

## PROCESS DEVELOPMENT NEAR PRODUCTION

This article describes how an esterification reaction has been transferred from pilot plant to commercial scale production at Lonza, Switzerland. The optimization of the process based on a simple statistical design of experiments and carried out in the FlexyLab laboratory automate is described. It is shown, how an optimization plan for large-scale production could rapidly be performed (in less than one week). Moreover, it is shown, how a good chemical understanding of the process could be achieved by allocating a minimum of resources.

Different drivers and features from each development stage are expected for a process being developed in the fine chemical / pharmaceutical business (see Table 1). At the beginning, when a molecule has been selected for preclinical / phase I trials, speed of delivery is by far the most important factor. At the other extreme, when a molecule is close to industrial production the value creation is the dominant factor. In such a situation, the processes are often changed / transformed into a so-called 'second generation process' where all the economical factors are taken into account. However, due to time to market pressure and regulation constraints, this second generation development must be fast and efficient. Appropriate laboratory automates are very useful in such a situation. Lonza has acquired recently the Flexilab laboratory automate (Systag) and this system has proven to be very useful for process development close to production. The advantage of Flexilab in comparison to other small scale automates is that it uses more quantities of reactant useful for a subsequent isolation and it works under realistic conditions (agitation, temperature control,

amount of catalyst. A second experimental design based on an optimization model was undertaken in the aforementioned Flexilab. The result of DOE is presented in this work.

### Equipment and Experimental

FlexyLab is a well-established tool for parallel process development in the chemical synthesis laboratory. Up to 6 independently working reactor units enable rapid and efficient screening of various process parameters to be performed in parallel and always under conditions near production. Slightly varying or completely different reactions can be investigated rapidly and reproducibly. In the case discussed here, four 250 ml standard reactor units with individual gravimetric dosages were implemented. A central thermostat is used as cooling unit. Due to the individual heating units of the FlexyLab reactors, chronologically shifted reactions can easily be applied.

When working with a parallel automated laboratory reactor, it is difficult to use a typical laboratory journal to protocol the experiment. To overcome this problem an electronic protocol is used. The protocol is a simple Excel sheet. This sheet is repeated for every experiment performed in the FlexyLab. It shows not just the experimental procedure but also all the results, curves, analytic etc. associated with each experiment.

We have chosen a full factorial design on 3 levels as the most appropriate design. Although a different design could be used (i.e. central composite) reducing the number of experiments, the availability of in-process analytic (GC) enabled to take samples at the appropriate reaction time reduced strongly the amount of experiments required for a 3 levels full factorial design. The design required normally 27 runs that can be reduced to 9 by the in-process analytics. Indeed 39 experimental points were obtained in 4 days by performing 13 experiments; some additional experiments were included. For this, a so-called basic recipe has been established to subsequently be altered slightly for the other experiments, according to the experiment matrix. The responses analyzed are conversion, yield and one dominant side product (SP). The experiments can also be performed without attendance over night, saving important time which is crucial to these kind of projects. Alarm limits and surveillance routines prevent from or simply record possible deviations from normal experiment runs. This provides a complete outline of the course of the reactions. Possible changes during running experiments provide necessary flexibility during the first few experiment runs.

### Results

Without water scavenger, the reaction time is long and a main detrimental side product (SP) is formed that must be kept low. It is to note that methanol is used here as reactant but also as solvent, limiting the maximum temperature. The main goal of the DOE is to maximize the response yield and the results are presented in Fig. 1 as a response surface area. The outcome is a powerful mapping of the experimental results which shows where to find the maximum but also its inherent robustness.

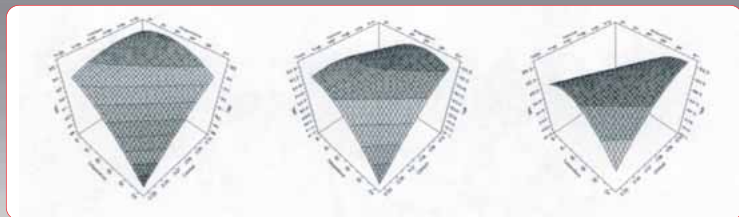
Phase of Development	• Preclinical • Phase I	• Phase II/III • Launch	• Industrial Production
Attrition Rate	Risk: 95-80%	Risk: 70-40%	Risk: > 10%
Features	Non-optimized process Process R&D Kg-Preparation	Validation NDA route cGMP API (HPAI)	COG Competition cGMP
Drivers	Speed Quality Value	Speed Quality Value	Speed Quality Value

**Table 1:**  
Customer expectations during process development in the pharmaceutical business

etc.).

Esterification reactions are routinely performed in the fine chemical / pharmaceutical business. However, although the principles are well established, every reaction must be independently optimized. In our case, the esterification reaction mentioned below had to be transferred from pilot plant to commercial scale production. The initial process had major drawbacks for large scale productions mainly due to the use of large quantities of trimethylorthoformate as water scavenger to shift the thermodynamic equilibrium towards the product. In an attempt to get rid of this reactant a design of experiments based on a screening model was undertaken to evaluate the feasibility of such a process. The initial results were promising and showed the influence of 3 main factors namely temperature, reaction time, and



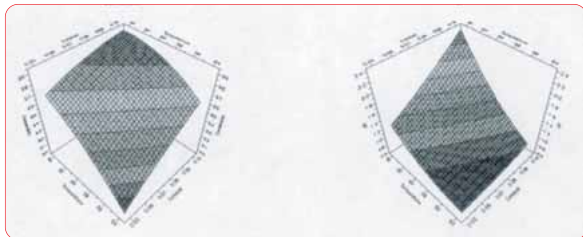


**Figure 1:** Response surface area for yield as a function of reaction duration [h], temperature [°C], and catalyst concentration [mol\_catalyst/mol\_reactant].

The quality of the experimental design can be appreciated by analysis of the reproducibility ( $R^2$ ) and the predictability ( $Q^2$ ) of the model. For all 3 responses (Conversion:  $R^2 = 0.99$ ,  $Q^2 = 0.98$ ; SP:  $R^2 = 0.99$ ,  $Q^2 = 0.97$ ; yield:  $R^2 = 0.98$ ,  $Q^2 = 0.97$ ) the values are over 97% which demonstrates the very good fit of the model. For an optimization design it is essential to have a high predictability ( $Q^2$ ) which is the main goal of such a design.

With a simple empirical model it was possible to model and determine the optimal operating range. The second question now is: How can we interpret in chemical terms the experimental results? By looking at the other two responses namely conversion and side product (SP) one realizes quickly the system behavior (Fig 2). The conversion increases with catalyst amount and temperature, however, when it starts to flat down the side product appears significantly.

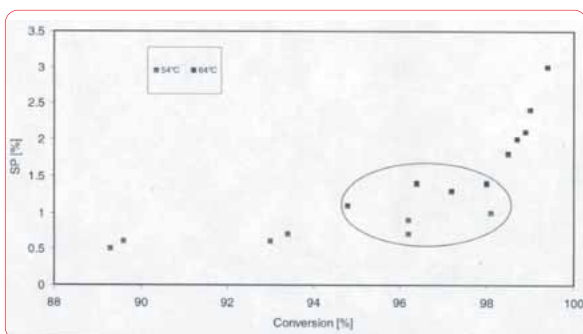
**Figure 2:** Response surface area for conversion (left) and SP (right). Duration = 8 h



The conversion is a measurement of the remaining quantity of reactant while the side product measures what is lost during the process. To sum up, the yield is an indirect measurement of the quantity of side product and reactant present in the process. Such a behavior can be interpreted in terms of 2 consecutive serial reactions as shown here:



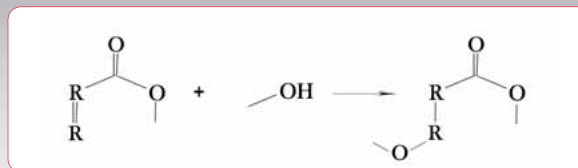
Such a reaction system is well analyzed when the formation of side product is plotted as a function of conversion. This is done in Fig. 3 for the high and low temperatures. For both temperatures, in the region where the conversion are similar, one can observe that at the higher temperature the formation of side product is more important. In other words, the activation energy of the side reaction is higher than the activation energy of the main reaction and higher temperatures favor the selectivity to more side products at high conversions. This confirms the trend observed in Fig. 1, where the optimal temperature range decreases when the reaction time increases. The longer you perform the reaction, the more can the reaction temperature be reduced to increase the yield.



**Figure 3:** Side product (SP) as a function of conversion.

The chemical identity of the side product was identified and indeed it refers to a product molecule that is further alkylated with methanol as represented in the following scheme:

It confirms the mechanism deduced by interpreting the results from the experimental design.



Once it was defined in which plant the process would be conducted it was possible to evaluate the duration of the full process. This step (esterification) is indeed not the bottleneck step unless the reaction time is more than 12 h. However, the usefulness to perform the reaction over a longer period was previously demonstrated: it permits to reduce the reaction temperature and increase the total yield. In such a case, it is valuable to perform the reaction in ca. 12 h.

The experimental design has been conducted with reaction times between 6 and 10 h. It is uncommon to extrapolate such a model in a domain outside the experimental design, as it is the case in Fig.2 with a reaction time (duration) of 12 h. We however believe, that such an extrapolation is acceptable in this case for the following 3 main reasons:

1. The model has a very high predictability ( $Q^2 = 0.97$ )
2. The extrapolating domain is close to the experimental matrix (extrapolating up to 20 h would make no sense).
3. A chemical understanding of the system has been worked out which gives a global comprehension of the system even outside the experimental domain.

The next step was than the direct transfer into production. The reaction is slow and no mass transfer limitations are envisioned – in other words, no scale-up effects are expected. The empirical model that has been evaluated gives a very good indication of the real situation even in production. Under such circumstances, an intermediate piloting was unnecessary and a direct technology transfer from FlexiLab (250 ml reactor) into a large-scale campaign (10 m<sup>3</sup> vessel) has been achieved. The yield during the campaign was the same as those predicted by the empirical model.

### Conclusion

With a simple experimental model it was possible to obtain a complete reaction optimization suitable for typical conditions in production. A good chemical understanding of the chemical process could also be achieved. Similar results could also have been obtained with a detailed kinetic analysis with the drawback of requiring quite more experiments and analysis time. The uses of a statistical design of experiments combined with a laboratory automate such as the FlexyLab permitted a very rapid reaction optimization giving a good qualitative understanding of the process by allocating a minimum of resources. The optimization plan for a large scale production was performed in less than one week.

Moreover, it seems important to mention the modularity of the applied automation solution. This is to enable a quick and cost effective adaptation of the system to often more complex reactions. The combination of statistical experimental planning with the Flexlab multi-reactor system for the optimization of reactions, represent a powerful tool to meet time to market pressure in development departments with fast project/turnover rates and needs for high value creation.

### Author:

Dr. D. Roberge  
Lonza Exclusive Synthesis  
R&D, Reaction engineering

### Contact:

Marcel Haugg  
SYSTAG – System Technik AG  
m.haugg@systag.ch