Optimising a Cham-Lam coupling reaction using automated synthesis.

Cham-Lam cross-coupling is a versatile reaction, which has advantages over Buchwald coupling as it can be done in air or room temperature and is catalysed by copper complexes, which is good as copper is inexpensive. The Technology Enhanced Chemical Synthesis (TECS) CDT 2021 cohort undertook the challenge to optimise the conditions for the Cham-Lam aminiation of *p*-tolylboronic acid with three chosen amines: *para*-anisidine, *meta*-anisidine and *para*-bromo aniline. In this blog post, we'll go through the process using the *meta*-product reaction scheme.

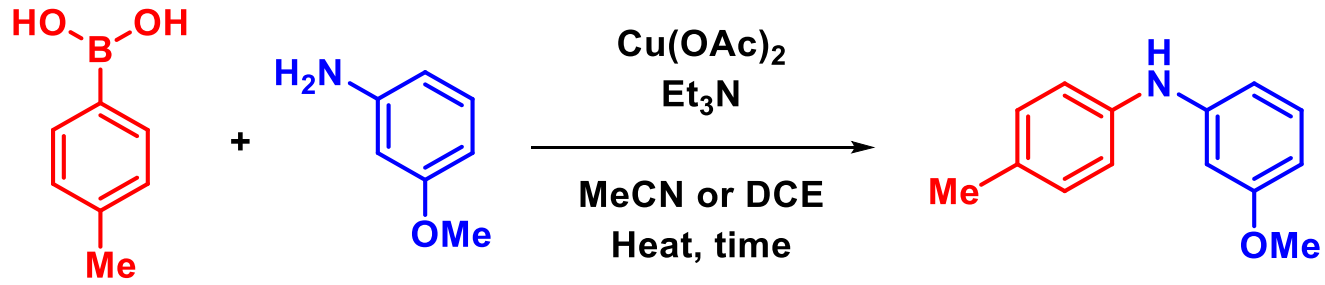

Figure 1. Cham-Lam coupling reaction scheme for the meta product.

The reaction proceeds via the somewhat complex boronic acid decomposition pathway, shown in Figure 2 (also see [1,2]). It took some time for people to discover this mechanism!

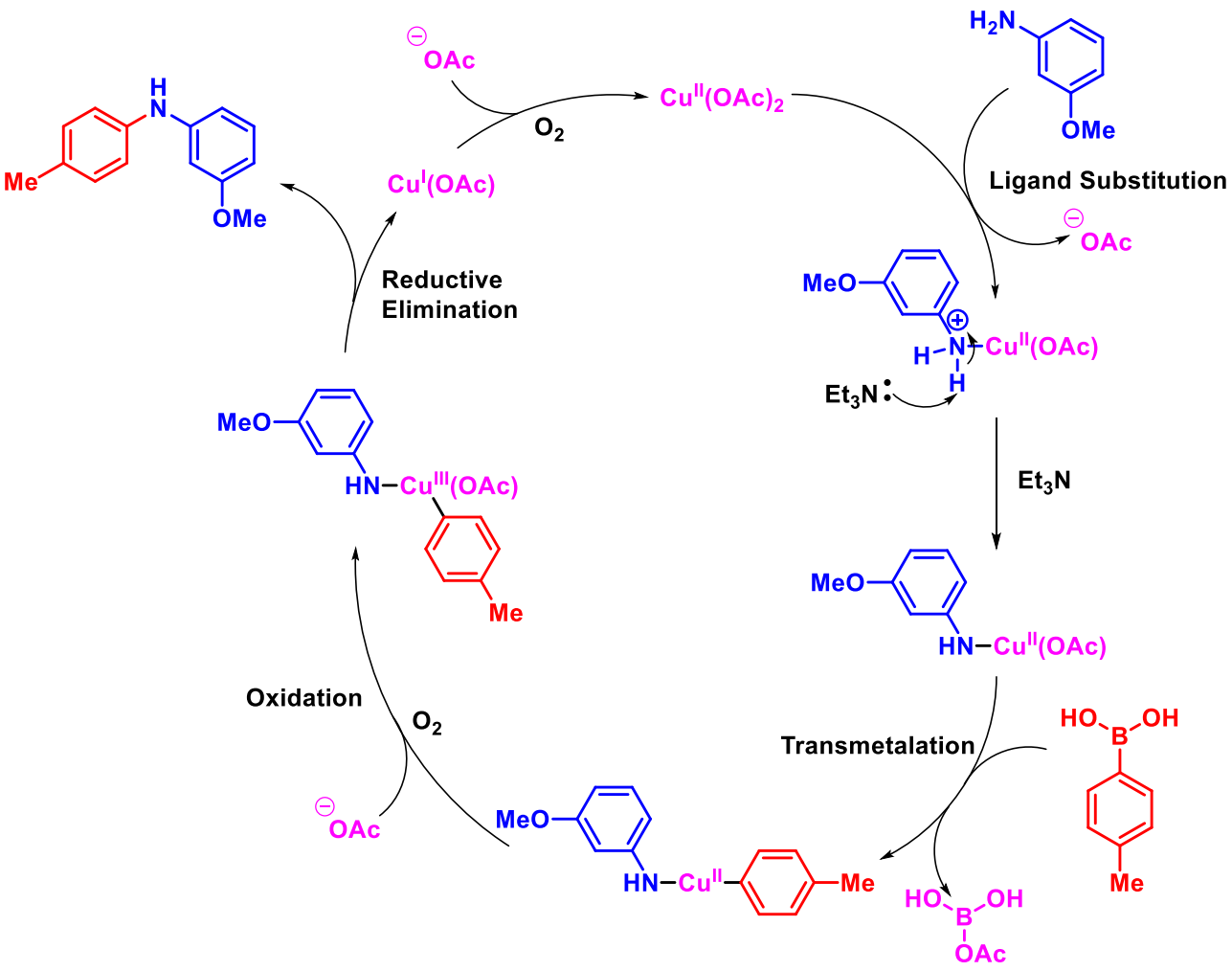

Figure 2. proposed mechanism…

The students first undertook a literature survey to decide which factors would be the most fruitful to investigate. There were many examples in the literature, as expected from such a versatile and cheap reaction, and yields were reported in the range of 21-99% for different reactions. On some substrates, there is a lot of room to improve this reaction! For the reaction optimisation we will discuss here, the yield was **50%**

The students decided to optimise:

1. **Boronic acid equivalent**
2. **Catalyst equivalent**
3. **Base equivalent**
4. **Concentration**
5. **Temperature**
6. **Solvent choice (our only qualitative factor).**

The students used Modde to set up a series of experiments that would map the reaction space and allow them to find the best conditions for each reaction they looked at. Performing these experiments happened on the ChemSpeed automated synthesis platform which is nicely set up for running multiple experiments in parallel and reducing human error.

Once they had their results, separate models were built for different time points (to investigate how reaction time might affect yield) and temperatures (as it is easier to run all reactions at the same temperature).

As there are quite a few factors under investigation, the students decided to use a fractional factorial model, which efficiently screens reaction space with only a few experiments, but as it is a linear model it can miss second order effects. However, if there are quadratic terms required to model the reaction space, it will be apparent in the results.

Optimised Linear Model

Models are trained and built in the software package Modde. For the meta-product tested at 23 degrees C, the optimised model exhibited high fitting ability (R^2), high predicting ability (Q^2) and high reproducibility (perfect results for these measurables would be 1.0, terrible results would be 0.0). However, the model validity as calculated by Modde was rather low for the 18 hour model (Modde recommends a model validity of at least 0.25), showing that the model was reproducible but not valid.

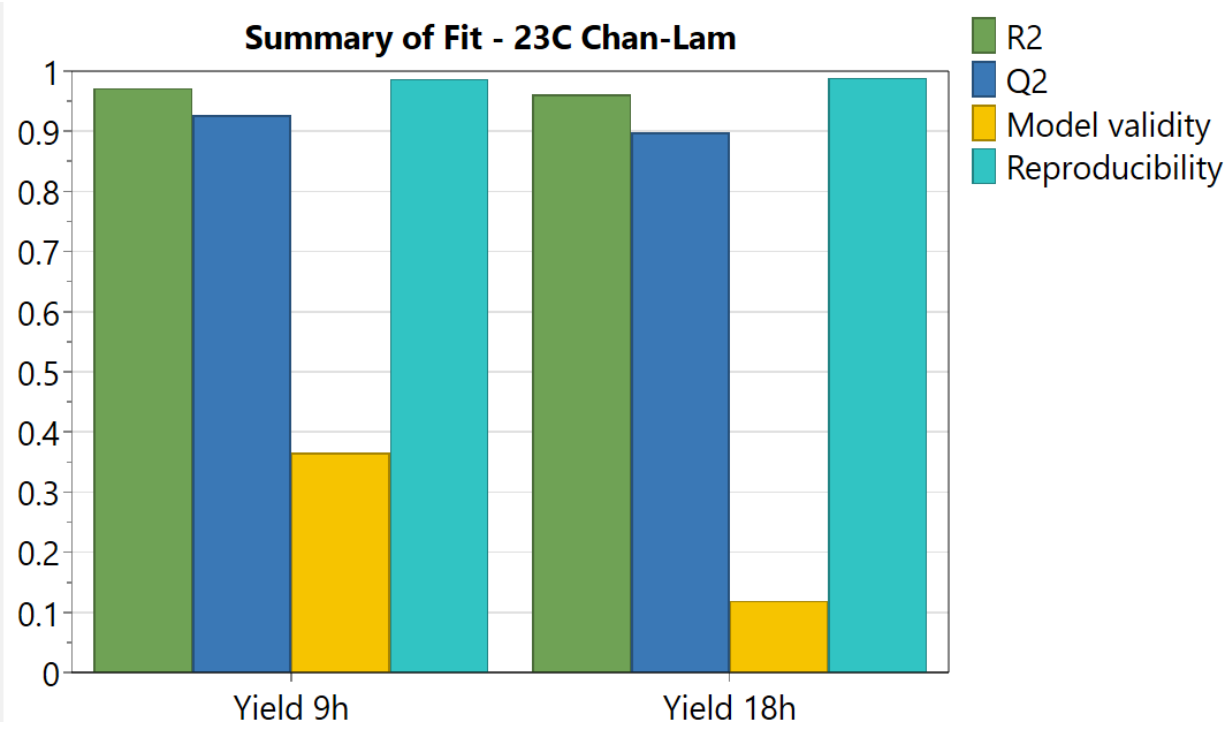

Figure 2. Optimised original model for the meta-product synthesis at low temperature.

Optimised Quadratic Model

As mentioned above, the original model was a linear one, i.e. it contained no square or cross terms. The students decided to see if adding in second order terms to the model would improve the model validity, and they managed to get phenomenally good results doing this! For the long reaction time, for example, they got slightly increased scores for the quadratic model as for the original model shown in figure 2: a fitting score of $R^2 = 0.97$, predicting score of $Q^2 = 0.94$ and reproducibility of 0.95**. But, the model validity was increased from around 0.12 to 0.90**! And as such is a much better model.

The original model surface was flat, the new model had a curved surface, an example of part of this surface is shown in Figure 3.

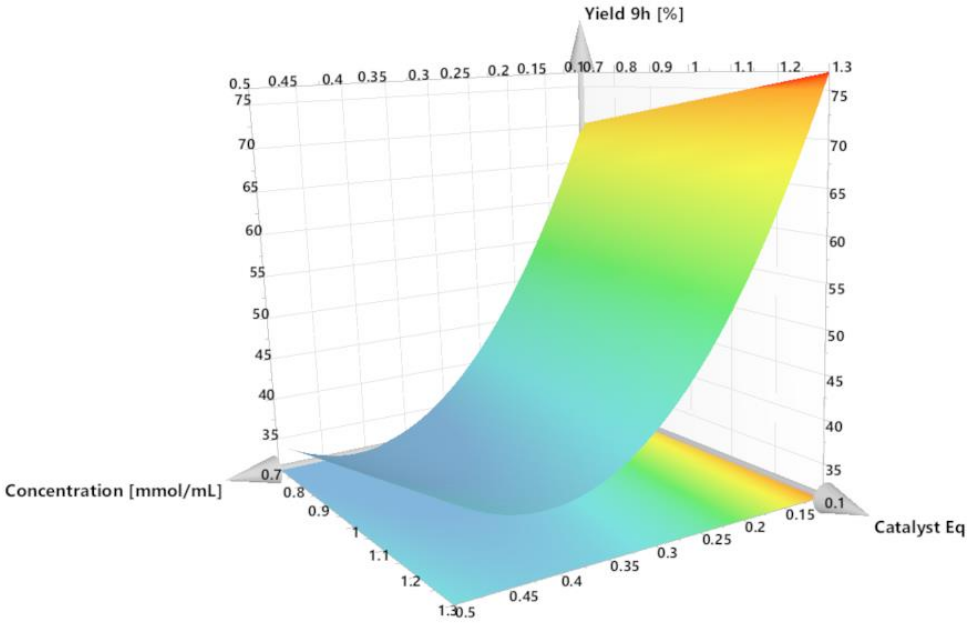

Figure 3. Curved model surface results from including quadratic terms in the model.

Due to the experimental design chosen (fractional factorial) it is not possible to know definitively which square term is responsible. One can find out by performing further experiments to try to disentangle (de-confound) the situation. However, in this field we are usually more interested in building **A sensible model** that can be used to predict reaction conditions to get the best yield, rather than building **THE best model** to explain how the reaction works. To find a valid model, it is a good idea to use your chemical intuition to guide you to pick which terms you think would be reasonable, or alternatively try out adding and removing terms to get the most parsimonious model (i.e. the one that balances getting the best results and using the fewest terms).

The students' investigation settled on adding two square terms to the model: **concentration*concentration** and **temperature*concentration**. The relative sizes of the model coefficients are shown in Figure 4. If the error bar on the scaled and centred coefficients crosses the x-axis then the coefficient is not significantly different from 0.0, so you can safely remove it from the model.

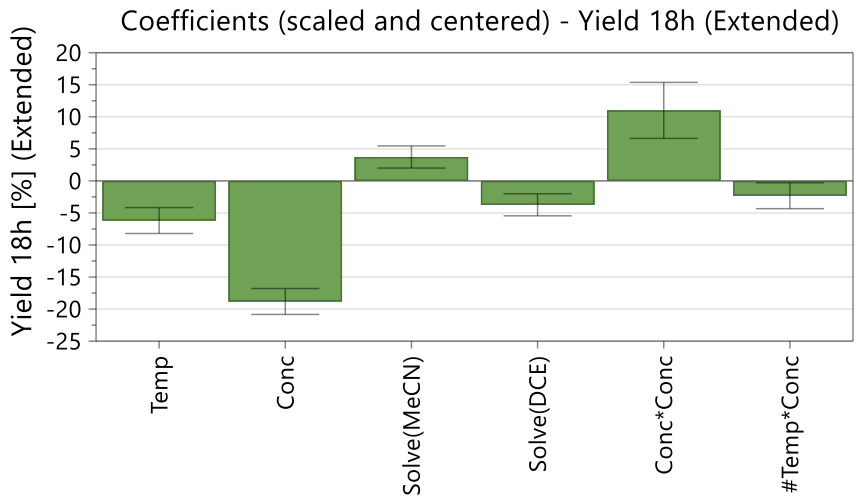

Figure 4. Quadratic model of the Cham-Lam reaction space at low temperature and long reaction time.

The model shows that the base equivalent is insignificant in the range tested over, note that this does not imply that the presence of base is insignificant, just that changing the amount of base won't particularly affect the yield. Similarly, the boronic acid equivalence is insignificant over this range. The quadratic model **predicted that a yield of 76%** was the maximum possible in the range tested over.

In design of experiments the test of a model is how well it predicts rather than how well it is fitted. To get predictions, the students plotted 4-D response contours as shown in Figure 5 (the four dimensions are temperature and concentration on the axes, catalyst equivalent for the separate slices through reaction space and the yield as indicated by the contours). The quadratic terms have added curvature, and as it seems that the concentration and temperature cause a large amount of variation in the predictions, using the concentration based square and cross terms chosen seems like a good choice.

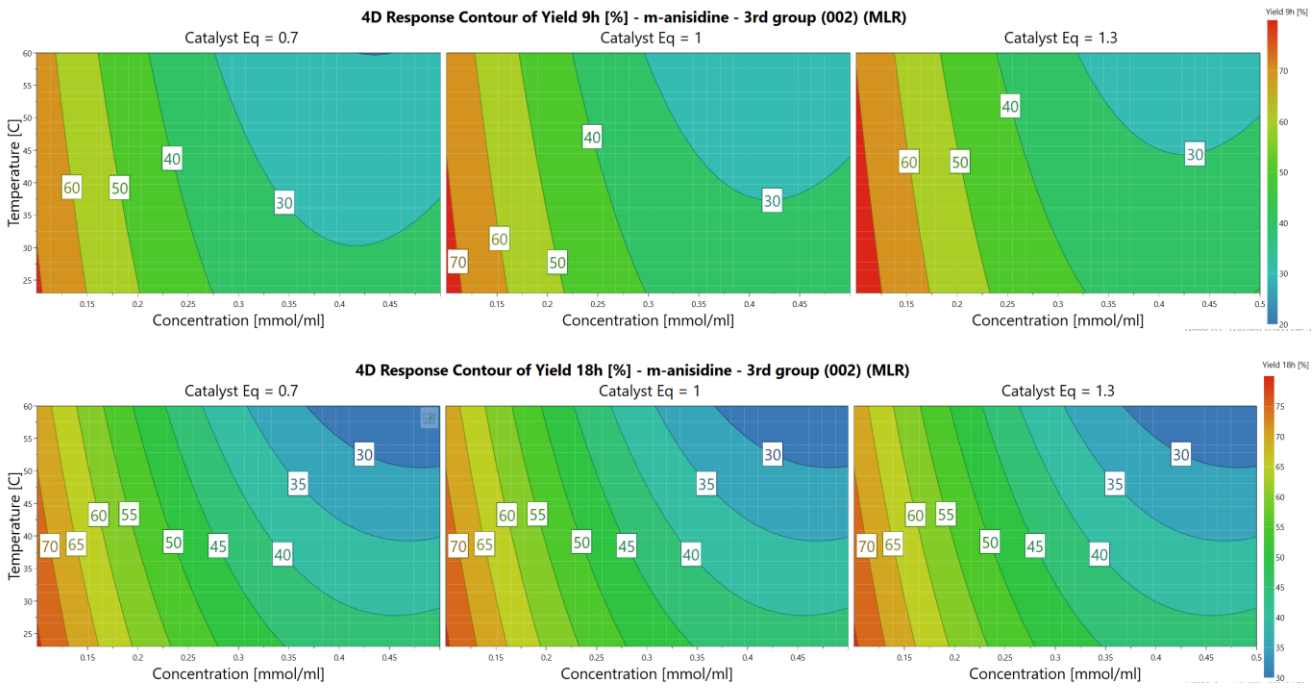

Figure 5. Response (yield) contours.

From the contours in Figure 5, it seems that the longer reaction time, lowest concentration, high catalyst equivalent and low temperature should give the highest yield of product.

The next step is to verify the model by seeing if its predictions will give an increased yield. Figure 6 shows the yields for the replicate experiments done at the conditions chosen. The yields for both reaction times have increased, as the best yield for this reaction in the literature was 50%, both reactions offer an improvement over the literature.

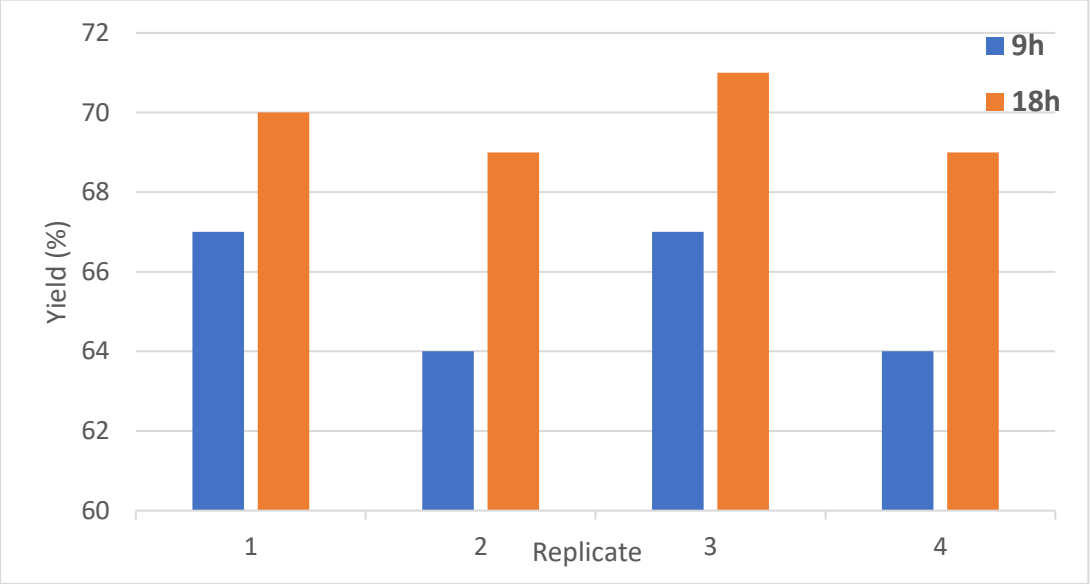

Figure 6. Trying out the predicted reaction conditions.

Results:

Reaction optimised from 50% yield to 70%!

Model proven correct as model predicted a yield of 76% and 70% was measured experimentally.

Further work

At this point the students finished the project, however, given that the best results were found at the edge of the range tested, if they had had more time the next move would be to repeat the optimisation process moving the range of tested temperature and concentration values even lower.

As the students were concerned with the best possible yield they picked the longer time, giving the optimised reaction conditions of low concentration, low temperature, high catalyst equivalent and 18 hour reaction time. There were other valid choices that could have been made here. For example, in industry, it might be more desirable to run the reaction for only 9 hours at the cost of reducing yield. Or, from figure 5, the model suggests that good yields could be found with a lower catalyst equivalent, which could be cheaper (although copper catalysts are cheap). It is possible to do DoE where you optimise more responses, so for example you could include cost, variability in yield as well as yield.

This blog post covers only one of the reactions optimised by the students of the TECS 2020 cohort, two other reactions were done and we are drafting a paper based on this work.

This work was done by: (students in alphabetical order) Grace Boden, Kaiman Cheung, Sarah Coppock, Malcolm George, Josie Harcourt, Emma Hollis, James Mortimer, Dylan Rigby, Isobel Scott Douglas and Borys Banecki (specialist technician). Blog post was written by Ella M Gale (ML subject specialist for TECS).

References:

[1] A.Reddy, K. Reddy, D. Rao, C. Jaladanki et. al., Org. Biomol. Chem., **2017**,*15*, 801-806

[2] G. Zhang, Y. Li, J. Liu, RSC Adv., **2017**, *7*, 34959