.1038/s41579-020-0323-1.
24. Reichart NJ, Jay ZJ, Krukenberg V, Parker AE, Spietz RL, Hatzenpichler R. Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. ISME J. 2020;14:2851–61. doi:10.1038/s41396-020-00749-1.
25. Pasulka AL, Thamatrakoln K, Kopf SH, Guan Y, Poulos B, Moradian A, et al. Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. Environ Microbiol. 2018;20:671–92. doi:10.1111/1462-2920.13996.
26. Hellwig P, Dittrich A, Heyer R, Reichl U, Benndorf D. Detection, isolation and characterisation of phage-host complexes using BONCAT and click chemistry; 2024.

640 27. Landor LAI, Bratbak G, Larsen A, Tjendra J, Våge S. Differential toxicity of bioorthogonal
641 non-canonical amino acids (BONCAT) in Escherichia coli. J Microbiol Methods.
642 2023;206:106679. doi:10.1016/j.mimet.2023.106679.
643 28. Feng D, Guo X, Lin R, Xia A, Huang Y, Liao Q, et al. How can ethanol enhance direct
644 interspecies electron transfer in anaerobic digestion? Biotechnol Adv. 2021;52:107812.
645 doi:10.1016/j.biotechadv.2021.107812.
646 29. Wessel D, Flügge UI. A method for the quantitative recovery of protein in dilute solution
647 in the presence of detergents and lipids. Anal Biochem. 1984;138:141–3.
648 doi:10.1016/0003-2697(84)90782-6.
649 30. Heyer R, Schallert K, Büdel A, Zoun R, Dorl S, Behne A, et al. A Robust and Universal
650 Metaproteomics Workflow for Research Studies and Routine Diagnostics Within 24 h
651 Using Phenol Extraction, FASP Digest, and the MetaProteomeAnalyzer. Front Microbiol.
652 2019;10:1883. doi:10.3389/fmicb.2019.01883.
653 31. Laemmli UK. Cleavage of structural proteins during the assembly of the head of
654 bacteriophage T4. Nature. 1970;227:680–5. doi:10.1038/227680a0.
655 32. Maus I, Tubbesing T, Wibberg D, Heyer R, Hassa J, Tomazetto G, et al. The Role of
656 Petrimonas mucosa ING2-E5AT in Mesophilic Biogas Reactor Systems as Deduced from
657 Multiomics Analyses. Microorganisms 2020. doi:10.3390/microorganisms8122024.
658 33. Wibberg D, Andersson L, Tzelepis G, Rupp O, Blom J, Jelonek L, et al. Genome analysis
659 of the sugar beet pathogen Rhizoctonia solani AG2-2IIIB revealed high numbers in secreted
660 proteins and cell wall degrading enzymes. BMC Genomics. 2016;17:245.
661 doi:10.1186/s12864-016-2561-1.
662 34. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node
663 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
664 Bioinformatics. 2015;31:1674–6. doi:10.1093/bioinformatics/btv033.

35. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119. doi:10.1186/1471-2105-11-119.
36. 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); 10/24/2019 - 10/26/2019; Da Nang, Vietnam: IEEE; 2019. p. 1–5. doi:10.1109/KSE.2019.8919295.
37. Berckx F, van Nguyen T, Bandong CM, Lin H-H, Yamanaka T, Katayama S, 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. doi:10.1186/s12864-022-08838-5.
38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. doi:10.1038/nmeth.1923.
39. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359. doi:10.7717/peerj.7359.
40. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing Genomic Data Quality and Beyond. Curr Protoc. 2021;1:e323. doi:10.1002/cpz1.323.
41. 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. doi:10.1038/s41467-019-10210-3.
42. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res. 2022;50:D801-D807. doi:10.1093/nar/gkab902.
43. Muth T, Behne A, Heyer R, Kohrs F, Benndorf D, Hoffmann M, et al. The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation. J Proteome Res. 2015;14:1557–65. doi:10.1021/pr501246w.

691 44. Schiebenhoefer H, Schallert K, Renard BY, Trappe K, Schmid E, Benndorf D, et al. A
692 complete and flexible workflow for metaproteomics data analysis based on
693 MetaProteomeAnalyzer and Prophane. Nat Protoc. 2020;15:3212–39. doi:10.1038/s41596-
694 020-0368-7.
695 45. Prestat E, David MM, Hultman J, Taş N, Lamendella R, Dvornik J, et al. FOAM (Functional
696 Ontology Assignments for Metagenomes): a Hidden Markov Model (HMM) database with
697 environmental focus. Nucleic Acids Res. 2014;42:e145-e145. doi:10.1093/nar/gku702.
698 46. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H.
699 KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score
700 threshold. Bioinformatics. 2020;36:2251–2. doi:10.1093/bioinformatics/btz859.
701 47. Somasekharan SP, Stoynov N, Rotblat B, Leprivier G, Galpin JD, Ahern CA, et al.
702 Identification and quantification of newly synthesized proteins translationally regulated by
703 YB-1 using a novel Click–SILAC approach. Journal of Proteomics. 2012;77:e1-e10.
704 doi:10.1016/j.jprot.2012.08.019.
705 48. Kovalchuk SI, Anikanov NA, Ivanova OM, Ziganshin RH, Govorun VM. Bovine serum
706 albumin as a universal suppressor of non-specific peptide binding in vials prior to nano-
707 chromatography coupled mass-spectrometry analysis. Anal Chim Acta. 2015;893:57–64.
708 doi:10.1016/j.aca.2015.08.027.
709 49. Suelter CH, DeLuca M. How to prevent losses of protein by adsorption to glass and plastic.
710 Anal Biochem. 1983;135:112–9. doi:10.1016/0003-2697(83)90738-8.
711 50. Nicora CD, Anderson BJ, Callister SJ, Norbeck AD, Purvine SO, Jansson JK, et al. Amino
712 acid treatment enhances protein recovery from sediment and soils for metaproteomic
713 studies. Proteomics. 2013;13:2776–85. doi:10.1002/pmic.201300003.
714 51. Kennell D, Riezman H. Transcription and translation initiation frequencies of the
715 Escherichia coli lac operon. J Mol Biol. 1977;114:1–21. doi:10.1016/0022-2836(77)90279-
716 0.

52. Roderick SL. The lac operon galactoside acetyltransferase. C R Biol. 2005;328:568–75. doi:10.1016/j.crvi.2005.03.005.
53. Zhao Y, Zhang W, Kho Y, Zhao Y. Proteomic analysis of integral plasma membrane proteins. Anal Chem. 2004;76:1817–23. doi:10.1021/ac0354037.
54. Clark DP. The fermentation pathways of Escherichia coli. FEMS Microbiology Letters. 1989;63:223–34. doi:10.1111/j.1574-6968.1989.tb03398.x.
55. Jin Du, Qidong Yin, Mengqi Gu, Guangxue Wu. New insights into the effect of ethanol and volatile fatty acids proportions on methanogenic activities and pathways. Environmental Research. 2021;194:110644. doi:10.1016/j.envres.2020.110644.
56. Cao H, Du Wei, Yang Y, Shang Y, Li G, Zhou Y, et al. Systems-level understanding of ethanol-induced stresses and adaptation in E. coli. Sci Rep. 2017;7:44150. doi:10.1038/srep44150.
57. Guan N, Li J, Shin H-D, Du G, Chen J, Liu L. Microbial response to environmental stresses: from fundamental mechanisms to practical applications. Appl Microbiol Biotechnol. 2017;101:3991–4008. doi:10.1007/s00253-017-8264-y.
58. Kornberg HL, Krebs HA. Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. Nature. 1957;179:988–91. doi:10.1038/179988a0.
59. Kornberg HL, Madsen NB. Synthesis of C4-dicarboxylic acids from acetate by a glyoxylate bypass of the tricarboxylic acid cycle. Biochim Biophys Acta. 1957;24:651–3. doi:10.1016/0006-3002(57)90268-8.
60. 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. doi:10.1099/mic.0.030858-0.
61. 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-

742 phellandrene) and nitrate. Syst Appl Microbiol. 1998;21:237–44. doi:10.1016/S0723-
743 2020(98)80028-3.
744 62. Kämpfer P, Denger K, Cook AM, Lee S-T, Jäckel U, Denner EBM, Busse H-J.
745 Castellaniella gen. nov., to accommodate the phylogenetic lineage of Alcaligenes
746 defragrans, and proposal of Castellaniella defragrans gen. nov., comb. nov. and
747 Castellaniella denitrificans sp. nov. Int J Syst Evol Microbiol. 2006;56:815–9.
748 doi:10.1099/ijs.0.63989-0.
749 63. Koch S, Kohrs F, Lahmann P, Bissinger T, Wendschuh S, Benndorf D, et al. RedCom: A
750 strategy for reduced metabolic modeling of complex microbial communities and its
751 application for analyzing experimental datasets from anaerobic digestion. PLoS Comput
752 Biol. 2019;15:e1006759. doi:10.1371/journal.pcbi.1006759.
753 64. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate
754 protein structure prediction with AlphaFold. Nature. 2021;596:583–9. doi:10.1038/s41586-
755 021-03819-2.
756 65. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold
757 Protein Structure Database: massively expanding the structural coverage of protein-
758 sequence space with high-accuracy models. Nucleic Acids Res. 2022;50:D439-D444.
759 doi:10.1093/nar/gkab1061.
760 66. Tang Y-Q, Matsui T, Morimura S, Wu X-L, Kida K. Effect of temperature on microbial
761 community of a glucose-degrading methanogenic consortium under hyperthermophilic
762 chemostat cultivation. J Biosci Bioeng. 2008;106:180–7. doi:10.1263/jbb.106.180.
763 67. Poddar A, Lepcha RT, Das SK. Taxonomic study of the genus Tepidiphilus: transfer of
764 Petrobacter succinatimandens to the genus Tepidiphilus as Tepidiphilus succinatimandens
765 comb. nov., emended description of the genus Tepidiphilus and description of Tepidiphilus
766 thermophilus sp. nov., isolated from a terrestrial hot spring. Int J Syst Evol Microbiol.
767 2014;64:228–35. doi:10.1099/ijs.0.056424-0.

68. Walter KA, Bennett GN, Papoutsakis ET. Molecular characterization of two Clostridium acetobutylicum ATCC 824 butanol dehydrogenase isozyme genes. J Bacteriol. 1992;174:7149–58. doi:10.1128/jb.174.22.7149-7158.1992.
69. 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. doi:10.1016/0014-5793(91)80358-a.
70. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2023;51:D418-D427. doi:10.1093/nar/gkac993.
71. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40. doi:10.1093/bioinformatics/btu031.
72. Burton RM, Stadtman ER. The oxidation of acetaldehyde to acetyl coenzyme A. J Biol Chem. 1953;202:873–90.
73. Fleischmann A, Darsow M, Degtyarenko K, Fleischmann W, Boyce S, Axelsen KB, et al. IntEnz, the integrated relational enzyme database. Nucleic Acids Res. 2004;32:D434-7. doi:10.1093/nar/gkh119.
74. Gómez-Manzo S, Escamilla JE, González-Valdez A, López-Velázquez G, Vanoye-Carlo A, Marcial-Quino J, 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. doi:10.3390/ijms16011293.
75. Hitschler L, Nissen LS, Kuntz M, Basen M. 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. doi:10.1186/s13068-021-02038-1.
76. Ramos AR, Grein F, Oliveira GP, Venceslau SS, Keller KL, Wall JD, Pereira IAC. 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. doi:10.1111/1462-2920.12689.

77. Wawrik B, Marks CR, Davidova IA, McInerney MJ, Pruitt S, Duncan KE, 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. doi:10.1111/1462-2920.13374.

78. Zhang B, Lingga C, Bowman C, Hackmann TJ. A New Pathway for Forming Acetate and Synthesizing ATP during Fermentation in Bacteria. Appl Environ Microbiol. 2021;87:e0295920. doi:10.1128/AEM.02959-20.

79. Schuchmann K, Chowdhury NP, Müller V. Complex Multimeric FeFe Hydrogenases: Biochemistry, Physiology and New Opportunities for the Hydrogen Economy. Front Microbiol. 2018;9:2911. doi:10.3389/fmicb.2018.02911.

80. Frimmer U, Widdel F. Oxidation of ethanol by methanogenic bacteria. Arch. Microbiol. 1989;152:479–83. doi:10.1007/BF00446933.

81. Song J, Wang J, Wang X, Zhao H, Hu T, Feng Z, et al. Improving the Acetic Acid Fermentation of Acetobacter pasteurianus by Enhancing the Energy Metabolism. Front Bioeng Biotechnol. 2022;10:815614. doi:10.3389/fbioe.2022.815614.

82. 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. doi:10.1007/s002030050514.

83. Maus I, Wibberg D, Stantscheff R, Eikmeyer F-G, Seffner A, Boelter J, et al. Complete genome sequence of the hydrogenotrophic, methanogenic archaeon Methanoculleus bourgensis strain MS2(T), Isolated from a sewage sludge digester. J Bacteriol. 2012;194:5487–8. doi:10.1128/jb.01292-12.

84. Kougias PG, Campanaro S, Treu L, Zhu X, Angelidaki I. 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. doi:10.1016/j.anaerobe.2017.02.009.
85. 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;54:20–9. doi:10.1128/aem.54.1.20-29.1988.
86. Manzoor S, Schnürer A, Bongcam-Rudloff E, Müller B. Complete genome sequence of Methanoculleus bourgensis strain MAB1, the syntrophic partner of mesophilic acetate-oxidising bacteria (SAOB). Stand Genomic Sci. 2016;11:80. doi:10.1186/s40793-016-0199-x.
87. 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. doi:10.1111/1462-2920.15388.
88. Reddy CA, Bryant MP, Wolin MJ. Characteristics of S organism isolated from Methanobacillus omelianskii. J Bacteriol. 1972;109:539–45. doi:10.1128/jb.109.2.539-545.1972.
89. Bryant MP, Wolin EA, Wolin MJ, Wolfe RS. Methanobacillus omelianskii, a symbiotic association of two species of bacteria. Arch Mikrobiol. 1967;59:20–31. doi:10.1007/BF00406313.
90. Morris BEL, Henneberger R, Huber H, Moissl-Eichinger C. Microbial syntrophy: interaction for the common good. FEMS Microbiol Rev. 2013;37:384–406. doi:10.1111/1574-6976.12019.
91. Seitz H-J, Schink B, Pfennig N, Conrad R. Energetics of syntrophic ethanol oxidation in defined chemostat cocultures. Arch. Microbiol. 1990;155:82–8. doi:10.1007/BF00291279.

92. Summers ZM, Fogarty HE, Leang C, Franks AE, Malvankar NS, Lovley DR. Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. Science. 2010;330:1413–5. doi:10.1126/science.1196526.
93. 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. doi:10.3389/fmicb.2019.00423.
94. Khesali Aghtaei H, Püttker S, Maus I, Heyer R, Huang L, Sczyrba A, et al. Adaptation of a microbial community to demand-oriented biological methanation. Biotechnol Biofuels Bioprod. 2022;15:125. doi:10.1186/s13068-022-02207-w.
95. Hatzenpichler R, Orphan VJ. Detection of Protein-Synthesizing Microorganisms in the Environment via Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT). In: McGenity TJ, Timmis KN, Nogales B, editors. Hydrocarbon and Lipid Microbiology Protocols. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016. p. 145–157. doi:10.1007/8623_2015_61.
96. Gómez-Varela D, Xian F, Grundtner S, Sondermann JR, Carta G, Schmidt M. Increasing taxonomic and functional characterization of host-microbiome interactions by DIA-PASEF metaproteomics. Front Microbiol. 2023;14:1258703. doi:10.3389/fmicb.2023.1258703.
97. Fekner T, Li X, Lee MM, Chan MK. A pyrrolysine analogue for protein click chemistry. Angew. Chem. Int. Ed. 2009;48:1633–5. doi:10.1002/anie.200805420.
98. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51:D587-D592. doi:10.1093/nar/gkac963.
99. Oliveros JC. Venny: An interactive tool for comparing lists with Venn's diagrams.; 2015.
100. Zhao Z, Wang J, Li Y, Zhu T, Yu Q, Wang T, et al. Why do DIETers like drinking: Metagenomic analysis for methane and energy metabolism during anaerobic digestion with ethanol. Water Res. 2020;171:115425. doi:10.1016/j.watres.2019.115425.