Organic Process Research & Development **2006**, *10*, 64–69

Optimization of the Mizoroki−Heck Reaction Using Design of Experiment (DoE)

Varinder K. Aggarwal,[*,†] Anne C. Staubitz,[†] and Martin Owen[‡]

University of Bristol, School of Chemistry, Cantock's Close, Bristol, BS8 1TS, UK, and GlaxoSmithKline, Stevenage, UK

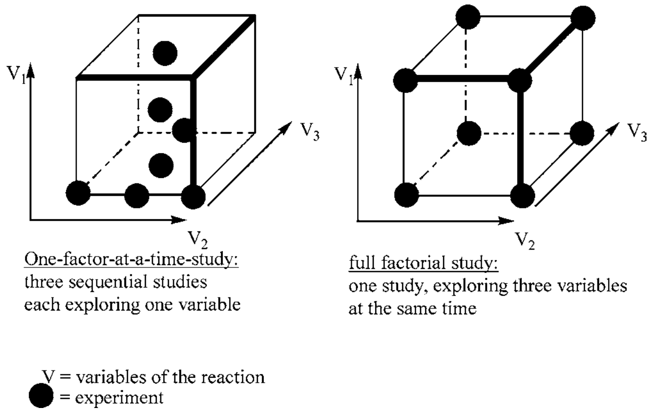
Abstract:

The influence of the variables concentration, temperature, time, and amounts of base and palladium on the yield of a Heck reaction was investigated using statistical methods (DoE) in a systematic sequential study. This revealed that temperature, concentration, and amount of palladium were the most important variables with a factor interaction between temperature and concentration. Through this approach, the yield of the Heck reaction was improved from 57% to 98% (89% isolated) with concomitant reduction of palladium loading from 3% to 0.5%.

Introduction

We recently required large amounts of 3-phenylcinnamaldehyde, which, because of its high cost, we elected to synthesize using the Heck reaction of iodobenzene with cinnamaldehyde. Such a reaction had been reported previously by Cacchi et al. but gave a mixture of starting material **1** and product **3** in a ratio of 41:59 (Scheme 1).[1]

Under their conditions we obtained similar results, but attempts to increase conversion by increasing the amount of palladium catalyst, base, and PTC were fruitless. Since a large number of variables can impact the success of this reaction (and palladium reactions in general), we decided to optimize the process by statistical methods (Design of Experiment, DoE).[2] Such methods are often practiced in industry for reaction optimization but very rarely in academia, where a one-factor-at-a-time approach is generally used. Although this intuitive approach can directly reveal how certain factors affect yield, it covers the parameter space very poorly and also reveals nothing about the interaction of certain factors. In contrast, using a DoE approach, the reaction space can be explored more efficiently, looking at as many dimensions as there are variables at the same time. Thus, a larger response space volume is covered, and at the same time factor interactions can be revealed (Figure 1). Also, several responses such as yield, residual starting material, and generated impurities can be modeled simultaneously at no extra experimental cost.



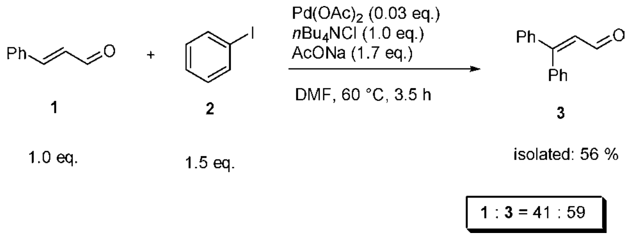
One-factor-at-a-time-study:
three sequential studies
each exploring one variable

full factorial study:
one study, exploring three variables
at the same time

V = variables of the reaction
● = experiment

Figure 1. **Comparison between the one-factor-at-a-time study and a full factorial study.**

Scheme 1. **Heck reaction for the preparation of 3**



In a sequential experimental design, the initial design will usually be a screening study (also referred to as a linear factorial study) which in many cases is refined by a more detailed second nonlinear study, such as a response surface study (RSM). By placing the settings for the variables to be examined of a factorial study at the extremes (low and high values) of the part of the parameter space one wishes to explore, a very large "volume" of this space is covered.

We thought that the variables concentration, temperature, time, amount of catalyst, and amount of base could all have an impact on yield (Table 1).[3] The rationale behind the selection of the ranges of the factors tested was that they should be sufficiently wide in order to get a clear picture of how these factors influence the conversion. Furthermore, when the values for the signal cover a wide range, it is more likely that the signal/noise ratio is acceptable even if the noise level is high. On the other hand, this range was restricted by practical considerations:

(1) The *volume of the solvent* could not be more than 5 mL because above this value sufficient heat transfer to the vessel in the parallel reactor could not be guaranteed.

(2) The *temperature* was set between 50.0 °C and 90.0 °C, to ensure that the reaction took place and stayed below the upper limit of the catalyst stability.[4]

---

[*] To whom correspondence should be addressed. E-mail: v.aggarwal@bristol.ac.uk.

[†] University of Bristol.

[‡] GlaxoSmithKline.

(1) Amorese, A.; Arcadi, A.; Bernocchi, E.; Cacchi, S.; Cerrini, S.; Fedeli, W.; Ortar, G. *Tetrahedron* **1989**, *45*, 813.

(2) For an introduction into DoE, see for factorial designs: (a) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experiments − An Introduction to Design, Data Analysis, and Model Building*; Wiley: New York, 1978. For response surface models, see: (b) Box, G. E. P.; Draper, N. R. *Empirical Model-Building and Response Surfaces*; Wiley: New York, 1987.

(3) Solvent, type of catalyst, and base might also have an effect. There are methods such as principle component analysis (PCA) to convert categorical variables into numerical variables, and had the current study not been successful, these would have been studied.

10.1021/op058013q CCC: $33.50

Table 1. **Ranges chosen for factors in the screening study (half factorial)**

factor	values chosen	initial values
volume/mL	1.6 to 5.0	2.0
T/°C	50.0 to 90.0	60.0
catalyst/equiv	0.005 to 0.050	0.030
base/equiv	1.5 to 4.0	1.7
time/h	4.0 to 9.0	3.5

(3) *Catalyst loading*: As the problem with this reaction might be the precipitation of the catalyst, the lower boundary was set very low. The upper value also reflected that the catalyst is expensive and therefore should be minimized as well.

(4) *Quantity of base*: The base had been varied by Cacchi et al., so the values chosen for this reflect a distribution of those used in the paper.[5]

(5) *Time*: It was not necessary to reduce the reaction time as this factor does not present a limitation in the laboratory. However, by including the factor time, information is gained on when the reaction is complete and the possible degradation of product. To establish when the samples should be taken, the reaction was run under the original conditions and samples were taken after 2 h 45 min, 4 h 30 min, and 8 h. There was no further conversion observed after 8 h. To be sure that in more extreme cases the end of the reaction could be observed, a longer time of 9 h was taken as the longest reaction time.

As strong factor interactions were suspected, a half factorial screening design was deemed appropriate. A full factorial on the other hand did not seem necessary because, with the half factorial, only three and four factor interactions are unresolvable. These are generally rare and are not likely to be significant. To get a clearer picture of the precision and the accuracy of the linear model, four centre points were run, with two together at the beginning of the study and two together at the end (to guard against reactor or time related trends).

All reactions were run in a parallel reactor. This not only allowed us to run a large number of reactions at the same time (though each run was constrained to be at the same temperature) but also helped to reduce any experimental bias, because the experimental setup was essentially the same for each experiment.

For statistical parameter optimizations it is vital that all experiments are performed under exactly the same conditions apart from those intentionally varied. In this case, this was done by using standardized reaction vessels in a parallel reactor and by always using the same batch of chemicals. Another important factor is randomization, wherever possible (e.g., in allocation to vessels or order of experiments and chemical analysis), to eliminate bias. All reactions were performed on a 1.0 mmol scale with respect to the aldehyde. With five variables and high and low values for each of them, 2^5 experiments are in principle required to resolve all effects by factors and factor interactions unambiguously.[6] However, as indicated above only a fraction (half or quarter) of these are required to generate a sufficient level of understanding of the factors and their interactions.

Results and Discussion

We elected to carry out a half factorial design with additional centre points.[7] These are design points that are placed in the center of the "volume" of the parameter space one wishes to explore and are not included in the model. Rather, they serve to determine the signal-to-noise ratio and can show whether curvature occurs, i.e., when they are not on the straight line connecting two design points, but significantly above or below.

These considerations led to a set of 20 reactions which were performed in parallel and yields were determined by GC−MS (calibrated). The results of this study can be visualized in a half normal plot[8] (Figure 2) where a statistically significant effect is shown by a factor or factor interaction straying from the indicator line.

From this it is clear that the most important factor is temperature (furthest away from the indicator line), followed by catalyst loading, concentration, and an interdependence of temperature and concentration (Figure 2). The finding of a significant interaction between two variables demonstrates the superiority of DoE compared to a one-factor-at-a time study where the interdependence of factors would have been missed. The amount of base and time of reaction[9] had no effect on yield. Interaction between variables can also be represented graphically by plotting yield against one of the variables overlaid with a second variable (Figure 3). If the lines run in parallel, there is no interaction between the two variables (temperature and catalyst loading), but if the lines divert, there is an interaction (volume and temperature).

Figure 4 shows the calculated model (2FI = Two Factor Interaction Model) for the yield in a cubic graph or a 3D surface graph. The four center-points gave yields between 44% and 50% which were above the value predicted by the linear 2FI model, thus indicating significant curvature. This was evidence that while the 2FI could be used as a rough guide to predict trends in the reaction, a more detailed study allowing for nonlinear terms would be needed to model the real situation more accurately. The fact that the center points give a rather broad range for the yield indicates that the reaction is very sensitive in this area to only subtle changes in the reaction conditions.

The following general conclusions were made:

(1) A higher temperature (variable D) results in higher yields, especially when the volume (variable A) is high

---

(4) Initially, we suspected that the catalyst stability would be reduced at higher temperatures. However, we were proven wrong, as the results of the factorial and later of the RSM study clearly showed.

(5) The rationale behind the values chosen in the paper is not clear since they vary with different bases.

(6) As the number of variables under consideration increase, the possible permutations of high and low settings increase rapidly.

(7) The calculations and randomisations for this optimisation were done using a commercially available computer program: *DesignExpert*, version 6.0.10; StatEase, Inc.: 2021 East Hennepin Ave., Suite 480, Minneapolis, MN 55413, USA, 2003.

(8) A half normal plot is a tool to visualize whether a factor or factor interaction has a significant effect. The red line is used to denote the factors that do not lead to a significant effect and follow the normal distribution. The further a factor falls to the right away from the line, the more pronounced the effect.

(9) Reactions were run for 4 and 9 h and showed no further conversion at the first time point. Product decomposition was not evident.

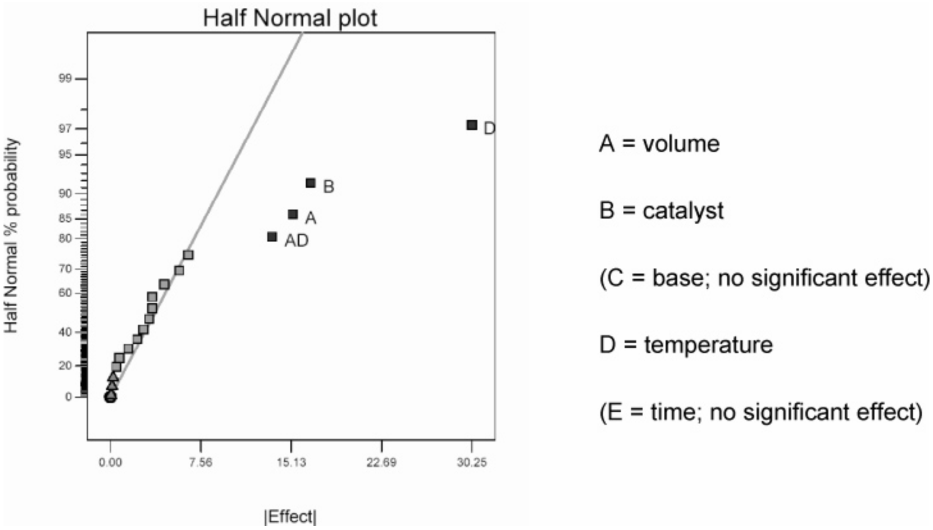
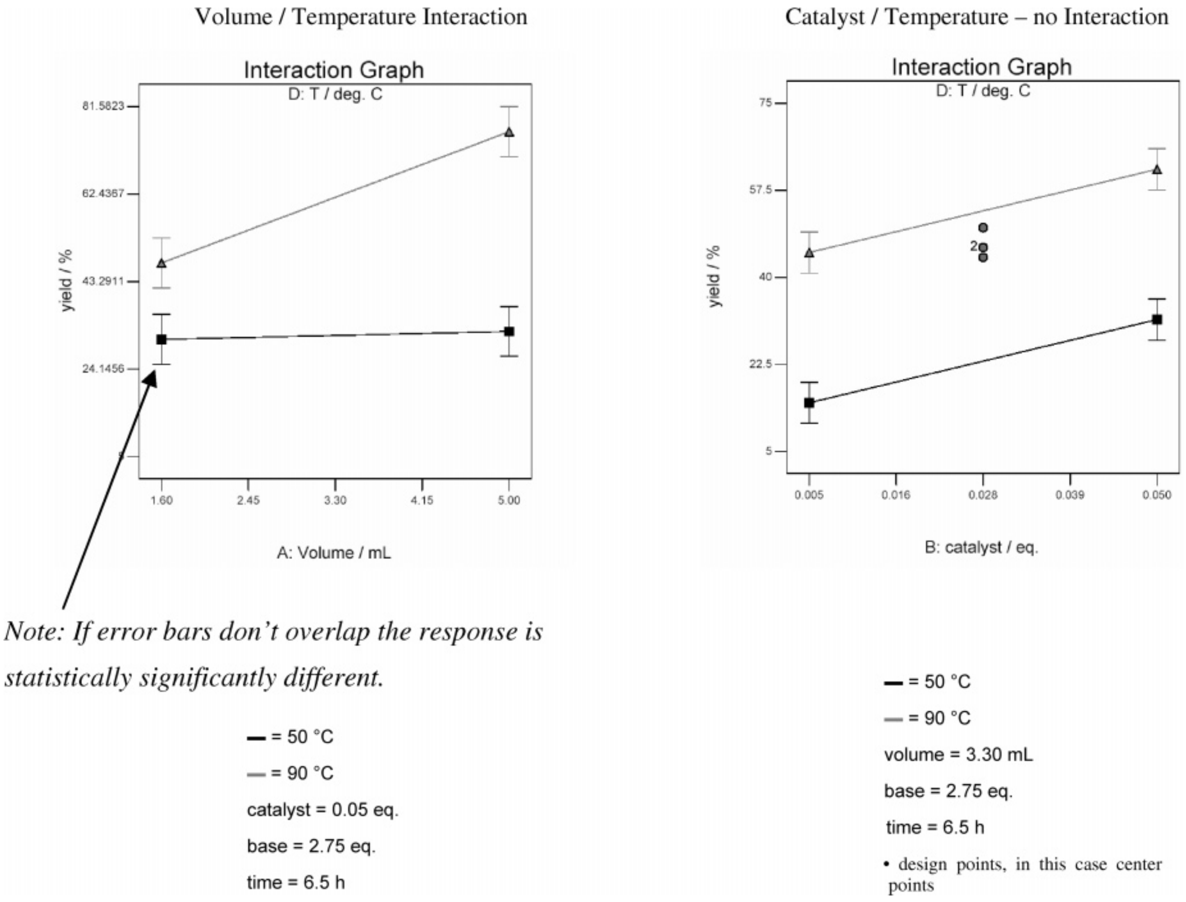


Figure 2. **Most important factors determining the yield of the Heck reaction.**


Volume / Temperature Interaction

Note: *If error bars don't overlap the response is statistically significantly different.*

Catalyst / Temperature – no Interaction

Figure 3. **Interaction graphs for the half factorial screening study.**

(factor interaction AD).

(2) Higher catalyst loading (variable B) gives higher yields, an effect which is more pronounced at higher temperatures. This effect is dependent on the volume and more pronounced if the volume is high.

(3) Equivalents of base (variable C) are virtually noninfluential (i.e., not significant) under all conditions tested and will not be included in further studies. Time (variable E) on the other hand was retained as a variable even though it was not significant in this study. Since the reaction did not go to completion, we wished to see if under different reaction conditions (see RSM study below) extending time had an effect, i.e., higher conversions or product degradation.

(4) The curvature found is significant. This is a clear indication that the linear assumptions made in a factorial model are not descriptive of the real response surface. Thus, any model derived from this can only give a rough indication for point predictions within the chosen boundaries. However, the overall trends as indicated by the values at the corners of the cube are valid (Figure 4).

We concluded from this study that high temperatures were required, probably higher than those we had set initially. At

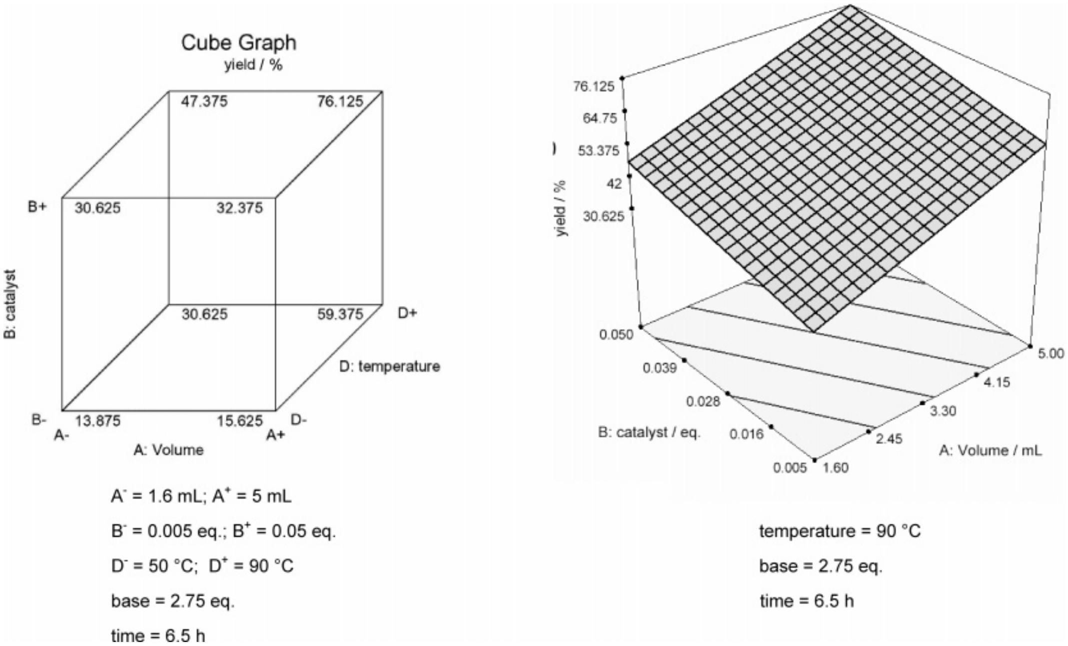


Figure 4. **Cubic and 3D view of the model for the half factorial model.**

this stage it seemed by extrapolation of the model that even better results might be obtained if temperature, dilution, and catalyst loading were increased. However, we were keen to keep loading of the costly Pd catalyst to a minimum. In general, high catalyst loading gave higher yields than low loading, but this effect was smaller at higher dilution. Thus, it was possible that, with further dilution and by operating at the optimum temperature, catalyst loading could be reduced without impacting negatively on yield.

Therefore, we initiated a second study, exploring higher dilutions together with higher temperatures, while again examining catalyst loading and time as a factor.

Since factorial studies are intrinsically linear (they only have two levels), they connect all points with straight lines and cannot model a mathematical optimum. When the center points are indicative of significant curvature, then the approximate linear model of the reaction can be refined by a more detailed second nonlinear study, such as a response surface study (RSM). RSM designs are particularly useful for process optimizations and are often used to augment a factorial design, if it shows significant curvature, to get a detailed image of the response surface. However, in this case, it was clear that the optimum area to operate this reaction was outside the one chosen for the factorial, as the highest yield was found in a corner of the cubic graph. Therefore, the RSM design chosen was set up to be analyzed independently from the screening study, i.e., without using those values for the model. The points for the RSM model were chosen according to a central composite design (CCD). This design provides estimations for all coefficients in a quadratic model, while still employing relatively few experiments. To estimate coefficients for second-order terms using a factorial design, it would be necessary to carry out a 3^k factorial design, with $k =$ number of variables tested. This would lead to a large number of experiments, even if k were to be only a small number. To reduce this experimental expenditure, the CCDs are composed of 2^k points in the design space augmented by $2k$ points on the k axis, equally spaced at $\pm\alpha$ units, and n center points (Figure 5). In the present study we used a CCD, which comprises a factorial design with additional axial points and six center points, leading to a set of 30 experiments. It is important that n should be a reasonably high number as this allows for good estimation of curvature and pure experimental error variance. While this is also true for pure factorial designs with center points, it becomes more important for estimation of higher order terms.

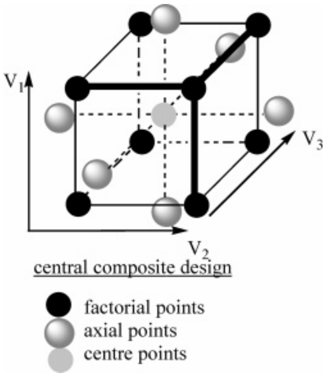


Figure 5. **Distribution of experiments for an RSM study in the parameter space.**

The factors studied in the second study were volume, temperature, catalyst loading, and time (Table 2):

(1) The *volume of the solvent* clearly proved to be one of the most important factors. Its range was set between 5 and 10 mL.[10]

(2) The *temperature* was set between 90.0 °C and 120.0 °C, as high-temperature had proved to be beneficial.

(3) *Catalyst loading*: From the factorial it was clear that higher catalyst loading contributed to higher conversions. But looking at the cubic graph, it seemed also possible to achieve good results even with lower catalyst loading and it

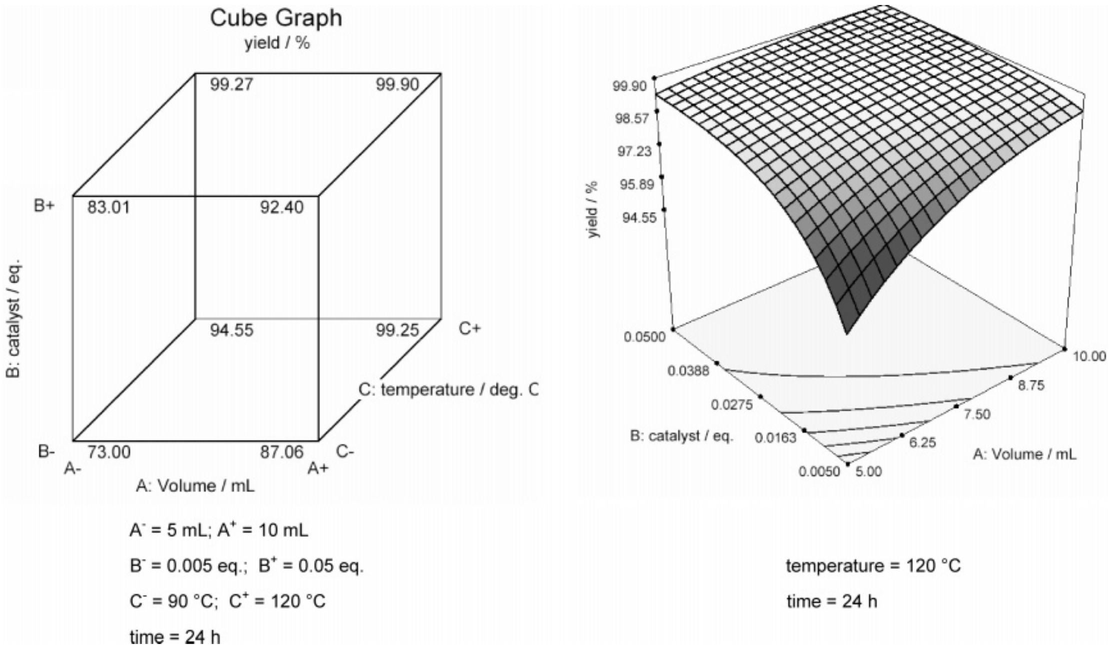


Figure 6. **Graphical analysis for the product.**

Table 2. **Values for the optimization (RSM) study**

factor	values chosen	best values from the factorial optimization (76%)
volume[a])/mL	5.0 to 10.0	5.0
temperature/°C	90.0 to 120.0	90.0
catalyst/equiv	0.005 to 0.050	0.050
time/h	8.0 to 24.0	8.2

[a] This is given relative to the half factorial study for better comparison. However, due to the dimensions of the reaction vessels in the carousel, every reaction was run on half the quoted scale.

was desirable to find out how low it was possible to go while still obtaining good yields. Therefore, the setting was again between 0.005 equiv and 0.05 equiv.

(4) *Time* was included as a factor since, although in the laboratory there is no strong need to reduce the time taken, it is desirable to achieve complete conversion, yet higher temperatures sometimes lead to decomposition of products or starting materials. Therefore, the low level was set at 8 h, and the high level, at 24 h.

For the present study the value for α was set to be 1, so that only three different temperatures instead of five had to be taken into account (face centered CCD) minimizing the number of times the parallel reactor block had to be used. For points which differed only in respect to time, the same sample was measured twice at the given times in order to reduce the number of actual experiments. While this will affect the validity of the statistical analysis (typically it will underestimate the background noise thus identifying more effects as significant than should truly be the case), it was practically very desirable and thought unlikely to appreciably distort the model. Analysis and calibration were done in the same manner as that for the screening study.

Viewing the quadratic model for the yield as cubic and 3D surface graphs (Figure 6), the following conclusions can be made:

(1) The higher the temperature, the better the yield, especially if the volume is large.

(2) The higher the catalyst loading, at high volumes, the higher the yield. BUT, if volume and temperature are high, even a low catalyst loading leads to excellent yields.

Table 3. **Verification experiments**

conditions	*T*/°C	catalyst loading/ equiv	volume/ mL	expected yield/%	GC−MS yield/%[a]	isolated yield/%
original conditions (base, 1.7 equiv; *t*, 24 h)	60	0.03	4	33[b] (59)[c]	56 59 58	39 41 40
conditions judging the trends shown in the half factorial (base, 1.5 equiv; *t*, 24 h)	120	0.05	10	>76	99 99 99	91 93 90
predicted best conditions after RSM (base, 1.5 equiv; *t*, 24 h)	120	0.005	10	98[d]	98 98 98	88 90 89

[a] Calibrated. [b] From the 2FI model. [c] Isolated yield of Cacchi et al. [d] RSM model.

(3) Time is again not very important, apart from the fact that some reactions were incomplete after 8 h.

Now, when reactions were conducted at 120 °C, very high yields were obtained, even at 0.5% catalyst loading, provided reactions were conducted at high dilution.

The models generated by both the screening and the optimization studies are approximations of reality. They are used to predict the favored settings and then these settings can be implemented in order to validate the model.

To substantiate the findings from both, the factorial and the RSM study, three verification experiments were performed three times: the original conditions, the conditions thought best after performing the half factorial design, and the conditions found in the RSM study (Table 3). Each of these experiments was measured three times by GC−MS in order to get a clearer picture of the precision of the analysis.

The optimum conditions arrived at through the statistical approach (high temperature, low catalyst loading, high dilution) can be rationalized in chemical terms. The high temperature gives faster rates of all the steps involved in the Heck reaction, and the low catalyst loading with high dilution prolongs the longevity of the Pd catalyst, thus resulting in full conversion. Indeed, the low catalyst loading with high dilution will reduce the likelihood for the Pd (0) nanoparticles,[11] which may be the catalytically active species, to conglomerate and precipitate out as unreactive Pd black. In fact, the use of "homeopathic" quantities of $Pd(OAc)_2$ has recently been advocated in related reactions.[12]

Conclusions

In conclusion, we have shown how DoE can be effectively used to rapidly (3 weeks) optimize a complex chemical process, the Heck reaction, involving five variables. Two variables, concentration and temperature, were found to be dependent on each other. The optimum conditions using low catalyst loading were found to require high temperatures and low concentrations. We believe that this results in rapid reaction rates with minimization of precipitation of Pd black through conglomeration of Pd (0) nanoparticles. Although these conditions are close to the Jefferey's conditions[13] employed for other Heck reactions, the sensitivity of such reactions to substrate structure invariably leads one to embark on reaction optimization from literature conditions of a closely related (or the same) substrate. However, the literature conditions in our case were far from optimum. The high temperature, low catalyst loading, and high dilution that were found to be optimum are beginning to emerge as a common set of conditions for the ligand-free Heck reaction, and we believe these should now be the starting point for most substrates.

Experimental Section

3,3-Diphenylacrylaldehyde 3[14] (Optimized Procedure): A mixture of NaOAc (123.0 mg, 1.5 mmol), $(n\text{Bu})_4\text{NBr}$ (322.4 mg, 1.0 mmol), iodobenzene (168 μL, 1.5 mmol), and *trans*-cinnamaldehyde (126 μL, 1.0 mmol) in DMF (7.5 mL) was heated to 120 °C. $Pd(OAc)_2$ (1.1 mg, 5 μmol) was dissolved in DMF (2.5 mL). After the catalyst had dissolved completely, this solution was added to the reaction mixture. After 24 h, the reaction mixture was cooled to rt and poured on $NaHCO_{3\text{half}-\text{sat.}}$ (80 mL) and extracted with EtOAc (3 × 40 mL). The combined organic phases were washed with brine (40 mL), dried over Na_2SO_4, and concentrated in the rotary evaporator. To remove further residual DMF, the crude mixture was dried at 60 °C under high vacuum. Then the product was purified by flash chromatography (petrol ether/EtOAc = 20/1). The product was isolated as a yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$): 9.53 (d, J = 8.3 Hz, 1H, =CH−C*H*O), 7.54−7.19 (2 m, 10H, Ar−*H*), 6.60 (d, J = 8.3 Hz, 1H, =C*H*−CHO); ^{13}C NMR (100 MHz, $CDCl_3$): 193.2, 162.0, 139.7, 136.7, 130.6, 130.4, 129.6, 128.6, 128.3, 127.3; MS (EI): 208 (M^+, 72), 207 (100), 179 (35), 178 (53), 177 (10), 176 (13), 165 (16), 152 (13), 102 (32), 77 (10).

Acknowledgment

We thank GSK (in particular Dan Tray and Marion Chatfield for useful discussions and Andy Reason for sponsoring a 2 day Experimental Design Workshop for the Chemical Education Community). A.S. thanks the Fonds der chemischen Industrie (FCI) for a Kekulé fellowship.

Supporting Information Available

Tables showing individual results of runs, details of statistical analysis, expanded discussions of peripheral findings, experimental procedure, analytical data. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review August 12, 2005.

OP058013Q

---

(10) As all reactions had to be done in the parallel reactor, this meant that to achieve this dilution the complete scale (and thus the volumes) had to be divided by two. However, to give some comparability to the factorial design, the tables will show the values in a molar volume scale (mL/mmol 18). For simplicity, this will be hinted at in the Supporting Information.

(11) (a) Reetz, M. T.; Westermann, E. *Angew. Chem., Int. Engl. Ed.* **2000**, *39*, 165. (b) Reetz, M. T.; Lohmer, G. *Chem. Commun.* **1996**, 1912.

(12) (a) de Vries, A. H. M.; Mulders, J. M. C. A.; Mommers, J. H. M.; Henderickx, H. J. W.; de Vries, J. G. *Org. Lett.* **2003**, *5*, 3285. (b) Reetz, M. T.; de Vries, J. G. *Chem. Commun.* **2004**, 1559.

(13) For a discussion of Jefferey's conditions, see: (a) Jeffery, T.; David, M. *Tetrahedron Lett.* **1998**, *39*, 5751. (b) Jefferey, T. *Tetrahedron* **1996**, *52*, 10113.

(14) First characterization (microanalysis, mp): (a) Kohler, E. P.; Larsen, R. G. *J. Am. Chem. Soc.* **1935**, *57*, 1448. For NMR, mass, IR, see: (b) Kurosawa, K.; Tsujita, T. *Bull. Chem. Soc. Jpn.* **1981**, *54*, 2391.