Organic Process Research & Development **2001**, *5*, 308−323

Efficiency by Design: Optimisation in Process Research

Martin R. Owen,* Chris Luscombe, Lai-Wah Lai, Sonya Godbert, Derek L. Crookes, and David Emiabata-Smith
Chemical Development, GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, United Kingdom

Abstract:
Experimental design (Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; Wiley: New York, 1978 and Carlson, R. *Design and Optimisation in Organic Synthesis*; Elsevier: Amsterdam, 1992) is an established and proven methodology for product and process improvement in the pharmaceutical industry. This paper presents a step-by-step approach to optimisation of a synthetic transformation using a central composite experimental design, in conjunction with automated on-line HPLC. Highly predictive models for the reaction were obtained using a commercially available software package. [There are many commercially available DOE packages. The software package we used was Design-Expert 5 (DX-5) (http://www.statease.com).] These mathematical models were interrogated to examine the effect on yield and quality under a variety of reaction conditions or constraints. The synergy of experimental design and automation is also discussed.

Introduction

For the development chemist, the optimisation of a synthetic transformation is a key deliverable in the process of developing a chemical reaction. The resulting process must be a robust procedure capable of operating routinely in a manufacturing environment. In this paper, we describe a stepwise procedure (summarised in Table 1) which aims to guide the chemist through the appropriate decision processes to allow a reaction to be progressed from the milligram scale to pilot or production scales. The chemist will be familiar with this decision pathway, but the principle objective of the guide is to

(a) highlight the advantages offered by applying experimental design to the chemist's strategy

(b) show how the information collected offers additional reaction and process modelling opportunities.

In the subsequent sections of this paper, each step is explained and a case study is used to more fully illustrate the ideas and approach. The case study presented describes our optimisation of the desilylation of a silyl ether **1** to give the corresponding alcohol **2**, a key step in the synthesis of a trinem antibiotic[4] (Figure 1).

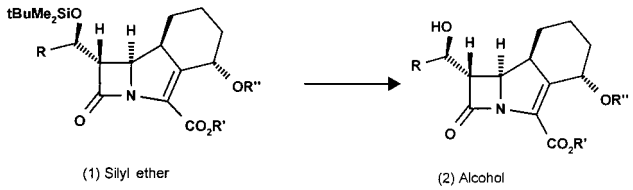


Figure 1. **Case study reaction scheme.**

Appendices 1, 2 are provided to give a glossary of the terminology used in experimental design and some additional comments on the statistical analysis employed.

Step 1: Choose the Synthetic Methodology for Optimisation

The first step is to decide whether to screen alternative chemistries or to optimise the current procedure. This is the most important decision point for the development of the synthetic transformation, irrespective of whether the strategy that follows is a traditional or an experimental design approach.

It is important to realise that any major limitations of your current procedure are unlikely to be significantly improved by experimental design alone. Experimental design is not a substitute for creative chemistry. However, if the only synthetic transformation available was flawed, then experimental design would have its role to play in obtaining the best from it.

After selecting the preferred chemical transformation (e.g., silyl deprotection) it is then recommended to apply screening design as early as possible, to identify the correct choice of discrete variables (e.g. solvent, reagent, and ligand). This is outside the scope of this paper and will be the subject of a future publication.

Case Study

Our original conditions involved deprotecting the silyl ether **1** using excess tetrabutylammonium bromide (TBAB), potassium fluoride (KF), and acetic acid. In the laboratory, these conditions gave us a yield of 75% of the alcohol **2** after product isolation. The yield dropped to ca. 68% when operating on up to a 1000-L scale in pilot plant. This drop in efficiency was attributed to the instability of the product during the protracted work up that was required to remove the acetic acid. An additional concern of the work up was that a considerable amount of waste was generated (150 kg of waste/kg of product) including quaternary amine by-products (5 kg/kg of product). These issues presented us with major long-term waste disposal problems as routine manufacture was expected to be several tens of tonnes per annum.

* Author for correspondence. E-mail: mro1220@glaxowellcome.co.uk.

(1) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; Wiley: New York, 1978.
(2) Carlson, R. *Design and Optimisation in Organic Synthesis*; Elsevier: Amsterdam, 1992.
(3) There are many commercially available DOE packages. The software package we used was Design-Expert 5 (DX-5) (http://www.statease.com).
(4) Ghiron, C.; Rossi, T. The Chemistry of Trinems. *Targets Heterocycl. Syst.* **1997**, *1*, 161−186.

 10.1021/op000024q CCC: $20.00
Published on Web 05/18/2001

Table 1: **Stepwise guide to the optimisation of synthetic reactions**

step	aim	desired outcome
1	choose the synthetic methodology for optimisation	a decision that the chemical transformation is carried out using a defined reagent(s) and solvent(s)
2	define the targets and goals for the optimisation study	a defined success criteria for the study
3	choose the strategic approach (traditional or experimental design)	a firm strategy that will provide the information/data that is required to meet project goals
4	select the critical factors for the procedure (e.g. temperature, concentration etc)	the list of the critical factors is identified
5	define the factor ranges (e.g. reaction time to be studied between 5 and 10 h)	the list of the upper and lower values for each of the critical factors is identified
6	define the scope of the study	a decision to study the whole process or to examine the reaction and work-up separately
7	choose the key responses to be measured (e.g., yield, quality etc)	response measurements are chosen that are meaningful and will indicate progress to the study's goal
8	choose the most appropriate experimental design	the adoption of a design strategy and plan of the experiments
9	consider the practical aspects of implementing experimental design	good reproducibility for measuring the responses (e.g. yield, particle size) to ensure the results are not confounded by experimental error process control of factor to be studied such as rate of addition, temperatures, etc. good control of the variables that are to be held constant
10	implement the design	the execution of experiments according to the design and an accurate record of the corresponding responses
11	consider the statistical implications to the practical implementation	an understanding of any errors or bias incurred whilst implementing the design
12	analyse the results	an insight into the statistical characteristics of the mathematical models (e.g. variability, lack of fit)
13	interpret and interrogate the model	an understanding and thereby control of the process under study. Identification of the important factors and any interaction effects. Identification of the optimal settings to carry out the experiment.
14	verify the mathematical model	a reality check of the predicted outcome of the mathematical model with that obtained experimentally
15	model the Process Deviation Ranges	an understanding of the robustness of the process
16	scale up the procedure.	an assessment of how the mathematical model generated on a laboratory scale, applies to experimental data generated on a larger scale

A program of work to optimise this transformation was initiated, commencing with a traditional laboratory-scale screen of alternative desilylation procedures. This identified triethylamine trihydrogen fluoride (TREAT.HF)[5] in *N*-methylpyrrolidone (NMP) as the most promising reagent/solvent to achieve the desired conversion. The other procedures screened relied on harsh conditions which, when applied to this particular substrate, resulted in significant or complete degradation.

A larger-scale scoping study showed that the potential advantages offered by this new methodology were:

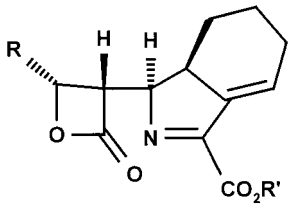
(a) the alcohol **2** was isolated initially in ca. 80% yield.

(b) because acetic acid was not used, it was not necessary to carry out extensive extractions. This reduced the waste generated by 60%.

(c) the amount of amine by-products were also significantly reduced (by 85%).

However, under the new reaction conditions, a new impurity, the lactone **3** (Figure 2), gradually forms which was found to be difficult to remove using the simpler isolation conditions.

(5) Franz, R. *J. Fluorine Chem.* **1980**, *15*, 423; McClinton, M. A. *Aldrichimica Acta* **1995**, *28*(2), 31.



(3) Lactone impurity

Figure 2. **Undesired side-product.**

The amount of impurity that forms was critically dependent on the reaction time. If the reaction is stopped too early the conversion is incomplete. Conversely, if the reaction is stopped too late, the level of impurity becomes unacceptable (see Figure 3 for reaction profile).

Outcome from Step 1. The use of TREAT.HF in NMP offered significant process improvements. However, we need to identify reaction conditions that give an optimal yield whilst ensuring that the generation of lactone **3** is minimized.

Step 2: Define the Targets and Goals for the Optimisation Study

Before embarking on the study, define your project targets and goals by looking at the typical Process Research concerns

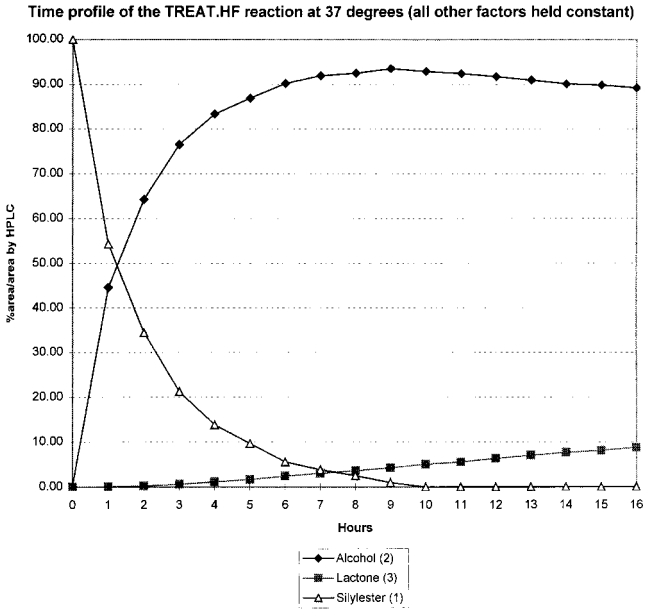


Figure 3. **Time profile of reaction.**

Table 2: **Case study goals**

effect on process	target/constraint
efficiency	maximize the yield of the alcohol **2** (target > 75%th)
product quality	the level of the by-product (lactone **3**) in the isolated product to be <0.5%.
cost of goods	minimize on the use of the TREAT.HF reagent maximize the throughput of the reaction by maximizing concentration and minimizing the reaction time.

(e.g. yield, throughput, cost of goods, environmental impact). It is also important to check that an optimisation study is appropriate at the current stage of the project life cycle. For example, if little is known about the reaction, it may be appropriate to carry out a scoping or a screening study before an optimisation study is attempted. Conversely, if the project goal has already been met, a robustness study would be more suitable. This aspect is covered in greater detail in step 8, Table 4.

Case Study

Outcome from Step 2. The goal for the case study was set as the optimisation of the TREAT.HF procedure using conditions consistent with a viable production procedure operating at tens of tonnes per annum.

The specific targets are quantified in Table 2.

Step 3: Choose the Strategic Approach (Traditional or Experimental Design)

Here, judgment is again crucial as to meeting the goals for the study. The requirement is to balance the type of information to be gained and the amount of resource available. At first glance, the designed approach may appear to require more resources than the traditional approach, because the plan of the experiments is carried out "up-front". However, the key issues for you to resolve are:

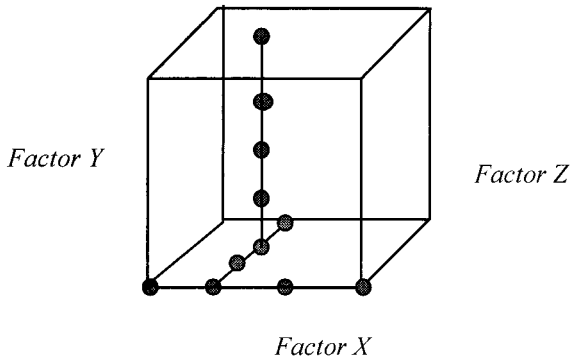


Figure 4. **Diagram of a traditional approach. This shows how the traditional "one-factor-at-a-time approach" is applied to the study of three factors (e.g., *X* = time, *Y* = concentration, *Z* = equivalents of reagent). The diagram illustrates that maintaining two factors constant, whilst varying the third allows the experimenter to obtain only a partial exploration of the overall reaction space.**

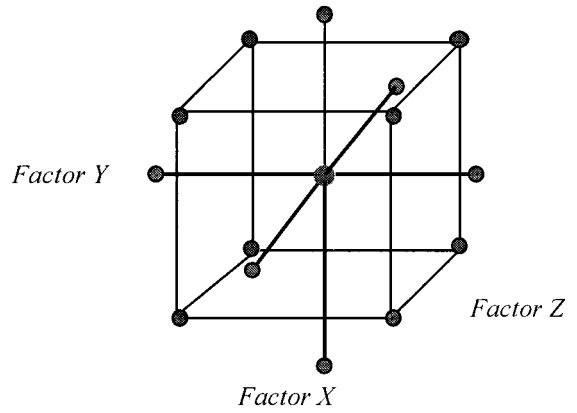


Figure 5. **Diagram of an experimental design approach. This shows a central composite design for three factors. There are 8 cube points, 6 star points, and 1 centre point, totalling 15 different reaction conditions. The centre point is repeated several times to obtain an estimate of background noise.**

(a) What is the risk that you will not obtain the quality of data you require to solve the problem by the traditional approach?

(b) How long will it take?

With experimental design, the difference is that, you will obtain the information (good or bad news) to make your decisions and that the resources and the time scales required can be defined at the outset of the study. Experimental design also allows for additional efficiencies as the reactions can be carried out in parallel.

A Comparison of a Traditional versus an Experimental Design Approach. The traditional approach assesses the effect of one particular factor by keeping all other conditions constant. Then this factor is held at its optimum setting, a different factor is examined, and so on. It is very difficult to apply the approach efficiently to obtain an overall picture of the reaction, as Figure 4 suggests. The result is that an incomplete picture is obtained that completely misses any synergistic effects between the factors. In addition, it contributes little to our understanding of robustness of the process.

In contrast, an approach based on experimental design (e.g., Figure 5, a central composite design) offers a more

systematic approach examining all factors, at several levels over the full range of the "reaction space" you have chosen. Importantly, the data collected from this design also allows the generation of a mathematical model (response surface model) of the chemical process based on the statistical analysis of this set designed experiments. This model then allows the user to analyse the interactions between the factors and hence offers the opportunity to achieve better understanding and control. This model can then be used to identify option and opportunities to optimise and balance conflicting processing constraints (e.g., economic, quality, and environmental).

Case Study

Outcome from Step 3. To learn more about the interactive effects of the factors and resolve conflicting requirements of yield and quality in an efficient manner, a decision was made to adopt an experimental design approach.

Step 4: Select the Critical Factors for the Procedure (Temperature, Concentration, etc.)

In this step it is necessary to predict what are the probable critical factors effecting your process and their relative importance. This is best achieved by carry out a brainstorming exercise, based on the existing level of knowledge.

If a critical factor is not included in the list of potential factors for investigation then diagnostic tools of experimental design (e.g. lack of fit tests and residual analysis) will allow the presence of a hidden variable to be detected.

Case Study

Outcome from Step 4. In the case study, the following four factors were considered to significantly impact yield and quality of the alcohol 2:

(a) equivalents of the TREAT.HF reagent

(b) temperature

(c) reaction time

(d) concentration (volumes of NMP)

Just as important, there were many factors that were held constant:

(e) process control factors: order and rate of addition, rate of stirring, nitrogen blanketing

(f) batches of the substrate, reagent, and solvent

(g) response measurement: the HPLC method was kept constant for all the experiments

(h) the chemist carrying out the experiments

Step 5: Define the Factor Ranges

For each critical factor, choose a range of settings to be studied, based on both chemical knowledge and intuition. Observations from limited scoping studies and pragmatism (e.g. practical considerations in scale-up) are also valuable in setting meaningful ranges.

Note that if the factor ranges chosen are too small, there is the danger that the optimum conditions will lie outside the area of study. Conversely if the factor ranges are too large, the model will not fit the data as well, and therefore will be less predictive.

Case Study

Outcome from Step 5. The following ranges for the factors, listed in Table 3, were chosen to deliver a procedure that would meet our success criteria outlined in Step 2 (Table 3).

Table 3: **Factors studied**

factor	range	units	factor
temperature	10	30	°C
time	19	31	hours
concentration (amount of NMP solvent)	3	7	volumes
equiv of $Et_3N{\cdot}3HF$ reagent	1	1.67	equivalents

Step 6: Define the Scope of the Study (Optimise the Reaction Conditions or Isolation or Both)

The optimisation of a synthetic transformation typically involves the development chemist in two principle areas of concern:

(a) the control of the reaction to maximize unisolated yield and minimize side-reactions

(b) maximizing the benefits of the work-up procedure by taking advantage of any opportunities to remove unwanted by-products, whilst minimizing the losses of the product on isolation

Experimental design offers the opportunity to look at the whole process or at issues (a) and (b) separately.

When it is early in the product's lifecycle, we recommend carrying out a study to establish the best reaction conditions based on the assumption that the purer the process stream, the easier the work-up. Sometimes work-up and reaction are inextricably linked in which case the study needs to cover both work-up and reaction. Consequently, the design will inevitably contain more variables to control and more time-consuming to implement. However, it is good practice then to look at both the work-up and reaction in the verification experiments (experiments that verify the validity of the model).

Case Study

Outcome from Step 6. The scope for the case study was limited to the optimisation of the reaction conditions. This decision reflects that the target for the subsequent isolation stage was to produce alcohol **2** with levels of the lactone impurity **3** at less than 0.5%. The initial pilot experiments showed that the key to achieve this was that the levels of the lactone **3** in the crude reaction mixture should be less than 2% (HPLC %area/area). This was because the lactone **3** co-crystallizes with the alcohol **2**.

Levels of the starting material **1** were not critical as it was readily removed during the purification steps.

Step 7: Choose the Key Responses to Be Measured (e.g., Yield, Quality, etc.)

The next important step is to establish the response measurements that will signify progress toward the goals or targets identified in step 2.

A key issue is also that the response measurements (e.g. yield, impurity level) are robust and meaningful. As with

Table 4: **Stepwise approach to design selection**

what is your goal?	suggested design	rationale
To check that the extreme combinations of factor levels will work and gain an insight into the repeatability of the system.	Pilot study (sizing or scoping experiments). 1. One experiment carried out at all the high factor settings 2. One experiment carried out at all the low factor settings 3. Two experiments carried out at the centre points.	This small experimental set performed before consuming larger resources on the chosen design. It should provide confidence that the proposed factor ranges and the response measurement method are appropriate. The centre-points will give an insight into the reproducibility of the transformation and the analysis system.
To identify the relative impact each of the factors has on the process. Typical continuous factors would be time, concentration, reagent equivalent, temperature, rate of stirring, rate of addition and pH. Typical discrete factors would be supplier of starting material, solvent A or solvent B.	fractional factorial screening design D-optimal designs	If very little is known about the process apply screening designs to reduce the number of factors to a manageable number in a subsequent optimisation study. If there are any discrete factors in your design, the decision to choose between options must be resolved before a subsequent optimisation study.
To optimise the process (e.g. identify the maximum yield, maximum throughput, or the minimization of an impurity)	Central composite	Apply response surface designs to optimise the key continuous factors against your defined goal(s).
To obtain a robust procedure once a process is identified. To establish allowable variations in the process operating conditions that will not impact on the yield and quality of your product	Fractional Factorial Plackett−Burman	If a more accurate assessment of robustness is required than that provided by the optimisation study it may be necessary to fine-tune (studying narrower factor ranges) in the area of interest. It may also be appropriate to reintroduce the apparently 'trivial' or less significant factors from the screening design. This will check whether the process deviation ranges for these factors are also valid.

all scientific studies, systematic errors (bias in the estimation of the response) and random errors (low precision of measurement or background noise) will present significant issues in the interpretation of the experimental results.

Case Study

Outcome from Step 7. The following responses were chosen to meet the objectives for the case study:

(a) yield as determined by HPLC analysis (% area/area)

(b) quality of the unisolated product **2** as determined by the peak % area/area of impurities as measured by HPLC. Specifically, the levels of residual silyl ether **1** and lactone **3** formation were recorded.

Step 8: Choosing the Most Appropriate Experimental Design

Check that your intended study is "fit for purpose" for the current point in the product's life cycle by using Table 4. If resources are limited, consider using a stepwise approach to applying the experimental design, that is, (a) pilot study, (b) screening design, (c) optimisation design, (d) robustness design. It is often possible to augment the data obtained from the initial screening design with additional data points to produce the optimisation study.

Case Study

Outcome from Step 8. For the case study, the agreed aim was to carry out a four-factor central composite design using 30 data-points. This allows us the opportunity to identify the optimised procedure for this transformation, working within the processing restraints agreed in step 2.

The experimental plan used is outlined in Table 5. As this is a four-factor design, there are 25 experimental data points. The centre point is replicated 5 times to estimate the level of experimental error, making a total of 30 data-points to be collected. In practice only 20 experiments were carried out as some experiments were sampled more than once at different time-points (see step 10 for further discussion of the implications of doing this).

Sufficient resource was available to go straight into the response surface design. It was expected that curvature would be seen in the response function over the region of interest. Therefore, a response surface design would be more appropriate than a screening design.

If resource were tighter, a more cautious approach would be to first carry out a fractional factorial study involving as few as 8 data-points. If the results from this were encouraging, the design could be augmented with an additional 22 data-points to convert it into the optimisation study (Table 5).

Note; it will be seen that although the case study deals with optimisation, additional information about the relative importance of the factors and robustness is also gained. Thus, Table 4 should be seen as a general, rather than an absolute, indicator of the information that can be derived from each stage.

Step 9: Consider the Practical Aspects of Implementing the Experimental Design

The key issues of process control to consider are those you would typically consider for good experimentation

Table 5: **Experimental plan**[a]

		factor settings				responses		
standard order	run order	temp °C	time h	NMP vol	TREAT.HF equiv	alcohol % area	silyl ether % area	lactone % area
1	4	15	22	4	1.17			
2	14	25	22	4	1.17			
3	8	15	28	4	1.17			
4	18	25	28	4	1.17			
5	3	15	22	6	1.17			
6	13	25	22	6	1.17			
7	7	15	28	6	1.17			
8	17	25	28	6	1.17			
9	2	15	22	4	1.5			
10	12	25	22	4	1.5			
11	6	15	28	4	1.5			
12	16	25	28	4	1.5			
13	1	15	22	6	1.5			
14	15	25	22	6	1.5			
15	5	15	28	6	1.5			
16	19	25	28	6	1.5			
17	20	10	25	5	1.33			
18	30	30	25	5	1.33			
19	9	20	19	5	1.33			
20	11	20	31	5	1.33			
21	26	20	25	3	1.33			
22	24	20	25	7	1.33			
23	22	20	25	5	1			
24	28	20	25	5	1.67			
control experiments (centre points)								
25	10	20	25	5	1.33			
26	21	20	25	5	1.33			
27	23	20	25	5	1.33			
28	25	20	25	5	1.33			
29	27	20	25	5	1.33			
30	29	20	25	5	1.33			

[a] Note that the table lists the experiments in standard order (a conventional ordering of the array of low- and high-factor levels versus runs), but they were actually carried out in the run order

whether by the traditional or the experimental design approach.

Case Study

Outcome from Step 9. Prior to this case study, control of two of the factors, temperature and time were identified as being particularly problematic over the factor ranges chosen. The issues were successfully overcome by introducing new technology to obtain better control.

Temperature. Commonly used small-scale laboratory equipment such as electrical heaters, oil-baths and cooling baths did not give the precision or the range of temperatures that was required. In the initial scoping studies, the reaction was carried out at "ambient" temperature but with the variations in the room temperature from the day and night, process control was poor. By employing a commercially available Peltier block,[6] we had convenient and accurate temperature control. Several reactions at the same temperature could be carried out in the block at the same time. The temperature range of 0 to 40 °C, although limited for routine chemistries, was ideal for the desilylation reaction under study.

Time. One of the factors in the study was the time to achieve reaction conversion. It was possible manually to monitor the reaction progressing during the day by HPLC, but not during the night, as there was the safety consideration of handling hydrogen fluoride solutions outside normal working hours.

To overcome this, the reaction was monitored automatically by HPLC, using a standard Gilson 231XL[6] auto-sampler that was capable of analytical sample preparation and injection. The reactions specified by the experimental design model were prepared by hand. Aliquots of reactions were placed in the Peltier block sample tray of the autosampler, the position and software control determining the time-point at which the reaction was automatically monitored. Therefore, using this semi-automated approach, the required datapoints were obtained.

Step 10: Implement the Design

Carry out the reactions with care, according to well-defined worksheets of the design plan. Avoid transcription errors whilst entering the associated responses.

Case Study

Outcome from Step 10. The reactions were carried out using the following procedure:

The silyl ether **1** (2 g) was dissolved in *N*-methylpyrrolidone (NMP, range 4−7 volumes [8−14 mL]) and then treated with triethylamine trihydrogen fluoride (TREAT.HF, 1−1.67 equiv [0.6−1.0 g]). The mixture was sonicated for 1 min to give a homogeneous solution. An aliquot (0.5 mL) was placed in a 2 mL vial in the Peltier autosampler rack (temperature range 10−30 °C) and sampled at the appropriate time-points.

Safety Note. Triethylamine trihydrogen fluoride is assumed to be of the same order of toxicity as a solution of hydrogen fluoride, that is, very toxic due to its extremely corrosive and irritant nature. Hydrogen fluoride was detected (13 ppm at 3 cm above the liquid surface) in the confined headspace of the reagent bottle. Therefore, the same precautions were taken as would have been for dealing with hydrogen fluoride (either in solution or gaseous). Unlike hydrogen fluoride, however, triethylamine trihydrogen fluoride does not etch glass. For further details, see the Materials Safety Data Sheet as supplied by FAR Research.[7]

Step 11: Consider the Statistical Implications of the Practical Implementation

Examine the experimental procedure for practices that may introduce error and bias into the study. An important requirement of experimental design is that, where possible, the reactions should be carried out in a random order. If there are any systematic errors (such a gradually degrading HPLC column), then this will be revealed during the analysis. If the experiments are not carried out in random order, it may be possible to confuse factor effects with systematic error. Inspection of the replicated control experiments will also reveal random error or noise due to unassignable causes, as these cannot be as a result of changes in the factors.

(6) The Gilson 231XL and the Peltier block are commercially available from Anachem Ltd, Charles Street, Luton, Beds, LU2 OEB UK.

(7) Far Research Inc., 307 Amherst Road, PO Box 2278, Morganton, NC 28680, U.S.A., Telephone: 704 438 0101.

Case Study

Outcome from Step 11. In this case study, there were two aspects of note − randomisation and repeated measures. The experiments were not carried out in a completely random order for two reasons. The first reason is that the temperature of the Peltier reaction block is uniform throughout. This meant that any reactions carried out in parallel had to be at the same temperature. For example, all the reactions at 15 °C were carried out on 1 day, and all the reactions at 25 °C were carried out on the next day. To carry out the reactions in random order using the available equipment would have meant that the reactions would have had to be carried out serially and hence the study would take longer to complete. The second reason was that where two design points had the same settings of temperature, solvent volume and equivalents of reagent, but differed in time, the same reaction was used and the response measured at the two time points. Clearly, this approach does not lend itself to randomisation, as the shortest time point always has to be measured first! In addition, it underestimates the experimental error in preparing the reaction over the whole study. Repeated sampling may also introduce a systematic error between the samplings − this would be the case if the mixture was not heterogeneous and sampling successively depleted the concentration of the supernatant. As a general point, sampling of a heterogeneous mixture, whether by an automated or a manual approach, should always be carried out with caution and awareness of the implications of the validity of the data.

Step 12: Analyse the Results

First, input the responses into the design table. Fit a full model by multiple linear regression and examine model terms for significance. Then modify the model, if necessary, to improve prediction by pooling nonsignificant model terms into error. ANOVA tables and *p*-values are useful for determining which terms can be removed from the model. The final model should maintain a hierarchical structure. Check the final model by examining studentised residuals and diagnostic plots. Finally generate contour and 3D plots to determine the region where predicted optimum process outcome occurs. Use the predicted model and confidence intervals to narrow the settings of the factors.

Case Study

Outcome from Step 12. The results are summarised in Table 6. The analysis is performed in Design-Expert 5 (DX-5).

Figure 6 shows the correlation between the experimental and the predicted data-points for the alcohol response. The data are arranged in standard order, not run order. The last six entries are the controls, which indicate the level of background noise is relatively small, compared to the variation in experimental data-points.

Similarly, Figures 7 and 8 show the same information for the silyl ether and the lactone responses. All three models show that the predicted responses generally lie within 0.5% or less of the experimental data. See the appendix 2 for more details of the statistical analysis.

Table 6: **Experimental results[a].**

		factor settings				responses		
standard order	run order	temp °C	time h	NMP vol	TREAT.HF equiv	alcohol % area	silyl ether % area	lactone % area
1	4	15	22	4	1.17	82.93	15.53	0.59
2	14	25	22	4	1.17	94.04	1.61	3
3	8	15	28	4	1.17	88.07	10.19	0.97
4	18	25	28	4	1.17	93.97	0.55	4.12
5	3	15	22	6	1.17	77.21	21.43	0.34
6	13	25	22	6	1.17	92.99	3.94	1.86
7	7	15	28	6	1.17	83.6	15.07	0.55
8	17	25	28	6	1.17	94.38	1.76	2.61
9	2	15	22	4	1.5	88.68	9.25	0.81
10	12	25	22	4	1.5	94.3	0.67	3.69
11	6	15	28	4	1.5	93	4.6	1.41
12	16	25	28	4	1.5	93.42	0	5.06
13	1	15	22	6	1.5	84.86	13.5	0.53
14	15	25	22	6	1.5	94.26	1.85	2.39
15	5	15	28	6	1.5	88.71	9.63	0.76
16	19	25	28	6	1.5	94.66	0.64	3.33
17	20	10	25	5	1.33	75.82	22.85	0.21
18	30	30	25	5	1.33	93.25	0	5.29
19	9	20	19	5	1.33	89.78	8.02	1.05
20	11	20	31	5	1.33	94.61	1.82	2.41
21	26	20	25	3	1.33	94.13	1.71	2.95
22	24	20	25	7	1.33	89.94	8.03	0.97
23	22	20	25	5	1	88.21	9.76	1.13
24	28	20	25	5	1.67	93.11	4.26	1.64
control experiments (centre points)								
25	10	20	25	5	1.33	93.32	3.83	1.77
26	21	20	25	5	1.33	92.32	5.13	1.54
27	23	20	25	5	1.33	93.68	3.58	1.72
28	25	20	25	5	1.33	93.27	4.19	1.62
29	27	20	25	5	1.33	92.87	4.42	1.65
30	29	20	25	5	1.33	92.96	4.45	1.65

[a] Note that the table lists the experiments in standard order (a conventional ordering of the array of low- and high-factor levels versus runs), but they were actually carried out in the run order.

Step 13: Interpret and Interrogate the Model

Examine the perturbation graph to see the effect of changing one factor while holding the rest constant. This plot can be useful to decide which axes to use on a contour or 3D plot. Pick the factors that have the most complex behaviour (most curved or steepest change rate) and use them as axes on the other plots. This will put the simplest (least interesting) dimensions off the graph. The contour plots show the interaction effects and the degree of robustness (the flatter the surface, the less susceptible the response is to changes in process conditions).

Case Study

Outcome from Step 13. Figure 9 shows that of the four factors, temperature has the most significant effect on all three responses (i.e. has the steepest change rate).

Although the model is in four dimensions, it is only possible to display two factors at once using the software. Figure 10 shows the contour maps for each of the responses for the two most interesting factors; time and temperature. The other two factor levels are held constant. The contours represent the level of the responses. Using these three plots in turn, it is possible to see that the conditions required to maximize the alcohol, to minimize the silyl ether and to minimize the lactone are in three different locations − i.e.,

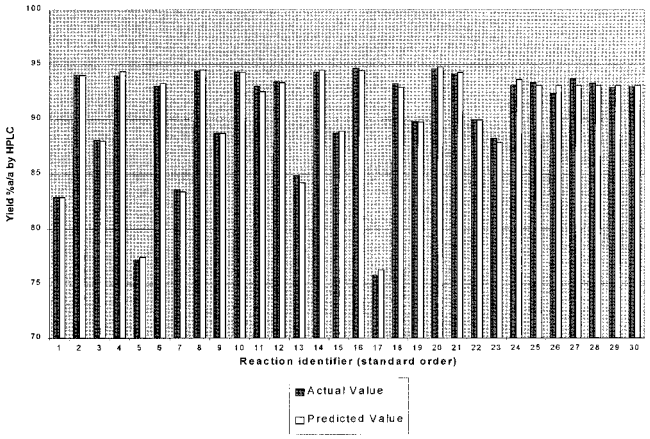


Figure 6. **Comparison of actual versus predicted yield (HPLC area/area %) for the alcohol (2).**

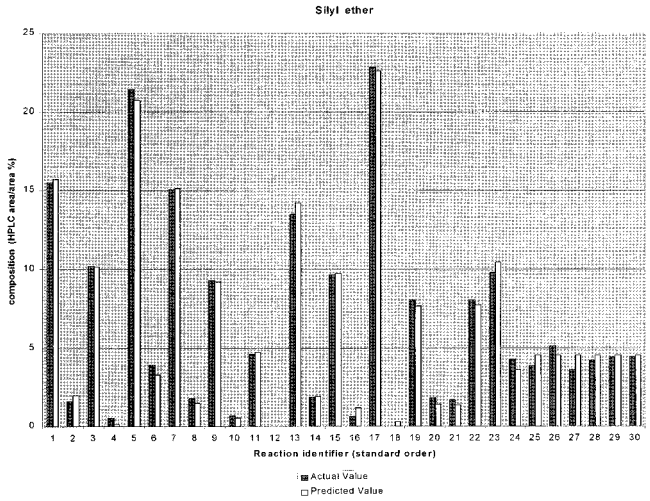


Figure 7. **Comparison of actual versus predicted composition (HPLC area/area %) for the silyl ether (1).**

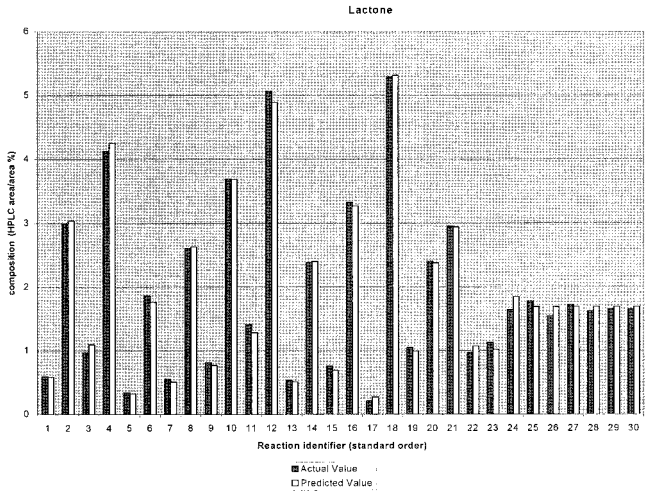


Figure 8. **Comparison of actual versus predicted composition (HPLC area/area %) for the lactone (3).**

have different preferred operating regions. Figure 11 shows how this conflict is resolved. If the requirements are superimposed on each other (i.e. maximize the yield, keep the lactone level <2%, no restriction on the silyl ether), then the permissible working area is reduced to the nonshaded portion (in Figure 11).

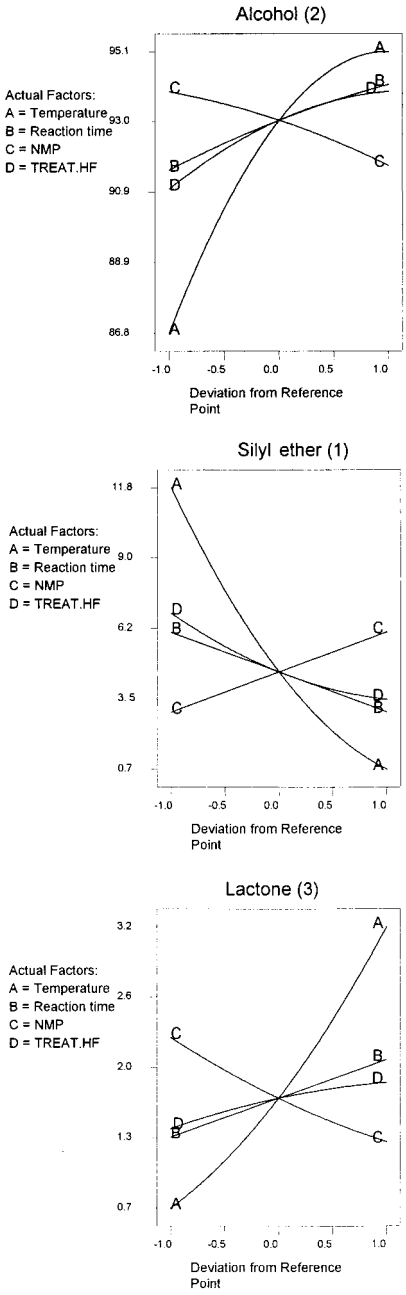


Figure 9. **Perturbation plots.**

Step 14: Verify the Mathematical Model

Now use the software package to interrogate the response models on all the factors studied. The goals can be refined to achieve alternative specified goals, such as increasing the throughput or modifying the acceptable quality specification. Carry out experiments using the predicted settings to confirm the validity of the model.

Case Study

Outcome from Step 14. The reaction modelling was highly predictive. The mathematical models were interrogated to examine the effect on yield and quality under a variety of conditions or constraints. In total, six different scenarios were evaluated (see Table 7, column 1). In each case, the model suggested the factor settings required to obtain such an outcome (column 2). The first three rows

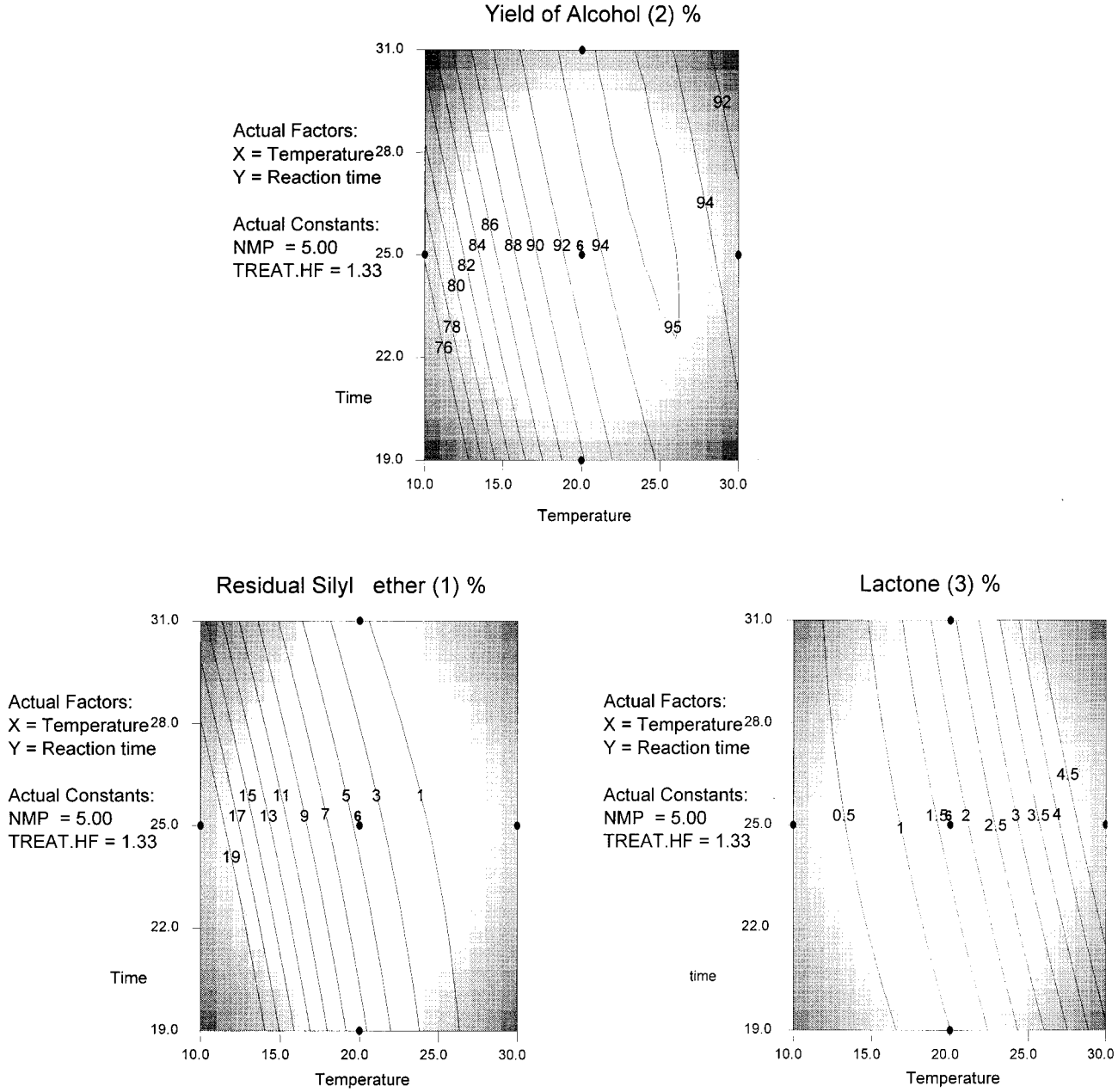


Figure 10. **Contour maps.**

show the conditions required to maximize quality within different quality constraints. The last three rows hold the quality constant and looks at altering the conditions to maximize the yield, whilst constraining factors with economic and environmental implications. When these six different processes were performed experimentally, the actual yield and predicted yields showed a high degree of correlation (on average accurate within 0.2% area/area by HPLC) (columns 3−6).

The analysis can also tell us the standard error and confidence intervals associated with each of the predictions. This helps us quantify the variability in the predicted response (see Appendix 2, Table 15 for further details).

Step 15: Model the Process Deviation Ranges

After identifying the manufacturing conditions, use the model to set process deviation ranges (PDRs). The process should be capable of operating within these process limits and still be able to deliver acceptable yield and quality of product.

Case Study

Outcome from Step 15. This is illustrated in Table 8. If the worst case scenarios are acceptable, then the PDRs can be extended to give more flexibility in manufacture. If the worst case scenarios are not acceptable, then the PDRs may be narrowed. Alternatively, the operating conditions could be moved to a more robust area, which may involve a loss of yield. In Table 8, the worst case scenario with the proposed PDR ranges is not acceptable. As a result of such modelling, the thermal control of the manufacturing plant was improved to bring the temperature control within ±1 °C. The cost of implementing this improved engineering control (ca. £10K) could be justified in light of these results. The model allows informed decisions to be made.

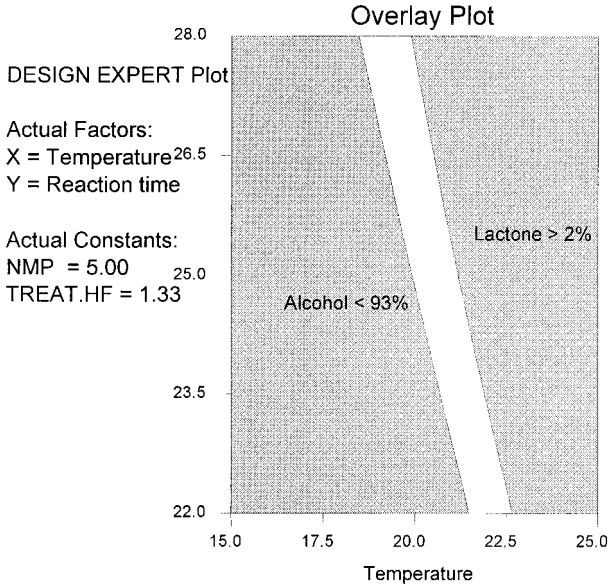


Figure 11. **Permissible working area.**

Step 16: Scale Up the Procedure

Where possible, carry out an experimental design study on an appropriate scale, using appropriate technology. In reality at the preliminary stages of exploration, amounts of starting material are very limited; therefore, carry out the initial designs easily and quickly on a small scale in the laboratory. Any options identified from the models should be verified as early as possible on a larger scale using jacketed vessels, reaction calorimeters, or small plant reactors. Ultimately, this will then provide a basis for industrial-scale production levels. Anticipate processes that may be scale-dependent due to bulk-transfer and heat-transfer effects. These include crystallisation, exothermic, and phase-transfer reactions. In these cases, it may be more appropriate to use small-scale experimentation as a guide to identifying the important factors and work on a larger scale to obtain the more predictive model.

Case Study

Outcome from Step 16. Fortuitously, the features of the desilylation reaction were compatible with the limited technical capabilities of the Gilson 231XL autosampler and Peltier rack. The reaction was homogeneous and hence need not be stirred. The reaction was carried out at near ambient temperatures and was not exothermic so it did not require condensing. It was not air- or moisture-sensitive, so an inert atmosphere was not critical. A set of the preferred conditions was first validated in traditional glassware in the laboratory and then subsequently on pilot plant. During the initial pilot-plant run, an aliquot was removed from the reaction vessel and placed in the Peltier block. The block containing this unstirred aliquot was set at the same temperature as that of the stirred, nitrogen-blanketed reaction in the 400-L vessel. After 26 h a second aliquot was removed from the reaction vessel, and both aliquots indicated the reaction had proceeded to the same degree of completion. Hence, for this reaction, the initial assumptions that stirring and nitrogen blanketing was not required for the reactions that were monitored automatically (10 mL), were verified both on a larger scale (500 mL) in the laboratory and in pilot plant (400 L).

After completing the manufacturing campaign, additional process investigation was carried out in the laboratory to further minimize waste in the isolation procedure. Perversely, with the new procedure it was now the silyl ether impurity that was difficult to remove, whereas the lactone impurity level was no longer an issue. Hence, it was now more important to drive the reaction on further before work-up.

It was possible to use the information derived from the original model of the reaction with the new work-up conditions, rather than having to start again from square one. This was possible, since the experimental design case study covered only the reaction conditions, rather than looking at the combined reaction condition/work-up.

Conclusions

The benefits of applying experimental design in process research are as follows:

(a) It produces powerful mathematical models of the chemical process or procedure to allow opportunities or constraints to be fully considered.

(b) Quality data allows better strategic decision making and faster scale-up of optimised processes into plant.

(c) The models are obtained for a quantifiable amount of resource.

(d) It is an efficient and effective method of choosing which experiments to perform.

(e) The strategy is compatible with running automated reactions in parallel.

(f) If circumstances change (e.g. the price of reagent increases dramatically or work-up alters), the model can be interrogated in different ways to take account of the new criteria.

(g) It can be used as a framework to capture and share information between project teams.

Outcome from Steps 1−16. The case study shows that reactions are often more complex than intuition would suggest. Woolly statements can be quantified, and the effect of several (often-conflicting) responses can be visualised.

This highly predictive model of reaction was generated in 6 days using readily accessible technology to monitor the reactions automatically. The reaction, which was originally carried out at “ambient” temperature, is actually very sensitive to variation in temperature, and it is the control of this factor that is most critical.

Implication for Automated Approach to Process Investigation

Noise due to unassignable causes can be minimized by keeping all other factors outside the design as constant as possible and changes in the environment should be minimized where possible. The use of automated and semi-automated processing and measurement systems provides extra confidence that this is the case, and is therefore encouraged.

Table 7: **Comparison of predicted and actual yields for six different conditions**

target/constraints	conditions suggested by the model	product yield[a]		impurity yield[a]	
		predicted	actual	predicted	actual
maximize product yield[a] no limit on lactone.	temperature 19 °C time 31 h solvent 3.6 vol $Et_3N{\cdot}3HF$ 1.42 equiv	95.3	95.8	3.3	3.3
maximize product yield limit on lactone < 2%	temperature 17 °C time 31 h solvent 4.8 vol $Et_3N{\cdot}3HF$ 1.50 equiv	94.2	94.0	1.9	1.7
maximize product yield limit on lactone < 1.1%	temperature 16 °C time 29 h solvent 5.3 vol $Et_3N{\cdot}3HF$ 1.68 equiv	92.4	93.1	1.1	1.1
maximize product yield limit on lactone < 2% solvent <3.5 vol	temperature 14 °C time 31 h solvent 3.45 vols. $Et_3N{\cdot}3HF$ 1.58 equiv	93.9	94.2	1.8	2.0
maximize product yield limit on lactone < 2% $Et_3N.3HF$ < 1.18 equiv	temperature 28 °C time 19.5 h solvent 7 vol $Et_3N{\cdot}3HF$ 1.17 equiv	93.7	93.4	1.9	2.0
maximize product yield limit on lactone < 2% time <23 h	temperature 24 °C time 23 h solvent 6.3 vol $Et_3N{\cdot}3HF$ 1.41 equiv	94.2	94.2	2.0	1.9

[a] Yield in this table refers to unisolated yield % area/area by HPLC.

Table 8: **Predicted deviation ranges**

					actual	
	temp °C	time h	NMP vol	TREAT.HF equiv	alcohol % area	lactone % area
factory setting	20	26	4.7	1.34	93.75	1.9
process deviation range	± 2	± 1	± 0.1	± 0.08		
worst case scenario derived from the model						
					predicted	
lowest yield outcome	18	25	4.8	1.26	90.4	1.2
highest impurity outcome	22	27	4.6	1.42	95.1	2.9

Automation is not critical to experimental design—it just makes it easier to obtain better control and more efficient to implement.[8] This proof of concept study showed that it was possible to carry out reactions in a "reaction station" located within the working envelope of an autosampler. Although adequate for this study, the Gilson 231XL and the Peltier rack had severe limitations in the range of other chemistry it was possible to perform. In addition, the ability to look at only one temperature per run became the rate-limiting step.

There already existed precedent for multizoned[9] or individual[10] thermal control using either resistive heating or Peltier control. The study also showed that having relieved the bottleneck of analysing the samples, the labour-intensive task became the reaction preparation and data-manipulation.

For all of these reasons, this case study led us to develop the development automated reaction toolkit (DART)[11] and process research optimisation screening parallel experimentation robot (PROSPER)[12] systems at GlaxoWellcome, which are specifically designed to accelerate process optimisation using automation and experimental design.

The DART system is based on the Gilson 233XL autosampler, (which has a larger workspace envelope than the Gilson 231XL) and custom-built reaction stations.

The DART allows automated reaction preparation, has better process control (range of temperature, concurrent reactions in different temperature zones, condensing, stirring, and nitrogen blanketing) and on-line HPLC monitoring. A refinement of this system is available commercially as the Anachem SK233 Workstation.[13]

PROSPER offers even greater control and flexibility, higher throughput and sophisticated software features to improve the quality of the acquired data.

Acknowledgment

We thank Sharon Filbey and Dennis Lendrem for discussions of the statistical analysis and Steve Skittrall, Geoff Smith, Simon Munt, and others for the scale-up of this reaction into plant.

(8) Owen, M. R. Laboratory Automation in Chemical Development. In *Process Chemistry in the Pharmaceutical Industry*; Gadamasetti, K., Ed.; Marcel Dekker Inc.: New York, 1998; pp 429−455.

(9) For example: Josses, P. *Adv. Lab. Autom. Rob.* **1989**, *6*, 463−475.

(10) Fujita, T.; Umemura, S.; Uchida, N. Jpn. Kokai Tokkyo. Patent Application JP 92-128553 920521.

(11) Emiabata-Smith, D. F.; Crookes, D.; Owen, M. R. *Org. Process Res. Dev.* **1999**, *3*, 281−288.

(12) Owen, M. R.; Smith, L. C. Presented at the *3rd Laboratoratory Automation in Process R&D Symposium*, Boston, 2000.

(13) The Anachem SK233 Workstation (formally referred to as the DART system, during its development at GlaxoWellcome) is commercially available from Anachem Ltd, Charles Street, Luton, Beds, LU2 OEB UK.

Appendix 1

Terminology.

ANOVA: analysis of variance is a statistical technique which subdivides the total variation of a set of data into component parts associated with the specified sources of variation.

Blocking: a technique used to reduce the noise and improve sensitivity to effects. It is used when there is a known factor that may influence the experimental result, but the effect itself is of no interest. A typical example would be if there is not enough material from one batch to complete the experimental design and therefore two batches of raw material have to be used.

Central composite design: type of experimental design used for optimisation, which allows a mathematical model of the process to be defined.

Centre points or controls: replicated centre points or controls are run at the points in the centre of each factor. The replicated centre points provide an estimate of background variability.

Continuous and discrete factors: factors can be quantitative or qualitative. Quantitative factors are those which can take numerical values on a continuous scale (temperature, time). Qualitative factors are usually assigned names and are discrete in nature (e.g., solvent, chemist, and site).

Duplication: occurs when several measurements are made on the same experimental run. Duplication is not the same as replication. Duplication gives information only on measurement error or product uniformity. Duplication almost always underestimates experimental error. Repeated sampling from the same reaction while considering time as a factor will also underestimate the experimental error.

Experimental design: a set of systematically designed experimental runs.

Experimental error: there are three main categories of experimental error:

(a) background variability or noise due to unassignable causes, quantify through replication.

(b) bias or systematic error due to assignable causes, minimize through blocking and randomisation.

(c) blunders due to mistakes in experimental practice, avoid through careful experiment practice and well-defined worksheets for experimental design.

If only one measurement for each experimental run is used, there is no direct estimate of the background variability. One approach to the analysis of such a design is to assume that certain high-order interactions are negligible and combine them to estimate the background variability.

Experimental run: an experiment with a specified combination of levels for each factor.

Factors: these are the controllable parameters used as inputs to the products and processes under evaluation. As they are varied, they may be expected to change the output of the response variable. Also referred to as: treatments, independent variables, predictors, x variables, and input variables. Examples of factors are: temperature, time, rate of addition, amount of reagent or solvent in a chemical reaction.

Factor effect: the change in response caused by varying the level of the factor.

Factor range: the difference between the highest and lowest levels for a given factor.

Fractional factorial: type of design used for screening a large number of factors.

Interaction: the measured change in response as a result of the combined effect of two or more factors. A two-factor interaction indicates how the effect of one factor changes as the level of the second factor is varied. This is typical of the relationship between temperature and time, for example if the reaction is run at a high temperature, then a shorter time is generally required and vice versa. Higher-order interactions, involving three or more factors, are also possible in which the effect of one factor depends on the levels of two or more factors. Fortunately these are less common in real life.

Lack of fit: measures the ability of the model to adequately predict the response within the design space. Note that the overall fit of the model may be good, but the lack of fit will detect specific regions of the design space where the model begins to break down. That is, the model may fit the design points, but will not be a very good predictor at other points.

Levels: the particular values or settings of a factor. For example, a temperature may be set at either 10 or 30 °C and the solvent may be acetone or methanol.

Method development: For a process chemist this refers to the investigation of the conditions required to bring about a particular transformation or isolation. Usually the study will be confined to one particular substrate. For the medicinal chemist, robust method development refers to the general applicability of reaction conditions for a wide range of substrates.

Model: All factors are varied simultaneously over a set of planned experiments and the results are then connected by means of a mathematical model (usually a polynomial regression model). This model is then used for interpretation, predictions and optimisation.

Optimisation of a process is more than simply obtaining the "best yield". There are many tradeoffs between other responses (such as quality) that need to be considered, and it is important to define and reevaluate the criteria for each study. For example, it is critical to control impurities to required specification levels. Economic and environmental benefits are also important. The throughput should be maximized by reducing reaction time or increasing concentration. Minimizing the input of starting materials, reagents, or solvents can reduce the material costs and waste disposal. Working at temperatures closer to ambient can reduce energy costs.

Parameter: See factor.

Pilot study: A small experimental set performed before embarking on the chosen design to check that the extreme combinations of factor levels will work or to determine the repeatability of the system. Also referred to as sizing or scoping experiments.

Process deviation ranges (PDRs): It is usual to define factor settings, with process deviation ranges (PDRs), within which the plant operator should work to ensure that the process and product are robust to minor variation of the manufacturing parameters.

Randomisation is an experimental technique that can be used to remove the effect of potential bias errors. There is always a risk that the experimental result may be influenced by nonrandom, often time-dependent errors. Such risks may be counteracted by randomisation. This means that in any situation where the experimenter has a choice as to the order he/she should do things, then a random choice should be made. For example, the order of executing the experimental runs should be randomised; the order of analysing samples drawn from the reaction should be randomised if several samples are analysed on the same occasion. Running the experiment in random order makes the interpretation of factor effects more straightforward.

Reaction space: an imaginary area bounded by the extremes of the tested factors.

Replicate is the complete repetition of an experimental run. Any differences between replicates are not a result of changes in the factors but are simply a reflection of background variability. Changes in the responses, as the factor levels are changed, can then be compared to this background variability.

Response surface designs: designs such as central composite, which allow models to be fitted which capture curvilinear relationships between factors and responses.

Responses: These are the outcomes of interest. Also referred to as output variables, dependent variables, *y* variables. Examples of responses are: the yield and level of impurities from a chemical reaction.

Robustness: The target for the Development chemist is to obtain a controlled process and consistent product quality.

Robustness designs: typically fractional factorial designs used to determine process deviation ranges.

Run order: the actual order in which the reactions were performed. This should preferably be in a random sequence.

Screening: the process of identifying the selection of "ingredients" (such as reagent and solvent) and the factors which have most impact on the process (i.e., to select the important few from the trivial many)

Screening designs: a group of experiments which identifies the relative impact each of the factors has on the process.

Standard order: a conventional ordering of the array of low- and high-factor levels versus runs.

Statistical significance (*p*-value): the probability that an effect at least as large as that observed would be obtained purely by chance if there were really no difference.

Appendix 2

Comments on the Statistical Analysis. Design-Expert 5(DX-5) provides a detailed glossary of the terms used in this section.

Fit a model with all linear, interaction, and quadratic terms by multiple linear regression. For four factors, this model will contain 15 parameters: 1 for the overall mean (intercept), 4 for the linear terms (A, B, C, D), 6 for the interaction terms (AB, AC, AD, BC, BD, CD), and 4 for the quadratic terms (A^2, B^2, C^2, D^2). This leaves 15 degrees of freedom for error.

Table 9: **Sequential model sum of squares**

source	sum of squares	degrees of freedom	mean square	*F* value	prob > *F*
mean	245952	1	245952		
linear	531.191	4	132.798	15.9808	< 0.0001
quadratic	204.782	10	20.4782	103.654	< 0.0001
cubic	1.68635	8	0.210794	1.1554	0.4309
residual	1.2771	7	0.182443		
total	246691	30	8223.03		

Table 10: **Lack of fit**

source	sum of squares	degrees of freedom	mean square	*F* value	prob > *F*
linear	206.657	20	10.3328	47.4331	0.0002
quadratic	1.87425	10	0.187425	0.860379	0.6089
cubic	0.1879	2	0.09395	0.43128	0.6717
pure error	1.0892	5	0.21784		

Table 11: **Model summary statistics**

source	root MSE	R^2	adjusted R^2	predicted R^2	PRESS
linear	2.88268	0.718858	0.673876	0.58693	305.233
quadratic	0.444481	0.99599	0.992247	0.983268	12.3641
cubic	0.427133	0.998272	0.99284	0.96126	28.626

The "Sequential model sum of squares" summary (Table 9) shows how terms of increasing complexity contribute to the model's predictive power. Using the column headed prob. > *F* select the highest-order polynomial where the additional terms are significant, in this case the quadratic model.

The target is to obtain a model with "insignificant lack of fit". The "Lack of fit" data (Table 10) compares the model residual error to the pure error from replicated design points. In this example, the replicated design points are the six centre points. If higher order model terms are needed to adequately explain the response surface, then the lack of fit p-value will be less than 0.05. In this example, the linear model is poor because it shows significant lack-of-fit (prob. > *F* is 0.0002), whereas the quadratic model is good because it shows insignificant lack of fit (prob. > *F* is 0.6089).

The "Model summary data" (Table 11) is useful in comparing different models. Focus on the model minimizing the "PRESS", or equivalently maximizing the "PRED R-SQR".

In this case, there is little discrepancy between the adjusted R^2 and the predicted R^2, which indicates the model does not need to be refined.

The value, R^2 in the ANOVA (Table 12) shows that the full quadratic model explains 99.5% of the variation in the data. This value represents the percent of variation in the data that can be *explained* by the fitted model. As an estimator, it usually overestimates how well the model fits the data because there is no penalty for adding additional terms to the model that make it more complex. The overall model fit is significant as evidenced by a *p*-value of <0.0001.

Table 12: **ANOVA for response surface quadratic model**

source	sum of squares	degrees of freedom	mean square	*F* Value	prob > *F*
model	735.973	14	52.5695	266.089	< **0.0001**
residual	2.96345	15	0.197563		
lack of fit	1.87425	10	0.187425	0.860379	**0.6089**
pure error	1.0892	5	0.21784		
cor total	738.937	29			

root MSE	dep mean	C.V.	PRESS	R^2	*a*djusted R^2	pred R^2	adeq precision
0.444481	90.545	0.490895	12.3641	**0.99599**	0.992247	0.983268	58.8459 desire > 4

Table 13: **Table of coefficients**

factor	coefficient estimate	DF	standard error	*t* for H_0 coeff = 0	prob > |*t*|	VIF
intercept	93.07	1	0.181459			
A - temperature	4.15917	1	0.090729	45.8415	<0.0001	1
B - reaction time	1.25833	1	0.090729	13.8691	<0.0001	1
C - NMP	−1.08833	1	0.090729	−11.9954	<0.0001	1
D - TREAT.HF	1.4375	1	0.090729	15.8438	<0.0001	1
A2	−2.11938	1	0.08487	−24.9722	<0.0001	1.05
B2	−0.20438	1	0.08487	−2.40811	0.0294	1.05
C2	−0.24438	1	0.08487	−2.87942	0.0115	1.05
D2	−0.58813	1	0.08487	−6.92976	<0.0001	1.05
AB	−1.17875	1	0.11112	−10.6079	<0.0001	1
AC	1.17875	1	0.11112	10.6079	<0.0001	1
AD	−1.38625	1	0.11112	−12.4752	<0.0001	1
BC	0.22	1	0.11112	1.97984	0.0664	1
BD	−0.3225	1	0.11112	−2.90226	0.0109	1
CD	0.245	1	0.11112	2.20482	0.0435	1

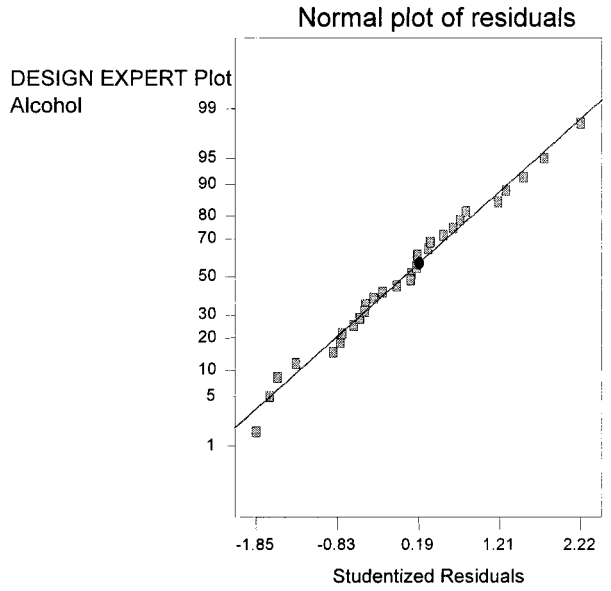


Figure 12. **Normal probability plot.**

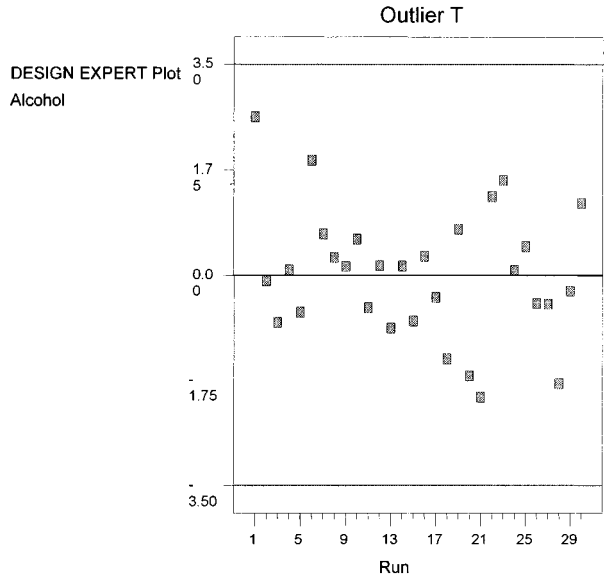


Figure 13. **Outlier *T* plotted against run number.**

Therefore, the response surface can be adequately modelled by the quadratic function selected.

Use the "Table of coefficients" for the full model (Table 13) to see whether the coefficient is that size by chance alone. When a coefficient has a *p*-value larger than 0.10, it is not significantly contributing to the model's predictive power. Some practitioners prefer to modify the model, if necessary, to improve prediction by pooling nonsignificant model terms into error. In this case study no terms have been removed in any of the three response models, because the six centre points contribute sufficient degrees of freedom.

Finally, use the "Diagnostics case statistics" (Table 14) based on the full quadratic model to assess the final model. Normal probability plots of the residuals are useful in examining the model and its assumptions. A studentised residual is the sample residual divided by the square root of its estimated variance. If the studentised residuals were the result of random noise (roughly normal), then they should

Table 14: **Diagnostics case statistics**

standard order	actual value	predicted value	residual	leverage	student residual	Cook's distance	outlier *t*	run order
1	82.93	82.90	0.026667	0.583	0.092944	0.000806	0.089818	4
2	94.04	93.99	0.045833	0.583	0.159747	0.002382	0.154462	14
3	88.07	87.98	0.0875	0.583	0.304972	0.008681	0.295549	8
4	93.97	94.36	−0.38833	0.583	−1.3535	0.170982	−1.3956	18
5	77.21	77.44	−0.22917	0.583	−0.79874	0.059545	−0.78861	3
6	92.99	93.24	−0.255	0.583	−0.88878	0.073726	−0.88218	13
7	83.6	83.40	0.201667	0.583	0.702889	0.046112	0.690522	7
8	94.38	94.49	−0.10917	0.583	−0.38049	0.013512	−0.36937	17
9	88.68	88.71	−0.02583	0.583	−0.09004	0.000757	−0.08701	2
10	94.3	94.25	0.048333	0.583	0.168461	0.002649	0.162903	12
11	93	92.49	0.505	0.583	1.76013	0.289151	1.90897	6
12	93.42	93.33	0.094167	0.583	0.328208	0.010054	0.318224	16
13	84.86	84.22	0.638333	0.583	2.22485	0.461994	2.62591	1
14	94.26	94.48	−0.2225	0.583	−0.7755	0.056131	−0.76469	15
15	88.71	88.89	−0.18083	0.583	−0.63028	0.037077	−0.61713	5
16	94.66	94.44	0.223333	0.583	0.778406	0.056552	0.767676	19
17	75.82	76.27	−0.45417	0.583	−1.58295	0.233869	−1.67562	20
18	93.25	92.91	0.339167	0.583	1.18213	0.130427	1.19928	30
19	89.78	89.73	0.044167	0.583	0.153938	0.002212	0.148836	9
20	94.61	94.77	−0.15917	0.583	−0.55476	0.028724	−0.54153	11
21	94.13	94.27	−0.13917	0.583	−0.48505	0.021959	−0.47232	26
22	89.94	89.92	0.024167	0.583	0.084231	0.000662	0.081394	24
23	88.21	87.84	0.3675	0.583	1.28088	0.153129	1.31124	22
24	93.11	93.59	−0.4825	0.583	−1.6817	0.263959	−1.80358	28
25	93.32	93.07	0.25	0.167	0.616137	0.005062	0.602924	10
26	92.32	93.07	−0.75	0.167	−1.84841	0.045555	−2.0321	21
27	93.68	93.07	0.61	0.167	1.50337	0.030135	1.57597	23
28	93.27	93.07	0.2	0.167	0.49291	0.003239	0.4801	25
29	92.87	93.07	−0.2	0.167	−0.49291	0.003239	−0.4801	27
30	92.96	93.07	−0.11	0.167	−0.2711	0.00098	−0.26255	29

Table 15: **Confidence intervals**

conditions suggested by the model	response	prediction	SE mean	SE Pred	95% PI low	95% PI high
temperature 19 °C time 31 h solvent 3.6 vol Et_3N·3HF 1.42 equiv	alcohol silyl ether lactone	95.30 0.17 3.27	0.51 0.59 0.11	0.68 0.78 0.15	93.85 −1.50 2.96	96.75 1.84 3.59
temperature 17 °C time 31 h solvent 4.8 vol Et_3N.3HF 1.50 equiv	alcohol silyl ether lactone	94.19 2.90 1.87	0.45 0.52 0.10	0.63 0.73 0.14	92.84 1.34 1.57	95.54 4.46 2.16
temperature 16 °C time 29 h solvent 5.3 vol Et_3N·3HF 1.68 equiv	alcohol silyl ether lactone	92.43 5.47 1.12	0.55 0.64 0.12	0.71 0.82 0.15	90.92 3.73 0.79	93.94 7.21 1.45
temperature 14 °C time 31 h solvent 3.45 vol Et_3N·3HF 1.58 equiv	alcohol silyl ether lactone	93.94 3.07 1.83	0.82 0.94 0.18	0.93 1.07 0.20	91.96 0.78 1.40	95.93 5.36 2.26
temperature 28 °C time 19.5 h solvent 7 vol Et_3N·3HF 1.17 equiv	alcohol silyl ether lactone	93.71 2.78 1.94	0.89 1.03 0.19	1.00 1.15 0.22	91.58 0.33 1.47	95.83 5.24 2.40
temperature 24 °C time 23 h solvent 6.3 vol Et_3N·3HF 1.41 equiv	alcohol silyl ether lactone	94.22 2.58 1.96	0.27 0.31 0.06	0.52 0.60 0.11	93.11 1.30 1.72	95.32 3.86 2.20

plot along a straight line. When data falls far off this line, the model should be examined. In Figure 12 the points on this plot lie fairly close to the straight line so the model seems appropriate.

Another diagnostic tool is the outlier *t* values plotted against the run number of the experiment. The outlier *t* value is the sample residual divided by the square root of its estimated variance; in this case, the estimated variance for the current residual is calculated based on all the other residuals, excluding the current one. A horizontal band of points symmetric about zero suggests the residuals are random with constant variance and a mean of zero. This is an underlying assumption of linear regression. Values larger in absolute value than 3.50 suggest that the model is not predicting the data well.

Figure 13 shows that these points are symmetric about zero within a band of ± 2.6. There is no pattern here that suggests the model needs to be changed.

For this case study, the terms in the final model are important in explaining the data. No higher order terms are needed in the model. The curvature in the response surface can be adequately modelled by the quadratic terms. Cubic terms are not necessary. These steps verified that the final model has met the critical assumptions.

Table 15 shows the standard error and confidence intervals associated with each of the predictions. A confidence interval helps quantify the variability of the predicted response for a particular combination of factor settings. The variability of the predicted response is in direct relationship to the width of the confidence level. The wider the interval, the more variability there will be at that particular setting. The standard error and confidence levels in the prediction are greatest for the penultimate set of conditions and lowest for the last set of conditions

Received for review March 3, 2000.

OP000024Q