THESIS OF COTUTELLE DOCTORATE

Presented to

The Faculty of the Graduate School *Science et Ingénerie des Molécules, des Procédés, des Produits et de l'Énergie (SIMPPÉ)*; de l'Université de Lorraine & the Faculty of the Graduate School *Unidad de Investigación y Desarrollo en Alimentos (UNIDA)*; del TecNM/Instituto Tecnológico de Veracruz (ITVer)

In Partial Fulfillment of the Requirements for the Degree of

Docteur de l'Université de Lorraine
Mention : « Procédés Biotechnologiques »

Doctor del TecNM/Instituto Tecnológico de Veracruz
Especialidad: Ciencias en Alimentos
By

Daniel Arturo ZAVALA ORTIZ

Application of *in situ* near-infrared spectroscopy (NIRS) for monitoring biopharmaceuticals production by cell cultures

---

Dissertation Defense Schedule : April 29, 2020.

Rapporteurs : CACHON Remy – Université de Bourgogne
GSCHAEDLER MATHIS Anne Christine – Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco – CIATEJ

Examinateurs: OLIART ROS Rosa María– TecNM/Instituto tecnológico de Veracruz
GONZALEZ RIOS Oscar– TecNM/Instituto Tecnológico de Veracruz
GUEDON Emmanuel – Université de Lorraine (Directeur de thèse)
EBEL Bruno – Université de Lorraine (Co-directeur de thèse)
AGUILAR USCANGA Maria Guadalupe – TecNM/Instituto Tecnológico de Veracruz (Directora de tesis)

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Veracruz
Subdirección Académica
División de Estudios de Posgrado e Investigación

"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

H. Veracruz, Ver., 13/marzo/2020
OFICIO No. ITV/DEPI/072/2020

Asunto: Autorización de impresión

C. DANIEL ARTURO ZAVALA ORTIZ
PRESENTE

Por este conducto se le comunica que su tesis de Doctorado en Ciencias en Alimentos titulada:

"Application of *in situ* near infra-red spectroscopy (NIRS) for monitoring biopharmaceuticals production by cell cultures"

Fue aprobada en su totalidad en cuanto a formato y calidad del contenido a satisfacción del H. Jurado del examen, por lo que está Usted autorizado a editar la presentación definitiva del trabajo.

ATENTAMENTE
Excelencia en Educación Tecnológica®
Aprendo y Lucho Fuego Permanente

BENIGNO ORTIZ MUÑIZ
JEFE DE LA DIV. EST. POSG. E INVEST

MARÍA GUADALUPE AGUILAR USCANGA
PRESIDENTE

ROSA MARÍA OLIART ROS
SECRETARIO

EMMANUEL GUEDON
VOCAL 1

BRUNO EBEL
VOCAL 2

REMY CACHON
VOCAL 3

ANNE CHRISTINE GSCHAEDLER MATHIS
SUPLENTE 1

ÓSCAR GONZÁLEZ RÍOS
SUPLENTE 2

SEP | TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE VERACRUZ
DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

Calz. Miguel Ángel de Quevedo 2779, Col. Formando Hogar
C.P. 91897, Veracruz, Ver. Tel. (229) 934 1500
www.veracruz.tecnm.mx

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Veracruz
Subdirección Académica
División de Estudios de Posgrado e Investigación

"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

El presente trabajo de tesis titulado **"Application of *in situ* near infra-red spectroscopy (NIRS) for monitoring biopharmaceuticals production by cell cultures"** presentado por el (la) C. **Daniel Arturo Zavala Ortiz** ha sido aprobado por su comité tutorial como requisito parcial para obtener el grado de:

DOCTOR EN CIENCIAS EN ALIMENTOS

DIRECTORA	**MARÍA GUADALUPE AGUILAR USCANGA**
CO-DIRECTOR	**EMMANUEL GUEDON**
CO-DIRECTOR	**BRUNO EBEL**
REVISOR	**REMY CACHON**
REVISORA	**ROSA MARÍA OLIART ROS**
REVISORA	**ANNE CHRISTINE GSCHAEDLER MATHIS**
REVISOR	**ÓSCAR GONZÁLEZ RÍOS**

Calz. Miguel Angel de Quevedo 2779, Col. Formando Hogar
C.P. 91897, Veracruz, Ver. Tel. (229) 934 1500
www.veracruz.tecnm.mx

DEDICATIONS

A mis padres, mi esposa y mi familia…

ACKNOWLEDGMENTS

The doctoral project was only possible with the support and facilities of the French and Mexican peoples through different institutions. In Mexico, to CONACyT and ITVer for a scholarship support and tuition scholarships. In France, to the Ministry of Europe and foreign affairs for an Eiffel excellence scholarship and to l'UL for a tuition scholarship for the last year.

In France, part of the work presented in this thesis was performed in the Laboratoire Réactions et Génie des Procédés (LRGP), in the Bioprocédés-Biomolécules team (BioProMo), at Université de Lorraine. I therefore express gratitude to Dr. Laurent Falk for hosting me at LRGP during these four years of thesis. In Mexico, part of the thesis work was performed in Bioingeniería Laboratory in the Unidad de Investigación y Desarrollo en Alimentos (UNIDA). I therefore express my gratitude to UNIDA for also hosting me.

I extend my deep gratitude to Remy CACHON, PhD, from Université de Bourgogne, Anne GSCHAEDLER, PhD from CIATEJ, for accepting judge this work as rapporteurs. I also appreciate the support of Prof. Rosa María OLIART ROS, and Prof. Oscar GONZALEZ RIOS from TecNM/ITVer for accepting to review the thesis manuscript.

Je souhaite aussi exprimer ma profonde reconnaissance à mes directeurs de thèse M. Bruno EBEL et M. Emmanuel GUEDON, pour leur encadrement et leurs conseils scientifiques et personnels, mais également pour la confiance et le temps qu'ils ont consacrés pour moi. Je dois exprimer ma profonde reconnaissance à Mme Annie Marc qui a également contribué à ce que ce travail porte ses fruits. Agradezco a mi directora de tesis, Sra. Maria Guadalupe Aguilar Uscanga, por sus inconmesurables muestras de apoyo, confianza y dedicación. A las sras. Patricia Hayward y Dulce María Barradas Dermitz por su dedicación, rigor científico y apoyo en la lengua inglesa.

.

The doctoral thesis has been an invaluable treasure that has changed my mind and my heart. During this journey I went through luminous and gloomy paths... Fortunately, I was always surrounded by good people on both sides of the Atlantic: Gaby, Fofo, Nito, Yerarli and Mauro in Mexico et en France, Léna, Juju le perroquet, Caro la divine, Amani, Aurelia la meilleure, M. Panini Poulet (Jonathan), et Céline. Je remercie M. Fabrice Blanchard pour son support technique sur les bioréacteurs. Je dois aussi remercier les amis que je me suis faits en France, et en particulier ma petite sœur Meng-Yao Li, et mon grand frère Chafik Bourkaib, qui m'ont beaucoup aidé, écouté, réconforté et accompagné durant cette thèse. Je vous remercie tous pour tous les bons moments passés ensembles, les soirées, le ski, les sorties en vélo, les brocantes, les randonnées, l'échange académique et l'aide avec la langue.

Agradezco a mis padres y mi familia, que siempre me han ayudado y cuidado. A mi esposa Mary, por tu comprensión, amor y apoyo incondicional, a Bolo y Nala que sufrieron la lejanía.

INDEX

Figure Index

Table Index

Abbreviations / Abréviations / Abreviaturas

1-Naphthaleneacetic acid (NAA)

2, 4-Dichlorophenoxyacetic acid (2, 4-D)

6-Benzylaminopurine (BAP)

Absolute value (Abs)

Active Pharmaceutical Ingredients (API)

Ammo (Ammonium)

Analysis of Variance (ANOVA)

Antibodies (Ab)

Antibody-dependent cellular cytotoxicity (ADCC)

antigen binding fragment (Fab)

Artificial Neural Net Regression (ANNR)

Cambial Meristem cells (CMCs)

Catharanthus roseus (*C.r.* or *Catharanthus r.*)

Cell dry weight (DW)

Chinese hamster ovary (CHO)

Coefficient of determination (R_2)

Critical Process Parameter (CPP)

Critical Quality Attributes (CQA)

Cross validation (CV)

Deoxyribonucleic acid (DNA)

Dihydrofolate reductase (DHFR)

Dimethylsulfoxide (DMSO)

Dissolved Oxygen (DO)

Electrospray ionization mode (ESI+)

Energy required for vibration (E)

European Medicines Agency (EMA)

Extended Multiplicative Signal Correction (EMSC)

Fetal bovine serum (FBS)

Fluorescence-activated cell sorting (FACS)

Fourier scores (FS)

Fragment crystallizable region (Fc)

Gaussian radial basis function (RBF)

Gibberellic acid (GA_3)

Gln (Glutamine)

Glu (Glucose)

Glycoforms containing Fucose (F-glycoforms)

Glycoforms containing galactose (G-glycoforms)

Glycoforms containing sialic acid (S-glycoforms)

Glycoforms of high mannose structures (HM-glycoforms)

Glycosylated mAb (G-mAb)

Goodness of estimation (Q2Y)

Green Fluorescent Protein (GFP)

High-performance liquid chromatography-Mass spectroscopy (HPLC-MS)

Host Cell Proteins (HCP)

Immunoglobulins (Ig)

Interference RNA (iRNA)

International Council for Harmonization (ICH)

Interval and iterative PLSR (iPLSR)

Kinetine (Kin)

Latent Variable (LV)

LDH (Lactate Deshydrogenase)

Linear Trap Quadripole (LTQ)

Local points (LP)

Locally weighted regression (LWR)

LOcally WEighted Scatter-plot Smoother (LOWESS)

Mean of residuals (MoR)

Methylerythritol 4-phosphate (MEP)

Monoclonal antibodies (mAb)

Multiplicative Scatter Correction (MSC)

Murashige and Skoog culture media (MS)

Near Infra-red (NIR)

Near infra-red Spectroscopy (NIRS)

Neurons (N)

Nicotinamide adenine dinucleotide (NAD or NADH)

Non-glycosylated monoclonal antibody (NG-mAb)

Orthogonal Scatter Correction (OSC)

Packed Cell Volume (PCV)

Partial Least Squares Regression (PLSR)

plant growth regulators (PGR)

Prediction (PRE)

Principal Component (PC)

Principal Component Analysis (PCA)

Principal Component Regression (PCR)

Probabilistic Quotient Normalization (PQN)

Process Analytical Technology (PAT)

Quality by Design (QbD)

Quality by testing (QbT)

Quality Target Product Profile (QTPP)

Relative centrifuge force (RCF)

Relative Error (RE)

Revolutions per minute (rpm)

Ribonucleic acid (RNA)

Root mean square error of cross validation (RMSECV)

Root mean squared error (RMSE)

Root mean squared error of prediction (RMSEP)

Spectra truncation (ST)

Standard error of cross-validation (SECV)

Standard error of prediction (SEP)

Standard Normal Variate (SNV)

Support Vector (SV)

Support Vector Regression (SVR)

U.S. Food and Drug Administration (FDA)

VCD (Viable cell Density)

Vinblastine (VB)

Vincristine (VC)

Wavelength (λ)

Nomenclature / Nomenclatura

A	Absorbance
I_0	Incident light
I	Transmitted light
ε	Molar attenuation coefficient or absorptivity
C	Concentration of the attenuating species
l	Optical path length
b_k	k-th regressor in-linear regression methods
x_k	k-th indetermined variable related to spectra in-linear regression methods
$\hat{y}_i$	Estimated or predicted concentration value by models for the i-th sample
y_i	Actual concentration value for the i-th sample
$\bar{y}_i$	Mean concentration value in calibration analysis
N_p	Number of calibration or prediction samples
X or [X]	Matrix of spectra
T	Scores matrix (for PCR) or pseudo-score matrix (for PLSR)
P	Matrix of loadings (for PCR) or matrix of pseudo-loadings (for [X] in PLSR)
[E], E or [F]	Error matrices in-linear models
Y or [Y]	Matrix of off-line concentration values
B	Regressor matrix for linear models
U	Pseudo-score matrix
Q	Matrix of pseudo-loadings for [Y] in PLSR
$w_i(x_j)$	i-th weight for the calibration sample j in LWR models
$\delta\ (x_j, x_i)$	Distance between prediction sample j and calibration sample i in LWR
$d(x_j)$	Maximum distance involved in each regression in LWR
ρ_{ip}	Distance function for LWR2 models
Y_{ip}	Distance in the chemical space for LWR2 models
X_{ip}	Distance in the spectral-related space for LWR2 models
$\pm\varepsilon$	SVR threshold or precision
$f(x)$	Hyperplane for SVR models

$\vec{w}$	Intuitive regressor for SVR model
b	Error for intuitive SVR model
ξ_i	Slack variables for soft margin
F	Feature space of SVR model
$\vec{p}$	Input for ANN
$\vec{w_i}$	Weight of neuron *i*
b_i	Bias value for neuron *i*
f_i	Transfer function for neuron *i*
a	Neuron output
$Y_{x/s}$	Yield of cell dry weight per total sugar consumed ($g.g^{-1}$)

ABSTRACT

The complexity of biopharmaceutical products implies that their approval is based on a specific process. Any further change, especially in the process, requires drug validation in terms of clinical effects and biosecurity. Because changes in the processes may be unavoidable, quality assurance by inspection at the end of the process (Quality by Testing-QbT) tends to be replaced by a new quality perspective called Quality by design (QbD) which builds drug Critical Quality Attributes (CQA) controlling key Critical Process Parameters (CPP) in real-time. However, QbD implementation has been limited by the complexity of cell culture processes and the need for multivariate methods that allow the use of complex signals from process analyzers as monitoring instruments.

Consequently, the objective of this work has been to develop new methodological and experimental applications, based on *in situ* NIR spectroscopy, for real-time monitoring of biopharmaceutical-producing cell cultures using two production platforms: animal cells (CHO-250-9) producing monoclonal antibodies (mAb) and plant cells (*Cantharanthus roseus*) producing antineoplastic molecules (vincristine-VC and vinblastine-VB).

First, a process capable of producing VC and VB was generated, cell differentiation was identified as CPP and the ability to monitor it by *in situ* NIR spectroscopy was firstly demonstrated using calibration models based on partial least squares regression (PLSR). Subsequently in CHO cell cultures, different regression techniques were evaluated to generate calibration models to monitor CPP and CQA. PLSR was inadequate because of the chemical and physical variability that CHO cell cultures present during the different phases of batch culture. Local Weighted Regression (LWR) was adequate to monitor classic CPP (concentration of glucose, lactate, and viable cells, amongst others) since it adequately handled the variability associated with the progression of cell culture. However, for the glycosylation profile (CQA), it was unable to properly handle the complex nonlinear relationships between NIR spectra and the concentration of various monoclonal antibody (mAb) glycoforms. This was overcome with the use of models based on support vector regressions (SVR), allowing the generation of models of different mAb glycoforms related to particular clinical effects. Globally, this work has contributed to the expansion of the capabilities of *in situ* NIR spectroscopy for the monitoring of classic CPP in a more precise

way, new innovative CPP such as cell physiological state in plant suspension cultures, and CQA such as mAb glycosylation profiles linked to clinical characteristics in animal cell cultures.

RÉSUMÉ

La complexité des médicaments biopharmaceutiques implique que leur validation par les instances réglementaires nécessite un processus spécifique. Tout changement, notamment du procédé, nécessite une nouvelle validation en termes d'efficacité clinique et de sécurité pour le patient. Puisque les changements dans le procédé sont parfois inévitables, la qualité du produit n'a plus à être évaluée uniquement en fin de procédé de production (Quality by testing, QbT), mais tout au long du procédé, et conceptualisée dans toutes les étapes de fabrication (Quality by Design, QbD). Cette démarche préconise de contrôler les paramètres critiques du procédé (CPP) en temps réel afin de maintenir les attributs de qualité critiques (CQA) dans une zone de confiance préalablement définie. Cependant, la mise en œuvre du QbD est actuellement limitée dans l'industrie biopharmaceutique en raison de la complexité des procédés de culture cellulaire ainsi que par la nécessité d'utiliser des méthodes d'analyse multivariée de données issues des analyseurs du procédé (*i.e.* méthodes spectroscopiques). **L'objectif de ce travail a dont été de développer de nouvelles applications méthodologiques et expérimentales, basées sur la spectroscopie proche infrarouge (NIR) *in situ*, pour le suivi en temps réel de cultures de cellules produisant des biopharmaceutiques. Pour cela, deux modèles cellulaires ont été étudiés : des cellules de hamster chinois (CHO) produisant un anticorps monoclonal (mAb) et des cellules de plantes (*Cantharanthus roseus*) produisant des molécules anticancéreuses (la vincristine, VC et la vinblastine, VB).**

Dans un premier temps, un procédé permettant la production de VC et de VB a été développé. La différentiation cellulaire de *Catharanthus roseus* ayant été identifiée comme un CPP, son suivi en ligne a été rendu possible grâce à l'utilisation combinée de la spectroscopie NIR et de modèles de calibration basés sur la régression des moindres carrés partiels (PLS). Dans un second temps, pour les cultures de cellules CHO, différentes techniques de régressions ont été évaluées pour générer des modèles de calibration permettant le suivi en ligne des CPP et des CQA. La régression PLS s'est révélée inadéquate en raison de la variabilité chimique et physique que les cellules CHO entrainent durant les différentes phases de culture. Au contraire, la régression LWR (Local Weighted Regression) a permis de suivre en temps réel des CPP conventionnels (concentration en glucose, en lactate, en cellules vivantes,…). Cette

régression permet de gérer de manière adéquate la variabilité associée à la culture cellulaire. Cependant, pour le suivi du profil de glycosylation des anticorps (CQA), cette régression n'est pas capable de gérer les relations non-linéaires existantes entre les spectres NIR et les concentrations en diverses formes d'anticorps glycosylés. Ce suivi en ligne des différentes glycoformes a été rendu possible uniquement par l'utilisation de la régression SVR (Support Vector Regressions). Ainsi, ce travail a ainsi permis l'amélioration du suivi en ligne de CPP par la spectroscopie NIR, mais également le suivi de nouveaux CPP comme l'état physiologique de cellules de plantes ou encore les différentes glycoformes des anticorps.

RESUMEN

La complejidad de los productos biofarmacéuticos implica que su aprobación sea en función de un proceso específico. Cambios o modificaciones posteriores en éste, requieren la validación del medicamento en términos de efectos clínicos y bioseguridad. Debido a que los cambios en los procesos pueden ser inevitables, el aseguramiento de la calidad por inspección-QbT (Quality by Testing) al final del proceso tiende a desaparecer por la perspectiva de calidad por diseño-QbD (Quality by Design), la cual construye la calidad del medicamento-PQA (Product Quality Attribute) o CQA (Critical Quality Attribute), controlando en tiempo real parámetros clave del proceso-CPP (Critical Process Parameter). Sin embargo, su implementación ha sido limitada por lo intrincado de los cultivos celulares y a la necesidad de métodos multivariados que permitan emplear las complejas señales de analizadores de proceso como instrumentos de seguimiento o monitoreo. **Consecuentemente, el objetivo de este trabajo ha sido desarrollar nuevas aplicaciones metodológicas y experimentales, basadas en la espectroscopía NIR *in situ*, para el seguimiento en tiempo real de cultivos celulares productores de biofarmacéuticos empleando dos plataformas de producción: células animales (CHO-250-9) productoras de anticuerpos monoclonales (mAb) y células vegetales (*Cantharanthus roseus*) productoras de antitumorales (vincristina-VC y vinblastinaVB).**

Primeramente, se generó un proceso capaz de producir VC y VB, se identificó como CPP a la diferenciación celular y se demostró de forma preliminar a nivel biorreactor, la capacidad de monitorearla por espectroscopía NIR *in situ* usando modelos de calibración basados en regresión de mínimos cuadrados parciales (PLSR). Posteriormente en cultivos de células CHO, se evaluaron diferentes técnicas de regresión para construir modelos de calibración para monitorear CPP y PQA. La PLSR resultó inadecuada por la variabilidad química y física que presentan los cultivos de células CHO durante las diferentes fases del cultivo en lote. La regresión local ponderada (LWR) fue adecuada para monitorear CPP clásicos (concentración de glucosa, lactato, células, entre otros) al manejar adecuadamente la variabilidad ligada a la progresión del cultivo celular. Sin embargo, para el perfil de glicosilación (CQA), fue incapaz de manejar adecuadamente las complejas relaciones no-lineales entre los espectros NIR y la concentración de diversas glicoformas del mAb. Esto fue superado con el uso de modelos basados en regresiones con vectores de soporte (SVR), permitiendo generar modelos de

diferentes glicoformas relacionadas con efectos clínicos particulares. Globalmente, este trabajo ha contribuido a expandir las capacidades de la espectroscopía NIR *in situ* para el monitoreo de CPP clásicos de una forma más precisa, a generar CPP innovadores como el estado fisiológico celular en cultivos vegetales y CQA como los perfiles de glicosilación del mAb ligados a características clínicas en cultivos de células animales.

INTRODUCTION (ENGLISH)

Biopharmaceuticals or biological medicines have become essential for the treatment of several diseases such as cancer, inherited deficiencies, immunological diseases, among others. These medicines include molecules such as monoclonal antibodies, genetic vectors, vaccines, anticancer drugs, among others. The high complexity of these molecules has limited their production by chemical synthesis, which is why cell cultures are currently used as production platforms for their industrial production.

The demand for biopharmaceuticals has increased in the last decade, particularly for anticancer drugs and monoclonal antibodies. Consequently, industry has had to supply the increasing demand of medicines while guaranteeing the quality of the products, which implies a technological and economic challenge. As result of the molecular complexity of biopharmaceuticals, these are approved by government regulatory agencies (U.S.A. FDA, European EMA, among others) based on an extremely specific process. Therefore, any modification to the production process requires new validation demonstrating the consistency of the drug in terms of clinical effects and biosecurity. Modifications in production processes may be inevitable in the biopharmaceutical industry, so two fundamental principles for the development of new processes for biopharmaceutical production have been proposed:

- ➢ The quality of the medicine should no longer be evaluated only at the end of the process (Quality by testing, QbT), but must be generated during the entire manufacturing process (Quality by design, QbD)
- ➢ The variability of the process and, thus of the product, must be controlled in real-time, which requires a deep understanding of the production process (culture media, kinetics, cell physiology, impact of cell environment and cell responses to it, etc.), as well as process real-time monitoring.

These principles led the International Council for Harmonization (ICH) to the presentation of a series of explanatory notes about the QbD (ICH Q8-11) for drug development, manufacturing, evaluation and registration. Subsequently, various

regulatory agencies published technical aspects for the establishment of QbD, particularly for advanced retro-control systems based on a deep knowledge of the production process. For this purpose, the Process Analytical Technology (PAT) initiative was proposed, so that the biopharmaceutical companies monitor in real-time critical parameters of the processes (Critical Process Parameter, CPP) and control them to reach the final quality specifications of the medicines, defined as Critical Quality Attributes (CQA). Currently, only some few physicochemical parameters as temperature, pH and dissolved oxygen, are monitored and controlled in real-time using in-line analyzers, while the monitoring of nutrients, by-products, products or cell concentrations are still commonly performed by off-line schemes. This has limited the implementation of real-time control strategies and the benefits of quality by design, such as continuous improvement and validation, adequate risk management, among others.

In recent decades, the use of in-line process analyzers based on vibrational spectroscopy (near infrared, Raman, among others) has gained popularity because they can provide multicomponent information without sample treatment and in real-time when used with in situ probes. However, their use has been limited and restricted to the monitoring of some classical compounds of the culture media as glucose, lactate or glutamine. This is due to the complexity of cell culture processes and the fact that multivariate methods are required so that analyzer's complex signals be used, in an estimated or predictive manner, as monitoring instruments. Consequently, real-time monitoring of parameters which could better describe the cell cultures processes, such as cell physiological state or the quality of the biopharmaceuticals produced, have not yet been properly addressed for successful implementation of the PAT strategies for control and optimization of cell culture processes.

In this context, the general objective of the thesis has been to develop new real-time monitoring methodological approaches for processing data from *in situ* near infrared spectroscopy (NIRS), in two cell cultures platforms producing biopharmaceuticals: animal cells (CHO) producing monoclonal antibodies (anti-Rhesus D) and plant cells (*Catharanthus roseus*) producing anticancer drugs (vincristine and vinblastine).

The first part of the thesis is the essential background on biopharmaceuticals and their production through cell cultures. Particular attention was given to the cultures of animal cells (CHO) producing monoclonal antibodies and plant cells (*C. roseus*) producing anticancer, as well as their challenges for the production of biopharmaceuticals. The nature and interest of the QbD-PAT initiative, to address such challenges, particularly in plant and animal cell cultures, are described. Subsequently, a detailed introduction on multivariate methods, which allow to use the complex data provided by process analyzers and probes in a predictive manner, is presented. Once the basics for the use of spectroscopic data have been addressed, a collection of QbD-PAT applications in cell cultures using process analyzers is shown, particularly applications of near-infrared spectroscopy in in-line modes. Finally, in the second part a general balance of the background is shown to clearly establish the hypotheses, objectives, and the general methodology adopted. The third part of the thesis, materials and methods, presents the experimental and numerical methodology used to carry out the research work. The fourth part of the thesis, results and discussion presents the results obtained throughout the four years of doctoral work and is divided into 4 chapters which included works as scientific articles already published, submitted or for submission, also complements that are not included within the articles. Chapter I concerns primarily the generation of plant cell culture processes for anticancer production and the identification of CPP, which will then be used to evaluate their monitoring by NIR spectroscopy. Chapter II adressed the multivariable data processing methodologies commonly used for animal cell culture monitoring (Partial Least Squares Regression-PLSR, Principal Component Regression-PCR) through classical parameters (concentrations of substrates, products and by-products), as well as a new approach (Locally Weighted Regression-LWR) of these methodologies is proposed to overcome their limitations. Chapter III proposes the new use of different multivariate data processing methodologies (Supported Vector Regression-SVR, Artificial Neural Network Regression-ANNR) seeking improvement of the monitoring procedures. Finally, Chapter IV takes up the results of the previous two chapters to propose new methodologies allowing the monitoring of innovative parameters related to cell physiology and the quality of the biopharmaceuticals. The last part of the

thesis, general conclusions, takes a global balance of the acquired results and provides perspectives that may be useful for future applications.

Les médicaments biologiques ou biopharmaceutiques sont devenus essentiels pour le traitement de diverses maladies telles que le cancer, les carences héréditaires, les maladies immunologiques. Ces médicaments comprennent des molécules telles que des anticorps monoclonaux, des vecteurs génétiques, des vaccins, des médicaments anticancéreux. La grande complexité de ces molécules a limité leur production par synthèse chimique, de sorte que les cultures cellulaires se sont imposées comme plateformes pour leur production industriel.

La demande de produits biopharmaceutiques a augmenté au cours de la dernière décennie, en particulier pour les anticorps monoclonaux et anticancéreux. En conséquent, l'industrie a dû répondre à cette demande croissante tout en garantissant la qualité de ces produits, impliquant de relever un doublé défi, à la fois technologique et économique. En raison de la complexité moléculaire des produits biopharmaceutiques, ceux-ci doivent être approuvés par les agences réglementaires de la santé (FDA, EMA, …) sur la base d'un processus extrêmement spécifique. Par conséquent, toute modification du procédé de production nécessite une nouvelle validation afin de montrer la consistance du médicament en termes d'éfficacité clinique et de biosécurité. Les modifications dans les procédés de production peuvent être inévitables dans l'industrie biopharmaceutique, si bien que deux principes fondamentaux pour le développement de nouveaux procédés pour la production biopharmaceutique ont été proposés afin d'en limiter l'impact :

- ➢ La qualité des médicaments ne doit plus être évaluée qu'à la fin du procédé (Quality by testing), mais tout au long du processus de fabrication (Quality by Design, QbD)
- ➢ La variabilité des procédés, et donc du produit, doit être contrôlée en temps réel, ce qui nécessite une compréhension détaillée du processus de production (milieu de culture, cinétique, physiologie cellulaire, impact de

l’environnement cellulaire et réponse associée…), ainsi que des moyens de mesure *in situ*.

Ces deux principes ont conduit le Conseil International d'Harmonisation (ICH) à présenter une série de notes explicatives sur le QbD (ICH Q8-11) concernant le développement, la fabrication, la qualification et l'enregistrement des médicaments. Par la suite, au niveau local, les agences réglementaires ont proposé d’approfondir les aspects techniques du QbD, en particulier par la mise en place de systèmes avancés de rétro-contrôle basés sur une connaissance approfondie du procédé de production. À cette fin, l'initiative Process Analytical Technology (PAT) a été proposée, afin que les opérateurs des sociétés bio-pharmaceutiques puissent suivre en temps réel les paramètres procédés critiques (Critical Process Parameter, CPP) afin d’atteindre les spécifications finales des biomédicaments, comme définies par les Attributs Qualité Critiques (Critical Quality Attribute, CQA). Actuellement, seuls quelques paramètres physico-chimiques tels que la température, le pH, ou encore la concentration en oxygène dissous sont mesurés et contrôlés en temps réel à l'aide d'analyseurs en ligne, tandis que le suivi des concentrations en nutriments, en métabolites ou en cellules vivantes est encore couramment effectuée par des méthodes analytiques hors ligne. Par conséquent, la mise en œuvre de stratégies de pilotage en temps réel sur la base du concept QbD, tels que l'amélioration et la validation continues de la qualité des produits et du risque associé ont été fortement limité.

Au cours de la dernière décennie, l'utilisation d'analyseurs en ligne basés sur la spectroscopie vibrationnelle (proche infrarouge, Raman, entre autres) a gagné en popularité car ces analyseurs, lorsqu’il sont utilisés *in situ*, ont la capacité de donner des informations multi-composantes de manière non invasive et en temps réel au cours du procédé. Cependant, leur utilisation ont été limitée au suivi de certains composés classiques du milieu de culture tels que le glucose, le lactate ou encore la glutamine, principalement en raison de la complexité des processus cellulaires et du fait que des méthodes d’analyse multivariées des données sont

nécessaires pour pouvoir utiliser les signaux spectroscopiques complexes issus de l'analyseur. Par conséquent, l'analyse en temps réel de paramètres permettant de mieux décrire les processus cellulaires, physiologiques, ou encore la qualité des produits biopharmaceutiques, n'a pas été réelement abordée au cours des procédés..

Dans ce contexte, l'objectif général de la thèse, est de développer de nouvelles approches méthodologiques de traitement de données spectrales, issues de la spectroscopie proche infrarouge (NIRS) *in situ*, pour le suivi en temps réel des cultures cellulaires productrices de molécules biopharmaceutiques sur deux plateformes de production : les cellules animales (CHO) productrices d'anticorps monoclonaux (anti-Rhésus D) et les cellules végétales (*Catharanthus roseus*) productrices d'anticancereux (vincristine et vinblastine).

La première partie de la thèse présente les notions essentielles sur les molécules biopharmaceutiques et leur production. Une attention toute particulière a été accordée aux cultures de cellules animales (CHO) produisant des anticorps monoclonaux et aux cellules végétales (*C. roseus*) produisant des anticancéreux, ainsi qu'aux défis à relever pour leur production à l'échelle industrielle. Afin de relever ces défis, l'approche QbD-PAT a été ensuite décrite, en particulier dans le contexte des procédés de cultures de cellules végétales et animales. Par la suite, une introduction détaillée des méthodes d'analyses multivariées permettant d'utiliser de manière prédictive les données complexes issues des sondes de mesures spectroscopiques est présentée. Une fois abordées les bases de l'utilisation des données spectroscopiques, un ensemble d'applications QbD-PAT dans les procédés de culture cellulaire utilisant en particulier la spectroscopie proche infrarouge *in situ*, est présentée. Enfin dans une deuxième partie, un bilan général du contexte scientifique et technique est réalisé afin de proposer les hypothèses de travail, les objectifs scientifiques et la méthodologie générale adoptée au cours de ce travail de thèse. La troisième partie du manuscrit

présente la méthodologie expérimentale et numérique employée pour mener à bien les travaux de recherche.

La quatrième partie du manuscrit de thèse présente les résultats obtenus au cours des quatre années de doctorat et est divisée en quatre chapitres. Le premier chapitre concerne principalement le développement d'un procédé de culture de cellules végétales pour la production d'anticancéreux et l'identification des CPP, qui seront ensuite utilisés pour le suivi en temps réel par spectroscopie NIR. Le deuxième chapitre présente les méthodologies de traitement de données multivariées couramment utilisées pour le suivi des procédés de culture de cellules animales (régression partielle par les moindres carré-PLSR, régression en composantes principales-PCR), ainsi qu'une nouvelle aproche (Régression pondérée localement-LWR) afin de surmonter les limites des deux précédentes méthodes. Le troisème chapitre propose une nouvelle utilisation de différentes méthodologies de traitement de données multivariées (Régressions par machines à vecteurs de supports -SVR, Régression par réseau de neurones artificiels-ANNR) afin d'améliorer les performances des modèles de prédiction. Enfin, le quatrième et dernier chapitre utilise les résultats des chapitres précédents afin proposer de nouvelles méthodologies permettant le suivi de paramètres innovants liés à l'état physiologique des cellules ou à la qualité des biopharmaceutiques. La dernière partie de ce manuscrit dresse un bilan global des résultats acquis et propose des perspectives qui pourraient être utiles pour de futures applications.

INTRODUCCIÓN (ESPAÑOL)

Las medicinas biológicas o biofarmacéuticos, se han vuelto imprescindibles para el tratamiento de diversas enfermedades como el cáncer, deficiencias hereditarias, enfermedades inmunológicas, entre otros. Estos medicamentos incluyen moléculas como anticuerpos monoclonales, vectores genéticos, vacunas, anticancerígenos, entre otros. La elevada complejidad de estas moléculas ha limitado su producción por síntesis química, por lo que actualmente se emplean cultivos celulares como plataformas de producción para su producción industrial.

La demanda de biofarmacéuticos ha aumentado en la última década, particularmente para los anticancerígenos y anticuerpos monoclonales. Consecuentemente, la industria ha debido satisfacer la creciente demanda garantizando la calidad de los productos, lo cual implica un reto tecnológico y económico. Como resultado de la complejidad molecular de los biofarmacéuticos, estos son aprobados por las agencias sanitarias gubernamentales sobre la base de un proceso extremadamente específico. Por lo tanto, cualquier modificación al proceso de producción requiere una nueva validación que demuestre la consistencia del medicamento en términos de efectos clínicos y bioseguridad. Las modificaciones en los procesos de producción pueden ser inevitables en la industria biofarmacéutica, por lo que se han propuesto dos principios fundamentales para el desarrollo de nuevos procesos para la producción de biofarmacéuticos:

- ➢ La calidad del medicamento ya no debe evaluarse sólo al final del proceso (Calidad por inspección, QbT), sino que debe generarse durante todo el proceso de manufactura (Calidad por diseño, QbD)
- ➢ la variabilidad del proceso y, por lo tanto, del producto, debe controlarse en tiempo real, lo que requiere una comprensión detallada del proceso de producción (medio de cultivo, cinética, fisiología celular, impacto del medio ambiente celular y respuestas asociadas, entre otros), así como seguimiento de éste en tiempo real.

Estos principios condujeron al Consejo Internacional para la Armonización (ICH) a presentar una serie de notas explicativas sobre la QbD (ICH Q8-11) para el desarrollo, la fabricación,

la evaluación y el registro de medicamentos. Posteriormente a nivel local, las agencias gubernamentales profundizaron en aspectos técnicos para la instauración de la QbD, particularmente para la instauración de sistemas avanzados de retrocontrol basados en un profundo conocimiento del proceso de producción. Para este fin, la iniciativa Tecnología Analítica de Procesos (PAT) fue propuesta, a fin de que las sociedades biofarmacéuticas puedan monitorear en tiempo real parámetros críticos de los procesos (Critical Process Parameter, CPP) y los controlen para alcanzar las especificaciones finales de los medicamentos, definidos como parámetros críticos de calidad (Critical Quality attributes, CQA).

Actualmente solo unos pocos parámetros fisicoquímicos como la temperatura, el pH, o la concentración de oxígeno disuelto, son monitoreados y controlados en tiempo real empleando analizadores in-line, mientras que el monitoreo de las concentraciones de nutrientes, metabolitos o células aún se realiza comúnmente mediante métodos off-line. Esto ha limitado la implementación de estrategias de control en tiempo real y los beneficios que conlleva el concepto de calidad por diseño, tales como la mejora y validación continua, adecuado manejo de riesgos, entre otros.

En las últimas décadas el uso de analizadores de procesos in-line basados en espectroscopía vibracional (infrarrojo cercano, Raman, entre otros) ha ganado popularidad debido a que tienen la capacidad de brindar información multicomponente sin tratamiento muestral y en tiempo real cuando se usan junto con sondas *in situ*. Sin embargo, su utilización ha estado limitada al seguimiento de ciertos componentes clásicos del medio de cultivo como la glucosa, el lactato y la glutamina, debido principalmente a la complejidad de los procesos de cultivo celular y al hecho de que se requiere métodos multivariados para poder utilizar las complejas señales espectroscópicas brindadas por estos analizadores. Consecuentemente, el seguimiento en tiempo real de parámetros que permitan describir mejor los procesos celulares, como aspectos fisiológicos o sobre la calidad de los productos biofarmacéuticos, no ha sido enteramente abordado.

En este contexto, el objetivo general de la tesis es desarrollar nuevos enfoques metodológicos para el procesamiento de datos, derivados de la espectroscopía de infrarrojo cercano (NIRS) in situ, para el monitoreo en tiempo real de cultivos celulares

productores de biofarmacéuticos en dos plataformas de producción: células animales (CHO) productoras de anticuerpos monoclonales (anti-Rhesus D) y células vegetales (*Catharanthus roseus*) productoras de anticancerígenos (vincristina y vinblastina).

La primera parte de la tesis presenta las nociones esenciales sobre los biofarmacéuticos y su producción mediante cultivos celulares. Se dio atención particular a los cultivos de células animales (CHO) productoras de anticuerpos monoclonales y células vegetales (*C. roseus, T. globosa*) productoras de anticancerígenos, así como sus retos para la producción de biofarmacéuticos. La naturaleza y el interés de la iniciativa QbD-PAT, para abordar los retos de los cultivos celulares, particularmente en cultivos de células vegetales y animales, son descritos. Posteriormente, una introducción detallada a los métodos multivariados, que permitan emplear los complejos datos brindados por los analizadores de proceso de manera predictiva, es presentada. Una vez abordados los aspectos básicos para el uso de datos espectroscópicos, se muestra una colección de aplicaciones QbD-PAT en cultivos celulares empleando analizadores de proceso, particularmente aplicaciones de la espectroscopía de infrarrojo cercano en modos in-line. Finalmente, en la segunda parte se muestra un balance general de los antecedentes para fijar claramente las hipótesis, objetivos, y la metodología general adoptada. La tercera parte de la tesis, materiales y métodos, presenta la metodología experimental y numérica empleada para realizar el trabajo de investigación. La cuarta parte de la tesis, resultados y discusión, presenta los resultados obtenidos a lo largo de los cuatro años de trabajo doctoral y se encuentra dividida en 4 capítulos. El capítulo I concierne principalmente el desarrollo de un proceso de cultivo de células vegetales para la producción de anticancerígenos y a la identificación de CPP, que posteriormente serán usados para evaluar su monitoreo mediante espectroscopía NIR. El capítulo II presenta las metodologías multivariables de tratamiento de datos empleadas actualmente para el monitoreo de cultivos celulares (Partial Leats Squares Regression-PLSR, Principal Component Regression-PCR) y sus limitaciones, así también se propone un nuevo enfoque (Locally Weighted Regression-LWR) de estas metodologías para superar sus limitaciones. El capitulo III propone el nuevo uso de diferentes metodologías de tratamiento de datos multivariados (Supported Vector Regression-SVR, Artificial Neural Network Regression-ANNR) para mejorar el desempeño de modelos de predicción. Finalmente, el capitulo IV retoma los resultados de los dos

capítulos anteriores para proponer nuevas metodologías que permitan monitorear parámetros innovadores relacionados con el estado fisiológico de las células o a la calidad de los biofarmacéuticos. La ultima parte de la tesis, conclusiones generales, hace un balance global de los resultados adquiridos y brinda perspectivas que puedan ser de utilidad para aplicaciones futuras.

1 BACKGROUND

1.1 BIOPHARMACEUTICALS, CELL CULTURE & QUALITY

Biopharmaceuticals are complex molecules produced by living organisms which have revolutionized the way medicine treats diseases such as immunologic illnesses, cancer, heredity deficiencies, among others. The term biopharmaceutical has been controversial, and the meaning changes according to scientific, regulatory, financial and popular context. It was first used in the 1980s to describe therapeutic proteins produced in biological systems (Walsh 1999), more recently, several proteins and non-protein substances such as nucleic acids for diagnostics, as opposed to therapeutic purposes, have been also called biopharmaceutics (Walsh 2002), revealing the trend and need of a broader definition which could be "biotechnology medicine" or "biotechnology product".

There are two main types of trends about biopharmaceuticals (Rader 2008): Broad biotech and New biotech. Broad biotech includes pharmaceuticals that are biological and manufactured by biotechnology, a broad view preferred in north America, where regulatory definitions such as *biological products* are used to regulate the market of biopharmaceuticals and related products (U.S. Food and Drug Administration 2008) (FDA); New biotech defines biopharmaceuticals such as those based on new technologies including genetic engineering. This definition is preferred in Europe, where those products are regulated as *biological medicinal products* through the European Medicines Agency (EMA).

According to these trends in biopharmaceuticals, there are several production platforms based on the nature of the producing living organisms, either wild or recombinant in nature, for instance, pluricellular organisms such as animals or plants cultures in farms and fields, or prokaryote and eukaryote cells in bioreactors. Such diverse production platforms involve different advantages and drawbacks in terms of the production process as well as in the properties of the biopharmaceutical product. For example, prokaryote-based processes could be more easily set up, although such medicines would lack post-translational modifications, which are inherent to eukaryotic cells such as insect, plant, animal or fungi cells. This is of great concern because translational modifications such as glycosylation or proper protein folding have great impact on clinical effects. Thus, there is an increasing interest for

eukaryotic, particularly mammalian, cell cultures for biopharmaceutical production with similar composition or post-translational modifications to those of humans. The biopharmaceutical industry is rapidly growing; in 2014, approximately 212 biological medicines received approval by sanitary regulatory agencies (Walsh 2014), while this number increased to 378 in 2018 (Walsh 2018). Most recently approved biopharmaceutics (2014-2018) included 68 monoclonal antibodies (mAb), 23 hormones, 16 clotting factors, 9 enzymes, 7 vaccines, 5 nucleic acid–based products and 4 engineered cell–based products and for the first time, an interference RNA (iRNA) (Walsh 2018). Much of the growing availability of biopharmaceutics is due to released biosimilars, although some of them have also been withdrawn from the market for commercial and sanitary reasons.

Nowadays, the application spectrum of biopharmaceutics is broad and covers several disease treatments as shown in Figure 1-a. Although the nature of biopharmaceutics is diverse including enzymes, hormones, among others, almost half (48 %) of traded biopharmaceutics are mAb (Walsh 2018) as indicated in Figure 1-b, with an estimated world market size of 125 billion US dollars by 2020 (Ecker et al. 2015). The main platforms are *Escherichia coli* (*E. col*i), yeast and mammalian cell cultures, particularly Chinese Hamster Ovary (CHO) cells (Figure 1-c). As these molecules are produced in living organisms and may also interact with risky compounds in the production process (dangerous Host Cell Proteins (HCP), virus, prions, incomplete-processed biopharmaceutics, etc.), complex and severe downstream processes are used for assuring safety issues. Thus, maintaining the integrity of biopharmaceutics particularly with post-translational modifications while limiting risks is a technological and economic challenge. Failures in the former described challenge could lead to lot withdrawals, compromising patients safety, and trust in health-care providers, governmental agencies, and purchasers (Bunniran et al. 2009). Thus, relentless control of production processes is needed compared with small pharmaceutical molecules as shown in Figure 1-d.

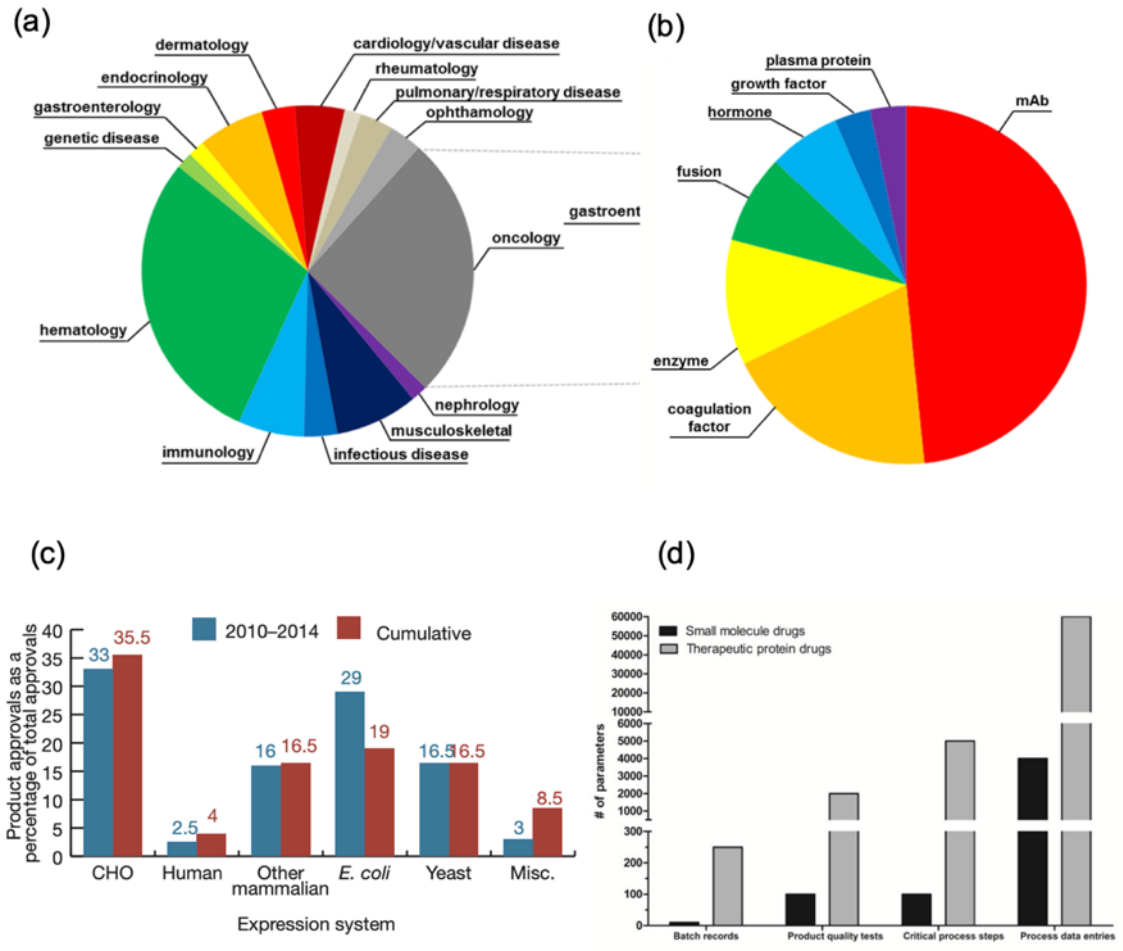

Figure 1. Biopharmaceuticals: market and manufacturing (Walsh 2014, 2018)

Pharmaceutical processes are usually operated under empirical approaches with fixed process conditions based on repeatability to achieve medicinal properties at the end of the production process. However, the complexity of biopharmaceuticals is a challenge to the biopharmaceutical industry which requires the integration of several disciplines from development to production processes (Alvi 2007) so as to ensure medicine properties and thus patient safety. A new initiative called Quality by Design (QbD) is intended to surpass the challenges of biopharmaceutical production (Rathore and Winkle 2009). QbD is based on three fundamental ideas: (i) risk management, (ii) knowledge management and (iii) process control. Risk management is a systematic process for the assessment of risks on the

quality of the medicine during the product lifecycle. Therefore, there is a need of innovative monitoring protocols so that retro control strategies could be put into operation. Finally, managing risks and process knowledge serves to execute changes in the manufacturing process to control and ensure medicine properties. This requires an extensive and increasing comprehension of the production process, i.e. knowledge management, for instance, the performance of the cell lines in particular cell culture media as well as the effect of the bioreactor operating parameters on biopharmaceutic properties.

1.1.1 Principles of bioreactor CHO cell cultures

1.1.1.1 Cell line

There are plenty of animal cell lines for industrial processes. However, CHO cell lines are particularly important in industry due to their advantages for biopharmaceutical production (Rita Costa et al. 2010): capable of high density cultures, relatively tolerant to adaptation in synthetic culture media, able to support process extrapolation due to a relative resistance to shear stress in large scale cultures, receptive to a broad spectrum of genetic engineering tools such as gene amplification systems, and similar to human profiles of post-translational modifications, among others. Consequently, more than 60 % of biopharmaceutical processes based on animal cell cultures use CHO cell lines (Walsh 2014). Development of productive cell lines requires extensive technical and economic resources. Thus, the gene of interest is often expressed in CHO cell lines to test efficacy and manufacturability suitability. Once proven, it is introduced into a CHO host cell line with effective amplification systems, such as DHFR-deficient CHO cell lines. It is important to use host cell lines already adapted to industrial culture media to save time and expression efficacy. High producing clones are then isolated. Since screening procedures are labor-intensive and time-consuming, novel cell screening systems such as fluorescence-activated cell sorting (FACS), the ClonePix™ system (Genetix), the LEAP™ system (Cyntellect) among others, are often used. Once promising clones are identified based on high expression levels, their performance in culture media is evaluated seeking ideal behavior for large-scale production (Kim et al. 2012).

1.1.1.2 Culture media

Animal cell cultures require a culture media for assuring proper cell metabolism. Moreover, culture media composition also plays an important role in protecting cells against shear stress and pH variation to some extent. Therefore, culture media must be optimized for cell growth as well as for product production, including titer and post-translational modifications.

At first, culture media developments included serum or tissue extractions from animals, which provided a rich environment for cell culture. However, new developments in the biopharmaceutical industry tend to exclude these compounds due to sanitary issues. Indeed, the use of animal-derived compounds are risky since there may also contain contamination

factors such as viruses or prions. Moreover, there is great lot-to-lot variation which may compromise process reproducibility and consistent product properties. Main components of animal cell culture media include (Yao and Asayama 2017):

- Hydrolysates: source of diverse compounds such as amino acids, salts, lipids, vitamins and low-molecular-weight peptides.
- Growth factors, hormones: required for inducing proliferation, differentiations, migration or export-import of nutrient compounds.
- Carrier proteins: particularly used for transportation into cells of non-dissolving compounds in aqueous solutions, such as lipids. Perhaps the most used is albumin, which also has toxin-neutralizing, antioxidant, and shear stress-reducing effects. Other examples are transferrin and lactoferrin for iron transportation into cells.
- Lipids and related components: serve various roles as membrane components, in nutrient storage and transport, in signal transduction and also as precursors if the cell lines lack the enzymes for particular metabolic pathways, such as for cholesterol synthesis.
- Transition metals: Se, Fe, Cu, Mn, and Zn; used for electron transfer in the active centers of enzymes or physiological proteins such as selenoproteins (glutathione peroxidase, thioredoxin reductase).
- Vitamins: necessary as precursors of diverse co-factors. Moreover, vitamin C and E have antioxidant effects.
- Polyamines: low-molecular-weight compounds having protein or nucleic-acid synthesis promoting effects.
- Reductants: important for maintaining intracellular redox environment, the most common reductant in culture media is cysteine.
- Protective additives, detergents: reduce shear stress. Pluronic F-68 or Tween 80 are commonly used for limited toxic activity and for capacity to solubilize lipophilic substances.

Although these compounds are of great importance for the success of cell culture processes, they are not usually monitored during continuous validation of processes. Indeed, only some

compounds such as the carbon and nitrogen source are often evaluated, as well as the byproducts of their metabolism such as ammonium and lactate. Glucose is the main energetic source, oxidized to pyruvate through the glycolysis pathway and then through the citric acid cycle in mitochondria. Glucose is commonly added at concentrations between 0.5 and 30 mM. Glutamine serves as carbon and nitrogen source, and also as an important energetic source for cells. It is usually added in concentration between 0.5 and 5 mM. Glucose and glutamine metabolism results in CO_2, lactate and ammonium production, which may have toxic effects for cells. Thus, there is also an interest in limiting such toxic effects by different strategies, as for example by substituting glucose and glutamine (Altamirano et al. 2000) in culture media, among others.

1.1.1.3 Engineering considerations & Operating parameters

In addition to culture media composition, the operating conditions of the reactor culture is essential for successful production processes. Temperature, dissolved oxygen, pH and osmolarity have received important attention since they may enhance the performance of cell culture processes. Optimal temperature for cell growth is usually 37 ºC but temperatures between 33 and 35 °C are frequently observed during production processes Indeed, decrease of temperature to 33 ºC may increase product synthesis or express anti-apoptosis genes (Gulevsky et al. 2017), which could enhance process productivity (Al-Fageeh et al. 2006).

Oxygen is required through respiration processes of energy metabolism. Because of the low solubility of oxygen in aqueous systems (≈7 $mg.L^{-1}$ at 37 ºC), it must be supplied continuously to satisfy the specific consumption rate of oxygen by cells (0.2 – 0.7 $mM.10^{-9}$ $cells.h^{-1}$). In bioreactor systems, oxygen is often supplied through bubbles diffusing into the culture media by sparges or membranes. It is usually controlled between 20 % and 50 % of dissolved oxygen. Oxygen availability is critical since a lack may cause over production of lactate, while high concentrations may have cytotoxic effects (Fleischaker and Sinskey 1981).

The pH is another important variable to control in CHO cell cultures. Indeed, cells tolerate only a small frame of pH range, between 6.8 and 7.8. The optimal pH value is usually reported as 7.2 for CHO cells, a deviation from this point may impact glucose consumption and lactate production (Schmid et al. 1990). Control of pH is usually performed using buffers,

addition of sodium hydroxyde or CO_2 into the culture system. It is also important to control osmolarity within optimal values (≈ 300 mOsm.kg^{-1}), which may be strongly affected by pH control or evaporation phenomena. In the case of hyper-osmolarity, there may be a decrease in cell growth, an increase in protein production or changes in post-translational modifications, particularly in glycosylation (Oyaas 2003; Konno et al. 2012). All these variables are critical, considering that in large-scale cultures there are problems commonly due to relatively low mixing intensity, for instance, limited CO_2 removal, oxygen, pH and nutrient gradients, among others (Xing et al. 2009).

1.1.1.4 Monoclonal antibodies: Structure and heterogeneity

Antibodies (Ab), also known as immunoglobulins (Ig), are large, Y-shaped proteins produced mainly by B lymphocytes which are used by the immune system for neutralizing pathogens in animals. Currently, the majority of therapies based on antibodies are immunoglobulin G (IgG1) mAb (Ryman and Meibohm 2017; Grilo and Mantalaris 2019), having the same basic structure as shown in Figure 2.

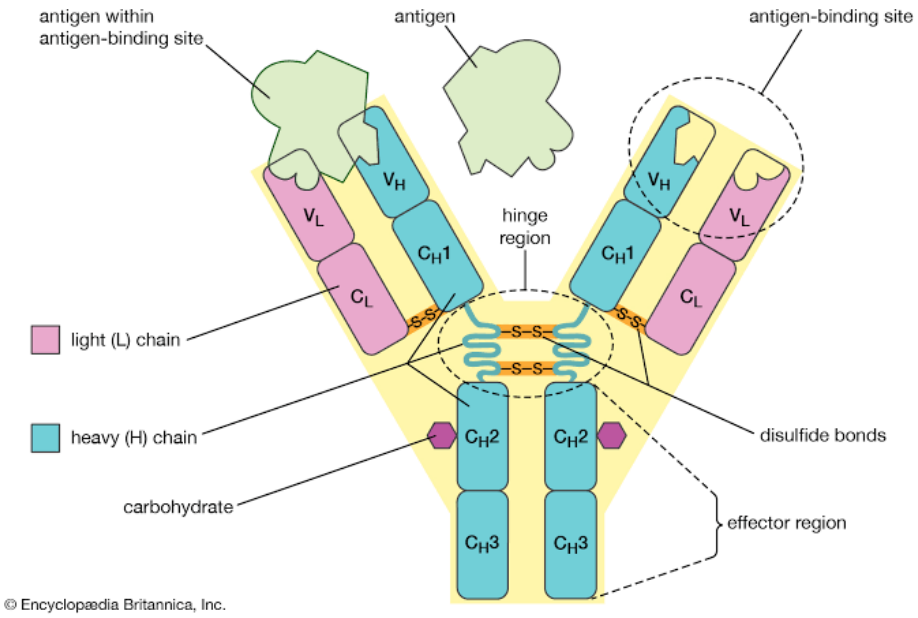

Figure 2. Structure of monoclonal antibodies

mAb are complex heterodimeric proteins with an approximate molecular weight of 150 kDa. Their composition includes four polypeptide chains, two identical heavy chains (50 kDa) and also two light chains (25 kDa). Both chains are linked by disulfide bonds forming a Y-shape, with constant domains (CL and CH) and variable domains (VL and VH). These variable domains, as well as the CH1 of the heavy chains constitute the antigen binding fragment (Fab) which also contains a hypervariable region of 5 to 10 amino acids with high specificity for the target antigen (McDonnell 2015). The CH2 and CH3 domains constitute the crystallizable fragment region (Fc) which can bind to diverse receptors on cells (Ryman and Meibohm 2017).

In 1975, the first attempts for mAb production involved fusing mouse lymphocyte and myeloma cells to produce murine hybridomas; an antigen is injected into a mouse and the resulting antigen-specific plasma cells are recovered from the spleen, isolated and fused with cancerous immune cell for immortality (Lerner 1981). Then these hybrid cells were cloned for identical daughter clones and cell line establishment. Originally, only murine hybridomas were produced through this technology; however, there were strong immune reactions in humans due to differences in the mAb effector region (Reichert et al. 2005). Therefore, new mAb types were generated through recombinant technology, for instance, chimeric mAb generated with murine VH and VL variable domains and human CH1, CH2 and CH3 domains, or humanized mAb that are totally human in nature with only the complementarity determining regions with a murine nature. More recently, primatized (Reichert et al. 2005) and fully human mAb has been developed seeking a reduction of undesirable immunogenic effects and enhanced patient tolerability (Ryman and Meibohm 2017). Perhaps the most important post-translational modification of mAb is glycosylation, consisting of glycan chain addition into the heavy chains, particularly into the highly conserved N-glycosylation locus on the fragment crystallizable region (Fc) in both heavy chains. The N-glycosylation position in mAb is on Asn-297 (Huhn et al. 2009) as schematically shown in Figure 2.

Although glycan chains within mAb represent only 2-3 % of total mAb molecular weight, they are essential for mAb conformation, stability, functionality (Zheng et al. 2011) and thus clinical effects, particularly complement-dependent cytotoxicity (CDC), antibody-dependent cell-mediated cytotoxicity (ADCC), immune modulation (del Val et al. 2010) and serum half-life (Dwek et al. 1995).

Glycosylation profile varies according to the cell line used and also by operating conditions. Thus, there is often reported some variability in glycosylation profiles that must be monitored and controlled in commercial processes (Schiestl et al. 2011) to ensure the conservation of mAb clinical profiles associated with glycosylation. This variability is due to the presence of mAb with different glycan chains attached to the same mAb structure; such variants have been referred to as glycoforms (Rademacher et al. 1988). There are diverse glycoforms as shown in Figure 3, although the great majority has the same basic skeleton of 5 sugar residues: Two N-acetylglucosamine residues attached to three mannose moieties. Such glycosylation heterogeneity is often analyzed under two main approaches: macro- and micro-heterogeneity. Macro-heterogeneity concerns the presence or absence of a glycan chain attached to the mAb, while micro-heterogeneity analyzes the structure of glycan chains in terms of sugar moieties.

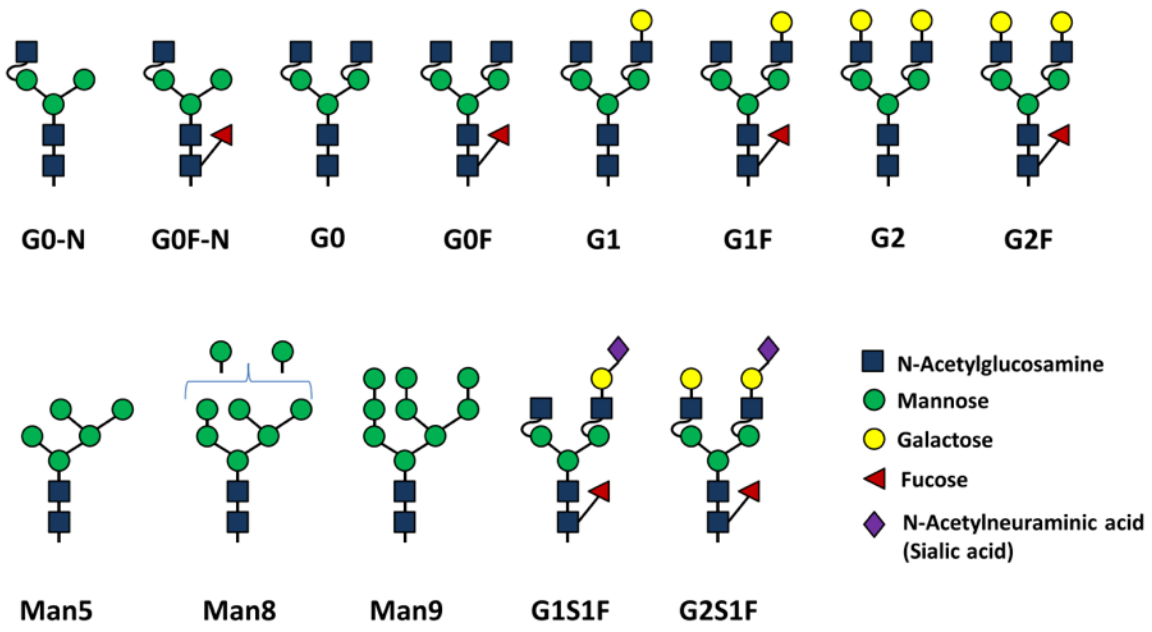

Figure 3. Common glycosylation forms in mAb (Ryman and Meibohm 2017)

Glycosylation heterogeneity depends on several factors such as the nature of the cell line, culture media, nutrient availability and by-product accumulation in culture media, cell physiological state, operating conditions such as pH, temperature, oxygen concentration and osmolarity, among others (Chotigeat et al. 1994; Yoon et al. 2005; del Val et al. 2010).

The choice of the production cell line also has a significant effect on the micro-heterogeneity of mAb mainly because the expression of glucosyltransferase enzymes and glycosidases varies from one to another (Raju et al. 2000). Commonly, CHO cells are reported to preferentially synthetize non-galactosylated (G0) glycoforms in contrast to other cell lines (Beck et al. 2013). Nutrient availability has been proven to affect both macro- and micro-heterogeneity when glucose is replaced by fructose, mannose or galactose (Tachibana et al. 1994). Effects of carbon source are likely due to different efficiency capacity of cells to channel sugar inputs through the nucleotide sugar pathway (del Val et al. 2010). Therefore, different strategies based on different sugar supplements have been explored seeking particular glycosylation profiles (Weikert et al. 1999; Hossler et al. 2009; Wong et al. 2010). For instance, the addition of 20 mM galactose into CHO cell culture media resulted in an increase in galactosylation from 14 % to 25 % (Kildegaard et al. 2016). Not only is the presence of sugars essential, but also their dynamics. Depletion of glucose is reported to increase the extent of non-glycosylated mAb, which can raise total mAb production values by as much as high as 45 %. This is likely due to the preference of cells to use glucose for energetic purposes instead of for supplying the precursor pool for glycosylation (Liu et al. 2013). Glutamine, an important energetic source has also important effects on glycosylation patterns, mainly related to high levels of non-glycosylated and Man5 forms in low glutamine concentration conditions (Fan et al. 2015).

1.1.2 Principles of bioreactor plant cell suspension cultures

Plants have historically been the source of clinical treatments. Their use has several variants based on cultural, knowledge and technological development stages of societies, primarily as herbal medicinal products, and more recently as the source of Active Pharmaceutical Ingredients (API) (Rates 2001; Niero et al. 2018) and biological medicines.

The development of plant-derived pharmaceuticals had firstly focused on solvent extraction of active molecules from plants, though low productivity, dependence on environmental factors and destruction of plants have remained as major issues to overcome. Thus, intensive efforts have been invested for increasing availability such as production via *in vitro* cultures, particularly suspension cultures feasible for scale-up. Nonetheless, commercial production using industrial scale suspension culture has been limited to a few molecules such as paclitaxel, shikonine, berberine, ginseng, among others (Bourgaud et al. 2001). When processes are intended for pharmaceutical issues, such as for the antileukemic agent paclitaxel, they can be considered as biopharmaceutical under the Broad biotech perspective since those processes use wild cell lines. The primary interest of plant biopharmaceutical processes was focused on native bioactive molecule production not industrially and economically feasible to be produced by chemical synthesis.

Nowadays, there has been an increasing interest for using plant recombinant cell lines as a promising platform due to their sanitary safety (Hellwig et al. 2004; Huang and McDonald 2009) and properties. Plant cells do not propagate mammalian virus, prions or pathogens, grow in chemically defined and economic culture media, and are capable of performing post-translational modifications, particularly glycosylation. Thus, plant cell suspension culture processes are now being considered as an alternative to animal cell culture processes, particularly to reduce production costs and make them more affordable to patients (Kaiser 2008).

Examples of the potential of plant suspension cultures as a new biotech biopharmaceutical production platform are the success of taliglucerase alpha (Elelyso®) for treating Gaucher disease (Grabowski et al. 2014) or the approval of Newcastle disease virus HN for treating Newcastle disease (Yusibov et al. 2011) and several proteins in pre- and clinical trials (Xu and Zhang 2014).

1.1.2.1 Cell lines

Development of a plant cell line often starts with callus induction from plant tissue, particularly from those suspected to be high producers and, ideally, also containing actively dividing cells. Then the callus is propagated through subcultures and placed into mixed liquid culture media for establishment of suspension cultures. Several rounds of subcultures are performed to discard callus clusters and finally fine suspension cultures with cell aggregates of only some few cells are generated, the whole process taking approximately 10 months (Mustafa et al. 2011).

Plant cells have diverse shapes and size profiles; spherical, cylindrical and geometric-like shapes are common with diameters of 50 – 100 μm. Their proliferation nature is characterized by 15 to 100 h doubling times, and by the fact that there is a tendency for cell aggregation. The size of cell aggregates may be small (composed of only few cells) or large (in the order of several millimeters), depending on the origin of the cell line and the operating parameters of culture (Chattopadhyay et al. 2002). *In vitro* plant cells can indefinitely proliferate and are totipotent when provided with proper plant growth regulator (PGR) regimen, can grow within wide pH (5 – 7) and temperature (20 ºC – 35 ºC) ranges (Maathuis 2013).

Genetic and phenotypic changes usually occur in terms of somoclonal variation (Deus-Neumann and Zenk 1984) and cell differentiation (Torrey 1975). The complex nature of cell proliferation in cell aggregates is not only a challenge during production cultures, but also for the establishment of monoclonal cell lines. Therefore recombinant biologic production is usually performed by polyclonal cultures (Nocarova and Fischer 2009).

1.1.2.2 Culture media

Plant culture media should be formulated considering the specific requirements of a cell line. Several studies were performed at the beginning of plant cell technology for determining the major needs of plant cells. Nowadays, most plant culture media is composed of macronutrients, micronutrients, vitamins, amino acids or nitrogen supplements, a carbon source, undefined organic supplements and growth regulators (I.M. and M. 2012).

Macro-nutrients are the source of nitrogen, phosphorus, potassium, calcium, magnesium and sulfur. The ideal concentration of nitrogen and potassium is around 25 mM, in the range of

1-3 mM for calcium, phosphorus, sulfur and magnesium. Nitrogen could be added as separated nitrates and ammonium salts or together as ammonium nitrate. Micro-nutrients, such as iron, manganese, zinc, boron, copper, molybdenum, among others, are required in minute quantities in the order of a few milligrams per liter. Among the microelements, the iron requirement is very critical. Chelated forms of iron and copper are commonly used in culture media for proper uptake by cells.

Plant cells are capable of synthesizing particular organic compounds such as vitamins, amino acids and organic acids and hormones, though in suboptimal quantities, therefore they are often added to culture media as supplements for improving culture performance. Vitamins added to culture media include thiamine, riboflavin, niacin, pyridoxine, folic acid, pantothenic acid, biotin, ascorbic acid among others. Supplementation of amino acids such as L-glutamine, L-asparagine, L- arginine, L-cysteine is also common for improving nitrogen uptake. Plant hormones or growth regulators are a group of organic compounds that promote growth, and cell development. Four broad classes of plant growth regulators (PGR) or hormones are used, auxins, cytokinins, gibberellins and abscisic acid derivatives.

In vitro cultures are usually heterotrophic and organic carbon sources are often needed. Sucrose being the most preferred, though fructose or even non-conventional sources such as lactose, maltose, galactose, raffinose, acetate, among others, have been used though with limited success.

Culture media is often autoclaved and thermolabile compounds such as PGR or vitamins, filter-sterilized and then added to the autoclaved medium. Nowadays, there are plenty of already-made plant culture media in the market which likely fit common necessities according to the nature of the cell line, for instance, White's, Murashige and Skoog, Gamborg B5, N6, Nitsch's media, among others.

1.1.2.3 Engineering considerations & Operating parameters

Plant cell nature implies particular considerations for bioreactor cultures, such as cell aggregation, mixing issues related to oxygen demand, rheological properties, shear sensitivity of plant cell cultures, and foaming.

Plant cells in suspension tend to form small cell aggregates or large clumps. This proliferation nature affects mass transfer and can lead to oxygen, nutrients or by-products inhomogeneities inside large aggregates, which may cause some cellular organization or differentiation with diverse effects on process performance. Mixing promotes growth by enhancing the transfer of oxygen and nutrients into cell aggregates. Mixing must consider rheological characteristics of the culture media (Scragg 1995), usually reported as highly viscous or with non-Newtonian behavior (Curtis and Emery 1993), relative high shear sensitivity to hydrodynamic stresses (Huang and McDonald 2009) and relatively low oxygen demand of cells, in the order of 0.1–0.5 mmol-O_2.$g_{DCW^{-1}}$.h^{-1} (Gao and Lee 1992). Thus, cultures are usually mixed at very low agitation speeds with high oxygen availability (Chattopadhyay et al. 2002). Aeration is also closely related to foaming, which has a remarkable influence on cell growth and both secondary metabolite and protein production. Several antifoaming agents have been tested, though their use is often related to reduction of cell growth and product formation (Wongsamuth and Doran 1994). Moreover, foaming may also trap cells within the foam phase and films of cells can be then attached to bioreactor walls, limiting the performance of the culture process.

1.1.2.4 Vincristine, vinblastine and suspension cultures

Vincristine (VC) and vinblastine (VB) are alkaloids with antileukemic properties which cause cell arrest during mitosis (Madoc-Jones and Mauro 1968). Vincristine is often used for treating diverse types of cancers such as acute lymphocytic leukemia, acute myeloid leukemia, Hodgkin's disease, neuroblastoma, and small cell lung cancer and Wilms tumor, among others. Vinblastine is often prescribed for Hodgkin's lymphoma, non-small cell lung cancer, bladder cancer, brain cancer, melanoma, and testicular cancer. Nowadays production is mainly carried out by extraction from *Catharanthus roseus* (*C.r.*) plant material grown in the United States, Spain, China, Africa, Australia, India and Southern Europe (Barkat et al. 2017). However, purification from whole plants, characterized by low product concentration remains a challenge to improving production processes.

The development of VC and VB as bioactive molecules against cancer started in the 1950. *C.r.* plants, used in traditional medicines acquired relevance as screening molecules with pharmaceutical potential. The medical research department of the Western Ontario

University decided to study the effects of plant extracts under controlled conditions. At first, the main interest was for diabetes treatment, but plant extracts did not show positive activity when it was intravenously supplied to rats. Animals died due to strong decrease in leucocytes, which finally moved attention to its anticancer properties. Several fractions were generated and analyzed, and the molecule showing this property was identified and called vincaleukoblastine (afterwards named VC) and later another molecule was also identified, VB. Parellely, Eli Lilly Company was also analyzing properties of *C.r.*, which finally lead to a cooperation agreement between both institutions (Noble 1990). VC and VB received FDA approval as cancer treatments in 1963 and 1965 respectively. Since then, their use in cancer therapies has been increased. Moreover, there is also an interest for developing VC and VB derivates with improved clinical effects, such as vindesine.

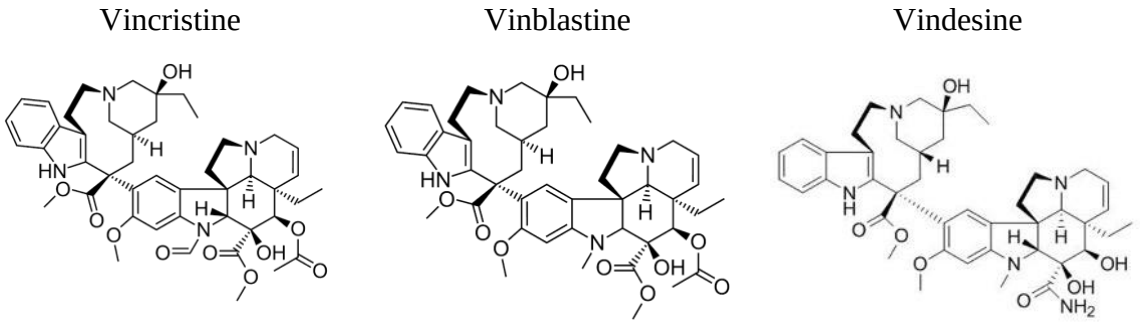

Figure 4. Structure of *Catharanthus roseus* alkaloids

Since total synthesis of these kinds of molecules is complex and economically unfeasible due to complex molecular structure (Figure 4), plants are the main industrial supply with low availability of VC and VB from plant material and thus, production is currently performed by large scale cultivars (Parry 2004). Precursor molecules such as catharanthine and vindoline are solvent-extracted and then VC and VB are produced by chemical semi-synthesis from precursors. Although large scale cultivars may appear as reliably supply, the low availability of molecules from plants and a complex purification processes have limited VC and VB provisions. Indeed, there has recently been shortages of vincristine in the USA causing severe shock in the health sector because there is no proper substitute for therapies (Roni 2019). Consequently, there have been intensive efforts for developing processes based

on plant suspension cultures. If so, production could be carried out using industrial reactors under controlled conditions that would allow the provision of the necessary conditions for the plant cells in suspension culture to produce VC and VB in a greater quantity. However, because *in vivo* synthesis of VC and VB requires a complex enzymatic process through various organelles in differentiated cells of aerial tissues (St-Pierre et al. 1999; Murata and Luca 2005), it has been widely believed that *in vitro* cultures are unable to produce VC and VB (Verpoorte et al. 1993).

More recently, several works have refuted former statements and have proven production of VC and VB using *in vitro* cultures (Miura et al. 1987; Kalidass et al. 2010; Taha et al. 2014; Ataei-Azimi et al. 2018), particularly from calluses with early differentiation into roots or shoots; however, no consensus has been achieved on the effect of culture conditions and the role of cell differentiation required for VC and VB production has received little attention. Moreover, suspension cultures have also been reported producing VC and VB (Taha et al. 2014), though no inference about this capacity was provided. Indeed, it was recently discovered that *in vitro* cultured cambial meristem cells (CMCs) contained complete sets of enzymes that are responsible for the production of VC and VB from vindoline, though vindoline supplementation was required (Zhang et al. 2015). On the other hand, there have been reports of vindoline producing cultures (Scott et al. 1980; Naaranlahti et al. 1989). Thus, there must likely be an alternative metabolic pathway for *in vitro* vindoline synthesis which could be active either during cell differentiation or in particular differentiated cells. *Catharanthus* technology should be developed to evolve novel production processes based on differentiated cell suspension cultures capable of producing biopharmaceuticals such as VC and VB.

1.1.3 Synthesis: Biopharmaceuticals, cell culture & quality

Advances in science and engineering have allowed the development of new pharmaceuticals from living cells, called biopharmaceuticals. Today more than ever, these medicines are essential for treating several diseases and therefore there is an increasing demand. Due to the complex chemical structure and complicated biological synthesis of such medicines, there is some heterogeneity in terms of chemical structure and thus final clinical effects caused by uncontrolled variability within the cell culture process. This is of great concern since industry must maintain regulatory and sanitary compliance, assuring patients safety and efficacy of medicines. Therefore, there is a need to embrace new and more efficient quality approaches for real-time monitoring and control of cell culture processes thus assuring medicines quality.

1.2 NEAR INFRARED SPECTROSCOPY AS A PAT MONITORING TOOL

The complexity and variability of cell cultures remain a great challenge for the monitoring and control of these bioprocess as for ensuring the right properties of biopharmaceutics and reduction of process costs. An important fraction of processes can yet be optimized to increase productivity while ensuring biopharmaceutical quality provided proper control of chemical and biological parameters are also performed (Claßen et al. 2017). However, in-line monitoring of such parameters has been claimed as a major limiting technology for ensuring proper retro-control systems (Clementschitsch and Bayer 2006). Nowadays, only some parameters are routinely monitored using in-line sensors during cell culture processes such as pH, temperature and dissolved oxygen and the majority of biochemical and physiological parameters for cell cultures are still performed with off-line methods (Claßen et al. 2017).

Off-line monitoring implies risks of contamination during manufacturing controls (Chu and Robinson 2001) and also delays due to sample analysis. These delays could likely fail in providing real-time data of process performance and thus limit proper control strategies for enhancing the processes. This fact is the major reason for the industry to explore novel monitoring approaches seeking maximization in economic and technical terms. Regulatory agencies such as the USA FDA and European EMA proposed the Quality by Design (QbD) approach strategy (Rathore and Winkle 2009; Yu et al. 2014), which could be combined with Process Analytical Technology (PAT) (Rathore et al. 2010). Briefly, the main objective is to monitor and control in real time the concentration of some process parameters such as viable cells, nutrients and by products, whose variability may have an impact on the Quality Target Product Profile (QTPP). Consequently, QTPP must be characterized by Critical Quality Attributes (CQA) of biopharmaceutics and also monitored in real time for continuous validation of the CPP control strategy (Teixeira et al. 2009a; Jenzsch et al. 2017).

In recent years, intensive efforts have been made to establish PAT as the means to monitor biochemical and physiological parameters of cell cultures, particularly using automatic at-line or on-line traditional biochemistry approaches with the aim of reducing analysis times, sample volumes and contamination risks. Such approaches allow the monitoring of substrates (glucose, glutamine) and by-products (ammonium, lactate, among others) in the order of minutes while for product CQA such as glycosylation in the order of hours (Burnina et al.

2013; Doherty et al. 2013; Henninot et al. 2015; Dong et al. 2016). In the last few decades, vibrational spectroscopy, in combination with multivariate analysis, has been proven to be a promising tool, particularly for cell culture monitoring (Li et al. 2016). However, it has mainly been restricted to some usual cell substrates and/or by-products (Arnold et al. 2003; Mattes et al. 2007; Henriques et al. 2009; Clavaud et al. 2013; Bhatia et al. 2018; Li et al. 2018b). Indeed, the major limitation for implementation of monitoring protocols using NIRS analyzers is likely the challenge of using complex multivariate analysis to use spectra in a predictive and accurate manner, to estimate concentration of key compounds within complex culture media.

In-line measurements for biopharmaceutical production by cell culture are challenging because of medium complexity in terms of chemical composition and dynamics during culture progression. Therefore, sensors with diverse principles are used for measurements of parameters sensitive to particular spectroscopic principles. For instance, in-line sensors based on dielectric, Raman, NIR and fluorescent spectroscopy, used when physical sampling from the bioreactor is not necessary, limiting contamination risks and providing signals related to process performance in real-time. Moreover, some spectroscopies such as NIR and Raman can provide multicomponent information within their signals or both physical and chemical information as in NIR spectroscopy. However, complex cell culture matrix also generates complex spectra that require deep knowledge of spectroscopy principles as well as of chemometric techniques for building correlations between spectrum and analyte values by reference methods. The term “chemometric” will be described later (1.2.2).

In the context of process monitoring, these in-line analyzers are intended for performing systematic measurements of the process performance in real-time, which require a sound comprehension of the relationship between spectra and process performance for building calibration models. Finally, spectra could be used in a predictive manner for estimating CPP or CQA routinely.

1.2.1 The NIRS challenge for monitoring: From basis to assumptions

The extent of NIR spectroscopy applications is broad since both chemical and physical information is contained within spectra. Thus, information extraction is complicated and is usually focused on either chemical or physical information. Though there are well developed analysis methods such as band assignment based on spectra-structure relationships, these methods may not be enough to unravel complex NIR spectra, particularly in actual process samples of complex matrices (Ozaki et al. 2006) with spectral distortions due to physical information in the form of scattering effects. Consequently, multivariate analysis is frequently, if not always, required for extracting information from complex NIR spectra.

The implications of NIR spectral complexity is more easily perceived when compared with classic colorimetric analysis for concentration determination. In such classic analysis only a single or few registered variables (absorption at particular wavelengths) are used for performing correlation. This is possible as the matrix is rather simple, generating relatively simple spectra with few peaks at particular wavelengths, the compound of interest strongly marked with a limited relationship to other wavelength absorptions. In contrast, NIR spectra of actual processed samples are complex as every registered variable is a mixture of energy combinations. Thus, the use of single or few variables could not depict the whole extension of the process dynamics affecting the whole spectrum.

Novel NIR analysers can provide a great number of observed variables when scanning, which can enhance the resolution of each in-line acquired spectrum and thus provide more detailed information. However, this availability of observed variables is usually greater than the available chemical information of the matrix (qualitative and quantitative composition) causing high collinearity. This scenario leads to a variety of relationship cases: exact linear relationships called exact multicollinearity and non-unique relationship with multiple solutions called near collinearity, a type of collinearity where variables can be written approximately as a linear function of other variables. This is one of the main reasons for the need for multivariate analysis and chemometrics.

1.2.1.1 NIRS generalities

Molecules can be considered to have two main properties: static and dynamic. Static ones include atomic composition, structure and stereochemistry while dynamic properties are related to vibrational and rotation patterns as well as resonance. Although NIRS users are often more interested in static properties of molecules (*i.e.* molecule structure) rather than their dynamic properties (vibrational movement), the later are directly responsible for the existence of NIR phenomena, besides such dynamic properties are based on the static ones and this provides the basis of NIRS analysis. In consequence, in order to analyse NIR spectra, it is mandatory to understand the relationship between both properties (Williams et al. 1987). Near-infrared is commonly referred to as wavelength region between λ 780 and 2,500 nm, or 1/λ 12,820 and 4,000 cm^{-1}. NIR analyses are mainly focused on vibrational energy, which refers to atomic bond oscillations in a molecule along the axis bond. The initial conceptual basis of vibrational spectroscopy (IR, NIR and Raman) has used the harmonic oscillation (based on Hooke's law) to explain the interaction between energy and matter (Figure 5). Therefore, a spring and two balls have been used to model a bond that connects two atoms. The oscillations may result into several vibrational states of a molecule.

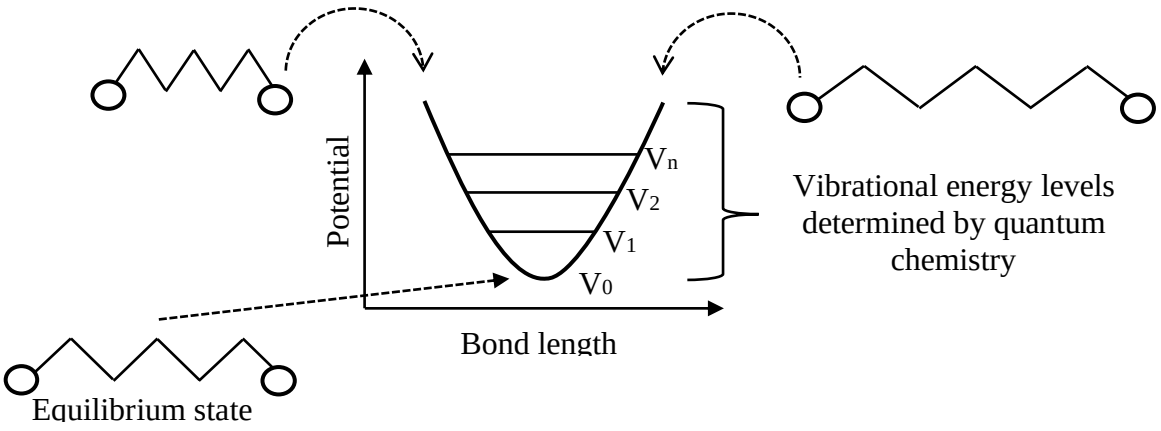

Figure 5. Concept of NIR vibrational spectroscopy

For instance, there may be three main levels depending on atom bond lengths. An equilibrium state with low potential energy and two states of high potential energy for the compressed and stretched states of the spring or bond. Changes in such states required gain or loss of

energy, which could be analyse by quantum theory for determining the specific energy levels that are possible for a particular vibration. Such gain or loss of energy is related to NIR spectroscopy equipment detection features and their analytical chemical applications, since molecules are selective in the electromagnetic radiation absorbed.

For changing the vibrational state of a molecule, the energy of the photon must be equal to the energy difference between two vibrational states. The two main modes of molecular vibrations (Burns and Ciurczak 2008) are: stretching (a change in bond length) and bending (a change in bond angle or a change in the location of a group of atoms in relation to the remaining part of the molecule). Four types of bend are recognised: scissoring and rocking (in-plane bending modes, symmetrical and asymmetrical, respectevey); wagging and twisting (out-of-plane bending modes, symmetrical and asymmetrical, respectively). Molecules have a finite number of possible vibrations and therefore molecules have characteristic transition profiles.

Indeed, the former example of the harmonic vibration model does not explain the nature of real molecules within complexor non-complex matrices, because it does not take into account changes in potential energy due to bond approach (Coulombic repulsion) or separation (near the breaking point). “Althogh bonds are elastic, they do not obey Hooke’s Law exactly” (Kalberg, 2006). Therefore, the anharmonic oscillator model is known as more precise for explaining vibrational spectroscopy of real molecules. The number of transitions levels or vibrational quantum numbers (V) in Figure 5, are called bands. Fundamental vibrational band refers to transitions from v=0 to v=1, overtone band from v=0 to any level greater than 1, combination band from v=1 to any other level (Williams et al. 1987).

Two main conditions favour absorption of IR electromagnetic radiation by a molecule: a change in the dipole moment because of vibrational movements or mode and that the energy of the incident radiation must match the difference between the two energetic levels. In relation to the absorption intensity this will increase with anharmonicity linked to difference in atomic masses of the bonded elements. This behaviour has been exemplified with the strong absorption of such bonds as C-H, N-H and O-H, compared to the less intense absorption of C=O and C-C (Williams et al. 1987).

Briefly, main chemical factors affecting NIR absorption are therefore (Williams et al. 1987):

The functional group effect: It is the most dominant effect since it explains much of the vibrational nature. As already has been mentioned, vibration can be depicted as oscillations of atoms through their bonds within a particular molecule. Then the energy required for vibration (E) could be analysed by Planck–Einstein relationship or the Planck equation (Equation 1):

$$E = h\nu = \frac{h}{2\pi}\sqrt{\frac{k}{\frac{m_1 m_2}{m_1 + m_2}}}$$

Where *k* is the force constant of the bond strength, *m* is masses of atoms, h the Planck constant (6.6×10^{-34} J.s) and v the photon frequency (s^{-1}).

Equation 1

According to equation 1, the fundamental vibration frequency increases with a reduction of atomic masses. Absorption of some common functional groups are shown in Figure 6. For instance, the first overtone bands for C-H stretching are at higher frequencies than those for N-H because the nitrogen atom is heavier than the carbon atom.

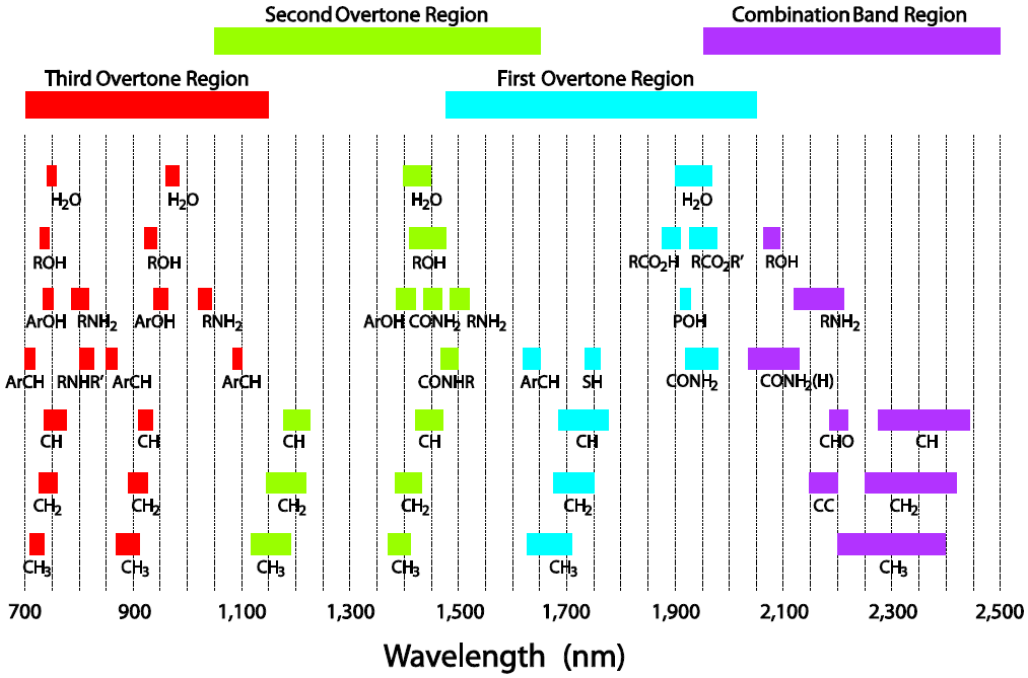

Figure 6. NIR bands assignment

The neighboring group effect: As shown in Figure 6, there is no assignment of a functional group to a particular wavelength but rather a frame. It is caused by the fact that molecules can be influenced by the identity of a neighboring functional group, particularly if it is strongly electron donating or withdrawing. This neighboring functional group can then affect the bound strength or the dipole moment of the main functional group of interest.

Hydrogen bonding: It does not involve a composition change though can have a strong influence in the spectrum. It is formed when a hydrogen atom, is attracted to a pair of electrons. The most common hydrogen bond donor groups are -OH and -NH mainly, though there are diverse accepting groups.

Macroscopic effects: The structure of the material may strongly influence absorption. The crystalline or amorphous nature of a material represent different molecular environments and thus different vibrational movements. Moreover, there is phenomenon called phase separation which involve the segregation of molecules into microscopic domains, such as micelles or aggregates. These domains have particular

interactions between functional groups, which are not the same as in the main bulk material.

NIRS as an absorption spectroscopy based also on the Bouguer-Lambert-Beer law (BLBL), relates the attenuation of light to the vibrational properties of the material through which the light is travelling, as shown in Equation 2:

$$A = \log\frac{I_0}{I} = \varepsilon \cdot C \cdot l$$

$Where\ A\ is\ absorbance,\ I_0$ and $I\ are\ the\ incident\ and\ transmitted\ light$

ε the molar absorptivity, C concentration (in mM)

and $l\ the\ optical\ path\ lenght\ (in\ cm)$

Equation 2

NIR spectra represent the first outcome of NIRS for analytical purposes. For instance, Figure 7 is the spectrum of a biscuit sample that obviously has different components so this spectrum contained hundreds if not thousands of different absorptions related to different vibrational movements though they all are contained in a single spectrum containing only some few broad bans.

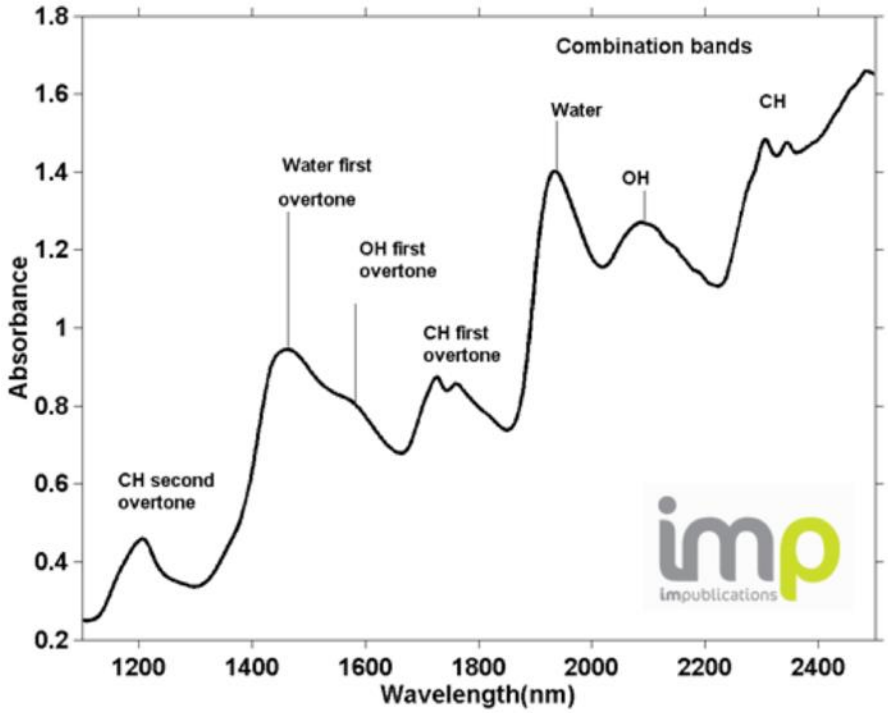

Figure 7. Spectrum of biscuit dough with the main absorption bands identified

All these phenomena on NIR spectroscopy may look overwhelming and give the impression of unfeasibility for practical implementations. However, applications of NIR spectroscopy is generally reported regarding analytical complex matrices. The success of NIRS relies on the fact that spectral complexity refers to the overall spectrum, while the validity of analytical methods is mainly based on detecting relevant variability in spectra. In this way, although the vibrational phenomena encoded within spectra is highly complex, generation of NIR calibration methods are feasible. However, such methods are often empirical in nature due mainly to the impossibility of totally isolating information of the compound of interest from the rest of the matrix information. Therefore, analysis of the nature of the matrix as well as the calibration method development require efforts in order to properly use relevant spectral variability ensuring that such variability actually characterizes the matrix in term of chemical composition with the highest possible confidence.

1.2.1.2 Cell culture and NIR analysis

In contrast to single component spectroscopic analysis, biopharmaceutic cell culture is a complex and challenging matrix, where absorbance/transmittance of light is not the only phenomena. There is a strong change in the matrix due to the progression of the cell culture process, involving chemical changes, for example due to the metabolism of cells (consumption of substrates, synthesis of products and by-products) and physical changes due to scattering compounds (mainly cells, cell debris, among others).

Culture media for cell culture is complex, particularly for animal cell culture. Nowadays, synthetic media, usually classified depending on their supplements: serum-containing media, serum-free media, protein-free media, and total chemically defined media (Yao and Asayama 2017) are preferred for sanitary reasons. Undefined ingredients in culture media are particularly challenging since they reduce the reproducibility of kinetic profiles of cell cultures and may require more effort to use spectra in a predictive manner. Furthermore, this initial variation impact on cell physiology may cause slight differences in by-product concentration in culture media. This is of great concern since molecular vibrations are also a function of the presence of other molecules adjacent to vibrating bonds. Thus, initial variability is accumulated through the cell culture process, which complicates the use of spectra for monitoring purposes. Moreover, in in-situ mode there may be by-products capable of producing biofilms on probes. All these non-ideal phenomena in cell culture processes, related to chemical variability, must be considered for proper use of spectra.

Physical variability within cell culture processes are also of great concern since it strongly affects the produced spectra. The most relevant phenomenon is light scattering, particularly due to cells, cell debris, protein aggregates, among others. This is mainly a function of two properties (Næs 2004): the number of light-surface interactions (size and shape of particles) and differences in refractive indices.

Light scattering alters the intensity of light absorbance/transmittance and subsequently the calculated concentration of the absorbent species; modeling scattering may become extremely difficult because of random variation of size and shape of particles in culture media, so it is essential to separate physical scattering effects from chemical vibrational light absorbance effects for developing proper spectra to be used in prediction. Light scattering

usually comprises three principal effects: multiplicative effect, additive effect and wavelength-dependent effect.

In addition to chemical and physical variability, cell line variability is also of concern. Cell variability may be from genetic or physiological origins, causing process yield impairments. Although it may be controlled to some extent by proper cell banking, during expansion of subcultures for production scale-up, there can be a loss of cell-specific productivity gradually or precipitately in only a few generations (Kim et al. 2011). This fact is particularly of great concern for long-term cell culture processes (Bailey et al. 2012) such as fed-batch, continuous and perfusion cultures. This cell line variability is also more evident for plant cell suspension culture where the proliferation nature, in the form of cell aggregates, may force biopharmaceutics production by sensitive cell lines into somoclonal variation (Torrey 1975; Deus-Neumann and Zenk 1984) or even into polyclonal cultures (Nocarova and Fischer 2009).

Although several studies in academia have shown successful application of NIR technology for monitoring cell cultures, its application in industry has been limited to production bioreactors (Jenzsch et al. 2017). This could likely be explained by limited variability in terms of raw material (Jenzsch et al. 2017), inoculum variation and also by some limited cell line instability seen in cell culture runs performed in academia, in comparison with variability commonly seen in industry.

1.2.2 Chemometrics for monitoring cell cultures using NIRS: scattering effects

Chemometrics as a discipline is relatively new; the term was coined in the 1970's and it was strongly related to scientific computing and multivariate statistical methods. Chemometrics main issue is the analysis and interpretation of instrumental data, which also include application of computational algorithms to handle data. Perhaps the major application of chemometrics is calibration, where one type of measurement is used to predict, calculate or estimate the value of a parameter, such as CPP or CQA. Process chemistry needs were perhaps the main motor for chemometrics development since a process chemist might have wished to monitor the quality of products (hydrocarbon compositions in refineries, among others) on a continuous basis by spectroscopy, then corrective action could immediately be taken if deviations from accepted limits were detected (Brereton 2007).

Briefly, a multivariate calibration could be seen as resolving the relationship [Y] = [M] [X], [Y] representing a concentration value set of a compound to be explained, and [X] a spectrum set of the corresponding culture media as an explanatory variable. Then once [M], a regressor expression representing the relationship between spectra and concentration, is solved, the calibration can be used for estimating the concentration value (y) at certain point using only a spectrum (x) each time the NIR analyzer performs a culture media scanning. There are several regression techniques for solving [M], such as Principal Component Regression, Partial Least Squares Regression, among others. Though regression is a critical phase for calibration, it comprises several stages that requires continuous enhancement; it is usually a loop process during model development and routine validation (Burns and Ciurczak 2008). Because of the dynamic nature of cell culture, there must be regression analysis between the columns of [X] (absorptions at different wavelengths) and rows (representing different samples). Samples are taken from different stages of cell culture progression; not only a particular compound concentration varies, but also the scattering nature of culture media. As calibration is often focused on concentration estimation, there is usually a previous stage for management of scattering effects within spectra of different samples. This management is performed by analysis and correction of spectra by techniques called spectral pre-treatments. The objective is to eliminate or minimize physical variability unrelated to the property of

interest, such as concentration, so that changes in spectra due to chemical information could be more effectively modeled.

The scattering effects in NIRS (Figure 8) are often summarized as: 1) Additive, 2) Additive and Multiplicative, and 3) Additive, multiplicative and Wavelength-dependent effects.

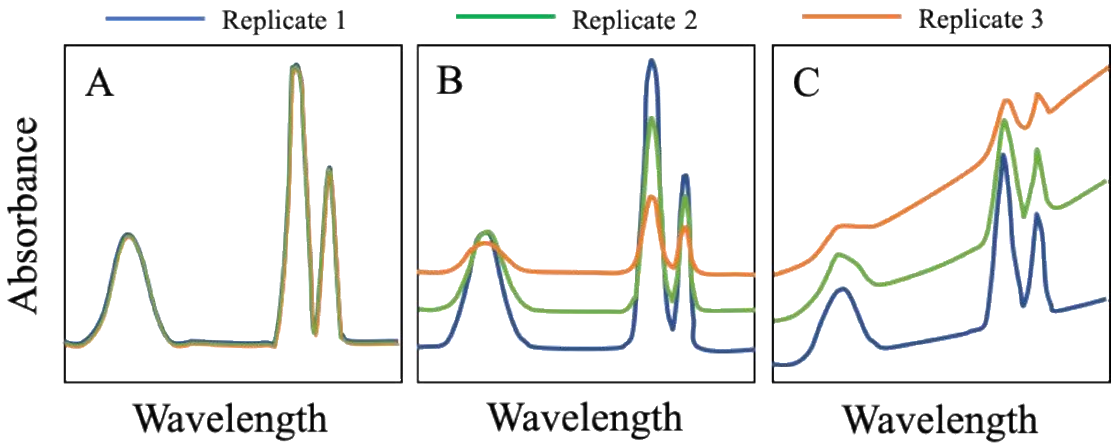

Figure 8. NIR spectra with scattering effects: A Spectra without scattering effects (replicates are perfectly overlapped). B Spectra with multiplicative (simple baseline offset) and additive (spectral scaling by a given factor as shown in the peaks) scattering effects. C Spectra with additive, multiplicative and wavelength-dependent (nonlinear baseline shift) scattering effects.

Scattering is a complex phenomenon and could be linear or non-linear, making it difficult to minimize or remove. Its correction can be made through explicit and implicit scatter correction, the former by particular spectra pre-treatment and the latter by the regression itself, which can compensate for unknown scatter variation at the expense of increasing model complexity. However, common linear regression methods (PCR/PLSR) are not effective for accommodating the impact of scattering, particularly multiplicative scattering (Martens et al. 2003). To minimize the impact of scattering effectively, explicit correction through appropriate data pre-treatment is needed.

Spectral treatment is not a trivial step when building calibration models. Although selection of best spectral pre-treatment is usually performed by assessing which pretreatment improves model performance by comparison with subsequent models, an approach that must be used

with care since it could also reduce the compound signal. Therefore, analysis of main scattering effects should firstly be performed, and primary selection of pretreatments undertaken; subsequent trial and error schemes would likely enhance model performance with better chances. The most common pretreatments are the use of spectral derivatives, spectral truncation (ST), and baseline corrections such as Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC) and Extended Multiplicative Signal Correction (EMSC).

Derivatives can resolve peak overlapping and enhance the resolution of spectra; they can also eliminate baseline drifts between samples though there could be noise enhancement which may complicate interpretation of spectra. Main tuning options for derivatives are specified gap distance, and Savitzky-Golay polynomial order fitting. SNV is basically an autoscaling of the rows instead of the columns of the spectral matrix; correction is performed individually and does not need data from all other spectra. MSC is often used for baseline offsets and multiplicative effect corrections. MSC regresses a measured spectrum against a reference (usually the mean of spectral matrix) and then corrects the measured spectrum using the slope and intercept of this linear fit. Wavelength-dependent light scattering variation is difficult to remove using either SNV or MSC. EMSC has been proven useful for management of such scattering effects. It is intended to separate physical light-scattering effects from chemical absorbance effects. ST is based on the fact that there may be spectral regions containing, or not, information of a particular compound. Only those closely related to compounds are used for further calibration. ST can bring several advantages such as eliminating unrelated variables and noise, more robust models with fewer components, among others. However, it requires previous knowledge of these particular regions or the use of iterative regression analysis (Huang et al. 2010).

Analysis of spectra is essential for building a proper NIRS monitoring procedure since it represents half of inputs for the regression method while the rest is compound concentration data. Once both data have been properly analyzed and verified, then the calibration model can be built using any regression method, briefly described in next sections below.

1.2.2.1 The calibration method

The final aim of an NIR method is to perform routine chemical analysis for controlling processes based on NIR spectra. Thus, multivariate techniques must be used to produce calibration methods. Although the calibration process may look basic, it becomes quite complex when dealing with biological matrices with overlapping and interfering bands, scattering effects, among others. It requires adequate and multidisciplinary knowledge of NIR technology since sophisticated regression methods, proper sample selection and design, instrument standardization and scatter correction, among other issues, may be needed (Burns and Ciurczak 2008).

In the case of cell culture processes, the first step for calibrating a method should address the design space, defined as the multidimensional combination and interaction of process parameters that have been demonstrated to provide quality assurance. Thus, calibration samples must address variability and all possible scenarios during routine control of processes to assure medicine quality. This could be an extremely difficult task considering the nature of cell culture-based processes. In this way, diverse options may be undertaken for preparing calibration samples with some confidence, for instance, by including the full range of compounds concentration as evenly distributed as possible while also considering a uniform matrix distribution. The calibration method would thus have been trained using samples representing the nature of samples expected during process routine analysis.

Once the nature of the calibration set has been established in term of process variability, caution must be also taken in off-line analysis, for instance, when calibrating for protein in biological samples a Kjeldahl procedure could be used as primary method for providing reference values. However, since NIR spectrum include information of peptide bonds directly, but not of reduced nitrogen, the NIR calibration procedure values would never perfectly agree with off-line data. Another example is about matrix interaction differences with concentration changes as the refractive index of liquids often change with concentration, which is particularly important when using volumetric or gravimetric techniques (Workman 2008).

According to the “garbage-in, garbage-out” adage, these facts must be carefully considered. If chemical and spectral calibrations sets lack quality and rigor, model performance would

be poor. Therefore, proper collection of samples representing the population for routine analysis is perhaps the most critical step during calibration. The analysis of calibration samples is also critical, for example for detecting outliers, data points that differ significantly from other observations due to variability in the NIR measurement (NIR instrumental or spectral aberrations) or experimental error (off-line analysis). Criteria for selecting outliers are often subjective and require expertise in both cell culture process and chemometrics prior to discarding any samples from the calibration set. Principal Component Analysis (PCA) is often used for analyzing spectra and detecting outliers in small Principal Component (PC) space. The aberrant variability of outliers can then be detected through two main phenomena: a large PC value, and a large residual value (Q).

The Hotelling's T_2 distribution can be performed at some confidence limit (*i.e.* 95 % or 99 %) seeking if there is any sample with PC values of different populations not corresponding to the calibration sample main population. This approach can also be performed for Q-residuals. Both phenomena (the Hotelling T_2/Q-residual criterion) can be simultaneously used to infer if the calibration sample is actually an outlier provided the calibration sample failed in both cases (Burns and Ciurczak 2008).

As formerly stated in section 0, calibration models are not totally causative but are also empirical in nature. Therefore, their performance must be evaluated to ensure a high level of confidence during routine analysis. Consequently, the main calibration sample is usually partitioned into a calibration set and a validation set. The calibration model is generated through multivariate statistics, which include regression itself and spectral pre-treatment for scattering effects. The confidence level of calibration is often evaluated through the validation set. The validation set must also consider all variability to ensure proper performance during routine analysis. Finally, successful calibration models are launched for routine analysis; during this period, models are also monitored to ensure the model validity and modified if required. The main steps for building a calibration method are summarized in Figure 9.

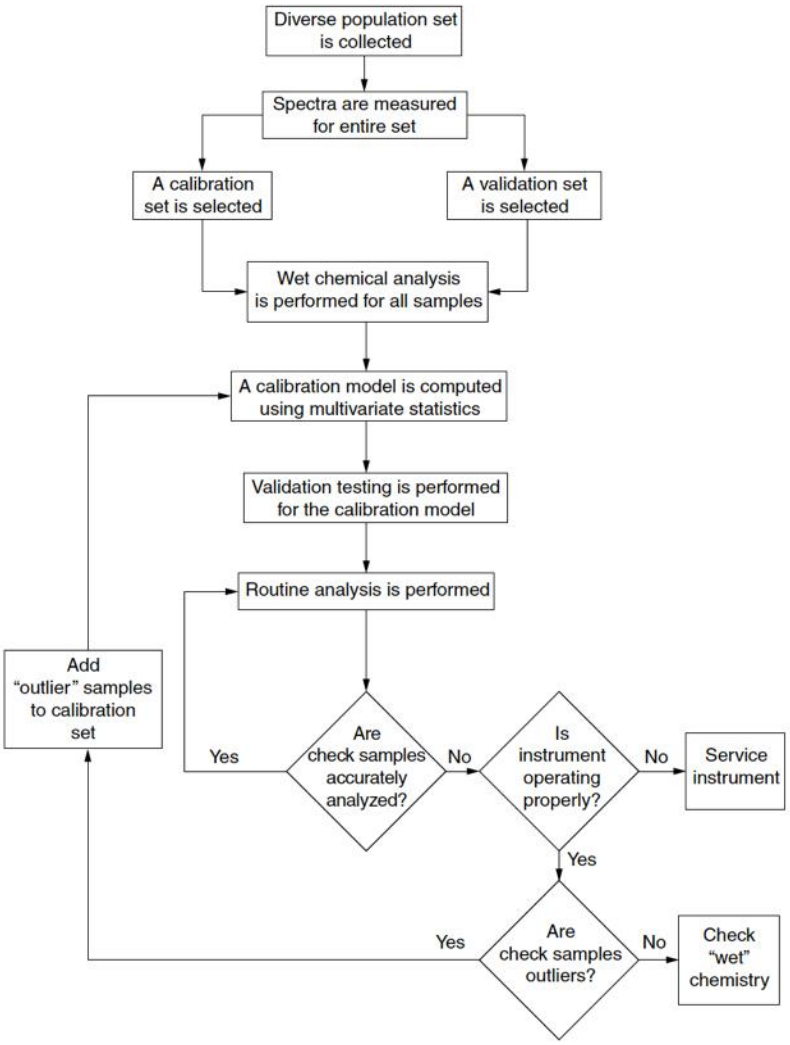

Figure 9. Flow chart for calibration (Workman 2008)

As can be noted from Figure 9, the main core of calibration is regression using multivariate analysis. Briefly, a multivariate calibration could be seen as resolving the expression Y = M.X + Error, Y representing a set of compound concentration values and X a spectrum set of culture media. Then the expression can be solved for M so that it could be used as a regressor parameter to estimate or predict particular concentration values (y) using only a spectrum (*x*). There are plenty of regression alternatives to solve M, each one with particular benefits and drawbacks. There are three main regression methods perceived as classic methods for spectroscopic data: Multiple Linear Regression (MLR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR). MLR has perhaps been the

most used regression method due to its simplicity and efficacy, analysing linear correlations between independent variables (absorptions) and a single dependent variable (compound concentration). However, its major constraint is its inability to handle collinearity where some of the independent variables are highly correlated as in NIR spectra. Moreover, as it is basically a system of polynomial equations, it then requires at least the same number of calibration samples as independent variables (absorptions at different wavelengths). This is of great concern since new NIR analysers can perform spectral scanning with hundreds or thousands of independent variables, which could lead to an undetermined nature of the regression (inconsistent regression or with infinitely many solutions). Therefore, PCR and PLSR which firstly reduce collinearity and the number of independent variables, are mostly used in NIRS applications. The principles of PCR and PLSR methods are discussed in sections 0 and 0, respectively.

An ideal calibration model would predict or estimate a reproducible, accurate and reliable concentration value of a compound within the analyzed matrix. It should be robust and properly handle endogenous instrumental variation, background interferences, temperature effects and at the same time, it should be quite sensitive to concentration changes. Although the regression can be solved for perfectly paired y values with spectral changes, the main purpose of calibration is to obtain a general mathematical expression allowing accurate estimation of concentration in any sample during the processes, a model property called model fitting or generalization. In other words, ideal models must extract the nature of the processes using only a small sample of it, which is the calibration set. This property is shown in Figure 10, where calibration samples are represented as points in two dimensions (time values and concentration values) and three different regression models with different fitting natures are shown. Underfitting and overfitting models did not capture the changes well, while the model with good fit properly captured the nature of the process. As the model with good fit is likely to properly pair variability of the process, it is likely to be more robust and perform more accurately during routine analyses.

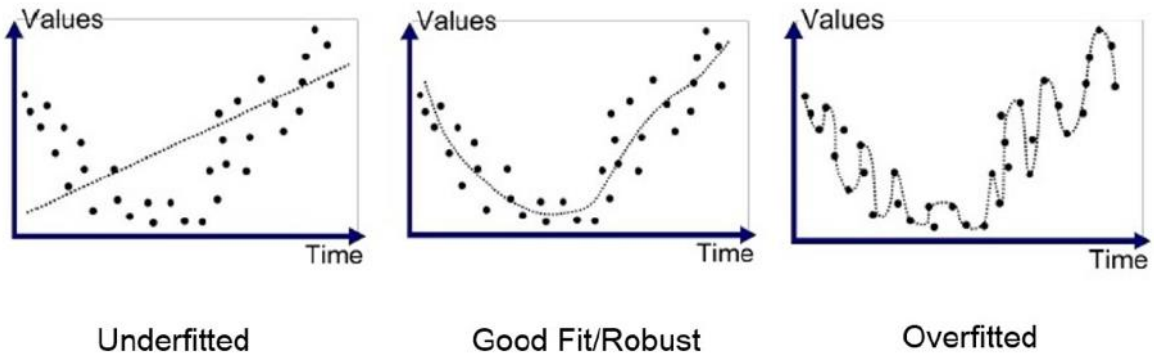

Figure 10. Fitting of a model

The capacity of a regression technique for matching values relies on the nature of the regression method and the relationships between dependent variables (compound concentration) and independent variables (spectra). Such relationships may be linear or nonlinear, which would require different regression treatments for pairing data. Moreover, there are two main types of nonlinear relationships: univariate and multivariate non-linearities (Næs 2004). As shown in

Figure **11**, there can be strong nonlinear relationships between variables X_1 and X_2 on one side, and with y on the other. This is called univariate nonlinearity as y could be set as a function of both X_1 and X_2, then the relationship between X_1-X_2 and y becomes linear. As univariate nonlinearity can be transformed into multivariate linearity, this phenomenon can easily be resolved with linear regression approaches. In most applications for vibrational spectroscopy calibration, the multivariate calibration model is linear; this means that the calibration model for each component is mainly based on an equation type represented by Equation 3:

$$Concentration = b_0 + \sum_{k=1}^{k} b_k x_k + error$$

Where b are regressor coefficients and x, variables related to absorption

Equation 3

On the other hand, there may be nonlinearity causing a totally nonlinear relationship between y and all the x-variables at the same time (multivariate non-linearity) and would then require nonlinear regression approaches (Perezmarin et al. 2007). The principal strategies to handle nonlinearity and to reduce the estimating errors are sophisticated spectral preprocessing, adding nonlinear terms to calibration equations, splitting data into subsets and deleting variables. Although such strategies may result in efficient calibration models, there may be cases where the only possibility to enhance prediction power is simply to try nonlinear multivariate calibration models.

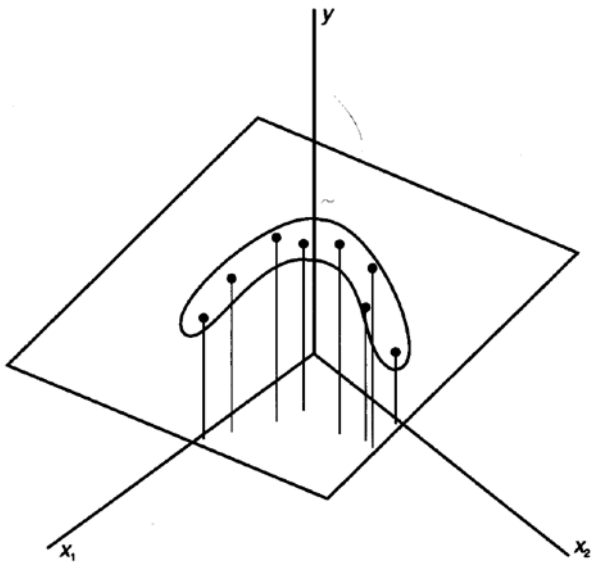

Figure 11. Nonlinear relationships (Næs 2004)

The ideal regression model should be as precise and accurate as possible; thus, evaluation criteria are essential not only for choosing between several alternative calibration models, but also on selecting the structure of the multivariate calibration technique such as number of latent variables for PLS, principal components for PCR, among others. Comparison of calibration models is not a trivial task and it could become difficult to compare in detail, especially when calibration models of different natures are involved. Perhaps the most robust and employed criterion for evaluating the acceptance of calibration models are several standard methods (Burns and Ciurczak 2008) such as the coefficient of determination (R^2), the root mean squared error (RMSE), the standard error of prediction (SEP), among others,

using calibration set, cross-validation set or external validation set approaches. The use of these parameters under different set approaches, is intended to assess how the calibration will generalize to an independent data set. Analysis of calibration performance with a calibration set approach only shows if the regression model can pair calibration spectra and concentration values, while using cross-validation set and external validation set approaches analysis the performance more rigorously. The validation set approach mostly is the best way to evaluate if the calibration procedure has properly assessed the cell culture process phenomenon since only independent-of-calibration samples are used. On the other hand, the cross-validation set approach involves partitioning the calibration data into complementary subsets, performing the analysis on one subset (the training set), and validating the analysis on the other subset (the testing set). Multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds as shown in Figure 12, to finally give an estimate of the predictive performance of the calibration model. Though it could be perceived that cross-validation would contribute less to confidence analysis than by using an external validation set, it is quite helpful since it analyses the whole variability extension of the cell culture process through the spectral calibration set, while the use of an external validation set could likely depict only a small fraction of the cell culture process variability.

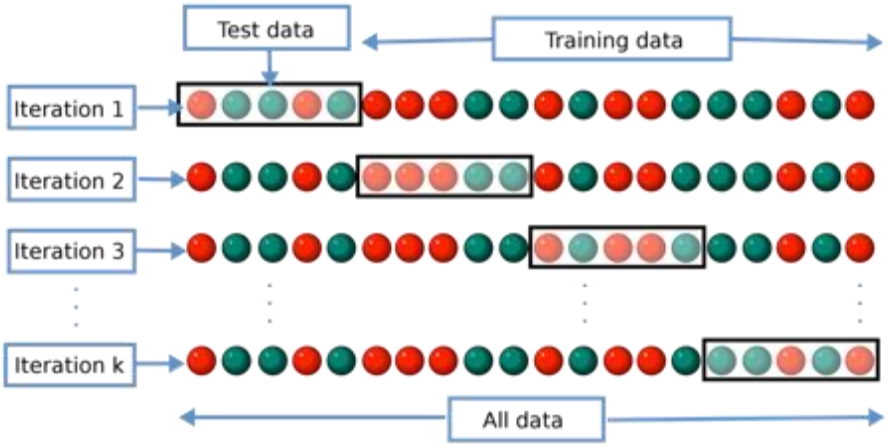

Figure 12. Cross-validation approach

All former standard methods are aimed at evaluating the acceptance of models from different perspectives, for instance, accuracy, precision and linearity. Accuracy refers to the difference between model predicted and actual y-values, while precision refers to the difference of repeated measurements, and linearity to the capability of a model to obtain results directly and linearly proportional to actual concentrations. Perhaps the most common standard methods for evaluating the model performance for cell culture issues is the root mean squared error (RMSE) either using external validation sets (usually reported as RMSEP) or during cross-validation (RMSECV) for accuracy, and R^2 for linearity, which are calculated using equations 4 and 5 (Workman 2008):

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{N_p}(\hat{y}_i - y_i)^2}{N_p}}$$

Equation 4

$$R^2 = \frac{\sum_{i=1}^{N_p}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{N_p}(y_i - \bar{y}_i)^2}$$

Equation 5

Where $\hat{y}$ is the estimated concentration value, $\bar{y}$ the mean y value and y_i a singular y value for the *i*-th sample, and N_p the number of calibration samples

Other important criteria when evaluating a calibration model are the grade of uncertainty of a prediction and the detection limit. The uncertainty of the model is usually characterised by the confidence intervals or confidence limits of the calibration; this interval or limit contains the unknown y-value (concentration) that we want to estimate with a high probability (≥ 95 %). Due to the nature of the calibration process, theoretical values are usually replaced by empirical correlations, one of the most employed being to establish a confidence interval of ± 2 RMSEP, approximately 95 % confidence interval, when residuals from a NIR calibration is close to t-distribution (Næs 2004). The detection limit is the lowest quantity of a substance that can be distinguished from the blank. In NIRS applications, it can be loosely

approximated as three times the SEP or twice the RMSE, an approach with approximately 99 % confidence that difference in *y*-value between one sample and the blank is correctly determined (Workman 2008).

Comparison of different calibration models is usually performed by comparison of performance variables such as RMSEP or R_2. Simple model comparisons employ validation sets (cross validation sets, internal validation sets or preferably external validation sets) and performance variables are compared, then the best values are selected (Balabin et al. 2007). However, this simple approach lacks significance of the differences and the effect of the validation sample remains unknown.

A good methodology to compare calibration models should consider not only prediction ability, but also aspects such as interpretation issues, outlier detection, and ease of use, among others. As it may become difficult to cover all these issues, the great majority of authors have focused on prediction ability. Some statistical methods for significance difference in prediction may be employed such as an ANOVA test, a sign test, a Wilcoxon test and a Friedman test, among others. Multiple comparison methods such as Tukey tests are feasible when comparing several calibration methods (Næs 2004).

Great attention to comparison issues in spectroscopic calibration models have been paid, which concluded that the most suitable method is the two-way ANOVA test with post-hoc Tukey test (Cederkvist et al. 2005). An alternative to the ANOVA test is a Friedman test, a preferred alternative when the assumption of normality is not met. In contrast to the ANOVA test which uses raw observation values, the Friedman test uses a data rank table to calculate the statistic, thus the test is based on the squared differences between the sums of the ranks of each treatment from the overall mean rank sum (Forthofer and Lee 1995).

When a confidence model is finally used in routine analysis, it must also be monitored to assuremodel validity through time and process variations. Validation schemes will depend on company politics and regulatory issues, though it is a common approach to use Shewhart charts with warnings and action limits at particular thresholds. Common rules for action are seven points in a row on the same side of the zero-line and one point outside the action line, among others (Næs 2004).

1.2.2.2 Principal Component Regression

Principal component regression intends to reduce collinearity by PCA compression of the X input matrix, and only PCA scores are employed as independent variables to build the model. Næs and Martens (1988) have described the PCR algorithm with focus on the NIR spectra context and part of their work is summarised here as an introduction to PCR calibration generation. The first step is to perform X matrix decomposition into principal component scores as described in equation 6:

$$X = TP^T + E$$

Equation 6

Where T is the score matrix, P_T the transpose matrix of loadings and E the residual matrix. Once the X matrix has been mapped into the score space and the number of principal components chosen, it can be represented by the score matrix to calculate the model according to equation 7:

$$Y = TB + E$$

Equation 7

where Y is the variable matrix (in this case the variable concentration matrix) and B the regressor matrix subject to be calculated by ordinary least squares as indicated in equation 8:

$$B = (T^T T)^{-1} T^T Y$$

Equation 8

Once regressors matrix is calculated, the model is generated, and problem samples may be calculated employing the models. To calculate the variable concentration, the spectral data is transformed into the principal component space and scores matrix of the samples (T*) is determined from spectra (X*) and loadings matrices as shown by Equation 9:

$$T^* = X^* P$$

Equation 9

Then the regressor matrix, calculated in the calibration process, is employed with the problem sample score matrix to determine variable concentration according to the PCR calibration equation depicted by Equation 10:

$$Y = T^{*}B$$

Equation 10

Collinearity problems completely disappear with PCR models since the PC are orthogonal and only a few PC are required to explain the majority of calibration set variability. However, there is one major constraint with PCR: the regression equation will consider all spectral variables (because each PC is a linear combination of the several absorptions at different wavelengths) and PC correlated to global process variability rather than to variability related to the variable being predicted (compound concentration). Consequently, new approaches taking this fact into consideration have been developed such as PLSR.

1.2.2.3 Partial Least Squares Regression

The basic algorithm for PLS regression was proposed by Wold et al. (1984). Nowadays, PLS regression is perhaps the most employed multivariate regression method, particularly in the NIR context. As with the PCR method, PLS is based in a reduction variable process for treating collinearity. In contrast to PCR which only reduces X-spectral matrix and then relates reduced X-matrix with Y-matrix, PLS finds components from X matrix that are also relevant for Y. PLS models with the constraint that these components explain as much as covariance between X and Y as possible and regression of matrices are then performed according to equations 11 and 12 (Höskuldsson 1988):

$$X = TP^T + E$$

Equation 11

$$Y = UQ^T + F$$

Equation 12

where X and Y are spectral and concentration matrices respectively, T and U are the pseudo-score matrices, P and Q are the pseudo-loading matrices and E and F are the residual matrices. Matrix decomposition of X and Y matrices are not independent, thus an internal relationship between the scores of X and Y are generated according to Equation 13:

$$U = BT$$

Equation 13

where U is the pseudo-scores of Y to be calculated, T the pseudo-scores of X, and B the regressor matrix. Once the regressor matrix has been determined, calculation of y-concentration value from the problem sample may be calculated according to the PLS calibration equation depicted as Equation 14:

$$Y = T^*BQ^T + F$$

Equation 14

where T* is the pseudo-score matrix of the problem sample, B the regressor matrix, Q_T the pseudo-loading matrix of the model and F the residual matrix.

1.2.2.4 Locally Weighted Regression

Locally weighted regression (LWR), also called LOWESS (LOcally WEighted Scatter-plot Smoother) was inspired by time series methodology where data points are smoothed by local fittings of polynomials, but the name of LWR and its application as a regression technique is attributed to Cleveland and Devlin (1988). In contrast to classic parametric multivariate calibration which generates a regression function considering all calibration points (Figure 13-A), LWR defines a neighborhood in the space of the independent variables that contains the sample to be predicted (Figure 13-B); each point of the neighborhood is weighted according to its distance from the sample to be predicted. Points close to the sample to be predicted are given more importance or weight, points far from the sample to be predicted are given less weight (weights represented by contour thickness of local calibration samples in Figure 13-C). A regression function is then generated employing local calibration samples and their weights (Cleveland and Devlin 1988).

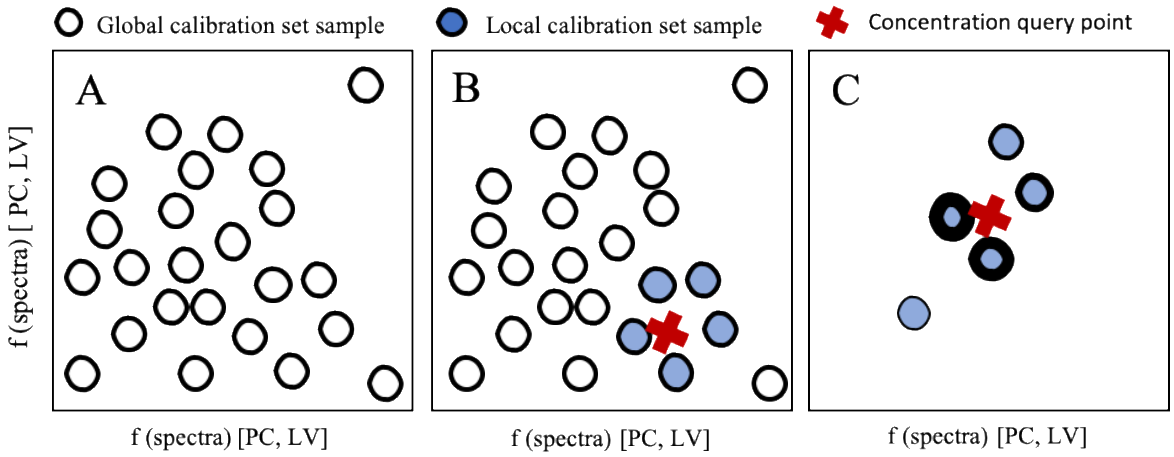

Figure 13. Intuitive concept of LWR. A Global calibration samples. B For a new query point (red cross), only a local set of relevant and near calibration samples are selected (blue circles). C calibration is generated with only local samples and weighted PLSR

In addition to the generation of the regression function for the neighborhood region, LWR requires additional parameters to be built, such as those that determine the neighborhood region and the weights of the points. The neighborhood region is determined by a distance

function and specified limit. Euclidean distance has been employed for LWR mainly in geography contexts, also in NIRS regression problems only when NIR spectra have been mapped into principal component space (Naes et al. 1990; Naes and Isaksson 1992; Centner and Massart 1998). Næs (2004) have also proposed several approaches related to the use the Mahalanobis distance in the principal component space. Since distance as a delimiting criterion may be inappropriate when lacking vast calibration samples in a wide calibration space, several authors have employed distance criteria in terms of a number of near calibration samples (Naes and Isaksson 1992; Næs 2004).

Once the local area is determined, each calibration point must be weighted for building the regression equation. Perhaps the most employed weight function is the LWR Cubic weight function proposed by Cleveland and Devlin (1988) and employed by the great majority of studies, defined below by Equation 15:

$$w_i(x_j) = W\left(\frac{\delta\left(x_j, x_i\right)}{d(x_j)}\right)$$

Equation 15

where

$$W(u) = \begin{cases} (1-u^3)^3 & if\ u \leq 1 \\ 0 & if\ u > 1 \end{cases}$$

$$\delta\left(x_j, x_i\right): Distance\ between\ prediction\ sample\ j\ and\ calibration\ sample\ i$$

$$d(x_j): maximum\ distance\ involved\ in\ each\ regression$$

As could be noted, the weights will be large (close to 1) for x_i close to x_j, and small (close to 0) from x_i far from x_j. Once the region and weights have been determined, the regression function in the local region is generated using weighted least squares, and the y-value for x_j is calculated. Although LWR Cubic weight function is perhaps the most employed weight function, various authors have proposed other functions such as the uniform weight function (Naes and Isaksson 1992; Centner and Massart 1998).

The nature of LWR makes it attractive to employ the general framework of least squares in order to model complex processes for which no theoretical model exists. On the other hand, as LWR is a local model, it requires fairly large and densely sampled data sets to produce efficient models, which increases experimental costs and makes it sensitive to outliers.

According to Naes and Isaksson (1992), LWR may reduce prediction error up to 15 % with respect to global and lineal multivariate regression. (Centner and Massart 1998) have worked on LWR parameter optimization (distance function, weighting function, nature of calibration sets, among others) for improving model prediction power and simplicity to reduce the requirements of computer processing power. Their results showed that for non-lineal heterogeneous data sets LWR yielded better results than global PLS and their work also showed that in extrapolation prediction cases, the use of LWR resulted in error reductions of up 75 % in reference to PLS prediction errors, thus, minimizing the risk of losing accuracy when non-linearity is presented in real-time process monitoring. One drawback of LWR is that only distances in independent variable space (spectral space) are considered. Wang *et al*. (1994) developed a new approach of LWR called LWR2, which included distance in the dependent variable space (chemical space).

Although LWR generates a neighborhood according to the closeness of calibration points to prediction samples in spectral space, they may not necessary be close in chemical space; LWR2 considers this issue and aims to be a more robust calibration technique. Wang et al. (1994) have proposed the following distance function which includes a distance function (Equation 16) in the spectral space (as LWR does) and a function corresponding to the distance in chemical space (concentration values):

$$\rho_{ip} = \alpha \rho Y_{ip} + \beta \rho X_{ip}$$

Equation 16

where

$$1 = \alpha + \beta$$

$$\rho Y_{ip} = \frac{\left[y_i - \ddot{y}_{p\,(PCR)}\right]}{\delta_p};\ \delta_p = \sum_{i=1}^{N} \rho Y_{ip};\qquad \rho Y_{ip}^{j+1} = \frac{\left[y_i - \ddot{y}_{p\,(LWR2)}{}^{j}\right]}{\delta_p}$$

ρ_{ip}: *Distance between calibration sample i and prediction sample p in*

$$both\ chemical\ and\ spectral\ space$$

$$\rho Y_{ip}: Distance\ in\ the\ chemical\ space;$$

$$\alpha: importance\ of\ distance\ in\ the\ chemical\ space$$

$$\beta: importance\ of\ the\ distance\ in\ the\ spectral\ space$$

$$\rho X_{ip}: Distance\ in\ the\ spectral\ space$$

$$y_i: concentration\ value\ for\ calibration\ sample\ i$$

$$\ddot{y}_{p\ (LWR2)}: concentration\ value\ for\ sample\ p\ by\ LWR2\ method$$

$$\ddot{y}_{p\ (PCR)}: initial\ concentration\ estimate\ for\ the\ prediction\ samplep\ from\ PCR$$

$$\delta_p: Parameter\ to\ make\ pY\ unitless\ and\ between\ 0\ and\ 1\ like\ \rho X$$

$$j: number\ of\ iterations$$

As could be noted in prediction samples, the neighborhood in chemical space remains unknown since no data of concentration is available, thus it is necessary to calculate it by other techniques (PCR, PLS), then an iterative process is launched until convergence. LWR2 also requires two more parameters to be optimized: contribution of distance measured in chemical space (α) and the number of iterations (j). According to Wang et al. (1994), LWR2 can handle non-linearity and noise in spectral space better than LWR. Moreover, LWR may also be used as a convenient platform to include new samples that were collected after primary calibration, allowing model adaptation to process changes while maintaining expected performance during process conditions (Chang et al. 2001).

1.2.2.5 Supported Vector Regression

A relatively novel alternative for nonlinear modelling of NIR spectra is SVR (Cogdill and Dardenne 2004). The main difference of SVR from other typical regression methods is that its objective is not merely to reduce the fitting error but to fit the error within a particular threshold (±ε). Dealing with calibration sample sets with complicated relationships between spectra and compound concentration may be an extremely difficult task for classic regression methods. For instance, the calibration sample distribution in Figure 14-A cannot be separable or regressed based on concentration, using only two dimensions. However, if data is properly mapped into a higher dimension space (called feature space), it could be separated by a regressor or hyperplane taking into account compound concentration as shown in Figure 14-B. In this case a regressor with a 3D spiral shape was generated (red line in Figure 14-B), requiring the addition of one more dimension, but it can occurr that hundreds or even thousands of dimensions may be required for proper error fitting within a particular threshold. A closer look at the hyperplane is shown in in Figure 14-C; this is defined by the support vectors (SV), which are the calibration samples in the positive and negative limits of the tolerated threshold. The hyperplane is then located in path half-way between positive and negative threshold while majority of samples remain in the threshold tube.

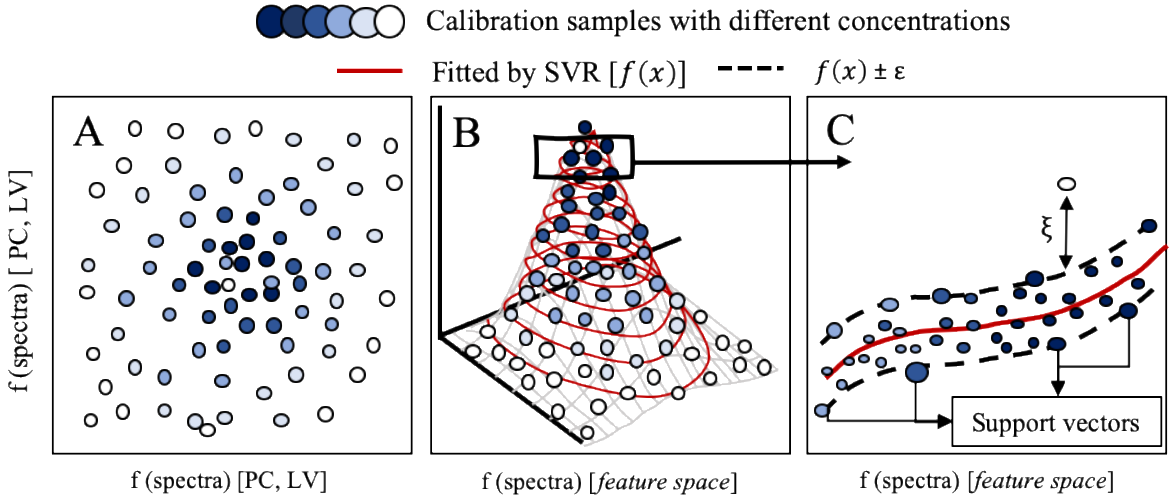

Figure 14. Intuitive concept of Support Vector Regression. A Calibration set. B Continuous classification of calibration samples by a kernel in a higher dimensional feature space. C Hyperplane or regressor structure showing the support vectors.

The goal of SVR is to generate a regression function, or hyper plane, that has a maximum number of calibration samples at most an ε deviation from an actual concentration, and at the same time keeping the function as flat as possible (Smola and Schölkopf 2004). The hyper plane could be considered as in Equation 17:

$$(x) = (wx) + b$$
$$with \quad w \in \mathrm{X}, \qquad \mathrm{b} \in \mathbb{R}, \qquad \text{and x being a variable related to spectra}$$

Equation 17

Flatness is then assured by minimization of w, for example by minimizing the norm as a convex optimization problem as shown by Equation 18:

$$minimise\ \frac{1}{2}|w|^2$$

Subject to

$$y_i - (wx) - b\ \leq \varepsilon \quad and \quad (wx) + b\ -\ y_i \leq \varepsilon$$

Equation 18

However, it may not be the case that f(x), which approximates all pairs (x_i, y_i) with ε precision, actually exists. Then a soft margin of slack variables ξi, ξ∗i, are introduced for coping with unfeasible constraints of optimization (Equation 19) as stated by Vapnik (Vapnik 2000):

$$minimise\ \frac{1}{2}|w|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

subject to

$$y_i - (wx) - b\ \leq \varepsilon + \xi_i\ , \quad (wx) + b\ -\ y_i \leq \varepsilon + \xi_i^* \quad \text{and} \quad \xi_i.\xi_i^* \geq 0$$

Equation 19

The constant C determines the compensation between the flatness of f(x) and the amount up to which deviations larger than ε are tolerated. This general procedure is depicted in Figure 14-C. As could be observed, SVR is then less vulnerable to outliers since it could properly generalize and leave the outliers in the soft margin (closed circle in Figure 14-C). In complex

multivariate data optimization, Equation 19 can be solved more easily in its dual formulation, which provides the possibility for extending the procedure to nonlinear functions. This could be achieved by mapping the x_i patterns into some feature space F (Nilsson 1965) as depicted by Equation 20:

$$\varphi: x \rightarrow F$$

Equation 20

Then standard SVR procedure is applied. Mapping into a higher, linear or nonlinear, dimensional space, may require exacerbated computational power, thus the majority of SVR use implicit mapping by kernels. The most common are linear, polynomial and Gaussian radial basis function (RBF) kernels.

The nature of the calibration set must be considered for properly selecting the kernel (Awad and Khanna 2015). The linear kernel is useful in large sparse data vectors with linear regularization, the polynomial may fit some soft non-linearity and RBF kernels are general-purpose, generally applied in strong non-linear regularization or in the absence of prior knowledge (Awad and Khanna 2015). This approach could be used for generalizing difficult-to-fit data in complex systems. As with LWR, monitoring procedure submission could be likely cumbersome for relating SVR parameters to specific chemical or physical properties of the cell culture, particularly in strong non-linear processes mapped into high dimensional feature space.

1.2.2.6 Neural Network Regression

Artificial neural networks (ANN) are inspired by biological neural networks. At first, they were focused on learning tasks by considering patterns in examples, but now they are used in diverse applications, especially for those that require difficult-to-express algorithms. In the context of NIRS calibration, ANN are mostly employed under the supervised learning paradigm (Naes et al. 1993). The core of an ANN is its basic unit, called the artificial neuron, which constitutes the building brick the network use for regression. The structure of a single neuron can be observed in Figure 15-A. The most common architectures for NIRS calibrations are single layer, multiple layer ANN and recurrent ANN (Figure 15). Since correction of PAT-NIRS calibration methods is a serious issue that usually requires official approval by regulatory agencies, recurrent ANN are not likely to be employed in the PAT-NIRS calibration process.

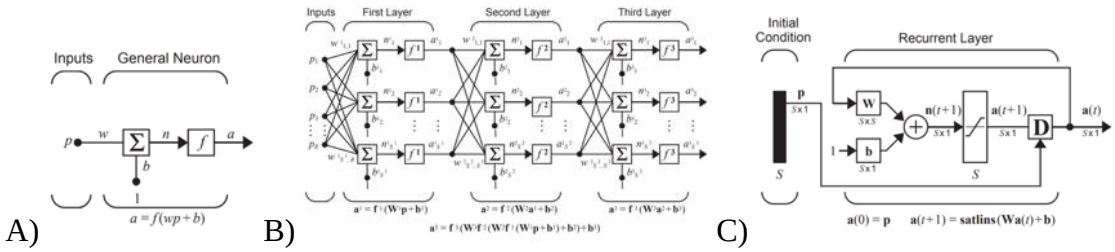

Figure 15: Artificial neural network structures: A) Single input neuron, B) Three layers ANN, C) Recurrent ANN (Hagan et al. 2004)

As can be seen in Figure 15-B and Figure 15-C, ANN could become as complex as necessary; however, the main idea of the networks rely on the neuron function (Figure 15-A; (Hagan et al. 2004)): The input p (absorbance at specific wavelength) is multiplied by the scalar weight (w) to form, one of the terms that is sent to the sum (Σ). The other input is a bias value b (1 in this case) and then passed to the sum. The sum output, often referred as the net input, goes into a transfer function (f), which produces neuron output a. Provided that a simple j multiple layer ANN with i inputs is employed, the neural network could be summarised in the equation 21:

$$y = f\left[\sum_{i=1}^{I} b_i f_i\left(\sum_{i=1}^{J} w_{ij} p_j + b_{i1}\right) + a_2\right] + error$$

Equation 21

According to Equation 21 of an ANN representation, w and b are adjustable parameters of the neural net; the ANN designer usually chooses the transfer function *f*. Thus, w and b must be tune by some learning protocol (also called learning algorithm) so that the neurons and then the whole ANN input-output relationship meets some specific goal (low error of prediction). Other parameters to be chosen are the number of neurons, the nature of the input data (compressed/uncompressed spectra) and specific neuron connections, among others.

There are several algorithms for training the networks; perhaps the first approach to accomplish this objective was the perceptron learning rule of Rosenblatt (Van Der Malsburg 1986) and the Least-Mean-Square algorithm of Widrow and Hoff. However, such protocols were design to train single layer networks only. To overcome this inconvenience, a learning rule called back-propagation was implemented and disclosed in the mid 1980s. Nowadays the back-propagation algorithm is perhaps the most widely used technique for training multilayer neural networks. Back-propagation calculates the error contribution of neurons, and then an enveloping optimization algorithm is used for adjusting the weight of each neuron. Technically it calculates the gradient of the loss function and is commonly used in the gradient descent optimization algorithm. It is also called backward propagation of errors, because the error is calculated at the output and distributed back through the network layers (Hagan et al. 2004). There are two main issues about ANN that are of great concern; topology and generalization of the network. Topology concerns the network structure (number of neurons, layers and connections). A small network may not be enough to fit the data. To make it clear Hagan et al. (2004) have given and explanatory example: An approximation to equation 22 by and ANNR model is desirable:

$$g(p) = 1 + \sin\left(\frac{i\pi}{4}p\right) \; for - 2 \leq p \geq 2$$

Equation 22

where i takes the values of 1, 2, 4 and 8. It can be noted that as i increases, the function becomes more complex, because it will have more periods of the sine wave over the interval. Thus, it will be more difficult for a neural network, with a fixed number of neurons in the hidden layers to approximate the function as *i* increases.

In Figure 16-A, the capacity of a neural network of one input, three neurons and one output (1-3-1) is shown. After proper training, it is clear that there is a limit to how complex a function this network can implement, i = 4 in this case. For i > 4, ANN tries to reduce the mean squared error between the network response and the function to be modelled, but is only able to match a small part of the function. In Figure 16-B, the same function is subject to fitting by several ANNs with different sizes and it can be observed how a complex function (or task) may require a minimum number of neurons to achieve the objective. This occurred because the number of hidden neurons it contained inherently limited ANN capabilities.

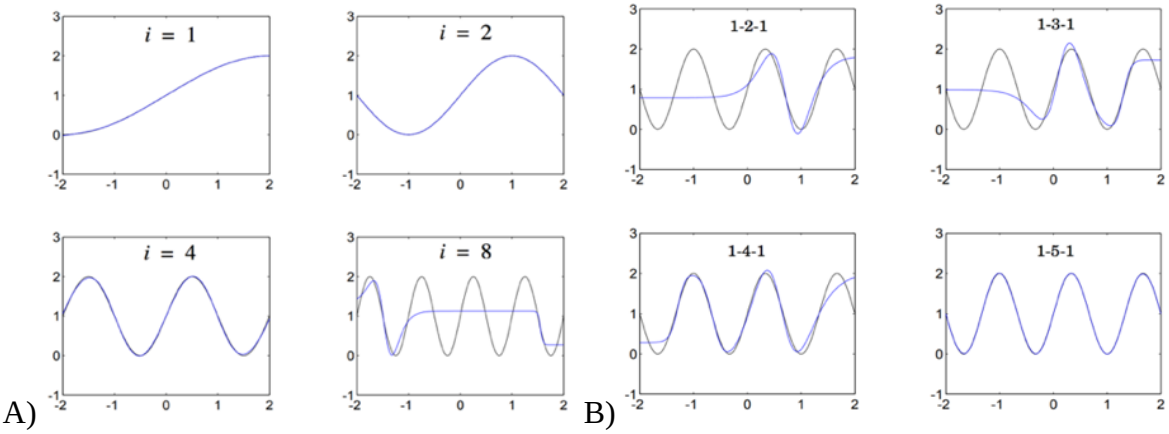

Figure 16: Capacity of ANN according to its size: A) Capacity of an ANN to fit a complex function, B) Required ANN size to fit a complex function. Network responses are shown by blue lines and black lines represent functions to be fitted by ANN (Hagan *et al.*, 2004)

It could be tempting to employ a large number of neurons, but it may lead to overfitting and in ANN context the antonym of overfitting is usually called generalization. In relation to NIRS calibration, the training set is to be representative of much larger samples of possible

input-output pairs (spectral data-concentration data), thus it is important that the network successfully generalizes the spectra-concentration relationship rather than performing a specific regression model for perfect pairing of calibration set data. In

Figure **17**-A, a function *g* is approximated by a 1-2-1 ANN trained with nine samples (+ symbols). The ANN shows good generalization, meaning that if the response of the ANN at a value that was not considered in the training set is to be found (-1), the ANN is capable of producing an output close to the modelled function. On the other hand, when a larger ANN (1-9-1) is employed (

Figure **17**-B) the network might produce an output far from the true response (black line). It models the function perfectly only in the training points, meaning that the ANN does not generalized well (Hagan et al. 2004).

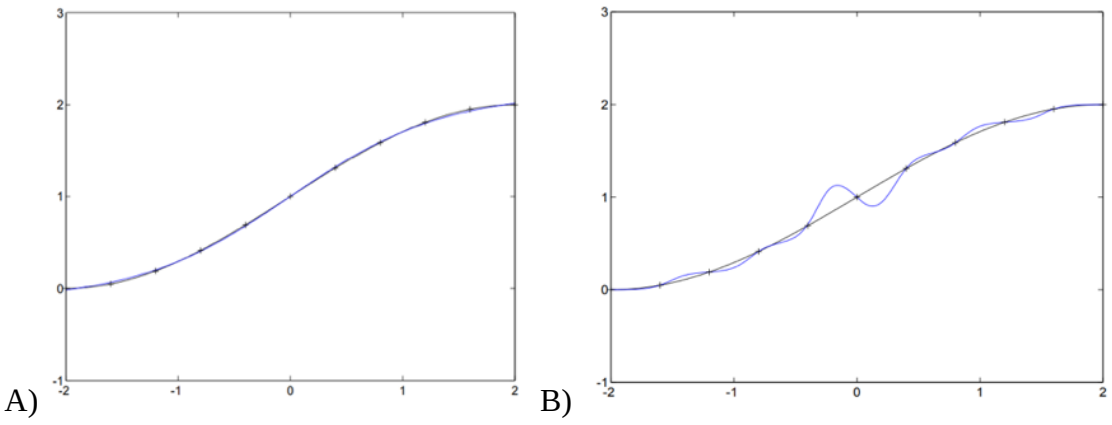

Figure 17: Generalization of ANN: A) 1-2-1 ANN approximation of a function *g*, B) 1-9-1 approximation of a function *g (*Hagen *et al*., 2004)

As seen in the former example, there is no guarantee of better performance by only increasing the neuron number; this is because as the number of neurons increases, the number of free parameters (weights and biases) also increases. A good ANN calibration model must be as accurate and as general as possible; thus, ANN generalization is not a trivial task.

Generalization is mainly performed by different approaches; the most applied techniques are early stopping, regularization, global searches and pruning and growing. Pruning and growing techniques eliminate or add neurons respectively until good generalization is

achieved. Global searches compare all possible network architectures to locate the simplest model that explains the data. In contrast to these techniques that try to keep the network simple in relation to the neuron number, the regularization and early-stopping techniques keep the network simple in relation of the weight magnitudes instead of reducing the neuron number (Hagan et al. 2004).

The concept of early stopping relies on the fact that as the training process progresses, the network uses more and more of its weight capacities, thus by increasing the training iterations the network becomes more complex. If the training is stopped before the minimum error is reached, then the network will effectively use fewer parameters and will be less likely to overfit. Regularization techniques modify the sum squared error performance index equation to include a term that penalizes network complexity, which forced the resulting function to be smooth and then less likely to overfit.

ANN have been used in the NIRS context as a calibration method by several authors. For instance, in latex-ethanol mixtures it has been reported that ANN achieved a reduction of the standard error of prediction between 50 % and 75 % compared to PLSR (Borggaard and Thodberg 1992). They also found that ANN requires less input data in the form of scores of principal component analysis (compared to PCR), providing better stability of ANN models. Moreover, the great advantage of ANN relied on the fact that ANN were able to learn non-linear relationships, which enable performing predictions in extrapolation situations with better results. On the other hand, calibration using UV/Visible spectra showed that ANNs provide results only slightly better than PLS and PCR only (Gemperline et al. 1991) and pointing out the importance of the transfer function, which has a critical role to model non-lineal relationships. This fact agreed with the results of Naes et al. (1993), which demonstrated that ANN based on sigmoid functions gave better predictions. However, it was concluded that the performance of an ANN relative to classic lineal methods such as PCR and PLS is strongly dependent on the situation under consideration.

1.2.3 Synthesis: Near Infrared spectroscopy as a PAT monitoring tool

Monitoring of cell culture processes by *in situ* spectroscopic approaches is an extremely complex task since not only chemical, but also physical and biological data and their dynamic must be considered within spectra, including scattering effects for physical information, confused relationships between correlated chemical variables (compounds concentrations) for chemical information, among other phenomena. Therefore, the use of NIR spectra in a predictive manner require sophisticated chemometric tools for building calibration models so that process analysers can be suitable PAT monitoring tools. Indeed, analysis of cell culture nature suggested that adoption of innovative regression methods, such as SVR, LWR or ANNR, could enhance the performance of monitoring by process analysers. However, calibration models for cell culture process have mostly been restricted to classic regression methods such as PCR and PLSR. This fact could explain why this promising technology has not been as widely adopted by the biopharmaceutical industry as first expected. Therefore, evaluation of new regression methods that enhance the comprehension of the relationship between spectra and cell culture properties, is required for going some way towards the implementation of new quality approaches assuring medicine consistency and patient safety.

1.3 MONITORING OF CELL CULTURES BY PAT & NIR SPECTROSCOPY

As previously described (section 0), cell culture processes are complex and require relentless control for assuring biopharmaceutical properties and patient safety while economically maximizing the benefits of processes. There are three main challenges for such monitoring and control: the high number of possible critical variables (physical, chemical, biological), inherent challenges for their monitoring (instrumentation) and deep process knowledge for control strategies. Although cell cultures are actually monitored in terms of several critical variables (Claßen et al. 2017) as shown in Table 1, they are mostly restricted to physical parameters and processes could be enhanced provided new critical variables be monitored, such as biological variables (Clementschitsch and Bayer 2006).

Table 1. Analytical techniques for monitoring cell cultures (Claßen et al. 2017)

Measurement category	Analytical technique/sensor	Process segment	Measurement mode	Note
Physical variables				
Temperature	Thermometer, thermistor, thermocouple	USP and DSP	In-line	Standard
Pressure	Membrane pressure sensors	USP and DSP	In-line	Standard
Viscosity	Viscometer	USP and DSP	Off-line	Standard
Redox potential	Redox (Pt) electrode	USP and DSP	In-line	Standard
Chemical variables				
pH	pH electrode	USP and DSP	In-line	Standard
Dissolved gases	Amperometric oxygen electrode, CO_2 electrode	Mostly USP Mostly USP	In-line In-line	Standard Standard
Gas phase	Paramagnetic (e.g., O_2), IR (e.g., CO_2)	Mostly USP Mostly USP	In-line In-line	Standard Standard
Volatile compounds	Mass spectroscopy	USP and DSP	Off-line	Standard
Dissolved components	Biosensors (e.g., glucose, lactate) HPLC, GC/MS (e.g., glucose, lactate) Spectroscopic sensors (e.g., glucose, lactate)	USP and DSP USP and DSP USP and DSP	Off-line or at-line At-line In-line	Standard Standard Mostly in research
Biological variables				
Biomass	Microscopy Spectroscopic sensors Impedance sensors	USP and DSP USP and DSP Mostly USP	Off-line In-line In-line	Standard Mostly in research Mostly in research
Turbidity	Spectroscopic sensors	USP and DSP	Off-line	Standard
Cell morphology	Microscopy Fluorescence sensors Impedance sensors In situ microscopy	USP USP USP USP	Off-line In-line In-line In-line	Standard Mostly in research Mostly in research Mostly in research
Viability	Spectroscopic sensors	USP	In-line	Mostly in research
Activities of enzymes	Biosensors	USP and DSP	Off-line or at-line	Standard
DNA/RNA content	Spectroscopic sensors	USP and DSP	In-line	Mostly in research

Critical biological variables are generally monitored by off-line measurements that may not be as fast as necessary to properly control the process. The lack of real-time monitoring constitutes a major limitation to further optimize the performance of these processes and guarantee product quality. Furthermore, sampling implies contamination risk. Consequently, in recent years biopharmaceutical companies have been paying increasing attention to the development of new techniques for online monitoring of bioprocesses. Therefore, regulatory agencies launched the Process Analytical Technology (PAT) initiative for real-time analysis of manufacturing stages through innovative process analyzers. The ultimate aim of the initiative is to gain deep understanding of process fundamentals, which would allow the control of process variability for assuring the Quality Target Product Profile (QTPP). Consequently, during the manufacturing process, Critical Process Parameters (CPP) affecting product Critical Quality Atributes (CQA) must be monitored and controlled, thus maintaining CQA within a particular confidence frame (Jenzsch et al. 2017).

The PAT initiative enforces the use of in-line process analyzers for monitoring and controlling bioprocesses without contamination risks. NIR spectroscopy-based analyzers with remote probes are of great interest since the probe may be placed *in situ* and suitable for steam sterilization. Moreover, they may provide multicomponent information in highly turbid matrices such as cell culture media. The application of this technology implies challenges related to the culture process itself (mixing, cell growth, cell physiology, compounds correlation, scattering, among others.) and technical challenges (instrumentation, development of calibration methods, among others.) (Cervera et al. 2009). Perhaps the main challenge of cell culture monitoring is the high number of possible critical variables (growth regulators, enzymes, substrates, byproducts, oligo-elements, products, cells, cell physiological states, among others). This could be overcome by proper selection of analyzer nature and development of accurate and precise calibration models (Scarff et al. 2006).

Several studies have been performed for the challenge of cell culture monitoring using different modes of operations, cell lines, culture media and calibration strategies as shown in

The great majority have focused on classic parameters such as glucose, lactate, glutamine or cells, while product or even product quality in terms of post-translational modifications, have received limited attention. The major advance of these studies is to achieve correlations within the cell culture process, such as compounds, metabolic rates, among others. They recommend using techniques such as compound spiking, or addition of synthetic calibration samples for breaking such correlations. However, other authors have pointed out that such techniques may increase noise in spectra if such samples are different in term of physical status, such as % DO, pH, temperature and culture mixing dynamics. Moreover, as far as can be ascertained, all these studies have only used linear regression techniques, such as PLSR, for calibration model building. Although there are plenty of regression techniques for building calibration models, only linear regressions such as PLSR have been reported. It is important since information in spectra may also be contained non-linearly as suggested by Li et al. (2018a), particularly for product quality.

Table 2. Cell cultures monitored by NIRS

Analysis mode, cell line, (Reference)	Compounds	Comments
In-line, CHO-k1, (Sandor et al. 2013)	Ala, Glc, Leu, Gln	Mis estimations at the end of culture
In-line, Not reported, (McShane and Coté 1998)	Glc, Lact, Ammo	Found correlation between chemical concentration and time
At line, CHO (Hakemeyer et al. 2012)	Gln, Glu, Glc, Lact, Ammo, VCD, mAb, LDH, Osmolality	Process dynamics must be taken into account for calibration to be built. Mis estimations at the start of culture
In-line, CHO k1, (Arnold et al. 2003)	Gln, Glc, Lact, Ammo	There could not only be correlations between metabolic rates affecting performances of models.
In-line, CHO, (Roychoudhury et al. 2007)	Glc, Lact	Calibration should consider different instrumentation variability (probes, bioreactors) for proper accuracy
In-line, Vero, (Petiot et al. 2010)	Glc, Lact	Accurate models can be achieved in cell cultures containing large scattering compounds such as micro-carriers during adherent Vero cell cultures.
In-line, a mammalian cell line, (Henriques et al. 2009)	Glc, Lact, Ammo	It was important to include the wavelength region of water peak since compound information was contained in the interactions with water molecules
In-line, CHO k1, (Milligan et al. 2014)	Glc	Difficulty in extrapolating predictions outside of model chemical range.
In-line, CHO (Courtès et al. 2016)	LDH activity	NIRS could differentiate the presence of the same protein inside and outside cells to some extent
In-line, CHO, (Clavaud et al. 2013)	Prot, Glc, PCV, VCD, osmolality	Mis predictions at the start and end of cultures
In-line, CHO, (Li et al. 2018a)	mAb, Glc, Lact, Gln, VCD	Information could be non-linearly contained within spectra
Glc: Glucose; Glu: Glutamate; Gln: Glutamine; Lact: Lactate; LDH: Lactate Dehydrogenase; PCV: Packed Cell Volume; VCD: Viable Cell Density		

Measurement of biological variables in plant cell suspension cultures has not received the same attention as for animal cell cultures. As far as can be ascertained, there is no report of NIRS for monitoring plant cell cultures. Indeed, monitoring strategies have been focused on CO_2, O_2, pH and biomass concentration, the latter being the most used. Thus, in-line monitoring of cells has implied correlations of cells with conductivity (Kwok et al. 1992), osmolarity (Madhusudhan et al. 1995) and turbidity (Zhong et al. 1993) of culture media. However, such approaches are likely vulnerable in the QbD-PAT frame since they are very sensitive to cell physiological state. Thus, variations in process operating conditions could compromise monitoring performance in real operating conditions.

More recently, new monitoring approaches based on real plant cell properties, instead of culture media compounds relationship with cells, have been analyzed for monitoring cell concentration. Dielectric properties of cells have been successfully used for performing correlation of permittivity and biomass in terms of packed cell volume (PCV) and cell dry weight (DW) (Markx et al. 1991; Matanguihan et al. 1994; Holland et al. 2013). These approaches are likely to be more robust though cell heterogeneity nature of plant cells has not been yet addressed. Plant cell morphology is dynamic and strong changes in cell volumes and shapes may occur during cultures, causing limited significance of classic monitoring variables such as DW or PCV. Indeed, major deviations of such permittivity-cells relationships are likely to be caused by cell heterogeneity (Holland et al. 2013). On the other hand, fluorescence of some compounds has been proven useful for the establishment of in-line or on-line monitoring strategies. For instance, NAD(P)H fluorescence correlation to biomass has been proposed for on-line monitoring of suspension cultures (Hisiger and Jolicoeur 2008; Srivastava et al. 2008). However, plant cells in suspension cultures are highly heterogenic and current approaches fail in providing information of cell heterogeneity such as somoclonal variation (Deus-Neumann and Zenk 1984) cell differentiation (Torrey 1975) or cell aggregation (Patil et al. 2013), which may limit the implementation of PAT strategies. Monitoring of substrates, products and by-products have mostly been restricted to off-line approaches. Only few studies have reported in-line or on-line monitoring, particularly using the fluorescence nature of some compounds. For instance, the alkaloids ajmalicine and serpentine could be monitored explicitly in suspension cultures of *C. r.*, though as for alkaloids presenting similar fluorescence profile monitoring, monitoring can only be

performed for total alkaloids (Hisiger and Jolicoeur 2008). In recombinant plant suspension cultures, the product DNA sequence could be added into plant DNA with a marker or reporter such as the Green Fluorescent Protein (GFP). It is expected that GFP concentration be directly proportional to product concentration, so GFP could be used for monitoring product titer in cell culture processes. This approach has been proven feasible in on-line modes for tobacco cell cultures also using fluorescent spectroscopy, though limited to secreted GFP, because of phenomena such as culture autofluorescence (Su et al. 2004). Novel calibration strategies considering such phenomena have proven these fluorescent approaches feasible for also monitoring intra-cellular GFP (Su et al. 2005).

1.3.1 Synthesis: Monitoring of cell cultures by PAT & NIRS

The complex nature of biopharmaceuticals requires relentless control of cell culture processes for limiting medicine heterogeneity and thus assuring clinical effects and patient safety. Currently in industry, cell culture processes are monitored in-line through several critical variables, though mostly restricted to physical parameters. Processes could be enhanced, provided control of innovative parameters with biochemical and biological nature be routinely implemented by new in-line process analyzers based on vibrational spectroscopy. However, ongoing developments for using spectra in a predictive manner have been limited to few biochemical parameters such as the concentrations of some substrates (*i.e.* glucose or glutamine) and by-products (*i.e.* lactate) while innovative parameters such as cell physiological state or medicine heterogeneity have received scarse attention. Indeed, it is worth noting that only classic regression methods have been employed for building monitoring procedures with limited results. Therefore, there is a need for the application of new chemometric strategies for better management of spectra so that monitoring procedures are capable of monitoring innovative biochemical and biological parameters within cell culture processes.

2 AIM OF THE THESIS, HYPOTHESIS & OBJECTIVES

Overall, the background pointed out the complexity of cell cultures and the need for new perspectives to assure proper process performances. Therefore, diverse physical, chemical and biological parameters must firstly be monitored and then controlled. In spite of the fact that new quality approaches such as PAT have been encouraged for 15 years with the idea of controlling functional key bioprocess parameters such as cell physiology, kinetics of metabolite products or viable cells as well as quality of recombinant proteins, only a few parameters are systematically monitored in bio-industries, mainly physical parameters such as pH, DO (dissolved oxygen) and temperature, among others. The potential of analyzers based on molecular vibrations, mainly NIR spectroscopy, has recently been demonstrated, although limited to some common classical parameters such as substrate (glucose, glutamine, among others) or by-product (lactate, among others) concentrations, only in some animal cell culture platforms, plant cells have not yet been addressed. Moreover, such monitoring approaches can still be enhanced as for reducing prediction errors.

Development of proper monitoring procedures based on NIRS is a multidisciplinary task requiring not only process engineering approaches for acquiring reliable and precise experimental measurements in large quantities, but also a strong chemometric background based on regression methods. Consequently, the aim of this thesis project has been to enhance the capabilities of *in situ* NIR spectroscopy for properly monitoring of cell cultures. It has been intended to go some way towards enhancing our understanding of the relationship between cell culture nature and NIR spectra by evaluation of diverse chemometric approaches.

Therefore, the main objectives were developed to test the following hypothesis:

1. As far as can be ascertained, *in situ* monitoring based on NIR spectroscopy has not been reported for biopharmaceutic plant suspension cultures – Is the nature of suspension culture suitable for *in situ* monitoring by NIR spectroscopy?
    - ➔ Objectives: To develop a plant suspension culture process capable of antineoplastic molecule production and then to evaluate *in situ* NIR spectroscopy for monitoring.

2. Current monitoring approaches are based on linear calibration methods such as PLSR and PCR – Are these linear methods always adequate whatever the nature of cell culture processes and functional parameter monitored? Could the use of other regression methods enhance the efficiency of calibration?
    - ➔ Objectives: To evaluate the adequacy of current linear regression methods and to explore the use of other regression approaches for handling non-linear relationships between process parameters and spectra, such as LWR, ANNR and SVR.
3. The use of *in situ* monitoring approaches based on NIR spectroscopy has been focused on the monitoring of some classical culture parameters (cells, substrates and by-products) – Could it be also used for monitoring innovative parameters such as mAb glycosylation or cell heterogeneity such as cell differentiation in plant suspension culture?
    - ➔ Objective: To develop and evaluate calibration methods for mAb produced in CHO cell culture processes, with different sugar moieties conferring the biological product clinical properties. To develop and evaluate calibration methods for monitoring cell differentiation in *Catharanthus roseus* cell culture processes, closely related to antileukemic molecule production.

For dealing with all these questions, results of the experimental approach have been organized into four chapters. Chapter I primarily concerns the generation of plant cell culture processes for anticancer production and the identification of the CPP related to this process, which were then used to evaluate their monitoring by NIR spectroscopy at bioreactor scale. Chapter II analyzed the multivariable data processing methodologies commonly used for cell culture monitoring (Partial Least Squares Regression-PLSR, Principal Component Regression-PCR) through classical parameters (concentrations of substrates, products and by-products). In addition, a new methodology approach (Locally Weighted Regression-LWR) is proposed to overcome limitations observed with PLSR and PCR. Chapter III proposes the new use of different nonlinear multivariate data processing methodologies (Supported Vector Regression-SVR, Artificial Neural Network Regression-ANNR) to improve the performance of prediction models for classical parameters. Finally, Chapter IV

takes up the results of the previous two chapters to propose new methodologies for monitoring innovative parameters related to heterogeneity sources within cell culture processes: product heterogeneity in term of mAb glycosylation in CHO cell culture processes, and cell heterogeneity in terms of cell differentiation within cell aggregates in suspension culture processes of *Catharanthus roseus*.

3 MATERIALS AND METHODS

3.1 CHO CELL CULTURES

3.1.1 Cell line

The cell line used was the CHO M250-9 cell line, capable of producing an anti-RhD antibody (mAb). It was developed in 2008 at the Bioprocessing Technology Institute in Singapore (Chusainow et al. 2009). It is the result of the transfection of the deficient in dihydrofolate reductase CHO-DG44 (dhfr -) cell line, with two expression vectors (phCMV-VHRhD-λ1C-neo and phCMV-VLRhD-KR-neo) containing the genes coding for human anti-rhesus D antibody and dihydrofolate reductase (Urlaub et al., 1983). The CHO M250-9 line was then adapted to serum-free medium for suspension culture.

3.1.2 Culture media

The culture medium used was a mixture (1: 1 v/v) of two serum- and protein-free commercial media: CD-CHO Medium (Thermo Fisher) and PF-CHO Medium (GE Healthcare Life Science). CD-CHO medium is a chemically defined medium optimized for CHO cell growth and recombinant protein expression in suspension cultures. It contains neither protein, peptide components from animal, plant or synthetic origin, nor undefined lysates or hydrolysates. PF-CHO medium is a protein-free medium containing soy hydrolysate and has been developed to support the growth of multiple CHO dhfr- cell clones. Their mixture was supplemented with 0.05 % Pluronic F68 (Sigma-Aldrich) and 4 mM L-glutamine (Sigma-Aldrich). In bioreactor cultures, an antimycotic and antibiotic solution (Gibco® Anti-anti, Sigma-Aldrich) was used at 1 % (v/v).

3.1.3 Conservation of cell lines and pre-culture propagation

The cell line was cryo-preserved in a primary and a working bank. Before freezing, exponential growing cells were resuspended in 1 mL culture medium containing 10 % dimethylsulfoxide (DMSO, Sigma-Aldrich) and 10 % (v/v) fetal bovine serum (FBS, Sigma -Aldrich), at a concentration of approximately $1.5x10^7$ cells.mL^{-1}. The cells were then frozen in cryogenic tubes (cryules, Corning) at -80 °C at approximately a 1 °C.min^{-1} cooling rate

(Mr. Frosty ™, Thermo Scientific). After 24 h at -80 °C, the cryogenic tubes were immersed and stored in liquid nitrogen at -196 °C. For starting a culture, a cryogenic tube was taken from liquid nitrogen and immediately thawed to 37 °C. The content of the tube was transferred to 10 mL culture media, then centrifuged (900 rpm, Jouan G 4.11), the supernatant discarded and the pellet suspended into 30 mL culture media and cultured in an incubator set at 37 °C, 5 % CO_2, 80 % relative humidity and 70 rpm. Cells were cultured for 10 days with 4 passages lasting 2 to 3 d each. Cells were always seeded at $3x10^5$ cells.ml^{-1} and cultured for 3 days before being transferred so as to ensure that the cells were in exponential growth phase before experiments.

3.1.4 Culture systems

During cell propagation cultures, single use Erlenmeyer flasks (Fisherbrand™ Shaker Flasks, Fisher Scientific) were used in an incubator (Kühner, Adolf Kühner AG) equipped with an orbital shaking table (5 cm diameter stirring). The incubator was set at 37°C, 5 % CO_2, 80 % relative humidity and 70 rpm. The volumes of the flasks used were 250, 500, or 1000 mL, containing 50, 100, and 300 mL culture medium respectively. For bioreactor cultures, a 2.5 L bioreactor (Tryton, Pierre Guérin, France) was used with a starting working volume of approximately 1.5 L. The dissolved oxygen percentage (% DO) was measured by an optic oxygen sensor (VisiFerm ™ OD, Hamilton) and controlled at 50 % by oxygen supply directly into the culture media through the sparger. Temperature was controlled at 37 ºC by recirculation of water into the bioreactor glass jacket. The pH was measured by an autoclavable sensor (Hamilton) and controlled at 7.2 by addition of either 0.5 N NaOH into the culture media or CO_2 into the space between culture media and the head-plate. Agitation was performed using an Elephant Ear impeller and controlled at 90 rpm

3.1.5 In-line NIR data capture

An *in situ* transflectance probe (Precision Sensing Devices, MA) with 2 mm effective pathlength, was coupled to the bioreactor through a single optic fiber and the probe was connected to an Antaris II NIR spectrophotometer analyzer (Thermo Scientific). A culture medium scan corresponded to an average of 128 scans with an 8 cm^{-1} resolution from 4,000

to 10,000 cm^{-1}. Spectra were automatically collected every 20 min during cultures. Spectra corresponding to the sampling points were selected to generate the calibration models using off-line data.

Spectra were collected using the analyzer software RESULT (Thermo Scientific) and transferred to the R2016a MATLAB® environment (MathWorks Inc.), which permitted the use of the chemometric software PLS-Toolbox 8.2.1 (Eigenvector Research Inc.) for building calibration models. For evaluating the capacity of the models to perform real-time and *in situ* monitoring, the NIR spectra automatically collected every 20 min during cultures were used as input for the calibration models. Off-line and in-line profiles were compared for internal validation.

3.1.6 Off-line analysis of culture compounds

3.1.6.1 Cells

Cells were analysed as viable and not viable using Trypan blue exclusion dye and a Vi-CELL cell counter (Beckman Coulter). Viable cells possess integral cell membranes that exclude the dye, whereas dead cells do not. Therefore, dead cells are detected as blue-colored while viable cells rest uncolored.
The cell counter adds trypan blue to the sample in a 1:1 (v/v) ratio. The sample is put into counting chambers and 50 pictures of cells are taken by a micro photo camera; concentrations of viable and non-viable cells are determined by image analysis.

3.1.6.2 Glucose, lactate, glutamine and ammonium

The Gallery™ Automated Photometric Analyser (Thermo Scientific) was used to measure the concentration of glucose, lactate, glutamine and ammonium, in the culture medium using various enzymatic kits. Samples were centrifuged (100 rcf, 5min) and the supernatant was then put in vials for analysis. Concentrations were determined using enzymatic kits from Thermo Scientific for glucose (981780) and lactic acid (984308), while enzymatic kits from Roche were used for ammonium (06343775001), and glutamine (073956655001) concentration determination.

3.1.6.3 Monoclonal antibody

Total mAb concentration determination was performed using an enzymatic kit from Roche (06681743001) in the Gallery™ Automated Photometric Analyser. and analysis of mAb glycosylation was performed at the end of the cultures. The procedure firstly consisted of purification and digestion of the mAb within the sample from the bioreactor (approx. 1.8 mL) for further analytical analysis. Purification consisted of mAb precipitation by cold acetone (-20 °C, 2 h) and then centrifugation (5 min, 13000 rcf, 4 °C). The supernatant was discarded, and the pellet re-suspended in 0.5 mL ammonium bicarbonate buffer solution (50 mM, pH 8). Protein denaturalization was carried out by heating the samples for 15 min at 95°C and the vials with sample were immediately put into ice to avoid re-naturalization. As the glycan chains represented approximately only the 2 % of the total mass of mAb, mAb were digested by a trypsin (T8003, Sigma-Aldrich) solution (1 gL^{-1} in 1mM HCl solution) into several glycol-peptides of smaller size, which were already indexed in databases (Kapur et al., 2014). Digestion consisted of the addition of 20 µL trypsin (Sigma T8003) solution (1 gL^{-1} in 1 mM HCl) into the sample contained in the buffer. The vials were incubated (37 °C, 12 h, orbital agitation 130 rpm) and then digestion was arrested by addition of 15 uL formic acid. The vials were then centrifuged (10000 rcf, 25 °C, 10 min) and the supernatant filtered for further analysis. Analysis of glycol-peptides were performed using a HPLC-MS (Thermo Fisher) equipment, consisting of a HPLC coupled to a photodiode detector and a Linear Trap Quadripole (LTQ) mass spectrophotometer in positive electrospray ionization mode (ESI+). The separation column was a C18 column (150mm x 2.1 mm) (Grace/ Alltech), 200 $\mu L.min^{-1}$ mobile phase flow rate. An elution gradient of phase A (0.1 % v/v trifluoracetic acid/water) and B (0.1 % v/v trifluoracetic acid/acetonitrile) was used: 4% B phase for 5 min, 4% to 15% B phase transition for 25 min, and 15 % to 98 % B transition for 5 min. Mass spectrometry conditions were as follows: 5 kV sputtering electric tension; flow rates of sheath gas, auxiliary gas and flushing gas set respectively at 40, 10 and 10 in $UA.min^{-1}$; capillary temperature at 300 °C; electric tension at 36 V, 80 V, -44 V and 3.25 V for capillary, tube lens, bottom lens and front lens respectively. The raw data from the analysis was processed using the XCALIBUR (Thermo-Fisher) software. The fraction of each glycol-peptide with a particular glycosylated chain was obtained dividing its area by the area of the total IgG signal

peak. The concentration of each glycol-peptide could thus be obtained semi-quantitatively by multiplying its fraction by the concentration of total mAb analysed by the Gallery Analyser.

3.2 PLANT CELL CULTURES

3.2.1 Cell line

The cell lines used were generated from plant material at the Francisco Javier Clavijero Botanical Gardens (Instituto de Ecología, A.C., Xalapa, Ver. Mexico). Then calluses were transferred to the Plant Cell Lines Bank of the Veracruz Institute of Technology and maintained under four different photoperiods. Calluses maintained in a 16 hour per day (hd^{-1}) photoperiod were used. Selected calluses with similar color and friability from stem, leaf and apical meristem tissues were aseptically transferred to 250 mL Erlenmeyer flasks containing 100 mL MS liquid medium for establishment of suspension cultures.

3.2.2 Culture media

The culture media used was Murashige and Skoog (MS) medium (Caisson) prepared from powder, supplemented with 2 mgL^{-1} glycine and 100 mgL^{-1} myo-inositol (Sigma-Aldrich). Media was supplemented with 30 gL^{-1} sucrose or glucose, 4.52 μM 2, 4-D, and 4.44 μM BAP. The pH was adjusted to 5.7 ± 0.1 with 0.5 N NaOH and 0.5 N HCl and the medium was autoclaved at 120 °C for 15 min. For bioreactor cultures, addition of Plant Preservative Mixture (Plant cell Technology) at 1 mLL^{-1} an antibiotic and antifungal solution was also used.

3.2.3 Conservation of cell lines and pre-culture propagation

Cell lines were conserved as suspension and callus cultures. Cells in suspension were subcultured every fifteenth day (10 mL suspension culture into 40 mL liquid media in 125 mL flasks). Stock calluses were maintained on media solidified with 0.9 % (w/v) agar and subcultured monthly. The cultures were kept in a 16 hour per day (hd^{-1}) photoperiod.

Inoculum preparation for experiments at flasks level required biomass concentration from several flasks subcultivated at least four times from callus. Biomass centrifuge-concentrated (5 min, 250 rcf, 15 ºC) pellets were re-suspended in fresh medium, and volumes were taken to inoculate flasks for the various experiments. For bioreactor cultures, suspension cultures were generated from 5 g fresh calluses with similar friability, color and age generated only

from leaf tissues and placed into 20 mL liquid MS media. Cultures were then incubated for six days in 125 mL ErlenMeyer flasks in a 16 $h.d^{-1}$ photoperiod (30 $\mu mol.m^{-2}$ s^{-1} photonic flux), 25 °C and shaken at 100 rpm on an orbital shaker. After the first culture, suspension culture was filtered using a sterile stainless steel mesh (2.25 mm^2 pore size) to retain callus clusters. Then approximately 20 mL culture volumes without the biggest clusters were used to equally inoculate two flasks containing 10 mL culture medium and both flasks were cultured under the same conditions. Two subsequent subcultures were performed for biomass propagation. Sixteen flasks containing a total of 320 mL were concentrated by repeated decantation until cells were concentrated in approximately 100 mL, which was used to inoculate bioreactors cultures. For both cases, inoculum age was always six days old.

3.2.4 Culture systems

Flask cultures were performed in 50, 125, 250 and 500 mL Erlenmeyer glass flasks containing 10, 50, 100 and 250 mL MS media respectively. While for the 125 mL flasks silicone sponge closures were used, aluminum foil closures were used for all the other flask sizes. Flasks were in orbital agitation (90 rpm) with a 5 cm orbital diameter. Flasks were maintained at room temperature, controlled at 24 to 26 °C.

Cell cultures were performed in 3 L benchtop bioreactors (Applikon, detalles) with a 2 L working volume. Agitation of suspension culture was performed by a spin-filter coupled to a marine impeller. The set points for all cultures were 30 °C and 90 rpm stirring. Aeration of cultures was by air flux at 200 $mL.min^{-1}$. DO and pH (5.6) were maintained at 50 % saturation and 5.6 respectively (Applikon Bio controller ADI 1010). Bioreactors were illuminated by a LED panel below the jar in a 16 $h.d^{-1}$ photoperiod (50 $\mu mol.m^{-2}$ s^{-1} photonic flux). For bioreactor cultures, culture media were filter-sterilized using a bottle top filter (0.20 µm, Thermo Scientific™ Nalgene™) and then tipped out into the autoclaved bioreactor.

3.2.5 In-line NIR data capture

An *in situ* transflectance probe (Precision Sensing Devices, Inc., Medfield, MA) with 6 mm effective pathlength was coupled to the bioreactor through a microbundle optic fiber containing 80 single fibers (40 fibers per radiation source and 40 fibers per signal receiving).

The probe was connected to a XDS Process Analytics near infrared spectrophotometer analyzer (Foss NIR Systems, Silver Spring, USA). A culture medium scan corresponding to an average of 64 scans with a 0.5 nm resolution from 800 nm to 2,200 nm was performed. Spectra were automatically collected every 30 min during cultures and those corresponding to the sampling points were selected to generate the calibration models using the off-line data. Spectra were collected using the analyzer software VISION (FOSS NIRSystems, V. 3.5,) and transferred to the R2016a MATLAB® environment (MathWorks Inc.), which permitted the use of the PLS-Toolbox 8.2.1 chemometric software (Eigenvector Research Inc.) for building the calibration models. For evaluating the capacity of the models to perform real-time and *in situ* monitoring, NIR spectra automatically collected every 30 min during cultures were used as input for the calibration models. Off-line and in-line profiles were compared for internal validation.

3.2.6 Off-line analysis of culture compounds

3.2.6.1 Biomass dry weight

In a laminar flow cabinet, 1.0 mL suspension culture from a thoroughly stirred medium was sampled and put in an already weighed 1.5 mL micro-centrifuge tube. The tube was centrifuged (10,000 rcf, 10 min; Eppendorf 5424), its content decanted and then 1.0 mL deionised water added. Centrifugation and decantation were repeated, then the washed pellet was vacuum-dried (65°C, ShellLab mod. 1410) until constant weight was reached, and then the pellet was weighed (Mettler H80) and analysed for calculating biomass dry weight concentration.

3.2.6.2 Differentiated cells concentration

Cell differentiation analysis of cells in suspension cultures was undertaken based on cell wall differences; parenchyma, collenchyma and sclerenchyma cells were counted. A 1 mL sample was put into a 1.5 mL micro-centrifuge tube and centrifuged (250 rcf, 15 min); 900 µL of the supernatant were put in a tube and 900 µL digestion enzyme (TrypLE™ Express Enzyme (1X), Thermo Fisher Scientific) were added to the pellet. The tube was agitated with a vortex (MX-S, Science MED) for 45 min. It was centrifuged under the same conditions, and the

pellet isolated in 300 µL by discarding the supernatant, resulting in a digested cell suspension suitable for analysis. Cells in digested cell suspensions were analysed and counted twice in Neubauer improved chamber, each count comprised between 105 and 180 single cells.

3.2.6.3 Carbohydrates concentration

Carbohydrate concentration in cell culture samples was determined as glucose and fructose concentration by HPLC (Alliance Waters 2695; Shodex 1110 column: H_2SO_4 0.5 M, 0.6 $mLmin^{-1}$, 50 °C; refractive index detector: Waters 2414, 55 °C, 10 µL volume injection).

Samples consisting of 500 µL cell-free supernatant were hydrolysed by adding 500 µL 4 M HCl solution and incubated at 60 ºC for 20 min. When the sample reached room temperature, 125 µL BaO 0.3 M and 125 µL $ZnSO_4$ 5 % p/v solutions were added. Then, after the sample was centrifuged (10000 rcf, 10 min), the supernatant was filtered (Polypropylene Pall GHP acrodisc 13, 0.45 µm) and transferred to an HPLC vial for analysis. Samples were compared against external standard curves for glucose and fructose.

3.2.6.4 Vincristine and vinblastine concentration

Anticancer compounds in cell cultures were analysed as VC and VB concentrations by HPLC analysis (Waters 600 HPLC System (Milford, Mass.U.S.A.); Chromolith® Performance RP-18 endcapped 100-4.6 HPLC column; mobile phase acetonitrile–0.1M/ phosphate buffer containing 0.5 % glacial acetic acid, pH 3.5 (21/79 v/v, pH 3.5); 1.2 $mL.min^{-1}$ flow rate; detector Waters 2487 detector (λ 205 nm).

Volume samples (20 – 30 mL) were taken from flasks and freeze-dried ($90x10^{-3}$ mbar, -43 ºC, 24 h) in 50 mL glass vials. Then approximately 100 mg lyophilised sample was taken and solubilised in a 25 mL glass tube containing 1 mL methanol. The content of the tubes was mixed and ultrasonically homogenised for 1 h at 56 °C (Cole-Parmer, cv33). Alkaloids in methanol were separated from insolubilized compounds by centrifugation (12,000 rcf, 10 min) and the supernatant was transferred to an HPLC vial for analysis. Samples were compared against external standard curves for VC and VB.

4 RESULTS AND DISCUSSION

4.1 CHAPTER I: PLANT SUSPENSION CULTURES FOR ANTILEUKEMIC AGENT PRODUCTION

4.1.1 Introduction

Plant-derived biopharmaceuticals are essential for treating several diseases. However, they are structurally complex which limits total chemical synthesis. On the other hand, their natural biosynthesis is extremely complex requiring several enzymatic steps, which limits their bioavailability in plants. During the last decades, efforts have been made to enhance bioavailability by the use of cell culture technology, particularly suspension cultures which are feasible for scaling up.

For instance, the industrial production of the antineoplastic molecule paclitaxel, which had been classically produced from plant material extraction are now moving toward plant suspension systems of *Taxus* species. The former example encouraged the interest of other plant-derived biopharmaceutics for their production using cell culture technology. Bioactive molecules from *Catharanthus roseus*, Madagascar periwinkle, are of great interest in this case since they exhibit important clinical effects, such as the anticancer molecules vincristine and vinblastine, or the antihypertensives ajmalicine and serpentine. However, such molecules have been extremely difficult to express in suspension cultures, particularly vincristine and vinblastine, which are *in vivo* synthetized through different differentiated cells within different tissues. Therefore, the production of a commercial suspension culture for *Catharanthus* alkaloids has not yet been set in motion. The objective of this chapter has been to analyze *Catharanthus* suspension cultures, which may contribute to the development of new processes based on differentiated cell suspension cultures capable of producing the anticancer molecules vincristine and vinblastine. Results have pointed out the suitability of such new preliminary processes, the basis for analyzing the suitability of *in situ* NIR spectroscopy monitoring on a bioreactor scale (Chapter IV 4.4.1.2). The great majority of results from this chapter are intended for publication and thus are accordingly organized for possible publication in Biotechnology Letters journal.

4.1.1.1.1 Interest of cellular differentiation in the production of vincristine and vinblastine in suspension cultures of *Catharanthus roseus* (L.) G Don.

Zavala-Ortiz Daniel Arturo[a,b], Martínez-Montero Mauro Josué[a], Guedon Emmanuel[b], Marc Annie[b], Ebel Bruno[b], Barradas-Dermitz Dulce María.[a], Hayward-Jones Patricia Margaret[a], Mata-Rosas Martín[c], Aguilar-Uscanga María Guadalupe[a]

[a] Tecnológico Nacional de México/Instituto Tecnológico de Veracruz: Calz. M.A. de Quevedo 2779, Veracruz, Ver. México.
[b] Laboratoire Réactions et Génie des Procédés, UMR 7274, CNRS-Université de Lorraine, 2 avenue Forêt de Haye, TSA 40602, 54518 Vandœuvre-lès-Nancy, France.
[c] Instituto de Ecología A.C.: Carretera antigua a Coatepec 351, Xalapa, México.

4.1.1.1.1.1 Abstract

Objective: Cell differentiation is needed for the *in vivo* synthesis of vincristine and vinblastine and thus *in vitro* cultures are usually considered as non-producing platforms for these antineoplastic molecules. However, several studies have recently detected these molecules in early differentiated calluses and also in suspension cultures. The degree of cell differentiation, nevertheless, has not been addressed for their production, particularly in suspension cultures which could be used as a large-scale producing platform. Therefore, the effect of different culture conditions on the production of vincristine and vinblastine, taking into account cytodifferentiation within cell aggregates, has been analysed for the first time. *Results:* Culture conditions such as light exposure and plant growth regulator regimes have been shown to affect cell differentiation. Moreover, this cell differentiation into collenchyma- and sclerenchyma-like cells was observed to be related to vincristine and vinblastine titers. *Conclusions*: Results demonstrated the use of cell differentiation for the establishment of advanced processes of differentiated cell suspension cultures for producing plant-derived biologicals such as vincristine and vinblastine.

4.1.1.1.1.2 Introduction

Catharanthus roseus (L.) G. Don, Madagascar periwinkle, has been used in traditional medicine against several diseases such as malaria and diabetes. It is one of the most

extensively investigated medicinal plants because of its capacity to produce high economic value phytochemicals such as the antineoplastic molecules vincristine (VC) and vinblastine (VB) and the antihypertensive ajmalicine (Tikhomiroff and Jolicoeur 2002). Due to the complexity and scarcity of VC and VB within plants, production is carried out from the coupling of the precursors catharanthine and vindoline present in higher amounts in plant material (Alam et al. 2017). Nowadays production is mainly carried out by extraction from plant material grown in the United States, Spain, China, Africa, Australia, India and Southern Europe (Barkat et al. 2017). However, expensive extraction procedures are required with very low recovery. Thus, intensive efforts have been invested for increasing availability such as production via *in vitro* cultures, particularly suspension cultures feasible for scale-up. Nonetheless, no commercial production using *C. roseus* suspension culture has been set in motion. Production of VC and VB in plants requires intracellular and intercellular translocation of pathway intermediates within differentiated cells in leaves (St-Pierre et al. 1999; Murata and Luca 2005). This complex synthesis partially explains why attempts for producing VC and VB by cell culture technology have failed, particularly in undifferentiated cell suspension cultures (Verpoorte et al. 1993).

This former conception has been challenged by several studies demonstrating VC and VB production capability by *in vitro* cultures, particularly of calluses with early differentiation into roots or shoots (Miura et al. 1987; Kalidass et al. 2010; Ataei-Azimi et al. 2018). These reports have claimed that production capability is closely related to cell differentiation, though no further details are provided. Moreover, suspension cultures with VC and VB production capability have also been reported (Taha et al. 2014; Zhang et al. 2015), although no inferential about this capacity was provided.

The conception that *in vitro* cultures are not capable for VC and VB production mainly relies on the fact that *in vivo* synthesis of the precursor vindoline has been localized in chloroplasts (De Luca and Cutler 1987) and thus *in vitro* heterotrophic cultures lacking functional chloroplasts could not synthetize the precursor. Autotrophic suspension cultures have been developed seeking production in functional chloroplasts but neither vindoline, VC nor VB was detected (Tyler et al. 1986). On the other hand, there have been reports of vindoline producing cultures under heterotrophic conditions (Scott et al. 1980; Naaranlahti et al. 1989). Indeed, it was recently discovered that suspension cultures of cambial meristem cells

contained complete sets of enzymes that are responsible for the production of VC and VB (Zhang et al. 2015). Thus, there must be alternative metabolic pathways in *in vitr*o heterotrophic cultures which could be activated by cell differentiation. Therefore, this study sought to analyze cell differentiation in suspension cultures and its eventual relationship with VC and VB production capability. A broad spectrum of culture conditions based on plant growth regulators (PGR) regimens and photoperiods was tested for inducing different cell differentiation and production capabilities. This study was exploratory and interpretative in nature, seeking to be the starting point for further research which could help to understand the nature of VC and VB *in vitro* synthesis for novel differentiated cell suspension culture production processes.

4.1.1.1.1.3 Materials and methods

Generation of suspension cultures

Callus lines were generated from a mature *C. roseus* plant at the Francisco Javier Clavijero Botanical Gardens (Instituto de Ecología, A.C., Xalapa, Ver. Mexico). Plant material passed through a sterilization procedure and then explants from stems, leaves, apical, intercalary and lateral meristem tissues were placed on MS (Murashige and Skoog) medium with 30 $g.L^{-1}$ sucrose, 6.5 gL^{-1} agar (Caisson, Micropropagation Powder Type I), 4.52 μM 2, 4-Dichlorophenoxyacetic acid (2, 4-D) and 4.44 μM 6-Benzylaminopurine (BAP) in darkness. Subsequently 374 calluses were induced, multiplied and used for further experiments. Calluses were transferred to the Plant Cell Lines Bank of the Veracruz Institute of Technology and maintained under four different photoperiods. Calluses maintained in a 16 hour per day (hd^{-1}) photoperiod (photonic flux of 30 $\mu mol.m^{-2}\ s^{-1}$) were used. Selected calluses with similar color and friability from stem, leaf and apical meristem tissues were aseptically transferred to 250 mL Erlenmeyer flasks containing 100 mL MS liquid medium supplemented with 2 mg L^{-1} glycine, 100 mg L^{-1} myo-inositol (Sigma-Aldrich), 30 g L^{-1} sucrose, 4.52 μM 2, 4-D, and 4.44 μM BAP (proliferation medium). The pH was adjusted to 5.7 ± 0.1 with 0.5 N NaOH and 0.5 N HCl and the medium was autoclaved at 120°C for 15 min. For subsequent subculturing in fresh medium, a volume of 10 mL of the original culture

was inoculated into a 125 mL Erlenmeyer flask containing 40 mL proliferation medium every 15 days. Cell aggregation was reduced with each subculture, negligible at the 4th subculture since the majority of aggregates freely passed through a needle (1.5 mm inner diameter), and therefore only suspension cultures with aggregates smaller than 1.5 mm were used for further experiments. For subsequent experiments, only leaf cell line was selected for enhanced capabilities (data not shown).

Once suspension culture was established, inocula for further experiments were always six-day-old biomass suspension cultures that were centrifuge-concentrated (5 min, 250 rcf, 15 ºC). Pellets were then re-suspended in agitated (magnetic stirrer, 90 rpm) fresh medium (≈ 50 mL) until a cell dry weight of approximately 3 gL^{-1} was achieved, then aliquots were taken to inoculate flasks for the various experiments. Initial biomass dry weight concentration for flask experiments was 0.06 gL^{-1} ± 0.01.

Effect of plant growth regulators and photoperiods on leaf cell line

Several attempts have been made to evaluate the effect of PGR on vincristine and vinblastine production with different results in plants (Srivastava and Srivastava 2007; Pan et al. 2010). Studies using callus cultures have underlined the importance of PGR mixtures (Kalidass et al. 2010; Mekky et al. 2018), and so this work focused on evaluating combinations of common auxins and citokines used for suspension cultures and also their effect on different photoperiods. Suspension cultures were initiated from calluses as previously described, then propagated during 12 subcultures in proliferation medium. For evaluating the effect of culture conditions, culture medium was based on basal proliferation medium with specific PGR composition. Firstly, the effect of three single PGR was studied: 2, 4-D (4.5 µM), 1-Naphthaleneacetic acid (NAA) (4.5 µM) and Gibberellic acid (GA_3) (4.5 µM). Secondly, mixtures of 2, 4-D (4.5 µM) with citokines (Kinetine (Kin) and BAP) were tested: 2, 4-D + BAP (2.25 µM); 2, 4-D + BAP (4.5 µM); 2, 4-D + Kin (2.25 µM) and 2, 4-D + Kin (4.5 µM). One treatment without any exogenous PGR was included as a control, thus eight PGR regimens were tested.

The same inoculum, a previously 12-fold subcultivated suspension culture, was used for all tested treatments to reduce the effect of common plant cell heterogeneity. Flasks with silicone

sponge closure (Chemglass, CLS-1490-038) were incubated for 25 days at room temperature with orbital agitation (0.22 rcf) in a rotary shaker (New Brunswick G70; 7 μmol.m^{-2} s^{-1} photonic flux). Four different photoperiods: 0 h (total darkness), 8 hd^{-1}, 16 hd^{-1} and 24 hd^{-1} (continuous light exposure) were used. As four different photoperiods and eight different PGR regimens were tested, the impact of 32 operating conditions on alkaloid production and cell differentiation was evaluated.

Production of VC and VB by differentiated cell culture in a bioreactor

Two bioreactor batch cultures were used to analyze the dynamics of cell differentiation and VC and VB production in suspension cultures. Differentiated cell subpopulations were measured during the whole culture, as also were the concentrations of VC and VB in culture media. Cultures were carried out in a 3 L benchtop bioreactor (Applikon, the Netherlands) with a 2 L working volume using marine impeller performed agitation (90 rpm; 30 °C). For the first culture, dissolved oxygen (DO) was controlled at 50 % air saturation, while for the second culture it was measured but not controlled. The bioreactor jars were illuminated in a 16 hd^{-1} photoperiod (50 μmol.m^{-2} s^{-1} photonic flux). A leaf cell line already subcultured four times during a 6-day period was used as inoculum. The culture media was MS medium supplemented with 2, 4-D (4.5 μM) + BAP (4.5 μM) for both bioreactor cultures.
For the generation of inoculums, calluses with similar friability, color and age were used to generate suspension cultures. Approximately 5 g fresh callus was placed into 20 mL liquid MS media supplemented with 2 mg.L^{-1} glycine, 100 mg.L^{-1} myo-inositol, 30 g.L^{-1} sucrose, 2, 4-Dichlorophenoxyacetic acid (2, 4-D), 4.44 μM 6-Benzylaminopurine (BAP). The pH was adjusted to 5.7 ± 0.1 with 0.5 N NaOH and 0.5 N HCl and the medium was autoclaved at 121°C for 15 min. Cultures were then incubated during six days in 125 mL Erlenmeyer flasks in a 16 h.d^{-1} photoperiod (30 μmol.m^{-2} s^{-1} photonic flux) at 25°C and shaken at 100 rpm on an orbital shaker. After the first culture, the suspension culture was filtered using a sterile stainless-steel mesh (2.25 mm^2 pore size) to retain callus clusters., then approximately 20 mL culture volumes were used to equally inoculate two flasks containing 10 mL culture media, which were subsequently cultured under the same conditions. Two further subcultures were performed for biomass propagation. Sixteen flasks containing a total of 320 mL were

then concentrated by repeated cell decantation until an approximately 100 mL concentration was achieved which was used to inoculate bioreactor cultures. This procedure was repeated for every bioreactor inoculation.

Determination of biomass concentration and growth

Biomass growth was measured by gravimetric method (changes in dry weight). In a laminar flow cabinet, 1.0 mL suspension culture from a thoroughly stirred medium was sampled and put in an already weighed 1.5 mL micro-centrifuge tube. The tube was centrifuged (10,000 rcf, 10 min; Eppendorf 5424), its content decanted and then 1.0 mL deionised water added. Centrifugation and decantation were repeated and the washed pellet vacuum-dried (65°C, ShellLab mod. 1410) until constant weight was reached, and then weighed (Mettler H80).

Cells differentiation analysis

Cell aggregates within cultures were histologically analysed by optical microscopy (Motic, MO-567, USA). Cell differentiation analysis of cells in suspension cultures was undertaken based on cell wall differences (Mauseth 2014). Cells with thin primary walls were defined as parenchyma cells, cells with thickened primary walls as collenchyma cells and cells with primary walls plus secondary walls as sclerenchyma cells and all cells counted.
A sample of 1 mL was put into a 1.5 mL micro-centrifuge tube and centrifuged (250 rcf, 15 min); 900 µL of the supernatant was put in a tube and 900 µL digestion enzyme (TrypLE™ Express Enzyme (1X), Thermo Fisher Scientific) was added to the pellet. The tube was agitated in a vortex (MX-S, Science MED) for 45 min. It was centrifuged with the same conditions, and the pellet was isolated in 300 µL by discarding the supernatant, resulting in a digested cell suspension suitable for analysis. Cells in digested cell suspensions were analysed and counted twice in a Neubauer improved chamber (Blaubrand), each count comprising between 105 and 180 single cells. Cell differentiation analysis was performed at the end of cultures in flask experiments and periodically in bioreactor cultures.

Determination of alkaloids, vincristine and vinblastine

For identification of alkaloids in cells, 100 µL digested cell suspension was mixed with 100 µL Wagner's reagent and incubated for 5 min. Then cell observation was carried out in a Neubauer improved chamber. Alkaloids within cells were detected by their reddish-brown coloration. (Motic, MO-567, USA).

Quantification of VC and VB in culture media was performed by an adapted and validated HPLC procedure (Gupta et al. 2005; Iskandar and Iriawati 2016), using a Waters 600 HPLC System (Milford, Mass., U.S.A.): Chromolith® Performance RP-18 end-capped 100-4.6 HPLC column; mobile phase acetonitrile–0.1M/ phosphate buffer containing 0.5 % glacial acetic acid, pH 3.5 (21/79 v/v), 1.2 $mLmin^{-1}$ flow rate; Waters 2487 detector, λ 205 nm. Briefly, volume samples (20 – 30 mL) of culture media with cells were taken and freeze-dried ($90x10^{-3}$ mbar, -43 ºC, 24 h) in 50 mL glass vials. Then approximately 100 mg lyophilised sample was taken for extraction after which alkaloids were separated from insolubilized compounds by centrifugation (12000 rcf, 10 min). The supernatant was transferred to an HPLC vial for analysis and samples were compared against external standard curves for VC and VB.

Statistics

Data were analyzed with a one-way ANOVA and post hoc Tukey comparison tests ($p \leq 0.05$) using MATLAB environment (R2016a version; MathWorks Inc.).

4.1.1.1.1.4 Results

Effect of culture conditions in flasks

Changing culture media or operating conditions will always show a change in culture performance, though it usually takes 5-10 subcultures for final effects. Culture conditions affected cell production capacity in the first subculture as shown in Table 4.1-1, and detection of VC and VB varied differently depending on culture conditions. Light exposure had a

strong effect since neither VC nor VB was detected in dark conditions independently of PGR regimen. Detection of VB only occurred when using an 8 hd^{-1} photoperiod.

Evaluation of the effect of light exposure (photoperiod) in each individual PGR regimen revealed that there was no statistical difference (One-way ANOVA at 95 % confidence level) in VB titers either for cultures with NAA, 2,4-D as sole PGR, those including BAP or control culture (without exogenous PGR). Thus, the culture conditions that effectively induced VB production were those including GA_3 and mixtures of 2,4-D and Kin. As for VC titers, there was no statistical difference (One-way ANOVA at 95 % confidence level) between photoperiods for treatments including NAA, GA_3, or control culture. Thus, the culture conditions that effectively induced VC production were those that contained 2,4-D.

Table 4.1-1. Effect of PGR on suspension cultures

			Exogenous plant growth regulator regimens							
			No PGR	2,4-D (4.5 µM)	NAA (4.5 µM)	GA_3 (4.5 µM)	2,4-D (4.5 µM) BAP (2.25 µM)	2,4-D (4.5 µM) BAP (4.5 µM)	2,4-D (4.5 µM) Kin (2.25 µM)	2,4-D (4.5 µM) Kin (4.5 µM)
Light exposure photoperiod (hd^{-1})	0	VC	0_a	0_a	0_a	0_a	0_a	0_a	0_a	0_a
		VB	0_A	0_A	0_A	0_A	0_A	0_A	0_A	0_A
	8	VC	1_a	19_b	0_a	3_a	2_a	1_a	3_a	2_a
		VB	1_A	1_A	0_A	21_B	0_A	0_A	17_B	13_B
	16	VC	0_a	34_d	0_a	0_a	23_b	27_b	25_b	30_b
		VB	0_A	0_A	0_A	0_A	0_A	0_A	0_A	0_A
	24	VC	0_a	25_c	0_a	0_a	0_a	38_c	35_c	37_c
		VB	0_A	0_A	0_A	0_A	0_A	0_A	0_A	0_A

2, 4-D: 2, 4-Dichlorophenoxyacetic acid, NAA: 1-naphthaleneacetic acid, BAP: 6-benzylaminopurine, Kin: Kinetin. VC: Vincristine titer (mg/L) in suspension culture, VB: vinblastine titer (mg/L) in suspension culture, VC and VB analyzed at 25 d; superscript letters indicating statistical difference between groups due to the effect of light exposure in a single PGR regimen, lowercase and uppercase letters for VC and VB respectively.

Cell differentiation and its relationship with VC and VB production in flasks

As VC and VB were detected, a possible relationship between *C. roseus* cell differentiation in suspension cultures and VC and VB titers was then investigated, after first classifying differentiated cells, based on cell wall differences, as parenchyma, collenchyma and sclerenchyma cells. Cell aggregates were observed, and their compositions analyzed seeking differentiated cells. During primary analysis, it was discovered that *C. roseus* aggregates contained several cell subpopulations in cell aggregates that achieved sizes up to 300 µm

Figure 4.1-1-a), although smaller sizes were more common (Figure 4.1-1-b). The three main types of differentiated cells were present in cell aggregates. Examples of parenchyma and collenchyma cells are shown in Figure 4.1-1-c, which also shows an example of cell differentiation from parenchyma to collenchyma status. Cell differentiation process to sclerenchyma status is also shown in Figure 4.1-1-d. The most common differentiated cell type was collenchyma, which usually comprised the majority of registered cells. Parenchyma cells were observed in three main subtypes: with nuclei and visible or invisible organelles, and chlorenchyma cells to a lesser degree. Sclerenchyma cells were observed in the form of tracheary elements mainly and sclereid cells to a lesser degree.

The presence of differentiated cells within cell aggregates may imply that there could be some kind of early specialization or at least cells with different metabolic characteristics which could be related to enhanced alkaloid production. Therefore, Wagner's reagent was used for staining cells and revealing the presence of alkaloids within cells. Parenchyma cells were negative for the stain (Figure 4.1-1-e2) while sclerenchyma cells in the form of tracheary elements (Figure 4.1-1-e1) and sclereid-like cells (Figure 4.1-1-f, white arrows) were colored reddish-dark brown indicating the presence of alkaloids, particularly in their massive cell walls. In contrast, deposition of alkaloids was limited for collenchyma cells (Figure 4.1-1-f, dark arrows). Summarizing, alkaloids were mainly detected in sclerenchyma cells, although there were a few sclerenchyma cells in the form of tracheary elements that were negative for Wagner's reagent (Figure 4.1-1-g, triangle), though negligible in extension.

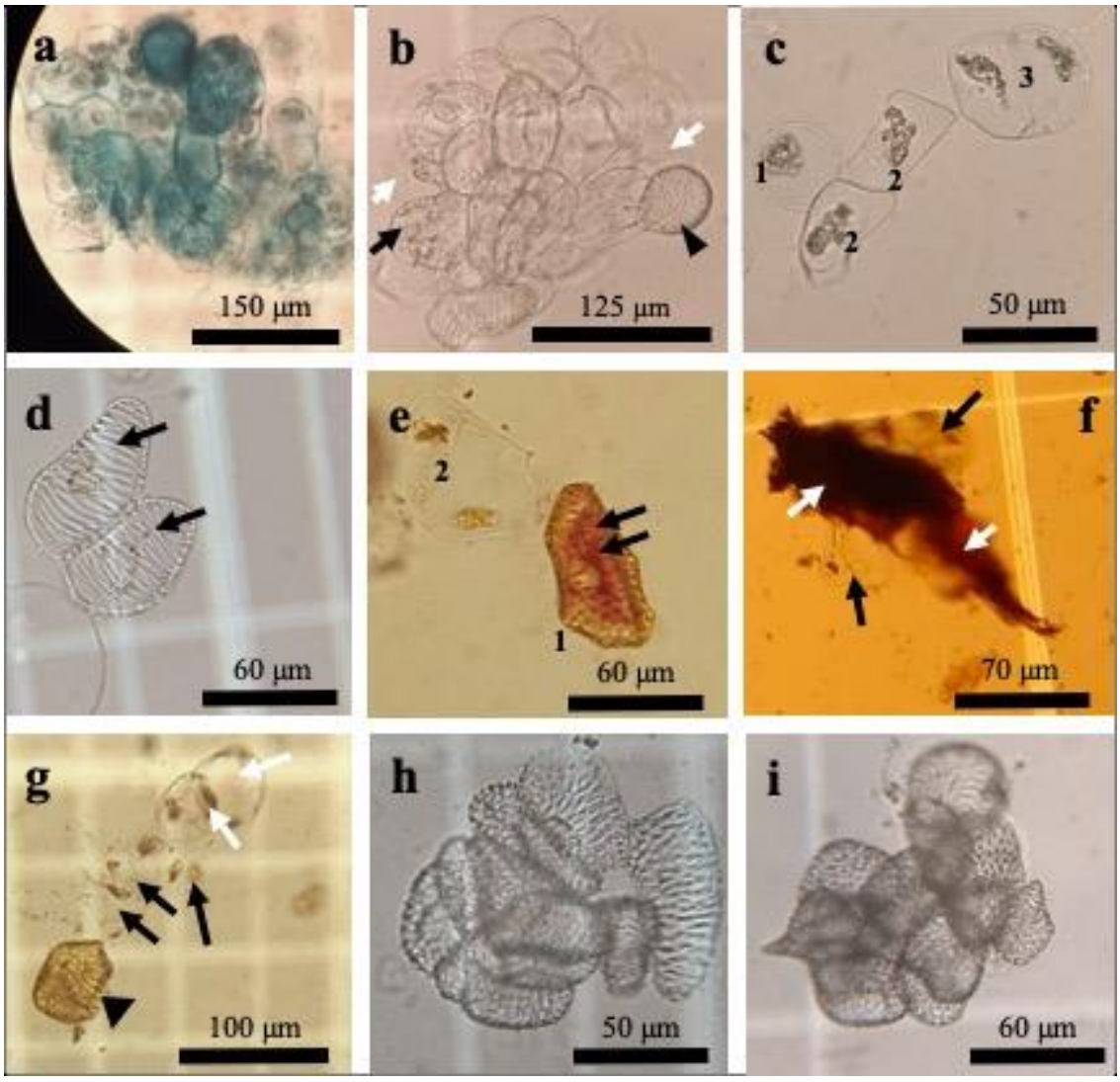

Figure 4.1-1. Micrographs of differentiated cells during *C. roseus* suspension cultures: a. Large aggregate with dead sclerenchyma cells stained by Evans blue. b. Typical size aggregate with differentiated cells: Parenchyma cells (white arrows), collenchyma cells (black arrows) and a sclerenchyma cell (triangle). c. Example of a parenchyma cell (1), collenchyma cells (2) and the differentiation process into collenchyma from parenchyma status (3). d. Example of differentiation process into sclerenchyma cells (arrows). e. a sclerenchyma cell as a tracheary element (1) stained reddish-dark brown, a positive reaction for alkaloids with Wagner's reagent (arrows) and parenchyma cells (2) negative for alkaloids. f. a sclerenchyma cell in the form of massive sclereid- or idioblast-like cells (white arrows) in front of collenchyma cells (dark arrows), positive for alkaloids with Wagner's reagent. g. Parenchyma (dark arrows), collenchyma (white arrows) and sclerenchyma (triangles) cells negative for alkaloids with Wagner's reagent. h. example of cell aggregate composed of only sclerenchyma cells in the form of tracheary elements. i. example of cell aggregate composed of only tracheary elements.

The presence of differentiated cells with different alkaloid contents could explain differences in VC and VB titers (Table 4.1-1). This hypothesis was addressed by evaluating the relationship between differentiated cell population distributions and VC and VB titers. The inherent nature of suspension cultures to form aggregates was not suitable for direct cell differentiation state identification, therefore, to study cell differentiation, subpopulation distribution was measured by using single cell count in enzymatic digested suspension cultures.

Table 4.1-2. Differentiated cell distribution in VC and VB producing and non-producing suspension cultures

			Exogenous plant growth regulator regimens			
			NAA (4.5 µM)		2,4-D (4.5 µM) + BAP (4.5 µM)	
		Differentiated cells type	% cells	VC (mg/L)	% cells	VC (mg/L)
Photoperiod (hd^{-1})	0	Parenchyma	23	0	17	0
		Collenchyma	72		79	
		Sclerenchyma	5		4	
	8	Parenchyma	25	0	21	1
		Collenchyma	70		74	
		Sclerenchyma	5		5	
	16	Parenchyma	31	0	25	27
		Collenchyma	64		59	
		Sclerenchyma	5		16	
	24	Parenchyma	38	0	11	38
		Collenchyma	56		61	
		Sclerenchyma	6		28	
Analysis at day 25						

A relationship between sclerenchyma cells concentration and VC concentration was found when data from all culture conditions was plotted, with a relative high determination coefficient (R_2) of 0.88 (data not shown). From all the operating conditions tested, the culture treated with 2, 4-D (4.5 µM) plus BAP (4.5 µM) was selected because consistent data in terms of small standard deviations for VC concentration was observed. This led to significant differences between VC productions obtained with the different photoperiods tested (One-way ANOVA, post hoc Tukey, $p \leq 0.01$). Results were compared to cultures with NAA (4.5

µM) where no production of VC or VB was observed. Differentiated cell distribution for these two PGR regimens at four different photoperiods is given in Table 4.1-2. The culture with NAA (4.5 µM) contained low levels of sclerenchyma cells independent of photoperiod used, while increased light exposure in the 2, 4-D (4.5 µM) plus BAP (4.5 µM) condition resulted mainly in sclerenchyma cell differentiation, up to 28 % (3 % sclereids, 25 % tracheary elements) of total cell population during continuous light cultures. This increase in sclerenchyma cell percentage was parallel to the increase in VC production (Table 4.1-2). These data strongly suggested that cell differentiation was responsible for different VC production capacity of the cultures. Composition of aggregates of NAA cultures was similar to the aggregate shown in Figure 4.1-1-b, while aggregates composed of only sclerenchyma cells (Figure 4.1-1-h, i) were commonly observed in cultures with relatively high VC and VB titers.

Dynamics of cell differentiation in bioreactors

Results in flask cultures confirmed that culture conditions, such as addition of PGR and light exposure, strongly affected production of VC and VB; moreover, such production was related to cell differentiation, particularly to sclerenchyma cells. Unfortunately, since large volumes of sample (≈ 25 mL) were required for VC and VB analysis, VC and VB concentrations were only analyzed at the end of the cultures. Consequently, two bioreactor cultures were used to analyze the dynamics of cell differentiation and its impact on VC and VB titers. The first bioreactor was pH and DO controlled, while in the second bioreactor culture, those parameters where monitored but not controlled.

For the first culture (Figure 4.1-2), analysis showed that parenchyma cells mainly supported culture growth while collenchyma cells were likely to have limited proliferation capacity. On the other hand, as sclerenchyma cells are programmed to die, particularly in the form of tracheary elements, increase of cell concentration was only caused by cell differentiation. Cell differentiation to sclerenchyma status was likely to be slow and constant during culture, though cell differentiation to collenchyma from parenchyma status was relatively fast, as demonstrated by the abrupt decrease and increase of parenchyma and collenchyma cells respectively around 100 h of culture. Analysis of VC and VB titer profiles in relation to differentiated cells revealed no positive relationship between parenchyma cells with VC and

VB production. During the first 24 h of culture, parenchyma cells increased two-fold but no increase in VC and VB was detected. This fact could explain why fast-growing suspension cultures of *C. roseus* mainly composed of parenchyma cells are usually reported as non-producing VC and VB processes. Direct comparison of collenchyma and sclerenchyma cell profiles with VC and VB profiles revealed a likely correlation of VC and VB production to collenchyma and sclerenchyma status respectively. This is in agreement to previous results in flasks experiments where VC titers were associated to sclerenchyma cells.

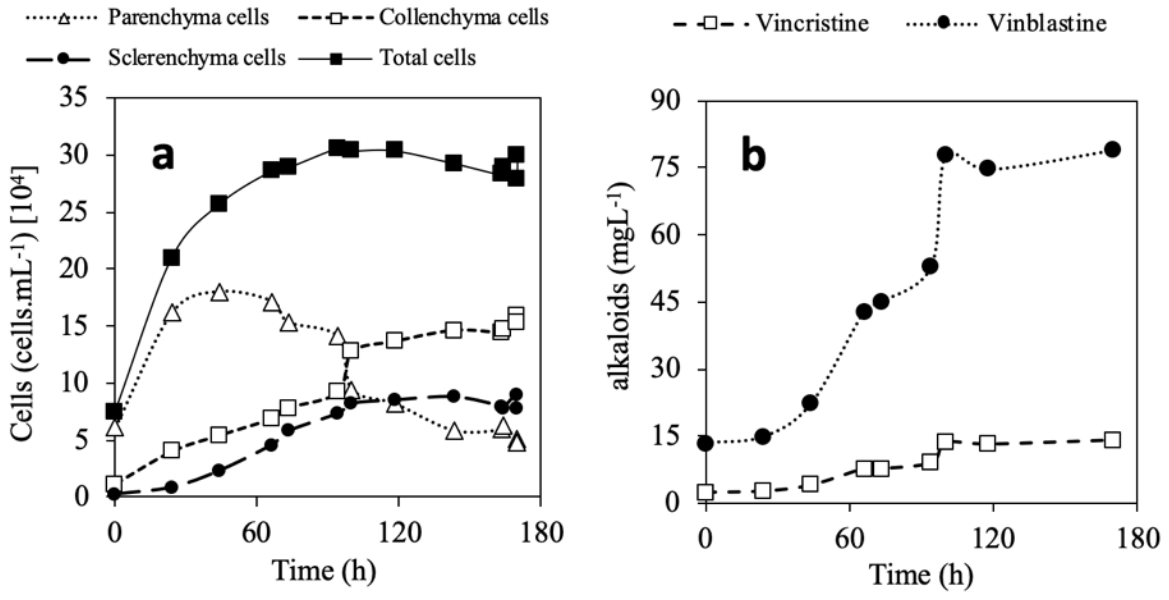

Figure 4.1-2. Kinetic profiles during bioreactor culture producing sclerenchyma cells, VC and VB (first bioreactor culture)

For the second culture (Figure 4.1-3), the increase of cell concentration was also supported by parenchyma cells, though the lag phase was longer than in the first culture. Concentration of collenchyma and sclerenchyma cells remained practically constant during the culture, except for collenchyma cells at the beginning and end of the culture; titers for VC and VB also remained constant with no important increase. Contrary to flask cultures, both bioreactor cultures overproduced VB instead of VC.

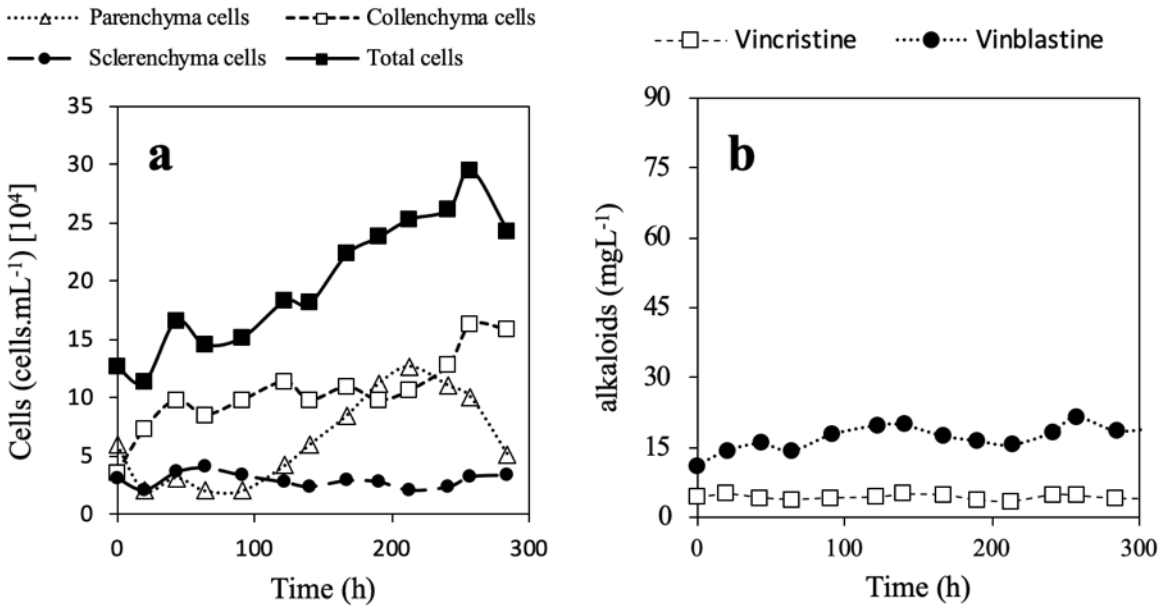

Figure 4.1-3. Kinetic profiles during bioreactor culture not producing either sclerenchyma cells, VC or VB (second bioreactor culture)

4.1.1.1.1.5 Discussion

Cell heterogeneity in callus or suspension cultures of *C. roseus* in terms of cell morphology or differentiation has only been described in a limited way, probably because plant cell growth analysis is usually performed by cell dry weight. In fact, although no consensus has been achieved for the term differentiation in suspension cultures, it has been widely stated that differentiation in *C. roseus* is favorable for producing molecules of interest (Lindsey and Yeoman 1983). The size and elongated form of cells have been related to a cell differentiated state and enhanced indole alkaloid production (Kim et al. 1994a). Other reports have defined cell differentiation in terms of chemical properties of cells. For instance, differentiation has been based on different vacuolar pH or on vacuole color due to the presence of alkaloids and other plant metabolites (Knobloch et al. 1982; Neumann et al. 1983). However, all these reports and the great majority of images provided in publications of *C. roseus* showed relatively homogeneous cell populations of parenchyma-like cells in terms of cell anatomy (Hall and Yeoman 1987; Kim et al. 1994b).

As far as it can be ascertained, sclerenchyma cells have not yet been reported in *C. roseus* suspension cultures and this is perhaps the first report of *C. roseus* suspension cultures

showing empirical data of cell differentiation within cell aggregates based on cell wall differences. Results clearly indicated that differentiation into sclerenchyma status is related to alkaloid expression and more importantly, to VC and VB synthesis. Analysis of cell differentiation analysis in bioreactors suggested that the key for VC and VB synthesis is the presence of sclerenchyma cells in the form of tracheary elements and sclereid-like cells and the presence of collenchyma cells to a lesser extent. These results are in agreement with a former study in suspension cultures of *Cinchona ledgeriana*, where tracheary elements were located at the periphery of aggregates and contained 90 % of total alkaloid titer (Hoekstra et al. 1990). Unfortunately, sclerenchyma differentiation in suspension cultures, particularly into tracheary elements, have mainly focused on woody issues and little is known about alkaloid production.

The PGR regimen and light exposure as operating conditions offered a viable way for inducing cell differentiation, particularly into sclerenchyma cells, and thus VC and VB synthesis (Table 4.1-2). For studying wood formation using suspension cultures, plant growth regulators are usually used to promote differentiation into tracheary elements (Fukuda et al. 1994; Roberts and Haigler 1994; Devillard and Walter 2014), though no clear consensus has been achieved.

Results also showed that light exposure was capable of inducing differentiation into tracheary cells as shown in Table 4.1-2. (Möller et al. 2006) found similar results in callus cultures of *Pinus radiata* where increased light exposure (continuous light or 16 hd^{-1} photoperiod) increased differentiation into tracheary elements. However, cultures in bioreactors suggested that there may be other important variables affecting cell differentiation such as gas composition and more importantly, the physiological state of inoculated cells. For instance, in the first bioreactor with DO control, gas phase composition was partially conditioned by cells and only modified when necessary for maintaining DO at set point, while in the second bioreactor without DO control and with continuous aeration, gas phase was the same as ambient atmosphere. These differences in gas composition and inoculum used for the bioreactor cultures explained difference performance, though the same PGR regimen was used for both cultures. Any cell culture made will be unique due to its physiological state, record of subculture, and inherent heterogeneity (somoclonal variation, aggregates sizes, among others) (Deus-Neumann and Zenk 1984; Patil and Roberts 2013; Bhatia and Sharma

2015). The findings reported here are in-line with this. Thus, for optimization of culture conditions seeking cell differentiation and VC and VB synthesis for production processes, the use of experimental design that enables optimizing the required number of variables in one experiment is recommended.

Results strongly suggested that there is a relationship between VC and VB synthesis capacity and sclerenchyma cells in suspension cultures. This fact implies that cells in suspension cultures globally contained the machinery for VC and VB synthesis provided some degree of differentiation within cell aggregates existed. Indeed, it was recently discovered that *in vitro* cultured cambial meristem cells contained complete sets of enzymes that are responsible for the production of VC and VB from vindoline, though vindoline supplementation was required (Zhang et al. 2015). Part of the *in vivo* metabolic pathway for vindoline has been localized in chloroplasts (De Luca and Cutler 1987) but not usually reported in heterotrophic cultures; however, there have been reports of vindoline-producing cultures under heterotrophic conditions (Scott et al. 1980; Naaranlahti et al. 1989). Thus, there must be alternative metabolic pathways for vindoline, VC and VB synthesis which could be active either during cell differentiation or in particular differentiated cells.

Vindoline, derived from the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway, is usually perceived as the limiting precursor for *in vitro* synthesis of VC and VB. Indeed, it is believed that the last *in vivo* steps for vindoline synthesis occur within idioblast and laticifer cells in aereal tissue, though the principal steps of the MEP pathway for producing tryptamine and secologanin occur in other differentiated epidermal cells, requiring translocation of intermediates for vindoline synthesis (St-Pierre et al. 1999). More recently, *in vivo* expression of enzymes of the MEP pathway that were thought to be restricted to aerial tissue were also detected in phloem cells (Burlat et al. 2004). This has implied that *in vivo* expression of a particular enzyme could be undertaken by several types of differentiated cells in different tissues. This fact is likely the case for the suspension cultures which produced VC and VB. The *in vitro* micro-environment of cell aggregates containing parenchyma cells and sclerenchyma cells (tracheary cells and sclereids) could have mimicked the *in vivo* dynamics of translocation between phloem-aerial tissue to some extent, leading to VC and VB production from vindoline in parenchyma cells (Zhang et al. 2015). These phenomena thus require more attention for further research, particularly the identification of the degree of cell

differentiation, by using biochemical markers preferably in synchronous cultures for a deeper understanding of cell differentiation phenomena. Moreover, the relationship of the progression of cell differentiation with the expression of key enzymes for vindoline, VC and VB synthesis, as well as the effect of the micro-environment within cell aggregates of differentiated cells, should be addressed.

Results provide important clues into the comprehension of *in vitro* culture performance for metabolites production requiring *in vivo* cell differentiation. They demonstrated the utility of taking into account cell differentiation for the further development of novel advanced processes of differentiated cell suspension cultures for producing valuable molecules, including biological medicines such as VC and VB.

4.1.1.1.2 Complement 1: Characterization of cell lines

4.1.1.1.2.1 Material and methods

For this characterization, the inoculum consisted of suspension culture already subcultivated four times. Volumes of 21 mL cell suspension culture were incubated in 50 mL flasks with aluminium foil closure, containing MS proliferation liquid medium for 14 days at room temperature (≈ 25 °C) and agitation (0.22 rcf) in a shaker, with either a 12 hd^{-1} photoperiod (photonic flux of 30 $\mu mol.m^{-2}\ s^{-1}$) or 0 hd^{-1} (total darkness). Duplicate samples were taken at 0, 3, 9, 11 and 14 days, and a single sample was taken on day 6.

4.1.1.1.2.2 Results and discussion

The apex, stem and leaf cell lines were first characterized during 14 days cultures, using either a 12 h-photoperiod or total darkness. The kinetic profiles of both conditions are presented in Figure 4.1-4. The maximum cell biomass was reached on day 11 for both treatments, with higher values for the 12 hd^{-1} photoperiod. It can be observed that the photoperiod drastically affected the pattern of *C. roseus* batch culture. However, although the light effect was quite evident on maximum cell biomass (One-way ANOVA, $p \leq 0.01$), the origin of the explant (leave, stem and apical meristem) had no effect neither in darkness nor in photoperiod (One-way ANOVA, $p \leq 0.05$).

Culture sugar profiles were also compared in Figure 4.1-4 for cultures subject to darkness and photoperiod conditions. All cell lines preferentially consumed glucose than fructose at the beginning of culture, especially under a photoperiod regimen, as was also observed earlier in another study on *C. roseus* suspension cultures (Sagishima et al. 1989). Furthermore, while several authors have reported that maximum biomass is reached after substrate total depletion, both tested conditions in this work reached maximum biomass concentration when sugars were still present in the culture media. A likely explanation is depletion of nutrients or other factors causing growth arrest, such as auxin depletion(King 1976) or accumulation of toxic by-products. Unfortunately, the reason for growth arrest has not yet been clearly determined.

The photoperiod strongly affected the way cells used carbon substrates as showed in Table 4.1-3. Biomass-substrate yields for cell lines cultured in darkness are statistically different to those cultured in photoperiod (One-way ANOVA, $p \leq 0.01$). While the origin of the cell line

(leave, stem and apical meristem tissues) had no impact on yields in darkness cultures (One-way ANOVA, $p \leq 0.05$), in light condition the apex cell line yield was different to those of stem and leaf lines (One-way ANOVA, post hoc Tukey, $p \leq 0.05$). All yields for darkness treatments (Table 4.1-3) are in agreement with those reported previously in *C. roseus* suspension cultures (Pareilleux and Vinas 1983; van Gulik et al. 1989; Rho and André 1991), 0.57, 0.6-0.78 and 0.31-0.35 [$g_{biomass} g^{-1}_{consumed\ sugars}$], respectively.

Table 4.1-3. *Catharanthus roseus* cell lines characterization during suspension cultures in darkness or with 12 h d^{-1} photoperiod

Cell line	Apex		Stem		Leaf	
Light condition (hd^{-1})	12	0	12	0	12	0
*$Y_{x/s}$ ($g_{biomass}\ g^{-1}_{consumed\ sugars}$)	0.89	0.57	0.96	0.54	0.97	0.51
VC ($mg_{VC}\ L^{-1}_{suspension\ culture}$)	0.8	0.7	0.9	1.8	1.5	1
VB ($mg_{VB}\ L^{-1}_{suspension\ culture}$)	31	19	29	43	60	29
* Yield based on total sugar uptake Cells after 4 subcultivations in suspension. All values at day 11_{th}, except for VC and VB at day 14_{th}						

The production of VC and VB by the three different lines cultured in darkness or photoperiod is shown in Table 4.1-3. All cell lines produced VB in greater amounts compared to the negligible observed productions of VC. Contrary to previous *C. roseus* calluses studies reporting loss of synthesis capabilities in darkness (Saiman et al. 2014), VC and VB were detected in all cultures independent of light exposure and cell line origin. This was particularly remarkable for the stem cell line cultured in darkness, which achieved the second largest amount of VB (43 mgL^{-1}) and the largest VC amount (1.8 mgL^{-1}). This result was partially in agreement with previous studies reporting that even in darkness conditions, *Catharanthus roseus* calluses with differentiated root cells were capable of producing VB (Miura et al. 1987). Indeed, some researchers have identified phloem and xylem systems in *C. roseus* callus cultures related to enhanced alkaloid production, especially vindoline, the main precursor of VC and VB (Zhao et al. 2001).

( ) Apex line, ( ) Stem line, ( ) Leaf line.

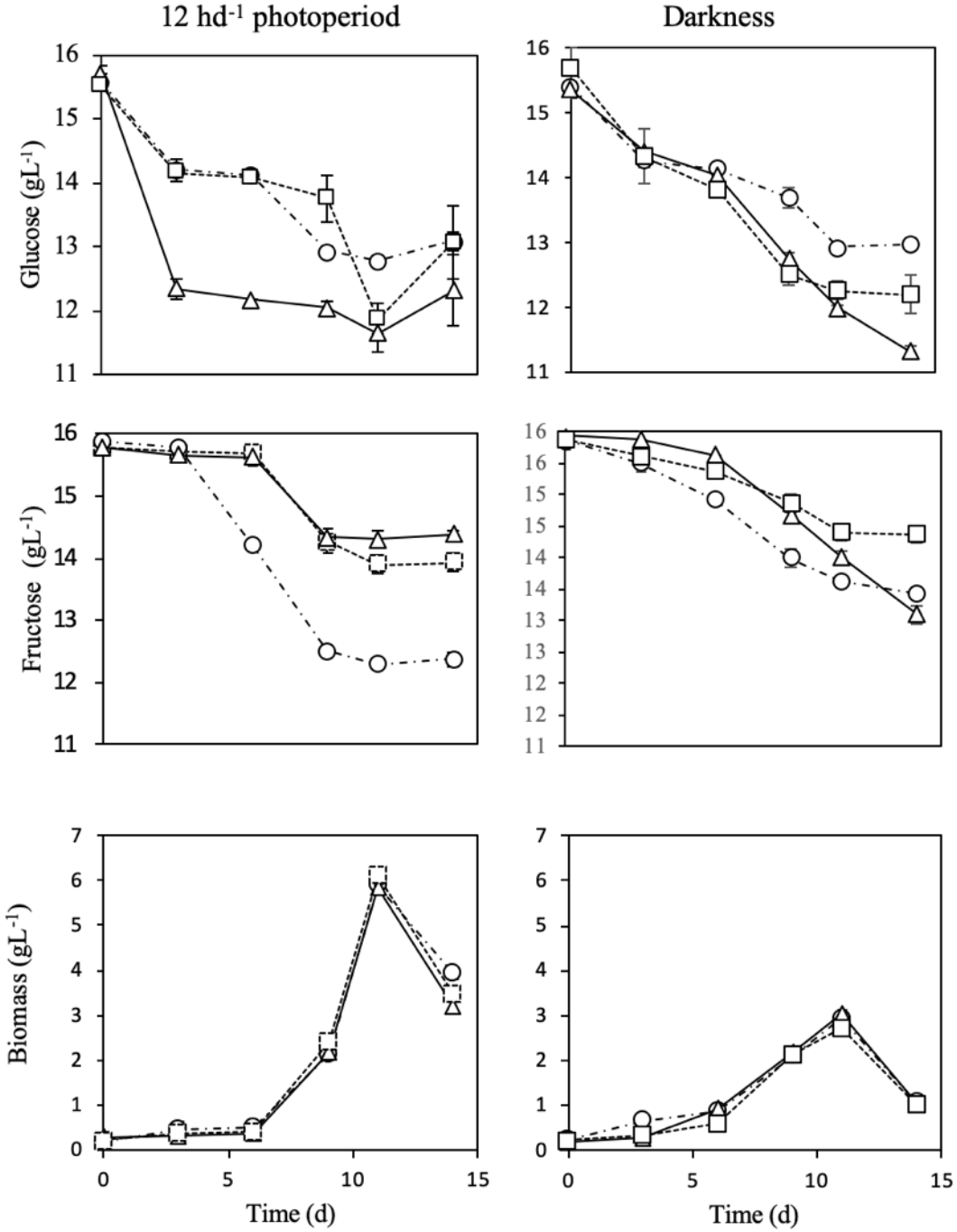

Figure 4.1-4. Kinetics of glucose, fructose and cell biomass concentrations during batch suspension cultures of apex, stem and leaf cell lines, performed in darkness (right) or with 12hd-1 photoperiod (left).

Based on these first results, further evaluation of these phenomena was carried out by performing long term-maintained suspension cultures of the leaf cell line, which achieved the largest production of VB (60 mgL^{-1}) and the second largest production of VC (1.5 mgL^{-1}). The main objective was to evaluate in more detail the effect of some of the culture conditions, such as growth regulators and photoperiods, on cell differentiation and the impact on VC and VB production. Therefore, the work in the intended scientific article in (4.1.1.1.1) only used suspension cultures of the leaf cell line.

4.1.2 Conclusions of Chapter I

The production of the biopharmaceutics VC and VB by *C. roseus* cells was investigated in suspension cultures, considering for the first time the cell differentiation analysis of cell aggregates. The effects of various operating conditions on culture performances, including the addition of PGR and light exposure, were evaluated. In particular, these culture conditions have been shown to affect cell differentiation. Moreover, cell differentiation into collenchyma and sclerenchyma cells was observed to be likely related in the production of VC and VB. Such conditions were then used to successfully produce them by a differentiated cell culture process.

Analysis of differentiation dynamics in bioreactor cultures revealed the need to understand the complex process of differentiation to design enhanced processes leading to high parenchymal cell concentration firstly, and then secondly promoting differentiation for VB and VC production.

These results provided important insights into the design of enhanced production processes. They demonstrated the utility of taking into account cell differentiation for the establishment of advanced processes in cell suspension cultures to produce valuable molecules, particularly biological medicines such as VC and VB. Therefore, cell differentiation is a promising Critical Process Parameter for monitoring and control, and then enhance the cell production processes. Consequently, monitoring of the cell culture process was focused on cell differentiation at bioreactor scale, which was undertaken in 4.4 Chapter IV (4.4.1.2).

4.2 CHAPTER II: EVALUATION OF CLASSIC CALIBRATION TECHNIQUES TO MONITOR CELL CULTURES

4.2.1 Introduction

Cell cultures provide a platform for biopharmaceuticals essential for treating several diseases. Such processes are complex and require strict control in order to assure drug qualities and properties and thus patient safety. Nowadays, only some physical parameters (temperature, pH, among others) are systematically monitored and controlled while chemical and biological parameters, which can also strongly impact drug properties, are usually monitored off-line; this implies delays between sampling and analysis that could compromise control strategies.

Monitoring based on process analyzers using NIRS are promising techniques since they can perform multicomponent analysis without sampling and in real-time when used in in-line modes. Indeed, several studies have shown the potential of this technology in animal cell culture processes. However, its application in industry has been limited in production bioreactors. Animal cell culture processes are highly dynamic and thus a challenging matrix for monitoring. As far as can be ascertained, such complex processes have only been addressed by the use of linear regression methods for building calibration models in order to estimate chemical or biochemical variables, based on NIR spectra.

Consequently, the aim of this chapter has been to evaluate the pertinence of current linear approaches for calibrating in cell culture-based processes. Then, a new approach has been adopted, seeking enhancement of estimating the power of models. This innovative approach tried to take into account the nature of the cell culture process to some extent. The great majority of results from this chapter have been reported in an article published in Biotechnology Progress and are thus organized in this format.

4.2.1.1 Interest of Locally Weighted Regression to overcome non-linear effects during *in situ* NIR monitoring of CHO cell culture parameters and antibody glycosylation

Daniel A. Zavala-Ortiz[1,2], Bruno Ebel[1], Meng-Yao Li[1], Dulce Ma. Barradas-Dermitz[2], Maria G. Aguilar-Uscanga[2], Patricia M. Hayward-Jones[2], Annie Marc[1], Emmanuel Guedon[1]

1- Laboratoire Réactions et Génie des Procédés, Université de Lorraine, CNRS, LRGP, 2 avenue Forêt de Haye, TSA 40602, 54518 Vandœuvre-lès-Nancy, France.
2- Instituto Tecnológico de Veracruz: Calz. M.A. de Quevedo 2779, Veracruz, Ver. Mexico.

4.2.1.1.1 Abstract

Animal cell culture processes have become the standard platform to produce therapeutic proteins such as recombinant monoclonal antibodies (mAb). Since the mAb quality could be subject to significant changes depending on manufacturing process conditions, real-time monitoring and control systems are required to ensure mAb specifications mainly glycosylation and patient safety. Up to now, real-time monitoring glycosylation of proteins has received scarce attention. In this paper, the use of near infrared (NIR) to monitor mAb glycosylation has been reported for the first time. Whereas monitoring models are mainly constructed using linear Partial Least Squares Regressions (PLSR) evidences presented in this study indicate nonlinearity relationship between *in situ* captured spectra and compound concentrations, compromising the PLSR performances. A novel and simple approach was proposed to fit non-linearity using the Locally Weighted Regression (LWR). The LWR models were found to be more appropriate for handling information contained in spectra so that real-time monitoring of cultures was accurately performed. Moreover, for the first time, the LWR calibration models allowed mAb glycosylation to be monitored, in a real-time manner, by using *in-situ* NIR spectroscopy. These results represent a further step towards developing active-control feedback of animal cell processes, particularly for ensuring properties of biologics.

4.2.1.1.2 Introduction

The production of biologicals, especially recombinant monoclonal antibodies (mAb), remains a challenge due to the structural complexity of these molecules and their sensitivity

to changes in the manufacturing process. That is why strict quality control systems are required to ensure mAb specifications and patient safety. For that, regulatory agencies proposed the quality by design (QbD) strategy (Rathore and Winkle 2009), which is rendered possible through the process analytical technology (PAT) approach (Yu 2008; Rathore et al. 2010; Yu et al. 2014). The main objective is to real-time monitor the concentrations of some process parameters, such as viable cell, nutrient and metabolite concentrations, whose variability may have an impact on mAb quality attributes (Teixeira et al. 2009a).

As one of the main quality attributes, glycosylation pattern confers chemical and therapeutic properties to mAb (serum half-life, immunogenicity, antibody-dependent cellular cytotoxicity and complement-dependent cytotoxicity) (del Val et al. 2010; Kayser et al. 2011; Lingg et al. 2012). Therefore, its control is essential to ensure efficacy of the product and safety for patients. However, glycosylation monitoring is usually performed at the end of the culture process because its analysis is time- and labour-consuming (Huhn et al. 2009). Even in advanced production processes supporting the PAT initiative, glycosylation-monitoring still requires some sample handling through the semi-automatic off-line analysis (Alvarez et al. 2011; Doherty et al. 2013; Wang et al. 2017). Moreover, delays due to off-line analysis may also compromise real-time control of the process. A new challenging objective for PAT is thus to control mAb glycosylation as well as cell metabolism using on-line spectroscopy (Hossler et al. 2009; Hossler 2011; Berry et al. 2016; Zhang et al. 2016). Consequently, accurate monitoring models must be developed so that advanced active feedback control systems for controlling processes could become feasible.

In recent decades, vibrational NIR and Raman spectroscopies in combination with multivariate data analysis have been proven to be promising tools for monitoring cell culture process parameters (Cervera et al. 2009; Abu-Absi et al. 2011; Li et al. 2016; Berry et al. 2016). The most widely used multivariate method for developing calibration models from spectroscopic data is Partial Least Squares Regression (PLSR) (Lourenço et al. 2012). PLSR maps linearly spectroscopic spectrum into a low-dimensional space of coordinates called latent variables (LV), which are employed to generate the regression or calibration equations using only linear combinations (Höskuldsson 1988). In this context, *in situ* spectroscopic monitoring has been claimed as an ideal analysis method since it provides real-time

multicomponent information directly without sample treatments, thus avoiding contamination risks and perturbation of compound properties (Arnold et al. 2003). Nevertheless, this on-line implementation remains a challenge because analyses can be subject to perturbations due to the dynamics of the cell culture. Indeed, scattering compounds are generated during cultures, such as cells, cellular debris or mAb aggregates. Their accumulation in the bioreactor may induce modifications of the scattered light reaching the detector, causing changes in apparent absorbance, and thus resulting in non-linear spectra slopes changes (Ge et al. 1994)[23]. These effects, which induce or increase the non-linear relationships between spectra and compounds, may limit the capabilities of classical linear regression methods, particularly PLSR.

Off-line spectroscopy and PLSR have been successfully used to perform monitoring of glycoproteins, particularly Raman spectroscopy (Brewster et al. 2011). However, clarified samples with limited scattering compounds, have been usually required to develop accurate calibration models for glycoproteins (Li et al. 2013). Such current off-line approaches cause monitoring delays that could compromise real-time control of the process. As far as can be ascertained, while only one study has reported *in situ* mAb quality monitoring by Raman spectroscopy and PLSR (Li et al. 2018b), such kind of monitoring using NIR spectroscopy has not been reported yet. This could be due to the complexity of the NIR spectra in scattering systems and the need of chemometric approaches adapted to such systems (Næs 2004; Huang et al. 2010). Taking into account the limitations of linear regression methods, several other methods and techniques are available to extract the information from spectra and overcome non-linearity during the calibration process. They include deletion of non-linear spectra variables, addition of non-linear extra-terms to calibration equations, or use of other regression methods (Verdú-Andrés et al. 1997; Blanco et al. 1999; Xiaobo et al. 2010; Mehmood et al. 2012). The studies dealing with animal cell culture monitoring have been mostly restricted to selection of the best spectra variables using linear regressions. However, such approaches fail to properly address potential non-linear relationships between spectra variabilities and compound concentration changes during cell cultures.

One way to solve non-linear behaviour using the widely known linear regression methods is to perform regression locally, such as the Locally Weighted Regression (LWR) method. This

method aims to model non-linearity using several local linear regressions (Cleveland and Devlin 1988). LWR requires defining the local area that contains the sample to be predicted. Then, each point in this local region is weighted according to its distance from the sample. Thus, in addition to the generation of the regression function for each local area, other parameters have to be determined in order to determine the local area and the point weights (Centner and Massart 1998). Consequently, a local linear calibration equation has to be generated using local data, each time a prediction is required. In contrast to global regressions that treat all the regression surface at the same time, as either linear or non-linear, LWR allows modelling the non-linear regions without compromising linear region predictions. This approach seems particularly adequate for animal cell culture processes, in which linear and non-linear behaviour may arise differently during exponential, stationary or declining growth phases of cell cultures, especially because of light scattering by cells, rendering parameter monitoring difficult.

Therefore, the aim of this paper was to propose an adequate regression method for in-line *monitoring* of CHO cell cultures in a bioreactor, including real-time analysis of mAb glycosylation site occupancy, by using *in-situ* NIR spectroscopy. Firstly, it assessed the capacity of PLSR method to handle eventual non-linear behaviour during animal cell cultures. Secondly, the LWR method was proven capable of dealing with PLSR limitations so that *in situ* monitoring of various culture parameters as well as mAb concentrations and glycosylation patterns throughout batch cultures was possible.

4.2.1.1.3 Materials and methods

Cell cultures and NIR spectra acquisition

The genetically modified DG44 CHO cell line (CHO M250-9), producing a human anti-Rhesus D mAb, was kindly provided by Bioprocessing Technology Institute (Singapore). The culture medium was a protein-free medium mixture consisting of PF-CHO (HyClone) and CD-CHO (Thermo Fisher Scientific) in a 1:1 volume ratio, supplemented with 4 mM L-glutamine (Sigma Aldrich) and 0.1 % pluronic F-68 (Sigma Aldrich). Cell cultures were performed in 2 L benchtop bioreactors (Pierre Guérin) with 1.5 L working volume. The set points for all cultures were 37 °C, 50 % dissolved oxygen, pH 7.2 and 90 rpm. Six cultures were performed in order to obtain off-line measurements, which cover bioprocess variability

(3 batch, 2 feed-harvest, 1 batch with glucose spiking). *In situ* spectral scanning of bioprocess culture media was carried out using a NIR transflectance probe with 1 mm path length (Precision Sensing Devices), connected to the Antaris II spectrometer (Thermo Fisher Scientific). Each NIR spectrum corresponded to an average of 128 scans with an 8 cm^{-1} resolution from 4,000 to 10,000 cm^{-1} wave number (*i.e.* 2,500 – 1,000 nm wavelength).

Off-line measurements

Viable cell density (VCD) was measured using the Vi-Cell XR™ cell counter (Beckman Coulter). Off-line concentrations of glucose, lactate, glutamine and mAb were determined with enzymatic kits using the automated photometric analyzer Gallery™ (Thermo Fisher Scientific). Using UHPLC-MS as previously described (Li et al. 2018b), mAb glycosylation heterogeneity was elucidated.

Development of NIR calibration models with chemometric methods

Calibration spectra was randomly partitioned into calibration (80 %) and validation (20 %) sets. The calibration set involved 134 samples, while the validation 34. A maximum of 4 observations were deleted when they were identified as influential outliers (Hotelling's T-square method, p: 0.95) (Næs 2004). Firstly, in order to generate the PLSR calibration models for concentration of viable cells, glucose, lactate, glutamine, mAb and non-glycosylated mAb, special attention was given to NIR spectra pre-processing. Selection of spectra pre-processing methods was an exhaustive qualitative process to determine and mitigate additive-multiplicative effects and wavelength-dependent baseline variation (Huang et al. 2010). Not less than 30 pre-treatments and their combinations were compared based on model performance evaluated by the root mean squared error of cross-validation (RMSECV) and the root mean squared error of prediction (RMSEP). RMSECV and RMSEP allow a direct measure of accuracy using calibration data and independent data respectively. A model with lower RMSECV or RMSEP is considered more accurate. The Relative Error (RE) was also used to evaluate the models. RE is the relationship of RMSECV or RMSEP with the maximum concentration of a compound during the calibration process. It is used as a

contextualised error of a model so that comparison of accuracy be more meaningful to expected measures during real-time calculations. The particular spectra pre-treatment that led to higher accuracy for a compound PLSR model was given in Table 4.2-1.

Table 4.2-1. Spectral pre-processing used for PLSR and LWR models

	PLS models	LWR models
Viable cells	1st derivative (15, 2,1), Abs, Smoothing (11, 6)	MSC (mean)
Glutamine	Detrend	Detrend, SNV
Glucose	SNV, 1st derivative (15, 1, 1), Abs	MSC (mean), 1st derivative (9, 1, 1)
Lactate	Detrend, SNV	Detrend
mAb	MSC (mean), Detrend, 2nd derivative (21, 2, 2), Abs	MSC (mean), Detrend
NG-mAb	Detrend, MSC (mean), 2nd derivative (15, 2, 2)	Detrend, MSC (mean), 2nd derivative (15, 2, 2)

MSC: Multiple Scatter Correction; SNV: Standard Normal Variate; Abs: Absolute value; OSC: Orthogonal Scatter Correction; Values for derivatives: filter width, polynomial order, derivative order

Subsequently, calibration models were generated using the LWR method. An optimization process was carried out to determine the local areas for regression in terms of number of local points (LPs) and the level of spectra compression in term of principal components (PC) (Næs 2004). LP is the number of nearest calibration samples in the principal component space to be used for a particular local regression, and it represents a good measure of non-linearity. A model with few LPs suggests the presence of strong non-linearity, since more local regressions based on only few samples are needed to properly outline the non-linear relationship between spectra and compounds. As PC is a linear spectrum mapping procedure, it will require more PC to fit a non-linear relationship, then a large number of PC used in a LWR model also suggest the presence of non-linearity. Parallel to a former optimization, a spectrum pre-treatment determination approach as used for PLRS was also performed. The particular spectrum pre-treatment that led to higher accuracy for a compound LWR model is given in Table 4.2-1.

As established for analytical procedures, linearity means the ability of a model to obtain results directly and linearly proportional to actual concentration within a compound concentration range (International Conference on Harmonisation 1996). Then non-linearity

and incapacity of models to handle it, can be detected by a systematic deviation of residuals from the zero line, usually with a curve tendency (Næs 2004). In order to properly detect non-linear behaviour during calibration, residual plots (off-line measured values against difference between off-line measurements and in-line calculated values) were analysed as suggested elsewhere (Slutsky 1998; Centner et al. 1998). The presence of non-linear relationships was confirmed by visual inspection of these plots. If models are capable to handle spectra linearly in relation to concentration, residuals have a random distribution around the zero line. On the contrary, non-linearity is detected by a correlation of residuals, usually in a curve profile. Once non-linear behaviour is visually found, its management by models is statistically taken into account by evaluation of residual correlation by the Durbin-Watson test (Mark and Workman 2007). This test evaluates the null hypothesis that there is no correlation between successive residuals (random distribution of residuals around the zero line), which is likely the case if the model estimates concentration from spectra linearly to actual concentration. Rejection of the null hypothesis indicates that correlation exists and that the model leads systematically to mis-estimation of concentrations because of non-linear relationship existing between spectra and concentration. Multivariate calibration models and statistical analysis were carried out using PLS-Toolbox 8.2.1 (Eigenvector Research Inc.) and Statistics and Machine Learning Toolbox R2016a in MATLAB® environment (MathWorks Inc.).

In situ monitoring of cell cultures

To evaluate the predictive capacity of the models, NIR spectra were automatically acquired *in situ* every 20 min throughout CHO cell cultures. The optimized NIR calibration models were then used to perform real-time calculation of the concentrations of main culture medium components (viable cells, glucose, lactate, glutamine, mAb) as well as of the non-glycosylated mAb.

4.2.1.1.4 Results and discussion

NIR spectroscopy has been proven as useful tool to monitor animal cell cultures under the PAT approach. However, the capabilities of NIR models to detect glycosylation of produced proteins *in situ* remain unknown. Moreover, *in situ* NIR spectroscopy monitoring still requires to be improved so that more accurate estimations of compound concentration could be used to also monitor cell metabolism. As far as can be ascertained, only PLSR has been used to monitor cell cultures *in situ* based on NIR spectroscopy. In this study, PLSR and LWR models were generated and compared to estimate CHO viable cells, glucose, lactate and glutamine concentrations. In addition, models for total mAb and non-glycosylated mAb (NG-mAb) concentrations were also generated with the aim of monitoring the quality of mAbs produced during a process.

Development and analysis of NIR models based on PLSR and LWR methods

Development of PLSR models for monitoring of animal cell culture process

Evaluation of PLSR models comprised different statistical estimators of accuracy, precision and linearity. The coefficient of determination (R^2) is a parameter used to evaluate how the model explains concentration variability based on spectrum variability. Coefficients with values close to one indicate that models relate spectra variability to concentration, while values close to zero suggest no relationship between spectra and concentration. In this context, PLSR was capable of relating spectral variabilities to concentrations for viable cells and glucose, as shown by the high R^2 values in Table 4.2-2. Consequently, PLSR models achieved high accuracy during the calibration and prediction steps as indicated by RMSECV and RMSEP respectively. However, PLSR showed limited capacity to relate spectral variabilities to lactate and glutamine concentrations as indicated by the low R^2 values. PLSR models for such compounds resulted in relative errors above 10 %, usually perceived as the maximum tolerated value for NIR calibration models (Burns and Ciurczak 2008). This phenomenon was also observed for the mAb and NG-mAb PLSR models, with approximately 10 and 20 % of relative errors of prediction (R.E$_{PRE}$) respectively, indicating

that PLSR is not the appropriate multivariate technique for monitoring mAb concentration and quality during animal cell culture processes.

Table 4.2-2. Statistical analysis of calibration models using PLSR or LWR method

		Viable cells	Glucose	Lactate	Glutamine	mAb	NG-mAb
PLSR	RMSECV	6.15	1.94	3.22	0.44	35	19
	RMSEP	8.34	2.51	4.38	0.50	53	18
	R^2_{CV}	0.95	0.95	0.78	0.85	0.84	0.39
	R^2_{PRE}	0.95	0.93	0.67	0.81	0.83	0.50
	R. E_{CV} (%)	6.47	7.19	10.73	10.23	9.21	19.00
	R. E_{PRE} (%)	8.78	9.30	14.60	11.63	13.95	18.00
	LV	5	6	5	5	5	3
LWR	RMSECV	4.97	3.91	1.78	0.29	37	7.1
	RMSEP	5.50	3.62	2.48	0.42	45	11
	R^2_{CV}	0.97	0.81	0.87	0.93	0.86	0.92
	R^2_{PRE}	0.97	0.86	0.82	0.87	0.84	0.80
	R. E_{CV} (%)	5.23	14.48	5.93	6.74	9.74	7.00
	R. E_{PRE} (%)	5.89	13.41	8.27	9.77	11.84	11.00
	LP - PC	13- 6	5 - 4	13 - 3	7 - 3	21 - 4	15 - 15

Units for RMSECV and RMSEP are the same: (cells.mL^{-1}) x10^5 for viable cells, mg.L^{-1} for mAb and NG-mAb, and mM for glucose, lactate and glutamine.

RMSECV: root mean square error of cross-validation; RMSEP: root mean square error of prediction; R^2_{CV}, R^2_{PRE}: correlation coefficients of cross-validation or prediction; R. E_{CV}, R. E_{PRE}: relative errors (in relation to maximum variable concentration) for cross validation (CV) or prediction (PRE); LV: number of latent variables; LP: number of local points; PC: number of principal components; NG-mAb: non-glycosylated mAb.

Nonlinear relationships between cell culture parameters and spectra

The way a model fits data induces some specific structure distribution of the residuals. If a linear model such as PLSR fits a curve relationship between spectra and compound concentrations, residuals will have a systematic deviation. In contrast, if the linear model fits a linear relationship, the residuals would scatter randomly from the zero line. This criterion was then used to graphically detect non-linear relationships and adequateness of PLSR (Figure 4.2-1). A clear hyperbolic profile was observed for NG-mAb residuals, indicating

that a strong non-linear relationship occurred between spectra and NG-mAb concentrations for the whole concentration range tested. In such a case, a linear PLSR model is inadequate and therefore non-linear regression approaches are required to generate proper calibration models. This non-linear behaviour was also observed for mAb to a lesser extent. In addition, a flattened parabola profile was observed for viable cells and lactate concentrations, while a curve tendency was detected for glutamine at concentrations over 3 mM. In these cases, PLSR is considered as adequate only for the concentration range where a linear relationship between actual and estimated compound concentrations is observed, corresponding to a random distribution of the residuals.

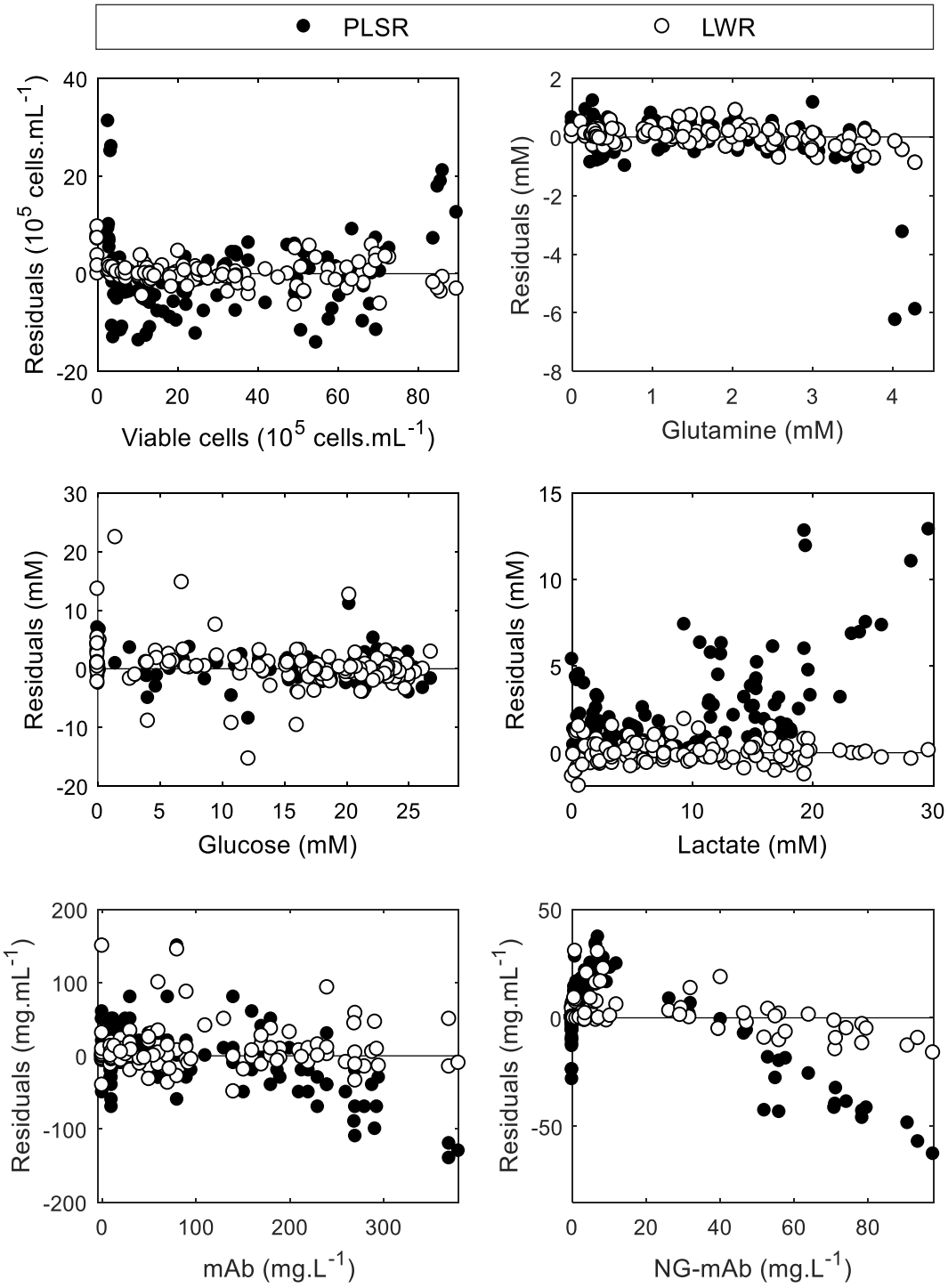

Figure 4.2-1. Detection of non-linearity by inspection of residual distribution within the concentration range tested during calibration for viable cells, glutamine, glucose, lactate, mAb and NG-mAb concentration

Non-linearity was then statistically analysed by the use of the Durbin-Watson test. The test provides a *d*-value related to the nature of the residual distribution, which is then used to evaluate correlation taking into account the nature of the calibration set. Results from this analysis revealed correlation of residuals for all PLSR models (Table 4.2-3), which particularly indicated a strong non-linearity for lactate and NG-mAb. These results suggested, from a statistical point of view, the inability of the PLSR to obtain results directly and linearly proportional to actual concentrations. Other widely used linear approaches, such as Principal Component Regression (PCR), has also been assessed with similar results (data not shown). This is in agreement with a former study reporting similar or even lower performances of these linear approaches (Khajehsharifi et al. 2017). Therefore, the novel implementation of other regression methods considering linearity analysis must be considered.

Table 4.2-3. Results of Durbin-Watson analysis

Compound	PLSR		LWR	
	d-value	*Conclusion*	*d-value*	*Conclusion*
Viable cells	0.73	Nonlinear	1.78	Linear
Glutamine	0.99	Nonlinear	1.71	Nonlinear
Glucose	1.65	Nonlinear	1.75	Linear
Lactate	0.40	Nonlinear	1.92	Linear
mAb	0.89	Nonlinear	1.86	Linear
NG-mAb	0.46	Nonlinear	0.8	Nonlinear

The critical *d* values for Durbin-Watson at α=0.05 are d_L = 1.72 and d_U = 1.74.
If $d < d_L$ the null hypothesis is rejected, the presence of correlation in residuals is indicated (non linearity).
If $d > d_U$ the null hypothesis cannot be rejected, the correlation in residuals is considered to be negligible (no nonlinearity).
If $d_L < d < d_U$ the test is inconclusive.

The presence of non-linear behaviour, which has limited the performance of PLSR, may be explained by the complexity of the cell culture medium used during the process. As NIR spectra contain both physical and chemical information of the samples, it is likely that non-linearity resulted from a wide variety of phenomena, such as variations of light diffusion profiles during cultures (Miller 1993). Chemical phenomena are mainly related to changes in the interaction of several absorbing functional groups, which may lead to shifts of absorption bands or to effects such as the Fermi and Darling-Dennison resonances (Siesler 2010). Such resonance phenomena might require management of non-linearity relationship to properly

extract the information within spectra (Agranovich and Kamchatnov 1999). Therefore, the use of the LWR method has been evaluated with the aim of overcoming the limitations of the linear PLSR method.

Development of LWR models to handle nonlinear relationships and comparison with PLSR models.

The development of LWR models firstly comprised an optimization process to select the size of local areas and the number of PC. If the size of local areas is small, more local regressions must be launched to fit global non-linearities, which also influence the way information within spectra should be handled. This is particularly related to the number of PC required to perform local regressions. Consequently, a compromise between local areas size, in terms of LPs and PC had to be found to avoid overfitting LWR models. The final structure of LWR models is given inTable 4.2-2. The size of local areas in term of number of local points varied from 5 to 21 for the different compounds, which represented approximately 10 % of the calibration set and depicted a strong non-linear behaviour. This non-linearity is likely attributed to the dynamics of the culture process since the locations of the LPs used for the local regressions were mainly determined as a function of culture progression. On the other hand, a high number of PC likely depicts a non-linearity caused by an inherent non-linear relationship between spectra and concentration. Once the global non-linearity was broken, relatively few PC were required to build the local regressions for viable cells, lactate, glutamine and mAb concentration. However, for NG-mAb and glucose, a relatively high number of PC, depending on the number of local points, was required. The LWR models related concentrations of lactate and glutamine with spectral variability more efficiently, as shown by the higher R_2 in contrast with results obtained using PLSR (Table 4.2-2). This enhanced management of spectral variability by the LWR method resulted in a reduction of RMSECV and RMSEP values, which corresponded to decreased relative errors (R.E) of about 35 % for lactate and glutamine. Such reductions allowed concentration estimates with R.E lower than 10 %. The LWR method also enhanced the accuracy for a viable cell model, resulting in a reduction of R.E of approximately 30 %. LWR displayed a similar performance to PLSR for mAb concentration whereas PLSR was higher for glucose (Table 4.2-2).

A remarkable characteristic of the LWR method was its capability to handle the strong non-linearity behaviour previously detected between spectra and concentration, as shown by the higher R^2, particularly during cross-validation, in comparison to PLSR (Table 4.2-2). Analysis of residual plots (Figure 4.2-1) confirmed that in general, LWR not only enhanced accuracy, but also drastically limited the effects of non-linearity, particularly for NG-mAb. This was statistically confirmed by the Durbin-Watson test (Table 4.2-3), which indicated that, excepting NG-mAb and glutamine models, all LRW models properly handled non-linearity and estimated compound concentration linearly to actual concentrations. In this context, the LWR method allowed the development of a calibration model with relative errors of approximately 9 %, opening the possibility to monitor the quality of mAb in terms of glycosylation site occupancy. As a general rule, the LWR method appeared to be the most appropriate model method for the majority of compounds, particularly NG-mAb.

Real-time monitoring of animal cell culture processes

Performances of prediction models based on both PLSR and LWR methods were evaluated during CHO cell culture processes producing mAb in a discontinuous mode. In-line monitoring of viable cells, glutamine, glucose, lactate, mAb and NG-mAb concentrations was carried out using NIR spectroscopy (Figure 4.2-2). In-line predictions were compared with off-line measurements to verify model accuracy. In all cases, the LWR method showed enhanced performance during real-time monitoring.

Contrary to what was expected on the basis of a relative high R^2 (Table 4.2-2) the PLSR model was not able to monitor viable cell concentrations efficiently from approximately 100 h of culture. From this moment, the model was unable to estimate viable cell concentrations from NIR spectra. This unsatisfactory result is in agreement with former reports (Henriques et al. 2009; Clavaud et al. 2013) that pointed out difficulties to model viable cell concentration from NIR spectra due to scattering effects. For example, only limited results of PLSR models of viable cell concentrations were obtained despite the high number of 10 LV required (Henriques et al. 2009). In addition, even with promising PLSR models established for viable cell concentrations, low precision and accuracy appeared once viable cell densities exceeded 80×10^5 cells.mL^{-1} (Clavaud et al. 2013) . The lack of accuracy of the

results could be attributed to the non-linear relationship between spectra and viable cell concentrations. This was previously observed by the right segment of the parabola in the residual plot for viable cell concentrations, which corresponds to concentration values over $75x10^5$ cells.mL^{-1} (Figure 4.2-1). Thereby, a more reliable comparison between PLSR and LWR models, with similar R^2 and relative errors, necessarily requires this linearity analysis to be performed in order to better evaluate the calibration process for in-line monitoring of animal cell cultures. A similar phenomenon was observed for the glucose PLSR model, which was even better than LWR during calibration. However, during prediction, LWR accurately monitored glucose concentration during the whole culture while the PLSR model miss-estimated the concentrations at the beginning and end of the culture. This behaviour was also observed for glutamine and lactate (Figure 4.2-2). The analysis of spectra location in the PLSR space revealed that for the first 10 h of culture they were totally outside the calibration space at 95 % confidence limit (data not shown). Thus, the observed mis prediction at the beginning of the culture could be attributed to an inappropriate extrapolation. On the contrary, the LWR method was shown to be more robust during this phase since culture spectra were always incorporated into the local PLSR space in order to perform a regression.

The mis estimations of lactate, glucose and glutamine concentrations after 120 h of the culture by PLSR model could also be partially explained by the non-linearity phenomenon as formerly observed in Figure 4.2-1. They could also be attributed to the lack of precision or to the presence of noise since even the LWR method showed prediction errors, although weaker, after 120 h. In addition, after 120 h, the culture is characterised by a decrease in cell viability, resulting in the release of intracellular metabolites. Such metabolites could interact with the monitored molecules causing discrepancies in NIR spectra.

This lack of precision of PLSR models at the end of cultures has already been reported, probably due to unknown components within some batches, while other batches showed good predictions (Henriques et al. 2009). Most of the studies, which present the use of PLSR as an efficient method to predict animal cell culture parameters using NIR spectroscopy, do not consider concentration ranges as broad as those reported here, neither the cell death-phase for the process monitoring (Arnold et al. 2003; Cervera et al. 2009; Qiu et al. 2014). Some

other authors have overcome non-linearity behaviour of NIR monitoring models by using at-line analyses of clarified culture medium, as already described (Rhiel et al. 2004; Hakemeyer et al. 2012). In this context, LWR has been proven as a promising method to perform *in situ* monitoring since it allowed calibration in a wide concentration frame, and also considered the nature of the culture medium.

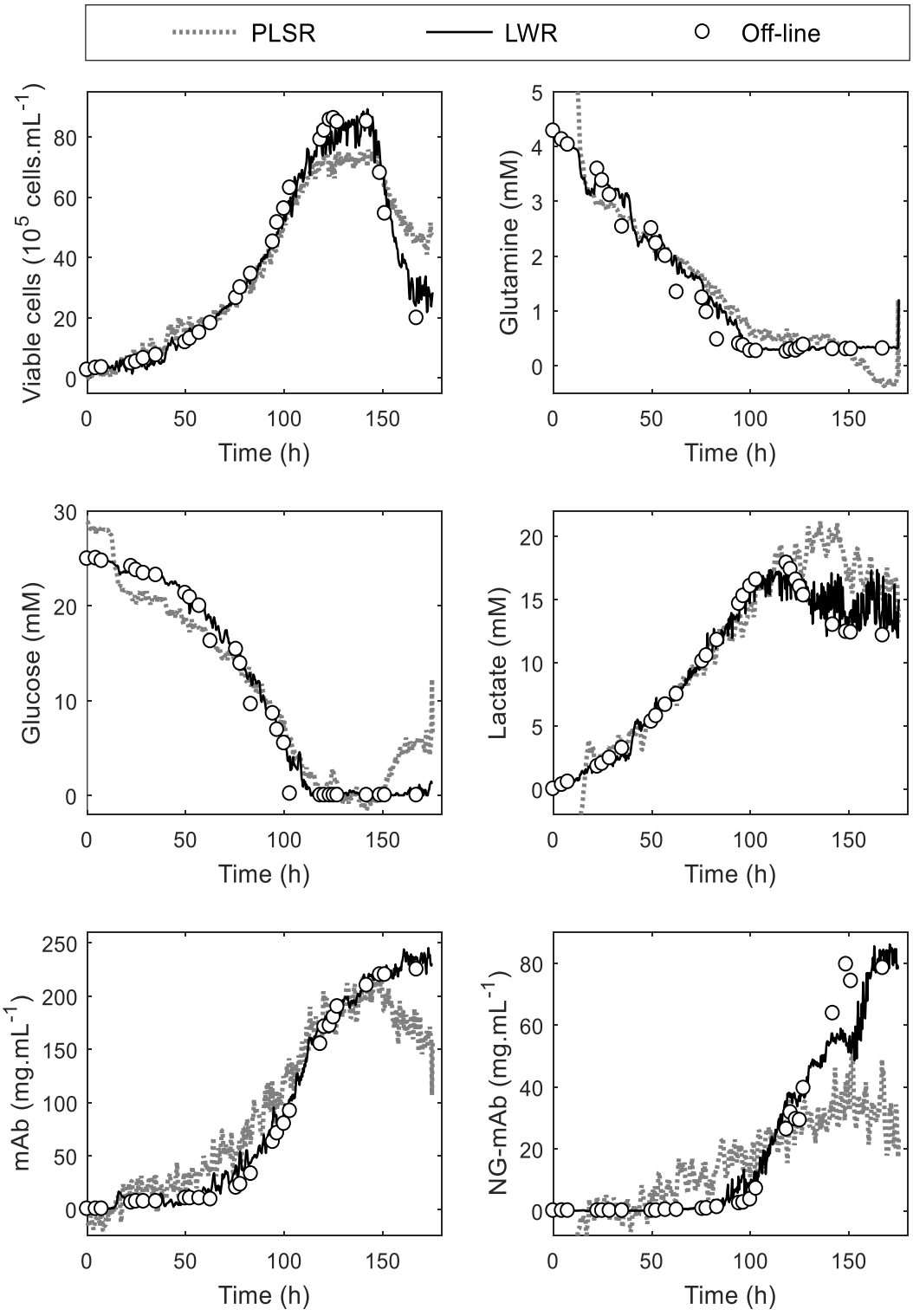

Figure 4.2-2. Real-time monitoring of a CHO cell culture in batch reactor by *in situ* NIR spectroscopy. Comparison of in-line prediction by models using PLSR or LWR regression methods with experimental off-line results.

Real-time monitoring of mAb concentration and quality during animal cell culture processes

Previously developed methods using PLSR and LWR were used to monitor total mAb concentration and non-glycosylated mAb. While the PLSR method showed a good trend for the evaluations of total mAb concentrations during the first days of cell culture, a reduced precision was observed from 140 h. This resulting mis estimation was likely due to the enrichment of NG-mAb within the total mAb molecules. Consequently, results indicated that even models apparently able to monitor properly total mAb concentration can be strongly influenced by changes in mAb properties. This justifies the use of other regression methods, as confirmed by the good performance of LWR prediction model of total-mAb (Figure 4.2-2).

The glycosylation pattern is a key quality parameter of mAb since it confers important properties such as ADCC or serum half-life (del Val et al. 2010; Kayser et al. 2011; Lingg et al. 2012). Therefore, its monitoring and control during process productions as proposed by the PAT initiative, is mandatory to ensure efficacy of mAb and safety of patients. Glycosylation analysis is time-consuming and usually performed by off-line approaches, that may induce a monitoring delay, and then compromise corrective action in order to maintain desired glycosylation properties. In this work, *in situ* NIR spectroscopy capability to real-time monitor non-glycosylated mAb concentrations has been proven, provided the LWR method was used in place of the PLSR.

The performance of PLSR to monitor NG-mAb concentrations was first evaluated. As shown in Figure 4.2-2, the PLSR method completely failed to predict accurately NG-mAb concentrations throughout the culture, accordingly to low R^2 values (Table 4.2-2). This result was expected because during calibration, a strong non-linear relationship between residuals and NG-mAb concentrations was observed (Figure 4.2-1), indicating the need for a more efficient regression method to monitor NG-mAb concentrations by *in situ* NIR spectroscopy. Indeed, the use of the LWR method made it possible to reduce the prediction errors and to obtain a good monitoring of NG-mAb concentrations (Figure 4.2-2).

4.2.1.1.5 Conclusions

In this study, experimental evidence of non-linear parameter behaviour in animal cell culture processes was provided. Consequently, the widely used PLSR method was incapable of relating spectra with compound concentrations, indicating that such a widely used regression methods is not always appropriate for the monitoring of animal cell culture processes. The novel use of the LWR method was shown to overcome PLSR limitations, which led to more accurate predictions of culture compound concentrations. Using NIR spectroscopy, the enhanced capability of LWR to handle non-linearities permitted for the first time, the *in-situ* monitoring of mAb glycosylation site occupancy. Overall, the results highlighted the fact that *in situ* NIR spectroscopy could have a broader potential as a PAT tool provided that effect of culture dynamics and nonlinearity be considered. In this context, NIR spectroscopy could be used to develop innovative spectroscopic calibration models so that effective control approaches to guarantee quality of antibodies could be implemented.

4.2.1.2 Complement 1: Performance of PCR models

As formerly stated, PCR regression was also analysed in terms of calibration itself and for real-time monitoring performance. PCR had a similar performance to PLSR either during calibration (Table 4.2-4) or during real-time monitoring. The same batch used for validation in the publication was used for analysing the nature of PCR regression for real-time analysis.

Table 4.2-4. Statistical analysis of PCR models

		Glucose	Lactate	Glutamine	mAb	NG-mAb	VCD
PCR	PC	6	5	5	5	5	5
	RMSECV	2.55	2.59	0.48	0.031	0.018	6.77
	RMSEP	2.65	4.08	0.58	0.045	0.016	8.34
	R^2_{CV}	0.91	0.84	0.81	0.88	0.41	0.93
	R^2_{PRE}	0.92	0.69	0.75	0.86	0.57	0.92
	R. $error_{CV}$ (%)	9.44	8.63	11.16	8.16	18.00	7.13
	R. $error_{PRE}$ (%)	9.81	13.60	13.49	11.84	16.00	8.78
	Spectral treatment	SNV, 1st derivative (11, 1,1)	Detrend, 2nd derivative (121,6,2), abs	MSC, detrend, abs	MSC, detrend	Detrend, MSC, 1st derivative (15,2,1)	MSC, 1st derivative (15,2,2)
	Units for RMSECV and RMSEP are the same: (cells.mL^{-1}) x10^5 for viable cells, mg.L^{-1} for mAb and NG-mAb, and mM for glucose, lactate and glutamine. RMSECV: root mean square error of cross-validation; RMSEP: root mean square error of prediction; R^2_{CV}, R^2_{PRE}: correlation coefficients of cross-validation or prediction; R. $error_{CV}$, R. $errir_{PRE}$: relative errors (in relation to maximum variable concentration) for cross validation (CV) or prediction (PRE); LV: number of latent variables; LP: number of local points; PC: number of principal components; NG-mAb: non-glycosylated mAb.						

Calibration based on PCR firstly reduced spectral dimensions in terms of principal components and then regression is performed. Thus, regression is completely based on the trajectory of the cell culture processes (Figure 4.2-3-a). Then misprediction of PCR in some situations as in the beginning of cultures (Figure 4.2-3-b) were due to extrapolation issues.

For instance, during the first 14 h of culture, PCR models were incapable of estimating lactate concentration, which was likely caused by the fact that spectra (violet dots in Figure 4.2-3-a) for that particular period were outside the calibration space (dashed circle in Figure 4.2-3-a).

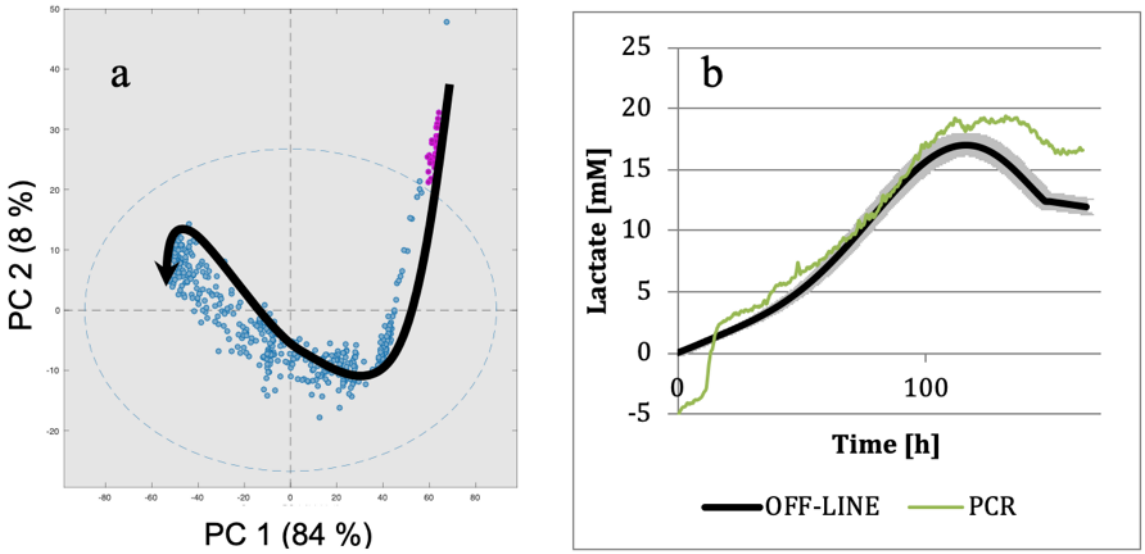

Figure 4.2-3. Analysis of PCR performance for real-time monitoring: a. Trajectory of culture process, b. performance of PCR model for real-time monitoring of lactate concentration

These results agreed with those for PLSR models in the article and pointed out the vulnerability of global linear models for fitting cell culture processes. Moreover, PCR was incapable of handling non-linear relationships, for example the non-linear relationships for lactate at the end of cultures (Figure 4.2-3-b). Overall, as for PLSR models, PCR seemed not a robust regression technique for cell culture monitoring using NIR spectroscopy.

4.2.1.3 Complement 2: Alternative PLSR models

A common approach for enhancing estimating power of models are spectral selection. It is intending for using only absorption values at particular wavelengths which are related to the compound of interest and excluding those with unrelated information and noise. However, the identification if particular wavelengths contain pertinent information for a model requires exhaustive methodology that may be undertaken by off-line analytical or chemometric approaches. Off-line approaches are based on previous knowledge of a compound, as in the use of samples with different concentration values of the compound of interest, which would eventually reveal changes in spectra at particular wavelengths as function of concentration. Then such a spectral region is likely to contain information mainly about the compound of interest. Although this approach may be feasible in relatively simple matrices, it becomes more difficult in complex matrices such as cell culture media where there may be complex interactions with other compounds. Another way of solving this challenge is the use of chemometrics techniques such as the interval and iterative PLSR method or the use of spectral compression seeking noise reduction, such as the use of Fourier transformations.

Interval and iterative PLSR (iPLSR)

This method is intended for selecting only a subset of variables (absorptions at particular wavelengths) that would eventually provide superior prediction power compared to using all the spectral NIR range. It performs a sequential and exhausitive search for the best combination of variables, seeking to reduce noise or variables without pertinent information. There are two main approaches for performing the research, which are called forward iPLSR, where variable intervals are successively included in the analysis, and the reverse iPLSR, where intervals are successively removed from the analysis. Though both approaches can be used in single variable or a range of adjacent variables, the latter is preferred in spectroscopic data since there may be spectroscopically correlated variables.

Fourier transformations:

In the 1990s, Fourier decomposition was particularly used for saving spectroscopic data in a compressed way since it is possible to represent NIR spectra with 99.999 % similitude with only 100 – 200 Fourier scores.

Fourier scors analysis relies on the Fourier series decomposition that states that any continuous function may be represented in the form of sin and cosine waves at different frequencies. The Fourier series are good to represent curves with the same degree of smoothness and they are not efficient to represent curves that are smooth in places and spiky in others. If the number of sine and cosine waves increases, the approximation to the function will be close. The great majority of the information of the function will be present in the first low frequencies waves, and the high frequencies waves will account for small variability and noise.

The use of Fourier scores (FS) for regression was intended for mimicking the nature of NIR absorption to some extent. Total absorption at a particular wavelength (λ) is caused by the sum of individual absorptions of several absorbing compounds as shown in Equation 23.

$$A_{\lambda 1} = l\,(\varepsilon_1 C_1 + \varepsilon_2 C_2 + \varepsilon_{..} C_{..} + \varepsilon_n C_n)$$

Equation 23

where ε is the molar attenuation coefficient or absorptivity of the attenuating species, C the concentration of the attenuating species and l the optical path length.

This approach sought to assign particular frequencies to particular compounds at low frequencies and eliminate noise by discarding high frequencies.

The use of PLSR with chemometric techniques related to spectral selection were developed for enhancing predictive power of models. Two main strategies were followed:

1) Use of interval and iterative PLSR:

   The iPLSR was used in forward mode, which first divided total spectral into frames (either 40 or 50 variables frames). It created individual PLS models (one for each frame) for the first iterative cycle. Cross-validation was then performed for each of these models and the interval which provided the lowest RMSECV was retained. For the second cycle, the already retained model in the first cycle was used in all models but was combined with each of the other remaining intervals, one at a time, when creating a new set of new PLSR models. The RMSECV was then used for selecting the best combination of intervals and the whole processes was repeated 30 times. This approach was used for building the models using up to eight latent variables. Then the best combination of variables and latent variables was selected for a particular compound.

2) Spectral compression by Fourier transformations & iPLSR:

   First, spectra were compressed using Fourier transformation and 770 Fourier scores were retained. PLSR models were then generated using only 200 FS which can be used for representing spectra with more than 99.99 % similitude. Then forward iPLSR was used using only the first 200 low frequency FS.

By using iPLSR some particular wavelengths and frequencies were identified as some particular compounds using either Fourier scores as inputs or spectra. However, no general enhancement was observed. Indeed, the use of iPLSR had detrimental effects for lactate, glutamine and glucose while no effect was observed for mAb or NG-mAb when evaluating models by either RMSCV or RMSEP. Only viable cells showed a slight enhancement (Table 4.2-5).

Table 4.2-5. Statistical analysis of alternative PLSR models

		Viable cells	Glucose	Lactate	Glutamine	mAb	NG-mAb
PLSR	RMSECV	6.15	1.94	3.22	0.44	35	19
	RMSEP	8.34	2.51	4.38	0.50	53	18
	R^2_{CV}	0.95	0.95	0.78	0.85	0.84	0.39
	R^2_{PRE}	0.95	0.93	0.67	0.81	0.83	0.50
	LV	5	6	5	5	5	3
iPLSR	RMSECV	5.77	1.93	3.44	0.44	37	19
	RMSEP	7.18	3.01	11.46	0.68	50	18
	R^2_{CV}	0.95	0.95	0.85	0.85	0.85	0.38
	R^2_{PRE}	0.94	0.90	0.56	0.66	0.84	0.42
	LV	5	5	3	7	7	3
iPLSR-Fourier	RMSECV	5.40	1.96	3.66	0.46	38	18
	RMSEP	7.21	2.98	4.41	0.49	51	19
	R^2_{CV}	0.96	0.95	0.72	0.84	0.82	0.38
	R^2_{PRE}	0.93	0.91	0.66	0.85	0.83	0.37
	LV	11	5	9	7	11	7

Units for RMSECV and RMSEP are the same: (cells.mL^{-1}) x105 for viable cells, mg.L^{-1} for mAb and NG-mAb, and mM for glucose, lactate and glutamine.
RMSECV: root mean square error of cross-validation; RMSEP: root mean square error of prediction; R^2_{CV}, R^2_{PRE}: correlation coefficients of cross-validation or prediction; LV: number of latent variables; NG-mAb: non-glycosylated mAb.
Spectral frames: 121-160, 241-320, 361-400, 441-520, 841-920, 1041-1160, 1281-1320 (Viable cells) ; 301-400, 451-500, 801-850, 1451-1500 (glucose) ; 441-495, 551-605, 771-825 (lactate) ; 401-440, 561-600, 681-720, 801-840, 1081-1120 (Glutamine) ; 151-200, 351-400, 451-600, 801-850 (mAb) ; 151-200, 351-400, 451-500, 1201-1250 (NG-mAb).
Fourier scores frame: 7-11, 13-17, 21-2,; 35-37, 41-43 (viable cells); 1-4, 17-20, 25-40 (glucose); 13-15, 28-30, 46-48, 61-66, 70-78 (lactate); 5-8, 17-24, 33-36 (glutamine); 1-6, 11-14, 20-30, 33-60, 65-66, 69-74, 79-82, 89-90 (mAb); 3-14, 17-40, 63-64, 67-76, 79-80, 89-118 (NG-mAb)

The iPLSR models using either spectra or FS as inputs, were then compared for viable cells monitoring as shown in Figure 4.2-4. Kinetic profiles performed by either iPLSR or FS-iPLSR for viable cells revealed no enhancement for real-time monitoring issues. Both mispredicted viable cell concentration at the stationary phase as formerly discussed in the article for PLSR models.

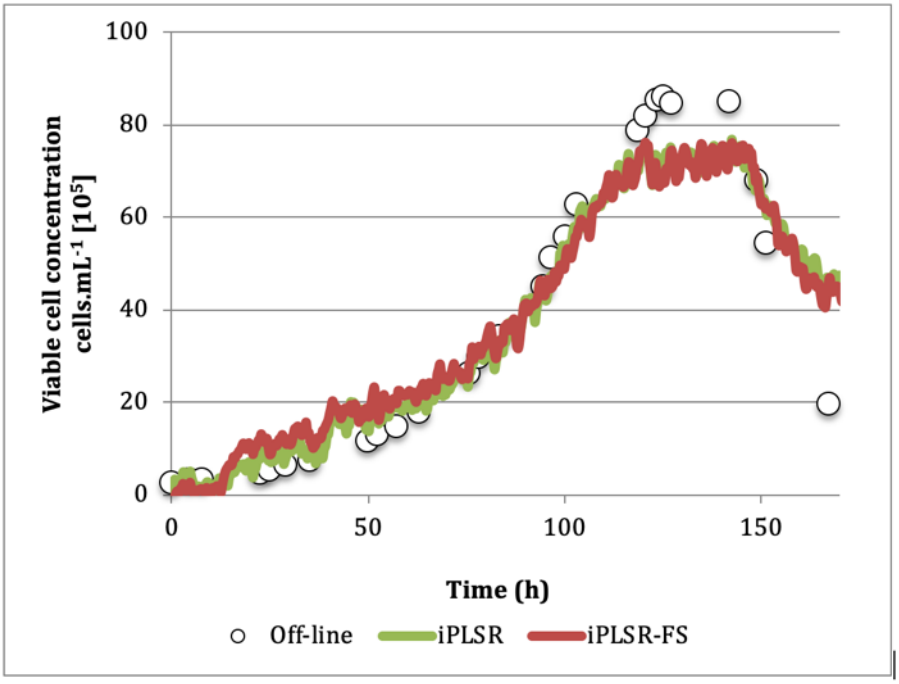

Figure 4.2-4. Performance of PLSR alternative models during real-time monitoring

Variable selection was intended for identifying wavelengths producing the smallest prediction error and thus better performance during real-time monitoring. Although many studies have shown the benefits of variable selection in PLSR models, this has not been the case in the present cell culture system. Lacking enhancement may have been due to remarkable interdependency of compound absorption or nonlinearity that cannot be properly managed by the classic PLSR method.

4.2.2 Conclusions of chapter II

In this chapter, the adequacy of PLSR models for cell culture processes was investigated. First, the performance of PLSR was evaluated and it revealed that cell culture processes are extremely dynamic and that diverse phenomena limited PLSR for accurate estimation of compound concentration. These phenomena were globally identified as nonlinear relationships between spectra and especially concentration of mAb. Nonlinearity compromised the most important assumption for PLSR calibration modelling: the relationship between spectra and compound concentrations should be constant during the cell culture process to be able to use the same regression equation to perform concentration estimations. The failure of this essential assumption had caused concentration misestimations for several compounds, particularly at the beginning and end of cell cultures where nonlinearity mainly arose. Moreover, there were cases where nonlinearity was not based on cell culture progression but were inherent during the whole calibration space, such as for NG-mAb concentration.

In contrast to PLSR which is a variable-based method, LWR is a sample-based method. In other words, PLSR generates calibration equations based on the relationships between variables (absorption at different wavelengths) and concentration, while LWR first accounts for similarity between calibration samples and uses only similar samples for calibration and a calibration equation is generated. Calibration under a local regression approach using LWR seemed appropriate for enhancing estimating power of models, particularly during real-time monitoring within different phases of batch cell culture. The key for this enhancement was the similarity analysis between calibration samples for creating the calibration model. Selection of similar samples for calibrating was closely related to cell culture progression. Using similar calibration samples enhanced scattering management by spectral pre-processing and also generated equations based on matrices with similar compound intercorrelations, leading to better generalization of information within spectra and thus enhanced estimating power by models.

Results have proven that the use of LWR could enhance the power of NIR monitoring procedures based on spectra captured using *in situ* probes. However, this approach also

requires consideration of more factors; LWR would require more attention in the calibration process since local calibrations would require more samples in a homogenized distribution in the calibration space. Besides, attention to outliers becomes essential since regression is based on few local data and the effect of an outlier in the local regression space may have an enormous effect. On the other hand, a perspective issue is also of concern when using calibration models based on sample-based regressions, since such concentration estimation is based more on empirical assumptions (sample similarity) than models based on variable-based regressions (specificity).

Overall, these results have shown the complex nature of cell culture processes and the need for more sophisticated calibration approaches for properly dealing with diverse phenomena depicted as nonlinear relationships between spectral and compound concentration. The undertaken local approach seemed appropriate for dealing with nonlinearity related to cell culture progression; however, there were other nonlinearity sources which could not be properly managed by the local regression approach, which likely required the novel use of nonlinear regression methods.

4.3 CHAPTER III: NOVEL IMPLEMENTATION OF NONLINEAR CALIBRATION APPROACHES

4.3.1 Introduction

In the former chapter, the vulnerability of PLSR models for monitoring of cell cultures was shown to be due to nonlinear relationships between spectra and concentration. Calibration based on local regressions using LWR, had shown enhanced performance, though there were still strong nonlinear relationships where even LWR failed. Therefore, in this chapter the novel use of nonlinear regression methods for generation of calibration models has been explored. Indeed, several studies have shown the potential of these regression methods in spectroscopic data. However, as far as can be ascertained its application for cell culture monitoring had yet not been addressed.

In this chapter, Partial Least Squares Regression (PLSR), Supported Vector Regression (SVR) and Artificial Neural Network Regression (ANNR) were used for the generation of calibration models. The performance of models was evaluated through different perspectives with focus on their performance under real-time monitoring conditions. Calibration methods were firstly evaluated for accuracy, precision and linearity using calibration data. Secondly, their performance during real-time monitoring was addressed with focus on the effect of inter-batch heterogeneity, specificity to the compound of interest and robustness. This study was exploratory and interpretative in nature, seeking to understand the behavior of calibration models in cell culture processes. Then it could be used as a friendly frame for basic interpretation issues facilitating proper management of the NIRS procedure lifecycle. Results for this chapter are intended for publication in Biotechnology Journal and they are accordingly organized for this purpose.

4.3.1.1 Evaluation of NIR calibration models for *in situ* CHO cell culture monitoring: PLSR, ANNR & SVR

Daniel A. Zavala-Ortiz[1,2], Bruno Ebel[1], Meng-Yao Li[1], Dulce Ma. Barradas-Dermitz[2], Maria G. Aguilar-Uscanga[2], Patricia M. Hayward-Jones[2], Annie Marc[1], Emmanuel Guedon[1]

1- Laboratoire Réactions et Génie des Procédés, Université de Lorraine, CNRS, LRGP, 2 avenue Forêt de Haye, TSA 40602, 54518 Vandœuvre-lès-Nancy, France.
2- Instituto Tecnológico de Veracruz: Calz. M.A. de Quevedo 2779, Veracruz, Ver. Mexico.

4.3.1.1.1 Abstract

The biopharmaceutical industry has to assure the consistency and biosafety of biological medicines which are quite sensitive to cell culture variability. Therefore, advanced retro-control strategies must be implemented in cell culture processes for proper control of biochemical process parameters, such as concentration of cells, substrates and by-products. Near Infrared spectroscopy coupled to multivariate analysis has been shown as a promising technique for monitoring cell culture processes. However, there are still some challenges for its successful establishment in industry. In this study the novel use of Neural Networks and Supported Vector regression nonlinear techniques has been explored for the generation of calibration models and compared with Partial Least Squares. Nonlinear approaches have been shown superior to PLSR during real-time monitoring by better management of inter-batch heterogeneity and enhanced specificity to particular compounds. Overall, the use of SVR and ANNR for the generation of calibration models has enhanced the potential of NIR spectra as a monitoring tool.

4.3.1.1.2 Introduction

The production of biologicals remains a challenge due to the structural complexity of these molecules and their sensitivity to changes in the manufacturing process. That is why the Quality by Design initiative is being established for building quality through the production processes. Accordingly, continuous monitoring of Critical Process Parameters (CPP) affecting the properties of biopharmaceuticals is required to establish advanced retro-control

systems which assure the final clinical effect of medicines. Among the diverse CPP affecting cell culture performance, composition of culture media has a wide effect since it contains substrates for cell proliferation and product synthesis or toxic by-products spoiling the cell physiological state and thus process performance. Thus, monitoring their NIRS is a promising tool since it is capable of providing multicomponent information directly without sample treatments using *in situ* analysis modes (Cervera et al. 2009; Li et al. 2016). Nevertheless, NIR spectra and culture media dynamics are complex and multivariate calibration methods are required to extract and relate the observed spectra, in an estimation manner, to a desired variable property, such as nutrient and by-product concentration. The design of the multivariate calibration model is not a trivial matter and responds to various technical and regulatory factors. Technical factors mainly concern the nature of the calibration process, while regulatory factors rely on proper validation (accuracy, precision, specificity, linearity, range of operation and robustness) and management of the NIRS calibration procedure lifecycle.

Concerning the technical factors, there are several challenges for successfully calculating compound concentration using NIRS calibration models: confused relationships between compounds, complicated relationships between spectra and compound concentration, inter-batch heterogeneity, noisy spectra and process changes during normal operation. Indeed, a complicated relationship between spectra and concentration is the reason why multivariate methods are required. Moreover, for *in situ* approaches in chemically and physically complex matrix such as cell culture media, extreme care must be taken to generate *in situ* calibration models, especially in the regression method.

Currently, Partial Least Squares regression (PLSR) is the most common regression method to perform calibration models for cell culture monitoring. PLSR maps spectral data linearly into low-dimensional space, and then low-dimensional coordinates are employed to generate the regression or calibration equations using only linear combinations (Höskuldsson 1988). Although PLSR is capable of addressing complicated linear relationships, it is incapable of properly addressing strong non-linear relationships. However, it may be modified to handle them by including non-linear regression coefficients in the calibration equation or by local modelling (Centner and Massart 1998). Several varieties of PLSR such as Poly-PLSR, Spline-PLSR, Linear-quadratic PLSR and others have been tested for NIR calibrations but

unfortunately such approaches have shown to be inferior in predictive capacity compared to Artificial Neural Network Regression (ANNR) and Supported Vector Regression (SVR) in spectroscopic data (Blanco et al. 1999; Balabin et al. 2007; Balabin and Lomakina 2011). Nevertheless, PLSR and other linear approaches have been proven proper methods in major cases and are broadly considered in guidelines for NIRS analytical procedures of regulatory agencies for cell culture monitoring. This can be explained by the fact that those semi-parametric methods also provide a friendly frame for interpretation issues that facilitate proper management of the NIRS procedure lifecycle.

Non-linear regression approaches have several advantages, though still require some advanced computer power when dealing with large amounts of data. A major advantage is their flexibility to model complex relationships, although model interpretation may become cumbersome due to the stochastic and generally non-parametric nature of such regressions, so that management of the NIR procedure may become difficult. This is exemplified for ANNRs, whose final structure may depend on initial training parameters and the fact that extremely different net architectures may achieve comparable results (Hagan et al. 2004). In addition, common tasks in parametric regression methods such as variable selection or input pre-treatments require exacerbated efforts and consequently selecting model inputs and their nature are often delegated to the regression process itself (Suzuki 2011).

Recently, non-linear and non-parametric regression approaches based on SVR and ANNR have gained popularity in developing NIR calibration models (Meyer and Weigelt 1992; Cogdill and Dardenne 2004; Brudzewski et al. 2006; Ferrão et al. 2007; Balabin et al. 2007; Wu et al. 2008; Liu et al. 2008; Balabin and Lomakina 2011), but as far as can be ascertained such regression methods have not been addressed in building calibration models for cell culture monitoring. To date, there has been no reliable evidence that provides a reference frame to evaluate the convenience of regression methods, either linear or nonlinear, for building calibration models of culture medium compounds during cell cultures. Therefore, the aim of this study is to provide an extensive empirical frame to evaluate the convenience of different regression methods (PLSR, ANNR and SVR) for monitoring CPP in CHO cell culture processes.

4.3.1.1.3 Materials and methods

Cell culture for data acquisition

Cell cultures of CHO cells were performed in 2 L bench-top bioreactors (Pierre Guérin, France) with a 1.5 L working volume. Six bioreactor cultures were performed for the calibration set, obtaining off-line measurements and for covering bioprocess variability: three batch cultures, two feed-harvest cultures with medium renewal and one batch culture with glucose spiking. A batch culture was also performed as an external validation culture, using ActiPro (HyClone) as the culture medium. The culture medium for all other cultures was a protein-free medium mixture consisting of a 1:1 volume ratio of PF-CHO (HyClone) and CD-CHO (Fisher Scientific) supplemented with 4 mM L-glutamine (Sigma Aldrich) and 0.1 % pluronic F-68 (Sigma Aldrich). The genetically modified DG44 CHO (CHO M250-9) cell line was used, kindly provided by the Bioprocessing Technology Institute (Singapore). Dissolved oxygen (DO) was controlled at 50 % air saturation; agitation rate was fixed at 90 rpm throughout the culture. Temperature was maintained at 37 °C and pH was set and controlled at 7.2 using 0.5 M sodium hydroxide and CO_2. *In situ* spectral scanning of bioprocess culture media was carried out with a NIR transflectance probe with a 1 mm pathlength (Precision Sensing Devices, MA). The sterilisable probe was connected to an Antaris II spectrometer (Thermo Scientific, USA). Each NIR spectrum corresponded to an average of 128 scans with an 8 cm^{-1} resolution from 4,000 to 10,000 cm^{-1} (2,500 – 1000 nm). Off-line concentrations of glucose, lactate, glutamine and mAb were determined using enzymatic kits (Roche 06681743001, 07395655001 for mAb and glutamine respectively; Thermo Scientific 981780, 984308 for glucose and lactate respectively) with an automatic spectrophotometer (Thermo Scientific GALLERY) against external standards. Viable cell density (VCD) was calculated by the trypan blue dye exclusion technique using an automatic cell counter (Beckman Coulter, Vi-CELL).

Development of calibration models (PLSR, SVR & ANNR)

A main set of calibration, consisting of 168 spectra, was acquired from bioreactor cultures, and divided into a calibration set (135 samples) and an internal validation set (33 samples).

The main calibration set, used for model development, was partitioned using the Kennard-Stone algorithm. The internal validation set was mainly used to analyse generalization of ANNR models by performing an analogous process of cross-validation. Calibration models for VCD, glucose, lactate, and glutamine were generated and spectral pre-processing for PLSR models was as simple as possible seeking conservation of maximum information contained. The main strategy was to normalise (MSC, PQN, SNV) and/or filter (derivative, EMSC) spectra and once a promising combination of spectra pre-processing was determined, only slight tuning was performed based on model performance so that predictive power of the models was enhanced. Predictive power was evaluated by the Root Mean Square Error of Cross-Validation (RMSECV). Determination of latent variable (LV) number for PLSR models was based on the goodness of estimation (Q2Y): the minimum number of LVs was obtained when Q2Y ceased to improve, using a venetian-windows cross-validation approach. The particular spectral pre-processing technique used for each compound with PLSR models was also used for all other regression methods for comparison of models to be mainly based on regression methods.

For the SVR models, the kernel was based on Radial Basis Function and optimization for gamma and epsilon values was also performed using a venetian-windows cross-validation approach. In the case of ANNR, the network size was firstly evaluated without cross-validation (100 iterations) for inferring the minimum number of neurons to be used in regressions, then the nature of the network structure was assessed. Early-stopping (maximum 20 training iterations using backpropagation with tanh as an activation function at 0.125 learn rate) was used to avoid overfitting and the internal validation set of 33 samples was used to validate the generalization power of the nets as a cross-validation procedure. The selection of the particular structures of ANNR and SVR models were those with the lowest RMSECV. The best model of each regression method (PLSR, ANNR and SVR) for each compound was retained for further evaluation.

Multivariate calibration models and statistical analysis were performed in MATLAB® (Statistics and Machine Learning Toolbox™, MATLAB R2016a, The MathWorks, Inc., Natick, Massachusetts, United States). Chemometrics software was also used (PLS_Toolbox® 8.2.1, Eigenvector Research, Inc., Manson, WA, United States).

Comparison of model performances

Comparison of models was based on requirements proposed by regulatory agencies for validation of NIR quantitative analytical procedures, such as accuracy, precision, linearity, specificity, range of operation and robustness. Because of such different structures of PLSR, ANNR and SVR models, a comparison was addressed in qualitative and quantitative approaches. Quantitative comparison of models comprised direct comparison of accuracy, precision and linearity. Qualitative approaches were undertaken to analyse the specificity of models and particularly for evaluating the performance of models under similar conditions expected during real-time monitoring. Discussion of some aspects of models, such as range of operation and robustness, is undertaken for the two types of comparison, particularly during qualitative analysis. In accordance to Anscombe (1973) quantitative analysis was focused on numerical calculations and qualitative analysis on graphical analysis of model performances. Finally, models were evaluated using a batch culture with different culture media (ActiPro (HyClone)) which was completely independent from the calibration process. For quantitative comparison, accuracy was measured as the RMSECV and statistically compared (One-way ANOVA with a post-hoc Tukey test) in terms of the absolute differences between the predicted and the real concentration values. Precision was measured as the mean of residuals (MoR) during calibration and statistically analysed in terms of homogeneity of variance of the residuals using multi-way Levene's test. Linearity was measured as the correlation coefficient (R_2) of the calibration plot and statistically evaluated in terms of correlation of the residuals from the calibration plot using the Durbin-Watson test.

Table 4.3-1. Methodological approach for comparison of models

Quantitative comparison	Qualitative comparison
- Accuracy: Measured as the RMSECV and statistically analysed by One-way ANOVA with a post-hoc Tukey test - Precision: Measured as the mean of residuals and statistically analysed by the variance of residuals using multi-way Levene's test - Linearity - Measured as R^2 and statistically analysed by the correlation of resdiuals using Durbin-Watson test	- Specificity of models: Detection of abnormal glucose spiking in calibration samples - Inter-batch heterogeneity effect: Comparison of the performance during real-time monitoring of three batch cultures for glucose - Performance under normal conditions: Analysis of model performances using a batch with mean inter-batch heterogeneity - Robustness: Analysis of model performances for real-time monitoring using a batch culture with a different culture media (HyClone ActiPro).

The three batch cultures were used for analysing the effect of inter-batch heterogeneity on predictive power of models. The batch culture with glucose spiking was used to evaluate the specificity of models to identify and relate glucose NIR signal within other NIR confused signals. The batch with mean inter-batch heterogeneity was used as internal validation, considered under usual operational conditions. Finally, an independent-of-calibration batch culture with a different culture media (ActiPro) was used to perform external validation and to discuss some aspects of model robustness. The general approach for analysing the model performance is summarised in Table 4.3-1.

4.3.1.1.4 Results and discussion

Spectra analysis

Analysis of spectra revealed noisy response caused by optic fibre noise (Clavaud et al. 2013) for the 2500 nm to 2550 nm range, which corresponded to the first NIR spectral variables and thus, this spectral section was not used for regressions. The presence of additive, multiplicative and wavelength-dependent effects due to scattering was evaluated within the calibration set spectra. The most common techniques to eliminate undesired spectra variations caused by light scattering (MSC, PQN, SNV, EMSC, derivatives) were evaluated. The standard deviation per each wavelength was used to elucidate the effect of scattering on calibration spectra. Raw spectra contained great variation within all the wavelength range, likely linked to multiplicative effect (offset of spectra). Analysis also revealed some scattering effects such as additive effect (baseline shift), and a likely wavelength-dependent effect (Rinnan et al. 2009). All spectra pre-processing techniques reduced the standard deviation caused by scattering and some of its effects. However only Extended Multiplicative Signal Correction (EMSC) (Martens et al. 2003) was capable of limiting the wavelength-dependent effect, though limiting the variance between spectra, which may compromise predictive model performance. Models based on PLSR are particularly sensitive to multiplicative effects (Martens et al. 2003) and proper spectra pre-treatment is essential for developing accurate calibration models. Final spectral pre-processing is shown in Table 4.3-2.

Table 4.3-2. Spectral pre-processing of models

Compounds	Spectral pre-processing
Viable cells	MSC + 1st derivative (15, 2,1)
Glutamine	EMSC
Glucose	MSC
Lactate	EMSC + 1st derivative (15, 2, 1)
mAb	EMSC
MSC: Multiple Scatter Correction; SNV: Standard Normal Variate; Abs: Absolute value; OSC: Orthogonal Scatter Correction; Values for derivatives: filter width, polynomial order, derivative order	

Decomposition of spectra by PCA allowed the analysis of data using only few spectral variables in terms of principal components. The first principal component (PC) explained 77 % of the spectral variance, and the second PC 9 %. This analysis revealed the evolution of the culture process in the PC space, also called process trajectory (Henriques et al. 2009; Clavaud et al. 2013) as shown in Figure 4.3-1.

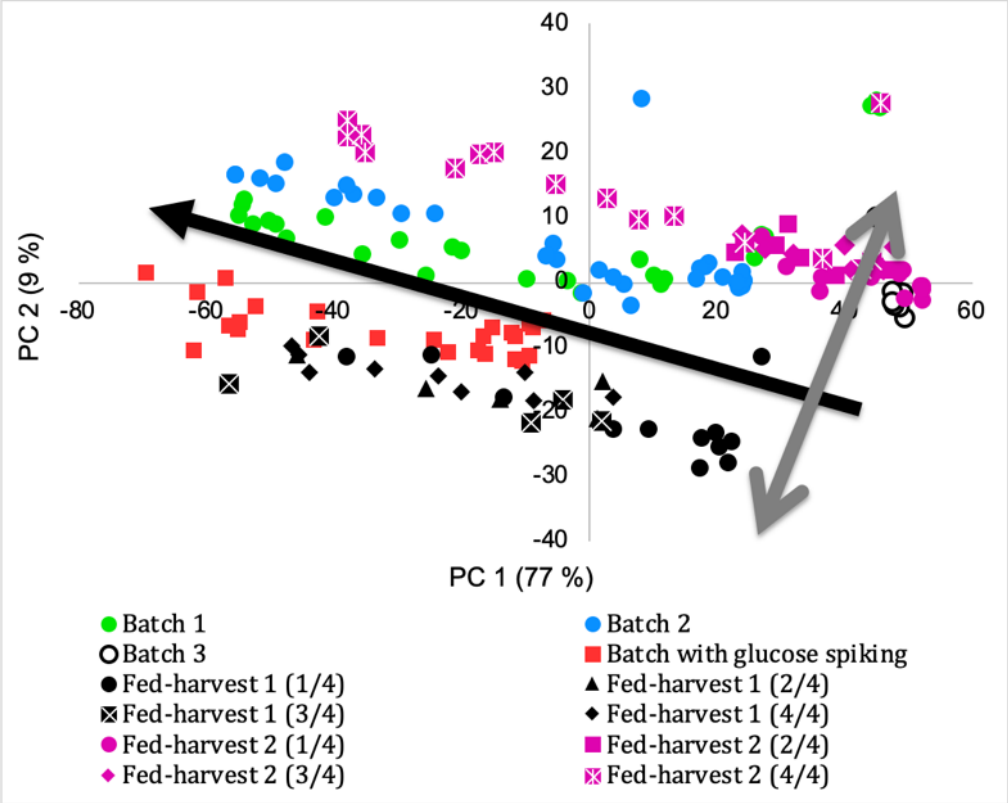

Figure 4.3-1. Calibration spectra set analysis

The main direction of the process was explained by PC1, which showed that as the cultures developed, spectra in the PC space go from right to left as shown by the dark arrow in Figure 4.3-1. Inter-batch variability is mainly represented by PC2 in the way that batch trajectories remained similar with only different offsets in the PC2 axis (grey arrow in Figure 4.3-1). PCA was also used to infer the nature of the relationship between spectra in the PC space and concentration as shown in Figure 4.3-2. Strong non-linear relationships were observed for ammonium ions and glutamine for all the concentration frame with hyperbolic and exponential profiles respectively.

There were compounds with a major linear relationship for a large concentration frame such as viable cells and glucose, but also for mAb and lactate to a lesser extent. For these compounds, nonlinearity was observed at relatively low and high concentrations, corresponding to the end of cell culture processes. For instance, when viable cell concentration surpassed $60x10^5$ cells.mL^{-1}, there was no clear linear relationship between spectra (in terms of PC 1) and concentration. This phenomena was even more marked for lactate and mAb, where an obvious linear relationship is completely lost when concentration surpassed 10 mM and 200 mg.L^{-1}, respectively.

The case of glucose and glutamine is perhaps the mostthought-provoking, where a linear response is observed at the same time as no response. In glucose profiles, there are three main obviously linear profiles, separated likely due to inter-batch heterogeneity. Even at low concentrations such a linear relationship is preserved. However, once concentration goes above 10 mM, there are samples corresponding null concentration (0 mM) with PC values above to 0, such PC values also including samples for concentration values up to 20 mM. Similar phenomena were observed for the mAb and glutamine relationship between spectra and concentration.

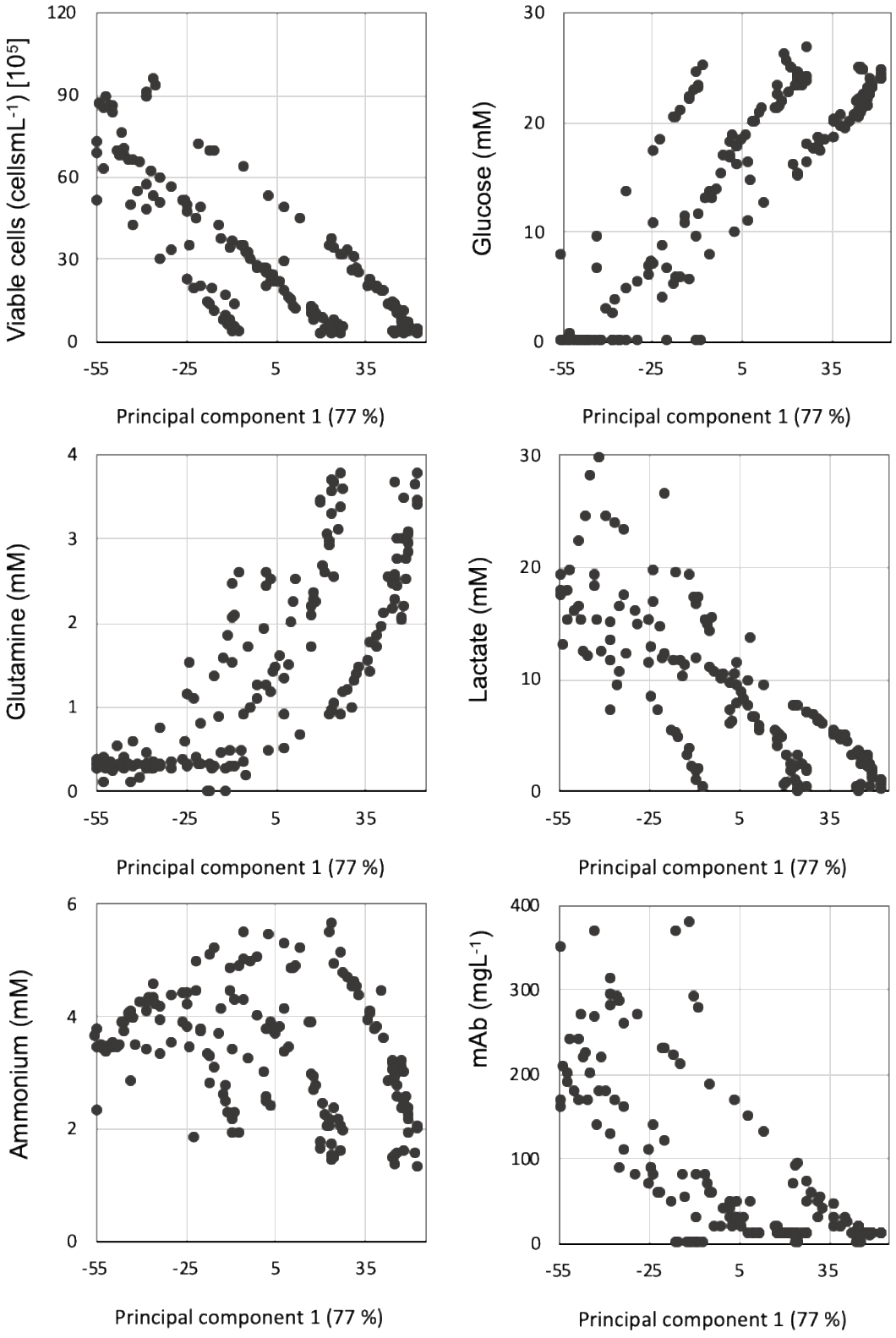

Figure 4.3-2. Main relationship between spectra and concentration

The parabolic and exponential profile relationship for ammonium and glutamine, respectively, may be a strong challenge for PLSR modelling. The nonlinear frames for all other compounds also represent a major challenge for linear calibration models. Therefore, calibration based on nonlinear regression techniques are likely required for properly management of calibration spectra within the whole concentration frames. Although it is common to make general assumptions of non-linearity based on the physical properties of the analysed matrix, data suggested that there may be inherent nonlinear relationships between spectra and some compounds concentrations, which may be independent of the physical properties of the matrix. This is evident since nonlinear relationships are different depending on the compound analysed. Therefore, treatment of all variables under the same approach seemed inappropriate to generate accurate calibration models. Consequently, two main regression techniques were used for addressing nonlinearity, SVR which is a sample-based method and ANNR which is a variable-based method. The main focus was on classic compounds (glucose, viable cells, glutamine and lactate) since mAb monitoring has been analysed separately in 4.4 Chapter IV (4.4.1.1).

Quantitative comparison

Direct comparison of model performances is summarized in Table 4.3-1. Accuracy, precision and linearity were measured through different perspectives. Accuracy, as the RMSECV is intended to depict the distance between actual and predicted concentration. For all compounds, nonlinear models achieved lower RMSECV than those by PLSR models, which suggested the limited performance of PLSR for cell culture monitoring. However, statistical analysis by one-way ANOVA with Tukey test revealed no significant difference. Two group means are significantly different if their intervals are disjoint and intervals overlapped for all variables using the three different regression models. This fact explained why even promising results using nonlinear models, are statistically equal to PLSR performance.

Table 4.3-3. Statistical analysis of PLSR, SVR and PLSR models

	Accuracy (RMSECV)	Precision (MoR)	Linearity (R_2)	Model structure
Glucose	PLSR: 3.377^{A} SVR: 2.29^{A} ANNR: 1.321^{A}	PLSR: 2.08^{C} SVR: 0.002^{A} ANNR: 0.012^{B}	PLSR: 0.85^{NL} SVR: 0.93^{L} ANNR: 0.98^{L}	PLSR: 5 LV SVR: 135 SV ANNR: 4 N
Lactate	PLSR: 3.5419^{A} SVR: 2.93^{A} ANNR: 3.23^{A}	PLSR: 1.462^{C} SVR: 0.622^{A} ANNR: 1.005^{B}	PLSR: 0.75 ^{NL} SVR: 0.83 ^{L} ANNR: 0.78 ^{NL}	PLSR: 5 LV SVR: 97 SV ANNR: 4 N
Glutamine	PLSR: 0.670^{A} SVR: 0.426^{A} ANNR: 0.43^{A}	PLSR: 0.270^{B} SVR: 0.16^{A} ANNR: 0.228^{B}	PLSR: 0.68 ^{NL} SVR: 0.87 ^{L} ANNR:0.87 ^{NL}	PLSR : 6 LV SVR : 131 SV ANNR : 5 N
Ammonium	PLSR: 0.8236^{B} SVR: 0.535^{A} ANNR:0.536^{A}	PLSR: 0.156^{C} SVR: 0.033^{A} ANNR: 0.18^{B}	PLSR: 0.45 ^{NL} SVR: 0.78 ^{L} ANNR:0.70 ^{L}	PLSR : 4 LV SVR : 135 SV ANNR : 4 N
Viable cells	PLSR: 6.88^{A} SVR: 7.18^{A} ANNR:5.29^{A}	PLSR: 3.287^{C} SVR: 0.253^{A} ANNR: 0.549^{B}	PLSR: 0.93 ^{NL} SVR: 0.92 ^{L} ANNR: 0.97 ^{L}	PLSR : 5 LV SVR : 128 SV ANNR : 2 N

Different letters as exponents indicate statistically significant differences between groups for accuracy and precision. Exponents for linearity indicate if models predict concentration linearly to actual concentration (L) or a non-linear deviation during prediction is detected (NL). Accuracy was measured as the RMSECV and statistically compared (One-way ANOVA with Tukey test) in terms of the differences between the predicted concentrations and the actual concentrations. Precision was measured as the mean of residuals (MoR) and statistically analysed in terms of homogeneity of variance of the predicted residuals using multi-way Levene's test. Linearity was measured as the correlation coefficient (R_2) of the calibration plot and statistically evaluated in terms of correlation of the residuals using the Durbin-Watson test.
LV: Latent variable, SV: Support vector for SVR models, N: Neurons for ANNR models; Units for RMSECV, RMSEP and SEP are the same: (cells.mL^{-1}) x10^{5} for viable cells, and mM for glucose, lactate, glutamine and ammonium.

In contrast to accuracy, which describes the distance between actual and predicted values, precision describes the variation on predicted values when the same (or extremely similar) sample is measured by the same model. The nature of calibrating using samples from heterogeneous cell culture processes is a great constrained to acquire data for precision analysis. Accuracy and precision are often analyzed using the same data (differences between actual and estimated values) rather than acquiring particular data for proper precision analysis (repeated measurements). Therefore, accuracy and precision are confused and then precision

has not received focused attention in cell culture monitoring studies. Then, it was proposed to analyzed precision by homoscedasticity analysis using the variance of the residuals.

The residuals for a particular compound were calculated using a particular calibration model based on either PLSR, SRV or ANNR. Then their distribution was computed and analyzed. If the variance of the residuals has the same distribution, then it is likely that PLSR, SVR and ANNR models have the same precision. Contrary, if models differ on distribution, it is clear that those with narrow distributions are likely more precise. The differences between those distributions were statistically analyzed by Leven's test. These analyses revealed that SVR and ANNR are likely more precise than PLSR as shown in Table 4.3-3. Moreover, SVR is slightly more precise than ANNR. This could be explained by the fact that SVR is a sample-based regression method, which could have also managed information of the cell culture progression and thus performed better than models based on ANNR.

Although R_2 is a parameter that shows how much variability is explained by the model, it is usually used as linearity term using the calibration plot data. Relative higher R_2 coefficients were achieved when both ANNR and SVR were used for generation of calibration models in contrast to PLSR. However, higher R_2 values do not guarantee that models predict concentration linearly to actual concentration. Analysis of correlation of the residuals revealed that there was correlation of the residuals for all compounds when using calibration models based on PLSR.

The use of SVR and ANNR for generation of calibration models improved the performance. In general, those nonlinear regression methods leaded models to linearly predict concentration based on actual concentration. However, this was not the case for particular compounds such as glutamine and lactate when ANNR was used. In general terms, SVR seemed not only as the more accurate and precise regression method for building calibration methods but was also capable to predict concentration linearly to actual concentration using spectra.

Qualitative comparison

After quantitative analysis, the models were evaluated qualitatively for analysing their performance during real-time monitoring. First, the effect of inter-batch heterogeneity on

glucose calibration model generated with PLSR was evaluated using three different batches. Then, the specificity of models was analysed using the batch with glucose spiking. For analysing the performance of models during routine monitoring, special attention was focused on batch with mean inter-batch heterogeneity since it is likely to depict the majority of cultures. Finally, a fed-batch culture totally independent from calibration, was monitored and analysed for discussing some robustness issues.

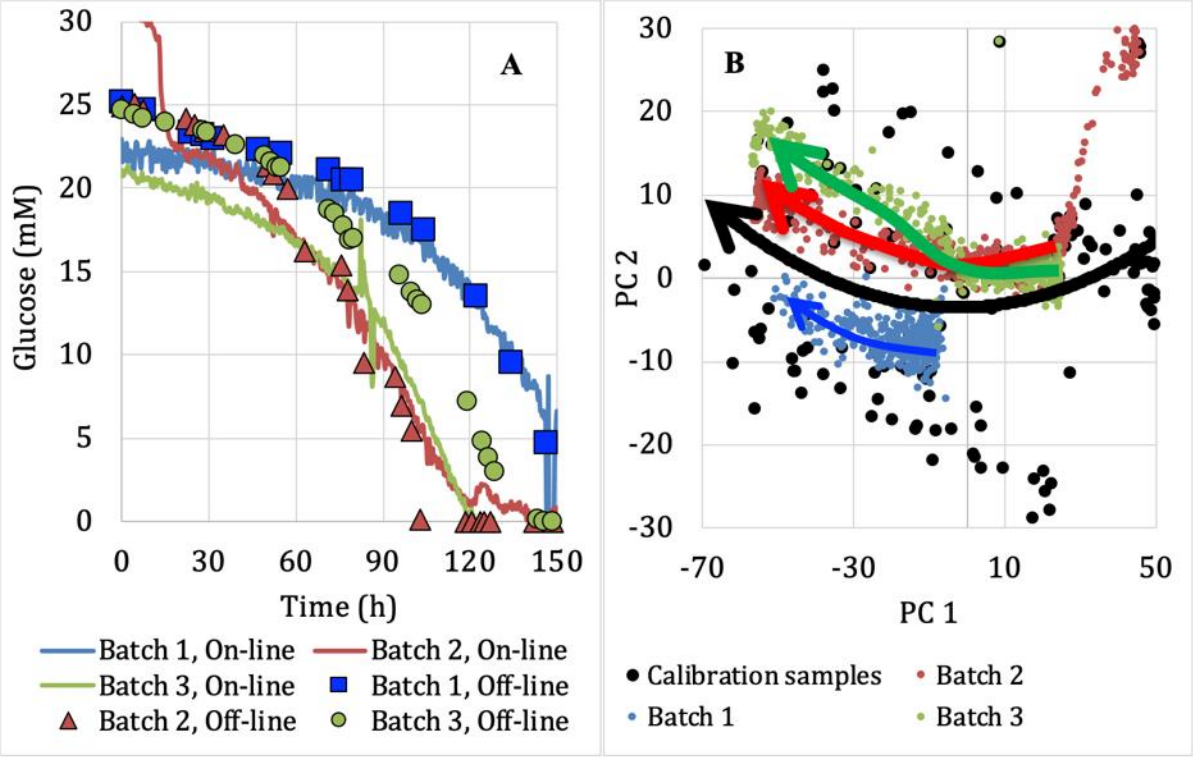

Figure 4.3-3. Effect of inter-batch heterogeneity on PLSR glucose model: A, performance of PLSR model for real-time monitoring of glucose concentration, B, inter-batch heterogeneity analysed by PCA

The three batches shown different glucose predicted profiles as shown in Figure 4.3-3-A. For instance, batch 3 has a marked offset during the whole culture while batch 1 and 2 had some frames with accurate predictions and some other frames with miss predictions, particularly at the beginning and at the end of the culture. The effect of inter-batch heterogeneity on prediction performance was then analyzed using PCA Figure 4.3-3-B. Though calibration set

is composed of several cell cultures with particular trajectories, a global trajectory may be obtained when considering all cultures within the calibration set, which is depicted in the dark arrow of Figure 4.3-3-B. It is worth to note that when a particular batch trajectory agreed with global trajectory, there are accurate predictions (Figure 4.3-3-A) as those for batch 1 and 2 in the frame within the blue and red arrows respectively (Figure 4.3-3-B), even if there are a strong offset from global trajectory. Contrary, there are mis predictions when the trajectory of a particular batch does not agree with the global trajectory. This phenomenon can be explained by the fact that a trajectory is a relationship between PC1 and PC2, which is then used for performing regression. If such relationship between PC1 and PC is not conserved, there will be mis predictions as those for Batch 3 (Figure 4.3-3-A).

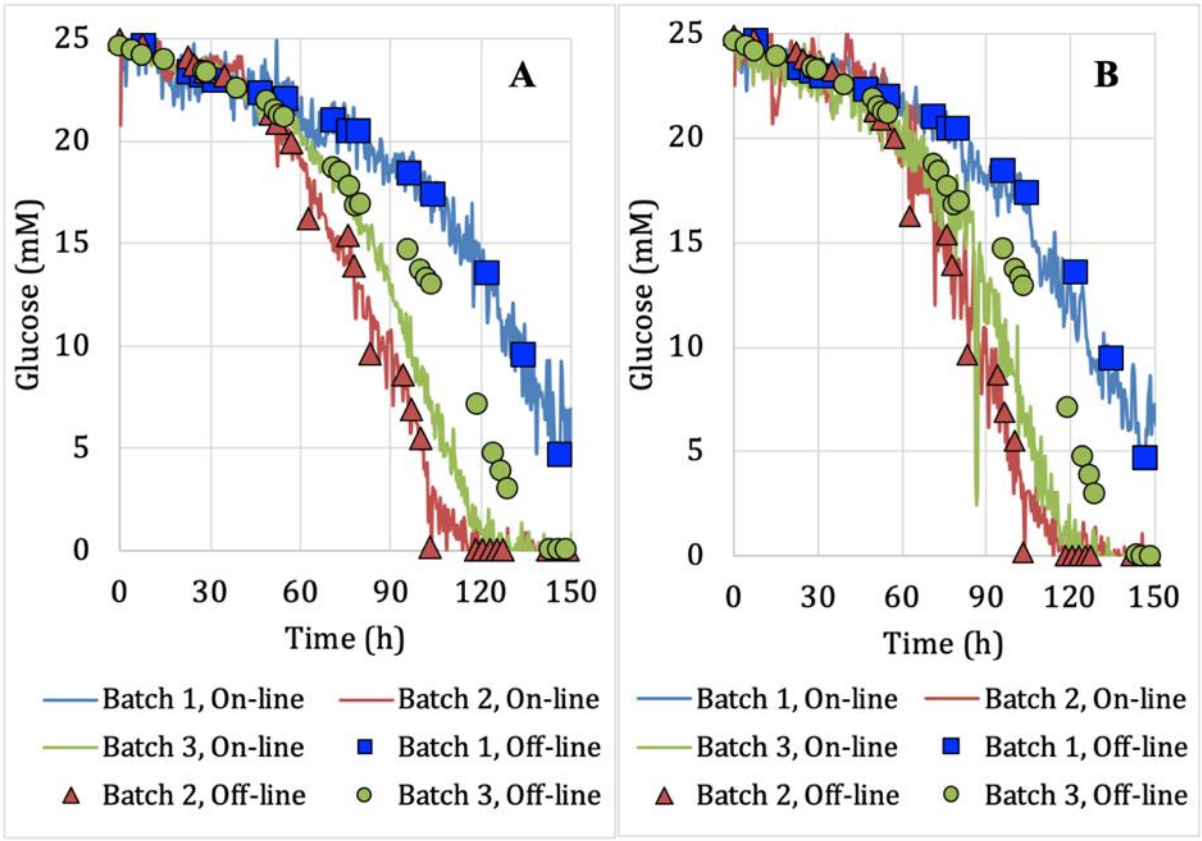

Figure 4.3-4. Effect of inter-batch heterogeneity on SVR and ANNR models: Real-time monitoring of glucose concentration by SVR (A) and ANNR (B) models

Preparation of the three batches was focused on repeatability and no special change on neither inoculation nor culture media was intended. Therefore, results have shown that PLSR is particularly vulnerable to inter-batch heterogeneity. Therefore, the new implementation of nonlinear regression methods was evaluated for the generation of calibration methods. The on-line profiles for SVR and ANNR were closer to off-line concentration data (Figure 4.3-4) than those for PLSR (Figure 4.3-3). Major improvements were observed for Batch 1 and 2 at the beginning of cultures, where glucose concentration was properly predicted. Inter-batch heterogeneity, which strongly affected Batch 3 using PLSR (Figure 4.3-3), was more efficiently managed by calibration models based on either SVR or ANNR. For instance, while the on-line concentration profile by PLSR was completely offset for all the culture, profiles by ANNR and SVR models were properly fitted to off-line profiles during the first 90 h of culture process. However, the effect of inter-batch heterogeneity still affected prediction power during the last part of the culture.

Prediction power by models are affected by inter-batch heterogeneity since models are not totally theorical but rather partially empirical. This means that prediction is based in confused and global changes in cell culture media rather than particular changes on glucose vibrational NIR movements. Inter-batch heterogeneity or variability can then be understood as changes in the composition of the cell culture processes that do not match with the variability pattern constructed by the calibration method. Enhancement of prediction power in inter-batch heterogeneity conditions would necessary require that models be highly specific for the compound of interest. The more specific the model is to the analyte of interest, the less the model will depend on the pattern of confused relationships in the cell culture media. Therefore, the specificity of models was evaluated using the batch with glucose spiking.

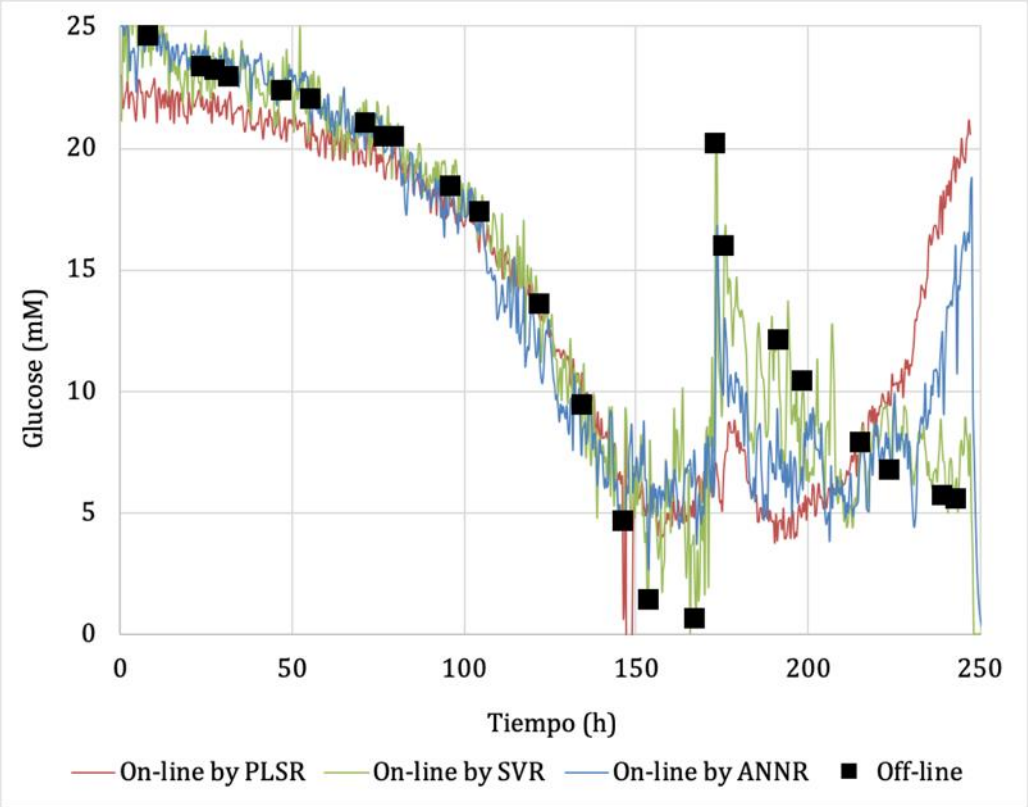

Figure 4.3-5. Specificity of PLSR, SVR and ANNR models for real-time monitoring of glucose concentration

The specificity of models was evaluated using the batch culture with glucose spiking. This batch was operated as other batches, but glucose was added at 173 h which completely interfered with the normal nature of batches and thus, the normal confused pattern of cell culture media used by calibration models to predict concentration. If a model is highly specific for glucose, then it would perform prediction more likely based on glucose signals rather than on confused pattern of cell culture media.

The performance of models for the batch with glucose spiking is shown in Figure 4.3-5. For the frame operated as normal batch (0 h – 172 h), the three models performed relatively good. However, once the glucose was spiked into the bioreactor the performance of models greatly differed. Addition of the concentrated glucose solution into the bioreactor caused an increase on concentration up to 20 mM, only the model based on SVR properly predicted such concentration, while both models based either on ANNR or PLSR showed a limited increase on glucose concentration of 16 and 7 mM respectively. Although the three models showed

later a decreasing profile which matched in nature off-line concentration profile, only SVR was close to actual values while predictions by PLSR and ANNR sub-estimated concentration. In addition, both PLSR and ANNR predicted an increase of glucose concentration after the 200 h, which is totally contradictory to actual decreasing tendency. For explaining this behavior, PCA was used.

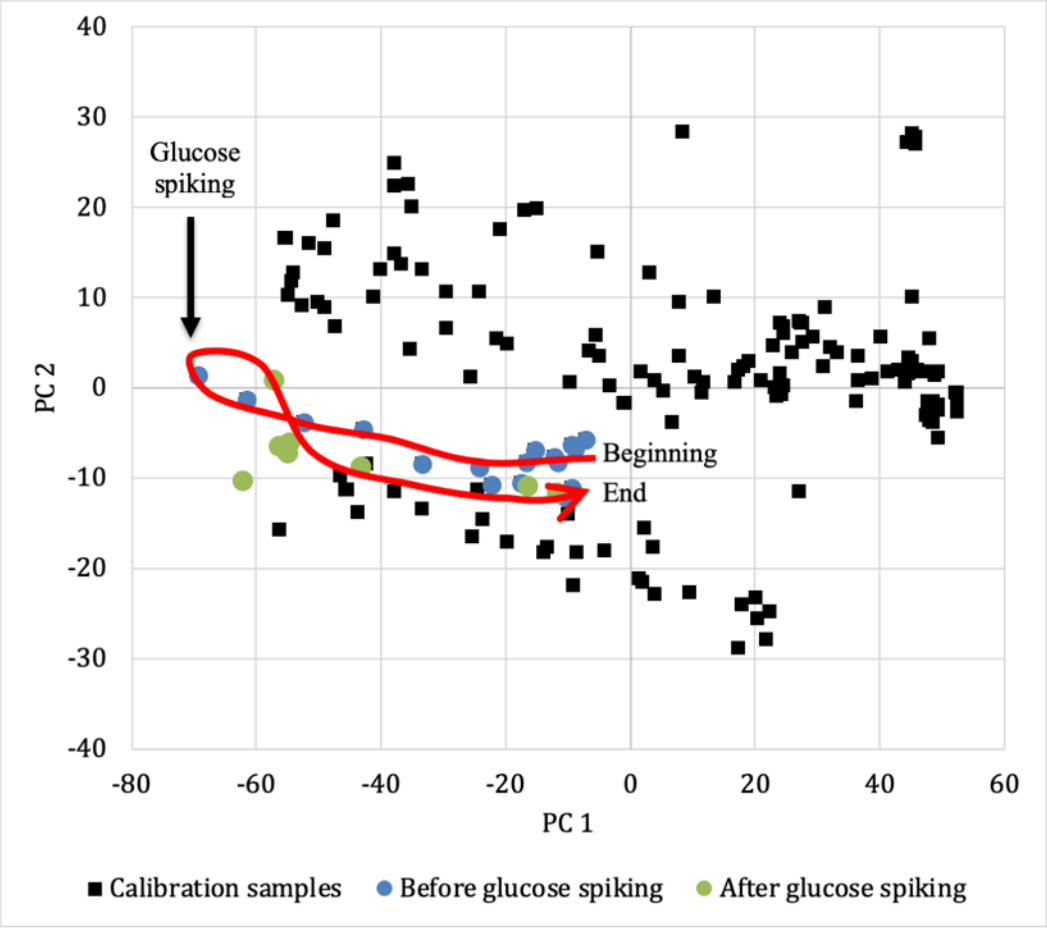

Figure 4.3-6. PCA of the batch with glucose spiking

The fact that model based on PLSR predicted increase of glucose concentration rather than the actual decrease, can be explained by the fact that addition of glucose strongly changed the trajectory of the culture process as shown by the red arrow in Figure 4.3-6. PCA had previously shown (Figure 4.3-1) that the progression of the culture was mainly explained by PC1 in a right to left direction, which agree with the decrease on glucose concentration. In

this context, while the trajectory of the batch is in agreement with global trajectory (blue circles in Figure 4.3-6), the model rightly predicted glucose concentration. Glucose spiking changed the trajectory of the culture, though the direction had no left tendency as expected. It changed into a right direction, as if the progression of the culture went backward up to states similar to those of the beginning of cultures where high concentrations of glucose are expected. This revealed the lack of specificity of the PLSR model, which performed prediction using mainly global changes of culture media. The lack of specificity for the model based on ANNR may be caused by the fact that calibration samples of this particular condition (glucose spiked culture) comprised only a negligible fraction from the calibration sample set. Thus, generalization of glucose information was mainly based on the majority of samples. Prediction of glucose concentration by SVR was likely enhanced by the fact that it is a sample-based regression method, which first detected abnormality in those few calibration samples with glucose spiked nature, separated them in the feature space and then relate them to concentration using a nonlinear relationship. This nature of SVR is likely advantageous as it could properly depicted the off-line concentration profile. However, bounce was important and thus accuracy and precision were limited, which could eventually compromise further control strategies.

The models were used for real-time monitoring using the batch culture with mean inter-batch heterogeneity. Comparison for viable cells, glutamine, lactate and ammonium concentration is shown in Figure 4.3-7. For viable cells, ANNR and PLSR did not predict concentration properly since a both models sub-estimated cells concentration at the stationary phase, only the model based on SVR properly predicted viable cell concentration. On the other hand, both ANNR and SVR accurately predicted glutamine concentration while the model based on PLSR mis estimated concentration at the beginning and end of cultures, likely due to extrapolation phenomena as discussed in chapter 4.2.

The lactate profile by the three methods were likely the same though PLSR and ANNR showed more bounce, during the last part of the culture the three models failed in accurately predict lactate concentration. Cell viability during this part of the culture was low and diverse intracellular compounds as well as cell debris were in the culture media, which could have limited generalization of models. This statement is in agreement to results from chapter II, where also LWR had limited predicted power in such conditions. In this context, LWR was

likely a better option even over SVR since LWR fitted data with mean errors of stochastic data rather than finding data using a nonlinear model such as SVR, which then predicted stochastically with new unseen data.

For monitoring ammonium ions concentration, the model based on PLSR completely failed at the beginning of the culture, while models based on SVR and ANNR accurately predicted concentration in this initial frame. Between 45 h and 60 h of culture, there was an abrupted concentration increase that was not properly predicted by any model. Only SVR predicted some actual concentration properly though the majority of prediction sub-estimated actual concentrations. Prediction by the three models was enhanced after the 90 h of culture, where bounce was more evident for PLSR models and limited for ANNR and particularly for SVR.

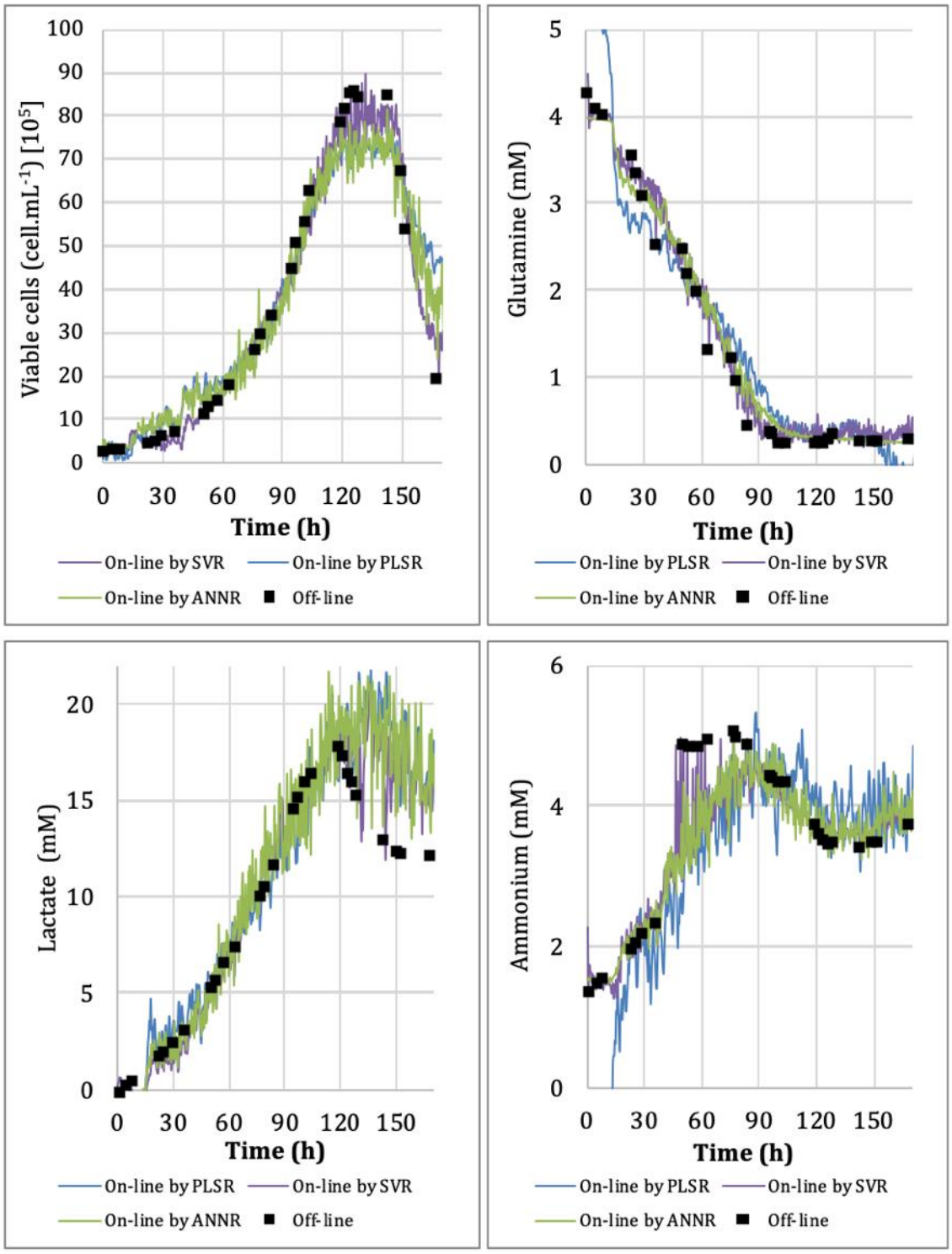

Figure 4.3-7. Comparison of PLSR, SVR and ANNR model performances for real-time monitoring of diverse compounds concentration

As calibration models are partially empirical, evaluation of robustness could then also be considered empirical and interpretative in manner. The performance of models based on PLSR, SVR and ANNR for glucose monitoring in a Batch culture with ActiPro culture media is shown in Figure 4.3-8. The batch culture first consisted in a batch culture and then glucose was spiked at 120 h. The three models predicted this binomial nature though with different closeness to off-line profile. The model based on ANNR had the best performance since it accurately predicted glucose during the first part of the culture (0 h – 60 h) and accurate trends were observed until glucose depletion, then it detected the increase on concentration though not accurately. The model based on PLSR had a similar performance though with a larger offset, sub-estimating actual concentration. After glucose spiking, the PLSR did not predict the abrupted increase on glucose concentration and instead, it predicted a decrease on concentration. On the other hand, SVR had likely the worst performance during the first part of the culture with a large offset from actual concentration profile. However, during the last part when glucose was spiked into the culture, it predicted a concentration increase as the model based on ANNR.

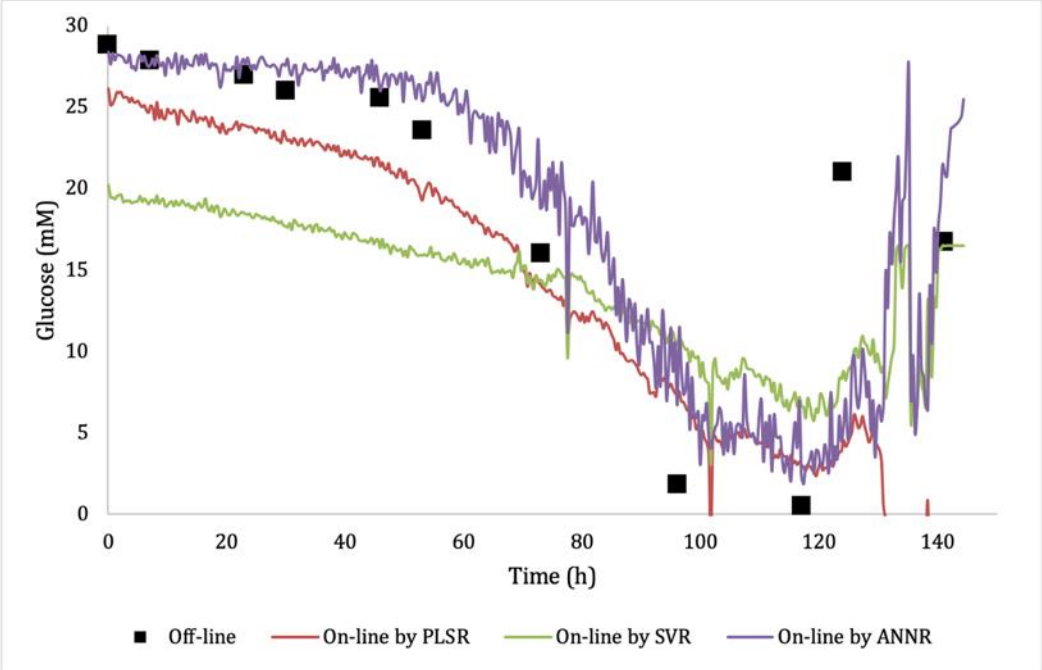

Figure 4.3-8. Robustness of PLSR, SVR and ANNR models to monitor under different culture media condition

4.3.1.1.5 Conclusion

Former analysis of model performances suggested that PLSR was vulnerable due to lack of specificity and una inability to handle inter-batch heterogeneity. However, analysis of the batch culture with ActiPro culture media revealed the PLSR model as more accurate than SVR. This could be explained by the fact that PLSR was more likely based on information about cell culture progression rather than specificity. As the direction of the culture progression using ActiPro also matched that from calibration, PLSR could predict the profile though with a strong offset. Once glucose was spiked, spectra rested outside the calibration space and PLSR was not able to properly perform prediction in extrapolation conditions, which explained why it predicted decrease on concentration when it actually increased. On the other hand, SVR had been shown to have more specific to glucose and it was less vulnerable to inter-batch heterogeneity. However, it is likely that such high specificity for glucose in the normal culture media was based on information of glucose couple to some particular compound(s) that was not as abundant in the ActiPro culture media as in the normal culture media. The model based on ANNR was likely to generalized specificity for glucose better in contrast to SVR, which only generalized using some samples rather than from all calibration samples since SVR is a sample-based regression method. Thus, for robustness issues ANNR is likely the best option if such great variations on culture media are intended during cell culture processes.

4.3.2 Conclusions of chapter III

In this chapter it has been shown that the nature of nonlinear relationship between spectra and concentrations is particular for each compound of interest. This fact revealed the need of different strategies for better management of nonlinearity according to particular compounds. Therefore, SVR and ANNR, two nonlinear regression methods for the generation of calibration models, were evaluated for fitting such nonlinearities. Then, the performance of nonlinear calibration models was compared against models generated using PLS as regression method.

Results from calibration data showed that accuracy between the three regression methods are statistically the same. However, the nonlinear regression methods demonstrated significance enhanced capacity for building calibration models for cell culture monitoring, particularly by improved accuracy and precision. During real-time monitoring analysis, the models based on PLSR showed limited performance, particularly for management of inter-batch heterogeneity and lack of specificity. These facts suggest that such linear models would likely fail in processes under the conventional QbD approach, where changes in the production processes and inter-batch heterogeneity is highly expected.

The novel use of SVR and ANNR showed to be a promising alternative to maximize the potential of NIR spectroscopy as monitoring tool. Both regression models over performed PLSR in normal operating conditions. SVR is likely the best option since it better managed inter-batch heterogeneity (more accurate and precise predictions) due to their more specificity capacity. This enhanced performance could be explained by the fact that SVR is a nonlinear sample-based regression method. However, this nature could also be counterproductive in particular cases as for example changes in the culture media or process operation under abnormal operating conditions, where ANNR is likely the most robust method since it globally generalized the process dynamics. Overall, these results provided a wide frame to follow up new chemometric strategies which enhanced the capacity of *in situ* analyzers for the establishment of controlled cell cultures processes.

4.4 CHAPTER IV: INTO PAT CHALLENGES THROUGH NIRS

4.4.1 Introduction

During the last decades, regulatory agencies have encouraged biopharmaceutical industry for adopting Quality by Design and Process Analytical Technology for monitoring and control of Critical Process Parameters of the processes affecting Critical Quality Attributes of the products bioproduction processes. Ideally, monitoring tools should provide as much of process status information as possible in real-time manner. Then advanced retro control strategies based on process knowledge, may be launched as for controlling process variability and thus ensuring medicine quality. Although some basic CPP are systematically monitored (pH, temperature, among others), cell culture processes also require monitoring of CPP and CQA with biochemical and biological nature. Consequently, monitoring tools must be updated considering the needs of cell cultures processes. Process analyzers based on NIR spectroscopy have been proven promising monitoring tools when couple to *in situ* probes as they can provide multicomponent information in real-time manner. However, most of the work has been focused to some few biochemical compounds such as glucose, lactate and glutamine concentration, while monitoring of innovative parameters such as mAb glycosylation, or more functional parameter such as the cell physiological state, have remained practically unaddressed. Therefore, this chapter has explored the feasibility of *in situ* NIR calibration models for monitoring some of such innovative parameters.

This chapter proposed the development of calibration models for monitoring common heterogeneity within animal and plant cell suspension cultures. In CHO cell cultures, calibration models were generated for monitoring mAb variants which differently impact clinical effects. The monitoring strategy was based on monitoring total mAb, non-glycosylated mAb (mAb macro-heterogeneity) as well as the presence of particular sugar moieties within the glycan chain (mAb micro-heterogeneity). On the other hand, heterogeneity was addressed in terms of cell differentiation in plant suspension cultures of *Catharanthus roseus*, which is likely linked to the synthesis of anticancer molecules as demonstrated in 4.1 Chapter I (4.1.1.1.1). Results for this chapter IV have been published in Biochemical Engineering Journal and Bioprocess and Biosystems Engineering Journal for CHO cells and *Catharanthus roseus* cultures, respectively. Thus, the chapter is composed of both publications.

4.4.1.1 Support Vector and Locally Weighted regressions to monitor monoclonal antibody glycosylation during CHO cell culture processes, an enhanced alternative to Partial Least Squares regression

Daniel A. Zavala-Ortiz[1,2], Bruno Ebel[1], Meng-Yao Li[1], Dulce Ma. Barradas-Dermitz[2], Maria G. Aguilar-Uscanga[2], Patricia M. Hayward-Jones[2], Annie Marc[1], Emmanuel Guedon[1]

1- Laboratoire Réactions et Génie des Procédés, Université de Lorraine, CNRS, LRGP, 2 avenue Forêt de Haye, TSA 40602, 54518 Vandœuvre-lès-Nancy, France.
2- Instituto Tecnológico de Veracruz: Calz. M.A. de Quevedo 2779, Veracruz, Ver. Mexico.

4.4.1.1.1 Abstract

Since monoclonal antibodies (mAb) are sensitive to the manufacturing process, several mAb variants can be the result of a single batch production. The most critical source of heterogeneity is glycosylation which has a profound impact on safety and efficacy of the final product. Implementation of monitoring and control of the process using the Quality by Design (QbD) approach may help to ensure mAb specifications, although its implementation is limited by the availability of real-time specific measurements. All current approaches to elucidate mAb glycoforms require sampling and labour-intensive efforts. Thus, glycosylation analysis is often performed with the objective of detecting quality defects at the end of the culture process. In this work, the capability of Near Infrared spectroscopy and chemometric treatment to accurately monitor mAb glycosylation during CHO cells cultures using *in situ* probes is shown for the first time. Real-time monitoring of glycosylation, in terms of high mannose isoforms, fucosylated, sialylated and galactosylated isoforms as well as non-glycosylated mAb, has been successfully performed by the novel use of Locally Weighted Regression (LWR) and Support Vector Regression (SVR). These encouraging results open the way for the implementation of control systems on the impact of cell culture operating parameters on mAb heterogeneity, particularly glycosylation, during CHO cell culture processes through the QbD approach.

4.4.1.1.2 Introduction

Monoclonal antibodies (mAb) produced in animal cell culture processes represent a success in terms of clinical benefit for patients and revenue generated by biopharmaceutical industries. Such molecules are quite sensitive to changes in manufacturing processes and thus several mAb variants could be produced within a single batch due to post-translational modifications. Glycosylation is the main source of mAb variability which can strongly impact mAb clinical properties. Therefore, the control of glycosylation specific profiles of mAb during the process is critical for therapeutic efficacy and patient safety. For this purpose, regulatory agencies proposed the Process Analytical Technology (PAT) strategy to control pharmaceutical manufacturing processes through the continuous adjustment of Critical Process Parameters (CPP) which affect Critical Quality Attributes (CQA) of the product. Accordingly, continuous monitoring of CQA, such as mAb glycosylation, is required to establish advanced retro-control systems to guarantee mAb specifications (Teixeira et al. 2009a).

However, continuous monitoring of mAb glycosylation is challenging since it requires analyses at a relatively high cost. Complete mAb glycosylation analysis must include identification of sugar chains attached to glycosylation sites (macro-heterogeneity) and of the sugars moieties within the glycan chain (micro-heterogeneity). Indeed, mAb macro- and micro-heterogeneity analyses involve several steps, such as enzymatic digestion, labelling, derivatization and separation, followed by structural analysis, usually using mass spectrometry (MS) (Sinha et al. 2008). The complexity of such analyses implies significant delays, 5 hours to 2 days, mainly for the acquisition of process information, thus mAb quality analysis is usually performed at the end of the cell culture process (Rathore and Winkle 2009).

In recent years, intensive efforts have been made to establish PAT as the mean to monitor mAb glycosylation during cell cultures, particularly using automatic at-line or on-line traditional biochemistry approaches with the aim of reducing analysis times and sample volumes. Such approaches allow the presence of mAb glycoforms to be known in a question of hours (Burnina et al. 2013; Doherty et al. 2013; Henninot et al. 2015; Dong et al. 2016). In the last few decades, vibrational spectroscopy, in combination with multivariate analysis, has been proven to be a promising tool (Moscetti et al. 2019), particularly for cell culture

monitoring (Li et al. 2016). However, it is mainly restricted to some usual cell substrates and/or by-products such as glucose, lactate or glutamine (Arnold et al. 2003; Mattes et al. 2007; Henriques et al. 2009; Clavaud et al. 2013; Bhatia et al. 2018; Li et al. 2018b).

More recently, a study showed the possibility of monitoring mAb concentration by using either Raman or NIR spectroscopy in real-time during CHO cell cultures (Li et al. 2018a). Although Raman spectroscopy led to a slightly better estimation for mAb concentration, NIR spectroscopy showed a higher signal-to-noise ratio, though in more complex spectra. The inferior capacity of the NIR model was thus mainly attributed to the lack of linear PLS regression for handling complex NIR spectra, likely containing information in non-linear ways (Li et al. 2018a). However, as far as can be ascertained, glycosylation micro-heterogeneity monitoring has not been addressed. In this study, we showed that *in situ* NIR spectroscopy can be applied for in-line monitoring of mAb glycosylation micro-heterogeneity.

As alterations in mAb glycoform patterns may result in strong changes in clinical profiles, manufacturers must guarantee glycosylation specifications to ensure reproducible and consistent clinical performance. The occurrence of several factors that may compromise conservation of clinical profiles is not uncommon in the pharmaceutical industry (Schiestl et al. 2011), and so monitoring and control systems are required to ensure mAb properties.

Conservation of such properties is due to proper combination of glycoforms with different sugar moieties within the glycan chain. Indeed, it has been widely reported that the presence of fucose, galactose and sialic acid strongly affect antibody dependent cellular cytotoxicity (ADCC), complement dependent cellular cytotoxicity (CDC) and immune modulation of the mAb, respectively (del Val et al. 2010). On the other hand, high mannose glycoforms are reported to reduce serum half-life (Dwek et al. 1995). In this context, producing mAb glycoforms with resulting clinical effects similar to those of the reference mAb is critical for batch approval (Schiestl et al. 2011). In this study, it was shown that NIR spectra can exhibit an estimated correlation to non-glycosylated mAb and total mAb concentration (mAb macro-heterogeneity), as well as to glycoforms containing fucose, galactose and high-mannose structures, including sialic acid within the glycan chains (mAb micro-heterogeneity). Furthermore, this approach could have an immediate application using a NIR

spectrophotometer in at-line or off-line modes, which could provide mAb glycosylation information in question of minutes.

4.4.1.1.3 Theory

Spectra, particularly from NIR in-line analysers, are complex since both physical and chemical information is contained, usually in a highly collinearity way. Thus, multivariate calibration particularly for regression, is needed for correlating complex spectra to desired quality attributes or analyte concentration.

Though many regression methods are available for building calibration models, as far as it can be ascertained, only Partial Least Squares Regression (PLSR) has been addressed in cell culture monitoring. In this work we analysed the adequateness of PLSR for cell culture monitoring and explored the performance of other regression methods such as Support Vector Regression (SVR) and Locally Weighted Regression (LWR). The intuitive concept of models as well as their characteristics for regression in cell cultures are discussed and shown in Figure 4.4-1.

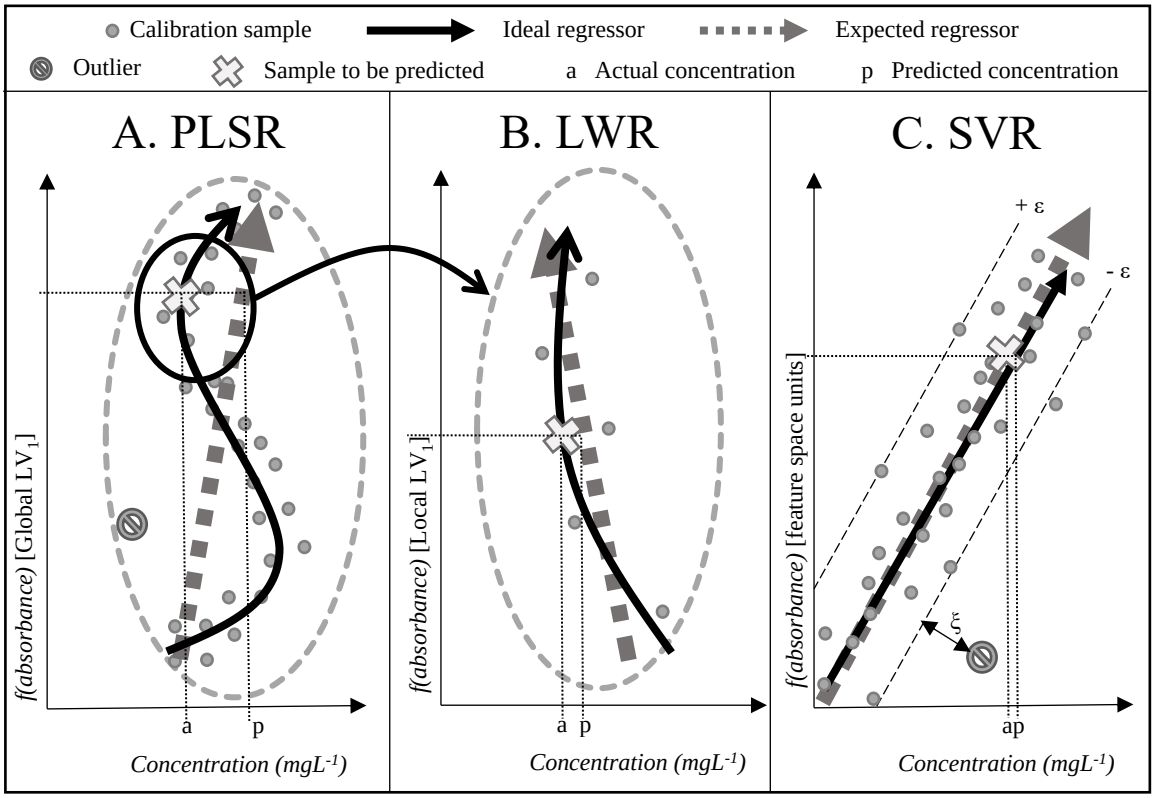

Figure 4.4-1. Intuitive concept of regression methods: PLSR creates a chemometric space in term of latent variables (LV) and then a linear regression is performed for estimating concentration. LWR enhance the fitting of non-linearities by performing local regression by weighted PLSR. In contrast to PLSR which generates the regressor by only minimising the error, SVR map spectra into a higher dimensional feature space as for fitting the error within a particular threshold (±ε) and then the regressor is generated.

Partial Least Squares Regression (PLSR)

The PLSR method is based in a reduction variable process in order to treat collinearity. Firstly the spectra (X) and the concentration (Y) matrix is decomposed as (Höskuldsson 1988):

$$X = TP^T + R_1 \ (1)$$

$$Y = UQ^T + R_2 \ (2)$$

where X and Y are spectra and concentration matrices respectively, T and U are the pseudo-scores matrices, P and Q are the pseudo-loadings matrices and R_1 and R_2 are the residuals

matrices. Matrix decomposition of X and Y matrices are not independent, thus an internal relationship between the scores of X and Y are generated accordingly:

$$U = BT \quad (3)$$

where U is the pseudo-scores of Y to be calculated, T the pseudo-scores of X and B the regressor matrix. PLS works with the constraint that these components explain as much as possible of the covariance between X and Y. Once the regressor matrix has been determined, calculation of y-concentration value from problem sample may be calculated as:

$$Y = T^{*}BQ^{T} + R_1 \quad (4)$$

Where T* is the pseudo-scores matrix of the problem sample, B the regressor matrix, Q_T the pseudo-loading matrix of the model and R_1 the residual matrix. Regression based on PLSR offers a relatively simple frame for analysing the relationship between spectral response and prediction by the model. However, it is particularly sensitive to scattering effects and may not properly handle information contained in a non-linear way. Therefore, for building quantitative calibration models using NIR spectroscopy, a common assumption has been focusing on chemical information and limiting the contribution of physical data contained in spectra by the use of spectral pre-treatment (Rinnan et al. 2009). However, manipulation of spectra for limiting scattering effects may also disturb chemical information (Huang et al. 2010) and thus a compromise must be adopted.

PLSR is a variable space-based regression method which calculates the relationship between each of the variables (absorptions at different wavelengths) and compound concentrations. Therefore, such a relationship should be relatively constant during the whole culture process in order to maintain accurate estimations. Though it is an obvious assumption, caution must be taken considering the strong physical and chemical variations of culture media during cell culture progression. Deviations from this assumption are usually observed as non-linear effects (Figure 4.4-1-A) and may limit the predictive capability of models. In such cases, the regression equation leads to a linear trajectory of predicted values (grey dotted arrow in Figure 4.4-1-A), limiting prediction on non-linear sections (cross in Figure 4.4-1-A). As far as can be ascertained, this fact has not been addressed yet in cell culture monitoring. Therefore, the novel use of space-based regression methods which firstly focus on affinity

and dissimilarity between samples (culture progression information), and secondly on the relationship between variables (absorptions at different wavelengths), has been evaluated.

Locally Weighted Regression (LWR)

LWR is mainly addressed for modelling complex relationships for which no theorical model may exist. In contrast to PLSR which generates a regression function considering all calibration points, LWR firstly compares the sample to be predicted with samples within the calibration set. Then only those calibration samples similar to the sample to be predicted are used (the local area) (black circle in Figure 4.4-1-A). Then each point of the local area is weighted according to its distance from the sample to predict: close points are given more importance or weight, far points are given less weight; then a regression function of the independent variables is generated employing the weights and in the local area (Cleveland and Devlin 1988) (Figure 4.4-1-B). Generation of accurate models then requires adjustment of key parameters such as similarity between samples, definition of the local area and the weights, and the nature of the regression itself.

Once spectra have been mapped into a chemometric space (principal component space, latent variable space, among others), the local area is determined by a distance function and specified limit. Since distance as a delimiting criterion may be inappropriate when lacking vast calibration samples in a wide calibration space, several authors have employed distance criteria in terms of near calibration samples (Næs and Isaksson 1992; Næs 2004). Then local calibration samples are weighted according to a weight function, such as the tricubic function (5). Then the weight for a calibration sample is calculated as:

$$W(u) = \begin{cases} (1-u^3)^3 & if\ u \leq 1 \\ 0 & if\ u > 1 \end{cases} \quad (5)$$

$$w_i(x_j) = W\left(\frac{\delta(x_j, x_i)}{d(x_j)}\right) \quad (6)$$

where

(x_j, x_i): Distance between prediction sample j and calibration sample i

$d(x_i)$: Maximum distance involved in each regression

As could be noted, the weights will be large (close to 1) for x_i close to x_j, and small (close to 0) for x_i far from x_j. Once the region and weights have been determined, regression function in the local region is generated generally using weighted PLSR. Finally, the concentration value for x_j is calculated using the local weighted PLSR regressor (grey dotted arrow in Figure 4.4-1-B). In contrast to global PLSR that treats all the regression surface at the same time, as either linear or non-linear, LWR models non-linear regions without compromising linear regions. This approach is particularly adequate for animal cell culture processes in which linear and non-linear behaviour may arise differently during different phases of cell cultures. Moreover, the use of similar samples in the local area could lead to better spectra pre-treatment and thus limiting the loss of information by attenuation of scattering effects in spectra. Drawbacks of LWR are the need of dense calibration samples, vulnerability to outliers (Naes et al. 1990) and the lack of a mechanistic model where fitted parameters specify particular physical or chemical properties of the cell culture. This is of great concern since regulatory agencies demand that NIRS signals be directly attributed to analytes or be an indirect measurement correlated with light scattering effects (European Medicine Agency 2014; U.S. Food & Drug Administration 2015). Then submission of monitoring procedures would eventually require efforts considering all possible combinations of local regressions.

Support Vector Regression (SVR)

A relatively novel alternative for non-linear modelling of NIR spectra is SVR (Cogdill and Dardenne 2004). The main difference of SVR from other typical regression methods is that its objective is not merely to reduce the fitting error but to fit the error within a particular threshold (±ε). Then the goal of SVR is to generate a regression function, or hyper plane, that has a maximum number of calibration samples at most an ε deviation from an actual concentration (y_i), and at the same time keeping the function as flat as possible (Smola and Schölkopf 2004). For instance, the hyper plane is considered as:

$$f(x) = (wx) + b \quad (7)$$

$$with \quad w \in X, \qquad b \in \mathbb{R}, \qquad \text{and x being a variable related to spectra}$$

Flatness is then assured by minimisation of w, for example minimising the norm as a convex optimization problem:

$$minimise\ \frac{1}{2}|w|^2 \quad (8)$$

Subject to

$$y_i - (wx) - b\ \leq \varepsilon \quad and \quad (wx) + b\ -\ y_i \leq \varepsilon \qquad (9)$$

However, it may not be the case that *f(x)*, which approximates all pairs (x_i, y_i) with ε precision, actually exists. Then a soft margin of slack variables ξ_i, ξ^*_i, are introduced for coping with unfeasible constraints of optimization (8) as stated by Vapnik (Vapnik 2000):

$$minimise\ \frac{1}{2}|w|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \qquad (9)$$

subject to

$$y_i - (wx) - b\ \leq \varepsilon + \xi_i\ , \quad (wx) + b\ -\ y_i \leq \varepsilon + \xi_i^* \quad \text{and} \quad \xi_i.\xi_i^* \geq 0 \quad (10)$$

The constant *C* determines the compensation between the flatness of *f(x)* and the amount up to which deviations larger than *ε* are tolerated. This general procedure is depicted in Figure 4.4-1-C. As could be observed, SVR is then less vulnerable to outliers since it could properly generalise and leave the outliers in the soft margin (forbidden symbol in Figure 4.4-1-C). In complex multivariate data optimization, (9) can be solved more easily in its dual formulation, which provides the possibility for extending the procedure to non-linear functions. This could be achieved by mapping the x_i patterns into some feature space *F* (Nilsson 1965):

$$\varphi: x \rightarrow F \quad (11)$$

Then standard SVR procedure is applied. Mapping into a higher, linear or non-linear, dimensional space, may require exacerbated computational power, thus the majority of SVR use implicit mapping by kernels. The most common are linear, polynomial and Gaussian radial basis function (RBF) kernels. The nature of the calibration set must be considered for properly selecting the kernel (Awad and Khanna 2015). The linear kernel is useful in large sparse data vectors with linear regularization, the polynomial may fit some soft non-linearity and RBF are general-purpose that are generally applied in strong non-linear regularization or in the absence of prior knowledge (Awad and Khanna 2015). This approach could be used

for generalizing difficult to fit data in complex systems. As well as for LWR, monitoring procedure submission could be likely cumbersome for relating SVR parameters to specific chemical or physical properties of the cell culture, particularly in strong non-linear processes mapped into high dimensional feature space.

4.4.1.1.4 Materials and methods

Cell cultures for NIR spectra acquisition

The bioreactor data set was designed with routinary monitoring for batch culture in mind. Several cultures of CHO cells were performed in 2 L bench-top bioreactors (Pierre Guérin, France) with a 1.5 L working volume: three batch cultures, two feed-harvest cultures with medium renewal and one batch culture with glucose spiking. The three batch cultures were intended to observe inter-batch heterogeneity as well as in-line and off-line expected routinary responses. Feed-harvest cultures were used for increasing the variance of mAb glycoforms within the calibration process, which could enhance model prediction capability. These were started after a first phase in batch mode, then 2/3 of cell culture was withdrawn and replaced by fresh culture medium. This procedure was repeated 2 and 4 times for these 2 feed-harvest cultures respectively. As relative abundance of particular mAb glycoforms is partially a function of cell culture progression, the use of feed-harvest cultures favoured not only samples with mAb at the beginning of the cultures, but also cell cultures with a wider variability of mAb glycoforms. Batch culture with glucose spiking was used for increasing mAb concentrations so that these values during routinely batch culture monitoring relied preferably within an appropriate concentration range.

The culture medium was a protein-free medium mixture consisting of a 1:1 volume ratio of PF-CHO (HyClone) and CD-CHO (Fisher Scientific) supplemented with 4 mM L-glutamine (Sigma Aldrich) and 0.1 % pluronic F-68 (Sigma Aldrich). The genetically modified DG44 CHO cell line was used (human anti-Rhesus D mAb-producing CHO M250-9), kindly provided by Bioprocessing Technology Institute (Singapore). Dissolved oxygen (DO) was controlled at 50 % air saturation and agitation rate was fixed at 90 rpm throughout the culture.

Temperature was maintained at 37 °C and pH was set and controlled at 7.2 using 0.5 M sodium hydroxide and CO_2.

In-situ (or in-line, invasive) spectral scanning of bioprocess culture media was carried out with a NIR transflectance probe with 1 mm pathlength (Precision Sensing Devices, MA). The autoclavable probe was connected to an Antaris II spectrometer (Thermo Scientific, USA). Each NIR spectrum corresponded to an average of 128 scans from 1,000 to 2,500 nm.

Off-line analyses

Off-line concentration of total mAb was determined using an enzymatic kit (Roche Life Science) with an automatic spectrophotometer (Thermo Scientific GALLERY) against external standards. The nature and concentration of mAb heterogeneity in the form of glycoforms was elucidated by HPLC/UHPLC-mass spectroscopy analysis, as previously described (Li et al. 2018b). Off-line concentration values for calibration included and exceeded those expected during routinary monitoring of batch cultures (0 – 240 $mg.L^{-1}$ and 0 – 75 $mg.L^{-1}$ for total mAb and NG-mAb respectively) and also the variability of mAb glycoforms. Off-line total mAb concentration range used for calibration was 0 – 380 $mg.L^{-1}$, off-line NG-mAb range was 0 – 98 $mg.L^{-1}$. Analysis of mAb glycoform relative abundance profiles revealed a significant difference between exponential and stationary-death phase of cultures (One-way ANOVA, $p \leq 0.05$), particularly for NG-mAb, G0F, G1F, G2F and Man5 mAb glycoforms (data not shown). Moreover, enhancement of prediction capacity is expected since the use of feed-harvest cultures increased mAb glycoform variability during the calibration process.

Development and analysis of calibration models

Firstly, in order to generate the calibration methods for mAb glycoforms, special attention was given to spectrum pre-processing according to Huang (Huang et al. 2010). The presence of additive, multiplicative and wavelength-dependent effects due to scattering was evaluated within the calibration set spectra. The most common techniques to eliminate undesired spectral variations caused by light scattering (Multiplicative Scatter Correction-MSC,

Probabilistic Quotient Normalization-PQN, Standard Normal Variate-SNV, Extended Multiplicative Scatter Correction-EMSC, derivatives) were evaluated. The standard deviation for each wavelength was used to elucidate the effect of scattering on calibration spectra; this data was considered for final spectral pre-treatment. The calibration set comprised 168 spectra collected from six bioreactor cultures. PLSR, LWR and SVR models and statistical analysis were performed in MATLAB® (Statistics and Machine Learning Toolbox™, MATLAB R2016a, The MathWorks, Inc., Natick, Massachusetts, United States) using chemometric software (PLS_Toolbox® 8.2.1, Eigenvector Research, Inc., Manson, WA, United States). Model performance was evaluated for accuracy by Root Mean Square Error of cross-validation (RMSECV) and square correlation coefficients (R_2). A low value of RMSECV is related to enhanced accuracy, while a high value R_2 value indicates that the model properly handles spectrum variability to perform concentration estimation.

Firstly, PLSR models were performed using a venetian blinds cross-validation. Determination of latent variable (LV) number was based on the goodness of estimation (Q2Y): the minimum number of LVs was obtained when Q2Y ceased to improve. LWR was applied to fit global non-linear relationships by local linear regressions using PLSR and the classic cubic weight equation. Determination of local areas in term of local points, and LV number, was performed by optimization of these parameters with RMSECV as the response variable. For SVR models, an epsilon-support vector regression using a Gaussian radial basis function kernel was used. SVR models were optimized using a random subset cross-validation approach with maximal error values corresponding to deviations up to 10 % from actual values.

In situ monitoring of mAb glycoforms

The focus of the work was primarily to analyse predicted or estimated kinetic profiles using batch culture. This provides a frame containing different physiological cell states within lag, exponential, stationary, and death phases of batch culture which dynamically impact the nature of mAb glycoform profiles. This strategy may be useful to infer model performance in different matrix compositions, which may help in future work.

Once calibration models were optimized, they were used independently to perform *in situ* monitoring of mAb glycosylation during a CHO cell culture. Characterization of mAb was carried out in terms of macro-heterogeneity (total mAb concentration, glycosylated mAb and non-glycosylated mAb) and micro-heterogeneity (high mannose glycoforms and glycoforms containing any fucose, sialic acid and galactose moiety). A NIR analyser was programmed for performing automatic *in situ* scanning of culture medium every 20 min. Batch culture monitoring produced 500 spectra from which only 27 were used for calibration. Thus approximately 95 % NIR data was not used to establish the models and may further depict the prediction performances of models. For the evaluation of monitoring, models mimick real-time monitoring of the mAb producing cell culture process. Spectra captured every 20 min were then used as inputs. Then calibration models returned mAb concentration values that were used to real-time generate the kinetic profiles of mAb macro- and micro-heterogeneity. As global therapeutic effects of mAb is mainly a function of the micro-heterogeneity profile of the lot, real-time glycosylation data from best models were used to real-time monitor the global glycosylation profile of the produced lot. Firstly, the macro-heterogeneity profile was determined as the relationship of NG-mAb concentration estimated by the SVR model, with total mAb concentration estimated by the LWR model. Secondly, the micro-heterogeneity profile was determined as the relationship of fucosylated, galactosylated, sialylated and high mannose glycoforms with glycosylated mAb concentration, using SVR models.

4.4.1.1.5 Results and discussion

Development and analysis of NIR models based on PLSR, LWR and SVR methods

Spectra for calibration were evaluated for scattering effects (Figure 4.4-2-A). The most common techniques to eliminate undesired spectral variations caused by light scattering (MSC, SNV, EMSC, derivatives) were evaluated. General analysis of spectra revealed some scattering effects such as additive effect (baseline shift), multiplicative effect (offset of spectra) and a likely wavelength-dependent effect from approximately 1000 to 1500 nm (Figure 4.4-2-B).

Table 4.4-1. Spectral pre-treatment used for models

Compound	Regression method	Pre-treatment
Total mAb	PLSR	EMSC + SNV
	LWR	SNV
	SVR	EMSC
NG-mAb	PLSR	EMSC
	LWR	Detrend + MSC
	SVR	MSC
Glycosylated-mAb	PLSR	EMSC
	LWR	EMSC
	SVR	MSC
F-glycoforms	SVR	MSC
G-glycoforms	SVR	MSC
S-glycoforms	SVR	MSC
HM-glycoforms	SVR	MSC
NG-mAb: Non-glycosylated mAb; G-mAb: Glycosylated mAb; F-glycoforms: Glycoforms containing Fucose; G-glycoforms: Glycoforms containing galactose; S-glycoforms: Glycoforms containing sialic acid; HM-glycoforms: Glycoforms of high mannose structures. EMSC: Extended Multiple Scatter Correction, SNV: Standard Normal Variate, MSC: Multiple Scatter Correction. Auto scale was always applied as last pre-treatment step		

These effects were particularly observed in spectra of stationary and cell death phases where maximum mAb concentration was achieved (data not shown). Only EMSC was effective for limiting the likely wavelength-dependent effect, also the use of derivatives with any normalization pre-processing, particularly 2nd order derivative with MSC (Figure 4.4-2-B). This analysis was firstly used for selecting spectra pre-treatments for models. After spectral analysis, random trials of promising pre-treatments and their combinations were assed for reducing RMSECV of models.

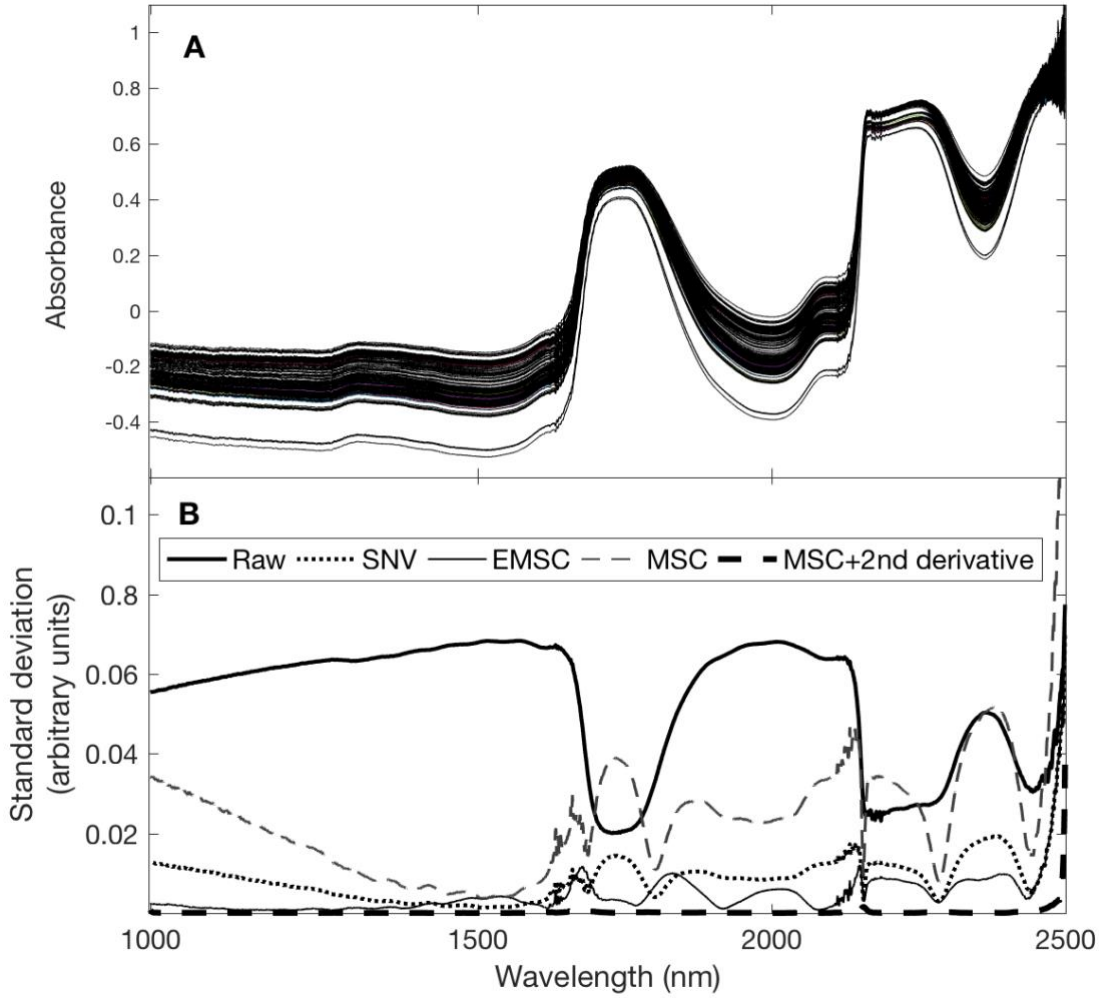

Figure 4.4-2. Nature of spectra for calibration: Process instrument raw spectra for calibration (A), effect of common spectral pre-treatments on spectra variability and scattering effects, MSC: Multiplicative Scatter Correction, SNV: Standard Normal Variate, EMSC: Extended Multiplicative Scatter Correction (B).

In general terms EMSC leaded to PLSR models with lower RMSECV, likely due to a proper compromise between reduction of spectra variability and scattering effects, particularly multiplicative and wavelength-dependent effects. The use of derivatives after normalization, limited the predictive power of PLSR models (data not shown), likely due to a strong reduction on spectra variability (Figure 4.4-2-B) and thus reduction on chemical information. On the other hand, LWR and SVR showed greater management of scattering effects, particularly wavelength-dependent effect since only MSC was sufficient for reduction of

RMSECV values in almost all models. Final spectral pre-treatments used for calibration are shown in Table 4.4-1.

Construction of mAb glycosylation models was performed using different regression methods (PLSR, LWR and SVR) as reported in Materials and Methods. Performances of models during calibration are summarized in Table 4.4-2. PLSR models leaded to poor estimation capability, even for total mAb concentration. PLSR is a variable space-based regression method which calculates the relationship between each of the variables (absorptions at different wavelengths) and compound concentrations. Such a relationship should be relatively constant during the whole culture process in order to maintain accurate estimations, including the scattering nature of the matrix. Perhaps the most evident deviation from this assumption is the fact that scattering effects occurred and impacted spectra in different ways, depending on the increase in scattering compounds (cells, cell debris, among others) according to batch culture progression. A plausible reason for poor performance is the limited capacity of PLSR for handling multiplicative and wavelength-dependent effects (Martens et al. 2003), likely caused by scattering compounds. Handling spectra with such different scattering natures with the same spectral pre-proccesing as commonly done in PLSR, would not only lead to correcting response in a narrow frame but also masking chemical information in the uncorrected frame.

As a result, LWR and SVR were, for the first time, evaluated for cell culture monitoring. LWR and SVR are sample space-based regression methods which firstly focus on affinity and dissimilarity between samples (batch progression information), and secondly on the relationship between variables (absorptions at different wavelengths) and compound concentrations. LWR uses only similar samples in the PLS space to perform local regression using weighted PLSR, while SVR consists of a number of support vectors corresponding to samples from the calibration set and non-linear model coefficients defining the relationships between spectra and compound concentrations. As LWR and SVR use only similar samples with a similar matrix nature, including similar scattering effects, it is likely that non-desirable effects of spectral pre-processing are limited.

Table 4.4-2. Performance of models during calibration

Model structure	Range ($mg.L^{-1}$)	RMSECV		R_2	Compound
		($mg.L^{-1}$)	(%)		
PLSR: 5 LV	0 - 380	34.1	9	0.85	Total mAb
LWR: 4 LV, 21 LP		31.1	8	0.90	
SVR: 120 SV		32.4	8	0.87	
PLSR: 3 LV	0 - 98	19.0	19	0.38	NG-mAb
LWR: 12 LV, 15 LP		15.0	15	0.61	
SVR: 148 SV		10.4	11	0.70	
PLSR: 3 LV	0-330	47.1	14	0.70	Glycosylated-mAb
LWR: 3 LV, 15 LP		37.5	11	0.81	
SVR: 158 SV		33.1	10	0.83	
PLSR: 4 LV	0 - 262	37.6	14	0.68	F-glycoforms
LWR: 24 LP, 3 LV		28.3	11	0.82	
SVR: 167 SV		27.9	11	0.83	
PLSR: 5 LV	0 - 147	17.6	12	0.71	G-glycoforms
LWR: 20 LP, 5 LV		13.7	9	0.83	
SVR: 166 SV		13.3	9	0.85	
PLSR: 4LV	0 - 19	2.4	13	0.70	S-glycoforms
LWR: 12 LP, 4 LV		2.2	12	0.74	
SVR: 148 SV		1.9	10	0.81	
PLSR: 5 LV	0 - 47	6.8	14	0.62	HM-glycoforms
LWR: 15 LP, 3 LV		5.4	12	0.77	
SVR: 166 SV		4.6	10	0.83	

NG-mAb: Non-glycosylated mAb; G-mAb: Glycosylated mAb; F-glycoforms: Glycoforms containing Fucose; G-glycoforms: Glycoforms containing galactose; S-glycoforms: Glycoforms containing sialic acid; HM-glycoforms: Glycoforms of high mannose structures. LV: Number of Latent Variable used for the PLSR and LWR models, LP: Number of local points used for the LWR models, SV: Number of Support Vectors used by the SVR models.

Results showed that, in contrast to PLSR performance, SVR and LWR were superior for estimating the concentration of all glycoforms (Table 4.4-2). This can be explained by the fact that SVR and LWR not only consider the relationship between spectra and compound concentrations, but also cell culture progression in terms of cell density, viability and metabolite concentrations. With the exception of total mAb which was better estimated by LWR, SVR was likely the best option for all mAb glycoforms, particularly glycosylated mAb.

Real-time monitoring of mAb glycosylation

Firstly, mAb glycosylation macro-heterogeneity monitoring was addressed as shown in Figure 4.4-3. As expected, evaluation of PLSR models revealed a limited capacity for monitoring mAb macro-glycoforms, particularly non-glycosylated mAb, due to non-linear relationships between spectra and non-glycosylated mAb concentration (data not shown). In fact, non-linear relationships are likely the result of physical (scattering, mass and heat dynamics) and chemical (chemical composition changes) phenomena that strongly change the interaction of NIR radiation with mAb during progression of batch cell cultures. LWR breaks global nonlinearity by performing several local regressions using only similar samples. In this context, LWR was successful in monitoring total and glycosylated mAb concentration (Figure 4.4-3) that display some non-linearity mainly associated to physical phenomena. However, for non-glycosylated mAb, only trends were observed. Limited capacity to estimate NG-mAb concentration by the LWR model is explained by the fact that an inherent nonlinear relationship between spectra and concentration existed (data not shown), which cannot be properly modelled by the local linear regressions. As shown in Figure 4.4-3, the novel use of LWR and SVR as enhanced regression methods, allowed the proper monitoring of total mAb and NG-mAb respectively. These results demonstrated the capability to monitor mAb glycosylation macro-heterogeneity in real-time, using *in situ* NIR spectroscopy. As more accurate and stable estimations of glycosylated mAb concentration were achieved using SVR, calibration for glycosylated glycoforms was addressed using the SVR approach.

As for mAb micro-heterogeneity, among a total of 25 potential glycosylated mAb glycoforms reported for mAb produced in CHO cell cultures (Sinha et al. 2008), only some glycoforms were detected off-line (data not shown). Thus, mAb micro-heterogeneity models were generated based on the detected glycoforms that contained particular sugar moieties conferring clinical properties. Detected glycoforms were classified into 4 groups for mAb micro-heterogeneity model development, corresponding to high mannose, fucosylated, sialylated and galactosylated isoforms (Table 4.4-3).

Table 4.4-3. Glycoforms considered for NIR mAb micro-heterogeneity calibration models

NIR models	Fucosylated	Galatosylated	Sialylated	High Mannosylated
Glycoforms	G0F, G1F, G2F, G0F-N, G1F-N, G0FN, G1FN, G2FN, G1FS, G2FS, G2FS2, G1FS-N	G1F, G2F, G1F-N, G1, G2, G1FN, G2FN, G2N, G1FS, G2FS, G2FS2, G1S, G2S, G1FS-N	G1FS, G2FS, G1S, G2S, G1FS-N	Man5, Man6, Man7, Man8

Oligosaccharide structures attached to the conserved site (Asn 297) in each heavy chain of the mAbs.
F a fucose moiety in the sugar chain; G a galactose moiety in the sugar chain: 0 no galactose, 1 a galactose in only one branch, 2 if both branches galactosylated; S a sialic acid in the sugar chain: no number if a S in only one branch, 2 if both branches contain a S; N N-acetylhexosamine in the sugar chain; Man mannose moieties in the sugar chain, number indicating the number of mannose moieties in the sugar chain.

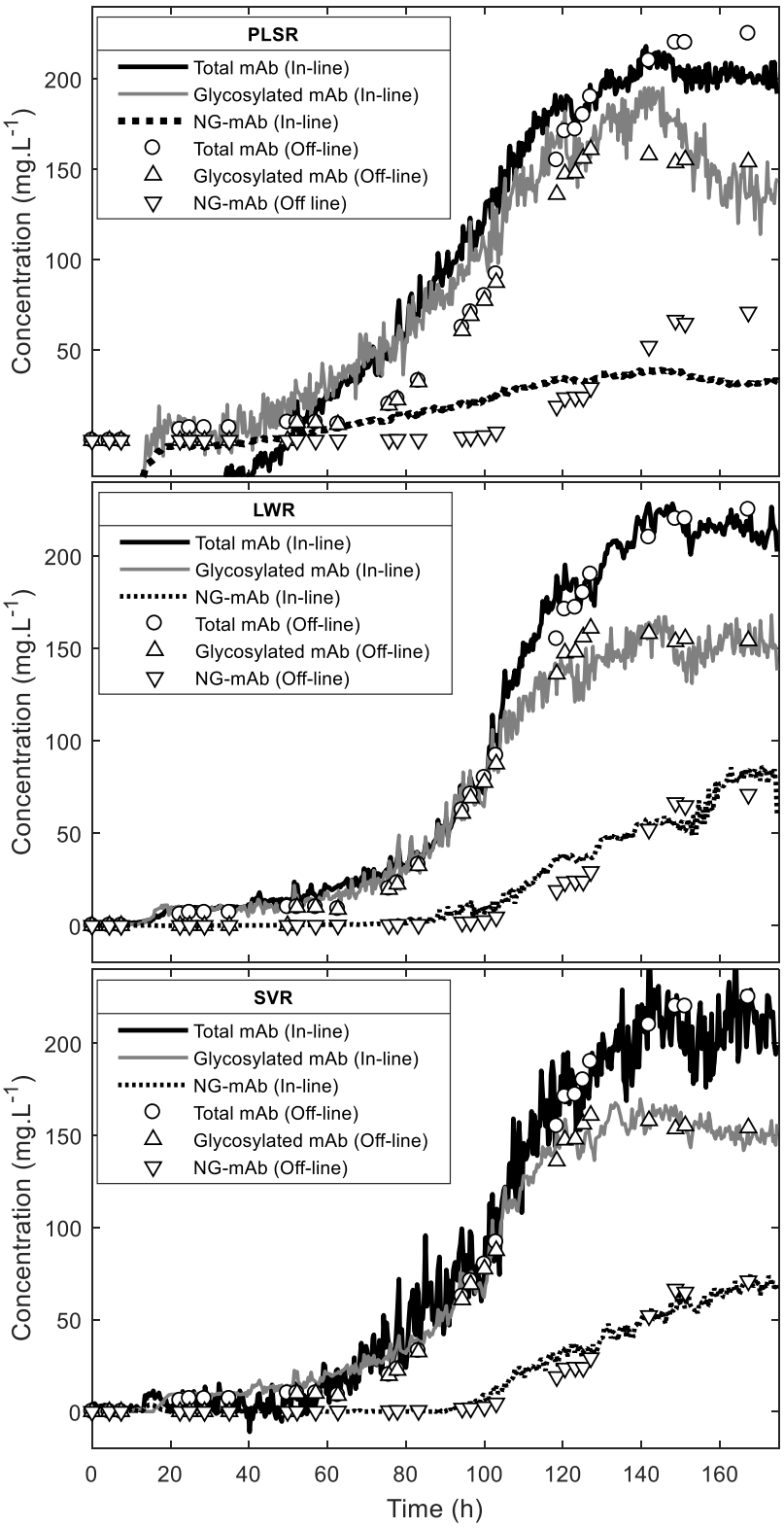

Figure 4.4-3. Comparison of PLSR, LWR and SVR models to monitor mAb macro-heterogeneity glycoforms concentration by *in situ* NIR spectroscopy, during batch CHO cell culture

As a glycosylated chain may contain different sugar moieties, it is possible that one particular glycoform be considered for two or more calibration models. SVR was capable of properly extracting mAb glycosylation information from NIR spectra, which allowed mAb micro-heterogeneity monitoring as shown in Figure 4.4-4. Even sialylated and high mannose glycoforms whose concentrations were low (<15 mg.L^{-1}), were specifically detected. These results demonstrated the capability of *in situ* NIR spectroscopy to quantitatively monitor mAb micro-heterogeneity.

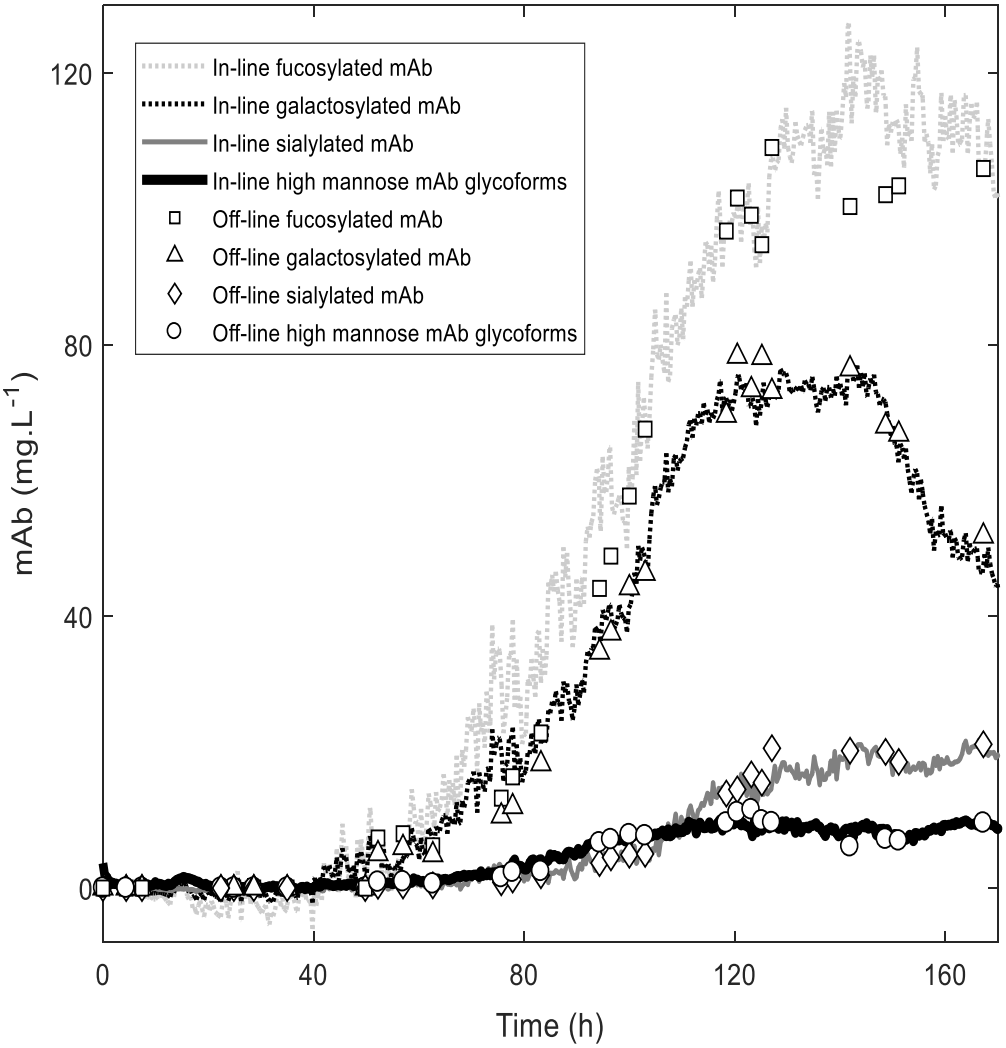

Figure 4.4-4. Performance of SVR models to monitor mAb micro-heterogeneity glycoforms concentration by *in situ* NIR spectroscopy, during batch CHO cell culture.

According to the QbD initiative, real-time monitoring should finally be used for performing advanced retro-control. Therefore, concentration values of mAb glycoforms must also be monitored using a reference frame as for setting target values for glycoforms ratios, which are related to mAb clinical effects. Monitoring of a process under this approach is a more challenging task since calculation of ratios could increase the bounce of glycoforms ratio profiles. For example, monitoring the extension of mAb macro-heterogeneity or the ratio of glycosylated mAb with total mAb concentration, would add the error of predictions of both glycosylated mAb model and total mAb model. This fact could compromise the resolution of the final mAb glycosylation profiles and so further control strategies. Therefore, the capability of models to clearly show these final profiles was investigated in the form of a control chart as shown in Figure 4.4-5.

Only best models were used for calculating the mAb macro- and micro-heterogeneity profiles. For the in-line mAb macro-heterogeneity profile, the fraction of glycosylated mAb was calculated as the ratio of glycosylated mAb (estimated by SVR) with total mAb concentration (estimated by LWR). Only SVR models were used for mAb micro-heterogeneity monitoring. For the in-line mAb micro-heterogeneity profiles, the fraction of either fucosylated, galactosylated, sialylated and high mannose glycoforms was calculated as the ratio of particular glycoform with glycosylated mAb concentration. Then micro-heterogeneity profiles could be monitored in real-time as the fraction of glycoforms containing particular sugar moieties within the glycosylated chain, which are closely related to clinical properties of mAb medicine.

Once mAb concentration was higher than 30 mgL^{-1}, models allowed proper monitoring of mAb glycosylation profiles. This approach was encouraging for monitoring mAb macro-heterogeneity since accurate tendencies were observed during the whole culture, particularly for the abrupt decrease of mAb glycosylation around 100 h after the beginning of the process. As for mAb micro-heterogeneity, particularly for the fucosylated glycoform profile, a limited capacity was observed between 45 h to 70 h of the culture. This behaviour was also observed, though to a lesser extention, for the galactosylated glycoform fraction profile. On the other hand, sialylated and high mannose isoforms profiles were properly estimated even at low concentrations. Results demonstrated the potential of SVR, LWR and NIR spectroscopy for real-time monitoring of mAb glycosylation properties during CHO cell culture processes.

Moreover, accuracy on concentration monitoring also permitted monitoring of accurate trends of mAb glycoforms ratios, closely related to mAb clinical effects. Then such mAb glycoforms ratios could be used as target values for later control.

Real-time estimation (lines) and off-line values (symbols) for the groups of (——, ○) glycosylated mAb, (——, △) fucosylated mAb, (·····, □) galactosylated mAb, (- - -,▽) sialylated mAb and (— -, ◇) high mannose mAb.

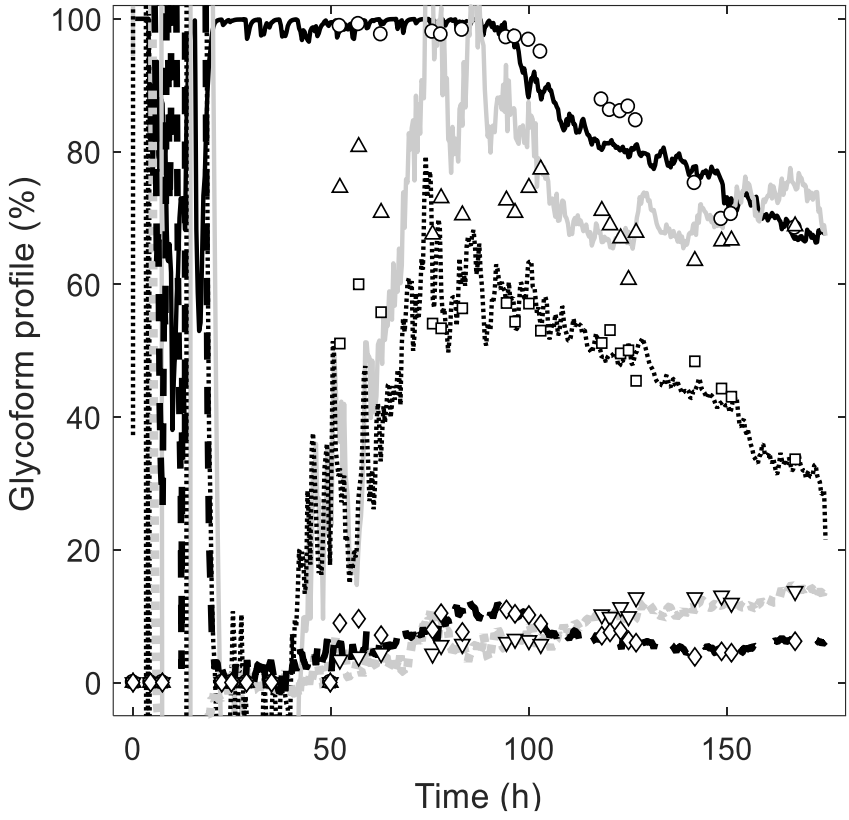

Figure 4.4-5. Example of model performances for automatic in-line monitoring of mAb glycosylation profiles: The mAb macro-heterogeneity profile was calculated as the concentration ratio between glycosylated mAb (estimated by SVR model) and total mAb (estimated by LWR model). The mAb micro-heterogeneity profiles were calculated as the ratio of fucosylated, galactosylated, sialylated and high mannose glycoforms in relation to glycosylated mAb concentration (all estimated by SVR models).

Overall, these are encouraging results for the use of NIR spectroscopy for developing new retro-control systems. However, caution must be taken when discussing eventual prediction capability of LWR and SVR in particular for mAb production processes. The same consideration should be shown in the case of NIR spectroscopy, as it contains both physical and chemical information linked to chemical and physical phenomena within processes. There is always the possibility of new variables in new production processes that have not been considered yet are critical to the performance of this newly developed prediction platform.

4.4.1.1.6 Conclusions

Data demonstrating the feasibility of NIRS to monitor mAb glycosylation *in situ* has been presented. In this study, the monitoring of both macro- and micro-heterogeneity of glycosylated mAb was improved by the novel use of sample space-based regression methods, particularly SVR, that could handle non-linear relationships between glycoforms and spectra. As far as it can be asserted, this is the first report of real-time and *in situ* monitoring of mAb macro- and micro-heterogeneity using NIR spectroscopy as well as the first report of LWR and SVR methods for cell culture monitoring. Such methods dealing not only with chemical but also some physical information contained within spectra, highlight the importance of considering the strongly dynamic nature of cell culture processes for accurate monitoring by calibration models.

There is an increasing number of new mAb producing processes including mAb biosimilars and biobetters (del Val et al. 2012), and one can assume that *in situ* spectroscopy methods will be implemented systematically to fulfil the demand in regard to quality. This study lays the foundation for future studies to expand the capabilities of *in situ* spectroscopy and multivariate analysis to monitor mAb properties so that enhanced retro-control strategies can be established, leading to a more efficient design and control of processes using PAT and the Quality by Design principles.

4.4.1.2 *In situ* cell differentiation monitoring of *Catharanthus roseus* suspension culture processes by NIR spectroscopy

Zavala-Ortiz Daniel Arturo[1,2], Ebel Bruno[2], Guedon Emmanuel[2], Marc Annie[2], Barradas-Dermitz Dulce María[1], Hayward-Jones Patricia Margaret[1], Aguilar-Uscanga María Guadalupe[1*]

[1] Tecnológico Nacional de México/Instituto Tecnológico de Veracruz: Calz. M.A. de Quevedo 2779, Veracruz, Ver. Mexico.
[2] Laboratoire Réactions et Génie des Procédés, Université de Lorraine, CNRS, LRGP, 2 avenue Forêt de Haye, TSA 40602, 54518 Vandœuvre-lès-Nancy, France.

4.4.1.2.1 Abstract

Plant suspension culture is attracting interest as a promising platform to produce biological medicines due to the absence of virus, prions or DNA related to mammals during the production process. However, the heterogenic plant cell proliferation nature is particularly challenging for establishing industrial processes based on innovative approaches currently used, particularly in the animal cell culture industry. In this context, while Process Analytical Technology (PAT) approaches have been used to monitor classical parameters such as biomass dry weight, its use in cells heterogeneity has received limited attention. Therefore, the feasibility of *in situ* monitoring of cell differentiation in plant cell suspensions employing NIR spectroscopy and chemometrics was investigated. Off-line measurements of cell heterogeneity in term of cell differentiation and in-line NIR spectra captured in 3 L bioreactor cultures were employed to generate calibration models. Then models were tested to estimate the population distribution of parenchyma, collenchyma and sclerenchyma cells during *Catharanthus roseus* suspension cultures. Results have proven *in situ* NIR spectroscopy as a capable PAT tool to monitor differentiated cells accurately and in real-time. These results are the starting point to follow up PAT systems so that plant cell culture heterogeneity may be better understood and controlled in biopharmaceutical plant cell cultures.

4.4.1.2.2 Introduction

Plant cell suspension cultures have attracted interest in biological medicine production due mainly to safety aspects (Fischer et al. 2004). In this context, some companies have started

to industrially produce biologicals using recombinant plant cell suspension cultures. Examples of the potential of this technology are the industrial production of human glucocerebrosidase and immune-protective proteins (Huang and McDonald 2009), also the production of paclitaxel, shikonine, berberine and cosmetics by wild species. The relatively novel application of plant suspension cultures for biologicals manufacturing has implied a delay on the implementation of biopharmaceutical trends such as the use of Process Analytical Technology (PAT). The main objective of PAT is to real-time monitor and control critical process parameters, such as cells and culture media compounds, whose variability have an impact on titer or quality of the biological medicine, as to ensure the best process performance (Teixeira et al. 2009b). Consequently, strategies for monitoring of cell culture processes must be firstly developed for the further application of PAT. As far as it can be ascertained, development of real-time monitoring tools for plant cell suspension cultures has only considered the packed cell volume or cell dry weight using conductivity and permittivity analyzers (Markx et al. 1991; Kwok et al. 1992; Holland et al. 2013). However, plant cells in suspension cultures are highly heterogenic and thus such approaches fail in providing information of cell heterogeneity, which may limit the implementation of PAT strategies.

Indeed, plant cells morphology is dynamic and strong changes in cell volumes and shapes may occur during cultures, causing limited significance of classic monitoring variables such as cell dry weight or packed cell volume (Kim et al. 1994a). Moreover, genetic and phenotypic changes usually occur in terms of somoclonal variation (Deus-Neumann and Zenk 1984) and cell differentiation (Torrey 1975). The complex nature of cell proliferation in cell aggregates is not only a challenge during production cultures, but also for the establishment of monoclonal cell lines. Therefore recombinant biologics production is usually performed by polyclonal cultures (Nocarova and Fischer 2009). All these sources of heterogeneity in plant cell suspension cultures must be understood, monitored and controlled to continuously improve the cell culture processes as encouraged by regulatory agencies. Therefore, the aim of this study was to evaluate the feasibility of *in situ* NIR spectroscopy for real-time monitoring of cell culture heterogeneity in term of cell differentiation. This study lays the foundation for future studies to expand the capabilities of PAT tools for *in situ* real-time monitoring of cell heterogeneity in biopharmaceutical production processes based on plant cell suspension cultures. In this context, NIR spectroscopy could be a monitoring

platform for real-time monitoring of the state of cells firstly, providing meaningful information of cell physiological state using calibration models, and then secondly performing retro-control strategies, leading to a more efficient design and control of plant suspension culture processes.

4.4.1.2.3 Material and methods

Heterogeneity analysis, calibration and in-line monitoring.

Heterogeneity analysis. Firstly, the cell culture process was subject to heterogeneity analysis using off-line data from three batch cultures in bioreactors. Analysis of the relationship between cell concentration and cell dry weight, as well as cell differentiation and morphology were performed. Secondly, Principal Component Analysis (PCA) was used to reduce the dimension of spectral data and then reveal dissimilarities, leading to qualitatively analysis of the impact of cell heterogeneity on NIR spectra.

Calibration. Three batch cultures were used to collect in-line (NIR spectra) and off-line (cells properties) data. Then both types of data were used to generate calibration models. Models for parenchyma, chlorenchyma and sclerenchyma cells, were created using Partial Least Squares Regression (PLSR) with venetian blinds cross-validation and the NIPALS algorithm. Determination of latent variable (LV) number was based on the goodness of estimation (Q2Y), which is the fraction of the captured variation of the cell concentration estimated by the model. The minimum number of LVs was obtained when Q2Y ceased to improve, which was in all cases when models captured over 90 % of cell concentration variation. Spectra data was pre-treated using Probabilistic Quotient Normalization (PQN), Savitzky–Golay second derivative algorithm and Standard Normal Variate (SNV), mean centering was used as last pre-treatment. The estimation capacity of calibration models was statistically assessed in terms of accuracy, precision and linearity using the Root Mean Square Error of Cross Validation (RMSECV), the relative error of estimation (REE) and the coefficient of correlation (R_2) respectively. Spectra acquisition was captured directly from the analyzer. Then multivariate calibration models and statistical analysis were carried out using PLS-Toolbox 8.2.1 (Eigenvector Research Inc.) in the R2016a MATLAB® environment (MathWorks Inc.).

In-line monitoring and validation of models. To evaluate the capacity of the models for performing real-time and *in situ* monitoring, NIR spectra collected every 30 min during batch cultures, were used as inputs for the calibration models. Then off-line and in-line kinetic profiles were compared for internal validation.

Bioreactor cultures and in-line spectra acquisition

Cell cultures were performed in 3 L benchtop bioreactors (Applikon, the Netherlands) with 2 L working volume. Three batch cultures were performed for obtaining off-line and in-line data of cultures. Agitation of suspension culture was performed by a spin-filter coupled to a marine impeller. The set points for all cultures were for temperature 30 °C and stirring 90 rpm. Aeration of cultures consisted in air flux of 200 mL.min^{-1}. Dissolved oxygen (DO) and pH (5.6) were maintained at 50 % saturation and 5.6 respectively in only one batch culture (batch 1) (Applikon Bio controller ADI 1010). For the other two batches (batch 2 and 3), pH and DO were monitored but not controlled. Bioreactors were illuminated by a LED light panel below the jar in a 16 h/d photoperiod (50 μmol.m^{-2} s^{-1} photonic flux). Culture media was the same as subsequent described but supplemented with 1000 ppm Plant Preservative Mixture (Plant Cell Technology), culture media was filter-sterilized using a bottle top filter (0.20 μm, Thermo Scientific™ Nalgene™) and then tipped out into the autoclaved bioreactor. Inoculation consisted in adding approximately 100 mL of 6-days-old suspension culture into the bioreactor containing 1.9 L of culture media.
An *in situ* or *in-line* transflectance probe (Precision Sensing Devices, Inc., Medfield, MA) with 6 mm effective pathlength was coupled to the bioreactor, the probe was connected to a XDS Process Analytics near infrared spectrophotometer analyzer (Foss NIR Systems, Silver Spring, USA). A culture media scanning corresponded to an average of 64 scans with a 0.5 nm resolution from 800 nm to 2200 nm.

Inoculum preparation and culture media

The suspension cultures of *Catharanthus roseus* were generated from callus cultures generated from leaf tissues. Calluses with similar friability, color and age were used to

generate suspension cultures. Approximately 5 g fresh callus was placed into 20 mL liquid MS media supplemented with 2 $mg.L^{-1}$ glycine, 100 $mg.L^{-1}$ myo-inositol, 30 $g.L^{-1}$ sucrose, 2, 4-Dichlorophenoxyacetic acid (2, 4-D), 4.44 µM 6-Benzylaminopurine (BAP). The pH was adjusted to 5.7 ± 0.1 with 0.5 N NaOH and 0.5 N HCl and the medium was autoclaved at 121°C for 15 min. Cultures were then incubated during six days in 125 mL Erlenmeyer flasks in 16 $h.d^{-1}$ photoperiod (30 $\mu mol.m^{-2}\ s^{-1}$ photonic flux), 25°C and shaken at 100 rpm on an orbital shaker. After the first culture, the suspension culture was filtered using a sterile stainless-steel mesh (2.25 mm^2 pore size) to retain callus clusters. Then approximately 20 mL culture volumes were used to equally inoculate two flasks containing 10 mL culture media, which were then cultured under the same conditions. Two subsequent subcultures were performed for biomass propagation. Then 16 flasks containing a total of 320 mL were concentrated by repeated cell decantation until an approximate 100 mL concentration was achieved which was used to inoculate bioreactors cultures. This procedure was repeated for every bioreactor inoculation.

Off-line analysis measurements

Cell differentiation. This was analysed by optical microscopy (Motic, MO-567, USA). Differentiation analysis of cells in suspension cultures was undertaken based on cell wall differences using botanical main classification (Mauseth 2014): parenchyma, collenchyma and sclerenchyma cells. Cells identification and counting were performed in enzymatically digested suspension culture aggregates. A sample of 1 mL was put into a 1.5 mL micro-centrifuge tube and centrifuged (250 rcf, 15 min); 900 µL of the supernatant were put in a tube and 900 µL digestion enzyme (TrypLE™ Express Enzyme (1X), Thermo Fisher Scientific) were added to the pellet. The tube was agitated in a vortex (MX-S, Science MED) for 45 min, centrifuged under the same conditions, and the pellet was isolated in 300 µL by discarding the supernatant, resulting in a digested cell suspension suitable for analysis. Cells in digested cell suspensions were analysed and counted twice in a Neubauer improved chamber, each count comprising between 75 and 120 cells.

Biomass dry weight. A volume of 1.0 mL suspension culture was put in an already weighed 1.5 mL micro-centrifuge tube. The tube was centrifuged (10,000 rcf, 10 min; Eppendorf

5424), its content decanted and then 1.0 mL deionised water added. Centrifugation and decantation were repeated, then the washed pellet was vacuum-dried (65°C, ShellLab mod. 1410) until constant weight was reached (Mettler H80).

4.4.1.2.4 Results and discussion

Heterogeneity analysis of the plant cell culture process

Heterogeneity of the culture process was firstly analysed. The relationship between cell concentration and biomass dry weight showed different profiles for the three cultures (Figure 4.4-6-A). This inter-batch heterogeneity was attributed to changes in morphology and weight of cells within cell aggregates, particularly due to differentiation of cells into parenchyma, collenchyma and sclerenchyma condition within cell aggregates (Figure 4.4-6-B). A plausible reason for non-consistent differentiation rate within cultures is the presence of diverse microenvironments in cell aggregates, which differently impact cell physiological state. Cell differentiation dynamics during culture is shown in Figure 4.4-6-C. Analysis showed that parenchyma cells mainly supported culture growth while collenchyma cells were likely to have limited proliferation capacity. On the other hand, as sclerenchyma cells are programmed to die, increase of cell concentration was only caused by cell differentiation. In this context, cell differentiation into sclerenchyma status was likely to be slow and constant during the culture, though cell differentiation into collenchyma from parenchyma status was relatively fast, as demonstrated by the abrupt decrease and increase of parenchyma and collenchyma cells respectively around 100 h of culture.

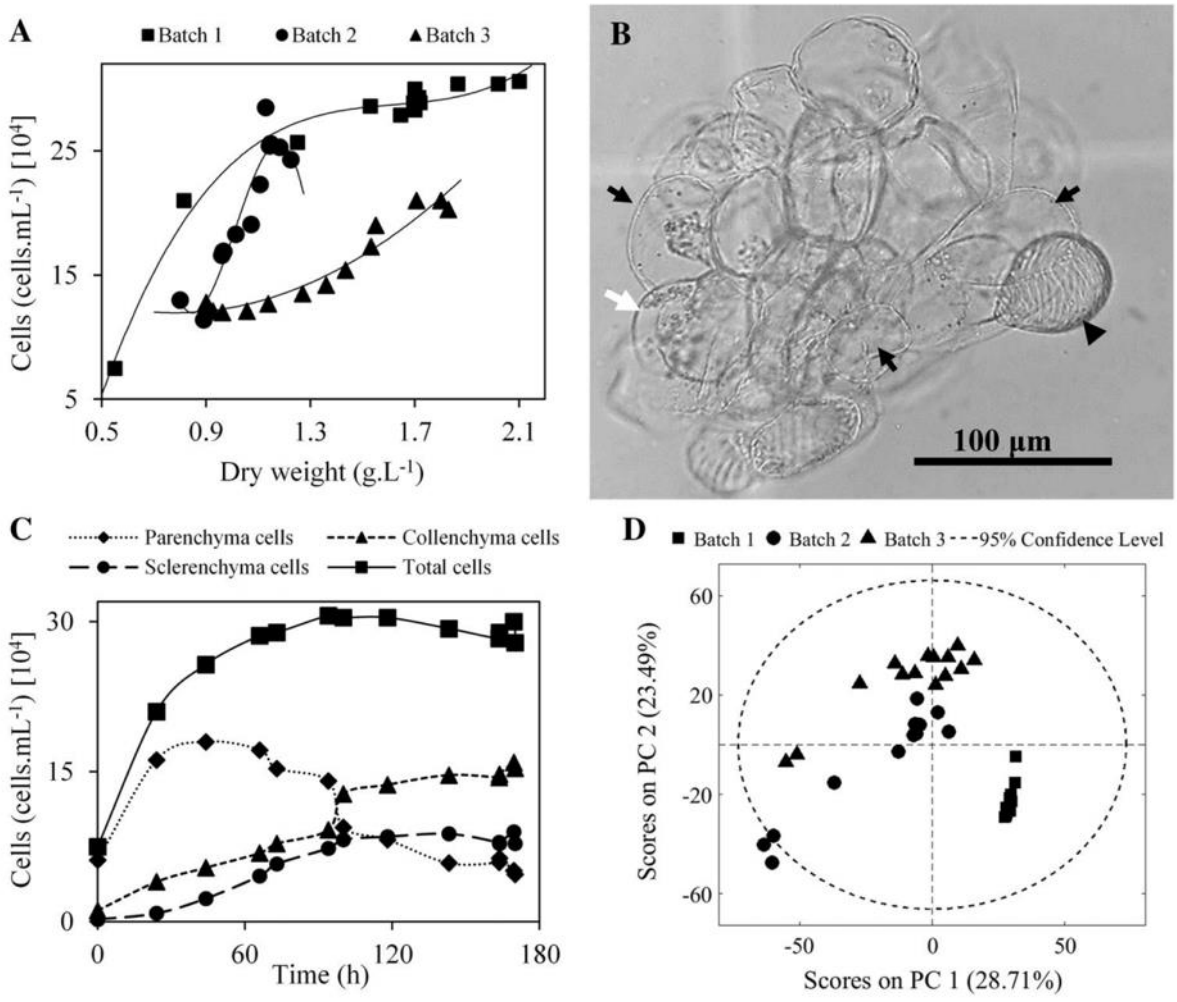

Figure 4.4-6. Heterogeneity within plant cell suspension cultures: (A) Inter-batch heterogeneity due to changes in cell properties. (B) Example of cell heterogeneity in term of cell differentiation within aggregates: Parenchyma cells (dark arrows), collenchyma cell (white arrow), and sclerenchyma cell (tringle). (C) Dynamics of cell differentiation with pH and DO controlled condition (Batch 1). (D) Inter-batch heterogeneity effect on calibration spectra analysed by Principal Component Analysis.

Differentiation can have a strong impact on process performances, such as growth arrest or expression of particular metabolic pathways (Lindsey and Yeoman 1983). Therefore its proper monitoring so as to detect early changes in differentiated cell subpopulations is mandatory for subsequent control action under the PAT approach (Teixeira et al. 2009b). However, current approaches based on daily sampling fail to provide real-time information

as, for example, to detect the abrupt change in parenchyma and collenchyma cell concentration at 100 h in batch 1. Consequently, real-time monitoring approaches are desirable to subsequently develop advanced retro-control systems. The impact of cell differentiation on spectra was carried out using PCA which revealed the trajectory (Clavaud et al. 2013) of the three cultures. In general, the trajectory of the culture process goes from left to right. Batch 1 resulted in a compact group while batch 2 and 3 were more widely spread (Figure 4.4-6-D). The distance between batch 1 and batch 2 and 3 was likely caused by the effect of pH and DO control. Punctual differences in culture trajectories between batch 2 and 3 were likely caused by differences in differentiated cell populations and their resulting culture media, which were adequately detected by NIR spectra.

Calibration and in-line monitoring of cell differentiation

In attempt to understand and monitor process heterogeneity, calibration models for the three main observed types of cells in cultures were performed. Characteristics of model performances are summarized in Table 4.4-4. The R_2 is a parameter used to depict the capability of the model to explain cell concentration variability based on spectra variability and is usually used to also evaluate calibration linearity. A coefficient with values close to one indicates that the model is capable to relate spectra variability to concentration well, close to zero lack capability. Models achieved relatively high R_2 values, particularly for collenchyma and sclerenchyma cells, demonstrating the feasibility of NIRS and multivariate analysis to measure differentiated cells within cultures in real-time manner. This fact is demonstrated by relatively high Q2Y values, particularly for collenchyma and sclerenchyma cells. The accuracy of models was characterized by low RMSECV values of approximately 11, 7 and 6 % of REE for parenchyma, collenchyma and sclerenchyma cells respectively.

Table 4.4-4. Statistical analysis of calibration models for plant cell cultures.

	Models for particular differentiated cell type		
	Parenchyma	Collenchyma	Sclerenchyma
RMSECV (cells.mL^{-1}) [10^4]	2.15	1.11	0.55
REE (± %)	12	7	6
R_{2CV}	0.76	0.92	0.97
PLSR model structure	5 LV	5 LV	5 LV
$Q2Y_{cv}$ (%)	77	91	94
RMSECV: Root mean square error of cross-validation; REE: Relative error of estimation of cross-validation; R_{2CV}:Correlation coefficient of cross-validation; QY2: Goodness of estimation of cross-validation.			

The models were then used to monitor the distribution of differentiated cell subpopulations in real-time by using the *in situ* probe as shown in the animation (Online Resource 1) within the three batch cultures as internal validation. The NIR calibration models were challenged differently since batch cultures were differently operated (pH and DO controlled for batch 1). Moreover, cell differentiation dynamics was different for the three batch cultures. For example, while collenchyma cells concentration increased at the beginning of the culture for batch 1 and 2, a decrease was observed for batch 3 (Figure 4.4-7); on the other hand, sclerenchyma cell concentrations remained mainly low and constant for batch 2 and 3 while an increase was observed for batch 1. Calibrating in such different differentiation dynamics can also be perceived as an advantage since any confused correlation between differentiated cell concentrations is limited, leading to a more accurate model performance (Riley et al. 1998).

The estimations of models based on *in situ* captured NIR spectra matched measured concentrations for parenchyma, collenchyma and sclerenchyma cells throughout the course of the three batch cultures as shown in Figure 4.4-7. Major deviations of model estimations from actual concentrations were evident for collenchyma cells during batch 2 at the beginning of the culture. This could be explained by the fact that spectra from the beginning of batch 2 were different from the majority of the calibration samples as suggested by PCA analysis (Figure 4.4-6-D).

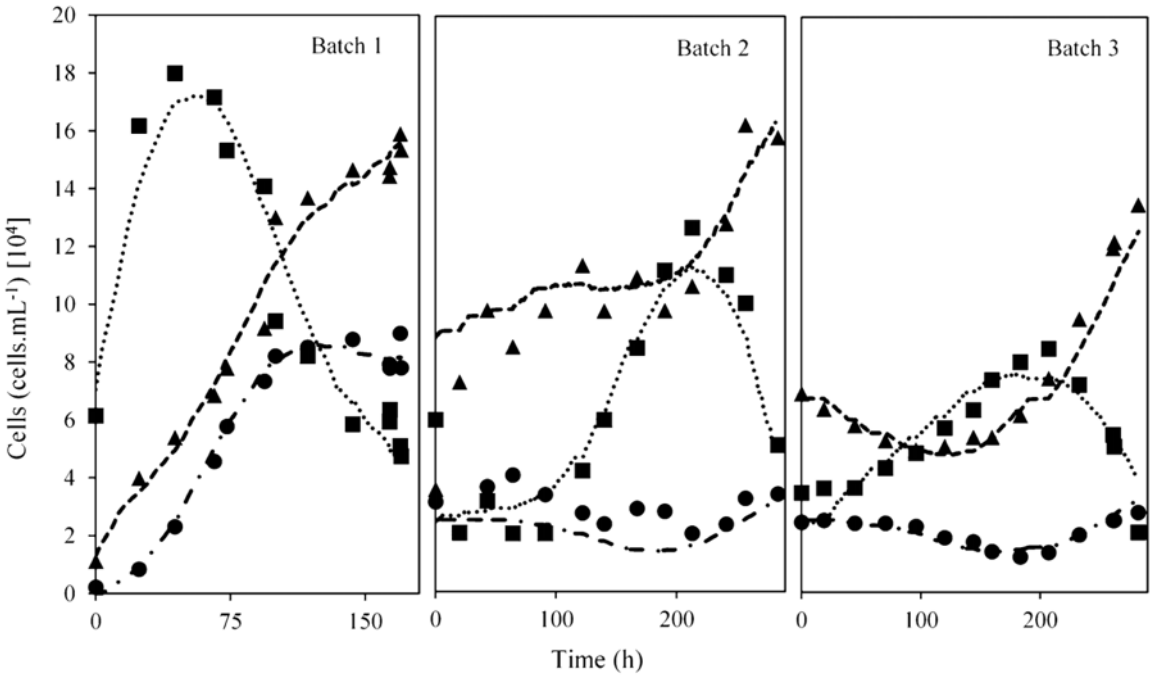

Figure 4.4-7. Performance of models to monitor differentiated cells using *in situ* captured NIR spectra in real-time: Real-time estimation by models (lines) and off-line measured values (symbols) for the groups of (·····,■) parenchyma cells, (-----,▲) collenchyma cells, and (– · – ,●) sclerenchyma cells.

Cell differentiation dynamics can have strong effects on process performance. Parenchyma cells are usually reported as fast growing cells in suspension cultures while collenchyma cells may have enhanced resistance to shear stress due to their thickened primary walls. On the other hand, sclerenchyma cells have been closely related to enhanced alkaloid production (Hoekstra et al. 1990) though also related to cell viability reduction of cultures (Twumasi et al. 2009). For instance, the extension of sclerenchyma cells should be maintained at minimum if cell accumulation is desired (beginning of cultures), collenchyma cells favoured in shear stressing conditions (high density cultures requiring strong mixing) and sclerenchyma cells induced for producing alkaloids once high density be achieved.

The calibration models developed could be used for monitoring cell differentiation dynamics firstly and then provide feedback control based on culture conditions changes such as addition of particular plant growth regulators (Twumasi et al. 2009). This scheme could

finally lead to establish PAT strategies on new enhanced processes based on plant suspension cultures of differentiated cells.

4.4.1.2.5 Current issues and future challenges

Based on current trends in biopharmaceutical industry, it seems likely that plant suspension culture will gain more importance as production platform. However, there are still several challenges to surpass for the successful establishment of production processes, particularly related to the heterogenic nature of plant cell proliferation in suspension cultures. The PAT initiative offers a frame for control of variability or heterogeneity based on real-time monitoring procedures and knowledge of the cell culture processes.

Though there are plenty of reports on monitoring, particularly with focus on some substrates and by-products concentration, key parameters for plant suspension cultures remains unaddressed since the growth and metabolism of plant cells considerably differ from those of microbial and animal cells. For instance, the size and shape of cells and cell aggregates, subpopulations of cells in G_0 phase, of cells with different production capacity in wild or polyclonal cultures, of cells with differentiated condition, among others.

Overall, the potential of *in situ* NIRS for cell differentiation monitoring in real-time has been demonstrated. As far as it can be ascertained, this is the first report applying NIR spectroscopy using an *in-situ* probe for real-time measurement of cell heterogeneity in plant cell suspension cultures. Nonetheless, sources of heterogeneity are diverse in plant cell suspension culture-based processes and it seems likely that the number and nature of components analysed using *in situ* spectroscopy will significantly expand. This study encourages future studies to expand the monitoring capabilities of analyzers for real-time monitoring of heterogeneity in plant cell suspension cultures.

4.4.1.2.6 Conclusions

The feasibility of *in situ* monitoring of some plant suspension culture heterogeneity in bioreactors through NIR spectroscopy has been demonstrated. This affirmation is based on the generation of calibration models for parenchyma, collenchyma and sclerenchyma differentiated cells of *Catharanthus roseus*. Models yielded appropriate accuracy and

precision to estimate the concentration of differentiated cells during bioreactor cultures in real-time, which could provide meaningful information of physiological state of cells. These are encouraging results to follow up in PAT approaches to better understand and control biological medicine production by plant suspension cultures.

4.4.2 Conclusions of chapter IV

The QbD-PAT initiative offers a frame for control of variability or heterogeneity based on real-time monitoring procedures and knowledge of the cell culture processes. Though there are plenty of reports on monitoring, particularly with focus on some substrates and by-products concentration, innovative parameters such as those characterizing heterogeneity had not been yet addressed. In this chapter, the potential of *in situ* NIR monitoring for monitoring of cell culture heterogeneity has been shown to some extension.

For mAb heterogeneity in CHO cell cultures, several regression strategies were used for generating calibration models of mAb CQA taking into account the nature of regression methods and the relationship between spectra and compounds concentration. Comparison of PLSR, LWR and SVR performance revealed that information of mAb heterogeneity contained in spectra is coded in a complex and mainly in a nonlinear way. Thus, common calibration approaches based on PLSR regression are likely to be not adapted for such a task. Real-time monitoring of glycosylation, in terms of high mannose isoforms, fucosylated, sialylated and galactosylated isoforms as well as non-glycosylated mAb, has been successfully performed by the novel use of Locally Weighted Regression (LWR) and Support Vector Regression (SVR) with relative error of predictions in the order of 10 %.

Plant suspension cultures are highly heterogenic in nature. The feasibility of *in situ* monitoring of heterogeneity in term of cell differentiation has preliminary been demonstrated. Calibration models were generated for differentiated cells within cell aggregates in suspension cultures of *C. roseus*. Models yielded appropriate accuracy and precision to estimate the concentration of differentiated cells during bioreactor cultures in real-time. Since cell differentiation is likely related to antileukemic agent synthesis, its monitoring could provide meaningful information for process control.

Overall, the potential of *in situ* NIRS for real-time monitoring of some heterogeneity in cell culture processes has been demonstrated. Nonetheless, sources of heterogeneity are diverse in cell culture-based processes and it seems likely that the number and nature of components analysed using *in situ* spectroscopy will significantly expand. This study encourages future studies to explore the monitoring capabilities of analyzers for real-time monitoring of heterogeneity within cell culture by the use of diverse chemometric tools.

5 GENERAL CONCLUSIONS AND PERSPECTIVES

The aim of this thesis project has been to enhance the capabilities of *in situ* NIR spectroscopy for properly monitoring of cell cultures. It has been addressed by evaluation of diverse chemometric approaches. The main conclusions have been summarized in four sections according to major contributions of the thesis work.

1. *Analysis of plant cell suspension cultures for real-time monitoring.*

As plant cell suspension cultures had not yet been addressed for monitoring using *in situ* NIRS probes, the first objective was to develop and then monitor a biopharmaceutical cell culture process. Analysis of suspension cultures revealed that production of biopharmaceutics under this cell platform involves particular challenges that are not commonly addressed in other cell platforms, particularly related to the complex proliferation nature of plant cells.

One particular challenge for the production of medicines using this cell platform is the different cell behavior under *in vitro* and *in vivo* conditions, which has limited the *in vitro* production of vincristine and vinblastine, first line chemotherapeutics, as their synthesis requires cell differentiation. Therefore, cell differentiation in suspension cultures was induced by changes in operating conditions, seeking different titers of the chemotherapeutics. Cell differentiation was observed within cell aggregates and seemed likely related to differentiated capacity of cell cultures for vincristine and vinblastine production.

This work has provided important clues into the comprehension of *in vitro* culture performance for metabolites production requiring *in vivo* cell differentiation. They demonstrated the utility of taking into account cell differentiation for the further development of novel advanced processes of differentiated cell suspension cultures for producing valuable molecules, including biological medicines such as VC and VB.

Perspectives

This contribution could go further by a deeper identification of the degree of cell differentiation by using biochemical markers, preferably in synchronous cultures for a deeper understanding of the cell differentiation phenomena. Moreover, the relationship of the cell differentiation progression with the expression of key enzymes for vincristine and vinblastine synthesis, as well as the effect of the micro-environment within cell aggregates of differentiated cells, should be addressed for the proper development of advanced differentiated cell cultures of plant cell cultures.

2. Evaluation of current monitoring approaches based on linear PLSR regressions

Several studies have shown the potential of NIR spectroscopy and multivariate analysis as a promising tool for monitoring of cell cultures processes. However, such complex processes have only been addressed by the use of linear regression methods for building calibration models for estimating chemical or biochemical variables based on NIR spectra. Therefore, the pertinence of current linear approaches for calibrating in cell culture-based processes was evaluated.

Cell culture processes have been shown challenging for monitoring by *in situ* NIR spectroscopy, particularly due to nonlinear relationships between spectra and concentration of particular compounds. Consequently, the widely used PLSR method was incapable of properly relating spectra with compound concentrations, indicating that such a widely used regression methods is not always appropriate for the monitoring of animal cell culture processes.

Indeed, it has been revealed that the relationship between spectra and concentration changes due to the strong dynamic nature of cell culture processes. Therefore, an approach which could consider this variability was used. Locally Weighted Regression-LWR was considered by the fact that it is a sample-based regression which firstly focus on similarity of calibration samples, related to cell culture progression, and then perform regression.

The novel use of the LWR method was shown to overcome PLSR limitations, which led to more accurate predictions of culture compound concentrations. The break of global

nonlinearity into several local linear frames was shown to be useful, providing more accurate and precise prediction using the same spectra. However, there were still nonlinearity that was not caused by cell culture progression where even the local approach failed. This fact suggested that there could be inherent nonlinearity relationships between spectra and compound concentration.

Perspectives

Overall, the results highlighted the fact that *in situ* NIR spectroscopy could have a broader potential as a PAT tool provided that the effect of culture dynamics and nonlinearity be considered. Therefore, new approaches based on different instrumentation modes (probe nature, different pathlengths for particular compounds) and chemometric tools (multivariate analysis) should be explored seeking pertinent monitoring of new key parameters providing meaningful information for advanced cell culture control protocols.

3. *Evaluation of new nonlinear regression methods for the generation of monitoring models*

As the vulnerability of PLSR models for monitoring of cell cultures has been shown due to nonlinear relationships between spectra and concentrations of substrates, products and viable cells, the use nonlinear regression methods for the generation of calibration models had been addressed seeking enhancement of predictive power compared with linear approaches currently used in biopharmaceutic industry.

Supported Vector Regression-SVR and Artificial Neural Network Regression-ANNR, sample-based and variable-based nonlinear methods respectively, were evaluated and compared to PLSR. The novel use of SVR and ANNR showed to be a promising alternative to maximize the potential of NIR spectroscopy as monitoring tool. Both regression models over performed PLSR in normal operating conditions. It was SVR which is likely the best option since it better managed inter-batch heterogeneity (more accurate and precise predictions) due to their more specificity capacity. This enhanced performance could be explained by the fact that SVR is a nonlinear sample-based regression method, which was

particularly useful for proper inter-batch heterogeneity management. However, this nature was also counterproductive in particular cases, as for example changes in the culture media or process operation under abnormal operating conditions, where ANNR is likely the most robust method since it globally generalized the process dynamics.

Perspectives

These results provided a wide frame to follow up new chemometric strategies which enhanced the capacity of *in situ* analyzers so that effective control approaches could eventually be implemented to guarantee quality of antibodies. In future investigations, it might be possible to use more chemometric tools such as local approaches of SVR and ANNR methods as well as more sophisticated versions them. Moreover, the use of such advanced monitoring procedures could allow monitoring of key compound concentration with high accuracy and precision so that metabolic rates and their relationships be also monitored in real time manner. This information could then be used for better characterization of the cell physiological state. Then, control strategies could be launched seeking maintenance of a particular cell physiological state enhancing cell culture processes yield and productivity.

4. *Monitoring of innovative parameters for a better understanding of cell cultures processes heterogeneity*

Currently, only some basic CPP are systematically monitored (pH, temperature, among others) by in-line process analyzers. However, cell culture processes also require monitoring of CPP and CQA with biochemical and biological nature. Most of the work has been focused to some few biochemical compounds such as glucose, lactate and glutamine concentration monitoring, while innovative parameters characterizing heterogeneity, such as mAb glycosylation or plant cell differentiation, had not been yet addressed. Consequently, monitoring tools must be updated considering the needs of cell cultures processes. Therefore,

the feasibility of *in situ* NIR calibration models for monitoring such innovative parameter has been explored.

Plant suspension cultures are highly heterogenic in nature and it strongly determined the process performance. The feasibility of *in situ* monitoring of some plant suspension culture heterogeneity in bioreactors through NIR spectroscopy has been preliminary demonstrated. This affirmation is based on the generation of calibration models for parenchyma, collenchyma and sclerenchyma differentiated cells of *Catharanthus roseus*. Models yielded appropriate accuracy and precision to estimate the concentration of differentiated cells during bioreactor cultures in real-time, which could provide meaningful information of physiological state of cells.

Perspectives

Plant cell suspension culture is gaining interest as biopharmaceutical production platform. However, plant cell cultures are highly heterogenic and there are still many parameters required for the adoption of QbD in suspension cultures, particularly due to particular plant cells proliferation nature and metabolism. There is then abundant room for further progress, as for example for monitoring cell aggregates size, the extension of polyclonal heterogeneity and of cells in G_0 phase, intra- or extracellular product concentration, among others.

Monitoring of mAb heterogeneity within CHO cell cultures has been a complex challenge for the implementation of real-time retro control systems since current approaches to elucidate mAb variants require sampling and labour-intensive efforts. Thus, glycosylation analysis is often performed at the end of the culture process, limiting real-time control strategies. Therefore, the development of *in situ* spectroscopic procedures for real-time monitoring of mAb heterogeneity is highly desired so that control strategies could be established once and for all. Results have proven that this can be possible provided proper generation of calibration methods based on *in situ* NIR spectroscopy. Indeed, common calibration approaches based on PLSR regression did not performed well as we expected.

While it has needed to use a combination of models based on LWR and SVR methods. Then, real-time monitoring of glycosylation, in terms of high mannose isoforms, fucosylated, sialylated and galactosylated isoforms as well as non-glycosylated mAb, was successfully performed with relative error of predictions in the order of 10 %.

Perspectives

This study provided important insights into new PAT applications through NIRS. It suggested that monitoring of innovative parameters characterizing cell culture processes heterogeneity is possible provided regression methods and chemometrics tools are properly employed. For instance, for monitoring more mAb properties, such as mAb coupling with risky host cell proteins, mAb aggregation, mAb reduction, among others.

Globally, this work has contributed for expanding the capabilities of *in situ* NIR spectroscopy for the monitoring of classic CPP in a more precise way, new innovative CPP such as cell differentiation in plant suspension cultures and innovative CQA such as mAb glycosylation profiles linked to mAb clinical characteristics in CHO cells cultures. This contribution is only a little part of an exciting and necessary investigation for the development of new spectroscopic approaches leading to the development of new optimized biopharmaceutic processes with automated control based on knowledge.

6 REFERENCES

Abu-Absi NR, Kenty BM, Cuellar ME, et al (2011) Real-time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. Biotechnol Bioeng 108:1215–1221. https://doi.org/10.1002/bit.23023

Agranovich VM, Kamchatnov AM (1999) Fermi Resonance Nonlinear Waves and Solitons in Organic Superlattices. In: Boardman AD, Pavlov L, Tanev S (eds) Advanced Photonics with Second-Order Optically Nonlinear Processes. Springer Netherlands, Dordrecht, pp 251–275

Alam MM, Naeem M, Khan MMA, Uddin M (2017) Vincristine and Vinblastine Anticancer Catharanthus Alkaloids: Pharmacological Applications and Strategies for Yield Improvement. In: Naeem M, Aftab T, Khan MMA (eds) Catharanthus roseus. Springer International Publishing, Cham, pp 277–307

Al-Fageeh MB, Marchant RJ, Carden MJ, Smales CM (2006) The cold-shock response in cultured mammalian cells: Harnessing the response for the improvement of recombinant protein production. Biotechnol Bioeng 93:829–835. https://doi.org/10.1002/bit.20789

Altamirano C, Paredes C, Cairo JJ, Godia F (2000) Improvement of CHO Cell Culture Medium Formulation: Simultaneous Substitution of Glucose and Glutamine. Biotechnol Prog 16:69–75. https://doi.org/10.1021/bp990124j

Alvarez M, Tremintin G, Wang J, et al (2011) On-line characterization of monoclonal antibody variants by liquid chromatography–mass spectrometry operating in a two-dimensional format. Analytical Biochemistry 419:17–25. https://doi.org/10.1016/j.ab.2011.07.033

Alvi K (2007) Cell Culture Technology for Pharmaceutical and Cell-Based Therapies. J Nat Prod 70:712–713. https://doi.org/10.1021/np078140a

Anscombe FJ (1973) Graphs in Statistical Analysis. The American Statistician 27:17. https://doi.org/10.2307/2682899

Arnold SA, Crowley J, Woods N, et al (2003) In-situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. Biotechnol Bioeng 84:13–19. https://doi.org/10.1002/bit.10738

Ataei-Azimi azra, Hashemloian BD, Ebrahimzadeh H, Majd A (2018) High in vitro production of ant-canceric indole alkaloids from periwinkle (Catharanthus roseus) tissue culture. Afr J Biotechnol 7:2834–2839

Awad M, Khanna R (2015) Support Vector Regression. In: Efficient Learning Machines. Apress, Berkeley, CA, pp 67–80

Bailey LA, Hatton D, Field R, Dickson AJ (2012) Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. Biotechnol Bioeng 109:2093–2103. https://doi.org/10.1002/bit.24485

Balabin RM, Lomakina EI (2011) Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. Analyst 136:1703. https://doi.org/10.1039/c0an00387e

Balabin RM, Safieva RZ, Lomakina EI (2007) Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. Chemometrics and Intelligent Laboratory Systems 88:183–188. https://doi.org/10.1016/j.chemolab.2007.04.006

Barkat MdA, Abul H, Rahman MdA (2017) Agricultural, Pharmaceutical, and Therapeutic Interior of Catharanthus roseus (L.) G. Don. In: Naeem M, Aftab T, Khan MMA (eds) Catharanthus roseus. Springer International Publishing, Cham, pp 71–100

Beck A, Wagner-Rousset E, Ayoub D, et al (2013) Characterization of Therapeutic Antibodies and Related Products. Anal Chem 85:715–736.

https://doi.org/10.1021/ac3032355

Berry BN, Dobrowsky TM, Timson RC, et al (2016) Quick generation of Raman spectroscopy based in-process glucose control to influence biopharmaceutical protein product quality during mammalian cell culture: Biotechnol. Prog. Biotechnol Progress 32:224–234. https://doi.org/10.1002/btpr.2205

Bhatia H, Mehdizadeh H, Drapeau D, Yoon S (2018) In-line monitoring of amino acids in mammalian cell cultures using raman spectroscopy and multivariate chemometrics models. Eng Life Sci 18:55–61. https://doi.org/10.1002/elsc.201700084

Bhatia S, Sharma K (2015) Technical Glitches in Micropropagation. In: Modern Applications of Plant Biotechnology in Pharmaceutical Sciences. Elsevier, pp 393–404

Blanco M, Coello J, Iturriaga H, et al (1999) Calibration in non-linear near infrared reflectance spectroscopy: a comparison of several methods. Analytica Chimica Acta 384:207–214. https://doi.org/10.1016/S0003-2670(98)00814-9

Borggaard Claus, Thodberg HHenrik (1992) Optimal minimal neural interpretation of spectra. Anal Chem 64:545–551. https://doi.org/10.1021/ac00029a018

Bourgaud F, Gravot A, Milesi S, Gontier E (2001) Production of plant secondary metabolites: a historical perspective. Plant Science 161:839–851. https://doi.org/10.1016/S0168-9452(01)00490-3

Brereton RG (2007) Applied Chemometrics for Scientists. John Wiley & Sons, Ltd, Chichester, UK

Brewster VL, Ashton L, Goodacre R (2011) Monitoring the Glycosylation Status of Proteins Using Raman Spectroscopy. Anal Chem 83:6074–6081. https://doi.org/10.1021/ac2012009

Brudzewski K, Kesik A, Kolodziejczyk K, et al (2006) Gasoline quality prediction using gas chromatography and FTIR spectroscopy: An artificial intelligence approach. Fuel 85:. https://doi.org/10.1016/j.fuel.2005.07.019

Bunniran S, McCaffrey DJ, Bentley JP, Bouldin AS (2009) Pharmaceutical product withdrawal: Attributions of blame and its impact on trust. Research in Social and Administrative Pharmacy 5:262–273. https://doi.org/10.1016/j.sapharm.2008.10.002

Burlat V, Oudin A, Courtois M, et al (2004) Co-expression of three MEP pathway genes and *geraniol 10-hydroxylase* in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. The Plant Journal 38:131–141. https://doi.org/10.1111/j.1365-313X.2004.02030.x

Burnina I, Hoyt E, Lynaugh H, et al (2013) A cost-effective plate-based sample preparation for antibody N-glycan analysis. Journal of Chromatography A 1307:201–206. https://doi.org/10.1016/j.chroma.2013.07.104

Burns DA, Ciurczak EW (eds) (2008) Handbook of near-infrared analysis, 3rd ed. CRC Press, Boca Raton

Cederkvist HR, Aastveit AH, Næs T (2005) A comparison of methods for testing differences in predictive ability. J Chemometrics 19:500–509. https://doi.org/10.1002/cem.956

Centner V, de Noord OE, Massart DL (1998) Detection of nonlinearity in multivariate calibration. Analytica Chimica Acta 376:153–168. https://doi.org/10.1016/S0003-2670(98)00543-1

Centner V, Massart DL (1998) Optimization in Locally Weighted Regression. Anal Chem 70:4206–4211. https://doi.org/10.1021/ac980208r

Cervera AE, Petersen N, Lantz AE, et al (2009) Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. Biotechnol Progress NA-NA. https://doi.org/10.1002/btpr.280

Chang S-Y, Baughman EH, McIntosh BC (2001) Implementation of Locally Weighted Regression to Maintain Calibrations on FT-NIR Analyzers for Industrial Processes. Appl Spectrosc 55:1199–1206. https://doi.org/10.1366/0003702011953207

Chattopadhyay S, Farkya S, Srivastava AK, Bisaria VS (2002) Bioprocess considerations for production of secondary metabolites by plant cell suspension cultures. Biotechnol Bioprocess Eng 7:138–149. https://doi.org/10.1007/BF02932911

Chotigeat W, Watanapokasin Y, Mahler S, Gray PP (1994) Role of environmental conditions on the expression levels, glycoform pattern and levels of sialyltransferase for hFSH produced by recombinant CHO cells. Cytotechnology 15:217–221. https://doi.org/10.1007/BF00762396

Chu L, Robinson DK (2001) Industrial choices for protein production by large-scale cell culture. Current Opinion in Biotechnology 12:180–187. https://doi.org/10.1016/S0958-1669(00)00197-X

Chusainow J, Yang YS, Yeo JHM, et al (2009) A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? Biotechnol Bioeng 102:1182–1196. https://doi.org/10.1002/bit.22158

Claßen J, Aupert F, Reardon KF, et al (2017) Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application. Anal Bioanal Chem 409:651–666. https://doi.org/10.1007/s00216-016-0068-x

Clavaud M, Roggo Y, Von Daeniken R, et al (2013) Chemometrics and in-line near infrared spectroscopic monitoring of a biopharmaceutical Chinese hamster ovary cell culture: Prediction of multiple cultivation variables. Talanta 111:28–38. https://doi.org/10.1016/j.talanta.2013.03.044

Clementschitsch F, Bayer K (2006) Improvement of bioprocess monitoring: Development of novel concepts. Microb Cell Fact 5:19. https://doi.org/10.1186/1475-2859-5-19

Cleveland WS, Devlin SJ (1988) Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association 83:596–610. https://doi.org/10.1080/01621459.1988.10478639

Cogdill RP, Dardenne P (2004) Least-Squares Support Vector Machines for Chemometrics:

An Introduction and Evaluation. Journal of Near Infrared Spectroscopy 12:93–100. https://doi.org/10.1255/jnirs.412

Courtès F, Ebel B, Guédon E, Marc A (2016) A dual near-infrared and dielectric spectroscopies strategy to monitor populations of Chinese hamster ovary cells in bioreactor. Biotechnol Lett 38:745–750. https://doi.org/10.1007/s10529-016-2036-0

Curtis WR, Emery AH (1993) Plant cell suspension culture rheology. Biotechnol Bioeng 42:520–526. https://doi.org/10.1002/bit.260420416

De Luca V, Cutler AJ (1987) Subcellular Localization of Enzymes Involved in Indole Alkaloid Biosynthesis in *Catharanthus roseus*. Plant Physiol 85:1099–1102. https://doi.org/10.1104/pp.85.4.1099

del Val IJ, Kontoravdi C, Nagy JM (2010) Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns. Biotechnol Progress 26:1505–1527. https://doi.org/10.1002/btpr.470

del Val IJ, M. P, Exley K, et al (2012) Application of Quality by Design Paradigm to the Manufacture of Protein Therapeutics. In: Petrescu S (ed) Glycosylation. InTech

Deus-Neumann B, Zenk M (1984) Instability of Indole Alkaloid Production in *Catharanthus roseus* Cell Suspension Cultures. Planta Med 50:427–431. https://doi.org/10.1055/s-2007-969755

Devillard C, Walter C (2014) Formation of plant tracheary elements in vitro – a review. NZ j of For Sci 44:22. https://doi.org/10.1186/s40490-014-0022-7

Devos O, Ruckebusch C, Durand A, et al (2009) Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation. Chemometrics and Intelligent Laboratory Systems 96:27–33. https://doi.org/10.1016/j.chemolab.2008.11.005

Doherty M, Bones J, McLoughlin N, et al (2013) An automated robotic platform for rapid

profiling oligosaccharide analysis of monoclonal antibodies directly from cell culture. Analytical Biochemistry 442:10–18. https://doi.org/10.1016/j.ab.2013.07.005

Dong J, Migliore N, Mehrman SJ, et al (2016) High-Throughput, Automated Protein A Purification Platform with Multiattribute LC–MS Analysis for Advanced Cell Culture Process Monitoring. Anal Chem 88:8673–8679. https://doi.org/10.1021/acs.analchem.6b01956

Dwek RA, Lellouch AC, Wormald MR (1995) Glycobiology: "the function of sugar in the IgG molecule". J Anat 187:279–292

Ecker DM, Jones SD, Levine HL (2015) The therapeutic monoclonal antibody market. mAbs 7:9–14. https://doi.org/10.4161/19420862.2015.989042

European Medicine Agency (2014) Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations

Fan Y, Jimenez Del Val I, Müller C, et al (2015) Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation: Process-Dependent IgG Glycosylation. Biotechnol Bioeng 112:521–535. https://doi.org/10.1002/bit.25450

Ferrão MF, Godoy SC, Gerbase AE, et al (2007) Non-destructive method for determination of hydroxyl value of soybean polyol by LS-SVM using HATR/FT-IR. Analytica Chimica Acta 595:114–119. https://doi.org/10.1016/j.aca.2007.02.066

Fischer R, Stoger E, Schillberg S, et al (2004) Plant-based production of biopharmaceuticals. Current Opinion in Plant Biology 7:152–158. https://doi.org/10.1016/j.pbi.2004.01.007

Fleischaker RJ, Sinskey AJ (1981) Oxygen demand and supply in cell culture. European J Appl Microbiol Biotechnol 12:193–197. https://doi.org/10.1007/BF00499486

Forthofer RN, Lee ES (1995) Tests of Hypotheses. In: Introduction to Biostatistics. Elsevier, pp 251–265

Fukuda H, Ito M, Sugiyama M, Komamine A (1994) Mechanisms of the proliferation and differentiation of plant cells in cell culture systems. Int J Dev Biol 38:287–299

Gao J, Lee JM (1992) Effect of oxygen supply on the suspension culture of genetically modified tobacco cells. Biotechnol Prog 8:285–290. https://doi.org/10.1021/bp00016a004

Ge Zhihong, Cavinato AG, Callis JB (1994) Noninvasive Spectroscopy for Monitoring Cell Density in a Fermentation Process. Anal Chem 66:1354–1362. https://doi.org/10.1021/ac00080a023

Gemperline PJ, Long JR, Gregoriou VG (1991) Nonlinear multivariate calibration using principal components regression and artificial neural networks. Anal Chem 63:2313–2323. https://doi.org/10.1021/ac00020a022

Grabowski GA, Golembo M, Shaaltiel Y (2014) Taliglucerase alfa: An enzyme replacement therapy using plant cell expression technology. Molecular Genetics and Metabolism 112:1–8. https://doi.org/10.1016/j.ymgme.2014.02.011

Grilo AL, Mantalaris A (2019) The Increasingly Human and Profitable Monoclonal Antibody Market. Trends in Biotechnology 37:9–16. https://doi.org/10.1016/j.tibtech.2018.05.014

Gulevsky AK, Akhatova YS, Shchenyavsky II (2017) Features of Apoptosis, Induced by Temperature Reduction. Probl Cryobiol Cryomed 27:97–109. https://doi.org/10.15407/cryo27.02.097

Gupta MM, Singh DV, Tripathi AK, et al (2005) Simultaneous Determination of Vincristine, Vinblastine, Catharanthine, and Vindoline in Leaves of Catharanthus roseus by High-Performance Liquid Chromatography. Journal of Chromatographic Science 43:450–453. https://doi.org/10.1093/chromsci/43.9.450

Hagan MT, Demuth HB, Beale MH, De Jesus O (2004) Neural network design, 2nd edition. Amazon Fulfillment Poland Sp. z o.o, Wrocław

Hakemeyer C, Strauss U, Werz S, et al (2012) At-line NIR spectroscopy as effective PAT monitoring technique in Mab cultivations during process development and manufacturing. Talanta 90:12–21. https://doi.org/10.1016/j.talanta.2011.12.042

Hall RD, Yeoman MM (1987) Intercellular and Intercultural Heterogeneity in Secondary Metabolite Accumulation in Cultures of *Catharanthus roseus* following Cell Line Selection. J Exp Bot 38:1391–1398. https://doi.org/10.1093/jxb/38.8.1391

Hellwig S, Drossard J, Twyman RM, Fischer R (2004) Plant cell cultures for the production of recombinant proteins. Nat Biotechnol 22:1415–1422. https://doi.org/10.1038/nbt1027

Henninot A, Terrier A, Charton J, et al (2015) Characterization of monoclonal antibodies by a fast and easy liquid chromatography–mass spectrometry time-of-flight analysis on culture supernatant. Analytical Biochemistry 491:52–54. https://doi.org/10.1016/j.ab.2015.08.006

Henriques JG, Buziol S, Stocker E, et al (2009) Monitoring Mammalian Cell Cultivations for Monoclonal Antibody Production Using Near-Infrared Spectroscopy. In: Rao G (ed) Optical Sensor Systems in Biotechnology. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 29–72

Hisiger S, Jolicoeur M (2008) Plant Cell Culture Monitoring Using an in Situ Multiwavelength Fluorescence Probe. Biotechnol Progress 21:580–589. https://doi.org/10.1021/bp049726f

Hoekstra SS, Harkes PAA, Verpoorte R, Libbenga KR (1990) Effect of auxin on cytodifferentiation and production of quinoline alkaloids in compact globular structures of Cinchona ledgeriana. Plant Cell Reports 8:571–574. https://doi.org/10.1007/BF00270055

Holland T, Blessing D, Hellwig S, Sack M (2013) The in-line measurement of plant cell biomass using radio frequency impedance spectroscopy as a component of process analytical technology. Biotechnology Journal n/a-n/a. https://doi.org/10.1002/biot.201300125

Höskuldsson A (1988) PLS regression methods. J Chemometrics 2:211–228. https://doi.org/10.1002/cem.1180020306

Hossler P (2011) Protein Glycosylation Control in Mammalian Cell Culture: Past Precedents and Contemporary Prospects. In: Hu WS, Zeng A-P (eds) Genomics and Systems Biology of Mammalian Cell Culture. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 187–219

Hossler P, Khattak SF, Li ZJ (2009) Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology 19:936–949. https://doi.org/10.1093/glycob/cwp079

Huang J, Romero-Torres S, Moshgbar M (2010) Practical considerations in data pre-treatment for NIR and Raman spectroscopy. Am Pharm Rev 13:116–127

Huang T-K, McDonald KA (2009) Bioreactor engineering for recombinant protein production in plant cell suspension cultures. Biochemical Engineering Journal 45:168–184. https://doi.org/10.1016/j.bej.2009.02.008

Huhn C, Selman MHJ, Ruhaak LR, et al (2009) IgG glycosylation analysis. Proteomics 9:882–913. https://doi.org/10.1002/pmic.200800715

I.M. A, M. A (2012) Plant Tissue Culture Media. In: Leva A (ed) Recent Advances in Plant in vitro Culture. InTech

International Conference on Harmonisation (1996) Guideline for Industry: ICH-Q2A, Validation of Analytical Procedures

Iskandar NN, Iriawati I (2016) Vinblastine and Vincristine Production on Madagascar Periwinkle (Catharanthus roseus (L.) G. Don) Callus Culture Treated with Polethylene Glycol. Makara J Sci 20:7–16. https://doi.org/10.7454/mss.v20i1.5656

Jenzsch M, Bell C, Buziol S, et al (2017) Trends in Process Analytical Technology: Present State in Bioprocessing. In: Kiss B, Gottschalk U, Pohlscheidt M (eds) New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins. Springer International Publishing, Cham, pp 211–252

Kaiser J (2008) Is the Drought Over for Pharming? Science 320:473–475. https://doi.org/10.1126/science.320.5875.473

Kalidass C, Mohan VR, Arjunan D (2010) Effect of auxin and cytokinin on vincristine production by callus cultures of Catharanthus roseus L.(Apocynaceae). Tropical and Subtropical Agroecosystems 283–288

Kayser V, Chennamsetty N, Voynov V, et al (2011) Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. Biotechnology Journal 6:38–44. https://doi.org/10.1002/biot.201000091

Khajehsharifi H, Pourbasheer E, Tavallali H, et al (2017) The comparison of partial least squares and principal component regression in simultaneous spectrophotometric determination of ascorbic acid, dopamine and uric acid in real samples. Arabian Journal of Chemistry 10:S3451–S3458. https://doi.org/10.1016/j.arabjc.2014.02.006

Kildegaard HF, Fan Y, Sen JW, et al (2016) Glycoprofiling effects of media additives on IgG produced by CHO cells in fed-batch bioreactors: Glycoprofiling Effects of Media Additives on IgG. Biotechnol Bioeng 113:359–366. https://doi.org/10.1002/bit.25715

Kim JY, Kim Y-G, Lee GM (2012) CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl Microbiol Biotechnol 93:917–930. https://doi.org/10.1007/s00253-011-3758-5

Kim M, O'Callaghan PM, Droms KA, James DC (2011) A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol Bioeng 108:2434–2446. https://doi.org/10.1002/bit.23189

Kim S, Jung K, Kwak S, Liu J (1994a) Relationship between cell morphology and indole alkaloid production in suspension cultures of Catharanthus roseus. Plant Cell Reports 14:. https://doi.org/10.1007/BF00233292

Kim SukW, Song NamH, Jung KyungH, et al (1994b) High frequency plant regeneration from anther-derived cell suspension cultures via somatic embryogenesis in Catharanthus roseus. Plant Cell Reports 13:. https://doi.org/10.1007/BF00232629

King PJ (1976) Studies on the Growth in Culture of Plant Cells: XX. UTILIZATION OF

2,4-DICHLOROPHENOXYACETIC ACID BY STEADY-STATE CELL CULTURES OF *ACER PSEUDOPLATANUS* L. J Exp Bot 27:1053–1072. https://doi.org/10.1093/jxb/27.5.1053

Knobloch K-H, Bast G, Berlin J (1982) Medium- and light-induced formation of serpentine and anthocyanins in cell suspension cultures of Catharanthus roseus. Phytochemistry 21:591–594. https://doi.org/10.1016/0031-9422(82)83146-4

Konno Y, Kobayashi Y, Takahashi K, et al (2012) Fucose content of monoclonal antibodies can be controlled by culture medium osmolality for high antibody-dependent cellular cytotoxicity. Cytotechnology 64:249–265. https://doi.org/10.1007/s10616-011-9377-2

Kwok KH, Tsoulpha P, Doran PM (1992) Limitations associated with conductivity measurement for monitoring growth in plant tissue culture. Plant Cell Tiss Organ Cult 29:93–99. https://doi.org/10.1007/BF00033613

Lerner EA (1981) How to make a hybridoma. Yale J Biol Med 54:387–402

Li B, Ray BH, Leister KJ, Ryder AG (2013) Performance monitoring of a mammalian cell based bioprocess using Raman spectroscopy. Analytica Chimica Acta 796:84–91. https://doi.org/10.1016/j.aca.2013.07.058

Li M, Ebel B, Chauchard F, et al (2018a) Parallel comparison of in situ Raman and NIR spectroscopies to simultaneously measure multiple variables toward real-time monitoring of CHO cell bioreactor cultures. Biochemical Engineering Journal 137:205–213. https://doi.org/10.1016/j.bej.2018.06.005

Li M, Ebel B, Courtès F, et al (2016) In Situ Infrared Spectroscopy as a PAT Tool of Great Promise for Real-Time Monitoring of Animal Cell Culture Processes. Austin J Anal Pharm Chem 3:1–5

Li M-Y, Ebel B, Paris C, et al (2018b) Real-time monitoring of antibody glycosylation site occupancy by *in situ* Raman spectroscopy during bioreactor CHO cell cultures. Biotechnol Progress 34:486–493. https://doi.org/10.1002/btpr.2604

Lindsey K, Yeoman MM (1983) The Relationship between Growth Rate, Differentiation and Alkaloid Accumulation in Cell Cultures. J Exp Bot 34:1055–1065. https://doi.org/10.1093/jxb/34.8.1055

Lingg N, Zhang P, Song Z, Bardor M (2012) The sweet tooth of biopharmaceuticals: Importance of recombinant protein glycosylation analysis. Biotechnology Journal 7:1462–1472. https://doi.org/10.1002/biot.201200078

Liu B, Villacres-Barragan C, Lattova E, et al (2013) Differential affects of low glucose on the macroheterogeneity and microheterogeneity of glycosylation in CHO-EG2 camelid monoclonal antibodies. BMC Proc 7:P112, 1753-6561-7-S6-P112. https://doi.org/10.1186/1753-6561-7-S6-P112

Liu F, He Y, Wang L (2008) Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy. Analytica Chimica Acta 610:196–204. https://doi.org/10.1016/j.aca.2008.01.039

Lourenço ND, Lopes JA, Almeida CF, et al (2012) Bioreactor monitoring with spectroscopy and chemometrics: a review. Anal Bioanal Chem 404:1211–1237. https://doi.org/10.1007/s00216-012-6073-9

Maathuis FJM (ed) (2013) Plant Mineral Nutrients: Methods and Protocols. Humana Press, Totowa, NJ

Madhusudhan R, Ramachandra Rao S, Ravishankar GA (1995) Osmolarity as a measure of growth of plant cells in suspension cultures. Enzyme and Microbial Technology 17:989–991. https://doi.org/10.1016/0141-0229(95)00011-9

Madoc-Jones H, Mauro F (1968) Interphase action of vinblastine and vincristine: Differences in their lethal action through the mitotic cycle of cultured mammalian cells. J Cell Physiol 72:185–195. https://doi.org/10.1002/jcp.1040720306

Mark H, Workman J (2007) Linearity in Calibration: Act III Scene I – Importance of

Nonlinearity. In: Chemometrics in Spectroscopy. Elsevier, pp 421–426

Markx GH, ten Hoopen HJG, Meijer JJ, Vinke KL (1991) Dielectric spectroscopy as a novel and convenient tool for the study of the shear sensitivity of plant cells in suspension culture. Journal of Biotechnology 19:145–157. https://doi.org/10.1016/0168-1656(91)90055-Z

Martens H, Nielsen JP, Engelsen SB (2003) Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. Anal Chem 75:394–404. https://doi.org/10.1021/ac020194w

Matanguihan RM, Konstantinov KB, Yoshida T (1994) Dielectric measurement to monitor the growth and the physiological states of biological cells. Bioprocess Engineering 11:213–222. https://doi.org/10.1007/BF00387695

Mattes RA, Root D, Chang D, et al (2007) In Situ Monitoring of CHO Cell Culture Medium Using Near-Infrared Spectroscopy. Bioprocess Int 46–50

Mauseth JD (2014) Botany: an introduction to plant biology, 5th ed. Jones & Bartlett Learning, Burlington, MA

McDonnell S (2015) Production of Antibodies in Hybridoma and Non-hybridoma Cell Lines. In: Al-Rubeai M (ed) Animal Cell Culture. Springer International Publishing, Cham, pp 65–88

McShane MJ, Coté GL (1998) Near-Infrared Spectroscopy for Determination of Glucose, Lactate, and Ammonia in Cell Culture Media. Appl Spectrosc 52:1073–1078. https://doi.org/10.1366/0003702981944968

Mehmood T, Liland KH, Snipen L, Sæbø S (2012) A review of variable selection methods in Partial Least Squares Regression. Chemometrics and Intelligent Laboratory Systems 118:62–69. https://doi.org/10.1016/j.chemolab.2012.07.010

Mekky H, Al-Sabahi J, Abdel-Kreem MFM (2018) Potentiating biosynthesis of the

anticancer alkaloids vincristine and vinblastine in callus cultures of Catharanthus roseus. South African Journal of Botany 114:29–31. https://doi.org/10.1016/j.sajb.2017.10.008

Meyer M, Weigelt T (1992) Interpretation of infrared spectra by artificial neural networks. Analytica Chimica Acta 265:183–190. https://doi.org/10.1016/0003-2670(92)85024-Z

Miller CE (1993) Sources of Non-Linearity in near Infrared Methods. NIR news 4:3–5. https://doi.org/10.1255/nirn.216

Milligan M, Lewin-Koh N, Coleman D, et al (2014) Semisynthetic model calibration for monitoring glucose in mammalian cell culture with in situ near infrared spectroscopy: Semisynthetic Model Calibration. Biotechnol Bioeng 111:896–903. https://doi.org/10.1002/bit.25161

Miura Y, Hirata K, Kurano N (1987) Isolation of Vinblastine in Callus Culture with Differentiated Roots of *Catharanthus roseus* (L). G. Don. Agricultural and Biological Chemistry 51:611–614. https://doi.org/10.1080/00021369.1987.10868049

Möller R, Ball RD, Henderson AR, et al (2006) Effect of Light and Activated Charcoal on Tracheary Element Differentiation in Callus Cultures of Pinus Radiata D. Don. Plant Cell Tiss Organ Cult 85:161–171. https://doi.org/10.1007/s11240-005-9065-z

Moscetti R, Massantini R, Fidaleo M (2019) Application on-line NIR spectroscopy and other process analytical technology tools to the characterization of soy sauce desalting by electrodialysis. Journal of Food Engineering 263:243–252. https://doi.org/10.1016/j.jfoodeng.2019.06.022

Murata J, Luca VD (2005) Localization of tabersonine 16-hydroxylase and 16-OH tabersonine-16-O-methyltransferase to leaf epidermal cells defines them as a major site of precursor biosynthesis in the vindoline pathway in Catharanthus roseus: Mapping biosynthesis of Catharanthus alkaloids. The Plant Journal 44:581–594. https://doi.org/10.1111/j.1365-313X.2005.02557.x

Mustafa NR, de Winter W, van Iren F, Verpoorte R (2011) Initiation, growth and

cryopreservation of plant cell suspension cultures. Nat Protoc 6:715–742. https://doi.org/10.1038/nprot.2010.144

Naaranlahti T, Lapinjoki S, Huhtikangas A, et al (1989) Mass Spectral Evidence of the Occurrence of Vindoline in Heterotrophic Cultures of *Catharanthus roseus* Cells. Planta Med 55:155–157. https://doi.org/10.1055/s-2006-961911

Næs T (ed) (2004) A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK

Naes T, Isaksson T (1992) Locally Weighted Regression in Diffuse Near-Infrared Transmittance Spectroscopy. Appl Spectrosc 46:34–43. https://doi.org/10.1366/0003702924444344

Næs T, Isaksson T (1992) Locally Weighted Regression in Diffuse Near-Infrared Transmittance Spectroscopy. Appl Spectrosc 46:34–43. https://doi.org/10.1366/0003702924444344

Naes T, Kvaal K, Isaksson T, Miller C (1993) Artificial Neural Networks in Multivariate Calibration. Journal of Near Infrared Spectroscopy 1:1–11. https://doi.org/10.1255/jnirs.1

Næs T, Martens H (1988) Principal component regression in NIR analysis: Viewpoints, background details and selection of components. J Chemometrics 2:155–167. https://doi.org/10.1002/cem.1180020207

Naes Tormod, Isaksson Tomas, Kowalski Bruce (1990) Locally weighted regression and scatter correction for near-infrared reflectance data. Anal Chem 62:664–673. https://doi.org/10.1021/ac00206a003

Neumann D, Krauss G, Hieke M, Gröger D (1983) Indole Alkaloid Formation and Storage in Cell Suspension Cultures of Catharanthus roseus. Planta Med 48:20–23. https://doi.org/10.1055/s-2007-969871

Niero R, Cechinel Filho V, Yunes RA (2018) Medicinal Plants and Phytomedicines. In:

Cechinel Filho V (ed) Natural Products as Source of Molecules with Therapeutic Potential. Springer International Publishing, Cham, pp 1–33

Nilsson NJ (1965) Learning machines: foundations of trainable pattern-classifying systems. McGraw-Hill

Noble RL (1990) The discovery of the vinca alkaloids—chemotherapeutic agents against cancer. Biochem Cell Biol 68:1344–1351. https://doi.org/10.1139/o90-197

Nocarova E, Fischer L (2009) Cloning of transgenic tobacco BY-2 cells; an efficient method to analyse and reduce high natural heterogeneity of transgene expression. BMC Plant Biol 9:44. https://doi.org/10.1186/1471-2229-9-44

Oyaas K (2003) Animal Cell Culture, Physiochemical Effects of Osmolality and Temperature. In: Spier RE (ed) Encyclopedia of Cell Technology. John Wiley & Sons, Inc., Hoboken, NJ, USA, p spi008

Ozaki Y, McClure WF, Christy AA (eds) (2006) Near-Infrared Spectroscopy in Food Science and Technology: Ozaki/Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, Inc., Hoboken, NJ, USA

Pan Q, Chen Y, Wang Q, et al (2010) Effect of plant growth regulators on the biosynthesis of vinblastine, vindoline and catharanthine in Catharanthus roseus. Plant Growth Regul 60:133–141. https://doi.org/10.1007/s10725-009-9429-1

Pareilleux A, Vinas R (1983) Influence of the aeration rate on the growth yield in suspension cultures of Catharanthus roseus (L.) G.Don. J Ferment Technol 61:429–433

Parry B (2004) Trading the Genome: Investigating the Commodification of Bio-Information. Columbia University Press, New York Chichester, West Sussex

Patil RA, Kolewe ME, Roberts SC (2013) Cellular aggregation is a key parameter associated with long term variability in paclitaxel accumulation in Taxus suspension cultures. Plant Cell Tiss Organ Cult 112:303–310. https://doi.org/10.1007/s11240-012-0237-3

Patil RA, Roberts SC (2013) Implications of Cellular Heterogeneity on Plant Cell Culture Performance. In: Chandra S, Lata H, Varma A (eds) Biotechnology for Medicinal Plants. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 207–239

Perezmarin D, Garridovaro A, Guerrero J (2007) Non-linear regression methods in NIRS quantitative analysis. Talanta 72:28–42. https://doi.org/10.1016/j.talanta.2006.10.036

Petiot E, Bernard-Moulin P, Magadoux T, et al (2010) In situ quantification of microcarrier animal cell cultures using near-infrared spectroscopy. Process Biochemistry 45:1832–1836. https://doi.org/10.1016/j.procbio.2010.08.010

Qiu J, Arnold MA, Murhammer DW (2014) On-line near infrared bioreactor monitoring of cell density and concentrations of glucose and lactate during insect cell cultivation. Journal of Biotechnology 173:106–111. https://doi.org/10.1016/j.jbiotec.2014.01.009

Rademacher TW, Parekh RB, Dwek RA (1988) Glycobiology. Annu Rev Biochem 57:785–838. https://doi.org/10.1146/annurev.bi.57.070188.004033

Rader R a (2008) (Re)defining biopharmaceutical. Nature biotechnology 26:743–751. https://doi.org/10.1038/nbt0708-743

Raju TS, Briggs JB, Borge SM, Jones AJS (2000) Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. Glycobiology 10:477–486. https://doi.org/10.1093/glycob/10.5.477

Rates SMK (2001) Plants as source of drugs. Toxicon 39:603–613. https://doi.org/10.1016/S0041-0101(00)00154-9

Rathore AS, Bhambure R, Ghare V (2010) Process analytical technology (PAT) for biopharmaceutical products. Anal Bioanal Chem 398:137–154. https://doi.org/10.1007/s00216-010-3781-x

Rathore AS, Winkle H (2009) Quality by design for biopharmaceuticals. Nat Biotechnol

27:26–34. https://doi.org/10.1038/nbt0109-26

Reichert JM, Rosensweig CJ, Faden LB, Dewitz MC (2005) Monoclonal antibody successes in the clinic. Nat Biotechnol 23:1073–1078. https://doi.org/10.1038/nbt0905-1073

Rhiel MH, Cohen MB, Arnold MA, Murhammer DW (2004) On-line monitoring of human prostate cancer cells in a perfusion rotating wall vessel by near-infrared spectroscopy. Biotechnol Bioeng 86:852–861. https://doi.org/10.1002/bit.10834

Rho D, André G (1991) Growth and stoichiometry of a *Catharanthus roseus* cell suspension culture grown under nitrogen-limiting conditions: GROWTH AND STOICHIOMETRY OF *C. ROSEUS* SUSPENSION CULTURE. Biotechnol Bioeng 38:579–587. https://doi.org/10.1002/bit.260380604

Riley MR, Arnold MA, Murhammer DW, et al (1998) Adaptive Calibration Scheme for Quantification of Nutrients and Byproducts in Insect Cell Bioreactors by Near-Infrared Spectroscopy. Biotechnol Prog 14:527–533. https://doi.org/10.1021/bp980022d

Rinnan Å, Berg F van den, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry 28:1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Rita Costa A, Elisa Rodrigues M, Henriques M, et al (2010) Guidelines to cell engineering for monoclonal antibody production. European Journal of Pharmaceutics and Biopharmaceutics 74:127–138. https://doi.org/10.1016/j.ejpb.2009.10.002

Roberts AW, Haigler CH (1994) Cell Expansion and Tracheary Element Differentiation Are Regulated by Extracellular pH in Mesophyll Cultures of Zinnia elegans L. Plant Physiol 105:699–706. https://doi.org/10.1104/pp.105.2.699

Roni CR (2019) Faced With a Drug Shortfall, Doctors Scramble to Treat Children With Cancer - Cancer Drug Shortage Imperils Children. The New York Times 21

Roychoudhury P, O'Kennedy R, McNeil B, Harvey LM (2007) Multiplexing fibre optic near

infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses. Analytica Chimica Acta 590:110–117. https://doi.org/10.1016/j.aca.2007.03.011

Ryman JT, Meibohm B (2017) Pharmacokinetics of Monoclonal Antibodies: Pharmacokinetics of Monoclonal Antibodies. CPT Pharmacometrics Syst Pharmacol 6:576–588. https://doi.org/10.1002/psp4.12224

Sagishima K, Kubota K, Ashihara H (1989) Uptake and Metabolism of Sugars by Suspension Cultured Catharanthus roseus Cells. Annals of Botany 64:185–193. https://doi.org/10.1093/oxfordjournals.aob.a087824

Saiman MZ, Mustafa NR, Pomahočová B, et al (2014) Analysis of metabolites in the terpenoid pathway of Catharanthus roseus cell suspensions. Plant Cell Tiss Organ Cult 117:225–239. https://doi.org/10.1007/s11240-014-0435-2

Sandor M, Rüdinger F, Bienert R, et al (2013) Comparative study of non-invasive monitoring via infrared spectroscopy for mammalian cell cultivations. Journal of Biotechnology 168:636–645. https://doi.org/10.1016/j.jbiotec.2013.08.002

Scarff M, Arnold SA, Harvey LM, McNeil B (2006) Near Infrared Spectroscopy for Bioprocess Monitoring and Control: Current Status and Future Trends. Critical Reviews in Biotechnology 26:17–39. https://doi.org/10.1080/07388550500513677

Schiestl M, Stangler T, Torella C, et al (2011) Acceptable changes in quality attributes of glycosylated biopharmaceuticals. Nat Biotechnol 29:310–312. https://doi.org/10.1038/nbt.1839

Schmid G, Blanch HW, Wilke CR (1990) Hybridoma growth, metabolism, and product formation in HEPES-buffered medium: II. Effect of pH. Biotechnol Lett 12:633–638. https://doi.org/10.1007/BF01088185

Scott AI, Mizukami H, Hirata T, Lee S-L (1980) Formation of catharanthine, akuammicine and vindoline in Catharanthus roseus suspension cells. Phytochemistry 19:488–489. https://doi.org/10.1016/0031-9422(80)83216-X

Scragg AH (1995) The problems associated with high biomass levels in plant cell suspensions. Plant Cell Tiss Organ Cult 43:163–170. https://doi.org/10.1007/BF00052172

Siesler HW (2010) Near-infrared spectroscopy: principles, instruments, applications

Sinha S, Pipes G, Topp EM, et al (2008) Comparison of LC and LC/MS methods for quantifying N-glycosylation in recombinant IgGs. J Am Soc Mass Spectrom 19:1643–1654. https://doi.org/10.1016/j.jasms.2008.07.004

Slutsky B (1998) Handbook of Chemometrics and Qualimetrics: Part A By D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, and J. Smeyers-Verbeke. Data Handling in Science and Technology Volume 20A. Elsevier: Amsterdam. 1997. Xvii + 867 pp. ISBN 0-444-89724-0. $293.25. J Chem Inf Comput Sci 38:1254–1254. https://doi.org/10.1021/ci980427d

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Statistics and Computing 14:199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Srivastava NK, Srivastava AK (2007) Influence of gibberellic acid on 14CO2 metabolism, growth, and production of alkaloids in Catharanthus roseus. Photosynthetica 45:156–160. https://doi.org/10.1007/s11099-007-0026-0

Srivastava S, Harsh S, Srivastava AK (2008) Use of NADH fluorescence measurement for on-line biomass estimation and characterization of metabolic status in bioreactor cultivation of plant cells for azadirachtin (a biopesticide) production. Process Biochemistry 43:1121–1123. https://doi.org/10.1016/j.procbio.2008.06.008

St-Pierre B, Vazquez-Flota FA, De Luca V (1999) Multicellular Compartmentation of *Catharanthus roseus* Alkaloid Biosynthesis Predicts Intercellular Translocation of a Pathway Intermediate. Plant Cell 11:887–900. https://doi.org/10.1105/tpc.11.5.887

Su WW, Guan P, Bugos RC (2004) High-level secretion of functional green fluorescent protein from transgenic tobacco cell cultures: Characterization and sensing. Biotechnol Bioeng 85:610–619. https://doi.org/10.1002/bit.10916

Su WW, Liu B, Lu W-B, et al (2005) Observer-based online compensation of inner filter effect in monitoring fluorescence of GFP-expressing plant cell cultures. Biotechnol Bioeng 91:213–226. https://doi.org/10.1002/bit.20510

Suzuki K (2011) Artificial Neural Networks Methodological Advances and Biomedical Applications. INTECH Open Access Publisher

Tachibana H, Taniguchi K, Ushio Y, et al (1994) Changes of monosaccharide availability of human hybridoma lead to alteration of biological properties of human monoclonal antibody. Cytotechnology 16:151–157. https://doi.org/10.1007/BF00749902

Taha H, Shams KS, Nazif NM, Seif-El-Nasr M (2014) In Vitro Studies on Egyptian Catharanthus roseus (L.) G.Don V: Impact of Stirred Reactor Physical Factors on Achievement of Cells Proliferation and Vincristine and Vinblastine Accumulation. Research Journal of Pharmaceutical, Biological and Chemical Sciences 5:330–330

Teixeira AP, Oliveira R, Alves PM, Carrondo MJT (2009a) Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative. Biotechnology Advances 27:726–732. https://doi.org/10.1016/j.biotechadv.2009.05.003

Teixeira AP, Oliveira R, Alves PM, Carrondo MJT (2009b) Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative. Biotechnology Advances 27:726–732. https://doi.org/10.1016/j.biotechadv.2009.05.003

Tikhomiroff C, Jolicoeur M (2002) Screening of Catharanthus roseus secondary metabolites by high-performance liquid chromatography. Journal of Chromatography A 955:87–93. https://doi.org/10.1016/S0021-9673(02)00204-2

Torrey JG (1975) Tracheary Element Formation from Single Isolated Cells in Culture. Physiol Plant 35:158–165. https://doi.org/10.1111/j.1399-3054.1975.tb03886.x

Twumasi P, Schel JHN, van Ieperen W, et al (2009) Establishing in vitro Zinnia elegans cell suspension culture with high tracheary element differentiation. Cell Biology International 33:524–533. https://doi.org/10.1016/j.cellbi.2009.01.019

Tyler RT, Kurz WGW, Panchuk BD (1986) Photoautotrophic cell suspension cultures of periwinkle (Catharanthus roseus (L.) G. Don): Transition from heterotrophic to photoautotrophic growth. Plant Cell Reports 5:195–198. https://doi.org/10.1007/BF00269117

U.S. Food & Drug Administration (2015) Development and Submission of Near Infrared Analytical Procedures Guidance for Industry

U.S. Food and Drug Administration (2008) FDA 101 : Biological Products. http://www.fda.gov/downloads/ForConsumers/ConsumerUpdates/ucm048367.pdf

Van Der Malsburg C (1986) Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. In: Palm G, Aertsen A (eds) Brain Theory. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 245–248

van Gulik WM, Meijer JJ, ten Hoopen HJG, et al (1989) Growth of a Catharanthus roseus cell suspension culture in a modified chemostat under glucose-limiting conditions. Appl Microbiol Biotechnol 30:270–275. https://doi.org/10.1007/BF00256217

Vapnik VN (2000) The Nature of Statistical Learning Theory. Springer New York, New York, NY

Verdú-Andrés J, Massart DL, Menardo C, Sterna C (1997) Correction of non-linearities in spectroscopic multivariate calibration by using transformed original variables and PLS regression. Analytica Chimica Acta 349:271–282. https://doi.org/10.1016/S0003-2670(97)00271-7

Verpoorte R, van der Heijden R, Schripsema J, et al (1993) Plant Cell Biotechnology for the Production of Alkaloids: Present Status and Prospects. J Nat Prod 56:186–207. https://doi.org/10.1021/np50092a003

Walsh G (1999) Biopharmaceuticals: Biochemistry and biotechnology. Pharmaceutisch Weekblad 134:551. https://doi.org/10.1590/S1516-93322005000200017

Walsh G (2002) Biopharmaceuticals and biotechnology medicines: An issue of nomenclature. Pharmaceutisch Weekblad 15:135–138. https://doi.org/10.1016/S0928-0987(01)00222-6

Walsh G (2014) Biopharmaceutical benchmarks 2014. Nat Biotechnol 32:992–1000. https://doi.org/10.1038/nbt.3040

Walsh G (2018) Biopharmaceutical benchmarks 2018. Nat Biotechnol 36:1136–1145. https://doi.org/10.1038/nbt.4305

Wang Y, Feng P, Sosic Z, Zang L (2017) Monitoring Glycosylation Profile and Protein Titer in Cell Culture Samples Using ZipChip CE-MS. J Anal Bioanal Tech 08: https://doi.org/10.4172/2155-9872.1000359

Wang Ziyi, Isaksson Tomas, Kowalski BR (1994) New approach for distance measurement in locally weighted regression. Anal Chem 66:249–260. https://doi.org/10.1021/ac00074a012

Weikert S, Papac D, Briggs J, et al (1999) Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. Nat Biotechnol 17:1116–1121. https://doi.org/10.1038/15104

Williams P, Norris KH, American Association of Cereal Chemists (eds) (1987) Near-infrared technology in the agricultural and food industries. American Association of Cereal Chemists, St. Paul, Minn., USA

Wold S, Ruhe A, Wold H, Dunn, III WJ (1984) The Collinearity Problem in-linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. SIAM J Sci and Stat Comput 5:735–743. https://doi.org/10.1137/0905052

Wong NSC, Wati L, Nissom PM, et al (2010) An investigation of intracellular glycosylation activities in CHO cells: Effects of nucleotide sugar precursor feeding. Biotechnol Bioeng 107:321–336. https://doi.org/10.1002/bit.22812

Wongsamuth R, Doran PM (1994) Foaming and cell flotation in suspended plant cell cultures and the effect of chemical antifoams. Biotechnol Bioeng 44:481–488. https://doi.org/10.1002/bit.260440411

Workman JJ (2008) NIR spectroscopy Calibration Basics. In: Handbook of Near-Infrared Analysis, 3rd ed. Taylor & Francis Group

Wu D, He Y, Feng S (2008) Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. Analytica Chimica Acta 610:232–242. https://doi.org/10.1016/j.aca.2008.01.056

Xiaobo Z, Jiewen Z, Povey MJW, et al (2010) Variables selection methods in near-infrared spectroscopy. Analytica Chimica Acta 667:14–32. https://doi.org/10.1016/j.aca.2010.03.048

Xing Z, Kenty BM, Li ZJ, Lee SS (2009) Scale-up analysis for a CHO cell culture process in large-scale bioreactors. Biotechnol Bioeng 103:733–746. https://doi.org/10.1002/bit.22287

Xu J, Zhang N (2014) On the way to commercializing plant cell culture platform for biopharmaceuticals: present status and prospect. Pharmaceutical Bioprocessing 2:499–518. https://doi.org/10.4155/pbp.14.32

Yao T, Asayama Y (2017) Animal-cell culture media: History, characteristics, and current issues. Reprod Med Biol 16:99–117. https://doi.org/10.1002/rmb2.12024

Yoon SK, Choi SL, Song JY, Lee GM (2005) Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37.0°C: Effect of pH on Erythropoietin Production by CHO Cells. Biotechnol Bioeng 89:345–356. https://doi.org/10.1002/bit.20353

Yu LX (2008) Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control. Pharm Res 25:781–791. https://doi.org/10.1007/s11095-007-9511-1

Yu LX, Amidon G, Khan MA, et al (2014) Understanding Pharmaceutical Quality by Design. AAPS J 16:771–783. https://doi.org/10.1208/s12248-014-9598-3

Yusibov V, Streatfield SJ, Kushnir N (2011) Clinical development of plant-produced recombinant pharmaceuticals: Vaccines, antibodies and beyond. Human Vaccines 7:313–321. https://doi.org/10.4161/hv.7.3.14207

Zhang P, Woen S, Wang T, et al (2016) Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. Drug Discovery Today 21:740–765. https://doi.org/10.1016/j.drudis.2016.01.006

Zhang W, Yang J, Zi J, et al (2015) Effects of Adding Vindoline and MeJA on Production of Vincristine and Vinblastine, and Transcription of their Biosynthetic Genes in the Cultured CMCs of *Catharanthus roseus*. Natural Product Communications 10:1934578X1501001. https://doi.org/10.1177/1934578X1501001220

Zhao J, Zhu W-H, Hu Q, Guo Y-Q (2001) Compact callus cluster suspension cultures of Catharanthus roseus with enhanced indole alkaloid biosynthesis. In Vitro CellDevBiol-Plant 37:68–72. https://doi.org/10.1007/s11627-001-0013-2

Zheng K, Bantog C, Bayer R (2011) The impact of glycosylation on monoclonal antibody conformation and stability. mAbs 3:568–576. https://doi.org/10.4161/mabs.3.6.17922

Zhong J-J, Fujiyama K, Seki T, Yoshida T (1993) On-line monitoring of cell concentration ofPerilla frutescens in a bioreactor. Biotechnol Bioeng 42:542–546. https://doi.org/10.1002/bit.260420420

7 ANNEX

7.1 ACCEPTED OR PUBLISHED WORKS

1. Daniel Arturo Zavala-Ortiz, Bruno Ebel, Emmanuel Guedon, Annie Marc, Dulce María Barradas-Dermitz, Patricia Margaret Hayward-Jones, María Guadalupe Aguilar-Uscanga (2019). In situ cell differentiation monitoring of *Catharanthus roseus* suspension culture processes by NIR spectroscopy. *Bioprocess and Biosystems Engineering*. https://doi.org/10.1007/s00449-019-02255-x

2. Daniel A. Zavala-Ortiz, Bruno Ebel, Meng-Yao Li, Dulce Ma. Barradas-Dermitz, Patricia M. Hayward-Jones, Maria G. Aguilar-Uscanga, Annie Marc, Emmanuel Guedon (2019). Interest of locally weighted regression to overcome nonlinear effects during in situ NIR monitoring of CHO cell culture parameters and antibody glycosylation. *Biotechnology Progress*. https://doi.org/10.1002/btpr.2924

3. Daniel A. Zavala-Ortiz, Bruno Ebel, Meng-Yao Li, Dulce Ma. Barradas-Dermitz, Patricia M. Hayward-Jones, Maria G. Aguilar-Uscanga, Annie Marc, Emmanuel Guedon (2020). Support Vector and Locally Weighted regressions to monitor monoclonal antibody glycosylation during CHO cell culture processes, an enhanced alternative to Partial Least Squares regression. *Biochemical Engineering Journal*. https://doi.org/10.1016/j.bej.2019.107457

7.2 WORKS TO BE SUBMITTED

1. Interest of cellular differentiation in the production of vincristine and vinblastine in suspension cultures of *Catharanthus roseus* (L.) G Don. Submitted to Biotechnology letters journal in January 2020.
2. Evaluation of NIR calibration models for *in situ* CHO cell cultures monitoring: PLSR, ANNR & SVR. Intended for publication in Biotechnology Journal.

7.3 COMMUNICATIONS IN CONGRESSES (POSTERS)

1. Mexican Academy of chemical engineering (AMIDIQ), 2018 (San José del Cabo, Mexico): Mejoramiento del proceso de producción de células de *Taxus globosa* S. por sincronización celular para la producción posterior del anticancerígeno paclitaxel.
2. European Symposium on Biochemical Engineering Sciences (ESBES), 2018 (Lisboa, Portugal): Interest of NIR spectroscopy to perform rapid analysis of medium compounds of plant cells cultivated in suspension
3. ESBES, 2018 (Lisboa, Portugal): Interest of the Locally Weighted Regression (LWR) to improve real-time monitoring of CHO cells and antibody quality by using in situ NIR spectroscopy
4. AMIDIQ, 2019 (Huatulco, Mexico): Interés de la heterogeneidad celular para la producción de los anticancerígenos vincristina y vinblastina por cultivos en suspensión de *Catharanthus roseus*
5. Séminaire de l'école doctorale RP2E, 2018 (Nancy, France): CHO cell culture quality monitoring by NIR Spectroscopy and LWR/LWR2
6. European society of animal science technology (ESACT), 2019 (Copenhagen, Denmark): Enhanced method to monitor cell cultures by dielectric Spectroscopy.
7. ESACT, 2019 (Copenhagen, Denmark): Chemometrics to monitor CHO cells cultures quality by in situ NIRS
8. 3éme Colloque Valorisation du vegetal, 2019 (Nancy, France): Surveillance de l'hétérogénéité cellulaire de Catharanthus roseus par spectroscopie NIR.
9. 17ème congrès de la société française de génie des procédés (SFGP) (Nantes, France): Amélioration du suivi en temps réel de l'état physiologique de cellules CHO par spectroscopie diélectrique in situ et régression à vecteurs de support
10. SFGP, 2019 (Nantes, France): Suivi en ligne de la qualité d'anticorps produits par cellules CHO par spectrométrie NIR in situ couplée à une régression à vecteurs de support

7.4 26º ESACT meeting Proceeding

Expected to be published as a supplement at BMC Proceedings.

Enhanced method to monitor cell cultures by dielectric spectroscopy

Daniel Arturo Zavala-Ortiz[1,2], Bruno Ebel[1*], Meng-Yao Li[1], Maria Guadalupe Aguilar-Uscanga[2], Javier Gomez-Rodriguez[2], Dulce Maria Barradas-Dermitz[2], Patricia Margaret Hayward-Jones[2], Annie Marc[1], Emmanuel Guedon[1]

[1]*Laboratoire Réactions et Génie des Procédés, CNRS-Université de Lorraine, Vandœuvre-lès-Nancy, France*
[2]*Laboratorio de Bioingeniería, Instituto Tecnológico de Veracruz, México*
*bruno.ebel@univ-lorraine.fr

Background

Physiological state of cells has a strong impact on post-translational modifications of biopharmaceuticals. Therefore, its accurate monitoring and control is mandatory to guarantee medicines properties and safety of patients. The specific cell growth rate (μ) may globally be used to depict the cells state during cultures. Using *in-situ* dielectric spectroscopy, μ can be used to control the feeding strategy of cell cultures so that glycosylation quality of monoclonal antibody (mAb) remained under proper levels [1]. However, the widely used simple linear regression (SLR) of measured permittivity to real-time estimate the viable cell density (VCD) and then calculate μ, can led to a lack of accuracy and precision. To avoid limitations, this study aimed to evaluate the novel implementation of Supported Vector Regression (SVR) on dielectrics.

Materials and methods

Capacitance spectra were collected every 12 min by using Biomass Evo 200 during several mAb producing CHO cells cultures in bioreactors (2L). Off-line measurements of VCD were performed using the Trypan Blue exclusion method (ViCell, Beckman Coulter). Then VCD values were related to measured permittivities by both regression methods (SLR and SVR) to build prediction models. The SLR model was generated using the linear equation:

$$VCD = a \times (permittivity_{1000\ kHz}) + b$$
where a and b are fitting coefficients

The SVR model was generated using an □-support vector regression with a Gaussian radial basis function kernel (PLS_Toolbox® 8.2.1, Eigenvector Research). Then, the *in-line* estimated VCD was used to calculate μ values in real-time, and compared them to off-line values for evaluation:

$$\mu = \frac{\Delta\, Ln\,(VCD)}{\Delta t}$$

Results

The SLR uses a variable (permittivity) space to perform regression. A consideration for such model is that cell properties, as well as relation of VCD to permittivity spectra, remain constant. This is not the case during the various phases of culture process. Consequently, the SLR cannot well relate cell subpopulations displaying different dielectric properties, mainly during late-stationary and dead phases. On the contrary, the SVR method, based on a sample space, demonstrated a remarkable robustness for tracking cells having different dielectric properties (Fig. 1-a). SVR creates a sample distribution based on dielectric properties before generating the regression equation, allowing considering variability of cell dielectric properties within samples. Then, both SLR and SVR models were used to in-line calculate μ. The SVR method leads to more accurate and stable calculations of μ during cultures (Fig. 1-b).

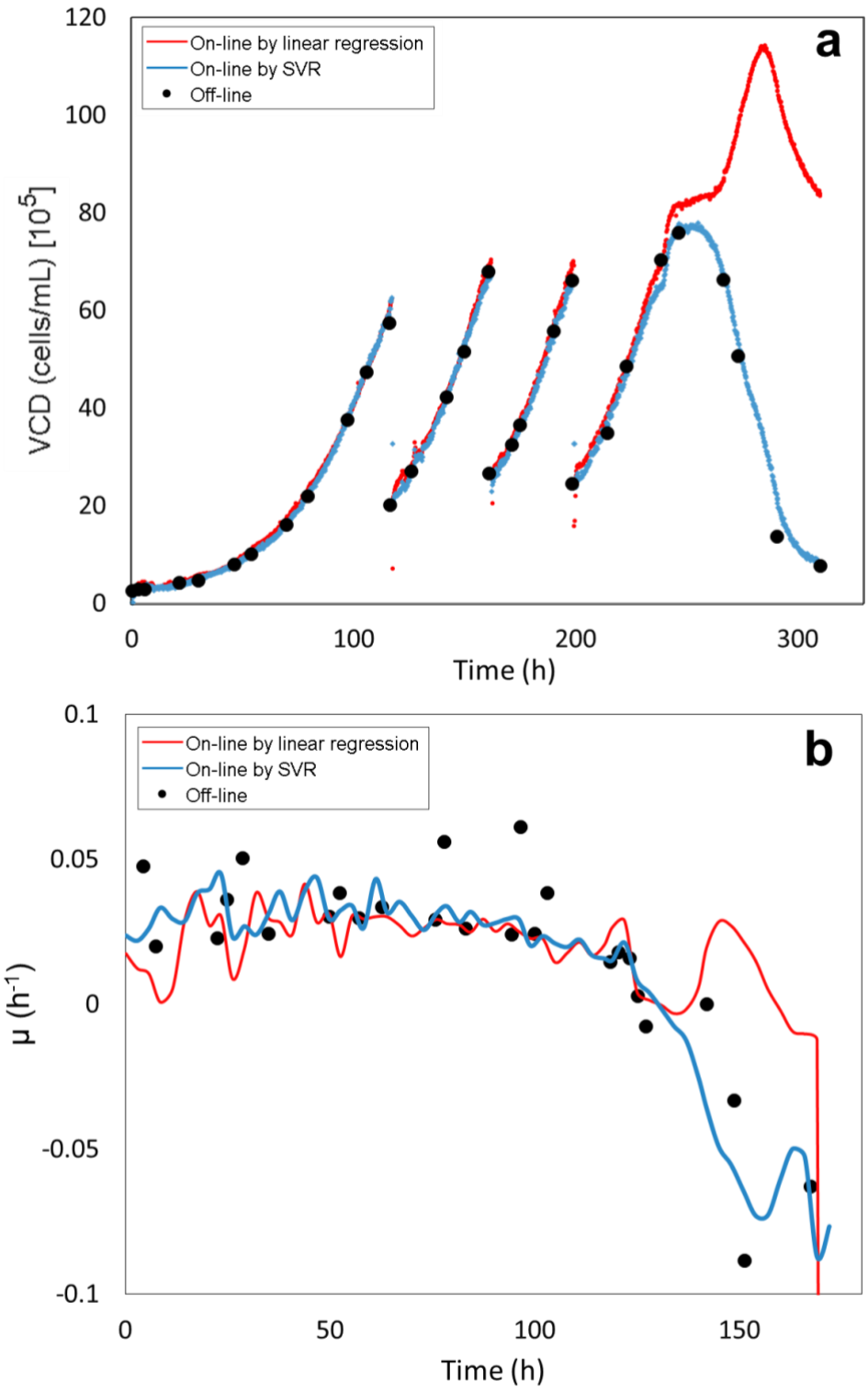

Figure 1. Use of SVR and SLR models to monitor (a) VCD in fed-harvest culture and (b) μ in batch culture.

Conclusions

On the basis of results, the SVR method must be favoured in order to improve the monitoring of animal cell culture processes by dielectric spectroscopy, especially when the composition of the culture medium changes significantly or when cells are subject to strong physiological changes. SVR also appears promising to develop new approaches taking into account the different dielectric subpopulations of cells rather than assuming a homogeneous dielectric population within cell cultures in bioreactor [2].

Acknowledgements

The authors acknowledge CONACyT and Campus France which granted Daniel. A. Zavala-Ortiz, and the French National Agency for Research (ANR) for supporting the ProCell-In-Line project.

References

[1] Li M-Y, Ebel B, Blanchard F, Paris C, Guedon E, Marc A. Control of IgG glycosylation by *in situ* and real-time estimation of specific growth rate of CHO cells cultured in bioreactor. Biotechnol Bioeng. 2019; 116(5):945-1252.

[2] Braasch K, Nikolic-Jaric M, Cabel T, Salimi E, Bridges GE, Thomson DJ, Butler M. The changing dielectric properties of CHO cells can be used to determine early apoptotic events in a bioprocess. Biotechnol Bioeng. 2013; 110(11): 2902–2914.

7.5 Transversal formation during the PhD

Scientific exchanges:

Short term internship in the laboratory of vegetal tissues, Insitute of Ecology, Xalapa (January 2016): Development of plant cell lines of *Taxus globosa* and *Catharanthus roseus*.

In the frame of the joint PhD between Université de Lorraine, and TecNM/Instituto Tecnológico de Veracruz, there were three exchanges between both institutions during the doctorate program. 22 months were spent in Université de Lorraine and 26 months in TecNM/Instituto Tecnológico de Veracruz.

Teaching:

- Guest Lecturer: Fundamentals of plant cell cultures, ITVer (Spring 2016, 20 h)
- Junior Lecturer: Fundamentals of Instrumental analysis, ITVer (February – June 2016; 50 h)

Doctoral school courses (110 h in total):

1. Français langue étrangère (CEFR B1 level) – Spring 2017
2. Culture de l'intégrité scientifique – Autumn 2019
3. Data science'- Initiation à l'analyse multivariée de données – Spring 2019.
4. Molecules, medicaments, matériaux: Elaboration, synthèse et characterization – Spring 2019.
5. Maîtriser les situations interculturelles : les comprendre, gérer les interactions au sein d'une équipe, prendre conscience des présupposés scientifiques de sa culture – Spring 2018.
6. Module Pole A2F: Techniques de caractérisation et d'analyses – Spring 2018.
7. Valorisation non alimentaire des productions vegetales – Spring 2018.

Industrial visits:

1. Ames research center, NASA, California U.S.A. - Autum 2017.
2. Grupo Bático, Veracruz, Mexico – Summer 2016.
3. Badische Anilin- & Soda-Fabrik (BASF Beauty), Pulnoy, France – Spring 2019
4. LUXINNOVATION, Esch sur Belval, Luxembourg – Spring 2019.
5. Novartis, Bâle, Switzerland – Summer 2019.

Awards:

- 1st place in the Evento Nacional Estudiantil de Innovación Tecnológica-ENEIT (National Student Technological Innovation Event, Mexico) 2017 in category “Industrial process design and financial proposal”
- 1st place in the Evento Nacional Estudiantil de Innovación Tecnológica-ENEIT (National Student Technological Innovation Event, Mexico) 2017 as best industrial project proposal.
- Laureate of the Eiffel Excellence Scholarship Programme 2018 by the French ministère de l'europe et des affaires étrangères
- 1st prize on poster contest during the 26th meeting of the European Society of Animal Cell Technology, Copenhagen, Denmark, 2019.
- Best thesis presentation of the Graduate School Unidad de Investigación y Desarrollo en Alimentos (UNIDA), January 2020.

Other:

- Participation in the Evento Nacional Estudiantil de Innovación Tecnológica -ENEIT (National Contest of Innovation), Veracruz, Veracruz & Pachuca, Hidalgo, Mexico. 2017 – 2018.
- Participation in the design and redaction of scientific proposals: ECOS Nord, Retos Nacionales (CONACyT).
- Participation in the design and redaction of two Mexican patent applications in 2017 and 2019.