Article:
Povey, JF, O'Malley, CJ, Root, T et al. (8 more authors) (2014) Rapid high-throughput characterisation, classification and selection of recombinant mammalian cell line phenotypes using intact cell MALDI-ToF mass spectrometry fingerprinting and PLS-DA modelling. Journal of Biotechnology, 184. pp. 84-93. ISSN 0168-1656

https://doi.org/10.1016/j.jbiotec.2014.04.028

Rapid High-Throughput Characterisation, Classification and Selection of Recombinant Mammalian Cell Line Phenotypes using Intact Cell MALDI-ToF Mass Spectrometry Fingerprinting and PLS-DA Modelling

Jane F. Povey[1^], Christopher J. O'Malley[2^], Tracy Root[3^], Elaine B. Martin[2], Gary A. Montague[2], Marc Feary[3], Carol Trim[1], Dietmar A. Lang[3,4], Richard Alldread[3], Andrew J. Racher[3], C. Mark Smales[1*]

[1]Centre for Molecular Processing and School of Bioscience, University of Kent, Canterbury CT2 7NJ, UK

[2]School of Chemical Engineering & Advanced Materials, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

[3]Lonza Biologics plc, 228 Bath Road, Slough SL1 4DX, UK

[4]Current address: Unilever Research and Development, Port Sunlight, Quarry Road East Bebington, Wirral, CH63 3JW, UK

^These authors contributed equally to this work

*To whom correspondence should be addressed

Abstract

Despite many advances in the generation of high producing recombinant mammalian cell lines over the last few decades, cell line selection and development is often slowed by the inability to predict a cell line's phenotypic characteristics (e.g. growth or recombinant protein productivity) at larger scale (large volume bioreactors) using data from early cell line construction at small culture scale. Here we describe the development of an intact cell MALDI-ToF mass spectrometry fingerprinting method for mammalian cells early in the cell line construction process whereby the resulting mass spectrometry data is used to predict the phenotype of mammalian cell lines at larger culture scale using a Partial Least Squares Discriminant Analysis (PLS-DA) model. Using MALDI-ToF mass spectrometry, a library of mass spectrometry fingerprints was generated for individual cell lines at the 96 deep well plate stage of cell line development. The growth and productivity of these cell lines were evaluated in a 10 L bioreactor model of Lonza's large-scale (up to 20,000 L) fed-batch cell culture processes. Using the mass spectrometry information at the 96 deep well plate stage and phenotype information at the 10 L bioreactor scale a PLS-DA model was developed to predict the productivity of unknown cell lines at the 10 L scale based upon their MALDI-ToF fingerprint at the 96 deep well plate scale. This approach provides the basis for the very early prediction of cell lines' performance in cGMP manufacturing-scale bioreactors and the foundation for methods and models for predicting other mammalian cell phenotypes from rapid, intact-cell mass spectrometry based measurements.

Keywords: Cell line development, Chinese hamster ovary cells, whole cell MALDI-ToF mass spectrometry, PLS-DA modelling, cell line prediction

1.0 Introduction

The majority of recombinant protein biopharmaceuticals are produced from cultured mammalian cells (Walsh, 2010), with the most commonly used industrial mammalian cell host being the Chinese hamster ovary (CHO) cell (Kim et al., 2012). Despite the development of high throughput methods that allow the screening of many recombinant cell lines to isolate those with desirable phenotypes (e.g. high growth and productivity), the ability of such methods to select or predict the performance of a given cell line at manufacturing scale remains limited with the best cell lines at a manufacturing scale often distributed across the phenotypic performance of cell lines at smaller-scale (Porter et al., 2010a; Porter et al., 2010b). As a result, early use of a simple productivity based approach does not necessarily allow the identification of high producers in the population of cell lines, and some potentially high producers are discarded early in the process (Porter et al., 2010b). In order to address this issue, a number of proteomic, transcriptomic and metabolic based studies have now been undertaken and the subsequent data used to develop models to predict the phenotype of a given cell line (e.g. Clarke et al., 2011; Clarke et al., 2012; Doolan et al., 2013; Jacob et al., 2010; Mead et al., 2009; Mead et al., 2012; Meleady et al., 2011; Sanchez et al., 2014; Selvarasu et al., 2012), however these models are usually developed from cell line data at the end of the cell line development process.

The goal of this work was to develop a screening system that would allow the selection of highly productive cell lines for monoclonal antibody (mAb) production early in the cell line development process that would use substantially less resource to achieve the same or better success rate as current methods. The vision was to be able to select a small number of cell lines based upon the analysis of data generated in multi-well plates, and take these straight to a lab-scale bioreactor-evaluation stage (10 L) with a high probability that the selected cell lines were highly productive. The approach was based upon the use of intact cell MALDI-ToF mass spectroscopy (MS) to create ‘fingerprints’ from cells in multi-well plates and then to use the information to predict their behaviour in lab-scale bioreactors.

Intact whole cell MALDI-ToF mass spectrometry combined with discriminant analysis is routinely used as a method for characterising and discriminating between bacterial species and is an established method both in the clinical setting and research laboratory (e.g. see (De Bruyne et al.,

2011; Franco et al., 2010; Lundquist et al., 2005; Munteanu and Hopf, 2013; Panda et al., 2013; Veloo et al., 2011; Warscheid and Fenselau, 2004) and has led to a fundamental shift in the characterisation of bacteria in the clinical microbiology setting (Clark et al., 2013). Commercial software has been developed that allows the comparison of test spectra with a library of known spectra to identify bacterial strains with high precision (Bright et al., 2002; Sogawa et al., 2011). Despite the sensitivity, success and simplicity of MALDI-ToF coupled with appropriate data analysis for characterising bacteria, this approach has not been routinely utilised for the analysis of mammalian cells. Whilst there have been several reports of applying intact cell MALDI-ToF for the analysis of different mammalian cell lines and processes (Hanrieder et al., 2011; Karger et al., 2010; Marvin-Guy et al., 2008; Munteanu et al., 2012; Zhang et al., 2006) and to the analysis of CHO cells (Feng et al., 2011; Feng et al., 2010), none of these studies has attempted to use such an approach to predict across scale (e.g. use data early in the cell line construction process to predict a cells productivity at larger scale). A major challenge therefore remains with regard to developing methodologies that allow the prediction of phenotype based on intact cell MALDI-ToF mass spectrometry analysis. Indeed, a recent review describes how the limitations on using MALDI-ToF fingerprinting for mammalian cells lies around the facts that there is not currently a rigorous standardization of the approach utilized, automation will be necessary, information on the influence of cell numbers is scarce, there is a lack of an extensive database of fingerprints and that the identification of mammalian cell types requires a classification algorithm based on a high number of distinct mammalian cell types (Munteanu and Hopf, 2013).

Here we set out to address these limitations with respect to Chinese hamster ovary (CHO) recombinant cell lines and described the development and evaluation of an intact cell MALDI-ToF mass spectrometry profiling and PLS-DA modelling method to predict the phenotype (productivity) of novel recombinant mammalian cell lines at the bioreactor scale (10 L) using data generated in multi-well plates. This approach facilitates the exploitation of the heterogeneity of phenotypes in the cell line population by allowing the probabilistic prediction, early in the development programme, of the phenotype of cell lines in bioreactors. The approach used to generate an initial mass spectrometry training data set for building a Partial Least Squares - Discriminant Analysis (PLS-DA) model for the

prediction of productivity at the 10 L bioreactor scale, using MALDI-ToF mass spectrometry intact cell fingerprints from cells at the 96 deep well plate stage, is described. The model is based on a PLS-DA algorithm, which is a binary formulation of the original PLS algorithm (Wold et al., 1987) whereby the output vector comprises the values 0 or 1 thereby classifying the training data into two groups. The resulting PLS-DA model can be used as a rapid screen to classify cell lines into high/low producers based on their MALDI-ToF mass spectrometry profile.

2.0 METHODS

2.1 Cell Line Construction

Cell line construction was undertaken using a standard industrial process (**Figure 1A**) with subsequent bioreactors run under standard industrial conditions as previously described (Porter et al., 2010a; Porter et al., 2010b). In this way the GS expression system (Lonza) containing the mAb genes of interest for each of the three cell line constructions was introduced into the Lonza Biologics CHOK1SV host cell line.

2.2 Collection of Cell Pellets

Cell pellets of 6.25 x 10^4 total viable cells were collected 5-7 days after cell lines were plated out into 96 deep well plates using a Vi-CELL Cell Viability Analyzer instrument (Beckman Coulter) to determine the concentration of viable cells. This number of cells gave good reproducible spectra whilst higher cell numbers gave poorer and less reproducible spectra and at lower cell numbers the number of peaks in the spectra was reduced (e.g. see **Figure 2**). 6.25 x 10^4 total viable cells was removed from each well and transferred to a 96 well plate format rack, centrifuged for 5 minutes at 960 rcf and the supernatant removed. The cell pellet was washed with 0.5 ml of PBS and centrifuged a second time. The supernatant was again removed before the cells were washed with 0.5 ml of 0.35 M sucrose, centrifuged for 5 minutes and the supernatant removed. Cell pellets were then stored at -80°C.

2.3 Preparation of Cell Pellets for Intact Cell MALDI-ToF Mass Spectrometry Analysis

A 20 mg/mL (saturated) solution of sinapinic acid was prepared in matrix buffer (40% acetonitrile,

0.06% TFA) that was then placed in a sonicating water bath for 15 minutes before centrifugation. The prepared and washed cell pellets were removed from -80°C storage and allowed to equilibrate to room temperature for 15 mins. Matrix solution (50 µL) was then added to each cell pellet sample in which the cells were resuspended, then incubated at 4°C for 3 hours at which time the cells were resuspended by gently tapping the tube. 1 µL of each sample was then spotted onto a 384 MTP ground steel MALDI-ToF plate (Bruker). Samples were allowed to air dry before the plate was placed into the MALDI-ToF mass spectrometry instrument (Bruker Ultraflex).

2.4 MALDI-ToF Mass Spectrometry Instrument Operation Conditions

The MALDI-ToF ms instrument (Bruker Ultraflex) was calibrated before use with the commercially available Bruker Calibration Standard 1 protein mixture (Bruker, part number 206355). The spectra of the intact cell pellets were then collected using the MALDI-ToF ms instrument settings described below and the inbuilt calibration program for this calibration mixture provided with the instrument. The spectra were collected using the following settings on the mass spectrometer; Laser frequency: 20 Hz; Polarity +ve; Ion sources: 1. 20 kV, 2. 17.25 kV, Lens 5.0 kV; Gating mode: maximum strength; Suppress @ 4000 Da; Pulsed ion extraction 550 nS; Range 5000-60000 Da; Sample rate 0.1 Gs/s; Resolution enhanced 100 mV electronic gain; Smooth high. For each sample 100 shots were summed and saved.

2.5 Pre-Processing of MALDI-ToF Mass Spectrometry Data for PLS-DA Analysis

The mass spectrometry data files were exported from the Bruker Flex Analysis software in ASCII format for pre-processing in MATLAB prior to the application of the PLS-DA technique. The following pre-processing steps were applied to the spectra prior to presenting the pre-processed data to the PLS-DA algorithm: 1. Resampling (up-sample to 50,000 to account for slight differences in m/z vector); 2. Baseline Correction (removes the effect of noise introduced by the matrix); 3. Filtering (Application of Savitzky -Golay filter to smooth the signal); 4. Alignment (Automatically select peak and align spectra based on height and over-segmentation filters); 5. Quality Control (Outlier detection and removal of “unusual spectra”); 6. Normalisation (normalisation of area under curve). Re-sampling

of the mass spectrometry profiles was performed using the *'msresample'* function from the MATLAB Bioinformatics Toolbox (http://tinyurl.com/msresample). As mass spectra profiles typically exhibit a varied baseline due to issues such as chemical noise in the MALDI matrix and ion overloading (Monchamp et al., 2007), which is undesirable when using data analysis techniques to compare profiles, baseline correction was performed using the *'msbackadj'* function in the MATLAB Bioinformatics Toolbox (http://tinyurl.com/msbackadj). As a typical mass spectra contains a mixture of both signal and noise, smoothing of the signal by use of a Savitzky-Golay filter, typically applied to mass spectrometry signals, was performed following baseline correction using the '*mssgolay'* function from the MATLAB Bioinformatics Toolbox (http://tinyurl.com/mssgolay). **Figure 3** shows a selection of representative 96DWP spectra before and after application of Savitzky-Golay filtering. **Figures 3A** and **3C** show the full representative spectra before and after pre-processing whilst **Figures 3B** and **3D** show a close-up area of the spectra before and after pre-processing. Peak alignment is then undertaken to correct for variation between the observed m/z value and true time of flight using the '*msalign*' function from the MATLAB Bioinformatics Toolbox, (http://tinyurl.com/msalign). The final pre-processing step was normalisation, which addresses the variation in the amplitude of the ion intensities. This was performed using the '*msnorm'* function from the MATLAB Bioinformatics Toolbox, (http://tinyurl.com/msnormal).

2.6 Generation of the Model and Analysis of Naïve Data to Predict Classification of Cell Lines

Following pre-processing of the data, a model was built using the training data set, consisting of all 96 Deep Well Plate MALDI-ToF spectra from cell lines with 10 L bioreactor productivity data, using PLS-DA (http://wiki.eigenvector.com/index.php?title=Plsda). We note that each cell line was only grown in 10 L bioreactors once to obtain productivity data. Training data was classified into high or low productivities depending on the classification boundary (typically set at 4000 mg/L in this work). Additional training samples are appended to the training data set as they become available from new cell line constructions. The model is then used to classify or predict naïve samples from their 96 deep well plate MALDI-ToF fingerprints as to whether they are 'high' or 'low' producers. The number of latent variables selected for each model was based on the total percentage variation in the y-block

captured in the model. The number of latent variables (LVs) for each model was selected such that the number of LVs that described more than 80% of the variation was chosen, this number being selected so as not to over-fit to the training data.

3.0 RESULTS

3.1 The MALDI-ToF Mass Spectrometry Intact Mammalian Cell Fingerprinting and Cell Line Phenotype PLS-DA based Prediction Workflow

A classical recombinant cell line construction (CLC) process (**see Figure 1** and (Porter et al., 2010b) aligned with sampling for MALDI-ToF mass spectrometry analysis forms the basis of the approach we outline here with the subsequent development of a partial least squares discriminant analysis (PLS-DA) model to predict the performance of recombinant cell lines at the 10 L bioreactor scale from samples taken at the multi-well plate stage of CLC. The classical approach to the selection of recombinant cell lines is based on the productivity of cell lines at multiple stages of cell line development, however various studies have demonstrated that productivity measurements alone at any one stage of CLC do not necessarily reflect the performance of a given cell line under different conditions further along the CLC process (e.g. switching between static and shaken cultures, controlled bioreactor and shake flasks, batch vs fed-batch and different culture durations;(Porter et al., 2010a; Porter et al., 2010b). The methodology here moves away from multiple rounds of screening for productivity characteristics to a 'collapsed' single round of mass spectrometry based screening where there is a high probability of selecting cell lines with the required phenotypic attributes. The approach involves sampling of cells at the 96 deep well plate (DWP) stage of CLC, this step being the preferred early stage of CLC for sampling as at this time all cell lines have already been selected for suitable growth in suspension cultures. The cell samples are then subjected to intact cell MALDI-ToF mass spectrometry analysis and the resulting data used to develop a PLS-DA model that maximizes the separation between samples in a training library based on a classification of the spectra into 'high' and 'low' producers at the 10 L stage. Thus, the historical database or library of spectra was created from cell lines whose MALDI-ToF profile was collected at the 96 DWP stage, ensuring that cell lines from across the productivity range were represented in these samples (i.e. we took cell lines from all

productivity classes as determined by Protein A ELISA at the 96 DWP scale to represent the widest range of productivity phenotypes which were classified into 'high' or 'low' producers at the 10 L scale; high being greater than 4000 mg/L) and these cell lines were subsequently carried through the CLC process to the 10 L bioreactor stage to assess productivity characteristics. Naïve cell lines were then assessed against the historical model to provide an early indication (at the 96 DWP stage) of their classification (high or low) without the need for extensive scale-up experiments.

3.2 Cell Sampling, MALDI-ToF Data Collection and Pre-Processing of MALDI-ToF Mass Spectrometry Data before use in PLS-DA Modelling

The time of cell sampling, nature of cell samples and their preparation could potentially all influence the subsequent MALDI-ToF data generated, which forms the basis of the PLS-DA model. Each of the steps involved were investigated sequentially, including the MALDI-ToF instrument parameters. For the MALDI-ToF mass spectrometry analysis of recombinant cell lines, cell samples (6.25×10^4 viable cells per pellet) are collected at the 96 DWP stage 5-7 days after the cells are placed in shaking 96 DWPs. This low number of cells means this analysis can be undertaken early in the CLC process. A range of viable cell numbers per pellet was investigated for analysis and at higher cell numbers (above $1\text{-}2 \times 10^6$ viable cells per pellet) spectra were poor in quality (see **Figure 2**). The cell pellets are washed upon collection once with PBS followed by a wash with 0.35 M sucrose, this washing procedure gave more reproducible spectra than those in the absence of the washing procedure or with PBS alone as determined by principal component analysis (PCA). It was found that incubating cell samples in the saturated matrix solution at 4^{o}C for 3 h prior to spotting onto the MALDI-ToF target plate decreased the variation between samples and resulted in more reproducible spectra.

Before the MALDI-ToF mass spectrometry data were analysed it was necessary to pre-process the data as described in the methods section. The data files were exported from the Bruker Flex Analysis software in ASCII format for pre-processing into MATLAB, before application of the PLS-DA algorithm. The goal of the pre-processing step is to remove uncontrolled sources of variation caused by factors including contamination from the MALDI matrix, 'drift' in the instrument and differences in the amount of absorbed energy from the laser during ionisation. **Figure 3** shows

representative spectra highlighting the difference between the spectra of samples prior to and post pre-processing.

3.3 Generation of the Initial Training Data Sets and Model using a Classical Cell Line Construction Approach

Using the CLC workflow described above, a naïve CLC (here after referred to as CLC1) was initially undertaken (see process outlined in **Figure 1A**) to generate recombinant GS-CHO cell lines expressing the IgG4 monoclonal antibody cB72.3 that is routinely used at Lonza as a model system (Smales et al., 2004; Tait et al., 2012). After the initial transfection procedure and FACS sorting, 119 unique cell lines were taken from the initial transfection stage to the multi-well plate stage whereby cell samples were collected for MALDI-ToF analysis. All of these samples were subsequently analysed by intact cell MALDI-ToF mass spectrometry analysis. It was not practical to carry all these cell lines through the remainder of the CLC process to assess their productivity at the 10 L bioreactor scale. As such, a selection of cell lines (n = 29) with different productivities at the 96 DWP stage as determined by Protein A HPLC analysis were expanded from the 96DWP stage through the work flow in **Figure 1A** to the 10 L fermenter scale. Within these 29 cell lines were included the highest producers, but also a pseudo-random selection of low and intermediate producers to ensure coverage of the original productivity distribution so that MALDI-ToF fingerprints representative of the range of productivities were available to generate the initial PLS-DA models. These cell lines were grown over a period of 14 days under fed-batch conditions in 10 L bioreactors before being harvested and then the productivity determined by Protein A HPLC to provide a training set for modelling. The productivity data spread for these cell lines at the 10 L scale is shown in **Figure 4A**.

The Partial Least Squares – Discriminant Analysis (PLS-DA) algorithm was then applied to the MALDI-ToF profiles of these cell lines to attain an initial model. The class boundary was set to indicate cell lines with ‘high’ productivity, i.e. 4000 mg/L as the antibody used in the cell line construction process was known to be well expressed. This binary classification is set to differentiate between MALDI-ToF mass spectrometry profiles of high producing cell lines from those classified as low producing (in practice this class includes cell lines producing at low and intermediate levels). The

actual expression level of high producers will differ between antibodies (see **Figure 4B**), and hence the class boundary can be adjusted for other molecules that are expressed at lower concentrations. However, the philosophy underlying the modelling approach assumes that there is information inherent in the spectra of high producing cell lines that differentiates these from low producing cell lines, regardless of the target molecule. To further allow for any target specific differences, after each CLC the MADLI-ToF data at the 96 DWP stage for those cell lines taken through to the 10 L bioreactor evaluation is appended to the historical database to potentially further improve the predictive power of the model for different targets.

The 10 L bioreactors were not all run concurrently but in blocks and as each bioreactor set of productivities became available, this data was appended to the training data set that formed the basis of the PLS-DA model. Twenty samples were initially included in the training data set with this increasing to a final training set comprising 29 cell line productivities at the 10 L bioreactor scale. Once the PLS-DA model had been generated, the MALDI-ToF data from the remaining CLC1 samples (those from the original 119 not used in building the model) were presented to the model to predict their position relative to the classification boundary. **Figure 5A** shows the PLS-DA latent variables (LV) scores plot of LV1 vs. LV2. These two latent variables capture the major sources of variation between the cell lines with concentrations ≥ 4000 mg/L and those < 4000 mg/L and hence have the strongest discriminatory power. **Figure 5B** shows the position of the discrimination boundary and the cell lines predicted as belonging to each class as determined by the PLS-DA model comprising 5 LVs. Using this model a further 9 cell lines that had not already been assessed at the 10 L bioreactor scale were selected to assess their productivity at this scale based upon their predicted productivity classification. It is noted that the majority of the highest producing cell lines will have already been selected and used to generate the model and thus in all likelihood selection was from the second and third rank of producers. Of these 9 cell lines, 4 were initially run as a batch at the 10 L scale and the data incorporated into the model and then the following 5 run as an independent group to evaluate the predictions of these. From this it was identified that 3/5 (60%) predicted to be high producers (≥ 4000 mg/L) from the PLS-DA model were correctly classified. This compares favourably with the 11/31 (35%) of cell lines that were classed as high producers (of which 29 were used for the building of the

model - 2 were not used as the productivity data came from a different day compared to the rest of the cell lines) even though high producers had been consciously selected for the building of the model alongside a range of other producers and thus the majority of high producers from the CLC1 cell lines had potentially been removed. Furthermore, previously studies have demonstrated that a classical approach based upon productivity screening, similar to that used to select the cell lines for the generation of the model here, preforms much better than the random selection of cell lines (Porter et al., 2010b). The data in this analysis suggests that even without the inclusion of some of the highest producers the MALDI-ToF PLS-DA model was able to perform at least as well as this approach.

3.4 Validation of Intact Cell MALDI-ToF MS Fingerprinting of Recombinant GS-CHO Cell Lines and Subsequent PLS-DA Productivity Model Predictions

Following the initial development of the PLS-DA model using data from CLC1 to predict productivity of recombinant cell lines, an experiment was undertaken to both assess the ability of the existing model to predict the performance of recombinant cell lines expressing a different monoclonal antibody molecule and to increase the size of the training data set. In this study a second CLC was undertaken with a second monoclonal antibody (IgG1, different antigen target). Once again, cell samples were collected from cell lines at the multi-well plate stage (n=84) and these were subjected to intact cell MALDI-ToF mass spectrometry analysis. A total of 28 cell lines were then progressed to the 10 L bioreactor stage with these grown in 4 blocks / sets of bioreactors. These 28 cell lines were selected to represent a cross-section of productivities based on the traditional productivity screen and the predictions from the MALDI-ToF PLS-DA model that was updated as new training data was made available from each of the 4 bioreactor blocks. In this way the 10 L bioreactor productivity and MALDI-ToF training data from the second CLC was combined with the first CLC data set to generate a new model (i.e. the model was trained on data from both the first and second CLC bioreactor productivities) using a classification of ≥4000 mg/L to denote a high producing cell line. The resulting model was then used to predict the productivity classification of the remaining cell lines in CLC2 and a further 5 cell lines then assessed at the 10 L bioreactor scale. **Figure 5C** shows the PLS-DA latent variables scores plot of LV1 vs. LV2 from this data set whilst **Figure 5D** indicates the position of the

discrimination boundary and the cell lines predicted as belonging to each class as determined by the model trained using the data from both CLC1 and CLC2 using an 8 LV model. **Table 1** reports the predicted classification of the 5 cell lines from CLC2 selected for assessment and validation at the 10 L bioreactor scale based upon the model predictions reported in **Figure 5C and 5D.** As indicated in **Table 1**, the final productivities determined agreed in each case with the classifications predicted by the model. Thus, the MALDI-ToF analyses and predictions from these reported in **Table 1** show that in each case the model correctly predicted the classification of each recombinant cell line at the 10 L bioreactor scale (as a high or low producer).

3.5 Experimental Validation of the Method in Classifying High Producing Recombinant Cell Lines at the 96 DWP Stage during a Naïve Cell Line Construction Process

The model was then experimentally validated by undertaking a completely naïve CLC (CLC3) with a third monoclonal antibody molecule (IgG4) whereby the intact cell MALDI-ToF mass spectrometry spectra at the 96 DWP scale of CLC was used to select those cell lines to take forward to the 10 L bioreactor scale. In this validation experiment, classification predictions were made based entirely on the training data set from CLC1 and CLC2 and a classification boundary of ≥4000 mg/L retained although the 'expressibility' of this molecule was unknown at this time. Thus, the PLS-DA model to determine which cell lines from CLC3 had intact cell MALDI-ToF spectra most similar to the high titre training data set was developed using the database from CLC1 and CLC2. We did not expect the cell lines to necessarily have productivities of 4000 mg/L or greater but that these should be representative of high producing cell lines for this particular molecule in comparison to the titres observed at smaller scale and against the classical method of cell line construction. For this analysis the model was developed from 76 training samples (MALDI-ToF data at the 96 DWP scale and productivity data at the 10 L bioreactor scale from CLC1 and 2) and 220 test samples from CLC3 at the 96 DWP scale in duplicate or triplicate. Alongside the selection of cell lines to validate the model, selection was undertaken in parallel using the classical approach of productivity measurements at different stages of the process. From the whole cell MALDI-ToF spectra of 220 cell lines at the 96 DWP scale, the PLS-DA model was used to predict the classification of these as high or low producers

at the 10 L bioreactor scale and from this 8 cell lines were selected to progress through to a 10 L fed-batch bioreactor evaluation. The results from the models are shown in **Figure 6** and the predicted and actual productivities of the cell lines assessed at the 10 L bioreactor scale are reported in **Table 1**.

When the 8 cell lines were assessed at the 10 L bioreactor scale the highest producing cell line selected using the intact cell MALDI-ToF method had a productivity of 4946 mg/L and 3 of the cell lines returned productivities above 4000 mg/L (**Table 1**). The highest producing cell line was only ranked 32nd in terms of productivity of all cell lines at the 96 DWP stage as determined by Protein A analysis and hence would not normally have been selected to progress from this scale in a CLC process. Of the entire 8 cell lines selected by the intact cell MALDI-ToF PLS-DA prediction method, 5/8 had titres >3600 mg/L (which within the variation of titres from industrial 10 L bioreactors if cell lines were run multiple times could be considered as within the range expected for a 4000 mg/L cell line) whilst 7/8 had titres >2400 mg/L (**Table 1**). The lowest antibody concentration from the MS selected cell lines was 1918 mg/L. The two highest cell line productivities at the 10 L bioreactor scale as determined using the classical approach were 5218 mg/L and 4917 mg/L, although neither of these were ranked as potentially top producers by the MALDI-ToF PLS-DA model. Despite this, the MALDI-ToF approach classified and selected recombinant cell lines with productivities in the same region as those selected using the classical approach, experimentally validating the MALDI-ToF approaches ability to predict and select high producing cell lines. Furthermore, all of the cell lines selected using the intact cell MALDI-ToF approach had industrially relevant high productivities, generally considered to be those greater than 2000 mg/L in fed-batch culture (Lonza Biologics, unpublished data). All cell lines selected using the MALDI-ToF approach came from across the ranking list at the 96 DWP scale as determined by Protein A analysis and none would have been considered for progressing further using the classical screening approach (Porter et al., 2010b).

4.0 DISCUSSION

Existing methods for the prediction of the productivity characteristics of recombinant mammalian cell lines in larger scale fed-batch bioreactors during early CLC rely heavily on the assessment of recombinant product concentrations in the early phases and subsequently throughout the remainder of

the CLC. Whilst these approaches generally result in the isolation of recombinant cell lines with acceptable productivities, the predictive power of this approach is limited with the productivities from cell lines at the earlier stages of CLC not necessarily reflecting those observed in the bioreactor at larger-scale. As such, the productivity ranking of clones early in cell line development is not usually maintained through to larger scale (Porter et al., 2010a; Porter et al., 2010b) and as a result, some potentially high producers are discarded early in the process and relatively poor producing clones can also be carried through the selection process. Thus, a sufficient number of clones must be taken through the cell line generation process to ensure that a clone with desirable productivity and growth characteristics in the bioreactor is ultimately produced. The goal of this research was to address this issue by developing a rapid, high throughput method that allows the characterisation and accurate prediction of multiple clones' phenotype in the bioreactor early in the cell line development process such that high producers in the bioreactor can be identified early. The aim was not necessarily to identify and isolate the highest producing cell line, but to be able to reproducibly predict and select high producing cell lines from a screen early in the CLC process, utilising less resource than current screening technologies.

Whilst a number of previous reports have shown on a small number of cell lines that whole cell MALDI-ToF fingerprints can be correlated to productivity, to date no study has developed a fully predictive model that has then been used in an industrial setting to select and develop high producing cell lines. An intact cell MALDI-ToF mass spectrometry approach aligned with PLS-DA modelling was developed to predict the productivity of recombinant cell lines at the 10 L bioreactor scale using MALDI-ToF data collected from cell lines at the 96 DWP scale. In the described method it is assumed that the spectra at the 96DWP scale contains information that allows the prediction of the productivity of the cell line at the 10 L bioreactor scale and makes no assumptions as to whether the MALDI-ToF fingerprint changes or is the same between the two scales for a given cell line. This approach has similarities to the use of intact cell MALDI-ToF profiling of bacterial strains that is used clinically, in biodefence and in the academic environment to classify and identify bacterial strains (De Bruyne et al., 2011; Lundquist et al., 2005; Panda et al., 2013; Veloo et al., 2011; Warscheid and Fenselau, 2004). The proposed approach is straightforward and rapid and based around the development of a

historical database of MALDI-ToF spectra associated with subsequent productivity data at the 10 L scale to predict the performance of cell lines from the 96 DWP scale. A number of different CLCs and recombinant monoclonal antibodies were used to develop, refine and evaluate the approach. The data shows that the approach can be used to predict the performance of recombinant cell lines expressing different monoclonal antibodies and hence may be broadly applicable to the production of these high value biotherapeutics. The major advantages lie in the potential reduction of time to development of a production cell line (see Figure 7) as although high producing cell lines can be selected/predicted, the actual productivities of these did not prove to be better than those generated using a classic cell line construction process. In the future we anticipate that this approach and the historical database could be applied to the prediction of the productivity of other recombinant molecules. Importantly, cell lines isolated using this MALDI-ToF approach gave similar productivities to the highest producing cell lines selected using a classical selection approach and identified highly productive cell lines for mAb production early in the CLC process using substantially fewer screening steps than the classical productivity method. The approach and database of CHO cell lines we have developed therefore address some but not all of the issues reported as outstanding to apply MALDI-ToF whole cell fingerprinting to mammalian cells (Munteanu and Hopf, 2013) (at least CHO cells), including development of a rigorous standardization of the approach utilized, information on the influence of cell numbers, an extensive database of fingerprints and development of a classification algorithm based on a high number of distinct mammalian cell lines.

Two rounds of CLC and MALDI-ToF analysis were carried out at an early stage after transfection (before significant scale up had occurred). The whole cell MALDI-ToF data and final productivities for these cell lines from both CLCs was used to derive a model to predict the productivities of other cell lines using the MALDI-ToF spectra data. Although the whole cell MALDI-ToF spectra of cell lines obtained from the second CLC appeared different to those from the first, the scores plot of PC1 vs PC2 showed that much of the data from the cell lines occupied the same space but that there some data fell outside this space. This suggests that as further historical data is added to the database, the method and predictive power of the PLS-DA model may further improve. Regardless of this, the model was able to successfully predict and hence allow the isolation of high producing

recombinant cell lines in a third CLC process.

Key drivers for industry with respect to CLC are to minimize timelines to first-in-human studies and to reduce costs. Potential benefits to industry of the method described here are (a) to reduce the time from transfection to having secured research cell banks and completed bioreactor evaluation; (b) to reduce the resource required by screening fewer cell lines that are more likely to be higher producers – the resulting resource could be used to search for cell lines making product with a particular set of critical quality attributes e.g. particular glycoprofiles; (c) being able to undertake a major screen on a small number of candidate cell lines in fully instrumented bioreactors. Figure 7(a) shows a representative timeline for the generation of a clonal GS-CHO cell line in a commercial environment. Research cell banks can be secured for a panel of candidate cell lines about 12 weeks after transfection. Further characterization against the production process in, for example, minibioreactors and a final process run-through of 1 to 3 cell lines at laboratory-scale takes a further 8 weeks. Overall, transfection to completion of the laboratory-scale takes about 20 weeks. Use of the MS-fingerprint analysis method to select cell lines predicted to be high producers in laboratory-scale bioreactors eliminates the need to screen large numbers of cell lines (for example, up to 48 in the Ambr bioreactor system) against the bioreactor process. The switch to use of the MS-fingerprint has the potential to reduce the timeline by up to 7 weeks (Figure 7(b). It is more likely that cell lines that work well in cGMP processes are identified early, reducing failure rate, a major contributor to project failure in terms of being on-time or cost.

Looking to the future, the approach described here allows for the rapid characterisation and prediction of the phenotypic performance of recombinant cell lines. We stress that this allows the early prediction of productivity unlike classical models of cell line development that require many more cell lines to be carried through the cell line construction process further. A common theme between such mammalian cell systems is the potential for large variations between cells and clones with inherent heterogeneity, making the phenotypic prediction or characterisation of specific cell lines or types difficult (Ouedraogo et al., 2010). We envisage that this methodology will not only be applied in the fields of CLC/clone selection and stem cell characterisation but also will be applied in high-throughput screens to follow cell line responses to various stimuli (including drug screening), to investigate cell

line resistance to drugs, and be applied in the wider areas of mammalian cell biology in the fields of basic biology, medicine and biotechnology.

ACKNOWLEDGEMENTS. This work was supported by funding from Lonza Biologics and the Biotechnology and Biological Sciences Research Council (BBSRC), UK via an Industry interchange award. The authors are grateful to Prof John Birch and Dr James Graham (Lonza Biologics), and Mr Kevin Howland (University of Kent), for advice and discussion. The authors also acknowledge the help of current and previous members of Process Development and Analytical Services, Lonza Slough, for their assistance.

REFERENCES

Bright JJ, Claydon MA, Soufian M, Gordon DB. 2002. Rapid typing of bacteria using matrix-assisted laser desorption ionisation time-of-flight mass spectrometry and pattern recognition software. J Microbiolog Methods 48:127-138.

Clark AE, Kaleta EJ, Arora A, Wolk DM. 2013. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev 26:547-603.

Clarke C, Doolan P, Barron N, Meleady P, O'Sullivan F, Gammell P, Melville M, Leonard M, Clynes M. 2011. Predicting cell-specific productivity from CHO gene expression. J Biotechnol 151:159-165.

Clarke C, Henry M, Doolan P, Kelly S, Aherne S, Sanchez N, Kelly P, Kinsella P, Breen L, Madden SF, Zhang L, Leonard M, Clynes M, Meleady P, Barron N. 2012. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. BMC Genomics 13:656.

De Bruyne K, Slabbinck B, Waegeman W, Vauterin P, De Baets B, Vandamme P. 2011. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. Syst Appl Microbiol 34:20-29.

Doolan P, Clarke C, Kinsella P, Breen L, Meleady P, Leonard M, Zhang L, Clynes M, Aherne ST, Barron N. 2013. Transcriptomic analysis of clonal growth rate variation during CHO cell line development. J Biotechnol 166:105-113.

Feng HT, Sim LC, Wan C, Wong NS, Yang Y. 2011. Rapid characterization of protein productivity and production stability of CHO cells by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 25:1407-1412.

Feng HT, Wong NS, Sim LC, Wati L, Ho Y, Lee MM. 2010. Rapid characterization of high/low producer CHO cells using matrix-assisted laser desorption/ionization time-of-flight. Rapid Commun Mass Spectrom 24:1226-1230.

Franco CF, Mellado MC, Alves PM, Coelho AV. 2010. Monitoring virus-like particle and viral protein production by intact cell MALDI-TOF mass spectrometry. Talanta 80:1561-1568.

Hanrieder J, Wicher G, Bergquist J, Andersson M, Fex-Svenningsen A. 2011. MALDI mass spectrometry based molecular phenotyping of CNS glial cells for prediction in mammalian brain tissue. Anal Bioanal Chem 401:135-147.

Jacob NM, Kantardjieff A, Yusufi FN, Retzel EF, Mulukutla BC, Chuah SH, Yap M, Hu WS. 2010. Reaching the depth of the Chinese hamster ovary cell transcriptome. Biotechnol Bioeng 105:1002-1009.

Karger A, Bettin B, Lenk M, Mettenleiter TC. 2010. Rapid characterisation of cell cultures by matrix-assisted laser desorption/ionisation mass spectrometric typing. J Virol Methods 164:116-121.

Kim JY, Kim YG, Lee GM. 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl Microbiol Biotechnol 93:917-930.

Lundquist M, Caspersen MB, Wikstrom P, Forsman M. 2005. Discrimination of Francisella tularensis subspecies using surface enhanced laser desorption ionization mass spectrometry and multivariate data analysis. FEMS Microbiol Lett 243:303-310.

Marvin-Guy LF, Duncan P, Wagniere S, Antille N, Porta N, Affolter M, Kussmann M. 2008. Rapid identification of differentiation markers from whole epithelial cells by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry and statistical analysis. Rapid Commun Mass Spectro 22:1099-1108.

Mead EJ, Chiverton LM, Smales CM, von der Haar T. 2009. Identification of the limitations on recombinant gene expression in CHO cell lines with varying luciferase production rates. Biotechnol Bioeng 102: 1593-1602.

Mead EJ, Chiverton LM, Spurgeon SK, Martin EB, Montague GA, Smales CM, von der Haar T. 2012. Experimental and in silico modelling analyses of the gene expression pathway for recombinant antibody and by-product production in NS0 cell lines. PLoS One 7:e47422.

Meleady P, Doolan P, Henry M, Barron N, Keenan J, O'Sullivan F, Clarke C, Gammell P, Melville MW, Leonard M, Clynes M. 2011. Sustained productivity in recombinant Chinese hamster ovary (CHO) cell lines: proteome analysis of the molecular basis for a process-related

phenotype. BMC Biotechnol 11:78.
Monchamp P, Andrade-Cetto L, Zhang JY, Henson R. 2007. Signal processing methods for mass spectrometry. In: Alterovitz G, Ramoni MF, eds. Systems Bioinformatics: An Engineering Case-Based Approach. London: Artech House Publishers.
Munteanu B, von Reitzenstein C, Hansch GM, Meyer B, Hopf C. 2012. Sensitive, robust and automated protein analysis of cell differentiation and of primary human blood cells by intact cell MALDI mass spectrometry biotyping. Anal Bioanal Chem 404:2277-2286.
Ouedraogo R, Flaudrops C, Ben Amara A, Capo C, Raoult D, Mege JL. 2010. Global analysis of circulating immune cells by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. PloS One 5:e13691.
Panda A, Kurapati S, Samantaray JC, Myneedu VP, Verma A, Srinivasan A, Ahmad H, Behera D, Singh UB. 2013. Rapid identification of clinical mycobacterial isolates by protein profiling using matrix assisted laser desorption ionization-time of flight mass spectrometry. Indian J Med Microbiol 31:117-122.
Porter AJ, Dickson AJ, Racher AJ. 2010a. Strategies for selecting recombinant CHO cell lines for cGMP manufacturing: realizing the potential in bioreactors. Biotechnol Prog 26:1446-1454.
Porter AJ, Racher AJ, Preziosi R, Dickson AJ. 2010b. Strategies for selecting recombinant CHO cell lines for cGMP manufacturing: improving the efficiency of cell line generation. Biotechnol Prog 26:1455-1464.
Sanchez N, Kelly P, Gallagher C, Lao NT, Clarke C, Clynes M, Barron N. 2014. CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. Biotechnol J 9:396-404.
Selvarasu S, Ho YS, Chong WP, Wong NS, Yusufi FN, Lee YY, Yap MG, Lee DY. 2012. Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnol Bioeng 109:1415-1429.
Smales CM, Dinnis DM, Stansfield SH, Alete DE, Sage EA, Birch JR, Racher AJ, Marshall CT, James DC. 2004. Comparative proteomic analysis of GS-NS0 murine myeloma cell lines with varying recombinant monoclonal antibody production rate. Biotechnol Bioeng 88:474-488.
Sogawa K, Watanabe M, Sato K, Segawa S, Ishii C, Miyabe A, Murata S, Saito T, Nomura F. 2011. Use of the MALDI BioTyper system with MALDI–TOF mass spectrometry for rapid identification of microorganisms. Anal Bioanal Chem:1-7.
Tait AS, Hogwood CE, Smales CM, Bracewell DG. 2012. Host cell protein dynamics in the supernatant of a mAb producing CHO cell line. Biotechnol Bioeng 109:971-982.
Veloo AC, Knoester M, Degener JE, Kuijper EJ. 2011. Comparison of two matrix-assisted laser desorption ionisation-time of flight mass spectrometry methods for the identification of clinically relevant anaerobic bacteria. Clin Microbiol Infect 17:1501-1506.
Walsh G. 2010. Biopharmaceutical benchmarks 2010. Nature Biotechnol 28:917-924.
Warscheid B, Fenselau C. 2004. A targeted proteomics approach to the rapid identification of bacterial cell mixtures by matrix-assisted laser desorption/ionization mass spectrometry. Proteomics 4:2877-2892.
Wold S, Hellberg STL, Sjostrom M, Wold H. 1987. PLS model building: theory and applications, PLS modeling with latent variables in two or more dimensions. Frankfurt am Main.
Zhang X, Scalf M, Berggren TW, Westphall MS, Smith LM. 2006. Identification of mammalian cell lines using MALDI-TOF and LC-ESI-MS/MS mass spectrometry. J Am Soc Mass Spectrom 17:490-499.

Table 1. PLS-DA predicted classification for samples as 'high' class (> 4000 mg/L) for cell line construction 2 using MALDI-ToF data at 96 DWP scale to predict 10 L bioreactor productivity as determined by HPLC Protein-A titres. Cell lines from CLC3 predicted as being 'high' producers by either the MALDI-ToF PLS-DA classification method or a classical ELISA based method at the 96 deep well plate stage and the subsequent productivities of these cell lines at the 10 L bioreactor scale are also reported.

Cell Line Construction 2 (CLC2)	**Y Pred Probability (High > 4000mg/L)**		**10 L Bioreactor Titre (mg/L)**
CL016B5	No		116
CL016F11	No		2519
CL033D5	Yes		6024
CL033G5	Yes		4155
CL948G2	No		1592
Cell Line Constriction 3 (CLC3)	**Method for Selection of Cell Line**	**Antibody Concentration in 96 DWP Screen (mg/L)**	**10 L Bioreactor Titre (mg/L)**
H03	Classical	6540	5218
D08	MALDI-ToF	4280	4949
D06	Classical	6430	4917
A09	MALDI-ToF	3870	4479
G12	MALDI-ToF	3720	4341
C11	MALDI-ToF	3010	3791
F10	MALDI-ToF	4270	3677
G05	MALDI-ToF	1460	2449
G04	MALDI-ToF	1820	2438
C04	MALDI-ToF	1250	1918

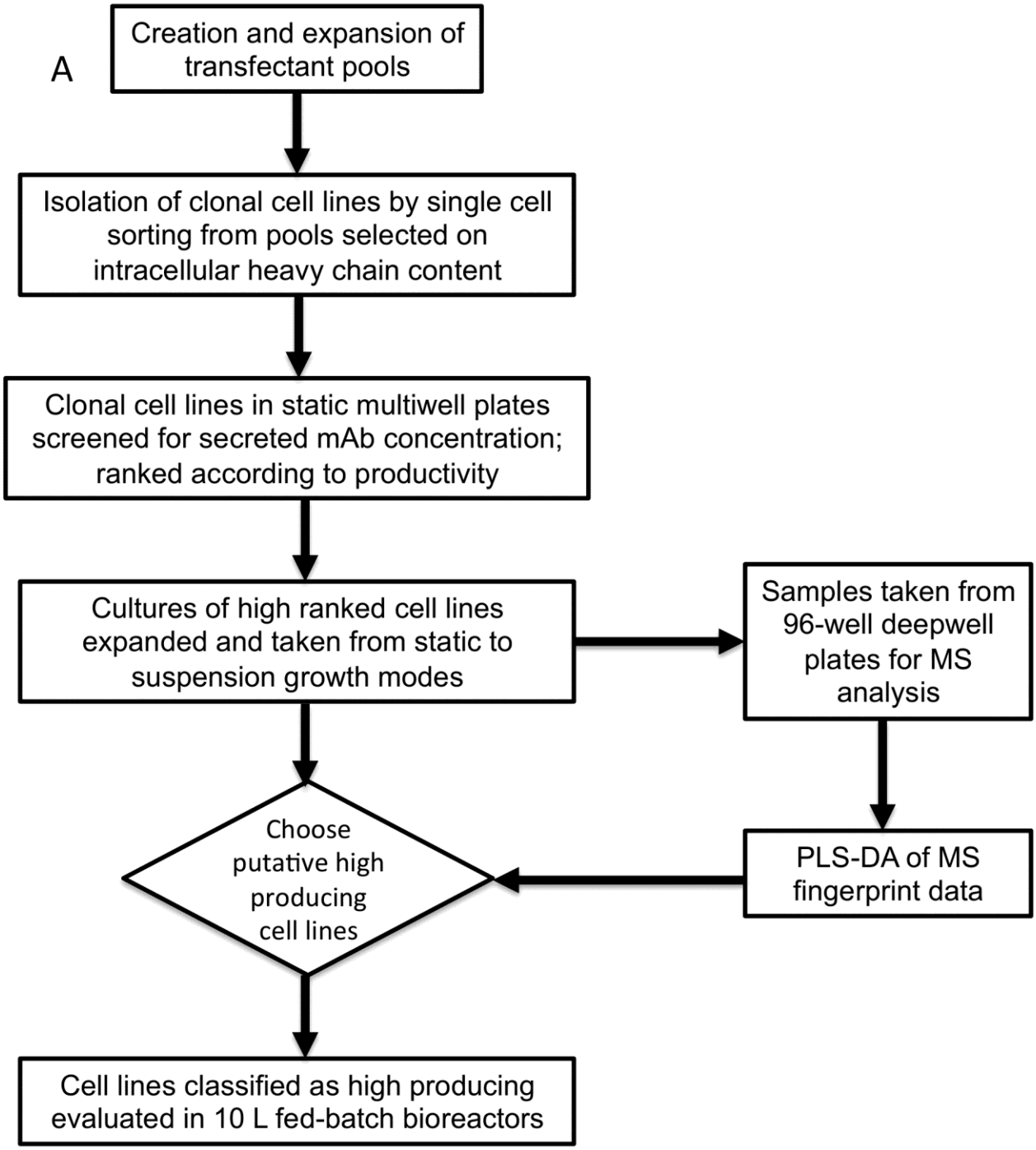

Figure 1. The classical cell line construction workflow utilzed for the construction of naïve recombinant GS-CHOK1SV cell lines expressing various monoclonal antibodies (mAb) in this study indicating where samples were taken for whole cell MALDI-ToF mass spectrometry analysis **(A)** and the subsequent application of a database PLS-DA modelling approach to classify unknown cell lines as either high or low mAb producers **(B)**.

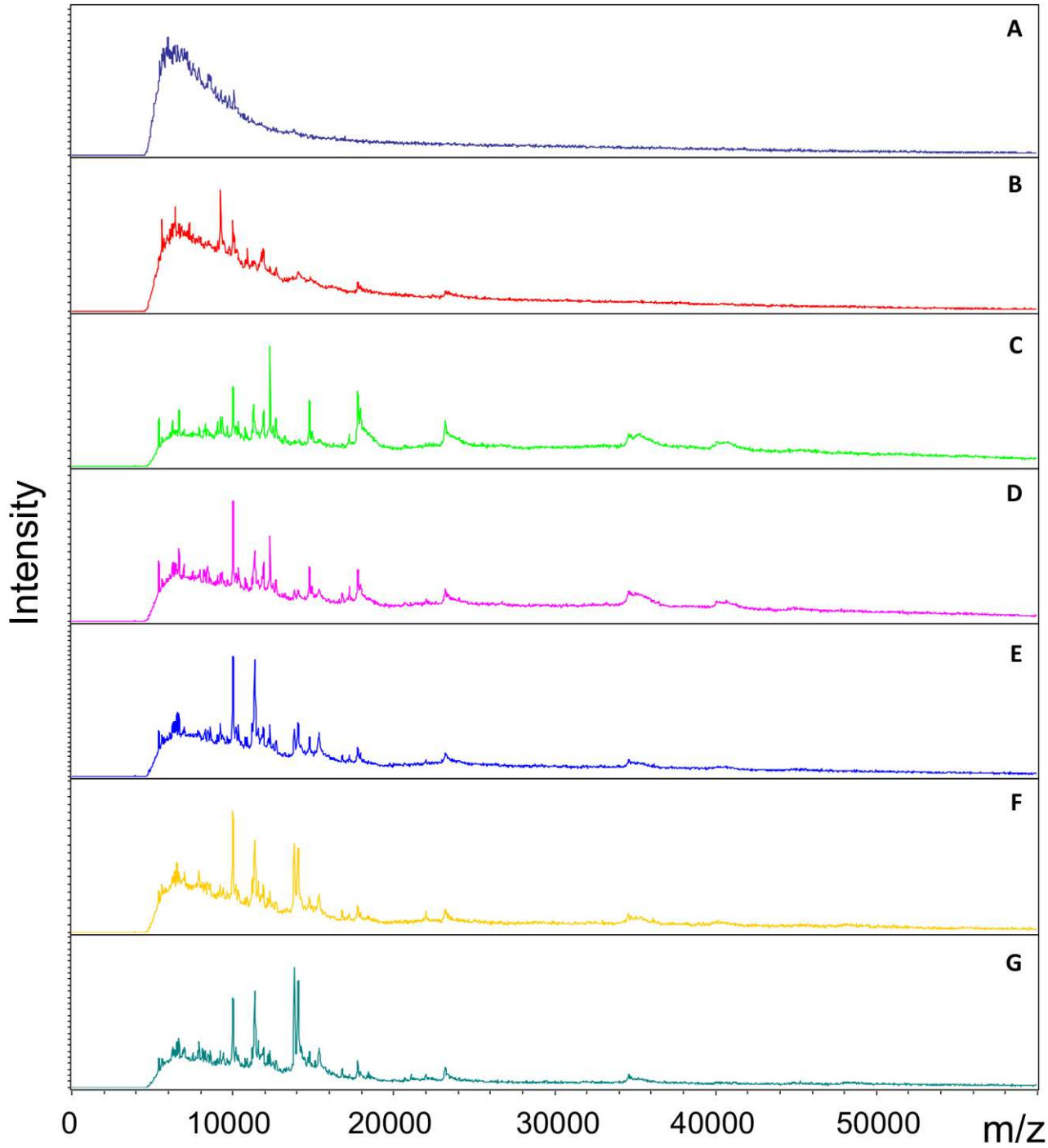

Figure 2. Representative raw MALDI-ToF mass spectrometry data with varying cell number. The cell numbers used for analysis are as follows; **(A)** $3x10^6$ viable cells, **(B)** $2x10^6$ viable cells, **(C)** $1x10^6$ viable cells, **(D)** $0.5x10^6$ viable cells, **(E)** $0.25x10^6$ viable cells, **(F)** $0.125x10^6$ viable cells, and **(G)** $0.0625x10^6$ viable cells. Each cell sample was taken up in 50 µl of matrix buffer and 1 µl was spotted on the MALDI-ToF target plate for analysis as described in the methods section.

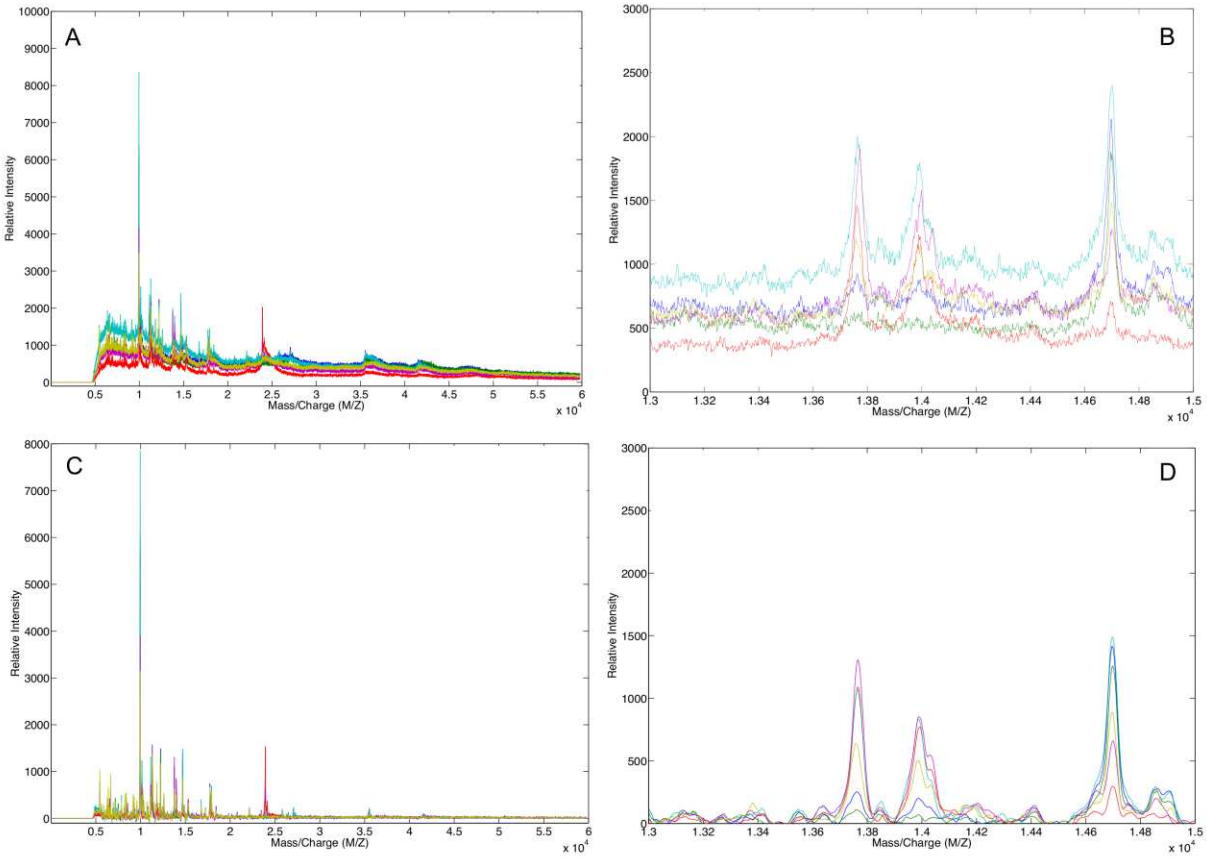

Figure 3. Representaive MALDI-ToF mass spectrometry data before and after pre-processing before analysis. Representative raw mass spectra (**Figure 3A**) were subjected to the following pre-processing steps in the order indicated. (1) Resampling (up-sample to 50,000 data points), (2) Baseline Correction (removes the effect of noise introduced by the matrix), (3) Filtering (application of Savitzky-Golay filter to smooth the signal), (4) Alignment (automatically selects peak and align spectra based on height and over-segmentation filters), (5) Normalisation (normalisation of area under curve for replicate samples). The influence of pre-processing on the entire spectrum is shown in **Figure (C)**. Figures **(B)** and **(D)** show a zoomed in area of a selection of spectra before (**B**) and after (**D**) pre-processing. Different representative spectra are shown in different colours.

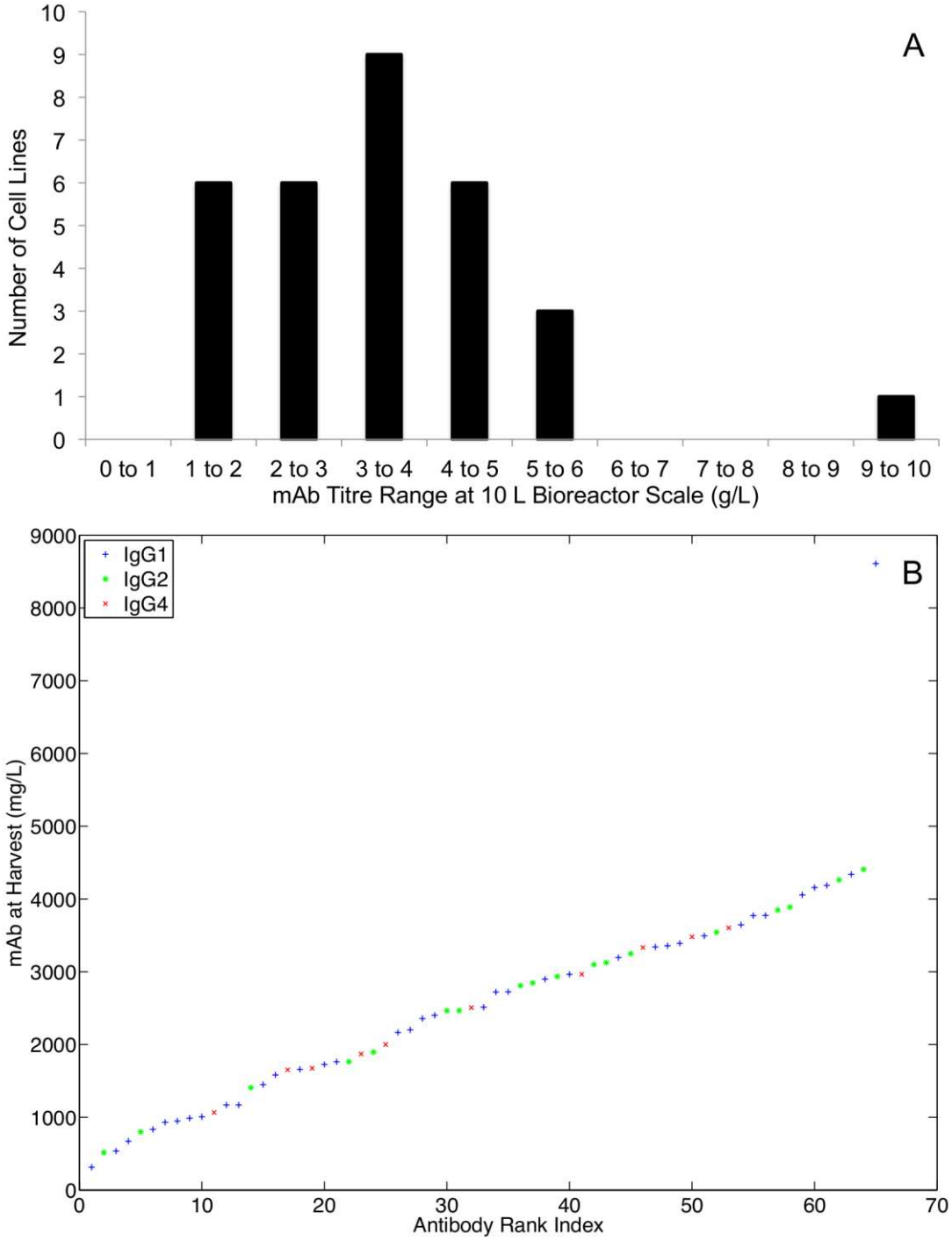

Figure 4. (A) The distribution of monoclonal antibody (mAb) productivity in 10 L fed-batch bioreactors of the recombinant CHO-K1SV cell lines from cell line construction 1 of which 29 were used to generate the initial PLS-DA models. These cell lines were selected at the 96 DWP stage to include the highest producers, but also a pseudo-random selection of low and intermediate producers to ensure coverage of the original productivity distribution so that MALDI-ToF fingerprints representative of the range of productivities were available to generate the initial PLS-DA models.

MALDI-ToF mass spectra were collected for each cell line at the 96 DWP stage and then these cell line progressed and grown over a period of 14 days under fed-batch conditions in 10 L bioreactors before being harvested and the productivity determined by Protein A HPLC to provide a training set for modelling. Two cell lines were excluded from the model, as the productivity data available was from later in culture than day 14. **(B)** Although we set a classification boundary to indicate cell lines with 'high' productivity at 4000 mg/L as the antibody used in cell line construction 1 was known to be well expressed, the actual expression level of high producers will differ between antibodies as demonstrated by historical fed-batch data from Lonza Biologics for different antibodies (see **Figure 4B**). Hence, the class boundary could be adjusted for other molecules that are expressed at lower concentrations.

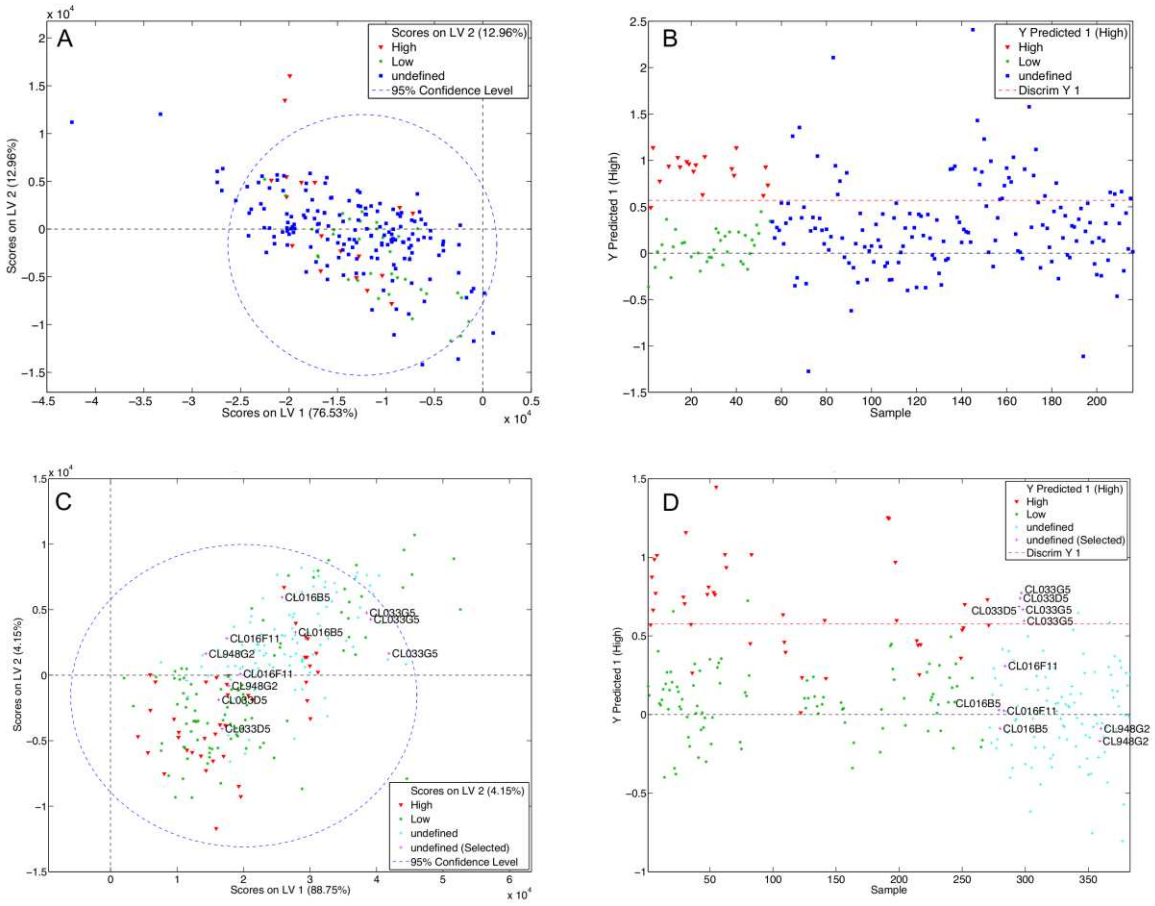

Figure 5. A PLS-DA model was initially generated using a training set comprising 29 cell lines from CLC1 whereby the 10 L bioreactor scale productivities and relevant 96 DWP whole cell MALDI-ToF data was available using a classification of high productivity as ≥4000 mg/L. Once the PLS-DA model had been generated, the MALDI-ToF data from the remaining CLC1 samples (those from the original 119 not used in building the model) were presented to the model to predict their position relative to the classification boundary. **Figure 5A** shows the PLS-DA latent variables (LV) scores plot of LV1 vs. LV2 and **Figure 5B** shows the position of the discrimination boundary and the cell lines predicted as belonging to each class as determined by the PLS-DA model comprising 5 LVs of the CLC1 cell lines of unknown productivity (shown in blue). **Figures 5C** and **5D** show the PLS-DA latent variables (LV) scores plot of LV1 vs. LV2 (**5C**) and position of the discrimination boundary and the cell lines predicted as belonging to each class as determined by the PLS-DA model (**5D**) for the CLC2 cell lines as determined by the model trained using the data from both CLC1 and CLC2 using an 8 LV model as described in the text. A selection of cell lines from CLC2 was then taken for analysis at the 10 L bioreactor scale based upon the class they were predicted to belong to and these are labelled in **5C** and

5D. Note that replicate MALDI-ToF data is represented in the figures, cell line data used to build the PLS-DA model is coloured red (high producers ≥4000 mg/mL) or green (low producers < 4000 mg/mL) whilst unknown samples are shown in blue.

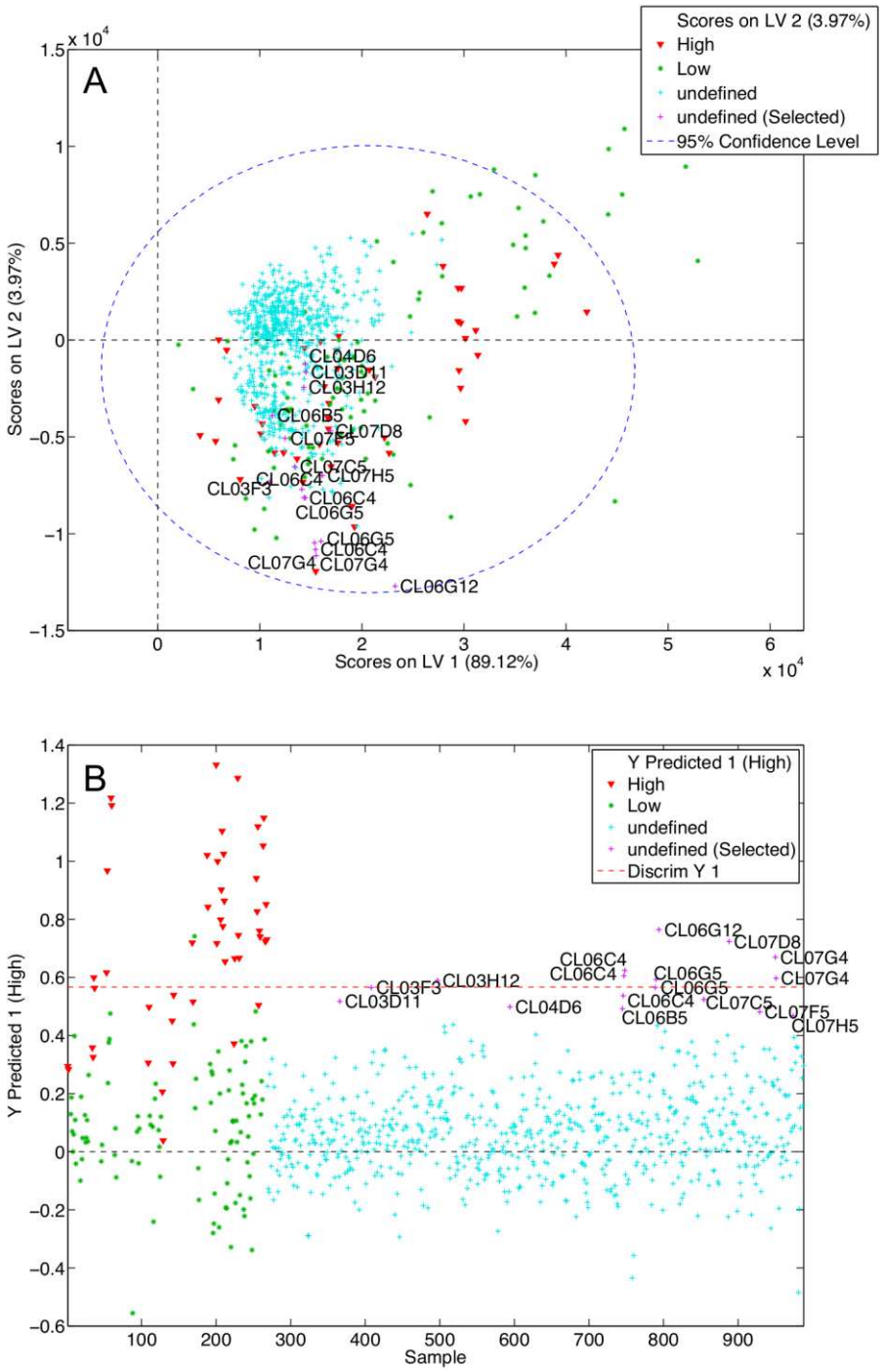

Figure 6. The PLS-DA latent variables (LV) scores plot of LV1 vs. LV2 (**A**) and position of the discrimination boundary and the cell lines predicted as belonging to each class as determined by the PLS-DA model (**B**) for the CLC3 cell lines as determined by the model trained using the data from both CLC1 and CLC2, a classification boundary of high as ≥4000 mg/L and an 8 LV model as described in the text. From the PLS-DA model 8 cell lines were selected to progress through to a 10 L fed-batch bioreactor evaluation (labeled on **Figure 6B**).

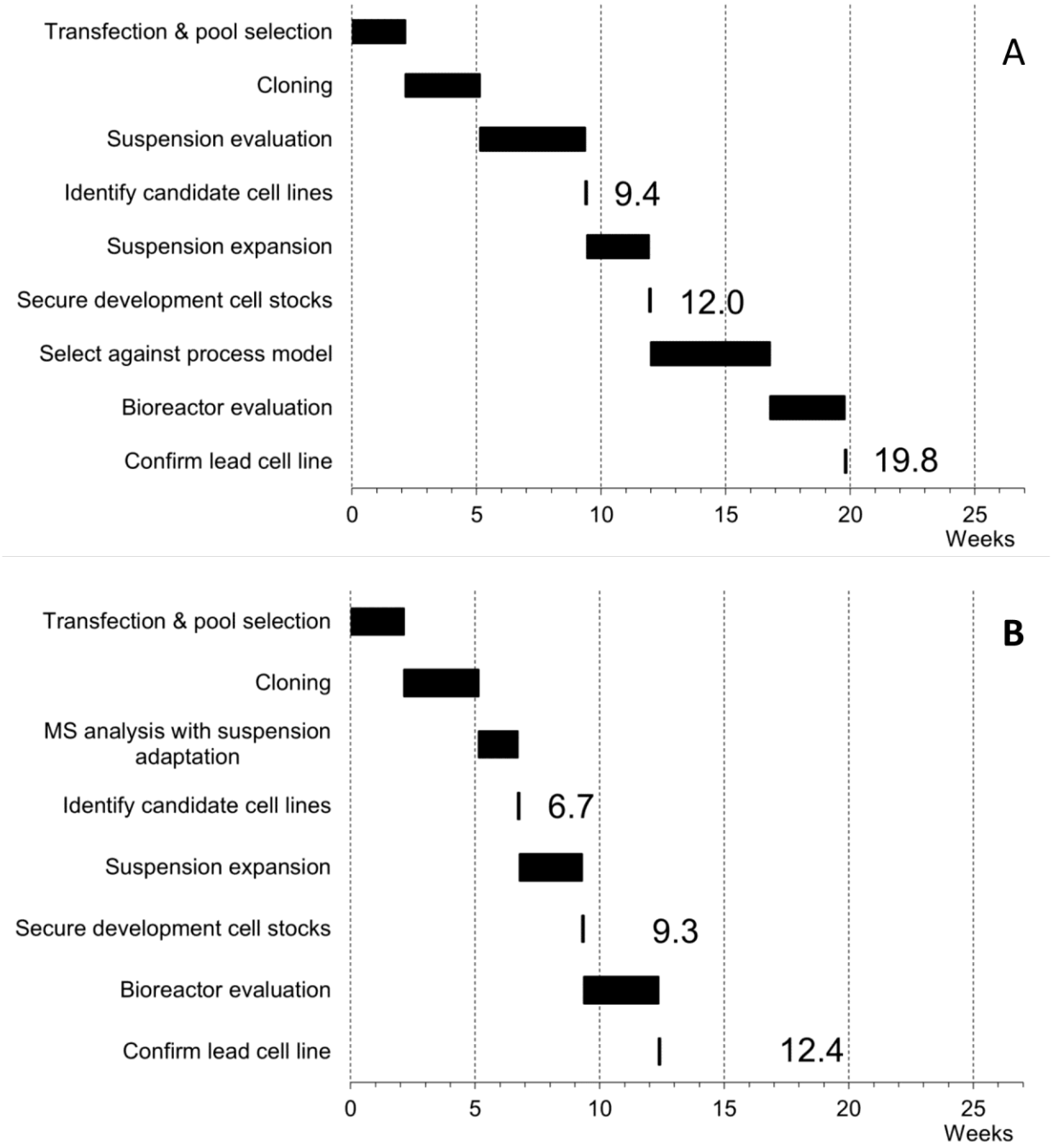

Figure 7. Representative timelines for the generation of clonal GS-CHO cell lines. (**A**) Timeline for generation of cell lines using the classical approach. (**B**) Timeline using the MALDI-ToF mass spectrometry fingerprinting approach described in this manuscript.