

Electronic Supplementary Information

Expanding Analytical Horizons: 3D HPLC Calibration Surfaces for Micro-Mole Scale Self-Optimizing Flow Reactors

Maëva Vallet,^a Daniel Cortés-Borda^{a,*} and François-Xavier Felpin^{a,*}

^a Nantes Université, CNRS, CEISAM UMR 6230, 2 rue de la Houssinière, 44322 Nantes, France.

* Corresponding authors.

Email: daniel.cortes-borda@univ-nantes.fr; [ORCID: 0000-0002-2593-0656](https://orcid.org/0000-0002-2593-0656)

Email: fx.felpin@univ-nantes.fr; [ORCID: 0000-0002-8851-246X](https://orcid.org/0000-0002-8851-246X); Website: <http://felpin.univ-nantes.fr/>

Table of content

| | |
|--|------------|
| 1. Construction of the 3D calibration surface | S2 |
| 2. Optimization results of the formal [3 + 3] cycloaddition of dihydroxyquinoline with citral | S4 |
| 3. Comparison 2D and 3D calibration models | S6 |
| 4. HPLC Chromatograms | S9 |
| 5. NMR spectra | S20 |
| 6. References | S21 |

1. Construction of the 3D calibration surface

The construction of the 3D model follows the general procedure explained in the results and discussion section of the manuscript. Here we present a step by step description of the procedure from diagram depicted in Fig. 5:

a/ User provided grid

For generating the 2D grid of 20 data points, we start by fixing the concentration of the internal standard (1.25 M), and ranging the relative concentration of product of interest at 100%, 50%, 10% and 0% of theoretical yield (Table S2). Next, we replicate this series of points for a set of decreasing concentrations by diluting the stock solution by a factor ranging from 0 to 90% (see Table S1 for the values of concentration). Importantly, the 0% yield point corresponds to a real analysis. Although the area ratio naturally remains zero, the intensity of the internal standard places this zero at varying levels of internal standard concentration. This grid is intended to explore all the range of concentrations, from very diluted to highly concentrated, and from very low yield to high yield.

TABLE S1. Concentration of compound **3** (100% theoretical yield) and *m*-xylene at 5 different levels of dilution.

| Dilution (%) | Product 3 (mM) | <i>m</i> -Xylene (M) |
|--------------|----------------|----------------------|
| 0 | 125 | 1.25 |
| 25 | 94 | 0.94 |
| 50 | 63 | 0.63 |
| 75 | 31 | 0.31 |
| 90 | 13 | 0.13 |

Table S2 compiles the data points (area ratio of product **3**/*m*-xylene) initially provided by the user. This grid includes two variable dimensions: 4 different concentrations of the main analyte at a fixed concentration of the internal standard (expressed as a % of yield) and a gradient concentration of the overall solutions (expressed as a % of dilution of the stock solution).

TABLE S2. Values of area ratio (product 3/*m*-xylene) used to generate the initial version of the 3D model.

| Dilution (%) \ Yield (%) | 0 | 25 | 50 | 75 | 90 |
|--------------------------|-------|-------|-------|-------|-------|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0.249 | 0.194 | 0.461 | 0.49 | 0.508 |
| 50 | 1.317 | 1.306 | 1.178 | 1.596 | 1.775 |
| 100 | 1.974 | 2.111 | 1.861 | 2.678 | 2.927 |

b/ Generate model

The collected data of x= area ratio, y= internal standard intensity, and z= yield is entered into the multi-polynomial regression method proposed by Agrawal et al.¹ to generate the calibration surface. We created an initial version of the 3D model stemming from 20 data points. It is worth mentioning that the model generation is a supervised process that allows the user to select the polynomial degree and manually evaluate the R² and overall error values by comparing real and calculated data.

c/ Accurate model?

To assess its accuracy, we generated an additional set of points, totaling 20% of the initial dataset, which comprised 4 points for testing (Table S3). After supervised analysis of prediction accuracy, we were able to decide whether to validate the initial model or enhance it by incorporating these validation points. In this case, we chose refinement, resulting in a second version of the model.

TABLE S3. Values of area ratio (product 3/*m*-xylene) used to validate the accuracy of the 3 D.

| Dilution (%) \ Yield (%) | 0 | 25 | 50 |
|--------------------------|-------|-------|-------|
| 25 | 0.698 | 0.503 | |
| 75 | | 1.554 | 1.147 |

d/ Duplicate points

Once the accuracy of the second version of the model is verified, we executed fractional replication, which involved repeating 1/3 of the points, resulting in 8 new points that duplicate some of the already sampled points (Table S4). We empirically determined that these new points should align within the initial 20-point grid.

TABLE S4. Values of area ratio (product *3/m*-xylene) used to refine the 3 D model with replicated data points.

| Dilution (%) \ Yield (%) | 0 | 25 | 50 | 75 | 90 |
|--------------------------|--------------|-------|--------------|-------|-------|
| 0 | | | | | |
| 10 | | | 0.464; 0.452 | 0.47 | |
| 50 | 1.367; 1.389 | | | 1.453 | |
| 100 | | 2.281 | | | 3.002 |

e/ Refine with duplicates

Once again, we could decide to further enhance the model using the replicated data points or keep the model as it is. After noting that the extra 8 points improve the model accuracy, we decided to include it. Overall, the last version of the model includes 32 points.

2. Optimization results of the formal [3 + 3] cycloaddition of dihydroxyquinoline with citral

We have used a deeply modified version of the Nelder-Mead optimization algorithm, which is described in detail in our previous reports.^{2,3}

The optimization began by placing the initial simplex in a 4-dimensional space at the starting point $X_0 = (36\text{ }^\circ\text{C}, 24\text{ minutes}, 1.2\text{ equivalents of citral } \mathbf{2}, 7\text{ mol\% of ethylenediamine})$ with delta values of $d = 6\text{ }^\circ\text{C}, 9\text{ minutes}, 0.2\text{ equivalents of citral } \mathbf{2}$ and 2 mol% of ethylenediamine. In the first experiment, a 13% yield was achieved (Table S5). By the 13th experiment, the optimization algorithm proposed reducing the problem dimensions from 4 to 3 by fixing the equivalents of citral **2** to 1 equiv. This 3D optimization continued until the 20th experiment, where the algorithm suggested further dimension reduction, going from 3D to 2D by fixing the

residence time to 60 minutes. However, the second-dimension reduction was not executed since the resulting 2D simplex already met the termination criterion of all being almost experimentally equivalent (experiments 18-20). At the 20th experiment, the yield was 70%. With 10 experiments remaining in the budget, the user decided to continue the optimization by accepting the algorithm's proposal to employ a diversification mechanism in an attempt to find a better local optimum.

A new 4D optimization began with a random starting point $X_0 = (74\text{ }^\circ\text{C}, 19.4\text{ minutes}, 1.3\text{ equivalents of citral } \mathbf{2}, 10.5\text{ mol\% of ethylenediamine})$ using the previously mentioned delta values. The new 4D optimization continued until the 30th experiment, depleting the experimental budget. A local maximum satisfying the user was found at the 29th experiment, with the following conditions: (82 °C, 31.2 minutes, 1.1 equivalents of citral **2**, 9 mol% of ethylenediamine).

TABLE S5. Maximization of the reaction yield of the formal [3 + 3] cycloaddition of dihydroxyquinoline with citral in DMSO with a modified version of the Nelder-Mead algorithm.

| Expt # | Residence time (min) | Temperature (°C) | Citral 2 (equiv) | Catalyst loading (equiv) | Yield (%) |
|----------------------------|----------------------|------------------|-------------------------|--------------------------|-----------|
| 1 | 24 | 36 | 1.2 | 0.07 | 13 |
| 2 | 33 | 36 | 1.2 | 0.07 | 21 |
| 3 | 24 | 42 | 1.2 | 0.07 | 19 |
| 4 | 24 | 36 | 1.4 | 0.07 | 15 |
| 5 | 24 | 36 | 1.2 | 0.09 | 18 |
| 6 | 28.5 | 39 | 1.3 | 0.08 | 17 |
| 7 | 30.8 | 41 | 1.1 | 0.09 | 25 |
| 8 | 31.7 | 41 | 1.0 | 0.09 | 32 |
| 9 | 27.9 | 39 | 1.0 | 0.08 | 23 |
| 10 | 34.3 | 43 | 1.0 | 0.06 | 28 |
| 11 | 34.3 | 39 | 1.0 | 0.08 | 23 |
| Dimension reduction | | | | | |
| 12 | 39.0 | 43 | 1.0 | 0.07 | 33 |
| 13 | 44.6 | 46 | 1.0 | 0.07 | 44 |

| | | | | | |
|------------------------|------|----|-----|------|----|
| 14 | 39.4 | 48 | 1.0 | 0.07 | 42 |
| 15 | 42.9 | 47 | 1.0 | 0.09 | 47 |
| 16 | 47.1 | 49 | 1.0 | 0.10 | 58 |
| 17 | 55.7 | 54 | 1.0 | 0.07 | 62 |
| 18 | 60.0 | 56 | 1.0 | 0.07 | 71 |
| 19 | 60.0 | 52 | 1.0 | 0.09 | 67 |
| 20 | 60.0 | 55 | 1.0 | 0.09 | 70 |
| Diversification | | | | | |
| 21 | 19.4 | 74 | 1.3 | 0.10 | 65 |
| 22 | 19.4 | 80 | 1.3 | 0.10 | 70 |
| 23 | 28.4 | 74 | 1.3 | 0.10 | 71 |
| 24 | 19.4 | 74 | 1.5 | 0.10 | 59 |
| 25 | 19.4 | 74 | 1.3 | 0.12 | 60 |
| 26 | 23.9 | 77 | 1.1 | 0.11 | 69 |
| 27 | 26.1 | 79 | 1.2 | 0.09 | 88 |
| 28 | 29.5 | 81 | 1.1 | 0.07 | 89 |
| 29 | 31.2 | 82 | 1.1 | 0.09 | 90 |
| 30 | 37.1 | 86 | 1.0 | 0.09 | 85 |

3. Comparison 2D and 3D calibration models

2D calibration curves were generated at five level of dilution with data used for the 3D calibration model (Fig. S1). A significant deviation of the slope can be observed resulting in large disparities in the yield determination.

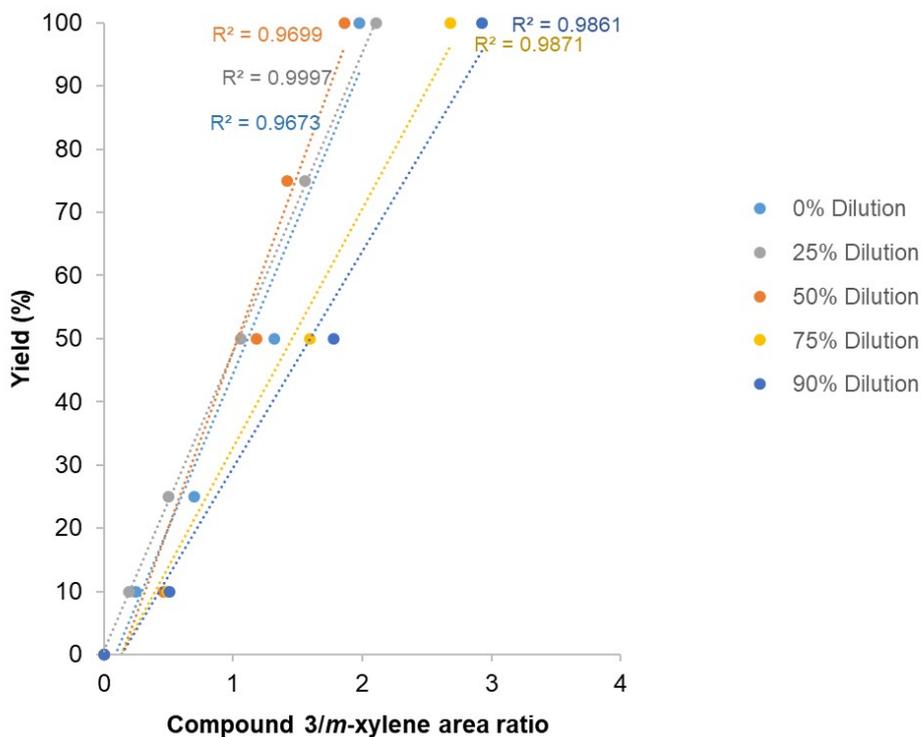


Fig. S1 2 D Calibration curves at different level of dilution of the stock solution

Fig. S2 illustrates the robustness of yield estimation achieved by the 3D calibration model in guiding the optimization process. In contrast, the utilization of 2D calibration models with distinct slope variations results in significant disparities. Error bars represent the outcomes from five distinct calibration curves obtained across various concentration ranges. These results highlight notable variability, which has the potential to introduce misleading feedback and complicate the optimization procedure. Notably, yield results from the 2D models were obtained post-optimization, with the optimization process itself conducted using the 3D model. If any of these 2D calibration models had been integrated into the optimization feedback loop, it would have led to a significantly divergent trajectory toward the optimal solution due to their distinct slope variations throughout the optimization process.

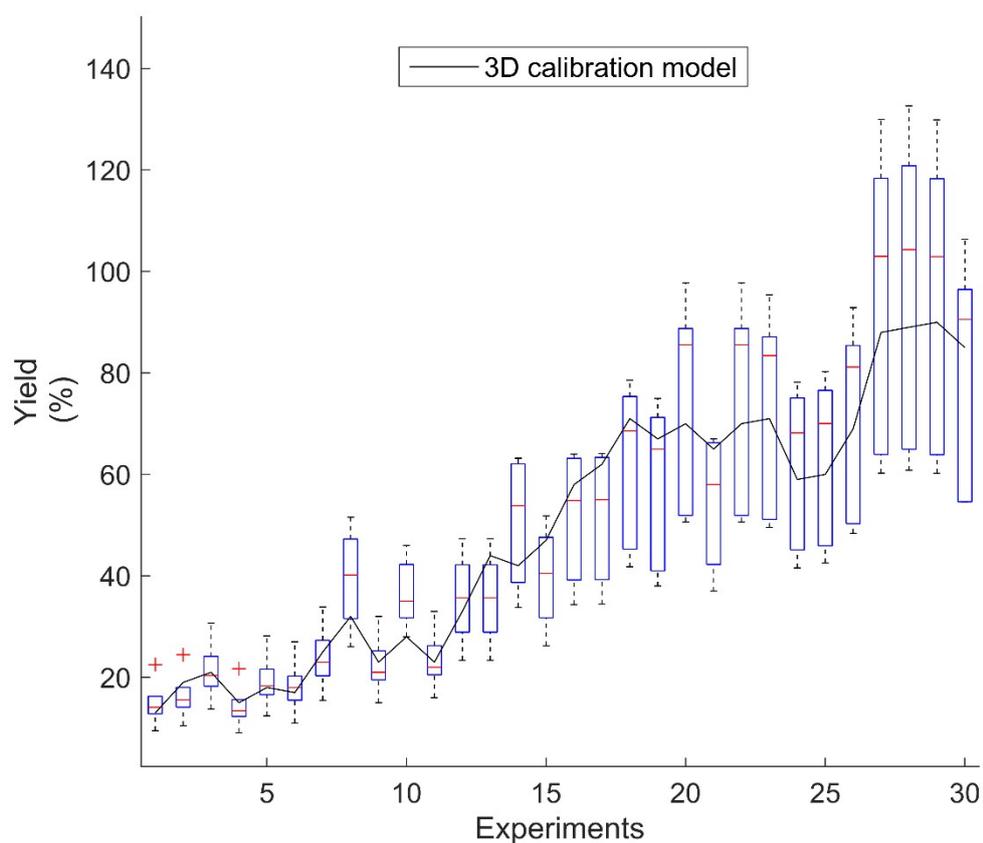


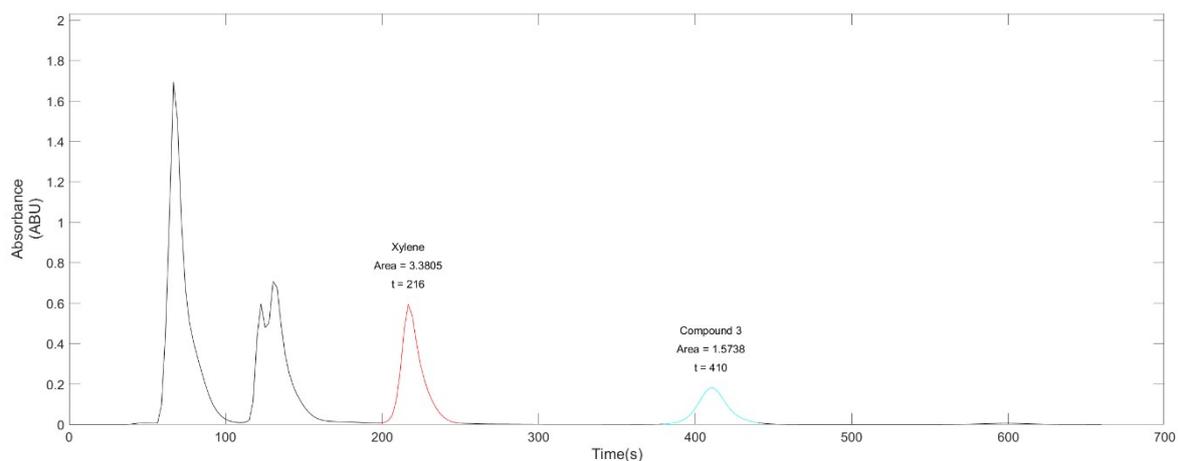
Fig. S2 Yield estimation with the 3D calibration model (solid line) and with five different 2D calibration models (error bars).

Note that error bars in Fig. S2 does not represent optimization using the 2D calibration curves. Error bars show, for each point of the optimization carried out using the 3D calibration surface, the variation in yield that would have been obtained with the five 2D calibration curves. It appears that the higher the yields, the greater the margin of error with the 2D approach.

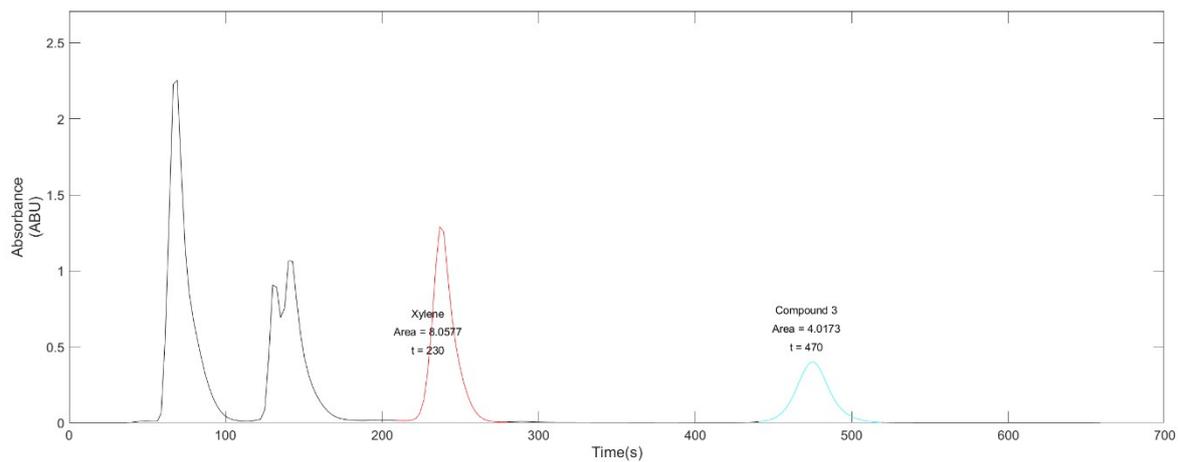
4. Chromatograms of the optimization

Note that chromatograms were not recorded with a temperature-controlled HPLC column, leading to changes in retention times throughout the days. When retention times are significantly altered the software prompts the chemist in charge of the optimization for confirmation. In such cases, the chemist's approach is to systematically duplicate the reaction to validate or invalidate the change in retention time. Therefore, all chromatograms with excessively altered retention times were duplicated or even triplicated before validation.

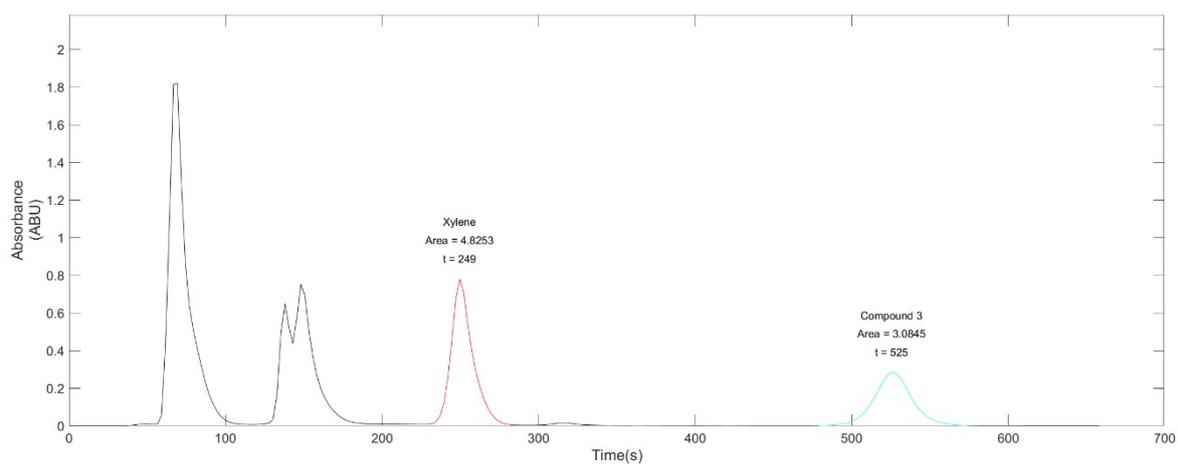
Iteration 1. 13% Yield



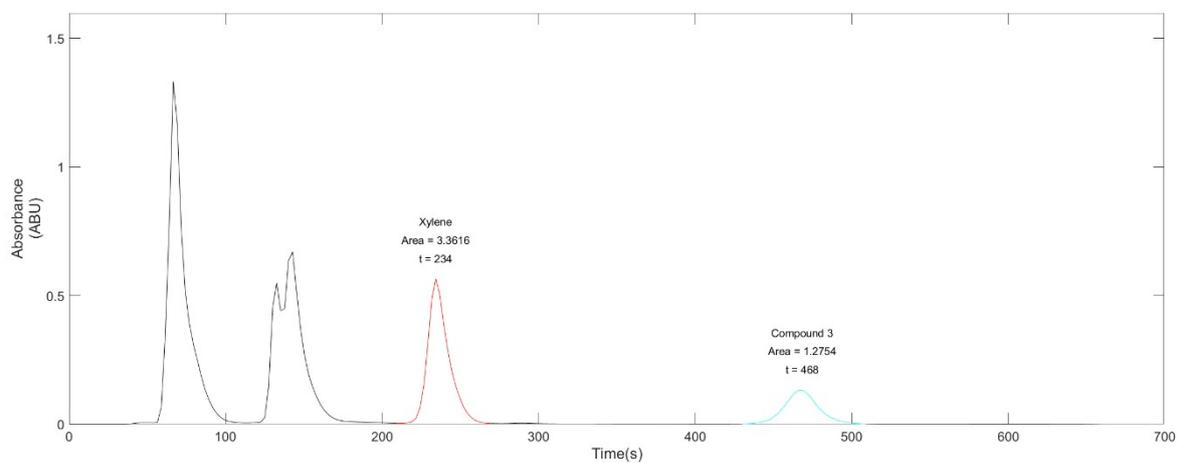
Iteration 2. 21% Yield



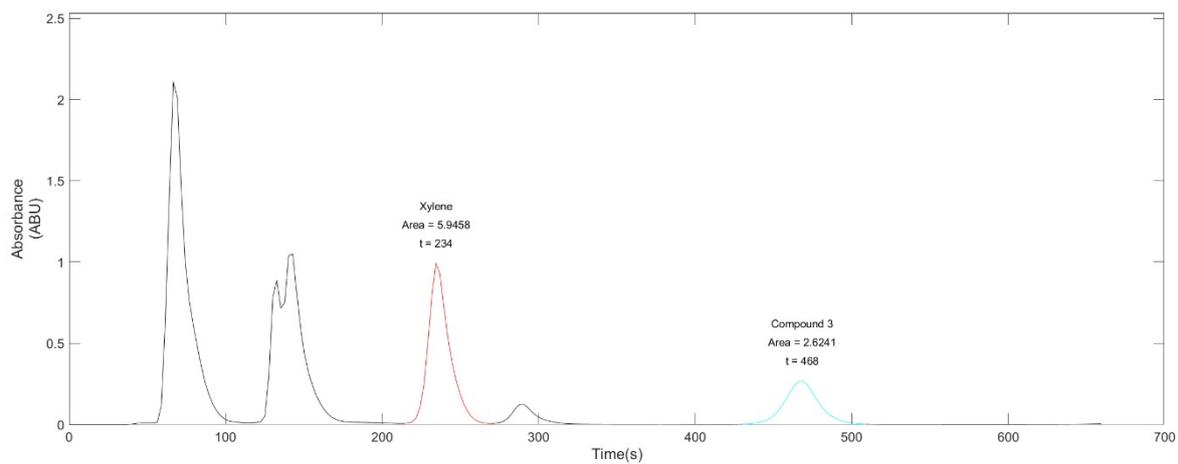
Iteration 3. 19% Yield



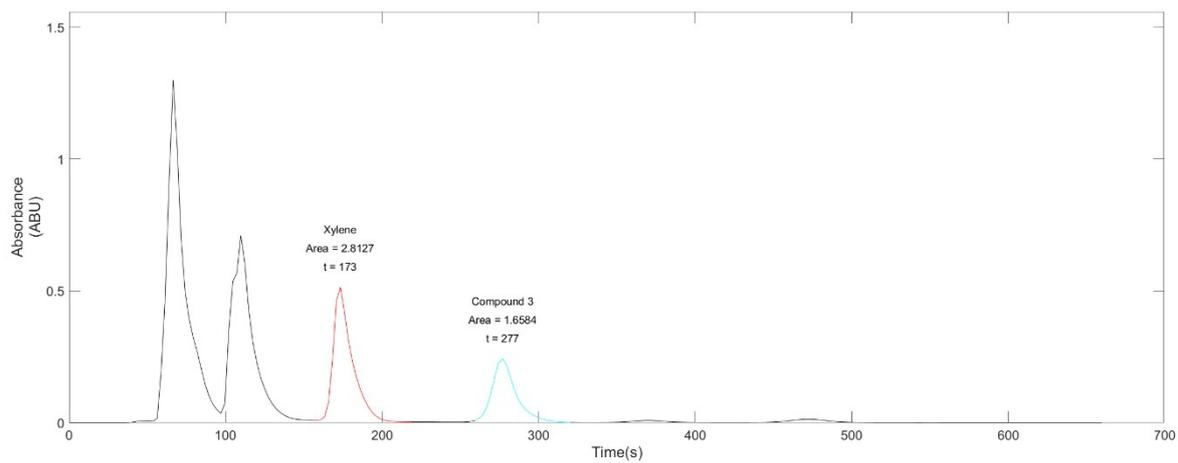
Iteration 4. 15% Yield



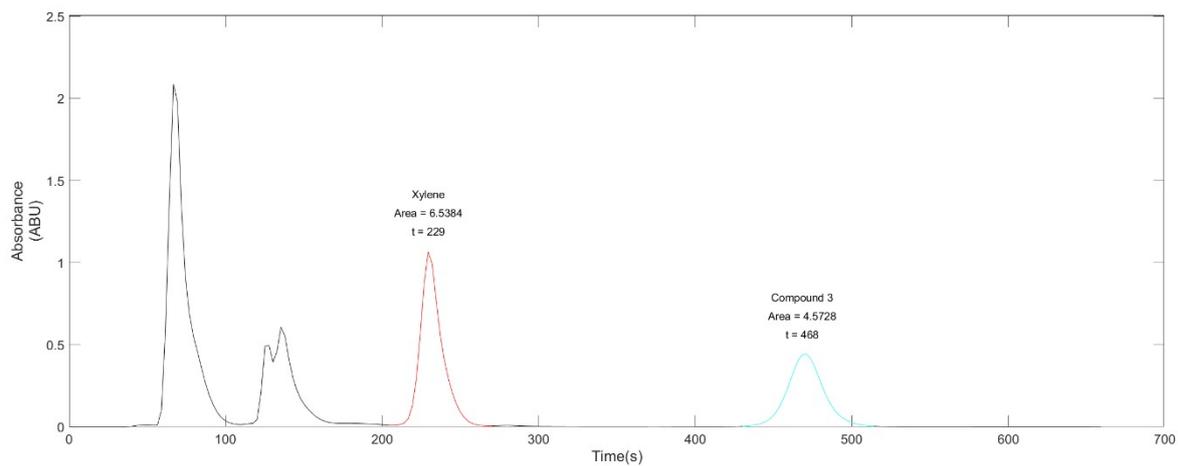
Iteration 5. 18% Yield



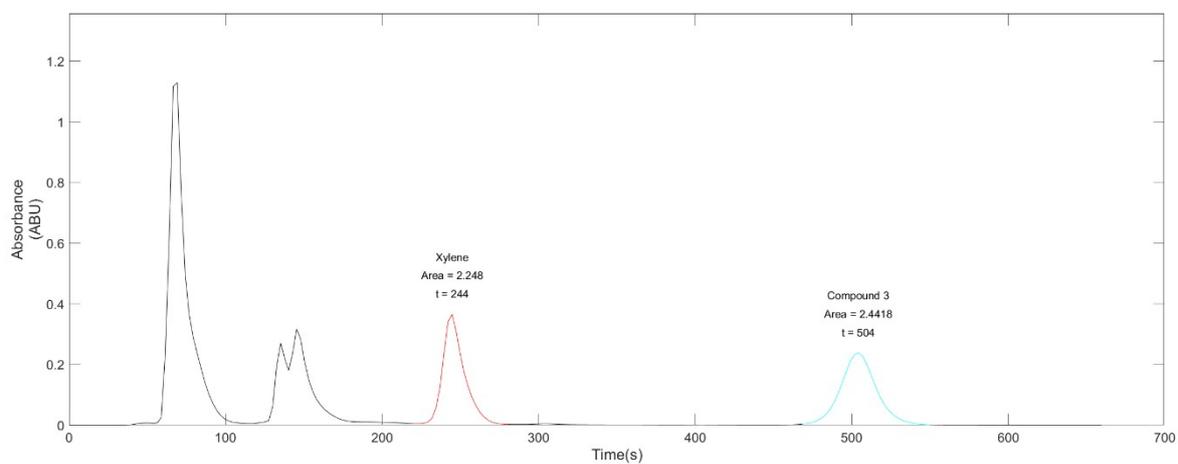
Iteration 6. 17% Yield



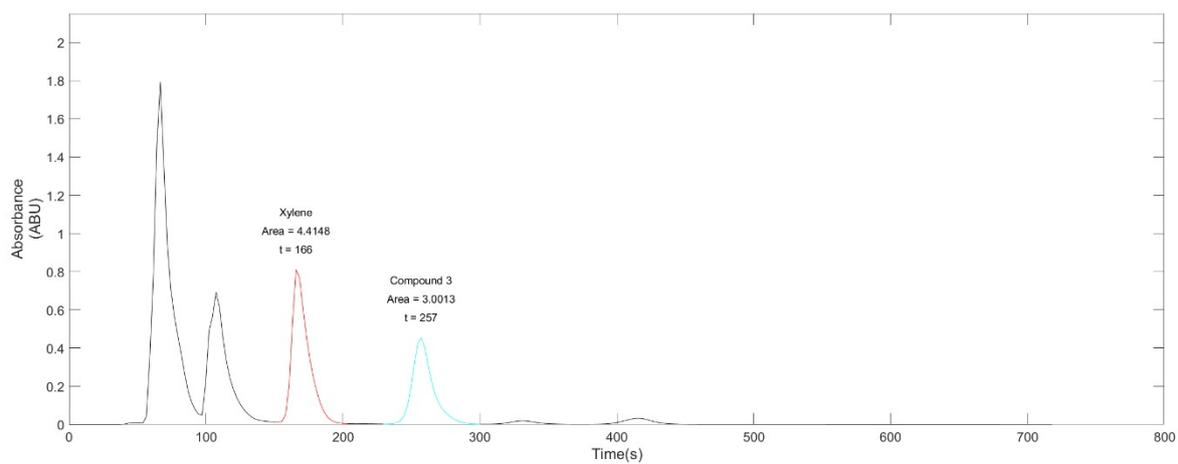
Iteration 7. 25% Yield



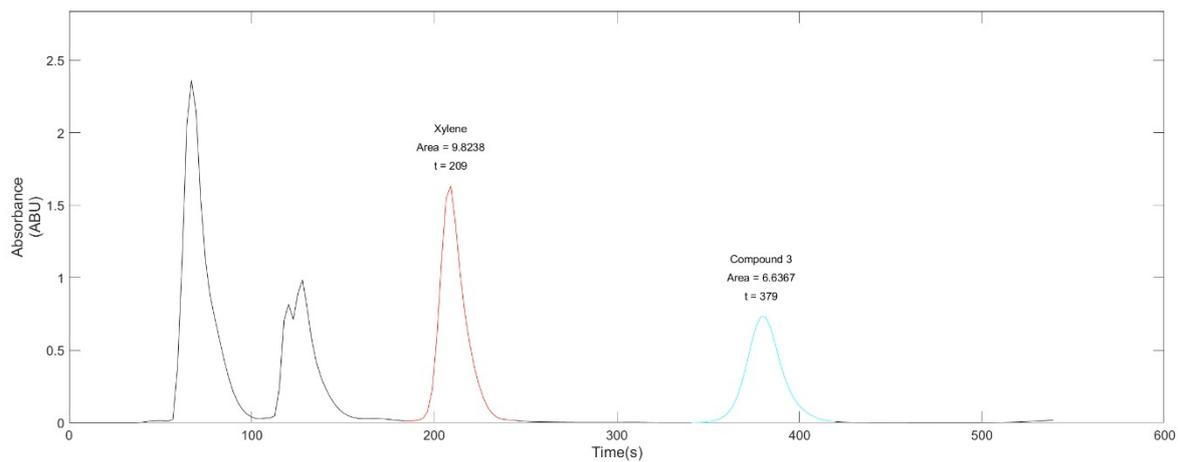
Iteration 8. 32% Yield



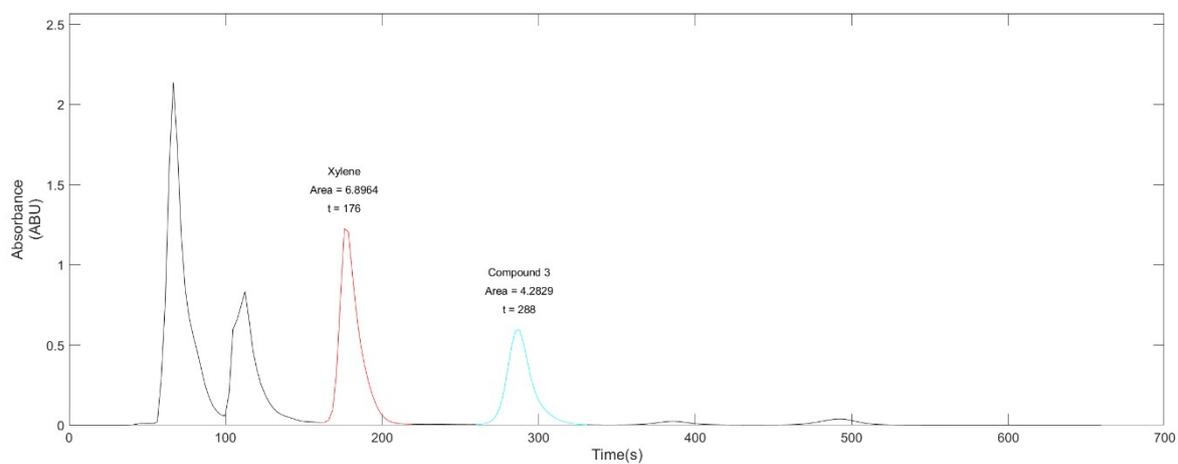
Iteration 9. 23% Yield



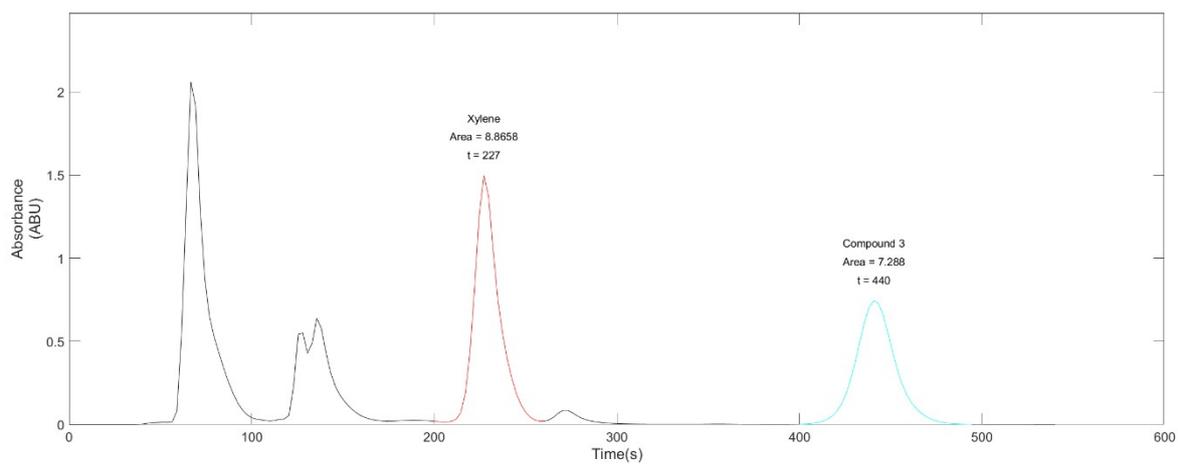
Iteration 10. 28% Yield



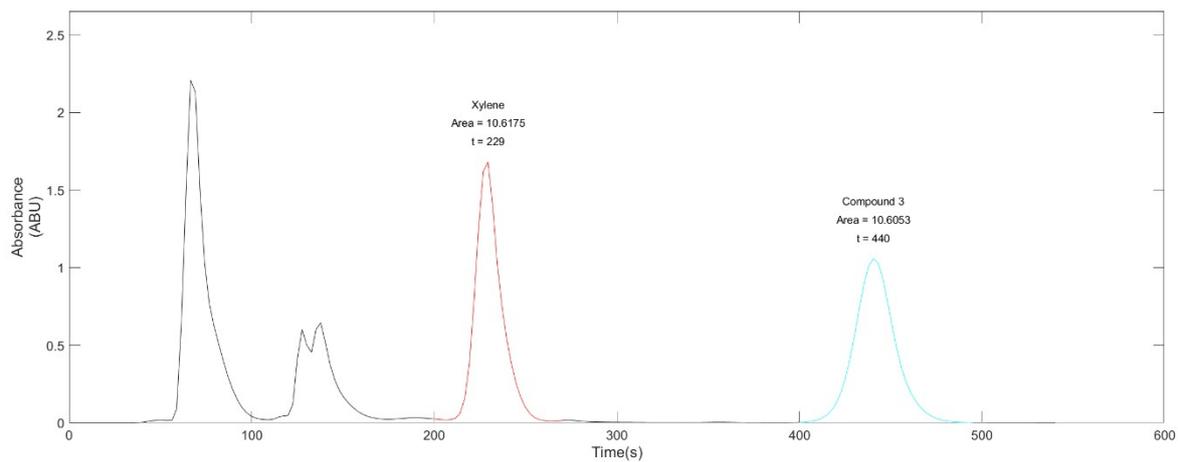
Iteration 11. 23% Yield



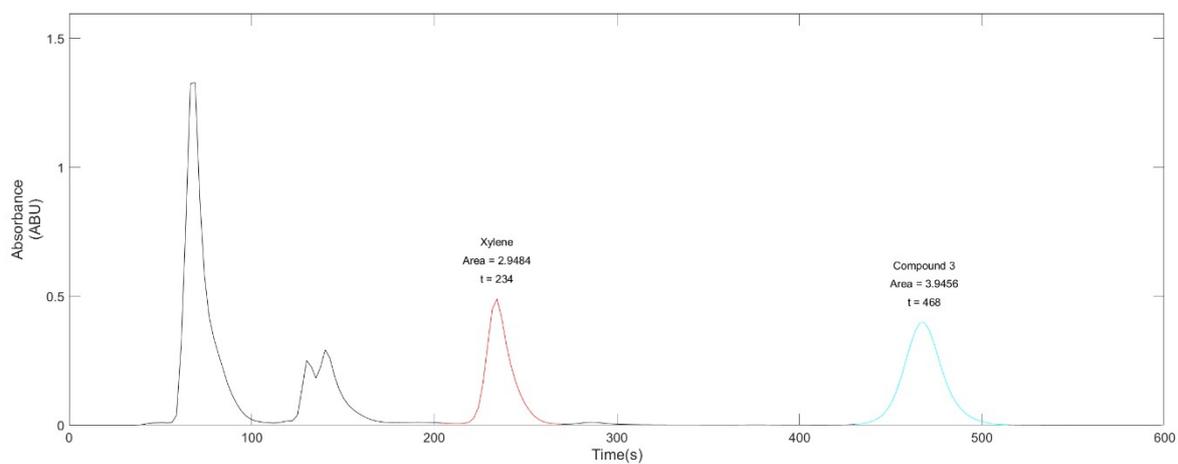
Iteration 12. 33% Yield



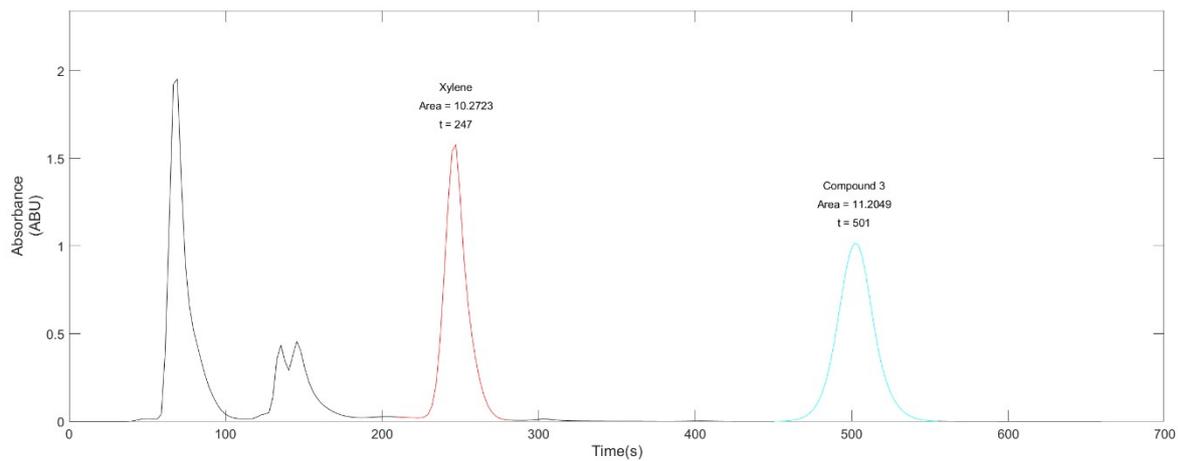
Iteration 13. 44% Yield



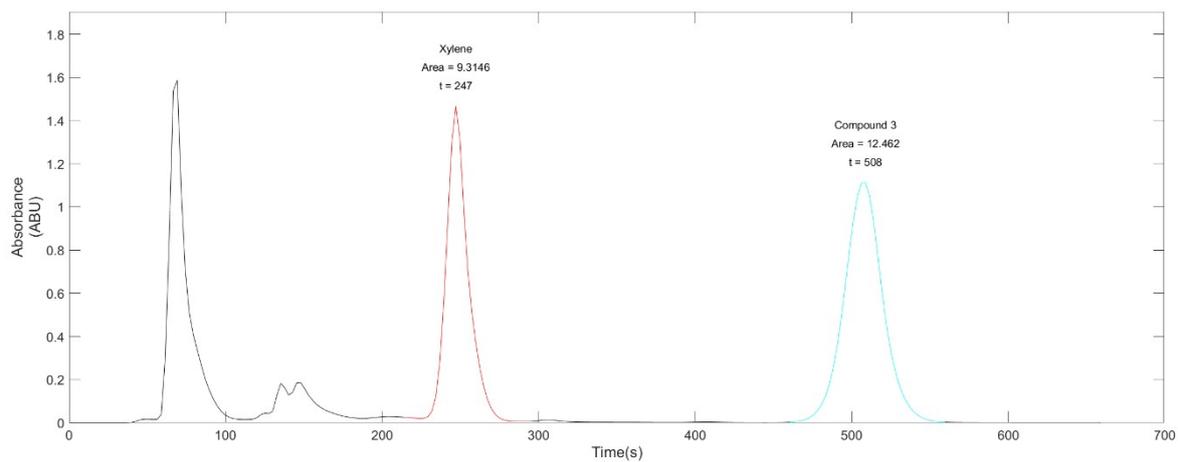
Iteration 14. 42% Yield



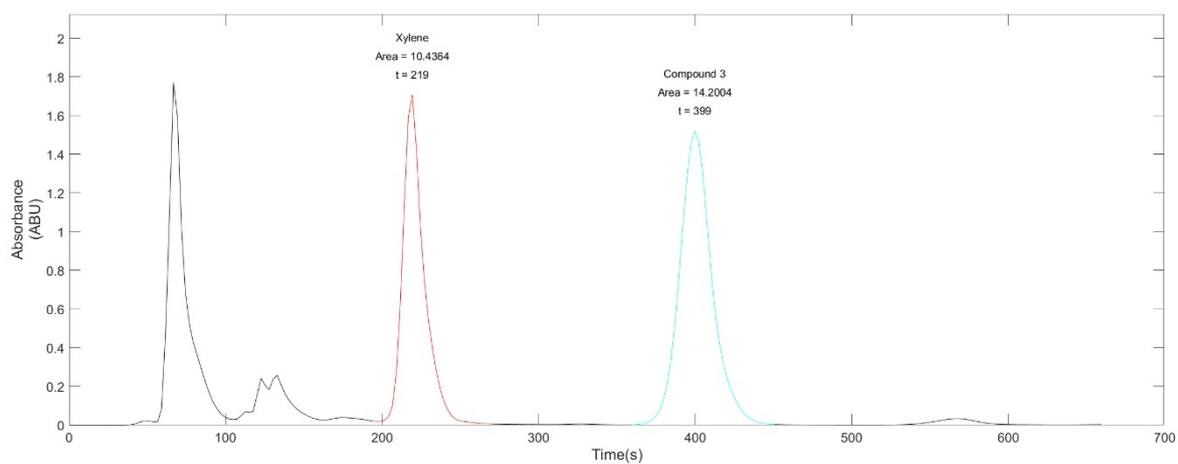
Iteration 15. 47% Yield



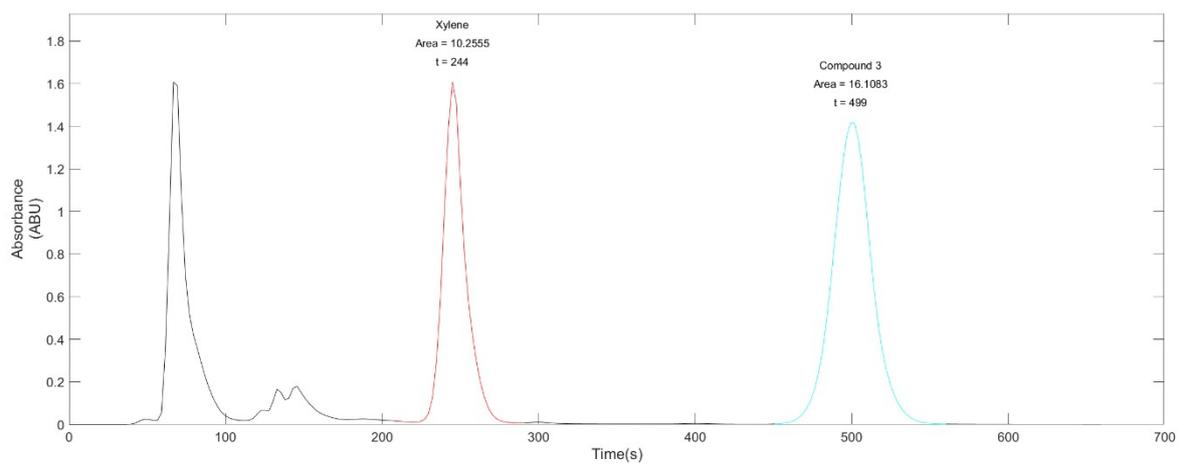
Iteration 16. 58% Yield



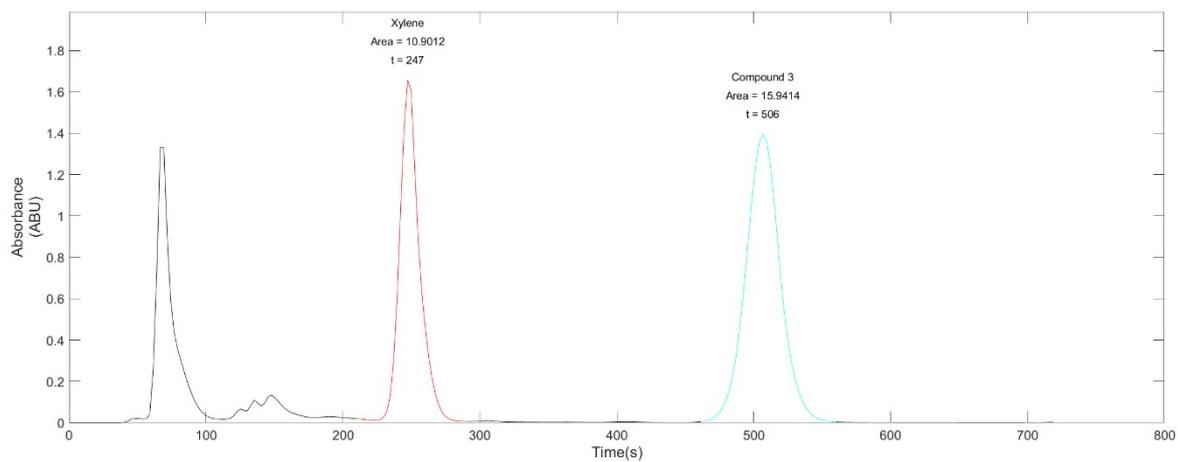
Iteration 17. 62% Yield



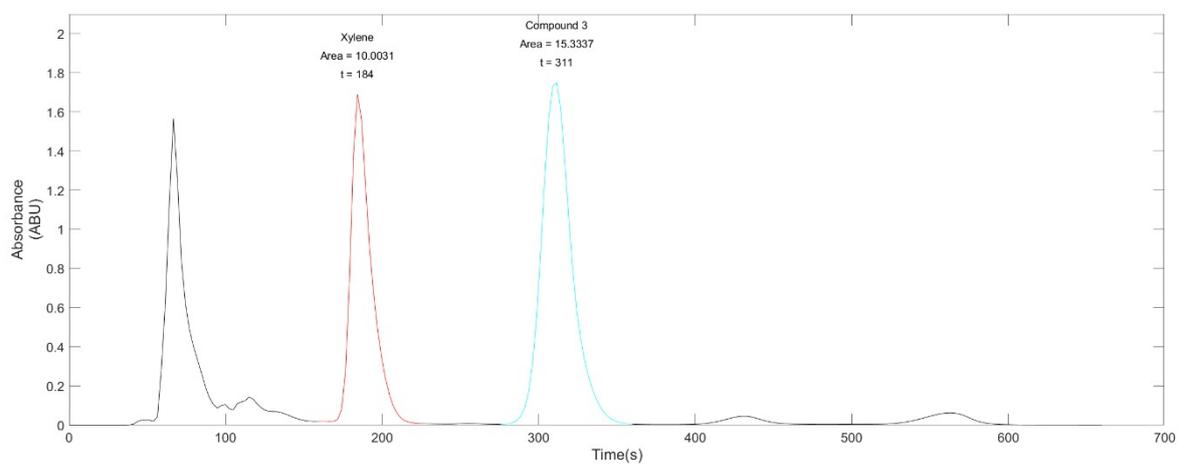
Iteration 18. 71% Yield



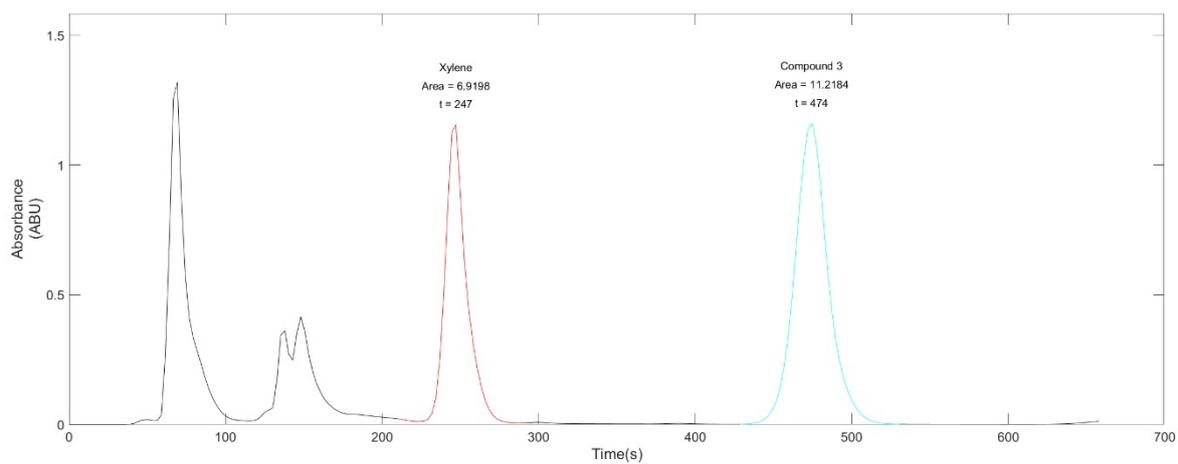
Iteration 19. 67% Yield



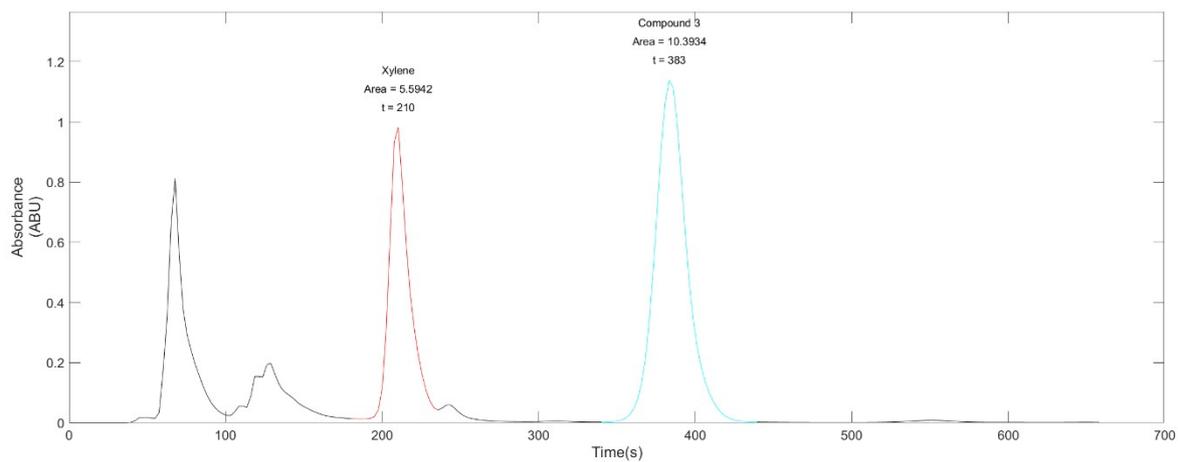
Iteration 20. 70% Yield



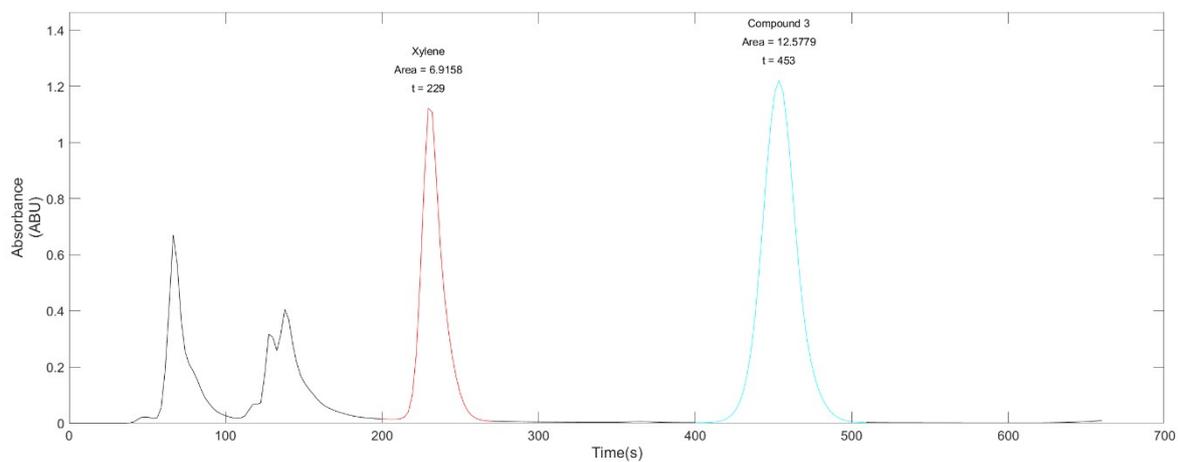
Iteration 21. 65% Yield



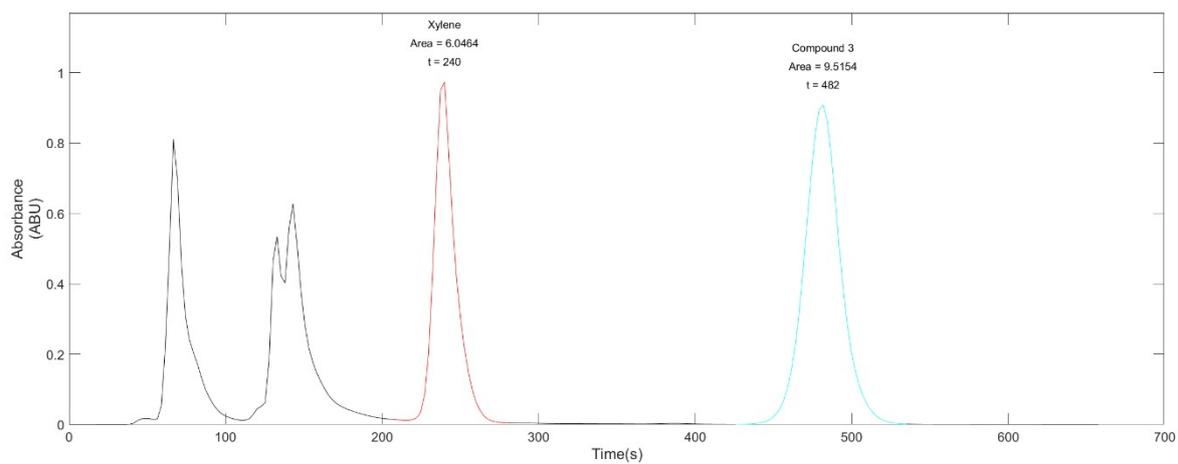
Iteration 22. 70% Yield



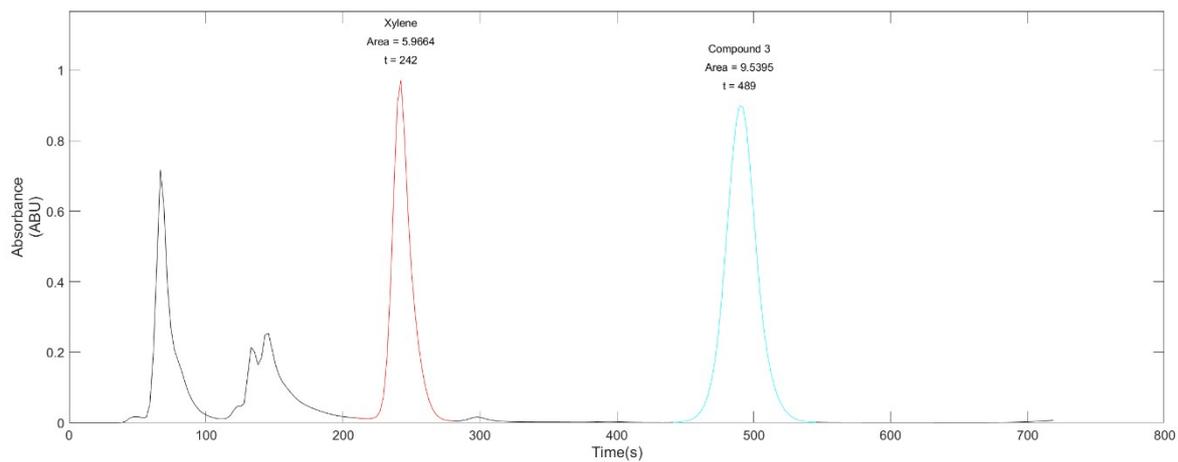
Iteration 23. 71% Yield



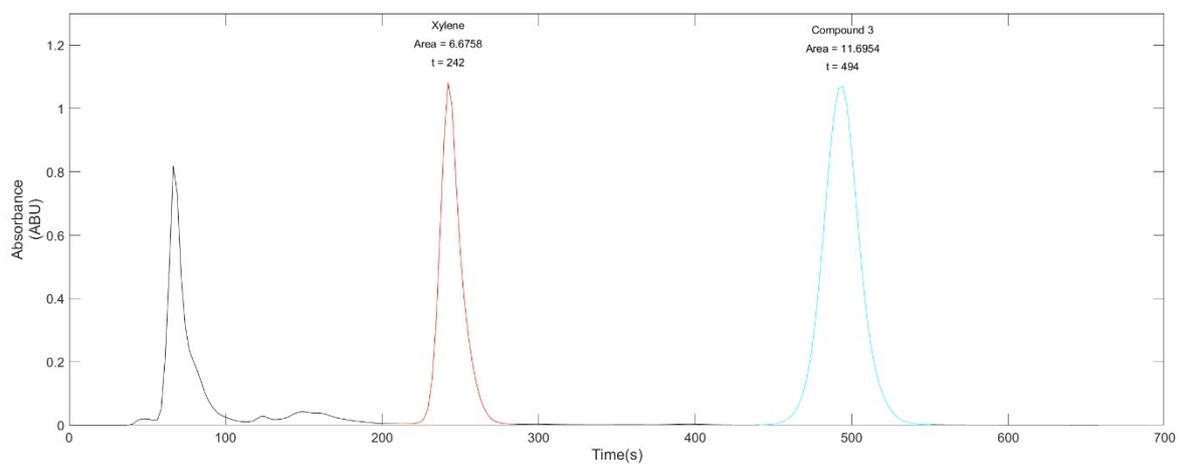
Iteration 24. 59% Yield



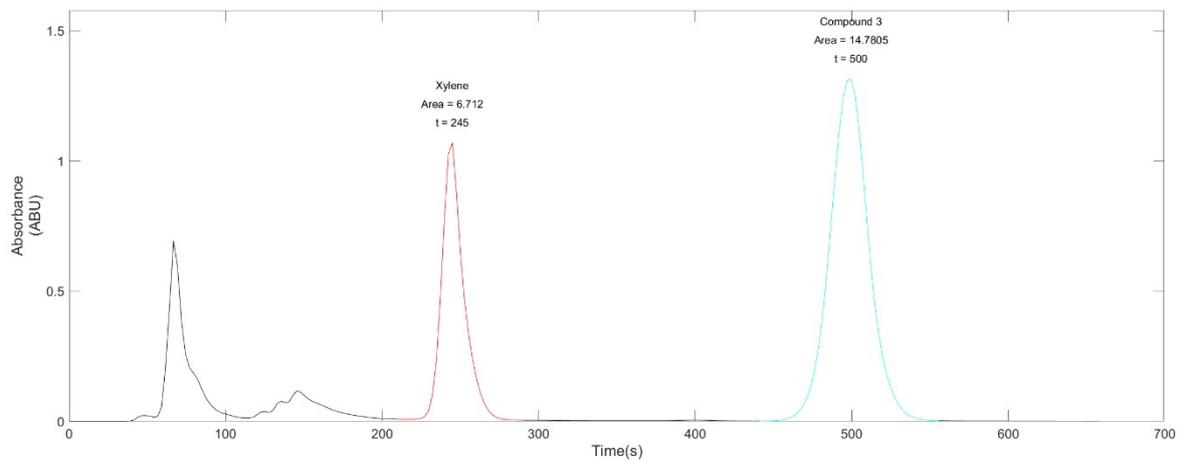
Iteration 25. 60% Yield



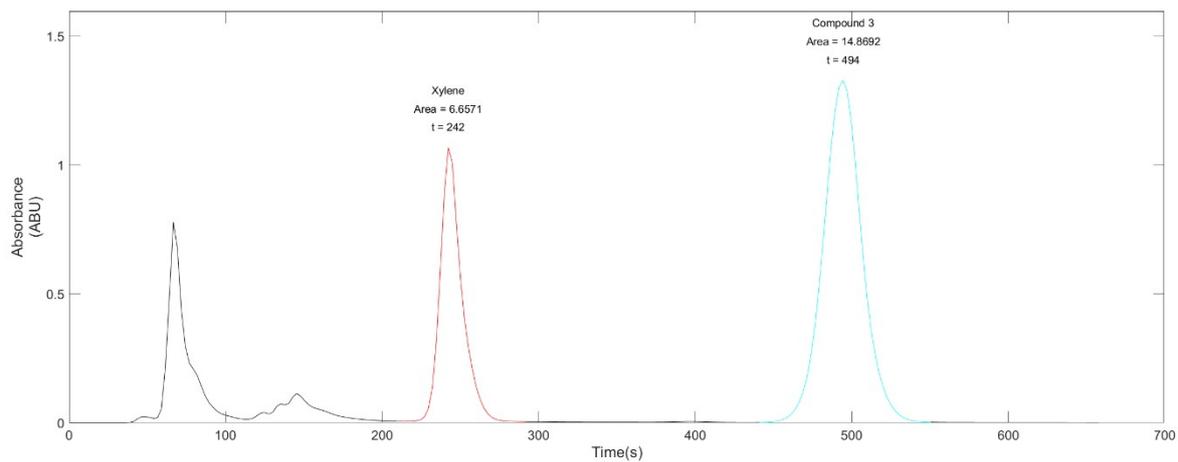
Iteration 26. 69% Yield



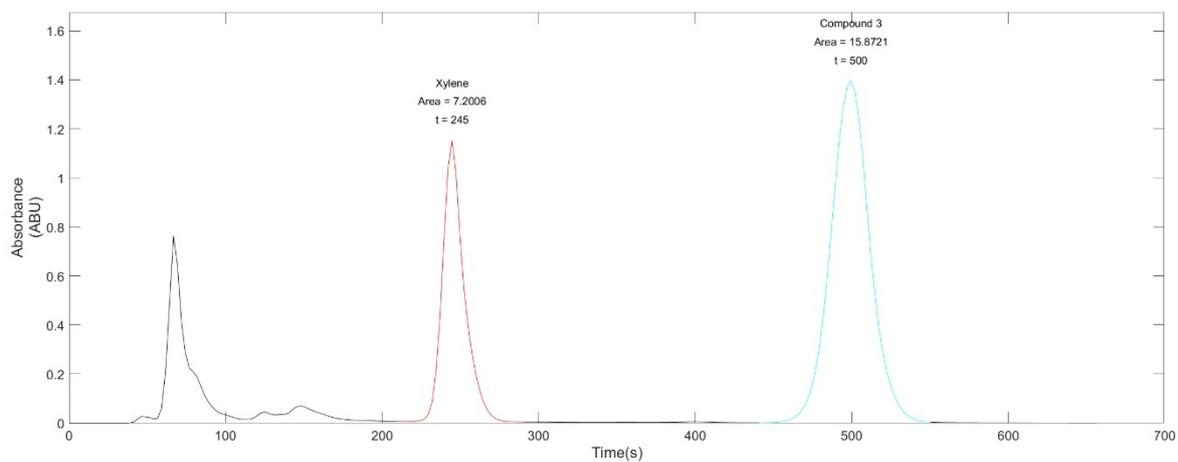
Iteration 27. 88% Yield



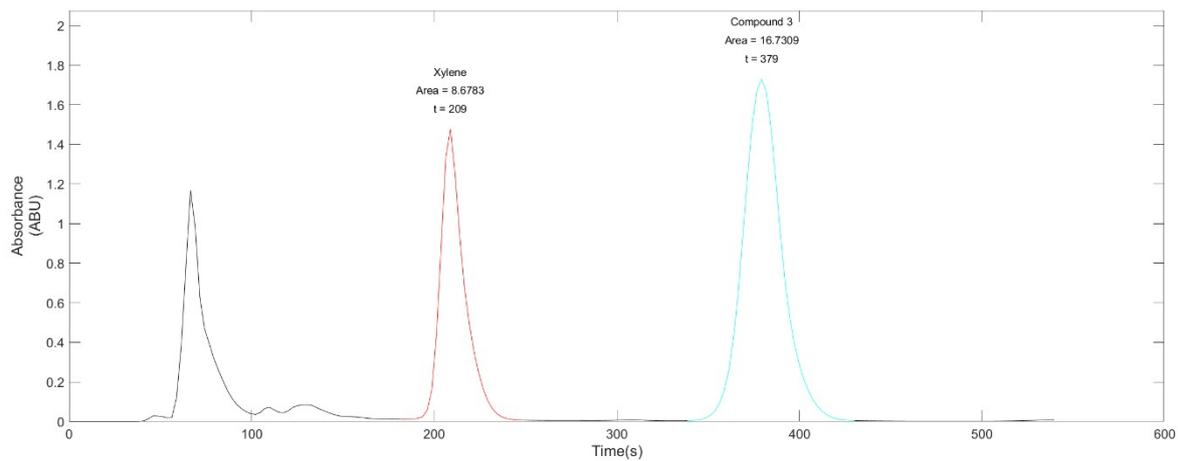
Iteration 28. 89% Yield



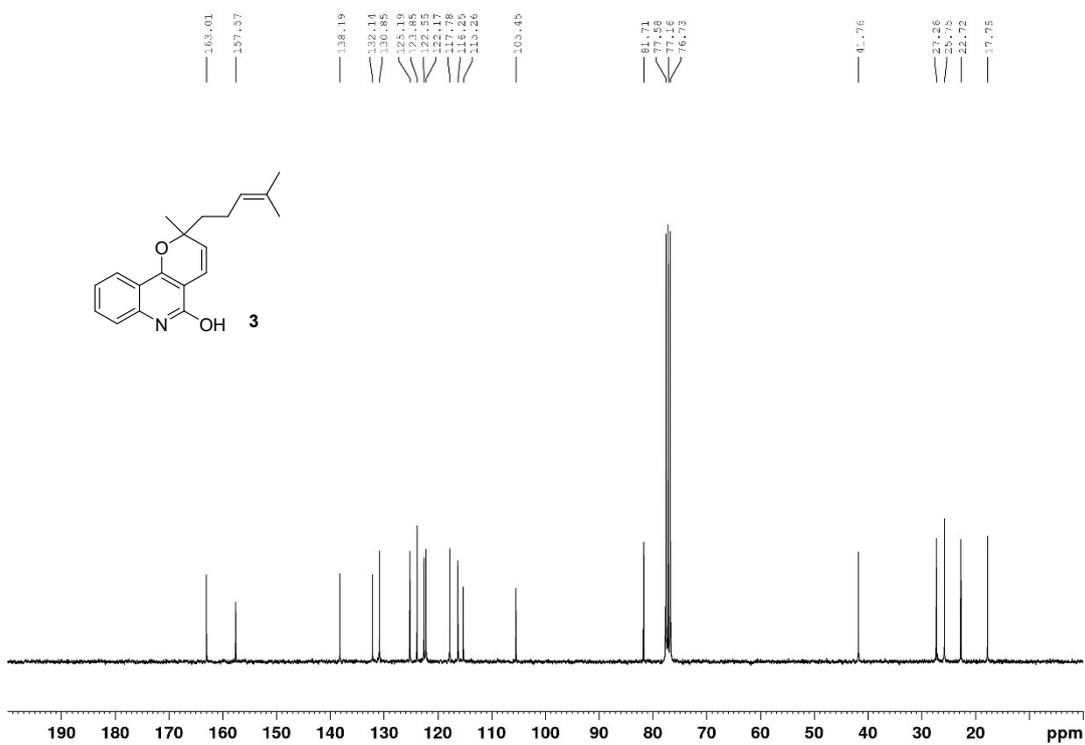
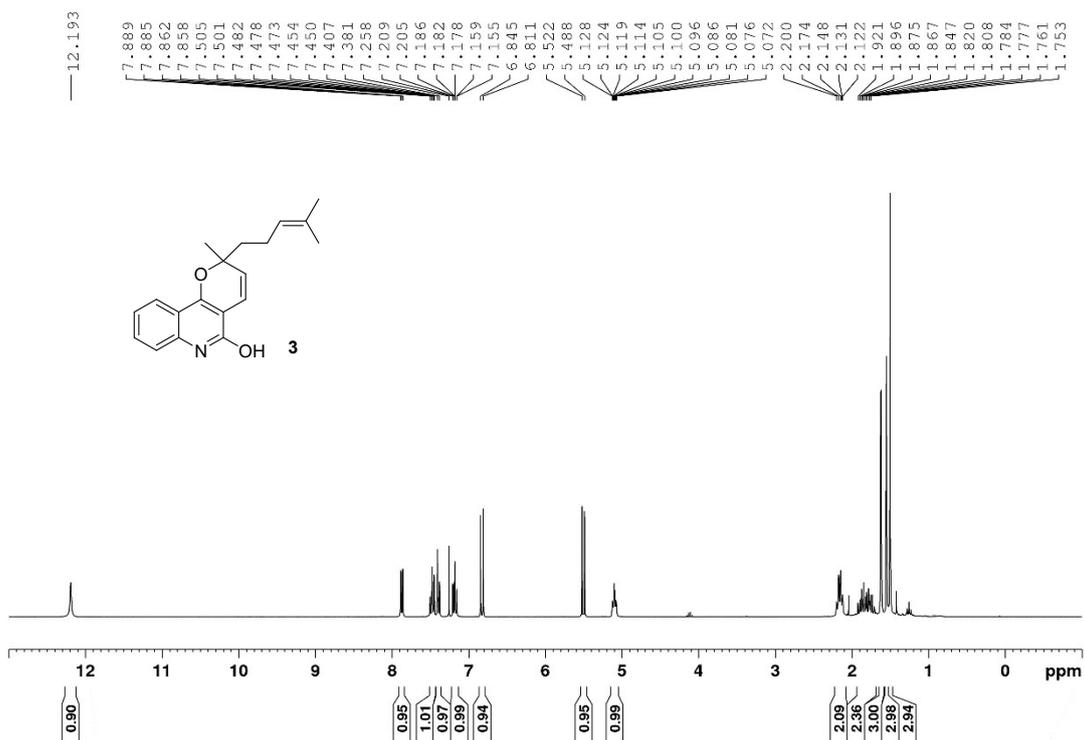
Iteration 29. 90% Yield



Iteration 30. 85% Yield



5. NMR spectra



6. References

- 1 A. Agrawal, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary and S. R. Kalidindi, *Integr. Mater. Manuf. I.*, 2014, **3**, 90–108.
- 2 D. Cortés-Borda, K. V. Kutonova, C. Jamet, M. E. Trusova, F. Zammattio, C. Truchet, M. Rodriguez-Zubiri and F.-X. Felpin, *Org. Process Res. Dev.*, 2016, **20**, 1979–1987.
- 3 D. Cortés-Borda, E. Wimmer, B. Gouilleux, E. Barré, N. Oger, L. Goulamaly, L. Peault, B. Charrier, C. Truchet, P. Giraudeau, M. Rodriguez-Zubiri, E. Le Grogneq and F.-X. Felpin, *J. Org. Chem.*, 2018, **83**, 14286–14299.