RapidRIP quantifies the intracellular metabolome of 7 industrial strains of *E. coli*

Douglas McCloskey\textsuperscript{a}, Julia Xu\textsuperscript{b}, Lars Schübbers\textsuperscript{a}, Hanne B. Christensen\textsuperscript{a}, Markus J. Herrgård\textsuperscript{b, a}

\textsuperscript{a} Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark
\textsuperscript{b} Department of Bioengineering, University of California - San Diego, La Jolla, CA 92093, USA

**A B S T R A C T**

Fast metabolite quantification methods are required for high throughput screening of microbial strains obtained by combinatorial or evolutionary engineering approaches. In this study, a rapid RIP-LC-MS/MS (RapidRIP) method for high-throughput quantitative metabolomics was developed and validated that was capable of quantifying 102 metabolites from central, amino acid, energy, nucleotide, and cofactor metabolism in less than 5 minutes. The method was shown to have comparable sensitivity and resolving capability as compared to a full length RIP-LC-MS/MS method (FullRIP). The RapidRIP method was used to quantify the metabolome of seven industrial strains of *E. coli* revealing significant differences in glycolytic, pentose phosphate, TCA cycle, amino acid, and energy and cofactor metabolites were found. These differences translated to statistically and biologically significant differences in thermodynamics of biochemical reactions between strains that could have implications when choosing a host for bioprocessing.

**1. Introduction**

Traditional high-throughput screening approaches in metabolic engineering have primarily relied upon basic physiology markers of strain performance. Physiological markers often include basic input/output parameters such as growth rate, substrate uptake rate, product excretion rate, yield, and productivity. While useful, these markers give little information on the underlying strain physiology. Omics data types (e.g., transcriptomics, metabolomics, fluxomics, etc.) could potentially provide a richer and deeper understanding of strain performance, which would allow more informed engineering decisions (McCloskey et al., 2013). However, -omics data types are often not utilized in routine screening of new production strain candidates because of high costs and low throughput (Hansen et al., 2017). Of the various -omics data types available, metabolomics provides the greatest potential to gain rich and deep insight on strain physiology at a lower cost and low throughput (Fuhrer et al., 2011; Guder et al., 2017; Link et al., 2015).

Metabolomics methods used for the absolute quantification of intracellular metabolites are often on the order of 30 min (or 48 samples per day) (Bennette et al., 2011; Buescher et al., 2010; McCloskey et al., 2016a, 2015). Longer run-times are often required for complete chromatographic separation of biologically important isomers that can not be resolved by differences in MS fragmentation. Longer run-times are also needed to allow for enough instrument scan time to acquire enough points across each detected metabolite peak in order to accurately and reproducibly measure a high number of transitions in a given run (McCloskey et al., 2016a). With the advent of shorter columns with decreased particle sizes and mass spectrometers with faster scan rates, the potential to reduce method run-times to less than 5 min with minimal compromise to chromatographic resolution or quality of acquired peaks is now possible (Guder et al., 2017). However, current fast separation methods often compromise on the ability to resolve important biological isomers (e.g., glucose 6 phosphate and glucose 1 phosphate). In addition, fast separation methods often compromise on the separation of structurally similar compounds with interfering signals (e.g., AMP, ADP, and ATP) that may compromise accurate quantitation. The ability to separate these species is critical for quantitative modeling of biological systems (Almquist et al., 2014; Henry et al., 2006; Jamshidi and Palsson, 2008; Miskovic et al., 2017; Saa and Nielsen, 2016). An ultra high-throughput metabolomics method less than 5 min that is able to resolve of biologically important isomers and provide complete separation of structural similar compounds with interfering transitions has yet to be demonstrated.

*Escherichia coli* are often used as a model prokaryote for genomic and physiological studies (Archer et al., 2011; Arifin et al., 2014; Ishii et al., 2007; Vijayendran et al., 2007; Yoon et al., 2012), and as host strains for industrial bioprocesses (Chae et al., 2017; S. Y. Choi et al.,...
owing to the rich toolset available for gene and expression modification (Datsenko and Wanner, 2000; Jiang et al., 2015; Ronda et al., 2016; Wang et al., 2009) as well as accumulated historical knowledge base of biochemical understanding (Guo et al., 2013; Kesheler et al., 2013; Keseler et al., 2013; Monk et al., 2017). Seven commonly used strains in biotechnology applications include K-12 MG1655, K-12 W3110, K-12 DH5α, BL21, C, Crooks, and W. It has been shown that particular strains are better suited for different applications (Monk et al., 2016). For example, DH5α is often preferred in cloning applications (Song et al., 2015; Taylor et al., 1993), while BL21 is often preferred for expression of recombinant proteins (Marisch et al., 2013a; Robichon et al., 2011). It has also been shown that choice of host strain for production of a given compound can significantly impact production titers (Na et al., 2013). Various comparisons between the seven industrial strains of E. coli under bioprocess conditions or as host strains for a particular metabolic engineering endeavor have been described (Chae et al., 2010; Fathi-Roudsari et al., 2016; Marisch et al., 2013a, 2013b; Monk et al., 2016). However, a comprehensive comparison of the metabolomes between the strains does not yet exist. An understanding of the starting metabolic levels of the strains would give researchers a key piece of information when deciding which E. coli strains to use as a host in their particular bioprocess.

In this study, a rapid reverse phase ion-pairing liquid chromatography triple quadrupole mass spectrometry (RapidRIP) for the rapid quantification of the intracellular metabolome for routine screening applications was developed and validated. It was shown that the method was capable of quantifying 102 metabolites in less than 5 min with comparable resolution, sensitivity, and reproducibility to a 33 min RIF-LC-MS/MS (FullRIP) method. The RapidRIP method was used to quantify the metabolome of seven industrial strains of E. coli. Significant differences in glycolytic, pentose phosphate, TCA cycle, amino acid, and energy cofactor metabolites were found. These differences translated to strain-specific differences in reaction thermodynamics.

2. Material and methods

2.1. Biological material and growth conditions

Escherichia coli strains E.coli C (DSMZ 4860), E. coli Crooks (DSMZ 1576), E. coli DH5α (DSMZ 6897) E. coli W (DSMZ 1116), E. coli W3110 (DSMZ 5911) were obtained from DSMZ-German Collection of Microorganism and Cell Cultures; E. coli BL21 (DE3) was purchased as competent cells from Agilent (Agilent Technologies), E. coli K-12 MG1655 (ATCC 700926). All cultures were grown in 25 mL of unlabeled or labeled glucose M9 minimal media (Sambrook and Russell, 2001) with trace elements (Fong et al., 2005) and sampled from a heat block in 50 mL autoclaved tubes that were maintained at 37 °C and aerated using magnets.

Growth and sampling procedures for Pseudomonas, Mouse, and CHO are described in the Supplemental methods.

2.2. Materials and reagents

Uniformly labeled 13C glucose was purchased from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA). Unlabeled glucose and other media components were purchased from Sigma-Aldrich (St. Louis, MO). LC-MS reagents were purchased from Honeywell Burdick & Jackson® (Muskegon, MI), Fisher Scientific (Pittsburgh, PA) and Sigma-Aldrich (St. Louis, MO).

2.3. LC-MS/MS instrumentation and data processing

Metabolites were acquired and quantified on an AB SCIEX Qtrap® 5500 mass spectrometer (AB SCIEX, Framingham, MA) using an ACQUITY UPLC HSS T3 Column (100 Å, 1.8 µm, 2.1 mm X 30 mm) and processed using MultiQuant® 3.0.1 as described previously (McCloskey et al., 2014a). The acquisition setting used for all transitions is given in Table S10.

2.4. Metabolomics

Internal standards were generated as described previously (McCloskey et al., 2014a, 2014b). All samples and calibrators were spiked with the same amount of internal standard taken from the same batch of internal standards. Calibration curves were run before and after all biological and analytical replicates. The consistency of quantification between calibration curves was checked by running a Quality Control sample that was composed of all biological replicates twice a day. Solvent blanks were injected every ninth sample to check for carryover. System suitability tests were injected daily to check instrument performance.

Metabolomics samples were acquired from triplicate cultures (1 mL of cell broth at an OD600 ~ 1.0) using a previously described method (Douglas McCloskey et al., 2014a, 2014b). A pooled sample of the filtered medium that was re-sampled using the FSF filtration technique and processed in the same way as the biological triplicates was used as an analytical blank. Extracts obtained from triplicate cultures and re-filtered medium were analyzed in duplicate. The intracellular values reported, unless otherwise noted, are derived from the average of the biological triplicates (n = 6). Metabolites in the pooled filtered medium with a concentration greater than 80% of that found in the triplicate samples were not analyzed due to high background interference. In addition, metabolites that were found to have a quantifiable variability (RSD > = 50%) in the Quality Control samples or any individual components with an RSD > = 80 were not used for analysis.

Missing values were imputed using a bootstrapping approach as coded in the R package Amelia II (Honaker et al., 2011) (version 1.7.4, 1000 imputations). Remaining missing values were approximated as ½ the lower limit of quantification for the metabolite normalized to the biomass of the sample. Prior to statistical analyses, metabolite concentrations were log transformed to generate an approximately normal distribution using the R package LMGene (Rocke et al., n.d.) (version 3.3, “multr = "TRUE", "lowessnorm" = "FALSE"). A Bonferroni-adjusted p-value cutoff of 0.01 as calculated from a Student’s t-test was used to determine significance between metabolic concentration levels. The glog-normalized values were used for downstream statistical analyses.

2.5. Thermodynamics

In vivo free energy change of reaction were calculated as described previously (McCloskey et al., 2014a, 2014b) using a recent genome-scale reconstruction of E. coli (Orth et al., 2011). In short, free energy of formation values taken from the eQuilibrator data base (Flamholz et al., 2012) were adjusted to physiological estimated values for the cytosolic, periplasmic, and extracellular space for temperature (37 °C), pH (7.5, 7.0, and 7.0, respectively), and ionic strength (0.2 M). The 95% confidence intervals of the metabolomics concentrations were used to calculate the 95% confidence intervals of the free energies of reaction. Free energy changes of reaction with non-overlapping confidence intervals between the strains were defined as statistically significant. Free energy changes of reaction with confidence intervals with different signs between the strains were defined as biologically significant. Free energy changes of reaction were only compared for those reactions with sufficient measured metabolomics and compound free energy of formation data coverage (50% and 99% of all reactants and products, respectively). Concentration values for phosphate, water, hydrogen, oxygen, and carbon dioxide were estimated as 1e-3, 55.0, 0.034e-3, 0.055e-3, and 1.4e-3M, respectively. Concentration confidence intervals for missing metabolites were estimated as 1.5e-3M and 1.5e-6M. Free energy of formation confidence intervals not measured were
estimated using the group contribution method (Flamholz et al., 2012).

Thermodynamic flux variability analysis (TFVA) was performed as previously described (Henry et al., 2007) using an open-source module for cobrapy called thermodynamics (https://github.com/dmccloskey/thermodynamics). Blocked reactions were those where the min and max flux was 0, essential reactions were those where the min and max flux were greater than 0, substitutable reactions were those where the min flux was 0 but the max flux was greater than 0, and constrained reactions were those where the min and max flux were the same.

3. Results and discussion

3.1. RapidRIP method development

A RapidRIP method was developed using the same mobile phase, temperature, and column chemistry to the FullRIP method (Fig. 1, Table 1). A shorter column with reduced particle size was used to maintain resolving power at a greatly reduced run-time (see Materials and Methods). A multitude of gradients, flow rates, and flow regimes were tested in order to optimize the method (Fig. S1). The optimized chromatographic parameters are given in the Material and Methods. Flow rate was modulated to allow for faster separation of later eluting compounds and faster column wash and equilibration, while maintaining the ability to retain early eluting compounds and separate hexose and pentose isomers. Percent mobile phase B was modulated to improve the separation of critical pairs, peak shape, and overall sensitivity (Fig. S1, Table S1). Interestingly, it was found that even within a greatly reduced method run time, substantial differences in separation between critical pairs could be obtained by careful tuning of the mobile phase gradient (Fig. S1, Table S1 and S2). Also, compound elution times
were evenly distributed and peak widths were minimized in order to maximize the number of transitions that could be analyzed in a given run.

### 3.2. RapidRIP method validation

The ability of the method to resolve biologically critical isomers was determined. Critical pairs included the hexose and pentose phosphates, mono-, di-, tri- adenine nucleotide phosphates, nad(p)(h), and citrate/isocitrate, ac(coa), and various organic acids and amino acids (Fig. S1, Tables S1 and S2). Baseline resolution (resolution greater than or equal to 1.0) was achieved for 75% of the representative critical pairs listed (9 of 12 pairs), while separation (retention time difference > 0.01 min) was achieved for all pairs. Importantly, compounds that were not baseline resolved have non-isotopic overlapping transitions, which allow them to be measured without isotopic interference from neighboring compounds.

Quantification accuracy and variation was assessed by measuring the average points across the peak, and the variation in peak height across multiple injects of a neat standard solution for five different acquisition methods (Table S3). The RapidRIP and FullRIP methods were found to have comparable linearity and sensitivity (Table S4). 54% of the compounds measured with RapidRIP had an LLOQ less than or equal to those measured with FullRIP, and 75% of the compounds measured with RapidRIP had an ULOQ greater than or equal to those measured with FullRIP. A decrease in sensitivity and LLOQ but increase in ULOQ was found primarily for early eluting metabolites consistent of amino acids, nucleosides and nucleotides. An increase in resolution and LLOQ was found for many later eluting metabolites. The increase in resolution can be explained by the decreased peak width using the RapidRIP method. The overall increase in ULOQ may be attributed to an overall increase in ion suppression due to the number of components eluting from the column in a given period of time. This is most noticeable for the early eluting compounds noted above where a loss in sensitivity was also found.

The RapidRIP and FullRIP methods were found to have comparable quantitative accuracy (Fig. 3A-G). A correlation coefficient (Pearson’s R) greater than or equal to 0.88 between glog normalized absolute metabolite concentrations (mmol*gDW-1) for all seven strains tested (Fig. 3A-G, Table S7) was found. Interestingly, the peak height ratio correlation coefficients between strains were much less than the absolute metabolite correlation coefficients (Table S7). This indicates that there was sufficient changes in analyte or internal standard peak heights from other components in the sample matrix that were well separated in the FullRIP method that are no longer well separated in the RapidRIP method to make the direct comparison between peak height ratios problematic. However, this also indicates that the use of a calibration curve was sufficient to compensate for the majority of these changes.

The RapidRIP and FullRIP methods were also found to have comparable quantitative precision (Fig. 3H). 61% and 58% of quantified metabolites for all strains had a %RSD less than or equal to 30 for RapidRIP and FullRIP methods, respectively. 7% and 4% of quantified metabolites for all strains had a %RSD greater than 60% RSD for RapidRIP and FullRIP, respectively. The distribution of %RSDs (Fig. 3H), and the percentages of metabolites below 30 and above 60%RSD between both methods indicates that the short method can retain the quantitative precision of the full method.
3.4. Quantification of the intracellular metabolome of seven industrial strains of E. coli

The RapidRIP method was used to quantify the intracellular metabolome of 7 industrial strains of E. coli (Fig. 4, Table S6, see Material and Methods for strains). Consistent with previous literature, the most abundant metabolites in the strains consisted of reduced glutathione, L-glutamate, and ATP (Bennett et al., 2009; Taymaz-Nikerel et al., 2011). Hierarchical clustering (Fig. 4A) reveals that similar to gene expression levels, the levels of intracellular metabolites do not reflect genomic distances between the strains (Monk et al., 2016). The hierarchical clustering is also consistent with how the strains group by Partial Least Squares Differential Analysis (PLS-DA, Fig. 4C). The primary mode of separation between the strains involve differences in glycolytic, pentose phosphate, and TCA cycle intermediates, and the secondary mode of separation include metabolites involved in nucleotide metabolism (Fig. 4D). These differences are discussed in greater detail below.

Physiological ratios are ratios of individual metabolites that often reflect broad physiological states (e.g., energy depletion, redox imbalance, etc.) that are tightly regulated by the cell. Physiological ratios,
including the nitrogen charge, energy charge, glutathione ratio, and redox ratio, differed between the strains (see methods for ratio definitions, Fig. 5A). *E. coli* C and DH5a had the lowest nitrogen and energy charges due to elevated levels of akg and amp, respectively. Non-significant differences in the glutathione and redox ratio, but significant differences in the metabolites that compose those ratios were found. The conservation of these ratios, but not necessarily the levels of the individual components may indicate the importance of the glutathione and redox ratios for maintaining normal aerobic physiology.

The intracellular concentration of 14 amino acids differed between the strains (Fig. 5B). *E. coli* DH5a was found to be a poor starting strain for L-tryptophan production. The intracellular concentrations of L-tryptophan were significantly lower than the other strains. *E. coli* MG1655, C, and W3110 were found to maintain significantly higher levels of D-aspartate. This is most likely due to utilization of the Phosphoenolpyruvate Carboxylase (PPC) in MG1655 and W3110 (McCloskey et al., 2016a, 2016b), and overall increased levels of TCA cycle intermediates in strain C. *E. coli* MG1655 and W3110 were also found to have the highest levels of L-arginine. It should be noted that when compared to the transcript levels or flux predictions of the amino acid producing pathways (Monk et al., 2016), little correlation between the absolute metabolite concentrations was found, which indicates the difficulty in predicting metabolite levels through indirect evidence as is consistent with previous works (Daran-Lapujade et al., 2007; Hackett et al., 2016).

Levels of central metabolism intermediates differed vastly between the strains (Fig. 6). *E. coli* MG1655, W, and W3110 were found to maintain significantly lower levels of glycolytic intermediates than other strains (i.e., g6p, f6p, fdp, 13dpg/23dpg, 2 pg/3 pg, and pep). However, the levels of pep in W3110 were found to be significantly higher than all other strains. *E. coli* MG1655 had the highest levels of ribulose and ribose 5-phosphate (rup5-D and r5p) compared to the other strains. This most likely contributed to the higher levels of L-histidine, a downstream product of r5p, in *E. coli* MG1655.

Based on the data collected here, *E. coli* C would serve as a good candidate for aerobic succinate and malate production due to high endogenous levels of succinate. *E. coli* C had significantly higher DL-lactate, succinate, malate, and fumarate concentrations compared to the other six strains. In addition, the levels of TCA cycle intermediates citrate and 2-oxoglutarate were among the highest. *E. coli* C has indeed previously been successfully used as the base strain for succinate production from glucose under aerobic and anaerobic conditions (Balzer et al., 2013; Jantama et al., 2008; Lin et al., 2005a, 2005b; Sánchez et al., 2005). Interestingly, *E. coli* DH5a had the highest levels of citrate, aconitate, isocitrate, and 2-oxoglutarate. This could indicate altered regulation or utilization of the TCA cycle in *E. coli* DH5a.

### 3.5. Thermodynamic analysis of the seven industrial strains

Change in free energies of reaction (dGr) for each of the industrial
Fig. 5. Physiological ratios (A) and intracellular amino acid levels (B). Selected ratios and metabolite levels that significantly changed are highlighted in the inset. The nitrogen charge (nc) is defined as (gln-L + 0.5*glu-L)* (gln-L + glu-L + akg)-1. The energy charge is defined as (atp + 0.5*adp)*(atp + adp + amp)-1. The glutathione ratio (Gth) is defined as (gthrd*gthox)-1. The redox ratio is defined as (nad(p)+nad(p)+nad(p)+nad(p)-1. Units are the log2 fold change of the median value of each strain to the mean of all 7 strains.

Fig. 6. Absolute metabolite concentrations of central metabolism intermediates for the 7 industrial strains of E. coli. A) Schematic of central carbohydrate metabolism. B) Heatmap of metabolite levels. Units are the log2 fold change of the median value of each strain to the mean of all 7 strains. Selected metabolites that changed significantly are shown in the inset.
strains were calculated and compared in order to identify statistically (i.e., non-overlapping confidence intervals) and biologically (i.e., a change in ΔG sign indicating a change in reaction directionality) significant differences in reaction thermodynamics between the strains (Table S8). Of the 475 reactions with satisfactory measured metabolomics and free energy of formation coverage (see Methods), between 22 and 58 statistically different reactions were found between the strains, between 4 and 17 biologically different reactions were found between the strains, and between 0 and 4 statistically and biologically different reactions were found between the strains (Fig. 7, Table S9).

Thermodynamics analysis revealed differences in the thermodynamic potential to push carbon towards glycogen biosynthesis instead of more economically valuable carbon endpoints between the strains. Biologically and statistically significant differences were found between glucose 1-phosphate adenyltransferase (GLGC) in E. coli C and all other strains except E. coli W and W3110 (Fig. 7). E. coli C had the highest levels of the glycogen precursor ADP-glucose (adpglc) in glucose 1-phosphate (GLGC). Note the high level of ADP-glucose (adpglc) in E. coli C. The change in reaction free energy differed for GLGC differed biologically and statistically between E. coli C and all other strains except E. coli W and W3110.

Fig. 7. Thermodynamic analysis of the 7 industrial strains of E. coli. A) Heatmap of pairwise counts of reaction that differed either statistically, biologically, or both between the strains. B) Schematic of the oxidative pentose phosphate pathway (oxPPP) with spontaneous and unknown (?) reactions shown for E. coli BL21. Note the high levels of oxPPP intermediates in E. coli BL21. C) Schematic of the first step in glycogen biosynthesis encoded by glucose 1-phosphate adenylyltransferase (GLGC). Note the high level of ADP-glucose (adpglc) in E. coli C. The change in reaction free energy differed for GLGC differed biologically and statistically between E. coli C and all other strains except E. coli W and W3110.

In order to demonstrate the applicability of RapidRIP for characterizing strain differences, od was used to quantify the metabolome of 7 industrial strains of E. coli during aerobic growth on glucose. Major differences in central, amino acid, nucleotide, energy, and redox metabolism metabolite levels were found. These differences translated to statistically significant differences in reaction thermodynamics; several of which were also found to be biologically significant through change
in reaction directionality. Knowledge of these differences provides researchers with valuable information when choosing which strain to use as host for a particular chemical production process. For example, researchers should not be dissuaded from using the BL21 strain based on theoretical predictions of compound yield derived from genomic information alone because the thermodynamics of the PPP in the absence of the pdh gene is such that flux and levels of NADPH are not significantly altered.

Acknowledgements

We would like to thank the Novo Nordisk Foundation for funding this work through the Novo Nordisk Foundation Center for Biosustainability. This work has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 686070.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jymem.2018.04.009.

References


McCloskey, D., Young, J.D., Xu, S., Palsson, B.O., Feist, A.M., 2016a. MID max LC–MS/MS method for measuring the precursor and product mass isotope distributions of