

# Raman based chemometric model development for glycation and glycosylation real time monitoring in a manufacturing scale CHO cell bioreactor process

Luke A. Gibbons<sup>1,2</sup>  | Carl Rafferty<sup>1</sup>  | Kerry Robinson<sup>3</sup> | Marta Abad<sup>4</sup> | Francis Maslanka<sup>4</sup> | Nikky Le<sup>4</sup> | Jingjie Mo<sup>3</sup> | Kevin Clark<sup>4</sup> | Fiona Madden<sup>1</sup> | Ronan Hayes<sup>1</sup> | Barry McCarthy<sup>1</sup> | Christopher Rode<sup>4</sup> | Jim O'Mahony<sup>2</sup> | Rosemary Rea<sup>2</sup> | Caitlin O'Mahony Hartnett<sup>1</sup>

<sup>1</sup>BioTherapeutics Development, Janssen Sciences Ireland UC, Cork, Ireland

<sup>2</sup>Department of Biological Sciences, Munster Technological University, Cork, Ireland

<sup>3</sup>Analytical Development, Janssen Pharmaceutical Companies of Johnson and Johnson, Malvern, Pennsylvania, USA

<sup>4</sup>BioTherapeutics Development, Janssen Pharmaceutical Companies of Johnson and Johnson, Malvern, Pennsylvania, USA

## Correspondence

Luke Gibbons, Biotherapeutics Development, Janssen Sciences Ireland UC, Barnahely, Cork P43 FA46, Ireland.

Email: lucasmacgiobuin@gmail.com

**Funding information**  
Irish Research Council

## Abstract

The Quality by Design (QbD) approach to the production of therapeutic monoclonal antibodies (mAbs) emphasizes an understanding of the production process ensuring product quality is maintained throughout. Current methods for measuring critical quality attributes (CQAs) such as glycation and glycosylation are time and resource intensive, often, only tested offline once per batch process. Process analytical technology (PAT) tools such as Raman spectroscopy combined with chemometric modeling can provide real time measurements process variables and are aligned with the QbD approach. This study utilizes these tools to build partial least squares (PLS) regression models to provide real time monitoring of glycation and glycosylation profiles. In total, seven cell line specific chemometric PLS models; % mono-glycated, % non-glycated, % GOF-GlcNac, % G0, % GOF, % G1F, and % G2F were considered. PLS models were initially developed using small scale data to verify the capability of Raman to measure these CQAs effectively. Accurate PLS model predictions were observed at small scale (5 L). At manufacturing scale (2000 L) some glycosylation models showed higher error, indicating that scale may be a key consideration in glycosylation profile PLS model development. Model robustness was then considered by supplementing models with a single batch of manufacturing scale data. This data addition had a significant impact on the predictive capability of each model, with an improvement of 77.5% in the case of the G2F. The finalized models show the capability of Raman as a PAT tool to deliver real time monitoring of glycation and glycosylation profiles at manufacturing scale.

## KEY WORDS

chemometrics, multivariate data analysis, Raman spectroscopy

## 1 | INTRODUCTION

The intensification of manufacturing in the biopharmaceutical industry has been mainly attributed to the growing use of monoclonal antibody

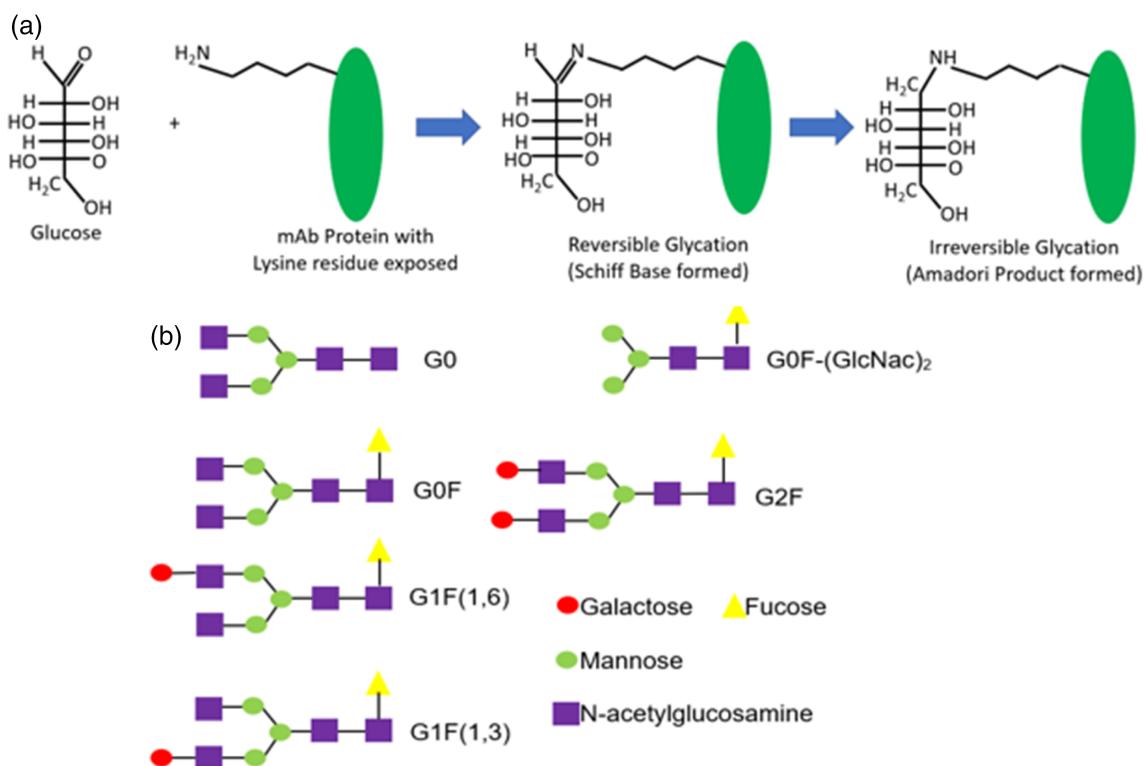
(mAb) based therapies.<sup>1</sup> These mAb therapies are produced in complex conditions by genetically engineered mammalian cells such as Chinese hamster ovary (CHO) cells.<sup>2</sup> Biological activity of recombinant proteins produced by CHO cells is dependent on several factors

including protein secondary structure and post translational modifications, such as glycosylation and glycation (non-enzymatic glycosylation).<sup>3,4</sup> Additionally, glycosylation of mAbs can influence the proteins' aggregation, solubility and stability both in vitro and in vivo.<sup>5</sup> Hence, glycation and glycosylation have been deemed critical quality attributes (CQAs) to be considered in the production of therapeutic proteins.<sup>6</sup> Glycation is a non-enzymatic glycosylation of mAb products governed by reducing sugar content in the bioreactor culture.<sup>7</sup> Glycation occurs according to the Maillard reaction<sup>8</sup> where reducing sugars, for example, Glucose interact with the side chains of lysine residues or the N-termini of proteins reversibly forming a Schiff base. This reversible reaction can progress further leading to the irreversible formation of a ketoamine via Amadori rearrangement<sup>9</sup> (Figure 1a). Glycosylation is a complex post translation modification which involves attachment of glycans to specific sites on the mAb structure. Here we consider N-linked glycosylation which involves attachment of glycans at the Asn-X-Ser/Thr sequence of Fc portion of mAb heavy chains, where X can be any amino acid except for proline.<sup>10</sup> A number of N-glycans exist which are capable of binding to the mAb molecule. These N-glycans include G0, G0F, G0F-GlcNac, G1F, and G2F, which differ slightly in structure (Figure 1b). The interactions between the translated protein structures and the glycans which can be bound to them are governed by multiple factors both intracellularly and extracellularly making it nearly impossible to replicate the batch to batch glycoform content in a production setting.<sup>11</sup> Instead, acceptable limits need to be determined during process development for these therapeutics to lessen the amount of glycoform variability in commercial

production.<sup>12</sup> Once in commercial production, CQA monitoring, and control are essential to ensure suitable product quality can be achieved.

Current methods for measuring glycation and glycosylation profiles include Boronate affinity chromatography, capillary isoelectric focusing colorimetric assays and liquid chromatography mass spectrometry (LC/MS).<sup>13</sup> LC/MS is the current method employed in Janssen Sciences Ireland UC (JSI). These methods, while capable of providing precise and accurate results are both time and resource consuming. The level of information gained is dependent on the number and frequency of samples taken from the bioreactor. A standard sampling plan for bioreactors in JSI is once per 24 h, more frequent sampling increases the available information but comes at the expense of product being removed from the bioreactor and a greater risk of contamination. In recent years there has been a trend to move away from product quality testing once a batch has been completed in favor of a Quality by Design (QbD) approach. Regulatory bodies such as the US Food and Drug Administration (FDA) promote the QbD approach as it focuses on the development of production processes where product quality is understood throughout, rather than confirmatory testing by retrospective study methods.<sup>14</sup>

To achieve this QbD approach, process analytical technology (PAT) is often implemented in the production process. PAT tools are used to control and monitor the production process in-line or at-line in real time. Spectroscopic techniques such as near-infrared spectroscopy, fluorescence spectroscopy, and Raman spectroscopy have potential for use in biotherapeutic production processes as PAT



**FIGURE 1** (a) Maillard reaction for the glycation of mAbs (b) Common N-Glycan structures associated with N-linked glycosylation of mAbs

tools.<sup>15–17</sup> While each technology has its merits, Raman spectroscopy presents itself as a technology particularly suited for use in bioreactor production processes as it provides clear, sharp spectra without suffering from some of the disadvantages of other technologies, for example, water interference in near-infrared spectroscopy. Raman spectroscopy is a vibrational spectroscopic technique which uses laser technology to provide a chemical fingerprint of a substance.<sup>18</sup> Moreover, it has been used as a PAT tool to provide non-destructive real time measurement of a number of biotherapeutic process variables including metabolites, growth profiles, product levels, product quality attributes, nutrient feeds and more recently, culture pH.<sup>15,19–21</sup> By combining Raman spectroscopic analysis with Chemometric modeling, the real time monitoring of these variables is achieved. The spectral peaks obtained by Raman spectroscopy are associated with the process variable of interest, measured by offline analysis, using chemometric modeling software such as SIMCA (Sartorius-Stedim, Germany). Partial Least Squares (PLS) regression models are often employed to model Raman data as PLS modeling suits the linear nature of Raman signal versus analyte concentration.<sup>22</sup>

Real time CQA analysis using Raman spectroscopy is a concept which would be fully aligned with the QbD approach to biotherapeutics manufacturing. Departing from current methods for CQA analysis in favor of an inline approach could fundamentally change how a commercial manufacturing process is controlled within its design space to ensure consistent quality and eliminate unwanted batch to batch variations.<sup>23</sup>

Manufacturing process development with this QbD approach to CQA monitoring embedded would allow for a more quality focused process design and implementation. The long term benefit of having real time CQA data introduces potential process improvements, at all stages of development through to commercial manufacturing, in a number of ways including a more informed process characterization, the introduction of real time product quality control strategies, more effective root cause investigation and the potential for implementing real time release of batches.<sup>24</sup> The potential benefits of this type of approach to CQA monitoring make it an alluring option from a manufacturing perspective.

Studies employing the off-line measurement of small proteins<sup>25</sup> and in-line measurement of large mAb proteins in laboratory scale bioreactors<sup>20</sup> show the potential for Raman spectroscopy as a PAT tool for the measurement of the glycosylation status of proteins. To the best of the authors knowledge, the real time monitoring of glycation and glycosylation profiles of mAbs in manufacturing scale bioreactors has not been described. The complex environmental conditions in a production scale bioreactor, coupled with the intricate nature of glycoprotein formation means that careful consideration must be made in the development of Raman based PLS models.

The purpose of this study was to develop PLS models of Raman spectra for real time monitoring of glycation and glycosylation (both CQAs), in a representative CHO cell culture at manufacturing scale. Model development from small scale data was initially assessed. Developing model robustness was then considered by supplementing the small scale data with manufacturing scale data. In order to fully

follow a QbD approach, product quality needs to be considered throughout the manufacturing process of a biotherapeutic mAb product. This study explores the capability of Raman Spectroscopy to achieve this by monitoring CQAs in real time throughout the upstream production process.

## 2 | MATERIALS AND METHODS

### 2.1 | Bioreactor operation

A total of 18 batches of a mAb-producing CHO cell line were used in the experimental design. The stirred bioreactor scales were single use bag (SUB) 2000 L (Thermofisher Scientific, Waltham, MA), glass 5 L (Applikon Inc., Schiedam, Netherlands) and glass 1 L (Eppendorf, Hamburg, Germany) and the fed-batch process lasted for 16–18 days. In total, nine 1 L batches, seven 5 L and two 2000 L batches were included. Each 2000 L batch was inoculated on day 0 using a commercial seed train, 10 small scale bioreactors were inoculated using day 1 cell culture from a 2000 L batch and six small scale bioreactors were inoculated on day 0 using a laboratory seed train. Proprietary basal media and daily feed media were used for each batch executed, and process controls for all were implemented as per standard practices at JSI. Two feeding strategies were employed during execution of the bioreactor runs, each consisted of two complex feeds, starting on day 3 to each bioreactor. A total of 12 batches ( $2 \times 2000$  L,  $7 \times 5$  L,  $3 \times 1$  L) were fed both complex feeds daily based upon a defined percentage of the vessel volume. The other six batches ( $6 \times 1$  L) were fed the first complex feed based upon a defined percentage of the vessel and the second complex feed was delivered multiple times in a day to maintain a predefined target level of glucose in the bioreactor as part of a feed strategy study. All bioreactors were inoculated within the same seeding density limits (proprietary information). Each batch had identical control targets for dissolved oxygen (DO), pH, and temperature. DO was controlled to 40% by aeration and sparged oxygen. A pH target of 6.95 was maintained using addition(s) of carbon dioxide and 2.0 M sodium carbonate. Temperature was controlled to the set point of 36.5°C throughout the cell culture. The scale-dependent process parameter agitation was transferred across scales using power per volume- calculated values. Offline samples were collected daily from each bioreactor and measured for a panel of metabolites. Each daily culture sample was then retained and frozen at –70°C for further testing once each batch had been completed.

### 2.2 | Glycation and glycosylation analysis

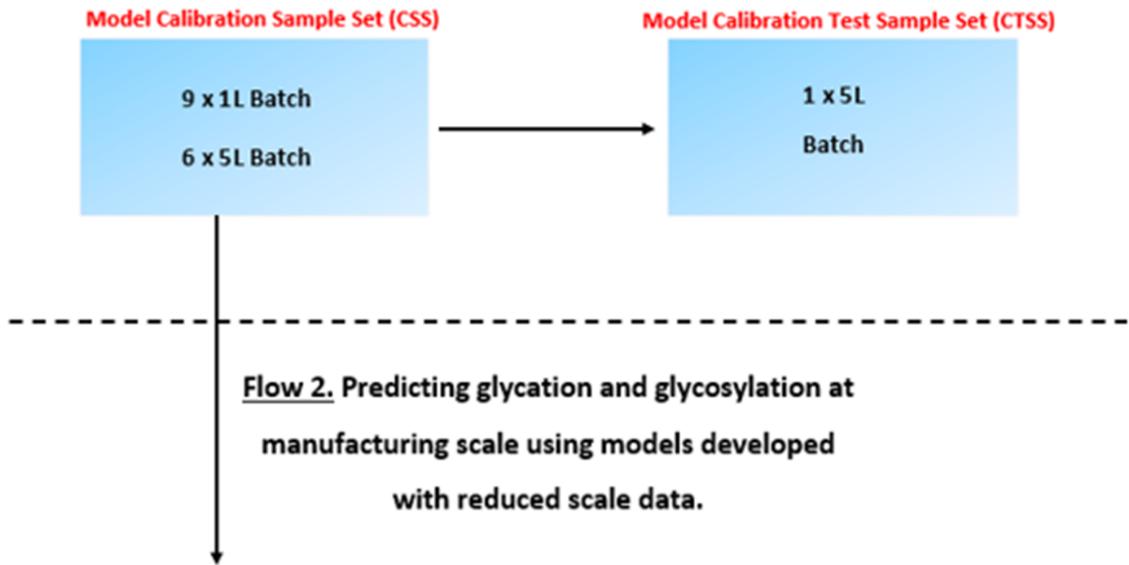
Daily culture samples from day 5 to day 16/18 for each of the 18 bioreactor batches were thawed for product quality analysis of the mAbs produced throughout the batch. Both glycation and glycosylation of the mAbs was characterized offline by LC/MS analysis.

All samples were initially purified using protein A purification as described previously.<sup>25</sup> Protein samples for glycation analysis were

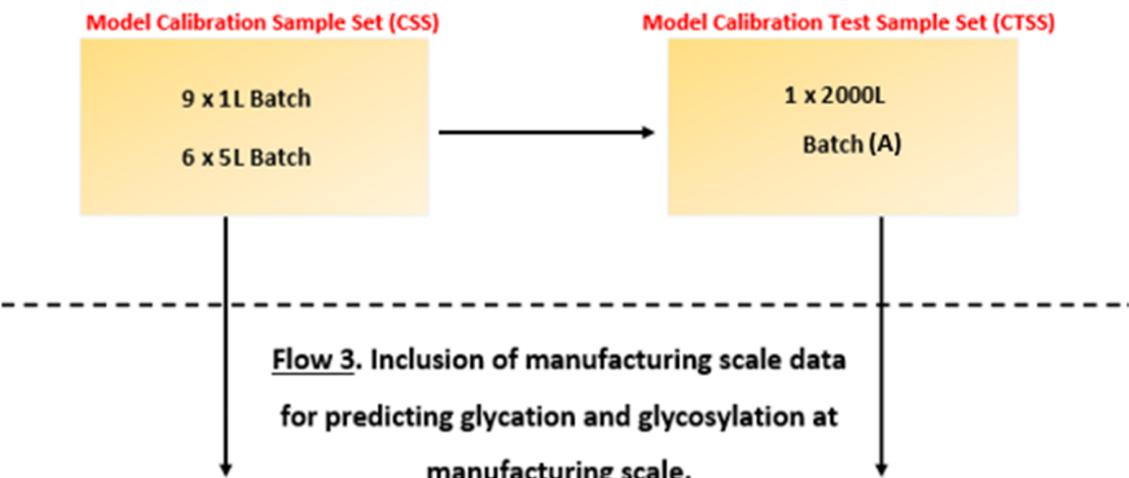
then pretreated with EndoS enzyme to remove N-linked carbohydrates in order to eliminate glycoform based sample heterogeneity. The protein samples were then separated by Ultra-High Performance Liquid Chromatography (HPLC) using Waters Acquity UPLC system (Waters Corp, Milford, MA) on a reverse phase column using a gradient of acetonitrile with trifluoroacetic acid and analyzed with a Xevo G2-XS Mass Spectrometer (Waters Corp, Milford, MA) by online

electrospray ionization quadrupole time-of-flight mass spectrometry. Mass/charge data collected across the chromatographic peak was then summed and deconvoluted using Masslynx (Agilent Technologies, Santa Clara, CA) or UNIFI software (Waters Corp, Milford, MA). Detected glycation isoforms were assigned based on deconvoluted mass spectra analysis and their relative abundances were calculated using peak intensities of centered deconvoluted mass spectra.

### **Flow 1. Development of Raman Spectroscopy based PLS models for glycation and glycosylation at reduced scale.**



### **Flow 2. Predicting glycation and glycosylation at manufacturing scale using models developed with reduced scale data.**



**FIGURE 2** Process flow for the development of and testing of Raman based PLS models

Protein samples for glycosylation were first pretreated with 1 M Dithiothreitol to separate individual mAb heavy and light chains. The protein samples were then analyzed as before by LC/MS and the glycosylation isoforms assigned and quantified as described previously.

### 2.3 | Raman spectral acquisition during bioreactor batches

Raman spectra were gathered using a multichannel Raman RXN2 system (Kaiser Optical Systems Inc., Ann Arbor, MI), which contained a 785-nm laser source and a charge-coupled device (CCD) at  $-40^{\circ}\text{C}$ . The detector was connected to an MR probe, which consisted of a fiber optic excitation cable and a fiber optic collection cable (Kaiser Optical Systems Inc.). Two instruments were used in data collection for this work. Data was collected by the MR probe attached to a bIO-Optic-220 stainless steel probe (Kaiser Optical Systems Inc.) inserted in the sterile bioreactor. The iCRaman software 4.1 (Mettler Toledo Autochem, Columbia, MD) was used to control the Raman RXN2 for spectral acquisition for all small scale batches (1 and 5 L). The Raman Runtime HMI (Kaiser Optical Systems Inc.) was used for spectral acquisition in all 2000 L batches. Collection for all Raman spectral data used the system settings of 10 s exposures for 75 scans, which resulted in a spectrum for a probe after 15 min including an overhead time of 2.5 min. Raman spectral acquisition spanned from wavenumbers  $100\text{--}3425\text{ cm}^{-1}$ . Reduced-scale vessels were protected from light interference by aluminum foil (due to the single-use bioreactor set up in the manufacturing suite this was not required as it was enclosed in stainless steel). Intensity calibration of the instrument was performed with the Hololab Calibration Accessory (Kaiser Optical Systems Inc.) prior to each use of the system and internal calibrations were set to occur every 24 h throughout the bioreactor process.

### 2.4 | Flow 1: Can Raman spectroscopy based PLS models be developed for glycation and glycosylation profiles of CHO cell bioreactor processes?

Raman based PLS models were developed as per Figure 2. In flow 1, both Raman and offline data from the cell culture process at small scale (1 and 5 L) was used to develop a panel of seven chemometric PLS models for the glycation and glycosylation profiles of the mAb. Two models were considered for glycation (mono-glycated, non-glycated), and five models were considered for the glycosylation profile (GOF-GlcNac, G0, GOF, G1F, and G2F). All chemometric modeling was performed with Simca 15.1 (Sartorius-Stedim, Germany). Offline measurements for glycation and glycosylation were aligned with Raman spectra based on the time at which they were taken, beginning at Day05 in each batch. The decision to build models from this time point was based on empirical knowledge of the process and its production of mAb product at detectable levels by HPLC. All data prior to day 05, Raman spectra and offline measurements, were excluded from model building and testing. Each model consisted of a calibration

sample set (CSS) of 15 batches ( $9 \times 1\text{ L}$  and  $6 \times 5\text{ L}$ ), for model development, and a calibration test sample set (CTSS) of 1 batch (5 L), used as a blind data set for testing with the PLS model generated for each CQA. This batch was chosen as the CTSS randomly from the available 5 L batch data. The X variables for each model in this flow were Raman spectra (centered) and the Y variables were offline values for: % mono-glycated, non-glycated, GOF-GlcNac, G0, GOF, G1F, and G2F (univariately scaled). Wavenumber selection of the Raman spectra for all models was  $415\text{--}1800\text{ cm}^{-1}$  and  $2800\text{--}3100\text{ cm}^{-1}$ . The spectral filters applied to all PLS models were Savitsky–Golay first derivative quadratic ( $31\text{ cm}^{-1}$  point) and standard normal variate (SNV; data not shown). Each PLS model was built and assessed for error by using a method of leave-batch-out cross validation (leave each batch out once in model development). The model error was averaged based upon prediction of the model against the omitted batch to identify the root mean square error of cross-validation (RMSEcv). The RMSEcv indicates the predictive power of the model based on the data used to build the model. A lower average error (RMSEcv) indicated an improved model. This enabled better informed decision-making on which component number was to be used for the models generated for testing against the blind data set. Testing the models' predictive capability versus the CTSS, using the predict function in Simca 15.1, identified the root mean square error of prediction (RMSEP), which indicates the predictive power of the model versus an unseen dataset. The Regression ( $R^2$ ) value, coefficient of variation was recorded for each PLS model. This  $R^2$  value is used to determine the amount of variation of the Y variable which the model predictors (X variables) can explain. The closer an  $R^2$  value is to 1, the greater a model explained the Y variable. Model performance was assessed based on each models' respective RMSEcv, RMSEP, and  $R^2$  values. The variable importance of projections (VIPs) of each model was also considered, this is a parameter which summarizes the importance of the X-variables in a model in predicting the Y-variable. X-Variables with a value greater than 1, are considered most relevant for explaining the Y-variable.

### 2.5 | Flow 2: Can Raman spectroscopy based PLS models for glycation and glycosylation accurately predict at manufacturing scale using only small scale data?

In this flow, each of the models developed using the CSS from flow 1. were investigated for their capability to accurately predict at manufacturing scale. No adjustments or additional data were used to develop the models in this flow. Each model (mono-glycated, non-glycated, GOF-GlcNac, G0, GOF, G1F, and G2F) was tested and evaluated by using the predict function in Simca 15.1 against a new CTSS comprised of data from a single manufacturing scale batch (2000 L Batch A) for this cell culture process. This batch was chosen as the CTSS randomly from the available 2000 L batch data. The CTSS was then used to investigate the ability of Raman models developed using only small scale data (1 and 5 L) to predict CQAs at a manufacturing scale.

A leave batch out cross validation was not repeated as there were no changes made to the models, and as such no change to RMSEcv was observed. Each of the seven model's RMSEP and  $R^2$  values were compared to the outputs of models generated in flow 1 to determine whether small scale data alone is sufficient for PLS model development.

## 2.6 | Flow 3: Can Raman spectroscopy based PLS model robustness for glycation and glycosylation be improved to accurately predict at manufacturing scale?

Flow 3 investigated the use of an updated CSS which is comprised of the CSS data from previous flow 1 supplemented with data from a single manufacturing scale (2000 L Batch B) batch to build each PLS model. The CSS of 16 batches ( $9 \times 1\text{ L}$ ,  $6 \times 5\text{ L}$ , and  $1 \times 2000\text{ L}$ ) was then used to develop seven updated models as per Section 3.4. Once completed, each of the seven models was then assessed by leave batch out cross validation in order to derive an updated RMSEcv and component number for each model. Testing the models' predictive capability versus the new CTSS, using the predict function in Simca 15.1, identified the root mean square error of prediction (RMSEP) and the  $R^2$  value for the updated models. As in previous flows, the RMSEcv, RMSEP,  $R^2$ , and VIP values output from each model were assessed to identify the models' ability to predict at manufacturing scale. They were also used to compare predictive capability against the models developed and used in flow 1 and flow 2 to determine the

necessity of including manufacturing scale data in the CSS of Raman PLS models for glycation and glycosylation.

## 3 | RESULTS AND DISCUSSION

In the development of a robust and effective process for production of therapeutic mAbs, there are a number of key CQAs that must be considered. The characterization of these CQAs across multiple batches ensures the production process is demonstrated to be robust and effectively controlled by the process parameters in place. The European Medicines Agency's guideline on the quality control of mAbs indicates specific structural features of mAbs which should be considered when determining CQAs, glycation and glycosylation (in particular, fucosylation) are two features identified in this document.<sup>26</sup>

Raman spectroscopy is a PAT tool capable of providing data using an in-situ probe, partnered with chemometric PLS models. By using Raman spectroscopy, there is the potential to deliver a predicted value for a range of process variables normally confined to a single daily offline measurement<sup>27,28</sup> as frequently as once every 15 min, when a single probe is in use, generating up to 96 times the amount of information versus a single daily offline measurement for a 16 days process. In this work, PLS models were developed to characterize glycation (mono-glycated, non-glycated) and glycosylation (G0F-GlcNac, G0, G0F, G1F, and G2F) of mAb product produced by a single proprietary CHO cell line at reduced (1 and 5 L) and manufacturing scales (2000 L).

**TABLE 1** Flow 1: Model development statistics for Raman PLS models for glycation and glycosylation

Cross-validation data set					
CQA (%)	Component number	$R^2$	RMSECV	Acceptable RMSECV	Acceptance criteria (pass/fail)
Mono-glycated	6	0.8036	0.5169	$\leq 0.554$	Pass
Non-glycated	6	0.8036	0.5169	$\leq 0.554$	Pass
G0F-GlcNac	6	0.9019	0.0312	$\leq 0.0322$	Pass
G0	5	0.9348	0.1315	$\leq 0.1453$	Pass
G0F	6	0.9546	2.1576	$\leq 2.6964$	Pass
G1F	6	0.9570	2.0707	$\leq 2.4149$	Pass
G2F	5	0.9138	0.2670	$\leq 0.3349$	Pass
Blind prediction data set (5 L)					
CQA (%)	Component number	$R^2$	RMSEP	Acceptable RMSEP	Acceptance criteria (pass/fail)
Mono-glycated	6	0.7131	0.5409	$\leq 0.554$	Pass
Non-glycated	6	0.7131	0.5409	$\leq 0.554$	Pass
G0F-GlcNac	6	0.6945	0.0303	$\leq 0.0322$	Pass
G0	5	0.8777	0.1059	$\leq 0.1453$	Pass
G0F	6	0.902	1.8996	$\leq 2.6964$	Pass
G1F	6	0.9163	1.756	$\leq 2.4149$	Pass
G2F	5	0.9115	0.2273	$\leq 0.3349$	Pass

This study was set up to determine if Raman spectroscopy can be used as a tool to deliver real time information on the glycation (mono-glycated, non-glycated) and glycosylation (G0F-GlcNAc, G0, G0F, G1F, and G2F) profiles of a mAb producing CHO cell process at both reduced and manufacturing scales.

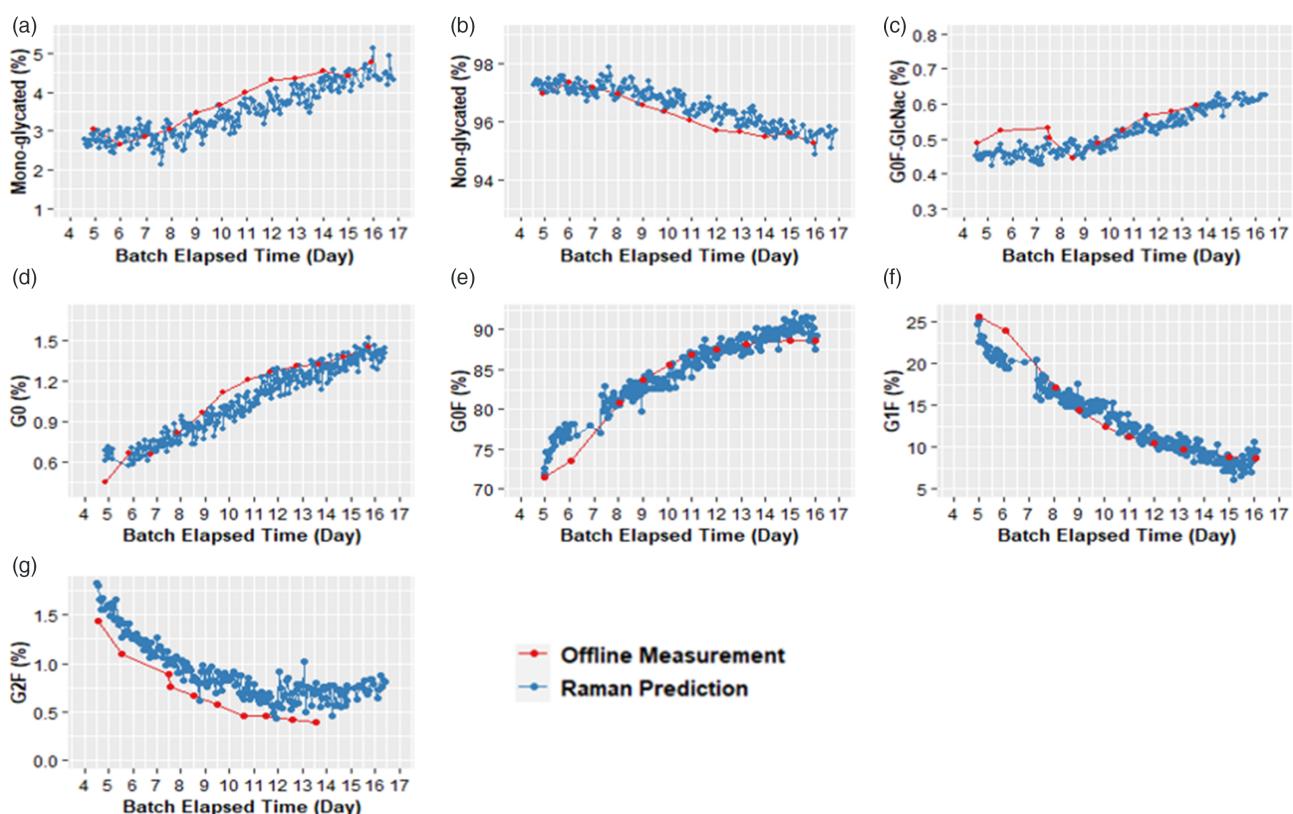
### 3.1 | Can Raman spectroscopy based PLS models be developed for glycation and glycosylation profiles of CHO cell bioreactor processes? (Flow 1)

Flow 1 successfully demonstrates the use of small scale data only as a calibration data set for PLS model building. Each of the seven models developed, two for glycation (mono-glycated, non-glycated) and five glycosylation profiles (G0F-GlcNAc, G0, G0F, G1F, G2F) were tested by predicting against a blind data set of small scale batch data. Each model in flow 1 consisted of 15 batches of small scale process data in the CSS with one full batch (5 L) being used for testing as a blind CTSS.

Leave batch out cross validation was used to determine the optimum component number for each model. This was decided as the lowest number of components with the lowest matching RMSEcv value. Each model was assessed individually for optimum component number. Average  $R^2$  results from cross validation show a strong degree of variability in each model, explained by an  $R^2$  value of >0.8 reported for each of the models considered (Table 1). The acceptance criteria for each

model's accuracy were determined based on the range of the measurements observed in the model CSS. Accuracy acceptance criteria was set as 10% of the range of values used in the CSS set as per chemometric model development standards at JSI. This was determined to be an adequate acceptance criterion based on potential sources of error for each model including the acceptable error of the method employed in offline analysis. The error in each model was assessed by comparing the root mean square errors (RMSEcv, RMSEP) against the acceptance criteria calculated (Table 1). The RMSEcv informs the optimum component number. This value also indicates the ability of the model to predict the variable of interest based on the calibration dataset. The component numbers chosen for each model gave an RMSEcv which fell within the acceptance criteria for prediction error and as such were deemed appropriate for model development.

Complimentary results were observed when each model's prediction accuracy was tested with the blind data set (5 L batch) (Table 1). The  $R^2$  was >0.85 for mono-glycated, non-glycated, G0F, G1F, G2F, G0, and <0.7 G0F-GlcNAc. All  $R^2$  values showed a decrease when tested against the blind data set, however, the RMSEP values for all models, except for G0, Mono-glycation and Non-glycation, showed a lower error in prediction than the values observed for the RMSEcv. An  $R^2$  of >0.9 does not necessarily indicate a better model and reinforces the need to consider multiple factors including RMSEcv and RMSEP in conjunction with  $R^2$  values during model assessment. While testing against the blind data set showed a decrease in  $R^2$  and slight



**FIGURE 3** Trends for flow 1 Raman PLS model predictions versus offline measurements for glycation (a,b) and glycosylation (c-g) in small scale (5 L) batch (CTSS)

increase in model error observed for the G0, mono-glycation and non-glycation models, it should be noted that these models still fall within the acceptance criteria and so can be deemed suitable for use in predicting at small scale. The trends observed for the prediction testing (Figure 3) appear to confirm the suitability of the models for use, the Mono-glycated and Non-glycated trends follow the offline data trends tightly and are complementary to one another for the process duration as would be expected. GOF-GlcNac, G0, GOF, G1F, and G2F follow the offline trends similarly and while slight deviations were observed in each, all were within acceptable error criteria. Raman spectral regions identified in each model to have a VIP score > 1 indicated an acceptable degree of model specificity (Appendix 1). Glycation models (mono-glycated, non-glycated) show a Raman spectral region association similar to that identified for glucose PLS models previously,<sup>29</sup> while some degree of secondary correlation is likely due to the structure of glycated proteins, unwanted correlations with glucose for these models was avoided by ensuring data used in the model CSS was acquired at intervals both pre and post glucose feeding. Different glucose feeding strategies were also employed when executing bioreactor batches for data collection in order to break unwanted glucose correlations. The glycosylation profile models (GOF-GlcNac, G0, GOF, G1F, and G2F) show VIP scores > 1 in regions which have previously been associated with glycan components, for example, mannose, fucose, *n*-acetylglucosamine.<sup>29–32</sup> This suggests the model identifying peaks in the Raman spectra associated with each of the glycans however it is unclear if these are true peaks and further study on the specificity of these peaks would be required to confirm. CHO cell bioreactors are complex and dynamic environments which can make specificity comparisons using pure analyte difficult. As well as this, confirmation of bond regions by analyte spiking may

not be an option due to the complex nature and formation of glycan species in a CHO cell bioreactor environment. Further research into the Raman bond associations in CHO cell bioreactors is necessary. It is likely that for each of the models presented there will be some degree of secondary correlations associated with product titer output as both glycation and glycosylation are product dependent variables.

While limited ICH guidelines governing the GMP implementation of models where secondary correlations are observed, it is likely that the implementation of such models as a GMP validated tool for process monitoring would require updates to existing regulatory guidelines. The relationship which explains the secondary correlation will need to be well described for models, such as those presented here, if they are to be implemented in a GMP manufacturing setting as a replacement for traditional test methods. As glycation and glycosylation detection will always be linked to titer output then this will need to be considered during the model development to ensure that the model design space is robust, and the influence of any secondary correlations can be minimized. Extensive work is still required to establish real time measurements of CQAs using modeling techniques such as PLS as replacement for offline measurements.

An initial assessment of the inner relation plot of each model indicated a satisfactory level of linearity to progress with model assessment, however it is possible for future work, given secondary correlations with product titer, a nonlinear-PLS method could potentially improve model accuracy. Model performance in each case is acceptable and a close agreement with daily offline samples with an expected profile supports model development decisions made in this flow.

The ability to monitor CQAs such as glycation and glycosylation in real time during a bioreactor process aligns with the key goals of

**TABLE 2** Flow 2: Model development statistics for Raman PLS models for glycation and glycosylation

Cross-validation data set					
CQA (%)	Component number	R <sup>2</sup>	RMSECV	Acceptable RMSECV	Acceptance criteria (pass/fail)
Mono-glycated	6	0.8036	0.5169	≤0.554	Pass
Non-glycated	6	0.8036	0.5169	≤0.554	Pass
GOF-GlcNac	6	0.9019	0.0312	≤0.0322	Pass
G0	5	0.9348	0.1315	≤0.1453	Pass
GOF	6	0.9546	2.1576	≤2.6964	Pass
G1F	6	0.957	2.0707	≤2.4149	Pass
G2F	5	0.9138	0.267	≤0.3349	Pass
Blind prediction data set					
CQA (%)	Component number	R <sup>2</sup>	RMSEP	Acceptable RMSEP	Acceptance criteria (pass/fail)
Mono-glycated	6	0.8699	0.3553	≤0.554	Pass
Non-glycated	6	0.8699	0.3553	≤0.554	Pass
GOF-GlcNac	6	0.8544	0.0272	≤0.0322	Pass
G0	5	0.8334	0.0997	≤0.1453	Pass
GOF	6	0.9317	2.667	≤2.6964	Pass
G1F	6	0.9447	2.3687	≤2.4149	Pass
G2F	5	0.8916	0.4074	≤0.3349	Fail

QbD, with the assessment of quality being built into the process.<sup>33</sup> Using in process CQA monitoring, a cause and effect relationship can be established between the critical process parameters (CPPs) and product quality. Establishing this approach early in product development when process characterization has not yet been fully completed can reduce the time needed for development and scale up<sup>34</sup> as well as reducing manufacturing inefficiencies which allow for tighter control of product quality and an improvement in yield.<sup>25</sup>

The results presented here support the decisions made in model development and show that Raman spectroscopy based PLS models are capable of predicting both the glycation profiles and glycosylation, with specific focus on glycan profiles, of CHO mAb producing bioreactor processes.

In order to fully validate each model's robustness, a second set of prediction testing was carried out on each model from flow 1 using batch data from a manufacturing scale (2000 L) process as the blind test dataset.

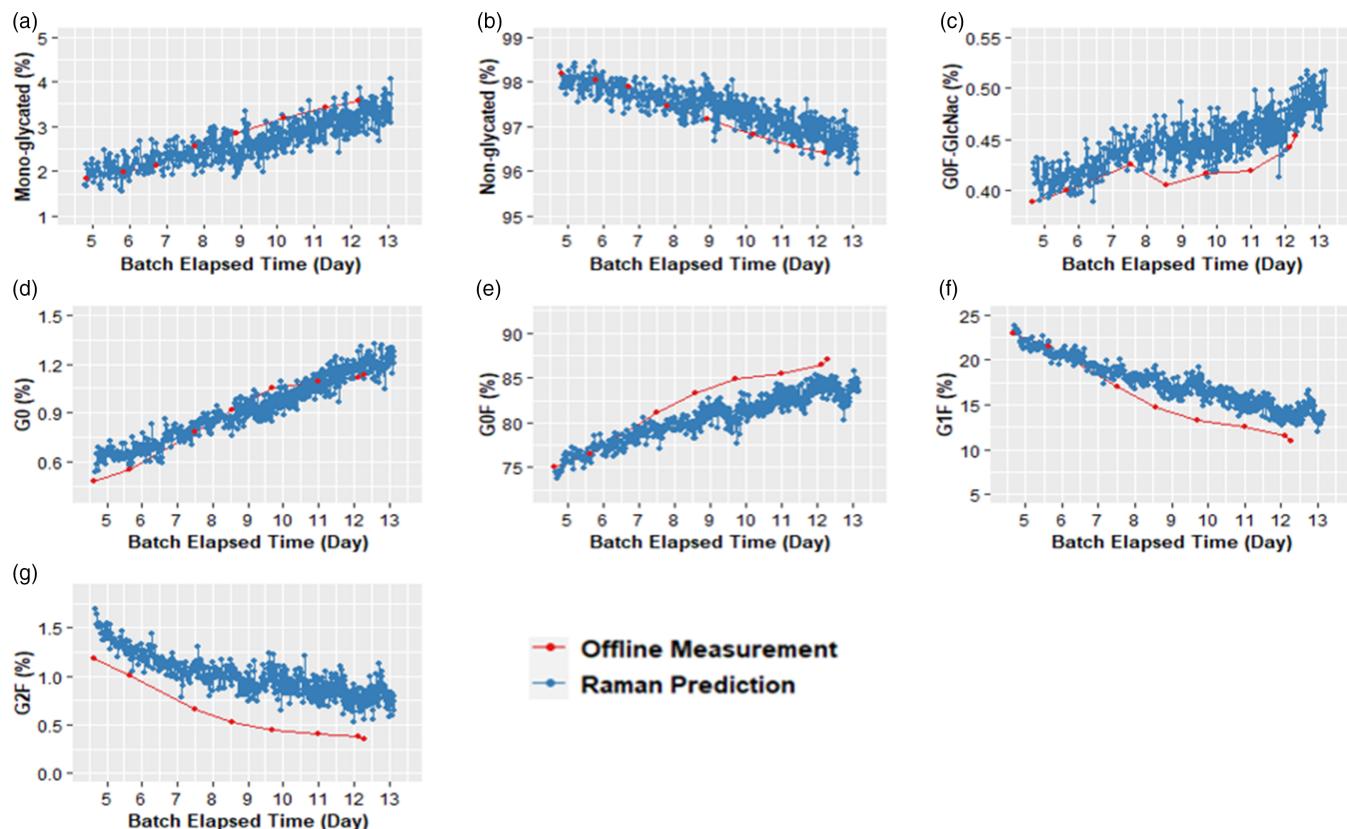
### 3.2 | Can Raman spectroscopy based PLS models for glycation and glycosylation accurately predict at manufacturing scale using only small scale data? (Flow 2)

In flow 2, a varying degree of accuracy was observed when model robustness was tested against manufacturing scale data (2000 L). Each

model from flow 1 was tested by predicting against a new blind CTSS comprised of process data from a 2000 L batch. The model statistics for RMSEcv were as observed in flow 1 as no additional data was added or removed from the models. Average  $R^2$  values of >0.80 were observed when predicted against the 2000 L blind dataset in the case of all models (Table 2) indicating good variability within each model.

The glycation models both performed well, with an RMSEP of 0.3553% observed in the case of each model (Mono-glycated and Non-glycated). Interestingly the RMSEP values here are lower than were observed in flow 1. This would indicate that small scale data alone is sufficient to develop a model for monitoring glycation profiles of mAb processes. The interactions that lead to protein glycation are dependent on the levels of reducing sugars, temperature and time in the bioreactor.<sup>35</sup> While the specific glycation site occupancy can be difficult to control,<sup>36</sup> the overall level of glycation can be maintained at certain levels during production scale up. As such, the process variables which typically contribute to glycation are maintained at a comparable level across all bioreactor scales. For this reason, using small scale data may be sufficient in developing robust glycation models.

Statistics relating to the glycosylation models show a more varied response when tested against the manufacturing scale data. The RMSEP value for the G2F model was observed to be 0.4074% which was outside of the acceptance criteria and accordingly the model was deemed unsuitable for use in monitoring at manufacturing scale. Interestingly, the  $R^2$  value for this model was >0.85, further reinforcing the



**FIGURE 4** Trends for flow 2 Raman PLS model predictions versus offline measurements for glycation (a,b) and glycosylation (c-g) in manufacturing scale (2000 L) batch (CTSS)

importance of consulting multiple statistics when evaluating PLS models. The trends showing the Raman model predictions versus the offline values for the 2000 L batch as it progressed (Figure 4) confirm the model statistics, with the G2F deviating from the offline measurement trends for the entirety of the batch. The G0F-GlcNac, G0, G0F, G1F models met the acceptance criteria and visually appear to follow the trends of the offline measurements. G0 and G0F-GlcNac performed best with a comparable or lower RMSEP value versus that obtained in flow 1. G0F and G1F showed a noticeable increase in RMSEP value and it was observed from the trends (Figure 4) that the predictive capability of these models begins to decline past day 07 in the process. Existing knowledge indicates that the process moves from the exponential growth phase to the stationary phase at day 07/08. As CHO cell cultures progress and move into the stationary phase of growth they reach peak antibody production. Cell size and metabolic rates change with the cells increasing in product output by as much as twice the rate in the exponential growth phase.<sup>37</sup> The interactions which give rise to mAb glycosylation are complex in nature and rely on numerous factors, often difficult to control.<sup>38</sup> The models presented in flow 2 would indicate that the process changes occurring at this time point have a different effect on the glycosylation profiles at small scale versus manufacturing scale, particularly as seen for G0F, G1F and G2F. Differences relating to scale are suspected here as the calibration data set for each model was constructed with robustness in mind. Variability in the data was introduced in each model by collecting spectra at multiple scales (1 and 5 L) and using different Raman probes/optics. Additionally, batches using different glucose feed strategies ( $6 \times 1$  L) and glucose spikes ( $1 \times 5$  L) were used in the calibration data set to break any spurious correlations with batch progression or glucose feeding events. Scale related differences are further suspected specifically as cell culture was taken from the 2000 L bioreactor at day 01 in the process and used to inoculate  $2 \times 5$  L bioreactors which formed part of the CSS of each model as well as the expected process range for each variable at 2000 L scale being built into the models.

It would be reasonable to accept that small scale data alone is sufficient for monitoring glycation and glycosylation profiles as each model, except G2F, meet the acceptance criteria. However, taking the models as a combined profile, glycosylation models built using small scale data perform worse when predicting against manufacturing data with an increased offset observed in G0, G1F, and G2F. G2F, in particular, had a significant increase (79%) in prediction error putting it outside the acceptable limits for prediction error (RMSEP). This offset is suspected to be due to the environment in the manufacturing scale reactor. While small scale bioreactor processes are designed to be representative of the manufacturing process, it is apparent here that the glycosylation kinetics at manufacturing scale differ to a degree whereby small scale data is not alone adequate for robust PLS model development.

To verify this claim, and increase the robustness of each model, the calibration dataset for models developed in flow 1 were supplemented with a single batch of manufacturing scale data (2000 L) and prediction testing was carried out as per flow 2, using the same 2000 L blind CTSS.

### 3.3 | Can Raman spectroscopy based PLS model robustness for glycation and glycosylation be improved to accurately predict at manufacturing scale? (Flow 3)

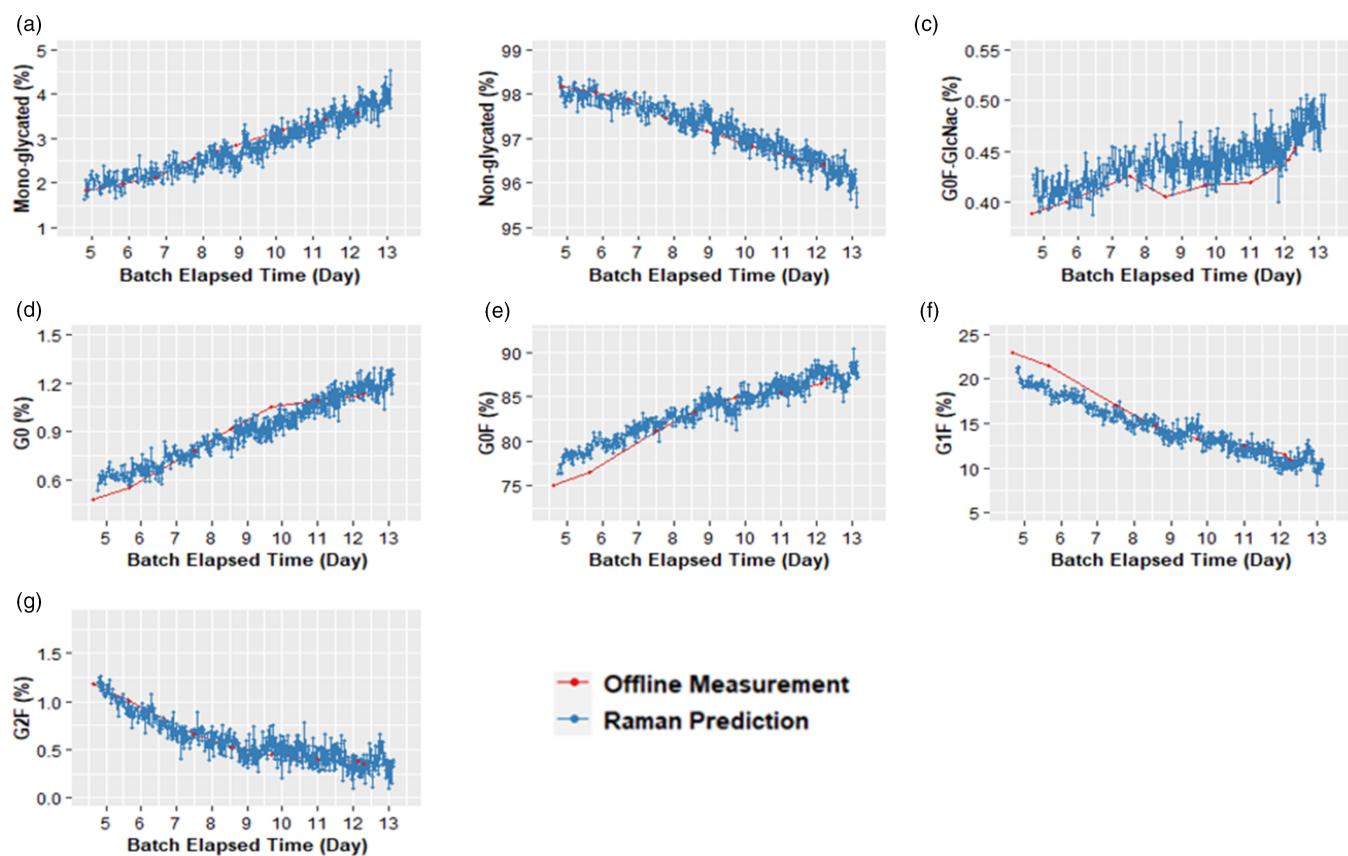
Flow 3 displays a significant improvement in model accuracy and robustness when compared to the models constructed in flow 1. The models from flow 1 were supplemented with process data from a 2000 L scale batch run. The addition of this 2000 L batch data meant that 16 batches were used to develop each model. Cross validation was again used here to determine the optimum component number for model development and prediction testing. An average  $R^2$  value of >0.8 was observed at cross validation for all models with each model also having an RMSEcv within the acceptable criteria (Table 3).

For this flow, the prediction testing was completed using the manufacturing scale blind test data set from flow 2. An average  $R^2$  of >0.80 was observed for each model. Notably, all models showed a decrease in the RMSEP value indicating a reduced error in prediction across all models. Also the RMSEcv of each model was observed to be similar or slightly reduced for each model which would signal that the addition of manufacturing scale data to these models did not impact the models ability to predict at small scale. The G2F model was greatly improved when the CSS was supplemented with manufacturing scale data, as evidenced by a reduction in RMSEP from 0.4074% to 0.0919% (a relative improvement in prediction error of 77.5%). The trends seen in Figure 5 show that the deviations which were observed in flow 2 have been accounted for and the models can now accurately predict throughout the entire manufacturing scale process. The regions of each model with VIP score > 1 were compared to the models created in flow 1, each model designated similar or the same regions as significant contributors to the model, in some cases the score for the regions identified changed as a result of the manufacturing scale data addition, which further supported the decisions made during model development (Appendix 1). It should be noted that during this work models for both mono-glycation and non-glycation were developed, each model shared the same results across all three flows. This is as these parameters split 100% of the analytical range, as a result a single model could be created, and the complimentary result inferred. For the purpose of this work, both models were created in an effort to support the decisions made in model development and show that acceptable values were obtained in each case when testing against an unknown dataset.

During process scale up, the bioreactor process for a specific mAb product will be tailored to a larger bioreactor, for example, 2000 L. However, there are a number of issues which may arise in the scale up process which lead to differences between the manufacturing and small scale process. Often, culture mixing time, CO<sub>2</sub> removal and oxygen transfer rates can differ with the manufacturing scale having a lower oxygen transfer coefficient and CO<sub>2</sub> removal rate, along with a greater mixing time.<sup>39</sup> A reduced oxygen transfer in the manufacturing scale process has the potential to influence the glycation and in particular the glycosylation profiles of the mAb products in a way that would not be seen at the small scale. A reduced

**TABLE 3** Flow 3: Model development statistics for Raman PLS models for glycation and glycosylation

Cross-validation data set					
CQA (%)	Component Number	R <sup>2</sup>	RMSECV	Acceptable RMSECV	Acceptance criteria (pass/fail)
Mono-glycated	6	0.8029	0.5132	≤0.554	Pass
Non-glycated	6	0.8029	0.5132	≤0.554	Pass
G0F-GlcNac	6	0.9007	0.0311	≤0.0322	Pass
G0	5	0.9312	0.1313	≤0.1453	Pass
G0F	6	0.9517	2.1541	≤2.6964	Pass
G1F	6	0.9535	2.0647	≤2.4149	Pass
G2F	5	0.9316	0.2669	≤0.3349	Pass
Blind prediction data set (2000 L)					
CQA (%)	Component number	R <sup>2</sup>	RMSEP	Acceptable RMSEP	Acceptance criteria (pass/fail)
Mono-glycated	6	0.8992	0.3279	≤0.554	Pass
Non-glycated	6	0.8992	0.3279	≤0.554	Pass
G0F-GlcNac	6	0.8171	0.0224	≤0.0322	Pass
G0	5	0.8424	0.0977	≤0.1453	Pass
G0F	6	0.9302	1.50925	≤2.6964	Pass
G1F	6	0.9438	1.43486	≤2.4149	Pass
G2F	5	0.9147	0.0919	≤0.3349	Pass

**FIGURE 5** Trends for flow 3 Raman PLS model predictions versus offline measurements for glycation (a,b) and glycosylation (c-g) in manufacturing scale (2000 L) batch (CTSS)

oxygen transfer leads to potential presence of “dead zones” in the bioreactor where no oxygen is present for brief periods.<sup>40</sup> This leads to an increased oxidative stress on the CHO cells in the bioreactor. Oxidative stress has been reported to have an effect on the glycosylation of mAbs as this stress reduces acetyl-CoA formation which in turn leads to a decreased N-acetylglucosamine (GlcNac),<sup>41</sup> a key amide which forms part of the backbone structure of glycosylation targets described here, apart from GOF-GlcNac.

This highlights the importance of model robustness and the considerations which must be taken into account when designing model calibration data sets. Manufacturing scale data, while in this study only representative of 4.4% of the total data observations available for use in the CSS, is a requirement in order to fully capture process variability and ensure robust model development. This is in keeping with previous research on scalability of Raman models<sup>29</sup> where a panel of common process variables were considered. Models developed for variables using only small scale data performed well in most cases, however a greater prediction error (RMSEP) was observed versus models for the same variables which included pilot scale or manufacturing scale data in the model CSS.

Process robustness is an essential consideration across the entire biopharmaceutical industry. Disruptive technologies such as in line monitoring with Raman spectroscopy stand to push process understanding forward, provide robust process control and improve product quality and reliability. A recent white paper outlining key gaps in pushing process robustness forward singles out in line monitoring, specifically highlighting process CQAs as a key gap in modern processes.<sup>4</sup> In developing PLS models capable of providing in line monitoring of glycation and glycosylation profiles, we present a viable solution to this problem. A roadmap published by the same group suggests the impact of implementing real time monitoring of CQAs in a biopharmaceutical manufacturing environment has the potential to impact nearly every stage of the process. In line monitoring increases process control as it provides the opportunity to identify potential quality issues before they impact the batch, reduce process variability, reduce supply costs due to yield improvements, carry out real time release of the product, reduce lead times and positively impact the technology transfer timelines of products due to a reduction in analytical method transfer and validation processes<sup>42</sup>. PLS models such as those presented here will help the biopharma industry realize a true QbD approach in biotherapeutics production.

## 4 | CONCLUSION

Raman spectroscopy has presented itself as a powerful PAT tool for in situ monitoring of cell culture variables. While one previous study has shown the potential for Raman monitoring of glycosylation in terms of site occupancy (macroheterogeneity),<sup>20</sup> this study represents, to the best of the authors knowledge, the first time in the literature where Raman spectroscopy has been used with chemometric PLS modeling to develop predictive models capable of monitoring glycation and glycosylation profiles, in particular individual glycoforms

(microheterogeneity), with a specific focus on manufacturing scale processes. The strength of Raman's ability to provide “real time” information is highlighted when the current methods of analyzing these CQAs and the influence which these CQAs have on mAb product quality is considered. Current methods are complex, time-consuming and reliant on the frequency of offline samples being taken, only giving a limited understanding of these CQAs. As a result, these methods are often only used after the bioreactor process has been completed. In situ Raman spectroscopic analysis of these CQAs can push the production process toward the QbD approach where quality is built into the process both during development and during the manufacturing process. Real time measurement of CQAs in a CHO cell bioreactor opens up opportunities for future real time control of these variables. Partnering real time product quality monitoring and real time advanced feedback control strategies creates a positive PAT feedback loop which will improve CHO cell bioreactor execution in both process development and commercial manufacturing for mAb products.

## ACKNOWLEDGMENTS

The authors would like to thank Janssen Science Ireland UC Operations team for their work during the manufacturing cell culture runs. Thanks to the Janssen Sciences Ireland UC USP BioTherapeutics Development team for their support in lab scale bioreactor runs, in particular Joseph McLaughlin, Lindsay Gallagher and Jessica Barrett. A special thanks also to the Irish Research Council for funding this research.

## AUTHOR CONTRIBUTIONS

**Luke A. Gibbons:** Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); writing – original draft (lead); writing – review and editing (equal). **Carl Rafferty:** Conceptualization (equal); investigation (supporting); writing – review and editing (equal). **Kerry Robinson:** Data curation (supporting); investigation (supporting). **Marta Abad:** Writing – review and editing (equal). **Francis Maslanka:** Resources (equal). **Nikky Le:** Resources (equal). **Jingjie Mo:** Resources (equal); writing – review and editing (equal). **Kevin Clark:** Funding acquisition (equal); resources (equal). **Fiona Madden:** Resources (equal); supervision (equal). **Ronan Hayes:** Resources (equal); supervision (equal). **Barry McCarthy:** Resources (equal); supervision (equal); writing – review and editing (equal). **Christopher Rode:** Funding acquisition (equal); resources (equal). **Jim O'Mahony:** Funding acquisition (equal); project administration (equal); supervision (equal); writing – review and editing (equal). **Rosemary Rea:** Funding acquisition (equal); project administration (equal); supervision (equal); writing – review and editing (equal). **Caitlin O'Mahony-Hartnett:** Conceptualization (equal); funding acquisition (equal); project administration (equal); supervision (equal); writing – review and editing (equal).

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## ORCID

Luke A. Gibbons  <https://orcid.org/0000-0002-9980-8119>  
Carl Rafferty  <https://orcid.org/0000-0003-3137-3515>

## REFERENCES

1. Andersen DC, Krummen L. Recombinant protein expression for therapeutic applications. *Curr Opin Biotechnol.* 2002;13:117-123.
2. Davis GC. Protein stability: impact upon protein pharmaceuticals. *Biologics.* 1993;21(2):105.
3. Wang W. Instability, stabilization, and formulation of liquid protein pharmaceuticals. *Int J Pharm.* 1999;185(2):129-188.
4. Wang T, Miller K, Robbins D, Springall J, Smelko J, Choo A, Swan H, Aich U, Shimuta T, Malmquist G, Ohrvik H, Vasi J, Adkins M, Arriola B, Haschke, Cuellar M, Tovey J, Beller J, Wilkes B, Patel B, Saucedo V, Wei B, Grimm C, Hsiao T, Lynch M, Green B. User Requirement Specification – Glycosylation Profile – Biophorum (12 May 2020). Available at: <https://www.biophorum.com/download/in-line-monitoring-real-time-release-testing-in-biopharmaceutical-processes-prioritization-and-cost-benefit-analysis/> (Accessed: 2 July 2020).
5. Wacker C, Berger C, Girard P, Meier R. (2011). Glycosylation profiles of therapeutic antibody pharmaceuticals, *Eu J Pharm Biopharm* 79. 503-7.
6. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS. Stability of protein pharmaceuticals: an update. *Pharm Res.* 2010;27(4):544-575.
7. Brady LJ, Martinez T, Balland A. Characterization of nonenzymatic glycation on a monoclonal antibody. *Anal Chem.* 2007;79(24):9403-9413.
8. Zhang Q, Ames JM, Smith RD, Baynes JW, Metz TO. A perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: probing the pathogenesis of chronic disease. *J Proteome Res.* 2009;8(2):754-769.
9. Jakas A, Horvat S. Study of degradation pathways of Amadori compounds obtained by glycation of opioid pentapeptide and related smaller fragments: stability, reactions, and spectroscopic properties. *Biopolymers.* 2003;69:421-431.
10. Ghaderi D, Zhang M, Hurtado-Ziola N, Varki A. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnol Gen Eng Rev.* 2012;28:147-176.
11. Patro SY, Freund E, Chang BS. Protein formulation and fill-finish operations. *Biotechnol Ann Rev.* 2002;8:55-84.
12. Krishnamurthy R, Manning M. The stability factor: importance in formulation development. *Curr Pharm Biotechnol.* 2005;3(4):361-371.
13. Wei B, Berning K, Quan C, Taylor Zhang Y. Glycation of antibodies: modification, methods and potential effects on biological functions. *mAbs.* 2017;9(4):586-594.
14. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. *Nat Biotechnol.* 2009;27:26-34.
15. Abu-Absi NR, Kenty BM, Cuellar ME, et al. Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. *Biotechnol Bioeng.* 2011;108: 1215-1221.
16. Capito F, Zimmer A, Skudas R. Mid-infrared spectroscopy-based analysis of mammalian cell culture parameters. *Biotechnol Prog.* 2015; 31(2):578-584.
17. Buckley K, Ryder AG. Applications of Raman spectroscopy in biopharmaceutical manufacturing: a short review. *Appl Spectrosc.* 2017;71: 1085-1116.
18. Smith E, Dent G. The theory of Raman spectroscopy. *Modern Raman Spectroscopy - A Practical Approach.* John Wiley & Sons, Ltd; 2005:71-92.
19. André S, Cristau LS, Gaillard S, Devos O, Calvosa É, Duponchel L. In-line and real-time prediction of recombinant antibody titer by in situ Raman spectroscopy. *Anal Chim Acta.* 2015;892:148-152.
20. Li M, Ebel B, Paris C, Chauchard F, Guedon E, Marc A. Real-time monitoring of antibody glycosylation site occupancy by in situ Raman spectroscopy during bioreactor CHO cell cultures. *Biotechnol Prog.* 2018;34:486-493.
21. Rafferty C, Johnson K, OMahony J, Burgoyne B, Rea R, Balss K. Analysis of chemometric models applied to Raman spectroscopy for monitoring key metabolites of cell culture. *Biotechnol Prog.* 2020;36:e2977.
22. Esmonde-White KA, Cuellar M, Uerpman C, Lenain B, Lewis IR. Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing. *Analyt Bioanal Chem.* 2017;409:637-649.
23. Yu LX, Amidon G, Khan MA, et al. Understanding pharmaceutical quality by design. *AAPS J.* 2014;16(4):771-783.
24. Yu LX. Pharmaceutical quality by design: product and process development, understanding, and control. *Pharm Res.* 2008;25(4):781-791.
25. Brewster VL, Ashton L, Goodacre R. Monitoring the glycosylation status of proteins using raman spectroscopy. *Anal Chem.* 2011;83(15): 6074-6081.
26. Dong J, Migliore N, Mehrman SJ, Cunningham J, Lewis MJ, Hu P. High-throughput, automated protein a purification platform with multiattribute LC-MS analysis for advanced cell culture process monitoring. *Anal Chem.* 2016;88(17):8673-8679.
27. EMEA, (2016) Guideline on development, production, characterisation and specification for monoclonal antibodies and related products. Available at: [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-development-production-characterisation-specification-monoclonal-antibodies-related\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-development-production-characterisation-specification-monoclonal-antibodies-related_en.pdf). (Accessed: 8 July 2020).
28. Rafferty C, O J, Burgoyne B, Rea R, Balss KM, Latshaw DC II. Raman spectroscopy as a method to replace off-line pH during mammalian cell culture processes. *Biotechnol Bioeng.* 2020;117(1):146-156.
29. Matthews TE, Smelko JP, Berry B, et al. Glucose monitoring and adaptive feeding of mammalian cell culture in the presence of strong autofluorescence by near infrared Raman spectroscopy. *Biotechnol Prog.* 2018;34(6):1574-1580.
30. DeGelder J, Gussem D, Vandenebeele K, Moens P. Reference database of Raman spectra of biological molecules. *J Raman Spectrosc.* 2007;38:1133-1147.
31. Wells HA, Atalla RH. An investigation of the vibrational spectra of glucose, galactose and mannose. *J Mol Struct.* 1990;224:385-424.
32. She CY, Dinh ND, Tu AT. Laser Raman scattering of glucosamine N-acetylglucosamine, and glucuronic acid general subjects. *Biochim Biophys Acta.* 1974;372:345-357.
33. PDA, Technical Report 56: Application of phase appropriate quality systems and cGMP to the development of therapeutic protein drug substance. Parenteral Drug Association: Bethesda, MD, 2012
34. Kozlowski S, Swann P. Current and future issues in the manufacturing and development of monoclonal antibodies. *Adv Drug Deliv Rev.* 2006;58(5-6):707-722.
35. Wei B, Berning K, Quan C, Zhang YT. Glycation of antibodies: modification, methods and potential effects on biological functions. *mAbs.* 2017;9(4):580-586.
36. Templeton N, Dean J, Reddy P, Young JD. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. *Biotechnol Bioeng.* 2013;110: 2013-2024.
37. Batra J, Rathore AS. Glycosylation of monoclonal antibody products: current status and future prospects. *Biotechnol Prog.* 2016;32 (5):1091-1102.
38. Xing P. Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Appl Microbiol Biotechnol.* 2017;101(22): 8101-8113.
39. Ast T, Mootha VK. Oxygen and mammalian cell culture: are we repeating the experiment of Dr. ox? *Nat Metabol Nat Res.* 2019;1: 858-860.
40. Chevallier V, Andersen MR, Malphettes L. Oxidative stress-alleviating strategies to improve recombinant protein production in CHO cells. *Biotechnol Bioeng.* 2020;117(4):1172-1186.
41. Ramasubramanyan N, Krinsek J, Pezzini J, Lu R, Smelko J, Swann P, Loladze V, Barnett J, Brophy L, Cuellar M, Strachan D, O'Riordan, Seiber

K. Higgins J, Zillmann M, Valery E, Doyle J, Jeffers P, Dillon P, Slade P, Harris R, Aich U, Dekner M, Philips M, Tummala S, Varghese J, Grimm C, Bell P, Josephs J, Dakin J, Green B (2020) In-line monitoring/real-time release testing in biopharmaceutical processes - prioritization and cost benefit analysis - BioPhorum Available at: <https://www.biophorum.com/download/in-line-monitoring-real-time-release-testing-in-biopharmaceutical-processes-prioritization-and-cost-benefit-analysis/> (Accessed: 2 July 2020).

**How to cite this article:** A. Gibbons L, Rafferty C, Robinson K, et al. Raman based chemometric model development for glycation and glycosylation real time monitoring in a manufacturing scale CHO cell bioreactor process. *Biotechnol. Prog.* 2022;38(2):e3223. doi:10.1002/btpr.3223

## APPENDIX 1

Overlay of trends for VIP scores of PLS models for (a) mono-glycation/non-glycation (b) G0F-GlcNac (c) G0 (d) G0F (e) G1F (f) G2F in flow 1 and flow 3.

