Improvement of Predictive Model to Reduce Impurity Presence Based on Raw Material Combination Pre-Selection

William Loubriel Nieves
Master of Engineering in Manufacturing Engineering
Advisor: Rafael A. Nieves Castro, PharmD.
Industrial and Systems Engineering Department
Polytechnic University of Puerto Rico

Abstract *— Biotechnology manufacture is composed of two main sections, Active Pharmaceutical Ingredient (API) and Drug Product (DP). The DS manufacturing most common steps are cell culture or fermentation (DSI), recovery and purification. The purification process purpose is to reduce impurities to acceptable levels. Depending on the purification raw materials, there will be different interactions with the DSI to be purified that could affect clearance capability. Multivariate analysis includes the effect of the considered variables and the response of interest. Therefore, it can be used to take into consideration combinations DSI solutions and purification raw materials available in inventory to predict the results of the impurity of interest. In this project, the improvement of a multivariate predictive model was performed by gathering recent manufacturing scale data and defining a more representative data set to improve the accuracy of the predictions. Upon implementation of the updated predictive model, several consecutive lots resulted in impurity results below the acceptable criteria.*

Key Terms *— Bioprocessing, Biotechnology, Raw Materials, Prediction, Predictive Model, Purification*

INTRODUCTION

Pharmaceutical Manufacture is a diverse but widely studied topic. There are many types of pharmaceutical products as oral, parenteral and semi soli formulations. Some of the therapy delivery technologies that have been in constant development through the recent years have been the once produce by means of biotechnology manufacture.

Biotechnology manufacturing is composed of two main sections, Active Pharmaceutical Ingredient (API) and drug product (DP) as shown in Figure 1. There are several steps for bioprocessing of a product, but the most common steps are fermentation/cell culture, recovery, purification, fill and finish. The fermentation or cell culture is the process of cultivating cells or bacteria so they can reproduce as shown in Figure 2 generating a Drug Substance Intermediate (DSI). During fermentation or cell culture, the protein of interest is generated. Impurities are also present in the DSI solution. They can be generated by the host cell, come from raw material traces, cell debris, within other.

The purification process purpose is to eliminate or reduce these impurities to acceptable levels for patients as well as isolating the protein of interest before moving the product to the Drug Product stage. The purification process consist of multiple orthogonal steps such as filtration, ultrafiltration/diafiltration, viral inactivation, chromatography purification, viral filtration, and bulk filtration as shown in Figure 2. Each of these steps are important for the purification as well as the raw materials used for each of them.

Depending on the raw materials used during the manufacturing process, different interactions with the fermentation or cell culture product to be purified can occur. Also, the fermentation or cell culture product (DSI) characteristics can have variation within acceptable ranges as well as the raw materials can have a lot-to-lot variation within acceptable ranges when manufactured by the vendor. Therefore, within the expected acceptable standards, there can be a different outcome of a purification step given the DSI and the purification raw materials characteristics variability. Taking into consideration that impurities should be lowered to a certain level at the end of the purification steps, it is ideal to increase process understanding by studying potential interactions of raw materials and which combination

is appropriate to achieve the desired impurity reduction.

To determine the optimal combination of these materials, multivariate analysis can be used. The multivariate analysis can take into consideration raw materials available to predict impurity results. Therefore, a preselection can be made, based on the predicted impurity results, of the best raw materials combination. In this way, the combinations with acceptable impurity results can be selected, thus reducing the possibility of having elevated results.

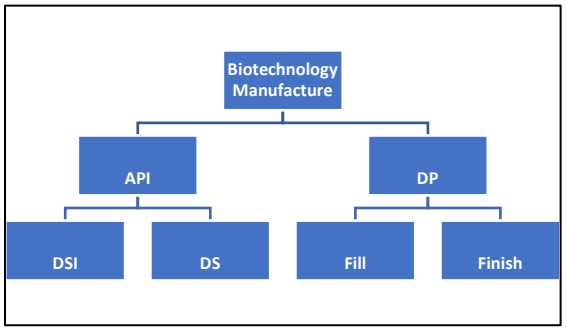

Figure 1
Biotechnology Manufacture Structure

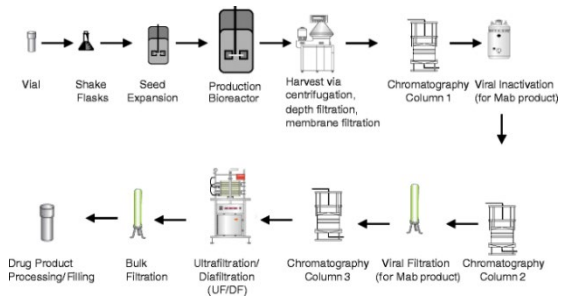

Figure 2
Biotechnology Drug Substance Manufacture Process [1]

LITERATURE REVISION

Biotechnology manufacturing is composed of two main sections, Drug Substance (DS) and Drug Product (DP). There are several steps within the Drug Substance for bioprocessing of a product. The most common DS steps are fermentation/cell culture, recovery and purification. During cell culture or fermentation, the protein of interest is generated, but this generation is not exclusive. Impurities like nucleic acid, cell membrane fragments, and host cell proteins (HCP) within others are also generated during cell culture [2, page 3] for example. Traces of impurities from raw materials can also be present. The purification's process purpose is to eliminate or reduce these impurities to acceptable levels for patients while avoiding the loss of product. The purification process can consist of filtrations, ultrafiltration/diafiltration, Viral inactivation, chromatography purification, viral filtration, and sterilizing filtration [2, page 9]. Filtration steps are mostly centered in size exclusion application. Chromatography purification consists of passing the cell culture product through cylinders filled with resin. There are different types of chromatography purification based on the resin and the characteristics of the solution to be purified. Some of the chromatography purification types are Affinity chromatography, Ion-Exchange chromatography, hydrophobic interaction chromatography, mixed-mode chromatography, among others. [2, page 405]. The efficacy of these chromatography operations can have variations and are very dependable on column packing [2, page 353]. Also, the raw materials that are used during purification and vendor manufacture consistency or variations can have an impact on final product and impurity amount. Comparable to a vendor raw material, the fermentation or cell culture product characteristics can also have variations within acceptable ranges.

The cell culture process can generate solutions with less or more product and less or more impurities [3, page 1], which later will have an impact on the outcome of the chromatography steps. Therefore, there can be different outcomes of a chromatography step given the introduced DSI product and the purification raw materials variability. Taking into consideration that impurities should be kept within a certain level at the end of a chromatography step to support product quality claim, it can be beneficial to understand which combination of DSI solution and raw materials is appropriate to achieve the desired impurity reduction or level outcome and avoid deviations. Simple statistical analysis or correlations cannot be applied to this kind of analysis since multiple variables from raw material and DSI product should be taken into consideration.

In the last century, statistical analysis has expanded from single variables and outcomes to multivariable data analysis (MVDA) and predictions. The multivariate analysis includes the effect of the considered variables and the response of interest [4, page 1]. Prediction of a response base on multivariable analysis has multiple applications. For example, it has been applied for consumer behavior prediction [5], wheat yield prediction [6], and the cell culture titer based on process variables behaviors [7]. Multivariable analysis and predictions can be performed in common software such as Minitab® [8] and JMP [9], among others.

With the application of MVDA software, a predictive approach like the examples above can be applied to a biotechnology process. The scope of this work is to use historical combinations of DSI solutions and raw materials to improve an impurity result predictive model. The impurity results are meant to be kept within the acceptance criteria. Therefore, a pre-selection can be made of the best raw material and DSI combination, based on the model predicted impurity results. In this way, the combination with lower impurity prediction result is selected, thus reducing the possibility of having a elevated result.

PROBLEM STATEMENT

While multivariate analysis can be beneficial for data processing and recommendation generation, developing an appropriate analysis can be difficult. As with any other data analysis, the resulting analysis will be dependent on the data set. Obtaining data to be used as foundation for the multivariate analysis can be a challenge when dealing with a large-scale manufacture. Commercial scale data can be limited since it will depend on the lots already manufactured and manufacturing schedules for subsequent lots. Therefore, data sets can be supported with laboratory data if the small-scale model is qualified to be representative of the large scale. Laboratory or small-scale data generation is easier than manufacturing scale since it does not depend on manufacturing schedules but rather on laboratory coordination. Therefore, some analysis initially utilizes laboratory scale data combined with the limited manufacturing scale data as a baseline for the data set and therefore, analysis or predictions.

The use of laboratory scale data can generate some variability between the predicted values and the manufacturing results since the small-scale data, in some cases, is not completely representative of the large-scale process behavior. The column packing, pool holds, human contribution (limited laboratory associates vs many manufacturing associates involved) can be some of the contributors to their difference. Also, large scale process improvements can result in changes that can drift the process behavior from the behavior contemplated in the analysis data set, thus creating differences between predictions and actual outcomes. In this project, the improvement of a predictive model will be performed by gathering recent manufacturing scale data and defining a more representative data set that can generate more accurate predictions. The predictive model will be improved to increase the probability of lower impurity level based on raw materials pre-use evaluation and combinations selection.

RESEARCH OBJECTIVES

As stated before, impurities are also generated along with the protein of interest in a fermentation or cell culture process. The purification processes' purpose is to eliminate or reduce these impurities to acceptable levels for patients to comply with product specification. As part of the product development, an impurity level limit is established based on clearance capability, potential impact to safety, efficacy etc. In this way impurities reduction is monitored and compliance with release specifications is ensured before distributing the product to patients.

There can be a different outcome of a purification step given the DSI and the raw materials characteristics variability. This research objective is that DSI product and raw material combinations characteristics can process through an improved predictive model, as part of a pre-use evaluation, to

define the adequate combinations to achieve lower impurity results.

RESEARCH CONTRIBUTIONS

There can be different consequences where impurity limit is exceeded. It depends on the importance of the impurity, the amount, how the product was developed, impact to safety, efficacy within other considerations. Also, not complying with the established limits will trigger what the manufacturing industry calls deviation or investigation.

The company quality system is used to investigate from what we are deviating, defining the problem, identifying the possible contributing factors, root cause, established product impact etcetera. Deviations management and completion can be very time-consuming, labor intensive, and, depending on the deviation, it can require support from most of the companies' functional areas. Also, a deviation consequence can be a delay of the product delivery or prevent the product from been distributed and consequently, unavailability of material for supply to the patients. Therefore, complying with impurity criteria based on correct material combination selection will contribute to reduce labor intensive activities and ensuring product supply for patients.

METHODOLOGY

Multivariate analysis can be of great benefit when solving problems. Predictions based on multivariate analysis can help to solve or avoid those problems. These prediction models based on multi variate analysis may need to be updated over time to capture new process data and to keep it current with changes or process improvements. In this project, the improvement of a predictive model will be performed by gathering recent manufacturing scale data and defining a more representative data set that can generate more accurate predictions. There are some steps that need to be performed to complete this project.

First, it is important to consider that a prediction model is as good as its foundation or the information and the amount of information that is considered for its construction. Therefore, the first step for this project will be processing data gathering. The data will be obtained from small scale runs, historical manufacturing runs, DSI intermediate characteristics and purification raw materials' certificates of analysis. The raw data cannot be directly introduced to the model. The collected data from small scale and historical manufacturing lots, needs to be processed. The data will be arranged and acclimated to the model structure so it can be used as input. Once the data is processed, it can be used to determine the best dataset combination.

To determine the best data set combination for model update and improvement, several predictive model iterations will be generated using different datasets, this using a data analysis application. Then, each predictive model generated with different data sets will be evaluated independently. The first criterion to be taken into consideration would be the permutation test result. This test will determine if the individual data set provide an unique solution when compared to 100 random combinations of the same data set. Models that do not pass this test will not be considered as an option since failing to pass this test implies that the solution provided by the data set is not unique when compared to 100 iterations of the data set.

After the models that do not comply with the permutation test are removed from the analysis, the remaining models will be evaluated considering other parameters like data linearity (R^2) and predictability (Q^2). Most importantly, the remaining models will be tested by comparing their predictions of already manufactured lots against their actual results. The manufactured lots' results that will be used for the test will have results on the low, middle and high side of the historical results. This will help to define accuracy of the remaining models across the spectrum of the results. The model that has the most realistic and representative data set, complies with permutation, has the best combination of R^2 and

Q^2 and predictions will be selected as the updated model.

The selected model will be used to replace and update the current model. The updated model will be introduced to the company official quality systems. Once the model is formally updated, it will be used as part of the readiness activities. To confirm its performance is acceptable, pre-use evaluation predictions will be compared against the actual lot's results. Also, it will be confirmed that the actual results comply with the impurities acceptance criteria.

RESULTS DISCUSSION

As defined in the methodology section, the historical data was gathered for manufactured lots from year 20XX to present. After data was gathered, it was properly processed to make it fit for use inside the model. The original data set was composed of xx lots/runs between small scale and full-scale manufacture data. For the model update, xx historical lots were considered. From all xx lots a total of 10 potential predictive models were generated using data analysis program and applying multivariate analysis. Each model was constructed upon different combinations of the available xx lots dataset. The combinations were based on time range, exclusion of less representative lots, inclusion of recently manufactured lots, within other reasons. The generated model options were individually analyzed considering different criteria.

The permutation test was used as the first tool to determine the acceptance or rejection of the proposed models. This test demonstrates that there is only one acceptable outcome or solution even if new data sets are created combining scrambled data from the original data set of the specific model option. From the total of 10 potential predictive models, the number of options was reduced to 7. For reference, Figure 3 and Figure 4 show unacceptance and accepted permutation tests examples. The plots show the R^2 and Q^2 resulting values of the original model on the right and the permutated models on the left. It is shown that similar R^2 and Q^2 between the original and permutated model were obtained for the discarded model (Figure 3). On the other hand, the non-discarded permutations do not show similar R^2 and Q^2 to the original model. All the permutations results are lower than the original model, indicating that it is a unique solution.

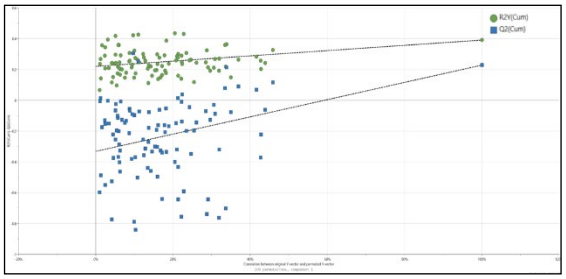

Figure 3
Example of Discarded Model Permutation Test Result

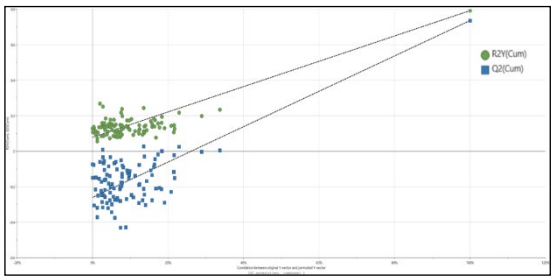

Figure 4
Example of Non-Discarded Model Permutation Test Result

Once the options were narrowed down, other model's characteristics were considered like R^2 and Q^2 which provide information about the predictability and the data fit of each individual model. Also, the remaining models were used to predict historical values (known results), to test how good were their predictions in addition to their Q^2 and R^2. All these factors were summarized and discussed with team subject matter experts (SMEs) to define the best approach for a final model selection.

Although all the previously mentioned factors were considered, the selection of the appropriate model was based on two major factors; model's data set was a better representation of the current manufacturing process, and the model showed the best predictions on the side that could trigger nonacceptable results. Having good predictions on the safe side of the results was considered non-essential since the major purpose of the model is to

help define combinations that would be within this zone (no matter where). On the other hand, more accurate predictions on the side that could trigger nonacceptable results, would help to avoid them by not selecting the raw materials combinations considered in that specific prediction. Once the model was selected, it was introduced to the company official process, replacing the previous version of the model. The updated model was used to define the raw material combinations for several manufacturing lots. For this purpose, all the lots of raw materials and the DSI product lots available were crossed against each other using the updated model. More than 70 combinations options were evaluated with the new model. Based on the predictions results, the best combinations options were narrowed down to the expected campaign amount of lots plus one spare lot.

Once the lots were manufactured following the raw materials combination guidance provided by the predictive model, the impurity results were obtained for the manufactured lots. The lots complied with the acceptance criteria as predicted by the model. When evaluated in detail, only one of the predictions was outside of the prediction range but, although close, it was within the acceptance criteria as shown in Figure 5 . This was expected since the model was not selected for its precision on the acceptable side, but for its accuracy on the non-acceptable side. Also, the updated model, although more representative, possesses a more limited data set than the previous model.

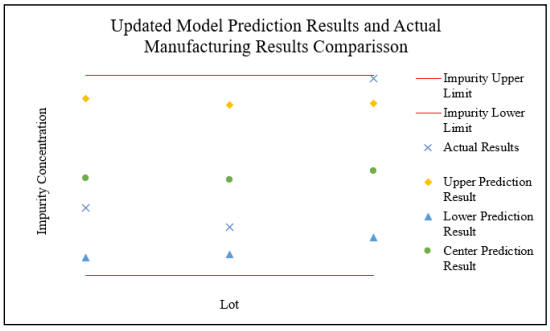

Figure 5
Comparison of Lower, Center and Upper Prediction Results against Actual Manufacturing Results for Three Lots

Therefore, the pre use raw material evaluation using the updated predictive model was effective on reducing the probability of obtaining impurity results outside of the acceptance criteria. This results on a successful campaign with 100% success rate, product availability and cost avoidance on resources to manage deviations.

CONCLUSION

This project aimed to update a raw material pre-use predictive model to generate impurity predictions that could reduce the probability of obtaining impurities results outside of the acceptance criteria. After implementing the updated predictive model, the results were satisfactory since consecutive lots obtained impurity results within the acceptable criteria. Consequently, no deviation was generated, and the product will be available for patient distribution.

One of the limiting factors for this project was data availability. Although more representative, the updated model possesses a more limited data set than the previous one. Manufacture data can only be obtained from completed lots. Therefore, the data can be improved at the same pace as the lots are manufactured. For future work and once more data is generated, the model can include these new lot's data to keep adding robustness to its predictions and continue having satisfactory results. It is also recommended that the automation of the data gathering, processing, and model generation to reduce the workload.

REFERENCES

[1] A. Rathore, R. Bhambure & V. Ghare, "Process analytical technology (PAT) for biopharmaceutical products," in *Analytical and Bioanalytical Chemistry*, May 18, 2010, vol. 398, pp. 137-154. Available: https://link.springer.com/article/10.1007/s00216-010-3781-x.

[2] M. C. Flickinger, *Downstream Industrial Biotechnology: Recovery and Purification*, John Wiley & Sons, Inc., 2013.

[3] B. Sissolak, N. Lingg, W. Sommeregger, G. Striedner & K. Vorauer-Uhl, "Impact of mammalian cell culture conditions on monoclonal antibody charge heterogeneity: an accessory monitoring tool for process development," in *Journal of*

Industrial Microbiology & Biotechnology, 2019, vol. 46, no. 8, pp. 1167-1178. Available: https://ezproxy.pupr.edu:2093/10.1007/s10295-019-02202-5.

[4] N. Mengual-Macenlle, P. J. Marcos, R. Golpe & D. González-Rivas, "Multivariate analysis in thoracic research," in *Journal of Thoracic Disease*, 2015, pp. 1-5.

[5] L. Guo, B. Zhang & X. Zhao, "A Consumer Behavior Prediction Model Based on Multivariate Real-Time Sequence Analysis," in *Mathematical Problems in Engineering*, May 2021, pp. 1-5. Available: https://ezproxy.pupr.edu:2093/10.1 155/2021/6688750.

[6] B. Nayana, K. R. Kumar & C. Chesneau, "Wheat Yield Prediction in India Using Principal Component Analysis-Multivariate Adaptive Regression Splines (PCA-MARS)," in *AgriEngineering*, 2022, vol. 4, no. 2, pp. 461-474. Available: https://ezproxy.pupr.edu:2093/10.3390/agriengineering4020030.

[7] M. Sokolov, et. al., "Fingerprint Detection and Process Prediction by Multivariate Analysis of Fed-Batch Monoclonal Antibody Cell Culture Data," in *Biotechnology Progress*, 2015, pp. 1633-1644. Available: https://ezproxy.pupr.edu:2093/10.1002 /btpr. 2174.

[8] Minitab, LLC. (April 5, 2023). *Minitab Software* [Online]. Available: https://support.minitab.com/en-us/minitab/20/help-and-how-to/statistical-modeling/multivariate/supporting-topics/basics/what-is-multivariate-analysis/.

[9] JMP Statistical Discovery LLC. (April 5, 2023). *JMP Software* [Online]. Available: https://www.jmp.com/en_nl/learning-library/topics/multi variate-methods.html.