



Provided by the author(s) and NUI Galway in accordance with publisher policies. Please cite the published version when available.

| | |
|-----------------------------|---|
| Title | Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. |
| Author(s) | Ryder, Alan G.; Li, Boyan; Ray, Bryan H. |
| Publication Date | 2013 |
| Publication Information | B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder (2013) 'Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy'. <i>Analytica Chimica Acta</i> , 796 :84-91. |
| Link to publisher's version | http://dx.doi.org/10.1016/j.aca.2013.07.058 |
| Item record | http://www.sciencedirect.com/science/article/pii/S0003267013010349 ; http://hdl.handle.net/10379/3938 |
| DOI | http://dx.doi.org/10.1016/j.aca.2013.07.058 |

Downloaded 2022-10-07T20:12:49Z

Some rights reserved. For more information, please see the item record link above.



Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

1 This is the author version. The definitive version is available on the *Analytica Chimica Acta*
2 website.

3 4 **PERFORMANCE MONITORING OF A MAMMALIAN CELL** 5 **BASED BIOPROCESS USING RAMAN SPECTROSCOPY.**

6
7 Boyan Li,¹ Bryan H. Ray,¹ Kirk J. Leister,² and Alan G. Ryder.^{1*}

8
9 ¹ Nanoscale Biophotonics Laboratory, School of Chemistry, National University of Ireland,
10 Galway, Galway, Ireland.

11 ² Bristol-Myers Squibb, Process Analytical Sciences, Syracuse, New York, USA.

12 * Corresponding Author: **Email:** alan.ryder@nuigalway.ie. **Tel:** +353 91 49 2943

13 14 **Abstract:**

15 Being able to predict the final product yield at all stages in long-running, industrial, mammalian
16 cell culture processes is vital for both operational efficiency, process consistency, and the
17 implementation of Quality by Design (QbD) practices. Here we used Raman spectroscopy to
18 monitor (in terms of glycoprotein yield prediction) a fed-batch fermentation from start to finish.
19 Raman data were collected from 12 different time points in a Chinese Hamster Ovary (CHO)
20 based manufacturing process and across 37 separate production runs. The samples comprised of
21 clarified bioprocess broths extracted from the CHO cell based process with varying amounts of
22 fresh and spent cell culture media. Competitive adaptive reweighted sampling (CoAdReS) and
23 ant colony optimization (ACO) variable selection methods were used to enhance the predictive
24 ability of the chemometric models by removing unnecessary spectral information. Using
25 CoAdReS accurate prediction models (relative error of predictions between 2.1–3.3%) were built
26 for the final glycoprotein yield at every stage of the bioprocess from small scale up to the final
27 5000L bioreactor. This result reinforces our previous studies which indicate that media quality is
28 one of the most significant factors determining the efficiency of industrial CHO-cell processes.
29 This Raman based approach could thus be used to manage production in terms of selecting
30 which small scale batches are progressed to large-scale manufacture, thus improving process
31 efficiency significantly.

32
33 **Keywords:** Raman spectroscopy, Bioprocess, Glycoprotein, Chemometrics, Variable Selection,
34 CHO cell.

35 36 **1. INTRODUCTION**

37 The manufacture of therapeutic proteins by mammalian cell culture based processes is
38 driving the development of a new generation of spectroscopic (primarily vibrational) based
39 analytical methodologies [1-8]. The need for rapid, reliable, robust, and non-destructive

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

40 analytical methods is of paramount importance to ensure efficient and reliable process control, to
41 improve fermentation performance and product quality, leading to decreases in cost-of-product.
42 Ideally it would be best if these methods could enable the accurate prediction of final yield (and
43 other product quality attributes) as early as possible in the process cycle, preferably in the seed
44 reactor.

45 A fed-batch fermentation process for recombinant protein production, starting with the cell
46 bank vial and ending with the final product, is a very complex system. Multiple process
47 parameters that determine product yield and other desired quality attributes include feed quality,
48 feeding strategy, inoculum age, and harvest point (to name but a few) [9-11]. Once a process
49 seed reactor has been transferred to the large-scale manufacturing bioreactor stage, most of the
50 process operational parameters will have been fixed, except for feed quality, which can vary
51 substantially. In bioreactors during growth and production phases, there is a complex
52 environment, comprising of materials that include feed media (a mixture of amino acids,
53 inorganic salts, carbohydrates, organic acids, vitamins, etc.), whole cells and cell debris, product
54 and host cell protein, and metabolites [9, 12, 13]. The analysis of these complex materials is
55 challenging, and chromatographic techniques (often coupled with mass spectrometry) offer the
56 necessary chemical resolution for detailed analysis [14, 15]. Alternatively one can consider the
57 use of high-field NMR which can generate extensive information about the constituents of cell
58 culture media [16]. However, these approaches are generally only implemented with a low
59 analysis frequency because of tedious sample preparation, high capital cost and often highly
60 skilled, labor-intensive/ time-consuming data analysis.

61 The potential of near-infrared (NIR) and mid-infrared (MIR) spectroscopies has been
62 documented for bioprocess analysis [1-3, 5, 6, 17]. These methods are however hindered by the
63 very strong water signal, so in aqueous solutions, much of the critical, analyte signal can be
64 masked. In the context of bioreactor broth analysis, Raman spectroscopy has significant
65 advantages, such as ease of implementation, ease of use, low maintenance, and high analysis
66 frequency, as an industrial process desires. Sample preparation in many cases is not required,
67 permitting *in-situ* sample analysis. Water has a weak Raman signal and so spectra can be easily
68 collected from aqueous solutions. Raman spectroscopy is generally implemented using
69 excitation sources in the visible to NIR regions of the spectrum which allows for the use of fiber
70 optic probes for remote or *in-situ* analysis [18, 19]. The use of Raman spectroscopy for the
71 analysis of complex systems like in-reactor bioprocess monitoring is a rapidly expanding [20-22].

72 One of the key technologies driving the adoption of Raman (and other multivariate
73 spectroscopic) based methods has been the increased use of chemometrics to extract useful
74 quantitative and qualitative information from data [23]. In the context of quantitative bioreactor
75 broth analysis, chemometrics has generally been used to specifically quantify metabolites or
76 nutrients [8], or more holistically predict the final yield. Partial least-squares regression (PLS)
77 [24, 25] is one of the most important chemometric tools and generally used to develop statistical
78 multivariate regression models within and between large and complex data matrices, and thus to

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

79 facilitate understanding of the important relationships between spectroscopic measurements and
80 the analyte or property of interest. To improve PLS regression performance, many methods have
81 been proposed for selecting the variables that carry higher information content regarding the
82 property of interest from a large number of spectral wavelengths/variables [26-29].

83 Some variable selection methods are based on the inspection of regression coefficients or
84 latent variables [30-32], such as the typical uninformative variable elimination [33, 34], variable
85 importance in projection [35], and competitive adaptive reweighted sampling (CoAdReS) [36].
86 Other methods involve the conduction of the minimum error searches, for example, interval-PLS
87 [37], moving window PLS [38], genetic algorithms [29, 39-41], particle swarm optimization
88 (PSO) [42], and ant colony optimization (ACO) [43, 44]. CoAdReS and ACO variable selection
89 methods both employ a Monte Carlo (MC) strategy to select a limited number of key variables
90 from the multivariate spectral data, and thus generate more accurate chemometric models [36,
91 44]. In this study we have used both methods because they are both better than the other
92 common variable selection methods (*e.g.* genetic algorithms) and second because the two
93 methods used intrinsically different methods of variable selection. Thus analyzing the complex
94 bioprocess derived Raman spectral data using both of these methods should provide a clear
95 indication of model reliability.

96 Here we used Raman spectroscopy to model/monitor a complete fed-batch, CHO cell based
97 process from the initial small, liter-scale right up to the final large-scale (5000L) fermenter.
98 Spectral data was collected from off-line, samples and the productivity of the process was
99 evaluated in terms of glycoprotein product yield. It was possible by the judicious use of
100 computational, variable selection methods to accurately predict process yield with small relative
101 errors of prediction (REP%). This ability to accurately predict final yield at all stages of the
102 process using a single analytical method is highly desirable because it provides a rapid quality
103 assurance tool for optimal operation of large-scale CHO bioreactors.

104
105

106 **2. MATERIALS AND METHODS**

107

108 **2.1 Materials**

109 An industrial bioprocess for the production of a recombinant protein using CHO cells in
110 bioreactors up to 5000L was sampled over a continuous 40+ batch, production campaign. For
111 each production run, the process was sampled at twelve different set time points over the course
112 of the fermentation process. The bioprocess was operated in fed-batch mode using proprietary
113 basal and feed media formulations. Samples were first centrifuged and sterile filtered to remove
114 any whole cells,¹ before being aliquotted under sterile conditions. Samples were then shipped to
115 Ireland at -70°C from the Bristol-Myers Squibb Company, Syracuse, US, with a maximum travel
116 time of two days. Sample temperature integrity was confirmed by the use of electronic

¹ For the sake of clarity we will refer to these specific samples as being bioprocess broths.

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

117 temperature sensors in each shipment. These samples were further aliquotted into smaller
 118 volumes and stored at -70°C. For analysis, the samples were randomly removed from cold
 119 storage and defrosted at room temperature [7].

120

121 **Table 1:** Details of the bioprocess samples used in this study obtained from a continuous 40+
 122 batch production campaign.

123

| Dataset | Bioreactor Content Description. | Bioreactor Volume | Sample size |
|---------|--|-------------------|-------------|
| DS1 | Media Start | 2L | 21 |
| DS2 | Media End | 2L | 17 |
| DS3 | Cells + spent basal media: <i>solutions of cells and spent basal media just prior to transfer to next, larger-sized bioreactor.</i> | 2L | 17 |
| DS4 | Cells + spent & fresh basal media: <i>contain the cells and spent basal (transferred from the previous bioreactor) plus new basal media added to advance process scale-up.</i> | 100-200L | 31 |
| DS5 | Cells + spent basal media: <i>solutions of cells and spent basal media just prior to transfer to next, larger-sized bioreactor.</i> | 100-200L | 31 |
| DS6 | Cells + spent & fresh basal media: <i>contain the cells and spent basal (transferred from the previous bioreactor) plus new basal media added to advance process scale-up.</i> | 1000L | 31 |
| DS7 | Cells + spent basal media: <i>solutions of cells and spent basal media just prior to transfer to next, larger-sized bioreactor.</i> | 1000L | 34 |
| DS8 | Cells + spent basal media | 5000L | 37 |
| DS9 | Cells + spent & fresh basal media: <i>contain the cells and spent basal (transferred from the previous bioreactor) plus new basal media added to advance process scale-up.</i> | 5000L | 29 |
| DS10 | Day 5 Post inoculation | 5000L | 35 |
| DS11 | Day 10 Post inoculation | 5000L | 34 |
| DS12 | Prior to transfer for harvest: <i>this is centrifuged harvest material, i.e., end of production material rich in cells and spent media. Some cells at this stage would have undergone apoptosis and thus expelled host cell protein and other cell debris into the supernatant. In every case, the material that was centrifuged to eliminate whole cell and large cell debris is considered clarified, but does still</i> | 5000L | 33 |

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

*have some cell components present such as membranes,
broken organelles, DNA/RNA, etc..*

124

125 The DS9–12 samples follow the final stages of fermentation up to the harvest point, and
126 during this phase, feed media was also added at specific times. Protein yield (titer) was
127 measured using the following method. The soluble media supernatant, which contains the Fc-
128 fusion protein, was first past over an affinity Protein A column. The captured product (usually
129 greater than 98% recovery) was then eluted by a low pH rinse. The product was then analyzed
130 using a spectrophotometric measurement (A_{280}) with an extinction coefficient of 1.0. The
131 extinction coefficient was determined both by theoretical and experimental amino acid
132 concentration. The protein concentration method is validated to ICH standards and is well
133 within 10% (2–3 % reference standard reproducibility, with EC within 2% of theoretical).

134

135 **2.2 Instrumentation and data collection/analysis**

136 Raman spectra were collected with 785 nm excitation using a RamanStation spectrometer
137 (AVALON Instruments Ltd, Belfast, NI, now acquired by Perkin-Elmer). A laser power of ~70
138 mW at the sample with an exposure time of 2×10 seconds was generally used and spectra were
139 recorded at a resolution of 8 cm^{-1} from 3311 to 250 cm^{-1} [7]. $100 \mu\text{L}$ of sample was pipetted
140 directly into a stainless steel multi-well plate for analysis [45]. For each measurement, a 3×3
141 sampling grid was used, and thus nine Raman spectra were generated. Each sample was
142 measured in triplicate and for each of the three measurements a fresh aliquot of sample was used.
143 Finally, the triplicate measurements were averaged to generate a single spectrum for each sample.
144 Raman data was collected over 38 months and a cyclohexane standard was used to ensure
145 wavelength accuracy during this period. To minimize the effects of baseline drift, scatter effects,
146 and uncontrolled fluctuations, Raman data were subjected to a series of sequential pre-processing
147 steps (baseline removal, normalization to water bending band, water band removal, and first
148 derivative transformation) prior to chemometric modeling. All calculations were performed
149 using MATLAB [46], PLS_Toolbox [47], and in-house-written MATLAB routines
150 (*supplemental information*). ACO MATLAB code was generously provided by Prof. A.C.
151 Olivieri (Universidad Nacional de Rosario, Argentina). See *supplemental information* for
152 sample spectra.

153

154 **2.3 Calibration and validation samples**

155 Twelve datasets (Table 1) were generated from the various samples; however, the first three
156 datasets (DS1–3) had low sample numbers and therefore were not used further. The remaining
157 sets comprised of samples acquired at different time points for the same CHO based process
158 (30+ lots) where each sample set describes a different stage of the process. All data sets were
159 mean-centered prior to PLS or PCA modeling. For PLS modeling, datasets were randomly split
160 into a calibration and test set (always five samples) in a ~80:20 split using an MC based

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

161 sampling protocol. To ensure robustness the calibration/test set selection was repeated 500 times
162 and a PLS model run on each unique selection. PLS model quality was assessed using a
163 combination of parameters including: root mean square error of calibration (RMSEC), root mean
164 square error of prediction (RMSEP) for validation/test set, relative error of prediction
165 ($\text{REP}\% = 100 \times \text{RMSEP} / \bar{y}_{\text{cal}}$, where \bar{y}_{cal} is the mean calibration value of the product titer), and the
166 square of the correlation coefficient (R^2) between predicted and measured titers for the validation
167 set. Finally to avoid potential overfitting, we used a randomization test method to determine the
168 proper number of PLS components to be used for each final model (see *supplemental*
169 *information* for details) [48]. This method enabled a clearer assessment of which components
170 were likely to contribute to overfitting, and resulted in the use of 20–45% fewer components
171 compared to standard cross-validation methods.

172

173 2.4 CoAdReS variable selection

174 CoAdReS was implemented on each individual dataset to select the spectral variables which
175 correlated most strongly with yield. These variables were then used to generate quantitative PLS
176 models (Table 2). 200 CoAdReS sampling runs were performed and for each sampling run, a
177 PLS model was constructed using 83% of the samples, which were randomly selected.
178 CoAdReS then generated sequentially 200 subsets of variables (182 in run 1, only 2 in run 200)
179 and regression coefficients for each variable were obtained from the PLS models. The variable
180 selection process was based first on the magnitude of the regression coefficients, and second on
181 the reduction rate, for example in the i^{th} sampling run, the ratio of variables/wavenumbers to be
182 kept (r_i) is given by: $r_i = ae^{-ki}$ ($a = 1.0234$, $k = 0.0232$, $i = 1, 2, \dots, 200$). Variables with low
183 regression coefficients were weighted to zero, and the significant variables to be retained were
184 weighted with a value related to their absolute regression coefficient value. These retained
185 variables were then used for PLS modeling in the next sampling run, and so on [36]. Once the
186 200 subsets were generated, the remaining samples (17%) were employed for cross validation on
187 each CoAdReS sampling run, and the RMSEP was calculated for this cross validation. The
188 optimal subset of variables (from the 200) is the subset with the lowest RMSEP value.

189 To ensure that we had a robust variable selection procedure, we reran CoAdReS 500 times
190 for each dataset using random calibration/test sample combinations (selected using MC), and as a
191 consequence, the key variables selected varied slightly. All 500 sets of key variables were then
192 statistically analyzed to generate a normalized histogram. To determine the optimal number of
193 the selected variables to be used for the final chemometric model, leave-one-out cross validation
194 [49] PLS modeling was performed with trial numbers of selected variables from 10 to 45. In
195 practice, all selected variables were ranked according to the magnitude of the histogram values
196 from largest to lowest. Then, a number of the selected variables (from 10 to 45) were picked for
197 PLS modeling and RMSEP values calculated. Plotting the RMSEP values *versus* variable
198 number allowed a minimum value to be determined and thus set a threshold limit for the optimal
199 number of selected variables. This rather computationally intensive approach was necessary

200 because of sample complexity, the low sample number per dataset, and because of the very weak
201 analyte bands. However, computational time is relatively inexpensive, so that it is feasible to
202 implement these methods in an industrial context without expensive IT infrastructure.

203
204

205 **3. RESULTS AND DISCUSSION**

206
207

207 **3.1 Spectral analysis**

208 Most of the Raman signal originates from water, with the media component signals being
209 relatively weak for both bioprocess broths and basal media samples (Figures 1 and 2). The O–H
210 stretching band above $\sim 3000\text{ cm}^{-1}$ shows the largest variation which is caused by a variety of
211 factors. Based on our previous experience with cell culture media analysis [7, 45] we omitted,
212 the $3311\text{--}1860\text{ cm}^{-1}$ spectral region from the chemometric analysis. The $400\text{--}250\text{ cm}^{-1}$ spectral
213 region was also excluded from chemometric analysis because it was compromised due to
214 Rayleigh light bleed through from the filters (Figure 1, inset graph) [7, 45]. The water bending
215 bands (1636 and 1364 cm^{-1}) dominate in a large proportion of the fingerprint region, making
216 specific analyte identification difficult (Figure 2). In addition, there are significant baseline
217 fluctuations and intensity variations present in the Raman spectra similar to those previously
218 observed for the media and raw materials used in this process [7, 45]. Most of the significant
219 spectral information is contained in the $1853\text{--}400\text{ cm}^{-1}$ range where we expect to observe bands
220 associated with the components of the media, cell constituents, and the protein product.
221 Providing definitive band assignments was not possible due to a combination of compositional
222 complexity, low Raman resolution, the unknown identity of many of the metabolites, and the
223 confidential nature of the basal media used in the process. In any event, we are seeking to use
224 Raman spectroscopy in a more holistic role rather than a precise diagnostic tool. One should
225 also note that in fed-batch operation the continual addition of fresh basal and feed media as one
226 progresses through (*i.e.* a longitudinal study) the production cycle makes it much more difficult
227 to track specific process changes, as these are more than likely swamped by the addition of
228 media. Thus it is more practical for process monitoring to only consider the changes at fixed
229 time points *i.e.* a cross-sectional approach. The downside of this approach is that one requires
230 access to a sufficient number of good quality samples (20–30 production cycles) in order to
231 extract useful data.

232 Figure 2A shows the normalized Raman spectra of clarified supernatant from the end
233 cultures (cells + spent basal media) of the small-scale bioreactors. When compared to Figure 2B
234 (normalized Raman spectra of clarified supernatant from the starting cultures: cells + spent &
235 fresh basal media) there are no significant differences. The exact formulation of these proprietary
236 media are commercial trade secrets and thus we cannot discuss in detail the origin of the
237 differences between the media, nor assign specific identities to the various spectral bands. Most

238 of the differences in these spectra are due to the increase in cell density and volume and to
239 changes in metabolite concentrations

240 Figure 2C shows the normalized Raman spectra of extracts from bioprocess broths from a
241 single production lot sampled at five time points over the last two bioreactors. The signal quality
242 is relatively good because of the sample preparation method. However, one has to be cautious
243 here with respect to spectral interpretation because a fed-batch strategy is employed, so the
244 chemical composition changes not only because of metabolic activity and protein production, but
245 also with the addition of the feed media. The DS7 material is used to seed the last large scale
246 bioreactor, and the DS9 sample is the seed material plus the newly added basal media used for
247 the final stage bioreactor for production. Thus if we consider the DS9–12 sequence of spectra
248 we can observe significant changes due to the bioprocess itself. DS9 contains exponentially
249 growing cells with spent and new basal media mixed, DS10 is from an exponential cell growth
250 phase with higher mass (10^6 mL^{-1} and viable) spent basal, feed media, DS11 is the stationary cell
251 phase (still viable) with spent feed media, and DS12 is the harvest material (rich in cells and
252 spent media).

253 The major visible changes with process time are the increase in band intensity at 534, 853,
254 1044, and 1413 cm^{-1} (see supplemental information for PCA study). Unfortunately, the
255 compositional complexity of the samples makes it very difficult to unambiguously assign any
256 bands in the spectra apart from water. However, it is quite possible that the 534 cm^{-1} band
257 originates from the nine disulphide bonds present in the product glycoprotein, and thus is a
258 marker for secreted product. The 534 cm^{-1} value is mid-way between the values reported for a
259 variety of similar proteins [50-52]. The identity of the other bands is much less certain. For
260 example, for the 853 cm^{-1} peak, strong bands at this wavenumber appear in both amino acids and
261 sugars and are ascribed to a variety of different vibrational modes [53].

262 However, changes in Raman spectra with process time are difficult to assign to specific
263 components because this difference is convoluted with the variations between the various
264 manufacturing runs, e.g., the lot-to-lot variation is much greater than the time-dependent changes
265 (Figure 3). The first plot shows the variation across 31 lots of a starting culture, DS4, and it is
266 clear that there is a large spectral variation. Most of this will be due to compositional changes,
267 some of which is due to dilution. The dilution with feed media is likely a significant variable
268 because the process has complex feed media criteria in which volume input is related to cell
269 density and growth rate, and thus nutrient consumption. This may be reflected and related to this
270 observation, i.e. some media samples look like they have a higher 1354/1635 band ratio
271 indicating a stronger water band. Similarly broth samples measured just prior to harvest (Figure
272 3B) also shows a lot of spectral variation, and we expect that a significant proportion of this
273 variation may be related to the yield of protein product and the degree of cell viability at harvest.

274
275

276

277 **3.2 Correlation with yield**

278 To correlate Raman spectra with the glycoprotein yield, PLS regression was applied to each
279 individual sample dataset using the pre-processed spectra (Table 2). The calibration models
280 were then validated using the test sets. The optimum number of latent variables (LVs) was
281 determined using Monte Carlo cross-validation [54] and randomization test [48]. These models
282 (using all 182 variables) were poor, $R^2 < 0.4$, RMSEP/RMSEC ratios were between 3.3 and 15.8,
283 and REP % values (8–13%) were high. Interestingly, RMSEC values were low, and thus we
284 surmised that the samples did contain intrinsic information that could be correlated with product
285 yield. However, the informative variables (Raman bands) are effectively swamped by the
286 presence of many bands (from all the other chemical species present) that do not have any
287 correlation with yield. The glycoprotein yield range for these samples is between 0.67–0.92 g L⁻¹
288 [55], while the dissolved solid concentration of the media alone is of the order of ~10–20 g L⁻¹,
289 thus the protein product bands will be very weak. The interference from uninformative variables
290 needs to be eliminated, and thus we needed to consider some strategies for eliminating
291 uninformative spectral data. If one has *a priori* knowledge about the analytes of interest in a
292 complex sample, then one can manually select variables [56], however, in this case the product
293 and samples are much more complex, and it is virtually impossible to definitively assign a
294 particular band to the protein product (apart from the disulphide stretch). Therefore we decided
295 to evaluate two different methods (CoAdReS and ACO) to select informative variables and then
296 use these selected variables for PLS regression.

297

298 **3.3 CoAdReS variable selection**

299 The quantitative PLS models generated using CoAdReS are shown in Table 2. For each
300 model, a normalized histogram (Figure 4A) was generated which showed the selected variables,
301 and then these variables were selectively used to generate the various PLS models, for which the
302 optimum variable number was selected by comparing the RMSEP *versus* variable number plot
303 (Figure 4B). For the example shown, the RMSEP decreased to a minimum of 0.018 g L⁻¹ using
304 15 variables and this corresponded to a histogram threshold of 0.26.

305 The improvement in model quality is dramatic compared to the case where the 1853–400
306 cm⁻¹ range was used. R^2 values are all >0.9, RMSEP:RMSEC ratios are ~2, and the REP%
307 values are low (2.1–3.3%). This large improvement is due to the removal of redundant variables
308 (or more correctly those with low information content relating to product yield). For example a
309 large proportion of the measured Raman signal originates from the glucose and other
310 carbohydrate energy sources which will be present in the highest concentration, and is unlikely
311 to show signal variances that correlate with yield. The high variable reduction factor of ~1 in 10
312 indicates that the vast majority of the Raman signal is as expected not related directly or
313 indirectly to the product yield. It's interesting to note that for both sets of PLS models (Table 2)
314 the RMSEC values are almost identical and the same numbers of LVs are used for each sample

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

315 set. This implies that the variables which has the greatest contribution were present in both
 316 datasets, but that their contribution to the PLS models when the full spectra were used, was
 317 swamped by the mass of irrelevant variables, leading to very poor RMSEP values.

318 In summary, CoAdReS seems to offer a very robust method for generating quantitative
 319 models that can be used to predict product yield at multiple stages over the 30+ day process.
 320 Very important to note is the fact that the sampling time points DS4/6/8 are the starting cultures
 321 for each bioreactor stage (*e.g.* transferred material plus fresh basal media) whereas the DS5/7
 322 samples are the materials prior to transfer that contains both cells and the *spent* media. This is
 323 significant because we have now established two separate yield correlations at the start and end
 324 of the small scale reactor stages. We have already established that for this process it is possible
 325 to correlate changes in feed media composition as observed by fluorescence EEM spectroscopy
 326 with product yield [55]. Thus we need to examine the variables selected to see if there is any
 327 information regarding the nature of the chemical components that give rise to these productivity
 328 correlations, and also whether or not the correlations are due entirely to the media. But first we
 329 need to validate the variable selection by using a different technique to see if the same variables
 330 are selected.

331
 332 **Table 2:** Summary of the PLS models and their performance using full spectral data (1853–400
 333 cm^{-1}), CoAdReS, and ACO selected variables for the 9 different sample sets. Figures are the
 334 mean values obtained from 500 different individual models (see main body text for details). In
 335 each case five samples were used for the test set. RMSEC/RMSEP errors are given in g L^{-1} of
 336 the final protein product titer. Dataset sample size in parentheses varied according to sample
 337 availability. See the *supplemental information* for measured *versus* predicted plots from selected
 338 PLS models.

339

| Data set | Variables selected | PLS Factors | RMSEC | RMSEP | REP% | R^2 |
|----------------------------------|--------------------|-------------|---------------------------------|-------------|-------|-------|
| | | | g L^{-1} protein titer | | | |
| <i>Full spectral data models</i> | | | | | | |
| DS4 (28) | 182 | 4 | 0.020±0.003 | 0.065±0.018 | 7.93 | 0.36 |
| DS5 (28) | 182 | 5 | 0.019±0.003 | 0.068±0.020 | 8.26 | 0.38 |
| DS6 (28) | 182 | 6 | 0.012±0.002 | 0.074±0.025 | 8.76 | 0.33 |
| DS7 (30) | 182 | 5 | 0.022±0.003 | 0.075±0.021 | 9.26 | 0.27 |
| DS8 (31) | 182 | 7 | 0.014±0.002 | 0.104±0.022 | 12.80 | 0.24 |
| DS9 (26) | 182 | 8 | 0.006±0.002 | 0.095±0.020 | 11.72 | 0.22 |
| DS10 (31) | 182 | 7 | 0.016±0.003 | 0.090±0.023 | 10.97 | 0.20 |
| DS11 (30) | 182 | 6 | 0.019±0.003 | 0.107±0.030 | 13.05 | 0.20 |
| DS12 (29) | 182 | 7 | 0.011±0.002 | 0.095±0.021 | 11.75 | 0.20 |
| <i>CoAdReS models</i> | | | | | | |
| DS4 | 15 | 4 | 0.011±0.001 | 0.018±0.007 | 2.15 | 0.965 |

| | | | | | | |
|-------------------|----|---|-------------|-------------|------|-------|
| DS5 | 16 | 5 | 0.021±0.008 | 0.029±0.016 | 3.04 | 0.915 |
| DS6 | 14 | 6 | 0.009±0.001 | 0.018±0.007 | 2.11 | 0.957 |
| DS7 | 18 | 5 | 0.012±0.001 | 0.026±0.009 | 3.08 | 0.932 |
| DS8 | 11 | 7 | 0.015±0.001 | 0.020±0.006 | 2.34 | 0.969 |
| DS9 | 25 | 8 | 0.012±0.002 | 0.024±0.008 | 2.95 | 0.941 |
| DS10 | 15 | 7 | 0.012±0.002 | 0.025±0.012 | 2.83 | 0.916 |
| DS11 | 17 | 6 | 0.014±0.001 | 0.028±0.010 | 3.30 | 0.902 |
| DS12 | 23 | 7 | 0.010±0.001 | 0.022±0.008 | 2.57 | 0.962 |
| <i>ACO models</i> | | | | | | |
| DS4 | 29 | 6 | 0.008±0.001 | 0.020±0.008 | 2.39 | 0.958 |
| DS5 | 26 | 6 | 0.021±0.006 | 0.029±0.013 | 3.25 | 0.919 |
| DS6 | 26 | 5 | 0.013±0.001 | 0.027±0.010 | 3.14 | 0.895 |
| DS7 | 32 | 5 | 0.019±0.002 | 0.038±0.015 | 4.43 | 0.846 |
| DS8 | 26 | 7 | 0.012±0.001 | 0.030±0.010 | 3.60 | 0.901 |
| DS9 | 30 | 5 | 0.013±0.002 | 0.036±0.012 | 4.30 | 0.852 |
| DS10 | 36 | 7 | 0.011±0.002 | 0.038±0.014 | 4.42 | 0.867 |
| DS11 | 23 | 5 | 0.017±0.002 | 0.031±0.011 | 3.74 | 0.935 |
| DS12 | 24 | 6 | 0.010±0.001 | 0.027±0.010 | 3.19 | 0.925 |

340

341

342 3.4 ACO variable selection

343 ACO was used because its basis of refinement is completely different to CoAdReS. ACO
344 was implemented using ρ (rate of pheromone evaporation) =0.65, N (number of ants) =100, w
345 (sensor width) =1, a maximum number of time steps of 50, and 50 repeated MC calculation
346 cycles to build a histogram of variable selection probability. The results (Table 2) reveal that in
347 general the ACO method selected approximately twice as many variables as CoAdReS except
348 for DS9/11/12 where variable numbers are very similar. When the ACO selected variables were
349 used for PLS modeling, the PLS models had the nearly same RMSEP/REP error as the CoAdReS
350 derived models, while the RMSEC errors were essentially the same. Taking DS4 as an example
351 (Figure 4C/D), the histogram generated by the 50 repeated calculation cycles shows the
352 importance assigned to each variable. As with CoAdReS, subsets with 10–45 variables were
353 generated and the data modeled by PLS. The RMSEP reached a minimum (0.017 g L⁻¹) at 29
354 variables (a corresponding threshold value of 0.43) and the selection of additional variables did
355 not improve the model any further.

356 Both the CoAdReS and ACO PLS methods (Table 2) significantly improved predictive
357 ability compared to the full spectrum models, with CoAdReS having a slightly better RMSEP
358 and R² values. This small improvement seems due to the fact that CoAdReS is better at
359 discriminating the good from the bad variables as shown by the green/black discrimination in
360 Figure 4. However, the differences are marginal, and when we consider that both variable

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

361 selection models yield prediction models with similar RMSEC/RMSEP values, similar numbers
362 of LVs, and %REP (Table 2), operationally there is little to separate the methods in terms of
363 predictive ability (for this limited sample number case). The key difference is that ACO is much
364 more time-consuming than CoAdReS, as it took ~200 times longer to run a single iteration using
365 a standard workstation, *i.e.*, 1.2 minutes *versus* 4 hours. In conclusion we would prefer the use
366 of the CoAdReS method for variable selection due to the fact that it is much more suited to rapid
367 analysis.

368

369 **3.5 PLS model quality**

370 One issue which needs to be addressed is the fact that the variable selection method
371 combined with the low sample number can generate PLS models which are overly optimistic
372 because of overfitting, particularly when CV method is used. Here we used the randomization
373 method to ascertain the proper number of PLS components to use, and we found that in
374 comparison to CV method (see *supplemental information*) the number of components was
375 reduced by 20–45%. The resulting models displayed RMSEP/RMSEC ratios that varied from
376 1.4–2.2 for CoAdReS and 1.4–3.5 for ACO which while not ideal, do show robustness of the
377 models. Improving the model quality further would require a doubling or tripling of the sample
378 numbers and unfortunately that is not feasible here at present. However, in the manufacturing
379 domain, one could easily increase the sample numbers year on year and revise/update the model
380 to generate much more robust models.

381

382 **3.6 Analysis of variables selected**

383 While both methods can extract relevant variables and generate good correlations, we now
384 have to consider if there is any useful composition information linked with the selected variables
385 and, more importantly what is the basis for the correlation models in these complex media.
386 Since the principles of operation for CoAdReS (PLS regression coefficients) and ACO
387 (minimum error search) are intrinsically different one expects that the selected variables will be
388 different, but that any common variables might be expected to be the ones with the greatest
389 correlation with process yield. Thus by looking at these common variables (Table S-4,
390 *supplemental information*) we could get some indication as to which molecular species may be
391 of significance.

392 Since the DS5/7/8 samples are the cells and spent media before new basal media has been
393 added, the variables selected should represent the key species in the spent media that correlate
394 with the final yield. The fact that the variables are very different in each case may indicate that
395 the important metabolites changes as the process scales up. When we next consider the DS4/6/9
396 samples where the fresh basal media is added, we see that the selected variables change very
397 significantly, indicating that the correlated bands are more likely to now be related to the new
398 basal media. This is not surprising since we have seen this type of process yield correlation to
399 the media variation of a feed before (actually the feed media used in this process), and using

400 fluorescence we were able to generate a predictive model [55]. At DS12, the final sampling time
401 point (just prior to harvest) should contain significant amounts of the glycoprotein product, the
402 protein product concentration should be relatively high ($0.67\text{--}0.92\text{ g L}^{-1}$) [55] and one might
403 expect that some of the selected variables/bands should be clearly related to protein bands of the
404 product. The variables selected here are very different from the preceding time points with what
405 looks like two clusters of significant variables in the $1600\text{--}1300\text{ cm}^{-1}$ and the $1250\text{--}920\text{ cm}^{-1}$
406 ranges. However, at this stage there is also an appreciable host cell proteins (HCP) concentration
407 (possibly $100\text{--}200\text{ mg L}^{-1}$) [57] which will be virtually indistinguishable from the glycoprotein
408 antibody in these complex samples. Thus, unfortunately, it is not feasible, at this stage to assign
409 these variables unambiguously to specific compounds using this low resolution, low signal-to-
410 noise quality Raman data.

411

412

413 4. CONCLUSIONS

414

415 Conventional Raman spectroscopy coupled with variable selection and standard PLS
416 modeling is an effective and inexpensive method for the quantitative characterization of
417 mammalian cell culture process in terms of product yield. We have shown the feasibility of
418 predicting product yield from the very early stages of the manufacturing process right through to
419 the final large-scale bioreactor. The use of clarified bioreactor supernatant in an off-line method
420 provides a good quality set of samples where scattering artifacts are minimized, thus generating
421 more reproducible spectral data. The key limitation is the inability to precisely identify the
422 molecular species that correlate most strongly with process yield. The variation in the selected
423 variables, indicate that at each process point the species which correlate most strongly with yield
424 change. For the starting cultures of each bioreactor, it may be that the correlation is linked to
425 specific media components. However, from the later stages (*i.e.* the cells and spent media) the
426 selected variables could be from metabolites and host cell proteins (secondary indicators) or the
427 glycoprotein (primary indicator). This then is a fundamental limitation of this low resolution (8
428 cm^{-1}) Raman method.

429 These results coupled with our previous work on cell culture feed media [55] are very
430 significant from an industrial standpoint because they suggest that one could design in
431 appropriate control measures to implement an effective quality assurance programme for
432 complex media and CHO based manufacturing using these Raman based methods. Furthermore,
433 if the appropriate calibration models are available [8] then one could also incorporate
434 quantitative measurements for a variety of specific components (*e.g.* glutamine, glucose, lactate)
435 at the same time. Thus a single Raman measurement can deliver multitude outputs which can be
436 used to control bioprocess operations.

437

438

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

439 **ACKNOWLEDGEMENTS**

440
441 This research was undertaken as part of the Centre for BioAnalytical Sciences, funded by
442 the IDA and BMS. We thank Ms. Lindy Smith (BMS) for her work in sample organization.
443

444 **REFERENCES**

- 445
446 [1] M. Rhiel, P. Ducommun, I. Bolzonella, I. Marison, U. von Stockar, *Biotechnol. Bioeng.*,
447 77 (2002) 174.
448 [2] S.A. Arnold, J. Crowley, N. Woods, L.M. Harvey, B. McNeill, *Biotechnol. Bioeng.*, 84
449 (2003) 13.
450 [3] P. Roychoudhury, L.M. Harvey, B. McNeil, *Anal. Chim. Acta*, 571 (2006) 159.
451 [4] T. Becker, B. Hitzmann, K. Muffler, R. Portner, K.F. Reardon, F. Stahl, R. Ulber, *White*
452 *Biotechnology (Advances In Biochemical Engineering / Biotechnology)*, Springer-Verlag Berlin,
453 Berlin, 2007, p. 249.
454 [5] C. Card, B. Hunsaker, T.A. Smith, J. Hirsch, *Bioprocess Int.*, 6 (2008) 58.
455 [6] A.E. Cervera, N. Petersen, A.E. Lantz, A. Larsen, K.V. Gernaey, *Biotechnol. Prog.*, 25
456 (2009) 1561.
457 [7] B. Li, P.W. Ryan, B.H. Ray, K.J. Leister, N.M.S. Sirimuthu, A.G. Ryder, *Biotechnol.*
458 *Bioeng.*, 107 (2010) 290.
459 [8] N.R. Abu-Absi, B.M. Kenty, M.E. Cuellar, M.C. Borys, S. Sakhamuri, D.J. Strachan,
460 M.C. Hausladen, Z.J. Li, *Biotechnol. Bioeng.*, 108 (2011) 1215.
461 [9] T. Chattaway, G.A. Montague, A.J. Morris, in: H.J. Rehm, G. Reed (Eds.), *Bioprocessing,*
462 *Biotechnology*, Wiley-VCH Verlag GmbH, Weinheim, Germany., 2008, p. 319.
463 [10] A.S. Rathore, R. Bhambure, V. Ghare, *Anal. Bioanal. Chem.*, 398 (2010) 137.
464 [11] J. Glassey, K.V. Gernaey, C. Clemens, T.W. Schulz, R. Oliveira, G. Striedner, C.F.
465 Mandenius, *Biotechnol. J.*, 6 (2011) 369.
466 [12] T. Cartwright, G.P. Shah, in: J.M. Davis (Ed.), *Basic Cell Culture*, Oxford University
467 Press Inc., New York 2002, p. 69.
468 [13] D.J. Newman, G.M. Cragg, *J. Nat. Prod.*, 70 (2007) 461.
469 [14] L. Olsson, U. Schulze, J. Nielsen, *Trac-Trends Anal. Chem.*, 17 (1998) 88.
470 [15] K.N. Baker, M.H. Rendall, A. Patel, P. Boyd, M. Hoare, R.B. Freedman, D.C. James,
471 *Trends Biotechnol.*, 20 (2002) 149.
472 [16] S.A. Bradley, A. Ouyang, J. Purdie, T.A. Smitka, T. Wang, A. Kaerner, *J. Am. Chem.*
473 *Soc.*, 132 (2010) 9531.
474 [17] A. Hashimoto, A. Yamanaka, M. Kanou, K. Nakanishi, T. Kameoka, *Bioprocess. Biosyst.*
475 *Eng.*, 27 (2005) 115.
476 [18] P. Marteau, F. Adar, N. ZanierSzydowski, *Am. Lab.*, 28 (1996) H21.
477 [19] U. Utzinger, R.R. Richards-Kortum, *J Biomed Opt*, 8 (2003) 121.

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

- 478 [20] A.C. McGovern, D. Broadhurst, J. Taylor, N. Kaderbhai, M.K. Winson, D.A. Small, J.J.
479 Rowland, D.B. Kell, R. Goodacre, *Biotechnol. Bioeng.*, 78 (2002) 527.
- 480 [21] C. Cannizzaro, M. Rhiel, I. Marison, U. von Stockar, *Biotechnol. Bioeng.*, 83 (2003) 668.
- 481 [22] H.L.T. Lee, P. Boccazzi, N. Gorret, R.J. Ram, A.J. Sinskey, *Vib. Spectrosc.*, 35 (2004)
482 131.
- 483 [23] T.J. Vickers, C.K. Mann, *Quantitative analysis by Raman spectroscopy.*, New York:
484 John Wiley and Sons, Inc., 1991.
- 485 [24] S. Wold, M. Sjöström, L. Eriksson, *Chemometr. Intell. Lab. Syst.*, 58 (2001) 109.
- 486 [25] M. Andersson, *J. Chemometr.*, 23 (2009) 518.
- 487 [26] J.P. Gauchi, P. Chagnon, *Chemometr. Intell. Lab. Syst.*, 58 (2001) 171.
- 488 [27] R.K.H. Galvao, M.C.U. Araujo, in: S.D. Brown, R. Tauler, B. Walczak (Eds.),
489 *Comprehensive Chemometrics*, Elsevier Amsterdam, 2009, p. 233.
- 490 [28] X.B. Zou, J.W. Zhao, M.J.W. Povey, M. Holmes, H.P. Mao, *Anal. Chim. Acta*, 667
491 (2010) 14.
- 492 [29] N. Sorol, E. Arancibia, S.A. Bortolato, A.C. Olivieri, *Chemometr. Intell. Lab. Syst.*, 102
493 (2010) 100.
- 494 [30] A.J. Burnham, J.F. MacGregor, R. Viveros, *J. Chemometr.*, 15 (2001) 265.
- 495 [31] R.F. Teofilo, J.P.A. Martins, M.M.C. Ferreira, *J. Chemometr.*, 23 (2009) 32.
- 496 [32] C.D. Brown, R.L. Green, *Trac-Trends Anal. Chem.*, 28 (2009) 506.
- 497 [33] V. Centner, D.L. Massart, O.E. deNoord, S. deJong, B.M. Vandeginste, C. Sterna, *Anal.*
498 *Chem.*, 68 (1996) 3851.
- 499 [34] Q.J. Han, H.L. Wu, C.B. Cai, L. Xu, R.Q. Yu, *Anal. Chim. Acta*, 612 (2008) 121.
- 500 [35] I.G. Chong, C.H. Jun, *Chemometr. Intell. Lab. Syst.*, 78 (2005) 103.
- 501 [36] H.D. Li, Y.Z. Liang, Q.S. Xu, D.S. Cao, *Anal. Chim. Acta*, 648 (2009) 77.
- 502 [37] L. Norgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, S.B. Engelsen, *Appl.*
503 *Spectrosc.*, 54 (2000) 413.
- 504 [38] J.H. Jiang, R.J. Berry, H.W. Siesler, Y. Ozaki, *Anal. Chem.*, 74 (2002) 3555.
- 505 [39] R. Leardi, A.L. Gonzalez, *Chemometr. Intell. Lab. Syst.*, 41 (1998) 195.
- 506 [40] H.C. Goicoechea, A.C. Olivieri, *J. Chem. Inf. Comp. Sci.*, 42 (2002) 1146.
- 507 [41] H.C. Goicoechea, A.C. Olivieri, *J. Chemometr.*, 17 (2003) 338.
- 508 [42] L. Xu, J.H. Jiang, H.L. Wu, G.L. Shen, R.Q. Yu, *Chemometr. Intell. Lab. Syst.*, 85 (2007)
509 140.
- 510 [43] M. Shamsipur, V. Zare-Shahabadi, B. Hemmateenejad, M. Akhond, *J. Chemometr.*, 20
511 (2006) 146.
- 512 [44] F. Allegrini, A.C. Olivieri, *Anal. Chim. Acta*, 699 (2011) 18.
- 513 [45] A.G. Ryder, J. De Vincentis, B.Y. Li, P.W. Ryan, N.M.S. Sirimuthu, K.J. Leister, *J.*
514 *Raman Spectrosc.*, 41 (2010) 1266.
- 515 [46] Mathworks Inc., Cambridge, MA, 1994-2008.
- 516 [47] Eigenvector Research Inc.: , 3905 West Eaglerock Drive, Wenatchee, WA.

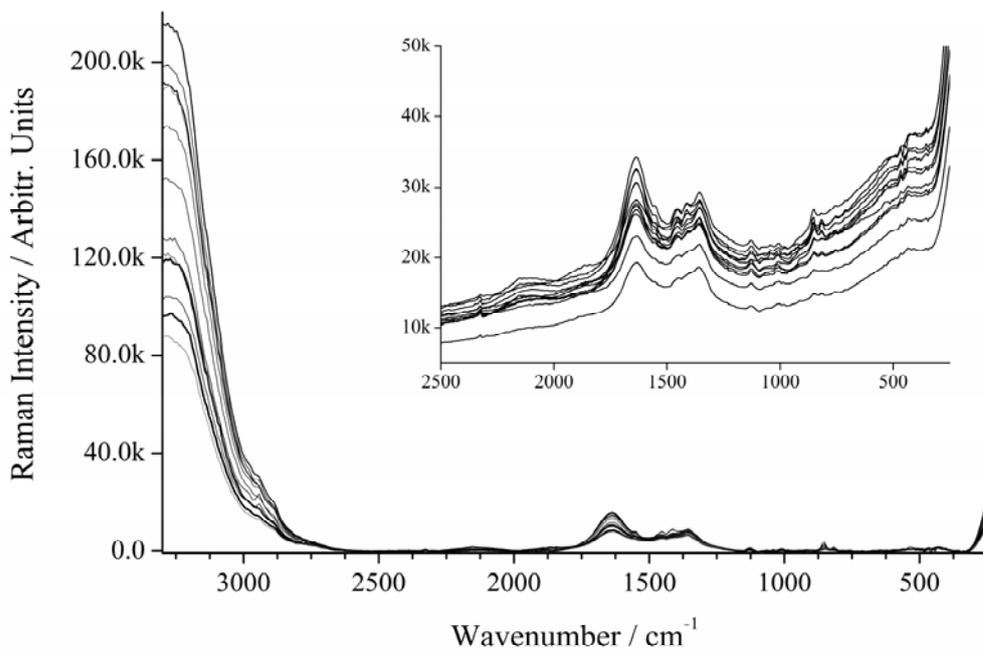
Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

- 517 [48] S. Wiklund, D. Nilsson, L. Eriksson, M. Sjostrom, S. Wold, K. Faber, J. Chemometr., 21
518 (2007) 427.
- 519 [49] D.M. Haaland, E.V. Thomas, Anal. Chem., 60 (1988) 1193.
- 520 [50] T. Kitagawa, T. Azuma, K. Hamaguchi, Biopolymers, 18 (1979) 451.
- 521 [51] C. David, S. Foley, C. Mavon, M. Enescu, Biopolymers, 89 (2008) 623.
- 522 [52] Z.Q. Wen, J. Pharm. Sci., 96 (2007) 2861.
- 523 [53] J. De Gelder, K. De Gussem, P. Vandenabeele, L. Moens, J. Raman Spectrosc., 38 (2007)
524 1133.
- 525 [54] Q.-S. Xu, Y.-Z. Liang, Chemometrics Intell. Lab. Syst., 56 (2001) 1.
- 526 [55] P.W. Ryan, B. Li, M. Shanahan, K.J. Leister, A.G. Ryder, Anal. Chem., 82 (2010) 1311.
- 527 [56] A.G. Ryder, J. Forensic Sci., 47 (2002) 275.
- 528 [57] A.S. Tait, C.E.M. Hogwood, C.M. Smales, D.G. Bracewell, Biotechnol. Bioeng., 109
529 (2012) 971.
- 530
- 531

Performance Monitoring of a Mammalian Cell Based Bioprocess using Raman spectroscopy. B. Li, B.H. Ray, K.J. Leister, and A.G. Ryder. *Analytica Chimica Acta*, 796, 84-91, (2013). DOI: <http://dx.doi.org/10.1016/j.aca.2013.07.058>

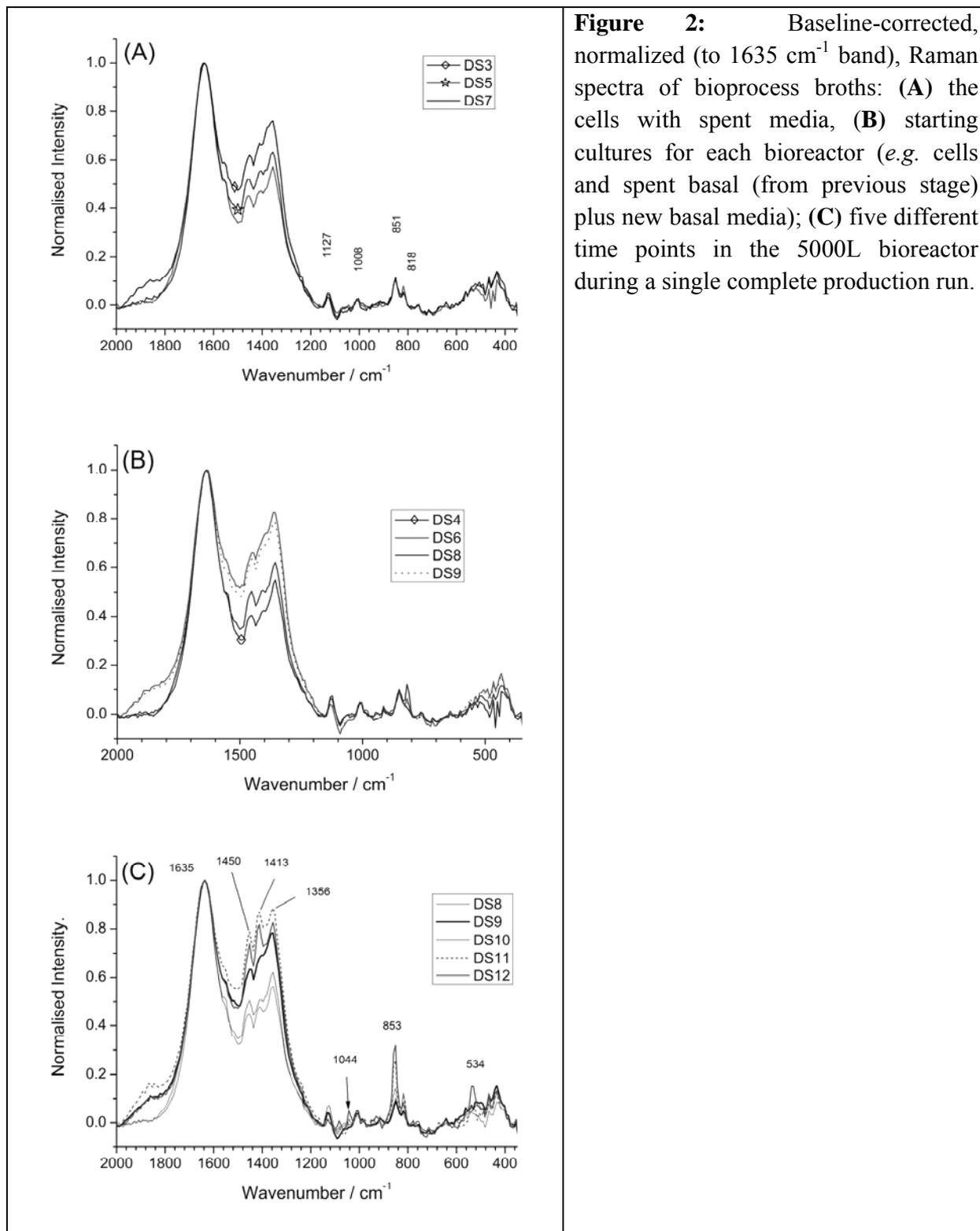
532 **FIGURES**

533



534

535 **Figure 1:** Raman spectra collected from bioprocess broth samples over the full spectral range.
536 Inset shows the low wavenumber range and the variation induced by excitation light bleed
537 through.



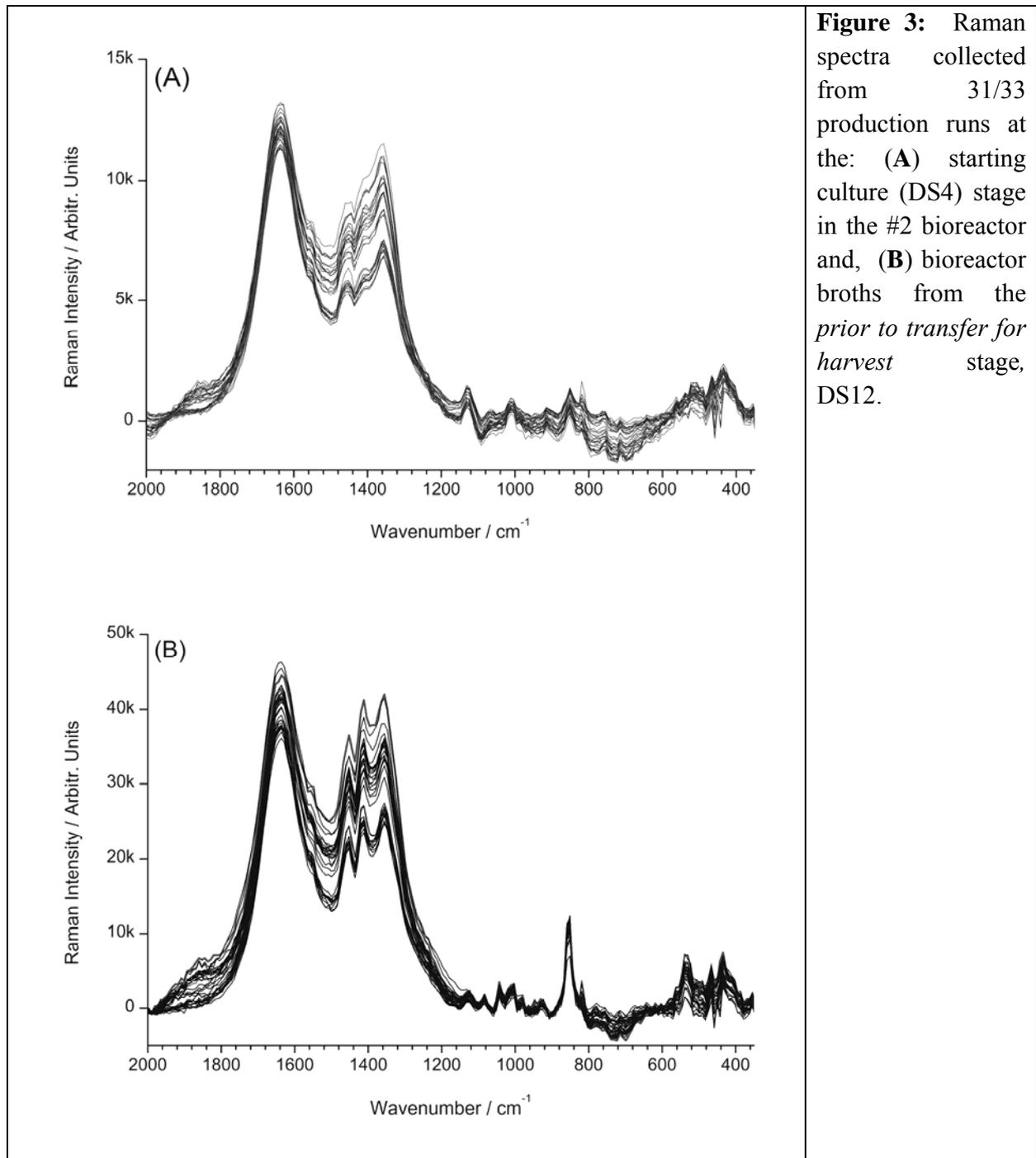
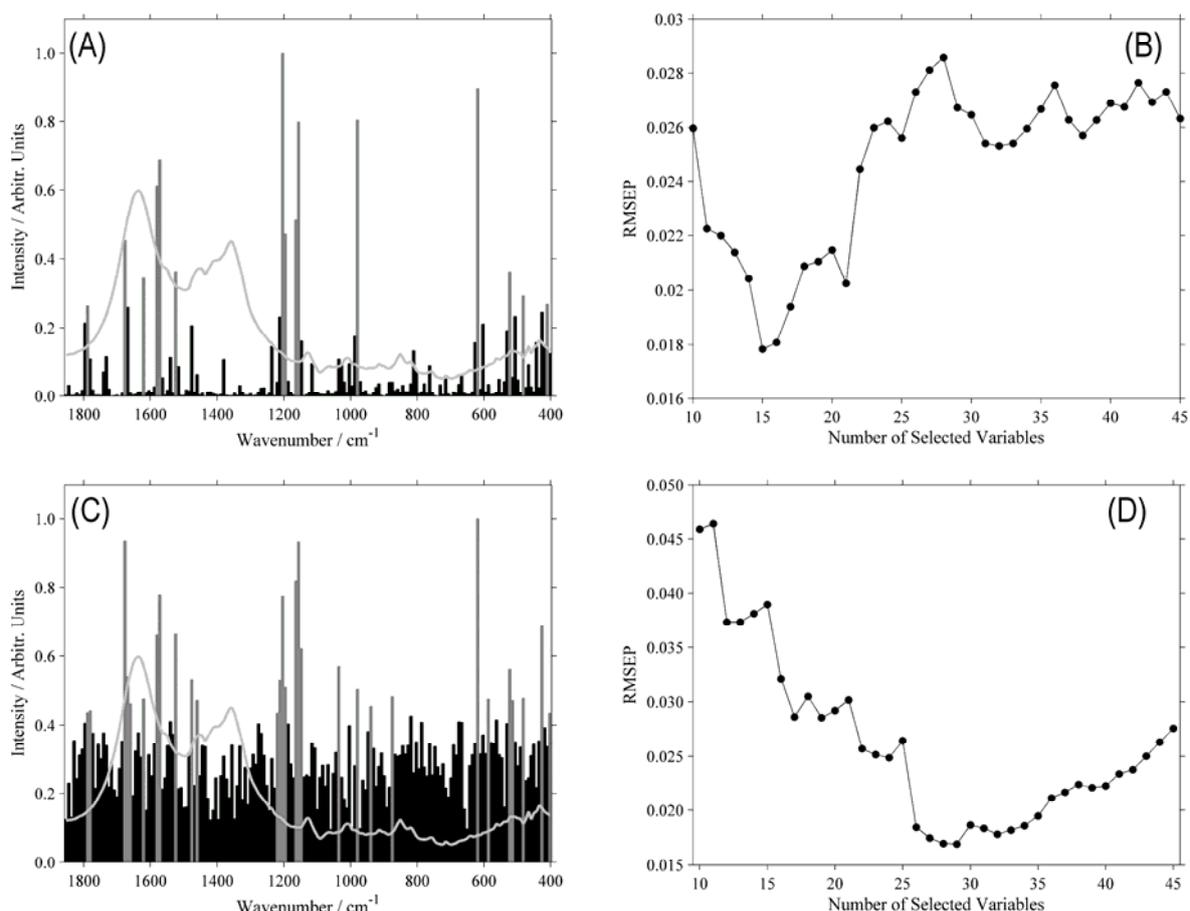


Figure 3: Raman spectra collected from 31/33 production runs at the: (A) starting culture (DS4) stage in the #2 bioreactor and, (B) bioreactor broths from the prior to transfer for harvest stage, DS12.



540
541 **Figure 4:** (A) CoAdReS variable selection result for DS4 (Histogram values, Grey ≥ 0.26 , Black
542 < 0.26). Superimposed is the mean baseline-corrected Raman spectrum (light grey trace,
543 arbitrary vertical scale). (B) Determination of number of the selected variables. (C) ACO
544 variable selection result for DS4 (Histogram values, Grey ≥ 0.43 , Black < 0.43). Superimposed is
545 the mean baseline-corrected Raman spectrum (arbitrary vertical scale). (D) Determination of
546 number of the selected variables.
547