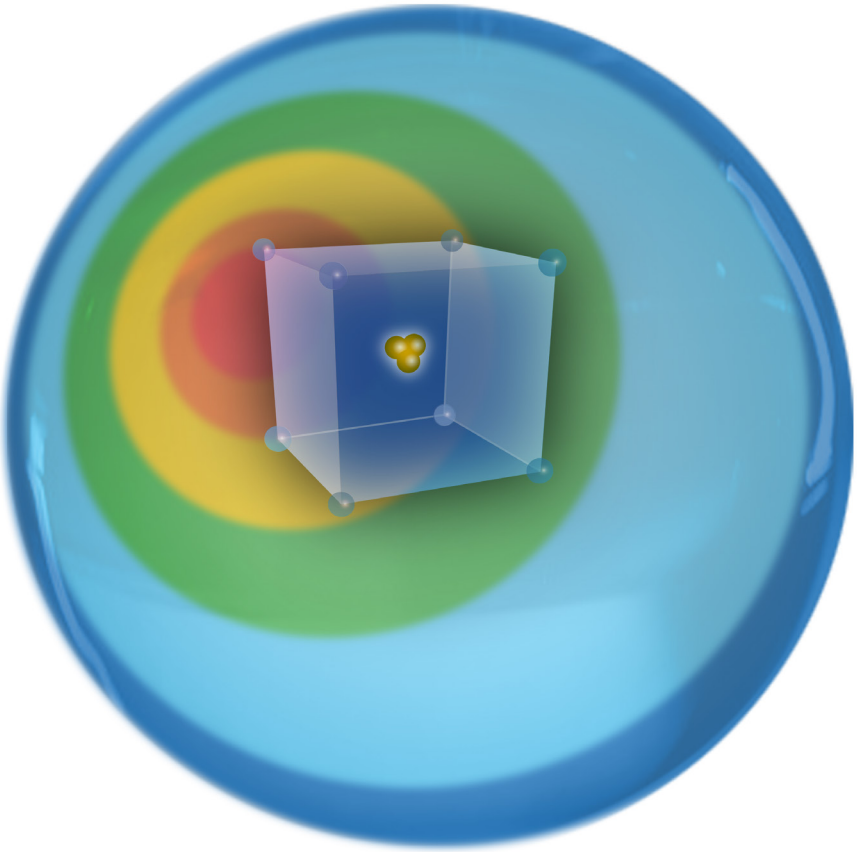
GE Healthcare
Life Sciences



Design of Experiments in Protein Production and Purification

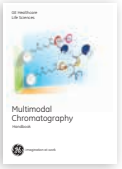
Handbook

Handbooks from GE Healthcare Life Sciences

For more information refer to www.gelifesciences.com/handbooks

2-D Electrophoresis using Immobilized pH gradients
Principles and Methods
80-6429-60

GST Gene Fusion System
Handbook
18-1157-58



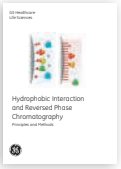
Multimodal Chromatography
Handbook
29-0548-08

Affinity Chromatography
Principles and Methods
18-1022-29

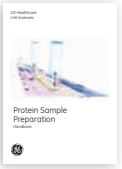
High-throughput Process Development with PreDictor Plates
Principles and Methods
28-9403-58

Nucleic Acid Sample Preparation for Downstream Analyses
Principles and Methods
28-9624-00

ÄKTA Laboratory-scale Chromatography Systems
Instrument Management Handbook
29-0108-31

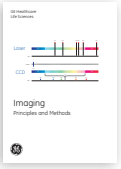


Hydrophobic Interaction and Reversed Phase Chromatography
Principles and Methods
11-0012-69

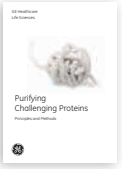


Protein Sample Preparation
Handbook
28-9887-41

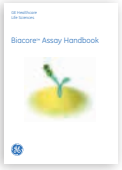
Antibody Purification
Handbook
18-1037-46



Imaging
Principles and Methods
29-0203-01



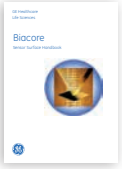
Purifying Challenging Proteins
Principles and Methods
28-9095-31



Biacore Assay
Handbook
29-0194-00

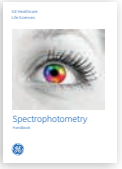
Ion Exchange Chromatography and Chromatofocusing
Principles and Methods
11-0004-21

Recombinant Protein Purification Handbook
Principles and Methods
18-1142-75



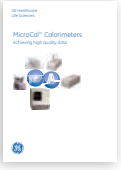
Biacore Sensor Surface
BR-1005-71

Isolation of Mononuclear Cells
Methodology and Applications
18-1152-69

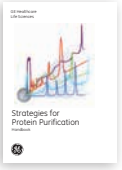


Spectrophotometry
Handbook
29-0331-82

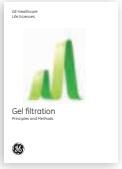
Cell Separation Media
Methodology and Applications
18-1115-69



MicroCal Calorimetry—Achieving High Quality Data
Handbook
29-0033-51

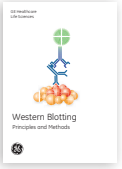


Strategies for Protein Purification
Handbook
28-9833-31



Gel Filtration
Principles and Methods
18-1022-18

Microcarrier Cell Culture
Principles and Methods
18-1140-62



Western Blotting
Principles and Methods
28-9998-97

Design of Experiments in Protein Production and Purification

Handbook

Contents

Chapter 1
Introduction

Design of experiments (DoE) is a technique for planning experiments and analyzing the information obtained. The technique allows us to use a minimum number of experiments, in which we systematically vary several experimental parameters simultaneously to obtain sufficient information. Based on the obtained data, a mathematical model of the studied process (e.g., a protein purification protocol or a chromatography step) is created. The model can be used to understand the influence of the experimental parameters on the outcome and to find an optimum for the process. Modern software is used to create the experimental designs, to obtain a model, and to visualize the generated information.

In a protein research lab or during process development, a DoE approach can greatly improve the efficiency in screening for suitable experimental conditions, for example, for cell culture, protein separation, study of protein stability, optimization of a process, or robustness testing. This handbook provides an introduction and an overview of DoE, followed by a step-by-step procedure that targets both newcomers and those with previous experience in DoE. The focus is on DoE for protein production and purification but the theory can be applied in many other applications.

Symbols used in this handbook

General advice

Warnings

Traditional experimental design versus DoE

DoE is not an alternative approach for experimental research. DoE is rather a methodology that provides stringency to the classical approach for performing research. How DoE can contribute to the statistical part of the research process is briefly illustrated in Figure 1.1.

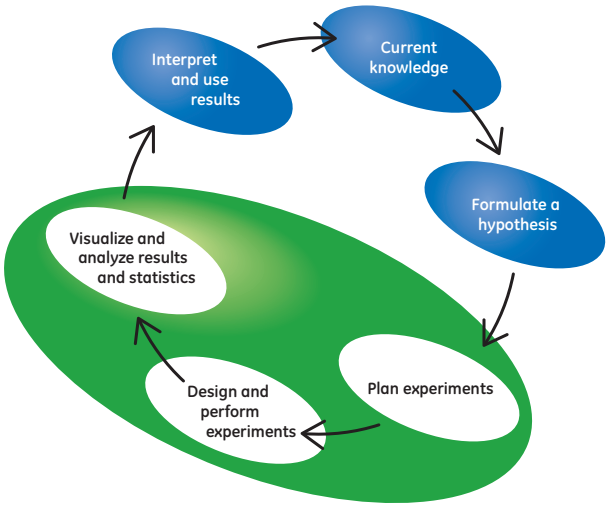


Fig 1.1. Schematic overview of the statistical contribution of DoE (the green oval) to the iterative research process.

Before describing the individual components of a DoE methodology, it is worthwhile to briefly consider the shortcomings of the traditional one-factor-at-a-time optimization approach. In the simplest traditional approach to optimize experiments, one parameter is varied while all others are fixed. In Figure 1.2, the traditional approach is exemplified with the optimization of yield in a purification step.

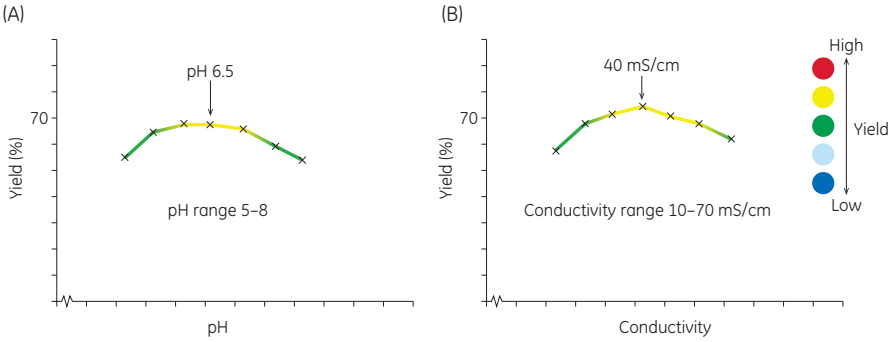


Fig 1.2. The traditional one-factor-at-a-time approach used for optimization of a protein purification step with respect to pH and conductivity. (A) A first series of experiments varying pH. (B) A second series of experiments varying conductivity at the pH value (pH 6.5) giving the highest response in (A). The experiments indicate that highest yield is obtained by using pH 6.5 and a conductivity of 40 mS/cm.

It can wrongly be assumed that the optimum levels for the factors (pH and conductivity in Fig 1.2) can simply be found by using the optimum levels of the factors obtained in the two series of experiments. As this setup only covers a limited part of the experimental space, this assumption is often incorrect. The reason is that experiments performed in the traditional way could be positioned out of scope, leading to no conclusions or sometimes even worse, the wrong conclusions. Further, the traditional setup does not take into account that experimental parameters can be dependent of each other (parameter interaction). In ion exchange chromatography, for example, the pH optimum will change when conductivity is changed. Thus, with the one-factor-at-a-time experimental setup, there is a great risk that the true optimum for the studied process is not identified. Ultimately, a study with the wrong setup cannot be saved or evaluated by even the most advanced statistical software programs.

In the DoE approach, on the contrary, process parameters are allowed to vary simultaneously, which allows the effect of each parameter, individually as well as combined to be studied. As shown in Figure 1.3, each parameter can have an optimum, but when combined, values might be found to give a different optimum.

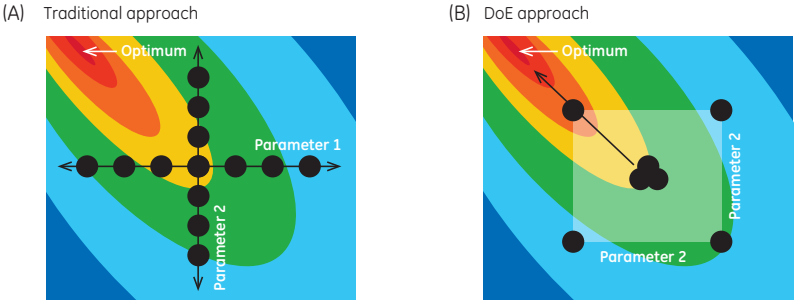


Fig 1.3. (A) The traditional one-parameter-at-a-time optimization approach versus (B) the DoE approach. With the DoE approach, the chances of finding the optimal conditions (in red) for the studied process increase. In addition, the combined effect of parameter 1 and 2 (interaction) on the response can be identified and evaluated. Although more experiments were performed than with the DoE approach, the traditional experimental setup failed to identify the optimum. Individual experiments are depicted with the filled black circles. Color scale: blue indicates a low response value and red a high response value, here the desired response (optimum).

When studying the effect of two or more factors on a process, the controlled arrangement of the DoE experimental setup allows collection of sufficient information with fewer experiments, compared to the traditional approach (Fig 1.4).

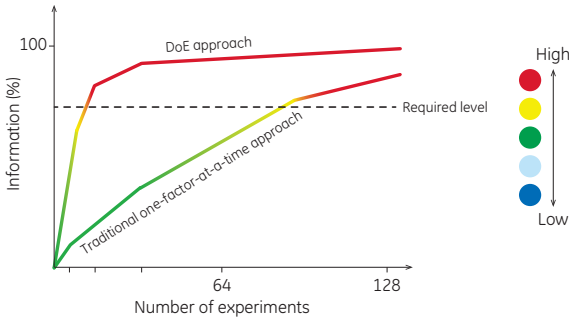


Fig 1.4. A schematic comparison of the number of experiments required to reach an acceptable level of information in an experimental study.

DoE nomenclature

Like all scientific disciplines, DoE has its own nomenclature. The nomenclature differs between the fields to which DoE is applied. Hence, there can be several terms for a particular phenomenon. For example, the model is also called the transfer function, the cause-and-effect relationship, or simply the relationship between our factors and the responses. In this handbook, we use the basic DoE terms introduced in Figure 1.5. Other terms are defined as they are introduced in the text. For some terms that are firmly established in protein research and in process development, we use that term and the corresponding DoE term alternately. In cases of uncertainty, please consult Appendix 2, Terminology.

DoE at a glance

DoE can be defined as a systematic way of changing experimental parameters (factors) to create results that can be methodically analyzed and that provide useful information about the process studied. Figure 1.5 provides an overview of the various steps of the DoE workflow.

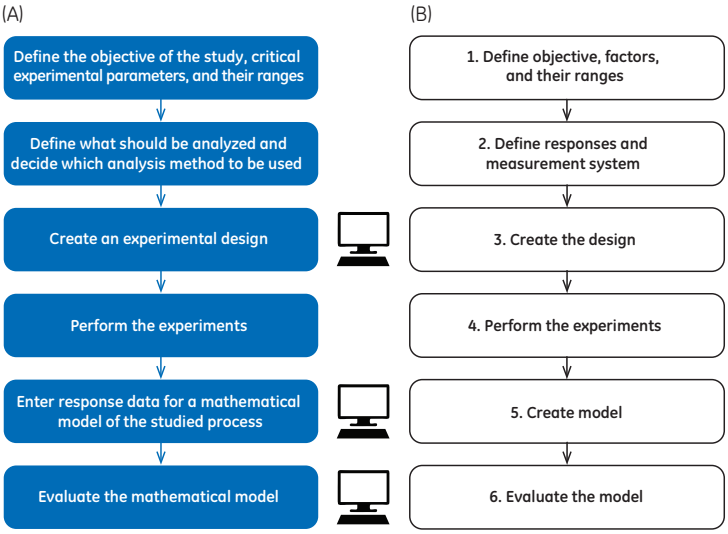


Fig 1.5. The DoE workflow summarized using both (A) the terminology used in protein purification and in (B) the corresponding DoE terminology used throughout this handbook. Predominately software-supported steps are indicated.

The different steps of the DoE workflow are described in more detail in Chapter 3. The first step involves defining the objective of the study and the factors that should be systematically varied. In the example in Figure 1.2, these factors are pH and conductivity. The range of variation (the upper and lower value of each factor) is also defined in this step. The second step involves defining relevant responses (type of analytical methods and data). Chapter 2 covers a comprehensive discussion on factors and responses that are relevant for DoE in protein production and purification.

A DoE experiment is set up in a rational way to cover the intended experimental space. As shown in Figure 1.6, the design can be visualized by a cube that represents the experimental space to be explored. The different factors are represented by the axes of the cube (x_1, x_2, and x_3 represent three different factors, e.g., pH, conductivity, and temperature). Using DoE, multiple factors handled in a single series of experiments can be viewed in arrangements called hypercubes as the setup becomes multidimensional. Depending on the study to be performed, different types of designs are available (see Chapters 3 and 6).

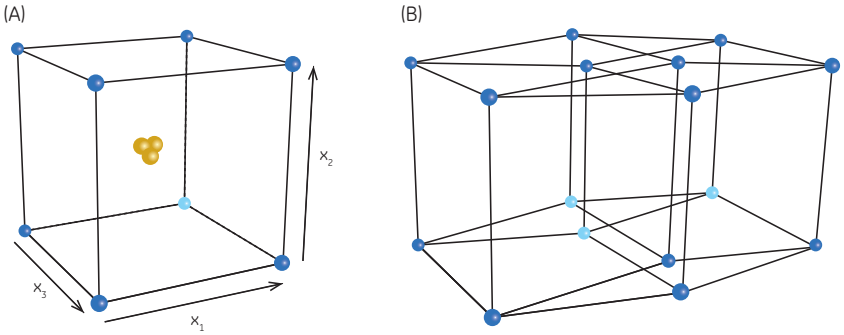


Fig 1.6. (A) A cube and (B) a hypercube representing the experimental space in a DoE setup. The cube represents an experimental design with three factors (x_1, x_2, and x_3), where the blue spheres correspond to experiments in which low and high settings for each factor are combined and the yellow spheres represents the replicated center point, that is, the midpoint setting for each factor. If more than three factors (e.g., four as in B) are studied, a hypercube can represent the set of experiments with combinations of all the factors. A center point (yellow), the mid value between the low and high setting for each factor, is replicated to estimate the experimental error and also to detect nonlinear relationships (see Chapter 3).

After performing the experiments according to the selected design, step 5 in the workflow (Fig 1.5) involves the use of DoE software for obtaining a mathematical model that describes the investigated process or system. A relevant model tells us, for example, which factors have a significant impact on the response and which factors do not. It is important to evaluate the model to determine its relevance, again using DoE software (step 6, Fig 1.5). The model is often visualized as a response surface plot and is used for evaluation of other parameter settings or process outputs within the experimental space (Fig 1.3 and Chapter 4). When performing a DoE study, it should always be carefully verified that the model is relevant. Verification of the model is preferably done through verification experiments within the experimental space. One important requirement for the model to be relevant is that there actually is a relationship between a factor and the response. UNICORN™ software, used together with the ÄKTA™ systems, provides support for the entire DoE workflow.

History of DoE

DoE methodology was proposed by Sir Ronald A. Fisher, a British statistician, as early as 1926. The pioneering work dealt with statistical methods applied to agriculture and the concepts and procedures are still in use today. In particular, Fisher and coworkers found that experimental design requires multiple measurements (i.e., replicates) to estimate the degree of variation in the measurements.

During World War II, DoE expanded beyond its roots in agricultural experiment as the procedure became a method for assessing and improving the performance of weapons systems. Immediately following World War II, the first industrial era marked a boom in the use of DoE. Total quality management (TQM) and continuous quality improvement (CQI) are management techniques that were later used also by the US armaments industry.

Some efficient designs for estimating several main effects simultaneously were developed by Bose and Kishen in 1940, but remained rather unknown until the Plackett-Burman designs were presented in 1946. About that time, Rao introduced the concepts of orthogonal arrays as experimental designs. This concept played a central role in the methods developed by Taguchi in the early 50s.

In 1950, Cox and Cochran published the book *Experimental Designs* that became the major reference work for statisticians for years afterwards. The development of the theory of linear models was considered and the concerns of the early authors were addressed. Today, the theory rests on advanced topics in linear algebra and combinatorics.

The history of DoE is briefly outlined in Figure 1.7.

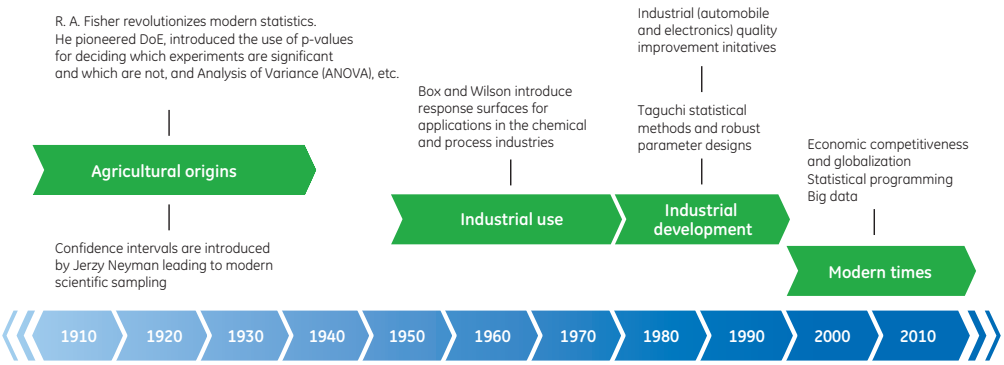


Fig 1.7. A summary of the history of DoE.

Industrial applications

The Taguchi methods were successfully applied and adopted by Japanese industries to improve quality and cost efficiency. In the 1960s, the quality of Japanese products started to improve radically as the Japanese car industry adopted statistical quality control procedures and conducted experiments that started a new era. Total quality management (TQM) and continuous quality improvement (CQI) are management techniques that were subsequently also embraced by US industry and are referred to as fractional factorial designs.

In the 1960s, randomized experiments became the standard for approval of new medications and medical procedures. Medical advances were previously based on anecdotal data, drawing conclusions from poor experimental setups and limited number of trials. The implementation of statistical procedures during this period was a move toward making the randomized, double-blind, clinical trial a standard method for approval of any new pharmaceutical product, medical equipment or procedure.

Since its beginnings in agriculture, DoE has been applied across many sectors of the industry. Around 1990, Six Sigma, a new way of representing CQI, became popular. Six Sigma is a technique that uses various tools, including DoE, to drive statistics-based quality improvements. Today, Six Sigma has been adopted by many of the large manufacturing companies including General Electric Company.

Chapter 2
DoE in protein production and purification

Optimization of protein purification is a multifactorial exercise, where factors such as sample composition, concentration and volume, purification technique, chromatography media (resins), binding, wash, and elution conditions (including pH, ionic strength, temperature, and additives) all might play a role. DoE is an excellent tool for identifying which factors are important and for finding an optimum for the process. The selection of factors and responses to explore in DoE for different protein purification situations is discussed in this chapter. A description of useful laboratory tools for performing DoE in the development of a protein purification process is also included. For a thorough description of protein purification strategies, please refer to the handbook *Strategies for Protein Purification* (28-9833-31).

For academic laboratories, the terminology used in the industry might sometimes be confusing. For example, the word "process" can refer to a method or protocol, but also to a complex workflow of different techniques and operations for a complete production and purification procedure. "Upstream" refers to the production part, for example, of cells or proteins in cell culture processes, whereas "downstream" involves processing of the material from the upstream process into a final product.

Biopharmaceutical protein production and purification differs from protein production and purification in the research laboratory, both in scale and in regulatory and industrial demands. The basic production, purification strategy, and technologies used, however, are the same. Therefore, the DoE approach in terms of selecting relevant factors and responses is similar in both the research and the industrial setting. An example from one area is generally applicable also to other areas. The main difference from a DoE perspective is that the industrial production setting requires stringent optimization to reach the most economical process. A comprehensive DoE effort is often a must in current process development. In research, good-enough purification schemes are often sought. In such case, a more limited DoE effort can be sufficient, comprising less coverage of the experimental space and thus requiring fewer experiments.

Upstream process development

In the biopharmaceutical industry, the production process for a target molecule is divided into upstream processing, including production of the target protein in a cell culture or fermentation process, as well as filtration steps; downstream processing, comprising yield of the target protein in a pure form; and final processing to gain product integrity and safety using techniques such as sterile filtration. The upstream process has a profound effect on the downstream process. Optimized upstream and downstream processes are essential for fast production of highly pure proteins.

Despite the long experience and widespread use of recombinant technology, many challenges remain when generating recombinant variants of native proteins. Critical cultivation conditions, such as temperature, pH, induction conditions, and medium composition, are carefully selected during process development to improve expression performance of the host cells. Because of the high number of impacting parameters and potential interaction between them, process optimization can be a tedious procedure. By applying a traditional trial-and-error approach, including changing one parameter at a time, the parameter responsible for the outcome can be identified. However, this approach requires numerous cultivations and is time-consuming.

As parameters interact, the identification of impacting factors is suboptimal and does not allow for detection of significant contributors to response changes. Using DoE helps to reduce necessary experimental burden, as settings of several parameters can be changed simultaneously. In addition, individual parameter effects as well as interactions and nonlinear effects can be identified and quantitated during data evaluation.

As an example of use in an upstream process, a DoE for optimization of monoclonal antibody (MAb) production was set up. Cultivation temperature and pH were selected as factors (Chapter 5.1). The responses were MAb monomer content, as monitored by gel filtration (size exclusion chromatography), and target protein concentration. For this MAb, a maximum for monomer content was found by cultivation at 32°C, at pH 6.8.

A strategy for protein purification

A protein purification process can be structured into different steps (Fig 2.1). In the capture, intermediate purification, and polishing (CIPP) model, sample preparation is followed by isolation and concentration of the target protein in the initial capture stage. During intermediate purification, bulk contaminants are removed. In the final polishing step, the most difficult impurities, such as aggregates of the target protein, are removed.

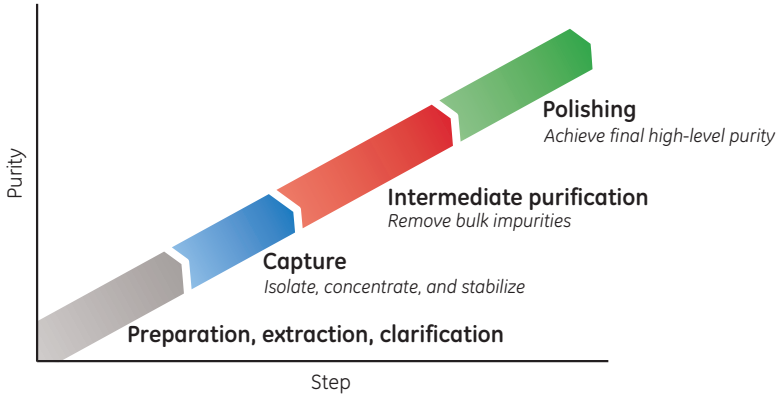


Fig 2.1. The different steps of a protein purification process. This way of structuring the protein purification process is referred to as the CIPP strategy. The goal of each step of the purification is indicated.

In a research setting, the purity requirements are often modest and, in such cases, only a capture step may be required. Also in cases where purity demands are higher, the capture step can be sufficiently efficient to be immediately followed by the polishing step. The number of purification steps will depend on the desired outcome of the overall purification process and on how efficient the individual steps are. Even in the simplest case, where our aim is a single purification step, a DoE approach can be beneficial. Increasing the number of purification steps will decrease the overall protein yield. Additional purification steps also means a longer purification time, which can be detrimental to the activity of the product. Thus, addition of purification steps will increase purity at the cost of decreased yield of active protein.

The individual purification steps

The target protein is separated from other sample components by its unique properties such as size, charge, hydrophobicity, isoelectric point (Ip), metal ion-binding property, and recognition of specific ligands (Table 2.1).

Table 2.1. Protein properties used during purification

Protein property	Technique
Specific ligand recognition	Affinity chromatography (AC)
Metal ion binding	Immobilized metal ion affinity chromatography (IMAC)
Charge	Ion exchange chromatography (IEX)
Size	Gel filtration (GF)
Hydrophobicity	Hydrophobic interaction chromatography (HIC)
	Reversed phase chromatography (RPC)
Size, charge, and hydrophobicity	Multimodal chromatography (MMC)
Isoelectric point	Chromatofocusing (CF)

The careful selection and combination of purification techniques is crucial for an efficient purification process.

Irrespective of technique, each purification step is a combination of several operations such as sample loading, wash, and elution. The experimental conditions for all of these operations are essential for achieving the desired purification goals. DoE for experimental planning, tools for modern high-throughput screening and optimization, and the automation capabilities of modern chromatography systems play a central role for simplification in resolving which conditions to be used in individual operations and in each purification step. With these tools and technologies, we can study essentially any key parameter in a multivariate automated approach.

Principles of combining purification steps

Often, the most efficient improvement in an overall protein purification strategy is to add a second purification step instead of optimizing a single-step protocol. Each purification technique has inherent characteristics, which determine its suitability for the different purification steps. As a rule of thumb, two simple principles are applied:

- Combine techniques that apply different separation mechanisms.
- Minimize sample handling between purification steps by combining techniques that omit the need for sample conditioning before the next step.

The strategy for combining purification techniques works well in both small laboratory- and large production-scale applications. Table 2.2 provides a brief overview of where different purification techniques are most suitable in a purification scheme. Applying these guidelines, Fig 2.2 shows some suitable and commonly used combinations of purification techniques for research applications.

Table 2.2. Suitability of purification techniques in a CIPP model

Typical characteristics	Purification phase					Conditions	
	Selectivity	Capacity	Capture	Intermediate	Polishing	Sample start conditions	Sample end conditions
Multimodal chromatography (MMC)	+++	++	+++	+++	+++	The pH depends on protein and type of ligand, salt tolerance of the binding can in some cases be expected	The pH and ionic strength depend on protein and ligand type
Affinity chromatography (AC)	+++++	+++	+++	++	+	Various binding conditions	Specific elution conditions
Immobilized metal ion affinity chromatography (IMAC)	+++	++	+++	++	+	Low concentration of imidazole and high concentration of NaCl	Intermediate to high concentration of imidazole, pH > 7, 500 mM NaCl
Gel filtration (GF)	+++	+	+		+++	Most conditions acceptable, limited sample volume	Buffer exchange possible, diluted sample
Ion exchange (IEX)	+++	+++	+++	+++	+++	Low ionic strength, pH depends on protein and IEX type	High ionic strength and/or pH changed
Hydrophobic interaction chromatography (HIC)	+++	++	++	+++	+++	High ionic strength; addition of salt required	Low ionic strength
Reversed phase chromatography (RPC)	++++(+)	++	(+)	+	+++	Ion-pair reagents and organic modifiers might be required	Organic solvents (risk of loss of biological activity)
Chromatofocusing (CF)	+++	++	+	+	++	Low ionic strength	Polybuffer, low ionic strength

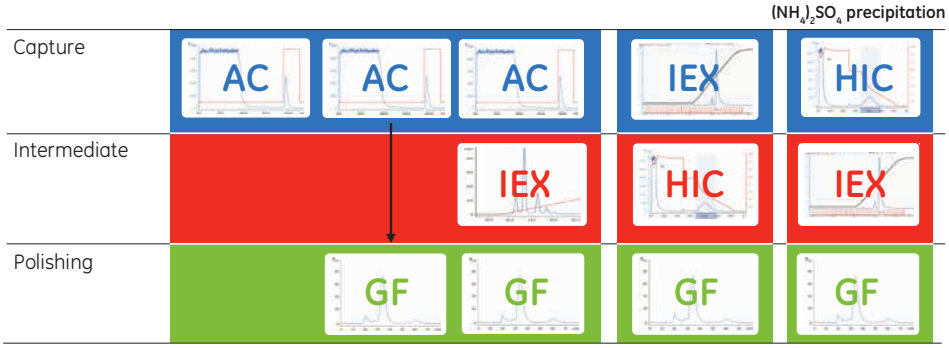


Fig 2.2. Example of suitable combinations of chromatographic techniques.

Keep in mind the interplay between purity and yield. Every added purification step will increase purity at the expense of overall process yield.

Principles of selection of chromatography media

The purification efficiency is also highly dependent of the chromatography medium selected for each technique. The efficiency, flow resistance, selectivity, and binding capacity differ between media. The particle size of the medium strongly affects efficiency and flow resistance. A medium with large beads gives chromatography columns with low resolution (broad peaks) but generates low backpressure, whereas small beads give higher resolution (narrow peaks) but also generates higher backpressure.

Figure 2.3 shows the general principle of choosing chromatography media with larger bead size for early purification steps, and smaller bead size for later steps, where demand on purity is increased. Inserted chromatograms show the separation results of applying a small, complex sample to columns with IEX media of different bead size. The importance of bead size is greater in large-scale purifications, where high flow rates are required for cost-efficient processing. At lab scale, intermediate-sized beads are commonly used for the initial capture step.

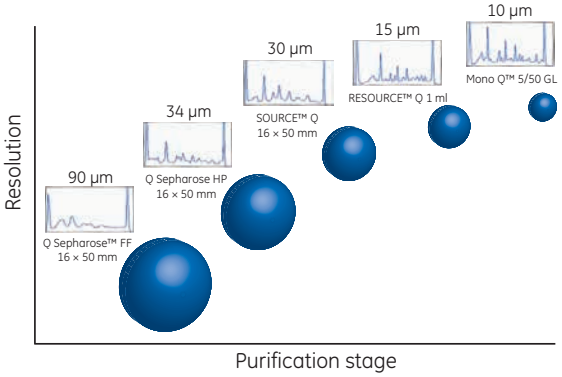


Fig 2.3. Principle for chromatography medium bead size selection. Large beads are primarily suitable for capture, while small beads offer the high resolution required for polishing.

Factors and responses when optimizing protein purification

Responses

The output of a protein purification scheme is traditionally described in terms of purity, homogeneity, and yield. Economy is an overall concern and some aspect of it is always targeted during optimization (Fig 2.4). These output parameters can be translated into a number of different responses and measurement systems for a DoE setup. Protein purity can be analyzed by measurement systems, such as electrophoresis or high-performance liquid chromatography (HPLC), and specified as the target protein-to-total protein (response) ratio. As a complement, host cell proteins (HCP), host cell DNA (hcDNA), and other impurities, can be monitored. Target protein homogeneity can be specified as monomer to aggregate content as analyzed by GF, or as correctly folded to incorrectly folded target protein as analyzed by HIC or RPC. Yield can be measured using a quantitative assay such as enzyme-linked immunosorbent assay (ELISA), or using an activity assay for the target protein, or both.

Overall economy depends on a number of parameters such as process time, buffer consumption, buffer additives, binding capacity and lifetime of the included chromatography media, the requirement for additional formulation steps (e.g., concentration and buffer exchange), as well as purchase cost for the components used. We also have to consider how much work (i.e., the number of experiments and responses) that can be conducted in the given time frame. Usually, all of these parameters need to be considered when planning a chromatographic purification. The importance of each parameter will vary depending on whether a purification step is used for capture, intermediate purification, or polishing.

Often, optimization of one of the output parameters can only be achieved at the expense of the other output parameters, and each purification step will therefore be a compromise. Purification techniques should be selected and optimized to meet the objectives for each purification step.

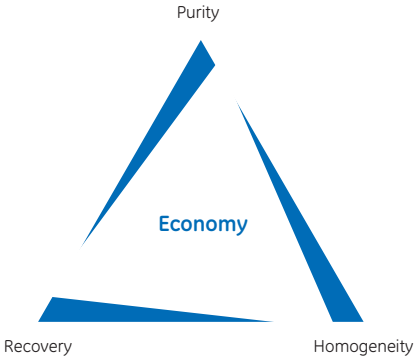


Fig 2.4. Common output parameters for a protein purification process.

Capture

In the capture step, the objectives are to rapidly isolate, concentrate, and transfer the target protein to an environment that will conserve its activity. A capture step is often a separation based on functional groups using, for example, IEX or AC, with step elution. Ideally, removal of critical impurities and contaminants is also achieved. With modern affinity chromatography media, it is often possible to achieve a very high level of purity during capture, for example, by using highly selective ligands for antibody purification such as protein A or protein L. In Chapter 5.5, the optimization of a capture step for an antibody fragment using Capto™ L is shown. Sample application and wash conditions (conductivity and pH) were varied and the effect on purity (measured as the reduction of HCP) and yield was estimated. An optimum was identified, at which HCP was reduced 25 000-fold at a high yield of the target protein (96%).

Key optimization parameters and responses for the capture step are outlined in Figure 2.5.

- Use a high-capacity, concentrating technique to reduce sample volume, enable faster purification and to allow the use of smaller columns.
- Focus on robustness and simplicity in the first purification step. Do not try to solve all problems in one step when handling crude material.
- For the capture step, select the technique that binds the target protein, while binding as few impurities as possible, that is, the technique that exhibits the highest selectivity and/or capacity for the target protein.
- Changing a protocol from gradient elution to step elution will increase speed at the expense of selectivity. Step elution will also increase the concentration of the target protein.

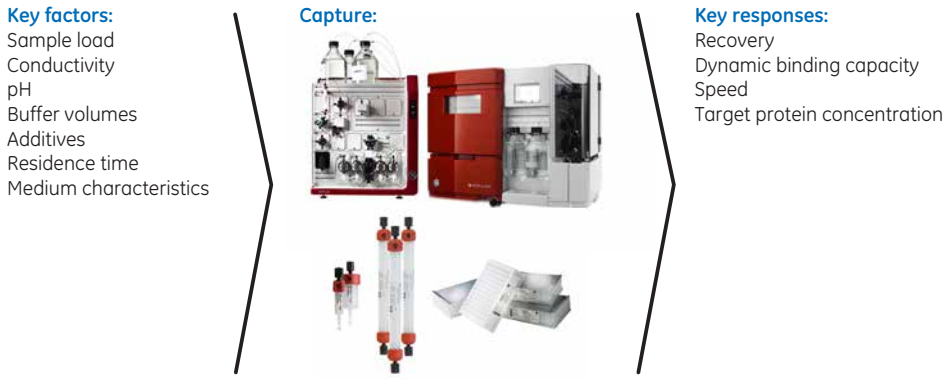


Fig 2.5. Initial purification of the target molecule from the source material is performed in the capture step, with the goals being rapid isolation, stabilization, and concentration of the target protein.

Intermediate purification

During the intermediate purification step, the key objective is to remove most of the bulk impurities, such as additional proteins, nucleic acids, endotoxins, and viruses. If the capture step is efficient, the intermediate purification step is often omitted in favor of one or more polishing steps. The ability to chromatographically resolve similar components is of increased importance at this stage (Fig 2.6).

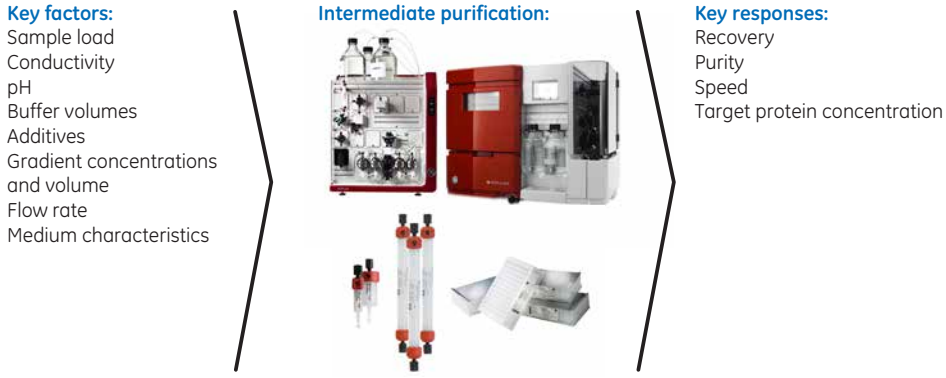


Fig 2.6. The intermediate purification step is characterized by the removal of bulk impurities, with the goals being purification and concentration.

For a maintained recovery, the capacity is still important in the intermediate purification step, as there might still be significant amounts of impurities in the sample. Speed is often less critical in the intermediate purification step as the sample volume is reduced and the impurities, causing proteolysis or other destructive effects, preferably have been removed in the capture step. The optimal balance between capacity and resolution should be defined for each case. In Chapter 5.4, the optimization of an intermediate purification step for a MAb using a cation exchanger is shown. The effect of pH and conductivity during sample loading was studied using dynamic binding capacity (DBC) of the chromatography medium as a response, and a global maximum was identified.

Polishing

In the polishing step, most impurities have already been removed. At this stage, only trace amounts of impurities remain, although these often consist of proteins closely related to the target protein, like fragments or aggregates of the target protein.

To achieve high purity and homogeneity, the focus is on chromatographic resolution in the polishing step (Fig 2.7). The technique chosen should discriminate between the target protein and any remaining impurities. To achieve sufficient resolution, it might be necessary to sacrifice sample load (overload might decrease purity) and yield by narrow-peak fractionation. On the other hand, product losses at this stage are more costly than at earlier stages. Preferably, the product should be recovered in buffer conditions suitable for the next procedure. In Chapter 5.7, a polishing step for a MAb was optimized. A key purpose of this step was to remove aggregates of IgG. Monomer content and purity were studied as a function of aggregate content in the starting sample, elution pH, and elution conductivity, and a sweet spot was identified.

Fig 2.7. Polishing involves removal of trace impurities and variants of the target protein as well as adjustment to conditions for final use or storage. The goals are high purity and homogeneity of the final product.

Chromatography step optimization

Each chromatographic step in a protocol can be split into several phases (Fig 2.8). Each phase can involve different factors (such as pH, conductivity, additives, flow rate), of which all can have a profound effect on the outcome of the chromatographic step. Because of the large number of tentative factors that can affect the outcome, a DoE approach is well-suited for screening and optimization of chromatographic purification steps.

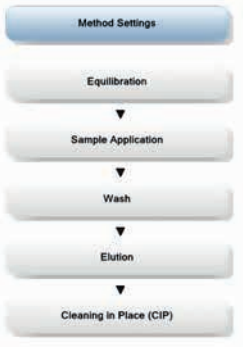


Fig 2.8. Typical phases of a chromatographic purification step.

Process development and characterization

During process development, where the aim is set at process optimization, the general strategy is to use initial screening designs, explore the cause-and-effect relationship, include additional factors, and explore wider ranges of factors settings. Usually, these initial DoE studies are performed in small scale. This initial screening could also give information about other process characteristics. For example, a lack of fit (model error), indicated in the model evaluation, can be due to a nonlinear relationship between factors and the responses (curvature effect) in the system (see Chapter 3 for more details).

In Chapter 3, a number of useful tools for process optimization are described, for example, fishbone diagrams, gage repeatability and reproducibility (gage R&R), and fractional factorial screening designs. When focusing on the detected critical factors, we use optimization type of designs comprising less factors, higher order models (including interaction and quadratic terms), and exploration of narrower ranges within the design space. The goal of process development is the specification of final operational parameter ranges in order to meet the quality criteria for the process or product. In this step, we try to reach both optimal settings for our parameters as well as process robustness. For example, we often try to reach the maximal product yield with minimal variation in product quality. The tools we choose often focus on accurate determination of curvature effects (i.e., response surface modeling (RSM) designs, see Chapter 3).

Process optimization is usually followed by verification of the process (robustness testing), at both small and large scale, using reduced designs (fewer studies of the important factors) in a range equal to an estimated factor variation. In this design space, we are looking at critical process parameters versus process/equipment capabilities.

Quality by design (QbD)

DoE provides a method for linking critical material attributes and process parameters to critical quality attributes of the product.

As stated in the FDA-issued guidelines for process validation and QbD (January 2011), the underlying principle of quality assurance is that a therapeutic product should be produced in a way so that it fits the intended use. This principle incorporates the understanding that quality, safety, and efficacy are designed or built into the product rather than obtained by inspection or testing. The guidance divides process validation into three stages: design, qualification, and continued verification. The QbD concept is described in Figure 2.9.

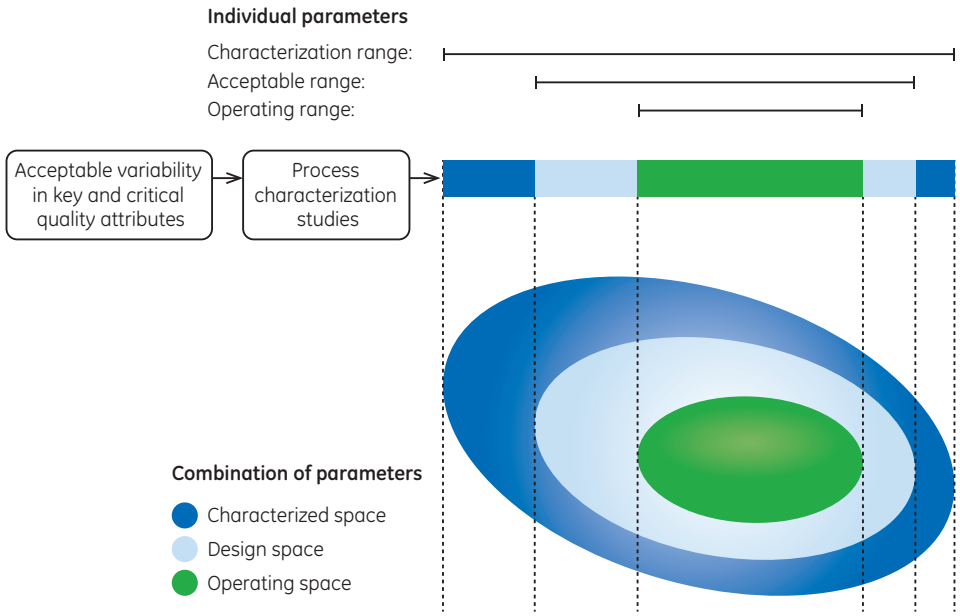


Fig 2.9. The concept of QbD. Process development aims at defining the design space (the experimental space that yields results within the set of responses specified for the process). An operating space is selected with conditions for operating the process in an optimal way. Data is collected while running the process. If the responses are found to be changed, or for any reason needs to be further optimized, changes of the running conditions can be made within the design space without revalidation of the process.

DoE is an integral part of the QbD concept. It is used, for example, for determination of the process design space and process parameters that are crucial for achieving the critical quality attributes for the final product. DoE also involves mapping of how the effects of process parameters depend on each other (interactions). The output of DoE for determination of design space forms a basis for setting acceptable and realistic variability ranges for the control space.

A structure for defining the process design space in QbD is outlined in Fig 2.10.

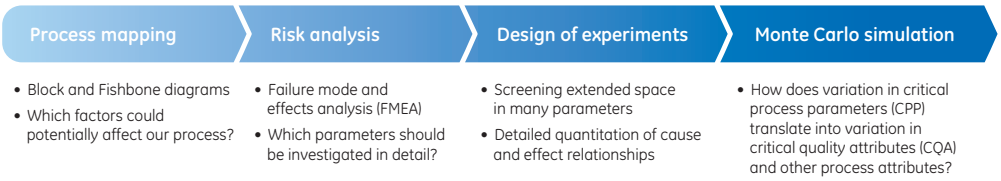


Fig 2.10. An example workflow for definition of a process design space.

Protein purification tools for DoE

Process development tools

Efficient development of the manufacturing process is a requirement in the biopharmaceutical industry as well as in other industries. A steadily increasing demand from regulatory authorities for a better understanding and control of manufacturing processes puts even more pressure on the development work. In high-throughput process development (HTPD), the initial evaluation of chromatographic conditions is performed in parallel, often using a 96-well plate format. Further verification and fine-tuning is typically performed using small columns before moving up to pilot and production scale. This approach to process development, using DoE, is illustrated in Figure 2.11.

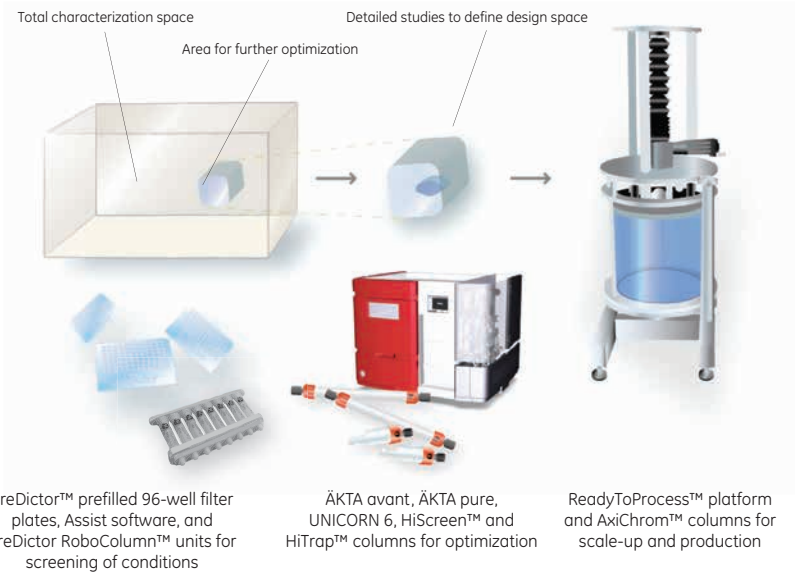


Fig 2.11. Conceptual workflow for process development using DoE.

PreDictor 96-well filter plates are prefilled with chromatography media. These plates can be used for the initial screening of process conditions, or for a more thorough investigation of a defined space as a basis for detailed process understanding and/or robustness studies. When using PreDictor plates, the fundamental interactions between the chromatography medium and the target molecule are the same as in chromatography columns. Basic concepts, such as mass balance, rate of uptake (at defined conditions), and adsorption isotherms, are the same in PreDictor plates as in chromatography columns. The Assist software helps chromatography process developers design and evaluate PreDictor plate experiments. The software provides guidance to experimental design and to handling experimental data, and also provides tools for data analysis. PreDictor RoboColumn units are prepacked, miniaturized columns that support HTPD using a robotic liquid handling workstation, such as the Freedom Evo™ platform from Tecan, for fully automated, parallel chromatographic separations. PreDictor RoboColumn units are available prepacked with a range of GE Healthcare chromatography media (Fig. 2.12).

(A) (B)

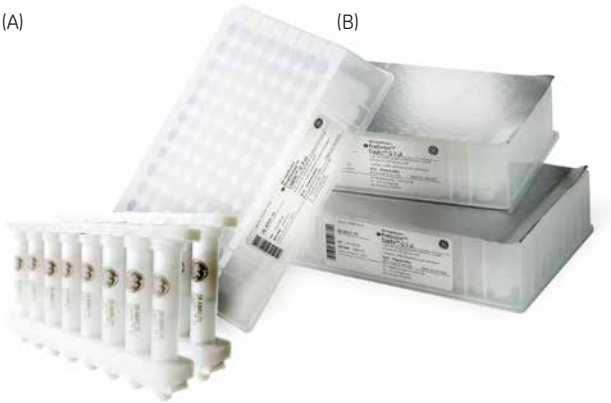


Fig 2.12. (A) PreDictor RoboColumn units are prepacked, miniaturized columns. (B) PreDictor 96-well filter plates prefilled with chromatography media.

After scouting and screening in a multiwell plate format such as PreDictor plates, verification and fine-tuning are performed in column format using ÄKTA systems such as ÄKTA avant or ÄKTA pure (Fig 2.13) with UNICORN 6 software.

Fig 2.13. ÄKTA avant and ÄKTA pure are high-performance chromatography systems designed for automated protein purification and process development. Prepacked columns, such as HiScreen 10 cm bed height columns, can be used with these systems.

UNICORN 6 software has integrated DoE functionality. As DoE is seamlessly integrated into UNICORN 6, methods are automatically generated from DoE schemes, allowing for fast and efficient process development. UNICORN 6 supports a number of different fractional factorial designs, full factorial designs, and response surface modeling (RSM) designs. Chapter 3 provides a comprehensive description of design types. Figure 2.14 illustrates a workflow for process development using GE Healthcare products.

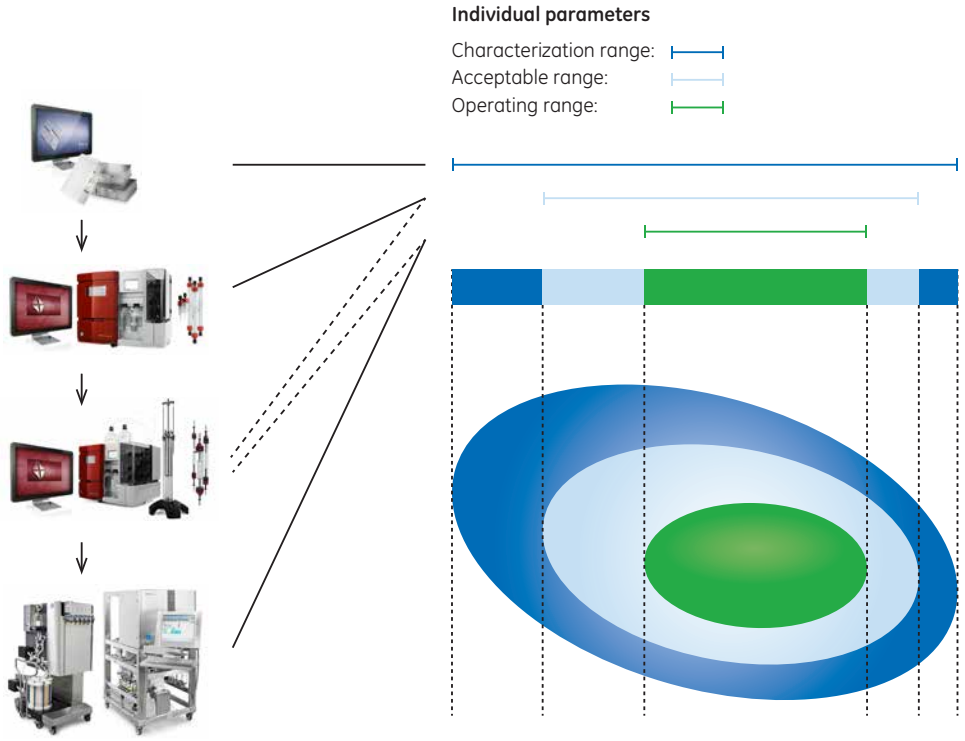


Fig 2.14. Process development workflow using GE Healthcare products for efficient applications of DoE under the QbD paradigm.

Other tools for parallel protein purification

DoE greatly improves efficiency in method development and optimization. Laboratory tools, enabling experiments to be performed in parallel, contribute to the efficiency gain. MultiTrap™ 96-well filter plates are prefilled with chromatographic media and are designed for efficient, parallel initial screening of chromatographic conditions (Fig 2.15). MultiTrap plates can be operated manually (using a centrifuge or a vacuum chamber to drive liquid through the wells) or in an automated fashion using a laboratory robot. Conditions defined in a screening study using MultiTrap plates can be verified and optimized using prepacked chromatography columns such as HiTrap or HiScreen columns.

Magnetic beads offer another possibility for small-scale, parallel protein purification, for example, for screening purposes (Fig 2.15). Mag Sepharose beads (substituted with protein A, protein G, and other affinity ligands) can either be operated manually or automated. An example of a DoE study for optimization of IgG purification on Protein A Mag Sepharose Xtra is given in Chapter 5.6.

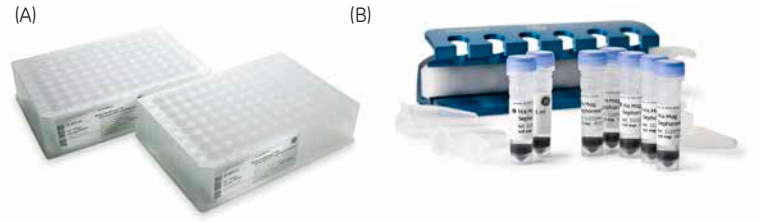


Fig 2.15. (A) MultiTrap filter plates prefilled with chromatography media and (B) and Mag Sepharose beads.

Chapter 3
Design of experiments, step-by-step

The individual DoE steps were briefly defined in Chapter 1. This chapter gives a step-by-step reference to performing DoE. Visualization of DoE results is described in Chapter 4. Application examples are given in Chapter 5 and additional information is found in Chapter 6.

Performing a DoE study is typically an iterative process. For example, an initial DoE could be set up for screening of process conditions (e.g., to identify relevant factors) before a more targeted DoE is conducted for careful optimization of the process.

In this chapter, DoE is described in a step-by-step fashion for clarity, but the iterative nature of the methodology should be kept in mind. Also, it should be understood that each of the individual steps described here typically have an impact on, or are impacted by, how later steps will be carried out. The design type, for example, should be briefly considered (create design, Step 3) while defining the objectives (Step 1) for the extent of the study to become realistic from a resource-availability perspective. Similarly, the response and measurement systems (Step 2) need to be considered when setting the objectives. For these reasons, each individual step described in this chapter will contain references to both the previous steps and also to the later steps.

Gather information before starting

For a DoE setup, some basic information regarding the process to be studied must be available. As a wealth of information is often available, both general (e.g., an ion exchange separation is always impacted by the ionic strength and pH of the solutions used) and specific (the size and sequence characteristics of a recombinant protein to be purified), having an idea of which factors could possibly impact the process is typically not an issue. A good starting point for a DoE setup is to collect all available information. If there are obvious gaps, initial experiments should be performed, for example, chromatographic runs using small column formats for indications on factor (starting) levels.

Step 1. Define objective, factors, and ranges

Step 1 contains the following elements:

- Define the overall project goal(s) and the study objective.
- Define process requirements (measurable) or issues that are not strictly part of the DoE study but that need to be fulfilled.
- Define the size of the study.
- Identify all parameters that have an effect on the end result and exclude irrelevant ones.
  - Pool all available information about the factors and responses.
  - Perform test experiments if necessary.
- Define factors and their levels.

The initial DoE step is critical. It lays the foundation for a good, conclusive experimental study. Most DoE failures can be traced to incomplete and/or incorrect identification of parameters (both input and output) and, hence, to an incorrect setup of the study. It could be tempting to move quickly to optimization (if that is the purpose) of a few parameters, but one has to be sure that the selected factors are the relevant ones. A screening DoE is a great tool for rapid identification of relevant factors and for screening of factor levels.

The first DoE step covers exactly the same elements as any experimental planning, irrespective of experimental design. The underlying logics will easily be recognized. During the first step, basic questions, such as study purpose, factors to include, and achievability within the given resource frame, are posed.

Objectives

The starting point for any DoE work is to state the objective, define the questions about the process to be answered, and choose the relevant factors and ranges.

The objective describes the purpose of the study. For a screening study, the objective could be to identify key parameters that impact purity and yield in an affinity chromatography capture step. Another objective could be to identify the most suitable chromatography medium for achieving high target protein homogeneity in a polishing step. These screening studies could subsequently be followed by optimization studies, again using DoE, with objectives such as maximizing purity and yield in antibody capture using a protein A medium or maximizing aggregate removal through a multimodal ion exchange purification step. It is important to define the requirements in detail and to make sure that they can be measured. For example, maximizing aggregate removal is linked to additional requirements such as the maximum allowed separation time or the minimum allowed yield of target protein. It is useful to start by listing a brief, overall study objective and the additional requirements. When this is listed, focus should be on how these requirements can be measured.

At a detailed level, there are as many study objectives as there are different studies. But at a higher level, there are three different categories of studies for which DoE is typically used. These are screening studies, optimization studies, and robustness testing. These studies differ not only in their objectives but also in the number of factors used, and the number of experiments involved (Fig 3.1).

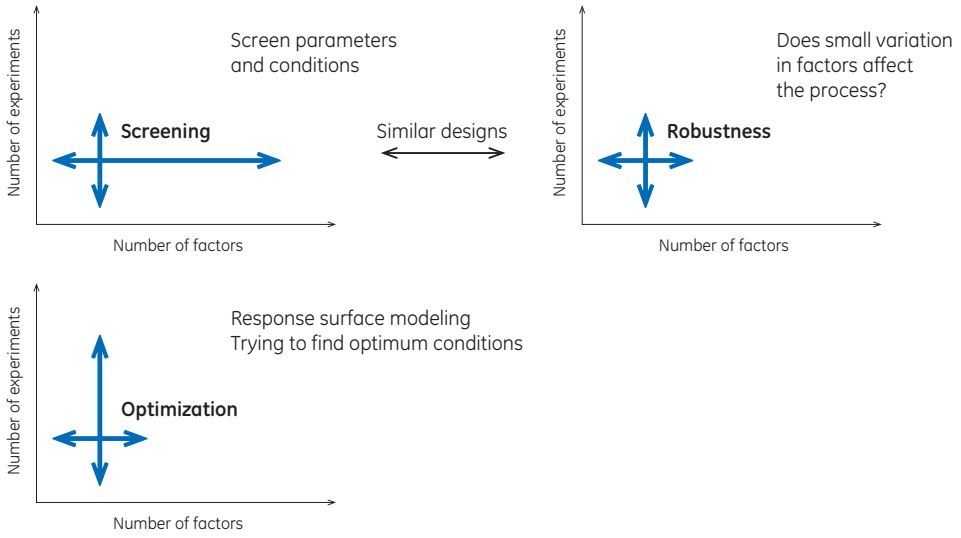


Fig 3.1. Common overall objectives for different types of DoE studies and some of their characteristics. The arrows indicate the number of experiments and the number of factors for these objectives. A screening DoE is used for obtaining information about factors and for selecting significant factors and their settings. Optimization is used for finding the optimal levels of a few relevant factors, and for obtaining a useful process model for future predictions. Robustness testing is used for verifying that the optimized conditions will give the expected response, with a variance that is sufficiently small. The different DoE objectives can be used sequentially to develop a process.

For future reference, a study protocol is useful for documentation of the experimental rationale, as well as the objectives, methods, and more.

Screening

Screening DoE explores the effects of a large number of factors in order to identify the ones that have significant effect on the response of a process or system and to determine which factors need to be further characterized or optimized. The law of the vital few (Pareto principle) states that for many events, roughly 80% of the effects come from 20% of the causes. By screening, we ensure that all critically important variables are considered before reducing number of variables. Screening DoE is also used when screening for conditions such as factor levels or ranges.

Optimization

Optimization is used for determination of optimal factor settings for a process or a system. The relationship between a selected response and the significant factors is determined in order to create a model. Several responses may be optimized simultaneously, but the final factor settings might be a compromise between the responses. Additional designs may be applied when the studied region does not contain the optimum. When additional factor ranges are required, this should be guided by the results from the previous design. The settings are varied to allow crawling across the experimental region towards optimum conditions. When the optimum has been identified, the chosen conditions should be verified.

Robustness testing

Robustness testing is used for determination of process robustness and is conducted by performing experiments based on a design, with only minor adjustments of the factor levels. The variation of the obtained responses should be within set specification limits. If the responses do not vary significantly when the factor levels are changed, the process is considered to be robust.

Design considerations

Although creation of the design is described in a later step, it is worthwhile to already at this point consider design selection, as the study objective will have an impact on which type of design is required. There is a balance between the amount of information needed and the number of experiments required (Fig 3.2). This balance is the initial, and one of the most important, considerations to be highlighted before conducting experiments. We should also keep in mind that the experimental process can be iterated and the initial screening results from the first DoE step can be used as input for the next DoE step.

Factors and factor ranges

Factors are the input parameters or conditions that we wish to control and vary for a process. We expect that a selected factor has an impact on the response that we intend to measure. Already when we decide what factors to control (see example in Fig 3.2), we also assume that other factors will not significantly affect the response and will be constant or left uncontrolled during the experiments. Constant variables are, if possible, fixed, measured, and recorded during the experiments. One common example of a constant variable in protein chromatography is temperature.

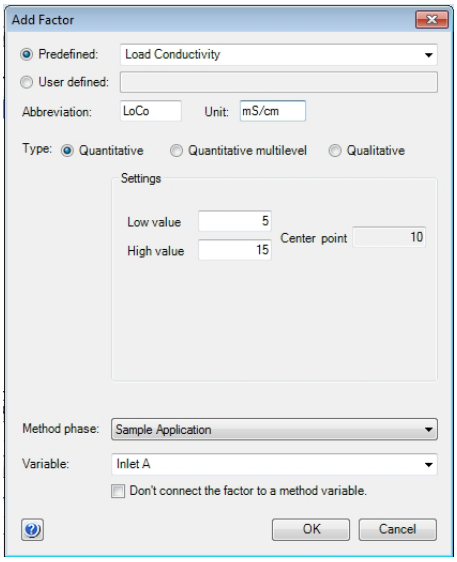


Fig 3.2. Factors and factor ranges are entered into the UNICORN software, for automated chromatography runs.

The factor range (i.e., the levels that are possible to set) depends on the experimental objective and the experimental noise (nonsystematic variability) for easier readability. In screening studies, the range should be large enough to increase the possibility of covering the optimum and to obtain effects above the noise. A broad range can also make the model more stable. In optimization DoE, the range should, and can usually be reduced as there is more information available at this stage. Determination of appropriate factor starting values (e.g., pH and conductivity in IEX) could include performing separate gradient runs (one for pH and one for conductivity) to find an approximate factor range for elution of the target.

Setting of factor values and ranges requires attention to avoid exceeding physical/chemical restrictions such as recommended maximum flow rate for a chromatography column.

DoE studies most often use two numerical levels for each factor. For each factor, a high and a low value are selected for the settings to cover the operating range for each variable. As an example, for optimization of an affinity chromatography separation, one selected factor can be flow rate during sample load. The range for this factor is selected from 1 to 5 mL/min. The low value is 1 mL/min and the high value is 5 mL/min. The design will thus cover experiments within these two factor levels. For the DoE model to be reliable (Step 5), however, we also need to detect possible nonlinear relationships between the factors and the response(s).

Therefore, it is useful to complement the design with center points to enable detection of the possible occurrence of a curvature effect (a nonlinear relationship). The center point experiments are also replicated for estimation of the experimental variation. In the example above, the center point will be at a flow rate of 3 mL/min. If a significant curvature is found, so called, star points can be added to quantitate the nonlinear effects of individual factors. This type of design is described in more detail under Step 3 in this chapter and in Chapter 6.

It is advisable and more economical to include more factors in the initial screening study than to add factors at a later stage. In some cases, this means that the screening study includes a very large number of experiments. With a large number of factors, the most common way to reduce the number of experiments is to use a reduced design (as described in Step 3). Alternatively, one can reduce the number of factors and runs by introducing dimensionless factors.

A dimensionless factor consists of the ratio between two parameters that each was initially considered to be factors of their own. For example, if both the concentration of NaCl and the concentration of the additive urea were considered as factors in a DoE study, the ratio between the molar concentrations of NaCl and urea could be used as a single factor to reduce the number of experiments. If we have three variables with the same unit in our study, the three variables x_1, x_2, and x_3 can sometimes be combined in the ratios x_1/x_2 and x_3/x_2 to give the same information as for two variables. The introduction of dimensionless factors adds a new level of complexity to the interpretation of the results, but is sometimes highly relevant, as interactions between dimensionless factors or between a dimensionless factor and a physical parameter can increase our understanding of the process.

Quantitative and qualitative factors

Quantitative factors are characterized by being on a continuous scale, for example, pH, flow rate, and conductivity. Qualitative factors are discrete (discontinuous), for example, column type, type of chromatography medium, and buffer substance.

Controllable and uncontrollable factors

Controllable factors can be managed in experiments. Uncontrollable factors might affect the response but are difficult to manage, for example, ambient temperature or target protein amount in the cell culture. Monitor uncontrollable factors when possible. When entered into the DoE software, the effects of the uncontrollable factors can be evaluated.

Structured approach to selecting factors

A fishbone diagram (also known as cause-and-effect or Ishikawa diagram) is a useful tool for documenting factors believed to have the greatest impact on the process, for example, in an initial cause-and-effect analysis. Fishbone diagrams also provide an easy reference for a preliminary risk assessment. The diagram is useful in process development, providing a structured way to list all possible causes of a problem or a desired process outcome before starting to think about a solution. An example of a fishbone diagram is given in Figure 3.3. The way this diagram is constructed includes the following steps:

- Identifying the process goal(s) (e.g., yield, purity, etc.), represented by the middle horizontal arrow.
- Identifying the major process steps (e.g., sample load, wash, elution, etc.), represented by the tilted lines.
- Identifying possible factors affecting the goal(s), represented by the numerous horizontal lines going from the tilted lines.
- Analyzing the diagram.

A fishbone diagram enables detection of the root causes (factors) to discover bottlenecks, or to identify where and why a process is not working.

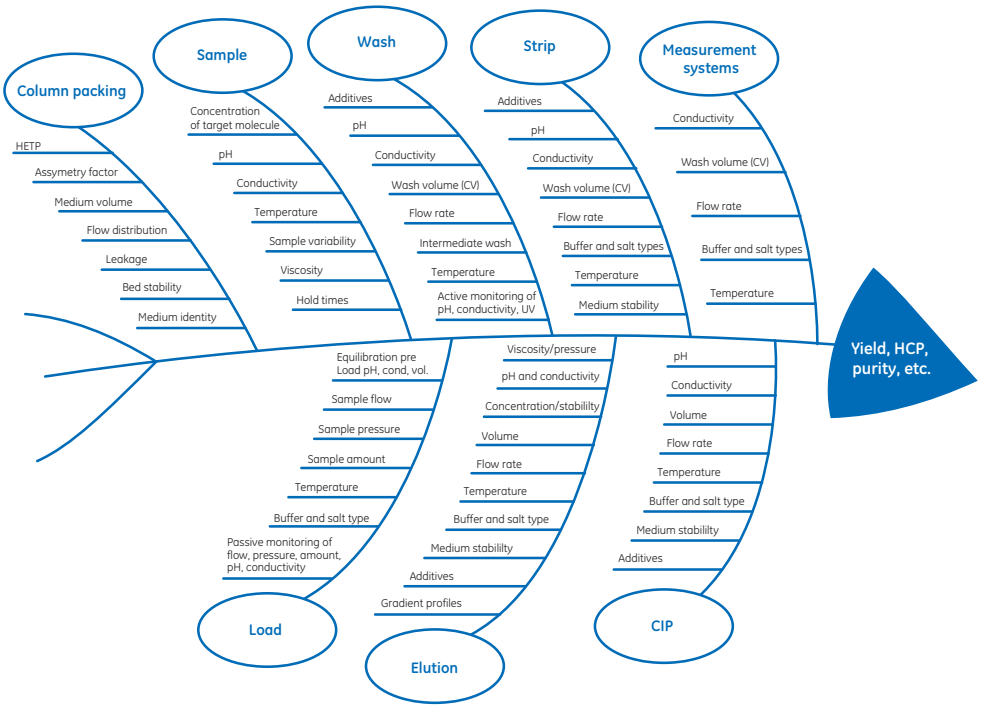


Fig 3.3. Example of a fishbone diagram showing possible contributors leading to a desired process output. The example shows the majority of process parameters in an ion exchange step. Only a fraction of these are relevant for a specific application. HETP = height equivalent to theoretical plate, CV = column volume, CIP = cleaning in place.

Independent measurements of factor effects

Orthogonality in an experimental design means that each factor can be evaluated independently of all other factors. For example, in a two-factor, two-level factorial design, this is achieved by arranging experiments so that each level of one factor is paired with each of an equal number of levels of the other factor. By setting the factor levels at the combined low and high values, we outline the factorial part of the experimental plan. For each factor to have an equal opportunity to independently affect the outcome, the factor levels are transformed. The coding is a linear transformation of the original measurement scale, so that if the high value is X_H and the low value is X_L, the factor level is converted to $(X - a)/b$, where $a = (X_H + X_L)/2$ and $b = (X_H - X_L)/2$, that is, the high value becomes +1 and the low value becomes -1. If k is the number of variables included in our DoE, we have 2^k experimental runs (in a full factorial design), of which each corresponds to a unique combination of factor settings. An example of a coded design matrix, where +1 and -1 are used for the high and low settings, respectively, is shown in Figure 3.4.

Exp. no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1
2	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1
3	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1
4	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1
5	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1
6	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1
7	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1
8	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1
9	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1
10	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1
11	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1
12	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1
13	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1
14	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1
15	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1
16	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1
17	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1
18	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1
19	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1
20	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

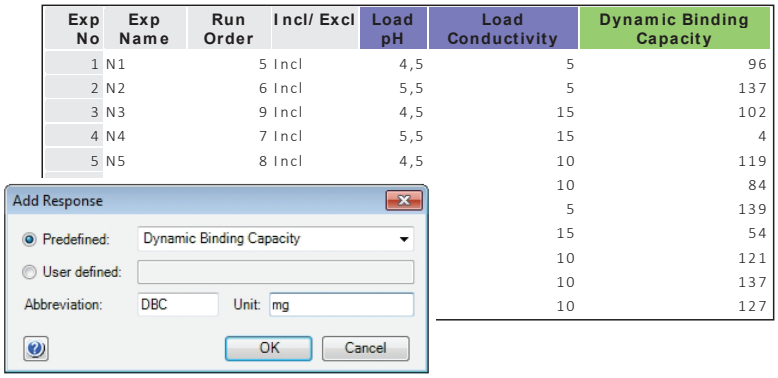
Fig 3.4. A coded data matrix with high (+1) and low (-1) settings for each factor. In this case, based on a Plackett-Burman screening design, 19 factors were investigated in only 23 runs. This corresponds to $23 << 2^{19}$ (524 288) runs (the number of runs in a full factorial design).

Step 2. Define responses and measurement systems

Step 2 contains the following elements:

- Define what to measure as an output from the process, and set the specification limits (i.e., our criteria for acceptable results) for these responses (critical-to-quality [CTQ] attributes)
- Define a reliable measurement methodology, and perform a measurement system analysis

Responses are the output parameter(s) of a process. In protein purification, responses can, for example, be binding capacity of the chromatography medium or purity and yield of the target protein (Fig 3.5). The strategy for defining a response is to find a response that maps the properties of a product, the performance characteristics of a process, or more specifically, the study objective. If several responses are required for describing the outcome, we need to analyze the advantages and disadvantages of having multiple responses in the evaluation. Having multiple responses often leads to compromises. A response should be appropriate (address the desired process characteristics effectively and coherently in a manner suited to the context) and feasible (we should be able to measure the response efficiently with available resources in a timeframe suitable to describe the system or process).



Exp No	Exp Name	Run Order	Incl/ Excl	Load pH	Load Conductivity	Dynamic Binding Capacity
1	N1	5	Incl	4,5	5	96
2	N2	6	Incl	5,5	5	137
3	N3	9	Incl	4,5	15	102
4	N4	7	Incl	5,5	15	4
5	N5	8	Incl	4,5	10	119
					10	84
					5	139
					15	54
					10	121
					10	137
					10	127

Fig 3.5. Responses and their values are entered in the UNICORN software for evaluation.

Quantitative and qualitative responses

Quantitative responses (e.g., protein purity and yield) are characterized by being on a continuous scale. These responses enable easy model interpretation. When response metrics are not possible to obtain, response judgments can instead be used. Such judgments, or qualitative responses (e.g., product quality), are discrete and result in outcomes such as yes/no, or low/medium/high. When possible, qualitative responses should be made semiquantitative by transforming the judgments to discrete values such as 1 = very low, 2 = low, 3 = medium, 4 = high, and 5 = very high.

Measurement system requirements

One major step, sometimes overlooked in scientific work, is to make sure that the measurements error is smaller than the variation in the process outputs. As managing a defined process requires knowledge of the progress and its end product or response, we need to define what to measure and the measurement system requirements. The selected output parameter should be valid for the process, that is, valid to detect the input-output relationship. A continuous variable is preferred and the response should be relatively easy to measure.

When we have defined what to measure, we can determine the quality of our measurement system, including measurement accuracy, acceptable measurement errors, stability, capability, and acceptable level of variation in the method. A measurement system needs to have good accuracy and precision, with minimal variability (noise), to enable reaching the correct optimal conditions for the process. A poor measurement system might overestimate or underestimate the process outcome and prevent detection of improvements made during screening and optimization. Measurement system analysis (MSA) refers to a method for determining accuracy and precision of a measurement system. The overall goal of MSA is to determine how much the variation of the measurement system contributes to the overall process variability (Fig 3.6).

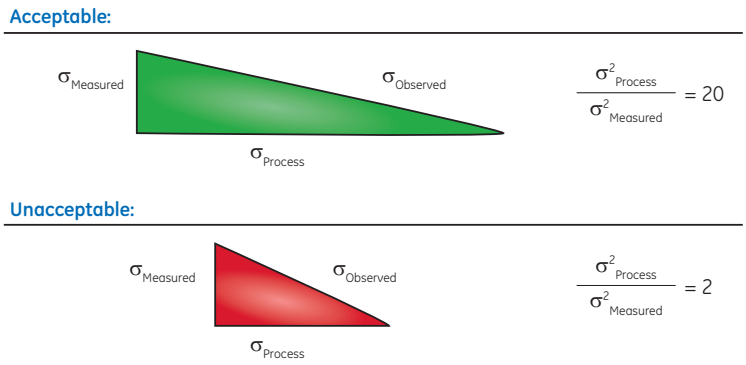


Fig 3.6. Analyzing measurement system capability, that is, measurement error (gage R&R) in relation to process output variation. The triangles show the relationship between the actual process variance ($\sigma_{Process}$), gage (measurement system) variance ($\sigma_{Measured}$), and the observed process variance ($\sigma_{Observed}$). A variance ratio of less than 2 is unacceptable, whereas a ratio greater than 20 is acceptable. Values between 2 and 20 indicate situations that should be handled cautiously.

Automatic peak integration and densitometry is supported by modern software. Manual operations, being more subjective, might sometimes introduce an abnormally large variation in the measurement system, such as a lack of accuracy and precision.

Step 3. Create the design

Step 3 contains the following elements:

- Use the information from Step 1 and 2 to create an experimental design, by hand or in a DoE software program (for easy setup).
  - The experimental design should include all factors believed to be relevant and to affect the CTQ responses.
  - The design setup can include all or a fraction of all high and low value combinations of all factors.
  - Decide the extent of the study. Are we interested in the main effects only or do we wish to include interactions and/or curvature; do we suspect a nonlinear relationship between the factors and the responses?
  - Setup of the experimental plan and review accuracy in the selection of design type, for example, by reviewing the correlation matrix and design power.
  - List constraints/assumptions/limitations.
- The defined experiments should represent a relevant sample size and an independent, unbiased, and randomized selection.
- Review the plan to make sure that every experiment is reasonable and feasible, and describe how the experiments are to be executed.

The corner points—high and low value combinations of a design—will provide information on how the factors, one by one or together, create a certain effect on the measured responses.

Creating an experimental design is a straightforward process performed by defining factors, factor ranges, and the objective. The experimental design is also related to the modeling. A specific design allows addition of specific mathematical terms to the model (mathematical description of the process), which depends on the complexity of the selected design (Fig 3.7). Terms used in modeling, and defined in the context of this handbook, are the main (linear), interaction (two-factor), and quadratic terms.

Screening designs are useful when we wish to measure the main effects or when we wish to disregard parameter interactions or nonlinear relationships. Screening designs are also useful for robustness testing when we wish to reduce the number of runs. In optimization designs, more experiments are added. These additional experiments, as defined by the factor settings, are used for quantitation of nonlinear cause-and-effect relationships and allow us to increase the complexity of the mathematical modeling by adding square terms, and hence, to spot a minimum or maximum for our process.

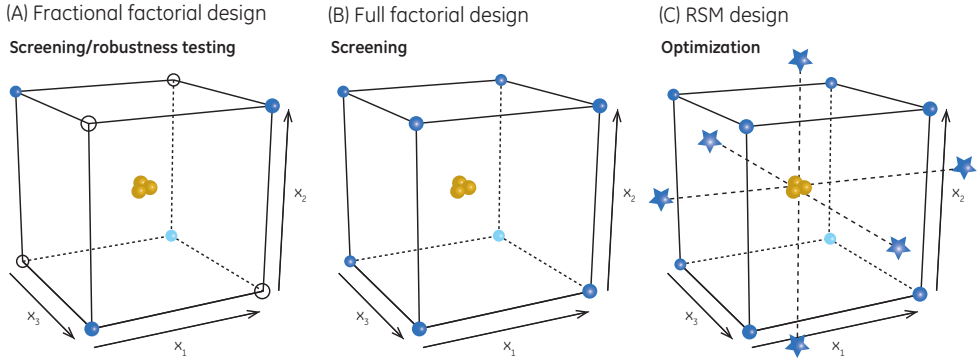


Fig 3.7. An overview of the key design types used in DoE: (A) fractional factorial design, with a fraction of corner point experiments excluded (white circles). Fractional designs are suggested for screenings, as the information provided using this design is often sufficient to find the factors (main effects) affecting the process; and for robustness testing, as the optimal factor settings have already been found and only minor changes in the factor settings are used to test the robustness of the process. (B) The full factorial design uses all corner point experiments. This design is often suggested for screening. Information about which factors that are important (main effects) and about factor interaction effects is obtained. (C) For optimization studies, especially if curvatures are detected, the basic full factorial design is extended with additional experiments outside the box, called star-point experiments, in the response surface modeling design (RSM). Star points enhance the detection capability for curvatures and give information about main factor effects, factor interaction, and curvature. Replicated center-point experiments (yellow) are always included in the designs.

The use of DoE does not eliminate the need for or purpose of single experiments. Careful selection and execution of single experiments based on a hypothetical experimental design can address initial questions. Single experiments can be used to test initial predictions and hypotheses on a subject. As an example, single experiments can be used to investigate the effect of two parameters at two levels by using the corner points of a design instead of the traditional one-factor-at-a-time approach.

Figure 3.8 displays an example of a DoE setup where the dynamic binding conditions for a chromatographic purification method is to be investigated. The low and high levels of each factor (in this case two) are entered into the DoE software and a worksheet is generated based on the selected design (Fig 3.9). Star-point values for quantitating curvature effects and a randomized run order (to ensure a nonbiased study) are generated. After the experimental series has been finished, the results (responses) for each experimental point are entered for evaluation of the DoE setup. The evaluation involves statistical analysis of the data generated.

Exp. no.	Exp. name	Run order	Load pH	Load conductivity	Dynamic binding capacity
Factorial part			4.5	5	96
			4.5	15	102
			5.5	5	137
			5.5	15	4
Star points			5	5	139
			5	15	54
			4.5	10	119
			5.5	10	84
Replicated center points			5	10	121
			5	10	137
			5	10	127

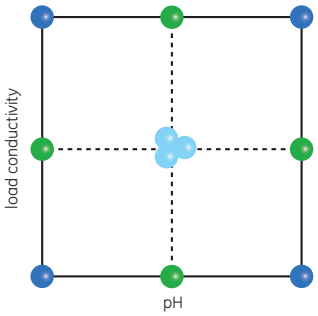


Fig 3.8. An example worksheet for the optimization of dynamic binding conditions for a chromatographic purification method. The upper part of the table shows the factorial part of the design (corner points) for estimating main effects, the middle part shows the star points for assessing curvature effects, and the bottom part shows that three replicated center points are used for determination of the true experimental error (noise).

Exp. no.	Exp. name	Run order	Incl./excl.	Load pH	Load conductivity
1	N1	5	Incl	4.5	5
2	N2	6	Incl	5.5	5
3	N3	9	Incl	4.5	15
4	N4	7	Incl	5.5	15
5	N5	8	Incl	4.5	10
6	N6	3	Incl	5.5	10
7	N7	2	Incl	5	5
8	N8	4	Incl	5	15
9	N9	10	Incl	5	10
10	N10	1	Incl	5	10
11	N11	1	Incl	5	10

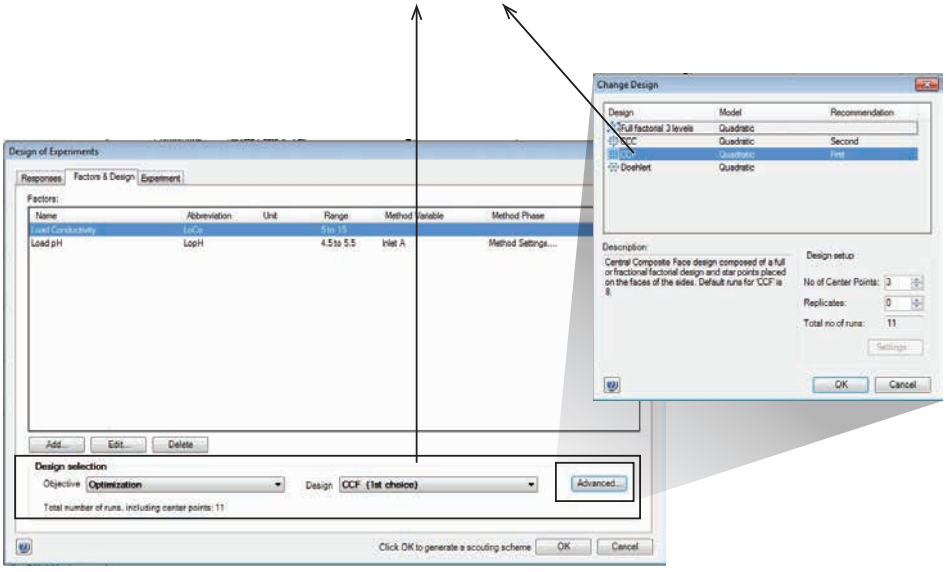


Fig 3.9. Design selection in the UNICORN software. The low and high levels of each factor are entered whereupon a worksheet is generated for the design, suggested based on the user-selected objective. The design can also be selected from an extended design list.

Design resolution

The term resolution in DoE tells us the type of effects that can be revealed (and the mathematical terms that can be added to the modeling) with the design in question. Resolution depends on the number of factors and the number of runs and the given resolution value refers to the confounding pattern (Fig 3.10). For example, Resolution III designs offer some support for linear main effects, but no support for two-factor interactions or nonlinear quadratic effects (more on these model terms in Step 5 of this chapter). With Resolution III designs, linear effects are pairwise uncorrelated (there is no correlation between them), but each linear effect is confounded (mixed-up) with a two-factor interaction effect and we cannot tell which one is affecting our response. With Resolution IV or higher, we can distinguish between linear and two-factor interaction effects. With Resolution V or higher we can quantitate two-factor interactions. As an example, the frequently used full factorial design offers good support for linear effects and all interaction effects, but no support for explaining nonlinear cause-and-effect relationships.

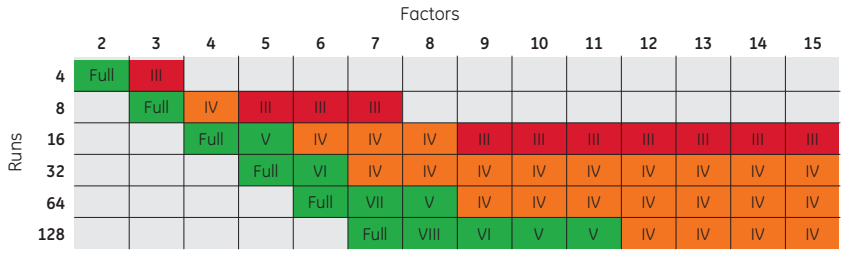


Fig 3.10. Resolution depends on the number of factors and runs. Three resolution levels are usually referred to—Resolution III, IV, and V.. Resolution III gives some support for linear effects, without support for two-factor interactions or nonlinear relationships. Resolution IV, on the other hand, gives good support for linear effects, limited support for two-factor interactions, but no support for nonlinear relationships. Resolution V, or higher, generally supports both linear and two-factor interaction effects, but does not give support for nonlinear relationships.

The reason for failing to detect the effects often relates back to a large variation in data. To find significant effects, the results should point out the direction and magnitude in relation to measurement errors and process variation.

Number of experiments needed

As shown in Table 3.1, the number of experiments required depends on the number of factors to be included and the level of detail needed.

Table 3.1. Number of runs required for some common experimental designs

		Number of runs								
		Effects explained by model			Number of factors					
Design	Objectives	Linear	Two-factor interaction	Curvature	2	3	4	5	6	7
Fractional factorial (Res III)*	S, R	■			–	7	–	11	11	11
Fractional factorial (Res IV)	S (R)	■	■		–	–	11	–	19	19
Fractional factorial (Res V)	S (O, R)	■	■		–	–	–	19	35	65
Rechtschaffner (Res V)	S (O, R)	■	■		–	10	14	19	25	32
Full factorial	S (O, R)	■	■		7	11	19	35	67	131
Central composite RSM	O	■	■	■	11	17	27	29	47	81
Rechtschaffner RSM	O	■	■	■	–	13	18	24	31	39
Box-Behnken RSM	O	■	■	■	–	15	27	43	51	59

S = screening, O = optimization, R = robustness test, Res = resolution.
** For example, Plackett-Burman design.*
Note! The level of support in a model depends on the design.

Systematic bias

Experimental investigations might be systematically biased, causing misinterpretation of the results. Bias is due to uncontrolled changes in known or unknown factors, also called "lurking variables" (Fig 3.11). As a practical example, sample material might degrade over time, causing uncontrolled changes in the series of experiments. For example, if the high level setting of a factor is performed early in the series, using fresh samples, and the low setting is performed later in the series, using aged samples, the difference in response for low and high level of the factor might be caused by a factor effect on the response or by the aging of the samples used. Hence, the effects of the factor and the age of the samples are confounded. To avoid such bias, materials and methods should be homogenous, data should be collected over a short period of time, and experiments performed in a random order.

Lurking variables are factors that we intentionally or unintentionally select not to be included in a study because we cannot control them (i.e., we have no data to enter into the experimental plan). We should, however, not exclude these factors from being confounded with the factors we are investigating. Lurking variables could in fact be confounded with our factor main effects due to the lurking variable variation ending up in the residuals.

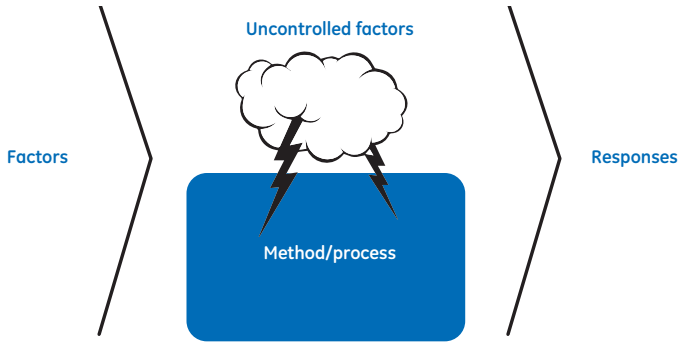


Fig 3.11. Uncontrollable factors can give systematic bias.

Avoid uncontrolled changes in the experiments by using homogenous materials and methods and by performing the experiments in a random order during a short period of time.

Order of experiments

We use our knowledge, experience, and judgment to decide what parameters to include in a study. However, it is difficult to remain unbiased when deciding in which order the experiments are performed. Randomization is the most reliable method of creating homogeneous treatment groups without involving potential biases or judgments; the effect of lurking variables can also be effectively avoided through randomization. Groups can be harmonized through randomizing the order in which the experiments are performed, so that the differences we see in the response variable can be attributed to the factors. A completely randomized experimental design is achieved when all parameter settings are randomly allocated among the experiments and is readily achieved by using any DoE software where this functionality is included.

Randomization should always be used to prevent systematic biases. First, the design is set up with systematic variation of the factor settings. The experiments and measurements are thereafter performed in a randomized order. Randomization thus reduces uncontrollable variability. Variability is very unlikely to vary with the same pattern as any of the factors. Randomization should always be used, even if all significant sources of errors have been eliminated, to ensure that no new uncontrollable factor turns up in the experiment.

A fully randomized study should always be the goal, although it is not always achievable because of the nature of the controlled parameters.

It is important to consider the use of appropriate conditions in steps adjacent to the ones that are subject for optimization of a chromatographic process (e.g., sometimes it is necessary to manually ensure that the right conditions are used in an equilibration step prior to a wash or elution step).

Replication of experiments

The tool for increasing the signal-to-noise ratio, for example, in order to quantitate the variability of the responses in a study, is replication. Replication, the repetition of experiments, helps to improve the significance of an experimental result and the quality of a model by increasing the reliability of the data for each point. If a performed experiment shows conditions necessary for affecting the response, replication will increase the credibility of the study. Thus, replication reduces the variability in the experimental results, thereby increasing their significance and the confidence level by which a researcher can draw conclusions about the study. It is important that new data is obtained by rerunning the entire experiment (experimental errors), not only the analyses (measurement errors). Replicating a series of experiments, or even the entire design, is often less expensive than to make a design with double the number of design points, and it can give a strong improvement of the quality of the investigation. Replication of an entire study gives us confidence in our results and helps us validate the process and is sometimes essential in order achieve statistical relevance in the data. Replication allows us to generalize to the larger population with more confidence.

Arranging experiments into blocks

To minimize the effect of lurking variables, we randomize designs when we can and block the designs when direct randomization is not feasible. In a block design, the experimental subjects are divided into homogeneous blocks (i.e., groups similar to each other) before randomly assigned a run number. Blocking reduces known, but assumed irrelevant, sources of variation between units and thereby enables greater control of the sources of variation. Blocking also allows us to map if any external source of variability related to the groups will influence the effects of the factors. An alternative to blocking is to keep the blocking factor constant during a study.

In blocking, we divide the runs into several sets performed group by group. As an example, consider a full factorial design, with five factors that equals 32 experiments, for investigation of the batch size of raw material or the conductivity range of the buffer used. If there are constraints that only allow you to perform eight runs per batch, we might wish to run the experiments in four blocks, each composed of eight runs using a homogeneous raw material or a single buffer system. The method of dividing, for example, 32 runs into four blocks of eight runs is called orthogonal blocking. In orthogonal blocking, each run is performed in a way that the difference between the blocks (the raw material or buffer salt) does not affect the responses. Blocking designs can readily be generated using DoE software.

When systematic bias cannot be avoided, blocking can be used to include the uncontrolled term into the model. Thus, blocking introduces an extra factor in the design and model, and the blocking variables result in reduction of the degree of freedom (see Appendix 2) and also affect the resolution of the design. The block size and the number of blocks of the two-level factorial designs are always a power of two; there is one blocking factor for two blocks, two for four blocks, and three for eight blocks, and so forth. The pseudo-resolution of the block design is the resolution of the design when all the block effects (blocking factors and all their interactions) are treated as main effects with the assumption that there are no interactions between blocks and main effects.

A blocking factor can be treated as a fixed or random effect. When the external variability can be set intentionally and the primary objective for blocking is to eliminate that source of variability, the blocking factor is considered a fixed effect. When the external variability cannot be controlled and set purposely and the primary blocking objective is to make a prediction without specifying the block level taking into account the external variability, the blocking factor is seen as a random effect.


Blocking in the UNICORN software is done manually by addition of a block factor to the design and dividing the experimental plan into the consequent blocks.


When blocking a design, additional center points divided between each block should be run.

Experimental design center points

Center points are added to a design for two reasons: to provide a measure of process stability and inherent process variability and to check for curvature. To allow estimation of the experimental error, it is common to add center points performed at least in triplicate. Center points can be excluded in resolution III designs as these designs are selected for keeping the number of experiments to a minimum. In the worksheet outlined in Figure 3.6, we have added three center point runs to the otherwise randomized design matrix, giving a total of eleven runs. True replicates, and not only repeated measurements, measures the process variation. The repetition of measurements on the same center point could result in essentially the same value for our response, thus inducing a significant lack of fit. In an experimental design, any replicated design point could be used for estimation of the process variation, not only the selected center point (if we wish to select another).

Evaluation of design quality (condition number)

The condition number, generally a software generated number based on the selected design, can be conceptually regarded as the sphericity of the design or as the ratio of the longest and the shortest diagonals of the design (Fig 3.12). Table 3.2 lists general guidelines for using condition number as a tool for evaluation of the quality of a design. The more symmetrical design, the greater the quality and the lower the condition number. Two-level factorial designs without center points are completely symmetrical as all of the design points are situated on the surface of a circle or a sphere and have condition number 1. Asymmetrical designs, such as designs with qualitative factors (or mixture factors), have condition numbers greater than 1.

Whenever a design is optimized or changed, the condition number obtained from the modeling should be checked and a value deviating from acceptable values (Table 3.2) should initiate re-evaluation of the entire experimental setup.

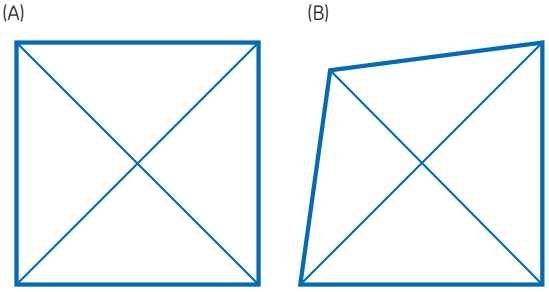


Fig 3.12. Conceptual schematic of condition numbers. (A) A completely symmetrical design has condition number 1. (B) A skewed design has a condition number greater than 1.

Table 3.2. Guidelines for use of condition number in evaluation of design quality (data from Eriksson *et al.* Design of Experiments. Principles and Applications. Umetrics Academy ISBN-10: 91-973730-4-4, ISBN-13: 978-91-973730-4-3 [2008])

	Condition number	
Design quality	**Screening and robustness testing**	**Optimization**
Good	< 3	< 10
Acceptable	3–6	8–12
Bad	> 6	> 12

The correlation matrix is another tool for estimating design quality. In the matrix in Figure 3.13, the linear correlation coefficients R between all the terms in the model and all the responses are displayed. The R value represents the extent of the linear association between two terms and ranges from -1 to 1. When R is near zero there is no linear relationship between the terms.

	LopH	LoCo	LopH*LopH	LoCo*LoCo	LopH*LoCo	DBC
LopH	1	0	0	0	0	-0.285047
LoCo	0	1	0	0	0	-0.656848
LopH*LopH	0	0	1	0.266667	0	-0.316677
LoCo*LoCo	0	0	0.266667	1	0	-0.362633
LopH*LoCo	0	0	0	0	1	-0.52746
DBC	-0.285047	-0.656848	-0.316677	-0.362633	-0.52746	1

Fig 3.13. The correlation matrix displays the correlation pattern in the data. In this case, the effect of load pH and conductivity on dynamic binding capacity (DBC) was examined. Each individual factor is correlated with itself, but we also see a slight correlation between the quadratic pH and conductivity terms (~ 0.27), and a negative correlation between the factors and the response (i.e., the DBC), which decreases with increasing factor levels.

Step 4. Perform experiments

Step 4 contains the following elements:

- Perform the experiments and analyses.
- Review the measurement system analysis and expected measurement variation as compared to the variation in the experiments.
- Check the data quality.
  - Make sure that all the experiments are well-defined, controlled, and validated.

In protein purification, the process studied can be one or several steps in the protocol, for example, for a chromatographic separation. Proceeding with this process in an efficient way requires automation.

DoE with UNICORN software and ÄKTA systems

The use of UNICORN software together with an ÄKTA system enables automatic execution of a sequence of runs following the setup of the DoE study (Fig 3.14). Based on the user-defined objective and number of factors and settings, DoE in the UNICORN software creates a designed set of experiments. The presented experimental plan contains the experiments to be performed. For minimal manual handling, the automated method events and sequential runs can be controlled automatically. Automatic control is achieved through the different modules (i.e., valves, pumps, etc.) and method events of the ÄKTA systems. ÄKTA systems offer high flexibility in configuration to meet the needs in factor and condition screening and optimization designs. Data from finalized runs are integrated in the UNICORN DoE software and can be complemented with results from external analyses for subsequent statistical evaluation. Statistical analyses include creating a model, refining the model, and displaying a number of plots to aid evaluation of the results. The model and visualized information is used to draw conclusions from the results, assist in decision making, predict responses for new parameter settings, and to optimize parameter settings for a desired combination of responses.

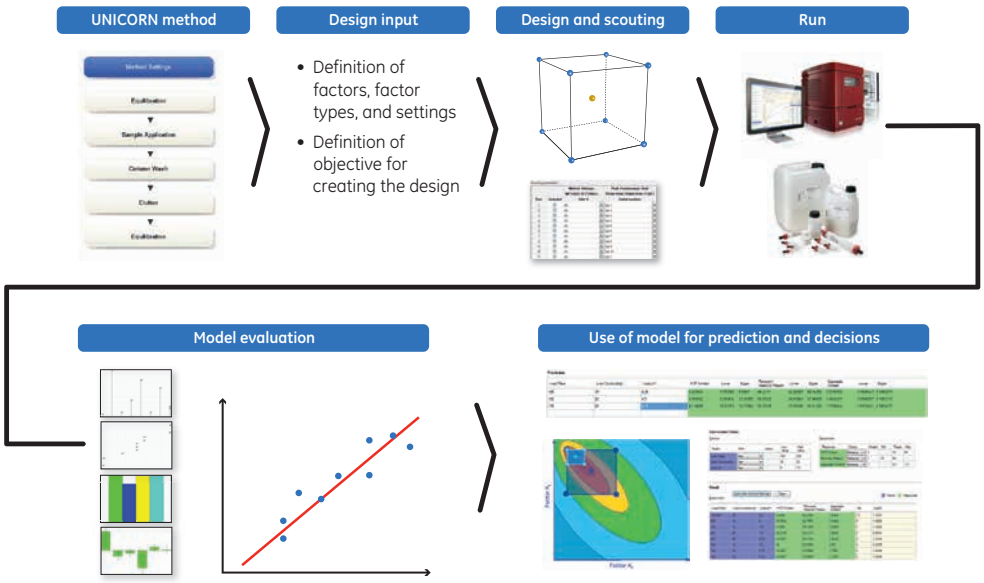


Fig 3.14. Using the UNICORN software to create a DoE workflow of which the main steps are creating a chromatography method for the process; setting up an experimental design, that is, defining the study objective (screening, optimization, or robustness testing), factors, and factor ranges; and performing the runs generated by the experimental design and entering the response values for subsequent statistical evaluation.

Configuration of ÄKTA systems

The wide variety of applications requires configurability of the chromatography system. When we wish to study different process conditions, flexibility in system configuration is important. ÄKTA systems can be equipped with valves for additional inlets and different flow directions, and with a range of columns and external equipment such as an autosampler via the I/O box. For determination of factor settings, such as buffer pH or conductivity, in a DoE study using ÄKTA avant, for example, is greatly facilitated by the possibility of using the quaternary valve for buffer mixing according to predefined or customized recipes or by using the user-defined quaternary gradient mixing ratio.

If several buffer concentrations or additional buffer components are required when using ÄKTA systems, it is possible to add an extra inlet prior to the quaternary valve.

Step 5. Create model

Step 5 contains the following elements:

- Enter the response data and use the DoE software to create a model.
- Use multiple linear regression for the mathematical modeling.

Experiments generate data that can later be transformed in to pictures for easy interpretation. In modeling, we move data to formulas that are mathematical descriptions of the relationships we are studying (Fig 3.15). DoE software is used for calculation of a mathematical model for the behavior of the process and helps reduce the effort required from the user. The model is used to investigate which factors have significant effects on the selected response. The level of the random variation (noise), a key attribute in modeling, is estimated and included in the model. To refine the model, factors that do not have significant effects on the response are deleted. A refined model can thereafter be used in predictions of responses to other factor levels and combinations not tested, allowing optimization of the process. Studies of more than one response could result in the introduction of more advanced graphs such as sweet-spot plots and a more advanced analysis (i.e., process simulations).

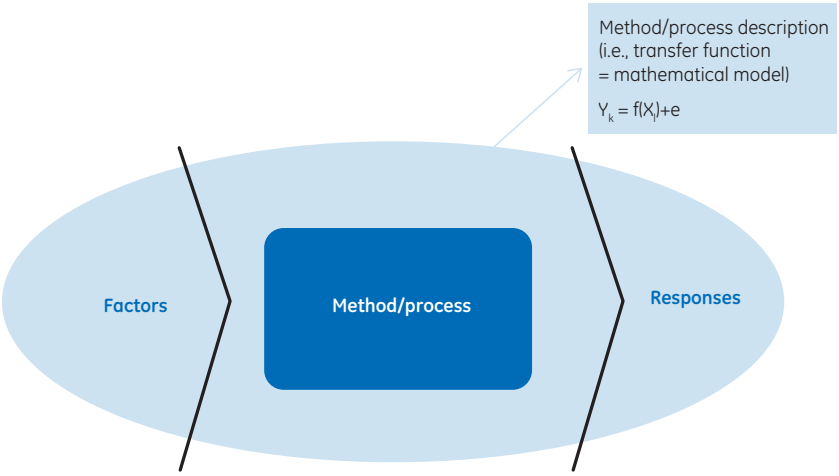


Fig 3.15. The DoE model concept. Different factors (inputs) may affect the response (output). Multiple factors and responses can be involved. The transfer function or model is the mathematical description of the cause-and-effect relationship.

To explain modeling, we use some basic calculus notations, in which f(x) represents the formula that describes the relationship between the factors and responses. This way of describing a model or transfer function is commonly used and helps us understand what we can obtain from different designs and models:

$y = f(x) + e$

Or exemplified in a linear, one-factor, cause-and-effect relationship:

$y = b_0 + b_1x_1 + e$

where

y = measured response

f(x) = function (i.e., the model) describing the relationship between factors and the response

e = residual (error) term

x_1 = factor (value)

b_0 = constant obtained at the y-axis intercept when x is zero

b_1 = the (correlation) coefficient, in this case for a linear term (main effect)

The residual term is also called error term or noise and is the random variation that cannot be explained by the model.

A simple set of data from a single response and a single factor can serve as an example of this regression analysis. In Figure 3.16, a model of the linear relationship between y and x is plotted. The difference between the observed value (data point) and predicted value on the line is termed (raw) residual. The residuals are minimized by using the least squares regression calculation.

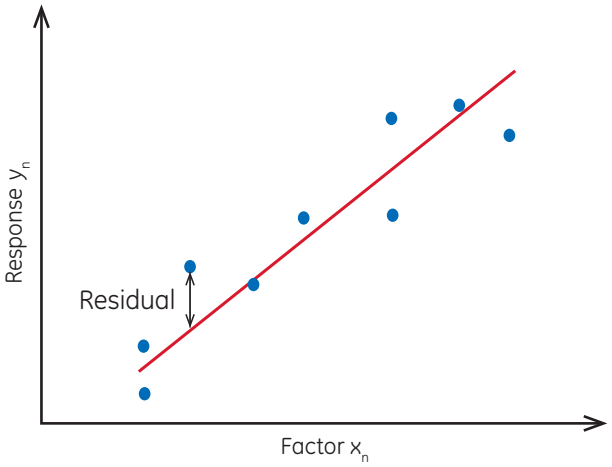


Fig 3.16. Linear regression analysis, where the red line corresponds to the model function y = f(x) + e. The residuals (e) (i.e., the minimized errors between the measured data) and the theoretical data, calculated according to the model, are indicated by the arrow between the line and the blue dots.

Signal and noise

The variation in response observed in an experimental series can be divided into systematic variability (signal) and random variability (noise) (Fig 3.17). The signal is the part of the response value that depends on the factor effect. That is, when the factor levels are changed, there will be a change in the response. The noise is additional variations in the response. The causes of the noise can be divided into model prediction error and replicate error. The replicate error in turn can be divided into errors related to the execution of the experiments (experimental error) and errors related to the execution of the response measurements (measurement error). If the noise is found to be large compared with the signals, it can be necessary to reduce the errors to obtain useful data. Experimental error can be reduced by improving the precision of the execution of the experiments, for example, by a more careful manual handling or by using adjusted equipment. While certain variables are either too difficult or too expensive to control in the actual process, it might still be possible to test or measure them using other resources and controls. Environmental conditions, for example, can be controlled to set levels for temperature and humidity during lab testing. The noise strategy involved in DoE planning includes selection of design type, factor selection and level control accuracy, type of response data, measurement of uncontrolled factors (if possible), and randomization of the experiments. We also need to take into account that there might be factors that have large effects on the response but that have been overlooked.

Fig. 3.17. The signal (systematic variability, i.e., the variance accounted for by each factor) and the noise (random variability) are both part of the measured response. Noise can be further divided into model prediction error (lack of fit) and replicate error. The replicate error in turn can be divided in to errors related to the execution of the experiments (experimental error) and execution of the response measurements (measurement error). The figure shows a hypothetical experiment with six factors x_1 to x_6.

Evaluation of the relevance of different signals compared with the residuals in the investigated process is the core of the statistical methods used in DoE. The purpose of the evaluation is to determine if the observed effect of a factor is relevant. The use of DoE software greatly simplifies this evaluation. As noise is defined as the variation derived from conducting the experiments or process runs and the variation in the analyses of the samples, this is what we are trying to differentiate from our signals. Hence, a critical step in conducting well-designed experiments is to first quantitate the total experimental error. The precision of the analysis method can be determined in a separate series of experiments (e.g., in a gage R&R study) by making repeated analyses of a single data point from the investigation. It is recommended to determine the measurement error before conducting the experiments.

- Make sure that the analysis method is sufficiently accurate for the investigation to be conducted.
- The replicates (often performed in the center point) are used for estimation of the experimental error.
- DoE resolves systematic variability (signal) from unsystematic variability (noise).

Model details

The relationship between input factors and the measured response needs to be described and displayed. These quantitative values may be arranged to display the correlation in the data, that is, the comparison of two paired sets of values to determine if an increase in one parameter corresponds to an increase or decrease in the other parameter. As illustrated in Figure 3.16, scatter plots are commonly used to compare paired sets of values. In the simplest case with linear correlations in the data, we express the relationship in a single value describing the correlation, the strength in the relationship (strong or weak) and the direction (positive or negative). The calculated value (the correlation coefficient) is a value between +1 and -1 where 0 indicates no correlation, +1 a complete positive correlation, and -1 a complete negative correlation. The greater the positive or negative value, the stronger the correlation.

Regression is the process of drawing a trend line representing the optimal fit through the data. A strong correlation has data points that all fall on the regression line. The trend could be either positive or negative. The trend line can be either linear or nonlinear. Nonlinearity could be due to a found maximum or minimum in the cause-and-effect relationship.

Multiple linear regression analysis is a technique for determining the linear relationship between one dependent response variable and two or more independent explanatory variables (e.g., yield depending on pH and conductivity). In DoE, the model usually incorporates multiple factors. Multiple linear regression analysis is commonly used for finding the transfer function and especially for estimation of the value of the β coefficients. This model is usually based on a polynomial function and a generic form is shown in Figure 3.18. The simplest model includes the constant term and linear terms. The model becomes more complex by inclusion of two-factor interaction or quadratic terms. Quadratic terms become valid in a model if we have a significant nonlinear relationship (curvature) between the factors and the response.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + e$$

Constant | Linear terms (main effects) | Two-factor interaction terms | Quadratic terms | Residual (error)

Fig 3.18. Regression model terms where β_0 is the constant term, β_n are the effects (regression coefficients), X_n are the factor values, and e is the residual term.

As data visualization is about creating images from numbers, the concept of modeling could be viewed as creation of formulas from data, which are then described visually. Modeling is one of the most fundamental aspects of DoE and all related software use modeling in the statistical data analysis. Modeling is a highly mathematical operation. However, we do not need to perform all calculations by hand. The available DoE software packages help us perform the regression analysis. What we need to understand is what the different plots and values tell us and to determine if the data is likely to make justified assumptions about the results. The results should make sense and be credible. For example, the regression coefficients that we obtain from our analysis are estimates of the positive or negative impact of changing the settings of our controlled factors on our response.

Additionally, we have to question the outcome and use our experience and knowledge to determine if the results are probable.

When performing factor screenings, we are principally interested in the main effects. We need to know which of the included factors are the most important. In most cases, we can assume that higher order effects such as two-factor interaction effects are negligible. In optimization studies in biotechnological applications (e.g., protein purification) and in most other industrial applications, there is no need for higher degree of polynomials than quadratic terms. Even higher degree terms are unlikely to become significant in an analysis. The design selection, and thus the possible model for us to use, is also important and depends on our factor-response relationship. If we expect the region of interest to be small enough to be described by a linear polynomial, we can select a design with only linear terms (Fig 3.19). For a larger region of interest, including a curvature effect, we might have to select a design that allows a quadratic polynomial to be fitted to the data. Describing very large areas in the cause-effect relationship is not realistic and is not generally needed in biology. Descriptions of large areas would also require a more complex model and too many experiments to be feasible.

Good predictive models that include main effects (first-order terms) as well as quadratic and two-factor interactions (second-order terms) are most often adequate enough to enable prediction of the behavior and optimization of the process. Higher-degree terms are unlikely to become significant in a statistical analysis.

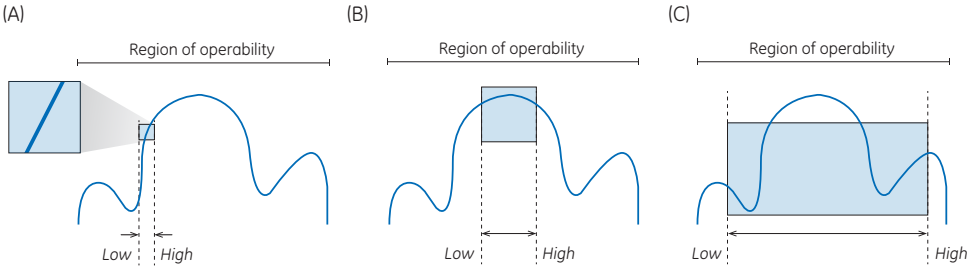


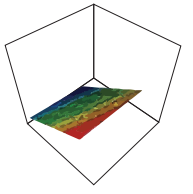
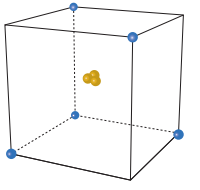
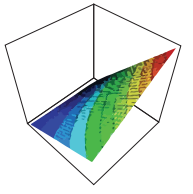
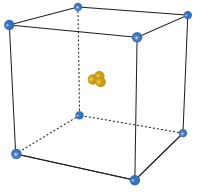
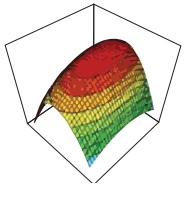
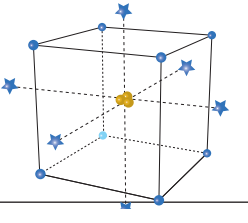
Fig 3.19. (A) For a small region of interest, a linear function can work well. (B) For a larger region of interest, a quadratic function will be suitable. (C) For a very large region of interest, accurate modeling of behavior might not be feasible with standard regression models.

Table 3.3 gives an overview of the relationship between experimental objectives, models, and designs.

The choice of the investigated factor region is crucial, as all experimental designs are local descriptions of a complex reality. If the region is not too large (and not too small), simple models including main effects, first-order interactions, and possibly nonlinear effects should work well.

Good process knowledge is necessary for performing good DoE.

Table 3.3. Relationship between experimental objectives, models, and design

Objective	Model description	Conceptual sketch	Design geometry	Design examples
Screening **Robustness testing**	**Linear polynomial model** $Y = \mathbf{b_0 + b_1x_1 + b_2x_2 + b_3x_3} + e$ The linear terms (bold) describe the main effects. In the graphical illustration, this part of the model can be viewed as an undistorted plane. The model gives an overall idea of the position (or direction to) the optimum for the process, but gives no details on how the sampling plane can be twisted or curved. The model usually gives sufficient information when the objective is screening or robustness testing. Fractional factorial designs will give enough input to create the model.	**Undistorted plane** 	**Fractional factorial designs** 	Plackett-Burman design (usually Res III) L-designs (Linear, quadratic, not all interactions)
Screening	Two-factor interaction polynomial model $Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \mathbf{b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3} + e$ The two-factor interaction terms (bold) describe how the effect of one factor depends on the level of a second factor. The model gives indications on twists in the plane. The model can be used when the objective is screening. Fractional and full factorial designs will give sufficient input to the model.	**Twisted plane** 	**Full factorial designs** 	Two-level full factorial design Rechtschaffner screening design (resolution V) Fractional factorials (resolution V)
Optimization	Quadratic polynomial model $Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 +$ $\mathbf{b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2} + e$ The quadratic terms (bold) describe any curvature of the sampling plane. These terms are added to the model when the objective is optimization. The optimization designs will give enough information to create the model.	**Curved plane** 	**Composite designs (RSM designs)** 	Three-level full factorial design Central composite circumscribed (CCC) design Central composite face-centered (CCF) design Rechtschaffner RSM design D-optimal RSM design Box-Behnken design Doehlert RSM design

Step 6. Evaluation of data

Step 6 contains the following elements:

- Analyze and visualize the results in plots and graphs.
  - Analyze raw data from the results.
- In the DoE software, perform analyses using computer modeling and statistical tests.
  - Display statistical plots and graphs to explain relationships and dependencies.
  - Use statistical methods to assure significance and interpretability of the results.
  - Evaluate each response individually and display significant factors (e.g., effect magnitude and p-values).
  - Review if there are any parameter interactions and what they mean.
  - Review if there are nonlinearities in the data and what that means.
  - Review the model(s) goodness of fit (R-square, Q-square, model validity, reproducibility).
  - With multiple responses, present global optimum and optimum range of each factor.
- Verify your results.
  - Verification runs should always be performed (i.e., check the results by experiments performed within the design space, with confirmatory runs and comparisons with the final process).
- Summarize the conclusions.
  - Data reports should include descriptions of statistical considerations (e.g., the sample determination and size), if (and which) data was excluded, any manipulations of data (e.g., transformations), effect sizes, and confidence intervals.
- Make recommendations for the next step, which could include follow-up on certain parameters and additional optimization studies.
- Implement, or optionally update, current process.
  - Add checklist, trend analysis, and data from failure modes and effects analysis (FMEA), for improved process control.
- Use the model.
  - The created model can be used for anticipation (prediction) of the outcome of runs using other parameter settings than used in the experimental design.
  - Process simulations (Monte Carlo) can be used for assessment of the effect(s) of random variability on the defined model.

Data evaluation is a highly iterative process. As soon as we start data visualization and evaluation, we will find that more questions often need to be answered before making any final conclusions and decisions.

DoE software will create a model, but it is critical to evaluate the significance and the relevance of that model.

In addition to the actual design of the experiments to be conducted, DoE also entails a systematic way to interpret results. Understanding the significance of these differences is the key to understanding the cause-and-effect relationship in a process, that is, why/how it is happening. As in all experiments, we have to consider several dimensions in the evaluation process:

- Are the factor levels appropriate?
- Are there correlations between the factors?
- Do the factors have a significant effect on the response variables?
- Are there significant interactions between factors?
- Are there nonlinearity in the data?
- Is there and how large is the uncertainty?

Most DoE software provide a systematic way to handle these questions and a statistical evaluation of data (Fig 3.20).

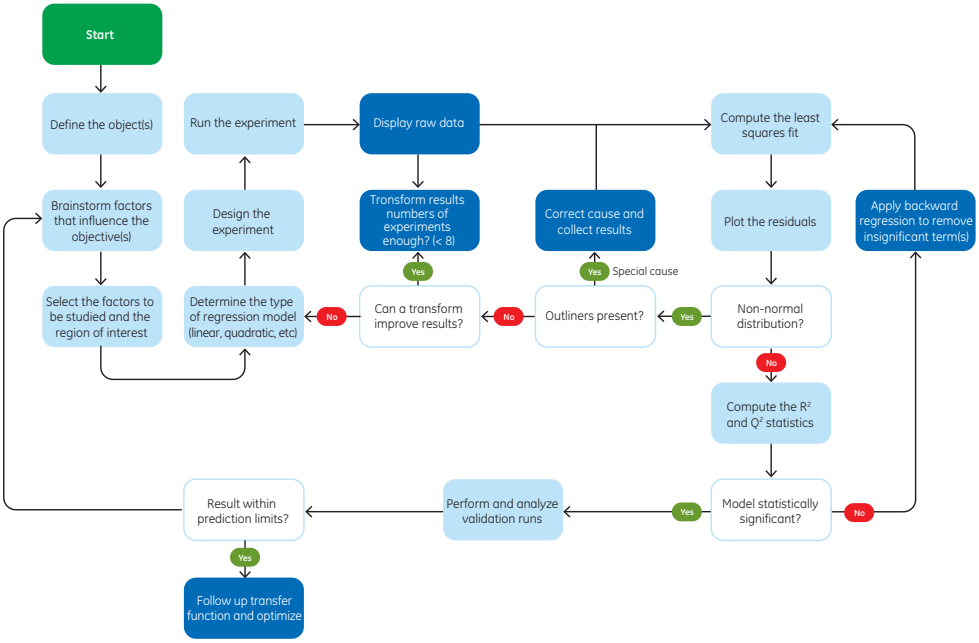


Fig 3.20. DoE data evaluation workflow from a statistical perspective.

Figure 3.21 shows a simplified description of the data handling and systematic evaluation process for data obtained from a DoE study. In the initial step, we plot raw data. A general rule for plotting data and also for displaying statistical results is to try to simplify the graphics although being conceptive in the design. Plotting and replotting raw data helps in data clean-up, and spotting of deviating values and patterns or abnormalities in order to speed up the subsequent statistical analysis and modeling step. The second step in the workflow includes regression analysis and model interpretation. In this step, our attempt is to derive and predict the most relevant model, to find out where there might be an optimum and what the factor settings are at that point. Residuals are used for determination of the validity of the linear regression. Residuals are also used for determining if the obtained results are independent of each other and if the responses have an equal variance for all factor values. Determination of the validity of the linear regression analysis is also performed by looking at the residuals. Assessing if residuals are normally distributed is important, as we also assume a similar behavior for our responses. If we at the end conclude that the model is correct, we try to reveal if the model makes sense from a logical perspective and if we are able to interpret the model.

Questions we ask are:

- What does the impact of the model mean?
- Where should we position new experiments?
- Where is my optimum?

The model needs to be verified before making predictions. In process development, we are likely to perform a robustness test around the optimal or most interesting region.

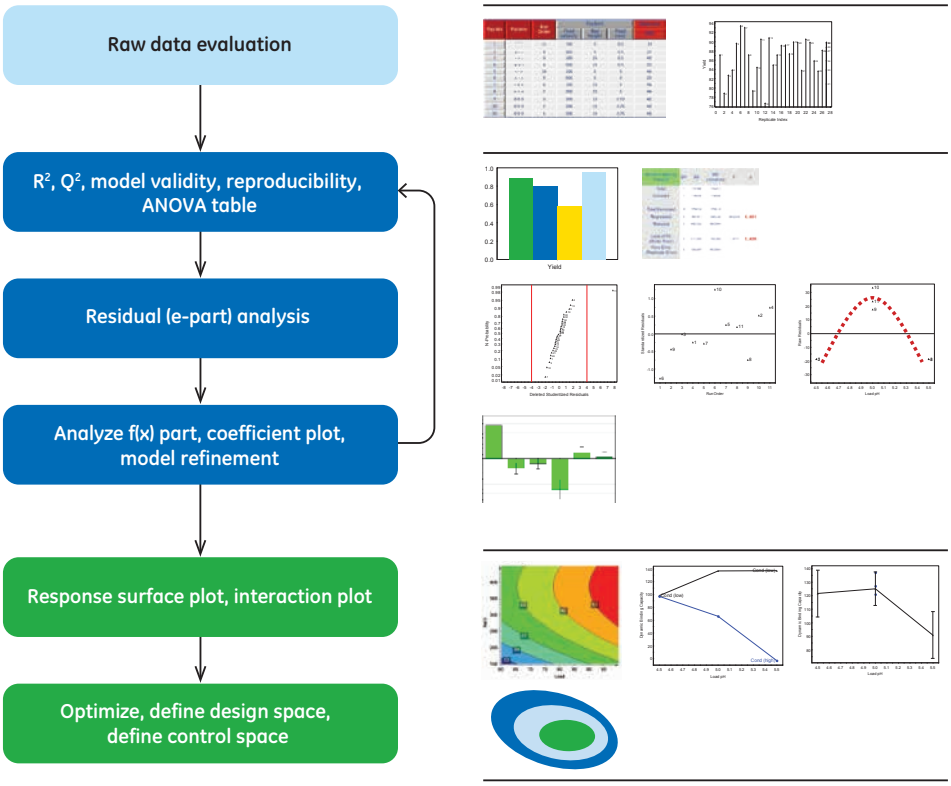


Fig 3.21. Any data evaluation is highly iterative and needs repeated user interactions during the process. In this workflow, we review raw data (Step 1), perform regression analysis and hypothesis testing as well as check the quality of our model (Step 2), and use the model for decisions and conclusions (Step 3).

Our questions (hypotheses) about the process

A hypothesis or question states the aim for a given process, for example, how to increase yield or purity. A hypothesis test investigates if the obtained response data is due to our effectuated change in factor settings. In statistics, the null hypothesis (H_0) states that the change in a process is uncorrelated to the change in the controlled parameters. The alternate hypothesis (H_a), on the other hand, states that the change in a process output is due to the change in the factors. Classic statistical tests give an indication on how well obtained data support the hypothesis by trying to disprove the null hypothesis. The alternate hypothesis is assumed being wrong until we find evidence for the contrary.

There are some basic steps for hypothesis testing of which some are similar to the strategy of setting up and performing a DoE study and some are purely statistical:

- Problem definition
- Objective statements
- Setting up the hypothesis
  - State the null hypothesis (H_0)
  - State the alternative hypothesis (H_a)

- Select the appropriate statistical test
- State the alpha-risk level
- Establish the effect size
- Create a sampling plan and determine sample size
- Collect samples
- Perform runs and record data
- Statistical analysis, calculating test statistics
- Determine p-value (probability)

If our statistical analysis reveals < 0.05 or a value equal to the alpha risk (see below) we reject the H_0 and accept H_a, and if $p > 0.05$, we do not reject the H_0.

The null hypothesis is technically never proven true. The null hypothesis is either "failed to reject" or "rejected". "Failed to reject" does not mean to accept the null hypothesis, as it is established only to be proven false by testing the sample of data.

Repeated hypothesis tests on specific data will ultimately result in finding a spurious effect, that is, an effect that is not real. What this means is that we should define a limited set of questions and use statistics to investigate them instead of pursuing with data analysis until a significant result is found.

It is possible to find unexpected patterns in data that are not based on any predefined hypotheses. However, the patterns can only be used for forming a new hypothesis to be tested with new data.

Analyzing model quality

Analysis of variance (ANOVA), in the context of this handbook, uses variance to investigate and statistically test experimental data. The first question we try to answer is if the obtained regression model is significant. To answer this, we compare the variances of the model (regression) and the residuals (error). If the obtained F-ratio (F-test, F for Fischer) is high, due to a much larger regression variance compared to the residual variance, we have a significant model. The second question is if we have a significant lack of fit (model error). Lack of fit can, for example, be analyzed in an F-test. In such case, the F-test compares the model error variance (residuals, replicate variation excluded) with the pure error variance (replicate variation). If the model error is approximately the same size as the pure error, we will get a low F-ratio between the two variances, and a high probability (p-value) that the model error and the pure error are not different, which would indicate that we do not have a lack of fit in our model. If, on the other hand, the model error is much larger than the pure error, the F-ratio will increase, the probability (p-value) will drop and we will have a statistical lack of fit. See Chapter 6 for more details on ANOVA calculations. With the third question, we try to answer if our factors are in fact related to the response variable (Fig 3.22).

ANOVA is commonly used in conjunction with DoE.

A probable cause for a significant lack of fit is a missing model term.

The sum of squares is defined as the squared distance between each measured data point and the predicted value, obtained from the model.

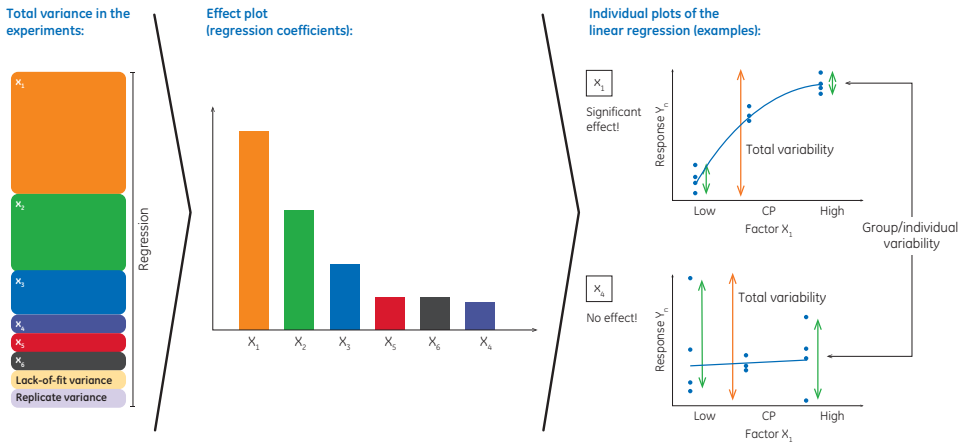


Fig 3.22. Linear regression analysis and ANOVA are essentially the same, and the included experiments possess a total variability of which each included parameter accounts for their specific part of this variability. In other words, we have a total study variability, and the included factors in the linear models account for parts of the total variance. Thus, the outcome variance can be related to the factors and the unexplained part (the residuals). Using ANOVA, we calculate the statistical significance of the variance accounted for by each factor of interest versus the total variance.

P-values

The p-value (probability) can be viewed as the probability that result properties, such as differences or relationships in data, are caused by chance. A significant (low) p-value summarizes data assuming a specific null hypothesis. If the p-value is significant, the results are most probably not caused by chance, indicating that there is an underlying relationship in the data. However, the p-value does not take in to account that a real effect was actually present. It only reflects the obtained data and aims to reject the null hypothesis. Usually the statistical significance level is set at 95% (alpha = 0.05), which means that the null hypothesis has one chance in twenty of being true and that we make the incorrect interpretation 1 out of 20 times. When evaluating experiments we should look at the size of the measured effects and the confidence intervals of the data and see if they are supported by a significant p-value.

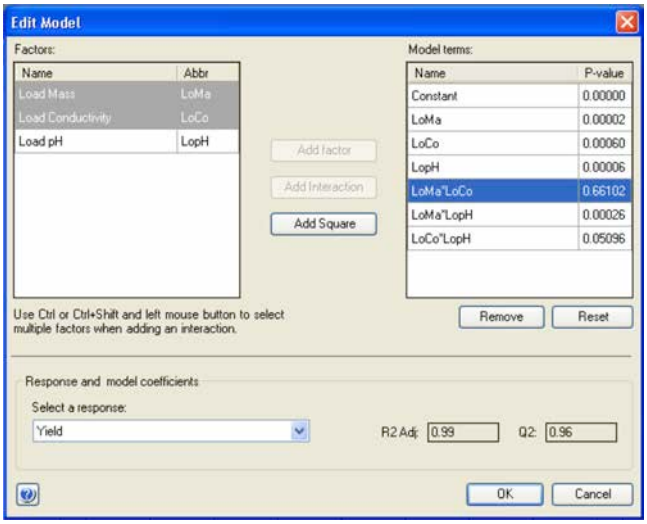


Fig 3.22. In the ***Edit model*** dialog in the UNCORN software, we use p-values to exclude insignificant model terms. When the term is removed from the model, the adjusted R^2 (***R2 Adj.***) and Q^2 (***Q2***) values are updated. If the refined model gives a higher Q^2 value, the refinement is justified.

- A response effect with $p > 0.05$ (> 5%) is referred to as not significant at that significance level. In this case, a significant effect refers to differences or relationships in statistical data analysis with a p-value less than 0.05. However, $p > 0.05$ is not a true evidence that the effect is not significant. The failure to detect a significant effect might just mean that the effect is small, that there is a lot of variability in the data, and/or that the sample size is too small.
- When we have large sample sizes, even small effects can be significant. The magnitude of the effect has to be viewed in the context of importance.
- P-values should always be presented together with other statistical numbers such as the effect size(s) and confidence intervals.

Lack of fit

Generally, a significant lack of fit is obtained when a model is missing a significant term, that is, if we have a process or system with a strong nonlinear relationship and a quadratic term is not included in the mathematical model. A lack of fit can also be obtained when having too many terms, that is, we have included insignificant terms in the model or if there is an outlier in the data. Thus, we have both a significant lack of fit test and an outlier test. Additionally, lack of fit can be more of a statistical problem than a real problem. If the number of replicated measurements is low (three to six) and measured values are similar, for example, we might obtain a lack of fit without there being a real problem with the process.

The differences between the actual and predicted values (the residuals) are made up of two components: pure error (experimental error based on the replicated experiments) and lack of fit (model error). The challenge in developing an adequate model is to find statistically significant predictors so that the lack-of-fit part of the residual is not significantly different from the pure error part. The pure error part is caused by the (random) variation between the center point runs (the true replicates) and not between repeated measures of individual trials.

The first thing we will need for a lack-of-fit test is a measure of how far away, in average, the model predictions are from the actual results (the lack of fit). If the model contains all the significant effects, the differences between predicted and actual values will only be due to experimental and measurement variability. The second thing we need for a lack-of-fit test is a measure of the experimental and measurement variability (the pure error), which can be obtained from any repeated experiments in the design. With the F-test, we can determine if the lack of fit is significantly larger than the pure error. Thus, a p-value of less than 0.05 in the lack-of-fit test is indicative of that we are missing a significant term (e.g., quadratic) in the model.

Confidence level—how sure can we be?

The confidence level and the estimation of the confidence interval when performing statistical analyses fall back on estimating the risk of making wrong decisions based on the obtained results, or in simpler words “how sure can we be?” The confidence level is expressed as a percentage and represents how often the true percentage of the population would be an outcome within the confidence interval (Fig 3.23). Usually, we report point estimates, such as means, but it is also important to indicate the amount of uncertainty in the data, for example, by showing confidence intervals. We usually use 5% as the confidence (or significance) level of hypothesis tests, meaning the risk of making the wrong decision is 5% or 1 out of 20 times. Another way to express this is to say that the process and the altered process parameters have 95% probability of producing an interval that includes the observed value. The confidence interval (also called margin of error) is the often reported plus-or-minus figure telling how sure we can be that if we have tested the entire relevant population, the outcome would fall within that interval.

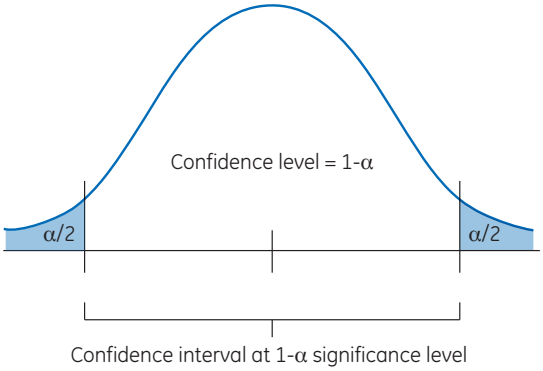


Fig 3.23. The relationship between alpha, confidence level, and the confidence interval.

Confidence interval calculations assume having a genuine random sample of the relevant population. If the sample is not truly random, intervals are not reliable.

A wider confidence interval means we can be more certain that the outcome from the whole population would be within that range.

Alpha-risk and confidence level

In statistical analysis, the alpha-risk gives us an understanding of how large the risk is that the conditions studied do not have an effect. The alpha-risk is the risk of incorrectly deciding to reject the null hypothesis or the risk of making the wrong decision. The confidence level is defined as 1-alpha. If the confidence level is 95%, the alpha-risk is 5% or 0.05, meaning there is a 5% chance that the observed variation is not actually reliable. In statistics, making an incorrect decision based on the alpha-risk is also called the alpha-error (or false positive or type I error). Generally, the confidence level (alpha) of a test is between 1% and 10% but can be any value depending on the desired level of confidence or need to reduce the alpha-error.

The used confidence level (i.e., the alpha-level) for quadratic terms in a model is usually the same as for main effects.

The confidence level, the risk we are willing to take in deciding which model terms to include, does not change the term's estimated coefficient (magnitude or direction) or variation.

The confidence level can be set based on the consequence making an incorrect decision. When the consequence increases, the risk decreases, and vice versa. In practice, for example, to keep the factor number of low in screening designs, used confidence level is often 0.10 rather than 0.05.

Outliers in the data

Outliers are results or experimental values that do not fit and are considered to be exceptions from the normal, that is, values that fall outside the quantitative range where most of the other values are located. When obtaining deviating values, we usually determine their relevance in relation to the process we are working with. If we have not obtained similar values in previous studies, the initial action would be to replicate the experiment to see if the deviating values remain. If the replicated experiment generates values more in line with previous data set, the action would be to exclude the previous value from the model. However, we always need to document, explain, and justify the action. The outlier could, for example, be due to measurement error where the solution would be to remeasure the data point. The outlier could also be due to an experimental (process) error where we would rerun the experiment, or a model error where we would need to build a more complex model (add terms).

The number of data points is important when evaluating an outlier. If we have a reasonable number of tests, removing the outlier that deviates largely from the rest of the data will not be questioned. However, if you have a large number of data points in a study, it is easier to accept that a few outliers remain in the model. A deviating value can be indicative of having a incorrect model, but if all other data points are reasonable and fits the model, a single value should not compel us to discard the rest of the data. In an outlier test, we try to find out if the result is significantly different from the rest of the data used in our model. The residuals are calculated from our measured, y_i, and predicted, $\hat{y}_{(i)}$, response values:

$$e_{(i)} = y_i - \hat{y}_{(i)}$$

where $e_{(i)}$ is our i^{th} prediction residual.

When normalizing a residual by dividing $e_{(i)}$ with the standard deviation S, this is referred to as the standardized residuals, d_i:

$$d_i = \frac{e_{(i)}}{S}$$

where S, the standard deviation is defined as:

$$S = \sqrt{\frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_{(I)})}{n - p}}$$

where is n the number of runs and p is the number of terms in the regression equation.

If we find the value to be unlikely, we eventually may exclude it from our model and repeat the modeling using the remaining data points. Removal of data points from our results, however, needs justification and thorough investigation. In Figure 3.24, we can easily detect an eventual outlier in the normal probability plots. The outlier is deviating from the normal probability line and has a large absolute residual value. When using deleted studentized residuals, a potential outlier has a d-value more than four standard deviations from the mean (indicated by red lines in the plot). Therefore, the i^{th} point is a potential outlier if $d_i < -4$ *or* $d_i > 4$. Other statistical tests for detecting outliers are Grubb's test (large sample set of 10 to 150), Tukey's hat, and Cook's distance.

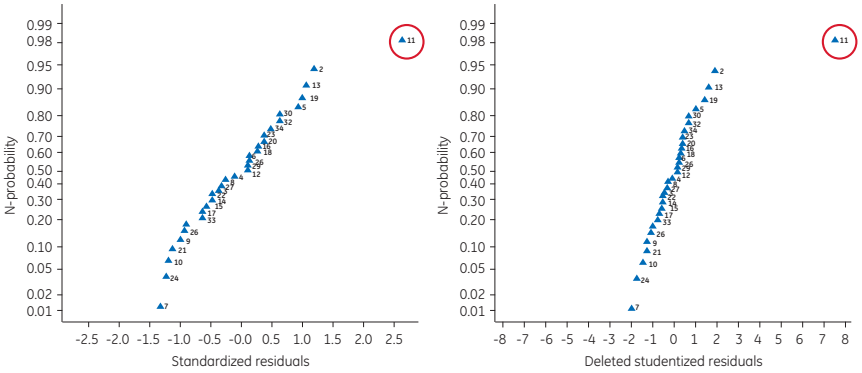


Fig 3.24. Normal probability plot of residuals when a suspected outlier is present. The left graph uses standardized residuals, whereas the right plot uses deleted studentized residuals (i.e., the outlier is truly predicted by the way the residuals are calculated, the suspected outlier is excluded in the calculations and has a predicted value).

The used regression model should be sufficiently robust so that excluding an obtained outlier from the model would not change the outcome of the analysis. However, it cannot be stressed enough that we always need to verify our results by performing confirmatory experiments based on the model predictions. Outliers are sometimes what bring us forward in learning about our process and allow us to make new discoveries that might show up as profit-making opportunities. If a data point really was improperly produced, it can be removed, or else, we need to keep it. Table 3.4 lists how to handle various outliers.

Table 3.4. Data deviations and how to handle them

Why outlier?	What to do
Bad replicates	Check the individual results, and that correct response values have been entered. If the run has failed, consider repeating the experiment and replace the run.
Deviating experiments	Check that the correct response values have been entered. Check the individual results. Consider performing new experiments to verify the deviation. If the results are indeed valid, the model may be inappropriate for the application.

Confounding—failing to see the difference

Variation can originate from many sources. All possible sources of variation (e.g., factor settings, uncontrolled variables, response measurements, etc.) need to be identified and controlled as far as possible. The simplest form of control is comparison. By comparing data, we can avoid confusing the effect of one factor with other influences (confounding or aliasing; Fig 3.25). Confounding data occurs when the effect of two or more factors (on a response variable of interest) cannot be distinguished from another effect. Confounding is the statistical terminology that describes when the unique effect of a factor or interaction cannot be separated from the unique effect of another factor or interaction.

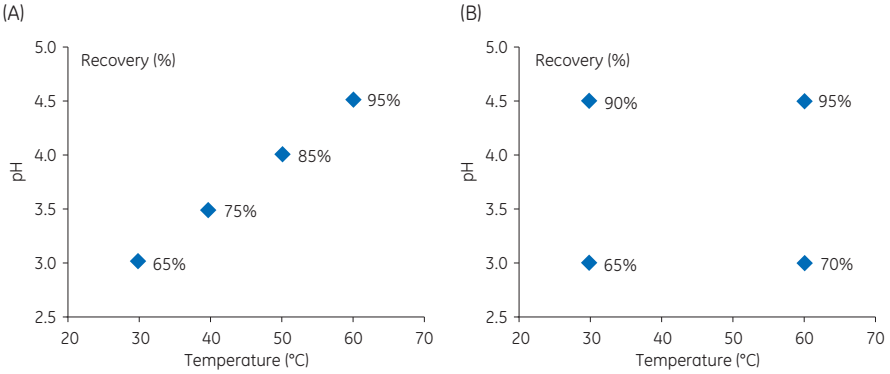


Fig 3.25. Example of confounding data. (A) It is impossible to conclude whether the effect on product recovery is caused by pH, temperature, or both. (B) Temperature and pH are not found to be confounded and it is possible to see that the recovery is affected by both pH and temperature, and that pH contributes most.

Depending on the chosen experimental design, some model terms can be completely aliased or have a high degree of correlation to other model terms. The correlation matrix is a detailed summary of the correlation patterns for different experimental designs, which is especially important if runs have been excluded from a design or if the used factor settings were not perfectly met. The condition number of a design is a summary of the correlation structure in the experimental plan.

For most screening purposes, the use of resolution IV designs is recommended, as the main effects are not confounded by two-factor interaction effects.

Chapter 4
Visualization of results

The information we obtain from experiments needs to be visualized to facilitate decision-making and to draw conclusions (Fig 4.1).

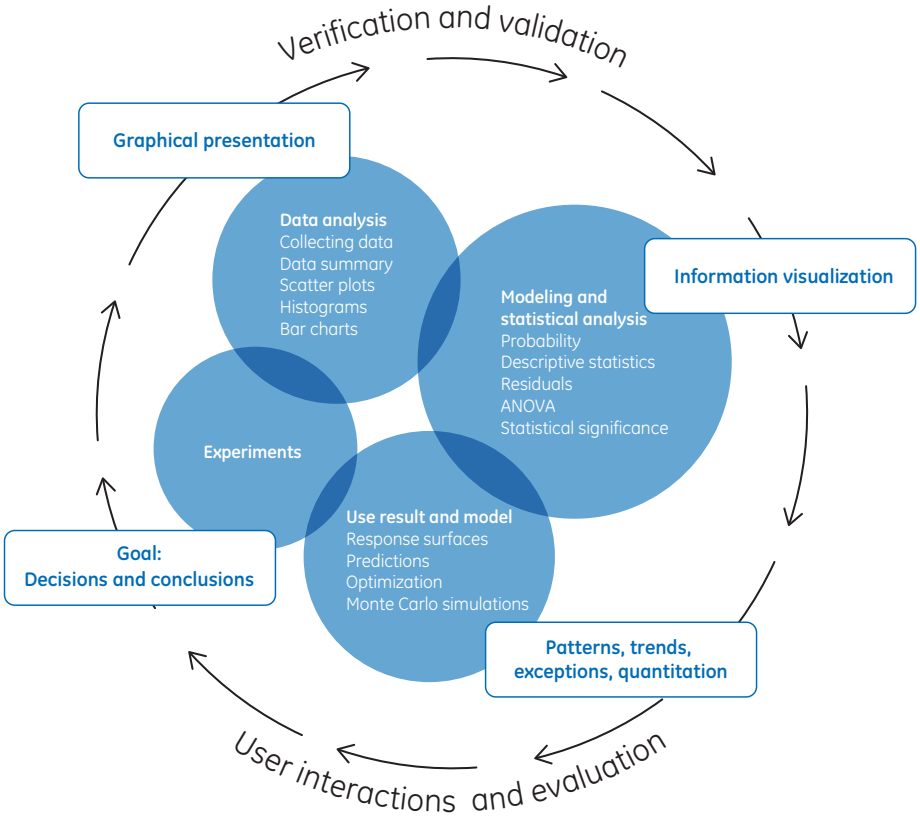


Fig 4.1. Brief description of the route from experiments to visualization and conclusions. The purpose of data evaluation is to detect response effects, induced by changes in process or system parameters. Visualization of these effects facilitates decision making. Besides the actual experiments and analyses, user interactions include: entering data, plotting data, replotting data, sorting, filtering, adding or removing variables, formatting plot layout, defining requirements for the statistical analysis, and editing the model.

Using a defined set of tools for general data analysis makes it easier to get a broad awareness of the results and to spot deviations that often become subject to a more detailed analysis. Figure 4.2 provides a striking example, where a graphic representation of data reveals relationships that are easily overlooked in a tabular format.

For many, the first contact with DoE is a colorful response surface plot. However, there are several other visualization tools in the DoE software that provide equally important information. In this chapter, some of these visualization tools are described.

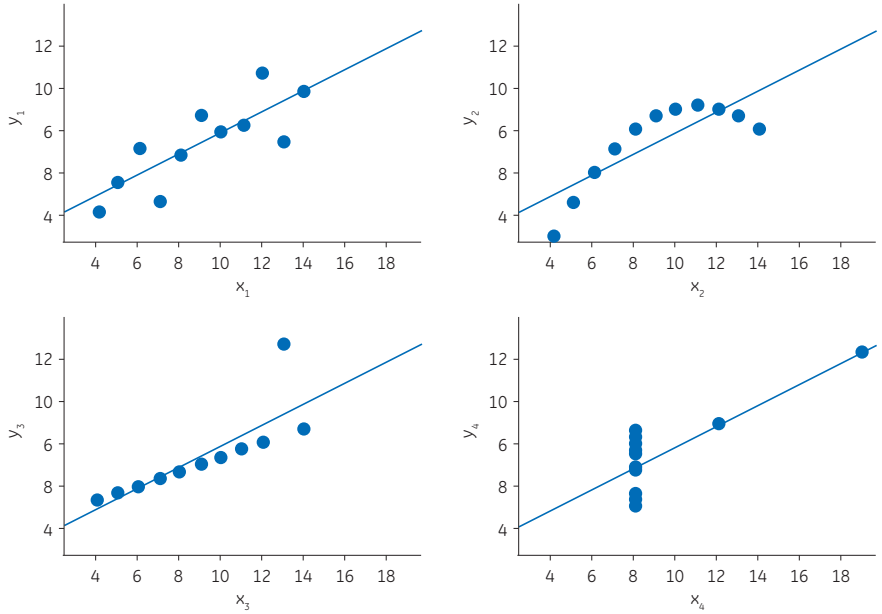


Fig 4.2. The four datasets, consisting of eleven (*x, y*) points, all have the same mean (x, 9; y, 7.5), variance (x, 11; y, 4.12), correlation coefficient (0.816), and regression line (y = 3 + 0.5x). The differences between the results, when visualized, illustrate the importance of viewing data graphically before data analysis and the inadequacy of basic statistical properties for describing realistic datasets. Modified from Anscombe, F. J. Graphs in Statistical Analysis. *American Statistician* **27**, 17–21 (1973).

Including statistical evaluation in the data analysis procedure allows quantifying how uncertain we should be about the obtained data and, ultimately, about our decisions and statements. We need to understand that the data is simply a description of the process or system of interest and need to be validated and verified.

The main steps of performing a statistical evaluation of the results from a DoE study using the UNICORN software are outlined in Figure 4.3. The evaluation steps are created to provide structure to this process. When we perform extended data analysis, we need to review both the raw data and the basic analysis plots.

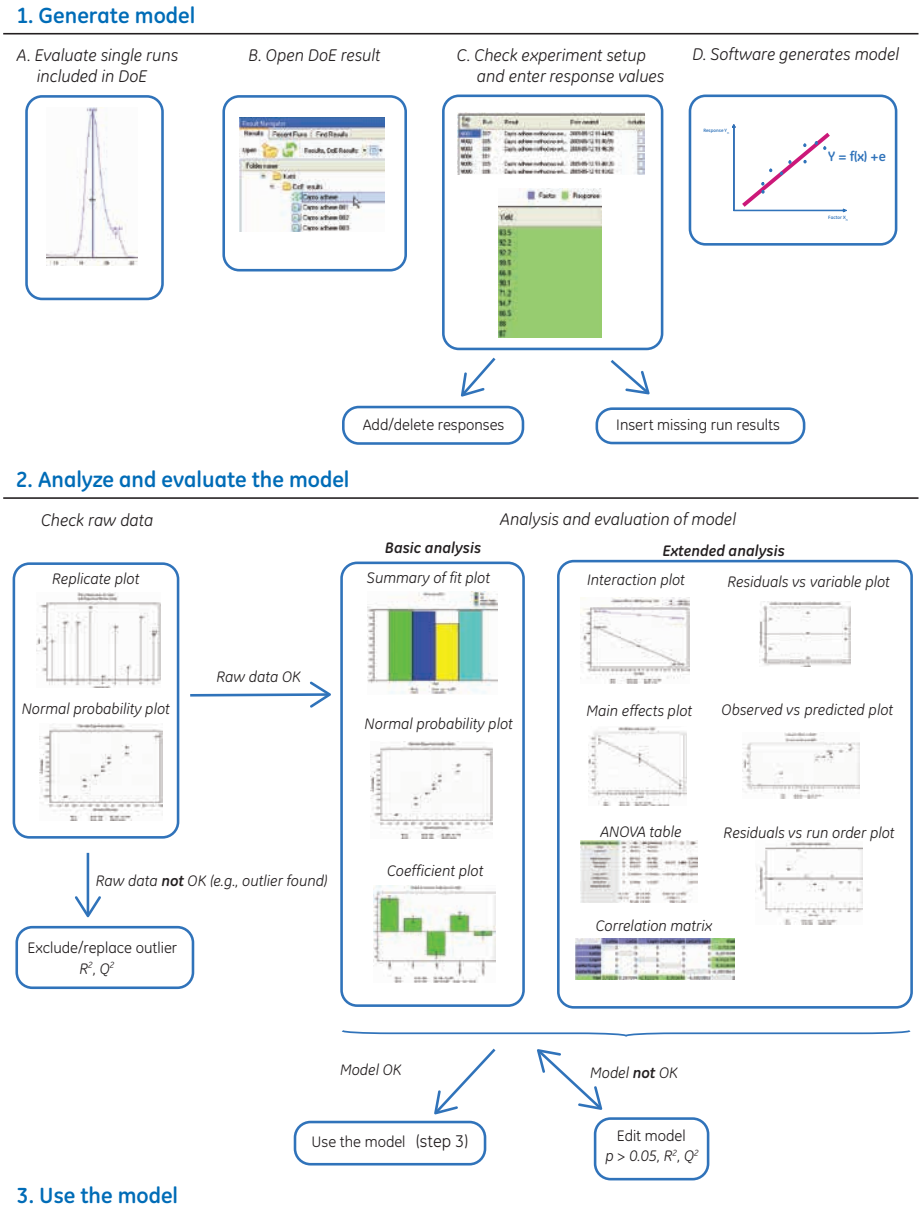


Fig 4.3. A range of visualization tools are included in the UNICORN software. In Step 1, the software calculates a model based on entered response data. The preliminary graphics of the statistically analyzed model are visualized in Step 2, for user judgment and model refinement. In Step 3, the information of the obtained model is used for decision making, drawing conclusions from our observations, and for making predictions.

Replicate plot

The replicate plot shows the variation among the replicates, our center point results, in relation to the variation across the entire experimental design (Fig 4.4). This plot is a useful tool for getting an understanding of the data before moving further to the actual statistical evaluation. Is the replicate variation probable? Is the response target range covered within the investigated experimental space? Do we seem to have any unlikely values? Do we see any indication of a second degree curvature?

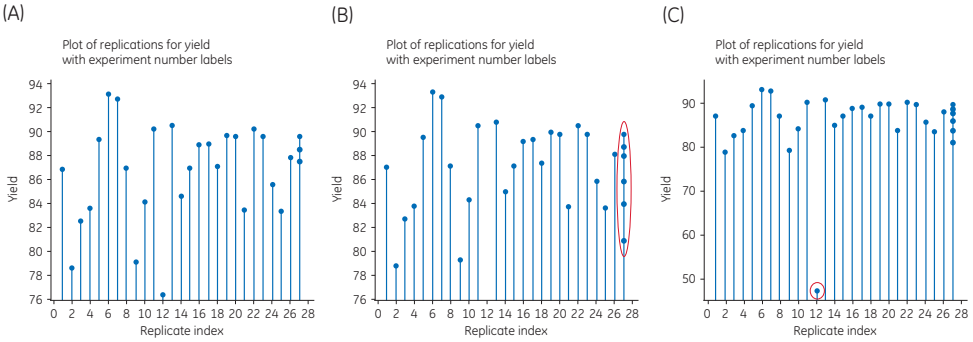


Fig 4.4. The replicate plot shows the variation among the replicates (the bars with the highest number [replicate index] of each plot represents the repeated center points) in relation to the variation across the entire experimental design (A, B, and C). In A and C, the replicate variation is much smaller than the overall variation. In A and B, we can also see that the response target range is covered within the investigated experimental space, which is not the case in C. In C, one of the experiments is deviating from the rest and could be a possible outlier. In A, we can see that the replicates are on the higher end of our response value, which is indicative of a curvature. In a linear relationship, we would expect the center point values to have a response value half-way between the low and high setting.

When a robustness test has been performed, the response variation is expected to be small and within our specification limits.

Normal probability plot of residuals

To review the normality of the data, we look at the normal probability plot of the residuals, that is, the minimized distance between the measured and the predicted data, calculated according to the model (Fig 4.5). If residuals are normally distributed, the residuals should be distributed along a straight line in the normal probability plot. A deviating pattern, s-shaped curve or grouping of data, occurs if there is a systematic variation in the residuals. This indicates either that the process is more complex than our current model (i.e., a nonlinear cause-and-effect relationship between our factor and the response is detected and the quadratic model term is not included) or that an unknown factor is affecting our results (e.g., the presence of insignificant terms).

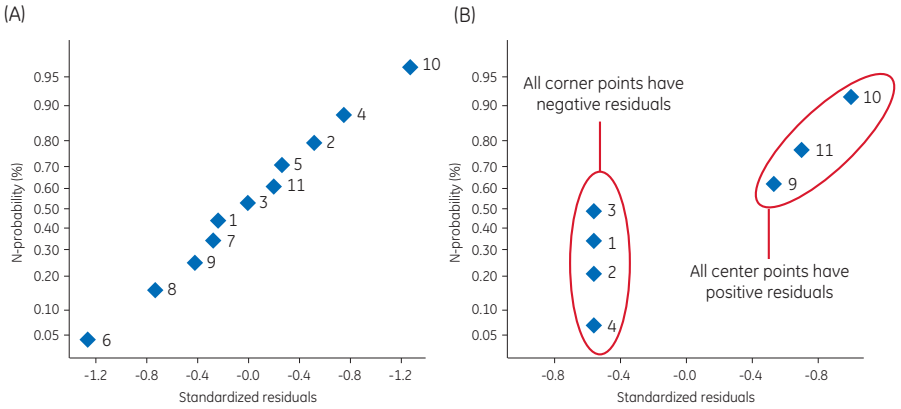


Fig 4.5. Normal probability plots. (A) normal data and (B) data with a clear pattern showing that quadratic terms are missing in the model.

- The residual standard deviation (RSD) shows the variation of the response that is not explained by the model, adjusted for degrees of freedom, and stated in the same units as the response.
- The RSD is lower near the center of the investigated experimental space and increases towards the outer limits of the factor levels.
- RSD is used in the determination of confidence interval for coefficients and predictions.

Summary-of-fit plot

How well did we do our modeling? Figure 4.6 shows an example of a summary-of-fit plot including four model statistics: the R^2 value, the Q^2 value, the model validity, and the reproducibility. The R^2 value gives a measure of how much of the overall data variance the model can explain. In general, R^2 values greater than 0.75 are considered acceptable. The Q^2 value is a measure of how well the model will work for future predictions. The model validity is a value representing the lack of fit (a low value indicates that the model suffers from lack of fit). The reproducibility compares the repeatability variation (replicates) with the overall variation (rest of the data).

If we exclude the star points from the data set used in Figure 4.6, the effect is radical (Fig 4.7). The model is of insufficient complexity, that is, we no longer have support for quadratic terms.

Table 4.2 lists the main considerations when evaluating the summary-of-fit plot.

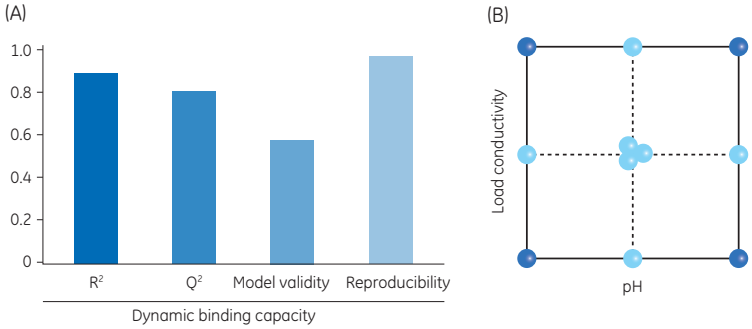


Fig 4.6. (A) A summary-of-fit plot showing a statistical summary of the regression modeling of the observed variation in dynamic binding capacity (DBC), caused by a controlled change in pH and conductivity (B), of a cation exchange chromatography medium for a monoclonal antibody. The plot shows excellent R^2 and a very good Q^2. There is no lack of fit as observed in the model validity, meaning that the residual variation is significantly lower than the replicate variation (the reproducibility of the center points). The reproducibility bar compares the center point variation with the overall variation, showing good results from the obtained data. In this case, the mathematical model corresponds to $DBC = k + b_1 \times pH + b_2 \times Cond + b_{11} \times pH^2 + b_{22} \times Cond^2 + b_{12} \times pH \times Cond + e$.

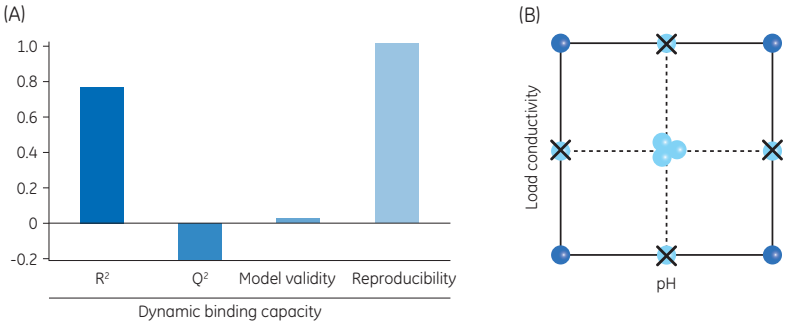


Fig 4.7. (A) A summary-of-fit plot for a model including linear and interaction model terms, which do not seem to fit into our data for DBC of the cation exchange chromatography medium for a monoclonal antibody. (B) Quadratic terms, supported by the star point experiments, are excluded in the data set and the effect is radical. Although the R^2 value is not critically low, the negative Q^2 is a clear indication of that the model cannot describe how DBC is affected by pH and conductivity. The low model validity indicates a strong lack of fit, that is, the residual variation is much greater than the variation in the replicated center points. The reproducibility is good, that is, the center point variation is low compared with the overall variation. The mathematical model in this case corresponds to $DBC = k + b_1 \times pH + b_2 \times Cond + b_{12} \times pH \times Cond + e$.

Table 4.2. Model statistics obtained in the summary-of-fit plot (Figures 4.6 and 4.7)

Coefficient value for...	Description
R^2	R^2 describes how well the model fits the current data. It can vary between 0 and 1, where 1 equals a perfect model and 0 corresponds to no model at all. A high R^2-value is necessary for a good model but not sufficient on its own. A value of 0.75 indicates a rough but stable and useful model and an R^2 of 0.5 is a model with rather low significance. Note: R^2 Adj is the fraction of variations in the response data that is explained by the model, adjusted for degrees of freedom. R^2 does not take into account degrees of freedom.
Q^2	Q^2 describes how well the model will predict new data. It can vary between infinity and 1. The higher Q^2-value, the better indicator of how well the model will predict new data. Q^2 should be greater than 0.1 for a significant model and greater than 0.5 for a good model. Q^2 is a better indicator of the usefulness of the model than R^2. Note: R^2 should not exceed Q^2 by more than 0.2–0.3 for a good model.
Model validity	Model validity tests a variety of problems and is only available if replicated experiments have been performed. A model validity > 0.25 indicates a good model. A model validity < 0.25indicates statistically significant model problems, such as the presence of outliers, an incorrect model, or a transformation problem. A low value here may also indicate that an interaction or square term is missing. When the pure error is very small (replicates almost identical), the model validity can be low even though the model is good and complete.
Reproducibility	A reproducibility < 0.5 indicates that there is a large pure error and poor control of the experimental setup (high noise level).

Although a good model (Q^2 > 0.9), the model validity can still be low because of high test sensitivity or extremely good replicates.

Coefficient plot

In the coefficient plot, we can view the effect and importance of each model term indicated by the height (positive or negative) of the response change as the factor changes from its low to high level (Fig 4.8). The coefficient plot is also useful for model refinement. Thus, nonsignificant terms are identified by checking the confidence intervals (the noise contained in the confidence intervals). If the confidence interval covers zero the term is not significant.

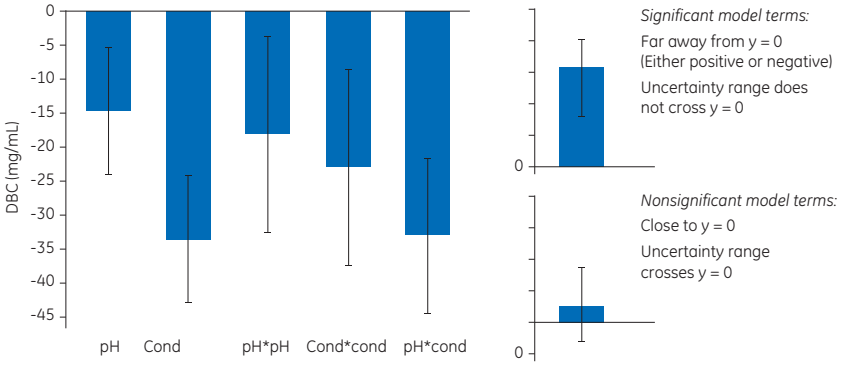


Fig 4.8. Example of a coefficient plot. Based on this plot, the interpretation of how pH and conductivity affects DBC of a chromatography medium is not straightforward because of the presence of strong quadratic and interaction effects. Model refinement involves removing insignificant terms from the model and is justified if it results in an increase in Q2. In this case, none of the model terms (pH, Cond, pH*pH, Cond*Cond, or pH*Cond) can be excluded as they are all significant. The significant mathematical model, using a 95% confidence level (or 5% risk), corresponds to $DBC = k + b_1 \times pH + b_2 \times Cond + b_{11} \times pH^2 + b_{22} \times Cond^2 + b_{12} * pH \times Cond + e$.

When editing a model, you cannot remove insignificant main terms if its interaction term is significant. For example, the main effect term cannot be removed if we have a significant interaction.

Main-effects plot

The main-effects plot displays the predicted response values when a factor varies from its low to its high level, all other factors in the design being set on their average levels (Fig 4.9).

The main-effects plot should be used for factors that are not involved (or involved in relatively small) interaction effects or used in direct combination with the interaction plot(s).

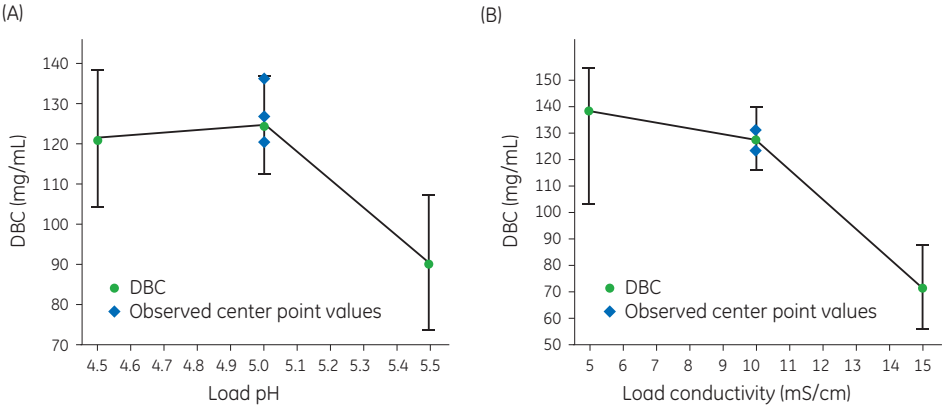


Fig 4.9. Example of main-effects plots: (A) the effect of load pH on medium DBC at the center point setting for load conductivity. (B) The effect of load conductivity on DBC at center point setting for load pH. The error bars indicate the width of the model prediction confidence intervals and the replicated center points are shown in blue.

Interaction plot

The interaction plot shows if there is any interaction between two factors (Fig 4.10).

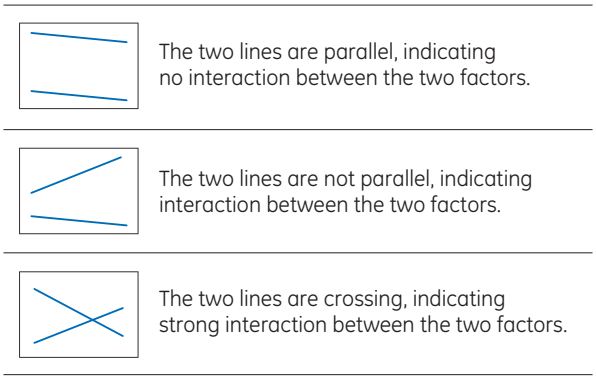


Fig 4.10. Interpretation of interaction plots.

The interaction plot should be used for interpretation of significant interaction effects.

Figure 4.11 shows examples of interaction plots. The plot in Figure 4.11A shows that the effect of load pH on the DBC of the chromatography medium depends on the load conductivity. At a low conductivity, an increase in pH results in an increase in DBC. On the other hand, at high conductivity, an increased pH results in a drastic drop in DBC. The plot in Figure 4.11B shows that the effect of load conductivity on DBC depends on the load pH. At low pH, increasing the conductivity results in a small increase in DBC, followed by a slight drop. At high pH, the effect of increasing the conductivity is a drastic drop in DBC.

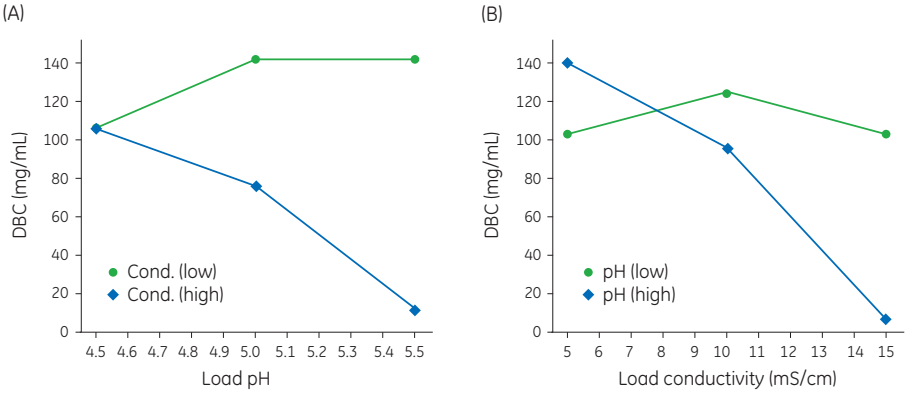


Fig 4.11. Interaction plots displaying factor interaction effects. (A) The effect of load pH on medium DBC depends on the load conductivity. At a low conductivity (5 mS/cm), a pH increase results in an increased DBC, whereas at high conductivity (15 mS/cm), increasing the pH results in a drastic drop in DBC. (B) The effect of load conductivity on DBC depends on the load pH. At low pH (pH 4.5), increasing the conductivity results in a slightly increased DBC followed by a slight drop, whereas at high pH (pH 5.5), the effect of increasing the conductivity is a drastic drop in DBC.

Residuals-versus-variable plot

In the residuals-versus-variable plot, standardized residuals are plotted against any factor included in the study. The aim is to detect if there are any indications of a curvature or other trends. In Figure 4.12, when plotting the residuals versus load pH, no curvature or trends can be detected when using a model that supports quadratic terms. Removing star points from the data will display missing model terms by a clear curvature indicated by the red dashed line.

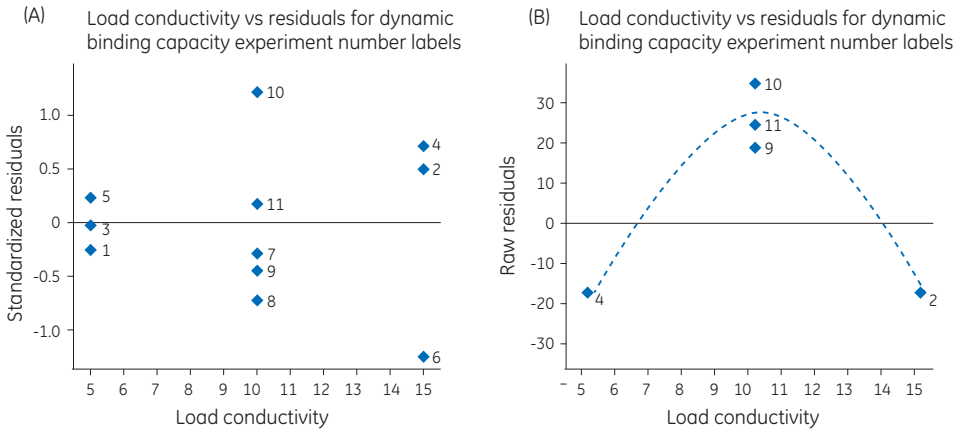


Fig 4.12. Example of a residuals-versus-variable plot. (A) No trend or pattern can be detected. The model corresponds to DBC = $k + b_1 \times pH + b_2 \times Cond + b_{11} \times pH^2 + b_{22} \times Cond^2 + b_{12} \times pH \times Cond + e$. (B) A clear indication of nonlinearity can be observed after removing star points (quadratic terms) in the model. The model corresponds to DBC = $k + b_1 \times pH + b_2 \times Cond + b_{12} \times pH \times Cond + e$.

Observed-versus-predicted plot

The observed-versus-predicted plot for a response can be used for estimation of the quality of a model. As illustrated in Figure 4.13, with a good model, all the data points will fall on a straight line.

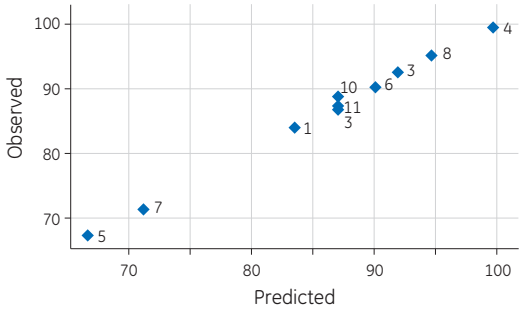


Fig 4.13. Observed (measured) values plotted against predicted model values.

Residuals-versus-run order plot

In the residuals-versus run order plot, any trends or patterns in the data can be detected. In Figure 4.14, we see the residuals plotted against the run order in which the experiments were performed. Here, no strong trends in the residuals or patterns in the data can be observed. The residuals are randomly distributed with no pattern. A pattern in this plot indicates a change in residuals over time, which could, for example, be the result when randomization errors exist in the experiment.

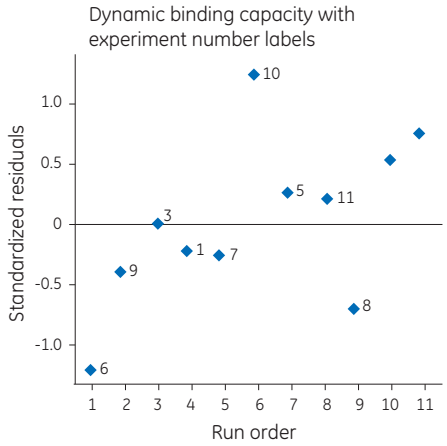


Fig 4.14. Example of a residuals-versus-run order plot. No visible pattern or trend is observed in the data.

ANOVA table

The ANOVA table summarizes and complements the data analyses described here in Chapter 4. As a reminder, we perform two statistical tests and compare the regression against the residuals as well as the model error against the replicate error. In both cases, we calculate the sum of squares followed by the mean square variance and divide these values to obtain the F-ratio. Finally, we calculate the probability value (p-value) for the F-ratio. Figures 4.15A and B illustrate two separate analyses of the same data set, with the difference that the star points in the data were removed in Figure 4.15B. By removing the star points (i.e., no support for the quadratic terms in the model), the variation described by the regression is not significantly larger than the variation in the residual. In other words, we conclude that the R^2 value and the model in whole are not significant. Removing the star points also results in a model error (residuals, replicate variation excluded) significantly larger than the replicate error. In other words, we can conclude that the model suffers from a statistically significant lack of fit.

(A) **Good model**
Significant R^2
No significant lack of fit

Dynamic Binding Capacity	DF	SS	MS (variance)	F	p
Total	11	131398	11945,3		
Constant	1	114036	114036		
Total Corrected	10	17361,6	1736,16		
Regression	5	16919,1	3383,82	38,2334	0,001
Residual	5	442,522	88,5043		
Lack of Fit	3	311,855	103,952	1,5911	0,408
(Model Error)					
Pure Error	2	130,667	65,3333		
(Replicate Error)					

(B) **Bad model**
No significant R^2
Significant lack of fit

Dynamic Binding Capacity	DF	SS	MS (variance)	F	p
Total	7	87944	12563,4		
Constant	1	74882,3	74882,3		
Total Corrected	6	13061,7	2176,95		
Regression	3	9674,75	3224,92	2,85646	0,206
Residual	3	3386,97	1128,99		
Lack of Fit	1	3256,3	3256,3	49,8413	0,019
(Model Error)					
Pure Error	2	130,667	65,3334		
(Replicate Error)					

Fig 4.15. The ANOVA statistical test shows two parallel analyses of the same data set from a study of the DBC of a chromatography medium. (A) The model (including the star points) is good, the variance in the model (regression) is smaller than the variance not captured (residual), and we obtain a significant R^2. The model error (residuals, replicate variance excluded) is not significantly larger than the replicate error, that is, the model has no lack of fit. The model corresponds to DBC = $k + b_1 \times pH + b_2 \times Cond + b_{11} \times pH_2 + b_{22} \times Cond^2 + b_{12} \times pH \times Cond + e$. (B) The star points are removed from the data, leaving a poor model with no significant R^2 and a significant lack of fit. The model corresponds to DBC = $k + b_1 \times pH + b_2 \times Cond + b_{12}$*$pH \times Cond + e$.

Response-surface plot

A response-surface plot is generated to get a graphical representation of the experimental region (Fig 4.16). From this plot, the most interesting area can be used to plan new experiments, verifying experiments, and to get a better understanding of the impact of large factor interactions. The response-surface plot is a great tool for visualizing interaction and curvature effects. The residual standard deviation (RSD) should always be considered when looking at this type of plot. Two factors are selected and displayed, and if more factors are included in the study, they will have constant values.

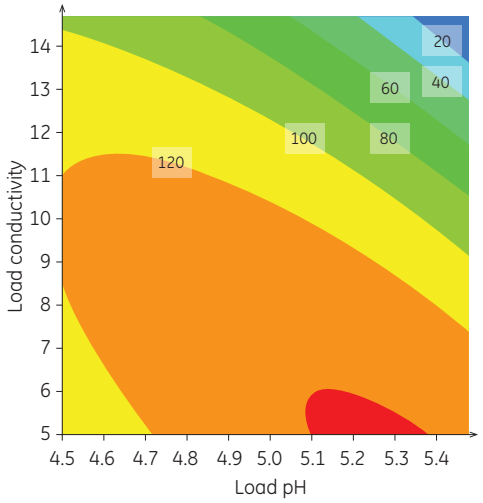


Fig 4.16. A response surface plot showing the quantitated effects (i.e., the f(x) part) of load pH and conductivity on medium DBC (gray squares). For this model, RSD = 9.4, meaning the model uncertainty for the DBC is about ± 18.8 mg/mL.

Sweet-spot plot

A sweet-spot plot highlights the areas were the included responses are within the user-specified ranges (Fig 4.17). In this plot, we set up the criteria for each response and the area corresponding to these criteria will be indicated.

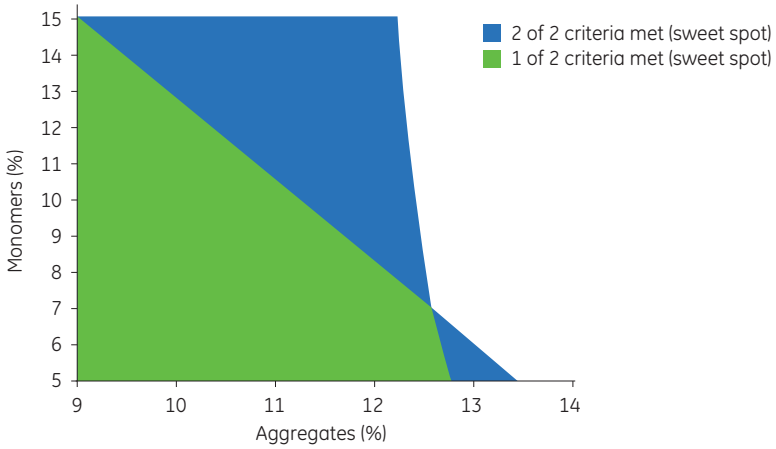


Fig 4.17. A sweet-spot plot for yield and purity of an antibody monomer. The blue region indicates the area where defined criteria for the yield and purity are met. The conditions for these plots were a sample load of 60 mg/L medium and elution with 300 mM of NaCl at pH 6.1.

Practical use of the obtained model

Like the data evaluation is based on and defined by the questions we ask, so is the use of a model. A model for predictions and optimization purposes is defined by questions about a desired output and how this is achieved in practice.

Predict outcome for additional factor values (prediction list)

In UNICORN software, it is possible to predict response values for additional factor levels based on the initially entered factors levels using the obtained model. This possibility is useful when you wish to find out how detailed factor settings influence the response(s) in an optimization experiment. Factor settings are entered and response values are calculated using the ***Prediction*** list (Fig 4.18).

Load pH	Load Conductivity	Dynamic Binding Capacity	Lower	Upper
4.25	7	121.5416	113.6...	129.4174
4.75	12	155.0053	152.2...	157.7551

Fig 4.18. The ***Prediction*** list of the UNICORN software.

Optimize results

It is possible to optimize the response values obtained in previous experiments using the ***Optimizer*** available in UNICORN software (Fig 4.18). When using the ***Optimizer***, criteria for the response values and factor settings are entered (e.g., yield > 90%). Based on entered settings, factor levels are calculated. In this way, the optimal region for a process can be estimated.

Result

Experiment

Calculate Optimal Settings | Clear | Factor | Response

Load pH	Load Conductivity	Dynamic Binding Capacity	Iter	Log(D)
4.7	6	158.4699	4	-1.1828
4.55	13	148.7036	4	-0.6498
5.5	5	163.5999	0	-1.6941
5.0361	6.9736	173.7244	60	-10
5	10	163.6107	0	-1.6955
4.5	10	160.0844	0	-1.3138
5	5	167.9511	0	-2.6834

Fig 4.19. In UNICORN ***Optimizer***, values are entered to minimize, maximize, or set a target value, or to exclude a specific response. The number of iterations for optimization is indicated in the Iter column, and a lower (or more negative) Log(D) (the logarithm of the distance to the target) value indicates improved results.

Chapter 5
Application examples

The case studies in this chapter comprise implementation of DoE in applications from upstream to downstream processing. The downstream examples are all according to the capture, intermediate purification, and polishing (CIPP) strategy.

5.1 Study of culture conditions for optimized Chinese hamster ovary (CHO) cell productivity

Case background

Recombinant proteins for therapeutic use are promising products for the biopharmaceutical industry. Different mammalian cell lines, such as CHO cells, mouse myeloma (NS0) cells, baby hamster kidney (BHK) cells, and human embryo kidney (HEK293) cells, are commonly used for recombinant protein production. For an optimized protein production process, many parameters need to be investigated. Because of the high number of parameters affecting the response (process outcome) and because of the potential interaction between them, optimization of the protein production process can be a tedious procedure. Changing one parameter at a time to identify factors responsible for the observed effects on the response would require numerous cultivations and be time-consuming. Applying DoE to the process development procedure can help reduce experimental burden, as settings of several parameters can be changed in parallel. In addition, existing parameter interactions can easily be identified during data evaluation. In this study, a biphasic cultivation strategy for a CHO cell line expressing an Epo-Fc fusion protein was developed. Special focus was on minimization of protein aggregation. The effects of cultivation temperature and pH on CHO cell productivity, as well as product aggregation were investigated using a DoE approach.

Methodology

Parallel batch cultivations of CHO cells expressing an Epo-Fc fusion protein were performed in stirred-tank bioreactors. As cultivation medium, DMEM/Ham's F12, supplemented with additional glucose, glutamine, and soy peptone, was used. To quantitate Epo-Fc monomer content, the recombinant protein was purified using MabSelect SuRe™ protein A medium and further analyzed by GF (also called size exclusion chromatography) using Superose™ 6 medium.

A central composite design was used for evaluation of the impact of pH and temperature on recombinant Epo-Fc production and aggregation. Three temperatures and three pH levels were investigated. As a nonlinear response to temperature was expected, the design was extended with two additional cultures at 28.5°C and 38.5°C to investigate curvature in the response. For this additional study, a fractional factorial design at resolution V was used. The experimental variability was assessed by three independent center-point experiments. The study outline is displayed in Table 5.1.1.

Table 5.1.1. Experimental outline with all experiments included for mathematical modeling

Exp. no	Exp. name	Run order	Temp.	pH
1	N1	4	30	6.75
2	N2	9	37	6.75
3	N3	3	30	7.05
4	N4	2	37	7.05
5	N5	10	28.5	6.9
6	N6	5	38.5	6.9
7	N7	8	33.5	6.75
8	N8	7	33.5	7.05
9	N9	1	33.5	6.9
10	N10	11	33.5	6.9
11	N11	6	33.5	6.9
12	N12	12	33.5	6.9

Results

Aggregation of Epo-Fc was strongly dependent of the cultivation pH and varied between 78% and 0.7% over the experiments. Figure 5.1.1 shows an example of monomer content from two different cultivation conditions analyzed by GF. A slightly acidic pH was clearly beneficial and resulted in low aggregate formation. A shift to lower cultivation temperatures further reduced Epo-Fc aggregation, indicating interaction between the two process parameters. The monomer peak contained more than 96% of the recombinant protein, and less than 1% of the protein was detected in the aggregate fraction.

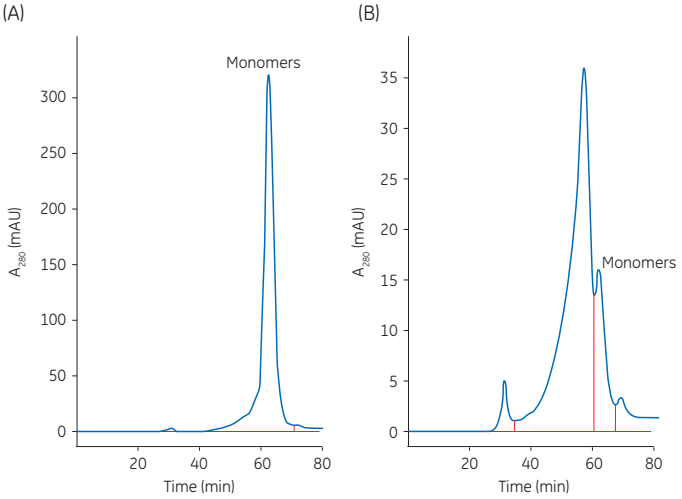


Fig 5.1.1. Gel filtration analyses after the protein A capture showing different Epo-Fc conformations depending on cultivation conditions: (A) 30°C, pH 6.75 and (B) 37°C, pH 7.05.

A response contour plot showing the monomer content as a function of cultivation temperature and pH is shown in Figure 5.1.2.

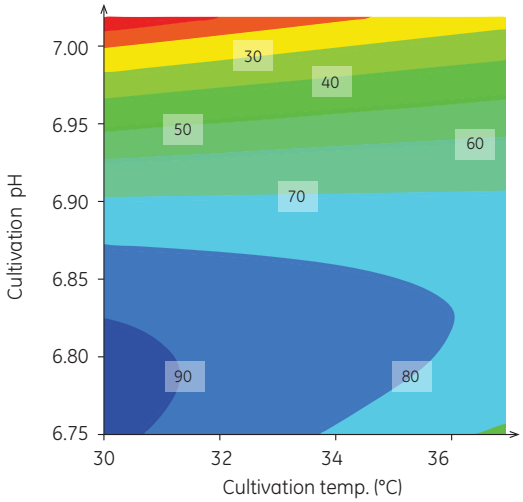


Fig 5.1.2. Response contour plot showing the monomer content (%, gray boxes) at various cultivation temperatures and pH values.

Conclusions

In this case study, the impact of cultivation temperature and pH on the productivity of CHO cells expressing a recombinant Epo-Fc fusion protein. In addition, protein aggregate formation was investigated. A central composite design was used for studying the effects of the cultivation parameters on the process outcome. As a curvature in the response was expected, two additional cultures at 28.5°C and 38.5°C were included in the study and a fractional factorial design at resolution V was used for data evaluation. Compared with cultivation at 37°C and pH 7.05, a parameter shift to a lower temperature and slightly more acidic pH resulted in a final 2.5-fold increase in product yield (96%) and a decrease in aggregate content (1%).

More detailed information on this study can be found in the poster by Kaisermayer *et al. Biphasic cultivation strategy for optimized protein expression and product quality* presented at the ESACT conference 2013.

5.2. Rapid development of a capture step for purifying recombinant S-transaminase from E. coli

Case background

This case study demonstrates a strategy for rapid process development of a capture step including anion exchange (AIEX) chromatography for purification of recombinant S-transaminase from *E. coli*. First, initial screening of chromatography media and process conditions was performed using prefilled PreDictor 96-well filter plates and Assist software. Next, UNICORN 6 software was used with ÄKTA avant 25 chromatography system to perform a DoE optimization of loading and elution conditions. DoE was also used for confirming that the optimized process conditions were robust.

Methodology

S-transaminase was recombinantly expressed in *E. coli*, and clarified cell suspension was used as sample in the development of an AIEX chromatography purification step for this enzyme.

Initial screening of media and loading conditions was performed using PreDictor AIEX Screening Plate, 20 µL prefilled with four different AIEX media: Capto Q, Capto DEAE, Q Sepharose Fast Flow, and Capto adhere. In the experimental design, generated using Assist software, pH and conductivity were varied for the four types of media tested: pH 6, 7, and 8 and NaCl concentrations of 0 mM, 50 mM, 75 mM, and 100 mM. Response data from the analysis of eluted samples were evaluated using Assist software. Purity and yield were analyzed by SDS-PAGE.

During screening, Capto DEAE was found to be the most suitable chromatography medium for this application, and HiTrap Capto DEAE 1 mL columns were used with ÄKTA avant 25 to optimize loading and elution conditions. The optimization was performed using DoE with UNICORN 6 software. To optimize loading conditions, the Rechtschaffner experimental design was used, allowing 13 different combinations of the tested parameters: loading pH, sample load volume, and flow rate. The sample (~ 4.0 mg/mL) was loaded in different volumes, 1.5, 8.25, and 15 mL at the different pH values 6.0, 6.5, and 7.0. Studied flow rates were 0.5, 2.25, and 4 mL/min. Step elution was achieved with 100 mM sodium phosphate buffer containing 0.5 mM pyridoxal 5'-phosphate and 1 M NaCl. Fractions were analyzed by SDS-PAGE and bicinchoninic acid (BCA) protein assay.

Elution conditions were optimized using the conditions obtained from the loading optimization. Elution pH and conductivity was varied according to a central composite circumscribed experimental design, allowing 11 different combinations of the parameters to be tested (Table 5.2.1).

Table 5.2.1. Run scheme for elution optimization on HiTrap Capto DEAE using a central composite circumscribed design

Run no.	Elution conductivity (mM NaCl)	Elution pH
1	200	6.5
2	600	6.5
3	900	6.8
4	600	6.5
5	600	6.5
6	1000	6.5
7	900	6.2
8	300	6.8
9	300	6.2
10	600	6.0
11	600	7.0

For robustness testing of the optimized purification step, two HiScreen Capto DEAE columns coupled in series for a 20 cm bed height were used on an ÄKTA avant 150 chromatography system. A fractional factorial experimental design generated by DoE in UNICORN 6 was used, allowing five different combinations of the controlled parameters—target elution conductivity (NaCl concentration) and sample load volume—to be tested. The samples (~ 4.0 mg/mL) were loaded in different volumes (14.5, 16, and 19 mL). A linear gradient using 100 mM sodium phosphate buffer with 0.5 mM pyridoxal 5'-phosphate and 1 M NaCl was used for elution at pH 6.0. Fractions were analyzed by SDS-PAGE and BCA protein assay.

Results

Chromatography media and binding conditions were screened using PreDictor AIEX screening plates containing four different types of media. Purity and yield were analyzed. Because of the complex nature of the sample, an overall purity of ~ 32% was obtained for the four different media types when varying the parameters salt and pH. However, the amount of recovered target protein was not highest in the samples that gave the highest purity. The media screening analysis revealed two candidate media: Capto DEAE and Capto adhere. Capto adhere is a strong anion exchanger and also enables multimodal interactions, whereas Capto DEAE is a weak anion exchanger. The results indicated an effect of load volume (data not shown), where the percentage of S-transaminase in the eluate increased with higher load.

This effect was more prominent for Capto adhere, which did not work well with a smaller sample load. Analyzing the amount of S-transaminase obtained when varying loading pH and conductivity revealed that a higher amount of target protein was obtained using Capto DEAE (Fig 5.2.1). Elution of the target protein from Capto adhere required 2 M NaCl at pH 3, but even at these conditions, the yield was lower than for Capto DEAE (data not shown). Thus, Capto DEAE was chosen for further optimization. The results also showed that low pH and lower conductivity favored a higher yield.

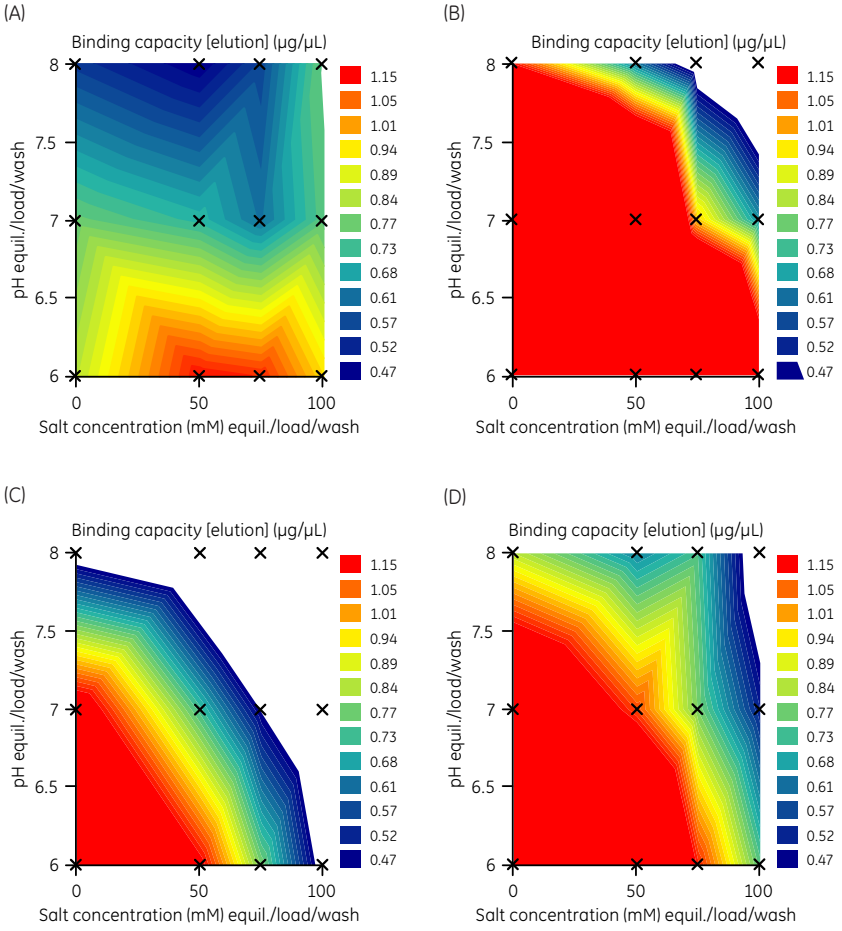


Fig 5.2.1. Surface plots generated using Assist software show the optimal region of the chosen parameters for (A) Capto Q, (B) Capto DEAE, (C) Q Sepharose Fast Flow, and (D) Capto adhere. Results indicate that the highest amount of target protein was obtained with Capto DEAE. Protein concentration is given in µg/µL.

Table 5.2.2 shows the run scheme and results from the DoE optimization of loading conditions using HiTrap Capto DEAE 1 mL columns. The study revealed that at increased load, more protein was recovered, although more protein was also lost in the flowthrough. The data indicate that the nonlinear relationship probably is due to displacement effects. The results from the robustness study indicate that an even higher load would have been beneficial (Fig 5.2.3). The overall purity increased to about 40% to 50%, which is sufficient for the purpose of immobilizing the protein onto a solid support.

Table 5.2.2. Experimental conditions and results from the loading study

Run no.	Loading pH	Flow rate (mL/min)	Sample load* (mL)	Purity (%)	Target protein in peak (mg)	Target protein in flowthrough fraction (mg)
1	6	0.5	15	51	6.8	3.3
2	6.5	2.25	8.25	45	3.1	1.2
3	7	0.5	1.5	37	0.5	0.1
4	7	0.5	15	49	7.4	3.0
5	7	4	1.5	41	0.5	0.2
6	6.5	2.25	8.25	48	3.2	1.3
7	6.5	2.25	15	51	6.4	3.2
8	7	2.25	8.25	40	2.5	2.7
9	6	0.5	1.5	42	0.7	0.2
10	6	4	15	47	6.6	4.7
11	6.5	4	8.25	38	2.8	2.6
12	6	4	1.5	54	0.9	0.2
13	6.5	2.25	8.25	44	3.3	2.5

** ~ 4.0 mg/mL*

Table 5.2.3. Experimental conditions and results from the DoE robustness test using the fractional factorial experimental design

Run no.	Loading conductivity (mM NaCl)	Loading volume (mL)	Purity (%)	Target protein (mg)
1	1.1	14.5	49	68
2	0.9	14.5	48	92
3	0.9	19	46	84
4	1	16.8	50	68
5	1.1	19	48	69
6	1	16.8	50	62
7	1	16.8	49	70

HiScreen Capto DEAE columns were used for robustness testing of the optimized process conditions using DoE. The SDS-PAGE analysis and BCA protein assay showed no significant change in the obtained purity or amount of target protein (calculated to 45% to 50% and 62 to 92 mg for the different runs). With an approximate 10% change in both loading salt concentration and loading volume, the statistical analysis displayed no significant model for either purity or the recovered amount of target protein (Fig 5.2.2). Thus, the process was considered to be robust for the tested conditions.

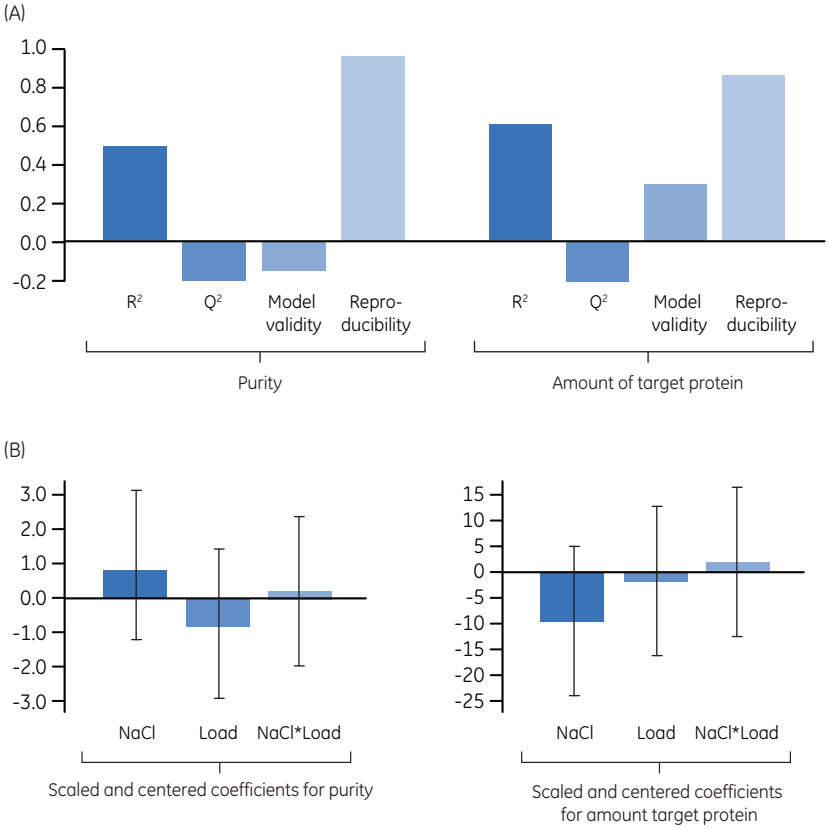


Fig 5.2.2. Statistical analysis of the robustness study. Purity and total amount (mg) of S-transaminase eluted from the column were determined using multiple linear regression analysis. (A) The summary of fit plot shows no significant model, and (B) no significant terms were obtained for the models generated. The tested conditions were therefore considered robust.

Conclusion

In this case study, DoE was used for rapid development of a single anion exchange chromatography capture step for the purification of S-transaminase. Due to the complex nature of the sample, a purity of only 40% to 50% was achieved. A second purification step may be required if a higher purity is needed. The higher yield obtained at a higher load was probably due to displacement effects on the column.

More detailed information on this study can be found in the application note *Fast process development of a single-step purification using ÄKTA avant systems* (28-9827-88).

5.3. *Optimization of conditions for immobilization of transaminase on NHS-activated Sepharose chromatography medium*

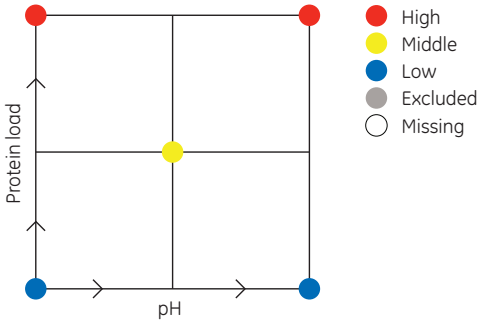
Case background

Transaminases are a group of enzymes that catalyze the transfer of an amino group from an amine to an acceptor, creating a ketone and an amino acid with a single chirality. There is great interest in the use of these enzymes for chiral production of chemical precursors. Chiral production is commonly performed by overexpressing the enzyme in a suitable host. The crude cell extract is thereafter mixed with the precursor to be modified. To some extent, the use of purified enzyme would reduce the required purification efforts for isolation of the product. An even more interesting approach is to use enzyme immobilized on a solid support for easy separation of the product from the enzyme. This case study, describes the optimization of conditions for coupling of purified *E. coli*-derived transaminase to NHS-activated Sepharose.

Methodology

For screening of coupling conditions, a full factorial design was used, with pH and amount of protein as factors and activity as the response (Fig 5.3.1). Different amounts of protein (5 to 20 mg/mL medium) at pH 7.5 to 9.5 were coupled in wells of a 96-well filter plate containing the equivalent of 55 µL NHS-activated Sepharose Fast Flow. Activity in each well was determined by measuring the amount of acetophenone, produced from (α)-MBA in the presence of sodium pyruvate in a reaction catalyzed by transaminase. Three series of wells were set up to allow three measurements (at different times) for each activity determination (each experimental point).

(A)



Investigation: DoE immobilization
Design: Full factorial (two levels)
Response: Activity

(B)

Center points

	1 ❶	2	3	4	5	6	7	❷
A	pH 7.5 5 mg 4 h	pH 7.5 20 mg 4 h	pH 9.5 5 mg 4 h	pH 9.5 20 mg 4 h	pH 8.5 12.5 mg 4 h	pH 8.5, 12.5 mg 4 h	pH 8.5 12.5 mg 4 h	20 min
B	pH 7.5 5 mg 4 h	pH 7.5 20 mg 4 h	pH 9.5 5 mg 4 h	pH 9.5 20 mg 4 h	pH 8.5 12.5 mg 4 h	pH 8.5 12.5 mg 4 h	pH 8.5 12.5 mg 4 h	10 min
C	pH 7.5 5 mg 4 h	pH 7.5 20 mg 4 h	pH 9.5 5 mg 4 h	pH 9.5 20 mg 4 h	pH 8.5 12.5 mg 4 h	pH 8.5 12.5 mg 4 h	pH 8.5 12.5 mg 4 h	5 min

NHS-activated Sepharose 4 Fast Flow

❶ Four conditions and center point = seven reactions
❷ Three time points used for activity assay after coupling

Fig 5.3.1. Screening model of the NHS coupling. (A) A full factorial design with two factors, pH and protein loading amount, was applied. Four conditions and a center point analyzed as triplicate were included. The corner colors represent the activity response level. (B) Protein load is stated in mg protein/mL medium.

Based on the findings from the condition screening experiment, a protocol for coupling of transaminase to HiTrap NHS-activated Sepharose HP 1 mL was designed. Immobilization of proteins on HiTrap NHS-activated HP 1 mL prepacked columns worked well above pH 8. The robustness of coupling of purified transaminase was evaluated by a DoE setup with the following factors at their high and low settings: pH 8 to 9, 30 to 120 min incubation time for coupling reaction, and a transaminase concentration of 3 to 5 mg/mL. Three center points were included in the design. The coupling was performed using ÄKTA pure chromatography system. The coupling efficiency was selected as response. The coupling efficiency was determined by quantitating protein in the flowthrough from the coupling column. A robustness test based on the DoE results was performed with the three factors coupling pH (pH 8 to 9), load concentration (3 to 5 mg/mL medium, and incubation time (30 to 120 min), and with coupling efficiency as the response.

On-column enzyme activity test was performed at 40°C using the model reaction for transaminase based on α-MBA as substrate. For optimization of the on-column conversion of α-MBA to acetophenone, conditions were screened with concentration of α-MBA (10 and 110 mM) and flow rate (0.1 and 0.5 mL/min) as factors, and yield (concentration of the product acetophenone) and conversion factor as responses (Fig 5.3.2). Center points were not included. For visualization of the screening results, yield and substrate conversion factor were subjected to multiple linear regression analysis and preparation of contour plots using SigmaPlot™ software (Systat Software).

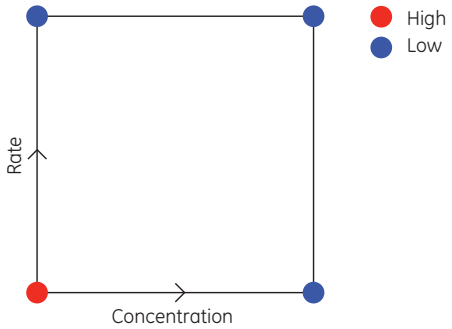


Fig 5.3.2. Screening model for the on-column reactions study. The factors included were concentration of α-MBA substrate and flow rate. The corner colors represent the product yield.

Results

The DoE summary-of-fit plot shows that the design model is good (Fig 5.3.3A), and the coefficient plot (Fig 5.3.3B) shows that the protein load is a valid factor. A higher transaminase activity was observed with high protein load during coupling. The results also suggest that there was no significant effect of pH on immobilized protein activity (results not shown).

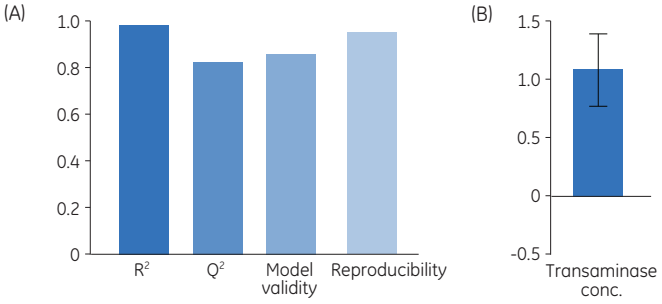


Fig 5.3.3. Evaluation of NHS coupling in a 96-well filter plate. (A) A summary-of-fit plot and (B) a coefficient plot.

A protocol for coupling of transaminase to HiTrap NHS-activated Sepharose HP 1 mL was designed based on the findings from the condition screening experiment (Fig 5.3.4).

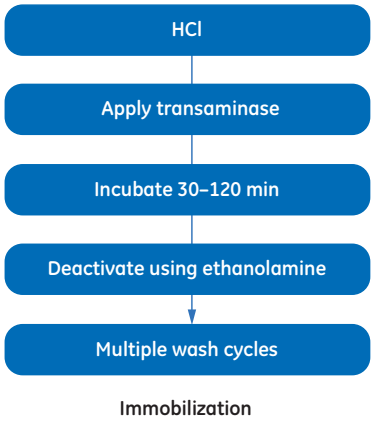


Fig 5.3.4. Workflow for coupling of transaminase on HiTrap NHS-activated HP 1 mL.

The coupling efficiencies obtained were in the range 87.5% to 98.3% for the robustness test experiment. The R^2 and Q^2 parameters in the summary-of-fit and coefficient plots showed weak or no relationship between factors and the response, indicating that the method was robust (Fig 5.3.5). The model reaction based on α-MBA and pyruvate as substrates was used for evaluation of the activity of the immobilized transaminase in the 1 mL HiTrap NHS-activated HP column.

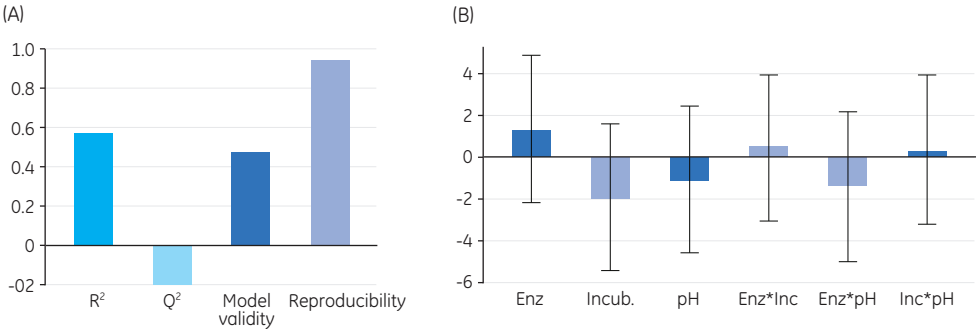


Fig 5.3.5. Evaluation of DoE results from robustness testing of the NHS coupling of transaminase on HiTrap NHS-activated HP 1 mL. (A) Summary-of-fit plot. (B) Coefficient plot.

Conditions for optimized on-column substrate conversion were determined in a screening study. The highest yield (2.0 mM acetophenone) was obtained using a low substrate concentration combined with a low flow rate, which might indicate product inhibition effects. The highest conversion factor was also observed for the same data point. As expected, the conversion factor plot indicated that a high substrate concentration reduces the conversion factor of the reaction. The highest conversion factor obtained was 20%. The results are visualized in Figure 5.3.6.

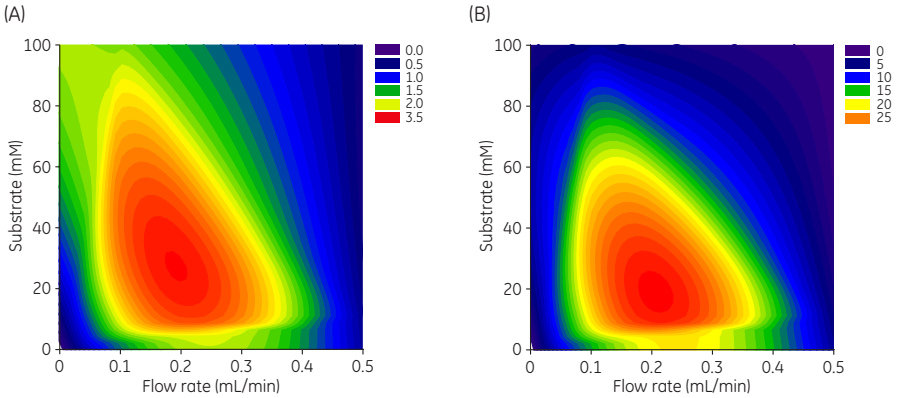


Fig 5.3.6. Contour plots of on-column reaction data obtained in the condition screening experiment. (A) Yield (mM). (B) Conversion factor (%). The plots serve as simplified visualization of the few data points, and should only be viewed as preliminary evaluation. More data is needed for a valid model of the reactions.

It was hypothesized that a longer (larger) column would increase the yield and conversion factor as this corresponds to an increased residence time on the column. This was tested by connecting five transaminase columns in series and applying a reaction mixture of 10 mM α-MBA and 17 mM sodium pyruvate at 0.1 mL/min. The yield was 3.9 mM acetophenone and the substrate conversion factor was 40%.

Conclusion

Investigation of on-column transaminase activity showed that application of 10 mM α-MBA (substrate) to the transaminase column at 40˚C with a residence time of 10 min gave a yield of 2 mM acetophenone (product). This yield corresponds to a substrate conversion factor of 20%. The conversion factor was doubled to 40% by using five HiTrap NHS-activated HP 1 mL columns in series (residence time 50 min) under the same conditions.

More detailed information on this study can be found in the application note *Purification and immobilization of a transaminase for the preparation of an enzyme bioreactor* (29-0211-99).

5.4. Optimization of dynamic binding capacity of Capto S chromatography medium

Case background

Capto S chromatography medium is based on high flow agarose. The medium combines high rigidity with high dynamic binding capacity (DBC) and fast mass transfer to allow faster purification, more flexible process design, and cost efficiency. Capto S is a strong cation exchanger and is often used as a second step in monoclonal antibody (MAb) purification.

High DBC of ion exchangers are typically obtained at low conductivities. As the conductivity of a sample increases, the DBC decreases. With Capto S medium, this traditional behavior is expected for most protein purification procedures. For some protein purifications, however, Capto S demonstrates a nontraditional behavior characterized by a DBC peak as the conductivity is increased (Fig 5.4.1).

To find optimize DBC of Capto S medium for a MAb, with regards to pH and conductivity, a thorough screening with a design capable of detecting curvature effects was performed.

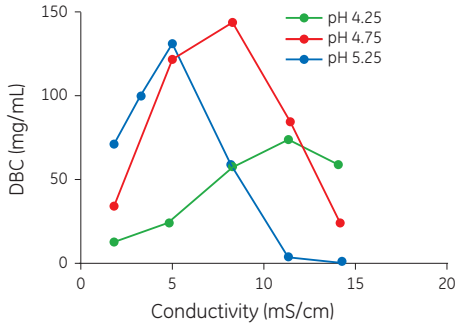


Fig 5.4.1. DBC versus sample conductivity at different pH for Capto S used in the purification of conalbumin. The DBC goes through its maximum and then declines as conductivity is increased.

Methodology

To study the effect of pH and conductivity on the DBC of Capto S medium when used in purification of a MAb, a central composite face-centered design with pH (range 4.5 to 5.5) and conductivity (range 5 to 15 mS/cm) was set up (Fig 5.4.2). Three center points were used to enable the detection of curvature in the data. In total, 11 experiments were conducted (Table 5.4.1). The response in this study was DBC for a pure MAb. All DBC measurements were conducted on an ÄKTA chromatographic system. The residence time was 4 min and the MAb concentration was kept constant at 5 mg/mL.

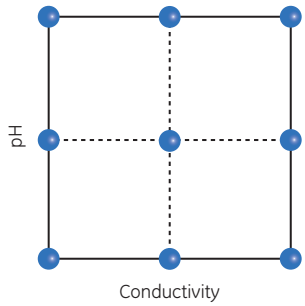


Fig 5.4.2. Central composite facial design used in optimization of the DBC of Capto S when used in a MAb purification.

Table 5.4.1. Experimental outline

Exp. no.	Run order	Cond.	pH	DBC
1	4	5	4.5	96
2	10	15	4.5	102
3	3	5	5.5	137
4	11	15	5.5	4
5	7	5	5.0	139
6	1	15	5.0	54
7	5	10	4.5	119
8	9	10	5.5	84
9	2	10	5.0	121
10	6	10	5.0	137
11	8	10	5.0	127

Results

Figure 5.4.3 displays a plot of raw data. At higher pH (pH 5.0, 5.5), a traditional behavior was observed, with decreased DBC when increasing the conductivity. At lower pH (pH 4.5), on the other hand, Capto S exhibited a nontraditional behavior, with DBC going through a maximum and then declines as conductivity was increased.

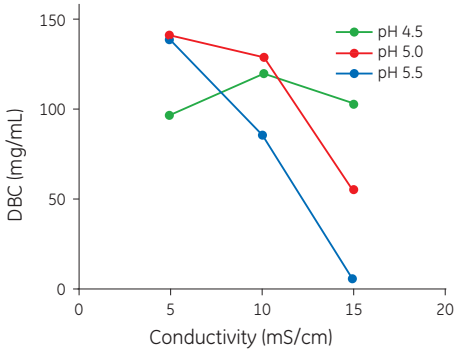


Fig 5.4.3. DBC versus conductivity, plotted at the three pH values.

Data analysis generated a well-explained model (R^2 = 0.98) with good stability according to a cross-validation study (Q^2 = 0.83). The coefficient plot shows that both pH and conductivity were significant factors affecting the DBC (Fig 5.4.4A). Curvature effects were found with both pH and conductivity. Additionally, an interaction effect between pH and conductivity was found. The coefficient plot therefore shows that:

- DBC decreases with increasing conductivity (Cond), with curvature (Cond*Cond).
- DBC decreases with increasing pH, with curvature (pH*pH).
- The effect of conductivity depends on pH (Cond*pH).
  - Traditional behavior at high pH (low net charge of the target protein).
  - Nontraditional behavior at low pH (high net charge of the target protein).

The response-surface plot, illustrating the coefficient plot graphically, shows that the highest DBC is predicted at low conductivity and at approximately pH 5.25. (Fig 5.4.4B).

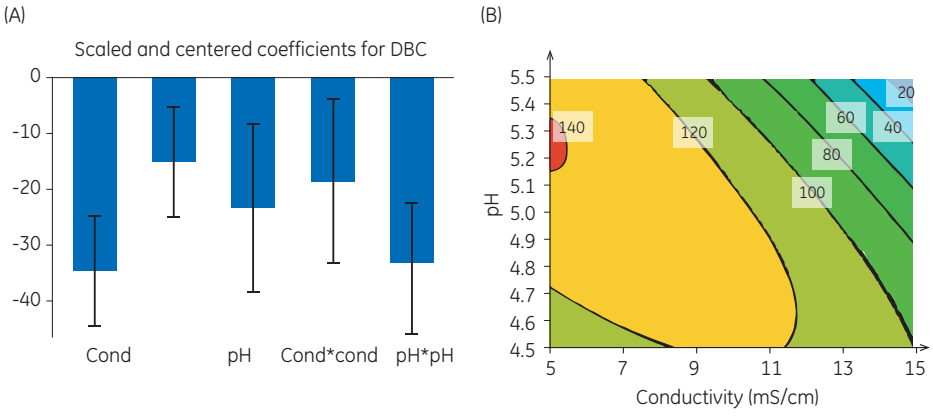


Fig 5.4.4. (A) Coefficient plot showing the model terms (curvature and interaction effects between pH and conductivity) that significantly affect the response (DBC). (B) A response-surface plot showing the predicted DBC (white squares = mg MAb/mL medium) as a function of pH and conductivity.

Conclusions

A central composite facial design enabled the detection of curvature and interaction effect between pH and conductivity in the optimization of the DBC of Capto S medium using a MAb. By data analysis using DoE software, an optimal DBC could be predicted.

More detailed information on this study can be found in the application note *Screening and optimization of the loading conditions on Capto S* (28-4078-16).

5.5. *Purification of an antibody fragment using Capto L chromatography medium*

Case background

Antibody fragments (e.g., Fab, scFv, and Dab) are expected to become the next important class of protein-based biotherapeutics after MAbs. With its recombinant protein L ligand, Capto L is a chromatography medium with affinity for a broad range of antibody fragments containing the kappa light chain (Fig 5.5.1).

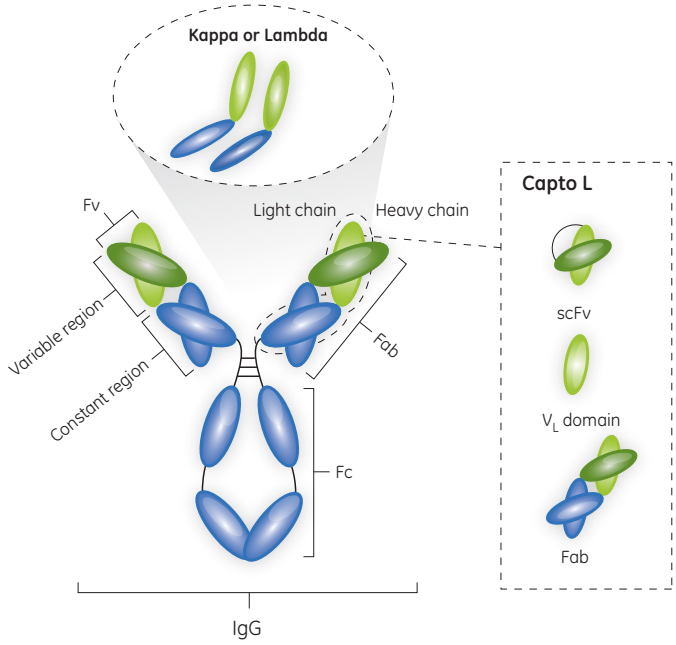


Fig 5.5.1. Structure of the IgG antibody. Capto L binds antibody fragments containing the kappa light chain.

This case study describes the use of Capto L medium in the capture step of the purification of a Fab fragment from an *E. coli* supernatant. Properties of the Fab fragment are listed in Table 5.5.1. The *E. coli* protein (ECP) content of the supernatant was approximately 300 000 ppm.

Table 5.5.1. Characteristics of the Fab fragment used in the Capto L capture step

Fab origin	*E. coli*
Theoretical isoelectric point (pI)	8.5
Molecular weight	M_r 48 000
Fab concentration in feed	1 mg/mL
Aggregate content	3.5%

The goal of a capture step is rapid isolation, stabilization, and concentration of the target molecule. Hence, this step requires a technique with high capacity and selectivity for the target. The capture step should also concentrate the target molecule, to enable fast processing in subsequent purification steps. Capto L medium is based on high-flow agarose and exhibits excellent flow property and high binding capacity, which makes it well-suited for the capture step of a Fab purification.

The requirements for the capture step described here were an ECP content of 2 to 30 ppm and a recovery of 95% to 100%.

Methodology

Wash and elution conditions were evaluated in PreDictor 96-well filter plates filled with Capto L medium. Studies of binding capacity were performed in a small-column format using purified Fab. To optimize wash and elution conditions, a three-factor central composite circumscribed design was used (Fig 5.5.2). This high-resolution design supports determination of quadratic interactions. Including star points eliminate confounding effects between factor interactions and quadratic terms.

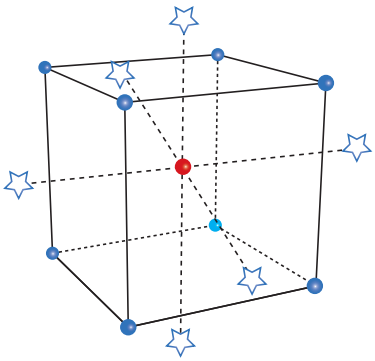


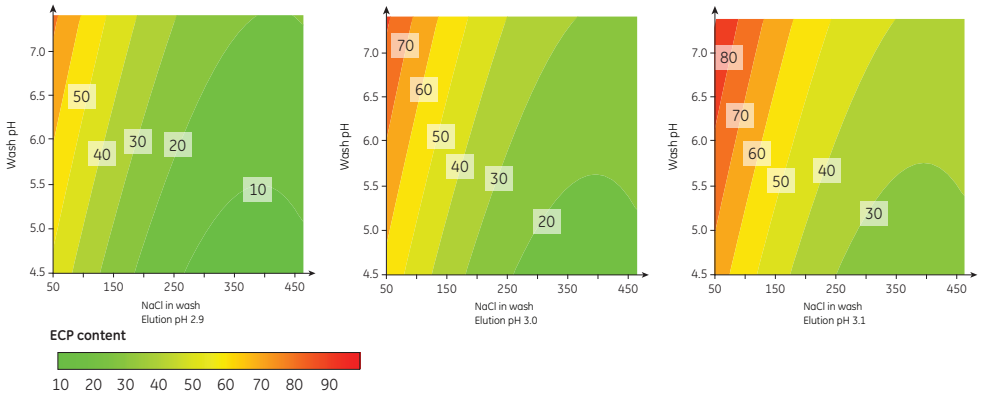
Fig 5.5.2. The central composite circumscribed design.

Studied factors were pH and NaCl concentration for wash and pH for elution. Other factors, such as sample load, were kept constant. A Fab sample concentration of 15 mg/mL corresponds to 70% of the DBC of Capto L and is a representative load for a production scenario. The responses in this study were ECP content and Fab recovery. ECP content was determined using antibodies from Cygnus Technologies on a Gyrolab™ workstation (Gyros AB) and Fab recovery was analyzed by GF.

Results

Figure 5.5.3 shows response-surface plots of ECP content and Fab recovery at three different pH values for elution.

(A) ECP content



(B) Fab recovery

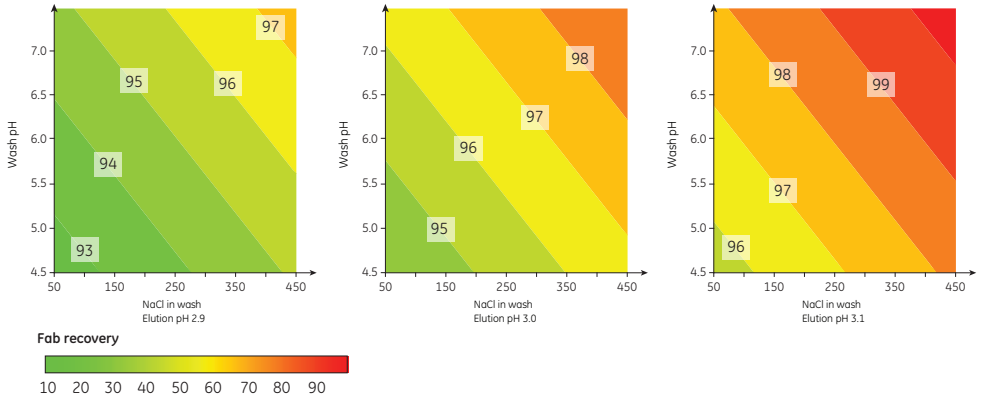


Fig 5.5.3. (A) Contour plots of ECP content (ppm, gray boxes) at three different elution pH values. (B) Contour plots of Fab recovery (%, gray boxes) at three different elution pH values.

The results show that optimum for Fab recovery and ECP content did not coincide. Optimal recovery was obtained at high salt content and high pH of the wash solution, while optimized removal of ECP was obtained at high salt content and low pH in the wash solution.

The criteria for the described process step were an ECP content of 2 to 30 ppm and a Fab recovery of 95% to 100%. Sweet-spot analysis was performed to combine the information retrieved for ECP content and Fab recovery to find the overall optimum.

The green surface in Figure 5.5.4 shows conditions where both criteria were met. A verification run was performed using a wash solution. As expected from the screening experiment, the resulting Fab recovery was 96% with an ECP content of 12 ppm.

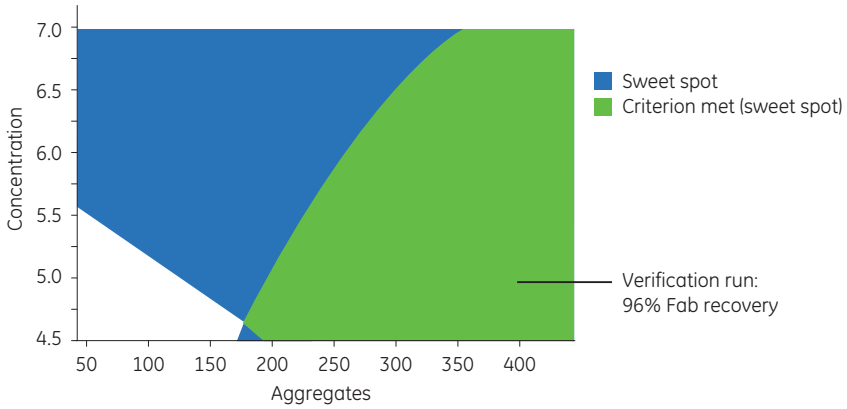


Fig 5.5.4. Sweet-spot analysis for Fab recovery and ECP content.

Conclusions

A three-factor central composite circumscribed design was used for optimization of recovery and purity in a Capto L capture step of a Fab purification process. With optimized conditions, EPC content could be 25 000-fold reduced, while achieving a high Fab recovery of 96%.

More detailed information on this study can be found in the application note *A platform approach for the purification of antibody fragments (Fabs)* (29-0320-66).

5.6. Optimization of elution conditions for a human IgG using Protein A Mag Sepharose Xtra magnetic beads

Case background

A well-known method for purification of antibodies from serum, ascites, or cell culture supernatant is to utilize the strong affinity of protein A from *Staphylococcus* or protein G from *Streptococcus* the for Fc part of IgG. Using protein A or protein G magnetic beads for antibody purification takes advantage of this highly specific interaction, and offers a simple, easy-to-use tool for rapid separations across the microliter to milliliter range. For high recovery, antibodies are normally eluted from protein A or protein G media using low-pH buffers. However, as certain antibodies precipitate at very low pH, additives such as arginine can enable elution at a higher pH. In this study, the effect of elution pH, conductivity (NaCl), and arginine concentration on IgG recovery was studied using a DoE approach.

Methodology

The response in this study was IgG recovery. A central composite facial design was used for evaluation of the effect of the three factors pH (3.0 to 4.0), arginine concentration (0 to 1.0 M), and NaCl concentration (0 to 750 mM) in the elution buffer, with three replicates at the center point (Fig 5.6.1). A total of 18 experiments were performed (Table 5.6.1). The sample load was 24 mg human IgG per milliliter of sedimented Mag Sepharose Xtra medium. The medium was eluted three times with elution buffer and then cleaned. The recovery and the mass balance were calculated. Data analysis was conducted using DoE software.

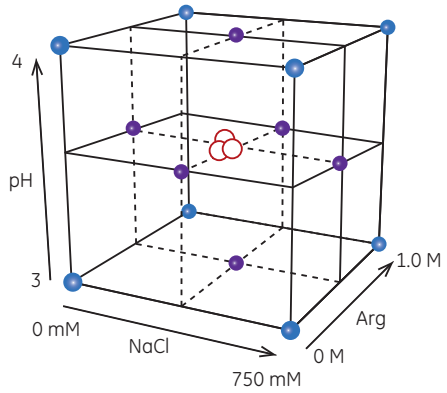


Fig 5.6.1. Experimental space of the central composite facial design used in this study.

Table 5.6.1. Experimental outline (color-coding according to Fig 5.6.1)

Exp. no.	pH	Conc. NaCl (mM)	Conc. arginine (M)
1	3.5	375	0.5
2	4.0	750	0
3	3.0	750	0
4	4.0	0	1.0
5	3.5	375	0.5
6	3.0	0	1.0
7	4.0	0	1.0
8	4.0	750	1.0
9	3.5	375	0.5
10	3.0	750	1.0
11	3.0	0	1.0
12	3.0	375	0.5
13	4.0	375	0.5
14	3.5	0	0.5
15	3.5	750	0.5
16	3.5	375	0
17	3.5	375	1.0
18	3.5	375	0.5

Results

The replicate plot of raw data shows the variation in IgG recovery for all experiments (Fig 5.6.2). For this study, the variability of repeated experiments was much less than the overall variability, which is optimal. The center points at replicate index 9 were all above the middle part of the response range, indicating a curvature in the model.

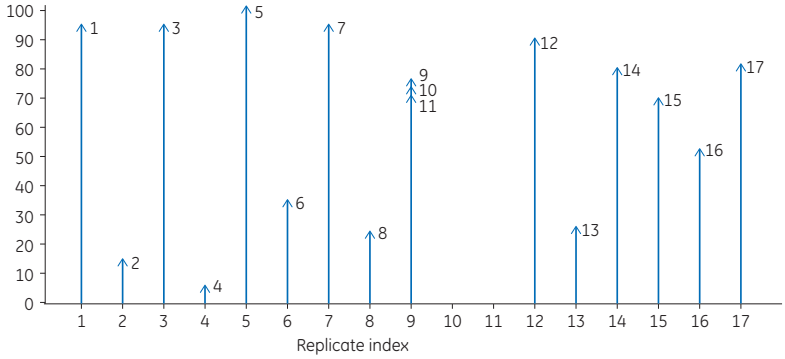


Fig 5.6.2. Replicate plot for the recovery (%). Experiments 9, 10, 11, and 18 represent the repeated center point and the values are all indicated at replicate index 9. Replicate index (i.e., the experiment number) is not the same as run order.

Yield data were fed into the DoE software and a mathematical model was fitted to this data. Figure 5.6.3 shows the coefficient plot, which is a graphical presentation of the significance of the model terms after removing nonsignificant terms. All three factors (pH, NaCl concentrations, and arginine concentration) were significant. Coefficients for pH and NaCl were negative, which means that high pH and high NaCl concentration in the elution buffer resulted in decreased recovery. Arginine concentration, on the other hand, was positive, meaning that a high arginine concentration resulted in an increased recovery. The pH was shown to exert the highest effect on the response and should be low for a good recovery. In addition, a curvature effect of pH (pH*pH) and interaction between pH and arginine concentration (pH*Arg) were observed. There is also a quadratic term in pH and an interaction term.

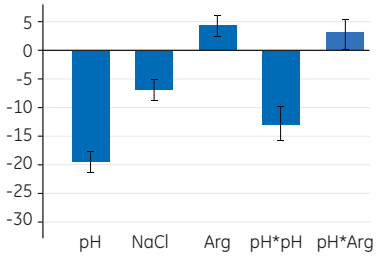


Fig 5.6.3. Coefficient plot showing that the factors pH, NaCl concentration, and arginine concentration significantly affect the response (recovery).

Figure 5.6.4 shows the summary plot of the model quality. R^2 shows the model fit and Q^2 gives an estimate of prediction precision. Q^2 should be greater than 0.1 for a significant model and greater than 0.5 for a good model. Model validity value represents the lack of fit of the model. A model validity value of less than 0.25 indicates a statistically significant lack of fit, for example, caused by the presence of outliers, an incorrect model, or a transformation problem. Reproducibility represents the variation of the replicates compared with the overall variability. A value for reproducibility that is greater than 0.5 is considered acceptable. The summary plot shows that the model is good with R^2, Q^2, and reproducibility close to 1 and model validity greater than 0.25.

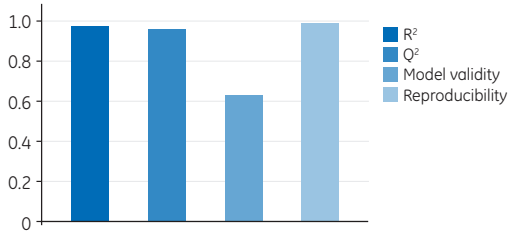


Fig 5.6.4. Summary plot of the model used.

The response contour plot in Figure 5.6.5 shows the IgG recovery at various pH and arginine concentrations in the elution buffer (NaCl concentration was set to 0). According to the results, an elution pH of less than 3.2 should be used for a good recovery (> 90%). Addition of arginine to the elution buffer slightly increases the recovery. An elution pH close to 4.0 gives a very low recovery (20% to 40%) without arginine and slightly higher (around 50%) by addition of 1.0 M arginine.

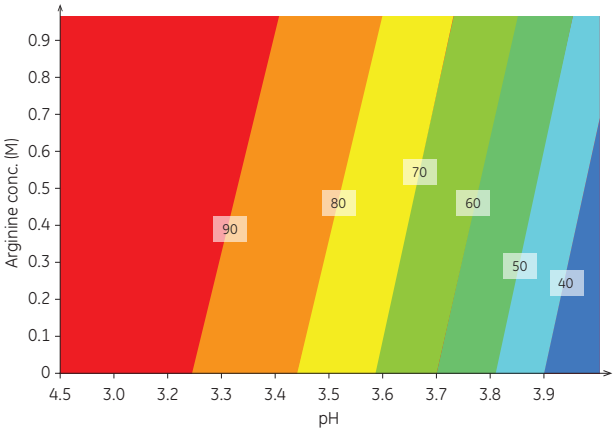


Fig 5.6.5. Response contour plot showing the recovery (%, gray boxes) at various pH and arginine concentrations in the elution buffer (NaCl concentration was set to 0).

Conclusions

A central composite facial design was used for studying the effect of the factors pH and concentrations of NaCl and arginine in the elution buffer on the recovery of a human IgG. Using this design, a model including both quadratic and interaction terms could be derived. Optimized IgG recovery (> 90%) was achieved by elution at low pH (pH < 3.2). By addition of arginine to the elution buffer, the recovery could be slightly increased.

More detailed information on this study can be found in the poster by Granér, T. *et al. High capacity, small-scale antibody purification using magnetic beads* presented at the ESBES conference 2010.

5.7. *Optimization of the multimodal polishing step of a MAb purification process*

Case background

Protein A affinity media, such as MabSelect SuRe chromatography media, are commonly used in the capture step of MAb purification processes as these media enable high MAb purity and yield after a single chromatography step. Subsequent downstream processing can be performed according to a variety of protocols including different combinations of chromatographic techniques such as IEX and HIC. Using a MabSelect SuRe medium in the capture step and a multimodal chromatography medium, capable of separation on both hydrophobic and ion exchange interactions, in the second step, MAbs purification with high recovery can be effectively performed in only two steps (Fig 5.7.1). The higher complexity of multimodal media requires somewhat more process optimization compared with traditional polishing media to take full advantage of the outstanding potential of this technology. This case study describes optimization of the polishing step of a MAb purification process including Capto adhere multimodal medium for selective removal of antibody aggregates.

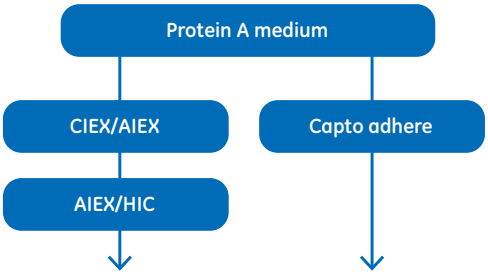


Fig 5.7.1. A three-step MAb purification process (left) can be reduced to a two-step process when using a multimodal chromatography medium, such as Capto adhere, in the second polishing step (right). CIEX = cation exchange chromatography, AIEX = anion exchange chromatography.

Methodology

PreDictor 96-well filter plates prefilled with Capto adhere were used for quick screening of a large experimental space. Favorable conditions from the plate study were further optimized on HiScreen columns using a DoE approach to establish the final process conditions. Column optimization was conducted and a central composite facial design generated. Investigated factors and factor ranges are summarized in Table 5.7.1. Responses used in the design were antibody monomer purity and recovery. The experiments were performed to optimize conditions for achieving an antibody monomer purity of > 99% and a recovery of > 85%. The process was verified using a 1 mL HiTrap Capto adhere column.

Table 5.7.1. Investigated factors and ranges (factor abbreviations used in DoE software are shown in parentheses)

Factor	Factor range
Aggregate content of the start sample (Aggr.)	9% to 14%
Protein concentration (Conc.)	5 to 15 mg/mL
Sample load (Load)	60 to 100 mg/mL
Elution pH (pH)	6.1 to 6.5
NaCl concentration in elution buffer (NaCl)	150 to 450 mM

Results

Models were built for the two responses—antibody purity and recovery. As shown in Figure 5.7.2, the purity decreased with increasing protein and aggregate amounts in the starting material, and by increased sample load and conductivity (NaCl) of the elution buffer.

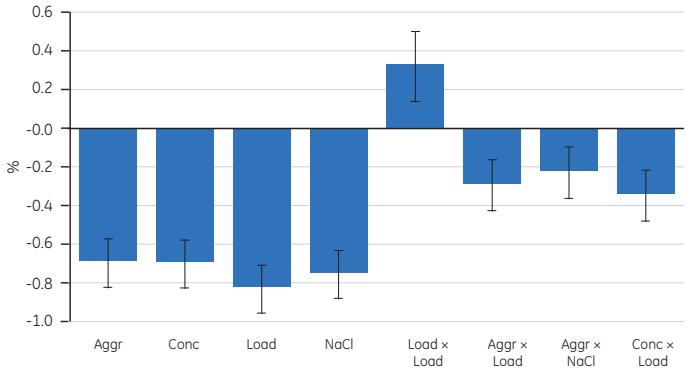


Fig 5.7.2. Coefficient plot for antibody monomer purity.

Figure 5.7.3 shows that the recovery decreased with increasing aggregate amount in the starting material, as well as by increasing pH of the elution buffer. However, the yield was increased by increases in sample load and NaCl concentration of the elution buffer. Although the effect of the sample load itself was not significant, it was included in the model, as one of the found interactions involved this factor. For this model, quadratic terms and other interactions were present.

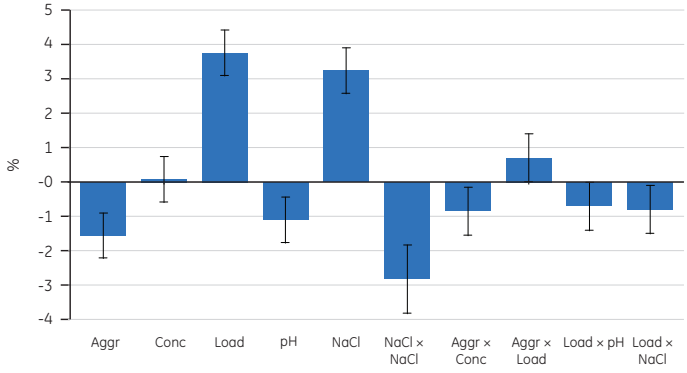


Fig 5.7.3. Coefficient plot for antibody monomer recovery.

The models for purity and recovery can be combined for production of a sweet-spot plot for a particular set of user-defined criteria. Fig 5.7.4 shows a sweet-spot plot for a sample load of 60 mg/mL and elution with 300 mM NaCl at pH 6.1. Here, the set criteria were monomer recovery > 85% and purity > 99% (equivalent to less than 1% of antibody aggregates).

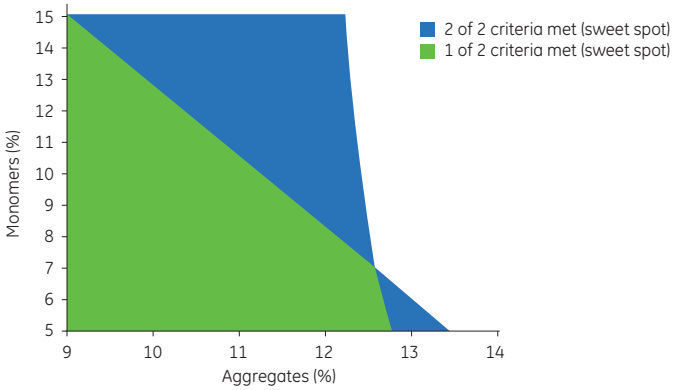


Fig 5.7.4. A sweet-spot plot for antibody monomer recovery and purity. Selected conditions were sample load 60 mg antibody/mL medium and elution with 300 mM NaCl at pH 6.1.

The process was verified using a 1 mL HiTrap Capto adhere column and similar run conditions to the conditions used for sweet-spot analysis:

- Sample load: 60 mg/mL
- IgG aggregate amount in the starting sample: 12.6%
- IgG concentration in the starting sample: 5 mg/mL
- Elution conditions: 250 mM NaCl, pH 6.1

The column verification study resulted in a recovery of eluted antibody monomer of 87% with a purity of 99.5%, both meeting the set criteria.

Conclusions

By using a DoE approach for screening of chromatographic run conditions, an IgG polishing step including Capto adhere multimodal medium could be effectively and rapidly developed. This polishing step can be used after initial IgG capture using MabSelect SuRe in a two-step process for efficient purification to high purity and yield.

More detailed information on this study can be found in the application note *High-throughput screening and optimization of a multimodal polishing step in a monoclonal antibody purification process* (28-9509-60).

Appendix 1

Parameter interactions

Clarifying parameter interactions (how multiple factors will affect a specific response) is an important aspect of process characterization. Understanding nonlinear relationships of chromatographic processes will ultimately lead to the establishment of an enhanced process. For the biochemist, the term "interaction" is often not as intuitive as the one-parameter-at-a-time paradigm has prevailed. Here, IEX is used as an example for explaining why we need to vary process parameters simultaneously in order to find these irregularities.

Ionized cysteine, aspartate, lysine, and histidine or an N-terminal amino acid typically accounts for positive charges, whereas aspartate, glutamate, or a C-terminal amino acid accounts for negative charges at certain pH values. At the isoelectric point (protein net charge 0) the total number of positive charges equals the number of negative charges. The protein charge is pH-dependent, so that at a pH lower than the protein isoelectric point, the protein has a positive net charge, and at a higher pH than the isoelectric point, the protein net charge is negative. Protein binding is not only pH dependent. Protein binding also depends on the ionic strength, so that proteins that are adsorbed by an ion exchanger at a low ionic strength can be desorbed at high ionic strengths.

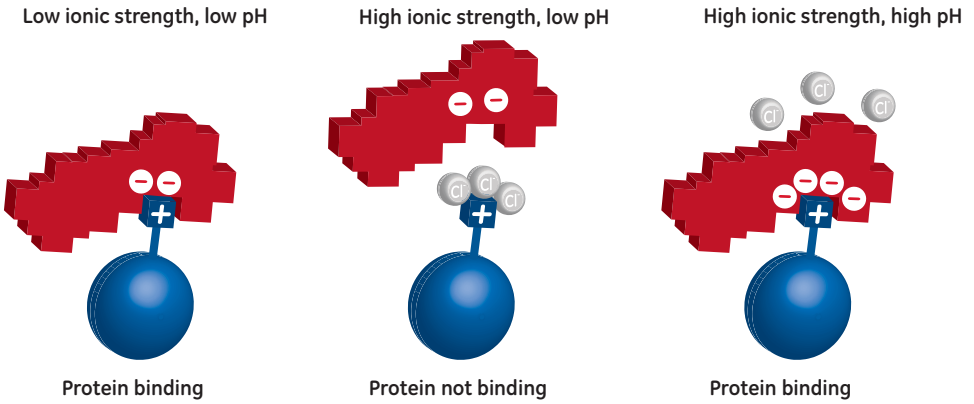


Fig 6.1. A schematic illustration of parameter interaction. The combination of pH and conductivity during sample application in anion exchange chromatography determines whether the protein (red) will bind to the chromatography medium (blue).

Somewhat simplified, we can vary pH to change the surface charge of proteins and hence to decrease or increase the interaction between proteins and the ion exchanger. We can vary the ionic strength to achieve a more selective binding. In this case, nothing implies that the factor-response relationships are solely linear. Clearly, there are combined effects of both pH and conductivity (parameter interactions) and possibly also an optimum to be found (i.e., a curvature). In IEX, pH elution is not common practice, as some proteins precipitate at pH value near their isoelectric point, with column clogging as a result. Elution pH will also change the ionic strength of the solution and impact the process reproducibility.

In IEX, pH elution is not common practice, as some proteins precipitate at pH value near their isoelectric point, with column clogging as a result. Elution pH will also change the ionic strength of the solution and impact the process reproducibility.

Monte Carlo simulation

Monte Carlo simulation is a stochastic (random) simulation procedure that makes use of internally generated (pseudo) numbers to compute variation in our process outputs or CTQ response parameters. If we know the mean values, standard deviation, and shape of the factors distribution in a study, the Monte Carlo method can be used for prediction of the shape of a response distribution as well as the mean and standard deviation. In the Monte Carlo simulation, the number of runs is independent of the number of factors. With this simulation technique, we can obtain a reasonable accuracy in a few thousand runs. The procedure is to iterate the creation of random numbers for all parameters (all with a specified statistical distribution) and to calculate and record response values using our (known) transfer function (Fig 6.2). The obtained data is used to calculate the response statistics (mean, standard deviation, and statistical distribution). The probability distribution will give us the possibility to add specification limits to the analysis (low, high, or both) in order to compute the response, for example, protein yield. The portion of the distribution that falls outside the specifications is the defect rate (a useful measure for CTQ optimization) for example, for obtaining consistent product performance with minimal variation.

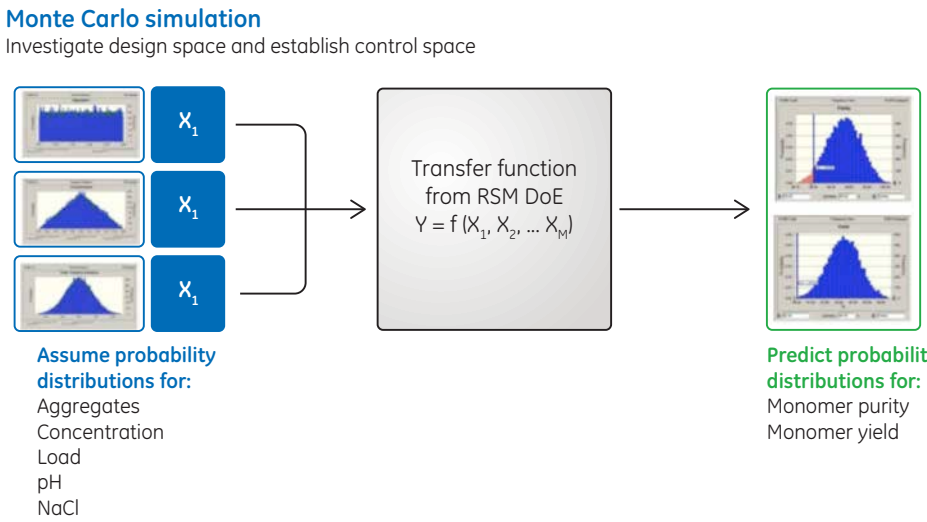


Fig 6.2. In the Monte Carlo simulation, response values are calculated and recorded using a transfer function.

Design considerations

There are different approaches for design selection depending on the amount of information required (Fig 6.3).

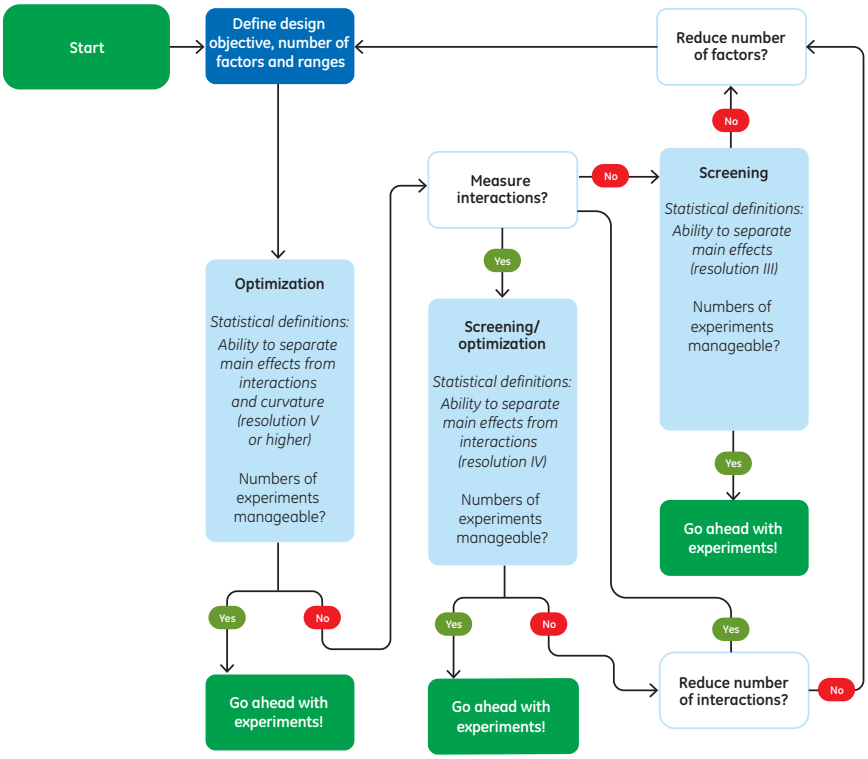


Fig 6.3. Selection of a design and determination of required number of experiments are based on the study objective, the number of factors, and factor levels. With the different resolutions, we can identify which effects can be estimated and the ability of a design to resolve these effects.

An alternative route is visualized in Figure 6.4. In this route, we will perform initial experiments to reveal the presence of interactions, confounding factors, and/or second degree curvature. As an example, the +++ and --- experiments, meaning the low-low-low and high-high-high settings in a three-factor study, and one center point are conducted, followed by evaluation of the results and review of the level settings in the design. This alternative route is especially useful when using two-level screening designs, where it is desired to prevent strong curvature in the response caused by too wide ranges of the factor levels. Among the benefits are an early indication of curvature and maximized information retrieval from the experiments used. An extension of the initial design could, for example, include expansion of a fractional factorial design to a full factorial design. So that instead of running a selected set of corner points, we include all parameters and their specification limits. This approach gives us the possibility to quantitate the main effects and all interaction effects, and identify whether or not a significant curvature is present. Simultaneously, a measurement system analysis (gage R&R) could be performed if we replicate the center point experiments in order to evaluate the resolution and capability of the analytical method used.

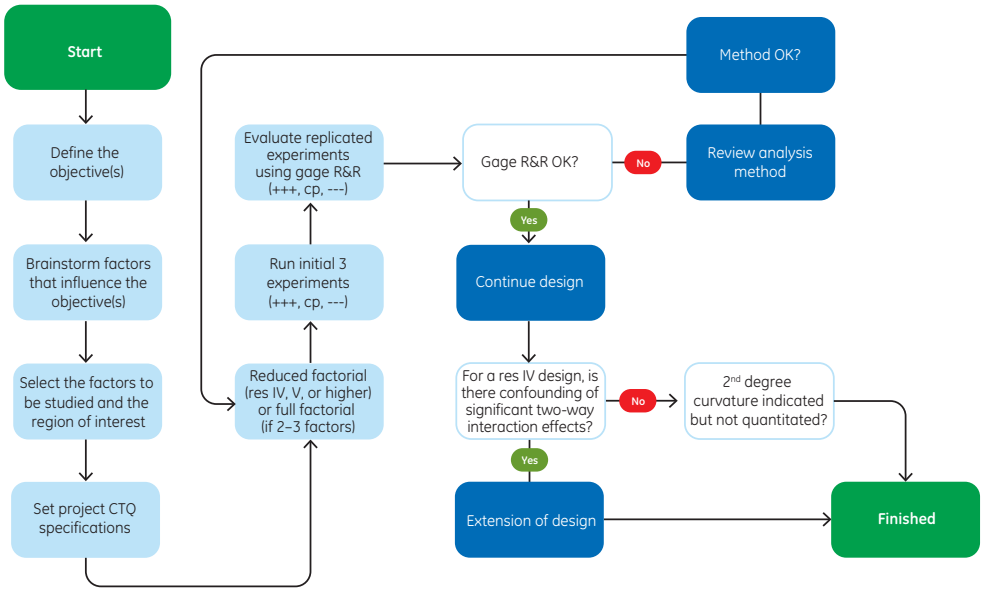


Fig 6.4. Performing experiments in a controlled way is about introducing variability. We need to compare the variability from the actual process with the variability of our measurement system. This alternative DoE route avoids wasting resources on performing a large study prior to having the variability measures under control. It also includes the possibility to detect, but not to quantitate, confounding of two-factor interaction effects and curvature, which would require an extension of the original design.

Design descriptions

Table 6.1 shows a summary of available designs in the UNICORN software. There are a number of different types of designs that can be applied in DoE. The designs are all based on multiple experiments using different settings (levels) of selected factors. The settings are combined in a highly structured way with an orthogonal pattern.

Table 6.1. Design types that are supported by the UNICORN software

Screening/robustness testing	
Fractional factorial designs	Two-level designs that are balanced subsets, or fractions, of the full factorials. The resolution of the design depends on the size of the subset (the number of runs selected). The possible resolutions are: • Resolution III designs, where main effects are confounded with two-factor interactions. • Resolution IV designs, where two-factor interactions are confounded with each other. • Resolution V designs, where main effects and all two-factor interactions are not confounded. With both resolution III and IV designs, you can only select the linear model. You may edit the model and enter selected interactions but in this case, you might have to edit the generators of the design.
L-designs	Sometimes we can have main effects and nonlinear relationships, but no two-factor interactions. In such cases, we need to consider designs more suitable for this purpose, such as the L9-design. Different variants are available in UNICORN software: L9 Fractional design at three levels for up to four factors. You can estimate square terms but not all interactions. L18 Fractional design with one factor at two levels and with up to seven factors at three levels. L27 Fractional design at three levels for up to 13 factors. You can estimate square terms but not all interactions. L36 Fractional design at three levels for up to 13 factors. You can estimate square terms but not all interactions. L-designs are useful when performing screening or robustness testing or when we suspect a nonlinear cause-and-effect relationship but no interactions present.
Plackett-Burman design	Suitable for screening many factors (main effects) but should be used only if there is no need to estimate any two-factor interactions or nonlinearity.
Screening	
Full factorial two-level designs	An orthogonal (balanced) design with all combinations of the factor levels. Main effects and all interactions are estimated and curvatures are detected but not quantitated. Useful when performing screening or robustness testing.
Optimization	
Rechtschaffner design	A saturated fraction of the 2^n and 3^n factorial designs that supports all the first-order interactions and quadratic terms. Useful when performing optimization with at least three factors in the experimental plan.
Full factorial three-level designs	A full factorial design with every factor varied at three levels. The full quadratic model can be estimated. Can be used for screening, optimization, or robustness testing, although not the primary choice for screening or robustness testing.
Central composite circumscribed (CCC) design	Composed of a full factorial design and star points. Useful when performing optimization.
Central composite face (CCF) design	Composed of a full or fractional factorial design and star points placed on the side faces. Useful when performing optimization.
Box Behnken design	A three-level (RSM) optimization design, with all design points (except the center points) located at the center of each edge of the hypercube, and on the surface of a sphere. The full quadratic model can be estimated. Useful when performing optimization and there are at least three factors in your experimental plan.
Doehlert design	An (RSM) optimization design constructed from regular polygons (e.g., hexagon) Useful when performing optimization.

Full factorial design

For studying the effect of two factors (e.g., conductivity and pH) on process outputs (response variables), and including all combinations of high and low settings for both of these factors, a full factorial design can be used. It is useful to think of a DoE study as taking place in a design space (optimization experiments). A visualization of a design space for a full factorial design using two variables would be a square with four corner experiments (Fig 6.5). The corners represent all combinations of the two factors at a high and a low level. A full factorial design also includes replicated center points in between the high and the low level ([high + low]/2) for both factors. The center point experiments are repeated at least three times and their main function is to measure variability. Center points will also detect curvature but cannot assign a specific factor level as the cause of the curvature.

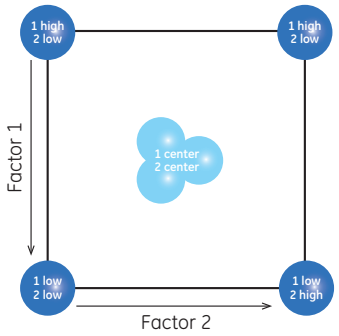


Fig 6.5. Two-factor full factorial design with three identical center-point experiments.

A visualization of the design space for three factors would be a cube with each corner representing an experiment (Fig 6.6). Again, at least three center points are included in the design. It is important that the levels (low/high) of the factors are set correctly so that the design space encompasses relevant areas.

Sometimes factor settings are not relevant or cannot be applied, for example, temperatures at which the protein of interest will denature or pH values that causes the protein to restructure.

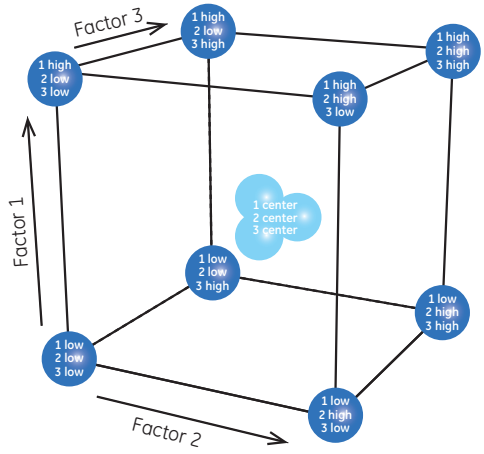


Fig 6.6. Three-factor full factorial design with three identical center-point experiments.

Increasing the number of factors from two to three doubles the number of experiments (excluding the center points) from four to eight. Adding yet another factor will require 16 experiments. The number of experiments can be calculated by the formula $N = 2^k$ where k is the number of factors and N is the number of experiments:

2 factors: $2^2 = 4$ experiments
3 factors: $2^3 = 8$ experiments
4 factors: $2^4 = 16$ experiments
5 factors: $2^5 = 32$ experiments
6 factors: $2^6 = 64$ experiments

Full factorial designs support linear effects and all interactions so that each factor can be evaluated separately. With full factorial designs, we rarely use more than four factors, as the number of experiments increases rapidly. For a two-level full factorial design the number of experiments is:

$N = 2^k + 3$

The center point allows detection of curvature, and is usually run in triplicate to estimate the noise.

2 factors: 22 = 4 experiments + center points (3) = 7
3 factors: 23 = 8 experiments + center points (3) = 11
4 factors: 24 = 16 experiments + center points (3) = 19
(5 factors: 25 = 32 experiments + center points [3] = 35)
(6 factors: 26 = 64 experiments + center points [3] = 67)

Fractional factorial design

The visualization of an experiment with more than three factors is more complex and not easily described as the dimensions required for the visualization equals the number of factors, but the principle of corner experiments is the same regardless of the number of factors. For studies where four or more factors are of interest, such as in a robustness test or a screening study, it is quite common to employ fractional factorial designs. As the name suggests, the fractional factorial design is a fraction of a full factorial design. A fractional factorial design is constructed in a way that it will still be possible to identify main effects without acquiring the detailed information that a full factorial design provides. The experiments are selected by using a symmetrical selection of corners, diagonals, and opposite diagonals. The half fraction of a three-factor experiment requires four trials. The design includes two sets of four trials each, exactly mathematically equivalent, where either set can be chosen (Fig 6.7). A full factorial design is easily completed using the other set of factor settings. The volume encompassed by either of the sets of four points is the maximum that could be encompassed by any set of four points. The design is balanced, meaning each factor is run the same number of times at each level.

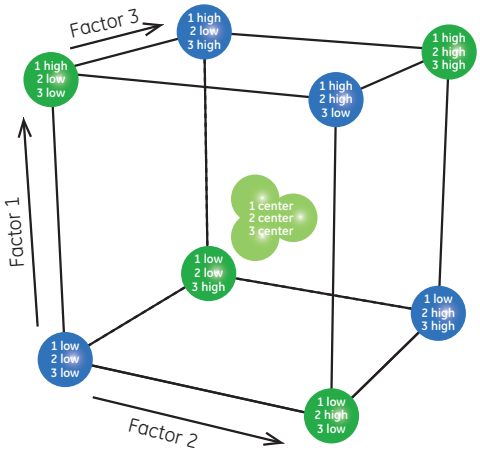


Fig 6.7. Example of a three-factor fractional factorial design. The blue and green circles are separate design sets that are mathematically equivalent in terms of experimental design and either set can be chosen.

In general, a fractional factorial design can be designated as $N = 2^{k-p}$, where N is the number of experiments, k is the number of factors to be investigated, and p the size of the fraction (1 = ½, 2 = ¼, 3 = 1/8, etc.). Hence, $N = 2^{4-1}$ signifies that four factors will be investigated in $2^3 = 8$ runs. Some fractional factorial designs can quantitate two-factor interaction effects. In the three-factor fractional factorial design displayed in Figure 6.7, four of the eight experiments from the full factorial design are selected in such a way that linear effects (main effects) are quantitated fully. A three-factor fractional factorial design would be effective for screening or robustness testing and could even be a first step in an optimization study. To be able to quantitate potential interaction effects, however, the remaining four corner points would also need to be run (effectively creating a full factorial design).

The strength of the fractional factorial design is that it allows screening of many factors using relatively few experiments. However, the drawback with fractional factorial design is that effects are confounded. Depending on the resolution (Res), the fractional factorial design supports linear and interaction effects.

Plackett-Burman design

Plackett-Burman is the most common screening design (Fig 6.8). This designs, with Res III using N = 4 × k number of runs to investigate up to (N-1) X factors, can only be used to fit linear models. However, these models are in general heavily confounded by interaction effects. Assuming these interactions are negligible, the Plackett-Burman design can be used for efficient detection of large main effects.

The Plackett-Burman design is highly useful in robustness testing designs for many factors (x). The design allows testing of up to 11 factors for main effects in 12 runs with center points.

Exp. no.	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	1	-1	1	-1	-1	-1	1	1	1	-1	1
2	1	1	-1	1	-1	-1	-1	1	1	1	-1
3	-1	1	1	-1	1	-1	-1	-1	1	1	1
4	1	-1	1	1	-1	1	-1	-1	-1	1	1
5	1	1	-1	1	1	-1	1	-1	-1	-1	1
6	1	1	1	-1	1	1	-1	1	-1	-1	-1
7	-1	1	1	1	-1	1	1	-1	1	-1	-1
8	-1	-1	1	1	1	-1	1	1	-1	1	-1
9	-1	-1	-1	1	1	1	-1	1	1	-1	1
10	1	-1	-1	-1	1	1	1	-1	1	1	-1
11	-1	1	-1	-1	-1	1	1	1	-1	1	1
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
13	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0

Fig 6.8. The Plackett-Burman design.

Rechtschaffner screening design

Rechtschaffner screening design, with Res V using a minimum number of runs, gives full support for linear effects and two-factor interactions.

Rechtschaffner screening design can have some degree of correlation between model terms, still with an acceptable condition number. The Rechtschaffner screening design can be used if the number of experiments needs to be kept to a minimum.

L-designs

L-designs are a family of designs that supports linear and quadratic model terms, but not all interactions. This design model is not a suitable choice in downstream optimization studies, as interaction effects often are present. For screenings, however, L-designs can be useful as they support identification of curvature effects:

- L9 design: up to 4 factors in 9 experiments
- L27 design: up to 13 factors in 27 experiments
- L36 design: up to 13 factors in 36 experiments

Composite factorial design for optimization and response surface modeling (RSM)

The designs displayed in Figure 6.9 can be used for RSM.

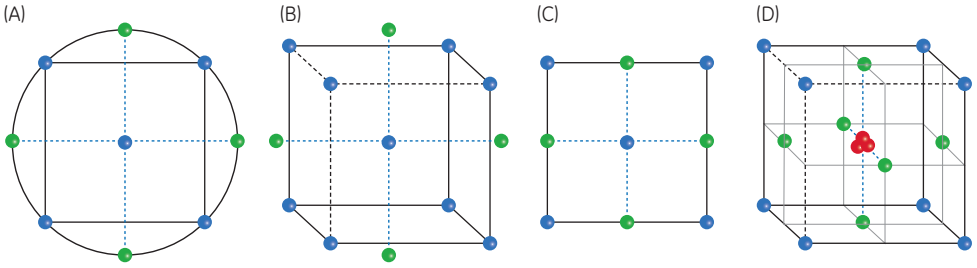


Fig 6.9. Central composite designs. (A) Two-factor central composite circumscribed (CCC), (B) three-factor CCC, and (C) two-factor central composite face-centered (CCF) design, and (D) three-factor central composite face-centered (CCF) design.

In Figure 6.9A, the blue corner points represent a two-factor full factorial design. The four green points can be seen as arranged in two pairs: one pair where factor 1 goes from low to high and one pair where factor 2 goes from low to high. Together with the center points, this will give us two series of runs where each factor is changed over three different settings (low, center, and high), while the other factor is set at its center. This design enables us to quantitate any second-degree curvature effects.

The number of star points that are needed to support identification of second-degree curvature is $2 \times x_n$ (number of factors). Figure 6.9C shows an optimization DoE using three factors with eight corner points in the full factorial part (enabling identification of linear and interaction effects) and an additional six star points that enable identification of second-degree curvature effects from the three factors.

The distance of the star points from the center should ideally be greater than the high/low settings in the factorial part (automatically handled by the DoE software), generating a central composite circumscribed (CCC) design. Setting the star points at the same distance as the corner points in the factorial part will generate a central composite-face centered (CCF) design.

A CCC design can be considered slightly more accurate than a CCF design, as the CCC design is more extended. Hybrids of CCC and CCF can be made to allow selection of certain points.

Check correlation matrix after selection of the design.

Box-Behnken RSM design

In the Box-Behnken design, experiments are performed on the edges instead of in the corners (Figure 6.10). This design avoids the corner settings with all factors simultaneously at high/low. Instead, the Box-Behnken design supports linear, interaction and quadratic effects for all model terms.

The Box-Behnken design is suitable for three to seven factors and is especially useful for investigations of many (five to seven) parameters. This design is also suitable to use when some corner-point settings are not feasible because of process limitations. For example, if the factors are temperature, pressure, and time, it might not be possible to simultaneously set all three factors at their high levels.

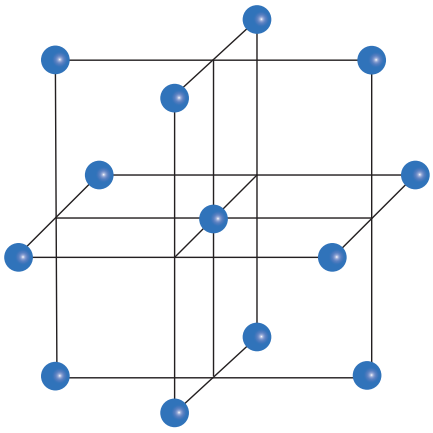


Fig 6.10. Box Behnken optimization design.

Three-level full (RSM) design

The three-level full factorial design shown in Figure 6.11 is available but rarely used. The use of a design that requires factors to be set at three levels anticipates that those factors will all have significant nonlinear effects, which is almost never the case. The three-level full factorial design can be used for response surface modeling.

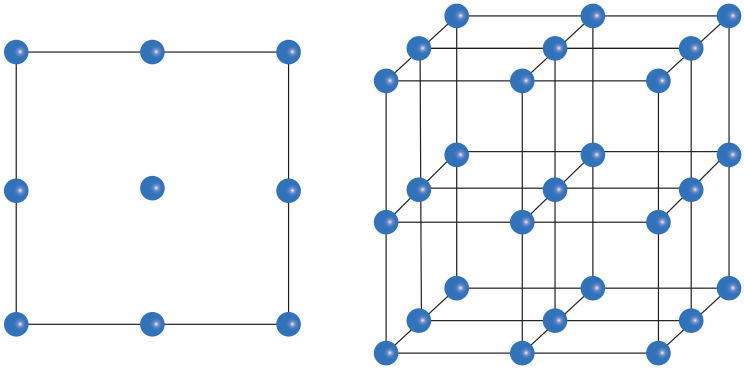


Fig 6.11. Three-level full factorial designs.

Rechtschaffner RSM design

The Rechtschaffner RSM design, with Res V, supports second-degree curvature using a minimum number of runs. The design supports linear, interaction, and quadratic effects for all model terms and can be used if the number of experiments needs to be minimized. The Rechtschaffner RSM design can have some degree of correlation between model terms, still with an acceptable condition number.

Doehlert RSM design

The Doehlert design is suitable for explorative DoE studies and optimization experiments using relatively few experiments (Fig 6.12). This design supports linear, interaction, and curvature effects and is easily expanded in any direction.

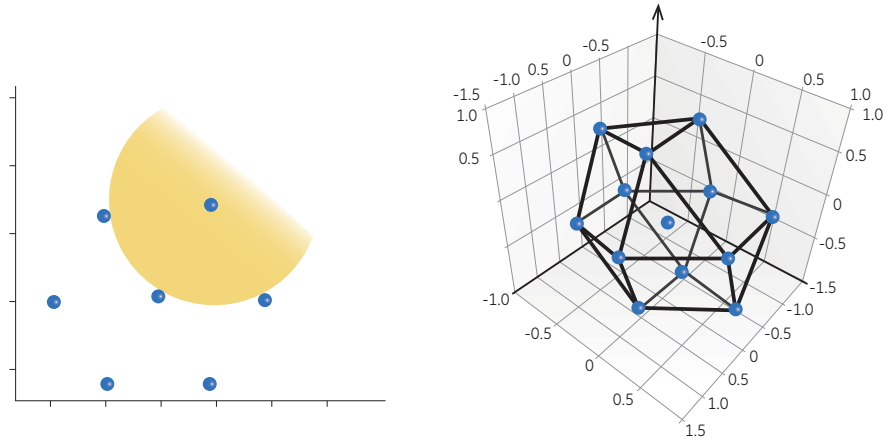


Fig 6.12. The Doehlert design.

Similar to other DoE software, the Doehlert design in the UNICORN software does not place the corner-point experiments at the user-defined high and low settings, which makes Doehlert designs slightly more sensitive to experimental noise. To compensate, it is recommended to place the high/low settings further apart for Doehlert designs than for other designs, especially if three or more parameters are investigated (Fig 6.13).

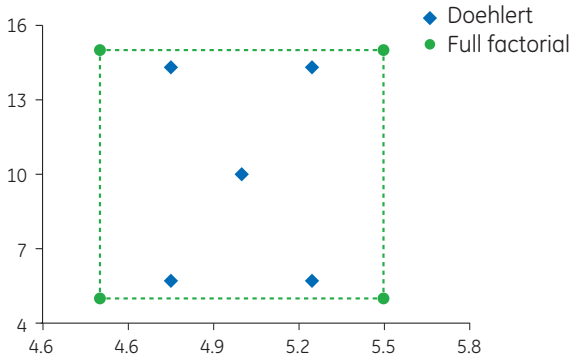


Fig 6.13. Comparison of Doehlert and full factorial designs. The Doehlert design does not include the user-defined high and low settings, which increases its sensitivity to noise. To compensate, it is recommended to place the high and low settings further apart when using a Doehlert design.

Mixture design

Screening and optimization of mixtures, in which the change in concentration of one component significantly affects the concentration of the other components, require mixture designs. Although exhibiting some unique features, many of the concepts that are valid for other designs are also valid for mixture designs. Mixtures requiring this design are common in chemical manufacturing, but less important in protein purification applications.

D-optimal design

The designs described so far can be used when working with regular experimental regions where the designs have regular geometries, such as squares, cubes, and hypercubes. In D-optimal design, the factors are varied without experimental restrictions. If the process has experimental restrictions, for example, if all factor settings in a tentative design cannot be used or where the outcome of the experiment is not acceptable, the experimental region is considered irregular. Irregular experimental regions can be addressed by reducing the design factor ranges to enable incorporation into a smaller region that can be regular, or by applying a D-optimal design (Fig 6.14). A D-optimal design can be used in both screening and optimization experiments. Generally, a larger number of experiments is needed.

The D-optimal RSM design supports linear effects, two-factor interactions, and curvature. This design can be used for:

- Asymmetric experimental regions
- Screening of multilevel qualitative factors
- Optimization of qualitative factors
- When the number of runs is to be minimized
- Updating of a model (can be used after screening for separation of confounding two-factor interactions)
- Inclusion of already performed experiments
- Both process and mixture factors in the design

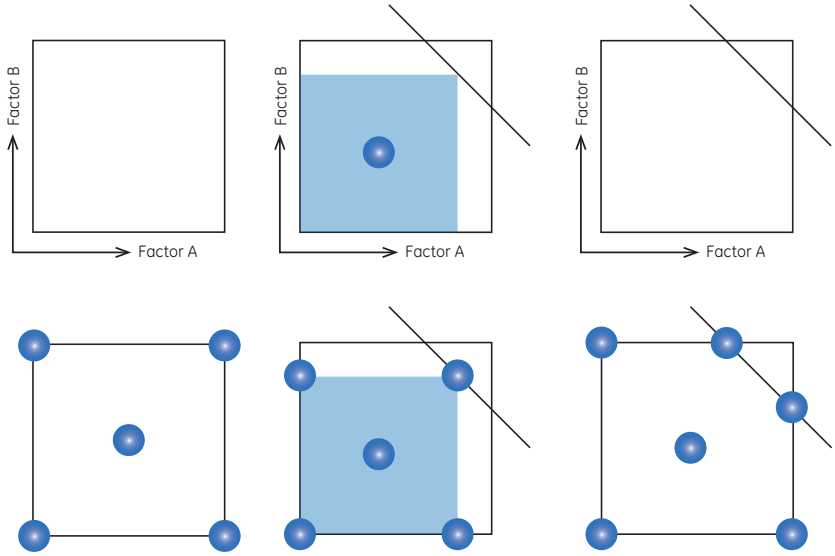


Fig 6.14. Regular and irregular experimental regions. If the experimental region is irregular, the factor ranges can be reduced to enable incorporation into a smaller regular region.

Calculating coefficients

The model polynomial in a popular format, including constant, linear, two-factor interaction, and quadratic terms.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + e$$

Constant | Linear terms (main effects) | Two-factor interaction terms | Quadratic terms | Residual (error)

A more generic expression of a second-degree model in k variables is:

$$Y = \beta_0 + \Sigma_{i=1}^{k} \beta_i X_i + \Sigma_{i=1}^{j} \Sigma_{j=1}^{k} \beta_{ij} X_i X_i + \Sigma_{k=1}^{k} \beta_{ii} (x_i)^2 + \text{error}$$

Constant | Linear terms (main effects) | Two-factor interaction terms | Quadratic terms | Residual (error)

x	individual data points.
i, j	refers to the individual observations in the data matrix, row, and column
e = error	unexplained variation
β	regression coefficients

The calculation and the number of unknown beta coefficients is:

$$\text{Number of } \beta \text{ coefficients} = 1 + 2k + \frac{k(k-1)}{2}$$

The standard methods for the regression analysis and of finding coefficients include matrix calculations:

$Y = X\beta + e$ (the matrix form of the model)

The ordinary least squares estimator of the vector of unknown model coefficient β is:

$$\beta = (X^T X)^{-1} X^T Y$$

$$\text{Size of } X = 2^k + 2k + n_0 \text{ by } 1 + 2k + \frac{k(k-1)}{2}$$

Estimated analytical formulas (assuming a response surface model that includes linear, interaction and quadratic terms) for the beta coefficients (assuming scaling to the interval [-1, 1]), when:

$N = 2^k + 2k + n_0$

(1) A two-level factorial arrangement (2^k points): $(x_1, x_2, \ldots x_k) = (1,1, \ldots 1)$
(2) 2k axial or "star" points: $(x_1, x_2, \ldots x_k) = (\pm a, \pm a, \ldots \pm a)$
(3) n0 center points: $(x_1, x_2, \ldots x_k) = (0, 0, \ldots, 0)$

$D = (2\alpha^4 + Mk)N - (M + 2\alpha^2)^2 k \; M = 2^k$

N is the number of data points in a design and M is the factorial part. Using the above definitions, the constant term estimate is:

$$\beta_0 = \frac{2\alpha^2(a^2 - k)}{D} \sum_{u=1}^{M} Y_u + \frac{M(k - \alpha^2)}{D} \sum_{u=M+1}^{M+2k} Y_u \frac{2\alpha^4 + Mk}{D} \sum_{u=M+2k+1}^{N} Y_u$$

The linear terms are:

$$\beta_i = \frac{1}{M + 2\alpha^2} \sum_{u=1}^{M} X_{u,i+1} Y_u + \frac{\alpha}{M + 2\alpha^2} (Y_{M+2i} + Y_{M+2i-1})$$

$1 \le i \le k$

And the interaction terms are:

$$\beta_{ij} = \frac{1}{M} \sum_{u=1}^{M} X_{u,2} X_{u,2} Y_u$$

Note that the n_0 term does not appear in either of the linear or the interaction term.

Finally the quadratic terms are given by:

$$\beta_{\ddot{U}} = \frac{(2k + n_0 - 2\alpha^2)}{D} \sum_{u=1}^{M} Y_u + \frac{M(2\alpha^2\, 2k - n_0)}{2\alpha^2 D} \sum_{u=M+1}^{M+2k} Y_u + \frac{1}{2\alpha^4} (Y_{M+2i} + Y_{M+2i-1}) - \left(\frac{M + 2\alpha^2}{D}\right) \sum_{u=M+2k+1}^{N} Y_u$$

$1 \le ii \le k$

Data transformation

When collected data or input variable values are strongly skewed, for example, when not normally distributed, or if we have a strong logarithmic relationship, we sometimes apply a mathematical function to rescale the data. The use of transformation should be justified by the actual physical behavior of a system. Indications of a response transformation need inclusion of a non-normal plot of residuals, undesired patterns in residual versus predicted plot (smile or frown), unexplained outliers, or unlikely interactions. With a response transformation, the data becomes much less skewed and eventual outliers become less extreme. The result is data that are easier to handle in a statistical modelling. Transformation of the response or input variables can improve the fit and correct violations, of model assumptions such as constant error variance (Fig 6.15).

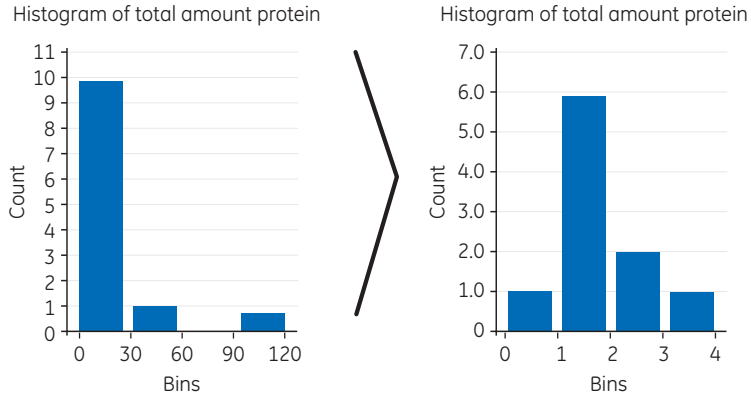


Fig 6.15. Transformation of response data from a cultivation DoE, where data becomes closer to a normal distribution.

In practice, it is usually not known how the errors enter the model: additively, multiplicatively, or otherwise. Therefore, the common approach for transformations is to try different functions and to check if the residuals satisfy the conditions required for linear regression, for example, in a normal probability plot. If we have a large number of data near zero (relative to the larger values in the data set) and all observations are positive, the data should be rescaled by a standard transformation using a logarithmic function. When a log scale is used, the regression coefficients can be interpreted in a multiplicative way rather than in the usual additive way. Other transformations include square root, inverse, linear, neg log, exponential, logit, and power functions.

Rescaling (i.e., a standard transformation) is defined by using a logarithmic function:

If $$\log \hat{Y} = \beta_0 + \Sigma_{i=1}^{k} \beta_i x_i + \Sigma_{i=1}^{j} \Sigma_{j=1}^{k} \beta_{ij} x_i x_j + \Sigma_{k=1}^{k} \beta_{ii} (x_i)^2 + \text{error}$$

then $$\hat{Y} = e^{\hat{\beta}_0} \times e^{\Sigma_i^k \hat{\beta}_i x_i} \times e^{\Sigma_i^j \Sigma_j^k \hat{\beta}_{ij} x_i x_j} \times e^{\Sigma_{k=1}^{k} \hat{\beta}_{ii} (x_i)^2} \times e^{\text{error}}$$

An increase in one input factor (x_i) will multiply the originally scaled predicted response by e^{β_i}. Using a log scale allows the regression coefficients to be interpreted in a multiplicative way rather than in the usual additive way.

Data transformation is readily performed using UNICORN software prior to establishing a DoE and entering factor values. Response values can be transformed after the chromatographic runs and added to the software for evaluation.

Blocking RSM designs

RSM designs can be blocked orthogonally when they fulfill the following two conditions:

1. Each block must be a first-order orthogonal block.
2. The fraction of the total sum of squares for each variable from individual blocks must equal the fraction of the total observations (experiments) allotted to the block.

The central composite circumscribed designs can be split into two blocks, the cube portion and the star portion, and will satisfy the above conditions when (the distance of the star points to the center) is equal to:

$$\alpha = \sqrt{[k(1 + p_s)/ (1 + p_c)]}$$

where

k	number of factors
$p_s = ns_0/ns$	proportion of center points in the star portion
$p_c = nc_0/nc$	proportion of center points in the cube portion
ns	number of star points runs
nc	number of runs from the cube portion

The cube portion of the central composite circumscribed design can be split into further blocks if the factorial or the fractional factorial part of the design can be split into orthogonal blocks of pseudo-resolution V and each block has the same number of center points. Box Behnken designs can be orthogonally blocked, as specified by Box and Behnken 1960 and Box and Draper 1987 (see Chapter 8). Central composite face designs cannot be blocked.

Appendix 2
Terminology

Six Sigma terminology and selected formulas

Accuracy

The differences between observed average measurement and the "truth" (the "truth" often obtained using a standard)

Alternative hypothesis

A tentative explanation which indicates that an event does not follow a chance distribution; a contrast to the null hypothesis.

ANOVA (analysis of variance)

is statistical method for evaluating the effect that factors have on process mean and for evaluating the differences between the means of two or more normal distributions. ANOVA is used in conjunction with DoE and fractional-factorial experimental designs. ANOVA includes the following calculation steps:

1. Sum of squares:

$$\text{Total sum of squares} = \text{SST} = \Sigma\,(x - \bar{x})^2 = \Sigma x^2 = \frac{(\Sigma x^2)}{n}$$

$$\text{Regression sum of squares} = \text{SSTR} = \sum_{j=1}^{N} n_j\,(\bar{x} - \bar{\bar{x}})^2 = \Sigma\,\frac{T_j^2}{n_j} - \frac{(\Sigma x^2)}{n}$$

$$\text{Error sum of squares} = \text{SSE} = \text{SST} - \text{SSTR} = \Sigma(n_j - 1)s_j^2$$

2. Mean square variance

$$\text{Mean squares variance regression} = \text{MSTR} = \frac{\text{SSTR}}{k - 1}$$

$$\text{Mean squares variance error} = \text{MSE} = \frac{\text{SSE}}{n - k}$$

3. Test statistic for one-way ANOVA (assuming independent samples, normal populations, and equal population standard deviations):

$$F = \frac{\text{MSTR}}{\text{MSE}}$$

with degrees of freedom.df = (k – 1, n – k) = (no. of rows)(no. of columns)

k	number of populations
n	total number of observations
$\bar{\bar{x}}$	(grand) mean of all n observations
n_j	size of sample from Population j
$\bar{x}$	mean of sample from Population j
s_j^2	variance of sample from Population j
T_j	sum of sample data from Population

Average

The average (central point value) calculated for a data set can be: the *Mean*, the value obtained from the sum of all values divided by the number of values; *Median*, the middle value in a sorted data set; *Mode*, the value that appears most often; or *Midrange*, the center value midway between the highest and lowest value in the data set.

Balanced factorial designs	When each level occurs equally often within each factor, which means that the intercept is orthogonal for each effect.
Blocking variables	A relatively homogenous set of conditions within which different conditions of the primary variables are compared. Used to ensure that background variables do not contaminate the evaluation of primary variables.
Experimental design	A set of predetermined levels of selected factors (process or system variables) for conducting experiments and identification of patterns and regularities in observational and experimental results.
Cause	That which produces an effect or brings about change.
Cause-and-effect diagram	A schematic sketch, usually resembling a fishbone, which illustrates the main causes and sub causes leading to an effect (symptom). Also known as a Fishbone diagram.
Characteristic	A definable or measurable feature of a process, product, or variable.
Combinatorics	A branch of mathematics dealing with combinations and permutations.
Composite	From the Latin compositus (put together).
Condition number	A measure of the sphericity or orthogonality of a design. If we describe the design as a matrix X consisting of -1's and +1's, the condition number is the ratio between the largest and smallest eigenvalue of a X'X matrix. All factorial designs without center points (the midpoint between the + and – levels) have a condition number 1 and all points are located on a sphere.
Confidence level	The probability that a randomly distributed variable "X" lies within a defined interval of a normal curve.
Confidence limits	The two values that define the confidence interval.
Confidence interval	The most likely range of the unknown population average or percentage. Confidence intervals does not give a likely range of all values; the interval implicates how much the average value is likely to fluctuate.
Confounding	Allowing two or more variables to vary together so that it is impossible to separate their unique effects.
Control chart	A graphical rendition of a characteristic's performance across time in relation to its natural limits and central tendency.
Continuous data	Data obtained from a measurement system that has an infinite number of possible outcomes.
Control chart	A graphical rendition of a characteristic's performance over time in relation to its natural limits and central tendency.
Control limits	Apply to both range and standard deviation and subgroup average (X) portions of process control charts and are used to determine the state of statistical control. Control limits are derived statistically and are not related to engineering specification limits in any way.
Correlation	The determination of the effect of one variable upon another in a dependent situation. The relationship between two sets of data such that when one changes, the other is likely to make a corresponding change. Also a statistical tool for determining the relationship between two sets of data.

Correlation coefficient; $r = \frac{1}{n-1} \Sigma_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_x} \right) \left(\frac{y_i - \bar{y}}{S_y} \right)$

Critical to quality (CTQ) An element of a design or a characteristic of a part that is essential to quality from the customer view.

Data Factual information used as a basis for reasoning, discussion, or calculation; often refers to quantitative information.

Degrees of freedom The number of independent measurements available for estimating a population parameter.

Design space The experimental space that yields results within the set of responses specified for the process.

Design of experiments A formal, proactive method of changing process parameters (i.e., controlled variables or factors), as well as establishing blocks, replications, and response variables associated with a planned experiment. Includes analyzing the resulting process outputs in order to quantitate the cause and effect relationship between them as well as the random variability of the process while using a minimum number of runs. It is the plan for conducting the experiment and evaluating the results in order to economically improve product and process quality. A major tool used in Six Sigma methodology.

Discrete data Data obtained from a measurement system that has a finite number of possible outcomes.

Distributions Tendency of large numbers of observations to group themselves around some central value with a certain amount of variation or "scatter" on either side.

Effect Experimental designs test if a factor influences the response. This cause –and-effect relationship is called "effect". There are three different types of effects: the factor effects the response directly (main effect), via an interaction with another factor, or by both mechanisms simultaneously. The significance of an effect is determined statistically with a probability (e.g., 95%) or a risk measure (usually 5%).

Error Once an error estimate is made, it is given the name "noise".

Experiment A test under defined conditions to determine an unknown effect; to illustrate or verify a known law; to test or establish a hypothesis.

Experimental error Variation in observations made under identical test conditions. Also called residual error. The amount of variation which cannot be attributed to the variables included in the experiment.

Experiments Manipulate some aspect of nature and observe the outcome.

F distribution Associated with hypothesis testing of standard deviation between two or more process distributions.

Factor These are the inputs (also called x's, independent variables, explanatory variables) for the experiment, which is then conducted in order to observe the measurable response or output. Factors are true variables when they are allowed to assume two or more values, called factor levels, during the course of experimentation.

Factorial (Full factorial and fractional factorial)	Fractional factorial experiments are experimental test programs that allow an engineer or technician to obtain statistically valid data using only a small fraction of the available test combinations.
Fishbone diagram	A schematic sketch, usually resembling a fishbone, which illustrates the main causes and subcauses leading to an effect (symptom). Also known as cause-and-effect diagram.
Failure mode effects analysis (FMEA)	A risk assessment technique useful for developing process knowledge when relating the effect of CTQ and other parameters of a process. The FMEA provides a team management tool for conducting quality risk analyses. FMEA generates a risk priority number (RPN) referring to the likelihood of occurrence of a quality defect. RPN is a measure of the ability to detect a quality defect (detectability) and the consequence of the quality defect (severity). In this analysis, we are converting qualitative information to quantitative numbers that tell us about the quality of our process. Usually, FMEA is performed at a later stage when we are capable of testing the robustness of a process and the probability of assessing failure in our quality. At the earlier stages of development, we are more concerned with finding which process parameters affect the overall quality of a process (cause-and-effect analysis).
Gage accuracy	The average difference observed between a gage under evaluation and a master gage when measuring the same parts over multiple readings.
Gage linearity	A measure of gage accuracy variation when evaluated over the expected operating range.
Gage repeatability	A measure of the variation observed when a single operator uses a gage to measure a group of randomly ordered (but identifiable) parts on a repetitive basis.
Gage reproducibility	A measure of average variation observed between operations when multiple operators use the same gage to measure a group of randomly ordered (but identifiable) parts on a repetitive basis.
Gage stability	A measure of variation observed when a gage is used to measure the same master over an extended period of time.
Gage repeatability and reproducibility (gage R&R)	A measurement system evaluation to determine equipment variation and appraiser variation. This study is critical to ensure that the collected data is accurate.
Histogram	Vertical display of a population distribution in terms of frequencies; a formal method of plotting a frequency distribution.
Hypothesis	A “tentative guess” about how the world works. Often several hypotheses are formed at once "multiple working hypotheses".
Independent variable	A controlled variable; a variable whose value is independent of the value of another variable in separate experiments.
Induction (inductive reasoning)	Generalizing from individual observations...to general conclusions.
Interaction	When the effects of a factor A are not the same at all levels of another factor B.

Lower control limit A horizontal dotted line plotted on a control chart which represents the lower process limit capabilities of a process.

Margin of error; $E = Z\alpha_{/2} \times \frac{\sigma}{\sqrt{n}}$

Mean μ (population), $\mu = \Sigma\, x/N \quad \bar{x}\ (\text{sample}),\ \bar{x} = \Sigma\, x_i/n$

Measurement systems analysis (MSA) Means of evaluating a continuous or discrete measurement system to quantitate the amount of variation contributed by the measurement system.

Model (or mathematical model) An abstract model that uses mathematical language.

Noise Another word for the total experimental (random) error that exits in any process from which measurements are taken - whether by DoE or "one-factor-at-a-time" (OFAT) experimentation. For example, if you were to repeat the same experiment over and over, the results would vary - in spite of your best attempts to keep everything "constant." If you were to then calculate the standard deviation of those results, that would be a measure of the "noise." In Statistical Process Control (SPC), the noise is referred to as common cause variation.

Normal distribution Described by $Y = \frac{1}{\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma}$ is a continuous symmetrical density function characterized by a bell-shaped curve, for example, distribution of sampling averages. The standard normal distribution is a continuous bell-shaped curve, with a mean of zero and a spread of 1, that is, μ=0 and σ=1 and an area under the curve that equals 1. This curve is obtained when displaying data in a frequency polygon showing the frequency of instances of each value, from lowest to highest. For any series of observations , that is, other normal distributions, the standard deviation (also known as the z-score) describes the variation in data compared to the mean as a percentage of the whole. This means that 1 standard deviation represents 68%, 2 represents 95% and 3 represents 98% of the data. Values that fall far outside the normal range could be indicative of underlying problems or extraordinary successes that should be investigated. Other distributions besides the normal are sometimes relevant and can lead to a higher success rate in data interpretation. In nature, any observations that are influenced by many small and unrelated random effects are approximately normally distributed, this phenomenon is also known as the fuzzy central limit theorem. In statistics this is expressed as: if we have a random sample of size n from a population with a mean μ and a standard deviation, σ. Then as n gets large, the sampling distribution approaches the normal distribution with mean μ and a standard deviation $\frac{\sigma}{\sqrt{n}}$

Null hypothesis An assertion to be proven by statistical analysis where two or more data sets are stated to be from the same population.

Observation A statement about something you have noticed.

Operating space Operation of a process within the operating space will yield a product meeting the defined quality attributes.

Optimization	In the simplest case, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function.
Orthogonal	When every pair of levels occurs equally often across all pairs of factors, the design is orthogonal. More generally, a design is orthogonal when the frequencies for level pairs are proportional or equal.
Parameter	A parameter is a numerical value that is equivalent to an entire population , that is, a functional constant such as the mean of a normal distribution.
Pareto diagram	A chart which ranks, or places in order, common occurrences.
Precision	A measure of the variation of the measurement system. (Note: you can be very precise but still not accurate).
Prediction test	Use hypotheses and theories to make predictions about how a particular system will behave, and then perform experiments to see if the system behaves as predicted.
Probability	The chance of something happening; the percent or number of occurrences over a large number of trials.
Process	A particular method of doing something, generally involving a number of steps or operations.
Process capability	The relative ability of any process to produce consistent results centered on a desired target value when measured over time.
Process control chart	Any of a number of various types of graphs upon which data are plotted against specific control limits.
Process map	Flow chart for analysis of a process by breaking it down into its component steps, and then gaining a better understanding of the process, step by step.
Pseudo-center points	In experiments, factors are often given nominal codes. For example, catalyst "A" might be the -1 setting, whereas catalyst "B" is coded +1. The choice of which is setting is "high" and which is "low" is arbitrary, but which catalyst setting is "standard" needs to be decided. The standard settings for the discrete input factors, together with center points for the continuous input factors, will be regarded as the "center points" for the purpose of the design.
Quality by design (QbD)	Guidelines that enhance pharmaceutical development to facilitate design of products and processes. QbD maximizes a product's efficacy and safety profile while at the same time enhances product manufacturability.
Random	Selecting a sample so each item in the population has an equal chance of being selected; lack of predictability; without pattern.
Random cause	A source of random variation; a change in the source will not produce a highly predictable change in the response (dependent variable), for example, a correlation does not exist; any individual source of variation results in a small amount of variation in the response; cannot be economically eliminated from a process; an inherent natural source of variation.

Random variation	Variations in data which result from causes that cannot be pinpointed or controlled.
Regression analysis	A statistical technique for determining the relationship between one response and one or more independent variables.
Resolution	The resolution identifies which effects (possibly including interactions) we can estimate and the ability of a design to resolve these effects.
Repeatability	Variation when one person repeatedly measures the same unit with the same measuring equipment.
Replication	Repeat observations made under identical test conditions.
Reproducibility	Variation when two or more people measure the same unit with the same measuring equipment.
Response	The output of a process.
Robust	The condition or state in which a response parameter exhibits a high degree of resistance to external causes of a nonrandom nature; impervious to perturbing influence.
Screening	A DoE that identifies which of many factors or conditions used that have a significant effect on the response.
Scientific method	(1) Statement of problem, (2) hypotheses about the cause of the problem, (3) experiments designed to test each hypothesis, (4) predicted results of experiments, (5) observed results of experiments, or (6) conclusions from the results of experiments.
Signal	The associated estimates of factor and interaction strengths are called “signals. The "signal," or signals, are the effects that you are trying to discover and measure - in DoE terms, signals are the main effects and interactions. Significant ANOVA will allow us to statistically examine the magnitude of signal-to-noise ratios in order to see if factor/interaction strengths are truly significant.
Sigma (σ)	Standard deviation; an empirical measure based on the analysis of random variation in a standard distribution of values; a uniform distance from the mean or average value such that 68.26% of all values are within 1 sigma on either side of the mean, 95.44% are within 2 sigma, 99.73% are within 3 sigma, 99.9% are within 4 sigma and so forth.
Sigma level	A statistical estimate of the number of defects that any process will produce equivalent to defects per million opportunities for that process.
Six sigma	A collection of tools and techniques for raising quality to worked-class levels. A combination of verified customer requirements reflected in robust designs and matched to the capability of production processes that creates products with fewer than 3.4 defects per million opportunities to make a defect. World-class quality. Specification limits represent how much variation is tolerated in the product or process.
Stable process	A process which is free of assignable causes, for example, in statistical control.

Standard deviation

The typical difference between each value and the mean value. Describing how broadly the sample values are distributed. Standard deviation is calculated from the square root of variance, thought of as the "average" deviation from the mean. The calculations include: calculating the mean of a data set, subtracting each value from the mean for a new set of values, square each of these new values and calculating the sum. Dividing the sum with the total number of values (-1) and finally calculating the square root of this value. The principle is the same for both calculating the standard deviation for a set of values as for an entire population although for the latter we do not subtract the number of values by 1:

Population: $\sigma = \sqrt{\frac{\Sigma_{i=1}^{N}(x_i - \mu)^2}{N}}$

Sample: $s = \sqrt{s^2} = \frac{\Sigma_{i=1}^{n}(x_i - x)^2}{n-1}$

Standard error of the mean

An estimate of variable means if the experiments are repeated multiple times. Inferring population mean, or whether samples sets are likely to come from the same population.

Statistic

Is a numerical value that represents a sample of an entire population, that is, a measured value of a parameter such as an average.

Statistical control

A quantitative condition that describes a process that is free of assignable/special causes of variation, for example, variation in the central tendency and variance. Such a condition is most often proven on a control chart.

Statistical process control

The application of statistical methods and procedures relative to a process and a given set of standards.

Symbols

Symbols used for sample statistics and population parameters:

Name	Sample statistic	Population parameter
Mean	$\bar{x}$	μ
Variance	s^2	σ^2
Standard deviation	s	σ
Correlation	r	p
Regression coefficient	b	β
Number of observations	n	N

Test of significance

A statistical procedure used to determine whether or not a process observation (data set) differs from a postulated value by an amount greater than that due to random variation alone.

T distribution

Associated with hypothesis testing of the means (averages) between two distributions (when sample sizes are less than 100).

Test statistic

A quantity calculated from our sample of data. Its value is used to decide whether or not the null hypothesis should be rejected in our hypothesis test. The choice of a test statistic will depend on the assumed probability model and the hypothesis.

Theory Refers to a description of the world that covers relatively large numbers of phenomena and has met observational and experimental tests.

Transfer function Transfer functions defines the dependency of the system performance on subsystem performance and of the subsystems on their components. Design of experiments and regression, and/or other analytical, engineering, numerical, and statistical methods are used to find relationships among a system CTQ's.

Transformation A mathematical technique used to create a near normally distributed data set out of a non-normal (skewed) data set.

Uncontrolled variables Variables that are of no experimental interest and are not held constant. Their effects are often assumed insignificant or negligible, or they are randomized to ensure that contamination of the primary response does not occur.

Upper control limit A horizontal line on a control chart (usually dotted) that represents the upper limits of process capability.

Variable A characteristic that may take on different values.

Variance The average of the squared deviations between values and the mean:

Population: $\sigma^2 = \frac{1}{N} \Sigma_{i=1}^{N} (x_i - \mu)^2$

Sample: $s^2 = \frac{\Sigma_{i=1}^{N} (x_i - \bar{x})^2}{n-1}$

Variation Any difference that can be quantitated between individual measurements; such differences can be classified as being due to common causes (random) or special causes (assignable). Variation in data is usually expressed in single numbers by looking at the Spread and Standard deviation. The *Spread*, a measure of how far from the center the data tend to range. The value is obtained by subtracting the lowest from the highest value or by dividing the data into groups to see how far the extreme groups are located from the median. The *Standard deviation* measures spread in relation to the mean where a higher value roughly indicates a higher average distance in the data from the mean of the data. This value can be used to compare the variation between different data sets.

X's Designation in Six Sigma terminology for variables that are independent root causes, as opposed to Y's, which are dependent outputs of a process. Six Sigma focuses on measuring and improving X's to see subsequent improvement in Y's.

X & R charts A control chart, which is a representation of process capability over time; displays the variability in the process average and range across time.

Y's Designation in Six Sigma terminology for variables that are dependent outputs of a process, as opposed to X's. which are independent root causes.

Chromatography terminology

Adsorption	Binding. The process of interaction between the solute (for example, a protein) and the stationary phase.
Affinity chromatography	A group of methods based on various types of specific affinities between target molecule(s), for example, a protein and a specific ligand coupled to a chromatography medium.
Asymmetry factor	Factor describing the shape of a chromatographic peak.
Backpressure	The pressure drop across a column and/or a chromatography system.
Band broadening	The widening of a zone of solute (for example, a protein) when passing through a column or a chromatography system. Gives rise to dilution of the solute and reduces resolution. Also often called peak broadening or zone broadening.
Binding	Adsorption. The process of interaction between a solute (e.g., a protein) and the stationary phase.
Binding buffer	Buffer/solution/eluent used for equilibration of the column before sample loading.
Binding capacity	The maximum amount of material that can be bound per milliliter of chromatography medium. See also Dynamic binding capacity.
Capacity factor	The degree of retention of a solute (e.g., a protein) relative to an unretained peak.
Chromatofocusing	Method that separates proteins on the basis of pI.
Chromatogram	A graphical presentation of detector response(s) indicating the concentration of the solutes coming out of the column during the purification (volume or time).
Chromatography	From Greek chroma, color, and the verb graphein, which means to write.
Chromatography medium/media	The stationary phase, also called resin. The chromatography medium is composed of a porous matrix that is usually functionalized by coupling of ligands to it. The matrix is in the form of particles (beads) or, rarely, a single polymer block (monolith).
Cleaning in place (CIP)	Common term for cleaning chromatography columns and/or systems, with the purpose of removing unwanted/nonspecifically bound material.
Column	Usually column hardware packed with chromatography medium.
Column equilibration	Passage of buffer/solution through the chromatography column to establish conditions suitable for binding of selected sample components. For example, to establish correct pH and ionic strength, and ensure that proper counter ions or counter ligands are present.
Column packing	Controlled filling of the column hardware with chromatography medium to obtain a packed bed.
Column volume	The geometrical volume of the column interior/the chromatography bed.

Counter ion	Ion of opposite charge that interacts with an ion exchange chromatography medium after the column equilibration. The counter ion is displaced by a protein that binds to the ion exchanger. If a high concentration of the counter ion is applied, it will compete with the bound protein and elute it from the chromatography column.
Dead volume	The volume outside the packed chromatography bed. Can be column dead volume or chromatography system dead volume. The dead volume contributes to band broadening.
Desorption	Elution. Release or removal of bound substances from the chromatography medium.
Dynamic binding capacity (DBC)	The binding capacity determined by applying the target using flow through a column, as opposed to equilibrium binding capacity determined by batch experiment.
Efficiency	Measured as number of theoretical plates. High efficiency means that sharp peaks will be obtained.
Effluent	The mobile phase leaving the column (= eluate).
Eluate	The mobile phase leaving the column (= effluent).
Eluent	The buffer/solution used during chromatography (= mobile phase).
Elution buffer	Buffer/solution used for elution (desorption) of bound solutes (for example, proteins) from a column.
Elution volume	The volume of buffer/solution (eluent) required to elute the solute, for example, a protein (= retention volume).
Elution time	The time required for elution of a solute (protein) (= retention time).
Flow rate	Volumetric flow (mL/min) or linear flow rate (cm/h). Measurement of flow through a column and/or chromatography system.
Flowthrough	Material passing the column during sample loading (without being bound).
Gel filtration (GF)	Size-exclusion chromatography. Separates solutes (e.g., proteins) according to size.
Gradient elution	Continuous increased or decreased concentration of a substance (in the eluent) that causes elution of bound solutes (e.g., proteins).
Homogeneity	Uniformity in composition or character.
Hydrophobic interaction chromatography (HIC)	Method based on the hydrophobic interaction between solutes (e.g., proteins) and the chromatography medium in the presence of high salt concentration.
Immobilized metal ion affinity chromatography (IMAC)	Method based on the affinity of proteins with histidine, cysteine, or tryptophan amino residues on their surface and metal ions on the chromatography medium.
Ion exchange chromatography (IEX)	Method based on electrostatic interactions between solutes (e.g., proteins) and chromatography medium.
Isocratic elution	Elution of the solutes without changing the composition of the buffer/solution (eluent).
Ligand	The specific molecular group that is coupled to the matrix to give some decided function to the chromatography medium.

Ligand density	Related to ligand concentration. The distribution of ligands on the surfaces (also surfaces inside pores) of the chromatography matrix.
Linear velocity	The flow rate normalized by the column cross-section (cm/h).
Mass transfer	Movement of a solute (e.g., a protein) in and out of the stationary phase. Important factor for column efficiency.
Matrix	The matrix is the nonfunctional base for the chromatography medium. The matrix has a porous structure that provides a large surface that can be modified with ligands that introduce possibilities for protein binding.
Mobile phase	The fluid (buffer/solution) carrying the solutes during chromatography (= eluent).
NHS	NHS or NHS-activated is a method to chemically "activate" the surface of a chromatography medium. NHS = N-hydroxysuccinimide is the substance to which a ligand can be attached.
Peak broadening	Same as band broadening.
Peak capacity	The number of peaks that can be separated using a chromatography column.
Peak tailing	Broadening at the end of a peak due to additional delay of a fraction of the solute. Results in increased asymmetry factor.
Pore	Cavity in a chromatography matrix.
Pore volume	The total volume of the pores in a chromatography medium.
Purity	Determined based on measuring the level of impurities or contaminants. In chromatography, this can be expressed as an area percentage.
Pressure over the packed bed	The pressure drop across the packed bed upon passage of solution through the column. Caused by flow resistance in the packed bed.
Yield	The relative amount of target protein that is retrieved after purification compared with amount loaded on the column.
Resin	The term is sometimes used instead of the more generic term, chromatography medium.
Resolution	Measurement of the ability of a packed column to separate two solutes (peaks).
Retention volume	Same as elution volume.
Retention time	Same as elution time.
Reversed phase chromatography (RPC)	Method based on hydrophobic interactions between solutes (sample components) and ligands coupled to the chromatography medium. Organic modifiers (e.g., acetonitrile) in the eluent are used for elution.
Sample	The material loaded on the chromatography column/medium, or to be analyzed.
Sample application	Applying/loading sample on the column.
Sample loading	Loading/applying sample on the column.
Sample volume	Usually the volume of the sample loaded on the chromatography column/medium.

Selectivity	Measure of the relative retention of two solutes in a column. Related to the distance between two peaks.
Solute	The dissolved substance (e.g., a protein) in, for example, the mobile phase.
Stationary phase	Often called resin, chromatography beads, chromatography material, chromatography medium.
Step gradient elution	Stepwise increase in concentration of the substance that affects elution of bound solutes.
Void volume	The elution volume of solutes that do not enter the pores or interact with the chromatography medium, thus passing between the beads in the packed bed.
Wash	Wash step. Removal of unbound or weakly bound material from a column after the sample loading.
Wash buffer	Buffer/solution used for washing the column after sample loading.
Wash volume	Volume of buffer/solution used for the wash step.
Yield	Amount of target protein (or other solute) obtained after a purification step, or after the entire purification (multiple steps).
Zone broadening	Same as peak broadening.

Literature list

Anscombe F. J. Graphs in statistical analysis. *American Statistician* **27**, 17–21 (1973).

Bose R. C. and Kishen, K. On the problem of confounding in the general symmetrical factorial design. *Sankhya* **5**, 21 (1940).

Box, G.E.P. and Behnken, D.W. Some new three level designs for the study of quantitative variables, *Technometrics* **2**, 455–475 (1960).

Box, G.E.P. and Draper, N.R. Empirical model building and response surfaces. John Wiley & Sons, Inc., New York (1987).

Box, G.E.P., Hunter, W.G., and Hunter, J.S., Statistics for experimenters, John Wiley & Sons, Inc., New York (1978).

Box, G.E.P. and Wilson, K.B. On the experimental attainment of optimum conditions. *J. Roy. Statist. Soc. Ser. B. Methods* **13**, 1–45 (1951).

Cox, G.M. and William Gemmell Cochran, W.G. Experimental designs. 1st Ed, John Wiley and Sons Inc, New York (1950).

Myers, R.H., *et al.* Variance dispersion of response surface designs. *J. Qual. Technol.* **24**, 1–11 (1992).

Eriksson *et al.* Design of Experiments. Principles and Applications. Umetrics Academy (ISBN-10: 91-973730-4-4, ISBN-13: 978-91-973730-4-3) (2008).

Granér, T. *et al.* High capacity, small-scale antibody purification using magnetic beads. Poster presented at the ESBES conference (2010).

Haaland, D.P. Experimental Design in Biotechnology, CRC Press; 1st Ed. (1989).

Kaisermayer, C. *et al.* Biphasic cultivation strategy for optimized protein expression and product quality. Poster presented at the ESACT Conference (2013).

Kunes, J. Dimensionless physical quantities in science and engineering, Elsevier; 1st Ed. (2012).

Montgomery, D.C. Design and analysis of experiments, 3rd Ed. John Wiley & Sons, Inc., New York, (1991).

Plackett, R.L. and. Burman, J.P. The design of optimum multifactorial experiments. *Biometrika* **33**, 305–325 (1946).

Taguchi, G. Robust technology development: bringing quality engineering upstream. ASME Press (ISBN 978-0791800287) (1992).

Ordering information

For more information, please contact your local sales representative or visit

www.gelifesciences.com/chromatographymedia
www.gelifesciences.com/predictor
www.gelifesciences.com/aktaavant
www.gelifesciences.com/aktapure
www.gelifesciences.com/unicorn

GE Healthcare UK Limited
Amersham Place
Little Chalfont
Buckinghamshire, HP7 9NA
UK

GE Healthcare Europe
GmbH, Munzinger Strasse 5
D-79111 Freiburg
Germany

GE Healthcare Bio-Sciences Corp.
800 Centennial Avenue, P.O. Box 1327
Piscataway, NJ 08855-1327
USA

GE Healthcare Japan Corporation
Sanken Bldg., 3-25-1, Hyakunincho
Shinjuku-ku, Tokyo 169-0073
Japan

For local office contact information,
please visit www.gelifesciences.com/contact

www.gelifesciences.com/bioprocess

GE Healthcare Bio-Sciences AB
Björkgatan 30
751 84 Uppsala
Sweden

imagination at work

29-1038-50 AA 05/2014