Absolute proteome quantification in the gas-fermenting acetogen *Clostridium autoethanogenum*

Kaspar Valgepea[1,2], Gert Talbo[1,3], Nobuaki Takemori[4], Ayako Takemori[4], Christina Ludwig[5], Alexander P. Mueller[6], Ryan Tappel[6], Michael Köpke[6], Séan Dennis Simpson[6], Lars Keld Nielsen[1,3,7] and Esteban Marcellin[1,3*]

[1]Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, 4072 St. Lucia, Australia

[2]ERA Chair in Gas Fermentation Technologies, Institute of Technology, University of Tartu, 50411 Tartu, Estonia

[3]Queensland Node of Metabolomics Australia, AIBN, The University of Queensland, 4072 St. Lucia, Australia

[4]Institute for Promotion of Science and Technology, Ehime University, 791-0295 Ehime, Japan

[5]Bavarian Center for Biomolecular Mass Spectrometry (BayBioMS), Technical University of Munich, 85354 Freising, Germany

[6]LanzaTech Inc., 60077 Skokie, USA

[7]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

[*]Correspondence: e.marcellin@uq.edu.au

ABSTRACT

Microbes that can recycle one-carbon (C1) greenhouse gases into fuels and chemicals are vital for the biosustainability of future industries. Acetogens are the most efficient known microbes for fixing carbon oxides CO_2 and CO. Understanding proteome allocation is important for metabolic engineering as it dictates metabolic fitness. Here, we use absolute proteomics to quantify intracellular concentrations for >1,000 proteins in the model-acetogen *Clostridium autoethanogenum* grown on three gas mixtures. We detect prioritisation of proteome allocation for C1 fixation and significant expression of proteins involved in the production of acetate and ethanol as well as proteins with unclear functions. The data also revealed which isoenzymes are important. Integration of proteomic and metabolic flux data demonstrated that enzymes catalyse high fluxes with high concentrations and high in vivo catalytic rates. We show that flux adjustments were dominantly accompanied with changing enzyme catalytic rates rather than concentrations. Our work serves as a reference dataset and advances systems-level understanding and engineering of acetogens.

INTRODUCTION

Increasing concerns about irreversible climate change are accelerating the shift to renewable, carbon-free energy production (e.g., solar, wind, fuel cells). However, many fuels and chemicals will stay carbon-based, and thus, technologies for their production using sustainable and renewable feedstocks are needed to transition towards a circular bioeconomy. Moreover, the rising amount of solid waste produced by human activities (e.g., municipal solid waste, lignocellulosic waste) will further endanger our ecosystems' already critical state. Both challenges can be tackled by using organisms capable of recycling gaseous one-carbon (C1) waste feedstocks (e.g., industrial waste gases [CO_2, CO, CH_4], syngas from gasified biomass or municipal solid waste [CO, H_2, CO_2]) into fuels and chemicals at industrial scale[1–3].

As we transition into a new bioeconomy, a key feature of global biosustainability will be the capacity to convert carbon oxides into products at industrial scale. Acetogens are the ideal biocatalysts for this as they use the most energy-efficient pathway, the Wood-Ljungdahl pathway (WLP)[4,5], for fixing CO_2 into the central metabolite acetyl-CoA[6–9] and accept gas (CO, H_2, CO_2) as their sole carbon and energy source[5]. Indeed, the model-acetogen *Clostridium autoethanogenum* is already being used as a cell factory in industrial-scale gas fermentation[3,10]. The WLP is considered the first biochemical pathway on Earth[7,11–13] and continues to play a critical role in the biogeochemical carbon cycle by fixing an estimated 20% of the global CO_2[6,14]. While biochemical details of the WLP are well described[4,6,15], a quantitative understanding of acetogen metabolism is just emerging[16,17]. Notably, recent systems-level analyses of acetogen metabolism have revealed mechanisms behind metabolic shifts[18–21], transcriptional architectures[22,23], and features of translational regulation[24,25]. However, we still lack an understanding of acetogen proteome allocation through the quantification of proteome-wide intracellular protein concentrations. This fundamental knowledge is required for advancing rational metabolic engineering of acetogen cell factories and for accurate in silico reconstruction of their phenotypes using metabolic models[1,2].

Quantitative description of an organism's proteome allocation through absolute proteome quantification is valuable in several ways. Firstly, it enables us to understand prioritisation of the energetically costly proteome resources among functional protein categories, metabolic pathways, and

single proteins[26,27]. This may also identify relevant proteins with unclear functions and high abundances. Secondly, some metabolic fluxes can be catalysed by isoenzymes and a comparison of their intracellular concentrations can indicate which are likely relevant in vivo and are thus targets for genetic perturbation experiments to validate in vivo functionalities[28]. Thirdly, integration of absolute proteomics and metabolic flux data enable the estimation of apparent in vivo catalytic rates of enzymes (k_{app})[26,29], which can be used to identify less-efficient enzymes as targets for improving pathways through metabolic and protein engineering. Absolute proteomics data also contribute to the curation of accurate genome-scale metabolic models.

Absolute proteome quantification is generally performed using label-free mass-spectrometry (MS) approaches without spike-in standards[30,31]. The major limitation of this approach is that accuracy of label-free estimated protein concentrations cannot be determined. Furthermore, the optimal model to convert MS signals (e.g., spectral counts, peak intensities) into protein concentrations remains unknown[32–34]. Label-based approaches using stable-isotope labelled (SIL) spike-ins of endogenous proteins are thus preferred for reliable absolute proteome quantification. This strategy relies on accurate absolute quantification of a limited set of intracellular proteins (i.e., anchors) using SIL spike-ins to establish a linear correlation between protein concentrations and their measured MS intensities[32]. Studies with the latter approach have determined a 1.5□2.4-fold error for label-free estimation of proteome-wide protein concentrations in multiple organisms[28,35–41].

The aim of our work was to perform reliable absolute proteome quantification for the first time in an acetogen. We employed a label-based MS approach using SIL-protein spike-in standards to quantify SIL-based concentrations for 16 key proteins and label-free-based concentrations for >1,000 *C. autoethanogenum* proteins during autotrophic growth on three gas mixtures. This allowed us to explore global proteome allocation, uncover isoenzyme usage in central metabolism, and quantify regulatory principles associated with estimated k_{app}s. Our work provides an important reference dataset and advances the systems-level understanding and engineering of the ancient metabolism of acetogens.

RESULTS

Absolute proteome quantification framework in the model-acetogen *C. autoethanogenum*. We performed absolute proteome quantification from autotrophic steady-state chemostat cultures of *C. autoethanogenum* grown on three different gas mixtures: CO, syngas ($CO+CO_2+H_2$), or $CO+H_2$ (termed "high-H_2 CO") described before[18,19]. Briefly, four biological cultures of each gas mixture were grown anaerobically on a chemically defined medium at 37 °C, pH of 5, and dilution rate ~1 day^{-1} (specific growth rate ~0.04 h^{-1}) without the use of heavy SIL substrates. The absolute proteome quantification framework (Fig. 1) was built on using 19 synthetic heavy SIL-variants of key *C. autoethanogenum* proteins covering central metabolism (Supplementary Table 1). The SIL-protein standards were spiked in for quantification of intracellular concentrations of their endogenous light counterparts. This framework ensures accurate absolute quantification compared to commonly used peptide spike-ins. Spiking cell lysates with protein standards before sample clean-up and protein digestion accounts for errors accompanying these critical steps[30,31,42,43]. Furthermore, selection of peptides ensuring accurate absolute protein quantification without prior MS data is challenging as its difficult to predict which peptides "fly" well[30,31,42,43]. In contrast, all proteotypic peptides from a protein spike-in can be used for quantification.

We synthesised heavy-labelled lysine and arginine SIL-proteins using a cell-free wheat germ extract platform as described previously[18,44,45] and quantified standard stocks using parallel reaction monitoring (PRM) MS. Next, proteins were extracted from culture samples using an optimised protocol maximising extraction yield[18] followed by spike-in of the 19 heavy SIL-proteins into light cell lysates. We then used a data-independent acquisition (DIA) MS approach[46] to quantitate 1,243 proteins of *C. autoethanogenum* across 12 samples (quadruplicate cultures of three gas mixtures) using a comprehensive spectral library consisting of whole-cell lysates, lysate fractions, and spike-in SIL-proteins. Finally, we quantified intracellular concentrations for 16 key *C. autoethanogenum* proteins using light-to-heavy ratios between endogenous and spike-in DIA MS intensities and further used these 16 as anchor proteins for label-free estimation of ~1,043 protein concentrations through establishing a linear correlation between protein concentrations and their measured MS intensities[32].

We express intracellular protein concentrations in nanomoles of protein per gram of dry cell weight (nmol/gDCW).

Absolute quantification of 16 anchor protein concentrations. To ensure high confidence absolute quantification of anchor protein concentrations from the DIA MS data, we employed stringent criteria on top of the automated mProphet peak picking algorithm[47] within the software Skyline[48]. We also performed a dilution series experiment for each SIL-protein to increase accuracy (see Methods for details). Briefly, we kept only peaks with Gaussian shapes and without interference and precursors with highest Skyline quality metrics. Importantly, only peptides whose signal were above the lower limit of quantification (LLOQ) and within the linear dynamic quantification range in the dilution series experiment were used for anchor protein quantification (Supplementary Table 2). We thus used 106 high-confidence peptides for the absolute quantification of 16 anchor protein concentrations (Table 1; see also Fig. 5). High confidence of the intracellular concentrations for these key *C. autoethanogenum* proteins of central metabolism is supported both by the low average 11% coefficient of variation (CV) between biological quadruplicate cultures (Table 1) and the average 22% CV between different peptides of single proteins (Supplementary Table S2).

Label-free estimation of proteome-wide protein concentrations. Both high quality proteomics data and suitable anchor proteins are required for reliable label-free absolute proteome quantification. Our proteome-wide DIA MS data were highly reproducible with an average Pearson correlation coefficient of R = 0.99 between biological replicates (Fig. 2a and Supplementary Fig. 1). We also found our anchor proteins suitable as their concentrations spanned across three orders of magnitude, and the summed mass accounted for ~⅓ of the peptide mass injected into the mass spectrometer (Table 1 and Fig. 2b). We used the 16 anchor proteins (with 106 peptides) to determine the optimal label-free quantification model with the best linear fit between anchor protein concentrations and their measured DIA MS intensities using the aLFQ R package[49] as described before for SWATH MS[28] (Fig. 2c). Notably, we detected an average 1.5-fold cross-validated mean fold-error (CV-MFE; bootstrapping) for the label-free estimated anchor protein concentrations across samples (Fig. 2d).

139 The errors were distributed normally (Supplementary Fig. 2) with an average 95% CI of 0.3 (Fig. 2d).
140 We then applied the optimal label-free quantification model to estimate ~1,043 protein concentrations
141 in *C. autoethanogenum* (Supplementary Table 3).

142 Prior to the detailed analysis of proteome-wide protein concentrations, we further evaluated
143 our label-free data accuracy beyond the 1.5-fold CV-MFE determined above. Firstly, the total
144 proteome mass (1.2±0.1 μg; average ± standard deviation) closely matched the 1 μg peptide mass
145 injected into the mass spectrometer (Fig. 2d). The data were also supported by a strong correlation
146 between estimated protein concentrations and expected stoichiometries for equimolar (Fig. 3) and
147 non-equimolar protein complexes (Supplementary Fig. 3). Notably, absolute protein concentrations of
148 syngas cultures correlated well (R = 0.65) with their respective absolute transcript expression levels
149 determined before[19] (Supplementary Fig. 4). This result is similar to the correlations of absolute data
150 seen in other steady-state cultures[26,50]. Altogether, we present the first absolute quantitative proteome
151 dataset for a gas-fermenting acetogen that includes SIL-based concentrations for 16 key proteins and
152 label-free estimates for over 1,000 *C. autoethanogenum* proteins during growth on three gas mixtures.
153
154 **C1 fixation dominates global proteome allocation.** Global proteome allocation amongst functional
155 gene classifications was explored using proteomaps[27] and KEGG Orthology identifiers (KO IDs)[51].
156 The "treemap" structure defining the four-level hierarchy of our proteomaps (Supplementary Table 4)
157 also included manually curated categories to accurately reflect acetogen metabolism (e.g., C1
158 fixation/WLP, Hydrogenases). As expected for autotrophic growth of an acetogen, the C1 fixation
159 (Fig. 4) or WLP (Supplementary Fig. 5) categories dominated the proteome allocation with a ~⅓
160 fraction, compared to Carbohydrate metabolism or Glycolysis/Gluconeogenesis. Notably, the data
161 show that two genes–dihydrolipoamide dehydrogenase (LpdA; CAETHG_RS07825) and glycine
162 cleavage system H protein (GcvH; RS07795)–encoded by the WLP gene cluster were translated at
163 very high levels (Fig. 4). This is important as both have unknown functions in *C. autoethanogenum*
164 metabolism. Significant investment in expression of proteins involved in acetate and ethanol
165 production (Supplementary Fig. 5) is consistent with ⅓-to-⅔ of fixed carbon channelled into these
166 two growth by-products across the three gas mixtures[18,19]. The 11% proteome fraction of category

Translation (Fig. 4) is expected for cells growing at a specific growth rate ~0.04 h^{-1} based on absolute proteomics data from *Escherichia coli*[26,39,52]. The notable proteome allocation for Amino acid metabolism and particularly the high abundance of ketol-acid reductoisomerase (IlvC; RS00580) are surprising since metabolic fluxes through 2,3-butanediol and branched-chain amino acid pathways were low under these growth conditions[18,19]. In addition, numerous proteins with unknown or unclear functions (coloured grey in proteomaps) are highly expressed (e.g., RS12590, RS08610, RS08145), highlighting the need for global mapping of genotype-phenotype relationships in acetogens. In general, proteome allocation was highly similar between the three gas mixtures (Supplementary Table 3). This result is unsurprising given the few relative protein expression differences detected previously[18].

Enzyme usage revealed in central metabolism. Next, we focused on uncovering enzyme usage in acetogen central metabolism (Fig. 5). This contains enzymes of the WLP, acetate, ethanol, and 2,3-butanediol production pathways, hydrogenases, and the Nfn transhydrogenase, which together carry >90% of the carbon and most of the redox flow in *C. autoethanogenum*[18–20]. Multiple metabolic fluxes in these pathways can be catalysed by isoenzymes and absolute proteomics data can indicate which of the isoenzymes are likely relevant in vivo. While the carbon monoxide dehydrogenase (CODH) AcsA (RS07861□62) that forms the bifunctional CODH/ACS complex with the acetyl-CoA synthase[53] (AcsB; RS07800) is essential for *C. autoethanogenum* growth on gas as confirmed in mutagenesis studies[54], the higher concentrations of the dispensable monofunctional CODH CooS1 (RS14775) suggest it may also play a role in CO oxidation (Fig. 4, 5), in addition to CO_2 reduction[54]. Additionally, our proteomics data show high abundance of the primary acetaldehyde:ferredoxin oxidoreductase (AOR1; RS00440) and this support the emerging understanding that in *C. autoethanogenum* ethanol is dominantly produced using the AOR1 activity via acetate, instead of directly from acetyl-CoA via acetaldehyde using mono- or bifunctional activities[18,19,55,56] (Fig. 5). Furthermore, the data suggest that the specific alcohol dehydrogenase (Adh4; RS08920) is responsible for reducing acetaldehyde to ethanol, a key reaction in terms of carbon and redox metabolism. The high abundance of the electron-bifurcating hydrogenase HytA-E complex (RS13745□70) compared to alternative hydrogenases confirms that it is the main H_2-oxidiser[57,58] (Fig. 5). This is consistent with

the fact that in the presence of H_2 all the CO_2 fixed by the WLP is reduced to formate using H_2 by the HytA-E and formate dehydrogenase (FdhA; RS13725) enzyme complex activity[18,19]. Despite the proteomics evidence, genetic perturbations are required to determine condition-specific in vivo functionalities of isoenzymes in acetogens unequivocally.

The overall most abundant protein was the formate-tetrahydrofolate ligase (Fhs; RS07850), a key enzyme in the WLP (Fig. 5, 4). Despite the high abundance, its expression might still be rate-limiting (see below). Another key enzyme for acetogens is AcsB because of its essentiality for acetyl-CoA synthesis by the CODH/ACS complex. AcsB is linked to the WLP by the corrinoid iron sulfur proteins AcsC (RS07810) and AcsD (RS07815) that supply the methyl group to AcsB. Interestingly, the ratio of AcsCD-to-AcsB increased from 1.7 (CO) to 2.3 (syngas) to 2.9 (high-H_2 CO), suggesting that the primary role of the CODH/ACS complex shifted from CO oxidation towards acetyl-CoA synthesis, likely because increased H_2 uptake could replace the supply of reduced ferredoxin from CO oxidation. Concurrently, the Nfn transhydrogenase (RS07665) levels that act as a redox valve in acetogens[20] are maintained high (Fig. 5), potentially to rapidly respond to redox perturbations. We conclude that absolute quantitative proteomics can significantly contribute to a systems-level understanding of metabolism, particularly in less-studied organisms.

Integration of absolute proteomics and flux data yields in vivo enzyme catalytic rates. Absolute proteomics data enable estimation of intracellular catalytic working rates of enzymes when metabolic flux rates are known[26,29]. We thus calculated apparent in vivo catalytic rates of enzymes, denoted as k_{app} (s^{-1})[26], as the ratio of specific flux rate (mmol/gDCW/h) determined before[18] and protein concentration (nmol/gDCW) (see Methods). This produced k_{app} values for 13 and 48 enzymes/complexes using either anchor or label-free protein concentrations, respectively (Supplementary Table 5 and Fig. 5, 6). The first two critical steps for carbon fixation in the methyl branch of the WLP (i.e., CO to formate) are catalysed at high rates (Fig. 5). Notably, FdhA showed a k_{app}~30 s^{-1} for CO_2 reduction without H_2 during growth on CO only, which is similar to in vitro k_{cat} data of formate dehydrogenases in other acetogens[59,60]. Interestingly, the next step of formate reduction was catalysed potentially by a less-efficient enzyme–Fhs–as its k_{app} of ~3 s^{-1} is significantly

lower compared to other WLP enzymes (Fig. 5). Overall, enzymes catalysing reactions in high flux pathways such as the WLP and acetate and ethanol production have higher k_{app}s than those of downstream from conversion of acetyl-CoA to pyruvate (Fig. 5). Indeed, enzymes catalysing high metabolic fluxes in *C. autoethanogenum* have both higher concentrations and higher catalytic rates compared to enzymes catalysing lower fluxes as both specific flux rates and enzyme concentrations (Kendall's $\tau = 0.56$, p-value = 5×10^{-9}) and flux and k_{app} ($\tau = 0.45$, p-value = 2×10^{-6}) were significantly correlated (Fig. 6a), as seen before for other organisms[26,61].

Having acquired absolute proteomics data for *C. autoethanogenum* growth on three gas mixtures with different metabolic flux profiles also allowed us to determine the impact of change in enzyme concentration and its catalytic rate for adjusting metabolic flux rates. Two extreme examples are the reactions catalysed by the HytA-E (Fig. 6b) and the Nfn (Fig. 5) complexes where flux adjustments were accompanied with large changes in k_{app}s rather than in enzyme concentrations. Flux changes in high flux pathways such as the WLP and acetate and ethanol production also coincided mainly with k_{app} changes (Fig. 6b). This principle seems to be dominant in *C. autoethanogenum* as 90% of flux changes were not regulated through enzyme concentrations (i.e., post-translational regulation; Supplementary Table 6) when comparing all statistically significant flux changes between the three gas mixtures with respective enzyme expression changes (see Methods).

DISCUSSION

The looming danger of irreversible climate change and harmful effects of solid waste accumulation are pushing humanity to develop and adopt sustainable technologies for renewable production of fuels and chemicals and for waste recycling. Acetogen gas fermentation offers great potential to tackle both challenges through recycling waste feedstocks (e.g., industrial waste gases, gasified biomass or municipal solid waste) into fuels and chemicals[1,2]. Although the quantitative understanding of acetogen metabolism has recently improved[16,17], a quantitative description of acetogen proteome allocation was missing. This is needed to advance their metabolic engineering into superior cell factories and accurate in silico reconstruction of their phenotypes[1,2]. Thus, we performed absolute

proteome quantification in the model-acetogen *C. autoethanogenum* grown autotrophically on three gas mixtures.

Our absolute proteome quantification framework relied on SIL-protein spike-in standards and DIA MS analysis to ensure high confidence of the determined intracellular concentrations for 16 key *C. autoethanogenum* proteins. We further used these proteins as anchor proteins for label-free estimation of >1,000 protein concentrations. This enabled us to determine the optimal label-free quantification model for our data to infer protein concentrations from MS intensities, which remains unknown in common label-free approaches not utilising spike-in standards[32,33]. More importantly, label-free estimated protein concentrations using the latter approach are questionable as their accuracy cannot be determined. We determined an excellent average error of 1.5-fold for our label-free estimated protein concentrations based on 16 anchor proteins and a bootstrapping approach. This error is in the same range as described in previous studies using SIL spike-in standards for absolute proteome quantification[28,35–41]. Further, we also observed a good match both between estimated and injected proteome mass into the mass spectrometer and between protein concentrations and expected protein complex stoichiometries. We conclude that label-free estimation of proteome-wide protein concentrations using SIL-protein spike-ins and state-of-the-art MS analysis is reasonably accurate.

Quantification of acetogen proteome allocation during autotrophic growth expectedly showed prioritisation of proteome resources for fixing carbon through the WLP, in line with transcript expression data in *C. autoethanogenum*[19,56]. The allocation of one third of the total proteome for C1 fixation is higher than proteome allocation for carbon fixation through glycolysis during heterotrophic growth of other microorganisms[26,36]. High abundances of other key enzymes of acetogen central metabolism were also expected as the WLP, acetate and ethanol production pathways, hydrogenases, and the Nfn transhydrogenase carry >90% of the carbon and most of the redox flow in *C. autoethanogenum*[18–20]. However, very high expression of the two genes–LpdA and GcvH–of the WLP gene cluster with unknown functions in *C. autoethanogenum* is striking, raising the question whether their function in *C. autoethanogenum* could also be to link WLP and glycine synthase-reductase pathways, as recently proposed for another acetogen[62]. Since many other proteins with unknown or

277 unclear functions were also highly abundant, global mapping of genotype-phenotype relationships in
278 acetogens is much needed.

279 The in vivo functionalities of isoenzymes are not clear for multiple key metabolic fluxes in
280 acetogen central metabolism and absolute proteomics data can indicate which isoenzymes are likely
281 relevant. Oxidation of CO or reduction of CO_2 is a fundamental step for all acetogens and known to
282 be catalysed by three CODHs in *C. autoethanogenum*[54]. Though only AcsA that forms the
283 bifunctional CODH/ACS complex with the acetyl-CoA synthase[53] is essential for growth on gas[54], we
284 detected higher concentrations of the monofunctional CODH CooS1, which deletion strain shows
285 intriguing phenotypes[54]. Concurrently, our data suggest that prioritisation of CODH/ACS activity
286 between CO oxidation and acetyl-CoA synthesis is sensitive to H_2 availability. Thus, further studies
287 are required to decipher condition-dependent functionalities of CODHs. In addition to CODHs, the
288 biochemical understanding of ethanol production is important in terms of both carbon and redox
289 metabolism. Our data confirm that in *C. autoethanogenum* ethanol is predominantly produced via
290 acetate by AOR1[18,19,55,56] and more importantly, indicate for the first time that AOR1 activity is
291 followed by Adh4 (previously characterised as butanol dehydrogenase[63]) for reduction of
292 acetaldehyde to ethanol. These observations call for large-scale genetic perturbation experiments to
293 determine unequivocally the condition-specific in vivo functionalities of isoenzymes in acetogens.

294 Absolute proteomics data offer a unique opportunity to estimate apparent in vivo catalytic
295 rates of enzymes (k_{app})[26,29] if also metabolic flux data are available. These data are particularly
296 valuable for more accurate in silico reconstruction of phenotypes using protein-constrained genome-
297 scale metabolic models[64,65]. While in vitro k_{cat} and in vivo k_{app} data generally correlate[29], models using
298 maximal k_{app} values show better prediction of protein abundances[66]. Furthermore, information of k_{app}s
299 can infer less-efficient enzymes as targets for improving pathways through metabolic and protein
300 engineering. For example, protein engineering of Fhs (catalysing formate reduction) might improve
301 WLP throughput and carbon fixation since its k_{app} was significantly lower compared to other pathway
302 enzymes. At the same time, the large change in k_{app}s of the abundant electron-bifurcating hydrogenase
303 HytA-E and the Nfn transhydrogenase complexes indicate capacity for the cells to rapidly respond to
304 H_2 availability and redox perturbations, which may be critical for metabolic robustness of acetogens[20].

Overall, we detected both higher concentrations and k_{app}s for enzymes catalysing higher metabolic fluxes, which is believed to arise from an evolutionary push towards reducing protein production costs for enzymes carrying high flux[61]. The observation that 90% of flux changes in *C. autoethanogenum* were not regulated through changes in enzyme concentrations is not surprising for a metabolism that operates at the thermodynamic edge of feasibility[16,17] since post-translational regulation of fluxes is energetically least costly. Further research is needed to identify which mechanism from post-translational protein modification, allosteric regulation, or substrate concentration change is responsible for post-translational regulation of fluxes.

We have produced the first absolute proteome quantification in an acetogen and thus provided understanding of global proteome allocation, isoenzyme usage in central metabolism, and regulatory principles of in vivo enzyme catalytic rates. This fundamental knowledge has potential to advance both rational metabolic engineering of acetogen cell factories and accurate in silico reconstruction of their phenotypes[1,2,64,65]. Our study also highlights the need for large-scale mapping of genotype-phenotype relationships in acetogens to infer in vivo functionalities of isoenzymes and proteins with unknown or unclear functions. This absolute proteomics dataset serves as a reference towards a better systems-level understanding of the ancient metabolism of acetogens.

METHODS

Bacterial strain and culture growth conditions. Absolute proteome quantification was performed from high biomass concentration (~1.4 gDCW/L) steady-state autotrophic chemostat cultures of *C. autoethanogenum* growing on three different gas mixtures with culturing conditions described in our previous works[18,19]. Briefly, four biological replicate chemostat cultures of *C. autoethanogenum* strain DSM 19630 were grown on a chemically defined medium (without yeast extract) either on CO (~60% CO and 40% Ar), syngas (~50% CO, 20% H_2, 20% CO_2, and 10% N_2/Ar), or CO+H_2, termed "high-H_2 CO", (~15% CO, 45% H_2, and 40% Ar) under strictly anaerobic conditions. The bioreactors were maintained at 37 °C, pH of 5, and dilution rate ~1 day^{-1} (specific growth rate ~0.04 h^{-1}).

Cell-free synthesis of stable-isotope labelled protein standards. Twenty proteins covering *C. autoethanogenum* central carbon metabolism, the HytA-E hydrogenase, and a ribosomal protein (Supplementary Table 1) were selected for cell-free synthesis of SIL-proteins as described in ref.[18]. Briefly, genes encoding for these proteins were synthesised by commercial gene synthesis services (Biomatik). Target genes were sub-cloned into the cell-free expression vector pEUE01-His-N2 (CellFree Sciences) and transformed into *Escherichia coli* DH5α from which plasmid DNA was extracted and purified. Correct gene insertion into the pEUE01-His-N2 was verified by DNA sequencing. Subsequently, cell-free synthesis of His-tag fused *C. autoethanogenum* proteins was performed using the bilayer reaction method with the wheat germ extract WEPRO8240H (CellFree Sciences) as described previously[44,45]. mRNAs for cell-free synthesis were prepared by an in vitro transcription reaction while in vitro translation of target proteins was performed using a bilayer reaction where the translation layer was supplemented with L-Arg-$^{13}C_6$,$^{15}N_4$ and L-Lys-$^{13}C_6$,$^{15}N_2$ (Wako) at final concentrations of 20 mM to achieve high efficiency (>99 %) for stable-isotope labelling of proteins. The in vitro synthesised SIL-protein sequences also contained an N-terminal amino acid sequence GYSFTTTAEK that was later used as a tag for quantification of the SIL-protein stock concentration. Subsequently, SIL-proteins were purified using the Ni-Sepharose High-Performance resin (GE Healthcare Life Sciences) and precipitated using methanol:chloroform:water precipitation in Eppendorf Protein LoBind® tubes. Lastly, precipitated SIL-proteins were reconstituted in 104 µL of 8 M urea ([UA]; Sigma-Aldrich) in 0.1 M Trizma® base (pH 8.5) by vigorous vortexing and stored at -80 °C until further use.

Absolute quantification of SIL-protein standards using PRM MS. Concentrations of the twenty synthesised SIL-protein standard stocks were determined using PRM MS preceded by in-solution digestion of proteins and sample desalting and preparation for MS analysis.

Sample preparation

Only Eppendorf Protein LoBind® tubes and pipette tips were used for all sample preparation steps. Firstly, 20 µL of UA was added to 4 µL of the SIL-protein standard stock used to determine the stock concentration, and the mix was vortexed. Then, 1 µL of 0.2 M DTT (Promega) was added, followed

by vortexing and incubation for 1 h at 37 °C to reduce disulphide bonds. Sulfhydryl groups were alkylated with 2 µL of 0.5 M iodoacetamide (IAA; Sigma-Aldrich), vigorous vortexing, and incubation for 30 min at room temperature in the dark. Next, 75 µL of 25 mM ammonium bicarbonate was added to dilute UA down to 2 M concentration. Subsequently, 2 pmol (2 µL of stock) of the non-labelled AQUA® peptide HLEAAKGYSFTTTAEKAAELHK (Sigma-Aldrich) containing the quantification tag sequence GYSFTTTAEK was added to enable quantification of SIL-protein stock concentrations using MS analysis based on the ratio of heavy-to-light GYSFTTTAEK signals (see below). Protein digestion was performed for 16 h at 37 °C with 0.1 µg of Trypsin/Lys-C mix (1 µL of stock; Promega) and stopped by lowering pH to 3 by the addition of 5 µL of 10% (v/v) trifluoroacetic acid (TFA).

Samples were desalted using C_{18} ZipTips (Merck Millipore) as follows: the column was wetted using 0.1% (v/v) formic acid (FA) in 100% acetonitrile (ACN), equilibrated with 0.1% FA in 70% (v/v) ACN, and washed with 0.1% FA before loading the sample and washing again with 0.1% FA. Peptides were eluted from the ZipTips with 0.1% FA in 70% ACN. Finally, samples were dried using a vacuum-centrifuge (Eppendorf) at 30 °C until dryness followed by reconstitution in 12 µL of 0.1% FA in 5% ACN for subsequent MS analysis.

LC method for PRM MS

A Thermo Fisher Scientific UltiMate 3000 RSLCnano UHPLC system was used to elute the samples. Each sample was initially injected (6 µL) onto a Thermo Fisher Acclaim PepMap C_{18} trap reversed-phase column (300 µm x 5 mm nano viper, 5 µm particle size) at a flow rate of 15 µL/min using 2% ACN for 3 min with the solvent going to waste. The trap column was switched in-line with the separation column (GRACE Vydac Everest C_{18}, 300Å 150 µm x 150 mm, 2 µm) and the peptides were eluted using a flowrate of 3 µL/min using 0.1% FA in water (buffer A) and 80% ACN in buffer A (buffer B) as mobile phases for gradient elution. Following 3 min isocratic of 3% buffer B, peptide elution employed a 3-40% ACN gradient for 28 min followed by 40-95% ACN for 1.5 min and 95% ACN for 1.5 min at 40 °C. The total elution time was 50 min including a 95% ACN wash and a re-equilibration step.

PRM MS data acquisition

The eluted peptides from the C18 column were introduced to the MS via a nano-ESI and analysed using the Thermo Fisher Scientific Q-Exactive HF-X mass spectrometer. The electrospray voltage was 1.8 kV in positive ion mode, and the ion transfer tube temperature was 250 °C. Full MS-scans were acquired in the Orbitrap mass analyser over the range m/z 550–560 with a mass resolution of 30,000 (at m/z 200). The AGC target value was set at 1.00E+06 and maximum accumulation time 50 ms for full MS-scans. The PRM inclusion list included two mass values of 552.7640 and 556.7711. MS/MS spectra were acquired in the Orbitrap mass analyser with a mass resolution of 15,000 (at m/z 200). The AGC target value was set at 1.00E+06 and maximum accumulation time 30 ms for MS/MS with an isolation window of 2 m/z. The loop count was set at 14 to gain greater MS/MS data. Raw PRM MS data have been deposited to Panorama at https://panoramaweb.org/Valgepea_Cauto_PRM.url (private reviewer account details: username: panorama+reviewer27@proteinms.net; password: hSAwNUAL) with a ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) dataset identifier PXD025760.

PRM MS data analysis

Analysis of PRM MS data was performed using the software Skyline[48]. The following parameters were used to extract PRM MS data for the quantification tag sequence GYSFTTTAEK: three precursor isotope peaks with a charge of 2 (++) were included (monoisotopic; M+1; M+2); five of the most intense y product ions from ion 3 to last ion of charge state 1 and 2 among the precursor were picked; chromatograms were extracted with an ion match mass tolerance of 0.05 m/z for product ions by including all matching scans; full trypsin specificity with two missed cleavages allowed for peptides with a length of 8-25 AAs; cysteine carbamidomethylation as a fixed peptide modification. Additionally, peptide modifications included heavy labels for lysine and arginine as $^{13}C(6)^{15}N(2)$/+8.014 Da (K) and $^{13}C(6)^{15}N(4)$/+10.008 Da (R), respectively. This translated into the SIL-proteins and the non-labelled AQUA® peptide possessing the tag GYSFTTTAEK with m/z of 556.7711 and 552.7640, respectively. Hence, the concentrations of SIL-protein stocks were calculated based on the ratio of heavy-to-light GYSFTTTAEK signals and the spike-in of 2 pmol of the non-labelled AQUA® peptide (see above). High accuracy of quantification was evidenced by the very high similarity between both precursor peak areas and expected isotope distribution (R^2>0.99; idotp in

Skyline) and heavy and light peak areas (R^2>0.99; rdotp in Skyline) for all SIL-protein standard stocks. No heavy GYSFTTTAEK signal was detected for protein CAETHG_RS16140 (an acetylating acetaldehyde dehydrogenase in the NCBI annotation of sequence NC_022592.1[67]), thus 19 SIL-proteins could be used for following absolute proteome quantification in *C. autoethanogenum* (Supplementary Table 1). PRM MS data with all Skyline processing settings can be viewed and downloaded from Panorama at https://panoramaweb.org/Valgepea_Cauto_PRM.url (private reviewer account details: username: panorama+reviewer27@proteinms.net; password: hSAwNUAL).

Absolute proteome quantification in *C. autoethanogenum* using DIA MS. We used 19 synthetic heavy SIL variants (see above) of key *C. autoethanogenum* proteins (Supplementary Table 1) as spike-in standards for quantification of intracellular concentrations of their non-SIL counterparts using a DIA MS approach[46]. Also, we performed a dilution series experiment for the spike-in SIL-proteins to ensure accurate absolute quantification. We refer to these 19 intracellular proteins as anchor proteins that were further used to estimate proteome-wide absolute protein concentrations in *C. autoethanogenum*. This was achieved by determining the best linear fit between anchor protein concentrations and their measured DIA MS intensities using the same strategy as described previously[28].

Preparation of spike-in SIL-protein standard mix and dilution series samples

Only Eppendorf Protein LoBind® tubes and pipette tips were used for all preparation steps. The 19 spike-in SIL-protein standards that could be used for absolute proteome quantification in *C. autoethanogenum* (see above) were mixed in two lots: 1) 'sample spike-in standard mix': SIL-protein quantities matching estimated intracellular anchor protein quantities (i.e., expected light-to-heavy [L/H] ratios of ~1) based on label-free absolute quantification of the same samples in our previous work[18]; 2) 'dilution series standard mix': SIL-protein quantities doubling the estimated intracellular anchor protein quantities for the dilution series sample with the highest SIL-protein concentrations.

To ensure accurate absolute quantification of anchor protein concentrations, a dilution series experiment was performed to determine the linear dynamic quantification range and LLOQ for each of the 19 spike-in SIL-proteins. Dilution series samples were prepared by making nine 2-fold dilutions

444 of the 'dilution series spike-in standard mix' (i.e., 10 samples total for dilution series with a 512-fold
445 concentration span) in a constant *C. autoethanogenum* cell lysate background (0.07 µg/µL; 10
446 µg/tube) serving as a blocking agent to avoid loss of purified SIL-proteins (to container and pipette tip
447 walls) and as a background proteome for accurate MS quantification of the linear range and LLOQ for
448 anchor proteins.

449 ***Sample preparation***

450 *C. autoethanogenum* cultures were sampled for proteomics by immediate pelleting of 2 mL of culture
451 using centrifugation (25,000 × *g* for 1 min at 4 °C) and stored at -80 °C until analysis. Details of
452 protein extraction and protein quantification in cell lysates are described previously[18]. In short, thawed
453 cell pellets were suspended in lysis buffer (containing SDS, DTT, and Trizma® base) and cell lysis
454 was performed using glass beads and repeating a 'lysis cycle' consisting of heating, bead beating,
455 centrifugation, and vortexing before protein quantification using the 2D Quant Kit (GE Healthcare
456 Life Sciences).

457 Sample preparation and protein digestion for MS analysis was based on the filter-aided
458 sample preparation (FASP) protocol[68]. The following starting material was loaded onto an Amicon®
459 Ultra-0.5 mL centrifugal filter unit (nominal molecular weight cut-off of 30,000; Merck Millipore): 1)
460 50 µg of protein for one culture sample from each gas mixture (CO, syngas, or high-H_2 CO) for
461 building the spectral library for DIA MS data analysis (samples 1-3); 2) 7 µg of protein for one
462 culture sample from either syngas or high-H_2 CO plus 'sample spike-in standard mix' for including
463 spike-in SIL-protein data to the spectral library (samples 4-5); 3) 15 µg of protein for all 12 culture
464 samples (biological quadruplicates from CO, syngas, and high-H_2 CO) plus 'sample spike-in standard
465 mix' for performing absolute proteome quantification in *C. autoethanogenum* (samples 6-17); 4) ten
466 dilution series samples with 10-15 µg of total protein (*C. autoethanogenum* cell lysate background
467 plus 'dilution series spike-in standard mix') for performing the dilution series experiment for the 19
468 spike-in SIL-proteins (see above) (samples 18-27).

469 Samples containing SIL-proteins (samples 4-27) were incubated at 37 °C for 1 h to reduce
470 SIL-protein disulphide bonds (cell lysate contained DTT). Details of the FASP workflow are
471 described before[18]. In short, samples were washed with UA, sulfhydryl groups alkylated with IAA,

proteins digested using a Trypsin/Lys-C mix, and peptides eluted from the filter with 60 µL of ammonium bicarbonate. Next, 50 µL of samples 1-3 were withdrawn and pooled for performing high pH reverse-phase fractionation as described previously[18] for expanding the spectral library for DIA MS data analysis, yielding eight fractions (samples 28-35). Subsequently, all samples were vacuum-centrifuged at 30 °C until dryness followed by reconstitution of samples 1-3 and 4-35 in 51 and 13 µL of 0.1% FA in 5% ACN, respectively. Finally, total peptide concentration in each sample was determined using the Pierce™ Quantitative Fluorometric Peptide Assay (Thermo Fisher Scientific) to ensure that the same total peptide amount across samples 1-17 and 28-35 (excluding samples 18-27, see below) could be injected for DIA MS analysis.

LC method for data-dependent acquisition (DDA) and DIA MS

Details of the LC method employed for generating the spectral library using DDA and for DIA sample runs are described previously[20]. In short, a Thermo Fisher Scientific UHPLC system including C_{18} trap and separation columns was used to elute peptides with a gradient and total elution time of 110 min. For each DDA and DIA sample run, 1 µg of peptide material from protein digestion was injected, except for dilution series samples (samples 18-27 above) that were injected in a constant volume of 3 µL to maintain the dilution levels of the 'dilution series spike-in standard mix'.

DDA MS spectral library generation

The following 13 samples were analysed on the Q-Exactive HF-X in DDA mode to yield the spectral library for DIA MS data analysis: 1) three replicates of one culture sample from each gas mixture (CO, syngas, or high-H_2 CO) (samples 1-3 above); 2) three replicates of one culture sample from either syngas or high-H_2 CO plus 'sample spike-in standard mix' (samples 4-5); 3) eight high pH reverse-phase fractions of a pool of samples from each gas mixture (samples 28-35).

Details of DDA MS acquisition for generating the spectral library are described before[20]. In short, eluted peptides from the C_{18} column were introduced to the MS *via* a nano-ESI and analysed using the Q-Exactive HF-X with an Orbitrap mass analyser. The DDA MS spectral library for DIA MS data confirmation and quantification using the software Skyline[48] was created using the Proteome Discoverer 2.2 software (Thermo Fisher Scientific) and its SEQUEST HT search as described previously[18]. The final .pd result file contained peptide-spectrum matches (PSMs) with q-values

estimated at 1% false discovery rate (FDR) for peptides ≥4 AAs. The generated spectral library file has been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the PRIDE partner repository[69] with the dataset identifier PXD025732 (private reviewer account details: username: reviewer_pxd025732@ebi.ac.uk; password: QdkJHXy0).

DIA MS data acquisition

Details of DIA MS acquisition are described before[20]. In short, as for DDA MS acquisition, eluted peptides were introduced to the MS *via* a nano-ESI and analysed using the Q-Exactive HF-X with an Orbitrap mass analyser. DIA was achieved using an inclusion list: m/z 395□1100 in steps of 15 amu and scans cycled through the list of 48 isolation windows with a loop count of 48. In total, DIA MS data was acquired for 22 samples (samples 6-27 defined in section *Sample preparation*). Raw DIA MS data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the PRIDE partner repository[69] with the dataset identifier PXD025732 (private reviewer account details: username: reviewer_pxd025732@ebi.ac.uk; password: QdkJHXy0).

DIA MS data analysis

DIA MS data analysis was performed with Skyline[48] as described before[18] with the following modifications: 1) 12 manually picked high confidence endogenous peptides present in all samples and spanning the elution gradient were used for iRT alignment through building an RT predictor; 2) outlier peptides from iRT regression were removed; 3) a minimum of three isotope peaks were required for a precursor; 4) single peptide per spike-in SIL-protein was allowed for anchor protein absolute quantification while at least two peptides per protein were required for label-free estimation of proteome-wide protein concentrations; 5) extracted ion chromatograms (XICs) were transformed using Savitszky-Golay smoothing. Briefly, the .pd result file from Proteome Discoverer was used to build the DIA MS spectral library and the mProphet peak picking algorithm[47] within Skyline was used to separate true from false positive peak groups (per sample) and only peak groups with q-value<0.01 (representing 1% FDR) were used for further quantification. We confidently quantitated 7,288 peptides and 1,243 proteins across all samples and 4,887 peptides and 1,043 proteins on average

within each sample for estimating proteome-wide absolute protein concentrations. For absolute quantification of anchor protein concentrations, we additionally manually: 1) removed integration of peaks showing non-Gaussian shapes or interference from other peaks; 2) removed precursors with similarity measures of R^2<0.9 between product peak areas and corresponding intensities in the spectral library (dotp in Skyline), precursor peak areas and expected isotope distribution (idotp), or heavy and light peak areas (rdotp). After analysis in Skyline, 17 spike-in SIL-proteins remained for further analysis as protein CAETHG_RS14410 was not identified in DIA MS data while CAETHG_RS18395 did not pass quantification filters (Supplementary Table 1). DIA MS data with all Skyline processing settings can be viewed and downloaded from Panorama at https://panoramaweb.org/Valgepea_Cauto_Anchors.url for anchor protein absolute quantification and at https://panoramaweb.org/Valgepea_Cauto_LF.url for estimating proteome-wide absolute protein concentrations (private reviewer account details: username: panorama+reviewer27@proteinms.net; password: hSAwNUAL).

Absolute quantification of anchor protein concentrations

We employed further stringent criteria on top of the output from Skyline analysis to ensure high confidence absolute quantification of 17 anchor protein concentrations. Firstly, precursor with highest heavy intensity for the highest 'dilution series spike-in standard mix' sample in the dilution series (DS01) was kept while others were deleted. Peptides quantified in less than three biological replicates cultures within a gas mixture, with no heavy signal for DS01 sample, or with heavy signals for less than three continuous dilution series samples were removed. Next, we utilised the dilution series experiment to only keep signals over the LLOQ and within the linear dynamic quantification range. For this, correlation between experimental and expected peptide L/H signal ratios for each peptide across the dilution series was made to determine the LLOQ and calculate correlation, slope, and intercept between MS signal and spike-in level (Supplementary Table 2). Only peptides showing correlation R^2>0.95, 0.95<slope<1.05, and -0.1<intercept<0.1 for the dilution series were kept. This ensured that we were only using peptides within the linear dynamic range. The remaining peptides were further filtered for each culture sample by removing peptides whose light or heavy signal was below the LLOQ in the dilution series. Subsequently, only peptides were kept with L/H ratios for at

556 least three biological replicates cultures for each gas mixture (i.e., ≥9 data points). Finally, we aimed
557 to detect outlier peptides by calculating the % difference of a peptide's L/H ratio from the average
558 L/H ratio of all peptides for a given protein for every sample. Peptides were considered outliers and
559 thus removed if the average difference across all samples was >50% or if the average difference
560 within biological replicate cultures was >50%. After the previous stringent criteria were applied, 106
561 high-confidence peptides remained (Supplementary Table 2) for the quantification of 16 anchor
562 protein concentrations since CAETHG_RS01830 was lost during manual analysis (Table 1 and
563 Supplementary Table 1). Proteins CAETHG_RS13725 and CAETHG_RS07840 were excluded from
564 the high-H_2 CO culture dataset as their calculated concentrations varied >50% between biological
565 replicates. Data of one high-H_2 CO culture was excluded from further analysis due to difference from
566 bio-replicates likely due to challenges with MS analysis.

567 ***Label-free estimation of proteome-wide protein concentrations***

568 We used the anchor proteins to estimate proteome-wide protein concentrations in *C.*
569 *autoethanogenum* by determining the best linear fit between anchor protein concentrations and their
570 measured DIA MS intensities using the aLFQ R package[49] and the same strategy as described for
571 SWATH MS[28]. Briefly, aLFQ used anchor proteins and cross-validated model selection by
572 bootstrapping to determine the optimal model within various label-free absolute proteome
573 quantification approaches (e.g., TopN, iBAQ). The approach can obtain the model with the smallest
574 error between anchor protein concentrations determined using SIL-protein standards and label-free
575 estimated concentrations. The models with the highest accuracy were used to estimate proteome-wide
576 label-free concentrations for all proteins from their DIA MS intensities (1,043 proteins on average
577 within each sample; minimal two peptides per protein; see above): summing the five most intense
578 fragment ion intensities of the most or three of the most intense peptides per protein for CO or high-
579 H_2 CO cultures, respectively; summing the five most intense fragment ion intensities of all quantified
580 peptides of the protein divided by the number of theoretically observable peptides (i.e., iBAQ[70]) for
581 syngas cultures.
582

583 **Expected protein complex stoichiometries.** Equimolar stoichiometries for the HytA-E/FdhA and
584 MetFV protein complexes were expected based on SDS gel staining experiments in *C.*
585 *autoethanogenum*[57] and the acetogen *Moorella thermoacetica*[71], respectively. Expected
586 stoichiometries for other protein complexes in Fig. 3 and Supplementary Fig. 3 were based on
587 measured stoichiometries in *E. coli* K-12 (Complex Portal; www.ebi.ac.uk/complexportal) and
588 significant homology between complex member proteins in *C. autoethanogenum* and *E. coli*. All
589 depicted *C. autoethanogenum* protein complex members had NCBI protein-protein BLAST E-
590 values<10^{-16} and scores>73 against respective *E. coli* K-12 proteins using non-redundant protein
591 sequences.
592
593 **Generation of proteomaps.** The distribution of proteome-wide protein concentrations among
594 functional gene classifications was visualised using proteomaps[27]. For this, the NCBI annotation of
595 sequence NC_022592.1[67] was used as the annotation genome for *C. autoethanogenum*, with
596 CAETHG_RS07860 removed and replaced with the carbon monoxide dehydrogenase genes named
597 CAETHG_RS07861 and CAETHG_RS07862 with initial IDs of CAETHG_1620 and 1621,
598 respectively. Functional categories were assigned to protein sequences with KO IDs[51] using the
599 KEGG annotation tool BlastKOALA[72]. Since proteomaps require a tree-like hierarchy, proteins that
600 were automatically assigned to multiple functional categories were manually assigned to one bottom-
601 level category (Level 3 in Supplementary Table 4) based on their principal task. We also created
602 functional categories "C1 fixation/Wood-Ljungdahl Pathway" (Level 2/3), "Acetate & ethanol
603 synthesis" (Level 3), "Energy conservation" (Level 3), "Hydrogenases" (Level 3) and manually
604 assigned key acetogen proteins to these categories to reflect more accurately functional categories for
605 an acetogen. Proteins without designated KO IDs were manually assigned to latterly created
606 categories or grouped under "Not Included in Pathway or Brite" (Level 1) with Level 2 and 3 as "No
607 KO ID". If BlastKOALA assigned multiple genes the same proposed gene/protein name, unique
608 numbers were added to names (e.g., pfkA, pfkA2). The final "treemap" defining the hierarchy for our
609 proteomaps is in Supplementary Table 4.
610

Calculation of apparent in vivo catalytic rates of enzymes (k_{app}). We calculated apparent in vivo catalytic rates of enzymes, denoted as k_{app} (s^{-1})[26], as the ratio of specific flux rate (mmol/gDCW/h) determined before[18] and protein concentration (nmol/gDCW) quantified here for the same *C. autoethanogenum* CO, syngas, and high-H_2 CO cultures. Gene-protein-reaction (GPR) data of the genome-scale metabolic model iCLAU786[18] were manually curated to reflect most recent knowledge and were used to link metabolic fluxes with catalysing enzymes. For reactions with multiple assigned enzymes (i.e,. isoenzymes), the enzyme with the highest average ranking of its concentration across the three cultures (Supplementary Table 3) was assumed to solely catalyse the flux. For enzyme complexes, average of quantified subunit concentrations was used (standard deviation estimated using error propagation). For the HytA-E hydrogenase, its measured protein concentration was split between reactions "rxn08518_c0" (direct CO_2 reduction with H_2 in complex with FdhA) and "leq000001_c0" (H_2 oxidation solely by HytA-E) proportionally to flux for syngas and high-H_2 CO cultures. The resulting enzymes or enzyme complexes catalysing specific fluxes are shown in Supplementary Table 5. Finally, we assumed each protein chain being catalytically active and only calculated k_{app} values for metabolic reactions with a non-zero flux in at least one condition, specific flux rate>0.1% of CO fixation flux in at least one condition, and label-free data with measured concentration for the associated enzyme(s) in all conditions (Supplementary Table 5). Membrane proteins were excluded from k_{app} calculations to avoid bias from potentially incomplete protein extraction. This produced k_{app} values for 13 enzymes/complexes using anchor protein concentrations and for 48 enzymes/complexes using label-free protein concentrations (Supplementary Table 5).

Determination of regulation level of metabolic fluxes. We used published flux and relative proteomics data[18] of the same cultures studied here to determine whether fluxes are regulated by changing enzyme concentrations or their catalytic rates by considering metabolic fluxes with non-zero specific flux rates in at least two conditions of CO, syngas, or high-H_2 CO cultures. The same manually curated GPRs and criteria for isoenzymes and protein complexes as described above for k_{app} calculation were used to determine flux-enzyme pairs (Supplementary Table 6). We first used a two-tailed two-sample equal variance Student's t-test with FDR correction[73] to determine fluxes with

639 significant changes between any two conditions (q-value<0.05). We then used the Student's left-tailed
640 t-distribution with FDR to determine if the significant flux change for every flux was significantly
641 different from the change in respective enzyme expression between the same conditions
642 (Supplementary Table 6). Flux with a q-value<0.05 for the latter test was considered to be regulated at
643 post-translational level (e.g., by changing enzyme catalytic rate).
644

645 **DATA AVAILABILITY**

646 All data generated or analysed during this study are in the main text, supplementary information files,
647 or public databases. Raw PRM MS data have been deposited to Panorama at
648 https://panoramaweb.org/Valgepea_Cauto_PRM.url (private reviewer account details: username:
649 panorama+reviewer27@proteinms.net; password: hSAwNUAL) with a ProteomeXchange
650 Consortium (http://proteomecentral.proteomexchange.org) dataset identifier PXD025760. Raw DIA
651 MS data have been deposited to the ProteomeXchange Consortium
652 (http://proteomecentral.proteomexchange.org) *via* the PRIDE partner repository[69] with the dataset
653 identifier PXD025732 (private reviewer account details: username: reviewer_pxd025732@ebi.ac.uk;
654 password: QdkJHXy0). PRM MS data with all Skyline processing settings can be viewed and
655 downloaded from Panorama at https://panoramaweb.org/Valgepea_Cauto_PRM.url. DIA MS data
656 with all Skyline processing settings can be viewed and downloaded from Panorama at
657 https://panoramaweb.org/Valgepea_Cauto_Anchors.url for anchor protein absolute quantification and
658 at https://panoramaweb.org/Valgepea_Cauto_LF.url for estimating proteome-wide absolute protein
659 concentrations (private reviewer account details: username: panorama+reviewer27@proteinms.net;
660 password: hSAwNUAL). Any other relevant data are available from the corresponding author on
661 reasonable request.
662

663 **REFERENCES**

664 1. Liew, F. et al. Gas Fermentation—A flexible platform for commercial scale production of low-
665 carbon-fuels and chemicals from waste and renewable feedstocks. *Front. Microbiol.* **7,** 694 (2016).

2. Claassens, N. J., Sousa, D. Z., dos Santos, V. A. P. M., de Vos, W. M. & van der Oost, J. Harnessing the power of microbial autotrophy. *Nat. Rev. Microbiol.* **14,** 692–706 (2016).

3. Fackler, N. et al. Stepping on the gas to a circular economy: accelerating development of carbon-negative chemical production from gas fermentation. *Annu. Rev. Chem. Biomol. Eng.* **12,** 1 (2021).

4. Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood–Ljungdahl pathway of CO_2 fixation. *Biochim. Biophys. Acta* **1784,** 1873–1898 (2008).

5. Wood, H. G. Life with CO or CO_2 and H_2 as a source of carbon and energy. *FASEB J.* **5,** 156–163 (1991).

6. Drake, H. L., Küsel, K. & Matthies, C. In *The Prokaryotes* Ch. Acetogenic Prokaryotes. 354–420 (2006).

7. Fuchs, G. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Ann. Rev. Microbiol.* **65,** 631–658 (2011).

8. Fast, A. G. & Papoutsakis, E. T. Stoichiometric and energetic analyses of non-photosynthetic CO_2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Curr. Opin. Chem. Eng.* **1,** 380–395 (2012).

9. Cotton, C. A., Edlich-Muth, C. & Bar-Even, A. Reinforcing carbon fixation: CO_2 reduction replacing and supporting carboxylation. *Curr. Opin. Biotechnol.* **49,** 49–56 (2018).

10. Köpke, M. & Simpson, S. D. Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Curr. Opin. Biotechnol.* **65,** 180–189 (2020).

11. Russell, M. J. & Martin, W. The rocky roots of the acetyl-CoA pathway. *Trends Biochem. Sci.* **29,** 358–363 (2004).

12. Weiss, M. C. et al. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1,** 16116 (2016).

13. Varma, S. J., Muchowska, K. B., Chatelain, P. & Moran, J. Native iron reduces CO_2 to intermediates and end-products of the acetyl-CoA pathway. *Nat. Ecol. Evol.* **2,** 1019–1024 (2018).

14. Ljungdahl, L. G. A life with acetogens, thermophiles, and cellulolytic anaerobes. *Annu. Rev. Microbiol.* **63,** 1–25 (2009).

693 15. Ragsdale, S. W. Enzymology of the Wood-Ljungdahl pathway of acetogenesis. *Ann. N. Y. Acad.*
694 *Sci.* **1125,** 129–136 (2008).
695 16. Schuchmann, K. & Müller, V. Autotrophy at the thermodynamic limit of life: a model for energy
696 conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* **12,** 809–821 (2014).
697 17. Molitor, B., Marcellin, E. & Angenent, L. T. Overcoming the energetic limitations of syngas
698 fermentation. *Curr. Opin. Chem. Biol.* **41,** 84–92 (2017).
699 18. Valgepea, K. et al. H_2 drives metabolic rearrangements in gas-fermenting *Clostridium*
700 *autoethanogenum*. *Biotechnol. Biofuels* **11,** 55 (2018).
701 19. Valgepea, K. et al. Maintenance of ATP homeostasis triggers metabolic shifts in gas-fermenting
702 acetogens. *Cell Syst.* **4,** 505–515.e5 (2017).
703 20. Mahamkali, V. et al. Redox controls metabolic robustness in the gas-fermenting acetogen
704 *Clostridium autoethanogenum*. *Proc. Natl. Acad. Sci. U. S. A.* **117,** 13168–13175 (2020).
705 21. Richter, H. et al. Ethanol production in syngas-fermenting *Clostridium ljungdahlii* is controlled by
706 thermodynamics rather than by enzyme expression. *Energy Environ. Sci.* **9,** 2392–2399 (2016).
707 22. de Souza Pinto Lemgruber, R. et al. A TetR-family protein (CAETHG_0459) activates
708 transcription from a new promoter motif associated with essential genes for autotrophic growth in
709 acetogens. *Front. Microbiol.* **10,** 2549 (2019).
710 23. Song, Y. et al. Determination of the genome and primary transcriptome of syngas fermenting
711 *Eubacterium limosum* ATCC 8486. *Sci. Rep.* **7,** 13694 (2017).
712 24. Al-Bassam, M. M. et al. Optimisation of carbon and energy utilisation through differential
713 translational efficiency. *Nat. Commun.* **9,** 4474 (2018).
714 25. Song, Y. et al. Genome-scale analysis of syngas fermenting acetogenic bacteria reveals the
715 translational regulation for its autotrophic growth. *BMC Genomics* **19,** 837 (2018).
716 26. Valgepea, K., Adamberg, K., Seiman, A. & Vilu, R. *Escherichia coli* achieves faster growth by
717 increasing catalytic and translation rates of proteins. *Mol. Biosyst.* **9,** 2344–2358 (2013).
718 27. Liebermeister, W. et al. Visual account of protein investment in cellular functions. *Proc. Natl.*
719 *Acad. Sci. U. S. A.* **111,** 8488–8493 (2014).

720 28. Schubert, O. T. et al. Absolute proteome composition and dynamics during dormancy and
721 resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* **18,** 96–108 (2015).
722 29. Davidi, D. et al. Global characterisation of in vivo enzyme catalytic rates and their correspondence
723 to in vitro kcat measurements. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 20167–20172 (2016).
724 30. Calderón-Celis, F., Encinar, J. R. & Sanz-Medel, A. Standardization approaches in absolute
725 quantitative proteomics with mass spectrometry. *Mass Spectrom. Rev.* **37,** 715–737 (2017).
726 31. Maaβ, S. & Becher, D. Methods and applications of absolute protein quantification in microbial
727 systems. *J. Proteomics* **136,** 222–233 (2016).
728 32. Ludwig, C., Claassen, M., Schmidt, A. & Aebersold, R. Estimation of absolute protein quantities
729 of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol. Cell. Proteomics* **11,**
730 M111.013987 (2012).
731 33. Blein-Nicolas, M. & Zivy, M. Thousand and one ways to quantify and compare protein
732 abundances in label-free bottom-up proteomics. *Biochim. Biophys. Acta* **1864,** 883–895 (2016).
733 34. Ahrné, E., Molzahn, L., Glatter, T. & Schmidt, A. Critical assessment of proteome-wide label-free
734 absolute abundance estimation strategies. *Proteomics* **17,** 2567–2578 (2013).
735 35. Schmidt, A. et al. Absolute quantification of microbial proteomes at different states by directed
736 mass spectrometry. *Mol. Syst. Biol.* **7,** 510 (2011).
737 36. Maier, T. et al. Quantification of mRNA and protein and integration with protein turnover in a
738 bacterium. *Mol. Syst. Biol.* **7,** 511 (2011).
739 37. Marguerat, S. et al. Quantitative analysis of fission yeast transcriptomes and proteomes in
740 proliferating and quiescent cells. *Cell* **151,** 671–683 (2012).
741 38. Zeiler, M., Moser, M. & Mann, M. Copy number analysis of the murine platelet proteome
742 spanning the complete abundance range. *Mol. Cell. Proteomics* **13,** 3435–3445 (2014).
743 39. Schmidt, A. et al. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat.*
744 *Biotechnol.* **34,** 104–110 (2015).
745 40. Beck, M. et al. The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7,** 549 (2011).
746 41. Malmström, J. et al. Proteome-wide cellular protein concentrations of the human pathogen
747 *Leptospira interrogans*. *Nature* **460,** 762–765 (2009).

748 42. Brun, V. et al. Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol.*
749 *Cell. Proteomics* **6,** 2139–2149 (2007).
750 43. Shuford, C. M. et al. Absolute protein quantification by mass spectrometry: not as simple as
751 advertised. *Anal. Chem.* **89,** 7406–7415 (2017).
752 44. Takemori, N. et al. High-throughput synthesis of stable isotope-labeled transmembrane proteins
753 for targeted transmembrane proteomics using a wheat germ cell-free protein synthesis system. *Mol.*
754 *Biosyst.* **11,** 361–365 (2015).
755 45. Takemori, N. et al. MEERCAT: multiplexed efficient cell free expression of recombinant
756 QconCATs for large scale absolute proteome quantification. *Mol. Cell. Proteomics* **16,** 2169–2183
757 (2017).
758 46. Gillet, L. C. et al. Targeted data extraction of the MS/MS spectra generated by data-independent
759 acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11,**
760 O111.016717 (2012).
761 47. Reiter, L. et al. mProphet: automated data processing and statistical validation for large-scale
762 SRM experiments. *Nat. Methods* **8,** 430–435 (2011).
763 48. MacLean, B. et al. Skyline: an open source document editor for creating and analysing targeted
764 proteomics experiments. *Bioinformatics* **26,** 966–968 (2010).
765 49. Rosenberger, G., Ludwig, C., Röst, H. L., Aebersold, R. & Malmström, L. aLFQ: an R-package
766 for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. *Bioinformatics*
767 **30,** 2511–2513 (2014).
768 50. Lahtvee, P.-J. et al. Absolute quantification of protein and mRNA abundances demonstrate
769 variability in gene-specific translation efficiency in yeast. *Cell Syst.* 4, 495–504.e5 (2017).
770 51. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference
771 resource for gene and protein annotation. *Nucleic Acids Res.* **44,** D457–D462 (2016).
772 52. Peebo, K. et al. Proteome reallocation in *Escherichia coli* with increasing specific growth rate.
773 *Mol. Biosyst.* **11,** 1184–1193 (2015).

774 53. Lemaire, O. N. & Wagner, T. Gas channel rerouting in a primordial enzyme: Structural insights of
775 the carbon-monoxide dehydrogenase/acetyl-CoA synthase complex from the acetogen *Clostridium*
776 *autoethanogenum*. *Biochim. Biophys. Acta* **1862,** 148330 (2021).
777 54. Liew, F. et al. Insights into CO_2 fixation pathway of *Clostridium autoethanogenum* by targeted
778 mutagenesis. *mBio* **7,** e00427-16 (2016).
779 55. Liew, F. et al. Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol
780 production. *Metab. Eng.* **40,** 104–114 (2017).
781 56. Marcellin, E. et al. Low carbon fuels and commodity chemicals from waste gases – systematic
782 approach to understand energy metabolism in a model acetogen. *Green Chem.* **18,** 3020–3028 (2016).
783 57. Wang, S. et al. NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex
784 with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO. *J. Bacteriol.* **195,** 4373–
785 4386 (2013).
786 58. Mock, J. et al. Energy conservation associated with ethanol formation from H_2 and CO_2 in
787 *Clostridium autoethanogenum* involving electron bifurcation. *J. Bacteriol.* **197,** 2965–2980 (2015).
788 59. Maia, L. B., Fonseca, L., Moura, I. & Moura, J. J. G. Reduction of carbon dioxide by a
789 molybdenum-containing formate dehydrogenase: a kinetic and mechanistic study. *J. Am. Chem. Soc.*
790 **138,** 8834–8846 (2016).
791 60. Schuchmann, K. & Müller, V. Direct and reversible hydrogenation of CO_2 to formate by a
792 bacterial carbon dioxide reductase. *Science* **342,** 1382–1385 (2013).
793 61. Bar-Even, A. et al. The moderately efficient enzyme: evolutionary and physicochemical trends
794 shaping enzyme parameters. *Biochemistry* **50,** 4402–4410 (2011).
795 62. Song, Y. et al. Functional cooperation of the glycine synthase-reductase and Wood–Ljungdahl
796 pathways for autotrophic growth of *Clostridium drakei*. *Proc. Natl. Acad. Sci. U. S. A.* **117,** 7516–
797 7523 (2020).
798 63. Tan, Y., Liu, J., Liu, Z. & Li, F. Characterization of two novel butanol dehydrogenases involved
799 in butanol degradation in syngas-utilising bacterium *Clostridium ljungdahlii* DSM 13528. *J. Basic*
800 *Microbiol.* **54,** 996–1004 (2014).

64. Davidi, D. & Milo, R. Lessons on enzyme kinetics from quantitative proteomics. *Curr. Opin. Biotechnol.* **46,** 81–89 (2017).

65. Nilsson, A., Nielsen, J. & Palsson, B. Ø. Metabolic models of protein allocation call for the kinetome. *Cell Syst.* **5,** 538–541 (2017).

66. Heckmann, D. et al. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat. Commun.* **9,** 5252 (2018).

67. Brown, S. D. et al. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant *Clostridia*. *Biotechnol. Biofuels* **7,** 40 (2014).

68. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362 (2009).

69. Vizcaíno, J. A. et al. The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41,** D1063–D1069 (2012).

70. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).

71. Mock, J., Wang, S., Huang, H., Kahnt, J. & Thauer, R. K. Evidence for a hexaheteromeric methylenetetrahydrofolate reductase in *Moorella thermoacetica*. *J. Bacteriol.* **196,** 3303–3314 (2014).

72. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterisation of genome and metagenome sequences. *J. Mol. Biol.* **428,** 726–731 (2016).

73. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).

ACKNOWLEDGEMENTS

We thank Jörg Bernhardt for help with proteomaps, Tim McCubbin for scripts, Andrus Seiman for statistics, and Olivier Lemaire for valuable discussions. This work was funded by the Australian Research Council (ARC LP140100213) in collaboration with LanzaTech. We thank the following investors in LanzaTech's technology: Sir Stephen Tindall, Khosla Ventures, Qiming Venture Partners, Softbank China, the Malaysian Life Sciences Capital Fund, Mitsui, Primetals, CICC Growth Capital

Fund I, L.P. and the New Zealand Superannuation Fund. The research utilised equipment and support provided by the Queensland node of Metabolomics Australia, an initiative of the Australian Government being conducted as part of the NCRIS National Research Infrastructure for Australia. There was no funding support from the European Union for the experimental part of the study. However, K.V. acknowledges support also from the European Union's Horizon 2020 research and innovation programme under grant agreement N810755.

AUTHOR CONTRIBUTIONS

Conceptualisation, K.V., C.L., M.K., L.K.N., and E.M.; Methodology, K.V., G.T., N.T., A.T., C.L., A.P.M., R.T., and E.M.; Formal Analysis, K.V., and G.T.; Investigation, K.V., G.T., N.T., and A.T.; Resources, N.T., A.T., M.K., S.D.S., L.K.N., and E.M.; Writing – Original Draft, K.V. and E.M.; Writing – Review & Editing, K.V., N.T., C.L., A.P.M., R.T., M.K., L.K.N., and E.M.; Supervision, M.K., L.K.N., and E.M.; Project Administration, R.T., and E.M.; Funding Acquisition, M.K., S.D.S., L.K.N., and E.M..

COMPETING INTERESTS

LanzaTech has interest in commercial gas fermentation with *C. autoethanogenum*. A.P.M, R.T., M.K., and S.D.S. are employees of LanzaTech.

MATERIALS & CORRESPONDENCE

Correspondence and request for materials should be addressed to E.M.

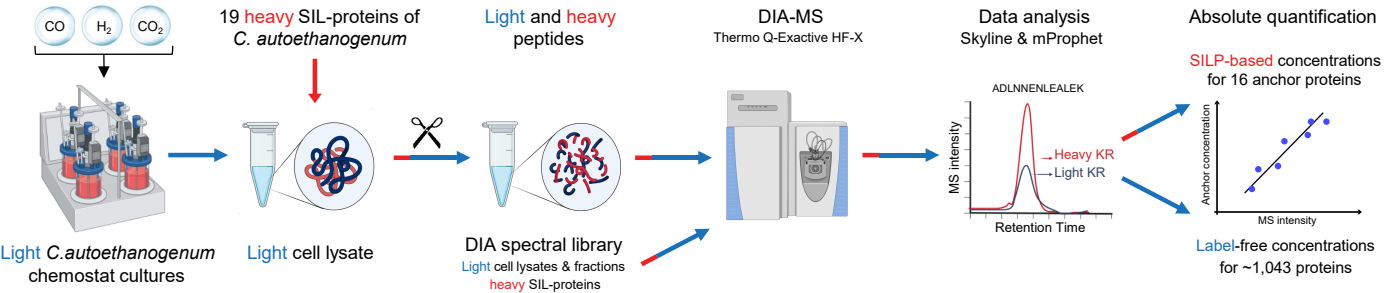
FIGURE LEGENDS

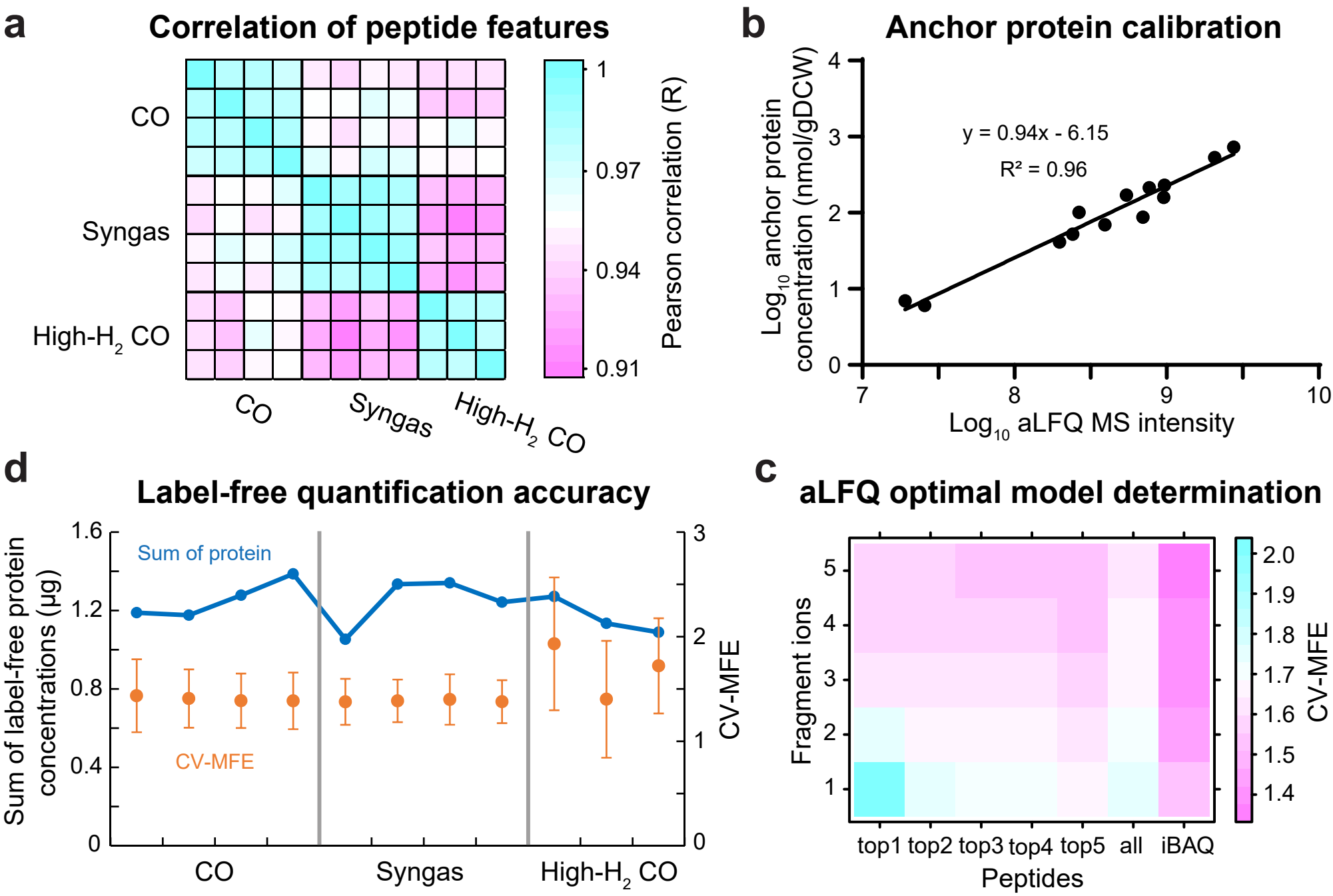
Fig. 1 Absolute proteome quantification framework in *C. autoethanogenum*. Absolute proteome quantification in light (no stable-isotope labelled [SIL] substrates) autotrophic *C. autoethanogenum* chemostat cultures was built on using 19 synthetic heavy SIL-protein spike-in standards and data-independent acquisition (DIA) mass spectrometry (MS) analysis. Culture samples with SIL-protein spike-ins and samples for DIA spectral library were analysed by DIA MS. Subsequent stringent data

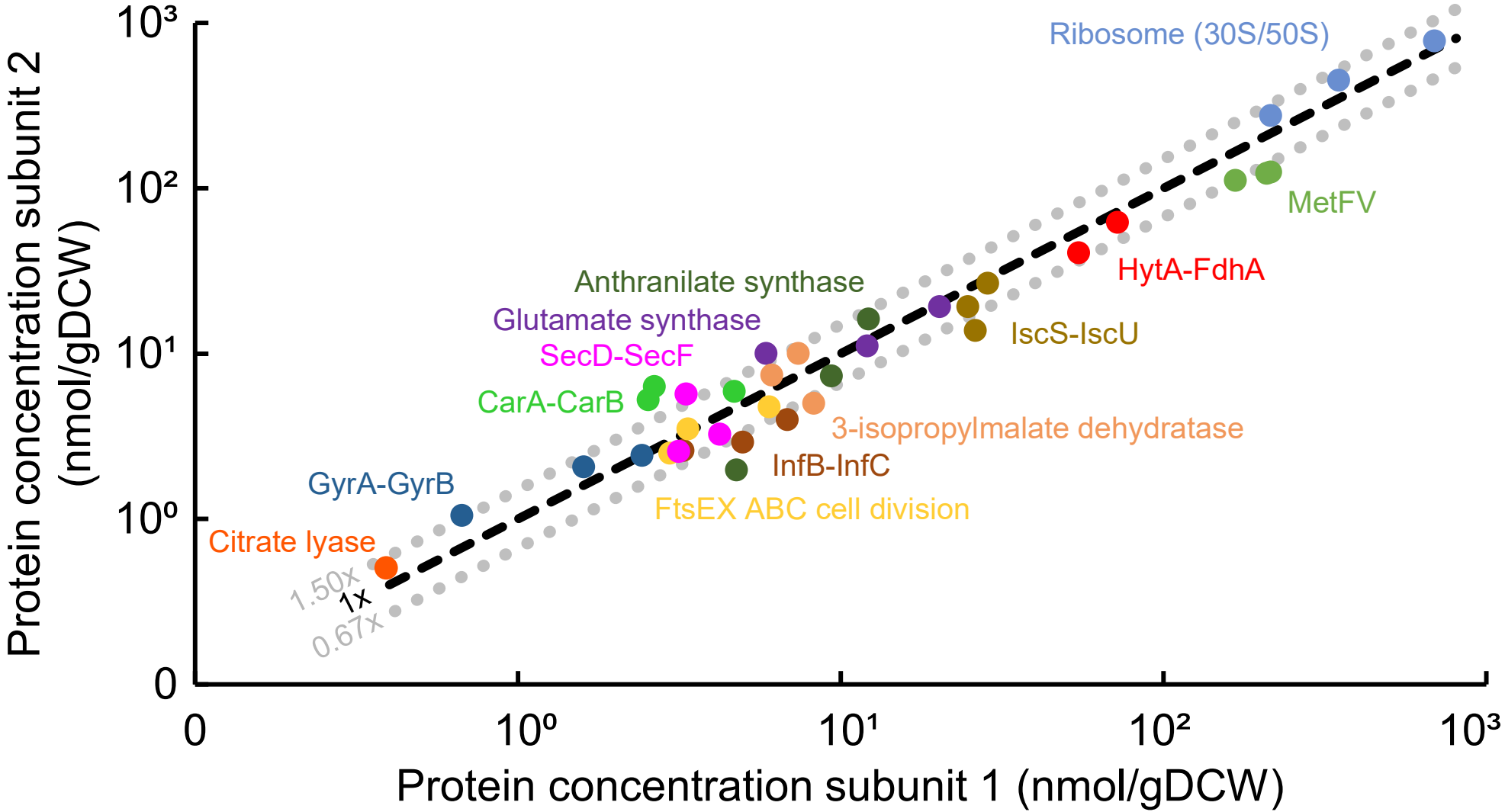
857 analysis allowed to quantify intracellular concentrations for 16 key *C. autoethanogenum* proteins
858 using light-to-heavy ratios between endogenous and spike-in DIA MS intensities. These 16 key
859 proteins were further used as anchor proteins for label-free estimation of ~1,043 protein
860 concentrations through establishing a linear correlation between protein concentrations and their
861 measured MS intensities. Some parts created with BioRender.com.
862
863 **Fig. 2 Label-free estimation of proteome-wide protein concentrations. a** Correlation of peptide
864 mass spectrometry (MS) feature intensities between biological replicate cultures of the three gas
865 mixtures. **b** Linear correlation between anchor protein concentrations and their measured MS
866 intensities for one syngas culture. *gDCW*, gram of dry cell weight, *aLFQ*, absolute label-free
867 quantification. **c** Errors of different label-free quantification models for the linear fit between anchor
868 protein concentrations and their measured MS intensities determined by bootstrapping using the aLFQ
869 R package[49] for one syngas culture. *CV-MFE*, cross-validated mean fold-error. **d** Label-free
870 quantification error of optimal model (orange) and total proteome mass (blue) across samples. Error
871 bars denote 95% CI.
872
873 **Fig. 3 Strong correlation between protein concentrations and expected stoichiometries for**
874 **equimolar protein complexes.** Grey dotted lines denote the average 1.5-fold cross-validated mean
875 fold-error (CV-MFE) of label-free protein concentrations. Label-free protein concentrations are
876 plotted, except for the HytA-FdhA complex, which was quantified using stable-isotope labelled
877 protein spike-ins. Data points of the same colour represent gas mixtures. See Methods for details on
878 expected protein complex stoichiometries. See Supplementary Table S3 for gene/protein ID, proposed
879 name, description, and label-free data. See Table 1 for HytA-FdhA data. *gDCW*, gram of dry cell
880 weight.
881
882 **Fig. 4 Proteomaps uncover global proteome allocation.** Left proteomap shows proteome allocation
883 amongst functional gene classification categories (KEGG Orthology identifiers [KO IDs] [51]) at level
884 two of the four-level "treemap" hierarchy structure (Supplementary Table S4). Right proteomap

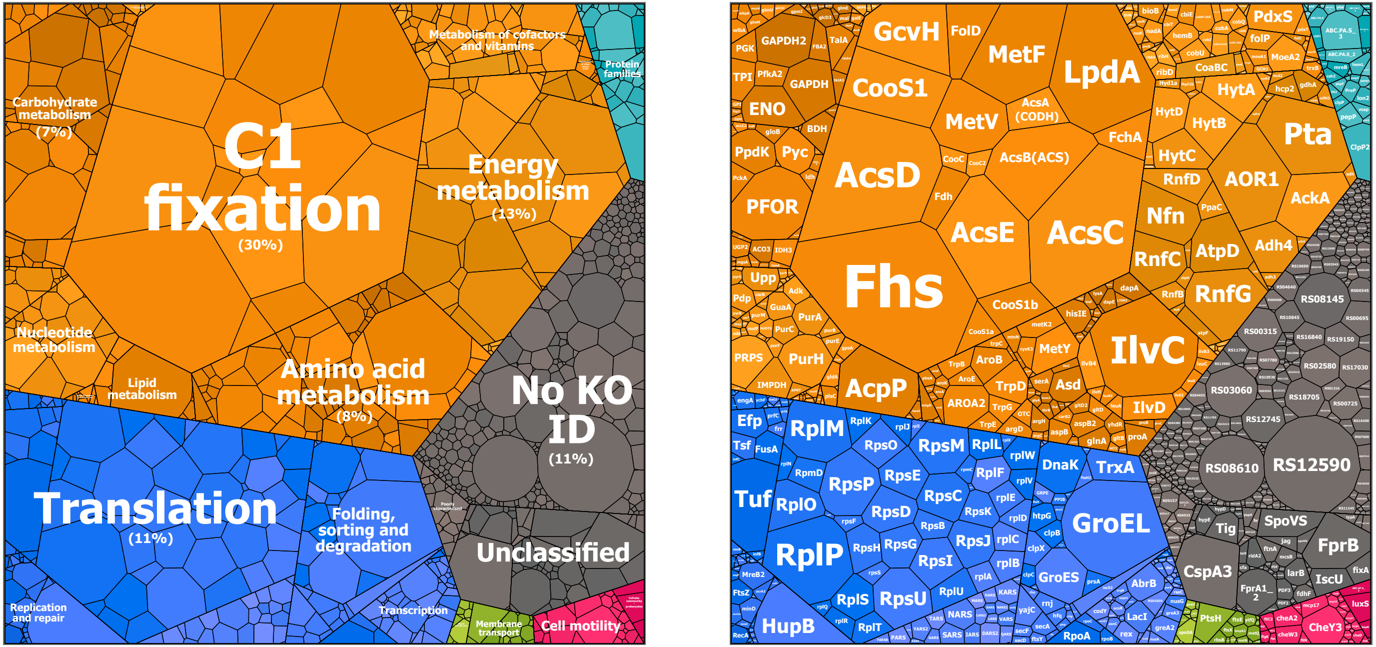
885 shows proteome allocation at the level of single proteins (level four of "treemap"). See Supplementary
886 Fig. S5 for proteomaps of levels one and three of "treemap". Area of the tile is proportional to protein
887 concentration. Colours denote level one categories of "treemap". Proteomaps visualise average
888 concentrations of syngas cultures while category percentages are average of three gas mixtures
889 (shown for categories with a fraction >5%). See Supplementary Table S3 for gene/protein ID,
890 proposed name, description, and label-free protein concentrations.
891
892 **Fig. 5 Quantitative systems-level view of acetogen central metabolism.** Enzyme concentrations
893 (nmol/gDCW), apparent in vivo catalytic rates of enzymes (k_{app}; s^{-1}), and metabolic flux rates
894 (mmol/gDCW/h) are shown for *C. autoethanogenum* steady-state chemostat cultures grown on three
895 gas mixtures. See dashed inset for bar chart and heatmap details. Enzyme concentration and k_{app} data
896 are average of biological replicates. Proteins forming a complex are highlighted with non-black
897 borders (FdhA forms a complex with HytA–E for direct CO_2 reduction with H_2; CooS1 is expected to
898 form a complex with CooS1a and b as they are encoded from the same operon). For reactions with
899 isoenzymes, k_{app} is for the enzyme with the highest concentration ranking (top location on enzyme
900 heatmap), see Methods for details. Flux data from ref.[18] are average of biological replicates and error
901 bars denote standard deviation. Arrows show direction of calculated fluxes; red arrow denotes uptake
902 or secretion. Gene/protein IDs right of enzyme concentration heatmaps are preceded with
903 CAETHG_RS and red font denotes concentrations determined using stable-isotope labelled (SIL)
904 protein spike-in standards (i.e., anchor proteins). Asterisk denotes data for redox-consuming CO_2
905 reduction to formate solely by FdhA without the use of H_2 during growth on CO. [a]Bifunctional
906 acetaldehyde/alcohol dehydrogenase (acetyl-CoA→ethanol); [b]Flux into PEP from OAA and pyruvate
907 is merged and k_{app} is for PEPCK. See Supplementary Table S3 for gene/protein ID, proposed name,
908 description, and label-free protein concentrations. See Table 1 for anchor protein concentrations. See
909 Supplementary Table S5 for k_{app} and flux data. See ref.[18] for cofactors of reactions and metabolite
910 abbreviations. *gDCW*, gram of dry cell weight, *NQ*, not quantified.
911

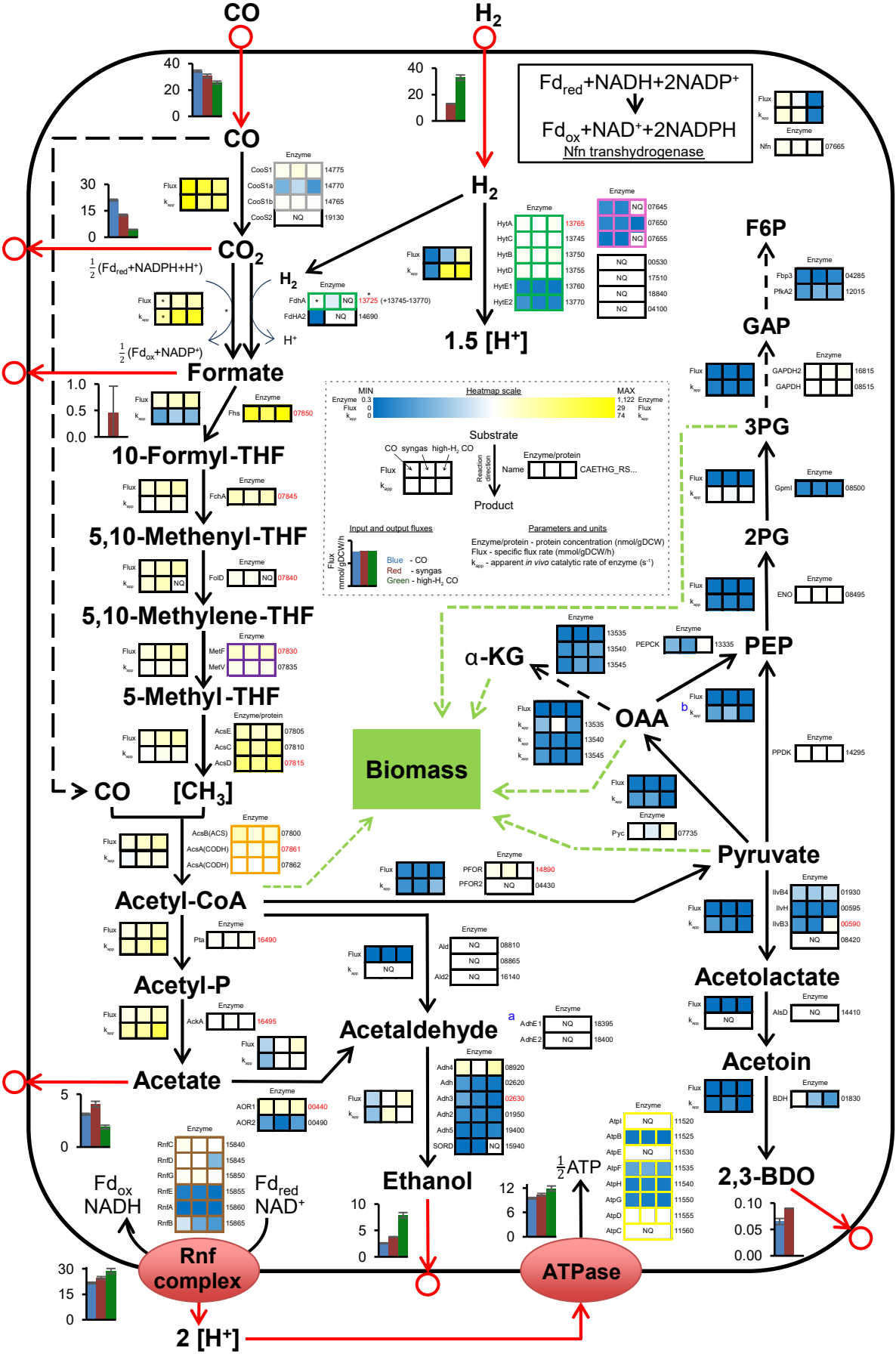
Fig. 6 Regulatory principles of apparent in vivo catalytic rates of enzymes (k_{app}) and metabolic flux throughput. a Enzymes catalysing higher metabolic flux rates have both higher concentrations and higher k_{app}s. Yellow and blue denote high and low values, respectively. Kendall's τ correlations with significance p-values between respective pairs are shown below heatmap. See Supplementary Table S5 for flux rate, enzyme concentration, and k_{app} data, and for description of reaction names (Rxn name) and gene-protein-reaction (GPR) associations. **b** Control of metabolic flux throughput through k_{app} changes for high flux pathways. See also Fig. 5. *gDCW*, gram of dry cell weight.

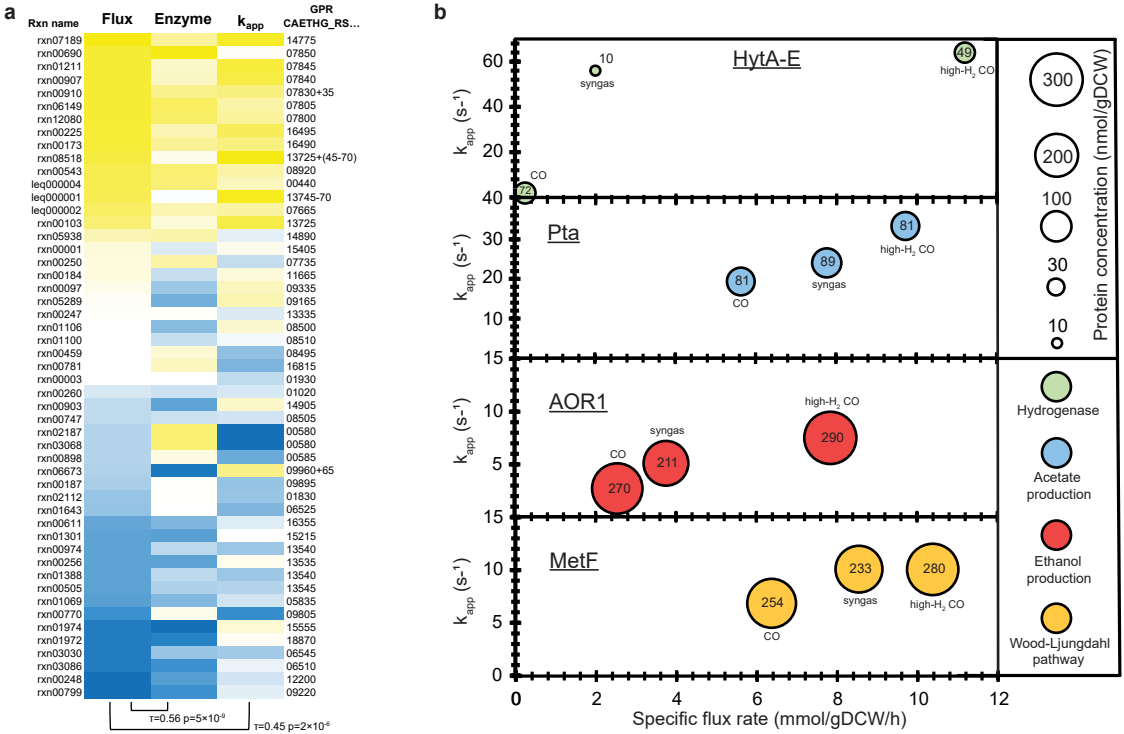

CO
H2
CO2
19 heavy SIL-proteins of
C. autoethanogenum
Light and heavy
peptides
DIA-MS
Thermo Q-Exactive HF-X
Data analysis
Skyline & mProphet
Absolute quantification
ADLNNENLEALEK
MS intensity
Heavy KR
Light KR
Retention Time
SILP-based concentrations
for 16 anchor proteins
Anchor concentration
MS intensity
Light C.autoethanogenum
chemostat cultures
Light cell lysate
DIA spectral library
Light cell lysates & fractions
heavy SIL-proteins
Label-free concentrations
for ~1,043 proteins


a
Correlation of peptide features
CO
Syngas
High-H2 CO
CO
Syngas
High-H2 CO
1
0.97
0.94
0.91
Pearson correlation (R)
b
Anchor protein calibration
y = 0.94x - 6.15
R² = 0.96
Log10 anchor protein concentration (nmol/gDCW)
Log10 aLFQ MS intensity
d
Label-free quantification accuracy
Sum of protein
CV-MFE
Sum of label-free protein concentrations (μg)
CV-MFE
CO
Syngas
High-H2 CO
c
aLFQ optimal model determination
Fragment ions
top1 top2 top3 top4 top5 all iBAQ
Peptides
CV-MFE


Ribosome (30S/50S)
MetFV
HytA-FdhA
IscS-IscU
Anthranilate synthase
Glutamate synthase
SecD-SecF
CarA-CarB
3-isopropylmalate dehydratase
InfB-InfC
FtsEX ABC cell division
GyrA-GyrB
Citrate lyase
1.50x
1x
0.67x
Protein concentration subunit 2 (nmol/gDCW)
Protein concentration subunit 1 (nmol/gDCW)
0
10⁰
10¹
10²
10³


Metabolism of cofactors and vitamins
Protein families
Carbohydrate metabolism (7%)
C1 fixation (30%)
Energy metabolism (13%)
Nucleotide metabolism
Lipid metabolism
Amino acid metabolism (8%)
No KO ID (11%)
Poorly characterised
Translation (11%)
Folding, sorting and degradation
Unclassified
Replication and repair
Transcription
Membrane transport
Cell motility
GcvH
FolD
MetF
LpdA
PdxS
folP
MoeA2
CoaBC
HytA
Pta
ClpP2
GAPDH2
PGK
TPI
PfkA2
GAPDH
ENO
BDH
PpdK
Pyc
PFOR
CooS1
MetV
AcsA (CODH)
FchA
HytD
HytB
HytC
RnfD
AOR1
AckA
AcsD
CooC
AcsB(ACS)
Fdh
AcsE
AcsC
Nfn
PpaC
Adh4
RnfC
AtpD
RnfB
RnfG
Fhs
CooS1b
dapA
hisIE
IlvC
Upp
Pdp
Adk
GuaA
PurA
PurC
PRPS
PurH
IMPDH
AcpP
AroB
TrpB
MetY
TrpD
Asd
AROA2
TrpG
IlvD
aspB
aspB2
glnA
proA
RS08145
RS00315
RS19150
RS02580
RS03060
RS18705
RS12745
RS08610
RS12590
SpoVS
Tig
FprB
CspA3
IscU
FprA1_2
Efp
Tsf
FusA
RplM
RplK
RpsO
RpsM
RplL
rplW
DnaK
TrxA
RpmD
RpsP
RpsE
RplF
RpsC
Tuf
RplO
RpsD
RpsK
GroEL
RpsB
RpsG
RpsH
RpsJ
rplC
RplP
RpsI
rplB
GroES
AbrB
MreB2
FtsZ
RplS
RpsU
RplU
HupB
RplT
NARS
SARS
RpoA
LacI
PtsH
cheA2
CheY3
luxS


CO
H2
CO
H2
CO2
Formate
10-Formyl-THF
5,10-Methenyl-THF
5,10-Methylene-THF
5-Methyl-THF
CO
[CH3]
Acetyl-CoA
Acetyl-P
Acetate
Acetaldehyde
Ethanol
Fdred+NADH+2NADP+
Fdox+NAD+2NADPH
Nfn transhydrogenase
1.5 [H+]
1/2 (Fdred+NADPH+H+)
1/2 (Fdox+NADP+)
H+
F6P
GAP
3PG
2PG
PEP
Pyruvate
α-KG
OAA
Biomass
Acetolactate
Acetoin
2,3-BDO
Fdox NADH
Fdred NAD+
Rnf complex
ATPase
1/2 ATP
2 [H+]
Heatmap scale
MIN
MAX
Substrate
CO syngas high-H2 CO
Product
Enzyme/protein
Input and output fluxes
Blue - CO
Red - syngas
Green - high-H2 CO
Parameters and units
Enzyme/protein - protein concentration (nmol/gDCW)
Flux - specific flux rate (mmol/gDCW/h)
kapp - apparent in vivo catalytic rate of enzyme (s-1)


a
Rxn name
Flux
Enzyme
k_{app}
GPR
CAETHG_RS...
rxn07189 14775
rxn00690 07850
rxn01211 07845
rxn00907 07840
rxn00910 07830+35
rxn06149 07805
rxn12080 07800
rxn00225 16495
rxn00173 16490
rxn08518 13725+(45-70)
rxn00543 08920
leq000004 00440
leq000001 13745-70
leq000002 07665
rxn00103 13725
rxn05938 14890
rxn00001 15405
rxn00250 07735
rxn00184 11665
rxn00097 09335
rxn05289 09165
rxn00247 13335
rxn01106 08500
rxn01100 08510
rxn00459 08495
rxn00781 16815
rxn00003 01930
rxn00260 01020
rxn00903 14905
rxn00747 08505
rxn02187 00580
rxn03068 00580
rxn00898 00585
rxn06673 09960+65
rxn00187 09895
rxn02112 01830
rxn01643 06525
rxn00611 16355
rxn01301 15215
rxn00974 13540
rxn00256 13535
rxn01388 13540
rxn00505 13545
rxn01069 05835
rxn00770 09805
rxn01974 15555
rxn01972 18870
rxn03030 06545
rxn03086 06510
rxn00248 12200
rxn00799 09220
τ=0.56 p=5×10-9
τ=0.45 p=2×10-6
b
HytA-E
Pta
AOR1
MetF
k_{app} (s^{-1})
Specific flux rate (mmol/gDCW/h)
10
syngas
49
high-H_2 CO
CO
72
81
high-H_2 CO
89
syngas
81
CO
high-H_2 CO
290
syngas
211
CO
270
233
syngas
280
high-H_2 CO
254
CO
Protein concentration (nmol/gDCW)
300
200
100
30
10
Hydrogenase
Acetate production
Ethanol production
Wood-Ljungdahl pathway