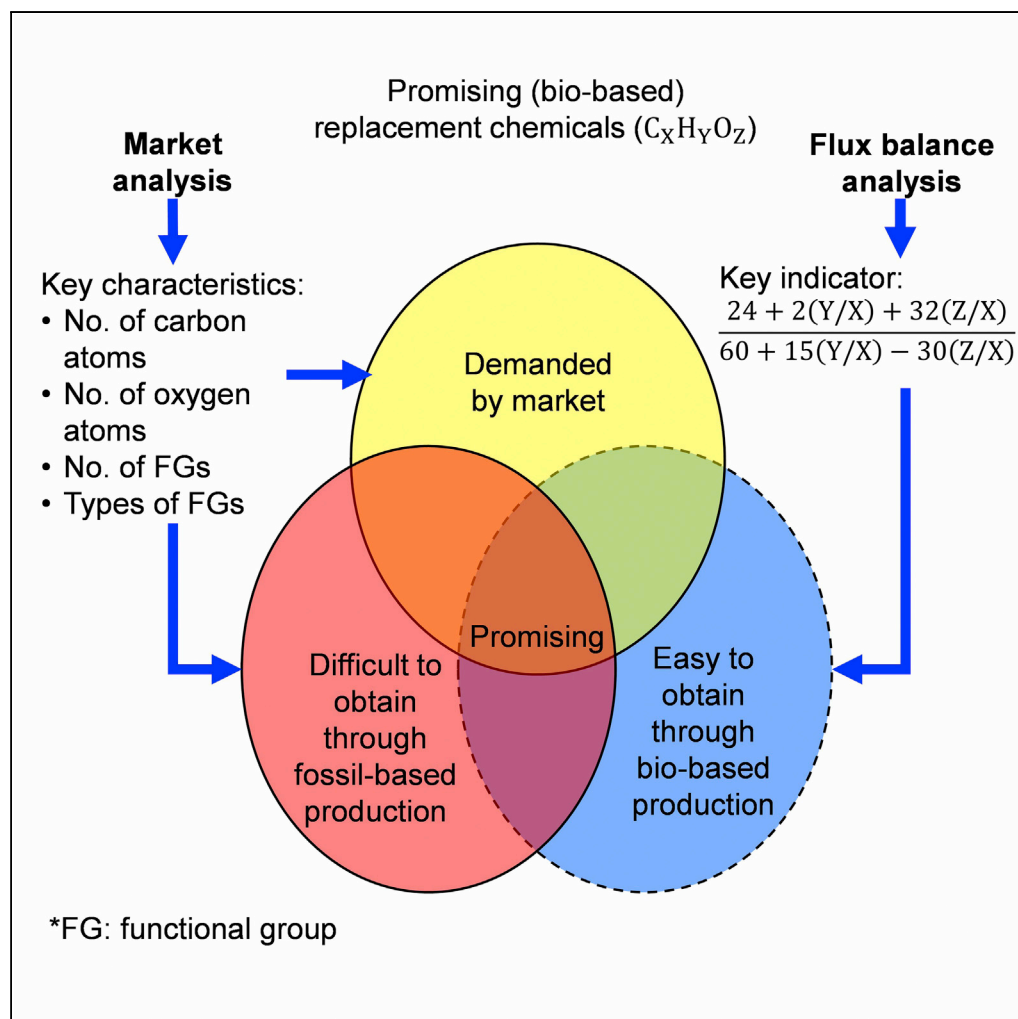


## Article

## Identifying the Characteristics of Promising Renewable Replacement Chemicals



Wenzhao Wu,  
Christos T.  
Maravelias

christos.maravelias@wisc.edu

## HIGHLIGHTS

Identified molecular characteristics of potentially promising replacement chemicals

Key characteristics: number of carbon and oxygen atoms and functional groups

Proposed metric for evaluating bio-based production efficiency of various chemicals

Wu & Maravelias, iScience 15, 136–146  
May 31, 2019 © 2019 The Author(s).  
<https://doi.org/10.1016/j.isci.2019.04.012>

## Article

# Identifying the Characteristics of Promising Renewable Replacement Chemicals

Wenzhao Wu<sup>1,2</sup> and Christos T. Maravelias<sup>1,3,\*</sup>**SUMMARY**

Although several “renewable” strategies have been recently proposed to produce high-volume as well as new (replacement) chemicals, the identification of good targets for such strategies remains challenging. Such chemicals that are expensive to obtain today from fossil fuel feedstocks would have an advantage if produced cheaply using alternative methods in the future. In this work we identify the characteristics of such potentially promising replacement chemicals. We also identify the characteristic of promising bio-based replacement chemicals that are relatively easy to obtain through bio-conversions. This work provides insights into the development of renewable chemicals to support a sustainable economy.

**INTRODUCTION**

Most fuels and organic chemicals are nowadays produced from fossil fuel feedstocks such as petroleum. In such processes, substantial amount of carbon originally stored underground is ultimately emitted into the atmosphere in the form of CO<sub>2</sub>, contributing to global warming. Therefore alternative production strategies, especially strategies employing biomass as feedstock, have been the focus of many studies (Ataman and Hatzimanikatis, 2017; Becker and Wittmann, 2015; Bordbar et al., 2014; Bornscheuer and Nielsen, 2015; Chundawat et al., 2011; Corma et al., 2007; Golden et al., 2015; Isikgor and Becer, 2015; Jong et al., 2012; Jordan et al., 2012; Julleson et al., 2015; Nielsen et al., 2014; Opgenorth et al., 2016; Woolston et al., 2013; Yadav et al., 2012). Biomass feedstocks (such as sugarcane and agricultural residues) capture CO<sub>2</sub> from the atmosphere through photosynthesis, and the carbon in the captured CO<sub>2</sub> is then transferred to the products during bio-production and finally returned back to the atmosphere after the products are used or disposed. Thus the overall process is often considered to be carbon neutral. Notably, some studies consider the overall process as carbon negative, because the carbon in the atmospheric CO<sub>2</sub> is assumed to be fixated in the form of chemicals (Brandao and Levasseur, 2011). Past studies have identified a number of existing compounds as promising targets for bio-production (Becker et al., 2015; Biddy et al., 2016; Bomtempo et al., 2017; Golden et al., 2015; Holladay et al., 2007; Jang et al., 2012; Matsumoto et al., 2013; Moncada et al., 2015; Shanks and Keeling, 2017; Straathof and Bampouli, 2017; Werpy and Petersen, 2004; Wu et al., 2018), because they have, generally speaking, large market demand, high price, and high potential to become valuable chemicals through further conversion. However, some novel renewable chemicals may currently have little or no demand but could potentially replace existing chemicals because of environmental or economic advantages, upon successful deployment in the future. We call these new chemicals *replacement chemicals*, and the existing chemicals that can be replaced *replaceable chemicals*. One goal of the research community therefore is to identify such promising replacement chemicals. However, directly evaluating the potential of such chemicals is a highly challenging task because they (1) have almost no market and production data and (2) have not even been discovered yet, and thus have unknown market prospect. Furthermore, it is usually a combination of multiple physical properties that makes a chemical more or less desirable. Properties typically considered include density, boiling and melting points, heat of vaporization, partition coefficients, viscosity, surface tension, and thermal conductivity and solubility (Gani, 2005; Shanks and Keeling, 2017). However, identifying the above combination of properties is difficult because (1) even if they were known, the total number of combinations that need to be studied for all major chemicals is combinatorial in nature and (2) many chemicals have multiple uses, and thus different sets of properties make them desirable.

Therefore, instead, in this article we aim to identify the molecular characteristics of potentially promising replacement chemicals as a surrogate of the original question. Our analysis is based on the following assumptions. First, molecular characteristics that are in high demand today will continue to be in high demand in the near future, because current commodity chemicals satisfy basic demands that have evolved

<sup>1</sup>Department of Chemical and Biological Engineering and DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706, USA

<sup>2</sup>Corporate Strategic Research, ExxonMobil Research and Engineering, 1545 US-22, Annandale, NJ 08801, USA

<sup>3</sup>Lead Contact

\*Correspondence: [christos.maravelias@wisc.edu](mailto:christos.maravelias@wisc.edu)  
<https://doi.org/10.1016/j.isci.2019.04.012>



over many years and are not expected to change at rapid pace. Second, if certain characteristics are difficult to obtain today, then chemicals with such characteristics, but produced via alternative technologies, are expected to have a comparative advantage. Third, the physical properties of a chemical are functions of molecular characteristics, such as chain length and functional groups. Therefore insights from the current analysis are applicable to a wide range of compounds, although we note that, in some cases, a specific compound with desirable physical properties may have molecular characteristics that lead to its classification as undesirable.

If the assumptions above do not hold for a specific candidate replacement chemical, then a more detailed assessment would be required.

## RESULTS

In general, potentially promising replacement chemicals should be (1) demanded by the market and (2) difficult to obtain using fossil-based strategies (i.e., have high production cost), thus favoring replacement. Furthermore, in the context of bio-production, replacement chemicals should be relatively easy to obtain from biomass.

We assume that if a characteristic in a replaceable chemical is difficult to obtain through fossil-based production today, then the same characteristic will also be difficult to obtain using the same method in the future. Therefore by studying the characteristics of *replaceable* chemicals with high demand and high production cost today, we can identify the characteristics of promising *replacement* chemicals (see Table S1 in the Supplemental Information).

Specifically, we analyze the following five molecular characteristics: (1) number of carbon atoms (abbreviated as #C hereafter), (2) number of oxygen atoms (#O), (3) number of functional groups (#FG), (4) number of distinct functional groups (#DFG; e.g., 2-hydroxyl FGs are counted as 1 DFG), and (5) existence of specific functional groups (FG), such as alkenyl, hydroxyl, and phenyl groups, as well as their combinations.

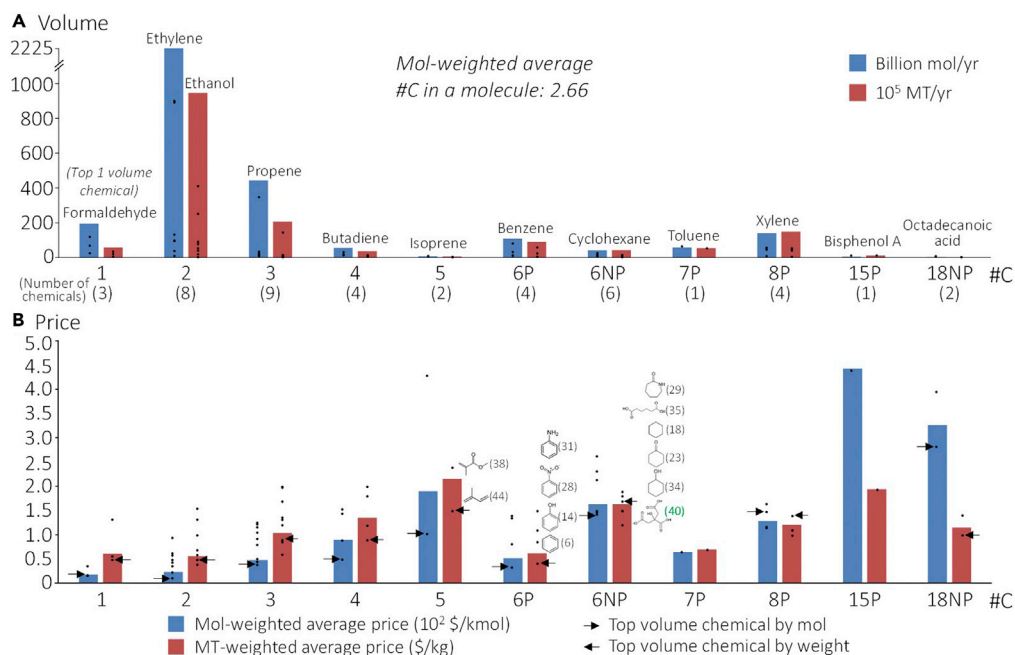
### Number of Carbon Atoms (#C)

The volume and price data for chemicals with different #C are shown in Figure 1, where “P” represents phenyl groups, e.g., 6P denotes a C<sub>6</sub> chemical with phenyl groups, and “NP” denotes the non-existence of phenyl groups.

The key insights from the analysis of Figure 1A are summarized as follows: (1) C<sub>2</sub> and C<sub>3</sub> chemicals have the highest demand, followed by C<sub>1</sub>, C<sub>8P</sub>, C<sub>6P</sub>, etc., and thus replacement chemicals with large #C (e.g., #C > 18) will likely have small demand; (2) on a mol-weighted basis, there are 2.66 carbon atoms in a molecule on average; (3) for #C ≥ 6, chemicals with phenyl groups have much larger demand than those without.

To facilitate the understanding of the insights from Figure 1B, we note the following for the industrial production methods of chemicals with different #C: (1) C<sub>1</sub> chemicals are produced from natural gas, which is relatively cheap; (2) C<sub>2</sub> and C<sub>3</sub> chemicals are produced from either natural gas or petroleum; (3) C<sub>4</sub> chemicals are mainly produced from petroleum; (4) C<sub>5</sub> chemicals are primarily by-products of the production of other chemicals from petroleum and usually have higher cost; (5) C<sub>6P</sub> chemicals are produced from benzene (or its derivatives), which is produced mainly through catalytic reforming of naphtha; (6) C<sub>6NP</sub> chemicals are mostly produced from cyclohexane or its derivatives produced from benzene, with the exception of citric acid, which is produced through microbial conversion of glucose; (7) C<sub>7P</sub>, C<sub>8P</sub>, and C<sub>15P</sub> chemicals are produced mainly through alkylation of benzene (or its derivatives); and (8) C<sub>18NP</sub> chemicals (octadecanoic acid and oleic acid) are produced from vegetable and animal oils.

There are three key insights from the analysis of Figure 1B. First, the price of #C (a surrogate for the difficulty of obtaining a specific attribute from fossil fuel feedstocks), on a mol-weighted average basis, is sequenced as 1 < 2 < 3 < 6P < 7P < 4 < 8P < 6NP < 5 < 15P. The sequence on the metric ton (MT)-weighted average basis is 1 < 6P < 7P < 2 < 3 < 8P < 4 < 6NP < 15P < 5, where chemicals with larger #C (e.g., 6P) have lower prices than on the mol-weighted average basis because they have larger molecular weights. Second, although a larger #C leads to a higher price in general, C<sub>4</sub> and C<sub>5</sub> chemicals are more expensive than



**Figure 1. Results for #C**

(A) Volume versus #C; blue (red) bars represent the total volume in billion mol/year ( $10^5$  MT/year) of chemicals with the corresponding #C; the number of chemicals with the corresponding #C are labeled at the bottom; the highest volume chemical for each #C is labeled above the two bars (only one is labeled if the highest volume chemical is the same for both bars); each black dot represents a chemical.

(B) Price versus #C; blue (red) bars represent the mol-weighted average price in  $10^2$  \$/kmol (MT-weighted average price in \$/kg) of chemicals with the corresponding #C; arrows represent top volume chemicals; molecular structures of chemicals with #C = 5, 6P, and 6NP are shown, ordered consistently with the MT-weighted average price (\$/kg); numbers in parentheses denote the chemical entry number in Table S1 and the supplementary excel sheet, and chemicals that are currently naturally produced are marked bold in green.

C6P chemicals (produced from benzene) because benzene cost is relatively low. Third, C6NP chemicals are more expensive than C6P chemicals because the former are produced from cyclohexane, which is produced from benzene, thus more expensive than C6P chemicals produced directly from benzene.

### Number of Oxygen Atoms(#O)

The volume and price data for the analysis of different #O are shown in Figure 2. The key insights from the analysis of Figure 2A are the following: (1) chemicals with larger #O have lower demand, and there is little demand for chemicals with more than four oxygen atoms except citric acid (O7), which is bio-produced today and (2) on a mol-weighted basis, there are 0.51 oxygen atoms in a molecule on average.

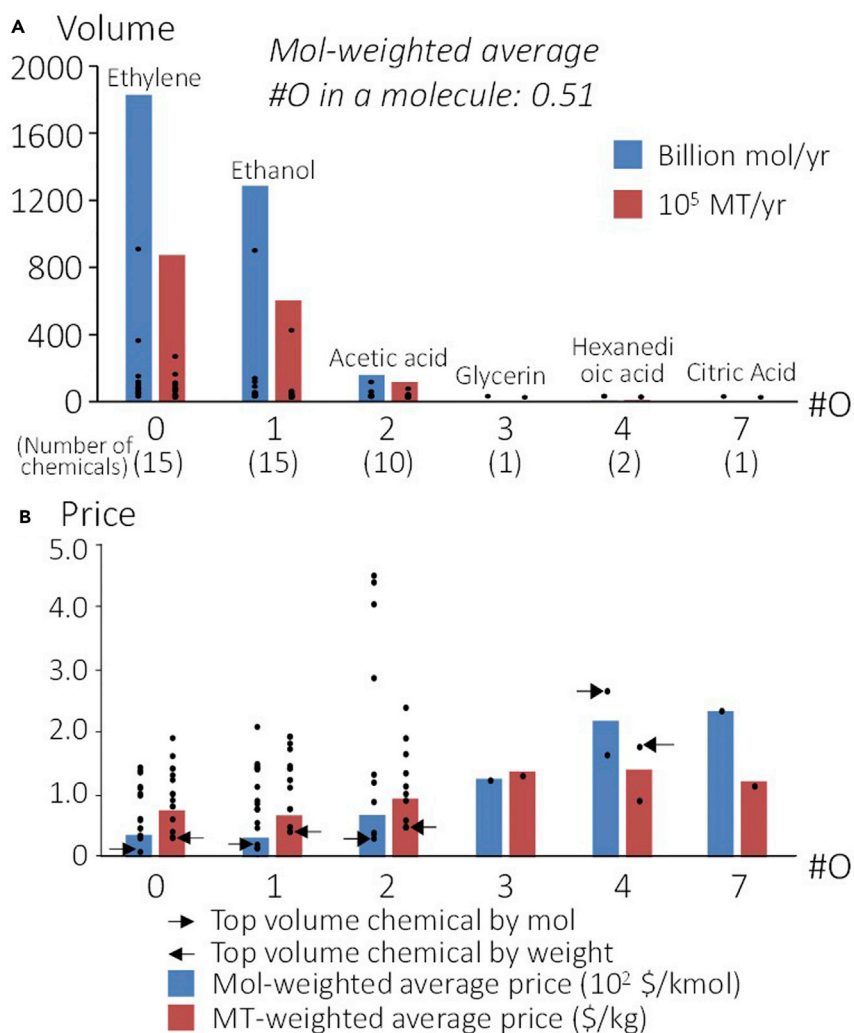
The key insight from the analysis of Figure 2B is that chemicals with larger #O have, in general, higher prices, which means that they can be potential replaceable chemicals. Presumably, this is because the oxygen content of fossil fuels is low, making the production of highly oxygenated chemicals more expensive.

### Number of (Distinct) Functional Groups (#FG and #DFG)

The volume and price data for the analysis of different #FG and #DFG are shown in Figure 3, and the key insights are the following: (1) chemicals with more FGs or DFGs have lower demand, and there is little demand for chemicals with more than 4 FGs or 2 DFGs; (2) on a mol-weighted basis, there are 1.14 FGs or 1.09 DFGs in a molecule on average; and (3) chemicals with more FGs or DFGs have higher prices.

### Specific Functional Groups (FG)

The volume and price data for the analysis of different specific FGs are shown in Figure 4. Here we draw insights based on attributes with at least three data points in the mono-DFG region (marked bold in



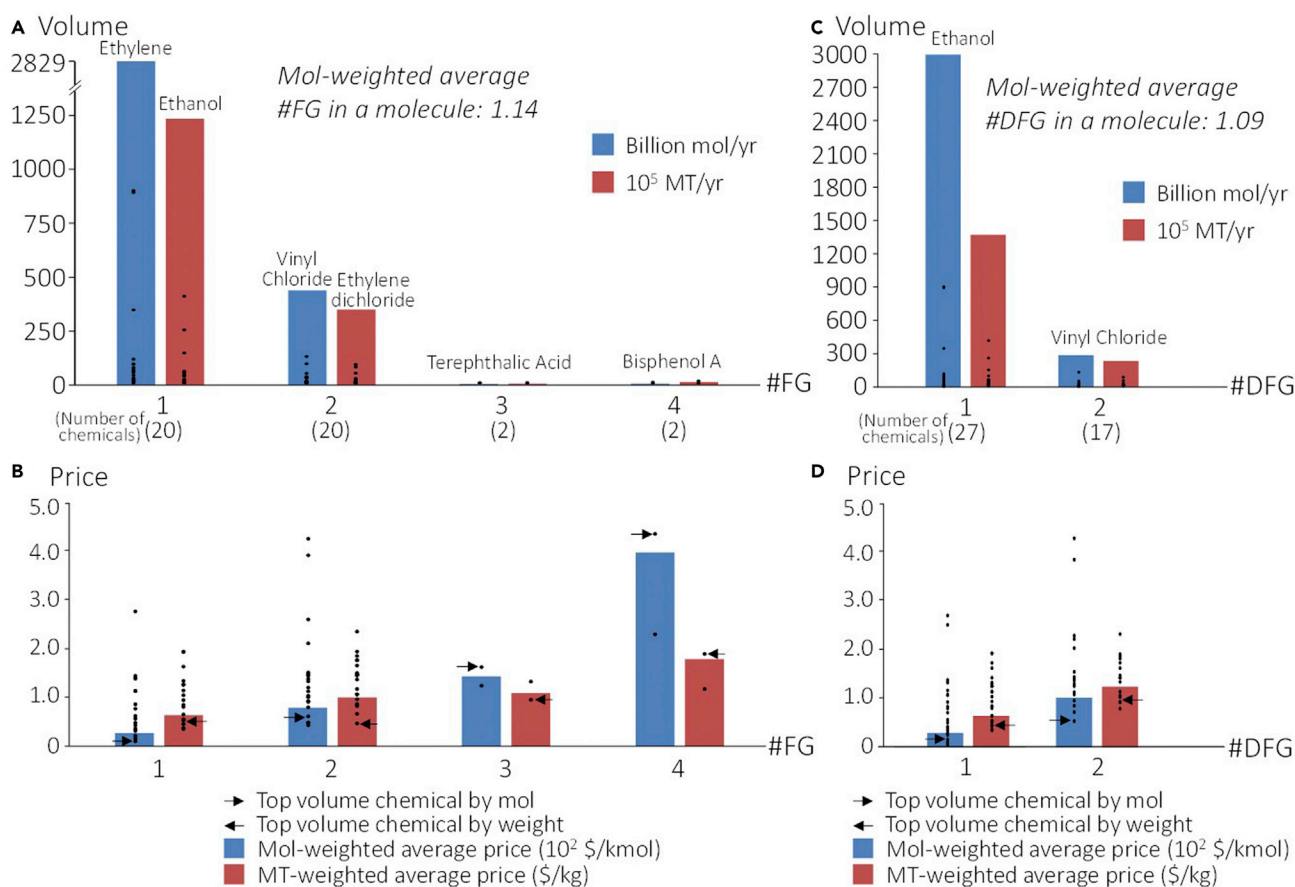
**Figure 2. Results for #O**

(A) Volume versus #O; blue (red) bars represent the total volume in billion mol/year ( $10^5$  MT/year) of chemicals with the corresponding #O; the number of chemicals with the corresponding #O are labeled at the bottom; the highest volume chemical for each #O is labeled above the two bars (each black dot represents a chemical).

(B) Price versus #O; blue (red) bars represent the mol-weighted average price in  $10^2$  \$/kmol (MT-weighted average price in \$/kg) of chemicals with the corresponding #O; arrows represent top volume chemicals.

red) and the combination of these attributes in the bi-DFG region (marked bold in blue). The key insights from the analysis of Figure 4A are the following: (1) the most demanded FGs are alkenyl (=), hydroxyl (OH), phenyl (Ph), and carboxyl (COOH) groups and (2) demand for all combinations of these four FGs exists except alkenyl (=) combined with hydroxyl (OH).

The key insight from the analysis of Figure 4B is that the mol-weighted average price of FGs is sequenced as follows: alkenyl (=) < hydroxyl (OH) < carboxyl (COOH) < phenyl (Ph), which is generally consistent with the price of the combinations of these FGs in the bi-DFG region (with exceptions that can be explained by the #C and #O). The alkenyl (=) < hydroxyl (OH) < carboxyl (COOH) sequence can be explained by the increased level of oxidation. Thus chemicals with carboxyl groups could be potential targets for replacement. Note that chemicals with phenyl groups have high prices partly because they also have large #C. Figures that separate the specific FGs and #C into two dimensions (with both  $10^2$  \$/kmol and \$/kg as units) can be found in the supplementary Excel file.



**Figure 3. Results for #FG and #DFG**

(A) Volume versus #FG; blue (red) bars represent the total volume in billion mol/year (10<sup>5</sup> MT/year) of chemicals with the corresponding #FG; the number of chemicals with the corresponding #FG are labeled at the bottom; the highest volume chemical for each #FG is labeled above the two bars (only one is labeled if the highest volume chemical is the same for both bars).

(B) Price versus #FG; blue (red) bars represent the mol-weighted average price in 10<sup>2</sup> \$/kmol (MT-weighted average price in \$/kg) of chemicals with the corresponding #FG; arrows represent top volume chemicals.

(C) Volume versus #DFG; blue (red) bars represent the total volume in billion mol/year (10<sup>5</sup> MT/year) of chemicals with the corresponding #DFG; the number of chemicals with the corresponding #DFG are labeled at the bottom; the highest volume chemical for each #DFG is labeled above the two bars (only one is labeled if the highest volume chemical is the same for both bars).

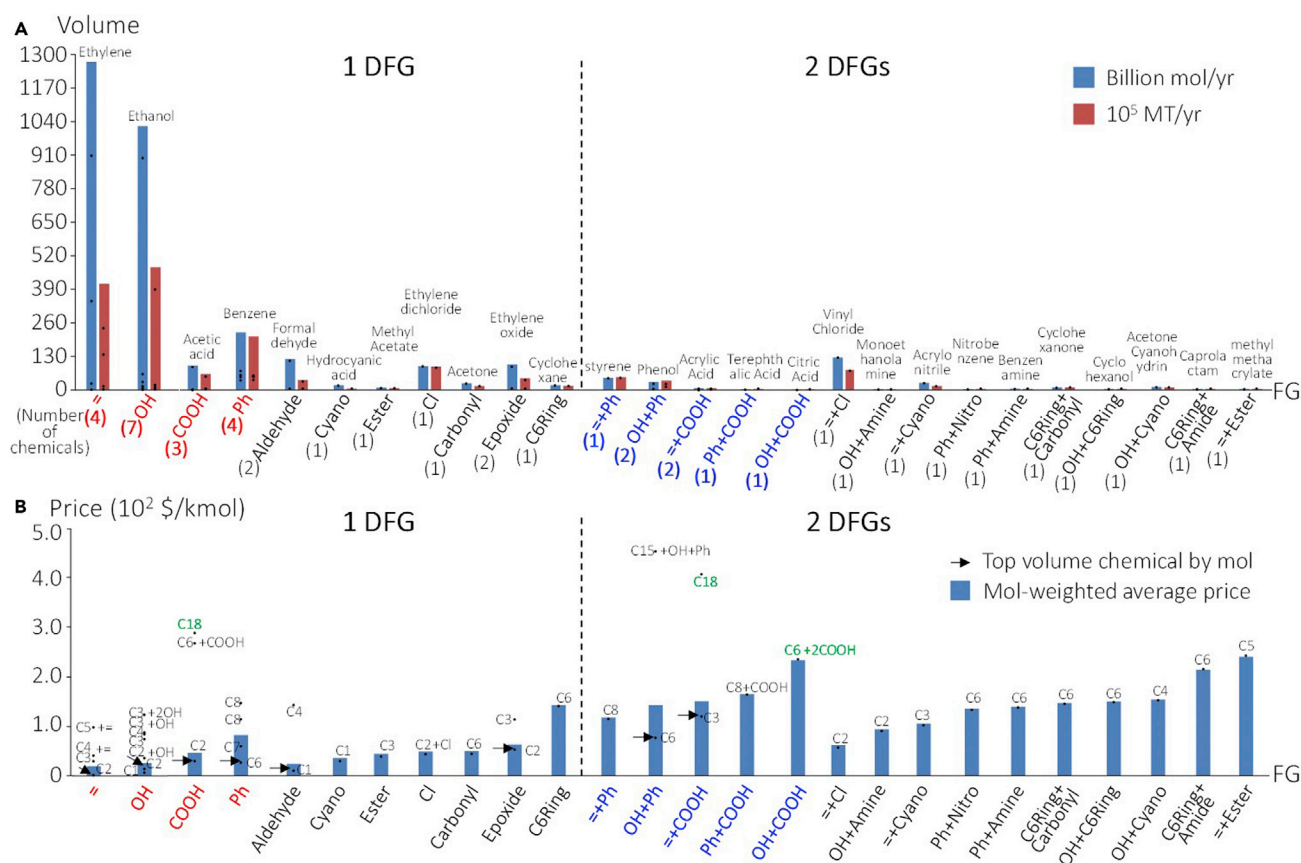
(D) Price versus #DFG; blue (red) bars represent the mol-weighted average price in 10<sup>2</sup> \$/kmol (MT-weighted average price in \$/kg) of chemicals with the corresponding #DFG; arrows represent top volume chemicals.

Each black dot represents a chemical.

## DISCUSSION

### Suitability of Bio-production

In this section, we discuss the characteristics of promising replacement chemicals that can be produced efficiently using microbial conversion of glucose, which can in turn be derived from biomass. We present the microbial conversion as an example production strategy. A similar analysis can be performed for alternative strategies such as the chemical transformation of natural oils. Also, importantly, we do not consider renewable strategies using CO<sub>2</sub> as feedstock (such as solar conversion), because they are based on C1 chemistry and thus will not have a “chemistry” advantage over existing natural-gas-based strategies. In other words, renewable strategies other than those based on alternative feedstocks do not likely lead to new (replacement) chemicals; they will simply lead to the production of existing high-volume chemicals. In the study by Wu et al. (Wu et al., 2018), the yields of chemicals produced through microbial conversion of glucose from *E. coli* and *S. cerevisiae* are estimated using flux balance analysis (FBA) (Papoutsakis, 1984), at both maximum yield (g product/g glucose fed) and maximum productivity (g product · L medium<sup>-1</sup> · day<sup>-1</sup>) conditions. It was also found that the total production cost (including



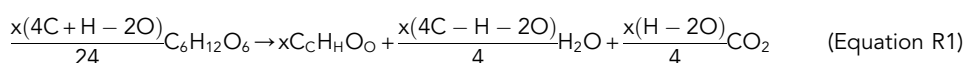
**Figure 4. Results for specific FGs**

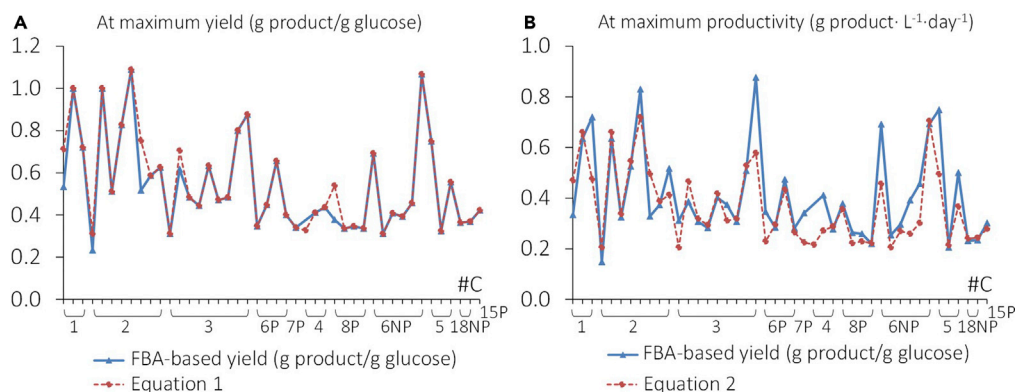
(A) Volume versus FGs; blue (red) bars represent the total volume in billion mol/year (10<sup>5</sup> MT/year) of chemicals with the corresponding FG (or a combination of FGs); the number of chemicals with the corresponding FG are labeled at the bottom; the highest volume chemical for each FG is labeled above the two bars.

(B) Price versus specific FGs; blue bars represent the mol-weighted average price in 10<sup>2</sup> \$/kmol of chemicals with the corresponding FG; arrows represent top volume chemicals; #C is marked next to each chemical, and if a chemical has more FGs than labeled on the x axis, then the additional FGs are also marked; chemicals that are currently bio-produced are marked bold in green. Each black dot represents a chemical; = denotes alkenyl; Cl denotes chloro; Ph denotes phenyl; COOH denotes carboxyl; OH denotes hydroxyl; C6Ring denotes non-aromatic ring with six carbon atoms.

both bio-conversion and downstream separation) is almost inversely proportional to yield, because higher yield leads to lower feedstock consumption (thus lower feedstock cost) and higher product titer (thus lower separation cost), which were identified to be the major cost drivers. Therefore chemicals that can be produced at high yield are more promising. Note that titer is, in fact, determined by both yield and sugar concentration in the feed, but the sugar concentration is assumed to be ~22 wt % according to the National Renewable Energy Laboratory (NREL) study (Humbird et al., 2011) for the cost analysis (Wu et al., 2018), and thus titer is proportional to yield. In detailed analyses, the sugar concentration can be considered a variable.

Based on FBA calculations (Wu et al., 2018), Figure 5A shows the maximum yield (representing minimum bio-production cost) and Figure 5B the maximum productivity, respectively, of all chemicals studied herein. The specific chemical corresponding to each #C is shown in Table 1. It can be seen that there is no clear correlation between #C and yield (thus bio-production cost). We further compare the FBA-based yield with the stoichiometric maximum yield (g product/g glucose), which can be expressed as  $\frac{12C+H+16O}{30C} \times \frac{4C}{4C+H-2O}$  in most cases, calculated from reaction (Equation R1):





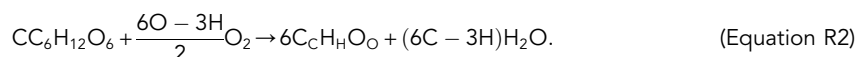
**Figure 5. Microbial Conversion Yield (g product/g glucose) of Chemicals**

(A) Maximum yield versus #C; blue curve represents the maximum yield estimated using flux balance analysis (FBA); red dashed curve represents the stoichiometric maximum yield using glucose as the carbon source, where C, H, and O represent the number of carbon, hydrogen, and oxygen atoms in a molecule, respectively, and MW (g/mol) is the molecular weight of the chemical.

(B) Yield at maximum productivity (g product · L medium<sup>-1</sup> · day<sup>-1</sup>) versus #C; blue curve represents the yield at maximum productivity estimated using FBA; red dashed curve represents the estimates based on the stoichiometric yield. The #C is ordered such that the mol-weighted average prices of chemicals increase from left to right.

where C, H, and O represent the number of carbon, hydrogen, and oxygen atoms in a molecule of the chemical;  $\frac{4C}{4C+H-2O}$  represents the molar carbon efficiency (mol of carbon in product/mol of carbon in glucose); and  $12C + H + 16O$  represents the molecular weight of the chemical. The specific derivation can be found in the [Supplemental Information](#).

Note that for 2 of the 44 chemicals (citric acid and terephthalic acid), their molar carbon yields, if calculated using  $\frac{4C}{4C+H-2O}$  based on (Equation R1), will be greater than 100% because CO<sub>2</sub> will be considered as feedstock to balance (Equation R1). Instead, we consider Equation R2 for such chemicals, where the yield can be expressed as  $\frac{12C+H+16O}{30C}$ .



Therefore, the stoichiometric maximum yield can be expressed as

$$\text{Stoichiometric maximum yield} = \frac{12C+H+16O}{30C} \times \text{Min}\left(1, \frac{4C}{4C+H-2O}\right) \quad (\text{Equation 1})$$

which is close to the maximum yield calculated using FBA, as shown in [Figure 5A](#). Also, at maximum productivity, the approximate yield can be expressed as follows (see [Figure 5B](#)):

$$\text{Approximate yield at maximum productivity} = 66\% \times \frac{12C+H+16O}{30C} \times \text{Min}\left(1, \frac{4C}{4C+H-2O}\right) \quad (\text{Equation 2})$$

Therefore in most cases, the stoichiometric yield (g/g),

$$\frac{12C+H+16O}{30C} \times \frac{4C}{4C+H-2O} = \frac{24+2(H/C)+32(O/C)}{60+15(H/C)-30(O/C)},$$

can be regarded as an indicator for the suitability of bio-production. Thus chemicals that are highly oxidized (with high O/C ratio) and/or moderately hydrogenated (with low H/C ratio, e.g., chemicals with alkenyl or phenyl groups) are likely suitable targets.

As an example, note that citric acid (C<sub>6</sub>H<sub>8</sub>O<sub>7</sub>), with a high O/C ratio and low H/C ratio and indicator value of 1.42 (numbered 40 and marked green in [Figure 1B](#)), is currently produced through microbial



| #C | Chemical            | #C | Chemical            | #C   | Chemical            |
|----|---------------------|----|---------------------|------|---------------------|
| 1  | Methanol            | 3  | 1,2-Propanediol     | 8P   | Xylene              |
| 1  | Formaldehyde        | 3  | Acrylonitrile       | 8P   | Terephthalic acid   |
| 1  | Hydrocyanic acid    | 3  | Propylene oxide     | 6NP  | Cyclohexane         |
| 2  | Ethylene            | 3  | Acrylic acid        | 6NP  | Cyclohexanone       |
| 2  | Acetic acid         | 3  | Glycerin            | 6NP  | Cyclohexanol        |
| 2  | Ethanol             | 6P | Benzene             | 6NP  | Caprolactam         |
| 2  | Ethylene glycol     | 6P | Phenol              | 6NP  | Citric acid         |
| 2  | Ethylene dichloride | 6P | Nitrobenzene        | 6NP  | Hexanedioic acid    |
| 2  | Vinyl chloride      | 6P | Benzenamine         | 5    | Isoprene            |
| 2  | Ethylene oxide      | 7P | Toluene             | 5    | Methyl methacrylate |
| 2  | Monoethanolamine    | 4  | 1-Butanol           | 18NP | Octadecanoic acid   |
| 3  | Propene             | 4  | Butanal             | 18NP | Oleic acid          |
| 3  | Methyl acetate      | 4  | Acetone cyanohydrin | 15P  | Bisphenol A         |
| 3  | Acetone             | 8P | Ethyl benzene       |      |                     |
| 3  | 2-Propanol          | 8P | Styrene             |      |                     |

**Table 1. Chemicals Corresponding to the #C axis in Figure 5**

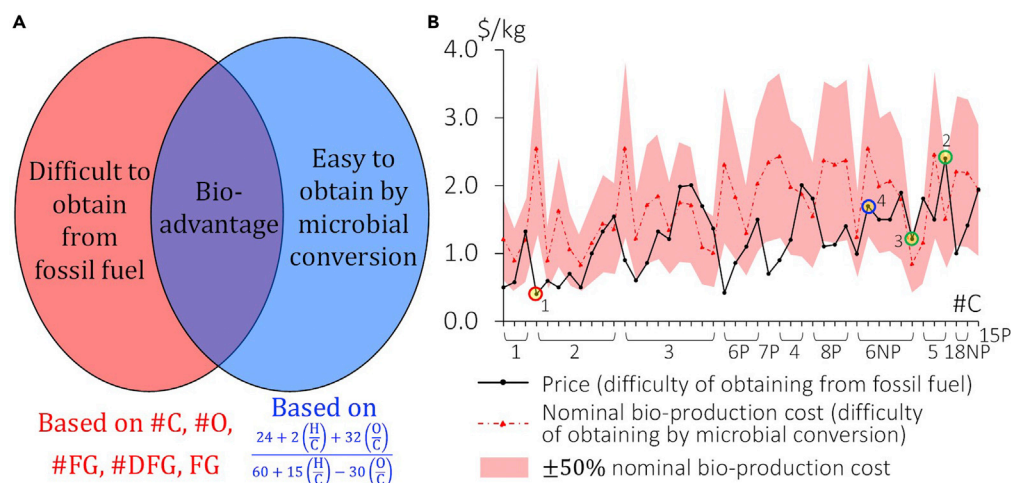
The order of entries is the same with the horizontal sequence of entries in the x axes in Figure 5.

conversion of glucose and is relatively cheap compared with other 6-NP chemicals with less complicated structures such as hexanedioic acid (C<sub>6</sub>H<sub>10</sub>O<sub>4</sub>, with an indicator value of 0.75; numbered 35 in Figure 1B). This suggests that  $\frac{24 + 2(H/C) + 32(O/C)}{60 + 15(H/C) - 30(O/C)}$  is a reasonable indicator for the suitability of bio-production.

### Identification of Promising Bio-based Replacement Chemicals

We present a conceptual analysis for the identification of potentially promising bio-based replacement chemicals in Figure 6A. A promising target should be difficult to obtain today from oil and natural gas (with the corresponding characteristics) and relatively easy to obtain through bio-production (highly oxidized and slightly hydrogenated). Potentially promising targets are at the intersection. In addition, the demand, which can also be estimated based on the characteristics discussed in the Results section, should be reasonably large.

To demonstrate how to compare fossil-based and bio-based production costs, we present the current market prices and bio-production cost estimates (including both bio-conversion and downstream separation) of the chemicals studied against #C as an example in Figure 6B. The specific chemical corresponding to each #C is shown in Table 1. Specifically, we calculate *nominal* bio-production costs based on the estimated maximum yields, assuming the chemicals are intracellular and soluble in water, which is a class of chemicals with intermediate separation costs (Wu et al., 2017, 2018; Yenkie et al., 2016). Further explanations and the specific formula used (Wu et al., 2018) can be found in the Supplemental Information. Note that the cost estimation presented here is based on the assumption that the maximum yield can be achieved, possibly through anaerobic fermentation, to minimize the number of reaction steps. If this is not possible, then the analysis can be repeated using different assumptions (e.g., experimentally achieved yield). Also note that the sugar feedstock can be obtained either from traditional sources such as sugarcane and sugar beets or from alternative sources such as cellulosic biomass. Thus the fluctuation in the biomass feedstock cost will also play a significant role in the economics. The red shaded band represents  $\pm 50\%$  of the nominal cost to account for uncertainty. There are few key observations. First, some chemicals have low market prices, with, typically, #C characteristic relatively easy to obtain from fossil fuel feedstocks and high



**Figure 6. Identification of Promising Bio-based Replacement Chemicals**

(A) Conceptual analysis framework: promising chemicals are at the intersection of the two sets.

(B) Current market prices and bio-production costs versus #C, where #C is ordered such that the mol-weighted average prices of chemicals increase from left to right. Specific equations used for the cost estimation can be found in the [Supplemental Information](#).

bio-production costs (with low  $\frac{24 + 2(H/C) + 32(O/C)}{60 + 15(H/C) - 30(O/C)}$  values), for example, Entry 1 marked by a red circle.

Such chemicals, especially if their prices are below the  $\pm 50\%$  band, are unlikely to be promising. Second, chemicals with prices above or close to the upper boundary of the band, which have either high prices and low bio-production costs (e.g., Entry 2 marked by a green circle) or relatively low prices and very low bio-production costs (e.g., Entry 3 marked by a green circle), are likely to be promising. Third, determining the potential of chemicals in the middle of the band (such as Entry 4 marked by a blue circle) needs further research.

Note that unlike existing chemicals, replacement chemicals do not have specific market prices. Therefore when evaluating a new replacement chemical, the average price of each attribute in [Figure 1B](#) can be used as an approximation of the actual price in [Figure 6B](#). The same analysis can be performed for the other characteristics, with the estimates based on all the characteristics indicating the overall prospect of the chemical. Finally, we note that fossil-based production and bio-production could occupy different market sectors.

## Conclusion

We developed a framework for the identification of molecular characteristics that make a chemical a promising replacement chemical. By studying the market volumes of major organic compounds, we identified desirable molecular characteristics; and by studying market prices, we identified molecular characteristics that are difficult to obtain from fossil fuel feedstocks today. Finally, we used a metric to quantify, at a high level, how expensive it is to produce a chemical biologically.

Specifically, replacement chemicals with one or more of the following characteristics are likely to have higher demand: (1) relatively few carbon atoms, oxygen atoms, or few functional groups and (2) contain alkenyl, hydroxyl, phenyl, or carboxyl groups. Chemicals with >18 carbon atoms, >4 oxygen atoms, >4 functional groups, or >2 distinct functional groups are likely to have small demand. Replacement chemicals with one or more of the following characteristics are likely to be expensive to obtain from fossil fuel feedstocks: (1) >6 carbon atoms but without phenyl groups, (2) 4–5 carbon atoms, and (3) carboxyl groups. Replacement chemicals that are suitable for bio-production typically have a large  $\frac{24 + 2(H/C) + 32(O/C)}{60 + 15(H/C) - 30(O/C)}$  ratio,

where C, H, and O represent the numbers of carbon, hydrogen, and oxygen atoms, respectively, in the molecular formula. A compound is likely to be a promising replacement chemical if (1) its characteristics have high demand and are difficult to obtain from fossil fuel feedstocks and (2) its bio-based production is expected to have low cost. Note that the analysis can be performed using different threshold values

for each criterion, thereby studying different scenarios. For example, if demand is not a major concern, then the characteristics corresponding to relatively lower demand may also be considered.

Finally, we note that the results in this work can be used to evaluate the sustainability implications of renewable replacement chemicals. For example, if all the fossil-based C3, C4, and C5 chemicals can be replaced by bio-derived counterparts, then ~22% of the fossil carbon in the US organic chemicals will be replaced by a renewable carbon source. However, using a renewable carbon source does not necessarily lead to lower overall greenhouse gas emissions (Montazeri et al., 2016), and thus strategically identifying the replacement chemicals with both positive economic and environmental prospects is essential.

### Limitations of the Study

First, although most chemicals with high prices appear to have low volumes, all the chemicals studied herein are in fact High-Production-Volume (HPV) chemicals (a chemical with the smallest volume considered has a volume of 200,000 MT/year). Second, we note that our goal is to identify general insights, and thus detailed analysis for specific chemicals or processes is deemed unnecessary due to the uncertainty in production strategies and the possibility of new technologies emerging. Such analysis is also highly challenging, if not impossible, because the specific chemicals that would become replacement chemicals are currently unknown. However, detailed analysis would become necessary when specific compounds and a detailed production process is identified. For example, additional considerations include process design for cost reduction of bio-conversion and separation, as well as microbial strain engineering to improve yield and productivity. Third, we have considered “platform” chemicals such as ethylene, which can be converted to polymer products. Although there is emerging research focusing on the direct utilization of bio-polymers from biomass components for the production of polymers such as carbon fibers (Ballinas-Casarrubias et al., 2016; Fang et al., 2017; Parsell et al., 2015), we do not directly study these strategies here. However, if polymers are the target replacement chemicals, then the polymerization process and the specific three-dimensional structure of the polymers will also be important. Fourth, we note that renewable chemicals today do not receive subsidies, as fuels do. However, because chemicals generally have higher prices than fuels, it can be envisioned that a set of carefully selected and effectively produced bio-based chemicals (given further basic science advances) may be less reliant on subsidies.

### METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2019.04.012>.

### ACKNOWLEDGMENTS

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409). Special thanks to Professor William Banholzer for fruitful discussions on production of chemicals.

### AUTHOR CONTRIBUTIONS

Conceptualization, Methodology, C.T.M. and W. W.; Software, Investigation, Writing – Original Draft, Visualization, W. W.; Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition, C.T.M.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 9, 2018

Revised: March 19, 2019

Accepted: April 8, 2019

Published: May 31, 2019

## REFERENCES

- Ataman, M., and Hatzimanikatis, V. (2017). lumpGEM: systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites. *PLoS Comput. Biol.* *13*, 1–21.
- Ballinas-Casarrubias, L., Camacho-Davila, A., Gutierrez-Méndez, N., Ramos-Sánchez, V.H., Chávez-Flores, D., Manjarrez-Nevárez, L., Zaragoza-Galan, G., and González-Sánchez, G. (2016). Biopolymers from Waste Biomass—Extraction, Modification and Ulterior Uses. In *Recent Advances in Biopolymers*. (IntechOpen), pp. 3–15. <https://www.intechopen.com/books/recent-advances-in-biopolymers/biopolymers-from-waste-biomass-extraction-modification-and-ulterior-uses>.
- Becker, J., and Wittmann, C. (2015). Advanced biotechnology: metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products. *Angew. Chem. Int. Ed.* *54*, 3328–3350.
- Becker, J., Lange, A., Fabarius, J., and Wittmann, C. (2015). Top value platform chemicals: bio-based production of organic acids. *Curr. Opin. Biotechnol.* *36*, 168–175.
- Biddy, M., Scarlata, C., and Kinchin, C. (2016). Chemicals from biomass: a market assessment of bioproducts with near-term potential. NREL. NREL/TP-5100-65509, <https://www.nrel.gov/docs/fy16osti/65509.pdf>.
- Bomtempo, J.-V., Chaves Alves, F., and de Almeida Oroski, F. (2017). Developing new platform chemicals: what is required for a new bio-based molecule to become a platform chemical in the bioeconomy? *Faraday Discuss.* *202*, 213–225.
- Bordbar, A., Monk, J.M., King, Z.A., and Palsson, B.O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* *15*, 107–120.
- Bornscheuer, U.T., and Nielsen, A.T. (2015). Editorial overview: chemical biotechnology: interdisciplinary concepts for modern biotechnological production of biochemicals and biofuels. *Curr. Opin. Biotechnol.* *35*, 133–134.
- Brandao, M., and Levasseur, A. (2011). Assessing temporary carbon storage in life cycle assessment and carbon footprinting outcomes of an expert workshop, [https://www.researchgate.net/publication/230674935\\_Assessing\\_Temporary\\_Carbon\\_Storage\\_in\\_Life\\_Cycle\\_Assessment\\_and\\_Carbon\\_Footprinting\\_Outcomes\\_of\\_an\\_expert\\_workshop](https://www.researchgate.net/publication/230674935_Assessing_Temporary_Carbon_Storage_in_Life_Cycle_Assessment_and_Carbon_Footprinting_Outcomes_of_an_expert_workshop).
- Chundawat, S.P.S., Beckham, G.T., Himmel, M.E., and Dale, B.E. (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* *2*, 121–145.
- Corma, A., Iborra, S., and Velty, A. (2007). Chemical routes for the transformation of biomass into chemicals. *Chem. Rev.* *107*, 2411–2502.
- Fang, W., Yang, S., Wang, X.L., Yuan, T.Q., and Sun, R.C. (2017). Manufacture and application of lignin-based carbon fibers (LCFs) and lignin-based carbon nanofibers (LCNFs). *Green. Chem.* *19*, 1794–1827.
- Gani, R. (2005). Integrated chemical product design: CAPE perspectives. *Comput. Aided Chem. Eng.* *20*, 21–30.
- Golden, J.S., Handfield, R.B., Daystar, J., and McConnell, T.E. (2015). An economic impact analysis of the U.S. biobased products industry: a report to the Congress of the United States of America. *127*, [https://www.biopreferred.gov/BPResources/files/EconomicReport\\_6\\_12\\_2015.pdf](https://www.biopreferred.gov/BPResources/files/EconomicReport_6_12_2015.pdf).
- Holladay, J.E., White, J.F., Bozell, J.J., and Johnson, D. (2007). Top value-added chemicals from biomass - volume II: results of screening for potential candidates from biorefinery lignin. Pacific Northwest Natl. Lab., Available: [https://www.pnnl.gov/main/publications/external/technical\\_reports/pnnl-16983.pdf](https://www.pnnl.gov/main/publications/external/technical_reports/pnnl-16983.pdf).
- Humbird, D., Davis, R., Tao, L., Kinchin, C., Hsu, D., Aden, A., Schoen, P., Lukas, J., Olthof, B., Worley, M., et al. (2011). Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: dilute-acid pretreatment and enzymatic hydrolysis of corn stover. NREL. <https://www.nrel.gov/docs/fy11osti/47764.pdf>.
- Isikgor, F.H., and Becer, C.R. (2015). Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polym. Chem.* *6*, 4497–4559.
- Jang, Y.S., Kim, B., Shin, J.H., Choi, Y.J., Choi, S., Song, C.W., Lee, J., Park, H.G., and Lee, S.Y. (2012). Bio-based production of C2-C6 platform chemicals. *Biotechnol. Bioeng.* *109*, 2437–2459.
- Jong, E., Higson, A., Walsh, P., and Wellisch, M. (2012). Bio-based chemicals - value added products from biorefineries. IEA Bioenergy. <http://www.ieabioenergy.com/wp-content/uploads/2013/10/Task-42-Biobased-Chemicals-value-added-products-from-biorefineries.pdf>.
- Jordan, D.B., Bowman, M.J., Braker, J.D., Dien, B.S., Hector, R.E., Lee, C.C., Mertens, J.A., and Wagschal, K. (2012). Plant cell walls to ethanol. *Biochem. J.* *442*, 241–252.
- Jullesson, D., David, F., Pflieger, B., and Nielsen, J. (2015). Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol. Adv.* *33*, 1395–1402.
- Matsumoto, N.M., González-Toro, D.C., Chacko, R.T., Maynard, H.D., and Thayumanavan, S. (2013). Synthesis of nanogel-protein conjugates. *Polym. Chem.* *4*, 2464.
- Moncada, J., Posada, J.A., and Ramirez, A. (2015). Early sustainability assessment for potential configurations of integrated biorefineries. *Screening of bio-based derivatives from platform chemicals. Biofuels, Bioprod. Biorefining* *6*, 246–256.
- Montazeri, M., Zaimes, G., Khanna, V., and Eckelman, M. (2016). Meta-analysis of life cycle energy and greenhouse gas emissions for priority biobased chemical. *ACS Sus. Chem. Eng.* *4*, 6443–6454.
- Nielsen, J., Fussenegger, M., Keasling, J., Lee, S.Y., Liao, J.C., Prather, K., and Palsson, B. (2014). Engineering synergy in biotechnology. *Nat. Chem. Biol.* *10*, 319–322.
- Opgenorth, P.H., Korman, T.P., and Bowie, J.U. (2016). A synthetic biochemistry module for production of bio-based chemicals from glucose. *Nat. Chem. Biol.* *12*, 393–395.
- Papoutsakis, E.T. (1984). Equations and calculations for fermentations of butyric acid bacteria. *Biotechnol. Bioeng.* *26*, 174–187.
- Parsell, T., Yohe, S., Degenstein, J., Jarrell, T., Klein, I., Gencer, E., Hewetson, B., Hurt, M., Im Kim, J., Choudhari, H., et al. (2015). A synergistic biorefinery based on catalytic conversion of lignin prior to cellulose starting from lignocellulosic biomass. *Green. Chem.* *17*, 1492–1499.
- Shanks, B.H., and Keeling, P.L. (2017). Bioprivileged molecules: creating value from biomass. *Green. Chem.* *19*, 3177–3185.
- Straathof, A., and Bampouli, A. (2017). Potential of commodity chemicals to become bio-based according to maximum yields and petrochemical prices. *Biofuels, Bioprod. Biorefining* *11*, 798–810.
- Werpy, T., and Petersen, G. (2004). Top value added chemicals from biomass: volume I – results of screening for potential candidates from sugars and synthesis gas. *Program* *1*, 69.
- Woolston, B.M., Edgar, S., and Stephanopoulos, G. (2013). Metabolic engineering: past and future. *Annu. Rev. Chem. Biomol. Eng.* *4*, 259–288.
- Wu, W., Yenkie, K., and Maravelias, C.T. (2017). A superstructure-based framework for bio-separation network synthesis. *Comput. Chem. Eng.* *96*, 1–17.
- Wu, W., Long, M.R., Zhang, X., Reed, J.L., and Maravelias, C.T. (2018). A framework for the identification of promising bio-based chemicals. *Biotechnol. Bioeng.* *115*, 2328–2340.
- Yadav, V.G., De Mey, M., Giaw Lim, C., Kumaran Ajikumar, P., and Stephanopoulos, G. (2012). The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metab. Eng.* *14*, 233–241.
- Yenkie, K.M., Wu, W.Z., Clark, R.L., Pflieger, B.F., Root, T.W., and Maravelias, C.T. (2016). A roadmap for the synthesis of separation networks for the recovery of bio-based chemicals: matching biological and process feasibility. *Biotechnol. Adv.* *34*, 1362–1383.

**ISCI, Volume 15**














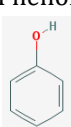


**Supplemental Information**



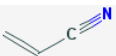


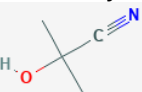
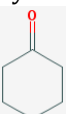
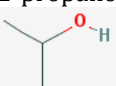

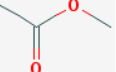
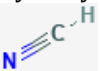
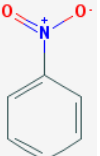

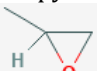
**Identifying the Characteristics of Promising  
Renewable Replacement Chemicals**

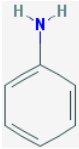
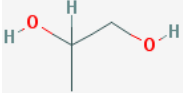
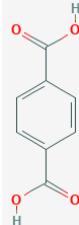
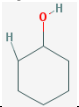

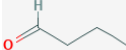
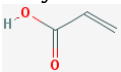
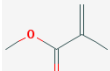
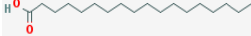
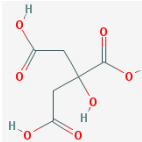
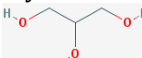
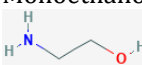

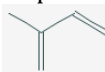
**Wenzhao Wu and Christos T. Maravelias**

## Structures of studied chemicals

Table S1. Structures of studied chemicals, related to Section 3. Entries marked in green are chemicals currently naturally produced. Urea is excluded. The entry numbers correspond to the numbers for chemicals in the supplementary excel sheets, ordered by volume MT/y (from high to low). C# represents the number of carbon atoms in a chemical.

| Entry# | C# | Chemicals order by MT/y  |
|--------|----|--|
| 1      | 2  | Ethanol<br>             |
| 2      | 2  | Ethylene<br>            |
| 3      | 3  | Propene<br>             |
| 4      | 2  | Ethylene dichloride<br> |
| 5      | 2  | Vinyl chloride<br>      |
| 6      | 6P | Benzene<br>             |
| 7      | 2  | Acetic acid<br>         |
| 8      | 8P | Xylene<br>            |
| 9      | 7P | Toluene<br>           |
| 10     | 8P | Styrene<br>           |
| 11     | 8P | Ethylbenzene<br>      |
| 12     | 2  | Ethylene oxide<br>    |
| 13     | 1  | Formaldehyde<br>      |
| 14     | 6P | Phenol<br>            |
| 15     | 1  | Methanol<br>          |
| 16     | 2  | Ethylene glycol<br>   |

|    |     |  |
|----|-----|--|
| 17 | 3   | Acetone<br>             |
| 18 | 6NP | Cyclohexane<br>         |
| 19 | 3   | Acrylonitrile<br>       |
| 20 | 4   | Butadiene<br>           |
| 21 | 15P | Bisphenol A<br>         |
| 22 | 4   | Acetone cyanohydrin<br> |
| 23 | 6NP | Cyclohexanone<br>     |
| 24 | 3   | 2-propanol<br>        |
| 25 | 4   | 1-butanol<br>         |
| 26 | 3   | Methyl acetate<br>    |
| 27 | 1   | Hydrocyanic acid<br>  |
| 28 | 6P  | Nitrobenzene<br>      |
| 29 | 6NP | Caprolactam<br>       |
| 30 | 3   | Propylene oxide<br>   |

|    |      |   |
|----|------|---|
| 31 | 6P   | Benzenamine<br>                |
| 32 | 3    | 1,2-propanediol<br>            |
| 33 | 8P   | Terephthalic acid<br>          |
| 34 | 6NP  | Cyclohexanol<br>               |
| 35 | 6NP  | Hexanedioic acid<br>           |
| 36 | 4    | Butanal<br>                    |
| 37 | 3    | Acrylic acid<br>             |
| 38 | 5    | Methyl methacrylate<br>      |
| 39 | 18NP | <b>Octadecanoic acid</b><br> |
| 40 | 6NP  | <b>Citric Acid</b><br>       |
| 41 | 3    | Glycerin<br>                 |
| 42 | 2    | Monoethanolamine<br>         |
| 43 | 18NP | <b>Oleic Acid</b><br>        |
| 44 | 5    | Isoprene<br>                 |

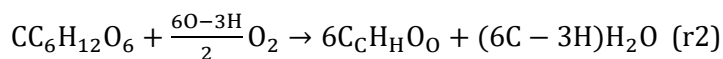
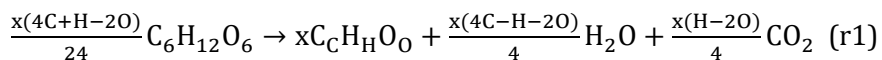


## Transparent methods

We first gather the molecular structures, market prices and market volumes of 44 organic chemicals (totaling 161 million MT/year; MT=metric ton; each has a volume between 0.2 and 41 million MT/year), including the top 20 commodity organic chemicals (Entries 1-20 in the Supplementary Material and the supplementary excel) by US volume (Essential Chemical Industry, 2016; Innovest, 2007; Wittcoff et al., 2012) and 24 other High-Production-Volume (HPV) chemicals (Entries 21-44) that can be produced through microbial conversion using *E. coli* and/or *S. cerevisiae* based on Flux Balance Analysis (Wu et al., 2018). We include the additional 24 chemicals to enlarge the data set we study. These 44 chemicals roughly represent 85% of the total US organic chemical market volume (FRED, 2017; Goldman, 2002). Note that while chemical prices and volumes are influenced by the price, availability and location of the oil and gas feedstocks, we do not consider such fluctuations in this work because the relative feedstock costs, which are correlated and fluctuate less, are more important although the shale gas development in recent years has somewhat weakened the correlation. The price and volume data obtained roughly reflect US values in recent years.

Next, we analyze five molecular characteristics: (1) number of carbon atoms (abbreviated as #C hereafter); (2) number of oxygen atoms (#O); (3) number of functional groups (#FG); (4) number of distinct functional groups (#DFG; e.g., 2 hydroxyl FGs are counted as 1 DFG); and (5) existence of specific functional groups (FG), such as alkenyl, hydroxyl and phenyl groups, as well as their combinations. Each characteristic corresponds to multiple attributes, e.g., #C=1 is an attribute of the #C characteristic. To analyze each of the five characteristics, we study the market volume and price data for all 44 chemicals. Market volume is used as a surrogate of the demand of a set of characteristics. price is used as surrogate for the cost of obtaining a chemical with a set of characteristics, because price is, in general, equal to the production cost plus a profit margin, and the margins across different commodity chemicals are (1) small (thus the use of price as a surrogate for cost); and (2) similar (9-14%, with commodity chemicals mostly near the 9% mark) (Maverick, 2015). Thus, we identify the molecular characteristics that are desirable (indicated by market volume) but difficult to obtain today from fossil fuel feedstocks (indicated by price). Finally, we discuss the characteristics of replacement chemicals that are relatively easy to obtain through bio-production.

### ***Explanations to the stoichiometric yield calculations***



From r1 (for most chemicals), we can calculate the stoichiometric yield (g product/g glucose) as:

$$\text{Yield} = \frac{x(12C+H+16O)}{180 \frac{x(4C+H-2O)}{24}} = \frac{4(12C+H+16O)}{30(4C+H-2O)} = \frac{12C+H+16O}{30C} \times \frac{4C}{4C+H-2O} \quad (S1)$$

From r2 (for two chemicals), we can calculate the stoichiometric yield as:

$$\text{Yield} = \frac{6(12C+H+16O)}{180C} = \frac{12C+H+16O}{30C} \quad (\text{S2})$$

Therefore, for all the chemicals, these two expressions can be combined:

$$\text{Stoichiometric maximum yield} = \frac{12C+H+16O}{30C} \times \text{Min}\left(1, \frac{4C}{4C+H-2O}\right) \quad (\text{S3})$$

### ***Explanations to nominal bio-production cost calculation***

Based on the work by Wu et al., bio-based chemicals that are extracellular, insoluble in water, and lighter than water (in terms of density) usually have the least expensive separation because this class of chemicals are naturally separated from the microbial cells, which are usually heavier than water, and thus less expensive centrifugal decantation is required for the separation. However, chemicals that are extracellular and soluble in water have the most expensive separation, because separating the product from large amount of water in the same phase is much more difficult, which usually involves the use of costly distillation and extraction. Therefore, we use chemicals that are intracellular and soluble in water as the basis for nominal cost calculation because they have a medium separation cost. The separation cost (\$/kg) - product titer (g/L) relation for intracellular soluble chemicals is as follows:

$$\text{Separation cost} = 9.09 \times \text{titer}^{-0.62} \quad (\text{S4})$$

where 240 g/L glucose is assumed in the feed, and thus titer can be calculated based on yield (g product/g glucose) as follows:

$$\text{Titer} = 240 \times \text{yield} \quad (\text{S5})$$

In the same study, the bio-conversion cost (\$/kg) is estimated based on yield as follows:

$$\text{Bioconversion cost} = \frac{0.6}{\text{yield}} \quad (\text{S6})$$

Therefore, the total bio-production cost can be calculated as follows:

$$\text{Bioproduction cost} = \frac{0.6}{\text{yield}} + 9.09 \times (240 \times \text{yield})^{-0.62} \quad (\text{S7})$$

where yield has been calculated in Equation S3.

## Supplemental References

Essential Chemical Industry. (2016). Available:  
<http://www.essentialchemicalindustry.org/chemicals.html>.

FRED. (2017). Industrial production organic chemicals. Available:  
<https://fred.stlouisfed.org/series/IPG32511A9N#0>.

Goldman, L. (2002). Toxic chemicals and pesticides. In *Stumbling toward Sustainability*: Environmental Law Institute.

Innovest (2007). Overview of the chemicals industry. Available:  
<http://www.precaution.org/lib/07/innovest.pdf>

Maverick, J. (2015). Which segments of the chemicals sector have the highest profit margins? Available: <https://www.investopedia.com/ask/answers/060115/which-segments-chemicals-sector-have-highest-profit-margins.asp>.

Wittcoff, H., Reuben, B., and Plotkin, J. (2012). *Industrial organic chemicals*. John Wiley & Sons.

Wu, W., Long, M.R., Zhang, X., Reed, J.L., and Maravelias, C.T. (2018). A framework for the identification of promising bio-based chemicals. *Biotechnol. Bioeng.* *115*, 2328–2340.