CHARACTERIZATION OF RAW MATERIAL INFLUENCE ON MAMMALIAN CELL CULTURE PERFORMANCE: CHEMOMETRICS BASED DATA FUSION APPROACH

Hae Woo Lee[a], Andrew Christie[b], Seongkyu Yoon[a]*
[a]University of Massachusetts, Lowell, MA 01854,
[b]SAFC Biosciences, Lenexa, KS, USA

Abstract

Near-infrared, Raman, fluorescence and X-ray fluorescence spectra of multiple soy hydrolysate lots manufactured by different vendors were analyzed for comprehensive characterization of raw materials used in mammalian cell culture processes. The variability of soy hydrolysates, as well as the correlation between multiple spectra and cell culture performance was addressed. The identified compositional variability was further analyzed in order to estimate the growth and protein production of two mammalian cell lines. Multiple spectral platforms were compared with each other in terms of their estimation capability, and finally integrated into a unifying prediction model using data fusion strategies. The performance of the resulting models demonstrated the potential of data fusion of multiple spectroscopies as a robust lot selection tool for raw materials while providing a biological link between the chemical composition of raw materials and cell culture performances.

Keywords

Raw materials characterization, Soy hydrolysate, CHO, Mammalian cells, data fusion, chemometrics.

Introduction

Therapeutic recombinant proteins, such as monoclonal antibodies, are often produced from mammalian cell culture process using Chinese Hamster Ovary (CHO) cells. CHO cell based platform provides a few attractive features: easy maintenance, safe use in humans, capability of post-translational modifications and acceptable regulatory standards. Along with a rapidly growing demand for biopharmaceuticals, considerable effort has been devoted to improving the productivity of CHO cells by developing proficient cell lines, formulating culture medium, and optimizing process conditions (Kim et al., 2004). Despite these significant advances, however, the performance of mammalian cell culture processes are highly variable, often resulting in inconsistent critical quality attributes for their final products (Rathore et al., 2010).

One of the most common sources of variability in mammalian cell culture processes is the composition of the raw materials or culture media, which are complex mixtures of nutrients. In many commercial processes that are producing biopharmaceuticals, variability in critical raw materials can have a great impact on the product quality as well as process, since most manufacturing processes are kept under tight controls with fixed operating conditions to meet strict regulatory requirements. Plant protein hydrolysates are often supplemented with culture medium to improve protein production from recombinant CHO cells in serum free environment (Lu et al., 2007). They have large influences on mammalian cell cultures. The drawback of using plant hydrolysates is that they often exhibit considerable variability in their growth-promoting and production-enhancing activities due to their compositional uncertainty. Furthermore, in many cases, a comprehensive analysis of these complex materials is rather time consuming, complicated and expensive, so it is impractical for routine use.

Application of spectroscopic techniques to raw materials can provide fast, simple and non-destructive ways to measure physicochemical properties or compositional variability of them (Kirdar et al, 2010; Ryan

et al., 2010). A simple and convenient characterization tool can eventually lead to the reduction of variability in the final product quality of the therapeutic proteins. Furthermore, the use of different spectroscopy platforms can provide complementary information regarding the chemical composition of raw materials. However, until now, only a few studies have been made for the use of multiple spectra in characterizing the raw materials.

Motivated by the importance of rapid media characterization in bioprocesses, a comprehensive data fusion strategy based on the multiple spectroscopic measurements of soy hydrolysates was employed in this study in order to distinguish the good and bad lots. Current methods to evaluate the quality of soy hydrolysate are mostly based on time consuming bioassays. Therefore, the application of multiple spectroscopic techniques, such as near-infrared, Raman, 2-D fluorescence and X-ray fluorescence (XRF) system can be a good alternative to these labor-intensive bioassays. Variability of soy hydrolysates, as well as the correlation between multiple spectra and cell-based assay results was first addressed using principal component analysis. Then, the spectral data sets were combined with chemometric tools to predict the cell growth and titer of two different cell lines. Different spectroscopic platforms were compared with each other in terms of their predictability, and the efficient methods to combine these multiple spectra were investigated using various data fusion strategies.

Materials and Methods

Samples

A total of 15 soy hydrolysate samples were obtained from different manufacturing lots produced by four vendors (A, B, C and D). The number of lots for each vendor was dependent on the availability of samples; therefore, nine, two, two and two lots are used for vendor A, B, C and D, respectively. Detailed composition, and specific vendor information, for each soy lot were not known for proprietary reason. All samples were stored in a refrigerator at 4^{o}C upon their arrival and were equilibrated at room temperature prior to the subsequent analysis.

Spectral Acquisition

Near-Infra spectra of soy hydrolysates were measured on a Bruker MPA FT-NIR spectrophotometer (Bruker Optics). To measure near-infrared spectra, all of the samples were packed into 22 mm glass vials and then scanned in the wavenumber range: 12500 - 4000 cm^{-1}, using the reflectance mode. Here, the number of co-added scans and resolution were 64 and 8 cm-1, respectively, which were sufficient to achieve a high signal to noise ratio for the given samples.

Raman spectra of soy hydrolysate were measured on a RXN3 Raman spectrophotometer (Kaiser Optical Systems, Inc.) equipped with an optical fiber probe. Due to the large fluorescence of soy hydrolysate powder samples, all spectra were measured after dissolving the solid powder into the distilled water with concentration of 10 g/l. Here, the spectral range was 100 – 1900 cm^{-1} with resolution of 1 cm^{-1}, and 32 co-added scans were taken for each sample with exposure time of 30 seconds.

2D-fluorescence spectra were measured on a LS45 fluorescence spectrometer (Perkin Elmer, Inc.) by employing multiple excitation wavelengths. All spectra were measured after dissolving the solid powder with concentration of 1 g/l. Excitation and emission wavelengths employed here were 200-450 nm (increment of 10nm) and 200-800 nm (increment of 1nm), respectively, and scanning speed was set to 1000 nm/min.

XRF spectra of soy hydrolysate were measured on a Niton FXL XRF analyzer (Thermo Scientific), equipped with geometrically optimized large area drift detector. To measure XRF spectra, all samples were packed into 8 mm cup and then scanned with sample spinner. Here, instead of utilizing raw XRF spectra, elemental analysis data provided by the built-in calibration curves of the Niton FXL analyzer were directly used. Note that all spectral measurements were conducted in triplicate except for XRF analysis.

Bioassays

To evaluate the culture performance of CHO cells grown from the different soy hydrolysate lots, biological assays were performed based on a dose response model. Here, two different CHO cell lines (A and B) expressing immunoglobulin (IgG) were employed. In the bioassays, the expanded seed cells were first washed and inoculated into replicate culture tubes, each of which contained 30 ml of the test medium supplemented with different soy concentrations (0, 1, 2.5, 5, 7.5, 10, 12.5 and 15 g/l) in duplicate. After seeding the cells, the cultures were grown in a shaking incubator at 200 rpm under the conditions of 37 ^{o}C and 5 % CO2; they were harvested after 7 days for analyzing the growth and productivity of the cells. Integral Viable Cell Density (IVCD) and IgG concentration were measured from the harvested cells using Cedex (Roche Innovatis, Germany) and Qctet QK (ForteBio Inc., USA), respectively as quantifiable culture performance indices.

Multivariate Data Analysis

Prior to conducting the multivariate analysis, all the spectra were preprocessed in order to suppress unwanted variations originating from light scattering, background drift and instrumental artifacts. For this, multiplicative scatter correction combined with the first derivative method was applied to the near-infrared spectra. On the other hand, baseline correction with polynomial fitting and total intensity normalization were utilized for Raman spectra. In addition, Savitzky-Golay smoothing with removal of Rayleigh and Raman scattering were carried

out for 2-D fluorescence spectra, followed by unfolding of 2-D map of each sample into one vector.

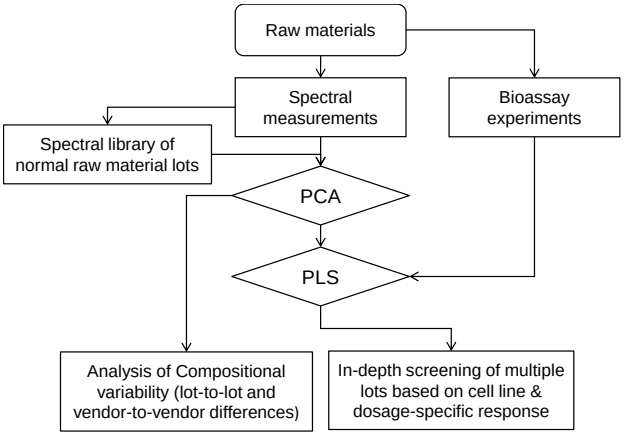

Figure 1. *Schematic diagram of multivariate data analysis conducted in this study*

Then, multivariate data analyses were conducted on the multiple experimental data independently using principal component analysis (PCA) and partial least squares (PLS) methods (Eriksson et al., 2001) as shown in Figure 1. PCA and PLS are the useful multivariate analysis tools for data visualization and dimensionality reduction, which can be achieved by projecting the raw data into a new low-dimensional principal component or latent variable space. Therefore, the datasets with thousands of variables can be easily characterized by a small set of new variables, called scores. PCA is usually used to analyze a single dataset, while PLS is utilized for the regression of two datasets which consist of predictor (**X**) and response variables (**Y**). Here, PCA was employed to identify the variability of the soy hydrolysate lots by the multiple spectra, and PLS is used to estimate the cell culture performance from each spectral measurement. In both methods, 5-fold cross validation was utilized to determine the optimal number of principal components (PC) or latent variable (LV) for PCA or PLS models. In addition, the prediction accuracy of the PLS model was quantified by either R^2 or Q^2 (≤ 1) of the leave-one-out cross-validation (Eriksson et al., 2001). Then, the validity and the statistical significance of the developed PLS models were confirmed by using permutation test (Eriksson et al., 2001).

To combine different spectra measured by near-infrared, Raman, 2-D fluorescence and XRF analyzers, consensus PCA (CPCA) and multi-block PLS (MBPLS) were employed (Westerhuis et al., 1998). In these multi-block methods, multiple blocks of either **X** or **Y** datasets are simultaneously modeled, thus four spectroscopic data can be efficiently fused to generate a single unified estimation model. To select the optimal number of latent variables in CPCA and MBPLS, 5-fold cross-validation was used again, and then permutation test was conducted to examine the validity of the constructed MBPLS model.

Note that all calculations for multivariate analyses were conducted using PLS toolbox (Eigenvector, Inc.) and home-written routines in MATLAB (Mathworks, Inc.).

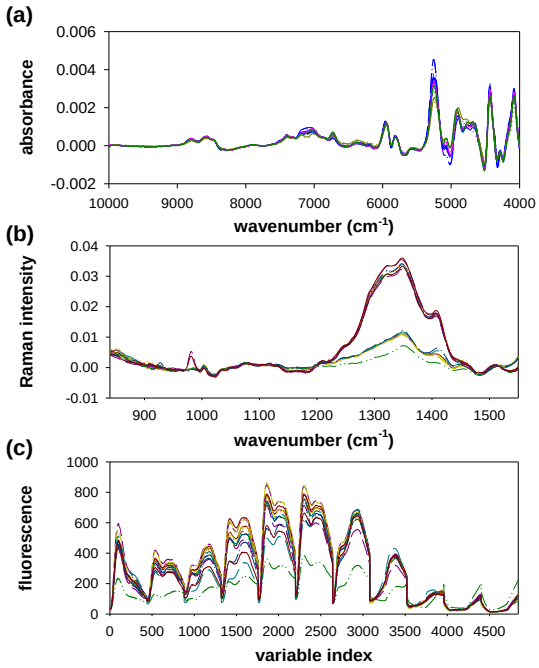

Figure 2. *Preprocessed spectra of different soy hydrolysate lots. (a) near-infrared; (b) Raman; (c) fluorescence spectra.*

Results and Discussion

Near-infrared, Raman, fluorescence and XRF spectra

Figure 2 shows the near-infrared, Raman and fluorescence spectra obtained from the different soy lots. In the near-infrared spectra, a spectral region below 1000 cm^{-1} was discarded from the entire spectra because the signal-to-noise ratio was very low in that region. In addition, the regions below 850 cm^{-1} and above 1550 cm^{-1} from the Raman spectra were also discarded due to the artifacts. The final spectra shown in Figure 1 clearly illustrate that there are noticeable variations in several regions, which correspond to the numerous chemical components contained in the soy hydrolysates.

Table 1 presents the summary of elemental analysis measured by XRF. In XRF data, chemical elements, whose concentrations (%) were smaller than the error range (provided automatically by instrument) with more than 50% of total number of the samples, were discarded, and only the elements having significant variations were included for the subsequent analysis. Pearson correlation coefficients were calculated to investigate the correlation between XRF data and bioassays, and the results revealed that chemical elements of tungsten, calcium, chloride and iron have a significant effect on the growth and productivity of two mammalian cell lines (p-value < 0.01).

***Table 1**. Summary of XRF analysis.*

element	Mean	Standard deviation	Max.	Min.
Ag	0.0033	0.0012	0.0070	0.0020
Balance*	78.2	4.617	84.3	66.2
Rb	0.0023	0.0008	0.0040	0.001
W	0.0210	0.0056	0.0330	0.0150
Zn	0.0039	0.0045	0.0130	0
Cu	0.0044	0.0029	0.0100	0.001
Fe	0.0136	0.012	0.0370	0
Ca	0.382	0.168	0.758	0.112
K	12.9	3.07	19.8	9.73
Al	0.0917	0.0575	0.238	0.0360
P	1.47	0.370	2.47	1.15
Si	0.0152	0.0229	0.0610	0
Cl	3.67	2.27	7.54	0.131
S	3.20	1.18	7.61	1.32

*Balance was calculated by [100 – sum of all species] (%)

Bioassay Data

Figure 3 shows results obtained from the cell culture-based bioassays, which were performed to evaluate the effect of different soy hydrolysate lots on mammalian cell culture processes. For this, after seven days of inoculation, IVCD and IgG quantities were measured from each of the culture tubes, where two CHO cell lines were cultivated in a medium supplemented by each of the soy hydrolysate lots with varying concentrations.

In the resultant response curves, shown in Figure 3, it could be observed that there is high variability in the multiple soy hydrolysate lots with each having different effects on the corresponding mammalian cell cultures. In overall, the variability of the soy lots in promoting growth and productivity of mammalian cells became more distinct at higher dosage region, although the IVCD profiles show more apparent dosage-dependent responses to the variability of the soy lots, compared to the IgG profiles. In addition, some lots even exhibited inhibitory cell growth and declined IgG production at high dosage region, indicating that these lots can potentially influence subsequent mammalian cell culture process in a negative way. On average, the soy hydrolysate lots from vendor A exhibited better performance by stimulating the cell growth and protein production than the remaining ones, indicating the effects of different manufacturing vendors. This was further confirmed by hierarchical clustering with Euclidean metric and average linkage method. The results revealed that good, intermediate and poor performing lots could be discriminated mostly by their vendors although more samples are needed to generalize this observation (data are not shown). The above results demonstrate that variations of soy hydrolysate lots supplemented in the cell culture medium can induce alterations in the growth and productivity profiles of mammalian cell cultures, stressing the necessity of efficient screening tools for the corresponding raw materials.

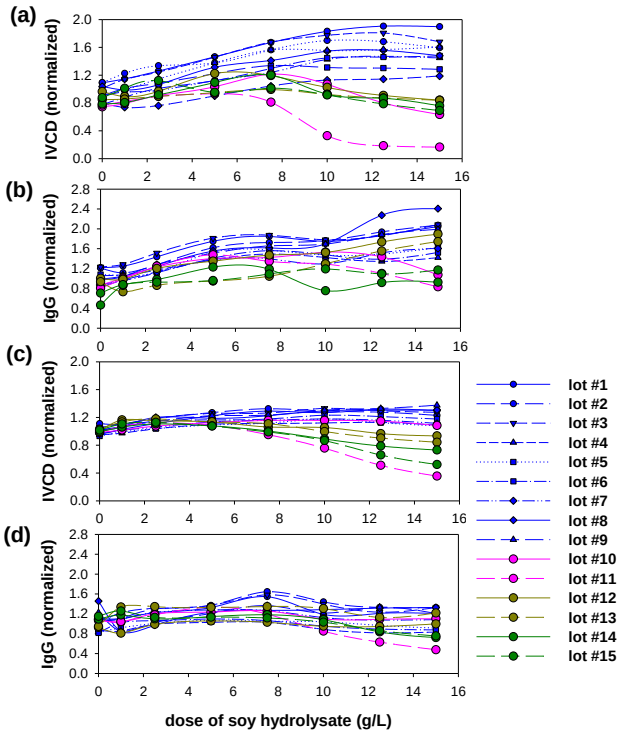

***Figure 3.** Dose-response curves of IVCD and IgG profiles obtained by bioassays. (A) IVCD profiles for cell line A; (b) IgG profiles for cell line A; (c) IVCD profiles for cell line B; (d) IgG profiles for cell line B.*

Identification of variability in soy hydrolysate lots

In order to examine the feasibility of each spectroscopic technique in characterizing the variability of the soy hydrolysate lots, PCA models were constructed separately on each spectral data set. Figure 4 shows the resultant score plots, which describe the variability of spectra by two major PCs. Several distinct groups of the soy lots were clustered together according to their manufacturing vendors, except for vendor B where one of the lots exhibited poorest performance in the bioassay results. For the comparison of these results with their explicit quality in mammalian cell cultures, additional PCA model was constructed on the bioassay profiles as shown in Figure 4 (e). Surprisingly, there was high similarity in the clustering patterns between the different spectral data sets and bioassays, especially for the near-infrared and Raman spectra. This might illustrate that the different spectroscopic techniques employed here have capability of capturing the compositional differences among different soy lots originating from lot-to-lot and vendor-to-vendor variability.

Additionally, to further understand whether there are overlapping features among the different spectral datasets, CPCA model was constructed on the combined dataset of

near-infrared, Raman, fluorescence and XRF measurements. Four PCs were selected as an optimum, explaining 89.2% of total variance. Figure 5 represents cumulative percent variance explained by each PC for each data block. In this model, first and second PCs (PC1 and PC2) mainly described the Raman and fluorescence spectra, indicating that they might capture common features from the soy hydrolysates. On the other hand, third PC mainly described the variance of near-infrared spectra, and forth PC explained XRF spectra, differentiating them from the remaining ones. This result illustrates that they might have complementary information about the compositional variability of the soy, and justifies the fusion of different analytical techniques employed here for screening the raw material lots.

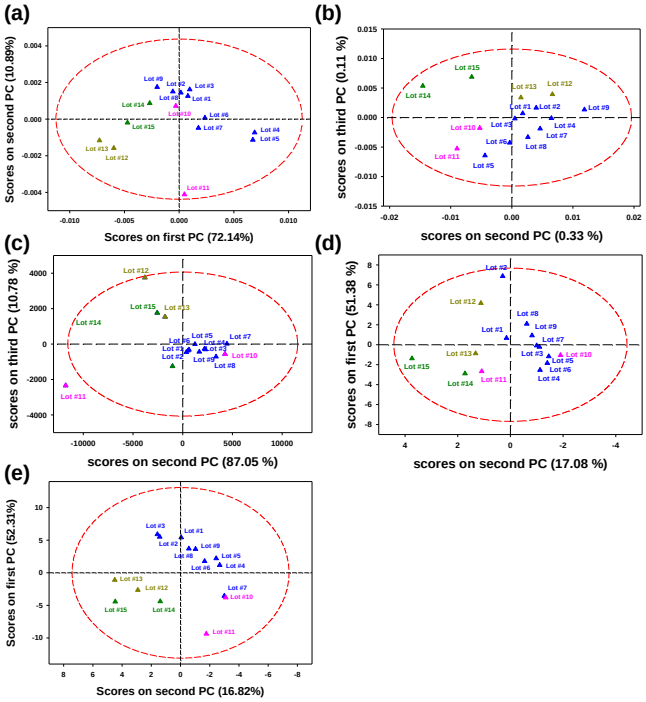

Figure 4. *Score plots obtained by separate PCA models on (a) near-infraed; (b) Raman; (c) fluorescence; (d) XRF; (e) bioassay profiles. blue-vendor A; pink-vendor B; yellow-vendor C; green-vendor D.*

Estimation of cell culture performance

In this section, the growth and productivity profiles of the different cell lines were estimated using combination of multiple spectra in order to evaluate the performance capability of raw materials. For this, several PLS models were constructed with different combinations of the multiple spectra in order to predict each of the IVCD and profiles of the two different CHO cell lines under the condition of varying soy dosages.

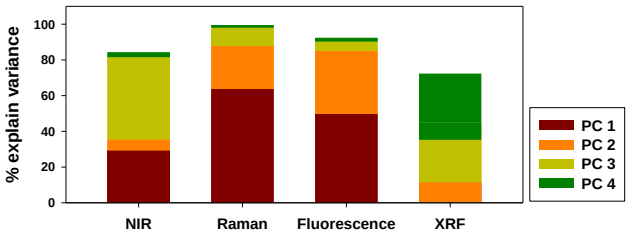

Figure 5. *Cumulative percent variance explained by CPCA model for each spectral data block.*

First, the prediction performance of the PLS models, which estimates either the IVCD or IgG values of different cell lines, was evaluated by employing a single spectroscopic technique. In Figure 6, Q^2 values of the models were displayed for two soy dosages (5 and 10 g/l) as a representative case. As can be seen from this figure, the prediction accuracy of PLS model was generally better at high dosage regions and declined as the soy dosage decreased. This is especially true in the IgG models of cell line B, where Q^2 of the most models exhibited negative values at low concentration ranges. These results are in line with the analysis of variance (ANOVA) of the bioassays. This also revealed that there were no significant effects from soy hydrolysate at lower concentration ranges (p-value>0.05, data are not shown). Thus, the variation of cell culture performance induced by different soy lots was less pronounced at lower concentration ranges, and the quality of soy hydrolysates in the cell culture processes might not be adequately predicted. The PLS models constructed here correctly captured these phenomena, illustrating the validity of the developed statistical models.

Among the different spectroscopic measurements, in general, near-infrared and Raman spectroscopy provided most reliable estimations compared to the other two spectra, regardless of differences in the cell lines and soy dosages. However, in some cases, fluorescence or XRF spectra gave more accurate prediction, highlighting the needs of combining the multiple spectra in order to obtain more robust estimation model.

To examine the different data fusion strategies for the multiple spectra, all possible combinations among four spectra, such as near-infrared (N) + Raman (R) or near-infrared (N) + fluorescence (F) + XRF (X) were made and their prediction accuracy represented by Q^2 was examined as shown in Figure 6, where only the cases of two soy dosages (5 and 10 g/l) are represented due to the space limit. In general, the advantages of combining the multiple spectra could be seen in most cases by improving the prediction accuracy of the estimation models, but there was no unique combination method which dominates over the others under the various conditions (i.e. two cell lines and different dosages). However, among different combinations, fusion of the near-infrared spectra with others generally showed the best prediction performance.

As illustrated in the previous section of CPCA model, near-infrared spectra did not share the common features with the other spectra, so the combinations with other datasets might provide some complementary information about the soy hydrolysates, resulting in the improved prediction accuracy. Therefore, incorporation of the near-infrared spectra with other sources of spectroscopic techniques might be an optimal data fusion strategy in constructing the prediction models for estimating the growth and productivity of mammalian cell cultures.

Overall, the prediction powers for most PLS models were acceptable at high dosage regions, showing there is a high correlation between the variability of raw materials and the resultant cell culture performance. In some models, the prediction accuracy was quite high (Q^2>0.8), suggesting that these models can be used to estimate cell culture performance directly from the multiple spectra instead of utilizing time-consuming bioassays. Considering that the bioassays implemented here took seven days to complete, the fast and simple nature of the spectroscopic techniques poses a great potential for the use of them as a real-time or near real-time inspection tool of the incoming raw material lots in mammalian cell cultures. At the same time, the procedures used in the identification of the lot or vendor differences can be ideally combined with real-time multivariate statistical control schemes, which might gain another benefit in the manufacturing processes of raw materials.

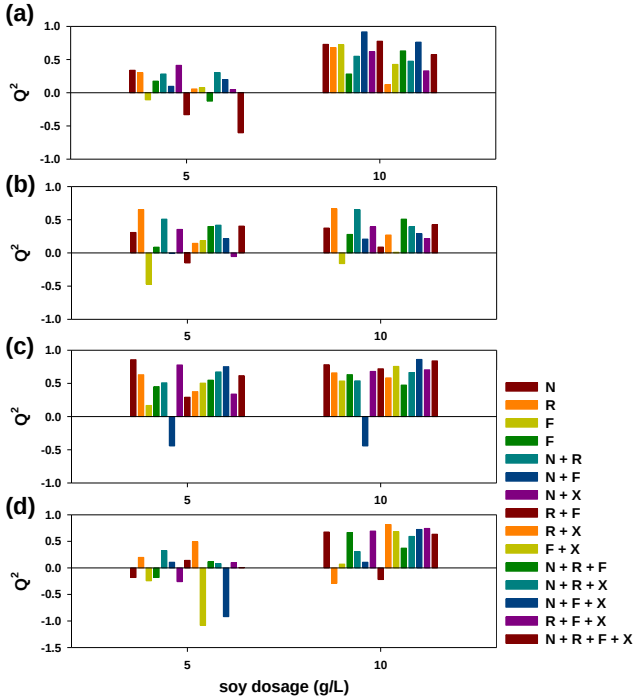

Figure 6. *Prediction performance (Q^2) of PLS models. (a) IVCD of cell line A; (b) IgG of cell line A; (c) IVCD of cel line B; (d) IgG of cell line B.*

Conclusion

In this study, the multiple spectra of different soy hydrolysate lots was analyzed in order to develop a fast screening tool for the raw materials in mammalian cell culture processes. By using a chemometric approach, it was demonstrated that data fusion of different spectroscopic technique can be used to reveal lot-to-lot variability, as well as vendor-to-vendor differences of soy hydrolysate, which cannot be avoided for these chemically undefined raw materials. At the same time, the prediction models for estimating cell growth and productivity of mammalian cell cultures from near-infrared spectra were constructed, providing estimation of the cell culture performance under conditions of varying soy dosages in a cell line-specific manner.

Acknowledgments

Financial support was provided by Massachusetts Life Science Center and University of Massachusetts.

References

Eriksson L, Johansson E, Kettaneh Wold N, Wold S. (2006). Multi and Megavariate Data Analysis, 2nd ed. Umetrics AB, Umea, Sweden.

ICH. (2008). Q8(R1): Pharmaceutical Development. Geneva, Switzerland: International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use.

Isaksson, T, Naes, T. (1998). The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. Appl. Spectrosc. 42, 1273.

Kim, J.M., Kim, J.S., Park, D.H., Kang, H.S., Yoon, J., Baek, K., Yoon, Y. (2004). Improved recombinant gene expression in CHO cells using matrix attachment regions. J. Biotechnol. 107, 95.

Kirdar, A.I., Chen, G., Weidner, J., Rathore, A.S. (2010). Application of Near-Infrared (NIR) spectroscopy for screening of raw materials used in the cell culture medium for the production of a recombinant therapeutic protein. Biotechnol. Prog. 26, 527.

Lu, C., Gonzalez, C., Gleason, J., Gangi, J., Yang, J.-D. (2007). A T-flask based screening platform for evaluating and identifying plant hydrolysates for a fed-batch cell culture process. Cytotechnology 55, 15.

Rathore, A.S., Bhambure, R., Ghare, V. (2010). Process analytical technology (PAT) for biopharmaceutical products. Anal. Bioanal. Chem. 398, 137.

Ryan, P.W., Li, B., Shanahan, M., Leister, K., Ryder, A.G. (2010). Prediction of cell culture media performance using fluorescence spectroscopy. Anal. Chem. 82, 1311-1317.

Westerhuis, J.A., Kourti, T., MacGregor, J.F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. J. Chemom. 12, 301-321.