Cell, Volume *166*

Supplemental Information

Sugar Synthesis from CO_2 in *Escherichia coli*

Niv Antonovsky, Shmuel Gleizer, Elad Noor, Yehudit Zohar, Elad Herz, Uri Barenholz, Lior Zelcbuch, Shira Amram, Aryeh Wides, Naama Tepper, Dan Davidi, Yinon Bar-On, Tasneem Bareia, David G. Wernick, Ido Shani, Sergey Malitsky, Ghil Jona, Arren Bar-Even, and Ron Milo

Supplemental Experimental Procedures

1. Genomic modifications

Multiple gene knockouts were obtained by iterative rounds of P1 transductions according to the following procedure: bacteria clone containing the desired gene deletion, derived from the Keio collection (Baba et al., 2006), was purified by single colony isolation. Gene deletion was confirmed by PCR and the clone was used as a donor for the preparation of the P1 transducing lysate using standard protocol (Lennox, 1955). Upon transduction, the recipient strain was plated on LB agar plate supplemented with Kanamycin (Km) as a positive selection marker for transduction. The introduction of the deletion into the recipient strain was confirmed by determining the size of a PCR product, amplified by primers located upstream and downstream of the targeted deletion. Due to the size of the transduced DNA fragment (up to 100 kbp), this method is not suitable for iterative deletions of adjacent genes. When the distance between two target loci was <20 kbp we used a lambda RED-mediated gene replacement using a chloramphenicol (Cm) cassette (Datsenko and Wanner, 2000). PCP20, a temperature sensitive plasmid encoding the FLP recombinase, was used according to standard procedures to eliminate the Km or Cm resistance markers, allowing iterative rounds of deletion. After recombinase expression, the loss of resistance markers was validated using PCR, and the resulting strain was used as the recipient in the next iteration of gene deletion. Whole genome sequencing was used to validate the genotype at the end of the process.

2. Recombinant expression

For recombinant expression of the CBB cycle related components, we constructed a synthetic operon encoding the His-tagged type-II RuBisCO from *Rhodospirillum rubrum ATCC 11170* (*cbbM*), His-tagged phosphoribulokinase (*prkA*) from *Synechococcus elongatus PCC 7942* and carbonic anhydrase (Rru_A2056) from *Rhodospirillum rubrum (CA)*. To differentially control the expression level of cbbM, *prkA* and *CA* we placed synthetic ribosome binding sites with varying translation efficiency (rbs-C, rbs-E and rbs-D, respectively) upstream to the open reading frame, as previously described (Zelcbuch et al., 2013). The synthetic operon was cloned into a pZ vector (Expressys, Germany) with the backbone containing a PLtetO-1 promoter and a p15A medium copy origin of replication to yield the final pCBB vector (accession number KX077536). The vector contains an inactive *tetR* sequence (due to transposon integration during initial propagation steps) and thus aTc induction is not needed. Construct assembly was based on "no-background" cloning methodology previously described (Zelcbuch et al., 2013), with Cm as the selective marker.

3. Computational analysis of RuBisCO dependent strains

To identify metabolically perturbed strains in which cell growth is coupled to RuBisCO dependent carbon fixation flux, we implemented an algorithm based on the principles of flux balance analysis (FBA) (Orth et al., 2010). We started by analyzing all combinations of up to three enzymatic reaction knockouts in central metabolism, and filtered out all those combinations that allow growth without any flux in RuBisCO. For those that cannot produce biomass without RuBisCO, we calculated a slope which is defined as the biomass production rate achieved by allowing a unit of flux in RuBisCO. All other constraints were the same as in standard FBA, where the rate of biomass production is maximized. In more technical terms, the slope is the shadow price (Reznik et al., 2013) of the upper bound on the RuBisCO flux, when that upper bound is set to 0. We implemented our algorithm using the COBRA for Python toolbox (Ebrahim et al., 2013) and using the *E. coli* core model (http://gcrg.ucsd.edu/Downloads/EcoliCore), augmented with the two reactions corresponding to RuBisCO and phosphoribulokinase. It is possible to create a bi-level optimization MILP problem to find the knockout with the highest slope, using the same mathematical principles used in OptKnock (Burgard et al., 2003; Xu et al., 2013) and RobustKnock (Tepper and Shlomi, 2010, 2011). However, since we only used the core *E. coli* model, and had to consider about 20 single knockouts, we chose to explicitly calculate the slope for all possible combinations of one to three knockouts. The total runtime was less than 3 hours on a single Intel™ Core-i7 CPU.

4. Chemostat evolution experiments

Chemostat based laboratory evolution was conducted in four independent experiments. Gas inflow composition was 25% CO_2, 5% O_2 and 70% N_2 and temperature was maintained at 37° C. We used Bioflo 110 chemostats (New Brunswick Scientific, USA) with working volume of 0.7 L except the second experiment which used a DASBox bacterial fermentation system (DASGIP - Eppendorf, Germany) with working volume of 0.1 L. In the first and third experiments dilution rate was set to 0.08 h^{-1} (equivalent to 9 hours doubling time) with feed input of M9 minimal media supplemented with 5 g/L sodium pyruvate, 100 mg/L xylose and 34 mg/L chloramphenicol. In the second and fourth experiments dilution rate was set to 0.035 h^{-1} (equivalent to 20 hours doubling time) with feed input of M9 minimal media supplemented with 5 g/L sodium pyruvate, 25 mg/L xylose and 34 mg/L chloramphenicol. Agitation speed was set to 500 rpm. Biomass density was measured daily and samples for glycerol stocks and the quantification of residual substrate concentrations were collected once a week. For the latter, culture samples were filtered immediately following sampling using a PVDF membrane syringe filter (0.22 μm) (Merck Millipore, Germany) and the filtrates were stored at -20°C for subsequent analysis. To rule out contaminations, culture samples were plated on positive (supplemented with pyruvate and xylose) and negative (glucose) control agar plates. Genetic markers of the ancestor strain were routinely validated in the colonies formed on pyruvate and

xylose. During the first experiment, the chemostat had a technical malfunction on day 20 and the experiment was restarted using a glycerol stock sample taken on day 18 as the inoculum. To avoid the possibility that pyruvate becomes the limiting nutrient instead of xylose, in the second and fourth experiments, xylose concentrations were lowered in the feed media whenever a significant drop in residual pyruvate was detected in the culture (as shown in Supplemental Figure 7). In the first experiment bacterial cells were inoculated following mutagenesis with AquaMutant kit (AquaPlasmid, USA). In the second experiment the ancestral strain inoculated contained a Δ*mutS* mutation that induces hyper-mutability. Several attempts to perform the evolutionary process resulted in culture contamination after one to 10 weeks and are not reported.

5. Pyruvate and xylose measurements
Reported results of the residual concentrations of pyruvate and xylose in the chemostat culture were quantified enzymatically with K-PYRUV and K-XYLOSE assay kits (Megazyme, Ireland) respectively. In addition, pyruvate concentrations were validated using an Agilent 1200 series high-performance liquid chromatography system (Agilent Technologies, USA) equipped with an anion exchange Bio-Rad HPX-87H column (Bio-Rad, USA). The column was eluted with 5 mM sulfuric acid at a flow rate of 0.6 mL/min at 45°C.

6. Isotopic analysis of intracellular metabolites
For mass isotopic distribution analysis, cells were cultured in M9 minimal media, either in the presence of a fully labeled organic carbon source and non-labeled CO_2, or in a reciprocal set-up with isotopically labeled $^{13}CO_2$ (Cambridge Isotope Laboratories, USA) and a non-labeled organic carbon source. For experiments in which gaseous $^{13}CO_2$ was used, culture tubes were placed in a transparent air-tight flask and flushed with 5 volumes of isotopically labeled gas mixture (10% $^{13}CO_2$, 10% O_2 and 80% N_2). When the OD600 reached ≈0.5 the cells were pelleted by centrifugation for 3 min at 5000g, the supernatant was aspirated immediately, and 400 µl of cold (-20°C) acetonitrile:methanol:water (40:40:20) extraction solution was added to the pellet. Following overnight extraction in -20°C, the extracts were centrifuged and the supernatants were transferred to vials and used for an LC-MS analysis. Metabolites were separated using a liquid chromatography system. A ZIC-pHILIC column (4.6 mm × 150 mm, guard column 4.6 mm × 10 mm; Merck) was used for liquid chromatography separation using gradient elution with a solution of 20 mM ammonium carbonate, with 0.1% ammonium hydroxide, and acetonitrile at 0.1 mL/min. Detection of metabolites was performed using a Thermo Scientific Exactive high-resolution mass spectrometer with electrospray ionization, examining metabolites in a polarity switching mode over the mass range of 75–1,000 *m*/*z*. Compound identities were verified by matching masses and retention times to a library of authenticated standards. Data analysis was performed using the Maven software suite (Melamud et al., 2010).

7. Isotopic analysis of hydrolyzed amino acids
For mass isotopologues distributions (MID) analysis of hydrolyzed amino acids, cells were grown as described above section and harvested during exponential growth. Protein biomass was hydrolyzed with 6N hydrochloric acid using standard protocols. Hydrolyzed amino acids were separated using ultra performance liquid chromatography (Acquity - Waters, USA) on a C-8 column (Zorbax Eclipse XBD - Agilent, USA) at a flow rate of 0.6 mL/min and eluted off the column using a hydrophobicity gradient. Buffers used were: A) H_2O + 0.1% formic acid and B) acetonitrile + 0.1% formic acid with the following gradient: 100% of A (0-3 min), 100% A to 100% B (3-9 min), 100% B (9-13 min), 100% B to 100% A (13-14 min), 100% A (14-20 min). Overall run time was 20 minutes. The UPLC was coupled online to a triple quadrupole mass spectrometer (TQS - Waters, USA). Data was acquired using MassLynx v4.1 (Waters, USA). Optimization of ionization parameters and determination of retention times was performed by direct infusion of amino acid commercial standards (Sigma-Aldrich, USA). Argon was used as the collision gas with a flow rate of 0.22 mL/min. Cone voltage was 25V, the capillary was set to 3 kV, source temperature was 150°C, desolvation temperature was 500°C, desolvation gas flow was 700 L/min, source offset 50, cone gas flow was 250 L/min and collision energy was 14eV. MIDs were detected using multiple reaction monitoring (MRM) with the known molecular masses and the neutral loss of carbonyl carbon as a daughter ion (either 47 or 46 m/z, labeled and unlabeled respectively). Data analysis was performed using TargetLynx (Waters, USA).

8. Isotopic carbon ratio in total biomass
Samples were grown in 500 mL glass flasks in M9 minimal media supplemented with sodium pyruvate [5 g/L]. Ancestor strain samples were supplemented with 0.2 g/L xylose in addition to pyruvate. Following inoculation flasks were sealed with rubber septa and flushed with 5 volumes of a gas mixture containing isotopically labeled $^{13}CO_2$. Cells were harvested during exponential growth, washed in M9 minimal media and lyophilized. Between 0.2 and 0.4 mg of dry biomass sample was weighed into tin capsules and the $^{13}C/^{12}C$ ratio was determined using an elemental analyzer linked to a Micromass (Manchester, UK) Optima IRMS.

9. Whole-genome sequencing
DNA was extracted from sampled cultures using DNeasy Blood & Tissue kit (QIAGEN, Germany) and the sequencing was performed as previously described (Blecher-Gonen et al., 2013) with the following modifications: 115-1000 ng of DNA was sheared using the Covaris E220X sonicator (Covaris Inc., USA). End repair was performed in 80ul reaction at 20°C for 30min. Following Agencourt AmPURE XP beads cleanup (Beckman Coulter, Inc., Indianapolis, IN, USA) in a ratio of 0.75X Beads/DNA volume, A bases were added to both 3' ends followed by adapter ligation in a final concentration of 0.125uM. A SPRI bead cleanup at a ratio of 0.75x beads/DNA volume was performed, followed by 8 PCR cycles using 2X

KAPA HiFi ready mix (Kappa Biosystems, Inc., USA) in a total volume of 25ul with the following program: 2 min at 98°C, 8 cycles of 20 sec at 98°C, 30 sec at 55°C, 60 sec at 72°C followed by 72°C at 10 min. The process yielded 100 bp paired-end reads. A minimum of 2.5 million reads was obtained per sample, with a mean of 3.5 million reads per sample. The Illumina short reads generated in this study have been deposited at the European Nucleotide Archive (ENA) with study accession number PRJEB13306.

10. Analysis of sequence data

A reference genome was used which is based on *E. coli* strain BW25113 (*GenBank*: CP009273, Grenier et al., 2014) and the pCBB plasmid containing the recombinant enzymes as a second contig (*GenBank*: KX077536). Sequence alignment and variant calling were achieved by the *breseq* pipeline (Deatherage and Barrick, 2014). The *breseq pro*gram was used to identify genomic variants including single nucleotide polymorphisms (SNPs) and insertion-deletion polymorphisms (INDELs). We note that our analysis does not capture certain types of mutations, such as copy number variation within the genome. Loci with poorly determined alternative alleles were revalidated by PCR amplification and Sanger sequencing. All mutations were validated manually using Geneious (Kearse et al., 2012) version 8.0.5 (Biomatters, New Zealand). The mean and standard deviation of genomic and plasmid DNA sequence coverage were calculated after read alignment in the Geneious software. Plasmid copy number of the various samples were estimated by taking the ratio between the mean coverage of the plasmid and genomic DNA sequence.

11. Plasmid curing and re-introduction of CBB enzymes

For curation of pCBB, clones displaying the hemiautotrophic phenotype were inoculated in permissive media in which growth is independent of carbon fixation activity (minimal media supplemented with pyruvate and glycerol) and incubated at 37°C. Antibiotics were omitted from the growth media to allow the propagation of plasmid-free cells. Stationary culture was plated on agar plates containing permissive media in the absence of antibiotics. Colonies were screened by PCR to identify plasmid loss with primers to the RuBisCO (*cbbM*) gene and for antibiotic sensitivity. Modified plasmids, in which either *cbbM*, *prkA* or CA have been removed, were constructed by PCR using pCBB plasmid as template. Cured cells were re-transformed with either the original or one of the modified plasmids, and hemiautotrophic growth was tested in the resulting transformants.

12. Reintroduction of wild-type *prs* sequence into an evolved genetic background

The reintroduction of the wild-type sequence to replace the mutated *prs* locus was accomplished via P1 transduction. P1 lysate was prepared from Keio collection (Baba et al., 2006) strains containing a Km resistance marker in an adjacent locus to *prs* (either *dauA* or *ychH*). Importantly, donor strains were prs^+. Evolved clones capable of hemiautotrophic growth were transduced using the P1 lysate. Due to the large size (≈100kb) of the DNA fragment delivered to the recipient, in addition to the selection marker, adjacent genes (such as *prs)* are often included in the transduction event, leading to the reintroduction of prs^+ locus in the recipient strain. Transduced colonies were sequenced to identify clones in which the desired replacement to wild-type sequence occurred. Clones in which the *prs* locus was successfully replaced with wild-type sequence were tested for hemiautotrophic growth in M9 minimal media containing only pyruvate and CO_2. As control experiments we performed P1 transductions of loci located over 100kb from the *prs* locus, which hence do not lead to the reintroduction of the prs^+ wild-type sequence.

13. Cloning, purification and kinetic measurements of *prs* variants

The native and mutated *prs* coding sequences were amplified via PCR (from the ancestor and evolved strains, respectively) with the addition of 5' His-tag. The protein was over-expressed using a pET41 plasmid (Novagen) in *E. coli* BL21(DE3) cultured in LB media and induced with IPTG. Tagged proteins were purified using Ni-NTA His•Bind Resin (Novagen) according to the manufacturer's instructions. Activity of purified proteins was determined using an *in-vitro* assay as previously described (Switzer and Gibson, 1978). To estimate the Michaelis constant (K_M), initial reaction rates were measured at various ribose-5-phosphate concentrations. Data was fitted to a Michaelis-Menten curve using MATLAB.

14. Simplified model for stability of an auto-catalytic carbon fixation cycle

The simple model for the CBB cycle focuses on the concentration of the metabolite at the branch point (R5P), and contains two reactions: (i) an effective carbon fixation generalized reaction, denoted as CBB with rate v_{CBB}, consuming 5 molecules of R5P and 5 molecules of CO_2 to produce 6 molecules of R5P, (ii) biomass producing reaction, denoted as prs, consuming one molecule of R5P to produce biomass. Both of these reactions were assumed to follow irreversible Michaelis Menten kinetics with corresponding kinetic parameters K_M^{prs}, V_{max}^{prs}, K_M^{CBB} and V_{max}^{CBB} (where $V_{max} = [E] \cdot k_{cat}$ for each of the reactions).

A steady state concentration of R5P must satisfy that:

$$v_{CBB} - v_{prs} = 0 \Rightarrow V_{max}^{CBB} \frac{[R5P]}{K_M^{CBB} + [R5P]} = V_{max}^{prs} \frac{[R5P]}{K_M^{prs} + [R5P]}$$

Which gives one solution at [R5P]=0 and another solution at:

$$[R5P] = \frac{K_M^{CBB} V_{max}^{prs} - K_M^{prs} V_{max}^{CBB}}{V_{max}^{CBB} - V_{max}^{prs}}$$

For a steady state concentration to be stable, the derivative of $v_{CBB} - v_{prs}$ has to be negative with respect to R5P at the steady state point, implying that if R5P increases above its steady state, the total flux will reduce its concentration and if R5P will decrease below its steady state, the total flux will increase its concentration to return to steady state. In terms of metabolic control analysis (Fell, 1997) this is equivalent to requiring that the elasticity of the *prs* reaction (defined as $\epsilon_{prs} = \frac{dv_{prs}}{d[R5P]} \cdot \frac{[R5P]}{v_{prs}}$) is greater than the elasticity of the CBB reaction at the steady state point ($\epsilon_{CBB} = \frac{dv_{CBB}}{d[R5P]} \cdot \frac{[R5P]}{v_{CBB}}$). As the mathematical derivation is somewhat elaborate, we give a simple argument to derive the stability criteria. First we note that for [R5P]=0:

$$\frac{dv_{CBB}}{d[R5P]}|_{[R5P]=0} = \frac{V_{max}^{CBB}}{K_M^{CBB}} \quad \text{and that} \quad \frac{dv_{prs}}{d[R5P]}|_{[R5P]=0} = \frac{V_{max}^{prs}}{K_M^{prs}}.$$

Therefore, if $\frac{V_{max}^{CBB}}{K_M^{CBB}} < \frac{V_{max}^{prs}}{K_M^{prs}}$ then [R5P]=0 is a stable steady state point, implying that if another steady state point exists for a positive concentration of R5P, it will be unstable. Thus, as can be understood most easily graphically, for a non-zero positive steady state point to exist, the kinetic parameters must satisfy that $\frac{V_{max}^{CBB}}{K_M^{CBB}} > \frac{V_{max}^{prs}}{K_M^{prs}}$. As this implies that $\frac{V_{max}^{prs}}{K_M^{prs}} - \frac{V_{max}^{CBB}}{K_M^{CBB}} < 0$, the concentration of the other steady state point will be positive if and only if:

$$V_{max}^{CBB} - V_{max}^{prs} < 0 \Rightarrow V_{max}^{CBB} < V_{max}^{prs}$$

leading us to conclude that for a positive steady state concentration of R5P to exist the kinetic parameters of CBB and *prs* must satisfy the two conditions:

$$V_{max}^{CBB} < V_{max}^{prs} \qquad (1)$$

$$\frac{V_{max}^{CBB}}{K_M^{CBB}} > \frac{V_{max}^{prs}}{K_M^{prs}} \qquad (2)$$

In our analysis we assume that the maximal rate of carbon fixation is lower than biomass synthesis ($V_{max}^{CBB} < V_{max}^{prs}$) and hence the stability criterion reduces to equation (2).

Supplemental References

Baba, T., Ara, T., Hasegawa, M., and Takai, Y. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol.

Burgard, A.P., Pharkya, P., and Maranas, C.D. (2003). Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. *84*, 647–657.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. *97*, 6640–6645.

Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst. Biol. *7*, 74.

Fell, D. (1997). Understanding the control of metabolism (Portland press London).

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Lennox, E.S. (1955). Transduction of linked genetic characters of the host by bacteriophage P1. Virology *1*, 190–206.

Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? Nat. Biotechnol. *28*, 245–248.

Reznik, E., Mehta, P., and Segrè, D. (2013). Flux imbalance analysis and the sensitivity of cellular growth to changes in metabolite pools. PLoS Comput. Biol. *9*, e1003195.

Switzer, R.L., and Gibson, K.J. (1978). Phosphoribosylpyrophosphate synthetase (ribose-5-phosphate pyrophosphokinase) from Salmonella typhimurium. Methods Enzymol. *51*, 3–11.

Tepper, N., and Shlomi, T. (2010). Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics *26*, 536–543.

Tepper, N., and Shlomi, T. (2011). Computational design of auxotrophy-dependent microbial biosensors for combinatorial metabolic engineering experiments. PLoS One *6*, e16274.

Xu, Z., Zheng, P., Sun, J., and Ma, Y. (2013). ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. PLoS One *8*, e72150.