Approche méthodologique innovante pour le suivi en ligne de procédés de production d'anticorps par cellules animales : apport des techniques spectroscopiques in situ à la stratégie PAT

Mengyao Li

▶ **To cite this version:**

Mengyao Li. Approche méthodologique innovante pour le suivi en ligne de procédés de production d'anticorps par cellules animales : apport des techniques spectroscopiques in situ à la stratégie PAT. Génie des procédés. Université de Lorraine, 2018. Français. NNT : 2018LORR0151 . tel-01943820

HAL Id: tel-01943820

https://theses.hal.science/tel-01943820

Submitted on 4 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

THÈSE

présentée à l'Université de Lorraine
Ecole Doctorale (anciennement R.P.2E. : Ressources, Procédés, Produits, Environnement)
Ecole Doctorale SIMPPE : Science et ingénerie des molécules, dees procédés, des produits et de l'énergie
Laboratoire Réactions et Génie des Procédés (UMR 7274 CNRS)
pour l'obtention du grade de

Docteur de l'Université de Lorraine
Spécialité : Procédés Biotechnologiques et Alimentaires

par

Mengyao LI

Approche méthodologique innovante pour le suivi en ligne de procédés de production d'anticorps par cellules animales : apport des techniques spectroscopiques *in situ* à la stratégie PAT

Soutenance prévue le 9 Novembre 2018 devant la Commission d'Examen

Rapporteurs : Gérald THOUAND – Université de Nantes, France
Carole MOLINA-JOUVE – INSA Toulouse, France

Examinateurs : Cécile LEGALLAIS – CNRS-Compiègne, France
Emmanuel GUEDON – CNRS-Nancy, France
Bruno EBEL – Université de Lorraine, France (Co-Directeur de thèse)
Annie MARC – CNRS-Nancy, France (Directeur de thèse)

Invités : Frantz FOURNIER – Université de Lorraine, France
Anne GREGOIRE – REVIMMUNE SAS, France

Remerciements

Ce projet de thèse a été réalisé au Laboratoire Réactions et Génie des Procédés (LRGP), et au sein de l'équipe Bioprocédés – Biomolécules (BioProMo). Je tiens à remercier les directeurs respectifs, M. Laurent Falk et M. Stéphane Delaunay de m'y avoir accueillie.

Je souhaite aussi exprimer ma profonde reconnaissance à ma directrice de thèse Mme. Annie Marc et mon co-directeur de thèse M. Bruno Ebel, pour leur encadrement et leurs conseils scientifiques, mais aussi pour la confiance et l'autonomie qu'ils m'ont accordée tout au long de cette thèse. Je remercie également, M. Emmanuel Guedon pour les discussions scientifiques (ou non), les conseils et le temps qu'il a consacré pour répondre à mes questions. L'implication scientifique de M. Frantz Fournier m'a été aussi précieuse pour réaliser ce travail.

Je remercie chaleureusement Mme. Carole Molina-Jouve et M. Gérald Thouand pour avoir accepté de juger ces travaux. Je tiens également à remercier Mme. Cécile Legallais et Mme. Anne Gregoire de faire partie du jury de cette thèse.

Ces trois années de thèse ont représenté une période très spéciale pour moi, une phase de transition entre la vie étudiante et la vie professionnelle, où j'ai eu des frustrations mais aussi beaucoup de joies. Un grand merci à mes collèges de travail : Arnaud, Xavier, Isabelle, Manu R., Latifa, Catherine, Fanny, Christelle, Irina, Joseph, Romain, Melody, Odile, Yann, Cédric, David… Je tiens à remercier tout particulièrement Fabrice, pour toujours m'avoir écouté et soutenu, et aussi pour être mon garent de l'appartement. Sans son aide, je ne pourrais jamais aller si loin. Je remercie également Eric qui m'a dépanné pas mal de fois, et aussi pour le jeu d'UNO avec lequel qu'on a passé des bons moments entre midi et deux. Mes pensées vont aussi aux doctorants de cette équipe, Léna qui m'a écouté, accompagné et réconforté (avec des bouteilles de vins), Aline avec ses fous rires qui m'a encouragé et m'a motivé d'aller au KeepCool (grâce à toi, j'avais le « ventre en béton » à l'époque), Caroline et nos chorés d'Hip-Hop, Julien et ses pâtisseries, Céline et les soirées Wiidance, Chafik, Amani, Claire, Sara, Sophie, Felipe… Merci surtout à Daniel, qui m'a beaucoup aidé et écouté surtout quand j'étais en fin de thèse. Je vous remercie pour tous les bons moments passés ensemble, vous m'avez chassé la solitude étant dans un pays étranger et loin de ma famille.

Enfin, un énorme merci à mon chéri, mon chauffeur, mon secrétaire personnel, Xiao, de m'avoir accompagné et soutenu tout au long même quand j'ai dû travailler les nuits et les weekends. Sans son amour et soutien, je ne pourrais jamais faire ceux que j'ai fait. Un grand cœur à ma lapine Saber de m'avoir accompagné et m'a laissé caresser pour me réconforter.

Sommaire général

Introduction générale 1

Etude bibliographique 5

1. Procédés de culture de cellules animales pour la production d'anticorps recombinants 7

2. Techniques spectroscopiques pour le suivi en ligne des procédés de culture de cellules animales 31

Objectifs des travaux 70

Matériel & Méthodes 74

1. Cellules 76

2. Analyses hors-ligne des cellules, substrats et produits 80

3. Analyses spectroscopiques en ligne 87

4. Techniques chimiométriques pour le traitement des données spectrales 90

Résultats & Discussion 95

Chapitre I. Suivi en ligne de la vitesse spécifique de croissance cellulaire par spectroscopie diélectrique 97

Chapitre II. Comparaison et combinaison des spectroscopies Raman et NIR pour le suivi en ligne de procédés de culture de cellules animales 120

Chapitre III. Suivi en ligne de la glycosylation des anticorps par spectroscopie Raman (et NIR) 152

Chapitre IV. Travaux préliminaires pour le pilotage automatisé d'une culture semi-continue par l'utilisation conjointe de différentes techniques spectroscopiques 194

Conclusion générale & Perspectives 226

Références Bibliographiques 239

Annexes 251

Abréviations

AA	Amino acid (acide aminé)
ABS	Antigen-binding site (site de couplage anticorps-antigène)
AcM	Anticorps monoclonal
ADCC	Antibody-dependent cell-mediated cytotoxicity (cytotoxicité à médiation cellulaire dépendante des anticorps)
AMM	Autorisation de mise sur le marché
ANN	Artificial Neural Networks (réseaux de neurones artificiels)
BCMESP	Bias corrected mean square error of prediction
CDC	Complement-dependent cytotoxicity (cytotoxicité par activation du système du complément)
CDR	Complementary determining region
CE	Capillary electrophoresis (électrophorèse capillaire)
CHO	Chinese hamster ovary (ovaire de hamster Chinois)
CPP	Critical process parameter (paramètre critique du procédé)
CQA	Critical quality attribute (attribut qualité critique)
DCD	Dead cell density (densité des cellules mortes)
DHFR	Dihydrofolate réductase
ELISA	Enzyme-linked immunosorbent assay (dosage d'immunoabsorption par enzyme liée)
EMSC	Extended multiplicative signal correction
En-PLS	Ensemble partial least square
EPO	Erythropoïétine
ESI	Electrospray ionization
FDA	Food and drug administration (agence fédérale américaine des produits alimentaires et médicamenteux)
FIR	Far infrared (Infrarouge lointain)
FOM	Figure of merit (facteur de mérite)
GA	Genetic algorithm (algorithme génétique)

GlcNAc	N-acétyglucosamine
GS	Glutamine synthétase
HCP	Host cell protein (protéine des cellules hôtes)
HEK	Human embryonic kidney (cellules rénales embryonnaires humaines)
HPLC	High-performance liquid chromatography (chromatographie en phase liquide à haute performance)
ICH	International council for harmonisation (conseil international d'harmonisation)
IgG	Immunoglobuline de type G
IgG-fuc	IgG fucosylée
IgG-gal	IgG galactosylée
IgG-man	IgG hautement mannosylée
IgG-NG	IgG non-glycosylée
IgG-sia	IgG sialylée
IgG-T	IgG totale
iPLS	Interval partial least square
KPLS	Kernel partial least square
LCD	Lysed cell density (densité des cellules lysées)
LDH	Lactate dehydrogenase (lactate déshydrogénase)
LOD	Limit of detection (limite de détection)
LV	Latent variable (variable latente)
LWPLS	Local weighted partial least square
MA	Moving everage (moyenne mobile)
MALDI	Matrix-assisted laser desorption
MC	Mean centering (centrer par la moyenne)
MIR	Mid infrared (moyen infrarouge)
MS	Mass spectrometry (spectrométrie de masse)
MSC	Multiplicative scatter correction (correction de dispersion multiplicative)
NIR	Near infrared (proche infrarouge)

NMPC	Nonlinear model predictive control
PAT	Process analytical technology
PC	Principal component (composante principale)
PCA	Principal component analysis (analyse en composantes principales)
PLS	Partial least square (moindres carrés partiels)
PNGase	Peptide-N-glycosidase
PTM	Post translational modification (modification post-traductionnelle)
QbD	Quality by design
QTPP	Quality target product profile (profil de qualité du produit recherché)
RER	Range error ratio (ratio de la gamme des erreurs)
RMSE (C, CV, P)	Root mean square error (calibration, validation croisée, prédiction)
RPD	Ratio of performance deviation (ratio de la déviation de performance)
SG	Savitzky-Golay
SNV	Standard normal variate (variation normale standard)
SP	Smoothing spline
S-PLS	Serial partial least square
SVF	Sérum de veau fœtal
SVR	Support vector regression
TCA	Tricarboxylic acid cycle (cycle des acides tricarboxyliques)
TCD	Total cell density (densité des cellules totales)
VCD	Viable cell density (densité des cellules viables)
VCV	Viable cell volume (volume des cellules viables)
VIP	Variable Importance in Projection (importance de la variable pour la projection)
vPCV	Packed viable cell volume (volume total des cellules viables)

Nomenclature

C_m	Capacitance membranaire des cellules	$F.m^{-2}$
f_c	Fréquence caractéristique de la courbe de β-dispersion	Hz
F_f	Débit d'alimentation final utilisé	$mL.h^{-1}$
F_t	Débit d'alimentation calculé à l'instant *t*	$mL.h^{-1}$
Glc_F	Concentration du glucose dans le milieu d'alimentation	mM
Glc_{target}	Consigne de la concentration du glucose résiduel	mM
$P_{IgG/X}$	Productivité spécifique d'anticorps IgG	$pg.cellule^{-1}.j^{-1}$
q_{Glc}	Vitesse spécifique de consommation du glucose	$mmole.h^{-1}.10^8 cellules^{-1}$
q_{IgG}	Vitesse spécifique de production d'anticorps IgG	$mg.h^{-1}.10^8 cellules^{-1}$
q_{Lac}	Vitesse spécifique de production du lactate	$mmole.h^{-1}.10^8 cellules^{-1}$
R	Différence résiduelle	
V	Volume de la suspension cellulaire en bioréacteur	L
Vit_P_{new}	Vitesse de la pompe d'alimentation	$tour.min^{-1}$
X_V	Concentration des cellules viables	10^5 $cellules.mL^{-1}$
$Y_{IgG/X}$	Rendement de production d'anticorps IgG par cellules viables produites	$mg.10^8 cellules^{-1}$
$Y_{Lac/Glc}$	Rendement de production de lactate par glucose consommé	$mmole.mmole^{-1}$
$Y_{X/Glc}$	Rendement de production de cellules par glucose consommé	$10^8 cellules.mmole^{-1}$
α	Paramètre empirique représentant la pente de la courbe de β-dispersion à f_C	
δF	Fonction de correction du débit d'alimentation	$mL.h^{-1}$
ε	Permittivité de la suspension cellulaire	$F.m^{-1}$
σ_i	Conductivité intracellulaire	$S.m^{-1}$
σ_m	Conductivité du milieu de culture	$S.m^{-1}$
μ	Vitesse spécifique de croissance cellulaire	h^{-1}

Introduction Générale

Introduction générale

La demande en produits biopharmaceutiques s'est fortement accrue au cours de la dernière décennie, en particulier en ce qui concerne les anticorps monoclonaux (AcM) thérapeutiques. Ces AcM sont essentiellement produits par des procédés mettant en œuvre la culture de cellules animales, telles que la lignée continue CHO (Chinese Hamster Ovary). Par conséquent, les industries biopharmaceutiques sont confrontées au double défi de répondre à cette demande croissante en minimisant les coûts de production tout en assurant la qualité des produits. Cependant, chaque produit biopharmaceutique est approuvé par les instances réglementaires sur la base d'un procédé bien précis. Aussi, toute modification significative du protocole de production nécessite une nouvelle validation de l'efficacité et de la sécurité du produit final, ce qui peut augmenter considérablement les coûts de production (Teixeira *et al.*, 2009a).

Pour répondre à cette problématique, deux principes fondamentaux ont été proposés pour le développement des nouveaux procédés de production de produits biologiques à usage thérapeutique (Jenzsch *et al.*, 2017) :

- la qualité du produit ne devrait plus être évaluée uniquement en fin de procédé (Quality by Testing), mais être suivie tout au long du procédé et conceptualisée dans toutes les étapes de fabrication (Quality by Design, QbD) ;
- la variabilité du procédé, et par conséquent du produit, devrait être mieux comprise et contrôlée en temps réel, ceci nécessitant un renforcement des connaissances avec une approche davantage basée sur des aspects scientifiques.

Ces deux principes ont conduit le ICH (Conseil international d'harmonisation des exigences techniques pour l'enregistrement des médicaments à usage humain) à rédiger une série de notes explicatives (ICH Q8-Q11), introduisant pour la première fois les concepts d'espace de conception (Design Space) et de QbD. D'autre part, l'initiative PAT (Process Analytical Technology), proposée en 2004 par les autorités règlementaires telles que la FDA (Food and Drug Administration), constitue un élément important dans la mise en place du QbD. En effet, elle incite les entreprises à adopter de nouveaux outils de suivi en temps réel à tous les stades du procédé de production, en se basant sur l'analyse en ligne de paramètres clés, en vue d'assurer la qualité du produit récolté en fin de procédé. A l'heure actuelle, seuls quelques paramètres physico-chimiques, tels que la température, le pH ou encore la teneur en oxygène dissous, sont habituellement mesurés en ligne dans le bioréacteur de production. Par contre, le suivi des concentrations des éléments nutritifs, des métabolites, ou des cellules reste encore principalement effectué par des méthodes hors ligne, ce

qui n'est pas en adéquation avec une potentielle stratégie de contrôle en temps réel du procédé (Abu-Absi *et al.*, 2014).

Au cours de ces dernières années, un engouement est apparu pour l'utilisation de nouveaux outils de suivi en ligne des procédés de culture de cellules animales. Dans ce cadre, les spectroscopies (diélectrique, proche infra-rouge (NIR), Raman, etc.), se sont progressivement imposées au sein des laboratoires de R&D industriels, même si elles sont encore loin d'être utilisées en routine sur les procédés de production. Ces méthodes spectroscopiques, couplées à des sondes stérilisables *in situ*, offrent l'avantage d'être non invasives, non destructirices (*e.g.* ne perturbent pas les cellules ni ne dégradent le milieu de culture), et de pouvoir fournir simultanément des informations en ligne sur plusieurs variables du procédé (Teixeira *et al.*, 2009a). Cependant, en raison de la complexité des procédés de culture de cellules animales en bioréacteur, notamment de la composition des surnageants de culture, et de la nécessité du traitement des données spectroscopiques multivariées, basé sur diverses méthodes chimiométriques, leur utilisation, dédiée pour le moment au seul suivi de la concentration des différents composés du milieu de culture, n'est pas encore bien développée (Claßen *et al.*, 2017). Par ailleurs, à l'heure actuelle, l'analyse en temps réel de paramètres innovants permettant de mieux décrire le procédé de production, tels que les paramètres cinétiques liés à l'état physiologique des cellules ou encore les paramètres décrivant la qualité du produit, reste très peu voire non abordée.

Il existe donc encore de nombreux défis à relever pour implémenter une stratégie PAT au sein des bioprocédés de culture de cellules animales, en vue de développer des stratégies de contrôle et d'optimisation des performances, tant en terme de productivité que de qualité des produits.

Dans ce contexte, l'objectif général de cette thèse s'inscrit dans le développement de nouvelles approches méthodologiques et expérimentales, basées sur les techniques spectroscopiques *in situ* (diélectrique, NIR et Raman), pour le suivi et le contrôle en temps réel de procédés de culture de cellules CHO productrices d'anticorps monoclonaux.

La première partie du mémoire (Etude bibliographique) établit tout d'abord l'état des connaissances sur les procédés de culture de cellules CHO en réacteur, avec une attention toute particulière sur la production d'anticorps recombinants, intégrant la description de leur structure, dont leur glycosylation, et présentant les outils d'analyse de ces produits. Les enjeux des approches PAT et QbD ainsi que l'intérêt des différentes méthodes spectroscopiques pour le suivi en ligne des bioprocédés sont ensuite présentés. Enfin, les méthodes chimiométriques pour le traitement des données spectrales multivariées ainsi que la notion de capteurs logiciels sont abordées. Une

présentation plus détaillée des enjeux scientifiques ainsi que des objectifs des travaux et de la démarche adoptée sera proposée juste après l'état de l'art bibliographique. La seconde partie du mémoire (Matériels & Méthodes) rassemble les matériels et les méthodes expérimentales et numériques utilisés pour exécuter les travaux de recherches. La troisième partie (Résultats & Discussion) présente l'ensemble des résultats obtenus au cours de la thèse. Le premier chapitre de résultats (Ch. I) concerne la mise en place du suivi en temps réel de la vitesse spécifique de croissance cellulaire par spectroscopie diélectrique. Le deuxième chapitre (Ch. II) propose une méthodologie permettant de comparer efficacement et en parallèle deux technologies spectroscopiques, le NIR et le Raman, sur la base d'un certain nombre de critères de performance. Le troisième chapitre (Ch. III) étudie la possibilité du suivi en ligne de la qualité des anticorps produits, et plus particulièrement de leur état de glycosylation. Enfin, le dernier chapitre de résultats (Ch. IV) concerne des essais préliminaires de cultures en mode semi-continu avec pilotage automatisé de l'alimentation, mettant en œuvre l'utilisation conjointe de différentes techniques spectroscopiques en ligne. La dernière partie du manuscrit (Conclusion générale & perspectives) propose une synthèse des acquis ainsi que des perspectives à ces travaux.

Etude bibliographique

1. Procédés de culture de cellules animales pour la production d'anticorps recombinants……7

1.1. Contexte général *7*

1.2. Procédés de culture de cellules CHO en bioréacteur *8*

1.2.1. Les cellules CHO 8

1.2.2. Composition des milieux de culture 10

1.2.3. Métabolisme des principales sources énergétiques 11

1.2.4. Impact des paramètres opératoires physico-chimiques 14

1.2.5. Différents modes de culture en bioréacteur 16

1.3. Anticorps thérapeutiques recombinants produits par les cellules animales *20*

1.3.1. Structure et mécanisme d'action des anticorps 20

1.3.2. Glycosylation des anticorps produits par des cellules animales 22

1.3.3. Facteurs susceptibles d'influencer la glycosylation des anticorps au cours du procédé 25

1.3.4. Méthodes d'analyse des anticorps et de leur glycosylation 28

2. Techniques spectroscopiques pour le suivi en ligne des procédés de culture de cellules animales 31

2.1. Enjeux des approches PAT et QbD *31*

2.2. Intérêt de la spectroscopie diélectrique *34*

2.2.1. Principe de fonctionnement 35

2.2.2. La spectroscopie diélectrique dans les procédés de culture de cellules animales 38

2.3. Intérêt de la spectroscopie infrarouge (IR) *41*

2.3.1. Principe de fonctionnement 42

2.3.2. Application de la spectroscopie IR au cas des procédés de culture de cellules animales 43

2.4. Intérêt de la spectroscopie Raman *51*

2.4.1. Principe de fonctionnement 51

2.4.2. Application de la spectroscopie Raman au cas des procédés de culture de cellules animales 53

2.5. Méthodes chimiométriques pour le traitement des données spectroscopiques *56*

2.5.1. Prétraitement des données spectrales 57

2.5.2. Méthodes chimiométriques pour l'analyse qualitative des données spectrales 59

2.5.3. Méthodes chimiométriques pour l'analyse quantitative des données spectrales 61

2.5.4. Evaluation des performances des modèles d'analyse quantitative des données spectrales 63

2.6. Intérêt des capteurs logiciels *65*

1. Procédés de culture de cellules animales pour la production d'anticorps recombinants

1.1. Contexte général

Le développement actuel rapide des connaissances et des techniques moléculaires a considérablement fait évoluer les industries pharmaceutiques vers les biotechnologies. Le secteur biopharmaceutique représente ainsi une proportion importante et croissante du marché pharmaceutique global. Jusqu'en 2014, 212 produits biopharmaceutiques ont obtenu une autorisation de mise sur le marché (AMM) par l'agence fédérale américaine des produits alimentaires et médicamenteux (FDA) (Walsh, 2014). Aujourd'hui, plus de 7000 produits biopharmaceutiques sont en développement clinique, et plus de 150 milliards USD sont dépensés chaque année dans le secteur R&D pour ces produits. La grande majorité de ceux actuellement sur le marché sont des protéines recombinantes, c'est à dire des protéines dont le gène a été introduit dans des cellules non productrices par recombinaison génétique. Plus précisément, parmi les 62 protéines thérapeutiques approuvées par la FDA entre les années 2011 et 2016, la moitié sont des anticorps monoclonaux (AcM) recombinants (Figure 1) (Lagassé *et al.*, 2017).

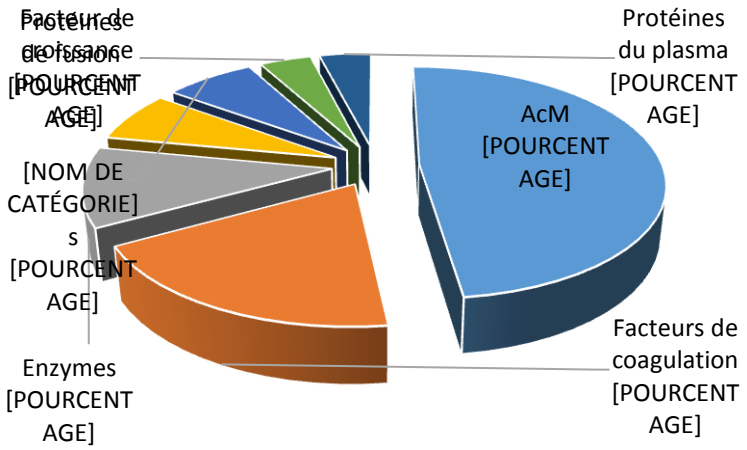


Figure 1 : Protéines recombinantes approuvées par la FDA entre 2011 et 2016 (d'après Lagassé et al., 2017)

Ces AcM sont utilisés pour traiter diverses maladies, en particulier des formes multiples de cancer, la sclérose en plaques ou encore la polyarthrite rhumatoïde. Grâce aux progrès de la biologie moléculaire, des améliorations significatives ont été obtenues pour la conception de ces AcM thérapeutiques, permettant d'améliorer leur biodisponibilité et leur spécificité, ou de fortement réduire les effets secondaires immunologiques (Reichert et Valge-Archer, 2007). Ainsi, le marché des AcM représente un marché en forte croissance qui devrait tendre, d'après certaines prévisions, vers un chiffre d'affaire d'environ 125 milliards USD en 2020 (Ecker *et al.*, 2015).

De nos jours, les bioprocédés industriels mettant en œuvre la culture de cellules animales sont devenus incontournables pour l'obtention de ces protéines thérapeutiques, et plus particulièrement des AcM. Ces procédés produisent en effet plus de la moitié des biomolécules thérapeutiques ayant reçu une AMM entre 2010 et 2014 (Walsh, 2014). Parmi les cellules animales les plus utilisées industriellement, les cellules d'ovaires de hamster chinois (Chinese Hamster Ovary - CHO) représentent le système de production le plus répandu (Figure 2).

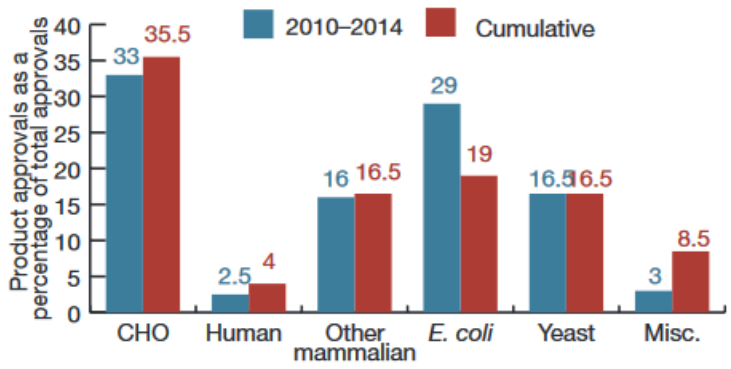


Figure 2 : Pourcentage des différents systèmes d'expression utilisés au cours des années 2010-2014 (bleu), et des années 1982-2014 (rouge), pour la production de produits biopharmaceutiques (Walsh, 2014)

1.2. Procédés de culture de cellules CHO en bioréacteur

1.2.1. Les cellules CHO

Il existe actuellement différents systèmes cellulaires d'expression permettant la production de protéines recombinantes : les cellules procaryotes comme la bactérie *Escherichia coli* ou les cellules eucaryotes telles que les levures (*Saccharomyces cerevisiae*), les cellules d'insectes (*SF-9*) ou encore les cellules de mammifères. Les organismes pluricellulaires comme les plantes et les animaux transgéniques peuvent également être utilisés. Néanmoins, tous ces systèmes ne permettent pas

d'obtenir les mêmes qualités de protéines recombinantes en terme de modifications post-traductionnelles (PTM) comme la glycosylation ou le repliement natif des chaînes peptidiques. C'est pourquoi, 60 à 70% des protéines recombinantes, dont 95% des anticorps thérapeutiques actuellement sur le marché, sont produits par des cellules de mammifères (Frenzel *et al.*, 2013). Ces cellules permettent, en effet, de réaliser de nombreuses PTM indispensables à la maturation et à la fonctionnalité des protéines produites. Ainsi, les protéines exprimées par ces cellules sont correctement assemblées, repliées et maturées, avec en particulier une glycosylation très proche de celle retrouvée dans les protéines d'origine humaine (Fallot, 2009).

Les lignées cellulaires mammaliennes les plus utilisées dans les procédés industriels sont d'origine humaine (HEK293 - Human Embryonic Kidney ; PER.C6 - cellules rétiniennes) ou murine (NS0, Sp2/0 et YB2/0 - myélomes murins ; BHK - Baby Hamster Kidney ; CHO) (Li *et al.*, 2010b). Parmi toutes ces lignées cellulaires, les cellules CHO sont les plus utilisées pour la production de protéines recombinantes thérapeutiques (Walsh, 2014). En effet, elles présentent de nombreux avantages pour les industries biopharmaceutiques (Rita Costa *et al.*, 2010) :

- elles peuvent atteindre des densités cellulaires élevées en bioréacteur, grâce à l'ingénierie cellulaire (par exemple pour réduire la mort cellulaire programmée ou apoptose), combinée à l'optimisation des technologies de réacteurs et des conditions opératoires du procédé de culture (afin, par exemple, de limiter la carence en nutriments au cours de la culture) ;
- la présence d'amplificateurs forts leur permet de produire une grande quantité de protéines d'intérêt, le maintien et l'amplification de l'expression recombinante étant assurée par un système de sélection basé par exemple sur la dihydrofolate réductase (DHFR) ou la glutamine synthétase (GS) ;
- elles peuvent être adaptées à la culture en suspension dans des milieux dépourvus de composés d'origine animale ;
- elles résistent assez bien à l'extrapolation des procédés aux grandes échelles et aux variations de leur environnement ;
- le profil des PTM des protéines obtenues par ces cellules est quasi identique à celui des protéines humaines, ce qui entraîne donc peu d'effets secondaires sur les patients.

En 1987, la première protéine recombinante produite par cellules animales ayant obtenu une AMM, fut l'activateur tissulaire du plasminogène (tPA), généré par culture de cellules CHO en bioréacteur à grande échelle chez Genentech. Aujourd'hui, face à la croissance exponentielle de la demande en ces protéines thérapeutiques, de nombreuses entreprises ont construit des installations de production à

grande échelle utilisant des bioréacteurs de 15 à 20 m^3. Par ailleurs, en moins de 20 ans, l'évolution conjointe des aspects biologiques (lignées cellulaires, milieux de cultures) et des conditions opératoires, a permis d'optimiser les procédés et d'augmenter la concentration des AcM produits par cellules CHO de quelques centaines de milligrammes par litre à près de 10 $g.L^{-1}$ (Li *et al.*, 2010b).

1.2.2. Composition des milieux de culture

Les cellules animales nécessitent un milieu de culture relativement complexe contenant un grand nombre d'éléments nutritifs pour assurer le fonctionnement de leur métabolisme. Le milieu de culture permet également de protéger les cellules de certaines contraintes hydrodynamiques et variations de pH. Ainsi, la composition du milieu de culture doit être maîtrisée pour assurer, non seulement, une bonne croissance cellulaire, mais aussi la qualité des protéines produites.

Les composants principaux de ces milieux de culture sont précisés ci-dessous :

- le glucose, en général entre 5 et 25 mM, représente la source principale d'énergie pour les cellules : il est oxydé par les cellules en pyruvate *via* la glycolyse pour être utilisé dans le cycle des acides carboxyliques ou réduit en lactate ;
- la glutamine, en général entre 0,5 et 5 mM, constitue la seconde source énergétique après le glucose : elle est aussi utilisée comme source de carbone et d'énergie, mais représentée également une source d'azote pour la synthèse des nucléotides, acides aminés et acides nucléiques. Elle est indispensable aux cellules qui sont incapables de la synthétiser à partir des ions ammonium et du glutamate, car dépourvues de glutamine synthétase (Neermann et Wagner, 1996) ;
- les acides aminés sont la base de la synthèse protéique : 12 d'entre eux sont dits essentiels car non synthétisables par les cellules (arginine, cystéine, histidine, isoleucine, leucine, lysine, méthionine, phénylalanine, thréonine, tryptophane, tyrosine et valine) (Eagle, 1955). Des acides aminés non essentiels peuvent aussi être ajoutés en cas de déplétion afin de prolonger la durée de vie des cellules. L'optimisation spécifique de la composition en acides aminés peut permettre une amélioration significative de la concentration maximale cellulaire et de la production en protéines recombinantes (Fan *et al.*, 2015) ;
- les lipides sont des éléments constitutifs des membranes cellulaires et jouent aussi un rôle dans le stockage d'énergie. Ils sont surtout ajoutés dans les milieux de culture sans sérum (cholestérol, acides gras, acide linoléique, ...) (Lima e Castro, 1993; Xu *et al.*, 2005) ;
- les nucléotides pourront être ajoutés pour la culture de lignées cellulaires déficientes en *dhfr* et donc incapables de les synthétiser. Les nucléotides mono-, di- et triphosphates participent au

métabolisme énergétique et sont également des précurseurs indispensables au processus de glycosylation (Carvalhal *et al.*, 2011; Kochanowski *et al.*, 2006) ;

- les vitamines sont requises en faible quantité et interviennent principalement comme coenzymes : elles ne peuvent pas être synthétisées par les cellules en quantité suffisante et doivent donc être apportées par le milieu de culture (Arora, 2016) ;

- les ions inorganiques et les éléments traces jouent plusieurs rôles comme le maintien du pH et de l'osmolarité du milieu, le transfert de molécules à travers les membranes ou encore comme cofacteurs enzymatiques. Les ions inorganiques les plus fréquents sont : Ca^{2+}, Mg^{2+}, K^{+}, PO_4^{3-}, Na^{+}, Cl^{-}, SO_4^{2-} et HCO_3^{-}. Les ions métalliques (Fe^{2+}, Cu^{2+}, Zn^{2+}, Co^{2+}, Mn^{2+}), apportés à l'état de traces, participent au site actif de certaines enzymes et interviennent dans les réactions de la chaîne respiratoire (Mizrahi et Lazar, 1988).

1.2.3. Métabolisme des principales sources énergétiques

a) Glucose

Le glucose est la source de carbone préférentielle pour les cellules animales, puisque les transporteurs d'hexoses présentent une affinité plus élevées pour le glucose que pour le galactose ou le fructose par exemple. Le glucose est oxydé au cours de la glycolyse où il est en grande partie converti en pyruvate, couplé à la formation d'ATP et de NADH (Figure 3). L'oxydation complète de 2 molécules de pyruvate dans le cycle des acides tricarboxyliques (TCA) permet alors de générer 36 molécules d'ATP supplémentaires. Cependant, le pyruvate peut aussi être transformé en lactate ce qui permet de régénérer le NAD^{+}, nécessaire au fonctionnement de la glycolyse. Cette voie est alors considérée comme une perte d'énergie (Hauser et Wagner, 2014). Par ailleurs, au-delà de 20 mM, le lactate produit est considéré comme toxique pour les cellules et peut influencer la production et la glycosylation des protéines (Lao et Toth, 1997). Cette production de lactate dépend fortement de la quantité de glucose disponible dans le milieu. Dans un milieu à concentration élevée en glucose, son intégration rapide dans la cellule provoque une forte activité glycolytique aboutissant à une grande quantité de pyruvate et de NADH. Afin de rééquilibrer rapidement le potentiel redox NAD^{+}/NADH, la voie de conversion du pyruvate en lactate est alors privilégiée (Feron, 2009). Par contre, à des niveaux réduits en glucose, les cellules utilisent un métabolisme plus efficace en énergie en réduisant la production de lactate. Généralement, un métabolisme présentant un rendement de conversion glycolytique ($Y_{lac/glc}$) inférieur à 1 mole.mole^{-1} est considéré comme énergétiquement efficace (Hauser et Wagner, 2014). A noter également que, pour limiter la production de lactate, le glucose peut

parfois être remplacé par d'autres glucides au métabolisme plus lent (fructose, maltose, galactose) (Altamirano *et al.*, 2006; Duval *et al.*, 1992; Imamura *et al.*, 1982).

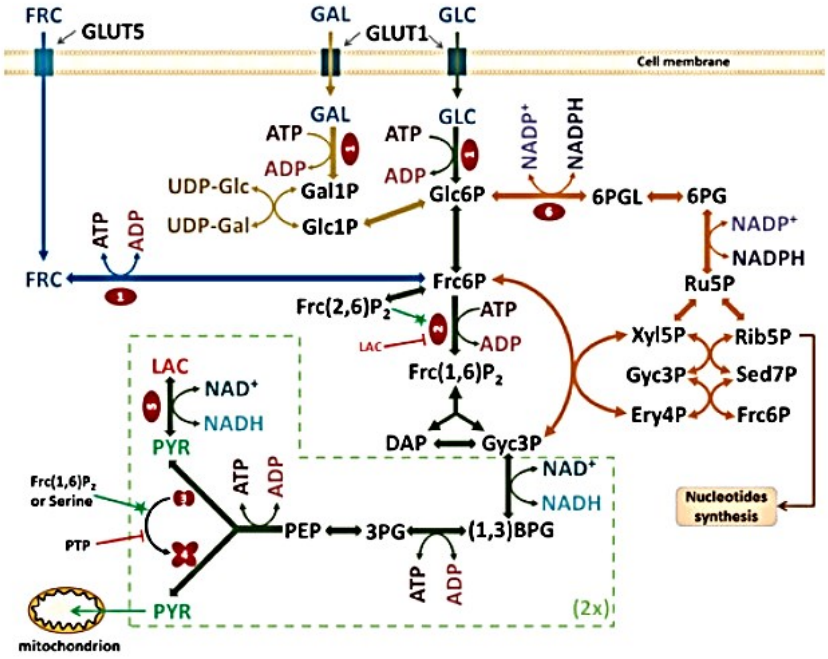


Figure 3 : Voie métabolique des principales sources de carbone dans les cellules animales. Les chiffres dans les ovales rouges représentent des enzymes clés, 1 : hexokinase, 2 : phosphofructo kinase, 3 : pyruvate kinase-M2 isoforme (dimère), 4 : pyruvate kinase-M2 isoforme (tétramère), 5 : lactate déshydrogénase, et 6 : glucose- 6-phophate déshydrogénase. ((1,3)BPG = 1,3-biphosphoglycérate; 3PG =3-phosphoglycerate; 6PGL = 6-phosphogluconolactone; 6PG = 6-phosphogluconate; ADP = adénosine diphosphate; ATP = adénosine triphosphate; DAP = dihydroxyacétone phosphate; Ery4P = erythrose-4-phosphate; FRC = fructose; Frc(1,6)P2 = Fructose-(1,6)-biphosphate; Frc(2,6)P2 = fructose-(2,6)-biphosphate; Frc6P = fructose-6-phosphate; GAL = galactose; Gal1P = galactose-1-phosphate; GLC = glucose; Glc1P = glucose-1-phosphate; Glc6P = glucose-6-phosphate; GLN = glutamine; GLUT1 = glucose transporter 1; GLUT5 = glucose transporter 5; Gyc3P = glyceraldéhyde-3-phosphate; LAC = lactate; PEP = phophoenolpyruvate; PTP = phosphorylated tyrosine peptide; PYR =pyruvate; Rib5P = ribose-5-phosphate; Ru5P = ribulose-5-phosphate; Sed7P = sedoheptulose-7-phosphate; Xyl5P = xylulose-5-phosphate) (Hauser et Wagner, 2014).

b) Glutamine

La glutamine est consommée à une vitesse beaucoup plus rapide que celles des autres acides aminés et peut fournir entre 35 à 70% de l'énergie requise par la cellule. Au cours de la première étape de la glutaminolyse, la glutamine est convertie en glutamate et en ions ammonium (NH_4^+). Puis, le glutamate est désaminé en α-cétoglutarate qui rentre dans le cycle TCA (Figure 4). Cette voie métabolique est énergétiquement efficace puisque l'oxydation complète de la glutamine génère 24 ATP et 3 NADPH. Elle est privilégiée quand le glucose ne fournit pas l'énergie nécessaire (Hauser et Wagner, 2014). Néanmoins, le métabolisme de la glutamine, auquel s'ajoute sa décomposition chimique spontanée à 37 °C, conduit à la formation d'ions ammonium dont l'accumulation peut être

toxique pour les cellules animales, tant pour leur croissance que pour la glycosylation des protéines produites (Schneider *et al.*, 1996; Yang et Butler, 2000). Cette accumulation peut être limitée en substituant la glutamine par d'autres acides aminés (asparagine, glutamate), par des intermédiaires réactionnels (pyruvate), ou encore par des dipeptides tels que l'alanine-glutamine (Glutamax®) (Eagle, 1959).

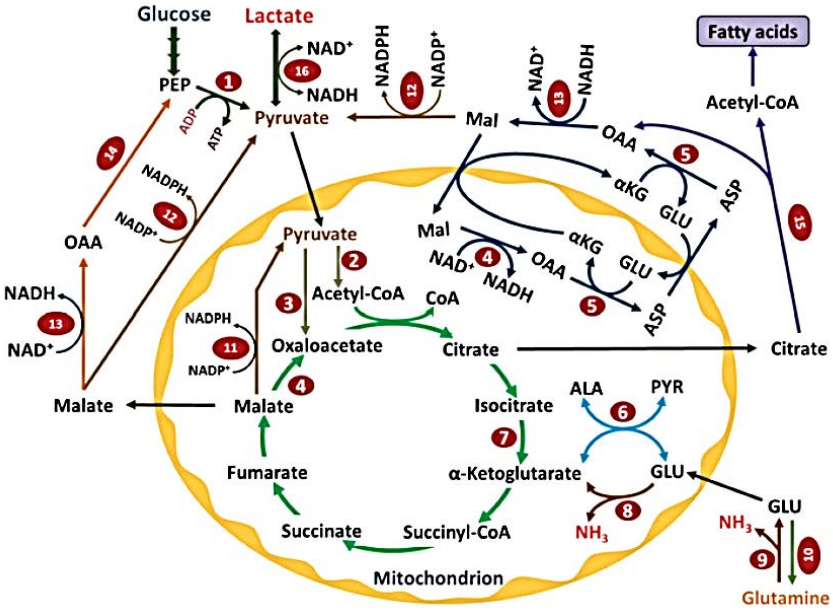


Figure 4 : Voie métabolique de la glutamine dans les cellules animales. Les chiffres dans les ovales rouges représentent des enzymes clés, 1 : pyruvate kinase (M2 isoforme), 2 : pyruvate déshydrogénase, 3 : pyruvate carboxylase, 4 : malate déshydrogénase II, 5 : aspartate aminotransaminase, 6 : alanine aminotransaminase, 7 : isocitrate déshydrogénase II, 8 : glutamate déshydrogénase, 9 : glutaminase, 10 : glutamine synthétase, 11 : enzyme malique II, 12 : enzyme malique I, 13 : malate déshydrogénase I, 14 : phosphoénolpyruvate carboxykinase, 15 : citrate lyase, 16 : lactate déshydrogénase. (αKG = α-cétoglutarate; ALA = alanine; ASP = aspartate; GLU = glutamate; Mal = malate; OAA = oxaloacétate; PEP = phosphoénolpyruvate; PYR = pyruvate) (Hauser et Wagner, 2014).

Alternativement, le glutamate peut être converti en α-cétoglutarate par amino-transamination en transformant du pyruvate en alanine et de l'oxaloacétate en aspartate. Contrairement aux réactions de désamination, cette transamination ne génère pas d'ions d'ammonium et peut constituer une réponse cellulaire pour limiter la production d'ions d'ammonium lors d'une consommation excessive de glutamine (Ozturk et Hu, 2005). La conversion de glutamine en lactate est caractéristique de beaucoup de cellules animales, jusqu'à 30% du lactate pouvant être générés à partir de la glutamine (Zielke *et al.*, 1980). Dans les conditions classiques, le rendement métabolique ammonium/glutamine

($Y_{NH4+/gln}$) est de l'ordre de 0,4 à 1 mole.mole^{-1}, mais il peut évoluer en fonction des conditions de culture (Clincke, 2010).

Pour mieux comprendre le métabolisme des cellules, les paramètres cinétiques tels que la vitesse spécifique de croissance (μ), les vitesses spécifiques de consommation/production (q), et les rendements (Y) sont également beaucoup étudiés. La vitesse spécifique de croissance est sensible aux lignées cellulaires utilisées, à l'environnement extracellulaire, et aux conditions opératoires physico-chimiques, sa valeur est de l'ordre de 0,005 à 0,04 h^{-1}. Lorsque les cellules CHO sont cultivées en mode discontinu, les vitesses spécifiques de consommation de glucose et de glutamine, ainsi que celles de production de lactate et d'ions ammonium, sont généralement élevées en début de culture, avant de décroître lors de la diminution de la concentration de glucose et de glutamine. Les valeurs moyennes des vitesses spécifiques de production/consommation sont de l'ordre de 0,03 mmole.10^8 cellules^{-1}.h^{-1} pour le glucose et le lactate (Hayter *et al.*, 1991), et de 0,005 mmole.10^8 cellules^{-1}.h^{-1} pour la glutamine et les ions ammonium (Farges *et al.*, 2008).

1.2.4. Impact des paramètres opératoires physico-chimiques

La maîtrise de certains paramètres opératoires, tels que le niveau d'oxygène dissous, la température, le pH ou l'osmolalité, au cours d'une culture de cellules animales, est essentielle pour optimiser la production des protéines recombinantes.

a) Température

La température est l'un des paramètres les plus importants pour la culture de cellules animales et doit être régulée de façon très précise. Un écart de l'ordre de 1 °C peut provoquer des changements métaboliques indésirables, un ralentissement de la croissance ou encore la diminution de production des protéines. La température optimale de croissance cellulaire est généralement de 37 °C, mais celle-ci peut être différente de la température optimale de production des protéines d'intérêt. Par exemple, dans certains procédés, une diminution brutale de la température autour de 33 °C pendant la phase de production de la molécule d'intérêt, peut fortement augmenter la productivité (Jenkins and Hovey, 1993). En bioréacteur, la température est mesurée avec précision à l'aide de dispositifs de température à résistance, bien que des thermocouples puissent également être utilisés (Li *et al.*, 2010b).

b) Oxygène dissous

L'oxygène dissous permet aux cellules de produire de l'énergie lors de la respiration. Néanmoins, du fait de la faible solubilité de l'oxygène dans l'eau (~ 7 $mg.L^{-1}$ à 37 °C), une alimentation en oxygène est nécessaire pour assurer la croissance cellulaire. La vitesse spécifique de consommation d'oxygène par les cellules animales est comprise entre 0,05 et 0,5 $mmole.10^9$ $cellules^{-1}.h^{-1}$ (Marc and Olmos, 2010). Pour les cultures à petite échelle en système non contrôlé, une aération surfacique est souvent suffisante, alors qu'à plus grande échelle, la surface de contact air-liquide devient trop faible pour assurer un échange diffusionnel suffisant. En bioréacteur, l'oxygène est alors apporté par la diffusion de bulles *via* différents systèmes d'aération (sparger, membrane, canne, etc.). En général, le point de consigne en oxygène dissous est réglé entre 20 et 50% de la saturation en air (Li *et al.*, 2010b). Un apport trop faible en oxygène peut entraîner une production excessive de lactate, tandis qu'un apport trop important est potentiellement cytotoxique (Fleischaker and Sinskey, 1981). Ceci nécessite un contrôle rigoureux du niveau d'oxygène dissous dans le milieu. Les électrodes ampérométriques recouvertes de membranes (électrodes de type Clark) sont les plus utilisées pour mesurer le niveau d'oxygène dissous dans les bioréacteurs (Biechele *et al.*, 2015). Il existe également des sondes optiques à oxygène dissous (Gao *et al.*, 2004; Hanson *et al.*, 2007).

c) pH

Les cellules animales ne tolèrent que de faibles variations de pH du milieu de culture (entre 6,8 et 7,8). La valeur optimale de pH pour la croissance cellulaire étant voisine de 7,2, un faible écart de 0,1 unité peut avoir un impact significatif sur la croissance des cellules et sur leur métabolisme, en particulier sur la consommation de glucose et la production de lactate (Schmid *et al.*, 1990). Le pH est ainsi un paramètre opératoire important à mesurer et à contrôler. Des régulations, basées sur le système tampon HCO_3^-/CO_2 à petite échelle, ou sur l'ajout de soude et de CO_2 en bioréacteur, permettent de limiter les variations du pH. Les sondes à pH instrumentant les bioréacteurs sont typiquement constituées d'une membrane en verre contenant du sodium qui forme une couche de gel hydraté sur la membrane lorsque la sonde est placée dans une phase liquide. Une électrode $Ag/AgCl_2$ dans un électrolyte de KCl saturé en $AgCl_2$ est utilisée pour mesurer la variation du potentiel de la surface externe de la membrane causée par les changements du pH du milieu (Li *et al.*, 2010b).

d) Osmolalité

La présence de sels permet de maintenir l'osmolalité du milieu de culture autour de sa valeur optimale (~ 300 $mOsm.kg^{-1}$). Cependant, l'ajout ponctuel de solutions nutritives concentrées ou de base pour contrôler le pH est susceptible d'augmenter cette valeur. En cas d'hyperosmolalité (jusqu'à 400 $mOsm.kg^{-1}$), il peut être observé un ralentissement de la vitesse spécifique de croissance

cellulaire, une augmentation de la production spécifique de la protéine recombinante, ainsi qu'une variation de la qualité des protéines produites, notamment de leur glycosylation (Konno *et al.*, 2012). Néanmoins, ce paramètre reste peu contrôlé au cours des procédés de culture (Øyaas, 2003).

1.2.5. Différents modes de culture en bioréacteur

a) Mode discontinu (batch)

En culture discontinue, le milieu initial et les cellules sont ajoutés dans le bioréacteur en début de culture. Aucun ajout ou soutirage n'est réalisé pendant toute la durée de la culture, le contenu du bioréacteur n'étant récolté qu'en une seule fois à la fin du procédé. On observe parfois une courte phase de latence juste après l'ensemencement, qui pourrait être attribuée au conditionnement du milieu ou à l'état physiologique de la préculture (Goergen *et al.*, 1997). Puis, les cellules entrent en phase exponentielle de croissance durant laquelle elles se multiplient en consommant les nutriments et en produisant les déchets métaboliques. Lorsqu'intervient l'épuisement d'un ou plusieurs nutriments et/ou l'accumulation trop importante des déchets, les cellules atteignent une phase stationnaire plus ou moins longue selon le type de cellules, avant de commencer à mourir pendant la phase de déclin (Figure 5.a). La culture en mode batch est simple et fiable, ce qui explique son utilisation pour de nombreuses applications industrielles (Noé *et al.*, 1992). Ses principaux avantages sont liés à des coûts fixes plus bas, ainsi qu'à la flexibilité et la simplicité d'utilisation. Néanmoins, les performances d'un tel procédé sont vite limitées, tant en croissance cellulaire qu'en productivité, du fait de la déplétion des nutriments et/ou l'accumulation des déchets présents dans le milieu de culture.

b) Mode recharge-récolte (feed-harvest)

En mode recharge-récolte, il est possible de réaliser plusieurs cultures discontinues successives en vidangeant, en fin de culture batch, un pourcentage important de milieu usagé (en général entre 70 à 90%) et en le remplaçant par du milieu neuf. Les cellules restant dans le réacteur reprennent alors leur croissance, ce qui permet d'initier une nouvelle culture (Figure 5.b). Ce mode de culture requiert une grande viabilité de la suspension cellulaire restante, et le choix adéquat du moment de changement du milieu (Ozturk et Hu, 2005). Il présente l'avantage de réduire fortement les phases d'arrêt du procédé liées au nettoyage et à la stérilisation du bioréacteur.

c) Mode semi-continu (fed-batch)

En mode semi-continu, le milieu de culture est ajouté en continu, par paliers ou ponctuellement, dans le bioréacteur, au moment où les cellules sont encore en phase de croissance active mais que les nutriments commencent à devenir limitants. Le volume de milieu dans le réacteur augmente au cours de la culture puisqu'aucun soutirage n'est réalisé avant la récolte finale. L'avantage principal du fed-batch par rapport au batch est de prolonger la phase de croissance cellulaire ainsi que la phase stationnaire, ce qui permet d'augmenter la concentration maximale des cellules et d'atteindre de plus hauts niveaux de production (Figure 5.c) (Jain et Kumar, 2008). Les bioprocédés de cultures semi-continues, avec une alimentation ponctuelle, sont actuellement parmi les plus couramment utilisés dans les industries pharmaceutiques jusqu'en réacteurs de 20 m^3 (Birch et Racher, 2006).

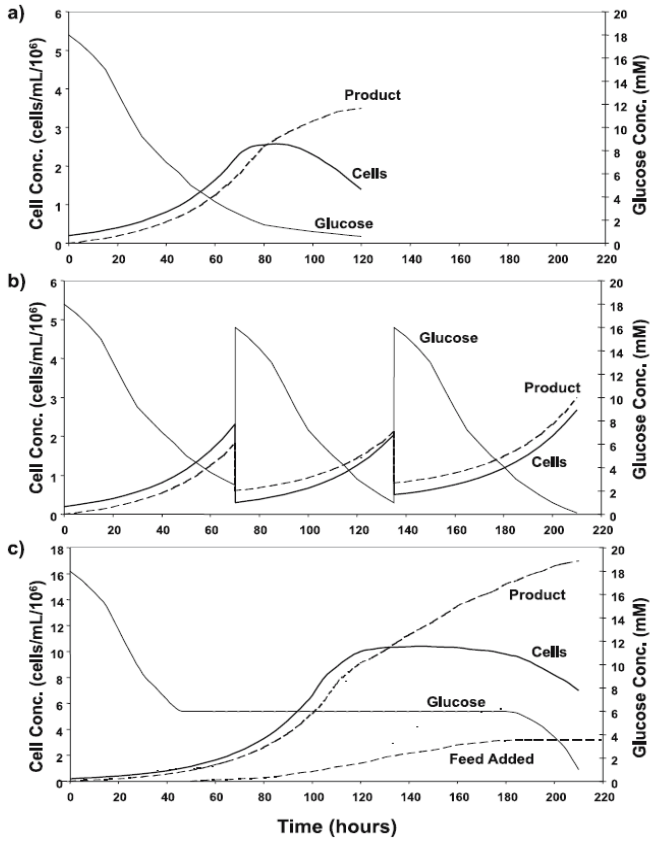


Figure 5 : Exemple de cinétiques des cultures en mode discontinu (a), recharge-récolte (b), et semi-continu (c) (Wlaschin et Hu, 2006)

La composition du milieu d'alimentation et le mode d'alimentation doivent être précisément maîtrisés pour contrôler le niveau des éléments nutritifs aux points de consigne choisis, ou pour suivre un profil optimal prédéterminé. Le mode fed-batch permet en effet, en régulant la concentration des composés clés, de contrôler certaines réactions métaboliques (Lim et Shin, 2013). Par ailleurs, comme pour les cultures en discontinu, l'accumulation de lactate, d'ions ammonium,

d'autres métabolites ou de molécules oxydantes dérivant de l'oxygène ainsi qu'une osmolalité trop élevée aboutissent à l'inhibition de la croissance cellulaire et à l'accélération de la mort cellulaire. Pour réduire cette accumulation, il est possible, par exemple, d'utiliser des stratégies d'alimentation qui maintiennent le glucose ou la glutamine à des concentrations très basses, de manière à provoquer un "shift" métabolique, une utilisation plus efficace de glucose/glutamine en produisant moins de lactate/ions ammmonium par exemple.

d) Mode continu simple et perfusé

En mode continu, du milieu de culture frais est ajouté en continu dans le bioréacteur alors que le contenu du réacteur, incluant les cellules, est soutiré au même débit, le volume du bioréacteur restant constant. Ce mode de culture permet de maintenir un environnement physico-chimique stable pendant une longue période de temps pouvant aller jusqu'à plusieurs mois. En raison de l'effet de dilution, la concentration cellulaire reste faible tout au long de ce type de culture, ce qui conduit à une faible productivité. De plus, il existe un risque d'expression instable par des cellules cultivées sur une longue période (Morrison *et al.*, 1997). Par conséquent, ce mode de culture est très peu utilisé dans l'industrie.

Un autre type de culture en mode continu est le mode perfusé, fonctionnant comme le mode continu simple mais pour lequel les cellules sont retenues ou recyclées dans le bioréacteur par un dispositif de rétention adéquat. De ce fait, la concentration cellulaire augmente progressivement, en fonction du débit choisi, jusqu'à ce qu'un état quasi stable soit atteint résultant d'une phase stationnaire (Mercille *et al.*, 2000). Ce mode de culture permet d'éviter l'accumulation de métabolites toxiques qui ont un impact négatif sur les cellules, et ainsi d'obtenir de hautes densités cellulaires et d'atteindre de fortes productivités en protéines produites. Cependant, certains inconvénients existent, comme le maintien du fonctionnement opératoire de la culture pendant des périodes prolongées, la prévention de contaminations , ou encore le maintien de la qualité du produit (Flickinger, 2013).

Pour synthétiser, les différents modes d'alimentation des procédés de culture de cellules animales sont comparés dans le Tableau 1 :

Tableau 1 : Comparaison des différents modes de cultures de cellules animales (Flickinger, 2013)

Culture Mode	Product Yield	Process Manipulation	Cost (Capital Investment and Labor)	Throughput
Batch	Low	Low	Low	Low
Repeated batch	Low	Medium	Medium	Medium
Fed-batch	High	Medium	Low	Low
Chemostat	Low	Medium	Low	Low
Perfusion	Medium	High	Medium	High

En résumé, la culture à grande échelle des cellules animales, et plus particulièrement des cellules CHO, est devenue un procédé incontournable à maîtriser et à optimiser pour la production industrielle de protéines recombinantes à usage thérapeutique de bonne qualité. Néanmoins, l'apport en substrats énergétiques et les paramètres physico-chimiques doivent être bien maîtrisés afin d'optimiser les procédés de culture.

Le mode opératoire des cultures est un choix très important et doit être bien réfléchi pour assurer la performance du procédé. Le mode discontinu est le mode de culture le plus simple à réaliser, mais sa productivité est relativement faible. Il est donc souvent le premier choix pour les recherches académiques. Pour les productions industrielles, le mode semi-continu et le mode perfusé sont les plus couramment utilisés. Une stratégie d'alimentation classiquement utilisée consiste à ajouter un volume de milieu prédéfini à des intervalles de temps assez larges (par exemple une ou deux fois par jour) en se basant sur des mesures hors ligne ou sur des données historiques (Wlaschin et Hu, 2006). Du fait de sa simplicité, cette stratégie est la plus utilisée dans les industries pharmaceutiques. Néanmoins, elle ne permet pas de bien répondre aux besoins cellulaires, ce qui limite les performances du procédé et peut influencer la qualité des protéines produites. Idéalement, la boucle de rétrocontrôle devrait être basée sur la mesure en ligne des vitesses de réactions cellulaires. Cependant, ce type de contrôle précis nécessite une très bonne connaissance des différentes voies métaboliques et de la physiologie des cellules, ainsi que des outils fiables de suivi en ligne. Par conséquent, il reste nécessaire d'optimiser la stratégie d'alimentation du procédé semi-continu par des mesures en ligne plus fréquentes des paramètres ainsi que par des schémas d'alimentation continus et bien contrôlés.

1.3. Anticorps thérapeutiques recombinants produits par les cellules animales

1.3.1. Structure et mécanisme d'action des anticorps

Les anticorps sont des glycoprotéines de la famille des immunoglobulines (Ig). Lorsqu'un antigène déclenche une réaction immunitaire dans un organisme, des anticorps sont produits par les cellules B. Ces anticorps reconnaissent spécifiquement l'antigène grâce à une région appelée épitope située à sa surface. Une fois que la molécule d'anticorps est liée à l'antigène, ce dernier peut être détruit par les cellules T. La structure générale de ces anticorps consiste en 2 copies symétriquement identiques d'une chaîne peptidique lourde (~ 55 kDa) et d'une chaîne peptidique légère (~ 25 kDa) reliées entre elles par des ponts disulfures et des liaisons non-covalentes (Figure 6 : Représentation générale de la structure d'un anticorps) (McDonnell, 2015).

Cette structure a été décrite pour la première fois en 1959 par Porter (1959). Cet auteur a montré qu'une hydrolyse par la papaïne permettait de libérer 3 fragments actifs : 2 fragments variables (Fab) intégrant la région de liaison avec l'antigène et 1 fragment constant (Fc) activateur des cellules du système immunitaire. Chaque chaîne lourde comporte trois domaines constants (C_{H1}, C_{H2}, C_{H3}) et un domaine variable (V_H), tandis que les chaînes légères ne possèdent qu'un domaine constant (C_L) et un domaine variable (V_L). Les domaines variables forment le site de couplage anticorps-antigène (ABS : Antigen-binding site), intégrant une région hypervariable de 5 à 10 acides aminés appelée CDR (Complementary Determining Region). Chez les mammifères, il existe 5 classes d'immunoglobulines (IgA, IgD, IgE, IgG, IgM) réparties selon la structure des domaines constants des chaînes lourdes (α, δ, ε, γ, μ). Les IgG et IgA peuvent être subdivisées en sous-classes (IgG1 à 4 ; IgA1 à 2), reflétant des différences plus fines entre chaînes lourdes (Burton et Woof, 1992). Il existe aussi des isotypes de chaînes légères du type κ ou λ. L'IgG1 est la classe d'anticorps la plus abondante sur le marché des anticorps à usage thérapeutique (Reichert, 2011).

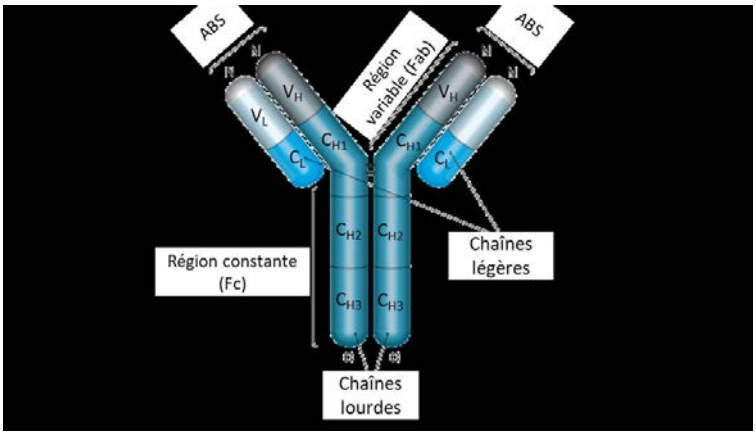


Figure 6 : Représentation générale de la structure d'un anticorps

Initialement, la production d'anticorps à usage thérapeutique a débuté par leur purification à partir de sérum de souris ou d'autres mammifères ayant subi des immunisations répétées avec un antigène ciblé. Cependant, ces anticorps étaient hétérogènes et dirigés vers plusieurs épitopes de l'antigène (anticorps polyclonaux). Ceci diminuait donc leur spécificité tout en augmentant la possibilité de liaison à l'antigène dans des conditions différentes. Actuellement, quelques anticorps polyclonaux (non recombinants) sont en phase clinique de développement pour le traitement de réactions allergiques, de maladies infectieuses ou de rejet de greffe.

Contrairement aux anticorps polyclonaux, les anticorps monoclonaux (AcM) permettent de cibler plus spécifiquement un épitope donné de l'antigène. Leur production stable a été rendue possible en 1975 grâce à la technique de cellules hybrides de rongeurs (Hybridomes) développée par Köhler et Milstein. Néanmoins, l'utilisation thérapeutique de ces AcM d'origine murine est limitée en raison de leur faible fixation sur les récepteurs humains et de leur immunogénicité. Grâce aux progrès de la biologie moléculaire et du génie génétique, les gènes codant pour les AcM humains ont pu être transfectés dans des lignées cellulaires continues, notamment les lignées CHO pour la production des AcM recombinants. Depuis le début des années 1990, sont alors progressivement arrivés sur le marché des AcM chimériques humain/souris (domaines variables murins et domaines constants humains), puis des AcM humanisés (CDR murins), et enfin des AcM humains, présentant une efficacité améliorée et une faible immunogénicité (Figure 7) (McDonnell, 2015).

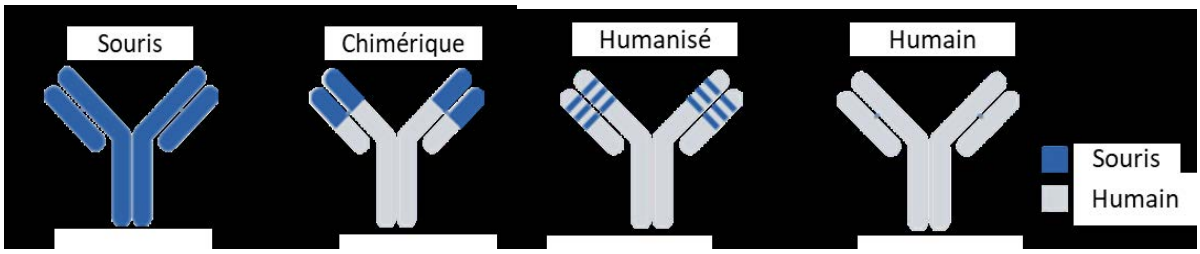


Figure 7 : Représentation des différents types d'AcM

Cas particulier de l'anticorps anti-RhD produit par la lignée CHO utilisée pour nos travaux :

L'anticorps anti-RhD est l'anticorps de type IgG, il est utilisé pour la prévention de la maladie hémolytique des nouveau-nés, représentant l'une des applications cliniques ayant le plus de succès dans l'immunosuppression médiée par les anticorps (Kumpel, 2008). Autrefois, ces anticorps étaient fournis par des mères précédemment allo-immunisées, ou par des donneurs RhD^- immunisés avec des globules rouges RhD^+. Cependant, cette méthode a fortement diminué en raison des problèmes éthiques et pratiques, et du risque de transmission de maladies infectieuses. C'est pourquoi, une alternative consiste à produire cet AcM sous forme recombinante par culture de cellules animales, même si l'efficacité d'un grand nombre d'AcM anti-RhD recombinants reste plus faible que celle des anticorps anti-RhD polyclonaux. Par ailleurs, il est apparu que certains AcM produits par cellules de mammifères pouvaient provoquer des réponses immunitaires indésirables. Ces effets semblent liés aux oligosaccharides fixés sur les peptides (glycosylation), qui peuvent varier en fonction de la lignée cellulaire productrice utilisée (Kumpel, 2008).

1.3.2. Glycosylation des anticorps produits par des cellules animales

La structure primaire de la chaîne polypeptidique de l'anticorps ne dépend que de la séquence nucléotidique de l'ADN de la cellule hôte et est donc parfaitement prévisible. Cependant, la plupart des protéines subissent des modifications co- ou post- traductionnelles, qui dépendent de la nature protéique et de l'équipement enzymatique de la cellule hôte, et sont donc plus difficilement prévisibles. La glycosylation représente l'une des PTM essentielles pour la fonctionnalité des AcM recombinants. Elle désigne les différentes réactions enzymatiques d'addition sur la chaîne peptidique par liaisons covalentes, de structures glucidiques appelées glycannes (Huhn *et al.*, 2009). Sur chaque chaîne lourde d'un anticorps de type IgG, le domaine constant C_{H2} porte le site de N-glycosylation au niveau du résidu Asparagine 97 (Figure 8).

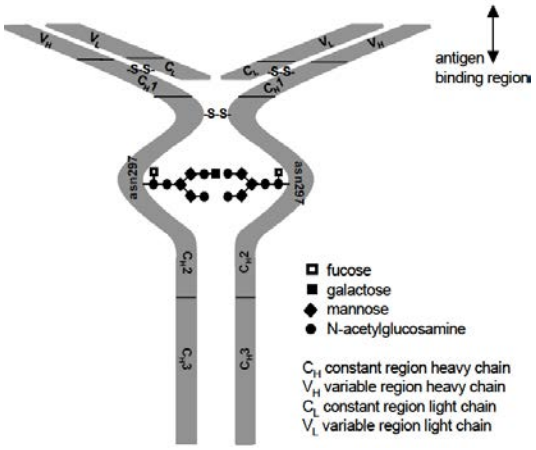


Figure 8 : Représentation de la glycosylation d'un anticorps (Nahrgang, 2002)

Les oligosaccharides fixés sur ce site représentent 2 à 3% du poids moléculaire de l'anticorps. Lorsque les anticorps sont produits dans des cellules de mammifères, ces structures glycanniques sont généralement de type complexe bi-antenné. Un noyau glycannique, constitué d'un penta-saccharide bi-antenné (Man3GlcNAc2), est caractéristique des N-glycannes (Figure 9).

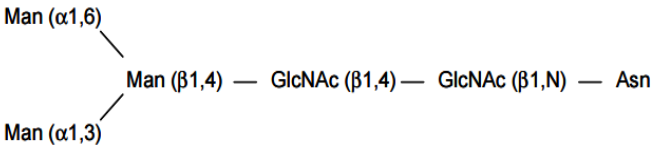
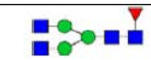
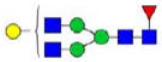
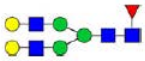
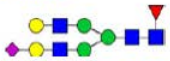
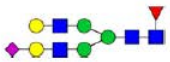
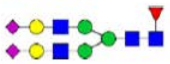
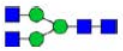
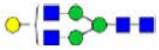
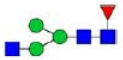
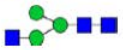
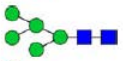
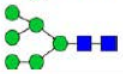


Figure 9 : Structure du noyau penta-saccharidique de la N-glycosylation (Nahrgang, 2002)

Selon les cas, différents glucides peuvent ensuite se fixer sur ce noyau : fucose, N-acétyglucosamine bissectée (GlcNAc), acide N-acétyl neuraminique (sialylation), mannose, ou encore galactose (Nahrgang, 2002). En effet, le profil de glycosylation des protéines recombinantes peut varier en fonction du système d'expression cellulaire utilisé. Le terme de glycoformes, introduit par Rademacher *et al.* en 1988, est alors utilisé pour décrire les différentes formes d'une même glycoprotéine qui diffèrent les unes des autres par leurs structures glycanniques. Les principales glycoformes des AcM produits par culture de cellules CHO et leurs pourcentages sont présentées Figure 10. Cette variation du profil de glycosylation d'une protéine donnée correspond à la glyco-hétérogénéité qui peut être classée en deux classes : macro et micro-hétérogénéité. La macro-hétérogénéité a trait à l'occupation ou pas des sites de glycosylation par un groupement glycannique. La micro-hétérogénéité s'intéresse aux séquences différentes des glycannes présents sur un site de glycosylation donné. Le plus souvent, les IgG produites par les cellules de CHO présentent un niveau

élevé de glycoformes non-galactosylée (G0) par rapport aux IgG produites par d'autres lignée cellulaires (Beck *et al.*, 2013).

Name and composition	Classification	Structure	Exemplary relative abundance for a therapeutic antibody[a]
G0F [H3N4F1]	Complex, fucosylated		35.5
G1F [H4N4F1]	Complex, fucosylated		43.4
G2F [H5N4F1]	Complex, fucosylated		9.5
G1FS [H4N4F1S1]	Complex, fucosylated		0.2
G2S1F [H5N4F1S1]	Complex, fucosylated		0.7
G2S2F [H5N4F1S2]	Complex, fucosylated		0.1
G0 [H3N4]	Complex, nonfucosylated		4.6
G1 [H4N4]	Complex, nonfucosylated		3.3
G2 [H5N4]	Complex, nonfucosylated		0.3
G0F-N [H3N3F1]	Hybrid (monoantennary)		0.5
G0-N [H3N3]	Hybrid (monoantennary)		0.4
M5 [H5N2]	High mannose		1.5
M6 [H6N2]	High mannose		0.1

■, *N*-acetylglucosamine; ●, mannose; ○, galactose; ▲, fucose; ◆, *N*-acetylneuraminic acid.

Figure 10 : Principales glycoformes retrouvées sur les anticorps recombinants produits par des cellules CHO (Reusch et al., 2015)

La glycosylation des AcM influence fortement, à la fois leur structure et leur stabilité, mais aussi leur fonctionnalité et leur activité biologique. Ainsi, les glycannes stabilisent la région C_{H2} des AcM et protègent la chaîne polypeptidique vis-à-vis de la protéolyse, tandis qu'une déglycosylation rend les AcM thermiquement moins stables et plus susceptibles de déplier leurs chaînes peptidiques. De plus, la glycosylation semble améliorer leur solubilité en réduisant leur propension à l'agrégation (Zheng *et al.*, 2011). Outre ces effets de stabilisation, les glycannes, et en particulier la galactosylation terminale, influencent la conformation de la partie Fc des AcM, et jouent un rôle essentiel dans les interactions entre l'AcM et les molécules réceptrices (Krapp *et al.*, 2003). Même s'il est connu que la glycosylation permet d'augmenter la demi-vie circulatoire de protéines comme l'interféron-γ, l'érythropoïétine, l'activateur tissulaire du plasminogène et les gonadotrophines, cet effet sur les

AcM reste peu clair, car probablement dépendant de leur isotype (Eon-Duval *et al.*, 2012). De plus, les glycannes des AcM diminuent les risques d'immunogénicité par le masquage d'épitopes. Enfin, la glycosylation est essentielle pour l'activité biologique des AcM. Son absence abroge à la fois la fonction de cytotoxicité à médiation cellulaire dépendante des anticorps (antibody-dependent cell-mediated cytotoxicity - ADCC), en diminuant l'affinité de couplage avec les récepteurs FcγRI, FcγRII, FcγRIII, et la fonction de cytotoxicité par activation du système du complément (complement-dependent cytotoxicity - CDC), en réduisant l'affinité de couplage avec le récepteur C1q (Leatherbarrow *et al.*, 1985; Patel *et al.*, 2010).

La micro-hétérogénéité présente aussi un impact non négligeable sur l'activité biologique des AcM. Par exemple, dans le cas de l'anticorps anti-RhD, un niveau élevé de galactosylation améliore les performances de ces AcM (Kumpel *et al.*, 1994). L'absence de fucose sur le noyau saccharidique augmente de 50 fois l'efficacité immunologique des AcM, notamment au niveau de l'ADCC. La présence de GlcNAc bissectée augmente aussi, mais plus faiblement, l'ADCC. Par ailleurs, l'ajout d'acide sialique terminal augmente l'effet anti-inflammatoire des AcM (Shade et Anthony, 2013). Cependant, un article récent montre que les anticorps anti-RhD galactosylés et sialylés sont immunosuppressifs et induisent une tolérance contre les globules rouges RhD^+ (Winkler *et al.*, 2013).

1.3.3. Facteurs susceptibles d'influencer la glycosylation des anticorps au cours du procédé

Etant donnée l'importance de la glycosylation des AcM sur leurs propriétés biologiques et physico-chimiques, il est essentiel de bien connaître l'influence des conditions opératoires du procédé de production susceptibles d'affecter cette glycosylation.

a) Lignée cellulaire

Puisque chaque lignée cellulaire possède son propre équipement enzymatique qui conditionne les réactions intracellulaires, la glycosylation peut varier significativement selon le type de cellules utilisées. Le choix de la lignée cellulaire est donc un paramètre essentiel à considérer initialement pour produire une protéine recombinante. Ces différences de glycosylation sont souvent attribuées à la présence ou à l'absence, ou encore à l'activité variable, des glucosaminyltransférases, sialyltransférases, α1-3 galactosyltransférases et fucosyltransférases, au sein des diverses lignées cellulaires. Les cellules de mammifères les plus utilisées pour la production des glycoprotéines sont les cellules CHO et les cellules de myélomes murines NS0, bien que d'autres lignées, telles que les cellules SP2/0 ou Y0 soient également utilisées dans une moindre mesure (del Val *et al.*, 2010). La

diversité des glycoformes en fonction de différents systèmes d'expression a été comparée en détail par Enon-Duval et ses collaborateurs (Tableau 2) (Eon-Duval *et al.*, 2012).

Tableau 2 : Répartition des glycoformes de protéines produites par différents systèmes d'expression cellulaire (Eon-Duval et al., 2012)

	Human Plasma				CHO Cell					NS0 Cell		Sp2/0 Cell	
						(106)					(106)	(106)	
	(103)	(105)	(104)	(106)	(105)	Xolair	Herceptin	Rituximab	4E10	(105)	Zenapax	Erbitux	Remicade
G0F*	20–40%	34%	34%	23–35%	61%	60%	45%	53%	19%	15%	44%	56%	31%
G1F*	40%	45%	45%	45–50%	36%	35%	45%	41%	44%	53%	43%	38%	51%
G2F*	20–40%	21%	21%	19–27%	3%	4%	10%	6%	37%	26%	13%	6%	17%
Total GxF				70–73%		87%	77%	96%	70%		88%	77%	88%
Fucose			97%	91–93%		88%	82%	96%	94%		98%	81%	90%
Sialic acid			11%	10–12%		0%	1%	0%	26%		0%	0%	0%
Bis GlcNAc			16%	9–12%		0%	0%	0%	0%		1%	0%	0%
High mannose			<0.1%	0%		4%	7%	1%	0%		1%	15%	3%
Hybrid			2%	0%		3%	1%	1%	0%		1%	2%	2%

G0F, biantennary, agalactosylated, fucosylated; G1F, biantennary, monogalactosylated, fucosylated; G2F, biantennary, digalactosylated, fucosylated; total GxF, G0F + G1F + G2F; Bis GlcNAc, bisecting *N*-acetylglucosamine.
*Reported proportions of G0F, G1F, and G2F glycoforms are relative to Total GxF. All other reported proportions are relative to total glycoforms.

b) Composés du milieu de culture

La disponibilité de la source de carbone au cours du procédé de culture peut provoquer un impact important sur la macro et la micro-hétérogénéité de la glycosylation des protéines. L'épuisement du milieu en glucose provoque une diminution du taux d'occupation des sites de glycosylation ainsi que du niveau de galactosylation et de sialylation des AcM (Liu *et al.*, 2014). Le pourcentage d'AcM non-glycosylé peut ainsi augmenter jusqu'à 45% des AcM totaux au cours d'une culture de cellules CHO. En effet, lorsque le niveau de glucose n'est plus suffisant, les cellules l'utilisent préférentiellement pour leur métabolisme énergétique, plutôt que pour reconstituer le pool de nucléotides-sucres servant de précurseurs de glycosylation (Miller *et al.*, 2011). D'autres sucres ont aussi été testés afin d'évaluer leur influence sur la glycosylation. L'ajout de galactose induit une augmentation de la galactosylation pour différentes glycoprotéines (Hossler *et al.*, 2009; Weikert *et al.*, 1999), comme dans le cas d'un AcM pour lequel l'ajout de 20 mM de galactose a permis d'augmenter la galactosylation de 14% à 25% (Kildegaard *et al.*, 2016). D'autres études montrent que la supplémentation du milieu par du galactose est responsable d'une augmentation de la sialylation des protéines (Clark *et al.*, 2005; Wong *et al.*, 2010). En outre, l'addition de mannose peut diminuer légèrement le niveau de GlcNAc (Kildegaard *et al.*, 2016). La glutamine, autre source importante de carbone et d'énergie, intervient également dans les voies de biosynthèse des nucléotides puriques et pyrimidiques. Différentes études ont montré son effet sur la glycosylation des protéines. En conditions limitantes, une diminution du taux d'occupation des sites de glycosylation a été observée pour l'IFN-γ, une diminution de la synthèse d'un précurseur de glycosylation, l'UDP-GlcNAc, pouvant

être responsable de cet effet (Nyberg *et al.*, 1999). Dans le cas d'un AcM, cette limitation en glutamine a provoqué une augmentation de la glycoforme Man5 (une des formes fortement manosylées), également liée à la diminution du niveau d'UDP-GlcNAc (Fan *et al.*, 2015). L'influence sur la glycosylation des protéines d'autres composés ajoutés au milieu de culture, comme les acides aminés, le manganèse, le butyrate de sodium et les nucléotides-sucres, a aussi été étudiée. Leurs effets sont regroupés dans le Tableau 3.

Tableau 3 : Exemples de l'influence de composés du milieu de culture sur la glycosylation de protéines recombinantes

Composé	**Effet sur la glycosylation**	**Référence**
acides aminé	augmentation de la sialylation d'EPO (Érythropoïétine)	(Crowell *et al.*, 2007)
manganèse	augmentation du taux d'occupation du site de glycosylation et de la galactosylation d'EPO. diminution de la forme G0F et augmentation du niveau de Man5	(Crowell *et al.*, 2007; St. Amand *et al.*, 2014)
butyrate de sodium	augmentation de la sialylation de l'IFN-γ diminution de la galactosylation de l'anticorps augmentation du taux d'occupation du site de glycosylation de tPA	(Andersen *et al.*, 2000; Hong *et al.*, 2014; Lamotte *et al.*, 1999)
glycérol	augmentation de la sialylation de l'IFN-β	(Rodriguez *et al.*, 2005)
lipides	augmentation du taux d'occupation du site de glycosylation de l'IFN-γ	(Castro *et al.*, 1995)

c) Ions ammonium et pH

Il a été montré un effet synergique de la concentration en ions ammonium et du pH sur la macro-hétérogénéité de la glycosylation des protéines. A pH élevé, l'augmentation de la concentration en ions ammonium provoque une augmentation plus importante de protéines non-glycosylées qu'à pH proche de 7,2. De plus, pour de faibles concentrations en ions ammonium, la glycosylation est relativement peu affectée dans la gamme de pH entre 6,9 et 8,2 (Borys *et al.*, 1993). D'autres chercheurs indiquent que l'accumulation d'ions ammonium réduit la galactosylation et la sialylation d'un AcM par l'augmentation du pH intracellulaire (Gawlitzek *et al.*, 2000).

d) Oxygène dissous

L'oxygène dissous, autre paramètre de culture essentiel pour les cellules, provoque des effets variables sur la glycosylation selon la glycoprotéine produite et la lignée cellulaire productrice. Il a été observé qu'au cours d'une culture de cellules CHO, le niveau de sialylation de la FSH (follicule stimulating-hormone) baisse lorsque la pression partielle d'oxygène dissous (pO_2) diminue (Chotigeat *et al.*, 1994). A l'inverse, d'autre chercheurs montrent qu'une pO_2 élevée (100% de la saturation en air) induit une diminution de la sialylation d'EPO produite par CHO (Trummer *et al.*, 2006). La galactosylation d'un anticorps produit par hybridomes diminue quand la pO_2 est inférieure à 10% de la saturation en air (Kunkel *et al.*, 1998). En revanche, une autre étude montre que la glycosylation de t-PA produit par cellules CHO n'est pas affectée par des conditions d'hypoxie sévère (Lin *et al.*, 1993). Ainsi, en dépit des efforts réalisés pour comprendre l'interaction entre l'oxygène dissous et la glycosylation des protéines recombinantes, les mécanismes mis en jeu n'ont pas encore été clairement définis.

e) Température, viabilité des cellules, et phase de croissance

L'influence de la température sur la glycosylation des protéines est liée à la viabilité et au cycle cellulaire. Une diminution de température de 37 à 32 °C est souvent utilisée pour maintenir la viabilité des cellules, prolonger la durée de culture et stimuler la vitesse de production, sachant que certains auteurs ont montré que le mécanisme de glycosylation est plus efficace sous conditions de viabilité cellulaire élevée (del Val *et al.*, 2010). Une étude indique que la diminution de température induit des augmentations corrélées du taux d'occupation des sites de glycosylation et de la fraction de cellules en phase G0/G1 du cycle celllulaire (Andersen *et al.*, 2000). D'autres travaux montrent, qu'à 32 °C, l'antennarité et la sialylation de l'EPO produite par les cellules CHO diminuent par rapport à 37 °C (Yoon *et al.*, 2005).

1.3.4. Méthodes d'analyse des anticorps et de leur glycosylation

Afin de contrôler et maîtriser un procédé de production des AcM, il est indispensable de pouvoir analyser quantitativement et qualitativement les AcM produits. Les méthodes d'analyse classiquement utilisées nécessitent quasiment toutes un prélèvement d'échantillon et une analyse hors ligne. Les méthodes permettant de quantifier les AcM sont principalement basées sur une approche immuno-enzymatique, comme la méthode ELISA (enzyme-linked immunosorbent assay). Cette méthode permet de visualiser une réaction antigène-anticorps grâce à une réaction colorée entre une enzyme préalablement fixée à l'anticorps et un substrat. Cependant, elle peut être longue

et coûteuse et ne permet d'analyser qu'une faible quantité d'échantillons à l'échelle du laboratoire (Baker *et al.*, 2002). Ces tests ELISA sont donc peu adaptés aux procédés industriels qui génèrent de très grandes quantités d'échantillons. Des techniques alternatives peuvent alors être utilisées pour la quantification hors ligne des AcM. La chromatographie en phase liquide à haute performance (HPLC) est de plus en plus utilisée en raison du développement de nouvelles colonnes d'affinité assurant une meilleure précision et un temps de mesure assez court de moins de 10 minutes. Malheureusement, elle est principalement adaptée à des protéines de petite ou moyenne taille. Son extension aux protéines de grande taille, telles que les AcM, reste difficile en raison de leur nature complexe et hydrophobe qui provoque une mauvaise récupération et une résolution limitée (Dillon *et al.*, 2006). Un autre inconvénient majeur de cette méthode réside dans le coût élevé des colonnes et la variabilité intra-colonnes.

Jusqu'à présent, aucune de ces techniques de quantification globale des AcM ne permet de fournir une information qualitative sur leur glycosylation. Or, la caractérisation de la glycosylation est devenue une exigence réglementaire pour les protéines thérapeutiques, en raison de ses effets potentiels sur l'activité et/ou l'immunogénicité de ces protéines (Berkowitz *et al.*, 2012). C'est pourquoi, ces deux dernières décennies ont vu des développements importants de techniques analytiques liées à la glycosylation des protéines. De manière générale, la glycosylation des protéines peut être caractérisée à trois niveaux, à savoir, la protéine intacte, les glycopeptides libérés par action d'une protéase, ou les glycannes libérés par action de PNGase et de glycosidases (Figure 11).

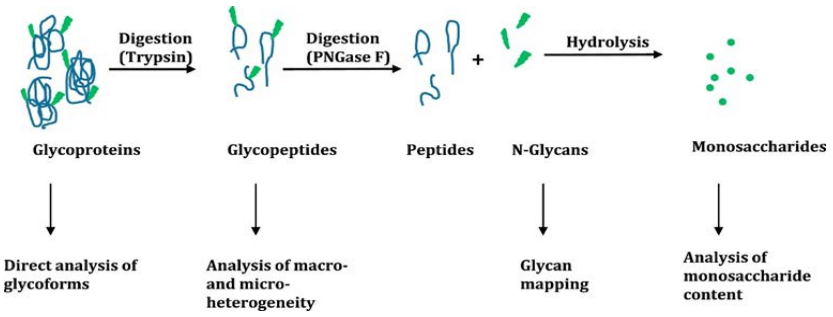


Figure 11 : Analyse de la glycosylation des protéines à différents niveaux (Planinc et al., 2016)

L'analyse de la glycosylation de la protéine intacte présente le grand avantage de réduire le temps de préparation de l'échantillon. Parmi la variété de méthodes analytiques existantes, l'électrophorèse capillaire (CE) et la chromatographie liquide (LC) couplées à la spectrométrie de masse (MS) à haute résolution sont les techniques le plus couramment utilisées (Planinc *et al.*, 2016). L'analyse à ce niveau permet de détecter les différents variants de la glycoprotéine tels que les variants oxydés et acétylés ainsi que les glycoformes majeures. Cependant, elle ne permet pas la détection de glycannes

peu abondants. Par conséquent, des informations structurelles plus complètes ne peuvent être obtenues qu'au niveau des glycopeptides et des oligosaccharides (Dotz *et al.*, 2015).

L'analyse des glycopeptides permet d'étudier à la fois la macro et la micro-hétérogénéité de glycosylation, ce qui est essentiel à la compréhension de la contribution des structures glycanniques à la sécurité et à l'efficacité de la protéine thérapeutique (Zhang *et al.*, 2016). Généralement, les glycoprotéines sont d'abord hydrolysées par des protéases spécifiques telles que la trypsine, l'endoprotéinase Lys-C ou la pepsine. Le mélange de glycopeptides obtenu est ensuite séparé par diverses méthodes, telles que LC ou CE, puis analysé par MS de type MALDI (matrix-assisted laser desorption) ou ESI (electrospray ionization). Néanmoins, le traitement des données, y compris l'analyse des spectres de masse des glycopeptides, requiert beaucoup de temps et une grande expertise (Planinc *et al.*, 2016).

L'analyse des glycannes, également appelée « glyco-profiling » ou « glycan-mapping », permet de fournir l'information structurelle la plus complète sur la glycosylation des protéines. Cependant, cette analyse est relativement compliquée, longue, et coûteuse. En général, les glycannes doivent d'abord être séparés de la chaîne peptidique par la peptide-N-glycosidase F (PNGase F). Les glycannes obtenus sont ensuite marqués par fluorescence pour être dérivatisés. Leur séparation est couramment réalisée par LC ou CE avec détection de fluorescence (Mariño *et al.*, 2010). Le marquage par fluorescence est une étape critique du « glyco-profilling » en raison de la possibilité d'une dérivatisation incomplète, ce qui entraîne une plus grande variabilité inter-laboratoires (Wada *et al.*, 2007). Le « glyco-profiling » est encore plus compliqué en présence d'isomères dont la masse est identique et qui sont co-élués en LC. Dans ce cas, l'hydrolyse des glycannes par des exoglycosidases peut être très bénéfique en facilitant la détermination : de la composition en monosaccharides, du type de liaison glycosidique et de l'abondance relative d'un glycanne dans le mélange. Une digestion par étapes utilisant différentes combinaisons d'exoglycosidases spécifiques est effectuée, ce qui élimine séquentiellement les monosaccharides terminaux, pour les identifier et connaître leur séquence dans les chaînes glycanniques (Planinc *et al.*, 2016).

En résumé, la glycosylation joue en rôle très important dans l'innocuité et l'efficacité des AcMs thérapeutiques. Néanmoins, le profil de glycosylation d'un AcM est susceptible d'être influencé par des conditions opératoires et des paramètres biochimiques de la culture. Il est ainsi essentiel de pouvoir suivre et contrôler ces paramètres afin d'assurer la qualité des AcMs produits. Actuellement, seuls certains paramètres physico-chimiques basiques sont suivis en routine et en temps réel dans

les industries pharmaceutiques. Le développement de nouveaux outils et méthodes s'avère donc nécessaire pour le suivi en ligne de paramètres biochimiques.

De plus, peu de techniques permettent d'analyser la quantité de l'anticorps en temps réel. Ceci est encore plus vrai en ce qui concerne sa qualité, dont en particulier sa glycoylation qui peut être fortement affectée par la mise en œuvre du procédé. C'est pourquoi, afin d'intégrer des stratégies de contrôle de la quantité et la qualité des anticorps produits par une meilleure maîtrise du procédé, il reste essentiel de développer également des méthodes de suivi en ligne de la qualité de ces produits.

2. Techniques spectroscopiques pour le suivi en ligne des procédés de culture de cellules animales

2.1. Enjeux des approches PAT et QbD

Compte-tenu de l'importance économique des AcM thérapeutiques, ainsi que de la complexité et de la grande variabilité des procédés de cultures des cellules productrices, le suivi et le contrôle de ces bioprocédés représentent actuellement un enjeu majeur. En effet, le développement de ces aspects est essentiel pour réduire les coûts de production, augmenter la productivité et assurer une qualité constante des protéines produites. Beaucoup de procédés actuels ne sont pas suffisamment optimaux en raison, principalement, du manque de technologies de suivi en ligne (Clementschitsch et Bayer, 2006). En effet, seuls quelques paramètres physiques et chimiques de base sont couramment suivis et contrôlés, à l'aide de sondes stérilisables, au cours des procédés industriels de cultures de cellules animales, comme le pH, la température ou l'O_2 dissous (Tableau 4).

Tableau 4 : Techniques analytiques utilisées pour le suivi des bioprocédés (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

La plupart des autres paramètres, notamment biochimiques ou physiologiques, nécessitent encore des analyses hors ligne, effectuées après un prélèvement stérile d'échantillon de culture (Claßen *et al.*, 2017). La lourdeur et la durée de ces analyses hors ligne empêchent, la plupart du temps, de fournir des informations en temps réel, informations pourtant essentielles pour mieux comprendre et contrôler le procédé. Par conséquent, ce manque de suivi précis en temps réel constitue un frein pour optimiser davantage les performances de ces procédés et garantir une bonne qualité du produit. C'est la raison pour laquelle, ces dernières années, les entreprises pharmaceutiques apportent une attention croissante au développement de nouvelles techniques de suivi en ligne des bioprocédés.

Dans ce contexte, la FDA a lancé en 2004 l'initiative "Process Analytical Technology" (PAT), qui encourage les industriels à développer et implémenter de nouvelles technologies d'analyse en temps réel au cours de toutes les étapes des procédés de production de molécules biopharmaceutiques (Food and Administration, 2004). L'objectif final est de permettre aux entreprises de mieux connaître leurs procédés pour mieux les contrôler et ainsi assurer une meilleure qualité du produit final. L'initiative PAT est donc directement liée à la méthodologie "Quality by Design" (QbD) (Quality Guidelines: ICH). Dans un premier temps, le QbD consiste à définir le profil de qualité du produit recherché (Quality Target Product Profile, QTPP) et à déterminer ses attributs critiques de qualité

(Critical Quality Arributes, CQAs). Le QTPP intègre, par exemple, la sécurité, l'efficacité, l'immunogénicité et la pharmacocinétique du produit final, qui doivent être caractérisés par les CQAs. Dans un second temps, il consiste à identifier les paramètres critiques du procédé (Critical Process Parameters, CPPs) qui affectent les CQAs. Au final, l'objectif du QbD est de contrôler les CPPs afin de maintenir les CQAs dans une zone de confiance préalablement définie (Figure 12) (Jenzsch *et al.*, 2017).

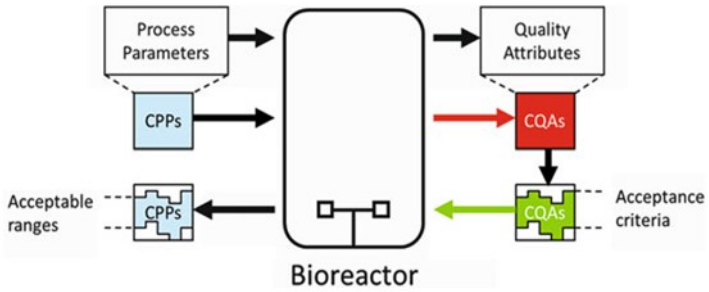


Figure 12 : Représentation schématique de l'approche QbD (Jenzsch et al., 2017)

Ainsi, l'initiative PAT devrait permettre de favoriser la démarche QbD en fournissant une mesure en temps réel des paramètres et attributs du procédé et du produit, et en assurant un meilleur contrôle du procédé. L'optimisation du procédé ainsi permet d'augmenter la probabilité d'obtention d'un produit biopharmaceutique répondant aux exigences de qualité. Les méthodes utilisées pour mesurer les variables du procédé de culture en bioréacteur peuvent être classées en quatre grands types : *in situ* (in-line), en ligne (on-line), hors ligne proche et rapide (at-line) et hors ligne (off-line). La mesure "in-line" est réalisée directement dans le bioréacteur sans perturber la culture suivie. La mesure "on-line" nécessite une boucle externe permettant un flux d'échantillon du réacteur vers les analyseurs, avant recyclage de l'échantillon dans le réacteur. Dans le cas de la mesure "at-line", l'échantillon est prélevé stérilement du réacteur et analysé hors ligne à proximité immédiate et avec une méthode rapide (le plus souvent la spectroscopie ou HPLC). Enfin, pour la mesure "off-line", l'échantillon est transféré sur un autre site géographique et l'obtention des résultats peut être notablement différée. Ainsi, les méthodes "in-line" et "on-line" assurent un suivi en temps réel de l'évolution des paramètres du procédé, ce qui permet d'envisager un contrôle automatique du procédé. Les méthodes "at-line" peuvent également donner des résultats avec un temps d'analyses relativement faible, à condition que l'analyse soit réalisée dans un délai très bref. Enfin, les méthodes "off-line" ne permettent généralement pas d'envisager un contrôle optimal du procédé (Figure 13) (Zhao *et al.*, 2015).

La mesure en ligne des paramètres critiques du procédé s'appuie sur divers outils analytiques utilisant des méthodes chimiques, physiques et mathématiques. Depuis une dizaine d'années, les méthodes spectroscopiques (diélectrique, NIR, Raman, etc.) commencent à occuper une place privilégiée dans le contexte de l'approche PAT. En effet, elles possèdent des avantages majeurs comme des mesures rapides, fréquentes et précises, et l'absence de besoin de prélèvements en conditions aseptiques. De plus, ces méthodes offrent l'intérêt d'être non destructives pour la culture. Enfin, elles peuvent fournir simultanément des informations en temps réel sur plusieurs variables clés du bioprocédé.

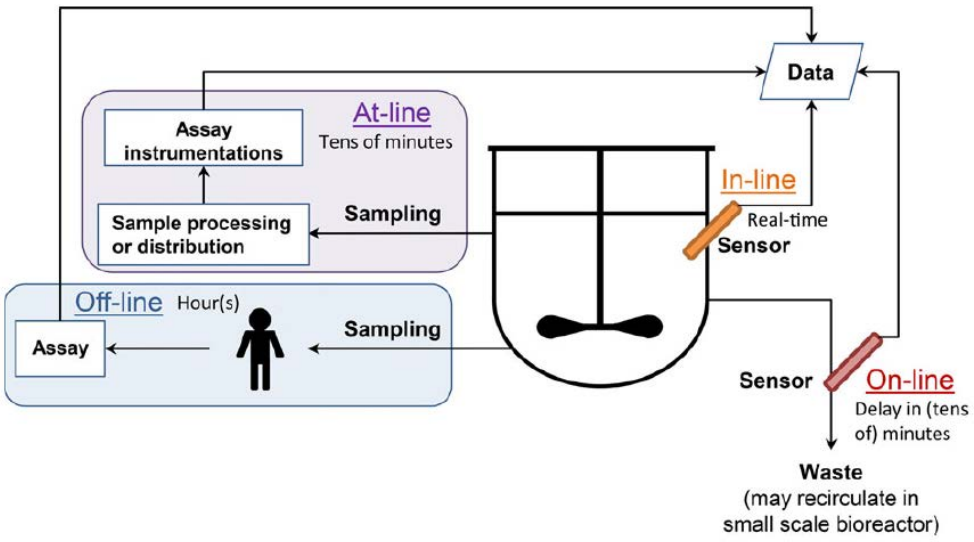


Figure 13 : Représentation schématique de différents types d'analyse des variables dans les bioprocédés (Zhao et al., 2015)

2.2. Intérêt de la spectroscopie diélectrique

Au cours des procédés de culture de cellules animales, la concentration des cellules viables ("viable cell density", VCD) représente l'un des paramètres les plus importants à connaître et à maitriser puisqu'il permet de suivre, non seulement, l'état de croissance cellulaire, mais également, la production de la protéine d'intérêt, celle-ci étant souvent couplée à la croissance. Plusieurs méthodes de mesure en ligne de la concentration cellulaire ont été proposées. Ce sont, soit, des mesures utilisant par exemple des méthodes optiques et spectroscopiques, soit, des méthodes basées sur l'analyse métabolique des cellules telles que l'analyse de la consommation de glucose ou d'oxygène (Höpfner *et al.*, 2010; Junker *et al.*, 1994; Kiviharju *et al.*, 2008). Ces méthodes doivent répondre à plusieurs impératifs comme être robustes, précises et faciles à calibrer, être exemptes de toute interférence biologique, et supporter les conditions de stérilisation.

Parmi ces méthodes, la spectroscopie diélectrique (ou spectroscopie d'impédance, spectroscopie à capacitance) offre de nombreux avantages comme notamment une bonne sensibilité, une rapidité de mesure, une analyse *in situ* stérile, et la capacité de mesurer certaines propriétés cellulaires difficilement accessibles par d'autres méthodes de mesure en ligne (la capacitance membranaire et la conductivité intracellulaire des cellules par exemple) (Carvell et Dowd, 2006; Justice *et al.*, 2011). De plus, la spectroscopie diélectrique est, a priori, moins sensible aux débris cellulaires et aux bulles d'air que les méthodes optiques, et elle donne accès à la concentration de cellules viables.

La spectroscopie diélectrique s'est développée au début du 20ème siècle (Wagner, 1914), mais ses premières applications en biologie n'apparaissent qu'à la fin des années 1950 (Schwan, 1957). Ce n'est qu'en 1987 qu'elle est proposée, pour la première fois, pour le suivi en ligne de la concentration de cellules (Harris *et al.*, 1987). Depuis, elle a été utilisée pour différentes applications intégrant la culture de cellules de mammifères (Cannizzaro *et al.*, 2003; Cole *et al.*, 2015; Downey *et al.*, 2014; Lee *et al.*, 2015), d'insectes (Ansorge *et al.*, 2007; Zeiser *et al.* 2000; Negrete, Esteban *et al.*, 2007), de plantes (Markx *et al.*, 1991), et microbiennes (Arnoux *et al.*, 2005; Neves *et al.*, 2000; November et Impe, 2000). Plusieurs appareils de spectroscopie diélectrique sont disponibles sur le marché depuis plus de 20 ans comme ceux proposés par Aber Instruments Ltd (Futura Biomass Monitor) ou Fogale Nanotech, racheté récemment par Hamilton (iBiomass).

2.2.1. Principe de fonctionnement

Lorsque des cellules viables, placées dans une solution ionique, sont soumises à un champ électrique alternatif, une polarisation des ions intracellulaires apparaît aux pôles de la cellule. En effet, le mouvement des ions dans le cytoplasme cellulaire hautement conducteur est bloqué par la membrane cellulaire non conductrice. Ainsi, chaque cellule vivante peut être considérée comme un mini-condensateur. A l'opposé, les cellules mortes ne peuvent pas se polariser puisque leur membrane est perméable et ne retient plus les ions (Figure 14).

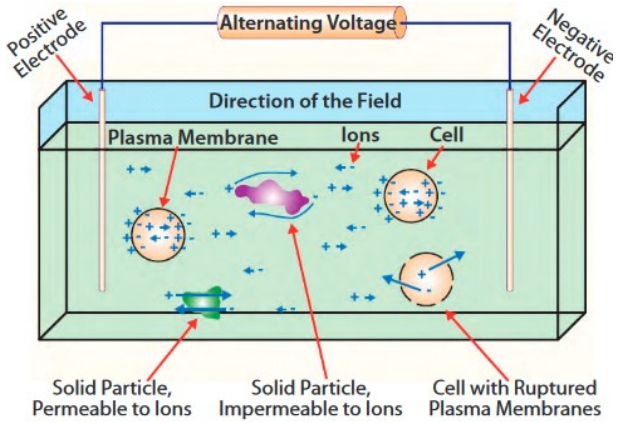


Figure 14 : Représentation schématique du principe de fonctionnement de la spectroscopie diélectrique (Logan et Carvell, 2011)

L'amplitude de cette polarisation est mesurée par la petmittivité, tandis que la vitesse de polarisation des cellules est fonction de la fréquence à laquelle le champ électrique est inversé. Lorsque la fréquence du champ électrique est faible, les ions ont assez de temps pour traverser le cytoplasme, atteindre la membrane plasmique et polariser les cellules avant l'inversion du champ électrique. Les cellules sont alors complètement polarisées et la petmittivité est élevée. Quand la fréquence du champ électrique augmente, la petmittivité diminue puisque de moins en moins d'ions ont le temps de traverser le cytoplasme, la polarisation des cellules étant alors incomplète. A fréquences très élevées (>10 MHz), la durée de polarisation des cellules est insuffisante, et, par conséquent, leur contribution à la petmittivité mesurée est négligeable devant la petmittivité du milieu de culture. Cette diminution de la polarisation cellulaire intervient généralement entre 0,1 et 10 MHz, et donne la courbe de β-dispersion (Figure 15).

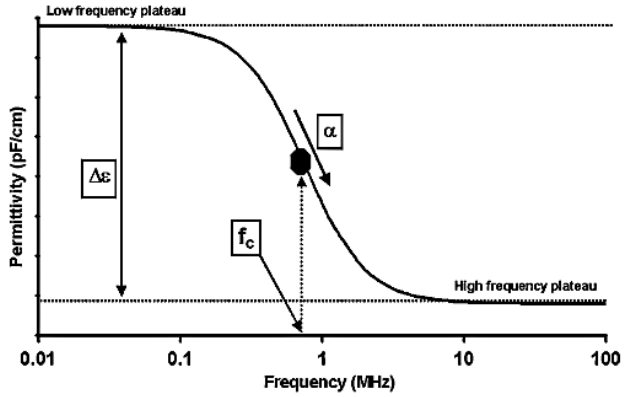


Figure 15 : Courbe de β-dispersion pour les cellules sphériques (Ansorge et al., 2007)

Cette courbe de β-dispersion est typiquement représentée par le modèle de Cole-Cole décrit par l'équation suivante (Cole and Cole, 1941) :

$$\varepsilon(f) = \frac{\Delta\varepsilon\left(1 + \left(\frac{f}{f_C}\right)^{1-\alpha} \sin\left(\frac{\alpha\pi}{2}\right)\right)}{1 + \left(\frac{f}{f_C}\right)^{2(1-\alpha)} + 2\left(\frac{f}{f_C}\right)^{1-\alpha} \sin\left(\frac{\alpha\pi}{2}\right)} + \varepsilon_\infty \quad \textbf{Eq 1}$$

$\varepsilon(f)$: permittivité en fonction de la fréquence f (F.m^{-1})
$\Delta\varepsilon$: différence de permittivité entre la fréquence la plus haute et la fréquence la plus basse (F.m^{-1})
f_C : fréquence caractéristique pour laquelle la permittivité est égale à $\Delta\varepsilon/2$ (Hz)
α : paramètre empirique représentant la pente de la courbe de β-dispersion à f_C
ε_∞ : permittivité plateau atteinte à très haute fréquence (F.m^{-1})

En outre, les paramètres de l'équation de Cole-Cole peuvent être liés aux propriétés diélectriques d'une suspension cellulaire en appliquant la théorie de Maxwell-Wagner et Bruggeman avec les équations simplifiées ci-dessous, la cellule étant assimilée à une sphère (Harris *et al.*, 1987; Schwan, 1957) :

$$f_C = \frac{2\sigma_i\sigma_m + rG_m(\sigma_i + \sigma_m)}{2\pi r C_m(\sigma_i + 2\sigma_m)} \approx \frac{1}{2\pi r C_m \times (\frac{1}{\sigma_i} + \frac{1}{2\sigma_m})} \quad \textbf{Eq 2}$$

$$\Delta\sigma_m = \frac{\sigma}{(1 - \frac{4\pi r^{3N}}{3})^{1,5}} \quad \textbf{Eq 3}$$

$$\Delta\varepsilon = \frac{9prC_m}{4(1 + rG_m\left(\frac{1}{\sigma_i} + \frac{1}{2\sigma_m}\right))^2} \approx \frac{9prC_m}{4} \approx 3\pi r^4 C_m N \quad \textbf{Eq 4}$$

σ_i : conductivité intracellulaire (S.m^{-1})
σ_m : conductivité du milieu de culture (S.m^{-1})
C_m : capacitance membranaire des cellules (F.m^{-2})
G_m : conductance membranaire des cellules (S)
r : rayon de la cellule (m)
p : fraction du volume des cellules
N : VCD (cellule.m^{-3})

En faisant l'hypothèse que les cellules sont sphériques et que le rayon des cellules est relativement constant au cours de la culture, l'équation **Eq *4*** peut être simplifiée pour aboutir à la relation linéaire de l'équation **Eq *5***.

$$\Delta\varepsilon \propto N \quad \textbf{Eq 5}$$

Ainsi, la valeur de $\Delta\varepsilon$ peut être corrélée à la VCD, et est souvent considérée comme un indicateur du biovolume total des cellules vivantes. Plus la quantité de cellules est importante, plus leur volume global est important et plus $\Delta\varepsilon$ est élevé (Figure 16a). La valeur de f_C peut, quant à elle, donner des informations sur la taille des cellules, avec une valeur plus élevée quand la taille cellulaire est plus petite (Figure 16b).

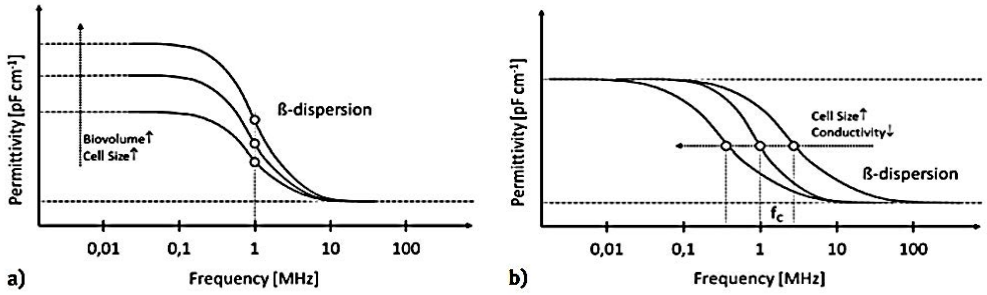


Figure 16 : Variations de la courbe de ß-dispersion en fonction du biovolume total (a) et de la taille (b) des cellules

2.2.2. La spectroscopie diélectrique dans les procédés de culture de cellules animales

La spectroscopie diélectrique a été largement utilisée pour le suivi en ligne de la concentration cellulaire au cours de différents types de procédés de cultures (réacteur à lit fixe, à bulles ascendantes, à cuve agitée), fonctionnant en mode discontinu, semi-continu, perfusé ou continu. Les travaux rapportés dans la littérature sont regroupés dans le Tableau 5.

La plupart du temps, la calibration de la méthode est réalisée par une corrélation linéaire entre la concentration des cellules viables et la permittivité mesurée. Cependant, certains travaux montrent que la précision de cette méthode diminue pendant la phase stationnaire de croissance ainsi que pendant la phase de déclin cellulaire. Une surestimation de la densité cellulaire est alors observée par rapport aux mesures de référence hors ligne (Figure 17) (Downey *et al.*, 2014).

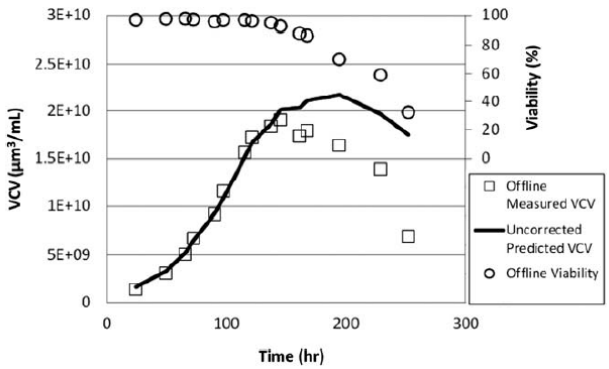


Figure 17 : Exemple de la prédiction de VCV (volume des cellules viables) par la spectroscopie diélectrique (Downey et al., 2014)

Tableau 5 : Application de la spectroscopie diélectrique aux procédés de culture de cellules animales

Lignée cellulaire	Conditions de culture	Méthodes de régression	Variables analysées	Ref.
CHO	Cellules immobilisées en réacteur à lit fixe	Linéaire	TCD	(Ducommun *et al.*, 2002)
CHO	Bioréacteur semi-continu	Linéaire, Cole-Cole, PLS	VCC, vPCV, diamètre des cellules, Cm, σi	(Opel *et al.*, 2010)
CHO K1/dhfr⁻CHO productrice d'anticorps	Bioréacteur 3L semi-continu	/	Permittivité, Fc, déplétion des nutriments	(Ansorge *et al.*, 2010)
CHO 320, HeLa	Fioles agitées (Spinners)	Linéaire	TCD	(Cerckel *et al.*, 1993)
CHO 320	Cellules adhérentes sur microporteurs, bioréacteur	Linéaire	VCD	(Guan and Kemp, 1998)
CHO 320	Bioréacteur	Linéaire	VCD	(Guan *et al.*, 1998)
CHO	Culture en mode perfusé	Inconnu	VCD	(Dowd *et al.*, 2003)
Hybridome	Cellules adhérentes sur microporteurs, spinners, bioréacteur à lit fluidisé	Linéaire	VCD, attachement cellulaire, vitesse de consommation de glutamine	(Noll et Biselli, 1998)
SP2/0	Culture en mode perfusé	Linéaire	VCD	(Carvell et Dowd, 2006)
Vero, productrice de virus	Cellules adhérentes sur microporteurs, bioréacteur discontinu et perfusé	Linéaire, Cole-Cole	VCD, lyse cellulaire	(Rourou *et al.*, 2010)
CHO	Bioréacteurs 2-80 L	Linéaire, PLS	VCD, extrapolation du modèle	(Konakovsky *et al.*, 2015)
CHO	Bioréacteur semi-continu	Linéaire, corrigée avec la demande en oxygène	VCD, LCD, DCD	(Kroll *et al.*, 2017)
Vero	Cellules adhérentes sur des microporteurs	Linéaire	VCD, détection d'apoptose par Fc	(Petiot *et al.*, 2012)
CHO	Bioréacteur	Linéaire	VCD, DCD (en fonction de la conductivité), LCD (à l'aide de la spectroscopie NIR)	(Courtès *et al.*, 2016)
HEK, Sf9, productrices de virus	Bioréacteur	/	Fc, Cm, σi pour l'identification des phases de production virale	(Petiot *et al.*, 2017)
NS/0	Bioréacteur semi-continu	Linéaire, corrigée par un ratio d'aire calculé à partir des données de multifréquences	VCV (viable cell volume)	(Downey *et al.*, 2014)
CHO	Bioréacteur perfusé	PLS	VCD	(Cannizzaro *et al.*, 2003)
CHO	Bioréacteur perfusé	Linéaire, Cole-Cole, PLS	VCD	(Párta *et al.*, 2014)
CHO	Bioréacteur	Linéaire, Cole-Cole, PLS (PLS, KPLS, LWPLS)	VCD	(Lee *et al.*, 2015)

PER.C6	Bioréacteur perfusé	Linéaire, corrigée avec le diamètre moyen des cellules	VCD	(Mercier *et al.*, 2016)

Abréviations : DCD dead cell density; KPLS Kernel partial lest square; LCD lysed cell density; LWPLS local weighted partial least square; PLS partial least square; TCD total cell density; VCD viable cell density; VCV viable cell volume; vPCV packed cell volume (viable).

Cette observation pourrait être liée aux changements de l'état physiologique des cellules ainsi qu'à l'altération des propriétés électrochimiques du milieu de culture (Opel *et al.*, 2010). Pour contourner ce problème, des auteurs ont proposé différentes méthodes pour caractériser la relation entre la permittivité et la VCD en utilisant, par exemple, les équations complètes de Cole-Cole ou des techniques de régression multivariée. Ainsi, Lee *et al.* (2015) ont montré que l'analyse multivariée de l'ensemble de données spectrales diélectriques permet de représenter correctement l'évolution de la VCD pendant toutes les phases de la culture, alors que la méthode linéaire surestimait la VCD à partir de la phase stationnaire. Dans cette étude, la méthode de régression multivariée LWPLS (local weighted partial least square) a donné la meilleure prédiction de la VCD (Figure 18).

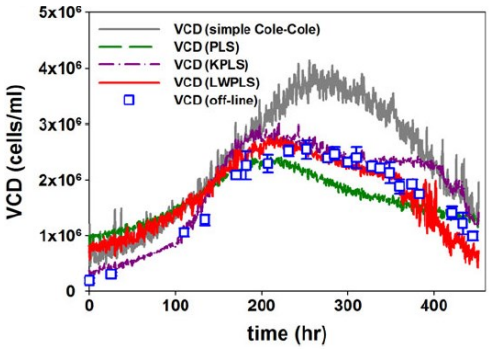


Figure 18 : Différentes méthodes de prédiction de la VCD par la spectroscopie diélectrique (Lee et al., 2015)

Dans une autre étude, les auteurs ont montré que cette déviation de la prédiction de la VCD par spectroscopie diélectrique pouvait être liée aux changements physiologiques des cellules, ce qui affecterait la permittivité du surnageant de culture. Dans ce cas, l'utilisation de la mesure hors ligne du rayon des cellules n'a pas réussi à corriger les écarts observés par rapport au modèle linéaire. Dans ce modèle, la conductance cellulaire interne, σ_i, et la capacitance membranaire par unité de surface, C_m, sont souvent considérées comme constantes. Mais, en calculant ces 2 paramètres à partir des équations de Cole-Cole, les auteurs ont montré qu'ils varient au cours d'une culture en mode discontinu ou semi-continu (Figure 19). Ces variations apparaissent plutôt à partir de la phase stationnaire, indiquant qu'une déviation de la permittivité peut être due à l'évolution de ces 2 paramètres (Opel *et al.*, 2010).

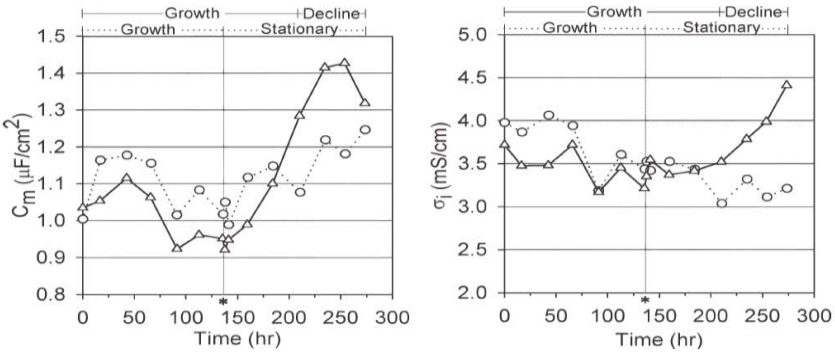


Figure 19 : Evolution de C_m et σ_i au cours d'une culture discontinue (cercles) et semi-continue (triangles) (Opel et al., 2010)

Différents travaux ont clairement montré que la spectroscopie diélectrique présente divers atouts pour le suivi en ligne de la concentration des cellules vivantes au cours des procédés de cultures cellulaires. Certaines études ont également montré le potentiel de la spectroscopie diélectrique pour l'analyse de l'état physiologique de différents types de cellules, mais avec une approche plutôt qualitative, le suivi quantitatif restant difficile en raison de la complexité du métabolisme cellulaire. La vitesse spécifique de croissance cellulaire (μ) est directement liée à l'état physiologique des cellules, qui influence lui-même potentiellement la quantité et la qualité de protéines produites. A ce jour, elle n'a jamais été suivie et contrôlée en temps réel au cours d'un procédé de culture de cellules animales.

Il serait ainsi intéressant d'étudier l'utilisation de la spectroscopie diélectrique en ligne pour le suivi et le contrôle en temps réel de μ, afin de mieux suivre l'état physiologique des cellules, et ainsi optimiser le procédé de production de protéines recombinantes de bonne qualité.

2.3. Intérêt de la spectroscopie infrarouge (IR)

La spectroscopie infrarouge comprend les zones spectrales du proche infrarouge (NIR, longueur d'onde entre 750 et 2500 nm), du moyen infrarouge (MIR, longueur d'onde entre 2500 et 25000 nm) et de l'infrarouge lointain (FIR, longueur d'onde entre 15 et 1000 μm), ce dernier étant rarement utilisé pour le suivi des bioprocédés (Biechele *et al.*, 2015). Les premières applications de la spectroscopie infrarouge pour l'étude de produits chimiques de base ont commencé dans les années 1800 (McClure, 2003). Au milieu du siècle dernier, les progrès en instrumentation et la disponibilité des spectromètres infrarouges commerciaux ont contribué à son implantation dans les industries de la chimie, et des polymères en particulier. A partir des années 1960, ces techniques ont été largement développées dans le domaine agricole pour analyser la composition des céréales, des

fruits, des légumes ou de la viande, etc. (Bellon-Maurel and McBratney, 2011). L'application de la spectroscopie infrarouge aux bioprocédés est relativement récente. Un premier rapport de 1983 décrit son intérêt pour l'analyse quantitative des substrats solides au cours d'un procédé de fermentation (Silman *et al.*, 1983). Son utilisation s'est ensuite généralisée à divers procédés de fermentation de levures, de champignons, et de bactéries (Arnold *et al.*, 2002; Cavinato *et al.*, 1990; Vaidyanathan *et al.*, 1999). Dans le cas des procédés de culture de cellules animales, son intérêt a été démontré plus tardivement vers la fin du XXe siècle, mais a pris rapidement de l'ampleur du fait de ses nombreux avantages, comme l'analyse rapide, simultanée et non-destructrice de la concentration des nombreux composés présents dans le surnageant des cultures.

La spectroscopie MIR offre une meilleure résolution spectrale avec des pics caractéristiques fins et bien définis. Cependant l'absorption de l'eau, très présente dans les bioprocédés, est importante dans cette région spectrale. En revanche, la spectroscopie NIR ne fournit que des bandes d'absorption faibles, larges et se chevauchant, qui correspondent aux harmoniques moléculaires et à la combinaison des vibrations des groupements fonctionnels (Arnold *et al.*, 2004). Cependant, elle est moins sensible à l'eau, plus robuste et moins coûteuse que la spectroscopie MIR, ce qui en fait la spectroscopie NIR la plus adaptée pour des bioprocédés (Vojinović *et al.*, 2006).

2.3.1. Principe de fonctionnement

La spectroscopie infrarouge exploite le fait que les molécules sont capables d'absorber de l'énergie et de tourner ou vibrer lorsqu'elles sont excitées à des fréquences définies. La lumière transmise révèle la quantité d'énergie absorbée à chaque longueur d'onde, fournissant ainsi des détails sur la structure moléculaire d'un mélange de molécules (Stuart, 2004). Les fréquences de la région MIR sont plus basses et génèrent donc une énergie plus faible que celle de la région NIR. La spectroscopie MIR détecte le changement de l'état énergétique lorsque la molécule est excitée. En conséquence, les bandes d'absorption du MIR peuvent être attribuées à des molécules spécifiques ou à des groupements fonctionnels de molécules. Par contre, l'énergie plus élevée de la région NIR implique que les molécules excitées peuvent passer sur plusieurs niveaux vibrationnels différents. Dans ce cas, la mesure est basée sur la combinaison des bandes et les harmoniques moléculaires (overtone) (Figure 20).

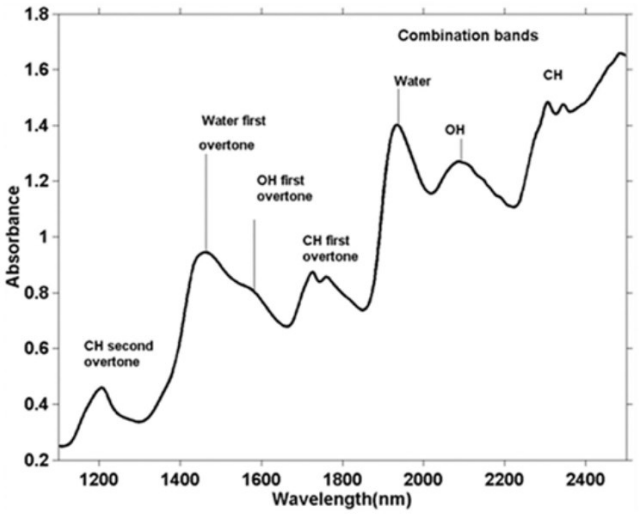


Figure 20 : Exemple de spectre NIR

Les spectres sont souvent présentés en reportant l'intensité ou l'absorbance en fonction du nombre d'onde ν, qui est l'inverse de la longueur d'onde. La relation entre l'absorbance et la concentration d'une molécule peut être calculée en utilisant la loi de Beer-Lambert :

$$A_\nu = -log_{10}\left(\frac{I_\nu}{I_{\nu,0}}\right) = \varepsilon bc \qquad \textbf{Eq 6}$$

A_ν : absorbance au nombre d'onde ν
I_ν : intensité de la lumière réémise au nombre d'onde ν (cd)
$I_{\nu,0}$: intensité de la lumière incidente au nombre d'onde ν (cd)
ε : coefficient d'absorption de la molécule ($cm^2.mole^{-1}$)
b : longueur du chemin optique (cm)
c : concentration de la molécule ($mole.L^{-1}$)

La spectroscopie NIR est particulièrement utile pour détecter les molécules qui contiennent des liaisons C-H aliphatiques, aromatiques ou alcènes, N-H des protéines et O-H des alcools (Biechele *et al.*, 2015). Cette technologie est ainsi bien adaptée à l'analyse : d'éléments nutritifs tels que le glucose, de produits du métabolisme cellulaire tels que le lactate, de produits d'intérêt tels que les protéines recombinantes, ou encore de cellules.

2.3.2. Application de la spectroscopie IR au cas des procédés de culture de cellules animales

Ce point a fait l'objet d'une mini revue publiée en 2016 dans la revue "Journal of Analytical and Pharmaceutical Chemistry", volume 3, issue 2, pages 1065-1070. Le manuscrit de cette publication est présenté ci-dessous.

Manuscrit de la publication

In situ infrared spectroscopy as a PAT tool of great promise for real-time monitoring of animal cell culture processes

Li Mengyao, Ebel Bruno, Courtès Franck, Guedon Emmanuel, Marc Annie

Abstract: Animal cell culture bioprocesses have become essential in pharmaceutical field for the production of recombinant therapeutic proteins, such as monoclonal antibodies (mAbs). Consequently, the Process Analytical Technology approach recommends the use of *in situ* characterization tools, such as infrared (IR) spectroscopy, to control production processes and ensure the quality of end-products. This review presents the current status of the implementation of IR spectroscopy in animal cell culture processes. IR spectroscopy appears to be a very promising tool due to its flexibility, simplicity, rapidity and its inherent ability to provide simultaneous multi-analyte information with a single spectrum. Taken into account the main characteristics of animal cell bioprocesses, the advantages and challenges of IR spectroscopy applications are discussed in this review. Then, published works underlining the interest of IR spectroscopy for animal cell bioprocesses are presented, which clearly demonstrated that IR spectroscopy is not only applicable for global supervision of process, but is also suitable for real-time monitoring of key parameters in animal cell cultures. Finally, this review also highlights some future improvement needed to strengthen IR spectroscopy as a reliable industrial PAT tool for feed-back control and process optimization, towards guaranteed end-product quality.

Introduction

The past decade witnessed a great expansion of the production of recombinant therapeutic proteins such as mAbs. Animal cells, mainly Chinese Hamster Ovary (CHO) cells, are the most used host cell lines because of their innate capacity to perform human-like PTMs, which are crucial for functionality and efficiency of the therapeutic proteins [1]. Therefore, to assure the quality of final products which are subject to high variabilities, the Quality by Design (QbD) approach has been outlined by the FDA in 2002, urging manufacturers to monitor their process critical quality attributs (CQAs) and optimize critical process parameters (CPPs) within a specified design space during the process [2]. In this context, the Process Analytical Technology (PAT) concept is considered as the most important tool for the implementation of QbD, since it encourages the biopharmaceutical industries to adopt modern monitoring tools based on real-time analysis of key variables during all stages of the

manufacturing processes [3]. To meet this need, spectroscopic methods, particularly infrared (IR) spectroscopies, have gained great attentions over the last decade. The aim of this review is to present the current status of IR spectroscopies in animal cell bioprocesses as monitoring tools.

Specificities for implementation of IR spectroscopy in animal cell bioprocesses

Industrial production of therapeutic proteins is mainly performed inside stirred and aerated reactors [4] (Figure 1). Several specificities of animal cell cultures in reactor must be taken into consideration for a successful IR spectroscopy implementation: (i) the nutritional requirements of animal cells is generally fulfilled with complex media [5]; (ii) strict sterility is needed to avoid any microbial contamination; (iii) animal cells can be cultured either in single cell suspension or after adhesion on spherical microcarriers [4].

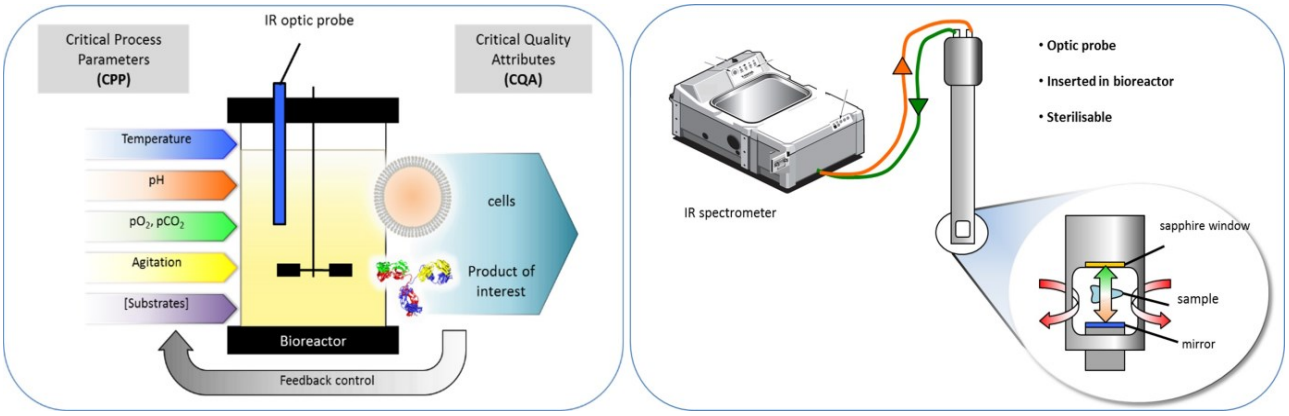


Figure 1: QbD approach using in situ IR probe as a PAT tool.

For bioprocesses performed in reactors, IR spectroscopy exhibits major advantages compared to other classical analytical techniques. It exploits the fact that molecules absorb specific frequencies that are characteristic of their structure. Near infrared (NIR) (750-2,500 nm) and mid infrared (MIR) (2,500-25,000 nm) are the main IR spectroscopies used in bioprocesses due to their simplicity and rapidity of measurement, great flexibility in application and the possibility to provide simultaneous multi-analyte information from a single spectrum. However, to perform analysis directly inside bioreactors, *in situ* optical probes providing information of key variables in real-time are needed. Strategies coupling IR spectroscopies with *in situ* probes have been recently developed, and represent a significant advance because none of the previously existing sensors were able to monitor medium compounds in real-time, therefore requiring sampling with increased risks of contamination [6]. Nevertheless, it has to be noticed that *in situ* probes can be affected by interfering with solid particles, gas bubbles, stirring rate, high cell densities or bulk viscosity, which often occur in bioprocesses [7], but these effects are generally reduced in animal cell culture since cell density as

well as agitation and aeration rates, are rather low. Despite these benefits, several limitations have to be considered before implementing IR spectroscopy in animal cell processes.

One challenge is related to complexity of animal cell culture media and to the fact that there is no single and obvious IR spectral region for one particular component. Consequently, IR spectra display overlapping peaks, leading to complex signals and hindering the assignment of specific features to individual compounds [8]. Moreover, the concentration of some compounds are often inversely correlated (*e.g.* glucose and lactate) which requires additional decorrelation experiments to perform real-time monitoring [9]. Additionally, culture medium compositions change greatly during the culture due to nutrient consumptions and product accumulations. Throughout the cell death phase, the spectroscopic characteristics of the culture bulk become even more complex due to cell debris, thereby increasing the challenge to monitor target bioprocess variables [8].

Another challenge lies in monitoring components that are present at only moderate concentrations over the time-course of mammalian cell bioprocess, since the molar absorptivity of molecules in NIR range is typically small. In MIR range, the vibration of the chemical bonds gives stronger signals and better degree of resolution [10]. However, water bonds absorption is also much stronger in MIR region, which often masks important information of key-molecule variations.

Most of the applications of IR spectroscopies for animal cell cultures reported in literature are applied to single cells cultivated in suspension. However, in the case of adherent cell culture, it is essential to take into consideration the presence of microcarriers in the sensing space, increasing greatly the complexity of spectral information analysis [9].

Finally, all these characteristics and complexities related to animal cell processes require sophisticated multivariate data analysis (MVDA) techniques, such as chemometrics, to reduce data dimensions and to relate IR spectral information with the target compounds. The chemometric methods most widely used for spectral data analysis in bioprocess monitoring are principle component analysis (PCA), principle component regression (PCR) and partial least square (PLS) regression [11].

Reported works highlighting the interest of IR spectroscopy for animal cell bioprocesses

In recent years, the use of IR spectroscopy in the field of bioprocesses has grown and evolved rapidly from raw material testing and final product quality testing to process monitoring [12]. Advances in instrumentation and chemometric techniques have largely contributed to this expansion in a wide

range of bioprocesses [11]. In the case of mammalian cell cultures, the use of IR spectroscopy mainly focuses on two different purposes: process supervision and process monitoring. Table 1 summarizes the reported works displaying off-line and *in situ* applications of IR spectroscopies in animal cell cultures.

IR spectroscopy as a global supervision tool for animal cell processes

Taking advantage of the multivariate nature of IR data, which contain both physical and chemical information, process trajectories can be analyzed using IR spectra coupled with MVDA techniques. This approach is based on a global comparison of spectra patterns with standard ones, without quantification of any particular medium component. It is mostly relevant to analyze differences in historical culture batches, to detect abnormal runs and to guarantee that process remains inside the design space. For example, NIR spectroscopy was used to characterize a mAb production process in which several cultures were compared according to their overall behavior and some batch deviations could be successfully identified [13]. Another study reported the possibility of using NIR combined with PCA to identify batch homogeneity between batches and to detect early stage contamination [14]. Additionally, NIR was also found to be able to detect variabilities in raw materials of cell culture media [15]. Indeed, such variabilities can lead to large unpredictability of cell culture process performances, which impacts the process CQA consequently. A combined NIR/MVDA approach, analyzing the fingerprint of raw materials, allows a robust selection of medium lot while providing a biological link between chemical composition of raw materials and cell culture performances [16].

Table 1: Reported works with off-line and in situ IR monitoring of animal cell culture processes.

Cell line	IR tool	Analyse mode	Monitored component	Reference
Insect cells	NIR	Off-line	Ala, Glc, Leu, Gln	[33]
N.A.	NIR	Off-line	Glc, Lact, Ammo	[20]
CHO	FT-NIR	Off-line	Glc, Gln, Glu, Lact, Ammo, Pyr, 14AA	[21]
CHO, NS0	FTIR	Off-line	mAb, Glc, Lact	[34]
CHO	NIR	Off-line	Gln, Glu, Glc, Lact, Ammo, VCD, mAb, LDH, Osmolality	[35]
NS0	FTIR-ATR	Off-line	mAb, HCP	[36]
CHO	MIR	Off-line	mAb, Lact, Glu, LDH, VCD, DCD, viability	[37]
MSCs	MIR	Off-line	Glc, Lact, Gln, Ammo	[25]
CHO	NIR	*In situ*	Gln, Glc, Lact, Ammo	[6]
CHO	NIR	*In situ*	Glu, Lact	[24]
HEK	NIR	*In situ*	Gln, Glu, Glc, Lact, Ammo, VCD, pH	[38]
Vero	NIR	*In situ*	Glc, Lact	[9]

N.A.	NIR	*In situ*	Glc, Lact, Ammo, TCD	[13]
CHO	NIR	*In situ*	Glc, mAb, PCV, ivPCV, VCD, iVCC, Osmolality	[14]
CHO	NIR	*In situ*	Glc	[23]
CHO	NIR	*In situ*	LDH	[27]
PER.C6®	NIR	*In situ*	Glc, Lact, TCD	[26]

FAA amino acid; Ammo ammonium; DCD dead cell density; Glc glucose; Glu glutamate; Gln glutamine; HCP host cell protein; iVCC integrated viable cell count; ivPCV integrated viable packed cell volume; Lact lactate; LDH lactate dehydrogenase; N.A. non available; PCV packed cell volume; Pyr pyruvate; TCD total cell density; VCD viable cell density.

IR spectroscopy as a real-time monitoring tool for animal cell process

In order to quantify concentrations of components in cell culture supernatants, it is crucial to establish beforehand calibration models based on off-line measurements with reference methods. Subsequently, these models must be validated in order to perform accurate predictions of target compound concentrations [17]. The first works reporting the use of IR spectroscopy for animal cell culture monitoring were published in late 1990s. Mostly based on off-line analysis, these preliminary studies demonstrated that IR spectroscopy can be a reliable tool to perform simultaneous measurement of different key-parameters in cell culture supernatant (*e.g*. nutrients and products concentration), with a relative high accuracy compared to reference methods [18, 19]. Some later works aimed to improve model predictions. For example, NIR spectral information of various samples, including cell culture samples and aqueous mixtures, were combined to a higher model accuracy [20]. In another work, NIR spectroscopy was used to quantify 19 cellular nutrients and waste products simultaneously in culture medium [21]. Formulated synthetic samples were prepared to provide a wider range of component concentrations and to prevent data from being influenced by correlated phenomena. All these results were proved to be an important step toward further *in situ* IR applications within animal cell processes.

In 2002, the first paper describing *in situ* IR monitoring of mammalian cell bioprocess was published: a MIR spectroscopy was used in association with an ATR diamond probe plunged into bioreactor to monitor real-time concentrations of glucose and lactate during a CHO cell culture [22]. Another work proposed NIR spectroscopy coupled with a fiber optic probe to monitor *in situ* concentrations of four key analyte, glucose, lactate, glutamine, and ammonia, over a CHO cell culture [6]. Some later works focused on calibration model constructions, and demonstrated that model performances can be improved using semi-synthetic sample data to extend the calibration to wider range of process conditions [23]. In another study, a calibration strategy employing a multiplexing NIR for the

monitoring of multiple reactors simultaneously has been proposed [24]. This method allowed improving and speeding up the model construction.

Latest studies have mainly focused on the expansion of IR spectroscopy applications in culture processes with wide range of operating conditions at different process scales, from laboratory to pilot and industrial ones. In one recent study, *in situ* NIR was applied on large production bioreactors of 12,500 L, and seven key-parameters were successfully monitored in real-time, including mAb concentration, which is one of the most important CQAs of protein production bioprocesses [14]. As far as we know, this was the first *in situ* mAb titer measurement. Furthermore, IR applications have been expanded to different cell types, from continuous to primary cell lines. Most recently, MIR was implemented in MSCs cultures, and glucose, lactate, and ammonia concentrations were monitored with low prediction errors [25]. IR monitoring of cell cultures have also been tested in the case of different feeding strategies and culture modes. One successful example of *in situ* NIR spectroscopy monitoring of a perfused cell culture was demonstrated, indicating that IR techniques have great potentials for future process control [26].

Conclusion and future challenges

This review highlights the potential of IR spectroscopy as a promising and reliable PAT tool to supervise and monitor animal cell culture bioprocesses. Besides global process supervision, it provides multiparametric, non-invasive, and real-time concentration measurements of key supernatant compounds throughout cell culture processes. In the near future, further developments should make it possible to monitor and control CPPs to ensure CQAs. This will lead to higher process stabilities, and thus better process performances and more stable product qualities.

Currently, the objective of the most published works remains in real-time estimation of medium compound concentrations. However, one future challenge lies, not only in additional compounds supervision but also in *in situ* monitoring of cell physiological state parameters as well as product qualities such as glycosylation of recombinant proteins. While IR based measurements are clearly useful, additional values for implementation of PAT approach can be obtained by combining different types of *in situ*/off-line monitoring tools, such as Raman spectroscopy, dielectric spectroscopy, flux cytometry, mass spectrometry, … For example, a new strategy based on NIR and dielectric spectroscopies combination has been recently proposed to monitor and characterize the different viable, dead and lysed cell populations of CHO cell cultures [27].

It should be noticed that, despite all the attractive benefits of IR spectroscopy, there are still little applications in biopharmaceutical industries. Indeed, while several successful examples have been

reported in the case of microbial fermentations, to our knowledge, there are still no application of IR spectroscopy for the control of animal cell bioprocesses [28, 29, 30]. However, many authors have suggested the great interest of IR implementation in control systems of animal cell cultures, in order to obtain a fully automatic and self-regulating production system [26, 31]. Future works should focus on IR spectroscopy adaptation to industrial bioprocesses, especially for close-loop feedback control of CPP, which is of great importance to control process CQA.

References *(pour alléger le document, les références complètes sont rassemblées à la fin du mémoire)*

1. Abès and Teillaud (2010)
2. Mandenius et al. (2009)
3 FDA (2004)
4. Chu and Robinson (2001)
5. Brunner etal. (2010)
6. Arnold et al. (2003)
7. Arnolda et al. (2000)
8. Teixeira et al. (2009)
9. Petiot et al. (2010)
10. Zhao et al. (2015)
11. Lourenço et al. (2012)
12. Roychoudhury et al. (2006)
13. Henriques et al. (2009)
14. Clavaud et al. (2013)
15. Kirdar et al. (2010)
16. Lee et al. (2012)
17. Miller and Miller (2010)
18. Harthun et al. (1997)
19. Riley et al. (1997)
20. Mcshane and Cote (1998)
21. Riley et al. (2001)
22. Rhiel et al. (2002)
23. Milligan et al. (2014)
24. Roychoudhury et al. (2007)
25. Rosa et al. (2016)
26. Mercier et al. (2016)
27. Courtès et al. (2016)
28. González-Vara et al. (2000)
29. Tosi et al. (2003)
30. Navrátil et al. (2005)
31. Cervera et al. (2009)
32. Sandor et al. (2013)
33. Riley et al. (1998)
34. Sellick et al. (2010)
35. Hakemeyer et al. (2012)
36. Capito et al. (2013)
37. Capito et al. (2015)
38. Card et al. (2008)

L'approche PAT encourage l'utilisation d'outils d'analyse *in situ*, tels que la spectroscopie infrarouge (IR), pour contrôler les procédés de production et assurer la qualité des produits finis. Il ressort de cette revue bibliographique que la spectroscopie IR représente une technique prometteuse en raison de sa flexibilité, de sa simplicité, de sa rapidité et de sa capacité inhérente à fournir des informations multiparamétriques simultanément. En effet, elle permet de suivre en ligne la composition du milieu de culture au cours des procédés de culture de cellules animales. Enfin, cette revue met également en évidence certaines améliorations futures nécessaires pour renforcer la spectroscopie IR en tant qu'outil industriel PAT fiable pour le rétrocontrôle et l'optimisation des processus, en vue d'assurer la qualité finale du produit.

A ce jour, les mesures sont plutôt focalisées sur la concentration des composés du milieu de culture. Aucune étude ne présente le suivi, ni de paramètres cinétiques, ni de la physiologie cellulaire, comme les vitesses spécifiques de croissance/consommation/production, et les rendements métaboliques. Par ailleurs, peu d'études rapportent le suivi en ligne de la production de protéines recombinantes, et aucune ne traite de la qualité de ces protéines, dont en particulier leur glycosylation.

2.4. Intérêt de la spectroscopie Raman

La spectroscopie Raman est basée sur la diffusion inélastique de la lumière monochromatique dans le visible, le proche infrarouge ou le proche ultraviolet. Cet effet est connu sous le nom de diffusion Raman, du nom du physicien l'ayant découvert en 1928. Initialement, la spectroscopie Raman était peu utilisée industriellement en raison d'une capacité limitée de la source lumineuse et d'une faible sensibilité de détection (Zhao *et al.*, 2015). Par la suite, elle est devenue plus robuste grâce à l'utilisation de lasers comme source de lumière et de caméras avec dispositif à transfert de charges (CCD, charg-coupled divice) sensibles comme détecteurs. Les spectres Raman permettent ainsi une analyse structurelle, qualitative et quantitative d'un échantillon, en accédant à des informations sur sa composition, son environnement chimique et sa structure. C'est pourquoi, au cours des trois dernières décennies, l'utilisation de la spectroscopie Raman a fortement augmenté dans différents domaines tels que les polymères, les produits pharmaceutiques, les bioprocédés et les analyses biomédicales (Esmonde-White *et al.*, 2017). Son potentiel a aussi été démontré au cours de toutes les étapes d'un bioprocédé, allant de l'analyse des matières premières et des milieux de culture, jusqu'au suivi des bioprocédés et à la caractérisation des macromolécules produites (Buckley et Ryder, 2017). Grâce au développement des fibres optiques et des sondes à immersion, elle a permis de réaliser des analyses *in situ*, en fournissant simultanément et en temps réel des informations sur de multiples paramètres fonctionnels d'un procédé (Abu-Absi *et al.*, 2011).

2.4.1. Principe de fonctionnement

La spectroscopie Raman fournit des informations sur les transitions vibrationnelles, rotationnelles et autres basses fréquences dans les molécules pour lesquelles la polarisabilité change lors de l'excitation par la lumière. Le signal Raman ne peut donc être détecté que s'il existe un changement de polarisabilité des vibrations moléculaires, cette polarisabilité étant directement liée à l'intensité du signal Raman (Larkin, 2011).

Lorsque la lumière monochromatique est incidente sur une molécule, la grande majorité des photons qui subissent une diffusion est dispersée de manière élastique (diffusion de Rayleigh). Cependant, une proportion infime (1 sur 10^9 ou 10^{10}) des photons est diffusée de manière inélastique (Lewis et Edwards, 2001). Celle-ci se produit lorsque la fréquence de la lumière monochromatique, provenant

d'une source laser, est modifiée lors de l'interaction avec l'échantillon. Les photons sont absorbés par l'échantillon puis réémis à une fréquence, soit avec plus d'énergie (anti-Stokes), soit avec moins d'énergie (Stokes), par rapport à la fréquence monochromatique d'origine (Figure 21). Ainsi, l'intensité des signaux Raman (Stokes et anti-Stokes) est proportionnelle à la concentration de molécules présentes, et varie en fonction du nombre d'onde (Raman shift).

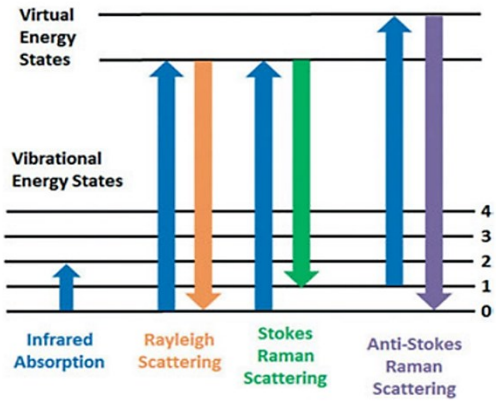


Figure 21 : Représentation schématique de l'effet Raman

Bien que les gammes spectrales des spectroscopies IR et Raman soient assez similaires, elles fournissent des informations différentes sur l'état énergétique des molécules. D'une part, la spectroscopie IR reflète l'absorption d'énergie liée au mouvement des liaisons spécifiques entre différents états vibrationnels. D'autre part, la spectroscopie Raman détecte la diffusion basée sur le transfert d'énergie entre les différents états vibrationnels (Figure 22).

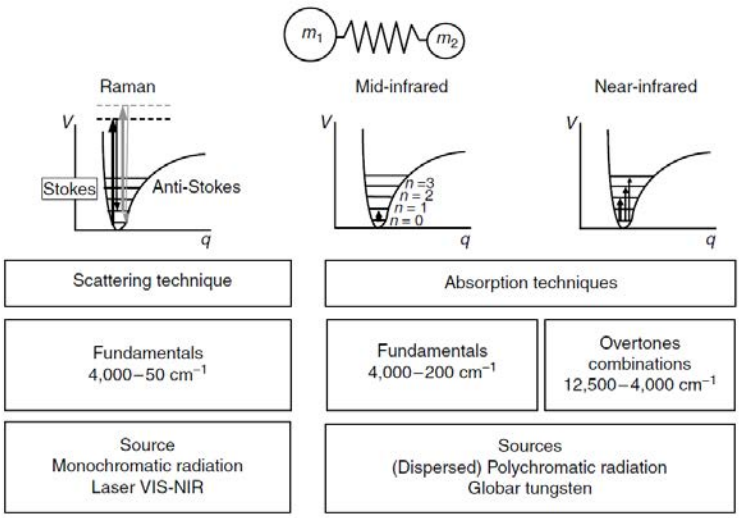


Figure 22 : Différences entre la spectroscopie Raman et les spectroscopies IR

Ainsi, un même composé peut donner un signal plus ou moins fort, selon qu'on utilise l'une ou l'autre des spectroscopies, les deux méthodes pouvant ainsi être complémentaires (Zhao *et al.*, 2015). L'utilisation de la spectroscopie Raman pour le suivi de bioprocédés en phase aqueuse présente un avantage certain car elle peut fournir des empreintes vibrationnelles hautement spécifiques de divers composés chimiques tout en étant relativement insensible à l'eau. En effet, l'impact de l'eau sur le signal Raman est faible et n'est observé que dans la zone de nombre d'onde de 2900 à 4000 cm^{-1}. Par contre, la fluorescence des systèmes biologiques représente un inconvénient majeur pour l'utilisation de la spectroscopie Raman, et nécessite une longueur d'onde d'excitation appropriée. Dans le cas des procédés de culture de cellules animales, une source laser d'excitation de 785 nm peut fournir un bon compromis entre sensibilité de détection et rejet de la fluorescence (Smelko *et al.*, 2014).

2.4.2. Application de la spectroscopie Raman au cas des procédés de culture de cellules animales

Les premiers travaux sur l'application de la spectroscopie Raman *in situ* ont porté sur la production de molécules simples, comme par exemple, l'éthanol produit par fermentation de levures, des hormones végétales et des caroténoïdes. Pour la plupart de ces études, la molécule produite provoquait un signal Raman raisonnablement fort avec des pics bien distincts du bruit de fond (Esmonde-White *et al.*, 2017). Les premiers résultats obtenus avec des procédés de culture de cellules animales n'ont été rapportés qu'en 2011 par Abu-Absi *et al.* Dans ce travail, les spectres Raman collectés pendant 14 jours au cours d'une culture de cellules CHO en bioréacteur de 500 L, ont permis de quantifier simultanément les concentrations de glucose, glutamine, lactate, ions ammonium, cellules totales et cellules viables. Puis, d'autres articles ont rapporté le potentiel de la technique en démontrant la transférabilité des modèles de prédiction entre différentes échelles : à l'échelle laboratoire de 3 à 15 L (Whelan *et al.*, 2012), ou de l'échelle laboratoire (3 L) et de l'échelle pilote (200 L) jusqu'à l'échelle de production (2000 L) (Berry *et al.*, 2015). En 2015, Mehdizadeh *et al.* ont proposé un modèle de régression générique permettant la prédiction par spectroscopie Raman de plusieurs paramètres au cours de divers procédés de culture utilisant différentes lignées cellulaires. Enfin, pour la première fois en 2015, la capacité de la spectroscopie Raman à prédire la concentration d'un anticorps recombinant produit au cours d'un procédé de culture de cellules CHO a été publiée (André *et al.*, 2015). La Figure 23 illustre le suivi en temps réel de la production d'anticorps par la spectroscopie Raman au cours de trois cultures en mode discontinu.

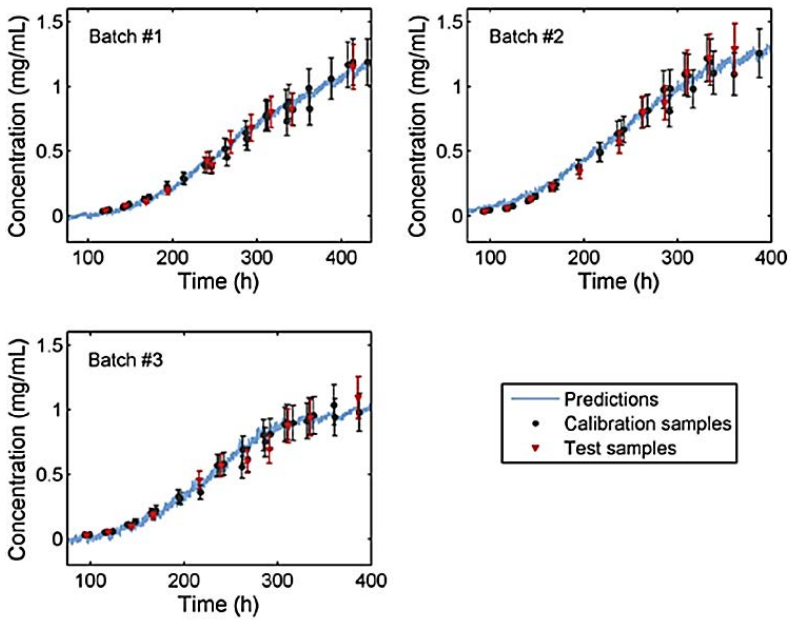


Figure 23 : Suivi en ligne de la concentration des anticorps produits pendant les cultures (André et al., 2015)

Dans une autre étude, l'utilisation de la spectroscopie Raman pour le rétrocontrôle du glucose dans un procédé de culture semi-continue de cellules CHO, a permis une réduction de la glycation de la protéine produite de 9 à 4% (Berry *et al.*, 2016). Par ailleurs, Matthews *et al.* (2016) ont utilisé la spectroscopie Raman pour suivre et contrôler en temps réel les concentrations de glucose et de lactate au cours de cultures semi-continues de cellules HEK. Le glucose était ajouté automatiquement quand le niveau de lactate était en-dessous des points de consigne (4 ou 2,5 $g.L^{-1}$), ou quand la concentration de glucose était inférieure à 0,5 $g.L^{-1}$. Ce contrôle en temps réel de la concentration de lactate produit a permis l'augmentation de la durée de la culture et de la productivité en protéine d'intérêt. Enfin, la spectroscopie Raman a également été utilisée, de manière plus qualitative, pour le contrôle de qualité des milieux de culture (Calvet et Ryder, 2014; Li *et al.*, 2010a), ou pour l'analyse détaillée de la cellule elle-même (détermination de son cycle cellulaire, analyse de sa composition, mort cellulaire) (Rangan *et al.*, 2018). L'ensemble des travaux de la littérature rapportant l'application de la spectroscopie Raman à des procédés de culture de cellules animales est regroupé dans le Tableau 6.

Tableau 6 : Application de la spectroscopie Raman à des procédés de culture de cellules animales

en ligne / hors ligne	cellule	paramètres analysés	quantitative / qualitative	références
hors ligne	**CHO**	celllules apoptotiques, nécrotiques, autophagiques	qualitative	(Rangan *et al.*, 2018)
	Milieu de	contrôle qualité		(Calvet et Ryder,

	culture			2014; Li *et al.*, 2010a)
	CHO	ADN, ARN des cellules		(Ashton *et al.*, 2015)
	A549	cellules vivantes, mortes, différentiation	qualitative et quantitative	(Notingher, 2007)
	A549, MLE-12	viabilité cellulaire, cycle cellulaire	quantitative	(Notingher *et al.*, 2002, 2003)
	MG63	cycle cellulaire		(Swain *et al.*, 2008)
	CHO	glucose, lactate, anticorps		(Ashton *et al.*, 2013)
en ligne	**CHO**	glucose, lactate, glutamine, glutamate, ammonium, VCD, DCD, viabilité	quantitative	(Abu-Absi *et al.*, 2011)
	HEK293	glucose, lactate, rétrocontrôle		(Matthews *et al.*, 2016)
	CHO	glucose, rétrocontrôle		(Berry *et al.*, 2016)
	CHO	anticorps		(André *et al.*, 2015)
	CHO	glucose, lactate, VCD (modèle générique)		(Mehdizadeh *et al.*, 2015)
	CHO, HeLa, sf9, HEK	glucose, lactate		(André *et al.*, 2017)
	CHO	glucose, lactate, glutamate, ammonium, osmolalité, VCD, TCD		(Berry *et al.*, 2015)
	CHO	tyrosine, tryptophane, phénylalanine, méthionine		(Bhatia *et al.*, 2018)
	CHO	glucose (modèle prédictif non linéaire)		(Craven *et al.*, 2014)
	CHO	surveillance du procédé		(Liu *et al.*, 2017)

La spectroscopie Raman présente un potentiel très attractif pour les procédés de culture de cellules animales, qu'il s'agisse de la caractérisation hors ligne des matières premières et de la formulation des produits, ou encore du suivi en ligne de certains paramètres du procédé. Pourtant, à l'heure actuelle, seules quelques études ont rapporté le suivi en ligne de la concentration des anticorps produits par les cellules. De plus, à notre connaissance, aucun article n'a montré le suivi en ligne de la qualité des anticorps, et de leur glycosylation en particulier. Par ailleurs, il n'existe pas d'étude sur le suivi en ligne par spectroscopie Raman de certains paramètres de la physiologie cellulaire, tels que les vitesses spécifiques et les rendements métaboliques. De même, l'intérêt de l'utilisation conjointe de plusieurs spectroscopies en ligne n'a pas été envisagé. Enfin, la comparaison en parallèle des performances des spectroscopies NIR et Raman pour le suivi d'un même procédé de culture cellulaire n'est pas disponible.

2.5. Méthodes chimiométriques pour le traitement des données spectroscopiques

Les mesures spectroscopiques génèrent une quantité importante de données spectrales, en particulier lorsqu'elles sont utilisées pour le suivi en temps réel de procédés avec une fréquence élevée d'acquisition des données. Ces données spectrales doivent pouvoir être corrélées, par des modèles mathématiques, à certaines variables du procédé, telles que la concentration des composés du milieu de culture ou l'état des cellules. Une fois établis, ces modèles pourront alors être utilisés pour prédire les variables concernées à partir de nouveaux spectres acquis en ligne. Cependant, les spectres contiennent souvent une information redondante, car un seul composé peut avoir une réponse dans plusieurs régions spectrales, ce qui rend les données hautement corrélées, et chaque pic ne peut donc pas être associé directement à la concentration d'un composé cible. Les données spectrales sont ainsi inexploitables en l'état sans l'établissement de modèles de calibration multivariés. Ceci requiert d'utiliser des méthodes chimiométriques afin d'établir des relations entre les spectres acquis en ligne et les concentrations des composés à mesurer (Lourenço *et al.*, 2012a).

La chimiométrie peut être définie comme une discipline de l'analyse chimique mettant en jeu l'utilisation de méthodes mathématiques, statistiques et logiques pour , (i) concevoir ou sélectionner les approches expérimentales optimales ; (ii) extraire un maximum d'informations chimiques pertinentes en analysant les données expérimentales ; et (iii) mieux connaître et comprendre le système étudié (Vandeginste et Massart, 1997).

Classiquement, l'approche méthodologique globale adoptée pour le traitement quantitatif des données multivariées issues des procédés biotechnologique est la suivante : dans un premier temps, les données spectrales sont prétraitées pour optimiser le rapport signal sur bruit, et réduire leur dimension en décrivant leur variabilité par un petit nombre de variables non corrélées (LVs, variables latentes) contenant les informations pertinentes pour la calibration des modèles (Teixeira *et al.*, 2009b). Une fois cette dimension réduite, la deuxième étape consiste à établir un modèle de calibration entre ces variables latentes et les valeurs des concentrations des composés mesurés hors ligne avec des méthodes de référence. Le modèle construit est ensuite optimisé par la minimisation des erreurs de calibration et de validation en faisant varier les méthodes de prétraitement et la zone spectrale sélectionnée. Après avoir été validé, le modèle est ultérieurement appliqué sur des spectres d'échantillons non utilisés pour la calibration, en analyse en temps réel, pour prédire la concentration inconnue des composés.

2.5.1. Prétraitement des données spectrales

L'objectif du prétraitement des spectres est d'éliminer les variations aléatoires (bruits de fond), les variations liées à l'instrumentation, ou les interférences physico-chimiques, afin de renforcer l'information pertinente dans l'ensemble des données (Lourenço *et al.*, 2012a). Les méthodes de prétraitement les plus couramment utilisées sont des méthodes de normalisation et de dérivation (Rinnan *et al.*, 2009).

Les méthodes de normalisation sont utilisées, en général, pour réduire la variabilité physique entre échantillons, liée aux phénomènes de diffusion, les informations qualitatives qui distinguent un échantillon d'un autre étant conservées. La normalisation de spectres permet ainsi de corriger, à la fois les effets additifs (offset), et les effets multiplicatifs (déplacements verticaux de la ligne de base en fonction de la longueur d'onde). Les méthodes de normalisation les plus couramment utilisées sont la correction de dispersion multiplicative (Multiplicative Scatter Correction - MSC), et la variation normale standard (Standard Normal Variate - SNV) (Rinnan *et al.*, 2009). La méthode MSC corrige chaque spectre dans un ensemble de données par le spectre moyen du lot. Les paramètres de correction a_i et b_i sont estimés à partir de l'équation suivante entre le spectre à corriger, le vecteur x_i, et le spectre de référence, le vecteur x_{ref} :

$$x_i = a_i + b_i x_{ref} + e_i \qquad \textbf{Eq 7}$$

e_i : partie d'erreur non modélisée.

Les spectres sont ensuite corrigés par l'équation suivante :

$$x_{corr,MSC} = \frac{x_i - a_i}{b_i} = x_{ref} + \frac{e_i}{b_i} \qquad \textbf{Eq 8}$$

La méthode SNV n'utilise pas de spectre de référence, les spectres étant corrigés par l'équation :

$$x_{corr,SNV} = \frac{x_i - \mu}{\sigma} \qquad \textbf{Eq 9}$$

μ : moyenne de l'ensemble du lot de spectres à prétraiter
σ : écart-type de l'ensemble du lot de spectres à prétraiter

La Figure 24 illustre l'effet d'un prétraitement de spectres NIR par la méthode de normalisation SNV.

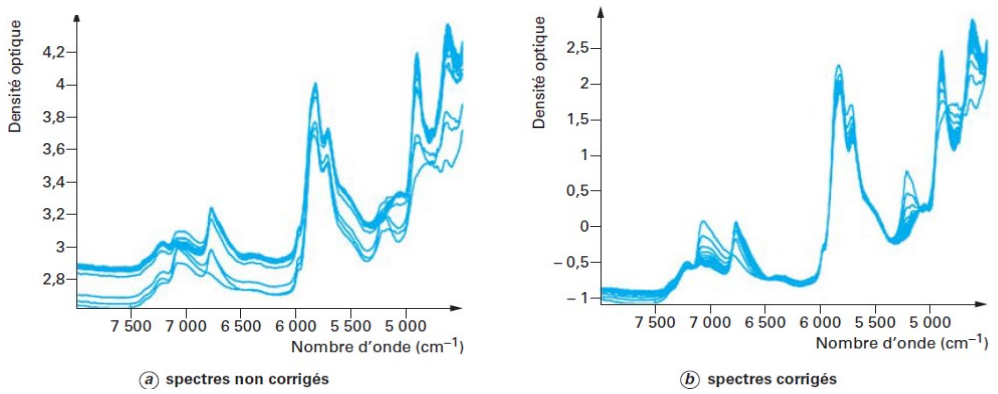


Figure 24 : Exemple de prétraitement de spectres NIR par la méthode SNV (a. Spectres non-traités; b. Spectres après traitement) (Bertrand, 2005)

Quant aux méthodes de dérivation, elles sont couramment utilisées pour réduire les effets des variations incontrôlées de la ligne de base, et pour séparer les régions actives des spectres se recouvrant partiellement (Williams et Norris, 1987). La dérivée première permet de mettre en avant les largeurs de bandes, alors que la dérivée seconde permet d'indiquer les positions des bandes (Figure 25).

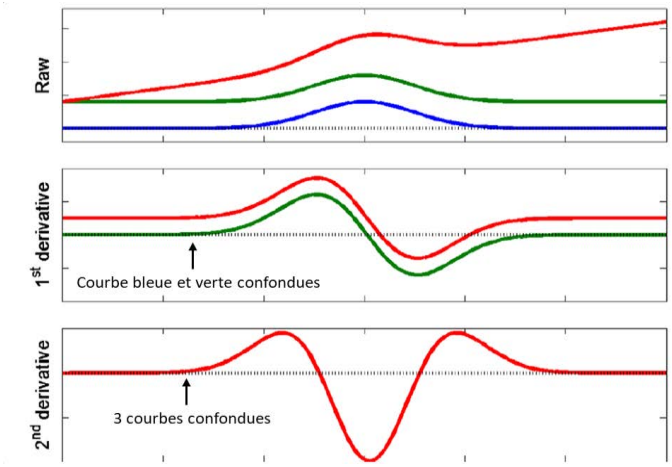


Figure 25 : Exemple de dérivation et de sa correction sur les effets additifs (courbe verte) et multiplicatifs (courbe rouge) (Rinnan et al., 2009)

La méthode la plus simple de dérivation consiste, pour obtenir la dérivée première, à calculer les différences entre des points adjacents du spectre, et, pour déterminer la dérivée seconde, à calculer les différences entre les points adjacents de la dérivée première, en utilisant les équations suivantes :

$$\frac{dx_i}{d\lambda} \approx \frac{\Delta x_i}{\Delta \lambda_i} = \frac{x_i - x_{i+1}}{\lambda_i - \lambda_{i+1}} \qquad \textbf{Eq 10}$$

$$\frac{d^2 x_i}{d\lambda^2} = \frac{x_{i-1} - 2x_i + x_{i+1}}{\Delta \lambda^2} \qquad \textbf{Eq 11}$$

λ : longueur d'onde (cm)

Cependant, cette approche génère souvent un niveau de bruit important. C'est pourquoi la méthode de dérivation par l'algorithme de Savizky-Golay a été proposée. Il calcule la dérivée de degré *n* au point i en appliquant un polynôme de degré *k* sur une fenêtre spectrale de *f* points ($f \geq k+1$) centrée autour de la variable i. La fenêtre f balaye ensuite la totalité du spectres afin de générer le spectre dérivé (Savitzky and Golay, 1964). Ces 3 paramètres *n*, *k*, *f* sont propres à la technique spectroscopique considérée, et doivent donc être optimisés pour chaque modèle développé.

Les méthodes de mise à l'échelle (scaling) et de centrage (centering) peuvent également s'appliquer à l'ensemble des données à traiter. La mise à l'échelle applique une division en chaque point des spectres par une valeur donnée, souvent l'écart-type. Cette technique est rarement utilisée en spectroscopie, car elle surévalue les zones de bruits, dans lesquelles les signaux spectroscopiques sont faibles, par rapport aux zones contenant les bandes spectrales d'intérêt. La méthode de centrage permet de centrer les données spectrales autour d'un spectre de référence. En général, le spectre de référence est le spectre moyen de l'ensemble des spectres à traiter. Cette technique est alors appelée le centrage par la moyenne (Mean Centering). Les spectres sont corrigés en soustrayant le spectre de référence de chaque spectre mesuré. Ce prétraitement permet de réduire l'influence spectrale du bruit afin de mettre en avant les informations spectrales les plus pertinentes (James, 2016).

D'autres méthodes de prétraitement peuvent également être appliquées telles que la sélection de zones spectrales pour analyser des variables spécifiques, la suppression des raies cosmiques et la soustraction du fond de fluorescence pour les spectres Raman, la soustraction de l'effet tampon pour les spectres IR, etc.

En fonction de l'application étudiée, le prétraitement des spectres intègre souvent séquentiellement une combinaison de plusieurs méthodes de prétraitement. Bien qu'il existe des études sur le choix optimal de ces méthodes de prétraitement (Bocklitz *et al.*, 2011; Lasch, 2012; Rinnan *et al.*, 2009), il reste encore difficile de les standardiser entre différents types de spectre. La plupart du temps, leur choix dépend de la nature des spectres, des observations, des problèmes à résoudre et de l'expérience antérieure (Gautam *et al.*, 2015).

2.5.2. Méthodes chimiométriques pour l'analyse qualitative des données spectrales

Afin de réaliser l'analyse qualitative des données spectrales, les méthodes de classification ont pour objectif de trouver ou d'identifier les tendances des échantillons pour les rassembler en sous-

groupes. La classification peut être supervisée (la catégorie d'appartenance des échantillons est requise à l'avance), ou non supervisée (les échantillons sont classés sans connaissance préalable de structures des groupes). L'analyse en composantes principales (PCA) est une méthode largement utilisée pour les analyses qualitatives des spectres, et représente souvent la première étape de ces méthodes de classification. Elle permet de réduire la dimension de données multivariées de très haute dimension, bruitées et fortement inter-corrélées, et de filtrer les informations redondantes. Les données sont projetées dans un espace de dimension inférieure pour créer des nouvelles variables contenant la plus grande partie de la variance des données d'origine. Ces nouvelles variables, appelées les composantes principales (PCs) sont des combinaisons linéaires des variables d'origine qui reflètent la structure inhérente de ces données d'origine (Camacho *et al.*, 2010). Ainsi, la première PC représente la variance maximale des données d'origine tandis que chaque PC successive prend en compte la plus grande partie possible de la variabilité restante.

Dans le cas de données spectrales, la PCA permet d'identifier les principales sources de variance dans une collection de spectres en définissant une série de PC et en attribuant à chaque spectre un score basé sur sa contribution relative à chaque PC. Les sources de variance peuvent aider à visualiser les principales tendances des données. De cette manière, les données multidimensionnelles complexes peuvent être représentées plus facilement par un espace bi ou tridimensionnel (score plots), décrivant les informations principales présentes dans les données (Meyers, 2012). Par exemple, la différence des spectres issus des 3 classes d'échantillons peut être mise en évidence par le « score plot », en projetant ces spectres sur différentes axes des PC (Figure 26).

Le clustering est une méthode de classification non supervisée pour détecter les groupements naturels dans les données, de telle manière que le degré d'association entre deux objets soit maximal s'ils appartiennent au même cluster. Le degré d'association est en général analysé par un dendrogramme, aussi appelé diagramme d'arbre (Lourenço *et al.*, 2012b) (Figure 27).

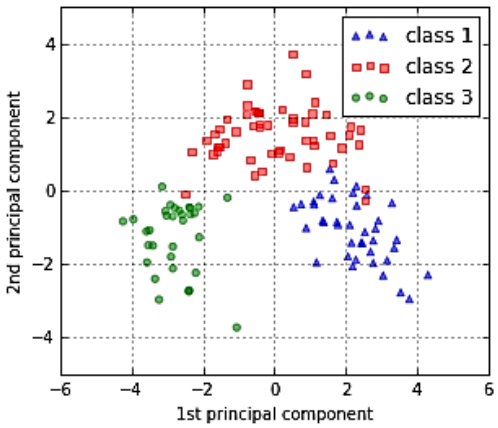


Figure 26 : Exemple d'un score plot d'une analyse PCA

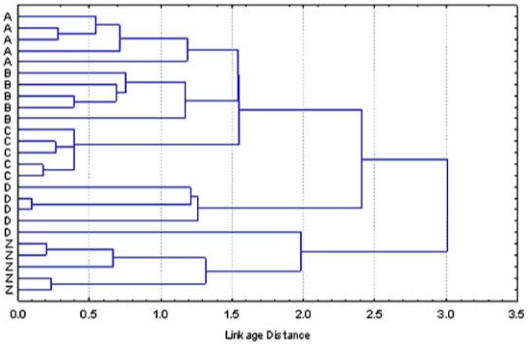


Figure 27 : Exemple d'un diagramme d'arbre

Pour la classification supervisée, les méthodes peuvent être basées sur la corrélation (LDA : Linear Discriminant Analysis), ou sur la distance (KNN : K-Nearest Neighbors). Les méthodes de modélisation, telles que SIMCA (Soft Independent Modeling of Class Analogy) et PLS-DA (Partial Least-Squares Discriminant Analysis), sont aussi couramment utilisées. Les méthodes de classification non linéaires, telles que ANN (Artificial Neural Networks), sont adaptées aux données spectrales non linéaires (Adams, 2007).

2.5.3. Méthodes chimiométriques pour l'analyse quantitative des données spectrales

L'analyse quantitative des données spectrales représenter un défi non négligeable, en particulier dans le cas de procédés de culture utilisant des milieux complexes contenant souvent plus de 50 composés. Ces nombreux composés peuvent générer des réponses dans des régions spectrales similaires, ce qui rend alors problématique l'identification et la quantification des composés individuels au sein d'un mélange complexe. Il devient donc nécessaire de faire appel à des méthodes

de régression multivariées pour établir des relations entre les données spectrales et les composés à quantifier.

La méthode de régression des moindres carrés partiels (PLS, Partial Least Squares) est la méthode de calibration la plus couramment utilisée dans le domaine de l'agro-alimentaire et des bioprocédés en général (Marison *et al.*, 2012). Elle consiste à générer un petit nombre de variables latentes (LVs) pour expliquer le maximum de variabilité dans un ensemble de spectres (*X*), tout en maximisant la corrélation entre les variables du bioprocédé qu'on souhaite prédire (*y*) et les données d'entrée des spectres, à travers un modèle de prédiction linéaire tel que :

$$y = \beta X + e \qquad \textbf{Eq 12}$$

β : matrice des coefficients de régression du modèle PLS
e : résidus

Pour la calibration des modèles PLS dans le cas d'un procédé de culture, il est indispensable de disposer de données représentatives du procédé considéré, qui comprennent, à la fois, les données spectrales (*X*) acquises en ligne et les valeurs des variables (*y*) mesurées hors ligne par des méthodes de référence. Les données doivent être réparties sur toute la durée du procédé pour un objectif de capturer le plus possible de variabilité du procédé et doivent permettre de représenter la variance, à la fois, au sein d'une même culture et entre différentes cultures successives. Ces deux variabilités doivent être considérées pour la calibration, afin d'obtenir des modèles de calibration PLS présentant ensuite une qualité de prédiction satisfaisante (Biechele *et al.*, 2015). Par la suite, les variables *y* pourront alors être prédites directement à partir de nouveaux spectres non reliés à des valeurs de référence. Cependant, quand une relation non-linéaire prononcée existe entre les données spectrales et les variables cibles, des techniques non-linéaires, telles que les réseaux de neurones artificiels (ANN) ou la méthode Support Vector Regression (SVR), doivent être utilisées (Roggo *et al.*, 2007).

Le nombre optimal de LVs à utiliser pour les modèles de calibration doit être recherché. D'un côté, pour un nombre de LVs élevé, la spécificité du modèle à l'ensemble de données utilisé est grande. Par conséquent, la capacité du modèle à estimer les concentrations de composés correspondant à des spectres en dehors de cet ensemble de données est fortement réduite. Ce phénomène est appelé « overfitting ». D'un autre côté, les modèles de calibration avec de faibles LVs sont généralement plus robustes mais peuvent provoquer le phénomène d' « underfitting », présentant

un risque si une variation importante du procédé n'est pas prise en compte lors de l'étape de calibration. (Figure 28).

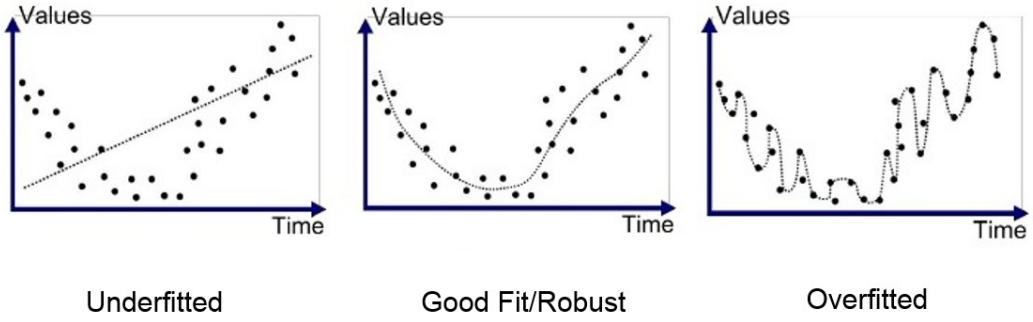


Figure 28 : Représentation schématique des phénomènes d'"underfitting" et d'"overfitting"

2.5.4. Evaluation des performances des modèles d'analyse quantitative des données spectrales

Les capacités prédictives d'un modèle de régression quantitatif sont généralement évaluées à l'aide des erreurs quadratiques moyennes (Root Mean Square Error - RMSE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \textbf{Eq 13}$$

y_i : valeur de référence mesurée
$\hat{y}_i$: valeur prédite par le modèle
n : nombre d'observations

Plusieurs types d'erreurs peuvent être ainsi calculés : l'erreur de calibration (RMSEC), l'erreur de prédiction (RMSEP), et l'erreur de validation croisée (RMSECV).

L'erreur de calibration RMSEC calcule l'erreur entre les valeurs prédites à partir de toutes les données utilisées pour la calibration du modèle PLS et les valeurs de référence correspondantes. Dans ce cas, les données spectrales étant déjà utilisées pour calibrer les modèles, cette erreur ne représente pas la réelle capacité de prédiction du modèle. Elle est plutôt utile pour caractériser la qualité de calibration du modèle.

L'erreur de prédiction RMSEP est calculée à partir de données, pseudo-internes ou externes, non utilisées pour développer le modèle de calibration. Dans le cas de la prédiction pseudo-interne, l'ensemble des données disponibles est séparé en deux parties, l'une pour la calibration et l'autre

pour la prédiction. Dans le cas de la prédiction externe, des données spectrales complètement indépendantes de celles utilisées pour la calibration, issues de nouvelles expériences, seront utilisées. Ainsi, le paramètre RMSEP permet d'évaluer le potentiel du modèle établi avec les données de calibration à prédire les futurs variables à partir de nouvelles données spectrales (Consonni *et al.*, 2010).

L'erreur de validation croisée (RMSECV) est calculée à l'aide de toutes les données utilisées pour la calibration, mais en extrayant successivement un certain nombre de ces données. Un sous-modèle est alors développé avec les données restantes, et validé par les données extraites. Cette opération est répétée plusieurs fois en faisant varier les données extraites. L'erreur RMSECV est ainsi obtenue en moyennant les erreurs déterminées pour chaque segment de données extraites. Cette méthode est particulièrement avantageuse quand il existe peu de données, car elle ne requiert pas de données supplémentaires. Elle est aussi souvent utilisée pour déterminer le nombre optimal de LVs. Le nombre minimal de LVs permettant une valeur minimum de RMSECV est retenu afin d'éviter les phénomènes d'« overfitting » et d'« underfitting » (Figure 29) (Cervera *et al.*, 2009).

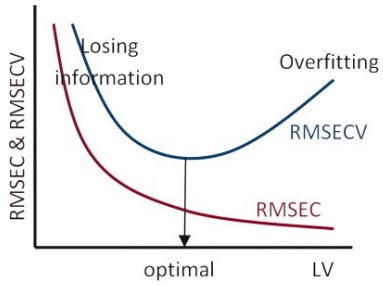


Figure 29 : Optimisation du nombre des LVs en minimisant la valeur de RMSECV (Cervera et al., 2009)

Il existe plusieurs stratégies de validation croisée, telles que les méthodes "leave-one-out", "k-fold", "venetian blinds" et la sélection aléatoire (Kuligowski *et al.*, 2016). La méthode "k-fold" est la plus utilisée pour un jeu de données relativement petit. *N/k* données sont extraites d'un ensemble de *N* données, ces données extraites étant ensuite prédites par le modèle établi avec les données restantes. Quand *k* est égal à *N*, la méthode est appelée "leave-one-out". La méthode "venetian blind" sépare les données en segments entrelacés, tous les $k^{ème}$ segments étant regroupés pour former le jeu de prédiction. Cette méthode est simple et pratique quand les échantillons sont répartis aléatoirement dans l'ensemble de données. La méthode de sélection aléatoire est, quant à elle, très utile pour des données contenant des répétitions, puisqu'elle crée une instabilité pour le modèle de calibration et augmente donc la fiabilité de l'évaluation des erreurs, surtout quand un grand nombre d'itérations est utilisé.

En outre, l'importance de la variable pour la projection (VIP, Variable Importance in Projection) est un autre paramètre important, permettant d'évaluer la contribution de chaque variable *j* à un modèle chimiométrique donné (PLS) selon la variance expliquée (Figure 30).

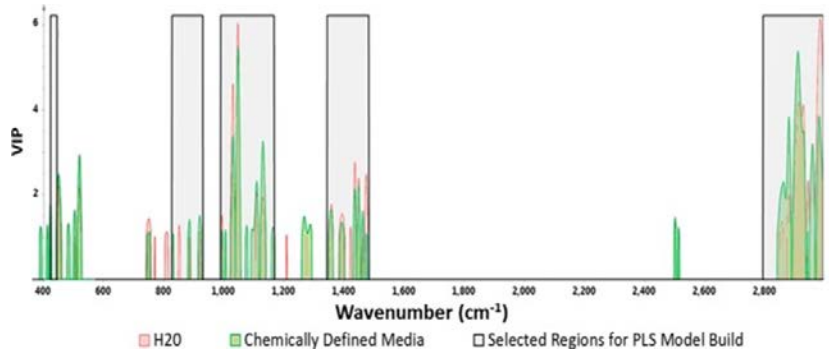


Figure 30 : Exemple des scores VIP d'un modèle de glucose dans un milieu de culture de cellules animales (Berry et al., 2016)

Les scores VIP peuvent être calculés à partir de l'équation suivante :

$$VIP_j = \sqrt{\frac{\sum_{f=1}^{F} w_{jf}^2 \times SSY_f \times J}{SSY_{total} \times F}} \qquad \textbf{Eq. 14}$$

w_{jf} : poids de la variable *j* et du composant *f*
SSY_f : somme des carrés de la variance expliquée par le $f^{ème}$ composant
SSY_{total} : somme des carrés de la variance expliquée totale
J : nombre de variables
F : nombre des composants

En général, les régions spectrales dans lesquelles les variables présentent des scores VIP supérieurs à 1 sont considérées comme ayant une contribution significative pour les modèles de prédiction. Ces régions spectrales identifiées peuvent être ensuite comparées avec les régions caractéristiques des molécules à analyser afin d'évaluer la fiabilité du modèle développé (André *et al.*, 2015).

2.6. Intérêt des capteurs logiciels

Le terme « capteur logiciel » est utilisé pour décrire l'utilisation d'une combinaison de différents capteurs et de modèles mathématiques en vue de prédire des variables difficilement accessibles par mesure directe (Figure 31) (Mandenius et Gustavsson, 2015).

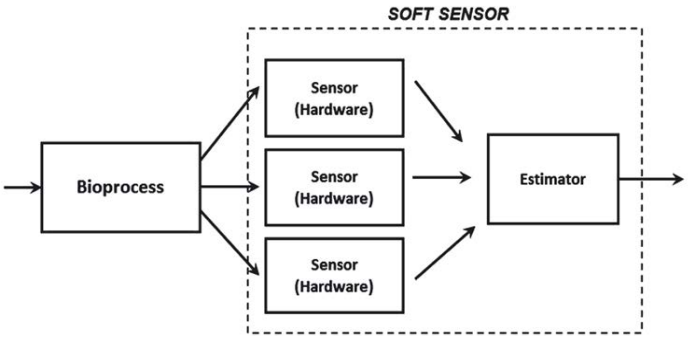


Figure 31 : Représentation schématique d'un capteur logiciel (Mandenius et Gustavsson, 2015)

D'un point de vue général, on distingue deux types de capteurs logiciels qui sont basés, soit sur les modèles (model-driven), soit sur les données (data-driven). Les capteurs logiciels "data-driven" utilisent des méthodes de régression simple ou des méthodes chimiométriques multivariées pour décrire les relations entre les données mesurées en ligne et les variables du procédé à prédire. Par exemple, pour les procédés de culture de cellules animales, les spectroscopies NIR, Raman ou diélectrique, décrites dans les parties précédentes (2.2-2.4), permettent de prédire, à l'aide de différentes méthodes chimiométriques, les concentrations des cellules, du glucose, de la glutamine, du lactate, des ions ammoniums, des protéines recombinantes produites, etc. Ces capteurs logiciels "data-driven" sont aussi appelés modèles boîte noire, puisque les modèles n'utilisent aucune connaissance préalable du procédé, en étant uniquement basés sur les observations empiriques du procédé. De tels modèles présentent une bonne précision de prédiction pour des conditions utilisées proches de celles de la calibration. Par contre, cette précision de prédiction diminue quand les variables s'éloignent de ces conditions. C'est pourquoi, les capteurs logiciels de type boîte noire manquent souvent de bonnes capacités d'extrapolation. Par conséquent, un grand nombre de données est nécessaire pour développer des modèles robustes (Ohadi *et al.*, 2015).

Les capteurs logiciels "model-driven", aussi appelés modèles boîte blanche, utilisent au contraire des connaissances phénoménologiques du procédé, telles que les réactions métaboliques, les bilans de matière, les bilans énergétiques, les cinétiques enzymatiques, etc. Ces modèles sont basés sur les connaissances a priori des relations entre les variables d'état, que sont, par exemple, les cellules, les substrats, les inhibiteurs et les produits d'intérêt. Dans ce cas, les paramètres des modèles ont alors une signification chimique ou biologique qui rend ces modèles utiles pour mieux comprendre les phénomènes sous-jacents du procédé ainsi que pour son suivi et son contrôle (Shirsat *et al.*, 2015). Cependant, en dépit des avantages des capteurs logiciels de type boîte blanche pour représenter la dynamique du procédé, il reste souvent difficile de développer et de calibrer de tels modèles en raison, en particulier, de la complexité du métabolisme des cellules animales. Les connaissances sur le système métabolique des cellules ne sont pas toujours disponibles, ce qui ne permet pas d'écrire

précisément toutes les réactions intervenant en cours du procédé. Une autre limite des modèles boîte blanche réside dans le fait qu'ils utilisent souvent des lois théoriques simplifiées et éloignées des conditions réelles du procédé, ce qui peut entraîner des niveaux de bruit élevés sur leurs prédictions (Kadlec *et al.*, 2009).

Pour contourner les limites respectives des modèles boîte noire et boîte blanche des modèles hybrides semi-paramétriques peuvent être développés. Cette approche, souvent appelée boîte grise, intègre les données mesurées expérimentalement par les différents capteurs aux différentes connaissances phénoménologiques du procédé (Teixeira *et al.*, 2009a). Son principe de fonctionnement est représenté dans la Figure 32. A titre d'exemple, les données spectrales obtenues par les mesures en ligne sur le surnageant de culture, peuvent fournir, non seulement, des informations en temps réel sur des variables du procédé, mais également des informations au niveau intracellulaire permettant de mieux décrire les voies métaboliques régissant le système biologique. Ainsi, les données spectrales peuvent être couplées aux analyses des flux métaboliques des cellules afin de mieux décrire les cinétiques du système. Un exemple réussi d'application de ces modèles hybrides semi-paramétriques a été démontré par Teixeira *et al.* Dans cet exemple, un modèle hybride a été développé en combinant des modèles cinétiques phénoménologiques et des modèles de réseaux de neurones à partir des données hors ligne pour décrire un procédé de culture des cellules BHK productrices de la protéine de fusion IgG1-IL2. L'optimisation du procédé à l'aide du capteur logiciel développé, a alors permis une augmentation considérable de sa productivité grâce à une meilleure stratégie d'alimentation en glucose et en glutamine (Teixeira *et al.*, 2005).

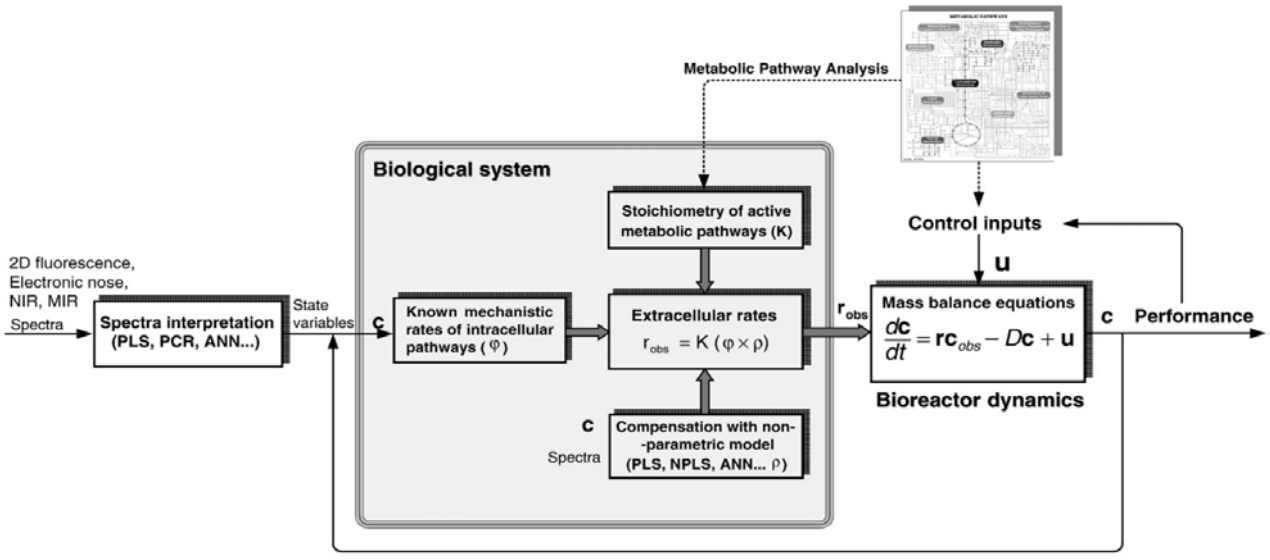


Figure 32 : Représentation schématique d'un capteur logiciel hybride (Teixeira et al., 2009a)

Les méthodes chimiométriques sont indispensables pour analyser les données spectrales, acquises en ligne par diverses méthodes spectroscopiques au cours des procédés de culture de cellules animales. Elles permettent de relier ces données spectrales à diverses variables du procédé, rendant ainsi possible la prédiction en temps réel des valeurs de ces variables. Cependant, l'analyse chimiométrique des données multivariées n'est souvent utilisée que pour traiter des données issues d'une seule méthode spectroscopique. Très peu d'études ont proposé de traiter les spectres combinés générés par différentes spectroscopies. Par ailleurs, le capteur logiciel de type boîte grise présente un potentiel intéressant pour le suivi et le contrôle en temps réel des bioprocédés. Or, à notre connaissance, aucune étude traitant des procédés de culture de cellules animales n'a combiné des données non paramétriques, obtenues par mesures spectroscopiques en ligne, à des modèles paramétriques basés sur les bilans de matière.

Objectifs des travaux

Objectifs de la thèse et démarche proposée pour les travaux

Le bilan de cette étude bibliographique met en exergue le fait que les procédés de culture de cellules animales producteurs d'anticorps thérapeutiques, sont complexes et présentent de nombreuses spécificités. En particulier, pour pouvoir assurer les performances souhaitées de quantité et de qualité du produit, de nombreux paramètres restent à maîtriser et à contrôler, qu'ils soient physiques, chimiques, et surtout biologiques. Or, en dépit de la stratégie PAT proposée il y a une quinzaine d'années pour le suivi en temps réel de ces procédés, seuls quelques paramètres sont actuellement contrôlés en ligne de manière quasi systématique, comme le pH, l'O_2 dissous et la température. Par ailleurs, le potentiel de méthodes spectroscopiques *in situ* (NIR, Raman, diélectrique) est étudié depuis plus récemment pour le suivi de paramètres biochimiques, comme la concentration des cellules et des composés du surnageant de culture (éléments nutritifs, métabolites, anticorps), même si l'utilisation de ces outils tarde à s'imposer en routine.

Sur la base de ces constats, l'objectif général de cette thèse est de développer une approche méthodologique innovante pour le suivi en ligne de procédés de production d'anticorps (de type IgG) par cellules animales CHO, en étudiant les apports des techniques spectroscopiques *in situ* permettant de contribuer à la stratégie PAT.

Dans ce cadre, certaines des questions scientifiques identifiées à l'issue de l'étude bibliographique sont formulées ci-dessous :

- Comment identifier la technologie spectroscopique la plus pertinente, NIR ou Raman, permettant de mieux caractériser en temps réel les concentrations des principaux composés d'une culture de cellules animales ? Quels indicateurs de performance peuvent être envisagés pour répondre à cette question ?
- Compte-tenu de la spécificité de chaque méthode spectroscopique, leur couplage *in situ* peut-il permettre d'améliorer les prédictions de ces différents paramètres de la culture en temps réel ? Quelles méthodes chimiométriques sont alors les plus performantes pour fusionner les diverses données spectrales ?
- La spectroscopie diélectrique *in situ*, classiquement utilisée pour le simple suivi de la concentration de cellules viables, est-elle susceptible d'apporter des informations complémentaires sur l'état physiologique des cellules, comme leur vitesse spécifique de croissance ? Comment peut-on alors prendre en compte les difficultés liées aux calculs en temps réel ? Quels sont les intérêts de suivi en ligne de cette vitesse spécifique de croissance cellulaire dans les procédés de production des IgG ?

- Peut-on envisager de suivre la qualité des anticorps produits en temps réel pour identifier le moment optimal de leur récolte ? Les spectroscopies NIR ou Raman peuvent-elles permettre de différencier, dans un milieu de culture complexe, des molécules de structure très semblables, telles que des IgG glycosylées ou non ? De manière plus fine, les très faibles modifications des motifs de glycosylation des IgG peuvent-elles être discriminées en temps réel ? Le mode de mise en œuvre de la culture influence-t-il la qualité de ces IgG ?
- Peut-on envisager un pilotage automatisé en temps réel d'une culture en mode semi-continu, en utilisant des mesures spectroscopiques *in situ* ? Quels peuvent être les effets de ce pilotage automatisé sur la quantité et/ou la qualité des IgG produites ?

Pour répondre à l'ensemble de ces questions, les démarches méthodologiques et expérimentales développées au cours de la thèse ont été organisées en quatre chapitres selon le plan présenté dans la Figure 33.

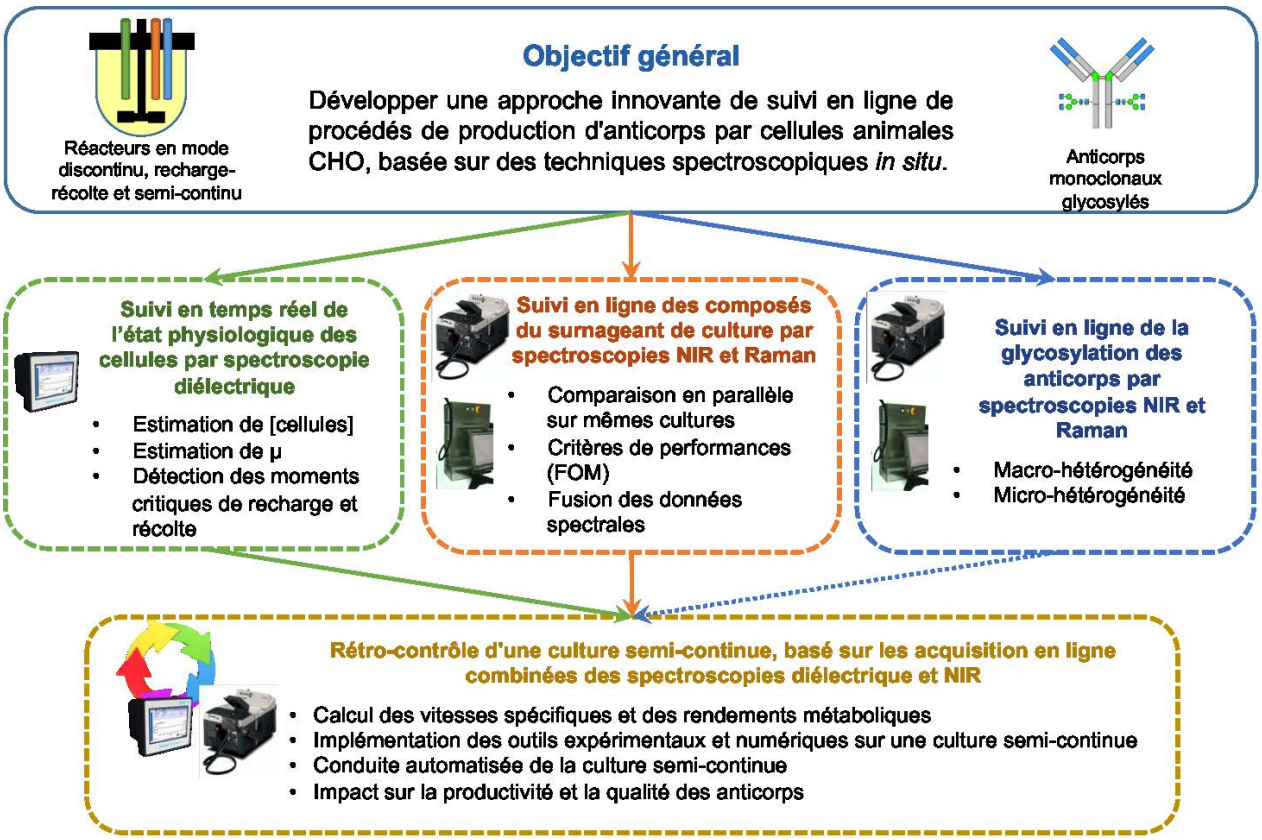


Figure 33 : Démarche globale des travaux de la thèse

- Le premier chapitre présente le développement d'un modèle d'estimation en temps réel de la vitesse spécifique de croissance cellulaire (μ) au cours de cultures réalisées en mode discontinu et recharge-récolte. Pour ce faire, la mesure en ligne de la permittivité par spectroscopie

diélectrique permet d'estimer la concentration des cellules vivantes, puis d'effectuer le calcul de μ en temps réel, moyennant des précautions de calcul. Ce chapitre présente aussi l'impact du suivi en ligne de μ sur le pourcentage de glycosylation des IgG produites au cours de cultures en mode recharge-récolte.

- Dans un second chapitre, une étude comparative des performances des spectroscopies Raman et NIR est proposée pour la prédiction en ligne des différentes concentrations (cellules viables, glucose, lactate, glutamine, ions ammonium, IgG) au cours des mêmes cultures de cellules CHO. De plus, des essais de combinaison des deux types de spectroscopies sont réalisés pour évaluer leur impact sur la performance des modèles de prédiction.
- Le troisième chapitre est consacré à l'évaluation de la capacité des spectroscopies Raman et NIR pour le suivi en ligne de la glycosylation des IgG produites. Plus précisément, des modèles sont développés pour prédire en temps réel, à la fois, la macro-hétérogénéité (IgG glycosylées et non-glycosylées) et la micro-hétérogénéité (IgG fucosylées, galactosylées, sialylées et hautement mannosylées) de la glycosylation des produits au cours des procédés de culture.
- Enfin, le dernier chapitre propose l'utilisation conjointe de la spectroscopie diélectrique et de la spectroscopie NIR pour le développement d'un capteur logiciel, permettant le suivi en temps réel de μ et du rendement des cellules formées par glucose consommé. Ce capteur logiciel est alors mis en œuvre au cours d'une culture en mode semi-continu, pour contrôler de manière automatique le débit d'alimentation du milieu de culture. Les performances de productivité et de qualité des anticorps sont enfin comparées à celles de cultures pilotées manuellement.

Matériel & Méthodes

1. Culture cellulaire 76

1.1. Lignée cellulaire 76

1.2. Milieu de culture 76

1.3. Conservation de la lignée cellulaire et propagation en pré-culture 77

1.3.1. Constitution des banques de cellules 77

1.3.2. Congélation des cellules 77

1.3.3. Décongélation des cellules 77

1.3.4. Propagation des cellules pour les pré-cultures 77

1.4. Mise en œuvre des cultures 78

1.4.1. Systèmes de culture 78

1.4.2. Échantillonnage 79

1.4.3. Modes d'alimentation des cultures 79

2. Analyses hors-ligne des cellules, substrats et produits 80

2.1. Cellules 80

2.2. Composés du milieu de culture 81

2.3. Anticorps 82

2.3.1. Analyses préliminaires des anticorps 82

2.3.2. Quantification des anticorps 83

2.4. Calcul des paramètres cinétiques des cultures 84

2.4.1. Vitesse spécifique de croissance 85

2.4.2. Vitesses spécifiques de consommation des substrats et de formation des produits 85

2.4.3. Rendements métaboliques 87

2.4.4. Productivité 87

3. Analyses spectroscopiques en ligne 87

3.1. Spectroscopie diélectrique 88

3.2. Spectroscopie proche Infrarouge (NIR) 88

3.3. Spectroscopie Raman 90

4. Méthodes chimiométriques pour le traitement des données spectrales 90

4.1. Répartitions des données spectrales entre calibration et prédiction 90

4.2. Modèles de prédiction basés sur la spectroscopie diélectrique 92

4.3. Modèles de prédiction basés sur les spectroscopies NIR et Raman 93

4.4. Evaluation des performances des modèles de calibration et prédiction 93

1. Culture cellulaire

1.1. Lignée cellulaire

La lignée cellulaire CHO M250-9 productrice de l'anticorps anti-RhD (IgG) a été développée en 2008 par des chercheurs du *Bioprocessing Technology Institute* de Singapour (Chusainow et al., 2009). Elle résulte de la transfection en électroporation de la lignée CHO-DG44 déficiente en dihydrofolate réductase (*dhfr* $^{-}$*)* par deux vecteurs d'expression (phCMV-VHRhD-λ1C-neo et phCMV-VLRhD-K_R-neo) qui contiennent les gènes codant pour l'anticorps anti-rhésus D humain et la *dhfr* (Urlaub et al., 1983). Les clones présentant des transfectants stables ont été sélectionnés en culture en présence de 600 µg.mL^{-1} de généticine (G418, Sigma-Aldrich). L'amplification des gènes a été induite par l'addition de méthotrexate suivant la méthode de Page et Sydenham (Page et Sydenham, 1991). La lignée CHO M250-9 a été ensuite adaptée à la culture en suspension et en milieu sans sérum.

1.2. Milieu de culture

Le milieu de culture utilisé au cours de ces travaux est un mélange (1:1 vol/vol) de deux milieux commerciaux sans sérum et sans protéines : le milieu CD-CHO (Thermo Fisher) et le milieu PF-CHO (GE Healthcare Life Sciences). Le milieu CD-CHO est un milieu chimiquement défini et optimisé pour la croissance des cellules CHO et pour l'expression de protéines recombinantes dans des cultures en suspension. Il ne contient ni protéines ni composants peptidiques d'origine animale, végétale ou synthétique, ni lysats ou hydrolysats non définis. Le milieu PF-CHO est un milieu sans protéine contenant des hydrolysats de soja. Ce milieu a été développé pour supporter la croissance de multiples clones de cellules CHO *dhfr*$^{-}$. Leur mélange a été supplémenté par 4 mM de L-glutamine (Sigma-Aldrich) et 0,05% de Pluronic F68 (Sigma-Aldrich). Dans certaines cultures en fiole, 50 nM de méthotrexate (Sigma-Aldrich) et 600 µg.mL^{-1} de généticine (Sigma-Aldrich) ont été également ajoutés pour maintenir la pression de sélection. Dans les cultures en bioréacteur, une solution antibiotique antimycotique (Sigma-Aldrich) a été utilisée à 1% (vol/vol).

1.3. Conservation de la lignée cellulaire et propagation en pré-culture

1.3.1. Constitution des banques de cellules

Dans un premier temps, une banque primaire de 6 cryules a été réalisée à partir d'une cryule d'origine après une semaine de culture comportant 2 passages en fiole d'Erlenmeyer. Puis, l'une des cryules de cette banque primaire a été remise en culture, et les cellules ont été cultivées pendant 8 jours, avec 2 passages, pour établir une banque de travail constituée de 36 cryules de 1 mL.

1.3.2. Congélation des cellules

Les cellules en phase de croissance exponentielle ont été resuspendues dans 1 mL de milieu de culture (CD-CHO/PF-CHO) contenant 10% de diméthylsulfoxyde (DMSO, Sigma-Aldrich) et 10% de sérum de veau fœtal (SVF, Sigma-Aldrich), à une concentration de 1.10^7 à 2.10^7 cellules.mL^{-1}. Les cellules ont été ensuite congelées dans des tubes cryogéniques (cryules, Corning) à l'aide d'un système de congélation (Mr. Frosty™, Thermo Scientific) contenant de l'isopropanol, dans le but d'atteindre une vitesse de refroidissement optimale (-1 °C.min^{-1}). Après 24 h à -80 °C, les cryules ont été immergées et conservées dans l'azote liquide à -196 °C.

1.3.3. Décongélation des cellules

Les cryules sorties de l'azote liquide ont été immédiatement réchauffées à 37 °C à l'aide d'un bloc chauffant pour limiter le temps de contact des cellules avec le DMSO. Une fois décongelées, les cellules ont été resuspendues dans 10 mL de milieu de culture frais préalablement thermostaté à 37 °C. Après 5 min de centrifugation à 900 rpm (Centrifugeuse Jouan G 4.11), le surnageant a été éliminé et le culot cellulaire remis en suspension dans 30 mL de milieu de culture.

1.3.4. Propagation des cellules pour les pré-cultures

Après décongélation, les cellules ont été cultivées en mode statique en fiole d'Erlenmeyer pendant 24 h, dans un incubateur (Kühner, Adolf Kühner AG) régulé à 37 °C, 5% de CO_2 et 80% d'humidité. Les bouchons des fioles étaient munis de filtres de 0,2 µm, permettant les échanges gazeux tout en gardant le milieu stérile. Au cours des 48 h suivantes, la vitesse d'agitation orbitale a été progressivement augmentée jusqu'à 70 rpm. Les cellules ont été cultivées pendant 10 jours avec 4 passages d'une durée de 2 à 3 jours chacun. Avant chaque étude cinétique, les cellules ont été ensemencées à 3.10^5 cellules.mL^{-1} et cultivées pendant 3 jours avant d'être transférées dans le récipient de l'étude, ceci afin de s'assurer que les cellules étaient bien en phase exponentielle de croissance à chaque début de culture.

1.4. Mise en œuvre des cultures

1.4.1. Systèmes de culture

- Fiole d'Erlenmeyer

Les cultures en fioles d'Erlenmeyer (Fisher Scientific) ont été réalisées dans l'incubateur Kühner muni d'une table à agitation orbitale (50 cm de diamètre d'agitation). L'incubateur est régulé à 37 °C, 5% de CO_2 et 80% d'humidité et 70 rpm pour la vitesse d'agitation. Les volumes des fioles utilisées étaient de 250, 500, ou 1000 mL, correspondant respectivement à 50, 100, et 300 mL de volume de milieu. La densité d'ensemencement était de 3.10^5 cellules.mL^{-1}, avec des cellules en phase exponentielle de croissance.

- Bioréacteur

Les cellules ont été aussi cultivées dans un bioréacteur de 2,5 L (Tryton, Pierre Guérin) (Figure 34).

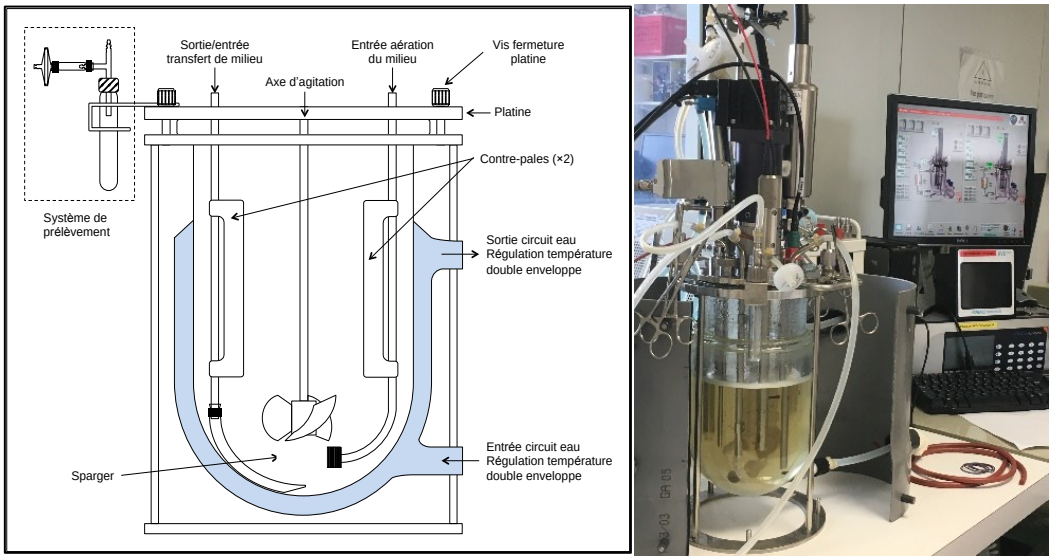


Figure 34 : Représentation schématique (gauche) et photographie du bioréacteur Tryton de 2,5 L

La culture en bioréacteur était régulée en O_2 dissous (50% de la saturation en air), température (37 °C) et pH (7,2). L'agitation était assurée par un mobile à pales inclinées (oreille d'éléphant), et la vitesse maintenue à 90 rpm. L'aération était réalisée par injection d'oxygène pur dans la phase liquide du réacteur à l'aide d'un sparger de type anneau. Le niveau d'oxygène dissous était mesuré par une sonde à oxygène optique (VisiFerm™ DO, Hamilton). La température était régulée par circulation d'eau chaude dans la double enveloppe du bioréacteur, à partir de la mesure d'une sonde de température. Le pH était contrôlé à partir de la mesure par une sonde à pH stérilisable (Hamilton),

soit par ajout de NaOH 0,5 M à l'aide d'une pompe, soit par injection de CO_2 dans le ciel du bioréacteur à l'aide d'une électrovanne. D'autres capteurs moins conventionnels ont été également utilisés pour le suivi des cultures en bioréacteur : la spectroscopie diélectrique (Evo200 iBiomass system, Hamilton), la spectroscopie Raman (BioViserion®, Indatech) et la spectroscopie NIR (Antaris II, Thermo Scientific). Ces équipements seront décrits par la suite.

Avant l'ensemencement des cellules, le bioréacteur a été stérilisé par autoclavage avec les sondes plongeant dans 1 L de tampon PBS (phosphate-buffered saline). Le PBS a été ensuite remplacé le milieu stérile maintenu à 37 °C pendant une nuit pour vérifier la stérilité du système. Puis, les cellules en phase exponentielle de croissance ont été ensemencées dans le bioréacteur à une concentration de 3.10^5 cellules.mL^{-1}.

1.4.2. Échantillonnage

Afin de réaliser le suivi cinétique des cultures cellulaires, des prélèvements ont été effectués 1 à 4 fois par jour au cours des cultures en fonction des besoins. Après chaque prélèvement, l'échantillon prélevé a été utilisé d'abord pour la numération cellulaire. Puis, il a été centrifugé 5 min à 900 rpm (Centrifugeuse Jouan G4.11) pour séparer le surnageant de culture des cellules. Le surnageant a été récupéré pour l'analyse des composés du milieu de culture, une partie étant aliquotée et conservée à -20 °C pour l'analyse de la glycosylation des IgG en fin de culture.

1.4.3. Modes d'alimentation des cultures

- Discontinu (batch)

Les cellules en phase exponentielle de croissance ont été ensemencées à 3.10^5 cellules.mL^{-1} en début de culture. Pendant la culture, aucun ajout ni soutirage (hormis les échantillonnages) n'est effectué. La suspension cellulaire est récoltée en une seule fois en fin de la culture.

- Recharge-récolte (feed-harvest)

Les cellules en phase exponentielle de croissance ont été ensemencées à 3.10^5 cellules.mL^{-1}. Après une première phase de culture en mode discontinu, 2/3 de la suspension cellulaire a été soutiré et remplacé par un même volume de milieu frais préchauffé à 37 °C. Ce cycle a été répété 2 à 4 fois, aux moments définis pour l'étude.

- Discontinu-alimenté (fed-batch)

Ce mode de culture consiste à ajouter du milieu d'alimentation sous forme concentrée (Cell Boost Supplement 5, Ficher Scientific) dans le réacteur sans effectuer de soutirage. L'alimentation débutait par l'ajout de milieu concentré lorsque les cellules étaient en phase exponentielle de croissance, au bout de 3 jours de culture. Différentes stratégies d'alimentation (ponctuelles, exponentielles ou adaptatives), ont été ensuite réalisées en fonction des études souhaitées.

2. Analyses hors-ligne des cellules, substrats et produits

2.1. Cellules

Le compteur automatique de cellules Vi-CELL™ (Beckman Coulter) a été utilisé pour la numération cellulaire à l'aide de la coloration au bleu Trypan. Les cellules viables relarguent le colorant et restent incolores, tandis que les cellules mortes retiennent le bleu Trypan et sont colorées en bleu. Le compteur automatique ajoute le bleu Trypan dans l'échantillon à un ratio de 1 :1 (vol/vol). Après le mélange, l'échantillon est délivré sous forme de fines gouttelettes dans une chambre de numération munie d'une caméra. 50 photos sont prises et l'analyse des images est faite par un logiciel qui permet de distinguer les cellules vivantes et les cellules mortes selon les niveaux de gris (Figure 35). La concentration de cellules viables (VCD) et la viabilité cellulaire sont déterminées par la moyenne des résultats de ces 50 photos. La gamme de concentration cellulaire qui peut être mesurée par cet appareil est comprise entre 1.10^5 et 1.10^7 cellules.mL^{-1}. Une étape de concentration ou de dilution est nécessaire en dehors de cette gamme.

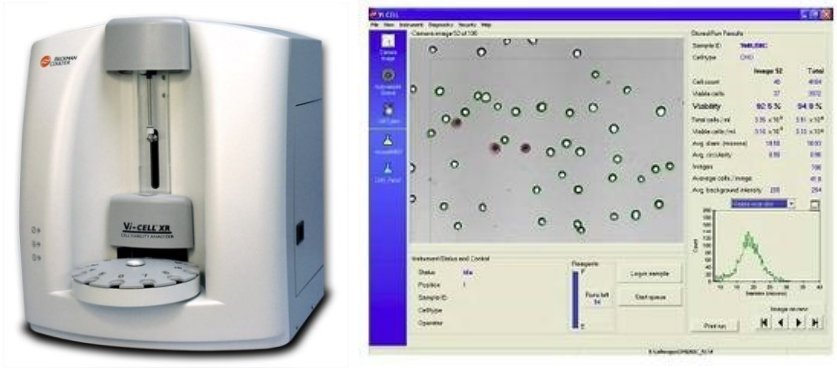


Figure 35 : Compteur de cellules Vi-CELL et logiciel de traitement d'images

2.2. Composés du milieu de culture

L'analyseur photométrique automatisé Gallery™ (Thermo scientific) a été utilisé afin de mesurer la concentration de glucose, lactate, glutamine, et ions ammoniumdans le milieu de culture à l'aide de divers kits enzymatiques (Figure 36).

Figure 36 : Analyseur multi-paramétrique Gallery

- Dosage du glucose

Le glucose a été analysé à l'aide du kit GOD-POD (Thermo Scientific), mettant en œuvre les réactions suivantes :

$$\beta\text{-}D\text{-}Glucose + H_2O + O_2 \xrightarrow{Glucose\ oxydase} D\text{-}Glucose + H_2O_2$$

$$2H_2O_2 + phénol + 4\text{-}amino\text{-}antipyrine \xrightarrow{Péroxydase} Quinonéimine + 4H_2O$$

L'apparition de la quinoéimine est détectée par spectrométrie à 510 et 620 nm après 10 min d'incubation.

- Dosage du lactate

Le lactate a été analysé avec le kit Acide L-Lactique (Thermo Scientific). Les réactions sont les suivantes :

$$L\text{-}lactate + NAD^+ \overset{L\text{-}LDH}{\longleftrightarrow} pyruvate + NADH + H^+$$

L'équilibre de la réaction est en faveur du lactate. L'élimination du pyruvate du milieu réactionnel déplace l'équilibre de la réaction dans le sens de la formation de pyruvate.

$$Pyruvate + L\text{-}glutamate \xleftrightarrow{Glutamate\text{-}pyruvate\text{-}transamidase} L\text{-}alanine + \alpha\text{-}cétoglutarate$$

La formation du NADH est mesurée à 340 nm après 10 min d'incubation.

- Dosage de la glutamine

La glutamine a été analysée à l'aide du kit Glutamine V2 Bio (Roche), mettant en œuvre les réactions suivantes :

$$Glutamine \xrightarrow{Glutaminase} Glutamate + NH_3$$

$$Glutamate \xrightarrow{L\text{-}glutamate\ oxidase} \alpha\text{-}cétoglutarate + NH_4^+ + H_2O_2$$

$$H_2O_2 + 4\text{-}AAP + TOOS \xrightarrow{Péroxidase} H_2O + Chromogène$$

La formation du composé chromogène est mesurée à 540 nm après 5 min d'incubation.

- Dosage des ions ammonium

Les ions ammonium ont été analysés par le kit NH3 Bio (Roche) selon le principe suivant :

$$NH_4^+ + 2\text{-}oxoglutarate + NADPH \xrightarrow{Glutamate\ déshydrogénase} L\text{-}glutamate + NADP^+ + H_2O$$

La disparition du NADPH, qui est proportionnelle à la concentration en ions ammonium, est mesurée à 340 nm après 5 min d'incubation.

2.3. Anticorps

2.3.1. Analyses préliminaires des anticorps

Dans un premier temps, la capacité de production de l'anticorps humain recombinant (IgG anti-rhésus D) par la lignée cellulaire CHO M250-9 a été caractérisée. Des analyses préliminaires ont été réalisées à l'aide de techniques d'électrophorèse de protéines avec divers surnageants de cultures prélevés au cours du temps.

- Analyse par la technique de SDS-PAGE

Les protéines présentes dans le surnageant de culture ont été séparées en fonction de leur taille sur un gel d'électrophorèse d'acrylamide en condition dénaturante. Elles ont été d'abord concentrées dans un gel d'acrylamide à 4%, puis séparées dans un gel d'acrylamide à 10%. Les échantillons de milieu de culture ont été dilués 3 fois dans la solution de dépôt contenant 10% de SDS (dodécylsulfate de sodium), 5% de β-mercaptoéthanol, du bleu de bromophénol et du glycérol, et

chauffés à 98 °C pendant 5 min. 15 µL ont été alors déposés dans chaque puit sur le gel d'électrophorèse. Par la suite, le gel a été placé dans le tampon de migration contenant 10% de SDS et mis sous tension (20 mA/gel).

- Analyse par la technique de Western-Blot

Après migration sur gel d'électrophorèse, les protéines ont été transférées sur une membrane de nitrocellulose pendant 5 à 15 min. La membrane a été ensuite saturée par un tampon contenant du lait écrémé à 5%, pendant une nuit, pour bloquer les sites d'interactions non spécifiques sur la membrane. Puis, après plusieurs étapes de rinçage avec du tampon PBS contenant 0,1% de Tween-20, un anticorps primaire anti-IgG conjugué avec la peroxydase de raifort (HRP) a été ajouté. Cet anticorps reconnait les sites spécifiques des chaînes lourdes et des chaînes légères de l'anticorps cible (IgG) produit par les cellules CHO. La membrane a été alors incubée pendant 1 h dans une solution de PBS contenant 5% de lait. Après une nouvelle série de rinçages, les IgG ont pu être révélées par une solution contenant 4 $mg.mL^{-1}$ d'o-phénylènediamine dihydrochloride (OPD) avec une formation de couleur orange-brune (Figure 37).

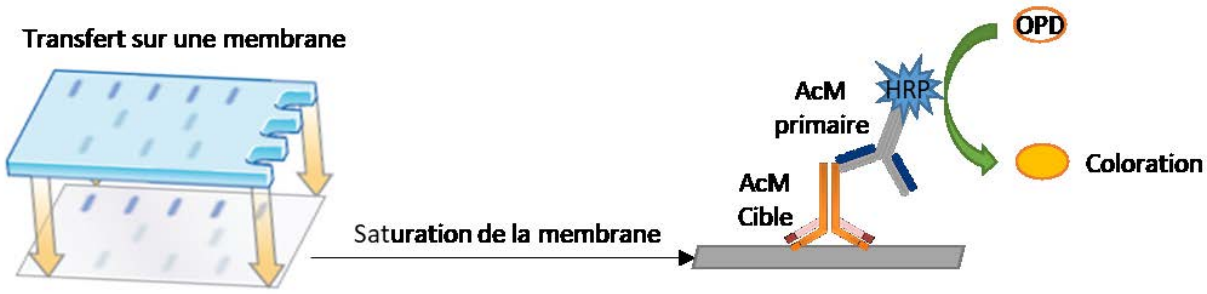


Figure 37 : Principe de fonctionnement de l'analyse de l'anticorps par Western-Blot

2.3.2. Quantification des anticorps

- Concentration des IgG totales

La concentration des IgG totales présentes dans le milieu de culture a été mesurée par l'analyseur Gallery directement après chaque prélèvement pour éviter la précipitation de ces IgG. L'analyse des IgG par immunoturbidimétrie a été réalisée par le kit IgG Bio (Roche). Les IgG présentes dans le milieu de culture précipitent en présence d'un antisérum spécifique. La turbidité provoquée par cette précipitation est mesurée à 340 nm après 2 min d'incubation.

- Analyse de la glycosylation des IgG

L'état de glycosylation des IgG produites a été analysé à la fin de chaque culture par un équipement HPLC-MS, (Thermo-Fisher), constitué d'un chromatographe liquide, possédant une pompe binaire de distribution de solvant connectée à un détecteur à photodiode (PDA), et d'un spectromètre de masse de type LTQ (Linear Trap Quadripole) en mode d'ionisation électrospray positif (ESI+). La colonne de séparation est une colonne C18 (150 mm * 2,1 mm) (Grace / Alltech). Les phases mobiles sont composées de 0,1% d'acide trifluoroacétique dans l'eau pour la phase A, et d'acétonitrile avec 0,1% d'acide trifluoroacétique pour la phase B. Le débit des phases mobiles est fixé à 200 $\mu L.min^{-1}$, avec le gradient d'élution suivant : 4% de phase B pendant 5 min, passage de 4% à 15% de phase B pendant 25 min, et passage de 15% à 98% de B pendant 5 min. Les conditions de la spectrométrie de masse étaient les suivantes : tension de pulvérisation à 5 kV ; débits de gaz de gaine, gaz auxiliaire et gaz de balayage réglés respectivement à 40, 10 et 10 en $UA.min^{-1}$; température capillaire à 300 °C ; tension capillaire à 36 V ; lentille à tube, lentille fendue et lentilles frontales respectivement à 80 V, -44 V et - 3,25 V.

Afin d'augmenter la sensibilité de l'analyse, les glycannes ne représentant que 2 à 3% de la masse totale des IgG, celles-ci ont d'abord été hydrolysés par la trypsine (T8003, Sigma-Aldrich). Cette réaction permet d'hydrolyser une glycoprotéine entière en plusieurs glycopeptides de taille beaucoup plus petite sans modifier les sites de glycosylation. La trypsine a été choisie pour ce projet car cette méthode d'hydrolyse enzymatique est beaucoup étudiée dans le cas de l'analyse de la glycosylation des anticorps de type IgG par MS. Les tailles et les séquences glycanniques et peptidiques sont ainsi déjà répertoriées dans des bases de données (Kapur et al., 2014). Les données brutes de l'analyse ont été traitées à l'aide du logiciel XCALIBUR (Thermo-Fisher). Le pourcentage de chaque glycopeptide recherché a été obtenu en divisant l'aire sous le pic de son signal MS par l'aire sous le pic du signal des IgG totales. La concentration de chaque glycopeptide peut être ainsi obtenue en multipliant son pourcentage par la concentration des IgG totales analysées par l'analyseur Gallery.

2.4. Calcul des paramètres cinétiques des cultures

Les courbes représentant les mesures hors ligne au cours du temps ont tout d'abord été lissées par une fonction de lissage « smooth spline » à l'aide du logiciel Matlab version R2016a (Mathworks, Inc). Les paramètres cinétiques ont alors été calculés à partir de ces courbes lissées en utilisant un

pas de calcul de 1 h, après écriture des bilans de matière tenant compte du mode d'écoulement du procédé.

2.4.1. Vitesse spécifique de croissance

- Mode discontinu. La vitesse de croissance r_x a été calculée à partir de la concentration de cellules viables, X_V, à l'aide de l'équation suivante, sachant que le volume de milieu est constant :

$$r_x = \frac{dX_v}{dt} \qquad (10^5 \text{ cellules.mL}^{-1}.\text{h}^{-1}) \qquad \textbf{Eq. 15}$$

La vitesse spécifique de croissance μ peut ainsi être calculée par :

$$\mu = \frac{1}{X_v} \times \frac{dX_v}{dt} \qquad (\text{h}^{-1}) \qquad \textbf{Eq. 16}$$

Les vitesses spécifiques instantanées de croissance ont été calculées, à chaque instant, par la pente de la courbe du logarithme de X_v en fonction du temps.

- Mode recharge-récolte. Pour ces cultures, les calculs de µ entre chaque cycle avant les échanges du milieu de culture, ont été les mêmes que ceux effectués en mode discontinu.

- Mode semi-continu. Dans ce cas, le volume V n'étant pas constant, les paramètres doivent être calculés par les équations suivantes :

$$r_x = \frac{1}{V} \times \frac{d(X_v V)}{dt} \qquad (10^5 \text{ cellules.mL}^{-1}.\text{h}^{-1}) \qquad \textbf{Eq. 17}$$

$$\mu = \frac{1}{X_v V} \times \frac{d(X_v V)}{dt} \qquad (\text{h}^{-1}) \qquad \textbf{Eq. 18}$$

Les vitesses spécifiques instantanées de croissance ont été calculées à chaque instant par la pente de la courbe du logarithme de $X_v V$ en fonction du temps

2.4.2. Vitesses spécifiques de consommation des substrats et de formation des produits

Les vitesses de consommation des substrats, r_s, et de production des métabolites, r_p, ont été calculées à partir des équations suivantes, en supposant qu'il n'y a pas de dégradation des substrats et des métabolites.

- Mode discontinu :

$$r_s = -\frac{dS}{dt} \qquad \text{(mmole.L}^{-1}\text{.h}^{-1}\text{)} \qquad \textbf{Eq. 19}$$

$$r_p = \frac{dP}{dt} \qquad \text{(mmole.L}^{-1}\text{.h}^{-1}\text{)} \qquad \textbf{Eq. 20}$$

Avec S la concentration d'un substrat donné, et P la concentration d'un produit donné.

Les vitesses spécifiques de consommation, q_s, et de production, q_p, peuvent ainsi être calculées par :

$$q_{s} = -\frac{1}{X_v} \times \frac{dS}{dt} \qquad \text{(mmole.h}^{-1}\text{.10}^{8}\text{cellules}^{-1}\text{)} \qquad \textbf{Eq. 21}$$

$$q_{p} = \frac{1}{X_v} \times \frac{dP}{dt} \qquad \text{(mmole.h}^{-1}\text{.10}^{8}\text{cellules}^{-1}\text{)} \qquad \textbf{Eq. 22}$$

- Mode semi-continu.

F étant le débit d'alimentation (L.h^{-1}) et S_F la concentration de substrat dans le milieu d'alimentation (mmole.L^{-1}), ces paramètres ont été calculés avec les équations suivantes :

$$r_s = \frac{FS_F}{V} - \frac{d(SV)}{dt \times V} \qquad \text{(mmole.L}^{-1}\text{.h}^{-1}\text{)} \qquad \textbf{Eq. 23}$$

$$q_s = \frac{FS_F}{X_v \times V} - \frac{1}{X_v \times V}\frac{d(SV)}{dt} = \frac{1}{X_v}\left(\frac{FS_F}{V} - \frac{dS}{dt} - \frac{S}{V}F\right) \qquad \text{(mmole.h}^{-1}\text{.10}^{8}\text{cellules}^{-1}\text{)} \qquad \textbf{Eq. 24}$$

$$r_p = \frac{d(PV)}{dt \times V} \qquad \text{(mmole.L}^{-1}\text{.h}^{-1}\text{)} \qquad \textbf{Eq. 25}$$

$$q_p = \frac{1}{X_v \times V}\frac{d(PV)}{dt} = \frac{1}{X_v}\left(\frac{dP}{dt} + \frac{P}{V}F\right) \qquad \text{(mmole.h}^{-1}\text{.10}^{8}\text{cellules}^{-1}\text{)} \qquad \textbf{Eq. 26}$$

2.4.3. Rendements métaboliques

Le rendement correspondant aux cellules formées par substrat consommé, $Y_{X/S}$, et le rendement correspondant au produit généré par substrat consommé, $Y_{P/S}$, ont été calculés selon les équations suivantes sur l'intervalle de temps entre *i* et *i-l*, *l* pouvant varier en fonction des études concernées :

$$Y_{X/S} = \frac{(X_i - X_{i-l})}{(S_{i-l} - S_i)} \qquad (10^8 \text{ cellules.mmole}^{-1}) \qquad \textbf{Eq. 27}$$

$$Y_{P/S} = \frac{(P_i - P_{i-l})}{(S_{i-l} - S_i)} \qquad (\text{mole.mole}^{-1}) \qquad \textbf{Eq. 28}$$

Le rendement de production correspondant aux IgG produites par cellules produites, $Y_{IgG/X}$, a été calculé avec l'équation suivante :

$$Y_{IgG/X} = \frac{(IgG_i - IgG_{i-l})}{(X_i - X_{i-l})} \qquad (\text{mg}.10^8 \text{ cellules}^{-1}) \qquad \textbf{Eq. 29}$$

Dans le cas des cultures en mode semi-continu, ces rendements ont été calculés de la même manière mais en utilisant la quantité cumulée des cellules, des substrats (résiduel et ajouté) et des produits pour tenir compte de la variation de volume de milieu de culture.

2.4.4. Productivité

La productivité spécifique en IgG est calculée par l'équation suivante :

$$P_{IgG/X} = \frac{1}{(X_i - X_{i-l})} \times \frac{(IgG_i - IgG_{i-l})}{(t_i - t_{i-l})} \qquad (\text{pg.cellules}^{-1}.\text{jour}^{-1}) \qquad \textbf{Eq. 30}$$

Dans le cas des cultures en mode semi-continu, la productivité spécifique des IgG a été calculée de la même manière mais en utilisant la quantité cumulée (au lieu des concentrations) des cellules, et des IgG produites, pour tenir compte de la variation de volume de milieu de culture.

3. Analyses spectroscopiques en ligne

Afin de suivre en ligne des paramètres caractéristiques de la culture de cellules CHO, plusieurs méthodes spectroscopiques (diélectrique, NIR et Raman) ont été mises en œuvre.

3.1. Spectroscopie diélectrique

Le système Evo 200 iBiomass (Hamilton) a été utilisé pour mesurer les propriétés diélectriques des cellules vivantes. Il comporte une sonde stérilisable avec 4 électrodes (2 négatives et 2 positives) reliée à un préamplificateur et à un boitier de contrôle (Figure 38).

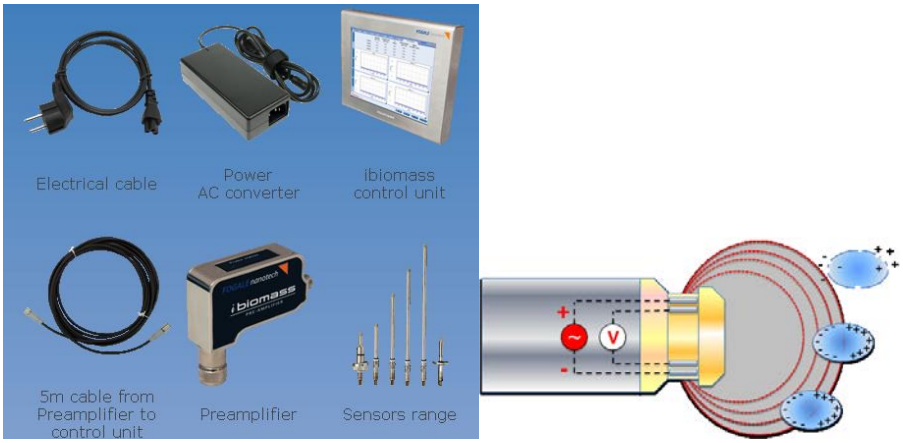


Figure 38 : Composants du système Evo 200 iBiomass, et représentation schématique de la tête de la sonde diélectrique

Avant l'ensemencement des cellules dans les cultures en bioréacteur, la sonde diélectrique est stabilisée dans du milieu de culture sans cellules pendant 12 h. L'acquisition de la permittivité en cours de culture est réalisée toutes les 12 min pour les cultures en mode discontinu et recharge-récolte, et toutes les 20 min pour les cultures semi-continues. Un filtre de lissage « Super-High » permet de lisser les signaux de permittivité obtenus. Une corrélation entre la permittivité enregistrée par la spectroscopie diélectrique et la concentration de cellules viables mesurées hors ligne par le Vi-CELL peut alors être déterminée, pour obtenir une mesure prédictive des concentrations inconnues.

3.2. Spectroscopie proche Infrarouge (NIR)

Une sonde stérilisable à fibre optique (Precision Sensing Devices, MA) couplée au spectromètre NIR Antaris II (Thermo Scientific) a été utilisée pour les mesures *in situ* en bioréacteur (Figure 39). Cette sonde fonctionne en mode transflectance (combinaison de la réflexion et de la transmission), avec un trajet optique de 1 mm (2 mm de trajet effectif). Les paramètres opérationnels de cette spectroscopie NIR ont été étudiés et fixés préalablement. La fréquence d'acquisition spectrale est de 20 minutes dans une zone de nombres d'onde allant de 4000 à 10000 cm^{-1} (résolution optique de 8 cm^{-1}). Chaque spectre correspond à 128 spectres moyennés. Le spectre de l'air a été choisi comme

référence. Une région spectrale spécifique (de 4300 à 4752 et de 5338 à 9000 cm^{-1}) a été sélectionnée afin d'éviter l'influence des molécules d'eau et de limiter le bruit lié à l'appareil et aux fibres optiques. Les spectres correspondant aux points d'échantillonnage ont été sélectionnés pour calibrer ultérieurement les modèles de prédiction avec les mesures hors ligne.

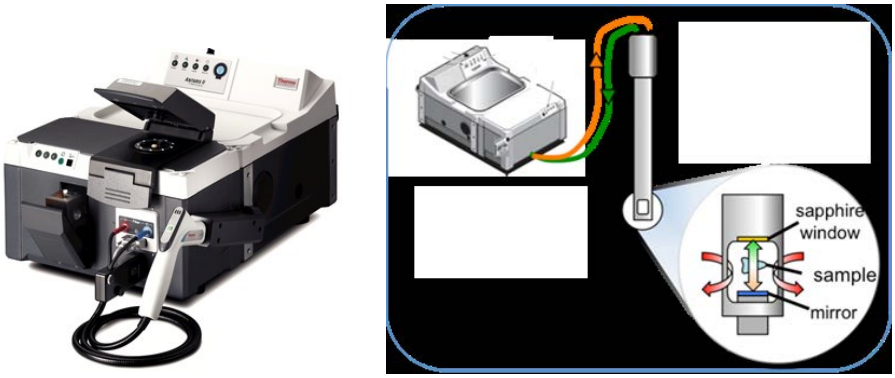


Figure 39 : Spectromètre NIR Antaris II et représentation schématique de la sonde stérilisable

3.3. Spectroscopie Raman

Le spectromètre Raman BioViserion® (Indatech, France) disposant d'une sonde à immersion reliée par fibres optiques a également été utilisé pour le suivi *in situ* en bioréacteur (Figure 40). La sonde est plongée dans le milieu de culture et fixée à environ 5 cm du fond du bioréacteur pour limiter l'influence de la paroi en verre sur les spectres. La source d'excitation est un laser à 785 nm avec une puissance de 400 mW. Le bioréacteur est protégé par une enveloppe en PVC opaque conçue spécifiquement pour notre étude afin d'éviter les perturbations liées à la lumière. De manière analogue à la spectroscopie NIR, les paramètres opérationnels ont été réglés préalablement afin d'optimiser le rapport signal/bruit. La fréquence d'acquisition spectrale était de 20 min, chaque spectre étant la moyenne de 6 spectres avec un temps d'acquisition de 60 s pour chaque spectre. L'acquisition a été réalisée entre 100 et 3425 cm-1 (résolution optique de 2 cm-1). La région spectrale d'intérêt a été choisie entre 800 et 1600 cm-1. Les spectres correspondant aux points d'échantillonnage ont été sélectionnés pour calibrer ultérieurement les modèles de prédiction avec les mesures hors ligne.

Figure 40 : Spectromètre Raman BioViserion® et sonde stérilisable

4. Méthodes chimiométriques pour le traitement des données spectrales

Le logiciel Matlab R2016a et le logiciel PLS_Toolbox 8.5 (Eigenvector, Inc) ont été utilisés pour le traitement des données spectrales et le développement des modèles chimiométriques.

4.1. Répartitions des données spectrales entre calibration et prédiction

Pour développer les modèles spectroscopiques de prédiction en ligne, les données spectrales (diélectrique, NIR et Raman) issues de six cultures en bioréacteur ont été utilisées. Les spectres

aberrants ont été détectés et éliminés de l'ensemble des données à l'aide d'une analyse en composante principales (PCA), seuls les spectres se situant dans la zone de confiance définie par les valeurs de T^2 Hotelling et Q résiduel étant retenus. Afin d'augmenter la variabilité des données et ainsi améliorer la robustesse des modèles, différents modes d'alimentation des cultures ont été utilisés pour le développement des modèles : discontinu, ajout ponctuel de glucose (glucose spiking) et recharge-récolte. La répartition des échantillons entre ces différents modes est présentée dans le Tableau 7 ci-dessous :

Tableau 7 : Echantillons utilisés pour le développement des modèles chimiométriques

Mode de culture	**Nombre d'échantillons**
Discontinu	59
Ajout ponctuel de glucose	22
Recharge-récolte	38

Les spectres ont ensuite été séparés, à l'aide de l'algorithme Kennard Stone, en deux ensembles, l'un (80 % des spectres) pour la calibration et la validation croisée, et l'autre (20 % des spectres) pour la prédiction. Dans cette étude, l'algorithme a été calculé à partir des spectres Raman car les points de prédiction étaient mieux répartis dans la gamme de mesure. Afin de pouvoir comparer la performance des différents modèles spectroscopiques, les mêmes ensembles de données ont été utilisés pour toutes les méthodes spectroscopiques. La Figure 41 présente la procédure globale suivie pour le développement de modèles chimiométriques.

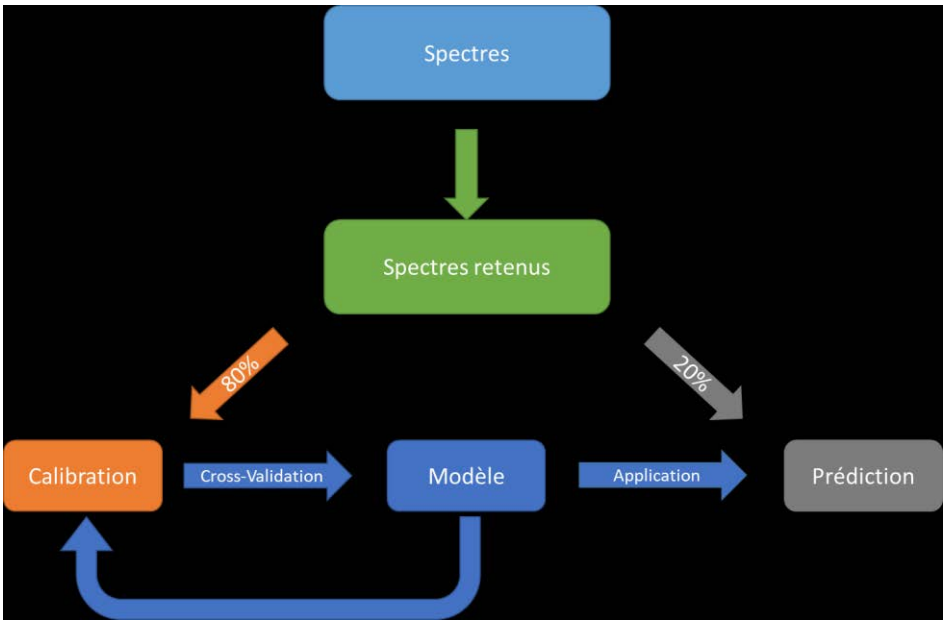


Figure 41 : Procédure de développement des modèles chimiométriques

4.2. Modèles de prédiction basés sur la spectroscopie diélectrique

La sonde diélectrique permet de mesurer la permittivité (Δε) du surnageant de culture, qui reflète la concentration des cellules viables (VCD). Cette dernière a été également mesurée hors ligne à l'aide de l'automate de comptage Vi-CELL. Les valeurs de VCD mesurées hors ligne ont alors permis de calibrer les valeurs de permittivité obtenues en ligne. Une méthode de calibration linéaire a été utilisées.

La calibration linéaire était basée sur les équations simplifiées de Maxwell-Wagner et Bruggeman :

$$\Delta\varepsilon = 3\pi r^4 C_m \cdot VCD \qquad \textbf{Eq. 31}$$

r : rayon de cellules
C_m : capacitance membranaire des cellules

En se basant sur les hypothèses que les cellules sont sphériques, que la moyenne du rayon des cellules est supposée constante tout au long de la culture et que la capacitance membranaire des cellules est constante, la permittivité peut alors être considérée comme proportionnelle à la VCD, selon la corrélation linéaire suivante :

$$\Delta\varepsilon = a \times VCD + b \qquad \textbf{Eq. 32}$$

4.3. Modèles de prédiction basés sur les spectroscopies NIR et Raman

Les spectres NIR et Raman enregistrés pendant les cultures ont été utilisés pour développer des modèles de prédiction des concentrations des différents composés du milieu de culture : cellules viables, glucose, lactate, glutamine, ions ammonium, IgG, ainsi que la glycosylation des anticorps (macro et micro-hétérogénéité). Les spectres ont été prétraités par la méthode SNV et centrés par la moyenne. Les valeurs de référence des différents composés mesurées hors ligne ont aussi été centrées par la moyenne. Ces prétraitements ont permis de minimiser le bruit de fond et les variations intrinsèques du spectromètre. La méthode de régression PLS a ensuite été utilisée pour calibrer les spectres prétraités avec les valeurs hors ligne de chaque composé. Le nombre optimal de variables latentes utilisé pour chaque modèle a été déterminé lorsque le maximum de variance était expliqué par un minimum de composantes principales.

4.4. Evaluation des performances des modèles de calibration et prédiction

La performance des modèles a été évaluée par les coefficients de régression, r^2, et par les erreurs quadratiques moyennes (Root mean square error, RMSE), calculés pour la calibration (r^2_C, RMSEC), la validation croisée (r^2_{CV}, RMSECV) et la prédiction (r^2_P, RMSEP). La valeur de r^2 est un indicateur de la qualité de la prédiction d'une régression linéaire, en traduisant la linéarité de la méthode développée. Les valeurs des RMSE permettent d'estimer l'exactitude de la prédiction en calculant la différence entre les valeurs prédites par le modèle et les valeurs réellement observées. La méthode « Venetian Blinds » a été utilisée pour la validation croisée, chaque sous-ensemble de données de prédiction est déterminé en choisissant chaque 9ème donnée dans l'ensemble de données de calibration à partir de la 1ère donnée.

Les performances des modèles ont aussi été évaluées par d'autres figures de mérite (FOM) comprenant la précision (BCMSEP : bias corrected MSEP), la sensibilité analytique (γ), la sélectivité (SEL), le rapport signal sur bruit (S/N), la limite de détection (LOD), le rapport de la déviation de performance (RPD), et le rapport d'erreur de gamme (RER) selon les équations suivantes :

$$BCMSEP = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2-\left[\sum_{i=1}^{n}(y_i-\hat{y}_i)\right]^2/n}{n-1} \quad \textbf{Eq. 33}$$

y_i : valeurs de référence mesurées
$\hat{y}_i$: valeurs prédites
n : nombre d'échantillons

$$\gamma = \frac{||(x_i \cdot b) \cdot (b^T \cdot b)^{-1} \cdot b^T / y_i||}{\delta r} \quad \textbf{Eq. 34}$$

x_i : spectres prétraités
b : coefficient de régression PLS
δr : bruit de l'instrument estimé (écart-type des réplicats des spectres des échantillons blancs prétraités)

$$SEL_i = \frac{||(x_i \cdot b) \cdot (b^T \cdot b)^{-1} \cdot b^T||}{||x_i||} \quad \textbf{Eq. 35}$$

$$S/N_i = \frac{||(x_i \cdot b) \cdot (b^T \cdot b)^{-1} \cdot b^T||}{\delta r} \quad \textbf{Eq. 36}$$

$$LOD = 3.3\delta r \frac{1}{||(x_i \cdot b) \cdot (b^T \cdot b)^{-1} \cdot b^T / y_i||} \quad \textbf{Eq. 37}$$

$$RPD = \frac{\text{écart-type des mesures de référence}}{RMSEP} \quad \textbf{Eq. 38}$$

$$RER = \frac{gamme\ de\ concentration\ des\ mesures\ de\ référence}{RMSEP} \quad \textbf{Eq. 39}$$

En outre, les scores de l'importance des variables en projection (VIP) ont été calculés pour permettre d'identifier les régions spectrales les plus pertinentes pour un modèle développé. Le score VIP de la j^{ème} variable (Raman shift ou nombre d'onde) est calculé avec la formule ci-dessous. En général, les variables ayant un VIP score supérieur à 1 sont considérées comme ayant une influence significative pour un modèle donné.

$$VIP_j = \sqrt{\frac{\sum_{f=1}^{F} w_{jf}^2 \times SSY_f \times J}{SSY_{total} \times F}} \quad \textbf{Eq. 40}$$

w_{jf} : poids de la variable j et du composant f
SSY_f : somme des carrés de la variance expliquée par le f^{ème} composant
SSY_{total} : somme des carrés de la variance expliquée totale
J : nombre de variables
F : nombre des composants

Résultats & Discussion

Chapitre I

Suivi en ligne de la vitesse spécifique de croissance cellulaire par spectroscopie diélectrique

I.1. Introduction **98**

I.2. Impact de l'estimation en temps réel de la vitesse spécifique de croissance cellulaire sur la glycosylation des anticorps produits **99**

I.3. Données supplémentaires **110**

I.3.1. Précisions sur le calcul de la vitesse spécifique de croissance cellulaire en temps réel *110*

I.3.2. Comparaison de cultures en mode recharge-récolte utilisant diverses stratégies de renouvellement du milieu de culture *113*

I.4. Conclusion **116**

I.1. Introduction

Les procédés de culture de cellules animales sont devenus incontournables pour la production d'AcM à usage thérapeutique. Ces bioprocédés mettant en œuvre des cellules mammaliennes sont particulièrement complexes, les cellules étant très sensibles aux changements environnementaux intervenant au sein du réacteur. De ce fait, pour répondre au principal enjeu de l'approche QbD, différents paramètres doivent être parfaitement contrôlés afin d'assurer la meilleure qualité possible du produit. Les cellules représentent sans aucun doute l'un des paramètres les plus importants à suivre au cours du procédé. En effet, elles conditionnent étroitement la production des AcM, et ainsi la performance globale du procédé.

Depuis quelques années, dans le cadre de l'initiative PAT, diverses techniques ont été proposées pour suivre en ligne la concentration des cellules vivantes. La spectroscopie diélectrique présente plusieurs avantages qui la distinguent des autres méthodes, dont, en particulier, une simplicité d'utilisation, une insensibilité aux débris cellulaires, et une absence d'encrassement des fenêtres de mesure. Actuellement, grâce à une sonde stérilisable immergée dans le milieu de culture cellulaire, elle est essentiellement utilisée pour le suivi en ligne de la concentration des cellules viables, cellules qui sont les seules à pouvoir se polariser sous l'effet d'un champ électrique. Cependant, il serait intéressant d'étudier la capacité de la spectroscopie diélectrique à suivre, non seulement, la concentration des cellules, et au-delà, une indication de leur état physiologique.

Dans ce cadre, l'objectif de ce chapitre est de développer une nouvelle approche, basée sur l'utilisation de la spectroscopie diélectrique *in situ,* pour le suivi en ligne de cultures de cellules CHO en bioréacteur, en mesurant, non seulement la concentration des cellules viables, mais également la vitesse spécifique de croissance cellulaire (μ). En effet, la mesure en temps réel de la concentration des cellules vivantes rend possible le calcul direct de cette vitesse, qui peut être considérée comme un indicateur direct de l'état physiologique des cellules, sa variation pouvant influencer à la fois la quantité et la qualité des anticorps produits (Jang and Barford, 2000). Or, à l'heure actuelle, les valeurs de μ sont généralement calculées en fin de culture à partir des analyses hors ligne des cellules. Cette approche est donc inadéquate pour envisager des stratégies de contrôle du procédé en temps réel. Pour répondre à ce défi, nous avons, dans un premier temps, développé un modèle d'estimation des valeurs de μ en temps réel à partir des permittivités mesurées en ligne par la spectroscopie diélectrique. L'intérêt de ce modèle pour détecter l'instant critique de la culture, où une diminution significative de μ est perçue, a ensuite été évalué. Puis, cette approche méthodologique a été utilisée dans le cas particulier d'une culture conduite en mode recharge-

récolte, afin de renouveler le milieu de culture lors de la détection de ces instants critiques, l'objectif étant de maintenir la valeur de μ proche de sa valeur maximale. Les performances de cette culture basée sur une stratégie de suivi en ligne de μ ont ensuite été comparées à celles d'autres cultures réalisées en mode recharge-récolte utilisant des stratégies hors ligne. Une attention particulière a été portée à l'impact de la stratégie en ligne sur la qualité de la glycosylation des IgG produites.

La majeure partie des résultats de ce chapitre fait l'objet d'un article scientifique soumis pour publication dans la revue Biotechnology and Bioengineering et présenté ci-dessous. Une petite partie supplémentaire permettra ensuite de détailler la méthode de calcul en ligne de μ et de présenter des résultats expérimentaux complémentaires.

I.2. Impact de l'estimation en temps réel de la vitesse spécifique de croissance cellulaire sur la glycosylation des anticorps produits

Manuscrit de la publication

Control of IgG glycosylation by *in situ* and real-time estimation of specific growth rate of CHO cells cultured in bioreactor

Meng-Yao Li, Bruno Ebel, Fabrice Blanchard, Cédric Paris, Emmanuel Guedon, Annie Marc

Abstract

The cell specific growth rate (μ) is a critical process parameter for antibody production processes performed by animal cell cultures, as it describes the cell growth and reflects the cell physiological state. When there are changes in these parameters, which are indicated by variations of μ, the synthesis and the quality of antibodies are often affected. Therefore, it is essential to monitor and control the variations of μ in order to achieve high product quality. In this study, a novel approach for on-line estimation of μ was developed based on the Process Analytical Technology (PAT) initiative by using an *in situ* dielectric spectroscopy. Critical moments, such as significant μ decreases, were successfully detected by this method, in association with changes in cell physiology as well as with an accumulation of non-glycosylated antibodies. Thus, this method was used to perform medium renewals at the appropriate times, in order to maintain the values of μ close to its maximum. Using this method, we demonstrated that the physiological state of cells remained stable, and the level of

non-glycosylated antibodies was kept inferior to 5% all along the cell culture, leading to better process performances compared to the reference feed-harvest cell culture carried out without real time monitoring of μ.

1 Introduction

The need of biopharmaceutical protein products, in particular recombinant monoclonal antibodies (mAbs), has been constantly increasing in the last decade (Walsh, 2014). Chinese hamster ovary (CHO) cells are the most commonly used cell lines for mAb production due to their capability of performing human-like post-translational modifications (PTMs) (Cole *et al.*, 2015). Glycosylation is one of the most important PTMs for mAbs, as it impacts effector functions, pharmacokinetics, antigenicity, safety, stability and solubility of the mAbs produced (Tharmalingam *et al.*, 2015). Therefore, glycosylation is a critical quality attribute (CQA) for mAb production processes that should be taken into consideration in all manufacturing steps, as outlined in the quality by design (QbD) approach initiated by the Food and Drug Administration (FDA) in 2004 (Quality Guidelines: ICH). The QbD initiative emphasizes the importance of using process analytical technologies (PAT) in order to monitor and control in real-time the critical process parameters (CPPs), which may influence the product CQAs in biopharmaceutical production processes (FDA, 2004).

Among various CPPs of mAb production bioprocesses, viable cell density (VCD) is undoubtedly one of the most important, as it is closely related to the cell growth, the mAb production, and overall process performances (Lee *et al.*, 2015). With increasing production potential of these bioprocesses, reliable in-line and real-time measurements of VCD are more and more required. Various PAT tools have been proposed to monitor VCD in real-time, including dielectric spectroscopy (Ansorge *et al.*, 2007; Cannizzaro *et al.*, 2003; Courtès *et al.*, 2016; Opel *et al.*, 2010; Yardley *et al.*, 2000), near-infrared and Raman spectroscopies (Abu-Absi *et al.*, 2011; Cervera *et al.*, 2009), *in situ* microscopy (Guez *et al.*, 2004), acoustic resonance densitometry (Kilburn *et al.*, 1989), soft-sensor-based approaches (Kiviharju *et al.*, 2008). Among these technologies, dielectric spectroscopy is probably one of the most reliable methodologies to monitor VCD due to its simplicity, robustness, and its capability of *in situ* fast signal acquisition, which is non-invasive and non-destructive for cell cultures. The dielectric spectroscopy is also insensible to air bubbles and cell debris which often cause problems for other methods (Justice *et al.*, 2011). Dielectric spectroscopy is based on measurements of the ability of viable cells to store electrical charge as a function of the frequency of the applied electrical field. The basic theoretical background of dielectric properties of biological cells was described elsewhere (Yardley *et al.*, 2000). However, on-line monitoring of VCD alone gives an

incomplete vision of cell growth and cell physiological state. To date, it remains challenging to monitor in real-time the cellular physiological state due to the complexity of animal cell system (Henry *et al.*, 2007).

The specific cell growth rate (μ) is the one of the most direct indicators of cell physiological state. When cells are in their active growth stage (μ close to its maximum, μ_{max}), they are able to produce enough energy for energy-consuming actions such as cell division and recombinant protein synthesis. On the contrary, when there are nutrients depletion, toxins accumulation, or other physico-chemical stresses appear, energy production of cells may be negatively influenced, resulting in a decrease in μ (Kondo *et al.*, 2000). In this case, not only the quantity but also the quality of the recombinant proteins produced could be affected, since non-correctly performed PTMs associated to mAb production could appear during the process (Kochanowski *et al.*, 2008). Therefore, μ appears as an essential parameter for mAb production bioprocesses, which should be carefully monitored and controlled throughout the process to ensure the quantity and the quality of the mAb synthesized. Classically, μ is often calculated indirectly from off-line measurements of cell densities, or by macroscopic kinetic modeling approaches, which could be inaccurate, resulting in delayed information which is not adequate for control strategies (Xiong *et al.*, 2015).

In the present work, a novel approach of real-time estimation of μ was developed based on on-line monitoring of VCD by using *in situ* dielectric spectroscopy and mass balance equations. This approach was then applied on a recombinant immunoglobulin G (IgG) production process with CHO cells cultured in bioreactor in feed-harvest mode, in order to investigate its potential for the improvement of process control. Quality of the IgG produced was assessed by analyzing the macro-heterogeneity of IgG glycosylation, which corresponds to the glycosylation site occupancy. Overall, the results of this work demonstrate that μ is an important process parameter allowing to detect early cell physiological state changes. On-line estimation of μ by using dielectric spectroscopy leads to a better process control for the feed-harvest cultures, and as a result, the IgG quality concerning the glycosylation is improved.

2 Materials and methods

2.1 Cell Culture

A genetically modified DG44 CHO cell line (CHO M250-9) kindly provided by Bioprocessing Technology Institute (Singapore) was used in this study. The culture medium employed during this work was a protein-free medium mixture consisting of a 1:1 ratio of PF-CHO (GE Healthcare, USA) and CD-CHO (Fisher Scientific, USA) supplemented with 4 mM L-glutamine (Sigma Aldrich, USA), and 0.1% Pluronic F-68 (Sigma Aldrich, USA). Cells were initially cultured in shake flasks (Fisher Scientific, USA) and were incubated in an agitated incubator (Kuhner, Switzerland) at 37 °C, 5% CO_2, 80% humidity with an agitation rate of 70 rpm. Bioreactor cell culture of CHO cells were performed in 2.5 L bench-top bioreactors (Pierre Guérin, France). The exponentially growing

cells were seeded at about 3×10^5 cell/ml in 1.5 L of working volume. Culture temperature was maintained at 37 °C and the pH was regulated at 7.2 using 0.5 M sodium hydroxide and CO_2. Dissolved oxygen (DO) was controlled at 50% of air saturation by delivering oxygen through the sparger. Agitation rate was fixed at 90 rpm throughout the culture.

2.2 Off-line analysis

Samples were taken 3-4 times per day. VCD, viability, and average diameter of cells were measured by Vi-CELL™ cell counter (Beckman Coulter, USA) on the basis of trypan blue dye exclusion of the viable cells. An automated photometric analyzer Gallery™ (Thermo Fisher Scientific, USA) was used to measure glucose, lactate, glutamine, ammonium ions and IgG titer.

For glycosylation analysis, IgGs were precipitated by cold acetone and were incubated at -20 °C for 2 h. 5 min of centrifugation at 13,000 g was carried out to recover the IgGs. Recovered IgGs were dissolved in 500 μL of 50 mM ammonium hydrogen carbonate buffer and was then heated to 95 °C for 10 min to denature IgGs. Next, 20 μL of prepared trypsin solution (Sigma-Aldrich, USA) was added in the cooled IgG samples at a concentration of 1 g/L. Samples were incubated overnight at 37 °C, and the reaction was stopped by adding 15 μL of formic acid (98%, Sigma-Aldrich, USA). Analysis of peptides compounds from IgG hydrolysate was performed on a UHPLC-MS system (ThermoFisher Scientific, USA) connected to a photodiode array detector (PDA) and a LTQ mass spectrometer equipped with an atmospheric pressure ionization interface operating in positive electrospray mode (ESI+). Details of the MS analyses have been described previously (Li *et al.*, 2018).

The abundance of non-glycosylated IgG form was obtained by dividing its MS peak integral to the total MS glycopeptide signal. The concentration of non-glycosylated IgGs was calculated by multiplying the percentage of non-glycosylated form of digested IgGs by total IgG titer measured by Gallery™ device.

2.3 *In-situ* dielectric spectroscopy

The cell culture permittivity was measured in real-time using a sterilizable dielectric probe connected to an Evo 200 iBiomass system (Hamilton, Switzerland) at a working frequency of 1,000 kHz. Measurements were performed every 12 min and the baseline was set by recording the permittivity of the cell-free medium, prior to cell seeding. The permittivity can be related to VCD using a linear relationship (Courtès *et al.*, 2016):

$$\varepsilon = \alpha \times VCD + \beta \tag{1}$$

In this study α equals to 13 and β equals to 3.

Based on VCD on-line estimation and mass balance equation, μ can be obtained in real-time by calculating the slope of a moving window linear regression on the logarithmic values of VCD in an interval of time (Eq. 2).

$$\mu = \frac{dVCD}{dt} \times \frac{1}{N_{vc}} = \frac{\Delta \ln(VCD)}{\Delta t} \tag{2}$$

To reduce the noise of data collected from dielectric spectroscopy, moving average filters were used in order to have a more stable on-line μ estimation. Meanwhile, the residual distance (R) was calculated in real-time to evaluate the variations of the μ values with respect to the value of μ_{max} (Eq. 3).

$$R = (\mu_{max} - \mu_t)^2 \tag{3}$$

Here μ_{max} is a predefined value of the maximum value of μ at the exponential growth phase obtained from previous cultures, and the μ_t is the on-line estimation of μ at the instant *t*.

Moreover, using multi-frequency dielectric measurements and off-line cell diameter measurements, a cell-specific dielectric property parameter, the intracellular conductivity (σ_i) can be calculated with Eq. 5-7:

$$p_m = \frac{4}{3}.\pi.\left(\frac{d}{2}\right)^3.VCD \tag{5}$$

$$\sigma_a = \frac{\sigma}{(1-p_m)^{1.5}} \tag{6}$$

$$\sigma_i = \frac{8.\pi.f_c.\Delta\varepsilon.\sigma_a}{9.p_m.\sigma_a - 4.\pi.f_c.\Delta\varepsilon} \quad (7)$$

Where p_m is the viable biovolume fraction, and σ_a (S/m) is the conductivity of the medium. The values used in these calculations are the critical frequency f_c (Hz), the magnitude of the β-dispersion drop $\Delta\varepsilon$ (F/m), and the suspension conductivity σ (S/m) measured by the dielectric spectroscopy; the average cell diameter *d* (m) measured off-line by Vi-CELL™ cell counter.

3 Results and discussion

3.1 Cell growth, metabolism, and IgG production kinetics in batch culture

To assess the influence of cell growth on IgG glycosylation during batch cultures, CHO cells were cultured in a 2-L bioreactor, and cell density, viability, nutrients and byproducts concentration, IgG production and IgG glycosylation were monitored off-line. In addition, the off-line values of μ were calculated at the end of the culture using VCD values measured off-line.

Although cells displayed excellent viabilities (> 95%) until 127 h, the value of μ started to decrease after about 100 h of culture (Fig. 1A). This decrease of μ is probably due to glutamine depletion (Fig. 1B), since glutamine is proved to be one of the most important key-metabolites in several biosynthetic pathways of CHO cells. Its availability can affect the purine and pyrimidine synthesis rate (Hayter *et al.*, 1991). Alternatively, it is possible that the accumulation of lactate and ammonia inhibits the growth of cells according to other authors (Glacken *et al.*, 1986; Hassell *et al.*, 1991; Yang and Butler, 2000). However, in our case, the concentration of ammonium ions and lactate were relatively low (< 5 mM and < 20 mM respectively) (Fig. 1B), which were proved to have little or negligible effect on the growth of cells (Hayter *et al.*, 1991; Lao and Toth, 1997). The intracellular conductivity (σ_i), which measures the ability of the cell's intracellular environment to conduct electrical current, was relatively stable during the first 95 h of culture, followed by a light decrease until 103 h before a significant increase after this period, indicating potential changes in cell physiological state (Fig. 1C). As suggested by other authors, the decrease of this parameter often correspond to nutrients limitations and depletions (Ansorge Sven *et al.*, 2009; Opel *et al.*, 2010). In our case, the decrease of σ_i is associated with glutamine exhaustion. In addition, the final increase of σ_i was reported to be linked to cell death since an alteration in cell membrane integrity could result in the entry of medium ions into cells, leading to significant changes in the intracellular environment (Opel *et al.*, 2010). As for IgG production, Fig. 1D shows that almost all the IgGs produced were fully glycosylated until 100 h of culture, then non-glycosylated IgGs began to accumulate in cell culture and reached about 30% of the total IgGs, resulting in an important product quality decrease.

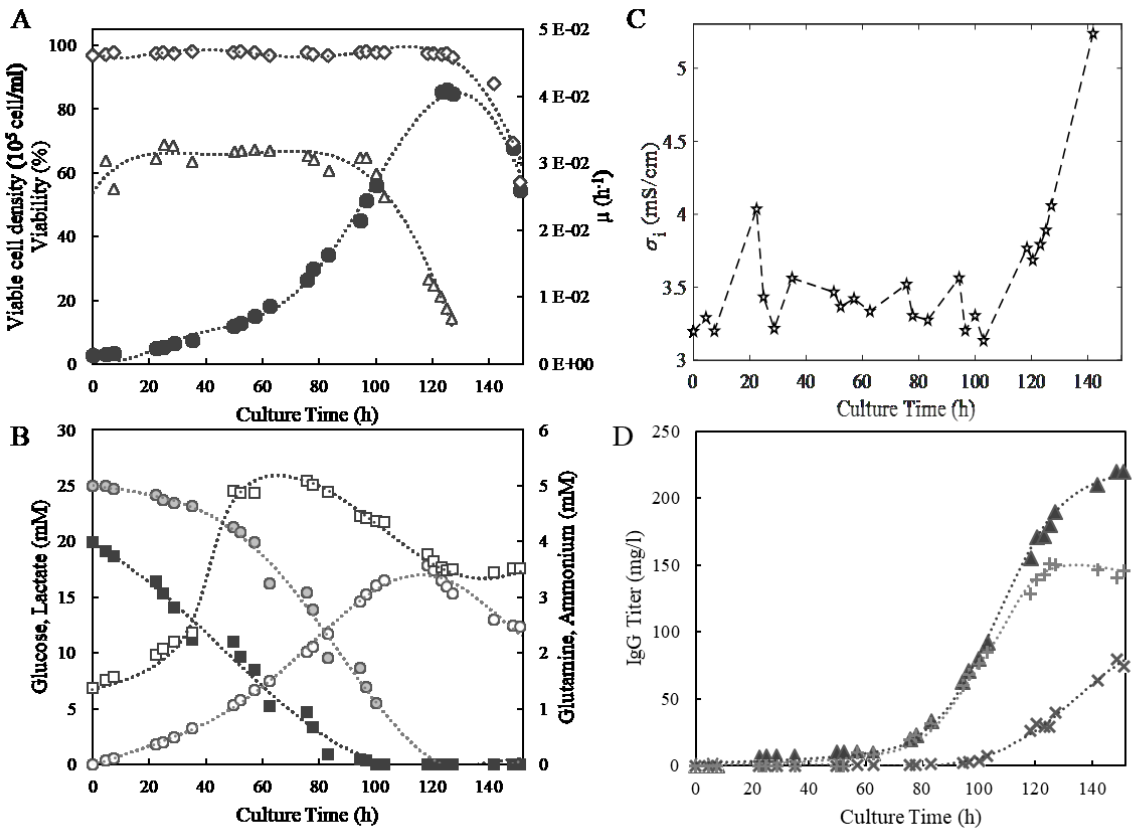


Figure 1. *Kinetics of cell growth (A): viable cell density (●), viability (◇), and specific growth rate (μ) (△); nutrients consumption and by-product accumulation (B): glucose (■), glutamine (■), lactate (■), and ammonium ions (■); cell physiological state (C): intracellular conductivity (σ_i) (☆); IgG production (D): total IgG (▲), glycosylated IgG (+), non-glycosylated IgG (×) throughout a representative batch culture of CHO cell in bioreactor.*

It seems that the diminution of μ coincided with the accumulation of non-glycosylated IgG form, which could be explained by an insufficient energy production of CHO cells due to the glutamine exhaustion after about 100 h, as suggested by Xie and Wang (1996). This decrease of μ led to cell physiological state changes as showed by σ_i calculations. Consequently, the recombinant protein production was also affected, resulting in a loss of ability for cells to perform correctly the PTMs for the proteins produced because of less available energy (Hayter *et al.*, 1993). In this study, we demonstrated that the global glycosylation of IgGs was influenced by the growth of cells, since the non-glycosylated IgG level increased abundantly after the decrease of μ, leading to a significant decrease in the final product quality at the end of the culture.

Therefore, as a direct indicator of the cell physiological state, μ is a key-parameter of cell culture processes and should be accurately monitored and controlled as rapidly as possible, in order to limit

the physiological changes of cells, and thus make sure that the glycosylation of IgGs produced is not altered.

3.2 *In situ* and real-time estimation of μ

In order to estimate in real-time the value of μ, an *in situ* probe connected with dielectric spectroscopy was used. The permittivity was measured every 12 minutes during cell cultures, and the on-line VCD was calculated from permittivity measured, by using a linear correlation. Estimation of on-line VCD is presented alongside with off-line VCD measurements in Fig. 2A. The values of VCD predicted were in good agreement with the off-line VCD values (r^2 0.99).

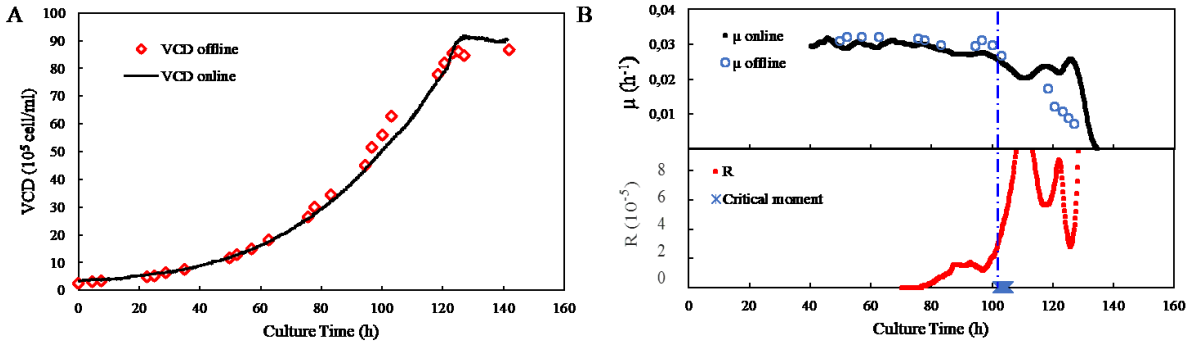


Figure 2. *On-line VCD prediction (A): VCD predicted (■) vs. VCD measured (◆); On-line μ estimation (B): μ estimated (■) vs. μ calculated off-line (◆), on-line calculation of R (■), and on-line critical moment (✕) detection (Blue dotted line).*

In order to calculate the value of μ, VCD predicted was firstly converted to logarithmic form (Ln(VCD)), then, the slope of the moving window linear regression of Ln(VCD) as a function of culture time was calculated to obtain the value of μ, as illustrated by equation (3). However, data collected from dielectric probe presented high level of noise, so smoothing filters were applied in order to make data more stable for μ estimation. In this work, two moving average filters were implemented to smooth the VCD predicted and the Ln(VCD) calculated. The on-line values of μ estimated are presented in Fig. 2B, which are validated by the off-line values of μ calculated at the end of the culture.

Variations of μ, with respect to μ_{max}, estimated in real-time were highlighted by the calculation of *R* (Eq. 4) (Fig. 2B). Nevertheless, these variations of μ could be caused by physiological changes of cells and also signal instabilities. The signal instabilities were evaluated during the middle stage of exponential cell growth phase (between 40-70 h in this study). The average values and the

derivatives of *R* during this period were calculated and defined as the threshold, since μ was supposed to be constant during this phase; therefore, here the variations of *R* were mainly caused by signal instabilities. When both the average values and the derivatives of *R* exceeded the thresholds during at least 36 min (3 recorded dielectric signals), the critical moment when the cell growth begins to slow down due to cell physiological changes can be determined. Here, the critical moment was detected at 102 h (blue dotted line). It should be noticed that after the first critical moment has been identified, the culture remained "critical" and the cell physiological state could be altered as long as no interventions have been carried out.

These results demonstrated the potential of using on-line estimation of μ to have better understandings of animal cell culture bioprocesses, which could lead to a better process control by rapid interventions at the critical moment to maintain an appropriate cell physiological state.

3.3 Feed-harvest cell culture using on-line estimation of μ

To demonstrate the interest of using on-line estimations of μ for animal cell process control, the approach described previously in this study was used to control feed-harvest cell cultures. This culture mode could extend the duration of the cell culture, increase the antibody production level, and could also maintain an appropriate cells physiological state associated with fully glycosylated IgG production by restoring the nutrients and removing the cell wastes. In this study, the feedings were performed by punctual medium renewals of 2/3 total medium volume each time. Classically, medium renewals are usually carried out at fixed time points or by off-line measurements of nutrients (glucose, glutamine) with little considerations of cell physiological states. However, without a perfect knowledge of cell metabolism, it is extremely difficult to predict the exact moment when medium renewals are needed. Although using off-line measurements could help to understand cell metabolic behaviors, there would still be risks of missing the moment for efficient medium renewals, as off-line measurements are performed only 2-3 times per day and the kinetics of nutrients consumption could vary rapidly especially at high cell concentrations. Consequently, when medium renewals are not performed at the right time, there could be nutrients depletions, resulting in a decrease in cell energy production, and changes in cell physiological states, that may affect the culture process and the production of antibodies as well as their glycosylation.

Two feed-harvest cell cultures were carried out in this work using different strategies for medium renewals. The first cell culture using an off-line strategy was designed to simulate late medium

renewals after total glucose and glutamine depletion (Fig. 3A). For the second strategy, medium renewals were performed using real-time estimations of μ (Fig. 3B).

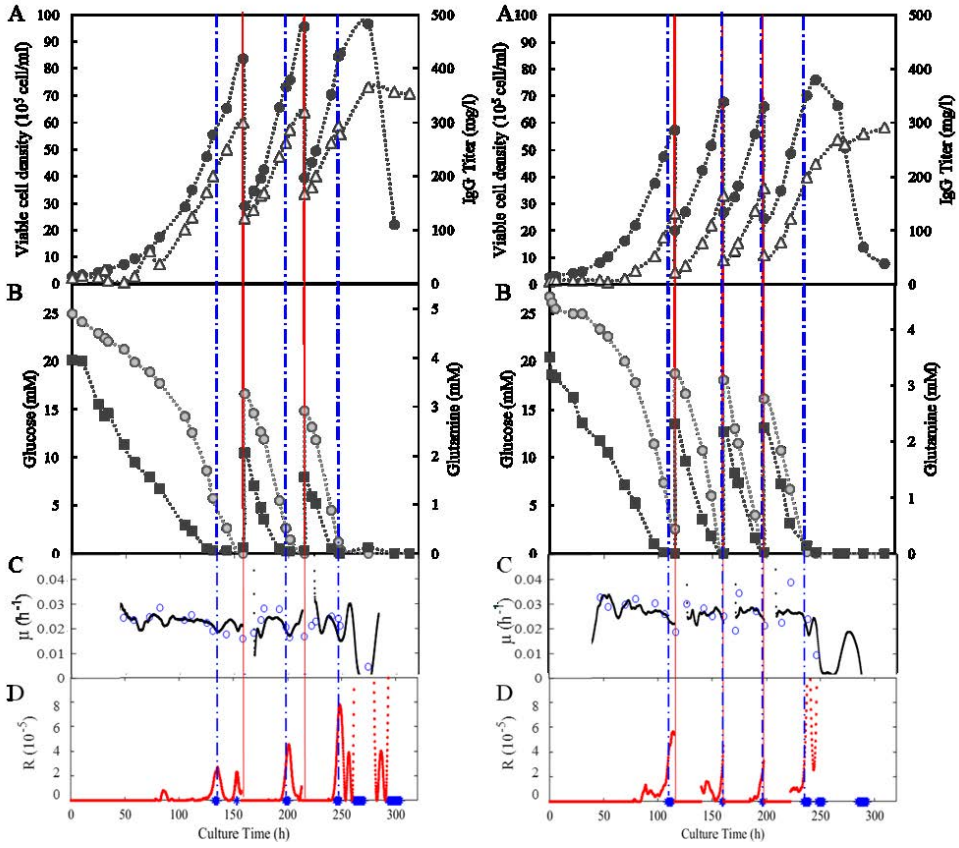


Figure 3. *Kinetics of feed-harvest cell cultures using off-line (left) and on-line (right) medium renewal strategies. A: viable cell density (●), IgG production (Δ); B: glucose (■), glutamine (■) consumption; C: On-line μ estimation (■), off-line μ calculated (■); D: On-line R calculation (■), critical moments detected (■). Blue dotted lines: critical moments; Red solid lines: moments of medium renewals.*

In the first cell culture process, two medium renewals were performed at 159 h and 215 h (red lines), when off-line measurements indicated that there was no more glucose, nor glutamine in the culture medium. However, on-line monitoring of μ and the calculation of R showed that the critical moments (blue dotted lines) were 130 h, 200 h and 250h of culture respectively. As indicated by the calculation of R, the value of μ began to decrease significantly after every critical moment, as nutrients depletions could cause non-sufficient energy production in cells to support the cell exponential growth, and an appropriate cell physiological state for fully glycosylated IgG production. Indeed, as showed in Fig. 4A, the σ_i of the cells increased after each critical moment, and decreased after the medium renewals, indicating that changes have appeared in cell physiological state for the cell culture using off-line strategy. Moreover, the non-glycosylated form of IgG increased dramatically

after the critical moments (Fig. 5A). Similar phenomenon has been observed in many other studies, where changes in protein glycosylation occurred at the end of the exponential cell growth phase (Castro *et al.*, 1995; Hooker *et al.*, 1995; Liu *et al.*, 2014). They suggested that the depletion in nutrients or in glycosylation precursors in the medium could be responsible for the change in the glycosylation pattern. However, another study suggested that protein glycosylation was not dependent on limitation in glucose, nor glutamine, but highly dependent on the energetic status of the cells (Kochanowski *et al.*, 2008). Therefore, when there were alterations in cell physiological state as indicated by the variations of σ_i, the non-glycosylated antibody level increased, although this level was reduced slowly after medium renewals.

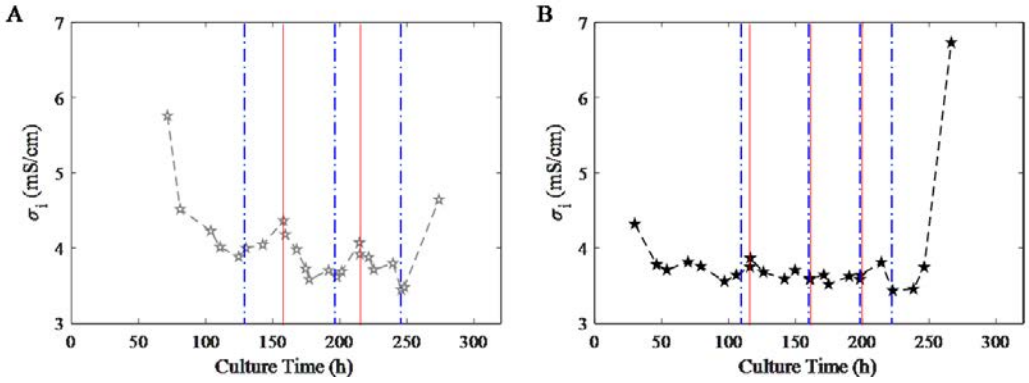


Figure 4. *Profile of cell intracellular conductivity (σ_i) in feed-harvest cell cultures using off-line (A) (☆) and on-line (B) (★) medium renewal strategies. Blue dotted lines indicate critical moments; Red solid lines indicate moments of medium renewals.*

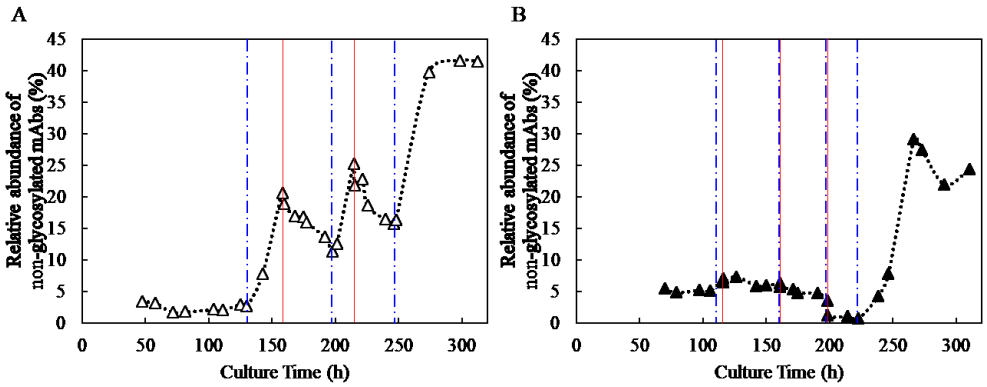


Figure 5. *Relative abundance of non-glycosylated IgGs produced in feed-harvest cell cultures using off-line (A) (△) and on-line (B) (▲) medium renewal strategies. Blue dotted lines indicate critical moments; Red solid lines indicate moments of medium renewals.*

In the second feed-harvest cell culture, medium renewals were performed at critical moments at about 110 h, 160 h and 200 h of the culture as indicated by the on-line estimation of μ and the on-line calculation of R (Fig. 3B). The decreases of μ were limited by timely performed medium renewals, and the value of μ remained nearly constant during the whole culture process before the final cell death phase. Consistently, a constant value of σ_i was maintained (Fig. 4B), indicating that cell physiological state remained stable, which led to better process performance regarding IgG glycosylation. As demonstrated in Fig. 5B, the level of non-glycosylated form of IgG produced was kept low (~ 5%) throughout the culture (before the final cell death phase) by using this on-line strategy.

Moreover, the critical moment would also be a good indicator for final culture harvesting, as it allows to harvest the cell culture before the accumulation of non-glycosylated antibodies due to the final cell death stage. Indeed, if the harvest is performed at the last critical moment, then the final level of the non-glycosylated antibodies produced during the process that used the on-line strategy was a lot lower (2% at 220 h) than that of the culture performed with the off-line strategy (16% at 245 h).

These results indicated that the on-line estimation of μ allowed to detect rapidly the critical moments, when the medium renewals must be performed in order to maintain an appropriate cell physiological state, and consequently to ensure the antibody glycosylation.

4 Conclusion

Cell specific growth rate (μ) is a key parameter for antibody producing animal cell culture processes, which reflects the cell physiological state. In order to optimize the antibody production, and especially, to ensure the quality of the final product, μ should be carefully monitored and controlled in real-time. In this study, a method of on-line estimation of μ was developed based on *in situ* dielectric spectroscopy measurements. As a result, the variations of μ were monitored in real-time using various mathematical methods to process these data, allowing a rapid detection of the critical moment when μ decreased significantly. This on-line detection of the critical moment was successfully implemented in feed-harvest cell cultures, and μ was maintained around its maximum value by performing medium renewals at critical moments. This on-line strategy allowed to maintain an appropriate cell physiological state. Consequently, a better antibody quality was obtained, as the non-glycosylated antibody level was kept low throughout the cell culture before the final cell death stage. When harvested at the last critical moment, the non-glycosylated antibody level was 8 times lower than to that obtained with the culture performed with off-line strategy. The method of on-line

estimation of μ shows great potentials for antibody production bioprocesses, and future efforts should be aimed at the implementation of this on-line method to fully automated feedback control scheme in animal cell culture processes for antibody quality control.

References *(pour alléger le document, les références complètes sont rassemblées à la fin du mémoire*

Abu-Absi, N. R. (2011)
Ansorge, S. (2007)
Ansorge S. (2009)
Cannizzaro, C. (2003)
Castro, P. M. L. (1995)
Cervera, A. E. (2009)
Cole, H. E. (2015)
Courtès, F. (2016)
Glacken, M. W. (1986)
Guez, J. S. (2004)
Hassell, T. (1991)
Hayter, P. M. (1993)
Hayter, P. M. (1991)
Henry, O. (2007)
Hooker, A. D. (1995)
Justice, C. (2011)
Kilburn, D. G. (1989)
Kiviharju, K. (2008)
Kochanowski, N. (2008)
Kondo, M. (2000)
Lao, M.S. (1997)
Lee, H. W. (2015)
Li, M.Y. (2018)
Liu, B. (2014)
Opel, C. F. (2010)
Quality Guidelines : ICH. (2017)
Tharmalingam, T. (2015)
Walsh, G. (2014)
Xie, L. (1996)
Xiong, Z.Q. (2015)
Yang, M. (2000)
Yardley, J. E. (2000)

I.3. Données supplémentaires

I.3.1. Précisions sur le calcul de la vitesse spécifique de croissance cellulaire en temps réel

Les valeurs de μ sont calculées en utilisant l'équation préalablement présentée (Equation 2 du manuscrit de publication ci-dessus), à partir des mesures en ligne de la permittivité réalisées tout au long de la culture. En pratique, ce calcul est réalisé sur des intervalles de temps Δt, avec une fenêtre glissante, en prenant la pente de la régression linéaire des valeurs logarithmiques de la concentration de cellules viables (*Ln(VCD)*) sur l'intervalle de temps Δt (Figure 42).

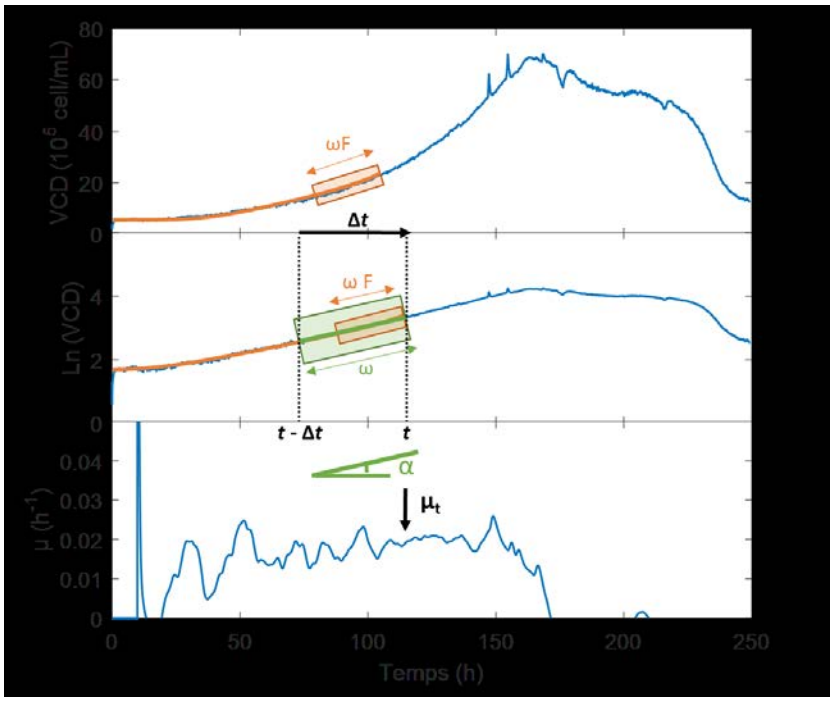


Figure 42 : Représentation schématique du calcul de μ et différentes fenêtres

Un premier défi pour le calcul de μ en temps réel consiste à choisir les filtres servant à lisser les données (*VCD* estimées en ligne et *Ln(VCD)*). Un compromis doit être trouvé pour assurer un lissage correct des données sans perdre les grandes tendances d'évolution. Trois types de filtres ont été testés dans cette étude : moyenne mobile (MA : moving average), Savitzky-Golay (SG), et smoothing spline (SP). L'algorithme de lissage le plus simple est le filtre MA. Il calcule la moyenne $\bar{y}_i$ d'un intervalle [*i-n, i+n*], puis cette intervalle glisse d'un point pour calculer la valeur de $\bar{y}_{i+1}$. L'intervalle a donc une largeur (ωF) de 2n+1. Concernant le filtre SG, il effectue une régression polynômiale locale de degré *x* sur un sous-ensemble de données en utilisant aussi une fenêtre glissante de largeur ωF. Le filtre SP lisse les données en fonction d'un paramètre spécifique de lissage *p*. Quand *p* tend vers 0, le filtre SP effectue une régression linéaire aux moindre carrés sur les données, tandis que quand *p* tend vers 1, il effectue une interpolation spline cubique (Wahba, 1990). La différence principale entre ces 3 filtres réside dans le fait que les filtres MA et SG fonctionnent sur des sous-ensembles de données dans une intervalle ωF, alors que le filtre SP utilise l'ensemble des données acquises. De ce fait, l'efficacité du filtre SP est fortement influencée par le nombre de données. Son lissage sera mauvais en début de culture avec peu de données, et trop fort en fin de culture (Figure 43 : courbe rouge). L'efficacité des filtres MA et SG est similaire (Figure 43 : courbes jaune et orange superposées). Pour la suite de l'étude, c'est le filtre MA qui a été choisi pour des raisons de simplicité, avec une ωF de 5 points (1 h de culture).

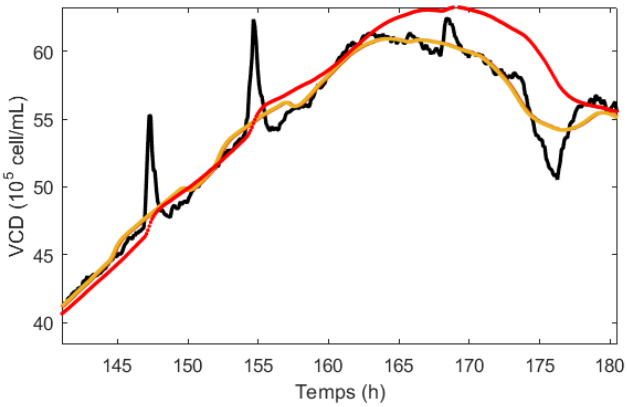


Figure 43 : Illustration de lissages effectués avec différents filtres. VCD non lissée (noire), VCD lissée par filtre SP (rouge), VCD lissée par filtre SG (jaune), VCD lissée par filtre MA (orange).

Après lissage, la valeur instantanée de μ à l'instant t étant calculée à partir des données lissées de *Ln(VCD)* de t-Δt à t, cette valeur est donc influencée par l'historique des *Ln(VCD)* dans la fenêtre. De ce fait, la largeur de la fenêtre de calcul (ω), correspondant au nombre de points dans l'intervalle de temps Δt, doit être bien choisie. Une largeur trop faible peut générer des instabilités au niveau des calculs, tandis qu'une largeur trop importante diminue la sensibilité du modèle de calibration et génère un retard sur l'obtention des résultats. La Figure 44 présente, dans le cas d'une culture en mode discontinu, l'influence de ω sur les valeurs de μ ainsi que sur la détection du moment critique de diminution significative de μ. Ici, 3 valeurs de ω (ω=10, 50, 100 points de mesure) sont évaluées. Comme observé dans la Figure 44, la ω de 10 points, correspondant à 2 h de culture, conduit à un calcul de μ très instable avec un niveau de bruit très important, ce qui rend quasi impossible le calcul du paramètre de distance résiduelle R servant à la détection du moment critique. Dans le cas d'un ω de 100 points (20 h de culture), les valeurs calculées de μ sont plus stables, mais le moment critique ne peut pas être détecté avant 128 h de culture (indicateurs verts sur l'abscisse). Enfin, la ω de 50 points (10 h de culture) permet une bonne stabilité des μ calculées avec un premier moment critique qui a pu être détecté dès 103 h de culture (indicateurs rouges sur l'abscisse). Ces résultats sont validés par les calculs à partir des mesures hors ligne (noir), puisque les premiers indicateurs rouges et noirs sont superposés. Dans l'article présenté dans la partie précédente (Chapitre I.2), la ω a donc été choisie égale à 50 points.

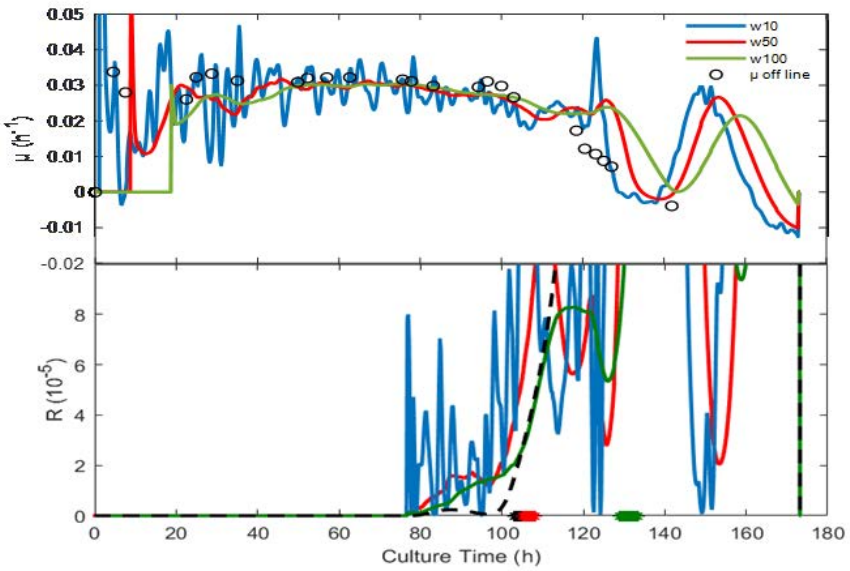


Figure 44: Influence de largeur de la fenêtre de calcul, à partir des données mesurées en ligne, sur les valeurs estimées de μ et sur la détection du moment critique. ω=10 (bleu) ; ω=50 (rouge) ; ω=100 (vert) ; calculs hors ligne (noir). Les moments critiques sont indiqués par l'apparition des premières étoiles sur l'abscisse.

I.3.2. Comparaison de cultures en mode recharge-récolte utilisant diverses stratégies de renouvellement du milieu de culture

Trois cultures différentes ont été réalisées en mode recharge-récolte (Figure 45). Dans la culture RR1, le milieu de culture a été renouvelé tardivement, plusieurs heures après la déplétion totale du glucose et/ou de la glutamine, composés analysés hors ligne. Pour la culture RR2, le changement de milieu a été basé sur le suivi en temps réel de μ à l'aide de la sonde diélectrique en ligne. Notons que les 2 cultures RR1 et RR2 ont déjà été présentées dans l'article qui précède cette partie. Enfin, la culture RR3 a mis en œuvre un changement de milieu anticipé, basé sur le moment où les concentrations de glucose et de glutamine atteignaient une valeur seuil respective de 5 mM ou 1 mM. Pour ces trois cultures en mode recharge-récolte, les valeurs hors ligne de μ ont ensuite été calculées en fin de culture. Comme recherché, les renouvellements de milieu dans la culture RR3 ont donc bien eu lieu avant l'apparition des moments critiques de diminution de μ (Figure 45). De ce fait, les cellules ont eu moins de temps pour se diviser avant d'être diluées par les renouvellements du milieu de culture. Les valeurs moyennes des VCD avant chaque échange de milieu sont donc plus faibles que dans les cultures RR1 (le tiers) et RR2 (la moitié).

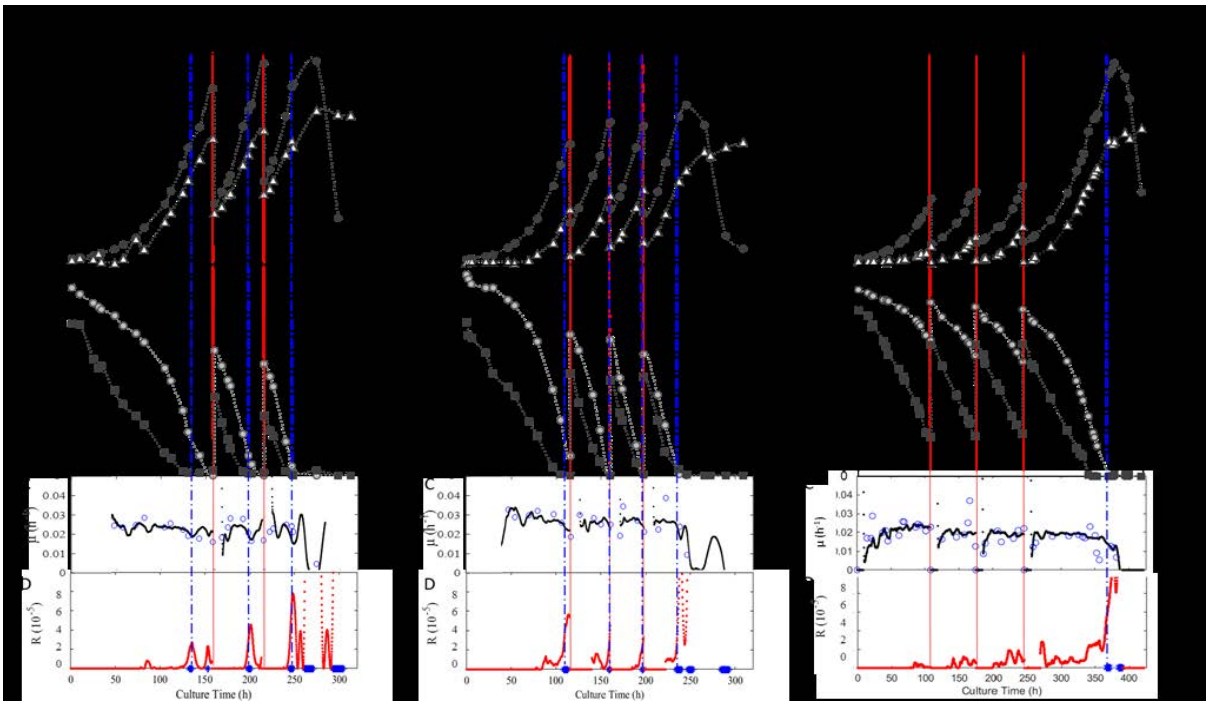


Figure 45 : Cultures en mode recharge-récolte pilotées par la stratégie : hors ligne en conditions tardives (RR1), en ligne (RR2), et hors ligne en conditions anticipées (RR3). A : VCD (●), AcM (Δ) ; B : glucose (■), glutamine (■) ; C : µ en ligne (■), off-line µ hors ligne (■) ; D : R en ligne (■), détection du moment critique (■). Ligne bleue pointillée : moment critique ; ligne rouge pleine : renouvellement du milieu.

La Figure 46 représente les évolutions des quantités cumulées d'IgG produites ainsi que celles du pourcentage d'IgG non-glycosylées au cours des trois cultures en mode recharge-récolte.

Concernant la quantité cumulée d'IgG, les cultures RR1 et RR2 permettent une production quasiment identique autour de 1 g après 310 h de culture. Par contre, la culture RR3 ne permet la production que de 0,74 g d'IgG au bout de 422 h de culture (Figure 46.A). La productivité spécifiqueglobale d'IgG, calculée en divisant la quantité cumulé d'IgG par la quantité cumulé de biomasse formée et par la durée de culture, est égale à 2,93 pg.cellule^{-1}.jour^{-1} pour RR1, 2,85 pg.cellule^{-1}.jour^{-1} pour RR2, et 1,72 mg.L^{-1}.jour^{-1} pour RR3.

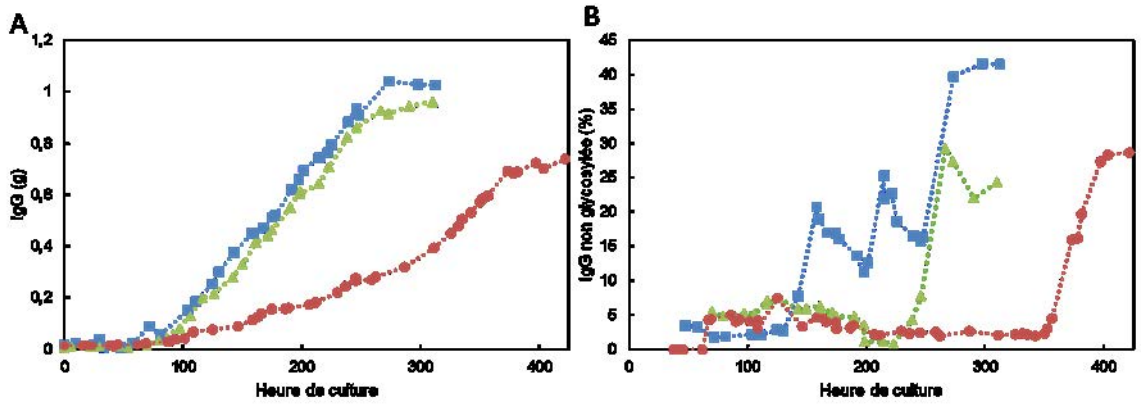


Figure 46: Quantité (A) et qualité (B) de l'IgG produite dans les cultures en recharge-récolte pilotées par la stratégie hors ligne en condition tardive (RR1 : bleu), par la stratégie en ligne (RR2 : vert), et par la stratégie hors ligne en condition anticipée (RR3 : rouge)

Concernant la qualité de glycosylation des IgG, le pourcentage d'IgG non-glycosylées reste très faible (< 5%) pendant toute la durée des cultures RR2 et RR3, pour n'augmenter que lorsque le renouvellement du milieu est interrompu en fin de culture et que la mort des cellules intervient. Le pourcentage maximal d'IgG non-glycosylées en fin de culture atteint alors 24% et 29% respectivement pour les cultures RR2 et RR3. Par contre, dans le cas de la culture RR1, ce pourcentage augmente brutalement après chaque instant critique, jusqu'à des valeurs de l'ordre de 25%. Ce niveau baisse ensuite un peu après chaque renouvellement du milieu jusqu'à environ 16%. Le pourcentage final d'IgG non-glycosylées après arrêt du renouvellement de milieu atteint 42% dans cette culture (Figure 46.B).

Les résultats hors ligne ou en ligne de μ permettent de confirmer les observations réalisées dans l'article précédent. En effet, dans le cas de la culture RR1, pilotée hors ligne avec des renouvellements tardifs du milieu, les augmentations du niveau d'IgG non-glycosylées coïncident avec les instants critiques de diminution significative de μ. Ce n'est que lors des changements de milieu, réalisés plusieurs dizaines d'heures après ces instants critiques, que les formes non-glycosylées diminuent à nouveau. Pour la culture RR3, les changements de milieu anticipés permettent de maintenir la valeur de μ relativement constante sans apparition d'instants critiques avant la fin de la culture. Ceci est bien corrélé avec le maintien d'une glycosylation maximale des IgG produites. Enfin, dans le cas de la culture RR2, le pilotage en ligne basé sur le maintien de la valeur de μ permet aussi de maintenir le pourcentage de formes non-glycosylées à leur niveau le plus bas. L'apport de milieu, qu'il soit réalisé par contrôle en ligne ou de manière anticipée, permet ainsi de répondre aux besoins énergétiques de la cellule et de ne pas remettre en cause la qualité de l'IgG

produite. Néanmoins, la quantité d'IgG produites est plus importante dans la culture RR2 avec le contrôle en ligne de μ comparativement à la culture RR3 avec l'apport du milieu anticipé.

Par ailleurs, l'estimation en ligne de μ présente l'avantage supplémentaire de pouvoir décider avec une meilleure précision du moment de la récolte des IgG afin de s'assurer de leur qualité de glycosylation maximale. Ainsi, dans la culture RR2, l'arrêt de la culture pourrait être réalisé à 238 h, soit juste avant le déclin cellulaire. La quantité d'IgG totale récoltée serait alors de 0,82 g, dont 0,79 g correctement glycosylés. Par contre, pour les cultures RR1 et RR3, la récolte avant déclin cellulaire devrait avoir lieu de manière plus empirique respectivement au bout de 274 h et 382 h, correspondant à une quantité totale d'IgG récoltée de 1,04 g et 0,69 g, dont 0,63 g et 0,55 g d'IgG glycosylées. Ainsi, la récolte des cultures RR1 et RR3 contiendrait un pourcentage non négligeable de formes non-glycosylées (40% et 20% respectivement) (Figure 46.B). En outre, dans le cas des stratégies de pilotage hors ligne (pour RR1 et RR3), l'identification des moments de renouvellement de milieu, basée sur les mesures hors ligne de glucose et glutamine, requièrent de nombreux et fréquents prélèvements, ce qui apparaît difficilement envisageable au cours d'un procédé industriel.

L'ensemble de ces résultats confirme l'intérêt de la stratégie de pilotage en ligne utilisant l'estimation de μ en temps réel, de manière à détecter rapidement les instants critiques de changement du milieu de culture en mode recharge-récolte. Cette stratégie, qui permet d'assurer un renouvellement de milieu ni trop tardif ni trop anticipé, assure une valeur de μ stable proche de sa valeur maximale, conduisant à une bonne quantité d'IgG avec un fort pourcentage de glycosylation.

I.4. Conclusion

Dans le cas des bioprocédés de culture de cellules animales, il apparaît essentiel de pouvoir suivre en temps réel la croissance cellulaire, puisque cette dernière conditionne la production des protéines recombinantes et ainsi les performances du procédé. La vitesse spécifique de croissance cellulaire représente en effet un bon indicateur de l'état physiologique des cellules. Sa diminution, du fait de conditions défavorables, indique une évolution de l'état physiologique des cellules qui peut affecter la production des anticorps recombinants et induire une altération au niveau de leur qualité. Il est ainsi très important de pouvoir suivre et contrôler les variations de μ au cours des procédés.

Dans ce chapitre, nous avons montré, dans un premier temps, la relation existant entre la variation de μ et la glycosylation de l'IgG produite au cours d'une culture réalisée en mode discontinu. Nous

avons constaté que lorsqu'une diminution de μ intervenait, les IgG non-glyosylées commençaient à s'accumuler pour atteindre plus de 30% des IgG totales en fin de culture, ce qui représente une baisse significative de leur qualité. Ces premiers résultats confirment bien l'importance de suivre et de contrôler en temps réel les variations de μ afin d'assurer la meilleure qualité possible de l'IgG produite.

Dans un second temps, un modèle a été développé pour estimer μ à partir des mesures de permittivité en temps réel réalisées en ligne par spectroscopie diélectrique. Après évaluation de la concentration des cellules viables, les valeurs de μ ont été calculées, moyennant une optimisation des paramètres de lissage et de calcul. Le calcul en temps réel du paramètre de distance résiduelle *R*, représentatif de la variation de μ, a alors permis de déterminer l'instant critique où la variation, issue du changement de l'état physiologique des cellules, dépassait la variation provoquée par les bruits intrinsèques à l'appareil.

Puis, le modèle développé a été implémenté dans une stratégie de pilotage d'une culture en mode recharge-récolte. Les renouvellements du milieu de culture ont été effectués aux instants critiques identifiés par le modèle d'estimation de μ, permettant ainsi de maintenir la valeur de μ proche de sa valeur maximale. Les performances de cette culture, tant en termes de quantité que de qualité des IgG produites, ont été comparées à celles de deux cultures utilisant des stratégies de pilotage hors ligne, pour lesquelles les renouvellements du milieu étaient déterminés par des mesures hors ligne des concentrations de glucose et de glutamine. La première de ces cultures simulait des conditions de recharge-récolte tardives après déplétion totale du glucose et de la glutamine. La deuxième culture pilotée hors ligne simulait des conditions de renouvellements de milieu anticipés au moment où la concentration des nutriments atteignait une valeur minimale seuil. L'utilisation de la stratégie de pilotage en ligne a permis de réaliser les renouvellements de milieu, ni de façon trop anticipée, ni trop tardivement, permettant ainsi de maintenir la valeur de μ stable. Comparée avec les stratégies de pilotage hors ligne, cette stratégie en ligne a présenté l'avantage d'assurer à la fois la quantité et la qualité des IgG produites, avec le niveau le plus élevé d'IgG glycosylées au moment de la récolte. De plus, cette stratégie en ligne ne nécessite pas de prélèvements fréquents, minimisant ainsi les besoins en personnels et les risques de contamination microbienne.

En conclusion, les résultats de ce chapitre présentent le potentiel du suivi en temps réel de l'état physiologique des cellules grâce à l'estimation de μ par spectroscopie diélectrique *in situ*. Ils peuvent être considérés comme une première étape vers le développement de systèmes de rétrocontrôle automatisés. Cette approche pourrait également être envisagée pour le suivi d'autres modes de

culture comme les modes semi-continus ou perfusés. Les principaux résultats de ce chapitre sont présentés de manière synthétique dans la Figure 47.

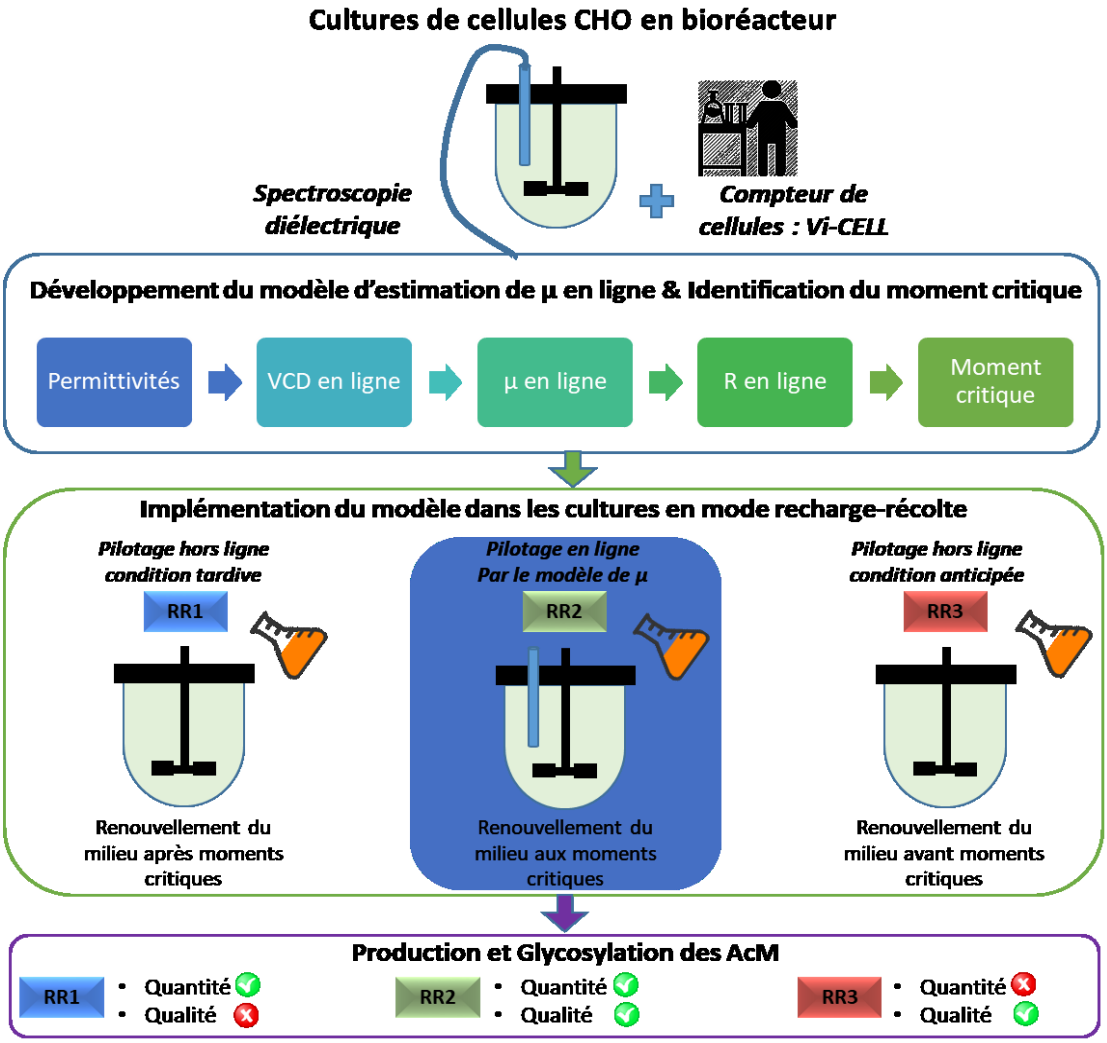


Figure 47 : Synthèse schématique des résultats du chapitre I

Chapitre II

Comparaison et combinaison des spectroscopies Raman et NIR pour le suivi en ligne de procédés de culture de cellules animales

II.1. Introduction 121

II.2. Etude comparative des spectroscopies Raman et NIR pour le suivi en ligne de procédés de culture de cellules animales 121

II.2.1. Introduction 122

II.2.2. Manuscrit de la publication 123

II.2.3. Principales contributions de la publication 138

II.3. Stratégies de combinaison des données spectrales NIR et Raman 139

II.3.1. Fusion des données de niveau bas 139

II.3.2. Fusion des données de niveau intermédiaire 141

II.3.3. Fusion des données de niveau haut 146

II.4. Conclusion 149

II.1. Introduction

Le suivi en ligne des paramètres fonctionnels du procédé est devenu incontournable dans le cas des procédés de culture de cellules animales, puisqu'il permet de mieux comprendre et contrôler ces procédés. Néanmoins, seuls quelques paramètres basiques (*e.g.* température, O_2 dissous, pH) sont actuellement mesurés en ligne. La plupart des mesures biochimiques sont, quant à elles, obtenues à partir d'analyses hors ligne nécessitant une prise d'échantillon. Le temps de retard entre l'échantillonnage et l'analyse est un inconvénient majeur pour contrôler et piloter le procédé de culture. Dans ce contexte, les techniques spectroscopiques, en particulier les spectroscopies Raman et NIR, offrent des avantages significatifs : analyse *in situ*, non-invasive, sensible, multiparamétrique et en temps réel. Bien que ces deux types de spectroscopies soient tous deux prometteurs pour le suivi en ligne des bioprocédés de culture de cellules animales, ils sont basés sur différents principes fondamentaux, chacun possédant ses propres avantages et inconvénients. Il n'existe pas de consensus clair quant à la technique spectroscopique la plus adaptée à ces bioprocédés. En effet, les méthodes de comparaison ne sont pas bien établies et les deux spectroscopies sont souvent utilisées séparément, ce qui rend impossible de les comparer directement, puisque les conditions de culture sont différentes. De plus, une autre approche intéressante consisterait à combiner ces deux spectroscopies, puisqu'elles fournissent des informations complémentaires sur les échantillons analysés. Cependant, il n'existe pas de stratégie universelle pour la combinaison des données spectrales issues de deux technologies spectroscopiques différentes.

Ainsi, l'objectif de ce chapitre est de développer, non seulement, une méthodologie de comparaison, mais aussi, une stratégie de combinaison des spectroscopies Raman et NIR pour le suivi *in situ* et en temps réel de procédés de culture de cellules CHO. Dans un premier temps, des modèles chimiométriques ont été développés individuellement à partir des spectres NIR et Raman acquis *in situ*, pour le suivi en ligne de la concentration des cellules vivantes, du glucose, du lactate, des anticorps IgG, de la glutamine et des ions ammonium simultanément. Nous avons ensuite comparé la performance des modèles de prédiction utilisant l'une ou l'autre des deux spectroscopies à l'aide de facteurs de mérite (Figure of Merit, FOM) multivariés. Ces résultats sont présentés sous la forme d'une publication. Nous avons enfin évalué l'intérêt de combiner les spectres NIR et Raman obtenus en ligne en utilisant différentes techniques de fusion de données.

II.2. Etude comparative des spectroscopies Raman et NIR pour le suivi en ligne de procédés de culture de cellules animales

II.2.1. Introduction

La spectroscopie NIR est une technique bien établie pour les analyses quantitatives et qualitatives dans les domaines de la fermentation et des cultures de cellules animales. Cependant, elle est grandement influencée par la présence des molécules d'eau et le chevauchement de larges bandes (Lourenço *et al.*, 2012). Bien que la spectroscopie Raman n'ait été utilisée pour la première fois qu'en 2011 pour le suivi en ligne de procédés de culture de cellules animales, son application a progressé très rapidement depuis. L'avantage principal de cette technique réside dans sa moindre sensibilité aux molécules d'eau, favorisant ainsi les analyses en milieu aqueux. Par contre, les molécules fluorescentes peuvent générer beaucoup d'interférences spectrales (Esmonde-White *et al.*, 2017). Bien que présentant des avantages et des inconvénients différents, ces deux types de spectroscopies restent utilisés séparément pour le suivi de bioprocédés. Il reste donc difficile de les comparer objectivement à partir du moment où les conditions opératoires de culture et les méthodes d'analyse diffèrent.

Dans l'article présenté ci-dessous, nous proposons de comparer en parallèle les performances de ces deux spectroscopies en les implémentant simultanément dans les mêmes cultures de cellules CHO en bioréacteur. Différents modèles chimiométriques ont été développés en utilisant la méthode de régression PLS, à partir des spectres Raman et NIR pour prédire les concentrations des cellules vivantes, du glucose, du lactate, des IgG, de la glutamine, et des ions ammonium. Puis, la performance des modèles a été caractérisée et comparée à l'aide de facteurs de mérites (FOM) multivariés, à savoir : l'exactitude (accuracy), la linéarité, la précision, la sensibilité analytique, la sélectivité, le ratio signal sur bruit, la limite de détection, le ratio de gamme des erreurs (RER : range error ratio) et le ratio de déviation de la performance (RPD : ratio of performance deviation). De plus, d'autres techniques chimiométriques, comme la PLS de type 2, et un premier essai de combinaison des données spectrales ont été aussi testés. Enfin, les deux spectroscopies ont été utilisées et comparées pour le suivi d'une culture en mode discontinu, en vue d'étudier leur capacité de suivi en temps réel des différents paramètres cités précédemment.

Ce point a fait l'objet d'un article scientifique publié dans la revue "Biochemical Engineering Journal" (137, 205-213, 2018). Pour faciliter la lecture, le manuscrit de cette publication est présenté ci-dessous.

II.2.2. Manuscrit de la publication

Parallel comparison of *in situ* Raman and NIR spectroscopies to simultaneously measure multiple variables toward real-time monitoring of CHO cell bioreactor cultures

Meng-Yao Li, Bruno Ebel, Fabien Chauchard, Emmanuel Guédon, Annie Marc

Abstract

For the first time, the ability of *in situ* Raman and Near Infrared (NIR) spectroscopies to predict in real-time the concentration of viable Chinese Hamster Ovary (CHO) cells, glucose, lactate, glutamine, ammonium and antibodies in bioreactors has been compared in parallel. PLS models were developed using Raman and NIR spectra recorded at same time points and in same cultures. Multivariate figures of merit were established to thoroughly compare the performances of models. Raman spectroscopy was found to better predict concentrations of glucose, lactate and antibody, while NIR spectroscopy predicted better glutamine and ammonium ion concentrations. By applying the models developed on spectra recorded continuously during a batch culture, both spectroscopies were able to monitor simultaneously the general trends of the concentration of different analytes, while Raman predictions were more accurate at the beginning of the culture. Moreover, other chemometric methods (partial least square regression of type 2 and spectral combination) were tested, but leading to only limited model performance improvements. Both spectroscopies are suitable tools for real-time monitoring of cell culture bioprocesses, while Raman spectroscopy provides several advantages concerning sensibility, selectivity and limit of detection. These results provide useful information for academic researchers and pharmaceutical industries to choose the most adequate spectroscopic technology for their cell culture bioprocesses.

1. Introduction

Mammalian cell culture processes, which are widely used for the production of biotherapeutic proteins, require a better control due to high variabilities of cultivation procedures and biological materials [1]. The process analytical technology (PAT) initiative encourages the use of innovative tools to monitor critical process parameters and ensure good process performances and consistent

product qualities. In the last ten years, vibrational spectroscopies such as near infrared (NIR) and Raman have gained great attention with the development of real-time monitoring techniques for bioprocesses, as they provide rapid and simultaneous information on multiple variables in a non-destructive way. Thanks to fiber optic probes that can be directly immersed in the culture broth, these spectroscopic techniques are capable to monitor *in situ* key-analytes without any sample withdraws, which may largely lower the contamination risks and the labor burden [2]. Recent studies have reported the efficiency of both NIR and Raman spectroscopies to accurately predict concentrations of different medium compounds during mammalian cell culture processes [3-7]. Since cell culture media contain a large variety of molecules leading to complex spectral features, multivariate data analysis tools are needed to extract the desired information on one specific medium component by calibrating on-line spectra with off-line reference analyses [2]. Nevertheless, to date, it remains difficult to objectively compare the advantages and limits of each of these techniques as NIR spectroscopy is based on optical absorption and Raman spectroscopy on the measurement of inelastic scattering. NIR spectroscopy is reported to be sensitive to water whose high signal intensity may mask important biochemical information in the spectrum [8]. As for Raman spectroscopy, although it is less sensitive to water, the fluorescence of biological compounds can be more intense than Raman scattering, reducing its sensibility of detection [9]. Two very recent papers demonstrated the comparison of these spectroscopic methods, either by off-line analysis of glucose, lactate and ammonium ion concentrations of cell cultures in 96-well plates [8], or by comparing the scalability of glucose concentration models developed in shake flask cultures [10]. However, to our knowledge, no study has already performed the complete comparison of calibration models developed for multiple variables by using *in situ* NIR and Raman probes implemented in parallel in the same cell culture bioreactors. This represents the real application condition for on-line monitoring of key culture variables by spectroscopy techniques. Despite a number of studies have reported separately the use of NIR or Raman spectroscopy in bioprocess monitoring, a proper comparison of their performances remains difficult, since the reported results come from separate bioreactor studies using various operating conditions. To give a more objective view and facilitate the choice of the most suitable technology according to the process needs, this study investigates NIR and Raman spectroscopies in parallel using *in situ* probes to develop models to predict in real-time the concentrations of viable cells, glucose, glutamine, lactate, ammonium ions and recombinant antibodies during CHO cell cultures performed in bioreactor. Then the models performances are characterized and compared by multivariate figures of merit (FOM). In addition, different chemometric methods are tested including classical partial least squares (PLS) regression, type 2 PLS regression and spectra combination strategy, in order to extract more information from the spectra data. Finally, spectroscopic models developed from both spectroscopies are applied to unknown

spectra recorded in a bioreactor cell culture to simulate the real-time monitoring of multiple cell culture variables.

2. Materials and Methods

2.1. Cell cultivation

The DG44 CHO cell line that produces an anti-Rhesus D antibody (CHO M250-9, kindly provided by Bioprocessing Technology Institute of Singapore) was cultured in a protein-free medium consisting of a 1:1 volume ratio of PF-CHO (GE Healthcare) and CD-CHO (Fisher Scientific) supplemented with 4 mM L-glutamine and 0.1% Pluronic F-68 (Sigma Aldrich). Cell cultures were performed in a 2.5 L bioreactor (Pierre Guérin) using a homemade light shield. Cultures were carried out at 37 °C, pH 7.2, 50% of air saturation and with a 90 rpm axial agitation. Six cultures were performed in batch, feed-harvest and punctual fed-batch (spiking) modes to increase cell environment variability and led to better robustness for spectroscopic model calibration. In particular, the culture with glucose spiking allowed to break the correlation between glucose and lactate, which is a common problem for spectroscopic model development.

2.2. Reference analyses

Culture samples were taken 3-4 times per day. Viable cell density (VCD) was measured using a Vi-CELL™ cell counter (Beckman Coulter). An automated photometric analyzer Gallery™ (Thermo Fisher Scientific) was used to measure the concentration of glucose, lactate, glutamine, ammonium ions and antibody.

2.3. *In situ* spectroscopies

Each culture was monitored in parallel by both Raman and NIR spectroscopies, with beforehand-optimized operating conditions (acquisition time, number of scans, working spectral region, etc.). The BioViserion® Raman spectrometer (INDATECH model type VSR-HR-1-785) was equipped with a 785 nm laser source linked by optic fibers to a stainless steel immersion probe (model type SPC-OH-785-2). An acquisition time of 6 min including 6 scans of 60 s was adopted. The working spectral region 800-1600 cm^{-1} was selected, while data in 1392-1422 cm^{-1} range were removed due to light interference. The NIR transflectance probe with a path length of 1 mm (Precision Sensing Devices) was connected by optic fibers to the Antaris II spectrometer (Thermo Scientific). Each NIR spectra corresponded to an average of 128 scans acquired with an 8 cm^{-1} resolution, from 4000 to 10000 cm^{-1} wavelength. The optimal working region was selected in 4300 - 4752 cm^{-1} and 5338 - 9000 cm^{-1} ranges.

2.4. Chemometric methods

Spectral data were separated into calibration (80%) and prediction (20%) datasets by Kennard-Stone algorithm. They were pre-processed by standard normal variates (SNV) normalization and mean centering, before performing model calibration using partial least squares (PLS1) regression. To minimize overfitting, a minimal latent variable (LV) was retained with a maximum explained variance for cross-validation using a Venetian blinds approach. PLS regression of type 2 (PLS2) was also tested, by calibrating the spectral data with off-line values of 2 culture variables simultaneously (glucose and lactate concentrations, glutamine and ammonium ion concentrations, VCD and antibody concentrations). Models were also developed based on combined Raman and NIR spectra within their individual working spectra regions after preprocessing steps.

2.5. Multivariate Figures of Merits

The model performances were evaluated by various statistic parameters including accuracy, linearity, precision, analytical sensitivity (γ), selectivity (SEL), signal to noise ratio (S/N), limit of detection (LOD), ratio of performance deviation (RPD), and range error ratio (RER).

Accuracy. According to International Conference on Harmonization (ICH) guidelines, accuracy describes the closeness of agreement between reference and predicted values [11]. Therefore, the root mean square error

(RMSE) for calibration (RMSEC), cross-validation (RMSECV) and prediction (RMSEP) are used to express the accuracy of each model developed. The RMSE is calculated as follows:

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

where y_i is the measured concentration, $\hat{y}_i$ the predicted concentration and n the number of sample in the dataset under consideration. A high accuracy can be obtained with low RMSE. RMSEP of models for each variable obtained from different spectroscopies were compared by using the *F*-test, which reveals if there is any significant difference between them [12]. A p-value inferior than 0.05 indicates a significant difference.

Linearity. The linearity of the models is usually evaluated by the correlation coefficient (R^2) of the prediction. High values of R^2 signify high linearity.

Precision. As accuracy is expressed as two components: trueness and precision, precision can be calculated as the part of accuracy without trueness [13]. Trueness is normally expressed in terms of bias, and precision is currently calculated over validation samples. Therefore, precision can be estimated with bias corrected MSEP (BCMSEP) [12]. A low BCMSEP corresponds to higher model precision.

$$BCMSEP = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 - [\sum_{i=1}^{n}(y_i - \hat{y}_i)]^2/n}{n-1} \quad (2)$$

Analytical sensitivity: Analytical sensitivity is a modified form of sensitivity, which characterizes the extent of signal variation due to changes in analyte's concentration. The sensitivity was calculated by using the multivariate Net Analyte Signal (NAS) analysis developed by Lorber *et al.* [14]. NAS is defined as a vector orthogonal to the interfering elements present within the spectra, and is therefore uniquely related to the analyte of interest. The NAS vector can be estimated by using following formula:

$$NAS_i = (x_i \cdot b) \cdot (b^T \cdot b)^{-1} \cdot b^T \quad (3)$$

where x_i is a sample spectrum after preprocessing, and b is a column vector of the PLS regression coefficient. Then, sensitivity (SEN) can be calculated as [15]:

$$SEN_i = \frac{NAS_i}{y_i} \quad (4)$$

where y_i is the measured concentration for the *i* sample. Sensitivity is reported in this study as the mean of the univariate sensitivity values of all samples under consideration. However, sensitivity is not applicable to models developed from two spectroscopies operating under different fundamental principles, because it incorporates the instrument's signal unit. The parameter γ is therefore proposed to provide an impartial assessment of different analytical techniques and is calculated as follows [15]:

$$\gamma = \frac{||SEN||}{\delta r} \quad (5)$$

where sensitivity is expressed by its univariate form by taking the Euclidean norm of the sensitivity vector, and δr is an estimate for instrumental noise, calculated as the standard deviation of the preprocessed replicate spectra of blank samples. A larger γ value corresponds to a higher sensitivity.

Selectivity. Selectivity (SEL) is a dimensionless measure of the portion of instrument signal that is not influenced by interfering elements and leading to spectra overlap. Its value is restricted between 0 and 1, and can be calculated as [16]:

$$SEL_i = \frac{||NAS_i||}{||x_i||} \quad (6)$$

with the NAS_i vector and the preprocessed spectrum x_i of the ith sample. SEL was reported in this study as the mean of all selectivity values for samples under consideration. A higher value of SEL will correspond to a more selective model.

Signal to noise ratio. S/N ratio is another useful term to evaluate the performance of an instrumental analysis, and is calculated as follows [14]:

$$S/N_i = \frac{||NAS_i||}{\delta r} \quad (7)$$

The S/N ratio is reported in this article as the mean of the S/N values of all samples under consideration.

Limit of detection. The LOD can be defined as a minimum detectable concentration for an analytical method in a given sample matrix, and can be calculated as follows [17]:

$$LOD = 3.3\delta r \frac{1}{||SEN||} \quad (8)$$

Ratio of performance deviation. The RPD represents the ratio of standard deviation for the reference values of the prediction dataset to the RMSEP. Higher values of RPD are desired to obtain a good model performance.

Range error Ratio. The RER is defined as the ratio of the range in the reference values from the prediction dataset to the RMSEP. Higher values of RER represents better model prediction capability.

Matlab 2016a (Mathworks, Inc.) and PLS Toolbox 8.1 (Eigenvector Research, Inc.) were used for model development and figures of merit calculations.

3. Results and discussion

3.1. Comparative evaluation of PLS1 models for Raman and NIR spectroscopies

After removing spectra outliers with Hotelling's T^2 test and Q residuals test within a 95% confidence interval, 92 samples were used for model calibration, while 23 others were tested as prediction samples. The same calibration/prediction datasets were used for both spectroscopies. PLS1 models of both Raman and NIR spectroscopies were developed to predict 6 culture variables: viable cell density (VCD), glucose, glutamine, antibody, lactate and ammonium ion concentrations. Their measured versus predicted values are presented in Fig. 1.

3.1.1 Accuracy and linearity of the PLS1 models

Tables 1 to 6 display the statistics describing the accuracy and the linearity of the models developed from both spectroscopies. Both Raman and NIR spectroscopies gave good predictions for VCD with high linearity (R^2> 0.97) and high accuracy showed by low level of related errors (RMSEC, RMSECV and RMSEP) representing less than 10% of the calibration range (Table 1).

Table 1: Comparison of linearity and accuracy of models developed for viable cell concentration prediction

VCD	LV	R^2_{Cal}	RMSEC	RMSEC	R^2_{CV}	RMSECV	RMSECV	R^2_P	RMSEP	RMSEP
(10^5 cell/ml)				%			%			%
Raman PLS1	5	0.97	3.95	4	0.89	8.24	9	0.94	5.31	6

Raman PLS2	5	0.97	3.95	4	0.89	8.24	9	0.94	5.31	6
NIR PLS1	5	0.98	3.76	4	0.96	5.16	6	0.93	6.12	7
NIR PLS2	3	0.9	7.82	9	0.89	8.3	9	0.90	6.92	8
Combination	5	0.98	4.04	4	0.93	6.68	7	0.93	6.1	7

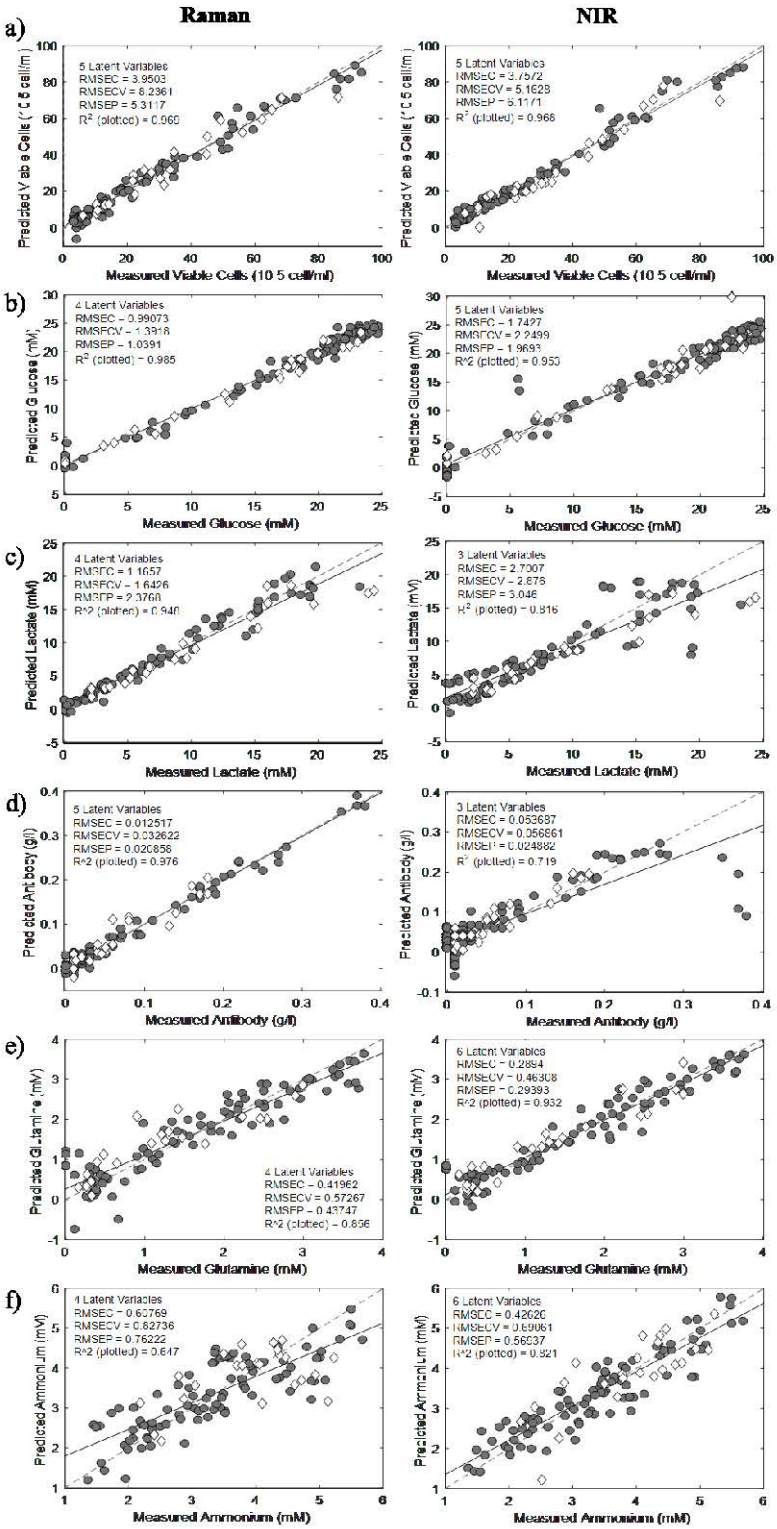


Figure 1: Raman (left column) and NIR (right column) PLS1 calibration models for concentrations of viable cells (a), glucose (b), lactate (c), antibody (d), glutamine (e), and ammonium ions (f). Solid lines: regression lines between measured and predicted values. Dashed lines: target 1:1 lines. Circles: calibration dataset. Diamonds: prediction dataset.

There was no statistical difference between accuracy of the two VCD models according to the F-test (Table 7). While glucose concentration was better predicted by Raman spectroscopy, as Raman model displayed a higher linearity and accuracy (R^2 = 0.99 and 4%, 6%, 4% for RMSEC, RMSECV and RMSEP respectively), the RMSEP being nearly 50% less than with NIR model (Table 2). NIR model largely overestimated glucose concentration of 3 samples (2 from calibration set, 1 from prediction set) of more than 6 mM, while Raman model correctly predicted these concentrations (Fig.1, b). Therefore, it can be suggested that this observation was more related to the intrinsic noises of NIR spectra instead of errors in reference values.

In the case of the lactate concentration prediction, a deviation between the model regression line and the 1:1 target line was observed with both spectroscopies, with a lower prediction accuracy when lactate concentration were higher than 10 mM (Fig.1, c). Similar observation was described by Mehdizadeh *et al.* [5], who pointed out that the model prediction could be affected by a lower number of samples in the upper end of the concentration range. Raman model appeared to give better predictions for lactate concentrations (RMSEC: 7%, RMSECV: 7%, RMSEP: 11%, R^2: 0.93), compared to the NIR model (RMSEC: 11%, RMSECV: 12%, RMSEP: 12%, R^2: 0.82) (Table 3). However, according to the F-test, there was no significant difference between accuracy of these 2 models (Table 7).

Table 2: Comparison of linearity and accuracy of models developed for glucose concentration prediction

Glucose			RMSEC			RMSECV			RMSEP	
	LV	R^2_{Cal}	RMSEC		R^2_{CV}	RMSECV		R^2_P	RMSEP	
(mM)				%			%			%
Raman PLS1	4	0.99	0.99	4	0.97	1.39	6	0.98	1.04	4
Raman PLS2	5	0.99	0.89	4	0.98	1.34	5	0.98	1	4
NIR PLS1	5	0.96	1.74	7	0.93	2.25	9	0.93	1.97	8
NIR PLS2	7	0.98	1.24	5	0.93	2.17	9	0.97	1.28	5
Combination	4	0.99	1.06	4	0.97	1.55	6	0.98	1.18	5

Table 3: Comparison of linearity and accuracy of models developed for lactate concentration prediction

Lactate (mM)	LV	R^2_{Cal}	RMSEC	RMSEC %	R^2_{CV}	RMSECV	RMSECV %	R^2_P	RMSEP	RMSEP %
Raman PLS1	3	0.94	1.52	7	0.93	1.68	7	0.90	2.51	11
Raman PLS2	5	0.97	1.06	5	0.93	1.62	7	0.92	2.38	10
NIR PLS1	3	0.81	2.7	11	0.78	2.88	12	0.90	3.05	12
NIR PLS2	7	0.97	1.02	4	0.81	2.69	12	0.81	3.25	14
Combination	5	0.98	0.95	4	0.94	1.58	7	0.93	2.36	10

Table 4: Comparison of linearity and accuracy of models developed for antibody concentration prediction

Antibody (g/L)	LV	R^2_{Cal}	RMSEC	RMSEC %	R^2_{CV}	RMSECV	RMSECV %	R^2_P	RMSEP	RMSEP %
Raman PLS1	5	0.98	0.01	3	0.89	0.03	9	0.89	0.02	5
Raman PLS2	5	0.89	0.03	9	0.87	0.04	9	0.87	0.02	6
NIR PLS1	3	0.71	0.05	14	0.67	0.06	15	0.90	0.02	7
NIR PLS2	3	0.7	0.05	14	0.67	0.06	15	0.91	0.02	6
Combination	5	0.97	0.02	4	0.92	0.03	8	0.92	0.02	5

Concerning the prediction of antibody concentrations, Raman spectroscopy was more accurate with significantly lower errors (R^2: 0.98, RMSEC: 3%, RMSECV: 9%, RMSEP: 5%) compared to those obtained with NIR spectroscopy (R^2: 0.7, RMSEC: 14%, RMSECV: 15%, RMSEP: 7%) (Table 4). Moreover, antibody concentration was predicted accurately by the Raman model over the whole calibration range, while large deviations were observed with the NIR model for the upper end of antibody concentrations (Fig.1, d). Thus, Raman spectroscopy appears to have an advantage over NIR for the prediction of glucose, lactate and antibody concentrations, while the related errors for these models developed from NIR spectra remained in an acceptable range.

Table 5: Comparison of linearity and accuracy of models developed for glutamine concentration prediction

Glutamine (mM)	LV	R^2_{Cal}	RMSEC	RMSEC %	R^2_{CV}	RMSECV	RMSECV %	R^2_P	RMSEP	RMSEP %
Raman PLS1	4	0.87	0.42	11	0.75	0.57	15	0.81	0.44	12
Raman PLS2	4	0.83	0.47	13	0.75	0.57	15	0.84	0.42	11
NIR PLS1	6	0.94	0.29	8	0.84	0.46	12	0.91	0.29	8
NIR PLS2	7	0.92	0.33	9	0.84	0.45	12	0.94	0.25	7
Combination	4	0.85	0.44	12	0.77	0.55	15	0.86	0.39	10

Regarding the prediction of glutamine and ammonium ion concentrations (Fig.1, e and f), both spectroscopies led to low accuracy compared to previous well-predicted variables, with RMSEP values of 12% and 18% for Raman models, and of 8% and 13% for NIR models (Tables 5 and 6). As the model error depends on both error of reference analyses and error of spectral data, the model with lower performance may be caused by larger errors in the off-line analyses of glutamine and ammonium ions. Furthermore, variations of glutamine and ammonium ion concentrations during cell cultures were relatively low, leading for example to a maximal range of 0-5 mM for glutamine instead of 0-25 mM for glucose. Therefore, it is likely that the calibration models of these two analytes were relying on correlations to other compounds rather than the direct detections of these two molecules [18].

Table 6: Comparison of linearity and accuracy of models developed for ammonium concentration prediction

Ammonium (mM)	LV	R^2_{Cal}	RMSEC	RMSEC %	R^2_{CV}	RMSECV	RMSECV %	R^2_P	RMSEP	RMSEP %
Raman PLS1	4	0.69	0.61	14	0.44	0.83	19	0.29	0.76	18
Raman PLS2	4	0.66	0.64	15	0.44	0.82	19	0.28	0.77	18
NIR PLS1	6	0.85	0.43	10	0.61	0.69	16	0.63	0.57	13
NIR PLS2	7	0.89	0.36	8	0.64	0.66	15	0.73	0.45	10
Combination	4	0.66	0.65	15	0.47	0.81	19	0.58	0.59	14

3.1.2. Additional FOM of the PLS1 models

Although the accuracy and linearity data provide a first mean of comparing these 2 spectroscopy methods, it remains difficult to give direct conclusions for the choice of instrument. Thereby, additional FOM could be useful to complete the comparison of the 2 spectroscopies (Table 7).

Table 7: Additional figures of merit for PLS1 models developed from Raman and NIR spectra (units of BCMSEP and LOD are that used for each variable, and other parameters are dimensionless)

Variable	Technique	F-test (p-value)	BCMSEP	Γ	SEL	S/N	LOD	RER	RPD
VCD	Raman	0.23	29.49	4.16	0.65	12.60	0.79	17	4.7
	NIR		34.37	0.47	0.09	8.37	1.49	14.8	4.1
Glucose	Raman	<0.001	1.05	7.16	0.72	14.36	0.46	23.8	8.1
	NIR		3.98	3.24	0.10	10.03	1.01	12.6	4.3

Lactate	Raman	0.018	5.24	16.86	0.74	14.61	0.20	9.3	2.5
	NIR		7.66	1.33	0.28	29.65	0.28	8.3	2
Antibody	Raman	0.37	4.43E-04	356.15	0.65	12.74	0.01	18.2	4.8
	NIR		4.07E-04	104.20	0.24	24.29	0.01	15.3	4
Glutamine	Raman	<0.001	0.18	35.83	0.67	13.04	0.09	8.6	2.6
	NIR		0.08	3.35	0.08	6.62	0.21	12.8	3.9
Ammonium	Raman	<0.001	0.60	110.42	0.65	12.48	0.03	5.7	1.4
	NIR		0.34	12.86	0.07	6.13	0.05	7.6	1.9

For model precision (BCMSEP) Raman predictions were found better for VCD, glucose and lactate concentrations, while NIR spectroscopy predicted more precisely the antibody, glutamine and ammonium ion concentrations. These results were coherent with the accuracy values, excepted for antibody predictions. The NIR spectroscopy predicted more precisely but less accurately the antibody concentration than the Raman spectroscopy. This is probably due to the small number of LV used in NIR model for the antibody prediction. Indeed, models with small number of LV trend to underfit the reference data by giving relatively stable but inaccurate predictions. Other authors have explained that the lower precision of the Raman spectroscopy had no negative effect on the PLS algorithm to resolve the covariance structure, since this inverse least squares regression method is less affected by precision than by sample leverage in the estimation of the true solution [15].

Analytical sensitivity, described by the parameter γ, allows to quantify and to compare the sensitivity with respect to analytical precision across different analytical technologies. Raman spectroscopy appeared to be more sensitive for all the 6 culture variables compared to NIR spectroscopy. This can be explained by the fact that Raman scattering is more specific to molecules measured, generating narrow peaks, while NIR spectroscopy relies on the absorption of combinations and the overtones of the chemical bands of the molecules [19]. Moreover, NIR spectroscopy is known to have lower absorptivities compared to other absorption based spectroscopy, leading to lower sensitivities [18]. Moreover, for both spectroscopies, antibody models exhibited the largest sensitivity, since the concentration was relatively low (< 0.35 g/L). Therefore, an increase in one unit (1 g/L) can lead to an important signal response.

Glucose and lactate models were the most selective (SEL values) with Raman spectroscopy, while NIR spectroscopy presented the highest selectivity for lactate and antibody. Nevertheless, Raman spectroscopy was more selective than NIR spectroscopy for all the 6 analytes measured. Indeed, it is reported that thanks to its basic measuring principle, Raman spectroscopy is highly selective and can

differentiate molecules that are very similar [20]. Furthermore, a principal drawback of NIR spectroscopy is the fact that NIR absorption bands are broad and overlapping with one another. Especially, the large water spectral feature in NIR spectroscopy masks the information of other components available to build the models, leading to a decreased selectivity [4].

Similar results were found for the S/N ratio. Raman spectroscopy had the highest S/N ratios for glucose and lactate, while the highest ratios of NIR spectroscopy were for lactate and antibody. Moreover, the NIR ratios for lactate and antibody were higher than the Raman ones, which is inconsistent with expected trends found in previous results. One possible reason could be the Fermi-resonance effects of these two analytes, generating spectra features with greater complexity particularly in the Raman spectra [21]. Since the relationship between the intensity of Fermi-resonances and the concentrations of corresponding molecules are often non-linear, the information embedded in the Fermi-resonances is not used in the linear PLS regression, but is considered as noise. Also, the Fermi-resonance trends to enhance the overtones of the IR bands, thereby, leading to superior S/N ratio for NIR spectroscopy [22].

From the results discussed so far, the resulting LOD estimations should be straightforward. For the 6 culture variables measured, Raman spectroscopy provided smaller LOD compared to NIR spectroscopy. These results are alongside with other reports, which demonstrated that Raman spectroscopy can detect the analytes concentrations below the level that can be detected by NIR [23].

Values of RPD and RER were also calculated to complete the evaluation of each calibration models. Generally, a successful calibration with good predictive performance is admitted for RPD greater than 3 and RER greater than 15, while it is considered as insufficient when RPD is lower than 2 and RER lower than 8 [24]. Therefore, on the basis of these two parameters, Raman spectroscopy provided successful calibrations for VCD, glucose, and antibody concentrations and moderate useful calibrations for lactate and glutamine concentrations. Otherwise, NIR spectroscopy displayed a good calibration for antibody concentration and moderate useful calibration for VCD, glucose, lactate, and glutamine concentrations. However, it is confirmed that both spectroscopies failed to give satisfactory predictions for ammonium ion concentrations.

3.2. Calibration models using PLS2 or Raman and NIR spectra combination

As it is known that some cell culture variables tend to be inter-correlated with each other (*i.e.* glucose and lactate), a PLS regression of type 2 (PLS2) was also tested for both spectroscopic technics. This method has been reported as suitable to improve some model performances, as they

can decrease model noises came from the measurements errors [25]. In this case, global models were obtained by calibrating the spectral data with reference data of two culture variables each time (glucose/lactate, glutamine/ammonium, and VCD/antibody). We found that, the PLS2 method, used to develop Raman models of glucose and lactate concentrations, displayed larger LV numbers and led to only about 1% improvement regarding RMESCV and RMESP (Tables 2 & 3). Same little improvements were found for glutamine and ammonium ion concentration models built with NIR spectra, while no differences were observed for other variables for both spectroscopic methods (Tables 1 to 6). This limited statistical benefit could be explained by the fact that the calibration settings were applied for all variables. Thereby, the regression needed to find a compromise between different variables, instead of optimizing each variable with one specific solution as the normal PLS regression [26].

Besides, since Raman and NIR spectroscopies provide complementary information due to different physical processes involved, as they explore different vibrational energy status of molecules, a strategy of spectra combination was tested. This method extends the information embedded in both types of spectra with the objective to improve the calibration model performances [27]. After the pre-processing of Raman and NIR spectra in their respective working regions, the spectra were combined to generate a new “hyper-spectra”. However, this new strategy provided only little improvement (decrease of about 1% of RMESCV and RMESP for lactate and antibody prediction) compared to individual models using separately Raman or NIR spectra (Tables 1-6). As the spectral combination was carried out by putting simply the pre-processed spectra together, the combined spectra may contain redundant information, and the intrinsic noises in each set of spectra may also add complexity for model calibration.

3.3. Comparison of real-time monitoring of multi-variables using *in situ* Raman and NIR spectroscopy

To evaluate the ability of both spectroscopies to monitor *in situ* and in real-time various cell culture variables, PLS1 models developed previously using PLS regression were applied on Raman and NIR spectra collected every 20 minutes during a CHO cell batch culture which were not involved in the calibration procedure. The predicted values were then compared with the off-line reference measurements (Fig. 2). For each variable, the time profile of on-line values predicted by both spectroscopies generally met the trends observed with off-line measurements. However, several deviations should be noticed. During the first 20 hours of the culture, NIR spectroscopy displayed large prediction errors compared to off-line values. Since models developed from NIR spectra had lower sensitivity and selectivity compared to models developed from Raman spectra, it is highly likely

that their prediction performance would be affected by minor spectral variations other than changes in the component concentrations [28]. Indeed, the water signal and the NIR spectra noises could be then predominant compared to the variation range of analytes, especially during the initial lag phase of the culture when the variations in the medium components were very small. Furthermore, as estimated previously, NIR models were more precise but less accurate (Table 7). Thereby, although the NIR predictions appeared visually smoother than Raman, their values displayed larger errors especially at the beginning of the culture. Otherwise, off-line measurements indicated a consumption of produced lactate from about 120 h (Figure 2-e), which allowed to decorrelate lactate from glucose. Although NIR spectroscopy had higher S/N ratio for lactate prediction, only Raman spectroscopy was able to predict this decrease of lactate concentration in real-time. It seems that the higher sensitivity and selectivity of Raman spectroscopy for lactate played a more important role in its real-time monitoring. Lastly, and as expected by the lower performances of calibration models previously evaluated for both spectroscopies, the real-time monitoring of ammonium ion concentrations was less accurate (Figure 2-f).

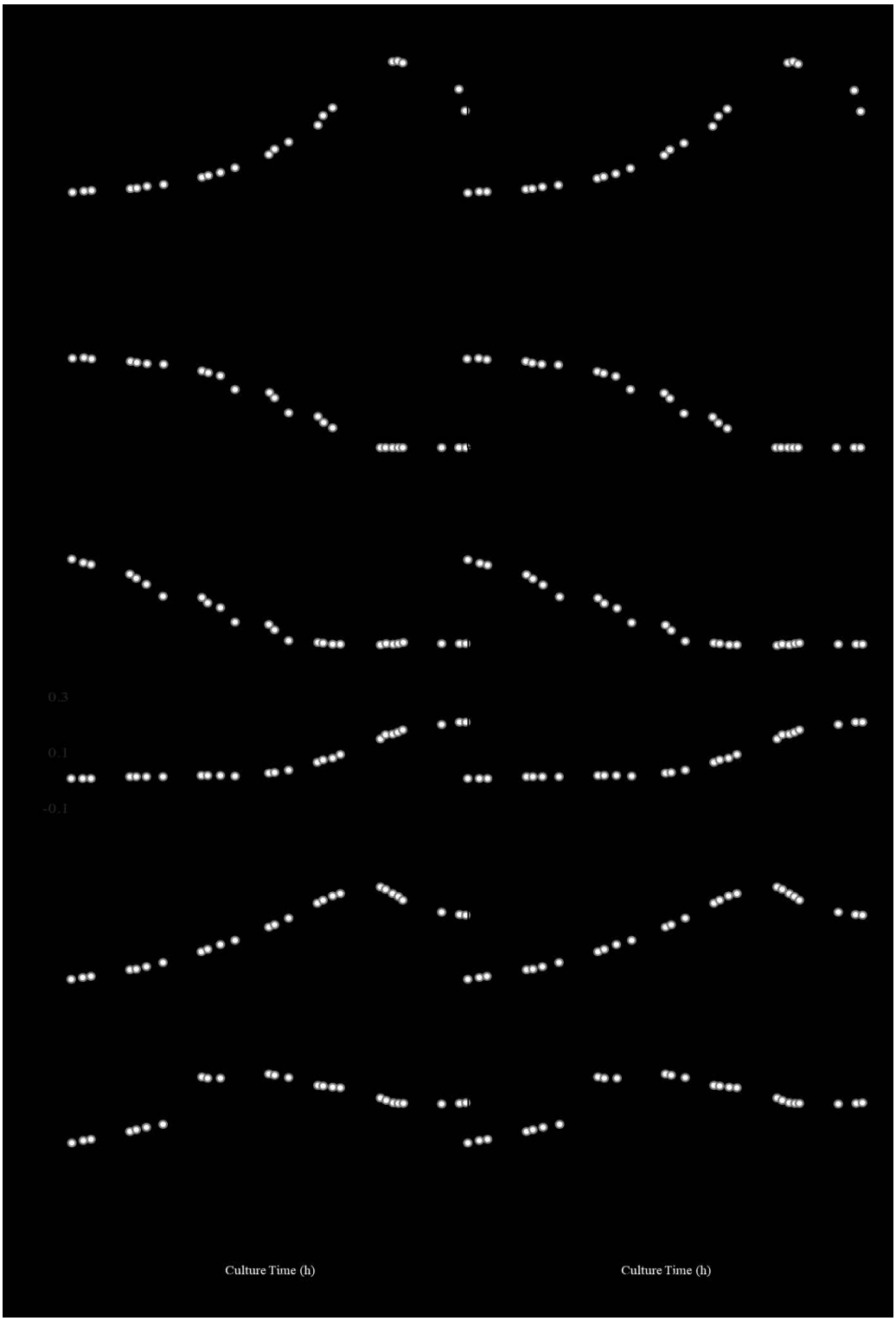


Figure 2: Comparison of real-time predictions of the concentrations of viable cells (a), glucose (b), lactate (c), antibody (d), glutamine (e), and ammonium ions (f) throughout a CHO cell culture in batch bioreactor, by using in situ Raman (left column) and NIR (right column) spectroscopy. Circles: off-line measurements.

4. Conclusion

PLS models developed from either Raman or NIR spectra demonstrated a good potential to simultaneously predict the concentration of 6 cell culture variables (VCD, glucose, lactate, glutamine, ammonium ions, and antibody concentrations) during CHO cell cultures. Beyond the error statistics, the multivariate FOM was used to evaluate and compare thoroughly the model performances. Each spectroscopy methods displayed advantages and limits, which demonstrates the complexity for the selection of the most adequate analytical method. However, Raman spectroscopy outperformed NIR in terms of sensitivity, selectivity and limit of detection. These benefits were demonstrated thereafter in the real-time monitoring of a cell culture in batch mode. Although the general trends of different variables were successfully monitored by both spectroscopies, some large deviations have been observed between NIR *in situ* predictions and off-line data during the first 20 hours of culture. The incapacity of NIR spectroscopy to detect the lactate consumption has also been showed at the end of culture process. For the first time, this paper compares the feasibility of real-time monitoring multiple variables of a mammalian cell culture processes by acquiring spectra in parallel with *in situ* implemented Raman and NIR spectroscopy sensors. As no single spectroscopy can satisfy the entirety of the FOM metrics, this indicates that the choice could need to be performed on a case-by-case basis. For the real-time monitoring, and eventually the control of the key-variables during animal cell culture processes, Raman spectroscopy seems to be a more adequate tool, since it gives accurate predictions of the variables especially at low levels of concentration. Indeed, a main process control objective during fed-batch or perfusion operations is to keep the nutrients (glucose, glutamine) at low level to prevent the accumulation of the toxic by-products. However, NIR spectroscopy remains a potential candidate, as further improvements of its real-time prediction ability could be expected from investigation of other chemometric methods, such as non-linear regression methods, and by using additional spectra data. Such *in situ* monitoring of key-parameters of mammalian cell cultures could advantageously contribute to the PAT approach by allowing feedback control strategies of the pharmaceutical processes.

References *(pour alléger le document, les références complètes sont rassemblées à la fin du mémoire*

[1] B.N. Berry *et al.* (2016)
[2] A.P. Teixeira *et al.* (2009) 726–732.
[3] A.E. Cervera *et al.* (2009)
[4] N.D. Lourenço *et al.* (2012)
[5] H. Mehdizadeh *et al.* (2015)
[6] B. Berry *et al.* (2015)
[7] M. Li *et al.* (2016)
[8] R.C. Rowland-Jones *et al.* (2017)
[9] S. Romero-Torres *et al.* (2009)
[10] B. Kozma *et al.* (2017)
[11] I.H.T. Guideline (2005)
[12] F. Sales *et al.* (1997)
[13] A. Menditto (2007)
[14] A. Lorber *et al.* (1997)
[15] S.M. Short (2007)
[16] G.D. Doddridge *et al.* (2015)
[17] R. Boqué *et al.* (2000)
[18] M. Sandor *et al.* (2013)
[19] S. Beutel *et al.* (2011)
[20] A.S. Franca *et al.* (2017)
[21] P. Larkin (2017)
[22] D.A. Burns *et al.* (2007)
[23] L. Zhao *et al.* (2015)
[24] C.A. Roberts *et al.* (2004)
[25] H. Martens *et al.* (1992)
[26] S.M. Mercier *et al.* (2016)
[27] S. Carlesi *et al.* (2016)
[28] J. Kim *et al.* (2008)

II.2.3. Principales contributions de la publication

Dans cette article, le potentiel des spectroscopies Raman et NIR pour le suivi en ligne de procédés de culture de cellules CHO a été présenté et comparé, au travers de la prédiction simultanée des concentrations des cellules viables, du glucose, du lactate, des anticorps (IgG), de la glutamine et des ions ammonium. Par la suite, des facteurs de mérites (FOM) multivariés ont été utilisés pour comparer plus en détail la performance des différents modèles développés. Cependant, aucune de ces deux techniques ne permet de satisfaire l'intégralité des critères des FOM, indiquant la complexité du choix de la technologie la plus adéquate pour les procédés concernés. La performance des différentes techniques spectroscopiques devra ainsi être évaluée au cas par cas. Globalement, la spectroscopie Raman surpasse la spectroscopie NIR pour les critères de sensibilité analytique, de sélectivité et de limite de détection. Ces avantages ont été confirmés par la suite lors du suivi cinétique en temps réel de différents paramètres au cours d'une culture en mode discontinu. En effet, la spectroscopie Raman prédit mieux que la spectroscopie NIR, d'une part, l'évolution des différents paramètres, surtout en début de culture, et, d'autre part, la consommation du lactate en fin de culture. Ainsi, pour les procédés de culture de cellules animales, bien que les deux spectroscopies présentent un grand potentiel, la spectroscopie Raman semble être plus en adéquation avec ce type de procédé. Elle est en effet plus sensible et plus sélective pour détecter des molécules de faible concentration dans un milieu aqueux complexe. Ceci est essentiel, puisque l'un des objectifs majeurs pour le contrôle de ces procédés en mode semi-continu ou continu sera de maintenir les éléments nutritifs à de faibles concentrations en vue d'éviter l'accumulation des sous-produits toxiques dérivant directement de leur métabolisme et pouvant provoquer potentiellement un "shift" du métabolisme.

Par ailleurs, d'autres approches ont été testées comme l'utilisation de la méthode chimiométrique PLS de type 2 ou la combinaison des données spectrales NIR et Raman. Néanmoins, ces approches n'ont que très peu amélioré les erreurs relatives (RMSEC, RMSECV, RMSEP) des modèles, car elles ne semblent pas assez optimisées pour notre application. Afin d'améliorer la performance de l'approche basée sur la combinaison des données spectrales, différentes stratégies sont proposées dans la suite de ce chapitre.

II.3. Stratégies de combinaison des données spectrales NIR et Raman

Les principes de fonctionnement des spectroscopies Raman et NIR étant différents, elles fournissent des informations différentes sur l'état énergétique des molécules. La spectroscopie Raman est basée sur la diffusion de lumières inélastiques, tandis que la spectroscopie NIR reflète l'absorption de l'énergie des liaisons des molécules. Ces deux spectroscopies peuvent donc s'avérer complémentaires. La fusion de leurs données serait alors un moyen efficace pour utiliser au mieux les avantages de chacune, en vue d'améliorer la performance du modèle de prédiction développé. Cette fusion de données implique le traitement d'un grand nombre de données multivariées de nature différente, nécessitant l'utilisation de stratégies chimiométriques spécifiques. Cependant, les études sur la fusion des données sont rares (Huang *et al.*, 2014). Les travaux existants relèvent plutôt du domaine de l'ingénierie robotique ou des analyses alimentaires et pétrolières (Godinho *et al.*, 2014). D'après la littérature, la fusion des données peut être classée selon trois niveaux : bas, intermédiaire et haut (Solano *et al.*, 2012). Dans ce cadre, l'objectif de notre étude a été de tester différentes techniques pour chacun des trois niveaux afin de prédire en ligne la concentration du glucose à partir des spectres NIR et Raman, acquis conjointement *in situ* au cours de cultures de cellules CHO. La performance des modèles, caractérisée par leur exactitude, a été évaluée et comparée à l'aide des erreurs relatives RMSEC, RMSECV et RMSEP.

II.3.1. Fusion des données de niveau bas

La fusion des données de niveau bas consiste à combiner directement les signaux originaux après les étapes de prétraitement des spectres. L'avantage principal de ce niveau de fusion est qu'il ne nécessite qu'un seul modèle de calibration établi à partir du spectre fusionné. Cette approche est ainsi la plus simple à mettre en place. Néanmoins, la fusion étant réalisée à partir des deux ensembles de données d'origine, la matrice des données fusionnées possède une taille très importante, ce qui entraîne une augmentation significative du temps de calcul nécessaire pour établir les modèles de calibration et de prédiction. De plus, une normalisation et une mise à l'échelle doivent être effectuées pour éviter que les données provenant d'un analyseur ne saturent les données provenant de l'autre analyseur (Figure 48 : Exemple de normalisation des données spectrales provenant de différentes sources : A, spectres NIR ; B, spectres Raman ; C, spectres normalisés par la méthode SNV. (Ovalles et Rechsteiner, 2015).

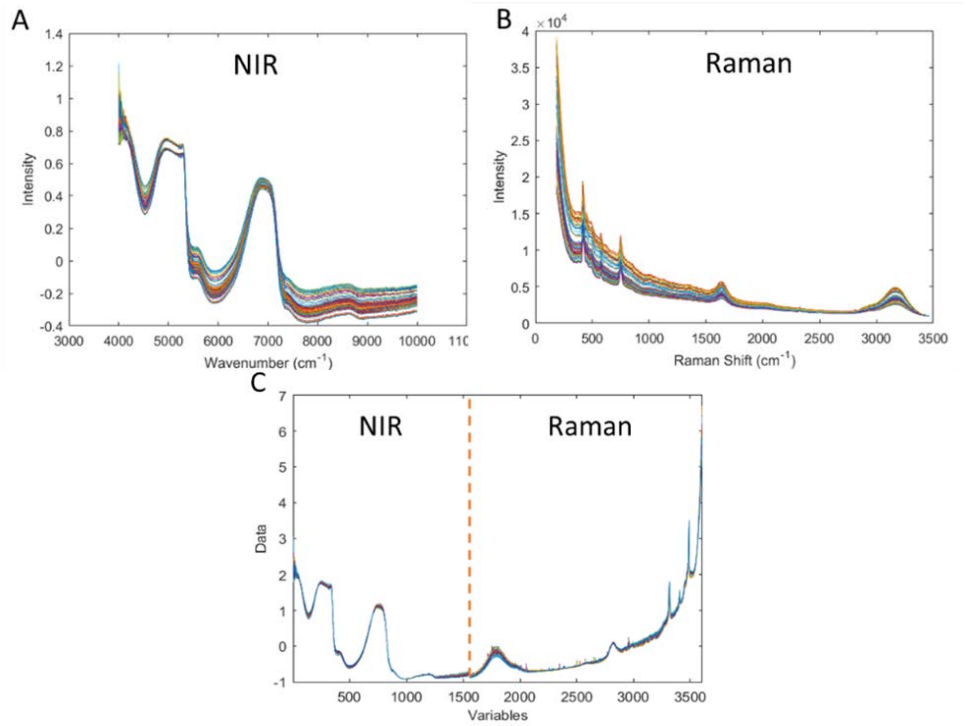


Figure 48 : Exemple de normalisation des données spectrales provenant de différentes sources : A, spectres NIR ; B, spectres Raman ; C, spectres normalisés par la méthode SNV.

L'approche de combinaison spectrale utilisée dans l'article précédent (Chapitre II, 2.2) correspond à une fusion des données de niveau bas, les spectres dans les régions de travail pour chaque type de spectroscopies étant fusionnés après prétraitement. Dans cette nouvelle partie, ce sont les spectres complets des spectroscopies Raman et NIR qui ont été fusionnés directement après avoir chacun été prétraité par la méthode SNV. Les données fusionnées sont ensuite centrées par leur moyenne (MC : mean centering). Les valeurs de glucose mesurées hors ligne sont également centrées par la moyenne avant d'être utilisées dans le modèle PLS pour la calibration des données fusionnées (Figure 49). Les résultats obtenus avec le modèle de fusion ainsi développé sont présentés dans le Tableau 8.

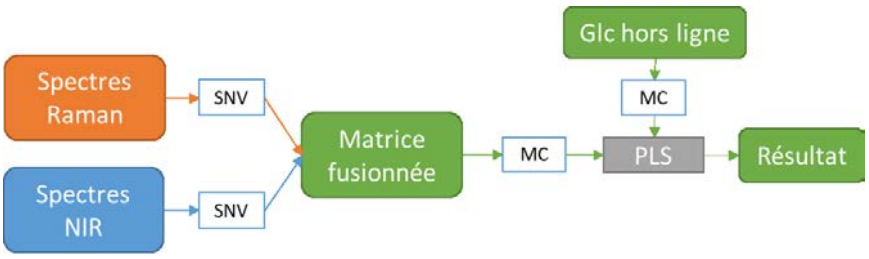


Figure 49 : Diagramme schématique de la fusion des données de niveau bas

Tableau 8 : Performance des modèles individuels et du modèle fusionné de niveau bas

Spectroscopie	Méthode	LV	RMSEC	RMSECV	RMSEP
			Glucose (mM)		
NIR	PLS	5	1,96	2,71	2,31
Raman	PLS	7	0,58	1,27	0,81
NIR + Raman	PLS	11	0,49	1,63	1,24

Le modèle développé à partir des données fusionnées de niveau bas ne permet d'améliorer que l'erreur relative de calibration, mais moyennant un nombre de LV beaucoup plus important par rapport aux modèles individuels, nombre résultant de la complexité accrue des données fusionnées. Cependant, les modèles utilisant un grand nombre de LV présentent des risques de surajustement (overfitting), c'est à dire une spécificité trop importante vis-à-vis du jeu de données utilisé pour la calibration, ce qui les rend moins robustes pour prédire d'autres jeux de données (Cervera *et al.*, 2009). C'est en effet le cas ici, puisque les valeurs de RMSECV et RMSEP, bien que plus faibles que celles du modèle NIR, sont plus élevées que celles du modèle Raman. Pour cette approche de fusion de données, les informations provenant des deux sources NIR et Raman pourraient potentiellement être redondantes. L'augmentation de l'information obtenue pourrait ne pas compenser la quantité d'information non pertinente ou les bruits apportés par l'addition des données spectrales (Biancolillo *et al.*, 2014).

II.3.2. Fusion des données de niveau intermédiaire

La fusion des données de niveau intermédiaire (aussi appelée « feature level fusion ») nécessite d'extraire au préalable des informations pertinentes de chacune des deux sources de données, avant de les rassembler en un seul jeu de données utilisé pour la régression multivariée. L'approche la plus couramment utilisée consiste en une réduction de la dimension des données spectrales. Pour y parvenir, des méthodes de sélection de variables (telles que la PLS intervalle (iPLS) ou les algorithmes génétiques) ou de compression (comme la PCA ou la PLS) peuvent être utilisées pour extraire les variables les plus pertinentes (Biancolillo *et al.*, 2014 ; Márquez *et al.*, 2016). Par conséquent, la fusion des données de niveau intermédiaire est généralement plus efficace que la fusion de niveau bas, puisque cette approche minimise la quantité de données et maximise les informations pertinentes dans les différentes matrices de données utilisées. Son principal inconvénient réside

dans le fait que des modèles préliminaires sont nécessaires (Li *et al.*, 2018). Le défi consiste ainsi à identifier la combinaison optimale d'informations extraites et de prétraitements pour les différentes méthodes, afin de fournir le meilleur modèle final. Pour notre étude, différentes méthodes d'extraction des données ont été testées.

- *Extraction des données par PLS itérative (iPLS)*

La méthode d'extraction par iPLS est utilisée en divisant les données spectrales en intervalles glissants de 50 variables avec un pas de 20 variables. En comparant les RMSECV des sous-modèles (régressions PLS locales dans chaque intervalle), les 20 intervalles présentant l'erreur la plus faible sont choisis. Les régions sélectionnées des spectres sont ensuite prétraitées et combinées pour former la matrice fusionnée (Figure 50).

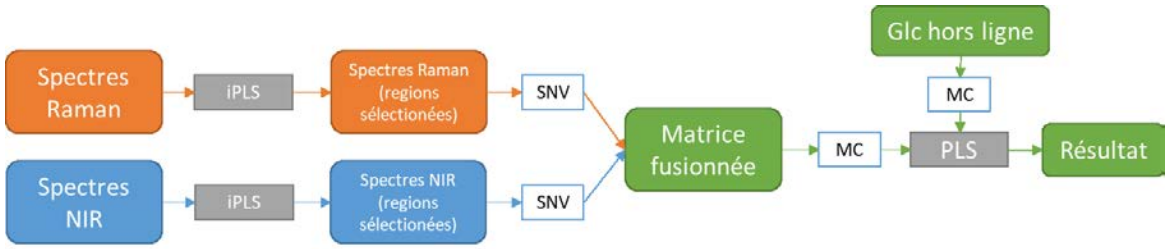


Figure 50 : Diagramme schématique de la fusion des données de niveau intermédiaire par l'approche iPLS

Les régions spectrales sélectionnées pour chaque spectroscopie sont présentées dans la Figure 51. Les attributions théoriques correspondant à différents groupements chimiques dans les zones sélectionnées sont présentées dans le Tableau 9.

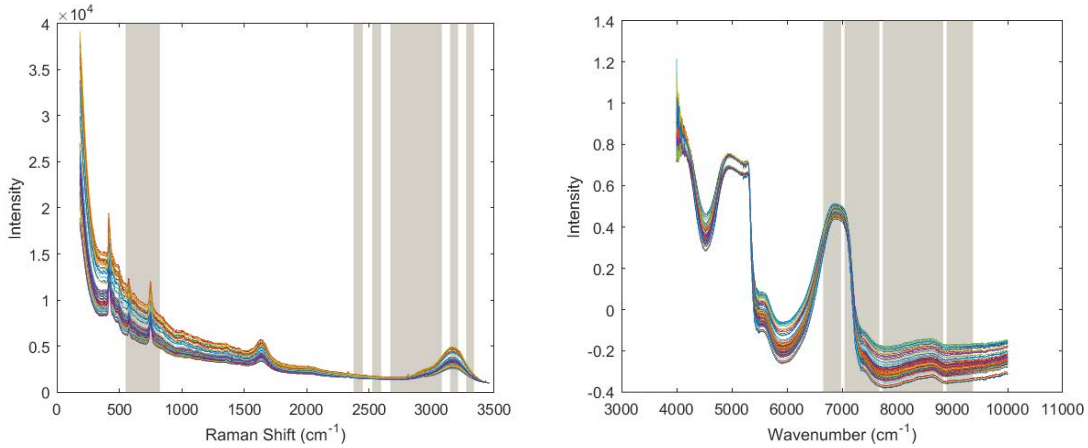


Figure 51 : Régions sélectionnées (zones grisées) dans les spectres Raman (gauche) et NIR (droite) par la méthode iPLS pour la prédiction de la concentration de glucose

Tableau 9 : Attributions théoriques des régions spectrales pour les différents groupements chimiques (Wiercigroch et al., 2017 et Salzer, 2018)

Régions spectrales Raman (cm^{-1})	**Attributions Raman**	**Régions spectrales NIR (cm^{-1})**	**Attributions NIR**
550 -824	élongation ν(C-C), déformations δ(C-C-C), δ(C-C-O) et δ(C-O-C)	**6657-7100**	première harmonique de l'élongation des bandes OH
2880 – 2962	élongation ν(CH) et ν(CH_2)	**6944, 7067 et 7168**	bandes de combinaison du C-H des groupements CH_2
3265 -3395	élongation ν(OH)	**7085 et 7194**	premières harmoniques et déformation des groupements CH et CH_2
		8163 – 8175	secondes harmoniques des groupements CH et CH_2
		9386	bandes de combinaison des groupements C-OH

Les résultats du modèle PLS avec la matrice fusionnée pour la prédiction de la concentration de glucose sont donnés dans le Tableau 10.

Tableau 10 : Performances comparées des modèles individuels et des divers modèles fusionnés de niveau moyen

Spectroscopie	**Méthode**	**LV**	**RMSEC**	**RMSECV**	**RMSEP**
			Glucose (mM)		
NIR	PLS	5	1,96	2,71	2,31
Raman	PLS	7	0,58	1,27	0,81
NIR + Raman	iPLS	8	0,28	1,15	0,83
	PCA	7	1,95	2,22	1,4
	PLS	8	0,57	0,69	0,84
	iPLS+PCA	6	2,16	2,6	1,3
	iPLS+PLS	7	0,36	0,43	0,86

Avec un nombre de LV légèrement plus important, le modèle fusionné issu de la sélection de données par iPLS permet de diminuer grandement la RMSEC et faiblement la RMSECV, par rapport aux modèles développés avec un seul type de spectroscopie. Par contre, la RMSEP reste très proche de celle du modèle Raman.

- *Extraction des données par la méthode PCA*

Dans ce cas, une analyse en composantes principales (PCA) est d'abord appliquée sur chaque groupe de spectres. Les scores de chaque modèle PCA sont ensuite combinés pour former la matrice fusionnée (Figure 52), qui servira à établir le modèle de calibration par régression PLS.

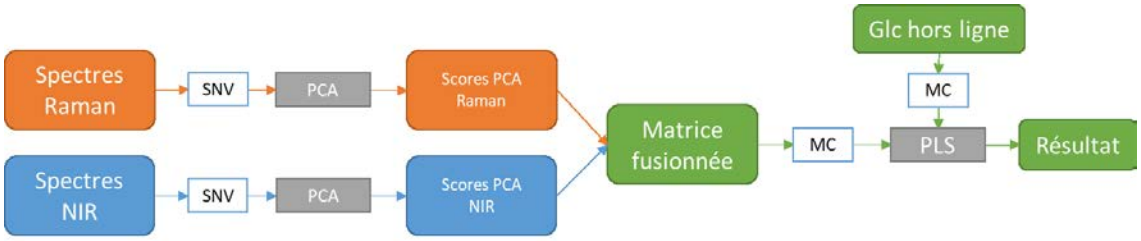


Figure 52 : Diagramme schématique de la fusion des données de niveau intermédiaire par l'approche PCA

Dans notre étude, la matrice fusionnée comporte 6 premières composantes principales des spectres Raman et 2 premières composantes principales des spectres NIR. Ceci confirme le fait que les spectres Raman contiennent plus d'information pertinente. Cette spectroscopie apparaît donc plus sensible et plus sélective pour le suivi de la concentration de glucose. Néanmoins, la performance du modèle développé à partir de la matrice fusionnée n'est pas supérieure à celle des modèles individuels (Tableau 10). L'une des raisons possibles serait que l'information extraite par PCA n'est pas assez pertinente pour le paramètre à prédire, le glucose. En effet, la PCA ne tient compte que des variations propres aux spectres, qui peuvent être influencées par les variations des autres composés du milieu de culture ou par les bruits des appareils. Afin de compenser cette limitation, l'extraction des informations et la réduction de la dimension des données initiales par la méthode PLS a alors été évaluée.

- *Extraction des données par la méthode PLS*

L'approche basée sur la PLS est similaire à celle utilisant la PCA (Figure 53).

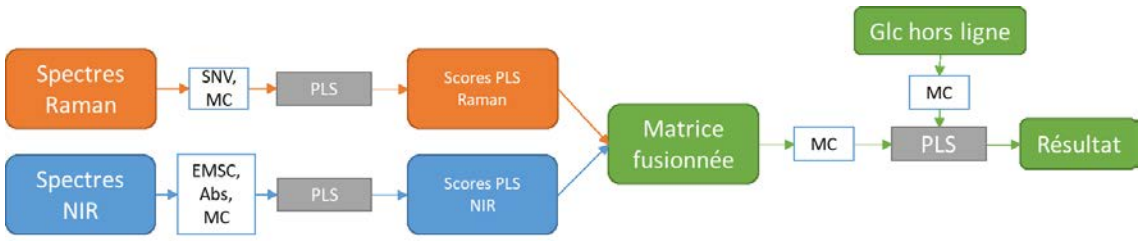


Figure 53 : Diagramme schématique de la fusion des données de niveau intermédiaire par l'approche PLS

Dans notre étude, le nombre de variables latentes (LV) utilisées est de 5 pour le modèle Raman et de 3 pour le modèle NIR. La matrice fusionnée est centrée par la moyenne ainsi que les valeurs expérimentales de la concentration de glucose. Une régression PLS est ensuite utilisée pour établir le

modèle de calibration fusionné. Dans ce cas, la RMSEC et la RMSECV sont largement réduites, contrairement à la RMSEP qui reste similaire à celle du modèle Raman (Tableau 10). Ces résultats indiquent que l'extraction de données par l'approche PLS permet d'éliminer efficacement les informations non pertinentes dans les spectres issus des deux types de spectroscopies. Pour tenter d'améliorer davantage la performance de la fusion de données de niveau moyen, ces diverses approches ont alors été combinées.

- *Extraction des données par des approches combinées iPLS+PCA et iPLS+PLS*

Pour ces 2 approches, les procédures sont similaires à celles proposées dans le cas de la fusion de données avec réduction spectrale par PCA ou PLS. Par contre, dans cette étude spécifique, la PCA ou la PLS sont appliquées sur des régions spectrales préalablement sélectionnées par la méthode iPLS (Figure 54).

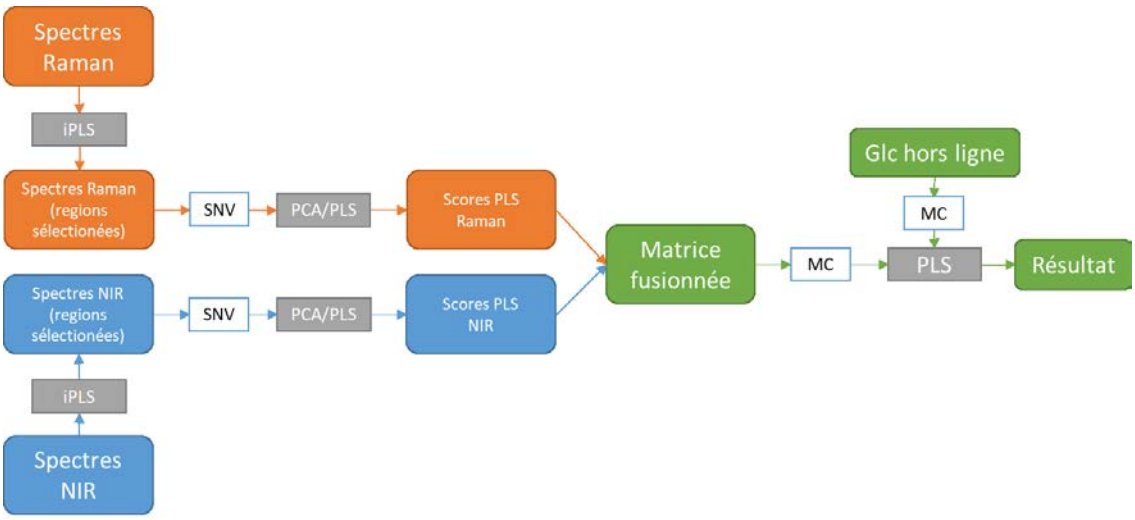


Figure 54 : Diagramme schématique de la fusion des données de niveau intermédiaire par les approches combinées iPLS+PCA ou iPLS+PLS

L'approche iPLS+PCA n'a pas réussi à améliorer la performance du modèle par rapport au modèle utilisant l'approche PCA seule. Au contraire, l'approche iPLS+PLS permet d'améliorer la performance du modèle au niveau de la RMSEC et de la RMSECV par rapport à l'approche PLS seule (Tableau 10). Comme discuté précédemment, la PCA n'analyse que les variations propres aux spectres. Une sélection spectrale non-optimale peut ainsi provoquer une perte d'information importante pour la molécule à prédire. La PLS permet, quant à elle, de lier les variations des spectres dans les régions sélectionnées avec la variation de la molécule à prédire, permettant ainsi de réduire les erreurs relatives du modèle.

Parmi les méthodes de fusion des données de niveau intermédiaire, l'extraction par l'iPLS et l'extraction par PLS sont les deux approches permettant de diminuer la RMSEC et la RMSECV. La combinaison de ces deux techniques d'extraction (iPLS+PLS) permet d'améliorer encore davantage la performance du modèle. Néanmoins, aucune de ces méthodes ne permet de diminuer la RMSEP significativement.

II.3.3. Fusion des données de niveau haut

Dans le cas de la fusion des données de niveau haut (également appelée fusion à niveau de décision), des modèles de régression sont développés séparément à partir de chacune des sources de données (NIR et Raman). Le résultat de chaque modèle individuel est alors combiné pour obtenir la matrice fusionnée finale. Les méthodes de régression appliquées peuvent varier considérablement et être adaptées aux besoins et objectifs spécifiques des analyses. Le défi, dans ce niveau de fusion de données, est de déterminer les modèles de régression les plus adéquats pour chaque groupe de données (NIR ou Raman) afin que leur combinaison apporte une plus-value de prédiction par rapport aux modèles individuels. Par ailleurs, il existe un risque de propagation des incertitudes, dans la mesure où les incertitudes de chaque modèle initial seront combinées dans le modèle fusionné. Ces deux aspects, dépendance par rapport au modèle initial choisi et propagation des incertitudes, ont conduit à une application limitée de la fusion des données de niveau haut. Cette approche est surtout utilisée pour des études de classification (Borràs *et al.*, 2015). Néanmoins, dans notre étude de régression quantitative, deux méthodes de fusion de données de niveau haut ont été testées : Ensemble-PLS (En-PLS) et Serial-PLS (S-PLS).

- *Fusion des données par la méthode Ensemble-PLS (En-PLS)*

Les techniques de régression Ensemble-PLS sont largement utilisées en apprentissage automatique pour créer des prédictions robustes et précises, en combinant des sous-modèles individuels (Mevik *et al.*, 2004). Dans notre étude, deux modèles PLS sont tout d'abord développés séparément à partir des spectres Raman ou NIR après prétraitements. Les valeurs prédites de chaque modèle sont alors combinées linéairement avec des coefficients de régression "Ensemble" afin d'obtenir la prédiction finale (Figure 55).

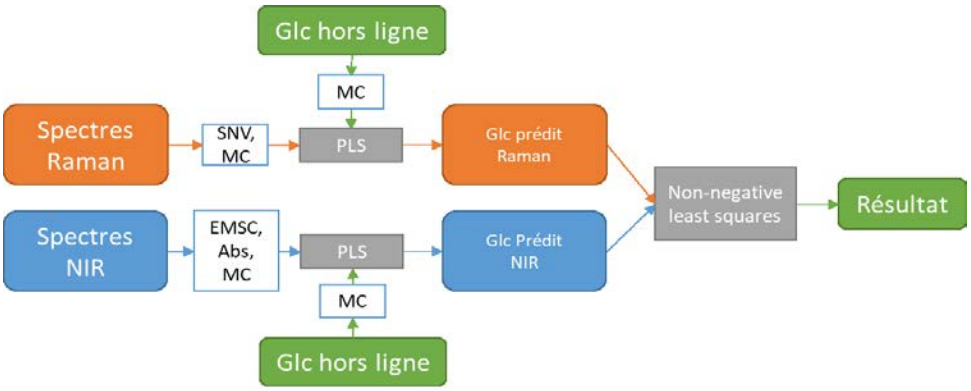


Figure 55 : Diagramme schématique de la fusion des données de niveau haut par l'approche En-PLS

Ces coefficients de régression "Ensemble" sont déterminés par la méthode des moindres carrés non négatifs (non-negative least squares) en raison de sa simplicité et de sa résistance au surajustement (overfitting) (Lee *et al.*, 2012). Les résultats de cette approche sont présentés dans le Tableau 11. Les erreurs relatives (RMSEC, RMSECV et RMSEP) observées sont similaires à celles obtenues dans le cas du modèle Raman individuel.

Tableau 11 : Performances des modèles individuels et des modèles fusionnés de niveau haut

Spectroscopie	Méthode	LV	RMSEC	RMSECV	RMSEP
			Glucose (mM)		
NIR	PLS	5	1,96	2,71	2,31
Raman	PLS	7	0,58	1,27	0,81
NIR + Raman	En-PLS	/	0,57	1,27	0,82
	S-PLS	8 ; 1	0,53	/	0,83

- *Fusion des données par la méthode Serial-PLS (S-PLS)*

Une seconde approche de fusion des données de niveau haut, la S-PLS, a alors été testée. Au lieu de combiner les données en parallèle, la S-PLS les combine en série. Le principe de la S-PLS consiste ainsi à développer, de manière itérative, une succession de modèles (par régression PLS) en utilisant les résidus des réponses du modèle précédent (*f1*, *f2*), jusqu'à obtenir une convergence des résidus obtenus (Figure 56). Ainsi, la variance dans le vecteur réponse (***y*** : concentrations de glucose mesurées hors ligne) qui n'est pas expliquée par les modèles précédents peut être décrite par les modèles suivants. La convergence de cet algorithme est atteinte lorsque *f2* (différence résiduelle entre *y'* et *y''*) de deux itérations successives est inférieure à un seuil (fixé à 0,02 pour cette étude). Le détail algorithmique de cette méthode a été décrit par Laxalde *et al.* (2014). Cette approche permet ainsi d'affecter un nombre de facteurs différent pour chaque modèle. Néanmoins, l'ordre des

modèles est important, puisque la première itération est réalisée sur les valeurs brutes de ***y***. Dans notre étude, nous avons choisi le groupe de spectres générés par spectroscopie Raman (*X1)* pour initier le processus, puisque, d'après les résultats obtenus précédemment, ils contiennent plus d'information que les spectres NIR. Le modèle final consiste en un seul cycle avec 8 LVs pour le premier modèle PLS développé à partir des spectres Raman (*X1*) après prétraitements (SNV+MC), et 1 seule LV pour le deuxième modèle développé à partir des spectres NIR (*X2*) après prétraitements (EMSC+Abs+MC). L'information restante, non expliquée par le premier modèle, est peu pertinente vis-à-vis de la molécule à prédire. Les performances de cette approche n'apparaissent pas meilleures, les erreurs relatives étant similaires à celles observées par le modèle Raman individuel (Tableau 11). Différentes combinaisons du nombre de LV pourraient être testées afin d'améliorer les performances de cette approche.

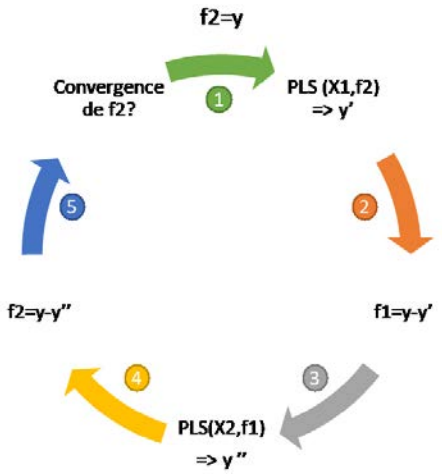


Figure 56 : Diagramme schématique de la fusion des données de niveau haut par l'approche S-PLS. X1, X2 sont les différents types de spectres. f2 et f1 sont des vecteurs de réponse dans les régressions PLS. Dans l'étape 1 d'initiation f2 est le vecteur de référence (concentration de glucose, y). y' et y'' sont les réponses prédites par les régressions PLS. Les étapes 2-5 sont répétées jusqu'à la convergence.

En résumé, différentes techniques de fusion des données de différents niveaux ont été évaluées dans cette étude. Ce sont des approches intéressantes permettant de combiner les informations qui sont souvent complémentaires à partir des différents types de spectres. En particulier, l'une des méthodes de fusion de données de niveau intermédiaire, l'approche iPLS+PLS, a permis d'améliorer significativement la performance du modèle concernant la RMSEC et la RMSECV, pour la prédiction de la concentration de glucose. Néanmoins, d'autres études plus complètes pourraient être envisagées afin d'optimiser les paramètres des différentes approches de fusion pour obtenir une meilleure performance des modèles. Cette approche de fusion des données pourrait être également

testée pour d'autres composés tels que la glutamine, les ions ammonium, les IgG etc. Il est possible que la fusion de données fonctionne mieux pour des composés présentant moins d'information redondante dans les deux spectroscopies.

II.4. Conclusion

Dans un premier temps, ce chapitre s'est attaché à comparer en parallèle la performance des spectroscopies Raman et NIR *in situ*, pour le suivi en ligne de la concentration de cellules viables, du glucose, du lactate, des IgG, de la glutamine et des ions ammonium, au cours de cultures de cellules CHO en bioréacteur. Les modèles de calibration ont été obtenus par la méthode PLS. A l'aide des facteurs de mérite multivariés (FOM). Nous avons évalué et comparé l'exactitude (accuracy), la linéarité, la précision, la sensibilité analytique, la sélectivité, le ratio signal/bruit, la limite de détection, ainsi que les ratios de gamme des erreurs (RER) et de déviation de la performance (RPD). Il en ressort que les deux méthodes spectroscopiques peuvent répondre aux critères de suivi en ligne de procédés de culture de cellules animales. Cependant, bien qu'aucune d'entre elles ne soit capable de satisfaire tous les FOMs pour toutes les variables analysées, la spectroscopie Raman surpasse la spectroscopie NIR au niveau de la sensibilité analytique, de la sélectivité et de la limite de détection. De plus, la spectroscopie Raman permet une meilleure prédiction des différentes variables au début de la culture, ainsi que de la consommation du lactate en fin de culture. La spectroscopie Raman semble donc un peu plus avantageuse que la spectroscopie NIR pour le suivi en ligne des variables, en particulier du fait des faibles concentrations rencontrées dans les procédés de culture de cellules animales. Elle représente donc un potentiel à envisager pour son implémentation dans les stratégies de rétrocontrôle et de pilotage automatique de ces bioprocédés, notamment en mode semi-continu ou continu-perfusé.

Dans un second temps, des combinaisons des spectres Raman et NIR ont été envisagées en utilisant différentes stratégies de fusion des données. L'objectif était d'identifier la possibilité d'améliorer la performance des modèles de prédiction, puisque ces deux spectroscopies peuvent être considérées comme complémentaires. Trois niveaux de fusion de données (bas, intermédiaire, haut) ont été testés dans le cas de la prédiction de la concentration en glucose. Les résultats obtenus montrent que la fusion des données de niveau intermédiaire, en particulier l'approche iPLS+PLS, s'avère la plus efficace. Elle permet de diminuer considérablement les erreurs relatives (RMSEC, RMSECV) et donc l'exactitude du modèle global NIR-Raman par rapport aux modèles individuels développés pour

chaque spectroscopie, même si la RMSEP est voisine du modèle Raman seul. Une étude chimiométrique plus complète sur les paramètres de chaque approche devrait être prolongée afin de mieux optimiser cette stratégie de fusion des données. De même, une démarche analogue devrait être réalisée pour évaluer ces approches dans le cas de la prédiction d'autres variables importantes du procédé.

Une synthèse des principaux résultats de ce chapitre est présentée dans la Figure 57.

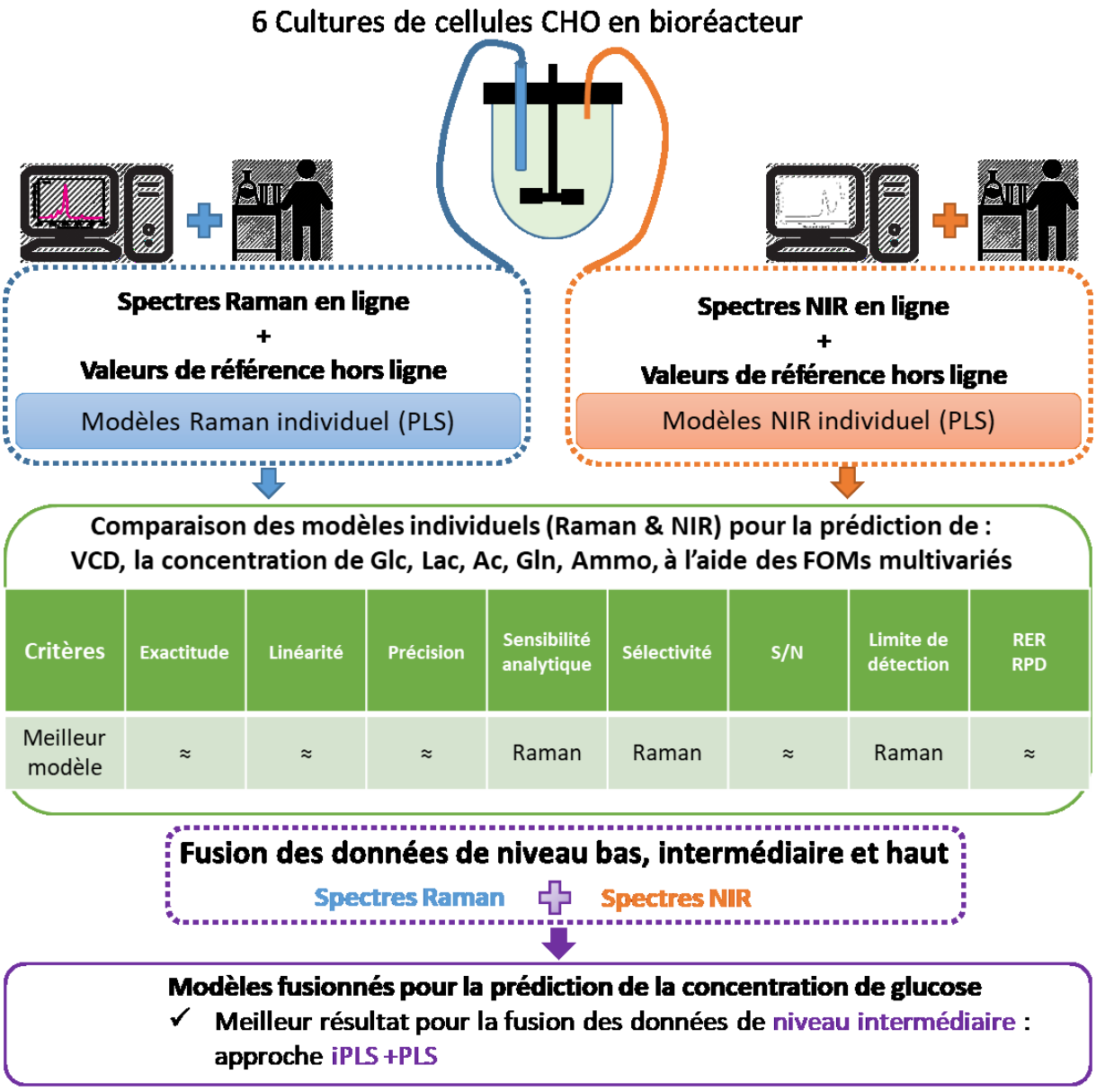


Figure 57 : Synthèse schématique des résultats du chapitre II.

Chapitre III

Suivi en ligne de la glycosylation des anticorps par spectroscopie Raman (et NIR)

III.1. Introduction 153

III.2. Analyses hors ligne de la glycosylation des IgG 153

III.3. Suivi en ligne de la macro-hétérogénéité de la glycosylation des IgG par spectroscopie Raman 160

III.3.1. Introduction 160

III.3.2. Manuscrit de la publication 161

III.3.3. Contributions de la publication 174

III.4. Suivi en ligne de la micro-hétérogénéité de la glycosylation des IgG par spectroscopie Raman 175

III.4.1. Introduction 175

III.4.2. Manuscrit de la publication 176

III.4.3. Contribution de la publication 182

III.4.4. Résultats supplémentaires non intégrés dans la publication 182

III.5. Suivi en ligne de la glycosylation des IgG par spectroscopie NIR 187

III.5.1. Modèles NIR de prédiction de la macro-hétérogénéité de glycosylation 187

III.5.2. Modèles NIR de prédiction de la micro-hétérogénéité de glycosylation 189

III.6. Conclusion 191

III.1. Introduction

Dans ce chapitre, nous nous proposons d'étudier le potentiel des spectroscopies Raman et NIR pour le suivi en ligne de la glycosylation des IgG produites au cours des cultures de cellules CHO. Nous avons montré, dans le chapitre II des résultats, que les spectroscopies Raman et NIR permettent de prédire en temps réel la concentration des anticorps IgG avec une assez bonne précision. Néanmoins, à l'heure actuelle, il n'existe pas de méthode rapportée dans la littérature pour le suivi en ligne de la qualité de l'IgG. Le plus souvent, les échantillons sont analysés en une seule fois, à la fin du procédé de culture, par des méthodes hors ligne, ce qui ne permet pas, notamment, d'effectuer de rétrocontrôle. Il apparaît donc intéressant d'étudier le potentiel de ces techniques spectroscopiques pour le suivi en temps réel de la qualité de l'IgG, et notamment de sa glycosylation (macro- et micro-hétérogénéité). L'objectif global de ce chapitre visera donc à développer des stratégies permettant, non seulement, de quantifier les IgG produites, mais aussi, de caractériser leur glycosylation en temps réel au cours d'un procédé de culture de cellules animales.

Pour ce faire, ce chapitre proposera tout d'abord la mise au point d'un protocole d'analyse hors ligne de la glycosylation de l'IgG. Cette méthode sera ensuite considérée comme la méthode de référence pour calibrer les spectres obtenus en ligne. Puis, nous développerons des modèles de prédiction pour la macro- et la micro-hétérogénéité de la glycosylation de l'IgG à l'aide de techniques chimiométriques. Enfin, ces modèles seront testés avec des cultures de cellules CHO pour vérifier la qualité de prédiction en temps réel de l'état de glycosylation des IgG.

III.2. Analyses hors ligne de la glycosylation des IgG

En général, pour pouvoir analyser la glycosylation d'une protéine, il convient tout d'abord de purifier celle-ci à partir des échantillons de culture. Cette purification est souvent réalisée, soit par chromatographie d'affinité, soit à petite échelle, par des kits contenant les protéines A ou G (Liu *et al.*, 2010). Pour vérifier la nécessité de cette étape dans notre étude, nous avons, dans un premier temps, comparé par électrophorèse des échantillons de culture non purifiés (Figure 58, colonnes A, C et D) à des échantillons purifiés par l'ultrafiltration (colonne B). L'électrophorèse SDS-PAGE (colonnes A, B, C) permet de séparer et de visualiser les monomères des protéines en condition dénaturante. Deux bandes largement majoritaires peuvent être observées (colonne A et C). Après transfert des

protéines sur une membrane de Western-Blot (colonne D), seuls les anticorps de type IgG sont révélés par un anticorps anti-IgG spécifique. A l'aide des marqueurs de taille, il peut être observé que les bandes les plus hautes correspondent à un poids moléculaire d'environ 50 kDa, tandis que les plus basses correspondent à environ 25 kDa. Ces valeurs sont en adéquation avec le poids moléculaire des chaînes lourdes et légères d'un anticorps de type IgG (Wang *et al.*, 2008). En comparant l'IgG purifiée (colonne B) à l'échantillon du surnageant de culture (colonne A) (Figure 58a), on peut conclure que ce dernier est quasiment exempt d'autres protéines extracellulaires, et que les IgG produites sont très majoritaires. Ainsi, une première étape de purification spécifique des IgG ne sera pas indispensable pour notre étude.

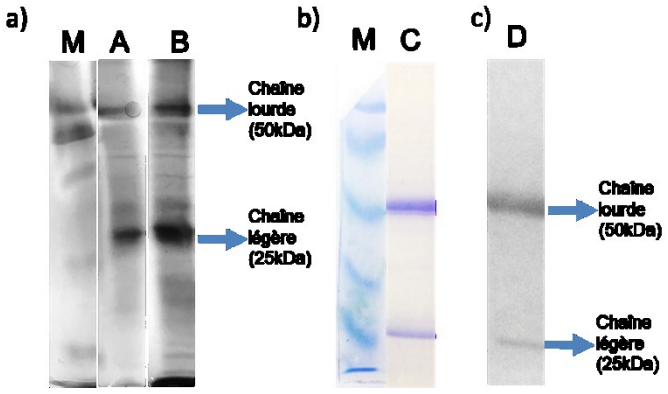


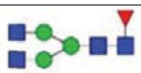
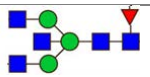
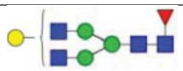
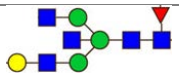
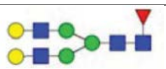
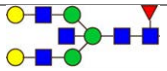
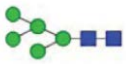
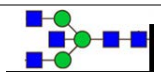
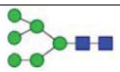
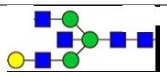
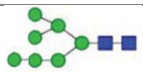
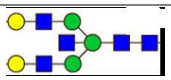
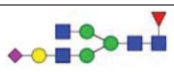
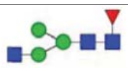
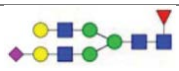
Figure 58 : Analyses d'électrophorèses de l'anticorps IgG produit dans le milieu de culture : a) SDS-PAGE révélation nitrate d'argent ; b) SDS-PAGE révélation bleu de Coomassie ; c) Westen-Blot. M : marqueur de taille ; A, C, D : échantillon de culture ; B : IgG purifiée

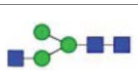
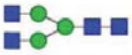
.

Pour analyser l'état de glycosylation des IgG produites, la chromatographie liquide couplée à la spectrométrie de masse (HPLC-MS) a été utilisée. Afin d'augmenter la sensibilité de l'analyse, les glycannes ne représentant que 2 à 3% de la masse totale, les IgG ont d'abord été hydrolysées par une enzyme, la trypsine. Cette réaction permet d'hydrolyser la glycoprotéine en glycopeptides de taille plus petite sans modifier le site de glycosylation. Les glycopeptides ont été analysés directement, sans hydrolyse à la PNGase qui libèrerait les groupements glycannes de la partie peptidique. En effet, cela permet d'analyser à la fois la macro- et la micro-hétérogénéité de la glycosylation de l'IgG (Planinc *et al.*, 2016). L'hydrolyse par la trypsine étant beaucoup étudiée pour l'analyse de la glycosylation des IgG par spectrométrie de masse, les tailles et les séquences glycanniques et peptidiques sont déjà répertoriées (Kapur *et al.*, 2014). Il est alors possible de connaître la masse molaire des différents glycopeptides analysés et d'en déduire leurs états de charges. De ce fait, leur séparation et détection par HPLC-MS sera donc bien spécifique et ne sera pas influencée par

d'éventuelles impuretés peptidiques. Nos résultats ont permis de distinguer 25 glycopeptides différents portant les groupements glycanniques typiquement observés dans les AcM du type IgG (Reusch *et al.*, 2015; Selman *et al.*, 2010; Sinha *et al.*, 2008). La structure de ces groupements glycanniques et leur masse molaire sont présentées dans le Tableau 12 ci-dessous. De plus, les peptides non-glycosylés, correspondant à la séquence connue du site de glycosylation, ont aussi été analysés et donneront accès à la concentration de la forme aglycone de l'IgG.

Tableau 12 : Structures et masses molaires des groupements glycanniques observés dans les IgG produites par les cellules CHO après analyse par HPLC-MS

Glycoformes	Structure	Masse molaire (Da)	Glycoformes	Structure	Masse molaire (Da)
Aglycone	/	0	**G2**		1623,5
G0F		1445,4	**G0FN**		1647,6
G1F		1607,5	**G1FN**		1809,7
G2F		1769,6	**G2FN**		1971,7
Man5		1217,2	**G0N**		1501,6
Man6		1379,3	**G1N**		1663,6
Man7		1541,4	**G2N**		1825,7
Man8		1703,5	**G1FS**		1897,7
G0F-N		1242,2	**G2FS**		2059,7
G1F-N		1404,3	**G2FS2**		2349,9

G0-N		1097,3	**G1S**		1751,6
G0		1299,3	**G2S**		1913,7
G1		1461,4	**G1FS-N**		1694,5

▼: fucose ; ■ : GlcNAc ; ● : mannose ; ● : galactose ; ◆ : acide sialique

A partir des aires sous pics du signal MS, le pourcentage de chaque glycopeptide sera calculé en divisant l'aire sous pic correspondante par la somme des aires sous pic. Puis, sa concentration sera obtenue en multipliant son pourcentage par la concentration en IgG totale obtenue par l'analyseur automatique Gallery™. Notons que chaque molécule d'IgG permet de générer seulement 2 glycopeptides identiques qui correspondent au site de glycosylation de chaque chaîne lourde. De ce fait, chaque glycopeptide analysé ne pourra provenir que d'une seule glycoforme d'IgG. Ceci permettra donc de relier directement la concentration ou le pourcentage du glycopeptide analysé à delui de la glycoforme de l'IgG correspondant.

Néanmoins, la structure spatiale polymérique de l'IgG et la composition du milieu de culture sont susceptibles d'influencer l'action de la trypsine. C'est pourquoi, trois protocoles de préparation des échantillons, incluant des étapes de précipitation, re-solubilisation et dénaturation des IgG pour faciliter l'accès de la trypsine, ont été comparés (Tableau 13). Pour ce faire, les aires sous pic du glycopeptide le plus abondant (pept-G0F), analysé dans un même échantillon de surnageant de culture traité par les différents protocoles, sont présentées dans la Figure 59. En comparant les résultats issus des protocoles 1 et 2, il apparait que l'étape de précipitation des protéines est nécessaire. En effet, sans cette étape, l'efficacité de digestion est environ 10 fois plus faible. Puis, en comparant les résultats issus des protocoles 2 et 3 qui diffèrent par leur méthode de dénaturation (urée ou chauffage), la dénaturation des IgG par chauffage apparaît beaucoup plus efficace.

Tableau 13 : Protocoles de préparation des échantillons pour l'analyse de la glycosylation des IgG

Protocole	Précipitation	Re-solubilisation	Dénaturation	Hydrolyse
1	/	/	98°C, 10 min	Trypsine, 37°C
2	Acétone, 2h, -20°C	Tampon ABC	98°C, 10 min	Trypsine, 37°C

3	Acétone, 2h, -20°C	Urée + DTT + IAM	Trypsine, 37°C

DTT : Dithiothréitol ; IAM : Iodoacétamide ; ABC : Ammonium Bicarbonate

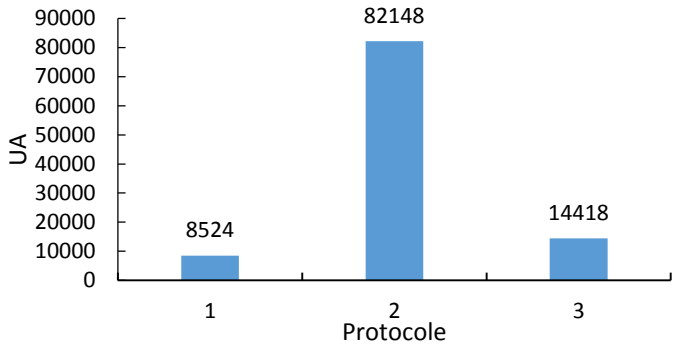


Figure 59 : Aire sous pic du glycopeptide pept-G0F analysé par HPLC-MS, après traitement des échantillons de culture par les 3 différents protocoles.

Sur la base de ces résultats, la Figure 60 synthétise le protocole retenu pour la préparation des échantillons et l'hydrolyse enzymatique des anticorps par la trypsine avant l'analyse des glycopeptides par HPLC-MS.

Figure 60 : Etapes du protocole de préparation des échantillons de culture, pour l'analyse de la glycosylation de l'IgG par HPLC-MS (ABC : bicarbonate d'ammonium)

Ce protocole a ensuite été utilisé pour analyser la glycosylation de l'IgG produite au cours d'une culture en mode discontinu. La distribution des différentes glycoformes de l'IgG, basée sur la distribution des différents glycopeptides, est représentée par la Figure 61, pour deux phases de la culture.

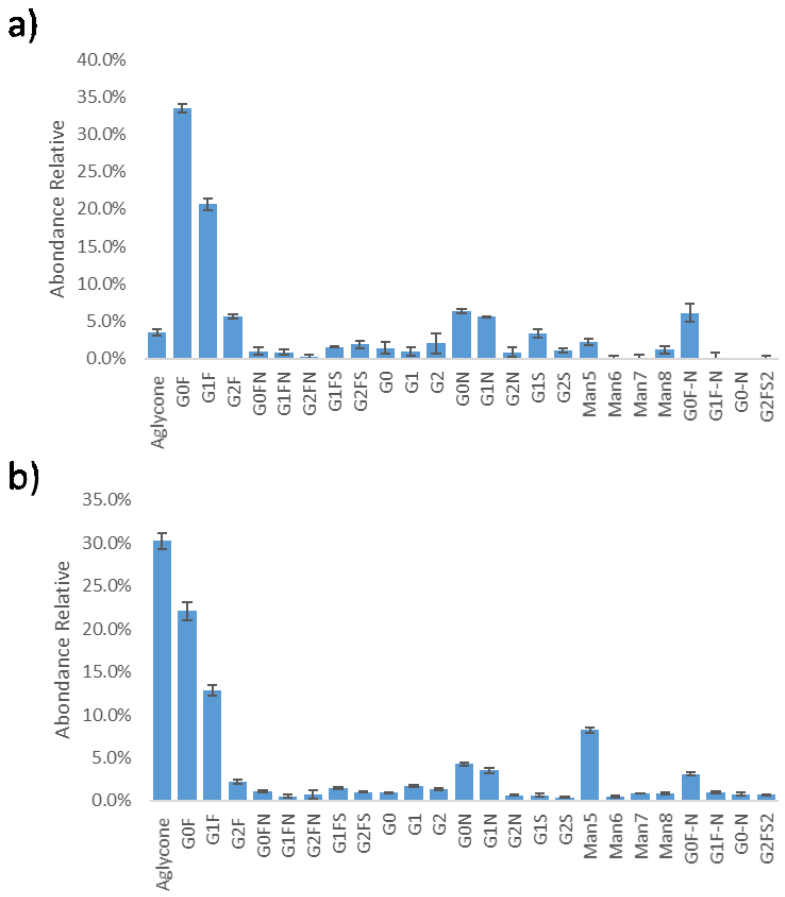


Figure 61 : Abondance relative des glycoformes des IgG produites : en phase exponentielle de croissance (a), et en phase de déclin cellulaire (b)

Les glycoformes présentant une abondance relative supérieure à 5% dans une de ces deux phases de culture sont considérées comme majoritaires. Elles correspondent aux formes suivantes : aglycone, G0F, G1F, G2F, G0N, G1N, Man5, et G0F-N. Nous constatons que les glycoformes des IgG produites pendant la phase de croissance de notre lignée cellulaire CHO, dans les conditions expérimentales utilisées, sont majoritairement fucosylées alors que leur niveau de galactosylation est relativement faible. La glycoforme G0F présente le pourcentage le plus élevé de toutes les glycoformes. Ces résultats sont en accord avec d'autres études sur la glycosylation des IgG produites par des cellules CHO (Eon-Duval *et al.*, 2012; del Val *et al.*, 2010a). En revanche, en fin de culture, la forme aglycone devient la plus abondante, en passant de 3,5% à 30,2%. Des résultats similaires ont été rapportés dans d'autres travaux, le pourcentage d'IgG aglycone pouvant monter jusqu'à 51% dans les cultures en cas d'épuisement du glucose (Villacrés *et al.*, 2015). Cette augmentation d'IgG non-glycosylée peut être liée à un niveau de concentration trop faible des substrats, dont le glucose ou la glutamine, en fin de culture. Dans ces conditions, la charge énergétique des cellules diminue et le substrat est

préférentiellement utilisé pour le métabolisme énergétique plutôt que pour réapprovisionner le pool de sucres nucléotidiques, précurseurs de la glycosylation (Nyberg *et al.*, 1999). La disponibilité de l'énergie pour la glycosylation de l'IgG est ainsi diminuée, et une glycosylation correcte ne peut plus être assurée (Kochanowski *et al.*, 2008). Par ailleurs, la glycoforme "high mannose", correspondant à une forme non maturée, augmente aussi de façon non négligeable en fin de culture (8,2%). Bien que le niveau de glycoformes "high mannose" soit généralement faible (< 5%) dans les IgG produites par des cellules animales (van Berkel *et al.*, 2009), certains auteurs montrent aussi une telle augmentation. Celle-ci pourrait être liée au niveau d'osmolarité du milieu de culture et à la durée de la culture (Pacis *et al.*, 2011), ou encore à une forte productivité spécifique en IgG (Zalai *et al.*, 2016).

Sur la base de ces premiers résultats, il est clair que la distribution des glycoformes des IgG évolue en cours de culture. Afin de mieux comprendre cette évolution de la distribution des glycoformes, les 8 glycoformes majoritaires ont été analysées pour chaque échantillon prélevé au cours d'une culture en mode discontinu (Figure 62). Entre 0 et 60 h, la concentration en IgG était trop faible pour l'analyse de glycosylation. L'abondance relative des formes aglycone et Man5 augmente au cours de la culture, tandis que celle des formes G0F, G1F et G2F diminue. Les pourcentages des 3 autres glycoformes (G0N, G1N, G0F-N) restent assez stables. C'est la forme aglycone, correspondant à l'IgG non glycosylée, qui présente la plus grande amplitude de variation (> 30%). Ces résultats indiquent donc clairement l'intérêt de suivre en ligne et de contrôler en temps réel la glycosylation de l'IgG, notamment sa forme aglycone, afin d'assurer une meilleure qualité du produit final.

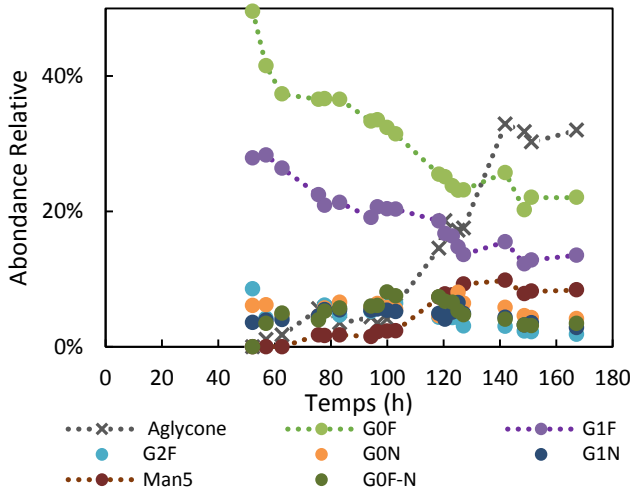


Figure 62 : Evolution des pourcentages des glycoformes majoritaires des IgG produites au cours d'une culture discontinue de cellules CHO

A l'heure actuelle, la glycosylation des IgG produites est généralement analysée en une seule fois en fin de procédé, par des méthodes hors ligne. Bien que pratique, cette approche peut masquer la

dynamique des changements intervenant dans la glycosylation des IgG en cours de culture. Ceci peut induire des risques de fortes déviations de la qualité du produit final par rapport au niveau de qualité attendu. L'observation de ces changements de glycosylation au cours du temps pourrait donc aider, non seulement, à mieux comprendre les mécanismes de glycosylation en cours de culture, mais aussi, à développer et optimiser les procédés de production pour assurer une meilleure qualité du produit. De plus, compte-tenu des exigences accrues en matière de comparabilité et de l'arrivée sur le marché des produits thérapeutiques biosimilaires, la capacité à suivre et à contrôler la qualité des produits en temps réel est devenue un objectif de plus en plus important pour les industries biopharmaceutiques (Tharmalingam *et al.*, 2015).

III.3. Suivi en ligne de la macro-hétérogénéité de la glycosylation des IgG par spectroscopie Raman

III.3.1. Introduction

Les résultats de la partie précédente ont montré que le profil de glycosylation des anticorps IgG, produits au cours d'une culture de cellules CHO en mode discontinu, peut varier de manière importante. En particulier, l'abondance relative de la forme non glycosylée (aglycone) augmente significativement durant la culture (de plus de 30%), conduisant à une diminution non négligeable des formes glycosylées et donc de la qualité de l'IgG produite. La présence ou l'absence de groupements glycanniques au niveau des sites de glycosylation de la molécule représente la macro-hétérogénéité de la glycosylation. C'est pourquoi, afin de pouvoir assurer le maintien de la qualité du produit au cours du procédé, il apparaît important de pouvoir suivre *in situ* et contrôler en temps réel la proportion de molécules d'IgG non-glycosylées, au cours des bioprocédés de culture de cellules animales. Néanmoins, comme indiqué dans l'étude bibliographique, les techniques d'analyses de la glycosylation des anticorps sont souvent longues et fastidieuses, et réalisées hors ligne ce qui ne permet pas de fournir des informations en temps réel pour réaliser un pilotage de la culture. Ainsi, il est essentiel de développer de nouvelles méthodes de suivi en ligne de la glycosylation des anticorps, afin de détecter d'éventuelles variations et de réagir rapidement pour éviter la détérioration de leur qualité.

La spectroscopie Raman a été proposée récemment comme étant un outil puissant pour le suivi en ligne de la concentration de différents composés présents dans le milieu de culture (cellules, glucose, lactate, glutamine, ions ammonium, anticorps, etc.) au cours de bioprocédés de culture de cellules

animales. Or, son intérêt pour le suivi de la qualité des produits d'intérêt n'a encore jamais été rapporté. C'est pourquoi, dans cette partie, nous nous proposons d'étudier le potentiel de la spectroscopie Raman pour le suivi en ligne de la macro-hétérogénéité de la glycosylation d'IgG produites au cours de cultures de cellules CHO. Deux modèles chimiométriques seront développés pour prédire respectivement la concentration totale des IgG et la concentration d'IgG non-glycosylées (aglycones). Les spectres Raman enregistrés en ligne seront tout d'abord calibrés par les valeurs de référence mesurées hors ligne. Puis, les performances de ces modèles seront évaluées par différentes techniques chimiométriques. Enfin, ces modèles seront utilisés pour prédire en temps réel l'évolution des concentrations des IgG totales et des IgG non-glycosylées au cours de deux cultures réalisées en mode discontinu et recharge-récolte.

Ce point a fait l'objet d'un article scientifique publié dans la revue "Biotechnology Progress" (34, 486-493,2018). Pour faciliter la lecture, le manuscrit de cette publication est présenté ci-dessous.

III.3.2. Manuscrit de la publication

Real-time monitoring of antibody glycosylation site occupancy by *in situ* Raman spectroscopy during CHO cell cultures in a bioreactor

Meng-Yao Li, Bruno Ebel, Cédric Paris, Fabien Chauchard, Emmanuel Guedon, Annie Marc

Abstract

The glycosylation of therapeutic monoclonal antibodies (mAbs), a known critical quality attribute, is often greatly modified during the production process by animal cells. It is essential for biopharmaceutical industries to monitor and control this glycosylation. However, current glycosylation characterization techniques involve time- and labor-intensive analyses, often carried out at the end of the culture when the product is already synthesized. This study proposes a novel methodology for real-time monitoring of the mAb glycosylation using Raman spectroscopy. It was first observed in CHO cell batch culture that when low nutrient concentrations were reached, a decrease in mAb glycosylation was induced, which made it essential to rapidly detect this loss of product quality. By combining *in situ* Raman spectroscopy with chemometrics tools, efficient prediction models were then developed for both glycosylated and non-glycosylated mAbs. By comparing Variable Importance in Projection profiles of the prediction models, it was confirmed that Raman spectroscopy is a powerful method to distinguish extremely similar molecules, despite the

high complexity of the culture medium. Finally, the Raman prediction models were used to monitor batch and feed-harvest cultures *in situ*. For the first time, it was demonstrated that the concentrations of glycosylated and non-glycosylated mAbs could be successfully and simultaneously estimated in real time with high accuracy, including their sudden variations due to medium exchanges. Raman spectroscopy can thus be considered as a promising PAT tool for feedback process control dedicated to on-line optimization of mAb quality.

Introduction

The market for therapeutic monoclonal antibodies (mAbs) has expanded remarkably over the past decade, reaching 85.4 billion USD in 2015, or nearly half the total sales of biopharmaceutical products (Monoclonal antibodies market size 2013). Among the host cells used to produce these recombinant mAbs, mammalian cells, in particular Chinese Hamster Ovary (CHO) cells, have emerged as the most often used expression system owing to their ability to perform human-like post-translational modifications (PTMs) (Zhu, 2012). Glycosylation is one of the most important PTMs for therapeutic mAbs. It corresponds to the addition of oligosaccharide structures to the polypeptide backbone at the conserved site (Asn 297) in each heavy chain of the mAbs. When produced in animal cells, glycan structures are usually of the biantennary and complex type (Higel *et al.*, 2016). Although the sugar moiety mass contribution can be very low compared with the entire mAb molecule (e.g. 2–3% of the molecular weight), it plays an important role in physical and chemical properties (e.g. folding, solubility, stability, and electrical charge) (Kayser *et al.*, 2011) and functionalities (e.g. *in vivo* half-life, immunogenicity, antibody-dependent cellular cytotoxicity and complement-dependent cytotoxicity) of the mAbs produced (Lingg *et al.*, 2012; del Val *et al.*, 2010a). Also, due to the operating conditions chosen for the cell culture processes, such as nutrient availability,(Villacrés *et al.*, 2015) dissolved oxygen level (Kunkel *et al.*, 1998), pH, temperature, and culture modes (Hossler *et al.*, 2009), changes occur in the intra- and extra-cellular environment that potentially influence glycosylation mechanisms and consequently the quality of mAbs (del Val *et al.*, 2010a; Villacrés *et al.*, 2015; Kunkel *et al.*, 1998; Hossler *et al.*, 2009; Butler, 2006). Glycosylation state is therefore considered as a critical quality attribute for therapeutical mAbs.

Quality by Design (QbD), a concept promoted by the regulatory authorities, encourages biopharmaceutical industries to build processes designed so as to yield consistent product quality throughout culture, instead of relying solely on quality testing at the end of the production process (Quality by Inspection) (Rathore and Winkle, 2009). To comply with this concept, the mAb production

processes must be carefully controlled and monitored in real time at all manufacturing steps using the Process Analytical Technology (PAT) approach to guarantee the safety and efficacy of the final product. However, off-line methods classically used to analyze mAb glycosylation in culture samples are time- and labor-intensive, involving several steps, such as glycan release by enzymatic digestion, subsequent labeling or derivatization, separation (often by chromatography-based methods) followed by analysis, usually by mass spectrometry (MS) (del Val *et al.*, 2010a)(Pais *et al.*, 2014) Huhn *et al.*, 2009). These methods delay results, which thus cannot be used to provide feedback information to control the cell culture process. The PAT approach therefore needs to be developed for real-time analysis of mAb glycosylation and better process control, while equaling or improving the accuracy and sensitivity of existing methods.

Spectroscopic technologies are being increasingly studied for the implementation of the PAT approach in the pharmaceutical industry. These sensitive, rapid, non-invasive, and easy-to-automate techniques can also provide simultaneous information on many process variables (Zhu, 2012 ; Teixeira *et al.*, 2009 ; Lourenço *et al.*, 2012). One of the most often used spectroscopic techniques for on-line monitoring of cell cultures is Raman spectroscopy, mostly for its well-known advantages, which include sharp spectral features and compatibility with aqueous systems. Signals from water molecules in Raman scattering are much weaker than in other spectroscopies such as infrared spectroscopy (Wen, 2007; Esmonde-White *et al.*, 2017). Its utility in real-time monitoring of nutrients, metabolites, by-products or cell densities during mammalian cell culture processes has been recently demonstrated in different studies (Abu-Absi *et al.*, 2011; Berry *et al.*, 2015, 2016; Matthews *et al.*, 2016; Whelan *et al.*, 2012a). Moreover, advances in Raman spectroscopic technique have led to predictive models for feedback process control by maintaining the concentration of glucose or lactate at set point levels (Matthews *et al.*, 2016; Berry *et al.*, 2016; Craven *et al.*, 2014). Yet despite this highlighted potential of Raman spectroscopy, only very few studies have demonstrated its application for monitoring recombinant antibody concentration (Ashton *et al.*, 2013; André *et al.*, 2015), and, to our knowledge, none has described real-time monitoring of mAb glycosylation inside a bioreactor equipped with an *in situ* sensor. Several authors have recently proposed off-line Raman spectroscopy as a promising tool for analyzing the glycosylation status of proteins (Brewster *et al.*, 2011; Cowcher *et al.*, 2016). However, these studies were carried out with small proteins, such as ribonucleases in solution in deionized water. Real-time monitoring of the glycosylation of large proteins, such as mAbs in a complex culture media, can thus still be very challenging for Raman spectroscopy: the cell culture media contain a broad range of entities whose concentrations vary in different ways in cell cultures, resulting in complex Raman spectral features. Consequently, it may be difficult to obtain accurate results for molecules such as mAbs present at

low concentrations. Moreover, it is still unclear whether the resolution of Raman spectroscopy is high enough to distinguish between glycosylated antibodies (G-mAbs) and non-glycosylated antibodies (NG-mAbs), as these molecules are extremely similar, with only 2–3% molecular weight differences in their structure. Considering these difficulties, careful application of appropriate chemometric methods is also required to extract the desired information from Raman spectral data.

In this context, the objective of this study was to explore and validate the potential of Raman spectroscopy for real-time monitoring of both concentration and glycosylation of mAbs in CHO cell cultures performed in a bioreactor. Adequate chemometric models were accordingly developed to allow the simultaneous estimation of G-mAbs and NG-mAbs concentration by using spectrometric data collected during cell cultures performed with various operating modes.

Materials and methods

Cell Culture

The genetically modified DG44 CHO cell line (CHO M250-9) producing a human anti-Rhesus D IgG, was kindly provided by Bioprocessing Technology Institute (Singapore). It was cultivated in a protein-free culture medium composed of a media mixture (1:1) of PF-CHO (HyClone) and CD CHO (Fisher Scientific), supplemented with 4 mM L-glutamine and 0.1% Pluronic F-68 (Sigma Aldrich). Cell cultures were performed in a 2.5 L bioreactor (Pierre Guérin) equipped with homemade light shield. Temperature was maintained at 37 °C, pH regulated at 7.2, dissolved oxygen at 50% of air saturation, and stirring rate set at 90 rpm. Six cultures were performed in a bioreactor under various operating conditions – batch, feed-harvest and spot glucose spiking – to increase environmental variability, to improve robustness for spectroscopic model calibration. Samples were taken 3–4 times per day, and a total of 115 samples were analyzed off-line for the calibration and validation of the Raman chemometric models.

Off-line analyses

The mAb concentration was quantified off-line using the automated photometric analyzer Gallery™ (Thermo Fisher Scientific). Turbidity of samples was measured after adding specific reagents containing anti-serum (Roche). For mAb glycosylation analyses, mAbs were first precipitated by cold acetone, incubated at −20 °C for 2 h, and centrifuged at 13000 *g* for 5 min. Recovered mAbs were dissolved in 500 μl of 50 mM ammonium hydrogen carbonate buffer, and heated to 95 °C for 10 min for denaturation. 20 μl of 1 $g.l^{-1}$ trypsin (Sigma-Aldrich) were then added to the cooled mAb samples. After incubation overnight at 37 °C, the reaction was stopped by adding 15 μl of formic acid (98%, Sigma-Aldrich). Peptides from mAb hydrolysis were analyzed by UHPLC-MS (ThermoFisher Scientific), combining a binary solvent delivery pump connected to a photodiode array detector and a Linear Trap Quadripole MS equipped with an atmospheric pressure ionization interface operating in positive electrospray mode. 10 μl samples were introduced into a C18 Alltima column (Grace/Alltech) with a flow rate of 200 $\mu l.min^{-1}$, using mobile phases of water modified with trifluoroacetic acid (0.1%) and acetonitrile modified with trifluoroacetic acid (0.1%). MS operating conditions were set as follows: spray voltage 5 kV; sheath, auxiliary and sweep gas 40, 10 and 10 arbitrary $units.min^{-1}$, respectively; capillary temperature 300 °C; capillary voltage 36 V; tube, split and front lens voltages 80 V, −44 V and −3.25 V, respectively. Full scan MS spectra (*m/z* 100–2000) and data-dependent MS/MS scans were then performed for structural investigation. Raw data were processed using the XCALIBUR software (Thermo Fisher Scientific). The percentage of each glycoform was obtained by dividing the area of its peak integral by the total glycopeptide signal. Analyzing the mAbs produced in this study at a concentration of around 300 mg/l, among a total of 25 potential mAb glycoforms reported in the literature (Sinha *et al.*, 2008; Selman *et al.*, 2010; Reusch *et al.*,

2015), the UHPLC-MS detected 7 glycoforms including non-glycosylated form, while 10 others were at the limit of detection (< 3% of total glycoforms), and 8 were not detected (data not shown). The concentration of NG-mAbs was calculated by multiplying the percentage of the non-glycosylated form by the total mAb (T-mAb) titer quantified off-line.

To gain fuller knowledge of the cell culture kinetics, additional culture parameters were also followed. Viable cell density (VCD) and cell viability were measured using the Vi-CELL XR™ cell counter (Beckman). A Gallery™ automated photometric analyzer (Thermo Fisher Scientific) was used to measure the concentrations of glucose, lactate, glutamine, and ammonium.

On-line acquisition of Raman spectra

The Raman system 'BioViserion®' (INDATECH model type VSR-HR-1-785) was used in this study for *in situ* and real-time analysis. The stabilizer 785 nm Laser source was set at 400 mW of power. The INDATECH Spectrograph sensor was a deep-cooled sensor camera (−50 °C) to provide long-term stability over weeks, but also low measurement noise with high integration time. This was particularly important since the media had a high concentration of water, which is not highly sensitive to the Raman effect. It was linked by optic fibers to an INDATECH probe body (model type SPC-OH-785-2) and an autoclavable GMP 200 mm immersion tube (model VSP-IMM-12S-05-200). The laser spot was focused shortly after the sapphire window (<0.5 mm working distance) so as not to be too sensitive to the turbidity parameter. The laser spot at the focal point was estimated at 100 μm (supplier information). The tube was immersed in the bioreactor about 5 cm from the bottom of the tank to avoid any artifacts. Raman spectra were acquired *in situ* with an acquisition duration time of 6 min (6 scans of 60 s), every 20 min for all the cultures. Spectral range of wavelengths was 100–3425 cm^{-1}; the region 800–1600 cm^{-1} was used as working spectral region, since it contained the major biochemical information (our process fingerprint region).

Chemometric methods for spectroscopic data processing

To develop models for T- and NG-mAbs, the chemometric analyses of the Raman spectra were carried out using PLS_Toolbox version 8.1 (Eigenvector Research Inc.) in MATLAB environment, version R2016a (MathWorks, Inc.). The dataset of spectra collected from six bioreactor cultures was separated into two independent sets using the Kennard-Stone algorithm: 80% for the calibration set and 20% for the prediction set. Among several techniques tested for spectra preprocessing, the one retained was as follows. The data were first treated using the standard normal variate method to scale the spectra of different cultures and reduce inter-batch variability. The spectra were then mean-centered to remove the variability of the devices. The values of off-line measurements of T- and NG-mAb concentrations were also mean-centered to improve model stability.

The Partial Least Squares (PLS) regression with venetian blinds cross-validation was adopted to develop the calibration models based on latent variables (LVs) linking the spectral data to the corresponding off-line measurements. A minimal LV number allowing a maximum explained variance for cross-validation was required to minimize overfitting. The performance of the calibration models was evaluated using various parameters: root mean square errors of calibration (RMSEC), cross-validation (RMSECV), and prediction (RMSEP), and square correlation coefficient (R^2) for calibration, cross-validation and prediction. In addition, to verify the robustness and applicability of the developed models, the importance of each variable (Raman shift) in the projection used in the PLS model was evaluated *via* the Variable Importance in Projection (VIP) score (André *et al.*, 2015). A variable with a VIP Score close to or greater than 1 can be considered as significant in a prediction model.

Results and discussion

Kinetics of antibody glycosylation during batch bioreactor culture

To ensure a better understanding of the evolution of mAb glycosylation during the cell culture process, a detailed kinetic study was first performed in a batch bioreactor. Off-line kinetics of the cell density and viability, of mAb concentrations (total, glycosylated and non-glycosylated forms), and of metabolite levels (glucose, glutamine, lactate, ammonium ions) are presented in Figure 1.

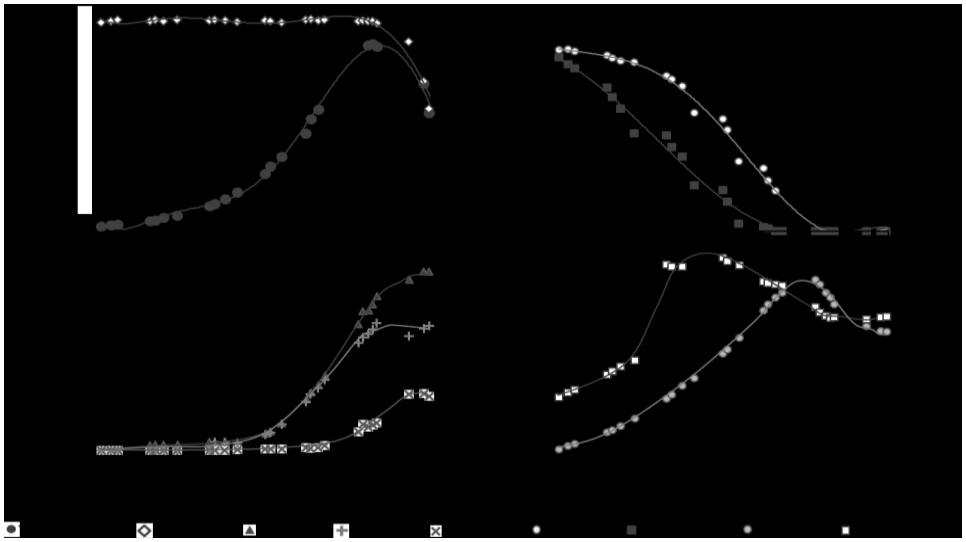


Figure 1. *Kinetics of CHO M250-9 cell culture in batch bioreactor: (A) viable cell density (●), viability (◇); (B) concentration of T-mAbs (▲), G-mAbs (✕) and NG-mAbs (+); (C): glucose (■) and glutamine (■) concentration; (D) lactate (■) and ammonium ion (■) concentration*

A maximal VCD of 8.5 × 10^6 cells.ml^{-1} was reached after 120 h of culture. During this period, the viability of cells was maintained near 100%. Both VCD and viability then decreased. Glutamine and glucose were progressively consumed and reached total depletion at 100 h and 120 h, respectively. The lactate concentration increased to a maximal value of 17.5 mM after 120 h, while ammonium ions accumulated up to a maximal concentration of 5 mM after 60 h of culture, before decreasing to 3.5 mM. Total mAbs produced increased with VCD, before slowing down at about 120 h of culture. Concerning the glycosylation of the mAbs, variations occurred at two major levels. The presence or absence of oligosaccharide structures at the glycosylation sites (Asn 297) is defined as macro-heterogeneity, while the presence of different oligosaccharide structures at one specific glycosylation site is referred to as micro-heterogeneity (del Val *et al.*, 2010 ; Eon-Duval *et al.*, 2012). In this study we found that the major variations in mAb glycosylation during the CHO cell culture in batch mode came from glycan site occupancy (>30%) compared with oligosaccharide structures (<15%) (data not shown). We therefore focused our study on macro-heterogeneity of the mAbs glycosylation, NG-mAbs being analyzed in each sample at the end of the cell culture. After about 100 h of culture, an accumulation of NG-mAbs began to appear, while cells still displayed a very good viability (>95%). At the end of the culture, after the depletion of both glutamine and glucose, the percentage of NG-

mAbs reached up to 32% of T-mAbs, resulting in a significant decrease in the quality of produced mAbs.

This macro-heterogeneity variability of the mAb glycosylation could be explained by a low energy metabolism of cells: the beginning of the accumulation of NG-mAbs after about 100 h coincided with the depletion of glutamine, one of the key energy carbon sources for cell growth. When glucose was completely consumed after 120 h, the production of G-mAbs stopped, while only NG-mAbs were still produced. A similar result has been observed for recombinant interferon-ϒ (IFN- ϒ) production by CHO cells (Kochanowski *et al.*, 2008). The authors suggested that the kinetic of IFN- ϒ production and its glycosylation were probably coupled with cell growth and cell energy status. When cells lacked energy for proliferation, they could begin to lose their capacity to correctly perform the post-translational modifications for synthesized extracellular proteins until protein synthesis finally stopped. Many other studies suggest that nutrient depletion during cell culture could be responsible for the decrease in the intracellular pool of nucleotides, resulting in variation in the macro-heterogeneity of protein glycosylation (Villacrés *et al.*, 2015; Nyberg *et al.*, 1999; Liu *et al.*, 2014; Xie *et al.*, 1997). Regarding lactate and ammonium accumulation, to our knowledge, no study has reported any inhibitory effect on glycosylation site occupancy when their concentrations are below 20 mM and 5 mM respectively.

These preliminary results thus clearly indicate that variability in macro-heterogeneity of recombinant mAb glycosylation occurred throughout CHO cell culture processes performed in batch mode. This variability mainly concerned the accumulation of non-glycosylated molecules, which could be related to low levels of energy substrates in the medium. To optimize the cell environment during cell culture processes and so avoid a sub-glycosylation of mAbs, ability to monitor this macro-heterogeneity of mAb glycosylation in real time may therefore be necessary.

Development of Raman prediction models for mAb glycosylation monitoring

Raman spectroscopy is recognized as a useful tool for the PAT approach applied to mammalian cell culture processes. However, it remains to be evaluated for mAb glycosylation. Development of efficient Raman calibration and prediction models is thus required. For this purpose, spectral data were collected *in situ* via a sterilizable sensor connected to a Raman spectrometer, during batch, feed-harvest and spot glucose spiking of CHO cell cultures in a bioreactor. In parallel, off-line analyses of the mAb concentration and glycosylation were performed 2–3 times a day, which resulted in a total of 115 samples. Firstly, the calibration models between on-line spectra and off-line data were built from the calibration set of 92 samples, using the PLS regression method after spectra pre-

processing as described in Materials and Methods. The calibration set included 46 samples from batch cultures, 17 samples from glucose spiking culture, and 29 samples from feed-harvest cultures. The number of LVs for both models was determined based on goodness of fit (R2Y) and goodness of prediction (Q2Y): the minimum number of LVs was obtained when both R2Y and Q2Y reached their maximal value. Figures 2-A and 2-B present the correlations observed between the concentrations of T-mAbs and NG-mAbs measured off-line and the values from on-line Raman spectra using the developed calibration models.

Both models performed well, on-line values being very close to the off-line measured ones, with almost no bias between the calibration regression line and the target 1:1 line. The remaining 23 samples were then used as the prediction dataset. Figures of merits for the models developed for T-mAbs (concentration range: 0-380 $mg.l^{-1}$) and NG-mAbs (concentration range: 0-98 $mg.l^{-1}$) are presented in Table 1.

The chosen LV numbers were 5 and 7 for T-mAbs and NG-mAbs respectively, consistent with LV numbers of Raman models for on-line prediction of different medium components reported in other studies (Abu-Absi *et al.*, 2011; Whelan *et al.*, 2012a; Berry *et al.*, 2015). Overall, both models displayed excellent performance with high square correlation coefficients, R^2, and low root mean square errors (RMSEC, RMSECV, RMSEP), representing less than 10% of the concentration ranges.

Table 1. *Statistical analysis of Raman models based on PLS regression for the prediction of T-mAb and NG-mAb concentrations.*

	Range	LV	R^2_C	RMSEC		R^2_{CV}	RMSECV		R^2_P	RMSEP	
	($mg.l^{-1}$)			($mg.l^{-1}$)	%		($mg.l^{-1}$)	%		($mg.l^{-1}$)	%
T-mAb	0–380	5	0.98	12.9	3	0.88	33.6	9	0.95	17.9	5
NG-mAb	0–97.6	7	0.99	1.8	2	0.89	7.0	7	0.95	5.7	6

The VIP scores were then calculated for each Raman shift to identify the most relevant spectral region for the two prediction models (Figure 2-C). VIP scores greater than 1 are typically considered significant, which means the corresponding spectra regions have most influence on the model (Chong and Jun, 2005).

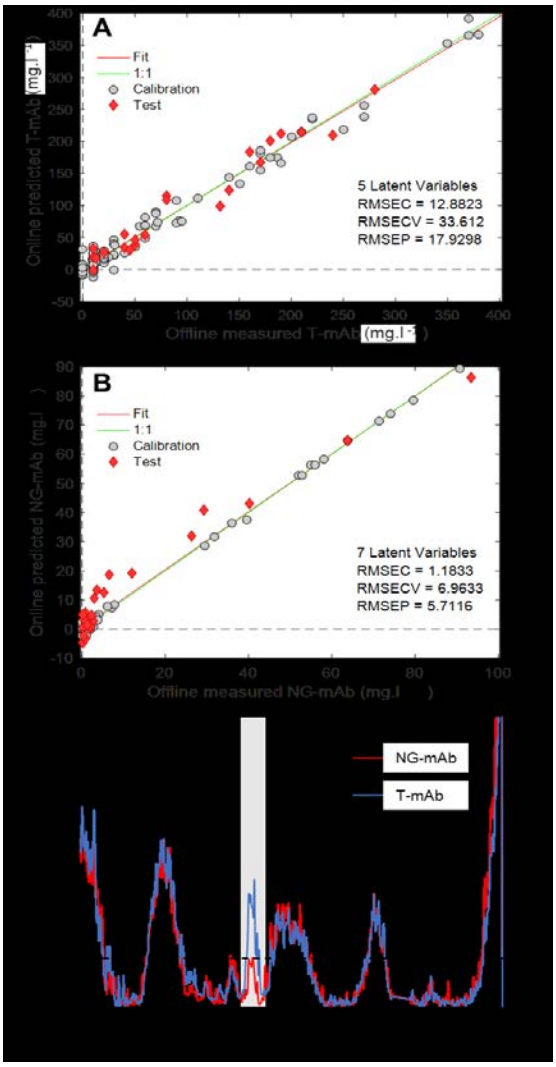


Figure 2. *Raman prediction models for the concentrations of (A) T-mAbs and (B) NG-mAbs (calibration set (●), prediction set (♦), linear regression of calibration set (——), target 1:1 line (——)); VIP scores of the models (C) for T-mAbs (blue line) and NG-mAbs (red line).*

Based on literature data, it was possible to relate the outlined spectral regions to the characteristic bands for some biological molecules. For example, the spectral region between 800 cm^{-1} and 850 cm^{-1} could be assigned to tyrosine (Cowcher *et al.*, 2016; Tuma, 2005; Wen, 2007), while the 930–980 cm^{-1} region and the 1209 cm^{-1} value could represent the skeleton stretching mode of the α-helix (Tuma, 2005; Wen, 2007). The vibration of phenylalanine may be seen at 950–1000 cm^{-1} and around 1205 cm^{-1}(Cowcher *et al.*, 2016; Wen, 2007). The amide III bands occurred in the interval 1235–1250 cm^{-1} (Wen, 2007; Weselucha-Birczyńska *et al.*, 2013). The stretching of the tryptophan indole ring could be found at around 1360 cm^{-1} (Lewis *et al.*, 2014; Wen, 2007), while the region 1554–1600 cm^{-1} could be assigned to amide II bands (Chi *et al.*, 1998; Weselucha-Birczyńska *et al.*, 2013). In the present study, the VIP score profiles for prediction models of T- and NG-mAb concentrations were

found to be very similar. This result could be expected, given that the biomolecules have very closely similar structures that can be distinguished only by a glycan group localized on one point of glycosylation. The major difference in the VIP profiles concerned the region centered at 1130 cm^{-1}, where VIP scores were less than 1 in the case of the NG-mAb prediction model, while VIP scores for the T-mAb prediction model were greater than 1. In this region, Raman spectroscopy is less influenced by characteristic bands of proteins and amino acids, and so spectral variations could be mainly attributed to certain glycan molecules (De Gelder *et al.*, 2007). According to the reference Raman database of biological molecules, the peaks at 1126 cm^{-1}, 1129 cm^{-1} and 1139 cm^{-1} can be assigned to *N*-acetyl glucosamine, fucose, and mannose respectively, three major glycan residues representative of mAb *N*-glycosylation (De Gelder *et al.*, 2007).

These results confirmed the ability of Raman spectroscopy to detect T-mAb and NG-mAb molecules separately, despite their extremely similar structures, which differ only by small glycan chains. Using PLS regression of Raman spectra acquired on-line during bioreactor cultures, it was then possible to develop well-performing prediction models for the concentrations of both T-mAbs and NG-mAbs in culture supernatant. Furthermore, by applying chemometric methodologies, the spectral region pertaining to the glycosylation differences between T-mAbs and NG-mAbs was identified.

On-line estimation of mAb glycosylation in various bioreactor cultures

The Raman models previously developed for T-mAb and NG-mAb concentrations were then used to estimate the mAb glycosylation on-line during CHO cell cultures performed in batch and feed-harvest modes. These culture modes induced progressive or point variations in the medium compounds, and so could be representative of industrial processes: industrial mAb production processes often use feed-harvest or fed-batch culture modes with spot medium replacement to increase final cell density, cell lifetime and productivity of recombinant proteins (Wlaschin and Hu, 2006). In this study, the feed-harvest culture was performed with three renewals of a 2/3 medium volume, corresponding to the time when the concentration of glucose and/or glutamine dropped below 5 mM and 1 mM respectively. Figure 3 details the kinetics of cell growth and antibody production observed in the two culture modes.

As shown in Figure 3.A, the feed-harvest mode allowed a slightly higher maximal VCD (9.8×10^6 $cells.ml^{-1}$) than the batch mode (8.8×10^6 $cells.ml^{-1}$). The maximal cumulated mAb level was increased by a factor of 2.2 in feed-harvest culture (750 mg) compared with batch culture (320 mg) (Figure 3.B). Concerning the evolution of mAb glycosylation, the cumulated quantity of NG-mAbs remained near zero until 350 h for the feed-harvest culture, while its accumulation corresponded to

the final phase when medium renewal was stopped. By contrast, accumulation of NG-mAbs began to increase after 100 h of culture in the batch mode. Therefore, to assure a better quality of harvested mAbs, it would be advisable to harvest the culture supernatant just before the increase in NG-mAbs, at 100 h and 350 h in batch and feed-harvest cultures respectively. This would ensure harvesting 140 mg in batch mode and 600 mg in feed-harvest mode of solely glycosylated mAbs. Thus these results also point out the need for such monitoring to anticipate the appropriate time for mAb harvesting, in addition to the utility of real-time monitoring of the mAb glycosylation time course to detect potential nutrient limitations.

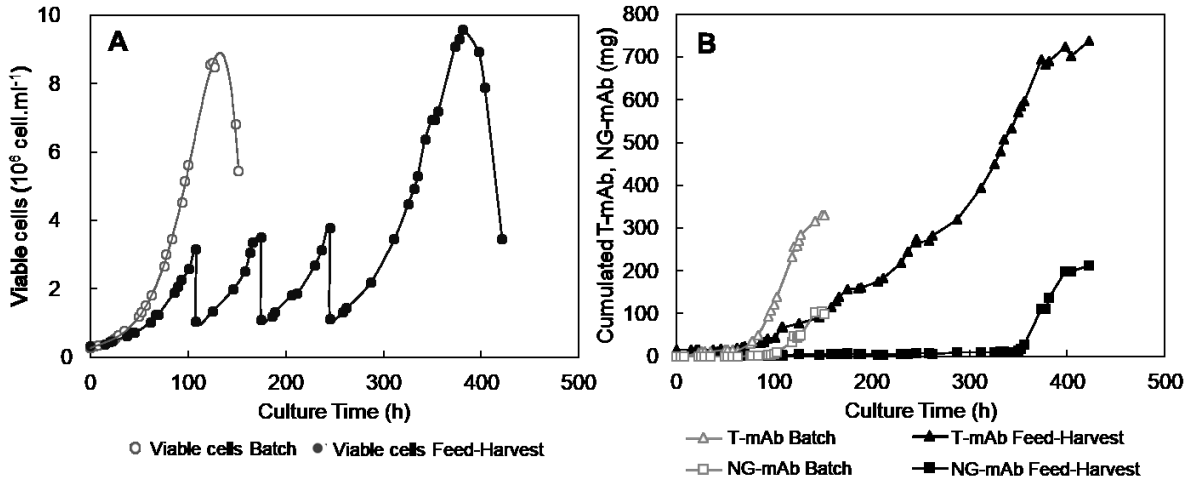


Figure 3. *Kinetics of mAb production and glycosylation during CHO M250-9 cell cultures: (A) viable cell density in batch (○) and feed-harvest (●) mode; (B) cumulated mAbs in batch (T-mAbs (△), NG-mAbs (□)), and in feed-harvest (T-mAbs (▲), NG-mAbs (■)) cultures.*

To validate the efficiency of the Raman models previously developed for the prediction of T-mAb and NG-mAb concentrations, we collected spectra every 20 min throughout a batch and a feed-harvest CHO cell culture, which amounted to a large number of spectra. Although these two cultures were involved in the development of the models, only new spectra without any reference off-line data were used in this part. After spectra processing, the models were used directly for the real-time estimation of T-mAb and NG-mAb concentrations, which were then compared with off-line analyses (Figure 4). Taking into consideration the accuracy of the off-line analyses for T-mAbs (about 10%) and NG-mAbs (15 - 20%), the trend of the concentration variations was correctly estimated from Raman data in both batch and feed-harvest cultures. Moreover, the rapid variations in T-mAb concentration induced by medium exchanges were also well monitored on-line.

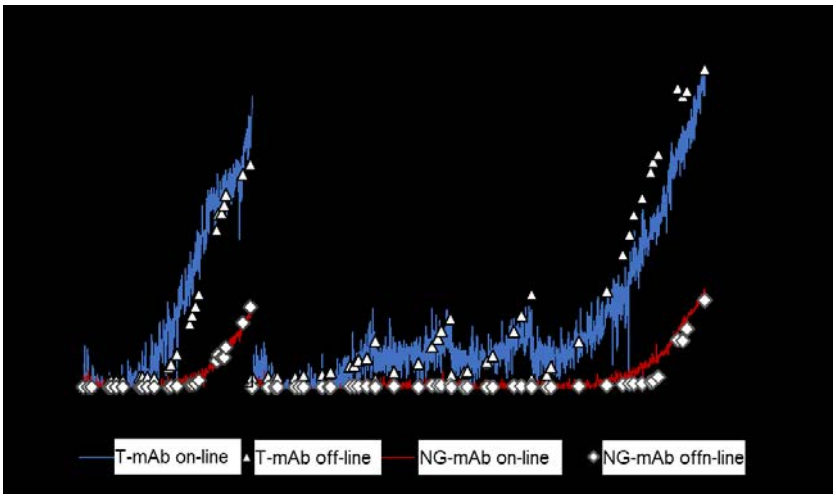


Figure 4. *On-line concentrations of T-mAbs (blue line) and NG-mAbs (red line) estimated from Raman spectroscopy data every 20 min, compared with off-line concentrations of T-mAbs (▲) and NG-mAbs (◆) in batch and feed-harvest cultures.*

These results validate the prediction accuracy and robustness of the Raman models developed in this study for the concentration and the glycosylation of the mAbs produced. They support the choice of working spectral region and preprocessing techniques. Furthermore, they indicate that the calibration set successfully captured the process variability. Only one paper in the literature recently demonstrated the prediction capability of Raman spectroscopy to monitor the mAb concentration in CHO cell cultures in real time. In this case, chemometric models were successfully applied to three batch cultures, and the mAb concentrations could be well predicted compared with those measured off-line (André *et al.*, 2015). However, the present study not only confirms that Raman spectroscopy is a powerful tool for mAb concentration monitoring, but also demonstrates, for the first time, its potential to monitor the glycosylation quality of produced mAbs in real time. Furthermore, we note that during the feed-harvest process, despite marked changes in the composition of culture medium occurring within a few minutes right after the medium renewal, and resulting in more complex spectral features (Petiot *et al.*, 2010), the models developed remained able to predict, accurately and simultaneously, the sudden variations in T-mAb level and the progressive increase in NG-mAbs at the end of the culture. Raman spectroscopy was able to differentiate between two extremely similar molecules, their variations displaying completely different profiles. Moreover, the critical time corresponding to the beginning of NG-mAb accumulation could be rapidly detected in both cultures.

Conclusion

Large variations in macro-heterogeneity of mAb glycosylation are observed throughout CHO cell culture bioprocesses, leading to NG-mAb accumulation and significant loss of quality in the mAbs produced. It was therefore essential to implement monitoring of the mAb glycosylation in real time. *In situ* Raman spectroscopy combined with appropriate chemometric methods proved to be a useful tool not only for real-time monitoring of mAb concentration throughout CHO cell culture processes, but also to distinguish between glycosylated and non-glycosylated molecules. Using an immersion probe connected to a Raman spectrometer by an optical fiber, prediction models of T-mAb and NG-mAb concentration were first developed. Their efficiency and robustness were then validated with batch and feed-harvest cultures. The trends of T-mAb and NG-mAb concentration were correctly predicted in both culture modes, despite the complexity of the culture medium and its composition variations during the processes. Furthermore, these on-line Raman estimations, performed at very short time intervals of 20 minutes, could greatly help process control to optimize the productivity and quality of therapeutic antibodies, either by real-time decision-making of medium feed, or by anticipated harvest of the product. By helping progress in the QbD and PAT approaches, these new results open a field of future investigations for the implementation of real-time control strategies dedicated to ensuring the quality of recombinant proteins produced in various bioprocesses performed in batch, fed-batch or perfusion modes.

References *(pour alléger le document, les références complètes sont rassemblées à la fin du mémoire)*

Abu-Absi *et al.* (2011)
André, S. *et al.* (2015)
Ashton, L. *et al.* (2013)
van Berkel, P.H.C. *et al.* (2009)
Berry, B. *et al.* (2015)
Berry, B.N. *et al.* (2016)
Brewster, V.L. *et al.* (2011)
Butler, M. (2006)
Chi, Z. *et al.* (1998)
Chong, I.G., and Jun, C.H. (2005)
Clavaud, M. *et al.* (2013)
Cowcher, D.P. *et al.* (2016)
Craven, S. *et al.* (2014)
De Gelder, J. *et al.* (2007)
Eon-Duval, A. *et al.* (2012)
Esmonde-White, K.A. *et al.* (2017)
Higel, F. *et al.* (2016)
Hossler, P. *et al.* (2009)
Huhn, C. *et al.* (2009)
Kayser, V. *et al.* (2011)
Kochanowski, N. *et al.* (2008)
Kunkel, J.P. *et al.* (1998)
Lewis, E.N. *et al.* (2014)
Lingg, N. *et al.* (2012)
Liu, B. *et al.* (2014)
Liu, H.F. *et al.* (2010)
Lourenço, N.D. *et al.* (2012)
Matthews, T.E. *et al.* (2016)
Nyberg, G.B. *et al.* (1999)
Pacis, E. *et al.* (2011)
Pais, D.A.M. *et al.* (2014)
Petiot, E. *et al.* (2010)
Planinc, A. *et al.* (2016)
Rathore, A.S., and Winkle, H. (2009)
Reusch, D. *et al.* (2015)
Selman, M.H. *et al.* (2010)
Sinha, S. *et al.* (2008)
Teixeira, A.P. *et al.* (2009)
Tharmalingam, T. *et al.* (2015)
Tuma, R. (2005)
del Val, I.J. *et al.* (2010)
Villacrés, C. *et al.* (2015)
Wen, Z.-Q. (2007)
Weselucha-Birczyńska, A. *et al.* (2013)
Whelan, J. *et al.* (2012)
Wiercigroch, E. *et al.* (2017)
Wlaschin, K.F., and Hu, W.S. (2006)
Xie, L. *et al.* (1997)
Zalai, D. *et al.* (2016)
Zhu, J. (2012)
MAbs Market Size, Ind Rep, 2013-2014

III.3.3. Contributions de la publication

Les travaux de cet article présentent le développement d'une méthode de suivi en temps réel de la macro-hétérogénéité de la glycosylation des IgG produites au cours de cultures de cellules CHO, en couplant la spectroscopie Raman avec des techniques chimiométriques.

Dans un premier temps, l'article reprend certains des résultats discutés dans la partie précédente montrant l'accumulation des IgG non-glycosylées pendant les cultures de cellules CHO en mode discontinu. Ceci entraîne une diminution significative de la qualité de l'IgG produite, ce qui démontre l'intérêt de suivre en temps réel la macro-hétérogénéité de la glycosylation de l'IgG.

En combinant la spectroscopie Raman avec les techniques chimiométriques, deux modèles ont été développés à partir des spectres Raman enregistrés *in situ* au cours de 6 cultures de cellules en bioréacteur. Il a été montré que ces 2 modèles présentent de bonnes capacités de prédiction de la concentration des IgG totales et des IgG non-glycosylées, avec des précisions relativement hautes. Les scores VIP des modèles ont montré que la spectroscopie Raman est assez résolutive pour distinguer des molécules très similaires (IgG glycosylées ou non).

Enfin, la méthode de suivi en ligne développée a été utilisée pour suivre en temps réel la concentration des IgG totales et des IgG non-glycosylées au cours de deux cultures réalisées en mode discontinu et recharge-récolte. Dans les deux cas, les évolutions des concentrations ont été correctement évaluées en ligne, par comparaison aux valeurs de référence mesurées hors ligne.

Ainsi, pour la première fois, cet article démontre le potentiel de la spectroscopie Raman pour le suivi en temps réel de la qualité de l'IgG produite, en particulier de la macro-hétérogénéité de sa glycosylation. La prochaine étape serait d'intégrer les modèles développés dans des stratégies de rétrocontrôle pour piloter en temps réel la qualité de l'IgG au cours des différents types de procédés de culture de cellules animales.

III.4. Suivi en ligne de la micro-hétérogénéité de la glycosylation des IgG par spectroscopie Raman

III.4.1. Introduction

Comme montré dans la partie précédente, la spectroscopie Raman s'est révélée un outil intéressant pour suivre en temps réel la macro-hétérogénéité de la glycosylation des IgG produites en culture, car elle parvient à distinguer les IgG glycosylées des IgG non-glycosylées. Pour approfondir cette étude, nous proposons d'étudier la faisabilité du suivi en ligne de la micro-hétérogénéité de glycosylation de l'IgG par spectroscopie Raman. Rappelons que la micro-hétérogénéité correspond à l'hétérogénéité des structures des groupements glycanniques fixés sur le site de glycosylation. De ce fait, les différences de structure et de masse molaire entre les glycoformes d'une IgG glycosylée sont encore plus faibles que celles existant entre une IgG glycosylée et une IgG non-glycosylée.

La qualité de glycosylation des IgG est souvent caractérisée en fonction de leur état de galactosylation, fucosylation, sialylation et la présence de groupes "high mannose" (Eon-Duval *et al.*, 2012). De ce fait, les 25 glycoformes analysées par HPLC-MS ont tout d'abord été classées en 4 groupes correspondant à ces états de glycosylation, chaque glycoforme pouvant participer à un ou plusieurs groupes (Tableau 14).

Tableau 14 : Participation des glycoformes d'IgG dans chaque groupe

Groupe de glycoformes	Fucosylées	Galactosylées	Sialylées	Fortement Mannosylées
glycoformes	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, G2FS2, G1S, G2S, G1FS-N	Man5, Man6, Man7, Man8

Des modèles de prédiction en ligne pour le suivi de la micro-hétérogénéité de glycosylation ont alors été obtenus en calibrant les spectres Raman acquis en ligne avec les concentrations des 4 groupes de glycoformes mesurées hors ligne. Les modèles développés ont ensuite été utilisés pour suivre en temps réel l'évolution de la micro-hétérogénéité de la glycosylation de l'anticorps au cours d'une culture en mode discontinu.

Ce point a fait l'objet d'un article scientifique (short note) prochainement soumis dans la revue "*Biotechnology Journal*". Pour faciliter la lecture, le manuscrit de cette publication est présenté ci-dessous.

4.2. Manuscrit de la publication

Real-time monitoring of different antibody glycoforms during cell culture processes by using *in situ* Raman spectroscopy

Meng-Yao Li, Bruno Ebel, Cédric Paris, Fabien Chauchard, Emmanuel Guedon, Annie Marc

Abstract

Glycosylation is a critical quality attribute of recombinant monoclonal antibodies since it impacts their safety and therapeutical efficacy. Subject to high variabilities during the manufacturing process, the antibody glycosylation should be carefully monitored and controlled. However, current off-line analysis practice requires a large time and only allows a quality by testing approach at the end of the process, which is inadequate for real-time control strategies. For the first time, we proposed a quality by design approach to monitor in real-time the various glycoforms of an antibody produced by CHO cells, by using *in situ* Raman spectroscopy combined with chemometric methods. Based on on-line Raman spectra and off-line HPLC-MS analyses, models were developed to predict with high accuracy the concentrations of four major groups of antibody glycoforms, which were fucosylated, galactosylated, sialylated or with presence of high mannose structures. Applied to independent Raman spectra recorded during a batch process, the developed models revealed efficient real-time monitoring of the antibody glycoforms time profiles. These results represent a new meaningful step towards an automated real-time feedback control of animal cell culture processes in order to better ensure the quality of produced antibody.

1 Introduction

During the last decades, production of therapeutic monoclonal antibodies (mAbs) have grown significantly to become dominant in the global biopharmaceutical market [1]. However, this commercial success relies on robust and cost-effective production platforms, which are currently based on mammalian cell culture systems, due to their ability to perform human-like post-

translational modifications [2]. One of the critical quality attributes (CQAs) for therapeutic mAbs is the glycosylation, which is known to modulate several key properties of the protein, such as efficacy, solubility, stability, bioactivity, immunogenicity and pharmacokinetics [3]. Yet, the glycosylation profile of mAbs, including the glycosylation site occupancy (macro-heterogeneity) and the sugar structure of attached glycans (micro-heterogeneity), is intrinsically subject to variabilities during the manufacturing processes [4]. Therefore, the glycosylation of mAbs should be monitored continuously during the cell culture. However, until this day, the general procedures used for evaluating the mAb glycosylation profile, are still mainly based on labor-intensive and time-consuming techniques that require withdraw and preparation of samples, which are only conducted at the end of the manufacturing process (quality by testing). Such approach bears a high risk of rejecting the product because of the non-conforming final product quality.

As a consequence, the Food and Drug Administration launched the Process Analytical Technology initiative in 2004, which aimed to guide the biopharmaceutical industries to move from the quality by testing to a more proactive quality by design (QbD) approach, by using on-line monitoring systems to provide better product and process understandings, thus mitigating risks by process control [5]. In recent years, serious efforts have been undertaken towards automated and at-line glycosylation analysis. As examples, automated robotic platforms have been reported for glycosylation profiling of mAbs within 90 min in a 96-well plate format by using micro sequential injection (μSI) systems coupled with ultra-performance liquid chromatography (UPLC) and mass spectrometry (MS) [6, 7]. While significantly reducing the time of analysis, these methods still need complex sample preparation steps and potentially increase the contamination risks. To avoid such inconvenient, *in situ* vibrational spectroscopies, Raman spectroscopy in particular, have been proved to be very promising tools due to non-invasive nature and rapid and sensitive measurements [8]. Nevertheless, the use of Raman spectroscopy has been mostly restricted to the monitoring of cell culture medium component concentrations (*e.g.* cell, glucose, glutamine, lactate, ammonium, antibody…) [9–11], while very few studies have addressed its ability to assess in real-time the CQAs of mAbs produced in animal cell cultures. Our recent work has demonstrated the ability of *in situ* Raman spectroscopy to distinguish between glycosylated and non-glycosylated mAb in complex cell culture media during bioreactor CHO cell cultures [12]. However, its capability to monitor the micro-heterogeneity of mAb glycosylation still needs to be confirmed.

In this paper, the real-time and simultaneous monitoring of different mAb glycoforms was developed by using an *in situ* Raman spectroscopy combined with chemometrics techniques, providing results with no time delays and no need for sample withdraws. In particular, we demonstrated the ability of

Raman spectroscopy to monitor *in situ* the mAb glycosylation in terms of fucosylation (mAb-fuc), galactosylation (mAb-gal), sialylation (mAb-sia) and high-mannose (mAb-man) forms.

2 Materials and methods

2.1 Cell culture

The DG44 CHO cell line (CHO M250-9, kindly provided by Bioprocessing Technology Institute of Singapore) producing an anti-Rhesus D antibody of type IgG was cultured in a 2.5 L bioreactor (Pierre Guérin) with a homemade light shield. Six cultures were performed under various operating conditions including batch, feed-harvest and spot glucose spiking modes. Details of cell culture conditions were described in a previous study [12].

2.2 Off-line glycosylation analysis

The glycosylation profile of the mAb produced was analyzed by the UPLC-MS system (ThermoFisher Scientific, USA) after trypsin (Sigma-Aldrich, USA) digestion. Details of this part can be found in our previous work [12]. The percentage of each glycoform was obtained by dividing the area of its peak integral by the total glycopeptide signal. The 25 glycoforms analyzed were gathered in 4 groups depending on the sugar moiety of their glycans: fucosylation (mAb-fuc), galactosylation (mAb-gal), sialylation (mAb-sia) and the presence of high mannose structures (mAb-man). Thereby, each glycoform could participate several groups. Total mAb concentrations were quantified off-line using the automated photometric analyzer Gallery™ (ThermoFisher Scientific, USA). The concentration of each glycoform group was calculated by multiplying the total of glycoform percentages by the total mAb concentration.

2.3 On-line Raman spectra acquisition

The BioViserion® Raman spectrometer (INDATECH model type VSR-HR-1-785) was equipped with a 785 nm laser source linked by optic fibers to a stainless steel immersion probe (model SPC-OH-785-2). An acquisition time of 6 min including 6 scans of 60 s was adopted. As previously described, the working spectral region 800-1600 cm^{-1} was selected, while data in 1392-1422 cm^{-1} range were removed due to light interference [12].

2.4 Chemometric analysis

After outlier removal, spectral data were separated into calibration (80%) and prediction (20%) datasets by Kennard-Stone algorithm. They were pre-processed by standard normal variates (SNV) normalization and mean centering, before performing model calibration using partial least squares (PLS) regression. A minimal number of latent variables (LVs) was retained with a maximum explained variance for cross-validation (using a Venetian blinds approach) and a minimum root mean square error of cross-validation (RMSECV), in order to minimize overfitting.

3 Results & Discussion

The primary goal was to develop an on-line analytical method of the mAb glycosylation micro-heterogeneity during cell culture processes by using Raman spectroscopy. Four groups of mAb glycoforms (mAb-fuc, mAb-gal, mAb-sia, mAb-man), having significant influence on the safety and efficacy of mAbs [13], were considered. The Raman spectra collected *in situ,* which coincided with off-line samples, were used to develop the PLS models to predict the concentration of the 4 mAb

glycoform groups in cell culture supernatant. Variations resulting from different culture modes (batch, glucose spiking, feed-harvest) were used to generate the calibration dataset in order to expand the design space. The Raman models were built with 92 calibration samples, while 23 other samples were used for the prediction.

Model calibration and prediction performances are summarized in Table 1. The numbers of LVs obtained for the 4 models were very moderate in the range of 3-5, indicating that the models developed used biochemical information of the culture medium rather than signal noises, and therefore, had less overfitting risks [14]. As can be seen by the high coefficients of determination (R^2_{Cal}), and low RMSE of calibration (RMSEC), the models displayed good calibration capabilities. The average RMSECV for the 4 groups of mAb glycoforms was 9% of their concentration ranges, demonstrating high model stability. The prediction abilities were evaluated by applying the models on the prediction dataset. The RMSE of prediction (RMSEP) was similar to the RMSECV for the 4 glycoform groups, indicating that the design space of the calibration dataset had successfully captured the process variability. Overall, the values of RMSEC, RMSECV, and RMSEP for all the models were comparable and in the same range, demonstrating their robustness. All these results show a significant predictive power of the developed models.

Table 1. *Performances of calibration and prediction models for the concentrations of four groups of mAb glycoforms produced during CHO cell cultures.*

	LV	R^2_{Cal}	R^2_{CV}	R^2_{Pred}	RMSEC	RMSECV	RMSEP
mAb-fuc	5	0.97	0.80	0.84	9.6 (4%)	25.8 (10%)	27.2 (10%)
mAb-gal	3	0.91	0.86	0.89	9.3 (6%)	11.4 (8%)	11.6 (8%)
mAb-sia	3	0.90	0.86	0.90	1.3 (7%)	1.6 (9%)	1.6 (9%)
mAb-man	5	0.98	0.86	0.87	1.1 (3%)	3.3 (9%)	3.9 (11%)

**RMSEC, RMSECV and RMSEP expressed in mg/L and in percentages*

Figure 1 compares the off-line measured values with the values obtained by the on-line prediction models. Overall, good agreements between measured and predicted values were obtained. However, the prediction for higher concentrations was not as successful as that for lower concentrations. In fact, fewer samples at higher concentrations were used in the model calibration. Moreover, the higher mAb glycoform concentrations corresponded to the culture stage at which cells began to lose viability. The increase of cell debris or waste products could trigger larger culture medium autofluorescence, and therefore influence the PLS regression results.

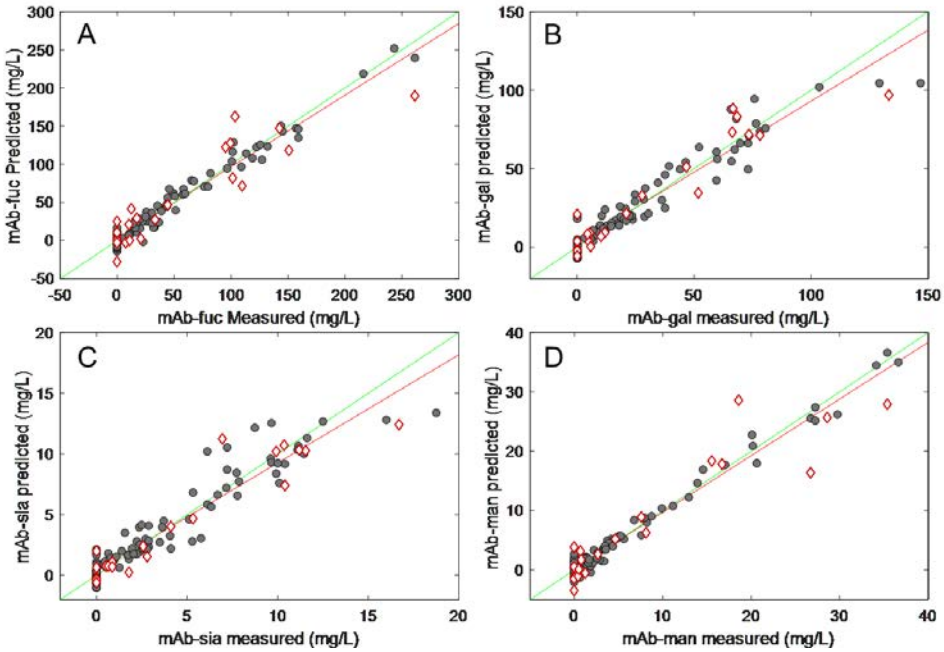


Figure 1. *Comparison of off-line measured data and on-line Raman predicted values provided by the PLS models for (A) fucosylated, (B) galactosylated, (C) sialylated, and (D) high mannose groups of mAb glycoforms. Regression line in red, target 1:1 line in green, calibration dataset (●), prediction dataset (◇).*

The next step was to challenge the PLS models developed previously by applying them to a new batch cell culture for evaluating the real-time monitoring of the product glycoforms. By recording Raman spectra *in situ* every 20 min, we examined the model performance to monitor variations of the micro-heterogeneity of the mAb glycosylation during cell culture process, in which 225 mg/L total mAb and 168 mg/L glycosylated mAb were achieved. Predictions from the PLS models based on Raman spectra were plotted along with the off-line measurements in Figure 2. Raman predictions appeared to follow the expected trends of the 4 mAb glycoform groups, since the predicted values were close to the off-line measured values. Nevertheless, the prediction accuracy of mAb-fuc concentrations was relatively lower than for the other groups. The concentration values of the mAb-

fuc seemed to be overestimated by the PLS model. One possible explanation could be that the mAbs produced by CHO cells were highly fucosylated (around 70%), and that little variations in mAb fucosylation were observed during cell cultures [15]. Therefore, the time profiles of total glycosylated mAbs and mAb-fuc were strongly correlated. This could lead to lower model prediction abilities. The quality of predictions could be improved, either experimentally, by increasing the number of calibration samples with more process variations, or numerically, by evaluating other chemometric methods such as local weighted regression, support vector regression or artificial neuron network.

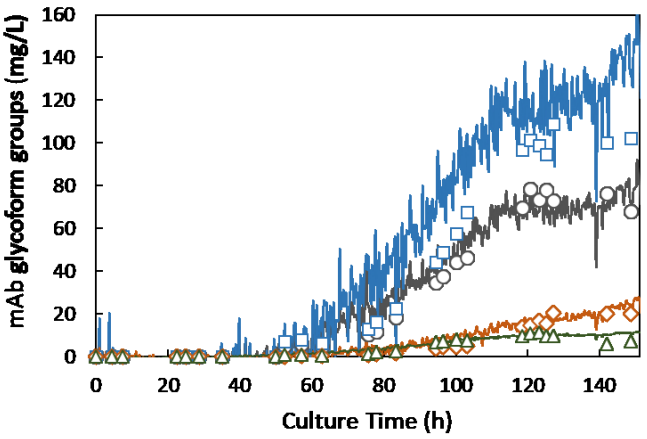


Figure 2. *Evolution with time of mAb glycoform concentrations during a CHO cell batch culture: real-time Raman prediction (lines) and off-line measured values (symbols) for the groups of fucosylated (—, □), galactosylated (—, ○), syalilated (—, ◇), and high mannose (—, △) mAb glycoforms.*

4 Concluding remarks

A novel method has been developed to monitor during a CHO cell culture the micro-heterogeneity of the produced mAb glycosylation by using *in situ* Raman spectroscopy. PLS models were built on the basis of Raman spectra to predict the concentrations of 4 major groups of mAb glycoforms (mAb-fuc, mAb-gal, mAb-sia and mAb-man) with high prediction accuracy. By applying the prediction models to new Raman spectra recorded in real-time during a batch culture, the general trends of the 4 mAb glycoform groups were successfully monitored. For the first time, the micro-heterogeneity of the mAb glycosylation could be monitored in real-time without any sample withdraws or preparations. By using *in situ* Raman spectroscopy, we were able to build a completely sealed system, decreasing the risk of contaminations, minimizing additional resource demand and shortening process development time. By providing product quality information in real-time, these results represent a

first step towards feedback control of animal cell culture processes, and could help for the broad adoption of QbD approach.

5 References *(pour alléger le document, les références complètes sont rassemblées à la fin du mémoire)*

[1] Lagassé, H. A. D. *et al.* (2017)
[2] Pais, D. A. M. *et al.* (2014)
[3] Higel, F. *et al.* (2016)
[4] Butler, M. *et al.* (2005)
[5] Rathore, A. S. *et al.* (2016)
[6] Burnina, I. *et al.* (2013)
[7] Tharmalingam, T. *et al.* (2015)
[8] Lourenço, N. D. *et al.* (2012)
[9] Abu-Absi, N. R. (2011)
[10] André, S. *et al.* (2015)
[11] Li, M-Y. *et al.* (2018)
[12] Li, M-Y. *et al.* (2018)
[13] Eon-Duval, A. *et al.* (2012)
[14] Whelan, J. *et al.* (2012)
[15] del Val, I. J. *et al.* (2010)

III.4.3. Contribution de la publication

Les résultats de cet article ont montré le potentiel de la spectroscopie Raman pour le suivi en ligne de la micro-hétérogénéité de la glycosylation de l'anticorps IgG dans les procédés de culture de cellules animales. Des modèles chimiométriques ont été développés à partir des spectres Raman pour prédire la concentration de 4 groupes de glycoformes de l'IgG (fucosylées, galactosylées, sialylées, et fortement mannosilées) avec une bonne précision (< 15% d'erreur). En appliquant cette méthode à une culture en mode discontinu, les tendances d'évolution de ces 4 groupes de glycoformes ont été suivies avec succès.

A notre connaissance, c'est la première fois que la micro-hétérogénéité de la glycosylation des IgG a pu être suivie *in situ* et en temps réel au cours de cultures de cellules animales en bioréacteur. En intégrant les résultats de la partie précédente, il devient ainsi possible d'analyser simultanément en ligne la macro- et la micro-hétérogénéité de la glycosylation des IgG. Ces résultats représentent un pas important vers une stratégie de rétrocontrôle automatique visant à optimiser les procédés de culture de cellules animales pour assurer la qualité du produit final.

III.4.4. Résultats supplémentaires non intégrés dans l'article

- Analyse des scores VIP des modèles de micro-hétérogénéité de glycosylation des IgG

Afin d'identifier les régions spectrales les plus pertinentes pour le suivi de la micro-hétérogénéité de la glycosylation des IgG, les scores VIP de chaque modèle de prédiction des 4 groupes de glycoformes ont été calculés (Figure 63).

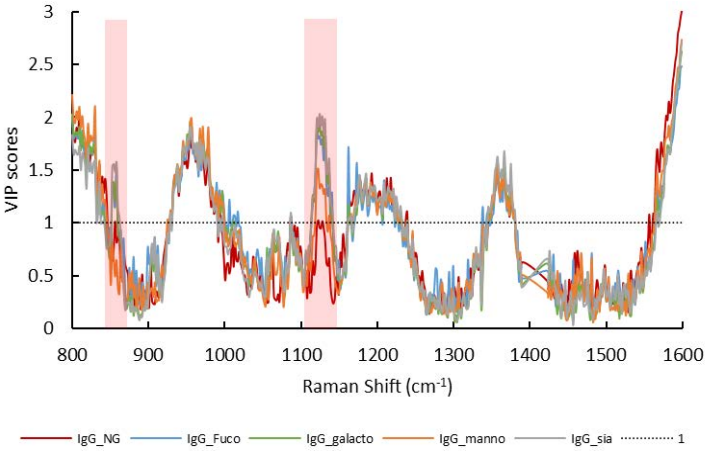


Figure 63 : Scores VIP des modèles de prédiction de la micro-hétérogénéité de la glycosylation des anticorps par spectroscopie Raman

En les comparant avec le profil VIP du modèle des IgG non-glycosylées (IgG-NG), nous pouvons d'abord confirmer que la zone spectrale autour de 1130 cm^{-1}, déjà identifiée dans la partie traitant de la macro-hétérogénéité de glycosylation, correspond bien aux bandes caractéristiques des résidus glucidiques de l'IgG, puisque les scores VIP des modèles de micro-hétérogénéité de la glycosylation sont supérieurs à 1 dans cette zone spectrale. Cependant, la distinction entre les profils de scores VIP pour les 4 modèles de micro-hétérogénéité de glycosylation est difficile, car ces 4 groupes de glycoformes peuvent contenir des résidus glucidiques en commun. En effet, toutes les glycoformes contiennent au moins 2 résidus de N-acétylglucosamine (GlcNAc) et 3 résidus mannose, ce motif étant souvent appelé le noyau de la N-glycosylation. De plus, certaines glycoformes peuvent appartenir à plusieurs groupes comme, par exemple, la glycoforme G1FS qui est, à la fois, fucosylée, galactosylée et sialylée. Seul le groupe des IgG de type "high mannose" ne contient que des résidus mannose et GlcNAc. Nous avons constaté que les scores VIP du modèle de prédiction pour ce type ("high mannose") sont quelque peu différents de ceux des modèles des autres glycoformes dans la zone autour de 856 cm^{-1} où ils sont inférieurs à 1. Selon la littérature, les signaux Raman dans cette zone spectrale peuvent être attribués aux élongations, ou stretchings, de certaines liaisons glucidiques (COC, CC, CH) des polysaccharides (Wiercigroch *et al.*, 2017). Dans notre cas, les signaux Raman dans cette zone pourraient donc correspondre aux liaisons C(1)-H(α) entre le fucose et le GlcNAc (Higel *et al.*, 2016), la fucosylation étant présente dans les 3 groupes de glycoformes (fucosylées, galactosylées, et sialylées) mais pas dans le groupe "high mannose". Ces résultats permettent ainsi de conforter les modèles de micro-hétérogénéité développés à partir des spectres Raman acquis en ligne au cours des cultures cellulaires.

- Prédiction de l'évolution des concentrations des glycoformes de l'anticorps au cours de cultures en mode discontinu et recharge-récolte

Comme pour la macro-hétérogénéité, les modèles développés pour la micro-hétérogénéité de l'anticorps ont ensuite été appliqués sur de nouveaux spectres Raman enregistrés toutes les 20 minutes au cours de deux cultures conduites en mode discontinu et recharge-récolte. La Figure 64 présente les résultats des prédictions en temps réel des diverses concentrations des IgG totales, fucosylées, galactosylées, sialylées et fortement mannosylées, et les compare aux analyses réalisées hors ligne sur des prélèvements.

Sur la base des résultats des analyses hors ligne par HPLC-MS, nous pouvons constater la forte fucosylation des IgG, puisque la concentration des formes fucosylées est très proche de celle des IgG totales hormis en fin de culture. Par ailleurs, le niveau de galactosylation des IgG est plus faible que celui de fucosylation. Comme pour les formes fucosylées, la concentration des formes galactosylées augmente parallèlement à celle des IgG totales, avant d'atteindre un plateau en fin de culture. Cette diminution du pourcentage des formes fucosylées et galactosylées en fin de culture est concomitante à l'apparition des IgG non-glycosylées à partir d'environ 100 h dans la culture discontinue et 350 h dans la culture recharge-récolte (Chapitre III.3.2, Figure 4). Dans la dernière phase de culture, une accumulation de la glycoforme immature "high mannose" est aussi observée. Enfin, le niveau des IgG sialylées reste relativement bas tout au longtemps des cultures.

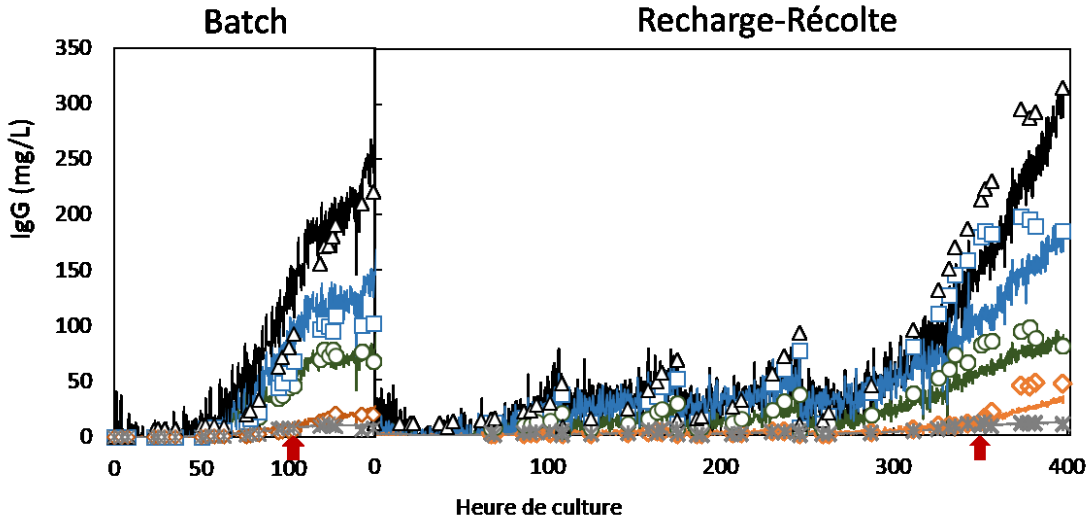


Figure 64 : Prédiction en temps réel de la concentration des IgG totales (noir), fucosylées (bleu), galactosylées (vert), sialylées (gris) et fortement mannosylées (orange), produites au cours de cultures en mode discontinu et recharge-récolte. Lignes continues : prédictions en ligne par Raman ; symboles : analyses hors ligne. Flèches rouges : début de l'accummulation des IgG non-glycosylées.

En ce qui concerne les prédictions en temps réel utilisant les modèles en ligne développés pour la micro-hétérogénéité de glycosylation, la tendance d'évolution des 4 groupes de glycoformes suit assez correctement les valeurs de référence obtenues hors ligne pour les deux cultures. Néanmoins,

après 300 h de culture, les prédictions pour les glycoformes fucosylées, galactosylées et fortement mannosylées apparaissent sous-estimées. Une explication possible serait que les modèles de micro-hétérogénéité sont développés à partir d'un jeu de données de calibration dont la variabilité de micro-hétérogénéité est faible (autour de 10%), ce qui diminue la capacité de prédiction du modèle.

- Prédiction de l'évolution des pourcentages des glycoformes d'IgG au cours d'une culture en mode discontinu

Cependant, suivre uniquement la concentration des différents groupes de glycoformes ne permet pas de connaître l'évolution de la distribution des glycoformes des IgG en fonction de leur fucosylation, galactosylation, sialylation et "high mannose" au cours des procédés de culture de cellules animales. En effet, l'évolution de la concentration des glycoformes est directement liée à l'augmentation de la concentration des IgG totale. Or, c'est plutôt la variation de la distribution des différentes glycoformes des IgG produites qui module les critères de qualité des IgG. Il apparaît donc également intéressant de pouvoir la suivre en temps réel.

Pour ce faire, le pourcentage de chaque groupe de glycoformes par rapport aux glycoformes totales, a été calculé en divisant la concentration de ce groupe par celle de toutes les IgG glycosylées. Comme toutes ces concentrations sont prédites en temps réel par spectroscopie Raman, grâce aux modèles conjoints de macro- et micro-hétérogénéité, les pourcentages de chaque groupe de glycoformes peuvent être aussi obtenus en temps réel. La Figure 65 présente les évolutions prédites en ligne et analysées hors ligne des pourcentages des différents groupes de glycoformes au cours de la culture discontinue. Il faut préciser que, dans cette figure le niveau de glycosylation des IgG a été calculé différemment en divisant la concentration de toutes les IgG glycosylées par la concentration des IgG totales. La concentration d'IgG étant très faible (< 20 mg/L) du début de la culture jusqu'à 70 h, les glycoformes sont quasi indétectables par la méthode d'analyse de glycosylation hors ligne. C'est pourquoi, le niveau de bruit sur les calculs étant alors très important, les résultats ne sont présentés qu'à partir de 70 h. La tendance d'évolution des pourcentages des différents groupes de glycoformes est bien prédite en général par comparaison aux analyses hors ligne. En tenant compte de la précision de la méthode hors ligne (15%), les valeurs de prédiction sont largement acceptables. De plus, cette figure permet de comprendre la moindre précision des modèles de micro-hétérogénéité par rapport au modèle de macro-hétérogénéité. En effet, l'évolution de la micro-hétérogénéité de glycosylation apparaît assez peu significative en cours de culture. Par conséquent, les concentrations des différents groupes de glycoformes sont proportionnelles à celle des IgG totales. La variation des glycoformes sont ainsi inter-corrélées avec celle des IgG totales, ce qui complique le développement des modèles distincts des glycoformes.

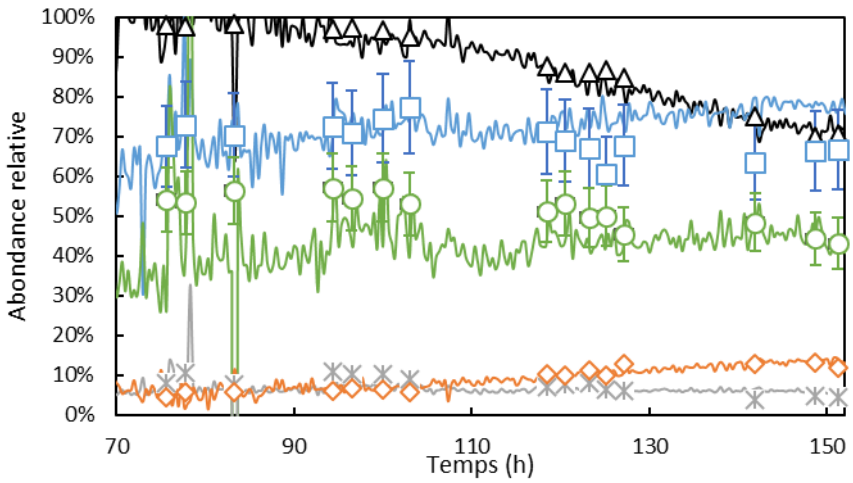


Figure 65 : Evolution des pourcentages des différents groupes de glycoformes (noir : glycosylation totale ; bleu : fucosylation ; vert : galactosylation ; orange : high mannose ; gris : sialylation) des IgG *produites à partir de 70 h d'une culture en mode discontinu*

Afin d'augmenter la performance de ces modèles de prédiction de la micro-hétérogénéité de la glycosylation des IgG, il serait intéressant de provoquer volontairement une augmentation de la variabilité de cette micro-hétérogénéité permettant d'améliorer les étapes de calibration. Ceci pourrait être réalisé par l'ajout, soit, de molécules supplémentaires dans le milieu de culture (galactose, manganèse par exemple), soit, d'hydrolases en cours de culture (fucosidase, galactosidase, sialydase, etc.). Parallèlement, d'autre méthodes chimiométriques pourraient être testées, telles que la technique de sélection de variables (iPLS, algorithme génétique, etc) ou les méthodes de régression non linéaire (SVR, ANN, etc). Dans notre étude, la zone spectrale choisie est assez générale pour pouvoir prédire toutes les molécules en éliminant juste les parties liées aux bruits intrinsèques à l'appareil et à la molécule d'eau. Grâce aux méthodes de sélection de variables, les informations spectrales pourraient être plus spécifiques à la micro-hétérogénéité de la glycosylation de l'IgG. Avec les méthodes de régression non linéaire, les relations non linéaires existantes entre les signaux Raman et la concentration des différentes glycoformes pourraient être mieux captées dans les modèles. De plus, les techniques de lissage en temps réel développées dans le chapitre I pourraient être utilisées pour améliorer la stabilité des signaux. Toutes ces améliorations permettraient éventuellement d'intégrer les prédictions de la micro-hétérogénéité de la glycosylation dans des stratégies de rétrocontrôle du procédé.

III.5. Suivi en ligne de la glycosylation des IgG par spectroscopie NIR

Bien que la spectroscopie NIR se soit révélée un peu moins performante que la spectroscopie Raman pour la prédiction de la concentration des IgG totales produites (Chapitre II), son potentiel pour le suivi en ligne de la macro- et micro-hétérogénéité de glycosylation des IgG a été également évalué. Comme présenté dans le chapitre précédent, la spectroscopie NIR a été utilisée en même temps que la spectroscopie Raman au cours de six cultures différentes de cellules CHO en bioréacteur. Afin de pouvoir comparer les performances des modèles développés avec les deux spectroscopies, les mêmes jeux de calibration et de prédiction ont été utilisés.

III.5.1. Modèles NIR de prédiction de la macro-hétérogénéité de glycosylation

Dans un premier temps, des modèles ont été développés à partir des spectres NIR pour prédire l'évolution de la macro-hétérogénéité de glycosylation *via* la concentration des IgG totales (IgG-T) et des IgG non-glycosylées (IgG-NG). Les valeurs prédites en ligne par les modèles ont été comparées avec les valeurs obtenues hors ligne par la méthode de référence (Figure 66). Visuellement, nous constatons que les points sont plus dispersés que ceux correspondant aux modèles développés à partir des spectres Raman (&Chapitre III.3.2), en particulier pour le modèle des IgG-NG. Par exemple, la valeur d'IgG-NG mesurée hors ligne à 12 $mg.L^{-1}$ est surestimée de plus de 400% par le modèle, tandis que celle mesurée à 93 $mg.L^{-1}$ est sous-estimée par le modèle avec une valeur prédite de 40 $mg.L^{-1}$.

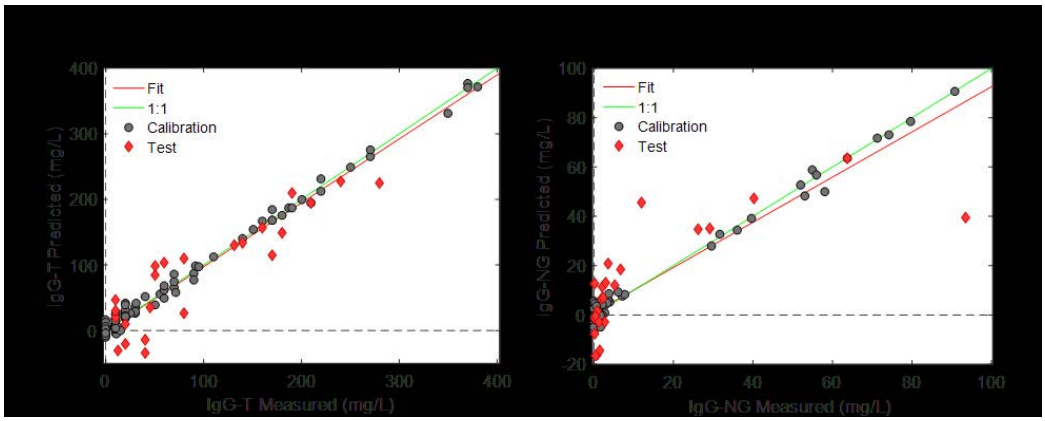


Figure 66 : Modèles de prédiction de la macro-hétérogénéité de glycosylation des IgG *par spectroscopie NIR en ligne : IgG-T (a), IgG-NG (b)*

De manière plus quantitative, l'analyse statistique présentée dans le Tableau 15 confirme le fait que la capacité de prédiction des modèles développés à partir des spectres NIR est moins bonne que celle des modèles Raman.

Tableau 15 : Analyse statistique des modèles de prédiction pour la macro-hétérogénéité de la glycosylation des IgG *par spectroscopie NIR*

	Range	LV	R^2_C	RMSEC		R^2_{CV}	RMSECV		R^2_P	RMSEP	
	($mg.L^{-1}$)			($mg.L^{-1}$)	%		($mg.L^{-1}$)	%		($mg.L^{-1}$)	%
IgG-T	0–380	8	0.99	8.3	2	0.72	50.6	13	0.82	36.8	10
IgG-NG	0–97.6	8	0.99	2.6	3	0.58	13.8	14	0.55	16.1	17

En effet, les erreurs relatives, en particulier, les erreurs de validation croisée et de prédiction, sont environ 2 fois plus importantes que celles des modèles Raman. L'analyse des FOM permet de confirmer que la performance de la spectroscopie NIR dans la prédiction de la glycosylation des IgG est bien inférieure à celle de la spectroscopie Raman (Annexe 1). Une hypothèse pour expliquer cette différence de performance pourrait être liée au fait que la spectroscopie NIR est probablement moins sensible pour détecter les liaisons peptidiques des IgG. De ce fait, comme le milieu de culture utilisé (surtout le milieu PF-CHO) contient de l'hydrolysat de protéines de soja, riche en peptides et acides aminés, les liaisons peptidiques pourraient alors se chevaucher et masquer les signaux des IgG. De plus, la sensibilité importante de la spectroscopie NIR vis-à-vis de la molécule d'eau pourrait également expliquer cette capacité de prédiction moindre des modèles NIR.

Par ailleurs, les profils des scores VIP des deux modèles NIR correspondant aux IgG totales et non-glycosylées ne montrent pas de différences significatives (Figure 67). Le pic observé autour de 7000 cm^{-1} correspond à la première harmonique de la molécule d'eau (Clavaud *et al.*, 2013). Il permet de confirmer la grande influence de l'absorption de la molécule d'eau sur les deux modèles car les signaux des IgG et des différentes glycoformes sont masqués.

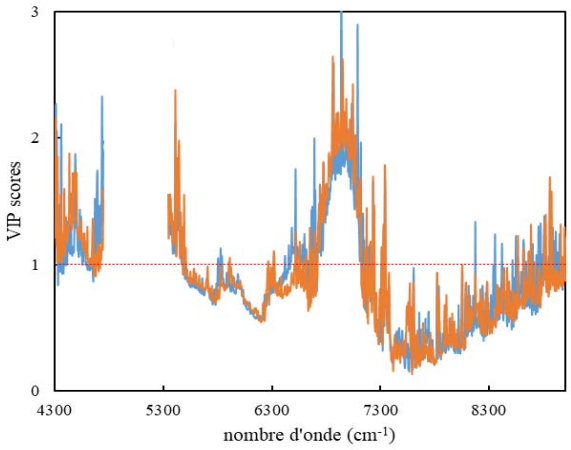


Figure 67 : Scores VIP des modèles de prédiction de la macro-hétérogénéité de la glycosylation des IgG *par spectroscopie NIR en ligne : IgG-T (traits oranges), IgG-NG (traits bleus)*

5.2. Modèles NIR de prédiction de la micro-hétérogénéité de glycosylation

Pour la prédiction en ligne de la micro-hétérogénéité de glycosylation des IgG, la même procédure que celle utilisée pour la spectroscopie Raman a été appliquée. Les modèles ont été développés à partir des spectres NIR pour prédire la concentration des IgG galactosylées, fucosylées, sialylés et fortement mannosylées. Les valeurs prédites en ligne par les modèles ont ensuite été comparées à celles mesurées hors ligne par la méthode de référence (Figure 68). Pour les 4 modèles, une bonne linéarité est observée entre les valeurs de prédiction et les valeurs mesurées, hormis 1 point (dans le jeu de prédiction) probablement aberrant.

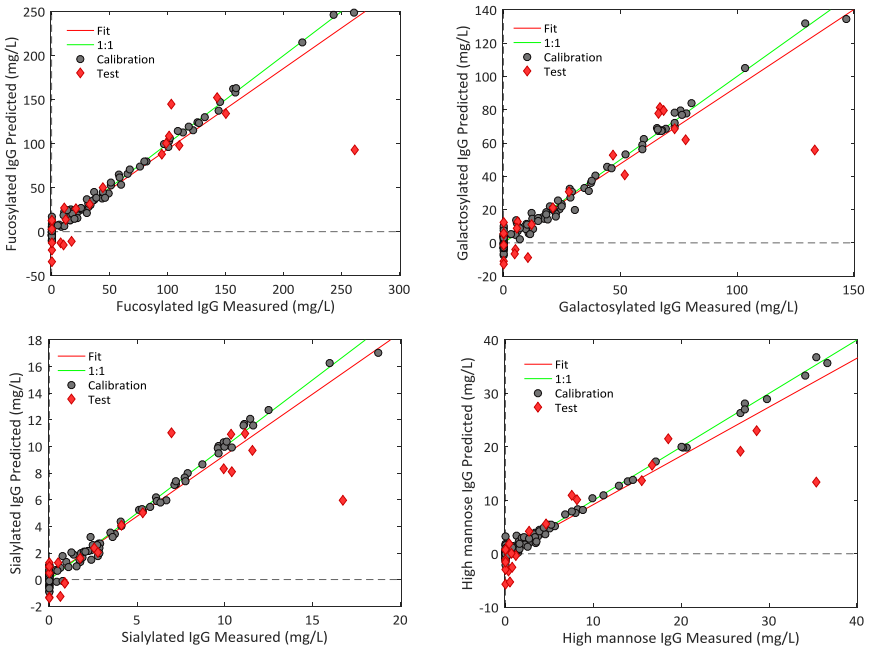


Figure 68 : Modèles de prédiction de la micro-hétérogénéité de la glycosylation des IgG *par spectroscopie NIR*

Ces bons résultats sont à mettre en regard avec un nombre élevé de LV (8 à 9) utilisés pour tous les modèles (Tableau 16). Dans ce cas, les modèles développés sont alors bien spécifiques des variables à prédire, mais ils présentent plus de risques de susciter un phénomène d'« overfitting ». C'est la raison pour laquelle leurs erreurs de validation croisée et de prédiction sont largement supérieures aux erreurs de calibration. Compte-tenu de ces résultats, les performances de la spectroscopie NIR pour prédire en temps réel la micro-hétérogénéité des IgG produits n'apparaissent pas suffisantes pour envisager un suivi en temps réel de cultures. L'analyse des FOM permet de confirmer cette conclusion (Annexe 1).

Tableau 16 : Analyse statistique des modèles de prédiction de la micro-hétérogénéité de la glycosylation des IgG *par spectroscopie NIR*

mg.L^{-1}	**LV**	**RMSEC**	**RMSECV**	**RMSEP**
IgG-fuc	9	5,4 (2%)	32,6 (12%)	39,9 (14%)
IgG-gal	8	3,9 (3%)	16,7 (11%)	19,1 (13%)
IgG-sia	8	0,5 (3%)	2,1 (11%)	2,7 (14%)
IgG-man	9	0,8 (2%)	5 (14%)	5,7 (16%)

III.6. Conclusion

Les travaux de ce chapitre ont étudié la faisabilité du suivi en ligne de l'un des attributs critiques de la qualité des IgG produites par des cellules CHO en bioréacteur. Plus spécifiquement, ils se sont attachés à l'analyse *in situ* de la glycosylation des IgG en utilisant les spectroscopies Raman et NIR couplées à des sondes stérilisables. L'approche mise en œuvre et les principaux résultats sont présentés Figure 69.

Tous d'abord, des méthodes de référence pour l'analyse hors ligne de la concentration des IgG produites ainsi que de leur glycosylation ont été développées par HPLC-MS. Concernant la macro-hétérogénéité de glycosylation, liée à l'occupation ou non des sites de glycosylation, nous avons constaté que la quantité relative des IgG non glycosylées n'était pas constante au cours de la culture. En effet, le pourcentage des IgG-NG augmente de 2 à 32% au cours d'une culture discontinue, ce qui induit une diminution significative de la qualité des IgG produites. De plus, pour la micro-hétérogénéité de glycosylation, nous avons montré que, parmi les 26 glycoformes analysées, 8 sont majoritaires. Parmi ces glycoformes, les formes fucosylées et galactosylées sont prépondérantes.

Dans un second temps, divers modèles ont été développés à partir des spectres Raman et NIR acquis *in situ* dans les réacteurs, pour prédire la concentration des IgG totales et non-glycosylées pour la macro-hétérogénéité de glycosylation, et la concentration des IgG fucosylées, galactosylées, sialylées et hautement mannosylées pour la micro-hétérogénéité. Les erreurs relatives (RMSEC, RMSECV, RMSEP) ont été calculées pour chaque modèle. Les modèles développés à partir des spectres Raman ont montré des erreurs inférieures à 10% pour la macro-hétérogénéité et légèrement plus élevées (voisines de 10%) pour la micro-hétérogénéité. En revanche, les modèles développés à partir des spectres NIR, même s'ils ont permis des erreurs de calibration faibles (RMSEC), ont montré des erreurs de validation croisée (RMSECV) et de prédiction (RMSEP) beaucoup plus élevées que celles des modèles développés avec les spectres Raman. De plus, les scores VIP calculés pour chaque modèle ont permis de confirmer que, au contraire de la spectroscopie NIR, la spectroscopie Raman était capable de distinguer entre les IgG-NG et les IgG-T. Ceci confirme que la spectroscopie Raman possède une meilleure capacité pour quantifier des molécules présentes à de faibles concentrations dans un milieu complexe.

Enfin, les modèles développés à partir des spectres Raman ont été appliqués sur de nouveaux spectres Raman enregistrés en temps réel (toutes les 20 min) au cours de cultures en mode discontinu et recharge-récolte. Pour la macro-hétérogénéité, l'évolution de la concentration des IgG-T et des IgG-NG a été bien prédite dans les 2 modes de cultures, ce qui démontre la robustesse et la

stabilité des modèles. De plus, par cette méthode, il est possible de détecter rapidement les moments critiques où les IgG-NG commencent à s'accumuler. Ceci permet d'accéder en temps réel à une information importante afin de pouvoir réagir rapidement pour éviter l'altération de la qualité du produit. Concernant la micro-hétérogénéité, les grandes tendances des différentes glycoformes ont été correctement suivies. En mode recharge-récolte, du fait de la faible variation de la distribution des glycoformes nécessaires pour bien calibrer les modèles, les prédictions étaient moins bonnes qu'en mode discontinu.

Ainsi, nos résultats montrent clairement le potentiel des méthodes spectroscopiques, telles que la spectroscopie Raman, non seulement, pour le suivi en ligne de la quantité des IgG produites, mais aussi pour la prédiction en temps réel de la qualité des IgG, notamment la macro- et micro-hétérogénéité de glycosylation. La performance des modèles développés pourrait être améliorée dans des perspectives futures, d'une part, en accumulant plus de données expérimentales, et, d'autre part, en élargissant la variabilité des données. D'autres méthodes chimiométriques telles que les techniques de sélection de variables (interval PLS, algorithme générique, etc.), et les méthodes de régression non-linéaires (SVR, ANN, etc.) pourraient également être testées. Ces résultats constituent une première avancée vers le contrôle et le rétrocontrôle en temps réel de la qualité des anticorps produits au cours de procédés de culture de cellules animales.

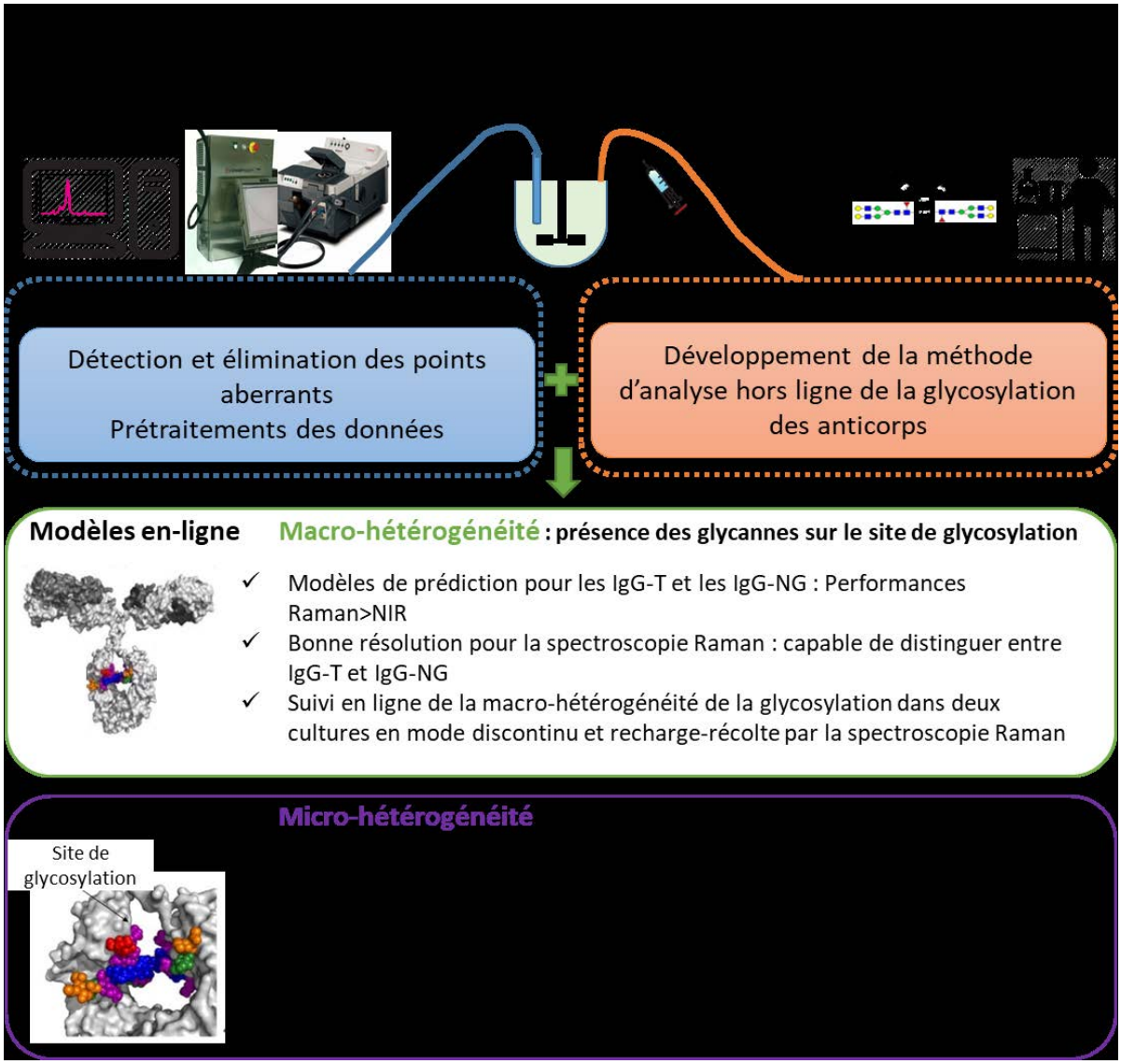


Figure 69 : Synthèse schématique des résultats du chapitre III

Chapitre IV

Travaux préliminaires pour le pilotage automatisé d'une culture semi-continue par l'utilisation conjointe de différentes techniques spectroscopiques

IV.1. Introduction 195

IV.2. Modèles de suivi en ligne des paramètres cinétiques par utilisation conjointe de différentes spectroscopies au cours d'une culture discontinue 196

IV.2.1. Evaluation en ligne des vitesses spécifiques de consommation et de production 197

IV.2.2. Evaluation en ligne des rendements métaboliques 200

IV.3. Mise au point d'une culture semi-continue avec pilotage automatisé de l'alimentation 202

IV.3.1. Mise au point expérimentale 204

IV.3.2. Développement du capteur logiciel 209

IV.3.3. Interface du programme de pilotage 212

IV.4. Résultats d'une culture semi-continue avec pilotage automatisé de l'alimentation 214

IV.4.1. Prédictions des paramètres cinétiques et calcul en temps réel du débit d'alimentation 215

IV.4.2. Comparaison de différentes stratégies d'alimentation d'une culture semi-continue 217

IV.5. Conclusion 222

IV.1. Introduction

Le procédé semi-continu (fed-batch) est, actuellement, le mode de culture le plus utilisé à l'échelle de la production industrielle pour répondre à la forte demande en protéines recombinantes thérapeutiques (Lu *et al.*, 2013). Cependant, le principal défi de ce mode de mise en œuvre réside dans la stratégie d'alimentation du bioréacteur en milieu de culture, celle-ci pouvant grandement affecter le niveau de production et la qualité des protéines recombinantes. Les pratiques actuelles des industries pharmaceutiques, dans le cas des procédés de culture de cellules animales, consistent à utiliser une alimentation ponctuelle (bolus feeding), unique ou répétée, basée sur des connaissances plus ou moins empiriques du procédé. Bien que largement utilisée, cette stratégie peut s'avérer limitée pour répondre aux demandes nutritionnelles réelles des cellules, et entraîner l'accumulation de sous-produits du métabolisme et/ou l'épuisement de certains nutriments, avec le risque d'affecter ainsi la productivité et la qualité des protéines recombinantes (Lu *et al.*, 2013). C'est pourquoi, plus récemment, des stratégies adaptatives utilisant divers algorithmes commencent à être employées pour mieux adapter l'apport en milieu de culture aux besoins des cellules. L'un des algorithmes les plus utilisés consiste à développer des modèles métaboliques, intégrant la croissance cellulaire et la consommation des éléments nutritifs, à partir de mesures expérimentales en temps réel, pour prédire la demande en nutriments au cours de la culture et ainsi adapter l'alimentation du milieu (Huang *et al.*, 2010; Aehle *et al.*, 2011). Cette approche, qui permet une meilleure connaissance dynamique du procédé, peut être considérée comme un capteur logiciel de type boîte blanche (Etude bibliographique, 2.6). Cependant, outre qu'elle demande une bonne connaissance métabolique antérieure des cellules, elle nécessite un volume de travail important de la part de l'opérateur car elle requiert, non seulement, des prélèvements stériles très fréquents pour assurer un suivi régulier, mais également, des calculs et des ajustements manuels du capteur logiciel après chaque prélèvement.

C'est pourquoi, ce chapitre propose la mise en place d'une nouvelle stratégie d'alimentation adaptative d'une culture semi-continue de cellules CHO productrices d'anticorps de type IgG. Cette stratégie sera basée sur l'utilisation d'un capteur logiciel de type boîte grise, intégrant des données mesurées en ligne par des techniques spectroscopiques aux modèles décrivant le métabolisme des cellules (*e.g.* bilans de matières). Pour ce faire, dans un premier temps, la combinaison de différentes spectroscopies sera proposée, dans le cas d'une culture discontinue, pour le suivi en temps réel de paramètres cinétiques clés du procédé, tels que les vitesses spécifiques et les rendements métaboliques. En effet, les résultats des chapitres précédents ont démontré le potentiel spécifique à chaque spectroscopie, utilisée séparément, pour le suivi en ligne du bioprocédé, comme la vitesse

spécifique de croissance cellulaire ou les concentrations des composés du milieu de culture. Néanmoins, l'intérêt de l'utilisation conjointe de ces méthodes n'a pas encore été envisagé pour le suivi en ligne plus détaillé du métabolisme des cellules à travers les vitesses spécifiques et les rendements métaboliques. Dans un second temps, la mise au point du pilotage automatisé de l'alimentation d'une culture semi-continue sera décrite. Cette partie intègrera la conception de la partie expérimentale, le développement du capteur logiciel, et l'interfaçage du programme de pilotage. Enfin, seront présentés les résultats concernant le pilotage automatisé d'une culture semi-continue, basé sur le calcul en temps réel des différents paramètres cinétiques, à partir des mesures *in situ* par spectroscopie diélectrique et NIR. Les performances de la culture semi-continue automatisée seront comparées à celles d'une culture discontinue ainsi qu'à celles de cultures semi-continues utilisant une stratégie d'alimentation manuelle, soit ponctuelle, soit exponentielle.

IV.2. Modèles de suivi en ligne des paramètres cinétiques par utilisation conjointe de différentes spectroscopies au cours d'une culture discontinue

En plus de la vitesse spécifique de croissance cellulaire (μ), les vitesses spécifiques de consommation des substrats (q_s) et de production des métabolites (q_p), de même que les rendements métaboliques (Y), sont des paramètres cinétiques essentiels pour comprendre et contrôler les procédés de culture de cellules animales, puisqu'ils renseignent directement sur l'état physiologique des cellules (Pörtner *et al.*, 1993). Dans le cas du pilotage de cultures semi-continues utilisant des modèles métaboliques de type boîte blanche, ces paramètres cinétiques sont souvent considérés comme constants, car ils sont difficilement accessibles en temps réel et ne peuvent donc être calculés qu'à partir de mesures hors ligne différées. Or, ces paramètres peuvent évoluer significativement en cours de culture, par exemple quand certains substrats deviennent limitants ou lors de "shifts" métaboliques. Utiliser des valeurs constantes dans les modèles métaboliques peut alors entraîner une sous- ou sur-alimentation en milieu nutritif lors de cultures semi-continues, et ainsi affecter les cellules et la production des protéines recombinantes (Aehle *et al.*, 2012). Le suivi en temps réel de ces paramètres cinétiques dans le cas des procédés de culture de cellules animales représente donc un réel défi. Comme montré dans les chapitres I et II, les concentrations des cellules viables et des différents composés du milieu de culture (glucose, lactate, IgG, ...), ainsi que la croissance cellulaire, peuvent être prédites en temps réel avec une assez bonne précision par spectroscopies *in situ*. En

combinant ces prédictions entre elles, il est alors envisageable de prédire en temps réel les paramètres cinétiques suivants :

- vitesse spécifique de consommation du glucose (q_{Glc}, mmole.h^{-1}.10^{8}cellules^{-1}),

- vitesse spécifique de production du lactate (q_{Lac}, mmole.h^{-1}.10^{8}cellules^{-1}),

- vitesse spécifique de production des anticorps IgG (q_{IgG}, mg.h^{-1}.10^{8}cellules^{-1}),

- rendement de production de cellules par glucose consommé ($Y_{X/Glc}$, 10^{8}cellules.mmole^{-1}),

- rendement de production de lactate par glucose consommé ($Y_{Lac/Glc}$, mmole.mmole^{-1}),

- rendement de production d'anticorps IgG par cellules viables produites ($Y_{IgG/X}$, mg.10^{8}cellules $^{-1}$).

Les vitesses spécifiques de consommation de glutamine et de production d'ions ammonium ne seront pas prises en compte dans cette partie, du fait des moindres précisions des modèles de prédiction respectifs de leur concentration.

IV.2.1. Evaluation en ligne des vitesses spécifiques de consommation et de production

Les modèles développés à partir des spectres Raman et NIR dans le chapitre II ont été utilisés pour la prédiction de la concentration en glucose, lactate et IgG. La spectroscopie diélectrique a été choisie pour la prédiction de la VCD, cette technologie étant plus directe et plus spécifique pour l'analyse des cellules viables que les spectroscopies NIR ou Raman. Une simple corrélation linéaire, présentée dans le Chapitre I, suffira pour prédire la VCD à partir des mesures de permittivité en ligne. En combinant ces différentes prédictions, les vitesses spécifiques de consommation (q_S) et de production (q_P) peuvent alors être calculées à partir des équations suivantes, dans le cas d'une culture discontinue :

$$q_{s} = -\frac{1}{X_v} \times \frac{dS}{dt} \qquad \textbf{Eq 41}$$

$$q_{p} = \frac{1}{X_v} \times \frac{dP}{dt} \qquad \textbf{Eq 42}$$

En pratique, q_s et q_p sont calculées en divisant la pente d'une régression linéaire locale (réalisée à l'aide d'une fenêtre glissante sur la courbe d'évolution de la concentration de substrat (*S*) ou de produit (*P*) prédite en fonction du temps par la valeur moyenne des VCD (X_v) prédites sur l'intervalle

de temps considéré. Des filtres sont appliqués à différents niveaux pour augmenter la stabilité des modèles. Tout d'abord, des filtres de type moyenne mobile (MA) sont appliqués en temps réel sur les courbes de concentrations de *S* et de VCD prédites en fonction du temps ; un filtre Savitzky-Golay (SG) est ensuite appliqué sur les valeurs calculées de q_s ou q_p (entre t_0 et t). Comme présenté dans le chapitre I.3.1, les largeurs de la fenêtre glissante (ω = 10 h) et des filtres (ωF = 2 h) ont été optimisés afin de trouver un bon compromis entre la stabilité et la sensibilité des modèles. La Figure 70 présente les résultats des calculs des vitesses spécifiques, obtenus à partir des mesures spectroscopiques en ligne, au cours d'une culture discontinue. Ils sont comparés à ceux obtenus à partir de mesures hors ligne. Pour faciliter la compréhension de l'évolution de ces vitesses spécifiques, les évolutions des diverses concentrations mesurées hors ligne sont rappelées dans la Figure 71.

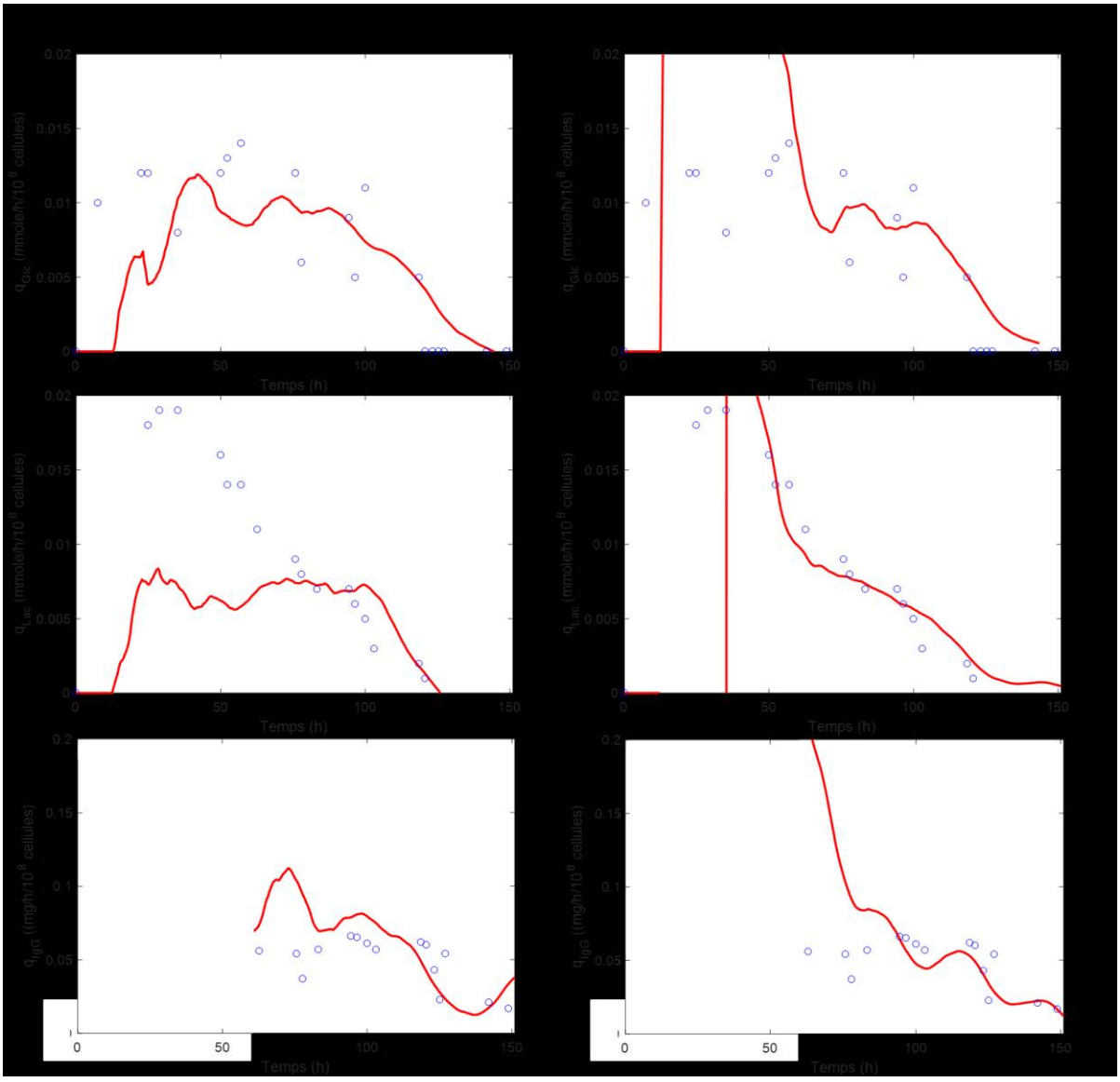


Figure 70. Prédictions en temps réel de vitesses spécifiques (courbes rouges) à partir des mesures de spectroscopie diélectrique couplées à la spectroscopie Raman (A, C, E) ou NIR (B, D, F) : q_{Glc} (A, B), q_{Lac} (C, D), q_{IgG} (E, F). Valeurs calculées à partir des mesures hors ligne (cercles bleus).

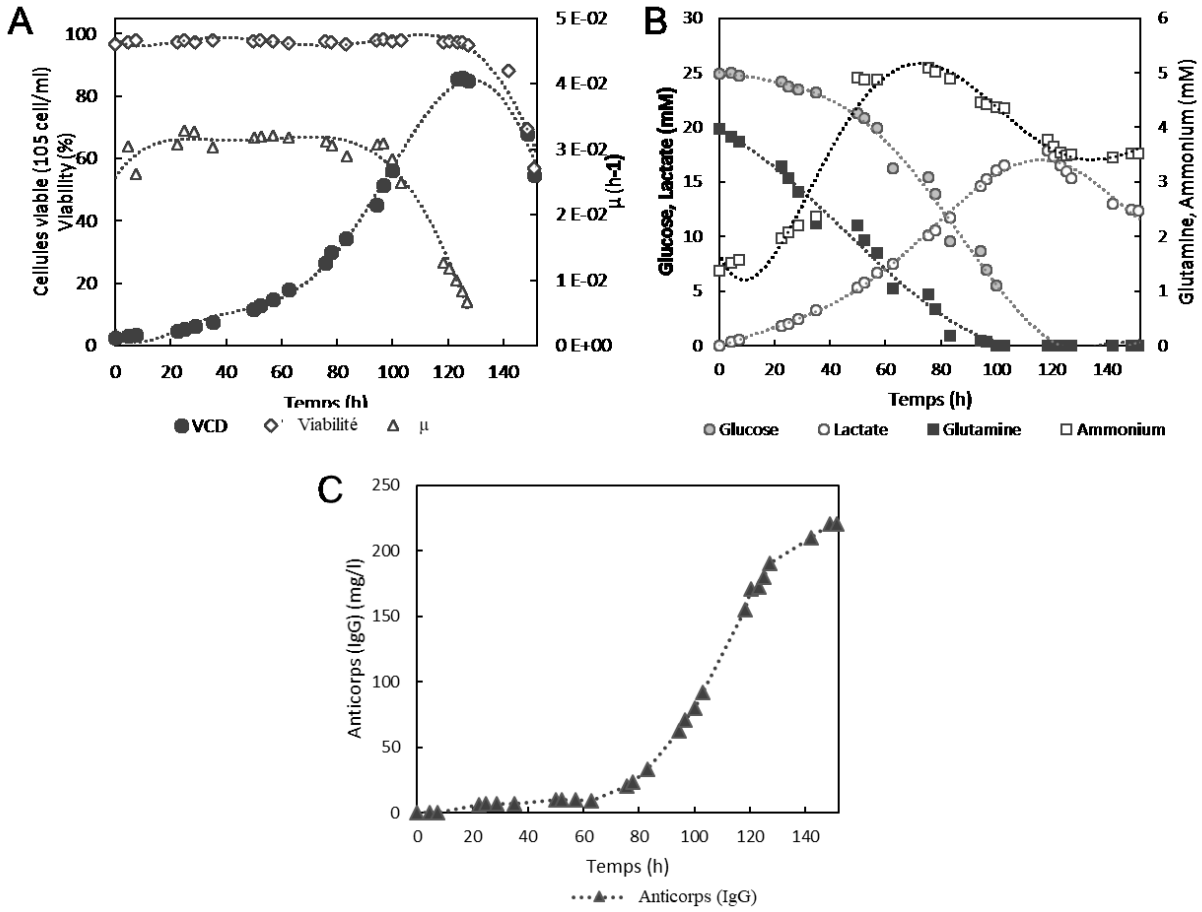


Figure 71 : Cinétique hors-ligne de la culture discontinue : croissance cellulaire (A), production/consommation des métabolites (B), production d'anticorps IgG (C).

Les valeurs moyennes des vitesses spécifiques q_{Glc} et q_{Lac} dans ces conditions de culture sont de l'ordre de 0,01 mmole.h^{-1}.10^8cellules^{-1} (Figure 70.A, B, C, D). Bien que ces valeurs dépendent de la lignée cellulaire et des conditions de culture, elles sont du même ordre de grandeur que celles trouvées dans la littérature (Le *et al.*, 2012). Dans cette culture, q_{Glc} est relativement stable en début de culture avant de diminuer aux alentours de 100 h. Cette diminution coïncide avec la diminution de μ (Figure 71.A). Ce couplage entre μ et q_{Glc} a été observé à de nombreuses reprises (López-Meza *et al.*, 2016). Pour q_{Lac}, sa valeur est importante en début de culture puis elle diminue au cours du temps. Après l'atteinte de la concentration cellulaire maximale (approximativement au moment de la déplétion en glucose vers 120 h), le lactate accumulé est en partie consommé par les cellules (Figure

70.C, Figure 71.B). Concernant q_{IgG}, son estimation n'est présentée qu'à partir de 60 h, quand la quantité d'IgG produites devient significative (Figure 71.C). Sa valeur est relativement stable jusqu'à 120 h, autour de 0,05 $mg.h^{-1}.10^{8}$ $cellules^{-1}$ avant de diminuer (Figure 70.E, F). Il est à noter qu'à ce moment-là, la valeur de μ devient quasiment nulle, ce qui correspond à la consommation totale de la glutamine et du glucose (Figure 71.A, B).

Ainsi, les vitesses spécifiques calculées en temps réel suivent globalement la tendance des valeurs calculées hors ligne. Néanmoins, les prédictions en temps réel par spectroscopie NIR de la concentration des divers composés présentent une forte déviation en début de culture (Chapitre II.2.2). Par conséquent, le suivi en ligne des vitesses spécifiques utilisant les données combinées des spectroscopies diélectrique et NIR présente également une forte instabilité sur cette même période initiale (Figure 70. B,D,F). Par ailleurs, contrairement aux vitesses spécifiques q_{Glc} et q_{IgG}, la vitesse q_{Lac} est mieux prédite en utilisant la spectroscopie NIR que la spectroscopie Raman.

IV.2.2. Evaluation en ligne des rendements métaboliques

Les prédictions des rendements métaboliques sont calculées à partir des modèles de prédiction de la VCD et de la concentration des composants du milieu de culture prédite (développés précédemment dans les chapitres I et II) à l'aide des équations suivante :

$$Y_{X/S} = \frac{(Xv_i - Xv_{i-\omega})}{(S_{i-\omega} - S_i)} \quad \textbf{Eq 43}$$

$$Y_{P/S} = \frac{(P_i - P_{i-\omega})}{(S_{i-\omega} - S_i)} \quad \textbf{Eq 44}$$

$$Y_{IgG/X} = \frac{(IgG_i - IgG_{i-\omega})}{(Xv_i - Xv_{i-\omega})} \quad \textbf{Eq 45}$$

Les modèles de calcul en ligne sont développés en traçant les évolutions des concentrations d'un paramètre en fonction d'un autre (les cellules en fonction du substrat par exemple). Les valeurs de $Y_{X/S}$, $Y_{P/S}$ et $Y_{IgG/X}$ peuvent être ainsi calculées à partir de la pente de la régression linéaire locale, obtenue à l'aide d'une fenêtre glissante (ω : largeur de fenêtre) sur la courbe tracée. De manière analogue aux calculs des vitesses spécifiques, des filtres MA sont utilisés pour lisser les valeurs

prédites de *X*, *S* et *P*, et le filtre SG est appliqué sur les valeurs de $Y_{X/S}$ et $Y_{P/S}$ calculées. Dans ce cas, la largeur de la fenêtre glissante (ω) est de 10 h, et la largeur des filtres (ωF) est de 2 h. Les résultats du suivi en ligne de $Y_{X/Glc}$, $Y_{Lac/Glc}$, et $Y_{IgG/X}$ sont présentés dans la Figure 72.

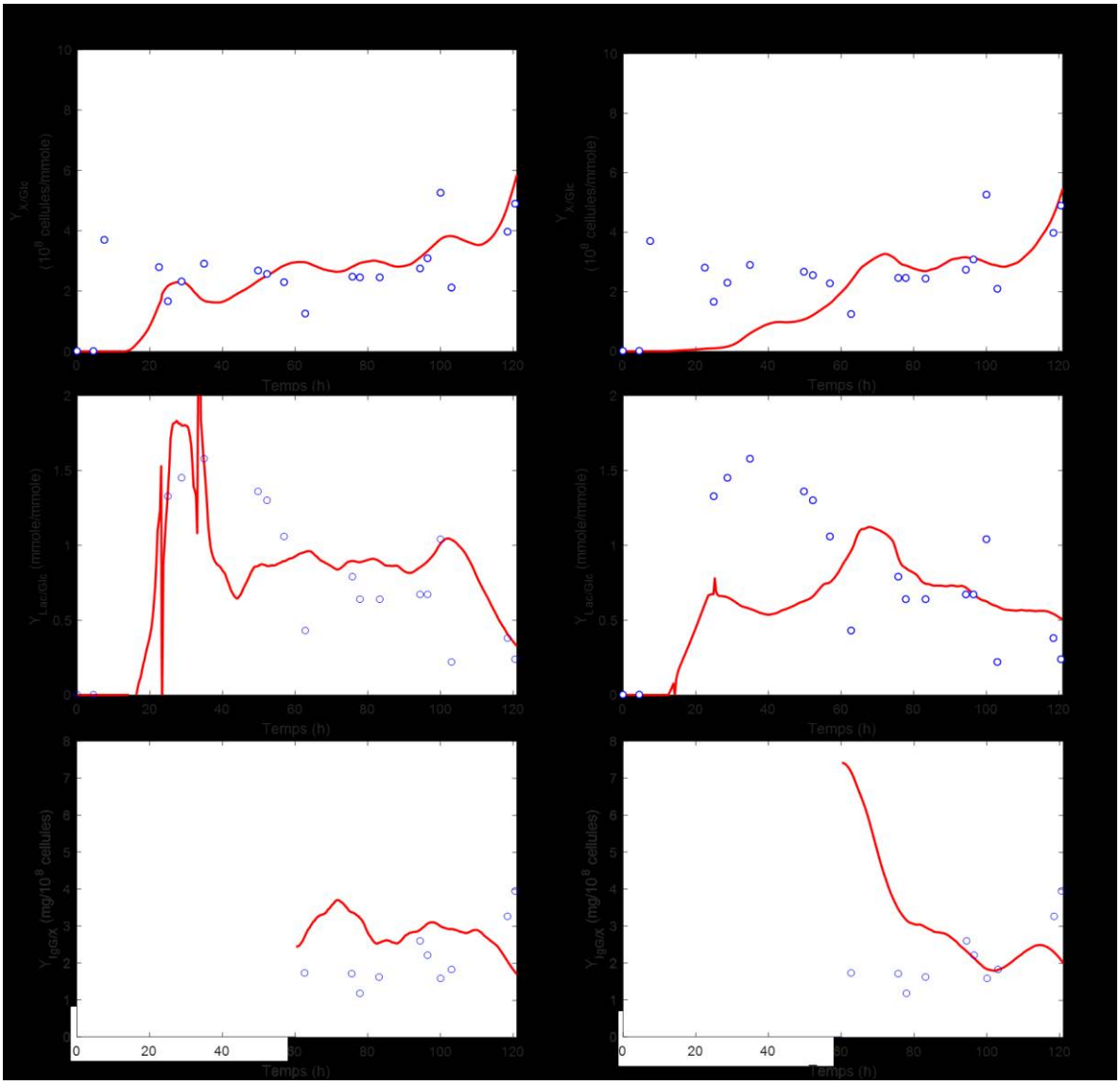


Figure 72 : Prédictions en temps réel des rendements métaboliques (courbes rouges) à partir des mesures de spectrocopie diélectrique couplée à la spectroscopie Raman (A, C, E) ou NIR (B, D, F) : $Y_{X/Glc}$ (A, B), $Y_{Lac/Glc}$ (C, D), $Y_{IgG/X}$ (E, F). Valeurs calculées à partir des mesures hors ligne (cercles bleus).

Le rendement $Y_{X/Glc}$ est un paramètre souvent utilisé lors de pilotages basés sur des modèles métaboliques pour déterminer le débit d'alimentation en substrat de cultures semi-continues (Konakovsky *et al.*, 2016). Nos résultats ayant été obtenus en mode discontinu, $Y_{X/Glc}$ reste quasiment constant (autour de 2.10^8 cellules.mmole^{-1}) avec une augmentation vers 120 h (autour de 5.10^8 cellules.mmole^{-1}) (Figure 72.A). Cependant, ce paramètre peut varier significativement dans le cas

d'une culture en mode semi-continu (Aehle *et al.*, 2012). Le rendement de conversion glycolytique, $Y_{Lac/Glc}$, représente un indicateur du métabolisme énergétique. Sa valeur est plus élevée en début de culture (de l'ordre de 1,5 mmole.mmole^{-1}), ce qui correspond à une faible efficacité d'utilisation de la source énergétique (ici le glucose). En effet, en présence d'un excès de glucose, ce qui est le cas en début de culture, le flux glycolytique est élevé et les cellules convertissent une grande partie du glucose consommé en lactate. Ce phénomène, correspondant à une glycolyse en condition aérobie, est une voie peu efficace pour la génération de molécules d'ATP (Vander Heiden *et al.*, 2009). La valeur de $Y_{Lac/Glc}$ diminue ensuite (autour de 0,7 mole.mole^{-1} entre 60 et 100 h), pour atteindre 0,3 mole.mole^{-1} après 100 h de culture. Cette valeur indique un changement métabolique avec une utilisation plus efficace du glucose. Concernant le rendement de production d'IgG $Y_{IgG/X}$, sa valeur est plutôt stable au cours de la culture (autour de 2 mg.10^{8} cellules^{-1}), avec une légère augmentation en fin de culture vers 120 h.

De manière globale, l'ordre de grandeur et les grandes tendances d'évolution des différents rendements métaboliques sont plutôt bien prédits en ligne par les méthodes spectroscopiques. Néanmoins, et de manière plus marquée que pour la prédiction des vitesses spécifiques, le suivi en ligne des rendements métaboliques est mieux prédit par la spectroscopie Raman en début de culture. De plus, la valeur de $Y_{Lac/Glc}$ est sous-estimée par la spectroscopie NIR pendant les premières 60 h de culture (Figure 72.B). Enfin, aucune des deux spectroscopies ne réussit à prédire la faible augmentation de $Y_{IgG/X}$ en fin de culture (Figure 72.E, F).

IV.3. Mise au point d'une culture semi-continue avec pilotage automatisé de l'alimentation

Actuellement, de nombreuses méthodes de pilotage de cultures semi-continues sont basées sur une connaissance préalable du procédé. L'alimentation, généralement réalisée avec un milieu concentré et avec des volumes fixes, peut avoir lieu, soit quotidiennement, soit de manière continue, soit encore à intervalles fixes. Néanmoins, il existe beaucoup de sources de variations au cours de ces procédés, augmentant considérablement la complexité du pilotage. Par conséquent, une stratégie d'alimentation adaptative, précise, et sûre, apparaît pertinente. Dans ce contexte, l'objectif de cette partie est de développer une nouvelle stratégie d'alimentation basée sur des analyses en temps réel, qui prendrait en compte à la fois les variations liées à la physiologie cellulaire (à travers μ) et les besoins en glucose (à travers $Y_{X/Glc}$), en utilisant un capteur logiciel de type boîte grise. Les mesures

spectroscopiques en ligne seront ainsi intégrées dans les équations de bilan de matière, afin de répondre plus précisément aux besoins cellulaires et de contrôler le procédé en temps réel. La transposition des modèles de prédiction des paramètres cinétiques, développés en mode discontinu, à des cultures semi-continues nécessitera de tenir compte, par l'écriture des bilans de matière, de la variation en temps réel du volume du milieu de culture dans le bioréacteur (*V*), ainsi que de la quantité de glucose introduite (Glc_F) par le milieu d'alimentation. Le débit d'alimentation (*F*) peut ainsi être calculé en fixant une consigne du glucose résiduel (Glc_{target}). *Xv* étant la concentration des cellules viables. Le bilan de matière du glucose peut être présenté par l'équation ce-dessous :

$$F \times Glc_F - r_{Glc} \times V = \frac{d(Glc_{target}V)}{dt} \quad \textbf{Eq 46}$$

En supposant que la croissance cellulaire est directement liée avec la consommation de glucose, la vitesse de consommation de glucose r_{Glc} peut être représentée par la vitesse de croissance r_X et le rendement $Y_{X/Glc}$:

$$r_{Glc} = \frac{r_X}{Y_{X/Glc}} = \frac{1}{Y_{X/Glc}} \times \mu \times X_v \quad \textbf{Eq 47}$$

Puisque la valeur de Glc_{target} est constante, mais la valeur de *V* est variable en fonction du temps, le terme dérivé $\frac{d(Glc_{target}V)}{dt}$ peut être représenté par $Glc_{target} \times \frac{dV}{dt}$. Sachant que $\frac{dV}{dt} = F$, le débit d'alimentation *F* peut être représenté par les équations suivantes :

$$F \times Glc_F - \frac{1}{Y_{X/Glc}} \times \mu \times X_v \times V = Glc_{target} \times F \quad \textbf{Eq 48}$$

$$F \cdot \left(Glc_F - Glc_{target}\right) = \frac{1}{Y_{X/Glc}} \cdot \mu \cdot X_v \cdot V \quad \textbf{Eq 49}$$

$$F = \frac{\mu . V . X_v}{Y_{X/Glc} . \left(Glc_F - Glc_{target}\right)} \quad \textbf{Eq 50}$$

Afin de calculer le débit d'alimentation pour un temps futur (F_t), nous devons connaître la concentration des cellules viables au temps 0 (X_{v0}) et au temps *t* (X_{vt}), ainsi que le volume du bioréacteur au temps 0 (V_0) et au temps t (V_t).

$$\mu = \frac{1}{X_v V} \times \frac{d(X_v V)}{dt} \quad \textbf{Eq 51}$$

$$X_{vt} \cdot V_t = X_{v0} \cdot V_0 \cdot e^{\mu.\Delta t} \quad \textbf{Eq 52}$$

$$F_t = \frac{\mu . V_t . X_{vt}}{Y_{X/Glc} . \left(Glc_F - Glc_{target}\right)} e^{\mu.\Delta t} \quad \textbf{Eq 53}$$

Ainsi, pour la suite des travaux, la prédiction en temps réel du rendement $Y_{X/Glc}$ développée précédemment sera utilisée dans le capteur logiciel pour le pilotage de la culture semi-continue à l'aide de la formule suivante, avec une largeur de fenêtre (ω) :

$$Y_{X/Glc} = \frac{quantité\ de\ cellules\ viables\ formées}{quantité\ de\ glucose\ consommé} = \frac{(Xv_i \times V_i - Xv_{i-\omega} \times V_{i-\omega})}{(Glc_{Conso\ i-\omega} \times V_{i-\omega} - Glc_{conso\ i} \times V_i)} \quad \textbf{Eq 54}$$

La quantité de glucose consommé ($Glc_{Conso} \times V$) peut être calculée à partir des équations suivantes :

$$Glc_{Conso} \times V = r_{Glc} \times V \times dt \quad \textbf{Eq 55}$$

$$r_{Glc} \cdot V = F \cdot Glc_F - \frac{d(Glc \cdot V)}{dt} \quad \textbf{Eq 56}$$

$$Glc_{Conso} \times V = F \cdot Glc_F \times dt - d(Glc \cdot V) \quad \textbf{Eq 57}$$

Ainsi, nous pouvons constater que dans les cultures en mode semi-continu, la quantité de glucose consommé comprend deux parties : la partie « visible » décrite par la variation de glucose résiduel mesurable dans le bioréacteur ($-d(Glc \cdot V)$), et la partie « invisible » fournie par le milieu d'alimentation ($F \cdot Glc_F \times dt$).

IV.3.1. Mise au point expérimentale

Pour cette étude, le glucose a été choisi comme l'élément nutritif de référence pour le contrôle de la culture. En effet, il constitue l'une des sources de carbone et d'énergie principales des cellules, et il est le composé dont la concentration est la mieux prédite par les techniques spectroscopiques (Chapitre II). Un niveau de glucose résiduel a été fixé d'une manière empirique à un point de

consigne, Glc_{target}, de 0,5 g.L^{-1} (2,78 mM). Cette valeur est un compromis entre une valeur faible pour éviter une forte accumulation de lactate, et un niveau suffisant pour éviter une limitation en glucose des cellules. D'après les cultures précédentes, lorsque la concentration en glucose est proche de 0,5 g.L^{-1}, la production de lactate s'arrête et la viabilité cellulaire reste élevée. Conjointement à la spectroscopie diélectrique, c'est la spectroscopie NIR qui a été utilisée puisque la spectrométrie Raman n'était pas disponible au moment des expérimentations. Le schéma de principe de la culture semi-continue pilotée automatiquement par une stratégie d'alimentation adaptative est présenté en Figure 73.

Comme pour les précédentes cultures en mode discontinu et recharge-récolte, cette culture semi-continue est réalisée dans un bioréacteur Pierre Guérin d'un volume utile de 2,5 L, à 37 °C, 50% de pO_2, et à pH 7,2. Les mesures spectroscopiques (permittivités et spectres NIR) acquises en temps réel à l'aide des sondes *in situ* sont transmises à un ordinateur (chaque 20 min). Un programme calcule alors le débit d'alimentation permettant de maintenir le niveau de glucose au point de consigne. Le débit calculé est assuré par une pompe péristaltique (Reglo Analog, Ismatec) qui apporte le milieu concentré dans le réacteur. De plus, une balance (Precisa) est placée sous le bioréacteur afin de mesurer en temps réel le volume de milieu de culture dans le réacteur.

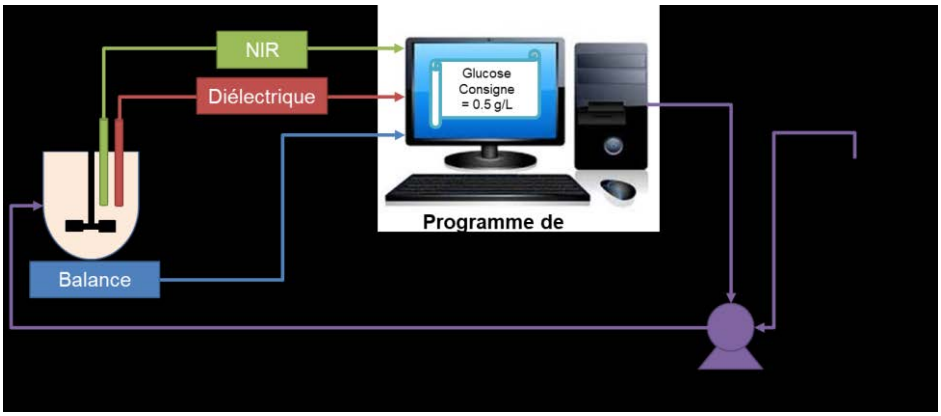


Figure 73 : Schéma du système de pilotage pour l'alimentation adaptative d'une culture en mode semi-continu.

La quantité de glucose consommé est calculée à l'aide de l'équation Eq 57 en faisant la somme de la variation du glucose résiduel prédit ($-d(Glc \cdot V)$) et la quantité du glucose ajouté en connaissant le débit de la pompe ($F \cdot Glc_F \times dt$). Afin de limiter des déviations importantes du calcul de débit liées au bruit et aux lectures aberrantes des différents appareillages, et s'assurer que le débit utilisé reste sur la bonne trajectoire, un régulateur proportionnel est incorporé dans le programme. Il est obtenu à partir d'une comparaison entre le glucose résiduel prédit dans le bioréacteur par la spectroscopie NIR (Glc_{res}) et la valeur de consigne du glucose (Glc_{target}), afin de s'assurer que le débit final utilisé (F_f) peut être corrigé en temps réel, et de garder un niveau de glucose résiduel constant (Figure 74).

Ainsi, la fonction de correction (δF) du débit d'alimentation, peut être obtenu par la différence (ΔGlc) entre le Glc_{res} et le Glc_{target}. Et le débit final utilisé pour piloter le bioréacteur peut être obtenu en faisant la somme du débit calculé à partir du modèle dévelopé précédemment (F_t) et le débit corrigé (δF) :

$$F_f = F_t + \delta F \qquad \textbf{Eq 58}$$

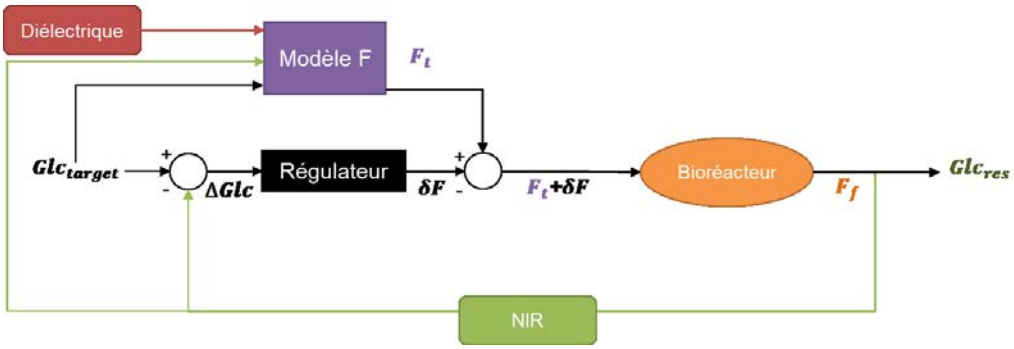


Figure 74 : Schéma de la boucle de régulation pour l'alimentation adaptative d'une culture en mode semi-continu.

Le milieu d'alimentation est un milieu concentré contenant différentes molécules telles que du glucose, des acides aminés, des vitamines, des éléments traces, des facteurs de croissance, de l'hypoxanthine/thymidine, des lipides et du cholestérol (Hyclone Cell Boost 5 supplement, GE lifescience). Il a été choisi du fait de sa richesse en éléments nutritifs pour éviter une limitation en nutriments autres que le glucose. Ce milieu contenant 118 mM de glucose mais étant exempt de glutamine, il est complété avec 15 mM de glutamine pour obtenir la même proportion entre glucose et glutamine que dans le milieu de culture initial.

Une photographie du montage expérimental est présentée en Figure 75. Les régulations classiques (température, pH, pO_2, agitation) sont contrôlées par le logiciel C-Bio sur l'ordinateur PC1. L'ordinateur PC2, sur lequel le programme de pilotage du débit d'alimentation est installé, permet aussi l'acquisition des données en temps réel (NIR, sonde diélectrique, balance, pompe). La balance 1 (Precisa), placée sous le bioréacteur, est tarée au préalable avec le bioréacteur vide complètement équipé. Le milieu d'alimentation concentré est conservé à 4°C dans un réfrigérateur afin d'éviter la dégradation de la glutamine et des vitamines par la chaleur ou la lumière. Une autre balance Precisa (balance 2) est placée sous la bouteille du milieu d'alimentation pour vérifier le volume ajouté. La cuve du bioréacteur est complètement recouverte de papier d'aluminium pour éviter les risques de dégradation des vitamines par la lumière. Un condenseur, relié à un générateur d'eau froide, est

installé sur la platine du bioréacteur afin d'éviter l'évaporation de l'eau du milieu de culture, ce qui pourrait fausser le pilotage. Un boîtier de de conversion (analogique/numérique) a été développé par le service électronique du LRGP pour la pompe péristaltique d'alimentation. Le tuyau de la pompe a un diamètre interne de 0,25 mm (PharMed Ismaprene). Tous les tuyaux et les fils sont calés et/ou fixés pour éviter au maximum des fluctuations non-souhaitées des mesures de la balance.

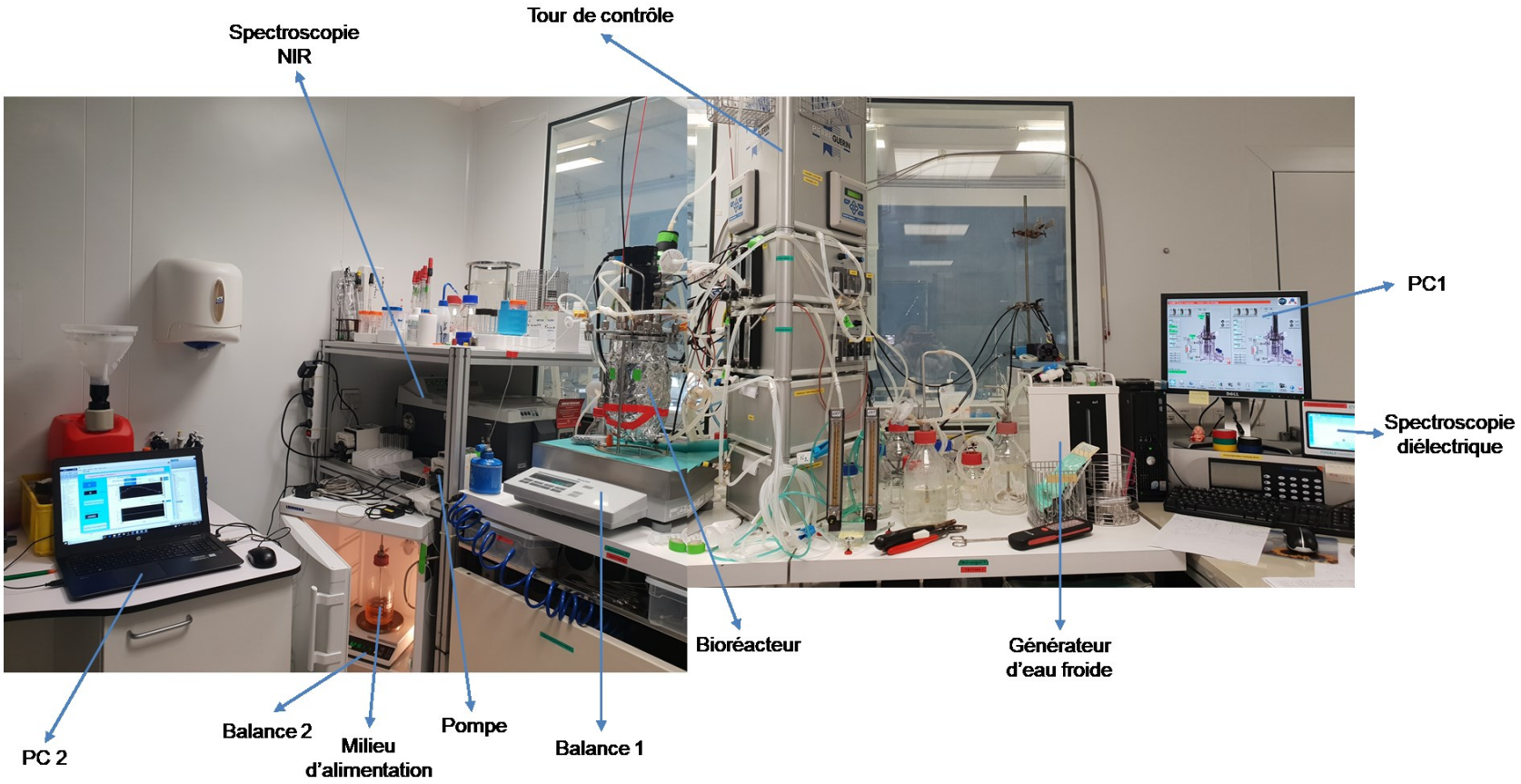


Figure 75 : Photographie de l'installation expérimentale permettant de réaliser une culture semi-continue avec pilotage automatisé de l'alimentation du milieu de culture

IV.3.2. Développement du capteur logiciel

Le capteur logiciel permettant le calcul du débit de la pompe d'alimentation du milieu de culture concentré met en œuvre 3 étapes (Figure 76).

- La première étape concerne l'acquisition des données en ligne. Des programmes Labview ont été développés par le service informatique du LRGP afin de pouvoir communiquer avec les différents appareils d'acquisition et récupérer les mesures en temps réel sur l'ordinateur. Trois types de données sont ainsi récupérés : les mesures de permittivités à partir de la spectroscopie diélectrique chaque 0,5 s ; les mesures de poids à partir de la balance chaque 2 s ; les mesures de la vitesse de pompe chaque 1 s. En raison des instabilités des signaux diélectriques et des variations intrinsèques du poids, ces deux types de mesures sont moyennés sur une base de temps de 2 mn et 1 mn respectivement. Une valeur de chaque type de données est enregistrée dans un fichier bloc-note toutes les 20 min. Le « Data 1 » correspond aux permittivités et au temps de culture en heure (*H_Incyte*). Le « Data 2 » regroupe les poids mesurés, correspondant approximativement au volume du milieu de culture dans le bioréacteur (*V_bio*) puisque la densité du milieu de culture est proche de 1 g.mL^{-1}. Le « Data 3 » correspond à la vitesse de pompe (*Vit_P*) mesurée. Concernant la spectroscopie NIR, les spectres sont enregistrés en format ".csv" toutes les 20 min. Le programme Matlab « Glc_NIR_online.m » permet de lire les fichiers et de récupérer les spectres

- La seconde étape constitue l'étape de calcul. Tout d'abord, le programme Matlab « Glc_NIR_online.m » permet d'appliquer le modèle de prédiction de la concentration de glucose, développé dans le Chapitre II, aux spectres NIR récupérés. Ainsi, le glucose résiduel (Glc_{res}) présent dans le bioréacteur peut être prédit en temps réel par la spectroscopie NIR (*Glc_NIR*). En parallèle, la VCD peut également être prédite en temps réel grâce à la spectroscopie diélectrique et à la régression linéaire développée dans le chapitre I. Par la suite, ces valeurs de VCD prédites sont intégrées dans le programme Matlab « Mu_online.m », utilisant l'approche d'estimation en ligne de μ, développée dans le Chapitre I. Les valeurs de *V_bio* enregistrées, permettent, quant à elles, de convertir les concentrations (mM ou cellules.mL^{-1}) en quantités (mmole ou cellules) pour des calculs ultérieurs. Pour la pompe, les valeurs de *Vit_P* enregistrées en tour.min^{-1} sont converties en débit d'alimentation (F, mL.min^{-1}) à l'aide du coefficient de calibration (*coeff_P*) de 0,0051 mL.tour^{-1}, établi préalablement. Connaissant la concentration de glucose dans le milieu d'alimentation, il est alors possible de calculer la quantité de glucose ajouté à chaque instant t ($F \times Glc_F \times dt$). A l'aide de l'équationEq 57, La quantité totale de glucose consommé peut être ainsi calculée en intégrant la quantité de glucose ajouté ($F \times Glc_F \times dt$) dans la partie de variation de glucose apparente

($-d(Glc \cdot V) = Glc_NIR_{t0} * V_bio_{t0} - Glc_NIR_t * V_bio_t$). Ainsi, en prédisant à chaque instant *t* la quantité de glucose consommé et la quantité de cellules viables formées, le rendement métabolique $Y_{X/Glc}$ peut être calculé à l'aide de l'équation Eq ***54*** (Chapitre IV.3) dans le programme « Y_NIR_online.m ». Enfin, le nouveau débit d'alimentation de référence à utiliser pour le temps *t* (F_t) est calculé par les équations Eq 46-Eq ***53***, basées sur l'écriture des bilans de matières en réacteur semi-continu. Dans ces équations, μ et $Y_{X/Glc}$ sont calculées à partir des programmes développés sous Matlab utilisant les données acquises en ligne, le temps *t* correspond au fichier temps H_Incyte enregistré dans « Data 1 ». V_0 et X_0 sont respectivement le volume du milieu (*V_bio*) et la VCD prédite au moment d'initiation d'alimentation. Glc_F est la concentration de glucose dans le milieu d'alimentation et Glc_{target} est la valeur de consigne du glucose résiduel.

- La dernière étape correspond au calcul du régulateur proportionnel permettant d'apporter une fonction de correction (δF) au débit calculé (F_t), afin de s'assurer que le débit final utilisé (F_f) reste sur la bonne trajectoire en maintenant une valeur constante pour le glucose résiduel. Les valeurs de *Glc_NIR* issues du programme « Glc_NIR_online.m » sont tout d'abord lissées en temps réel par un filtre SG (degré 1 avec une fenêtre de 15 points (3 h environ)). Puis, toutes les 20 min, la valeur lissée ($Glc_{NIRliss}$) à l'instant *t* est comparée à la valeur de Glc_{target}. Le δF est alors calculé avec l'équation suivante :

$$\delta F = -\frac{\left(Glc_{NIRliss} - Glc_{target}\right)}{Glc_{target}} \times 0{,}3 \times F_t \qquad \textbf{Eq 59}$$

La valeur 0,3 est choisie empiriquement pour cette étude sur la base des résultats de simulation. La nouvelle consigne (*Vit_Pnew*) pour la vitesse de la pompe est alors égale à :

$$Vit_Pnew = \frac{F_{ref} + \delta F}{Coef f_P} \qquad \textbf{Eq 60}$$

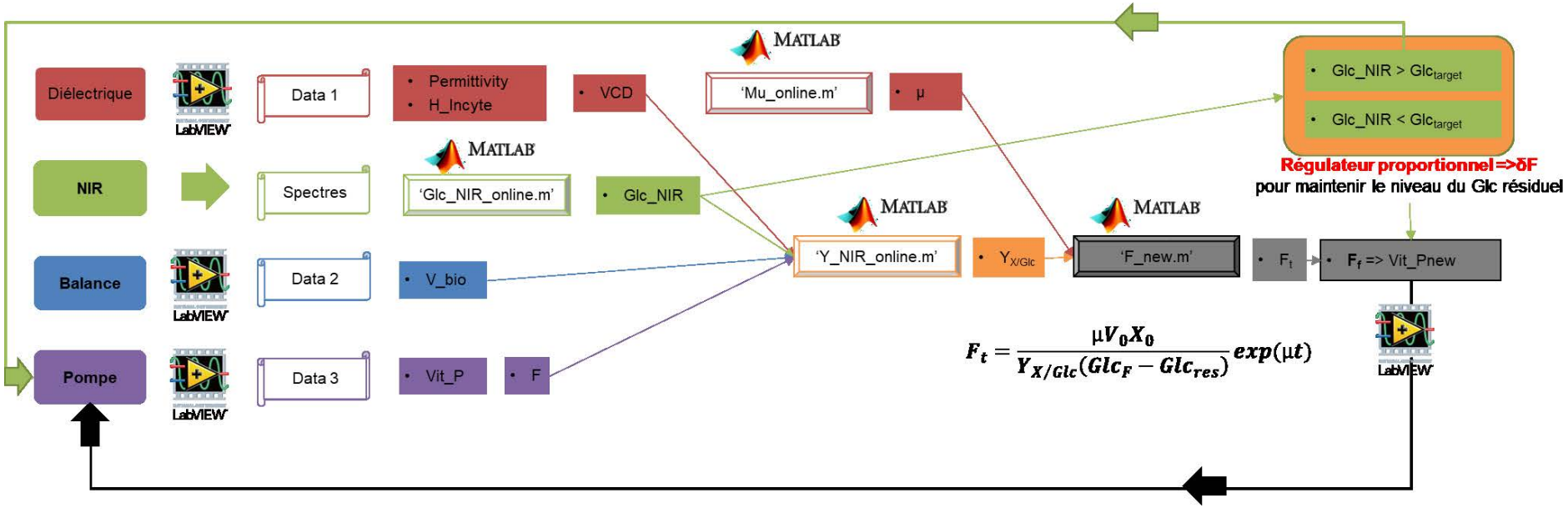


Figure 76: Développement du capteur logiciel pour le pilotage de la culture semi-continue

IV.3.3. Interface du programme de pilotage

L'interface du programme de pilotage de l'alimentation du milieu de culture a été développée sous Labview avec des scripts Matlab intégrés. Elle permet de visualiser les données mesurées, de tracer des courbes, ainsi que d'entrer ou de calculer manuellement les consignes (Figure 77-Figure 79).

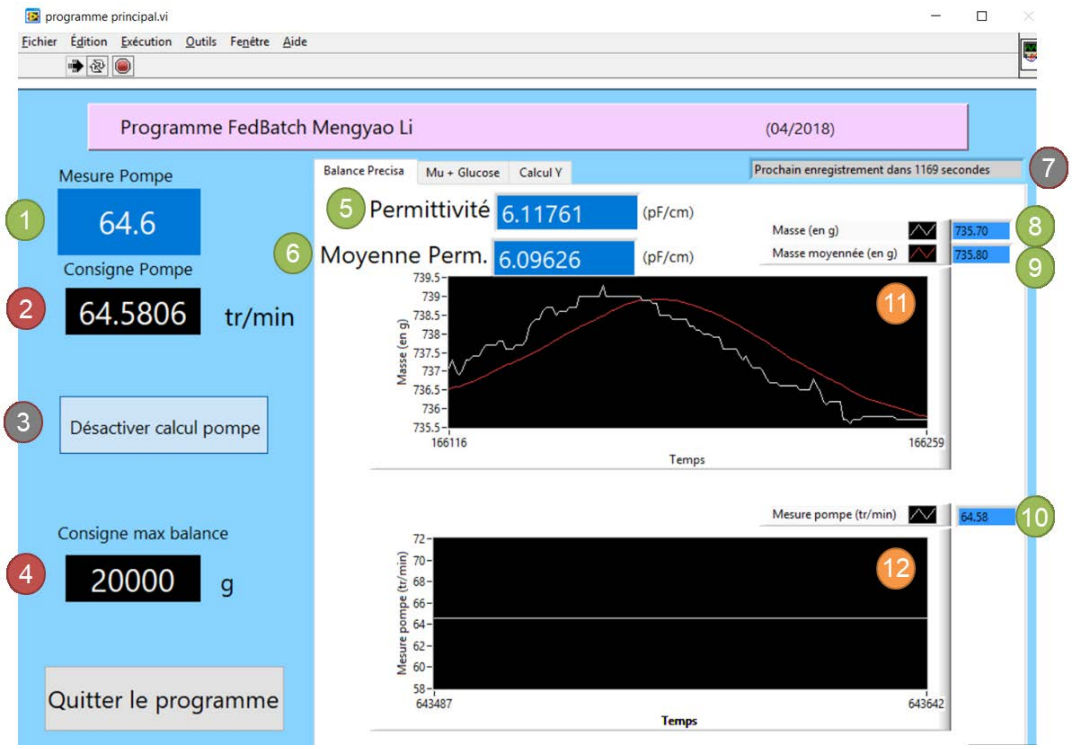


Figure 77 : Interface du programme de pilotage (onglet 1)

•Dans le 1er onglet, les cases bleues (1, 5, 6, 8, 9, 10) correspondent à l'affichage des données acquises en ligne. 1 et 10 : vitesse actuelle de la pompe en tour.min^{-1} ; 5 : permittivité instantanée mesurée en pF.cm^{-1} ; 6 : valeur moyenne des permittivités mesurées pendant 2 mn ; 8 : poids instantané de la suspension cellulaire en g ; 9 : valeur moyenne des mesures du poids de la suspension cellulaire pendant 1 mn.

•Les cases noires (2, 4) correspondent à des valeurs de consigne. 2 : consigne pour la vitesse de la pompe d'alimentation correspondant à *Vit_Pnew* calculée à partir des programmes Matlab. Cette valeur de consigne peut potentiellement être renseignée manuellement. La valeur maximale autorisée pour cette consigne est de 66 tour.min^{-1}. 4 : consigne maximale pour la balance positionnée sous le réacteur. Quand la mesure de poids dépasse cette valeur, le programme envoie

une valeur 0 pour la consigne de la pompe et le programme s'arrête. En début de culture, une valeur élevée (20000 g) est entrée manuellement pour éviter de bloquer le programme de pilotage par des variations de poids provenant des manipulations, par exemple au moment des prélèvements des échantillons. Aux derniers jours de culture, lorsque le volume du bioréacteur s'approche du volume final (2,3 L pour cette étude), la consigne est alors ajustée (2300 g) pour que le volume total du bioréacteur ne dépasse pas la valeur seuil.

●Les tracés de courbes (11, 12, 13, 14, 15) sont affichés en temps réel. 11 : poids mesuré (blanc) et poids moyenné (rouge) sur un intervalle de 2 mn. 12 : vitesse de pompe mesurée et moyennée sur un intervalle de 2 mn. Dans le second onglet (Figure 78), 13 : calcul en ligne de μ (h^{-1}) en fonction du nombre de points enregistrés. 14 : prédiction de la concentration du glucose (mM) par la spectroscopie NIR en fonction du nombre de points enregistrés. Dans le dernier onglet (Figure 79), 15 : calcul de $Y_{X/Glc}$ (10^8 cellules.mmole^{-1}) en fonction du nombre de points enregistrés.

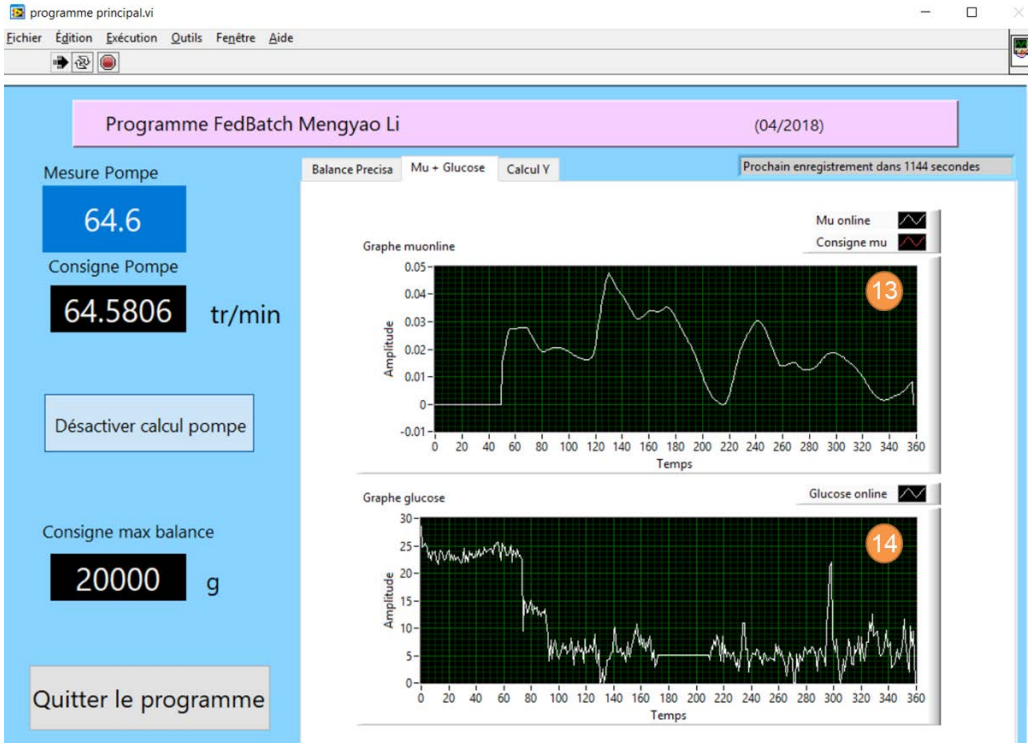


Figure 78 : Interface du programme de pilotage (onglet 2)

●Le reste des affichages correspond aux modules de fonctionnement. Dans le 1er onglet (Figure 77), 3 : bouton d'activation du programme. Avant son activation, seul l'enregistrement des différents types de données est possible. Une fois activé, le programme démarre les calculs pour obtenir la nouvelle consigne de vitesse de pompe (*Vit_Pnew*) toutes les 20 min. 7 : décompteur de temps

permettant de connaître le temps restant avant le prochain enregistrement/calcul, afin d'éviter de perturber les mesures de balance par des prélèvements. Dans le 3ème onglet (Figure 79), 16 : fenêtre indiquant les erreurs éventuelles venant des scripts Matlab, ce qui permet de les corriger rapidement afin de ne pas bloquer le programme de pilotage pendant l'expérience. 17 : voyant permettant de visualiser le bon fonctionnement des scripts Matlab.

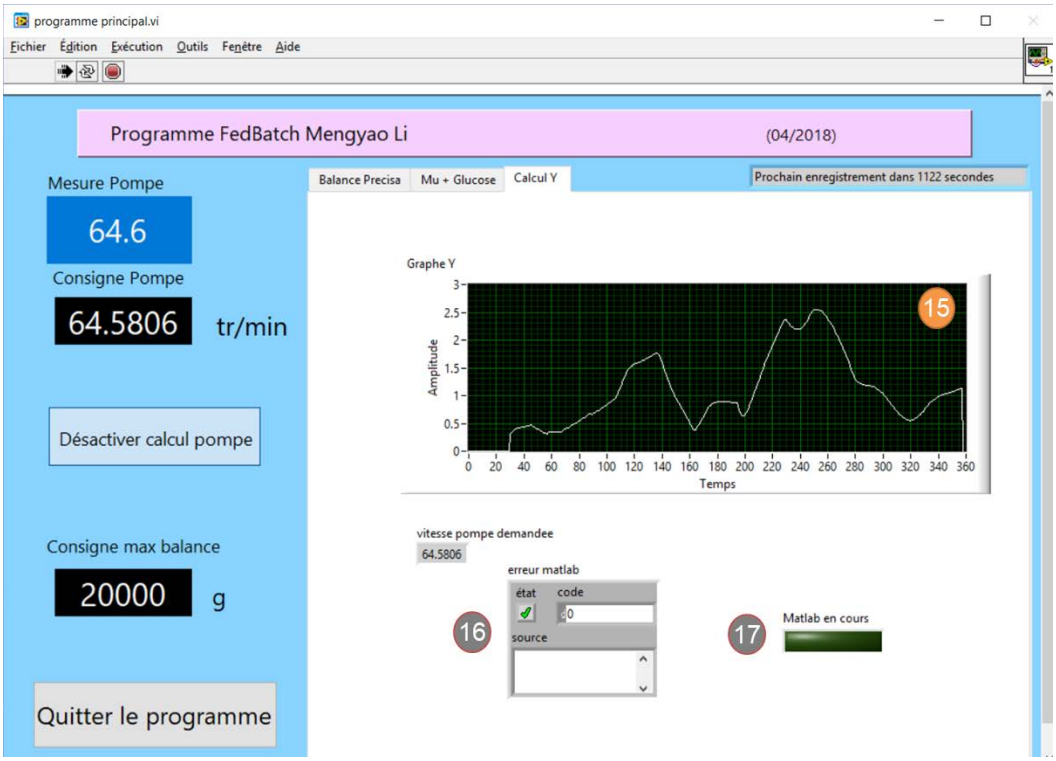


Figure 79 : Interface du programme de pilotage (onglet 3)

IV.4. Résultats du pilotage automatisé de l'alimentation d'une culture semi-continue

La culture semi-continue est démarrée par une phase discontinue de 3 jours, avec un volume de milieu d'1 litre. Cette phase discontinue est effectuée pour 3 raisons principales :

- le milieu de culture initial contenant une forte concentration de glucose, celle-ci doit d'abord diminuer pour se rapprocher du point de consigne afin de ne pas perturber les calculs ;
- les programmes nécessitent une certaine accumulation de données pour initier les calculs ;

- la durée de la phase discontinue doit être fixée pour faciliter la comparaison avec les autres cultures semi-continues pilotées en mode manuel.

IV.4.1. Prédictions des paramètres cinétiques et calcul en temps réel du débit d'alimentation

Les prédictions en ligne de la concentration de glucose résiduel Glc_{res}, du rendement $Y_{X/Glc}$, et de μ, ainsi que le calcul en ligne du débit de la pompe *F* sont présentés dans la Figure 80. L'alimentation du milieu concentré est démarrée à 72 h de culture (par le bouton « Activer calcul pompe ») et s'arrête automatiquement lorsque le poids de la suspension cellulaire atteint 2300 g (ici au bout de 213 h).

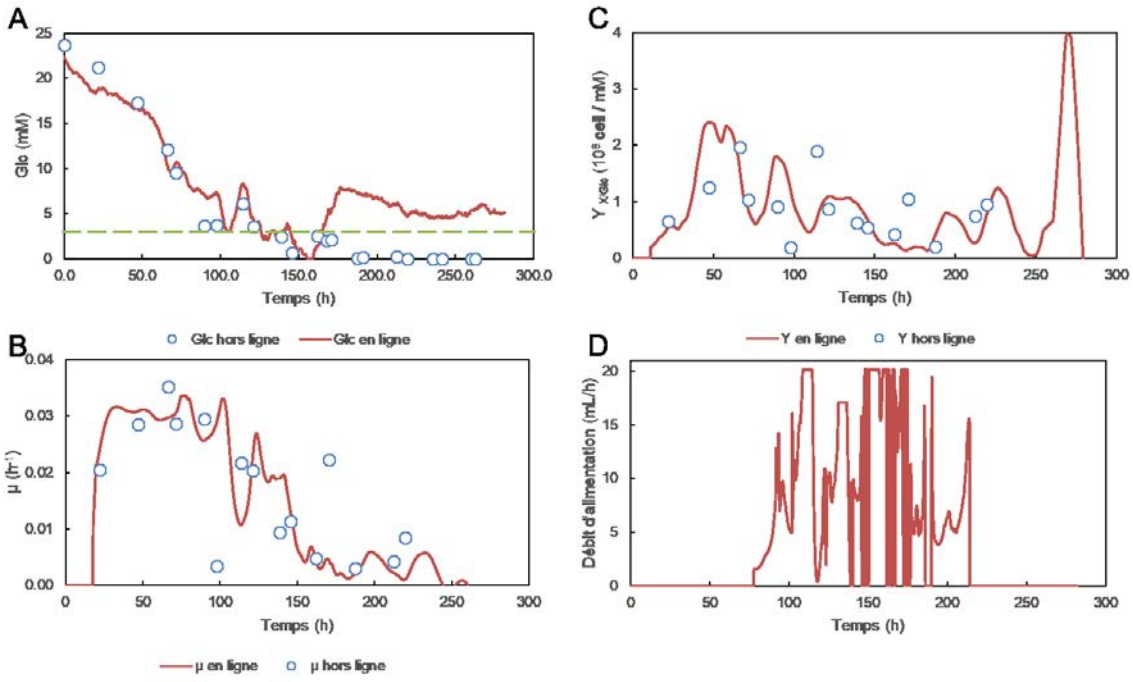


Figure 80 : Prédictions en ligne (courbes rouges) : glucose résiduel (A), μ (B), et $Y_{X/Glc}$ (C). Calcul en ligne du débit de la pompe (D). Valeurs calculées à partir des mesures hors ligne (cercles bleus)

Concernant la prédiction de la concentration de glucose résiduel par spectroscopie NIR, les valeurs sont proches de celles obtenues hors ligne pendant la première partie de la culture. A partir de 160 h et jusqu'à la fin de la culture, les valeurs prédites présentent un décalage d'environ 5 mM de plus par rapport aux valeurs mesurées hors ligne (Figure 80.A). Néanmoins, la tendance de ces deux séries de valeurs reste cohérente, puisqu'on observe dans les deux cas une diminution entre 175 et 200 h, suivie d'une stabilisation jusqu'à la fin de culture. Une raison possible de ce décalage à partir d'environ 160 h pourrait être liée au fait que le modèle de prédiction du glucose a été développé

à l'aide de données de cultures réalisées en mode discontinu ou recharge-récolte avec une matrice de milieu de culture homogène, correspondant au milieu initial. Or, dans le cas de la culture en mode semi-continu, un milieu d'alimentation de composition différente de celle du milieu de culture initial est ajouté progressivement dans la matrice d'origine, introduisant ainsi une hétérogénéité de plus en plus importante. Cette hétérogénéité de la matrice pourrait engendrer une modification de la dispersion de la lumière en générant des variations dans les spectres NIR (Cervera *et al.*, 2009). Entre 150 et 160 h de culture, on observe une vitesse très importante de la pompe, ce qui se traduit par une quantité importante de milieu d'alimentation concentré ajoutée dans le bioréacteur. Vers 160 h, le volume du milieu d'alimentation représente ainsi presque 50% du volume total. Par conséquent, l'hétérogénéité de la matrice devient importante et réduit probablement la capacité de prédiction du modèle NIR. De ce fait, la régulation de la concentration en glucose résiduel est plutôt stable autour du point de consigne (2,78 mM) entre 100 et 160 h, mais elle dévie après cette période. En effet, alors que la concentration résiduelle du glucose chute à 0 mM vers 187 h d'après les mesures hors ligne, les prédictions sont de l'ordre de 5 mM. Pour améliorer cette prédiction de la concentration du glucose par spectroscopie NIR au cours de cultures semi-continues, il serait intéressant d'inclure, dans le développement du modèle de prédiction, des données spectrales présentant une plus grande hétérogénéité au niveau de la matrice de milieu de culture, ou bien d'utiliser des méthodes de prétraitement afin d'atténuer cet effet de matrice.

Les valeurs de μ sont plutôt bien prédites en ligne (Figure 80.B). Ainsi, μ atteint rapidement sa valeur maximale (autour de 0,03 h^{-1}) au début de culture (vers 25 h) et s'y maintient jusqu'à 100 h. Puis, elle diminue pour atteindre une valeur assez constante (environ 0,005 h^{-1}) entre 150 et 250 h, avant de devenir nulle après 250 h.

Le calcul en ligne du rendement $Y_{X/Glc}$ est présenté dans la Figure 80.C. Les valeurs prédites sont assez proches des valeurs calculées à partir des mesures hors ligne en dépit du fait que, la prédiction de $Y_{X/Glc}$ nécessite des données en ligne issues de deux spectroscopies (diélectrique et NIR). L'évolution générale de $Y_{X/Glc}$ semble être bien suivie par le modèle développé, avec une valeur maximale autour de $2{,}2.10^8$ cellules.mmole^{-1} vers 50 h. Puis, ce rendement diminue progressivement pour atteindre $0{,}2.10^8$ cellules.mmole^{-1} vers 190 h. Par la suite, sa valeur remonte jusqu'à 1.10^8 cellules.mmole^{-1} avant de redescendre après 220 h. Il est à noter que cette dernière diminution de la valeur de $Y_{X/Glc}$ n'est pas bien prédite par le modèle de calcul en ligne, un artefact étant observé entre 250 h et la fin de culture. La raison principale réside dans le fait que, après 220 h, la concentration de cellules viables est en déclin et que des cellules mortes s'accumulent. La régression linéaire déterminée avec une forte viabilité n'est donc plus adaptée pour prédire la VCD à partir de la permittivité, ce qui

génère une surestimation de la valeur de VCD. De plus, la pompe arrête de tourner après 214 h (Figure 80.D). Les variations de glucose ne proviennent alors que du glucose résiduel prédit par la spectroscopie NIR (Eq 57 : F=0, $Glc_{Conso} \times V = -d(Glc \cdot V)$), alors que ces valeurs prédites ne sont plus assez précises sur cette période. De ce fait, le calcul de $Y_{X/Glc}$ n'est plus fiable en fin de culture.

Enfin, le débit d'alimentation calculé en ligne est présenté Figure 80.D. Les variations de ce débit semblent assez brutales, en particulier sur la période entre 150 et 176 h. Cette observation peut provenir de deux effets différents : une fonction de correction δF mal calculée en raison de la déviation sur la prédiction du glucose résiduel, et/ou des artefacts dans les mesures de permittivités qui influencent la prédiction de VCD.

IV.4.2. Comparaison de différentes stratégies d'alimentation des cultures semi-continues

Dans ce chapitre, les performances de la culture semi-continue conduite avec un pilotage automatisé (culture FB3) sont comparées à celles de 2 cultures semi-continues utilisant différentes stratégies d'alimentation manuelles (FB1 et FB2) et de la culture discontinue de référence (B0). L'alimentation en milieu de la culture FB1 est réalisée en mode manuel avec un ajout ponctuel de 5% (v/v) de milieu concentré : 1 fois par jour aux jours 3, 4, 5 et 6, puis 2 fois par jour à 8 h d'intervalle aux jours 7, 8, 9 et 10. Pour la culture FB2, l'apport de milieu est réalisé en continu à l'aide d'une pompe à vitesse variable. Le débit de la pompe est calculé et modifié manuellement 2 fois par jour à 8 h d'intervalle, à partir du 3ème jour de culture. L'augmentation de ce débit est de type exponentiel, le débit étant calculé à l'aide de la même équation que celle utilisée pour la culture FB3 (Eq 53), en fixant des valeurs constantes pour μ (0,015 h^{-1}) et $Y_{X/Glc}$ (1,88.10^{8} cellules.mmole^{-1}). Cette valeur de μ correspond à environ la moitié de la valeur maximale observée en culture discontinue. En effet, d'après certains auteurs, la productivité en anticorps peut être améliorée lors du ralentissement de μ lors de cultures semi-continues ou continues (Suzuki and Ollis, 1990; Jang and Barford, 2000). La valeur fixée pour $Y_{X/Glc}$ est calculée de manière empirique à partir des données des cultures réalisées en mode discontinu. Les valeurs initiales de X_0 et V_0 sont également fixées dans l'équation. Elles correspondent aux valeurs initiales de VCD et de volume du bioréacteur au moment où l'alimentation est démarrée.

Les résultats des performances des différentes cultures sont rassemblés dans le Tableau 17. Globalement, les cultures semi-continues avec alimentation par un milieu concentré permettent, par

rapport à la culture discontinue, de prolonger la durée de la culture de 4 et 6 jours supplémentaires, d'atteindre une concentration de cellules viables plus élevée (sauf pour FB3) et d'augmenter la quantité des IgG produites. De plus, il est intéressant de noter que les conditions de la culture FB3, avec alimentation automatisée, permettent de doubler la productivité spécifique globale, en pg d'IgG produit par cellule et par jour, par rapport aux autres cultures. De même, le rendement global $Y_{IgG/X}$ passe de 4,6 et 4,2 mg.10^8 cellule^{-1}, respectivement pour les cultures FB1 et FB2, à 7,7 mg.10^8 cellule^{-1} pour la culture FB3.

Tableau 17 : Comparaison des performances des cultures en mode discontinu et semi-continu

Culture	Mode	Stratégie d'alimentation	Volume ajouté (mL)	Durée de culture (j)	VCD max (10^5 cellule.mL^{-1})	IgG max (mg.L^{-1})	Productivité spécifique globale (pg.cellule^{-1}.j^{-1})
B0	Discontinu	/	0	7	86	225	3,7
FB1	Semi-continu	Ponctuelle + Manuelle	841	13	154	700	3,5
FB2	Semi-continu	Continue exponentielle + Manuelle	1000	12	160	663	3,5
FB3	Semi-continu	Continue +Automatisée	1000	11	54	416	7

Les courbes d'évolution en fonction du temps des concentrations de cellules viables et d'IgG, de la productivité spécifique en IgG, et du pourcentage de molécules d'IgG glycosylées sont présentées dans la Figure 81. La Figure 82 rapporte les évolutions au cours du temps des quantités cumulées, en mmole, de glucose et de glutamine consommés, ainsi que de lactate et d'ions ammonium produits. Enfin, la Figure 83 donne les évolutions au cours du temps des concentrations de glucose et de glutamine, en mM, ainsi que du volume de milieu de culture au sein du réacteur.

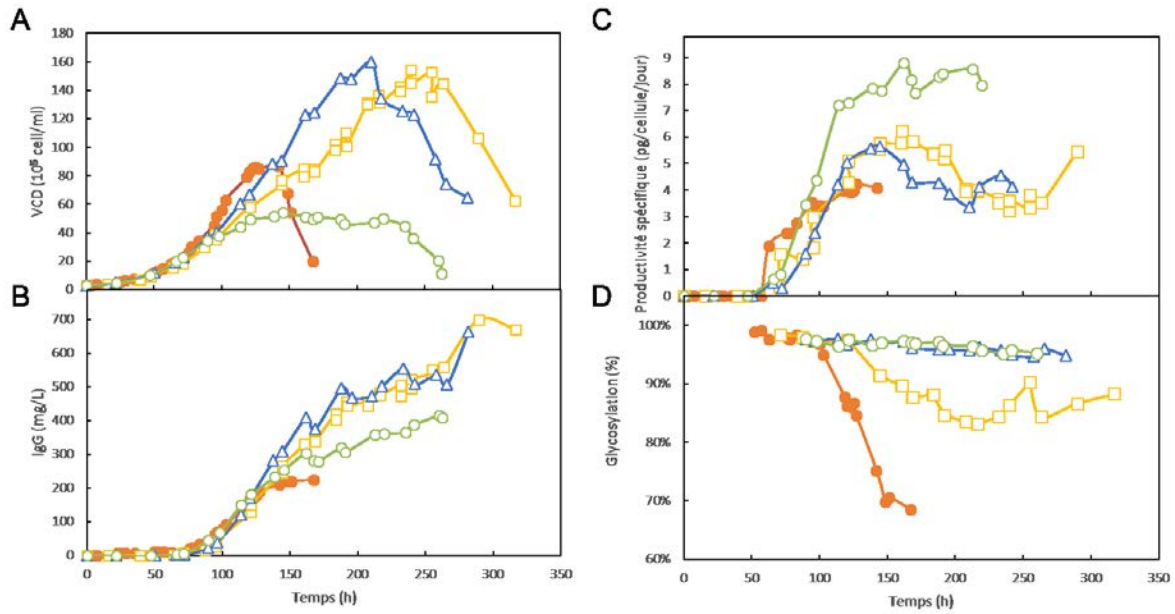


Figure 81 : Evolution de la concentration de cellules viables (A) et d'IgG (B), de la productivité spécifique (C) et de la glycosylation des IgG (D) au cours des différentes cultures : B0 (●), FB1 (□), FB2 (△), FB3 (○)

Les cultures FB1 et FB2 qui permettent d'atteindre la VCD la plus élevée, suivent un profil de croissance cellulaire similaire à celui de la culture discontinue B0, avec 4 phases : latence, exponentielle, ralentissement et déclin (Figure 81.A). Par contre, la courbe de VCD pour la culture FB3 présente une phase stationnaire de plus d'une centaine d'heures, entre la phase de ralentissement et la phase de déclin, ce qui est souvent caractéristique des cultures semi-continues. Lors de cette phase stationnaire (entre 120 et 230 h), μ chute fortement pour atteindre la valeur très faible proche de 0 (Figure 80.B). De plus, au cours de cette phase stationnaire de VCD, les cellules consomment plus de glucose et de glutamine que dans les autres cultures FB1 et FB2 (Figure 82.A, C).

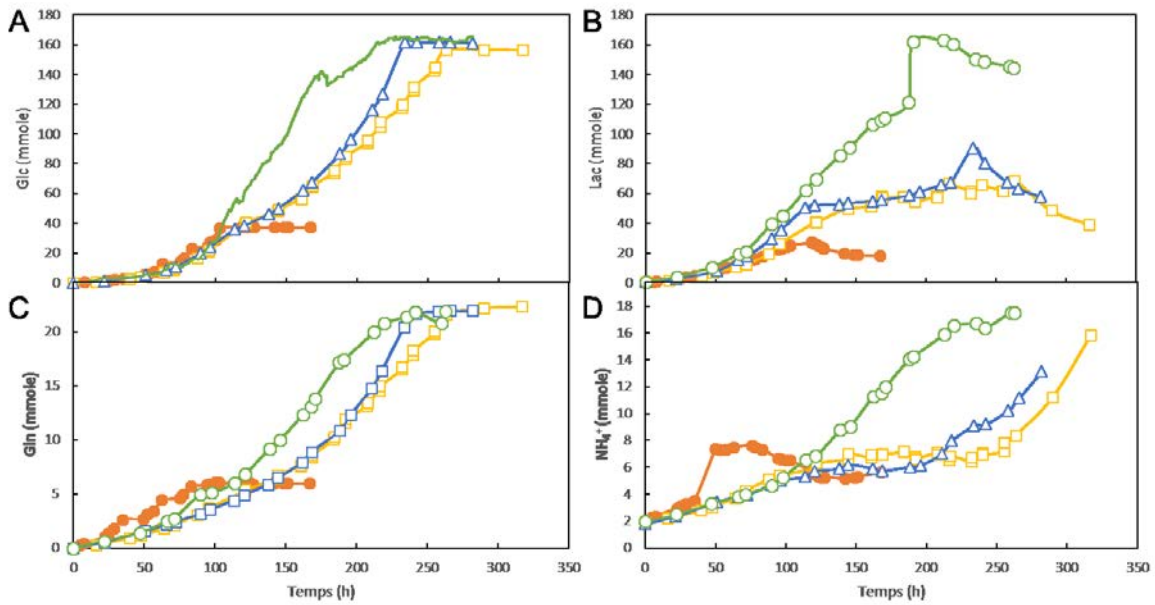


Figure 82 : Evolution de la quantité cumulée de glucose consommé (A), lactate produit (B), glutamine consommée (C) et ions ammonium produits (D) au cours des cultures : B0 (●), FB1 (□), FB2 (△), FB3 (○)

Ainsi, pendant cette phase stationnaire, il semble que l'énergie apportée par les substrats consommés et le pool des acides aminés, est utilisée en priorité pour la production des IgG (Figure 81.B), la production de cellules et la synthèse des IgG entrant en compétition (Jang et Barford, 2000). Cette hypothèse peut être confirmée par la Figure 81.C, puisque la productivité spécifique des cellules calculée d'une manière instantanée au cours de la culture FB3, est plus importante que celle des autres cultures pendant la phase stationnaire de VCD, en se maintenant à des valeurs entre 1,5 et 2 fois plus élevées que celles des cultures FB1 et FB2 jusqu'à la fin de culture.

Concernant la glycosylation des IgG produites, les cultures semi-continues permettent d'augmenter notablement le pourcentage des IgG glycosylées par rapport à la culture discontinue B0. En particulier, dans les cultures FB2 et FB3, les IgG produites sont quasiment glycosylées à 100% jusqu'à la fin de la culture, ce qui entraîne une amélioration significative de la qualité globale des IgG récoltées. L'évolution des 4 principaux groupes de glycoformes a aussi été analysée pour chaque culture, sans qu'une différence significative ne soit observée.

Cependant, en dépit de l'intérêt de la culture FB3, en termes de productivité spécifique et de qualité des IgG, il est observé une production de lactate et d'ions ammonium plus importante que dans les cultures FB1 et FB2 (Figure 82.B, D). Ceci laisse supposer une utilisation peu efficace des sources énergétiques.

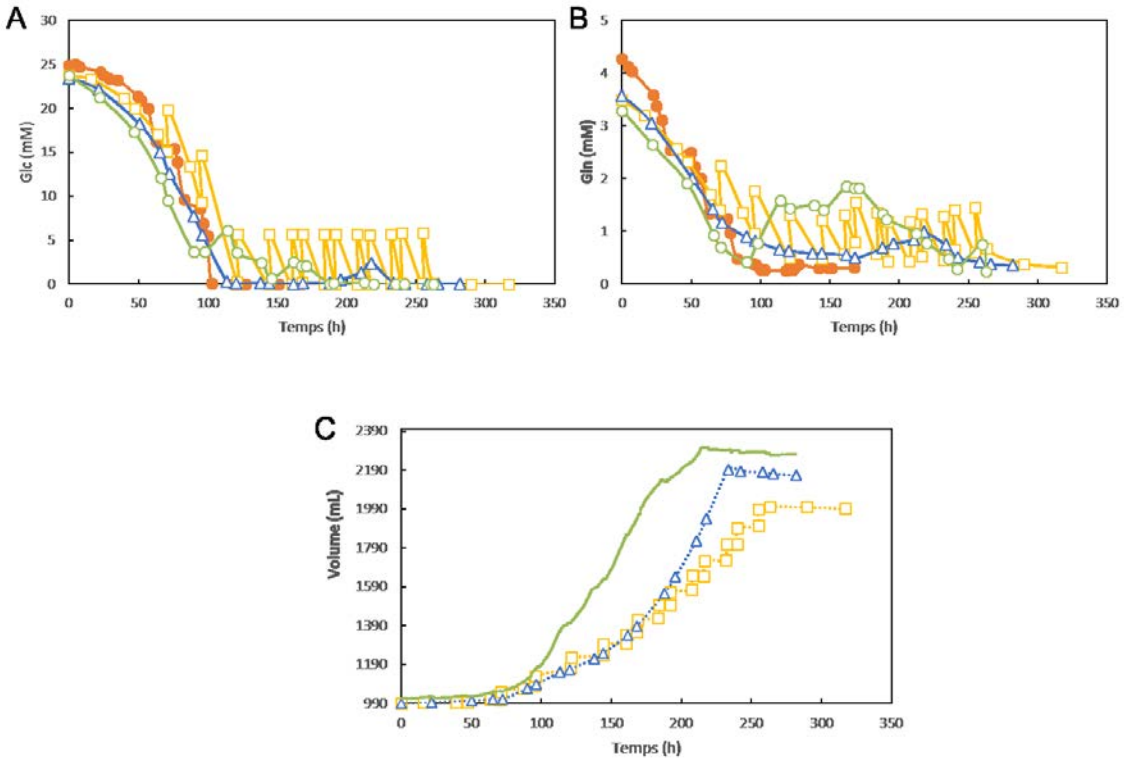


Figure 83 : Evolution de la concentration de glucose (A) et de glutamine (B), ainsi que du volume de milieu dans le bioréacteur (D) au cours des différentes cultures : B0 (●), FB1 (□), FB2 (△), FB3 (○)

Cette production importante de produits issus du métabolisme cellulaire est probablement l'une des raisons de l'inhibition de la croissance cellulaire à partir de 120 h pour la culture FB3 (Lao and Toth, 1997). Une autre raison pourrait être liée à une osmolalité trop élevée du milieu de culture FB3, puisque le volume du milieu concentré, ajouté entre 100 et 200 h, est 2 fois plus élevé que pour les cultures FB1 et FB2 (Figure 83.C). Ainsi, afin de diminuer la production de lactate et d'ions ammonium tout en limitant l'augmentation de l'osmolalité du milieu, la consigne de la concentration de glucose résiduel pourrait être fixée à une valeur plus basse. Une autre suggestion serait de faire une double alimentation : l'une avec une solution concentrée de glucose et glutamine, et l'autre avec un milieu d'alimentation exempt de glucose et de la glutamine. Dans ce cas, le pilotage automatique ne serait effectif que pour la solution concentrée, le milieu de culture étant ajouté plus modérément afin d'éviter un excès avec les autres nutriments tels que les acides aminés, les vitamines et les éléments traces.

En synthèse de ces résultats comparatifs, il peut être observé qu'une stratégie d'alimentation d'une culture semi-continue, mettant en œuvre un pilotage automatisé basé sur des mesures

spectroscopiques en ligne combinées, peut présenter des avantages en permettant de doubler la productivité spécifique du procédé et d'assurer la qualité de glycosylation des IgG produites. Cependant, ces résultats ne doivent être considérés que comme tout à fait préliminaires. Des études complémentaires sont en effet nécessaires pour optimiser les paramètres de pilotage en vue d'obtenir une concentration plus élevée de cellules et d'IgG tout en conservant la qualité des IgG produites.

IV.5. Conclusion

L'objectif principal de ce chapitre était de tirer profit des études précédentes en utilisant différentes techniques spectroscopiques pour développer un capteur logiciel de type boîte grise en vue de contrôler le débit d'alimentation en milieu de culture au cours d'une culture de cellules CHO en mode semi-continu.

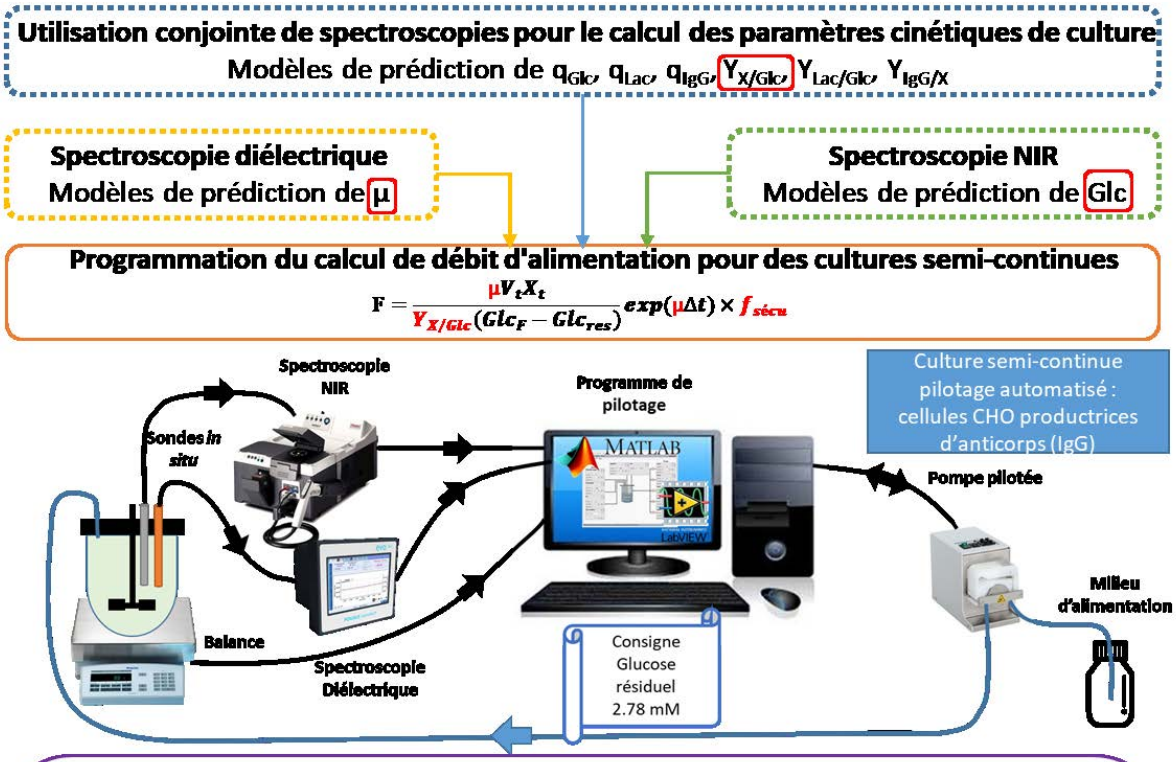
Dans un premier temps, il a été montré que, grâce à l'utilisation conjointe de différentes techniques spectroscopiques *in situ* (NIR/diélectrique et Raman/diélectrique), il était possible de prédire en temps réel certains paramètres cinétiques d'une culture en mode discontinu, tels que les vitesses spécifiques de consommation/production (q_{Glc}, q_{Lac}, q_{IgG}) et les rendements métaboliques ($Y_{X/Glc}$, $Y_{Lac/Glc}$, $Y_{IgG/X}$). De manière générale, les grandes tendances d'évolution ont été plutôt bien prédites par les modèles précédemment développés pour la calibration des mesures spectroscopiques. Il est à noter que les modèles utilisant les spectres Raman permettent de mieux prédire ces différents paramètres, en particulier en début de culture. Néanmoins, ces prédictions de paramètres cinétiques ayant été développées en culture discontinue, le transfert des modèles développés à des cultures semi-continues devra intégrer la variation du volume du bioréacteur et l'ajout de substrat concentré par le milieu d'alimentation.

Dans un second temps, la conception et la mise au point technique d'une culture semi-continue avec pilotage automatisé ont été décrites à travers 3 parties : le montage expérimental, le développement du capteur logiciel et l'interfaçage du pilotage. Un algorithme simple a été choisi pour le capteur logiciel permettant de calculer le débit de la pompe d'alimentation. Pour ce premier essai, seul les modèles permettant la prédiction en temps réel de μ, de la concentration de glucose résiduel et du rendement $Y_{X/Glc}$ ont été utilisés, à partir des données en ligne des spectroscopies diélectrique et NIR. Afin de limiter les erreurs intrinsèques liées aux outils spectroscopiques et les erreurs de calculs, il a

été appliqué un régulateur proportionnel, basé sur la comparaison entre la concentration de glucose prédite et la valeur de consigne pour la concentration de glucose résiduel fixée à 0,5 $g.L^{-1}$ (2,78 mM).

La troisième partie a abordé la mise en œuvre de la culture semi-continue automatisée. Les résultats des calculs en temps réel des paramètres cinétiques indiquaient que les variations de μ et $Y_{X/Glc}$ étaient bien suivies par le capteur logiciel développé. Néanmoins, pendant la seconde partie de la culture, probablement en raison d'une augmentation de l'hétérogénéité de la matrice du milieu de culture, la prédiction de la concentration du glucose résiduel par la spectroscopie NIR s'est écartée des valeurs mesurées hors ligne. Par conséquent, la régulation de la concentration de glucose résiduel a été perturbée pendant cette période, du fait des variations brutales du débit calculé pour la pompe. Les performances de la culture semi-continue automatisée ont alors été comparées à celles d'une culture de référence en mode discontinu et de deux cultures semi-continues pilotées manuellement (alimentation ponctuelle ou exponentielle). De façon très intéressante, le pilotage automatisé a permis d'augmenter la productivité spécifique en IgG tout en assurant une qualité constante de leur glycosylation. Néanmoins, cette culture automatisée ne représente qu'un essai préliminaire. Les différents paramètres de pilotage devront être réévalués de manière plus approfondie dans le futur afin d'améliorer la stratégie d'alimentation de telles cultures semi-continues, en permettant de répondre à la double exigence de productivité et de qualité.

Les résultats principaux de ce chapitre sont présentés dans la Figure 84.



Comparaison des stratégies d'alimentation des cultures semi-continues :

Mode de culture	Productivité spécifique (pg.cellule.jour)	Glycosylation des anticorps (%)
Discontinu	-	- -
semi-continu manuel (ajout ponctuel)	-	-
semi-continu manuel (ajout exponentiel)	-	+
semi-continu automatisé	++	+

Figure 84 : Synthèse schématique des résultats du chapitre IV

Conclusion générale

&

Perspectives

Conclusion générale

L'objectif général de ce travail de thèse était de développer, dans le cadre d'une démarche PAT, de nouvelles approches de suivi en ligne, par des méthodes spectroscopiques, des procédés de culture de cellules CHO en bioréacteur, en vue d'améliorer les performances de production des anticorps (IgG) monoclonaux. Trois types de spectroscopies (diélectrique, NIR et Raman) ont été utilisés et combinés avec des méthodes chimiométriques pour suivre en ligne, soit des paramètres classiques (éléments nutritifs, métabolites), soit des paramètres plus innovants (IgG, glycosylation des IgG, paramètres cinétiques), au cours de procédés réalisés en mode discontinu, recharge-récolte et semi-continu. Les performances des spectroscopies NIR et Raman pour le suivi en ligne ont été comparées et combinées. Les modèles développés à partir des spectroscopies diélectrique et NIR ont été intégrés dans un capteur logiciel de type boîte grise en vue de piloter de manière automatique le débit d'alimentation d'une culture semi-continue. Une synthèse des principales contributions des travaux est présentée dans la Figure 85 et détaillée ci-dessous, avant de proposer différentes perspectives visant à confirmer, compléter et élargir la démarche engagée.

Suivi en ligne de l'état physiologique des cellules en réacteurs discontinus et recharge-récolte

La vitesse spécifique de croissance cellulaire (μ) affecte la glycosylation des IgG produites

Au cours de cultures de cellules CHO réalisées en mode discontinu, il a été observé que lorsque les principaux éléments nutritifs (glucose, glutamine) devenaient limitants, μ commençait à diminuer, indiquant un changement de l'état physiologique défavorable à la croissance cellulaire. De plus, cette diminution de μ coïncidait avec la production d'IgG non-glycosylées (jusqu'à 30 % en fin de culture), conduisant ainsi à une diminution significative de la qualité du produit final. En effet, dans des conditions défavorables à la croissance cellulaire, les modifications post-traductionnelles des IgG produites ne peuvent plus être assurées. Ainsi, afin de permettre une glycosylation des IgG la plus correcte possible, le suivi en ligne de μ, comme indicateur direct de l'état physiologique des cellules, représente un réel défi.

Le suivi en ligne de l'état physiologique des cellules, via l'estimation de μ, peut être réalisé par spectroscopie diélectrique

A partir des permittivités mesurées en ligne par spectroscopie diélectrique, la concentration des cellules viables (VCD) a été évaluée par régression linéaire. La vitesse spécifique de croissance

cellulaire a alors été calculée, en temps réel, à l'aide d'une régression linéaire locale avec fenêtre glissante à partir des valeurs logarithmiques des VCD prédites. Différents filtres de lissage ont été utilisés pour stabiliser les calculs. Divers paramètres de calcul, tels que la largeur de fenêtre de régression et la largeur de fenêtre de lissage, ont été optimisés pour limiter l'effet du bruit de fond tout en conservant la sensibilité de détection du modèle. Les résultats obtenus en temps réel à partir des données spectroscopiques ont été comparés à ceux issus des calculs réalisés en fin de culture à partir de la mesure hors ligne des VCD.

La détection en temps réel du moment critique où μ diminue significativement est possible

Afin de pouvoir contrôler μ en temps réel en agissant sur l'alimentation en éléments nutritifs, le moment critique où μ commence à diminuer significativement doit être rapidement détecté. Néanmoins, deux sources de variations impactent les valeurs de μ déterminées en ligne par spectroscopie diélectrique : la "variation non-biologique" qui résulte de l'instabilité des signaux, et la "variation biologique" qui est liée au changement de l'état physiologique des cellules. La variation totale des valeurs de μ a été déterminée en temps réel en calculant la différence résiduelle entre les valeurs estimées de μ et une valeur maximale de μ préalablement définie. La "variation non-biologique" de μ a alors été évaluée pendant la première partie de la culture correspondant à la phase exponentielle de croissance cellulaire, durant laquelle aucune "variation biologique" n'intervient théoriquement. Le moment critique où μ diminue de manière significative a pu alors être déterminé en temps réel quand la variation totale dépasse la "variation non-biologique".

Suivi en ligne de procédés de culture de cellules CHO par spectroscopies NIR et Raman

La concentration des cellules et des composés du milieu de culture peut être prédite en temps réel et avec une bonne précision par les spectroscopies NIR et Raman

Les spectroscopies NIR et Raman ont été utilisées pour le suivi en ligne de cultures de cellules CHO à l'aide de sondes stérilisables *in situ*. Différents modèles chimiométriques ont été développés pour prédire simultanément et en temps réel la concentration des cellules viables, du glucose, du lactate, de la glutamine, des ions ammonium et des IgG. Les spectres NIR et Raman acquis en ligne ont été calibrés à l'aide d'une régression PLS par des valeurs de références mesurées hors ligne. Avec les deux méthodes spectroscopiques, les modèles développés présentent une bonne capacité de prédiction, avec des valeurs d'erreur de calibration, de validation croisée, et de prédiction, inférieures ou égales à 15 % (sauf dans le cas des ions ammonium).

Les performances des spectroscopies NIR et Raman pour le suivi en ligne des procédés de culture peuvent être comparées par les facteurs de mérite (FOM) multivariés

Une comparaison en parallèle des performances des deux types de spectroscopie a été réalisée de manière plus exhaustive. En plus de l'exactitude (erreurs relatives de calibration, validation croisée et prédiction), d'autres FOM multivariés (précision, sensibilité analytique, sélectivité, ratio signal sur bruit, limite de détection, rapport de la déviation de performance, et rapport d'erreur de gamme) ont été également évalués. Aucune des deux spectroscopies n'a présenté de meilleurs résultats pour tous les FOM, ce qui indique la complexité du choix de la technologie la plus adaptée pour les procédés concernés. Cependant, la spectroscopie Raman surpasse la spectroscopie NIR pour les critères de sensibilité analytique, de sélectivité, et de limite de détection. Ainsi, la spectroscopie Raman démontre des avantages pour détecter des molécules de faible concentration dans un milieu de culture complexe.

Les stratégies de fusion de données permettent d'identifier des pistes de combinaison des données spectrales

Basées sur des principes fondamentaux différents, les spectroscopies Raman et NIR fournissent des informations complémentaires sur les molécules présentes dans un échantillon. Il peut être ainsi intéressant de combiner ces deux spectroscopies afin d'améliorer la performance de suivi des bioprocédés. Des stratégies de fusion des données ont été développées à trois niveaux (bas, intermédiaire et haut) pour prédire la concentration de glucose dans le milieu de culture. La fusion des données de niveau bas, plus simple à mettre en œuvre, ne permet pas une amélioration significative de la capacité de prédiction de l'une ou l'autre spectroscopie. La fusion de données de niveau haut, surtout utilisée pour des analyses de classification, ne montre pas d'amélioration significative dans le cas de notre étude essentiellement quantitative. Par contre, l'une des méthodes de fusion des données de niveau intermédiaire, combinant iPLS et PLS, permet l'amélioration la plus importante des erreurs relatives de calibration et de validation croisée.

Suivi en ligne par spectroscopie Raman (et NIR) de la glycosylation des IgG produites

L'analyse de référence hors ligne de l'état de glycosylation des IgG produites peut être réalisée par HPLC-MS après digestion par la trypsine

L'état de glycosylation est l'un des attributs de qualité les plus importants pour les anticorps thérapeutiques, puisqu'il conditionne leur innocuité et leur efficacité. Dans notre étude, les anticorps

(de type IgG) ont été analysés par HPLC-MS après digestion par la trypsine. Cette méthode permet de générer les données de référence hors ligne concernant, à la fois la macro-hétérogénéité (sites de glycosylation), et la micro-hétérogénéité (structures glycanniques) de la glycosylation des IgG. Les IgG produites sont hautement fucosylées avec un niveau de galactosylation relativement faible. En fin de culture, une augmentation non négligeable du niveau des formes aglycones (IgG non-glycosylées) est observée, indiquant une diminution significative de la qualité des IgG. Ces résultats montrent ainsi l'importance de disposer d'une méthode de suivi en temps réel de la glycosylation des anticorps au cours des procédés de culture de cellules animales.

La spectroscopie Raman permet le suivi en ligne de la macro- et de la micro-hétérogénéité de la glycosylation des IgG produites

Le potentiel de la spectroscopie Raman a été évalué pour le suivi en ligne, non seulement, de la quantité, mais aussi, de la qualité des IgG produites. Dans un premier temps, pour le suivi de la macro-hétérogénéité de glycosylation, deux modèles ont été développés avec une bonne précision pour prédire simultanément et en temps réel la concentration des IgG totales (IgG-T) et des IgG non-glycosylées (IgG-NG), au cours de cultures en mode discontinu et recharge-récolte. En outre, les analyses des scores VIP ont montré que la spectroscopie Raman est capable de distinguer des molécules très similaires (IgG-T et IgG-NG) dans un milieu de culture très complexe. Dans un second temps, concernant la micro-hétérogénéité de glycosylation, la spectroscopie Raman a permis de suivre en temps réel les quatre groupes des principales glycoformes (fucosylées, galactosylées, sialylées et hautement mannosylées) au cours des différentes cultures.

Contrôle de procédés de production d'anticorps par la culture de cellules CHO

L'estimation de μ en ligne permet d'améliorer la quantité et la qualité des IgG produites grâce au pilotage en temps réel de l'alimentation de cultures en mode recharge-récolte

L'approche développée pour déterminer en ligne le moment critique où μ diminue significativement peut être utilisée pour piloter en temps réel des cultures en mode recharge-récolte. A chaque moment critique détecté, 2/3 de volume du milieu de culture usagé était remplacé par du milieu frais pour maintenir la valeur de μ. Les performances, en termes de quantité et de glycosylation des IgG produites, ont été comparées à celles de cultures pilotées hors ligne, soit de manière anticipée, soit de façon tardive. Le pilotage en ligne a permis de produire plus d'IgG que la culture avec recharge-récolte anticipée, et d'obtenir une meilleure glycosylation des IgG que dans la culture de recharge

récolte tardive. Par ailleurs, cette approche d'estimation de µ en ligne offre également la possibilité de mieux définir le moment de récolte des produits en fin de culture afin d'éviter toute dégradation de la qualité de glycosylation des IgG récoltées.

Les vitesses spécifiques de consommation/production et les rendements métaboliques peuvent être suivis en temps réel par utilisation conjointe de plusieurs spectroscopies

Les vitesses spécifiques de consommation des substrats et de production des métabolites, ainsi que les rendements métaboliques sont aussi des paramètres cinétiques importants caractéristiques de l'état physiologique des cellules. Ils sont essentiels pour comprendre et contrôler les procédés de culture de cellules animales. Etant difficiles d'accès en temps réel, ils sont souvent considérés comme constants durant toute la culture. Ceci constitue l'une des sources principales d'imprécision pour le pilotage des cultures en mode semi-continu. A l'aide de l'utilisation conjointe de différentes techniques spectroscopiques (diélectrique, NIR/Raman), et des modèles de prédiction en temps réel développés, le suivi en ligne de l'évolution de différents paramètres cinétiques a pu être réalisé au cours d'une culture discontinue (vitesses spécifiques de consommation de glucose, de production de lactate, et de production d'IgG ; rendements de production de cellules par glucose consommé, de production de lactate par glucose consommé, et de production d'IgG par cellules produites). Ces suivis en ligne ont été validés en comparant les valeurs estimées en ligne à celles mesurées hors ligne.

Le contrôle automatique par capteur logiciel du débit d'alimentation d'une culture semi-continue améliore la productivité et la glycosylation des IgG

Classiquement, l'alimentation des cultures de cellules animales en mode semi-continu est réalisée avec un volume fixe de milieu concentré, en se basant sur une connaissance préalable du procédé. Cette stratégie ne permet cependant pas de répondre de manière adaptative aux besoins nutritifs des cellules, ce qui conduit souvent à une sous- ou sur-alimentation qui risque d'influencer la quantité et la qualité du produit final. C'est pourquoi, dans notre étude, une stratégie d'alimentation adaptative a été proposée, basée sur une meilleure connaissance de l'état physiologique des cellules, permettant ainsi de mieux répondre aux besoins cellulaires en éléments nutritifs. Un capteur logiciel de type boîte grise a été développé en intégrant des paramètres (μ, $Y_{X/Glc}$), prédits en ligne par spectroscopies diélectrique et NIR, dans les modèles basés sur les bilans de matières. Ce capteur logiciel permet d'analyser en temps réel l'état physiologique des cellules, et de contrôler la pompe d'alimentation pour répondre aux besoins des cellules en glucose. La trajectoire du débit d'alimentation, calculée par le capteur logiciel, est alors corrigée en temps réel en imposant une

valeur de consigne de 0,5 $g.L^{-1}$ (2,78 mM) pour la concentration de glucose résiduel, mesurée par spectroscopie NIR. Une première expérience utilisant cette stratégie d'alimentation adaptative a été effectuée, permettant le pilotage automatisé de la culture semi-continue sans aucun besoin de prélèvements ni d'interventions humaines. La performance de cette culture a été comparée à celles de cultures semi-continues pilotées manuellement, utilisant une stratégie d'alimentation ponctuelle ou exponentielle. En dépit d'une plus faible concentration maximale de cellules vivantes, cette stratégie de pilotage automatisé a abouti à une meilleure productivité spécifique en IgG tout en assurant une glycosylation quasiment complète des IgG produites.

Une liste des publications et communications issues de la thèse est présentée dans l'annexe 2.

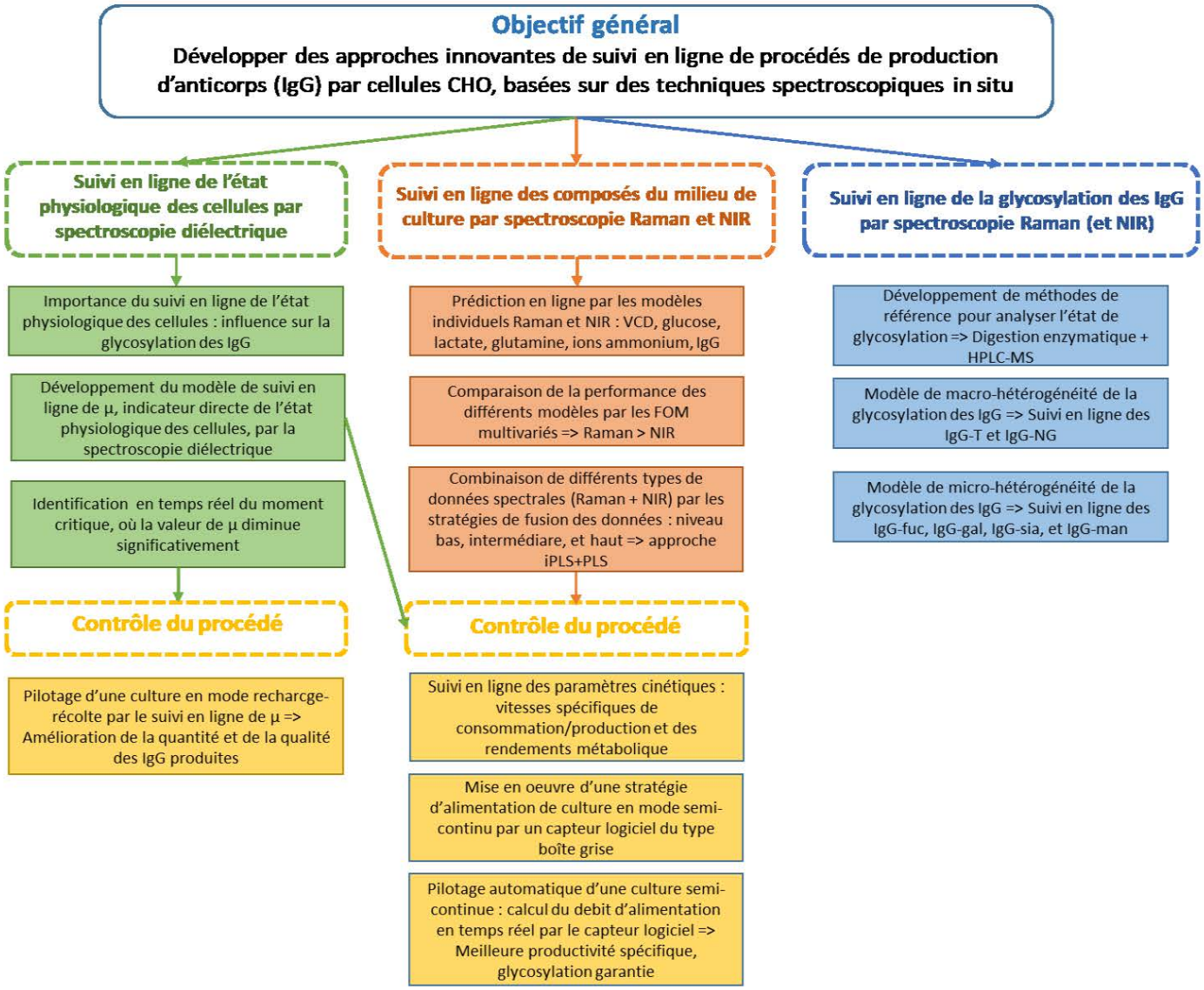

Objectif général
Développer des approches innovantes de suivi en ligne de procédés de production d'anticorps (IgG) par cellules CHO, basées sur des techniques spectroscopiques in situ
Suivi en ligne de l'état physiologique des cellules par spectroscopie diélectrique
Suivi en ligne des composés du milieu de culture par spectroscopie Raman et NIR
Suivi en ligne de la glycosylation des IgG par spectroscopie Raman (et NIR)
Importance du suivi en ligne de l'état physiologique des cellules : influence sur la glycosylation des IgG
Développement du modèle de suivi en ligne de µ, indicateur directe de l'état physiologique des cellules, par la spectroscopie diélectrique
Identification en temps réel du moment critique, où la valeur de µ diminue significativement
Contrôle du procédé
Pilotage d'une culture en mode recharge-récolte par le suivi en ligne de µ => Amélioration de la quantité et de la qualité des IgG produites
Prédiction en ligne par les modèles individuels Raman et NIR : VCD, glucose, lactate, glutamine, ions ammonium, IgG
Comparaison de la performance des différents modèles par les FOM multivariés => Raman > NIR
Combinaison de différents types de données spectrales (Raman + NIR) par les stratégies de fusion des données : niveau bas, intermédiare, et haut => approche iPLS+PLS
Contrôle du procédé
Suivi en ligne des paramètres cinétiques : vitesses spécifiques de consommation/production et des rendements métabolique
Mise en oeuvre d'une stratégie d'alimentation de culture en mode semi-continu par un capteur logiciel du type boîte grise
Pilotage automatique d'une culture semi-continue : calcul du debit d'alimentation en temps réel par le capteur logiciel => Meilleure productivité spécifique, glycosylation garantie
Développement de méthodes de référence pour analyser l'état de glycosylation => Digestion enzymatique + HPLC-MS
Modèle de macro-hétérogénéité de la glycosylation des IgG => Suivi en ligne des IgG-T et IgG-NG
Modèle de micro-hétérogénéité de la glycosylation des IgG => Suivi en ligne des IgG-fuc, IgG-gal, IgG-sia, et IgG-man

Figure 85 : Synthèse des résultats obtenus au cours de la thèse

Perspectives

Au cours de nos travaux, dans le cadre d'une démarche PAT, de nouvelles approches de suivi en ligne par diverses méthodes spectroscopiques ont été étudiées, afin de mieux comprendre et contrôler des procédés de culture de cellules CHO productrices d'anticorps recombinants de type IgG. Toutefois, les résultats obtenus suscitent encore beaucoup d'interrogations qui nécessiteraient des études complémentaires.

Il a été montré que la glycosylation des anticorps produits pouvait être influencée par le changement de l'état physiologique des cellules, correspondant à une diminution significative de la vitesse spécifique de croissance cellulaire. L'hypothèse est, qu'en cas de limitation en éléments nutritifs, la charge énergétique des cellules ne peut plus assurer la croissance cellulaire. A fortiori, les modifications post-traductionnelles des produits synthétisés, telles que la glycosylation, assez coûteuses du point de vue énergétique, seraient aussi affectées. Ainsi, il serait intéressant d'analyser plus précisément la charge énergétique des cellules tout au long du procédé de culture, et d'évaluer une éventuelle relation entre celle-ci et la glycosylation des anticorps. Pour aller plus loin, il pourrait être aussi très informatif de pouvoir suivre en temps réel l'évolution de cette charge énergétique à l'aide des techniques spectroscopiques NIR ou Raman.

La nouvelle approche d'estimation en ligne de la vitesse spécifique de croissance cellulaire proposée ici, utilisant la spectroscopie diélectrique, mériterait d'être optimisée davantage. En particulier, une évaluation plus complète des différents filtres et des paramètres de chaque filtre utilisé permettrait probablement d'augmenter à la fois la sensibilité et la stabilité du modèle. De plus, il pourrait être aussi très intéressant d'évaluer différents modèles non linéaires pour relier les mesures de permittivité à la concentration de cellules viables (*e.g.* équation Cole-Cole, régression multivariée, etc.), afin d'améliorer la prédiction de la VCD en particulier au cours de la phase de mort cellulaire. En outre, les informations sur les propriétés diélectriques des cellules (fréquence caractéristique (f_c), diminution de la permittivité ($\Delta\varepsilon$), paramètre de l'équation Cole-Cole (α)), pourraient être utilisées pour mieux comprendre l'état physiologique des cellules. Ceci permettrait éventuellement de mieux caractériser les différentes phases métaboliques des cellules. Par ailleurs, alors que l'intérêt de l'estimation en ligne de μ a été démontré dans le cas de cultures en mode recharge-récolte, il serait pertinent d'évaluer également cet intérêt du contrôle de la valeur de μ pour des cultures en mode semi-continu ou perfusé.

Des modèles chimiométriques ont été développés à partir des spectroscopies Raman et NIR pour prédire en temps réel la concentration de différents composés du milieu de culture. Pour améliorer la performance de prédiction de ces modèles, une optimisation des méthodes de prétraitement des spectres, et/ou une sélection de variables spécifiques à chaque molécule analysée (intervalle PLS, algorithme génétique, méthode de réconciliation etc.) seraient envisageables. De plus, il serait pertinent d'évaluer des méthodes de régression non-linéaires telles que la régression SVM (support vector machine) ou les réseaux de neurones artificiels (ANN). Par ailleurs, la capacité de prédiction et la robustesse de ces modèles devraient être renforcées grâce à l'obtention de nouvelles données expérimentales. Ces données devraient être générées avec des conditions opératoires diversifiées pour augmenter la variabilité de chaque molécule analysée, et des ajouts ponctuels (spiking) permettraient de limiter les corrélations internes entres les différentes molécules. Il serait aussi intéressant de suivre en ligne à l'aide des spectroscopies Raman et NIR d'autres molécules que celles étudiées au cours de ce travail, telles que les acides aminés, les vitamines, les lipides, etc. Enfin, il serait utile de valider la transférabilité de ces modèles à de nouvelles conditions opératoires : lignées cellulaires, compositions du milieu de culture, modes de culture, échelles de culture.

Dans notre étude, différentes stratégies de fusion des données ont été proposées pour combiner les différents spectres (NIR et Raman) afin d'améliorer la performance de prédiction des modèles. Néanmoins, une optimisation des paramètres de chaque méthode reste indispensable pour mieux évaluer le potentiel de ces stratégies. Par ailleurs, ces modèles fusionnés n'ont été testés que dans le cas du suivi de la concentration de glucose, dont la précision de prédiction était déjà très bonne avec les modèles individuels Raman et NIR. C'est pourquoi, il serait intéressant d'évaluer les performances de prédiction avec ces stratégies de fusion dans le cas d'autres molécules du milieu de culture, dont la concentration était moins bien prédite par les modèles individuels (glutamine, ions ammonium).

Des modèles de prédiction en ligne de la qualité des IgG produites ont été proposés en utilisant la spectroscopie Raman. En effet, la macro- et la micro-hétérogénéité de la glycosylation des IgG ont pu être suivies en temps réel au cours des cultures de cellules CHO. Toutefois, il serait pertinent de valider ces modèles avec des expériences complémentaires réalisées avec des conditions opératoires différentes (semi-continu, perfusé, ...), supplément par des éléments tels que le manganèse pour obtenir une plus grande variabilité des glycoformes, ou mettant en œuvre des ajouts ponctuels de molécules d'anticorps digérés et purifiés (IgG déglycosylées, défucosylées, dégalactosylées, désialylées, hautement mannosylées) dans le milieu au cours des cultures. Dans ce cas aussi, il serait intéressant d'améliorer les méthodes de prétraitement des spectres, de sélectionner les variables les plus pertinentes, et d'évaluer le potentiel des modèles de régressions non-linéaires. Dans ces

conditions, la performance de prédiction de la glycosylation des IgG par spectroscopie NIR pourrait également être améliorée. Ces améliorations permettraient alors d'envisager l'intégration de ces modèles dans une stratégie de rétrocontrôle du procédé en vue d'assurer une meilleure qualité des anticorps produits au cours du procédé. Par ailleurs, le potentiel de la spectroscopie Raman pourrait être évalué pour le suivi d'autres attributs de qualité des anticorps (agrégation, oxydation, fragmentation, glycation), ou pour la détection d'impuretés comme des protéines de cellules hôtes (HCP). S'il s'avère que la spectroscopie Raman permet de suivre ces paramètres de qualité, il serait alors utile d'implémenter cette technologie dans les procédés de séparation/purification (downstream processing) afin de contrôler plus efficacement les procédés de purification, ainsi que les étapes de polissage et de conditionnement.

Notre étude a aussi évalué la possibilité de suivre en temps réel certains des paramètres cinétiques comme les vitesses spécifiques de consommation/production et les rendements métaboliques, par utilisation conjointe de différentes techniques spectroscopiques (diélectrique + Raman/NIR). La précision de l'approche développée mériterait d'être améliorée par une optimisation des méthodes de calcul. De plus, lorsque la prédiction des modèles de glutamine et des ions ammonium aura été améliorée, d'autres paramètres cinétiques (vitesse spécifique de consommation de glutamine et de production des ions ammonium, rendement de croissance cellulaire par glutamine consommée, rendement de production des ions ammonium par glutamine consommée) pourraient également être suivis en ligne.

Enfin, nous avons développé une stratégie d'alimentation adaptative, basée sur l'utilisation d'un capteur logiciel de type boîte grise, pour piloter automatiquement l'apport de milieu de culture dans une culture réalisée en mode semi-continu, en vue de contrôler la concentration en glucose. Néanmoins, faute de temps, une seule culture a pu être réalisée, les résultats obtenus restant ainsi préliminaires. D'autres expériences de cultures semi-continues pilotées sont nécessaires pour optimiser les différents paramètres de calcul et mieux définir la valeur de consigne de la concentration de glucose. D'autre part, des améliorations seraient envisageables pour la partie en boîte blanche (modélisation macroscopique, analyse des flux métaboliques, modélisation par agent, etc.) afin de mieux décrire les comportements des cellules CHO. En ce qui concerne les prédictions issues des méthodes spectroscopiques en ligne (boîte noire), les propositions d'améliorations évoquées précédemment devraient augmenter les performances du capteur logiciel. Par ailleurs, la composition du milieu de culture d'alimentation est un élément prépondérant pour la maîtrise d'un procédé de culture semi-continue. Ce point devrait être analysé de manière plus systématique pour assurer simultanément une augmentation de la croissance cellulaire et de la productivité en

anticorps. Il pourrait aussi être pertinent d'évaluer d'autres stratégies de pilotage, en boucle ouverte ou fermée, basées sur le contrôle de paramètres différents tels que la vitesse spécifique de croissance cellulaire ou le rendement de lactate produit par glucose consommé. Dans l'idéal, un algorithme d'optimisation en temps réel pourrait être proposé pour maximiser la productivité en anticorps tout en maintenant leur qualité.

Ainsi, nos travaux ouvrent un large champ d'investigations pour le développement des nouvelles approches spectroscopiques dans le cadre du suivi en ligne des procédés de culture de cellules animales. Pour répondre à la récente stratégie PAT appliquée au cas des procédés de culture de cellules animales, il reste nécessaire de développer de nouvelles connaissances, à la fois biologiques et algorithmiques, afin de proposer de nouveaux procédés de production de protéines recombinantes par cellules animales qui soient mieux maîtrisés, plus automatisés, mieux contrôlés et ainsi plus optimisés.

Références Bibliographiques

Abès, R., and Teillaud, J.-L. (2010). Impact of Glycosylation on Effector Functions of Therapeutic IgG. Pharmaceuticals 3, 146–157.

Abu-Absi, N.R., Kenty, B.M., Cuellar, M.E., Borys, M.C., Sakhamuri, S., Strachan, D.J., Hausladen, M.C., and Li, Z.J. (2011). Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. Biotechnol. Bioeng. 108, 1215–1221.

Abu-Absi, N.R., Martel, R.P., Lanza, A.M., Clements, S.J., Borys, M.C., and Li, Z.J. (2014). Application of spectroscopic methods for monitoring of bioprocesses and the implications for the manufacture of biologics. Pharm Bioprocess 2, 267–84.

Adams, M.J. (2007). Chemometrics in Analytical Spectroscopy, Royal Society of Chemistry.

Aehle, M., Bork, K., Schaepe, S., Kuprijanov, A., Horstkorte, R., Simutis, R., and Lübbert, A. (2012). Increasing batch-to-batch reproducibility of CHO-cell cultures using a model predictive control approach. Cytotechnology 64, 623–634.

Aehle, M., Kuprijanov, A., Schaepe, S., Simutis, R., and Lübbert, A. (2011). Increasing batch-to-batch reproducibility of CHO cultures by robust open-loop control. Cytotechnology 63, 41–47.

Altamirano, C., Illanes, A., Becerra, S., Cairó, J.J., and Gòdia, F. (2006). Considerations on the lactate consumption by CHO cells in the presence of galactose. J. Biotechnol. 125, 547–556.

Andersen, D.C., Bridges, T., Gawlitzek, M., and Hoy, C. (2000). Multiple cell culture factors can affect the glycosylation of Asn-184 in CHO-produced tissue-type plasminogen activator. Biotechnol. Bioeng. 70, 25–31.

André, S., Lagresle, S., Da Sliva, A., Heimendinger, P., Hannas, Z., Calvosa, É., and Duponchel, L. (2017). Developing global regression models for metabolite concentration prediction regardless of cell line. Biotechnol. Bioeng. 114, 2550–2559.

André, S., Saint Cristau, L., Gaillard, S., Devos, O., Calvosa, É., and Duponchel, L. (2015). In-line and real-time prediction of recombinant antibody titer by in situ Raman spectroscopy. Anal. Chim. Acta 892, 148–152.

Ansorge Sven, Esteban Geoffrey, and Schmid Georg (2009). Multifrequency permittivity measurements enable on-line monitoring of changes in intracellular conductivity due to nutrient limitations during batch cultivations of CHO cells. Biotechnol. Prog. 26, 272–283.

Ansorge, S., Esteban, G., and Schmid, G. (2007). On-line monitoring of infected Sf-9 insect cell cultures by scanning permittivity measurements and comparison with off-line biovolume measurements. Cytotechnology 55, 115–124.

Ansorge, S., Esteban, G., and Schmid, G. (2010). On-line monitoring of responses to nutrient feed additions by multi-frequency permittivity measurements in fed-batch cultivations of CHO cells. Cytotechnology 62, 121–132.

Arnold, M.A., Small, G.W., Xiang, D., Qui, J., and Murhammer, D.W. (2004). Pure component selectivity analysis of multivariate calibration models from near-infrared spectra. Anal. Chem. 76, 2583–2590.

Arnold, S.A., Crowley, J., Woods, N., Harvey, L.M., and McNeil, B. (2003). In-situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. Biotechnol. Bioeng. 84, 13–19.

Arnold, S.A., Gaensakoo, R., Harvey, L.M., and McNeil, B. (2002). Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch Escherichia coli process. Biotechnol. Bioeng. 80, 405–413.

Arnolda, S.A., Crowley, J., Vaidyanathan, S., Matheson, L., Mohan, P., Hall, J.W., Harvey, L.M., and McNeil, B. (2000). At-line monitoring of a submerged filamentous bacterial cultivation using near-infrared spectroscopy. Enzyme Microb. Technol. 27, 691–697.

Arnoux, A.S., Preziosi-Belloy, L., Esteban, G., Teissier, P., and Ghommidh, C. (2005). Lactic Acid Bacteria Biomass Monitoring in Highly Conductive Media by Permittivity Measurements. Biotechnol. Lett. 27, 1551–1557.

Arora, M. (2016). Cell Culture Media: A Review. Mater. Methods.

Ashton, L., Hogwood, C.E.M., Tait, A.S., Kuligowski, J., Smales, C.M., Bracewell, D.G., Dickson, A.J., and Goodacre, R. (2015). UV resonance Raman spectroscopy: a process analytical tool for host cell DNA and RNA dynamics in mammalian cell lines. J. Chem. Technol. Biotechnol. 90, 237–243.

Ashton, L., Xu, Y., Brewster, V.L., Cowcher, D.P., Sellick, C.A., Dickson, A.J., Stephens, G.M., and Goodacre, R. (2013). The challenge of applying Raman spectroscopy to monitor recombinant antibody production. Analyst 138, 6977–6985.

Baker, K.N., Rendall, M.H., Patel, A., Boyd, P., Hoare, M., Freedman, R.B., and James, D.C. (2002). Rapid monitoring of recombinant protein products: a comparison of current technologies. Trends Biotechnol. 20, 149–156.

Beck, A., Wagner-Rousset, E., Ayoub, D., Van Dorsselaer, A., and Sanglier-Cianférani, S. (2013). Characterization of Therapeutic Antibodies and Related Products. Anal. Chem. 85, 715–736.

Bellon-Maurel, V., and McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. Soil Biol. Biochem. 43, 1398–1410.

Berkowitz, S.A., Engen, J.R., Mazzeo, J.R., and Jones, G.B. (2012). Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. Nat. Rev. Drug Discov. 11, 527.

Berry, B., Moretto, J., Matthews, T., Smelko, J., and Wiltberger, K. (2015). Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis. Biotechnol. Prog. 31, 566–577.

Berry, B.N., Dobrowsky, T.M., Timson, R.C., Kshirsagar, R., Ryll, T., and Wiltberger, K. (2016). Quick generation of Raman spectroscopy based in-process glucose control to influence biopharmaceutical protein product quality during mammalian cell culture. Biotechnol. Prog. 32, 224–234.

Bertrand, D. (2005). Étalonnage multidimensionnel : application aux données spectrales, Techniques de l'ingénieur.

Beutel, S., and Henkel, S. (2011). In situ sensor techniques in modern bioprocess monitoring. Appl. Microbiol. Biotechnol. 91, 1493–1505.

Bhatia, H., Mehdizadeh, H., Drapeau, D., and 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.

Biechele, P., Busse, C., Solle, D., Scheper, T., and Reardon, K. (2015). Sensor systems for bioprocess monitoring. Eng. Life Sci. 15, 469–488.

Birch, J.R., and Racher, A.J. (2006). Antibody production. Adv. Drug Deliv. Rev. 58, 671–685.

Bocklitz, T., Walter, A., Hartmann, K., Rösch, P., and Popp, J. (2011). How to pre-process Raman spectra for reliable and stable models? Anal. Chim. Acta 704, 47–56.

Boqué, R., Faber, N. (Klaas) M., and Rius, F.X. (2000). Detection limits in classical multivariate calibration models. Anal. Chim. Acta 423, 41–49.

Borys, M.C., Linzer, D.I., and Papoutsakis, E.T. (1993). Culture pH affects expression rates and glycosylation of recombinant mouse placental lactogen proteins by Chinese hamster ovary (CHO) cells. Biotechnol. Nat. Publ. Co. 11, 720–724.

Brewster, V.L., Ashton, L., and Goodacre, R. (2011). Monitoring the glycosylation status of proteins using Raman spectroscopy. Anal. Chem. 83, 6074–6081.

Brunner, D., Frank, J., Appl, H., Schöffl, H., Pfaller, W., and Gstraunthaler, G. (2010). Serum-free cell culture: the serum-free media interactive online database. ALTEX 27, 53–62.

Buckley, K., and Ryder, A.G. (2017). Applications of Raman Spectroscopy in Biopharmaceutical Manufacturing: A Short Review. Appl. Spectrosc. 71, 1085–1116.

Burnina, I., Hoyt, E., Lynaugh, H., Li, H., and Gong, B. (2013). A cost-effective plate-based sample preparation for antibody N-glycan analysis. J. Chromatogr. A 1307, 201–206.

Burns, D.A., and Ciurczak, E.W. (2007). Handbook of Near-Infrared Analysis, Third Edition, CRC Press.

Burton, D.R., and Woof, J.M. (1992). Human antibody effector function. Adv. Immunol. 51, 1–84.

Butler, M. (2005). Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. Appl. Microbiol. Biotechnol. 68, 283–291.

Butler, M. (2006). Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. Cytotechnology 50, 57.

Calvet, A., and Ryder, A.G. (2014). Monitoring cell culture media degradation using surface enhanced Raman scattering (SERS) spectroscopy. Anal. Chim. Acta 840, 58–67.

Camacho, J., Picó, J., and Ferrer, A. (2010). Data understanding with PCA: structural and variance Information plots. Chemom. Intell. Lab. Syst. 100, 48–56.

Cannizzaro, C., Gügerli, R., Marison, I., and von Stockar, U. (2003). On-line biomass monitoring of CHO perfusion culture with scanning dielectric spectroscopy. Biotechnol. Bioeng. 84, 597–610.

Capito, F., Skudas, R., Stanislawski, B., and Kolmar, H. (2013). Matrix effects during monitoring of antibody and host cell proteins using attenuated total reflection spectroscopy. Biotechnol. Prog. 29, 265–274.

Capito, F., Zimmer, A., and Skudas, R. (2015). Mid-infrared spectroscopy-based analysis of mammalian cell culture Parameters. Biotechnol. Prog. 31, 578–584.

Card, C., Hunsaker, B., Smith, T., and Hirsch, J. (2008). Near-infrared spectroscopy for rapid, simultaneous monitoring. BioProcess Int.

Carlesi, S., Ricci, M., Cucci, C., Lofrumento, C., Picollo, M., and Becucci, M. (2016). Multivariate analysis of combined reflectance FT-NIR and micro-Raman spectra on oil-paint models. Microchem. J. 124, 703–711.

Carvalhal, A.V., Santos, S.S., and Carrondo, M.J.T. (2011). Extracellular purine and pyrimidine catabolism in cell culture. Biotechnol. Prog. 27, 1373–1382.

Carvell, J.P., and Dowd, J.E. (2006). On-line measurements and control of viable cell density in cell culture manufacturing processes using radio-frequency impedance. Cytotechnology 50, 35–48.

Castro, P.M.L., Ison, A.P., Hayter, P.M., and Bull, A.T. (1995). The macroheterogeneity of recombinant human interferon-gamma produced by Chinese Hamster Ovary cells is affected by the protein and lipid-content of the culture medium. Biotechnol. Appl. Biochem. 21, 87–100.

Cavinato, A.G., Mayes, D.M., Ge, Z., and Callis, J.B. (1990). Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near-infrared spectroscopy. Anal. Chem. 62, 1977–1982.

Cerckel, I., Garcia, A., Degouys, V., Dubois, D., Fabry, L., and Miller, A.O. (1993). Dielectric spectroscopy of mammalian cells. 1. Evaluation of the biomass of HeLa- and CHO cells in suspension by low-frequency dielectric spectroscopy. Cytotechnology 13, 185–193.

Cervera, A.E., Petersen, N., Lantz, A.E., Larsen, A., and Gernaey, K.V. (2009). Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. Biotechnol. Prog. 25, 1561–1581.

Chi, Z., Chen, X.G., Holtz, J.S., and Asher, S.A. (1998). UV resonance Raman-selective amide vibrational enhancement: quantitative methodology for determining protein secondary structure. Biochemistry 37, 2854–2864.

Chong, I.-G., and Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. Chemom. Intell. Lab. Syst. 78, 103–112.

Chotigeat, W., Watanapokasin, Y., Mahler, S., and Gray, P.P. (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.

Chu, L., and Robinson, D.K. (2001). Industrial choices for protein production by large-scale cell culture. Curr. Opin. Biotechnol. 12, 180–187.

Chusainow, J., Yang, Y.S., Yeo, J.H., Toh, P.C., Asvadi, P., Wong, N.S., and Yap, M.G. (2009). A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? Biotechnol. Bioeng. 102, 1182–1196.

Clark, K.J.-R., Griffiths, J., Bailey, K.M., and Harcum, S.W. (2005). Gene-expression profiles for five key glycosylation genes for galactose-fed CHO cells expressing recombinant IL-4/13 cytokine trap. Biotechnol. Bioeng. 90, 568–577.

Claßen, J., Aupert, F., Reardon, K.F., Solle, D., and Scheper, T. (2017). Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application. Anal. Bioanal. Chem. 409, 651–666.

Clavaud, M., Roggo, Y., Von Daeniken, R., Liebler, A., and Schwabe, J.-O. (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.

Clementschitsch, F., and Bayer, K. (2006). Improvement of bioprocess monitoring: development of novel concepts. Microb. Cell Factories 5, 19.

Clincke, M.-F. (2010). Influence des conditions de culture sur la quantité de l'INF-γ recombinant produit par des cellules CHO au cours de procédés discontinus. Vandoeuvre-les-Nancy, INPL.

Cole, H.E., Demont, A., and Marison, I.W. (2015). The Application of Dielectric Spectroscopy and Biocalorimetry for the Monitoring of Biomass in Immobilized Mammalian Cell Cultures. Processes 3, 384–405.

Cole, K.S., and Cole, R.H. (1941). Dispersion and absorption in dielectrics I. Alternating current characteristics. J. Chem. Phys. 9, 341–351.

Collin, M., and Ehlers, M. (2013). The carbohydrate switch between pathogenic and immunosuppressive antigen-specific antibodies. Exp. Dermatol. 22, 511–514.

Consonni, V., Ballabio, D., and Todeschini, R. (2010). Evaluation of model predictive ability by external validation techniques. J. Chemom. 24, 194–201.

Courtès, F., Ebel, B., Guédon, E., and 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.

Cowcher, D.P., Deckert-Gaudig, T., Brewster, V.L., Ashton, L., Deckert, V., and Goodacre, R. (2016). Detection of protein glycosylation using tip-enhanced Raman scattering. Anal. Chem. 88, 2105–2112.

Craven, S., Whelan, J., and Glennon, B. (2014). Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. J. Process Control 24, 344–357.

Crowell, C.K., Grampp, G.E., Rogers, G.N., Miller, J., and Scheinman, R.I. (2007). Amino acid and manganese supplementation modulates the glycosylation state of erythropoietin in a CHO culture system. Biotechnol. Bioeng. 96, 538–549.

De Gelder, J., De Gussem, K., Vandenabeele, P., and Moens, L. (2007). Reference database of Raman spectra of biological molecules. J. Raman Spectrosc. 38, 1133–1147.

del Val, I.J., Kontoravdi, C., and Nagy, J.M. (2010). Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns. Biotechnol. Prog. 26, 1505–1527.

Dillon, T.M., Bondarenko, P.V., Rehder, D.S., Pipes, G.D., Kleemann, G.R., and Ricci, M.S. (2006). Optimization of a reversed-phase high-performance liquid chromatography/mass spectrometry method for characterizing recombinant antibody heterogeneity and stability. J. Chromatogr. A 1120, 112–120.

Doddridge, G.D., and Shi, Z. (2015). Multivariate figures of merit (FOM) investigation on the effect of instrument parameters on a Fourier transform-near infrared spectroscopy (FT-NIRS) based content uniformity method on core tablets. J. Pharm. Biomed. Anal. 102, 535–543.

Dotz, V., Haselberg, R., Shubhakar, A., Kozak, R.P., Falck, D., Rombouts, Y., Reusch, D., Somsen, G.W., Fernandes, D.L., and Wuhrer, M. (2015). Mass spectrometry for glycosylation analysis of biopharmaceuticals. TrAC Trends Anal. Chem. 73, 1–9.

Dowd, J.E., Jubb, A., Kwok, K.E., and Piret, J.M. (2003). Optimization and control of perfusion cultures using a viable cell probe and cell specific perfusion rates. Cytotechnology 42, 35–45.

Downey, B.J., Graham, L.J., Breit, J.F., and Glutting, N.K. (2014). A novel approach for using dielectric spectroscopy to predict viable cell volume (VCV) in early process development. Biotechnol. Prog. 30, 479–487.

Ducommun, P., Kadouri, A., von Stockar, U., and Marison, I.W. (2002). On-line determination of animal cell concentration in two industrial high-density culture processes by dielectric spectroscopy. Biotechnol. Bioeng. 77, 316–323.

Duval, D., Demangel, C., Miossec, S., and Geahel, I. (1992). Role of metabolic waste products in the control of cell proliferation and antibody production by mouse hybridoma cells. Hybridoma 11, 311–322.

Eagle, H. (1955). The specific amino acid requirements of a mammalian cell (strain L) in tissue culture. J. Biol. Chem. 214, 839–852.

Eagle, H. (1959). Amino acid metabolism in mammalian cell cultures. Science 130, 432–437.

Ecker, D.M., Jones, S.D., and Levine, H.L. (2015). The therapeutic monoclonal antibody market. In MAbs, Taylor & Francis.

Eon-Duval, A., Broly, H., and Gleixner, R. (2012). Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach. Biotechnol. Prog. 28, 608–622.

Esmonde-White, K.A., Cuellar, M., Uerpmann, C., Lenain, B., and Lewis, I.R. (2017). Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing. Anal. Bioanal. Chem. 409, 637–649.

Fallot, S. (2009). Elaboration de vecteurs polycistroniques à stoechiométrie variable et application à la production d'anticorps recombinants. Université de Toulouse, Université Toulouse III.

Fan, Y., Jimenez Del Val, I., Müller, C., Wagtberg Sen, J., Rasmussen, S.K., Kontoravdi, C., Weilguny, D., and Andersen, M.R. (2015). Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. Biotechnol. Bioeng. 112, 521–535.

Farges, B., Chenu, S., Marc, A., and Goergen, J.-L. (2008). Kinetics of IFN-γ producing CHO cells and other industrially relevant cell lines in rapeseed-supplemented batch cultures. Process Biochem. 43, 945–953.

Feron, O. (2009). Pyruvate into lactate and back: from the Warburg effect to symbiotic energy fuel exchange in cancer cells. Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol. 92, 329–333.

Fleischaker, R.J., and Sinskey, A.J. (1981). Oxygen demand and supply in cell culture. Eur. J. Appl. Microbiol. Biotechnol. 12, 193–197.

Flickinger, M.C. (2013). Upstream Industrial Biotechnology, 2 Volume Set , John Wiley & Sons.

Food, and Administration, D. (2004). Guidance for industry: PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance. DHHS Rockv. MD.

Franca, A.S., and Nollet, L.M. (2017). Spectroscopic Methods in Food Analysis, CRC Press.

Frenzel, A., Hust, M., and Schirrmann, T. (2013). Expression of recombinant antibodies. Front. Immunol. 4.

Gao, F.G., Jeevarajan, A.S., and Anderson, M.M. (2004). Long-term continuous monitoring of dissolved oxygen in cell culture medium for perfused bioreactors using optical oxygen sensors. Biotechnol. Bioeng. 86, 425–433.

Gautam, R., Vanga, S., Ariese, F., and Umapathy, S. (2015). Review of multidimensional data processing approaches for Raman and infrared spectroscopy. EPJ Tech. Instrum. 2, 8.

Gawlitzek, M., Ryll, T., Lofgren, J., and Sliwkowski, M.B. (2000). Ammonium alters N-glycan structures of recombinant TNFR-IgG: degradative versus biosynthetic mechanisms. Biotechnol. Bioeng. 68, 637–646.

Glacken, M.W., Fleischaker, R.J., and Sinskey, A.J. (1986). Reduction of waste product excretion via nutrient control: Possible strategies for maximizing product and cell yields on serum in cultures of mammalian cells. Biotechnol. Bioeng. 28, 1376–1389.

Goergen, J.-L., Marc, A., and Engasser, J.-M. (1997). 3.3 Kinetics and simulation of animal cell Processes. in mammalian cell biotechnology in protein production, Walter de Gruyter.

González-Vara Y R, A., Vaccari, G., Dosi, E., Trilli, A., Rossi, M., and Matteuzzi, D. (2000). Enhanced production of L-(+)-lactic acid in chemostat by *Lactobacillus casei DSM 20011* using ion-exchange resins and cross-flow filtration in a fully automated pilot plant controlled via NIR. Biotechnol. Bioeng. 67, 147–156.

Guan, Y., and Kemp, R.B. (1998). The Viable Cell Monitor: A dielectric spectroscope for growth and metabolic studies of animal cells on macroporous beads. in new developments and new applications in animal cell technology, Springer, Dordrecht, 321–328.

Guan, Y., Evans, P.M., and Kemp, R.B. (1998). Specific heat flow rate: An on-line monitor and potential control variable of specific metabolic rate in animal cell culture that combines microcalorimetry with dielectric spectroscopy. Biotechnol. Bioeng. 58, 464–477.

Guez, J.S., Cassar, J.P., Wartelle, F., Dhulster, P., and Suhr, H. (2004). Real time *in situ* microscopy for animal cell-concentration monitoring during high density culture in bioreactor. J. Biotechnol. 111, 335–343.

Guideline, I.H.T. (2005). Validation of analytical procedures: text and methodology Q2 (R1). In International Conference on Harmonization, Geneva, Switzerland, 11–12.

Hakemeyer, C., Strauss, U., Werz, S., Jose, G.E., Folque, F., and Menezes, J.C. (2012). At-line NIR spectroscopy as effective PAT monitoring technique in Mab cultivations during process development and manufacturing. Talanta 90, 12–21.

Hanson, M.A., Ge, X., Kostov, Y., Brorson, K.A., Moreira, A.R., and Rao, G. (2007). Comparisons of optical pH and dissolved oxygen sensors with traditional electrochemical probes during mammalian cell culture. Biotechnol. Bioeng. 97, 833–841.

Harris, C.M., Todd, R.W., Bungard, S.J., Lovitt, R.W., Morris, J.G., and Kell, D.B. (1987). Dielectric permittivity of microbial suspensions at radio frequencies: a novel method for the real-time estimation of microbial biomass. Enzyme Microb. Technol. 9, 181–186.

Harthun, S., Matischak, K., and Friedl, P. (1997). Process control of antithrombin III production by near-infrared spectroscopy. In Animal Cell Technology: from vaccines to genetic medicine, M.J.T. Carrondo, B. Griffiths, and J.L.P. Moreira, eds, 417–421.

Hassell, T., Gleave, S., and Butler, M. (1991). Growth inhibition in animal cell culture. Appl. Biochem. Biotechnol. 30, 29–41.

Hauser, H., and Wagner, R. (2014). Animal cell biotechnology: In biologics production, Walter de Gruyter.

Hayter, P.M., Curling, E., Gould, M.L., Baines, A.J., Jenkins, N., Salmon, I., Strange, P.G., and Bull, A.T. (1993). The effect of the dilution rate on CHO cell physiology and recombinant interferon-γ production in glucose-limited chemostat culture. Biotechnol. Bioeng. 42, 1077–1085.

Hayter, P.M., Curling, E.M.A., Baines, A.J., Jenkins, N., Salmon, I., Strange, P.G., and Bull, A.T. (1991). Chinese hamster ovary cell growth and interferon production kinetics in stirred batch culture. Appl. Microbiol. Biotechnol. 34, 559–564.

Henriques, J.G., Buziol, S., Stocker, E., Voogd, A., and Menezes, J.C. (2009). Monitoring mammalian cell cultivations for monoclonal antibody production using near-infrared spectroscopy. In Optical Sensor Systems in Biotechnology, Springer, 29–72.

Henry, O., Kamen, A., and Perrier, M. (2007). Monitoring the physiological state of mammalian cell perfusion processes by on-line estimation of intracellular fluxes. J. Process Control 17, 241–251.

Higel, F., Seidl, A., Sörgel, F., and Friess, W. (2016). N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. Eur. J. Pharm. Biopharm. 100, 94–100.

Hong, J.K., Lee, S.M., Kim, K.-Y., and Lee, G.M. (2014). Effect of sodium butyrate on the assembly, charge variants, and galactosylation of antibody produced in recombinant Chinese hamster ovary cells. Appl. Microbiol. Biotechnol. 98, 5417–5425.

Hooker, A.D., Goldman, M.H., Markham, N.H., James, D.C., Ison, A.P., Bull, A.T., Strange, P.G., Salmon, I., Baines, A.J., and Jenkins, N. (1995). N-glycans of recombinant human interferon-γ change during batch culture of chinese hamster ovary cells. Biotechnol. Bioeng. 48, 639–648.

Höpfner, T., Bluma, A., Rudolph, G., Lindner, P., and Scheper, T. (2010). A review of non-invasive optical-based image analysis systems for continuous bioprocess monitoring. Bioprocess Biosyst. Eng. 33, 247–256.

Hossler, P., Khattak, S.F., and Li, Z.J. (2009). Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology 19, 936–949.

Huang, Y.-M., Hu, W., Rustandi, E., Chang, K., Yusuf-Makagiansar, H., and Ryll, T. (2010). Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. Biotechnol. Prog. 26, 1400–1410.

Huhn, C., Selman, M.H.J., Ruhaak, L.R., Deelder, A.M., and Wuhrer, M. (2009). IgG glycosylation analysis. Proteomics 9, 882–913.

Imamura, T., Crespi, C.L., Thilly, W.G., and Brunengraber, H. (1982). Fructose as a carbohydrate source yields stable pH and redox parameters in microcarrier cell culture. Anal. Biochem. 124, 353–358.

Jain, E., and Kumar, A. (2008). Upstream processes in antibody production: Evaluation of critical parameters. Biotechnol. Adv. 26, 46–72.

James, W., Jr Jerome (Jerry) (2016). Concise handbook of analytical spectroscopy, the: theory, applications, and reference materials (In 5 Volumes), World Scientific.

Jang, J.D., and Barford, J.P. (2000). Effect of feed rate on growth rate and antibody production in the fed-batch culture of murine hybridoma cells. Cytotechnology 32, 229–242.

Jenkins, N., and Hovey, A. (1993). Temperature control of growth and productivity in mutant Chinese hamster ovary cells synthesizing a recombinant protein. Biotechnol. Bioeng. 42, 1029–1036.

Jenzsch, M., Bell, C., Buziol, S., Kepert, F., Wegele, H., and Hakemeyer, C. (2017). Trends in Process Analytical Technology: Present State in Bioprocessing. SpringerLink 1–42.

Junker, B.H., Reddy, J., Gbewonyo, K., and Greasham, R. (1994). On-line and *in-situ* monitoring technology for cell density measurement in microbial and animal cell cultures. Bioprocess Eng. 10, 195–207.

Justice, C., Brix, A., Freimark, D., Kraume, M., Pfromm, P., Eichenmueller, B., and Czermak, P. (2011). Process control in cell culture technology using dielectric spectroscopy. Biotechnol. Adv. 29, 391–401.

Kadlec, P., Gabrys, B., and Strandt, S. (2009). Data-driven soft sensors in the process industry. Comput. Chem. Eng. 33, 795–814.

Kapur, R., Della Valle, L., Sonneveld, M., Hipgrave Ederveen, A., Visser, R., Ligthart, P., Haas, M., Wuhrer, M., Schoot, C.E., and Vidarsson, G. (2014). Low anti-RhD IgG-Fc-fucosylation in pregnancy: a new variable predicting severity in haemolytic disease of the fetus and newborn. Br. J. Haematol. 166, 936–945.

Kayser, V., Chennamsetty, N., Voynov, V., Forrer, K., Helk, B., and Trout, B.L. (2011). Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. Biotechnol. J. 6, 38–44.

Kilburn, D.G., Fitzpatrick, P., Blake-Coleman, B.C., Clarke, D.J., and Griffiths, J.B. (1989). On-line monitoring of cell mass in mammalian cell cultures by acoustic densitometry. Biotechnol. Bioeng. 33, 1379–1384.

Kildegaard, H.F., Fan, Y., Sen, J.W., Larsen, B., and Andersen, M.R. (2016). Glycoprofiling effects of media additives on IgG produced by CHO cells in fed-batch bioreactors. Biotechnol. Bioeng. 113, 359–366.

Kim, J., Hwang, J., and Chung, H. (2008). Comparison of near-infrared and Raman spectroscopy for on-line monitoring of etchant solutions directly through a Teflon tube. Anal. Chim. Acta 629, 119–127.

Kirdar, A.O., Chen, G., Weidner, J., and Rathore, A.S. (2010). Application of near-infrared (NIR) spectroscopy for screening of raw materials used in the cell culture medium for the production of a recombinant therapeutic protein. Biotechnol. Prog. 26, 527–531.

Kiviharju, K., Salonen, K., Moilanen, U., and Eerikäinen, T. (2008). Biomass measurement online: the performance of in situ measurements and software sensors. J. Ind. Microbiol. Biotechnol. 35, 657–665.

Kochanowski, N., Blanchard, F., Cacan, R., Chirat, F., Guedon, E., Marc, A., and Goergen, J.-L. (2006). Intracellular nucleotide and nucleotide sugar contents of cultured CHO cells determined by a fast, sensitive, and high-resolution ion-pair RP-HPLC. Anal. Biochem. 348, 243–251.

Kochanowski, N., Blanchard, F., Cacan, R., Chirat, F., Guedon, E., Marc, A., and Goergen, J.-L. (2008). Influence of intracellular nucleotide and nucleotide sugar contents on recombinant interferon-γ glycosylation during batch and fed-batch cultures of CHO cells. Biotechnol. Bioeng. 100, 721–733.

Köhler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256, 495.

Konakovsky, V., Clemens, C., Müller, M.M., Bechmann, J., Berger, M., Schlatter, S., and Herwig, C. (2016). Metabolic control in mammalian fed-batch cell cultures for reduced lactic acid accumulation and improved process robustness. Bioengineering 3.

Konakovsky, V., Yagtu, A.C., Clemens, C., Müller, M.M., Berger, M., Schlatter, S., and Herwig, C. (2015). Universal capacitance model for real-time biomass in cell culture. Sensors 15, 22128–22150.

Kondo, M., Yamaoka, T., Honda, S., Miwa, Y., Katashima, R., Moritani, M., Yoshimoto, K., Hayashi, Y., and Itakura, M. (2000). The rate of cell growth is regulated by purine biosynthesis via ATP production and G1 to S phase transition. J. Biochem. 128, 57–64.

Konno, Y., Kobayashi, Y., Takahashi, K., Takahashi, E., Sakae, S., Wakitani, M., Yamano, K., Suzawa, T., Yano, K., Ohta, T., 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.

Kozma, B., Hirsch, E., Gergely, S., Párta, L., Pataki, H., and Salgó, A. (2017). On-line prediction of the glucose concentration of CHO cell cultivations by NIR and Raman spectroscopy: Comparative scalability test with a shake flask model system. J. Pharm. Biomed. Anal. 145, 346–355.

Krapp, S., Mimura, Y., Jefferis, R., Huber, R., and Sondermann, P. (2003). Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. J. Mol. Biol. 325, 979–989.

Kroll, P., Stelzer, I.V., and Herwig, C. (2017). Soft sensor for monitoring biomass subpopulations in mammalian cell culture processes. Biotechnol. Lett. 39, 1667–1673.

Kuligowski, J., Pérez-Guaita, D., and Quintás, G. (2016). Application of discriminant analysis and cross-validation on proteomics data. Methods Mol. Biol. Clifton NJ 1362, 175–184.

Kumpel, B.M. (2008). Lessons learnt from many years of experience using anti-D in humans for prevention of RhD immunization and haemolytic disease of the fetus and newborn. Clin. Exp. Immunol. 154, 1–5.

Kumpel, B.M., Rademacher, T.W., Rook, G.A., Williams, P.J., and Wilson, I.B. (1994). Galactosylation of human IgG monoclonal anti-D produced by EBV-transformed B-lymphoblastoid cell lines is dependent on culture method and affects Fc receptor-mediated functional activity. Hum. Antibodies Hybridomas 5, 143–151.

Kunkel, J.P., Jan, D.C., Jamieson, J.C., and Butler, M. (1998). Dissolved oxygen concentration in serum-free continuous culture affects N-linked glycosylation of a monoclonal antibody. J. Biotechnol. 62, 55–71.

Lagassé, H.A.D., Alexaki, A., Simhadri, V.L., Katagiri, N.H., Jankowski, W., Sauna, Z.E., and Kimchi-Sarfaty, C. (2017). Recent advances in (therapeutic protein) drug development. F1000Research 6.

Lamotte, D., Buckberry, L., Monaco, L., Soria, M., Jenkins, N., Engasser, J.M., and Marc, A. (1999). Na-butyrate increases the production and alpha2,6-sialylation of recombinant interferon-gamma expressed by alpha2,6- sialyltransferase engineered CHO cells. Cytotechnology 29, 55–64.

Lao, M.-S., and Toth, D. (1997). Effects of ammonium and lactate on growth and metabolism of a recombinant Chinese hamster ovary cell culture. Biotechnol. Prog. 13, 688–691.

Larkin, P. (2011). Infrared and Raman Spectroscopy: Principles and Spectral Interpretation, Elsevier.

Lasch, P. (2012). Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. Chemom. Intell. Lab. Syst. 117, 100–114.

Le, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., Karypis, G., and Hu, W.-S. (2012). Multivariate analysis of cell culture bioprocess data--lactate consumption as process indicator. J. Biotechnol. 162, 210–223.

Leatherbarrow, R.J., Rademacher, T.W., Dwek, R.A., Woof, J.M., Clark, A., Burton, D.R., Richardson, N., and Feinstein, A. (1985). Effector functions of a monoclonal aglycosylated mouse IgG2a: binding and activation of complement component C1 and interaction with human monocyte Fc receptor. Mol. Immunol. 22, 407–415.

Lee, H.W., Carvell, J., Brorson, K., and Yoon, S. (2015). Dielectric spectroscopy-based estimation of VCD in CHO cell culture. J. Chem. Technol. Biotechnol. 90, 273–282.

Lee, H.W., Christie, A., Liu, J.J., and Yoon, S. (2012). Estimation of raw material performance in mammalian cell culture using near infrared spectra combined with chemometrics approaches. Biotechnol. Prog. 28, 824–832.

Lewis, E.N., Qi, W., Kidder, L.H., Amin, S., Kenyon, S.M., and Blake, S. (2014). Combined dynamic light scattering and Raman spectroscopy approach for characterizing the aggregation of therapeutic proteins. Molecules 19, 20888–20905.

Lewis, I.R., and Edwards, H. (2001). Handbook of Raman spectroscopy: from the research laboratory to the process line, CRC Press.

Li, B., Ryan, P.W., Ray, B.H., Leister, K.J., Sirimuthu, N.M.S., and Ryder, A.G. (2010a). Rapid characterization and quality control of complex cell culture media solutions using raman spectroscopy and chemometrics. Biotechnol. Bioeng. 107, 290–301.

Li, F., Vijayasankaran, N., Shen, A. (Yijuan), Kiss, R., and Amanullah, A. (2010b). Cell culture processes for monoclonal antibody production. MAbs 2, 466–477.

Li, M., Ebel, B., Courtès, F., Guedon, E., and Marc, A. (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, 1065.

Li, M.-Y., Ebel, B., Paris, C., Chauchard, F., Guedon, E., and Marc, A. (2018). Real-time monitoring of antibody glycosylation site occupancy by in situ Raman spectroscopy during bioreactor CHO cell cultures. Biotechnol. Prog. 34, 486–493.

Lim, H.C., and Shin, H.S. (2013). Fed-batch cultures: principles and applications of semi-batch bioreactors. Cambridge University Press.

Lima e Castro, P.M. (1993). Optimisation of CHO cell growth and recombinant interferon-γ production. University of London.

Lin, A.A., Kimura, R., and Miller, W.M. (1993). Production of tPA in recombinant CHO cells under oxygen-limited conditions. Biotechnol. Bioeng. 42, 339–350.

Lingg, N., Zhang, P., Song, Z., and Bardor, M. (2012). The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. Biotechnol. J. 7, 1462–1472.

Liu, B., Spearman, M., Doering, J., Lattová, E., Perreault, H., and Butler, M. (2014). The availability of glucose to CHO cells affects the intracellular lipid-linked oligosaccharide distribution, site occupancy and the N-glycosylation profile of a monoclonal antibody. J. Biotechnol. 170, 17–27.

Liu, H.F., Ma, J., Winter, C., and Bayer, R. (2010). Recovery and purification process development for monoclonal antibody production. In MAbs, (Taylor & Francis), pp. 480–499.

Liu, Y.-J., André, S., Saint Cristau, L., Lagresle, S., Hannas, Z., Calvosa, É., Devos, O., and Duponchel, L. (2017). Multivariate statistical process control (MSPC) using Raman spectroscopy for in-line culture cell monitoring considering time-varying batches synchronized with correlation optimized warping (COW). Anal. Chim. Acta 952, 9–17.

Logan, D., and Carvell, J. (2011). A biomass monitor for disposable bioreactors. 9, 48–54.

López-Meza, J., Araíz-Hernández, D., Carrillo-Cocom, L.M., López-Pacheco, F., Rocha-Pizaña, M. del R., and Alvarez, M.M. (2016). Using simple models to describe the kinetics of growth, glucose consumption, and monoclonal antibody formation in naive and infliximab producer CHO cells. Cytotechnology 68, 1287–1300.

Lorber, A., Faber, K., and Kowalski, B.R. (1997). Net analyte signal calculation in multivariate calibration. Anal. Chem. 69, 1620–1626.

Lourenço, N.D., Lopes, J.A., Almeida, C.F., Sarraguça, M.C., and Pinheiro, H.M. (2012). Bioreactor monitoring with spectroscopy and chemometrics: a review. Anal. Bioanal. Chem. 404, 1211–1237.

Lu, F., Toh, P.C., Burnett, I., Li, F., Hudson, T., Amanullah, A., and Li, J. (2013). Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. Biotechnol. Bioeng. 110, 191–205.

Mandenius, C.-F., and Gustavsson, R. (2015). Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics. J. Chem. Technol. Biotechnol. 90, 215–227.

Mandenius, C.-F., Graumann, K., Schultz, T.W., Premstaller, A., Olsson, I.-M., Petiot, E., Clemens, C., and Welin, M. (2009). Quality-by-Design for biotechnology-related pharmaceuticals. Biotechnol. J. 4, 600–609.

Marc, A., and Olmos, E. (2010). Procédés de culture en masse de cellules animales. Tech. Ing. 6 800, 1–17.

Mariño, K., Bones, J., Kattla, J.J., and Rudd, P.M. (2010). A systematic approach to protein glycosylation analysis: a path through the maze. Nat. Chem. Biol. 6, 713.

Marison, I., Hennessy, S., Foley, R., Schuler, M., Sivaprakasam, S., and Freeland, B. (2012). The choice of suitable online analytical techniques and data processing for monitoring of bioprocesses. In Measurement, monitoring, modelling and control of bioprocesses, Springer, Berlin, Heidelberg, 249–280.

Markx, G.H., ten Hoopen, H.J., Meijer, J.J., and Vinke, K.L. (1991). Dielectric spectroscopy as a novel and convenient tool for the study of the shear sensitivity of plant cells in suspension culture. J. Biotechnol. 19, 145–157.

Martens, H., and Naes, T. (1992). Multivariate Calibration, John Wiley & Sons.

Matthews, T.E., Berry, B.N., Smelko, J., Moretto, J., Moore, B., and Wiltberger, K. (2016). Closed loop control of lactate concentration in mammalian cell culture by Raman spectroscopy leads to improved

cell density, viability, and biopharmaceutical protein production. Biotechnol. Bioeng. 113, 2416–2424.

McClure, W.F. (2003). 204 Years of near Infrared Technology: 1800–2003. J. Infrared Spectrosc. 11, 487–518.

McDonnell, S. (2015). Production of antibodies in hybridoma and non-hybridoma cell lines. In Animal Cell Culture, Springer, Cham, 65–88.

McShane, M.J., and Coté, G.L. (1998). Near-infrared spectroscopy for determination of glucose, lactate, and ammonia in cell culture media. Appl. Spectrosc. 52, 1073–1078.

Mehdizadeh, H., Lauri, D., Karry, K.M., Moshgbar, M., Procopio-Melino, R., and Drapeau, D. (2015). Generic Raman-based calibration models enabling real-time monitoring of cell culture bioreactors. Biotechnol. Prog. 31, 1004–1013.

Menditto, A., Patriarca, M., and Magnusson, B. (2007). Understanding the meaning of accuracy, trueness and precision. Accreditation Qual. Assur. 12, 45–47.

Mercier, S.M., Rouel, P.M., Lebrun, P., Diepenbroek, B., Wijffels, R.H., and Streefland, M. (2016). Process analytical technology tools for perfusion cell culture. Eng. Life Sci. 16, 25–35.

Mercille, S., Johnson, M., Lanthier, S., Kamen, A.A., and Massie, B. (2000). Understanding factors that limit the productivity of suspension-based perfusion cultures operated at high medium renewal rates. Biotechnol. Bioeng. 67, 435–450.

Meyers, R.A. (2012). Encyclopedia of Analytical Chemistry, Wiley.

Miller, J. N., Miller, J. C. (2000). Statistics and chemometrics for analytical chemistry, 4th. Essex, UK: Pearson Education Limited.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 29, 143.

Milligan, M., Lewin-Koh, N., Coleman, D., Arroyo, A., and Saucedo, V. (2014). Semisynthetic model calibration for monitoring glucose in mammalian cell culture with in situ near infrared spectroscopy. Biotechnol. Bioeng. 111, 896–903.

Mizrahi, A., and Lazar, A. (1988). Media for cultivation of animal cells: an overview. Cytotechnology 1, 199–214.

Morrison, C.J., McMaster, W.R., and Piret, J.M. (1997). Differential stability of proteolytically active and inactive recombinant metalloproteinase in Chinese hamster ovary cells. Biotechnol. Bioeng. 53, 594–600.

Nahrgang, S. (2002). Influence of cell-line and process conditions on the glycosylation of recombinant proteins.

Navrátil, M., Norberg, A., Lembrén, L., and Mandenius, C.-F. (2005). On-line multi-analyzer monitoring of biomass, glucose and acetate for growth rate control of a *Vibrio cholerae* fed-batch cultivation. J. Biotechnol. 115, 67–79.

Neermann, J., and Wagner, R. (1996). Comparative analysis of glucose and glutamine metabolism in transformed mammalian cell lines, insect and primary liver cells. J. Cell. Physiol. 166, 152–169.

Negrete, A., Esteban, G., and Kotin, R.M. (2007). Process optimization of large-scale production of recombinant adeno-associated vectors using dielectric spectroscopy. Appl. Microbiol. Biotechnol. 76, 761–772.

Neves, A.A., Pereira, D.A., Vieira, L.M., and Menezes, J.C. (2000). Real time monitoring biomass concentration in *Streptomyces clavuligerus* cultivations with industrial media using a capacitance probe. J. Biotechnol. 84, 45–52.

Noé, W., Bux, R., Werner, R.G., Berthold, W., and an der Riss, B. (1992). Aspects of continuous culture versus batch mode for animal cells in industrial processes. In Animal cell technology: developments, processes and products, Elsevier, 354–364.

Noll, T., and Biselli, M. (1998). Dielectric spectroscopy in the cultivation of suspended and immobilized hybridoma cells. J. Biotechnol. 63, 187–198.

Notingher, I. (2007). Raman spectroscopy cell-based biosensors. Sensors 7, 1343–1358.

Notingher, I., Verrier, S., Haque, S., Polak, J.M., and Hench, L.L. (2003). Spectroscopic study of human lung epithelial cells (A549) in culture: living cells versus dead cells. Biopolymers 72, 230–240.

Notingher, I., Verrier, S., Romanska, H., Bishop, A.E., Polak, J.M., and Hench, L.L. (2002). *In situ* characterisation of living cells by Raman spectroscopy. Spectrosc. Int. J. 16, 43–51.

November, E.J., and Impe, J.F.V. (2000). Evaluation of on-line viable biomass measurements during fermentations of *Candida utilis*. Bioprocess Eng. 23, 473–477.

Nyberg, G.B., Balcarcel, R.R., Follstad, B.D., Stephanopoulos, G., and Wang, D.I. (1999). Metabolic effects on recombinant interferon-gamma glycosylation in continuous culture of Chinese hamster ovary cells. Biotechnol. Bioeng. 62, 336–347.

Ohadi, K., Legge, R.L., and Budman, H.M. (2015). Development of a soft-sensor based on multi-wavelength fluorescence spectroscopy and a dynamic metabolic model for monitoring mammalian cell cultures. Biotechnol. Bioeng. 112, 197–208.

Opel, C.F., Li, J., and Amanullah, A. (2010). Quantitative modeling of viable cell density, cell size, intracellular conductivity, and membrane capacitance in batch and fed-batch CHO processes using dielectric spectroscopy. Biotechnol. Prog. 26, 1187–1199.

Øyaas, K. (2003). Animal Cell Culture, Physiochemical Effects of Osmolality and Temperature. In Encyclopedia of cell technology, John Wiley & Sons, Inc.

Ozturk, S., and Hu, W.-S. (2005). Cell culture technology for pharmaceutical and cell-based therapies, CRC Press.

Pacis, E., Yu, M., Autsen, J., Bayer, R., and Li, F. (2011). Effects of cell culture conditions on antibody N-linked glycosylation--what affects high mannose 5 glycoform. Biotechnol. Bioeng. 108, 2348–2358.

Page, M.J., and Sydenham, M.A. (1991). High level expression of the humanized monoclonal antibody Campath-1H in Chinese hamster ovary cells. Biotechnol. Nat. Publ. Co. 9, 64–68.

Pais, D.A.M., Carrondo, M.J., Alves, P.M., and Teixeira, A.P. (2014). Towards real-time monitoring of therapeutic protein quality in mammalian cell processes. Curr. Opin. Biotechnol. 30, 161–167.

Párta, L., Zalai, D., Borbély, S., and Putics, A. (2014). Application of dielectric spectroscopy for monitoring high cell density in monoclonal antibody producing CHO cell cultivations. Bioprocess Biosyst. Eng. 37, 311–323.

Patel, D., Guo, X., Ng, S., Melchior, M., Balderes, P., Burtrum, D., Persaud, K., Luna, X., Ludwig, D.L., and Kang, X. (2010). IgG isotype, glycosylation, and EGFR expression determine the induction of antibody-dependent cellular cytotoxicity in vitro by cetuximab. Hum. Antibodies 19, 89–99.

Petiot, E., Ansorge, S., Rosa-Calatrava, M., and Kamen, A. (2017). Critical phases of viral production processes monitored by capacitance. J. Biotechnol. 242, 19–29.

Petiot, E., Bernard-Moulin, P., Magadoux, T., Gény, C., Pinton, H., and Marc, A. (2010). *In situ* quantification of microcarrier animal cell cultures using near-infrared spectroscopy. Process Biochem. 45, 1427–1431.

Petiot, E., El-Wajgali, A., Esteban, G., Gény, C., Pinton, H., and Marc, A. (2012). Real-time monitoring of adherent Vero cell density and apoptosis in bioreactor processes. Cytotechnology 64, 429–441.

Planinc, A., Bones, J., Dejaegher, B., Van Antwerpen, P., and Delporte, C. (2016). Glycan characterization of biopharmaceuticals: Updates and perspectives. Anal. Chim. Acta 921, 13–27.

Porter, R.R. (1959). The hydrolysis of rabbit γ-globulin and antibodies with crystalline papain. Biochem. J. 73, 119–127.

Pörtner, R., Bohmann, A., Lüdemann, I., and Märkl, H. (1993). Metabolic parameters of a hybridoma cell line in batch and continuous cultivation. In Animal cell technology: basic & applied aspects, Springer, Dordrecht, 471–478.

Quality Guidelines : ICH.

Rademacher, T.W., Parekh, R.B., and Dwek, R.A. (1988). Glycobiology. Annu. Rev. Biochem. 57, 785–838.

Rangan, S., Kamal, S., Konorov, S.O., Schulze, H.G., Blades, M.W., Turner, R.F.B., and Piret, J.M. (2018). Types of cell death and apoptotic stages in Chinese Hamster Ovary cells distinguished by Raman spectroscopy. Biotechnol. Bioeng. 115, 401–412.

Rathore, A.S., and Singh, S.K. (2016). Production of protein therapeutics in the Quality by Design (QbD) paradigm. In Protein therapeutics, Springer, Cham, 41–67.

Rathore, A.S., and Winkle, H. (2009). Quality by design for biopharmaceuticals. Nat. Biotechnol. 27, 26–34.

Reichert, J.M. (2011). Antibody-based therapeutics to watch in 2011. MAbs 3, 76–99.

Reichert, J.M., and Valge-Archer, V.E. (2007). Development trends for monoclonal antibody cancer therapeutics. Nat. Rev. Drug Discov. 6, 349–356.

Reusch, D., Haberger, M., Falck, D., Peter, B., Maier, B., Gassner, J., Hook, M., Wagner, K., Bonnington, L., Bulau, P., et al. (2015). Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 2: Mass spectrometric methods. MAbs 7, 732–742.

Rhiel, M., Ducommun, P., Bolzonella, I., Marison, I., and Von Stockar, U. (2002). Real-time *in situ* monitoring of freely suspended and immobilized cell cultures based on mid-infrared spectroscopic measurements. Biotechnol. Bioeng. 77, 174–185.

Riley, M.R., Arnold, M.A., Murhammer, D.W., Walls, E.L., and DelaCruz, N. (1998). Adaptive Calibration Scheme for Quantification of Nutrients and Byproducts in Insect Cell Bioreactors by Near-Infrared Spectroscopy. Biotechnol. Prog. 14, 527–533.

Riley, M.R., Crider, H.M., Nite, M.E., Garcia, R.A., Woo, J., and Wegge, R.M. (2001). Simultaneous Measurement of 19 Components in Serum-Containing Animal Cell Culture Media by Fourier Transform Near-Infrared Spectroscopy. Biotechnol. Prog. 17, 376–378.

Riley, M.R., Rhiel, M., Zhou, X., Arnold, M.A., and Murhammer, D.W. (1997). Simultaneous measurement of glucose and glutamine in insect cell culture media by near infrared spectroscopy. Biotechnol. Bioeng. 55, 11–15.

Rinnan, Å., Berg, F. van den, and Engelsen, S.B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal. Chem. 28, 1201–1222.

Rita Costa, A., Elisa Rodrigues, M., Henriques, M., Azeredo, J., and Oliveira, R. (2010). Guidelines to cell engineering for monoclonal antibody production. Eur. J. Pharm. Biopharm. Off. J. Arbeitsgemeinschaft Pharm. Verfahrenstechnik EV 74, 127–138.

Roberts, C.A., Workman, J., and Reeves, J.B. (2004). Near-infrared spectroscopy in agriculture. American Society of Agronomy Madison.

Rodriguez, J., Spearman, M., Huzel, N., and Butler, M. (2005). Enhanced production of monomeric interferon-beta by CHO cells through the control of culture conditions. Biotechnol. Prog. 21, 22–30.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., and Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J. Pharm. Biomed. Anal. 44, 683–700.

Romero-Torres, S., Huang, J., and Hernandez, P. (2009). Practical considerations in PAT analyzer selection: NIR vs. Raman. Am Pharm Rev 12–19.

Rosa, F., Sales, K.C., Carmelo, J.G., Fernandes-Platzgummer, A., da Silva, C.L., Lopes, M.B., and Calado, C.R.C. (2016). Monitoring the ex-vivo expansion of human mesenchymal stem/stromal cells in xeno-free microcarrier-based reactor systems by MIR spectroscopy. Biotechnol. Prog. 32, 447–455.

Rourou, S., Gaumon, S., and Kallel, H. (2010). On-line monitoring of Vero cells cultures during the growth and rabies virus process using biomass spectrometer. In Cells and culture, Springer, Dordrecht, 829–832.

Rowland-Jones, R.C., van den Berg, F., Racher, A.J., Martin, E.B., and Jaques, C. (2017). Comparison of spectroscopy technologies for improved monitoring of cell culture processes in miniature bioreactors. Biotechnol. Prog. 33, 337–346.

Roychoudhury, P., Harvey, L.M., and McNeil, B. (2006). The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. Anal. Chim. Acta 571, 159–166.

Roychoudhury, P., O'Kennedy, R., McNeil, B., and Harvey, L.M. (2007). Multiplexing fibre optic near infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses. Anal. Chim. Acta 590, 110–117.

Sales, F., Callao, M.P., and Rius, F.X. (1997). Multivariate standardization techniques using UV-Vis data. Chemom. Intell. Lab. Syst. 38, 63–73.

Sandor, M., Rüdinger, F., Bienert, R., Grimm, C., Solle, D., and Scheper, T. (2013). Comparative study of non-invasive monitoring via infrared spectroscopy for mammalian cell cultivations. J. Biotechnol. 168, 636–645.

Sandor, M., Rüdinger, F., Solle, D., Bienert, R., Grimm, C., Groß, S., and Scheper, T. (2013). NIR-spectroscopy for bioprocess monitoring & control. BMC Proc. 7, P29.

Savitzky, A., and Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.

Schmid, G., Blanch, H.W., and Wilke, C.R. (1990). Hybridoma growth, metabolism, and product formation in HEPES-buffered medium: II. Effect of pH. Biotechnol. Lett. 12, 633–638.

Schneider, M., Marison, I.W., and von Stockar, U. (1996). The importance of ammonia in mammalian cell culture. J. Biotechnol. 46, 161–185.

Schwan, H.P. (1957). Electrical properties of tissue and cell suspensions. Adv. Biol. Med. Phys. 5, 147–209.

Sellick, C.A., Hansen, R., Jarvis, R.M., Maqsood, A.R., Stephens, G.M., Dickson, A.J., and Goodacre, R. (2010). Rapid monitoring of recombinant antibody production by mammalian cell cultures using Fourier transform infrared spectroscopy and chemometrics. Biotechnol. Bioeng. 106, 432–442.

Selman, M.H., McDonnell, L.A., Palmblad, M., Ruhaak, L.R., Deelder, A.M., and Wuhrer, M. (2010). Immunoglobulin G glycopeptide profiling by matrix-assisted laser desorption ionization Fourier transform ion cyclotron resonance mass spectrometry. Anal. Chem. 82, 1073–1081.

Shade, K.-T.C., and Anthony, R.M. (2013). Antibody glycosylation and inflammation. Antibodies 2, 392–414.

Shirsat, N.P., English, N.J., Glennon, B., and Al-Rubeai, M. (2015). Modelling of mammalian cell cultures. in animal cell culture, Springer, Cham, 259–326.

Short, S.M., Cogdill, R.P., and Anderson, C.A. (2007). Determination of figures of merit for near-infrared and raman spectrometry by net analyte signal analysis for a 4-component solid dosage system. AAPS PharmSciTech 8, 109–119.

Silman, R.W., Black, L.T., and Norris, K. (1983). Assay of solid–substrate fermentation by means of reflectance infrared analysis. Biotechnol. Bioeng. 25, 603–607.

Sinha, S., Pipes, G., Topp, E.M., Bondarenko, P.V., Treuheit, M.J., and Gadgil, H.S. (2008). Comparison of LC and LC/MS methods for quantifying N-glycosylation in recombinant IgGs. J. Am. Soc. Mass Spectrom. 19, 1643–1654.

Smelko, J.P., Ryll, T., and Müller, D. (2014). 7.1 Concepts and technologies for advanced process monitoring and control. In Animal cell biotechnology in biologics production, Walter de Gruyter.

St. Amand, M.M., Radhakrishnan, D., Robinson, A.S., and Ogunnaike, B.A. (2014). Identification of manipulated variables for a glycosylation control strategy. Biotechnol. Bioeng. 111, 1957–1970.

Stuart, B.H. (2004). Infrared spectroscopy: fundamentals and applications. Wiley.

Suzuki, E., and Ollis, D.F. (1990). Enhanced antibody production at slowed growth rates: experimental demonstration and a simple structured model. Biotechnol. Prog. 6, 231–236.

Swain, R.J., Jell, G., and Stevens, M.M. (2008). Non-invasive analysis of cell cycle dynamics in single living cells with Raman micro-spectroscopy. J. Cell. Biochem. 104, 1427–1438.

Teixeira, A., Cunha, A.E., Clemente, J.J., Moreira, J.L., Cruz, H.J., Alves, P.M., Carrondo, M.J.T., and Oliveira, R. (2005). Modelling and optimization of a recombinant BHK-21 cultivation process using hybrid grey-box systems. J. Biotechnol. 118, 290–303.

Teixeira, A.P., Oliveira, R., Alves, P.M., and Carrondo, M.J.T. (2009). Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative. Biotechnol. Adv. 27, 726–732.

Tharmalingam, T., Wu, C.-H., Callahan, S., and T. Goudar, C. (2015). A framework for real-time glycosylation monitoring (RT-GM) in mammalian cell culture. Biotechnol. Bioeng. 112, 1146–1154.

Tosi, S., Rossi, M., Tamburini, E., Vaccari, G., Amaretti, A., and Matteuzzi, D. (2003). Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes. Biotechnol. Prog. 19, 1816–1821.

Trummer, E., Fauland, K., Seidinger, S., Schriebl, K., Lattenmayer, C., Kunert, R., Vorauer-Uhl, K., Weik, R., Borth, N., Katinger, H., et al. (2006). Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors. Biotechnol. Bioeng. 94, 1033–1044.

Tuma, R. (2005). Raman spectroscopy of proteins: from peptides to large assemblies. J. Raman Spectrosc. 36, 307–319.

Urlaub, G., Käs, E., Carothers, A.M., and Chasin, L.A. (1983). Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. Cell 33, 405–412.

Vaidyanathan, S., Macaloney, G., Vaughan, J., McNeil, B., and Harvey, L.M. (1999). Monitoring of Submerged Bioprocesses. Crit. Rev. Biotechnol. 19, 277–316.

van Berkel, P.H.C., Gerritsen, J., Perdok, G., Valbjørn, J., Vink, T., van de Winkel, J.G.J., and Parren, P.W.H.I. (2009). N-linked glycosylation is an important parameter for optimal selection of cell lines producing biopharmaceutical human IgG. Biotechnol. Prog. 25, 244–251.

Vandeginste, B., and Massart, D.L. (1997). Handbook of chemometrics and qualimetrics, Elsevier.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the warburg effect: The metabolic requirements of cell proliferation. Science 324, 1029–1033.

Villacrés, C., Tayi, V.S., Lattová, E., Perreault, H., and Butler, M. (2015). Low glucose depletes glycan precursors, reduces site occupancy and galactosylation of a monoclonal antibody in CHO cell culture. Biotechnol. J. 10, 1051–1066.

Vojinović, V., Cabral, J.M.S., and Fonseca, L.P. (2006). Real-time bioprocess monitoring: Part I: *In situ* sensors. Sens. Actuators B Chem. 114, 1083–1091.

Wada, Y., Azadi, P., Costello, C.E., Dell, A., Dwek, R.A., Geyer, H., Geyer, R., Kakehi, K., Karlsson, N.G., Kato, K., et al. (2007). Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. Glycobiology 17, 411–422.

Wagner, K.W. (1914). The after effect in dielectrics. Arch. Für Elektrotechnik 2, 371–387.

Wahba, G. (1990). Spline models for observational data. Siam.

Walsh, G. (2014). Biopharmaceutical benchmarks 2014. Nat. Biotechnol. 32, 992–1000.

Wang, W., Wang, E.Q., and Balthasar, J.P. (2008). Monoclonal antibody pharmacokinetics and pharmacodynamics. Clin. Pharmacol. Ther. 84, 548–558.

Weikert, S., Papac, D., Briggs, J., Cowfer, D., Tom, S., Gawlitzek, M., Lofgren, J., Mehta, S., Chisholm, V., Modi, N., et al. (1999). Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. Nat. Biotechnol. 17, 1116–1121.

Wen, Z.-Q. (2007). Raman spectroscopy of protein pharmaceuticals. J. Pharm. Sci. 96, 2861–2878.

Wese\lucha-Birczyńska, A., Kozicki, M., Czepiel, J., and Birczyńska, M. (2013). Raman micro-spectroscopy tracing human lymphocyte activation. Analyst 138, 7157–7163.

Whelan, J., Craven, S., and Glennon, B. (2012). *In situ* Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. Biotechnol. Prog. 28, 1355–1362.

Wiercigroch, E., Szafraniec, E., Czamara, K., Pacia, M.Z., Majzner, K., Kochan, K., Kaczor, A., Baranska, M., and Malek, K. (2017). Raman and infrared spectroscopy of carbohydrates: A review. Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 185, 317–335.

Williams, P., and Norris, K.H. (1987). Near-infrared technology in the agricultural and food industries, American Association of Cereal Chemists.

Winkler, A., Berger, M., and Ehlers, M. (2013). Anti-rhesus D prophylaxis in pregnant women is based on sialylated IgG antibodies. F1000Research 2.

Wlaschin, K.F., and Hu, W.-S. (2006). Fedbatch culture and dynamic nutrient feeding. In Cell culture engineering, Springer, 43–74.

Wong, N.S.C., Wati, L., Nissom, P.M., Feng, H. t., Lee, M. m., and Yap, M.G.S. (2010). An investigation of intracellular glycosylation activities in CHO cells: Effects of nucleotide sugar precursor feeding. Biotechnol. Bioeng. 107, 321–336.

Xie, L., and Wang, D.I.C. (1996). Energy metabolism and ATP balance in animal cell cultivation using a stoichiometrically based reaction network. Biotechnol. Bioeng. 52, 591–601.

Xie, L., Nyberg, G., Gu, X., Li, H., Möllborn, F., and Wang, D.I. (1997). Gamma-interferon production and quality in stoichiometric fed-batch cultures of Chinese hamster ovary (CHO) cells under serum-free conditions. Biotechnol. Bioeng. 56, 577–582.

Xiong, Z.-Q., Guo, M.-J., Chu, J., Zhuang, Y.-P., and Zhang, S.-L. (2015). On-line specific growth rate control for improving reduced glutathione production in *Saccharomyces cerevisiae*. Biotechnol. Bioprocess Eng. 20, 887–893.

Xu, F., Rychnovsky, S.D., Belani, J.D., Hobbs, H.H., Cohen, J.C., and Rawson, R.B. (2005). Dual roles for cholesterol in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. 102, 14551–14556.

Yang, M., and Butler, M. (2000). Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation. Biotechnol. Bioeng. 68, 370–380.

Yardley, J.E., Kell, D.B., Barrett, J., and Davey, C.L. (2000). On-line, real-time measurements of cellular biomass using dielectric spectroscopy. Biotechnol. Genet. Eng. Rev. 17, 3–36.

Yoon, S.K., Choi, S.L., Song, J.Y., and Lee, G.M. (2005). Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37.0 degrees C. Biotechnol. Bioeng. 89, 345–356.

Zalai, D., Hevér, H., Lovász, K., Molnár, D., Wechselberger, P., Hofer, A., Párta, L., Putics, Á., and Herwig, C. (2016). A control strategy to investigate the relationship between specific productivity and high-mannose glycoforms in CHO cells. Appl. Microbiol. Biotechnol. 100, 7011–7024.

Zeiser, A., Elias, C.B., Voyer, R., Jardin, B., and Kamen, A.A. (2000). On-line monitoring of physiological parameters of insect cell cultures during the growth and infection process. Biotechnol. Prog. 16, 803–808.

Zhang, P., Woen, S., Wang, T., Liau, B., Zhao, S., Chen, C., Yang, Y., Song, Z., Wormald, M.R., Yu, C., et al. (2016). Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. Drug Discov. Today 21, 740–765.

Zhao, L., Fu, H.-Y., Zhou, W., and Hu, W.-S. (2015). Advances in process monitoring tools for cell culture bioprocesses. Eng. Life Sci. 15, 459–468.

Zheng, K., Bantog, C., and Bayer, R. (2011). The impact of glycosylation on monoclonal antibody conformation and stability. MAbs 3, 568–576.

Zhu, J. (2012). Mammalian cell protein expression for biopharmaceutical production. Biotechnol. Adv. 30, 1158–1170.

Zielke, H.R., Sumbilla, C.M., Sevdalian, D.A., Hawkins, R.L., and Ozand, P.T. (1980). Lactate: a major product of glutamine metabolism by human diploid fibroblasts. J. Cell. Physiol. 104, 433–441.

Annexe 1. Facteurs de mérite des modèles PLS pour la glycosylation des anticorps IgG

Tableau 18: Facteurs de mérites des modèles PLS développés à partir des spectres Raman ou NIR pour la prédiction de la glycosylation des anticorps IgG

Variable	Technique	F-test (p-value)	BCMSEP[1]	γ[2]	SEL[3]	S/N[4]	LOD[5]	RER[6]	RPD[7]
IgG-NG	Raman	<0.001	26.7	1.30	0.62	11.82	2.54	16.41	3.36
	NIR		271.4	0.64	0.07	5.71	24.16	5.81	1.19
IgG-fuc	Raman	<0.001	775.9	1.66	0.67	12.32	1.99	9.60	2.13
	NIR		1553.70	0.69	0.07	0.69	22.37	6.54	1.45
IgG-gal	Raman	<0.001	142.7	1.18	0.68	12.57	2.79	12.65	2.69
	NIR		363.4	0.53	0.07	5.59	29.46	7.68	1.63
IgG-sia	Raman	<0.001	2.6	22.73	0.80	14.91	0.15	11.71	2.63
	NIR		7.0	10.91	0.07	5.74	1.42	6.94	1.56
IgG-man	Raman	<0.001	15.7	5.17	0.67	12.48	0.64	9.39	2.26
	NIR		28.6	2.55	0.07	5.58	6.07	6.42	1.55

IgG-NG : IgG non-glycosylées ; IgG-fuc : IgG fucossylées ; IgG-gal : IgG galactosylées ; IgG-sia : IgG sialylées ; IgG-man : IgG fortement mannosylées.

[1] Précision

[2] Sensitivité analytique

[3] Sélectivité

[4] Rapport signal sur bruit

[5] Limite de détection

[6] Ratio de la gamme des erreurs

[7] Ratio de la déviation de performance

Annexe 2. Liste des publications et communications issues de la thèse

Ce travail de thèse a donné lieu à : 3 articles publiés dans des journaux scientifiques internationaux à comité de lecture, 2 manuscrits soumis pour publication, 1 acte de congrès, 4 communications orales et 3 communications affichées.

Articles publiés :

- Li M-Y., Ebel B., Chauchard F., Guedon E., Marc A. **2018**. 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, 207-213.
- Li M.-Y, Ebel B., Chauchard F., Guedon E., Marc A. **2018**. Real-time monitoring of antibody glycosylation site occupancy by *in-situ* Raman spectroscopy during bioreactor CHO cell cultures. *Biotechnology Progress*, 34, 486-493.
- Li M-Y., Ebel B., Courtès F., Guedon E., Marc A. **2016**. *In situ* infrared spectroscopy as a PAT tool of great promise for real-time monitoring of animal cell culture processes. *Journal of Analytical and Pharmaceutical Chemistry*, 3(2), 1065-1069.

Article soumis ou à soumettre :

- Li M-Y., Ebel B., Chauchard F., Guedon E., Marc, A. Real-time monitoring of different antibody glycoforms during cell culture processes by using *in situ* Raman spectroscopy. A soumettre à *Biotechnology Journal*.
- 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. Soumis à *Biotechnology and Bioengineering*.

Acte de congrès international

- Li M-Y., Ebel B., Fournier F., Guedon E., Marc A. **2017**. Online and real-time control of CHO cell specific growth rate throughout cultures in bioreactor. 25th ESACT meeting (European Society of Animal Cell Technology), Lausanne, May 14-17, 2017. *BMC Proceedings* 2017, 12 (Suppl 1):3-P134.

Communications orales dans des congrès internationaux :

- Li M-Y., Ebel B., Guedon E., Marc A. **2018**. *In situ* Raman Spectroscopy as a promising PAT tool for monitoring of macro and micro-heterogeneity of antibody glycosylation. 5th BioProScale Symposium: Innovative scale up and scale down for bioprocess intensification, Berlin, March 20-22, 2018.
- Chauchard F., Li M-Y., Ebel B., Marc A., Broecks W., Uyttersprot S., de Beer T., Brouckaert D. **2018**. Improving the use of Raman Spectroscopy for CHO cell culture using experiences and procedure from continuous chemistery processes. Annual meeting of International Foundation Process Analytical Chemistry (IFPAC) : Advancing the understanding and control of manufacturing processes, Washington, February 11-14, 2018.
- Li M.Y., Ebel B., Paris C., Chauchard F., Guedon E., Marc A. **2018**. Innovative methodological approach for online monitoring of animal cell culture processes for antibody production: contribution of in situ spectroscopic techniques to the PAT strategy. Conference at Novartis Pharma AG, Basel, 20 august 2018 (conférence invitée).
- Li M.Y., Ebel B., Paris C., Chauchard F., Guedon E., Marc A. **2018**. Real-time analysis of antibody glycosylation using in-line Raman spectroscopy in cell culture. 3rd Conference on Process Optimization and Safety in Industry 4.0, Paris, 25 septembre 2018 (conférence invitée).
-

Communications affichées dans des congrès internationaux ou nationaux :

- Zavala-Ortiz D. A., Ebel B., Li M-Y., Barradas-Dermitz D. M., Aguilar-Uscanga M. G., Marc A., Guedon E. **2018.** Interest of the Locally Weighted Regression to improve real-time monitoring of CHO and antibody quality by using *in situ* NIR spectroscopy. 12th European Symposium on Biochemical Engineering Sciences, Lisbonne, 9-12 septembre 2018.
- Li M-Y., Ebel B., Marc A., Chauchard F., Alexandris D. **2017**. Improving monitoring of cell culture using Raman spectroscopy on CHO cell culture. 4th European Conference on Process Analytics and Control Technology, Europact, Berlin, May 10-12, 2017.
- Li M-Y., Ebel B., Fournier F., Guedon E., Marc A. **2017**. Intérêt de l'estimation en temps réel de la vitesse spécifique de croissance pour le pilotage de culture de cellules animales. 16ème Congrès de la Société Française de Génie des Procédés (SFGP), Nancy, July 11-13, 2017

Résumé : Approche méthodologique innovante pour le suivi en ligne de procédés de production d'anticorps par cellules animales : apport des techniques spectroscopiques in situ à la stratégie PAT

Les bioprocédés industriels mettant en œuvre la culture de cellules animales sont devenus incontournables pour la production d'anticorps monoclonaux (AcM). Cependant, l'état physiologique des cellules et la qualité des AcM produits, en particulier leur glycosylation, sont tributaires des variations intervenant au cours du procédé. Il en découle des risques d'altération de l'efficacité et de la sûreté des AcM. C'est pourquoi, depuis quelques années, l'initiative Process Analytical Technology (PAT) encourage le développement du suivi en ligne de ces procédés, avec l'objectif de mieux les maîtriser et d'assurer la qualité finale des produits.

Dans ce contexte, cette thèse propose des approches innovantes pour le suivi en ligne de procédés de culture de cellules CHO (Chinese Hamster Ovary) en bioréacteur, basées sur l'utilisation de trois types de spectroscopies *in situ* (diélectrique, Raman, proche infrarouge(NIR)).

Le premier chapitre présente une nouvelle démarche permettant de prédire en temps réel l'état physiologique des cellules, au travers de la vitesse spécifique de croissance cellulaire (µ). A partir de la mesure en ligne de la permittivité grâce à la spectroscopie diélectrique, la µ a été calculée en temps réel, permettant de détecter le moment critique correspondant au moment où µ diminue. Comparée à une démarche hors ligne, l'utilisation de cette méthode pour le pilotage de cultures en mode recharge-récolte, permet d'assurer à la fois la quantité et la qualité de glycosylation des AcM.

Le second chapitre aborde l'utilisation des spectroscopies NIR et Raman *in situ* combinées à des méthodes chimiométriques. Les performances de ces deux spectroscopies ont été comparées en parallèle. Des modèles en ligne ont été développés pour prédire la concentration de différents paramètres (cellules viables, glucose, lactate, glutamine, ions ammonium, anticorps). L'évaluation de ces modèles par facteurs de mérite (FOM), a révélé certains avantages de la spectroscopie Raman. La combinaison de ces deux spectroscopies par diverses stratégies de fusion de données a été également évaluée.

Dans le troisième chapitre, l'intérêt de la spectroscopie Raman a été démontré pour le suivi en ligne, non seulement, de la concentration, mais aussi, de la glycosylation des AcM. Des modèles ont été développés pour la prédiction en ligne, à la fois, de la macro-hétérogénéité (sites de glycosylation), et de la micro-hétérogénéité (structures glycanniques) de la glycosylation des AcM dans le cas de cultures en mode discontinu et recharge-récolte.

Le dernier chapitre a utilisé les spectroscopies NIR et diélectrique, en les intégrant à un « capteur logiciel » combinant des équations de bilans de matière. Ce « capteur logiciel », mis en œuvre au cours d'une culture en mode semi-continu pour le contrôle automatique du débit d'alimentation, a conduit à une augmentation de la productivité du procédé ainsi qu'à une meilleure glycosylation des AcM produits.

Mots clés : Bioprocédés de culture de cellules animales, Suivi en temps réel, Spectroscopies *in situ*, Chimiométrie, Anticorps monoclonaux, Glycosylation

Abstract: Innovative methodological approach for online monitoring of animal cell culture processes for antibody production: contribution of *in situ* spectroscopic techniques to the PAT strategy

Bioprocesses of mammalian cell culture have become essential for the production of therapeutic recombinant proteins, such as monoclonal antibodies (mAb). However, the physiological state of the cells and the quality of the mAb produced, in particular their glycosylation, may vary during the process, and may lead to the alteration of the safety and efficacy of the final product. Consequently, the Process Analytical Technology (PAT) initiative has encouraged the development of online monitoring techniques, with the aim to better control the process and ensure the quality of the final product.

In this context, this thesis proposes innovative approaches for online monitoring of CHO (Chinese Hamster Ovary) cells bioreactor cultures, by using three types of *in situ* spectroscopic measurements (dielectric, Raman, near infrared (NIR)).

The first chapter presents a novel approach to predict in real-time one of the major cell physiological state parameters, the specific growth rate (μ). Based on online permittivity measured by *in situ* dielectric spectroscopy, the cell concentration was estimated and μ was calculated in real-time, making possible to detect the critical moment when μ begins to decrease significantly. Compared to an offline approach, this online approach allowed to maintain the cells in a stable physiological state, ensuring the glycosylation of the mAb produced in feed-harvest cultures.

The second chapter shows the use of in situ NIR and Raman spectroscopies combined with chemometric methods. For the first time, the performances of these two spectroscopies were compared in parallel in the same cultures. Online models were developed to predict in real-time the concentration of different parameters (viable cells, glucose, lactate, glutamine, ammonium ions and antibodies). The evaluation of these models by the multivariate Figures of Merit (FOM) revealed some of the advantages of Raman spectroscopy. The combination of the two spectroscopies by various data fusion strategies has also been evaluated.

In the third chapter, the interest of Raman spectroscopy for the online monitoring of both the quantity and the glycosylation of the mAb was demonstrated. Models were developed for online prediction of both macroheterogeneity (glycosylation site occupancy) and microheterogeneity (glycan structures) of mAb glycosylation in batch and feed-harvest cultures.

The last chapter used models previously developed for NIR and dielectric spectroscopies, to integrate into a “soft sensor” by combining with cell metabolic and mass balance equations. This “soft sensor”, implemented in a fed-batch cell culture for the automatic control of the feed rate, leads to an increased mAb productivity and better mAb glycosylation.

Key words: Animal cell culture bioprocesses, Real-time monitoring, In situ spectroscopies, Chemometrics, Monoclonal antibodies, Glycosylation