Experimental design for high-throughput screening

Michael W. Lutz, J. Alan Menius, Tony D. Choi, Rebecca Gooding Laskody, Paul L. Domanico, Aaron S. Goetz and David L. Saussy

Novel methods in molecular biology and advanced technologies have given pharmaceutical research laboratories the capability to test combinatorial libraries rapidly against large numbers of potential targets. Methods to identify optimal assay conditions efficiently are very useful in the development of robotic screening assays where there are numerous variables and potential interactions between the variables. This review discusses the use of statistical experimental design in the development and optimization of high-throughput screening assays. The authors provide a brief introduction to the theoretical basis for experimental design and discuss practical aspects of using these methods in a research laboratory. Two case studies demonstrate the power of the method in solving problems in assay development and illustrate the diversity of potential applications.

Recent scientific and technological advances have introduced new paradigms for drug discovery research. The availability of chemical libraries and robotic systems for bioassay allows synthesis and testing of hundreds or even thousands of compounds in one day. These developments present great challenges and opportunities for assay development and automation. As advances in molecular biology and bioinformatics identify more potential biochemical targets, and combinatorial libraries provide a large number of compounds for testing, it is critical that methods for the efficient development and optimization of assays are used. Specifically, the time to identify optimal assay conditions must be shortened and the use of scarce proteins and reagents must often be minimized in high-throughput screening laboratories. Assay variability must be minimized in order that experimental results can be used for SAR analysis and for the design of subsequent combinatorial libraries.

Experimental design is a classical statistical method that has found utility in many disciplines that require the finding of optimal conditions and modeling of response surfaces. The theoretical basis for experimental design is described in detail in numerous statistical texts[1–4]. Application case studies can be found in the literature associated with a given field. Examples include industrial process optimization[5,6] and chemistry[7,8]. A recent book by Haaland, *Experimental Design in Biotechnology*[9], is an excellent introduction to some of the problems and methodologies associated with using experimental design for biological and chemical applications. Experimental design nomenclature used in this review is explained in Box 1.

In all applications, the underlying issue is how to optimize a process in a systematic way in order to extract the most information in the least number of runs. In many cases, identification of the optimal conditions is sufficient. In other cases, the relationship between the factors and the response is modeled mathematically; this is termed response surface modeling. Interactions are particularly important because they are more

Michael W. Lutz*, J. Alan Menius, Tony D. Choi, Rebecca Gooding Laskody, Paul L. Domanico, Aaron S. Goetz and **David L. Saussy**, GlaxoWellcome Research Institute, Research Triangle Park, NC 27709, USA. *tel: +1 919 483 6106, fax: +1 919 483 2494, e-mail: Michael_Lutz@glaxo.com

  1359-6446/96/$15.00. PII: S1359-6446(96)10025-8

Box 1. Experimental design nomenclature used in this review

Factor: an experimental variable that can be changed by the investigator (e.g. temperature, time in incubator, pH or choice of buffer)
Levels: specific settings of factors for a given experiment
Interactions: the interplay of combinations of factors (e.g. an interaction between pH and temperature)
Response: the variable or quantity being optimized (e.g. yield in a chemical reaction or variability of a measurement in an assay)
Effect: the magnitude of the change in response as factors are varied
Run: an experiment performed at specific settings of all the factors
Main effects: relates to the parameters associated with single factors, an estimate of their standard deviation or significance in the model

the rule than the exception in biology and can greatly increase the complexity of a problem or, in the worst case, make interpretation of the data impossible. Variance-reduction experimental designs constitute a special subset of experimental designs. In these designs, the variance of a response is optimized rather than the response itself. Such designs are important in biology because by minimizing variance the assay becomes more robust. Cost is often an important factor, and optimization is desired to minimize use of scarce resources, such as proteins. In many cases, multiple responses are to be considered: for example, minimizing variance while keeping receptor concentration low in order to conserve protein. Finally, the advantage of having a model, even a simple one such as the linear model that forms the basis of classical experimental designs, is that you can make predictions.

In this review, space-filling designs, a recent addition to the array of statistical experimental designs, are discussed in the light of their potential application to intrinsically nonlinear problems in the optimization of complex processes, such as biochemical assays.

Although not discussed here, experimental design has utility in the design of combinatorial libraries[10,11]. For such applications, the factors are either substituents or collections of monomers attached at a particular site on a molecule. The substituents or monomers must be characterized by some set of properties – electronic, steric or topological. The design process selects a set of compounds for synthesis that will allow interpretation of the chemical properties' space using a linear model. From this model, specific questions can be answered about the effect of a set of substituents at a specific site and interactions between sites. Young and Hawkins[12] published an analysis of a full factorial combinatorial peptide library designed to study stereochemical binding properties of NK_1 receptors[13]. Their work illustrates the use of analytic and visualization tools in the analysis of data from combinatorial libraries.

This review illustrates the use of statistical experimental design for the development and optimization of assays designed for high-throughput screening. Two examples are discussed which demonstrate the diversity of problems in assay development that experimental design can address:

- Case study 1: Optimization of assay conditions for a scintillation proximity assay (SPA).
- Case study 2: Choice of a sampling and data analysis protocol for a radioligand binding experiment.

The first example is concerned with optimization of an assay and involves identification of relevant factors, choosing an appropriate design and collecting data. The second example is a computer simulation study using retrospective data. Taken together, they illustrate the power of experimental design to enable the rapid solution of problems in biological assay development.

Classical experimental designs

In classical experimental design, investigators choose a number of factors that they believe may have an effect on the response. Depending on the number of factors and their possible ranges, different types of designs are chosen to sample the space defined by these factors. All of the factors are varied simultaneously in a systematic way, which later allows the significance of a specific factor or interaction between factors to be measured. Typically, the factors are sampled at the extremes of the operating range and a linear or polynomial interpolating function is used to model the region between the design points.

If estimates of the coefficients associated with each term in the model (main effects and interactions; see Box 1) are independent, the design is said to be *orthogonal*. The mathematical form of the linear model, in matrix terms, is as follows:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \quad (1)$$

where $\mathbf{X}$ is the design matrix (independent variables or predictors), β is a vector of parameter estimates obtained by linear regression, $\mathbf{Y}$ is the vector of observed responses and ε is

a vector of error terms, assumed to be normally distributed with a mean of zero and a standard deviation of σ.

Full factorial design

Full factorial designs are conceptually the easiest experimental designs to set up and to interpret. Data are obtained for all combinations of factors, allowing the effect of each factor and the interaction between factors to be studied. Often, two levels are chosen for each factor, representing the upper and lower limits of the operating range, although more than two levels can be chosen to better understand the relationship between a factor and the response. Graphically, simple full factorial designs can be described by geometric shapes: for example, a square for a two-factor, two-level design (Figure 1) and a cube for a three-factor, two-level design (Figure 2). If each factor is sampled at two levels, 2^k runs will be needed for k factors in a full factorial design.

Although is it clear that, if there are many factors and/or levels, these designs will require a large number of runs, we have found them to be very useful in practice, particularly for assay development. Interpretation of the design can usually be performed graphically, often simply by recording the response at the vertex of each design point (see Figures 1 and 2). For factor levels that are under software control in a robotic assay system, large numbers of runs are easily set up and an assay optimized in a short time.

Fractional factorial design

A fractional factorial design is a compromise between the information on all interactions that can be obtained in a full factorial design and the total number of runs. Fractional factorial designs can be grouped by their resolution, the types of effects and interactions that can be estimated by a class of designs. Higher resolution designs allow more interactions and effects to be studied than lower resolution designs but require more runs. The term 'confounding' is used to refer to effects (factors and/or interactions) that cannot be estimated independently.

- Resolution V designs allow all main effects and two-factor interactions to be estimated, sacrificing only the ability to estimate higher order interactions.
- Resolution IV designs do not allow any two-factor interaction to be confounded with a main effect, but allow confounding among the two-factor interactions.
- Resolution III designs allow main effects to be confounded with interactions but do not allow confounding among main effects.

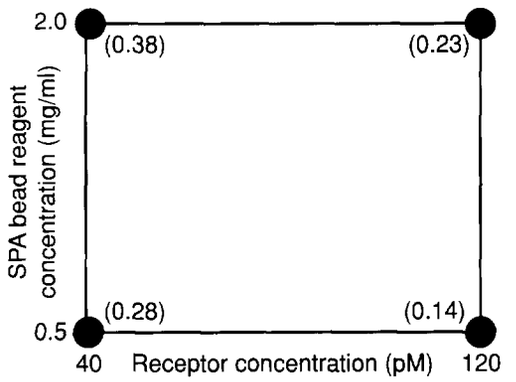

Figure 1. *Two-factor, two-level full factorial experimental design. The two factors and their levels are concentration of receptor (40 pM, 120 pM) and SPA (scintillation proximity assay) bead reagent concentration (0.5 mg/ml, 2.0 mg/ml). The response values, variability of the pK_i estimate in standard deviation units, are given in parentheses.*

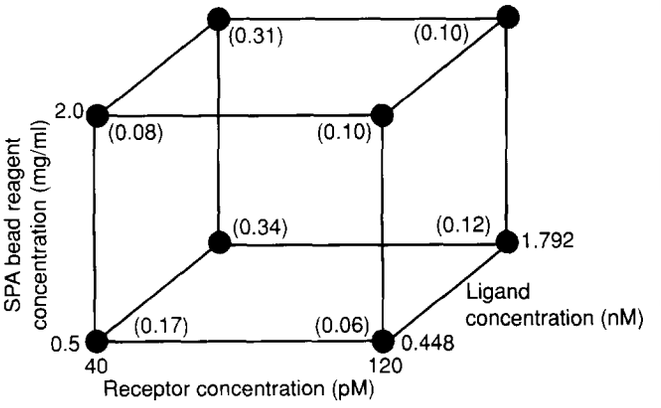

Figure 2. *Three-factor, two-level full factorial experimental design. The three factors and their levels are concentration of receptor (40 pM, 120 pM), SPA (scintillation proximity assay) bead reagent concentration (0.5 mg/ml, 2.0 mg/ml) and concentration of ligand (0.448 nM, 1.792 nM). The response values, variability of the pK_i estimate in standard deviation units, are given in parentheses.*

A variation on the fractional factorial designs are Plackett–Burman[14] designs. These are low-resolution designs that require even fewer runs than fractional factorial designs. They are most useful for screening large numbers of factors rather than testing for interactions.

The simplest way to set up a fractional factorial design is to use experimental design software or to consult a listing of designs. The design digest in Haaland's book[9] provides a list of the most common designs encountered in practice.

Response surface modeling

Response surface modeling involves the use of a regression equation to describe mathematically the relationship between factors and response. Often, graphs such as contour or surface plots are used to describe such relationships. Response surface modeling can be based on both the full and the fractional designs already described, but additional data must be collected in order to improve definition of the curvature. First, a center point is established at the centroid of all of the design points. Points are then added to cover regions away from the design points. These are commonly called star points (Figure 3). These designs are known as central composite designs. In the case of linear models, terms for quadratic effects are included for interpretation of results.

Box–Behnken designs[15] are another class of designs used for response surface modeling. These designs allow estimation of all main effects, two-factor interactions and quadratic effects. Box–Behnken designs are useful when a set of factor settings can be considered to be the standard operating conditions, perhaps after some factorial screening designs, and allow a more detailed analysis of the region spanning out from the central design point.

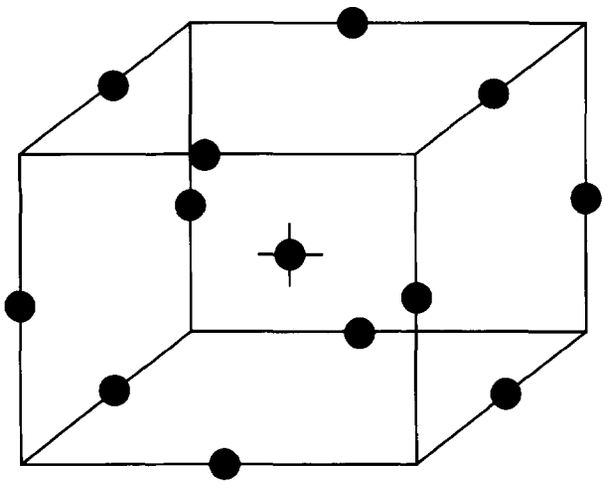

Figure 3. *Star points for a central composite design. These points augment the corner points of factorial designs to give a better estimate of the curvature of response and are used for response surface modeling.*

D-optimal designs

D-optimal designs[16–18] are an example of a computer-generated design. Many experimental design software programs, such as JMP® (SAS Institute Inc., Cary, NC) and DISCOVER (BBN Software Products, Boston, MA), include D-optimal designs as an option. In using a D-optimal design or any other similar computer-generated design, the user chooses an acceptable number of runs, identifies the main effects and interactions of interest and allows the computer to generate the design. In a D-optimal design, this is done iteratively by maximizing the determinant $|(\mathbf{X}'\mathbf{X})^{-1}|$. The determinant can always be maximized by increasing the number of points or by expanding the range of values for the predictors. These methods will better define a linear relationship between a set of predictors and a response. In a D-optimal design, however, the number of runs and ranges for the predictors is fixed. Computationally, the algorithm chooses design points that maximize the determinant and minimize the region that bounds the parameters of the linear model used to interpret the design by iteratively adding and deleting points from a set of candidate design points. The main disadvantage of such designs is that they assume that the form of the mathematical model is known *a priori*. D-optimal designs are not necessarily orthogonal. If there are interactions present that are not included in the design, the other effects may not be estimated properly.

Computer-generated designs are useful in the biochemical sciences since the software often allows substitution for points that are inaccessible because of physical constraints. For example, it may not be possible to run an experiment at the factor settings for a particular design point in a full factorial design. A computer-generated design, however, can substitute a different point and still choose a design that maximizes the D-optimal criteria, giving the best design for that series.

Some practical considerations

Identification of factors and levels

Identifying critical factors and associated levels is very important and an area in which many designs fail. Often, factor levels are chosen that either fail to cover a suitable range or cover such a broad range that they are not adequately modeled by a linear model. Factors and settings are usually determined from prior experience and knowledge of the assay system. Some practitioners argue that it is worth running a series of undesigned experiments to define suitable ranges for the factors, whereas others prefer to start with a low-resolution screening design. Ideally, the settings are chosen

such that the investigator expects to see a difference in response over the range of factor settings.

Different experimental designs are chosen for specific objectives. When a robotic assay is first established, there are often many factors that can potentially influence the robustness of the assay. To find the important factors, a low-resolution fractional factorial design may be run; this is known as factor screening. After the important factors are identified, a full factorial design might be initiated to find the optimal settings for the factors. Response surface modeling could be performed in order to explore the region around the optimal settings, to assess the sensitivity of the factors and to test for higher order interactions.

Assay variability and replication

It is important to have a good estimate of assay variability, both to interpret the results of the experimental design and, later, to be able to identify significant differences between compounds tested in the assay. When assay variability itself is the response being optimized, each design point must be replicated. For other designs, a replicated center point is an efficient means of estimating assay variability. The center point is obtained from a run in which all factors are set in the middle of their range. This point is replicated several times (typically 3–10) in order to provide an estimate of the variance of the response. It is important to recognize that the replicates must be independent. For example, observations obtained from adjacent wells on a 96-well microtiter plate are probably not independent and will measure only the error in pipetting. Although the center point is a reasonable initial estimate of variability, it is also possible that the factors themselves may have an effect on variability requiring replication at other design points.

As with most statistical analysis, the interpretation of an experimental design using a linear model assumes that the error in the response is random variation. All systematic effects are assumed to be part of the model. The investigator must guard against the introduction of inadvertent systematic effects by examining the experimental protocol for potential sources, such as time effects, lack of homogeneous reaction conditions and differences between reagents and biological materials. In many cases, if a systematic effect is suspected, it is possible to modify the design and add terms to the model to account for the effect – a technique known as 'blocking'. Most standard experimental design software allows for blocking effects. In practice, they are terms that are allowed to be confounded with high order interactions.

Analysis and interpretation of experimental designs

Analysis and interpretation of experimental designs can be as simple as sorting a table of results to find the conditions that give the lowest assay variability, or as complex as fitting a model to the data and using interactive graphics to study the effect of interactions.

For designs that have few factors, simple plots that resemble the shapes of the design (squares and cubes) are useful to identify general trends. These plots can easily be drawn with pencil and paper. The response value is simply filled in at the design points. Tabulating the data in a statistical analysis or spreadsheet program allows sorting by different criteria in order to find trends in the data. Sorting by size of effect is often helpful for two-level factors. The average response at each factor setting is calculated and subtracted to yield size of effect.

Because a statistical model is used to interpret an experimental design, all of the tools appropriate for the evaluation of the model apply, including standard deviations and significance tests on the parameters, residual plots, normal probability plots and analysis of variance tables[19]. It is particularly important to compare the variance of estimated effects to assay variability.

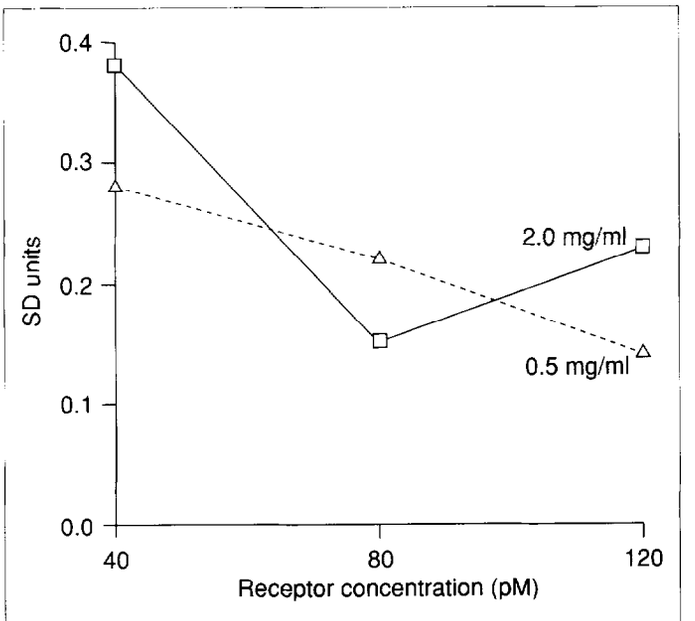

Figure 4. *Example of an interaction plot for a two-factor interaction. The ordinate is the response value: variability of the pK_i expressed in standard deviation (SD) units. The abscissa represents factor settings for one of the factors (receptor concentration); the second factor, SPA bead concentration, delineates the curves.*

Although the statistical analysis will calculate a significance test for each interaction term, it is often helpful to make interaction plots for two-factor interactions as well. Figure 4 is an example of an interaction plot. The ordinate is the response value, and the abscissa is the factor settings for one of the two factors. The second factor is used to group points on the plot; all points that share a factor setting are connected by a line. If there is no interaction, the lines are parallel. If the lines are not parallel, the viewer can determine whether the interaction is working to their benefit (e.g. lowering variability while keeping the first factor at a preferred level), or whether it is a detrimental interaction.

JMP® has visualization capabilities that allow investigators to interpret an experimental design interactively. A graph pane is created for each effect in the model where the abscissa reflects the factor settings and the ordinate is the response. The vertical slider, originally set at the mean for each factor, can be moved to any factor setting. As the slider is moved, all panes update to show the effect on response. It is important to realize that the effect on response is calculated from the full model of the data including interactions. The error bars allow the viewer to identify significant differences between response at different factor levels and differences from the mean response (horizontal line). JMP® also allows contour plots to be easily constructed from experimental design results. Depending on the number of factor levels used, contour plots can be helpful in the identification of optimal operating regions. However, their usefulness is limited by their bivariate nature.

Space-filling experimental designs

Space-filling designs are a new area of interest to statisticians. When experimental designs fail, it is often because the response surface is irregular or contains many spikes and valleys that are difficult to model with linear or polynomial functions. In a space-filling design, the factor settings are chosen to distribute the design points evenly throughout the region or space of possible values, rather than to sample the *extrema*. Such designs may be more particularly appropriate for use in biological and chemical experimental applications than the classical designs which work well for optimizing reaction yield or an industrial process. The designs tend to require many data points; unlike classical designs, in which factors are often sampled at two levels, each factor may be sampled at 10–20 levels. For high-throughput screening and robotic assay systems, it is possible to achieve the sampling rates required by these designs. Although algorithms to create the designs have existed for some time[20], good software has only recently become available. Programs include ALEX (Refs 21,22) and SAS/QC (Ref. 23). Cluster analysis[24] can also be used to create the designs.

Interpretation of results from a space-filling design requires different methods from those used for traditional designs. In general, a nonparametric regression model is necessary for interpretation of the data because of the possibility of curvature in the design space. Examples of methods used to analyze the results of space-filling designs include polynomial regression models, Gaussian stochastic processes, thin-plate splines and neural networks[25]. The simplest models are the polynomial models that can be fit by ordinary linear regression. Neural network models can be fit using a variety of commercial software packages. For Gaussian stochastic processes and thin-plate splines, the model has a component in addition to the linear term of Equation 1:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z} + \varepsilon \qquad (2)$$

The $\mathbf{Z}$ term has a distribution that is normal with a mean of zero and a standard deviation of $\sigma^2 R$, where R is the correlation function. The key point is that a flexible function is used rather than a strict linear model. In some cases, only the intercept term of the linear model is used and all of the flexibility is put into the $\mathbf{Z}$ term. The idea behind the correlation function is a simple concept – points located together in space should be more highly correlated, an assumption that is valid for many problems in biology and chemistry.

For space-filling designs, choice of an analysis method depends both on the experimental objective and on the nature of the surface itself. For screening problems, simple graphical methods are often sufficient. For response surface modeling, choice of an analysis method is based on the plausibility and stability of the parameter estimates and the model fit.

Some recent examples of space-filling designs in biology include a design to determine optimal conditions for protein construct storage conditions[20] and a study of buffer conditions on DNA strand displacement amplification[25].

Case study 1: optimization of a scintillation proximity assay

SPAs are often considered as an alternative to traditional radioligand binding assays because they are more amenable to high throughput and automation. It is not necessary to separate bound and free radioligand in an SPA. In this example, human α_{1B}-adrenoceptors were expressed in mammalian cells and the membranes were partially purified. SPA beads with a wheat germ agglutinin coating were used to bind the glycosylated

proteins of the membranes containing the expressed receptors. A competition binding assay was performed in which concentration–response curves were generated in response to 5-methylurapidil displacement of tritiated prazosin.

The objective for this experimental design was to minimize variation in the K_i (inhibitor binding dissociation constant) estimate. A secondary objective was to minimize the amount of receptor needed. Several responses were considered: variability of the K_i, signal to noise ratio (S/N), range of counts and adequacy of the nonlinear regression fit to the data. In this case, the multiple response variables were considered separately. Three factors were considered in the design:

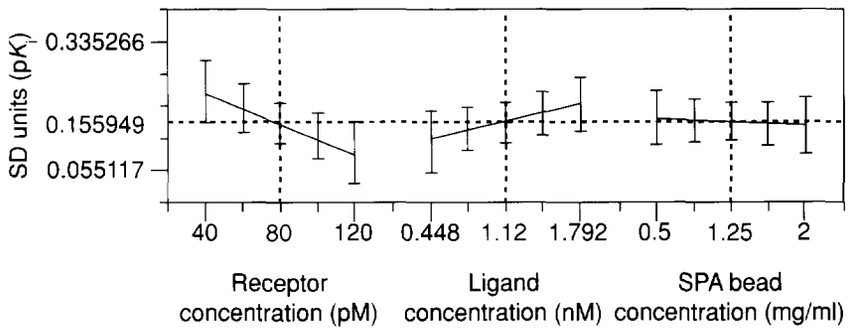

Figure 5. *Experimental design results for a full factorial optimization of an SPA (scintillation proximity assay). The ordinate is variability of the pK_i (expressed in standard deviation units) calculated from three replicate measurements. The abscissae are factor levels for each factor in the design: receptor concentration, ligand concentration and SPA bead reagent concentration.*

- receptor concentration (40 pM, 80 pM, 120 pM),
- SPA bead reagent concentration (0.5 mg/ml, 2.0 mg/ml) and
- ligand concentration (0.448 nM, 1.792 nM).

Because the number of factors and levels was small, it was possible to use a full factorial design of 12 runs ($2 \times 2 \times 3$). Since variability in the K_i was the primary response, the design was repeated three times in order to estimate the variance at each design point.

Variability

The results for variability are shown in Figure 5. Each panel is obtained using the JMP® statistical analysis/visualization software package. The ordinate in each case is variability of the negative log of the K_i (pK_i), the abscissae are the factor levels used in the design, and the vertical lines represent the values for particular factor settings that can be changed by the user. When the factors are changed, the plots are updated to show the effect on response predicted by the linear model, including interactions. The horizontal line shows the response obtained when all factors are set at their middle level. The error bars show the pooled standard deviation estimate for the model.

From the plots in Figure 5, it is seen that, for the factor levels chosen, there is a significant effect of receptor concentration (Figure 5, left panel). The highest receptor concentration minimizes variance. An inverse relationship is observed between ligand concentration and variability; the lowest ligand concentration has the smallest variance (Figure 5, center panel). SPA bead reagent concentration does not have a significant effect on variability in this experiment (Figure 5, right panel).

A contour plot (Figure 6) is an alternative way of viewing these results. This type of plot is helpful in visualizing operating regions, particularly if many factor levels are used, but it is limited by its bivariate nature.

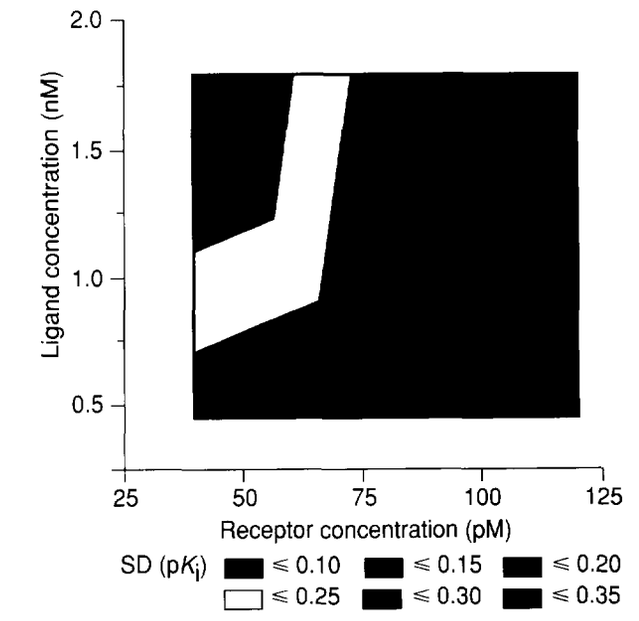

Figure 6. *Contour plot for experimental design results of the full factorial optimizaton of an SPA (scintillation proximity assay). The ordinate is ligand concentration. The abscissa is receptor concentration. Contours are variability of the pK_i (expressed in standard deviation units) calculated from three replicate measurements.*

Signal to noise

The optimal factor settings for S/N are slightly different from variability. The best S/N is obtained at the highest receptor concentration (Figure 7, left panel). Lowering the ligand concentration (Figure 7, center panel) and SPA bead reagent concentration (Figure 7, right panel) also improves S/N.

pK_i trend

Figure 8 and Table 1 show the trend of pK_i values for the different design points. It is seen that pK_i is not invariant with respect to receptor and ligand concentrations. Inspection of Table 1 shows that, while most combinations of factor levels have a variance that is typical of this type of assay, the two rows with the lowest receptor concentration have unusually high variance.

Case study 1: summary

The best conditions are low ligand and SPA bead reagent concentrations and high receptor concentration. These conditions minimize variance and provide the best S/N. If the amount of receptor were to be minimized, for example because of limited supply, it would be desirable to use the lowest ligand concentration level. For the best S/N, the lowest concentration of SPA bead reagent should be used. Although data are not shown for lack of fit, the higher SPA bead reagent concentrations appear to give the best fits. These conclusions address the trade-off between the different factors. After completing the design and constructing a model, it is possible to make decisions on a more quantitative basis. For example, if lack of fit does not seem to be a problem, it may be possible to lower the SPA bead reagent concentration in order to obtain a better signal.

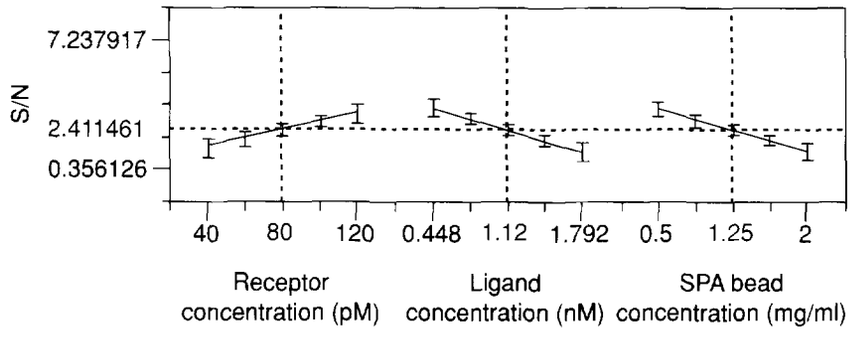

Figure 7. *Experimental design results for a full factorial optimization of an SPA (scintillation proximity assay). The ordinate is signal to noise (B_0 – NSB; B_0 is the concentration of radioligand bound with no inhibitor added and NSB is nonspecific binding). The abscissae are factor levels for each factor in the design: receptor concentration, ligand concentration and SPA bead reagent concentration.*

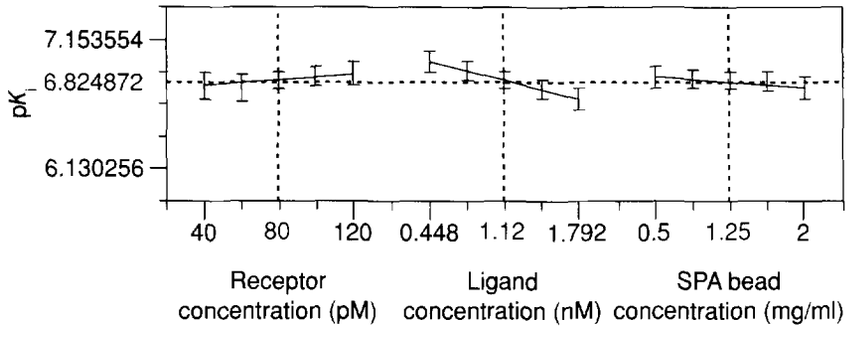

Figure 8. *Consistency of the pK_i estimates for a full factorial optimization of an SPA (scintillation proximity assay). The ordinate is pK_i. The abscissae are factor levels for each factor in the design: receptor concentration, ligand concentration and SPA bead reagent concentration.*

Case study 2: choice of a sampling and data analysis protocol for a radioligand binding assay

The choice of a sampling protocol for assays that involve nonlinear curve fitting, such a radioligand binding experiments, is important in ensuring the quality and reproducibility of the resulting K_i estimates or precision of interpolated concentrations if the experiment is a radioimmunoassay or SPA. The objective is to sample a sufficient number of concentrations over a suitable range to define the location and scaling parameters of the curve (IC_{50}, exponent if used, upper and lower asymptotes). Each point can also be replicated. Munson and Rodbard[27] suggest that, if the within-sample variance is smaller than that between samples (response at different concentrations), the mean value of replicates should be used; otherwise, all replicates should be used.

Table 1. pK_i trends for full factorial experimental design

Receptor (pM)	Ligand (nM)	SPA concentration (mg/ml)[a]	*n*	Mean (pK_i)	SD (pK_i)[b]
40	0.448	0.5	3	6.99	0.17
40	0.448	2.0	3	6.96	0.08
40	1.792	0.5	3	6.72	0.34
40	1.792	2.0	3	6.38	0.31
80	0.448	0.5	3	7.03	0.13
80	0.448	2.0	3	6.89	0.19
80	1.792	0.5	3	6.71	0.17
80	1.792	2.0	3	6.81	0.12
120	0.448	0.5	3	6.93	0.06
120	0.448	2.0	3	7.07	0.10
120	1.792	0.5	3	6.72	0.12
120	1.792	2.0	3	6.68	0.10

[a]SPA, scintillation proximity assay
[b]SD, standard deviation

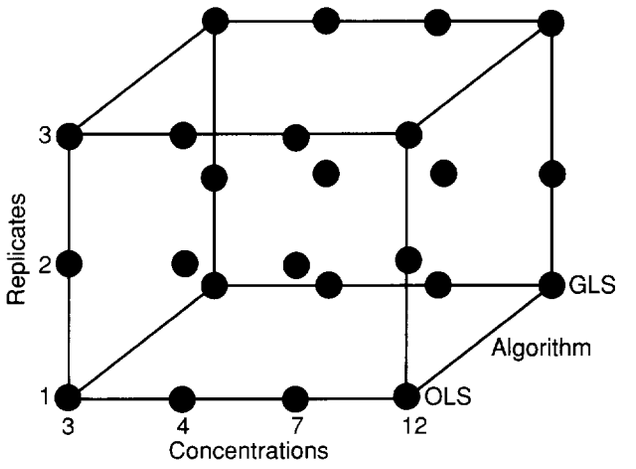

Figure 9. *Experimental design for choice of a sampling and data analysis protocol for a radioligand binding assay. GLS, generalized least squares; OLS, ordinary least squares.*

The experimental design used to address the question of optimal sampling and fitting of a concentration–response curve is shown in Figure 9. The analysis is different from traditional experimental designs in which experimental runs are performed to collect data after the design points have been selected. Because a large collection of concentration–response curves for a variety of different compounds was available, it was possible to perform a retrospective analysis by sampling the curves in a Monte Carlo simulation experiment.

The data were obtained from competition binding curves in a radioligand binding assay for three subtypes of α_1-adrenoceptors[28]. The receptors were human clones expressed in mammalian cells. In all cases, the radioligand was [^{125}I]HEAT. Approximately, 200 inhibitors were chosen at random covering a potency range of 6 log units. Each concentration–response curve was constructed by resampling the data for the number of concentrations and replicates at the design point[29–31]. This approach allowed us to pool results across different drugs.

The objective of this experiment was to simulate calculations from a calibration curve. Three factors were considered: number of concentrations, number of replicates and the choice of nonlinear regression algorithm. The nonlinear regression algorithms were either unweighted ordinary least squares or derived from a generalized least squares utilizing a power function to model the variance of the response[32,33]. Ordinate values were chosen at random within the range of values associated with a particular curve. These values were used to calculate concentrations and were compared to the true values (known from making the same calculation using all of the data available for the curve). The response used in the design was the standard deviation of the difference between the true and observed values for at least 100 simulations.

The results are presented in the cross-classification tables (Table 2). The cells are in standard deviation units. Differences of 5 are not statistically significant ($P > 0.05$). Differences of 10 are marginally significant. In much of the data, there is statistically little difference; for example, 12 doses in singlet give as consistent results as 12 doses in triplicate. Replication has less effect on reproducibility than the number of concentrations tested in this set of simulations. The results suggest that 12 concentrations in singlet or 7 concentrations in duplicate are optimal. Better performance was achieved using the generalized least squares algorithm rather than least squares for all of the combinations of replicates and number of doses, except in the cases where the curve is severely undersampled (lower right corner of each table).

Case study 2: summary

The results make intuitive sense. In curve fitting, points are necessary to define the asymptotes and region around the IC_{50}. Replication is less important if there are neighboring points to influence the fit. The superiority of the variance function model is also reasonable; the variance of the binding data

Table 2. Results of experimental design for sampling and data analysis protocol for a radioligand binding assay

Concentrations	OLS replicates[a]			GLS replicates[b]		
	3	2	1	3	2	1
12	44	45	45	28	30	31
7	49	55	59	29	33	37
4	67	67	–	37	45	–
3	273	275	–	383	379	–

[a]OLS, ordinary least squares
[b]GLS, generalized least squares

is noticeably a function of signal. The combination of a simple experimental design with a Monte Carlo simulation in this case illustrates an effective method for determining an optimal sampling scheme for the most consistent results while minimizing use of resources.

Conclusions

High-throughput screening of combinatorial libraries is a major trend in drug discovery. Statistical experimental design is a powerful tool, both for the development and optimization of the biological assay and for the design of the combinatorial library itself. This review has described both classical methods for experimental design and a more recent methodology, that of space-filling designs, which show promise for applications in biology and chemistry.

The case studies illustrate typical problems in the development of biological assays: namely, choosing optimal conditions (reagent concentrations, kinetics, volumes) and sampling protocols to provide maximal information for structure–activity determination. The availability of robotic systems for biological screening is well suited to experimental design: completion of runs under different assay conditions and randomization is often possible under software control.

Drug discovery in the 1990s is faced with explosions in the numbers of compounds to test and of potential biological targets. Experimental design is an efficient method both to optimize an assay and to understand the quality of the data it will provide.

REFERENCES

1 Cochran, W.G. and Cox, G.M. (1957) *Experimental Designs*, Wiley
2 Hicks, C.R. (1974) *Fundamental Concepts of Design of Experiments*, Holt, Rhinehart and Winston
3 Box, G.E., Hunter, W.G. and Hunter, S. (1978) *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*, Wiley
4 Montgomery, D.C. (1984) *Design and Analysis of Experiments*, Wiley
5 Daniel, C. (1976) *Applications of Statistics to Industrial Experimentation*, Wiley
6 Murphy, T.D. (1977) *Chem. Eng.* 84 (June 6), 168–182
7 Deming, S.N. and Morgan, S.L. (1987) *Experimental Design: A Chemometric Approach*, Elsevier
8 Austel, V. (1983) in *Quantitative Approaches to Drug Discovery* (Dearden, J.C., ed.), Elsevier
9 Haaland, P.D. (1989) *Experimental Design in Biotechnology*, Marcel Dekker
10 Gallop, M.A. *et al.* (1994) *J. Med. Chem.* 37, 1233–1251
11 Gordon, E.M. *et al.* (1994) *J. Med. Chem.* 37, 1385–1401
12 Young, S.S. and Hawkins, D.M. (1995) *J. Med. Chem.* 38, 2784–2788
13 Wang, J. *et al.* (1993) *Bioorg. Med. Chem. Lett.* 3, 451–456
14 Plackett, R.L. and Burman, J.P. (1946) *Biometrika* 33, 305–325
15 Box, G.E.P. and Behnken, D.W. (1960) *Technometrics* 2, 455–475
16 Kiefer, J. (1959) *J. R. Stat. Soc. Ser. B* 21, 272–319
17 Galil, Z. and Kiefer, J. (1980) *Technometrics* 21, 301–313
18 Snee, R.D. (1985) *J. Qual. Technol.* 17, 222–236
19 Snedecor, G.W. and Cochran, W.G. (1989) *Statistical Methods*, Iowa State University Press
20 Kennard, R.W. and Stone, L.A. (1969) *Technometrics* 11, 137–148
21 Sacks, J. *et al.* (1989) *Stat. Sci.* 4, 409–435
22 Welch, W.J. *et al.* (1992) *Technometrics* 34, 15–25
23 SAS/QC Software Reference (1989) Version 6, First edn, SAS Institute Inc.
24 Zemrock, P.J. (1986) *Technometrics* 28, 39–49
25 Haaland, P. *et al.* (1994) *Comput. Sci. Stat.* 26, 111–120
26 Menius, J.A. *et al.* (1994) *Comput. Sci. Stat.* 26, 106–110
27 Munson, P.J. and Rodbard, D. (1980) *Anal. Biochem.* 107, 220–239
28 Goetz, A.S. *et al.* (1994) *J. Pharmacol. Exp. Ther.* 271, 1228–1233
29 Efron, B. and Tibshirani, R.J. (1991) *Science* 253, 390–395
30 Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*, Chapman and Hall
31 Lutz, M.W. *et al.* (1995) *J. Pharmacol. Toxicol. Methods* 34, 37–46
32 Carroll, R.J. and Ruppert, D. (1982) *J. Am. Stat. Assoc.* 77, 878–882
33 Giltinan, D.M. and Ruppert, D. (1989) *J. Pharmacokin. Biopharm.* 17, 601–614