**Efficient utilization of pentoses for bioproduction of the renewable two-carbon compounds
ethylene glycol and glycolate**

Brian Pereira^{1§}, Zheng-Jun Li^{1,2§}, Marjan De Mey^{1,3}, Chin Giaw Lim^{1†}, Haoran Zhang^{1†}, Claude
Hoeltgen^{1,4}, Gregory Stephanopoulos^{1*}

1. Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA,
02139, U.S.A.

2. Beijing Key Laboratory of Bioprocess, College of Life Science and Technology, Beijing
University of Chemical Technology, Beijing 100029, P. R. China

3. Center for Industrial Biotechnology and Biocatalysis, Department for Biochemical and
Microbial Technology, Ghent University, 9000 Ghent, Belgium

4. Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences,
ETH Zurich, 8093 Zurich, Switzerland

§Authors contributed equally to this study

†Present addresses:

(C.G.L.) Manus Biosynthesis, Cambridge, MA, 02138, U.S.A.

(H.Z.) Rutgers University, New Brunswick, NJ, 08901, USA

*Corresponding author (G. Stephanopoulos)

Mailing address: Department of Chemical Engineering, Massachusetts Institute of Technology,
Room 56-469, 77 Massachusetts Avenue, Cambridge, MA 02139

24 Phone: +1-617-253-4583

25 Fax: +1-617-253-3122

26 E-mail: gregstep@mit.edu

27

28 Running title: Bioproduction of ethylene glycol and glycolate from pentoses

Abstract

The development of lignocellulose as a sustainable resource for the production of fuels and chemicals will rely on technology capable of converting the raw materials into useful compounds; some such transformations can be achieved by biological processes employing engineered microorganisms. Towards the goal of valorizing the hemicellulose fraction of lignocellulose, we designed and validated a set of pathways that enable efficient utilization of pentoses for the biosynthesis of notable two-carbon products. These pathways were incorporated into *Escherichia coli*, and engineered strains produced ethylene glycol from various pentoses, including simultaneously from D-xylose and L-arabinose; one strain achieved the greatest reported titer of ethylene glycol, 40 g/L, from D-xylose at a yield of 0.35 g/g. The strategy was then extended to another compound, glycolate. Using D-xylose as the substrate, an engineered strain produced 40 g/L glycolate at a yield of 0.63 g/g, which is the greatest reported yield to date.

Keywords: ethylene glycol, glycolate, renewable, metabolic engineering, xylose, hemicellulose

1. Introduction

Most of today's chemicals and fuels are derived primarily from crude oil, natural gas, and coal, i.e., from fossil fuels, but as concerns about the use of fossil fuels continue to mount, there is a strong impetus to institute alternative, sustainable resources. Biomass, in particular from plants, is one such promising and abundant resource. Exploiting biomass as a feedstock for biobased chemicals and fuels necessitates technology for the conversion of raw plant material into useful products. This conversion can be effected by catalysts in chemical processes (Ji *et al.*, 2009; Pang *et al.*, 2011; Sun and Liu, 2011) and/or by microorganisms in biological processes (Cesário *et al.*, 2014; Olson *et al.*, 2012). Though chemical processes are faster, they generally exhibit low selectivities, thereby requiring greater energy input and greater capital investment; therefore, an effective, single-step biological process with high selectivity has the potential to lead the market. Consequently, there is a critical need to engineer microorganisms that efficiently yield beneficial products from plant-derived substrates.

Current biotechnology chiefly utilizes plant-derived sugar and starch, but future technologies will likely make use of lignocellulosic feedstocks (Nieves *et al.*, 2015). The most abundant constituents of lignocellulose are D-glucose and D-xylose, which are derived principally from cellulose and hemicellulose, respectively. D-Glucose is easily metabolized by most microorganisms, but only a smaller fraction of microorganisms can metabolize D-xylose; for this reason, metabolic engineering research has concentrated on D-glucose as the substrate (Kallio *et al.*, 2014; Lee *et al.*, 2012; Xu *et al.*, 2013; Yim *et al.*, 2011), with significantly fewer studies employing D-xylose (Kim *et al.*, 2013; Kurosawa *et al.*, 2013; Wasylenko and Stephanopoulos, 2015). However, D-xylose is a significant constituent of lignocellulose and can be obtained from biomass at a lower cost; therefore, D-xylose represents an important resource

for valorization. Here we propose a scheme by which D-xylose, and other hemicellulose-derived pentoses, can be utilized efficiently for the production of two-carbon products.

Various two-carbon compounds can be produced through biosynthetic pathways branching from central metabolism intermediates. These intermediates are generally three-carbon intermediates, and one carbon is lost for each two-carbon product that is formed; the theoretical yield for such pathways is therefore 2:3 C-mol-product/C-mol-substrate. Alternatively, if D-xylose, instead of being metabolized through the pentose phosphate pathway, is cleaved into two- and three-carbon intermediates and if each intermediate can be subsequently converted into the product, this new approach can yield 4:5 C-mol-product/C-mol-substrate (2:1 mol-product/mol-D-xylose), a significant, 20% improvement. Implementation of this strategy requires pathways for the cleavage of D-xylose into two- and three-carbon intermediates, for transforming the two-carbon intermediate into the target product, and for the conversion of the three-carbon intermediate into product. These pathways will ostensibly depend on the particular product, and for the initial targets we have selected ethylene glycol (EG) and glycolate.

EG is a commodity chemical used in multiple everyday applications; the two most prominent uses are as automotive antifreeze and as one of the precursors for the polymer poly(ethylene terephthalate). As such, EG is produced in large volumes, greater than 15 million tons globally, and the vast majority is generated from fossil fuels. There is a small fraction of the market that is indirectly derived from sugar, but this biobased EG is achieved through the fermentation of sugar into biobased ethanol followed by multiple energy-intensive chemical processes. There is a clear opportunity for a single-step bioprocess for renewable EG production. On this basis, Liu *et al.* investigated the direct production of EG from D-xylose, and they designed a pathway for the cleavage of D-xylose into two- and three-carbon intermediates (Liu *et*

al., 2013). The engineered strain utilized the two-carbon intermediate for EG production and achieved 11.7 g/L EG at 0.29 g/g yield. In a recent analogous study, Alkim *et al.* employed another pathway to improve production to 20 g/L EG with a yield of 0.38 g-EG/g-D-xylose (2015). Here, we design and implement an alternative, novel set of pathways that enable EG production from various pentoses, including superior titer from D-xylose: 40 g/L EG at 0.35 g/g yield. We then show that the approach can be extended to another product, glycolate (the conjugate base of glycolic acid).

Glycolic acid is a notable α -hydroxy acid that has applications primarily in skin care and personal care products, but also in dyeing and tanning, as a food additive, as a cleaning agent, and as a precursor for biocompatible polymers. With increasing demand, the global market is expected to reach USD 277.8 million by 2020 (<http://www.grandviewresearch.com/press-release/global-glycolic-acid-market>). Though glycolic acid is found in natural sources including sugarcane and sugar beet, it is chiefly manufactured from fossil fuels, via the carbonylation of formaldehyde by synthesis gas. In addition to the issues surrounding fossil fuels, there are concerns about formaldehyde being associated with the final glycolic acid product, especially as used in personal care products. Accordingly, there is again an opportunity for a single-step bioprocess. Direct bioproduction of renewable glycolate has been established previously (Koivistoinen *et al.*, 2013; Martin *et al.*, 2013; Zahoor *et al.*, 2014; Cam *et al.*, 2015; Deng *et al.*, 2015); these studies encompass various substrates and organisms, but, with the exception of the recently published work by Cam *et al.* (2015), all of the studies employ the glyoxylate shunt for glycolate production. In this study, we demonstrate that cleavage of D-xylose can be utilized for glycolate production and that the pathway can complement the glyoxylate shunt pathway for very efficient production of glycolate, achieving 0.63 g/g yield.

2. Materials and methods

2.1. Strains and plasmids

Escherichia coli K-12 MG1655 (DE3) $\Delta recA \Delta endA$ was used as the parent for all strain engineering, and *E. coli* K-12 MG1655 (DE3) $\Delta recA \Delta endA \Delta aldA$ was constructed previously (Pereira *et al.*, 2015). Site-directed chromosomal deletions were introduced by homologous recombination according to established methods (Datsenko and Wanner, 2000). For deletion of *xylB* and *araB* genes, linear DNA for homologous recombination was generated by PCR-amplifying the FRT-flanked antibiotic resistance cassette from pKD4 and pKD3, respectively; the corresponding primer pairs were KOxylB_P1 & KOxylB_P2 and KOaraB_P1 & KOaraB_P2 (see Supplementary Table I for all oligonucleotide sequences). For deletion of *fucO*, *glcD*, *gcl*, *aceB*, and *glcB* genes, strains containing these disruptions were purchased from The Coli Genetic Stock Center (Yale University, New Haven, CT), and linear DNA was generated by PCR-amplifying the corresponding genomic DNA using KOgene_fw & KOgene_rv primer pairs. The various modified strains were used as production hosts. *E. coli* NEB 5-alpha (NEB, Ipswich, MA, USA) was used for plasmid maintenance.

Plasmid construction and DNA manipulations were performed following standard molecular cloning protocols, and all associated enzymes were purchased from NEB. The D-tagatose 3-epimerase gene from *Pseudomonas cichorii* (*Pc.dte*) (Ishida *et al.*, 1997) was codon optimized for *E. coli* (see Supplementary Information for sequence) and synthesized by Bio Basic Inc. (Markham, Ontario, Canada). From the synthesized template, *Pc.dte* was PCR-amplified using primers Pc.dte_fw & Pc.dte_rv, and using primers fucA_fw1 & fucO_rv1 and fucK_fw1 & fucK_rv1, *fucA-fucO* and *fucK*, respectively, were amplified from *E. coli* K-12

136 MG1655 genomic DNA. The PCR products were digested with *NcoI/SacI*, *FseI/SacI*, and
137 *SacI/SalI*, respectively; these digested fragments were successively cloned into the *NcoI/SalI* site
138 of p10_T5 (Ajikumar *et al.*, 2010), *FseI/SacI* site of p10_T5-*Pc.dte*, and the *SacI/SalI* site of
139 p10_T5-*Pc.dte-fucA-fucO*, yielding p10_T5-*Pc.dte-fucA-fucO-fucK*. Next, p5_T5-*Pc.dte-fucA-*
140 *fucO-fucK* was constructed by subcloning the *XbaI/AccIII* fragment of p10_T5-*Pc.dte-fucA-*
141 *fucO-fucK* into the *XbaI/AccIII* site of p5_T7 (Ajikumar *et al.*, 2010).

142 In order to construct p5_T7-*Pc.dte*, the T7 promoter was amplified from p10_T7
143 (Ajikumar *et al.*, 2010) using primers T7_fw & T7_rv, digested with *MluI/NcoI*, and cloned into
144 the *MluI/NcoI* site of p5_T5-*Pc.dte*. From *E. coli* genomic DNA, *rhaB*, *rhaD*, and *fucO* were
145 amplified using primers *rhaB_fw1* & *rhaB_rv*, *rhaD_fw* & *rhaD_rv1*, and *fucO_fw* & *fucO_rv2*,
146 respectively. The PCR products were digested with *FseI/SalI*, *SacI/NotI*, and *NotI/XhoI*,
147 respectively; the digested fragments were successively cloned into the *FseI/SalI* site of p5_T7-
148 *Pc.dte*, the *SacI/NotI* site of p5_T7-*Pc.dte-rhaB*, and the *NotI/XhoI* site of p5_T7-*Pc.dte-rhaB-*
149 *rhaD*, yielding p5_T7-*Pc.dte-rhaB-rhaD-fucO*. Plasmid p5_T5-*Pc.dte-fucA-fucO-fucK-rhaB-*
150 *rhaD* was constructed by first PCR-amplifying *rhaB-rhaD* from p5_T7-*Pc.dte-rhaB-rhaD-fucO*
151 using primers *rhaB_fw2* & *rhaD_rv2* and digesting p5_T5-*Pc.dte-fucA-fucO-fucK* with *SalI*;
152 these products were combined using Gibson assembly (Gibson *et al.*, 2009; Gibson *et al.*, 2010).

153 Furthermore, from *E. coli* genomic DNA, *fucA*, *fucK*, and *aldA* were amplified using
154 primers *fucA_fw2* & *fucA_rv*, *fucK_fw2* & *fucK_rv2*, and *aldA_fw* & *aldA_rv*, respectively.
155 The PCR products were digested with *FseI/SalI*, *SacI/NotI*, and *NotI/XhoI*, respectively; the
156 digested fragments were successively cloned into the *FseI/SalI* site of p5_T5-*Pc.dte-fucA-fucO-*
157 *fucK*, the *SacI/NotI* site of p5_T5-*Pc.dte-fucA*, and the *NotI/XhoI* site of p5_T5-*Pc.dte-fucA-fucK*,
158 yielding p5_T5-*Pc.dte-fucA-fucK-aldA*. To replace the T5 promoter with a constitutive one, we

searched the Registry of Standard Biological Parts (<http://parts.igem.org>), and we selected the strongest member of a family of constitutive promoters, part BBa_J23119, herein designated as CP1. A 537-bp DNA fragment (see Supplementary Information for sequence) containing *SpeI*-CP1-RBS-*NcoI* was synthesized; the synthesized fragment was digested with *SpeI/NcoI* and used to replace the 1699-bp *XbaI-NcoI* DNA fragment of p5_T5-*Pc.dte-fucA-fucK-aldA* by digestion and ligation, giving p5_CP1-*Pc.dte-fucA-fucK-aldA*.

The plasmid, pHHD-*ycdW-aceA-aceK* (Martin *et al.*, 2013), was generously shared by Prof. Kristala Prather. The 5.0-kb DNA fragment amplified from pBBR1MCS-2 (Kovach *et al.*, 1995) using primers MCS2_fw & MCS2_rv was purified and self-ligated to generate pMCS-CP1, harboring the constitutive promoter CP1. The *ycdW-aceA-aceK* fragment was isolated from pHHD-*ycdW-aceA-aceK* by *EcoRI/BamHI* digestion and then cloned into pMCS-CP1 digested with the same enzymes to generate the constitutive expression plasmid pMCS-*ycdW-aceA-aceK*. The myriad plasmids were transformed into the various hosts to give production strains; the relevant plasmids and strains are summarized in Table I.

Table I: Select plasmids and strains used in this study.

Name	Description
<i>Plasmids</i>	
pEGx	p5_T5- <i>Pc.dte-fucA-fucO-fucK</i>
pEGa	p5_T7- <i>Pc.dte-rhaB-rhaD-fucO</i>
pEGxa	p5_T5- <i>Pc.dte-fucA-fucO-fucK-rhaB-rhaD</i>
pGAx1	p5_T5- <i>Pc.dte-fucA-fucK-aldA</i>
pGAx2	pHHD- <i>ycdW-aceA-aceK</i> ¹⁶

pGAx3	p5_CP1- <i>Pc.dte-fucA-fucK-aldA</i>
pGAx4	pMCS- <i>ycdW-aceA-aceK</i>
<i>Strains</i>	
WT	<i>E. coli</i> K-12 MG1655 (DE3) $\Delta endA \Delta recA$
$\Delta aldA$	WT $\Delta aldA$
EG-X	WT $\Delta xylB \Delta aldA$ / pEGx
EG-A	WT $\Delta araB \Delta aldA$ / pEGa
EG-XA	WT $\Delta xylB \Delta araB \Delta aldA$ / pEGxa
GA-01	WT $\Delta xylB$ / pGAx1
GA-02	WT $\Delta xylB \Delta fucO$ / pGAx1
GA-03	WT $\Delta xylB \Delta glcD$ / pGAx1
GA-04	WT $\Delta xylB \Delta fucO \Delta glcD$ / pGAx1
GA-05	WT $\Delta xylB \Delta glcD$ / pGAx1 / pGAx2
GA-06	WT $\Delta xylB \Delta glcD \Delta gcl$ / pGAx1 / pGAx2
GA-07	WT $\Delta xylB \Delta glcD \Delta aceB$ / pGAx1 / pGAx2
GA-08	WT $\Delta xylB \Delta glcD \Delta glcB$ / pGAx1 / pGAx2
GA-09	WT $\Delta xylB \Delta glcD \Delta aceB \Delta glcB$ / pGAx1 / pGAx2
GA-10	WT $\Delta xylB \Delta glcD \Delta aceB \Delta glcB \Delta gcl$ / pGAx1 / pGAx2
GA-11	WT $\Delta xylB \Delta glcD \Delta aceB \Delta glcB \Delta gcl$ / pGAx3 / pGAx4

175

176

177 2.2. Media and cultivation conditions

Chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) unless otherwise stated. Strain construction and fermentation seed-cultures were performed in lysogeny broth (LB), which was prepared per instructions (BD, NJ, USA). Experiments conducted in tubes (50-mL capped, conical centrifuge tubes) or shake flasks (250-mL flasks) utilized minimal medium (MM1) comprising 2.0 g/L NH_4Cl , 5.0 g/L $(\text{NH}_4)_2\text{SO}_4$, 3.0 g/L KH_2PO_4 , 7.3 g/L K_2HPO_4 , 8.4 g/L MOPS, 0.5 g/L NaCl, 2 mL/L 1M MgSO_4 , 1 mL/L mineral solution, 0.1 mL/L 4 mM Na_2MoO_4 , and specified sugar; pH was adjusted to 7 with 6 M KOH. The mineral solution consisted of 3.6 g/l $\text{FeCl}_2 \cdot 4\text{H}_2\text{O}$, 5.0 g/l $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 1.3 g/l $\text{MnCl}_2 \cdot 2\text{H}_2\text{O}$, 0.38 g/l $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, 0.50 g/l $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 0.94 g/l ZnCl_2 , 0.031 g/l H_3BO_3 , 0.40 g/l $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$, and 1.0 g/l thiamine-HCl. Bioreactor fermentations were carried out in a 3-L Bioflo culture vessel (New Brunswick, CT, USA) with a 2-L working volume. These fermentations utilized minimal medium (MM2) consisting of 2.0 g/L NH_4Cl , 5.0 g/L $(\text{NH}_4)_2\text{SO}_4$, 2.0 g/L KH_2PO_4 , 0.5 g/L NaCl, 2 mL/L 1M MgSO_4 , 1 mL/L mineral solution, 0.1 mL/L 4 mM Na_2MoO_4 , and specified sugar; additionally, silicone antifoaming B emulsion was used to prevent foaming. Aerobic conditions were maintained by sparging air at 0.5 or 1 L/min, and the pH was maintained at 7.0 with 6 N NaOH. For all cultures, 50 mg/L spectinomycin, 34 mg/L chloramphenicol, and 50 mg/L kanamycin were added as appropriate. To induce gene expression, 0.1 mM IPTG (Zymo Research, Irvine, CA, USA) and/or 250 $\mu\text{g/L}$ aTc were added.

To test EG production from D-arabinose or L-lyxose, WT and ΔaldA strains were first cultivated in LB overnight (o/n) at 37 °C, 250 rpm. These seed cultures were used to inoculate to 1% v/v 10-mL tube cultures containing MM1 with 10 g/L L-fucose or L-rhamnose; these cultures were incubated at 37 °C, 250 rpm. The subsequent cultures were used to inoculate to 1%

v/v 10-mL tube cultures containing MM1 with 10 g/L D-arabinose or L-lyxose (Carbosynth, Berkshire, UK), respectively, and the cultures were incubated at 37 °C, 250 rpm.

For fed-batch bioreactor production of EG from D-xylose, an initial culture of strain EG-X was grown in LB at 37 °C, 250 rpm o/n. This culture was used to inoculate to 1% v/v two 50-mL flask cultures of MM1 with 15 g/L D-glucose. After growing o/n at 37 °C, 250 rpm, these seed cultures were combined and used to inoculate to 5% v/v a bioreactor containing MM2 with approximately 35 g/L D-xylose. Temperature was maintained at 37 °C, and dissolved oxygen content was maintained at 30% by altering agitation from 400 to 650 rpm. At 24 h, we initiated pumping of a feed solution into the bioreactor through a reactor port. The feed solution consisted of 600 g/L D-xylose, 10 g/L (NH₄)₂SO₄, 5.0 g/L MgSO₄, 0.1 mM IPTG, and spectinomycin. The flow rate was varied from 0.05 mL/min to 0.15 mL/min so as to maintain a concentration of D-xylose between 0 and 10 g/L. Similarly, an initial culture of strain EG-A was grown in LB at 37 °C, 250 rpm o/n and used to inoculate to 1% v/v two 50-mL flask cultures of MM1 with 15 g/L D-glucose. After growing o/n at 37 °C, 250 rpm, these seed cultures were combined and used to inoculate to 5% v/v a bioreactor containing MM2 with approximately 50 g/L L-arabinose (Chem-Impex International Inc., Wood Dale, IL, USA). Temperature was maintained at 37 °C, and dissolved oxygen content was maintained at 30% by altering agitation from 400 to 650 rpm. Flask cultures containing MM1 with 15 g/L D-xylose and 15 g/L L-arabinose were inoculated from LB cultures of strain EG-XA. These were cultivated at 30 °C, 250.

To assess glycolate production, strains were first grown in LB at 37°C, 250 rpm o/n. These cultures were used to inoculate to 1% v/v 10-mL tube cultures containing MM1 with 10 g/L D-xylose. The tubes were cultivated at 37 °C, 250 rpm for strains GA-01, GA-02, GA-03,

and GA-04, and at 30 °C, 250 rpm for strains GA-05, GA-06, GA-07, GA-08, GA-09, and GA-10. For batch bioreactor production of glycolate from D-xylose, an initial culture of strain GA-03, GA-10, or GA-11 was grown in LB at 37 °C, 250 rpm o/n. This culture was used to inoculate to 1% v/v two 50-mL flask cultures of MM1 with 15 g/L D-glucose. After growing o/n at 37 °C, 250 rpm, these seed cultures were combined and used to inoculate to 5% v/v a bioreactor containing MM2 with D-xylose; the initial D-xylose concentration was approximately 110 g/L for GA-03 and 65 g/L for GA-10 and GA-11. Temperature was maintained at 37 °C for strain GA-03 and 30 °C for strains GA-10 and GA-11. Dissolved oxygen content was maintained at 20% by altering agitation from 200 to 800 rpm.

2.3. Analytical methods

Culture samples were taken at several times throughout the experiments. Cell growth was assayed by measuring optical densities at 600 nm (Ultrospec 2100 pro, Amersham (GE Healthcare), NJ, USA). Metabolites in the culture supernatants were analyzed in an Agilent 1200 Infinity series HPLC system (Agilent, CA, USA). Analytes were separated by an Aminex HPX-87H column (Bio-Rad, CA, USA) and measured by a refractive index detector. The column was heated to 50 °C, and a mobile phase of 14 mM H₂SO₄ was run at 0.7 mL/min. Concentrations of arabinose, lyxose, xylose, EG, glycolate, and other metabolites were calculated based on standard curves.

3. Results

3.1. Production of ethylene glycol from D-arabinose and L-lyxose

Studies of *E. coli* K-12 metabolism have found that D-arabinose is metabolized by the same enzymes that degrade L-fucose (LeBlanc and Mortlock, 1971). These enzymes, FucI, FucK, and FucA, effect the cleavage of L-fucose into lactaldehyde and dihydroxyacetone phosphate (DHAP) and the cleavage of D-arabinose into glycolaldehyde and DHAP. While lactaldehyde is oxidized to lactate by AldA under aerobic conditions, it is reduced to propanediol by FucO under anaerobic conditions (Baldomà *et al.*, 1988). We expected that FucO and AldA similarly catalyze the reduction and oxidation of glycolaldehyde into EG and glycolate, respectively (Caballero *et al.*, 1983). Therefore, as shown in Figure 1, D-arabinose should yield EG via native metabolism under anaerobic conditions or if AldA is attenuated. The same rationale was applied to the analogous metabolism of L-lyxose, which can be acted on by enzymes responsible for L-rhamnose metabolism, RhaA, RhaB, and RhaD (Badía *et al.*, 1991); EG can be potentially produced through native L-lyxose metabolism. To test these hypotheses, we grew *E. coli* K-12 (see Table I for strain info) in minimal medium with D-arabinose or L-lyxose as the carbon source. In both cases, approximately 1.8 g/L EG was produced (Fig. 2A), thereby validating the pathways; however formation of glycolate was also observed (Supplementary Table II). When the *aldA* gene was deleted to attenuate glycolaldehyde oxidation, glycolate formation was effectively prevented, and EG production was significantly improved, surpassing 3.4 g/L. Furthermore, these titers of EG were paralleled by the corresponding yields (Fig. 2B); WT achieved less than 0.22 g/g (theoretical pathway yield is 0.41 g/g) from either substrate, compared to greater than 0.35 g/g yield for the $\Delta aldA$ strain. These compelling results demonstrate effective EG-biosynthesis pathways from D-arabinose and L-lyxose, but these two pentoses make up very little of hemicellulose and are consequently not very abundant.

Figure 1: Metabolic network for production of EG and/or glycolate from pentoses. Enzymes associated with the reactions are shown. Blue reactions and intermediates are associated with pentose cleavage into two- and three-carbon intermediates, green with conversion of the two-carbon intermediate, and red with conversion of the three-carbon intermediate. Gray reactions and intermediates correspond to pathways attenuated in specified strains.

Figure 2: Production of EG from pentoses. WT and $\Delta aldA$ *E. coli* strains were cultivated in tube cultures with D-arabinose or L-lyxose as the substrate; final EG titers and yield are shown in **a** and **b**, respectively. Columns are the average of triplicate experiments, and error bars represent standard deviation. **c**) Strain EG-X was cultivated on D-xylose in a fed-batch bioreactor; cell growth, EG concentration, and D-xylose concentration were measured over time. After the initial batch, D-xylose was maintained at low levels. Data points are the average of duplicate fermentations, and error bars represent standard deviation. **d**) Strain EG-A was cultivated on L-arabinose in a batch bioreactor; cell growth, EG concentration, and arabinose concentration were measured over time. The prolonged lag phase is not shown. **e**) Strain EG-XA was cultivated on D-xylose and L-arabinose in shake flasks; cell growth, EG concentration, D-xylose concentration,

and L-arabinose concentration were measured over time. Data points are the average of triplicate fermentations, and error bars represent standard deviation. The prolonged lag phase is not shown.

3.2. Production of ethylene glycol from D-xylose and L-arabinose

As discussed above, D-xylose is the most abundant pentose, so EG production from D-xylose is significantly more beneficial. In *E. coli* K-12, D-xylose is metabolized first by isomerization to D-xylulose followed by phosphorylation to D-xylulose-5-phosphate which then continues through the pentose phosphate pathway. Instead of exploring a means to cleave one of these intermediates, we sought an enzyme that could connect the pathway for D-xylose metabolism to that for D-arabinose metabolism. In 1993, Izumori *et al.* identified an enzyme from *Pseudomonas cichorii* that catalyzes the epimerization of D-tagatose (Izumori *et al.*, 1993); this D-tagatose epimerase (DTE) was also found to interconvert xylulose and ribulose (Itoh *et al.*, 1994; Izumori *et al.*, 1993), so we hypothesized that DTE can be used to enable D-xylose to proceed into the D-arabinose degradation pathway and thereby yield EG (Fig. 1).

Toward this end, in the $\Delta aldA$ strain we additionally knocked out *xylB* to prevent native D-xylose metabolism and overexpressed the gene encoding DTE (*Pc.dte*) from a low-copy plasmid (p10) (Ajikumar *et al.*, 2010). When this strain was cultivated in minimal medium with D-xylose as the carbon source, the strain grew and exhibited concomitant EG production; however, the culture experienced a long lag phase and very slow growth (data not shown). We then incorporated more pathway genes (*fucK*, *fucA*, and *fucO*) into the plasmid-borne operon and tested variations of the gene order; several of these variations resulted in faster performance, with p10_T5-*Pc.dte-fucA-fucO-fucK* providing the best results. It is worth noting that, although the FucO enzyme becomes inactivated in aerobic conditions (Cabiscol *et al.*, 1990),

overexpression has been shown to positively affect EG production (Alkim *et al.*, 2015; Pereira *et al.*, 2015). Performance on xylose was further improved by changing to a five-copy plasmid (p5) (Ajikumar *et al.*, 2010) (data not shown), and we next evaluated this high-performing strain (EG-X) in a fed-batch bioreactor. The initial batch of 35 g/L D-xylose was mostly consumed in approximately 24 h, and D-xylose was fed for an additional 48 h; the final titer reached 40 g/L at a yield of 0.35 g/g, which is 85% of the theoretical yield from only the two-carbon pathway, and productivity was 0.58 g/(L*h) (Fig. 2C). The titer and productivity are the best reported metrics for microbial EG production, and the yield compares favorably to the greatest published yield, 0.39 g-EG/g-D-xylose (Alkim *et al.*, 2015).

We next engineered a strain to utilize the second most abundant pentose, L-arabinose. Analogous to the D-xylose strain, the arabinose strain contains a gene deletion ($\Delta araB$) to prevent native metabolism and a DTE-expressing plasmid (p5_T7-*Pc.dte-rhaB-rhaD-fucO*) to support L-arabinose metabolism through the pathway for L-lyxose metabolism. This strain (EG-A) was cultivated in a batch bioreactor with 50 g/L L-arabinose as the substrate. There was a significant lag phase, but the strain then grew quickly and exhibited concomitant EG production, achieving a titer of 20 g/L corresponding to a yield of 0.38 g/g (Fig. 2D). Having demonstrated production from each of D-xylose and L-arabinose, we next engineered a single strain that could generate EG from both substrates, as both will likely be present in a hemicellulose-based feedstock. In this new strain, *aldA*, *xylB*, and *araB* are deleted, and *Pc.dte*, *fucA*, *fucO*, *fucK*, *rhaB*, and *rhaD* are all expressed in a single, plasmid-borne operon; this strain (EG-XA) was grown in shake flasks with 15 g/L D-xylose and 15 g/L L-arabinose. There was again a significant lag phase, but eventually, the strain grew and produced EG, attaining a titer of 10.5 g/L (Fig. 2E). These results indicate that EG is derived from both pentoses; furthermore, it is

interesting to note that D-xylose and L-arabinose were consumed concurrently rather than exhibiting diauxic growth (Desai and Rao, 2010). Though further optimization is needed, the designed metabolic network provides a promising foundation for hemicellulose-based EG biosynthesis, and it can serve as the basis for the production of other compounds as well.

3.3. Production of glycolate from D-xylose using the two-carbon intermediate

As described above, glycolaldehyde can be oxidized to glycolate by AldA; this reaction was attenuated for EG production in order to prevent by-product formation. However, glycolate is itself a valuable product, so we next pursued glycolate production, focusing on D-xylose as the substrate. We thus started with *E. coli* K-12 $\Delta xylB$ as the host background, and to this strain we made additional modifications. In one strain, the *fucO* gene was deleted in order to prevent glycolaldehyde reduction while in another strain *glcD*, which encodes a subunit of glycolate oxidase, was deleted to prevent oxidation of glycolate to glyoxylate; there is also a strain in which both genes are knocked out. These host strains were used to overexpress a p5 plasmid comprising *Pc.dte*, *fucA*, *fucK*, and *aldA*, yielding strains GA-01, GA-02, GA-03, and GA-04. When the four strains were cultivated in D-xylose, glycolate production was observed for all of them, validating the glycolate-biosynthesis pathway (Table II). Deletion of *fucO* had little effect on the final glycolate titer, but deletion of *glcD* resulted in significant improvement. Based on these results, strain GA-03 was selected for scale-up; the strain was grown in a bioreactor with an initial batch of D-xylose. Nearly 100 g/L D-xylose was consumed within 48 h, and the glycolate titer reached 44 g/L, corresponding to a yield of 0.44 g/g (theoretical pathway yield is 0.50 g/g) and a productivity of 0.92 g/(L*h) (Fig. 3); moreover, there was very little EG byproduct formed despite the strain not having *fucO* deleted (Supplemental Figure 1). In this

strain, only the two-carbon fraction of D-xylose was utilized, so additional utilization of the three-carbon fraction should further improve the yield of glycolate.

Table II: Glycolate titer and yield after 48 h of cultivation on minimal medium with D-xylose. Data are averages of triplicate experiments, and errors represent standard deviation (errors less than 0.01 not specified)

Strain	Glycolate Titer (g/L)	Yield (g-glycolate/g-D-xylose)
GA-01	2.14±0.03	0.26
GA-02	2.16±0.07	0.26±0.01
GA-03	3.43±0.02	0.46
GA-04	3.58±0.06	0.47±0.01

Figure 3: Production of glycolate using the two-carbon pathway. Strain GA-03 was cultivated on minimal medium with D-xylose in a batch bioreactor; cell growth, glycolate concentration, and D-xylose concentration were measured over time.

3.4. Production of glycolate from D-xylose using both two-carbon and three-carbon intermediates

We have shown the effectiveness of the pathways for pentose cleavage and for product biosynthesis from the consequent two-carbon intermediate, but for full utilization of a pentose, there also needs to be a pathway for product biosynthesis from the three-carbon intermediate. In our designed metabolic network, the initial three-carbon intermediate is DHAP, which is metabolized through central metabolism. Because glycolate production via the glyoxylate shunt has been established, we can adopt this complementary glycolate-biosynthesis pathway to demonstrate efficient pentose use. A plasmid (generously provided by Prof. Kristala Prather) comprising the pathway genes *ycdW*, *aceA*, and *aceK*, was incorporated into strain GA-03, yielding new strain GA-05, and GA-05 was grown in a 10-mL culture of minimal medium with D-xylose as the carbon source. Glycolate was produced at a yield of 0.52 g/g (Fig. 4A), exceeding the yield for GA-03 (Table II), which only utilized the two-carbon pathway for glycolate biosynthesis, and thus confirming the concerted activity of both pathways.

Figure 4: Production of glycolate using both two-carbon and three-carbon pathways. **a)** Strains were cultivated in tube cultures containing minimal medium and D-xylose, and final yields were calculated after 48 h. Columns are the average of triplicate experiments, and error bars represent standard deviation. **b)** Strain GA-10 and **c)** strain GA-11 were cultivated in batch bioreactors

containing minimal medium with D-xylose; cell growth, glycolate concentration, and D-xylose concentration were measured over time.

In an effort to further improve the yield, we next endeavored to prevent loss of glyoxylate through side reactions. The gene *gcl* encodes glyoxylate carboligase which catalyzes the condensation of two glyoxylate molecules to give tartronate semialdehyde, and the genes *aceB* and *glcB* each encode an isozyme of malate synthase, which catalyzes the reaction between glyoxylate and acetyl-CoA. Some or all of these genes were deleted as further modifications to the glycolate-producing strain, and we evaluated the performance of these new strains (Fig. 4A). Deletion of *gcl* (strain GA-06) had little effect while the yield was significantly increased by deletion of either *aceB* or *glcB* (strains GA-07 and GA-08, respectively), though combination of the gene deletions did not further improve the yield. Hence, we decided to proceed with strain GA-10 for scale up studies.

Production of glycolate by strain GA-10 was scaled up by cultivating it in a bioreactor. A single batch of approximately 65 g/L D-xylose was used as the substrate, and all D-xylose was consumed after 85 h (Fig. 4B). From this amount of D-xylose, the strain produced 40 g/L glycolate, corresponding to a yield of 0.63 g/g (theoretical pathway yield is 1.0 g/g); compared to the two-carbon pathway alone, the combination of two- and three-carbon pathways generated approximately the same titer while increasing the yield by roughly 40%. This is a remarkable advance toward the efficient use of hemicellulose, but the use of inducer molecules, isopropyl β -D-1-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTc), could be cost prohibitive for an industrial process; therefore, we adapted the system to instead use constitutive promoters. For both plasmids, the promoters were each changed to a constitutive promoter. These new plasmids

were incorporated into new strain GA-11 which was then cultivated in a batch bioreactor. Similar levels of glycolate production were realized, achieving 41 g/L glycolate and a yield of 0.62 g/g without the use of inducer molecules (Fig. 4C). These metrics are encouraging for an industrial process, and they exemplify high-efficiency use of D-xylose for the production of a two-carbon compound.

4. Discussion

Lignocellulosic biomass is an abundant, sustainable resource that can be utilized as raw material for the production of chemicals and fuels, and whereas lignin and cellulose are recalcitrant, the hemicellulose fraction is more easily degraded; the resulting hydrolysate is rich in pentoses, primarily D-xylose, and can serve as a meaningful feedstock for bioprocesses. To demonstrate the valorization of such pentoses, we designed and implemented a metabolic network by which several pentoses are first cleaved into glycolaldehyde and DHAP. Because glycolaldehyde production is directly coupled to pentose metabolism, subsequent single-step conversion to either EG or glycolate occurs at very high yield; considering only this two-carbon fraction, the engineered strains achieved at least 85% of the theoretical yield. Validated by such positive results, this approach could possibly be extended to efficiently produce other two-carbon compounds. Glycolaldehyde itself can be a desirable product, and targeted production would be a matter of attenuating both oxidation and reduction. Oxidation of glycolate by glycolate oxidase (GlcD, GlcE, and GlcF subunits) can be exploited to give glyoxylate (Pellicer *et al.*, 1996). Moreover, ethylene glycol can be converted by diol dehydratase to acetaldehyde (Hartmanis and Stadtman, 1986), which can then be transformed into ethanol, acetate, or acetyl-CoA. For any of

these products, or still others, this strategy and metabolic network provides a good foundation for complementary biosynthetic pathways.

The other three carbons of the substrate pentose enter central metabolism as DHAP, and from central metabolism, they can be diverted to the above-mentioned glycolate, glyoxylate, ethanol, and acetate, but also to many other products derived from central metabolism. These myriad pathways can be used for efficient co-production of different compounds; as a proof of concept, we have co-produced glycolaldehyde-derived EG and glyoxylate-derived glycolate, at yields of 0.35 g/g and 0.17 g/g, respectively (data not shown). However, the greatest value will likely come from focused production of a single product. A novel pathway for the production of EG from central metabolism has only been recently developed (Chen *et al.*, 2015; Pereira *et al.*, 2015), so incorporation of this pathway into strain EG-X is in progress. For glycolate, our engineered strains were able to cleave D-xylose and effectively use both the two-carbon and three-carbon fractions. Although the use of D-xylose for glycolate production has been reported previously, the yield reached a maximum of 0.45 g/g (Cam *et al.*, 2015) because only the two-carbon fraction was utilized; from D-glucose, the greatest yield is 0.52 g/g (Deng *et al.*, 2015). The engineered strains in this study have surpassed these metrics, achieving 0.63 g-glycolate/g-xylose. Therefore, this study represents an encouraging demonstration of a future single-step bioprocess for the production of renewable glycolate and establishes a set of pentose-specific pathways that can be adapted for efficient production of two-carbon compounds.

Acknowledgments

We acknowledge partial support from the US Department of Energy, award number 0130-G-PA-291. Z.J.L. acknowledges funding from the National Natural Science Foundation of China

(21476014). We thank J.L. Avalos and D. Dugar for helpful discussions. We thank M.A. Keibler and T.M. Wasylenko for their careful reading of this manuscript and valuable comments. We thank K.L.J. Prather for providing plasmid pHHD-*ycdW-aceA-aceK*.

Competing Financial Interests

B.P., M.D.M., and G.S. are co-inventors on patent application PCT/US2013/027351: “Engineering microbes and metabolic pathways for the production of ethylene glycol.” B.P. and G.S. are co-founders of Advanced Renewable Technologies which has licensed the reported technology; Advanced Renewable Technologies did not play a role in the design or conduct of the work presented here.

References

- Ajikumar, P.K. *et al.* Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70-74 (2010).
- Alkim, C. *et al.* Optimization of ethylene glycol production from (D)-xylose via a synthetic pathway implemented in *Escherichia coli*. *Microb Cell Fact.* **14**, 127 (2015).
- Badía, J. *et al.* L-Lyxose metabolism employs the L-rhamnose pathway in mutant cells of *Escherichia coli* adapted to grow on L-lyxose. *J. Bacteriol.* **173**, 5144-5150 (1991).
- Baldomà, L., Aguilar, J. Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: aerobic-anaerobic regulation of L-lactaldehyde dissimilation. *J. Bacteriol.* **170**, 416-421 (1988).
- Caballero, E., Baldomà, L., Ros, J., Boronat, A., Aguilar, J. Identification of lactaldehyde dehydrogenase and glycolaldehyde dehydrogenase as functions of the same protein in *Escherichia coli*. *J. Biol. Chem.* **258**, 7788-7792 (1983).

479 Cam, Y. *et al.* Engineering of a synthetic metabolic pathway for the assimilation of (D)-xylose
 480 into value-added chemicals. *ACS Synth Biol.* Advance online publication. doi:
 481 10.1021/acssynbio.5b00103 (2015).

482 Cabiscol, E. *et al.* Oxygen regulation of L-1,2-propanediol oxidoreductase activity in
 483 *Escherichia coli*. *J. Bacteriol.* **172**, 5514-5515 (1990).

484 Cesário, M.T. *et al.* Enhanced bioproduction of poly-3-hydroxybutyrate from wheat straw
 485 lignocellulosic hydrolysates. *N. Biotechnol.* **31**, 104-113 (2014).

486 Chen, Z., Huang, J., Wu, Y., Liu, D. Metabolic engineering of *Corynebacterium glutamicum* for
 487 the de novo production of ethylene glycol from glucose. *Metab. Eng.* Advance online publication.
 488 doi: 10.1016/j.ymben.2015.10.013 (2015).

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

491 Deng, Y., Mao, Y., Zhang, X.J. Metabolic engineering of *E. coli* for efficient production of
 492 glycolic acid from glucose. *Biochem. Eng. J.* **103**, 256-262 (2015).

493 Desai, T.A., Rao, C.V. Regulation of arabinose and xylose metabolism in *Escherichia coli*. *Appl.*
 494 *Environ. Microbiol.* **76**, 1524-1532 (2010).

495 Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat.*
 496 *Methods* **6**, 343-345 (2009).

497 Gibson, D.G., Smith, H.O., Hutchison, C.A., 3rd, Venter, J.C., Merryman, C. Chemical synthesis
 498 of the mouse mitochondrial genome. *Nat. Methods* **7**, 901-903 (2010).

499 Hartmanis, M.G.N., Stadtman, T.C. Diol metabolism and diol dehydratase in *Clostridium*
 500 *glycolicum*. *Arch. Biochem. Biophys.* **245**, 144-152 (1986).

501 Ishida, Y., Kamiya, T., Itoh, H., Kimura, Y., Izumori, K. Cloning and characterization of the D-
502 tagatose 3-epimerase gene from *Pseudomonas cichorii* ST-24. *J. Ferment. Bioeng.* **83**, 529-534
503 (1997).

504 Itoh, H. *et al.* Purification and characterization of D-tagatose 3-epimerase from *Pseudomonas* sp.
505 ST-24. *Biosci. Biotechnol. Biochem.* **58**, 2168-2171 (1994).

506 Izumori, K., Khan, A.R., Okaya, H., Tsumura, T. A new enzyme, D-ketohexose 3-epimerase,
507 from *Pseudomonas* sp. ST-24. *Biosci. Biotechnol. Biochem.* **57**, 1037-1039 (1993).

508 Ji, N. *et al.* Catalytic conversion of cellulose into ethylene glycol over supported carbide
509 catalysts. *Catal. Today* **147**, 77-85 (2009).

510 Kallio, P., Pásztor, A., Thiel, K., Akhtar, M.K., Jones, P.R. An engineered pathway for the
511 biosynthesis of renewable propane. *Nat. Commun.* **5**, 4731 (2014).

512 Kim, S.R., Park, Y.C., Jin, Y.S., Seo, J.H. Strain engineering of *Saccharomyces cerevisiae* for
513 enhanced xylose metabolism. *Biotechnol. Adv.* **31**, 851-861 (2013).

514 Koivistoinen, O.M. *et al.* Glycolic acid production in the engineered yeasts *Saccharomyces*
515 *cerevisiae* and *Kluyveromyces lactis*. *Microb. Cell Fact.* **12**, 82 (2013).

516 Kovach, M.E. *et al.* Four new derivatives of the broad-host-range cloning vector pBBR1MCS,
517 carrying different antibiotic-resistance cassettes. *Gene* **166**, 175-176 (1995).

518 Kurosawa, K., Wewetzer, S.J., Sinskey, A.J. Engineering xylose metabolism in triacylglycerol-
519 producing *Rhodococcus opacus* for lignocellulosic fuel production. *Biotechnol. Biofuels.* **6**, 134
520 (2013).

521 LeBlanc, D.J., Mortlock, R.P. Metabolism of D-arabinose: a new pathway in *Escherichia coli*. *J.*
522 *Bacteriol.* **106**, 90-96 (1971).

523 Lee, J.W. *et al.* Systems metabolic engineering of microorganisms for natural and non-natural
 524 chemicals. *Nat. Chem. Biol.* **8**, 536-546 (2012).

525 Liu, H. *et al.* Biosynthesis of ethylene glycol in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **97**,
 526 3409-3417 (2013).

527 Martin, C.H. *et al.* A platform pathway for production of 3-hydroxyacids provides a biosynthetic
 528 route to 3-hydroxy- γ -butyrolactone. *Nat. Commun.* **4**, 1933 (2013).

529 Nieves, L.M., Panyon, L.A., Wang, X. Engineering sugar utilization and microbial tolerance
 530 toward lignocellulose conversion. *Front. Bioeng. Biotechnol.* **3**, 17 (2015).

531 Olson, D.G., McBride, J.E., Shaw, A.J., Lynd, L.R. Recent progress in consolidated
 532 bioprocessing. *Curr. Opin. Biotechnol.* **23**, 396-405 (2012).

533 Pang, J., Zheng, M., Wang, A., Zhang, T. Catalytic hydrogenation of corn stalk to ethylene
 534 glycol and 1,2-propylene glycol. *Ind. Eng. Chem. Res.* **50**, 6601-6608 (2011).

535 Pellicer, M.T., Badía J., Aguilar J., Baldomà, L. *glc* locus of *Escherichia coli*: characterization of
 536 genes encoding the subunits of glycolate oxidase and the *glc* regulator protein. *J. Bacteriol.* **178**,
 537 2051-2059 (1996).

538 Pereira, B., *et al.* Engineering a novel biosynthetic pathway in *Escherichia coli* for production of
 539 renewable ethylene glycol. *Biotechnol. Bioeng.* Advance online publication. doi:
 540 10.1002/bit.25717 (2015).

541 Sun, J., Liu, H. Selective hydrogenolysis of biomass-derived xylitol to ethylene glycol and
 542 propylene glycol on supported Ru catalysts. *Green Chem.* **13**, 135-142 (2011).

543 Wasylenko, T.M., Stephanopoulos, G. Metabolomic and ^{13}C -metabolic flux analysis of a xylose-
 544 consuming *Saccharomyces cerevisiae* strain expressing xylose isomerase. *Biotechnol. Bioeng.*
 545 **112**, 470-483 (2015).

546 Xu, P. *et al.* Modular optimization of multi-gene pathways for fatty acids production in *E. coli*.
547 *Nat. Commun.* **4**, 1409 (2013).
548 Yim, H. *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol.
549 *Nat. Chem. Biol.* **7**, 445-452 (2011).
550 Zahoor, A., Otten, A., Wendisch, V.F. Metabolic engineering of *Corynebacterium glutamicum*
551 for glycolate production. *J. Biotechnol.* **192**, 366-375 (2014).