A Design Space Exploration for Control of Critical Quality Attributes of mAb

Hemlata Bhatia^a, Erik Read^{b+}, Cyrus Agarabi^{b+}, Kurt Brorson^{b+}, Scott Lute^{b+} and Seongkyu Yoon^{a*}

^aDepartment of Chemical Engineering, University of Massachusetts Lowell, Lowell, MA

^b Division II, Office of Biotechnology Products, Office of Pharmaceutical Quality, CDER, FDA,
Silver Springs, MD, USA

*Corresponding Author, Department of Chemical Engineering, University of Massachusetts Lowell, 1 University Ave, Lowell MA - 01854, USA; E-mail: seongkyu_yoon@uml.edu; Tel: 978 934 4741; Fax: 978 934 3047

+ Disclaimers

This article reflects the views of the authors and should not be constructed to represent official FDA's views or policies.

Running Title: “Design Space Exploration Strategy”

ABSTRACT

A unique “design space (DSp) exploration strategy,” defined as a function of four key scenarios, was successfully integrated and validated to enhance the DSp building exercise, by increasing the accuracy of analyses and interpretation of processed data. The four key scenarios, defining the strategy, were based on cumulative analyses of individual models developed for the Critical Quality Attributes (23 Glycan Profiles) considered for the study. The analyses of the CQA estimates and model performances were interpreted as (1) Inside Specification/Significant Model (2) Inside Specification/Non-significant Model (3) Outside Specification/Significant Model (4) Outside Specification/Non-significant Model. Each scenario was defined and illustrated through individual models of CQA aligning the description. The R^2 , Q^2 , Model Validity and Model Reproducibility estimates of G2, G2FaGbGN, G0 and G2FaG2, respectively, signified the four scenarios stated above. Through further optimizations, including the estimation of Edge of Failure and Set Point Analysis, wider and accurate DSps were created for each scenario, establishing critical functional relationship between Critical Process Parameters (CPPs) and Critical Quality Attributes (CQAs). A DSp provides the optimal region for systematic evaluation, mechanistic understanding and refining of a QbD approach. DSp exploration strategy will aid the critical process of consistently and reproducibly achieving predefined quality of a product throughout its lifecycle.

Key words: Design space, critical quality attribute, and critical process parameter.

1. INTRODUCTION

The Pharmaceutical Good Manufacturing Practice (cGMP) for the 21st Century Initiative of the Food and Drug Administration (FDA, 2004; Yu, 2008) has laid the responsibility of product quality on the industry, along with critical reformations in regulations of pharmaceutical manufacturing over the last decade. In addition to these reforms, guidelines from International Conference on Harmonization (ICH) Q8, Q9 and Q10 (ICH 2005; ICH 2008) have dramatically changed the review processes for assuring product quality. Together, these updated regulatory efforts provide the pharmaceutical and biotech industry a description of a quality drug product, and approaches to produce such a quality drug product, along with a model for an effective quality management system driven by current scientific and risk-based approaches. However, the pace of implementation of these regulatory guidelines has differed due to the complexity of biopharmaceutical products, resulting in increased documentation and prolonged review processes (Rathore and Winkle, 2009). In order to collectively assist both FDA and the industry in a more effective and consistent way, and to ensure efficiency in producing quality products and focused decision-making, the principles of Quality by Design (QbD) and Process Analytical Technology (PAT) were introduced (FDA, 2004).

QbD was a logical shift away from the traditional Quality by Testing (QbT) that largely relied on end product testing by restricting flexibility in manufacturing processes, resulting in inherent limitations in implementation of new technology, decreased predictability of scale-up processes and increased product waste (as high as 50% in some industry segments although not biotech) (Rathore and Winkle, 2009). Conversely, the QbD approach emphasized the quality being intrinsically built into the product by prompting the industry to move toward an empirical, preemptive system approach of drug development and manufacturing. The concept of QbD

fosters improved understanding of the process knowledge to steer efficient manufacturing of a product with defined product quality and performance (Sangshetti et al., 2014).

A typical life cycle of a pharmaceutical process involves defining, designing, characterizing, validating, monitoring and control of the process. The QbD paradigm seeks to build quality into the product at designing and developmental stages, based on the process improvement opportunities identified that are either introduced new or identified during previous process monitoring, to ensure a predefined quality and clinical performance at the end of the manufacturing cycle. PAT (Hinz, 2006) provides any necessary improvements to the QbD process that may effectively resolve previous quality issues and hence play a complementary role to QbD.

The concept of “DSp” (Design Space) has gained prominence in the past few years as an effective tool for implementation of QbD. ICH guidelines define DSp as “the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality” (ICH 2012). Even though the DSp was initially conceptualized in context of pharmaceutical processes, known as "Process DSp," it has expanded its utility to represent the clinical (Clinical DSp) and the product-quality aspects (Product DSp) of a product. The expanded concept of DSp establishes the relationship between a process, product quality and clinical properties of the product, with each Critical Quality Attribute (CQA) serving as a dimension within the multidimensional space. Hence, establishing a well-defined relationship between CQAs, Critical Process Parameters (CPPs) and Critical Material Attributes (CMAs) is at the core of understanding the application and utility of a DSp.

QbD paradigm and DSp strategy typically involve the following eight key steps i) Define target product profile (TPP), ii) Determine CQAs and CPPs, iii) Define product design space, iv)

85 Define process design space, v) Define control strategy, vi) Process validation, vii) Regulatory
86 filings and viii) Process monitoring, life-cycle management and continuous improvement. These
87 steps are usually in alignment with ICH guidelines (ICH guidance Q8 and ICH guidance Q8
88 Annex –2006) and have been implemented for various products ranging from small molecule
89 manufacture to monoclonal antibodies (mAbs), and various other biopharmaceuticals (Adam et
90 al., 2011; Ghirlando et al., 1999; Lionberger et al., 2008; Wu et al., 2011; Yu et al., 2014).
91 Defining a process DS_p further involves risk analysis, design of experiments (DOE) and
92 determining the process parameters likely to have greatest impact on the product quality.
93 “Working within this DS_p is not considered as a change. Movement out of the DS_p is
94 considered to be a change and would normally initiate a regulatory post-approval-change
95 process. DS_p is proposed by the applicant and is subject to regulatory assessment and
96 approval.” (ICH , 2012; Sellberg, 2013)

97 To establish such a DS_p that assures quality, various statistical analyses, model building
98 exercises, set-point analyses and optimization studies need to be carried out, with each step
99 producing large amounts of data. The decision to venture into further steps should be based on
100 careful understanding and interpretation of such data, which unfortunately has not been the norm
101 within the industry and has largely been neglected. This challenge has been further aggravated
102 due to inherent uncertainties within a DS_p, involving variability in experimental measurements,
103 model estimations, range of studies, and even the scale of experiments. Such uncertainties can
104 only be marginally addressed with statistical treatments and performance monitoring. In order to
105 aid the process comprehension and interpretation of such data, we have built a “DS_p exploration
106 strategy,” defining “four key scenarios” of decision making based on the parameters of modeling
107 (R^2 , Q^2 , Model Validity and Model Reproducibility). We have applied this strategy to an existing

dataset of inputs and outputs from parallel bioreactor mammalian cell cultures (Agarabi et al., 2015).

Key elements of this study included: (1) **Risk assessment** to identify the high impact process variables influencing the selected quality attributes (23 Glycan Profiles) of our model mAb; (2) **Plackett-Burman screening** performed to determine the most significant factors affecting quality of the model murine IgG3:K antibody producing hybridoma; (3) **Model development** and DSp exploration with "four scenarios," predefined based on summary of fit and model parameters (R^2 , Q^2 , model validity and reproducibility), to comprehend and interpret the DOE output and to define the exact relationship between various bioreactor process variables and the CQAs; (4) Finally, the intended DSp building by doing set point optimization, validating the robustness and accuracy of the generated model. The roadmap for design space building and the steps to improve the model and experimentation has been explained.

2. MATERIALS AND METHODS

2.1 Preliminary Risk analysis

This study was approached as DSp establishment initiative (Agarabi et al., 2015), beginning with risk assessment of bioreactor cell culture to identify potential high impact process variables influencing the quality of mAbs. Eleven process parameters were considered for a DOE to test and confirm the hypotheses of the risk analysis. For the risk analysis, the Fc N-glycosylation modulates key biological activities of mAbs. Essential variations of protein glycosylation patterns are rooted in the micro-heterogeneity of terminal galactose and mannose residues, which range from G0F (no galactose), G1F (one galactose) and G2F (2 galactoses) to various residues of mannose (M3, M4, M5, M6 etc.)(Ghirlando et al., 1999; Liu et al., 2006; Wu

et al., 2011). Such varying levels of terminal galactose and mannose, of the N-linked glycans, require careful characterization at the drug substance level. Varying levels of galactose and mannose of N-linked glycans have been reported to play a key role in humoral immune responses of various therapeutic mAbs, as well as *in vivo* effector functions, immunogenicity, plasmatic clearance, protease resistance and overall shelf life of formulated product. (Dowling et al., 2007; Hossler et al., 2009; Liu et al., 2015; Raju and Scallon, 2007; Zheng et al., 2011) These vital biological considerations guided us to select the glycan profiles of mAbs as our CQAs for this study, as a focus for identifying CPPs through DOE. Consequently, all likely process factors were systematically considered based on CDER's in-house experience and a risk analysis was performed from literature review to consider only likely parameters that could significantly affect the glycosylation profile of mAbs, with variables that could affect glycan profile and thus *in vivo* performance generally scoring high. Potential variables affecting glycosylation include, Dissolved Oxygen (DO), Temperature (Temp), Cell Density, pH Shift (pHS), Temperature Shift (TempS), Hydro Cortisone (hCort) etc. This exercise led to a total of 11 process parameters previously shown to have potential impact on the glycosylation profiles of mAbs and were chosen as input variables (Table I) for the Plackett-Burman DOE screening (Ghirlando et al., 1999).

2.2 Plackett-Burman Design of Experiments

All the process variables identified above as possibly having impact on quality were screened through Plackett-Burman DOE for their influence on the glycan profiles of the model mAbs produced in the culture. An "11 factor - 2 level - 12 run" Plackett-Burman screening study was performed using MODDE software (v11.0.1, Umetrics, San Jose, CA). The experimentation was conducted in 2 sets. These 12 runs were accompanied with 4 control runs (numbered 13-16)

including 2 controls in each experimentation set. Eleven input variables and 23 glycan profiles considered as CQAs for the study are listed in Table I. The structure of experimental design is described in Table II, with each input variable evaluated at two levels. DO, Sparge, Agitation, Temp and Density had specific high (+) and low (-) numerical values; pH_S and Temp_S varied according to their shift (+) or no-shift (NS)(-); bioreactors were fed either as Bolus (-) or Drip (+); and finally, Non-Essential Amino Acids (NEAA), Fatty Acids (FA) and hCort were either added (+) or not-added (-). Due to logistics, the whole experiment was run in two different phases, with each phase containing six test conditions and two control conditions run at the same time in a parallel bioreactor system.

All runs were performed in individually controlled parallel bioreactors, in a fed batch mode at the FDA/CDER/OBP (FDA White Oak Facility, Silver Spring, MD) using the same model murine IgG3:K hybridoma described in Read et al., 2013 and Agarabi et al., 2015. The seed culture was prepared in 2L spinner flasks and was then transferred to 7.5L Bioflo110 bioreactor (New Brunswick Scientific, Edison, NJ) or 1.2L DasGip parallel bioreactors (DasGip Biotools, Shrewsbury, MA) that contained either 4 or 0.8L of media, respectively. The 7.5L bioreactor was equipped with an Aber bio-capacitance probe connected to an Aber Biomass Monitor 200 (Aber Instruments, Aberystwyth, UK) to provide a measure of live cell bulk biomass and metabolic activity. Concentration of viable and non-viable cells was measured periodically using trypan blue staining and manual hemacytometer counts. The values of viability and total cell concentration were calculated from those manual counts. The DOE runs were inoculated at a low cell density of 5×10^4 cells/mL or at a high cell density of 1×10^5 cells/mL. All other culture conditions of the bioreactors were maintained according to conditions described in Table II.

2.3 Glycoform analysis

Glycans were liberated and analyzed as described in a previous work (Agarabi et al., 2015). Briefly, each antibody sample containing 0.5 mg IgG3 was taken and made up to 400 μ L with sodium phosphate buffer (50mmol/L, pH 7.4). Glycans were released by incubating with 10 μ L of PNGase F (New England Biolabs, Ipswich, MA) for 17 hours at 37°C after the dialysis against sodium phosphate (50mmol/L, pH 7.4) using Slide-A-Lyzer cartridges (Thermo Scientific, Pittsburgh, PA). Glycans were purified using GlycoPrep H cartridges (Prozyme, Hayward, CA) and the purified glycans were dried using speedvac (centrifugal evaporator). The glycan labeling was then completed with 2-AB by incubating them at 37°C for 17 hours in 10 μ L of acetic acid/DMSO solution containing 5 mg 2-AB/100 μ L and 6 mg NaCNBH₃/100 μ L. The samples were then diluted with 190 μ L of acetonitrile before purifying them with GlykoPrep 2-AB cleanup cartridges. Then 0.5 mg of digested glycoprotein from each sample was dissolved in acetonitrile (ACN:WATER::80:20). Glycan analysis was then performed by LC-MS/MS based assay. All the instrument settings and data analysis procedures were used as described previously (Schiel et al., 2012). The samples were analyzed with a glycan column (Waters, Waltham, MA) set-up with a binary gradient of 0.1% (v/v) formic acetic acid in water and 0.1% formic acid (v/v) in acetonitrile. The gradient was linear from 80% of former to 63% of later from 2 to 32 minutes followed by washing and equilibration.

2.4 Statistical Analysis and Design Space Exploration Strategy

The results of the IgG3 Plackett-Burman screening study were analyzed using MODDE software (v11.0.1, Umetrics, San Jose, CA), in an attempt to define functional relationships between the process variables and the CQAs (23 glycan profiles). Variables with significant

effect on the distribution of terminal galactose and mannose residues, and their associated p-values, were determined employing analysis of variance (ANOVA) and General Regression Model (GRM) through MODDE. Since GRM considers the influence of individual experimental inputs as well as interaction between inputs, it was utilized to predict the impact of the variable inputs on the CQAs, with the $p \leq 0.05$ cut-off established to define significant effects. To ensure solidity of the results of the DOE runs, comprehensive validity control, outlier detection and data interpretation steps were carried out for each of the design constraints. The relative strength of each factor influencing CQAs was assessed using a set of interpretations involving raw data analysis, factor influence, residual control and a summary fit with estimates of R^2 , Q^2 , Model Validity and Model Reproducibility (e.g. Figure 3B). Further, we conceived a robust “design space exploration strategy” with four key scenarios. The goal of the strategy is to aid cumulative understanding and interpretation of the data generated from inexpensive small scale DOE runs before venturing into more costly production scale studies. To demonstrate its utility, specification ranges for all the CQAs (glycan profiles) were hypothetically set. A cumulative summary fit of the model was created, summing up the estimates (R^2 , Q^2 , Validity and Reproducibility of the model) of all CQAs (Figure 2). Individual models for all the CQAs were carefully analyzed. Based upon CQA results and model performance, they were categorized into four key scenarios: (1) Inside specification/significant model; (2) Inside specification/Non-significant model; (3) Outside specification/significant model; (4) Outside specification/Non-significant model. Implementation of the conceived exploration strategy was used to logically analyze data with Edge of Failure, Set Point Analysis and other optimizations to finally create hypothetical DSps for each scenario with greater robustness and degrees of freedom to accommodate variations.

3. RESULTS AND DISCUSSION

The Plackett-Burman process DOE was originally conceptualized to screen for significant process variables with high efficiency and accuracy (Plackett and Burman 1946). The goal is a quick reduction in the number of risk factors with high impact on product quality to narrow down the list for further analysis. In the current model-set, total of 11 process parameters considered likely to have potential impact on the glycan profiles of mAbs were identified through risk analysis. These were chosen as input variables (Table I) for the Plackett-Burman DOE screening (Table II, Agarabi et al., 2015). The dataset from this DOE screening study (See Supplementary Data, Bhatia H et. al. 2016) was analyzed using MODDE software (v11.0.1, Umetrics, San Jose, CA). The relative strength of each process variable influencing CQAs (23 Glycan Profiles, Table I) was expressed using a set of interpretations involving raw data analysis, factor influence, residual control and a summary fit. The adequacy of the developed model was estimated by R^2 , Q^2 , Model Validity and Model Reproducibility values expressed in summary fit of each CQA. R^2 value is the “maximum squared regression coefficient that can be achieved by a model using only the variables in it,” which is an indication of how well the model fits the experimental data (Nagelkerke, 1991). R^2 value indicates the amount of variability in the data that has been captured by the model and the closer the value of R^2 is to 1, the more superior is the model. Q^2 (Model Predictability), Positive Q^2 preferably closer to 0.5 indicates better model predictability. Models with Q^2 value > 0.1 can be considered as a significant model (Bravo Linares et.al. 2013). Model Validity indicates the measurement of appropriateness of the model. Model validity >0.25 signifies the appropriate model, while <0.25 indicate the lack of fitness of the model. Model Reproducibility assesses the variation between the replicates. Reproducibility

values below 0.5 indicate the large error and poor control on the experimentation. (Bravo-Linares et al., 2013)

In our proposed approach, variables assessed to be statistically significant will further be used to create desirability plots in order to confirm the most appropriate variable to be considered for model confirmation and process refinement studies. The decision to transition into production at large scale should be based on a thorough and cumulative understanding and interpretation of the above described data. Further, in a multidimensional process DSp, numerous process variables interact as input parameters to create the best possible combination of process values in order to assure quality and minimize the chance of process and product failure. Since “Working within the DSp is not considered as a change for regulatory reporting purposes; however, the movement out of the DSp is considered a change,” (ICH, 2012) it is logical to have a DSp as wide as possible. This consequently imparts more robustness and degrees of freedom to accommodate variations in cases of process deviations, for example. Also, it minimizes regulatory post-approval reporting requirements, freeing manufacturing, regulatory and quality personnel for more critical tasks.

In order to aid interpretation of such a multitude of data produced by a DOE, we conceived a robust “DSp exploration strategy” describing four different scenarios (pictorially represented in Figure 1). The data from a Plackett-Burman DOE is used to demonstrate the design space building exercise. Once the Summary of Fit created by the model provided a cumulative review of estimates (R^2 , Q^2 , Validity and Reproducibility) of all CQAs (23 glycan profiles)(Figure 2) individual models for each CQA were carefully analyzed. Based upon hypothetically defined specification ranges, CQA estimates and model performances, the DSp exploration strategy was conceived as a function of four scenarios. The approach to explore

design space or the potential actions to be taken to improve the process has been explained in detail. The four possible cases are defined based on the result outcome from the experimental design screening data. The case type determination was based upon whether the CQA results were lying in the specification range specified for the process and model significance, which in turn was determined by model parameters (R^2 and Q^2).

3.1 Inside Specification/Significant model

This category is defined as “a scenario where (1) all the CQAs from all the experimental runs are in alignment with the specification ranges allowed for a particular product and (2) the model developed was superior” (R^2 , Q^2 values closer to 1) (Eriksson et. al. 2008). This describes the most favorable scenario, one that denotes a predictable and well-built model, which could capture a decent amount of variability and establish a functional correlation between the CPPs and the CQAs. In this case, the screened CPPs have a significant impact on the CQAs, preserving all of the antibody quality attributes within the specification range. Moreover, it also suggests that the ranges chosen for the CPP are appropriate as they are giving the CQA in a specified range.

Partial least square models for all the CQAs were generated and analyzed carefully to categorize the models into one of the above-mentioned scenarios. One of our models derived from the bioreactor dataset, CQA-G2 (Figure 3), was observed to be significant, with an $R^2=0.9$, $Q^2=0.43$, validity=0.65, reproducibility=0.7 (Figure 3B). All process replicates were well within the predefined hypothetical specification ranges (represented by the red horizontal bars in Figure 3A), making it an ideal candidate for this scenario of DSp exploration. Typically, such cases, where all the experimental results strongly indicate the model developed to be genuinely

predictive and significant, eliminate the need for any further evaluations like evaluation of the edge of the failure. A DSp can easily be established for models that meet this scenario even without set point analysis or major optimizations. This, in turn, approves the use of screened input variables for further process validation. While it is possible to accept this model with no further alterations, given that a strong correlation has already been established between the input parameters and CQAs, our study suggests that further optimization may be warranted. For example, the primary DSp created for CQA-G2 (Figure 3E) has an estimated chance of failure as high as 50%. In theory, a desirable chance of failure is expected to be lower, insisting caution in interpretation of such data to delve into further conclusions. Such error can be attributed to the inherent limitation in quantity and quality of the modeling data from the bioreactors, which is indicated by wide confidence intervals in the co-efficient plot (Figure 3C). We propose a Response Surface Methodology (RSM) to be a superior option to gather more reliable process information in this context (Myers et al., 2009). Set point optimization feature was used to perform the CPP optimization using MODDE. The important input parameters having significant impact on CQA's were chosen from the co-efficient plot (figure 3C) for the optimization. As a result of which, optimization of two CPPs, DO and Temperature, significantly reduced the chance of failure of this model to about 10%, indicating the requirement and possible impact of such optimizations for this scenario.

3.2 Inside Specification/Non- Significant Model

This scenario is defined as “where all the CQAs from all the experimental runs are in alignment with the specification ranges allowed for a particular product, while the model developed is insignificant” (Eriksson et. al. 2008). The results for CPP inside the specification

range suggests that this scenario recognizes an impact of CPPs on quality of the product, but the insignificant model ($Q^2 = 0.05$), the predictability of the model isn't verified. This implies the existence of underlying issues and a need for improvement in model robustness.

Of all the models generated, the individual model, CQA-G2FaGbGN (Figure 4), was observed to be only a moderately good fit, with $R^2=0.65$, $Q^2=0.05$, validity=0.27, reproducibility=0.59 (Figure 4B). The process replicates well within the predefined specification ranges (represented by the red horizontal bars in Figure 4A), making it an ideal example for this scenario of DSp exploration. The CQA estimates from all the experimental runs were within the specification range, thus this model needed no further validation for edge of the failure. However, based on the outputs, the relatively poor fit of the model can be attributed to the following reasons. (1) The clustered replicates (replicates 13,15 and 14,16) in Figure 4A indicated the poor reproducibility of the model. The DOE runs performed in two different groups could have caused clustered replicates due to inherent lot-to-lot experimental variations, diminishing the robustness of the bioreactor data from the screening study. (2) Also, a Plackett-Burman screening study has an inherent disadvantage over other DoE designs wherein the interactions between variables are not measured in this approach. As there are few degrees of freedom in this DoE design, the poor fit of this particular model may result. On the other hand, a DSp and its optimizations can still be performed in a similar way as in the previous, better fit model scenario. The optimization of the key process variables in the DSp for this scenario, initially created based on the model, significantly reduced the earlier estimate of 100% risk of failure (Figure 4E) to 5% (Figure 4B). As mentioned above, the accuracy of such DSp building and optimization can further be improved through RSM (Myers et al., 2009).

3.3 Outside Specification/Significant Model

Defined as “a scenario where some values of CQAs from few experimental runs are outside the defined specification range for a particular product but the model developed is superior in entirety,” (Eriksson et. al. 2008) this indicates that the model built could capture a decent amount of variability ($R^2 = 0.9$) and could predict the correlation between the CPPs and CQAs fairly well ($Q^2 = 0.4$), making it a favorable scenario. However, presence of a few values outside the specification range insists on further improvements in the process, as the significant model suggests the strong correlation between the input parameters chosen and the product quality. So by optimizing the ranges of input parameters, the product quality values may be brought into the specification range. Those optimized input parameter ranges can be identified by design space.

The individual model, CQA-G0 (Figure 5), was observed to be significant with $R^2 = 0.9$, $Q^2 = 0.4$, Validity = 0.8, Reproducibility = 0.57 (Figure 5B) and most process replicates within the predefined specification range (represented by the red horizontal bars in Figure 5A), making it an ideal candidate for this scenario of DSp exploration. Even though most CPPs screened through the DOE were shown to have significant impact on the CQAs, deviation of a few key estimates outside the specification range reduces its overall utility. Hence, a DSp established using the optimum strategy for DSp building, should bring the CQAs inside the specification range through manipulation of the CPPs that were deemed most important for the product CQAs during risk analysis. To achieve this, a set-point validation was performed with MODDE (v11.0.1), which predicts the success rate (Predicted Response Profile) of a model based on the CPP ranges (Estimated Acceptable Range) used for the DOE. The DSp built after set point validation and other optimizations reduced the overall risk of failure from the initial estimate of

100% (Figure 5E) to 5% (Figure 5F), which is still a high risk of failure but can be attributed to the intrinsic limitation of quality and quantity of data gathered through a screening study (explained above).

3.4 Outside Specification/Non-Significant Model

This model type would be defined by “a scenario where some values of CQAs from several experimental runs are outside the defined specification range for a particular product and the model developed is also insignificant.” (Eriksson et. al. 2008) This model could not capture either variability, or could not predict the correlation between the CPPs and CQAs with certainty; this scenario is completely undesirable. Such scenarios raise various concerns, starting from CPPs chosen for the experimentation as the insignificant model suggests poor correlation between the input parameters and the results, which in turn questions the methods employed for preliminary risk analysis, screening studies performed based on such ambiguously chosen CPPs, and even the variability of CPPs considered for the study. Under such circumstances, identifying and solving the uncertainty of a specific step and/or a specific CPP in focus, should precede DSp exploration.

The individual model, CQA-G2FaG2 (Figure 6), was observed to be insignificant, with $R^2 = 0.58$, $Q = -0.05$, Validity = 0.5, Reproducibility = 0.1 (Figure 6B) and most process replicates outside the predefined specification range (represented by the RED horizontal bars in Figure 6A), making it an example of this scenario of DSp exploration. The replicative index clearly demonstrates the scenario, with only 6 out of 16 runs (runs 1, 6, 7, 11, 12, 16) aligning within the acceptable range and poor reproducibility (runs 13, 14 and 15, 16) (Figure 6A) indicating the insignificant model. The summary of fit further illustrates the poor predictability

and reproducibility of the model, confirming weak correlation between the CPPs and CQA-G2FaG2 (Figure 6B). Overall, it is not viable to build a DSp for such a scenario indicating a strong need for caution in selection and experimentation of CPPs, to start over and recreate a better model that can accurately and reproducibly predict the functional relationship between CPPs and CQAs. This could be attributed to either inappropriate factors chosen for screening, lesser variability during experimentation or experimental error (as explained in Fig.1).

4. CONCLUSIONS

The study successfully integrates and validates a unique DSp exploration strategy with four different scenarios. Enhancing the QbD approach one step further, this strategy proves to increase the accuracy of data interpretation, and makes both design and handling of large number of qualitative and quantitative modalities a logical transition. Every step of data interpretation can be guided in more detailed, logical ways; and the complexity of DSp exploration can be dramatically minimized. In this study, data from a Plackett-Burman screening design, having bioreactor design input parameters affecting the quality of the product, was categorized into four scenarios to guide the design space building. In the first 2 cases, the design space building is fairly straight forward without the need to do further validation for the edge of failure. In the case similar to case 3, it is most important to do the CPP manipulation to get the CQA's in the specified range. Data similar to case 4 from experimental run suggests further improvements with the experimentation. The causes and the approaches to resolve these kinds of issues have been explained in the results and discussion section. Implementation and further improvement of this approach will provide a unique control over a DSp to understand and validate experiments and make sound process-related decisions. Understanding of functional relationships and effect

of manufacturing parameters on product CQAs, prediction and validation of their ranges and in totality multidimensional comprehension of products and processes defines the fundamentals of a QbD approach. It is the DSp that provides the optimal region for such a systematic evaluation, mechanistic understanding and refining of the QbD approach. In this context, the DSp exploration strategy will aid the process of consistently and reproducibly achieving predefined quality of a product and facilitates continual improvement and innovation throughout a product lifecycle.

5. ACKNOWLEDGEMENTS

The authors gratefully acknowledge scholarship support for this research from the Biopharmaceutical Product and Quality Consortium, University of Massachusetts Lowell. The bioreactor runs were funded by FDA/CDER Critical Path Project 1500. The authors thank Pfizer for the partial funding of this research. This project was also supported, in part, by an appointment to the Research Participation Program at the Center for Drug Evaluation and Research administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

6. References

- Adam S, Suzzi D, Radeke C, Khinast JG. 2011. An integrated Quality by Design (QbD) approach towards design space definition of a blending unit operation by Discrete Element Method (DEM) simulation. *Eur. J. Pharm. Sci.* **42**:106–15.
<http://www.sciencedirect.com/science/article/pii/S0928098710003581>.
- Agarabi CD, Schiel JE, Lute SC, Chavez BK, Boyne MT, Brorson KA, Khan MA, Read EK. 2015. Bioreactor Process Parameter Screening Utilizing a Plackett-Burman Design for a Model Monoclonal Antibody. *J. Pharm. Sci.* **104**:1919–1928.

433 <http://www.ncbi.nlm.nih.gov/pubmed/25762022>.

434 Bhatia H, Read EK, Agarabi CD, Brorson KA, Lute SC, Yoon S 2016 Glycan Profile data at
 435 variable bioreactor engineering parameters for hybridoma cell line Intl J Pharm “Submitted”.

436

437 Bravo-Linares C, Ovando-Fuentealba L, Mudge SM, Loyola-Sepulveda R. 2013. Application of
 438 response surface methodology to oil spill remediation. *Fuel* **103**:876–883.

439 Department of Health and Human Services U.S Food and Drug Administration 2004
 440 Pharmaceutical cGMPs for the 21ST Century — A Risk-Based Approach
 441 (<http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/Manufacturing/QuestionsandAnswers/CurrentGoodManufacturingPracticescGMPforDrugs/UCM176374.pdf>)
 442
 443

444 Department of Health and Human Services, Food and Drug Administration 2004 PAT Guidance
 445 for Industry – A Framework for Innovative Pharmaceutical Development, Manufacturing
 446 and Quality Assurance

447 Dowling W, Thompson E, Badger C, Mellquist JL, Garrison AR, Smith JM, Paragas J, Hogan
 448 RJ, Schmaljohn C. 2007. Influences of glycosylation on antigenicity, immunogenicity, and
 449 protective efficacy of ebola virus GP DNA vaccines. *J. Virol.* **81**:1821–37.
 450 <http://jvi.asm.org/content/81/4/1821.abstract>.

451 Eriksson L, Johansson E, Keetaneh-Wold N, Wikstrom C and Wold S Design of Experiments
 452 Principles and Applications Umetrics Academy p177-180.

453 Ghirlando R, Lund J, Goodall M, Jefferis R. 1999. Glycosylation of human IgG-Fc: Influences
 454 on structure revealed by differential scanning micro-calorimetry. In: . *Immunol. Lett.*, Vol.
 455 68, pp. 47–52.

456 Hinz DC. 2006. Process analytical technologies in the pharmaceutical industry: The FDA’s PAT
 457 initiative. *Anal. Bioanal. Chem.*

458 Hossler P, Khattak SF, Li ZJ. 2009. Optimal and consistent protein glycosylation in mammalian
 459 cell culture. *Glycobiology* **19**:936–949.

460 International Conference of Harmonisation (2008) ICH Harmonised Tripartite Guideline:
 461 Q8(R1) Pharmaceutical Development ([http:// www.ich.org/LOB/media/MEDIA4986.pdf](http://www.ich.org/LOB/media/MEDIA4986.pdf)) 4

462 International Conference of Harmonisation (2005) ICH Harmonised Tripartite Guideline: Q9
 463 Quality Risk Management ([http:// www.ich.org/LOB/media/MEDIA1957.pdf](http://www.ich.org/LOB/media/MEDIA1957.pdf)) 5

464 International Conference of Harmonisation (2008) ICH Harmonised Tripartite Guideline: Q10
 465 Pharmaceutical Quality Systems ([http:// www.ich.org/LOB/media/MEDIA3917.pdf](http://www.ich.org/LOB/media/MEDIA3917.pdf)) 6

466 International Conference of Harmonisation (2012) ICH Harmonised Tripartite Guideline: Q11
 467 Development and Manufacture of Drug Substances

468 http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q11/Q11_Ste
469 [p_4.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q11/Q11_Ste)

470 Li X, Xu T, Ma X, Guo K, Kai L, Zhao Y, Jia X, Ma Y. 2008. Optimization of culture conditions
471 for production of cis-epoxysuccinic acid hydrolase using response surface methodology.
472 *Bioresour. Technol.* **99**:5391–6.

473 Lionberger RA, Lee SL, Lee L, Raw A, Yu LX. 2008. Quality by design: concepts for ANDAs.
474 *AAPS J.* **10**:268–76.

475 Liu H, Wang S, Jia Y, Li J, Huang Z, Lu S, Xing Y. 2015. N-Linked Glycosylation at an
476 Appropriate Position in the Pre-S2 Domain Is Critical for Cellular and Humoral Immunity
477 against Middle HBV Surface Antigen. *Tohoku J. Exp. Med.* **236**:131–138.

478 Liu H, Bulseco GG, Sun J. 2006. Effect of posttranslational modifications on the thermal
479 stability of a recombinant monoclonal antibody. *Immunol. Lett.* **106**:144–153.

480 Myers RH, Montgomery DC, Anderson-Cook C. 2009. Response Surface Methodology: Process
481 and Product Optimization Using Designed Experiments . *Wiley Ser. Probab. Stat.* 704 p.

482 Nagelkerke NJD. 1991. A note on a general definition of the coefficient of determination.
483 *Biometrika*.78:691-92.

484 Plackett, R. L. and Burman, J. P. 1946. The design of optimum multifactorial experiments.
485 *Biometrika* 33:305-325.

486

487 Raju TS, Scallan B. 2007. Fc Glycans Terminated with N-Acetylglucosamine Residues Increase
488 Antibody Resistance to Papain. *Biotechnol. Prog.* **23**:964–971.

489 Rathore AS, Winkle H. 2009. Quality by design for biopharmaceuticals **27**:26-34.

490 Read EK, Bradley SA, Smitka TA, Agarabi SC, Lute SC, Brorson KA 2013. Fermentanomics
491 informed amino acid supplementation of an antibody producing mammalian cell culture.
492 *Biotechnol. Prog.* 29(3): 749-53.

493

494 Sangshetti JN, Deshpande M, Zaheer Z, Shinde DB, Arote R. 2014. Quality by design approach:
495 Regulatory need. *Arab. J. Chem.*

496 Schiel JE, Au J, He HJ, Phinney KW. 2012. LC-MS/MS biopharmaceutical glycoanalysis:
497 Identification of desirable reference material characteristics. *Anal. Bioanal. Chem.*
498 **403**:2279–2289.

499 Sellberg A. 2013. Design Space in pharmaceutical industry.

500 Wu H, White M, Khan MA. 2011. Quality-by-Design (QbD): An integrated process analytical
501 technology (PAT) approach for a dynamic pharmaceutical co-precipitation process
502 characterization and process design space development. *Int. J. Pharm.* **405**:63–78.

503 Yu LX, Amidon G, Khan M a, Hoag SW, Polli J, Raju GK, Woodcock J. 2014. Understanding
504 pharmaceutical quality by design. *AAPS J.* **16**:771–83.

505 Yu LX. 2008. Pharmaceutical quality by design: Product and process development,
506 understanding, and control. *Pharm. Res.* **25**:781–791.

507 Zheng K, Bantog C, Bayer R 2011. The impact of glycosylation on monoclonal antibody
508 conformation and stability. *MAbs* **3**:568–76.

509

510

Table I: Critical Process Parameters (CPPs) selected through preliminary risk analysis and Critical Quality Attributes (CQAs) considered for Plackett-Burman Design of Experiments (DOE)

Input Variables (X)		
Process Variables	Low (-)	High (+)
Dissolved Oxygen (DO) (%)	20	40
Sparging (SU)	0.3	0.7
Agitation (rpm)	90	170
Temperature (°C)	35.5	37
Cell Density (Cells/mL)	50000 (5×10^4)	100000 (1×10^5)
pH Shift (pHS)	No-Shift (NS)	Shift (Sh)
Temperature Shift (TempS)	No-Shift (NS)	Shift (Sh)
Feed	Bolus	Drip
Non-Essential Amino Acids (NEAA)	No (N)	Yes (Y)
Fatty Acids (FA)	No (N)	Yes (Y)
Hydro Cortisone (hCort)	No (N)	Yes (Y)
Output Variables (Y)		
Glycan Nomenclature	Description	
M3, M4, M5, M6, M7	Glycans with varying number of mannose units	
G0, G0F, G0GN, G0F-GN	Glycans with no Gal unit	
G1F3, G1F6, G1FS, G13, G16	Glycans with one Gal unit	
G2, G2F, G2FAG, G2FAGBGN, G2FSA, G2FG2, G2FSG, G2FAG2, G2FA2GSG	Glycans with 2 Gal units	

522

523

524

Table II: Experimental design for Plackett-Burman screening study

High and low factor values of manipulated variables in “11 factor - 2 level - 12 run” Design

of Experiments

PB No	Pattern	DO (%)	Sparge (SU)	Agitation (rpm)	Temp (°C)	Density (cells/ml)	pHS	TempS	Feed	NEAA	FA	hCor
1	--+---+-----	20	0.3	170	35.5	50000	Sh	NS	Drip	Y	Y	N
2	+---+-----	40	0.3	90	37	50000	Sh	Sh	Drip	N	N	N
3	+++---+---+	40	0.7	170	35.5	50000	NS	Sh	Bolus	N	Y	N
4	+++++---+--	40	0.3	170	37	100000	NS	NS	Bolus	Y	N	N
5	+-----+---+	40	0.3	90	35.5	100000	NS	NS	Drip	N	Y	Y
6	+++++++-----	40	0.7	170	37	100000	Sh	Sh	Drip	Y	Y	Y
7	--+-----+	20	0.3	170	35.5	100000	Sh	Sh	Bolus	N	N	Y
8	-+-----+--	20	0.7	90	37	100000	Sh	NS	Bolus	N	Y	N
9	-+-----+---	20	0.7	90	35.5	100000	NS	Sh	Drip	Y	N	N
10	--+---+-----	40	0.7	90	35.5	50000	Sh	NS	Bolus	Y	N	Y
11	--+---+-----	20	0.3	90	37	50000	NS	Sh	Bolus	Y	Y	Y
12	--+---+-----	20	0.7	170	37	50000	NS	NS	Drip	N	N	Y

525

526

Figure 1. Pictorial representation of the conceived “DSp Exploration Strategy” defined by 4 key scenarios.

Figure 2. Summary of Fit created by MODDE Software (v11.0.1, Umetrics, San Jose, CA). Represents a cumulative illustration of R^2 , Q^2 , Model Validity and Model Reproducibility estimates of all CQAs (glycan profiles). N = Number of process runs, DF = Degrees of Freedom, R^2 = Indication of fit of the model, Description of the glycan nomenclature = Refer Table I. Missing in red indicates the validity for 2 CQA’s (M3 and M4) was not generated by the model as many data points for these 2 CQA’s were not available.

Figure 3. Inside Specification / Significant Model - CQA-G2; Fig. 3(A) Replicate index illustrating all process replicates well within the predefined hypothetical specification range, represented by the RED horizontal bars, the blue boxes numbered 13, 14, 15 and 16 are the control conditions; Fig. 3(B) Summary of fit for CQA-G2 illustrating R^2 , Q^2 , Model Validity and Model Reproducibility estimates; Fig. 3(C) Coefficients (scaled and centered) for CQA-G2; Fig. 3(D) Residuals Normal Probability for CQA-G2; Fig. 3(E) Design of space CQA-G2 – pre-optimization with estimated 50% risk of failure; Fig. 3(F) Design of space CQA-G2 – post-optimization with estimated 10% risk of failure.

Figure 4. Inside Specification / Non-Significant Model - CQA-G2FaGbGN; Fig. 4(A) Replicate index illustrating all process replicates well within the predefined hypothetical specification range, represented by the RED horizontal bars; Fig. 4(B) Summary of fit for CQA-G2FaGbGN illustrating R^2 , Q^2 , Model Validity and Model Reproducibility estimates; Fig. 4(C) Coefficients (scaled and centered) for CQA-G2FaGbGN; Fig. 4(D) Residuals Normal Probability for CQA-G2FaGbGN; Fig. 4(E) Design of space CQA-G2FaGbGN – pre-optimization with estimated 100% risk of failure; Fig. 4(F) Design of space CQA-G2FaGbGN – post-optimization with estimated 5% risk of failure.

Figure 5. Outside Specification / Significant Model - CQA-G0; Fig. 5(A) Replicate index illustrating most process replicates within the predefined hypothetical specification range, represented by the RED horizontal bars; Fig. 5(B) Summary of fit for CQA-G0 illustrating R^2 , Q^2 , Model Validity and Model Reproducibility estimates; Fig. 5(C) Coefficients (scaled and centered) for CQA-G0; Fig. 5(D) Residuals Normal Probability for CQA-G0; Fig. 5(E) Design of space CQA-G0 pre-optimization with estimated 100% risk of failure; Fig. 5(F) Design of space CQA-G0 post-optimization with estimated 5% risk of failure.

Figure 6. Outside Specification / Non-Significant Model - CQA-G2FaG2; Fig. 6(A) Replicate index illustrating most process replicates outside the predefined hypothetical specification range, represented by the RED horizontal bars; Fig. 6(B) Summary of fit for CQA-G2FaG2 illustrating R^2 , Q^2 , Model Validity and Model Reproducibility estimates; Fig. 6(C) Coefficients (scaled and centered) for CQA- G2FaG2; Fig. 6(D) Residuals Normal Probability for CQA-G2FaG2;

Figures:

Figure1:

589 Figure2:

590
591
592 Figure 3(ABCD):

593
594
595 Figure3(E):

Figure 3(F)

605 Figure4:

606
607
608 Figure 4(E)
609

614 Figure4(F)

615
616
617
618 Figure5(ABCD)

619
620 Figure 5(E)
621
622

623

624
625
626
627
628
629

Figure5(F):

630
631

Figure6(ABCD)

1. Experimental setup and data

- Temperature
- Mixing
- Aeration
- pH
- Nutrients
- Harvest

Batch	Input Variables-CPP			Results-CQA		
	CPP1	CPP2	CPP11	CQA1	CQA2	CQA11
1	Low	Low	Low	0.22	0.24	0.35
2	High	Low	Low	0.25	0.34	0.32
3	Low	High	Low	1.12	0.54	0.34
4	High	High	Low	0.78	0.64	0.82
5	Low	Low	High	0.64	0.72	0.24
8	High	High	High	0.34	1.23	0.78

3. Classification

A Guide to Design Space Exploration

2. Regression Analysis

4. Design space exploration