PCR Process Optimized via Split-Plot DOE

Pat Whitcomb
Principle
Stat-Ease, Inc.
2021 E. Hennepin Ave.
Suite 480
Minneapolis, MN 55413
E-mail: pat@statease.com
Web: http://www.statease.com

Shari Kraber
Statistical Consultant
Stat-Ease, Inc.
2021 E. Hennepin Ave.
Suite 480
Minneapolis, MN 55413
E-mail: shari@statease.com
Web: http://www.statease.com

SUMMARY:

ASQ members are often looking for talks that illustrate the successful application of a quality tool. This paper illustrates the use of design of experiments (DOE) and split-plot design to quickly and effectively determine the factor settings that maximize amplification in a polymerase chain reaction (PCR) experiment. DOE dramatically increases the amount of information that can be obtained from any set of experimental runs. The addition of the split-plot design allows accurate evaluation of the factor effects by assigning the experimental error correctly. This paper and the associated presentation assume the audience has basic knowledge of DOE.

INTRODUCTION TO POLYMERASE CHAIN REACTION (PCR):

Polymerase chain reaction (PCR) is a relatively new method of generating unlimited copies of any fragment of DNA. Developed in the mid-1980's, we can consider it to be a biological copying machine. The process has revolutionized the exploration and understanding of DNA. PCR can take an extremely small amount of DNA, say from a fingerprint left on a glass, or even from an ancient skull, and reproduce it so that it can be analyzed and studied more easily. The method has a wide range of application, from identifying the presence of genetic defects, to DNA fingerprinting, to understanding the relationships between various animal species.[1]

The PCR method is quick and easy to perform compared to older DNA replicating methods. But, like any process, there is always room for improvement. Although the method is generally pre-determined, there are variables in the process that can be changed to produce optimal results for a specific situation.

There are three basic steps in PCR[2]:

1. A small quantity of target DNA is added to a test tube containing things like primers (short chains of the four different chemical components that make up any strand of genetic material), Taq polymerase (a bacterium that survives the heating process), and the cofactor $MgCl_2$. Once the test tube is prepared with the necessary reagents, the DNA is denatured by heating it. This breaks the classic DNA double helix into two separate strands.
2. Then cooling takes place, which allows the primers to hybridize, or anneal, to their complementary sequences.
3. Finally, the polymerase binds together the primers with the original DNA, resulting in a complementary DNA strand from each primer.

This process creates double the original amount of DNA. It is repeated as necessary to obtain the desired quantity of DNA.

DESIGNING THE EXPERIMENT:

A research company wanted to optimize the PCR method for their particular application. Traditional experimentation would have dictated they try varying factors one at a time (OFAT.) This keeps things simple, but it's inefficient and reveals nothing about interaction effects. Interactions occur when two things changed at the same time cause results that differ from varying each factor individually. For instance, in the PCR process, the results depend on both the time allowed for denaturing and its temperature. Interactions like this can only be studied in a factorial experiment.

The PCR researchers chose nine factors to study. All other controllable variables were held constant. Three factors (A, B, C) deal with the thermocycler settings. This is a machine that heats and cools samples in a pre-designated sequence. The other six factors (D, E, F, G, H, J) are associated with the sample preparation. The objective of the DOE was to find factor settings that would maximize the amplification of the DNA sample.

	Factor	Units	Low Level	High Level
Thermocycler	A: Anneal Temperature	°C	55	65
	B: Denature Temperature	°C	90	97
	C: Denature Time	seconds	10	20
Sample Preparation	D: Forward Primer	nM	200	900
	E: Reverse Primer	nM	200	900
	F: DNA probe	nM	100	400
	G: MgCl2	nM	3	9
	H: Tween (PCR) buffer	%	0.005	0.020
	J: Taq polymerase	u/µl	0.01	0.04

Due to the inherent complexity of this process, it is desirable to get complete information on all nine main effects, plus all two-factor interactions. This means that a resolution V design is needed. For nine factors, a half-fraction two-level factorial design requires 256 runs (2^{9-1}). Ideally, all combinations are done randomly. But, for the PCR process, this would require preparing one sample at a time and then processing it through the thermocycler. The normal randomization process would require far more time than researchers are willing to give. A restriction on the randomization is required.

A SHORT PROCESS EXPLANATION:

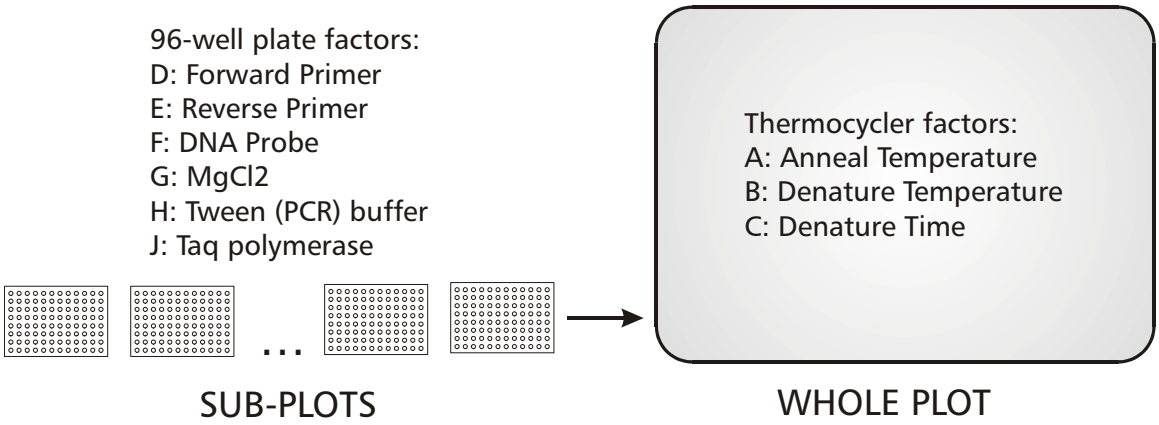
At this point, a bit more explanation of the experimentation process is required. In the past, researchers had to make test samples by hand, combining chemicals one at a time in a single test tube. Now, many companies are taking advantage of automated liquid handling systems, such as Beckman

Coulter's Biomek® FX, that can automatically transfer the correct amount of reagents into individual wells on a plate. 96-well plates are commonly used. Within these 96-well plates, the edge wells are often restricted from use due to edge effects (e.g. heat transfer, evaporation, plate reader limitations, etc.) and a percentage of the other wells may be reserved as quality control wells. A quality control well contains a known chemical standard, often the base reagents that will not be changing in the experiment. Because of these restrictions, there are not 96 wells available for the experiment, but only a designated subset.

The experiment on the PCR process contained 256 runs, or samples that had to be processed. Each combination of the three thermocycler factors must be run using a different plate, so eight (2^3) plates are required. As a result of splitting the 256 runs between the 8 plates, each plate contained 32 experimental runs, in addition to the quality control wells and restricted wells. The randomization of the design is restricted by the processing of the plates through the thermocycler.

CREATING THE SPLIT-PLOT DESIGN:

A split-plot design[3] is one in which there is a restriction on the randomization of all the factors. The sample preparation factors are randomized within each plate. The thermocycler conditions are randomized across the plates. Thus, the amount of error in the thermocycler factors will be different than the amount of error in the sample prep factors. This situation can be properly taken into account during the data analysis. Since the three thermocycler factors (A, B, and C) are varied between plates, they are called the "whole plot" (in this case "whole plate") factors. The six sample prep factors (D, E, F, G, H, J) are varied within plates, so they are called the "sub-plot" (in this case "sub-plate") factors.



ANALYSIS OF A SPLIT-PLOT DESIGN:

A standard two-level factorial design is typically analyzed by first generating a normal or half-normal probability plot of effects. This is a tool that allows the experimenter to determine which of the factor effects do not fit a normal distribution, and thus are likely to be the statistically significant effects. This method assumes that the error for all the effects is approximately constant. In the PCR (split-plot) situation, this assumption is not valid. The randomization of the experiment was restricted by the fact that many samples are processed in the thermocycler at the same time. Since the samples on any one 96-well plate only encounter one set of thermocycler conditions, the data from those samples will not include the error associated with the setup of the thermocycler.

The solution to this problem is to split the factors into their natural groups (whole plot and sub-plot) and test them using two separate half-normal probability plots. This distributes the error in the same way that it was distributed during the actual procedure.

Here's how the factors and all the interactions are divided:

- Whole Plot Group: A, B, C, and interactions involving only these factors
- Sub-Plot Group: D, E, F, G, H, J, and all interactions involving these factors, eg AD, DE, etc.

Figure 2 shows the two half-normal probability plots. The points not on the line (named with letters) are the terms that are statistically significant. We can see that the significant factors are A, B, (from Figure 2a) and D, F, J, interactions AD and FJ (from Figure 2b.) If all the terms had been pooled together into a single half-normal probability plot, the results would have been slightly different and can easily lead to incorrect conclusions. On this combined plot the error is too small for the whole plot factors, leading to over-selection. (Find the two extra interactions on the incorrect plot in Figure 3.) The error is too large for the sub-plot factors, often hiding significant effects.

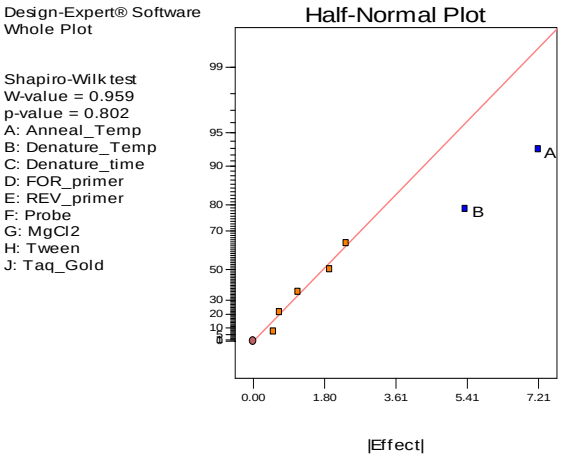


Figure 2a: Whole Plot Effects

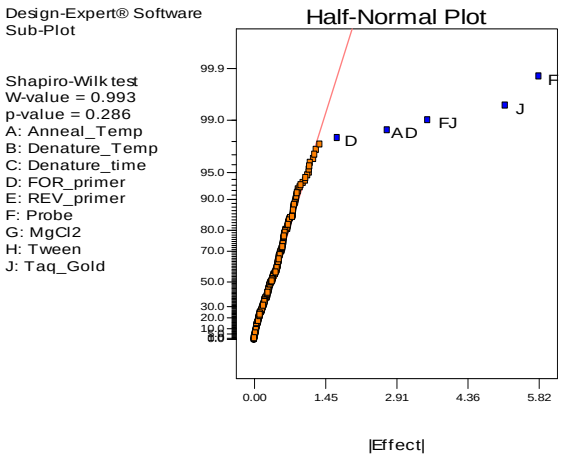


Figure 2b: Sub-plot Effects

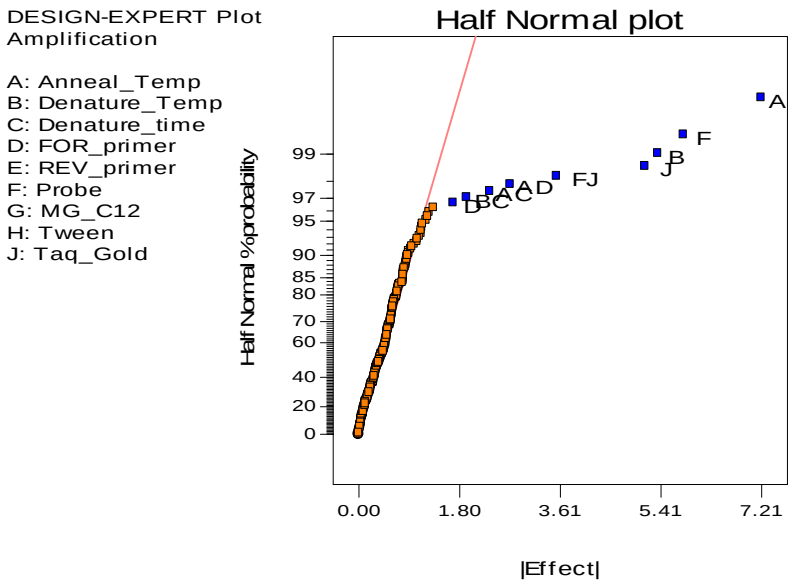


Figure 3: Incorrect Effect Plot (not taking into account split-plot nature of this design)

Once the significant effects are identified, they are combined into a prediction equation (Equation 1.) From this equation, main effect and interaction graphs can be generated that identify the optimal process settings (Figures 4-6.)

Eqn. 1: Amplification = 1.84 – 3.61A + 2.68B + 0.85D – 2.91F – 2.57J + 1.36AD + 1.77FJ

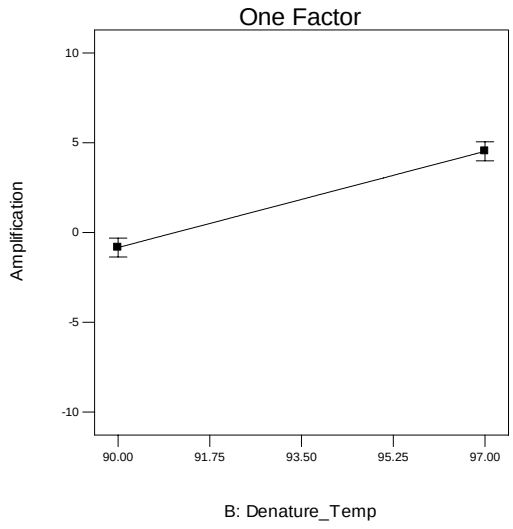


Figure 4: Main Effect of B

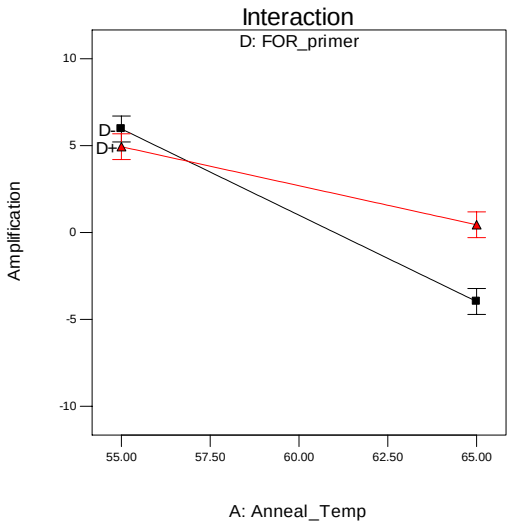


Figure 5: AD Interaction Effect

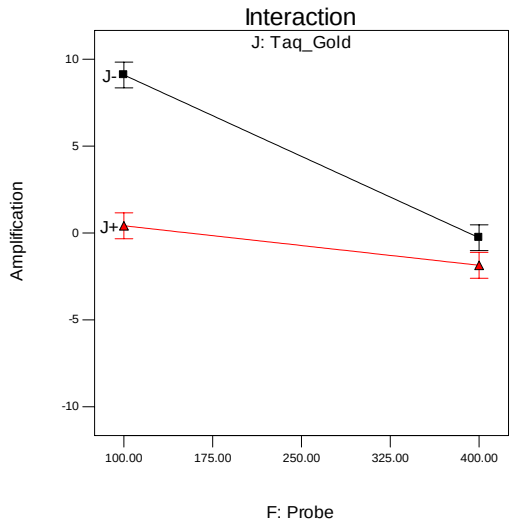


Figure 6: FJ Interaction Effect

From these graphs, the maximum amplification is obtained at the higher denatured temperature (B = 97 °C), the lower anneal temperature (A = 55 °C), the lower amount of DNA probe (F = 5 ng/µL), and the lower amount of Taq polymerase (J = 0.1 units). Factor D, the forward primer, is not significant when factor A (anneal temperature) is set to its lower value.

CONCLUSION:

The application of design of experiments and especially the introduction of the split-plot design, allowed these researchers to optimize a complex process with many interactive factors. As a result, their DNA research was accomplished more rapidly and more accurately. Design of experiments has become an essential tool in both pharmaceutical and biological research companies. It is a key tool that leads to understanding extremely complex situations. Often these companies are not only delving into complex areas of research, but they are achieving productivity gains by using processing equipment such as robotics and thermocyclers. These pieces of equipment add both complexity and accuracy to the research process. DOE can simultaneously sort through the processing conditions to discover the combinations that hold the most potential. The application of split-plot designs becomes necessary when complete randomization is not practical. Split-plot analysis provides improved testing for significant effects, ultimately leading to more accurate conclusions.

REFERENCES:

1. Powledge, Tabitha M., “The Polymerase Chain Reaction”, Federation of American Societies for Experimental Biology. http://www.faseb.org/opa/break.
2. Bloom, Mark V., “Polymerase Chain Reaction”, DNA Learning Center. http://www.accessexcellence.org/RC/CT/polymerase_chain_reaction.html.
3. Montgomery, Douglas C., “Design and Analysis of Experiments”, John Wiley & Sons, 2001.
4. Design-Expert® software. Stat-Ease, Inc., http://www.statease.com.

PRESENTATION EXPERIENCE:

- 2002, 51st International Wire & Cable Symposium, Lake Buena Vista, FL
- 2002, Fall Technical Conference, Valley Forge, VA.
- 2000, Fall Technical Conference, Minneapolis, MN