Computer-aided Assay Development for Charge Heterogeneity Analysis by iCE

Introduction

Unlike chemically synthesized drugs, protein therapeutics are a dynamic heterogeneous mix of active compounds[1]. Due to their complexity, analytical techniques like isoelectric focusing have become indispensable tools in evaluating biologic preparations. The resulting surge in charge isoform analysis has led to major advances in instrumentation, such as Imaged Capillary Electrophoresis (iCE™)[2]. However, to obtain the full benefit from improved instrumentation requires the coinciding development of robust assays.

Initially implemented in biopharmaceutical manufacturing, the holistic process characterization philosophy known as Quality by Design (QbD) has the potential to transform assay development[3, 4, 5]. Proper adaptation of these techniques will provide a tremendous benefit to the robustness and predictability of assay performance. Key to QbD is comprehensively gauging the effects of process inputs on critical to quality (CTQ) attributes of the output[3]. To this end, the Design of Experiments (DOE) methodology has proven itself to be a highly efficient tool in modeling the relationship between input and output. Though statistical analysis packages such as SAS JMP and Minitab® have lowered the computational barriers to executing DOE, generating meaningful results still requires a working knowledge of the model building process.

The goal of this note is to promote the successful application of DOE tools in the assay development process by offering a stepwise example. The road map contained in the following pages has purposely captured enough technical detail to provide a comprehensive reference guide for both the statistician and analytical biochemist. The subjects that will be covered include initial factor screening, construction of a central composite DOE, response surface modeling, assay optimization, model validation and assay performance.

Materials

Capillaries and Reagents

- IgG 1 Kappa analysis runs were performed on FC (fluorocarbon-coated) cIEF Cartridges (P/N 101701, ProteinSimple).
- Pharmalyte 3 to 10 Ampholytes (P/N 17-0456-01, GE Life Sciences) and Pharmalyte 8 to 10.5 Narrow Range Ampholytes (P/N 17-0455-01, GE Life Sciences) were provided as 25 mL volumes.
- pI 8.18 marker (P/N 102408, ProteinSimple) and pI 9.46 marker (P/N 102059, ProteinSimple) were provided as 200 µL volumes at a 300X concentration.
- 100 mL volumes of Anolyte solutions containing 80 mM phosphoric acid in 0.1% methyl cellulose and Catholyte solutions containing 100 mM sodium hydroxide in 0.1% methyl cellulose were provided together as an electrolyte kit (P/N 102056, ProteinSimple).

- 10 M urea solutions were made fresh by dissolving 6 g of urea (P/N U6504, Sigma Aldrich) in 10 mL of distilled deionized (DDI) water and vortexing until all solids solubilized.
- The IgG 1 Kappa buffer replacement solution which contained 20 mM Tris-HCl (pH 7.0) was made by diluting 0.8 mL of 1 M Tris-HCl buffer at pH 7.0 (P/N T2413, Sigma Aldrich) with 39.2 mL of DDI water.
- 100 mL bottles of 0.5% Methyl Cellulose (P/N 102505, ProteinSimple) and 1% Methyl Cellulose (P/N 101876, ProteinSimple,) were provided as ready to use reagents.
- **cIEF master mix:** A 220 µL volume per sample of cIEF master mix containing 0.44% v/v methyl cellulose, 5 M urea, 2.5% v/v Pharmalyte 3 to 10, 1.25x% v/v Pharmalyte 8 to 10.5 Narrow Range, and 0.0031X pI 9.46 and pI 8.18 markers was prepared fresh and vortex mixed. The amount of x% Pharmalyte 8 to 10.5 ranged between 0 and 4%.
- **Buffer exchanged IgG 1 Kappa solution:** A 500 µL volume of a Human IgG 1 Kappa at 1 mg/mL (P/N 523, Protos Immunoresearch) was loaded into an Amicon Ultracel 50K Membrane Centrifugal Filter (P/N 4311, Millipore). Following centrifugation for 5 minutes at 12K rcf, the filtered volume was replaced with 20 mM Tris buffer (pH 7.0). Two additional cycles of centrifugation and buffer replacement were performed and the buffer exchanged sample was stored at -20 °C or below.
- **IgG 1 Kappa working stock solution:** A 50 µL volume per sample of 1.25 mg/mL Human IgG 1 Kappa was prepared by diluting buffer exchanged IgG 1 Kappa stocks with DDI water.
- **iCE separation sample:** A 200 µL volume of cIEF master mix was added to each 50 µL of IgG 1 Kappa working stock solution. The resulting 250 µL volume was vortex mixed and then centrifuged at 13K rcf for 3 minutes. The top 200 µL was then transferred into a plastic sample vial and placed into a 2 mL glass vial, which was then capped and loaded into a PrinCE Next MicroInjector or Alcott 720 NV Autosampler.

Methods

iCE3 Separations

Separations were performed on an iCE3 system equipped with either a PrinCE Next MicroInjector or an Alcott 720 NV Autosampler. Samples were injected into the FC cIEF cartridge for a duration determined by a transfer time measurement step performed during the cartridge installation procedure. After pressure in the sample loading path stabilized, a pre-focusing period of 1 minute was performed at 1500 V followed by a focusing period of y minutes at 3000 V. The focusing time (y) ranged between 9 and 11 minutes.

Circumscribed Central Composite Design

A two-factor, circumscribed central composite design was generated in Minitab 16 using the DOE tools under the Stat Menu. The resulting design was balanced but not symmetrical as it was slightly modified to allow for a smaller span of % Narrow Range to be tested than that for focusing time. A basic schematic is shown in **Figure 3**. The first factor, % Narrow Range v/v, had axial points at 0 and 4%, cube points at 1 and 3%, and a center point at 2%. The second factor, focusing time min, had axial points at 9 and 11 minutes, cube points at 9.5 and 10.5 minutes and a center point at 10 minutes. The center point setting at 2% Narrow Range and 10 minutes of focusing time was run in triplicate to estimate the pure error of the model. Samples for each point in the central composite design were made independently and run in random order assigned by Minitab. (See *Construction of a Central Composite Design* on page 4 for more information).

Model Generation, Optimization, and Validation

The charge heterogeneity profiles were calibrated, confirmed, and converted into a standard ANDI file format using iCE CFR Software Version 3.0. The standard ANDI files were then imported into Chrom Perfect, where the charge heterogeneity profile peaks were integrated to assess pI, percent composition, and peak-specific resolution. The peak-specific resolution factors for peaks 3b, 4b, and 4c (see **Figure 1** for peak profile)

were then input into the central composite worksheet in Minitab and analyzed to generate response surfaces and contour plots. A contour plot overlay was also generated to select for a robust setting. The model was validated by comparing the predicted resolution values at this setting with actual values generated by running three independently prepared IgG 1 Kappa samples at this setting.

Results

Factor Screening

It is always important to remember that though the computational tools are powerful, they are not intelligent, and they will need your knowledge to perform at their best. Simplicity is key to an efficient experimental design. As the numbers of factors increases, so do the logistics and complexity of the DOE. Utilizing prior experience is a great way to isolate for important factors. However, if the assay process is rather unknown, an initial low-resolution DOE can be performed as a selection process on a wider array of factors.

To better understand the charge heterogeneity profile, Human IgG 1 Kappa was initially screened using a typical 4% v/v Pharmalyte 3 to 10 gradient with 4 M urea added to avoid aggregation and/or precipitation. The results from this isoelectric focusing run indicated that convoluted microheterogeneity was present in the charge profile as tightly packed peak clusters. Accelerating the screening process, settings from a prior assay optimization project were implemented in a second run to investigate the profile further (**Figure 2**).

The results from these two screening runs were sufficient to identify the percentage of Pharmalyte 8 to 10.5 Narrow Range in the sample solution as a critical factor for resolving these tightly spaced peaks. Since this cut of ampholyte has been extensively characterized and found to be primarily comprised of slowly focusing ampholyte species, focusing time also needed to be included as the other factor for the central composite design[6].

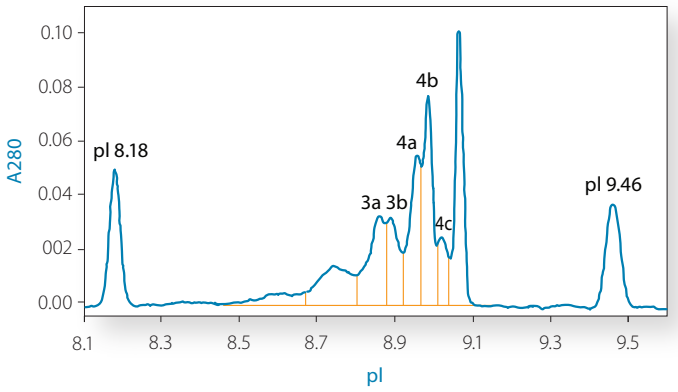


FIGURE 1. Human IgG 1 Kappa profile generated using 2% Pharmalyte 3 to 10, 2% Pharmalyte 8 to 10.5 gradient with 4 M urea, focused for 1 min at 1500 V and 10 min at 3000 V. Peaks 3a through 4c were identified as resolution indicating and assigned as critical to quality attributes.

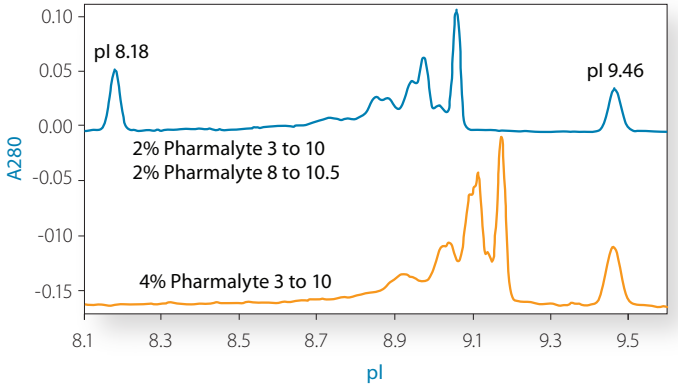


FIGURE 2. Two iCE profiles of Human IgG 1 Kappa. The top trace was generated using 2% Pharmalyte 3 to 10, 2% Pharmalyte 8 to 10.5 Narrow Range pH gradient with 4 M urea, focused for 1 min at 1500 V and 10 min at 3000 V. The bottom trace was generated using a 4% Pharmalyte 3 to 10 pH gradient with 4 M urea focused for 1 min at 1500 V and 7 min at 3000 V.

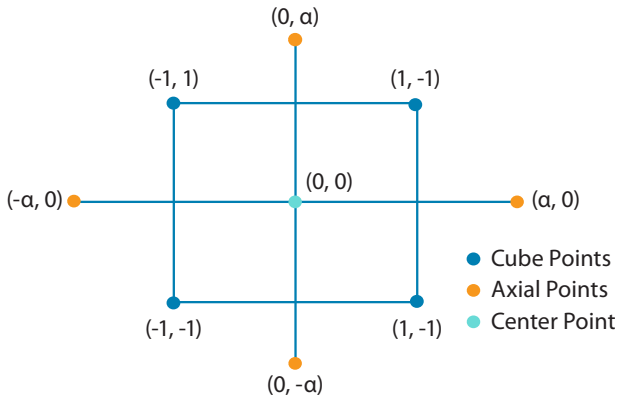


FIGURE 3. Plot of a two-factor circumscribed central composite design.

The high resolution charge profile obtained in the 2% v/v Pharmalyte 3 to 10, and 2% v/v Pharmalyte 8 to 10.5 Narrow Range gradient showed the presence of five separate peaks with maximums spaced by 0.03 to 0.05 pH units apart in isoelectric point (pI). Consistent resolution of these peaks will have a great effect on downstream peak integration and quantitation of the charge profile's Peak Cluster Percent Composition (PCPC), so their specific resolution was selected as the CTQs to evaluate assay performance.

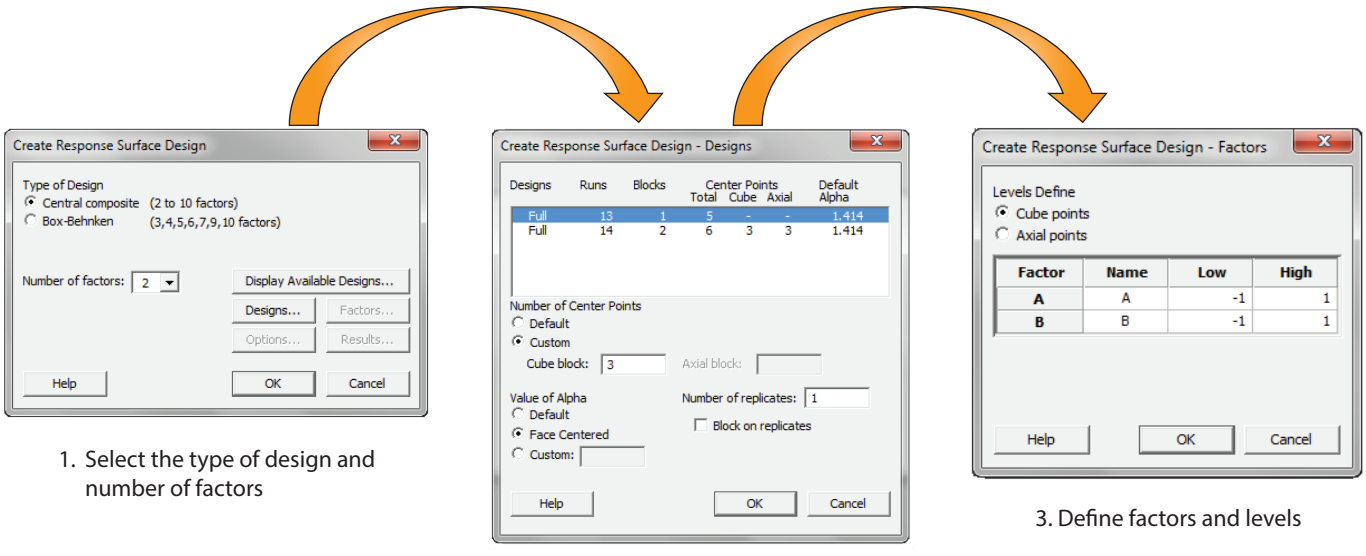
Construction of a Central Composite Design

Unlike traditional experimental design approaches that test the effects of a single factor at a time, DOE applies methodology that simultaneously tests all factors. In DOE, combinations of multiple factor settings are contained in the experimental design space. Comprised of a balanced distribution of experimental points surrounding a center point, regions of the design space are defined by discrete values called levels (see **Figure 3** for a basic schematic). In factorial designs these outer points, called cube points, test conditions above (High Level: 1) and below (Low Level: -1) the center point (0). The three levels in this design space are sufficient to model a factor's first order (Linear) effects, combined effects with other factors (Interactions), and test for the presence of second order (Curved) effects on output.

Central Composite Designs (CCD), through the utilization of a second set of experimental levels is capable of modeling first order, second order (Quadratic), and combined effects. In addition to the cube points and center points from fraction factorial designs, most CCD types incorporate a second set of experimental levels in the form of axial points. Axial points, as their name alludes to, lie on the axis of the design space placed a distance from the center point which is determined by the value alpha (α).

The value of α also gives rise to the three types of CCD. Circumscribed design axial points are outside the design space with $-1 < \alpha > 1$. Face centered designs have axial points on the surface of the design space with $\alpha = \pm 1$. Inscribed designs occur when $\alpha = \pm 1$ and the factorial levels of the cube points are set > -1 and < 1.

The face centered and inscribed central composite designs are usually implemented to address boundary conditions with experimental settings. Of these two



1. Select the type of design and number of factors

2. Select design specifics

3. Define factors and levels

FIGURE 4. The highly streamlined experimental design workflow and easy to use user interface typical of advanced statistical analysis packages like Minitab 16.

design types, inscribed central composite designs are more preferable to apply because they still contain five separate experimental levels to better model curvature. Circumscribed central composite designs should be implemented whenever possible as they tend to produce the most accurate models.

The CCD utilized in the optimization of the Human IgG 1 Kappa iCE assay was automatically generated in Minitab 16 using the DOE tool under the Stat Menu. The experiment design process in **Figure 4** generates a rotatable CCD with a single default alpha of 1.414. This setting of alpha places all points in the design space equidistant from the center point allowing for better variance modeling. Modifications needed to be made to the design after construction in order to incorporate two alphas. This was done to avoid failure modes as the sensitivity of the assay to the factor of Narrow Range % was much less than that of focusing time. Though the design still remained balanced (*e.g.*, the center point of each factor is equidistant from the axial points), the asymmetry in the design does pose the risk of errors in the variance estimation of model. This risk was deemed acceptable in order to execute the more powerful circumscribed design while remaining accommodating for potential boundary conditions.

Response Surface Modeling

The various points of the CCD were run on the iCE3 system in a random order determined by Minitab 16 during construction. These runs were calibrated, converted and imported into Chrom Perfect to be integrated. Specific peak resolutions were calculated for peaks 3a through 4c, but only the values for 3b, 4b, and 4c were inputted into the CCD worksheet for analysis as they were much lower than 3a and 4a.

Once populated with the specific resolution values, the CCD worksheet was analyzed by Minitab 16 using the workflow shown in **Figure 5** and response surface models built. The coefficients in the resulting regression model output were used to determine the magnitude of the effect on specific resolution for each factor (X), its square (X^2), and interaction (XY) with other factors.

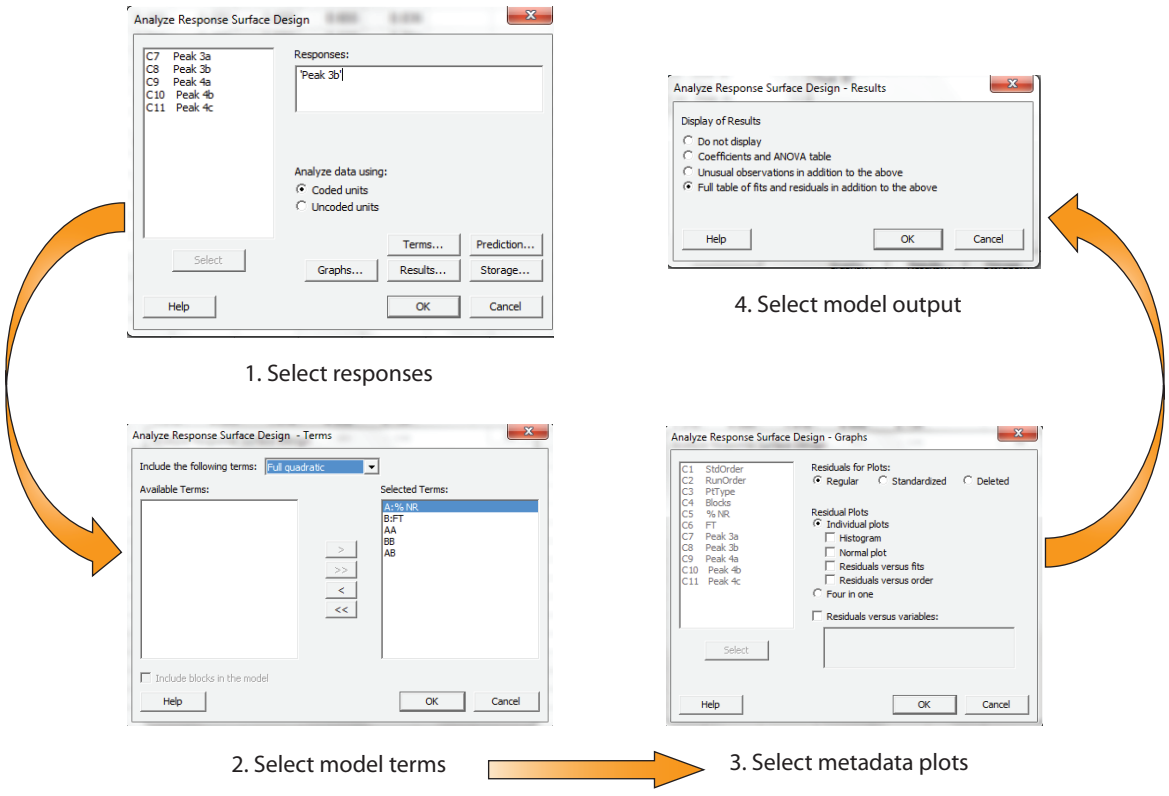


FIGURE 5. The highly streamlined workflow and easy to use user interface for response surface analysis in Minitab 16.

Significance of these effects was determined to a 0.95 confidence interval (C.I.) with corresponding p values. After evaluating the models, a second round of analysis was performed to build a simplified model by eliminating non-significant terms. These reduced models were then used to generate surface plots to better visualize the effects of the factors.

For all models, four in one residual plots were also generated by Minitab. These plots were useful in examining the distribution of error throughout the model. Deviations from normal distributions or trends in the residual error indicate potential problems with the model's predictive behavior.

The regression model in **Table 1** indicates that between 79 to 83% of the observed variation in peak 3b resolution can be explained by linear effects of % Narrow Range (NR) and focusing time (FT). Residual plots of the regression model in **Figure 6** show a normal distribution of error through the model, and there appears to be no bias of error throughout the

TERM	COEF	SE COEF	T	P
Constant	0.5132	0.0115	44.657	0.000
% NR	0.0773	0.0147	5.272	0.001
FT	-0.0491	0.0147	-3.348	0.010
S = 0.0381 PRESS = 0.0242 R-Sq = 82.98% R-Sq(adj) = 78.72%				

Table 1. Reduced regression model for peak 3b in the Human IgG 1 Kappa charge profile. The R-Sq (adj), which takes into consideration model complexity, predicts that approximately 79% of the variance in peak 3b resolution can be explained by the linear effects of % Narrow Range (NR) and focusing time (FT).

TERM	COEF	SE COEF	T	P
Constant	0.6908	0.0249	27.697	0.000
% NR	0.1299	0.0254	5.114	0.001
% NR*% NR -0.0941 0.0245 -3.838 0.005 S = 0.0660 PRESS = 0.2625 R-Sq = 83.63% R-Sq(adj) = 79.54%				

Table 2. Reduced regression model for peak 4b in the Human IgG 1 Kappa charge profile. The R-Sq (adj) predicts that approximately 80% of the variance in peak 4b resolution can be explained by the linear effects and the square effects of % Narrow Range (NR).

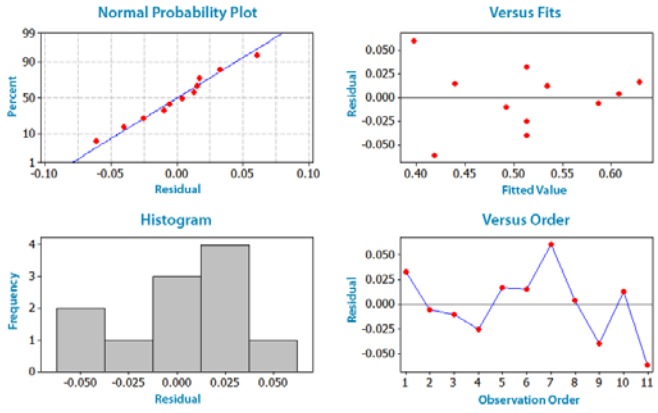


FIGURE 6. Four in one residual plot for the regression model of peak 3b specific resolution. The residual plots indicate that the error of the regression model has a normal distribution (Normal Probability plot), is balanced across the predictive range of the model (0.400 to 0.650, Versus Fits plot) and is not effected by run order (1 to 11, Versus Order plot).

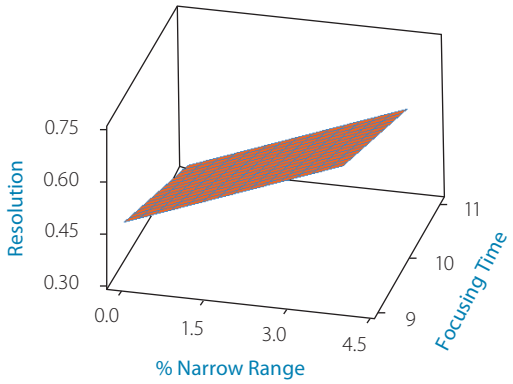


FIGURE 7. Surface plot for the regression model of peak 3b specific resolution. The plot shows a linear improvement in peak 3b resolution as % Narrow Range is increased and a linear decline in resolution as focusing time is increased. The rate of resolution improvement due to increasing % Narrow Range is much greater than the rate of resolution loss due to increasing focusing time.

predictive range (Versus Fits) of the model, and there are no apparent time course effects through the data collection (Versus Order).

Both the regression model and surface plot (**Figure 7**) for peak 3b resolution highlight a positive correlation to % Narrow Range and a negative correlation with focusing time. Since the response in peak 3b resolution is linear for both factors, there is no optimum setting for peak 3b.

According to **Table 2**, between 80 to 84% of the observed variation in peak 4b resolution can be explained by linear and square effects of % Narrow Range. There appear to be no effects on resolution explained by focusing time. Like peak 3b, the residual plots for peak 4b in **Figure 8** indicate no issues with the predictive capability of the model. Peak 4b's resolution has a positive correlation to the linear effects of % Narrow Range and a negative correlation to the squared effects of % Narrow Range.

The surface plot in **Figure 9** shows that there are quadratic effects to peak 4b resolution in response to changes in % Narrow Range. The curvature is a downward shaped parabola with a maximum just before 3.0% Narrow Range. Peak 4b's resolution is flat across the range of focusing times.

TERM	COEF	SE COEF	T	P
Constant	0.8387	0.0190	44.081	0.000
% NR	0.0299	0.0194	1.543	0.161
% NR*% NR	-0.0721	0.0187	-3.854	0.005
S = 0.0503 PRESS = 0.0914 R-Sq = 68.29% R-Sq(adj) = 60.37%				

Table 3. Reduced regression model for peak 4c in the Human IgG 1 Kappa charge profile. The R-Sq (adj) predicts that approximately 60% of the variance in peak 4c resolution can be explained by the linear and the square effects of % Narrow Range (NR).

The regression model in **Table 3** estimates that between 60 to 68% of the observed variation in peak 4c resolution can be explained by the linear and square effects in % Narrow Range. Though the linear effects of % Narrow Range are not significant with a p value > 0.05, limitations in Minitab require linear effects remain when modeling their corresponding square effects. There is no evidence that focusing time effects peak 4c resolution.

Like the two prior models generated before it, the residual plots in **Figure 10** indicate that the model is clear of predictive defects. Peak resolution has a

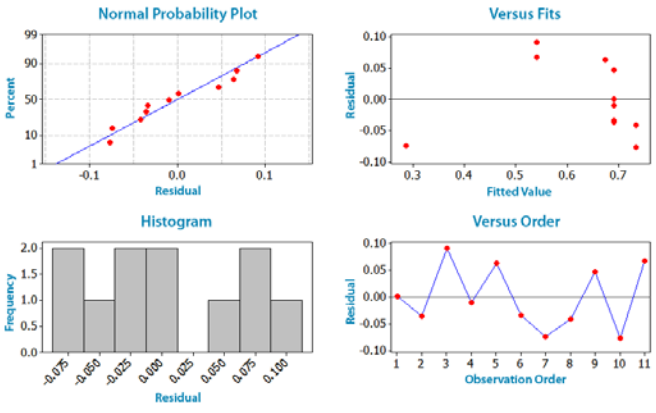


FIGURE 8. Four in one residual plot for the regression model of peak 4b specific resolution. The residual plots indicate that the error of the regression model has a normal distribution (Normal Probability plot), is balanced across the predictive range of the model (0.300 to 0.750, Versus Fits plot) and is not effected by run order (1 to 11, Versus Order plot).

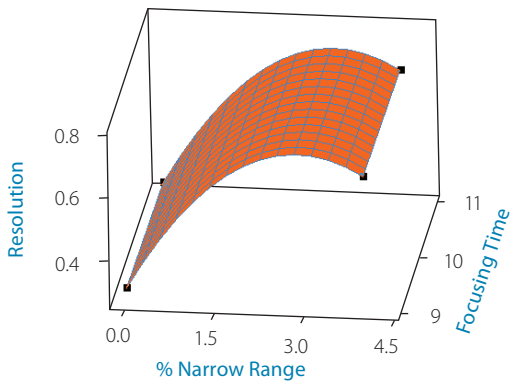


FIGURE 9. Surface plot for the regression model of peak 4b specific resolution. The plot shows quadratic effects in peak 4b resolution in response to changes in % Narrow Range. There appears to be a maximum setting for resolution near 3.0% Narrow Range.

positive correlation to % Narrow Range and a negative correlation to the squared effects of % Narrow Range (**Figure 11**). There is a resolution optimum between 1.5 to 3.0% Narrow Range for peak 4c.

The three models for peak 3b, 4b and 4c show individually how each peak responds to changes in % Narrow Range and focusing time. Understanding these responses will not only allow for optimization of the method, but will also provide indicators that can be later applied to assay training and troubleshooting.

Assay Optimization and Model Validation

Finding the robust region of factor settings requires combining these individual responses into a single graphical representation in order to investigate overlapping conditions for system performance. The best way to observe this is through an overlay contour plot, which is comprised of a composite overhead view of the response surfaces.

Creating an overlay contour plot in Minitab only requires selecting the responses that you wish to combine and then inputting the upper and lower limits of the range for each of those responses. The contour provides a comprehensive view of the performance landscape and highlights the robust region for system operation.

For defining the robust region and optimizing the Human IgG 1 Kappa iCE assay, a contour overlay plot with the following specific resolution targets was set for the final assay. Peak 3b was set from 0.5 to 0.6, peak 4b from 0.7 to 0.8 and peak 4c from 0.8 to 0.9. The resulting contour plot in **Figure 12** contained a large robust region of operation (shown in the white area) between the desired resolution ranges. In Minitab, the contour plot is an active display allowing the user to set down flags in the design space and generate predicted values for each response. Placing the flag away from the failure boundaries at factor values at 2.5% Narrow Range and 10 minutes of focusing time, the contour model predicted the resolution performance for peak 3b to be 0.527, peak 4b at 0.744 and peak 4c at 0.823.

To assess the real-world accuracy of the combined response surface model, separate validation runs were performed and the actual values from the integrated data were compared to the predicted values. Three samples containing 2% Pharmalyte 3 to 10, 2.5% Pharmalyte 8 to 10.5 Narrow Range, 4 M urea and 250 mg/mL Human IgG 1 Kappa were separately

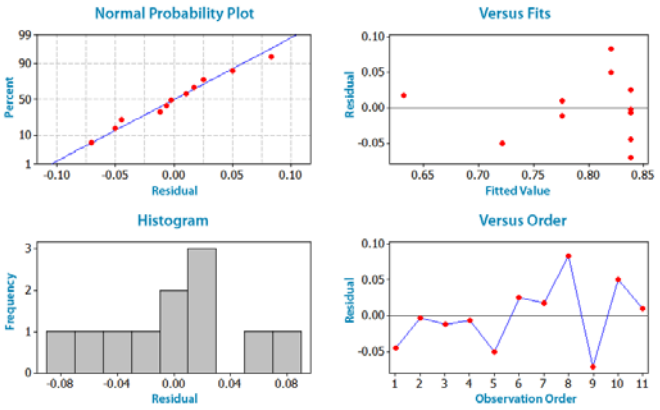


FIGURE 10. Four in one residual plot for the regression model of peak 4c specific resolution. The residual plots indicate that the error of the regression model has a normal distribution (Normal Probability plot), is balanced across the predictive range of the model (0.600 to 0.850, Versus Fits plot) and is not effected by run order (1 to 11, Versus Order plot).

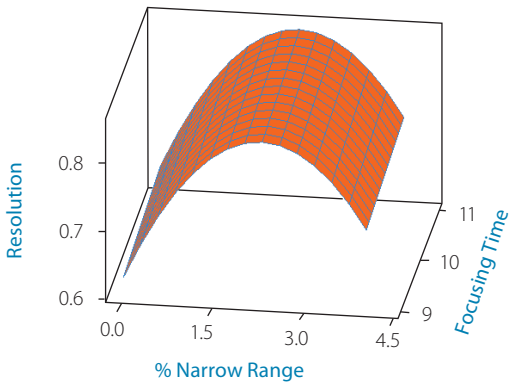


FIGURE 11. Surface plot for the regression model of peak 4b specific resolution. The plot shows quadratic effects in peak 4b resolution in response to changes in % Narrow Range. There appears to be a maximum setting for resolution near 2.5% Narrow Range.

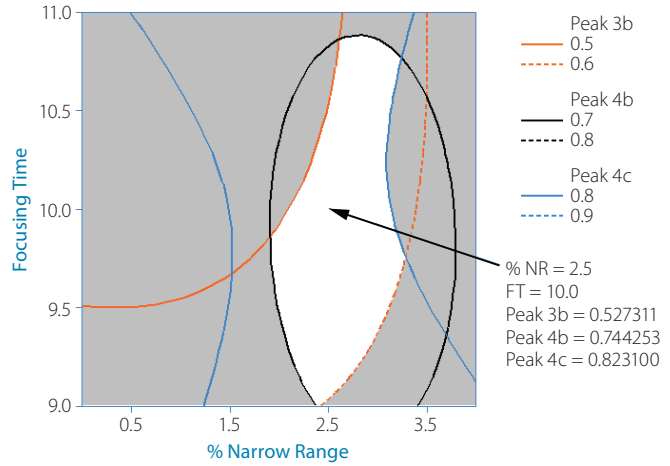


FIGURE 12. Overlay contour plots for peaks 3b, 4b, and 4c.

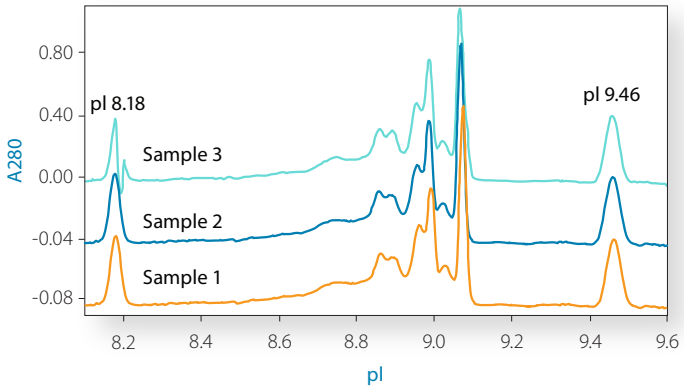


FIGURE 13. Three iCE profiles of Human IgG 1 Kappa model validation runs. Each sample was individually prepared containing 2% Pharmalyte 3 to 10, 2.5% Pharmalyte 8 to 10.5 Narrow Range, and 4 M urea. The samples were pre-focused for 1 min at 1500 V followed by a focusing step for 10 min at 3000 V.

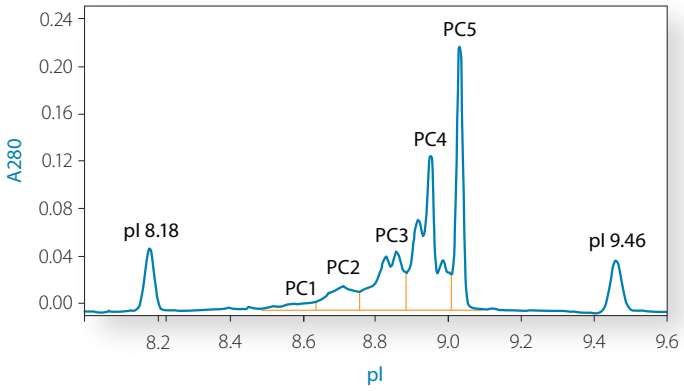


FIGURE 14. The five peak clusters (PC) used in the percent composition analysis of Human IgG 1 Kappa.

prepared and separated by isoelectric focusing with the iCE3 system using a pre-focusing period of 1 minute at 1500 V followed by a focusing period of 10 minutes at 3000 V. The generated charge profiles were then integrated and resolution values for peaks 3b, 4b, and 4c were extracted (**Figure 13**).

Comparison between the predicted and experimentally obtained resolution values highlights how even complex separations like isoelectric focusing by iCE can be modeled successfully by CCD. Peak 3b's three experimental results were all higher than predicted with percent deviation for the experimental values ranging between 6.06 and 11.73%. Separations of peak 4b generated a percent deviation from the predicted values between -2.76 and 7.58%. Peak 4c had the closest experimental values with a percent deviation between -3.13 and 1.20%. The three higher experimental results generated by peak 3b does point to a potential predictive bias in the model, however it is deemed acceptable since it is in favor of the desired performance. The more balanced distribution of error in peak 4b's and 4c's three experimental results are ideal, indicating no predictive bias.

Assay Performance

Using CCD for setting controlled assay conditions far away from failure mode boundaries allows the system to ingest more variation from less controlled factors without effecting results. To demonstrate this, the CCD-optimized Human IgG 1 Kappa iCE assay was performed on two instruments over 11 days using 12 FC cIEF cartridges from four separate build lots.

The results from routine analysis of the complex Human IgG 1 Kappa charge profile clearly illustrate the benefit achieved through CCD conditions settings (**Figure 14**). Though average peak cluster percent compositions were rather low at values between 2.93 to 39.02%, their corresponding coefficients of variance remained respectable ranging between 11.54 and 2.05%.

Conclusion

Computer-aided method development techniques like DOE clearly reduce the cost of method development, especially when leveraged with prior knowledge and/or experimental results. These factorial approaches provided for simultaneous multidimensional characterization of the assay performance envelope. This holistic understanding of the relationship between system inputs and output provides for more predictable assays and easier avoidance of non-robust condition settings.

The results from a properly executed central composite experimental design generated data that allowed for accurate modeling of both linear and nonlinear effects of system inputs on peak resolution. Using these experimentally validated models, robust conditions were identified for the Human IgG 1 Kappa iCE assay. An intermediate precision study of these robust assay settings highlighted that the resulting iCE method was capable of ingesting uncontrolled variation while maintaining stable output.

References

1. (Re)defining biopharmaceutical, R Rader, *Nat. Biotechnol.*, 2008; 26:743-751.
2. Protein analysis using imaged capillary isoelectric focusing, J Wu, A Watson and A Torres, *American Biotechnology Laboratory,* 1999.
3. FDA (2006) Guidance for industry. Q8 pharmaceutical development.
4. Characterization of a monoclonal antibody cell culture production process using a quality by design approach, B Horvath, M Mun and M W Laird, *Mol. Biotechnol.*, 2010; 45(3):203-6.
5. An industry perspective on quality by design, R Johnston, J Lambert and E Stump, *BioProcess International*, 2012; 10(3):26-35.
6. Carrier ampholytes for IEF, on their fortieth anniversary (1967-2007), brought to trial in court: the verdict, P G Righetti, C Simo, R Sebastiano and C Attilio, *Electrophoresis*, 2007; 28:3799-3810.

proteinsimple

Toll-free: (888) 607-9692
Tel: (408) 510-5500
Fax: (408) 510-5599
info@proteinsimple.com
proteinsimple.com

98-4001-00 / Rev A.