

1 **Isotopomer distributions in amino acids from a highly expressed protein as a proxy for**
2 **those from total protein**

3

4 Afshan S. Shaikh^{†1}, Yinjie J. Tang^{†1,2,3}, Aindrila Mukhopadhyay^{2,3}, Jay D. Keasling^{1,2,3,4*}.

5

6 Department of Chemical Engineering, University of California, Berkeley, USA¹, Physical
7 Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, USA², Virtual Institute
8 of Microbial Stress and Survival³, Department of Bioengineering, University of California,
9 Berkeley, USA⁴

10

11

12 *Corresponding author

13 Prof. Jay D. Keasling

14 Berkeley Center for Synthetic Biology,

15 717 Potter Street, Berkeley, CA, 94720 USA

16 Email: keasling@berkeley.edu

17 Phone: 510-495-2620, Fax: 510-495-2630

18 †: These authors contributed equally in this study

19 **ABSTRACT**

20 ¹³C-based metabolic flux analysis provides valuable information about bacterial
21 physiology. Though many biological processes rely on the synergistic functions of
22 microbial communities, study of individual organisms in a mixed culture using existing
23 flux analysis methods is difficult. Isotopomer-based flux analysis typically relies on
24 hydrolyzed amino acids from a homogenous biomass. Thus metabolic flux analysis of a
25 given organism in a mixed culture requires its separation from the mixed culture. Swift
26 and efficient cell separation is difficult and a major hurdle for isotopomer-based flux
27 analysis of mixed cultures. Here we demonstrate the use of a single highly-expressed
28 protein to analyze the isotopomer distribution of amino acids from one organism. Using
29 the model organism *E. coli* expressing a plasmid-borne, his-tagged Green Fluorescent
30 Protein (GFP), we show that induction of GFP does not affect *E. coli* growth kinetics or
31 the isotopomer distribution in nine key metabolites. Further, the isotopomer labeling
32 patterns of amino acids derived from purified GFP and total cell protein are
33 indistinguishable, indicating that amino acids from a purified protein can be used to infer
34 metabolic fluxes of targeted organisms in a mixed culture. This study provides the
35 foundation to extend isotopomer-based flux analysis to study metabolism of individual
36 strains in microbial communities.

37

38

39 Key words: ¹³C based metabolic flux analysis, *E. coli*, his-tagged GFP, central metabolic
40 pathways, microbial communities.

41

42 INTRODUCTION

43 Metabolic flux analysis allows insight into a cell's overall carbon metabolism,
44 energy production, and the relationship of the genotype to its phenotype¹. Development
45 of typical flux balance models requires assumption of an objective function (often
46 maximizing the specific growth rate) and a series of physicochemical constraints
47 (thermodynamic directionality, enzymatic capacity, and reaction stoichiometry)². For a
48 complicated metabolic network, there are not always enough constraints to find a unique
49 solution. Thus, such an approach alone may not accurately describe the reversible
50 reactions or reactions that may form a futile cycle, especially when the cell's metabolism
51 is suboptimal or the annotated pathways are latent under experimental conditions^{3, 4}. For
52 determination of actual carbon fluxes, ¹³C based flux analysis is an advanced approach
53 whereby a ¹³C-labeled carbon source, such as glucose, is fed to the cells, and the labeling
54 pattern in resulting metabolites is measured by NMR or GC-MS⁵⁻⁷. Amino acids arise
55 from central metabolic pathways and reflect the carbon backbones of their precursors
56 (Figure 1). The labeling pattern in amino acids provides information about the ¹³C
57 distribution in their corresponding precursor metabolites, and thus allows the unique
58 determination of carbon fluxes in branched pathways or metabolic cycles. In a typical
59 isotopomer based flux analysis assay, labeled biomass provides the total protein, which is
60 hydrolyzed to yield the amino acid analyte. Recent advances in isotopomer-based flux
61 analysis have been applied to numerous microorganisms^{3, 4, 8-10}. Results from such studies
62 not only quantified fluxes through central metabolic pathways but also provided
63 experimental evidence for predicted pathways in recently sequenced genomes^{8, 11, 12}.
64 Since analytes comprise of amino acids derived from total biomass of a culture, this

65 strategy is only applicable to defined monocultures; little is known about the cellular
66 metabolic network of individual organisms in complicated mixed cultures.

67 Most naturally-occurring biological processes, ranging from the degradation of
68 complex organics to elemental cycling, involve the synergistic action of multiple
69 microorganisms¹³⁻²¹. Understanding how multiple organisms interact and function in a
70 community is essential to improve their use in bioremediation of contaminated
71 environments and production of renewable energy²²⁻²⁴. However, it is difficult to monitor
72 metabolic fluxes of individual organisms in a community because it is not possible to
73 assign amino acid isotopomer data to the source organism when the sample is derived
74 from a hydrolysate of the total culture biomass. Traditional separation techniques such as
75 cell sorting or gradient centrifugation are not efficient for enriching cell types for
76 isotopomer analysis. Currently, flux analysis for microbial communities is limited to the
77 study of the entire mixed culture to provide an overall view of the carbon metabolism²⁵.
78 Here we describe experiments to test the assumption that an enriched pure protein from
79 an organism can provide the same information as that from total cell protein. While the
80 origin of an amino acid is ambiguous and cannot be assigned to a sub-population in a
81 mixed culture, the origins of a protein on the other hand can be fully assigned. As a
82 result, the use of a single protein for isotopomer-based flux analysis would allow the
83 study of a microbe in its community.

84

85

86

87

88 **EXPERIMENTAL SECTION**

89 *E. coli* growth and metabolite assays for isotopomer studies. Difco M9 minimal
90 salts (5X) were purchased from Becton, Dickinson and Company (Franklin Lakes, NJ).
91 Unlabeled glucose was obtained from Fisher Scientific (Fair Lawn, NJ) and labeled
92 glucose (1-¹³C, 99%) from Cambridge Isotope Laboratories (Andover, MA). Isopropyl β-
93 D-1-thiogalactopyranoside was procured from Sigma-Aldrich (St. Louis, MO). Chelating
94 sepharose fast flow resin was purchased from GE Healthcare Bio-Sciences (Piscataway,
95 NJ).

96 The gene encoding green fluorescence protein (*gfp*) was cloned in the *NcoI* and
97 *EcoRI* restriction sites in a pET-30 vector (Novagen, Madison, WI) to create the
98 expression vector pET30-GFP. *E. coli* DH10B was used for cloning and plasmid
99 maintenance while *E. coli* BLR(DE3) was used as the host for protein production. Total
100 protein and purified GFP in this study were isolated from the wild type (*E. coli*
101 BLR(DE3)) or the GFP expression strain (*E. coli* BLR(DE3) pET30-GFP). All cultures
102 were grown in M9 minimal medium supplemented with 1% glucose, 0.01 mM FeSO₄,
103 and a trace metal solution²⁶. Shake flasks (250 mL) were used to grow these cultures at
104 37°C with shaking at 200 rpm. Kanamycin (50 µg/mL) was used for strains containing
105 pET30-GFP.

106 To avoid excessive lag time during adaptation to minimal medium, cultures were
107 subjected to the following procedure. A single colony was used to inoculate 10 mL of
108 Luria Broth and this culture was grown overnight. An aliquot (500 µL) of this culture
109 was used to inoculate 20 mL of M9 minimal medium, and this culture was incubated
110 overnight at 37°C. If the optical density at a wavelength of 600 nm (OD₆₀₀) of this culture

111 was greater than 1.0/mL, the cells were subcultured (2%) into fresh M9 minimal medium
112 and allowed to grow for 24 hours. This procedure was repeated until the OD₆₀₀ of
113 overnight cultures was greater than 3.5/mL for at least two subcultures to adapt cells to
114 defined medium. One-mL aliquots were stored at -80°C for future use. A thawed aliquot
115 was used to inoculate overnight cultures in M9 minimal medium, and this culture (0.1%)
116 was used to inoculate 50 mL of M9 minimal medium for subsequent studies. When the
117 OD₆₀₀ of the cultures containing the plasmid reached 0.7/mL, IPTG was added to a final
118 concentration of 1 mM to induce **production of** recombinant his-tagged GFP. Cells were
119 incubated at 37°C for three hours and cell pellets were harvested by centrifugation
120 **(10,000 × g for 20 minutes)**. Cell growth was monitored by measuring the absorbance at
121 OD₆₀₀. The concentrations of glucose and acetate in the culture medium were measured
122 using enzyme kits **as per the manufacturer's instructions** (r-Biopharm, Darmstadt,
123 Germany).

124

125 ***Protein purification.*** The cell pellets were resuspended in ice-cold, 50 mM
126 HEPES buffer, pH 8. The cells were lysed by sonication and the total soluble protein was
127 separated from the insoluble debris by centrifugation (10,000 × g for 30 minutes). The
128 his-tagged recombinant GFP was purified from the total soluble protein using
129 immobilized nickel ion affinity chromatography. Briefly, 0.5 mL resin was used in each
130 case and the resin was prepared according to the manufacturer's instructions (GE
131 Healthcare, Piscataway, NJ). Total soluble protein was added to the resin. The resin was
132 washed with increasing concentrations of imidazole; the protein was eluted with 250 mM
133 imidazole. Protein samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel

134 electrophoresis (SDS-PAGE). To confirm the presence of tagged GFP, a Western blot
135 was carried out using the S-tag alkaline phosphatase Western blot kit as per the
136 manufacturer's instructions (Novagen, Madison, WI).

137

138 *Sample preparation for GC-MS.* One-mL purified protein (containing about 3 mg
139 his-tagged GFP) was precipitated using 0.3 mL 60 wt % trichloroacetic acid (TCA). After
140 adding TCA, samples were incubated at 4°C for 10 min and then centrifuged at 20,000 ×
141 g for 5 min at 4°C. The supernatant was removed, and the protein pellet was washed three
142 times using 1 mL deionized water followed by three washes with 0.5 mL cold acetone.
143 These washes minimized carry forward of any residual phosphate salts from the growth
144 medium and eliminate imidazole and other organic compounds that may interfere with
145 GC-MS signals. *To hydrolyze proteins to amino acids, the protein pellet was dried at*
146 *100°C for 2 min, resuspended in 1.5 mL 6 M HCl, and transferred to a clear glass, screw-*
147 *top GC vial. Vials were capped and incubated at 100°C for 24 hours. After hydrolysis,*
148 *the caps were removed and the samples were dried overnight under a stream of air.*

149 GC-MS samples were prepared as previously described^{8, 11}. Briefly, samples were
150 dissolved in 100 µL tetrahydrofuran (THF) and 100 µL N-(tert-butyldimethylsilyl)-N-
151 methyl-trifluoroacetamide (Sigma-Aldrich, St. Louis, MO). All samples were derivatized
152 in a water bath at 65-80°C for 1 hour, resulting in tert-butyldimethylsilyl (TBDMS)
153 derivatives. One µL of the derivatized sample was injected into a gas chromatograph
154 (Agilent model 6890, Wilmington, DE) equipped with a DB5-MS column (J&W
155 Scientific, Folsom, CA) and analyzed using *a quadrupole mass selective detector (EI)*
156 *operated at 70 eV (Agilent 5973, Wilmington, DE). The MS was operated in scan mode*

157 (started after 4 min, mass range 60-550 a.m.u. at 2.94 s/scan). The sample injection
158 volume was 1 μ L at a carrier gas flow of 2 mL/min helium with a split ratio of 1:20. The
159 GC operation conditions were as follows: the GC column was held at 150°C for 2 min,
160 heated at 3°C per minute to 280°C, heated at 20°C per minute to 300°C, and held for 5
161 min at that temperature. Two types of positively charged ions were clearly observed by
162 GC-MS: the derivatized amino acids after the loss of a tert-butyl group $[M-57]^+$, and the
163 ion after fragmentation at the α -carboxyl group, $[M-159]^+$. The natural abundance of
164 isotopes was corrected using a published algorithm before using the data for calculating
165 the label distribution¹².

166

167 **RESULTS AND DISCUSSION**

168 To test the hypothesis that a pure protein could be used to obtain the isotopomer
169 data to compute flux for the source organism, two important requirements had to be
170 fulfilled. First, in using a recombinant protein it was important to assess if production of
171 GFP affects *E. coli* central metabolism. Second, it was important to determine if the
172 isotopomer labeling pattern in amino acids from GFP reflected the labeling pattern in
173 amino acids from total protein.

174 The experimental strategy is shown in Figure 2: sample 1 is *E. coli* BLR(DE3)
175 cultured with un-labeled glucose; sample 2 is *E. coli* BLR(DE3) cultured with 99% $1-^{13}\text{C}$
176 glucose, samples 3, 4 and 5 are *E. coli* BLR(DE3) pET30-GFP cultured with 99% $1-^{13}\text{C}$
177 glucose and induced for his-tagged GFP production. For isotopomer analysis, amino
178 acids derived from either hydrolyzed total protein or hydrolyzed purified GFP were used.
179 Samples 2 and 3 were used for analyzing amino acids from total protein in un-induced

180 and induced cultures, respectively. Sample 4 was used for analyzing amino acids from
181 pure, his-tagged, GFP. Sample 6 was a 1:1 mixture of sample 1 (un-labeled wild-type
182 biomass) and sample 5 (labeled induced biomass) and represented an artificial mixed
183 culture from which his-tagged GFP was purified. To prove that protein induction does not
184 affect central metabolism, the isotopomer data obtained from samples 2 and 3 were
185 required to be comparable. To establish that a purified protein can be used for an
186 isotopomer-based flux analysis of the source organism, data from the his-tagged GFP
187 from sample 6 were required to be equivalent to the data from sample 2.

188 Under aerobic conditions, all cultures outlined above had a doubling time of ~1
189 hr. In order to determine the isotopomer patterns in amino acids, samples were taken at
190 mid-log phase ($OD_{600} \sim 2$), which represents a (quasi) steady state that is often used for
191 flux analysis²¹. At the time of sampling, glucose consumption and acetate production
192 were measured (Table 1). No significant difference was observed between the various
193 cultures, indicating that GFP production did not significantly affect *E. coli* BLR(DE3)
194 growth kinetics. Total soluble protein was obtained from the cells as described in the
195 Materials and Methods section. His-tagged GFP was obtained from total soluble protein
196 of samples 4 and 6. Total protein was loaded on Ni-charged resin, and bound GFP was
197 eluted with 250 mM imidazole. All the samples were analyzed by SDS-PAGE. The
198 elution fractions corresponding to samples 4 and 6 indicated that GFP was more than
199 99% pure (Figure 3A). A Western blot was carried out to confirm the presence of
200 recombinant GFP (Figure 3B).

201 GC-MS analysis is a high-throughput and sensitive ($\leq 2\%$ error) method widely
202 used for ¹³C isotopomer analysis. Application of GC-MS to resolve derivatized protein

203 hydrolysate gave chromatographic peaks corresponding to 16 amino acids (asparagine,
204 cysteine, glutamine, and tryptophan could not be determined, possibly due to sample
205 degradation) (Figure 4A). Several amino acid pairs derived from the same precursor, such
206 as proline and glutamate (from precursor oxoglutarate), threonine and aspartate (from
207 precursor oxaloacetate), tyrosine and phenylalanine (from precursors
208 phosphoenolpyruvate and erythrose-4-phosphate), had similar isotopomer distribution
209 patterns. Nine key amino acids, representing their precursor's labeling that are often used
210 for isotopomer analysis are listed in Table 2, and the remaining amino acids provide
211 redundant isotopomer information³. Two types of ions were used (Table 2 and Figure
212 4B): fragment [M-57]⁺ is the *tert*-butyldimethylsilyl (TBDMS)-derivatized amino acid
213 with a loss of the *tert*-butyl group; fragment [M-159]⁺ is the TBDMS-derivatized amino
214 acid with a loss of the COO-TBDMS portion due to a break between the α - and β -carbon
215 atoms of the amino acid¹².

216 The GC-MS spectra for key amino acids from total protein of wild-type and
217 induced *E. coli*, as well as purified GFP (samples 2, 3 and 4) had identical isotopomer
218 distributions (Table S-1). Isotopomer data of amino acids derived from GFP isolated
219 from the artificial mixed culture (sample 6) was also equivalent to data from the control
220 sample (sample 2) (Table 2 and Figure S-1). The difference in isotopomer distribution is
221 <2%, well within the measurement noise. These results indicate that this method can be
222 used to investigate the metabolic fluxes in an individual microorganism in a microbial
223 community using a purified protein. The same strategy may be extrapolated to any
224 protein that can be isolated in adequate levels and then used to quantify the isotopomer
225 distribution in the corresponding amino acids.

226 While the method outlined in this study is fairly universally applicable and uses
227 generally accessible reagents and instrumentation, certain precautions were necessary
228 prior to obtaining accurate spectra for isotopomer analysis. Media used for culturing
229 bacteria contains phosphate that may remain in the total protein samples. Residual
230 phosphate can be derivatized and its products interfere with other compounds of interest
231 during GC-MS separation²⁷. Moreover, reagents such as urea and imidazole that are part
232 of typical cell lysis and protein purification buffers may also be derivatized by TBDMS
233 and complicate the GC-MS data acquisition and spectral quality. Examples of urea,
234 imidazole and phosphate contaminated spectra are provided in Figure S-2. To overcome
235 this problem, TCA precipitation of proteins followed by extensive washing with water
236 and acetone proved effective. Another factor to be considered is the quantity of pure
237 protein required. Despite the sensitivity of GC-MS to detect low sample levels, due to the
238 multiple clean up steps, efficiency of protein hydrolysis and efficiency of derivatization,
239 approximately 1 mg of GFP was required to enable detection of all key amino acids
240 essential for flux analysis. Lower amounts of protein may cause the loss of signal of rarer
241 amino acids like methionine and histidine.

242

243 **CONCLUSIONS**

244 Flux analysis provides key information about cellular metabolism. This
245 information is central to many studies: physiological characterization of stress responses,
246 impact of metabolic engineering, to name a few. However, isotopomer-based flux
247 analysis methods cannot be currently applied to assess the core physiology of an
248 organism unless it is present in a mono-culture. In order to address this hurdle, we show

249 here that the isotopic labeling in key amino acids derived from purified over-expressed
250 protein in an organism serves as a proxy for total protein of that organism. As it is the
251 isotopomer data from these key amino acids that are used to determine flux through
252 central metabolic pathways, the flux distribution information of a target organism in a
253 mixed culture can be obtained from a single purified protein.

254

255 **ACKNOWLEDGEMENTS**

256 We thank Christopher Petzold (University of California, Berkeley) for reviewing
257 our manuscript and Yen Hsieh (Molecular and Cell biology Department, University of
258 California, Berkeley) for helping with the experiments. We also thank Wesley D. Marner
259 II (Chemical Engineering Department, University of California, Berkeley) and Sydnor T.
260 Withers (National Research Council of Canada) for their help in making the pET30-GFP
261 construct. This work is part of the Virtual Institute for Microbial Stress and Survival
262 (<http://vimss.lbl.gov>) supported by the U.S. Department of Energy, Office of Science,
263 Office of Biological and Environmental Research, Genomics:GTL Program through
264 contract DE-AC02-05CH11231 between the Lawrence Berkeley National Laboratory and
265 the US Department of Energy.

266 **REFERENCES**

- 267 (1) Stephanopoulos, G. N.; Aristidou, A. A.; Nielsen, J. *Metabolic Engineering*
268 *Principles and Methodologies*; Academic Press: San Diego, 1998.
- 269 (2) Edwards, J. S.; Palsson, B. O. *Proc. Natl. Acad. Sci.* **2000**, *97*, 5528–5533.
- 270 (3) Fischer, E.; Sauer, U. *Nat. Genet.* **2005**, *37*, 636-640.
- 271 (4) Fong, S. S.; Nanchen, A.; Palsson, B. O.; Sauer, U. *J. Biol. Chem.* **2006**, *281*,
272 8024-8033.
- 273 (5) Szyperski, T. *Q. Rev. Biophys.* **1998**, *31*, 41-88.
- 274 (6) Wiechert, W. *Metab. Eng.* **2001**, *3*, 195-206.
- 275 (7) Wittmann, C.; Heinzle, E. *Biotechnol. Bioeng.* **1999**, *62*, 739-750.
- 276 (8) Tang, Y. J.; Chakraborty, R.; Martin, H. G.; Chu, J.; Hazen, T. C.; Keasling, J. D.
277 *Appl. Environ. Microbiol.* **2007**, *73*, 3859-3864.
- 278 (9) Tang, Y. J.; Meadows, A. L.; Kirby, J.; Keasling, J. D. *J. Bacteriol.* **2007**, *189*,
279 894-901.
- 280 (10) Tang, Y. J.; Hwang, J. S.; Wemmer, D.; Keasling, J. D. *Appl. Environ. Microbiol.*
281 **2007**, *73*, 718-729.
- 282 (11) Tang, Y.; Pingitore, F.; Mukhopadhyay, A.; Phan, R.; Hazen, T. C.; Keasling, J.
283 D. *J. Bacteriol.* **2007**, *189*, 940-949.
- 284 (12) Wahl, S. A.; Dauner, M.; Wiechert, W. *Biotechnol. Bioeng.* **2004**, *85*, 259-268.
- 285 (13) Eun, J. S.; Fellner, V.; Burns, J. C.; Gumpertz, M. L. *J. Anim. Sci.* **2004**, *82*, 170-
286 178.
- 287 (14) Icgen, B.; Harrison, S. *Res. Microbiol.* **2006**, *157*, 922-927.

- 288 (15) Moons, P.; Van Houdt, R.; Aertsen, A.; Vanoirbeek, K.; Engelborghs, Y.;
289 Michiels, C. W. *Appl. Environ. Microbiol.* **2006**, *72*, 7294-7300.
- 290 (16) Puzyr, A. P.; Mogil'naia, O. A.; Gurevich Iu, L.; Babkina, E. A. *Mikrobiologiya*
291 **2001**, *70*, 98-105.
- 292 (17) Schrap, S. M.; van den Heuvel, H.; van der Meulen, J.; Ruiter, H.; Parsons, J. R.
293 *Chemosphere* **2000**, *40*, 1389-1397.
- 294 (18) Shah, S.; Thakur, I. S. *Curr. Microbiol.* **2003**, *47*, 65-70.
- 295 (19) Zhang, Y.; Frankenberger, W. T., Jr. *Sci. Total Environ.* **2005**, *346*, 280-285.
- 296 (20) Zwolinski; Harris, R. F.; Hickey, W. J. *Biodegradation* **2000**, *11*, 141-158.
- 297 (21) Sauer, U.; Lasko, D. R.; Fiaux, J.; Hochuli, M.; Glaser, R.; Szyperski, T.;
298 Wuthrich, K.; Bailey, J. E. *J. Bacteriol.* **1999**, *181*, 6679-6688.
- 299 (22) Gilbert, E. S.; Walker, A. W.; Keasling, J. D. *Appl. Microbiol. Biotechnol.* **2003**,
300 *61*, 77-81.
- 301 (23) Leschine, S. B.; Canale-Parola, E. *Curr. Microbiol.* **1984**, *VII*, 129-135.
- 302 (24) Fischer, E.; Sauer, U. *Eur. J. Biochem.* **2003**, *270*, 880-891.
- 303 (25) Stolyar, S.; Van Dien, S.; Hillesland, K. L.; Pinel, N.; Lie, T. J.; Leigh, J. A.;
304 Stahl, D. A. *Mol. Syst. Biol.* **2007**, *3*, 92.
- 305 (26) Neidhardt, F. C.; Bloch, P. L.; Smith, D. F. *J. Bacteriol.* **1974**, *119*, 736-747.
- 306 (27) Yon, C.; Han, J. S. *Exp.Mol. Med.* **2000**, *32*, 243-245.

307

308

309

310

311 **FIGURE CAPTIONS**

312 **Figure 1. Central metabolic and amino acid biosynthetic pathways.** E4P, erythrose-
313 4-phosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; PEP,
314 phosphoenolpyruvate; PGA, 3-phosphoglycerate; C5P, ribose-5-phosphate; T3P, Triose-
315 3-phosphate; and S7P, sedoheptulose 7-phosphate.

316

317 **Figure 2. Experimental strategy outlining the various samples used for proof of**
318 **concept isotopomer analyses.** Samples were numbered as shown. Sample 6 is a mixture
319 of un-labeled culture (sample 1) and labeled induced culture (sample 5).

320

321 **Figure 3. Total protein and pure recombinant GFP used in isotopomer analysis. (A)**
322 Coomassie stained SDS-PAGE analysis of protein samples. Lane 1 contains the ProSieve
323 color protein marker (Cambrex, Rockland, ME). Lanes 2 and 3 show total soluble
324 protein, while lanes 4 and 5 show pure, recombinant GFP from samples 4 and 6
325 respectively. **(B)** Anti S-tag Western blot to confirm the presence of GFP. Lane 1
326 contains the ProSieve color protein marker. Lanes 2 and 4 show total soluble protein
327 while lanes 3 and 5 show pure recombinant GFP obtained from samples 4 and 6,
328 respectively.

329

330 **Figure 4. Gas chromatograph-mass spectrum of amino acids from purified GFP**
331 **(sample 6). (A)** 16 of the 20 amino acids can be clearly observed in the spectrogram
332 (glutamine, asparagine, tryptophan, and cysteine were not observed). Standard
333 abbreviations of amino acids are used to label the corresponding peaks. **(B)** Mass

334 spectrum of a TBDMS-derivatized amino acid. An amino acid is usually derivatized at
335 both the carboxylic acid and amino group. The spectrum shown is for glutamic acid,
336 where ions with $m/z = 432 - 437$ are the isotopomers of $[M-57]^+$ fragments and the ions
337 with $m/z = 330 - 334$ are the isotopomers of $[M-159]^+$ fragments.

Table 1. Growth kinetics of *E. coli* after IPTG induction for GFP production.

	Sample 2	Samples 3, 4 and 5
Glucose remaining, g/L	7.4 ± 0.3	7.6 ± 0.1
OD ₆₀₀	~2	~2
Acetate, g/L	0.15 ± 0.02	0.17 ± 0.05
Doubling time	~1 hr	~1 hr

Table 2. Comparison of GC-MS data, total protein and his-tagged GFP.

Amino acid	Precursor/Pathway	Ion ^a	Total protein from BLR(DE3) (sample 2)		Purified GFP (sample 6)	
			[-57] ^a	[-159]	[-57]	[-159]
Alanine	Pyruvate	M0	0.56±0.01	0.56±0.02	0.55±0.02	0.55±0.01
		M1	0.43±0.02	0.41±0.01	0.44±0.01	0.42±0.02
		M2	0.01±0.0	0.03±0.01	0.01±0.0	0.03±0.01
Glycine	Serine	M0	0.97±0.03	0.99±0.01	0.99±0.0	1.0±0.0
		M1	0.02±0.01	0.01±0.0	0.01±0.0	0.0±0.0
Valine	Pyruvate	M0	0.34±0.01	0.33±0.02	0.33±0.01	0.32±0.02
		M1	0.48±0.01	0.47±0.02	0.48±0.01	0.48±0.01
		M2	0.17±0.01	0.19±0.01	0.18±0.01	0.19±0.01
Leucine	Pyruvate & Acetyl-CoA	M0	Peak overlap	0.20±0.02	Peak overlap	0.19±0.02
		M1		0.41±0.02		0.41±0.02
		M2		0.30±0.01		0.31±0.03
Isoleucine	Pyruvate & oxaloacetate	M0	Peak overlap	0.23±0.02	Peak overlap	0.22±0.01
		M1		0.42±0.02		0.43±0.0
		M2		0.27±0.01		0.27±0.0
Serine	3-P-Glycerate	M0	0.57±0.03	0.61±0.02	0.56±0.02	0.60±0.01
		M1	0.42±0.02	0.38±0.02	0.42±0.01	0.39±0.02
		M2	0.01±0.0	0.01±0.01	0.01±0.0	0.01±0.0
Phenylalanine	P-enolpyruvate & Erythrose-4-P	M0	0.25±0.01	0.24±0.02	0.25±0.02	0.24±0.02
		M1	0.44±0.02	0.45±0.01	0.43±0.02	0.45±0.01
		M2	0.25±0.01	0.23±0.02	0.26±0.01	0.24±0.02
Aspartic acid	Oxaloacetate	M0	0.37±0.02	0.43±0.01	0.37±0.01	0.43±0.02
		M1	0.47±0.02	0.44±0.0	0.47±0.02	0.44±0.01
		M2	0.10±0.02	0.10±0.01	0.10±0.01	0.10±0.02
Glutamic acid	2-oxo-glutarate	M0	0.26±0.03	0.29±0.02	0.25±0.01	0.29±0.02
		M1	0.42±0.03	0.46±0.02	0.43±0.02	0.45±0.03
		M2	0.23±0.01	0.21±0.02	0.23±0.02	0.22±0.01
Histidine	Ribose-5-P	M0	0.29±0.03	0.37±0.02	0.28±0.02	0.36±0.02
		M1	0.39±0.02	0.45±0.03	0.39±0.02	0.46±0.02
		M2	0.24±0.01	0.15±0.01	0.24±0.01	0.16±0.01

^aM0, M1, M2 represent unlabeled, singly ¹³C and doubly ¹³C labeled ions, respectively, of a given fragment.

Note: the standard error of each measurement was 0~3% (n=2). Ions [-57] of leucine and isoleucine could not be resolved because their peaks were overlapped by other peaks, and thus their isotopomer data are not clear.