

# Using Design of Experiments To Assess *Escherichia coli* Fermentation Robustness

Arun Tholudur, Todd Sorensen, Xiaoyue Zhu, Scot Shepard

**A**s initial bench-scale processes progress to full-scale commercial manufacturing, it is important to identify critical parameters and assess their robustness. If manufacturing deviations occur, researchers with process robustness data in hand have critical information with which to evaluate the effects of those deviations. With this in mind, we recently conducted a statistical study to assess fermentation process robustness in the manufacture of an active pharmaceutical ingredient.

Design of experiments (DOE) is a structured and efficient methodology for planning experiments such that statistically valid relationships between factors affecting a process and its outputs can be established. DOE was a key element in our study because it is ideally suited to establishing critical parameters and their acceptable ranges.

To begin the study, we first scaled down an *Escherichia coli* fed-batch fermentation process that was already in full-scale operation producing a recombinant protein in the form of inclusion bodies. Next, using historical

data that was thus available, we identified parameters that were likely to affect process performance. Finally, we performed statistically designed experiments to assess the impact of those process parameters. As a result, we were able to establish a data set to support broader parameter ranges for routine manufacturing.

## FERMENTATION PROCESS AND FACTORS

A two-stage seed-expansion procedure provided 100-mL of culture that we used to inoculate 15-L production vessels containing 9.1 L of complex fermentation medium with glycerol as the primary carbon source. During the fermentation, we maintained temperature at 37 °C, pH at 7.10, and dissolved oxygen (DO) at 15% of air saturation. Agitation and airflow rates remained constant throughout fermentation while pure oxygen was sparged to maintain DO at the desired set point. The pH of the medium was maintained at its set point by means of base (NaOH) addition.

Following an initial batch growth phase, when optical density ( $OD_{600}$ ) reached  $10 \pm 1$  absorbance unit (AU), two constant-rate feed solutions began: one solution a yeast extract/glycerol (YEG) feed and the other a lactose feed. Adding the lactose solution initiated product expression. After 4.5 hours, the feeds were stopped and the cells were harvested.

We studied process robustness by focusing on the six factors below. The first three specifically affect cellular growth rates and product formation.



WWW.PHOTOS.COM

**Factor A:** culture-medium temperature

**Factor B:** fermentation-medium pH

**Factor C:** DO concentration

**Factor D:** induction cell density ( $OD_{600}$ ) — cell density at which feeds initiate

**Factor E:** yeast extract/glycerol feedrate — carbon/nitrogen mixture

**Factor F:** lactose feedrate — induces product formation

We evaluated the following three responses.

**Response 1:** harvest cell density ( $OD_{600}$ ) — acceptable harvest range compatible with downstream processing capacity

**Response 2:** product band purity — desired product protein amount

**PRODUCT FOCUS:** FERMENTATION PRODUCTS

**PROCESS FOCUS:** PRODUCTION

**WHO SHOULD READ:** PROCESS DEVELOPMENT, MANUFACTURING, AND ANALYTICAL STATISTICIANS

**KEYWORDS:** ROBUSTNESS, OPTIMIZATION, DESIGN OF EXPERIMENTS, FED-BATCH, *E. COLI*

**LEVEL:** INTERMEDIATE

**Figure 1:** DOE design table. Six factors and three responses appear in their standard run order to summarize the experimental design. This matrix is the initial step in identifying critical process parameters and ranges. To protect proprietary data, actual responses are omitted here.

| Std | Run | Block | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Response 1 | Response 2 | Response 3 |
|-----|-----|-------|----------|----------|----------|----------|----------|----------|------------|------------|------------|
| 1   | 2   | 1     | 35       | 6.8      | 5        | 6        | 3.2      | 4.8      |            |            |            |
| 2   | 10  | 2     | 39       | 6.8      | 5        | 6        | 4.8      | 7.2      |            |            |            |
| 3   | 11  | 3     | 35       | 7.4      | 5        | 6        | 4.8      | 7.2      |            |            |            |
| 4   | 17  | 4     | 39       | 7.4      | 5        | 6        | 3.2      | 4.8      |            |            |            |
| 5   | 18  | 4     | 35       | 6.8      | 40       | 6        | 4.8      | 4.8      |            |            |            |
| 6   | 15  | 3     | 39       | 6.8      | 40       | 6        | 3.2      | 7.2      |            |            |            |
| 7   | 8   | 2     | 35       | 7.4      | 40       | 6        | 3.2      | 7.2      |            |            |            |
| 8   | 3   | 1     | 39       | 7.4      | 40       | 6        | 4.8      | 4.8      |            |            |            |
| 9   | 19  | 4     | 35       | 6.8      | 5        | 20       | 3.2      | 7.2      |            |            |            |
| 10  | 13  | 3     | 39       | 6.8      | 5        | 20       | 4.8      | 4.8      |            |            |            |
| 11  | 6   | 2     | 35       | 7.4      | 5        | 20       | 4.8      | 4.8      |            |            |            |
| 12  | 5   | 1     | 39       | 7.4      | 5        | 20       | 3.2      | 7.2      |            |            |            |
| 13  | 4   | 1     | 35       | 6.8      | 40       | 20       | 4.8      | 7.2      |            |            |            |
| 14  | 9   | 2     | 39       | 6.8      | 40       | 20       | 3.2      | 4.8      |            |            |            |
| 15  | 14  | 3     | 35       | 7.4      | 40       | 20       | 3.2      | 4.8      |            |            |            |
| 16  | 20  | 4     | 39       | 7.4      | 40       | 20       | 4.8      | 7.2      |            |            |            |
| 17  | 1   | 1     | 37       | 7.1      | 15       | 13       | 4.0      | 6.0      |            |            |            |
| 18  | 7   | 2     | 37       | 7.1      | 15       | 13       | 4.0      | 6.0      |            |            |            |
| 19  | 12  | 3     | 37       | 7.1      | 15       | 13       | 4.0      | 6.0      |            |            |            |
| 20  | 16  | 4     | 37       | 7.1      | 15       | 13       | 4.0      | 6.0      |            |            |            |

relative to other cellular proteins being made (Ideally, batch-to-batch variation should remain relatively low.)

**Response 3:** inclusion bodies (IB)  
— actual product concentration

### EXPERIMENT DESIGN

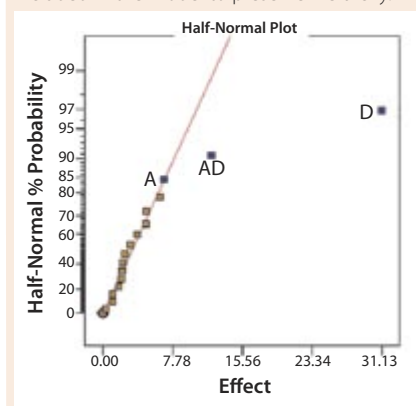
Factorial designs are a type of DOE that allow researchers to efficiently explore potentially interacting effects of multiple factors on process outputs. Fractional factorial designs are a subset of full factorial designs (using fewer runs) that can be useful when attempting to establish process robustness — in which the desired outcome is a lack of effect of various factors on process outputs.

With six factors, several fractional two-level factorial designs are possible. A one-half fractional factorial requires 32 experiments, whereas one-quarter and one-eighth designs using six factors require 16 and 8 experiments, respectively. With higher fractional factorial designs, fewer experiments are required; however, estimation of effects becomes complicated because of aliasing (a confounding of effects that prevents us from unambiguously estimating some effects and/or interactions), which is the price we pay for running a reduced number of experiments.

We applied a 1/4-fractional factorial design that allowed us to estimate main effects unambiguously. That proved to be a good compromise between number of experiments and data quality, allowing us to draw reasonable conclusions. The high and low levels we selected for this study represented the extremes of normal operating ranges, which accentuated all effects while we tested 16 unique factor combinations. Those 16 combinations ran in four blocks of five experiments, with a center point included in each block. Thus, 20 fermentations were performed in all. The entire design matrix appears as Figure 1. To protect proprietary data, all response columns in the figure have been left blank.

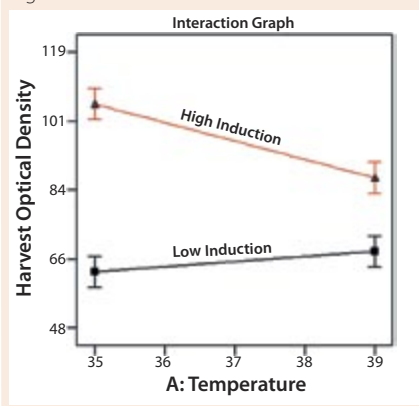
We used a DOE software package (Design-Expert software from Stat-Ease, Inc., [www.statease.com](http://www.statease.com)) to calculate all effects that could be estimated from the data, and then we assessed their significance. Subsequent software analysis found our fermentation process to be robust with respect to purity percentage of inclusion bodies and protein concentration. However, induction OD and temperature affected harvest OD, as depicted in Figure 2. Equally important, an interaction occurred

**Figure 2:** Half-normal probability plot. Induction OD (D) and the temperature-induction OD interaction (AD) significantly affect harvest OD. Although temperature, by itself, is not a significant parameter, it is included in the model to preserve hierarchy.



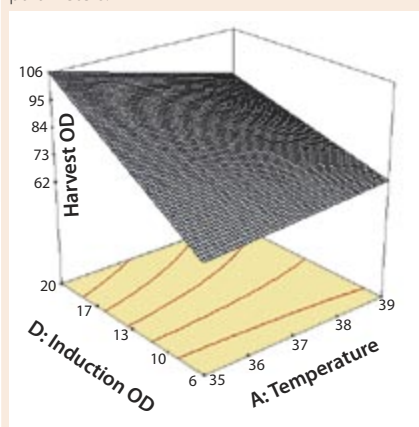
**Aliasing (a confounding of effects that makes results ambiguous) is the PRICE we pay for running a reduced number of experiments.**

**Figure 3:** Interaction graph. When cells are induced at the low level, harvest OD is insensitive to temperature. However, wide variations in harvest OD, with respect to temperature, occur when cells are induced at higher ODs.

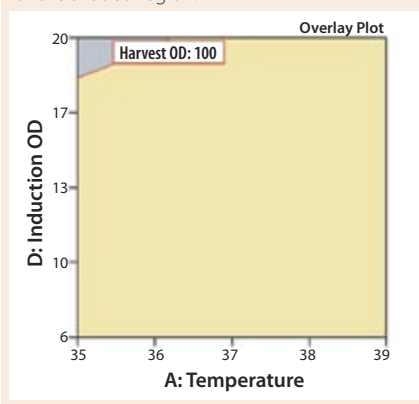


between induction OD and

**Figure 4:** Harvest OD 3-D plot. Harvest OD appears as a function of induction OD and temperature. This graphic analysis indicates harvest OD's lack of sensitivity to specific parameters.



**Figure 5:** Overlay plot. Shaded area in upper left shows temperature and induction OD combinations resulting in harvests greater than 100 AU. Because specifications are set at 50–100 AU, any batch resulting in harvest OD > 100 AU could be rejected. This plot reveals that temperature and induction OD ranges are almost entirely within range except for the shaded region.



temperature. (Because the induction OD–temperature interaction term is integral to our model, the parent temperature term is included to preserve hierarchy even though it does not significantly affect harvest OD on its own.)

Figure 3 shows the induction OD–temperature interaction effect on harvest OD. Harvest OD is relatively insensitive to temperature at low induction ODs, but the graph reveals the temperature effect to be more dramatic at high induction ODs. Harvest OD insensitivity to temperature is also revealed in Figure 4, which is a three-dimensional plot of harvest OD as a function of induction OD and temperature. The plot indicates insensitivity of harvest OD to temperature excursions.

Figure 5 is an overlay plot generated by setting ranges on input and output variables in the DOE software. Because our manufacturing process has a specification of 50–100 AU for harvest OD, we sought factor combinations resulting in harvest OD specification violations. The overlay plot indicates that within the ranges of induction OD and temperature studied, harvest OD is always >50 AU. However, at high induction ODs and low temperatures, harvest OD can exceed 100 AU.

Although current manufacturing specifications state that temperature should be maintained at  $37 \pm 1$  °C, our study indicated that the process is robust with respect to temperatures over  $37 \pm 2$  °C for a wide range of induction ODs. As a result, broader acceptable operating ranges have been established for temperature. Although manufacturing specifications were not changed, data from this study can be used in technical evaluations and to close out any deviations that may occur during routine manufacturing.

### THE DOE HORIZON

We have developed processes that are in use at our company for producing clinical and commercial recombinant protein products. Our goal is to develop cGMP production processes that are economically viable, scalable, and

robust. As well as developing processes “from scratch,” we are significantly enhancing existing processes for transfer to full-production facilities.

DOE is helping us select and develop serum-free and protein-free media formulations in addition to nutrient feed strategies that optimize cell growth and protein productivity. DOE is also being used for process-robustness pilot studies in preparation for process validation and routine commercial manufacturing. Finally, as demonstrated in this article, DOE also can be used for further characterization of process ranges even as routine manufacturing continues.

### FOR FURTHER READING?

1 Anderson MJ, Whitcomb PJ. *DOE Simplified: Practical Tools for Effective Experimentation*. Productivity Press: New York, NY, 2000.

2 Montgomery DC. *Design and Analysis of Experiments*, 6th Edition. John Wiley & Sons, Inc.: Hoboken, NJ, 2004. 🌐

Corresponding author **Arun Tholudur** is a senior scientist in the upstream development and production group at Diosynth Biotechnology, 3000 Weston Parkway, Cary, NC 27513; 1-919-388-5721, fax 1-919-678-0366; arun.tholudur@diosynth-rtp.com; www.diosynthbiotechnology.com. **Todd Sorensen** and **Xiaoyue Zhu** are scientists, and **Scot Shepard** is director of the upstream process development group at Diosynth Biotechnology. The authors' primary focus is on developing and implementing scalable microbial fermentation and animal cell culture processes for manufacturing active pharmaceutical ingredients. Their company is a contract manufacturer with process development and manufacturing operations in the United States and Europe experienced in biopharmaceutical production using a wide range of cell types including mammalian, hybridoma, and insect cells, E. coli, and Pichia pastoris.