Exploratory analysis of 3D Fluorescence spectra of raw material and link to the process yield in Bioproduction

A. Maléchaux[1] A. Kapitan[2] A. Croguennoc[3] S.Preys[4] A. Restivo[5] S. Hansen[6]

[1] Ondalys, 4 rue Georges Besse, 34830 Clapiers, France, amalechaux@ondalys.fr

[2] GSK, 89 Rue De L'Institut, B-1330 Rixensart, Belgium, andre.x.kapitan@gsk.com

[3] Ondalys, 4 rue Georges Besse, 34830 Clapiers, France, acroguennoc@ondalys.fr

[4] Ondalys, 4 rue Georges Besse, 34830 Clapiers, France, spreys@ondalys.fr

[5] GSK, 1 via Fiorentina, 53100 Siena, Italy, antonino.x.restivo@gsk.com

[6] GSK, 89 Rue De L'Institut, B-1330 Rixensart, Belgium, sylvain.x.hansen@gsk.com

Keywords: 3D Fluorescence, PARAFAC, unfolded PCA, N-PLS, unfolded PLS.

1 Introduction

Raw materials used in cell fermentation can be complex. Because cell culture process conditions are highly controlled in biopharmaceutical manufacturing, the raw material lot-to-lot variability can therefore have a high impact on the yield of the vaccine antigen production processes. 3D fluorescence spectroscopy can be used to characterize complex raw materials and thus is a useful option for the holistic analysis. The combination of spectroscopy and chemometric methods like principal component analysis (PCA), parallel factor analysis (PARAFAC) or partial least squares (PLS) and multilinear PLS (N-PLS) modeling has been commonly used as a tool for the identification or classification of raw materials. In our presentation, we show the usefulness of 3D fluorescence to characterize raw materials and ultimately predict antigen yield based on the batch raw material.

2 Material and methods

The first objective was to identify the best pretreatment combination to improve as much as possible the reproducibility of fluorescence measurements. Forty-three multiway 3D fluorescence spectra from 16 batches of raw material with 2 to 4 repetitions per batch were used. The reproducibility of measurements was evaluated by visualization of the score plots and by calculation of the Wilks' lambda on the scores of the unsupervised models developed either on 3D spectra with PARAFAC or on unfolded spectra with PCA. The second objective was to evaluate the link between the fluorescence spectra of a key raw material and the end-product yield of antigen. The quality attribute to predict was optimized as the median of antigen yield normalized by the average number of viable cells. Seventeen fluorescence spectra from 6 batches of the key raw material were used for the development of supervised models, either by N-PLS using 3D spectra or by PLS using unfolded spectra, with a cross-validation strategy by raw material batch. The performances of the prediction models were evaluated based on their root mean squared error of cross-validation (RMSECV) and their determination coefficient (R^2). Finally, variable selection by interval PLS (i-PLS) or by knowledge from the PARAFAC loadings was carried out to improve the performances of the supervised regression models.

3 Results and discussion

The unsupervised analysis of fluorescence spectra, using unfolded PCA or PARAFAC, demonstrated the importance of combining several pretreatments to improve the intra-batch

reproducibility and inter-batch separation. PCA on unfolded spectra allowed a better visualization of the effect of pretreatments on the groups of samples, whereas PARAFAC on 3D spectra favored the interpretation of loadings and the identification of important variables. The optimal pretreatment combination consisted of removing the Rayleigh scattering areas[1,2], applying a baseline correction, normalizing by the amplitude of the water Raman band[3], and normalizing by the raw material weight.

The supervised analysis emphasized the importance of selecting variables that contain the most relevant information to predict this parameter of interest, in addition to choosing the appropriate combination of spectral pretreatments. In the optimal conditions, classic PLS models on unfolded fluorescence spectra reached similar performances compared to more complex multiway N-PLS models on 3D spectra. The best predictions were obtained after reducing the Emission range to 400-428 nm and the Excitation range to 310-378 nm, which corresponds to a band identified on the third component of PARAFAC during unsupervised analysis (Figure 1).

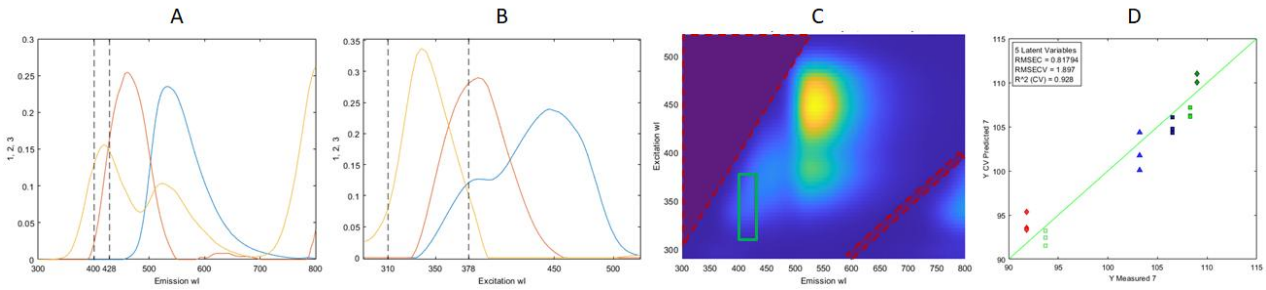

Figure 1: Results of the prediction of the protein yield using PARAFAC knowledge and PLS. (A-B) Emission and excitation components obtained by PARAFAC. (C) Area selected. (D) Summary of PLS performance.

4 Conclusion

We have developed a 3D fluorescence method which can be used to characterize a complex raw material and ultimately establish a link to the efficacy of individual batch of raw material in terms of protein yield. After the optimization of the pretreatments used to improve the intra-batch reproducibility and inter-batch separation, and the selection of the most important variables by knowledge from the PARAFAC loadings, the optimal PLS model gives a RMSECV of 1.90 for the normalized antigen yield corresponding to a relative error of 1.86%. Compared to the Standard Error of Laboratory of about 3% for this attribute, the prediction model seems to be very promising. However, these results should be taken with caution as the number of batches used to calibrate the model was very low and a real independent test could not yet be performed. This ability to predict product yield before the use of the raw material has the potential to control one of the major process variables increasing the robustness of process performance and identifying leverage to improve product yield.

5 References

[1] Thygesen, L. G., Rinnan, Å., Barsberg, S., & Møller, J. K. (2004). Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area. Chemometrics and Intelligent Laboratory Systems, 71(2), 97-106.

[2] Bahram, M., Bro, R., Stedmon, C., & Afkhami, A. (2006). Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. Journal of Chemometrics: A Journal of the Chemometrics Society, 20(3- 4), 99-105.

[3] Lawaetz, A. J., & Stedmon, C. A. (2009). Fluorescence intensity calibration using the Raman scatter peak of water. Applied spectroscopy, 63(8), 936-940.