



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

DOTTORATO DI RICERCA  
IN  
AREA DEL FARMACO E TRATTAMENTI INNOVATIVI  
CICLO XXX

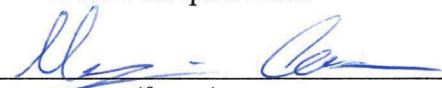
COORDINATORE Prof. Elisabetta Teodori

**Designing Quality:  
Quality by Design in the analytical  
pharmaceutical development**

Settore Scientifico-Disciplinare CHIM/01

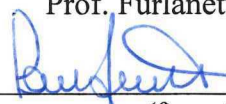
**Dottorando**

Dott. Nompari Luca

  
(firma)


**Tutore**

Prof. Furlanetto Sandra

  
(firma)

**Coordinatore**

Prof. Teodori Elisabetta

  
(firma)

Anni 2014/2017



Scientific Tutor: *Prof. Sandra Furlanetto*

University of Florence  
Chemistry department "Ugo Schiff"  
Analytical Chemistry (CHIM/01)  
Sesto Fiorentino, Florence, 50019  
Via Ugo Schiff, 6  
sandra.furlanetto@unifi.it

Theoric Tutor: *Prof. Serena Orlandini*

University of Florence  
Chemistry department "Ugo Schiff"  
Analytical Chemistry (CHIM/01)  
Sesto Fiorentino, Florence, 50019  
Via Ugo Schiff, 6  
serena.orlandini@unifi.it

PhD coordinator: *Prof. Elisabetta Teodori*

University of Florence  
Pharmaceutical Chemistry department (NEUROFARBA)  
Pharmaceutical Chemistry (CHIM/08)  
Sesto Fiorentino, Florence, 50019  
Via Ugo Schiff, 6  
elisabetta.teodori@unifi.it



## Preface

---

Analytical chemistry is the science of separation, identification and measurement of analytes [1]. The results provided by analytical chemistry are crucial for understanding characteristics of natural materials or synthetic products and for making right decisions when they have to be assessed as appropriate for a certain application.

In developing and producing pharmaceutical products, specifically vaccines, analytical methods and the related data have a key role to ensure the required and predefined product quality, as well as monitoring product and process consistency. In that context, implementation of Quality by Design (QbD) approach has been assessed as a powerful approach to provide significant advantages in terms of (I) data quality, (II) method knowledge and (III) robustness in a prospective method life cycle.

The present thesis study describes an integrated approach for the implementation of Quality by Design principles to screening and development of analytical methods applied to vaccines physical-chemical testing, according to Analytical QbD (AQbD) published in literature.

In particular, the present study is focused on application of the risk assessment and the experimental design tools to be used to build knowledge about the analytical methods performance and to mitigate the risks of failure. The risk management concept is one of the most relevant values of the AQbD framework. The recent years have seen an expansion in developing and using of management tools for risk assessment, risk control (design of experiment) and risk review (control strategy). In the experimental study here described, risk management tools and mathematical/statistical models have been applied to study and define well characterized analytical methods (based on understanding of method parameters-performance relationships), in order to establish an analytical procedure able to ensure the desired performance and to identify the method design space. The AQbD structured approach clearly mitigates the risk of failure and provides advantages in a life-cycle management prospective.



*Designing Quality:*  
*Quality by Design in the analytical pharmaceutical development*



*Chapter 1* contains an overview of the research activities and introduces the liquid chromatography and the chemometrics for experimental designs, in the context of the analytical quality by design for vaccines.

*Chapter 2* focuses on vaccines and provides a brief overview of the Bexsero product, with respect to the application of the AQbD principles developed in the present study refers.

*Chapter 3* describes in details the AQbD framework and how its principles can be applied to vaccine analytical methods.

*Chapter 4* lists the AQbD applications to different Bexsero analytical methods. Experimental design and risk assessment tools employed for method development are shown.

*Chapter 6* summarizes the study experimental activities and provides remarks and statements for AQbD application to vaccines.

*Chapter 7* deepens the conflict of interest statements.





# Table of Contents

---

Preface.....	2
Abbreviations list.....	7
Table of Figures.....	10
List of Tables.....	12
Table of Equations.....	14
Chapter 1: Introduction.....	15
UHPLC.....	17
<i>Reverse Phase Chromatography</i> .....	21
<i>Proteins analysis by Reverse Phase Chromatography</i> .....	24
Chemometrics.....	27
Experimental Design.....	28
<i>Quantitative factors</i> .....	30
<i>Qualitative factors</i> .....	30
<i>Factor space and experimental domain</i> .....	31
<i>Responses</i> .....	32
<i>Mathematical models</i> .....	32
<i>Multi-linear regression (MLR)</i> .....	33
<i>Alternative chemometric regression models</i> .....	36
Screening of Factors (Screening DoEs).....	38
<i>Factorial design</i> .....	40
<i>D-Optimal design</i> .....	41
<i>Plackett–Burman design</i> .....	42
Optimization DoEs.....	43



<i>Response Surface Methodology (RSM)</i> .....	45
<i>Central Composite design (CCD)</i> .....	46
Design variability and quality .....	47
Chapter 2: Vaccines .....	49
Bexsero vaccine (4CMenB) .....	51
Chapter 3: Analytical QbD .....	54
ATP .....	57
Method Scouting .....	58
Risk Assessment .....	59
Knowledge Space by DoEs.....	61
RSM for MODR.....	62
Model Confirmation & Robustness.....	65
Control Strategy & Life Cycle .....	66
Chapter 4: Experimental activities .....	67
RP-UHPLC Adsorption percentage.....	67
<i>Chemicals and reagents</i> .....	68
<i>Solutions and sample preparation</i> .....	68
<i>Chromatographic equipment and analysis</i> .....	69
<i>Calculations and software</i> .....	70
<i>Method development and results</i> .....	71
<i>Method qualification</i> .....	96
AAA Hydrolysis Design Space .....	97
<i>Materials and methods</i> .....	97
<i>AAA hydrolysis development and results</i> .....	102
OMV Protein Pattern Profile by RP-UHPLC.....	116
<i>Material and methods</i> .....	116



<i>RP-UPLC scouting and results .....</i>	<i>118</i>
<i>LC-MS characterization of RP-UHPLC OMV protein pattern.....</i>	<i>128</i>
Chapter 5: Bibliography .....	135
Chapter 6: Conclusions and remarks .....	153
Chapter 7: Conflict of interest .....	155
Trademark Statement .....	155
Acknowledgment.....	156
Publication Annexes.....	157
Annex 1.....	157
Annex 2.....	157



## Abbreviations list

4CMenB: four components MenB vaccine (registered as Bexsero)

A.A: Amino Acid

AAA: Ammino-Acidic Analysis

ACN or MeCN: Acetonitrile

ACQ: 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate

aka: also called as

AMQ: 6-amminoquinolone

AQbD: Analytical Quality by Design

ARD: Analytical Research and Development

AU: Adsorption Unit

BCA: Bicinchoninic Acid assay

BSA: Bovine Serum Albumin

C&E: cause and effect

C&E: cause and effect matrix

CCC: Central Composite Circumscribed design

CCF: Central Composite Faced design

Chap: Chapter

CMA: Critical Method Attribute

CMP: Critical Method Parameter

CPP: Critical Process Paramenter

CPS: capsular polysaccharide

CQA: Critical Quality Attribute

D<sub>F</sub>: Dilution Factor

DoE: Design of Experiment

d<sub>p</sub>: diameter of particles

DP: Drug Product

DS: Design Space

DS: Drug Substance

EMA: European Medicines Agency

FDA: Food and Drug Administration



FFD: full/fractional Factorial Design  
fHbp: factor H binding protein  
FMEA: failure mode and effect analysis  
GAVI: Global Alliance for Vaccines and Immunization  
GNA1030: Genome-derived Neisseria Antigen 1030  
GNA2091: Genome-derived Neisseria Antigen 2091  
GSK: GlaxoSmithKline  
HEPT: Height Equivalent of Theoretical Plates  
ICH: International Conference on Harmonization  
ID: Internal Diameter  
IEX: Ionic Exchange  
IUPAC: International Union of Pure and Applied Chemistry  
LC-FC: Liquid Chromatography - Fraction Collection  
LC-MS: Liquid Chromatography - Mass Spectroscopy  
KS: Knowledge Space  
MenB: *Neisseria meningitidis* serogroup B  
MeOH: methanol  
MLR: Multi Linear Regression  
MODR: Method operable design region (*aka* analytical design space)  
MW: Molecular Weight  
N: Number of theoretical plates  
NadA: Neisseria adhesin A  
NHBA: Neisseria Heparin Binding Antigen  
NIST: National Institute of Standard and Technology  
OFAT: One Factor At Time  
OMV: Outer Membrane Vesicles  
PCA: Design of Experiment  
pCMP: potential-Critical Method Parameter  
PFPA: Poly Fluoro Pentanoic Acid  
PhOH: phenol (C<sub>6</sub>H<sub>6</sub>O)  
PLS: Partial List Square



PorA: porine protein A of OMV

PorB: porine protein A of OMV

RA: Risk Assessment

RA: risk assessment

rMenB: recombinant protein of MenB

RP: Reverse Phase

RP-UHPLC: Reverse Phase UHPLC

RSD: relative standard deviation

RSM: Response Surface Methodology

SDS-PAGE: sodium dodecyl sulphate - polyacrylamide gel electrophoresis

SE: Size Exclusion

STD: standard

TFA: Tri-Fluoro Acetic acid

$t_R$ : Retention Time

TRD: Technical Research and Development

UHPLC: Ultra-high performance Liquid Chromatography

UniFi: University of Florence



## Table of Figures

<b>Figure 1:</b> Van Deemter curves for different particle sizes [33].	19
<b>Figure 2:</b> Viscosity of reverse phase solvent mixture [36].	19
<b>Figure 3:</b> Effect of temperature on Van Deemter curve [37].	20
<b>Figure 4:</b> Chromatography resolution dependencies [36].	22
<b>Figure 5:</b> Schematic representation protein binding to RP sorbent.	25
<b>Figure 6:</b> DoEs Vs OFAT experimentation.	29
<b>Figure 7:</b> Factor space ( <b>U</b> ) and experimental domain ( <b>X</b> ).	31
<b>Figure 8:</b> Example of graphic effects for screening by NEMROD-W.	39
<b>Figure 9:</b> Example of screening DoE output for robustness by MODDE.	40
<b>Figure 10:</b> Full factorial design $2^2$ (a) and $3^2$ (b).	41
<b>Figure 11:</b> Fractional factorial design $2^2$ (a) and $3^2$ (b).	41
<b>Figure 12:</b> D-Optimal design for two factor and 9 runs.	42
<b>Figure 13:</b> Plackett-Burman design.	43
<b>Figure 14:</b> Response surface examples.	44
<b>Figure 15:</b> CCC design for two (a) and three (b) factors.	47
<b>Figure 16:</b> CCF design for two (a) and three (b) factors.	47
<b>Figure 17:</b> Schematic representation of the Bexsero antigens on the surface of <i>N. meningitidis</i> [82].	53
<b>Figure 18:</b> Schematic representation of the Bexsero vaccine formulation [89].	53
<b>Figure 19:</b> AQbD flowchart [7].	56
<b>Figure 20:</b> Typical quality risk management process [3].	60
<b>Figure 21:</b> Graphical presentation of MODR by probabilistic surface plot.	64
<b>Figure 22:</b> Fishbone diagram for RP-UHPLC method risk assessment.	74
<b>Figure 23:</b> Graphic analysis of effects of VIAL and CONC on chromatographic area responses.	77
<b>Figure 24:</b> Graphic analysis of effects for investigation of chromatographic resolutions and area responses.	83



<b>Figure 25:</b> Graphic analysis of effects for resolution and area responses .....	84
<b>Figure 26:</b> Isoresponse surfaces drawn by plotting RAMP vs. ACN% for: (a) NHBA-GNA1030 capacity factor $K'$ , (b) $R_1$ , (c) $R_2$ , (d) $R_3$ , (e) $R_4$ at three different values of temperature: 50°C, 60°C, 70°C. ....	88
<b>Figure 27:</b> Isoresponse surfaces drawn by plotting RAMP vs. ACN% for: (a) A1, (b) AB, (c) AA, (d) A2, (e) A3 at three different values of temperature: 50°C, 60°C, 70°C. ....	90
<b>Figure 28:</b> Response Surface Methodology: graphic analysis of effects of the CMPs on the CMAs.....	91
<b>Figure 29:</b> Sweet spot plot obtained plotting RAMP vs. ACN% at three different values of T.....	92
<b>Figure 30:</b> RP-UHPLC design space map by plotting RAMP vs. ACN% three different values of temperature: 55°C, 60°C, 65°C.....	93
<b>Figure 31:</b> RP-UHPLC chromatogram in the working conditions for a mock solution. ....	94
<b>Figure 32:</b> Robustness study: graphic analysis of effects of CMPs on CMAs.....	95
<b>Figure 33:</b> AccQ•Tag Ultra reaction for amino acid derivatization [133].....	98
<b>Figure 34:</b> Standard amino acids chromatogram by RP-UHPLC.....	98
<b>Figure 35:</b> Screening DoE-graphic effect analysis for gas-hydrolysis .....	105
<b>Figure 36:</b> Ishikawa diagram for AAA assay.....	106
<b>Figure 37:</b> Isoresponse surfaces drawn by plotting HT vs. T, at three different values of VOL (100 $\mu$ l, 300 $\mu$ l, 500 $\mu$ l) and three different values of PhOH (0 $\mu$ l, 5 $\mu$ l, 10 $\mu$ l): (a) BSA, (b) $R_1$ , (c) $R_2$ , (d) $R_3$ .....	111
<b>Figure 38:</b> Response Surface Methodology: graphic analysis of effects of the CMPs on the CMAs.....	111
<b>Figure 39:</b> Sweet spot plot obtained plotting HT vs. T at three different values of VOL and PhOH .....	112
<b>Figure 40:</b> Hydrolysis design space map by plotting HT vs. T, at three different values of VOL: 150 $\mu$ l, 60 $\mu$ l, 65 $\mu$ l and three different values of PhOH: 2 $\mu$ l, 5 $\mu$ l, 8 $\mu$ l. ....	113
<b>Figure 41:</b> RP-UHPLC scouting results.....	120





<b>Figure 42:</b> RP-UHPLC analytical method map .....	122
<b>Figure 43:</b> Fishbone diagram for RP-UHPLC OMV protein pattern .....	122
<b>Figure 44:</b> RP-UHPLC gradient.....	123
<b>Figure 45:</b> D-Optimal DoE-graphic effects analysis.....	127
<b>Figure 46:</b> Fractional factorial DoE-graphic effects analysis .....	127
<b>Figure 47:</b> RP-UHPLC profile of OMV protein pattern .....	128
<b>Figure 48:</b> Gaussian charge-states distribution of PorA protein.....	129
<b>Figure 49:</b> Intact mass analysis of opcA and nspA.....	130
<b>Figure 50:</b> Intact mass analysis of PorB and fetA.....	131
<b>Figure 51:</b> Intact mass analysis of ompA and PorA.....	131
<b>Figure 52:</b> Intact mass analysis of fbpA and yaeT.....	132
<b>Figure 53:</b> LC-FC scheme.....	133
<b>Figure 54:</b> Peptide Mapping experimental design .....	133
<b>Figure 55:</b> Peptide Mapping protein matching.....	134

## List of Tables

<b>Table 1:</b> Viscosity of common HPLC solvents at various temperatures [37] [38] ....	20
<b>Table 2:</b> Organic solvents for reverse phase chromatography [44].....	26
<b>Table 3:</b> Ion-pair agents for reverse phase [45] [46] [47] .....	27
<b>Table 4:</b> Experimental matrix for Equation 10.....	36
<b>Table 5:</b> Process and analytical QbD comparison [98]. .....	57
<b>Table 6:</b> RP-UHPLC critical method attributes and selected requirements .....	73
<b>Table 7:</b> Asymmetric screening matrix for sample preparation with two replicates for each run. ....	76
<b>Table 8:</b> Experimental domain/optimized values for critical method parameters in the screening phase, in Response Surface Methodology, and in the definition of the design space. ....	79



<b>Table 9:</b> Screening of knowledge space: 16-run asymmetric screening matrix with three replicates for each run. ....	81
<b>Table 10:</b> Response surface methodology: 15-run Central Composite Design experimental plan with two replicates for each run.....	86
<b>Table 11:</b> Response surface methodology: quality of the calculated models for all the CMAs.....	87
<b>Table 12:</b> Robustness study .....	95
<b>Table 13:</b> Linearity data.....	96
<b>Table 14:</b> AAA critical method attributes and selected requirements .....	103
<b>Table 15:</b> 24-run symmetric screening matrix ( $2^3//8$ ) for gas-hydrolysis.....	104
<b>Table 16:</b> C&E matrix for hydrolysis performance.....	107
<b>Table 17:</b> Response surface methodology: 15-run Central Composite Design experimental plan with two replicates for each run.....	108
<b>Table 18:</b> Response surface methodology: quality of the calculated models for all the CMAs.....	109
<b>Table 19:</b> AAA validation summary results .....	115
<b>Table 20:</b> Reverse phase columns selected for scouting.....	120
<b>Table 21:</b> D-Optimal screening experimental design .....	125
<b>Table 22:</b> Fractional factorial experimental design.....	126



## Table of Equations

<b>Equation 1:</b> The Van Deemter equation.....	18
<b>Equation 2:</b> Fundamental resolution equation.....	22
<b>Equation 3:</b> N / HEPT relationship and equation.....	22
<b>Equation 4:</b> Capacity factor equation.....	23
<b>Equation 5:</b> Selectivity equation.....	23
<b>Equation 6:</b> Equation of linear models .....	32
<b>Equation 7:</b> Firs-order multi-linear equation with interaction for two factors .....	34
<b>Equation 8:</b> General matrix equation .....	34
<b>Equation 9:</b> General matrix for firs-order multi-linear model with interaction of two factors .....	35
<b>Equation 10:</b> Matrix of the 2-level 2-factors ( $2^2$ ) Fractional Design.....	35
<b>Equation 11:</b> AAA formula for conversion of amino acids concentration to protein amount .....	101
<b>Equation 12:</b> Equations system for intact protein mass deconvolution.....	129



## Chapter 1: Introduction

---

Over the past 10 years Quality by Design (QbD) concepts have been increasingly appreciated and applied by the pharmaceutical and biopharmaceutical industry, following the overall guidance from International Conference on Harmonization (ICH) [2] [3] [4] [5]. QbD approach is now largely adopted by the industry as a common practice for connecting knowledge of product attributes to drug safety and efficacy and in understanding how the control of manufacturing process permits to consistently ensure the product quality.

In such product development context, the analytical methods are definitely critical elements due to their roles in assisting product and process development and assuring product batches quality by means of suitable assays. Inadequate analytical methods can lead to inaccurate outcome, resulting in misleading information, detrimental for the drug development program [6]. As well, in the commercial phase, high quality methods ensure reliable product release, also reducing testing operational costs. Moreover, the deep and structured knowledge gained during QbD method development provides robust information and rationales related to regulatory submission activities, for justifying the parameters selected for assays validation and supporting method changes or removal of unnecessary tests from product specifications [6] [7].

Although, the Quality by Design principles are extensively applied in the pharmaceutical field, especially for small molecules. Quality by design application to analytical methods development has been well adopted by the pharmaceutical analytical professionals, with extensive or partial implementation of the systematic AQbD approach [6] [8] [9] [10] [11]. However, regarding biopharmaceutical molecules, as the vaccine antigens, the QbD principles even if enough consolidated in supporting process development and understanding, are not yet largely applied to analytical aspects. The aim of the present thesis is to improve the application of the Quality by Design (QbD) principles to analytical method development and in particular to investigate the application of Analytical QbD (AQbD) to vaccine



analysis. The application of AQbD to vaccines represents an opportunity for method development, to achieve a higher degree of confidence and knowledge on the method performance (i.e. precision and trueness of data) and method robustness. The application of Quality by Design principles to analytical method development is specifically based on the concept to build quality during the development stage itself [2] [7]. A good AQbD developed method can generate reliable data for product and process development; it allows confident decision making and contributes to improve the product safety and efficacy.

Applications of Quality by Design for analytical method development is reported in literature for several, different, analytical techniques. Mostly of recent examples of AQbD refer to the pharmaceutical field, mainly concerning the development of separation methods as high-performance liquid chromatography (HPLC) [12] [13] [14] [15], ultra-high-performance liquid chromatography (UHPLC) [16] [17], hydrophilic interaction liquid chromatography [18], supercritical fluid chromatography [19], capillary zone electrophoresis [20] [21] [22] [23] [24], micellar electrokinetic chromatography [25] [26] and microemulsion electrokinetic chromatography [27] [28] [29]. However, there are not analytical examples, or case studies, published for bio-molecules such as vaccine products. In this context the present study represents a pilot experimentation for AQbD application on vaccines.

The most popular assays used as release tests and for characterization by pharma and bio-pharma industries are chromatographic methods. The liquid chromatography (LC) is certainly one of the most employed, due to its feasibility, reproducibility, transferability to Quality Control labs and flexibility for a broad spectrum of applications. Many LC techniques, such as reverse phase (RP), size exclusion (SE), ionic exchange (IEX) and many others, are routinely used for testing very different product attributes, e.g. identity and purity of active ingredients, process related impurities, molecular dimension, aggregation, concentration, etc. In this context, the development of a LC method based on UHPLC technique was considered and investigated in this study by applying the AQbD framework.



The basic principles of UHPLC are provided below in bottom heading. In addition, a smattering of the basic concepts of chemometrics and experimental design will be provided. Design of Experiment (DoE) is a strategic tool for AQbD risk management and control and it is based on chemometric methods and mathematical models. The DoE is crucial to know the analytical system under development and to drive the method design space.

## UHPLC

According to IUPAC definition, "*chromatography is a physical method of separation, in which the components to be separated are distributed between two phases, one of which is stationary whilst the other moves in a definite direction*". In chromatography the stationary phase is either a solid, porous, surface active material in small particle form, or a liquid which is coated onto micro-particulate beads of an inert solid support (usually silica, but not limited to). The mobile phase is a liquid that moves through the packed bed of stationary phase in the column under pressure [30].

As it is very well known from Van Deemter equations (*Equation 1*), the efficiency of chromatographic process is inversely proportional to particle size. This model describes relationship between height equivalent of theoretical plate (HETP) and linear velocity ( $u$ ), one of the terms (path dependent term  $C$ ), is dependent on a diameter of particle packed into the analytical column. Smaller particle diameter can significantly reduce HETP which results in higher efficiency and the flatter profile of Van Deemter curve (*Figure 1*). Consequently, the mobile phase flow-rate increase does not have negative influence to the efficiency for sub-2 $\mu$ m particles [31] [32]. It is possible to increase the linear velocity and maintaining good chromatography performance, namely low HEPT values. Furthermore, increasing the linear velocity, the chromatographic peaks band boarding (diffusion dependent term  $B$ ) is decreased, hindering the longitudinal



column diffusion of analytes, contributing to increase the chromatographic efficiency. The negative aspect of small particle packed columns for LC is the high back-pressure generating. For this reason, new Ultra-High Performance Liquid Chromatography (UHPLC) systems and new column technologies able to withstand the back pressure due to sub-2 $\mu$ m particles [33] were developed by several suppliers. Sharper peaks and higher resolutions are obtained in lower analysis time using UHPLC system; enhancing the LC throughput respect to traditional HPLC. UHPLC could be considered to be a new direction of liquid chromatography. UHPLC chromatographic system is designed in a special way to withstand high system back-pressures [34] [35]. Additionally, the analyst can investigate the temperature effect, solvent type and composition to find the best operating combination of these parameters (Figure 2, Figure 3 and Table 1) to reduce mobile phase viscosity and back-pressure.

$$HETP = A + \frac{B}{u} + C u$$

**Equation 1:** *the Van Deemter equation*

Where:

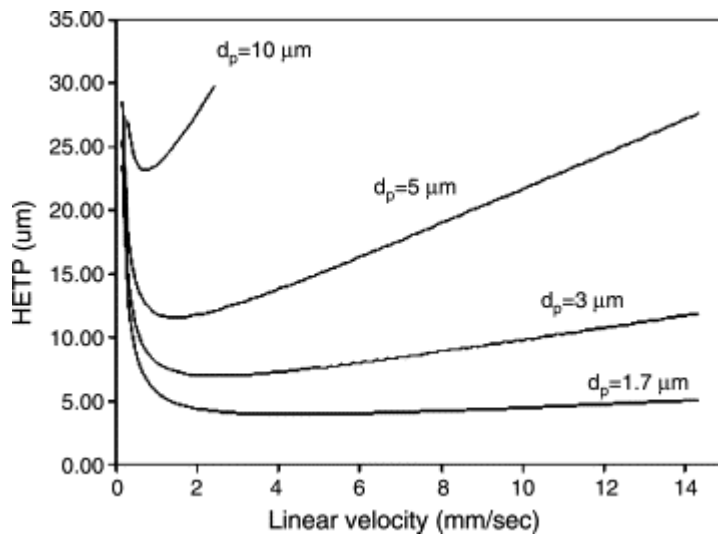
*HETP = height equivalent of theoretical plate. It is a measure of the column resolving power [m]*

*A = Eddy-diffusion parameter, related to channelling through a non-ideal packing [m]*

*B = diffusion coefficient of the eluting particles in the longitudinal direction, resulting in dispersion [m<sup>2</sup> s<sup>-1</sup>]*

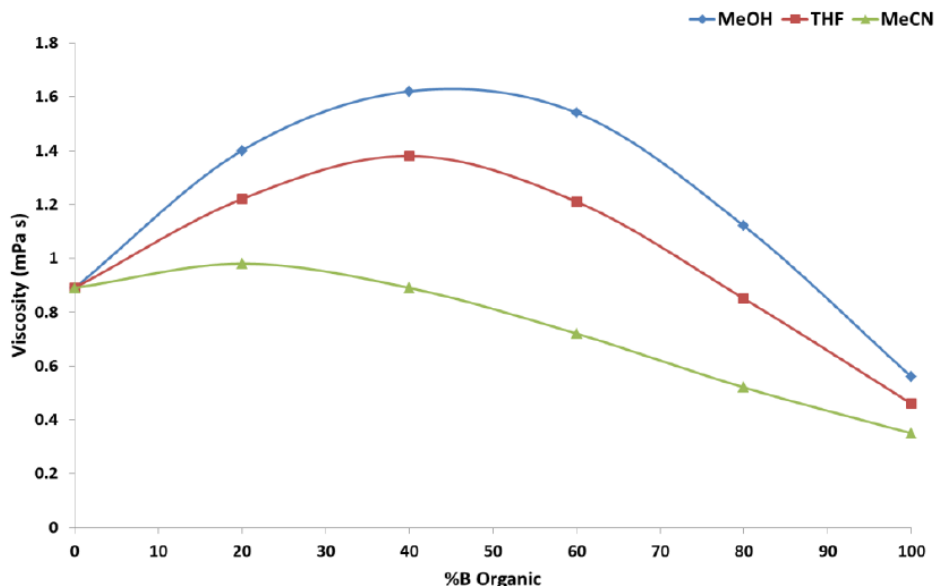
*C = Resistance to mass transfer coefficient of the analyte between mobile and stationary phase [s]*

*u = Linear Velocity [m s<sup>-1</sup>]*



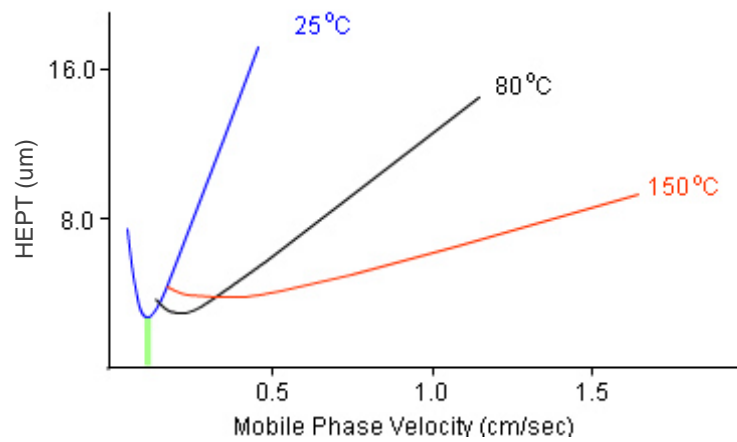
**Figure 1:** Van Deemter curves for different particle sizes [33].

The separation mechanism is still the same between HPLC and UHPLC, chromatographic principles are maintained while speed, sensitivity and resolution are improved. The result is a higher performance coupled with a lower analysis time; a very important aspect for analytical laboratories in term of solvent consumption and analysis cost.



**Figure 2:** viscosity of reverse phase solvent mixture [36]





**Figure 3:** effect of temperature on Van Deemter curve [37]

Solvent	Viscosity (mPa s)					
	-25 °C	0 °C	25 °C	50 °C	75 °C	100 °C
Methanol	1.258	0.793	0.544	-	-	-
THF	0.849	0.605	0.456	0.359	-	-
Acetonitrile	-	0.400	0.369	0.284	0.234	-
Water	-	1.793	0.890	0.547	0.378	0.282

**Table 1:** Viscosity of common HPLC solvents at various temperatures [37] [38]



### *Reverse Phase Chromatography*

Reversed Phase (RP) chromatography involves the separation of molecules on the basis of their hydrophobicity. Historically it came second after the silica polar separation (called “*normal*” or “*direct*” phase chromatography, NPC) and by this chronology takes the name of “*reverse*”. In RP chromatography a solute molecule in a polar solvent (mobile phase) binds to immobilized hydrophobic molecules (stationary phase). This partitioning occurs as a result of the solute molecule tending to have hydrophobic patches at its surface, and binding via those patches to the matrix [39]. Typical reversed phase stationary phases are hydrophobic molecules bonded to the surface of a silica support particle. Recently, other support materials and bonded phases became commercially available (monolithic phases, solid-core particles, fused-particles and many others), increasing the potentiality of RP chromatographic technique. The hydrophobic interaction between the exposed patches and the immobilized matrix is less favourable than the interaction between the bound molecule and the solvent, using buffer of increasing hydrophobicity [39]. Each on-off partition is called theoretical plate (N). The molecule releases from the matrix and elutes with a specific retention time ( $t_R$ ) due to the specific hydrophobic / hydrophilic ration (partition coefficient, **K**).

$$\mathbf{K} = \frac{[\textit{stationary phase analyte amount}]}{[\textit{mobile phase analyte amount}]}$$

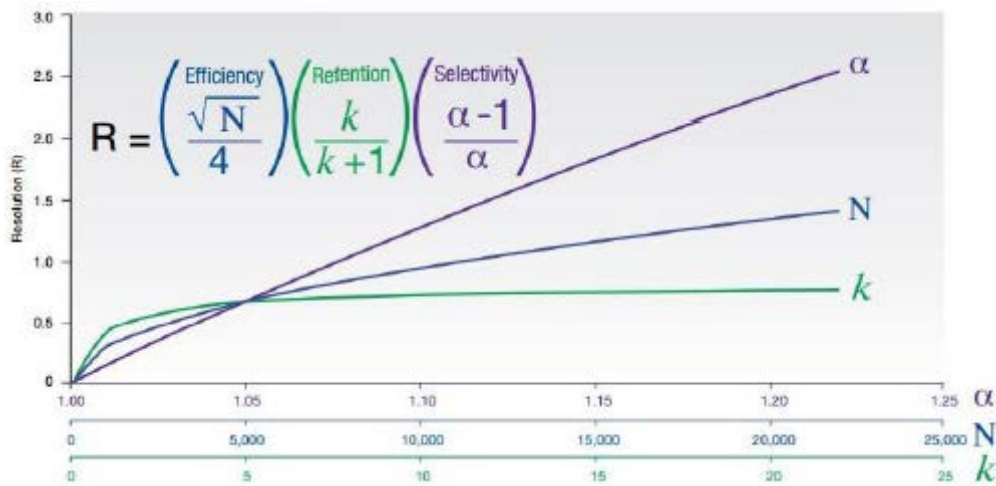
Each molecule has a specific K value and is a measure of the time the sample component resides in the stationary phase relative to the time it resides in the mobile phase: it expresses how much longer a sample component is retarded by the stationary phase than it would take to move through the column with the velocity of the mobile phase [30].

The resolution between analytes ( $R_s$ ) depends by retention times, due to K values of the chemical species, selectivity ( $\alpha$ ) and efficiency (N). The fundamental resolution equation is expressed by Equation 2.

The influence of the three terms on resolution performance is plotted in Figure 4.

$$R_s = \left( \frac{\sqrt{N}}{4} \right) \left( \frac{\alpha - 1}{\alpha} \right) \left( \frac{k'}{1 + k'} \right)$$

**Equation 2: fundamental resolution equation**



**Figure 4: chromatography resolution dependencies [36]**

**Efficiency (N):**

N is the number of theoretical plates and physically represents the number of partition between stationary and mobile phases for the analyte. It is a dimensionless number computed as the ration between column length (L) and HEPT (L / HEPT). All the considerations made for HEPT (Equation 1) are still valid and the Van Deemter equation terms explain the influence of mobile phase linear velocity and stationary particles size on chromatography efficiency (pick width) [30] [37].

$$N = \frac{L}{HEPT} = 16 \left( \frac{t_R}{W} \right)^2 = 5.54 \left( \frac{t_R}{W_{1/2}} \right)^2$$

**Equation 3: N / HEPT relationship and equation**

Were W is the base width of the peak and  $W_{1/2}$  is the width at half peak height;  $t_R$  is the peak retention time as stated above.



**Retention (k')**:

Capacity factor (k') is the retention factor and depends by the **K** partition coefficient of the chemical species. It is a means of measuring the retention of an analyte on the chromatographic column. The retention factor is equal to the ratio of the retention time of the analyte on the column ( $t_R$ ) to the retention time of an unretained compound. The unretained compound has no affinity for the stationary phase ( $t_0$ ), which is also known as the hold-up time [30] [37]. It is a dimensionless number compute according to Equation 4.

$$k' = \frac{t_R - t_0}{t_0}$$

**Equation 4:** *capacity factor equation*

**Selectivity ( $\alpha$ )**:

Selectivity is the separation factor and expresses the difference of retention between peaks. The selectivity factor  $\alpha$  is the ability of the chromatographic system to distinguish between different components. It is measured as the ratio of the capacity factors k' of the two peaks in question ( $k'_2$  and  $k'_1$ ) and can be visualized as the distance between the apices of the two peaks. By definition, the selectivity is always greater than one, as when  $\alpha$  is equal to one the two peaks are co-eluting [30] [37]. Selectivity is a dimensionless number according to Equation 5.

$$\alpha = \frac{k'_2}{k'_1} = \frac{t_2 - t_0}{t_1 - t_0}$$

**Equation 5:** *selectivity equation*

Were  $t_1$  and  $t_2$  are the retention time of the two peaks;  $t_0$  the hold-up time;  $k'_1$  and  $k'_2$  the related retention factors.

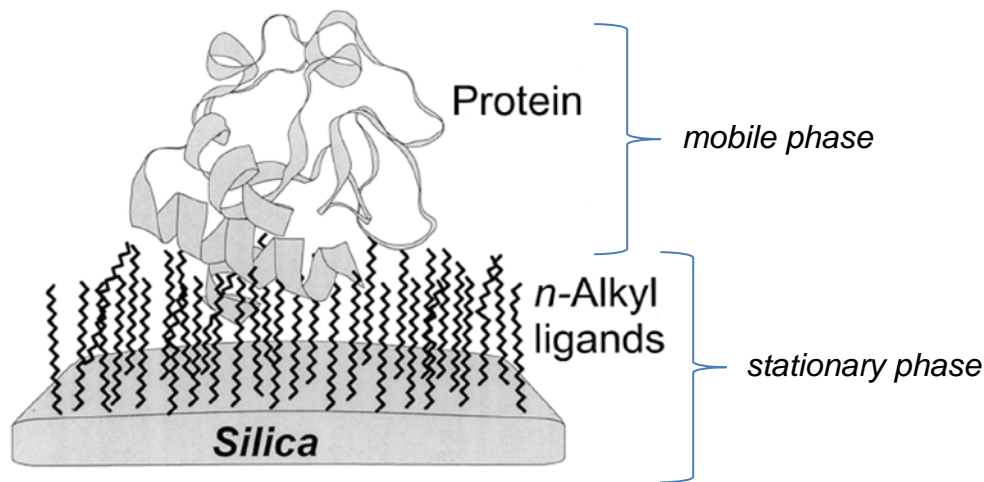


### ***Proteins analysis by Reverse Phase Chromatography***

Reverse Phase chromatography is a very powerful technique for the analysis of proteins and bio-molecules in general due to a number of advantages as excellent resolution, selectivity and reproducibility.

These advantages make reverse phase the election chromatography to monitor and control important vaccine attributes such as purity, identity and antigen content. Moreover, the RP is a suitable technique to couple with a broaden spectra of detectors: photo diode array for UV-Vis. detection, fluorescence, evaporative light scattering and mass spectroscopy (MS); thus expanding the RP technique potentiality also for characterizing antigens [40] [41] [42]. However, RP technology can cause the irreversible denaturation of proteins and bio-samples thereby reducing the potential recovery of material in a biologically active form [39].

The reverse phase chromatography for the analysis of protein antigens usually consists of an n-alkyl stationary phases (C4 or C8 to prevent adsorptions due to strong affinity for stationary phase), links to silica-based sorbent [39]. The experimental system is visualized on Figure 5 below. Considering the bigger dimension of protein antigens, respect to small drugs, there is the technology necessity to use particles with higher porous dimensions for stationary phase. The most frequently available columns have 100 Å pore particles, designed to analyse small molecules. 100 - 200 Å pore particles are not suitable for large macromolecules and can be used only for small recombinant proteins with a narrow molecular weight (MW: 50 - 100 kDa). Protein antigens and higher macromolecules in general (MW > 100 kDa) needs a broader porosity (300 Å pore particles) to permeate the stationary phase.



**Figure 5:** Schematic representation protein binding to RP sorbent

An organic solvent gradient is recommended for complex mixtures that cannot be easily separated by isocratic methods because of their wide K range. In gradient elution the eluent strength is increased during the separation by changing the composition of the mobile phase. As a result, the analysis time is reduced and the quality of the separation is improved as well as the detection limit. Binary linear gradients are the most common and they are obtained by mixing an aqueous phase and an organic phase such as acetonitrile. An ionic modifier such as trifluoroacetic acid (TFA) or formic acid (HCOOH) can be also used. The selection of the right acid agent could be critical for the chromatographic resolution of mixture products, because of it could enhance lipophilic interactions by the following action mechanisms:

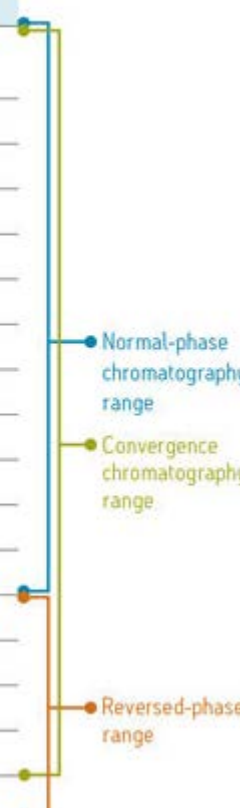
*acid modifier* – inducing the denaturation of the protein structures and exposing the lipophilic amino-acid to the external environment.

*ion-pair agent* – binding to the solute by ionic interaction with the protein charges, which results in the modification of the solute hydrophobicity. For this purpose long-chain and poly-fluoro acids are preferable to TFA, especially for low lipophilic compounds.

Typical examples of organic solvents and ion-pair agents used for reverse phase are reported on Table 2 and Table 3.

The RP separations can be easily influenced by changing the gradient elution, the operating temperature, the ionic-pair agent, the organic solvent composition, the linear velocity of the mobile phase and by choosing the right chemistry and technology of the stationary phase [39] [43]. Complex macromolecular-antigen mixtures (proteins, protein-conjugates, polysaccharides, lipopolysaccharides and DNA / RNA) can be routinely separated and picomolar or even femtomolar amounts of material can be analysed with a good sensitivity using reverse phase chromatography.

Solvent strength		
Solvent	Eluotropic value [E°]	Polarity [P°]
Pentane, Hexane, Heptane	0	0.1
Xylene	0.22	2.5
Toluene	0.22	2.4
Diethyl ether	0.29	2.8
Dichloromethane	0.30	3.1
Chloroform	0.31	4.1
Acetone	0.43	5.1
Dioxane	0.43	4.8
THF	0.48	4.0
MTBE	0.48	2.5
Ethyl acetate	0.48	4.4
DMSO	0.50	7.2
Acetonitrile	0.52	5.8
Isopropanol	0.60	3.9
Ethanol	0.68	4.3
Methanol	0.73	5.1
Water		10.2



**Table 2:** organic solvents for reverse phase chromatography [44]

Acid	formula	pKa (25°C)	H <sub>2</sub> O solubility (25°C)	name
Formic Acid	HCOOH	3.75	1.0 M	FA
Trifluoroacetic Acid	CF <sub>3</sub> -COOH	0.52	0.5 M	TFA
Perfluoropropionic Acid	CF <sub>3</sub> -CF <sub>2</sub> -COOH	0.18	0.5 M	-
Perfluorobutyric Acid	CF <sub>3</sub> -(CF <sub>2</sub> ) <sub>2</sub> -COOH	0.4	0.5 M	PFBA
Perfluoropentanoic Acid	CF <sub>3</sub> -(CF <sub>2</sub> ) <sub>3</sub> -COOH	- 2.29	0.5 M	PFPA
Perfluorohexanoic Acid	CF <sub>3</sub> -(CF <sub>2</sub> ) <sub>4</sub> -COOH	- 0.16	5 mM	-
Perfluoroheptanoic Acid	CF <sub>3</sub> -(CF <sub>2</sub> ) <sub>5</sub> -COOH	0.31	5 mM	-
Perfluorooctanoic Acid	CF <sub>3</sub> -(CF <sub>2</sub> ) <sub>6</sub> -COOH	2.8	5 mM	PFOA
Perfluorodecanoic Acid	CF <sub>3</sub> -(CF <sub>2</sub> ) <sub>11</sub> -COOH	- 5.2	5 mM	PFDA

**Table 3:** ion-pair agents for reverse phase [45] [46] [47]

## Chemometrics

The application of Quality by Design Approach to analytical method development leads to datasets that require sophisticated mathematical tools to efficiently extract information and understand the physical-chemical problem.

Chemometrics is the science of extracting information from chemical systems by data-driven means. Chemometrics utilizes methods frequently employed in core data-analytic disciplines such as multivariate statistics, applied mathematics, and computer science, in order to address problems in chemistry, biochemistry, medicine, biology and chemical engineering. Chemometrics can be applied to obtain knowledge about the analytical method [48] [49] [50].

The chemometrics guiding principles that are applied for analytical method development data interpretation are the following:

- I) *sampling* – doing a representative sampling of the sample population under study is the first step for the right experiment design. Wrong sampling could lead to wrong conclusions and untrustworthy.





- II) *experimental designing* – set up the right design for the experimentation is crucial for the extrapolation of the system information.
- III) *information extraction* – analyse the data critically, interpreting the system input-output relationships for conclusion making.

Setting an experiment by using multivariate technique could help obtaining more information on the effect of inputs (method variables) on outputs (responses) [51]. The quality of the information is higher because it is possible:

- 1) to understand not only the effect of each single variable, but to gain additional knowledge about variables interaction and the dependences between factors and their impact on responses.
- 2) to reduce the number of experiments, with time and cost saving.

All the above mentioned aspects, especially the latter, are important for the analytical application field and pharmaceutical in general, because the knowledge and the quality of the analytical / process systems are improved and the project costs and the time required for information gathering decrease.

## **Experimental Design**

A well designed experiment is an efficient method to improve knowledge about the analytical assay under development. The statistical models are crucial for an efficient design and analysis. The Design of Experiment (DoE) methodology makes possible to acquire more information about experimental factors and their interactions, in shorter time and minimizing the number of experiments respect to traditional One Factor at Time (OFAT) approach. Figure 6 provides a picture of the above mentioned concept. Considering two parameters  $X_1$  and  $X_2$  (e.g. *time* and *temperature*), the following experimental design opportunities can be built:

- traditional OFAT experimentation by which a first experiment to find the optimum for  $X_1$ , keeping  $X_2$  factor fixed at an arbitrary level (a), and a subsequently second experiment for optimization of  $X_2$  factor, keeping fixed  $X_1$  factor at the best (b), are performed. The resulting point found (blue point) does not take into account the interaction between the two variables and therefore depends on the condition from which the OFAT experiment is conducted. As a consequence, the identified point (blue point) can be a relative maximum within the whole performance surface of the experiment and not necessarily the absolute maximum point within the performance surface.
- DoE approach by which the absolute optimization of  $X_1$  and  $X_2$  factors is targeted. In particular (c) evidence an interaction effect (orange arrow) and the best performance is in the orange circle. By only one DoE experiment (indeed two by OFAT) you get more information and the absolute maximum of the experimental performance surface is found. In this context, it is possible to find the whole set of possible  $X_1$ - $X_2$  combinations (orange space) by which the absolute maximum is found; a maximum performance surface is reached instead of only a single point. This concept is key and will be deeper discussed in the AQBd chapter (Chap. 3).

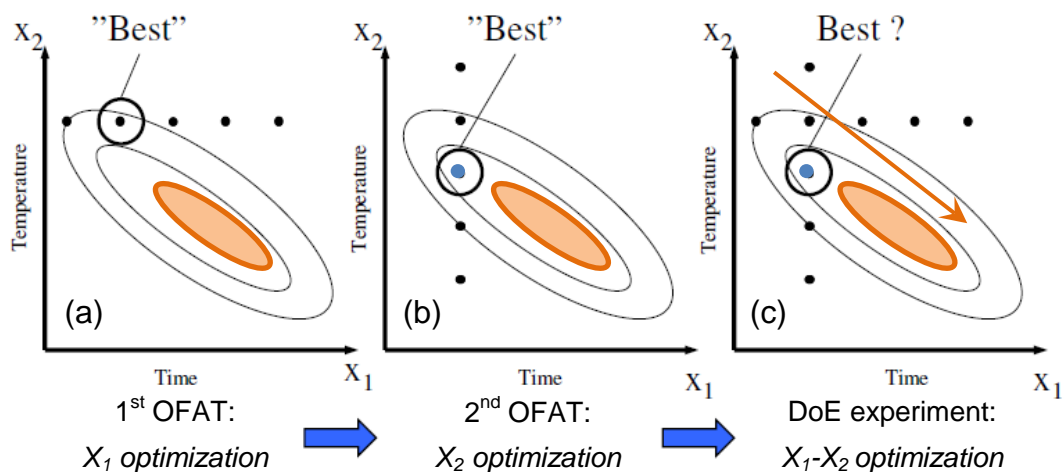


Figure 6: DoEs Vs OFAT experimentation



As proclaims by A.C. Atkinson and A.N. Donev in the Optimum Experimental Design book, “a well-designed experiment is an efficient method of learning about the world” [51]. In pharmaceutical analytical development, the assay purpose and the answer to give are often known before starting the method development. The Analytical Target Profile (ATP) is crucial for that purpose and will be discussed in the dedicated section in chapter 3. To know the intended purpose of the method is an essential starting point to define the analytical development strategy and to set up the right experimental design.

### *Quantitative factors*

Quantitative factors are those acting on the analytical system that can take numerical value (e.g. rate, time, percentage, amount, etc...). Among all, the majority are continuous variables and can therefore be set at any value. However, basing on physical-chemical understanding of the phenomenon that is intended to be studied, the analytical technique, physical principles limits and their impact on analytical method application (e.g. the pH range restriction suitable for a chromatographic column), the continuous factors / variables are studied within a defined range or at discrete levels. Unless otherwise stated, a quantitative factor is assumed continuous [52].

### *Qualitative factors*

Qualitative factors take only discrete value. In pharmaceutical development, an example of analytical qualitative factors might be the type of column for a screening (e.g. pore-, monolithic- and solid-core column; RP-C4 and RP-C8), organic eluent type (e.g. ACN, THF and MeOH), ion-pair agent (e.g. all the acid substances listed on Table 3) and many other development parameter for different analytical techniques (e.g. spectroscopy, electrophoresis, bio-assay, etc...).

### Factor space and experimental domain

The *factor space* is the  $n$ -dimensional space defined by the coded variables  $X_i$  for the continuous quantitative factors being investigated. If only two variables (factors) are investigated it is a bi-dimensional space, for  $n$  factors it is a  $n$ -dimensional space. The factor space is defined in term of  $n$  independent variables. Only a part of the *factor space*, which is defined as *experimental domain*, is of interest for pharmaceutical applications. Sometimes it is also called *region of interest*, it is the part of the factor space enclosed by upper and lower limits of the coded variables. As an example of this definition a simplified one dimensional space for pH can be considered. The factor space is the pH range possible (0 – 14), while the experimental domain can be any space contained within it, for example the space 6 – 12, if referred to the pH range of use of a silica chromatographic column.

For qualitative factors, the factor space consists of discrete points and the experimental domain is equal in number to the product of the available levels of all the qualitative factors [52]. Exemplificative illustrations of bi-dimensional experimental domains of continuous and qualitative factors are reported in Figure 7 below.

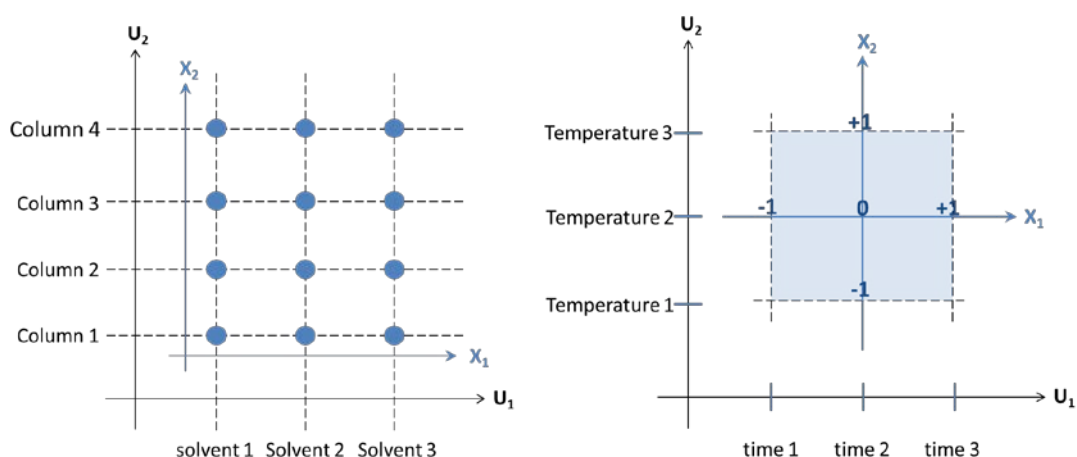


Figure 7: factor space ( $U$ ) and experimental domain ( $X$ )

## *Responses*

Responses (Y) are measured properties of the analytical method, sometime referred to as dependent variables. The measured response for experiment n<sup>th</sup> will be written as Y<sub>n</sub> [52]. The identification of the responses to be measured (method out-puts) is depending by the intended purpose of the assay and in this context it is important what is defined in the analytical target profile (ATP). The discussion concerning of the ATP is postponed to AQbD chapter 3. Now it is important to understand that ATP drives the selection of the responses to control the method performance.

## *Mathematical models*

A mathematical model is an expression defining the dependence of responses (dependent variables) on the independent variables: factors or parameters. The models most frequently used are the linear models (Equation 6), that should not be confused with first order models. A linear model is a model in which all terms may be represented by a constant coefficient  $\beta_i$  multiplied by a variable ( $x_i$ ;  $x_i x_j$ ;  $x_i^2$ ;  $x_i^3$ ; etc.) [52]. The linear model, also called linear regression models, could be a polynomial of the first-order without interaction (Eq.A) or with interaction between variables (Eq.B); polynomial of the second-order (Eq.C), known as quadratic models; more rarely third-order (cubic models) or superior order models are used. The purpose is to use the simplest possible model fitting the analytical problem.

$$\text{Eq. A: } Y = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + \varepsilon$$

$$\text{Eq. B: } Y = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + \beta_{12} X_1 X_2 + \dots + \beta_{ij} X_i X_j + \varepsilon$$

$$\text{Eq. C: } Y = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + \beta_{12} X_1 X_2 + \dots + \beta_{ij} X_i X_j + \beta_{11} X_1^2 \dots + \beta_{ii} X_i^2 + \varepsilon$$

### **Equation 6:** *equation of linear models*

Where:

Y = *response*



$\beta_0$  = model constant term

$\beta_i$  = constant coefficient of variable (factor effect)

$X_i$  = variable

$\varepsilon$  = experimental error

The linear models are the most used to represent relationship between variables and responses in an analytical experiment. Nevertheless, there are some situations when models with non-linear parameters have to be preferred; particularly if the error assumptions are violated by the transformed model [37]. Non-linear modelling can be utilized efficiently in a vast number of situations where traditional modelling is impractical or impossible to apply. Nonlinear modelling approaches are generally formulated by differential or exponential equations and include non-parametric methods, such as feed forward neural networks, kernel regression, multivariate splines and many others.

### ***Multi-linear regression (MLR)***

During the development of an analytical method as well as for chemical process, the factors to be studied are more than one ( $X_1, X_2 \dots X_i$ ) as well as the responses ( $Y, Y' \dots Y^k$ ) and each  $Y^k$  output should be studied at different levels ( $Y_1, Y_2 \dots Y_n$ ;  $Y'_1, Y'_2 \dots Y'_n$ ;  $Y^k_1, Y^k_2 \dots Y^k_n$ ). As an example, for separation method, responses to be controlled have to be peaks resolution, column capacity, run time, peaks symmetry/purity. The type of responses to be controlled depends on the basic principle of the assay, (e.g. bio-assay, spectroscopic method).

The assumption behind the modelling of an analytical system is that each experimental result can be represented by the same equation (Equation 6) for every response and each level ( $Y^k_n$ ).

The overall analytical system will then be described by the combination of all equations  $Y^k_n = f(X_{n1} \dots X_{ni})$ . In the Equation 7 below, as application example, the analytical problem is represented for a first-order two-factor linear-model with interactions, for one response. For a more sophisticated

analytical system, one equation for each of responses of the analytical system under study must be defined [52].

$$Y_1 = \beta_{10}X_{01} + \beta_1X_{11} + \beta_2X_{12} + \beta_{12}X_{11}X_{12} + \varepsilon_1$$

...

$$Y_2 = \beta_{20}X_{02} + \beta_1X_{21} + \beta_2X_{22} + \beta_{12}X_{21}X_{22} + \varepsilon_2$$

...

$$Y_n = \beta_{n0}X_{0n} + \beta_1X_{n1} + \beta_2X_{n2} + \beta_{12}X_{n1}X_{n2} + \varepsilon_n$$

**Equation 7:** *first-order multi-linear equation with interaction for two factors*

Where:

$Y_1$  = the value of the response  $Y$  in the 1<sup>st</sup> experiment condition

$Y_2$  = the value of the response  $Y$  in the 2<sup>nd</sup> experiment condition

$Y_n$  = the value of the response  $Y$  in the  $N^{\text{th}}$  experiment condition

$X_{n1}$  = the level of the  $X_1$  factor in the  $N^{\text{th}}$  experiment condition

$X_{n2}$  = the level of the  $X_2$  factor in the  $N^{\text{th}}$  experiment condition

$\beta_1$  = effect coefficient of  $X_1$  factors for each  $N^{\text{th}}$  condition ( $\beta_1X_{n1}$ )

$\beta_2$  = effect coefficient of  $X_2$  factors for each  $N^{\text{th}}$  condition ( $\beta_2X_{n2}$ )

$\beta_{12}$  = effect coefficient of  $X_1$ - $X_2$  interaction for each  $N^{\text{th}}$  condition ( $\beta_{12}X_{n1}X_{n2}$ )

$\beta_{10}$ ;  $\beta_{20}$ ;  $\beta_{n0}$  = model constant terms for each  $N^{\text{th}}$  measurement

$\varepsilon_1$ ;  $\varepsilon_2$ ;  $\varepsilon_n$  = experimental error for each  $N^{\text{th}}$  measurement

Always considering only the response  $Y$ , the Equation 7 could be written in a matrix form to represent the system of equations as a table for a direct and visual interpretation of the multivariate conditions tested. Applying Equation 8, Equation 7 is translated in matrix Equation 9.

$$Y = X\beta + \varepsilon$$

**Equation 8:** *general matrix equation*

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} x_{11} & x_{12} & x_{11}x_{12} \\ x_{21} & x_{22} & x_{21}x_{22} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n1}x_{n2} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{12} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

**Equation 9:** general matrix for first-order multi-linear model with interaction of two factors

Where:

**X** is a  $N \times p$  matrix, called the **model matrix** and known also as **effect matrix**, having as many columns as there are the coefficients  $\beta$  in the model ( $p$ ) and as many rows as there are the experiments ( $N$ : the experiments number depend by the design chosen and the number of factors and levels) [52].

**Y** is the **column matrix** and is the vector of the experimental responses.  $\varepsilon$  is the vector of the experimental errors and  $\beta$  is the vector of the factors coefficients [52].

Applying a fractional design to the 2 factors ( $X_1; X_2$ ) and studying at two levels each factor (-1; +1), with the inclusion of a central point (0; 0) and two replicates for all conditions tested, we obtained the following experimental matrix showed by Equation 10 and related Table 4.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & +1 \\ -1 & +1 & +1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_{12} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

**Equation 10:** matrix of the 2-level 2-factors ( $2^2$ ) Fractional Design



The model matrix  $\mathbf{X}$  (blue dash box on Equation 10) has been and will normally to be represented as a table; as in following Table 4. The column for the model matrix corresponds for a 2-level fractional design corresponds to the linear combinations of factors for calculating its coefficients [52].

$X_1$	$X_2$	$X_1X_2$	measured Y response
+1	-1	-1	$Y_1$
+1	-1	-1	$Y_2$
-1	+1	+1	$Y_3$
-1	+1	+1	$Y_4$
0	0	0	$Y_5$
0	0	0	$Y_6$

**Table 4:** experimental matrix for Equation 10

The MLR model is the foundational of regression models applied to the experimental design discussed in this thesis.

### *Alternative chemometric regression models*

Different regression models could be used for the factor assessment and responses prediction, e.g. partial least squares (PLS) and principal component regression (PCR) methods.

*Principal component regression* (PCR) is a regression analysis technique that is based on principal component analysis (PCA). Typically, it considers regressing the responses on a set of covariates (independent variables) based on a standard linear regression model, while uses principal component analysis (PCA) for estimating the unknown regression coefficients in the model. PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PC). In PCR, instead of regress the dependent variables, the principal components of the explanatory variables are used as repressors. One major advantages of PCR lies in



overcoming the multicollinearity problem. For analytical application, PCR does not represent a viable solution since all variables have a different effect on the analytical system and multicollinearity is not relevant [53] [54] [55] [56].

*Partial least squares regression* (PLS regression) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of maximum variance between the response and independent variables. PLS finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the X and Y data are projected to new spaces, the PLS family of methods are known as bilinear factor models. PLS is used to find the fundamental relations between two matrices (X and Y), i.e. a latent variable approach to modelling the covariance structures in these two spaces. A PLS model will try to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. PLS regression is particularly suited when the matrix of predictors has more variables than responses [53] [57].



## Screening of Factors (Screening DoEs)

The screening purpose is to select from the factors which may possible influence the analytical process being studied those which factors have a real effect on responses, an influence that is unequivocally undistinguishable from the background noise [52].

Screening DoE is normally done very early in the life of the method development in order to simplify the analytical problem and thus to concentrate attention and resources to optimize the main important factors. Although, many factors can often be listed as possibly important, it is not unusual that a large part of the experimental variation for responses can be explained by a small number of the factors studied. The objective of a screening study is not to obtain numerical data for method factors effect, but rather to discriminate between factors that have effect on analytical responses and factor that do not have a significant effect. Both qualitative (typically) and quantitative (discrete) factors could be studied by screening DoEs [52]. The screening goal is to perform a minimum number of experiments exploring the highest possible number of factors [58].

As previously mentioned, screening DoEs are generally performed before and in preparation to optimization DoE, to identify the “trivial many” vs. the “vital few” parameters and subsequently deprioritize the first and continue to study the latter. It is possible that the experiments carried out in the screening study might be exhaustive to identify the optimal value for some parameters, typically for the qualitative factors. A practical analytical example could be the column type selection or the ion-pair agent screening for liquid chromatography.

A second employment of Screening DoEs for analytical pharmaceutical purposes is for testing assay robustness (small changes in set point conditions) and ruggedness (e.g. replicates of set point conditions) [58]. The intended purpose of these studies is to demonstrate that the explored changes are not significant for responses. It is particularly useful (I) during method development, for the identification of a suitable control space and

a good control strategy, e.g. system suitability test, and (II) during validation to demonstrate robustness and ruggedness by multivariate techniques.

Two examples of screening DoEs are reported as graphic effect analyses on Figure 8 and Figure 9. Graphic effect analysis is a simply histogram representation of the model-factor coefficients, with related variance estimation. It is generally used to understand if the change of level for the selected factors and interactions has influence on the selected response. The example on Figure 8 shows the effects of the five factors b1...b5 on response Y. It appears clear that when b2 and b5 are at their higher levels the response increases, contrariwise to b1 that prefers the lower factor level. b3 factor does not impact the response, while b4 shows an effect close to pure error of the model.

Figure 9 shows the robustness of the Y response against the method factors (ACN, time and temperature). The estimation of the model coefficient parameters are not significant by ANOVA and the effects (green bars) are lower than model variability.

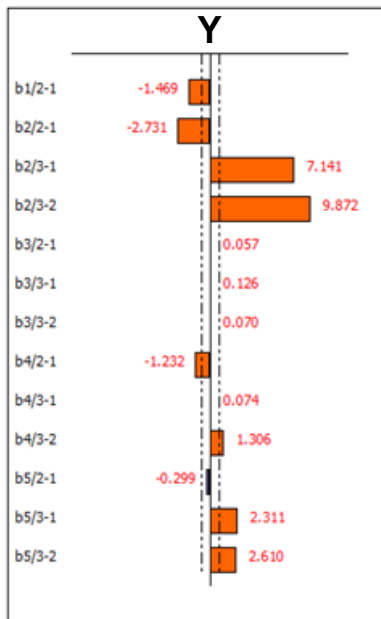
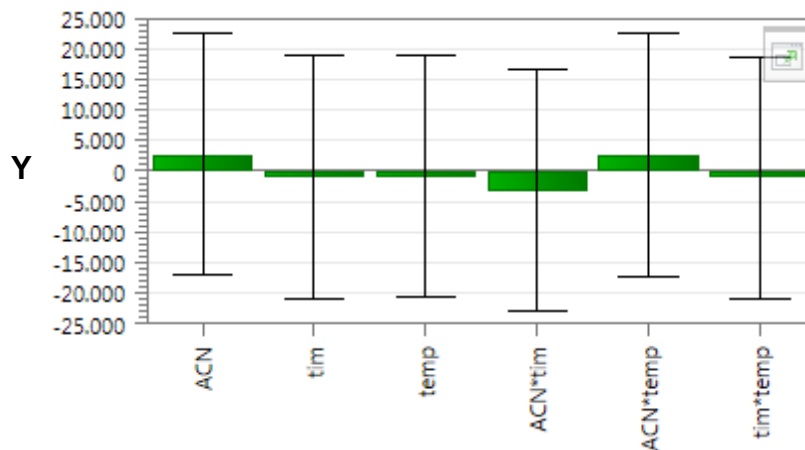


Figure 8: example of graphic effects for screening by NEMROD-W



**Figure 9:** example of screening DoE output for robustness by MODDE

The DoE possibilities are not only restricted to Factorial, D-Optimal and Plackett-Burman design; others could be available for different analytical purposes, e.g. Rechtschaffner design and many others.

### *Factorial design*

Factorial designs identify experiments at every combination of factor levels. There are  $L^k$  combinations of  $L$  levels of  $k$  factors. In full factorial designs (see example in Figure 10) every experiment is performed, while for fractional factorial designs a specific subset is performed that allows calculation of certain coefficients of the model (see example in Figure 11). Two-level designs are typically chosen for screening factors since can provide main and interaction effects, without providing indications on higher orders interactions. Higher number of levels are used for the screening design (3, 5, n levels) mainly as fractional design in order to provide indications on main effects and interactions with fewer runs respect to full factorial.

Calculation of effects in two levels designs is easy. If the two levels are coded +1 and -1, then the column of +1 and -1 under each factor is multiplied by the response for each experiment. The result of this product is then added by the half of the experiments number and finally divided by the half of the number of the performed experiments, leading to the main effect for the factor. For the interaction effect, a new column

is created, it is the product of the level codes and the procedure outlined above is applied to this column [58].

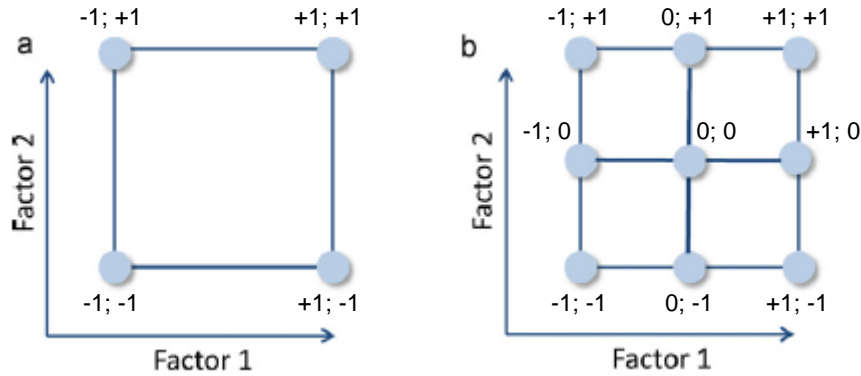


Figure 10: full factorial design  $2^2$  (a) and  $3^2$  (b)

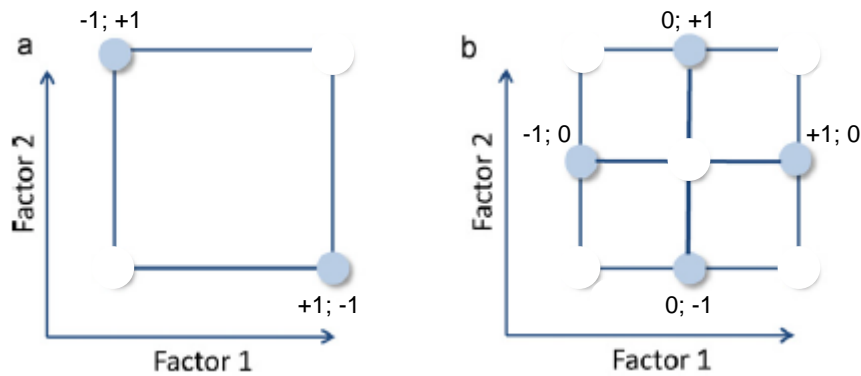


Figure 11: fractional factorial design  $2^2$  (a) and  $3^2$  (b)

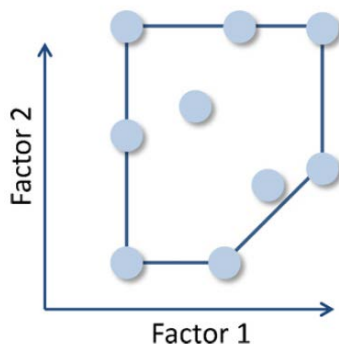
Full factorial design is typically chosen for screening few parameters or for robustness studies. Fractional factorial design with high resolution are generally chosen when the number of factor to screen is high (e.g. for early analytical screening) and a high number of information are required from a restricted number of experiments.

### D-Optimal design

The D-optimal design is able to define the optimal distribution of the experimental points based on a defined number of design points [59]. This occurs when the determinant of the matrix  $\mathbf{X} \mathbf{X}^T$  is maximised, where  $\mathbf{X}$  is the matrix of design points and  $\mathbf{X}^T$  denotes the transpose.

The approach allows stepwise refinement: upon a minimal design output is analysed, further points can be added to refine knowledge [58]. D-optimal designs might represent an advantageous solution in the following situations:

- a) When the factor space is not uniformly accessible, for example when combinations of solvent composition and solute concentration are not possible.
  - b) To pre-define a limited number of experiments that guarantees adequate analysis of the output (that corresponds to the number of effects plus a constant term).
  - c) When there are limitations in the sample set, (e.g. the maximum runs possible within the same analytical session) or when values or combinations of parameters are not possible for practicability limits.
- An example of a D-Optimal design for two factors and 9 runs is reported in Figure 12 below. The combinations of high value for factor 1 and low value for factor 2 are excluded by the design.



**Figure 12:** *D-Optimal design for two factor and 9 runs*

### ***Plackett-Burman design***

Plackett-Burman design requires  $4n$  experiments to be performed to investigate a maximum of  $4n - 1$  factors at two levels. Plackett and Burman design has become particularly popular for robustness tests because one of the runs requires the base level of each factor. When deciding the levels, if ‘-’ is allocated to the base level of the factor, then ‘+’ is this base plus a small change that is being investigated as part of

the robustness study (see the example on Figure 13 below for a three factors design). Note that the change can be an increase or a decrease. The main effect that is obtained from the analysis of the DoE is an estimate of the change in response as the factor goes from the ‘-’ level to the ‘+’ level [58] [60]. This peculiarity makes Plackett-Burman design an useful tool for confirmation and validation of the design space model and/or to identify the so called control space, a robust and controlled space within the design space where the analytical performance is not affect by the factors variation.

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>
1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
2	-1	+1	-1	+1	+1	+1	-1	-1	-1	+1	-1
3	-1	-1	+1	-1	+1	+1	+1	-1	-1	-1	+1
4	+1	-1	-1	+1	-1	+1	+1	+1	-1	-1	-1
5	-1	+1	-1	-1	+1	-1	+1	+1	+1	-1	-1
6	-1	-1	+1	-1	-1	+1	-1	+1	+1	+1	-1
7	-1	-1	-1	+1	-1	-1	+1	-1	+1	+1	+1
8	+1	-1	-1	-1	+1	-1	-1	+1	-1	+1	+1
9	+1	+1	-1	-1	-1	+1	-1	-1	+1	-1	+1
10	+1	+1	+1	-1	-1	-1	+1	-1	-1	+1	-1
11	-1	+1	+1	+1	-1	-1	-1	+1	-1	-1	+1
12	+1	-1	+1	+1	+1	-1	-1	-1	+1	-1	-1

**Figure 13:** *Plackett-Burman design*

## Optimization DoEs

Optimisation is the process of determining where the desired target values for factors lie. The graphical representation of the relationship between responses and factors ( $Y^k = f_{X_1, X_2, \dots, X_i}$ ) are so called *response surfaces* (Figure 14). There is not always a single maximum (or minimum, depending on assay purpose) in the response surface. Often the response plateaus and there is an area of response surface with approximately the



same value (minimization example *b*). Sometimes the response function shows a saddle with maximum values at the edges (maximization example *a*). For chromatographic separations the goal is to find conditions whose responses meet the target criteria, by operating in a range that can be considered applicable from an operational point of view (e.g. for time and temperature, according to Figure 14 example), rather than find the absolute optimum. This makes DoE very powerful when the polynomial function does not fit the data perfectly, but does describe the response sufficiently to locate an acceptable region, in order to have the possibility to identify a maximization or minimization on responses. Mathematically, to find the maxim/minimum optimum at least quadratic terms are needed, refer to previous Equation 6 for models discussion [52] [58].

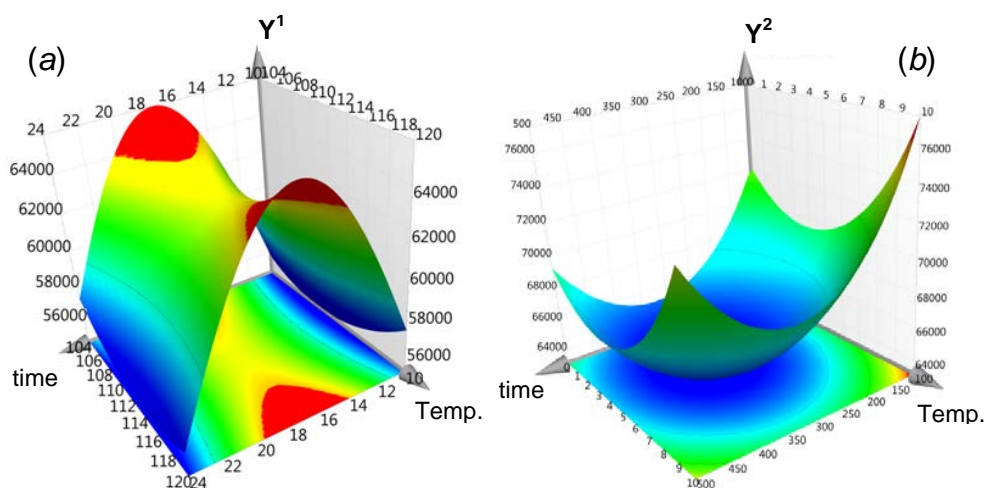


Figure 14: response surface examples

In the reality, during the development of an analytical method (and a process in general) there is not only one response, but more responses to be optimize at the same time: e.g. chromatographic outputs could be resolution ( $R_1, R_2 \dots R_n$ ), capacity ( $k'$ ) or selectivity ( $\alpha$ ) parameters between peaks. For this reason, it is of crucial importance the definition of the target criteria (refer to ATP paragraph on chapter 3). It is important to define not only a target but an acceptable region for responses, where the output values are suitable. Certain responses could oppose another one; changes of a factor that improve a response may have negative effect on a second [52]. In that context it is crucial the “a-priory” definition of the



method intended purpose by formalization of the analytical target profile (ATP). The ATP is the key tool for the prioritization between method responses, and is also useful to identify the desirability values and ranges for the specific method-output responses. The definition of the desirability values/ranges is crucial and leads the definition of the final design space by the implementation of the desirability function to response surfaces. Refer to AQbD chapter 3 for a deeper discussion on desirability, while response surface methodology is presented below.

### ***Response Surface Methodology (RSM)***

Response surface analysis is a graphical optimization technique of multiple responses. The polynomial models estimated by measured responses are graphically represented by response surfaces plots. The graphical representation of the responses behaviour helps to understand and visualize the effect of the method parameters. It is a relatively simple technique for finding the optimum combination of factors to satisfy the desirability values. Designing the optimization DoE for response surface methodology (RSM) it is important to consider that, to estimate a model we need to carry out as many experiments as many coefficient there are in the model. More replicates, central points and additional test points help to improve the quality of the predictive RSM model. RSM is typically applied to quantitative and continuous factors and the resulting response surface can be visualised by contour plots or three-dimensional diagrams. As can be easily understood, RSM cannot be used, at least not directly, for discontinuous responses. A good RSM model should be predictive over the whole experimental domain and the prediction goodness should be validated, e.g. by Plackett-Burman design as previously discussed [52].

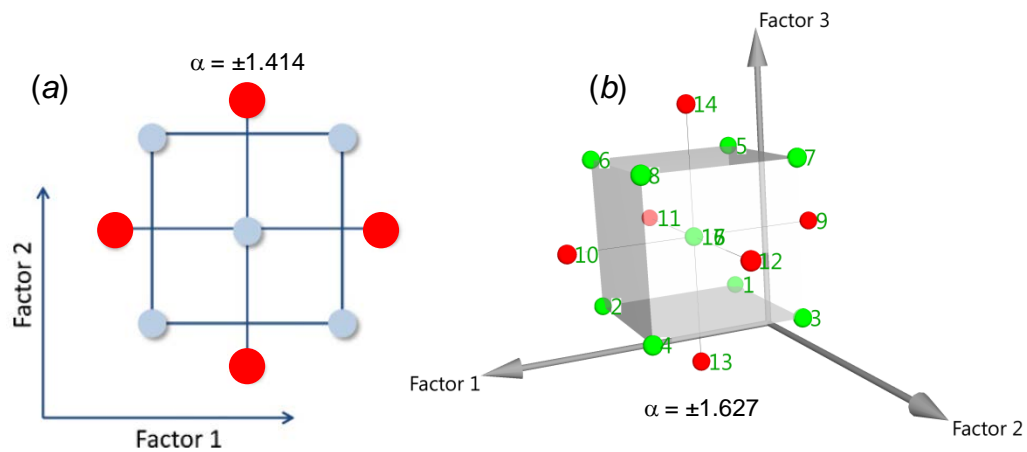
The DoE possibilities are not only restricted to Face Centred and Central Composite designs, e.g. Box-Behnken, D-Optima, Doehlert and Full Factorial designs, could be available and chosen for different optimization analytical purposes. In the following sections, the two



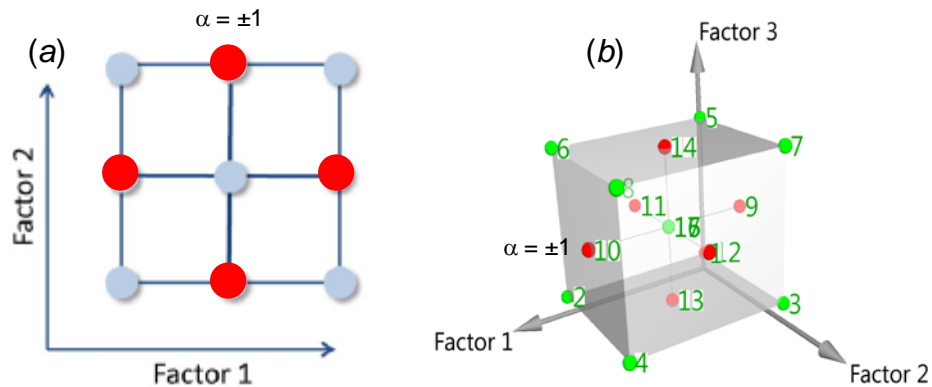
experimental designs used for the factors optimization in this PhD thesis will be described in details.

### ***Central Composite design (CCD)***

Two-level designs can only lead to linear models of responses and so cannot give information about maxima or any non-linear relationships. Moreover, for full factorial designs conducted at levels greater than two, the number of required experiments to achieve a suitable description of the model becomes challenging for a concrete application. Designs that allow greater numbers of levels cover the factor space near the centre with more points than at the periphery, without performing experiments at every combination of factors and levels. One such design is the central composite design, so named because it combines a two-level factorial design with a star design and centre points. The star ( $\pm\alpha$ ) and factorial points can lie equidistant from the centre (*circumscribed* design - CCC, see Figure 15), or the star points can lie within the space of the factorial design (*inscribed* design), or they can lie on the faces of the factorial design points (*faceted* design - CCF, see Figure 16). Central composite designs require  $L^k + Lk + n_c$  experiment [58]. Thus it results composed by a full factorial design ( $L^k$ ), or a fraction factorial design if the numbers of runs is higher ( $L^{k-p}$ ), to estimate the linear coefficient of factors ( $\beta_i$ ) and their interactions ( $\beta_{ij}$ ). Additional  $2k$  experiments, also called star point, symmetrically distributed at  $\pm\alpha$  along the factors axis and used for the estimation of the quadratic effect coefficients ( $\beta_{ii}$ ). The  $\alpha$  value depends by the number of experiment due to factorial term. Namely, for a 2-factor design  $\alpha = \pm 1.414$  (half diagonal of the square experimental domain); for a 3-factor design  $\alpha = \pm 1.627$  (half diagonal of the cubic experimental domain); and so on for larger factor numbers. Finally,  $n_c$  are the number of replicate centre points chosen. The centre point replicates are introduced for the experimental variance estimation. On the following example figures, the star points ( $\alpha$ ) of CCC and CCF design are coloured by red.



**Figure 15:** CCC design for two (a) and three (b) factors



**Figure 16:** CCF design for two (a) and three (b) factors

## Design variability and quality

The model error estimation is used to estimate if a factor has a higher / lower effect respect the pure model error. In this context, the variability associated to the measurement of the response must be taken into account and used to determine the number of replicates of each measurement of the response to obtain reliable information from the model. Predictive models, determined by multi-linear regression, are tested by ANOVA. Analysis of variance (ANOVA) should be used while analysing RSM designs. ANOVA is used to analyse the results depending by simultaneous factors variation, evaluating significance and lack of fit (validity) of the regression model used, by partition of the total variation of



a selected response SS (SS - *Sum of Squares corrected for the mean*) into a part due to the regression model and a part due to the residuals:  $SS_{Tot} = SS_{regr} + SS_{resid}$ . If there are replicated observations (experiments), the residual sum of squares is further partitioned into pure error ( $SS_{pe}$ ) and Lack of fit ( $SS_{lof}$ ):  $SS_{Tot} = SS_{regr} + SS_{pe} + SS_{lof}$  [52] [53] [61].

Analysing the quality of the design and regression model, it is to be controlled as well determination coefficient ( $R^2$ ), adjusted- $R^2$  ( $R^2_{adj}$ ) and  $Q^2$  values [62]. Few general concept are following describe for the R and Q square values to understand the meaning and the using on experimental section, without deeper discussions on computation.

$R^2$  estimate goodness of fitting for the multi-linear regression. Determination coefficient is the proportion of the variance in the responses (dependent variables) that is predictable from the method variables (independent factors). A  $R^2$  value equal to 1 corresponds to a perfect fit.

$Q^2$  estimate the prediction goodness of the identified regression model.  $Q^2$  value come from cross-validation and wants to estimate how accurately a predictive model will perform in practice. Dimensionally it could assume values from 0 to 1, rather than absolute value it is to consider in relationship of adjusted- $R^2$ .

$R^2_{adj}$  gives an idea of how many data points fall within the multi-linear regression model identify.  $R^2_{adj}$  is an adjustment of  $R^2$  taking into account the number of factors (predictors) and the total multivariate conditions tested, including replicates (sample size).



## Chapter 2: Vaccines

---

Vaccine administration has been demonstrated as one of the most effective and large scale applicable health-care approach in preventing infection disease. Due to successful immunization programs, some of the common diseases of the early 20<sup>th</sup> century almost disappeared [63]. Vaccines available today have an outstanding impact on human health, preventing every year over 2.5 million deaths caused by infection diseases worldwide [64] [65]. Development of vaccines against a variety of diseases, including diphtheria, tetanus, polio, measles, mumps, rubella, hepatitis B, and meningitis, have reduced the associated mortality by 97–99% [66]. In the near future the beneficial impact of vaccination will further increase thanks to several factors including: the enhanced coverage of children vaccination in developing countries promoted by Expanded Program on Immunization and by the Global Alliance for Vaccines and Immunization (GAVI); the increased use of vaccines in elderly and pregnant women; the full implementation of vaccines that have been recently licensed (e.g., Meningococcus B and Dengue); the launch of novel vaccines that have successfully concluded phase III trials (e.g., Malaria and Zoster) [65]. However, even with multiple successful vaccination campaigns, infectious diseases remain the second leading cause of death worldwide, disproportionately affecting children under the age of 5 and people in low income countries [67] [68]. A large number of infectious diseases caused by viruses, bacteria (often resistant to multiple antibiotics) or parasites are not yet preventable by vaccination and are responsible for millions of death every year. For the majority of these targets, traditional approaches have failed and there is a need to apply innovative science and novel technologies to develop effective vaccines. In addition, people in the world continue to experience pathogen outbreaks such as Pandemic Influenza, Ebola and, more recently, Zika virus, which have highlighted the need of more rapid vaccine discovery and development strategies supported by breakthrough scientific innovation [65]. In fact, five of the top ten leading causes of death in low income countries are caused by infectious agents: lower respiratory infections (e.g. pneumonia), HIV/AIDS,



diarrheal disease, malaria, and tuberculosis. While some of these pathogens currently lack a vaccine necessary for disease control, an estimated 20% of these deaths still results from vaccine-preventable diseases, indicating the need for substantial improvement in vaccine technology and administration [67] [69] [70]. Hence, innovative strategies are essential in development of novel vaccines [63] to rationally design effective vaccines where drug-based conventional approaches have failed [71]. Most of the promising strategies include design of an appropriate recombinant antigens as well as development of properly formulated vaccines to obtain the appropriate immune-response and the sufficient potency. Any case, a well-designed analytical control strategy is important to ensure an appropriate monitoring of the vaccine product quality.

In the present thesis work the AQbD framework was applied for the development of new analytical methods related to some critical quality attributes (CQAs) of *Neisseria meningitidis* serogroup B vaccine. Bacterial meningitis is an infection of the membranes and cerebrospinal fluid surrounding the brain and spinal cord and it is a major cause of death and disability worldwide. Three organisms are responsible for most cases of bacterial meningitis: *Neisseria meningitidis*, *Haemophilus influenzae* type b and *Streptococcus pneumoniae* [72] [73] [74] [75]. *Neisseria meningitidis* is a pathogen bacterium that is transmitted through contact with respiratory droplets. Transmission and colonization typically results in asymptomatic carriage in the upper respiratory tract, leading to bacteraemia that can quickly become life-threatening invasive meningococcal disease, which most often presents meningitis and/or septicaemia, and less commonly pneumonia, septic arthritis, otitis media and epiglottitis [76]. *Neisseria meningitidis* is classified into serogroups based on the immunological reactivity of the bacteria capsular polysaccharide. Meningococcal serogroups A, B, C, W, Y and recently X account for the majority of meningococcal diseases [77], with serogroup B (MenB) being now the most prominent cause of infant bacterial meningitis and septicaemia in Europe, Latin America, US and Canada [78] [79]. Serogroup B capsular polysaccharide (CPS), that consists of a homopolymer of  $\alpha$  (2-8)-linked polysialic acid, is immunologically similar to that of neural-cell adhesion molecules and thus is poorly immunogenic,



hindering its use in the traditional polysaccharide conjugate-vaccine approach [79]. To overcome this problem the new genome-based approach of reverse vaccinology was employed to identify new candidates for the development of a vaccine against Meningitidis B disease [80]. The application of reverse vaccinology to MenB vaccine development allowed the identification of bacterial proteins able to induce bactericidal antibodies [65].

The genomic era has radically changed the vaccine development approach. Reverse vaccinology defines the process of antigen discovery starting from genome information, by the availability of whole genome sequences. From its first application to *Neisseria meningitidis* group B, this approach has gradually evolved and is now accepted as a successful method of vaccine discovery, as it can be exploited to develop vaccines against many types of pathogens. Current reverse vaccinology approaches include comparative in silico analyses of multiple genome sequences, in order to identify conserved antigens within a heterogeneous pathogen population. The purpose is the identification of antigens that are unique to pathogenic isolates but not present in commensal strains. In addition, transcriptomic and proteomic data sets are integrated into a selection process that yields a short list of candidate antigens to be tested in animal models, thus reducing the costs and time of downstream analyses [81].

## **Bexsero vaccine (4CMenB)**

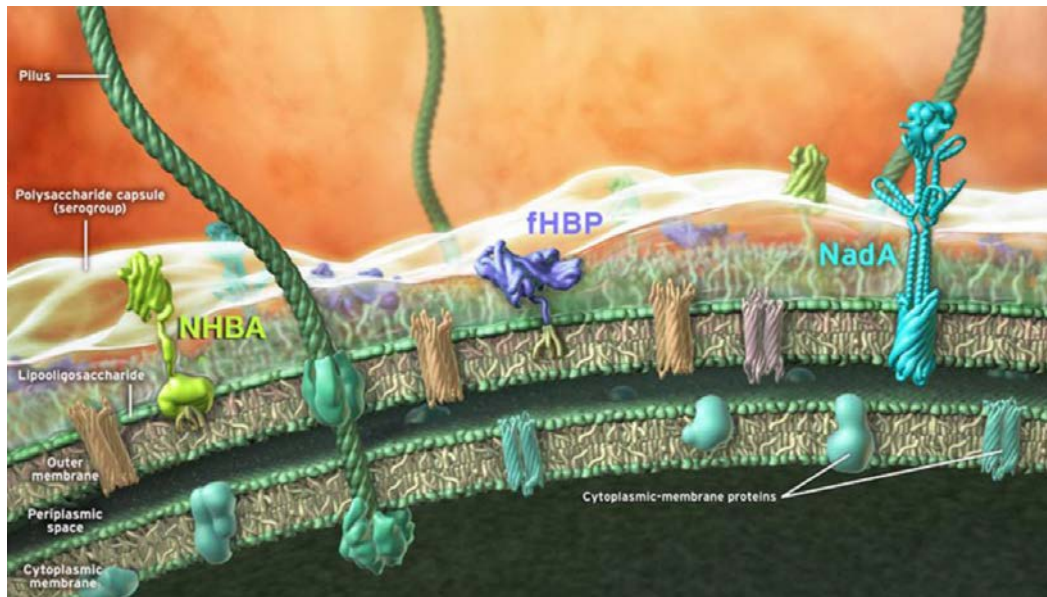
Bexsero is the first approved vaccine for active immunization of individuals from 2 months of age and older to prevent invasive disease caused by *Neisseria meningitidis* serogroup B. Using reverse vaccinology, three MenB surface-exposed proteins were identified as potentially immunogenic antigens: Neisseria Heparin Binding Antigen (NHBA), factor H binding protein (fHbp) and Neisseria adhesin A (NadA), constituting the three core proteins of recombinant meningococcal B vaccine (rMenB). NHBA and fHbp have been fused to two additional antigens (Genome-derived Neisseria Antigen: GNA1030 and GNA2091, respectively) to increase their immunogenicity [82] [83] [84].





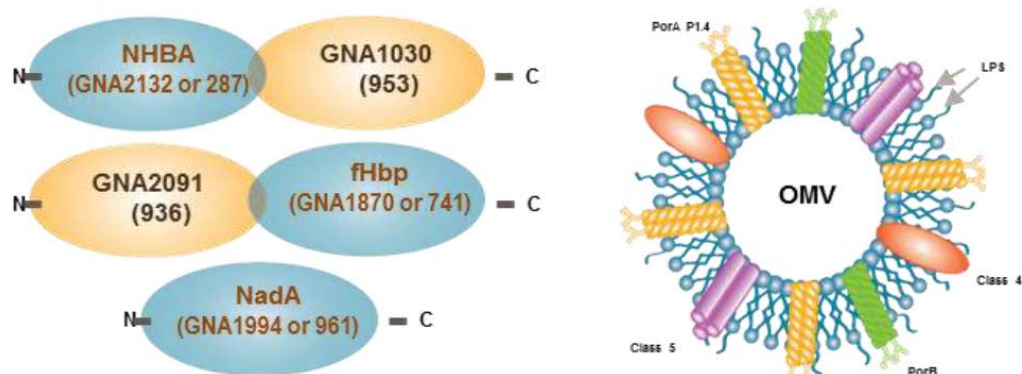
NHBA-GNA1030, fHbp-GNA2091 and NadA, have been combined with Outer Membrane Vesicles (OMV), mimicking the vesicles naturally released by *Neisseria meningitidis* and displaying protein antigens in a context similar to their native environment. In particular OMV contains Porin A (PorA) and Porin B (PorB), the most abundant outer membrane proteins. OMV have been shown to be safe and efficacious in many clinical trials and effective to MenB outbreaks, by inducing immunity that is mostly due to the highly variable PorA outer membrane protein, but can also involve outer membrane proteins PorB, OmpC, FetA, and Lipooligosaccharides [84] [85]. The rMenB components of the vaccine (HBA-GNA1030, fHbp-GNA2091 and NadA) are produced in *Escherichia coli* by recombinant DNA technology, while Outer Membrane Vesicles (expressing Porin A and Porin B proteins) are produced by fermentation of *Neisseria meningitidis* strain NZ98/254. The three rMenB antigens and OMV are formulated in a four component vaccine, named 4CMenB (Bexsero, GSK), where all these active components are adsorbed on aluminium hydroxide and the unadsorbed antigens content is a product critical quality attribute. Bexsero vaccine is the first MenB vaccine based on recombinant proteins able to elicit a robust bactericidal immune response in adults, adolescents and infants against a broad range of isolated serogroup B [82]. Bexsero has been recently licensed for use in Europe, US and elsewhere [86] [87] [88].

A schematic representation of the Bexsero vaccine antigens on the surface of *Neisseria meningitidis* is provided in Figure 17. The different bacterial compartments (outer membrane, periplasmic space, cytoplasmic membrane) and the main antigens identified through reverse vaccinology approach (NHBA, fHbp and NadA) are depicted. The representation of NHBA and fHbp in the picture is derived from the NMR structural data available and reported as cartoon. NadA is a model based on the structural homology with other members of the Oca family. Other components of the meningococcal membranes are also shown (pilus, polysaccharide capsule, lipooligosaccharide and integral inner and outer membrane proteins) [82].



**Figure 17:** schematic representation of the Bexsero antigens on the surface of *N. meningitidis* [82].

The structure of the recombinant antigens in the Bexsero vaccine are well characterized by spectroscopy techniques (NMR, X-Ray, etc...) and several publications are available for a deep understanding of the vaccine components [82] [89] [90] [91] [92] [93] [94] [95]. A schematic representation of the Bexsero vaccine formulation is following provided by Figure 18. The image shows the main antigens identified through Reverse Vaccinology approach (NHBA, fHbp and NadA) on the left and the OMV on the right [89].



**Figure 18:** schematic representation of the Bexsero vaccine formulation [89].



## Chapter 3: Analytical QbD

---

The application of Analytical Quality by Design (AQbD) to methods for testing vaccines is the core scope of the present experimental thesis study, in order to build quality into the assay during its development. The objective should be achieved by controlling method performances within predefined boundaries that ensure quality expectations are met. AQbD approach represents an useful tool to understand and develop the assay quality from the early method development stages, without waiting the final performance verification by method qualification and validation activities. The knowledge expected from this methodology provides several advantages, in terms of data confidence, method robustness and, additionally, in terms of assay failure probability evaluation. The application of AQbD approach during development contributes to decrease the probabilities of faults both during the validation and the routine use, with time and costs saving.

In 2004, the US Food and Drug Administration outlined a new science- and risk-based approach that encourages manufacturers to develop robust processes and appropriate control strategies, thus supporting continuous improvement and product quality [96]. Over the years this approach has been evolved by the regulatory authorities and the pharmaceutical industry to establish a core concept called Quality by Design (QbD), now formally recommended and supported by International Conference on Harmonisation (ICH) guidelines [2] [3] [4] [5]. The ICHQ8(R2) guidance defines QbD as “*a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management*” [2]. So far, in the field of vaccine production, QbD principles have been mainly applied to accelerate process development to manufacture a vaccine candidate at commercial scale [97].

Many of the concepts associated with QbD for the product and the manufacturing process can be as well considered in the analytical method development [7] [98] [99]. The analytical control strategy for vaccine development starts from the definition of the *Critical Quality Attributes*



(CQAs) of the vaccine, namely the characteristics of the product that could affect the safety and the efficacy of the vaccine. However, as the CQAs may not yet be known in the early development stage, a risk-based approach should be followed in developing the control strategy [100]. The starting point of the AQbD method development is the *Analytical Target Profile* (ATP). For each product attribute to be tested, a specific ATP should be generated and formalized in order to define (I) the intended purpose of the assay measurement, (II) the *Critical Method Attributes* (CMAs) to be measured to ensure the assay performance and (III) the confidence and quality level with which we want to control the CMAs. Hence, starting from the ATP, AQbD emphasizes the need to thoroughly understand the candidate analytical system by an in-depth understanding of *Critical Method Parameters* (CMPs) based on risk assessment tools and experimental multivariate studies for managing the risk of failure [52] [101]. Such AQbD framework leads to the definition of the analytical design space, also known as *Method Operable Design Region* (MODR). MODR is the core of the AQbD approach. Adapting the ICHQ8(R2) definition set for the “*process design space*”, MODR could be defined as the multidimensional region of successful operating combinations and interactions of CMPs (method input variables), which lead to desired values for critical method attributes (CMAs), providing assurance of method quality [2] [7]. The MODR represents a step up for building the method life cycle and control strategy, since ensures the intended performance and the quality of the method. MODR prevents the generation of inaccurate results and help to identify the method failure modes. The concept of failure probability is the added value of the AQbD approach on pharmaceutical method development. AQbD systematically investigates CMPs leading to an increased knowledge of their effects on the CMAs and establish a reduced variability by controlling assay conditions and risks [98]. Additional advantages, due to MODR knowledge, are related to regulatory flexibility since supports that [102]: modifications of operative conditions within the MODR are not considered method changes.

The conceptual AQbD flowchart that has been proposed and adapted in the present thesis work is reported in Figure 19. A parallelism between application of QbD for process and AQbD is also described in Table 5.

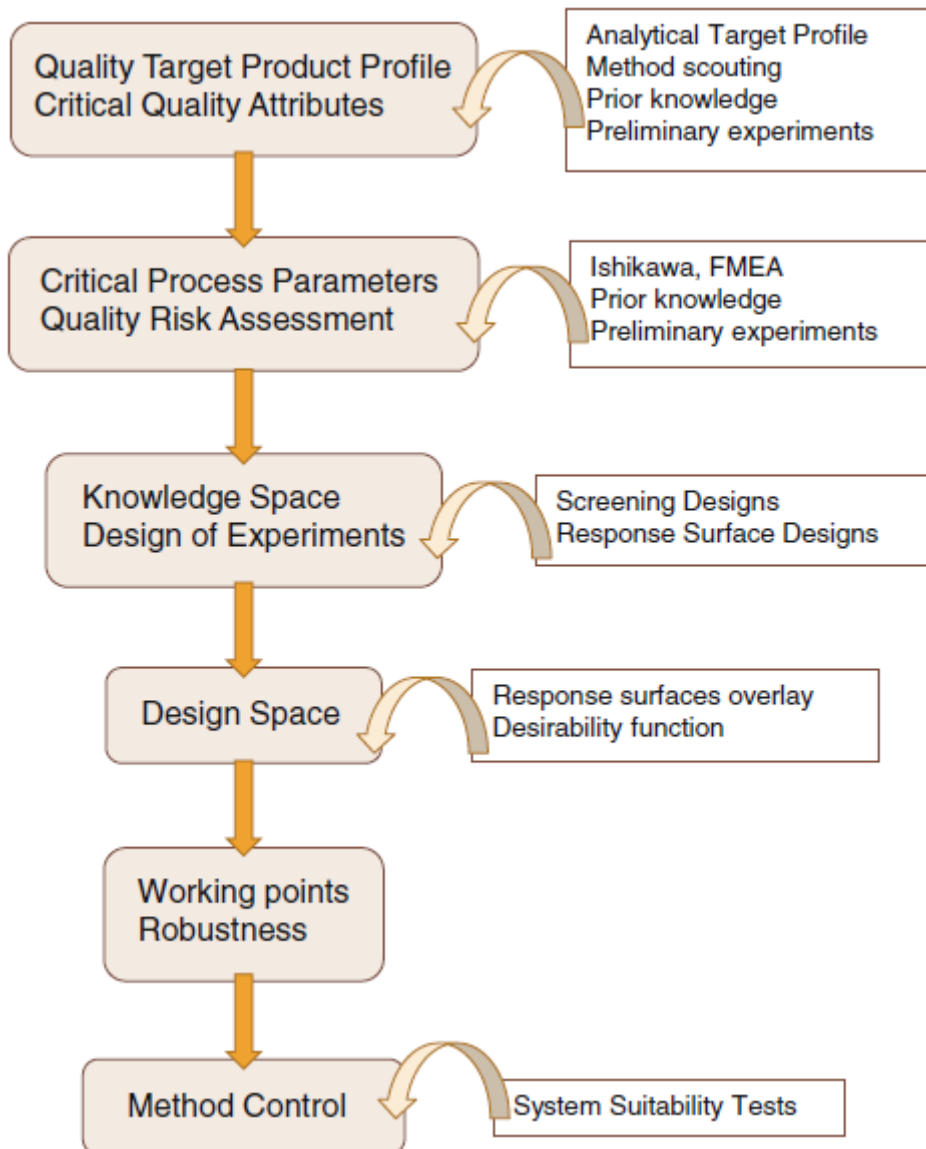


Figure 19: AQbD flowchart [7]

Recent examples of AQbD applications to pharmaceutical field, mainly concerning the development of separation methods, adapting a similar flowchart are reported in literature for HPLC [103] [104] [105], UHPLC [106] [107], hydrophilic interaction liquid chromatography [108], supercritical fluid chromatography [109], capillary zone electrophoresis [110] [111] [112] [113], micellar electrokinetic chromatography [114] [115] [116] and microemulsion electrokinetic chromatography [117] [118].



Process concept	Analytical counterpart
Target Product Profile (TPP)	Analytical Target profile (ATP)
<ul style="list-style-type: none"> <li>• Target clinical performance, manufacturing, and commercial requirements</li> </ul>	<ul style="list-style-type: none"> <li>• Target analytical performance, testing laboratory, and customer requirements</li> </ul>
Critical Quality Attributes (CQAs)	Critical Method Attributes (CMAs)
<ul style="list-style-type: none"> <li>• Potency</li> </ul>	<ul style="list-style-type: none"> <li>• Precision</li> </ul>
<ul style="list-style-type: none"> <li>• Aggregation</li> </ul>	<ul style="list-style-type: none"> <li>• Sensitivity</li> </ul>
<ul style="list-style-type: none"> <li>• Purity</li> </ul>	<ul style="list-style-type: none"> <li>• Accuracy</li> </ul>
Specifications (acceptance criteria)	Acceptance criteria
<ul style="list-style-type: none"> <li>• 80–125 % potency</li> </ul>	<ul style="list-style-type: none"> <li>• %GCV &lt; 10 %</li> </ul>
<ul style="list-style-type: none"> <li>• Purity &gt; 95 %</li> </ul>	<ul style="list-style-type: none"> <li>• LLOQ &gt; 1 ng/mL</li> </ul>
Critical process parameters (CPPs)	Critical method parameters (CMPs)
<ul style="list-style-type: none"> <li>• pH, time, temperature</li> </ul>	<ul style="list-style-type: none"> <li>• pH, time, temperature</li> </ul>
Process control strategy	Method control strategy
<ul style="list-style-type: none"> <li>• Comparability protocols</li> </ul>	<ul style="list-style-type: none"> <li>• Method comparability or bridging protocols</li> </ul>
<ul style="list-style-type: none"> <li>• Process technology transfer</li> </ul>	<ul style="list-style-type: none"> <li>• Method transfer</li> </ul>
Continuous process verification	Analytical method maintenance
<ul style="list-style-type: none"> <li>• Continuous review and updating of process knowledge</li> </ul>	<ul style="list-style-type: none"> <li>• Continuous review and updating of analytical knowledge</li> </ul>

LLOQ lower limit of quantification; GCV geometric coefficient of variation

**Table 5:** Process and analytical QbD comparison [98].

## ATP

In the AQbD context the upfront definition of the desired method performance, by formalization of the Analytical Target Profile (ATP), is crucial. Independently by the assay that will be chosen by screening of different candidate technologies to approach the method development, it is necessary to define the requirements of the measurement itself. The ATP includes, beyond a tight definition of what has to be measured, a set of criteria that define the target assay performance in terms of accuracy, precision, range, specificity, etc. ATP requirements are defined in consideration of the context and objective of the test, in order to ensure that an assay will be able to appropriately measure the product attribute. In other words, the ATP defines the critical method attributes (CMAs) and the respective acceptance ranges / limits, including the associated confidence levels (probabilistic confidence that method actually ensure those acceptance criteria).





However, for the analytical scientist approaching the method development, it may be important to define surrogate attributes for the assay that will be developed. Similar concepts concerning CMAs surrogate measurements and reportable values are introduced also by USP chapter <1220> [119].

The definition of the desirability values and acceptance ranges for the method requirements is crucial to lead the definition of the final analytical design space, by impacting the desirability function for response surface methodology. Furthermore, the ATP should represent the key mechanism for the prioritization between method responses, avoiding different degree of subjectivity. Hence, ATP gives robustness to decision making process.

The ATP should be drafted as soon as possible in the method development life-cycle, as a life document for guiding method evolution, amended as the method requirements evolve in line with improvements in product knowledge or in manufacturing process development. In parallel with ATP (that is focus on method performance related to the quality of the data) practical and feasibility requirements (focus on method applicability at the final lab, e.g. throughput, time / costs, simplicity, expected percentage of invalid test, etc...) should be also defined. Depending on the method scope (e.g. release, stability, characterization, in-process control) the practical and feasibility requirements can be crucial for driving method selection and assay development.

## **Method Scouting**

Starting from ATP and keeping also in consideration the feasibility requirements, as appropriate, the method scouting is the next step to approach the development for a CQA measure and control. The method selection process begins with an evaluation of all the possible assays potentially applicable to the measurement, identified on the base of previous experiences and the literature data. After an appropriate scouting to identify the most promising methods among the previously listed analytical technologies, an experimental screening has to be



executed to preliminarily verify the candidate methods performance vs. the criteria set for predefined critical method attributes.

Different approaches can be applied for screening phase to identify method/s potentially suitable for the intended use. One-factor-at-time (OFAT) or more structured experimental designs (DoEs) could be employed. Both method performance (ATP requirements) and feasibility criteria have to be considered to lead the method selection. During the product life cycle, the same assay could be screened several times to verify its fitting for purpose in different project phases and contemporary with respect to technological innovations; allowing the identification of technological innovations and their implementation.

The technical innovations help the analytical scientist to optimize the methods screening phase. High-throughput and automatized instruments facilitate the simultaneous testing of multiple technologies.

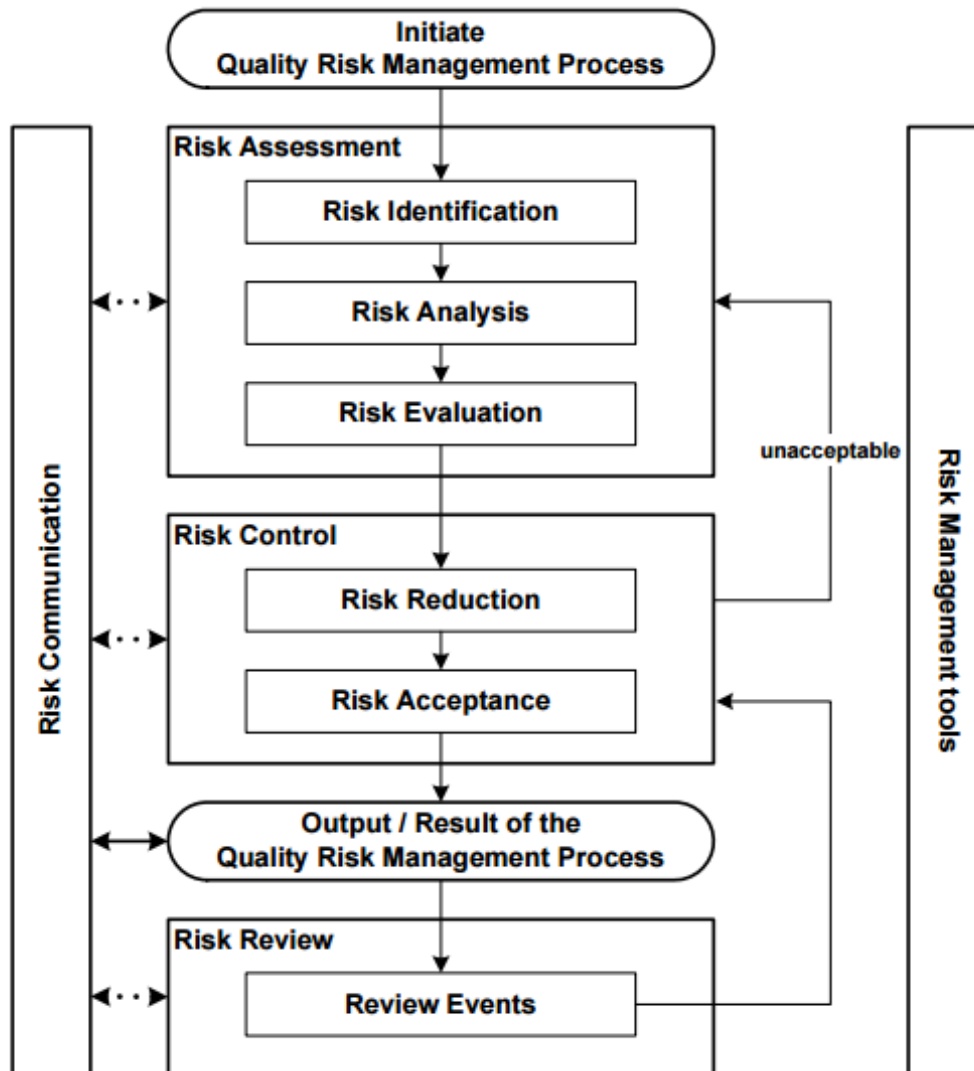
## **Risk Assessment**

Risk assessment (RA) is the first stage of a risk management process, as defined in ICHQ9 guideline [3] (Figure 20). The guidance defines the risk assessment as “*A systematic process of organizing information to support a risk decision to be made within a risk management process. It consists of the identification of hazards and the analysis and evaluation of risks associated with exposure to those hazards*” [3]. The concept is applicable to AQbD framework, where analytical methods hazards correspond to “unreliable results” [98].

The risk assessment goal is to identify, analyse and evaluate the method factors with respect to a potential impact on the measurement, by theoretical evaluations based on experience, literature and previous knowledge to finally provide a list of potential critical method parameters (pCMPs). Criticality will be then confirmed (or not) by appropriate experimental activities (see next section). By the risk assessment exercise, the initial list of parameters related to the method and which can affect CMAs can be quite extensive. The parameters list can be



significantly and confidently reduced, by using the proper risk assessment tools [4], for example Ishikawa diagrams (aka fishbone), cause and effect matrix (C&E), process map, failure mode effects analysis (FMEA) and many others.



**Figure 20:** typical quality risk management process [3]

In the present thesis work combinations of fishbone diagram, process map and C&E matrix tools has been selected for the risk assessment exercise (identification, analysis and evaluation of risks / impacts). The fishbone diagram is helpful to identify, list and cluster all the method parameters. The design of a process map could be useful to brainstorm the method. The C&E matrix is powerful for ranking the impact of each method parameter and identifying the potentially critical (pCMPs) to be



studied by experimental design. Practical examples of those tools have been provided in the chapter 4.

## **Knowledge Space by DoEs**

After the theoretical exercise of risk assessment, the method quality risk control continues with the screening experimentation to evaluate the criticality of the identified potentially critical method parameters. Screening studies, by means of DoE-effect analysis (see the example Figure 8), can be used to evaluate the effect of the higher-ranked method parameters (pCMPs) and to finally establish the factors that have a greater effect on CMAs and / or their surrogates. The objective of the screening phase is to identify the most important pCMPs (the most impacting, expected to be a limited number) with respect to method parameters that have not or small impact. Thus allows the identification of the pCMPs that will be later studied in detail by Response Surface Methodology (RSM). A first assessment of the method performance and factors criticality is obtained at this step.

The ideal approach for screening is to perform a minimum number of experiments on a maximum number of factors to build knowledge about the method, by determining cause-effect relationships between pCMPs and CMAs, including factors interactions. DoE approaches yield three major advantages over OFAT: (I) the total number of runs is typically smaller and (II) the statistical power (i.e. the ability of detecting significant effects) of the study is higher, (III) information about factor interactions is also obtained. Nevertheless, OFAT studies can also be used in appropriate cases, for example when no method parameters interactions are supposed, or when it is practically unfeasible to combine different levels for different parameters.

The design region explored by DoEs (ranges studied for each parameter) should be as broader as possible, compatibly with the assay feasibility, applicability and previous experiences on similar assays. The



experimental region investigated by DoE, is so called *knowledge space* (KS).

Screening designs are also applicable to discontinuous parameters, e.g. column, buffer and solvent choice. As deeply discussed on experimental design chapter, Screening DoEs generally involve testing of only two levels of the parameters under study (i.e. a low and an high values) and should always include several repeats of a central condition (also called “centre point”, chosen as reference conditions), run in order to obtain a first estimate of the residual noise.

## **RSM for MODR**

As the influence of the method parameters onto the CMAs has been qualitatively understood by the screening study performed to identify the CMPs, the method development proceeds with the definition of the analytical design space (MODR: Method Operable Design Region), that is the core of the AQbD approach. Such multi-dimensional space defines the combination of the method parameters within which the method requirements are satisfied, with predefined confidence, to ensure the desired method performance and the data quality. The design space boundaries are the so-called edges of failure, outside which the method performances are not acceptable. This approach supports the idea that the analytical method should not be designed (and potentially qualified) in one fixed condition, but verified under a range of conditions around the set point. In this context, any method deviations or modifications occurring within the design space, where the method performance and robustness are ensured, should not consider “method changes”. Hence allows regulatory advantages and flexibility.

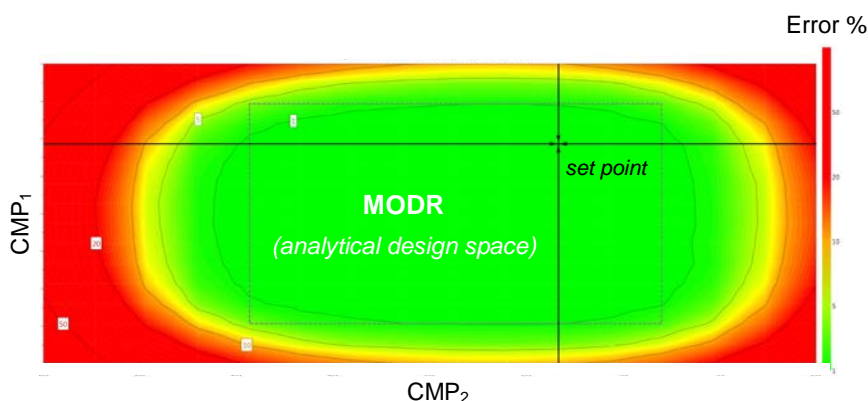
Usually, the MODR can be identified using an optimization DoE (also called response surface methodology). RSM models explore the relationships between CMPs and one or more method responses (an attribute or a surrogate). The purpose of the RSM is to obtain experimental knowledge on responses for an appropriate set of



multivariate conditions of method parameters to predict the responses over the whole experimental domain.

With respect to the space investigated by screening DoE, generally the space to be considered for establishing the MODR can be restricted around the best performing sub-region previously identified. All the predefined method requirements need to be satisfied within the MODR. For this purpose, the most useful approaches in elaborating the experimental data to identify a combined design space are response overlay and desirability functions, taking also into account the uncertainty of the RSMs predictions (e.g. via tolerance intervals). The uncertainty cut-off must also be defined in order to ensure the desired confidence of the identified design space. Monte-Carlo simulations of the propagation of the model's predictive errors (as it has been applied in the present thesis work), or alternative Bayesian approaches, may be considered for estimating the risk of not meeting the requirements.

The analytical design space can be mathematically represented by equations describing the effects of method parameters on the response. [2] (e.g. by Equation 7 for multi-linear regression models). In the real analytical development, multiple CMAs and other responses are controlled at the same time and the mathematical model could be relatively complex and difficult to be interpreted from a physical perspective. For the analyst scientist it is more practical and easier to present the MODR graphically by using a probabilistic surface plot, such as shown in the example Figure 21. The surface plot makes an immediate visualization of the design *hyper*-space and the possible combination of method parameters (CMPs) able to satisfy the ATP requirements. i.e. to provide quality for data with the required confidence level. In the example, the green area represents to the MODR for  $CMP_1$  and  $CMP_2$  with a confidence level of 95% (the probability of failure allowed is lower or equal to 5%).



**Figure 21:** graphical presentation of MODR by probabilistic surface plot

The RSM study provides also the opportunity to refine the criticality assessment for each parameter under study. Namely, the criticality level is finally estimated by the response surface model for MODR, through the evaluation of effects and interactions of method parameters on responses.

Although the MODR is in principle expected as a curve space, from practical point of view, MODR is conveniently expressed as a set of ranges, one range for each CMP. For the MODR definition (I) the method performance (based on ATP criteria) is a “must do” requirement. However, we need to also consider other important factors such as (II) analytical feasibility and practicability considerations, (III) assay needs and peculiarities and (IV) robustness and controllability of the method parameters, to finally restrict MODR to a more practical space.

All the points within the MODR hyperspace represent potential acceptable working conditions. However, definition of a set point for describing the assay standard operative conditions has to be defined for routine use. The selection of the optimal set point conditions, and the surrounding control space in which the method should be operated during routine application, can be based on:

- desirability function (possibly weighing different responses);
- multiple response regression models (taking into account the correlation between CMAs);
- practical considerations (e.g. robustness indicated by plateau in response surface);



- feasibility considerations (e.g. feasibility for preparations, costs, toxicity of materials, et...)
- analytical requirements for routine use (e.g. high-throughput and/or time of analysis for in-process test, etc...);
- failure probability (e.g. low risk of failure for the prediction model).

## **Model Confirmation & Robustness**

The statistical models used to drive the MODR should be experimentally validated by executing a subset of runs through the identified analytical design space.

In order to validate the RSM model, the MODR predictions are verified by testing the boundary ranges of the identified hyperspace, since the risk of failure of the response surface prediction model is higher at the edges of the space. It is a good rule to also include an experimental verification of the selected working point (candidate set point for routine use). The verification of the MODR can be achieved by a new designing for testing different multivariate conditions respect to conditions tested by Optimization DoE. The MODR is so confirmed by comparing the verification data (experimentally obtained) with the design space model predicted data.

The analytical design space itself can be considered as a theoretical robust area, since within MODR variation of the factors condition does not significantly influence the quality of the method [99] [120] [121]. Robustness is an integral part of development, contrarily to past traditional OFAT approach, where robustness testing was performed at the end of the development process (during the assay validation) with a higher risk of failure. Therefore, the final step of the AQBd approach is the verification of the method robustness within the MODR boundaries. The DoE robustness test is not redundant, but it is still necessary to verify if the chosen factor ranges are suitable and could be applied as control space for the assay routine use. The robustness DoE makes possible to identify the so called control space, a robust and controlled space within



the design space where the analytical performance is not affected by the method factors variation. An example of robustness by DoE-effect analysis has been provided on previous Figure 9. Thus, the control space boundaries should be moved inside the design space to accommodate robustness and optimize the operating space. As discussed in chapter 1, Plackett-Burman design has become particularly popular for the robustness tests and with the same experimental design it is possible to perform also the confirmation/validation of the design space model.

## **Control Strategy & Life Cycle**

System suitability testing (SST) is an integral part of analytical procedures. System suitability parameters to be established for a particular procedure depend on the type of procedure being validated. For chromatographic analysis, USP <621> [122] and European Pharmacopoeia (EP) chapter 2.2.41 [123] have both specified requirements for SST to demonstrate that a chromatograph is fit for the analysis. System suitability are run each time an analysis is undertaken and each SST is specific for an individual method with pre-defined acceptance criteria (e.g. precision, peak shape, resolutions, and many others parameters on the base of assay peculiarities).

Definition of a SST and the related acceptance criteria is the first step to design an adequate control strategy for routinely applying the developed method. The responses obtained from the worst-case results of the robustness DoE can be used to determine the SST limits. The SST ensures the confidence of the method results only in case the method has been demonstrated robust. For a long term control strategy, the SST supports the monitor of the method performance. A control strategy based on SST needs to be completed and consolidated during the method life cycle through a strategic Continued Procedure Performance Verification, to ensure that the method performance (quality of the data produced by the assay) and the control of critical method parameters still remain under control during the method life cycle [98].



## Chapter 4: Experimental activities

---

In the present thesis study, the quality by design approach was applied to design analytical methods aimed to the quality control of a vaccine product. The scope of this research work is the implementation and application of QbD principles to analytical method screening and development for vaccines analysis. It is the first time in literature that AQbD is applied to a vaccine product. Bexsero is the vaccine chosen for the analytical researches as application of AQbD framework.

AQbD principles were applied to development of new UHPLC methods for the control of the Bexsero vaccine CQAs. Three different analytical applications were investigated to improve the analytical control strategy of Bexsero and identify new methods potentially applicable for:

- 1) Not adsorbed antigen content determination by RP-UHPLC;
- 2) rMenB quantification assay by Amino Acid Analysis (A.A.A);
- 3) OMV protein pattern analysis by RP-UHPLC.

The AQbD quality risk management framework, used for the three analytical methods, builds adequate method knowledge to ensure regulatory flexibility during lifecycle, as well as method robustness and safety for product quality control.

### **RP-UHPLC Adsorption percentage**

A fundamental critical quality attribute of Bexsero vaccine is the unadsorbed antigens content. The actual method used for unadsorbed antigens determination in the commercial vaccine product is a sodium dodecyl sulphate - polyacrylamide gel electrophoresis (SDS-PAGE) test. Following the analytical target profile requirements, the AQbD framework was applied increasing the assays throughput by reducing the manual operations, improving selectivity and sensitivity with respect to the current SDS-PAGE assay used to control the above CQA.





### ***Chemicals and reagents***

For the development of the chromatographic method a mock standard solution was used. The mock solution was a mixture with the same composition of antigens and excipients of the vaccine product, apart from the aluminium hydroxide adsorbent which was removed for analytical needs and apart from the addition of 0.15% (p/v) Zwittergent 3-14 detergent. The mock solution was used as calibration for the analysis and its composition enabled the target concentration range to be achieved. The drug substances used for mock formulation were produced by GSK group of companies (Siena, Italy). L-histidine  $\geq 99\%$  (ReagentPlus grade), trifluoroacetic acid (TFA)  $\geq 99\%$  (LCMS & HPLC grade), methanol (CH<sub>3</sub>OH)  $\geq 99.9\%$  (HPLC grade), tween 80 (cell Biology grade) and sucrose (C<sub>12</sub>H<sub>22</sub>O<sub>11</sub>)  $\geq 99.5\%$  (BioUltra grade) were purchased by Sigma-Aldrich (Saint Louis, MO, USA). Sodium chloride (NaCl) 99.99% (Suprapur grade), potassium di-hydrogen phosphate (KH<sub>2</sub>PO<sub>4</sub>, EMSURE ISO grade), Hydrochloric acid fuming (HCl) 37% (ACS ISO Reag.Ph.Eur. grade), Potassium hydroxide (KOH) 45% w/w solution (Reagent grade) and Zwittergent 3-14 detergent (EMSURE ISO grade) were purchased by Merck KGaA (Darmstadt, Germany). Acetonitrile (ACN) 99.8% (LC-MS grade) was purchased by Panreac (Radnor, PA, USA). Ultrapure water was produced by Millipore Milli-Q system (Billerica, MA, USA) and filtered on a nylon membrane of 0.22  $\mu\text{m}$  porosity using Nalgene clepsydra filters (Nalgene, Rochester, NY, USA).

### ***Solutions and sample preparation***

The L-Histidine buffers and the phosphate buffers were adjusted to the proper pH by adding HCl and KOH solutions, respectively. A 5% Zwittergent 3-14 detergent solution was prepared in pH 6.5 phosphate buffer 1M.

The mock standard solutions were prepared each day by dilution of the proper volumes of the drug substance bulks (each aliquot of rMenB proteins stored at -20°C and of OMV sample stored at 4-8°C) in 200 mM L-Histidine pH 6.3 buffer plus 90 mg ml<sup>-1</sup> NaCl and 5% (p/v) sucrose solutions up to 100  $\mu\text{g ml}^{-1}$  rMenB proteins and 50  $\mu\text{g ml}^{-1}$  OMV sample. The working solutions were obtained by diluting the samples to the final concentration of 10  $\mu\text{g ml}^{-1}$



rMenB proteins and 5  $\mu\text{g ml}^{-1}$  OMV sample, maintaining the same matrix composition, with 0.15% (p/v) Zwittergent 3-14 detergent in pH 6.5 phosphate buffer 1M, added for the analytical purpose of applying the UHPLC procedure also to the antigen content determination. The working solutions were stored at 4-8°C before the analysis. The samples were stored in the autosampler using three different vials type: Clear Glass Total Recovery (Waters Corp., Milford, MA, USA), LCMS Certified Total Recovery vials (Waters Corp., Milford, MA, USA) and Polypropylene Total Recovery vials (Thermo Fisher Scientific, Waltham, MA, USA).

For mobile phase preparation, 500  $\mu\text{l}$  of TFA  $\geq 99\%$  were diluted up to 500 ml ultrapure water to obtain a 0.1% (v/v) TFA aqueous solution and 500  $\mu\text{l}$  of TFA  $\geq 99\%$  were added to 450 ml of ACN plus 49.5 ml of ultrapure water to prepare a 0.1% (v/v) TFA and 90% (v/v) ACN organic phase. All buffers and solutions were filtered by a nylon membrane of 0.22  $\mu\text{m}$  porosity using Nalgene filters (Nalgene).

### ***Chromatographic equipment and analysis***

Different chromatographic columns were screened: Acquity RP-C4 BEH 300Å, 1.7 $\mu\text{m}$ , 2.1x150 mm (C4pore) and Acquity UHPLC BEH 130Å C8, 1.7 $\mu\text{m}$ , 2.1x150 mm (C8pore) from Waters Corp. (Milford, MA, USA) and Aeris WIDEPOR C4 200Å, 3.6  $\mu\text{m}$ , 4.6x150 mm (C4shell) from Phenomenex (Torrance, CA, USA) [124].

For the method screening the NexeraX2 method scouting UHPLC series 30 system was used, equipped with LC-30AD pump, DGU-20A5R degasser unit and LPGE-unit, SIL-30AD autosampler, CTO-20AC oven with 180  $\mu\text{l}$  mixer and FCV-34AH UHPLC switching valve, SPD-M30A PDA detector with high sensitive flow-cell (85 mm; 9  $\mu\text{l}$ ) from Shimadzu Corp. (Kyoto, Japan). For response surface methodology (RSM) the Acquity H-Class Bio UHPLC system (Waters Corp.) was used, equipped with bio-Quaternary Solvent Manager (bioQSM) with 100  $\mu\text{l}$  mixer, bio-Sample Manager (bioSM-FTN) with 15  $\mu\text{l}$  injector needle and 50  $\mu\text{l}$  extension loop, column oven CH-A with pre-heater and photodiode array detector (ACQ-PDA) with analytical flow-cell (10 mm; 500 nL). The detection wavelength was 210 nm. Sample injections



were done by 30 µl of the working solution stored in the autosampler at 4-8°C. A new column was conditioned with the mobile phase for 60 min before starting the analysis. After the analysis the column was stored in pure ACN filtered on 0.22 µm nylon membrane.

The optimal separation of the antigens was achieved using the C4pore column. The working conditions (with the interval corresponding to the MODR) were as follows: starting organic phase concentration, 33.0% (32.0-34.6%) (%v/v); ramp time, 4.0 min (4.0-5.6 min) to 75% (v/v) organic phase final concentration; column temperature, 60°C (60-68°C). After each injection the column was washed with 90% of ACN for one minute and equilibrated for 3 min in the starting conditions.

### *Calculations and software*

The chromatographic resolutions (R) between two adjacent peaks were calculated using the retention times ( $t_R$ ) and the peak widths at half height (w), according to the following formula:

$$1.18(t_{R2} - t_{R1})/(w_1 + w_2).$$

The capacity factor ( $K'$ ) of NHBA-GNA1030 antigen, measurement of the retention time relative to column void volume ( $V_0$ ), was calculated according to the formula:

$$K' = (t_R - V_0) / V_0.$$

LabSolution Version 5 software equipped with Method Scouting start-up kit and licensed by Shimadzu Corp. [125] was used for the NexeraX2 UHPLC instrument control and for the chromatographic data computation in the screening phase. Empower3 software [126] licensed by Waters Corp. was used for the Acquity H-Class Bio UHPLC instrument control and for the chromatographic data computation in RSM.

Nemrod-W software [127] was used to generate the two asymmetric screening matrices used for investigating the knowledge space and the two Full Factorial Designs used for selecting the verification points at the edges of the design space and for testing robustness. MODDE software [128] was purchased from S-IN (Vicenza, Italy) and was employed to



generate the Central Composite circumscribed design (CCD) used for RSM, to perform data analysis and to find the design space by means of risk failure maps calculated using the Monte-Carlo simulations. The runs of all the DoE plans were carried out in a randomized order.

### ***Method development and results***

The method development followed the systematic approach of AQbD workflow for separation methods [7], involving the following steps:

- I) Analytical target profile definition, method scouting and definition of the CMAs;
- II) Quality risk assessment and identification of potential CMPs;
- III) Investigation of knowledge space by screening DoE;
- IV) RSM and definition of MODR;
- V) Working point definition and robustness testing;
- VI) Method control.

### ***Analytical target profile, method scouting and critical method attributes***

The analytical target profile is the intended purpose of the method and is defined by the selection of the analytes and of the analytical performances to be achieved [128] [129]. In this study, it consisted in obtaining the accurate quantitation of the five proteins NHBA-GNA1030, fHbp-GNA2091, NadA, PorA and PorB. Moreover, general validation requirements according to ICHQ2(R1) guideline [129] [130] had to be fulfilled, including an adequate selectivity and sensitivity, which corresponded in obtaining baseline resolution of the peaks and adequate peak areas, to be able to monitor the unadsorbed antigens content identified as a product CQA.

In order to reach this target, different preliminary aspects had to be considered. First, scouting of different UHPLC operative modes for obtaining the separation of the five antigens in the vaccine was



performed. Prior knowledge and experimental studies led to choose reverse phase UHPLC (RP-UHPLC) as analytical technique for method development. Different RP-UHPLC operative modes were tuned and screened by the NexeraX2 method scouting system, operating at 60°C and changing type of organic phase (mixtures of ACN/TFA or methanol/TFA), organic ramp (%/min), starting concentration (from 30% to 40%) and ending concentration (from 75% to 100%) of the organic phase. The aim was to approach to the experimental conditions leading to good selectivity and fast analysis. The best results were achieved using an organic mixture of 90/0.1/9.9 ACN:TFA:H<sub>2</sub>O (v/v) as organic phase and a 4 minutes gradient from 34% to 75% of organic phase. These conditions constituted the starting point for further in-depth optimization by DoE. In these conditions, the retention order of the peaks was the following: NHBA-GNA1030, PorB, PorA, fHbp-GNA2091, NadA.

The second aspect was the definition of the surfactant needed for the analysis, which was based on prior knowledge of antigens desorbing from aluminium hydroxide and on preliminary experimental runs. Different surfactants were tested, i.e. Tween 80 and Zwittergent 3-14 detergents in phosphate buffers. The selected surfactant was 0.15% (p/v) Zwittergent 3-14, since it made it possible to maintain consistent over time the chromatographic area of the proteins in the mock solution without aluminium hydroxide.

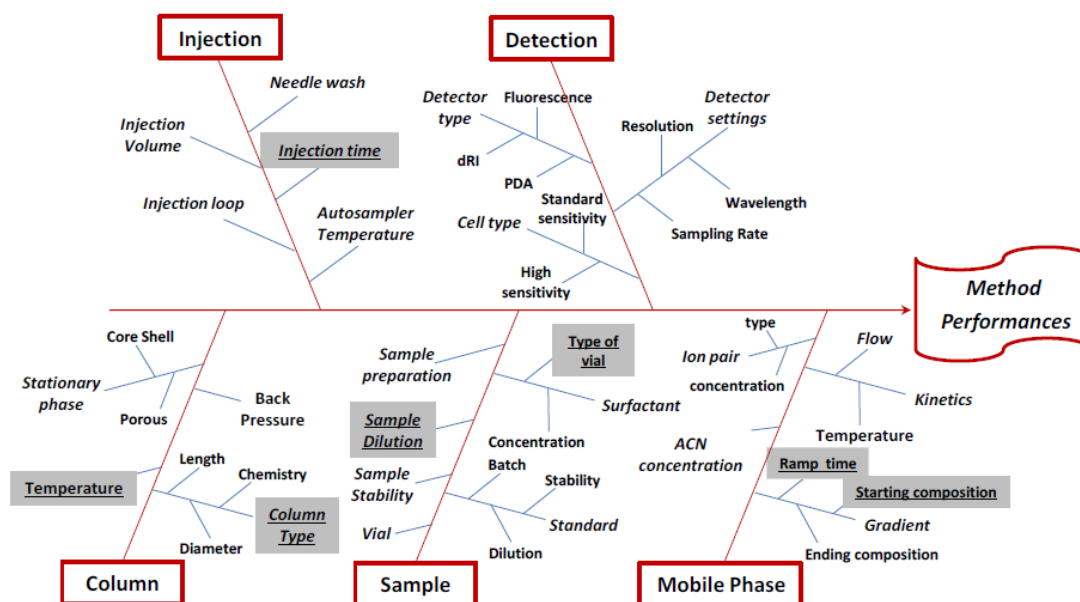
The selected CMAs, reported in Table 6, were the resolution values between the peak pairs, named as  $R_1$  (NHBA-GNA1030/PorB),  $R_2$  (PorB/PorA),  $R_3$  (PorA/fHbp-GNA2091),  $R_4$  (fHbp-GNA2091/NadA), the peak areas  $A_1$  (NHBA-GNA1030),  $A_B$  (PorB),  $A_A$  (PorA),  $A_2$  (fHbp-GNA2091),  $A_3$  (NadA), and NHBA-GNA1030 capacity factor  $K'$ , considered in the RSM for controlling the elution of the first antigen peak with respect to column void volume.

Abbreviation	CMA	CMA requirement
$K'$	NHBA-GNA1030 capacity factor	$1.5 \leq K' \leq 3$
$R_1$	NHBA-GNA1030/PorB resolution	$\geq 1.5$
$R_2$	PorB/PorA resolution	$\geq 1.5$
$R_3$	PorA/fHbp-GNA2091 resolution	$\geq 1.5$
$R_4$	fHbp-GNA2091/NadA resolution	$\geq 1.5$
$A_1$	NHBA-GNA1030 area	$\geq 6 \times 10^5 \mu\text{V s}$
$A_B$	PorB area	$\geq 8 \times 10^4 \mu\text{V s}$
$A_A$	PorA area	$\geq 5 \times 10^4 \mu\text{V s}$
$A_2$	fHbp-GNA2091 area	$\geq 8 \times 10^5 \mu\text{V s}$
$A_3$	NadA area	$\geq 5 \times 10^5 \mu\text{V s}$

**Table 6:** RP-UHPLC critical method attributes and selected requirements

### Risk assessment and potential critical method parameters

The objective of a risk assessment is to develop understanding of procedure variables and their impact on the method reportable values for the identification of hazards and the analysis and evaluation of risks associated with exposure to those hazards [119] [131]. Tools such as process maps and fishbone diagrams may be used, in addition to prior knowledge, to provide structure to a brainstorming and information-gathering exercise to identify pCMPs [119]. In this study, a fishbone diagram (Figure 22) was used to formalize the risk assessment and point out the risk factors associated with the characteristics of the RP-UHPLC analysis and thus to highlight the potential CMPs which were supposed to potentially affect the selected CMAs. Some of the CMPs, including detector type and settings, sample surfactant, ion pair type and autosampler temperature, had been already studied and fixed by preliminary experiments and scouting tests. Other pCMPs, underscored and with grey background in Figure 22, needed to be risk managed and in-depth studied by DoE to enhance knowledge on their effects on method performances.



**Figure 22:** Fishbone diagram for RP-UHPLC method risk assessment.

### Screening experimental designs

As a result of RP-UHPLC scouting and risk assessment, the selected pCMPs to be investigated by DoE were represented by vial type (VIAL), sample concentration (CONC), injection volume (INJ), column type (COL), starting organic phase concentration (ACN%), elution ramp time (RAMP) and column temperature (T).

The first screening study involved the investigation of the effects of VIAL and CONC for optimizing sample preparation, with the aim of reaching adequate sensitivity before starting the optimization of the RP-UHPLC method conditions. The considered CMAs were the peak areas of the five antigens  $A_1$ ,  $A_B$ ,  $A_A$ ,  $A_2$ ,  $A_3$ . The vial type was studied at 3 levels (Clear Glass Total Recovery vial, Polypropylene Total Recovery vial and LCMS Certified Total Recovery vial). Sample concentration was examined at 4 levels (1-4-10-20  $\mu\text{g ml}^{-1}$ ) in order to evaluate the possibility of aggregation and/or aspecific absorption of antigens on vials walls. The injection volume was adapted to inject, at each different sample concentration, the same amount of sample in column.



Nemrod-W software [127] was employed to generate the asymmetric screening matrix used to estimate the coefficients of the following Free-Wilson model with interactions [52]:

$$Y=b_0+b_{1A}(X_{1A})+b_{1B}(X_{1B})+b_{2A}(X_{2A})+b_{2B}(X_{2B})+b_{2C}(X_{2C})+b_{1A2A}(X_{1A}X_{2A})+b_{1A2B}(X_{1A}X_{2B})+b_{1A2C}(X_{1A}X_{2C})+b_{1B2A}(X_{1B}X_{2A})+b_{1B2B}(X_{1B}X_{2B})+b_{1B2C}(X_{1B}X_{2C})$$

where  $X_1$  is VIAL and  $X_2$  is CONC,  $b_0$  is the constant term and  $b_i$  are the linear and interaction coefficients.

The model contains one constant term plus, for each factor, a number of terms equal to its number of levels minus one. The 12-run experimental plan ( $2^13^1//12$ ) is reported in Table 7. Each analysis was duplicated in order to obtain a reliable estimate of the experimental variance.

The graphical plots describing the effects of changing the levels of the factors on antigens areas are shown in Figure 23, where A stands for the type of vial (A1, Clear Glass Total Recovery vial; A2, Polypropylene Total Recovery vial; A3, LCMS Certified Total Recovery vial) and B stands for level of sample concentration (B1,  $1 \mu\text{g ml}^{-1}$ ; B2,  $4 \mu\text{g ml}^{-1}$ ; B3,  $10 \mu\text{g ml}^{-1}$ ; B4,  $20 \mu\text{g ml}^{-1}$ ).

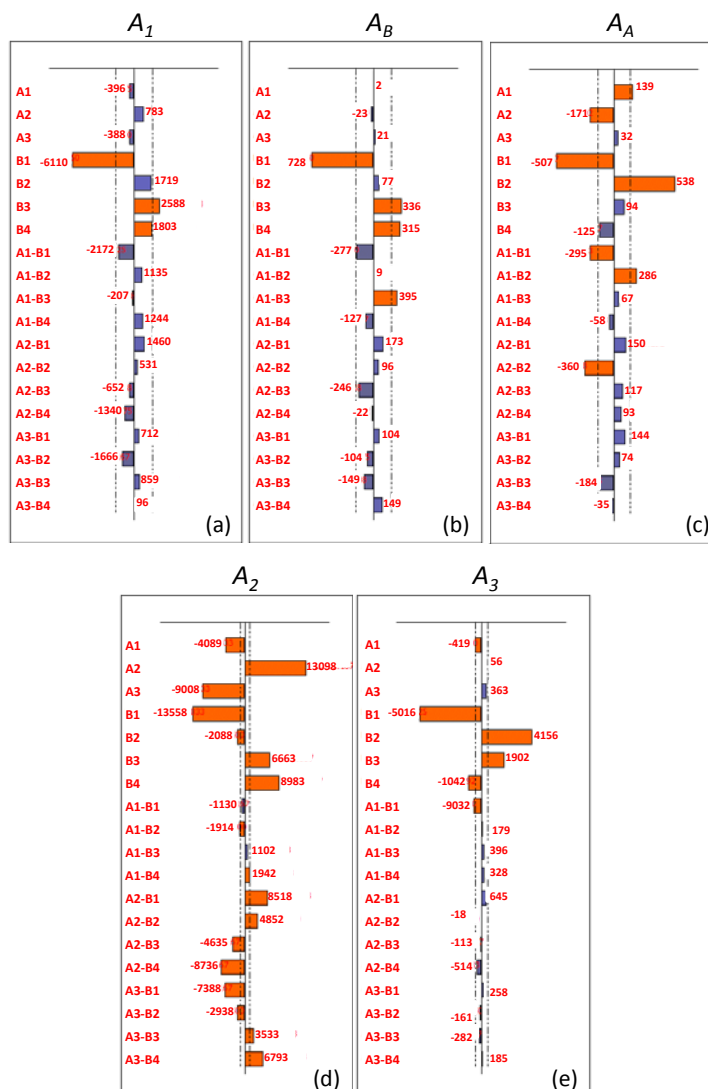




Exp. No	VIAL <sup>[a]</sup>	CONC ( $\mu\text{g ml}^{-1}$ )	$A_I$	$A_B$	$A_A$	$A_2$	$A_3$
1	CGTR	1	111548	15795	6969	106500	96167
2	CGTR	1	103531	15496	6681	112164	94654
3	PTR	1	112098	16092	7096	136009	96749
4	PTR	1	112603	16047	6823	136327	98119
5	LCTR	1	113487	15834	6998	96654	97152
6	LCTR	1	107375	16255	7316	99655	97554
7	CGTR	4	120501	16409	8255	120172	105465
8	CGTR	4	116850	17062	8646	119865	105864
9	PTR	4	119772	16986	7227	144156	106183
10	PTR	4	118731	16606	7761	143789	105702
11	LCTR	4	117253	16704	8181	114013	106466
12	LCTR	4	114513	16579	8083	114136	105747
13	CGTR	10	118815	17499	7837	130611	104159
14	CGTR	10	117589	17263	7739	132962	103094
15	PTR	10	119313	17011	7589	143451	103054
16	PTR	10	118560	16416	7467	143023	104131
17	LCTR	10	119206	16768	7353	129208	103752
18	LCTR	10	119346	16942	7508	129389	103708
19	CGTR	20	119720	16641	7380	133652	100225
20	CGTR	20	118017	17035	7507	136242	101005
21	PTR	20	118058	17152	7236	141458	100299
22	PTR	20	116870	16683	7332	141455	100198
23	LCTR	20	117089	17337	7408	135222	101717
24	LCTR	20	118369	16930	7312	134535	100791

**Table 7:** Asymmetric screening matrix for sample preparation with two replicates for each run.

<sup>[a]</sup>CGTR, Clear Glass Total Recovery vial; PTR, Polypropylene Total Recovery vial; LCTR, LCMS Certified Total Recovery vial.



**Figure 23:** Graphic analysis of effects of VIAL and CONC on chromatographic area responses

(a)  $A_1$ , NHBA-GNA1030; (b)  $A_B$ , PorB; (c)  $A_A$ , PorA; (d)  $A_2$ , fHbp-GNA2091; (e)  $A_3$ , NadA. A1, A2 and A3 indicate the different levels of vial type (A1, Clear Glass Total Recovery vial; A2, Polypropylene Total Recovery vial; A3, LCMS Certified Total Recovery vial), while B1, B2, B3, B4 indicate the different levels of sample concentration (B1, 1  $\mu\text{g ml}^{-1}$ ; B2, 4  $\mu\text{g ml}^{-1}$ ; B3, 10  $\mu\text{g ml}^{-1}$ ; B4, 20  $\mu\text{g ml}^{-1}$ ).

The type of vial had a significant influence only on area responses of PorA and fHbp-GNA2091. The effects were opposite, as the use of Clear Glass Total Recovery vial slightly enhanced PorA area but reduced fHbp-GNA2091 area. The use of Polypropylene Total Recovery vial had a negative effect on PorA area and an important positive effect on fHbp-



GNA2091 area. Glass Total Recovery Silicone Coated vial caused the decrease of fHbp-GNA2091. As for the effects of CONC, in general a decrease of the areas was observed by using the lower concentration value. The maximization of the areas was obtained by using  $4 \mu\text{g ml}^{-1}$  for PorA and NadA,  $10 \mu\text{g ml}^{-1}$  for NHBA-GNA1030 and PorB, and  $20 \mu\text{g ml}^{-1}$  for fHbp-GNA2091. Some interactions were also noticed, evidencing in particular an important positive interaction between  $1 \mu\text{g ml}^{-1}$  concentration value and polypropylene vial on A2 (A2-B1) and other positive interactions involving AB (A1-B3) and A2 (A1-B2). Hence, the vial type selected was Polypropylene Total Recovery vial and sample concentration was fixed at  $1 \mu\text{g ml}^{-1}$ , taking into account the presence of the positive interaction A2-B1 for fHbp-GNA2091, the importance of quantitation of fHbp-GNA2091 antigen (a rMenB protein) with respect to the OMV PorA and PorB antigens, and the possibility to reach a lower range concentration required by analytical target profile in terms of sensitivity gain.

Once the sample conditions were selected, the AQbD framework continued with the screening study of the chromatographic parameters INJ, COL, ACN%, RAMP and T. The following Free-Wilson model was postulated [52]:

$$Y = b_0 + b_{1A}(X_{1A}) + b_{2A}(X_{2A}) + b_{2B}(X_{2B}) + b_{3A}(X_{3A}) + b_{3B}(X_{3B}) + b_{4A}(X_{4A}) + b_{4B}(X_{4B}) + b_{5A}(X_{5A}) + b_{5B}(X_{5B})$$

where  $X_1$  is INJ,  $X_2$  is COL,  $X_3$  is ACN%,  $X_4$  is RAMP and  $X_5$  is T,  $b_0$  is the constant term and  $b_i$  the linear coefficients of each factor. INJ was studied at two levels, while the other four factors were studied at three levels, as reported in Table 8. A new asymmetric screening matrix was designed for obtaining preliminary information throughout the knowledge space on the effects of the five selected factors on chromatographic CMA, i.e. resolutions  $R_1$ ,  $R_2$ ,  $R_3$  and peak areas  $A_B$  and  $A_A$ . In this step  $R_4$  was excluded from the study, as it presented values above 2 in all the runs, while the other resolutions were critical. Only the OMV areas were taken into account because PorB and PorA present the lowest areas among the

five antigens peaks, also due to the lower concentration in the sample with respect to rMenB components.

Critical method parameter	Screening levels	RSM experimental domain	Design space limits
Injection volume (INJ)	20-50 $\mu$ l	30 $\mu$ l	30 $\mu$ l
Column (COL)	C4pore-C4shell-C8pore	C4pore	C4pore
Organic phase starting concentration (ACN%)	24.0-29.0-34.0%	28.0-38.0%	32.0-34.6%
Ramp time (RAMP)	2.0-4.0-6.0 min	3.0-6.0 min	4.0-5.6 min
Temperature (T)	50-60-70 $^{\circ}$ C	50-70 $^{\circ}$ C	60-68 $^{\circ}$ C

**Table 8:** *Experimental domain/optimized values for critical method parameters in the screening phase, in Response Surface Methodology, and in the definition of the design space.*

The screening asymmetric matrix used to estimate the model coefficients was made by 16 runs ( $2^1 3^4 // 16$ ), with three replicates for each experiment in order to obtain a reliable estimate of the experimental variance. Table 9 reports the experimental matrix with the measured responses. The graphic analysis of effects made it possible to obtain two types of plots. The first type of plot, reported in Figure 24, shows the effects of the different levels of the factors on the responses, with the bar length proportional to the effects (a longer bar corresponds to a maximization of the response). The second type of plot, shown in Figure 25, shows the difference of the effects between the two considered levels, and the bars coloured in orange, exceeding the reference line, correspond to the pair of factors for which a change of level is significant on the response.



**Designing Quality:**  
**Quality by Design in the analytical pharmaceutical development**



No. Exp.	INJ ( $\mu$ L)	COL <sup>[a]</sup>	ACN% (% v/v)	RAMP (min)	T (°C)	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	A <sub>B</sub>	A <sub>A</sub>
1	20	C4pore	24.0	2.0	50	0.46	0.76	1.43	74105	15775
2	20	C4pore	24.0	2.0	50	0.51	0.82	1.43	72825	14619
3	20	C4pore	24.0	2.0	50	0.76	1.20	1.60	56603	12740
4	20	C4shell	29.0	4.0	60	0.98	n.m.	n.m.	n.m.	n.m.
5	20	C4shell	29.0	4.0	60	0.90	n.m.	n.m.	n.m.	n.m.
6	20	C4shell	29.0	4.0	60	0.74	n.m.	n.m.	n.m.	n.m.
7	20	C8pore	34.0	6.0	70	2.61	5.04	3.38	38037	10110
8	20	C8pore	34.0	6.0	70	2.46	4.25	3.05	36104	11306
9	20	C8pore	34.0	6.0	70	2.57	4.25	3.05	30549	36104
10	20	C4pore	24.0	2.0	50	0.61	0.99	1.47	61427	17336
11	20	C4pore	24.0	2.0	50	0.60	0.98	1.58	57619	12284
12	20	C4pore	24.0	2.0	50	0.51	0.84	1.36	68829	17482
13	50	C4pore	29.0	6.0	50	2.06	3.29	3.59	137639	47219
14	50	C4pore	29.0	6.0	50	2.13	3.26	3.55	148624	45943
15	50	C4pore	29.0	6.0	50	1.84	3.33	3.69	149765	41934
16	50	C4shell	24.0	2.0	70	0.33	0.80	0.52	252144	72987
17	50	C4shell	24.0	2.0	70	0.32	0.78	0.55	282327	60408
18	50	C4shell	24.0	2.0	70	0.32	0.77	0.51	288066	62828
19	50	C8pore	24.0	2.0	60	0.74	1.01	1.15	147498	30831
20	50	C8pore	24.0	2.0	60	0.64	0.94	1.14	145162	33618
21	50	C8pore	24.0	2.0	60	0.70	0.96	1.13	151450	36951
22	50	C4pore	34.0	4.0	50	2.21	2.80	2.86	131703	49740
23	50	C4pore	34.0	4.0	50	2.21	3.02	3.00	111981	46155
24	50	C4pore	34.0	4.0	50	2.21	2.78	2.77	126437	47672
25	20	C4pore	34.0	2.0	60	1.41	2.42	2.12	53342	13901
26	20	C4pore	34.0	2.0	60	1.41	2.44	2.10	50820	13877
27	20	C4pore	34.0	2.0	60	1.72	2.44	2.11	38824	13348
28	20	C4shell	24.0	6.0	50	n.m.	n.m.	n.m.	n.m.	n.m.



29	20	C4shell	24.0	6.0	50	n.m.	n.m.	n.m.	n.m.	n.m.
30	20	C4shell	24.0	6.0	50	n.m.	n.m.	n.m.	n.m.	n.m.
31	20	C8pore	24.0	4.0	50	2.02	2.30	1.88	20868	11633
32	20	C8pore	24.0	4.0	50	2.03	2.28	2.11	27053	12302
33	20	C8pore	24.0	4.0	50	1.96	2.42	2.15	29004	13175
34	20	C4pore	29.0	2.0	70	0.56	1.35	1.80	66099	11488
35	20	C4pore	29.0	2.0	70	0.60	1.51	1.81	63891	12028
36	20	C4pore	29.0	2.0	70	0.62	1.55	1.85	80214	11253
37	50	C4pore	24.0	4.0	70	0.89	1.51	3.05	198604	37440
38	50	C4pore	24.0	4.0	70	0.98	1.99	3.40	150100	33814
39	50	C4pore	24.0	4.0	70	0.93	1.93	3.56	163760	4483
40	50	C4shell	34.0	2.0	50	1.25	2.00	1.65	119055	23262
41	50	C4shell	34.0	2.0	50	1.33	2.18	1.71	119690	19659
42	50	C4shell	34.0	2.0	50	1.28	2.06	1.65	126697	24961
43	50	C8pore	29.0	2.0	50	1.17	1.37	1.41	114977	25790
44	50	C8pore	29.0	2.0	50	1.16	1.36	1.42	114421	24496
45	50	C8pore	29.0	2.0	50	1.25	1.37	1.40	87813	18504
46	50	C4pore	24.0	6.0	60	1.89	3.54	3.59	114771	35997
47	50	C4pore	24.0	6.0	60	2.45	3.35	3.39	121570	41358
48	50	C4pore	24.0	6.0	60	2.44	3.25	3.32	101252	42953

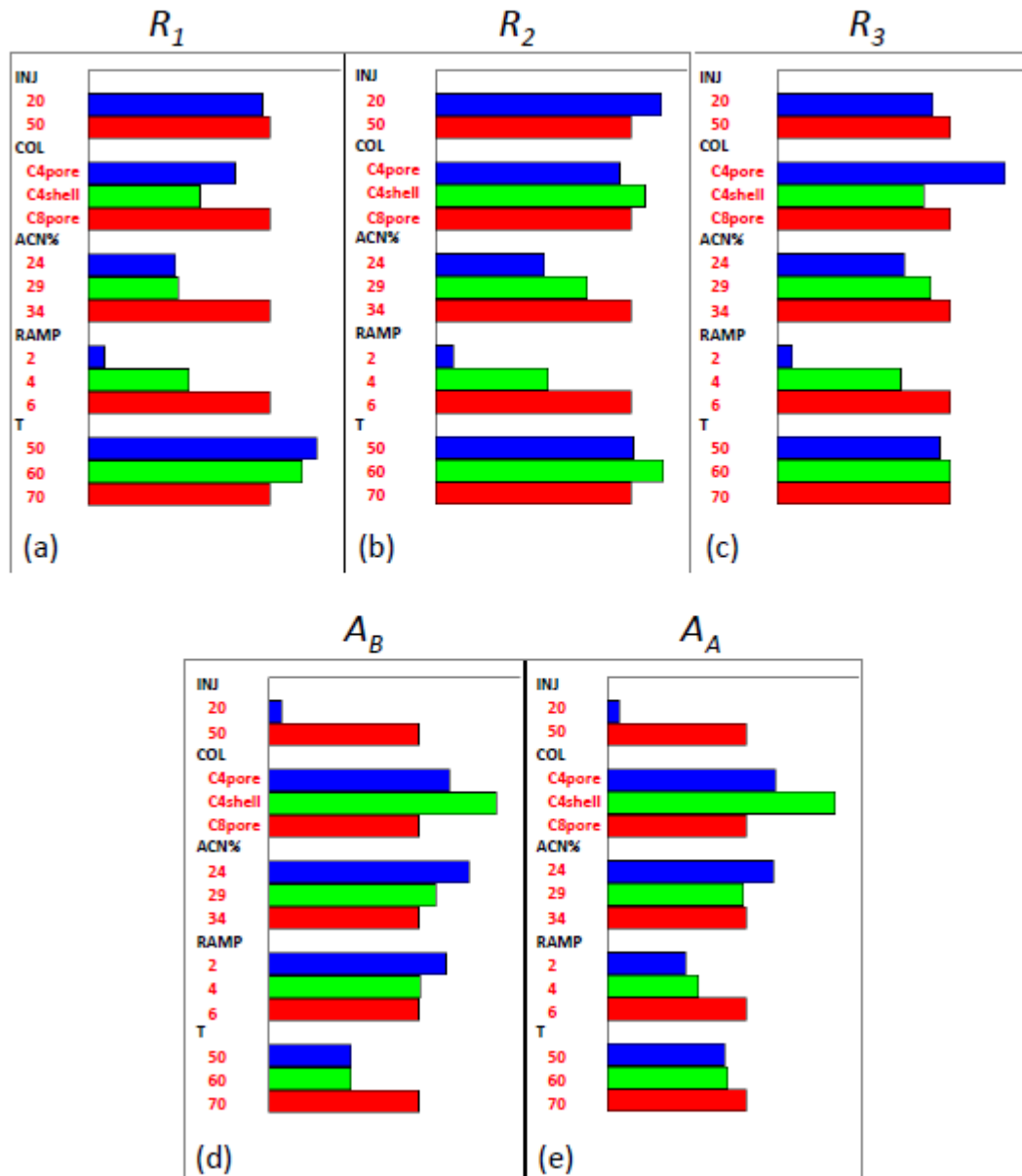
**Table 9:** Screening of knowledge space: 16-run asymmetric screening matrix with three replicates for each run.

n.m.: not measured; due to particular peak patterns, the result of the experiment was excluded from data treatment.

<sup>[a]</sup> C4pore, Acquity RP-C4 BEH 300Å, 1.7µm, 2.1x150 mm; C4shell, Aeris™ WIDEPOR C4 200Å, 3.6 µm, 4.6x150 mm; C8pore, Acquity UPLC BEH 130Å C8, 1.7µm, 2.1x150 mm.



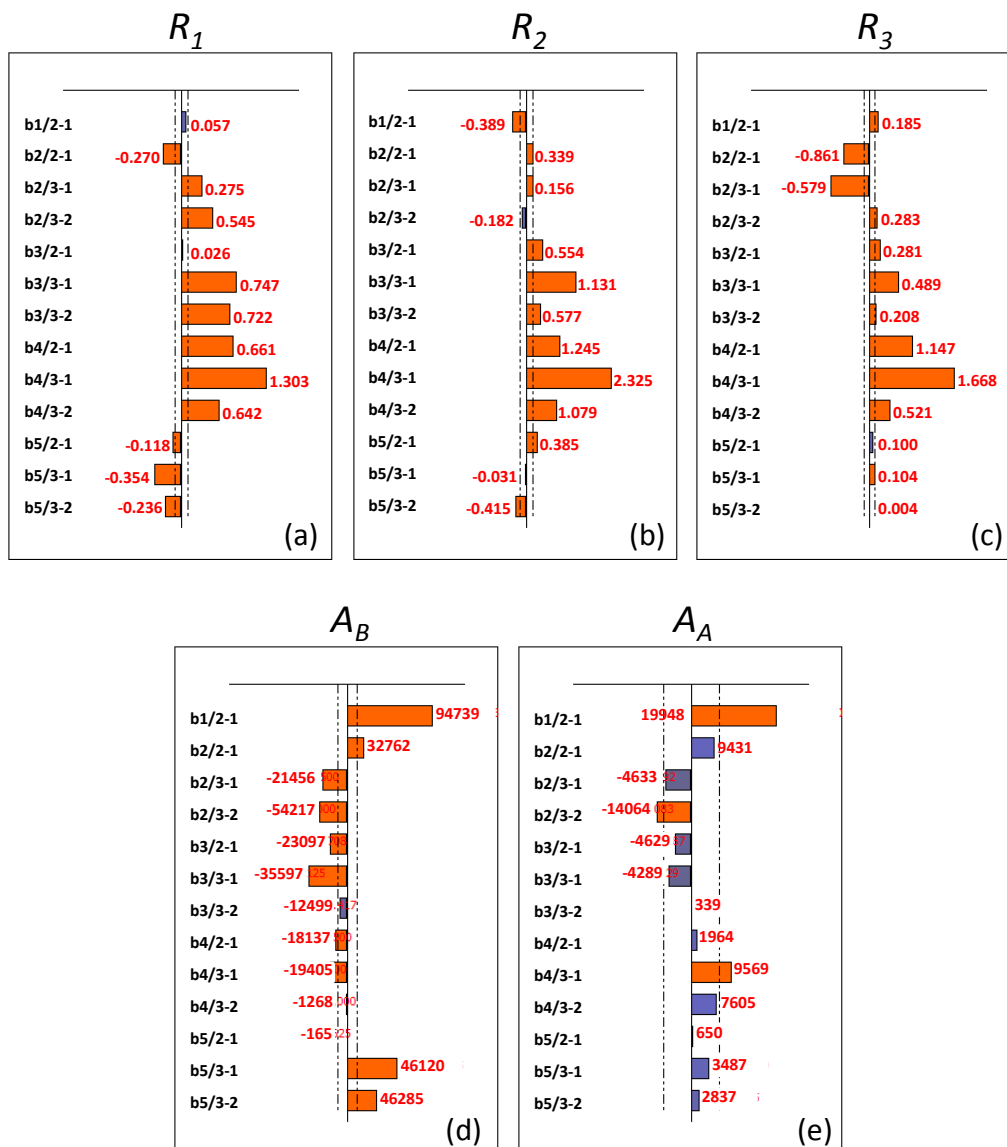
By examining the following plots, it was possible to select the optimal value for some CMPs and to decide which others should be further studied more deeply by RSM. As for injection volume, its change exerted a limited or no effect on resolution values, but the 50  $\mu$ l value was definitely better for increasing both the areas, as expected. Hence, this factor was fixed at an intermediate value of 30  $\mu$ l, in order to find a compromise between selectivity and sensitivity. As regards the column, C4pore was selected for further studies, since it gave the maximization of  $R_3$  value and led to good results also for the other responses, taking into account that in general this column is preferred for protein studies and for avoiding absorption. The highest level for ACN% led to maximize all the resolution values, while low-medium values were preferred for increasing area responses. Hence, the new domain to be investigated was moved towards the values 28.0-38.0%. When considering RAMP, the value of 6 min led to the maximization of all the responses apart from  $A_B$ , and it was decided to further study this CMP in the range 3-6 min, in the perspective of keeping low analysis time. As concerns T, in general it presented a lower influence on the CMAs with respect to the other CMPs, but it showed conflicting effects on the different responses, thus it was decided to continue to study this factor by RSM in the domain 50-70°C.



**Figure 24:** *Graphic analysis of effects for investigation of chromatographic resolutions and area responses.*

Resolutions: (a)  $R_1$ , NHBA-GNA1030/PorB; (b)  $R_2$ , PorB/PorA; (c)  $R_3$ , PorA/fHbp-GNA2091. Areas: (d)  $A_B$ , PorB; (e)  $A_A$ , PorA. The length of each bar indicates the effect of the each level of each factor under study.





**Figure 25:** Graphic analysis of effects for resolution and area responses

The length of each bar indicates the effects of changing level of the factors. Resolutions: (a)  $R_1$ , NHBA-GNA1030/PorB; (b)  $R_2$ , PorB/PorA; (c)  $R_3$ , PorA/fHbp-GNA2091. Areas: (d)  $A_B$ , PorB; (e)  $A_A$ , PorA.



### *Response surface methodology and design space*

RSM was applied for in-depth investigations of the effects of ACN%, RAMP and T on the CMAs in the new experimental domain reported in Table 6. A three-factor circumscribed CCD, fractionating the experimental domain for each factor into 5 levels, was employed for building a quadratic model with interactions making it possible to study all the selected chromatographic CMAs, namely resolutions ( $R_1, R_2, R_3, R_4$ ) and peak areas ( $A_1, A_A, A_B, A_2, A_3$ ) as above described. Additionally, NHBA-GNA1030 capacity factor  $K'$  was also considered among the CMAs in order to control the elution of the first antigen peak with respect to column void volume. As a matter of facts, from screening DoE it was observed that high organic phase starting concentration values anticipate the chromatographic pattern of the antigens.

The 15-run CCD experimental plan with the measured responses is reported in Table 10, where each condition was twice replicated, including a central point. The responses  $R_1$  and  $A_B$  were respectively reverse ( $Y^{-1}$ ) and logarithmic ( $\text{Log}_{10}Y$ ) transformed to stabilize the variance and makes the data more normal distributed-like. The others CMA models were obtained without mathematical transformation. All the ten models, calculated by multiple-linear regression, were significant in terms of ANOVA, and in general the goodness of fitting (expressed by determination coefficient  $R^2$ ) and of prediction (expressed as cross validated  $Q^2$ ) were good, as reported in bottom Table 11. Only for  $A_A, A_2$  and  $A_3$  low  $Q^2$  coefficients were obtained, but in these cases the validity of the models was satisfactory (lack of fit  $p > 0.050$ ).

Hence, contour plots were drawn reporting the calculated isoresponse curves, in order to obtain detailed information on the behaviour of each CMAs throughout the experimental domain investigated. The contour plots are shown in Figure 26 for  $K'$  and resolution responses and in Figure 27 for area responses. The investigation of these plots, combined with the analysis of the coefficients reported in Figure 28, allowed understanding the method performance behaviour in function of the selected CMPs.



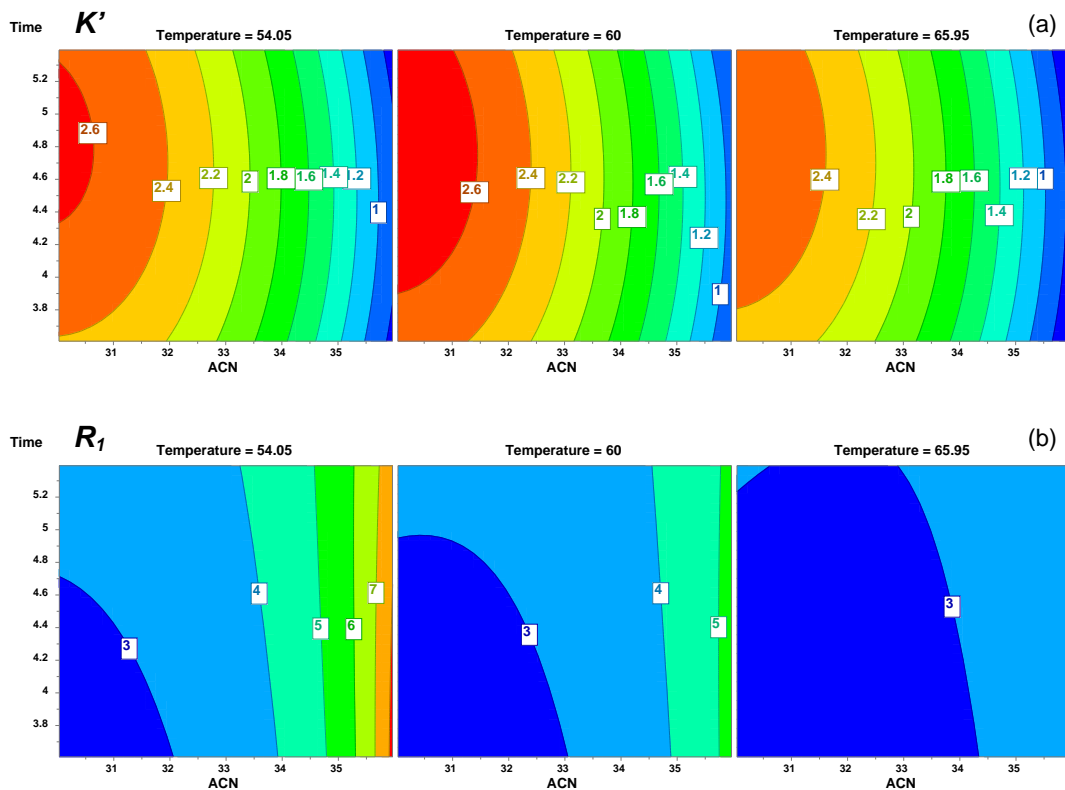
Exp No.	ACN% (v/v)	RAMP (min)	T (°C)	K'	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	A <sub>I</sub>	A <sub>B</sub>	A <sub>A</sub>	A <sub>2</sub>	A <sub>3</sub>
1	30.0	3.6	54	2.33	2.74	4.38	4.40	12.34	827072	120153	61532	954229	695159
2	36.0	3.6	54	0.40	8.87	3.07	4.97	13.69	192518	77833	61966	946606	679564
3	30.0	5.4	54	2.54	3.47	5.52	5.58	15.60	744264	112250	57937	874339	639877
4	36.0	5.4	54	0.38	7.62	3.78	6.26	17.04	172650	65528	64237	918216	664891
5	30.0	3.6	66	2.28	2.44	4.69	4.31	11.84	786301	99240	60278	920390	689505
6	36.0	3.6	66	0.28	4.11	18.45	4.89	12.82	512063	91635	61987	927961	690779
7	30.0	5.4	66	2.47	2.96	5.81	5.43	14.44	815790	107843	62408	883211	688962
8	36.0	5.4	66	0.27	3.65	22.61	6.02	15.79	533282	79175	60218	929630	693485
9	28.0	4.5	60	2.55	2.91	5.08	4.81	13.39	834298	103214	61695	957885	688684
10	38.0	4.5	60	0.28	n.m.	n.m.	n.m.	15.17	n.m.	63088	n.m.	n.m.	691094
11	33.0	3.0	60	2.04	2.64	4.37	4.17	11.47	831260	117138	60337	961178	692585
12	33.0	6.0	60	2.22	3.77	6.53	6.25	16.78	732744	97817	55479	801380	621264
13	33.0	4.5	50	2.18	3.50	5.33	5.30	14.88	815698	119205	60706	900552	648673
14	33.0	4.5	70	2.06	2.80	5.68	4.95	13.19	570167	81423	45961	692391	503270
15	33.0	4.5	60	2.13	3.23	5.56	5.21	14.28	856186	109521	62729	964840	685766
16	30.0	3.6	54	2.34	2.77	4.34	4.43	12.52	691623	100595	53493	812879	588620
17	36.0	3.6	54	0.43	12.48	3.98	4.90	13.47	295265	88196	64728	959019	688298
18	30.0	5.4	54	2.55	3.49	5.7	5.66	15.83	763914	115626	56247	897206	657636
19	36.0	5.4	54	0.33	8.27	4.73	6.26	16.58	197803	60159	64686	924911	662633
20	30.0	3.6	66	2.28	2.40	4.76	4.26	11.75	813741	98845	61369	947682	688511
21	36.0	3.6	66	0.27	3.59	23.19	4.98	12.89	517853	80352	61431	898508	691444
22	30.0	5.4	66.	2.47	2.91	5.83	5.27	14.35	816902	108987	62429	895531	688421
23	36.0	5.4	66	0.27	3.75	21.94	5.94	15.37	533917	83108	61752	930497	693814
24	28.0	4.5	60	2.56	2.82	4.94	4.76	13.1	773774	100973	58040	898587	636369
25	38.0	4.5	60	0.24	n.m.	n.m.	n.m.	15.64	n.m.	48332	n.m.	n.m.	680147
26	33.0	3.0	60	2.04	2.68	4.51	4.26	11.78	789085	111594	60534	917929	671487
27	33.0	6.0	60	2.21	3.66	6.50	6.04	16.57	849670	111872	62340	912363	702199
28	33.0	4.5	50	2.19	3.47	5.15	5.30	14.95	723611	114196	55352	812261	590733
29	33.0	4.5	70	2.08	2.73	5.48	4.79	13.11	647595	90466	52648	779225	562611
30	33.0	4.5	60	2.14	3.20	5.47	5.16	14.08	754701	98261	55638	854503	607587

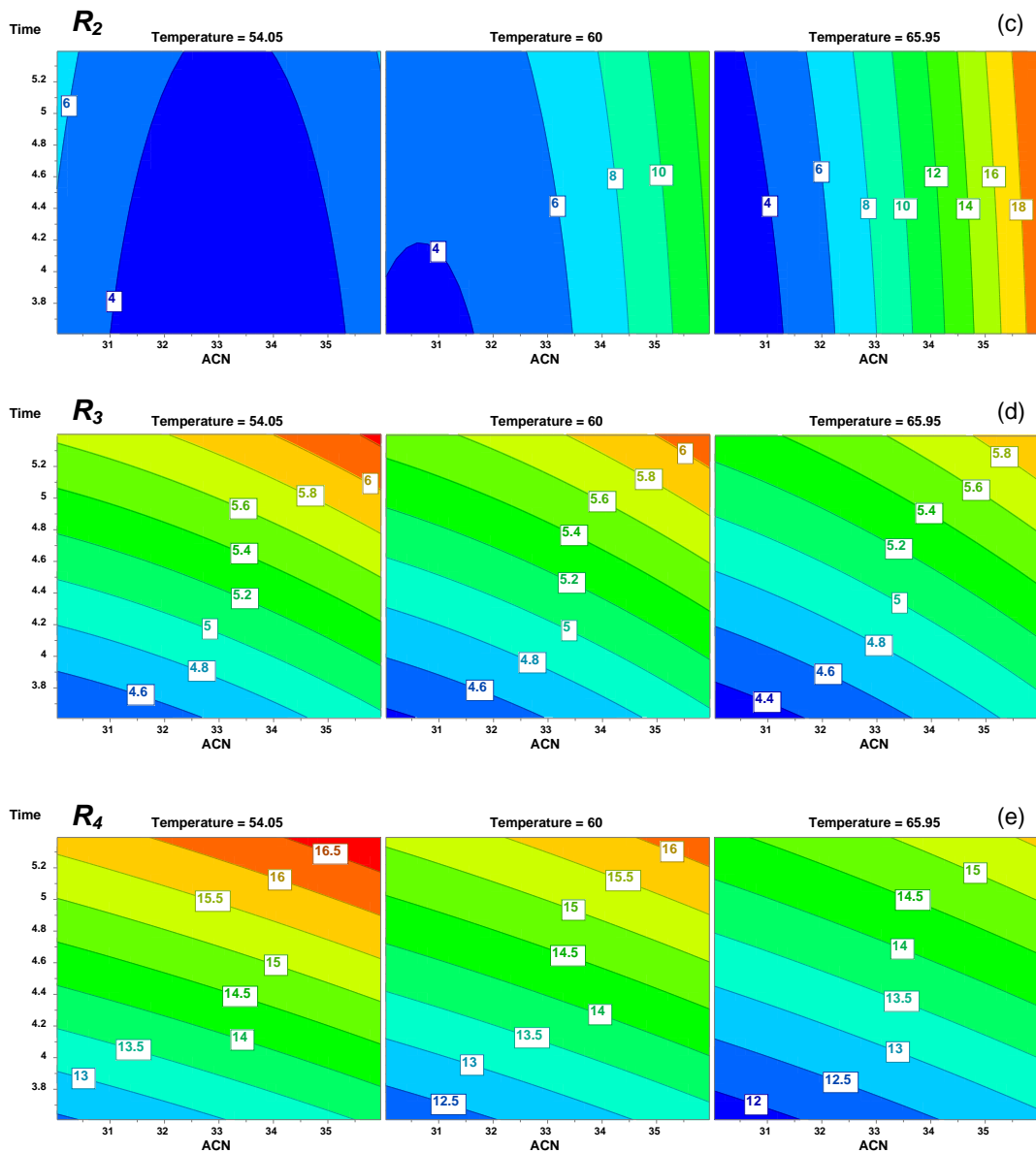
**Table 10:** Response surface methodology: 15-run Central Composite Design experimental plan with two replicates for each run.

n.m.: not measured; due to particular peak patterns, the result of the experiment were excluded from data treatment.

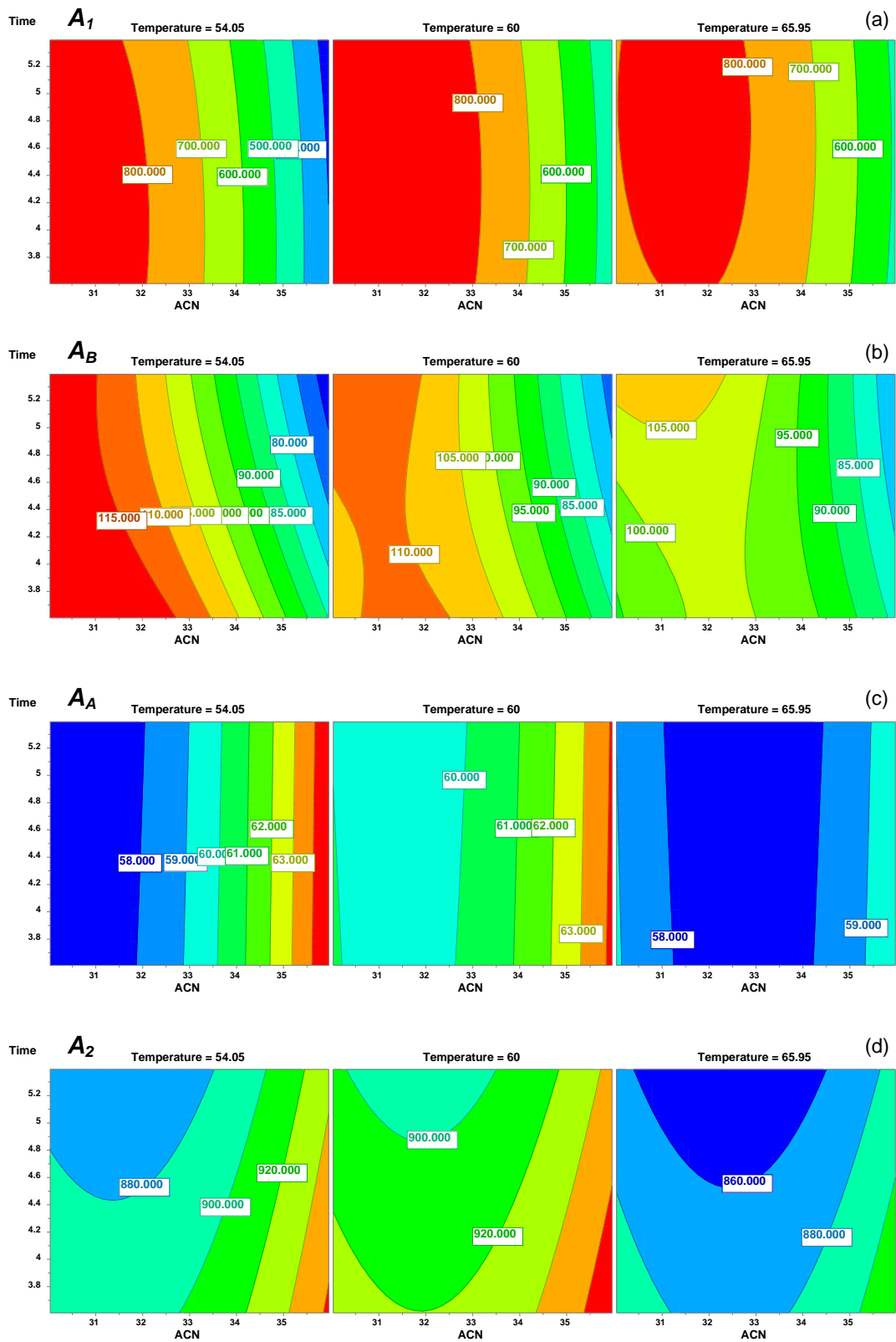
CMA	$R^2$	$Q^2$	Regression	Lack of fit
			$p$ value	$p$ value
$K'$	0.907	0.800	0.000	0.000
$R_1$	0.950	0.885	0.000	0.000
$R_2$	0.844	0.722	0.000	0.000
$R_3$	0.992	0.982	0.000	0.366
$R_4$	0.993	0.983	0.000	0.852
$A_1$	0.850	0.652	0.000	0.000
$A_B$	0.876	0.725	0.000	0.056
$A_A$	0.514	0.105	0.011	0.233
$A_2$	0.523	0.135	0.010	0.388
$A_3$	0.425	-0.121	0.033	0.092

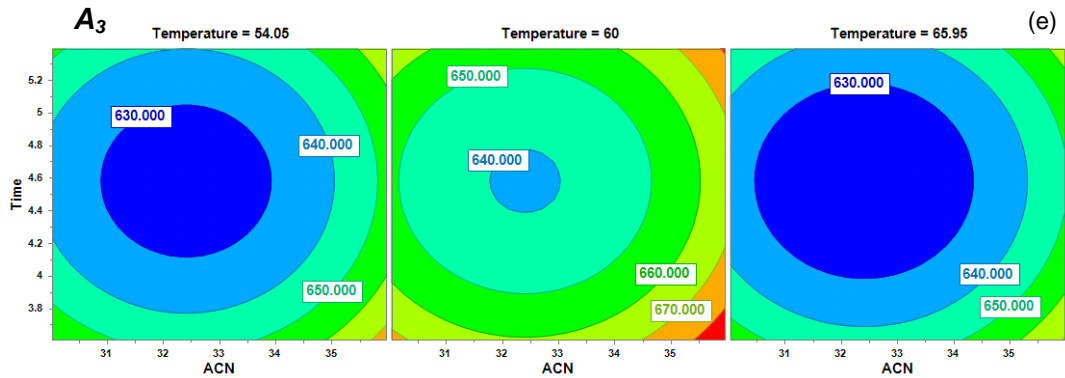
**Table 11:** Response surface methodology: quality of the calculated models for all the CMAs.



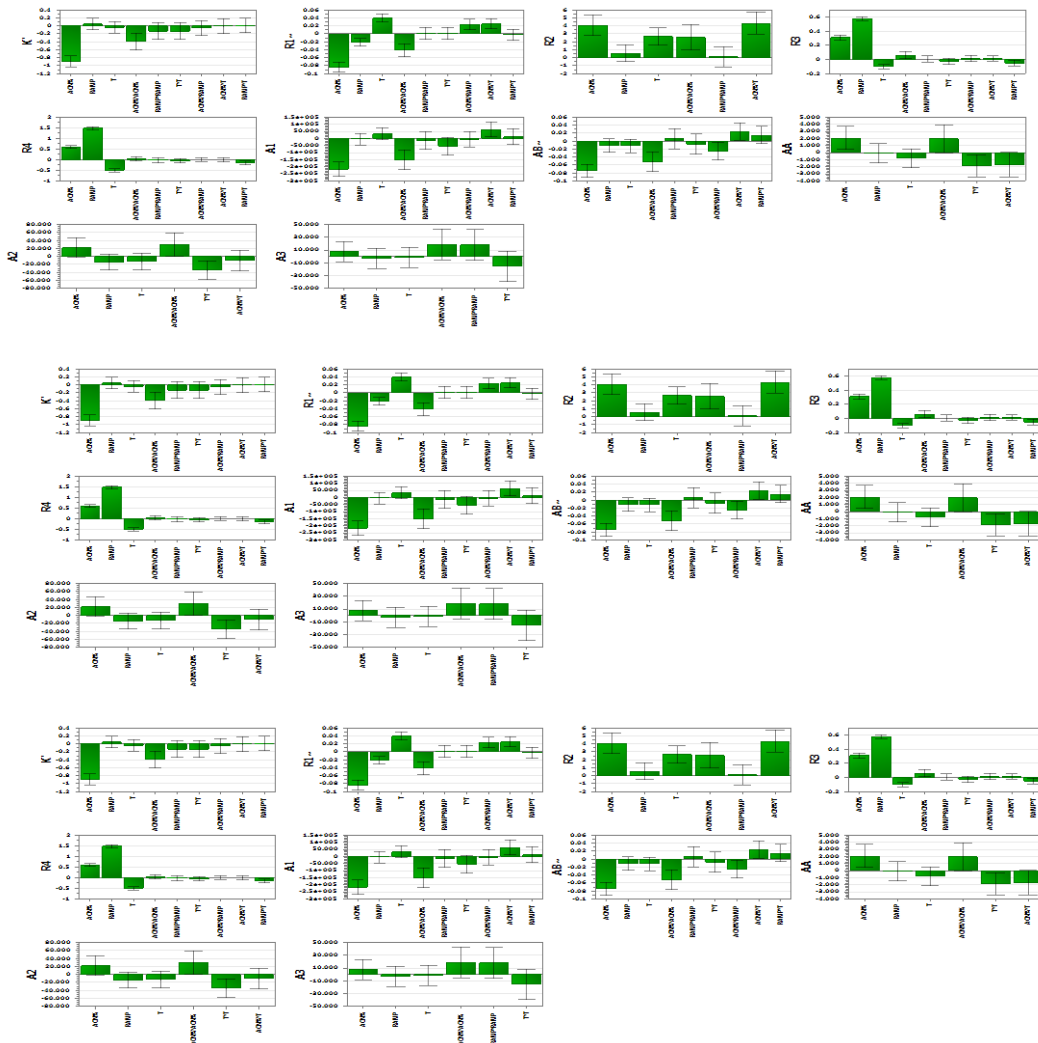


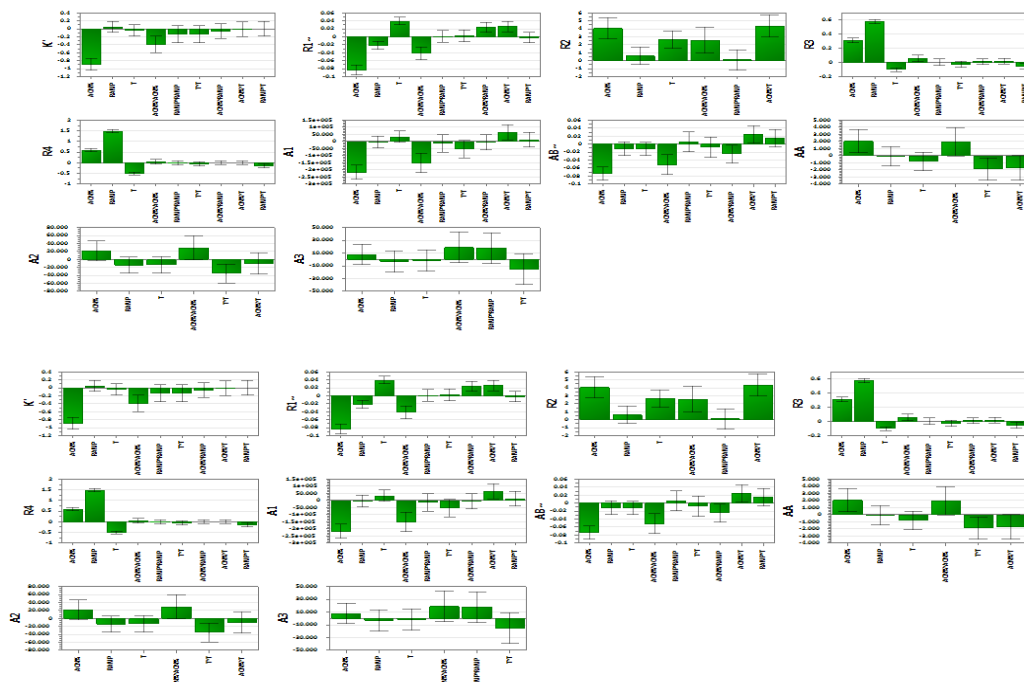
**Figure 26:** Isoresponse surfaces drawn by plotting RAMP vs. ACN% for: (a) NHBA-GNA1030 capacity factor  $K'$ , (b)  $R_1$ , (c)  $R_2$ , (d)  $R_3$ , (e)  $R_4$  at three different values of temperature: 50°C, 60°C, 70°C.





**Figure 27:** Isoresponse surfaces drawn by plotting RAMP vs. ACN% for: (a) A1, (b) AB, (c) AA, (d) A2, (e) A3 at three different values of temperature: 50°C, 60°C, 70°C.





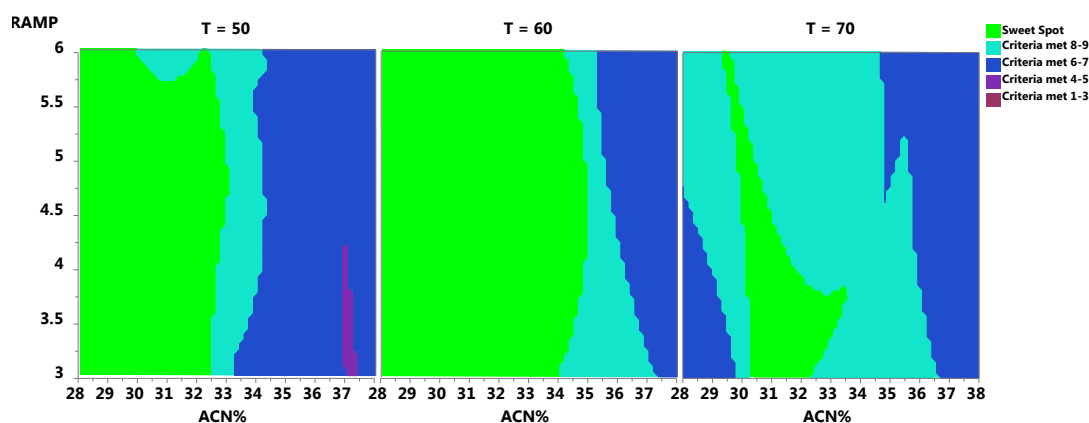
**Figure 28:** Response Surface Methodology: graphic analysis of effects of the CMPs on the CMAs

The precision of the regression coefficient was calculated as the 95% confidence interval and was superimposed as an error on each coefficient bar.

As for  $K$  (Figure 26 a), the only significant effect was exerted by ACN%, which showed both a negative linear and a negative quadratic effect; high values of this factor anticipate the NHBA-GNA1030 peak toward the column void volume and the zone corresponding to the maximization of this factor was located at medium levels of all the CMPs. As concerns the four resolution CMAs, they were all maximized by high levels of ACN% (Figure 26 b-e). RAMP presented a strong positive effect on resolution  $R_3$  and  $R_4$ . T had a negative effect on  $R_1$  and  $R_4$  and a positive effect on  $R_2$ . Quadratic effects of ACN% were evidenced on all the resolution CMAs apart from  $R_4$ . Some significant interactions were also noticed: ACN%-RAMP and ACN%-T on  $R_1$  and ACN%-T on  $R_2$ . As for the CMAs related to areas (Figure 27), different trends and curvatures in the contour plots were noticed. The graphical analysis of effects in Figure 28 evidenced negative effects for ACN% on  $A_1$  and  $A_B$  and a positive effect on  $A_A$ , while T showed a positive interaction with ACN% on  $A_1$  and  $A_B$ . RAMP showed no influence on area responses.



The sweet spot plots, shown in Figure 29, highlight by different colours the areas where one or more predicted CMAs fulfil the related requirements. For all the CMAs related to resolution, a desired minimum value of 1.5 was set, while for the other CMAs the limits were set as reported in Table 6. The different colours had the following meaning: purple, from 1 to 3 CMAs criteria met; different blue gradations, from 4 to 5 and from 6 to 7 criteria met; light blue, from 8 to 9 criteria met; and finally, green where all the ten CMAs criteria were fulfilled, namely the zone corresponding to the sweet spot. Anyway, the green plot area does not constitute the design space, but it only shows where the overlay of the CMAs response surfaces meets the desired performances, providing additional information by highlighting the zone where the maximum number of criteria is fulfilled.

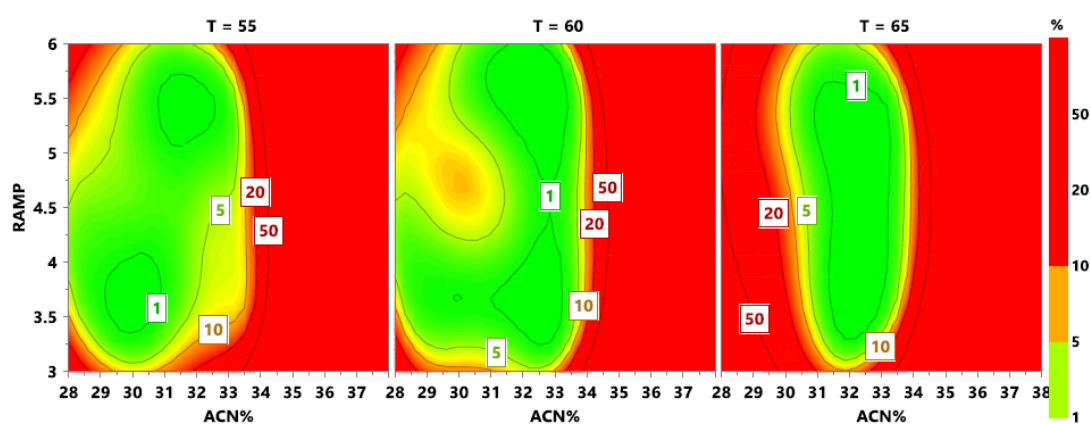


**Figure 29:** Sweet spot plot obtained plotting RAMP vs. ACN% at three different values of  $T$ .

The zones colored in green correspond to the zones where the requirements for all the ten CMAs are fulfilled.

On the other hand, the MODR takes into account the concept of probability that the requirements are met [2] [7], and it was computed by MODDE software [128]. Considering all the response surface models and the settings of CMAs requirements, the risk of failure map (Figure 30) was drawn using Monte-Carlo simulations for risk analysis, taking into account the model parameters uncertainty, propagating the prediction error from parameters to responses and giving access to the responses distributions

for each RSM condition [132]. The selected level of probability ( $\pi$ ) and the original set point for MODR identification were:  $\pi \geq 90\%$ ; ACN%, 33.3%; RAMP, 4.8 min; T, 64°C. Hence, the design space was graphically represented in the risk of failure map by the area coloured in green, within the iso-error curves corresponding to the probability failure  $\leq 10\%$ . Within the MODR, the following limits for CMPs ranges were identified: ACN%, 32.0-34.6 %; RAMP, 4.0-5.6 min; T, 60-68°C.

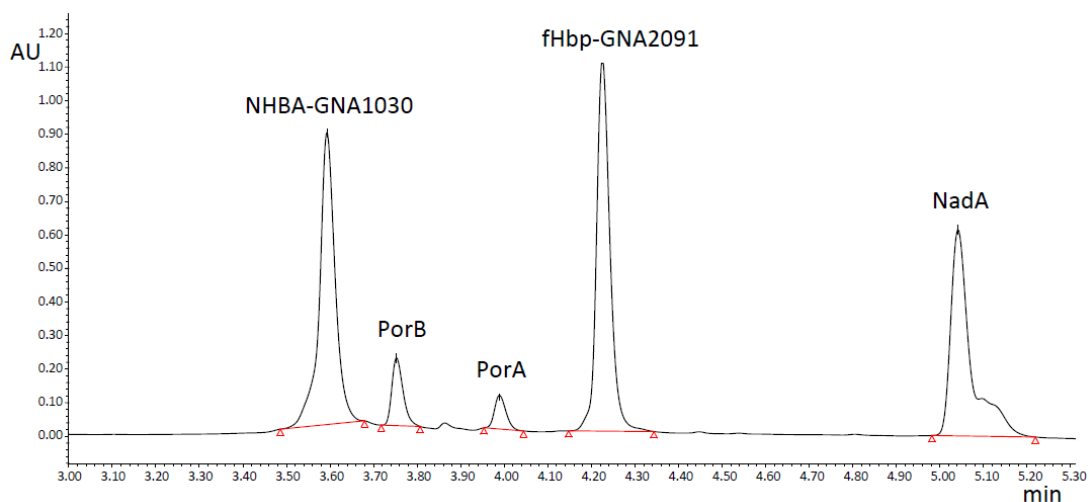


**Figure 30:** RP-UHPLC design space map by plotting RAMP vs. ACN% three different values of temperature: 55°C, 60°C, 65°C.

In order to validate the MODR, the lower and higher limits of the design space ranges were selected as the -1 and +1 levels of a 23 Full Factorial Design [52], with 3 replicates of the original set point to estimate the experimental variance, and the agreement between RSM predicted values and experimental responses at the edges of failure was verified. Once the model validity was confirmed, a working point was chosen inside the lower risk region ( $\pi \geq 99\%$ ), taking into account some practicability factors such as the advantages of implementing an high-throughput method and of choosing a temperature value as low as possible to avoid proteins damages or modifications.

The selected working point for routine use was: ACN%, 33.0%; RAMP, 4.0 min; T, 60°C. Under the selected conditions, a completely separation of the five antigens in about 5 minutes, with the desired selectivity and sensitivity, was obtained. The RP-UHPLC chromatogram of the mock

solution in the selected operative conditions is reported on bottom Figure 31.



**Figure 31:** *RP-UHPLC chromatogram in the working conditions for a mock solution.*

10  $\mu\text{gml}^{-1}$  NHBA-GNA1030, fHbp-GNA2091, NadA and 5  $\mu\text{gml}^{-1}$  Por B, PorA; Acquity H-ClassBio UPLC system; Acquity RP-C4 BEH 300 $\text{\AA}$ , 1.7 $\mu\text{m}$ , 2.1x150 mm; organic phase, ACN plus 0.1% TFA ion pair; injection volume, 30  $\mu\text{l}$ ; temperature, 60 $^{\circ}\text{C}$ ; ramp time, 4 min (from 33.0% to 75.0% ACN); photodiode array detector, 210 nm, 1.2 nm resolution, 20 points  $\text{s}^{-1}$ .

### **Robustness and control strategy**

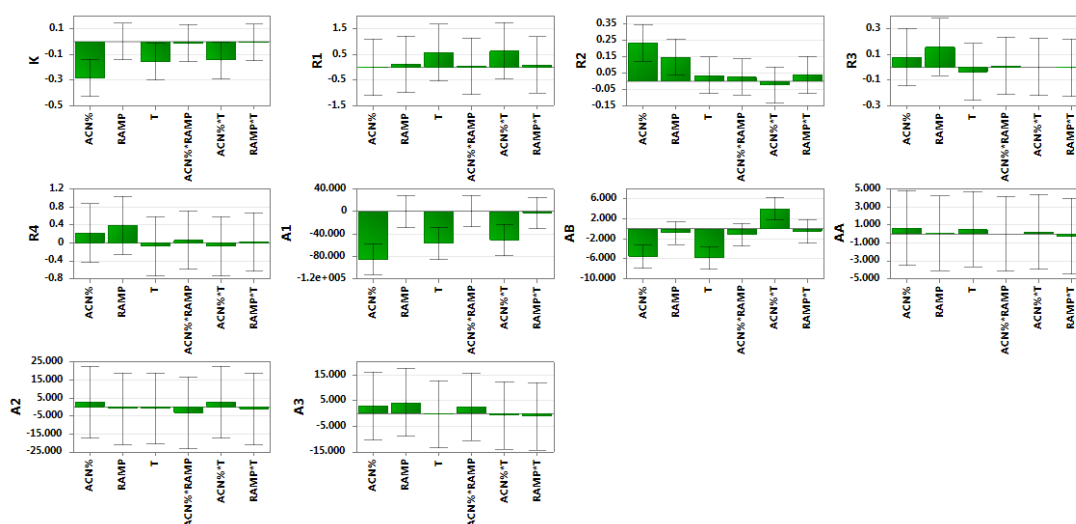
For robustness study, a 8-run Full Factorial Design was employed to calculate the main effects on the CMAs of small CMPs changes around the chosen working conditions [130] and the experimental plan with the responses is shown in Table 12. The resulting graphical analysis of effects is shown Figure 32. The analysis of effects evidenced that the method could be considered robust, including a precautionary statement for ACN%. As a matter of facts, ACN% exerted a significant effect on  $K'$ ,  $R_2$ ,  $A_1$  and  $A_B$  also in this small interval. As for the other factors, RAMP was significant only on  $R_2$  and T on  $A_1$  and  $A_B$ . Anyhow, all the results were within the CMAs desired limits for all the experiments.

Finally, a preliminary control strategy [2] for the RP-UHPLC method was accomplished on the basis of all the development data, consisting in a continuous performance verification plan to control recombinant proteins

resolutions, efficiency and peak symmetry, setting the following system suitability limits. A standard mixture of the rMenB proteins ( $100 \mu\text{g ml}^{-1}$ ) should fulfil these requirements: resolution (NHBA-GNA1030/fHBp-GNA2091)  $\geq 9$ , resolution  $R_4$  (fHBp-GNA2091/NadA)  $\geq 10$ , fHBp-GNA2091 number of theoretical plate  $\geq 50000$  and  $0.9 \leq$  fHBp-GNA2091 peak symmetry  $\leq 1.6$ .

Exp. No.	ACN% (v/v)	RAMP (min)	T (°C)	$K'$	$R_1$	$R_2$	$R_3$	$R_4$	$A_1$	$A_B$	$A_A$	$A_2$	$A_3$
1	33.0	3.8	58	2.13	3.09	5.16	4.95	13.7	813447	132413	65189	979825	731861
2	35.0	3.8	58	1.86	1.89	5.64	5.05	14.07	747892	115959	66659	989529	732180
3	33.0	4.2	58	2.15	3.23	5.34	5.22	14.21	823484	134370	66505	989657	735569
4	35.0	4.2	58	1.87	1.85	5.89	5.43	15.02	749966	112523	66679	980334	750354
5	33.0	3.8	62	2.10	2.96	5.22	4.85	13.56	811993	113951	66916	978123	733266
6	35.0	3.8	62	1.29	3.95	5.57	5.02	13.83	534618	113006	68008	992237	734331
7	33.0	4.2	62	2.13	3.13	5.52	5.17	14.35	800452	113288	65846	977280	736316
8	35.0	4.2	62	1.22	4.63	6.01	5.31	14.65	532770	108393	68268	984843	743607

**Table 12: Robustness study**



**Figure 32: Robustness study: graphic analysis of effects of CMPs on CMAs.**

The precision of the regression coefficient was calculated as the 95% confidence interval and was superimposed as an error on each coefficient bar.

### Method qualification

Qualification of the RP-UHPLC developed method was carried out to ensure the suitability of the method for the quality control of the vaccine product. The method was qualified following ICH guidelines [130] for validation, using an experimental plan for testing the effect of noise factor on results variability: operator, column batch, sample preparation, instrument, day of analysis and analytical session. The qualification results showed the suitability of the method for the intended purpose, identity of antigens respect to standard solutions, selectivity, linearity, accuracy (recovery for antigens spike) and precision (data variation of the titred amount). Linearity was evaluated by preparing and analysing ten samples, two for each of five concentration values, ranging from 1 to 20  $\mu\text{g ml}^{-1}$  for rMenB antigens and from 0.5 to 10  $\mu\text{g ml}^{-1}$  for OMV antigens; the related data are shown in Table 13. The recovery values for antigens spike were measured at three concentration values (1.5, 10 and 20  $\mu\text{g ml}^{-1}$ ) performing 6 replicates and obtaining values included in the following intervals: NHBA-GNA1030, 92.3-113.9%; PorB, 93.1-115.9%; PorA, 96.1-108.6%; fHBp-GNA2091, 99.3-106.3%; NadA, 90.9-104.7%. Precision was assessed for fHBp-GNA2091 (10  $\mu\text{g ml}^{-1}$ ): as for repeatability, six analyses were run obtaining a RSD = 1.7% for the titrated amount; as for intermediate precision, calculated by performing 6 replicates at three concentration values, the following RSD values for the titrated amount were found: 1.5  $\mu\text{g ml}^{-1}$ , RSD = 8.4%; 10  $\mu\text{g ml}^{-1}$ , RSD = 7.9%; 20  $\mu\text{g ml}^{-1}$ , RSD = 2.8%.

Antigen	Range ( $\mu\text{g ml}^{-1}$ )	<i>a</i>	<i>b</i>	<i>R</i> <sup>2</sup>
NHBA-GNA1030	1-20	$8.91 \times 10^4$	$1.59 \times 10^4$	0.9996
PorB	0.5-10	$2.32 \times 10^4$	$1.74 \times 10^3$	0.9996
PorA	0.5-10	$1.29 \times 10^4$	$6.21 \times 10^2$	0.9994
fHBp-GNA2091	1-20	$1.02 \times 10^5$	$1.82 \times 10^3$	0.9998
NadA	1-20	$7.56 \times 10^3$	$7.33 \times 10^4$	0.9999

**Table 13:** Linearity data



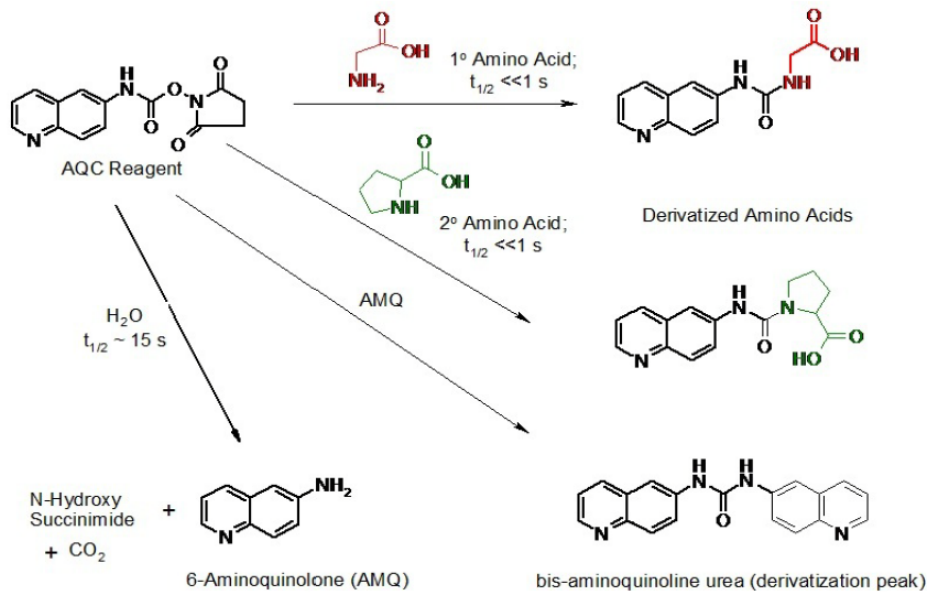
The regression equation of above table is  $y = ax + b$ , where  $x$  is the antigen concentration ( $\mu\text{g ml}^{-1}$ ),  $y$  is the area of the antigen peak and  $R^2$  the coefficient of determination or the linearity regressions.

## AAA Hydrolysis Design Space

A fundamental critical quality attribute of Bexsero vaccine is the antigens content of the final drug product. The drug substances quantification is necessary to qualify the reference rMenB calibration standards for antigen content determination in Bexsero vaccine. The actual method for protein content determination of recombinant MenB proteins (rMenB) in the drug substances is the microBCA test. The bicinchoninic acid assay is a standard method for proteins titration, relatively old and with certain data variability. Following the Analytical Target Profile requirements, the AQbD framework was applied increasing the assays throughput, by reducing the manual operations, improving sensitivity and data precision with respect to the current microBCA assay used to control the above CQA.

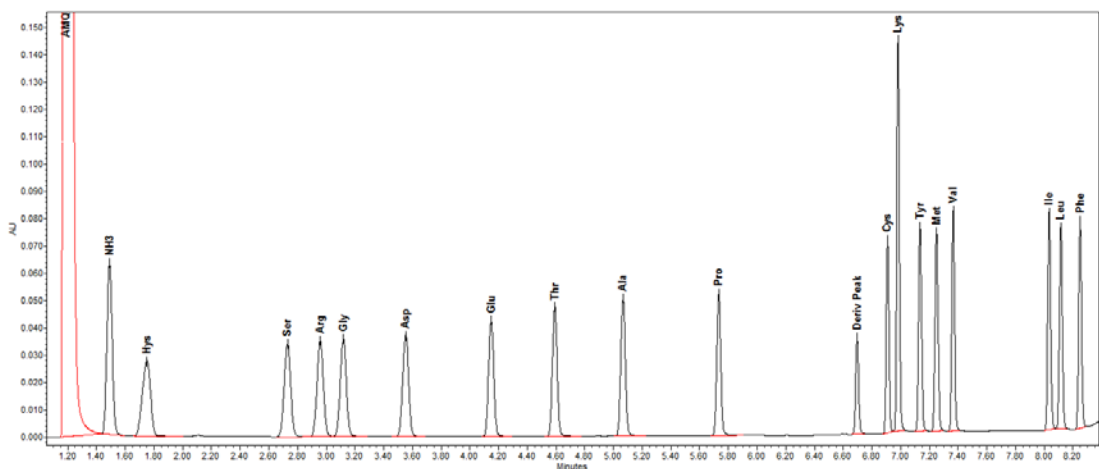
### *Materials and methods*

The Amino Acid Analysis (AAA) is a titration assay for the quantification of the protein content of analyte samples. AAA refers to a two-step procedure that quantitates amino acid molecules in proteins. The first step involves the digestion of a protein into free-amino acids, constituting the protein primary structure. The second step involves the chromatographic separation of the free-amino acids for the detection and quantification. AAA assay is based on the AccQ•TagUltra method, commercially available by Waters Corp. [133]. The method used the ACQ reagent (6-aminoquinolyl-N-hydroxysuccinimidyl carbamate) to add UV absorbance to free-hydrolysate amino acids. ACQ is an amine-derivatization reagent specific for primary and secondary amino acids (Figure 33). AAA measures the concentration of each free-amino acid after the protein hydrolyses by acid conditions. A commercial free-amino acids standard solution is used for external standard calibration.



**Figure 33:** *AccQ•Tag Ultra reaction for amino acid derivatization [133]*

The hydrolysis procedure is crucial for the quality of the titration data and must be properly optimized in order to identify all the amino acids constituting the protein structure; avoiding at the same time the degradation. Hence, our study was focused to apply AQBd framework to develop and optimize a robust protein hydrolysis step. The RP-UHPLC method and the derivatization procedure (Figure 33) is already developed and verified by Waters Corp. AccQ•Tag kit [133]. An example of the RP-UHPLC separation for the standard amino acids is reported on Figure 34.



**Figure 34:** *standard amino acids chromatogram by RP-UHPLC*





### *Chemicals and reagents*

For the hydrolysis step optimization and AAA method validation, the rMenB drug substances and a Bovine Serum Albumin (BSA) standard solution were employed.

NHBA-GNA130, fHbp-GNA2091 and NadA antigens are produced by GSK group of companies (Siena, Italy). BSA 7% water solution is certified by National Institute of Standard and Technology (NIST, SRM 927e) and was purchased by NIST (U.S. Department of Commerce, USA). 6 M hydrochloric acid (HCl) water solution (amino acid analysis grade) and 2.5 M amino acid hydrolysate standard mix solution in 0.1 M HCl (amino acid analysis grade) were purchased by Sigma-Aldrich (Saint Louis, MO USA). Phenol (C<sub>6</sub>H<sub>6</sub>O) solution 90% in water (ACS reagent grade) was purchased by Merck KGaA (Darmstadt, Germany). AccQ•Tag Ultra Derivatization kit (Eluent A AccQ•Tag; Eluent B AccQ•Tag; AccQ•Tag borato buffer; AccQ•Tag powder reagent; AccQ•Tag dilution reagent) and total recovery glass vials (*12x32mm glass screw neck vial, Quick thread, Lectrabond cap, PTFE/silicone septa Total Recovery*) were purchased by Waters Corp. (Milford, MA, USA). Ultrapure water was produced by Millipore Milli-Q system (Billerica, MA, USA). All solutions used have been filtered on a nylon membrane of 0.22 µm porosity using Nalgene clepsydra filter (Nalgene, Rochester, NY, USA).

### *Solutions and sample preparation*

The NHBA-GNA130, fHbp-GNA2091 and NadA antigen aliquots are stored at -20°C and thawed at room temperature before analysis. Such as positive control for hydrolysis performance, the BSA 7% water solution titrated and certified by NIST was used. The commercial available BSA NIST solution was portioned in 125 µl aliquots and stored at -20°C for two months. The working solutions were obtained by dilution, with ultrapure waters, of the protein solutions within the amino acid standard calibration range; considering the primary structures of each protein and the seven most abundant amino acids. The calibration curve is built using the 2.5 M hydrolysate standard amino acidic solution, diluted with ultrapure water to





cover the 15-250 nM concentration range. The protein working solutions were prepared each day and have been showed to be stable for 1 week after hydrolysis. After derivatization step, the hydrolysate solutions are stored in autosampler at 20°C using the total recovery, Lectabond cap, glass vials and are stable for 24 hours.

### *Equipment and analysis*

The rMenB proteins diluted in the working range are dried using a vacuum SpeedVac (thermo-fisher scientific, Waltham, MA, USA), at least for 2 hours at room temperature. The dry samples are reconstituted with 6 M HCl solution, adding of phenol solution such as antioxidant. The new sample mixture is left to hydrolyse overnight in a heating chamber using 4 ml clean glass vial with Teflon vite-cap, purchased by Wheaton Corp. (Millville, NJ, USA). Hydrolysis has been performed in a Binder oven (Bohemia, NY, USA). The hydrolysis working conditions (with the interval corresponding to the MODR) were as follows: 6 M HCl volume 300  $\mu$ l (233-420  $\mu$ l); PhOH volume 5  $\mu$ l (1.3-6.0  $\mu$ l); time 17 hours (16-21 hrs); temperature 112°C (105-118°C). Whereupon, the hydrolysate samples are neutralized by evaporation of the hydrochloric acid in a vacuum SpeedVac at least for 4 hours applying 45°C to accelerate the process. Hence, the dried free-amino acids samples are reconstituted with 0.1 M NaCl aqueous solution, for a better solubility.

Both for hydrolysate free-amino acids samples and standards, the derivatization procedure was performed mixing 10  $\mu$ l of working solution with 20  $\mu$ l of ACQ reagent and 70  $\mu$ l of AccQ•Tag borato buffer [133]. The mixing has been done directly in total recovery glass vials, sealing with not pre-slit lectabond caps, and derivatized in a thermoblock or an oven for 10 minutes (8-15 min) at controlled temperature, 55°C (50-60°C).

The chromatographic configurations for the Screening DoEs and RSM are: Acquity UPLC BEH RP-C18 1.7  $\mu$ m (2.1 x 150 mm) column and Acquity H-Class Bio UPLC system (Waters Corp.) equipped with bio-Quaternary Solvent Manager (bioQSM) with 100  $\mu$ l mixer, bio-Sample Manager (bioSM-FTN) with 15  $\mu$ l injector needle, column oven CH-A with



activated pre-heater and photodiode array detector (ACQ-PDA) with analytical flow-cell (10 mm; 500 nL) and a specific 0.0025 ID inlet peek for bioFTN and PDA connection. The detection wavelength was 260 nm (4.8 nm resolution and 10 points sec<sup>-1</sup> sampling rate frequency). Sample injections were done by 1 µl of the working solution stored in the autosampler at 20°C [133].

### Computations and software

In order to verify the hydrolyse conditions the qualified BSA standard purchased by NIST was used. The BSA protein amount certified by NIST, considering also the related amount confidence interval such as desirability range, has been reported in Table 14. Empower 3 software [124] licensed by Waters Corp. was used for the Acquity H-Class Bio UHPLC instrument control and for the Chromatographic data computation. According to European Pharmacopoeia 8.0 indications [134], by using Empower 3 custom fields for automatic computation, the formula applied for total protein amount determination (µg/ml), starting from free-amino acids concentration (nM) found by analysis, has been following reported on Equation 11;

$$\text{Protein average amount} = \frac{1}{N} \sum_{i=3}^{17} \frac{AA \text{ Conc}_i}{n^\circ AA_i} D_F \frac{MW_i}{1000}$$

**Equation 11:** AAA formula for conversion of amino acids concentration to protein amount

Were:

*Protein average amount* (µg/ml): is the average concentration of the protein sample

*N*: is the number of amino acids (*i*) considered for the computation of the average protein concentration, from a minimum of 4 to a maximum of 17.

*AA Conc<sub>i</sub>* (nM): is the concentration found for the *i*-esimo amino acid

*n°AA<sub>i</sub>*: is the abundance of the *i*-esimo amino acid in the protein primary structure

*D<sub>F</sub>*: is the starting dilution value of the protein sample for analysis

*MW<sub>i</sub>* (g/mol): is the molecular weight of the *i*-esimo amino acid

Empower custom fields have been programmed to automatically exclude by computations the free-amino acids out of the standard calibration



range (concentration values < 15 nM and > 250 nM) and the A.A. based protein amounts with a variability higher to 5 % respect to the average protein amount, according to European Pharmacopoeia 8.0 indications [134].

MODDE software [128] was purchased from S-IN (Vicenza, Italy) and was employed to generate the full Factorial Design (FFD) for screening DoE and to generate the graphic effects analysis and the Central Composite faced design (CCD) used for RSM. Moreover, to perform data analysis and to find the design space by means of risk failure maps computed using the Monte-Carlo simulations. The runs of all the DoE plans were carried out in a randomized order.

### ***AAA hydrolysis development and results***

The hydrolysis procedure development agreed the systematic AQbD framework described on Figure 19, previously applied to RP-UHPLC separation method, highlighting the flexibility and the goodness of the proposed approach for different analytical purposes. The following steps were applied:

- I) ATP definition, method scouting and CMAs;
- II) Risk assessment and pCMPs;
- III) MODR identification by RSM;
- IV) Working point definition;
- V) Method control strategy and validation.

### ***Analytical target profile, method scouting and critical method attributes***

For the AAA hydrolysis study, ATP consists to obtain the accurate quantification of the three Bexsero recombinant components NHBA-GNA1030, fHbp-GNA2091 and NadA. Moreover, general validation requirements according to ICHQ2(R1) guideline [129] [130] had to be fulfilled, including an adequate method throughput and feasibility to be

able to monitor rMenB antigen content in the drug substances. The AAA is important to quantify the protein concentration of rMenB reference materials to be used as external calibration standards for the antigen content determination in Bexsero vaccine. The component content is a product CQA of Bexsero. The selected CMAs are reported in Table 14 and were: BSA NIST amount (BSA),  $R_1$  (fHbp-GNA2091 amount),  $R_2$  (NHBA-GNA1030 amount) and  $R_3$  (NadA amount).

Abbreviation	CMA	CMA requirement ( $\mu\text{g/ml}$ )
BSA	BSA NIST amount	$59085 \leq BSA \leq 71204^\#$
$R_1$	fHbp-GNA2091 amount	$1351 \leq R_1 \leq 2027^*$
$R_2$	NHBA-GNA1030 amount	$615 \leq R_2 \leq 923^*$
$R_3$	NadA amount	$1671 \leq R_3 \leq 2507^*$

**Table 14:** AAA critical method attributes and selected requirements

# BSA amount by NIST certification; \* range target by microBCA variability.

In order to reach the method target, different preliminary aspects had to be considered. First, scouting of different hydrolysis types was performed. On the base of literature research [134] [135] [136] [137] gas- and liquid-hydrolysis were investigated to identify the best promising procedure for protein digestion. The gas-hydrolysis was investigated first by DoE experimentation for the following factors: 6M HCl volume (VOL), hydrolysis time (HT) and sample-support (SUPP). Each factor was studied at two levels ( $2^3$ ), 10-20 ml for VOL, 24-48 hours for HT and Backer-Pasteur for SUPP. MODDE software [128] was employed to generate the symmetric full factorial screening matrix, used to estimate the coefficients of the following linear model with interactions:

$$Y = b_0 + b_{1A}(X_{1A}) + b_{1B}(X_{1B}) + b_{2A}(X_{2A}) + b_{2B}(X_{2B}) + b_{3A}(X_{3A}) + b_{3B}(X_{3B}) + b_{1A2A}(X_{1A}X_{2A}) + b_{1A2B}(X_{1A}X_{2B}) + b_{1A3A}(X_{1A}X_{3A}) + b_{1A3B}(X_{1A}X_{3B}) + b_{1B2A}(X_{1B}X_{2A}) + b_{1B2B}(X_{1B}X_{2B}) + b_{1B3A}(X_{1B}X_{3A}) + b_{1B3B}(X_{1B}X_{3B}) + b_{2A3A}(X_{2A}X_{3A}) + b_{2A3B}(X_{2A}X_{3B}) + b_{2A3A}(X_{2B}X_{3A}) + b_{2A3B}(X_{2B}X_{3B})$$

where:  $X_1$  is VOL,  $X_2$  is CONC and  $X_3$  is SUPP,  $b_0$  is the constant term and  $b_i$  are the linear and interaction coefficients.



The model contains one constant term plus, for each factor, a number of terms equal to its number of levels and interactions. The 8-run experimental plan ( $2^3//8$ ) is reported in Table 15. Each multivariate condition was triplicated in order to obtain a reliable estimation of the experimental variance.

Exp. No	Volume (ml)	HT (hrs)	SUPP	BSA ( $\mu\text{g/ml}$ )	R <sub>1</sub> ( $\mu\text{g/ml}$ )
1	10	24	Backer	57333,8	1297,7
2	20	24	Backer	57735,3	1303,2
3	10	48	Backer	58717,9	1283,7
4	20	48	Backer	56472,6	1290,5
5	10	24	Pasteur	57926,7	1283,9
6	20	24	Pasteur	57757,8	1324,5
7	10	48	Pasteur	58345,0	1305,3
8	20	48	Pasteur	56328,5	1279,0
9	10	24	Backer	57851,8	1305,0
10	20	24	Backer	57948,9	1274,1
11	10	48	Backer	58008,7	1275,1
12	20	48	Backer	54544,0	1268,5
13	10	24	Pasteur	58064,5	1338,4
14	20	24	Pasteur	57655,8	1294,3
15	10	48	Pasteur	57212,8	1310,7
16	20	48	Pasteur	57255,9	1203,7
17	10	24	Backer	60232,2	1314,0
18	20	24	Backer	58254,5	1319,6
19	10	48	Backer	57906,9	1288,2
20	20	48	Backer	56375,4	1253,6
21	10	24	Pasteur	57811,1	1284,7
22	20	24	Pasteur	57507,0	1297,4
23	10	48	Pasteur	56384,3	1299,8
24	20	48	Pasteur	57748,8	1214,9

Table 15: 24-run symmetric screening matrix ( $2^3//8$ ) for gas-hydrolysis

For the hydrolysis screening, in order to manage a lower number of samples, only BSA and R<sub>1</sub> CMAs were investigated. For all the conditions tested lower protein amounts and data reproducibility were obtained for gas-hydrolysis with respect to liquid-hydrolysis, both for BSA NIST and fHbp-GNA2091 samples. The DoE-graphic effects analysis (Figure 35) shows a minimization of recovery increasing both VOL and HT factors.

No SUPP and interactions effects are identified. It means a low probability to fulfil CMAs requirement by optimization DoE, considering the low recovery. Contrariwise, the liquid hydrolysis was tested by OFAT experimentation, adapting literature instructions [134] [135] [136] [137], and providing good recovery results within the CMAs requirement ranges (Table 14). Moreover, considering the lower HCl volume needed for liquid-hydrolysis (cost saving) the faster reaction rate ( $\leq 24$ h, time saving) more compliant with the analyst's time organization (assay feasibility) and finally the higher system reproducibility (assay ruggedness), the liquid-hydrolysis was chosen for further proteolysis optimization.

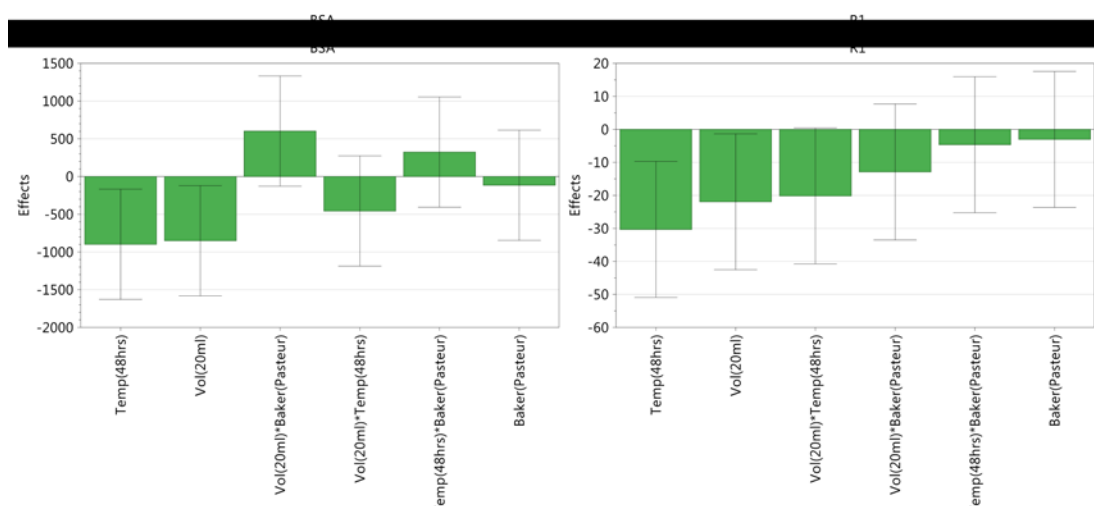


Figure 35: screening DoE-graphic effect analysis for gas-hydrolysis

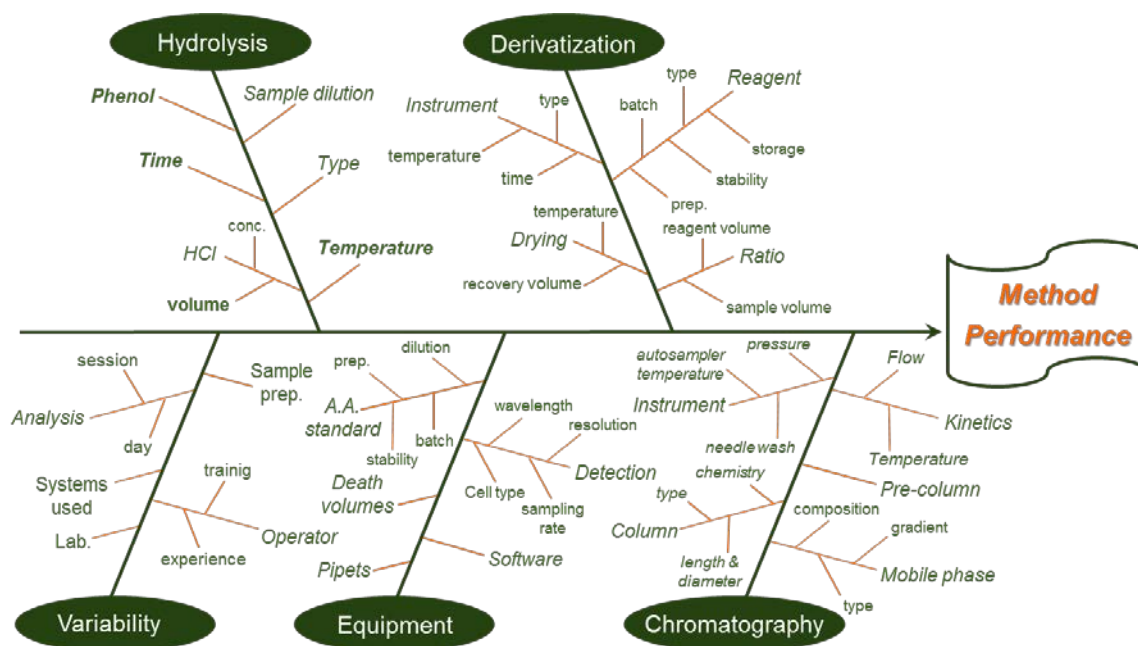
### Risk assessment and potential critical method parameters

The objective of a risk assessment is to develop understanding of procedure variables and their impact on the method reportable values for the identification of hazards and the analysis and evaluation of risks associated with exposure to those hazards [119] [131]. Tools such as process-map, fishbone diagram and cause-and-effect matrix are used, in addition to prior knowledge, to brainstorm the method factors and gain information-gathering to identify pCMPs [119]. In this study, a fishbone diagram (Figure 36) with the support of C&E matrix (Table 16:) were used to formalize the risk assessment and point out the risk factors associated with the characteristics of the AAA assay and thus to highlight the

potential CMPs which were supposed to potentially affect the selected CMAs.

Some of the CMPs, generally not related to hydrolysis step and associated to AAA chromatographic system and amino acid derivatization, had been already studied and fixed by Waters Corp. AAA RP-UHPLC procedure [133]. For this reason the Ishikawa diagram was formalized for all AAA assay factors, step I (Hydrolysis) and step II (derivatization and chromatography), while the C&E matrix was prevalently focused to investigate the effect of the hydrolysis factors on proteolysis performance (AAA step I); considering the others category factors (AAA step II) well known and controlled (high knowledge level).

The hydrolysis pCMPs identified, coloured by red in the Table 16, needed to be risk managed and in-depth studied by DoE to enhance knowledge on their effects on hydrolysis performances.



**Figure 36:** Ishikawa diagram for AAA assay

Variables Category	Parameters	Units	Accuracy	Precision	Specificity	Range	Analysis time	throughput	n° valid run	Severity score (S)	Knowledge (K)	Criticality
Hydrolysis	Sample dilution	N/A	7	7	4	10	1	1	1	10,0	10,0	MP
	Hydrolysis type	N/A	10	7	7	4	10	7	10	10,0	10,0	MP
	Hydrolysis time	hrs	7	4	4	4	7	1	7	7,0	1,0	pCMP
	Hydrolysis temperature	°C	10	4	4	1	1	1	1	10,0	4,0	pCMP
	HCl volume	µl	7	7	1	10	1	1	7	10,0	4,0	pCMP
	Phenol volume	µl	7	7	1	4	1	1	7	7,0	1,0	pCMP
Derivatization	Derivatization temperature	°C	7	7	1	1	1	1	4	7,0	10,0	MP
	Derivatization time	min.	7	7	1	1	1	1	4	7,0	10,0	MP
	Instrument for derivatization	N/A	1	1	1	1	1	1	4	4,0	10,0	MP
	sample recovery after hydrolysis	µl	4	4	4	1	1	1	1	4,0	10,0	MP
Chromatography	Column type	N/A	7	7	1	10	1	1	7	10,0	10,0	MP
	Mobile Phases composition	N/A	7	10	4	1	1	1	4	10,0	10,0	MP
	Mobile Phases gradient	%/min.	7	10	4	1	1	1	4	10,0	10,0	MP
	auto-sampler stability	hrs	7	10	1	1	1	1	7	10,0	10,0	MP
Variability & Equipment	Sample preparation	N/A	4	10	1	1	1	1	7	10,0	10,0	MP
	hydrolysis session	N/A	4	10	4	1	1	1	7	10,0	10,0	MP

**Table 16: C&E matrix for hydrolysis performance**

Severity score (S) defines the impact of the method parameters on CMAs. S score is computed as the maximum impact factor assessed by C&E matrix, using a severity discrete ranking from 1 to 10, were: 1 (no impact), 4 (low impact), 7 (impact) and 10 (high impact). Method parameters with S score  $\geq 7$  are considered pCMPs.

Knowledge score (K) defines the knowledge/uncertainty about the factor-CMA relationship and the controllability of the parameters. K score is used to prioritize actions for factors effect investigations and to demote pCMPs to not-critical if they are full known and controlled. For knowledge assessment was used a discrete ranking from 10 to 1, were: 10 (high knowledge and control), 7 (knowledge on CMAs effect), 4 (low factor knowledge) and 1 (no factor knowledge and high uncertainty). Method parameters with K = 10 are demote to not-critical, because known and controlled, not necessary to be studied further more.

### Design space definition by response surface methodology

Due to low number of factors identify as pCMPs by risk assessment, all the hydrolysis parameters were investigated directly by optimization DoE. The four parameters identify as potentially critical for the hydrolysis performance were represented by hydrolysis time (HT), temperature (T), 6 M HCl volume (VOL) and antioxidant phenol volume (PhOH).

RSM was applied for in-depth investigations of the effects of HT, T, VOL and PhOH on the selected CMAs (Table 14), applying the new experimental domain reported in Table 17.



A three-level four-factor faced CCD ( $3^4$ ) was employed for building a quadratic model with interactions, making it possible to study all the selected chromatographic CMAs: BSA,  $R_1$ ,  $R_2$  and  $R_3$  (Table 14).

The 25-run CCD experimental plan ( $3^4//25$ ) with the measured responses is reported in Table 17, including three repetition of the central point to estimate the response variability.

Exp No.	HT (hrs)	T (min)	VOL ( $\mu$ l)	PhOH ( $\mu$ l)	BSA ( $\mu$ g/ml)	$R_1$ ( $\mu$ g/ml)	$R_2$ ( $\mu$ g/ml)	$R_3$ ( $\mu$ g/ml)
1	10	104	100	0	62661,8	1349,5	560,6	1642,8
2	24	104	100	0	63035,3	1683,4	637,7	1734,1
3	10	120	100	0	64018,0	1088,4	628,5	1738,6
4	24	120	100	0	62286,1	1544,4	624,9	1749,6
5	10	104	500	0	62237,0	1167,5	628,2	1608,8
6	24	104	500	0	63892,4	1660,7	607,2	1675,1
7	10	120	500	0	63399,4	1316,1	614,5	1664,1
8	24	120	500	0	57709,4	1487,6	670,7	1672,0
9	10	104	100	10	62127,3	989,2	617,7	1635,5
10	24	104	100	10	65165,5	1783,5	566,9	1717,2
11	10	120	100	10	64175,9	1470,7	656,1	1767,5
12	24	120	100	10	64264,1	1625,0	663,2	1728,9
13	10	104	500	10	62650,9	1047,1	629,9	1623,0
14	24	104	500	10	64090,8	1599,0	735,9	1718,2
15	10	120	500	10	64583,3	1451,2	616,1	1758,6
16	24	120	500	10	64121,8	1585,7	634,1	1756,9
17	10	112	300	5	52241,1	1472,5	634,6	1707,0
18	24	112	300	5	55927,0	1524,8	646,6	1825,3
19	17	104	300	5	63387,7	1487,0	598,4	1709,4
20	17	120	300	5	65443,0	1536,6	635,5	1742,0
21	17	112	100	5	76184,4	1536,2	620,5	1740,8
22	17	112	500	5	60046,0	1530,6	630,3	1739,8
23	17	112	300	0	58191,9	1500,5	631,9	1733,5
24	17	112	300	10	76436,3	1574,7	642,9	1760,9
25	17	112	300	5	64458,5	1545,8	642,3	1717,6
26	17	112	300	5	69565,5	1500,1	635,7	1728,6
27	17	112	300	5	65024,1	1507,6	630,3	1765,6

**Table 17:** Response surface methodology: 25-run Central Composite Design experimental plan with three central points.

Only the  $R_1$  responses were analysed without mathematical transformation. All the others responses were mathematical transformed to stabilize the variance and makes the data more normal distributed-like.

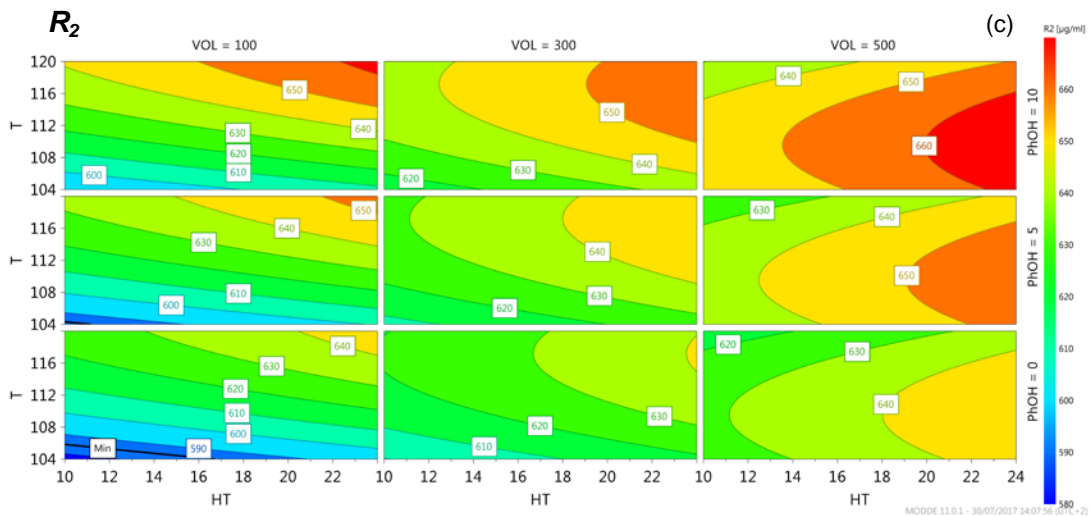
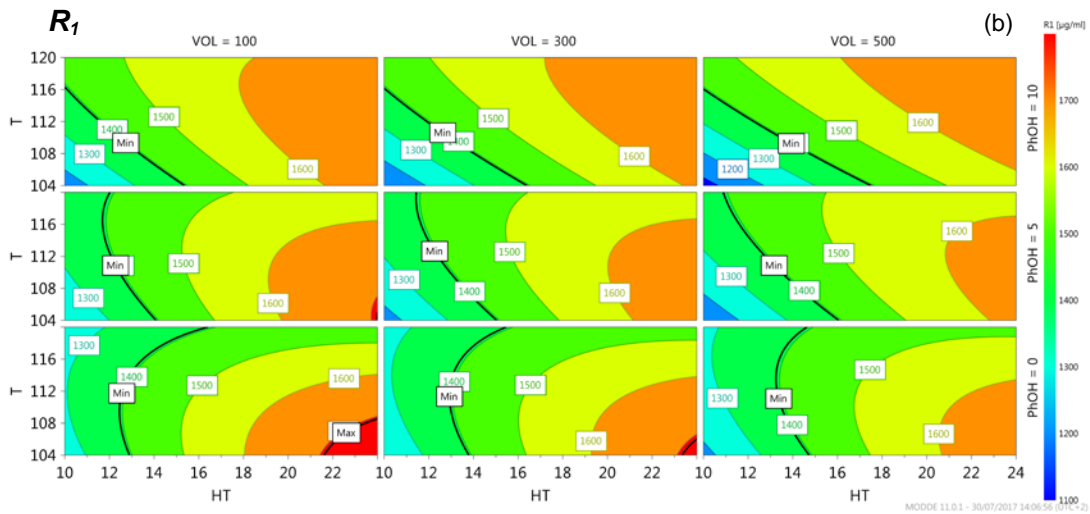
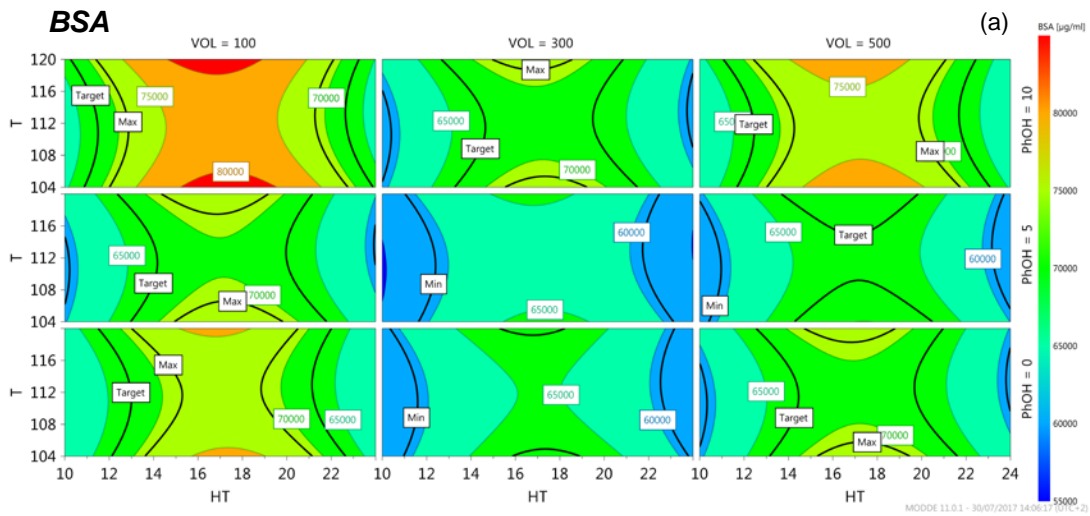
Following the Box-Cox plot indications, BSA and  $R_3$  responses were  $Y^{-2}$  transformed, while  $R_2$  responses were  $Y^{-1.5}$  transformed. All the four models, calculated by multiple-linear regression, were significant in terms of ANOVA, without lack of fit. The goodness of fitting (expressed by determination coefficient  $R^2$ ) and of prediction (expressed as cross validated  $Q^2$ ) of the four identified models are reported in Table 18. BSA and  $R_2$  have a low  $Q^2$  due to the low data variability. It means that the effect on output responses of the hydrolysis factors variation is near to the pure model error (response variability estimated by central point repetitions). Additionally, as observable by CCD experimental matrix (Table 17) and response surfaces (Figure 37), the BSA and  $R_2$  results are always within the CMAs desirably ranges (Table 14).

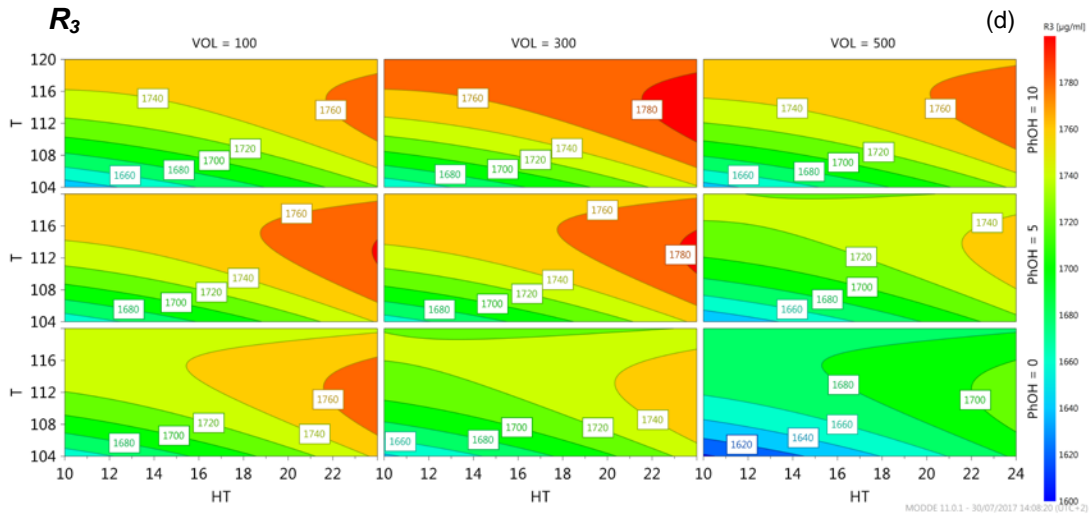
CMA	$R^2$	$Q^2$	Regression	Lack of fit
			<i>p</i> value	<i>p</i> value
BSA	0.598	0.386	0.032	0.285
$R_1$	0.842	0.504	0.000	0.058
$R_2$	0.452	0.294	0.038	0.044
$R_3$	0.910	0.710	0.000	0.737

**Table 18:** *Response surface methodology: quality of the calculated models for all the CMAs.*

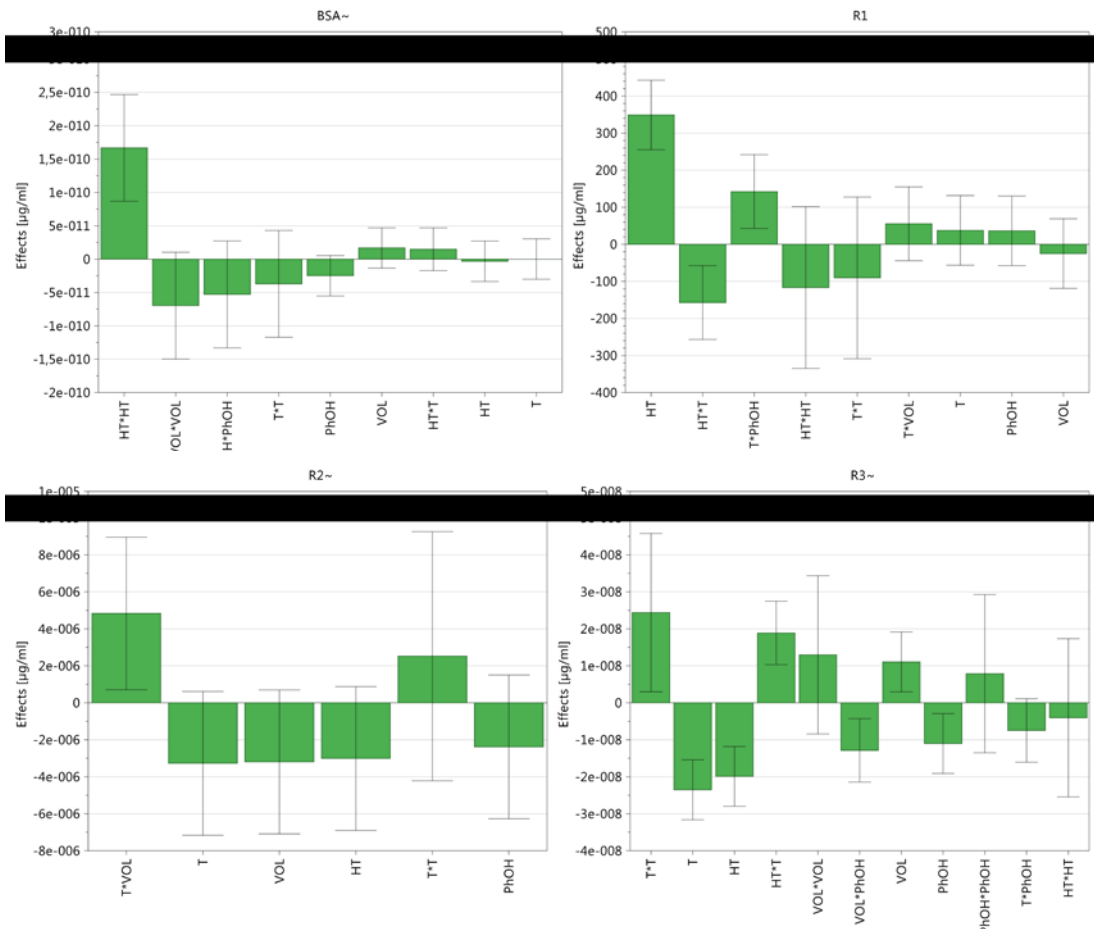
In order to obtain detailed information of the behaviour of each CMAs throughout the experimental domain investigated, the contour plots were drawn reporting the calculated isoresponse curves (Figure 37).

The analyses of contour plot response surfaces and models coefficients (Figure 38) allow understanding of the method performance behaviour in function of selected CMPs.





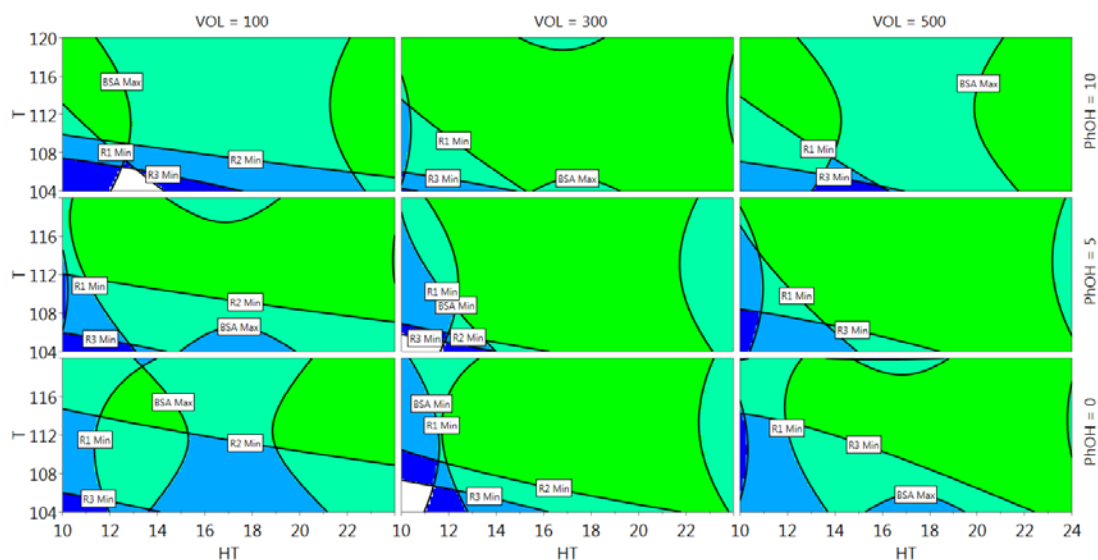
**Figure 37:** Isoresponse surfaces drawn by plotting HT vs. T, at three different values of VOL (100  $\mu$ l, 300  $\mu$ l, 500  $\mu$ l) and three different values of PhOH (0  $\mu$ l, 5  $\mu$ l, 10  $\mu$ l): (a) BSA, (b) R<sub>1</sub>, (c) R<sub>2</sub>, (d) R<sub>3</sub>.



**Figure 38:** Response Surface Methodology: graphic analysis of effects of the CMPs on the CMAs

The sweet spot plots (Figure 39) highlight by different colours the areas where one or more predicted CMAs fulfil the related requirements (Table 14). The different colours had the following meaning: blue only one CMA meet; brilliant blue two CMAs meet; light blue three CMAs meet; green where all the four CMAs criteria were fulfilled, namely the zone corresponding to the sweet spot.

The green plot area does not constitute the design space. Sweet spot only shows where the response surfaces overlay meets the target performance for all CMAs, providing information about the space where the maximum numbers of criteria are fulfilled.



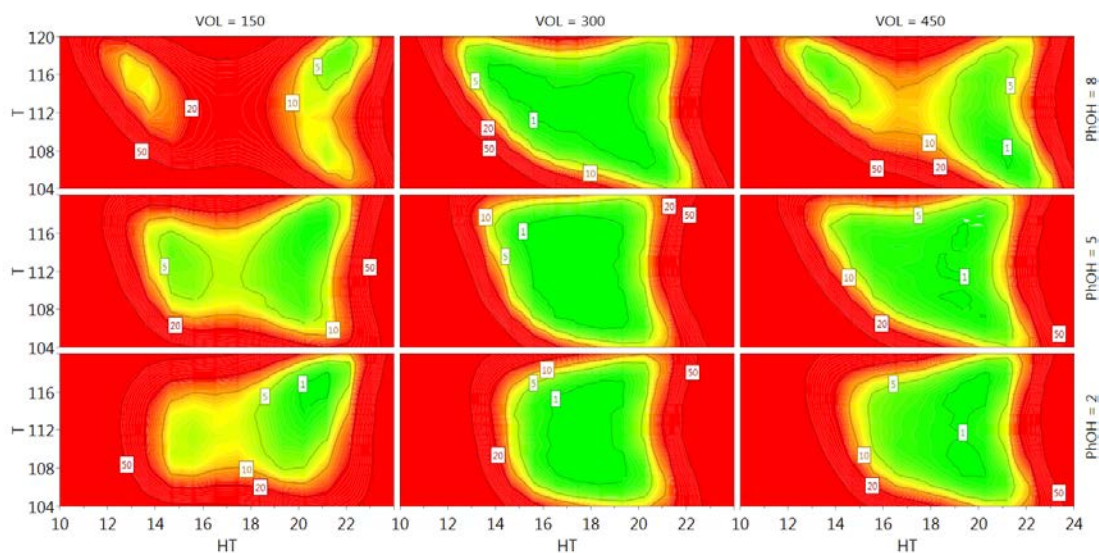
**Figure 39:** Sweet spot plot obtained plotting HT vs. T at three different values of VOL and PhOH

The green colour corresponds to the zone where the requirements for all the four CMAs are fulfilled.

The sweet spot plot is identify by desirability function, considering all the response surface models and the CMAs target requirements. To achieve the MODR needs to be taken into account also the concept of probability that the requirements are met [2] [7]. The analytical design space was computed by MODDE software [128] and the risk of failure map (Figure 40) was drawn using Monte-Carlo simulations. Monte-Carlo risk analysis takes into account the model parameters uncertainty, propagating the

prediction error from parameters to responses and giving access to the responses distributions for each RSM condition [132].

The selected level of probability ( $\pi$ ) and the original set point for MODR identification were:  $\pi \geq 95\%$ ; HT, 18.4 hrs; T, 107°C; VOL, 393  $\mu\text{l}$ ; PhOH, 2.7  $\mu\text{l}$ . The design space was graphically represented in the risk of failure map by the area coloured in green, within the iso-error curves corresponding to the probability failure  $\leq 5\%$ . Within the MODR, the following limits for CMPs ranges were identified: HT, 16.0-21.0 hrs; T, 105-118°C; VOL, 233-420  $\mu\text{l}$ ; PhOH, 1.3-6.0  $\mu\text{l}$ .



**Figure 40:** Hydrolysis design space map by plotting HT vs. T, at three different values of VOL: 150  $\mu\text{l}$ , 300  $\mu\text{l}$ , 450  $\mu\text{l}$  and three different values of PhOH: 2  $\mu\text{l}$ , 5  $\mu\text{l}$ , 8  $\mu\text{l}$ .

The selected working point for routine use was: HT, 17 hrs; T, 112°C; VOL, 300  $\mu\text{l}$ ; PhOH, 5  $\mu\text{l}$ . The selection was performed taking into account the assay feasibility (17 hours of hydrolysis are compatible with the overnight time) and additional practicability considerations, such as a suitable phenol volume to be withdrawn by automatic pipets (5  $\mu\text{l}$ ) and to reduce the exposition to the chemical substance. The MODR plateau response and the related risk of failure were additionally considered: 112°C and 300  $\mu\text{l}$  of HCl volume (combined with 5  $\mu\text{l}$  of phenol and 17 hours of hydrolysis) falls within the lower iso-error curve ( $\pi < 1\%$ ), with



good response robustness (area plateau). The MODR model and the selected set point for routine use were validated by the agreement of the RSM predicted value with the experimental data.

### *Method control strategy and validation*

The qualification of the *Neisseria meningitidis* recombinant antigens by AAA method was carried out to ensure the suitability of the quality control of the vaccine components in the pharmaceutical drug product for humane use. The AAA assay is employed for the quantification of rMenB standard proteins to be used for calibrations in antigen content methods, e.g. the RP-UHPLC method for unadsorbed antigen content determination. In order to ensure that the standard proteins are well titrated, there is the necessity to be sure that AAA assay steps, hydrolysis (step I) and chromatography (step II), are well controlled and are performing according to quality requirements (Table 14).

The control strategy for hydrolysis step has been planned by using the BSA NIST as positive control of the hydrolysis performance. On the base of the data collected by RSM, if the hydrolysis is complete and no issues are occurred during step I, for the BSA sample needs to be finding the NIST certificate amount values (concentrations used as CMAs requirements for RSM). The titration result of BSA positive control is used such as system suitability for the performance of the hydrolysis reaction in each analytical session. Additionally, each protein sample and control is prepared in triplicate and the sample preparation variability (expressed as minimum / maximum ratio) need to be equal or lower respect to amount ratio variability found in validation:  $\geq 0.95$  for fHbp-GNA2091,  $\geq 0.90$  for NHBA-GNA1030 and  $\geq 0.94$  for NadA.

An additional chromatographic system suitability based on amino acids separation and peaks area precision is performed according to RP-UHPLC AccQ•Tag method instructions [133] to control the AAA step II. The chromatographic SST is performed before each analysis session to be sure that the RP-UHPLC system has the required chromatographic performance.



Finally, the AAA assay was validated according to ICH guidelines [130], quantitative test for titration. For the assay validation it was used two experimental plans for testing the effect of noise factor on results variability: operator, column batch, sample preparation, instrument, day of analysis and analytical session. The validation results showed the suitability of the method for the intended purpose: identity respect to standard solutions, specificity (for the free-hydrolysed amino acids respect to sample matrix), accuracy (recovery for antigens spike), precision (data variation of the titrated amount) and sample linearity (the same protein amount is obtained for different protein dilutions within the amino acids working range and the area response is linear). The AAA assay satisfies the required data quality and the summary results of the method validation are reported on Table 19.

Validation parameter	Acceptance criteria	Result	Conclusion
Accuracy	Recovery percentage should be in the range 80-120% with 90% CI	90% CI ranges between 83% to 107%	PASS
Repeatability	RSD% < 8%	RSD% < 3%	PASS
Intermediate precision	RSD% < 10%	RSD% < 4%	PASS
Sample linearity	$R^2 > 0.98$	$R^2 > 0.996$	PASS
Specificity	Specificity for free-hydrolysate amino acids	No drug substance matrices effects; specificity for hydrolysate amino acids	PASS
Range	the data quality are valid in the following working ranges:		
<i>fHBp-GNA2091</i>	34 – 145 µg/ml		
<i>NHBA-GNA1030</i>	35 – 110 µg/ml		
<i>NadA</i>	34 – 187 µg/ml		

Table 19: AAA validation summary results





## OMV Protein Pattern Profile by RP-UHPLC

The OMV component of Bexsero vaccine, previously introduced and illustrated in Figure 18, is a stable colloidal suspension that consists of small membranous spherical vesicles in which the native complex antigen composition of the sub-capsular cell surface of *Neisseria meningitidis* serogroup B is highly preserved (Figure 17). OMV contains the most abundant proteins of the outer membrane, PorA and PorB proteins are the main expressed, but also involve some minor outer membrane proteins such as OmpC, FetA, OmpA, fbpA (and many others) and Lipooligosaccharides (LPS) [41] [84] [85] [89].

The analytical method currently applied for the quality control of OMV proteins identity and purity is a densitometry SDS-PAGE assay. With the aim of developing a more state of the art method to replace the SDS-PAGE, an ATP was developed (see following ATP section) and subsequently a method screening exercise was conducted that led to the selection of the RP-UHPLC (data not shown). In this context, the screening of commercially available UHPLC reverse phase columns was conducted using elements of the AQbD framework. Results are shown in this chapter.

### *Material and methods*

The OMV bulk solution is produced by GSK group of company (Siena, Italy) and is stored at  $4 \pm 2^\circ\text{C}$  controlled temperature. The OMV bulk was injected as it, without pre-dilution.

### *Chemicals and reagents*

Trifluoroacetic acid  $\geq 99\%$   $\text{CF}_3\text{COOH}$  (TFA, LCMS & HPLC grade), formic acid  $\text{HCOOH}$  (FA, LCMS & HPLC grade), perfluoro-pentanoic acid 97%  $\text{CF}_3(\text{CF}_2)_3\text{COOH}$  (PFPA, HPLC grade) were purchased by Sigma-Aldrich (Saint Louis, MO, USA). Methanol  $\text{CH}_3\text{OH} \geq 99.9\%$  (HPLC grade) was purchased by Merck KGaA (Darmstadt, Germany). Acetonitrile (ACN) 99.8% (LC-MS grade) was purchased by Panreac (Radnor, PA, USA).



Trypsin gold (Mass spectrometry grade) was purchased by Promega Corp. (Madison, WI, USA). Ultrapure water was produced by Millipore Milli-Q system (Billerica, MA, USA). All solutions used were filtered on a nylon membrane of 0.22  $\mu\text{m}$  porosity, using Nalgene clepsydra filters (Nalgene, Rochester, NY, USA).

### *Chromatographic equipment*

Chromatographic columns tested:

- Acquity RP-C4 BEH 300Å, 1.7 $\mu\text{m}$ , 2.1x150 mm (*BEH C4*) from Waters Corp. (Milford, MA, USA);
- Aeris WIDEPORE C4 200Å, 3.6  $\mu\text{m}$ , 4.6x150 mm (*AWPC4\_150*);
- Aeris WIDEPORE C4 200Å, 3.6  $\mu\text{m}$ , 4.6x250 mm (*AWPC4\_250*);
- Aeris WIDEPORE XB-C8 200Å, 3.6  $\mu\text{m}$ , 4.6x100 mm (*AWP C8*);
- Jupiter C5 300Å, 5  $\mu\text{m}$ , 2.0x150 mm (*Jupiter*) from Phenomenex (Torrance, CA, USA);
- ProSwift RP-3U monolithic column, 4.6x50 mm (*RP-3U*) from Thermo-Fischer scientific (Waltham, MA, USA).

The chromatographic configurations used for columns scouting and experimental designs were: NexeraX2 method scouting UHPLC series 30 system equipped with LC-30AD pump, DGU-20A5R degasser unit and LPGE-unit, SIL-30AD autosampler, CTO-20AC oven with 180  $\mu\text{l}$  mixer, FCV-34AH UHPLC switching valve, SPD-M30A PDA detector (detection wavelength 280 nm; 4 nm resolution) and high sensitive flow-cell (85 mm; 9  $\mu\text{l}$ ) from Shimadzu Corp. (Kyoto, Japan).

The LC-MS configurations for mass analysis were: Acquity H-Class Bio UHPLC system (equipped with bioQSM, bioSM-FTN and ACQ-PDA) from Waters Corp. (Milford, MA, USA) coupled with Exactive EMR system from Thermo Scientific (Waltham, MA, USA). The TriVersa NanoMate system (Advion, Inc. Ithaca, NY, USA) was adapted for LC-MS interface in order to reduce UHPLC flow and collect fractions.



### *Computation and software*

LabSolution Version 5 software equipped with Method Scouting start-up kit and licensed by Shimadzu Corp. [125] was used for the NexeraX2 UHPLC instrument control and for the chromatographic data computation of purity and resolution of each chromatographic peak. Chromatographic resolutions (R) between two adjacent peaks were calculated using the retention times ( $t_R$ ) and the peak widths at half height (w), according to the following formula:

$$1.18(t_{R2} - t_{R1})/(w_1+w_2).$$

Purity of each chromatographic peak (A%) was calculated by the percentage ratio of the peak Area ( $A_i$ ) respect to total Area ( $A_{TOT}$ ), according to the following formula:

$$A\% = (A_i/A_{TOT}) \times 100$$

Peak capacity was controlled counting the number of resolved peaks for each chromatographic run.

BioPharma Finder Version 2.0 (Thermo Scientific) [138] and PEAKS Studio Version 8.0 (BioInformatics Solutions) [139] software were employed for computations of intact mass and MS peptide mapping respectively.

### *RP-UPLC scouting and results*

Elements of the AQbD approach were applied to RP-UPLC technologies screening for ensuring clear information are obtained with a limited number of experiments. The following steps of the AQbD approach [7] have been applied to RP-UHPLC OMV protein pattern:

- I) Analytical target profile and critical method attribute definition;
- II) Method scouting;
- III) Quality risk assessment and identification of potential critical method parameters;
- VI) Investigation of knowledge space by screening DoE.



### *ATP definition and CMAs*

Upfront definition of the desired method performance is crucial to define the intended purpose of the method, by formalization of the Analytical Target Profile. For the SDS-PAGE OMV protein pattern replacement, the ATP main requirements were the following:

- At least the same performance of the SDS-PAGE assay in terms of number of outer membrane proteins detected;
- Method selectivity and specificity for PorB and FbpA proteins (not resolved by SDS-PAGE assay);
- Improved method throughput and assay robustness;
- General validation requirements according to ICHQ2(R1) guideline [129] [130].

For the scouting of the OMV protein pattern RP-UPLC columns, the ATP requirements were translated to the following CMAs surrogates:

- (I) *Peaks capacity*: capability to separate the maximum possible number of peaks. At least the eight outer membrane proteins identified by SDS-PAGE.
- (II) *Columns selectivity*: capability to identify the outer membrane proteins with base line resolution.

### *Column scouting*

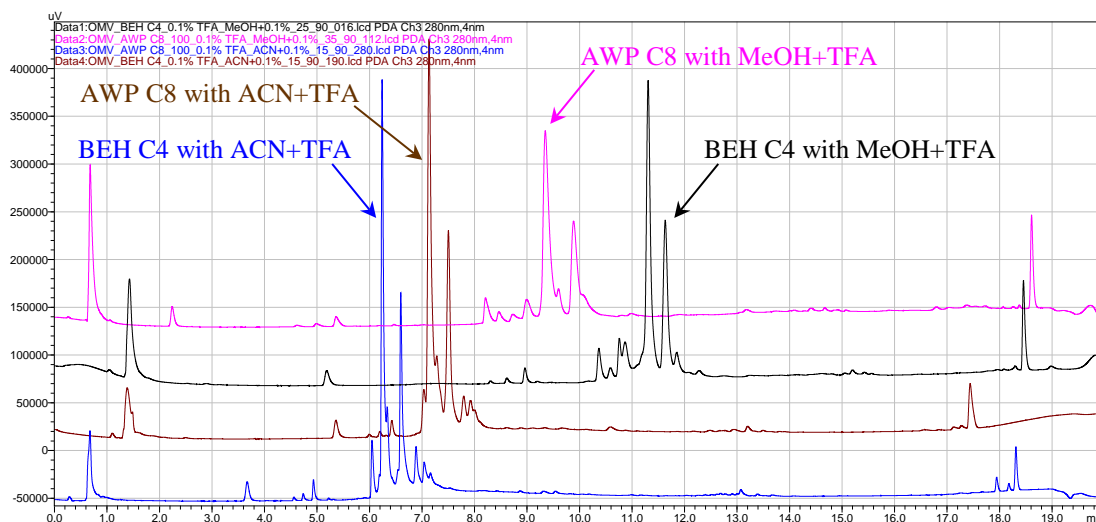
The selection process of RP column technology starts with the listing all the commercially available columns, with technology compatible with the goal of the method to be developed. Focusing prevalently to UHPLC columns, the selected stationary phases (Table 20) have been chosen to among the following chemistry (RP-C4, RP-C5 and RP-C8), supports (silica-particles and polypropylene-particles), column lengths (100, 150, 250 mm), particles porosity (130, 200, 300 Å) and column technology (pore-particles, solid-core and monolithic).

Column	Description	Supplier	Abbreviation
Acquity RP-C4 BEH 300Å, 1.7µm, 2.1x150 mm	silica pore	Waters Corp.	BEH C4
Aeris WIDEPORE C4 200Å, 3.6µm, 4.6x150 mm	silica solid-core	Phenomenex	AWPC4_150
Aeris WIDEPORE C4 200Å, 3.6µm, 4.6x250 mm	silica solid-core	Phenomenex	AWPC4_250
Aeris WIDEPORE XB-C8200Å, 3.6µm, 4.6x100 mm	solid-core endcap	Phenomenex	AWP C8
ProSwift RP-3U, 4.6x50 mm	PP monolithic	Thermo-Fischer	RP-3U
Jupiter C5 300Å, 5µm, 2.0x150 mm	silica pore	Phenomenex	Jupiter

**Table 20:** reverse phase columns selected for scouting

The NexeraX2 method scouting system allows the simultaneous screening of all the six columns in only one analytical session by automation testing of different mobile phases and linear gradients from low to high organic concentrations. The organic phases tested by OFAT experimentation were ACN + 0.1% TFA and MeOH + 0.1% TFA.

The starting and final organic concentrations studied for linear gradient were: 5, 15, 25, 35, 45 % for starting and 5, 100, 90, 80, 60 % for final ramp concentration. Always, after the linear gradient, two minutes of washing step and 2 minutes of re-equilibration (in starting conditions) were performed. The results achieved by RP-UHPLC method scouting are reported on the following Figure 41, using the BEH C4 and the AWP C8 columns and represents the closest to the target.



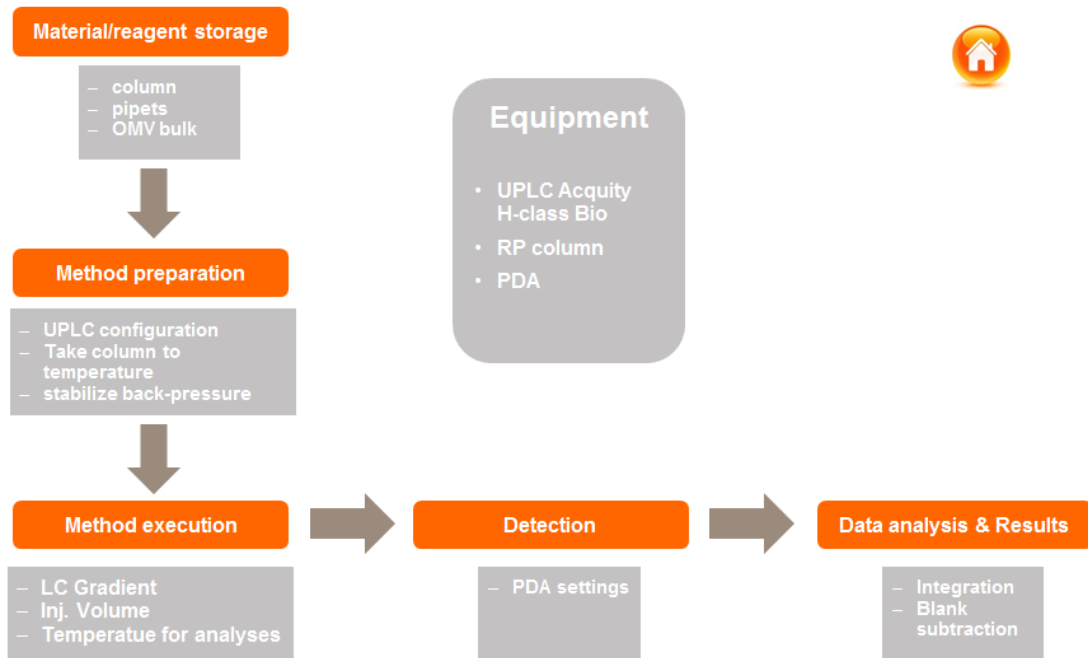
**Figure 41:** RP-UHPLC scouting results



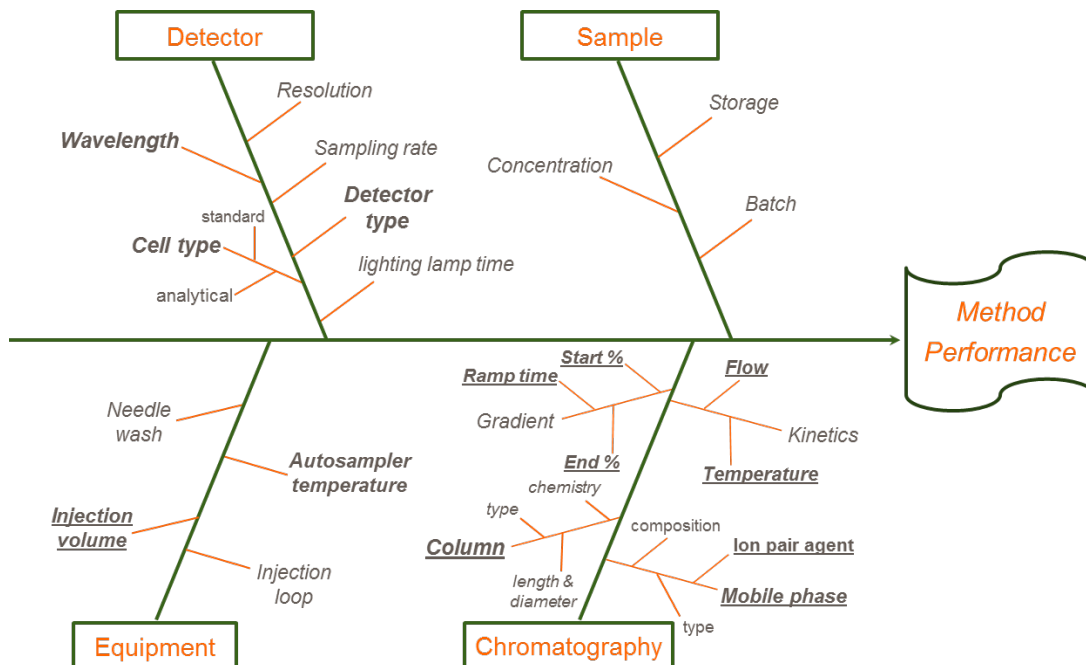
As a consequence of the preliminary screening, the *BEH C4* and the *AWP C8* were selected to be further studied by the following screening DoEs.

### *Quality Risk Assessment and pCMPs*

After the method scouting, the process mapping [98] and fishbone diagram [7] were employed to develop understanding of method variables and their impact on the method reportable values (peak capacity and selectivity), for the identification of hazards and the analysis and evaluation of associated risks [119] [131]. The risk assessment exercise has been formalized in the following analytical method map (Figure 42) and Ishikawa diagram (Figure 43). The pCMPs, highlighted bold in Figure 43, including detector and cell type, wavelength and autosampler temperature, are out of the scope of the optimization exercise since already studied and fixed by preliminary scouting experimentation (Figure 41). All the rest of the pCMPs, highlighted in bold and underlined in Figure 43, are in the scope of the risk management exercise and of the DoE with the objective of finding optimal method condition and achieve relevant knowledge on their effects on method performances (confirmation of criticality of pCMPs and definition of the design space).



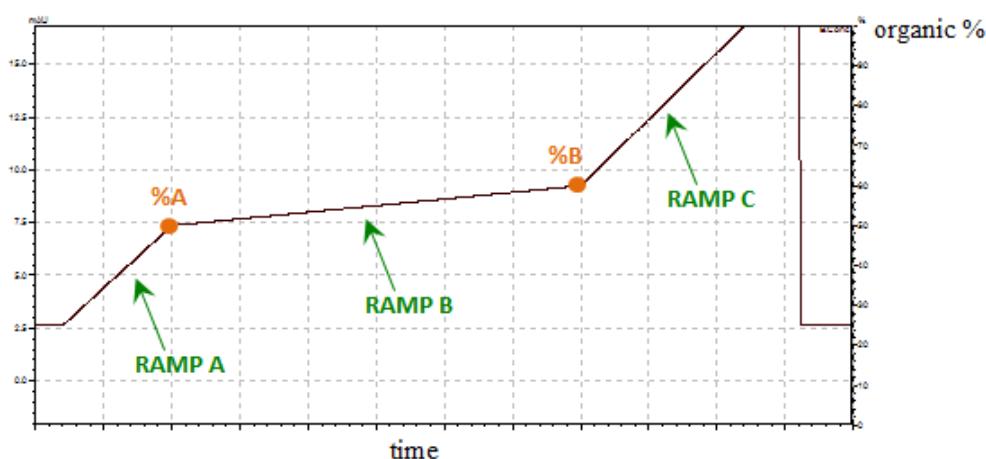
**Figure 42:** RP-UHPLC analytical method map



**Figure 43:** fishbone diagram for RP-UHPLC OMV protein pattern

### Screening experimental designs

As a result of RP-UHPLC scouting, the gradient shape reported on Figure 44 has been designed to optimize the chromatographic selectivity. As a result of quality risk assessment (Figure 42 and Figure 43) the selected pCMPs to be investigated by DoE were represented by mobile phase type (MP), column type (COL), column temperature (TEMP), injection volume (VOL), time for gradient ramp A (RAMP A), time for gradient ramp B (RAMP B), time for gradient ramp C (RAMP C), starting organic % of ramp B (%A), ending organic % of ramp B (%B), mobile phase flow (FLOW) and ion-pairing agent (ION).



**Figure 44:** RP-UHPLC gradient

As per ATP, the CMAs selected as output for the study are the following: the total numbers of peaks (N), the number of peaks in the gradient ramp B (nB), the PorB capacity factor ( $K'$ ) and the resolution between PorB and PorA outer membrane protein ( $R_1$ ).

Considering the high number of method parameters, the effect pCMPs onto CMAs were investigated by two different experimental designs. In a first 20-run asymmetric ( $2^{33^5}/20$ ) D-optimal design, the effect of MP, TEMP, VOL, COL, ION, RAMP A, RAMP B and RAMP C was investigated (Table 21). A centre point was included in the design and each run was duplicated in order to obtain a reliable estimation of the experimental variance (42 total runs, G-efficiency: 81%) and identifying the best conditions for the following factors: COL, AWP C8; TEMP, 70°C;





SOL, ACN; RAMP A, 6 min.; RAMP C, 10 min.; INJ, 10  $\mu$ l. The DoE-graphic effects analysis has been shown in Figure 45.

Whereupon, a new 16-run fractional factorial ( $2^5//16$ ) resolution V design, was employed to in depth study the effects of RAMP B and ION, with addition of %A, %B and FLOW factors, onto selected CMAs. Each run was duplicated in order to obtain a reliable estimation of the experimental variance (Table 22). By the DoE-graphic effects analysis (Figure 46) the best factors condition identified were: ION, TFA; RAMP B, 15 min.; %A, 30 %; %B, 40 % and FLOW, 0.5 ml/min.



Exp No	MP	ION	TEMP °C	VOL µl	COL	RAMP A min.	RAMP B min.	RAMP C min.	R <sub>1</sub>	N	nB	K'
1	ACN	TFA	60	6	BEH C4	2	2	2	2,353	11	4	2,600
2	ACN	AF	50	2	AWP C8	6	2	2	1,868	8	3	3,432
3	MeOH	TFA	50	10	AWP C8	2	6	2	2,231	15	7	11,629
4	ACN	TFA	70	2	BEH C4	10	6	2	8,351	13	9	3,587
5	MeOH	AF	70	6	BEH C4	6	10	2	0,044	8	1	5,513
6	MeOH	AF	60	10	AWP C8	10	10	2	0,691	10	2	16,956
7	MeOH	TFA	70	10	BEH C4	6	2	6	1,596	14	8	5,684
8	ACN	AF	60	10	AWP C8	10	2	6	0,678	7	2	3,568
9	ACN	AF	70	6	AWP C8	2	6	6	0,001	5	1	2,500
10	MeOH	TFA	60	2	AWP C8	6	6	6	3,104	15	9	15,992
11	ACN	AF	50	10	BEH C4	2	10	6	0,162	7	3	1,606
12	ACN	AF	50	2	BEH C4	10	10	6	0,648	7	2	2,568
13	MeOH	TFA	60	6	AWP C8	10	10	6	3,370	13	8	23,554
14	MeOH	AF	70	2	AWP C8	2	2	10	0,632	9	2	6,036
15	MeOH	TFA	50	6	BEH C4	10	2	10	1,026	12	5	8,483
16	ACN	AF	60	10	BEH C4	6	6	10	0,601	7	2	1,277
17	MeOH	AF	50	6	BEH C4	10	6	10	0,001	9	1	8,747
18	MeOH	TFA	60	2	BEH C4	2	10	10	3,453	16	11	5,592
19	ACN	TFA	50	6	AWP C8	6	10	10	5,189	15	10	6,171
20	ACN	TFA	70	10	AWP C8	10	10	10	7,083	15	11	6,559
21	MeOH	TFA	70	10	AWP C8	10	10	10	4,459	15	9	20,251
22	ACN	TFA	60	6	BEH C4	2	2	2	2,360	11	4	2,613
23	ACN	AF	50	2	AWP C8	6	2	2	2,013	8	3	3,293
24	MeOH	TFA	50	10	AWP C8	2	6	2	2,211	15	7	10,886
25	ACN	TFA	70	2	BEH C4	10	6	2	7,696	13	9	3,499
26	MeOH	AF	70	6	BEH C4	6	10	2	0,001	8	1	5,585
27	MeOH	AF	60	10	AWP C8	10	10	2	0,985	10	2	17,141
28	MeOH	TFA	70	10	BEH C4	6	2	6	1,799	14	8	5,887
29	ACN	AF	60	10	AWP C8	10	2	6	0,613	7	2	3,523
30	ACN	AF	70	6	AWP C8	2	6	6	0,001	5	1	2,502
31	MeOH	TFA	60	2	AWP C8	6	6	6	3,129	15	9	15,835
32	ACN	AF	50	10	BEH C4	2	10	6	0,164	7	3	1,605
33	ACN	AF	50	2	BEH C4	10	10	6	0,755	7	2	2,398
34	MeOH	TFA	60	6	AWP C8	10	10	6	3,340	13	8	23,550
35	MeOH	AF	70	2	AWP C8	2	2	10	0,316	11	4	6,054
36	MeOH	TFA	50	6	BEH C4	10	2	10	1,005	12	5	8,805
37	ACN	AF	60	10	BEH C4	6	6	10	0,776	7	2	1,685
38	MeOH	AF	50	6	BEH C4	10	6	10	0,001	9	1	8,612
39	MeOH	TFA	60	2	BEH C4	2	10	10	3,542	16	11	5,564
40	ACN	TFA	50	6	AWP C8	6	10	10	5,176	15	10	5,999
41	ACN	TFA	70	10	AWP C8	10	10	10	7,088	15	11	6,527
42	MeOH	TFA	70	10	AWP C8	10	10	10	4,422	15	9	20,316

**Table 21:** *D-Optimal screening experimental design*



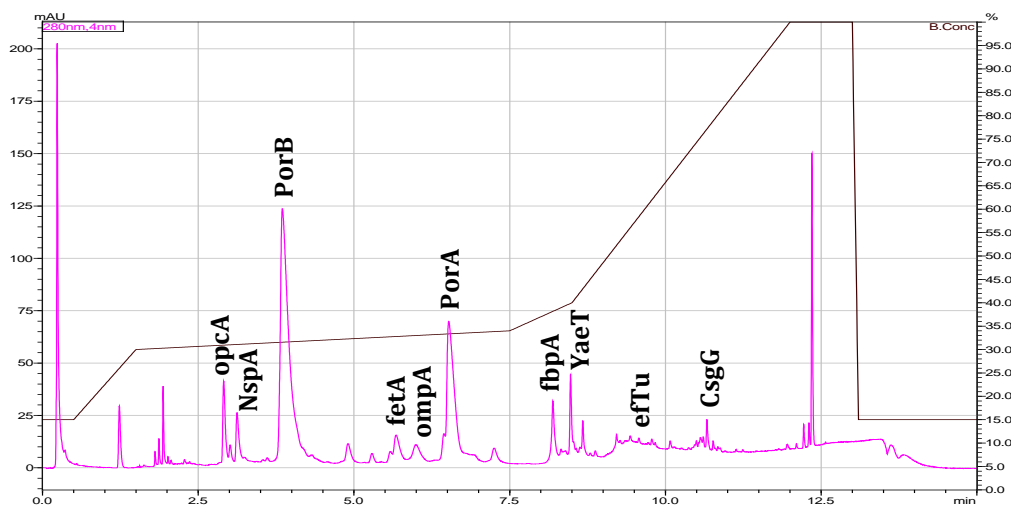
Exp No	RAMP B min.	%A %	%B %	Flow ml/min	ION	N	nB	K'	R <sub>1</sub>
1	10	20	40	0.3	PFPA	11	5	25,000	1,576
2	15	20	40	0.3	TFA	20	13	22,012	2,698
3	10	30	40	0.3	TFA	20	13	10,681	3,125
4	15	30	40	0.3	PFPA	11	5	32,206	1,66
5	10	20	50	0.3	TFA	17	10	14,829	2,08
6	15	20	50	0.3	PFPA	16	10	29,010	4,612
7	10	30	50	0.3	PFPA	14	8	13,281	2,357
8	15	30	50	0.3	TFA	19	12	10,534	2,761
9	10	20	40	0.5	TFA	19	12	28,021	3,194
10	15	20	40	0.5	PFPA	11	5	52,354	2,364
11	10	30	40	0.5	PFPA	11	5	40,449	2,323
12	15	30	40	0.5	TFA	22	15	17,615	6,714
13	10	20	50	0.5	PFPA	16	10	35,528	5,458
14	15	20	50	0.5	TFA	20	13	28,279	3,186
15	10	30	50	0.5	TFA	20	13	15,690	2,996
16	15	30	50	0.5	PFPA	16	10	37,907	1,22
17	10	20	40	0.3	PFPA	11	5	24,620	1,569
18	15	20	40	0.3	TFA	21	14	17,615	3,338
19	10	30	40	0.3	TFA	20	13	10,735	3,337
20	15	30	40	0.3	PFPA	11	5	32,112	1,686
21	10	20	50	0.3	TFA	16	10	15,005	1,868
22	15	20	50	0.3	PFPA	16	10	29,525	6,232
23	10	30	50	0.3	PFPA	14	8	13,207	2,252
24	15	30	50	0.3	TFA	20	13	10,507	2,741
25	10	20	40	0.5	TFA	20	13	28,015	3,138
26	15	20	40	0.5	PFPA	11	5	53,285	2,294
27	10	30	40	0.5	PFPA	11	5	39,907	2,409
28	15	30	40	0.5	TFA	21	14	17,293	7,067
29	10	20	50	0.5	PFPA	16	10	33,755	5,641
30	15	20	50	0.5	TFA	21	14	28,006	3,34
31	10	30	50	0.5	TFA	21	14	15,638	3,211
32	15	30	50	0.5	PFPA	16	10	37,500	1,252

**Table 22:** fractional factorial experimental design



The method screening has been then refined by an OFAT aimed to reduce dead times that lead to the final RP-UHPLC chromatography profiling for OMV protein pattern (Figure 47). With the identified conditions, the ATP was satisfied since:

- a) the proteins quantified in the SDS-PAGE method used to release the OMV bulk are indeed identified by the new RP-HPLC method;
- b) PorB and FbpA proteins are selectively separated and identified;
- c) Further proteins are detectable and have been identified;
- d) Sample preparation requires a single step;
- e) Throughput of the assay has been increase.



**Figure 47:** RP-UHPLC profile of OMV protein pattern

Peaks identity provided by LC-MS analysis

### ***LC-MS characterization of RP-UHPLC OMV protein pattern***

Peaks identity of RP-UHPLC profile for OMV protein pattern (Figure 47) has been attributed with two different LC-MS approaches:

- I) LC-Intact Protein Mass Spectrometry analysis;
- II) LC-Fraction Collection-Peptide Mapping.

### ***LC-Intact Protein Mass Spectrometry analysis:***

Intact OMV proteins, eluted by the RP-UHPLC chromatography, pass through the TriVersa NanoMate system [140], which makes a split of the

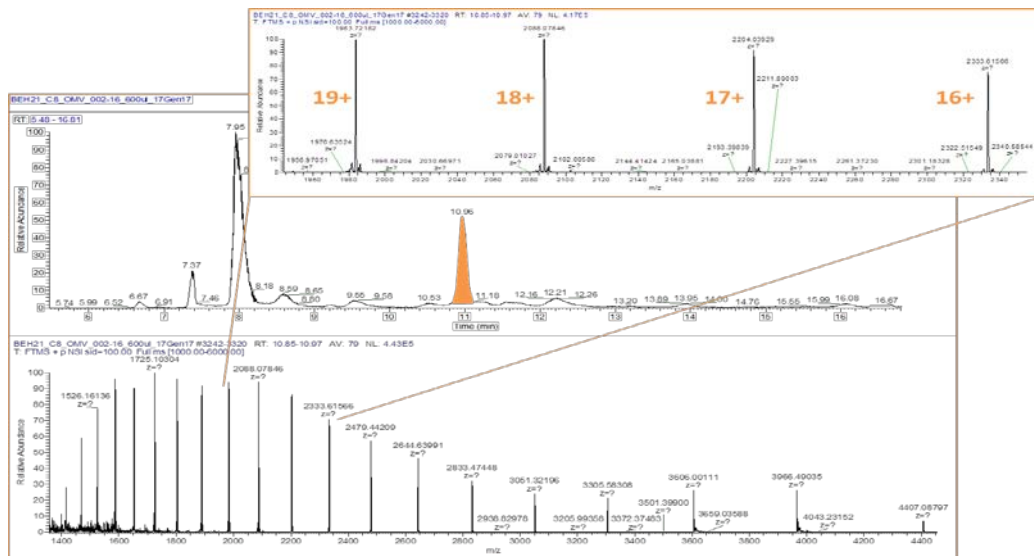
flow, realizing a nano-flow to have a more stable signal in Mass Spectroscopy (MS). This flow goes to an interface chip to which is applied a voltage that ionizes the proteins and realizes the spray. Since the mass spectrometer reveals the mass to charge ratio, mass spectra does not see directly the mass of protonated protein, but multiple charge-states that need to be deconvolved. Mass deconvolution has been done applying the specific algorithms reported in Equation 12. The deconvolution process is based on a system of two equations that allows defining the charge-state of a signal and consequently to calculate its molecular weight.

$$\begin{cases} \left(\frac{m}{z}\right)_2 = \frac{M+z}{z} \\ \left(\frac{m}{z}\right)_1 = \frac{M+z+1}{z+1} \end{cases} \text{ when solved, it brings to: } \begin{cases} z = \frac{m_1-1}{m_2-m_1} \\ M = \frac{(m_1-1)(m_2-1)}{m_2-m_1} \end{cases}$$

**Equation 12:** equations system for intact protein mass deconvolution

Were  $z$  is for the charge determination of a charge-state,  $M$  is the mass determination taking into account 2 adjacent charge-state ( $m_1, m_2$ ).

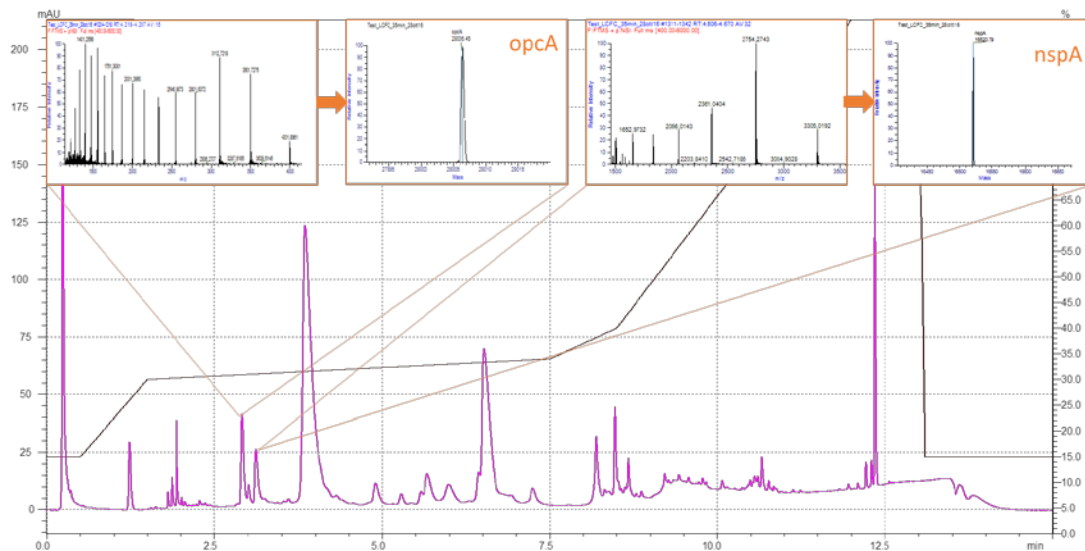
The deconvolution has been done for all the identified and assigned charge-state, coupled by 2, and the final protein molecular weight has been assigned making the average of all results. BioPharma Finder 2.0 software [138] has been used for molecular weight computations.



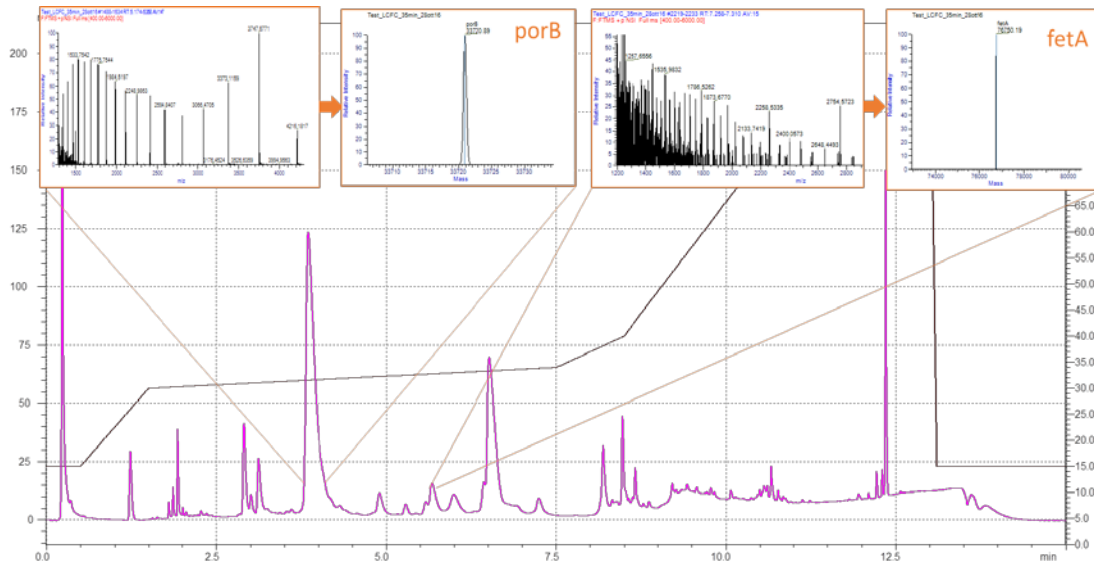
**Figure 48:** Gaussian charge-states distribution of PorA protein

The PorA example in Figure 48 (orange peak) makes possible to observe a typical mass spectrum of an intact protein with a Gaussian distribution of charge-states. Figure 48 allows seeing the  $m/z$  value of adjacent charge-states. Applying the deconvolution algorithm (Equation 12) has been calculated the deconvolved mass from the 17+ and 16+ charge-states, with a resulting value of 39658.80 Da. In order to have a measurement precision of 5-10 ppm, the same computation has been done for all the charge-states identified. By the average of the deconvolved masses, the final value obtained corresponds to the intact mass of PorA outer membrane protein (theoretical MW = 39657 Da).

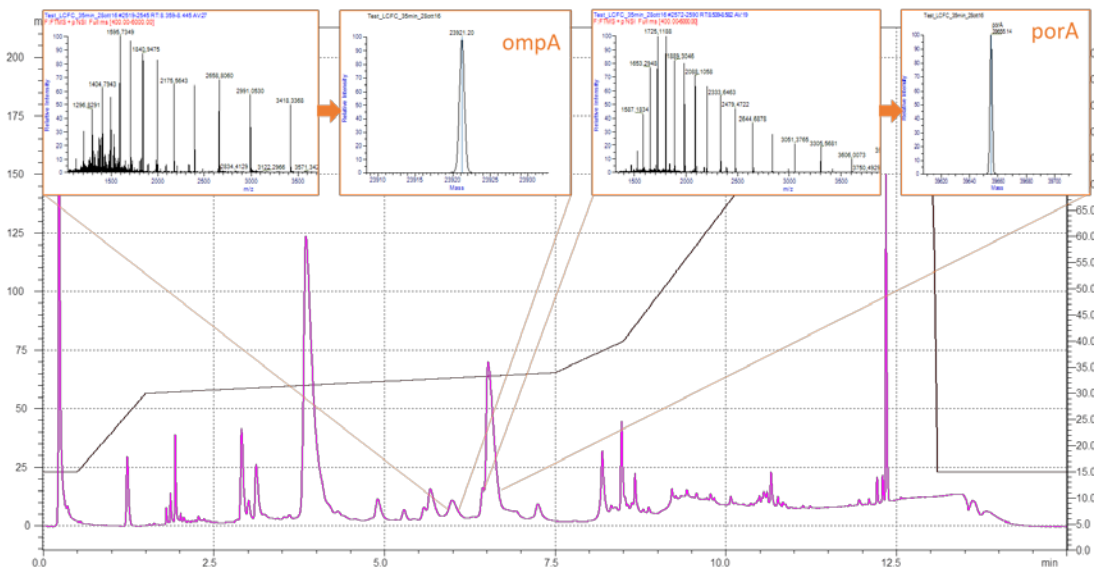
This kind of MS approach requires no sample treatment and is quite immediate. In order to have the correlation between the chromatographic peak and the protein identity, the calculated averaged molecular weight has to be compared with theoretical MW of OMV proteins. The WM computation has been made over all the protein signals found in MS spectra and it has been possible to assign the identity of the most abundant proteins/peaks. The identifications of the eight most abundant OMV proteins are shown from Figure 49 to Figure 52.



**Figure 49:** intact mass analysis of *opcA* and *nspA*

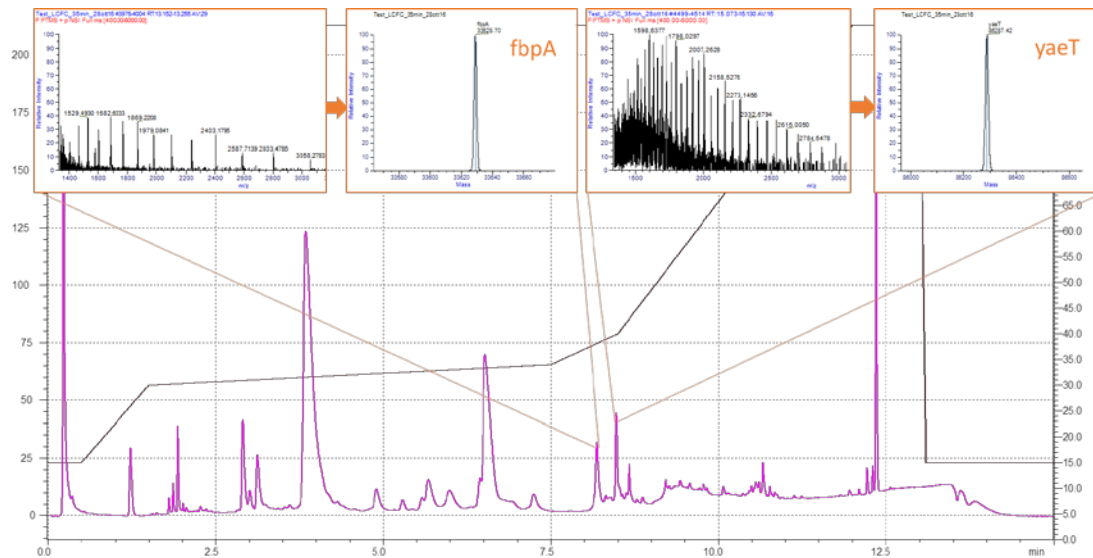


**Figure 50:** intact mass analysis of PorB and fetA



**Figure 51:** intact mass analysis of ompA and PorA





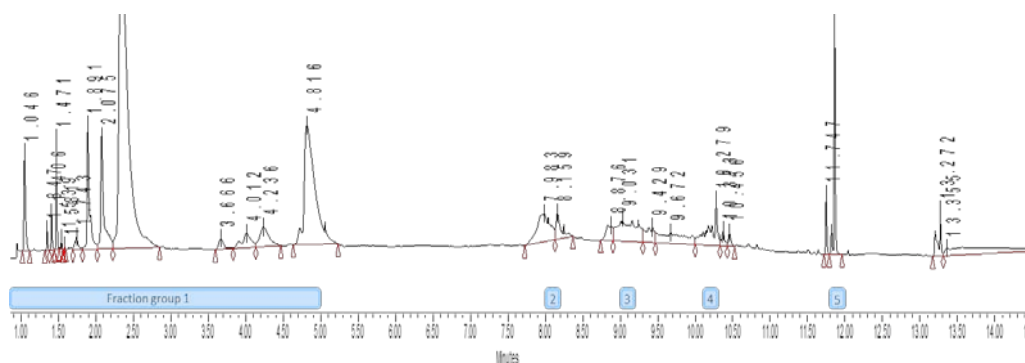
**Figure 52:** intact mass analysis of *fbpA* and *yaeT*

The on-line LC-MS approach is very fast and accurate. The technique allows to drive chromatographic separation optimization directly, without wasting time in sample preparation. Since LC-Intact MS assigns protein identity on the base of correspondence between theoretical and calculated MW, the approach is applicable especially when there are few and well characterized antigens (which relative abundances are known); because of possible wrong assignment due to sample complexity are possible. For this reason, it was decided to complete the characterization study with LC-FC Peptide mapping approach to have an higher assignment confidence of proteins' identity.

### ***LC-Fraction Collection-Peptide Mapping***

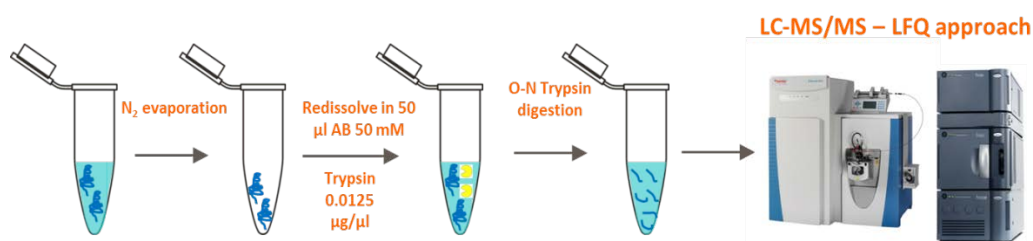
The final confirmation of OMV protein pattern identity was done by fraction collection of the most abundant peaks in the chromatogram and consequently proteins digestion for peptide mapping. Peptide mapping by LC-MS/MS is one of the most valuable methods for verifying the amino acid sequence of protein. LC-FC Peptide mapping approach assigns the proteins' identity on the basis of the correspondence of experimental and theoretical amino acidic sequence found after tryptic digestion of collected fractions [141]. The TriVersa NanoMate system

[140] was employed for the fraction collection of the RP-UHPLC peaks. The fraction collection scheme has been reported in Figure 53.



**Figure 53:** LC-FC scheme

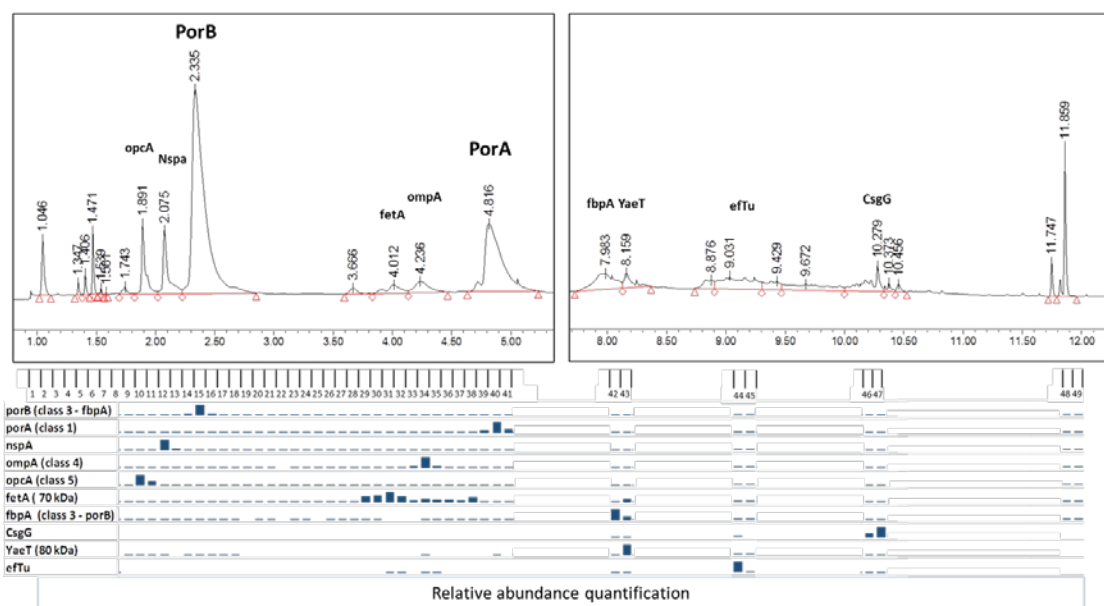
Peptide mapping is a MS technique consisting in two steps: (I) enzymatic digestion of the collected fraction and (II) analysis of the protein digests using LC-MS/MS technique. The experimental design scheme for peptide mapping has been reported in Figure 54.



**Figure 54:** Peptide Mapping experimental design

Each LC-FC fraction has undergone over a N<sub>2</sub> evaporation process to remove the organic phase due to chromatographic elution. Afterwards, protein fractions have been dissolved in 50 mM ammonium bicarbonate buffer, compatible with tryptic digestion, to which 0.0125 µg/µl of trypsin has been added to have a digestion of collected proteins. The samples digestion are left overnight at 37°C. Finally, each sample has been acidified (0.1% HCOOH to stop the enzymatic reaction) and injected to LC-MS system to have the LC peptides separation and the fragmentation MS/MS in a Q-Orbitrap MS configuration. In particular, over the chromatographic separation, the peptide signal is selected, filtered by the quadrupole, fragmented in collision cell and revealed with high-resolution in Orbitrap.

PEAKS Studio 8.0 software [139] has been used for the data analysis of each protein fraction by MS/MS spectra interpretation for the amino acidic sequences. The peptide sequences have been matched with a database of proteins belonging to a specific bacterial strain, previously undergone to an in-silico tryptic digestion. After the collection of all correspondences, it was possible to assign the identity of all identified OMV proteins to their correspondent fraction, as reported in Figure 55 below.



**Figure 55:** Peptide Mapping protein matching



## Chapter 5: Bibliography

---

- [1] S. A. Douglas, W. M. Donald, H. F. James and C. R. Stanley, *Fundamentals of Analytical Chemistry*, Belmont, CA: Brooks/Cole, 2014.
- [2] ICHQ8(R2), “ICH Harmonised Tripartite Guideline. Pharmaceutical Development Q8(R2) (2009) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use”.
- [3] ICHQ9, “ICH Harmonised Tripartite Guideline. Quality risk management Q9 (2008) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use”.
- [4] ICHQ10, “ICH Harmonised Tripartite Guideline. Pharmaceutical quality system Q10 (2008) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use”.
- [5] ICHQ11, “ICH Harmonised Tripartite Guideline. Development and manufacture of drug substances (chemical entities and biotechnological/biological entities) Q11 (2012)”.
- [6] J. Kochling, W. Wu, Y. Hua, Q. Guan and J. Castaneda-Merced, “A platform analytical quality by design (AQbD) approach for multiple UHPLC-UV and UHPLC–MS methods development for protein analysis,” *J. Pharm. Biomed. Anal.*, vol. 125, pp. 130 - 139, 2016.
- [7] S. Orlandini, S. Pinzauti and S. Furlanetto, “Application of quality by design to the development of analytical separation methods,” *Analytical and Bioanalytical Chemistry*, vol. 405, pp. 443-450, 2013.



- [8] J. Kochling, J. Bridgewater and R. Najji, "Introducing a science-based quality by design concept to analytical methods development," in *Pharmaceutical Stability Testing to Support Global Markets, Biotechnology: Pharmaceutical Aspects*, New York, Springer, 2010, pp. 169 - 179.
- [9] C. Hubert, P. Lebrun, S. Houari, E. Ziemons, E. Rozet and P. Hubert, "Improvement of a stability-indicating method by quality-by-design versus quality-by-testing: a case of a learning process," *J. Pharm. Bio. Anal.*, vol. 88, pp. 401 - 409, 2014.
- [10] F. Vogt and A. Kord, "Development of quality-by-design analytical methods," vol. 3, p. 797–812, 2011.
- [11] G. Reid, J. Morgado, K. Barnett, B. Harrington, J. Harwood and D. Fortin, "Analytical quality by design (AQbD) in pharmaceutical development," *Am. Pharm. Rev.*, vol. 16, p. 49–59, 2013.
- [12] K. E. Monks, H. -J. Rieger and I. Molnár, "Expanding the term "Design Space" in high performance liquid chromatography (I)," *J. Pharm. Biomed. Anal.*, vol. 56, pp. 874-879, 2011.
- [13] T. Tol, N. Kadam, N. Raotole, A. Desai and G. Samanta, "A simultaneous determination of related substances by high performance liquid chromatography in a drug product using quality by design approach," *J. Chromatogr. A*, vol. 1432, pp. 26-38, 2016.
- [14] J. Pantović, A. Malenović, A. Vemić, N. Kostić and M. Medenica, "Development of liquid chromatographic method for the analysis of dabigatran etexilate mesilate and its ten impurities supported by quality-by-design methodology," *J. Pharm. Biomed. Anal.*, vol. 111, pp. 7 - 13, 2015.
- [15] Y. Li, G. J. Terfloth and A. S. Kord, "A systematic approach to RP-HPLC method development in a pharmaceutical QbD environment," *Am. Pharm. Rev.*, vol. 12, pp. 87 - 95, 2019.



- [16] A. H. Schmidt and I. Molnár, "Using an innovative Quality-by-Design approach for the development of a stability indicating UHPLC method for ebastine in the API and pharmaceutical formulations," *J. Pharm. Biomed. Anal.*, Vols. 78-79, pp. 65-74, 2013.
- [17] C. Boussès, L. Ferey, E. Vedrines and K. Gaudin, "Using an innovative combination of quality-by-design and green analytical chemistry approaches for the development of a stability indicating UHPLC method in pharmaceutical products," *J. Pharm. Biomed. Anal.*, vol. 115, pp. 114 - 122, 2015.
- [18] J. Terzić, I. Popović, A. Stajić, A. Tumpa and B. Jančić-Stojanović, "Application of Analytical Quality by Design concept for bilastine and degradation impurities determination by hydrophilic interaction liquid chromatographic method," *J. Pharm. Biomed. Anal.*, vol. 125, pp. 385-393, 2016.
- [19] B. Andri, P. Lebrun, A. Dispas, R. Klinkenberg, B. Streel, E. Ziemons, R. Marini and P. Hubert, "Optimization and validation of a fast supercritical fluid chromatography method for the quantitative determination of vitamin D3 and its related impurities," *J. Chromatogr. A*, vol. 1491, pp. 171 - 181, 2017.
- [20] E. Van Tricht, L. Geurink, H. Backus, M. Germano, G. W. Somsen and C. E. Sängers-van de Griend, "One single, fast and robust capillary electrophoresis method for the direct quantification of intact adenovirus particles in upstream and downstream processing samples," *Talanta*, vol. 166, pp. 8 - 14, 2017.
- [21] S. Furlanetto, S. Orlandini, B. Pasquini, C. Caprini, P. Mura and S. Pinzauti, "Fast analysis of glibenclamide and its impurities: quality by design framework in capillary electrophoresis method development," *Anal. Bioanal. Chem.*, vol. 407, pp. 7637-7646., 2015.



- [22] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, S. Pinzauti and S. Furlanetto, "Analytical Quality by Design in pharmaceutical quality assurance: Development of a capillary electrophoresis method for the analysis of zolmitriptan and its impurities," *Electrophoresis*, vol. 36, pp. 2538 - 2545, 2014.
- [23] S. Orlandini, B. Pasquini, R. Gotti, A. Giuffrida, F. Paternostro and S. Furlanetto, "Analytical quality by design in the development of a cyclodextrin-modified capillary electrophoresis method for the assay of metformin and its related substances," *Electrophoresis*, pp. 2538 - 2545, 2014.
- [24] S. Orlandini, B. Pasquini, M. Del Bubba, S. Pinzauti and S. Furlanetto, "Quality by design in the chiral separation strategy for the determination of enantiomeric impurities: Development of a capillary electrophoresis method based on dual cyclodextrin systems for the analysis of levosulpiride," *J. Chromatogr. A*, vol. 1380, pp. 177 - 185, 2015.
- [25] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, M. Douša, S. Pinzauti and S. Furlanetto, "Enantioseparation and impurity determination of ambrisentan using cyclodextrin-modified micellar electrokinetic chromatography: Visualizing the design space within quality by design framework," *J. Chromatogr. A*, vol. 467, pp. 363-371, 2016.
- [26] B. Pasquini, S. Orlandini, C. Caprini, M. Del Bubba, M. Innocenti, G. Brusotti and S. Furlanetto, "Cyclodextrin- and solvent-modified micellar electrokinetic chromatography for the determination of captopril, hydrochlorothiazide and their impurities: A Quality by Design approach," *Talanta*, vol. 160, pp. 332 - 339, 2016.



- [27] S. Orlandini, B. Pasquini, M. Stocchero, S. Pinzauti and S. Furlanetto, "An integrated quality by design and mixture-process variable approach in the development of a capillary electrophoresis method for the analysis of almotriptan and its impurities," *J. Chromatogr. A*, vol. 1339, pp. 200 - 209, 2014.
- [28] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, L. Squarcialupi, V. Colotta and S. Furlanetto, "A comprehensive strategy in the development of a cyclodextrin-modified microemulsion electrokinetic chromatographic method for the assay of diclofenac and its impurities: Mixture-process variable experiments and quality by design," *J. Chromatogr. A*, vol. 1446, pp. 189 - 198, 2016.
- [29] G. Piepel, B. Pasquini, S. Cooley, A. Heredia-Langner, S. Orlandini and S. Furlanetto, "Mixture-process variable approach to optimize a microemulsion electrokinetic chromatography method for the quality control of a nutraceutical based on coenzyme Q10," *Talanta*, vol. 97, pp. 73 - 82, 2012.
- [30] L. S. Ettre, "Nomenclature for chromatography (IUPAC Recommendations 1993)," *Pure & Appl. Chem.*, vol. 56, no. 4, pp. 819 - 872, 1993.
- [31] J. J. Van Deemter, F. J. Zuiderweg and A. Klinkenberg, "Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography," *J. Chem. Eng. Sci.*, vol. 5, p. 272, 1956.
- [32] C. E. Meloan, "Chemical separation principles," in *Techniques and Experiments*, John Wiley & Sons, Canada, 1999.
- [33] S. M. Chesnut and J. J. Salisbury, "The role of UHPLC in pharmaceutical development," *J. Sep. Sci.*, vol. 30, pp. 1183 - 1190, 2007.





- [34] L. Nováková, L. Matysová and P. Solich, “Advantages of application of UPLC in pharmaceutical analysis,” *Talanta*, vol. 68, p. 908–918, 2006.
- [35] “Ultra Performance LCTM by design, 2004. Waters Corporation, USA”.
- [36] L. R. Snyder, J. J. Kirkland and J. L. Glajch, *Practical HPLC Method Development*, New York: John Wiley & Sons, Inc., 1997.
- [37] C. (LC-GC), *Reversed Phase Chromatography*, Crawford scientific, [www.chromacademy.com](http://www.chromacademy.com).
- [38] D. R. Lide, *CRC Handbook of Chemistry and Physics*, Internet Version 2005, Boca Raton, FL: CRC Press, 2005.
- [39] M. -I. Aguilar, “Reversed-Phase High-Performance Liquid Chromatography,” in *HPLC of Peptides and Proteins: Methods and Protocols*, Totowa, NJ, Humana Press Inc, 2004, pp. 9-22.
- [40] I. U. Antia, D. R. Yagnik, L. P. Munoz, A. J. Shah and F. A. Hills, “Heparan sulfate disaccharide measurement from biological samples using pre-column derivatization, UPLC-MS and single ion monitoring,” *Anal. Biochem.*, vol. 530, pp. 17 - 30, 2017.
- [41] C. Tani, M. Stella, D. Donnarumma, M. Biagini, P. Parente, A. Vadi, C. Magagnoli, P. Costantino, F. Rigat and N. Norais, “Quantification by LC–MSE of outer membrane vesicle proteins of the Bexsero vaccine”.
- [42] A. O. Nwokeoji, A. Kung, P. M. Kilby, D. E. Portwood and M. J. Dickman, “Purification and characterisation of dsRNA using ion pair reverse phase chromatography and mass spectrometry,” *J. Chromat. A*, vol. 1484, pp. 14 - 25, 2017.
- [43] S. Fekete, R. Berky, J. Fekete, J. L. Veuthey and D. Guillarme, “Evaluation of recent very efficient wide-pore stationary phases for the reversed-phase separation of proteins,” *J. Chromatogr. A*, vol. 1252, pp. 90-103, 2012.



- [44] H. B. Hewitson, T. E. Wheat and D. M. Diehl, Alternative solvents for the reversed-phase separation of proteins, Waters, [https://www.waters.com/webassets/cms/category/media/primers/primer\\_cc\\_table\\_02.jpg](https://www.waters.com/webassets/cms/category/media/primers/primer_cc_table_02.jpg).
- [45] TCI Chemicals. Ion-Pair Reagents for HPLC, TCI, <http://www.tcichemicals.com/pdf/A1084E.pdf>.
- [46] “The PubMed Project,” [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/>.
- [47] Merck, “sigma-Aldrich on-line catalog,” [Online]. Available: <https://www.sigmaaldrich.com/united-kingdom.html>.
- [48] E. R. Malinowski, Factor Analysis in Chemistry, 3rd Edition, 2002.
- [49] K. R. Beebe, R. J. Pell and M. B. Seasholtz, Chemometrics: A Practical Guide, 1998.
- [50] B. Slutsky, Handbook of Chemometrics and Qualimetrics: Part A, Amsterdam: Elsevier, 1997.
- [51] A. C. Atkinson and A. N. Donev, Optimum Experimental Design, New York: Oxford University Press, 1992.
- [52] G. A. Lewis, D. Mathieu and R. Phan-Tan-Luu, Pharmaceutical Experimental Design, Wilmington, North Carolina: James Swarbrick AAI, Inc., 1999.
- [53] R. Tedeschi, Introduzione alla chemiometria, Edises, 1998.
- [54] J. E. Jackson, A User's Guide to Principal Components, John Wiley & Sons, Inc., 2004.
- [55] I. T. Jolliffe, “A note on the Use of Principal Components in Regression,” *J R Stat Soc*, vol. 31, no. 3, p. 300–303, 1982.
- [56] Y. Dodge, The Oxford Dictionary of Statistical Terms, OUP Oxford; 6 edition, 2003.



- [57] R. Kramer, *Chemometric Techniques for Quantitative Analysis*, CRC Press; 1 edition, 1998.
- [58] B. D. Hibbert, "Experimental design in chromatography: A tutorial review," *J. Chromatogr. B*, vol. 910, p. 2– 13, 2012.
- [59] R. W. Kennard and L. A. Stone, *Technometrics* 11 (1969) 137.
- [60] R. L. Plackett and J. B. Burman, "The Design of Optimal Multifactorial Experiments," *Biometrika*, vol. 33, no. 4, pp. 305-325, 1946.
- [61] S. N. Deming and S. L. Morgan, *Experimental Design: a Chemometric Approach*, Amsterdam: Elsevier, 1993.
- [62] H. R. Cederkvist, A. H. Aastveit and T. Naes, "The importance of functional marginality in model building – A case study," *Chemometrics and Intelligent Laboratory Systems*, vol. 87, pp. 98-106, 2007.
- [63] S. Thomas, R. Dilbarova and R. Rappuoli, "Future challenges for vaccinologists," in *Vaccine Design-Methods and Protocols- Volume 1: Vaccines for Human Diseases, Volume 1403 of the Series Methods in Molecular Biology*, Humana Press, Springer Protocols, 2016, pp. 41-55.
- [64] WHO, "Global vaccine action plan 2011–2020, World Health Organization,"  
[www.WHO.int/immunization/global\\_vaccine\\_action\\_plan](http://www.WHO.int/immunization/global_vaccine_action_plan).
- [65] R. Rappuoli and E. De Gregorio, "Editorial overview: Vaccines: novel technologies for vaccine development," *Curr. Opin. Immunol.*, vol. 41, pp. V-VII, 2016.
- [66] R. Rappuoli, H. I. Miller and S. Falkow, "The intangible value of vaccination," *Science*, vol. 297, p. 937–939, 2002.
- [67] WHO, "The Top 10 Causes of Death, World Health Organization," 2014.



- [68] D. J. Irvine, M. A. Swartz and G. L. Szeto, "Engineering synthetic vaccines using cues from natural immunity," *Nat. Mater.*, vol. 12, p. 978–990, 2013.
- [69] WHO, "World Health Statistics, World Health Organization," 2012.
- [70] J. E. Vela-Ramirez, L. A. Sharpe and N. A. Peppas, "Current state and challenges in developing oral vaccines," *Adv. Drug Deliv. Rev.*, vol. Article in Press, no. <http://dx.doi.org/10.1016/j.addr.2017.04.008>, 2017.
- [71] M. J. Bottomley, R. Rappuoli and O. Finco, "Vaccine design in the 21st century," in *The Vaccine Book (2nd Ed.)*, Academic Press, Elsevier, 2016, pp. 45-65.
- [72] A. J. Pollard, "Global epidemiology of meningococcal disease and vaccine efficacy," *Pediatr. Infect. Dis. J.*, vol. 23, pp. S274-S279, 2004.
- [73] L. H. Harrison, "Prospects for vaccine prevention of meningococcal infection," *Clin. Microbiol. Rev.*, vol. 19, pp. 142-164, 2006.
- [74] L. Jódar, I. M. Feavers, D. Salisbury and D. M. Granoff, "Development of vaccines against meningococcal disease," *The Lancet*, vol. 359, pp. 1499-1508, 2002.
- [75] D. S. Stephens, "Conquering the meningococcus," *FEMS Microbiol. Rev.*, vol. 31, pp. 3-14, 2017.
- [76] A. Gandhi, P. Balmer and L. J. York, "Characteristics of a new meningococcal serogroup B vaccine, bivalent rLP2086 (MenB-FHbp; Trumenba®)," *Postgrad. Med.*, vol. 128, pp. 548-556, 2016.
- [77] S. Bardotti, G. Averani, F. Berti, S. Berti, V. Carinci, S. D'Ascenzi, B. Fabbri, S. Giannini, A. Giannozzi, C. Magagnoli, D. Proietti, F. Norelli, R. Rappuoli, S. Ricci and P. Costantino, "Physicochemical characterisation of glycoconjugate vaccines for prevention of meningococcal diseases," *Vaccine*, vol. 26, pp. 2284-2296, 2008.



- [78] P. S. Watson and D. P. J. Turner, "Clinical experience with the meningococcal B vaccine, Bexsero®: Prospects for reducing the burden of meningococcal serogroup B disease," *Vaccine*, vol. 34, pp. 875-880, 2016.
- [79] T. Vesikari, S. Esposito, R. Prymula, E. Ypma, I. Kohl, D. Toneatto, P. Dull and A. Kimura, "EU meningococcal serogroup B vaccine (4CMenB) administered concomitantly with routine infant and child vaccinations: results of two randomised trials," *Lancet*, vol. 381, pp. 825-835, 2013.
- [80] V. Masignani, R. Rappuoli and M. Pizza, "Reverse vaccinology: a genome-based approach for vaccine development," *Expert Opin Biol Ther.*, vol. 2, no. 8, pp. 895-905, 2002.
- [81] E. Del Tordello, R. Rappuoli and I. Delany, "Chapter 3 – Reverse Vaccinology: Exploiting Genomes for Vaccine Design," in *Human Vaccines: Emerging Technologies in Design and Development*, Kayvon Modjarrad and Wayne C. Koff, 2017, pp. 65-86.
- [82] D. Serruto, M. J. Bottomley, S. Ram, M. M. Giuliani and R. Rappuoli, "The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens," *Vaccine*, vol. 30, pp. B87-B97, 2012.
- [83] M. Pizza, V. Scarlato, V. Masignani, M. Giuliani, B. Aricò, M. Comanducci, G. Jennings, L. Baldi, E. Bartolini, B. Capecchi, C. Galeotti, E. Luzzi, R. Manetti, E. Marchetti, M. Mora, S. Nuti, G. Ratti, L. Santini, S. Savino, M. Scarselli and E. Storni, "Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing," *Science*, vol. 287, pp. 1816-1820, 2000.
- [84] C. Tani, M. Stella, D. Donnarumma, M. Biagini, P. Parente, A. Vadi, C. Magagnoli, P. Costantino, F. Rigat and N. Norais, "Quantification by LC–MSE of outer membrane vesicle proteins of the Bexsero® vaccine," *Vaccine* 32, vol. 32, pp. 1273-1279, 2014.



- [85] M. M. Giuliani, A. Biolchi, D. Serruto, F. Ferlicca, K. Vienken, P. Oster, R. Rappuoli, M. Pizza and J. Donnelly, "Measuring antigen-specific bactericidal responses to a multicomponent vaccine against serogroup B meningococcus," *Vaccine*, vol. 28, pp. 5023-5030, 2010.
- [86] EMA, "European Medicines Agency. Bexsero assessment report," [Online]. Available: [http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002333/human\\_med\\_001614.jsp&](http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002333/human_med_001614.jsp&). [Accessed 3 May 2017].
- [87] FDA, "U.S. Food and Drug Administration. Bexsero product information," [Online]. Available: <http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm431374.htm>. [Accessed 3 May 2017].
- [88] M. V. Pinto, S. Bihari and M. D. Snape, "Immunisation of the immunocompromised," *Child, J. Infect.*, vol. 72, pp. S13-S22, 2016.
- [89] I. Peschieri, "Structural Investigation of Antigens Using Electron Microscopy, [Dissertation thesis]," Alma Mater Studiorum (AMS) Università di Bologna, 2016. [Online]. Available: <http://amsdottorato.unibo.it/id/eprint/7667>. [Accessed 22 April 2016].
- [90] Mascioni, A., B. E. Bentley, R. Camarda, D. A. Dilts, P. Fink and V. Gusarova, "Structural Basis for the Immunogenic Properties of the Meningococcal Vaccine Candidate LP2086," *J Biol Chem.*, vol. 284, no. 13, p. 8738–8746, 2009.
- [91] L. Cendron, D. Veggi, E. Girardi and G. Zanotti, "Structure of the uncomplexed Neisseria meningitidis factor H-binding protein fHbp (rLP2086)," *Acta Crystallogr Sect F Struct Biol Cryst Commun*, vol. 67 (Pt 5), p. 531–535, 2011.
- [92] M. C. Schneider, B. E. Prosser, J. J. Caesar, E. Kugelberg, S. Li and Q. Zhang, "Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates," *Nature*, vol. 458, no. 7240, p. 890–



893, 2009.

- [93] M. Carselli, B. Arico, B. Brunelli, S. Savino, F. Di Marcello and E. Palumbo, "Rational design of a meningococcal antigen inducing broad protective immunity," *Sci Transl Med*, vol. 3, no. 91, p. 91ra62, 2011.
- [94] S. Jacobsson, S. T. Hedberg, P. Molling, M. Unemo, M. Comanducci and R. Rappuoli, "Prevalence and sequence variations of the genes encoding the five antigens included in the novel 5CVMB vaccine covering group B meningococcal disease," *Vaccine*, vol. 27, no. 10, p. 1579–1584, 2009.
- [95] F. Cantini, D. Veggi, S. Dragonetti, S. Savino, M. Scarselli and G. Romagnoli, "Solution structure of the factor H binding protein, a survival factor and protective antigen of *Neisseria meningitidis*," *J Biol Chem*, vol. 284, no. 14, pp. 9022-9026, 2009.
- [96] FDA, "Guidance for Industry. PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance," U.S Food and Drug Administration, 2004. [Online]. Available: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070305.pdf> . [Accessed 2017 May 3].
- [97] J. Haas, A. Franklin, M. Houser, D. Maraldo, M. Mikola, R. Ortiz, E. Sullivan and J. M. Otero, "Implementation of QbD for the development of a vaccine candidate," *Vaccine*, vol. 32, pp. 2927-2930, 2014.
- [98] B. Junker, E. Zablackis, T. Verch, T. Schofield and P. Douett, "Quality-by-Design: As Related to Analytical Concepts, Control and Qualification (Chapter 12)," in *Vaccine Analysis: Strategies, Principles, and Control*, Berlin Heidelberg, Springer-Verlag, 2015, pp. 479-520.
- [99] E. Rozet, P. Lebru, B. Debrus, B. Boulanger and P. Hubert, "Design spaces for analytical methods," *Trends Anal. Chem.*, vol. 42, pp. 157-167, 2013.



- [100] L. Chang, J. T. Blue, J. Schaller, L. Khandke and B. A. Green, "Applicability of QbD for vaccine drug product development," in *Quality by Design for Biopharmaceutical Drug Product Development, Volume 18 of the Series AAPS Advances in the Pharmaceutical Sciences*, Springer, 2015, pp. 443-473.
- [101] S. Orlandini, R. Gotti and S. Furlanetto, "Multivariate optimization of capillary electrophoresis method: a critical review," *J. Pharm. Biomed. Anal.*, vol. 87, pp. 290-307, 2014.
- [102] T. Kourti and B. Davis, "The business benefits of Quality by Design (QbD)," *Pharmaceutical Engineering*, vol. 32, pp. 1-4, 2012.
- [103] K. E. Monks, H. -J. Rieger and I. Molnár, "Expanding the term "Design Space" in high performance liquid chromatography (I)," *J. Pharm. Biomed. Anal.*, vol. 56, pp. 874-879, 2011.
- [104] T. Tol, N. Kadam, N. Raotole, A. Desai and G. Samanta, "A simultaneous determination of related substances by high performance liquid chromatography in a drug product using quality by design approach," *J. Chromatogr. A*, vol. 1432, pp. 26-38, 2016.
- [105] J. Pantović, A. Malenović, A. Vemić, N. Kostić and M. Medenica, "Development of liquid chromatographic method for the analysis of dabigatran etexilate mesilate and its ten impurities supported by quality-by-design methodology, *J. Pharm. Biomed. Anal.* 111 (201)".
- [106] A. H. Schmidt and I. Molnár, "Using an innovative Quality-by-Design approach for the development of a stability indicating UHPLC method for ebastine in the API and pharmaceutical formulations," *J. Pharm. Biomed. Anal.*, Vols. 78-79, pp. 65-74, 2013.
- [107] C. Boussès, L. Ferey, E. Vedrines and K. Gaudin, "Using an innovative combination of quality-by-design and green analytical chemistry approaches for the development of a stability indicating UHPLC method in pharmaceutical products," *J. Pharm. Biomed. Anal.*, vol. 115, pp. 114-122, 2015.





- [108] J. Terzić, I. Popović, A. Stajić, A. Tumpa and B. Jančić-Stojanović, "Application of Analytical Quality by Design concept for bilastine and degradation impurities determination by hydrophilic interaction liquid chromatographic method," *J. Pharm. Biomed. Anal.*, vol. 125, pp. 385-393, 2016.
- [109] B. Andri, P. Lebrun, A. Dispas, R. Klinkenberg, B. Streel, E. Ziemons, R. Marin and P. Hubert, "Optimization and validation of a fast supercritical fluid chromatography method for the quantitative determination of vitamin D3 and its related impurities," *J. Chromatogr. A*, vol. 1491, pp. 171-181, 2017.
- [110] E. Van Tricht, L. Geurink, H. Backus, M. Germano, G. W. Somsen and C. E. Sängers-van de Griend, "One single, fast and robust capillary electrophoresis method for the direct quantification of intact adenovirus particles in upstream and downstream processing samples," *Talanta*, vol. 166, pp. 8-14, 2017.
- [111] S. Furlanetto, S. Orlandini, B. Pasquini, C. Caprini, P. Mura and S. Pinzauti, "Fast analysis of glibenclamide and its impurities: quality by design framework in capillary electrophoresis method development," *Anal. Bioanal. Chem.*, vol. 407, pp. 7637-7646, 2015.
- [112] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, S. Pinzauti and S. Furlanetto, "Analytical Quality by Design in pharmaceutical quality assurance: Development of a capillary electrophoresis method for the analysis of zolmitriptan and its impurities," *Electrophoresis*, vol. 36, pp. 2642-2649, 2015.
- [113] S. Orlandini, B. Pasquini, M. Del Bubba, S. Pinzauti and S. Furlanetto, "Quality by design in the chiral separation strategy for the determination of enantiomeric impurities: Development of a capillary electrophoresis method based on dual cyclodextrin systems for the analysis of levosulpiride," *J. Chromatogr. A*, vol. 1380, pp. 177-185, 2015.



- [114] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, M. Douša, S. Pinzauti and S. Furlanetto, "Enantioseparation and impurity determination of ambrisentan using cyclodextrin-modified micellar electrokinetic chromatography: Visualizing the design space within quality by design framework," *J. Chromatogr. A*, vol. 1467, pp. 363-371, 2016.
- [115] B. Pasquini, S. Orlandini, C. Caprini, M. Del Bubba, M. Innocenti, G. Brusotti and S. Furlanetto, "Cyclodextrin- and solvent-modified micellar electrokinetic chromatography for the determination of captopril, hydrochlorothiazide and their impurities: A Quali," *Talanta*, vol. 160, p. 332–339, 2016.
- [116] B. Pasquini, S. Orlandini, M. Del Bubba, E. Bertol and S. Furlanetto, "The successful binomium of multivariate strategies and electrophoresis for the Quality by Design separation of a class of drugs: the case of triptans," *Electrophoresis*, vol. 36, pp. 2650-2657, 2015.
- [117] S. Orlandini, B. Pasquini, M. Stocchero, S. Pinzauti and S. Furlanetto, "An integrated quality by design and mixture-process variable approach in the development of a capillary electrophoresis method for the analysis of almotriptan and its impurities," *J. Chromatogr. A*, vol. 1339, pp. 200-209, 2014.
- [118] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, L. Squarzialupi, V. Colotta and S. Furlanetto, "A comprehensive strategy in the development of a cyclodextrin-modified microemulsion electrokinetic chromatographic method for the assay of diclofenac and its impurities: Mixture-process variable experiments and quality by design," *J. Chromatogr. A*, vol. 1466, pp. 189-198, 2016.
- [119] USP, "Proposed New USP General Chapter: The Analytical Procedure Lifecycle (1220)," Pharmacopea, U.S., 10 October 2016. [Online]. Available:  
[http://www.usp.org/sites/default/files/usp\\_pdf/EN/USPNF/revisions/s20](http://www.usp.org/sites/default/files/usp_pdf/EN/USPNF/revisions/s20)



1784.pdf.

- [120] E. Rozet, E. Ziemons, R. D. Marini, B. Boulanger and P. Hubert, "Quality by Design Compliant Analytical Method Validation," *Anal. Chem.*, vol. 84, no. 1, p. 106–112, 2012.
- [121] K. Monks, I. Molnár, H. -J. Rieger, B. Bogáti and E. Szabó, "Quality by Design: Multidimensional exploration of the design space in high performance liquid chromatography method development for better robustness before validation," vol. 123, pp. 218-230, 2012.
- [122] USP, "USP <621> Chromatography," Pharmacopea, US, [Online]. Available: <https://hmc.usp.org/sites/default/files/documents/HMC/GCs-Pdfs/c621.pdf>. [Accessed 31 August 2017].
- [123] EP, "European Pharmacopoeia 5.0 (EP) chapter 2.2.41," [Online]. Available:  
[http://library.njucm.edu.cn/yaodian/ep/EP5.0/02\\_methods\\_of\\_analysis/2.2.\\_\\_physical\\_and\\_physicochemical\\_methods/2.2.46.%20Chromatographic%20separation%20techniques.pdf](http://library.njucm.edu.cn/yaodian/ep/EP5.0/02_methods_of_analysis/2.2.__physical_and_physicochemical_methods/2.2.46.%20Chromatographic%20separation%20techniques.pdf). [Accessed 31 August 2017].
- [124] S. Fekete, R. Berky, J. Fekete, J. -L. Veuthey and D. Guillarme, "Evaluation of recent very efficient wide-pore stationary phases for the reversed-phase separation of proteins," *J. Chromatogr. A*, vol. 1252, pp. 90-103, 2012.
- [125] Shimadzu, "LabSolutions Analytical Data System, C191-E016," Shimadzu Corp., 2012. [Online]. Available: <https://www.ssi.shimadzu.com/products/literature/cds/C191-E016.pdf>. [Accessed 3 May 2017].
- [126] "Empower 3, Product code 720001257EN," Waters Corp., [Online]. Available:  
[http://www.waters.com/waters/library.htm?locale=it\\_IT&cid=10190669&lid=1529289..](http://www.waters.com/waters/library.htm?locale=it_IT&cid=10190669&lid=1529289..) [Accessed 3 May 2017].



- [127] D. Mathieu, J. Nony and R. Phan-Tan-Luu, " NEMROD-W," LPRAI sarl, Marseille. [Online].
- [128] Umetrics, "MODDE v.10," MKS Umetrics AB, Sweden.. [Online].
- [129] S. Furlanetto, S. Orlandini, B. Pasquini, M. Del Bubba and S. Pinzauti, "Quality by Design approach in the development of a solvent-modified micellar electrokinetic chromatography method: Finding the design space for the determination of amitriptyline and its impurities," *Anal. Chim. Acta*, vol. 802, pp. 113-124, 2013.
- [130] ICHQ2(R1), "ICH Harmonised Tripartite Guideline. Validation of Analytical Procedures: Text and Methodology Q2(R1) (2005) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use," [Online].
- [131] USP, "Lifecycle Management of Analytical Procedures: Method Development, Procedure Performance, Qualification, and Procedure Performance Verification," Pharmacopea, U.S., [Online]. Available: [https://www.usp.org/sites/default/files/usp\\_pdf/EN/USPNF/revisions/lifecycle\\_pdf.pdf](https://www.usp.org/sites/default/files/usp_pdf/EN/USPNF/revisions/lifecycle_pdf.pdf). [Accessed 3 May 2017].
- [132] M. A. Herrador, A. G. Asuero and A. G. Gonzalez, "Estimation of the uncertainty of indirect measurements from the propagation of distributions by using the Monte-Carlo method: an overview," *Chemom. Intell. Lab. Syst.*, vol. 79, pp. 115-122, 2005.
- [133] "Waters Corp. Application note: Acquity UPLC H-Class and H-Class Bio Amino Acid Analysis System Guide," [Online].
- [134] EP, "European Pharmacopoeia 8.0, Chap. 2.2.56. (Amino Acid Analysis)," 2010. [Online].
- [135] B. J. Smith, Protein Sequencing Protocols, Vol. 211 (2nd edition) in *Methods in Molecular Biology*, Totowa, NJ: Humana Press Inc., 2003.



- [136] Dionex, "Determination of Protein Concentrations Using AAA-Direct (Application Note 163, Thermo Scientific)," [Online]. Available: [https://tools.thermofisher.com/content/sfs/brochures/7442-AN163\\_LPN1634.pdf](https://tools.thermofisher.com/content/sfs/brochures/7442-AN163_LPN1634.pdf).
- [137] M. Fountoulakis and H. -W. Lahm, "Hydrolysis and amino acid composition analysis of proteins," *J Chromat. A*, vol. 27, pp. 109-134, 1998.
- [138] "BioPharma Finder User Guide, Software Version 2.0," Thermo Scientific, [Online]. Available: <https://tools.thermofisher.com/content/sfs/manuals/Man-XCALI-97813-BioPharma-Finder-User-ManXCALI97813-EN.pdf>. [Accessed 04 September 2017].
- [139] "Introduction to PEAKS," BioInformatics Solutions, [Online]. Available: [https://www.msi.umn.edu/sites/default/files/Introduction\\_to\\_PEAKS\\_0.pdf](https://www.msi.umn.edu/sites/default/files/Introduction_to_PEAKS_0.pdf). [Accessed 04 September 2017].
- [140] "Advion, Ltd. TriVersa NanoMate LESA with ESI chip technology," [Online]. Available: [https://advion.com/wp-content/uploads/TVNM\\_FINAL.pdf](https://advion.com/wp-content/uploads/TVNM_FINAL.pdf). [Accessed 02 September 2017].
- [141] C. Perrin, W. Burkitt, X. Perraud, J. O'Hara and C. Jone, "Limited proteolysis and peptide mapping for comparability of biopharmaceuticals: An evaluation of repeatability, intra-assay precision and capability to detect structural change," *J Pharmaceut Biomed*, vol. 123, pp. 162-172, 2016.



## Chapter 6: Conclusions and remarks

---

The present PhD thesis describes the theoretical context and the experimental activities performed to investigate the application of the Analytical Quality by Design (AQbD) approach in developing analytical methods, applicable to release and characterization of vaccines. In particular, the AQbD framework was successfully applied to the development of:

- a fast RP-UHPLC method for the simultaneous determination of the four Bexsero components in the vaccine supernatant.
- a robust and reliable method for testing the concentration of the Bexsero recombinants proteins (rMenB) in monovalent bulks by Amino Acidic Analysis, in order to use rMenB products as standards for the quantification of the components in the final vaccine.
- an assay based on RP-UHPLC technology for characterizing the OVM protein pattern.

The implementation of AQbD has been demonstrated powerful in achieving a deep understanding of such analytical methods. The development approach was performed by application of appropriate risk assessment tools and multivariate experimental designs.

The experimental data were generated through preliminary screening experimentation and then by structured DoE studies for obtaining the Response Surface Models. The final MODR was identified (for the selected CMPs) by implementation of Monte-Carlo simulations to RSM models. The MODR controls the method performances and manages the related risks of failure, with a selected degree of confidence. Identification of a robust MODR represents a decisive added value in ensuring reliable and controlled analytical methods based on scientific knowledge and in potentially providing regulatory flexibility.

All the developed methods can be considered suitable to be formally validated according to ICH and FDA requirements. For the first time (no



*Designing Quality:*  
***Quality by Design in the analytical pharmaceutical development***



example reported in scientific literature), QbD was applied to drive the definition of analytical design space for a vaccine product, making progress in safety for both product quality and control strategy.



## Chapter 7: Conflict of interest

---

Funding and Conflict of Interest Statement: This study was funded by GlaxoSmithKline Biologicals SA. Luca Nompari is an employee of the GSK group of companies.

### Trademark Statement

Bexsero is a trademark owned by or licensed to the GSK group of companies.





## Acknowledgment

---

The authors wish to thank Dr. Amin Khan for the strong sponsorship of QbD application and implementation inside GSK, supporting the PhD program at the basis of this study.

We would also like to thank QbD Senior Advisors Dr. Timothy Schofield and Dr. William Egan, the QbD integration head Dr.ssa Cristiana Campa (that provide a fundamental contribute to allow the AQbD PhD opportunity). All the SMEs of QbD team for the good level of discussion and collaboration, in particular to Dr. Dominique Labbé, Dr. Gaël de Lannoy, Dr. Ghislain Delpierre and Dr. Olivier Vandeputte.

Hence, Dr. Davide Serruto and Dr. Rino Rappuoli, GSK heads, for the possibility to shows Bexsero figures.

Finally, the ARD organization heads Dr. Simona Cianetti, Dr. Francesco Norelli and Dr. Michele Rovini for the time and energy spent in supporting PhD activities.

Special thanks to Prof. Sandra Furlanetto, Prof. Serena Orlandini, Dr.ssa Benedetta Pasquini, Dr.ssa Claudia Caprini and Prof. Sergio Pinzauti of the “Ugo Shiff” chemistry department at the University of Florence. Three year of wonderful collaboration and an high knowledge gathered on AQbD; very good scientists and very good people.

Last but not least all the GSK ARD scientists that have collaborated and supported the thesis studies, sharing experiences, problems and successes during the PhD period. In particular, my special colleagues Dr. Guido Perra and Dr.ssa Letizia Fontana for separation sciences support and Dr. Nicola Messuti and Dr. Alessandro Vadi for LC-MS support.



## Publication Annexes

---

The AQbD framework used for the method development of the analytical techniques discussed in the present thesis work are published in literature and/or under paper submission. The publication articles are annexes in the present section.

### **Annex 1**

Title: Quality by Design approach in the development of an ultra-high performance liquid chromatography method for Bexsero Meningococcal Group B Vaccine

Authors: Luca Nompari; Serena Orlandini; Benedetta Pasquini; Cristiana Campa; Michele Rovini; Massimo Del Bubba; Sandra Furlanetto

Reference: *Talanta* 178 (2018) 552–562 (DOI: 10.1016/j.talanta.2017.09.077)

### **Annex 2**

Title: Design space identification of hydrolysis conditions for amino acid analysis applying Quality by Design approach

Authors: Luca Nompari; Serena Orlandini; Benedetta Pasquini; Letizia Fontana; Guido Perra; Michele Rovini; Sandra Furlanetto

Status: *article under drafting*



# Quality by design approach in the development of an ultra-high-performance liquid chromatography method for Bexsero meningococcal group B vaccine



Luca Nompri<sup>a,b</sup>, Serena Orlandini<sup>b,\*</sup>, Benedetta Pasquini<sup>b</sup>, Cristiana Campa<sup>c</sup>, Michele Rovini<sup>a</sup>, Massimo Del Bubba<sup>d</sup>, Sandra Furlanetto<sup>b,\*</sup>

<sup>a</sup> GSK, Analytical Research and Development (ARD), Via Fiorentina 1, 53100 Siena, Italy

<sup>b</sup> Department of Chemistry "U. Schiff", University of Florence, Via U. Schiff 6, 50019 Sesto Fiorentino, Florence, Italy

<sup>c</sup> GSK, Technical Research & Development (TRD), Via Fiorentina 1, 53100 Siena, Italy

<sup>d</sup> Department of Chemistry "U. Schiff", University of Florence, Via della Lastruccia 3-13, 50019 Sesto Fiorentino, Florence, Italy

## ARTICLE INFO

### Keywords:

Bexsero  
Experimental design  
*Neisseria meningitidis*  
Quality by design  
Ultra-high-performance liquid chromatography  
Vaccine

## ABSTRACT

Bexsero is the first approved vaccine for active immunization of individuals from 2 months of age and older to prevent invasive disease caused by *Neisseria meningitidis* serogroup B. The active components of the vaccine are Neisseria Heparin Binding Antigen, factor H binding protein, Neisseria adhesin A, produced in *Escherichia coli* cells by recombinant DNA technology, and Outer Membrane Vesicles (expressing Porin A and Porin B), produced by fermentation of *Neisseria meningitidis* strain NZ98/254. All the Bexsero active components are adsorbed on aluminum hydroxide and the unadsorbed antigens content is a product critical quality attribute. In this paper the development of a fast, selective and sensitive ultra-high-performance liquid chromatography (UHPLC) method for the determination of the Bexsero antigens in the vaccine supernatant is presented. For the first time in the literature, the Quality by Design (QbD) principles were applied to the development of an analytical method aimed to the quality control of a vaccine product. The UHPLC method was fully developed within the QbD framework, the new paradigm of quality outlined in International Conference on Harmonisation guidelines. Critical method attributes (CMAs) were identified with the capacity factor of Neisseria Heparin Binding Antigen, antigens resolution and peak areas. After a scouting phase, aimed at selecting a suitable and fast UHPLC operative mode for the vaccine antigens separation, risk assessment tools were employed to define the critical method parameters to be considered in the screening phase. Screening designs were applied for investigating at first the effects of vial type and sample concentration, and then the effects of injection volume, column type, organic phase starting concentration, ramp time and temperature. Response Surface Methodology pointed out the presence of several significant interaction effects, and with the support of Monte-Carlo simulations led to map out the design space, at a selected probability level, for the desired CMAs. The selected working conditions gave a complete separation of the antigens in about 5 min. Robustness testing was carried out by a multivariate approach and a control strategy was implemented by defining system suitability tests. The method was qualified for the analysis of the Bexsero vaccine.

## 1. Introduction

Vaccine administration is one of the cheapest health-care interventions that have saved more lives than any other drug or therapy. Due to successful immunization programs, some of the common diseases of the early 20th century almost disappeared. The goal of new approaches in the development of novel vaccines is to rationally design effective vaccines where drug-based conventional approaches have

failed, and innovative strategies can effectively support a successful development [1,2]. The design of an appropriate recombinant antigen has a fundamental role to develop an effective vaccine, and the antigen must be properly formulated with an adjuvant that helps triggering B and T cells responses of right quality and sufficient potency [3]. Moreover, a well-designed analytical control strategy is also important to ensure an appropriate monitoring of the vaccine product quality.

Bacterial meningitis is an infection of the membranes and

\* Corresponding authors.

E-mail addresses: [luca.x.nompri@gsk.com](mailto:luca.x.nompri@gsk.com) (L. Nompri), [serena.orlandini@unifi.it](mailto:serena.orlandini@unifi.it) (S. Orlandini), [benedetta.pasquini@unifi.it](mailto:benedetta.pasquini@unifi.it) (B. Pasquini), [cristiana.x.campa@gsk.com](mailto:cristiana.x.campa@gsk.com) (C. Campa), [michele.x.rovini@gsk.com](mailto:michele.x.rovini@gsk.com) (M. Rovini), [massimo.delbubba@unifi.it](mailto:massimo.delbubba@unifi.it) (M. Del Bubba), [sandra.furlanetto@unifi.it](mailto:sandra.furlanetto@unifi.it) (S. Furlanetto).

<http://dx.doi.org/10.1016/j.talanta.2017.09.077>

Received 10 July 2017; Received in revised form 25 September 2017; Accepted 28 September 2017

Available online 29 September 2017

0039-9140/ © 2017 Elsevier B.V. All rights reserved.

cerebrospinal fluid surrounding the brain and spinal cord and it is a major cause of death and disability worldwide. Three organisms are responsible for most cases of bacterial meningitis: *Neisseria meningitidis*, *Haemophilus influenzae* type b and *Streptococcus pneumoniae* [4–7]. *N. meningitidis* is a pathogen bacterium that is transmitted through contact with respiratory droplets. Transmission and colonization typically results in asymptomatic carriage in the upper respiratory tract, leading to bacteremia that can quickly become life-threatening invasive meningococcal disease, which most often presents meningitis and/or septicemia, and less commonly pneumonia, septic arthritis, otitis media and epiglottitis [8]. *N. meningitidis* is classified into serogroups based on the immunological reactivity of the capsular polysaccharide. Meningococcal serogroups A, B, C, W, Y and recently X account for the majority of meningococcal diseases [9], with serogroup B (MenB) being now the most prominent cause of infant bacterial meningitis and septicemia in Europe, Latin America, US and Canada [10,11]. Serogroup B polysaccharide is immunologically similar to that of neural-cell adhesion molecules and thus is poorly immunogenic, hindering its use in the traditional polysaccharide conjugate-vaccine approach [11].

The application of reverse vaccinology, a new genome based approach [3], to MenB vaccine development allowed the identification of new proteins able to induce bactericidal antibodies. Three highly immunogenic protein antigens, Neisseria Heparin Binding Antigen (NHBA), factor H binding protein (fHbp) and Neisseria adhesin A (NadA), constituting the three core proteins of recombinant meningococcal B vaccine (rMenB), were combined with Outer Membrane Vesicles (OMV) and formulated for human use in a multicomponent vaccine, named 4CMenB (Bexsero, GSK) [12]. NHBA and fHbp have been fused to two additional antigens (Genome-derived Neisseria Antigen: GNA1030 and GNA2091, respectively) to increase their immunogenicity [12–14]. NHBA-GNA1030, fHbp-GNA2091 and NadA have been combined with OMV (expressing Porin A (PorA) and Porin B (PorB), the most abundant outer membrane proteins), mimicking the vesicles naturally released by *Neisseria meningitidis*, and displaying protein antigens in a context similar to their native environment [14]. OMV have been shown to be safe and efficacious in many clinical trials and able to combat MenB outbreaks [14], inducing immunity that is mostly due to the highly variable PorA outer membrane protein, but can also involve outer membrane proteins PorB, OmpC, FetA, and Lipooligosaccharides [15]. All the Bexsero active components are adsorbed on aluminum hydroxide. Bexsero vaccine is the first MenB vaccine based on recombinant proteins able to elicit a robust bactericidal immune response in adults, adolescents and infants against a broad range of isolated serogroup B [12] and has been recently licensed for use in Europe, US and elsewhere [16–18].

The analytical control strategy for vaccine development should cover all the critical quality attributes (CQAs) of the vaccine, including component interactions, and should be able to evaluate factors that could affect the safety, identity, strength, purity, and efficacy of the vaccine. However, as the CQAs may not be known early in development, a risk-based approach should be adopted in developing the control strategy [19]. In 2004, the US Food and Drug Administration outlined a new science- and risk-based approach that encourages manufacturers to develop robust processes and appropriate control strategies, thus supporting continuous improvement and product quality [20]. Over the years this approach has been evolved among the regulatory authorities and the pharmaceutical industry to a core concept called Quality by Design (QbD), recommended and supported by International Conference on Harmonisation (ICH) guidelines [21–23]. So far, in the field of vaccine production QbD principles have only been applied to accelerate process development to manufacture a vaccine candidate at commercial scale [24].

Many of the concepts associated with QbD for the manufacturing process can be mapped to similar concepts in the analytical method development [25–27] and the usefulness and the advantages of the QbD methodology as an adaptive optimization tool have been demonstrated

[28]. Starting from the analytical target profile, that defines the intended purpose of the measurement, analytical QbD (AQbD) emphasizes the need to thoroughly understand the analytical system by an in-depth study of critical method parameters (CMPs) based on risk assessment and multivariate tools. The design space (DS) is determined as the multidimensional region of successful operating ranges for the CMPs, which lead to desired values for critical method attributes (CMAs) [29]. Recent examples of AQbD mostly refer to the pharmaceutical field, and concern mainly the development of separation methods as HPLC [30–32], ultra-high-performance liquid chromatography (UHPLC) [33,34], hydrophilic interaction liquid chromatography [35], supercritical fluid chromatography [36,37], capillary zone electrophoresis [38–42], micellar electrokinetic chromatography [43,44] and microemulsion electrokinetic chromatography [45–47].

The AQbD approach intends to build quality of the analytical performances during method development rather than testing it at the end of the process, by the use of risk assessment tools and of design of experiments (DoE) [48,49]. The result is consistent with controlled method performances within predefined boundaries that ensure predetermined quality expectations are met. AQbD systematically investigates and controls CMPs leading to an increased knowledge of their effects on the CMAs and to a reduced variability by controlling assay conditions and risks. This systematic approach helps in achieving these objectives by identifying, mitigating and controlling method risks, as well as by a thorough examination of the method DS with respect to required method performances to meet quality and business targets [27]. Additional advantages in prospective method life cycle and regulatory flexibility are also provided [50].

A fundamental CQA of Bexsero vaccine is the unadsorbed antigens content. rMenB and OMV components are characterized at the drug substance level by the combination of multiple techniques [14,51,52] and the actual state of art for unadsorbed antigens determination in the commercial vaccine product is a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) test.

In this study, for the first time in the literature, the QbD approach was applied to design an analytical method aimed to control the quality of a vaccine product. AQbD principles were applied to the development of a new UHPLC method for the control of the mentioned product CQA. UHPLC technique was selected due to its characteristics of efficiency, selectivity and rapidity of analysis. By applying the AQbD approach to the development of the UHPLC method the analytical target profile requirements were reached, the assay throughput was increased by reducing the manual operations, and selectivity and sensitivity with respect to the current SDS-PAGE method were improved. In this way it was possible to build an adequate method knowledge to ensure regulatory flexibility during lifecycle, as well as method robustness and safety for product quality control. For sake of clarity, a list of abbreviations has been added in the Supplementary Information.

## 2. Materials and methods

### 2.1. Chemicals and reagents

For the development of the chromatographic method a mock standard solution was used. The mock solution was a mixture with the same composition of antigens and excipients of the vaccine product, apart from the aluminum hydroxide adsorbent which was removed for analytical needs and apart from the addition of 0.15% (p/v) Zwittergent 3–14 detergent. The mock solution was used as calibration for the analysis and its composition enabled the target concentration range to be achieved. The bulk drug substances used for mock formulation were produced by GSK (Siena, Italy). L-histidine, trifluoroacetic acid (TFA)  $\geq 99\%$ , methanol (HPLC grade), Tween 80 and sucrose were purchased by Sigma-Aldrich (Saint Louis, MO, USA). NaCl,  $\text{KH}_2\text{PO}_4$ , HCl, KOH 45% and Zwittergent 3–14 detergent were purchased by Merck KGaA (Darmstadt, Germany). Acetonitrile (ACN) was purchased by Panreac

(Radnor, PA, USA). Ultrapure water was produced by Millipore Milli-Q system (Billerica, MA, USA) and filtered on a nylon membrane of 0.22 µm porosity using Nalgene clessidra filters (Nalgene, Rochester, NY, USA).

## 2.2. Solutions and sample preparation

The L-Histidine buffers and the phosphate buffers were adjusted to the proper pH by adding HCl and KOH, respectively. A 5% Zwittergent 3–14 detergent solution was prepared in pH 6.5 phosphate buffer 1 M.

The mock standard solutions were prepared each day by dilution of the proper volumes of the drug substance bulks (each aliquot of rMenB stored at –20 °C and of OMV stored at 4–8 °C) in 200 mM L-Histidine pH 6.3 buffer plus 90 mg ml<sup>-1</sup> NaCl and 5% (p/v) sucrose solutions up to 100 µg ml<sup>-1</sup> rMenB and 50 µg ml<sup>-1</sup> OMV. The working solutions were obtained by diluting the samples to the final concentration of 10 µg ml<sup>-1</sup> rMenB and 5 µg ml<sup>-1</sup> OMV, maintaining the same matrix composition, with 0.15% (p/v) Zwittergent 3–14 detergent in pH 6.5 phosphate buffer 1 M, added for the analytical purpose of applying the UHPLC procedure also to the antigen content determination. The working solutions were stored at 4–8 °C before the analysis. The samples were stored in the autosampler using three different vials type: Clear Glass Total Recovery (Waters Corp., Milford, MA, USA), LCMS Certified Total Recovery vials (Waters Corp.) and Polypropylene Total Recovery vials (Thermo Fisher Scientific, Waltham, MA, USA).

For mobile phase preparation, 500 µl of TFA ≥ 99% were diluted up to 500 ml ultrapure water to obtain a 0.1% (v/v) TFA aqueous solution and 500 µl of TFA ≥ 99% were added to 450 ml of ACN plus 49.5 ml of ultrapure water to prepare a 0.1% (v/v) TFA and 90% (v/v) ACN organic phase. All buffers and solutions were filtered by a nylon membrane of 0.22 µm porosity using Nalgene filters (Nalgene).

## 2.3. Chromatographic equipment and analysis

Different chromatographic columns were tested: Acquity RP-C4 BEH 300 Å, 1.7 µm, 2.1 × 150 mm (C4pore) and Acquity UPLC BEH 130 Å C8, 1.7 µm, 2.1 × 150 mm (C8pore) from Waters Corp. and Aeris WIDEPOR C4 200 Å, 3.6 µm, 4.6 × 150 mm (C4shell) from Phenomenex (Torrance, CA, USA) [53].

For the screening phase the Nexera X2 method scouting UHPLC series 30 system was used, equipped with LC-30AD pump, DGU-20A5R degasser unit and LPGE-unit, SIL-30AD autosampler, CTO-20AC oven with 180 µl mixer and FCV-34AH UHPLC switching valve, SPD-M30A PDA detector with high sensitive flow-cell (85 mm; 9 µl) from Shimadzu Corp. (Kyoto, Japan). For Response Surface Methodology (RSM) the Acquity H-ClassBio UPLC system (Waters Corp.) was used, equipped with bio-Quaternary Solvent Manager (bioQSM) with 100 µl mixer, bio-Sample Manager (bioSM-FTN) with 15 µl injector needle and 50 µl extension loop, column oven CH-A with pre-heater and photodiode array detector (ACQ-PDA) with analytical flow-cell (10 mm; 500 nL). Both the systems used low pressure mixing (quaternary pumps), and no difference was observed in the chromatographic profile and selectivity. Only a few seconds shift in retention times was observed, corresponding to the difference of the void volumes when applying the same linear velocity of the mobile phase.

The detection wavelength was 210 nm. Sample injections were done by 30 µl of the working solution stored in the autosampler at 4–8 °C. After each injection the column was washed with 90% of ACN for one minute and equilibrated for 3 min in the starting conditions. A new column was conditioned with the mobile phase for 60 min before starting the analysis. After the analysis the column was stored in pure ACN filtered on 0.22 µm nylon membrane.

The optimal separation of the antigens was achieved using the C4pore column. The working conditions (with the interval corresponding to the DS) were as follows: starting organic phase concentration, 33.0% (32.0–34.6%) (%v/v); ramp time, 4.0 min

(4.0–5.6 min) to 75% (v/v) organic phase final concentration; column temperature, 60 °C (60–68 °C).

## 2.4. Calculations and softwares

The chromatographic resolutions (*R*) between two adjacent peaks were calculated using the retention times (*t<sub>R</sub>*) in min and the peak widths at half height (*w*) in min, according to formula (1)

$$1.18x(t_{R2} - t_{R1})/(w_1 + w_2) \quad (1)$$

The capacity factor (*K'*) of NHBA-GNA1030 antigen, measurement of the retention time relative to column void volume (*V<sub>o</sub>*), was calculated according to formula (2)

$$K' = (t_R - V_o)/V_o \quad (2)$$

LabSolution Version 5 software equipped with Method Scouting startup kit and licensed by Shimadzu Corp [54]. was used for the Nexera X2 UHPLC instrument control and for the chromatographic data computation in the screening phase. Empower3 software [55] licensed by WatersCorp. was used for the Acquity H-ClassBio UPLC instrument control and for the chromatographic data computation in RSM.

Nemrod-W software [56] was used to generate the two asymmetric screening matrices used for investigating the knowledge space, and the two Full Factorial Designs used for selecting the verification points at the edges of the design space and for testing robustness. MODDE software [57] was purchased from S-IN (Vicenza, Italy) and was employed to generate the Central Composite Design (CCD) used for RSM, to perform data analysis and to find the design space by means of risk failure maps calculated using the Monte-Carlo simulations. Multiple Linear Regression (MLR) was used for fitting the models to the data, hence each CMA was modeled independently. For estimating the probability map, and thus calculating the DS, the possible factor ranges are expanded from a set-point (optimum) to the largest possible range where all response predictions still fulfill the requirements. The distributions of predictions simulate a real situation with a random combination of factor setting disturbances within a given range. The distribution for each factor is expanded symmetrically around the set-point until one or more response limits are exceeded according to the specified defects per million opportunities (DPMO) target, set as 100,000 in this study (10% risk of failure). Model error was also included in the predictions of the response distributions. The runs of all the DoE plans were carried out in a randomized order.

## 3. Results and discussion

The method development followed the systematic approach of AQBd workflow for separation methods [25], involving the following steps: (i) Analytical target profile definition, method scouting and definition of the CMAs; (ii) quality risk assessment and identification of potential CMPs; (iii) investigation of knowledge space by screening DoE; (iv) RSM and definition of DS; (v) working point definition and robustness testing; (vi) method control.

### 3.1. Analytical target profile, method scouting and critical method attributes

The analytical target profile is the intended purpose of the method and is defined by the selection of the analytes and of the analytical performances to be achieved [25,58]. In this study, it consisted in obtaining the accurate quantitation of the five proteins NHBA-GNA1030, fHbp-GNA2091, NadA, PorA and PorB. Moreover, general validation requirements according to ICH Q2(R1) guideline [58,59] had to be fulfilled, including an adequate selectivity and sensitivity, which corresponded in obtaining baseline resolution of the peaks and adequate peak areas, to be able to monitor the unadsorbed antigens content identified as a product CQA.

In order to approach this target, different preliminary aspects had to



be considered. First, scouting of different UHPLC operative modes for obtaining the separation of the five antigens in the vaccine was performed. Prior knowledge and experimental studies led to choose reverse phase UHPLC (RP-UHPLC) as analytical technique for method development. Different RP-UHPLC operative modes were tuned and tested by the Nexera X2 method scouting system, operating at 60 °C and changing type of organic phase (mixtures of ACN/TFA or methanol/TFA), organic ramp (%/min), starting concentration (from 30% to 40%) and ending concentration (from 75% to 100%) of the organic phase. The aim was to approach to the experimental conditions leading to good selectivity and fast analysis. The best results were achieved using an organic mixture of 90/0.1/9.9 ACN:TFA:H<sub>2</sub>O (v/v) as organic phase and a 4 min gradient from 34% to 75% of organic phase; these conditions constituted the starting point for further in-depth optimization by DoE. In these conditions, the retention order of the peaks was the following: NHBA-GNA1030, PorB, PorA, fHbp-GNA2091, NadA.

The second aspect was the definition of the surfactant needed for the analysis, which was based on prior knowledge of antigens desorbing from aluminum hydroxide and on preliminary experimental runs. Different surfactants were tested, i.e. Tween 80 and Zwittergent 3–14 detergents in phosphate buffers. The selected surfactant was 0.15% (p/v) Zwittergent 3–14, since it made it possible to maintain consistent over time the chromatographic area of the proteins in mock solution without aluminum hydroxide.

The selected CMAs are reported for clarity in [Supplementary Table S1](#) and were the resolution values between the peak pairs, named as  $R_1$  (NHBA-GNA1030/PorB),  $R_2$  (PorB/PorA),  $R_3$  (PorA/fHbp-GNA2091),  $R_4$  (fHbp-GNA2091/NadA), the peak areas  $A_1$  (NHBA-GNA1030),  $A_B$  (PorB),  $A_A$  (PorA),  $A_2$  (fHbp-GNA2091),  $A_3$  (NadA), and NHBA-GNA1030 capacity factor  $K'$ , considered in the RSM for controlling the elution of the first antigen peak with respect to column void volume.

### 3.2. Risk assessment and critical method parameters

The objective of a risk assessment is to develop understanding of procedure variables and their impact on the method reportable values for the identification of hazards and the analysis and evaluation of risks associated with exposure to those hazards [60,61]. Tools such as process maps and fishbone diagrams may be used, in addition to prior knowledge, to provide structure to a brainstorming and information-gathering exercise to identify CMPs [60]. In this study, a fishbone diagram ([Fig. 1](#)) was used to formalize the risk assessment and point out the risk factors associated with the characteristics of the RP-UHPLC analysis and thus to highlight the potential CMPs which were supposed to potentially affect the selected CMAs. Some of the CMPs, including detector type and settings, sample surfactant, ion pair type and auto-sampler temperature, had been already studied and fixed by preliminary experiments and scouting. Other CMPs, underscored and with grey background in [Fig. 1](#), needed to be risk managed and in-depth studied by DoE to enhance knowledge on their effects on method performances.

### 3.3. Screening experimental designs

AQbD emphasizes the need to thoroughly understand the analytical system by an in-depth study of CMPs and their interactions using DoE to find cause-and-effect relationships [25,60,61]. As a result of RP-UHPLC scouting and risk assessment, the selected CMPs to be investigated by DoE were represented by vial type (VIAL), sample concentration (CONC), injection volume (INJ), column type (COL), starting organic phase concentration (ACN%), ramp time (RAMP), column temperature (T). For sake of clarity, ACN% refers to the percentage of the organic phase (made by 0.1% (v/v) TFA and 90% (v/v) ACN) in the whole mobile phase. The other part of the mobile phase is the aqueous phase (made by 0.1% (v/v) TFA aqueous solution), thus the concentration of TFA was kept constant (0.1% v/v).

The first screening study involved the investigation of the effects of VIAL and CONC for optimizing sample preparation, with the aim of reaching adequate sensitivity before starting the optimization of the RP-UHPLC conditions. The considered CMAs were the peak areas of the five antigens  $A_1$ ,  $A_B$ ,  $A_A$ ,  $A_2$ ,  $A_3$ . The vial type was studied at 3 levels (Clear Glass Total Recovery vial, Polypropylene Total Recovery vial and LCMS Certified Total Recovery vial). Sample concentration was examined at 4 levels (1–4–10–20  $\mu\text{g ml}^{-1}$ ) in order to evaluate the possibility of aggregation and/or aspecific absorption of antigens on vials walls. The injection volume was adapted to inject, at each different sample concentration, the same amount of sample in column. Nemrod-W software [56] was employed to generate the asymmetric screening matrix used to estimate the coefficients of the Free-Wilson model with interactions (3) [49]:

$$Y = b_0 + b_{1A}(X_{1A}) + b_{1B}(X_{1B}) + b_{2A}(X_{2A}) + b_{2B}(X_{2B}) + b_{2C}(X_{2C}) + b_{1A2A}(X_{1A}X_{2A}) + b_{1A2B}(X_{1A}X_{2B}) + b_{1A2C}(X_{1A}X_{2C}) + b_{1B2A}(X_{1B}X_{2A}) + b_{1B2B}(X_{1B}X_{2B}) + b_{1B2C}(X_{1B}X_{2C}) + e \quad (3)$$

where  $Y$  is the experimental response,  $X_1$  is VIAL and  $X_2$  is CONC,  $b_0$  is the constant term,  $b_i$  are the linear and interaction coefficients and  $e$  is the experimental error. The model contains one constant term plus, for each factor, a number of terms equal to its number of levels minus one. The 12-run experimental plan ( $2^{1 \times 3} / 12$ ) is reported in [Supplementary Table S2](#); each analysis was duplicated in order to obtain a reliable estimate of the experimental variance. The graphical plots describing the effects of changing the levels of the factors on antigens areas are shown in [Supplementary Fig. S1](#), where A stands for the type of vial ( $A_1$ , Clear Glass Total Recovery vial;  $A_2$ , Polypropylene Total Recovery vial;  $A_3$ , LCMS Certified Total Recovery vial) and B stands for level of sample concentration ( $B_1$ , 1  $\mu\text{g ml}^{-1}$ ;  $B_2$ , 4  $\mu\text{g ml}^{-1}$ ;  $B_3$ , 10  $\mu\text{g ml}^{-1}$ ;  $B_4$ , 20  $\mu\text{g ml}^{-1}$ ). The type of vial had a significant influence only on area responses of PorA and fHbp-GNA2091. The effects were opposite, as the use of Clear Glass Total Recovery vial slightly enhanced PorA area but reduced fHbp-GNA2091 area. The use of Polypropylene Total Recovery vial had a negative effect on PorA area and an important positive effect on fHbp-GNA2091 area. Glass Total Recovery Silicone Coated vial caused the decrease of fHbp-GNA2091. As for the effects of CONC, in general a decrease of the areas was observed by using the lower concentration value. The maximization of the areas was obtained by using 4  $\mu\text{g ml}^{-1}$  for PorA and NadA, 10  $\mu\text{g ml}^{-1}$  for NHBA-GNA1030 and PorB, and 20  $\mu\text{g ml}^{-1}$  for fHbp-GNA2091. Some interactions were also noticed, evidencing in particular an important positive interaction between 1  $\mu\text{g ml}^{-1}$  concentration value and polypropylene vial on  $A_2$  ( $A_2$ - $B_1$ ) and other positive interactions involving  $A_B$  ( $A_1$ - $B_3$ ) and  $A_2$  ( $A_1$ - $B_2$ ). Hence, the vial type was selected as Polypropylene Total Recovery vial and sample concentration was fixed at 1  $\mu\text{g ml}^{-1}$ , taking into account the presence of the positive interaction  $A_2$ - $B_1$  for fHbp-GNA2091, the importance of quantitation of fHbp-GNA2091 antigen (a rMenB component) with respect to the OMV PorA and PorB antigens, and the possibility to reach a lower range concentration required by analytical target profile in terms of sensitivity gain.

Once the sample conditions were selected, the AQbD framework continued with the screening study of the chromatographic parameters INJ, COL, ACN%, RAMP and T. The Free-Wilson model (4) was postulated [49]:

$$Y = b_0 + b_{1A}(X_{1A}) + b_{2A}(X_{2A}) + b_{2B}(X_{2B}) + b_{3A}(X_{3A}) + b_{3B}(X_{3B}) + b_{4A}(X_{4A}) + b_{4B}(X_{4B}) + b_{5A}(X_{5A}) + b_{5B}(X_{5B}) + e \quad (4)$$

where  $Y$  is the experimental response,  $X_1$  is INJ,  $X_2$  is COL,  $X_3$  is ACN%,  $X_4$  is RAMP and  $X_5$  is T,  $b_0$  is the constant term,  $b_i$  the linear coefficients of each factor and  $e$  is the experimental error. INJ was studied at two levels, while the other four factors were studied at three levels, as reported in [Supplementary Table S3](#). A new asymmetric screening matrix was designed for obtaining preliminary information throughout the

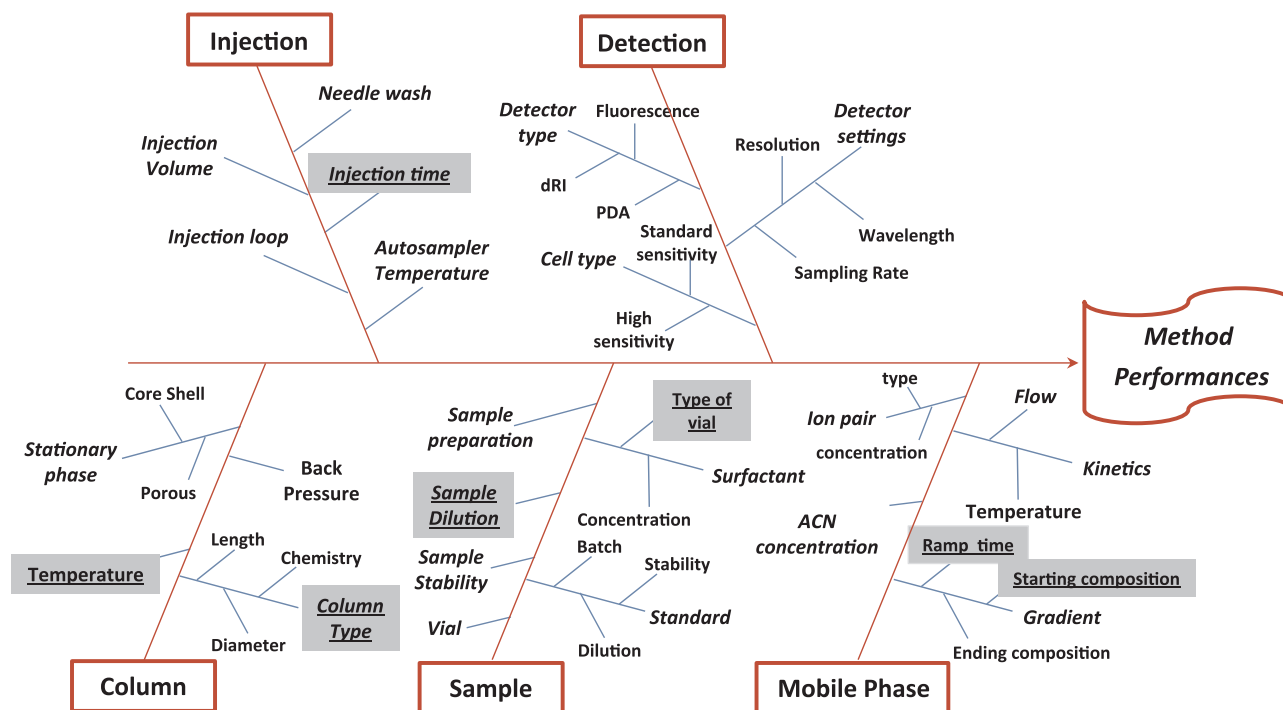


Fig. 1. Fishbone diagram for RP-UHPLC method risk assessment. The factors considered as CMPs and evaluated by DoE are underscored with grey background.

knowledge space on the effects of the five selected factors on chromatographic CMAs, i.e. resolutions  $R_1$ ,  $R_2$ ,  $R_3$  and peak areas  $A_B$  and  $A_A$ . In this step  $R_4$  was excluded from the study, as it presented values above 2 in all the runs, while the other resolutions were critical. Only the OMV areas were taken into account because PorB and PorA present the lowest areas among the five antigens peaks, also due to the lower concentration in the sample with respect to rMenB components.

The screening asymmetric matrix used to estimate the model coefficients was made by 16 runs ( $2^{13}4^{1/16}$ ), with three replicates for each experiment in order to obtain a reliable estimate of the experimental variance, and is reported in Supplementary Table S4 with the measured responses. The graphic analysis of effects made it possible to obtain two types of plots; the first type of plot, reported in Fig. 2, shows the effects of the different levels of the factors on the responses, with the bar length proportional to the effects (a longer bar corresponds to a maximization of the response). The second type of plot, shown in Supplementary Fig. S2, shows the difference of the effects between the two considered levels, and the bars colored in orange, exceeding the reference line, correspond to the pair of factors for which a change of level is significant on the response. By examining these plots, it was possible to select the optimal value for some CMPs and to decide which others should be further studied more deeply by RSM. As for injection volume, its change exerted a limited or no effect on resolution values, but the 50  $\mu$ l value was definitely better for increasing both the areas, as expected. Hence, this factor was fixed at an intermediate value of 30  $\mu$ l, in order to find a compromise between selectivity and sensitivity. As regards the column, C4pore was selected for further studies, since it gave the maximization of  $R_3$  value and led to good results also for the other responses, taking into account that in general this column is preferred for protein studies and for avoiding absorption. The highest level for ACN% led to maximize all the resolution values, while low-medium values were preferred for increasing area responses. Hence, the new domain to be investigated was moved towards the values 28.0–38.0%. When considering RAMP, the value of 6 min led to the maximization of all the responses apart from  $A_B$ , and it was decided to further study this CMP in the range 3–6 min, in the perspective of keeping low analysis time. As concerns T, in general it presented a lower influence on the CMAs with respect to the other CMPs, but it

showed conflicting effects on the different responses, thus it was decided to continue to study this factor by RSM in the domain 50–70  $^{\circ}$ C.

### 3.4. Response surface methodology and design space

RSM [49] was applied for in-depth investigations of the effects of ACN%, RAMP and T on the CMAs in the new experimental domain reported in Supplementary Table S3. A three-factor CCD, fractionating the experimental domain for each factor into 5 levels, was employed for building a quadratic model with interactions making it possible to study all the selected chromatographic CMAs, namely resolutions ( $R_1$ ,  $R_2$ ,  $R_3$ ,  $R_4$ ) and peak areas ( $A_1$ ,  $A_A$ ,  $A_B$ ,  $A_2$ ,  $A_3$ ) as above described. Additionally, NHBA-GNA1030 capacity factor  $K'$  was also considered among the CMAs in order to control the elution of the first antigen peak with respect to column void volume. As a matter of facts, from screening DoE it was observed that high organic phase starting concentration values anticipate the chromatographic pattern of the antigens. The 15-run CCD experimental plan with the measured responses is reported in Supplementary Table S5, where each condition was twice replicated, including a central point. The responses  $R_1$  and  $A_B$  were reverse ( $Y^{-1}$ ) and logarithmic ( $\log_{10}Y$ ) transformed, respectively, to improve the model goodness of prediction  $Q^2$ , while the others CMP models were obtained without mathematical transformation. All the models, calculated by MLR, were refined by deleting some of the interaction and/or quadratic terms, which were not significant, in order to increase the values of  $R^2$  and  $Q^2$ . The resulting performances indicators of the refined models are reported in Supplementary Table 6. All the models were significant in terms of ANOVA, and were valid apart from  $K'$ ,  $R_1$ ,  $R_2$  and  $A_1$ , for which a lack of fit was evidenced ( $p$  values < 0.025). In the case of  $K'$  model, the lack of validity could be explained by the extremely high value of reproducibility observed (equal to 0.9998). The parameter of reproducibility indicates the variation of the response under the same conditions (pure error), compared to the total variation of the response, and such a high value for  $K'$  makes it difficult to obtain a valid model due to the extremely low experimental variance. Anyway, all the other performances indicators were very good for  $K'$  model and also for  $R_1$ ,  $R_2$  and  $A_1$  models: the values of  $R^2$  and  $Q^2$  ranged from 0.844 to 0.950 and from 0.652 to

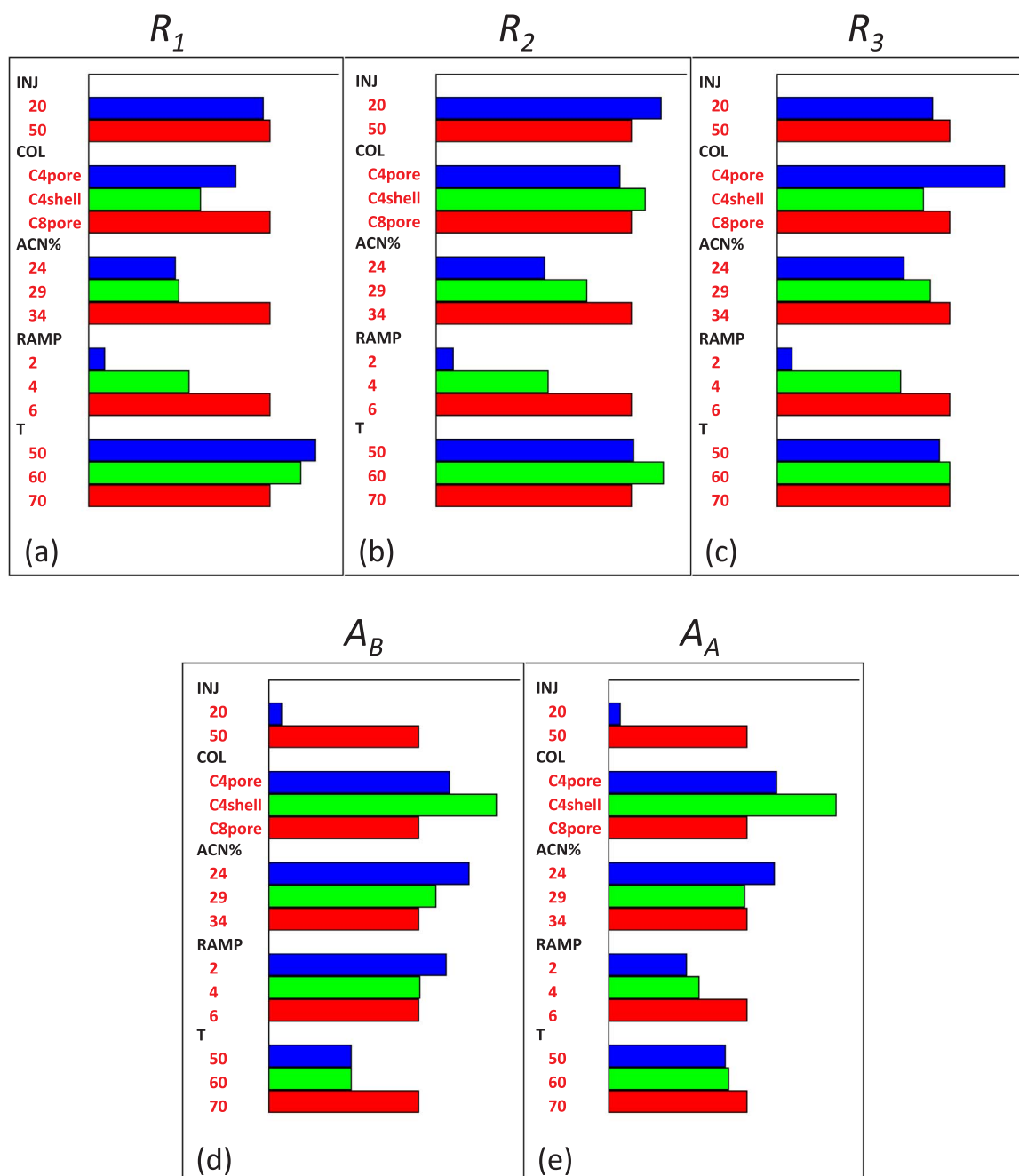


Fig. 2. Graphic analysis of effects for investigation of chromatographic resolutions and area responses. Resolutions: (a) resolution  $R_1$  between NHBA-GNA1030 and PorB; (b) resolution  $R_2$  between PorB and PorA; (c) resolution  $R_3$  between PorA and fHbp-GNA2091. Areas: (d)  $A_B$ , area of PorB peak; (e)  $A_A$ , area of PorA peak. The length of each bar indicates the effect of each level of each factor under study.

0.885, respectively: the difference between  $R^2$  and  $Q^2$  was lower than 0.2; reproducibility was  $> 0.9$ , thus confirming that the models can be effectively used [62]. Hence, contour plots were drawn reporting the calculated isoresponse curves, in order to obtain detailed information on the behavior of each CMA throughout the experimental domain investigated.

The contour plots are shown Fig. 3 for  $K'$  and resolution responses and in Fig. 4 for area responses. The graphic analysis of coefficients is reported in Supplementary Fig. S3, making it possible to point out the terms for each model which were maintained after model refining. The investigation of these two types of graphs allowed understanding the method performance behavior in function of the selected CMPs. As for  $K'$  (Fig. 3a), the only significant effect was exerted by ACN%, which showed both a negative linear and a negative quadratic effect; high

values of this factor anticipate the NHBA-GNA1030 peak toward the column void volume and the zone corresponding to the maximization of this factor was located at medium levels of all the CMPs. Concerning the four resolutions, they were all maximized by high levels of ACN% (Fig. 3b-e). RAMP presented a strong positive effect on resolution  $R_3$  and  $R_4$ . T had a negative effect on  $R_1$  and  $R_4$  and a positive effect on  $R_2$ . Quadratic effects of ACN% were evidenced on all the resolution CMAs apart from  $R_4$ . Some significant interactions were also noticed: ACN%–RAMP and ACN%–T on  $R_1$  and ACN%–T on  $R_2$ . As for the CMAs related to areas (Fig. 4), different trends and curvatures in the contour plots were noticed. The graphical analysis of effects in Supplementary Fig. S3 evidenced negative effects for ACN% on  $A_1$  and  $A_B$  and a positive effect on  $A_A$ , while T showed a positive interaction with ACN% on  $A_1$  and  $A_B$ . RAMP showed no influence on area responses.



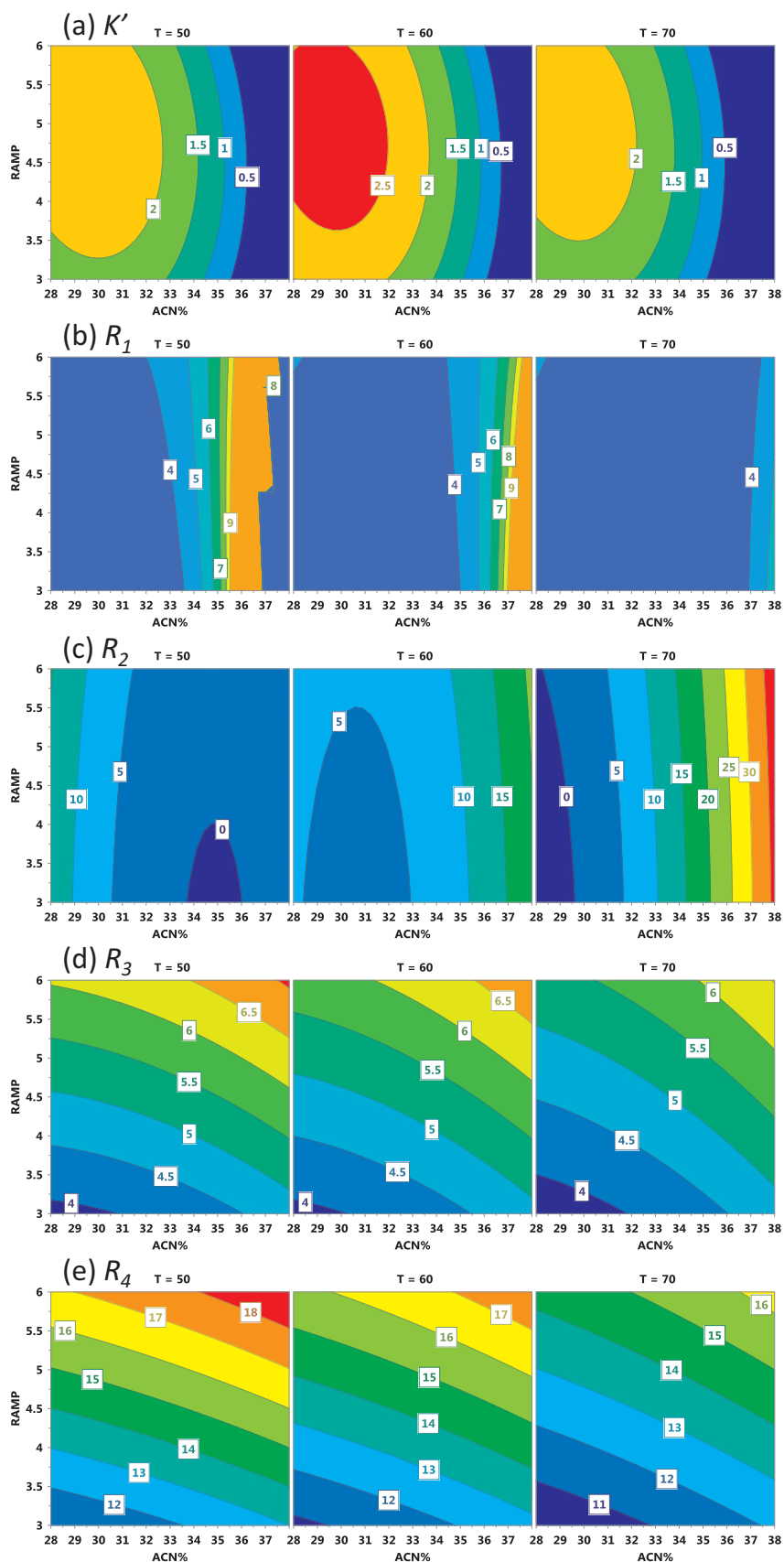


Fig. 3. Isoresponse surfaces drawn by plotting ramp time (RAMP) vs. organic phase starting concentration (ACN%) for: (a) NHBA-GNA1030 capacity factor  $K'$ ; (b) resolution  $R_1$  between NHBA-GNA1030 and PorB; (c) resolution  $R_2$  between PorB and PorA; (d) resolution  $R_3$  between PorA and fHbp-GNA2091; (e) resolution  $R_4$  between fHbp-GNA2091 and NadA at three different values of temperature: 50 °C, 60 °C, 70 °C.

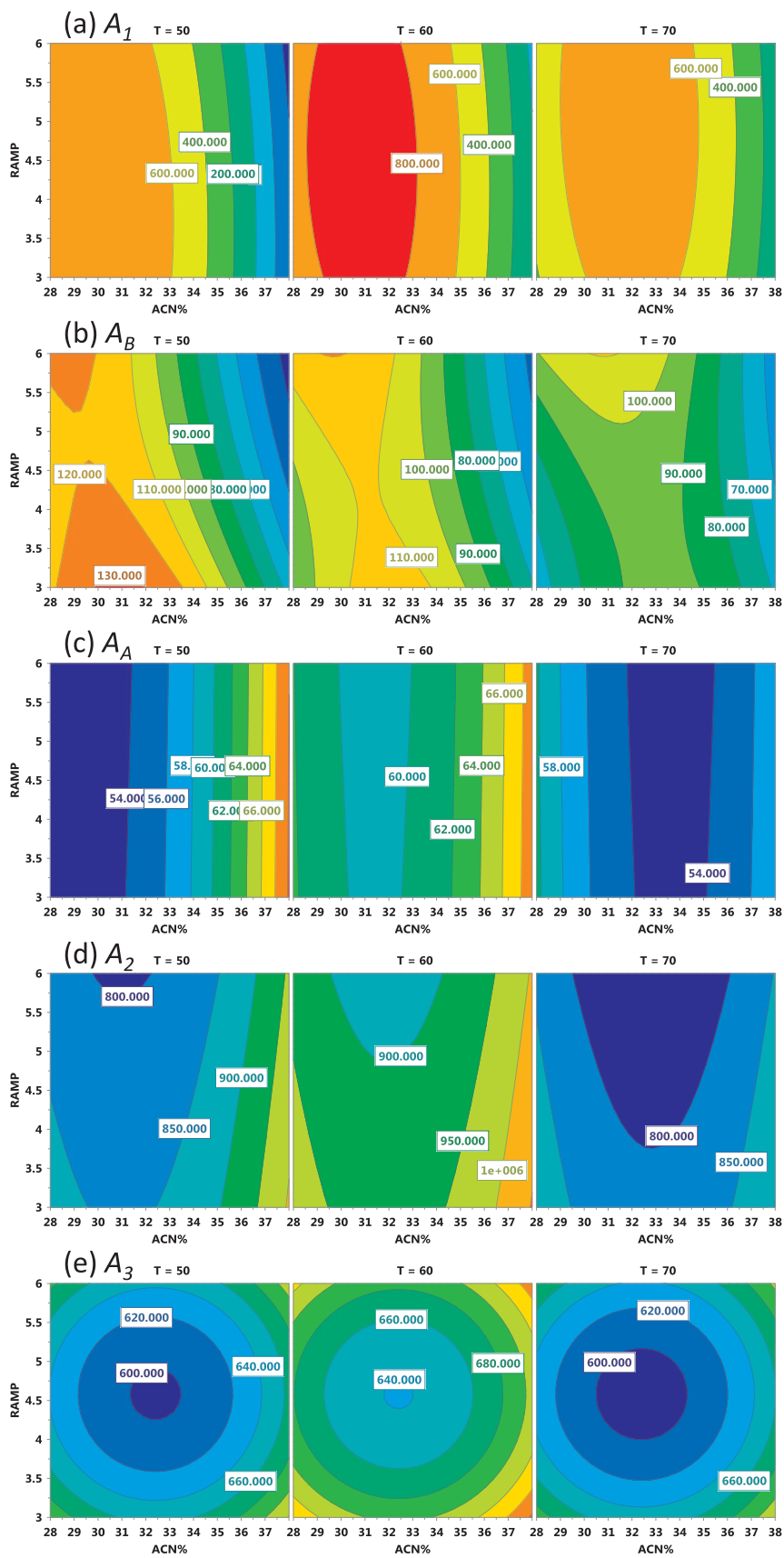


Fig. 4. Isoresponse surfaces drawn by plotting ramp time (RAMP) vs. organic phase starting concentration (ACN%) for: (a)  $A_1$ , area of NHBA-GNA1030 peak; (b)  $A_B$ , area of PorB peak; (c)  $A_A$ , area of PorA peak; (d)  $A_2$ , area of fHbp-GNA2091 peak; (e)  $A_3$ , area of NadA peak at three different values of temperature: 50 °C, 60 °C, 70 °C.

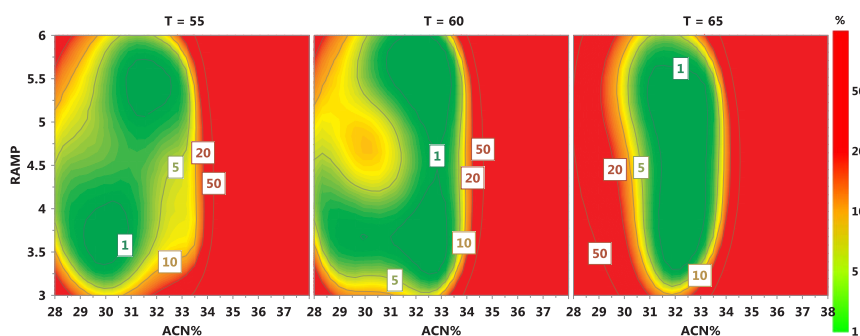


Fig. 5. Probability map by plotting ramp time (RAMP) vs. organic phase starting concentration (ACN%) at three different values of temperature: 55 °C, 60 °C, 65 °C, obtained by setting the desired requirements as reported in Supplementary Table S1.

The sweet spot plots shown in Supplementary Fig. S4 made it possible to highlight by different colors the areas where one or more predicted CMAs fulfill the related requirements. For all the CMAs related to resolution, a desired minimum value of 1.5 was set, while for the other CMAs the limits were set as reported in Supplementary Table S1. The different colors had the following meaning: purple, from 1 to 3 CMAs criteria met; different blue gradations, from 4 to 5 and from 6 to 7 criteria met; light blue, from 8 to 9 criteria met; and finally, green where all the ten CMAs criteria were fulfilled, namely the zone corresponding to the sweet spot.

Anyway, the green plot area does not constitute the DS, but it only shows where the overlay of the CMAs response surfaces meets the desired performances, providing an additional information by highlighting the zone where the maximum number of criteria is fulfilled. On the other hand, the DS takes into account the concept of probability that the requirements are met [21,25], and it was computed by MODDE software [57]. Considering all the response surface models and the settings of CMAs requirements, the risk of failure map (Fig. 5) was drawn using Monte-Carlo simulations for risk analysis, taking into account the model parameters uncertainty, propagating the prediction error from parameters to responses and giving access to the responses distributions for each RSM condition [63]. The selected level of probability ( $\pi$ ) and the original set point for DS identification were:  $\pi \geq 90\%$ ; ACN%, 33.3%; RAMP, 4.8 min, T, 64 °C. Hence, the DS was graphically represented in the risk of failure map by the area colored in green, within the iso-error curves corresponding to the probability

failure  $\leq 10\%$ . Within the DS, the following limits for CMPs ranges were identified: ACN%, 32.0–34.6%; RAMP, 4.0–5.6 min; T, 60–68 °C. In order to validate the DS, the lower and higher limits of the DS ranges were selected as the  $-1$  and  $+1$  levels of a  $2^3$  Full Factorial Design [49], with 3 replicates of the original set point to estimate the experimental variance, and the agreement between RSM predicted values and experimental responses at the edges of failure was verified. Once the model validity was confirmed, a working point was chosen inside the lower risk region ( $\pi \geq 99\%$ ), taking into account some practicability factors such as the advantages of implementing an high-throughput method and of choosing a temperature value as low as possible to avoid proteins damages or modifications. The selected working point for routine use was: ACN%, 33.0%; RAMP, 4.0 min; T, 60 °C, and made it possible to completely separate the five antigens in about 5 min with the desired selectivity and sensitivity, as from the chromatogram reported in Fig. 6. From the observation of the final chromatogram, it is worthwhile to note that NadA peak does not present a gaussian symmetry. This is due to the fact that the most intense peak is related to the intact form of the core protein Neisseria Adhesin A, while the hump at the end of the chromatographic signal is due to the truncated (C-terminal deleted) forms [52]. All these forms are determined together, and the quantification refers to the entire signal of NadA.

### 3.5. Robustness and control strategy

For robustness study, a 8-run Full Factorial Design was employed to

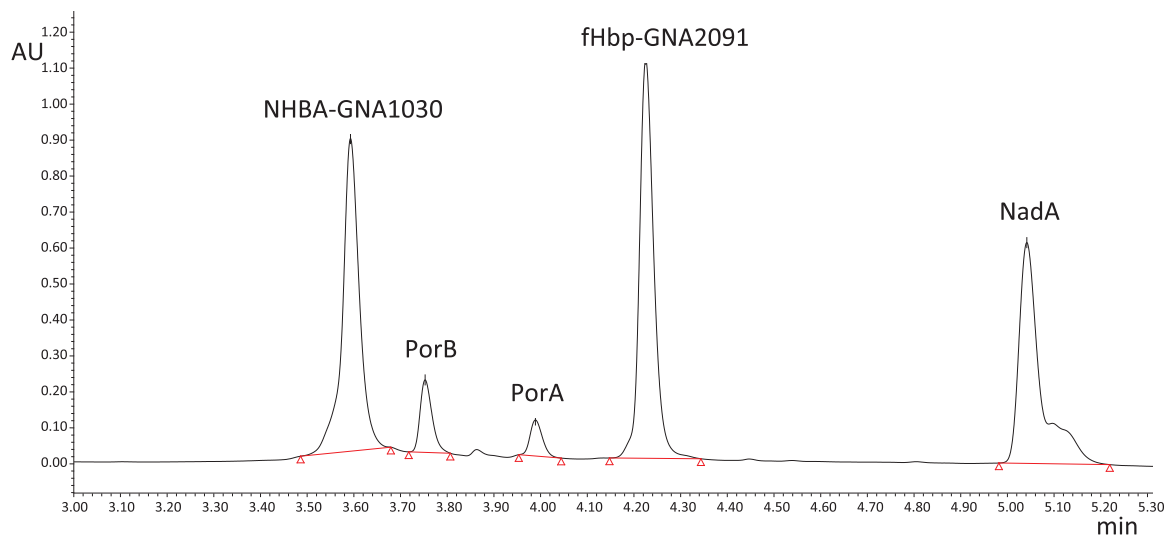


Fig. 6. RP-UHPLC chromatogram in the working conditions for a mock solution ( $10 \mu\text{gml}^{-1}$  NHBA-GNA1030, fHbp-GNA2091, NadA and  $5 \mu\text{gml}^{-1}$  Por B, PorA). Acquity H-ClassBio™ UPLC system; Acquity RP-C4 BEH 300 Å, 1.7  $\mu\text{m}$ ,  $2.1 \times 150$  mm; organic phase, ACN plus 0.1% TFA ion pair; injection volume, 30  $\mu\text{l}$ ; temperature, 60 °C; ramp time, 4 min (from 33.0% to 75.0% ACN); photodiode array detector, 210 nm, 1.2 nm resolution, 20 points  $\text{s}^{-1}$ .

calculate the main effects on the CMAs of small CMPs changes around the chosen working conditions [59] and the experimental plan with the responses is shown in [Supplementary Table S7](#). The resulting graphical analysis of effects is shown in [Supplementary Fig. S5](#). The analysis of effects evidenced that the method could be considered robust, including a precautionary statement for ACN%. As a matter of facts, ACN% exerted a significant effect on  $K'$ ,  $R_2$ ,  $A_1$  and  $A_B$  also in this small interval. As for the other factors, RAMP was significant only on  $R_2$  and  $T$  on  $A_1$  and  $A_B$ . Anyhow, all the results were within the CMAs desired limits for all the experiments.

Finally, a preliminary control strategy [21] for the RP-UHPLC method was accomplished on the basis of all the development data, consisting in a continuous performance verification plan to control recombinant proteins resolutions, efficiency and peak symmetry, setting the following system suitability limits. A standard mixture of the rMenB proteins ( $100 \mu\text{g ml}^{-1}$ ) should fulfill these requirements: resolution (NHBA-GNA1030/fHBp-GNA2091)  $\geq 9$ , resolution  $R_4$  (fHBp-GNA2091/NadA)  $\geq 10$ , fHBp-GNA2091 number of theoretical plates  $\geq 50,000$  and  $0.9 \leq \text{fHBp-GNA2091 peak symmetry} \leq 1.6$ .

### 3.6. Qualification of the method

Qualification of the RP-UHPLC developed method was carried out to ensure the suitability of the method for the quality control of the vaccine product. The method was qualified following ICH guidelines [59] for validation, using an experimental plan for testing the effect of noise factor on results variability: operator, column batch, sample preparation, instrument, day of analysis and analytical session. Specificity of the method was evaluated by preparing and analyzing different mock standard solutions where the antigens were removed one by one, checking for the absence of the corresponding peak in the chromatogram. Also, separate solutions of the single drug substances and a mock standard solution containing all the analytes were analysed and compared, verifying the correspondence of the peaks in the chromatograms. The qualification results showed the suitability of the method for the intended purpose, identity of antigens respect to standard solutions, selectivity, linearity, accuracy (recovery for antigens spike), precision (data variation of the titred amount), quantitation limits. Quantitation limits were determined by the analysis of samples with known concentrations of analyte and establishing the minimum level at which each analyte can be quantified with acceptable accuracy and precision, and corresponded to the lower limits of the linearity curves ( $1 \mu\text{g ml}^{-1}$  for rMenB antigens and  $0.5 \mu\text{g ml}^{-1}$  for OMV antigens). Linearity was evaluated by preparing and analyzing ten samples, two for each of five concentration values, ranging from 1 to  $20 \mu\text{g ml}^{-1}$  for rMenB antigens and from 0.5 to  $10 \mu\text{g ml}^{-1}$  for OMV antigens, and the related data are shown in [Supplementary Table S8](#). The recovery values for antigens spike were measured at three concentration values (1.5, 10 and  $20 \mu\text{g ml}^{-1}$ ) performing 6 replicates and obtaining values included in the following intervals: NHBA-GNA1030, 92.3–113.9%; PorB, 93.1–115.9%; PorA, 96.1–108.6%; fHBp-GNA2091, 99.3–106.3%; NadA, 90.9–104.7%. Precision was assessed for fHBp-GNA2091 ( $10 \mu\text{g ml}^{-1}$ ): as for repeatability, six analyses were run obtaining a RSD = 1.7% for the titred amount; as for intermediate precision, calculated by performing 6 replicates at three concentration values, the following RSD values for the titred amount were found:  $1.5 \mu\text{g ml}^{-1}$ , RSD = 8.4%;  $10 \mu\text{g ml}^{-1}$ , RSD = 7.9%;  $20 \mu\text{g ml}^{-1}$ , RSD = 2.8%.

## 4. Conclusions

A fast and fully QbD-compliant RP-UHPLC method was developed for the simultaneous determination of the Bexsero components in the vaccine supernatant. The implementation of AQbD allowed enhanced understanding of the analytical method, whose development was effectively supported by experimental design for quality risk management and performance evaluation. For the first time in the literature, QbD

was applied to drive to an analytical design space for a vaccine product. The data gathered through well-suited experimental designs, which were planned first in a screening phase and then in a RSM study, made it possible to identify the DS, defined on the basis of the calculated models for the CMPs and on the basis of Monte-Carlo simulations. The DS consisted of a set of CMPs conditions which provided satisfactory values for all the CMAs, with a selected degree of probability. QbD applied to analytical method development, leading to DS identification, represents an added value, providing advantages for adequate method knowledge, ensuring regulatory flexibility during lifecycle, improving method robustness and making progress in safety for both product quality and control strategy. The developed method successfully passed the qualification process demonstrating to be useful for routine analysis of Bexsero vaccine.

## Acknowledgements

The authors wish to thank Dr. Amin Khan for the strong sponsorship of QbD application and implementation inside GSK, supporting the PhD program at the basis of this study. We would also like to thank QbD Senior Advisors Dr. Timothy Schofield and Dr. William Egan and the GSK organization program heads Dr. Davide Serruto and Dr. Rino Rappuoli at GSK.

## Conflict of interest

L. Nompari, C. Campa and M. Rovini are employees of GSK, Siena, Italy. The other authors declaim no conflict of interest.

## Trademark statement

Bexsero is a trademark owned by or licensed to the GSK group of companies.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.talanta.2017.09.077>

## References

- [1] S. Thomas, R. Dilbarova, R. Rappuoli, Future challenges for vaccinologists, in: S. Thomas (Ed.), *Vaccine Design-Methods and Protocols- Volume 1: Vaccines for Human Diseases*, Volume 1403 of the Series Methods in Molecular Biology, Humana Press, Springer Protocols, 2016, pp. 41–55.
- [2] M.J. Bottomley, R. Rappuoli, O. Finco, Vaccine design in the 21st century, in: B.R. Bloom, P.-H. Lambert (Eds.), *The Vaccine Book*, 2nd ed., Academic Press, Elsevier, 2016, pp. 45–65.
- [3] R. Rappuoli, E. De Gregorio, Editorial overview: Vaccines: novel technologies for vaccine development, *Curr. Opin. Immunol.* 41 (2016) V–VII.
- [4] A.J. Pollard, Global epidemiology of meningococcal disease and vaccine efficacy, *Pediatr. Infect. Dis. J.* 23 (2004) S274–S279.
- [5] L.H. Harrison, Prospects for vaccine prevention of meningococcal infection, *Clin. Microbiol. Rev.* 19 (2006) 142–164.
- [6] L. Jódar, I.M. Feavers, D. Salisbury, D.M. Granoff, Development of Vaccines Against Meningococcal Disease, 359, 2002, pp. 1499–1508.
- [7] D.S. Stephens, Conquering the meningococcus, *FEMS Microbiol. Rev.* 31 (2007) 3–14.
- [8] A. Gandhi, P. Balmer, L.J. York, Characteristics of a new meningococcal serogroup B vaccine, bivalent rLP2086 (MenB-FHbp; Trumenba<sup>®</sup>), *Postgrad. Med.* 128 (2016) 548–556.
- [9] A. Bardotti, G. Averani, F. Berti, S. Berti, V. Carinci, S. D'Ascenzi, B. Fabbri, S. Giannini, A. Giannozzi, C. Magagnoli, D. Proietti, F. Norelli, R. Rappuoli, S. Ricci, P. Costantino, Physicochemical characterisation of glycoconjugate vaccines for prevention of meningococcal diseases, *Vaccine* 26 (2008) 2284–2296.
- [10] P.S. Watson, D.P.J. Turner, Clinical experience with the meningococcal B vaccine, Bexsero<sup>®</sup>: Prospects for reducing the burden of meningococcal serogroup B disease, *Vaccine* 34 (2016) 875–880.
- [11] T. Vesikari, S. Esposito, R. Prymula, E. Ypma, I. Kohl, D. Toneatto, P. Dull, A. Kimura, EU meningococcal serogroup B vaccine (4CMenB) administered concomitantly with routine infant and child vaccinations: results of two randomised trials, *Lancet* 381 (2013) 825–835.
- [12] D. Serruto, M.J. Bottomley, S. Ram, M.M. Giuliani, R. Rappuoli, The new multi-component vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens, *Vaccine* 30 (2012)

- B87–B97.
- [13] M. Pizzi, V. Scarlato, V. Masignani, M.M. Giuliani, B. Aricò, M. Comanducci, G.T. Jennings, L. Baldi, E. Bartolini, B. Capecci, C.L. Galeotti, E. Luzzi, R. Manetti, E. Marchetti, M. Mora, S. Nuti, G. Ratti, L. Santini, S. Savino, M. Scarselli, E. Storni, P. Zuo, M. Broecker, E. Hundt, B. Knapp, E. Blair, T. Mason, H. Tettelin, D.W. Hood, A.C. Jeffries, N.J. Saunders, D.M. Granoff, J.C. Venter, E.R. Moxon, G. Grandi, R. Rappuoli, Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing, *Science* 287 (2000) 1816–1820.
- [14] C. Tani, M. Stella, D. Donnarumma, M. Biagini, P. Parente, A. Vadi, C. Magagnoli, P. Costantino, F. Rigat, N. Norais, Quantification by LC–MS<sup>2</sup> of outer membrane vesicle proteins of the Bexsero<sup>®</sup> vaccine, *Vaccine* 32 (2014) 1273–1279.
- [15] M.M. Giuliani, A. Biolchi, D. Serruto, F. Ferlicca, K. Vienken, P. Oster, R. Rappuoli, M. Pizzi, J. Donnelly, Measuring antigen-specific bactericidal responses to a multicomponent vaccine against serogroup B meningococcus, *Vaccine* 28 (2010) 5023–5030.
- [16] M.V. Pinto, S. Bihari, M.D. Snape, Immunisation of the immunocompromised child, *J. Infect.* 72 (2016) S13–S22.
- [17] European Medicines Agency. Bexsero Assessment Report. <[http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002333/human\\_med\\_001614.jsp&](http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002333/human_med_001614.jsp&)>. (Accessed 3 May 2017).
- [18] U.S. Food and Drug Administration. Bexsero Product Information. <<http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm431374.htm>>. (Accessed 3 May 2017).
- [19] L. Chang, J.T. Blue, J. Schaller, L. Khandke, B.A. Green, Applicability of QbD for vaccine drug product development, in: F. Jameel, S. Hershenson, M.A. Khan, S. Martin-Moe (Eds.), *Quality by Design for Biopharmaceutical Drug Product Development*, Volume 18 of the Series AAPS Advances in the Pharmaceutical Sciences, Springer, 2015, pp. 443–473.
- [20] Guidance for Industry, PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, U.S. Food and Drug Administration, 2004 (Accessed 3 May 2017), <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070305.pdf>.
- [21] ICH Harmonised Tripartite Guideline. Pharmaceutical development Q8(R2), in: Proceedings of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2009.
- [22] ICH Harmonised Tripartite Guideline. Quality risk management Q9, in: Proceedings of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2005.
- [23] ICH Harmonised Tripartite Guideline. Pharmaceutical quality systems Q10, in: Proceedings of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2008.
- [24] J. Haas, A. Franklin, M. Houser, D. Maraldo, M. Mikola, R. Ortiz, E. Sullivan, J.M. Otero, Implementation of QbD for the development of a vaccine candidate, *Vaccine* 32 (2014) 2927–2930.
- [25] S. Orlandini, S. Pinzauti, S. Furlanetto, Application of Quality by Design to the Development of Analytical Separation Methods, 405, 2013, pp. 443–450.
- [26] E. Rozet, P. Lebrun, B. Debrun, B. Boulanger, Ph Hubert, Design spaces for analytical methods, *Trends Anal. Chem.* 42 (2013) 157–167.
- [27] B. Junker, E. Zablackis, T. Verch, T. Schofield, P. Douette, Quality-by-Design: As Related to Analytical Concepts, Control and Qualification, Springer-Verlag, Berlin Heidelberg, 2015, pp. 479–520.
- [28] C. Hubert, P. Lebrun, S. Houari, E. Ziemons, E. Rozet, Ph Hubert, Improvement of a stability-indicating method by quality-by-design versus quality-by-testing: a case of a learning process, *J. Pharm. Biomed. Anal.* 88 (2014) 401–409.
- [29] B. Pasquini, S. Orlandini, M. Del Bubba, E. Bertol, S. Furlanetto, The successful binomium of multivariate strategies and electrophoresis for the Quality by Design separation of a class of drugs: the case of triptans, *Electrophoresis* 36 (2015) 2650–2657.
- [30] K.E. Monks, H.-J. Rieger, I. Molnár, Expanding the term “Design Space” in high performance liquid chromatography (I), *J. Pharm. Biomed. Anal.* 56 (2011) 874–879.
- [31] T. Tol, N. Kadam, N. Raotole, A. Desai, G. Samanta, A simultaneous determination of related substances by high performance liquid chromatography in a drug product using quality by design approach, *J. Chromatogr. A* 1432 (2016) 26–38.
- [32] J. Pantović, A. Malenović, A. Vemić, N. Kostić, M. Medenica, Development of liquid chromatographic method for the analysis of dabigatran etexilate mesilate and its ten impurities supported by quality-by-design methodology, *J. Pharm. Biomed. Anal.* 111 (2015) 7–13.
- [33] A.H. Schmidt, I. Molnár, Using an innovative quality-by-design approach for the development of a stability indicating UHPLC method for ebastine in the API and pharmaceutical formulations, *J. Pharm. Biomed. Anal.* 78–79 (2013) 65–74.
- [34] C. Boussès, L. Ferey, E. Vedrines, K. Gaudin, Using an innovative combination of quality-by-design and green analytical chemistry approaches for the development of a stability indicating UHPLC method in pharmaceutical products, *J. Pharm. Biomed. Anal.* 115 (2015) 114–122.
- [35] J. Terzić, I. Popović, A. Stajić, A. Tumpa, B. Jančić-Stojanović, Application of analytical quality by design concept for bilastine and degradation impurities determination by hydrophilic interaction liquid chromatographic method, *J. Pharm. Biomed. Anal.* 125 (2016) 385–393.
- [36] B. Andri, P. Lebrun, A. Dispas, R. Klinkenberg, B. Streeel, E. Ziemons, R.D. Marini, Ph Hubert, Optimization and validation of a fast supercritical fluid chromatography method for the quantitative determination of vitamin D3 and its related impurities, *J. Chromatogr. A* 1491 (2017) 171–181.
- [37] A. Dispas, V. Desfontaine, B. Andri, P. Lebrun, D. Kotoni, A. Clarke, D. Guillaume, P. Hubert, Quantitative determination of salbutamol sulfate impurities using achiral supercritical fluid chromatography, *J. Pharm. Biomed. Anal.* 134 (2017) 170–180.
- [38] E. van Tricht, L. Geurink, H. Backus, M. Germano, G.W. Somsen, C.E. Säger-van de Griend, One single, fast and robust capillary electrophoresis method for the direct quantification of intact adenovirus particles in upstream and downstream processing samples, *Talanta* 166 (2017) 8–14.
- [39] S. Furlanetto, S. Orlandini, B. Pasquini, C. Caprini, P. Mura, S. Pinzauti, Fast analysis of glibenclamide and its impurities: quality by design framework in capillary electrophoresis method development, *Anal. Bioanal. Chem.* 407 (2015) 7637–7646.
- [40] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, S. Pinzauti, S. Furlanetto, Analytical quality by design in pharmaceutical quality assurance: development of a capillary electrophoresis method for the analysis of zolmitriptan and its impurities, *Electrophoresis* 36 (2015) 2642–2649.
- [41] S. Orlandini, B. Pasquini, R. Gotti, A. Giuffrida, F. Paternostro, S. Furlanetto, Analytical quality by design in the development of a cyclodextrin-modified capillary electrophoresis method for the assay of metformin and its related substances, *Electrophoresis* 35 (2014) 2538–2545.
- [42] S. Orlandini, B. Pasquini, M. Del Bubba, S. Pinzauti, S. Furlanetto, Quality by design in the chiral separation strategy for the determination of enantiomeric impurities: development of a capillary electrophoresis method based on dual cyclodextrin systems for the analysis of levosulpiride, *J. Chromatogr. A* 1380 (2015) 177–185.
- [43] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, M. Douša, S. Pinzauti, S. Furlanetto, Enantioseparation and impurity determination of ambrisentan using cyclodextrin-modified micellar electrokinetic chromatography: visualizing the design space within quality by design framework, *J. Chromatogr. A* 1467 (2016) 363–371.
- [44] B. Pasquini, S. Orlandini, C. Caprini, M. Del Bubba, M. Innocenti, G. Brusotti, S. Furlanetto, Cyclodextrin- and solvent-modified micellar electrokinetic chromatography for the determination of captopril, hydrochlorothiazide and their impurities: a quality by design approach, *Talanta* 160 (2016) 332–339.
- [45] S. Orlandini, B. Pasquini, M. Stocchero, S. Pinzauti, S. Furlanetto, An integrated quality by design and mixture-process variable approach in the development of a capillary electrophoresis method for the analysis of almotriptan and its impurities, *J. Chromatogr. A* 1339 (2014) 200–209.
- [46] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, L. Squarzialupi, V. Colotta, S. Furlanetto, A comprehensive strategy in the development of a cyclodextrin-modified microemulsion electrokinetic chromatographic method for the assay of diclofenac and its impurities: Mixture-process variable experiments and quality by design, *J. Chromatogr. A* 1466 (2016) 189–198.
- [47] G. Piepel, B. Pasquini, S. Cooley, A. Heredia-Langner, S. Orlandini, S. Furlanetto, Mixture-process variable approach to optimize a microemulsion electrokinetic chromatography method for the quality control of a nutraceutical based on coenzyme Q10, *Talanta* 97 (2012) 73–82.
- [48] S. Orlandini, R. Gotti, S. Furlanetto, Multivariate optimization of capillary electrophoresis method: a critical review, *J. Pharm. Biomed. Anal.* 87 (2014) 290–307.
- [49] G.A. Lewis, D. Mathieu, R. Phan-Tan-Luu, *Pharmaceutical Experimental Design*, Marcel Dekker, New York, 1999.
- [50] T. Kourti, B. Davis, The business benefits of Quality by Design (QbD), *Pharm. Eng.* 32 (2012) 1–4.
- [51] A. Martino, C. Magagnoli, G. De Conciliis, S. D’Ascenzi, M.J. Forster, L. Allen, C. Brookes, S. Taylor, X. Bai, J. Findlow, I.M. Feavers, A. Rodger, B. Bolgiano, Structural characterisation, stability and antibody recognition of chimeric NHBA-GNA1030: an investigational vaccine component against *Neisseria meningitidis*, *Vaccine* 30 (2012) 1330–1342.
- [52] C. Magagnoli, A. Bardotti, G. De Conciliis, R. Galasso, M. Tomei, C. Campa, C. Pennatini, M. Cerchioni, B. Fabbri, S. Giannini, G.L. Mattioli, A. Biolchi, S. D’Ascenzi, F. Helling, Structural organization of NadA<sub>351–405</sub>, a recombinant MenB vaccine component, by its physico-chemical characterization at drug substance level, *Vaccine* 27 (2009) 2156–2170.
- [53] S. Fekete, R. Berky, J. Fekete, J.-L. Veuthey, D. Guillaume, Evaluation of recent very efficient wide-pore stationary phases for the reversed-phase separation of proteins, *J. Chromatogr. A* 1252 (2012) 90–103.
- [54] LabSolutions Analytical Data System, C191-E016, Shimadzu Corp., 2012. <<https://www.ssi.shimadzu.com/products/literature/cds/C191-E016.pdf>>. (Accessed 3 May 2017).
- [55] Empower 3, Product Code 720001257EN, Waters Corp. <[http://www.waters.com/waters/library.htm?locale=it\\_IT&cid=10190669&lid=1529289](http://www.waters.com/waters/library.htm?locale=it_IT&cid=10190669&lid=1529289)>. (Accessed 3 May 2017).
- [56] D. Mathieu, J. Nony, R. Phan-Tan-Luu, NEMROD-W, LPRAI sarl, Marseille.
- [57] MODDE v. 10, MKS Umetrics AB, Sweden.
- [58] S. Furlanetto, S. Orlandini, B. Pasquini, M. Del Bubba, S. Pinzauti, Quality by design approach in the development of a solvent-modified micellar electrokinetic chromatography method: Finding the design space for the determination of amitriptyline and its impurities, *Anal. Chim. Acta* 802 (2013) 113–124.
- [59] ICH Harmonised Tripartite Guideline. Validation of Analytical Procedures: Text and Methodology Q2(R1), in: Proceedings of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2005.
- [60] USP Validation and Verification Expert Panel: G.P. Martin, MS (Chair), K.L. Barnett, C. Burgess, P.D. Curry, J. Ermer, G.S. Gratzl, J. P. Hammond, J. Herrmann, E. Kovacs, D.J. LeBlond, R. LoBrutto, A.K. McCasland-Keller, P.L. McGregor, P. Nethercote, A.C. Templeton, D.P. Thomas, M.L.J. Weitzel, H. Pappa, Proposed New USP General Chapter: The Analytical Procedure Lifecycle <1220>. <[http://www.usp.org/sites/default/files/usp\\_pdf/EN/USPNF/revisions/s201784.pdf](http://www.usp.org/sites/default/files/usp_pdf/EN/USPNF/revisions/s201784.pdf)>. (Accessed 3 May 2017).
- [61] USP Validation and Verification Expert Panel: G.P. Martin, MS (Chair), K.L. Barnett, C. Burgess, P.D. Curry, J. Ermer, G.S. Gratzl, J. P. Hammond, J. Herrmann, E. Kovacs, D.J. LeBlond, R. LoBrutto, A.K. McCasland-Keller, P.L. McGregor, P. Nethercote, A.C. Templeton, D.P. Thomas, M.L.J. Weitzel, Lifecycle Management of Analytical Procedures: Method Development, Procedure Performance Qualification, and Procedure Performance Verification. <[http://www.usp.org/sites/default/files/usp\\_pdf/EN/USPNF/revisions/lifecycle.pdf](http://www.usp.org/sites/default/files/usp_pdf/EN/USPNF/revisions/lifecycle.pdf)>. (Accessed 3 May 2017).
- [62] L. Eriksson, E. Johansson, N. Kettaneh-Wold, C. Wikström, S. Wold, Design of Experiments – Principles and Applications, MKS Umetrics AB, Umeå, Sweden, 2008.
- [63] M.A. Herrador, A.G. Asuero, A.G. Gonzalez, Estimation of the uncertainty of indirect measurements from the propagation of distributions by using the Monte-Carlo method: an overview, *Chemom. Intell. Lab. Syst.* 79 (2005) 115–122.