

Father and son	106
For some, total adoption	107

High-throughput screening: designer screens

Nathan Blow

Some researchers say an eighty-year-old statistical method can make setting up and analyzing high-throughput screens and large-scale experiments faster and more efficient. So why are more biologists not flocking to use this tool?

In his 1935 book, *The Design of Experiments*¹, British mathematician Ronald A. Fisher developed a mathematical framework for designing experiments. Fisher explored how experiments could be most efficiently set up to survey the interactions among different experimental factors in an effort to identify optimal combinations. His approach, along with several different mathematical underpinnings that have evolved since, has come to be known as design of experiments (DoE).

Since Fisher's book¹ was published, chemists, engineers and social scientists have relied on DoE approaches and software packages for setting up research models and complex experimental designs in fields ranging from clinical trials to petrochemical manufacturing. But surprisingly, in biology, where today's high-throughput screens often use multiple conditions, variables and reagents, the story has been quite different. "I am not sure why more people are not using DoE in biology," says Seth Cohen, director of Microfluidic Applications at Caliper Life Sciences in Hopkinton, Massachusetts, USA.

At Caliper, Cohen and his team swear by the use of DoE format when it comes to finding optimal biochemical assay conditions, using it 100% of the time now to design reagents and optimize enzyme performance. They have developed an integrated suite of tools for approaching DoE methodology including a commercially available DoE software package they adapted for their enzyme assay development efforts, Caliper's Sciclone ALH3000 automated liquid handling platform to set up the DoE software-generated experiment and the LabChip EZ Reader to analyze the results.



Adam Hill, who has been using DoE in his research for more than a decade, would like to see wider adoption of DoE in biology.

After the results are generated, the software analysis package can identify important experimental interactions between assay components as well as provide the optimal conditions for the assay. Using this pipeline, Cohen and his assay development group can explore up to 600,000 different combinations of conditions in a single DoE experiment: 192 different conditions such as salt concentration and pH, for example, with 250 unique combinations. He says that the amount of information that can be extracted from a single experiment can often hook researchers: "Once scientists use DoE, they don't turn back."

Waiting for the tipping point

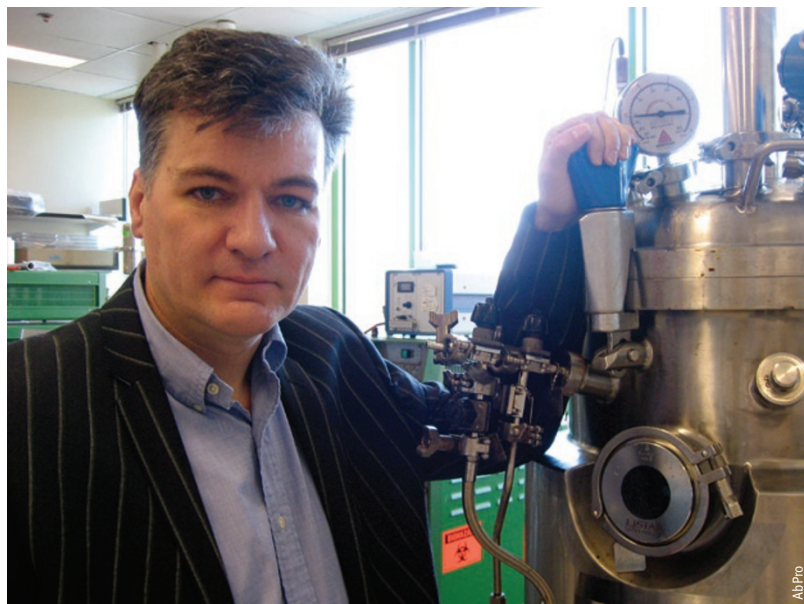
But getting to the point of using DoE methodology is the issue. And in biology, the

implementation of DoE is often met with resistance. "At the onset, lead balloons have more luck," says Adam Hill, director of the Hits Discovery group at the Novartis Institutes for Biomedical Research in Cambridge, Massachusetts, USA, who has been using DoE in his work for the past ten years.

It was Hill who championed the DoE approach in assay development for high-throughput screening at Novartis when he first joined the company in 2004. "They were basically not doing it when I arrived," he says. But Hill, whose group develops cell-based and biochemical screening assays using technologies that include liquid-handling robots and microplate scanners, wanted to speed up the screening process, and he knew from previous work that DoE was a powerful technique that could enable better use of the available equipment for rapid assay development.

The quantitative nature of most biochemical assays, like the ones Hill develops at Novartis, are particularly well suited to DoE; in fact, many researchers say a quantitative output is almost a prerequisite to setting up an effective experiment using DoE. Hill says that many cell-based assays are not particularly amenable to a fractional factorial DoE approach, as researchers like to consider every condition and the controls tend to be specific to each condition. When it comes to using DoE, having an optimal number or value to attain makes building models and testing for experimental interactions within a dataset easier.

But Hill and others suspect that the quantitative nature of DoE, along with its statistical underpinnings, could be one of the reasons for its limited use in biology. "For



Stephen Chambers says protein expression studies can often be done faster using DoE approaches.

biologists it can be intimidating; biologists and math usually don't mix," says Stephen Chambers, vice president of the Cambridge-based Abpro, a protein reagent company that has adopted DoE approaches. Still, Chambers notes that there are now many good software options available for those interested in applying DoE in setting up their high-throughput screens or assay-optimization problems.

For designing experiments, several commercial packages exist for researchers to try. SAS, a software development company in Cary, North Carolina, USA, now offers JMP 8, a statistical package with programs that allow users to design their experiments with several different variables as well as offering a diagnostic module to improve the data analysis. Stat-Ease, located in Minneapolis, is also devoted to furthering the use of DoE approaches through software development with their Design-Ease program for designing screening experiments and Design-Expert software for optimization of experiments based on response-surface methods, as well as providing training in the use of DoE methodology to interested researchers.

Kinnelon, New Jersey, USA-based Umetrics is another software developer whose DoE software is being used for both initial screen designs and follow-up optimization experiments. "It is a two-step process: first you are doing more traditional DoEs, which are used to find the most important variable and identify nonlinear interactions,

but then in the second pass you can use [those] data to train a system to stay within an optimal range," says Chris McCready, director of Global Process Analytical Technology at Umetrics. Currently Umetrics offers MODDE 8.0 for DoE and SIMCA-P for data analysis. McCready says a researcher inputs the number of experimental variables and what type of model they expect, either a response surface or screening design similar to the Stat-Ease packages, and the MODDE 8.0 program will tell the researcher what types of runs to make, whereas SIMCA-P can perform data analysis in cases when many variables are present.

Although DoE software packages are proving effective for setting up experiments, the other issue novice researchers tend to encounter when first working with DoE is integration. "If someone came out with a robust software package that allowed you to design an experiment, feed that design to an automated liquid-handling instrument and analyze the results from your assay, that would be a good start towards more widespread use," says Hill.

The ability to integrate DoE software with automation could be one reason that up to this point it has really been pharmaceutical and biotechnology companies who adopted DoE and applied it to biology. "You do not have to use automation, I started off doing it by hand, but in most cases people would like to use it," says Hill.

But with more and more liquid-handling platforms being developed at lower price

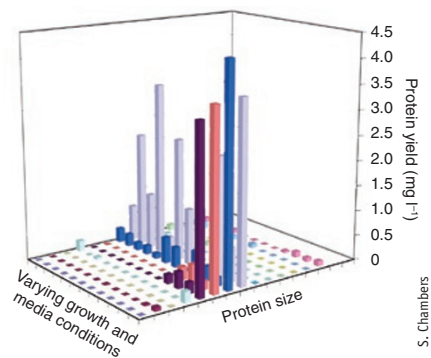
points, along with an ever increasing number of researchers interested in using high-throughput screens in their research, one has to wonder whether more use of DoE or even variants of DoE could migrate into the academic world.

Father and son

When Charlie Carter, a biochemist at the University of North Carolina in Chapel Hill, started as an assistant professor in the late 1970s, it was his father who introduced him to the idea of using DoE approaches to grow crystals for his crystallography studies. "We used a randomized balanced factorial design, which had previously been through the ringer," says Carter.

Randomized balanced factorial design was initially used by Carter's father for his work in the field of quality control. Carter's father and his colleagues tried a modified fractional factorial design to identify engineering malfunctions, but instead of analyzing the specific fractions individually they went back to the advice of Fisher who said that sampling should be random and balanced across the entire experimental space to best search for interactions and effects. Carter thought this approach could provide the rigor he was searching to explore large experimental spaces when it came to setting up crystal growth conditions. In their first application, 34 experiments produced a crop of different crystals that Carter found map out the entire catalytic cycle of the enzyme tryptophanyl-tRNA synthetase².

Carter's work turned out to be a precursor to the work of Sung-Hou Kim, who instead of randomized balanced factorial design, used a sparse matrix sampling technique to identify sets of initial conditions to be used for protein crystalliza-



Three-dimensional plots can be used to illustrate interacting conditions that produce the optimal protein expression when using DoE.

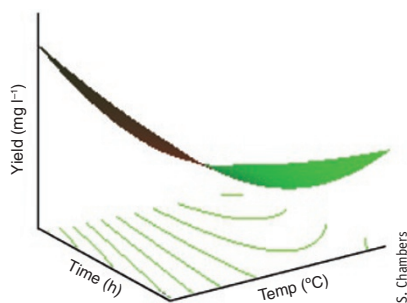
tion trials³. Kim's group initially set up 50 different crystallization conditions based on conditions known to generate crystals. Although it was not a true DoE according to Carter because the conditions did not follow Fisher's random and balanced models, and "it would be very difficult to extract the information that the ensemble of experiments provides," the simplicity and effectiveness of Kim's sparse matrix sampling approach has taken hold for the design and testing of crystallization conditions in structural biology.

Several companies now offer crystallization kits based on the sparse matrix sampling. Jena Biosciences in Jena, Germany, offers crystallization screening kits that cover 240 different conditions, along with specific subset screening conditions for specific protein classes such as kinases, phosphatases and membrane proteins. Other companies such as Emerald BioSystems of Bainbridge Island, Washington, USA, Hampton Research in Aliso Viejo, California USA and Qiagen in Valencia, California, USA also offer crystallization kits for researchers.

Carter is continuing to use DoE for both crystallization work and a new research project in his lab: high-throughput mutagenesis screens of proteins to explore patterns of energetic coupling between side chains. Although in the early stages now, Carter's exploration of protein mutants can be performed in a high-throughput manner that provides quantitative answers just like biochemical assays. "I think this approach opens new frontiers to accumulate a bigger picture about how proteins work," he says.

For some, total adoption

Back at Caliper, Cohen says that by using the DoE approach, the development and optimization of their enzyme assays has gone



The use of DoE allows models to be generated that can be used to describe and predict protein expression.

from taking months on average to a single day in most cases.

"And we are able to find the best conditions, regardless of the individual performing the assay," says Cohen. This is another benefit that many researchers bring up when talking about using DoE; the results can be standardized and presented in a language that can cut across disciplines.

"At meetings we can talk about the big experiment and then show results quantitatively and the significance of those results in quantitative terms," notes AbPro's Chambers. For Chambers, who works on high-throughput protein expression, having hard numbers for a large screen or an optimization experiment can, in some cases, make for easier discussions with his colleagues than showing denaturing polyacrylamide gels of all the results.

At AbPro, where they can express thousands of proteins with their high-throughput methodology, the use of the DoE methodology to guide their protein expression work was the result of a perfect storm. "What is happening in biology is that people do not want to do the single experiment anymore looking at one protein; they

want to look at one protein in context with other proteins," says Chambers. Whether it is screening or enzyme optimization assays, he says the experiments have become much bigger and more quantitative, the conditions that bring DoE into play. So AbPro, like Caliper with their enzyme optimizations, has used advances in automation and microfluidics, along with DoE, to move their protein expression efforts into a large-scale, industrial operation, an important development according to Chambers as more and more researchers need large numbers of purified proteins to advance technologies like protein chips.

In the end, the question of wider adoption still remains. Although DoE is making headway in several fields, including crystallization, protein expression, biochemical assay development, mass spectrometry and metabolomics, even those researchers devoted to its use question just how widely and fully the technique can invade biology. "Without improvements in education of scientists, ease of implementation and interpretation of DoE data, I think it will probably permeate to the same level as it has at Novartis," says Hill. Although he adds that in the end that is only around 20%, even that proportion would bring DoE to the attention of a much wider range of researchers.

1. Fisher, R.A. *The Design of Experiments* (Oliver and Boyd, London, 1935).
2. Carter, C.W. Jr. & Carter, C. W. Protein crystallization using incomplete factorial experiments. *J. Biol. Chem.* **254**, 12219–12223 (1979).
3. Jancarik, J & Kim, S.H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24**, 409–411 (1991).

Nathan Blow is the technology editor for *Nature* and *Nature Methods* (n.blow@boston.nature.com).

TECHNOLOGY FEATURE

SUPPLIERS GUIDE: COMPANIES OFFERING AUTOMATED LABORATORY SYSTEMS AND APPROACHES

Company	Web address
Abbott Molecular	http://www.abbottmolecular.com/
Abpro	http://www.abpro-labs.com/
Agilent	http://www.agilent.com/
Applied Biosystems	http://www.appliedbiosystems.com/
Aurora Biomed	http://www.aurorabiomed.com/
BD Biosciences	http://www.bdbiosciences.com/
Biometra	http://www.biometra.de/
Bio-Rad Laboratories	http://www.bio-rad.com/
BioTec	http://www.biotec.co.jp/
BioTek	http://www.biotek.com/
Bruker Daltonics	http://www.bdal.com/
Caliper Life Sciences	http://www.caliperls.com/
Corbett Life Science	http://www.corbettlifescience.com/
CyBio	http://www.cybio-ag.com/
Deerac Fluidics	http://www.deerac.com/
Emerald Biosystems	http://www.emeraldbiosystems.com/
Eppendorf	http://www.eppendorf.com/
Genetix	http://www.genetix.com/
Genomic Solutions	http://www.genomicsolutions.com/
Gilson	http://www.gilson.com/
GNF Systems	http://www.gnfsystems.com/
Hampton Research	http://www.hamptonresearch.com/
Hamilton Robotics	http://www.hamiltoncomp.com/
Hudson Control Group	http://www.hudsoncontrol.com/
Jena Bioscience	http://www.jenabioscience.com/
Kalypsys Systems	http://www.kalypsys-systems.com/
Komax	http://www.komaxgroup.com/
Labcyte	http://www.labcyte.com/
Millipore	http://www.millipore.com/
Molecular Devices	http://www.moleculardevices.com/
Nikon Instruments	http://www.nikoninstruments.com/
Perkin Elmer	http://las.perkinelmer.com/
Promega	http://www.promega.com/
Protodyne	http://www.protodyne.com/
Qiagen	http://www.qiagen.com/
Rigaku	http://www.rigaku.com/
Roche Applied Science	http://www.roche-applied-science.com/
SAS	http://www.sas.com/
Stat-Ease	http://www.statease.com/
Tecan Group, Ltd.	http://www.tecan.com/
Thermo Electron	http://www.thermo.com/
Titertek	http://www.titertek.com/
Torcon Instruments	http://www.torconinstruments.com/
Umetrics	http://www.umetrics.com/
Velocity 11	http://www.velocity11.com/
Xceed Molecular	http://www.xceedmolecular.com/
Xiril	http://www.xiril.com/
Zinsser Analytic	http://www.zinsser-analytic.com/