THE 2015 HOEFER PRIZES
FOR EXCELLENCE IN UNDERGRADUATE WRITING

In Recognition of Writing Achievement in the
Undergraduate Field of Study

stanford university
May 21, 2015
Quantifying Microbial Production of Fatty-Acid-Derived Biofuels

Samuel Cady

Chemical Engineering 185A

Chemical Engineering Laboratory A

Elizabeth Sattely

Chemical Engineering
Quantifying Microbial Production of Fatty-Acid-Derived Biofuels

Abstract

Over the past decade, dwindling reserves of traditional fuel sources and the threat of climate change have motivated the search for alternative sustainable fuels. While renewables such as wind and solar are promising, it is likely that biofuels will play a key role in the future energy economy due to their similarity to traditional fuels and their compatibility with existing distribution systems. However, to produce large amounts of these molecules economically, it is necessary to first adopt a system of microbial conversion of renewable feedstocks. Here, we take the first steps by quantifying the ability of engineered *Escherichia coli* to produce the desired biofuel, fatty acid ethyl esters (FAEEs). Through the use of liquid and gas chromatography, we track the consumption of glucose and the production of FAEEs over time. The concentration of FAEEs, ranging from 12-18 carbons, is found to increase to 88.0 mg/L, achieving a maximum percentage of theoretical yield of 7.39 % after 23 hours. While much work remains to be done to increase the efficiency of this process in order to improve the economics and scalability, this study demonstrates the viability of the microbial conversion approach to biofuel production.

Introduction

Fatty acid ethyl esters (FAEEs) are organic molecules composed of a long hydrophobic carbon chain attached to an ethyl ester group and are produced through the esterification of fatty acids with ethanol. Because FAEEs contain a long hydrocarbon chain, they are structurally very similar to the major components of diesel, 10-15 carbon-long saturated hydrocarbons. As a result, FAEEs are a highly desirable biofuel since they have nearly the same properties as diesel and therefore can be used with existing diesel infrastructure with minimal modification. But
unlike diesel, FAEEs can be produced from renewable feedstocks. However, challenges remain. The most promising method of FAEE biofuel production is microbial conversion. This method is explored here with the *E. coli* strain A2A that metabolizes glucose into FAEEs. Unfortunately, yields remain low. It is therefore critical to have a rapid method for monitoring biofuel yields to enable higher throughput screening of newly engineered microbes to identify a more efficient pathway. Rapid monitoring is also useful in certifying that biofuels meet quality standards.\(^4\) For this experiment, gas chromatography/mass spectrometry was selected to monitor FAEE production and liquid chromatography/mass spectrometry was selected to monitor glucose consumption. These techniques were selected as they are commonly practiced in the literature, are fully automated, and have high sensitivity,\(^5\) which is critical when first evaluating a strain to ensure that all products are identified. To further improve performance, the GC/MS method was optimized to minimize run time.

**Results & Discussion**

*GC/MS Method Development*

A number of preliminary operations were carried out in order to enable the rapid characterization and quantification of FAEEs in the cell cultures using gas chromatography/mass spectrometry. First, the run method was optimized in order to minimize run time while still achieving high peak separation. To this end, a commercial FAEE standard (Sigma Aldrich 49454-U) was run using a standard GC/MS method (see Methods).
The resulting chromatograph displayed seven peaks, corresponding to seven of the nine FAEEs contained in the standard. Figure 1 shows the chromatograph and peak assignments. Peaks were assigned by examining the accompanying mass spectrum. All peaks showed the characteristic m/z ratios of 88 and 101, identifying them as FAEE6 and were assigned by matching the molecular ion m/z signal to the molecular weight of the standard components. Peak assignments, as well as the m/z ratios and molecular weights used for assignment, are given in Figure 1B (mass spectra for individual peaks can be found in the Appendix 2.1). As expected, the lighter standard components eluted first since they are more volatile and less hydrophobic.

Having identified the standard components, it was possible to optimize the method to reduce the run time while ensuring that the same seven compounds were observed. First, as the peaks showed significant separation under the default method, the temperature ramp speed was increased to compress the spectrum, ramping from 70 °C to 240 °C by 30 °C per min rather than 25 °C. Second, as the default method included empty collection time at the end, the final hold at 2.134 min reduction.
240 °C was decreased from 3 to 2 min. As a result, run time was reduced by 2.134 min, while still achieving good separation as seen in Figure 1A for the improved method. The improved method was further validated when, using the peak assignment process described above, it was determined that the same seven FAEE were observed when using the improved method.

The final prerequisite to FAEE quantification was the creation of standard curves used to relate chromatograph peak area to concentration. Five dilutions of known concentration (2.5, 5.0, 10, 25, 50 mg/L) of the commercial standard were processed using the improved method. For each of the seven observed standard components, peak areas were found at each concentration. Observing that the relationship between peak area and concentration is linear, linear regression was performed on the data for each component to derive a ladder of equations relating peak area to concentration for FAEEs of varying length (as seen in Figure 2 for ethyl butyrate, see Appendix 2.2).

Resolution of Separation

The improved GC/MS method was further validated by calculating the resolution of separation between each consecutive peak for the 25 mg/L commercial FAEE standard (see Appendix 1.1). A resolution of separation greater than 1.5 indicates a minimum separation; all peak pairs had values far exceeding this minimum value (see Figure 3). We observe that lower molecular weight compounds, corresponding to early peaks, show a greater resolution of
separation. This is expected as for larger compounds the mass increase resulting from the addition of a two carbons is less significant compared to the overall compound mass. Thus, adjacent long FAEEs in the standard ladder will interact more similarly with the column than would adjacent short FAEEs, rendering the heavy compounds more difficult to separate. However, in all cases it is clear that the resolution of separation far exceeds the requirement, suggesting that this method could be further optimized by increasing the temperature ramp speed.

Theoretical Plates

As a final validation of the improved GC/MS method, the number of theoretical plates was calculated for each FAEE peak appearing in the 25 mg/L commercial standard. Theoretical plates is an important metric for characterizing the effectiveness of a column and method, with a larger number of theoretical plates corresponding to a more effective separation. The number of theoretical plates for each FAEE was calculated using the half peak height method (see Appendix 1.2). The results are summarized in Table 1.
The GC column used in this lab, an Agilent HP-5ms Ultra Inert (19091S-433UI), has a manufacturer reported number of theoretical plates of 142,500. Our calculated values are significantly larger than this number for all seven FAEEs (ranging from 2-18 times larger). Since the number of theoretical plates is unique to each compound, without knowing which compound the manufacturer used it is difficult to explain this variation. However, it appears that this column is particularly efficient at separating FAEEs since the calculated number of theoretical plates is much larger than the manufacturer reported value.

The number of theoretical plates was also calculated for the PMP-glucose complex, which was separated from the cell culture media using liquid chromatography. Again, the half peak height method was used (see Appendix 1.2), yielding a value of 68,000 theoretical plates for the 11-hour incubation sample, replicate 2. The LC column used in this lab, a Zorbax SB-C18 (827700-902), has a manufacturer reported number of theoretical plates of 11,689. In comparison, our calculated value is about six times larger. This column appears to be well chosen for this procedure as it is highly efficient at separating PMP-derivatized glucose. The high number of theoretical plates can also be attributed to the derivatization procedure, which was carried out to ensure that glucose could be successfully retained and separated by the column.

**FAEE Quantification**

Having validated our improved GC/MS method, it was possible to quantify the ability of the engineered *E. coli* to produce FAEEs from glucose. Cell cultures were sampled at three time points: prior to addition of cells (T=0), prior to cell induction (Uninduced), and eleven hours after induction (T=11); two replicates were taken at each time point for a total of six samples (for details, see Methods). To facilitate GC/MS analysis, FAEEs were extracted from each of the
samples (see Methods). Each sample was analyzed using the improved GC/MS method (see Methods). The resulting data for the 11-hour incubation sample (replicate 2) is summarized in Figure 4.

![Figure 4](image)

**Figure 4** | (A) Labeled total ion chromatograph (black) of extracted FAEEs after 11 hour incubation (STT_11_2) and total ion chromatograph (red) of 25 mg/L commercial FAEE standard; (B) Peak assignments for spectrum in panel A. Assignment performed by matching molecular ion m/z signal (M+) to component of correct molecular weight (M_w).

The total ion chromatograph of the cell culture (upper line, Figure 4A) shows a large number of peaks, of which ten were deemed significant upon further inspection of their corresponding mass spectra (either fatty acid or FAEE). Peak assignment began with the internal standard dodecane, known to be present in the sample. This compound was assigned to peak 1 as this peak appeared consistently in all tested samples and displayed a signal at an m/z of 170, the molecular weight of dodecane, and another signal at an m/z of 85 corresponding to the double ion. Next, the chromatograph of the cell culture was compared with that of the commercial FAEE standard (Figure 4A). As all samples were run with the same method, peaks with the same elution time were assumed to contain the same compound. This was confirmed by comparison of the mass spectra, and allowed peaks 2, 4, and 7 to be assigned (see Figure 4B). The remaining
peaks cluster near peaks present in the commercial FAEE standard. Examination of the mass spectra of these unassigned peaks revealed that they have molecular ion m/z signals (M+) either 2 or 28 less than the molecular ion peak of the nearest eluting standard component. Peaks with molecular ion signals 2 less than the closest standard peak also displayed m/z signals at 88 and 101, characteristic of FAEEs. Therefore, these peaks were assigned as the unsaturated version of the corresponding standard component as the presence of a double bond accounts for the mass reduction. Peaks with molecular ion signals 28 less than the closest standard peak displayed additional m/z signals at 60 and 73, characteristic of fatty acids. Therefore, these peaks were assigned as the unesterified fatty acid version of the corresponding standard component as the lack of the ethyl group accounts for the mass reduction (see Figure 4B). All peak assignments were validated by comparison with known mass spectra. A similar procedure was followed to assign peaks in the other samples (peak mass spectra found in Appendix 2.4).

Following peak assignment, the concentrations of major FAEEs were determined for each sample. The area of each peak corresponding to an FAEE was integrated using default parameters. Using our FAEE standard curves, peak areas were converted to sample concentrations. For unsaturated FAEEs not appearing in the commercial standard, the standard curve of the most similar saturated FAEE was used. Normalization was performed using the internal standard dodecane (see Appendix 1.3). The results are summarized in Table 2.

<table>
<thead>
<tr>
<th>Concentrations of Major FAEEs</th>
<th>Cell Culture Concentrations (mg/L)</th>
<th>Elution Time</th>
<th>FAEE Assignment</th>
<th>T = 0 hours</th>
<th>Uninduced</th>
<th>T = 11 hours</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.217</td>
<td>ethyl dodecanoate</td>
<td>6.989</td>
<td>9-tetradecanoate</td>
<td>7.067</td>
<td>ethyl myristate</td>
<td>7.938</td>
<td>ethyl 9-hexadecanoate</td>
<td>8.03</td>
<td>ethyl palmitate</td>
<td>9.238</td>
</tr>
<tr>
<td>FAEE Total Concentration (mg/L)</td>
<td>0</td>
<td>0</td>
<td>8.81</td>
<td>8.12</td>
<td>59.6</td>
<td>50.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As expected, there are no FAEEs present in the \( t = 0 \) samples. This is explained by the fact that no FAEE-producing *E. coli* were present when these samples were taken, and the media contains no FAEE. Both the uninduced replicates show low levels of longer chain FAEEs (C14-C18). This is likely due to a leaky pathway in the engineered *E. coli*. While these cells had not yet been induced, it is possible that the FAEE synthesis pathway was operating at a low level. This hypothesis is strengthened by the fact that there are no short chain (< C14) FAEEs present. Since this pathway has a preference for longer chain FAEEs, if it were operating at a low level we would expect to see only longer chain FAEE in measurable quantities.\(^{13}\) In the 11-hour incubation sample, we observe the highest concentrations of FAEEs. This is expected as the FAEE synthesis pathway had been fully active for eleven hours when this sample was taken. Again, heavier FAEEs are present in higher concentrations, indicating the pathway’s preference for longer chain FAEEs. In all cases, there is good correspondence between the two replicates, suggesting consistent extraction.

**Glucose Quantification**

Having measured the production of fatty acid ethyl esters, it was necessary to quantify the glucose concentration in our samples through the use of liquid chromatography/mass spectrometry. The consumption of glucose over time can then be compared to the production of FAEEs in order to gauge process efficiency. Prior to analysis, the glucose in each of the previously described samples was derivatized with PMP (see Methods, Figure 5 for reaction scheme) to ensure retention.

\[\text{Figure 5} \mid \text{PMP-derivatization of D-glucose}\]
on the hydrophobic column and to prepare it for analysis with UV/VIS spectroscopy and mass spectrometry. All six derivatized samples were run using the standard LC/MS method (see Methods).

![Figure 6](A) Total ion chromatograph (TIC, upper line) and extracted ion chromatograph (EIC, lower line) for PMP-derivatized E. coli media after 11 hours incubation (STT_11_2); (B) Mass spectrum for PMP peak; (C) UV spectrum for PMP peak; (D) Mass spectrum for PMP-Glucose peak; (E) UV spectrum for PMP-Glucose peak

The resulting data for the 11-hour incubation sample (replicate 2) is summarized in Figure 6. The total ion chromatograph (see Figure 6A) shows two major peaks corresponding to free PMP and the derivatized glucose. These peaks were assigned by examining their corresponding mass and UV/VIS spectra. The first peak showed a strong signal at an m/z of 175.1, close to 174.2, the molecular weight of PMP (see Figure 6B). It also absorbed strongly at 240 and 280 nm, which is characteristic of PMP (see Figure 6C). Thus, this peak was assigned to free PMP. The second peak showed a strong signal at an m/z of 512.2, close to 510.6, the
molecular weight of the PMP-glucose complex (seen in Figure 5). This is confirmed by a peak at 256.2 corresponding to the double ion (see Figure 6D). UV absorbance is similar to that of the PMP peak, suggesting PMP is present (see Figure 6E). Therefore, this peak was assigned to the PMP-glucose complex. A similar procedure was followed to assign peaks in the other samples.

Having identified the peak corresponding to the glucose-PMP complex, an extracted ion chromatogram was generated for this species, including possible isotopes up to an m/z of 514 (see Figure 6A). These curves were integrated using the default parameters. Using the provided standard curve data, peak areas were converted to sample concentrations (see Appendix 1.4). The results are summarized in Table 3 below.

As expected, the concentration of glucose is highest at t = 0 since at this point the culture contained no cells. We note that the measured concentration of glucose at t = 0 using this technique is approximately 25% lower than the actual concentration of glucose in M9 minimal media (4 g/L)\textsuperscript{16}, indicating that not all glucose in the sample was derivatized successfully. However, this does not impede the analysis of trends over time. The results for the 11-hour samples also conform to expectations. The concentration of glucose decreases by roughly 35% over this time period. This is explained by the metabolism of glucose by \textit{E. coli} and its conversion into FAEE and other products. However, the result for the un-induced samples is less easily accounted for. We find that the glucose concentration for these samples is as low or lower as that of the 11-hour samples, even though these cells lacked the ability to produce FAEE. One explanation is that most glucose metabolism occurs before the cells are induced, resulting in a negligible difference in glucose concentrations between the two

<table>
<thead>
<tr>
<th>Glucose Concentrations (g/L)</th>
<th>T = 0 hours</th>
<th>T = 11 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate 1</td>
<td>Replicate 2</td>
<td>Replicate 1</td>
</tr>
<tr>
<td>3.01</td>
<td>2.86</td>
<td>1.77</td>
</tr>
</tbody>
</table>
samples. However, this is unlikely as the induced cells produce measurable quantities of FAEEs by 11 hours, and this requires significant additional glucose metabolism. Another explanation is that glucose derivatization was less successful for the un-induced samples, resulting in artificially lowered glucose concentrations. But since both replicates showed these low concentrations, this explanation is not entirely satisfying. This phenomenon will need to be studied further. Nonetheless, combined with the data showing FAEE production, this analysis confirms that the engineered *E. coli* are successfully converting the glucose feedstock into the desired fatty acid ethyl ester product.

*Theoretical Yield and Trends in FAEE and Glucose Concentrations*

Drawing on the data collected by other groups\textsuperscript{17} who measured glucose and FAEE concentrations at other time points, an effort was made to track changes in FAEE and glucose concentrations over time. The titer\textsuperscript{s} for each major FAEE, total FAEE, and glucose for each of our samples, reported earlier in this paper (see Table 2 for FAEE data and Table 3 for glucose data), were compared with those of other groups (see Appendix 2.8 for all data). The same previously described procedures for calculating FAEE and glucose concentrations were applied to the new data (see Appendix 1.3-4).

The resulting data is summarized in Figure 7. To aid in visualization, all FAEEs of the same chain length were aggregated, regardless of saturation, and concentrations were averaged across the two replicates at each time point (see Appendix 1.5). As expected, total FAEE concentration increases with time following incubation, reaching a maximum 88.0 mg/L at 23 hours (Figure 7A). Concentration appears to have a non-linear dependence on time as there is a sharp increase in concentration between the 7- and 11-hour samples. It is possible that this is the result of some lag in the metabolic pathway for FAEE synthesis, as it is fairly complex.\textsuperscript{18} It is
also possible that this non-linearity is exacerbated by error in measurement of the 7-hour FAEE concentration. This sample, unlike the others, does not contain any C18 compounds, suggesting the FAEE extraction may have been incomplete or that the GC/MS method used did not capture larger FAEEs, resulting in an artificially lowered concentration. The concentrations of the individual FAEEs increase over time in the same manner and exist in roughly the same proportions at each time. These proportions are likely determined by the propensity of the engineered pathway to produce FAEEs of a certain length.

The trend in glucose concentration over time is less expected. Since there is a clear increase in the concentration of FAEEs over time, and glucose must be metabolized to produce FAEEs as well as sustain the *E. coli*, it was expected that the glucose concentration would steadily decrease over time. However, this is not the case. While a sharp reduction in glucose concentration is observed between the t = 0 and un-induced samples, the measured glucose concentration actually increases following incubation and remains roughly constant over the remaining sample times (Figure 7B). As previously discussed, it is possible that the majority of glucose metabolism occurs before induction, resulting in negligible differences in concentration.
in later samples. However, this seems unlikely as the *E. coli* were producing FAEE, indicating continued metabolism over a period of 23 hours. A more likely explanation is the failure of the glucose derivatization procedure to achieve consistent conversion of glucose to glucose-PMP across samples, introducing a large degree of error into the measured concentrations. We previously observed that only ~75% of glucose was converted to glucose-PMP in the t = 0 samples and the yield may have fluctuated for other samples. Since these samples lacked any internal standard, it was not possible to normalize the samples. A final possible source of error is the samples themselves. Across all time points (post-induction), a degree of magnitude variation in glucose-PMP peak area was noted between replicates. As this variation was independently observed by four groups, it suggests some problem with the provided cell culture samples as one replicate had a much higher concentration of glucose than the other. For the purpose of this study, the values were averaged, but it would be useful to repeat this experiment with new cell cultures to eliminate this source of error.

Finally, to gauge the overall efficiency of the FAEE synthesis process, the theoretical yield was calculated using the maximum observed FAEE titer of 88.0 mg/L (23-hour incubation sample). This titer is nearly an order of magnitude lower than the maximum FAEE titer reported by Keasling et al. of 674 mg/L.\(^{20}\) This discrepancy is unexpected as this experiment employed the same *E. coli* strain A2A from the Keasling lab and used a comparable induction time (23 hrs. v. 24 hrs.). It appears that Keasling et al. observed anomalous behavior during their experiment or that our procedures for FAEE extraction and quantification were poorly designed or executed. The most likely explanations are incomplete extraction of FAEEs from the cell cultures, failure of the GC/MS method to capture all present FAEEs, or a problem with the *E. coli* culture. Nonetheless, this titer, and the concentration of glucose in the media at t = 0, were used to
calculate a total FAEE theoretical yield of 1190 mg/L, of which we achieved 7.39 % (see Appendix 1.6). This value is close to that reported by Keasling et al. of 9.4 %. However, these values cannot be directly compared as they were calculated using different methodologies. Using the Keasling methodology, which reported a total FAEE titer of 647 mg/L as 9.4 % of the theoretical yield, we achieved only 1.3 % yield. Again, we observed a much lower production of FAEEs than that reported by Keasling et al.

**Conclusion**

In conclusion, this work demonstrates both the potential and the challenges of microbial fatty acid ethyl ester production. The development of an optimized GC/MS method allowed for the rapid characterization of sample compounds. Using this method, the production of FAEEs and the consumption of glucose were successfully tracked over time, to a maximum of FAEE titer of 88.0 mg/L and a percentage of theoretical yield of 7.39 %, proving the validity of this metabolic pathway for synthesizing FAEE from glucose. However, these values remain low and differ significantly from those reported in the literature. This suggests two avenues for future improvement. First, additional metabolic pathways and microbes need to be explored to achieve higher yields of FAEEs from glucose. Second, the procedures outlined in this report for quantifying FAEE and glucose concentrations need to be improved to achieve greater reproducibility. The procedures outlined in this report can serve as a basis for both of these efforts. This report reveals the vast quantity of work that remains to be done before microbial biofuel synthesis will become competitive with fossil fuels, but it also highlights the promise: a consolidated bioprocess for converting sugar to usable fuel.
9 Value provided by David Ayala-Lindeman
10 Value provided by Joshua Willis
17 Data provided by other members of Wednesday lab section: Georgy, Austin, Chris, Pepito, Tom, Mesgana. Eric.
19 Ibid
20 Ibid, 561.
21 Ibid