UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Chemometrics and Process Analytical Technology: Applications in Pharmaceutical and Biopharmaceutical Industries

PhD thesis • 2018

Francisca Folque Gouveia

4TUNEENGINEERING
DO IT BETTER

Title
Chemometrics and Process Analytical Technology:
Applications in Pharmaceutical and Biopharmaceutical Industries

Author
Maria Francisca Dias Folque de Gouveia

Date of Submission
27 November 2017

Date of Defense
23 March 2018

Supervisors
Assoc. Prof. Rasmus Bro (principal supervisor)
Spectroscopy and Chemometrics
Department of Food Science, Faculty of Science
University of Copenhagen

Jesper P Rahbek, PhD (co-supervisor)
Chemical Production Development, Chemical Production Denmark,
H. Lundbeck A/S

Pedro M Felizardo, PhD (co-supervisor)
Technical Operations Department
4Tune Engineering, Ltd.

Assessment Committee
Assoc. Prof. Thomas Skov
Spectroscopy and Chemometrics
Department of Food Science, Faculty of Science
University of Copenhagen, Denmark

Aux. Prof. Marco S Reis
Department of Chemical Engineering
University of Coimbra, Portugal

Principal Scientist Erik Skibsted, PhD
Novo Nordisk A/S, Denmark

ISBN 978-87-7209-109-9
Printed by SL grafik, Frederiksberg, Denmark (slgrafik.dk)

Preface

This dissertation is submitted as a requirement for obtaining the Ph.D. degree at the University of Copenhagen. The presented work was conducted at 4Tune Engineering Ltd., Lisbon, Portugal and at the Spectroscopy & Chemometrics group (SPECC), Department of Food Science, Faculty of Science, University of Copenhagen, in collaboration with the Process Optimization Laboratory, Chemical Production Development Department, H Lundbeck A/S and Roche Diagnostics GmbH – Pharmaceutical Biotech Production and Development, Penzberg, Germany. The PhD programme was supervised by Professor Rasmus Bro and funded by 4Tune Engineering, without any supplementary public contribution.

After working for several years at 4Tune Engineering Ltd., a company dedicated to industrial processes optimization, I felt the need to push myself to new heights and contribute to the current state of knowledge in the field of pharmaceutical sciences. Undergoing a doctorate program seemed the right move. At that point, I was far from imagining that I would gain so many things other than scientific knowledge. Unfortunately, this thesis does not summarize all achievements and lessons learnt, both professionally and at a personal level. Along the way, there were many contributions from inspiring people that made this work possible and to whom I shall forever remain grateful.

My sincere acknowledgments to my supervisor, Rasmus Bro and co-supervisors, Jesper Rahbek and Pedro Felizardo. Rasmus, thank you for your scientific insight, your enthusiasm that energizes everyone around, and for all wise suggestions and corrections. A very special word to Jesper for sharing his expertise and practical problem-solving attitude, and for making me feel so welcome to “far-distant” Lumsås. Pedro, thank you for bringing things into broader perspective, for your unconditional support and for sharing all the ups and downs of working in a Ph.D.

I am deeply indebted to H. Lundbeck A/S for supporting the experimental work, particularly to Michael Mealy for setting up the contracts and agreements and providing all working conditions. I would like to thank all team members of POL, specially Asmus for all support at the lab and joyful moments (the photo session is for sure one memory to keep!).

To all my colleagues at 4Tune Engineering, a big thank you for creating a unique working atmosphere. Specially, I wish to thank, Tiago, João, Henrique, Marcos, Maria, Luísa, Ricardo, Sofia, Ana, Sara, Sílvia, Madalena and Catarina for all support, friendship and teamwork. I am utterly grateful to Prof. José Cardoso Menezes for sponsoring and encouraging this project from the very beginning. His many years of industrial and scientific experience combined with his forward-looking vision inspire all who have the privilege of working with him.

Among the key people that contributed to this work, a special thanks to the project leaders I had the privilege to work at Roche Diagnostics GmbH, particularly Dr. Ulrike Strauss, Dr. Stefan Buziol and Dr. Marco Strohmeier. I am truly appreciated for all thoughtful discussions and for sharing your insider's perspective on the industry needs and challenges.

Finally, I would like to thank my friends and family. I will always be grateful to my dear parents, brothers and sisters for their unconditional love and support. A special word to my father and mother in-law for being positive and helpful. To my son Miguel, my "partner in crime", a very special thank you! You literally grew up with this project and showed me how to move forward so many times.

I dedicate this thesis to my husband Jorge for putting up with me even when work and this project turned me into a ghost. Thanks for believing in us!

Abstract

The concept of Process Analytical Technology (PAT) was introduced in pharmaceutical and biopharmaceutical manufacturing over 10 years ago. Yet, the conventional manufacturing paradigm focused on *quality by testing* (QbT) is still deep-rooted in the quality culture of pharmaceutical companies and, has considerably slowed down the introduction of new technologies in manufacturing. This thesis specifically concerns this issue by focusing on the development of PAT applications combined with chemometrics to *design*, *analyse*, *control* and *optimize* pharmaceutical processes. The challenges underpinning the lack of data-driven decision making in (bio)pharmaceutical manufacturing are discussed in three different perspectives:

- Gathering data and extracting information from PAT to build process understanding, fasten product development or introduce improvements in existing processes;
- Setting appropriate workflows for analytical method development and lifecycle management of PAT procedures in drug manufacturing, including knowledge transfer and knowledge management in a global environment.
- Establishing knowledge and data-driven approaches where all process unit operations are linked and the product lifecycle perspective is considered under an evolving regulatory framework.

The emergence of a data-driven mind set and the advancements in data analytics and computer science are an opportunity for pharmaceutical companies to gain novel insights to improve drug development and manufacturing efficiency. **PAPER I** describes how PAT combined with different chemometric approaches can be used to support the conversion of a conventional batch process for API manufacturing into a continuous one. The work describes a roadmap for PAT screening and details the scientific basis to develop monitoring and control strategies for continuous reactions based on PAT information. **Poster I** and **Poster II** further detail on the experimental work required to characterize the reaction system investigated in the study (i.e., reactants and products).

In a very competitive landscape for the launch of biopharmaceutical products, it is crucial to expedite development timelines while maximizing the efficiency of process characterization studies. **PAPER II** conveys how 2D fluorescence and advanced data analysis methods can shed some light into biologic drug process development. Due to limited sensor capabilities and/or first-principles understanding, 2D fluorescence is a promising technology for accelerating bioprocesses development and evaluating control strategies.

When developing a new process or addressing a problem in an existing one, it is fundamental to adopt consistent procedures and practices, particularly in a globalized environment. Perhaps one of the major constraints delaying the adoption of PAT in (bio)pharmaceutical manufacturing is the inexistence of a systematic workflow for PAT method development and maintenance in routine use. **BOOK CHAPTER I** aims to address this gap by proposing a systematic procedure for PAT-based methods development and lifecycle management, aligned with current regulatory expectations, and applicable to the production of any (bio)pharmaceutical product.

Many of the existing solutions to deal with multivariate data within PAT and more general within Quality by Design (QbD), tend to focus on *local* data analysis: data is generated and analysed in the context of a microsystem (viz., sample or unit operation). However, the true benefits of innovative process technologies or advanced data analysis methods can only be realized if the knowledge is properly transferred and maintained throughout the process flowsheet (i.e., linkage of steps) and over the product lifecycle, including development and commercial manufacturing. **BOOK CHAPTER II** sets continued process verification (CPV) at the center of the process/product lifecycle approach. To fully accomplish the CPV concept, the process must perform at any point of its existence as well and as consistently as it did when filed and approved in the first place. A workflow to streamline the information hidden in complex databases is provided to elevate *legacy* product validations to a higher level, in terms of compliance with current regulations/guidelines, robustness and operational performance. Under the framework of pharmaceutical quality systems (cf. ICH Q10, 2009), that incorporate quality risk management (QRM) and data-based justifications to develop a good grasp of all important variability sources, different stakeholders can challenge process owners to make evidence of product quality and consistency which is a true indicator of in-depth process understanding and efficient knowledge dissemination.

Resumé

Process Analytical Technology (PAT) blev introduceret i den farmaceutiske og biofarmaceutiske industri for mere end ti år siden. MDden traditionelle kvalitetstankegang: ”quality by testing” er dog stadig meget udbredt i de farmaceutiske firmaer, hvilket har dæmpet udbredelsen af PAT teknologier. Denne afhandling omhandler denne problematik ved at fokusere på udvikling af PAT applikationer kombineret med kemometri til design, analyse, kontrol og optimering i den farmaceutiske produktion. Udfordringen i at få større gavn af databaserede beslutninger i den (bio)farmaceutiske produktion vil blive diskuteret ud fra tre forskellige perspektiver:

- Opsamling af data og ekstrahering af relevant information fra PAT teknologier til bedre procesforståelse, hurtigere produktudvikling eller forbedringer af eksisterende processer;
- Opstilling af workflows til analytisk metodeudvikling af lifecycle management af PAT procedurer i den farmaceutiske produktion inklusiv vidensdeling og -management i et globalt miljø;
- Etablering af en videns- og data-baseret tilgang hvor alle processens enhedsoperationer er sammenkædet under hensyntagen til produktets lifecycle perspektiv i et dynamisk regulatorisk setup.

Den øgede fokus på en data-baseret tankegang i kombination med den teknologiske udvikling i dataanalyse og computer science giver nye muligheder for de farmaceutiske firmaer for at opnå indsigt til forbedring af produktudvikling og produktionseffektiviseringer. **PAPER I** beskriver hvordan PAT teknologier i kombination med forskellige kemometriske metoder kan anvendes til at understøtte transitionen af en proces fra batch til kontinuert produktion. Yderligere beskrives et roadmap til screening af PAT teknologier og nødvendig videnskabelig viden til udvikling af monitorerings- og kontrol-strategier i kontinuerte processer baseret på PAT teknologier. **Poster I og II** beskriver i yderligere detaljer det eksperimentelle arbejde, der er nødvendigt for at karakterisere det undersøgte kemiske system beskrevet i **PAPER I** (dvs. reaktanter og produkter).

Det er vigtigt at optimere udviklingstiden samtidig med at effektiviteten i udviklingsstudierne maksimeres. **PAPER II** beskriver hvordan 2D fluorescens-spektroskopi og avancerede data analytiske metoder kan bidrage med nye viden i procesudvikling af biologiske farmaceutiske produkter. Fluorescens giver nogle unikke muligheder for at accelerere udviklingen og styringen af bioprocesser.

Under udvikling af nye processer, eller under troubleshooting af en eksisterende, er det vigtigt at efterleve konsistente procedurer og bedste praksis - specielt i et globalt miljø. En af grundene til den forsinkede udbredelse af PAT teknologier i den (bio)farmaceutiske industri er måske netop den manglende beskrivelse af systematiske arbejdsgange i forbindelse med metodeudvikling og vedligehold af PAT applikationer under rutinebrug. **BOOK CHAPTER I** beskriver en løsning på dette, ved at foreslå en systematik for udvikling og lifecycle management af PAT applikationer som er tilpasset til de aktuelle regulatoriske forventninger, og som er anvendeligt under fremstilling af (bio)farmaceutiske produkter generelt.

Mange eksisterende multivariate dataanalytiske løsninger inden for PAT, og mere generelt inden for Quality by Design (QbD) fokuserer på lokal dataanalyse, Det vil sige at data er genereret og analyseret inden for et mikrosystem (f.eks. prøve- eller enhedsoperation). Men for at få fuldt udbytte af fordelene ved innovative procesanalytiske teknologier eller avanceret dataanalyse, så bør sammenkædning af viden tilsikres, både mellem forskellige procestrin og mellem forskellige livs-cykler af produktet. **BOOK CHAPTER II** sætter "continued process verification" (CPV) i centrum i en sådan proces/produkt livscyklus tilgang. CPV konceptet kan kun gennemføres hvis processen er valid til enhver tid og præcist som den var valid, da den blev registeret første gang. Her gives et workflow til at strømline den information, der er gemt i databaser på kommercielle produkter. Dette vil kunne løfte produktvalidering til et højere niveau – både når det gælder regulativ compliance, robusthed og operationel performance. Inden for rammerne af de farmaceutiske kvalitetssystemer (ICH Q10, 2009) anvendes "quality risk management" (QRM) og databaserede redskaber til at forstå variationskilder. Og her kan forskellige interessenter udfordre procesejere til at demonstrere produktkvalitet og robusthed - begge som gode indikatorer for dybdegående procesforståelse og effektiv vidensformidling.

List of Publications

Paper I

Gouveia F.F., Rahbek J.P., Mortensen A.R., Pederson M.T., Felizardo P.M., Bro R., Mealy M.J. (2016). Using PAT to accelerate the transition to continuous API manufacturing. *Analytical Bioanaytical Chemistry*, 409(3), 821-832.doi 10.1007/s00216-016-9834-z.

Paper II

Gouveia F.F., Felizardo P., Menezes J.C., Jung C., Buziol S. (2017). 2D-Fluorescence Spectroscopy as a Process Analytical Technology (PAT) Tool to Support CHO Cultivation Process Monitoring, Optimization & Troubleshooting Activities (submitted to *Biochemical Engineering Journal*).

Book Chapter I

Gouveia F. F. & Felizardo P. M. (2017). Lifecycle Management of PAT Procedures: Applications to Batch and Continuous Processes. Ed. Ferreira, Menezes, Tobyn. In Multivariate Analysis in the Pharmaceutical Industry, *Elsevier* (*in press*).

Book Chapter II

Strohmeier M., Pradines C., **Gouveia F.F**., Menezes J.C. (2017). Knowledge-based Product and Process Lifecycle Management for Legacy Products. Ed. Calnan, Kane, Lipa, Menezes, in Lifecycle Approach to Knowledge Excellence in the Pharmaceutical Industry, *Taylor & Francis*.

Poster I

Gouveia F.F., Rahbek J.P., Mortensen A.R., Felizardo P.M., Bro R. (2015). Moving from Batch to Continuous Flow Production with PAT. International Conference on Near Infrared Spectroscopy (NIR2015), Foz do Iguassu, Brazil.

Poster II

Gouveia F.F., Rahbek J.P., Mortensen A.R., Pederson M.T., Felizardo P.M., Bro R. (2016). In-Depth Understanding of an API Synthesis: a Combined PAT and Chemometric Approach. XVI Chemometrics in Analytical Chemistry (CAC-2016), Barcelona, Spain.

Additional publications by the author

Jose GE, **Folque F**, Menezes JC, Werz S, Strauss U, Hakemeyer C (2011). Predicting mAb product yields from cultivation media components, using NIR and 2D-Fluorescence spectroscopies. *Biotechnol. Prog*. 27(2), 1339–1346.

Hakemeyer C, Strauss U, Werz S, Jose GE, **Folque F**, Menezes JC (2012). At-line NIR spectroscopy as effective PAT monitoring technique in mAb cultivations during process development and manufacturing. *Talanta* 90, 12–21.

Felizardo P, **Folque, F**, Machado JE, Menezes JC (2012). Process analytical technology: a common approach across different industries. *NIR News* 23(7),10-15.

Hakemeyer C., Strauss U., Werz S., Jose G.E., **Folque F**., Menezes J.C. (2013). Near-infrared and two-dimensional fluorescence spectroscopy monitoring of monoclonal antibody fermentation media quality: aged media decreases cell growth. *Wiley Biotechnol. J*. 8(7), 835-846.

Menezes J., **Gouveia, F**., Felizardo, P. (2014). Science-based data-driven technologies in Pharma: development and manufacturing with QbD 2.0. *European Pharmaceutical Review* 19(3) 50-55.

Tirado S.S, **Gouveia F.F**., Menezes J.C. (2015). PAT Paves the Way for Continuous Manufacturing. *Pharmaceutical T*. 39 (18): 33-34.

List of Abbreviations

API: active pharmaceutical ingredient
ASTM: Amarican Society for Testing and Materials
BHK: Baby Hamster Kidney
CAPA: corrective and preventive action
CHO: Chinese Hamster Ovary
COGs: cost of goods sold
CMC: chemistry, manufacturing and control
CpK: process capability index
CPP: critical process parameter
CPs: continuous processes
CPV: Continued Process Verification
CQA: critical quality attribute
DoE: design of experiments
DQ: design qualification
EFA: evolving factor analysis
EMA: European Medicines Agency
EQ: equivalence point
FAD: flavin adenine dinucleotide
FDA: Food and Drug Administration
FMEA: failure mode and effect analysis
FT-IR: Fourier transformed mid-infrared
GFP: Green Fluorescent Protein
HPLC: high performance liquid chromatography
ICH: International Conference on Harmonization
IPC: in-process control
iPLS: iteractive partial least squares
IQ: installation qualification
KPI: key performance indicator
LCL: lower control limit
LDH: lactate dehydrogenase
LV: latent variable
MCR-ALS: multivariate curve resolution-alternating least squares
MCR: multivariate curve resolution
MDVA: multivariate data analysis
NADPH: Nicotinamide Adenine Dinucleotide Phosphate Hydrogen
NIR: near-infrared
NIRS: near-infrared spectroscopy
NOR: normal operating ranges
OOS: out-of-specification
OOT: out-of-trend
OP: operational qualification
OPV: Ongoing Process Verification
PACA: preventive and corrective action
PAR: proven acceptable ranges
PARAFAC: parallel factor analysis
PAT: process analytical technology
PC: principal component
PCA: principal components analysis
PLS: partial least squares or projections to latent structures
PP: process parameter
PpK: process performance index
PQ: performance qualification
PQS: Pharmaceutical Quality System
PSE: process systems engineering
QA: quality attribute
QbD: *quality by design*
QbT: Quality by Testing
QTPP: quality target product profile
RA: risk assessment
RMSEC: root mean square error of calibration
RMSECV: root mean square error of cross validation
RMSEP: root mean square error of prediction
SME: subject matter expert
SNV: standard normal variate
SPC: statistical process control
SQT: signal quality test
TQM: Total Quality Management
UCL: upper control limit
VCD: viable cell density
VIP: variable importance projection

Table of Contents

PREFACE ..... **III**

ABSTRACT ..... **V**

RESUMÉ ..... **VII**

LIST OF PUBLICATIONS ..... **IX**

LIST OF ABBREVIATIONS ..... **XI**

TABLE OF CONTENTS ..... **I**

PART I – QUALITY BY DESIGN, PAT AND CHEMOMETRICS **1**

1 INTRODUCTION **3**

1.1 MOTIVATION 3

1.2 SCOPE 4

1.3 THESIS OUTLINE 4

2 THE QUALITY BY DESIGN CONCEPT: AN INTEGRATED APPROACH TO PHARMACEUTICAL MANUFACTURING **7**

2.1 QBD FOUNDATIONS: OVERCOMING THE QUALITY MYOPIA 7

2.2 THE CRITICAL PATH FOR PHARMACEUTICAL INNOVATION: REQUIREMENTS AND ENABLERS 9

3 BRIDGING PAC AND PAT DOMAINS IN PHARMACEUTICAL MANUFACTURING **11**

3.1 UNDERSTANDING THE "5 WHY'S" 11

3.2 EVOLUTION OF THE REGULATORY FRAMEWORK FOR PHARMACEUTICAL PRODUCTS 12

PART II - PAT IN PHARMACEUTICAL AND BIOPHARMACEUTICAL PRODUCTION **14**

4 USING PAT TO ACCELERATE THE TRANSITION TO CONTINUOUS MANUFACTURING **16**

4.1 INTRODUCTION 17

4.2 MATERIALS AND METHODS 18

4.2.1 Chemistry *18*

4.2.2 Technology screening roadmap *19*

4.2.3 Instrumentation *21*

4.2.4 Experimental procedure and setup *21*
4.2.4.1 Laboratory scale: batch reactions 21
4.2.4.2 Full-scale: continuous mode 23
4.2.5 Multivariate data analysis *24*
4.3 RESULTS AND DISCUSSION 25
4.3.1 Qualitative analysis *25*
4.3.2 Predictive models *28*
4.3.3 NIR implementation in full-scale continuous processing *31*
4.4 CONCLUSIONS 33
5 IN-DEPTH UNDERSTANDING OF AN API SYNTHESIS: A COMBINED PAT AND CHEMOMETRIC APPROACH 34
5.1 PURPOSE 35
5.2 MATERIAL AND METHODS 35
5.3 RESULTS AND DISCUSSION 36
5.4 CONCLUDING REMARKS 38
6 FLUORESCENCE SPECTROSCOPY AS A PROCESS ANALYTICAL TECHNOLOGY (PAT) TOOL TO SUPPORT CHO CULTIVATION PROCESS MONITORING, OPTIMIZATION & TROUBLESHOOTING ACTIVITIES 40
6.1 INTRODUCTION 41
6.2 MATERIAL AND METHODS 43
6.2.1 Cell culture *43*
6.2.2 Reference analysis *43*
6.2.3 Fluorescence spectroscopy *44*
6.2.4 Data analysis *44*
6.3 RESULTS AND DISCUSSION 45
6.3.1 Fluorometric properties of CHO cell cultivations *45*
6.3.2 Batch-to-batch variability based on fluorescence profiles: investigation of optimization and fault diagnostics potential *47*
6.3.3 Inline monitoring of CHO cell cultivations *53*
6.3.3.1 Product concentration (PRO) 53
6.3.3.2 Lactate dehydrogenase (LDH) 54
6.3.3.3 Viable Cell Density (VCD) 55

6.3.3.4 Lactate concentration (LAC) 56
6.3.3.5 Glucose concentration (GLU) 56
6.3.4 Improvement Opportunities Based on 2D-fluorescence spectroscopy 57
6.3.5 CONCLUSIONS 59

PART III - SCALING-UP KNOWLEDGE MANAGEMENT IN THE PHARMACEUTICAL INDUSTRY .. 60

7 LIFECYCLE MANAGEMENT OF PAT PROCEDURES: BATCH AND CONTINUOUS PROCESS APPLICATIONS 62

7.1 INTRODUCTION 63
7.2 A THREE STAGE APPROACH TO PAT PROCEDURES DEVELOPMENT AND LIFECYCLE MANAGEMENT 65
7.2.1 Stage 1 – PAT Procedure Design 66
7.2.2 Stage 2 – PAT Performance Qualification 70
7.2.3 Stage 3 – PAT Ongoing Performance Verification (OPV) 72
7.3 ONGOING PROCESS VERIFICATION (OPV) OF PAT PROCEDURES: EXAMPLES FROM BATCH AND CONTINUOUS PROCESSES 77
7.4 CONCLUSIONS 81

8 KNOWLEDGE-BASED PRODUCT AND PROCESS LIFECYCLE MANAGEMENT FOR LEGACY PRODUCTS 84

8.1 INTRODUCTION 85
8.2 EVALUATION OF THE VALIDATION STATUS OF A LEGACY PHARMACEUTICAL PRODUCT 87
8.2.1 Criticality analysis of the manufacturing process 89
8.2.2 Risk assessment and filtering 89
8.2.3 Gap analysis of the current control strategy 92
8.2.4 Revised control strategy: opportunities for improvement 92
8.2.5 Continued process verification and lifecycle management 93
8.3 CONCLUSIONS 94
8.4 ANNEX: A STREAMLINED-WORKFLOW FOR LIFECYCLE MANAGEMENT OF BIOPHARMACEUTICAL LEGACY PRODUCTS 95

9 CONCLUSIONS AND PERSPECTIVES 106

10 REFERENCES 108

11 ANNEXES 118

PART I – Quality by Design, PAT and Chemometrics

"Quality can be planned and most quality crises and problems relate to the way in which quality was planned in the first place."

Joseph Juran

1 Introduction

1.1 Motivation

The Process Analytical Technology (PAT) initiative launched by the Food and Drug Administration (FDA) and the guidelines issued by the International Conference on Harmonization (ICH) set a shift in the definition of *quality* applied to pharmaceutical products. A comprehensive approach to improve the understanding of drug product attributes and the corresponding relationship with process unit operations has been recommended since. More specifically, ICH Q8(R2) states that "Product and process understanding in combination with quality risk management will support the control of the process such that the variability can be compensated for in an *adaptable* manner to deliver consistent product quality" (ICH, 2009). In order to became *adaptable*, process analytical technologies are envisioned as true enablers of innovation and continuous improvement initiatives.

The goal of this academic contribution is to demonstrate the value of PAT and multivariate data analysis to the (bio)pharmaceutical industry and its potential to improve the practical knowledge and general understanding of manufacturing processes. This study was driven by the expectation that these approaches can overcome specific challenges often encountered in (bio)pharmaceutical production, such as poor first-principles model understanding, management of complex data structures, high degree of correlation between process variables and unit operations interdependency. Furthermore, the aim of this project is to propose a structured workflow for effective lifecycle management of PAT-based procedures, particularly important when these methods take part of manufacturing control strategy. It is our expectation to provide additional guidance to companies on how to demonstrate the "fit-for-purpose" of PAT-based procedures throughout the manufacturing lifecycle.

To date, very few companies have been successful in performing low-cost improvements to processes developed without a QbD rationale. Here, we demonstrate how data-driven methods can be used to improve the understanding of the existing relationships between process parameters and product quality attributes, supporting a Continuous Process Verification (CPV) strategy towards continuous improvement of existing/legacy pharmaceutical products.

"Data do not yield information except with the intervention of the mind. Information does not yield meaning except with the intervention of imagination."

Theodore Levitt

1.2 Scope

This thesis aims to demonstrate the potential of PAT and multivariate data analysis to characterize complex processes for chemical and biological drug substance production. The value of these approaches is demonstrated for a diverse set of analytical tools and measuring principles. Furthermore, the importance of chemometrics to uncover the relationships between spectral information, process dynamics and product quality are highlighted. In our perspective, extending the lifecycle management concept to analytical method development (particularly to PAT) is a key initiative to increase the prevalence of PAT as integrating elements of the process control strategy. Therefore, one of the most important outcomes of the present work is the development of a structured workflow to effectively manage and disseminate the information and knowledge generated during PAT development. With the same purpose, a workflow focused on knowledge assessments and data-driven approaches is presented with the aim to identify root-causes for process variations in already established (bio)processes and, consequently, improvement opportunities.

1.3 Thesis outline

The dissertation is organized in three parts. In the first part, the foundations and the evolving perspective of the *quality* concept are presented along with the building blocks (i.e., enablers) necessary to realize the *Quality by Design* vision. The second and the third parts detail the research work. First, the practical aspects of PAT applications are discussed, followed by a generalisation of such approaches into a conceptual workflow for effective knowledge management and transfer. Each chapter is based either on a published/submitted paper to a peer reviewed journal or a poster presentation in a scientific conference. For consistency purposes, minor changes were made to the original documents, both in content and format. These changes are meant to guide the reader through the focal points of this dissertation but do not change the key messages disclosed in the original publications. In the case of **Book Chapter II,** the introductory section describing the regulatory landscape for pharmaceutical development and manufacturing was moved to **Chapter 3**. The emphasis was given to the practical application of the concepts to improve a commercial biopharmaceutical product well established in the market. A graphical representation of the thesis structure is presented in Figure 1-1 and will be further used to draw the reader's attention to the key subjects discussed within each chapter.

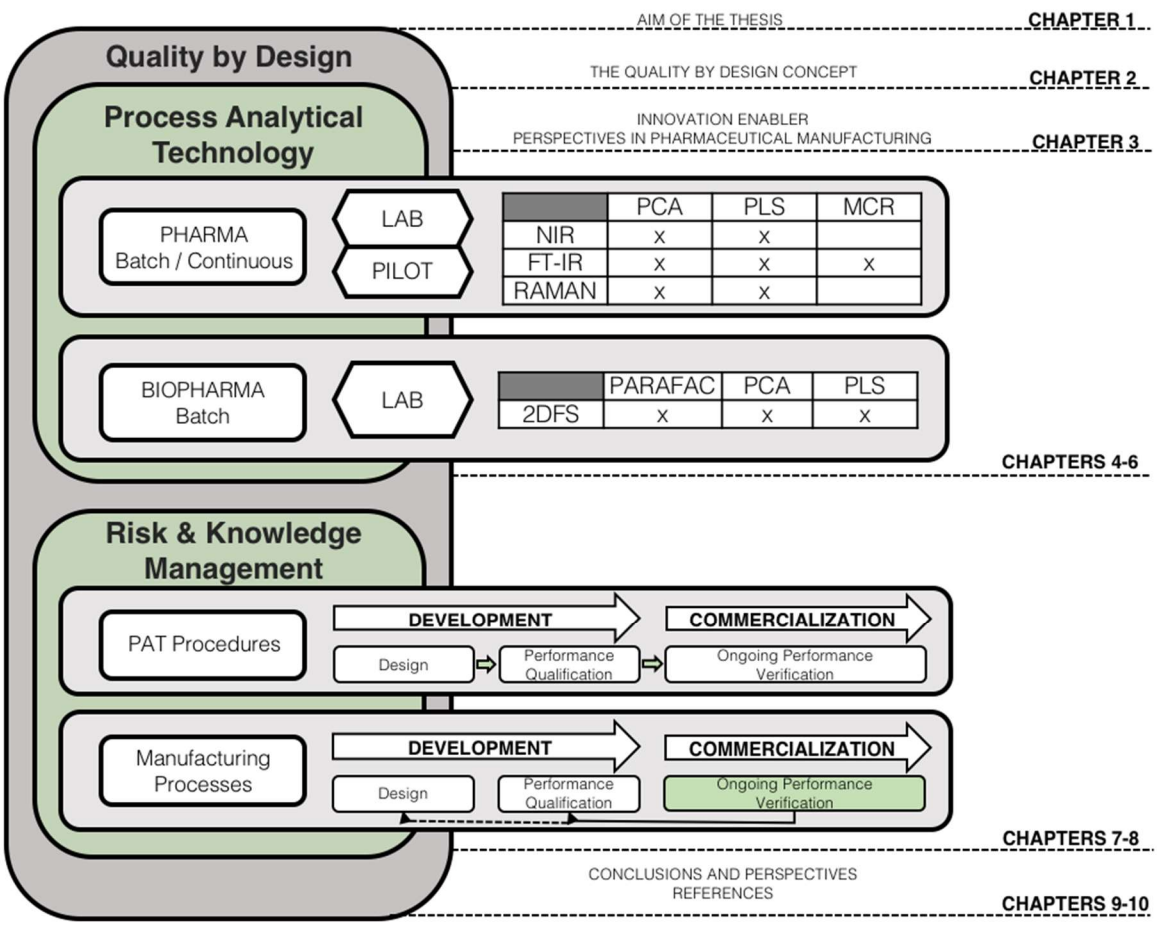


Figure 1–1: Thesis structure and contents.

2 The Quality by Design concept: an integrated approach to pharmaceutical manufacturing

2.1 QbD foundations: overcoming the quality myopia

The concept of quality has existed in different formats for many hundreds of years. Modern quality principles have evolved considerably since the beginning of the 20th century, making its first pivot away from "inspection-centered" when the theory of Scientific Management was introduced by Frederick Taylor. His philosophy was centered on the standardization of working methods to improve productivity and less in the common problems of inefficiency and slow rate of work. Other major breakthrough occurred in 1924 when Shewhart and others (e.g., Deming) at Western Electric introduced the concept of statistical process control (SPC) setting the beginning of the quality control era. During this period, important quality concepts were established, namely the demonstration that variations in the process lead to variations in product quality attributes. The foundations behind modern sampling plans and sampling guidelines are rooted in this period (Hyde, 1998).

Later in 1950, Feigenbaum, Cosby, Juran, and Taguchi pioneered the quality assurance era. Their theories were based on the previous quality standards (i.e., inspection and control principles) but the focus considerably shifted from a *product-oriented* to a *systems-oriented* quality approach. The introduction of quality manuals, quality planning and documentation control in this period, illustrates the broader spectrum of *quality* within the organization (Mazumder et al., 2011).

The next level of development was brought by Total Quality Management (TQM) focused on improving all organizational processes through the people involved. TQM is part of a wider concept, that addresses organizational performance and the importance of structured workflows to integrate quality into the total organisation which are the very primary foundations of the QbD concept.

Over the past decades, the pharmaceutical industry developed a very complex system (i.e., the Pharmaceutical Quality System – PQS) to ensure that high-quality products are accessible to patients. Despite these efforts and considerable improvement in technology and production methods, the number of incidents and drug shortages has increased in recent years (Gonce et al., 2014). To understand the underlying factors in the chain of events leading to the increasing number of quality issues, it is important to understand the wave of competing challenges pharmaceutical companies are currently facing. The increase in demand, and consequently in production run rate, combined with the increasing complexity of the pharmaceutical supply chain, raises the risk of failure considerably. Moreover, new products (e.g., biopharmaceuticals) feature more complex production infrastructures, requiring improved process controls to achieve

high performance and consistency levels. A more fundamental challenge is the prevailing quality culture (the “quality myopia”), focused on compliance, rather than on truly understanding the root causes and impacts of process incidents and deviations. The difficulty in conducting proper root cause investigations and trace-back to process or system issues is easily surpassed by the difficulty to shift the mind-sets of the quality and operation groups (Gerstner & Rutten, 2014).

Quality excellence relies on fully understanding the relationships between process settings and material attributes, from development up to routine manufacturing. This means that quality issues can be anticipated in the first place or promptly mitigated, relieving the burden of compliance bureaucracy created by the interplay of regulators and manufacturers. In this context, the QbD initiative has emerged as a science-based approach to product development and manufacturing, according to which, the manufacturing process should be developed to meet the desired quality specifications and to increase the scientific understanding of product attributes. The two main aspects of QbD are:

1. Identify and manage the critical quality attributes (CQAs) of the product;
2. Establish the *design space* of the process and its boundaries defined as the appropriate acceptable ranges for desired product quality.

Although final product testing is an important element of quality control, “quality cannot be tested into products; quality should be built in by design” (ICH, 2009). To realize the full benefits of QbD it is necessary to enable a thorough understanding of the relationship underlying the supply chain, process parameters and final product attributes (Rathore, 2009). Successful QbD implementation for a (bio)pharmaceutical product involves the following steps: 1) identification of quality target product profile (QTPP) and critical quality attributes (CQAs) that are essential to product safety and efficacy; 2) design of the process in order to deliver product specifications – risk-based identification of critical process parameters (CPPs) and assessment of interacting effects; 3) implementation of an efficient quality control strategy to guarantee consistent process performance; 4) implementation of an ongoing process verification (OPV) program to ensure process reproducibility over the lifecycle of the product.

Strategies for handling sources of variability, scientific understanding of the process based on PAT applications and, risk evaluation in order to establish appropriate operating ranges, provide a foundation for QbD implementation (Figure 2-I).

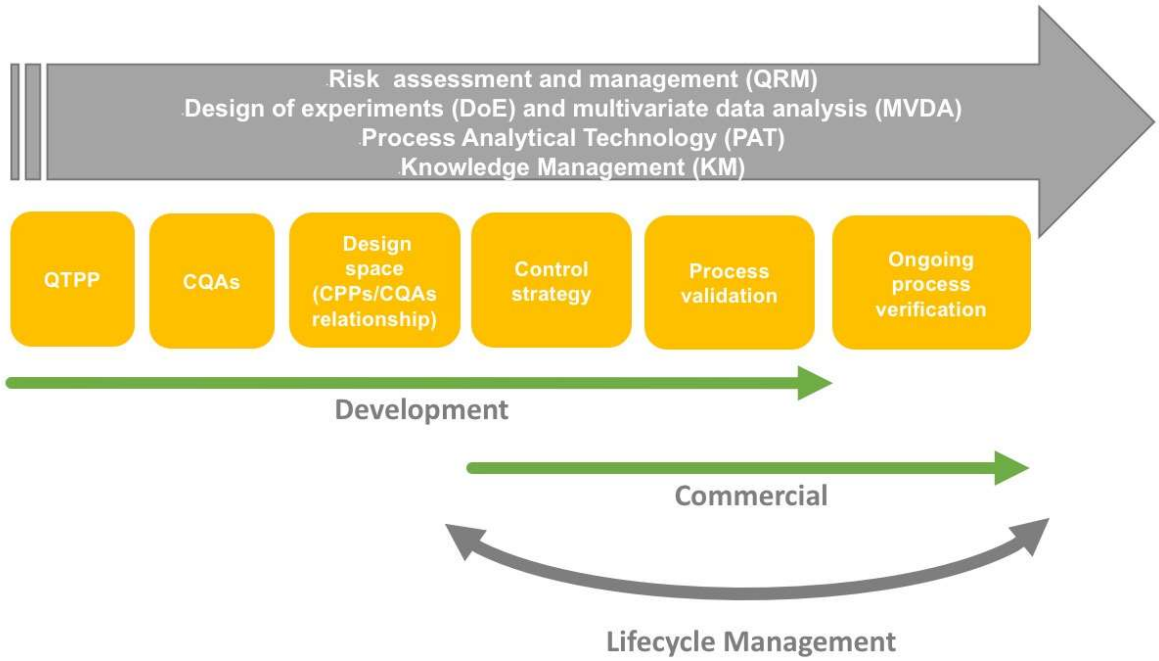


Figure 2–1: Quality by Design approach to (bio)pharmaceutical products development and manufacturing (adapted from Rathore, 2009).

2.2 The critical path for pharmaceutical innovation: requirements and enablers

High-quality products are the current standard of the pharmaceutical industry. Yet, *high-quality* requires much more than complying with a set of rules prescribed by regulators or having high-tech production processes and quality systems in place. Despite the significant impact of risk events, most companies still not have in place a comprehensive quality risk management (QRM) program to proactively assess risk sources and mitigate them on a continual basis.

QRM is not a one-time-only activity, but a mechanism that ensures that risks are continuously evaluated during the product lifecycle and remain within an acceptable level (Figure 2-2). The principles for designing a comprehensive approach to manage quality risks in pharmaceutical production are presented in ICH guidelines (ICH, 2006; ICH, 2009). Some of the tools recommended to effectively rank, monitor and manage quality risks are (non-exhaustive) Ishikawa diagrams, 5 Whys method, Failure Mode Effects Analysis (FMEA) Fault Tree Analysis (FTA), risk ranking and filtering and supporting statistical tools. The applicability of each tool depends on the prior knowledge available, the risk assessment stage (i.e., Risk Assessment, Risk Control or Risk Review, according to Figure 2-2) and the complexity of the quality issue to be addressed.

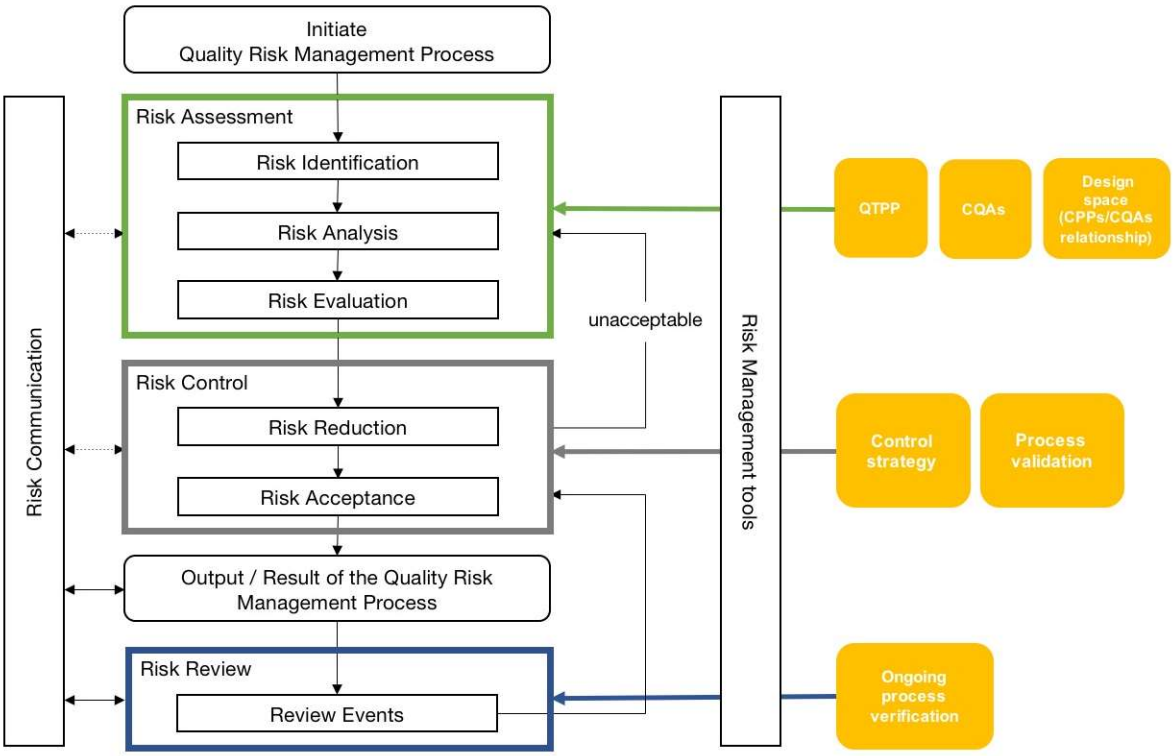


Figure 2–2: Integration of the Quality Risk Management workflow (ICH, 2009) within the QbD framework for product/process development and manufacturing.

Many production processes for pharmaceuticals were initially validated with limited knowledge from few production lots. Once processes are established, the flexibility for changes either to improve product quality or the efficiency of the process is significantly limited. In a very competitive landscape such as the one (bio)pharmaceuticals currently face, time-to-market is one of the key drivers for success. Therefore, strategic planning of development efforts and a comprehensive evaluation of potential risks are key elements to maximize operational and business efficiency. A step-wise approach is recommended, starting with a clear understanding of the patient needs and market constraints that is translated into product and intermediate attributes, leading to the definition of unit operations and in-process controls. During these activities, collaboration across functions within the organization is essential to identify, address and manage all critical risks to guarantee high-quality product and process performance levels during commercial manufacturing (De Boeck et al., 2014).

QRM is the basis for the transfer of process knowledge and for continuous improvement of the process control strategy robustness allowing to reduce the likelihood and impact of quality issues, to conduct faster and robust corrective-and-preventive-action (CAPA) investigations and to foster a preventive quality culture – i.e., a preventive-and-corrective-action (PACA) mind set – within the organization.

3 Bridging PAC and PAT domains in pharmaceutical manufacturing

3.1 Understanding the "5 Why's"

The quest of mankind to defeat illness has made the pharmaceutical industry one of the most profitable and successful businesses in the world. The continuous growth of the sector is fuelled by the constant demand of the world's population for longer and healthier lives. To fully understand the foundations of the industry it is important to contextualize the business environment. Research and innovation are key drivers, therefore a deep understanding of the technical side is mandatory (Beynon & Porter, 2000). But is also essential to realize that once a medicine reaches the market (i.e., is accessible to patients) the strongly regulated environment slows down the pace of innovation. As such, the lack of innovation in commercial manufacturing is usually accepted as a consequence of the many rules and regulations enforced by health authorities. The ingrained regimen of time-based quality inspections has discouraged companies to adopt innovative technologies in order to avoid post-approval changes and the associated costs, time and efforts required to obtain approval by regulatory agencies (Hinz, 2006). However, the large number of sequential steps and the many variability sources in pharmaceutical production result either in low performance or in quality events that jeopardize the final product quality and timely release. The situation was recognized by FDA and other regulatory bodies that initiated a collaboration with manufacturers to evaluate process analytical technologies (PAT) as suitable methods to gain a better control and understanding (the 5 why's) of manufacturing processes.

It is important to understand that the term analytical in PAT goes far beyond the process analyser domain (i.e., the PAC domain) combining chemical, physical, microbiological, mathematical, risk analysis and information technologies in an integrated and timely manner (US FDA, 2004b). The emphasis is given to process understanding, predictability and to enhance the overall efficiency of manufacturing processes. The opportunity to perform "timely measurements" in "whole process samples" is key under the PAT initiative to understand the process dynamics comprehensively. The industrial process is a sequence of unit operations with inputs, outputs and disturbances. Not only observability and state estimation is overlooked at the unit operation level, but also almost no use is made of the strong interconnected nature of pharmaceutical manufacturing. There is for example, a strong carry over of fingerprints (e.g., impurities present in raw-materials or abnormal process conditions during upstream production) that may have a strong effect later in the process or in the product. By linking unit operations in an overall description, it is possible to correlate cause and effect and act by anticipation to avoid or mitigate different events and disturbances impacting process performance and/or product quality (Felizardo et al., 2012). Only with this perception a true holistic process characterization can be accomplished.

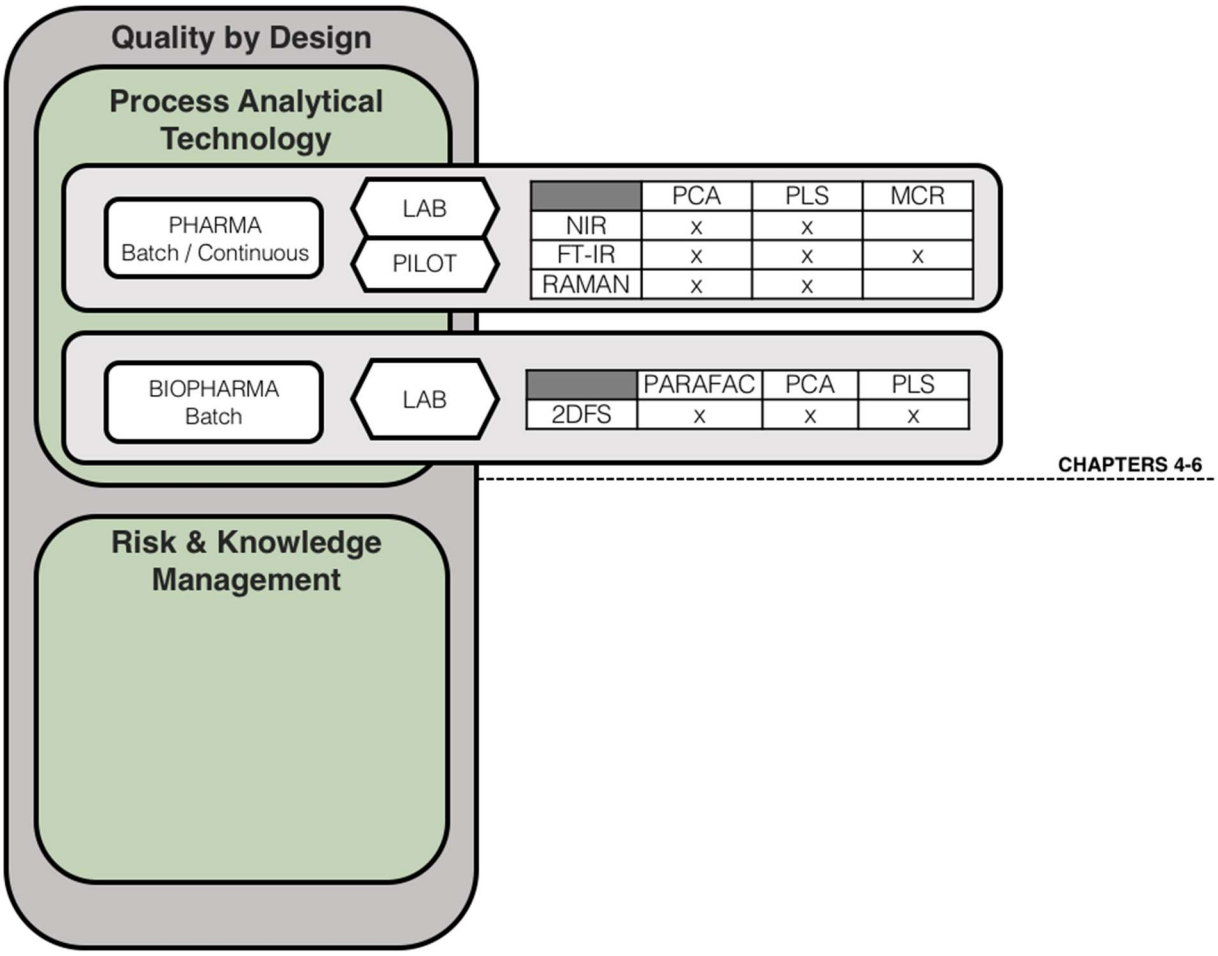
3.2 Evolution of the regulatory framework for pharmaceutical products

In 2004, FDA launched the 2-year initiative “Pharmaceutical cGMPs for the 21st Century – a Risk-Based Approach” which was to signal a shift in FDA’s regulatory practice (US FDA, 2004a). Soon after, a document defining “Process Analytical Technology – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance” (US FDA, 2014b) – was issued that addressed how the shift from quality by testing to the new paradigm of quality-built-into the processes producing the products –Quality by Design – should occur. That new FDA thinking was adopted by the European Medicines Agency (EMA) and culminated in the International Council for Harmonisation for Technical Requirements for Pharmaceuticals for Human Use (ICH) issuing over a period of less than five years three guidelines, Q8, Q9 and Q10 (ICH, 2009; ICH, 2005, ICH, 2008) and more recently ICH Q11 (ICH, 2012) that make up the foundation of modern pharmaceutical quality systems (PQS) based on QbD (ICH, 2009a; ICH, 2005; ICH, 2009b; ICH, 2012).

In 2011, FDA updated its nearly 25-year-old process validation guidance (US FDA, 1987) and proposed the alignment of process validation activities with the lifecycle concept, integrating all process design phases through commercial and routine manufacturing (US FDA, 2011). QbD is at the core of the guidance, endorsing a science-based approach to product development and manufacturing supported by prior knowledge and enhanced process understanding. Under the QbD framework, Process Analytical Technologies (PAT) are true enablers of the QbD vision, bridging process understanding, state estimation and control goals.

An organization’s capabilities in establishing a comprehensive science-based process design effort – focused on understanding all potential sources of variability and supported by knowledge gained during product commercialization fed-back in a continuous improvement effort – is the key element to a successful implementation of the validation lifecycle concept. For those organizations that have been embedding quality by design and risk management principles as part of their drug and process development strategies, the integration will be nothing but logical. For those who have relied on the agencies to prescribe the necessary requirements, the revision of the validation procedures against the new recommendations that in time will be mandatory may seem unachievable. The new guidance introduces a new mind set regarding the foundations of quality and its definition, challenging manufacturers to stand for what is appropriate to provide “scientific evidence” of process performance and robustness. The adoption of this approach will among other benefits, improve process capability and reduce cost of goods (COGs), facilitate and standardize post-approval changes and reduce regulatory burden while ensuring that products are of the highest quality, safety and effectiveness. An organization’s proficiency in implementing these approaches may dictate in the future who will remain competitive and in the long-term sustainable.

PART II - PAT in Pharmaceutical and Biopharmaceutical Production



4 Using PAT to Accelerate the Transition to Continuous Manufacturing

Significant improvements can be realized by converting conventional batch processes into continuous ones. The main drivers include reduction of cost and waste, increased safety, and simpler scale-up and tech transfer activities. Re-designing the process layout offers the opportunity to incorporate a set of process analytical technologies (PAT) embraced in the Quality-by-Design (QbD) framework. These tools are used for process state estimation, providing enhanced understanding of the underlying variability in the process impacting quality and yield.

This work describes a roadmap for identifying the best technology to speed-up the development of continuous processes while providing the basis for developing analytical methods for monitoring and controlling the continuous full-scale reaction. The suitability of in-line Raman, FT-infrared (FT-IR) and near-infrared (NIR) spectroscopy for real-time process monitoring was investigated in the production of 1-bromo-2-iodobenzene. The synthesis consists of three consecutive reaction steps including the formation of an unstable diazonium salt intermediate, which is critical to secure high yield and avoid formation of byproducts.

All spectroscopic methods were able to capture critical information related to the accumulation of the intermediate with very similar accuracy. NIR spectroscopy proved to be satisfactory in terms of performance, ease of installation, full-scale transferability and stability to very adverse process conditions. As such, inline NIR was selected to monitor the continuous full-scale production. The quantitative method was developed against theoretical concentration values of the intermediate since representative sampling for off-line reference analysis cannot be achieved. The rapid and reliable analytical system allowed: speeding up the design of the continuous process and a better understanding of the manufacturing requirements to ensure optimal yield and avoid unreacted raw materials and byproducts in the continuous reactor effluent.

Key-words: continuous processes, in-line monitoring, spectroscopies, process development.

This paper has been published in modified format as: **Gouveia F.F**., Rahbek J.P., Mortensen A.R., Pederson M.T., Felizardo P.M., Bro R., Mealy M.J. (2016). Using PAT to accelerate the transition to continuous API manufacturing. Analytical Bioanaytical Chemistry, 409(3), 821-832.doi 10.1007/s00216-016-9834-z.

4.1 Introduction

The pharmaceutical industry is now facing unprecedented challenges in their business environment. This dictates the need to change or modify strategies to mitigate the loss in revenue caused by patent expiration and lack of blockbuster medicines to replace them. To address these issues, pharmaceutical companies are reinventing themselves in the way research and development activities are conducted as well as optimization of new and already established processes (Baines, 2010). Continuous manufacturing is often seen as one potential approach towards *operational excellence* via production-on-demand and consequently many companies are shifting from traditional batch to continuous processes (CPs). Particularly, in active pharmaceutical ingredient (API) manufacturing, implementation of CPs holds several advantages. Among them, the possibility to operate in conditions that would not be viable under normal batch settings, improved yields and minimum by-product formation, achievements that are particularly relevant for reactions in which toxic and hazardous by-products can be generated (Xiang et al., 2012). Additional drivers, to move from batch-wise to continuous production, rely on the need to reduce time-to-market, simplify scale-up and technology transfer activities, while promoting innovation and continuous improvement. This is aligned with regulatory expectations advocating the need for more efficient and flexible manufacturing processes (US FDA, 2004a). Adopting CPs requires however a truly multidisciplinary effort as companies have to consolidate in-house deep knowledge of the chemical pathways, reaction kinetics and a thorough understanding of process systems engineering (PSE) principles to establish performance requirements for the continuous production set-up. Further hurdles include technical difficulties with available equipment and technologies, particularly when developing the control strategy for routine production (Poechlauer, 2012). Analytical technologies such as chromatographic methods have historically been used to support API synthesis, process development and commercial manufacturing. Although accurate and extremely reliable, on/in-line analysis and real time feedback are more difficult to accomplish with these techniques (Chen, 2011). Spectroscopic-based methods (viz., PAT tools) on the other hand, offer the opportunity to monitor real-time CPs, providing information of both reaction media composition and kinetics (Sellick et al., 2010, Cervera-Padrell et al., 2012; Roberto et al., 2013). Easy-to-use instrumentation is well developed in the mid-infrared (IR), near-IR (NIR), and visible regions of the spectrum. Both instrumentation and interfaces (e.g., fiber optics) have been optimized contributing to its increasing implementation in industrial-scale production (Vieira et al., 2003; Schaefer et al., 2013; Knop et al., 2013; Saerens et al., 2014). Deciding which technique to select will depend on sample format, measurement environment and understanding which tool provides the most relevant and accurate information to address the required purpose. In early development, the main interest relies in building up process understanding to speed scale-up activities, while in routine manufacturing, robustness (e.g., to production environment) as well as the ability to monitor and control the process is essential. Systems that benefit the most from PAT-based approaches are the ones dealing with transient/unstable intermediates, critical endpoints and sampling constraints due to high-energy reactants, extreme temperature and/or pressure conditions. In such cases, even combining several PAT tools (e.g., Raman, FT-IR, NIR) in the same reaction system can be a strategy towards

fast process and PAT-based system development. While Raman is a scattering measurement, NIR and FT-IR are absorption based. Such vibrational techniques are sensitive to dipole vibrations as in O-H, C-H and N-H bonds whereas Raman is sensitive to polarizable vibrations like those from the molecular backbone (e.g., C=C, C=N, aromatics). In general, strong bands in NIR and FT-IR spectrum of a compound correspond to weak bands in the Raman spectrum and vice-versa (De Beer et al., 2009; De Beer et al., 2011). These distinct spectral characteristics present a unique opportunity to investigate all information gathered, opening a window into the reaction chemistry and kinetics. In fact, using such complementary information can provide a better understanding and control of the underlying variability sources affecting the manufacturing process (i.e., disturbances]) and selection of the best in-process control (IPC) method to implement in industrial scale.

There are challenges, however, in adopting PAT tools to pharmaceutical real-time process monitoring and control ranging from equipment handling, calibration development, and lifecycle management of the PAT-based method. In addition, large and complex datasets are generated as a result of PAT tools implementation. Hence, multivariate techniques are required for information extraction and interpretation, enabling the formulation of mathematical models that can be used for process optimization and supervision of routine production (Chen et al., 2011; Rajalahti & Kvalheim, 2011) Such models can also be used for predicting process behavior, which is a clear evidence of increased process knowledge and understanding – a key objective of PAT approaches (US FDA, 2004b). In the present work, we present a technology screening roadmap to evaluate and develop a PAT approach from the earliest stage of development until transfer to a continuous full- scale unit. The major focus is given to build up process understanding and set the requirements for the control of the continuous setup.

4.2 Materials and Methods

4.2.1 Chemistry

A brief description of the reaction process in study is outlined in Figure 4-1, where 2-bromoaniline (**1**, Figure 4-1) reacts with 3M hydrochloric acid to form a suspension of fine 2-bromobenzeneammonium chloride (**2,** Figure 4-1) crystals. A nitration of primary aromatic amines with nitrous acid (generated *in situ* from sodium nitrite and hydrochloric acid) leads to 2-bromobenzenediazonium chloride (**3,** Figure 4-1), a diazonium salt that can be isolated upon subsequent displacement with a nucleophile (viz., I^-). The product of the reaction is 1-bromo-2-iodobenzene (**4,** Figure 4-1), a key starting material in the synthesis of an API commercialized by H. Lundbeck A/S. In this contribution the focus will be put on the diazotization reaction, which has been found to be the most critical step in order to optimize 1-bromo-2-iodobenzene yield and quality.

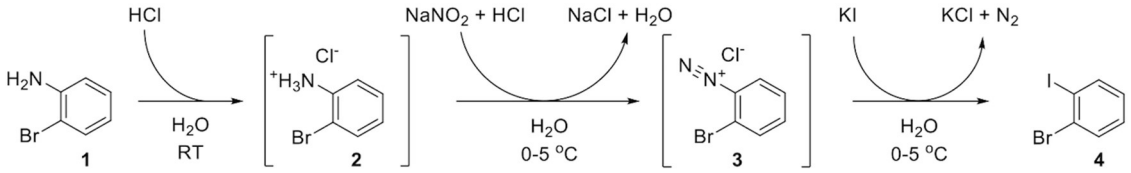


Figure 4–1: Synthesis of 1-bromo-2-iodobenzene: precipitation, diazotization and iodination reactions.

4.2.2 Technology screening roadmap

The true benefits of continuous processes can only be realized when a high level of control is achievable, i.e., when implemented technologies result in tight control of the process outputs, through a combination of equipment and control strategies. Spectroscopic measurements based on Raman, FT-IR and NIR technologies can be acquired in a very fast and non-destructive way, making all three tools eligible for control purposes. Each spectroscopic method holds advantages and limitations and the choice will depend upon the specific application.

Considering the highly diluted nature of the applied 1-bromo-2-iodobenzene synthesis, NIR spectroscopy could tentatively have been excluded from the technology assessment, as it is a relatively insensitive technique. The detection limits for specific compounds may be in the range of g/L. However, the sensitivity can be considerably enhanced using appropriate instrument configuration (e.g., detector type and source intensity) and settings (measurement principle, optical pathlength) while combining different wavelengths that contribute useful information for the prediction of a certain property (e.g., viscosity, density, concentration). Such calibrations can be very accurate (i.e., as typical reference analytical methods) and therefore useful as long as ‘good modeling practices’ are employed. This means proper use of spectral preprocessing and wavenumber selection, a parsimonious model structure not over-fitting the available data and appropriate statistical figures of merit used throughout (Menezes, 2009; Felizardo et al., 2012). In addition, NIR holds several advantages from a process implementation point of view, in terms of available interfaces (i.e., depending on the physical properties of the given sample/flow), simplicity and low cost of maintenance procedures, and long-term stability to very adverse conditions such as elevated temperature, pressure, and corrosive environments (Reich, 2005). FT-IR and Raman spectroscopies, on the other hand, potentially provide better-resolved and chemical-rich information from the reaction system (i.e., typically sharp and well defined peaks). There are some challenges, however, in terms of FT-IR and Raman implementation in process streams. The development of fiber optics in the mid-infrared region has improved significantly in later years, yet working distances are still constrained to relatively short lengths. Fiber-optics’ bending sensitivity can introduce unwanted variation and affect measurement reproducibility. Oppositely, Raman process analyzers can be very easily coupled with fiber optics and maintenance of fiber integrity is straightforward. However, from a practical standpoint, Raman implementation in industrial environments can be challenging due to fluorescence

effects derived from impurities or from the sample itself, that can interfere with the Raman signal and, most important, the cost of instrumentation which is significantly high (Smith & Dent, 2005).

In order to assess the performance of all available tools and interfaces, a technology screening strategy was put in place (Figure 4-2) to select the best method to introduce to the continuous full-scale setup. Firstly, process requirements were established by the development team, considering the inputs fed into the reaction system (flow-rates, preset ratios of reactants, energy) and the expected outputs (products, by-products, energy). This step is extremely important in order to select candidate tools to (1) enhance process understanding and later, (2) define the control system. Another key aspect of technology selection is the risk assessment component, which can be envisioned as the probability of failing to translate into a robust, reliable and effective control system element over the process lifecycle. Having a broad understanding of the inherent risks is extremely important to establish an optimal business case for the PAT tool, since all the following steps will require the involvement of a multidisciplinary team (development, manufacturing), resource allocation and experimental planning.

The use of the PAT tools as process-fingerprinting techniques was first explored to better describe the stoichiometry of reactants and products (PAT method selection - Screening Phase). Instead of developing quantitative calibrations, the spectral data was analyzed through multivariate projection methods (such as Principal Component Analysis) to obtain process trajectories and investigate the impact of chemical and physical variations in the spectra. This step is particularly valuable to enhance process understanding and also to evaluate the potential of each spectroscopy to be used for process controlling purposes. The most promising PAT methods were then selected for calibration development to investigate their suitability for setting the necessary control of the critical intermediate (Method development phase). Several probing designs (immersion vs. flow-cell)) and experimental setups (step-wise and continuous addition of reactants) were tested at this stage in light of the continuous process requirements. Moving the method to the full-scale setup will depend not only on method performance, but also on technical aspects (implementation in the manufacturing plant, equipment stability during routine production) and economic considerations. As an output from the Development Phase, an initial version of a semi-quantitative data-driven model (i.e., Partial Least Squares Regression method) was developed and used to establish appropriate control of a critical quality attribute control to the continuous reaction (Method Suitability evaluation). Depending on results the method can be later implemented in production or it might be necessary to further investigate additional methods in order to address control requirements.

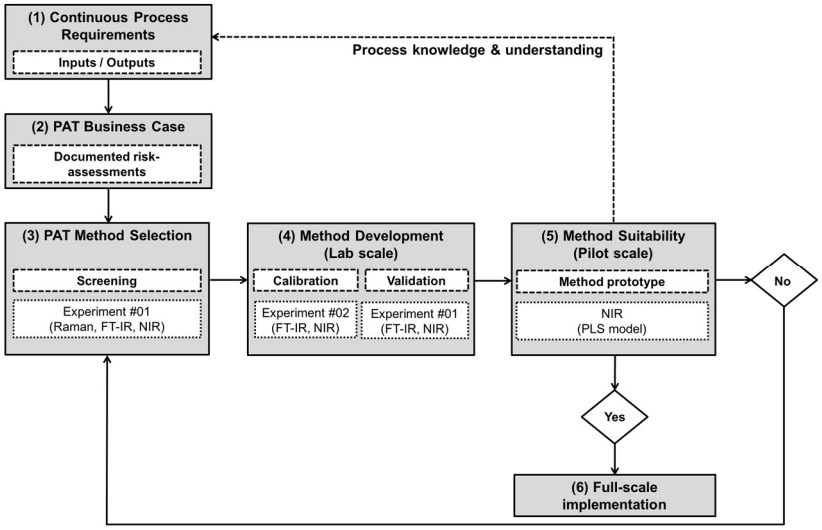


Figure 4–2: Technology screening and development of the PAT-based control system to implement in the full-scale facility.

4.2.3 Instrumentation

A NIR FTPA 2000-260 spectrometer (ABB Bomem Inc., USA) equipped with a transmission immersion probe (Q-Interline A/S, Denmark) with a 2 mm optical path length was used in Experiment #01. In Experiment #02, the probing system was switched to a NIR transmission flow-cell (Ocean Optics Inc., USA) with an optical path length of 1 mm, which was later implemented in the continuous full-scale reactor. Each recorded spectrum was obtained by averaging 128 scans with 16 cm-1 resolution over the range from 4500 to 15700 cm-1.

The FT-IR process analyzer used in both experiments was a ReactIR 45m (Mettler Toledo Inc., USA) spectrometer equipped with a 9,5 mm DiComp Fiber connected Gold sealed ATR probe. Spectra were acquired over the range from 4000 to 650 cm-1, averaging 64 scans with 8 cm-1 resolution.

Raman spectra acquisition was performed in Experiment #01 with a Kaiser RXN1 spectrometer, equipped with fiber optic connected immersion MR probe (1/4” - short focal point) (Kaiser Optical Systems Inc., USA). The excitation wavelength was 785 nm with a laser power of 400mW. Spectra were collected with 5 cm-1 resolution 5 seconds exposure time and averaging across 5 consecutive spectra. In the case of NIR and FT-IR a background spectrum was taken in air before each experiment.

4.2.4 Experimental procedure and setup

4.2.4.1 Laboratory scale: batch reactions

Experiment #01 (Figure 4-3a) consisted of a proof of concept sequence to evaluate the potential of each PAT tool to fulfil the established requirements, namely, a clear description of the reaction

stoichiometry, selectivity and capability of detecting a key-intermediate in the continuous production of 1-bromo-2-iodobenzene. Thus, NIR, FT-IR and Raman immersion probes were inserted in a magnetically stirred 250 mL glass reactor as illustrated in Figure 4-3a. The setup was cooled in an ice bath (5±2ºC) and the temperature monitored with an in situ probe throughout the experiment.

Experiment #02 (Figure 4-3b) was redesigned integrating the insights gained during the first feasibility run. As such, a NIR flow cell was implemented in a “by-pass loop” configuration to continuously recirculate the reaction media through the flow cell. The FT-IR immersion probe was directly inserted in the 500 mL jacketed glass reactor with temperature controlled through a thermo-stated oil bath (Julabo Inc., USA). All reagents were commercial products (Sigma-Aldrich, St. Louis, MO, USA) from the highest purity available (>98%) and used without further purification. Due to the multi-phase nature of the reaction system, the agitation rate was optimized in both setups to improve conversion efficiency and minimize the impact in the PAT measurements.

2-bromoaniline was suspended in water followed by step-wise addition of 3M hydrochloric acid (precipitation step). A solution of sodium nitrite was then added following two different procedures: i) in experiment #01 it was added stepwise via syringe until 50% stoichiometric excess was reached; ii) in experiment #02 the set up was redesigned in order to maintain a continuous feed (1 mL/min) of sodium nitrite, as in the first experiment, until 50% stoichiometric excess (diazotization step). In both feasibility studies, the initial reaction mixture consisted of thick slurry, turning into a clear solution as the reaction progressed (Figure 4-3c-e). Thus, the system was initially monitored with FT-IR and/or Raman only, as these techniques are reflectance based. In experiment #01, initial NIR spectra were removed from the dataset due to a very low signal-to-noise ratio, whereas in experiment #02, when the turbidity of the suspension was considered small enough to cope with the flow system design, sodium nitrite addition was interrupted and the mixture recirculated across the NIR flow cell. Some particles were still observed in the reactor and in the tubing, however, without interfering with the flow. Due to this constraint, the NIR dataset available is less comprehensive, compared to FT-IR and Raman, both in terms of spectral and chemical information.

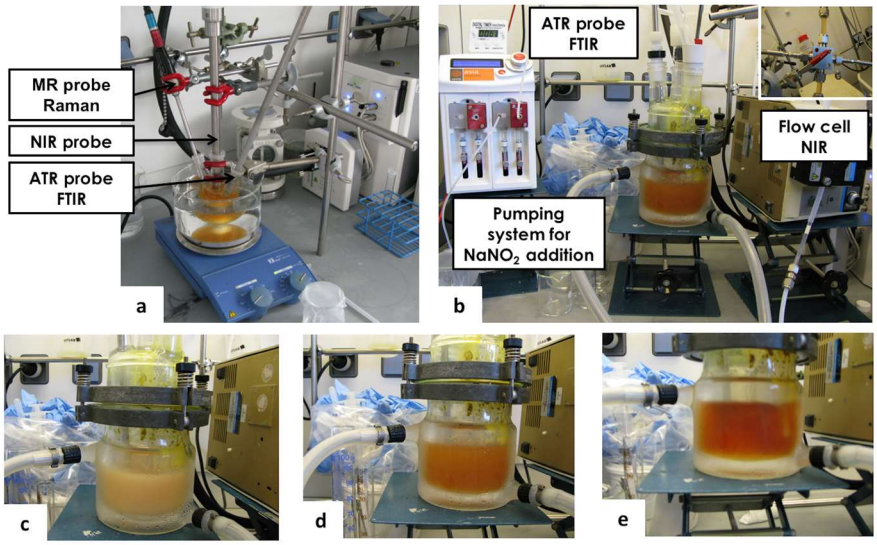


Figure 4–3: Laboratory setup for batch reactions: **a)** Experiment #01 (step-wise sodium nitrite addition); **b)** Experiment #02 (constant sodium nitrite addition); **c)** precipitation, **d**) early diazotization; **e)** late diazotization reaction.

4.2.4.2 Full-scale: continuous mode

The full-scale continuous reactor was customized in-house and implemented at H. Lundbeck's API production facility in Lumsås, Denmark. As part of the continuous manufacturing strategy, preliminary laboratory work on the synthesis of 1-bromo-2-iodobenzene was conducted and the continuous reactor designed and implemented in order to meet experimental criteria. First, accurate dosing and temperature control of the reactants leading to the formation of the first intermediate was found essential to limit precipitation and avoid clogging the system. Second, it is important to control the reaction in order to secure a high yield and avoid unreacted raw materials and by-products in the reactor effluent. Adding for instance too little sodium nitrite leads to unreacted 2-bromoaniline in the reactor output, which reduces yield and complicates the following work-up procedures. On the other hand, adding an excess of sodium nitrite both reduces iodide to iodine and produces NOX-gasses, which in turn demands an increased addition of potassium iodide and complicates work-up procedures. As illustrated in Figure 4- 4a, the continuous reactor is divided in two continuous sections. 2-bromoaniline is continuously pumped into a mixing tee, where it is protonated with a flow of 40°C 3 M hydrochloric acid to form 2-bromoammonium chloride. The formed 2-bromoammonium chloride flows into Loop A, where it is diazotized with a continuous feed of a 2 M aqueous sodium nitrite solution. The reacting medium is recirculated through a static mixer consisting of helical shaped PTFE elements to ensure fast mixing and a plate heat exchanger to remove the heat of the reaction mixture. The reaction loop provides enough residence time for the reaction to finish before entering Loop B. The NIR flow cell was integrated in-between the two loops in order to determine the concentration of 2-bromobenzenediazonium chloride, providing a basis for optimization and control of the process. The second loop (Loop B) was designed in a similar way as to Loop A, to convert the intermediate into 1-bromo-2-iodobenzene through a continuous addition of potassium iodide

solution. Sodium bisulfite ($NaHSO_3$) was added in the stream to improve the iodination reaction (Urbach, 1977). Finally, the effluent from Loop B was led into a reactor containing an aqueous sodium hydroxide solution for further batch work-up. An illustration of the full-scale prototype is shown in Figure 4-4b.

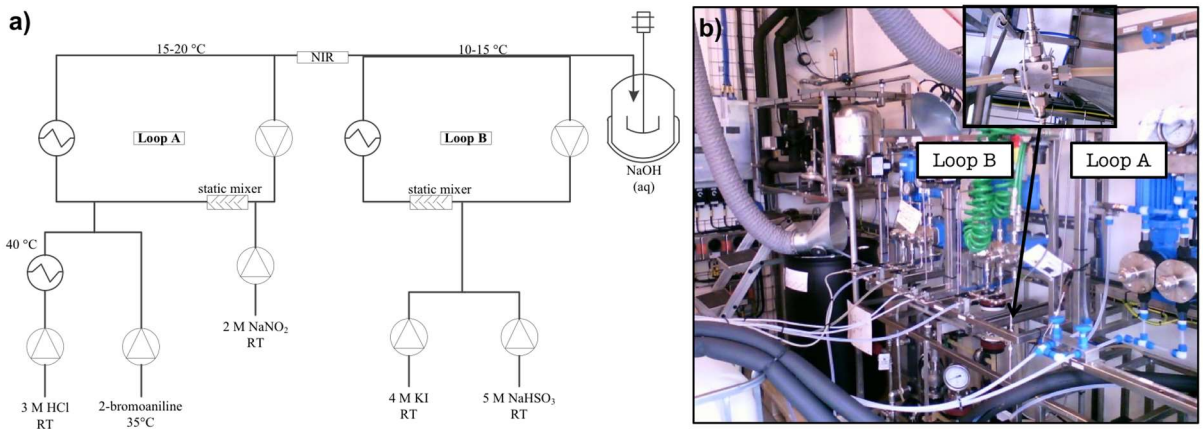


Figure 4–4: a) Diagram and **b)** reactor prototype used for continuous synthesis of 1-bromo-2-iodobenzene. Loop A: precipitation and diazotization steps; loop B: iodination step. The insertion of the NIR flow-cell in the process stream (i.e., in-between loops) is highlighted.

4.2.5 Multivariate data analysis

Data analysis including preprocessing and multivariate calibrations was performed using PLS_Toolbox® version 7.5.2 (Eigenvector Research, Inc., USA) for Matlab version 8.1 for Mac (Mathworks, U.S.A.). Different mathematical pretreatments were applied and tested to eliminate baseline effects caused by light scattering and/or instrument noise (Igne & Hurburgh, 2010; Xiaobo et al., 2010). Visual inspection of spectra over time and spectra-compound relation information was used to guide the preliminary selection of wavenumber intervals for the chemometric model development. Principal component analysis (PCA) was applied to identify major sources of variability in the datasets and to obtain process trajectories and reaction fingerprints directly from spectral data (Montague et al., 2008; Sandor et al., 2013). Partial least squares (PLS) regression with venetian blinds cross validation was adopted for all quantitative regression models. A combination of spectral range and variable selection methods such as Interval Partial Least Squares (iPLS) and Variable Importance Projection (VIP), were used to select spectral regions that are most informative with respect to the parameters under consideration. The number of PLS factors was chosen as the prime factor for which no significant variation in the root-mean square error of cross-validation (RMSECV) value was observed, while minimizing the number of significant latent variables to avoid model over-fitting. More details about the algorithms can be found in specialty literature (Bro, 2003; Rajalahti & Kvalheim, 2011). NIR models used to monitor full-scale reactions were implemented in FTSW100 (version 2.71) process software, integrated in the NIR FTPA 2000-260 spectrometer installation.

4.3 Results and Discussion

4.3.1 Qualitative analysis

Reaction spectra for Raman, IR and NIR are shown in Figure 4-5. Both Raman and IR characteristic peaks can be linked to specific functional groups in the sample. The most important changes in Raman spectra take place in the region at 650-485 cm-1 and are assigned to C-Br stretching vibrations, while C-H in-plane and out-of-plane bending vibrations are usually observed around 1300-1000 cm-1. The diazonium salt shows clear absorption bands at 1110, 1030 and 480 cm^{-1}, while 2-bromobenzeneammonium chloride disappearance can be followed at 540 cm^{-1} (Wiss & Zilian, 2003; Neal et al., 2013). Both FT-IR and NIR spectra are strongly influenced by the water signal, which absorbs broadly over the full spectrum. The diazonium formation manifests in the IR spectrum at 1462 cm^{-1}, assigned to the stretching of the azo linkage (N=N). At 1641cm^{-1} the carbonyl stretch can be identified (both in 2-bromobenzeneammonium and diazonium salt). The peak near 1466 cm^{-1} and 1566 cm^{-1} most likely matches the aromatic C-C stretching, whereas the peak arising at 1111 cm^{-1} can be assigned to C-N vibrations. Finally, the strong band at 965 cm^{-1} might be indicative of a di-substituted mode (Li et al., 2013). NIR peaks on the other hand, are broader and weaker resulting from combinations and overtones of those functional groups, and therefore peak assignment is not straightforward. In the present study, N-H overtone band (~ 5900 cm^{-1}) can be identified and change depending on substituents on the benzene ring.

Finally, disturbing features such as particle size, varying system turbidity and overlapping absorbing bands from different compounds may jeopardize qualitative and quantitative analysis. The compounds involved in the diazotization reaction resemble in a significant part of their chemical structure, thus several overlapping bands can be identified in all the applied spectroscopic techniques. Nonetheless, the formation of the diazonium salt manifests in specific bands in mid-, near-IR and visible (Raman) regions of the spectrum, as illustrated in Figure 4-5.

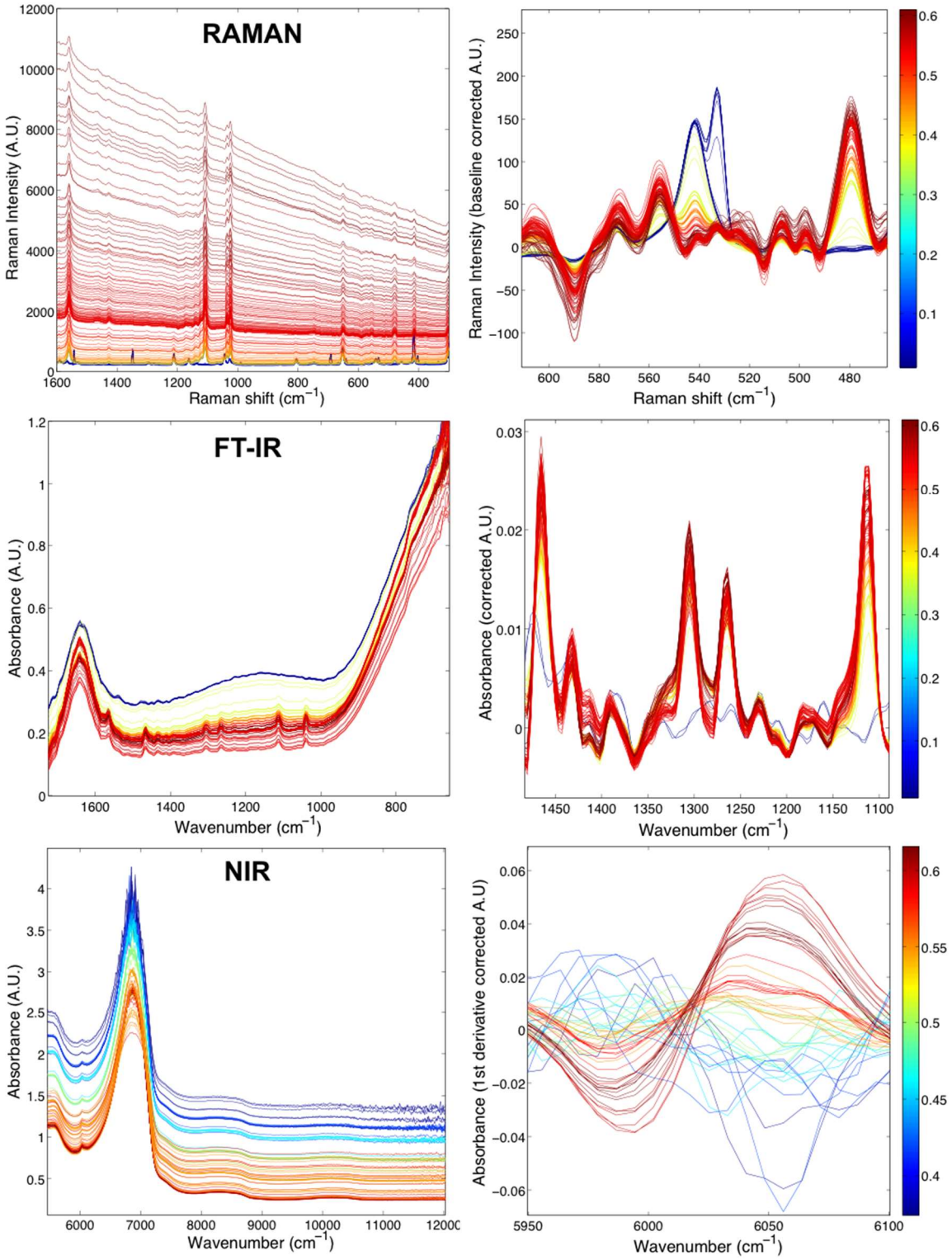


Figure 4–5: Raw (left) and preprocessed (right) Raman, IR and NIR reaction spectra. Colour scheme indicates 2-bromobenzenediazonium chloride theoretical concentration (M) over reaction time (specific bands selected for illustration purposes – left side plots).

Three major questions were addressed in this work: 1) How much information can each spectroscopic method capture from the reaction? 2) How accurate are the predictions? and 3) Which technology should be moved to the full-scale plant in order to monitor and establish appropriate control of a critical to quality attribute for the continuous reaction?

As to the first, Figure 4-6 illustrates process changes observed in the stepwise batch experiment captured by the different spectroscopic methods (Screening Phase). The same general conclusions can be retrieved by analyzing the different datasets although the amount of variation retained by the first two principle components decreases in the order Raman>FT-IR>NIR. Useful information content from the molecular structure also decreases in this order while the influence of physical properties (interferences) is ranked inversely.

The diazotization reaction can be split into two distinct phases, i.e., before and after the stoichiometric equivalence point (EQ). In the first phase, PCA component one captures the conversion of 2-bromobenzeneammonium chloride into the diazonium salt until maximum conversion. In the case of NIR, the spread in the scores indicates that PCA component one is affected by particle size effects. All spectroscopic techniques identify the same inflection point, in agreement with the equivalence point determined from the stoichiometric model. In the second phase, changes in PCA component one are no longer significant, except for a slight inversion in the score values due to dilution effect and perhaps, some instability of diazonium salts in aqueous solutions. At this point, PCA component two assumes a more dynamic role, shifting considerably towards negative (Raman) or positive (FT-IR and NIR) values. In PCA models, there is so-called rotational and scaling freedom which means that components may change sign, order or even be mixed when comparing different datasets. This explains the inverse evolution in PCA component 2 between Raman and FT-IR/NIR. The interpretation, though is kept unchanged. As such, the second PCA component is possibly capturing information related to accumulation of unreacted sodium nitrate (added in excess to the mixture) and changes in particle size during the reaction (i.e., slurry to a thin particle size mixture) that could not be completely eliminated by spectral preprocessing, particularly in the case of FT-IR and NIR.

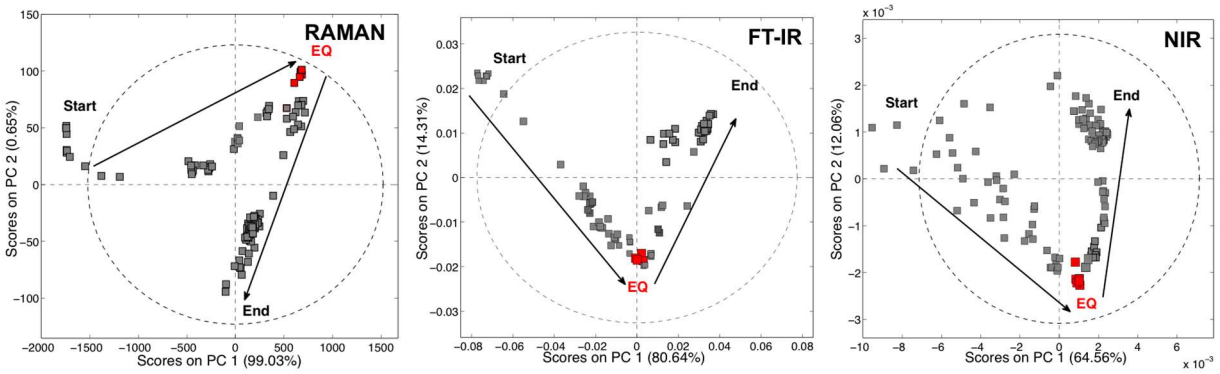


Figure 4–6: PCA component one vs. PCA component two (PCA model) of stepwise batch reaction of 1-bromo-2-iodobenzene (diazotization). Reaction start, equivalence (EQ) and end points are highlighted.

4.3.2 Predictive models

Conventional off-line HPLC analysis is not applicable for monitoring the reaction because the diazonium salt is not stable under HPLC conditions and representative sampling for off-line reference analysis cannot be achieved. Thus, theoretical concentration values based on reaction stoichiometry (1:1) were used to develop quantitative models with the different spectroscopic methods and address the second fundamental question arising in this study. Although far from ideal, it is expected that the absorbing species present in the calibration set are to be representative of those in future data sets assuming an acceptable lot-to-lot reagent variation. Nevertheless, in order to keep or even improve the model attributes over time a model maintenance plan should be implemented to accommodate for changing conditions (e.g., in raw materials or/and process settings) while preserving predictive capability (Wise & Roginski, 2015). Under these conditions, the current developed model might present an offset from true concentration values of the reaction intermediate. Even so, its utility is of great value since it is not possible to withdraw stable samples from the process stream. Furthermore, the calibration allows quantifying the intermediate in the concentration ranges planned for the continuous setup and overcomes two of the greatest challenges of this study, which are the lack of a reference method for developing the models and a rather difficult sampling procedure.

All the present spectroscopic techniques have the potential to characterize the diazotization reaction, however obtaining accurate and robust composition estimations of the intermediate proved to be challenging. Changes both in chemical and physical properties over the reaction had a very pronounced effect on the quality of spectra.

Hence, different preprocessing methods were tested in order to remove interferences undermining the linear relationship between the spectral data and the target chemical component. The accuracy and long-term reliability of multivariate calibration models for processes subjected to variations in physical properties (such as particle size and shape) highly depends on the ability to remove such undesired spectral variations, e.g., by preprocessing. Table 4-1 summarizes the best PLS models developed for the quantification of 2-bromobenzenediazonium chloride (stoichiometric) for each spectroscopic method. A comparative analysis demonstrates that all technologies are eligible to setup the monitoring system for the continuous process (Figure 4-7).

Table 4–1: Summary of Raman, FT-IR and NIR comparative performance (Screening Phase).

Parameter	Raman	FT-IR	NIR
Range (M)	0.00-0.62	0.00-0.62	0.45-0.62
Preprocessing	Baseline weighted least squares correction + 1st derivative, 2nd polynomial, 50 points window	Baseline weighted least squares correction	1st derivative, 2nd order polynomial, 9 points window
Variable selection	511-458 cm^{-1} 1389-1475 cm^{-1}	1570-1518 cm^{-1} 1458-1406 cm^{-1} 1234-1182 cm^{-1} 1011-996 cm^{-1}	5978-6164 cm^{-1} 8231-8671 cm^{-1}
Latent variables	2	2	2
Cumulative X variance (%)	85.3	97.9	93.4
Cumulative Y variance (%)	99.6	99.7	96.1
RMSEC (mol/L)	0.010	0.009	0.009
RMSECV (mol/L)	0.014	0.018	0.015

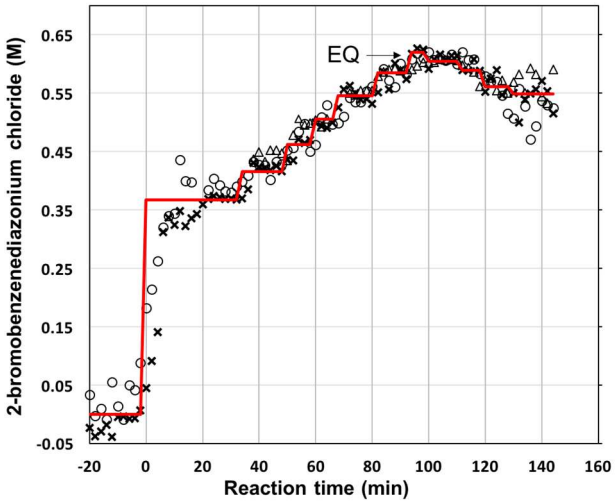


Figure 4–7: Theoretical (red line) versus predicted 2-bromobenzenodiazonium chloride concentration by Raman (x), FTIR (o) and NIR (Δ) in Experiment #01. Equivalence point derived from stoichiometric model highlighted.

The choice of the method will therefore rely upon considerations related with technical requirements and economic considerations. Raman holds several advantages considering the diluted environment in which the reaction takes place. Water is a weak Raman scatterer and therefore the technique has minimal sensitivity towards its interference. However, as the reaction progresses a strong background arises affecting the spectral signal (Figure 4-5). This might be linked to the laser source, as shorter laser wavelengths are prone to increased background signals, particularly when fluorescent compounds derived from impurities or from the media

build up over time (Strachan et al., 2007). This evidence combined to the higher cost of Raman instrumentation in comparison to the other two spectroscopic techniques supported the decision of ruling Raman out of the control system development.

FT-IR and NIR methods proved capable of monitoring the second experiment (Method Development Phase) where the system was redesigned to integrate the NIR flow-cell implemented in the recirculation loop (Figure 4-3B). The change in the NIR interface aimed to achieve improved sensitivity (i.e., the NIR flow cell probe has a longer path length) and cope with the highly diluted reaction stream. Optical path length is a key parameter of the NIR instrument and its choice depends upon the physical properties of the sampling material, such as viscosity and optical density. Turbid liquids normally require short path lengths for sufficient light penetration, whereas measurements in clear solutions can be performed with longer ones. In order to detect low absorption species, shorter path lengths may be required, however this will lead to higher noise levels. A compromise between the two effects must therefore be achieved (Jensen & Bak, 2002). Results show that both NIR equipped with the transmission flow-cell and FT-IR switched to the ATR probe are able to monitor the reaction and estimate the intermediate formation with good accuracy as shown in Figure 4-8.

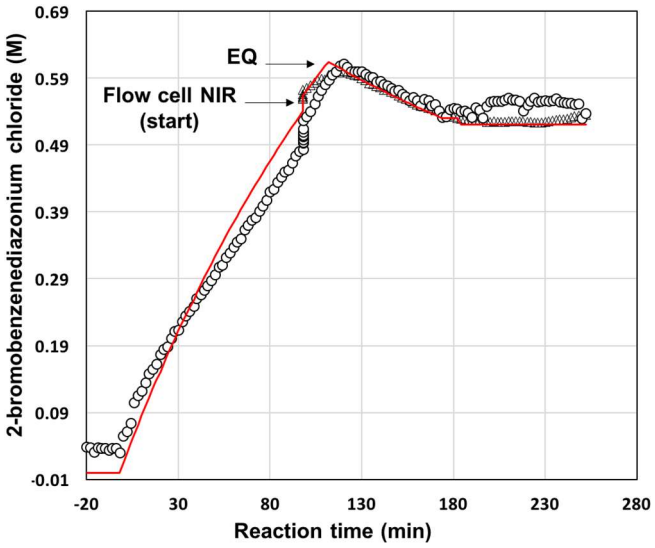


Figure 4–8: Theoretical (red line) versus predicted 2-bromobenzenodiazonium chloride concentration by FTIR (o) and NIR (Δ) in Experiment #02. Equivalence point derived from stoichiometric model highlighted.

The NIR calibration dataset comprehends a narrow range of samples as data was only collected for the second part of the reaction (i.e., when the particle size was small enough to pass through the probe path length). Nevertheless, in terms of accuracy, results are similar to the ones obtained with FT-IR (Table 4-2). In the case of FT-IR, a subset of wavelengths that produced the smallest RMSECV was identified by *i*PLS and the model was improved accordingly.

It is important to highlight that NIR spectra from Experiment #01 (used as a validation) was collected with a different probing system and extrapolates the current calibration range, which explains the higher relative error. However, within the acceptable range for the continuous reaction (ca. 0.49-0.61M), the accuracy of both methods is very similar.

NIR is very stable to adverse process conditions such as the highly corrosive environment of the reaction system, relatively inexpensive and there is a consolidated in-house knowledge about its industrial implementation. As such, NIR technology was chosen to integrate the continuous reaction and establish appropriate control of the key intermediate, addressing the third fundamental question arising in the present study.

Table 4–2: Summary of FT-IR and NIR comparative performance (Method Development Phase).

Parameter	FT-IR	NIR
Range (M)	0.00-0.62	0.52-0.62
Preprocessing	Baseline weighted least squares correction	1st derivative, 2nd order polynomial, 9 points window
Variable selection	1200-1163 cm^{-1} 1118-1040 cm^{-1}	5978-6164 cm^{-1} 8231-8671 cm^{-1}
Latent variables	1	2
Cumulative X variance (%)	90.2	97.9
Cumulative Y variance (%)	95.3	95.8
Calibration set	Experiment #02 (N=150)	Experiment #02 (N=82)
Validation set	Experiment #01 (N=12)	Experiment #01 (N=9)
RMSEC (M)	0.032	0.006
RMSECV (M)	0.032	0.008
RMSEP (M)	0.054	0.014
RMSEP/Range (%)	8.7	14.0

4.3.3 NIR implementation in full-scale continuous processing

In addition to monitoring intensification and obtaining a broad understanding of all input parameters, in a [continuous] pharmaceutical facility it is also important to maintain the in-process material in a state that will ensure meeting the specifications of the end-product. The steady state is often defined as the operational setting in which none of the individual parameters varies as a function of time, encompassing process variables, measured and manipulated variables. From a process engineering standpoint, steady-state is a slightly overrated definition in the sense that an industrial system seldom operates at perfect steady-state because of fluctuations

in instrumentation (pumps, flow-meters), lot-to-lot variability of input materials and environmental conditions. Most important from a control perspective, is to allow some flexibility in specific parameters (input parameters) to improve process efficiency (yields or to minimize impurities formation) and guarantee that specific critical quality attributes (CQAs) stay within specification. The ultimate goal is to ensure that the process is running in a “state-of-control” rather than achieving “steady-state” (Myerson et al., 2015).

An important aspect in the optimization of the full-scale setup is optimization of residence time distribution, a probability function describing the time the fluid element (in this study, a mixture between fluid and particles) spends in the reactor. Having the in-line NIR application and real-time detection of the intermediate conversion, give the production engineers the opportunity to better understand the process dynamics and a better description of residence time distribution for the two reaction steps [precipitation and diazotization]. The combined flow rate of 2-bromoaniline, hydrochloric acid and sodium nitrite was set between 46.4 and 50.1 L/h with a ratio of 2-bromoaniline/sodium nitrate of 1.1-1.5. At these settings, the concentration of 2-bromobenzenediazonium chloride is expected to be between 0.49 and 0.61 M, seen from 275-307 and 345-425 minutes, and within the acceptable performance range (Figure 4-9). In the period in-between, the concentration decreased below the acceptable range, however the conversion rate was constant and the turbidity was moving to lower levels. By this time (ca. 341 min) an adjustment was made in the reactant flows to maximize the conversion rate. The change is promptly detected by the NIR application, within 4 to 5 minutes.

Initial tests demonstrated that the turbidity of the reaction mixture varied considerably, depending on hydrochloric acid and 2-bromoaniline individual flow-rates and temperature. The ability of NIR to detect physical changes over reaction time was used as an early-warning indicator for mixing and ultimately, clogging issues. As such, a very simple model based on the baseline level of absorption over time (measured at 9500 cm^{-1}) was implemented in the plant software as an indicator of turbidity issues. This application was extremely useful from a control point of view as it gave the operator the opportunity to adjust the reactant flows, preventing tube clogging and driving the process within the acceptable concentration window (>80% intermediate yield).

Batches produced during the full-scale campaign were not run at completely “steady-state”. And since no reference method is established for the diazonium intermediate measured by this application, a direct verification of the NIR predictions was not possible. However, the predictions made sense both in terms of the intermediate diazonium concentration (compared with the calculated concentrations based on flow rates of the reactants) and in terms of turbidity. The real-time display of process measurements was crucial to understand the individual reaction phases and their interactions, the actions required to suppress the disturbances and a faster development of the continuous process.

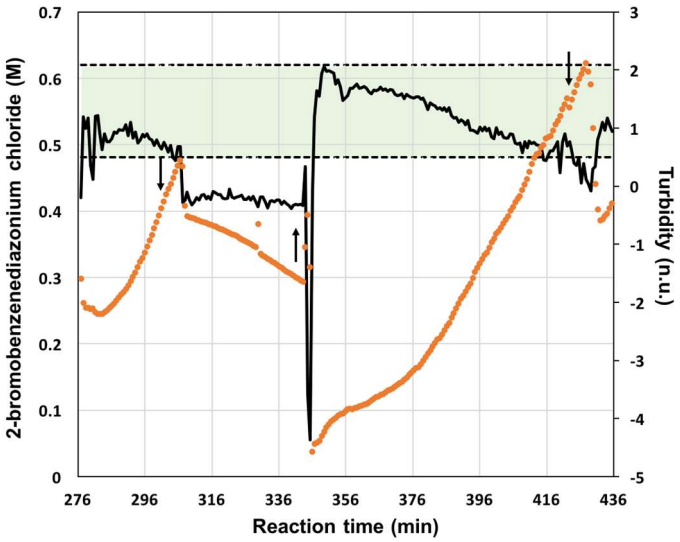


Figure 4–9: NIR online prediction of 2-bromobenzenediazonium in the continuous full-scale prototype (black) and spectra baseline absorption level over time (orange) as a surrogate measure of turbidity. Green area represents the acceptance concentration range of the intermediate; arrows indicate adjustments to 2-bromoaniline and/or hydrochloric acid flow rates.

4.4 Conclusions

In this work, feasibility studies for development of an in-line monitoring tool based on spectroscopic methods for the control of a diazotization reaction are described. The reaction is the second synthesis step towards formation of 1-bromo-2-iodobenzene, a starting-material in the production of an API developed by H. Lundbeck A/S. An accurate dosing of reactants on adequate stoichiometry was found essential to guarantee yield, quality and success of subsequent unit operations. First the potential of Raman, FT-IR and NIR spectroscopies was critically evaluated under batch reaction conditions. Second, a monitoring strategy with NIR was proposed for the continuous industrial-scale reaction, which is able to monitor the trend of intermediate production and identify abnormal situations such as tube clogging and disrupts in the flow. This approach presents a clear advantage over an otherwise difficult sampling procedure due to the system thermodynamics and characteristics (e.g., non-ideal multiphasic system, unstable intermediates, highly corrosive media stream). The results achieved are encouraging in that the spectroscopic method can support the process engineering goal to define more reliable processes able to deliver consistently high quality product. In order to select which technology should be employed in a particular environment it is fundamental to have a holistic understanding of the process. As such, the establishment of a cross-functional team encompassing several knowledge areas such as chemometrics, spectroscopy, reactor and process design as well as information technologies is highly recommended.

5 In-Depth Understanding of an API Synthesis: A Combined PAT and Chemometric Approach

A comprehensive regulatory framework endorsing the use of Quality by Design in pharmaceutical manufacturing is now in place. These documents promote a science-based approach supported by prior knowledge and enhanced process understanding obtained through Process Analytical Technologies (PAT). To fully realize the QbD vision, PAT tools need to be used in-situ, enabling process state estimation and enhanced understanding of the manufacturing requirements (Menezes et al., 2014). In the present study, in-line IR spectra collected from a complex, multi-phase reaction system were combined with chemometrics to enhance the understanding of the reaction mechanism. Different modeling strategies were applied such as, multivariate projection methods, partial least squares regression and multivariate curve resolution to (1) describe the stoichiometries between reactants and products, (2) develop a real-time monitoring system able to detect variations derived from process inputs manipulation and (3) identify improvement opportunities in the current manufacturing process.

A systematic procedure for exploiting the information provided by IR spectroscopy is highlighted, demonstrating how these [PAT] tools can support continuous improvement and innovation of commercial processes.

Key-words: batch to continuous, PAT, FT-IR, chemometrics; process development

This poster has been presented in modified format as: **Gouveia F.F**., Rahbek J.P., Mortensen A.R., Pederson M.T., Felizardo P.M., Bro R. (2016). In-Depth Understanding of an API Synthesis: a Combined PAT and Chemometric Approach. XVI Chemometrics in Analytical Chemistry (CAC-2016), Barcelona, Spain.

5.1 Purpose

Process Analytical Technologies have been widely adopted by the pharmaceutical industry and have been transforming the development and manufacture of drug substances (APIs) and drug products. The methodologies are mainly based on spectroscopy in combination with data-driven approaches (Liu et.al., 2006). In the present study, in-line IR spectra collected from a complex, multi-phase reaction system were combined with chemometrics to enhance the knowledge about the reaction mechanism. The aim of the study serves a twofold purpose: first, to improve the batch process currently in place and, second to establish the business case for setting the continuous process layout.

Development of an in-line monitoring system was found essential to fulfil the above-mentioned requirements. Different modeling strategies were applied to enhance information extraction, such as principal component analysis (PCA), partial least squares regression (PLS) and multi-variate curve resolution (MCR).

5.2 Material and Methods

A graphical representation of the experimental protocol is depicted in Figure 5-1. The FT-IR process analyzer used in all six experiments was a ReactIR 45m (Mettler Toledo Inc., USA) spectrometer equipped with 9.5-mm DiComp Fiber connected Gold sealed ATR probes. The system integrates a multiplex optical interface module to monitor two different reactions simultaneously. Spectra were acquired over the range from 4000 to 650 cm^{-1}, averaging 64 scans with 8 cm^{-1} resolution, throughout the reaction time.

Each experiment was conducted in an air-tight jacketed glass reactor equipped with an oil circulator unit connected to an in-situ thermocouple for temperature control, overhead stirrer, condenser, gas bubbler and the FT-IR probe.

250 ml of toluene were first added to the reactor followed by the catalytic system, reactant **A** and reactant **C**. The catalytic system was previously mixed for 5 min at room temperature for activation purposes. Addition and sampling operations were performed through a side arm. Reactant mixture **B** was added stepwise to the reactor with a glass pipette. Temperature increased by 35º-55ºC and the control set point readjusted to 85ºC, 95ºC and 105ºC, according to the experimental protocol in Figure 5-1. The viscosity increased significantly therefore, stirrer intensity was increased from 300 to 400 rpm to improve homogeneity. The system was continuously stirred at temperature set-point for approximately 24h.

2ml samples were acquired with a plastic pipette at timely intervals, from beginning until end of the reaction (in average, 28 samples per experiment). To stop the reaction, samples were immediately quenched with water. From the organic phase, 100 µL of each sample was transferred to a volumetric flask and diluted to 10 mL with acetonitrile (CH_3CN) of HPLC grade. Finally, 50

µL were transferred to a vial and diluted to 1 mL with a 20 %v/v aqueous solution of acetonitrile of HPLC grade. HPLC results were determined using in-house methods.

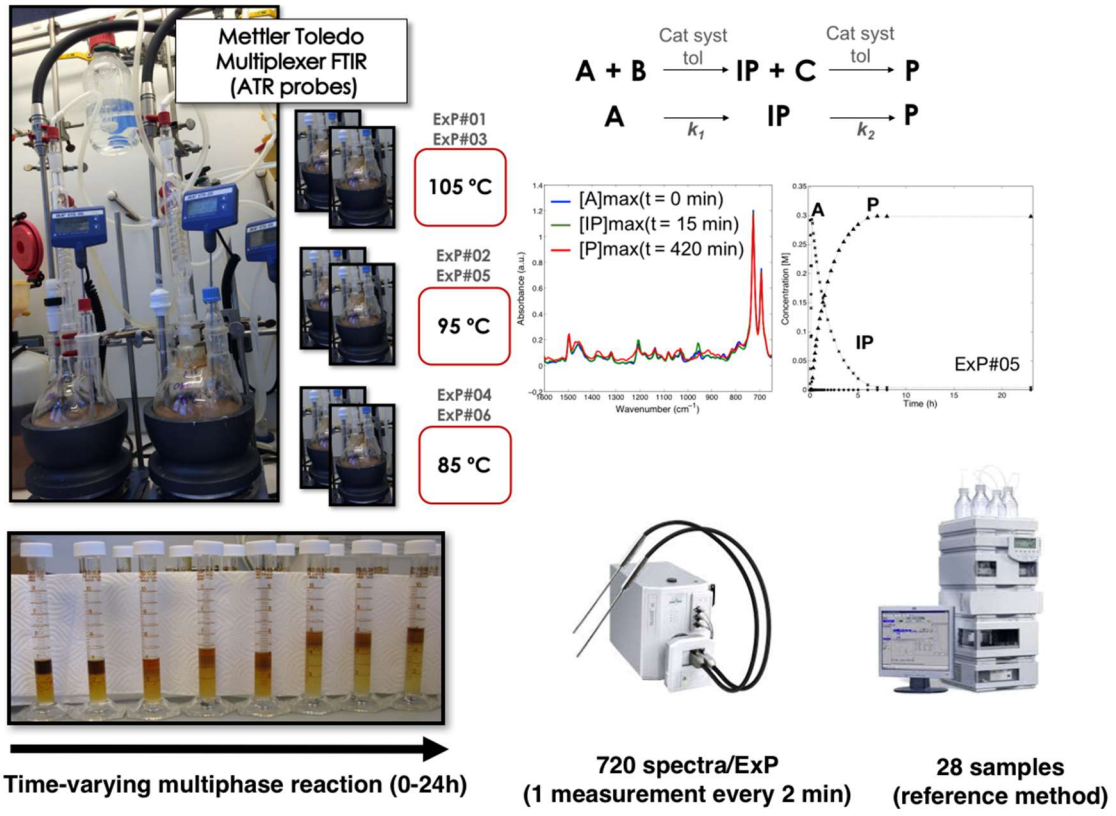


Figure 5–1: Graphical representation of the experimental procedure.

5.3 Results and Discussion

The purpose of the present work was to maximize information extraction from spectral datasets using different chemometric methods.

Principal Component Analysis (PCA) was used to examine the dominant patterns in spectral data (e.g., data integrity and batch-to-batch variability, temperature effects). It was possible to identify batch-to-batch variations but they were not related with temperature effects. The predominant changes in the collected spectra were baseline shifts resulting from light scattering due to different particle sizes (data not shown). An important outcome of the analysis is that the three main reaction species can be seen to be spectrally different as they show up in different locations in the main score plot (Figure 5-2). Hence, FTIR spectra include information from all reaction species and the analytical method can potentially be used to monitor all reaction phases.

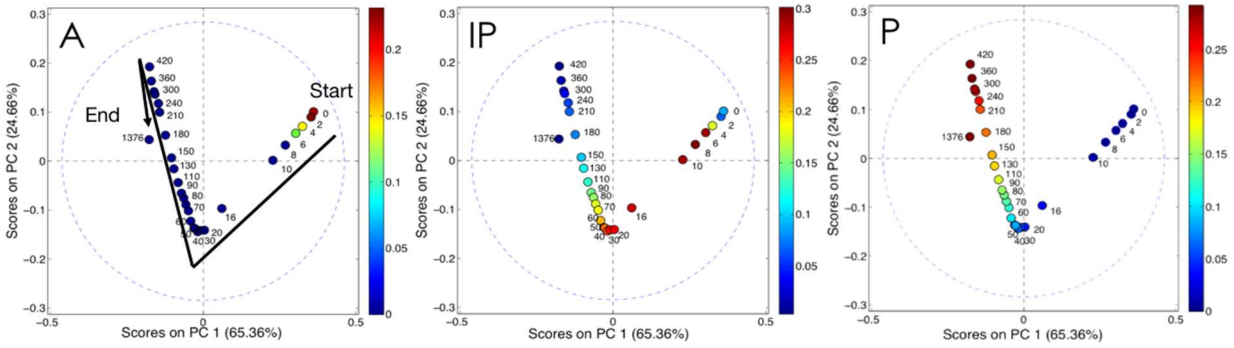


Figure 5–2: PCA scores of IR spectra showing ExP#04 reaction dynamics (A - reactant, IP - intermediate product, P – product) – colour scheme corresponding to component concentration [M] along reaction time measured by HPLC.

Partial Least Squares (PLS) models for real-time process monitoring and supervision of batch runs are important tools to characterize the reaction system and fasten process development (Wiss, 2015). In the present study, it was possible to develop accurate models for all chemical species. The model for product monitoring will be further detailed for illustration purposes (Figure 5-3a). After variable selection (Figure 5-3b), venetian blinds cross-validation was used to establish the number of latent variables (LVs) to be retained in the model. Calibration performance was determined relatively to the reference method by computing RMSECV. The external validation of the model comprised ExP#05 samples not considered in the calibration phase. Additional benefits of the presented study include the possibility to determine the reaction end-point based on spectral information instead of reaction time (Figure 5-3c). This additional knowledge can be translated into enhanced operational flexibility, allowing cycle time reduction and improved operational planning.

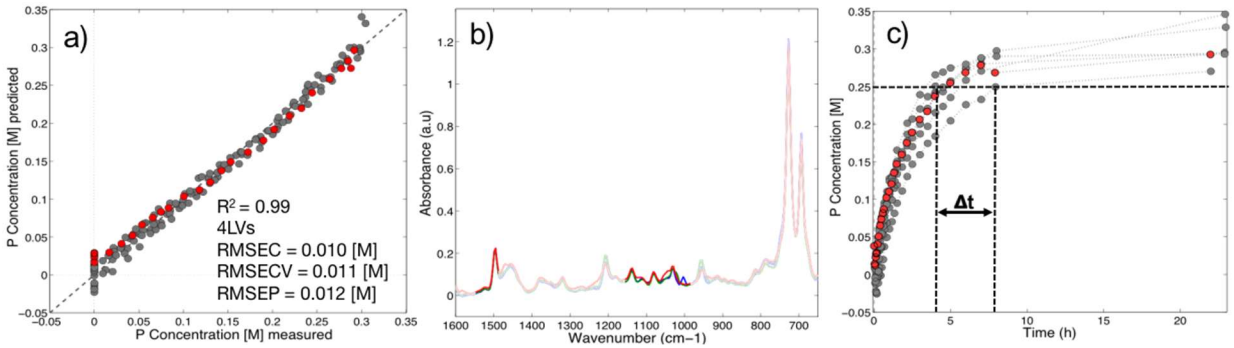


Figure 5–3 a) PLS model for product (P) concentration prediction; **b)** FTIR spectra where variable selection is highlighted; **c)** real-time monitoring for end-point determination.

Multiple Curve Resolution (MCR) was applied to the experimental dataset to investigate the process dynamics and the behavior of the different species over time. In the studied system,

the concentration of each individual compound is a function of both time and temperature. The main advantage of MCR applied to spectroscopy data is the recovery of the concentration profiles and the pure spectra of the compounds, even in the presence of unknown interferences (i.e., physical and chemical). First, the sequential structure of the reaction was determined by evolving factor analysis (EFA) to estimate the number of species spectrally active and the appearance-decay profiles throughout reaction time. Second, based on prior knowledge equality and non-negativity constraints were used to obtain a quantitative description of each experimental trial (Figure 5-4a). More details about the method can be found elsewhere (Juan et. al., 2000).

Prediction of experimental runs (ExP#03) with reference runs deconvoluted by MCR-ALS showed very similar results to HPLC-derived profiles (Figure 5-4b). With such approach, it was possible to significantly reduce sampling and lab testing during kinetic studies and most important, to identify at an early phase deviations from the desired kinetic pathway.

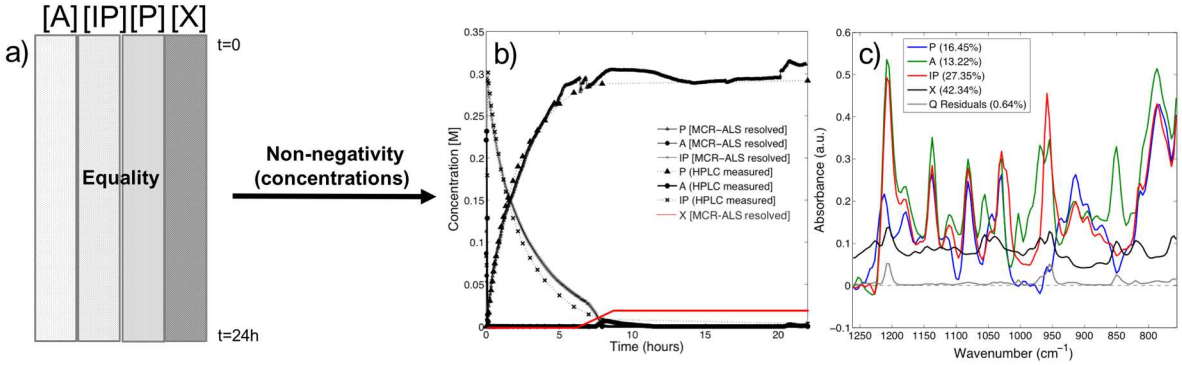


Figure 5–4 a) MCR_ALS model based on FTIR spectra (ExP#01-#02-ExP#04-#06) to investigate ExP#03 synthesis; a) number of components estimated by Evolving Factor Analysis; **b)** kinetic profiles MCR-ALS resolved; **c)** "pure spectra" exhibiting absorbance peaks in the same spectral regions as spectra acquired from pure solutions of reactant A, intermediate (IP) and product (P).

5.4 Concluding remarks

The extent of information existing in complex data structures such as the ones provided by spectroscopy methods may not be apparent from a cursory evaluation. Chemometric techniques such as PCA, PLS and MCR have the power to unveil minor differences in complex and limited sample sets. The in-depth knowledge gained through the application of such techniques

is of great value to overcome time and resource-related challenges often encountered in pharmaceutical applications. In the current study, FT-IR in combination with chemometrics was very effective to guide the experimentation leading to a better understanding of the batch reaction and technical information about the synthetic pathway to establish the business case for continuous manufacturing of the API.

6 Fluorescence Spectroscopy as a Process Analytical Technology (PAT) Tool to Support CHO Cultivation Process Monitoring, Optimization & Troubleshooting Activities

The last decade has witnessed an unprecedented use of advanced process analytical technologies (PAT) to efficiently perform (near) real-time monitoring and control of bioprocesses.

Here we report on the use of 2D-fluorescence spectroscopy and chemometrics (viz., PARAFAC, PCA, and PLS methods) for rapid, reliable and effective (near) real-time monitoring of CHO cell cultivations. Inline 2D-fluorescence data was acquired throughout the fermentation and analyzed with two different purposes. First, to investigate the underlying chemical structure of the multivariate fluorescence spectra, extract process information and find optimization potentials (i.e., characterization of the physiological state of cultivations, identification of variability sources affecting batch reproducibility and evaluation of 2D fluorescence fault diagnosis potential). Second, the same spectral dataset was used for indirect in-line monitoring of relevant cell culture variables, such as product, lactate and glucose, viable cell density (VCD) and lactate dehydrogenase (LDH) activity. Preliminary work establishes 2D-fluorescence combined with multivariate methods as a robust method delivering high frequency real-time process state information thereby supporting process optimization and troubleshooting. The opportunities for using 2D-fluorescence spectroscopy as an element of the control system in commercial biomanufacturing are considerable and are highlighted in the present study.

Keywords: 2D-fluorescence, CHO cultivations, process optimization, chemometrics, monitoring & supervision

This paper has been submitted in modified format as **Gouveia F.F**. ., Felizardo P., Menezes J.C., Jung C., Buziol S. (2017). 2D-Fluorescence Spectroscopy as a Process Analytical Technology (PAT) Tool to Support CHO Cultivation Process Monitoring, Optimization & Troubleshooting Activities (submitted to Biochemical Engineering Journal).

6.1 INTRODUCTION

Over the last ten years, a significant increase in approved biopharmaceutical products for use in an extensive range of diagnostic and therapeutic applications has been observed. These include proteins, peptides, nucleic acids, whole cells and viruses whose typical manufacturing process consists of cell cultivation, product recovery, purification and drug product formulation (Hakemeyer et al., 2013). Indeed, commercial production of biopharmaceuticals presents several unique challenges related to the high complexity of biological systems, especially if mammalian cells are used in multicomponent media mixtures, and the target product represents a small fraction of the bulk culture fluid. Furthermore, worrisome impurities are difficult to detect, raw materials are complex and subject to variability, and the functional/clinical significance of product variants are often incompletely characterized or understood (Read et al., 2010). The ability to improve and optimize bioprocesses under conventional development and manufacturing practices can be, in general, limited by the lack of reliable real-time process state information of all process stages. The Process Analytical Technology (PAT) initiative launched by FDA in 2004 has emerged as a science-based approach to product development, according to which the manufacturing process should be designed and implemented with the goal of ensuring final product quality and consistency (US FDA, 2004b). More comprehensively, the Quality by Design (QbD) framework embraces the creation of a manufacturing knowledge base established on risk-management principles, process design spaces, and scientific understanding of product attributes. PAT can be viewed as a QbD enabler, offering the opportunity to increase process knowledge and transfer it consistently, beginning from R&D up to production (Rathore, 2014).

With such a broad scope of applications, it is expected the role of PAT during product and process lifecycle to be very comprehensive. Generally, during process development, PAT applications are focused on establishing a measurement system for monitoring potentially critical quality attributes and key parameters, identify what is critical through process optimization studies, fill in knowledge gaps as to process parameters and dynamics, and test control strategies to ensure process robustness. In contrast, during routine manufacturing, the realization of PAT approaches is more focused on control and supervision of the process, which implies that a significant part of process understanding was already established during the development phase. Nevertheless, inherent and non-intended variability (e.g. in raw materials) will be observed over the process lifecycle. As such, the use of PAT tools should be extended during routine manufacturing to support ongoing process verification, improvement initiatives, and root cause analysis of deviations.

NIR, MIR and Raman spectroscopies have been used for process fingerprinting or monitoring of several process variables, media components and end product, whereas dielectric spectroscopy has been used to determine biomass volume/concentration throughout the entire cultivation time (Carvell & Dowd, 2006; Holm-Nielsen et al., 2008; Read et al., 2010; Triadaphillou et al., 2007; Ulber et al., 2001; Veale et al., 2007, Cole et al., 2015).

Particularly, 2D-Fluorescence spectroscopy has received considerable attention for high-throughput analysis of bioprocesses since it is a noninvasive technique with relatively high sensitivity and selectivity (Teixeira et al., 2011). Using a multi-wavelength excitation (λex) / emission (λem) spectrofluorometer, it is possible to simultaneously measure fluorescent signals derived from cells (e.g., cofactors) and medium components such as amino acids and vitamins (Rossi et al., 2012). However, the interpretation of a fluorescence map is not always straightforward as it is influenced by the cellular metabolic state, by overlapping signals from other components (medium and cell produced/derived) that might fluoresce in the same wavelengths, and by a comprehensive set of interferences, such as cell debris accumulation (particularly in death phase), gas bubbles, stirring, and high cell densities and viscosities. Moreover, the sensitivity of the measurements can decrease due to inner filter effects i.e., when fluorescent compounds absorb the exciting radiation or the one emitted by other fluorophores, or cascade effects observed when emission from one fluorophore excites a second one. However, using appropriate spectral handling and data processing methods, valuable information and robust multivariate prediction models can be retrieved for process monitoring and optimization purposes (Teixeira et al., 2009a).

The benefits of combining 2D-fluorescence spectroscopy with chemometric tools have been demonstrated in several studies (Hitzmann & Faassen, 2015). The use of such tools for bacterial and yeast culture monitoring was demonstrated with the development of a 96-well plate fermentation system with in-line monitoring of scattered light and fluorescence at specific wavelengths (Kensy et al., 2009). This strategy allowed selecting the best producer clone by directly correlating biomass growth with scattered light intensity for standard bacterial and yeast expression systems. Furthermore, PARAFAC modeling was successfully used to monitor profiles for enzyme and tryptophan concentrations in bacterial (Bacillus) cultures, and multivariate prediction models were developed for biomass, ethanol, and glucose of batch cultures of Saccharomyces cerevisiae (Mortensen & Bro, 2006; Ödman et al., 2009).

Even though the use of 2D-fluorescence spectroscopy to monitor mammalian cell fermentation is less covered in the literature, several applications have proven to be extremely valuable to reduce variability and enhance process understanding. Successful prediction models have been established for Green Fluorescent Protein (GFP) quantification of a mouse myeloma cell line (NS0), to effectively monitor viable cell density and recombinant glycoprotein concentration in recombinant BHK (Baby Hamster Kidney) cell cultures, and for monitoring cellular growth in CHO (Chinese Hamster Ovary) cell cultures based on changes in tryptophan, coenzymes and vitamins profiles (Hisiger & Jolicoeur, 2005; Teixeira et al., 2009b).

Lot-to-lot variability in raw materials can lead to significant large variations in the yield and quality of recombinant proteins. 2D-fluorescence spectroscopy coupled with multivariate data analysis tools can be used to complement existing media analytics with additional information about the quality of cell culture media in terms of compositional changes due to prolonged storage (Ryan et al., 2010). Other study demonstrated that it is possible to establish predictive models

for protein yield by fusing 2D-fluorescence with NIRS and MIRS data from different media components (Jose et al., 2011). Moreover, it was shown that 2D-fluorescence spectroscopy allowed to detect media composition changes upon storage that resulted in a decrease of the integral of viable cells (IVC) in CHO cultures, whereas product titer was not significantly affected (Hakemeyer et al., 2013).

2D-fluorescence spectroscopy combined with chemometrics has proven to be a robust and reliable method for handling different variation sources impacting bioprocesses. Monitoring of fermentation processes using inline 2D-fluorescence spectroscopy can potentially be complementary or even reduce the sampling workload for at-line analytical methods. The information can be correlated with the physiological state of the cells for the investigation of critical process steps or identification of process phases with specific requirements (e.g., feeding), supporting process understanding, troubleshooting, and optimization activities. For that reason, 2D-fluorescence encompasses a broad range of opportunities in terms of end-to-end optimization and continuous improvement over the process and product lifecycle.

6.2 MATERIAL AND METHODS

6.2.1 Cell culture

One recombinant CHO clone (Chinese Hamster Ovary cells, suspension culture) constitutively expressing an antibody (herein, designated as product) was cultivated in suitable chemically defined cell culture media. All cell cultivations (herein, CC#01 to CC#06) took place in a 14 L stirred tank bioreactor (Biospectra AG, Schlieren, Switzerland) employing state of the art process monitoring and control technologies, for approximately 14 days cultivation time. The working volume at inoculation was 10.0 L for CC#01-03 and 10.5 L for CC#04-06. The temperature of the biosuspension was kept constant at 36.5°C, and the pO2 was controlled at 35% air saturation. The process strategy consisted of an initial batch process followed by a fed batch phase to achieve higher cell density. The feeding strategy differed for the fermentation runs considered in this study. It comprised several separate feeding solutions starting at different times of the fermentation process. Bolus feeding of glucose and other nutrients was used for CC#01 and continuous feeding for CC#02-06.

6.2.2 Reference analysis

Samples were drawn from the bioreactor once or twice a day and either analyzed directly or centrifuged to remove cells and debris. Supernatants were collected and frozen until analysis, whenever necessary. Standard reference analytics included Cedex Bio HT Analyzer (Roche Diagnostics GmbH, Custom Biotech, Penzberg, Germany) for the concentrations of product (mg/L), glucose (mg/L), lactate (mg/L), and lactate dehydrogenase (U/L), while a Cedex Automated Cell Counter (Roche Diagnostics GmbH, Custom Biotech, Penzberg, Germany) was used for analysis of viable cell density (105 cells/mL).

For proprietary reasons, all analytical results were normalized by its maximum value according to Equation 6-1.

$$\text{Normalized Property (NP)} = p_i / p_{max} \qquad \text{Equation 6-1}$$

where p*i* is the *i*th row of the original dataset and p_{max} is the maximum value for each property considering cell cultivation experiments C#01 to C#06.

6.2.3 Fluorescence spectroscopy

Fluorescence spectroscopy analyzes the emission of light after absorption of ultraviolet or visible light by a fluorescence molecule called fluorophore. The fluorophore absorbs energy at a given wavelength and emits at a higher wavelength. During the first step, excitation, light is absorbed by the fluorophore, which is transferred to an electronically excited state. This is followed by a vibrational relaxation, where the molecule undergoes a transition from an upper electronically excited state to a lower one. When the electron returns to its ground state, it emits light at a specific wavelength according to the difference in energy between the two electronic states. Since emission and excitation spectra can be obtained from 2D fluorescence, qualitative and quantitative information about fluorophores present in complex media can be retrieved comprehensively. In other words, the technique can provide information about fluorescent molecules in a wide variety of biological samples with very high sensitivity (100 up to 1000 times more than other spectroscopic methods) (Haack et al., 2004; Jain et al., 2011; Lindemann et al., 1998).

In this study, on-line measurements were performed with the BioView® sensor (DELTA, Light & Optics, Denmark) for simultaneous detection of fluorescence at different wavelengths. The system employs two independent filter wheels with 15 different filters in the wavelength range from 270 to 550 nm for excitation and 310 to 590 nm for emission with a step width of 20 nm, and a photomultiplier for detection of the emission light. The fluorescence spectrometer is interfaced to the bioprocess via optical light guides and a probe inserted into a standard 25 mm port with an optical well containing a bottom surface sapphire window. A whole 2D-fluorescence spectrum is acquired after a complete cycle of both excitation and emission filter wheels. During the cultivation, the instrument was set to collect a full spectrum every 2 minutes with high sensitivity.

6.2.4 Data analysis

Multiwavelength fluorescence produces excitation (λ_{ex}) and emission (λ_{em}) spectra, gathered in a two-dimensional matrix per sample. When these variables are monitored over the culture time, a three-dimensional data array is produced requiring the use of specific data analysis techniques. Parallel Factor Analysis (PARAFAC) was used to explore the chemical structure of the dataset and to identify underlying phenomena over culture time (Figure 6-1). Several publications describe the theory and PARAFAC applications in bioprocesses (Bro, 1997; Mortensen &

Bro, 2006; Surribas et al., 2006). In the present study, visual inspection of emission and excitation loadings, split-half analysis and core consistency test were used for validation purposes (Andersen & Bro, 2003; Bro and Kiers, 2003).

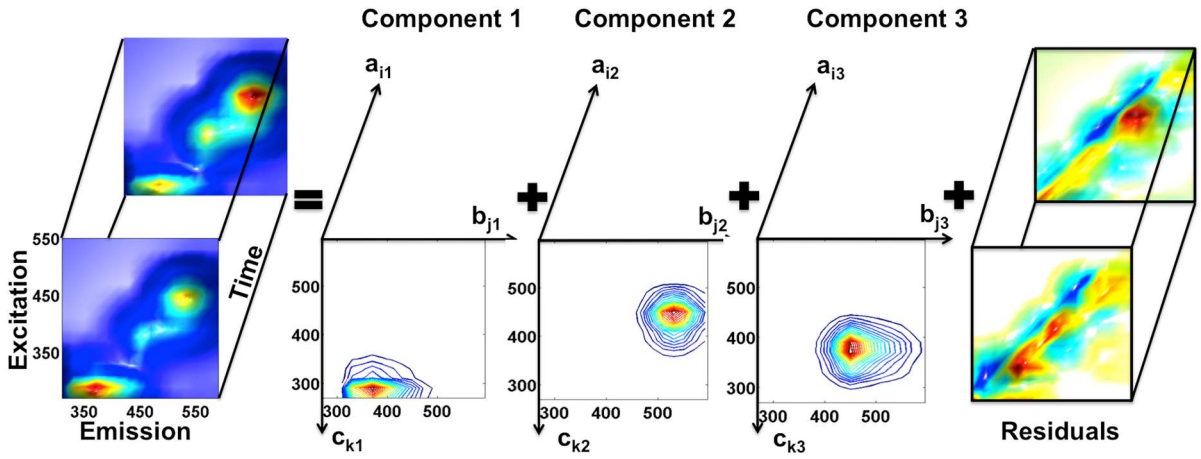


Figure 6–1: Graphical representation of a three-component PARAFAC model of 2D-fluorescence data from CHO cell cultivations.

To further investigate the properties of the fermentation process, three-dimensional fluorescence maps were unfolded into a matrix (sample vs. excitation/emission pair) and analyzed by Principal Component Analysis (PCA). The PCA method decomposes the array into sets of scores and loadings that can be used to interpret the dynamics of the fermentation process and describe the fluorescence "fingerprints" of the system under study (e.g., batch-to-batch variability) (Rhee et al., 2006, Ödman et al., 2009).

To verify the dependency between biogenic fluorophores and cell culture parameters in CHO batch cultures, two-way unfold-PLS models on centered data were developed and discussed in terms of accuracy and prediction performance considering Root Mean Square Error of Cross-Validation (RMSECV) and Root Mean Square Error of Prediction (RMSEP). For all models developed two cell cultivations (CC#03 and CC#06) where used as an external dataset (*i.e.*, not used in calibration) to fully challenge model performance and predictability. All calculations were performed using MATLAB® R2014a (8.3.0.532) (The MathWorks, MA, USA) and PLS_Toolbox ver. 7.9.4 (Eigenvector Research, Inc., WA), supported by in-house written codes.

6.3 RESULTS AND DISCUSSION

6.3.1 Fluorometric properties of CHO cell cultivations

Mammalian cell-based fermentation processes are very complex systems, requiring detailed monitoring of both chemical and physiological properties. Hence, fluorescence measurements

performed in a wide spectral range hold the potential to characterize the most biologically relevant fluorophores and provide important information about the cell metabolic state over cultivation time. The culture media (*i.e.*, prior inoculation) was first analyzed regarding its fluorometric properties. Three main regions with significantly distinct fluorescence intensities could be distinguished and assigned to flavins, NAD(P)H and amino acids, *i.e.,* phenylalanine, tyrosine and tryptophan, the last two being more dominant in the collected 2D spectral range (Teixeira et al., 2009b, Faassen & Hitzmann, 2015) (Figure 6-2).

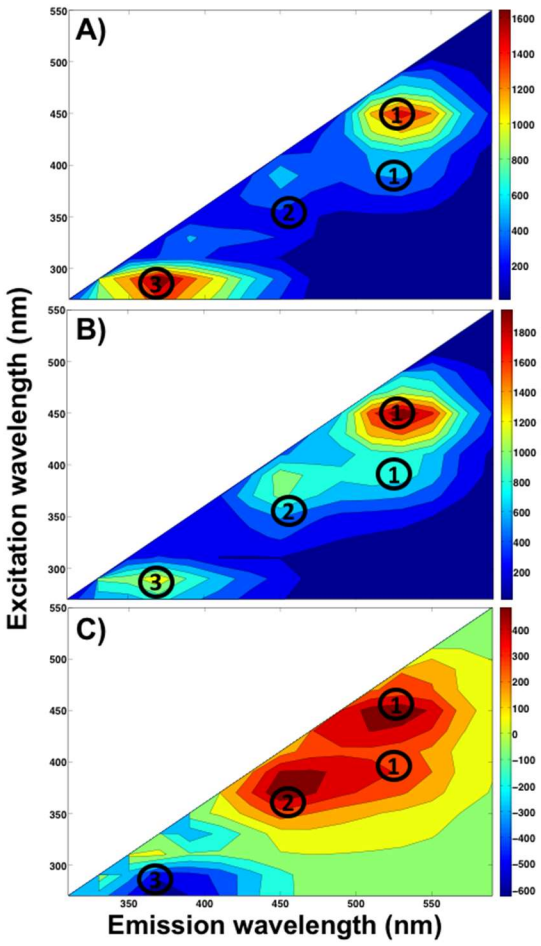


Figure 6–2: Biogenic fluorophores in 2D fluorescence map of CC#01: (**A**) prior inoculation; (**B**) end of fermentation; (**C**) difference spectrum between end and prior inoculation fluorescence maps. 1 – flavins (riboflavin, FAD, FMN); 2 – NAD(P)H; 3 – amino acids (tyrosine and tryptophan).

The fluorescence intensity in the amino acid region ($\lambda_{ex}\sim 290nm; \lambda_{em} \sim 370nm$), exhibited the strongest signal prior to cell inoculation (Figure 6-2A). Over cultivation time, a significant decrease in the intensity of the amino acids signal was observed due to cell uptake and integration into proteins (potentially both recombinant and cellular proteins, the last reflecting the increase in biomass).

Concerning the region where vitamins such as pyridoxine (vitamin B6 derivative) exhibit fluorescence ($\lambda_{ex}\sim 310nm; \lambda_{em} \sim 410nm$), a decrease in signal intensity, particularly during exponential cell growth was recorded (data not shown). The signal increases once again, remaining stable until the end of the cultivation. The shift to higher intensity was not expected since pyridoxine is not in the composition of the feed media recipes. Therefore, other unknown fluorescent compounds are most likely contributing to the signal intensity.

In opposition to amino acids, the fluorescence signal in the region of NAD(P)H ($\lambda_{ex}\sim 350nm; \lambda_{em} \sim 450nm$), was negligible prior inoculation and increased over time due to the metabolic activity of cells (Figure 6-2A and 6-2B). Some variations related to feed additions were observed over cell cultivation time. These might also reflect changes in environmental conditions to which fluorophores are exposed, namely in pH and ionic strength (Svendsen, Skov, & van den Berg, 2016).

As previously reported by Teixeira et. al. 2011, the fluorescence signal in the region of flavins ($\lambda_{ex}\sim 450nm; \lambda_{em} \sim 530nm$), correlates well with cellular growth. Notably, during exponential growth, it decreases steadily until end of this phase (data not shown) and then increases until the end of cultivation, presumably accumulating in the media. Fig. 6-2C depicts the difference spectrum between the end of CC#01 and the corresponding media preparation (*i.e.*, before inoculation) highlighting the substantial decrease in the amino acids fluorescence intensity while NAD(P)H and flavins regions evidence the opposite pattern.

Although very informative, the interpretation of fluorescence spectra can be hindered by overlapping signals and process-related interferences. Therefore, exploring the multivariate properties of fluorescence maps in combination with chemometric modeling can contribute to enhanced data quality and promote the use of 2D fluorescence for qualitative and quantitative monitoring of fermentation performance.

6.3.2 Batch-to-batch variability based on fluorescence profiles: investigation of optimization and fault diagnostics potential

For all six CHO cultivations, the fluorescence signal was recorded continuously and process samples characterized by off-line analytical methods for product, LDH, VCD, glucose, ammonia and lactate quantification. In the context of bioprocess optimization, the aim of this work was to extract information from fluorescence landscapes correlated to metabolically related compounds to characterize changes occurring during CHO cell fed-batch cultivations. PARAFAC was employed for the purpose as it can provide chemically interpretable information from process variations (*i.e.*, each PARAFAC component represents the concentration of intrinsic culture fluorophores) potentially affecting growth and cell metabolism. A three-component model was selected based on residual inspection, core consistency testing and split-half analysis, indicating that the scores and loadings from the resulting models are very similar, and confirming an appropriate description of the data (Table 6-I). The underlying chemical structures identified by the PARAFAC model can be assigned to pure solutions of biogenic fluorophores (Figure 6-

3) such as tryptophan (component 1), flavins (component 2) and NAD(P)H (component 3)(Faassen & Hitzmann, 2015).

Table 6–1: PARAFAC model diagnostics based on split-half quality assessment and core consistency test.

Cell cultivations	Components Overview Ex/Em (nm)	Explained Variance (%)	Core consistency
All (CC#01 – CC#06)	1 - 290/370 2 - 450/530 3 – 370/450	96.5	99.0
Odd (CC# 01, #03 and #05)	1 - 290/370 2 - 450/530 3 – 370/450	96.6	99.0
Even (CC# 02, #04 and #06)	1 - 290/370 2 - 450/530 3 – 370/450	96.4	100.0

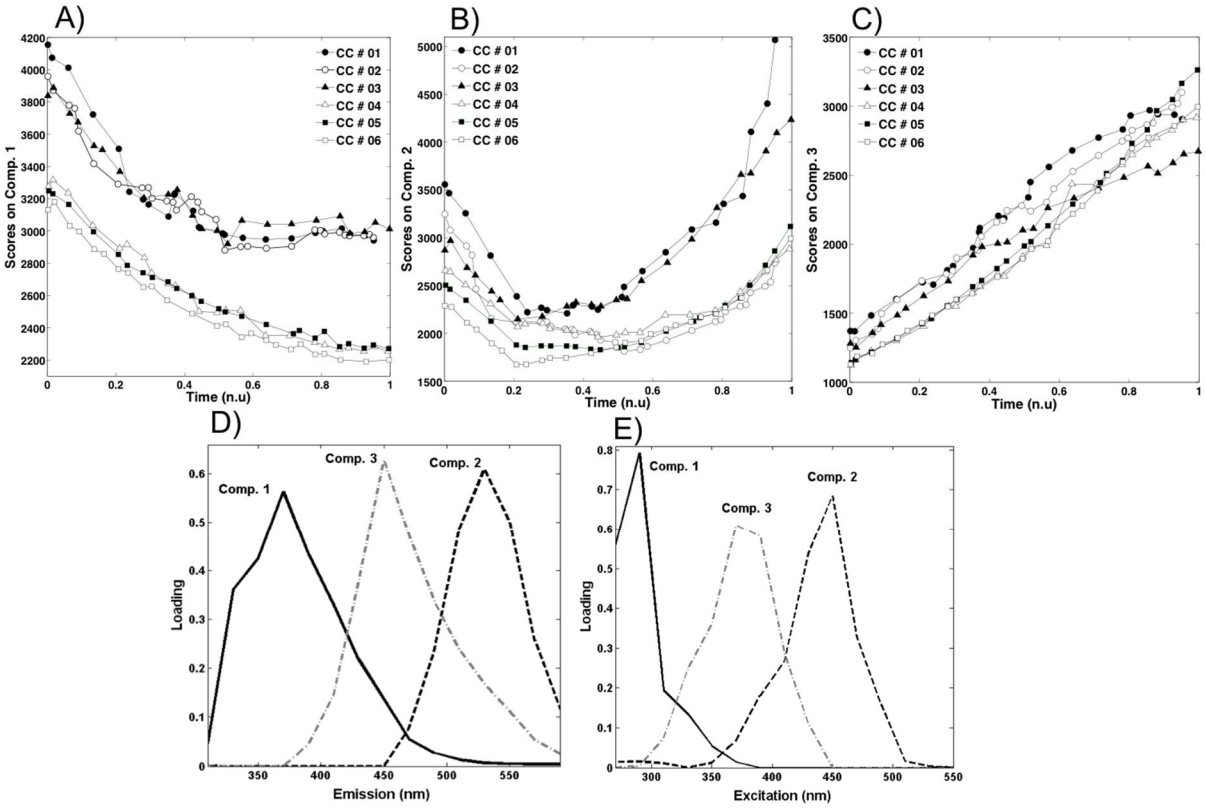


Figure 6–3: Overview of PARAFAC scores (**A**, **B** and **C**), emission (**D**) and excitation (**E**) loadings characterizing CC#01 to CC#06.

Component 1 (Figure 6-3D and 6-3E) captures the emitted fluorescence in the amino acids region where tryptophan exhibits its fluorescence signal ($\lambda_{ex} \sim 290nm; \lambda_{em} \sim 370nm$). As expected, a more pronounced consumption during exponential growth phase takes place, as cells require amino acids to incorporate in cellular and/or production proteins. Following, a stabilization phase is observed in which the feed stream adjusts amino acids concentrations levels, particularly in the case of cell cultivations CC#01 to CC#03. In CC#04 to CC#06 the decreasing trend is observed until the end of cultivation time (Figure 6-3A).

Amino acids play a central role in cellular metabolism for compound synthesis or degradation in a variety of pathways. Tryptophan is an essential amino acid formulated in chemically defined media used in CHO cultivations, *i.e.,* it cannot be synthesized by mammalian cells. Differences in the amino acids fluorescence signal are due to the use of different supplementation of media within the current study. The influence of media components in cell culture performance has been reported in several publications. Specifically, unbalanced amino acid formulations (concerning tryptophan and other amino acids) can promote accumulation of toxic by-products such as lactate and ammonia affecting cell growth and product yield (Xing et al., 2011). The present technique has the potential to rapidly assess variations in the culture media, particularly regarding aromatic amino acids, providing added value from a quality control point of view (Calvet et al., 2012).

Growing cells use riboflavin mainly for FAD biosynthesis, which denotes a fluorescent quantum yield approximately ten times lower than the one exhibited by riboflavin (Eitenmiller et al., 2007). Hence, the fluorescence intensity in the region of flavins (Figure 6-3B, component 2 Figures 6-3D and 6-3E) shows an almost linear decrease from inoculation (acceleration phase) until the end of exponential growth phase. Thereafter, variations can be likely explained by the fact that the feed-batch phase is initiated and riboflavin formulated in the feed media composition accumulates in the fermentation media, as cells require less substrate than in the later growth phase. It is thus possible to distinguish two different cell culture profiles starting at 50% total fermentation time, characterized by faster (CC#01 and CC#03) and slower accumulation of flavins (CC#02, CC#04 - CC#06). The root cause for this observation is not entirely clear, but it might be related to a slowdown in cell growth and protein synthesis capacity observed for CC#03. In the case of CC#01, the faster accumulation of flavins is most likely related to the bolus feeding strategy. Also, an accidental overfeed due to a fault in the feed pump control system was promptly detected in the fluoresce signal of CC#01 at 0.85 total fermentation time, characterized by a sharp and unexpected increase in component 2 score values. As such, a control strategy based on 2-D fluorescence signal can be used for fault-detection of similar events.

The third PARAFAC component (Figure 6-3C and 6-3D) describes NAD(P)H variations over culture time (Figure 6-3C). The signal increases progressively over time until the end of cultivation. NAD(P)H fluorescence can provide qualitative and quantitative information about cell growth and metabolic state. Indeed, the differences observed in NAD(P)H time profiles correlate

with two different metabolic signatures. Typically, lactate is preferentially produced in exponential growth phase, but the metabolism of cells can shift to lactate consumption, an event that can correlate with optimal process performance (Le, et al., 2012). Alternatively, cells can continue to produce lactate until glucose is fully depleted. Time profiles presented in Fig. 4 show that higher lactate concentration (Figure 6-4E) correlates in general with lower growth (Figure 6-4A) and protein yield (Figure 6- 4B) at harvest. Since 2D-fluorescence spectroscopy can provide insights from media/broth composition and cell metabolic state, it holds huge potential to increase overall understanding of desirable growth and metabolic phenotypes. Besides, 2D-fluorescence carries indirect information about the physiologic state pointing to process improvement in terms of product yield and quality. This information can be used to encircle the onset of metabolic events and to identify early-warnings for events/shifts in the process, very useful for process optimization and troubleshooting.

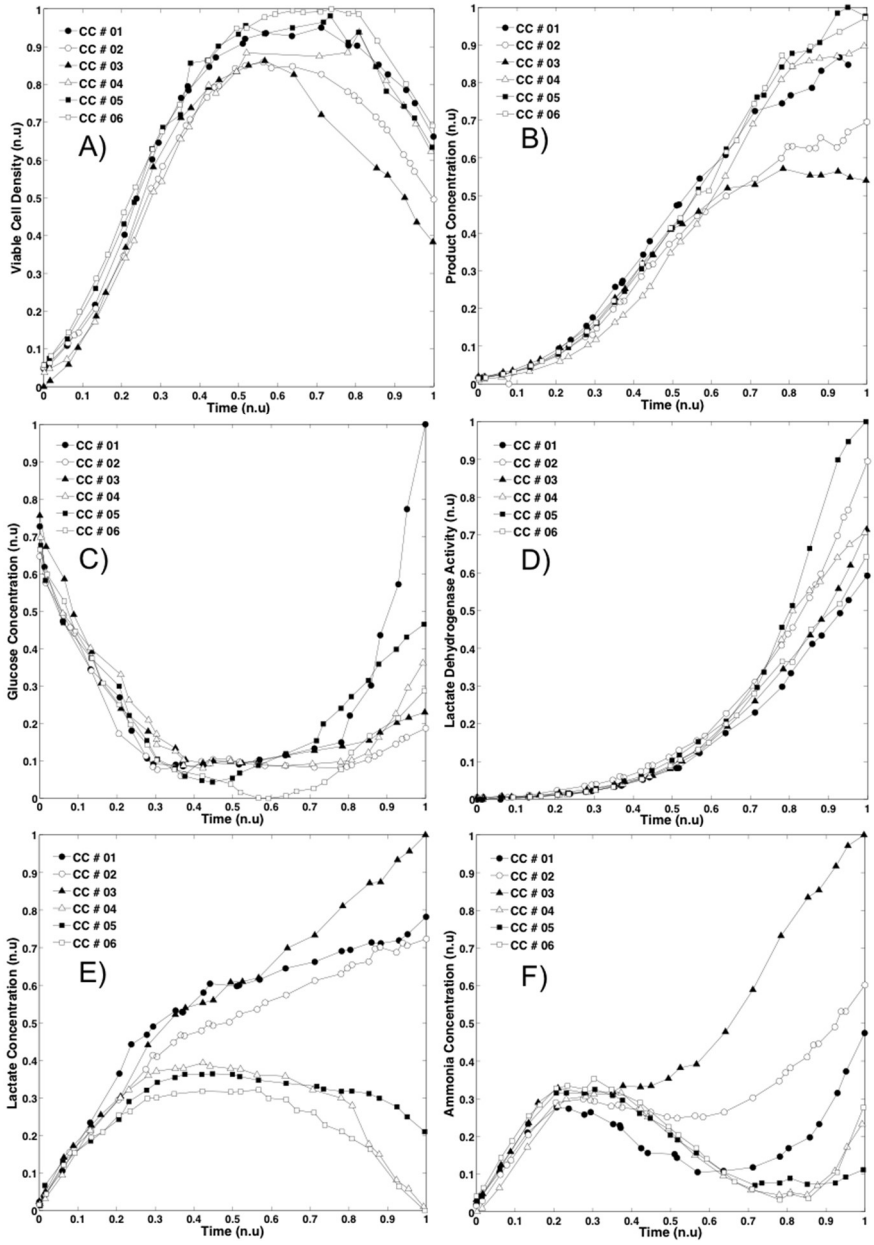


Figure 6–4: Time profiles of (**A**) VCD, (**B**) product, (**C**) glucose, (**D**) LDH activity, (**E**) lactate and (**F**) ammonia for six CHO cell cultivations (normalized units so that the maximum across all runs is one).

To better understand the dynamics of the process under study, a two-component PCA model of 2D-fluorescence unfolded spectra enabled the identification of three main phases over cultivation (Figure 6-5C). Phase I (growth phase – batch mode) is characterized by intensive cell division and high flux of glycolysis from glucose to lactate. During Phase I, fermentations CC#01, CC#02 and CC#03 showed larger variation in PC2 score values relatively to fermentations CC#04, CC#05 and CC#06 (Figure 6-5C – red dots) which may be an early indication for differences in the cultures metabolic state. The decrease in the second PC is mainly capturing the amino acids and vitamins consumption (Figure 6-5D)

Phase II (stationary phase – feed-batch mode) starts around 0.2 fermentation when the feed stream is initiated, which is clearly captured in the score plot by the sharp inflection of the scores trajectory, making PC1 score variation predominant over PC2 (Figure 6-5C – green samples). PC1 which is most influenced by flavins and NAD(P)H signals increases during this phase, while PC2 score slightly decreases over time, reaching its minimum value when the maximum viable cell density is reached for each cell cultivation. In CC#01, CC#02 and CC#03 the extension of this phase was significantly shorter and a sharp inflection in PC2 scores could be captured around 0.5 total fermentation time (Figure 6-5A and Figure 6-5C) (transition to Phase III – death phase, blue dots), while for fermentations CC#04, CC#05 and CC#06 this event happened much later, around 0.8 fermentation time (Figure 6-5B and Figure 6-5C). This change in the slope of the process signature, captured by 2DFS, most likely reflects a significant change in cell physiology due to accumulation of lactate and ammonia, which are known for inhibiting growth and promoting cell death. Once reaching critical inhibitory levels, these factors lead to earlier cell death and consequently lower product concentration at harvest for those fermentations (Figure 6-4 – CC#01, CC#02 and CC#03).

As showed by PARAFAC and PCA models, the most significant events captured by 2DFS are either related to cell physiology or process changes including variations in media composition (e.g., overfeeds) and feeding strategy, demonstrating its potential for process optimization purposes and to design more efficient control strategies for the manufacturing process.

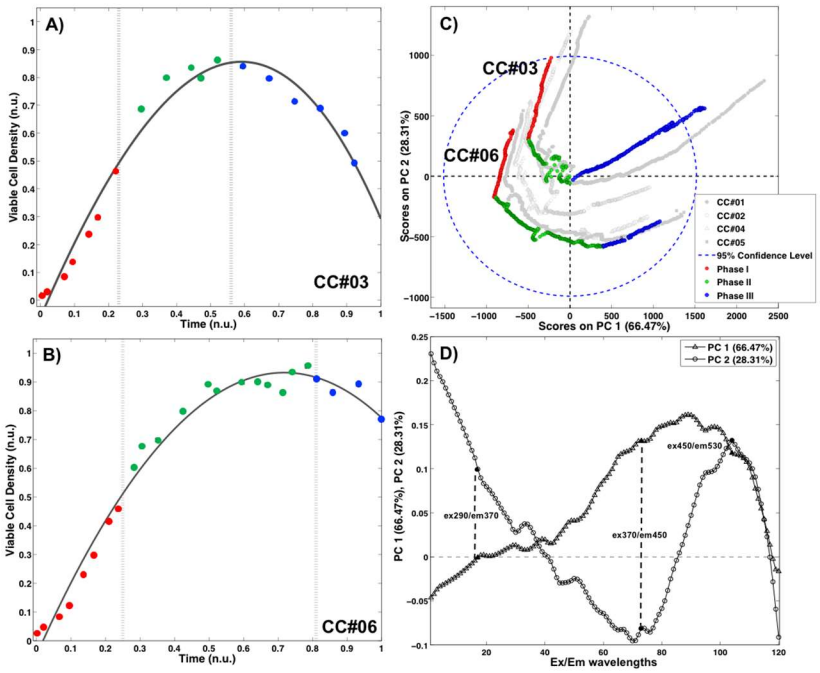


Figure 6–5: VCD profiles of CC#03 (**A**) and CC#06 (**B**), and PCA model of unfolded 2DFS spectra for all six CHO cell cultivations: PC1 vs. PC2 (**C**) and loadings (**D**) plot – Tryptophan (ex290/em370), NAD(P)H (ex370/em450) and flavins (ex450/em530) excitation/emission maxima. Phase I, II and III colored red, green and blue, respectively.

6.3.3 Inline monitoring of CHO cell cultivations

Several inline process variables (*e.g.*, temperature, pH, dissolved oxygen) and at-line variables reflecting more directly the quality status of the process (*e.g.*, product concentration, viable cell density) are routinely measured during cell cultivations. There is a considerable time-lag between the availability of inline measured process variables and variables reflecting the physiological state of the cell, pointing to the quality status of the process. Information about the quality of the product is only available after process completion, and usually, it takes several days or even weeks to become available. Thus, relating input and output parameters is often done retrospectively and statistically relevant only when several production batches are available. Of particular interest in this work is to evaluate the potential of 2D-fluorescence to quantify real-time several cell culture variables able to describe the physiological state of the culture. These, in turn, indicate far better whether production requirements will be met at the end of the process compared to relying only on the classical inline process variables.

For quantitative modeling, fluorescence measurements from all cell cultivations corresponding to offline reference analytics were included in the analysis. It is worth mentioning that these models rely on correlations with culture fluorophores which are in turn involved in cell growth and productivity (Ödman et al., 2009; Rowland-Jones, van den Berg, Racher, Martin, & Jaques, 2017).

6.3.3.1 Product concentration (PRO)

Product concentration was predicted based on positive correlations with several excitation/emission pairs in NAD(P)H fluorescence region (Figures 6-6A and 6-6B). Negative correlations with emission/excitation pairs in the region of tryptophan were identified when developing the model. However, the inclusion of these variables did not improve prediction performance. Figure 6-6 depicts the optimal PLS model built with 4 latent variables (Table 6-II) presenting very accurate predictions throughout the calibration range. For higher product concentration, the so called inner-filter effects likely contribute to the deviation between measured and predicted values (Figure 6-6C).

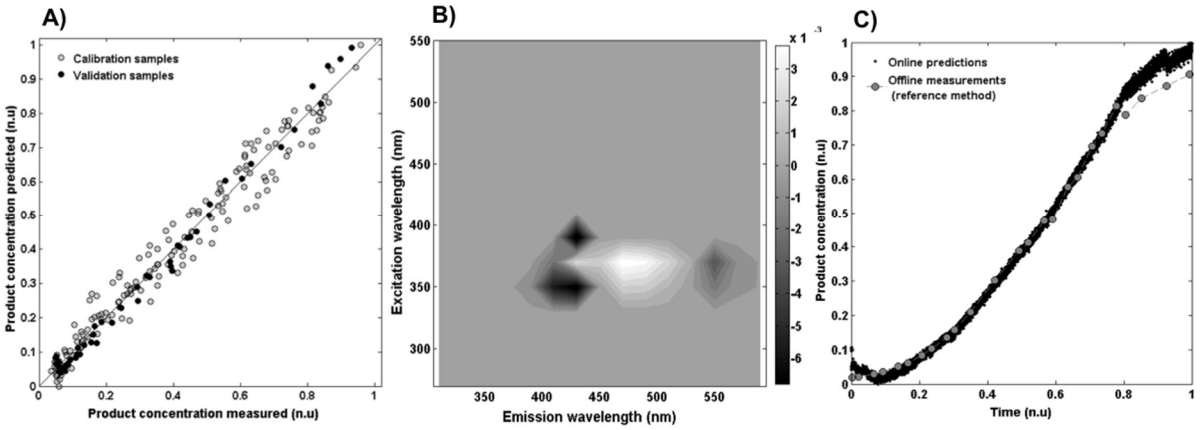


Figure 6–6: (**A**) Measured vs. PLS model predictions for PRO monitoring (normalized units); (**B**) regression coefficients indicating the contribution of emission/excitation pairs in the model; (**C**) Online predictions and PRO measurements by reference analytics (CC#06).

6.3.3.2 Lactate dehydrogenase (LDH)

Lactate dehydrogenase (LDH) is an oxidoreductase enzyme involved in growth, product biosynthesis and cell maintenance, that is present in a wide variety of organisms. LDH catalyzes the interconversion of pyruvate and lactate accompanied by the interconversion of NAD(P)H and NAD(P)+ (i.e., NAD or NADP as an acceptor). In agreement with the metabolic cell function of LDH, the spectral regions that most correlate with LDH profile are in the NAD(P)H and flavins region (Figure 6-7B). Even though LDH activity was indirectly predicted based on its stoichiometric relation with these biogenic fluorophores, for higher enzyme activities (> 0.8 cultivation time), the absorption phenomena could have been disturbed by chemical and/or physical interferences such as higher concentrations and cell densities (Figure 6-7C). A similar effect is seen for all other models.

Four latent variables were selected as the optimal rank for the PLS model (Table II), with a relative prediction error of 4.7% (Figure 6-7A). It is important to highlight that the model is valid only for the last 70% of total cultivation time since the sensitivity of the reference method in the lower concentration range is considerably low.

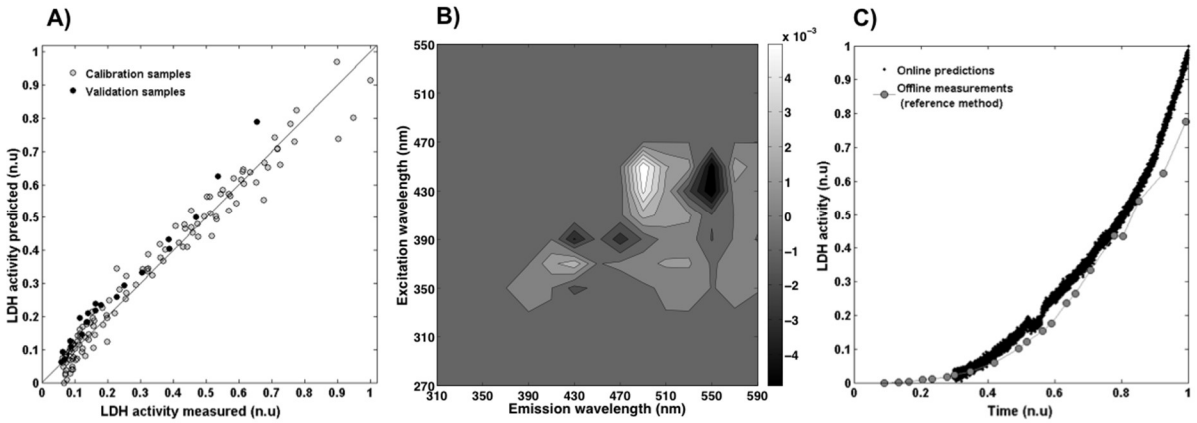


Figure 6–7: (**A**) Measured vs. PLS model predictions for LDH monitoring (normalized units); (**B**) regression coefficients indicating the contribution of emission/excitation pairs in the model; (**C**) online predictions and LDH measurements by reference analytics (CC#06).

6.3.3.3 Viable Cell Density (VCD)

Important parameters in biomanufacturing processes include specific growth rate and viable cell density, whose control is hampered by the difficulties in measuring accurately the concentration of biomass online (Dabros, Schuler, & Marison, 2010). The prediction of viable cell density (VCD) through 2D fluorescence was established on correlations with several fluorophores: *i)* tryptophan which is inversely correlated with VCD (i.e., cells consume amino acids for growth; *ii)* a positive correlation with NAD(P)H which increases continuously along culture time and); *iii)* flavins exhibiting a negative correlation most likely explained by the decrease in the flavins signal during cell exponential growth compensated afterwards by the feed media (which includes riboflavin in the formulation) and culture metabolic slowdown (Figure 6-8B). A model with three latent variables showed acceptable predictions throughout the calibration range (Figure 8-8A), enabling inline monitoring of viable cell density throughout cell cultivation time (Figure 6-8C).

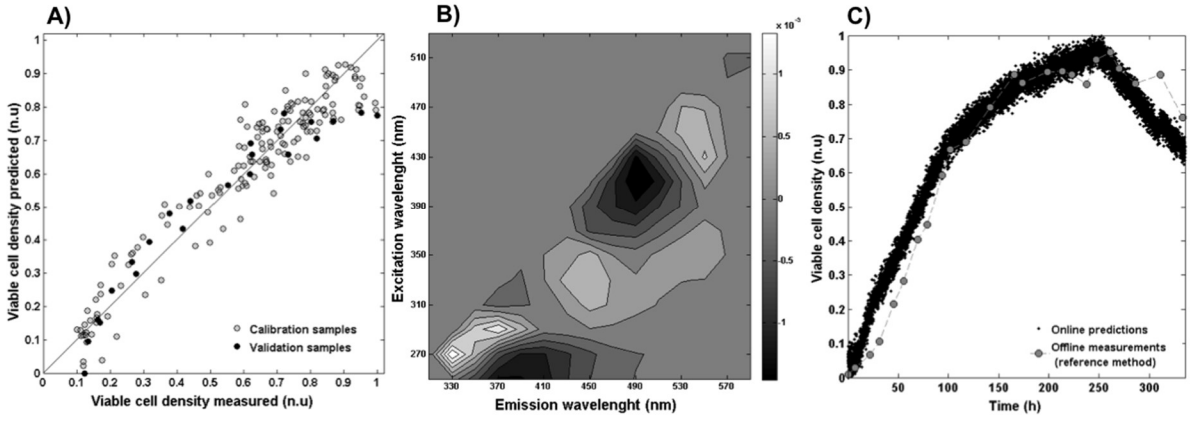


Figure 6–8: Fig. 8. (**A**) Measured vs. PLS model predictions for VCD monitoring (normalized units); (**B**) regression coefficients indicating the contribution of emission/excitation pairs in the model; (**C**) online predictions and VCD measurements by reference analytics (CC#06).

The ability to measure viable cell density inline offers the opportunity to monitor the cell growth in a timely manner, control the fermentation process by adjusting glucose feed rate towards optimal performance, and to benchmark process conditions of 'good' performing batches, such as feed media composition and feeding strategy (e.g., starting of feed media).

6.3.3.4 Lactate concentration (LAC)

Lactate production is a critical parameter in mammalian cells cultivations. Accumulation during initial cell growth is generally observed in CHO cultivations. In the present study, after glutamine depletion, lactate can follow one of two possible pathways: cells may switch to net lactate consumption or continue to produce lactate until glucose depletion. High lactate production can be related to a reduced oxidative metabolism or/and with media composition/feeding. A PLS model with five latent variables (Table II) enabled to predict lactate concentration with very good accuracy until the end of the exponential growth phase (i.e., until 0.25 total fermentation time) (Figure 6-9A). From this phase onwards, the degree of correlation between the fluorescent signal and lactate net concentration decreased considerably. Nevertheless, based on the fluorescence signal (Figure 6-9B) it was possible to distinguish different lactate profiles at an early process phase (Figure 6-9C), which is a very important outcome of the present study. The use of such model in early development can be very useful for feeding optimization purposes, accounting the metabolic behavior of cells.

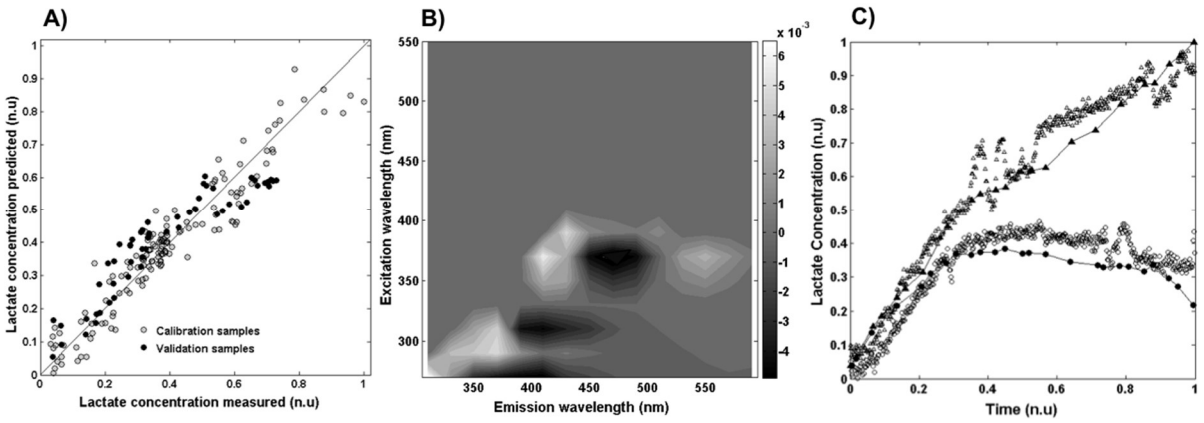


Figure 6–9: (**A**) Measured vs. PLS model predictions for LAC monitoring (normalized units); (**B**) regression coefficients indicating the contribution of emission/excitation pairs in the model; (**C**) Online predictions and LAC measurements by reference analytics (CC #03 (▲) and CC #06 (•)).

6.3.3.5 Glucose concentration (GLU)

Glucose does not exhibit intrinsic fluorescence, however, glucose consumption is related to the production of fluorescence molecules which enables to monitor glucose based on signals related to cell growth (Ödman et al., 2009; Jain et al., 2011). The emission/excitation pairs that are more relevant for glucose prediction in the current fermentation process are in NAD(P)H and flavins regions (Figure 6-10B). The model (Figure 6-10A and Table II) captured the process

trend until 0.25 total fermentation time (i.e., during exponential growth), presenting, however, a significant offset from this phase onwards and until the end of the cultivation (Figure 6-10 C). This offset can be explained by the start of the glucose feed which disrupted the correlation between analytical measurements and the fluorescence signal in the region of NAD(P)H and flavins. It is important to highlight that the later are included in the composition of the feed media.

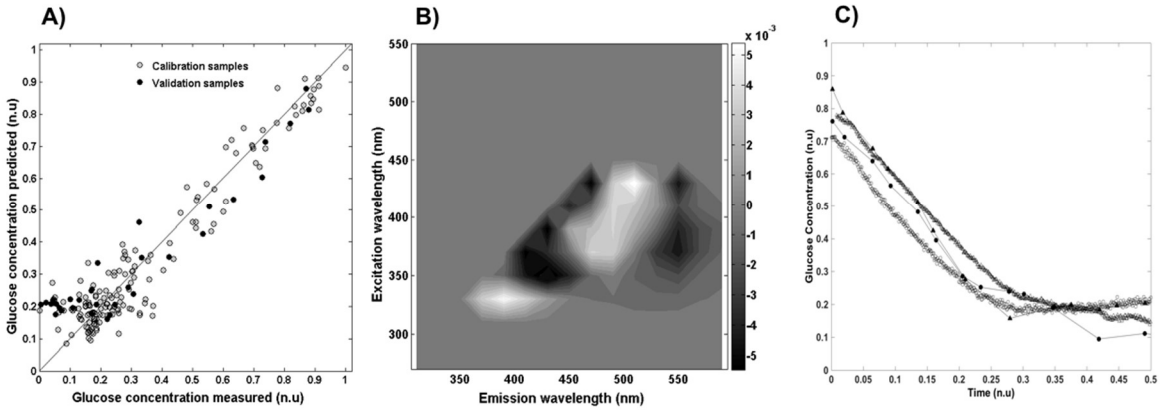


Figure 6–10: A) Measured vs PLS model predictions for GLU monitoring (normalized units); (**B**) regression coefficients indicating the contribution of emission/excitation pairs in the model; (**C**) Online predictions and GLU measurements by reference analytics (CC #03 (▲) and CC #06 (•)).

Table 6-2: Unfold-PLS correlations (R^2) between fluorescence and reference analytics measured in CHO cell cultivations. *LV* latent variables in the model, *RMSECV*, root mean square error of cross validation, *RMSEP*, root mean square error of prediction.

Process Parameter	Calibration Range	LV	RMSECV (n.u)	RMSEP (n.u.)	R^2
Product concentration	0.04 - 0.96	4	0.05	0.03	0.97
LDH activity	0.06-1.00	4	0.04	0.06	0.97
Viable Cell Density	0.09 – 0.96	3	0.07	0.09	0.92
Lactate concentration	0.03 – 1.00	5	0.06	0.08	0.91
Glucose concentration	0.00 – 1.00	5	0.08	0.11	0.89

6.3.4 Improvement Opportunities Based on 2D-fluorescence spectroscopy

In fed-batch processes, basal media composition supports the initial cell growth and production, while the feed stream accommodates metabolic requirements across different production phases (Le, 2012). Along with the feeding strategy (i.e., feed start, glucose plateau, fixed ratio vs. adaptive feeding) the above are the cornerstone for cell culture optimization during development and for establishing the control strategy to be implemented in routine manufacturing.

In the present study, 2D-fluorescence spectra combined with PARAFAC enabled a deeper characterization of media composition variation (i.e., prior to cell inoculation), and its potential impact on metabolic behavior and productivity. During routine production, PARAFAC scores for component 1 and 2 (tryptophan and flavins, respectively) could be used to supervise media composition and to target feed recipes to the ones used in better performing batches (CC#04 to CC#06). Additional experiments are however necessary to establish a thorough understanding about the “golden” media recipes. Nevertheless, implementing such approach would decrease batch-to-batch variations and enhance process understanding to support continuous improvement initiatives over the process lifecycle.

Additionally, it is important to maintain appropriate levels of nutrients and metabolites in the biosuspension over cultivation time. In the present study, monitoring PARAFAC scores for component 2 (flavins) allowed to immediately detect the overfeed, providing operators with a fault detection system to correct the feed stream for optimal performance.

The combination of 2D-fluorescence signal with PCA is another effective strategy to identify differences in cell culture metabolic state throughout the fermentation. Variations in the PCA scores trajectories can be used for fingerprint recognition (e.g., phase transition) and related to performance. Establishing a trigger for initiating the glucose feed based on the physiological state of the cultivation instead of a time-based approach, could be an option for an improved control strategy, integrating 2D-fluorescence spectroscopy.

Another key element for process optimization is to ensure a proper design of the feed medium, for optimal cell concentration and product titer at harvest. As opposed to the predefined feeding strategy used in this study, lactate production can be controlled through adaptive feeding by limiting the amount of carbon source (glucose or any other substrate) available (Konakovsky et al., 2016). 2D-fluorescence combined with PLS regression can take part of such control system, to supervise cell concentration real-time and adjust the glucose feed rate accordingly.

In addition to the elements presented above, the PLS models can be used for monitoring intensification of key performance indicators (such as lactate profiles) to enhance process understanding and identify what is critical and needs to be controlled in routine manufacturing.

6.3.5 Conclusions

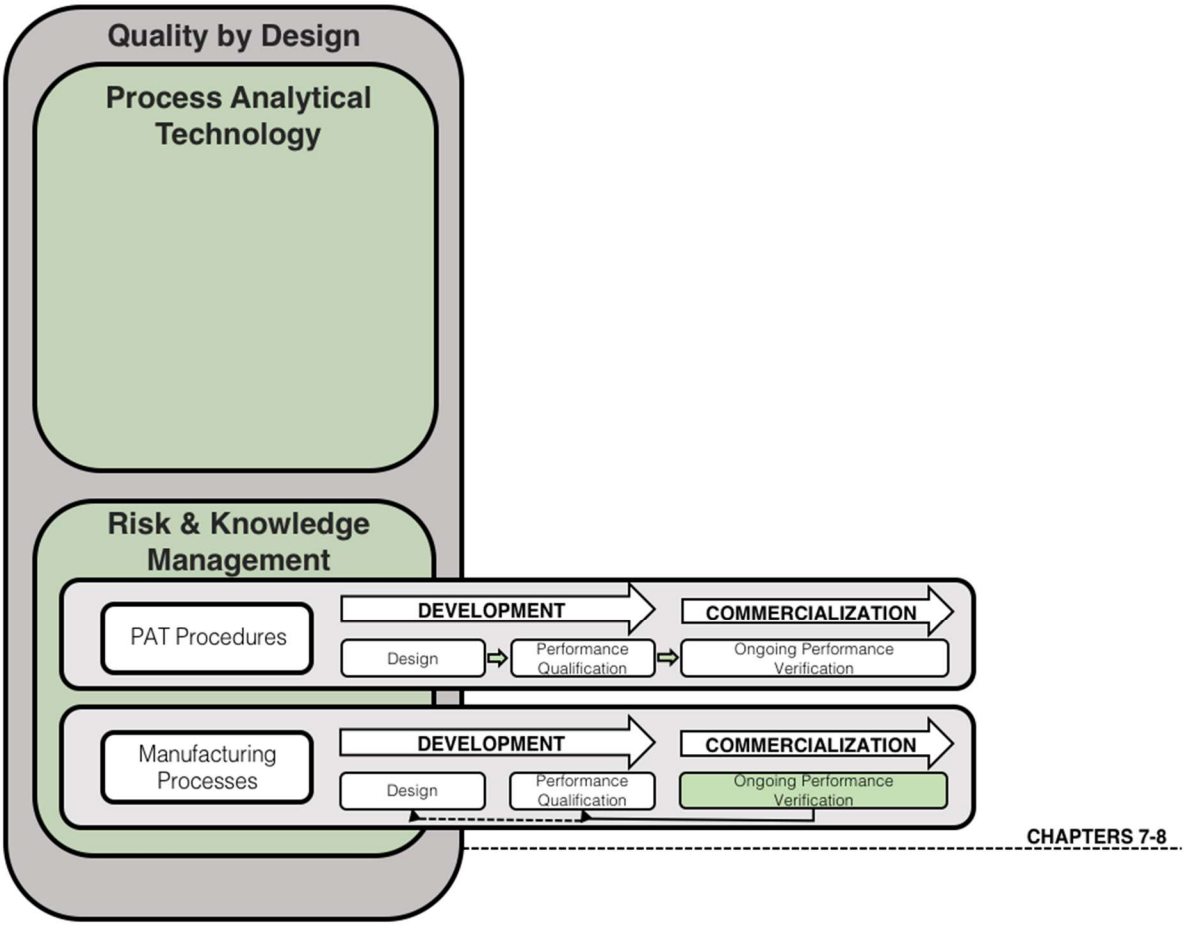
This work highlights the suitability of 2D - fluorescence as a true PAT tool in the context of bioprocess monitoring and optimization, enabling enhanced process understanding intertwined with supervisory and control goals. The present study demonstrated the capabilities of 2D-fluorescence spectroscopy combined with different chemometric methods to monitor real-time multiple parameters and to support science-based control strategies, as advocated in the QbD initiative. In addition, 2D-fluorescence can be used in bioprocess end-to-end assessments in the sense that the variations in starting materials can be identified and linked with further operations (e.g., feeding strategies, metabolic fingerprints) during the time course of bioprocesses, enabling mitigation actions to be implemented as feed-forward control initiatives.

Due to its unique features, 2D-fluorescence is a powerful tool for process monitoring and optimization, during development and potentially routine manufacturing of biotechnology products.

6.3.6 Acknowledgements

The authors are grateful for the support provided by 4Tune Engineering team, Elena Stocker, Late Stage Cell Culture Development Team and Josef Gabelsberger for the sponsorship by Roche Pharma Biotech Development Fermentation, Penzberg, Germany.

PART III - Scaling-up Knowledge Management in the Pharmaceutical Industry



7 Lifecycle management of PAT procedures: batch and continuous process applications

In 2011, FDA updated the existing guideline on manufacturing process validation to align modern quality concepts (i.e., Quality by Design) with the product lifecycle approach. Several authors have highlighted that many of the concepts and principles used to demonstrate continued product quality could be applied to guarantee ongoing quality of data produced by analytical procedures (Pohl et al., 2010; Nethercote & Ermer, 2012).

The QbD initiative applied to analytical method development is expected to foster Chemistry, Manufacturing, and Control (CMC) innovation while enabling the concept of "right analytics at right time" to support all stages of the manufacturing process lifecycle. Extending the lifecycle concept to analytical methods is essential to ensure its suitability during routine production. This is even more critical to process analytical technology (PAT) procedures that may also be part of the control strategy or used for real-time release testing (RTRT). In this chapter, a framework for PAT procedures lifecycle management will be discussed and illustrated with examples from industrial batch and continuous applications.

Keywords: PAT, method development & validation, lifecycle management, ongoing performance verification.

This chapter has been published in modified format as: **Gouveia F. F**. & Felizardo P. M. (2017). Lifecycle Management of PAT Procedures: Applications to Batch and Continuous Processes. Ed. Ferreira, Menezes, Tobyn. In Multivariate Analysis in the Pharmaceutical Industry, *Elsevier* (in press).

7.1 Introduction

The change in the process validation paradigm (US FDA, 2011) made pharmaceutical companies rethink analytical testing as to method development, validation, transfer and continued performance verification, to support all aspects of the product and process lifecycle (Nethercote & Ermer, 2014). Analytical testing and the manufacturing process are interdepending elements as changes in product profile may require modifications to existing analytical procedures. On the other hand, changes in analytical methods performance due to aging of components or replacement of parts during routine production may entail additional validation activities. Extending the lifecycle concept to analytical methods (particularly, to PAT procedures[1]) is imperative to holistically align analytical procedure variability with product requirements over lifecycle (Figure 7-1). This will broaden the scope for technological innovation and enhance the understanding of the sources of variability impacting product quality attributes, encompassing the ones arising from the manufacturing process as well as the analytical procedure (Martin et al., 2013).

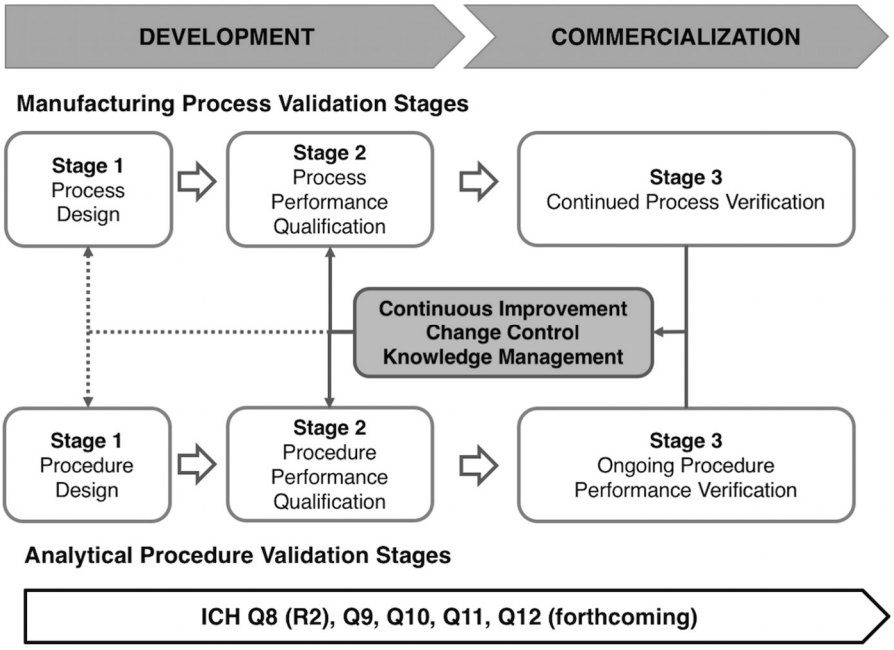


Figure 7–1: A common approach to manufacturing processes and analytical procedures lifecycle management.

[1] In this Chapter PAT Procedure is used in accordance with EMA Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations (EMA, 2014), i.e., it describes how the PAT infrastructure (defined as method in the same document) and model are used for the intended purpose, within the defined scope. The definition of procedure is also referenced in FDA Guidance for Development and submission of Near Infrared Analytical Procedures (US FDA, 2015).

As the pharmaceutical industry continues to overcome new manufacturing strategies with promising technical and economic benefits, such as single-use disposable technologies and adoption

of continuous processes, the challenges facing process control are permanently evolving. The ability to improve and optimize manufacturing processes is, in general, limited by the lack of reliable real-time information from all process stages. The adoption of new technologies and dynamic approaches able to handle process variations by means of changing the control offsets is therefore, necessary. Spectroscopy-based PAT have become widely used analytics so given its capabilities to measure both physical and chemical properties directly at the process streams, either by probing systems (in-line) or via sampling devices (on-line), enabling to reach the aims of process understanding, process control and manufacturing flexibility to consistently generate products of predetermined quality (Read et al., 2010).

There are challenges, however in adopting PAT in pharmaceutical environments relatively to conventional lab-based analytical methods, due to the twofold nature of PAT procedures: the PAT infrastructure integrating the equipment, the sampling interface and the software for data acquisition and modeling as well as connectivities to the control system, and the PAT model usually requiring a chemometric approach for representative sample selection, model calibration and validation of the entire setup as "fit for its intended purpose".

To fulfil the regulatory gap as to PAT procedures development and validation requirements, and to foster the adoption of PAT by the pharmaceutical industry, several publications by regulatory authorities and scientific groups have been recently issued. In 2011, the American Society for Testing and Materials (ASTM) International published a "Standard guide for verification of PAT enabled control systems" establishing the principles and verification activities necessary to ensure that PAT-enabled control systems are fit for purpose, properly implemented and perform as expected in routine use (ASTM, 2011). The document details on the use of process models based on first principles understanding or empirical models derived from experimental investigations applied to batch and continuous processing. In the same year, ICH released the Guide for ICH Q8/Q9/Q10 (US FDA, 2011) implementation describing the role of mathematical modeling in the QbD framework including models for process design (e.g. formulation optimization, design space determination and scale-up), analytical characterization (e.g., PAT-based models), process monitoring and process control (e.g., MSPC). The implementation working group recommends to perform model categorization according to the impact of the reported result to product quality. The extent of verification, validation and proper documentation of model-related information is dependent on such categorization. Later, the 2014 EMA "Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requirements for new submissions and variations" described the requirements for the development, calibration and validation of NIRS-based methods, both for qualitative and quantitative PAT applications (EMA, 2014a). Few months after, an addendum to the guideline (EMA, 2014b) was published to clarify the scope of application and to provide guidance on change management over NIRS-based applications lifecycle use. The principles and concepts illustrated in the guideline can be extrapolated to other spectroscopy based methods, such as FT-IR and Raman, widely used in PAT applications. Recently, the European Directorate for the Quality of Medicines and

Healthcare (EDQM) issued a new chapter in the European Pharmacopeia entitled “Chemometric methods applied to analytical data (5.21)” (EP, 2016). For the first time, the use of chemometrics is referenced in a Pharmacopeia, providing guidance on good chemometric practices and encouraging the use of multivariate data analysis methods for evaluating data generated by spectroscopy and chromatography-mass spectrometry methods (e.g., LC-MS) as integral components of PAT applications. The use of quality risk management (QRM) tools and principles to select critical parameters and attributes to be measured, design of experiments (DoE), multivariate data analysis (MVDA), modeling approaches and statistical process control (SPC), are all recommended to be used in combination so that “scientific evidence” of analytical procedures consistency and performance is obtained throughout its lifecycle use.

In this context, this contribution encompasses a systematic workflow for lifecycle management of PAT procedures, fully aligned with the QbD initiative as well as with validation requirements established in ICH Q2 (ICH, 2005), EMA NIRS guideline (EMA, 2014a; EMA, 2014b), European and United States Pharmacopoeias (EP, 2016; USP 38, n.d.). The approach comprises a series of activities taking place over the lifecycle of the PAT procedure to guarantee ongoing assurance that reportable results are in a state of control over its life use, even if there are changes in materials, equipment, modeling approach, production environment or personnel. Special focus will be given to ongoing performance verification strategies of PAT procedures, used in batch and continuous processes applications.

7.2 A Three Stage Approach to PAT Procedures Development and Lifecycle Management

The domain of PAT is the manufacturing process itself (Felizardo et al., 2012). Performance requirements of PAT applications should be defined within the context of the control strategy for the product. As such, it is not possible to rely on the quality of reportable results of PAT procedures if its conception is not fully understood or if its lifecycle is not properly managed.

The three-stage approach depicted in Figure 7-2 enables the necessary alignment between PAT procedures and manufacturing requirements to assure that quality commitments are met over the entire product lifecycle. The required activities that are inherent to procedure design (Stage 1), performance qualification (Stage 2) ongoing performance verification (Stage 3), including change control, monitoring programs and re-validation initiatives will be further detailed in the next sections.

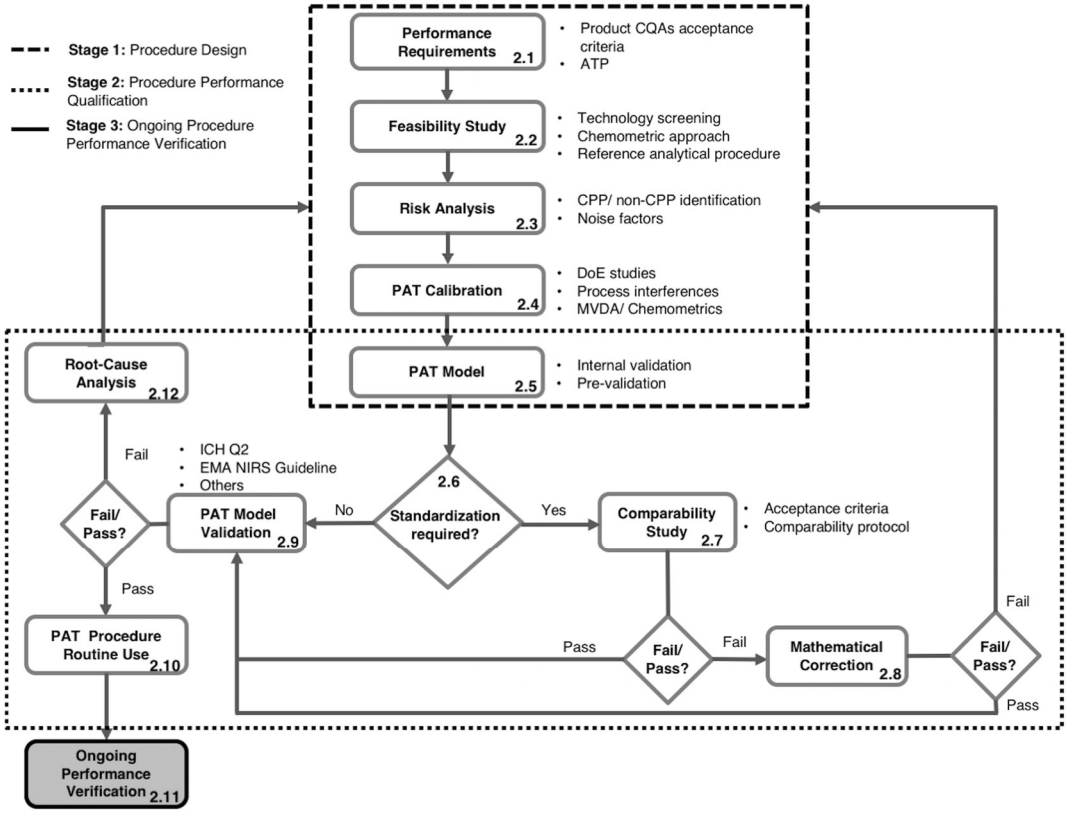


Figure 7–2: General workflow for lifecycle management of PAT procedures.

7.2.1 Stage 1 – PAT Procedure Design

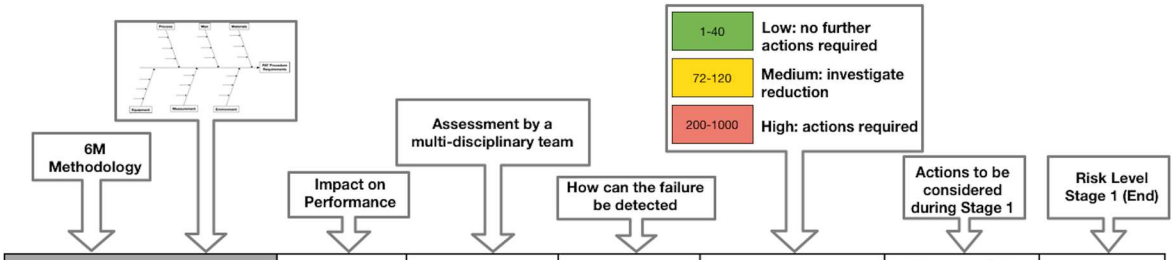
The main goal of this stage is to concentrate efforts in defining the PAT procedure in enough detail to consistently control critical parameters and sources of variability in the manufacturing scale which may undermine the quality of the reportable data.

In agreement with the QbD construction described in ICH Q8, defining the quality target product profile (QTPP) is the starting point for process development. Analogously, establishing performance requirements (viz., Analytical Target profile – ATP) for the PAT procedure is the first step for an integrated lifecycle approach (Weitzel, 2014). Such requirements (Figure 7-2 – 2.1) relate to the specifications for a given product quality attribute and their expected variation during the manufacturing process lifecycle (viz., acceptance criteria that are stage-specific and can be revised over time). The ATP defines the objective of the test and quality requirements, specifically the expected level of confidence of the reportable result which is the driver for the selection of the PAT procedure (i.e., PAT infrastructure and PAT model).

Feasibility studies (Figure 7-2 – 2.2) are important in the development process for technology screening (i.e., each spectroscopy method holds advantages and limitations, and the choice depends upon the specific application), selection of the chemometric approach (i.e., model-free,

qualitative or quantitative) and reference analytical method, when applicable. Having a broad understanding at an early stage of the suitability of the PAT procedure is extremely important to establish an optimal PAT business case, since the following steps will require considerable resource allocation and experimental planning (Gouveia et al., 2016)

Another key aspect is the risk assessment component (Figure 7-2 – 2.3) which consists in the identification, assessment and ranking of parameters that may adversely impact the suitability of the PAT procedure to fit the intended purpose (Figure 7-3). Risk assessments are iterative throughout the lifecycle of the method, they should take into consideration all steps from development and validation to method transfer and on-going performance verification during routine use. The critical re-appraisal and re-evaluation of the PAT procedure on a regular basis will support continuous improvement and appropriate change control when necessary. In addition, the risk assessment can guide experimentation to de-risk the method and identify critical parameters having the greatest effect on performance (Figure 7-2 – 2.4). The use of testing and design of experiments (DoE) can be leveraged to yield important method understanding that ultimately leads to a robust control strategy and reduced variability. Simultaneously, appropriate DoE approaches are recommended to efficiently build a robust calibration ensuring that expected variation involving the parameters identified during the risk assessment will be covered in the PAT calibration model (Figure 7-2 – 2.5) (Schaefer et al., 2014).



FMEA		Severity		Occurrence		Detectability		Risk Evaluation		Recommended Actions	Revised		
Risk ID	Potential failure mode	Failure effect	S	Failure cause	O	Current controls	D	RPN	Risk evaluation (CNX method)	Actions required	O	D	RPN
Operator (MAN)	Incorrect handling of PAT infrastructure or model	Reported results not reliable	10	Operator not trained	2	Check-list, SOP	2	40	Low (controlled)	No further actions required	-	-	40
Probe (MACHINE)	Probe position sub-optimal	Impact in reportable results	6	Incorrect probe setup	6	Probe verification tests (comparison to initial tested scans)	2	72	Medium (noise)	Probe position to be tested during method development as a ruggedness factor	2	2	40
Analyte concentration (MATERIALS)	Spectra not representative of process variability	Reported results not reliable	10	Concentration range not properly represented in the calibration	10	SQT, check process specifications (to be implemented)	6	600	High (experimental)	Parameter to be tested during model development (DoE)	6	2	120
PQ tests (METHOD)	Equipment performance not tested	Reported results not reliable	10	Operator not trained	2	Check-list, SOP	2	40	Low (controlled)	No further actions required	-	-	40
Temperature (MOTHER NATURE)	Variations in room temperature	Impact on PAT accuracy	6	Failure in room temperature control system	2	Check-list Maintenance plan	2	24	Low (controlled)	No further actions required	-	-	24
Temperature (MEASUREMENT)	Changes in spectral properties (e.g., wavelength shifts)	Impact on PAT accuracy	6	Process NOR/PAR (30-37±2°C)	6	Manufacturing controls	2	72	Low (noise)	Temperature effect to be tested during procedure development	2	2	24

Figure 7–3: Risk assessment exercise for a PAT procedure development (Stage 1) based on Failure Mode and Effect Analysis (FMEA). An example of a specific failure mode is presented for each 6M category (man, machine, material, method, mother-nature and measurement). Risk evaluation based on CNX methodology (control, noise, experimental) (Martin et al., 2013).

As an outcome of Stage 1, the comprehensive description of the PAT procedure should contain the following elements (EMA, 2014a):

- Purpose of the PAT procedure within the context of the control strategy including the location in the process flow, matrix composition, the intended use (e.g., material identification, in-process control, end-point release testing) and the mode of measurement (e.g., in-line, on-line, at-line, off-line);
- Full description of the PAT infrastructure, referring to the equipment, sampling interface, probe setup, sampling protocol and connectivities to the control system;
- Step description of the PAT data acquisition process including details about e.g., background acquisition conditions, frequency and storage, sample presentation and conditioning (when required);

- Detailed description of the PAT model including the software package and the chemometric algorithm used. Specific parameters such as pre-processing, number of latent variables in the model, cross-validation method and, the required justifications for choosing such criteria (e.g. rationale for preprocessing, and wavelength selection) should be detailed and justified;
- Description of the reference method (when applicable);
- An initial plan for ongoing process verification, including quality checks during routine operation (e.g. equipment performance verification and spectral quality tests).

The above-mentioned information must be gathered during the PAT procedure design stage, documented in standard operating procedures (SOPs) or equivalent documents, and managed under the company's pharmaceutical quality system (PQS). For illustration purposes, a summary of relevant information is depicted in Table 7-1.

Table 7–1: Comprehensive characterization of the PAT procedure, including requirements, description of the PAT infrastructure and PAT model.

PAT Procedure Requirements	
The purpose of the in-line near NIR procedure is to quantify A in the presence of B and C over a range of 75 to 125% of the nominal concentration for determination of the reaction end-point (in-process-control) in the process flow step XXX. The decision criterion is based on five consecutive measurements at target concentration of X %(m/m) for A, max. Y % residual concentration of B and max. Z % (m/m) for C. The reportable result must fall within ± 10 % of the true value with at least 90% probability determined with 95% confidence.	
PAT Infrastructure	
Parameter	**Description**
Instrument	ABB FTPA2000-260
Software	ABB FTSW100 (CFR-P11 compliant)
Communication link to Control System	OPC XML DA
Light dispersion principle of the optical system	FT-NIR
Detector type	InGaAs detector
Mode	Transmission
Wavelength range	3800 cm^{-1} to 14 000 cm^{-1}
Resolution and scans	8 cm^{-1} resolution; 64 scans
Sample preparation/ presentation/ sampling device	On-line measurements Insertion probe (fiber optics)
Sample population	-Calibration set: lab scale (according to DoE plan) + 2 pilot batches -Calibration test set (pre-validation): 1 pilot batch -Validation set: 5 manufacturing batches

PAT Model	
Parameter	**Description**
Software	Matlab Version 8.1 (MathWorks, Natick, MA) PLS Toolbox Version 7.8 (Eigenvector Research Inc., USA)
Spectral pre-processing	Standard Normal Variate (SNV) Savitsky-Golay smoothing (15 points) + 1st Derivative (2nd order polynomial, 15 points window) Mean center
Spectral quality check statistics	Q-residuals and Hotellings T^2
Spectral quality check statistics threshold	Confidence limits: 95 % ($\mu \pm 2\sigma$)
Chemometric algorithm	PLS
PLS model parameters	PLS spectral range: 5978-6164 cm-1; 8231-8671cm-1 Number of latent variables: 3 Cross-validation method: Venetian blinds (15 data splits)
Statistical attributes	Standard Error of Calibration (SEC) Standard Error of Cross-Validation (SECV) Standard Error of Prediction (SEP) Bias, Slope, Intercept SEP/Standard Error of Laboratory (SEL)
Reference method	HPLC method with UV detection (description of the analytical procedure according to Module 3.2.P.5.2)
Method validation protocol	Specificity, Linearity, Range, Accuracy, Precision, Robustness Detection and Quantification limits (for impurities, only)

7.2.2 Stage 2 – PAT Performance Qualification

The demonstration that PAT procedures are fit for their intended purpose before use, involves a considerable amount of effort, time and resource allocation. The sources of changes that occur during the lifecycle of the PAT procedure can be classified as sample-related or equipment related. The first category refers to physical or chemical variations of the sample presented to the measurement system due to changes in the manufacturing process or sampling system (e.g. changes in viscosity, particle size distribution, chemical composition, drift of the manufacturing process to a new steady state, sampling location/ frequency). The second case is verified when the measurement system response function has changed because of (non-exhaustive) instrument ageing (e.g. light source, probing system, detector), repairs and maintenance activities, equipment changes or shifts in environmental conditions (e.g., temperature, humidity).

When such changes are expected (e.g., instrument, probe type or path length, scale of the manufacturing process), strategies for model standardization must be used before validation of the PAT procedure according to ICH Q2 requirements. In many situations, standardization issues can be significantly minimized through proper method design or simple mathematical correction. PAT model robustness can be enhanced by including in the PAT calibration, samples

acquired with different instruments, probing systems or other sources of variability with potential impact in the quality of the data produced by the PAT procedure, *e.g.*, probe position and bending, temperature – designated ruggedness factors. Another strategy consists in selecting spectral variables and preprocessing methods that minimize the impact of sample or equipment-related changes in the PAT procedure performance.

In several occasions, the transfer process is straightforward and the performance of the transferred PAT model is found to be satisfactory. The evidences should be gathered under a comparability protocol where the acceptance criteria for the model transfer are clearly indicated (e.g., chemometric parameters as statistical performance indicators) (Figure 7-2-2.7). Although similar to the validation exercise, this step does not need to fit entirely with the method validation requirements found in ICH Q2 document.

Under certain circumstances, mathematical correction is not enough to effectively transfer the PAT model (Figure 7-2-2.8), and more complex standardization or calibration transfer methods must be applied to guarantee the necessary performance of the PAT procedure. Most of these methodologies aim to reproduce the response of the measurement system during the PAT calibration phase in the PAT infrastructure to be used in the manufacturing process. The first step consists in measuring a set of representative samples in both PAT infrastructures referred to as "standardization samples" and then computing standardization parameters to correct for instrumental differences. The most common standardization methods can be included in three different categories: *i)* standardization of the predicted values (DiFoggio, 1995) consisting in post-processing the predictions with a slope and bias adjustment; *ii)* standardization of spectral responses, establishing a transfer function between the two instruments. The three most common methods for spectral correction are the direct standardization (DS), the piecewise direct standardization (PDS), and the Shenk and Westerhaus method (SW) (Wang et al., 1991; Wang et al., 1992); *iii)* standardization of the model coefficients involving transferring a regression equation between different instruments by means of a two-step PLS approach (Galvão et al., 2015; Setarehdan et al., 2002).

When the measurement of standardization samples on both PAT infrastructures is not an option (e.g. calibration samples are no longer available, sampling is not feasible) the use of dedicated pre-processing methods can compensate for spectral differences between equipment while preserving the common features (Wise & Roginski, 2015). The list of available preprocessing methods is extensive, and typically includes Orthogonal Signal Correction (OSC) and Generalized Least Squares (GLS). To guarantee successful transfer of the PAT procedure, it is recommended to perform a pre-validation test using an external set of samples, consisting in a collection of data not previously included in the calibration model (Figure 7-2-2.7). The performance requirements are similar to the ones considered in the PAT model development phase (Figure 7-2-2.5) to demonstrate the quality of the data provided by the transferred PAT procedure.

Whenever the standardization approach requires expanding the calibration set with new samples, the PAT procedure calibration phase must be initiated and all documentation re-evaluated, including risk-assessments, SOPs and development reports.

Finally, to assess the validity of the PAT procedure as part of the control system of the manufacturing process, a formal validation protocol (Figure 7-2-2.9) following regulatory requirements must be in place. Traditional chemometric performance measures may not be sufficient to evidence that the reportable results provided by the PAT procedure are of adequate quality and reliability during routine use of the method (De Bleye et al., 2012). Several approaches for validation of PAT procedures can be found in the literature (Schaefer et al., 2014; Feng & Hu, 2006; Bodson et al., 2007). Depending on the pharmaceutical application (e.g., qualitative vs. quantitative), the validation of the PAT procedure may include all or part of the criteria found in ICH Q2. If the reportable results are considered reliable and provide confidence that the product has the required quality, the method can be used in routine (Figure 7-2-2.10). An ongoing performance verification plan should be established at this point (Figure 7-2-2.11). Otherwise, a root-cause investigation must be performed to identify the causes for failing the validation exercise and ultimately, a full re-development of the PAT procedure must be undertaken (Figure 7-2-2.12).

7.2.3 Stage 3 – PAT Ongoing Performance Verification (OPV)

PAT procedures used in routine production of pharmaceuticals are critical elements of the overall quality system. To guarantee that the data generated by the PAT procedure is fit for purpose during its lifecycle use, it is essential to have systems in place to monitor the performance and for detecting and addressing unplanned departures from the designed procedure (US FDA, 2011). Ongoing performance verification (OPV) of a PAT procedure consists in collecting and analyzing data that is related to procedure performance during routine use. The OPV strategy is a science and risk based approach, used to evaluate if the PAT infrastructure (equipment, sampling device and data acquisition software) and, when applicable, the PAT model continues to operate according to the specified requirements for the intended use. This strategy includes the following steps:

- Performance Qualification (PQ) tests, or performance checks as commonly designated by equipment vendors;
- Spectral Quality Tests (SQTs);
- Investigation and handling of outlier measurements and out-of-specification (OOS) results during routine use;
- PAT procedure performance monitoring by parallel testing (periodic review).

An optimal PAT procedure design and development should anticipate significant sources of variability and establish appropriate detection, control and mitigation actions. Whenever the performance of the method is considered unacceptable, an investigation to determine the possible

root causes must be carried out which can lead to model-update and re-validation of the PAT procedure, if necessary.

Performance Qualification (PQ) tests

The first requisite for ongoing monitoring of procedure performance is the routine conduction of the PAT infrastructure performance qualification/verification tests, to demonstrate that the equipment is performing within specifications during routine use.

Equipment qualification (Stage 2) is the output of several activities that can be grouped into four phases: design qualification (DQ), installation qualification (IQ), operational qualification (OQ), and performance qualification (PQ). The purpose of equipment qualification is to demonstrate, through documented evidence, that the equipment is suitable for its intended use, and all requirements are applicable to the PAT infrastructure. The use of a qualified instrument in analyses contributes to maximize the confidence in the validity of the generated data. A comprehensive description of PAT equipment qualification is out of the scope of the present contribution but can be found elsewhere (USP 1119, n.d; Harrington, 2010).

Periodic tests for both routine performance verification and after preventive maintenance and repairs (e.g., replacement due to aging of instrument components, deterioration of equipment parts, trends or drifts in PQ tests) are key elements of an OPV strategy (Figure 7-4). A suite of procedures to ensure that the equipment is performing adequately before and during ongoing production should be established. Wavelength accuracy and repeatability checks, response repeatability, photometric linearity, signal-to-noise ratio and baseline stability are examples of most common PQ tests (Harrington, 2010). Statistical analysis techniques are a useful toolkit to trend equipment performance over time so that any change in performance can be monitored. In addition, changes to the PAT infrastructure (e.g., detector, lamp) can be assessed against historical data which is a key element of the PAT procedure lifecycle management strategy. Any changes to the system hardware (i.e. spectrometer and computer system) arising from either maintenance or modifications should be reviewed against the original IQ/OQ/PQ criteria. Appropriate action and testing should be completed to ensure that the instrumentation operates in an equivalent or improved manner. The frequency and protocol to be followed for equipment qualification should be documented and justified in specific SOPs, as well as specifications and acceptance criteria. All applicable documentation obtained during instrument qualification should be adequately managed, e.g., through PQS.

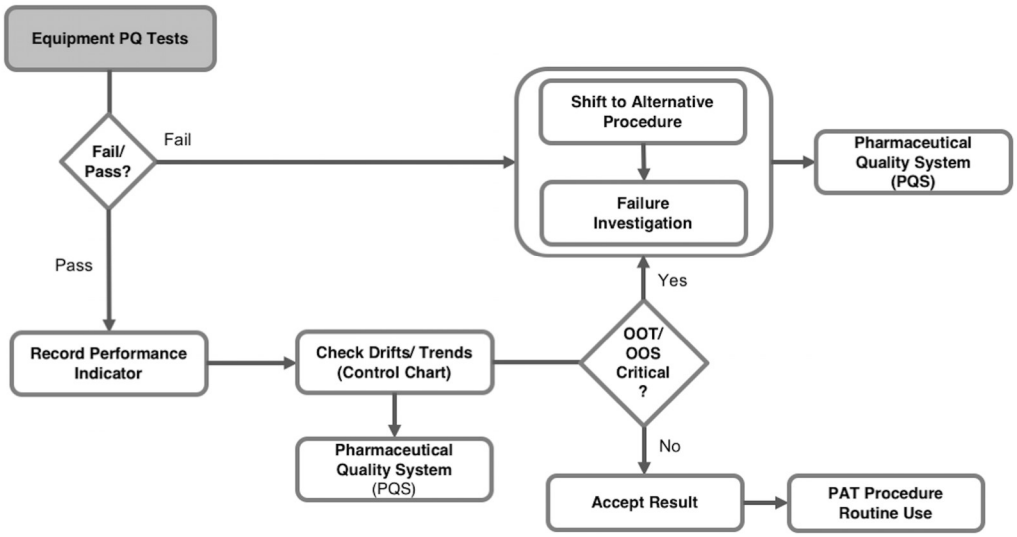


Figure 7–4: Workflow for ongoing performance verification of the PAT infrastructure performance in routine use.

Spectral Quality Test (SQT)

Analogously to the PAT procedure ongoing performance monitoring, the PAT model is expected to undergo the same kind of OPV strategy. Spectral quality tests aim to determine whether the characteristics of the samples presented to the PAT procedure fall within the range of variation for which the model was calibrated and validated. A very comprehensive roadmap for multivariate model maintenance is proposed by Wise & Roginski, 2015. Model diagnostic measures such as residuals (measurement of the orthogonal difference between a sample and the modeled data) and leverages (measurement of how far a sample is from the center of the data set, frequently in a type of weighted form) are valuable model performance indicators. These statistics are implemented in common chemometric software packages and may be named differently, e.g., Q residuals or DModX and T2 or Hotelling's T2, respectively. Samples with high residuals usually indicate new sources of variation not present during the calibration and validation phases, while high leverages signal that the concentration ranges are being extrapolated or there is an unusual concentration ratio of components in the sample.

The definition of the criteria depends upon the nature and intended purpose of the PAT procedure, and should be established based of knowledge and data-driven principles. Setting hard statistical limits a priori may lead to a significant number of false alarms and, consequently, unnecessary shutdowns for investigation of potential process OOT/OOS or analytical outliers. A timeframe should be established for assessment of the spectral uncertainty during initial use of the PAT procedure for definition of appropriate SQT criteria. The criteria should also include the number of consecutive SQT failures allowed before triggering an outlier investigation (depending on system dynamics, e.g., three consecutive samples failing SQT). An automatic outlier alert system can be defined and implemented to generate alarms and/or warning signals when

a pre-defined number of new spectral samples do not satisfy the acceptance criteria defined for the SQT. As with PQ tests, SPC rules can be applied to monitor model residuals and leverages variability during the PAT procedure routine use.

Investigation and handling of spectral outliers and out-of-specification (OOS) results during routine use

The terms out-of-specification (OOS) and out-of-trend (OOT) are classical designations related to deviating analytical measurements. An OOS result is generated when a reportable result, collected at a single point in time exceeds a predetermined specification. The OOT result on the other hand, can assume several entities such as a drift resulting from a gradual upward or downward in the data or a sudden change in the average value (Harrington, 2010). In general, OOT results account for historical data and have a very relevant practical significance from a regulatory and business standpoint. The concept of OOS and OOT results can be applied during routine use of the PAT procedure but must be handled differently because the root-cause may or may be not related to the PAT procedure performance (e.g., an OOS result due to a manufacturing process deviation) (Figure 7-5).

Spectral outliers designate PAT procedure results that do not encompass the experimental variability observed during development and validation phases, which does not necessarily indicate an OOS result. These samples may be representative of acceptable material but new sources of variability arising from changes in the process or materials are not satisfactorily described by the PAT model. In such cases, reportable results are not reliable and additional investigations are required to identify the root cause. Whenever the root cause is not identified in the initial assessment, the sample should be tested using the analytical reference method or any appropriate alternative procedure. After confirmation of authenticity, the sample may be included in the spectral database and the model should be re-calibrated and re-validated to include this source of variation. The possibility of including additional similar samples should be considered, if needed. Alternatively, it can be concluded that the SQT criteria should be reviewed, to reduce the number of false alarms. This means the PAT procedure is fit for purpose, reliable and further updates to the PAT model are not necessary.

In case the collection of samples for authentication is not feasible, rejection or acceptance of the PAT procedure results should be evaluated based on risk assessment and prior knowledge gained during method development and validation, as well as based on evidences gathered during the investigation. While the spectral outlier is under investigation, the protocol to change to an alternative procedure (reference method or non-PAT method) should be clearly defined, documented and justified, as part of the control system.

Whenever the PAT procedure is found to perform adequately for its intended purpose, the OOS reportable result should be managed according to the PQS of the company. Rejection/acceptance of the product should be based on the outcome of the failure investigation (which is

not related with the PAT procedure performance), and may or may not include analysis by the analytical reference method.

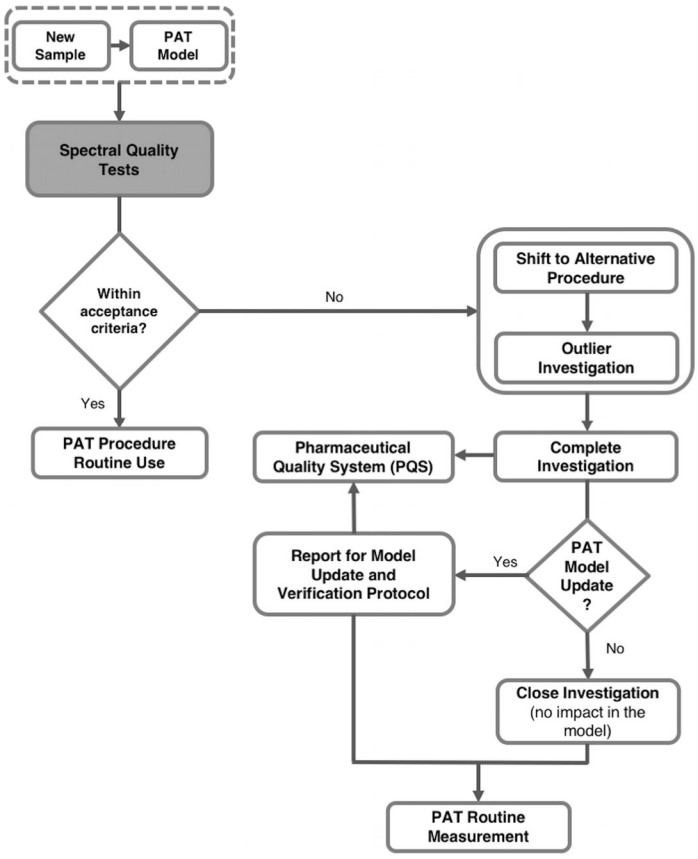


Figure 7–5: Decision tree to manage Spectral Quality Tests deviations and implementation mitigation actions.

PAT procedure performance monitoring by parallel testing

Parallel testing consists in comparing the PAT model predictions (e.g., from ongoing production) with the correspondent measured values by the reference analytical method (when available), and is a critical stage in the PAT procedure ongoing process verification strategy. The integration of parallel testing (Figure 7-6) in the ongoing process verification strategy proposed in the current approach provides higher assurance as to the PAT procedure lifecycle validity as is considered best practice. The criteria and methods used to monitor the prediction accuracy of the model should be documented and justified (e.g. Student t-test). Statistical Process Control (SPC) charts combined with a set of rules (e.g., Western Electric 8) can be implemented to monitor reportable results (e.g. RMSEP) throughout the lifecycle. The testing frequency should be defined based on number of lots or at regular time intervals, depending of the existing knowledge baseline. The recurrence of the testing should be reviewed as part of the lifecycle management plan and must account for changes in the manufacturing process and changes in the PAT procedure, both infrastructure and model-related.

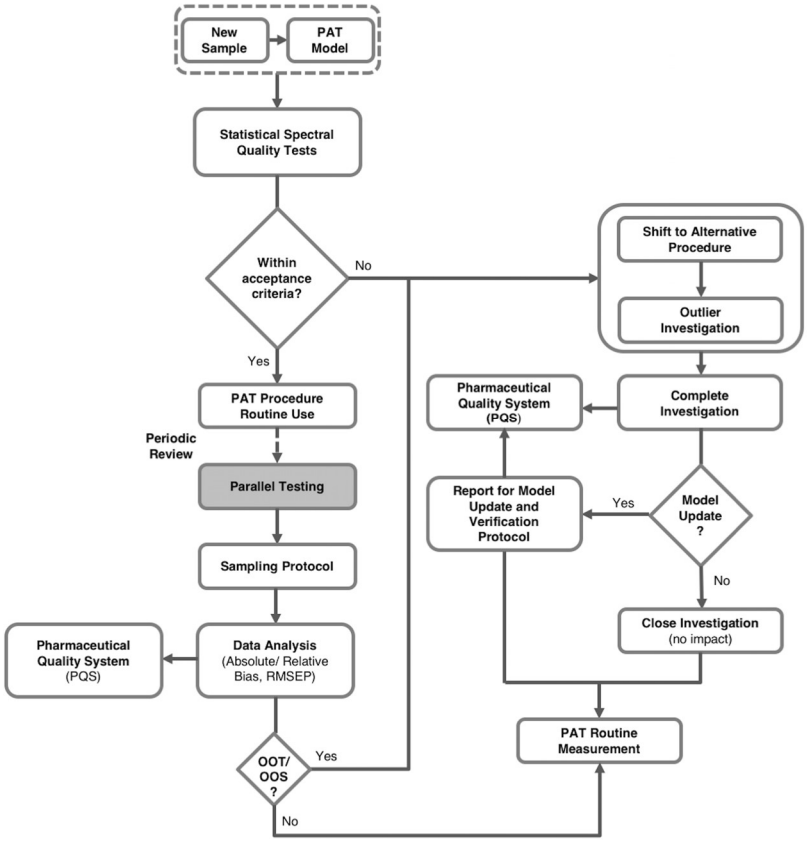


Figure 7–6: Integration of parallel testing within the PAT procedure ongoing process verification workflow.

7.3 Ongoing process verification (OPV) of PAT procedures: examples from batch and continuous processes

The underlying goal of PAT is to raise process understanding up to a level that the final product consistently conforms to initially planned quality standards. The benefits of PAT can be fully realized both in batch and continuous processing. Systems that benefit the most from PAT-based approaches are the ones dealing with transient/unstable intermediates, critical endpoints, sampling constraints or those where first-order understanding is not fully elucidated (e.g., several biopharmaceutical processes) (Myerson et al., 2015).

PAT procedures are valuable tools in bioprocessing monitoring and control, since they can be a critical element in the control strategy of quality and performance parameters through manipulation of input variables (e.g., air flow, agitation rate, temperature, and subtract, acid and base feeds) based on the current state of the process (e.g., pH, pO2, temperature, biomass, product, substrates and key metabolites concentrations). The use of PAT-based procedures as key elements of the process control strategy demand robust and reliable design of the control system, demonstration that the system is fit for its intended purpose and that this state is kept over the lifecycle of the manufacturing process.

In the current example, the control approach for a fed-batch cell cultivation process consisted in measuring real-time the viable cell density (VCD) to adjust the substrate feed rate towards an optimal growth profile. The development of the current control strategy was investigated in the same experimental setup described in **PAPER II**. The control system integrates the PAT infrastructure (Incyte DN12-220, Hamilton Co., Switzerland) for spectral acquisition, the PAT model that transforms the measured spectra into cell concentration values, the controller which compares the measured signal with a pre-determined control set point and finally, the actuator controlling the amount of feed media to be supplied to the fermentation unit (Figure 7-7). As the PAT procedure is a critical part of the control system, the spectra signal quality is checked routinely prior sending the predictions to the controller, through implementation of spectral quality tests (Q-residuals and Hotelling T2). Warning and action alarms are displayed whenever a predefined threshold value is surpassed, or when trends over time are identified. Data from these deviations is automatically recorded in a database for review of the control system performance and root-cause investigation. Finally, whenever the criteria for quality checks is not fulfilled, the control shifts to and alternative method based on manual adjustment of the feed stream. The calculations are, in this case, based on samples measured by reference analytics.

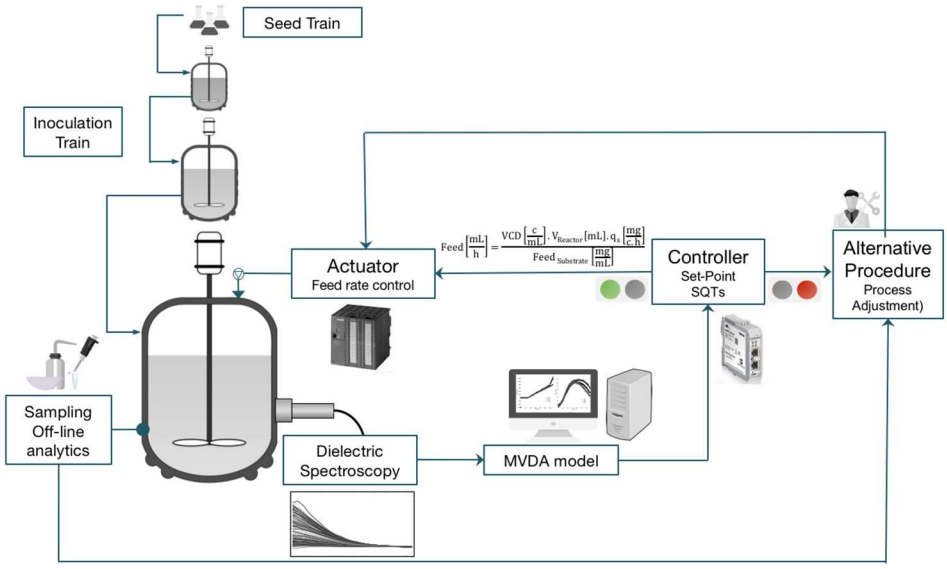


Figure 7–7: Control strategy of the feed-batch cell cultivation process using dielectric spectroscopy and a multivariate quantitative model.

Since a quantitative model is used in the PAT procedure, the model prediction accuracy is timely verified against offline analytics. As such, parallel testing comprising a set of fermentation runs was used during the initial implementation phase to confirm the model performs acceptably during its routine use. The number of parallel test runs to be performed has been progressively reduced over time, after successful model performance reviews.

In addition, a software platform for model lifecycle management is currently in place, allowing the visualization, in customized dashboards, of key indicators representative of the PAT procedure performance (i.e., equipment PQ tests, quality spectral tests and parallel testing results over time). This way, the need for model re-development or infrastructure maintenance in verified on a continual basis. The platform enables operators and process experts to report deviations, register corrective actions, and is used as a knowledge repository for reconciling evidences from both process and procedure performance.

While in batch processing, local control of each piece of equipment is in many occasions considered sufficient, in continuous manufacturing not only is local control mandatory, but also the entire process flow must be aligned and kept in a control state (Santos et al., 2015).

Continuous manufacturing has been implemented in many industries to overcome limitations related to batch processing (e.g., safety, scale-up constraints), to improve product quality and reduce costs. Particularly, in active pharmaceutical ingredient (API) manufacturing, adoption of continuous processes can hold huge potential for quality improvements and significant reduction of the manufacturing facility footprint (Xiang et al., 2012).

An example from a specific API production will be given here (Gouveia et al., 2016b – **PAPER I**). To overcome the real-time feedback limitations of lab-based analytical procedures, NIR spectroscopy (NIRS) was selected to integrate the control strategy of an API synthesis operated in continuous mode. The continuous reactor setup comprised two continuous reaction loops were a 3-step reaction takes place (Figure 7-8). A NIRS flow cell was implemented in-between the loops to determine: i) the concentration of the unstable intermediate resulting from the reaction of intermediate AB with raw material (RM) C ii) the homogeneity of the reaction mixture (reflected in particle size distribution), providing operators the means to adjust the flow rates to achieve desired intermediate concentrations. The measurement system comprised a NIR FTPA 2000-260 equipped with a transmission flow cell (Ocean Optics Inc, USA). Each recorded spectrum is communicated through an OPC protocol to the analyzer controller (ABB FTSW100) where the NIR model and the SQT control sequence are implemented. Predictions and key indicators of the quality of such predictions (Figure 7-9) are made available to the operator real-time, providing the opportunity to manually adjust the reactant flow rates, when necessary, to drive the process within acceptable concentration window. The data collected is stored in a dedicated station, the PAT data manager, ensuring data security, easy access to historians and connectivities with other data bases, when required.

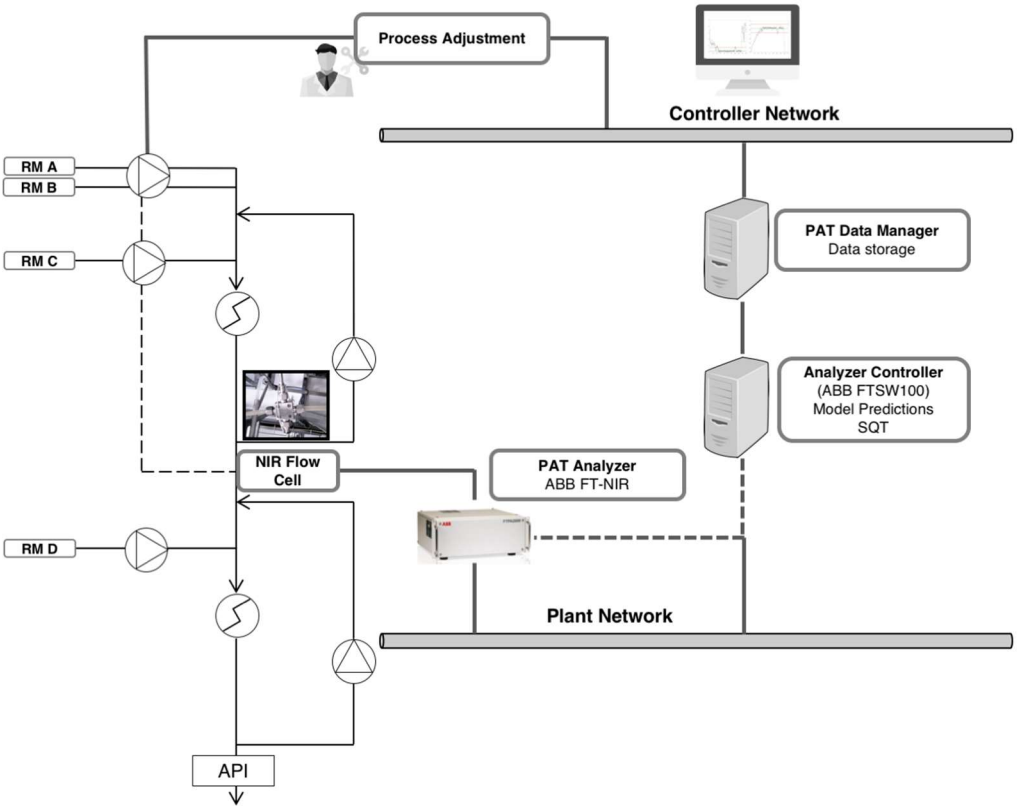


Figure 7–8: Integration of the NIR-based procedure within the continuous process flow architecture.

In continuous mode operation, it is essential to guarantee that all disturbances are controlled and unable to force the output parameters outside the targeted ranges. In the current application, SQT tests were very important to control the continuous reaction within the desired concentration window, and to access routinely the quality of reportable results from the PAT procedure. Having the in-line NIR-based application allowed to enhance the understanding of the process dynamics and a better description of residence time distribution for the continuous reaction (e.g., start-up and shutdown operations) (Figure 7-9A). Setting limits on SQT diagnostics was a critical step considered the NIR model calibration phase. The control charts were developed based on T2 and Q residual variability (Figure 7-9B and 7-9C, respectively) observed in calibration samples representative of desired concentration ranges (> 80% intermediate yield). Trending T2 values enabled operators to drive the process within such conditions and adjust the reactant flow rates to promptly surpass disturbances. The Q residuals control chart, on the other hand, is used in routine operation to identify if a new variation has been introduced in the system. In this example, when the disturbance occurred, the Q-residuals slightly increased for the respective samples, but within the variation ranges observed during calibration development, indicating that the PAT procedure is fit for purpose, reliable and further updates to the NIR model are not necessary.

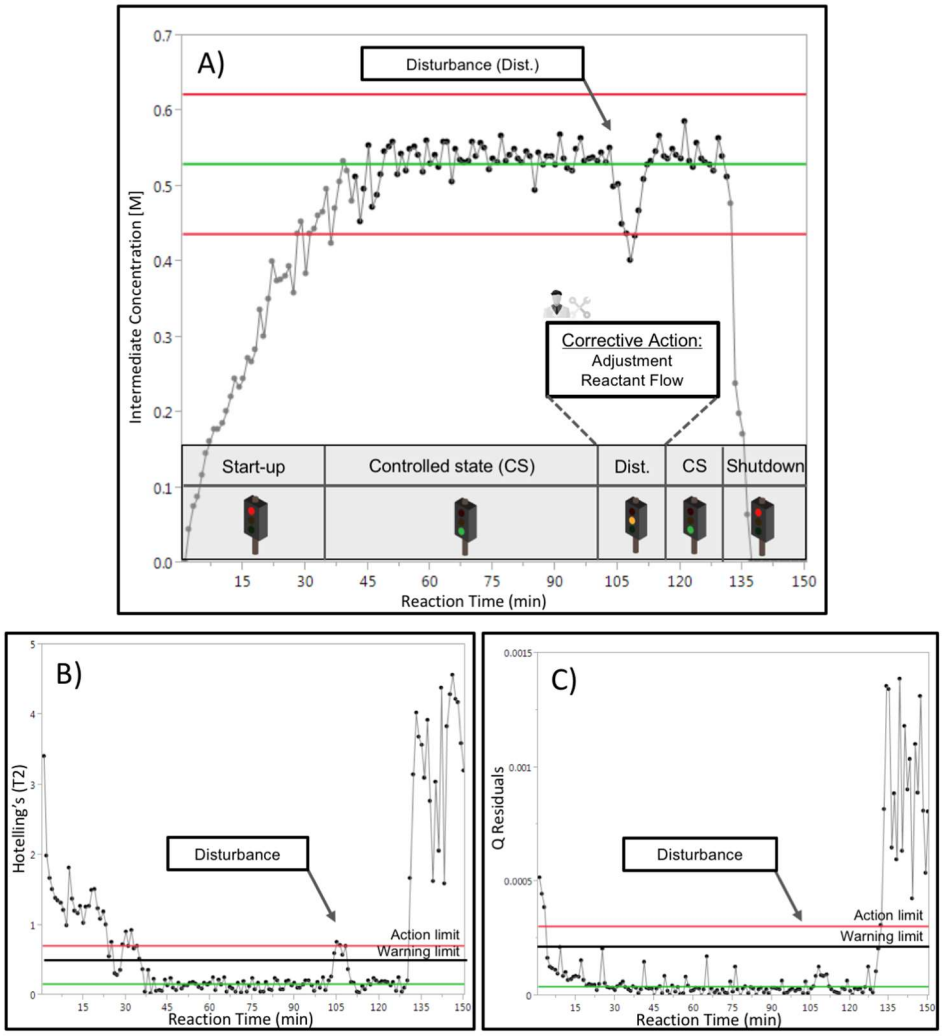


Figure 7–9: Ongoing performance verification of a NIR-based procedure. Statistical control of NIR online predictions (**A**), model Hoteling's T2 (**B**) and Q residuals (**C**). Warning and action limits established based on maximum Hotelling's T2 and Q residuals registered for calibration samples representative of acceptable concentration values, ±2σ and ±3σ, respectively.

7.4 Conclusions

Manufacturers are moving away from the Quality-by-Testing (QbT) mindset and attempting different approaches to implement PAT procedures into production and quality processes. The path is not a straight line. Different disciplines must be in place to implement PAT in its full potential, including spectroscopy (or equivalent), chemometrics, process design, data management systems and information technologies. As challenging as it might seem, specially taken into consideration the initial capital and resource investment, successful development validation and implementation of PAT procedures in pharmaceutical manufacturing is expected to be a major trend in the next years, with companies aiming at (not extensive):

- Reducing R&D costs and product time-to-market;
- Improving the efficiency of tech transfers and scale-up activities;

- Increasing product throughput, improving yields and reducing process variability;

Many [companies], have already initiated their journey in the implementation of PAT programs and are starting to realize the benefits. These include, cycle time reduction, optimization of sampling protocols and analytical testing and, enhanced process understanding resulting in a significant decrease of incidents and process deviations (Schaefer et al. 2014; BioPhorum Operations Group, 2017).

From a practical standpoint, successful implementation of PAT procedures depends upon risk mitigation, detailed planning, team commitment and a structured approach for development, validation and ongoing procedure verification for effective lifecycle management. The current regulatory framework clearly indicates how companies should demonstrate the fitness of the analytical procedure on a regular basis, throughout the procedure lifecycle. The workflow presented in the current contribution is an interpretation of such recommendations, combined with the experience implementing PAT procedures in industrial environments.

8 Knowledge-based Product and Process Lifecycle Management for Legacy Products

Legacy biological/biotechnology products have a wealth of historical data on process parameters, quality attributes and manufacturing information of various types that from a knowledge management perspective are largely unused over the lifecycle. The Subject Matter Experts (SME) involved in the development, launch and in the commercial life of those products have in-depth knowledge (viz., at all 5M levels of an Ishikawa diagram - machines, methods, materials, manpower, measurements) that could be used in a formal knowledge driven risk identification exercise. The ability to challenge and integrate the experience generated during commercial life is a fundamental requirement to ensure continued stability and capability of the manufacturing process.

Here we outline general principles and approaches necessary to align process validation activities with the product and process lifecycle concept for biological/biotechnology legacy products which include scientific, data-driven and risk-based approaches to continuously improve control strategies and the assurance of final product consistency and quality.

Keywords: legacy products, continuous process validation, knowledge and data-driven assessments.

This chapter has been published in modified format as: Strohmeier M., Pradines C., **Gouveia F.F**., Menezes J.C. (2017). Knowledge-based Product and Process Lifecycle Management for Legacy Products. Ed. Calnan, Kane, Lipa, Menezes, in Lifecycle Approach to Knowledge Excellence in the Pharmaceutical Industry, Taylor & Francis.

8.1 Introduction

The practical concept of process validation in drug manufacturing has been widely adopted since its origin in the 1970s. For many years, the emphasis in process validation was on full compliance with initially established operating procedures, leading to a perception that process validation was essentially a documentation exercise of quality assurance (*QbT*, quality by testing). Good Manufacturing Practices (GMPs) as defined by the US Food and Drug Administration in 1987 and in ICH Q7 enforced the notion of an one-off type of activity (US FDA, 1987; ICH, 2000). According to the last, “Process Validation as part of GMP is the documented evidence that the process (…) can perform effectively and reproducibly (...) meeting specifications and quality attributes (…)”. However, since “cGMPs for the 21st Century – a Risk-Based Approach” was launched by FDA in 2002, a debate was started about QbT and specially, regarding a systematic approach to confirm the effectiveness of all process steps and conditions that was needed to ensure production of drug products with necessary quality, safety and efficacy levels (US FDA, 2002).

In January 2011, FDA published a revised Process Validation Guidance (US FDA, 2011) that evolved from its earlier 2002 initiative. That document sets a shift from a documentation focused to a science and risk-based end-to-end approach – not only across process flowsheet but also across product lifecycle - integrating process/product development, scale-up activities and continued process verification (CPV) during routine manufacturing. In other words, the impact of process input parameters (alone or in combination) on product quality and process performance (the outputs) should be established from end to end of the production process and, throughout its lifecycle use. In 2012, ICH adopted this concept in its Q11 (ICH, 2012) guideline and EU changed Vol.4 of GMPs to include in Annex 15 (EU, 2015), all aspects of a risk-based and over lifecycle validation procedure as new GMP requirements (ICH, 2011; EMA, 2014).

Before that paradigm shift, process validation for commercial products was performed following a conservative approach. After identification of an appropriate process outline, repeated executions of a process would be performed to demonstrate the ability to produce product of intended quality in three non-consecutive batches. Such one-off validation approach is still accepted by health authorities. However, for those three consecutive batches only sparse data is generated as the validation exercise is performed once before commercial production. That created significant challenges as to capturing small process variations over longer periods of time and/or unexperienced during development. The lack of risk-based elements and the limitations (amount and type) of data generated in the classical approach, had a significant impact on:

- understanding of process input and output correlations, and thus subsequent CPP definition depending on the criticality of their impact on CQA, as required for definition of process robustness indexes,
- ability to evaluate batch-to-batch consistency at each process step,
- ability to evaluate the existing control strategy based on data generated during routine production,

- identification of in-process variability (as indicated by CQAs and key performance indicators - KPIs) that could in turn indicate the need for additional or improved in-process controls (i.e., related to current levels of process observability and controllability),
- definition of meaningful acceptable ranges for IPCs,
- definition of key performance indicators based on statistical data,
- maintaining process consistency (robustness and reproducibility) during process monitoring as an integral element of the third process validation stage, and
- evaluation of changes and deviations.

With these limitations, the content of a dossier for a legacy product was assembled based on general expert knowledge but with limited anticipation of problems to be experienced over the lifecycle. In addition, data generated in the validation exercise followed a conservative 'one-factor-at-a-time' development approach to establish Proven Acceptable Ranges (PAR) for each process parameter. Such data does not map the knowledge space well as process parameters (PPs) interactions are not considered. The resulting PAR typically contain the Nominal Operating Ranges (NOR), but may not entirely overlap with rigorous derived and formally defined design- and operating-spaces (cf. QbD), respectively. There are instances in which simultaneous changes inside PAR of several interacting critical PPs – a situation not considered in classical development – may lead to undesirable events (i.e., out-of-trend (OOT) or even out-of-specifications (OOS) results).

However, legacy products have two advantages over new ones. First, they hold massive amounts of highly informative data on final product quality attributes available that can be related to existing chemical and manufacturing controls in the CMC filing (i.e., IPCs and PPs). Second, there is a wealth of process and product knowledge accumulated in SMEs, throughout manufacturing history.

The current document describes an integrated approach for legacy products, developed without any QbD elements (viz., typically filed before 2005 or ICH Q8 and Q9) that aggregates the data- and knowledge-based aspects above. The goal is to apply concepts of risk- and knowledge-based process validation, well-established over the past decade in regulatory documents, to elevate legacy product validations to the next level, in terms of compliance with current regulations, robustness and operational performance. Under the framework of PQS (cf. ICH Q10, 2009), that incorporates Quality Risk Management (QRM) and science-based justifications – cf. ICH Q11, 2012, the main elements of this approach are supported by an initial risk assessment to identify:

- gaps within existing validation studies and in process controls;
- gaps in the dossier;
- identification of meaningful in process control acceptable ranges,

through:

- process input and output correlations;
- criticality assessment of PPs;
- a comprehensive description of known root-causes for variations in PPs and QAs.

When using quality risk management (QRM) principles as described in ICH Q9 (ICH, 2006), the identified gaps and associated activities can be prioritized according to the risk level, ensuring transparency to senior management within the context of their responsibility for aligning the business strategy with such perceived risks (cf. ICH Q10, 2009). The current framework integrates the principles covered by the International Conference on Harmonization (ICH) quality guidelines for industry, EU Guide to GMP Annex 15 (EU, 2015), ASTM E55 committee standard guides (namely, E2500-07, E2281-03 E2537-08 and E2709-09) (ASTM, 2007; ASTM, 2008a, ASTM, 2008b, ASTM, 2014), FDA and EMA guidelines on CPV (US FDA, 2011, EMA, 2014), and can be summarized under the concept of risk-based process validation applied to legacy products, a general framework providing the basis for QbD realization (US FDA, 2015, ICH, 2014).

Many companies find it easier to postpone improvements to facilities, processes and analytics or simply refrain from planning for improvements at all in order to avoid the intricate nature of implementing such changes, especially for product registered in multiple countries (Seymour et al., 2015). However, a knowledge and risk-based approach to process validation can significantly support the value-adding focus and contribute to enhanced control and understanding of the manufacturing process. A discussion on how to align process validation activities with the process and product lifecycle concept is provided herein.

8.2 Evaluation of the validation status of a legacy pharmaceutical product

In the current contribution, a stepwise approach (Figure 8-1) is used to improve the understanding of the manufacturing process of a legacy/commercial product. The approach includes the following elements/steps:

- identification of relevant outputs (i.e., critical quality attributes – CQAs - and key performance indicators - KPIs) for each process step influenced by that specific unit operation;
- knowledge-based identification and data collection of all process parameters and material attributes for each unit operation, that might influence relevant outputs;
- determining the functional relationship through data-driven approaches (e.g., correlation analysis, multivariate methods) between material attributes and process parameters to process relevant outputs;

- CPP definition based on the joint assessment of their variability and of their criticality through a functional relationship with all relevant outputs then,
- identification of gaps and/or improvement opportunities and validation requirements to enhance the robustness of the control strategy (i.e., quality controls, manufacturing controls and/or validation report).

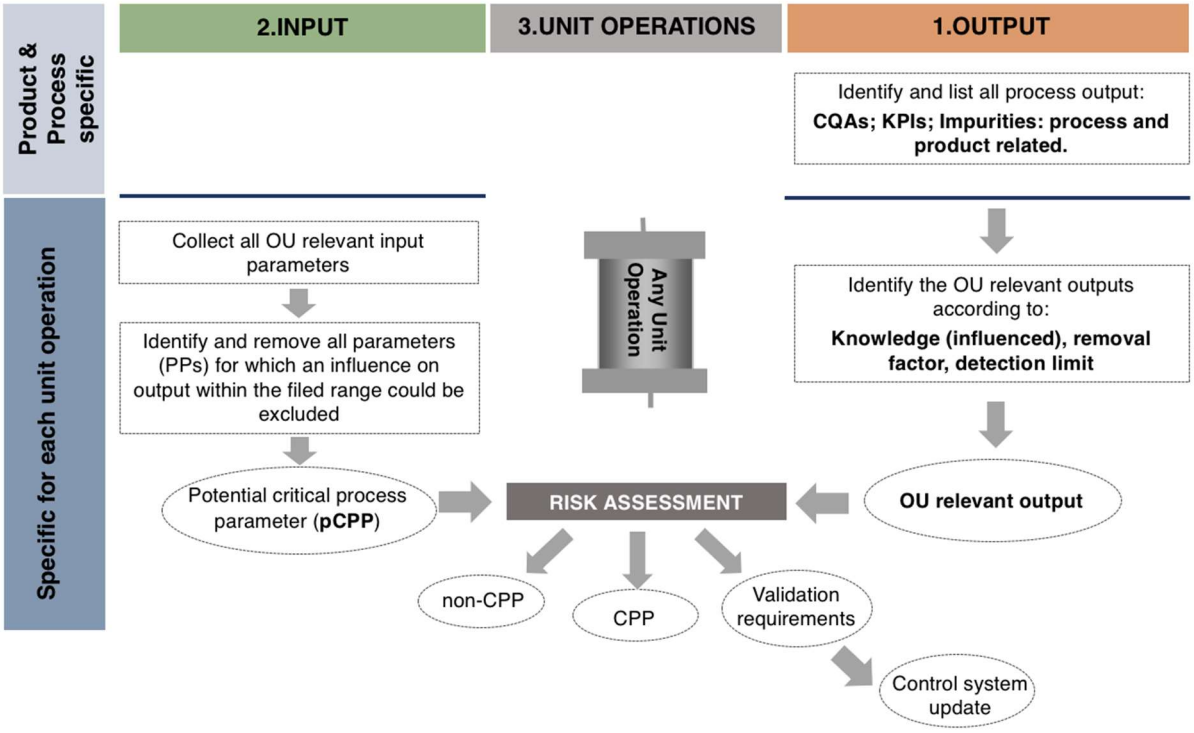


Figure 8–1: Approach for continuous evaluation of the validation status of commercial/legacy products (as presented at QRM Summit, 10-11 May, 2017 Lisbon, Portugal).

A short overview of existing published alternatives to the presented approach is provided below for this very recent area. A discussion on parameter and attribute selection based on prior knowledge and data to support CPV programs was done by Boyer et al., 2016. The authors detail how to use process capability indexes (*Cpk*) and process performance indexes (*Ppk*) based on parameters and attributes, and they provide a discussion about when these indexes should be included in the CPV plan. In addition, a comprehensive roadmap on how to implement and update a CPV plan is presented both for new and commercial products. Gouveia et al., 2016a discussed a general approach to include QbD elements into legacy filings following FDA's regulatory initiative on 'established conditions' as enabler, according to which improvement opportunities are not restricted to the available CMC history and to the current knowledge available from SMEs. The authors propose that scaled-down experiments in qualified process models should also be used with the existing lifecycle aspect. Agarwal & Hayduk, 2016 describe a comparable approach.

Assessing CQAs for a legacy product is relatively straightforward, but assessing the criticality of PPs can be much more challenging. Identification of cause-effect relationships is essential to identify CPPs and rank critical relationships between input and output parameters to prioritize improvement actions. Bozzone, 2016 described the use of the *Z score* - viz., a measure of distance from the limit of specification - and its use over lifecycle. It is perhaps a better metrics that could also be used to evaluate input parameters when the size or the distribution of the data raise concerns. De Long et al., 2016 defined an *Rp* index defined as the proportion of the allowable range used by the process (*Rp* is between 0 and 1, where *Rp* close to 0 indicates high process performance). The same authors also point out that *Cpk* and *Ppk* can be misleading when the data are not normal.

The main objective of the proposed approaches is to understand what is already known about critical interactions between input and output parameters and which elements should be further included in the validation plan to be able to consistently meet the established goals under the defined operating conditions.

8.2.1 Criticality analysis of the manufacturing process

Output parameters reflect the performance of a given unit operation and indicate whether the process gave the desired outcome for each intermediate critical quality attribute (CQA). As such, output parameters reflect the step contribution in terms of performance (e.g., yield, impurity removal factor) or influence in the properties of the final-product (e.g., purity, isoform distribution).

Under the current framework, an output filtering exercise is performed by SMEs to identify relevant parameters for each unit operation (UO) according to the following criteria:

- Existing knowledge (SMEs or corporate technology/platform knowledge);
- Introduced/not introduced by the UO;
- Data availability (e.g., validation studies);
- Detection limit before the UO;
- Covered [correlated] by other output parameter.

The UO relevant outputs and the final process outputs are used to provide evidence of process robustness and consistency.

8.2.2 Risk assessment and filtering

The knowledge and risk assessment methodology takes part of the five-stage roadmap that makes up the approach. As general requirements for each UO:

- All available inputs from the different data sources (i.e., online data, trend and offline process data, analytical measurements, experimental data) are collected and listed for a number of production batches considered to be representative of process performance over time (e.g., 30 end-product lots, lots manufactured within three production years);
- The necessary descriptors to estimate the observability [Certainty] and controllability [Occurrence] scores are available and provided in an appropriate template document.

In the risk analysis stage, the team should discuss and rank the likely impact [Severity] that deviations in process inputs might have in process relevant outputs for each UO according to predetermined criteria (Table 8-1). The observability component [Certainty] should be evaluated based on the scientific demonstration of a given relationship between the input and output parameter. Whenever possible, a quantitative threshold should be established for score ranking (e.g., degree of correlation, parameter importance in a multivariate model). The controllability component [Occurrence] is intended to demonstrate the process capability to operate within specified limits to control the input, providing evidence if additional control measures on input are required. In all situations where PAR/NOR are not specified in the batch record or in the dossier, the team should evaluate if those limits could be specified based on existing knowledge or derived from historical data.

Table 8–1: Established criteria for Severity (S), Certainty (C) and Occurrence (O) ranking.

Score	Severity	Score	Certainty	Score	Occurrence
2	Variation in process input across the acceptable range (PAR/filed) alone, or if affected by an interaction, causes no measurable/detectable variation in process output	**2**	Supported by data (correlation or variation over history/ without correlation)	**2**	CpK>= 1,2 *
6	Variation in process input across the acceptable range (PAR/filed) alone, or if affected by an interaction, causes variation in process output	**6**	Based on experience/common knowledge	**6**	CpK 1,2<x>=0,8 *
10	Variation in process input across the normal operating range (NOR) alone, or if affected by an interaction, causes variation in process output	**10**	Nothing is known (no validation or no parameter variability during correlation analysis)	**10**	CpK <0,8 * or CpK could not be calculated → low detectability)

* or another appropriate metric. Calculated based on historical data and compared to PAR (depending on what is provided in the dossier).

CPP identification should be determined by multiplying the Severity (S) score by the Certainty (C) score in order to evaluate if the input parameter is non-critical, critical or potentially critical from a process control perspective, as depicted in Table 8-2. The filtering exercise will provide a first indication of the current validation status as well as an indication of potential misalignments between the Knowledge and Evidence components – specifically, the observability score

- characterizing the actual manufacturing process. It should be highlighted that a CERTAINTY score of 10 will immediately classify an input parameter as potentially critical (pCPP) since the implementation of a revised control strategy cannot be based on uncertainty. This directly triggers additional experiments and investigations to close this gap in knowledge.

Table 8–2: SC score matrix for CPP identification.

		Input parameter classification (non-CPP/pCPP/CPP)		
Severity (S)	**10**	**CPP**	**CPP**	***p*CPP**
	6	**CPP**	**CPP**	***p*CPP**
	2	**Non-CPP**	**Non-CPP**	***p*CPP**
		2	**6**	**10**
		Certainty (C)		

The control strategy can be evaluated by combining the SC score with the controllability component [Occurrence] (Table 8-3), providing a reliable indication of the current status and a procedure to prioritize necessary follow-up actions (i.e., SCO score) in agreement with:

- Low SCO score [8-40]: well controlled parameter; the impact in output parameter is expected to be negligible;
- Medium SCO score [72-120]: process is capable to control CPP within PAR; activities to improve process control might be required if there are at-scale evidences of an optimal operating window within PAR/NOR: (1) studies to improve knowledge-based component; (2) at-scale verification if improvements have high likelihood.
- High SCO score [200-1000]: process capability to control CPP is questionable; activities to improve the control system are required: (1) process design studies to revise NOR/PAR followed by at-scale verification studies; (2) in-process controls (IPCs) update (e.g., redundant measurements) due to high variability.

Table 8–3: SCO score matrix to evaluate the current validation status and control strategy.

		SCO score		
SC score	100	200	600	1000
	60	120	360	600
	36	72	216	360
	20	40	120	200
	12	24	72	120
	4	8	24	40
		2	6	10
		Occurrence (O)		

8.2.3 Gap analysis of the current control strategy

Once the initial CPP identification has been completed, there remains the critical part of follow-up actions. The formal knowledge and risk assessment results in (1) a documented characterization of the product/process by a multidisciplinary team and (2) a prioritization of parameters or critical areas/UOs that need to be addressed. Based on the SCO score, the team should agree on a cut-off value (e.g., Pareto analysis) for prioritizing action items depending on time and resource availability. The criteria to establish the cut-off value should be recorded in the assessment report. All follow-up actions are to be addressed by the responsible person and should be documented in separate technical or validation reports (according to the PQS structure).

8.2.4 Revised control strategy: opportunities for improvement

According to ICH Q11 (ICH, 2012) a control strategy consists in a planned set of controls derived from current product and process understanding that assures process performance and product quality. Change control programs are considered essential elements of PQSs as stated in EU GMP Annex 15 (EU, 2015): “A formal system by which qualified representatives of appropriate disciplines review proposed or actual changes that might affect the validated status of facilities, systems, equipment or processes. The intent is to determine the need for action that would ensure and document that the system is maintained in a validated state.”

As part of the control system revision, the global criticality assessment can potentially trigger:

- inclusion of additional output parameters (e.g., UO relevant outputs);
- revision of output parameter criticality;
- inclusion of additional input parameters (e.g. derived from existing measurements);

- resetting operating ranges for existing input parameters or defining PAR/NOR for new ones;
- revised performance indicators target ranges;
- changes in starting materials specifications.

All these elements should be considered in the Validation Master Plan (VMP) or in the Control Strategy Document in agreement with the company's internal procedures.

8.2.5 Continued process verification and lifecycle management

After considering all relevant scientific and technical aspects involved at each critical unit operation, a systems engineering perspective is used to integrate the different components. A whole process analysis connecting raw-materials, unit operations and end-product properties is adopted comprehensively, in parallel to a time-wise integration of data, information and knowledge acquired throughout process development, industrialization and commercialization. Continuous improvement can only happen if these two perspectives are present and combined under proper data-, information- and knowledge- management systems.

Under the current framework, it is expected to (1) comply with the requirement that all manufactured product lots (batches) should be as good as any of the batches previously required for validation, even after manufacturing changes have been introduced; and (2) manage any deviation backed by deep process understanding and scientific knowledge. In this context, the resulting quality risk management report, a living document, will become the knowledge base enabling continuous improvement.

Results of the herein described assessments for each UO should be described in a comprehensive report that should include:

- scope and date;
- list of team members (name and functional role);
- raw data used in the assessment (input and output parameters);
- references to documentation used to support the *Knowledge*-based component (e.g., reference to validation studies, literature, internal data sources);
- calculations to estimate the *Evidence*-based component (observability and controllability scores);
- formal CPP assessment exercise (reference to template) and its outputs (i.e., SC scores for CPP identification and SCO scores for assessing the validation status);
- cut-off SCO value and criteria to prioritize follow up actions;
- report-out and action list (with indication of the responsible to further investigate the items assigned and write subsequent technical reports).

8.3 Conclusions

The current proposal aims to address a regulatory expectation, relevant to establishing a process validation lifecycle concept for legacy products, fully aligned with the Pharmaceutical Quality System (PQS, cf. ICH Q10) within each company and following fundamental science-based and knowledge management principles.

It enables a comprehensive post-approval lifecycle management approach that is both comprehensive, supported by a formal criticality assessment of potential risks and evidence-based backing – i.e., contains strong science- and knowledge-based components. The example provided in the Annex section demonstrates how the approach can be applied to evaluate and improve the control system of a legacy biologic product.

8.4 Annex: A streamlined-workflow for lifecycle management of biopharmaceutical legacy products

The current framework has been created with the purpose of conducting end-to-end criticality assessments of existing commercial processes influenced by a complex matrix of input/output parameters across several unit operations (UOs). Upstream (USP) and downstream (DSP) processing of biopharmaceutical products involves multiple steps and each step introduces potential critical relationships between input and output parameters that must be evaluated and controlled to ensure successful production.

A general flow of a Chinese hamster ovary-based (CHO) cell culture process for production of a recombinant protein will be further used for demonstration purposes and to illustrate the principles and tools deployed in the current document. (Figure 8-2).

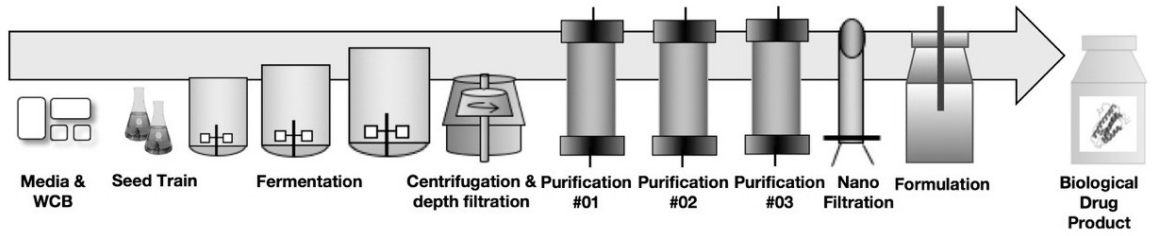


Figure 8–2: General representation of the manufacturing flowsheet for a biological/biotechnology drug product.

A vial from the working cell bank (WCB) is thawed in culture media. The cell culture is then generated and expanded through successive transfers in shake flasks. When the cell density reaches the target value, the culture is used to inoculate intermediate bioreactors to generate enough cell mass to inoculate the production bioreactor (i.e., main fermentation bioreactor). Both intermediate and main production bioreactors are monitored and controlled in a similar way to ensure production requirements. At the end of the production process, cells are harvested and clarified using centrifugation and depth-filtration. As indicated by the process layout, considerable variations can be propagated over the flowsheet including among others, variability in media components and in the feeding strategy, temperature and pH shifts, variability of gas exchange parameters and shear stresses (Gronemeyer et al., 2014). Downstream processing combines a sequence of operations to purify the harvested material. Purification steps are in general chromatography based and require sensitive and sophisticated equipment which can be object of considerable variations over the lifecycle of commercial processes. As such, performance of each chromatographic step should be evaluated in a timely manner against process outputs to mitigate variations.

A primary goal of process validation is to demonstrate that each step or unit operation is in a state of control and that all variation sources are identified, monitored as regards relevant process outputs and in a state of control. The following example is provided for illustrative purposes and exemplifies how the current approach should be applied to a purification step (Purification #01) of a legacy recombinant protein (Figure 8-2).

Stage 1 – Identification of process relevant outputs

By utilizing extensive SME and platform knowledge a complete assessment was conducted in order to identify process step relevant outputs influenced by Purification#01 operations. The assessment was conducted taking into consideration the likely impact that deviations in process inputs within the PAR/NOR ranges might have in overall process outputs.

Product specific impurities and CQAs

The product specific impurities and CQAs indicated in Table 8-4 were considered in Purification #01 assessment, to be potentially affected by input parameters variability.

Table 8–4: Identification of Product Specific CQAs potentially affected by Purification #01.

Category	CQA description	Relevant for process step CPP identification?
Product variants	Size-related variants	Yes
	Charge-related variants (Isoforms)	Yes
	Glycosylated variants	Yes
	Structural variants	Yes
Adventitious agents	Viral Purity	No
	Bioburden	No
	Endotoxins	No
DS composition	Appearance	No
	pH	No
	Protein content	Yes
	Sialic acid	Yes
	Purity	Yes
DS Strength	Potency	No
	Bioassay	No

Process specific impurities

Typically, process specific impurities are introduced by cell disruption and raw materials used throughout the production flow. The purification steps must assure that all critical process or product-related impurities are controlled/reduced to be within acceptable levels in final bulk. If a criticality assessment based on available data is not possible, further studies to demonstrate removal of identified critical impurities would be required. For those impurities already identified (Table 8-5) the validation status is assessed following the knowledge and risk assessment methodology previously described.

Table 8–5: Identification of process related impurities potentially affected by Purification #01.

Impurity	Point of Introduction	Relevant for process step CPP identification?
DNS (DNA) (pg/IU)	USP/Cell Culture	Yes
CHO Protein (ppm)	USP/Cell Culture	Yes
Antifoam (µg/mL)	USP/Fermentation	No
Methotrexate (µg/mL)	USP/Fermentation	No
Solvent A (ppm)	DSP/Purification #01	Yes
Solvent B (ppm)	DSP/Purification #02	No

Key performance indicators (KPIs)

Performance outputs are defined as indicators of how well the unit operation performed but are not directly linked with product quality. An assessment of KPIs based on historical data, process validation data and SME knowledge is performed to derive a complete list, relevant to monitor the performance and consistency of each unit operation (Table 8-6).

Table 8–6: Identification of performance Indicators potentially affected by Purification #01.

Performance indicator	Process step	Relevant for process step CPP identification?
Viability at transfer (%)	USP/Seed Train	No
Doubling Time (h)	USP/Seed Train	No
IVC (cells/mL·day)	USP/Fermentation	No
Viability (%)	USP/Fermentation	No
Yield at harvest (g/L)	USP/Fermentation	No
Max VCD (cells/mL)	USP/Fermentation	No
Culture duration (days)	USP/Fermentation	No
Specific productivity (pg/cell·day)	USP/Fermentation	No
Specific oxygen uptake rate (mmol/cell h)	USP/Fermentation	No
Max pCO_2 (mmHg)	USP/Fermentation	No
Step Yield (%)	DSP/Purification #01	Yes
Step Yield (%)	DSP/Purification #02	Yes
Step Yield (%)	DSP/Purification #03	Yes

Stage 2 – Risk impact assessment and prioritization

Identification of process parameters and material attributes

For each unit operation, all available process and material parameters are listed and collected for a number of production batches considered to be representative of process performance over time (Table 8-7). Subsequently, the data are analyzed by uni- and multivariate statistical methods to investigate if meaningful correlations with outputs parameters can be established. In the current assessment production data spanning three production years were considered (90 production batches).

Table 8–7: Collection of input parameters for Purification #01(not extensive). Initial scientific evaluation supported by SME knowledge and validation studies.

Input category	Process Parameters	PAR/NOR (dossier)	Scientific Rationale
Equipment	Temperature	5 ± 4 (ºC)	Evaluation of temperature effect in specific outputs not fully documented
	Column Volume	30-40 L	Description and harmonization of column packing available
	Column Diameter	42-46 cm	
	Compression Factor	ND	
	-//-		
Materials	Resin lot-to-lot variability	ND	Assessed by functional testing to ensure proper performance within load density limits
	Resin capacity	1g protein/L resin	
	Buffer stock solution Molarity	5 mM	Potentially critical effect on product quality/process performance
	-//-		
Process	Flow rate	15-31 cm/h	Potentially critical effect on product quality/process performance
	pH Equilibration buffer	6.9 ± 0.2	Potentially critical effect on product quality/process performance
	Volume Wash buffer	1.5-3.0 CV	Potentially critical effect on product quality/process performance
	pH Wash buffer	6.9 ± 0.2	Potentially critical effect on product quality/process performance
	pH Elution buffer	6.9 ± 0.2	Potentially critical effect on product quality/process performance
	Protein pooling (Start UV Signal)	ND	Evaluation of pooling criteria effect in specific outputs not evaluated. Moderate impact in step yield expected
	Protein pooling (End UV Signal)	ND	
	-//-		

ND – Not determined

Knowledge and data-driven components estimation

To illustrate the risk impact assessment for a input parameter in particular, the protein pooling procedure will be further detailed. While analysing potential risks the team should discuss and rank the likely impact of the Start/End protein collection triggers in process relevant outputs – Severity score. The data-driven component (Certainty and Occurrence) are simultaneously evaluated and used to demonstrate a given relationship between pooling parameters and each process step relevant output. This way, the data-driven component leverages the extensive

SMEs knowledge supported by process-specific data to provide insights if additional control measurements are required to enhance the overall control system.

Figure 8-6 describes the approach used to rank the Certainty component in this case. From the online UV elution profiles the absorbance values at which the protein collection was initiated and ended were retrieved for all production batches considered in the assessment (Figure 8-6a and 6b).

A correlation analysis was performed against all relevant process outputs according to the formula:

$$\boldsymbol{Correlation}\ (\boldsymbol{X},\boldsymbol{Y}) = \frac{\boldsymbol{\Sigma}(x-\bar{x})(y-\bar{y})}{\sqrt{\boldsymbol{\Sigma}(x-\bar{x})^2\ \boldsymbol{\Sigma}(y-\bar{y})^2}} \text{ (Equation 2)}$$

where $\bar{X}$ and $\bar{Y}$ are input and output sample means, respectively. The correlation coefficient (Figure 6c) was used to rank the relationship between the pooling criteria and all relevant process outputs together with multivariate data analysis (Figure 8-6d) to identify potential interaction effects within input parameters potentially critical to specific outputs. The consistency of the protein pooling procedure was evaluated through process capability index (c_{pK}) calculation (Figure 8-7). Since no PAR/NOR are defined in the dossier, LCL and UCL were estimated as:

$$\boldsymbol{LCL = \bar{X} - k\sigma;\ UCL = \bar{X} + k\sigma}\ \text{(Equation 3)}$$

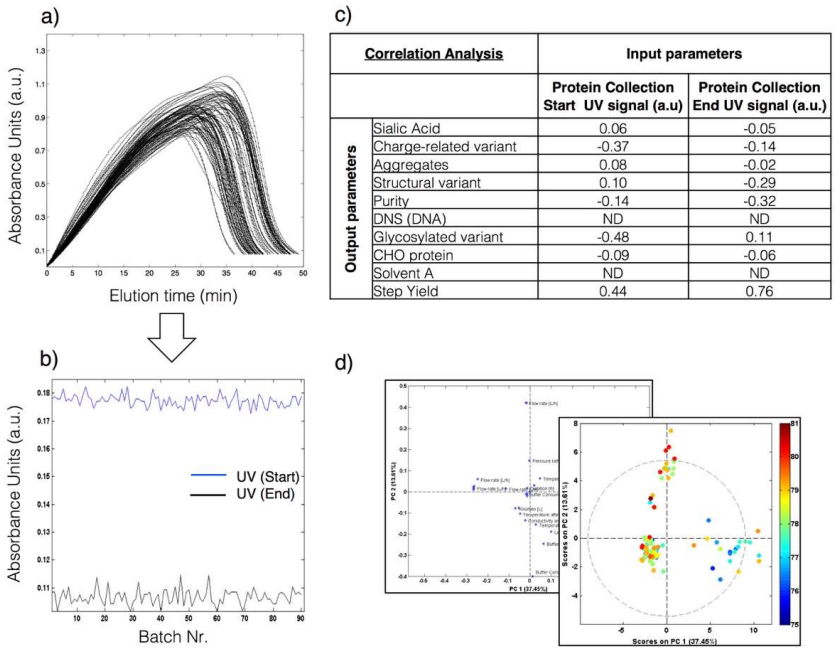
where:

$\bar{X}$= the mean of individual measurements;

k = number of standard deviations

σ= process standard deviation;

With all elements in place it is thus possible to rank Severity, Certainty and Occurrence scores towards CPP identification, perform a gap analysis of the validation status and then prioritize follow-up actions, if required (Table 8-8).



Correlation Analysis		Input parameters	
		Protein Collection Start UV signal (a.u)	Protein Collection End UV signal (a.u.)
Output parameters	Sialic Acid	0.06	-0.05
	Charge-related variant	-0.37	-0.14
	Aggregates	0.08	-0.02
	Structural variant	0.10	-0.29
	Purity	-0.14	-0.32
	DNS (DNA)	ND	ND
	Glycosylated variant	-0.48	0.11
	CHO protein	-0.09	-0.06
	Solvent A	ND	ND
	Step Yield	0.44	0.76

Figure 8–3: Estimation of the observability component for protein pooling. **a)** Online elution UV-based profiles from Purification #01 step; **b)** Protein collection Start / End absorbance units retrieved from online measurements; **c)** Correlation analysis with relevant output parameters; **d)** Principal Component Analysis (PCA) of Purification #01 input parameters (score plot colour scheme refers to step yield% over production history).

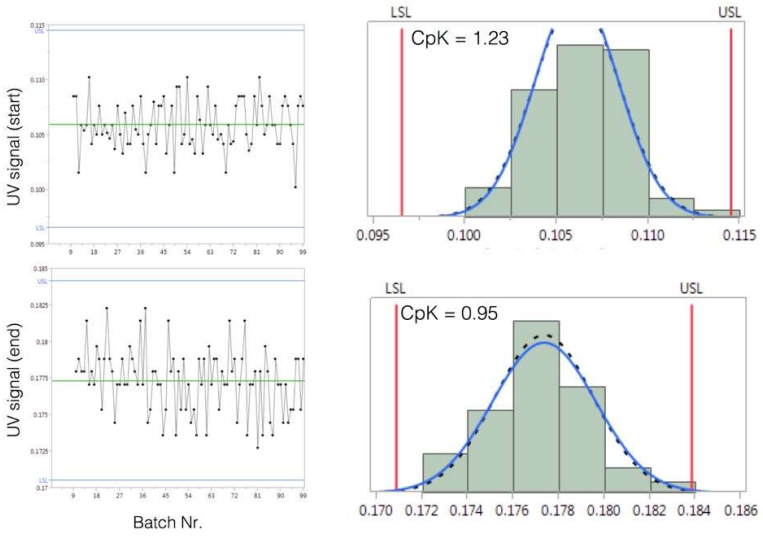


Figure 8–4: Estimation of the controllability component score value for protein pooling (Fractionation Start/End).

Table 8–8: CPP identification for process relevant outputs based on SC scores. Risk ranking and filtering of SCO scores for gap analysis of the current control system and prioritization of follow-up actions.

Process Phase	Input Parameter	Output Parameter	Severity [S]	Knowledge source	Certainty [C]	Knowledge source	Occurrence [O]	Knowledge source	SC Score	SCO score
Purification #01 (Elution)	Protein Collection Start	Size-related variants	6	SME/Platform knowledge	2	Internal Project_REF (Correlation Analysis and Multivariate Data Analysis). Solvent A clearance demonstrated by spiking studies - Validation Report	2	Internal Project_REF (CpK=1.229)	CPP	24
		Charge-related variants (Isoforms)	10	SME/Platform knowledge	2		2		CPP	40
		Glycosylated variants	10	Internal Project_REF	2		2		CPP	40
		Structural variants	6	Internal Project_REF	2		2		CPP	24
		Sialic Acid	6	Internal Project_REF	2		2		CPP	24
		Purity	2	SME/Platform knowledge	2		2		Non-CPP	8
		DNS (DNA)	2	SME/Platform knowledge	2		2		Non-CPP	8
		CHO Protein	2	SME/Platform knowledge	2		2		Non-CPP	8
		Solvent A	2	Validation report_REF	2		2		Non-CPP	8
		Step yield	6	SME/Platform knowledge	2		2		CPP	24
Purification #01 (Elution)	Protein Collection End	Size-related variants	6	SME/Platform knowledge	2	Internal Project_REF (Correlation Analysis and Multivariate Data Analysis). Solvent A clearance demonstrated by spiking studies - Validation Report	6	Internal Project_REF (CpK=0.952)	CPP	72
		Charge-related variants (Isoforms)	10	SME/Platform knowledge	2		6		CPP	120
		Glycosylated variants	10	Internal Project_REF	2		6		CPP	120
		Structural variants	6	Internal Project_REF	2		6		CPP	72
		Sialic Acid	6	Internal Project_REF	2		6		CPP	72
		Purity	6	SME/Platform knowledge	2		6		CPP	72
		DNS (DNA)	2	SME/Platform knowledge	2		6		Non-CPP	24
		CHO protein	2	SME/Platform knowledge	2		6		Non-CPP	24
		Solvent A	2	Validation report_REF	2		6		Non-CPP	24
		Step yield	10	SME/Platform knowledge	6		6		CPP	360

Stage 3 – Gap analysis and follow up actions for Purification #01

The knowledge and risk assessment should be documented and used to support a designated revision of the control system while investigations and further studies are taking place. Results of the studies should be summarized in follow-up reports, including a brief outline on the risk evaluation and mitigation actions implemented (Table 8-9).

Table 8–9: Risk evaluation and follow-up actions for protein pooling criteria optimization (Purification #01).

Process Phase	Input Parameter	Output Parameter	Severity [S]	Certainty [C]	Occurrence [O]	SC Score	SCO score	Risk Evaluation	Risk Mitigation Strategy/Recommended Actions
Purification #01 (Elution)	Protein Collection End	Step yield	10	6	6	CPP	360	Platform and process knowledge indicate a moderate influence of protein pooling criteria in quality attributes. Due to the steepness of the end part of the elution peak variations in the end-point protein collection significantly impact protein yield	Improvement of fraction collection trigger is recommended. Further investigate baseline checks and auto-zero procedures to minimize variability in protein collection
	Protein Collection End	Charge-related variants (Isoforms)	10	2	6	CPP	120		
	Protein Collection End	Glycosylated variants	10	2	6	CPP	120		
	Protein Collection End	Size-related variants	6	2	6	CPP	72		
	Protein Collection End	Structural variants	6	2	6	CPP	72		
	Protein Collection End	Sialic Acid	6	2	6	CPP	72		
	Protein Collection End	Purity	6	2	6	CPP	72		
	Protein Collection Start	Charge-related variants (Isoforms)	10	2	2	CPP	40	Input parameter not expected to have an effect in process outputs	No further actions required
	Protein Collection Start	Glycosylated variants	10	2	2	CPP	40		
	Protein Collection Start	Size-related variants	6	2	2	CPP	24		
	Protein Collection Start	Structural variants	6	2	2	CPP	24		
	Protein Collection Start	Sialic Acid	6	2	2	CPP	24		
	Protein Collection Start	Step yield	6	2	2	CPP	24		
	Protein Collection End	DNS (DNA)	2	2	6	Non-CPP	24		
	Protein Collection End	CHO protein	2	2	6	Non-CPP	24		
	Protein Collection End	Solvent A	2	2	6	Non-CPP	24		
	Protein Collection Start	Purity	2	2	2	Non-CPP	8		
	Protein Collection Start	DNS (DNA)	2	2	2	Non-CPP	8		
	Protein Collection Start	CHO Protein	2	2	2	Non-CPP	8		
	Protein Collection Start	Solvent A	2	2	2	Non-CPP	8		

Stage 4 – Revised control system – opportunities for improvement

A specific mitigation action derived from a high priority SCO score on Table 8-9, with indication of CPP to be manipulated, associated ranges, means of control, expected impacted CQAs and respective action plan, aligned with the company PQS. The current approach contributes to efficient risk management by targeting process robustness while providing the tools for continuous monitoring and ongoing risk assessment.

Stage 5 – Continued process verification and lifecycle management

The capabilities to: collect all CPPs, CQAs and control actions taken according to the revised control strategy; aggregating all that information, monitoring the correlations found during the

legacy criticality assessment, for each new lot produced; revisiting and updating the risk management and ranking; computing performance statistics (e.g., Cpk) and deciding on improvement opportunities within the current revised control strategy and company PQS.

Bridging scientific knowledge and process-derived information will increase the effectiveness of the validation program, providing the necessary evidence that the process is fully understood, well-controlled and performing in a consistent manner. The resulting documents should represent a body of work to support non-conformance investigations, post-approval changes and to address questions arising during inspections or regulatory reviews.

9 Conclusions and Perspectives

An effective development and manufacturing strategy for pharmaceutical products should be built on a solid foundation of knowledge and requirements. Just as important as the creation and implementation of high-throughput applications for the quality control of pharmaceuticals is the maintenance and improvement of such strategies. As described in **Chapters 4 to 6**, product knowledge and process understanding are essential to design the manufacturing process from a QbD perspective. The use of PAT and advanced data analysis strategies in very different production settings was a key element to: *i)* speed the development of a continuous process for an API manufacturing (**PAPER I**); *ii)* improve the knowledge about starting materials and process parameters influencing fed-batch bioprocess performance for production of a monoclonal antibody (**PAPER II**) and, *iii)* create a flexible program to respond to variability sources based on product and process understanding (**BOOK CHAPTER II**). In order to establish a flexible program, *enablers* of the overall control strategy (such as PAT and other analytical methods) must be integrated and evaluated throughout the product and process lifecycle (**BOOK CHAPTER I**).

ICH Q10 defines *enablers* as “a tool or process which provides the means to achieve an objective” and generally identifies the pharmaceutical quality system, facilities and equipment, the supply chain and the analytical control system as key elements. If continuous improvement is not built into each of these components, the overall manufacturing process will not be sustainable in the long-term. Information technology (IT) tools are essential elements to foster knowledge transfer and management over the product lifecycle phases, including development, transfer and commercialization.

As a primary goal, the research presented in this thesis aims to encourage the adoption of PAT approaches by pharmaceutical manufacturers and to demonstrate how chemometrics and advanced data analysis techniques can increase the understanding of the relationships between process parameters and product quality attributes to enhance process performance and drug product quality and consistency throughout the product lifecycle. A comprehensive demonstration of the value of PAT and multivariate data analysis in (bio)pharmaceutical development and manufacturing was presented together with a general framework for lifecycle management of PAT procedures in highly regulated production environments. It is our expectation that the detailed description of the lifecycle management of PAT models can be a valuable information source for practitioners in the field to implement the approaches described.

The work developed in this thesis opens perspectives for future work and challenges in the use of PAT methods and multivariate data analysis in pharmaceutical manufacturing. The benefits of QbD implementation depend upon considering all relevant scientific and technical aspects

involved at each phase and unit operation – a cross-functional team and a system engineering perspective must be in place to integrate the different components. In more detail, a whole process analysis connecting raw-materials to processing to end-product properties must be adopted comprehensively, in parallel to a time-wise integration of data, information and knowledge acquired throughout process development, industrialization and commercialization. Continuous improvement can only happen if these two perspectives are present and combined under proper data, information and knowledge management systems. Retrieving, visualizing and managing data sources of very different complexity – from simple univariate data (e.g., process parameters) to complex multivariate data (PAT data on quality attributes requiring further processing), both acquired with different sampling frequencies is a current challenge that has not yet been fully addressed. The range of information technologies (IT) and SME competences will demand a significant change in the structure of pharmaceutical companies. Embedding experts from outside the organization and recognizing the importance of mind-sets and behaviours to a successful transformation of the quality culture from compliance-driven to process understanding oriented, are key to realize the QbD/PAT vision.

The new process validation guidance from FDA clearly shows that companies will have to address the above challenges to be able to (1) comply with the requirement that all manufactured product lots (batches) should be as good as any of the “three golden-batches” previously required for validation; and (2) to be able to control any deviation and explain it in scientific terms backed by deep process understanding. The success of the PV approach will depend upon the efficiency of organizations to manage the very different *knowledge* sources. To embed a strong knowledge culture within the organization, top managers must encourage sharing and collaboration across functions while investing in a solid IT infrastructure, leveraging multiple knowledge sources to deploy a comprehensive PQS-KM platform.

10 References

Agarwal A, Hayduk E (2016). Retrospective implementation of quality by design for legacy commercialized enzyme replacement therapies. *Cell Culture Engineering XV* - Palm Springs (CA), May 8-13, USA.

Andersen, C. M., & Bro, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics*, 17(4), 200–215.

ASTM E2281-03 (2008). Standard Practice for Process Measurement and Capability Indices.

ASTM E2500-07 (2007). Guide for Specification, Design and Verification of Pharmaceutical and Biopharmaceutical Manufacturing Systems and Equipment.

ASTM E2537-08 (2008). Standard Guide for the Application of Continuous Quality Verification to Pharmaceutical and Biopharmaceutical Manufacturing.

ASTM E2629-11 (2011). Standard Guide for Verification of Process Analytical Technology (PAT) Enabled Control Systems, ASTM International, West Conshohocken, PA, www.astm.org.

ASTM E2709-09 (2014). Standard Practice for Demonstrating Capability to Comply with a Lot Acceptance procedure.

Baines, D (2010). Problems facing the pharmaceutical industry and approaches to ensure long term viability. *Sch. Commons*, 5(4),1–59.

Beynon K & Porter A (2000). The global pharmaceutical industry: pharmaceuticals and the treatment of disease. In: Valuing Pharmaceutical Companies, Ed. Woodhead Publishing Limited, Cambridge, England, 3-14.

BioPhorum Operations Group (2017). BPOG Technology Roadmap: In-Line Monitoring and Real-Time Release, United Kingdom, 2017.

Bodson C, Rozet E, Ziemons E, Evrard B, Hubert PH, Delattre L (2007). Validation of manufacturing process of Diltiazem HCl tablets by NIR spectrophotometry (NIRS), *Journal of Pharmaceutical and Biomedical Anal*ysis, 45, 356-361.

Boyer M, Gampfer J, Zamamiri A, Payne R (2016). A Roadmap for the Implementation of Continued Process Verification. *PDA J. Pharm. Sci. Tech*., 70, 282-292.

Bozzone S (2016). Continuous Use of Development Data in Validation Plans and the Process Lifecycle. *PDA Annual Meeting San Antonio (TX*) March 14-16, USA.

Bro R (1997). PARAFAC. tutorial and applications.pdf.ashx.pdf. *Chemometrics and Intelligent Laboratory Systems*, 38, 149–171.

Bro R (2003). Multivariate calibration: What is in chemometrics for the analytical chemist? *Anal. Chim. Acta*, 500(1), 185-194.

Bro, R, & Kiers, HAL (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, 17(5), 274–286.

Calvet, A, Li, B, & Ryder, AG (2012). Rapid quantification of tryptophan and tyrosine in chemically defined cell culture media using fluorescence spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, 71, 89–98.

Carvell JP, & Dowd JE (2006). On-line measurements and control of viable cell density in cell culture manufacturing processes using radio-frequency impedance. *Cytotechnology,* 50(1-3), 35-48.

Cervera-Padrell AE, Nielsen JP, Jønch Pedersen M, Müller Christensen K, Mortensen AR, Skovby T, Dam-Johansen K, Kiil, S, Gernaey KV (2012). Monitoring and control of a continuous Grignard reaction for the synthesis of an active pharmaceutical ingredient intermediate using inline NIR spectroscopy. *Org. Process Res. Dev*, 16(5), 901–914.

Chen R, Bowles D, Antosz F, Xiang Y, Li S, Barrila M, Coutant M (2011). Practical Approaches in Applying Process Analytical Technology (PAT) Tools to Early Active Pharmaceutical Ingredient (API) Development. *Pharm. Outsourcing*,12(4).

Chen Z, Lovett D, Morris J (2011). Process analytical technologies and real time process control a review of some spectroscopic issues and challenges. *J. Process Control*, 21(10),1467–1482.

Cole H, Demont A, & Marison I (2015). The Application of Dielectric Spectroscopy and Biocalorimetry for the Monitoring of Biomass in Immobilized Mammalian Cell Cultures. *Processes*, 3(2), 384–405.

Dabros M, Schuler MM, & Marison IW. (2010). Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass. *Bioprocess and Biosystems Engineering*, 33(9), 1109–1118.

De Beer T, Burggraeve A, Fonteyne M, Saerens L, Remon JP, Vervaet C (2011). Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int. J. Pharm*., 417(1-2),32–47.

De Beer T, Vercruysse P, Burggraeve A, Quinten T, Ouyang J, Zhang X, Vervaet C, Remon JP, Baeyens WRG (2009). In-line and real-time process monitoring of a freeze drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. *J. Pharm. Sci*. 98(9),3430–3446.

De Bleye C, Chavez P-F, Mantanus J, Ziemons EM (2012). Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications, *Journal of Pharmaceutical and Biomedical Anal*ysis, 69, 125-132.

De Boeck P, Huber U, Jahn K, Nielsen HJ (2014). Reinventing the QMS: Adding agility and robustness to control and compliance, in in (eds. Gonce A, Positano L, Rutten P, Telpis V) Flawless: from measuring failure to building quality robustness in pharma, McKinsey.

De Long M, Blaettler D, Goerke AR (2016). Development of Process Performance as a Quality Metric: Looking Back to Indicate Future Risk. *PDA Annual Meeting San Antonio (TX*) March 14-16, USA.

DiFoggio R (1995). Examination of some misconceptions about near-infrared analysis, *Applied Spec*troscopy, 49, 67-75.

Eitenmiller RR, Landen WO, & Ye L (2007). Vitamin Analysis for the Health and Food Sciences (2nd Ed). CRC Press.

EMA Addendum to EMA/CHMP/CVMP/QWP/17760/2009 Rev 2 (2014). Defining the Scope of an NIRS Procedure, June 2014.

EMA Guideline on Process Validation for Finished Products – Information and data to be provided in regulatory submissions, November 2016.

EMA Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations, January 2014.

EudraLex, Vol. 4 - EU Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use. Annex 15: Qualification and Validation, 2015.

European Pharmacopoeia Supplement 8.7 Chemometric methods applied to analytical data (5.21), April 2016.

Faassen, S, & Hitzmann, B (2015). Fluorescence Spectroscopy and Chemometric Modeling for Bioprocess Monitoring. *Sensors*, 15(5), 10271–10291.

Felizardo P, Folque, F, Machado JE, Menezes JC (2012). Process analytical technology: a common approach across different industries. *NIR News* 23(7),10-15.

Feng Y-C, Hu & C-Q (2006). Construction of universal quantitative models for the determination of roxithromycin and erythromycin ethylsuccinate in tablets from different manufacturers using near-infrared reflectance spectroscopy, *Journal of Pharmaceutical and Biomedical Anal*ysis, 41, 373-384.

Galvão RKH, Soares FCS, Martins MN, Pimentel MF, Araújo MCU (2015). Calibration transfer employing univariate correction and robust regression, *Analytical Chimica Acta*, 864, 1-8.

Gerstner W-C and Rutten P (2014). Quality 2030: Where is the frontier of flawless performance, in (eds. Gonce A, Positano L, Rutten P, Telpis V) Flawless: from measuring failure to building quality robustness in pharma, McKinsey.

Gonce A, Positano L, Rutten P, Telpis V (2014). Flawless: from measuring failure to building quality robustness in pharma, in (eds. Gonce A, Positano L, Rutten P, Telpis V) Flawless: from measuring failure to building quality robustness in pharma, McKinsey.

Gouveia FF, Campos MB, Felizardo PG (2016). Quasi-QbD Frameworks for Commercial Products: Established-Conditions. IFPAC Annual Meeting Washington DC, Jan 24-27, USA.

Gouveia FF, Rahbek JP, Mortensen AR, Pederson M, Felizardo PG, Bro R, (2016). Using PAT to accelerate the transition to continuous API manufacturing. *Analytical Bioanaytical Chemistry*, 409(3), 821-832.

Gronemeyer P, Ditz R, Strube J (2014). Trends in Upstream and Downstream Process Development for Antibody Manufacturing. *Bioengineering*, 1, 188-212.

Haack, MB, Eliasson, A, & Olsson, L. (2004). On-line cell mass monitoring of Saccharomyces cerevisiae cultivations by multi-wavelength fluorescence. *Journal of Biotechnology*, 114(1–2), 199–208.

Hakemeyer, C, Strauss, U, Werz, S, Folque, F, & Menezes, JC (2013). Near-infrared and two-dimensional fluorescence spectroscopy monitoring of monoclonal antibody fermentation media quality: Aged media decreases cell growth. *Biotechnology Journal*, 8(7), 835–846.

Harrington ACE (2010). Validation of process Analytical Technology Applications, in (eds. Chung C. Chan, Herman Lam and Xue M. Zhang) Practical Approaches to Method Validation and Essential Instrument Qualification, John Wiley & Sons, Ltd.

Hinz DC (2006). Process analytical technologies in the pharmaceutical industry. *Anal. Bioanal. Chem*., 384(5),1036-1042.

Hisiger, S, & Jolicoeur, M (2005). A multiwavelength fluorescence probe: Is one probe capable for on-line monitoring of recombinant protein production and biomass activity? *Journal of Biotechnology*, 117(4), 325–336.

Holm-Nielsen, JB, Lomborg, CJ, Oleskowicz-Popiel P, Esbensen KH (2008). On-line near infrared monitoring of glycerol-boosted anaerobic digestion processes: Evaluation of process analytical technologies. *Biotechnology and Bioengineering*, 99(2), 302–313.

Hyde JC (1998). Theoretical ffoundations of TQM and CQI: or, does quality really have its own theory? *Top Health Inf. Manage*, 18(3), 60-67.

ICH Guideline Q6B (1999). Specifications: Test Procedures and Acceptance Criteria for Biotechnology/Biologic Products.

ICH Guideline Q7 (2000). Good Manufacturing Practice Guide for Active Principle Ingredients.

ICH Harmonized Tripartite Guideline Q2(R1) (2005). Validation of analytical procedures: text and methodology.

ICH Guideline Q9 (2006), Quality Risk Management. Int. Conf. Harmonization.

ICH Guideline Q8R2 (2009a), Pharmaceutical Development. Int. Conf. Harmonization.

ICH Guideline Q10 (2009b). Pharmaceutical Quality System, Int. Conf. Harmonization.

ICH Quality Implementation Working Group Points to Consider (R2) (2011). ICH-Endorsed Guide for ICHQ8/Q9/Q10 Implementation. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_9_10_QAs/PtC/Quality_IWG_PtCR2_6dec2011.pdf. (accessed 13 September 2017).

ICH Guideline Q11 (2012). Development and Manufacture of Drug Substances.

ICH Guideline Q12 (2014). Technical and Regulatory Considerations for Pharmaceutical Product Lifecycle (concept paper).

Igne B, Hurburgh, CR (2010). Local chemometrics for samples and variables: optimizing calibration and standardization processes. *J. Chemom*., 24(2), 75-86.

J Wiss (2015). Process Analytical Technology: tools and applications in pharmaceutical manufacturing. *Chemistry Today*, 33(2), 30-33.

Jain G, Jayaraman G., Kökpinar Ö., Rinas U, Hitzmann B. (2011). On-line monitoring of recombinant bacterial cultures using multi-wavelength fluorescence spectroscopy. *Biochemical Engineering Journal,* 58–59(1), 133–139.

Jensen PS, Bak J (2002). Near-infrared transmission spectroscopy of aqueous solutions: Influence of optical pathlength on signal-to-noise ratio. *Appl. Spectrosc*., 56(12),1600–1606.

Jose, GE, Folque, F, Menezes, JC, Werz, S, Strauss, U, Hakemeyer C. (2011). Predicting Mab Product Yields from Cultivation Media Components Using Near-Infrared and 2D-Fluorescence Spectroscopies. *Biotechnol Prog*., 27(5), 1339-1346.

Juan A, Maeder M, Martinez M, Tauler R (2000). Combining hard- and soft-modelling techniques to solve kinetic problems. *Chem Int Lab Syst*, 54(2), 123-141.

Kensy F, Zang E, Faulhammer C, Tan R-K, Büchs J. (2009). Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. *Microbial Cell Factories*, 8, 31 doi: 10.1186/1475-2859-8-31.

Knop K, Kleinebudde P (2013). PAT-tools for process control in pharmaceutical film coating applications. *Int. J. Pharm.*, 457(2), 527–536.

Konakovsky V, Clemens C., Müller M, Bechmann J, Berger M, Schlatter S, Herwig, C. (2016). Metabolic Control in Mammalian Fed-Batch Cell Cultures for Reduced Lactic Acid Accumulation and Improved Process Robustness. *Bioengineering*, 3(1), 5 doi:10.3390/bioengineering3010005.

Li T, Huang B, Wang S, Chen C, Tang L, Lu Q, Lin W (2013). Covalent grafting of organic molecules onto activated carbon by a single step. *Bio Resources*, 8(2), 2300–2309.

Lindemann, C, Marose, S, Nielsen, H O, Scheper, T. (1998). 2-Dimensional fluorescence spectroscopy for on-line bioprocess monitoring, *Sens. Actuators, B. Chem*., 51, 273–277.

Liu Z, Zhou L, Mahajan A, Song S, Wang T, Ge Z, Ellison D (2006). Real-time endpoint monitoring and determination for pharmaceutical salt formation process with in-line FTIR spectroscopy. *J. Pharm. Biom. Anal.*, 41, 99-104.

Martin GP, et al., (2013). Stimuli to the revision process: lifecycle management of analytical procedures: method development, procedure performance qualification, and procedure performance verification, *Pharmacopeia Forum* 39(5).

Mazumber B, Shattacharya S and Yadav A (2011). Total Quality management in Pharmaceuticals: A Review. Int. Journal PharmaTech Research, 3(1), 365-375.

Menezes JC, Ferreira AP, Rodrigues LO, Brás LP, Alves TP (2009). Chemometrics role within the PAT context: Examples from primary pharmaceutical manufacturing. *In*: Brown S, Tauler R, Walczak eds. *Comprehensive Chemometrics*; Oxford:Elsevier, 4, 313–355.

Misturini Rossi D., Solle,D., Hitzmann B, Ayub, MAZ (2012). Chemometric modeling and two-dimensional fluorescence analysis of bioprocess with a new strain of *Klebsiella pneumoniae* to convert residual glycerol into 1,3-propanediol. *Journal of Industrial Microbiology & Biotechnology*, 39(5), 701–708.

Montague GA, Martin EB, Malley CJO (2008). Forecasting for Fermentation Operational Decision Making. *Biotechnol. Prog*. 24, 1033-1041.

Mortensen, PP, Bro, R (2006). Real-time monitoring and chemical profiling of a cultivation process. *Chemometrics and Intelligent Laboratory Systems,* 84(1):106-113.

Myerson AS, Krumme M, Nasr M, Thomas H, Braatz RD (2015). Control systems engineering in continuous pharmaceutical manufacturing. *J Pharm Sci.,* 104(3), 813-820.

Neal NS, Orefuwa AS, Overton TA, Staples JR, Mohamed AA (2013). Synthesis of Diazonium Tetrachloroaurate(III) Precursors for Surface Grafting. *Inorganics*, 1(1), 70-84.

Nethercote P, Ermer J (2012). Quality by Design for analytical methods. Implications for method validation and transfer., *Pharmaceutical Technology,* 36(10), 74-79.

Nethercote P, Ermer J (2014). Analytical Validation within the Pharmaceutical Lifecycle, in *Method Validation in Pharmaceutical Analysis: A Guide to Best Practice*, Second, Completely Revised and Updated Edition (eds. J. Ermer and P. Nethercote), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany doi: 10.1002/9783527672202.ch1.

Ödman P, Johansen CL, Olsson L, Gernaey KV, Lantz AE (2009). On-line estimation of biomass, glucose and ethanol in *Saccharomyces cerevisiae* cultivations using in-situ multi-wavelength fluorescence and software sensors. *Journal of Biotechnology*, 144(2), 102–112.

Poechlauer P, Manley J, Broxterman R, Gregertsen B, Ridemark M (2012). Continuous Processing in the Manufacture of Active Pharmaceutical Ingredients and Finished Dosage Forms: An Industry Perspective. *Org. Process Res. Dev*., 16(10),1586–1590.

Pohl M, et al. (2010). Pharmaceutical Research and Manufacturers of America (PhRMA), Analytical technical Group (ATG), the European Federation of Pharmaceutical Industries and Associations (EFPIA), and Analytical Design Space (ADS) Team. Implications and opportunities of applying QbD principles to analytical measurements. *Pharmaceutical. Technology*, 34(2), 29-37.

Rajalahti T & Kvalheim OM (2011). Multivariate data analysis in pharmaceutics: A tutorial review. *Int. J. Pharm*., 417(1-2), 280–290.

Rathore AS (2009). Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends in Biotechnology*, 27 (9):546-553.

Rathore, AS, Mittal, S, Pathak, M, Arora, A. (2014), Guidance for performing multivariate data analysis of bioprocessing data: Pitfalls and recommendations. *Biotechnol Progress*, 30: 967–973.

Read EK, Park JT, Shah RB, Riley BS, Brorson KA, Rathore AS (2010). Process analytical technology (PAT) for biopharmaceutical products: Part I. Concepts and applications. *Biotechnology and Bioengineering,* 105(2), 276-284.

Reich G (2005). Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews*, 57, 1109-1143.

Roberto M, Dearing T, Branham C, Bleie O, Marquardt B (2013). Rapid Determination of Optimal Conditions in a Continuous Flow Reactor Using Process Analytical Technology. *Processes*, 2(1), 24–33.

Rowland-Jones, RC, van den Berg F, Racher, AJ, Martin, EB, & Jaques C (2017). Comparison of spectroscopy technologies for improved monitoring of cell culture processes in miniature bioreactors. *Biotechnology Progress*, 33(2), 337–346.

Ryan PW, Li B, Shanahan M, Leister KJ, Ryder AG (2010). Prediction of Cell Culture Media Performance Using Fluorescence Spectroscopy. *Anal. Chem*., 82(4), 1311–1317.

Saerens L, Segher N, Vervaet C, Remon JP, De Beer T (2014). Validation of an in-line Raman spectroscopic method for continuous active pharmaceutical ingredient quantification during pharmaceutical hot-melt extrusion. *Anal. Chim. Acta*, 806, 180–187.

Sandor M, Rüdinger F, Bienert R, Grimm C, Solle D, Scheper T (2003). Comparative study of non-invasive monitoring via infrared spectroscopy for mammalian cell cultivations. *J Biotechnol.*, 168(4):636-45.

Santos ST, Gouveia FF, Menezes JC (2015). PAT Paves the Way for Continuous Manufacturing, Supplement to *Pharmaceutical Technology,* 39, 33-34.

Schaefer C, Clicq D, Lecomte C, Merschaert A, Norrant E, Fotiadu F (2014). A Process Analytical Technology (PAT) approach to control a new API manufacturing process: development, validation and implementation, *Talanta*, 120, 114-125.

Schaefer C, Lecomte C, Clicq D, Merschaert A, Norrant E, Fotiadu F (2013). On-line near infrared spectroscopy as a Process Analytical Technology (PAT) tool to control an industrial seeded API crystallization. *J. Pharm. Biomed. Anal*. 83,194–201.

Sellick CA, Hansen R, Jarvis RM, Maqsood AR, Stephens GM, Dickson AJ, Goodacre R (2010). Rapid Monitoring of Recombinant Antibody Production by Mammalian Cell Cultures Using Fourier Transform Infrared Spectroscopy and Chemometrics. *Biotechnol. Bioeng*., 106(3),432–442.

Setarehdan, SK, Soraghan JJ, Littlejohn D, Sadler DA (2002). Maintenance of a calibration model for near infrared spectrometry by a combined principal component analysis–partial least squares approach. *Anal. Chim. Acta*, 452, 35–45.

Seymour M., Ramnarine E., Baker, D., Vinther, A. (2015). High at the End of the Tunnel for PAC (Post-Approval Change) Complexity. PDA Letter Nov/Dec 48-49.

Smith E, Dent G (2005). Modern Raman Spectroscopy - A Practical Approach, John Wiley & Sons; 2005.

Strachan CJ, Rades T, Gordon KC, Rantanen J (2007). Raman spectroscopy for quantitative analysis of pharmaceutical solids. *J Pharm Pharmacol*., 59(2):179-92.

Surribas A, Amigo JM, Coello J, Montesinos JL, Valero F, Maspoch S (2006). Parallel factor analysis combined with PLS regression applied to the on-line monitoring of Pichia pastoris cultures. *Anal Bioanal Chem.,* 385(7):1281-8.

Svendsen, C, Skov T, van den Berg FWJ (2016). Weighted PARAFAC and Nonlinear Regression for Handling Intensity Changes in Fluorescence Spectroscopy Caused by pH Fluctuations. *Applied Spectroscopy*, 70(10), 1739–1750.

Teixeira AP, Duarte TM, Oliveira R, Carrondo MJT, Alves PM (2011). High-throughput analysis of animal cell cultures using two-dimensional fluorometry. *Journal of Biotechnology*, 151(3), 255–60.

Teixeira AP, Oliveira R., Alves PM, Carrondo MJT (2009). Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative. *Biotechnology Advances*, 27(6), 726–32.

Teixeira AP, Portugal C a M, Carinhas N, Dias JML, Crespo JP, Alves PM, Oliveira R (2009). In situ 2D fluorometry and chemometric monitoring of mammalian cell cultures. *Biotech. Bioeng.*, 102(4), 1098–106.

Triadaphillou S, Martin E., Montague G, Norden A, Jeffkins P, Stimpson S (2007). Fermentation process tracking through enhanced spectral calibration modeling. *Biotech. Bioeng.*, 97(3), 554–567.

Ulber R, Protsch C, Solle D, Hitzmann B, Willke B, Faurie R, Scheper, T (2001). Use of bioanalytical systems for the improvement of industrial tryptophan production. *Engineering in Life Sciences*, 1(1), 15–17.

US Pharmacopoeia 1225 Validation of Compendial Methods. Available at: http://www.pharmacopeia.cn/v29240/usp29nf24s0_c1225.html#usp29nf24s0_c1225 (accessed 12 March 2017).

US FDA (1987). Guideline on General Principles of Process Validation. http://www.variation.com/files/fda/FDA%20Process%20Validation%20Guidance%201987.pdf (accessed 23 March 2017).

US FDA (2004a). Pharmaceutical cGMPs for the 21st century – A Risk Based Approach, Final Report. https://www.fda.gov/downloads/drugs/developmentapprovalprocess/manufacturing/questionsandanswersoncurrentgoodmanufacturingpracticescgmpfordrugs/ucm176374.pdf (accessed 23rd March 2017).

US FDA (2004b) Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. Available at: https://www.fda.gov/downloads/drugs/guidances/ucm070305.pdf (accessed 11th September 2017).

US FDA (2011) Guidance for industry – Process Validation: General Principles and Practices. Available at: http://www.fda.gov/downloads/drugs/guidances/ucm070336.pdf (accessed 23 March 2017).

US FDA (2015) Guidance for Industry – Development and Submission of Near Infrared Analytical Procedures. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM440247.pdf (accessed 23 March 2017).

Veale EL, Irudayaraj J, Demirci A (2007). An on-line approach to monitor ethanol fermentation using FTIR spectroscopy. *Biotechnology Progress*, 23(2), 494–500.

Vibrational Spectroscopy in Pharmaceutical Development. *Particle Sciences*, Technical Brief 2011,7.

Vieira RM, Embiruçu M, Sayer C, Pinto JC, Lima EL (2003). Control strategies for complex chemical processes. Applications in polymerization processes. *Comput. Chem. Eng.*, 27(8-9),1307–1327.

Wang Y, Lysaght MJ, Kowalski BR (1992). Improvement of Multivariate Calibration through Instrument Standardization. *Analytical Chem*istry, 64(5), 562-565.

Wang Y, Veltkamp DJ, Kowalski BR (1991). Multivariate Instrument Standardization. *Analytical. Chem*istry, 63(23), 2750-2756.

Weitzel, JML. (2014). Establishment of Measurement Requirements – Analytical Target Profile and Decision Rules, in Method Validation in Pharmaceutical Analysis: A Guide to Best Practice, Second, Completely Revised and Updated Edition (eds J. Ermer and P. Nethercote), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: 10.1002/9783527672202.ch3.

WHO Technical Report Series No. 996, 2016, Annex 3 WHO good manufacturing practices for biological products.

Wise BM, Roginski RT (2015). A Calibration Model Maintenance Roadmap, IFAC-PapersOnLine, 48(8), 260-265.

Wiss J, Zilian A (2003). Online Spectroscopic Investigations (FT-IR/Raman) of Industrial Reactions: Synthesis of Tributyltin Azide and Hydrogenation of Chloronitrobenzene. *Org. Process Res. Dev.*, 7 (6), 1059–1066.

Xiang, Y, Lucas J, Van Alsten J, Li B, Preston B, Lovdahl M, Hayward CM (2012). Using process analytical technology (PAT) tools to support flow chemistry development and production. *Am. Pharm. Rev*., 15(3), 1–10.

Xiaobo Z, Jiewen Z, Povey MJW, Holmes M, Hanpin M (2010). Variable selection methods in near-infrared spectroscopy. *Anal. Chim. Acta*, 667 (1-2), 14-32.

Xing, Z, Kenty, B., Koyrakh, I., Borys, M., Pan, S. H., & Li, Z. J. (2011). Optimizing amino acid composition of CHO cell culture media for a fusion protein production. *Process Biochemistry*, 46(7), 1423–1429.

11 Annexes

POSTER I

Gouveia F.F., Rahbek J.P., Mortensen A.R., Felizardo P.M., Bro R.

Moving from Batch to Continuous Flow Production with PAT

International Conference on Near Infrared Spectroscopy (NIR 2015), Foz do Iguassu, Brazil

POSTER II

Gouveia F.F., Rahbek J.P., Mortensen A.R., Pederson M.T., Felizardo P.M., Bro R.

In-Depth Understanding of an API Synthesis: a Combined PAT and Chemometric Approach

XVI Chemometrics in Analytical Chemistry (CAC-2016), Barcelona, Spain.