

RESEARCH

Open Access



Perspectives in modeling and model validation during analytical quality by design chromatographic method evaluation: a case study

Yongzhi Dong^{*} , Zhimin Liu, Charles Li, Emily Pinter, Alan Potts, Tanya Tadey and William Weiser

Abstract

Design of experiments (DOE)-based analytical quality by design (AQbD) method evaluation, development, and validation is gaining momentum and has the potential to create robust chromatographic methods through deeper understanding and control of variability. In this paper, a case study is used to explore the pros, cons, and pitfalls of using various chromatographic responses as modeling targets during a DOE-based AQbD approach. The case study involves evaluation of a reverse phase gradient HPLC method by a modified circumscribed central composite (CCC) response surface DOE.

Solid models were produced for most responses and their validation was assessed with graphical and numeric statistics as well as chromatographic mechanistic understanding. The five most relevant responses with valid models were selected for multiple responses method optimization and the final optimized method was chosen based on the Method Operable Design Region (MODR). The final method has a much larger MODR than the original method and is thus more robust.

This study showcases how to use AQbD to gain deep method understanding and make informed decisions on method suitability. Discoveries and discussions in this case study may contribute to continuous improvement of AQbD chromatography practices in the pharmaceutical industry.

Keywords: AQbD, DOE, MODR, Statistical model validation, Scientific model validation, Multiple responses optimization

Introduction

Drug development using a quality by design (QbD) approach is an essential part of the Pharmaceutical cGMP Initiative for the twenty-first century (FDA Pharmaceutical cGMPs For The 21st Century — A Risk-Based Approach, 2004) established by the FDA. This initiative seeks to address unmet patient needs, unsustainable rise of healthcare costs, and the reluctance to adopt new technology in pharmaceutical development and manufacturing. These issues were the result of old regulations that are very rigid and made continuous improvement of

previously approved drugs both challenging and costly. The International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) embraced this initiative and began issuing QbD relevant quality guidelines in 2005. The final versions of ICH Q8–Q12 [(ICH Q8 (R2) 2009) (ICH Q9 2005) (ICH Q10 2008) (ICH Q11 2012) (ICH Q12 2019)] have been adopted by all ICH members. The in-progress version of ICH Q14 (ICH Q14 2018) will offer AQbD guidelines for analytical procedures and promote the use of QbD principles to achieve a greater understanding and control of testing methods and reduction of result variability.

* Correspondence: yongzhi.dong@thermofisher.com
Thermo Fisher Scientific Inc., Durham, NC, USA

Product development using a QbD approach emphasizes understanding of product and process variability, as well as control of process variability. It relies on analytical methods to measure, understand, and control the critical quality attributes (CQA) of raw materials and intermediates to optimize critical process parameters and realize the Quality Target Product Profile (ICH Q8 (R2) 2009). Nevertheless, part of the variability reported by an analytical test can originate from the variability of the analytical measurement itself. This can be seen from Eq. 1.

$$\sigma^2_{\text{reported}} = \sigma^2_{\text{product}} + \sigma^2_{\text{measurement}} \quad (1)$$

The reported variability is the sum of intrinsic product variability and extrinsic analytical measurement variability (NIST/SEMATECH e-Handbook of statistical methods 2012a, 2012b, 2012c, 2012d). The measurement variability can be minimized by applying QbD principles, concepts, and tools during method development to assure the quality and reliability of the analytical method can meet the target measurement uncertainty (TMU) (EURACHEM/CITAC 2015). High-quality analytical data truthfully reveal product CQAs and thus enables robust, informed decisions regarding drug development, manufacturing, and quality control.

ICH Q14 introduces the AQbD concepts, using a rational, systematic, and holistic approach to build quality into analytical methods. The Method Operable Design Region (MODR) (Borman et al. 2007) is a multidimensional space based on the critical method parameters and settings that provide suitable method performance. This approach begins with a predefined analytical target profile (ATP) (Schweitzer et al. 2010), which defines the method's intended purpose and commands analytical technique selection and all other method development activities. This involves understanding of the method and control of the method variability based on sound science and risk management. It is generally agreed upon that systematic AQbD method development should include the following six consecutive steps (Tang 2011):

1. ATP determination
2. Analytical technique selection
3. Method risk assessment
4. MODR establishment
5. Method control strategy
6. Continuous method improvement through a life cycle approach

A multivariate MODR allows freedom to make method changes and maintain the method validation (Chatterjee S 2012). Changing method conditions within

an approved MODR does not impact the results and offers an advantage for continuous improvement without submission of supplemental regulatory documentation. Establishment of the MODR is facilitated by multivariate design of experiments (DOE) (Volta e Sousa et al. 2021). Typically, three types of DOE may be involved in AQbD. The screening DOE further consolidates the potential critical method parameters determined from the risk assessment. The optimization DOE builds mathematical models and selects the appropriate critical method parameter settings to reach to the target mean responses. Finally, the robustness DOE further narrows down the critical method parameter settings to establish the MODR, within which the target mean responses are consistently realized. Based on this AQbD framework, it is very clear DOE models are essential to understanding and controlling method variability to build robustness into analytical methods. Although there have been extensive case studies published regarding AQbD (Grangeia et al. 2020), systematic and in-depth discussion of the fundamental AQbD modeling is still largely unexplored. Methodical evaluation of the pros, cons, and pitfalls of using various chromatographic responses as modeling targets is even more rare (Debrus et al. 2013) (Orlandini et al. 2013) (Bezerra et al. 2019). The purpose of this case study is to investigate relevant topics such as data analysis and modeling principles, statistical and scientific validation of DOE models, method robustness evaluation and optimization by Monte Carlo simulation (Chatterjee S 2012), multiple responses method optimization (Leardi 2009), and MODR establishment. Discoveries and discussions in this case study may contribute to continuous improvement of chromatographic AQbD practices in the pharmaceutical industry.

Methods/experimental

Materials and methods

C111229929-C, a third-generation novel synthetic tetracycline-class antibiotic currently under phase 1 clinical trial was provided by KBP Biosciences. A reverse phase HPLC purity and impurities method was also provided for evaluation and optimization using AQbD. The method was developed using a one factor at a time (OFAT) approach and used a Waters XBridge C18 column (4.6 × 150 mm, 3.5 μm) and a UV detector. Mobile phase A was composed of ammonium acetate/ethylenediaminetetraacetic acid (EDTA) buffer at pH 8.8 and mobile phase B was composed of 70:30 (v/v) acetonitrile/EDTA buffer at pH 8.5. Existing data from forced degradation and 24-month stability studies demonstrated that the method was capable of separating all six specified impurities/degradants with ≥ 1.5 resolution.

A 1.0 mg/mL C111229929-C solution was prepared by dissolving the aged C111229929-C stability sample into

10 mM HCl in methanol and used as the method evaluation sample. An agilent 1290 UPLC equipped with a DAD detector was used. In-house 18.2 M Ω Milli-Q Water was used for solution preparations. All other reagents were of ACS equivalent or higher grade. Waters Empower[®] 3 was used as the Chromatographic Data System. Fusion QbD v 9.9.0 software was used for DOE design, data analysis, modeling, Monte Carlo simulation, multiple responses mean, and robustness optimization. Empower[®] 3 and Fusion QbD were fully integrated and validated.

A method risk assessment was performed through review of the literature and existing validation and stability data to establish priorities for method inputs and responses. Based on the risk assessment, four method parameters with the highest risk priority numbers were selected as critical method parameters. Method evaluation and optimization was performed by a modified circumscribed central composite (CCC) response surface DOE design with five levels per parameter, for a total of 30 runs. The modifications were the extra duplicate replications at three factorial points. In addition to triplicate replications at the center point, the modified design had a total of nine replicates. See Table 1 for the detailed design matrix. A full quadratic model for the four-factor five-level CCC design has a total of fourteen potential terms. They include four main linear terms (A, B, C, D), four quadratic terms (A², B², C², D²), and six two-way interaction terms (A*B, A*C, A*D, B*C, B*D, and C*D).

Pre-runs executed at selected star (extreme) points verified that all expected analytes eluted within the 35 min run time. This mitigated the risk of any non-eluting peaks during the full DOE study, as a single unusable run may raise questions regarding the validity of the entire study. Based on the pre-runs, the concentration of the stock EDTA solution was decreased four-fold to mitigate inaccurate in-line mixing of mobile phase B caused by low volumes of a high concentration stock. The final ranges and levels for each of the four selected method parameters are also listed in Table 1.

Each unique DOE run in Table 1 is a different method. As there were 25 unique running conditions, there were 25 different methods in this DOE study. The G-Efficiency and the average predicted variance (NIST/SEMATECH e-Handbook of statistical methods 2012a, 2012b, 2012c, 2012d) (Myers and Montgomery 1995) of the design were 86.8% and 10.6%, respectively, meeting their respective design goals of $\geq 50\%$ and $\leq 25\%$. Some of the major advantages of this modified CCC design include the following:

- Established quadratic effects
- Robust models that minimize effects of potential missing data

- Good coverage of the design space by including the interior design points
- Low predictive variances of the design points
- Low model term coefficient estimation errors

The design also allows for implementation of a sequential approach, where trials from previously conducted factorial experiments can be augmented to form the CCC design. When there is little understanding about the method and critical method parameters, such as when developing a new method from scratch, direct application of an optimizing CCC design is generally not recommended. However, there was sufficient previous knowledge regarding this specific method, justifying the direct approach.

DOE data analysis and modeling principles

DOE software is one of the most important tools to facilitate efficient and effective AQbD chromatographic method development, validation, and transfer. Fusion QbD software was employed for DOE design and data analysis. Mathematical modeling of the physicochemical chromatographic separation process is essential for DOE to develop robust chromatographic methods through three phases: chemistry screening, mean optimization, and robustness optimization. The primary method parameters affecting separation (e.g., column packing, mobile phase pH, mobile phase organic modifier) are statistically determined with models during chemistry screening. The secondary method parameters affecting separation (e.g., column temperature, flow rate, gradient slope settings) are optimized during mean optimization using models to identify the method most capable of reaching all selected method response goals on average. During robustness optimization, robustness models for selected method responses are created with Monte Carlo simulation and used to further optimize method parameters such that all method responses consistently reach their goals, as reflected by a process capability value of ≥ 1.33 , which is the established standard for a robust process (NIST/SEMATECH e-Handbook of statistical methods 2012a, 2012b, 2012c, 2012d).

Models are critical to the AQbD approach and must be validated both statistically and scientifically. Statistical validation is performed using various statistical tests such as residual randomness and normality (NIST/SEMATECH e-Handbook of statistical methods 2012a, 2012b, 2012c, 2012d), regression R-squared, adjusted regression R-squared, and prediction R-squared. Scientific validation is achieved by checking the terms in a statistical model against the relevant established scientific principles, which is described as mechanistic understanding in the relevant literature (ICH Q8 (R2) 2009).

Table 1 The experimental matrix of the modified CCC design

Run No.	(A) Pump flow rate (mL/min)	(B) Final strong solvent (%)	(C) Oven temperature (°C)	(D) EDTA additive concentration (mM)	Predicted response variance (%)
1 ^a	0.9	35.0	26.0	0.50	16.67
2 ^b	0.8	33.0	28.0	0.45	10.80
3 ^b	1.0	33.0	28.0	0.45	10.87
4 ^b	0.8	37.0	28.0	0.45	10.87
5 ^b	1.0	37.0	28.0	0.45	16.67
6 ^b	0.8	33.0	28.0	0.55	16.85
7 ^b	1.0	33.0	28.0	0.55	17.02
8 ^b	0.8	37.0	28.0	0.55	17.02
9 ^b	1.0	37.0	28.0	0.55	17.21
10 ^d	0.8	33.0	28.0	0.45	10.80
11 ^d	1.0	33.0	28.0	0.45	10.87
12 ^d	0.8	37.0	28.0	0.45	10.87
13 ^a	0.7	35.0	30.0	0.50	16.91
14 ^a	1.1	35.0	30.0	0.50	17.14
15 ^a	0.9	31.0	30.0	0.50	16.91
16 ^a	0.9	39.0	30.0	0.50	17.14
17 ^a	0.9	35.0	30.0	0.40	16.67
18 ^a	0.9	35.0	30.0	0.60	17.41
19 ^c	0.9	35.0	30.0	0.50	10.00
20 ^c	0.9	35.0	30.0	0.50	10.00
21 ^c	0.9	35.0	30.0	0.50	10.00
22 ^b	0.8	33.0	32.0	0.45	16.85
23 ^b	1.0	33.0	32.0	0.45	17.02
24 ^b	0.8	37.0	32.0	0.45	17.02
25 ^b	1.0	37.0	32.0	0.45	17.21
26 ^b	0.8	33.0	32.0	0.55	17.32
27 ^b	1.0	33.0	32.0	0.55	16.16
28 ^b	0.8	37.0	32.0	0.55	16.16
29 ^b	1.0	37.0	32.0	0.55	16.16
30 ^a	0.9	35.0	34.0	0.50	17.41
Parameter ranges	0.7–1.1	33–37	26–34	0.4–0.6	

Each run is labeled with the type of design point: ^a, star points; ^b, factorial points; ^c, triplicate replications at the center point (the nominal conditions); and ^d, model robustness points by run 10, 11, and 12, which are the duplicate replications of the factorial points by run 2, 3, and 4, respectively

Fusion uses data transformation analysis to decide whether data transformation is necessary before modeling, and then uses analysis of variance (ANOVA) and regression to generate method response models. ANOVA provides objective and statistical rationale for each consecutive modeling decision. Model residual plots are fundamental tools for validating the final method response models. When a model fits the DOE data well, the response residuals should be distributed randomly without any defined structure, and normally. A valid method response model provides the deepest

understanding of how a method response, such as resolution, is affected by critical method parameters.

Since Fusion relies on models for chemistry screening, mean optimization, and robustness optimization, it is critical to holistically evaluate each method response model from all relevant model regression statistics to assure model validity before multiple method response optimization. Inappropriate models will lead to poor prediction and non-robust methods. This paper will describe the holistic evaluation approach used to develop a robust chromatographic method with Fusion QbD.

Results

Representative chromatogram under nominal conditions

Careful planning and pre-runs executed at select star points allowed for successful execution of the DOE with all expected peaks eluting within the running time for all the 30 runs. A representative chromatogram at the nominal conditions is shown in Fig. 1. The API peak (C1112299299-C) and the Epimer peak (C1112299299-C-epimer) can be seen, as well as seven minor impurity peaks, among which impurity 2 and impurity 3 elute at 8.90 and 10.51 min, respectively. The inset shows the full-scale chromatogram.

Results for statistical validation of the DOE models

ANOVA and regression data analysis revealed many DOE models for various peak responses. The major numeric regression statistics of the peak response models are summarized in Table 2.

MSR (mean square regression), MSR adjusted, and MS-LOF (mean square lack of fit) are major numeric statistics for validating a DOE model. A model is statistically significant when the $MSR \geq$ the MSR significance threshold, which is the 0.0500 probability value for statistical significance. The lack of fit of a model is not statistically significant when the $MS-LOF \leq$ the MS-LOF significance threshold, which is also the 0.0500 probability value for statistical significance. The MSR adjusted statistic is the MSR adjusted with the number of terms in the model to assure a new term improves the model fit more than expected by chance alone. For a valid model, the MSR adjusted is always smaller than the MSR and the difference is usually very small, unless too

many terms are used in the model or the sample size is too small.

Model Term Ranking Pareto Charts for scientific validation of DOE models

DOE models are calculated from standardized variable level settings. Scientific validation of a DOE model through mechanistic understanding can be challenging when data transformation before modeling ostensibly inverts the positive and negative nature of the model term effect. To overcome this challenge, Model Term Ranking Pareto Charts that provide the detailed effects of each term in a model were employed. See Fig. 2 for details.

The chart presents all terms of a model in descending order (left to right) based on the absolute magnitude of their effects. The primary y -axis (model term effect) gives the absolute magnitude of individual model terms, while the secondary y -axis (cumulative percentage) gives the cumulative relative percentage effects of all model terms. Blue bars correspond to terms with a positive effect, while gray bars correspond to those with a negative effect. The Model Term Ranking Pareto Charts for all models are summarized in Fig. 2, except the two “customer” peak area models with a single term and the two C_{pk} models.

Discussion

AQbD relies on models for efficient and effective chemistry screening, mean optimization, and robustness optimization of chromatographic methods. It is critical to “validate” the models both statistically and scientifically, as inappropriate models may lead to impractical methods. As such, this section will discuss statistical and

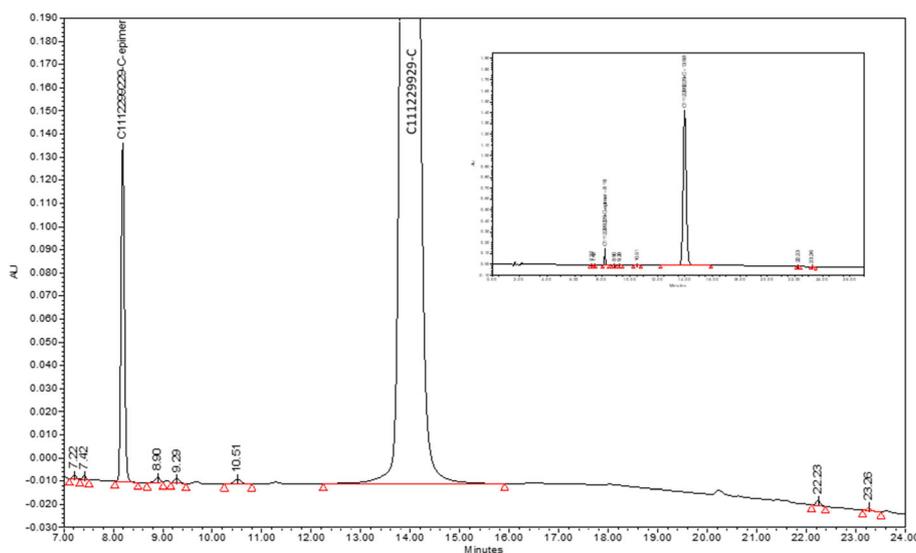


Fig. 1 A representative chromatogram under nominal conditions

Table 2 Major model regression statistics of various chromatographic responses

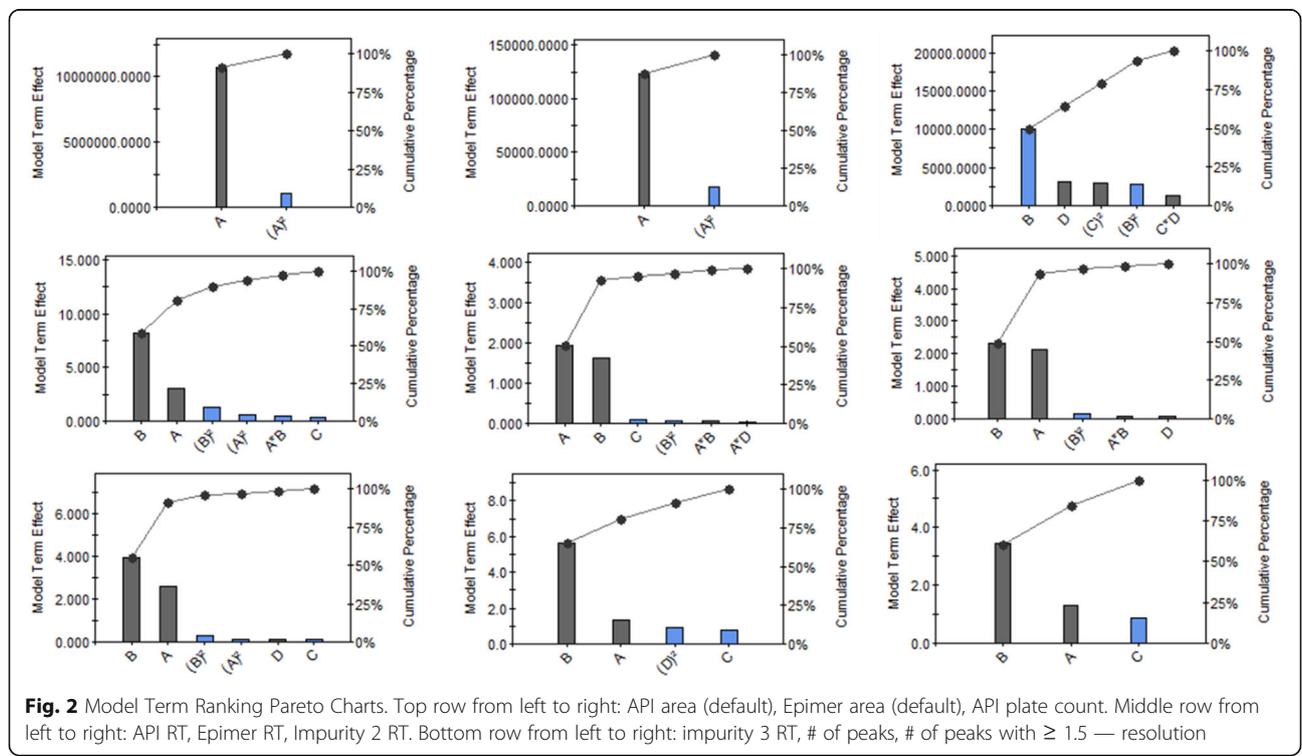
Response	MSR	MSR adjusted	MSR threshold	MS-LOF	MS-LOF threshold
API area (default)	0.9913	0.9906	0.0022	0.0006	0.0041
Epimer area (default)	0.9902	0.9895	0.0024	0.0008	0.0026
API area (customer)	0.8191	0.8124	0.0282	0.0081	0.0026
Epimer area (customer)	0.8481	0.8425	0.0237	0.0068	0.0017
API plate count	0.9101	0.8906	0.0516	0.0206	0.0716
API RT (min)	0.9988	0.9984	0.0009	0.0003	0.0028
Epimer RT (min)	0.9986	0.9982	0.0009	0.0004	0.0014
Impurity 2 RT (min)	0.9988	0.9986	0.0006	0.0002	0.0014
Impurity 3 RT (min)	0.9983	0.9979	0.0011	0.0004	0.0022
# of peaks	0.8582	0.8356	0.0626	0.0250	0.0613
# of peaks with ≥ 1.5 — resolution	0.7928	0.7889	0.0711	0.0275	0.0400
API plate count C_{pk}	0.9949	0.9924	0.0059	N/A	N/A
# of peaks with ≥ 1.5 — resolution C_{pk}	0.9999	0.9999	0.0001	N/A	N/A

scientific validation of the DOE models. After the models were fully validated for all selected individual method responses, the method MODR was substantiated by balancing and compromising among the most important method responses.

Statistical validation of the DOE models

As shown in Table 2, the MSR values ranged from 0.7928 to 0.9999. All MSR values were much higher

than their respective MSR threshold, which ranged from 0.0006 to 0.0711, indicating that all models were statistically significant and explained the corresponding chromatographic response data. The MSR adjusted values were all smaller than their respective MSR values, and the differences between the two was always very small (the largest difference was 0.0195 for the API plate count model), indicating that there was no overfitting for the models. There was slight



lack of fit for the two customer models due to very low pure errors, and the MS-LOF cannot be calculated for the two C_{pk} model because the Monte Carlo simulation gives essentially zero pure error. Other than that, the MS-LOF \leq the MS-LOF significance threshold for all other models, indicating the lack of fit was not statistically significant.

In addition to the above numeric statistical validation, various model residual plots were employed for graphical statistical model validation. The parameter-residual plots and the run number-residual plots for all models showed no defined structure, indicating random residual distribution. The normal probability plots showed all residual points lay in a nearly straight line for each single model, indicating normal residual distribution for all models. The randomly and normally distributed residuals provided the primary graphical statistical validation of the DOE models. See Fig. 3 for representative residuals plots for the “# of Peaks” model.

Scientific validation of the DOE models

With all models statistically validated, the discussions below will focus on scientific validation of the models by mechanistic understanding.

Peak area models for API and Epimer peaks

Peak areas and peak heights have been used for chromatographic quantification. However, peak area was chosen as the preferred approach as it is less sensitive to peak distortions such as broadening, fronting, and tailing, which can cause significant variation in analyte quantitation. To use peak area to reliably quantify the analyte within the MODR of a robust chromatographic method, the peak area must remain stable with consistent analyte injections.

Peak area models can be critical to the method development and validation with multivariate DOE approach. Solid peak area models were revealed for the API and Epimer peaks in this study. See the “API (default)” and “Epimer (default)” rows in Table 2 for the detailed model

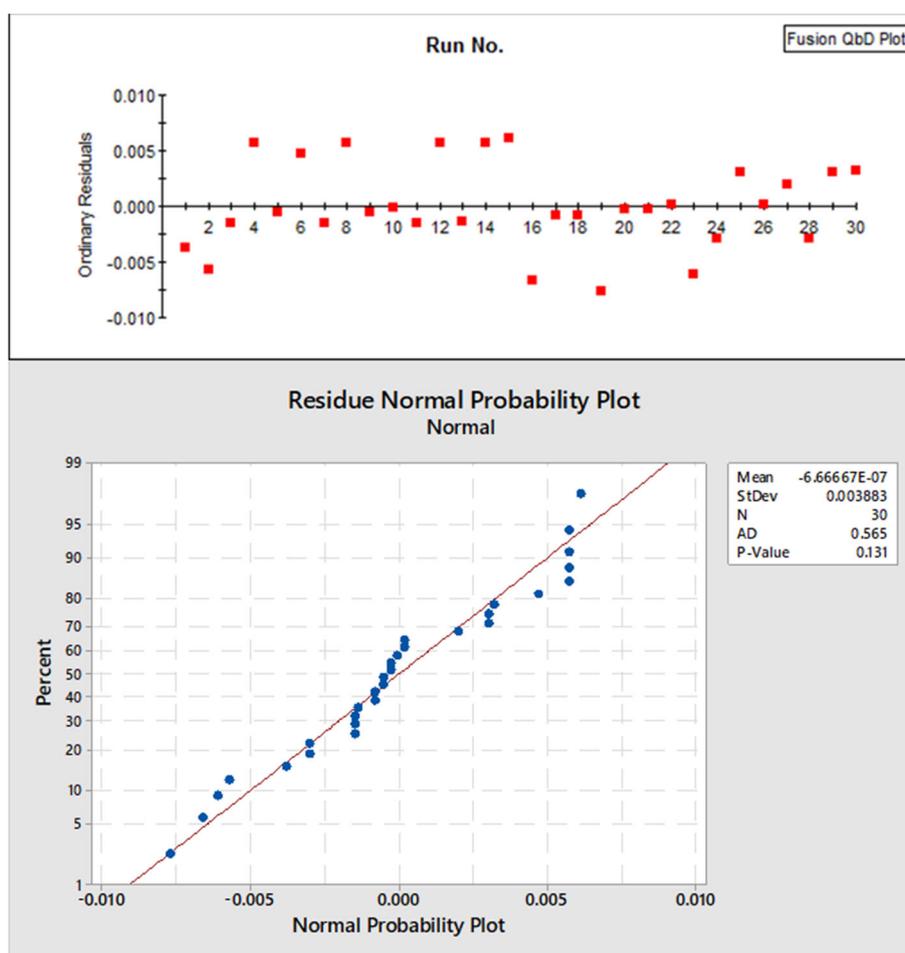


Fig. 3 Representative residuals plot for the “# of Peaks” model. Upper: run no – residuals plot; lower: residues normal probability plot

regression statistics. See Fig. 2 for the Model Term Ranking Pareto Charts. See Eqs. 2 and 3 below for the detailed models.

$$\begin{aligned} \text{API Peak Area} = & +23,438,293.01 - 2,672,900.45(A) \\ & + 268,590.01(A)^2 \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Epimer Peak Area} = & +270,051.96 - 30,896.51(A) \\ & + 4,332.26(A)^2 \end{aligned} \quad (3)$$

Although a full quadratic model for the four-factor five-level CCC design has a total of fourteen potential terms, multivariate regression analyses revealed that only two of the fourteen terms are statistically significant for both the API and Epimer peak area models. In addition, the flow rate term and flow rate squared-terms are identical for the two models, indicating the other three parameters (final percentage strong solvent, oven temperature, and EDTA concentration) have no significant effect on peak area for both peaks.

Oven temperature and EDTA concentration have negligible effect on peak area and thus were not significant terms in the peak area models. The percentage of strong solvent was also not a significant term in the peak area models even though it did appear to influence peak height almost as much as flow rate, but not the peak area, as seen in Fig. 4. It was hypothesized that the two flow rate terms in the model consisted of a strong negative first order term and a weak positive second order term, but more investigation was needed.

Peak purity and peak integration are the primary factors affecting peak area. Partial or total peak overlap (resolution < 1.5) due to analyte co-elution can impact the peak purity resulting in inaccurate integration of both peaks. Peak integration may also be affected by unstable baseline and/or peak fronting and tailing due to

uncertainty in determining peak start and end points. In this DOE study, the API and Epimer peaks were consistently well-resolved (resolution ≥ 2.0) and were also significantly higher than the limit of quantitation, contributing to the strong peak area models. In contrast, no appropriate peak area models could be developed for other impurity peaks as they were either not properly resolved or were too close to the limit of quantitation. For peaks with resolution ≤ 1.0 there will likely never be an area model with reliable predictivity as the peak area cannot be consistently and accurately measured.

The importance of a mechanistic understanding of the DOE models for AQbD has been extensively discussed. The API and Epimer peak area models were very similar in that they both contained a strong negative first order flow rate term and a weak positive second order flow rate term.

The strong negative first order term can be explained by the exposure time of the analyte molecules to the detector. The UV detector used in the LC method is non-destructive and concentration sensitive. Analyte molecules send signals to the detector when exposed to UV light while flowing through the fixed length detecting window in a band. As the molecules are not degraded by the UV light, the slower the flow rate, the longer the analyte molecules are exposed to the UV light, allowing for increased signal to the detector and thus increased analyte peak area. Simple direct linear regression of the peak area against inverse flow rate confirmed both the API and Epimer peak areas were proportional to the inverse flow rate, with R^2 values ≥ 0.99 (data not included).

As there was no obvious mechanistic explanation of the weak positive second order term in the models, more investigation was needed. Multivariate DOE customer models were pursued. The acquired customer models, listed in Eqs. 4 and 5, used inverse flow rate "1/A" in place of the flow rate "A" for all pertinent terms among the fourteen terms. The major model regression

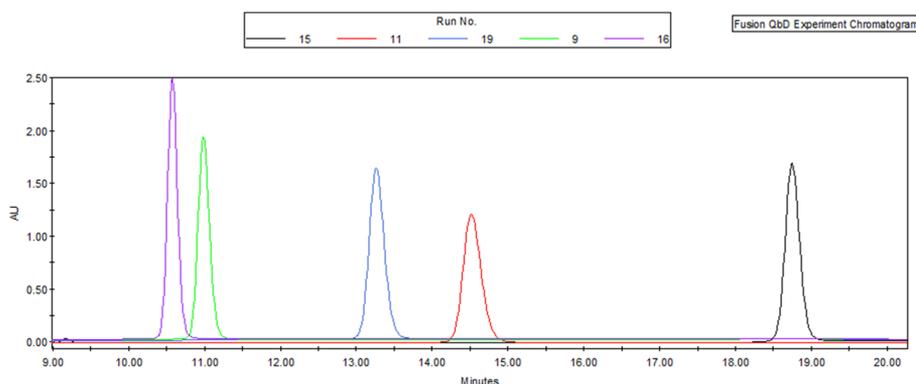


Fig. 4 Effects of final percentage of strong solvent and flow rate on the API peak area and peak height: run 15 (black) = 31%/0.9 mL/min; run 11 (red) = 33%/1.0 mL/min; run 19 (blue) = 35%/0.9 mL/min; run 9 (green) = 37%/1.0 mL/min; run 16 (purple) = 39%/0.9 mL/min

statistics of the customer models are summarized in the “API (customer)” and “Epimer (customer)” rows in Table 2. Both customer models contain a single inverse flow rate term, confirming the negative effect of flow rate on peak area for both peaks. The customer models in Eqs. 4 and 5 provide more intuitive understanding of the flow rate effects on peak area than the “default” models in Eqs. 2 and 3. The weak positive second order flow rate term in Eqs. 2 and 3 contributes less than 15% effect to the peak area and is very challenging to explain mechanistically. This kind of model term replacing technique may be of general value when using DOE to explore and discover new scientific theory, including new chromatographic theory.

Additionally, the peak area models in Eqs. 2–5 revealed that the pump flow rate must be very consistent among all injections during a quantitative chromatographic sequence. Currently, the best-in-industry flow rate precision for a binary UPLC pump is “< 0.05% RSD or < 0.01 min SD” (Thermo Fisher Scientific, Vanquish Pump Specification, 2021).

$$\text{API Peak Area} = +21,403,748.42 + 100,731.30((1/A)) \quad (4)$$

$$\text{Epimer Peak Area} = +247,286.48 + 1,173.90((1/A)) \quad (5)$$

API peak plate count model

Column plate count is potentially useful in DOE modeling as it is a key parameter used in all modes of chromatography for measuring and controlling column efficiency to assure separation of the analytes. The equation for plate count (N) is shown below. It is calculated using peak retention time (t_r) and peak width at half height ($w_{1/2}$) to mitigate any baseline effects and provide a more reliable response for modeling-based QbD chromatographic method development.

$$N = 5.54 \frac{(t_r)^2}{(w_{1/2})^2}$$

The peak plate count model for the API peak can be seen in Eq. 6. It was developed by reducing the fourteen terms. The major model quality attributes are summarized in Table 2.

$$\begin{aligned} \text{API Peak Plate Count} = & +19,991.86 \\ & +3,322.41(B) - 767.71(D) \\ & +720.10(B)^2 - 741.62(C)^2 - 664.70(C*D) \end{aligned} \quad (6)$$

The flow rate was not a critical factor in the plate count model. This seemingly goes against the Van

Deemter equation (van Deemter et al. 1956), which states that flow rate directly affects column plate height and thus plate count. However, the missing flow rate term can be rationalized by the LC column that was used. According to the Van Deemter equation, plate height for the 150 × 4.6 mm, 3.5 μm column will remain flat at a minimum level within the 0.7–1.1 mL/min flow rate range used in this DOE study (Altiero 2018). As plate count is inversely proportional to the plate height, it will also remain flat at a maximal level within the 0.7–1.1 mL/min flow rate range.

The most dominating parameter in the API plate count model was the final percentage of strong solvent. Its two terms B and B^2 provided more than 60% positive effects to the plate count response (see the Model Term Ranking Pareto Chart in Fig. 2) and could be easily explained by the inverse relationship between plate count and peak width when the gradient slope is increased.

Retention time models

Retention time (RT) and peak width are the primary attributes for a chromatographic peak. They are used to calculate secondary attributes such as resolution, plate count, and tailing. These peak attributes together define the overall quality of separation and subsequently quantification of the analytes. RT is determined using all data points on a peak and is thus a more reliable measurand than peak width, which uses only some data points on a peak. As such, peak width cannot provide the same level of RT accuracy, especially for minor peaks, due to uncertainty in the determination of peak start, end, and height. Consequently, RT is the most reliably measured peak attribute.

The reliability of the RT measurement was confirmed in this DOE study. As listed in Table 2, well-fitted RT models were acquired for the major API and Epimer peaks as well as the minor impurity 2 and impurity 3 peaks. The retention time models are listed in Eqs. 7–10 (note: reciprocal square for the Epimer and impurity 2, and reciprocal for impurity 3 retention time data transformation before modeling inverted the positive and negative nature of the model term effect in Eqs. 8–10, see the Model Term Ranking Pareto Charts in Fig. 2 for the actual effect). The four models shared three common terms: flow rate, final percentage of strong solvent, and the square of final percentage of strong solvent. These three terms contributed more than 90% of the effect in all four RT models. Furthermore, in all four models the flow rate and final percentage of strong solvent terms consistently produced a negative effect on RT, whereas the square of the final percentage of strong solvent term consistently produced positive effects. While the scientific rationale for the negative effects of the first two terms is well-established, the rationale for

the positive effects of the third term lies beyond the scope of this study.

$$\begin{aligned} \text{API RT} = & 13.37632 - 0.99845(A) - 2.05511(B) \\ & + 0.08980(C) + 0.14804(A)^2 \\ & + 0.33645(B)^2 + 0.21846(A*B) \end{aligned} \quad (7)$$

$$\begin{aligned} (\text{Epimer RT})^{-1/2} = & 0.01520 + 0.00176(A) \\ & + 0.00149(B) - 0.00008(C) - 0.00007(B)^2 \\ & + 0.00014(A*B) - 0.00006(A*D) \end{aligned} \quad (8)$$

$$\begin{aligned} (\text{Imp 2 RT})^{-1/2} = & 0.01306 + 0.00154(A) \\ & + 0.00165(B) \\ & + 0.00005(D) - 0.00011(B)^2 \\ & + 0.00012(A*B) \end{aligned} \quad (9)$$

$$\begin{aligned} (\text{Imp 3 RT})^{-1} = & 0.09801 + 0.00619(A) \\ & + 0.00918(B) - 0.00022(C) \\ & + 0.00025(D) - 0.00026(A)^2 - 0.00072(B)^2 \end{aligned} \quad (10)$$

As RT is typically the most reliable measured peak response, therefore, it produces most reliable models. One potential shortcoming of RT modeling-based method optimization is that the resolution of two neighboring peaks is not only affected by the retention time, but also by peak width and peak shape, such as peak fronting and tailing.

Peak number models

A representative analytical sample is critical for AQbD to use DOE to develop a chromatographic method capable of resolving all potential related substances. Multivariate DOE chromatography of a forced degradation sample may contain many minor peaks, which may elute in different orders across the different runs of the study, making tracking of the individual peaks nearly impossible. One way to solve this problem is to focus on the number of peaks observed, instead of tracking of individual peaks. Furthermore, to avoid an impractical method with too many partially resolved peaks, the number of peaks with ≥ 1.5 resolution could be an alternative response for modeling.

Excellent models were acquired for both the number of peak responses and the number of peaks with ≥ 1.5 resolution in this DOE study. See Table 2 for the major model statistics, Fig. 2 for the Model Term Pareto Ranking Chart, and Eqs. 11 and 12 for the detailed models (*note*: reciprocal square data transformation before modeling reversed the positive and negative nature of the model term effect in Eqs. 11–12; see the Model Term Ranking Pareto Charts in Fig. 2 for the actual effect). Of the 14 terms, only

four were statistically significant for the peak number model and only three were statistically significant for the resolved peak number model. Additionally, it is notable that the two models share three common terms (final percentage of strong solvent (B), flow rate (A), and oven temperature (C)) and the orders of impact for the three terms is maintained as $(B) > (A) > (C)$, as seen in the Model Term Ranking Pareto Chart. The models indicated that within the evaluated ranges the final percentage of strong solvent and flow rate have negative effects on the overall separation, while column temperature has a positive effect. These observations align well with chromatographic scientific principles.

$$\begin{aligned} (\text{No. of Peaks})^{-1/2} = & +0.028 + 0.003(A) \\ & + 0.009(B) - 0.002(C) - 0.002(D)^2 \end{aligned} \quad (11)$$

$$\begin{aligned} (\text{No. of Peaks} \geq 1.5 - \text{USP Resolution})^{-1/2} \\ = & +0.039 + 0.005(A) + 0.015(B) - 0.003(C) \end{aligned} \quad (12)$$

Challenges and solutions to peak resolution modeling

No appropriate model was found for the API peak resolution response in this study, possibly due to very high pure experimental error (34.2%) based on the replication runs. With this elevated level of resolution measurement error, only large effects of the experiment variables would be discernable from an analysis of the resolution data. There are many potential reasons for the high pure experimental error: (1) error in the resolution value determination in each DOE run, especially with small peak size or tailing of the reference impurity peaks; (2) the use of different reference peaks to calculate the resolution when elution order shifts between DOE runs; (3) the column is not sufficiently re-equilibrated between different conditions (note: Mention of column equilibration was hypothetical in this case and only to stress the importance of column conditioning during DOE in general. As Fusion QbD automatically inserts conditioning runs into the DOE sequence where needed, this was not found to be an issue in this case study). The respective solutions to overcome these challenges are (1) when reference materials are available, make a synthetic method-development sample composed of each analyte at concentrations at least ten times the limit of quantitation; (2) keep the concentration of analytes in the synthetic sample at distinguishably different levels so that the peaks can be tracked by size; and (3) allow enough time for the column to be sufficiently re-equilibrated between different conditions.

Table 3 Critical method parameter settings for the optimized method

Variable	Level setting
Pump flow rate (mL/min)	0.78
Final % strong solvent (%)	34.2
Oven temperature (°C)	30.8
EDTA concentration (mM)	0.42

Method robustness evaluation and optimization by Monte Carlo simulation

The robustness of a method is a measure of its capacity to remain unaffected by small but deliberate variations in method parameters. It provides an indication of the method's reliability during normal usage. Robustness was demonstrated for critical method responses by running system suitability checks, in which selected method parameters were changed one factor at a time. In comparison, the AQBd approach quantifies method robustness with process robustness indices, such as C_p and C_{pk} , through multivariate robustness DOE, in which critical method parameters are systematically varied, simultaneously. Process robustness indices are standard statistical process control matrices widely used to quantify and evaluate process and product variations. In this AQBd case study, method capability indices were calculated to compare the variability of a chromatographic method response to its specification limits. The comparison is made by forming the ratio between the spread of the response specifications and the spread of the response values, as measured by six times standard deviation of the response. The spread of the response values is acquired through tens of thousands of virtual Monte Carlo simulation runs of the corresponding response model, with all critical method parameters varied around their setting points randomly and simultaneously according to specified distributions. A method with a process capability of ≥ 1.33 is considered robust as it will only fail to meet the response specifications 63 times out of a million runs and thus is capable of providing much more reliable measurements for informed decisions on drug development, manufacturing, and quality control. Due to its intrinsic advantages over the OFAT approach, multivariate DOE robustness evaluation was recommended to replace the OFAT approach in the latest

regulatory guidelines (FDA Guidance for industry-analytical procedures and methods validation for drugs and biologics. 2015).

In this DOE study, solid C_{pk} models were produced for the "API Plate Count" and "Number of Peaks ≥ 1.5 USP Resolution". See Table 2 for the detailed model regression statistics.

Multiple responses method optimization

Once models have been established for selected individual method responses, overall method evaluation and optimization can be performed. This is usually substantiated by balancing and compromising among multiple method responses. Three principles must be followed in selecting method responses to be included in the final optimization: (1) the selected response is critical to achieve the goal (see Table 4); (2) a response is included only when its model is of sufficiently high quality to meet the goals of validation; and (3) the total number of responses included should be kept to a minimum.

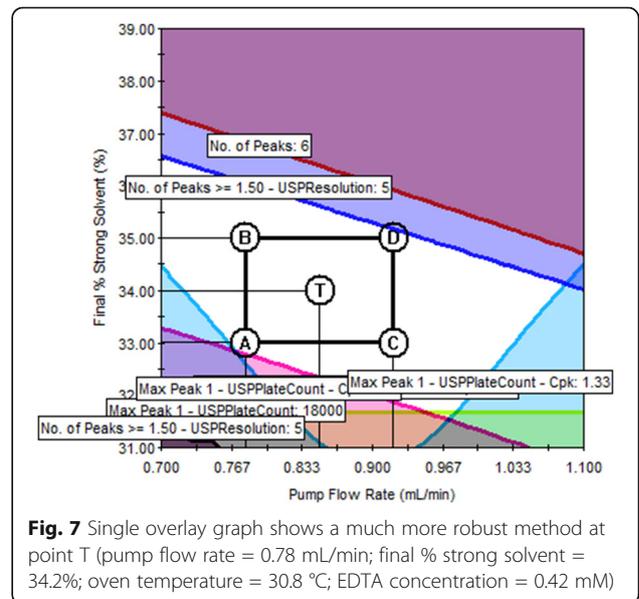
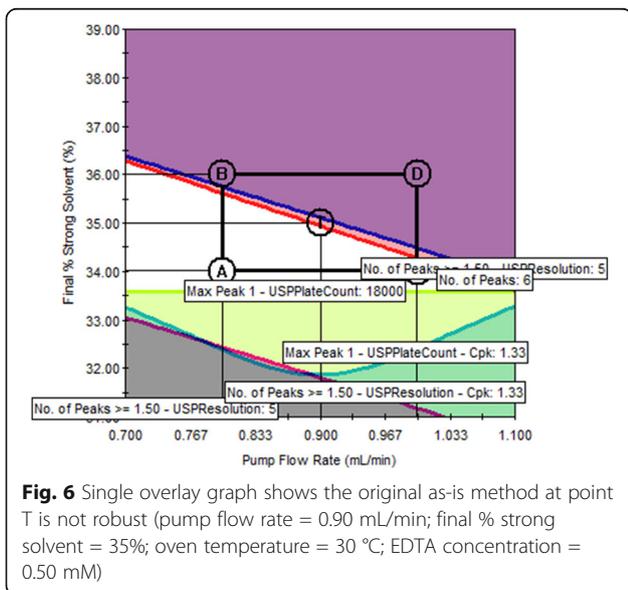
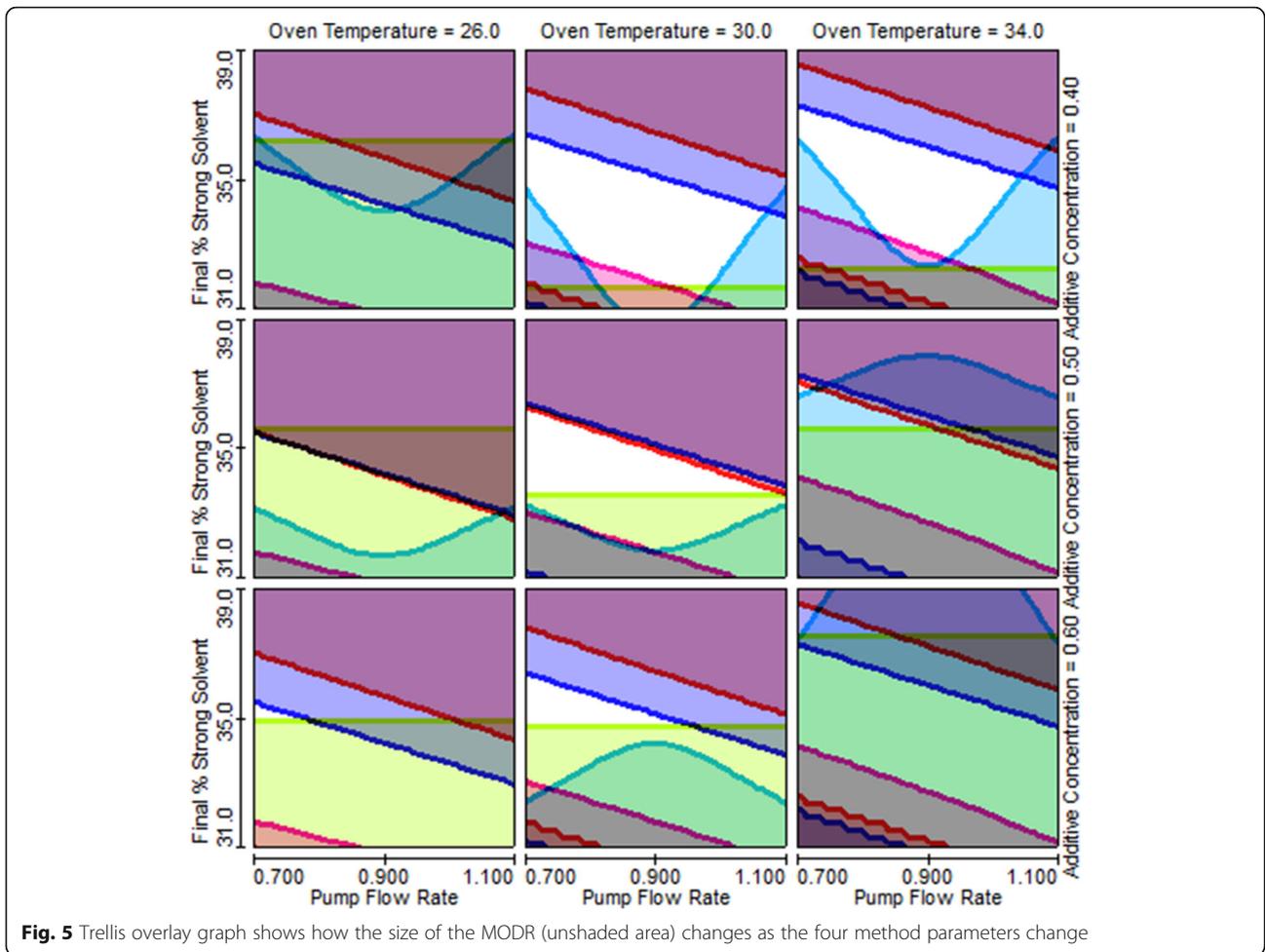
Following the above three principles, five method responses were selected for the overall method evaluation and optimization. Best overall answer search identified a new optimized method when the four critical method parameters were set at the specific values as listed in Table 3. The cumulative desirability for the five desired method response goals reached the maximum value of 1.0. The desirability for each individual goal also reached the maximum value of 1.0, as listed in Table 4.

Method Operable Design Region (MODR)

The critical method parameter settings in Table 3 define a single method that can simultaneously fulfill all five targeted method goals listed in Table 4 to the best extent possible. However, the actual operational values of the four critical parameters may drift around their set points during routine method executions. Based on the models, contour plots for method response can be created to reveal how the response value changes as the method parameters drift. Furthermore, overlaying the contour plots of all selected method responses reveal the MODR, as shown in Figs. 4, 5, and 6. Note that for each response, a single unique color is used to shade the region of the graph where the response fails the criteria;

Table 4 Predicted results and confidence intervals for the selected responses of the optimized method

Response	Goal	Predicted result	Desirability	- 2 Sigma conf. limit	+ 2 Sigma conf. limit
No. of peaks	Maximize	8.0	1.0000	6.3	10.5
No. of peaks ≥ 1.5 — USP resolution	Maximize	6.2	1.0000	4.8	9.5
Max peak 1 — USP plate count	Maximize	20,313	1.0000	17,560	23,067
Max peak 1 — USP plate count — C_{pk}	Maximize	2.01	1.0000	1.78	2.24
No. of peaks ≥ 1.5 — USP resolution — C_{pk}	Maximize	3.59	1.0000	3.54	3.63



thus, criteria for all responses are met in the unshaded area.

The Trellis overlay graph in Fig. 5 reveals the MODR from the perspectives of all four critical method parameters, among which flow rate and final percentage of strong solvent change continuously while oven temperature and EDTA additive concentration were each set at three different levels. Figure 5 clearly demonstrates how the size of the MODR changes with the four method parameters. The single overlay graph in Fig. 6 shows that the original as-is method (represented by the center point T) is on the edge of failure for two method responses, number of peaks (red) and number of peaks ≥ 1.5 resolution (blue), indicating that the original method is not robust. Conversely, point T in the single overlay graph in Fig. 7 is at the center of a relatively large unshaded area, indicating that the method is much more robust than the original method.

Conclusion

Through the collaboration of regulatory authorities and the industry, AQbD is the new paradigm to develop robust chromatographic methods in the pharmaceutical industry. It uses a systematic approach to understand and control variability and build robustness into chromatographic methods. This ensures that analytical results are always close to the product true value and meet the target measurement uncertainty, thus enabling informed decisions on drug development, manufacturing, and quality control.

Multivariate DOE modeling plays an essential role in AQbD and has the potential to elevate chromatographic methods to a robustness level rarely achievable via the traditional OFAT approach. However, as demonstrated in this case study, chromatography science was still the foundation for prioritizing method inputs and responses for the most appropriate DOE design and modeling, and provided further scientific validation to the statistically validated DOE models. Once models were fully validated for all selected individual method responses, the MODR was substantiated by balancing and compromising among the most important method responses.

Developing a MODR is critical for labs that transfer in externally sourced chromatographic methods. In this case study, method evaluation using AQbD produced objective data that enabled a deeper understanding of method variability, upon which a more robust method with a much larger MODR was proposed. The in-depth method variability understanding through AQbD also paved the way for establishing a much more effective method control strategy. Method development and validation from a multivariate data driven exercise led to better and more informed decisions regarding the suitability of the method.

Abbreviations

AQbD: Analytical quality by design; DOE: Design of experiments; CCC: Circumscribed central composite; MODR: Method Operable Design Region; QbD: Quality by design; ICH: The International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use; CQA: Critical quality attributes; TMU: Target measurement uncertainty; ATP: Analytical target profile; ANOVA: Analysis of variance; EDTA: Ethylenediaminetetraacetic acid; MSR: Mean square regression; MS-LOF: Mean square lack of fit; OFAT: One factor at a time

Acknowledgements

The authors would like to thank KBP Biosciences for reviewing and giving permission to publish this case study. They would also like to thank Thermo Fisher Scientific and S Matrix for the Fusion QbD software, Lynette Bueno Perez for solution preparations, Dr. Michael Goedecke for statistical review, and both Barry Gujral and Francis Vazquez for their overall support.

Authors' contributions

YD designed the study and performed the data analysis. ZL was the primary scientist that executed the study. CL, EP, AP, TT, and WW contributed ideas and information to the study and reviewed and approved the manuscript. The authors read and approved the final manuscript.

Funding

Not applicable; authors contributed case studies based on existing company knowledge and experience.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 27 May 2021 Accepted: 29 July 2021

Published online: 01 September 2021

References

- Altiero, P. Why they matter, an introduction to chromatography equations. Slide 21. https://www.agilent.com/cs/library/eseminars/public/Agilent_Webinar_Why_They_Matter_An_Intro_Chromatography_Equations_Nov262018.pdf. Accessed 13 May 2021. (2018).
- Bezerra MA, Ferreirab SLC, Novaesa CG, dos Santoset AMP, Valasquesal GS, da Mata Cerqueira UMF, et al. Simultaneous optimization of multiple responses and its application in Analytical Chemistry – a review. *Talanta*; 194: 941-959. (2019).
- Borman P, Chatfield M, Nethercote P, Thompson D, Truman K (2007) The application of quality by design to analytical methods. *Pharma. J Technol* 31(12):142–152 (n.d.)
- Chatterjee S, CMC Lead for QbD, ONDQA/CDER/FDA. Design space considerations, AAPS Annual Meeting, 2012. (n.d.)
- Debrus B, Guillaume D, Rudaz S (2013) Improved quality-by-design compliant methodology for method development in reversed-phase liquid chromatography. *J Pharm Biomed Anal* 84:215–223 (n.d.)
- EURACHEM / CITAC. Setting and using target uncertainty in chemical measurement. 2015. (n.d.)
- FDA Guidance for industry-analytical procedures and methods validation for drugs and biologics. 2015. (n.d.)
- FDA pharmaceutical cGMPs for the 21st century — a risk-based approach. 2004. (n.d.)
- Grangeia HB, Silva C, Simões SP, Reis MS (2020) Quality by design in pharmaceutical manufacturing: a systematic review of current status, challenges and future perspectives. *Eur J Pharm Biopharm* 147:19–37 (n.d.)
- ICH Q10 - Pharmaceutical quality system. 2008. (n.d.)
- ICH Q11 - Development and manufacturing of drug substances (chemical entities and biotechnological/biological entities). 2012. (n.d.)
- ICH Q12 - Technical and regulatory considerations for pharmaceutical product lifecycle management. 2019. (n.d.)

- ICH Q14 - Analytical procedure development and revision of Q2(R1) analytical validation - final concept paper. 2018. (n.d.).
- ICH Q8 (R2) - Pharmaceutical development. 2009. (n.d.).
- ICH Q9 - Quality risk management. 2005. (n.d.).
- Leardi R (2009) Experimental design in chemistry: a tutorial. *Anal. Chim. Acta* 652(1–2):161–172 (n.d.)
- Myers RH, Montgomery DC (1995) *Response surface methodology: process and product optimization using designed experiments*, 2nd edition. John Wiley & Sons, New York, pp 366–404 (n.d.)
- NIST/SEMATECH e-Handbook of statistical methods. 2012a. <http://www.itl.nist.gov/div898/handbook/ppc/section1/ppc133.htm>. Accessed May 13, 2021. (n.d.).
- NIST/SEMATECH e-Handbook of statistical methods. 2012b. <https://www.itl.nist.gov/div898/handbook/pmc/section1/pmc16.htm>. Accessed May 13, 2021. (n.d.).
- NIST/SEMATECH e-Handbook of statistical methods. 2012c. <https://www.itl.nist.gov/div898/handbook/pmd/section4/pmd44.htm>. Accessed May 13, 2021. (n.d.).
- NIST/SEMATECH e-Handbook of statistical methods. 2012d. <https://www.itl.nist.gov/div898/handbook/pri/section5/pri52.htm>. Accessed May 13, 2021. (n.d.).
- Orlandini S, Pinzauti S, Furlanetto S (2013) Application of quality by design to the development of analytical separation methods. *Anal Bioanal Chem* 2:443–450 (n.d.)
- Schweitzer M, Pohl M, Hanna-Brown M, Nethercote P, Borman P, Hansen P, Smith K et al (2010) Implications and opportunities for applying QbD principles to analytical measurements. *Pharma. Technol* 34(2):52–59 (n.d.)
- Tang YB, FDA/CDER/ONDQA (2011) *Quality by design approaches to analytical methods -- FDA perspective*. AAPS, Washington DC (n.d.)
- Thermo Fisher Scientific, Vanquish pump specification. 2021. <https://assets.thermofisher.com/TFS-Assets/CMD/Specification-Sheets/ps-73056-vanquish-pumps-ps73056-en.pdf>. Accessed May 22, 2021. (n.d.).
- van Deemter JJ, Zuiderweg FJ, Klinkenberg A. Longitudinal diffusion and resistance to mass transfer as causes of non ideality in chromatography. 1956. (n.d.).
- Volta e Sousa L, Gonçalves R, Menezes JC, Ramos A (2021) Analytical method lifecycle management in pharmaceutical industry: a review. *AAPS PharmSciTech* 22(3):128–141. <https://doi.org/10.1208/s12249-021-01960-9> (n.d.)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
