



**HAL**  
open science

**A new combined LC (ESI+) MS/MS QTOF impurity fingerprinting and chemometrics approach for discriminating active pharmaceutical ingredient origins: example of simvastatin**

Dominique Hirth

► **To cite this version:**

Dominique Hirth. A new combined LC (ESI+) MS/MS QTOF impurity fingerprinting and chemometrics approach for discriminating active pharmaceutical ingredient origins: example of simvastatin. Analytical chemistry. 2011. dumas-00960820

**HAL Id: dumas-00960820**

**<https://dumas.ccsd.cnrs.fr/dumas-00960820>**

Submitted on 18 Mar 2014

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

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

# CONSERVATOIRE NATIONAL DES ARTS ET METIERS

Centre d'enseignement de Montpellier

## MEMOIRE

*Présenté en vue d'obtenir*

Le **DIPLOME d'INGENIEUR CNAM**

SPECIALITE : MESURE-ANALYSE

OPTION : SCIENCES ET TECHNIQUES ANALYTIQUES  
APPLIQUEES A LA CHIMIE ET AU VIVANT

Par

**M. Dominique HIRTH**

---

**A new combined LC (ESI+) MS/MS QTOF impurity fingerprinting and chemometrics approach for discriminating active pharmaceutical ingredient origins: example of simvastatin.**

---

**Soutenu le 8 juillet 2011**

**JURY**

**PRESIDENT : Pr. Christine PERNELLE**

**MEMBRES : Pr. Torbjörn ARVIDSSON  
Pr. Pierre-Antoine BONNET  
Pr. Christophe MOULIN  
Dr. Nathalie MARCOTTE**

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Torbjörn Arvidsson for having welcomed me within the Swedish Medical Products Agency Laboratory. He always considered me as a full member of his staff, he placed a total trust in me, while demanding, so that it was really pleasant and enlightening to work on his sides. His supervision and guidance was of great teaching for me. Thanks to him for having given me the opportunity to attend the “Analysdagarna 2010” lectures at the University of Uppsala.

My deepest gratefulness goes also to Professor Monika Johansson for her invaluable advices and relevant views indispensable to the progress and success of this project.

I extend my genuine thanks to Dick Fransson for having taught me so much knowledge about the LC-MS technology. His expertise, availability, kindness and patience were deeply appreciable.

I thank, in particular, Marianne Ek, Anette Silvàn, Ahmad Amini, Stefan Jönsson and Ian Mac Even for the numerous discussions and skills that they shared with me about their respective activities, such as the European Pharmacopoeia for Marianne, the quality assurance for Anette, the capillary electrophoresis and MALDI-TOF technologies for Ahmad, the nuclear magnetic resonance for Ian and the high performance liquid chromatography for Stefan.

I would like to address my grateful thanks to all my Swedish colleagues for the warm and exceptional welcoming that they showed me during all my stay within their laboratory. Their sympathy, kindness and benevolence were constant, so that I will never forget this experience. Many of them have become friends.

I thank Professor Curt Petersson and, in his name, all the staff of the Analytical Pharmaceutical Chemistry Department of the Uppsala University Pharmacy Faculty for their support and help with the comprehension and manipulation of the multivariate data analysis software. Curt contributed to facilitate my registration at the University of Uppsala and gave me the opportunity to validate a Degree Project (30 credits) in Analytical Pharmaceutical Chemistry.

I would like to thank my company, the French Health Products and Safety Agency, and specially, Professor Alain Nicolas, Professor Pierre-Antoine Bonnet and Denis Chauvey for having supported me in this initiative and approach to increase my professional experience. I thank more particularly Professor Alain Nicolas for his investment in time in rereading and correction of this dissertation.

I would like to thank the CNAM teachers in helping me to broaden my skills and knowledge all along my training cursus. This dissertation is the result and culmination of their instruction and work. I thank more particularly Professor Christine Pernelle and Professor Claudine David for their devotion and professionalism. Thanks to Michel Evers for his valuable help in the correction of this work. They were all of great support for me.

Finally, I express my gratefulness to my parents, my family and all of my closest friends who showed me indefectible support and presence during this venture.

## GLOSSARY OF SYMBOLS AND ABBREVIATIONS

AFSSAPS:	French Health Products Safety Agency.
API:	Active Pharmaceutical Ingredient.
APCI:	Atmospheric Pressure Chemical Ionization.
APPI:	Atmospheric Pressure Photo Ionization.
BRIC:	Brazil, Russia, India, China.
C:	Coulomb.
CEP:	Certificates of Suitability to the Monographs of the European Pharmacopoeia.
CID:	Collision Induced Dissociation.
CAD:	Collision Activated Decomposition.
CRS:	Chemical Reference Substance.
DC:	Direct Current.
EDQM:	European Directorate for the Quality of Medicines and Healthcare.
EEC:	European Economic Community.
EIC:	Extracted Ion Chromatogram.
ESI:	Electrospray Ionization.
FWHM:	Full Width at Half Maximum.
HCA:	Hierarchical Clustering Analysis.
HETP:	Height Equivalent to a Theoretical Plate.
hMG-CoA:	3-hydroxy-3-methylglutaryl coenzyme A
HPLC:	High Performance Liquid Chromatography.
IUPAC:	International Union of Pure and Applied Chemistry.
kHz:	kilo Herz.
LC-MS:	Liquid Chromatography coupled to Mass Spectrometry.
LC-MS/MS:	Liquid Chromatography coupled to Mass Spectrometry in tandem.
LOD:	Limit of Detection.
LOQ:	Limit of Quantification.
M+H <sup>+</sup> :	Pseudo molecular ion.
mM:	Millimolar.
mDa:	milli Dalton.
MVDA:	MultiVariate Data Analysis.
MS:	Mass Spectrometer.
m/z:	Mass to charge ratio.
nm:	Nanometer.

## GLOSSARY OF SYMBOLS AND ABBREVIATIONS (continued)

OMCLs:	Official Medicine Control Laboratories.
PC:	Principal Component.
PCA:	Principal Component Analysis.
pg:	Picogram.
Ph. Eur.:	European Pharmacopoeia.
ppm:	Part Per Million.
QTOF:	Hybrid Quadrupole - Time-of-Flight analyzer.
RF:	Radio Frequency.
RPLC:	Reversed Phase Liquid Chromatography.
RRLC:	Rapid Resolution Liquid Chromatography.
Rs:	Resolution.
RSD%:	Relative Standard Deviation expressed in percent.
SIM:	Selected Ion Monitoring.
SMPA:	Swedish Medical Products Agency.
S/N:	Signal to Noise ratio.
SVT:	Simvastatin.
TIC:	Total Ionic Chromatogram.
t <sub>R</sub> :	Retention time.
TOF:	Time-of-Flight.
t <sub>M</sub> :	Hold-up time.
TWC:	Total Wavelength Chromatogram.
UHPLC:	Ultra High Performance Liquid Chromatography.
UV-DAD:	Ultra Violet Diode Array Detection.
v/v:	Volume to volume.

## FIGURE INDEX

II-1	Graphical representation of Gaussian peaks in a typical chromatogram.....	16
II-2	LC-MS - the marriage between the bird and the fish.....	18
II-3	Principle of LC-MS system.....	19
II-4	Combination of two analyzers in space tandem mass spectrometry.....	20
II-5	Ionization range by ESI, APCI, and APPI as a function of analyte polarity and molecular weight.....	22
II-6	Diagram of an electrospray ionization source functioning in positive mode .....	23
II-7	Photograph of the electrospray process.....	24
II-8	Diagram of an atmospheric pressure chemical ionization source.....	24
II-9	Ionization mechanism in an APCI source.....	25
II-10	Diagram of an atmospheric photo-ionization source .....	26
II-11	Examples of Mathieu stability diagrams for three different masses (upper diagram) and corresponding mass peak widths when applying different linear scan lines (diagram below).....	28
II-12	Schematic diagram of ion trajectories in a quadrupole mass analyzer.....	29
II-13	Schematic of a hybrid quadrupole-time-of-flight mass analyzer.....	30
II-14	Schematic of a reflectron-ToF.....	32
II-15	Resolving power.....	32
II-16	Variance explained by the first principal component.....	37
II-17	Variance explained by the second principal component .....	37
II-18	Score plots of principal component 1 and principal component 2 (top graph) and related scatter plots of principal component 1, 2, 3 (graph below) describing the relationships between raw materials and finished products originated from five different API providers (A, B, C, D and E) present on the French market.....	38
II-19	Example of a dendrogram plot.....	39

## FIGURE INDEX (continued)

III-1	Molecular representation, empirical formula, molecular weight, pKa and log $P_{\text{octanol/water}}$ partition coefficient of simvastatin.....	42
III-2	Cholesterol endogenous synthesis pathway.....	42
III-3	Molecular representation, empirical formula, molecular weight, estimated pKa and log $P_{\text{octanol/water}}$ partition coefficient of simvastatin specified impurities.....	43-44
III-4	Typical UV-chromatogram of a mixture of simvastatin and its specified impurities.....	45
III-5	Performance of Kinetex™ Core-Shell particles compared to fully porous sub-2 $\mu\text{m}$ and 3 $\mu\text{m}$ particles.....	48
III-6	Total ionic chromatogram of 2 $\mu\text{L}$ simvastatin peak for identification CRS solution injected in chromatographic system using various mobile phase buffer concentrations, formic acid 0.1% (top left), formic acid 0.05% (bottom left), formic acid 0.025% (top right) and formic acid 0.001% (bottom right).....	51
III-7	Mass spectrometer signal to noise ratio for various mobile phase buffer concentrations in formic acid (0.1%, 0.05%, 0.025% and 0.001%) of a 2 $\mu\text{L}$ simvastatin for peak identification CRS solution injection .....	52
III-8	Plots of the decimal logarithm of capacity factor $k'$ ( $\log k'$ ) <i>versus</i> mobile phase composition (%B) for simvastatin and major impurities.....	53
III-9	Plots of the decimal logarithm of capacity factor $k'$ ( $\log k'$ ) <i>versus</i> mobile phase composition (%B) for impurities E and F.....	54
III-10	Plots of the decimal logarithm of capacity factor $k'$ ( $\log k'$ ) <i>versus</i> mobile phase composition (%B) for impurities F and G, and simvastatin.....	54
III-11	Plots of the decimal logarithm of capacity factor $k'$ ( $\log k'$ ) <i>versus</i> mobile phase composition (%B) for impurities A and unknown at $m/z = 391.2479$ .....	55
III-12	Plots of the decimal logarithm of capacity factor $k'$ ( $\log k'$ ) <i>versus</i> mobile phase composition (%B) for impurities B and C.....	56
III-13	Bar charts of the retention times against pH values at 2.7 and 6.7 of the mobile phase, for simvastatin, European Pharmacopoeia impurities A, E, F, G and unknown at $m/z = 391.2479$ .....	57
III-14	Plots of the decimal retention times against column temperature for simvastatin, impurities A, E, F, G, B, C and unknown at $m/z = 391.2479$ .....	58
III-15	Total ionic chromatograms of a sample prepared a) in pure acetonitrile. b) in an acetonitrile/water 40:60 (v/v) mixture. c) in pure methanol.....	61

## FIGURE INDEX (continued)

III-16	Detector response when using an electrospray ion source in positive mode (upper diagram). Detector response when using an electrospray ion source in negative mode (lower diagram) corresponding to the injection of the identical solution.....	63
III-17	Plots of simvastatin main impurities peaks area (counts.s) against nebulising gas pressure (psi).....	65
III-18	Plots of the areas (counts.s) against drying gas temperature (°C) for impurities A, E, F, G, B and C and unknown at m/z = 391.2479 and m/z = 421.2949.....	66
III-19	Plots of the areas (counts.s) against drying flow rate (L.min <sup>-1</sup> ) for impurities A, E, F, G, B and C and unknown at m/z = 391.2479 and m/z = 421.2949.....	67
III-20	Plots of the areas (counts.s) against capillary voltage (V) for impurities A, E, F, G,B and C and unknown at m/z = 391.2479 and m/z = 421.2949.....	68
III-21	Plots of the areas (counts.s) against fragmentor voltage (V) for impurities A, E, F, G, B and C and unknown at m/z = 391.2479 and m/z = 421.2949.....	69
III-22	Linearity of the LC-MS signal of simvastatin specified impurities A, E, F, G, B and C and unknown at m/z = 391.2479 and m/z = 421.2949.....	70
III-23	Linearity of simvastatin LC-MS signal.....	71
III-24	Extracted ion chromatogram, displaying abundance and peak to peak signal to noise ratio of a low 8.25 ng.mL <sup>-1</sup> simvastatin concentration solution.....	72
III-25	Agilent 6520 AA QTOF .....	76
III-26	UV-DAD chromatogram of the “simvastatin for peak identification” CRS solution (upper graphic) and gradient profile (lower graphic).....	82
III-27	Blank solution chromatogram .....	83
III-28	Placebo solution chromatogram .....	84
III-29	Example of a finished product solution mass chromatogram.....	85
III-30	a) Extracted ion chromatogram of impurity C - b) Extracted ion chromatogram of impurity B' - c) Extracted ion chromatogram of impurity B - d) Overlaid extracted ion chromatograms of impurities C, B' and B.....	86
III-31	Simvastatin in-tandem mass spectrum at 5 eV collision energy.....	88
III-32	Impurity A' in-tandem mass spectrum at 10 eV collision energy.....	90
III-33	Impurity B' in-tandem mass spectrum at 5eV collision energy.....	91



## FIGURE INDEX (continued)

III-34	Initial PCA calibration model score scatter plot component 1 versus component 2 (left) and PCA calibration model score scatter plot component 1 versus component 3 (right) built up with 15 variables.....	96
III-35	Cross validation of the 15-variable model.....	97
III-36	Contribution intra group E in projection plane component 1 versus component 2.....	97
III-37	Contribution inter groups D and E in projection plane component 1 versus component 3....	98
III-38	Score scatter plots and corresponding loading scatter plots of the final API origin discriminating training model component 1 versus component 2 (upper), component 1 versus component 3 (middle) and component 2 versus component 3 (lower).....	100
III-39	Loadings and uncertainty of the loadings' calculation of the first component (left), the second component (center) and the third component (right).....	101
III-40	Calibration model cross validation.....	102
III-41	PCA predicted validation set in projection plane P1P2 (left) and corresponding HCA three-dimensional predicted validation set (right).....	103
III-42	Predictive three component HCA (upper left) and PCA models (projection planes P1P2 upper right, P1P3 lower left and P2P3 lower right) for 5 unknown samples.....	104

## TABLE INDEX

II-1	Performance comparison of different mass spectrometers.....	33
III-1	Gradient conditions reported in the European Pharmacopoeia monograph on simvastatin (7 <sup>th</sup> edition).....	45
III-2	Final gradient conditions of the developed in-lab method.....	56
III-3	Mass spectrometer starting settings before optimization.....	64
III-4	Mass spectrometric detector linearity for main simvastatin impurities.....	70
III-5	Intra-day (n=6) and inter-day (n=18) instrument precision considering peak areas.....	73
III-6	Intra-day (n=6) and inter-day (n=18) instrument precision considering internal area normalization.....	74
III-7	Unknown impurity information.....	87
III-8	Simvastatin major fragment ions.....	89
III-9	Proposed molecular representations and IUPAC names for impurity A'.....	91
III-10	Proposed molecular representation and IUPAC name for impurity B'.....	92
III-11	Proposed molecular representations and IUPAC names for unknown impurities located at 435.2741 m/z, 433.2585 m/z, 403.2479 m/z and 421.2949 m/z.....	93

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	1
GLOSSARY OF SYMBOLS AND ABBREVIATIONS.....	2
FIGURE INDEX.....	4
TABLE INDEX.....	8
<b>I. INTRODUCTION</b> .....	<b>13</b>
<b>II. MEASUREMENT PRINCIPLE and DATA ANALYSIS:</b> .....	<b>15</b>
High performance liquid chromatography coupled to mass spectrometry in tandem using a hybrid quadrupole - time-of-flight analyzer in conjunction with multivariate data analysis.	
II.1. Reminder.....	15
II.2. High Performance Liquid Chromatography.....	15
II.3. Liquid chromatography hyphenated to mass spectrometry.....	18
II.3.1 LC-MS analysis.....	18
II.3.2 Tandem mass spectrometry.....	20
II.3.3 Atmospheric pressure ionization sources.....	21
II.3.3.1 Electrospray ionization source.....	22
II.3.3.2 Atmospheric pressure chemical ionization source.....	24
II.3.3.3 Atmospheric pressure photo-ionization source .....	26
II.3.3.4 Atmospheric pressure high vacuum interface .....	27
II.3.4 Mass analyzers.....	27
II.3.4.1 Single quadrupole mass analyzer.....	28
II.3.4.2 Hybrid quadrupole – time-of-flight mass analyzer.....	29
II.3.4.3 High resolution and mass accuracy measurements.....	31
II.3.4.4 QTOF operating modes.....	34
II.4. Multivariate data analysis.....	35
II.4.1 Principal component analysis.....	36
II.4.2 Hierarchical clustering analysis.....	39

<b>III. APPLICATION TO SIMVASTATIN AND RELATED SUBSTANCES IN ORDER TO DISCRIMINATE BETWEEN DIFFERENT PROVIDER ORIGINS, ROUTES OF SYNTHESIS OR MANUFACTURING AREAS</b>	41
III.1 Enhanced impurity profiling of simvastatin by LC (ESI+) MS/MS QTOF.....	43
III.1.1 Chromatographic system optimization for an efficient separation of simvastatin and related substances.....	46
III.1.1.1 Choice of the analytical column.....	46
III.1.1.2 Flow rate adjustment.....	49
III.1.1.3 Impact of the mobile phase buffer ionic strength.....	50
III.1.1.4 Effect of the mobile phase organic modifier concentration.....	53
III.1.1.5 Effect of the mobile phase pH.....	57
III.1.1.6 Influence of the column temperature .....	58
III.1.1.7 Autosampler carryover and contaminations.....	60
III.1.1.8 Sample solvent investigation.....	60
III.1.2 Optimization of the mass spectrometer parameters.....	62
III.1.2.1 Choice of the ionization source and functioning mode.....	62
III.1.2.2 Effect of the nebulizer gas pressure.....	64
III.1.2.3 Influence of the drying gas temperature .....	66
III.1.2.4 Drying gas flow rate adjustment.....	67
III.1.2.5 Role of the capillary voltage .....	68
III.1.2.6 Impact of the fragmentor voltage.....	70
III.1.2.7 Response linearity of the mass spectrometric detector.....	72
III.1.2.8 Measurement precision of the mass spectrometer response.....	75
III.1.3 Experimental Disposal.....	75
III.1.3.1 Chemicals and Reagents.....	75
III.1.3.2 Material and apparatus.....	75
III.1.3.3 Preparation of sample solutions.....	77
III.1.3.3.1 Solution of simvastatin for peak identification CRS.....	77
III.1.3.3.2 Starting material solutions.....	78

III.1.3.3.3	Finished product solutions.....	78
III.1.3.3.4	Blank and Placebo solutions.....	78
III.1.3.4	Analytical conditions.....	79
III.1.3.4.1	HPLC experimental conditions.....	79
III.1.3.4.2	Mass spectrometer experimental conditions .....	80
III.1.3.5	Measurement protocol .....	80
III.1.4	Results.....	81
III.1.4.1	UV-DAD Chromatogram.....	82
III.1.4.2	Identification of new impurities by LC-MS/MS.....	83
III.1.4.2.1	Example of a blank injection chromatogram.....	83
III.1.4.2.2	Example of a placebo injection chromatogram .....	84
III.1.4.2.3	Example of a finished product impurity profile.....	84
III.1.4.3	Structure elucidation of new impurities by LC-MS/MS.....	87
III.1.4.3.1	MS/MS spectrum of simvastatin.....	88
III.1.4.3.2	MS/MS spectrum of impurity A'.....	89
III.1.4.3.3	MS/MS spectrum of impurity B'.....	91
III.1.4.3.4	Structure elucidation for impurities located at 435.2741 m/z 433.2585 m/z, 403.2479 m/z and 421.2949 m/z.....	93
III.2.	Chemometric discrimination between different simvastatin API origins.....	94
III.2.1	Development of the calibration model.....	95
III.2.2	Results.....	99
III.2.2.1	Calibration model score scatter plots and associated loading scatter plots.....	99
III.2.2.2	Uncertainty of the PCA calibration model loading calculation.....	101
III.2.2.3	Validation.....	102
III.2.2.3.1	Cross validation.....	102
III.2.2.3.2	Internal validation set.....	103
III.2.2.4	Identification of API origins of unknown pharmaceuticals.....	104

<b>IV. DISCUSSION</b>	105
<b>V. CONCLUSION and PERSPECTIVES</b>	107
APPENDIX A	Structure and physic-chemical data on simvastatin and impurities..... 109
APPENDIX B	Intra-day and inter-day instrument precision considering individual components' absolute peak areas ..... 111
APPENDIX C	Intra-day and inter-day instrument precision considering internal peak area Normalization..... 112
APPENDIX D	Liquid chromatographic parameters..... 113
APPENDIX E	Mass spectrometer parameters..... 114
APPENDIX F	(Fragment pathway and in-tandem mass spectra at 5eV collision energy of molecular ion located at 435.2725 m/z corresponding to (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl]ethyl]-3,7-dimethyl-1,2,3,7,8,8a-hexahydronaphtalen-1-yl-3-hydroxy-2,2-dimethyl-butanoate).  (Fragment pathway and in-tandem mass spectra at 5eV collision energy of molecular ion located at 433.2565 m/z corresponding to (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl]ethyl]-3,7-dimethyl-1,2,3,7,8,8a-hexahydronaphtalen-1-yl-3-hydroxy-2,2-dimethyl-but-3-enoate.... 115
APPENDIX G	(Fragment pathway and in-tandem mass spectra at 5eV collision energy of molecular ion located at 403.2951 m/z corresponding to (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl]ethyl]-3,7-dimethyl-1,2,3,7,8,8a-hexahydronaphtalen-1-yl-2-methyl-but-3-enoate).  (Fragment pathway and in-tandem mass spectra at 10eV collision energy of molecular ion located at 421.2949 m/z corresponding to (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl]ethyl]-3,7-dimethyl-1,2,3,7,8,8a-octahydronaphtalen-1-yl-2,2-dimethyl-butanoate..... 116
APPENDIX H	Reporting, identification and qualification thresholds of related substances in active substances according to the European Pharmacopoeia 7 <sup>th</sup> edition general monograph "Substances for pharmaceutical use (2034)" ..... 117
REFERENCES	..... 118
ABSTRACT	..... 123
KEY WORDS	..... 123

## I. Introduction

Regulatory agencies like the Swedish Medical Products Agency (SMPA) or the French Health Products Safety Agency (AFSSAPS) are competent national authorities responsible for the protection of public health by controlling and guaranteeing the safety, efficacy and quality of medicines [1] and [2]. Both are active and dynamic recognized members of the European network of the Official Medicines Control Laboratories (OMCLs), which is coordinated by the European Directorate for the Quality of Medicines and Healthcare (EDQM) [3]. Among many various missions entrusted to OMCLs, one of their most essential roles encompasses the supervision of medicinal products for human use available on their respective national market and within the European area.

However, over the past few years, those institutions have to face to profound and significant changes in the market organization of active pharmaceutical ingredients (API) and finished products. Indeed, since the enlargement of the European Economic Community (EEC) to twenty seven members in 2007, and furthermore, in a context of an increasingly globalized world economy, all the tendencies in the pharmaceutical industry converge on greater and more systematic internationalization. This results in the outsourcing of pharmaceutical manufacturing to new emerging markets and low-wages countries, such as the BRICs for instance, Brazil, Russia, India and China [4]. Such low-cost alternatives are likely to involve novel concerns over the quality and efficiency of raw materials and finished products, due, sometimes, to an absence of regulation or lesser controls in these lands. Thus, verifying and ensuring the good quality of safe and effective medicines imported into Europe is subject of ever increasing attention, as well as combating illegal and counterfeit medicinal and medical products [5].

Consequently, inspecting manufacturing areas and collaborating with national, European and International organizations have become necessary options for the new strategies within the regulatory bodies. And in the same way, the development in their laboratories of more specific and sensitive analytical methods, by using innovative and powerful techniques, has become a top priority for controlling the pharmaceutical drug compounds. At this prospect, the main challenge of the present work aims to evaluate the possibility of perfecting and developing a generic classification method able to collect chemical fingerprint information for pharmaceutical starting materials, and corresponding finished products, that allows discriminating between different API providers, routes of synthesis, or manufacturing areas, as well as detecting any quality change or purity contamination, or pinpointing counterfeit medicines.

The first objective of this work was to experiment and investigate all the advanced performances, in terms of ultra trace-level sensitivity, increased specificity, high resolution and mass measurement accuracy of high performance liquid chromatography coupled to mass spectrometry in tandem, using a hybrid quadrupole – time-of-flight analyzer (LC-MS/MS QTOF), in order to establish the identification and the impurity profiling of drug substances [6]. In addition to the attractive QTOF instrumentation, modern liquid chromatography technologies, like recent generations of columns packed with superficially porous particles and demonstrating high separation efficiencies, were used in this study [7].

The second objective consisted in exploring multivariate data analysis (MVDA), like principal component analysis (PCA) or hierarchical clustering analysis (HCA), as statistical tools to interpret the datasets and classify the APIs and finished products according to their origins.

Simvastatin, a lipid lowering agent used in the treatment against cholesterol [8-9], was chosen as test molecule for this survey because numerous formulations, containing simvastatin and coming from many manufacturers, are available on the Swedish and French markets, and most of those manufacturers call for several API furnishers. Moreover, prescriptions are required to benefit from treatment based on this medicine, increasing the offer over less regarding internet sites and, by the way, the risk of finding fake pharmaceutical drugs for this molecule.

The first part of this document introduces the basic principles encountered in chromatographic separations using high performance liquid chromatography (HPLC) technology, as well as the coupling of HPLC with mass spectrometry (MS). Several source interfaces, like electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI), are reviewed and the characteristic features of the hybrid quadrupole time-of-flight analyzer are reported. In a second phase, material, apparatus and experimental implementation of the measurement used during the study will be described. The perfecting and the development of the analytical method will be discussed and, especially, the optimization of the chromatographic separation, the setting adjustments of the mass spectrometer parameters and the building of the PCA model will be theorized. Finally, the results obtained from the discrimination between 49 samples by combining LC/MS QTOF impurity fingerprinting with principal component analysis and hierarchical clustering analysis will then be presented in order to confirm the capacity of the developed training model to define the API origins in both starting materials and finished products.



## **II. Measurement principle and data analysis**

High performance liquid chromatography coupled to mass spectrometry in tandem using a hybrid quadrupole time-of-flight analyzer in conjunction with multivariate data analysis.

### **II.1. Reminder**

High performance liquid chromatography hyphenated to mass spectrometry (LC-MS) is an extremely versatile and powerful instrumental technique which has become, during the recent years, an essential investigation tool in trace and ultra trace-level compound analysis. A lot of quantitative and qualitative methods based on LC-MS and LC-MS/MS find their applications in many fields as varied as pharmaceutical industry, proteomics and metabolomics, food-processing industry, environmental protection, forensics and toxicology, etc [10-14]. Numerous stakes, such as detection and quantification of infinitely low quantities of components in very complex matrices, or identification and structural elucidation of molecules, and also molecular composition or functional groups determination, result from the outstanding performances of this technology. The instrumentation comprising a high performance liquid chromatography system in combination, via a suitable interface, with a mass spectrometer, will be presented in the next two chapters, as well as the governing principles and theoretical aspects of both techniques.

### **II.2. High performance liquid chromatography**

High performance liquid chromatography has gained in popularity over the decades in most of analytical laboratories, owing to its suitability for separating and analysing almost all types of complex multi-component mixtures, allowing identification and quantitative determination of targeted molecules. Indeed, advent of compact and automated equipments, composed of high performance modules, like accurate flow delivery systems, on-line degassers, efficient injectors and multiple types of sensitive and selective detectors have led to precise and reproducible analytical results. In the same time, emergence of a wide range of highly effective columns, with many various polarity properties, conducting to enhanced separation, and development of more and more powerful computers have also made their contribution to the growing success of HPLC.

The basic principle of the chromatographic separation lies in the characteristic distribution ratio of each species of the sample mixture between two non miscible phases: the stationary phase, packed in the column, and the liquid mobile phase, which is forced through the column at high pressure by the mean of the pump system. The mobile phase tends, in its motion, to carry away the components to separate, while the stationary phase tends to retain and slow down the components during their migration through the system [15]. Therefore, the separation results from the differences between the specific migration rates of each analyte within the column matrix. More precisely, the separation depends on the solubility differences of the solutes in the mobile phase, and on the relative molecular interactions of those same solutes with the chemical coating of the particles, i.e. it is the more or less great affinity, in terms of polarity, of each injected compound with one of the phases that will determine the time at which the compound will elute from the column. The time for migration of a retained substance is called retention time ( $t_R$ ), while the time for migration of an unretained substance is called hold-up time ( $t_M$ ) [16]. The resulting chromatogram, or plotted detector response versus time, corresponds ideally to a series of Gaussian peaks, as illustrated in figure II-1.

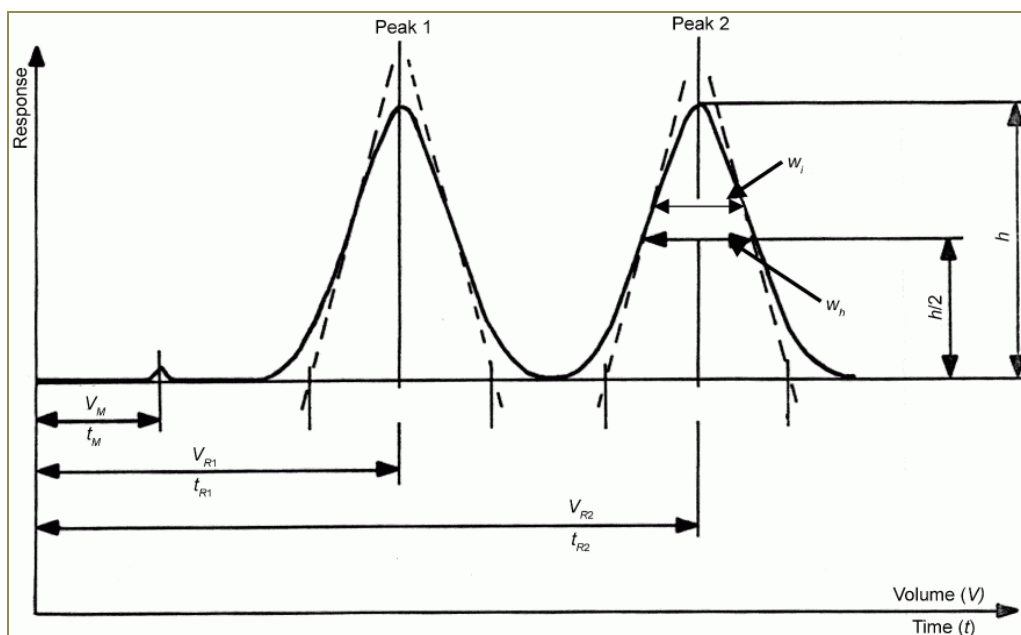


Figure II-1: Graphical representation of Gaussian peaks in a typical chromatogram [16]

In pharmaceutical analysis, reversed phase liquid chromatography (RPLC), is widely used. RPLC is characterized by polar mobile phases, typically mixtures of aqueous solutions with methanol or acetonitrile, and non-polar stationary phases, typically spherical silica particles bonded to hydrophobic alkyl chains, made up of 18 carbons (C18) or 8 carbons (C8), for instance. The

performance of a chromatographic system is measured by the resolution ( $R_s$ ) between two adjacent peaks, according to the following equation:

$$R_s = 1.18 \times (t_{R2} - t_{R1}) / (\omega_{h1} + \omega_{h2}) \quad (1)$$

Where  $t_{R2} > t_{R1}$ , i.e. the second analyte elutes from the column after the first analyte, and where  $\omega_{h1}$  and  $\omega_{h2}$  correspond, in the chromatogram, to the respective peak widths at half height. However, the resolution can be expressed as in equation (2), known as Purnell relation:

$$R_s = \frac{1}{4} \times k_2 / (1 + k_2) \times (\alpha - 1) / \alpha \times \sqrt{N} \quad (2)$$

Where

$$k \text{ is defined as the retention factor: } k = (t_R - t_M) / t_M \quad (3)$$

$$\alpha \text{ is defined as the selectivity: } \alpha = k_2 / k_1 \quad (4)$$

$N$  corresponds to the plate number of the column and conveys the efficiency of the column:

$$N = 5.54 (t_R / \omega_h)^2 \quad (5)$$

The Purnell relation emphasizes the importance of parameters like the retention factor, the selectivity and the column efficiency for chromatographic separation optimization. The retention factor and the selectivity are principally governed by the chemical nature of the stationary phase, the mobile phase composition, the eluent pH or the column temperature. The column efficiency, as for it, is related to the column length, the particle size of the column packing materials and the mobile phase flow rate. Moreover, the selection of a gradient elution instead of an isocratic elution may also be an important criterion when optimizing the chromatographic system. A gradient elution consists in changing the mobile phase composition during the chromatographic run in order to speed up the analysis of late eluting compounds or modify peak shapes and impact the separation mechanisms.

A common idea consists in thinking that liquid chromatography may be simplified when it is combined to a very selective detector such as a mass spectrometer. However, the great capabilities of the MS instrument to separate ions in mass, even if they are not separated in time, should not conceal the extreme importance and the contribution of an efficient chromatography to the quality of the mass spectrometer response. Indeed, a previous chromatographic separation minimizes the

signal mass spectral complexity by reducing the number of the sample matrix co-eluting substances. Thus, it contributes to eliminate or restrain phenomena as ion suppression or signal enhancement. In another hand, liquid chromatography is, unlike mass spectrometry, able to separate isobaric, which means with the same mass, isomeric structures like enantiomers [17].

In the next chapter, the attention is focused on the operating principles and the properties of the online combination of both techniques high performance liquid chromatography and mass spectrometry.

## II.3. Liquid chromatography hyphenated to the mass spectrometry

### II.3.1 LC-MS analysis

The principle of mass spectrometry in LC-MS systems consists in measuring the mass-to-charge ratio of charged particles issued from compounds previously separated by high performance liquid chromatography. Now, mass spectrometry and liquid chromatography are ostensibly not compatible since mass spectrometry needs ultra high vacuum when HPLC operates at high pressure. This, a priori incompatibility of combining both techniques is ideally depicted in figure II-2. This picture called “LC-MS - the marriage between the bird and the fish” was first proposed by Professor P. Arpino and then re-worked by E. Potyrala [18].

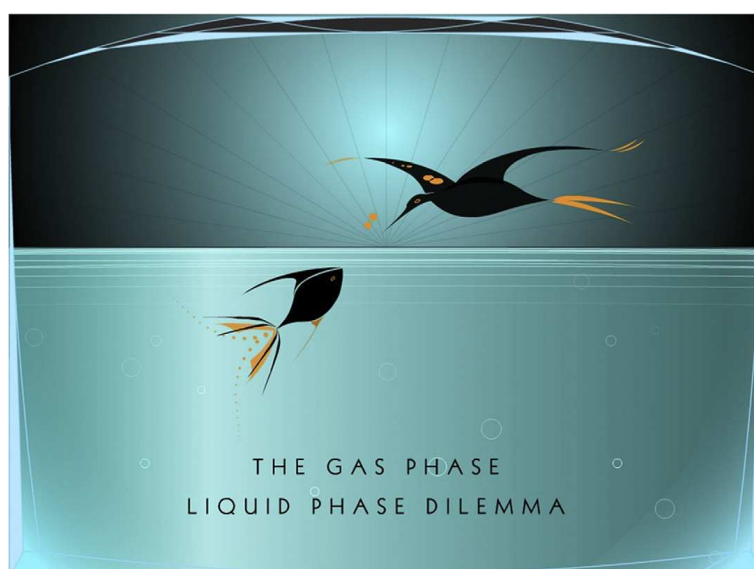


Figure II-2: LC-MS - the marriage between the bird and the fish [18]

The first step in analysis based on liquid chromatography coupled to mass spectrometry consists in generating gas phase ions from analytes dissolved in liquid solutions by using atmospheric pressure ionization sources [19]. The ionic charge is produced either by protonation, i.e. proton addition, or by deprotonation, i.e. a loss of a proton, either by cationic or anionic adduct formation or else, by ejection or capture of an electron. The produced ions are then transmitted via the interface to the mass analyzer where they are separated and measured, according to their mass-to-charge ratio ( $m/z$ ). The ions passing through the mass analyzer are counted and transformed into an electric signal when they strike the detector. The generated signal is amplified, recorded and converted, after reprocessing, into total ionic currents (TIC), intensities, mass spectra, relative abundances, or else extracted ion chromatograms by the computer.

A schematic diagram describing the general feature of LC-MS instrumentation is presented in figure II-3.

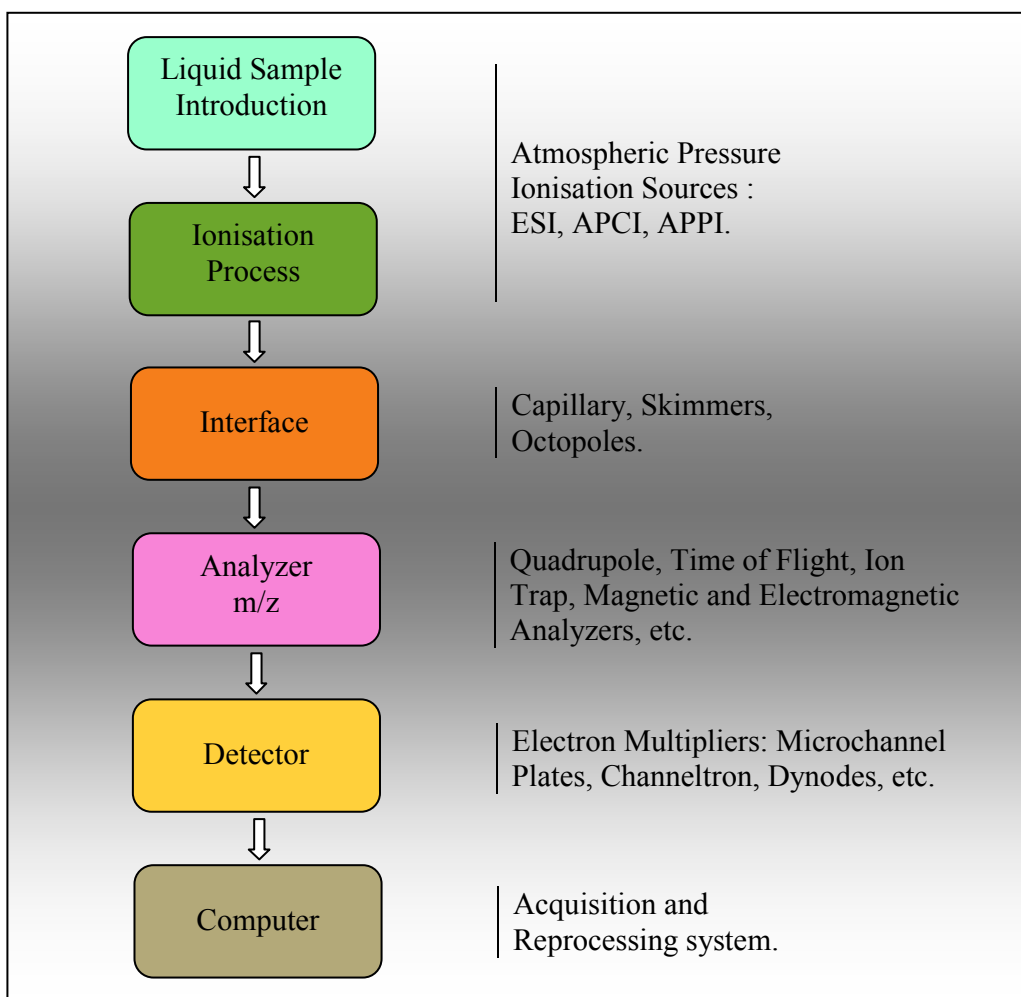


Figure II-3: Principle of LC-MS system

### II.3.2 Tandem mass spectrometry

Tandem Mass analysis, also called MS/MS analysis, is mentioned when ions produced in the ion source, are scanned across a preset  $m/z$  range and isolated as parent ions, or ions of interest, in a first mass analyzer, such as a quadrupole, before being fragmented in a collision cell. The fragmentation ions, or product ions, generated by collision are then separated and measured, as last step, in a second mass analyzer [20]. The combination of two distinct instruments in order to perform MS/MS experiments, as illustrated in figure II-4, is referred as in-space tandem mass spectrometry.

In contrast to in-space tandem mass spectrometers, in-time mass spectrometers, like ion traps, performed the selection of the precursor ion, the fragmentation process and the ion fragment measurements in an identical and unique analyzer. This functioning mode allows applying several fragmentation steps to the original ion species in order to realize  $MS^n$  experiments, where  $n$  corresponds to the number of MS stages performed [21]. MS/MS techniques are particularly recommended in quantitative analysis, when increased sensitivity is requested, for determining molecule empirical formulae or when looking for structural information of the original ion.

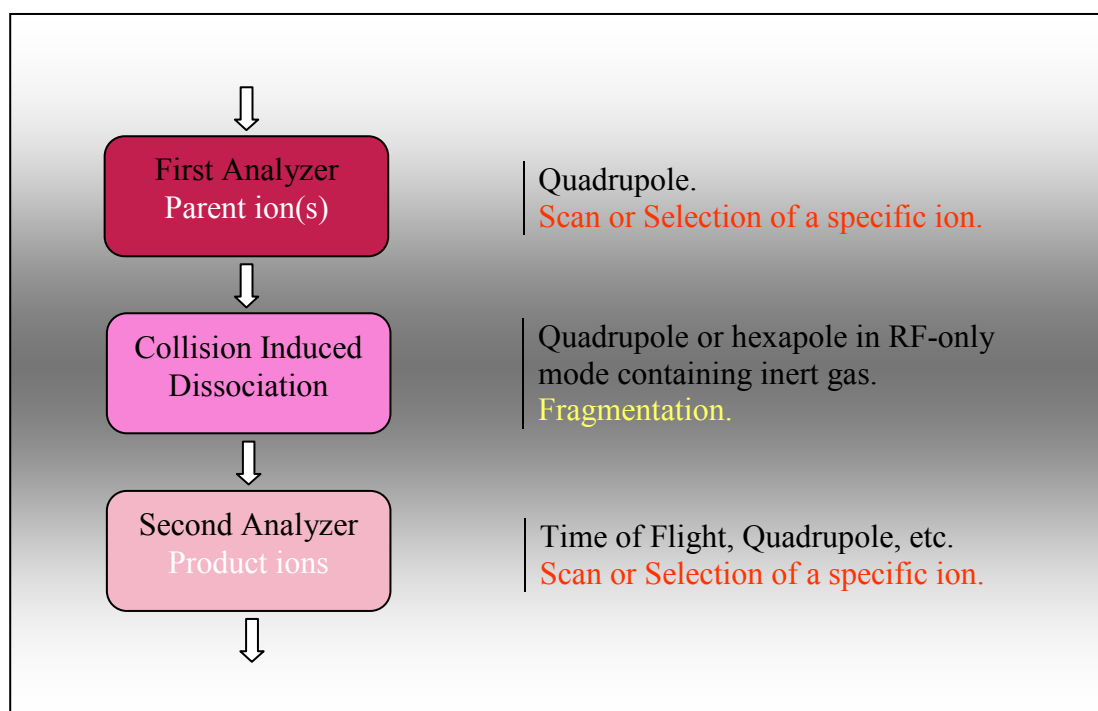


Figure II-4: Combination of two analyzers in space tandem mass spectrometry

Collision induced dissociation (CID), sometimes named Collision activated decomposition (CAD), represents the core of the tandem mass spectrometry process. The selected ion is collided with neutral molecules like inert gas molecules, Nitrogen, Argon, Xenon or Helium. The collision cell used during the experiment can be a simple quadrupole or hexapole analyzer functioning in RF-only mode, which means that all the ions are just focused along the x-axis, guided and transmitted towards the second mass analyzer, without mass discrimination.

During the collision, the kinetic energy acquired by the accelerated ions, due to the electric field corresponding to the specified collision energy, is converted into potential energy in the molecules. If this internal energy exceeds the fragmentation threshold, precursor ions will undergo bond cleavages into smaller fragments and, sometimes, molecular rearrangements, leading to the most stable ion forms [22]. The types of fragment ions depend, obviously, on the nature and the structure of the precursor ion, but also on the collision energy applied. Low energies, close to the fragmentation threshold, rather induce neutral losses, like water molecules, methanol, carbon monoxide, and carbon dioxide, for example. Higher energies lead to carbon-carbon bond breakages and more uncontrolled fragmentation processes. The resulting fragmentation pattern can be used for structural information or quantitative analysis [23].

### **II.3.3 Atmospheric pressure ionization sources**

The main inconvenience when coupling high performance liquid chromatography system to mass analyzer lies in eliminating the liquid solvent and converting the solute into gas phase ions in order to carry out mass spectrometry. The last fifteen years have seen considerable breakthrough in developing atmospheric pressure ionization sources. Emergence of reliable, robust and efficient electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) sources, and more recently atmospheric pressure photo ionization sources (APPI), has contributed to democratize the use of LC-MS applications in modern laboratories. Figure II-5 shows the theoretical applicability domains of each ionization source in function of the polarity and the molecular weight of the analytes [24].

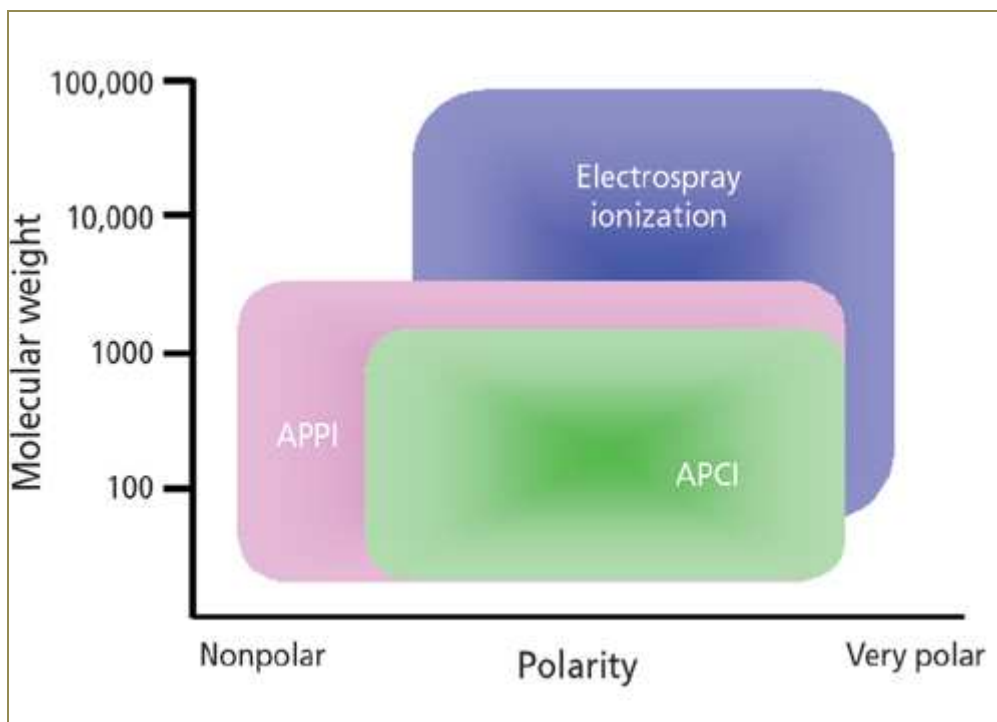


Figure II-5: Ionization range by ESI, APCI, and APPI as a function of analyte polarity and molecular weight [24]

As illustrated by the diagram, ESI and APCI sources are, in most applications, the sources of choice because of their ability to ionize a large range of compounds. ESI sources are particularly suited for the ionization of very large molecules, as well as smaller molecules. Furthermore, ESI sources, as APCI sources, offer the possibility of ionizing a wide range of compounds from very polar to less polar, when APPI shows the advantage in ionizing low-polar and non polar substances.

### II.3.3.1 Electrospray ionization source

J.B. Fenn is considered as being the inventor of ESI. He published, in 1989, an identification method of biological macromolecules based on the ionization properties of this type of source [25]. In 2002, he was rewarded with the attribution of the Nobel Prize in Chemistry for the results of his scientific researches and works. Nowadays, the electrospray ionization technique can be found in many applications in extremely varied domains and stands for the most widely used sources when analyzing polar components, like drugs, and large molecules, like peptides and proteins.



To understand the functioning of this ionization source, the detailed principle of the electrospray ionization, used in positive mode in this case, is presented in figure II-6, while a photograph of the ESI process is shown in figure II-7.

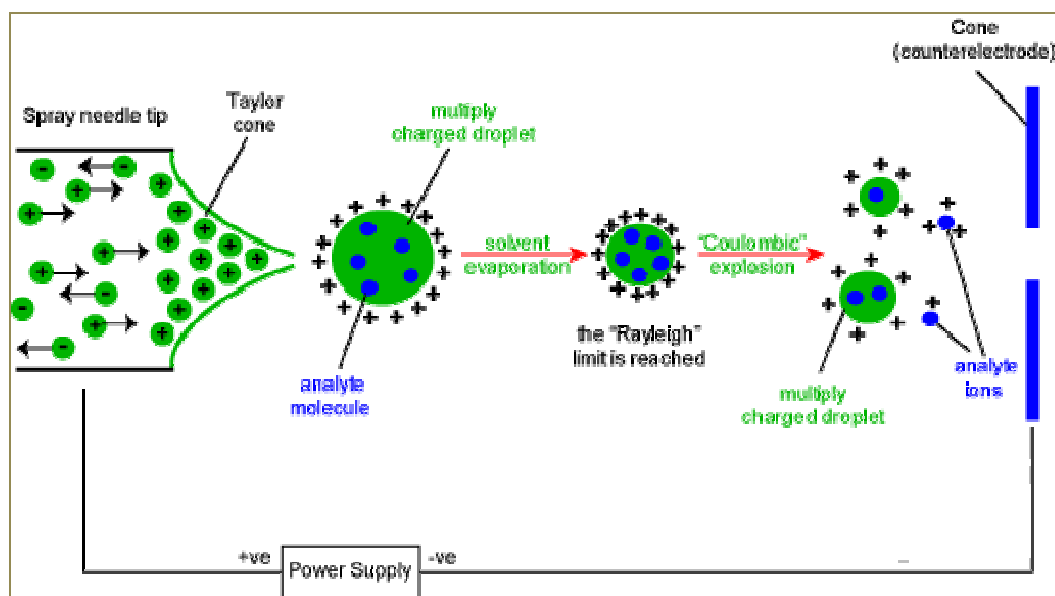


Figure II-6: Diagram of an electrospray ionization source functioning in positive mode [26]

ESI is a liquid phase ionization process which demands low flow rates, below  $1\text{mL}\cdot\text{min}^{-1}$ . The solution of the analytes is sprayed from the tip of a metalized silica capillary to which a high potential of about +4 to +6 kV, in positive mode, or -4 to -6 kV, in negative mode, is applied. Under the combined effects of the electric field and a co-axial nebulizing gas, the electrically charged liquid emerged from the capillary by forming a Taylor cone before changing shape into a fine jet, which finally disintegrates into a plume of tiny and highly charged droplets (see figure II-7).

Those fine droplets shrink then progressively by evaporation of the solvent due to another gas current, called the drying gas. The density of charge increasing dramatically on the surface of the micro-droplets, electric repulsion reaches a critical state, called the Rayleigh limit, leading to Coulomb explosion and apparition of gas phase ions. The generated ions are then accelerated towards the analyzer through the interface. Some ESI sources, called Dual spray source, are equipped with a second sprayer through which a continuous low-level introduction of a reference mass solution is operated, minimizing interferences with the analyte molecules.

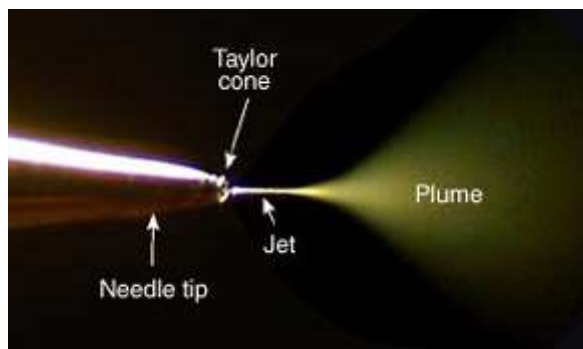


Figure II-7: Photograph of the electrospray process [27]

This type of source is particularly adapted to polar molecules. ESI is a soft ionization method for chemical analysis since it does not induce a severe fragmentation of the ionized species. It produces single charged ions and sometimes dimers, like  $(2M+H)^+$  or  $(2M-H)^-$ , and also multiple charged ions, extending considerably the mass range of the instrument. In positive mode, it generates protonated molecules  $(M+H)^+$ , and cationic molecules, like sodium adducts  $(M+Na)^+$ , ammonium adducts  $(M+NH_4)^+$  or potassium adducts  $(M+K)^+$ . In negative mode, ESI conducts to the formation of deprotonated molecules  $(M-H)^-$  and other anionic species.

### II.3.3.2 Atmospheric pressure chemical ionization source

Unlike ESI process, atmospheric pressure chemical ionization is a gas phase ionization process. It was developed in the seventies by Professor E.C. Horning and collaborators in order to hyphenate liquid chromatography to mass spectrometry [28]. The principle governing an atmospheric pressure chemical ionization source is illustrated in figure II-8 and detailed in figure II-9.

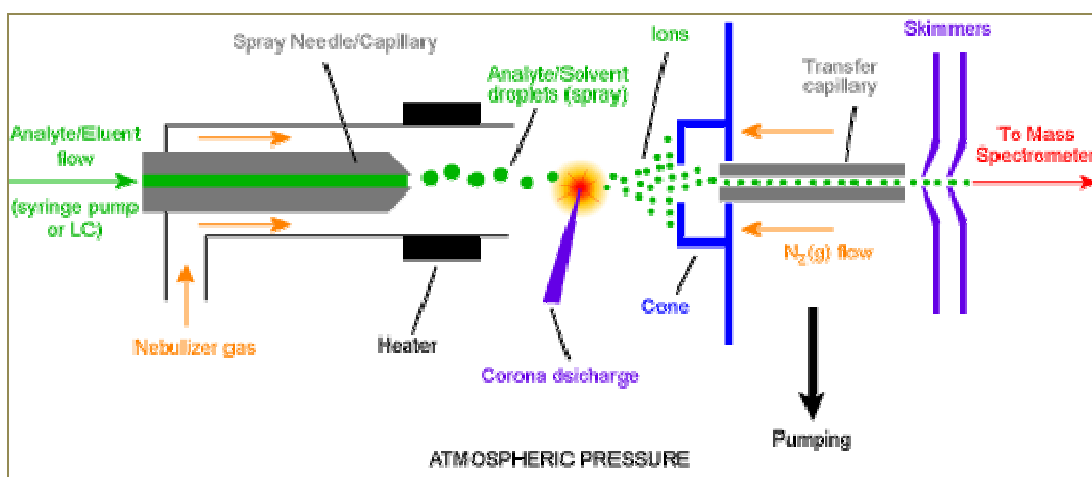


Figure II-8: Diagram of an atmospheric pressure chemical ionization source [26]

The principle of APCI is based on gas phase ion-molecule reactions. The solution containing the analyte sprayed from the tip of the pneumatic nebulizer in a fine aerosol cloud. The spray droplets are then heated to relatively high temperatures, between 100 and 400 degrees Celsius, and displaced by high flow rates of nitrogen to the region of reaction. Heating combined with nitrogen nebulization induce vaporization and desolvation of the micro-droplets, so that the reaction zone contains analyte molecules, solvent molecules, nitrogen and water vapor, oxygen and carbon dioxide. The corona needle discharge due to high potential produces electrons which generate a primary ionization plasma of  $N_2^{+}$ ,  $CO_2^{+}$ ,  $O_2^{+}$  and  $H_2O^{+}$ . Those primary ions interact with the polar molecules of the solvent and the water vapor to form reactive ions, as described in figure II-9. The high collision frequency between those reactive ions and the compounds of interest leads to the observation of a gain, in positive mode, or of a loss of proton, in negative mode. According to the respective proton affinities of the species present in the reaction zone, the proton will be transferred from the species with the lowest proton affinity to the species with the greatest proton affinity [21].

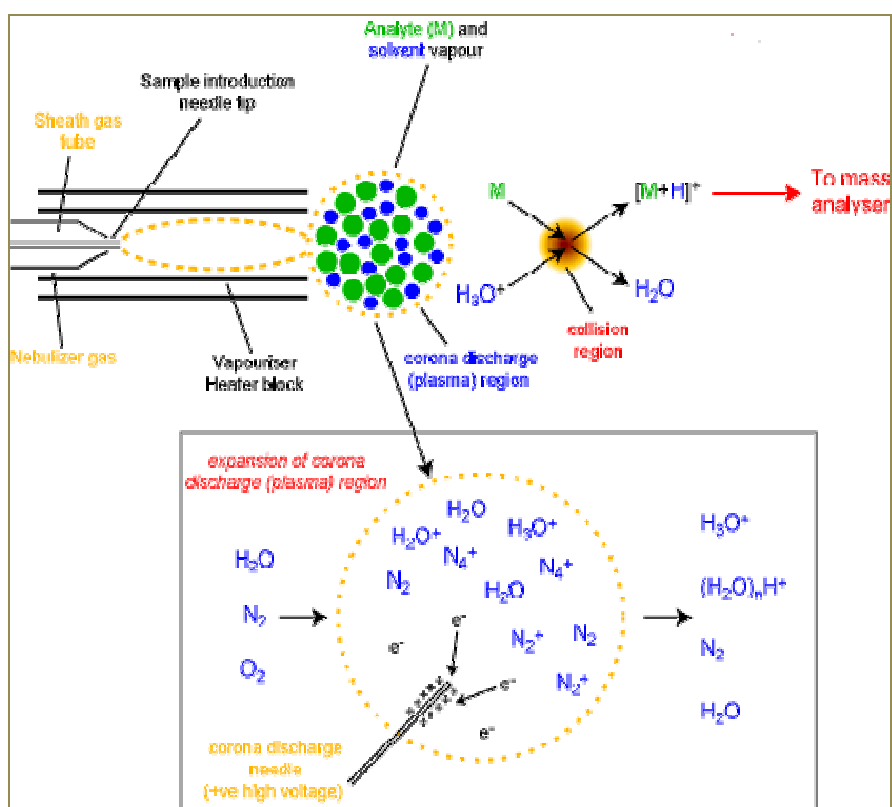


Figure II-9: Ionization mechanism in an APCI source [26]

APCI allows for flow rates from  $0.1 \text{ mL}\cdot\text{min}^{-1}$  to  $2.0 \text{ mL}\cdot\text{min}^{-1}$ . Compared to electrospray ionization, APCI is a less soft ionization technique, i.e. it generates more fragment ions relative to the parent ion. Moreover, it produces only single charged ions and is suited for less polar molecules.

### II.3.3.3 Atmospheric pressure photo-ionization source

A brief description of the atmospheric pressure photo-ionization source will be given in this paragraph despite of its non-utilization during the study. Indeed, APPI is a new emerging technology which allows to expand the range of applicability of the LC-MS instrumentation to very-low and non-polar compounds (see figure II-5). The atmospheric pressure photo-ionization source is very similar to the APCI source in its design, with the difference that solvent and analyte molecules, previously reduced in a fine spray by pneumatic nebulization and preheated, are irradiated by photons emitted from a UV lamp source, instead of electrons, to induce primary ionization, as illustrated in figure II-10. Several UV lamp sources, commercially available, provide selective ionization in regard of the energy emitted, like for instance, Krypton arc lamps (10 eV), Argon lamps (11.7 eV) or Xenon lamps (8.4 eV).

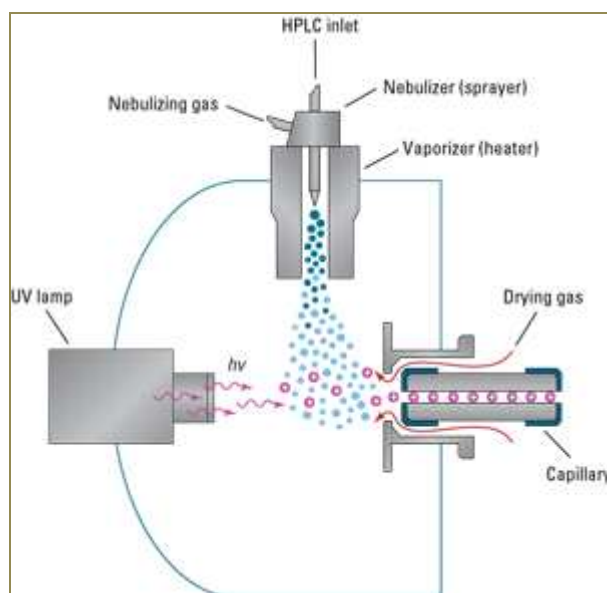
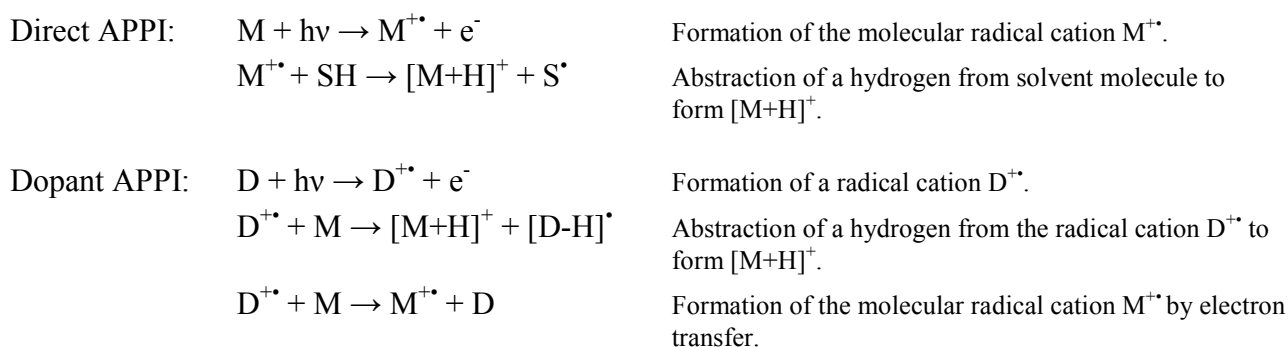


Figure II-10: Diagram of an atmospheric photo-ionization source [29]

The primary ionization can also occur by post-column addition of dopants, such as acetone or toluene. Thus, two mechanisms rule the APPI process:



#### **II.3.3.4 Atmospheric pressure high vacuum interface**

Ions are steered from the ion source to the analyzer through the interface. The interface design is essential for the atmospheric pressure ionization sources. The role of the interface consists in transferring a maximum number of ions from the ion source, where they are generated, to the analyzer, where they are separated and measured [23]. It must be noticed that only 0.01% to 1% of the ions produced in the ion source enter the analyzer prior detection. In addition, the interface permits the ion transfer from the atmospheric pressure source compartment towards the very high vacuum analyzer compartment ( $10^{-7}$  Torr), by series of ion optics like, the sample capillary, the skimmers, which are electronic lenses with very small orifices, and focusing octapole lenses.

Furthermore, a counterflow of dry and preheated nitrogen gas, at the entrance of the capillary, is used to improve the removal of solvent molecules. Moreover, the adjustment of the first skimmer voltage, or fragmentor voltage, contributes to the final ion desolvation and ion declustering within the capillary, by provoking collisions between residual gas molecules and the accelerated ions. However, if the fragmentor voltage is too high, the energy transferred to the ions during those collisions may result in a sufficient increase in their internal energy to induce fragmentation. This phenomenon is known as in-source fragmentation. The first octapole acts as an ion guide and an energy distribution homogenizer by focusing the ion beam near the x-axis before getting into the mass analyzer. Furthermore, in this region, remaining neutral molecules are pumped away by the turbomolecular pumps, which provide ultra low pressure, necessary to avoid further decomposition, direction change, or else, charge neutralization of the ions before they get into the first mass analyzer [30].

#### **II.3.4 Mass analyzers**

Breakthrough in mass analyzer technology has been observed during the last two decades, offering most relevant laboratory solutions, particularly for leading-edge applications requiring ultra trace level sensitivity. Combining a quadrupole to a TOF analyzer in a mass spectrometer provides selectivity, flexibility for collision experiments, high resolving power, accurate mass measurements, great sensitivity and speed in scan mode. The technical features of both instruments, as well as their combination, will be introduced and discussed in the following chapters.

### II.3.4.1 Single quadrupole mass analyzer

A quadrupole mass filter is made up of four strictly parallel metallic electrodes with circular or hyperbolic section [31], electrically connected together in diagonally opposite pairs. Positive and negative oscillating electrostatic fields, constituted by radiofrequency components (RF) superimposed on direct-current potentials (DC), are applied to each rod pair [32]. The complex ion trajectories within the quadrupole are illustrated by the Mathieu stability diagram. The trajectories depend on the precise voltage sets applied to the rods, and particularly to the DC to RF voltage ratios chosen, represented by linear scan lines, as shown in figure II-11.

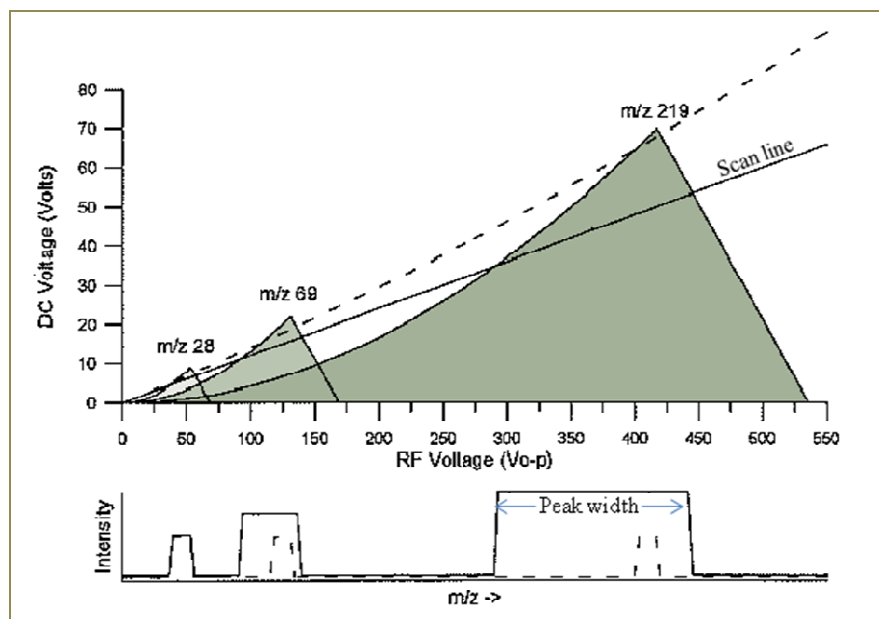


Figure II-11: Examples of Mathieu stability diagrams for three different masses (upper diagram) and corresponding mass peak widths when applying different linear scan lines (diagram below) [33]

The Mathieu stability diagram is a plot of a parameter related to the RF voltage versus a parameter related to the DC voltage. The stable trajectories, corresponding to the grey-shaded triangular areas, represent all the possible combinations of RF and DC voltages allowing the ions of a certain mass to pass through the analyzer. Unstable trajectories, outside the grey-shaded triangular areas, result in ions being neutralized by striking the rods, as illustrated in the following figure by the blue dashed line.

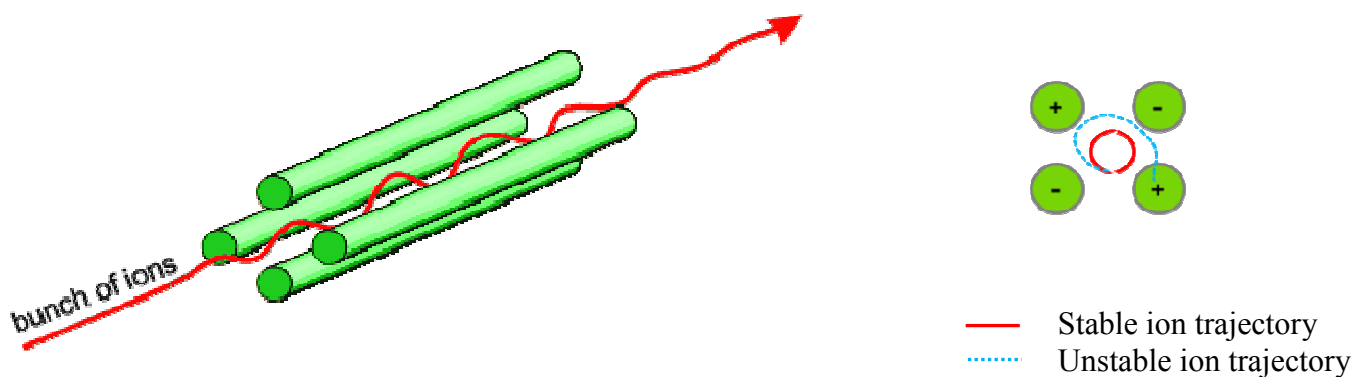


Figure II-12: Schematic diagram of ion trajectories in a quadrupole mass analyzer [34]

Varying the voltage set by increasing or decreasing the magnitude of the RF and DC voltages contributes to scan all the mass range. Similarly, changing the slope of the scan lines determines the mass peak width and the value of the resolution across the mass range. Generally, unit mass resolutions, corresponding of a decimal mass accuracy, in the hundreds of parts-per-million (ppm), are obtained with quadrupole mass filters. As the quadrupole mass filters present, on top of that, the advantages of high sensitivity, due to their elevated ion transmission capabilities, and also rapid switching between the selected ions, they are the best choice as first mass analyzer in tandem mass spectrometry analysis.

#### II.3.4.2 Hybrid quadrupole-time-of-flight mass analyzer

A hybrid quadrupole - time-of-flight mass analyzer corresponds to the combination of a quadrupole mass filter with a time-of-flight mass analyzer, both separated by a quadrupole or a hexapole analyzer, functioning in RF-only mode, in which ions undergo collisions with inert gas molecules, inducing partial or total fragmentation. A schematic of the instrumentation is shown in figure II-13.

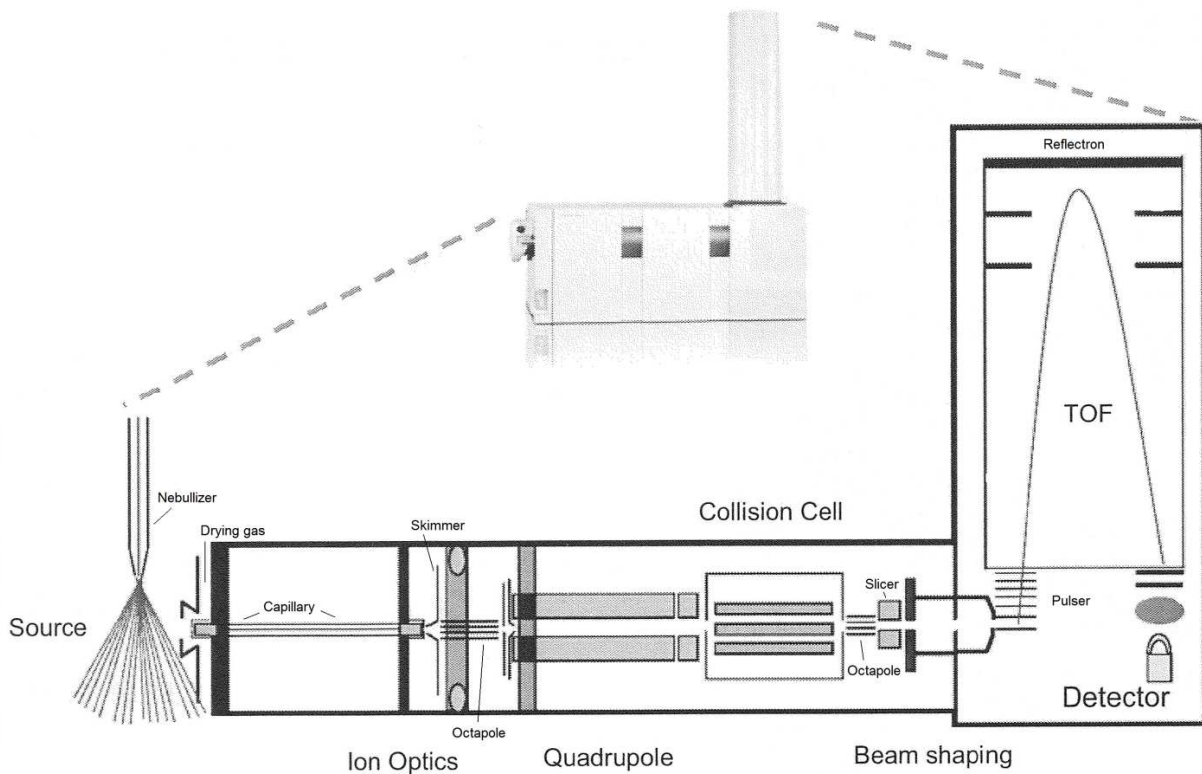


Figure II-13: Schematic of a hybrid quadrupole - time-of-flight mass analyzer [30].

Time-of-flight mass analyzers stand for the simplest mass separation devices, since the basic principle is based on the difference of velocities between ions moving in a field-free region and initially accelerated with an identical kinetic energy [35]. More precisely, ions reaching the TOF, are orthogonally and simultaneously accelerated by a high voltage pulser into the flight tube. A 10 kilovolts electric potential difference is applied every 100 microseconds, corresponding to a pulser timing of 10 kHz. The ions enter then the electric field-free zone with a kinetic energy equivalent to their potential energy due to the voltage differential applied in the pulser assembly [30].

$$E_k = \frac{1}{2} mv^2 = E_p = zeU \quad (6)$$

$$m/z = 2 eU / v^2 \quad (7)$$

Where

- e is the charge of an electron ( $e = 1.6 \cdot 10^{-19}$  C).
- z is the number of ion charges.
- U is the extraction pulse potential.
- v is the ion velocity when no electric field is applied.
- m is the ion mass.



Moreover, the ion velocity is equal to the flight path length (L) divided by the time t of the ion flight from the pulser to the detector.

$$v = L / t \quad (8)$$

As a result from equations (7) and (8), it can be concluded that the mass to charge ratio for a given ion is proportional to the square of the flight time, as expressed in the following equation:

$$m/z = 2 eUt^2 / L^2 \quad (9)$$

Therefore, by measuring precisely the time separating the acceleration pulse and the detection of the ions, it is possible to determine accurately the mass to charge ratio for each ion. From equation (9), it can be also deduced that the lightest ions are detected first, since they travel faster across the TOF analyzer than the heaviest ions.

### **II.3.4.3 High resolution and mass accuracy measurements**

Normally all the ions undergo the same acceleration energy from the extraction pulse, but in fact, slight differences in kinetic energy distribution exist, resulting in slight differences in arrival times of isobaric ions at the detector constituted of a micro-channel plate in combination with a scintillator and a photomultiplier tube [36]. In addition, all the ions of a given mass don't leave the pulser at exactly the same position, leading in a spatial distribution and tiny gaps in detector striking times. Those both physical phenomena are at the origin of peak width broadening and thus, lower resolution and mass accuracy measurements. The spatial distribution is considerably minimized by positioning a slicer at the entrance of the TOF analyzer, which shapes the ion bunch into a narrow parallel beam. This slicer is made up of a long tube ended by fine rectangular slits intended to retain and eliminate the ions getting off the horizontal axis [30]. On the other hand, the kinetic energy distribution is corrected by using a reflectron, or ion mirror, consisting in series of increasing electric fields which discriminatingly slows down and refocused the ions with same m/z, before repulsing them as a single group towards the detector, as depicted in figure II-14.

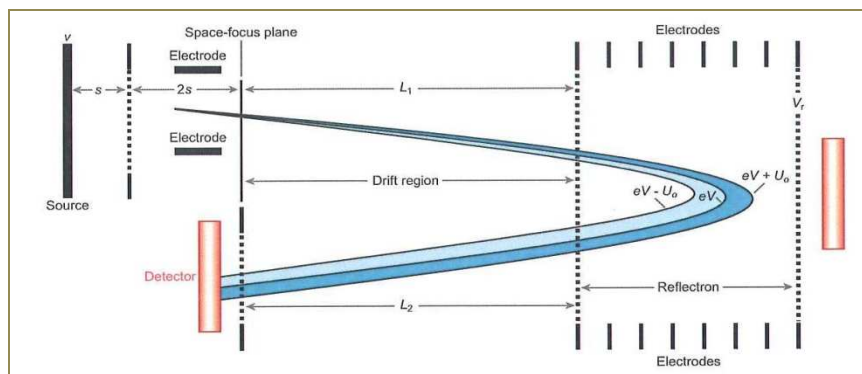


Figure II-14: Schematic of a reflectron-TOF [37].

Ions with higher kinetic energy will penetrate deeper in the reflectron and will be repelled in the same time as the slightly less rapid ions and with lower kinetic energy, so that all the isobaric ions are finally regrouped in compressed ion packets. Besides narrowing the time-of-flight distribution for each ion mass, the reflectron contributes, by reversing the direction of the ion travel, to extend the time-of-flight path length and thus, the separation time, without increasing the bulk of the flight tube [37]. As a result of the homogenization of the kinetic energy distribution by the reflectron, the peak width measured at 50% height level on the mass scale, also called full width at half maximum (FWHM), is considerably reduced and consequently, the resolution is increased, according to the definition of this latter.

$$R_s = M / \Delta M \quad (10)$$

The resolution, or resolving power, corresponds to the ability of the mass spectrometer to distinguish between two ions with close mass-to-charge ratios [21].  $\Delta M$  stands for the smallest gap between two resolved peaks at masses  $M$  and  $M + \Delta M$ , as illustrated in figure II-15.

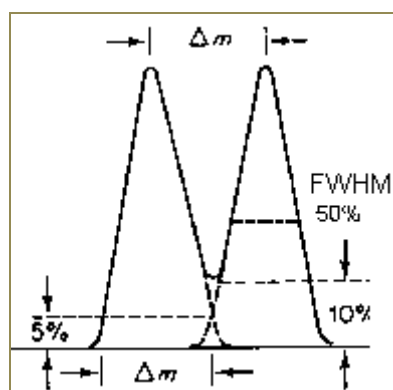


Figure II-15: Resolving power [38].

High resolution instruments, like TOF based instruments, achieve resolution up to 10,000. Such resolutions contribute to considerably narrow the peak width, allowing the determination of the peak centroid with greater precision and accuracy, so that the instrumental mass resolving power has a great incidence on the mass measurement accuracy. Accuracy, or mass error, is expressed in parts per million (ppm) and defined as the difference between the calculated mass-to-charge ratio ( $m/z_{\text{calculated}}$ ) and the measured mass-to-charge ratio ( $m/z_{\text{measured}}$ ) divided by the calculated mass-to-charge ratio [39]:

$$\text{Error (ppm)} = (m/z_{\text{measured}} - m/z_{\text{calculated}}) / m/z_{\text{calculated}} \times 10^6 \quad (11)$$

TOF mass spectrometers perform outstanding mass measurements in the milli-Dalton range (mDa), representing errors in the order of 3 to 5 ppm and allowing the determination of exact masses. This constitutes, with scan speed and extended dynamic range, one of the main advantages of QTOF instruments over unit mass resolution detectors like triple-quadrupoles and ion traps, for example. Table II-1 summarizes comparatively the performance characteristics of the principal mass analyzers available in the analytical chemistry field and illustrates the great capabilities of the quadrupole – time-of-flight.

Table II-1: Performance comparison of different mass spectrometers [30].

	Quadrupole	Ion Trap	Time-of-Flight	Time-of-Flight Reflectron	Magnetic Sector	FTMS	Quadrupole - TOF
<b>Accuracy</b>	0.01% (100 ppm)	0.01% (100 ppm)	0.02 to 0.2% (200 ppm)	0.001% (10 ppm)	<0.0005% (<5 ppm)	<0.0005% (<5 ppm)	0.001% (10 ppm)
<b>Resolution</b>	4000	4000	8000	15,000	30,000	100,000	10,000
<b>m/z Range</b>	4000	4000	>300,000	10,000	10,000	10,000	10,000
<b>Scan Speed</b>	seconds	seconds	milliseconds	milliseconds	seconds	seconds	Seconds
<b>Tandem MS</b>	MS <sup>2</sup> (triple quad)	MS <sup>n</sup>	MS	MS <sup>2</sup>	MS <sup>2</sup>	MS <sup>n</sup>	MS <sup>2</sup>
<b>Tandem MS Comments</b>	Good accuracy Good resolution Low-energy collisions	Good accuracy Good resolution Low-energy collisions	Not generally applicable	Precursor ion selection is limited to a wide mass range; growing number of applications	Limited resolution High-energy collisions	Excellent accuracy & resolution of product ions	Excellent accuracy Good resolution Low-energy collisions High sensitivity

The exact mass information obtained with the quadrupole – time-of-flight analyzer provides useful indications about the isotopic pattern and, more specifically, about the isotopic spacing, which contribute first, to the determination of the ion charge state, and second, to the prediction of a reduced number of possible empirical formulas for the investigated substance.

This propensity to eliminate unlikely or incorrect molecular formulas and to limit the choice of elemental compositions to only few options is particularly interesting and appropriate in cases of structure elucidation. Moreover, narrowing the mass window conduces to filter out co-eluting matrix interferences and competing compounds. Consequently, the chemical noise is decreased, inducing an increase of the sensitivity. Thus, QTOF technology offers significantly high sensitivity, and particularly when functioning in scan mode. For instance, QTOF demonstrates superior sensitivity compared to triple-quadrupole mass spectrometer technology when operating in full scan mode.

However, in order to maintain the accuracy of the instrument and avoid mass shifting generated by small temperature variations and vacuum or electronic instability, a continuous mass-axis calibration is performed during the analysis. This mass assignment is performed continuously using a solution containing known reference masses (see part III-1.3 for more details about those reference masses). Additionally, an operation called “tuning” is carried out every second week. “Tuning” consists in adjusting ion optics, quadrupole and time-of-flight parameters to achieve the most efficient ion transmission and the optimum signal intensity and resolution. Those adjustment operations are done automatically by the instruments.

#### **II.3.4.4 QTOF operating modes**

The quadrupole – time-of-flight mass spectrometer can operate in different modes which are the TOF mode and the product ion scan mode, comprising auto MS/MS and targeted MS/MS functions. In TOF mode the quadrupole works in total transmission ion mode, which means that no collision energy is applied in the collision cell. All the ions are focused near the axis, through the quadrupole and the collision cell, and transmitted from the interface to the time-of-flight mass detector without undergoing any fragmentation. Then the TOF analyzes the ions in scan mode and provides the MS spectra.

In targeted MS/MS analysis, the quadrupole works in selected ion monitoring mode (SIM mode). Specific precursor ions, as defined in the “target mass list” table, are isolated by the quadrupole and transmitted towards the collision chamber where they are fragmented. The fragment ions generated are analyzed by the TOF in scan mode, providing MS/MS spectra. This operating mode is particularly adapted for quantitative analysis, identification and structural elucidation of known compounds [30].

In auto MS/MS analysis, the instrument performs analysis in SIM mode. Precursor ions are chosen by the instrument among the most abundant ions, according to criteria previously entered, like the maximum number of ions to consider, the charge state, the absolute and the relative threshold values, and the preferred/exclude ions table. The collision cell generates fragment ions by colliding the selected precursor ions with nitrogen and the TOF analyzes the fragment ions in scan mode and provides the MS/MS spectra. This operating mode is particularly adapted when investigating the identification and the structural composition of unknown compounds [30].

The amount of information gathered during a series of tests containing many samples, will be so substantial, that a tool for data processing may be necessary, or even indispensable, to interpret the results and facilitate decision support. Principal component analysis and hierarchical clustering analysis are the data exploratory analysis methods of choice used in this study. The next chapter is consecrated to the explanatory description of these two useful statistical techniques.

## **II.4 Multivariate data analysis**

Multivariate data analysis, also called chemometrics, refers to extremely powerful statistical decision tools like principal component analysis or hierarchical clustering analysis, for example. Nowadays largely applied in the field of modern analytical chemistry, the term “chemometrics” was first introduced in 1972, by Swante Wold, a Swedish professor of organic chemistry. Those kinds of data analysis techniques are modeling sciences based on sophisticated mathematical methods, and particularly matrix calculations, with the aim of retrieving the significant information from a signal [40]. Indeed, an instrumental signal results in a combination of two components, firstly, a descriptive information, which can be assimilated to variation, specific and characteristic of the signal, and second, a residual part, the noise. Thus, the major interest of multivariate data analysis consists in separating the information from the noise and consequently, simplifying the interpretation of complex and huge datasets, helping to make insightful decision.

### II.4.1 Principal component analysis

Principal component analysis is without contest one of the actual most largely used multivariate exploratory data analysis techniques in modern laboratories. The success of PCA is linked to its ability to reduce the complexity of large datasets, characterized by high dimensionality, into simplest but significant information with smaller dimensionality and consequently, easier to work out.

In practical terms, PCA consists first in transforming the original data matrix constituted in “n” observations or samples, as rows, and “k” variables or measurements, as columns, into a covariance matrix  $C$  or  $\text{Cov}(X,Y)$ . The covariance is a measure of the simultaneous variation of two random variables,  $X$  and  $Y$  for example. It corresponds to the summation of the differences between the  $X_i$  values and the mean of  $X$  multiplied by the differences between the  $Y_i$  values and the mean of  $Y$ , divided by the number of observations minus 1, as expressed by equation (12).

$$C = \text{Cov}(X,Y) = 1/(n-1) \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (12)$$

The covariance matrix  $C$  is then transformed into a diagonal matrix, the matrix of eigenvalues  $\lambda_i$ , and the related matrix of eigenvectors  $v_i$ , as expressed by equation (13):

$$C v_i = \lambda_i v_i \quad (13)$$

When the eigenvalues of the covariance matrix  $C$  are the solutions of the following equation (14):

$$\det [C - \lambda I] = 0 \quad (14)$$

Where

| I is the identity matrix.  
|  $\det [ ]$  stands for the determinant of the matrix.

Eigenvalues and eigenvectors are closely linked. Eigenvalues denote the variability within the corresponding eigenvectors. Eigenvectors are called principal component (PC) and there are as many principal components as dimensions in the original matrix, generally, only few of them are sufficient to describe significantly the relationships among the data.

Accordingly, if only two or three principal components are considered, representing the directions with maximum variability, the original dimensionality of the dataset will be reduced to the number of PC chosen, simplifying consequently the data investigation. The eigenvector with the highest eigenvalue represents the first principal component and characterizes the largest variation in the dataset as shown in figure II-16.

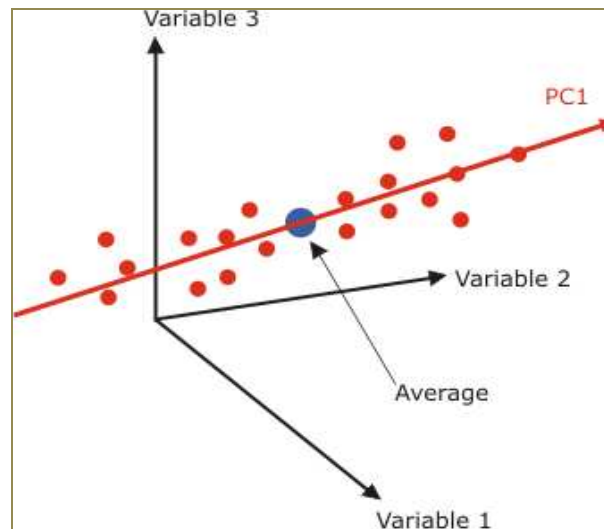


Figure II-16: Variance explained by the first principal component [41]

The second principal component stands for the eigenvector with the second largest eigenvalue. This component is of lesser significance and explains lesser dispersion as illustrated in figure II-17. The second principal component is orthogonal to the first principal component, so that both constitute the new axes of a projection plane.

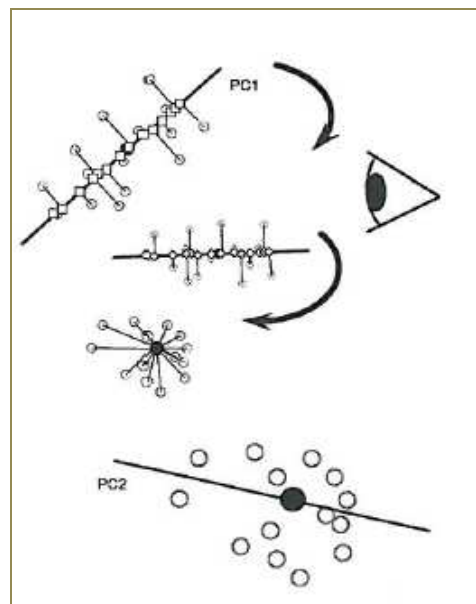


Figure II-17: Variance explained by the second principal component [42]

In the plane constituted by both principal components, the objects will be assigned with new coordinates called the scores corresponding to the distance from the mean along the axes PC1 and PC2. The loading is another factor useful to analyze the influence of a variable in the model. It is measured by the cosine of the angle between the observation and the axis, and indicates the importance of the link between the variable and the PC. This means that a high value of loading indicates a high impact of the variable on the model. Examples of score plots are given in figure II-18: at the top, a two dimensional graphics PC1 versus PC2, and at the bottom, the corresponding three dimensional graphics, both illustrating the relationships between raw materials and finished products originated from five different API providers present on the French market. The variation explained by the principal components, as well as the Hotelling T<sup>2</sup> ellipse, are reported on the graphs. The Hotelling T<sup>2</sup> ellipse is the 95% confidence region which enables to reveal outliers.

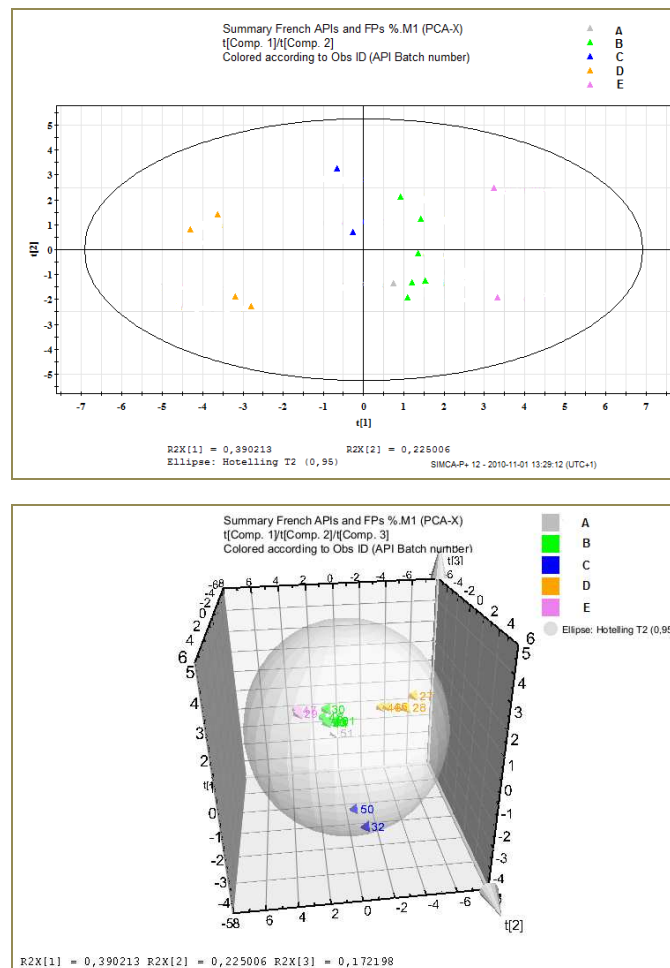


Figure II-18: Score plots of principal component 1 versus principal component 2 (top graph) and related scatter plots of principal components 1, 2, 3 (graph below) describing the relationships between raw materials and finished products originated from five different API providers (A, B, C, D and E) present on the French market



PCA is a linear projection method which constitutes the ideal means to spot trends and correlations between samples. It allows to detect outliers and to identify patterns and groups among all individual points from the datasets. However, the principal advantage remains its capacity to provide a graphical representation of the data structure without any loss of essential information. The visualization in a two or three dimensional space of multivariate elements contributes to facilitate the comprehension and the interpretation of the data correlations. Another interesting multivariate data analysis tool applied during this project is the hierarchical clustering analysis. It will be briefly developed hereafter.

### II.4.2 Hierarchical clustering analysis

Hierarchical clustering analysis is complementary to the principal component analysis. It is a convenient tool to demonstrate and understand the grouping of observations in regard to their similarities and singular characteristics. HCA provides a graphical inspection of the relationships between large amounts of data, as PCA. However, the generation of clusters is based on the Euclidean distance between the objects. Basically, HCA starts with as many clusters as there are observations. Then the two closest observations will be regrouped together in a same cluster. Afterwards the two closest clusters or objects are merged again, and so on, until only one cluster remains. The results are displayed as a dendrogram plot, also called tree diagram, as depicted in figure II-19.

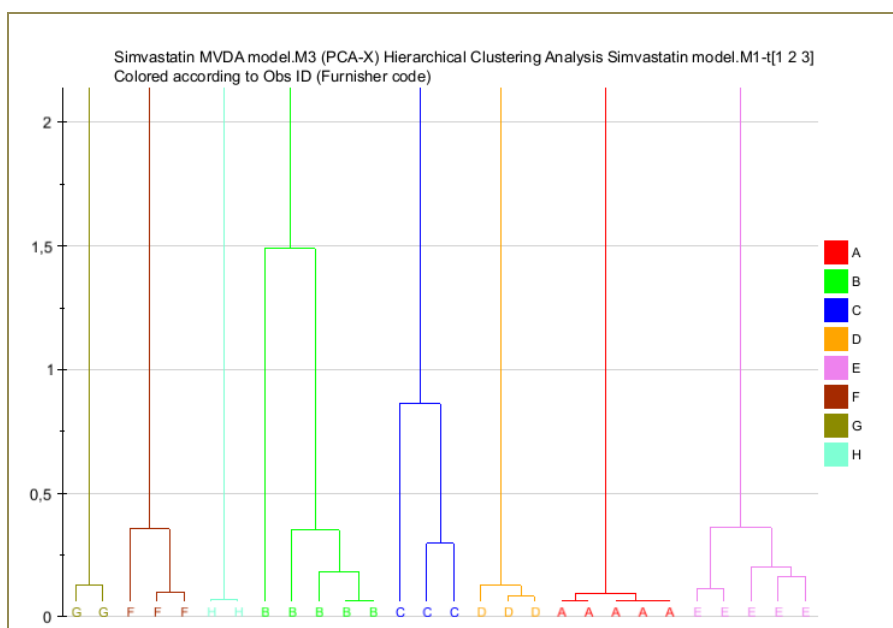


Figure II-19: Example of a dendrogram plot

The plot shows the different clusters as a function of the vertical coordinate representing the distance between clusters, so that the higher the bars and the higher the distance between the clusters. Now, from the variation magnitude in the dataset depends the distance between the samples. Therefore, the dispersion within groups can be assessed by the height of the bars.

In this study, the powerful data treatment capacity of PCA and HCA were used in combination with LC-MS impurity fingerprinting with the aim to determine the origin of raw materials and finished products. The experimental development and the results are presented in the next chapter.

### **III. Application to simvastatin and related substances in order to discriminate between different provider origins, routes of synthesis or manufacturing areas**

The monitoring of drug substance impurities constitutes, among others, an extremely important challenge for ensuring an adequate quality for users and guaranteeing the best public health protection. For that purpose, specific monographs on chemical substances for pharmaceutical use are described in the European Pharmacopoeia (Ph. Eur.). These monographs are regularly verified, improved and revised. Testing methods and acceptance criteria are given for specified and unspecified impurities. For example, thresholds for reporting, identification and qualification are required in regard to safety at the maximum daily dose, as defined in the European Pharmacopoeia general monograph “Substances for pharmaceutical use (2034)” (see appendix H).

The organic impurities may originate from degradation processes and/or from routes of synthesis, including either by-products, remaining intermediates or chiral impurities. Degradation products arise from particular environmental conditions like pH, heat, water, light and oxidation, while synthesis by-products arise from minor side reactions of starting materials and intermediates with reagents. Formation of dimers, for example, may occur during the chemical synthesis. In the same way, reactions between an early stage intermediate and a later stage reagent may also take place, leading to by-products. Therefore, the chromatographic impurities profiles are unique and specific to each source of active pharmaceutical ingredients, and consequently, may be used to characterize and identify them.

The objective of this study consisted in exploiting all the outstanding performances, in terms of high sensitivity and specificity, of high performance liquid chromatography coupled to mass spectrometry in tandem using a hybrid quadrupole – time-of-flight analyzer in order to establish impurity profiling of API, in both raw materials and finished products, allowing, in conjunction with multivariate data analysis, the discrimination between their origins, synthetic routes or production sites.

The drug substance simvastatin was chosen as test molecule to evaluate this new API generic classification method. The major reason is that simvastatin is commercialized on the Swedish and French markets under a large number of pharmaceutical formulations and originating from numerous manufacturers. And above all, most of these manufacturers call for several active pharmaceutical ingredient furnishers.

Simvastatin (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetrahydro-2H-pyran-2-yl]ethyl] - 3,7-dimethyl -1,2,3,7,8,8a-hexahydro naphthalen-1-yl-2,2-dimethylbutanoate (figure III-1), is a lipid regulating agent belonging to the family of statins and successfully employed in the treatment of hypercholesterolaemia [43].

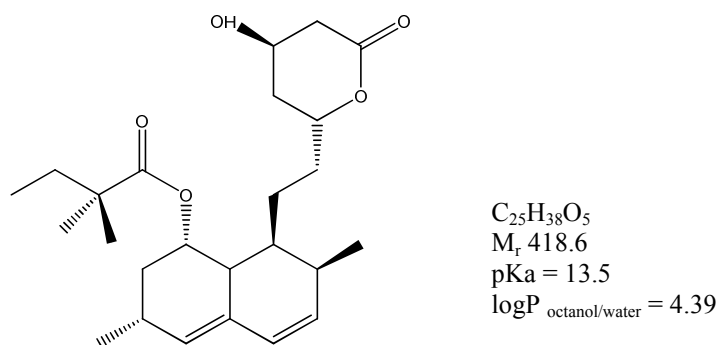


Figure III-1: Molecular representation, empirical formula, molecular weight, pKa and  $\log P_{\text{octanol/water}}$  partition coefficient of simvastatin

Once assimilated in the organism, simvastatin is transformed into its active metabolite form through a rapid hydrolysis of the lactone ring, leading to simvastatin hydroxy acid (Ph. Eur. impurity A). This metabolite is a specific competitive inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A reductase (hMG-CoA reductase), an early stage rate-determining liver enzyme taking place in cholesterol endogenous synthesis [44], as described in figure III-2.

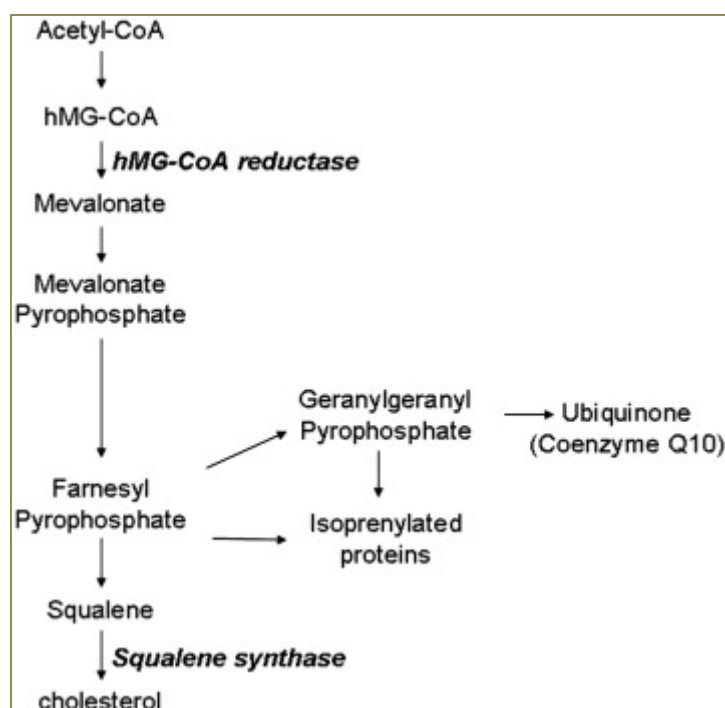
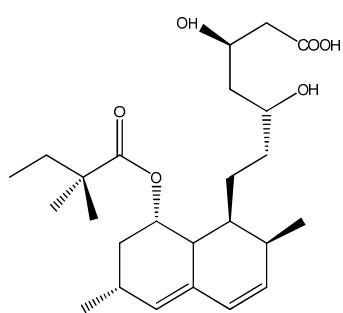


Figure III-2: Cholesterol endogenous synthesis pathway [9]

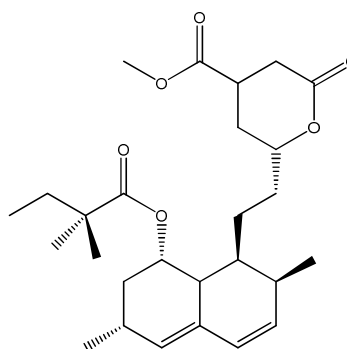
### III.1 Enhanced impurity profiling of simvastatin by LC (ESI+) MS/MS QTOF

The development of the test method was aimed to establish the API impurities fingerprinting and was based on the monograph described in the European Pharmacopoeia, 7<sup>th</sup> edition, for simvastatin. The proposed Pharmacopoeia method for the determination of simvastatin and related substances consists in a high performance liquid chromatography with an ultraviolet spectrophotometric detection at 238 nm. Seven impurities among, simvastatin hydroxy acid (Ph.Eur. impurity A), simvastatin methyl ester (Ph. Eur. impurity B), dehydro simvastatin (Ph. Eur. impurity C), simvastatin dimer (Ph. Eur. impurity D), lovastatin and epilovastatin, which are stereoisomers (respectively Ph. Eur. impurities E and F), and Ph. Eur. impurity G are reported in the simvastatin monograph [45]. Figure III-3 sums up the molecular representation, the empirical formula, the molecular weight, the estimated pKa and calculated  $\log P_{\text{octanol/water}}$  partition coefficient for each specified impurity.



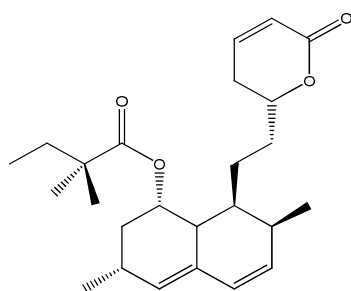
$C_{25}H_{40}O_6$   
 $M_r$  436.3  
 $pK_a = 4.31$   
 $\log P_{\text{octanol/water}} = 3.85$

Ph. Eur. Impurity A  
Simvastatin hydroxy acid



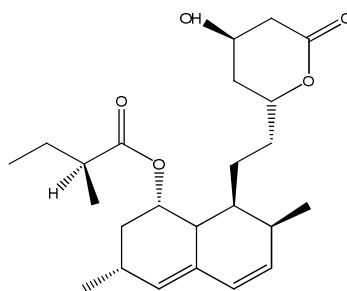
$C_{27}H_{40}O_6$   
 $M_r$  460.3  
 $pK_a \approx 13.5$   
 $\log P_{\text{octanol/water}} = 4.94$

Ph. Eur. Impurity B  
Simvastatin methyl ester



$C_{25}H_{36}O_4$   
 $M_r$  400.3  
 $pK_a \approx 13.5$   
 $\log P_{\text{octanol/water}} = 5.5$

Ph. Eur. Impurity C  
Dehydro Simvastatin



$C_{24}H_{36}O_5$   
 $M_r$  404.3  
 $pK_a \approx 13.5$   
 $\log P_{\text{octanol/water}} = 3.68$

Ph. Eur. Impurity E  
Lovastatin

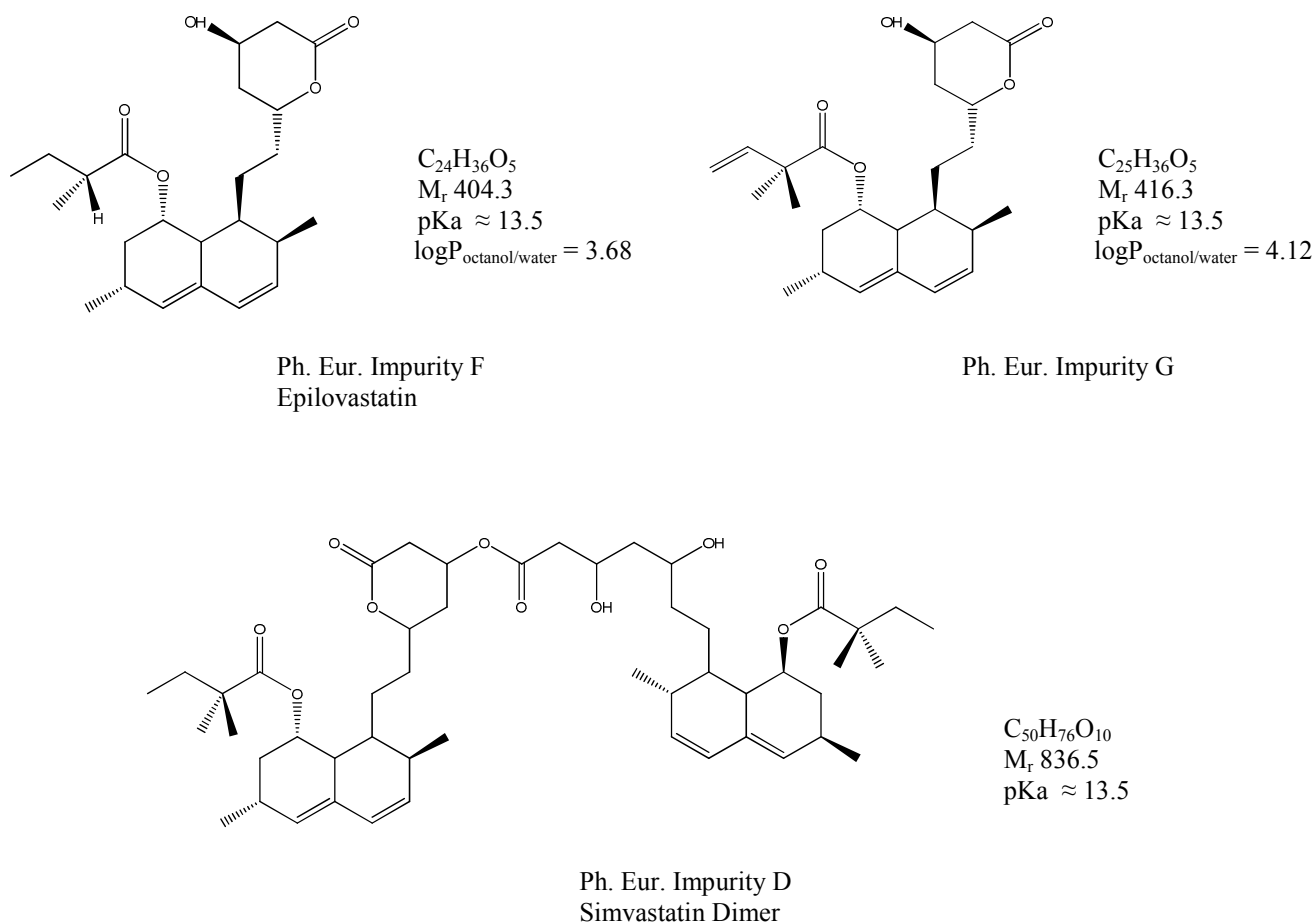


Figure III-3: Molecular representation, empirical formula, molecular weight, estimated  $pK_a$  and  $\log P_{\text{octanol/water}}$  partition coefficient of simvastatin specified impurities

In the analytical method described in the European Pharmacopoeia, the separation of simvastatin and related substances is performed on a 33 mm x 4.6 mm end-capped octadecyl-bonded silica column packed with 3 $\mu$ m particles. A Perkin Elmer Pecosphere cartridge is proposed as chromatographic column in the EDQM “Knowledge Database”. Injection volume is set to 5 $\mu$ L. The binary gradient elution corresponds to a mix of 50 volumes of acetonitrile and 50 volumes of a 0.1% phosphoric acid solution, as mobile phase A, and a 0.1% phosphoric acid solution in acetonitrile as mobile phase B, at a flow rate of 3.0 mL.min<sup>-1</sup>. Gradient conditions are reported in table III-1 herein after.

Table III-1: Gradient conditions reported in the European Pharmacopoeia monograph on simvastatin (7<sup>th</sup> edition) [45]

Time (min)	Mobile phase A (per cent V/V)	Mobile phase B (per cent V/V)
0 - 4.5	100	0
4.5 - 4.6	100 → 95	0 → 5
4.6 - 8.0	95 → 25	5 → 75
8.0 - 11.5	25	75

The chromatogram obtained with these monograph's chromatographic conditions is represented in figure III-4. The chromatogram highlights the lack of selectivity of the method employed due to the presence of several co-eluting peaks. Indeed, the separation of two pairs of impurities, corresponding to Ph. Eur. impurities E and F, and Ph. Eur. impurities B and C, is not effective.

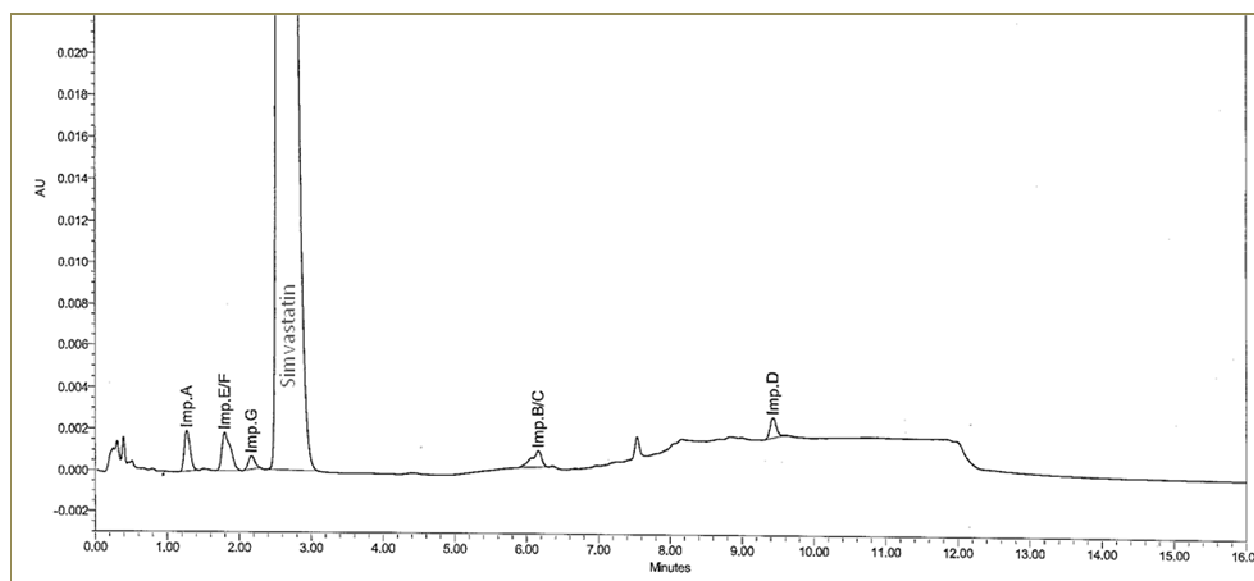


Figure III-4: Typical UV-chromatogram of a mixture of simvastatin and its specified impurities (excerpt from EDQM Lab report PA/PH/LAB 10A (08) 30 - study 5446, July 2008)

Firstly, the development of the new LC-MS analytical method had to take into account the necessity of using more “MS-friendly” chromatographic conditions. The strategy consisted in adapting the flow-rate values between 0.1 and 1.0 mL.min<sup>-1</sup>, by reducing the internal diameter of the column, and in replacing non-volatile buffers, as phosphoric acid, with volatile buffers such as formic acid.

Secondly, the method development was focused on the resolution improvement between the impurities constituting the both critical pairs, Ph. Eur. impurities E and F, on one hand, and Ph. Eur. impurities B and C, on another hand, by using a more selective packing material for the analytical column and by changing the gradient conditions. Accordingly, the optimization of the chromatographic separation was conducted by adapting the liquid chromatographic system to suitable non damaging conditions for the mass spectrometer and emphasizing the quality of the separation. Numerous parameters were expected to influence the selectivity and the retention performance of the method, like for instance, the column efficiency, the column temperature, effects of the mobile phase organic strength, the buffer ionic strength, the mobile phase pH or the composition of the sample eluting solvent. All those factors were investigated and tested with the objective to achieve appropriate separations in a reasonable time scale. Investigation results are presented and discussed in the next paragraphs.

### **III.1.1 Chromatographic system optimization for an efficient separation of simvastatin and related substances**

As suggested by the Purnell relation (cf. equation (2) paragraph II.2) the ability to control parameters like the selectivity " $\alpha$ ", the retention factor " $k$ " and the efficiency, or N-term, will critically affect the column resolving power. Selectivity and retention factor predominantly depend on factors related to the nature of the molecular interactions between the analyte and both, the mobile phase and the stationary phase, i.e. parameters such as the column temperature, pH value of the mobile phase and gradient mode elution, etc. The efficiency of the separation is governed more particularly by the mobile phase flow rate, the column length and the packing materials' characteristics, like particle size and particle size distribution. All the parameters previously listed, which greatly affect the chromatographic resolution, were subjected to enhancement and development in this study.

#### **III.1.1.1 Choice of the analytical column**

Column choice is the cornerstone step in successfully improving and optimizing chromatographic separation methods. Simvastatin and related substances are extremely weak acids with pKa values around 13.5, except for impurity A which is characterized by a pKa value of 4.3.



Therefore, interaction mechanisms between these compounds and the mobile phase and the stationary phase, are correlated to their hydrophobicity proprieties. The  $\log P_{\text{octanol/water}}$  partition coefficient is a suitable indicator to estimate the solubility characteristics of a substance and is therefore helpful in the estimation of the elution order of organic molecules by reversed phase liquid chromatography.  $P_{\text{octanol/water}}$  is defined as the ratio of the concentration of the molecule neutral form in octanol divided by the concentration of the molecule neutral form in water.

$$P_{\text{octanol/water}} = \frac{[\text{neutral species}]_{\text{in octanol}}}{[\text{neutral species}]_{\text{in water}}} \quad (12)$$

The  $\log P_{\text{octanol/water}}$  values were calculated using ChemDraw Ultra version 11.0 for simvastatin and related impurities. They are reported in figures III-1 and III-3, as well as in appendix A.

Instead of the Perkin Elmer Pecosphere cartridge (33 mm x 4.6 mm – 3 $\mu\text{m}$ ) recommended in the European Pharmacopoeia monograph (7<sup>th</sup> edition), our choice went to a Kinetex<sup>TM</sup> C18 column (50 mm x 2.1 mm - 2.6  $\mu\text{m}$ ) in order to improve the selectivity of the analytical method. This choice was dictated by the performances inherent in this non conventional column [46-49]. Indeed, Kinetex<sup>TM</sup> columns are filled with partially porous particles made up of a solid silica core, with a diameter of 1.9  $\mu\text{m}$ , and coated with a 0.35  $\mu\text{m}$  thick permeable shell. As the fused core is non porous and impermeable to analytes, these latter cannot penetrate deeply into the particles. Consequently, diffusion path is considerably shortened during the migration of the analytes through the column matrix. This feature results in faster mass transfer kinetics between the mobile phase and the stationary phase and contributes to lower the C-term in the Van Deemter equation. The Van Deemter equation stands for the expression of the height equivalent to a theoretical plate (HETP), expressed in  $\mu\text{m}$ , versus the linear velocity (u), expressed in  $\text{mm}\cdot\text{s}^{-1}$ , as stated hereafter:

$$\text{HETP} = A + B/u + Cu \quad (13)$$

$$= L/N \quad (14)$$

Where

- A corresponds to eddy diffusion.
- B corresponds to longitudinal diffusion.
- C corresponds to mass transfer resistance.
- L is the column length.
- N is the column efficiency.
- u is the linear velocity.

Faster mass transfer kinetics induces lower HETP values and, therefore, increased efficiency and chromatographic resolution. Moreover, homogenous particle size distribution of the packing material contributes to reduce the eddy diffusion (the A-term in the Van Deemter equation), so that the column efficiency is accordingly improved. Figure III-5 shows performance comparison between Kinetex™ column and fully porous sub-2µm and 3µm particle columns.

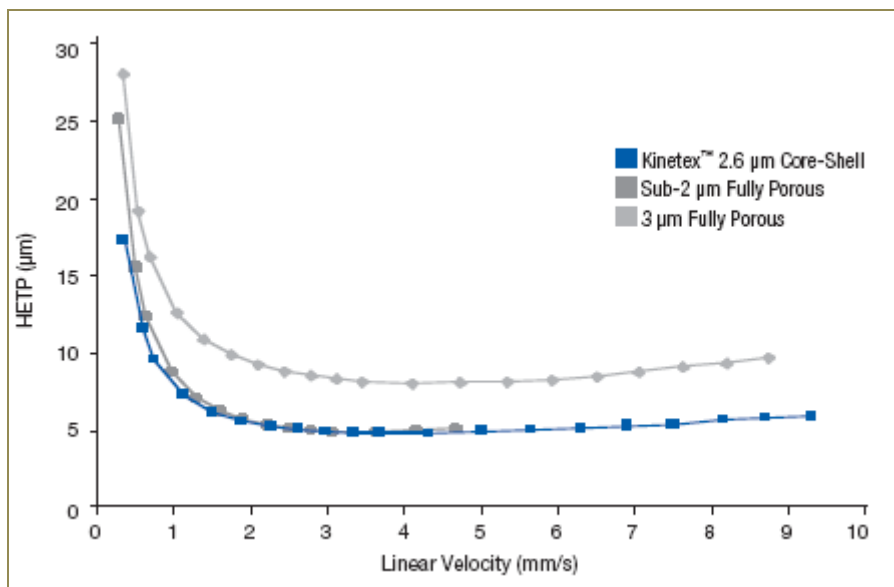


Figure III-5: Performance of Kinetex™ Core-Shell particles compared to fully porous sub-2µm and 3µm particles [46]

In this kind of diagram representing Van Deemter curves, the lower is the plate height the higher are the efficiency and the resolution of the column. The figure clearly suggests that the efficiency of fused core columns is quite equivalent to the efficiency of columns filled with sub-2µm fully porous particles and significantly better than the efficiency of columns filled with 3µm fully porous particles, and this, over a wider range of linear rate. Indeed, a further advantage with that type of columns is that, according to the expression of the backpressure, as suggested by Darcy law (see following equation) [50], Kinetex™ columns, despite of small particle size, do not induce very high backpressure, due to lesser flow resistance, and can be therefore easily used with traditional HPLC instrumentations.

$$\Delta P = \eta L u / K^0 \quad (15)$$

$$= \eta L u \phi / d_p^2 \quad (16)$$

Where

$\Delta P$  is the backpressure.  
 $\eta$  is the mobile phase viscosity.  
 $L$  is the column length.  
 $u$  is the linear velocity.  
 $K^0$  is the column permeability.  
 $\phi$  is the flow resistance.  
 $d_p$  is the average particle size.

Nevertheless, column length was limited to 50 mm in order to minimize the analysis time, solvent consumption and waste generation, and above all, to be consistent with the pressure limitations of the instrumentation to 600 bars. Another advantage of the Kinetex™ Core-Shell technology was that band broadening of the peaks was narrowed. Reducing the band broadening led to significantly sharpen peak shapes, with an increase in their height and consequently an improvement in the method sensitivity.

The next step in the method development concerned the optimization of the chromatographic separation by varying the selectivity of the mobile phase system. Experiments on parameters affecting the selectivity and the sensitivity of the method were performed demonstrating the importance of the mobile phase buffer ionic strength and the role of the mobile phase organic strength. Other essential factors such as the influence of the mobile phase pH and the effect of the temperature of the column were also investigated in this part of the study. First the adjustment of the flow rate will be discussed, aiming to adapt the high performance liquid chromatography characteristics to suitable mass spectrometer conditions.

### III.1.1.2 Flow rate adjustment

The change of a column with an internal diameter of 4.6 mm to a narrow bore column with an internal diameter of 2.1 mm, needed an adjustment of the flow rate in accordance with the following equation [51]:

$$F/d_c^2 = u \pi \varepsilon / 4 = \text{constant} \quad (17)$$

Where

$F$  represents the flow rate.  
 $d_c$  is the column internal diameter.  
 $u$  is the linear velocity.  
 $\varepsilon$  corresponds to the column porosity.

So that 
$$F_2 = F_1 \times d_{c2}^2 / d_{c1}^2 \quad (18)$$

When

$$\left| \begin{array}{l} F_1 = \text{the original method flow rate (3 mL}\cdot\text{min}^{-1}\text{).} \\ d_{c1} = \text{the original column internal diameter (4.6 mm).} \\ d_{c2} = \text{the new column internal diameter (2.1 mm).} \end{array} \right.$$

Accordingly, the new flow rate was set at a value of 0.5 mL $\cdot$ min<sup>-1</sup>, compatible with the technical limitations of the electrospray ionization source requiring flow rates in the range of 0.1 mL $\cdot$ min<sup>-1</sup> to 1.0 mL $\cdot$ min<sup>-1</sup>. Working with higher flow rate values than 1.0 mL $\cdot$ min<sup>-1</sup> could severely damage the mass analyzer by clogging and blocking the capillary, provoking indubitably the failure of expensive machine parts. Moreover, correct flow rate setting is not only essential to avoid material damage but also to maximize the desolvation of the droplets in the mass spectrometer spray chamber and, by way of consequence, the number of ions generated and the signal intensity.

### III.1.1.3 Impact of the mobile phase buffer ionic strength

Selection and concentration of the mobile phase buffer are important factors in LC-MS, especially of ionisable molecules. For example, formate and acetate buffers favor the formation of charged species while non-volatile phosphate and sulfate buffers induce ion suppression by forming ion pairs. Furthermore, the selection of the mobile phase buffer characterizes the pH zone where the interaction mechanisms are obtained, and the concentration determines the buffer capacity of the solvent. The buffer capacity is the propriety of a solution to resist to small addition of acids or basis without alteration of its pH value [52]. Controlling the pH of the mobile phase allows to determine the selectivity of the chromatographic method and also to avoid strong modifications in retention times. However, mobile phase ionic strength has a huge influence on the response of the mass spectrometer. Indeed, buffers and other additives impact the ionization process in the ion source. Consequently, various experimentations were undertaken in order to highlight the role of the buffer concentration in the ion detection.

Figure III-6 shows the total ionic chromatograms (TIC) obtained after injection into the chromatographic system, at different formic acid concentrations, of 2 $\mu$ L of simvastatin for peak identification chemical substance reference solution. The mass spectrometer signals were studied for compounds like simvastatin, Ph. Eur. impurities A, E, F and G, and unknown at  $m/z = 391.2479$  (cf paragraph III.1.4.2 – “Identification of new impurities by LC-MS/MS” for characterization), for various mobile phases containing formic acid ranging from 0.001% to 0.1%. Total ionic chromatograms correspond to intensities, expressed in number of counts detected by the mass spectrometer, *versus* the acquisition time, in minutes. Peak areas and corresponding signal to noise ratio are reported in the graphics.

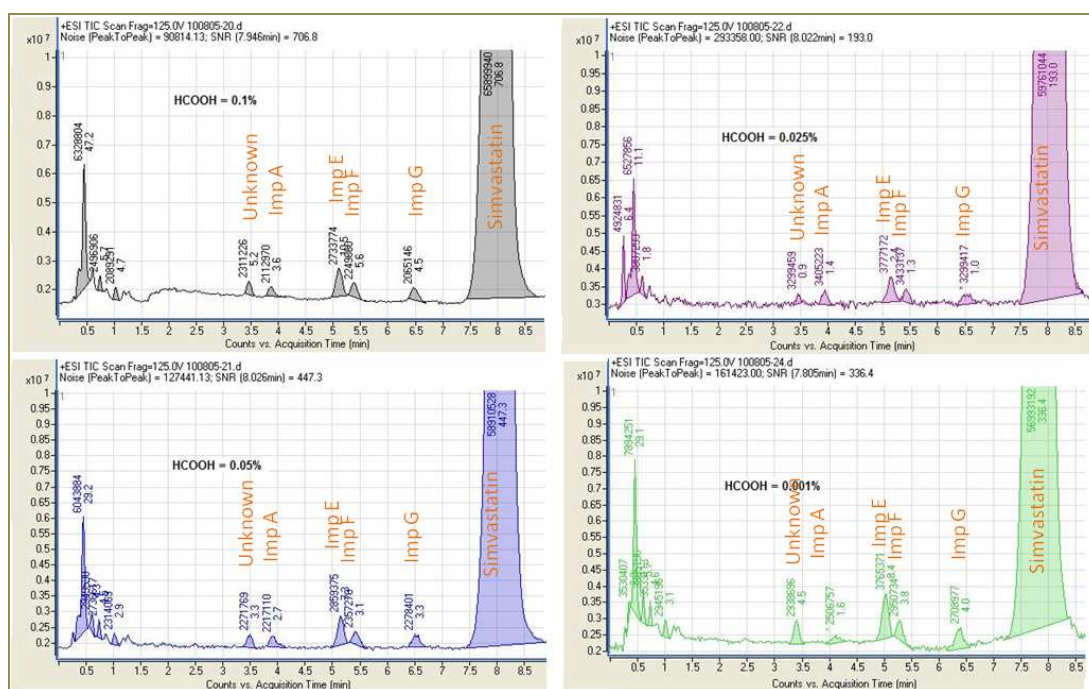


Figure III-6: Total ionic chromatograms of 2 $\mu$ L simvastatin for peak identification CRS solution injected in chromatographic systems using various mobile phase buffer concentrations, formic acid 0.1% (top left), formic acid 0.05% (bottom left), formic acid 0.025% (top right) and formic acid 0.001% (bottom right)

Globally, there was a slight increase in intensities (in counts) and areas (in counts.s) when the concentration in formic acid of the mobile phase was decreased from 0.1% to 0.001%. However, the noise rose dramatically at the same time, so that the signal to noise ratio decreased readily when reducing the ionic strength of the buffer solution, as illustrated in the following and corresponding bar charts.

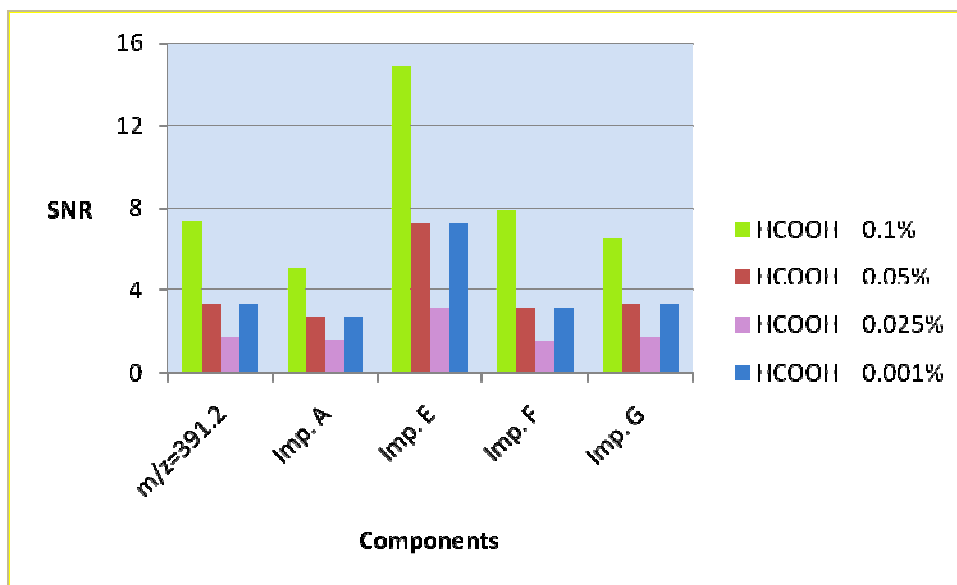


Figure III-7: Mass spectrometer signal to noise ratio for various mobile phase buffer concentrations in formic acid (0.1%, 0.05%, 0.025% and 0.001%) of a 2 $\mu$ L simvastatin for peak identification CRS solution injection

The signal noise resulted in severe fluctuation of the baseline. Baseline fluctuations represent major drawbacks when addressing the method sensitivity. This phenomenon was particularly observed when formic acid was used at a concentration of 0.025%, inducing an intrinsic signal to noise ratio logically lower. Signal to noise ratios corresponding to mobile phase buffer concentrations in formic acid of 0.05% and 0.001% were quite equivalent but, nevertheless, at a half value when compared to the signal to noise ratio reached with the system using the same buffer, at a concentration of 0.1%. Consequently, the chromatographic separation was carried out with mobile phases containing formic acid at a concentration of 0.1%. It should be noticed that such a concentration in formic acid induced a mobile phase buffer pH value of approximately 2.7.

The next paragraph spotlights the necessity to adjust the proportion of the mobile phase organic modifier in order to reach the best selectivity of the chromatographic system. Several experiments were performed and will be described in this part of the dissertation. The results are also presented aiming at demonstrating the importance of the mobile phase organic strength in the separation process.

### III.1.1.4 Effect of the mobile phase organic modifier concentration

In reversed phase interaction mechanisms, the retention of the compounds on the column is related to their more or less hydrophobic properties and is highly depending on the percentages of organic modifiers contained in the mobile phase. Therefore, in order to optimize the chromatographic separation of the solutes, the volume fraction of acetonitrile containing 0.1% acid formic (mobile phase B), was tested in isocratic mode within different ranges from 35% (no elution) to 55% (rapid elution). Capacity factor  $k'$  is a key indicator when estimating the quality of compound elution. Indeed, if the retention factor of the chemical substance is less than 2, elution will be considered as too fast, and if the retention factor is greater than twenty, the elution will be reckoned as too late [53]. Ideally, the numerical values of retention factors for analytes are comprised between 5 and 15, corresponding to decimal logarithm values from 0.7 to 1.2 because at those values, the term  $k'/1+k'$  of the Purnell relation (2) is maximal, inducing optimal resolution conditions. Consequently, plots of decimal logarithm values of capacity factor  $k'$  ( $\log k'$ ) versus various compositions of the mobile phase (%B) were drawn for each compound and reported in figure III-8. Outcomes were resulting from injection onto the column of 2 $\mu$ L simvastatin for peak identification chemical substance reference solution.

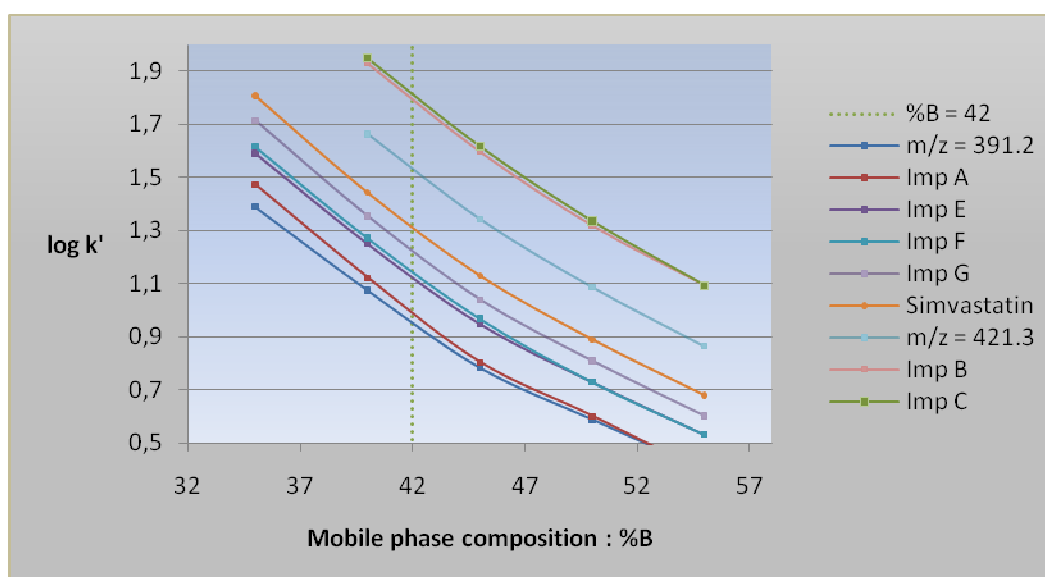


Figure III-8: Plots of the decimal logarithm values of capacity factor  $k'$  ( $\log k'$ ) versus mobile phase composition (%B) for simvastatin and major impurities

The plots of decimal logarithm values of capacity factor  $k'$  ( $\log k'$ ) versus different compositions in organic modifier of the mobile phase (%B) allowed to define the best chromatographic conditions to separate the components present in the mixture. The objective consisted in focusing on co-eluting compounds in order to determine the ideal proportions of

acetonitrile. Particular attention was given to critical pairs made up of impurities E and F, on one hand, impurities A and unknown at  $m/z = 391.2479$ , on another hand, and impurities B and C. The first critical pair investigated was the isomer pairs formed by impurities E and F. The plots  $\log k'$  against %B for those both impurities are reported in figure III-9.

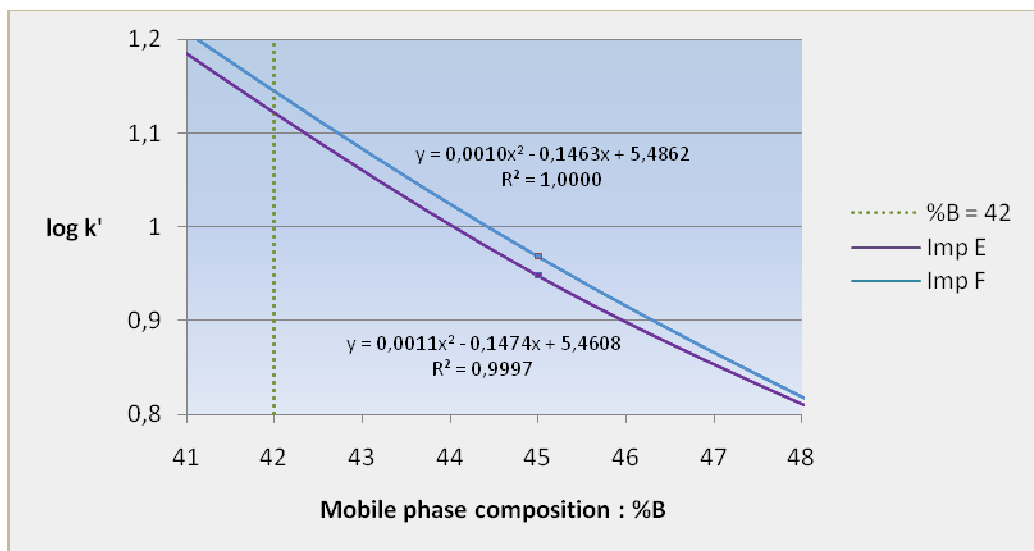


Figure III-9: Plots of the decimal logarithm of capacity factor  $k'$  ( $\log k'$ ) versus mobile phase composition (%B) for impurities E and F

The diagram indicated that the best separations between impurities E and F were reached for mobile phase compositions ranging within 41% and 44% of organic solvent. Below a proportion of acetonitrile of 41%, the peaks were late eluting, while the selectivity started to slightly decrease above a proportion of acetonitrile of 44%. Identical conclusions could be drawn when examining the lines  $\log k'$  against %B for impurities F (epilovastatin) and G, and simvastatin (figure III-10).

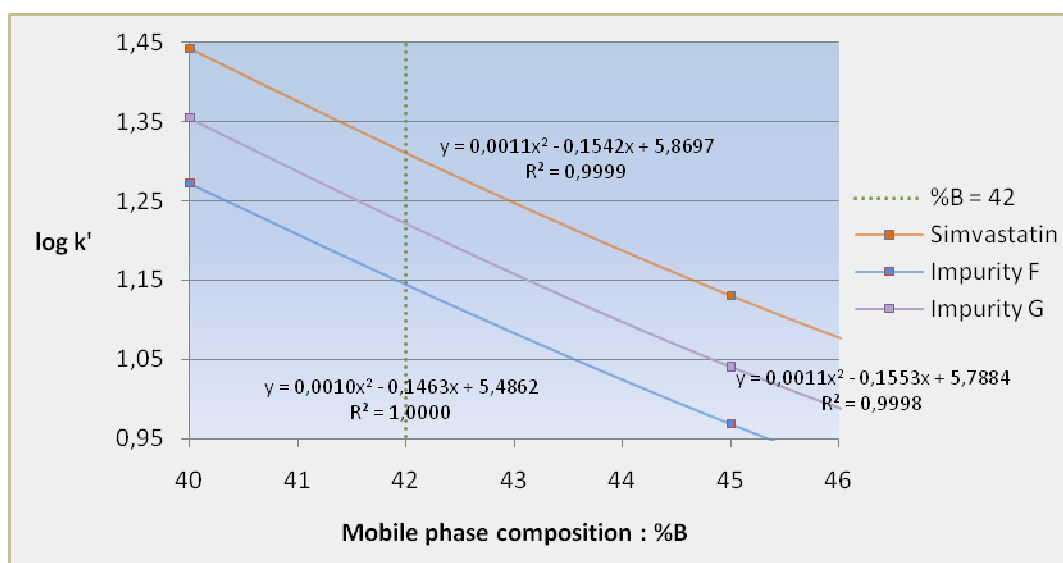


Figure III-10: Plots of the decimal logarithm of capacity factor  $k'$  ( $\log k'$ ) versus mobile phase composition (%B) for impurities F and G and simvastatin



The curves  $\log k'$  against %B for impurities A and unknown at  $m/z = 391.2479$  (figure III-11) showed a gradual decrease in selectivity and resolution between those species when changing the proportion of acetonitrile from 41% to 45 %.

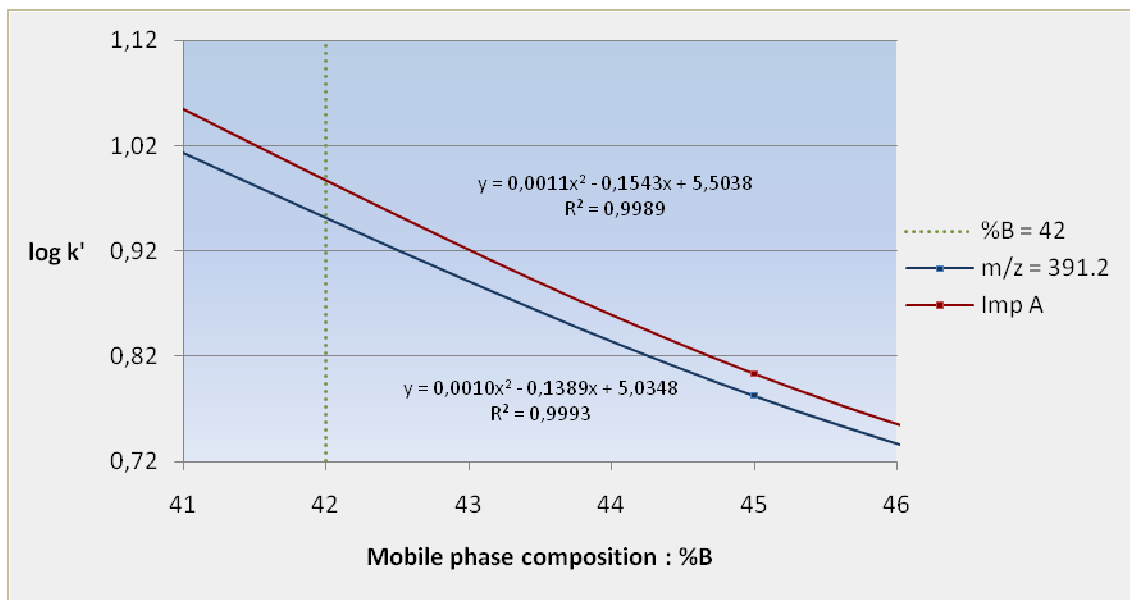


Figure III-11: Plots of the decimal logarithm of capacity factor  $k'$  ( $\log k'$ ) versus mobile phase composition (%B) for impurities A and unknown at  $m/z = 391.2479$

Therefore, a compromise between selectivity and sufficient compound resolution, within a reasonable analysis time, for simvastatin, Ph. Eur. impurities A, E, F, G and unknown species at  $m/z = 391.2479$ , was reached by setting the gradient starting conditions at a mobile phase composition in organic modifier of 42%. At these conditions, simvastatin should theoretically emerge from the column with a retention time corresponding to a decimal logarithm value of capacity factor equal to 1.3, which stands for a capacity factor of 20 and is then regarded as a limit value.

Consequently, critical pair constituted by Ph. Eur. impurities B and C, which eluted later than simvastatin, needed a much higher volume fraction of organic modifier to be separated in an appropriate total analysis time, as illustrated in figure III-12.

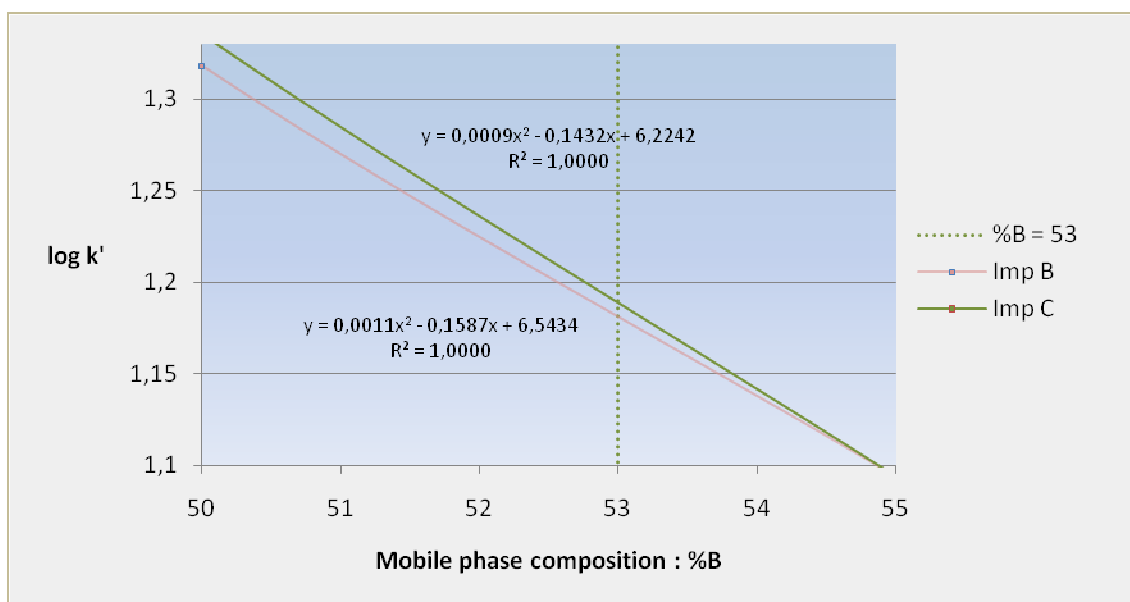


Figure III-12: Plots of the decimal logarithm of capacity factor  $k'$  ( $\log k'$ ) versus mobile phase composition (%B) for impurities B and C

The examination of this graph led us to conclude that the best resolution between Ph. Eur. impurities B and C, for fastest elution conditions, was obtained when a 53% volume proportion of organic modifier was used. Concerning the most strongly retained components, like Ph. Eur. impurity D, imposing a high proportion of acetonitrile (87.5%) removed them from the stationary phase and contributed to clean the column. Several linear gradients were performed during the optimization phase of the chromatographic separation, using varied steepness and considering the dwell volume of the system (0.5 mL). The following gradient was eventually implemented to separate simvastatin and its related substances (Table III-2).

Table III-2: Final gradient conditions of the developed in-lab method

Time (min)	Mobile phase A (per cent V/V)	Mobile phase B (per cent V/V)
0	58	42
6.5	58	42
6.5 - 7.0	58 → 47	42 → 53
9.5	47	53
9.5 - 14.0	47 → 12.5	53 → 87.5
17	12.5	87.5
17 - 17.2	12.5 → 58	87.5 → 42
20	58	42

### III.1.1.5 Effect of the mobile phase pH

The retention and separation properties were also investigated at isocratic conditions at pH 2.7 and pH 6.7, according to the manufacturer recommended pH range (2.0 – 8.0) for maximum column life. Indeed, silica based stationary phases are particularly unstable under both, low acidic and alkaline conditions, due to the chemical properties of silica. Silica hydrolyzes and dissolves at pH above 9.0, while a pH below 2.0 causes the loss of the functional group bonded to the silica particle by siloxane linkage. In a first experimentation, the pH of the mobile phase was set at a value of 2.7 by adding 0.1% (v/v) of concentrated formic acid in a mixture of acetonitrile and water 40:60 (v/v), whereas in a second experimentation, the pH of the mobile phase was set at a value of 6.7 by adding 25 mM of ammonium acetate in a mixture of acetonitrile and water 40:60 (v/v). The bar charts below (see figure III-13) represents the retention times, in minutes, obtained under both tested pH values of the mobile phase, for the following compounds: simvastatin, Ph. Eur. impurities A, E, F and G, and unspecified impurity with a mass to charge ratio equal to  $m/z = 391.2479$ .

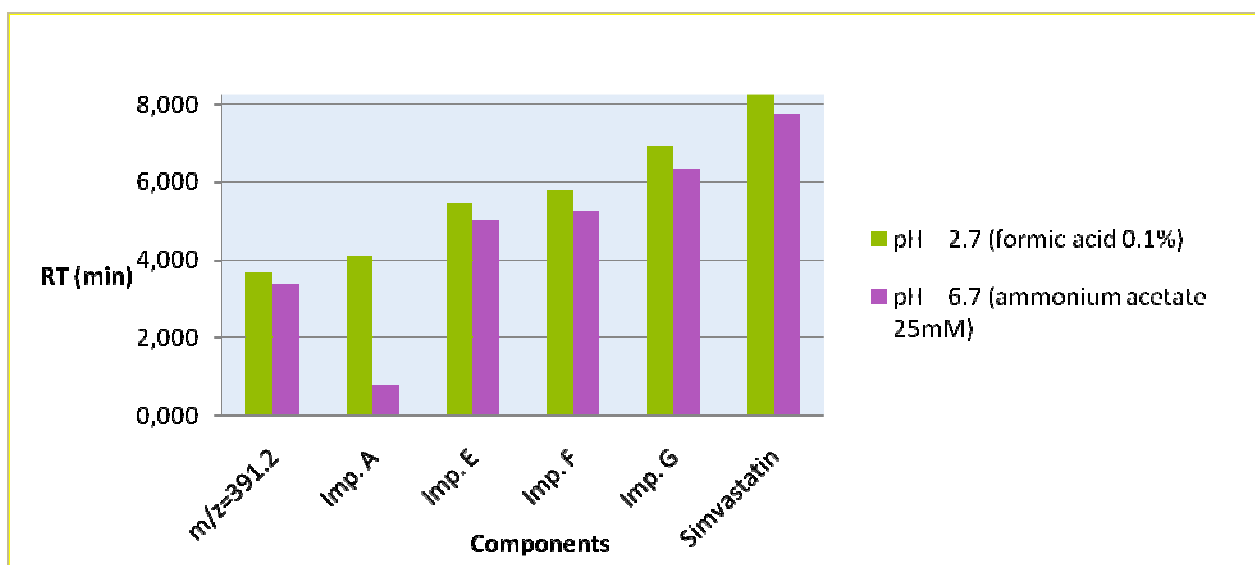


Figure III-13: Bar charts of the retention times against pH values at 2.7 and 6.7 of the mobile phase, for simvastatin, European Pharmacopoeia specified impurities A, E, F, G and unknown at  $m/z = 391.2479$

As illustrated in Figure III-13, retention times were not significantly affected by the different pH conditions studied, except for Ph. Eur. impurity A, which eluted at a retention time close to the hold-up time of the column,  $t_0 = 0.274$  minutes, when eluting with the mobile phase containing the ammonium acetate 25mM buffer.

According to their high pKa values around 13.5, the majority of the compounds related to simvastatin were not affected by a rise of the pH value from 2.7 to 6.7, because at those pH values they still remained under their neutral form. Actually, in this case, the buffer pH had only prominent impact on the retention of Ph. Eur. impurity A (simvastatin hydroxy acid). Indeed, with a pKa value of 4.31, Ph. Eur. impurity A went through the column under its neutral form at pH2.7 and was retained by the lipophilic stationary support. At pH6.7, Ph. Eur. impurity A migrated through the column in its ionized form having a low retention. Consequently, formic acid 0.1% v/v at pH2.7 was chosen as mobile phase buffer.

### III.1.1.6 Influence of the column temperature

By increasing the diffusivity of the analytes, the temperature of the column is an important variable toward reducing analysis time. Furthermore, it is a major factor for lowering mobile phase viscosity and consequently system backpressure, as suggested by Darcy law (cf. equation 14). It also impacts the polarity and the pH of the mobile phase by decreasing both factors, and thus, the selectivity of the column, so that predicting retention mechanisms, when changing column temperature, is a tricky task. Controlling the temperature has also a great influence on the column efficiency, contributing thus to increase the signal to noise ratio [51]. Experimentations were carried out in order to establish the influence of the column temperature over the retention mechanisms. Figure III-14 corresponds to the plots of retention time, expressed in minutes, against column temperature, expressed in °C, for simvastatin and its major impurities.

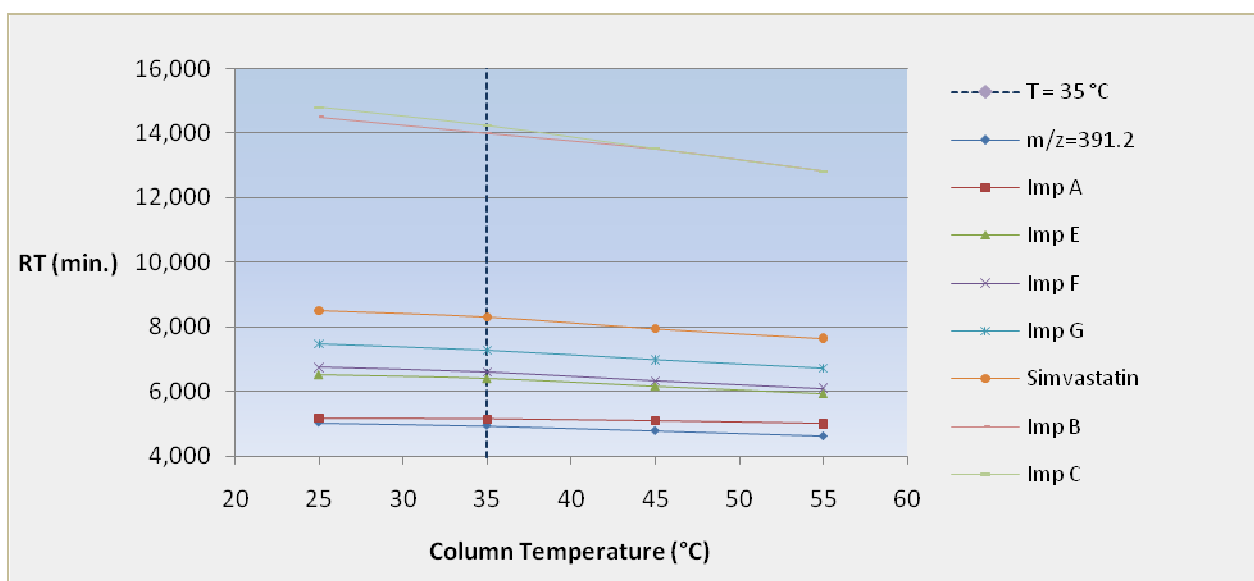
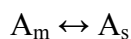


Figure III-14: Plots of retention time against column temperature for simvastatin, impurities A, E, F, G, B, C and unknown at m/z = 391.2479

The chromatographic separation of chemical species in a mixture within the analytical column is related to the differential partition coefficient (K) of each compound between the mobile and the stationary phases, resulting in distribution equilibrium of analyte A between both phases:



Where

$A_m$  represents the analyte in the mobile phase.  
 $A_s$  represents the analyte in the stationary phase.

That is

$$K = [A]_s / [A]_m \quad (19)$$

Now

$$\ln K = -\Delta G_0 / RT \quad (20)$$

With

K is the partition coefficient.  
 $\Delta G_0$  corresponding to the Gibbs free energy.  
R is the ideal gas constant.  
T is the thermodynamic temperature.

Equation (20) shows that partition coefficient K is inversely proportional to temperature changes. Indeed, raising or decreasing temperature respectively generates a drop or an increase in partition coefficient value. Equation (19) points out that a fall in partition coefficient value corresponds to simultaneous rise in analyte concentration in mobile phase  $[A]_m$  and decrease in analyte concentration in stationary phase  $[A]_s$ . Therefore, according to both equations (19) and (20), a downward trend in retention times was observed for each compound when the temperature of the column was gradually increased (see figure III-14).

However, it was interesting to note that the decreasing in retention times was less pronounced for simvastatin hydroxy acid (impurity A) than for the unknown impurity at  $m/z = 391.2479$ . Hence, increasing temperature improved the resolution between these two compounds of interest. On the other hand, above a temperature of 40°C, Ph. Eur. impurities B and C were no more satisfyingly separated and emerged finally with the same retention time when the temperature was over 45°C.

A compromise between minimizing the analysis time and conserving the best separation for all compounds was reached by setting the column temperature at 35°C. It should be noted that symmetry factors and plate numbers were not significantly changed over the investigated temperature range (results are not presented here).

The next two parts of this dissertation are devoted first, to the comprehension of autosampler carryover and contaminations occurring during the injection process, and second, to the importance of the choice of an adequate sample solvent. Solutions and means to avoid drawbacks linked to such intrusive phenomena are exposed in those paragraphs.

### **III.1.1.7 Autosampler carryover and contaminations**

Autosampler carryover is caused by residual analyte from preceding injections, ensnared within the injection system. Several parts of the injection system can cause carryover and especially, needle outside and inside, needle seat, sample loop, needle seat capillary and injection valve. It can dramatically affect the quality of the results by impacting the reliability and the performances of the analytical method, in terms of accuracy and precision. And this is particularly noticeable with very sensitive and critical LC-MS applications. Hence, it was very important to remove, even infinite, traces of previously injected sample solutions. In order to dismiss that inconvenience, a post injection rinsing of the device was introduced as part of the programming of the injector (see appendix D). The device was programmed to trigger first a needle wash for 10 seconds in the flush port containing a mixture in equal proportions of acetonitrile and water, in order to rinse the outer part of the needle and to prevent a possible contamination of the needle seat. One minute after the sample was injected, the valve unit switched to the bypass position. In that position, the mobile phase flew directly to the column without passing through the sample loop, the needle and the needle seat capillary. This contributed to reduce the system delay volume and to shorten the analysis cycle times. After 14.5 minutes, successive switching of the injection valve between the positions “mainpass” and “bypass” led to remove the eventually trapped analytes from the rotor. Finally the injection system was rinsed with the highest proportion of organic modifier in mobile phase.

### **III.1.1.8 Sample solvent investigation**

An important factor to take into account when running HPLC analysis concerns the strength of the dilution solvent. Indeed, it is well-known that the nature and the composition of the dilution eluent have a significant impact on chromatographic peak shapes. Peak fronting and peak broadening can be observed, for example, when the sample solvent is stronger than the mobile phase, as illustrated in figure III-15. Band broadening is detrimental to resolution as demonstrated in following diagrams. For instance, the critical pairs constituted either of Ph. Eur. impurities A and

unknown at  $m/z = 391.2479$ , or Ph. Eur. impurities E and F, or Ph. Eur. impurities B and C, were henceforth not resolved when the sample was dissolved in pure acetonitrile or pure methanol, respectively diagrams a) and c) in figure III-15. Moreover, peak doubling may also arise when the sample is diluted in a solvent incompatible with the mobile phase. Thus, a particular attention was focused on the sample preparation and dissolution step. Typically, samples are prepared in mobile phase whenever possible, or in a solvent of lower eluting strength. However, simvastatin demonstrated rapid degradation into simvastatin hydroxy acid when prepared in the mobile phase, due to the low pH value of the solution. Actually, in this case, simvastatin underwent an oxidation reaction in presence of an acid. Consequently, samples were diluted in a mixture of ultrapure water and acetonitrile in proportion 60:40 (v/v) in order to obtain satisfying peak shapes as illustrated in graph b) in figure III-15.

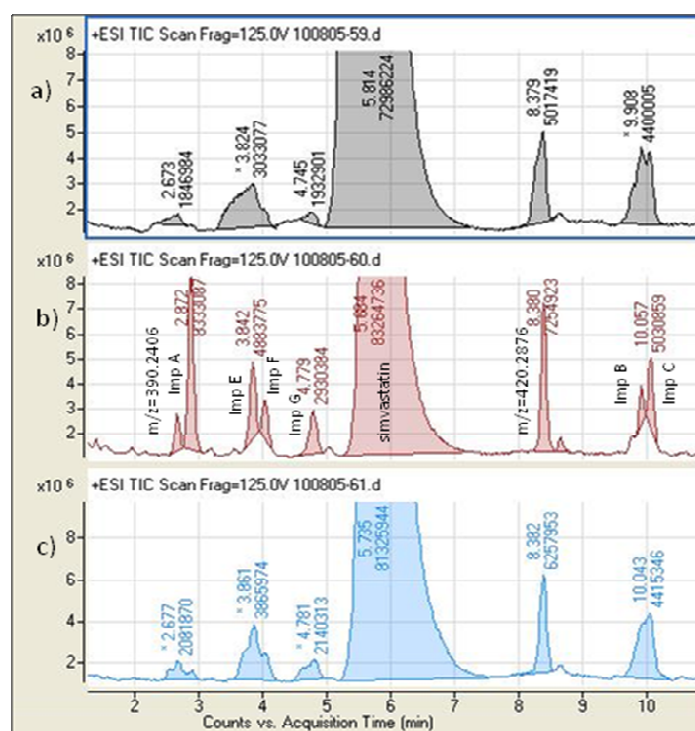


Figure III-15: Total ionic chromatograms of a sample prepared  
a) in pure acetonitrile  
b) in a water/ acetonitrile 60:40 (v/v) mixture  
c) in pure methanol

### **III.1.2 Optimization of the mass spectrometer parameters**

Once the perfecting and the development of the chromatographic separation were carried out inducing reasonable method selectivity, significant and focused attention was turned to the optimization of the mass spectrometer parameters. Indeed, correct spray chamber and interface settings, like the ionization modes, positive or negative, the nebulizer gas pressure, the drying gas temperature or the drying gas flow rate, the capillary voltage, or the fragmentor voltage, favor ion formation and result in maximized sensitivity. Consequently, many different experimental conditions were tried out and tested with the objective to set properly the parameters of the ion source. The parameters were tested one after another by keeping successively the previously optimized adjustments.

#### **III.1.2.1 Choice of the ionization source and functioning mode**

Two different sorts of ion sources were available at the laboratory for the mass spectrometer, the electrospray ionization source and the atmospheric pressure chemical ionization source. Nevertheless, the selection of the type of source was not based on experimentations but on criteria found in the literature [54-55]. Firstly, all the papers studied concerning simvastatin analysis describe the electrospray ionization interface between the liquid chromatographic system and the mass spectrometer as the best alternative to obtain signals with high sensitivity. Secondly, in order to detect and identify potential drug substances in products of dubious origin, like counterfeit medicines, the laboratory of the Swedish Medical Products Agency had developed screening methods based on LC-MS and using preferentially the electrospray ion source. And since the requested analysis were mainly coming from customs seizures, they were treated for urgent most of the time, so that, with regard to the Swedish Medical Products Agency activities, it was not possible to realize extended trials on the APCI source. Accordingly, our choice went to the use of the electrospray source in order to generate analyte ions.

However, the electrospray ionization mode was investigated during this study. Trials were run aiming at determining for which type of electrospray ionization mode, positive or negative, the most enhanced signal intensities were achieved. In positive mode, a strong electrostatic field of about +4000V to +6000V is applied to the spray needle to engender the ionization process of the analytes. In that mode, only the cations enter the mass analyzer in order to be detected.



In negative mode, on the contrary, an electric potential of about -4000V to -6000V is applied to the spray needle, and only the anions enter the mass analyzer, inducing then different response signals. Figure III-16 illustrates the variations in detector responses when using the electrospray ion source either in positive mode (upper diagram) or in negative mode (lower diagram). The diagrams corresponding to the numbers of counts recorded by the detector *versus* the acquisition time, expressed in minutes, for both ion source modes, negative and positive, were represented with the same scale so that a visual comparison between the signal intensities was immediate.

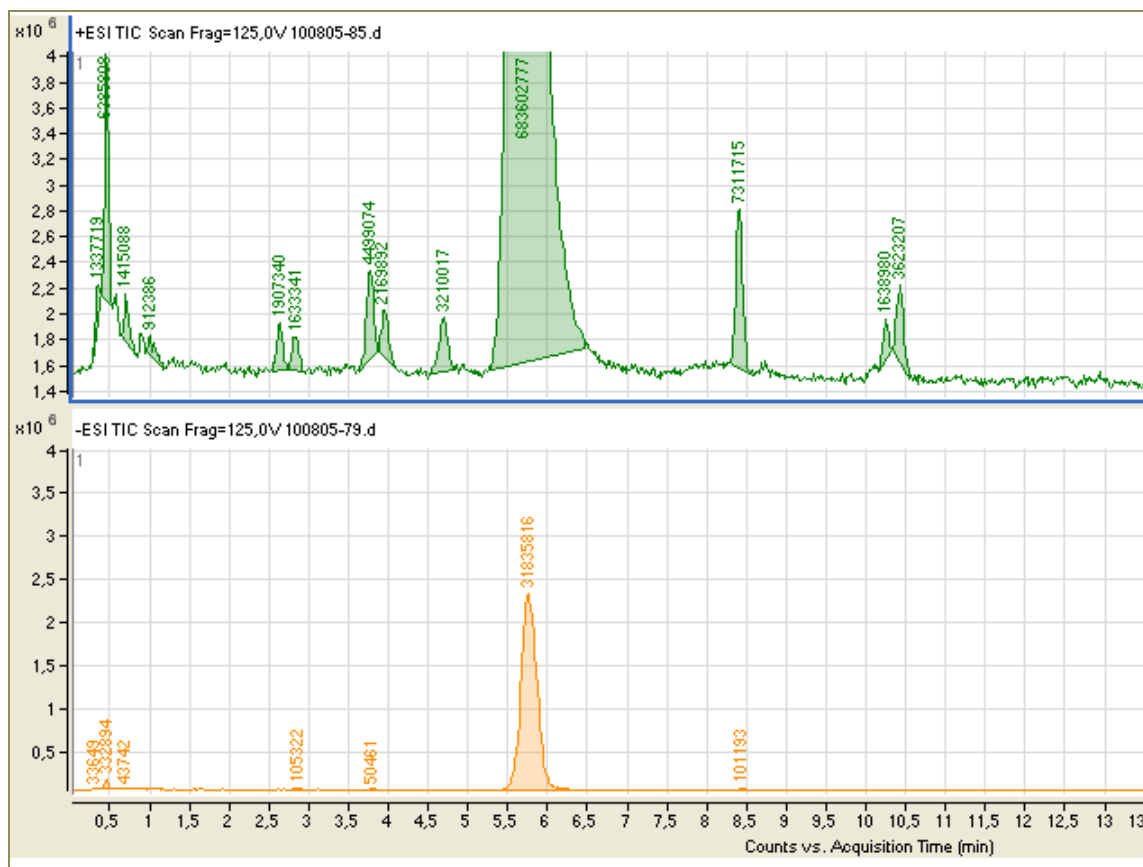


Figure III-16: Detector response when using an electrospray ion source in positive mode (upper diagram). Detector response when using an electrospray ion source in negative mode (lower diagram) corresponding to the injection of the identical solution

The diagrams show obviously that the responses were quite different depending on whether the electrospray was used either in positive mode or in negative mode. Globally, signal intensities corresponding to simvastatin and related impurities drop dramatically, more than twentyfold, when using the electrospray in negative mode. Furthermore, some of the species ionized in positive mode were not ionized in negative mode, so that they could not be detected by the mass spectrometer. Consequently, all the study was realized with an electrospray ionization source functioning in positive mode.

### III.1.2.2 Effect of the nebulizer gas pressure

The first parameter investigated in order to improve the intensities of the mass chromatogram peaks was the gas pressure of the nebulizer. Nitrogen was used as nebulizing gas and its role consisted principally in generating a stable spray, without fluctuation, as fine as possible, at the tip of the needle, so as to induce a symmetrical plume.

The starting settings of the other parameters of the mass spectrometer such as the drying gas temperature, the drying gas flow rate, the capillary voltage and the fragmentor voltage were adjusted at typical values. These values are described in table III-3.

Table III-3: Mass spectrometer starting settings before optimization

Mass spectrometer parameters	Value
Drying gas temperature	300 °C
Drying gas flow rate	10 L.min <sup>-1</sup>
Capillary voltage	3100 V
Fragmentor voltage	190 V

The variation of the area (counts.s), in function of the nebulizing gas pressure (psi), is displayed in figure III-17 for simvastatin main impurities, comprising the impurities specified in the European Pharmacopoeia (7<sup>th</sup> edition) and some new impurities such as the impurity located at  $m/z = 391.2479$  and the impurity located at  $m/z = 421.2949$ . The description and the characterization of these unknown impurities are carried out in paragraph III.1.4.2 of this dissertation, entitled “Identification of new impurities by LC-MS/MS”.

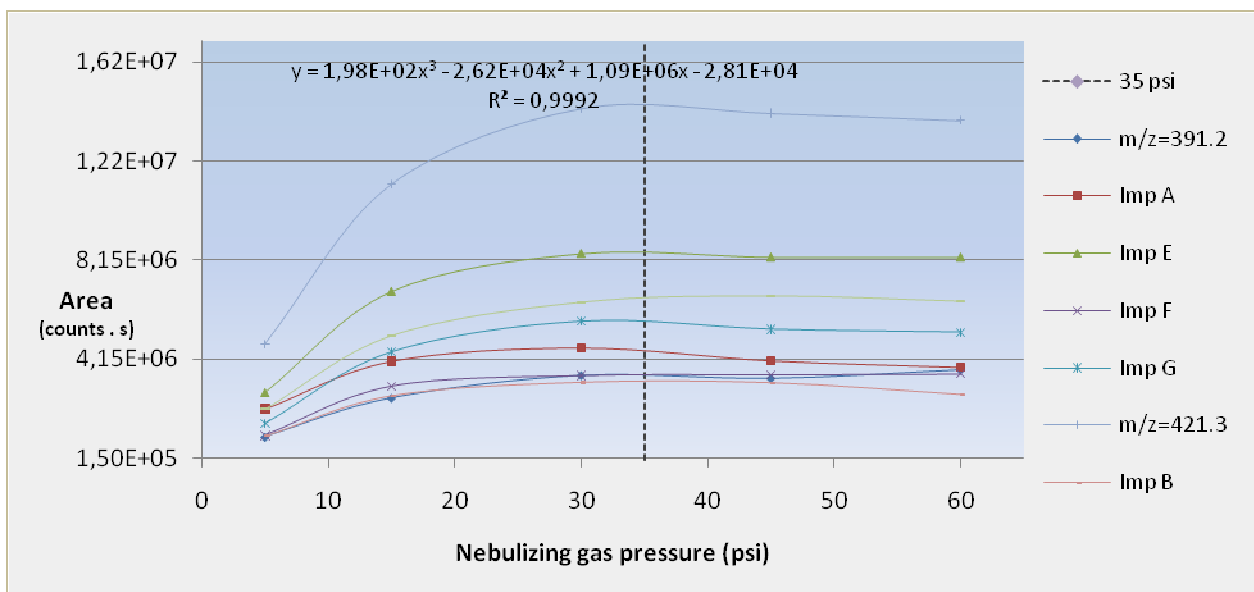


Figure III-17: Plots of simvastatin main impurities peaks area (counts.s) against nebulizing gas pressure (psi)

All the curves plotted in figure III-17 present a graphical look-shaped plateau between nebulizing gas pressure values of 30 psi and 60 psi. Therefore a typical value of 35 psi was assigned to the nebulizing gas pressure parameter.

The nebulizing gas pressure is not the only factor impacting on the intensity of the chromatographic signals. Other factors, like the drying gas temperature or the drying gas flow rate, were dependent on the composition of the mobile phase and on the flow rate of the mobile phase. Indeed, all those parameters adjustments contributed to help the desolvation of droplets and to help the ionization process of the compounds of interest. Desolvation is facilitated when the percentage of organic modifier in the mobile phase is elevated or when the flow rate is lower. Therefore, in concrete terms, higher drying gas flow rate and temperature are needed when the organic proportion in the mobile phase is decreased or when the flow rate of the chromatographic system is increased.

Once the optimum value of the peak areas was obtained using a nebulizer gas pressure of 35 psi the impact of the drying gas temperature and the drying gas flow rate was examined. The results are presented in the next two divisions.

### III.1.2.3 Influence of the drying gas temperature

Drying gas settings, like temperature and gas flow rate, are also, as specified previously, critical factors to study when proceeding with the optimization of the mass spectrometer. The drying gas temperature is the temperature of warm nitrogen gas current intended to provide for efficient solvent evaporation. Incomplete drying can induce spikes and noise in the mass spectra caused by remaining solvent droplets in the ion source. Conversely, high temperatures can have a detrimental effect on the signal intensities when the thermal stability of the samples are reached, or exceeded. Increasing or reducing the drying gas temperature can also provoke the decrease or the rise in sodium adducts generation and in neutral loss. So, the plots of the areas (counts.s), against drying gas temperatures (°C), were drawn for simvastatin main impurities. Corresponding plots are reported in figure III-18.

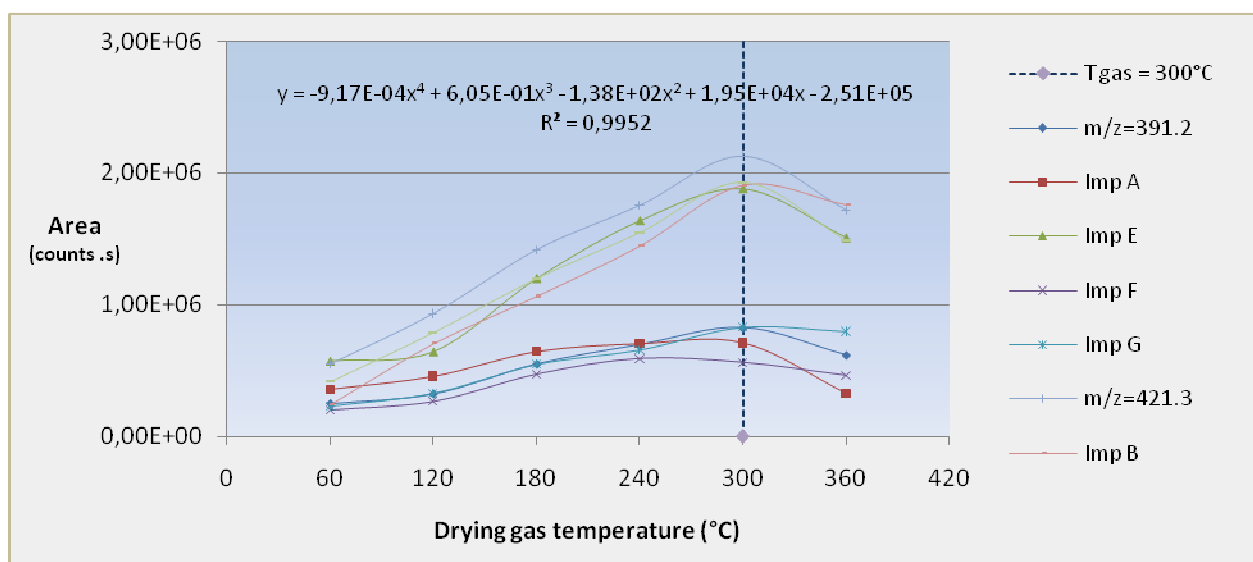


Figure III-18: Plots of the areas (counts.s) against drying gas temperatures (°C) for impurities A, E, F, G, B and C and unknown at m/z = 391.2479 and m/z = 421.2949

The signals of each impurity steadily rose in the range of 60°C to 300°C. However there was a dramatic increase in signal for Ph. Eur. impurities B, C, E and unknown at m/z = 421.2949, while there was a slighter increase for Ph. Eur. impurities A, F, G and unknown at m/z = 391.2479. Above 300 °C, a general drop in area was observed probably due to more neutral sodium adduct formation or thermal degradation of the analytes, so that the drying gas temperature was set to a value of 300°C in order to achieve optimum conditions.

### III.1.2.4 Drying gas flow rate adjustment

The adjustment of the drying gas flow rate can contribute, for its part, to minimize the formation of clusters. Drying gas flow rate helps shrinking droplets of the sample flow by evaporating the spray solvent. As a result, it prevents liquid entering the system and contaminating the ion optics so that it can be quite rightly considered as a barrier against sample pollution. The plots of the areas (counts.s) against drying gas flow rates ( $\text{L}\cdot\text{min}^{-1}$ ) were produced for the principal impurities of simvastatin (figure III-19).

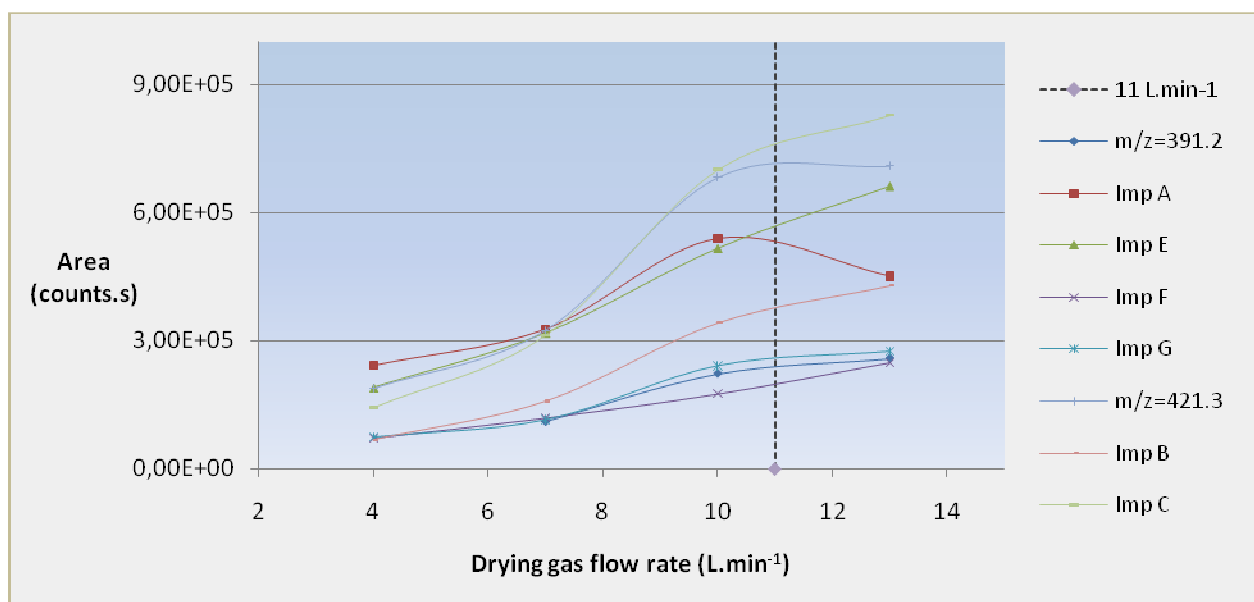


Figure III-19: Plots of the areas (counts.s) against drying gas flow rates ( $\text{L}\cdot\text{min}^{-1}$ ) for impurities A, E, F, G, B and C and unknown at  $m/z = 391.2479$  and  $m/z = 421.2949$

The response of the signal was checked between the extreme parameter settings of the instrument concerning the drying gas flow rates, which were  $4 \text{ L}\cdot\text{min}^{-1}$ , as lowest value, and  $13 \text{ L}\cdot\text{min}^{-1}$  as maximum value. A general increasing in area was noticed for each analyte, except for Ph. Eur. impurity A, which demonstrated a slight decrease in signal above a drying gas flow rate value of  $10 \text{ L}\cdot\text{min}^{-1}$ . A compromise solution between optimized response and reasonable nitrogen gas consumption was found by setting the drying gas flow rate parameter to a value of  $11 \text{ L}\cdot\text{min}^{-1}$ .

### III.1.2.5 Role of the capillary voltage

The capillary voltage ( $V_{\text{cap}}$ ) corresponds to the voltage applied to the entrance of the interface capillary. In theory,  $V_{\text{cap}}$  constitutes an essential parameter of the ionization process by maximizing the ion transmission. Indeed, the role of the capillary voltage consists in drawing the charged species into the source. Thus the capillary voltage was investigated over a wide range of values, from 2500 V to 5000 V, in order to determine the value equivalent to the maximum signal enhancement and sensitivity (figure III-20). It should be noted that too high voltages, like voltages above 5000 V, may induce corona discharge in the electrospray chamber, causing signal fluctuations and leading to latent instrument damages.

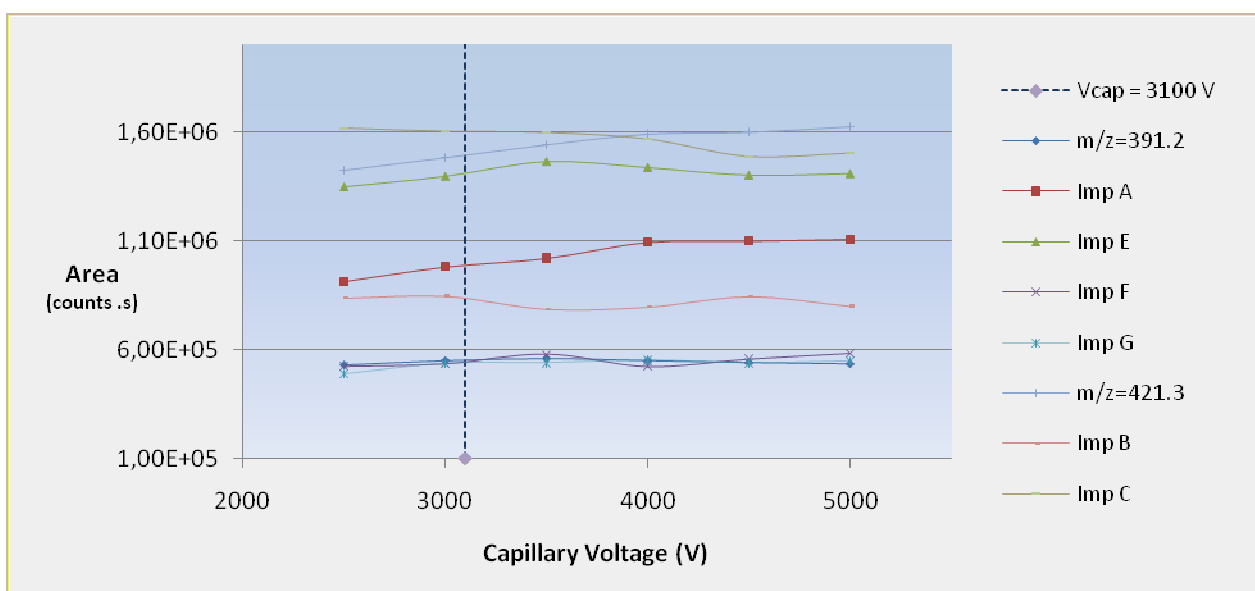


Figure III-20: Plots of the areas (counts.s) against capillary voltage (V) for impurities A, E, F, G, B and C and unknown at  $m/z = 391.2479$  and  $m/z = 421.2949$

The overall trends of the plots show clearly that the capillary voltage adjustment did not severely impact the response of the detector, so that a value of 3100 V was chosen for the analysis.

### III.1.2.6 Impact of the fragmentor voltage

A series of experiments evaluating the influence of the fragmentor voltage ( $V_{\text{frag}}$ ) on the detector response was performed in order to complete the optimisation of the mass spectrometer. The fragmentor voltage corresponds to the voltage applied to the exit of the transfer capillary. It impacts therefore the ion transmission. The ionization behaviour is particularly affected by this parameter, as well as the in-source molecule fragmentation and the adduct formation. In-source collision induced dissociation, or in-source CID, stands for molecular ion splitting up into smaller ion fragments in the ion source. Consequently, in-source dissociation may trigger a decrease in the intensity of the molecular ion.

Conversely, in-source CID may also reduce the emergence of solvent adducts or dimer formation, increasing therefore the sensitivity. Therefore, the fragmentor voltage was examined in a range of values starting from 50 V to 400 V. Results are reported in figure III-21 describing the trends in area, expressed in counts.s, of simvastatin impurities, when varying the fragmentor voltage over a scale of 350V.

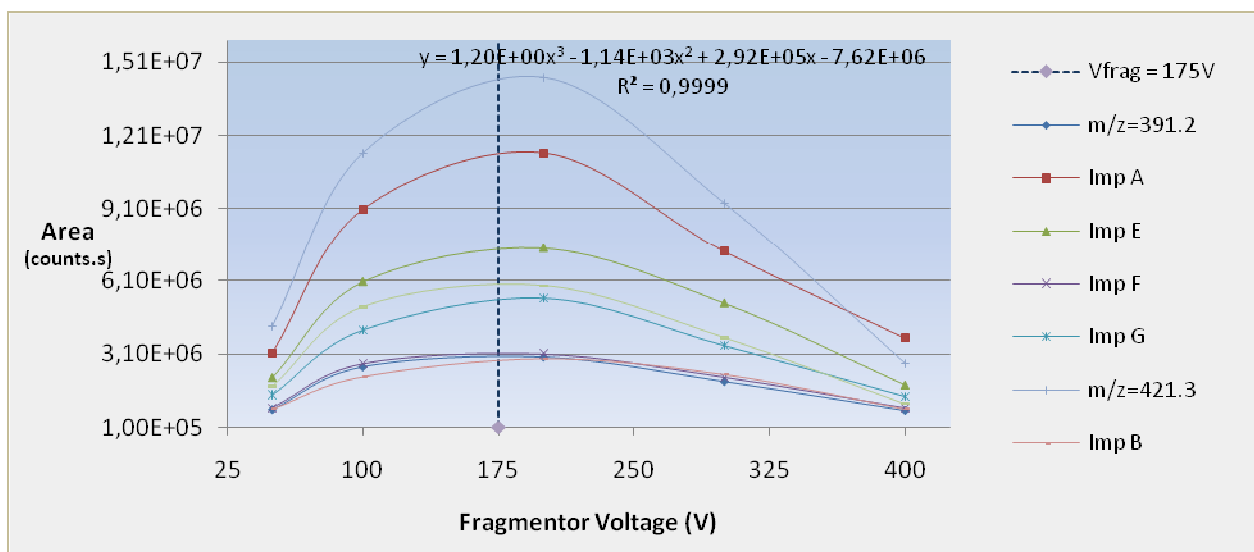


Figure III-21: Plots of the area (counts.s) against fragmentor voltage (V) for impurities A, E, F, G, B and C and unknown at  $m/z = 391.2479$  and  $m/z = 421.2949$

All the graphs representing the area of the compounds versus the fragmentor voltage showed an optimum for  $V_{\text{frag}}$  equivalent to 175V. Accordingly, the fragmentor voltage was tuned to a value of 175V in order to achieve the highest sensitivity of the mass spectrometer in regard to the respective molecular ion intensities.

Once all the ESI parameters were optimized, the response linearity and the sensitivity of the mass spectrometric detector were checked.

### III.1.2.7 Response linearity of the mass spectrometric detector

Assuming that the linearity of the HPLC injection system had been periodically checked during equipment qualification, the response linearity of the mass spectrometer detector was investigated by varying the injection volume of mixture containing simvastatin and impurities from 2  $\mu$ L to 30  $\mu$ L. Lines, equations and determination coefficients of the calibration curves are presented in figure III-22 and summarized in table III-4 for main simvastatin specified impurities, i.e. Ph. Eur. impurities A, B, C, E, F and G and unknown impurities at  $m/z = 391.2479$  and  $m/z = 421.2949$ .

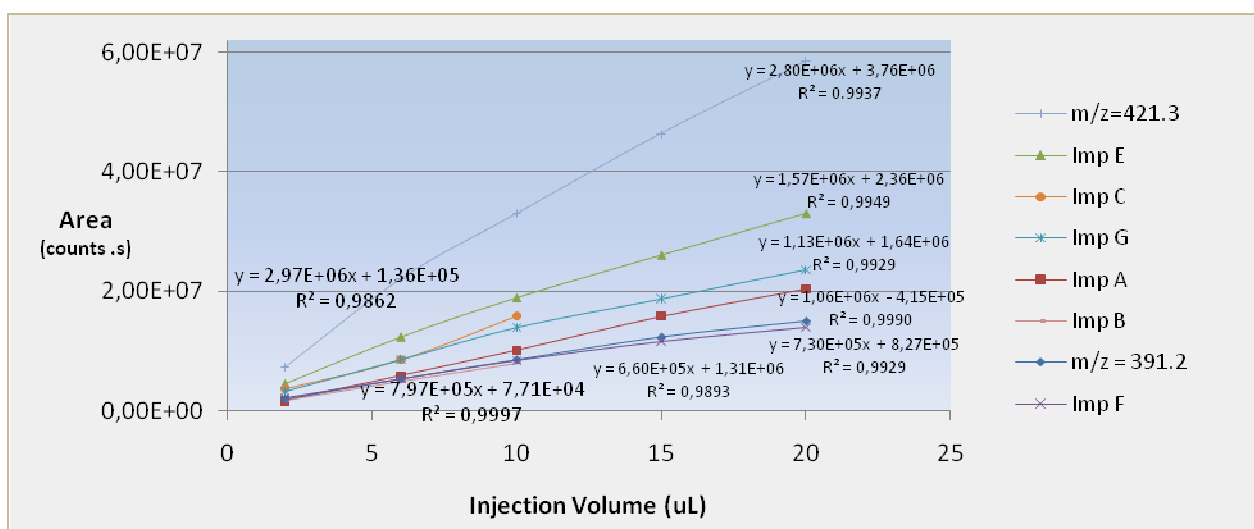


Figure III-22: Linearity of the LC-MS signal of simvastatin specified impurities A, B, C, E, F and G, and unknown impurities at  $m/z = 391.2479$  and  $m/z = 421.2949$

The obtained results implied that the detector response expressed as peak area (counts.s) *versus* injection volume ( $\mu$ L) showed a linear trend for most of the components in the range between 2  $\mu$ L and 20  $\mu$ L.

Table III-4: Mass spectrometric detector linearity for main simvastatin impurities

Compound	Equation	Determination Coefficient R <sup>2</sup>	Linearity Range ( $\mu$ L)
m/z=421.2949	$Y = 2.80 E+06 x + 3.76 E+06$	0.9937	2 - 20
Imp E	$Y = 1.57 E+06 x + 2.36 E+06$	0.9949	2 - 20
Imp C	$Y = 2.97 E+06 x + 1.36 E+05$	0.9862	2 - 10
Imp G	$Y = 1.13 E+06 x + 1.64 E+06$	0.9929	2 - 20
Imp A	$Y = 1.06 E+06 x + 4.15 E+05$	0.9990	2 - 20
Imp B	$Y = 7.97 E+05 x + 7.71 E+04$	0.9997	2 - 10
m/z=391.2479	$Y = 7.30 E+05 x + 8.27 E+05$	0.9929	2 - 20
Imp F	$Y = 6.60 E+05 x + 1.31 E+06$	0.9893	2 - 20



As shown in Table III-4, the coefficients of determination were determined to be comprised between 0.9893 and 0.9990 over the range of 2 $\mu$ L and 20 $\mu$ L for Ph. Eur. impurities A, E, F and G and unknown impurities at  $m/z = 391.2479$  and  $m/z = 421.2949$ , whilst they were found to be equal to 0.9862 and 0.9997 over the range of 2 $\mu$ L and 10 $\mu$ L for Ph. Eur. impurities C and B, respectively.

The linearity of the signal corresponding to the active pharmaceutical ingredient simvastatin was similarly verified and a determination coefficient equal to 0.9856 was found over the range between 2 $\mu$ L and 10 $\mu$ L (figure III-23).

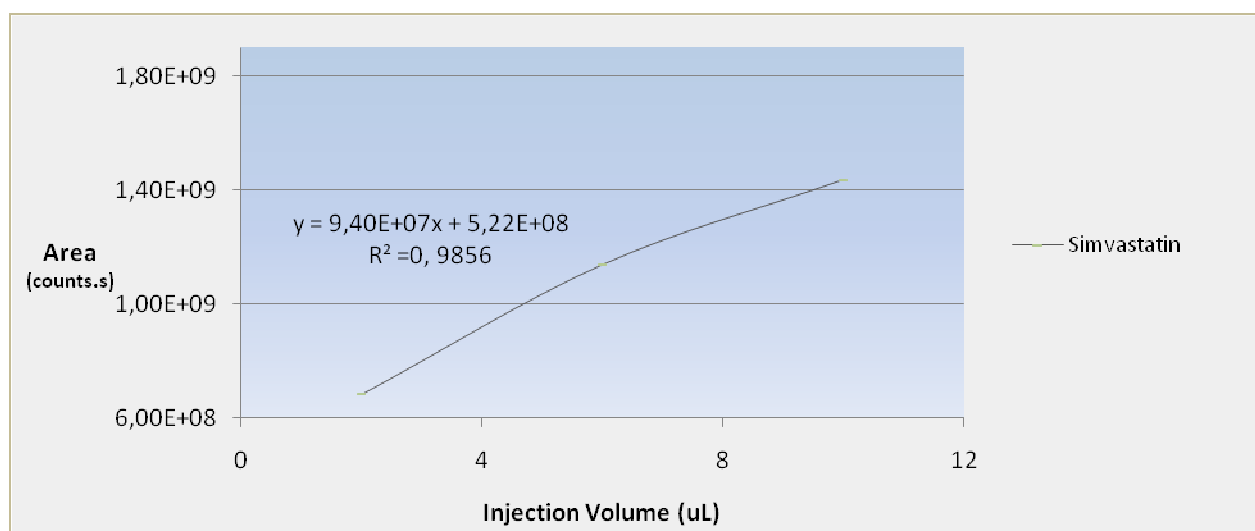


Figure III-23: Linearity of simvastatin LC-MS signal

Given the results obtained for the linearity of simvastatin and its main impurities between 2  $\mu$ L and 10  $\mu$ L, the injection volume was set to 5  $\mu$ L.

The quantitation limit (LOQ) of the method was equally investigated by injecting a low concentration solution of simvastatin corresponding to a signal to noise ratio defined as 10:1. The noise was determined by the Agilent MassHunter software by measuring the peak to peak height (h) of the baseline observed over a distance equal to 5 times the width at half-height on both sides of simvastatin chromatographic peak. However, according to the European Pharmacopoeia (7<sup>th</sup> edition) the noise is determined by considering the half peak to peak height (h/2) of the baseline, so that the quantitation limit of the analytical method was found equal to 4.12 ng.mL<sup>-1</sup> [16].

The extracted ion chromatogram (EIC) of a serially diluted solution containing simvastatin at a concentration of 8.25 ng.mL<sup>-1</sup>, standing for a 41.25 pg API injection and presenting a signal to noise ratio of 10, is represented in figure III-24.

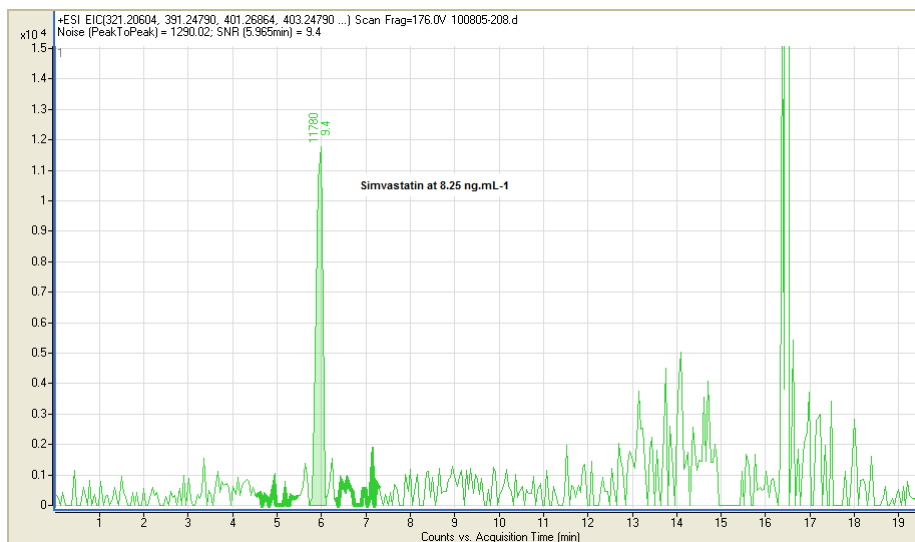


Figure III-24: Extracted ion chromatogram, displaying abundance and peak to peak signal to noise ratio of a low 8.25 ng.mL<sup>-1</sup> simvastatin concentration solution

The results showed that the method based on LC (ESI+) MS QTOF detection offered sensitivity at a picomole level for simvastatin compound (about 10 pM). Thus, compared to traditional ultraviolet diode array detection (UV-DAD), the sensitivity of the quadrupole - time of flight detector was about 10 – 25 times higher [56]. That noteworthy sensitivity, in addition of the QTOF high specificity, was of great importance to improve the capacity of the method for discriminating API origins.

### III.1.2.8 Measurement precision of the mass spectrometer response

A linear coefficient of determination equivalent to 0.9856, as obtained for simvastatin calibration curve, could be regarded as too low in a strictly scientific point of view. However, considering the intrinsic variability of the mass spectrometric response, that kind of values was quite satisfying. Indeed, differences in intensity distributions might result from instrument thermal sensitivity inducing flight tube expansion or else, inner electrode voltage variations. Slight vacuum and electronic instabilities or fluctuation in spray nebulising might also contribute to increase the signal variability.

A study was designed to assess this response variability by estimating the measurement repeatability (or intra-day precision) and the intermediate measurement precision, also called inter-serial precision [57] by using relative standard deviations expressed in percent (RSD%). The obtained results are summarized in table III-5 and whole raw data are presented in appendix B.

Table III-5: Intra-day (n=6) and inter-day (n=18) instrument precision considering peak areas

Compounds	m/z = 391.2406	Impurity A	Impurity E	Impurity F	Impurity G	SVT	m/z = 421.2876	Impurity B	Impurity C
<i>Day 1</i> RSD% (n=6)	3.1	4.8	3.3	2.9	3.2	0.5	3.4	3.4	3.5
<i>Day 2</i> RSD% (n=6)	3.3	2.9	2.2	0.6	3.6	1.1	1.8	3.3	2.8
<i>Day 3</i> RSD% (n=6)	4.1	3.1	3.3	2.1	2.1	1.1	3.0	3.2	3.6
<i>Inter-day</i> RSD% (n=18)	6.1	30.8	4.8	5.6	5.5	1.6	4.7	5.1	5.4

Intra-day RSD% values between six replicated measurements of a same “simvastatin for peak identification” chemical reference substance solution were all comprised between 0.5 and 4.8 for each compound, simvastatin and main impurities. This indicated low instability of the mass spectrometer response when repeating measurements over a short period of time.

Similarly, the inter-day RSD% values between eighteen replicated measurements of the same solution were comprised between 1.6 and 6.1 for all solutes, except for Ph. Eur. impurity A (RSD% = 30.8). This conveyed, at the same time, low signal variability and solution stability.

Moreover, it was important to notice the highly elevated intermediate measurement precision value RSD% = 30.8 (between eighteen measurement readings) corresponding to Ph. Eur. impurity A signal. The Ph. Eur. impurity A peak areas increased dramatically over the three days of observation from  $7.74 \cdot 10^{+5}$  to  $14.68 \cdot 10^{+5}$  counts.s. This rise in peak areas, from one to two, resulted from simvastatin degradation by oxidation reaction over time leading to simvastatin hydroxy acid, as mentioned in the corresponding stability study of the marketing authorization dossier (confidential information non produced). Therefore, given the high level of variability introduced by the degradation of simvastatin into simvastatin hydroxy acid, it was subsequently of great use to dismiss information coming from this impurity in our classification model.

In addition, it was also wise to take into account the ratio of each impurity area to simvastatin area, expressed in percent, in order to minimize the degree of instrument variability, as it is suggested in table III-6 hereafter (see corresponding raw data in appendix C).

Table III-6: Intra-day (n=6) and inter-day (n=18) instrument precision considering internal area normalization

Compounds	m/z = 391.2406	Impurity A	Impurity E	Impurity F	Impurity G	m/z = 421.2876	Impurity B	Impurity C
<i>Day 1</i> RSD% (n=6)	2.7	4.9	3.0	2.9	3.2	3.4	3.1	3.4
<i>Day 2</i> RSD% (n=6)	3.5	3.5	2.0	1.0	2.7	1.3	3.6	3.3
<i>Day 3</i> RSD% (n=6)	3.2	3.4	2.2	1.2	1.8	2.0	2.2	2.6
<i>Inter-day</i> RSD% (n=18)	4.9	30.6	3.5	4.4	4.2	3.4	4.1	4.2

On the whole, intra-day RSD% values between the six replicated measurements of the same “simvastatin for peak identification” chemical reference substance solution decreased when considering internal area normalization. Besides, the inter-day RSD% values between the eighteen replicated measurements of the solution were also improved and comprised between 3.4 and 4.9 for all solutes excepting Ph. Eur. impurity A. These results were very interesting and essential to take into account in order to decrease the inherent variation induced by the instrument. Variability in detector response was particularly observed after a long period of non use or after maintenance, for example.

Accordingly, internal area normalization, corresponding to the impurity peak areas to simvastatin peak area ratio, was used in order to build the principal component analysis training model.

Once the description of the system optimization completed, our attention will be focused on the characterization of the experimental disposal in one hand, and on the presentation of the sample solutions preparation on the other hand, before outlining the set of results.

### **III.1.3 Experimental Disposal**

#### **III.1.3.1 Chemicals and Reagents**

All reagents used in this study were of analytical grade and were chosen with meticulous care for their high degree of purity. Acetonitrile gradient grade for chromatography (ref: 34851 / batch: 9131A), methanol gradient grade for chromatography (ref: 34885 / batch: 92105) and uracyl (ref: U0750-5G / batch: 098K0165) were purchased from Sigma Aldrich (Sigma Aldrich Co., Steinheim, Germany). Formic acid 98 – 100 % (ref: 1.00264.1000 / batch: K39331464847) and ammonium acetat (ref: 1.01116.1000 / batch: A973616838) were obtained from Merck (Merck and Co., Darmstadt, Germany). The ESI – TOF Tuning mix solution with ten different standards (ref: G1969-85000 – Serial Nr: LB53883) and the ESI – TOF reference solutions intended to calibrate with exact masses the quadrupole - time-of-flight mass analyzer, containing Purine 5mM in a mixture of acetonitrile/water 90:10 (v/v) (ref.: I8720242 – Serial Nr: LB53605) and HP-0921 2.5 mM solution, corresponding to hexakis(1H, 1H, 3H-tetrafluoropropoxy) phosphazine in acetonitrile (ref.: I8720241 / Serial Nr: LB53604), were obtained from Agilent (Agilent Corporation, Santa Clara, USA). Chemical reference substance (CRS) of simvastatin for peak identification (ref: 32332 / batch: B070717) was provided by the European Directorate for the Quality of Medicines (EDQM, Strasbourg, France). The finished products and their corresponding active pharmaceutical ingredients were obtained directly from the respective manufacturers. All aqueous solutions were prepared with ultrapure water supplied by a Maxima Analytica system from USF Elga (Elga Inc. Northbrook, USA). The water delivered by that production system entailed concomitant specifications like, a measured resistivity not less than 18 M $\Omega$ .cm, a total organic carbon value less than 10 ppb and a final ultra filtration assured by a 0.05  $\mu$ m filter. The 0.45  $\mu$ m GHP Acrodisc® syringe filters (ref: 4556T / batch: 21733074), containing a hydrophilic polypropylene GH polypro membrane and used during the sample preparation, were purchased from Pall Life Sciences (Pall Life Sciences, Ann Arbor, USA).

#### **III.1.3.2 Material and apparatus**

The weighing of the chemicals was realized on an analytical semi-micro balance Sartorius (Sartorius AG, Goettingen, Germany), Genius series model ME215P (Serial Nr: 14707429). The pH of the solutions was measured with a pH-meter Beckman (Beckman Instruments Inc, Fullerton, USA), model 360 (ref.: 511212 / Serial Nr: 1286) associated to a combination electrode Beckman (ref.: 511080 / Serial Nr: S112D).

Chromatographic analysis were performed using an Agilent HPLC system (Agilent Corporation, Santa Clara, USA), model 1200 Series RRLLC, rapid resolution liquid chromatography, equipped with a binary gradient pump SL (ref.: G1312B – Serial Nr: DE63058385), an online degasser (ref.: G1379B – Serial Nr: JP73107608), an autosampler HIP-ALS SL (ref.: G1367C – Serial Nr: DE64557270), a cooling/heating module FC-ALS Therm, from 4°C to 40°C, using Peltier elements (ref.: G1330B – Serial Nr: DE60562911), a thermostatted column compartment TCC SL (ref.: G1316B – Serial Nr: DE60558165) and an ultraviolet-visible diode array detector DAD SL (ref.: G1315C – Serial Nr: DE73457383). Chromatographic separations were operated on a 50 mm x 2.1 mm Kinetex™ octadecyl reversed phase analytical column, filled with 2.6 µm fused core particles, purchased from Phenomenex (Phenomenex, Inc., Torrance, USA).

The HPLC system was hyphenated in series to an Agilent 6520 AA QTOF mass spectrometer detector illustrated in figure III-25 (ref.: G6520A – Serial Nr: US74420206). The very low pressure of the mass spectrometer was provided by a high vacuum pump from Edwards (Edwards Ltd, Crawley, England), model E2M28 (ref.: A373-24-930 - Serial Nr: 086005119). The high purity nitrogen, above 99.999 %, for collision cell gas, drying gas, nebulizing gas and used to pressurize the calibrant delivery system, was supplied by a nitrogen gas generator from Granzow (Granzow AB, Enköping, Sweden) model N2MID600 EDB-CXE (ref.: 636280256 – Serial Nr: 0500302). Different types of ion sources were at disposal in the laboratory, a Dual ESI source (ref.: G3251B – Serial Nr: US75000231) and an APCI source (ref.: G1947B – Serial Nr: US80300355). The monitoring of the instruments and the acquisition of the data were performed by using the Agilent MassHunter Workstation Acquisition software version B.02.01, while the Agilent MassHunter Workstation Qualitative Analysis software version B.03.01 was dedicated to the processing of the analysis data.

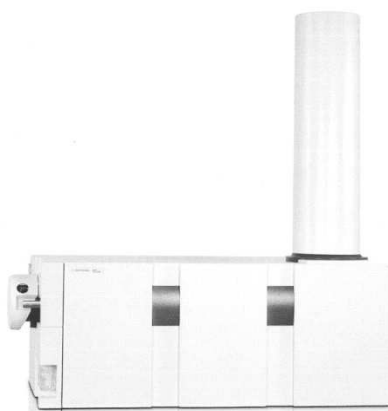


Figure III-25 : Agilent 6520 AA QTOF [30]

The automated searches on multiple compounds and empirical formula from the Agilent MassHunter Qualitative Analysis software allowed the easy identification of the analytes of interest by using an in-house simvastatin component library, created with the Agilent METLIN Personal Metabolite Database for MassHunter Workstation software, version B.02.00 (Agilent Corporation, Santa Clara, USA). ChemDraw Ultra 11.0 from CambridgeSoft (CambridgeSoft Corporation, Cambridge, USA) was used to define correct molecules stereochemically structures and determine information about Log  $P_{\text{octanol/water}}$  and mass weight. The molecule exact masses were calculated with the Isotope Distribution Calculator from Agilent MassHunter Workstation Data Analysis Core software (version 3 1.346.6). The data exploration, leading to the discrimination between different groups of samples, was carried out with the help of the Umetrics Simca P+ 12 software (Umetrics AB, Umeå, Sweden).

### **III.1.3.3 Preparation of sample solutions**

In order to protect the solutions from light degradation, amber color glassware was used for the preparation of the samples. Storage of the solutions were realized at 2 – 4 °C (particularly, during inter-day precision investigation).

#### **III.1.3.3.1 Solution of simvastatin for peak identification chemical reference substance**

The chemical reference substance “simvastatin for peak identification” corresponded to a mixture of the active pharmaceutical ingredient simvastatin spiked with its related impurities, which are specified and described in the European Pharmacopoeia monograph on simvastatin (04/2009:1563). The mixture contained Ph. Eur. impurity A (simvastatin hydroxy acid), Ph. Eur. impurity B (simvastatin methyl ester), Ph. Eur. impurity C (dehydro simvastatin), Ph. Eur. impurity D (simvastatin dimer), Ph. Eur. impurities E and F, which are stereoisomers (respectively lovastatin and epilovastatin), and Ph. Eur. impurity G. The preparation of the solution consisted in dissolving 5 mg of simvastatin for peak identification CRS in 5,0 mL of a mixture of acetonitrile and ultrapure water in the proportion 40:60 (v/v) followed by alternating sonication with mixing for 5 minutes, before filtering through 0.45  $\mu\text{m}$  Acrodisc® and transferring approximately 2 mL in an adequate amber color HPLC vial. This solution was particularly used for, firstly, the identification of the specified impurities and secondly, the optimization of the chromatographic separation and the adjustment of the mass spectrometer parameters.

#### **III.1.3.3.2 Starting material solutions**

Transfer 25.0 mg accurately weighed of simvastatin substance to be examined in a 25.0 mL volumetric flask. Dissolve by adding 20 mL of a mixture of acetonitrile and ultrapure water in the proportion 40:60 (v/v). Sonicate for 5 minutes. Cool to room temperature and complete with the same solvent mixture to volume. Mix for 5 minutes. Filter through 0.45 µm Acrodisc® and transfer approximately 2 mL in an adequate amber color HPLC vial.

#### **III.1.3.3.3 Finished product solutions**

Grind finely 10 tablets of the finished product to be investigated in a mortar. Then transfer an accurately weighed portion of the fine powder, equivalent to about 25.0 mg of simvastatin to a 25.0 mL volumetric flask. Add 20 mL of a mixture of acetonitrile and ultrapure water in the proportion 40:60 (v/v) and sonicate for 5 minutes. Cool to room temperature and complete with the same solvent mixture to volume. Mix for 5 minutes. Filter through 0.45 µm Acrodisc® and transfer approximately 2 mL in an adequate amber color HPLC vial.

#### **III.1.3.3.4 Blank and Placebo solutions**

The Blank solution corresponded to a mixture of acetonitrile and ultrapure water in the proportion 40:60 (v/v), filtered through 0.45 µm Acrodisc® and transferred in a 2 mL HPLC vial.

The Placebo solutions were prepared by dissolving 200.0 mg, accurately weighed, of placebo mixture containing all the excipients, i.e. all the components entering in the finished product composition, except the active pharmaceutical ingredient, in a 20.0 mL volumetric flask. Add 15 mL of a mixture of acetonitrile and ultrapure water in the proportion 40:60 (v/v) and sonicate for 5 minutes. Cool to room temperature and complete with the same solvent mixture to volume. Mix for 5 minutes. Filter through 0.45 µm Acrodisc® and transfer approximately 2 mL in an adequate HPLC vial.



### **III.1.3.4 Analytical conditions**

The purpose of this study was to determine to which extent the correlation between the active pharmaceutical ingredients impurity profiling and corresponding finished product impurity profiling could be characterized by using chemometrics. The technique employed to establish the impurity fingerprinting of both, APIs and finished products, was high performance liquid chromatography hyphenated, through an electrospray atmospheric pressure ionization source, to a mass spectrometer in tandem, using a hybrid quadrupole – time-of-flight analyzer. In addition to the information collected with the mass spectrometer detector, data were retrieved from the UV-DAD signal and capitalized as supplementary information in order to clearly identify, by their specific retention times, the known components, i.e. simvastatin and its reported impurities described in the European Pharmacopoeia monograph. UV spectra of unknown compounds were also useful to confirm or deny their similarity in molecular structure with simvastatin and to establish a related connection with the parent drug molecule.

The next two parts of this chapter describe the liquid chromatography and the mass spectrometry experimental conditions.

#### **III.1.3.4.1 HPLC experimental conditions**

All the chromatographic parameters chosen to achieve the most efficient separation of the components were subject of discussion in previous paragraph III.1.1 of this dissertation. The high performance liquid chromatography method was developed using a partially porous Kinetex™ octadecyl column, 50 mm x 2.1 mm, packed with 2.6 µm core shell particles. A twenty minutes long, binary, multi-segmented linear gradient was implemented, including two isocratic steps with increasing proportion of organic modifier, optimized at 42% from 0 to 6.5 minutes, and 53% from 7 to 11 minutes. The binary gradient elution system was composed of formic acid 0.1% in ultrapure water, as mobile phase A, and formic acid 0.1% in acetonitrile, as mobile phase B. It also comprised a washing step of the column, by using a high proportion of organic modifier, 87.5% from 14 to 17 minutes, in order to elute the most hydrophobic components and remove them from the stationary phase. A three minutes-long step permitted finally the system to re-equilibrate to the starting conditions.

The flow rate was set at a value of  $0.5 \text{ mL}\cdot\text{min}^{-1}$  and the temperature of the column was maintained at a fixed value of  $35 \text{ }^\circ\text{C}$ . The autosampler cooler were kept at a temperature of  $10 \text{ }^\circ\text{C}$ , and a fixed volume of  $5 \text{ }\mu\text{L}$  was injected into the system. An injector program was developed in order to first, minimize sample carryover by cleaning, before injection, the outside of the sampling needle in a flush port with a wash solvent containing a mixture of acetonitile and ultrapure water in the proportion 40:60 (v/v), and second, deplete the delay volume by bypassing the damper and the sample loop after injection. Appendix D lists all the chromatographic parameters of the final generic method.

#### **III.1.3.4.2 Mass spectrometry experimental conditions**

All the parameters of the mass spectrometer were chosen and deemed satisfying after optimization experimentations described and argued in chapter III.1.1 of this dissertation. The ionization of the components was performed using a Dual electrospray ion source functioning in positive mode. The nebulization of the liquid solutions was carried out under a nitrogen gas current stabilized at a pressure value of 35 psi. To optimize the evaporation of the solvent droplets, the drying gas flow rate was set at  $11 \text{ mL}\cdot\text{min}^{-1}$  while the drying gas temperature was brought to  $300^\circ\text{C}$ . Capillary voltage and fragmentor voltage were respectively adjusted at 3100 V and 175 V to facilitate the molecular ion transferring. The instrument operated in auto MS/MS acquisition mode. The complete set of mass spectrometer adjustments is described and summarized in appendix E.

#### **III.1.3.5 Measurement protocol**

Five micro-liters of each sample solutions were injected into the chromatographic system in order to be analyzed. The acquisition of the ultraviolet signal as well as the mass spectrometric response was carried out using the MassHunter Acquisition software. The data were compiled and then analyzed off line through the MassHunter Qualitative Analysis software. Single and total wavelength chromatograms (TWC), as well as total ionic chromatograms (TIC) or extracted ion chromatograms (EIC) were displayed for investigation. Unexpected peaks emerging from the TIC were analyzed, for example, by extracting and comparing their mass spectra with simvastatin mass spectrum or by editing with the “Generate Formulas” algorithm a short list of the most likely associated formulas for each compound ion.

A compound identification based on both molecular formulas and retention times was then performed using the “Search library” algorithm. This algorithm compared the components detected in the examined mass chromatogram with the Metlin personal library list of compounds dedicated to simvastatin. The simvastatin Metlin personal library was a self made database, based on the results obtained after the injections, in the optimized LC-MS/MS QTOF system, of the chemical reference substance “simvastatin for peak identification” solution and completed with the LC-MS/MS investigation analysis (cf. chapter III.1.4.2). This library was made up of a list of 16 compounds comprising simvastatin and substances related to simvastatin, summarized in appendix A, including: the names of the chemicals, their empirical formulas, exact masses, retention times and molecular structures. The results were reported in a table as individual component peak areas (in counts.s) and mass error assessments (in ppm). Subsequently, ratios, expressed in percent, of each impurity peak area relative to simvastatin peak area were calculated from this table. The values obtained from this internal normalization were then used to fill in the multivariate data analysis matrix intended to build up the classification model.

### **III.1.4 Results**

By using the method developed in the SMPA laboratory, three different kinds of analytical information were obtained. First, the ultraviolet diode array detector provided signals which, once converted into chromatograms, could be compared to the chromatogram obtained from the European pharmacopoeia monograph’s chromatographic conditions. Second, the mass detection led to acquire either total ionic chromatograms or extracted ion chromatograms for each species. Especially, all or few individual extracted ion chromatograms could be selected and merged into a single chromatogram, easier to inspect. And last, MS/MS experiments were performed in order to get structural information, particularly about unknown compounds detected in the sample mixture solutions. The structure elucidation was based on the MS instrument capacity to provide accurate mass measurements and to generate molecular formulas.

### III.1.4.1 UV-DAD chromatogram

A typical UV-DAD chromatogram of the chemical reference substance “simvastatin for peak identification” solution was performed by applying the optimized in-lab method. The chromatogram is reproduced in figure III-26 (upper diagram). The lower diagram stands for a representation of the corresponding gradient profile.

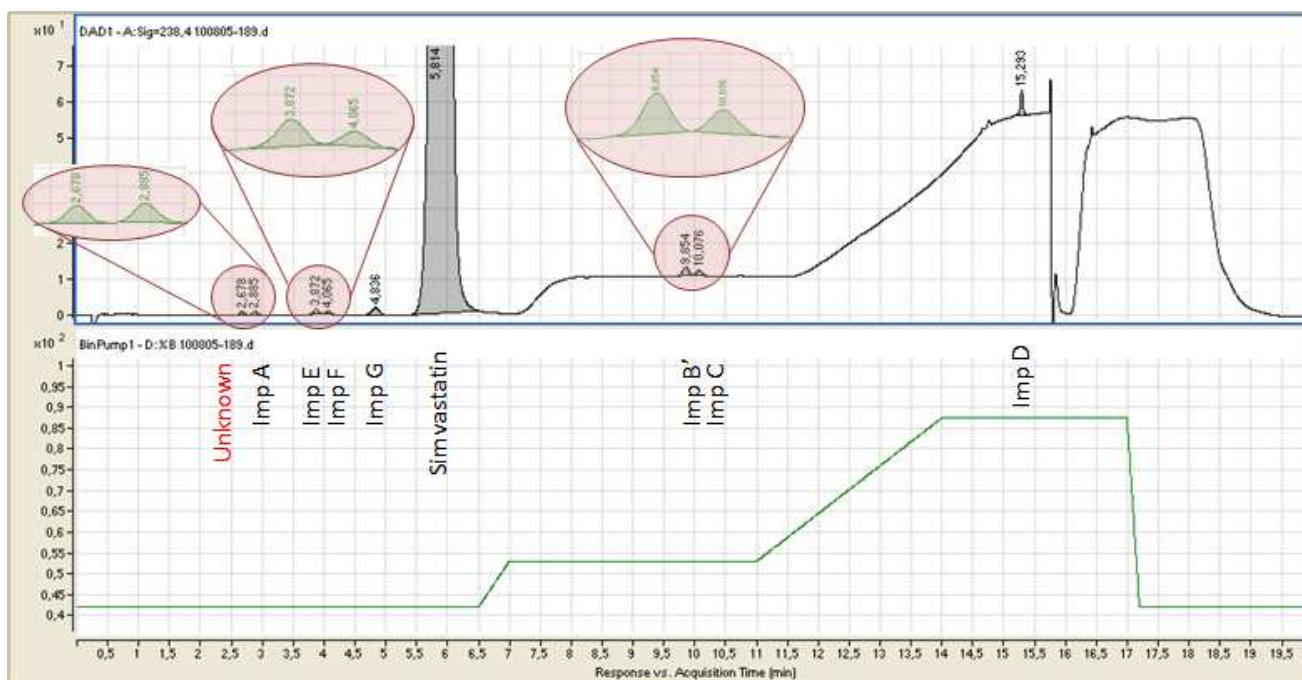


Figure III-26: UV-DAD chromatogram of the “simvastatin for peak identification” CRS solution (upper graphic) and gradient profile (lower graphic)

With the in-lab chromatographic conditions, the separations between Ph. Eur. impurities E and F, in one hand, and Ph. Eur. impurities B and C, in another hand, were considerably improved. The corresponding resolutions were respectively 1.2 between Ph. Eur. impurities E and F, and 1.3 between Ph. Eur. impurities B and C.

Moreover, a new impurity at a retention time close to the retention time of Ph. Eur. impurity A, which was henceforth designated as impurity A', could be detected and identified. The resolution between this unspecified impurity A' and Ph. Eur. impurity A was found equal to 1.6 when using the new method features, while a co-elution between both impurities is observed when applying the European Pharmacopoeia monograph conditions. Further investigations about the molecular structure of this new component were performed by realizing MS/MS experiments.

### III.1.4.2 Identification of new impurities by LC-MS/MS

The mass chromatograms obtained from the injections of blank solutions, finished product solutions and corresponding placebo solutions, are shown and analyzed in this chapter.

#### III.1.4.2.1 Example of a blank injection chromatogram

The injection of a blank solution aimed at determining if possible chromatographic peaks could be induced by the solvent eluent used for the sample preparation (see figure III-27). In the case of solvent eluent peaks, those latter should be simply dismissed when interpreting the impurity fingerprinting of the finished products.

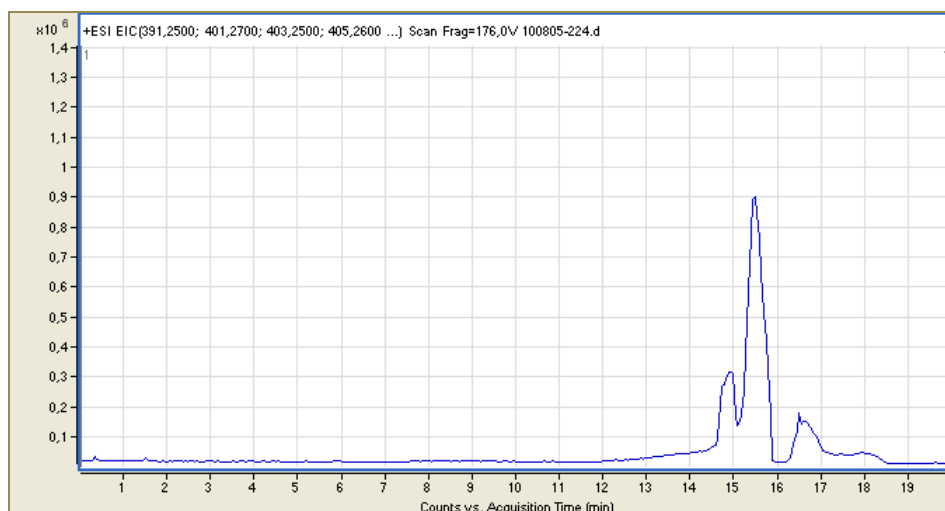


Figure III-27: Blank solution chromatogram

The chromatogram of the “Blank solution” shows that no interfering peaks coming from the solvent could be observed between 0 minute and 15 minutes.

### III.1.4.2.2 Example of a placebo injection chromatogram

The purpose of injecting a placebo solution consisted in defining all the chromatographic peaks specifically related to the finished product excipients (see figure III-28) so that they could be disregarded when analyzing the series of drugs.

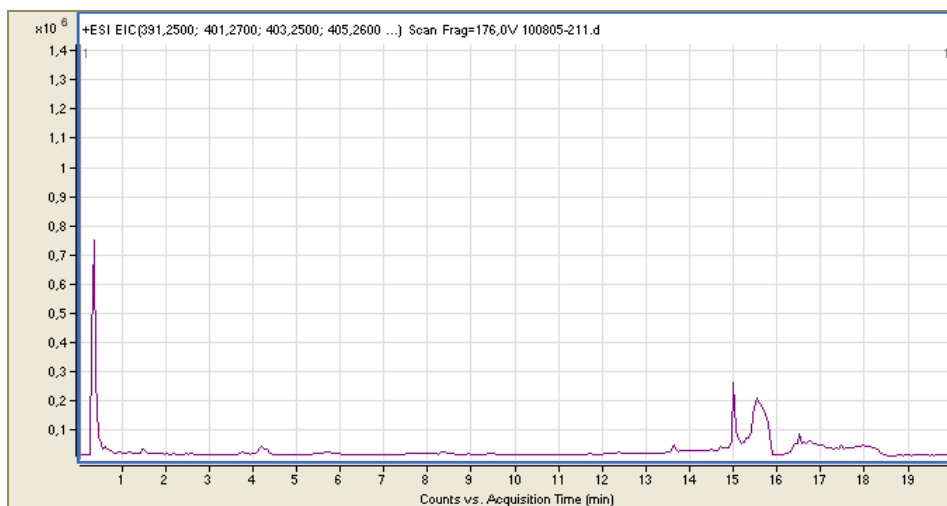


Figure III-28: Placebo solution chromatogram

As well as for the blank injection, the placebo injection chromatogram showed no noticeable interfering peaks, except at 0.274 minutes corresponding to the column hold-up time.

### III.1.4.2.3 Example of a finished product impurity profile

An example of a mass chromatogram attributable to the injection of a finished product solution is given in figure III-29. The chromatogram is a merger of a set of extracted ion chromatograms selected for different mass to charge ratios (compounds described in appendix A): 321.2060 (diol lactone), 435.2741, 433.2585, 391.2479, 437.2898 (simvastatin hydroxy acid), 403.2479, 405.2636 (lovastatin and epilovastatin), 417.2636 (impurity G), 419.2792 (simvastatin), 421.2949 (dihydro simvastatin, 465.3211 (simvastatin ethyl ester), 433.2949 (simvastatin methyl ether), 461.2898 (simvastatin methyl ester), 401.2686 (dehydro simvastatin) and 837.5511 (simvastatin dimer).

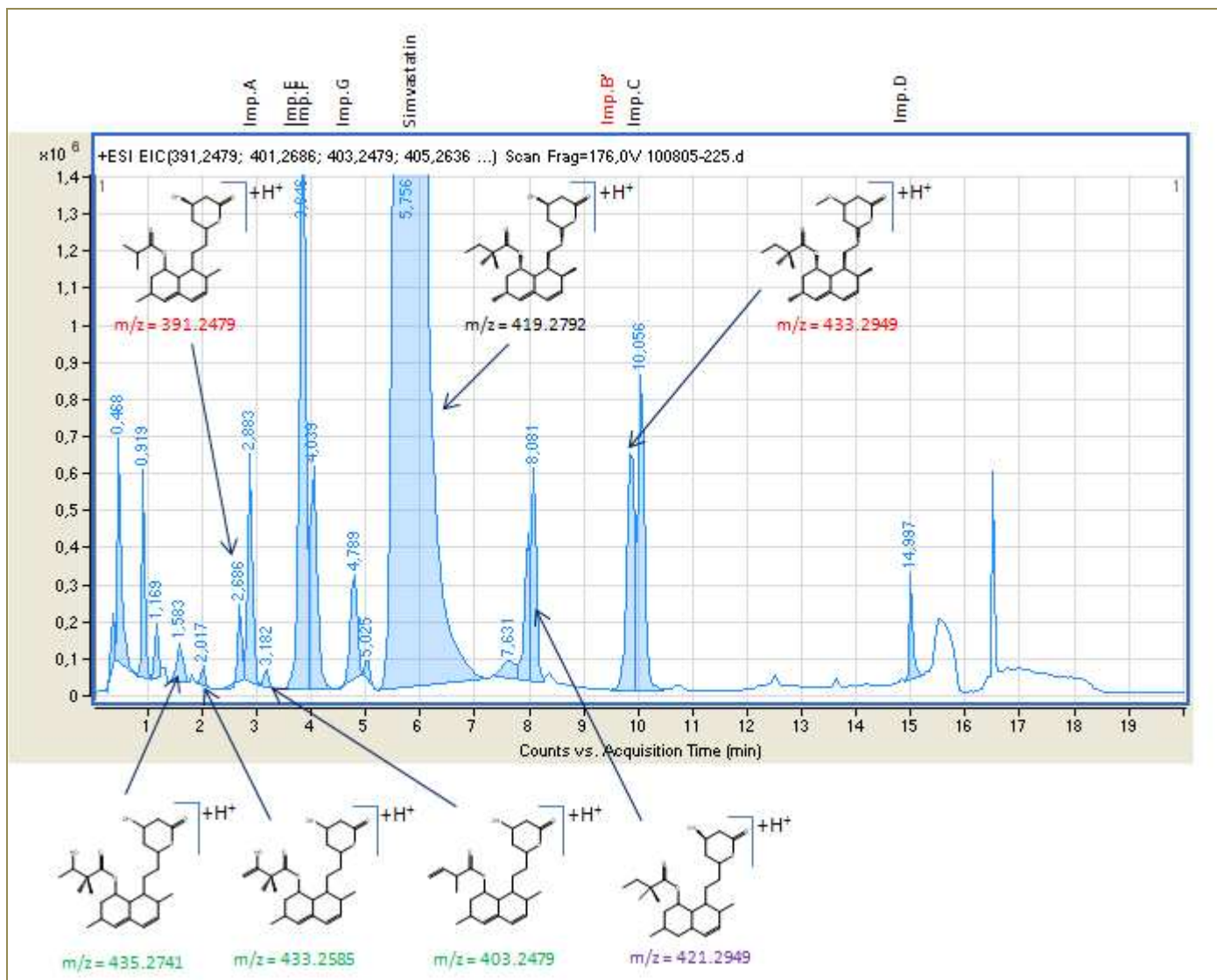


Figure III-29: Example of a finished product solution mass chromatogram

The chromatogram illustrated in figure III-29 demonstrated that the new method transposed from the European pharmacopoeia and using a mass spectrometric detection allowed, first of all, the detection of each specified impurity with a higher degree of sensitivity. Second, the analytical method enabled the detection and the characterization of additional simvastatin related substances (see part III 1.4.3 for the compound structure elucidation).

For example, impurity A', which co-eluted with Ph. Eur. impurity A, eluted then at a retention time of 2.686 minutes so that it was clearly separated from Ph. Eur. impurity A ( $t_R = 2.883$  minutes). This impurity presented a pseudo molecular ion ( $M+H^+$ ) with a measured mass to charge ratio equal to 391.2479. The empirical formula generated by the MassHunter Qualitative Analysis software, for this ion, was given as  $C_{23}H_{35}O_5$ , with a significant score of 82.25 (see table III-7) despite a relatively weak mass match percentage (63.64%). A low mass match value indicated a larger number of

empirical formulas proposed by the software for the corresponding mass to charge ratio. Nevertheless, some of the suggested formulas were containing elements like nitrogen or sulfur so that, given the corresponding UV spectrum, they could be easily dismissed.

On the other hand, the diagram demonstrated that an unknown compound presenting a molecular ion mass to charge ratio of 433.2949 appeared at a retention time of 9.869 minutes and was therefore separated from Ph. Eur. impurity C ( $t_R = 10.056$  minutes). However, this compound, that henceforth was designated as impurity B', eluted at a retention time quite close to Ph. Eur. impurity B ( $m/z_{M+H^+} = 461.2898 - t_R = 9.83$  minutes) as shown in figure III-30. The proposed diagrams represent a detailed investigation of each individual extracted ion mass chromatogram and overlaying for the three components B, B' and C.

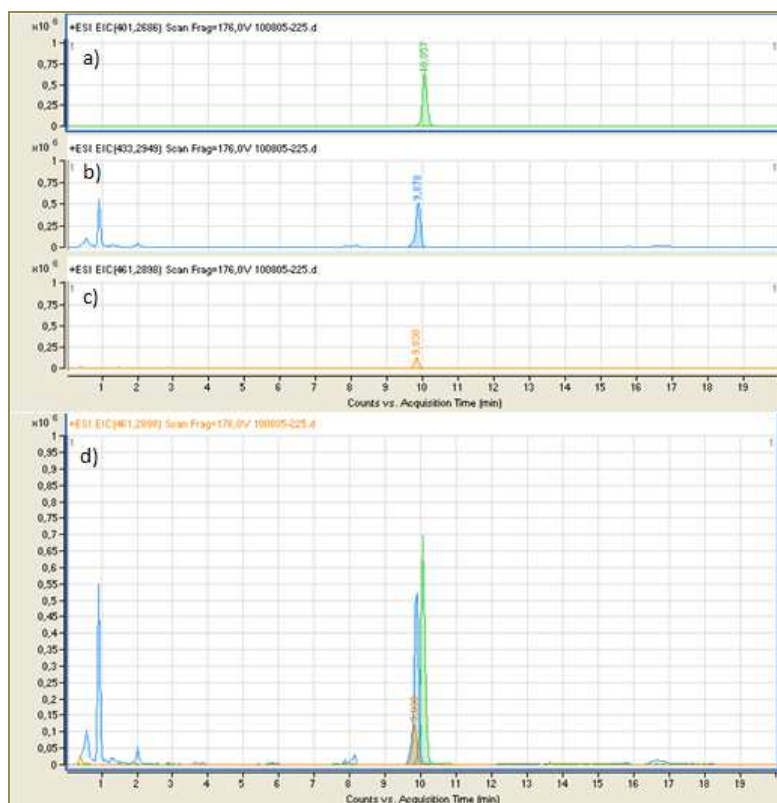


Figure III-30: a) Extracted ion chromatogram of impurity C  
b) Extracted ion chromatogram of impurity B'  
c) Extracted ion chromatogram of impurity B  
d) Overlaid extracted ion chromatograms of impurities C, B' and B



The fact that impurities B and B' overlapped led to a possible overestimation of the Ph. Eur. impurity B with both, the monograph and the in-lab methods, when using only ultraviolet detection, especially as impurity B' was the most abundant species.

Furthermore, the mass analyzer enabled also the observation of chromophoreless compounds, such as dihydro simvastatin ( $m/z_{M+H^+} = 421.2949 - t_R = 8.081$  minutes), which were not detectable with a traditional ultraviolet diode array detector. And finally, the presence of ultra trace level impurities was highlighted at several retention times, at 1.583 minutes ( $m/z_{M+H^+} = 435.2741$ ), 2.017 minutes ( $m/z_{M+H^+} = 433.2585$ ) and 3.186 minutes ( $m/z_{M+H^+} = 403.2479$ ). All the new detected impurities are summarized in table III-7, specifying their retention times, their calculated and measured mass to charge ratios, the induced mass errors and the corresponding generated molecular ion formulas, with score and mass match.

Table III-7: Unknown impurity information

Compound mass	$t_R$ (min)	$m/z_{M+H^+}$ calculated	$m/z_{M+H^+}$ measured	Mass error $\Delta m$	Generated Formulas (M+H <sup>+</sup> )	Score	Mass match (%)
434.3	1.606	435.2741	435.2736	1.19	C <sub>25</sub> H <sub>39</sub> O <sub>6</sub>	89.28	98.66
432.3	1.991	433.2585	433.2590	1.17	C <sub>25</sub> H <sub>37</sub> O <sub>6</sub>	78.65	98.69
390.2	2.676	391.2479	391.2507	7.22	C <sub>23</sub> H <sub>35</sub> O <sub>5</sub>	82.25	63.64
402.2	3.188	403.2479	403.2483	0.87	C <sub>24</sub> H <sub>35</sub> O <sub>5</sub>	99.14	99.33
420.3	8.082	421.2949	421.2960	2.71	C <sub>25</sub> H <sub>41</sub> O <sub>5</sub>	96.6	93.38
432.3	9.894	433.2949	433.2959	2.37	C <sub>26</sub> H <sub>41</sub> O <sub>5</sub>	96.32	94.79

The investigation of the additional impurities detected with the more specific and sensitive LC-MS method were completed with a series of MS/MS experiments in order to propose molecular structures for those components.

### III.1.4.3 Structure elucidation of new impurities by LC-MS/MS

The structural elucidation information of the newly identified compounds was based on the MS/MS spectra acquired after fragmentation of the singular molecular ions in the collision cell. The mass spectrometry technology using a hybrid quadrupole - time of flight analyzer enabled very high mass measurement accuracy, proving mass errors mostly less than 5 ppm (cf table III-7).

The obtained fragmentation patterns were compared with that of simvastatin in order to confirm the connection between the impurities and simvastatin substance. The empirical formulas of each fragment ion were obtained from the MassHunter Qualitative Analysis software. This software generated molecular formulas by taking into account the isotopic patterns and more particularly, the isotopic relative abundances and the isotopic spacing of the compounds. The knowledge of exact masses combined to the knowledge of molecular formulas for each specific mass peak, contributed to the identification and the structure elucidation of the main fragment ions and molecular ions.

### III.1.4.3.1 MS/MS spectrum of simvastatin

The in-tandem mass-spectrum at 5eV collision energy of simvastatin molecule ( $m/z = 419.2792$ ) and proposed fragment pathway are represented in figure III-31.

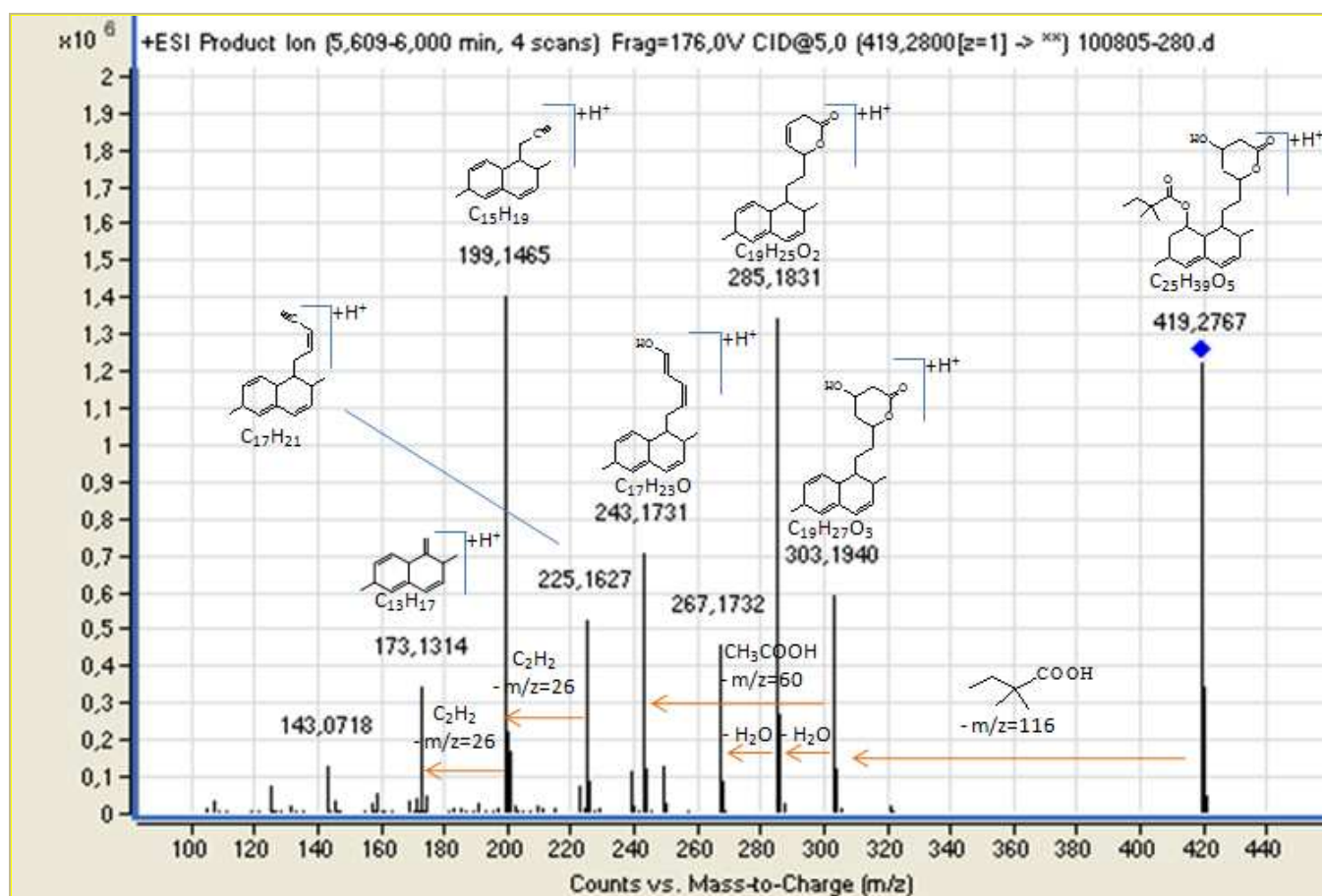


Figure III-31: Simvastatin in-tandem mass spectrum at 5 eV collision energy

Simvastatin MS/MS spectrum featured a specific fragmentation pattern because of the presence of major fragment ions at the following mass to charge ratios summarized in table III-8:

Table III-8: Simvastatin major fragment ions

Fragment ion m/z	Relative Abundance (%)	Proposed empirical formula	Score
173.1314	24.38	C <sub>13</sub> H <sub>17</sub>	98.09
199.1465	100	C <sub>15</sub> H <sub>19</sub>	94.39
225.1627	37.31	C <sub>17</sub> H <sub>21</sub>	97.50
243.1750	50.30	C <sub>17</sub> H <sub>23</sub> O	97.85
267.1732	32.70	C <sub>19</sub> H <sub>23</sub> O	98.37
285.1831	95.59	C <sub>19</sub> H <sub>25</sub> O <sub>2</sub>	96.72
303.1940	42.24	C <sub>19</sub> H <sub>27</sub> O <sub>3</sub>	97.82

The first transition, 419.2767 m/z → 303.1940 m/z, resulted from the neutral loss of a 2,2-dimethyl butanoic acid molecule (m/z = 116), characterizing the ester moiety of simvastatin molecule.

Afterwards, the fragment ion 303.1940 m/z underwent two different fragmentation processes on its lactone ring. First it might generate a fragment ion at 285.1831 m/z by neutral loss of a water molecule (m/z = 18) and second, it might generate a fragment ion at 243.1731 m/z due to the neutral loss of an ethanol molecule (m/z = 60).

Similarly, fragment ion located at 285.1831 m/z underwent either a neutral loss of water (m/z = 18) or a neutral loss of ethanol (m/z = 60) into respectively fragment ions 267.1732 m/z and 225.1627 m/z. Then, the latter mass ion was subjected to successive fragmentations, by neutral losses of acetylene (m/z = 26), into fragment ions 199.1465 m/z and 173.1314 m/z.

#### III.1.4.3.2 MS/MS spectrum of impurity A'

The structure elucidation of impurity A' (m/z = 391.2479) and fragment pathway proposal were based on the interpretation of the in-tandem mass spectrum of this compound at 10 eV collision energy, as represented in figure III-32.

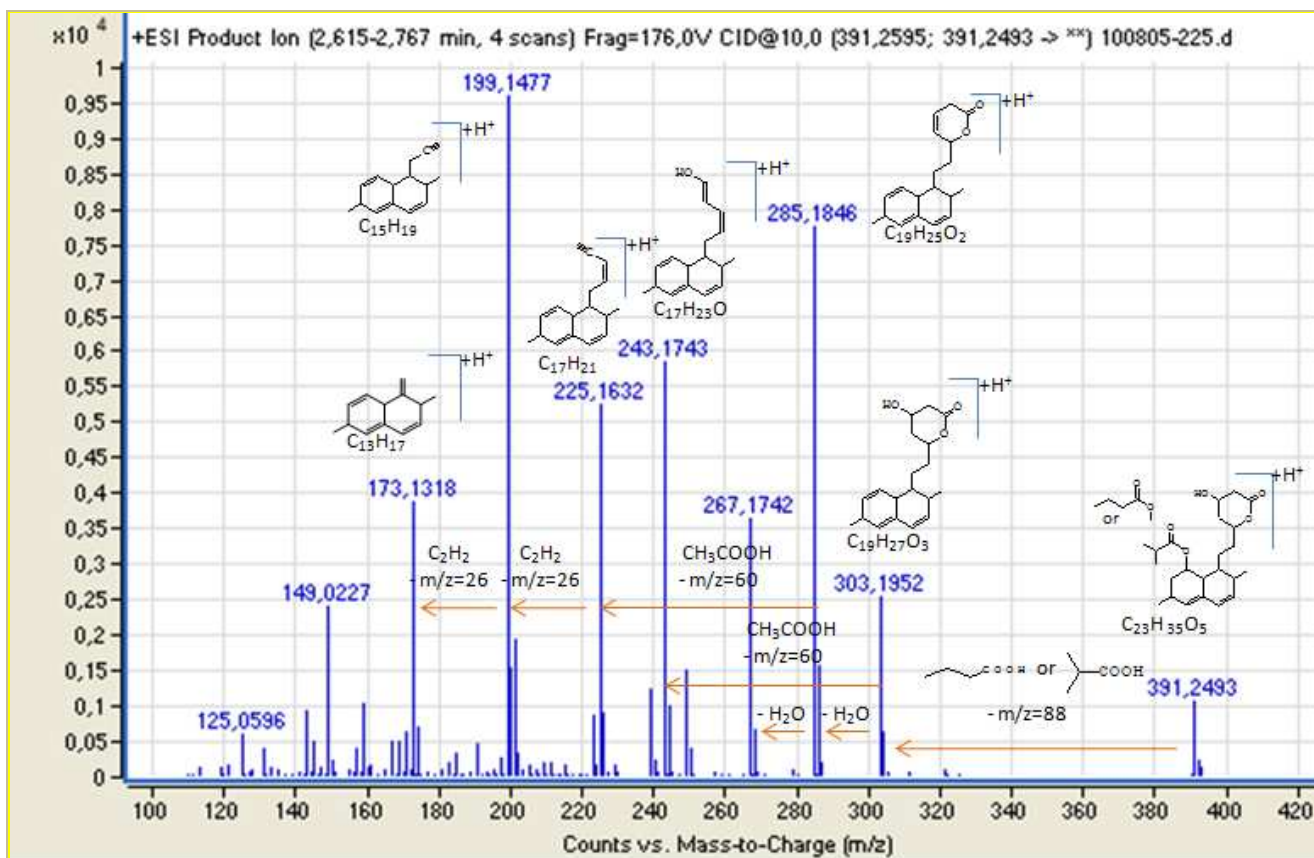


Figure III-32: Impurity A' in-tandem mass spectrum at 10 eV collision energy

The MS/MS spectrum of impurity A' showed up quite identical fragment ion data as that of simvastatin, demonstrating unambiguously the link between both components. The numerous common ion mass peaks corresponded to the lactone ring structure and the double ring structure of simvastatin.

The major difference in the fragmentation pattern lied in the first transition, i.e. from the pseudo molecular ion, at mass to charge ratio of 391.2493, to the ion at mass to charge ratio of 303.1952. The MassHunter Qualitative Analysis software generated formulas for those both ions as  $C_{23}H_{35}O_5$  and  $C_{19}H_{27}O_3$  respectively, so that this transition stood for the neutral loss of a molecule presenting  $C_4H_8O_2$  as elemental composition ( $m/z = 88$ ). Considering the number of oxygen atoms and the degree of unsaturation confirming the presence of a single double bond, the chemical structure of this molecule was strongly related to the ester moiety of impurity A'. Indeed, two different skeletal isomers might be assigned to this species: a long carbon chain structure corresponding to butanoic acid, or a branched carbon chain structure corresponding to isobutyric acid.

In conclusion, two molecular structures might be suggested for impurity A', (see table III-9):

Table III-9: Proposed molecular representations and IUPAC names for Impurity A'

	Structure I	Structure II
Molecular representation		
IUPAC Name	(1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetrahydro-2H-pyran-2-yl] ethyl] -3,7-dimethyl -1,2,3,7,8,8a-hexa-hydro naphthalen-1-yl-butanoate	(1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl] ethyl]-3,7-dimethyl -1,2,3,7,8,8a-hexa-hydro naphthalen-1-yl-2-methyl-propanoate

### III.1.4.3.3 MS/MS spectrum of impurity B'

Similarly, the structure elucidation of impurity B' ( $m/z = 433.2949$ ) and reaction pathway proposal were determined from the interpretation of the in-tandem mass spectrum of this compound at 5eV collision energy, as represented in figure III-33.

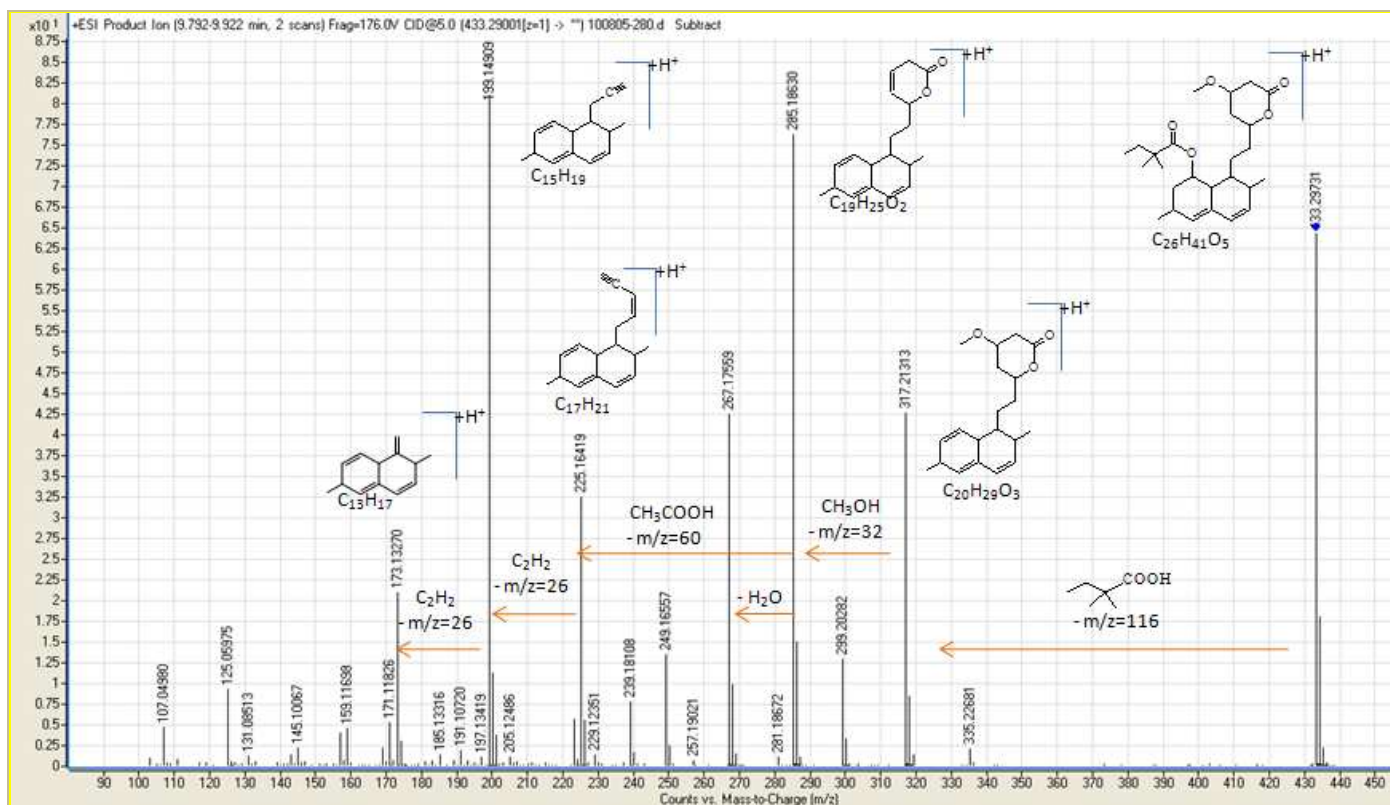


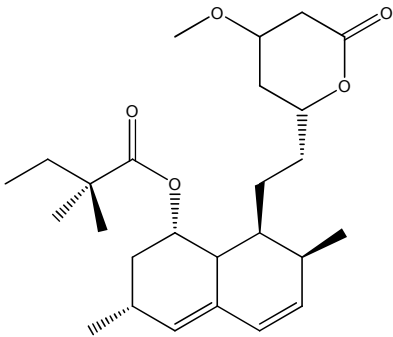
Figure III-33: Impurity B' in-tandem mass spectrum at 5eV collision energy

The chemical dissociation of impurity B' in the collision cell, arisen from a 5eV fragmentation energy, led to the specific mass signature related to simvastatin major daughter peaks. Those ion peaks located at 173.1329 m/z (C<sub>13</sub>H<sub>17</sub>), 199.1482 m/z (C<sub>15</sub>H<sub>19</sub>), 225.1645 m/z (C<sub>17</sub>H<sub>21</sub>), 267.1752 m/z (C<sub>19</sub>H<sub>23</sub>O) and 285.1855 m/z (C<sub>19</sub>H<sub>25</sub>O<sub>2</sub>) were specific of simvastatin naphthalen and lactone ring structures (cf. table III-8).

Moreover, the first transition from molecular ion, located at 433.2973 m/z, to the fragment ion located at 317.2131 m/z, corresponded to a neutral loss of a 2,2-dimethyl butanoic acid molecule (m/z = 116). This neutral loss, already present in the MS/MS spectrum of simvastatin, was specific of simvastatin ester moiety.

Consequently, the main difference between impurity B' and simvastatin was observed for the second transition 317.2131 m/z → 285.1863 m/z which was characteristic of the branching on the lactone ring. As it corresponded to a neutral loss of methanol (m/z = 32), the following molecular structure was proposed for impurity B' (table III-10):

Table III-10: Proposed molecular representation and IUPAC name for Impurity B'

Molecular representation	
IUPAC Name	(1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-methoxy-6-oxo-tetrahydro-2H-pyran-2-yl] ethyl] -3,7-dimethyl -1,2,3,7,8,8a-hexahydronaphtalen -1-yl-2,2-dimethylbutanoate

Analogous deductive reasoning was applied for the interpretation of the MS/MS spectra obtained from fragmentation experiments of supplementary unknown compounds.

### III.1.4.3.4 Structure elucidation for impurities located at 435.2741 m/z, 433.2585 m/z, 403.2479 m/z and 421.2949 m/z

Structural elucidation information and fragment pathways, based on in-tandem mass spectra at 5eV or 10 eV collision energy of 4 molecular ions, located at 435.2741 m/z, 433.2585 m/z, 403.2479 m/z and 421.2949 m/z, are reproduced in appendices F and G. Inferred molecular representations and IUPAC names are given in table III-11 hereafter.

Table III-11: Proposed molecular representations and IUPAC names for unknown impurities located at 435.2741 m/z, 433.2585 m/z, 403.2479 m/z and 421.2949 m/z

Unknown Impurity	435.2741 m/z	433.2585 m/z
Molecular representation		
IUPAC Name	(1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetrahydro-2H-pyran-2-yl] ethyl] -3,7-dimethyl -1,2,3,7,8,8a-hexa-hydro naphthalen-1-yl-3-hydroxy-2,2-dimethylbutanoate	(1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetrahydro-2H-pyran-2-yl] ethyl] -3,7-dimethyl -1,2,3,7,8,8a-hexa-hydro naphthalen-1-yl-3-hydroxy-2,2-dimethylbut-3-enoate
Unknown Impurity	403.2479 m/z	421.2949 m/z
Molecular representation		
IUPAC Name	(1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetrahydro-2H-pyran-2-yl] ethyl] -3,7-dimethyl -1,2,3,7,8,8a-hexa-hydro naphthalen-1-yl-2-methylbut-3-enoate	(1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetrahydro-2H-pyran-2-yl] ethyl] -3,7-dimethyl -1,2,3,4,4a,7,8,8a-octa-hydro naphthalen-1-yl-2,2-dimethylbutanoate

### **III.2 Chemometric discrimination between different simvastatin API origins**

As mentioned earlier, the classification model intended to pinpoint the active pharmaceutical ingredient sources was based on powerful statistical techniques such as, principal component analysis and hierarchical clustering analysis.

These methods were of paramount importance in our study because of their ability to give an immediate and straightforward insight of the relationships and natural groupings in the huge and high dimensional dataset. More particularly, the graphical projection in a limited number of new components, called score plots, represented easy means to reveal the structure among the observations, or samples. Similarly, loading plots were used to highlight the correlation and the relevance of the variables.

Thus, the visual impression of clustering was usefully explored in order to distinguish between the different API origins, including the routes of synthesis or else the production sites. However, the desired objective to apply the method simultaneously to both, raw materials and finished products, implied some adjustments in the model construction. Indeed, the whole information provided by the mass spectrometer could not be incorporated as such, i.e. as the entire LC-MS ion chromatograms, into the model. For instance, information due to excipients entering in the drug formulation was susceptible to induce a bias, in the form of a gap, between the groups of active substances and finished products of an identical origin.

Excipients are pharmacologically inactive ingredients added during the drug preparation in order to stabilize the API (like coatings or preservatives), or to simplify the manufacturing process (lubricants and glidants) and to improve the hardness as well as the taste of the tablets (binders, sweeteners and flavours). Consequently, only proper information, like specific extracted ion chromatograms, common to both, raw substances and final medicinal products, and stated among the list of 15 related substances detected with the LC-MS/MS in-lab method, was selected and taken into account to develop the discriminatory analysis. Hence, it was necessary to conduct a development and a perfecting of the calibration model aiming at carrying out an appropriate variable selection and building-up a simple, but resolute and discriminating model. This optimization approach is described in the next paragraphs, as well as the final calibration model implementation and validation process.



In all, during the project course, over 60 samples, comprising 39 raw materials coming from 9 various API furnishers, and 21 simvastatin based medicines coming from 14 different manufacturers, were analyzed.

### **III.2.1 Development of the calibration model**

The purpose of the optimization step consisted in defining progressively a restricted number of relevant variables to include in the calibration model, and in establishing a simplified, but nevertheless discriminatory, multivariate statistical tool for the differentiation of 8 API furnishers among 28 observations representing 20 raw materials and 8 finished products. From each observation a matrix of 15 variables, corresponding to the impurities detected with the LC-MS in-lab method and listed in the appendix A, was generated. The relative peak areas of these impurities, expressed in percentage compared to the peak area of simvastatin, were used to fill in the data matrix. The aforementioned peak areas were obtained from the respective extracted ion chromatograms, as described in parts III.1.4.2.3 and III.1.5 above. All variables were pretreated by the SIMCA P+ 12 software in order to give them equal importance and weight. This preprocessing consisted first in a mean centering, i.e. variables were centered by subtracting the mean value to each data, and second in an autoscaling to unit variance, which referred to a homogenization of each variable contribution by dividing the centered values by the standard deviation.

The comparison between the different model performances was realized, first by visual inspection of the cluster separation and second, by considering the cross-validation results. Besides, improvements in the cluster resolution capacities were achieved by estimating the uncertainty of the loading calculations, on one hand, and by interpreting the contributions between the observations or groups of observations, on another hand.

Figure III-34 displays the initial PCA calibration model score scatter plot component 1 *versus* component 2 and the PCA calibration model score scatter plot component 1 *versus* component 3. Both were obtained from the data treatment of all the 15 variables collected and corresponding to the investigation of the 28 samples.

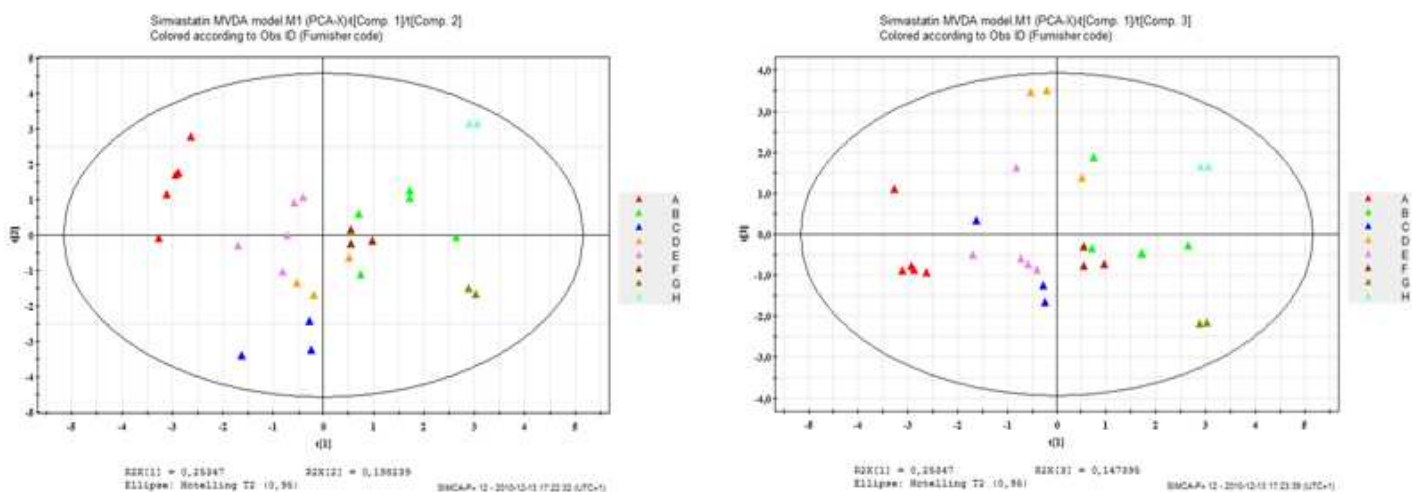


Figure III-34: Initial PCA calibration model score scatter plot component 1 *versus* component 2 (left) and PCA calibration model score scatter plot component 1 *versus* component 3 (right) built up with 15 variables

Each mark of the score scatter plots corresponds to an observation which might refer either to an active pharmaceutical ingredient or to a final medicinal product. The color code is related specifically to the API origins studied, which were designated as furnishers A, B, C, D, E, F, G and H in this study.

The cumulative fraction of the variation explained for this calibration model after 3 selected components rose to 59.8 % of the global data variation. The first component expressed 25.3 % of the variation, the second component 19.8 % of the variation and the third component expressed 14.7 % of the variation. A characteristic pattern might be already noticed for 3 out of 8 API furnishers (groups A, G and H). Nevertheless, sample spreading within a same furnisher group was observed, inducing sometimes clusters' overlapping (groups B, D, and F) and thus, bad discriminatory power.

In order to improve the cluster resolution and the percentage of explained variation, various tools such as cross validation or variable intra-group and inter-group contributions were helpfully exploited. Cross validation was used to test the significance of both the components and the variables. In cross validation, parts of the data were kept out of the model development and then successively predicted and compared with the initial values. Cross validation results are given in the form of two coefficients, R2VX and Q2VX. The representativeness R2VX measures the goodness of fit, i.e. how well the model fits the data after the selected component. A useful model should have a coefficient R2VX as large as possible, at least higher than 0.5. A value over 0.9 indicates excellent representativeness of the model.

The reliability Q2VX measures the goodness of prediction that can be attributed to the model. Q2VX should be higher than the value of R2VX minus 0.4. Figure III-35 displays the analysis by cross validation of the model containing 15 variables.

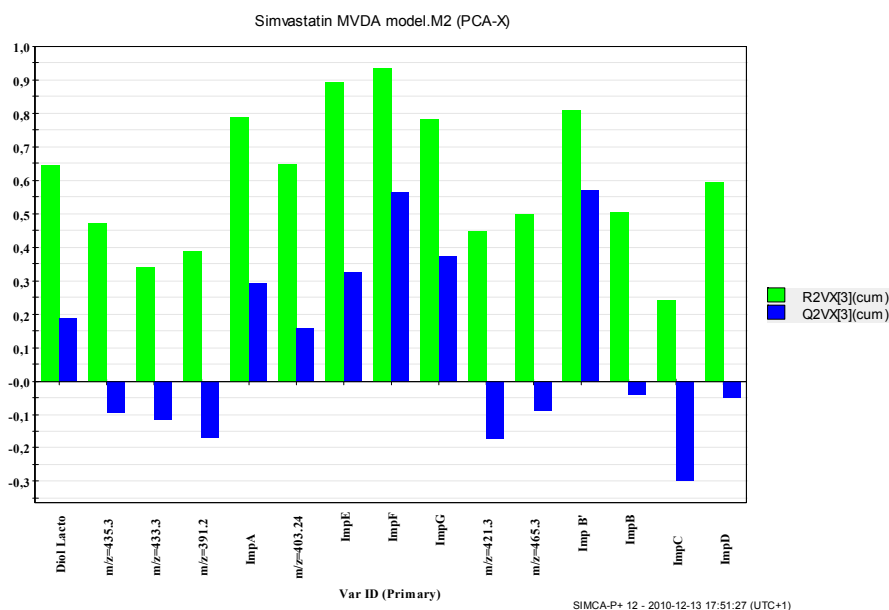


Figure III-35: Cross validation of the 15-variable model

The bar charts of representiveness R2VX and reliability Q2VX highlighted the irrelevance of some variables like, for instance, the variables corresponding to the 435.2741 m/z, 433.2585 m/z, 391.2479 m/z, 421.2949 m/z, 465.3211 m/z and impurity C ions. Those variables showed both, low goodness of fit and very low (negative) goodness of prediction. Impurities B and D demonstrated also bad prediction properties but had the advantage to present a coefficient of representiveness R2VX higher than 0.5, so that further tests like, for example, the contributions intra and inter groups, were necessary to consider.

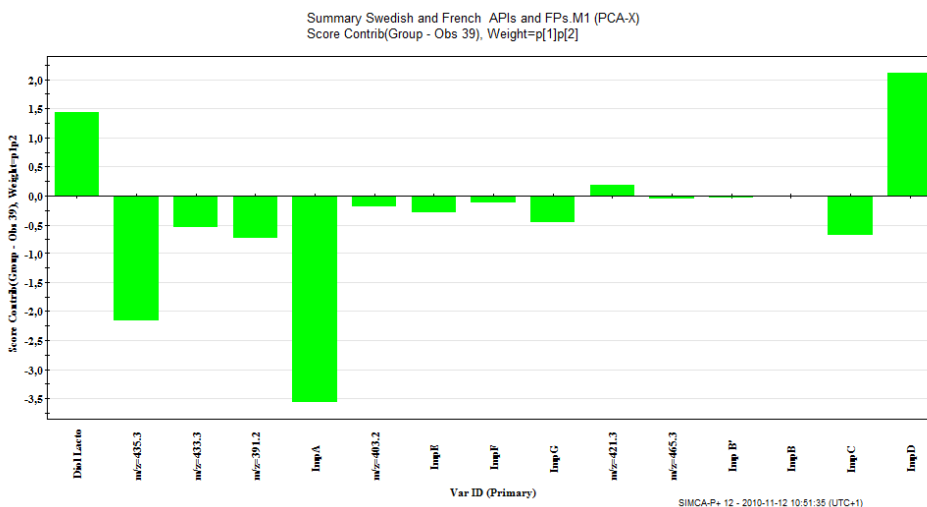


Figure III-36: Contribution intra group E in projection plane component 1 versus component 2

Figure III-36 is an example of a bar chart representation of the contribution within a group, group E in this case. The bars are representative, in this particular instance, of the differences induced by an observation corresponding to a finished product, compared to the rest of the group, composed exclusively of APIs. The higher the bars are, the more the cluster spreads.

Accordingly, the variables related to impurities Diol lactone, A, D, 435.2741 m/z and to a lesser extent impurities 433.2585 m/z, 391.2479 m/z, and C, generated too much dispersion and were then removed from the model. Similarly, considering the contributions inter groups led also to exclude irrelevant variables from the calibration model. An example of the inter group contributions between clusters D and E is given in figure III-37.

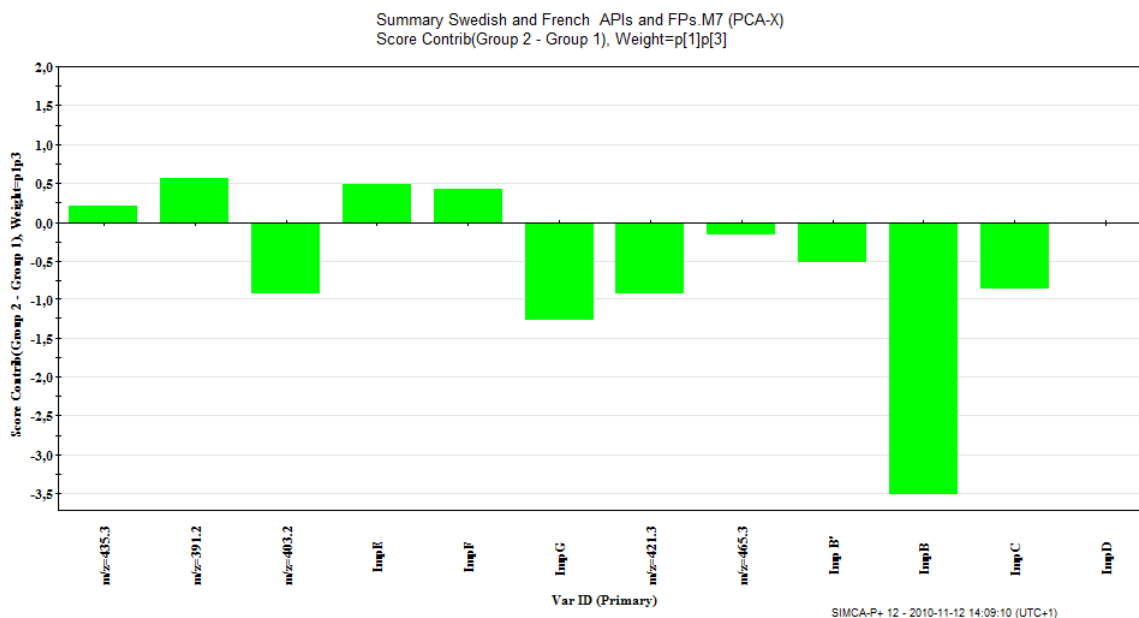


Figure III-37: Contribution inter groups D and E in projection plane component 1 *versus* component 3

This bar chart emphasized the role of the variables in the model discriminatory properties to separate 2 adjacent or 2 overlapping groups. For instance, the variable linked to impurity B was of great importance in the differentiation process between groups D and E. Therefore, impurity B was kept as a meaningful variable in the final calibration model.

To sum up, the optimization process of the calibration model was realized step by step, by eliminating one irrelevant variable after another, and by considering a set of information about the significance of the variable and the connection between the variables and the observations. The potential irrelevance of some variables was estimated by using different tools partly presented in this paragraph. Information derived from score scatter plots, cross validation analysis, bar charts of the contributions intra and inter groups, but also indication derived from loading plots and uncertainty of the loading calculations, were interpreted and collapsed in order to define the proper training model intended to discriminate between API origins in starting materials and finished products.

The following chapter is dedicated to the results obtained by applying this advanced model, and more particularly, the analysis of 5 samples of unknown origins will be presented. A validation process, covering the cross validation and an external validation testing, is also proposed in this next part.

## **III.2.2 Results**

### **III.2.2.1 Calibration model score scatter plots and associated loading scatter plots**

Score scatter plots and loading scatter plots of the previously developed PCA calibration model using optimal number of variables are depicted in figure III-38. This three-component PCA training model was constructed of 28 observations, containing information coming from the analysis of 20 starting materials and 8 finished products, and 6 variables composed of impurities E, F, G, B, B' and 403.2479 m/z. The data of the training set corresponded to the impurity relative area percentages compared to simvastatin area. Those data were auto scaled to unit variance prior to the classification analysis.

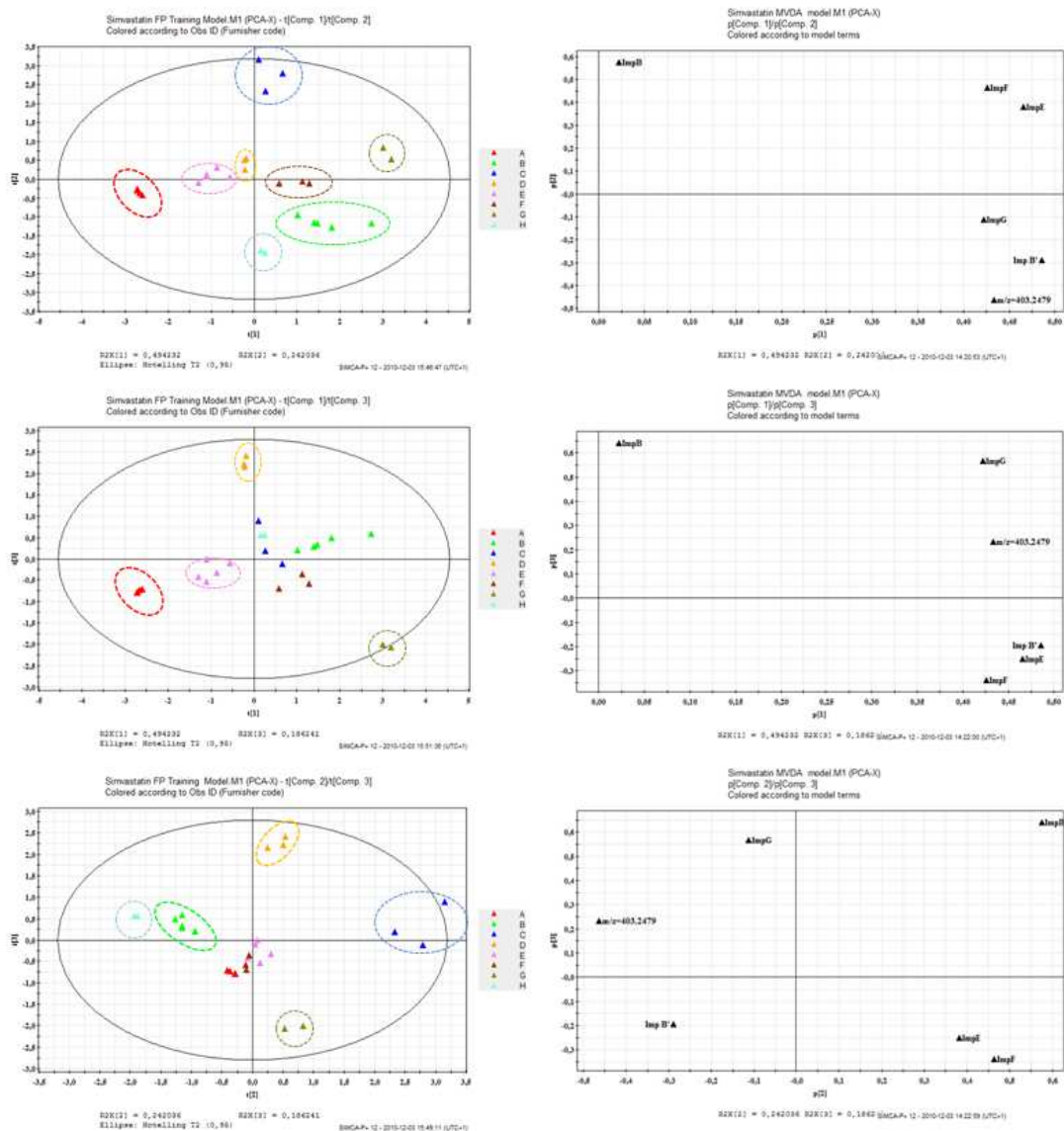


Figure III-38: Score scatter plots (left) and corresponding loading scatter plots (right) of the final API origin discriminating training model component 1 *versus* component 2 (upper), component 1 *versus* component 3 (middle) and component 2 *versus* component 3 (lower)

Henceforth, the three-component PCA training model explained cumulatively 92.2 % of the variation. More precisely, components 1, 2 and 3 accounted respectively for 49.4 %, 24.2 % and 18.6 % of the variation. It was noticeable that all of the eight API furnishers were unambiguously distinguished in the projection plane composed by components 1 and 2. However, in order to refine the PCA discrimination, it was possible to consider the two additional projection planes P1P3 and P2P3, respectively constituted by components 1 and 3, on one hand, and components 2 and 3, on another hand. The first aforementioned projection plane enabled to distinguish more specifically between the furnishers D and E whereas the second projection plane permitted to separate more particularly the furnishers B and F.

### III.2.2.2 Uncertainty of the PCA calibration model loading calculation

The loading scatter plots in figure III-38 show the correlation structure of the different variables within the model. Close variables were positively correlated and underwent simultaneous increase or decrease, like impurities E and F, for example, in projection plane P1P2 (upper diagram). Whilst variables opposite to each other, such as impurities B and B' in projection plane P2P3 (lower diagram), were negatively correlated, which meant that an increase of the first variable value was accompanied by a decrease of the second one.

The uncertainty of the loadings' calculation was an indicator of the variable relevance or irrelevance. It was given by the confidence interval derived from jackknifing as expressed in figure III-39.

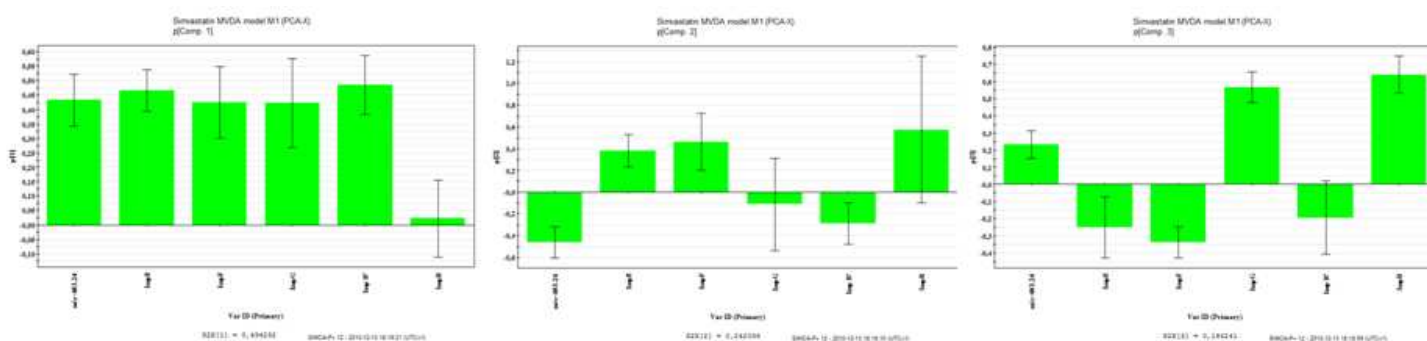


Figure III-39: Loadings and uncertainty of the loadings' calculation of the first component (left), the second component (center) and the third component (right)

The significance of variables E, F, G, B' and 403.2479 m/z was demonstrated for almost each principal component. The relevance of impurity B was particularly evidenced in principal component P3, and to a lesser extent in principal component P2.

### III.2.2.3 Validation

Two validation tests were implemented in order to estimate the significance of the predictive model. The first test consisted in a cross validation procedure included in the SIMCA P+ 12 software and the second test lied in a graphical simulation on a personal external prediction set.

#### III.2.2.3.1 Cross validation

Cross validation procedure implied that one or more observations at a time were held out from the sample set, then predicted and compared to the original values. Cross validation referred to representativeness R2VX and reliability Q2VX coefficients. The goodness of fit (illustrated as green bars in the chart below) and the goodness of prediction (blue bars) are given in figure III-40 for the calibration model used in this study.

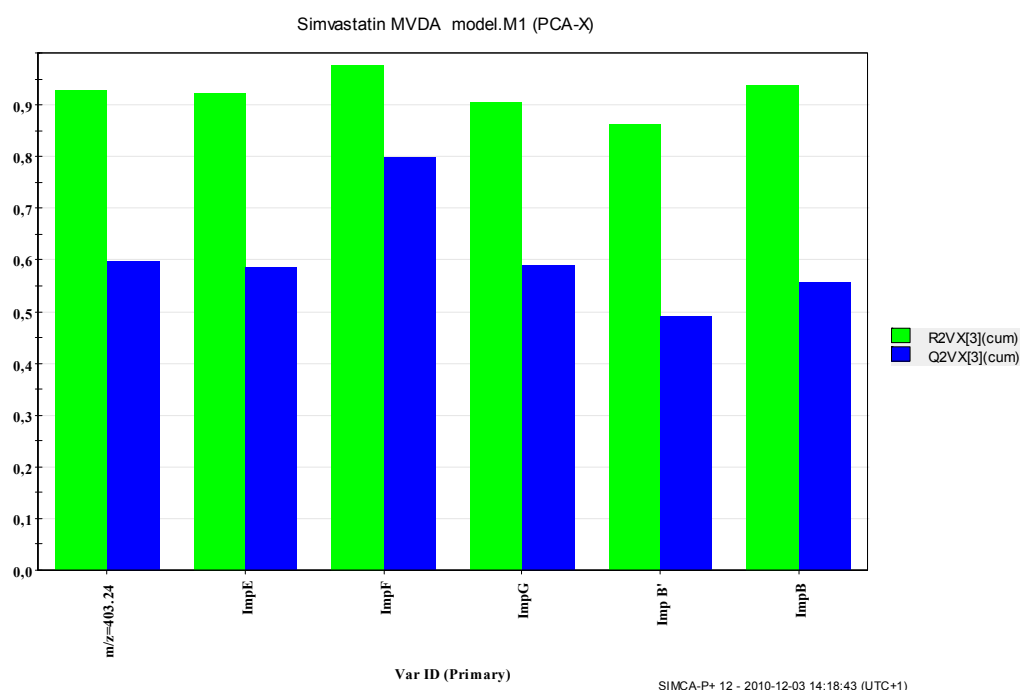


Figure III-40: Calibration model cross validation

With that enhanced 6 impurities calibration model, the variation explained cumulatively after 3 components more than 90.0 % for most of the variables, except for compound B' (about 87.0 %), and the predicted coefficients were all satisfyingly comprised between 50.0 and 80.0 %.



### III.2.2.3.2 Internal validation test

Figure III-41 represents the graphical visualization of the validation set prediction when applying the final PCA calibration model and executing the resulting hierarchical clustering analysis.

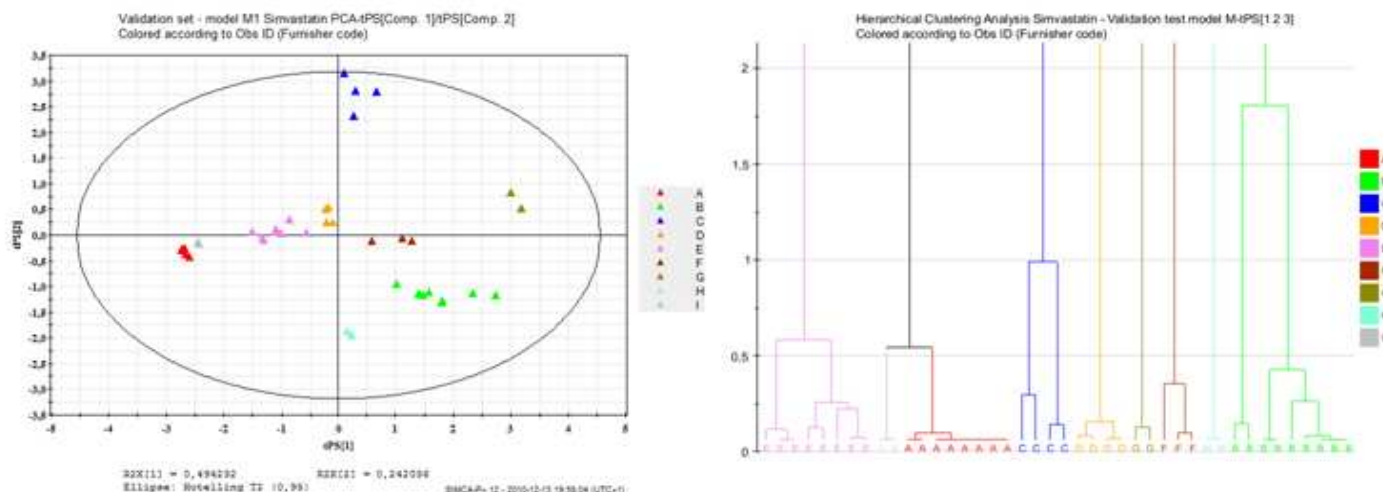


Figure III-41: PCA predicted validation set in projection plane P1P2 (left) and corresponding HCA three-dimensional predicted validation set (right)

The validation set encompassed 14 supplementary samples of starting materials and finished products among the 8 already known API furnishers. Two additional observations of a new API provider, named provider I, were also added in order to confirm the resolution power of the training model, even with samples of unknown origins.

Thus, nine non overlapping clusters could be distinctly observed corroborating, in that way, our expectations of discrimination abilities of the model to distinguish between several API sources. HCA is particularly adapted to detect dispersion within observations of a same group. Especially, clusters B and C, for instance, showed large spreading and batch to batch variation (high bars) when compared to groups A or D (low bars). Moreover, the compact aspect of group A allowed us to distinguish unambiguously between the clusters A and I despite of their proximity.

An interesting application of the training model in order to pinpoint the origins of five unknown samples is shown in the next part.

### III.2.2.4 Identification of API origins of unknown pharmaceuticals

The 6-variable training model was then applied to a test set of 5 unknown finished products (marked Ux in figure III-42) with the objective to predict their API origins.

Two of the unknown samples came from the Icelandic market (U1 and U2), one from the French market (U3) and another one from the Morocco market (U4). The last sample (U5) was sent by the Swedish customs to the SMPA laboratory for quality monitoring due to counterfeit suspicion, and was consequently used as test sample in our study with special interest.

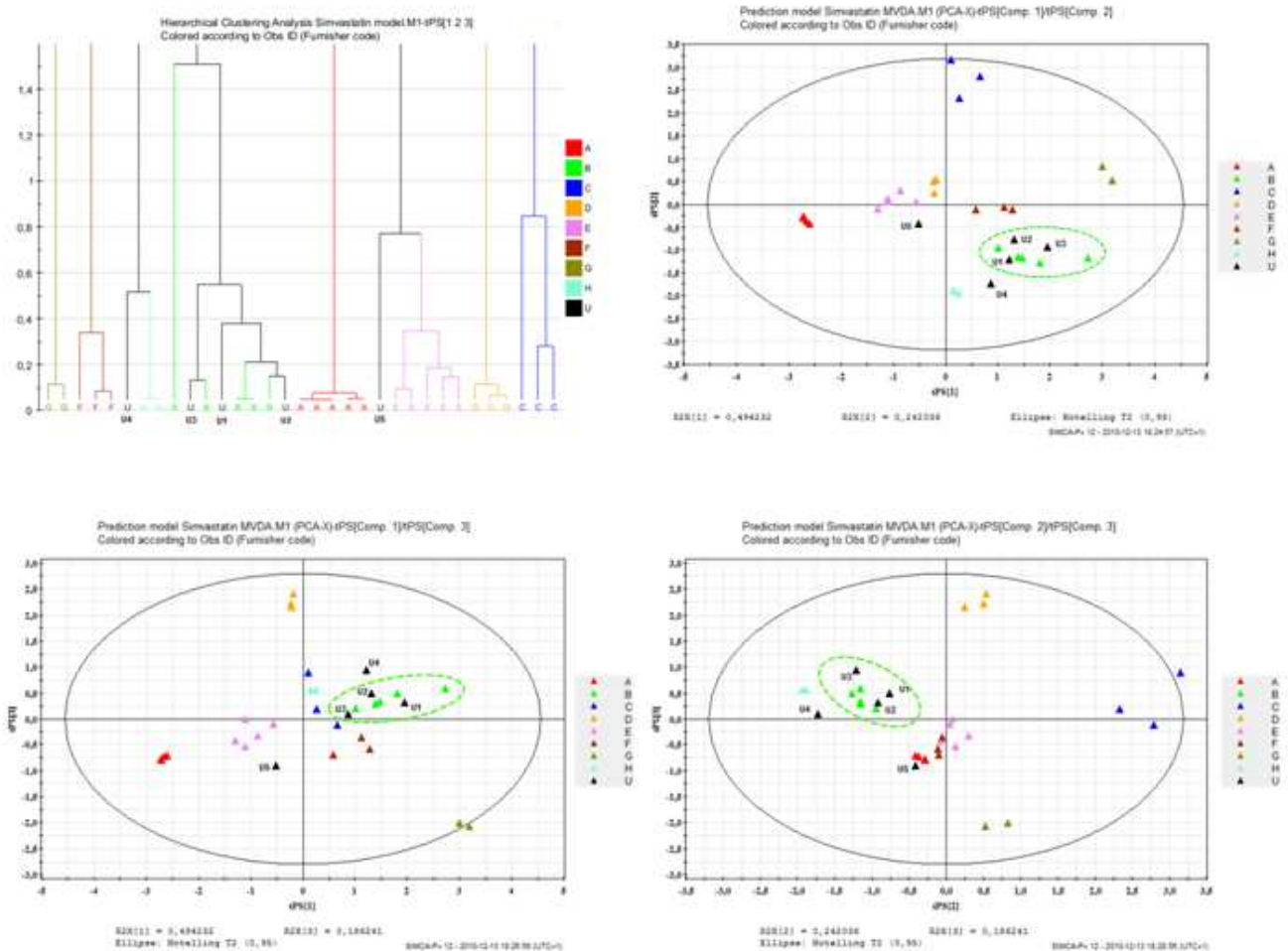


Figure III-42: Predictive three component HCA (upper left) and PCA models (projection planes P1P2 upper right, P1P3 lower left and P2P3 lower right) for 5 unknown samples

Proceeding to the combined PCA and HCA predicted experimentations in order to classify 5 unknown samples led us to conclude first that the Icelandic (U1 and U2) and the French (U3) samples were likely originated from provider B (zone illustrated by a green dotted circle). Second, the graphical inspection of PCA projection plane composed of component 2 versus component 3 was particularly conclusive and obvious to confirm that the Moroccan sample (U4) did not belong

to any of the 9 studied groups of API furnishers and constituted a new one. And lastly, PCA information tallied with HCA information to similarly classify sample U5 as coming from an unidentified API provider.

All of those hypotheses were confirmed after enquiry. Indeed, the Icelandic and the French finished drug products were effectively originated from provider B. The Moroccan sample was manufactured in India by an unknown furnisher and the sample sent by the customs was coming from an unidentified Thai manufacturer.

#### **IV. Discussion.**

The results achieved in this project confirmed that the approach to the combination of multivariate data analysis with characteristic impurity fingerprinting obtained from high performance liquid chromatography coupled to mass spectrometry in tandem using a hybrid quadrupole time-of-flight analyzer led to discriminate between various active pharmaceutical ingredient sources both in starting materials and in finished drug products. Combining chemometrics with LC-MS QTOF impurity profiling has been already reported in the literature for the determination, for example, of the variety of red wines [58] and for the discrimination between different coffee beans origins [59] or traditional herbal medicines manufacturers [60], but so far, never with the objective of pinpointing the origins of API in pharmaceuticals.

The in-lab developed MVDA – LC-MS QTOF method was optimized and reduced to its most minimalist form, while maintaining its classification potential efficacy. Indeed, the presence of excipients in the drug formulations induced differences in the mass spectra and total ion chromatograms of starting materials compared to finished products of a same origin. Thus, the principal component analysis calibration model were built with 6 relevant variables corresponding to specific extracted ion chromatograms and 28 observations, based on the analysis of 20 starting materials and 8 finished products originated from 8 different API providers.

The 6 variables, including impurities E, F, G, B, B' and 403.2479 m/z, were scrupulously chosen among a list of 15 impurities related to simvastatin. The training model explained then cumulatively, after internal cross-validation and 3 principal components, 92.2 % of the variation, which represented an excellent goodness of fit. At the same time, the goodness of prediction came to a satisfying average goodness of prediction coefficient of 60 %. Those high values of

representativeness and reliability coefficients were verified and monitored by the right fitting of the external validation set of 16 further samples originated from the 8 reference suppliers plus one new supplementary supplier. Accordingly, the tallying results suggested that few carefully selected extracted ion chromatogram peaks contained information sufficiently relevant to distinguish between different manufacturers.

The application of that method to the determination of 5 unknown simvastatin sample origins concluded that, according to the calibration model, 3 out of 5 finished products (U1, U2 and U3) were originated from a same provider, known as provider B, and both of them (U4 and U5) were coming from two new unidentified furnishers. The confirmation after enquiry of these conclusions corroborated the effectiveness and the significance of the proposed classification model.

In addition, the study showed that the results were strongly correlated to the instrument performances. The liquid chromatographic system was essential to separate isobaric compounds, like the diastereoisomers Ph. Eur. impurity E (lovastatin) and Ph. Eur. impurity F (epilovastatin). The high resolving power combined to a high-degree detection sensitivity of the mass spectrometer using a QTOF analyzer, enabled to detect and identify co-eluted species, such as impurities B and B', as well as ultra trace level components, like impurity 403.2479 m/z.

Moreover, the mass measurement accuracy of the instrument allowed elemental composition deduction so that structures and fragmentation patterns were proposed for 4 new detected impurities of simvastatin. The structural elucidation information was more particularly based on the comparison between the mass spectra of these impurities, on one hand, and simvastatin mass spectrum, on another hand. As far as we know, the related substances, impurities A', 435.2741 m/z, 433.2585 m/z and 403.2479 m/z, were never described in the literature before. Nonetheless, our work is in accordance with the results published in previous papers [54] and [55] concerning the structure description of impurities B' and 421.2949 m/z.

However, this method presented some limitations due particularly to batch to batch variation. The dispersion within the clusters, like it was observed for example for the furnishers B, C and E, could induce a decrease in the precision of the classification characteristics, so that a perfect knowledge of the model was necessary.

Moreover, the fact that few groups of the training model were constituted of a low number of samples (providers G and H were characterized with only two samples), contributed also to limit the precision of the model. Therefore, investigation of more samples should be considered in order to incorporate supplementary observations to the calibration model and increase its representativeness and precision.

## V. Conclusion and perspectives.

In this dissertation we have presented a novel generic approach intended to pinpoint the origins of active pharmaceutical ingredients in raw materials and finished products. The new analytical method is based on high performance liquid chromatography coupled to mass spectrometry in tandem, using a hybrid quadrupole time-of-flight analyzer, in conjunction with multivariate data analysis statistical tools, such as principal component analysis or hierarchical clustering analysis. The test molecule chosen in this project was simvastatin, a hypolipidemic substance.

The objectives of this study were to optimize and define a discrimination method able to distinguish between different API providers, routes of synthesis or production sites, using leading edge and powerful technologies. Indeed, besides the noteworthy separation selectivity abilities of HPLC, the hybrid quadrupole time-of-flight analyzer offers great possibilities in terms of enhanced specificity and ultra trace level sensitivity. Exceptional characteristics, like mass measurement accuracy, below 5 ppm, and high resolution up to 10,000, have contributed to considerably minimize the signal noise and thus, to increase the method sensitivity.

The mass spectrometric detection enabled an  $4 \text{ ng.mL}^{-1}$  limit of quantitation for simvastatin, whilst a value of only  $100 \text{ ng.mL}^{-1}$  could be reached with the ultraviolet diode array detection, for that same compound. Moreover, inter-day precision ( $\text{RSD}\% \leq 4.1\%$ ,  $n=6$ ) and intra-day precision ( $\text{RSD}\% \leq 6.1\%$ ,  $n=18$ ) were found absolutely satisfactory.

Consequently, the great performances of the LC-MS/MS QTOF have led us first to detect four supplementary low-level impurities, all related to simvastatin. It was then possible to establish and propose the corresponding molecular structures in order to identify the compounds. Second, one of the degradation products, presenting a mass to charge ratio of 391.2479, and which co-eluted with simvastatin hydroxy acid when applying the original European Pharmacopoeial monograph's HPLC

conditions, could be clearly separated and identified with the proposed analytical method using a high efficiency column filled with partially porous particles.

And lastly, we have demonstrated that the combination of hyphenated techniques, such as the LC-MS/MS QTOF technology, and computational multivariate data analysis was able to determine API sources from designed models by simple visualization.

In conclusion, all these latter features could make this analytical method an outstanding tool in identifying eventual contamination of the API batches and, above all, in verifying the compliance with the certificates of suitability to the monographs of the European Pharmacopoeia (CEP). This method showed also a potential interesting aspect in defining and combating pharmaceutical counterfeit by pinpointing the origins of the API. It could be, for instance, complementary to near infrared analysis, largely used in counterfeit detection, since it presents the supplementary advantage of allowing assay determination and impurity profiling of the active pharmaceutical ingredient.

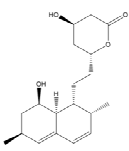
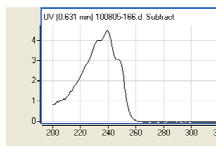
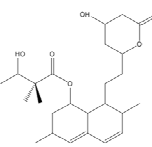
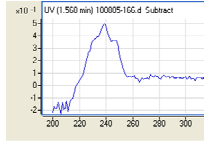
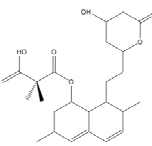
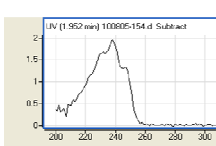
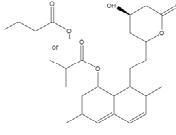
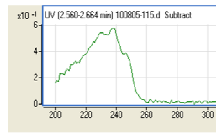
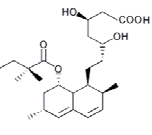
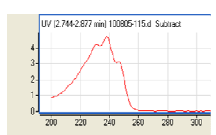
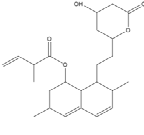
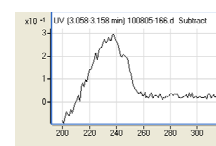
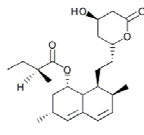
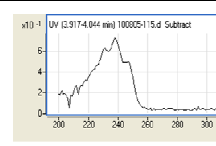
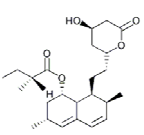
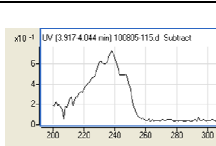
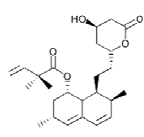
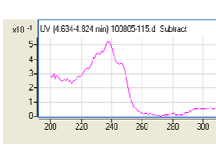
As perspectives of this work, it would be of great interest to confirm in further analysis the identification of the unknown degradation products by a complementary technique, such as nuclear magnetic resonance. The targeted compounds should be isolated and purified and then characterized by  $^1\text{H}$  and  $^{13}\text{C}$  NMR.

It could be also of great usefulness to develop a more selective chromatographic method able to separate impurities B, B' and C by using, for instance, sub- $2\mu\text{m}$  particle size column with ultra high performance liquid chromatography (UHPLC). Another useful experimentation could be led in order to determine the respective response factors of the new impurities detected. But this suggests implementing preparative HPLC, or else, synthesizing the different compounds in the laboratory.

Finally, the model should incorporate more samples and should be tested with other molecules in order to confirm the great discriminating capacities of the proposed combined approach using chemometrics, such as principal component analysis supported by hierarchical clustering analysis, in conjunction with impurity fingerprinting based on LC-MS QTOF signals. Another axis of progress in order to improve the model resolving power could be introduced by considering the information resulting from the analysis of the inorganic impurities, by using, for example, inductively coupled plasma mass spectrometry.

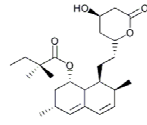
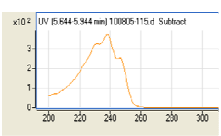
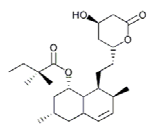
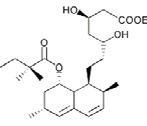

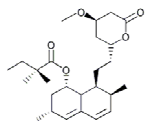
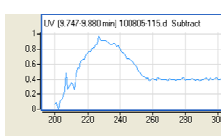
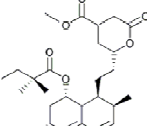
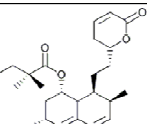

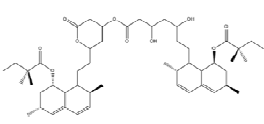
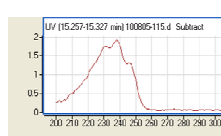
# APPENDIX A

## Structure and physic-chemical data on simvastatin and impurities

Compounds	Formula	Exact Mass	Molecular Structure	pKa	LogP	UV spectrum at 238.4nm	RT (min)	(M+H) <sup>+</sup> and abund (%largest)																
Diol Lactone	C19H28O4	320.19876			1.82		0.710	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>321.20604</td><td>100</td></tr> <tr><td>322.20944</td><td>21.04</td></tr> <tr><td>323.21213</td><td>2.92</td></tr> <tr><td>324.21481</td><td>0.31</td></tr> <tr><td>325.21745</td><td>0.03</td></tr> <tr><td>326.22005</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	321.20604	100	322.20944	21.04	323.21213	2.92	324.21481	0.31	325.21745	0.03	326.22005	0		
m/z	Abund (% largest)																							
321.20604	100																							
322.20944	21.04																							
323.21213	2.92																							
324.21481	0.31																							
325.21745	0.03																							
326.22005	0																							
Impurity m/z = 434.3	C25H38O6	434.26684			3.23		1.606	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>435.27412</td><td>100</td></tr> <tr><td>436.27752</td><td>27.72</td></tr> <tr><td>437.28029</td><td>4.93</td></tr> <tr><td>438.28301</td><td>0.86</td></tr> <tr><td>439.28566</td><td>0.07</td></tr> <tr><td>440.28827</td><td>0.01</td></tr> <tr><td>441.29086</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	435.27412	100	436.27752	27.72	437.28029	4.93	438.28301	0.86	439.28566	0.07	440.28827	0.01	441.29086	0
m/z	Abund (% largest)																							
435.27412	100																							
436.27752	27.72																							
437.28029	4.93																							
438.28301	0.86																							
439.28566	0.07																							
440.28827	0.01																							
441.29086	0																							
Impurity m/z = 432.3	C25H36O6	432.25119			2.78		1.991	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>433.25847</td><td>100</td></tr> <tr><td>434.26107</td><td>27.69</td></tr> <tr><td>435.26464</td><td>4.92</td></tr> <tr><td>436.26736</td><td>0.66</td></tr> <tr><td>437.27</td><td>0.07</td></tr> <tr><td>438.27261</td><td>0.01</td></tr> <tr><td>439.2752</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	433.25847	100	434.26107	27.69	435.26464	4.92	436.26736	0.66	437.27	0.07	438.27261	0.01	439.2752	0
m/z	Abund (% largest)																							
433.25847	100																							
434.26107	27.69																							
435.26464	4.92																							
436.26736	0.66																							
437.27	0.07																							
438.27261	0.01																							
439.2752	0																							
Impurity m/z = 390.2	C23H34O5	390.2406					2.676	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>391.2479</td><td>100</td></tr> <tr><td>392.2513</td><td>25.47</td></tr> <tr><td>393.2541</td><td>4.14</td></tr> <tr><td>394.2568</td><td>0.5</td></tr> <tr><td>395.2595</td><td>0.05</td></tr> <tr><td>396.2621</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	391.2479	100	392.2513	25.47	393.2541	4.14	394.2568	0.5	395.2595	0.05	396.2621	0		
m/z	Abund (% largest)																							
391.2479	100																							
392.2513	25.47																							
393.2541	4.14																							
394.2568	0.5																							
395.2595	0.05																							
396.2621	0																							
Impurity A Simvastatin hydroxy acid	C25H40O6	436.28249		4.31	3.85		2.876	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>437.28377</td><td>100</td></tr> <tr><td>438.29310</td><td>27.74</td></tr> <tr><td>439.29595</td><td>4.93</td></tr> <tr><td>440.29867</td><td>0.86</td></tr> <tr><td>441.30132</td><td>0.07</td></tr> <tr><td>442.30393</td><td>0.01</td></tr> <tr><td>443.30652</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	437.28377	100	438.29310	27.74	439.29595	4.93	440.29867	0.86	441.30132	0.07	442.30393	0.01	443.30652	0
m/z	Abund (% largest)																							
437.28377	100																							
438.29310	27.74																							
439.29595	4.93																							
440.29867	0.86																							
441.30132	0.07																							
442.30393	0.01																							
443.30652	0																							
Impurity m/z = 402.2	C24H34O5	402.24062			3.41		3.188	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>403.2478</td><td>100</td></tr> <tr><td>404.25131</td><td>26.55</td></tr> <tr><td>405.25412</td><td>4.41</td></tr> <tr><td>406.25695</td><td>0.55</td></tr> <tr><td>407.25952</td><td>0.06</td></tr> <tr><td>408.26216</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	403.2478	100	404.25131	26.55	405.25412	4.41	406.25695	0.55	407.25952	0.06	408.26216	0		
m/z	Abund (% largest)																							
403.2478	100																							
404.25131	26.55																							
405.25412	4.41																							
406.25695	0.55																							
407.25952	0.06																							
408.26216	0																							
Impurity E Lovastatin	C24H36O5	404.25627			3.68		3.849	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>405.26265</td><td>100</td></tr> <tr><td>406.26696</td><td>26.57</td></tr> <tr><td>407.26977</td><td>4.42</td></tr> <tr><td>408.2725</td><td>0.55</td></tr> <tr><td>409.27519</td><td>0.06</td></tr> <tr><td>410.27782</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	405.26265	100	406.26696	26.57	407.26977	4.42	408.2725	0.55	409.27519	0.06	410.27782	0		
m/z	Abund (% largest)																							
405.26265	100																							
406.26696	26.57																							
407.26977	4.42																							
408.2725	0.55																							
409.27519	0.06																							
410.27782	0																							
Impurity F Epilovastatin	C24H36O5	404.25627			3.68		4.042	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>405.26265</td><td>100</td></tr> <tr><td>406.26696</td><td>26.57</td></tr> <tr><td>407.26977</td><td>4.42</td></tr> <tr><td>408.2725</td><td>0.55</td></tr> <tr><td>409.27519</td><td>0.06</td></tr> <tr><td>410.27782</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	405.26265	100	406.26696	26.57	407.26977	4.42	408.2725	0.55	409.27519	0.06	410.27782	0		
m/z	Abund (% largest)																							
405.26265	100																							
406.26696	26.57																							
407.26977	4.42																							
408.2725	0.55																							
409.27519	0.06																							
410.27782	0																							
Impurity G	C25H36O5	416.25627			4.12		4.798	<table border="1"> <thead> <tr> <th>m/z</th> <th>Abund (% largest)</th> </tr> </thead> <tbody> <tr><td>417.26355</td><td>100</td></tr> <tr><td>418.26696</td><td>27.65</td></tr> <tr><td>419.26981</td><td>4.71</td></tr> <tr><td>420.27255</td><td>0.6</td></tr> <tr><td>421.27525</td><td>0.06</td></tr> <tr><td>422.2779</td><td>0.01</td></tr> <tr><td>423.28053</td><td>0</td></tr> </tbody> </table>	m/z	Abund (% largest)	417.26355	100	418.26696	27.65	419.26981	4.71	420.27255	0.6	421.27525	0.06	422.2779	0.01	423.28053	0
m/z	Abund (% largest)																							
417.26355	100																							
418.26696	27.65																							
419.26981	4.71																							
420.27255	0.6																							
421.27525	0.06																							
422.2779	0.01																							
423.28053	0																							

## APPENDIX A (continued)

### Structure and physic-chemical data on simvastatin and impurities

Compound Name	Chemical ID	Retention Time (min)	Chemical Structure	Retention Time (min)	Retention Time (min)	Chromatogram	Retention Time (min)	m/z	Abund. (% largest)
Simvastatin	C25H38 O5	418.27192		13.5	4.39		5.770	418.2792	100
Impurity m/z = 420.3 dihydro Simvastatin	C25H40O5	420.28757			4.86	No signal	8.082	421.29485	100
Impurity m/z = 464.3 Simvastatin ethyl ester	C27H44O6	464.31379			4.46		8.351	465.32107	100
Impurity m/z = 432.3 Simvastatin methyl ether	C26H40O5	432.28757			4.75		9.844	433.29465	100
Impurity B Simvastatin methyl ester	C27H40O6	460.28249			4.94	Not available	9.941	461.26977	100
Impurity C Dehydro Simvastatin	C25H36O4	400.26136			5.5		10.068	401.26884	100
Impurity D Simvastatin dimer	C50H76O10	836.54385					15.354	837.55113	100



## APPENDIX B

Intra-day and inter-day instrument precision considering individual components' absolute peak areas

Injection										
Day 1	m/z=390.2 RT=2.68	ImpA m/z=436.3 RT=2.88 Hydroxy acid	ImpE m/z=404.3 RT=3.85 Lovastatin	ImpF m/z=404.3 RT=4.04 Epilovastatin	ImpG m/z=416.3 RT=4.80	Simvastatin m/z=418.3 RT=5.77	m/z=420.3 RT=8.08 Dihydro SVT	m/z=432.3 RT=9.84 SVT methyl ether	ImpC m/z=400.3 RT=10.07 Dehydro SVT	ImpD m/z=836.5 RT=15.35 SVT Dimer
1	714342	804225	8991563	3753726	1867615	370566848	1898272	2568468	2616524	285494
2	744078	761741	8872587	3595447	1817702	374529765	1818473	2546035	2475953	314728
3	720663	755677	8602265	3618419	1876633	372444816	1747619	2508552	2453717	289519
4	702114	734889	8446604	3581971	1798747	370000000	1753035	2469328	2404130	286236
5	733501	834533	8410401	3446727	1716983	370675710	1737969	2366719	2499332	278016
6	682565	755758	8257714	3510483	1836611	369046004	1771782	2385233	2369756	282378
Mean (counts)	716211	774471	8596856	3584462	1819049	371210524	1787858	2474056	2469902	289395
SD (counts)	22044	37215	284327	104282	57990	1969350	61168	83328	86044	13005
RSD (%)	3,1	4,8	3,3	2,9	3,2	0,5	3,4	3,4	3,5	4,5

Injection										
Day 2	m/z=390.2 RT=2.68	ImpA m/z=436.3 RT=2.88 Hydroxy acid	ImpE m/z=404.3 RT=3.85 Lovastatin	ImpF m/z=404.3 RT=4.04 Epilovastatin	ImpG m/z=416.3 RT=4.80	Simvastatin m/z=418.3 RT=5.77	m/z=420.3 RT=8.08 Dihydro SVT	m/z=432.3 RT=9.84 SVT methyl ether	ImpC m/z=400.3 RT=10.07 Dehydro SVT	ImpD m/z=836.5 RT=15.35 SVT Dimer
1	636955	893089	8175002	3199373	1631328	362140096	1624480	2330352	2271524	274321
2	672068	863740	7867032	3177606	1673040	362834285	1685966	2322093	2290333	254642
3	623939	842411	7809300	3165337	1641276	361075594	1625045	2185717	2192776	263824
4	612779	830194	7777569	3154606	1631033	360840854	1626988	2177417	2190039	270339
5	623565	852751	7681525	3152550	1671591	360193590	1638040	2189864	2145187	280523
6	645564	888929	7746034	3159312	1512917	352000000	1599896	2303282	2293565	280670
Mean (counts)	635812	861852	7842744	3168131	1626864	359847403	1633403	2251454	2230571	270720
SD (counts)	21130	25201	174181	17760	58926	3959164	28628	74158	62576	10142
RSD (%)	3,3	2,9	2,2	0,6	3,6	1,1	1,8	3,3	2,8	3,7

Injection										
Day 3	m/z=390.2 RT=2.68	ImpA m/z=436.3 RT=2.88 Hydroxy acid	ImpE m/z=404.3 RT=3.85 Lovastatin	ImpF m/z=404.3 RT=4.04 Epilovastatin	ImpG m/z=416.3 RT=4.80	Simvastatin m/z=418.3 RT=5.77	m/z=420.3 RT=8.08 Dihydro SVT	m/z=432.3 RT=9.84 SVT methyl ether	ImpC m/z=400.3 RT=10.07 Dehydro SVT	ImpD m/z=836.5 RT=15.35 SVT Dimer
1	687753	1522189	8667990	3448870	1760944	373478462	1799114	2398371	2540567	386986
2	679544	1414534	8477058	3381863	1775465	371084781	1783121	2408582	2440490	385786
3	691963	1426845	8533281	3297655	1767502	369813290	1744850	2352289	2383559	383262
4	630131	1442103	8192287	3295672	1678422	364699182	1699496	2258891	2372680	351511
5	633238	1504053	8128959	3291378	1768680	364433601	1725417	2238365	2329256	344201
6	656994	1497740	7967122	3263398	1734195	363906353	1660062	2270376	2298026	346249
Mean (counts)	663271	1467911	8327783	3329806	1747535	367902612	1735343	2321146	2394096	366333
SD (counts)	27302	45483	271374	70671	36770	4077575	51943	74679	86729	20997
RSD (%)	4,1	3,1	3,3	2,1	2,1	1,1	3,0	3,2	3,6	5,7

Inter-day repeatability										
Mean (n=18)	671764	1034745	8255794	3360800	1731149	366320180	1718868	2348885	2364856	308816
SD (n=18)	40895	319195	396812	189359	95226	5895153	80553	120215	126895	44973
RSD (n=18)	6,1	30,8	4,8	5,6	5,5	1,6	4,7	5,1	5,4	14,6

## APPENDIX C

### Intra-day and inter-day instrument precision considering internal peak area normalization

Injection										
Day 1	m/z=390.2 RT=2.68	ImpA m/z=436.3 RT=2.88 Hydroxy acid	ImpE m/z=404.3 RT=3.85 Lovastatin	ImpF m/z=404.3 RT=4.04 Etilovastatin	ImpG m/z=416.3 RT=4.80	Simvastatin m/z=418.3 RT=5.77	m/z=420.3 RT=8.08 Dihydro SVT	m/z=432.3 RT=9.84 SVT methyl ether	ImpC m/z=400.3 RT=10.07 Dehydro SVT	ImpD m/z=836.5 RT=15.35 SVT Dimer
1	0,1928%	0,2170%	2,4264%	1,0130%	0,5040%	100,0000%	0,5123%	0,6931%	0,7061%	0,0770%
2	0,1987%	0,2034%	2,3690%	0,9600%	0,4853%	100,0000%	0,4855%	0,6798%	0,6611%	0,0840%
3	0,1935%	0,2029%	2,3097%	0,9715%	0,5039%	100,0000%	0,4692%	0,6735%	0,6588%	0,0777%
4	0,1898%	0,1986%	2,2829%	0,9681%	0,4861%	100,0000%	0,4738%	0,6674%	0,6498%	0,0774%
5	0,1979%	0,2251%	2,2689%	0,9298%	0,4632%	100,0000%	0,4689%	0,6385%	0,6743%	0,0750%
6	0,1850%	0,2048%	2,2376%	0,9512%	0,4977%	100,0000%	0,4801%	0,6463%	0,6421%	0,0765%
Mean (counts)	0,1929%	0,2086%	2,3157%	0,9656%	0,4900%	100,0000%	0,4816%	0,6664%	0,6654%	0,0779%
SD (counts)	0,0051%	0,0102%	0,0700%	0,0276%	0,0155%	0,0000%	0,0163%	0,0206%	0,0227%	0,0031%
RSD (%)	2,7	4,9	3,0	2,9	3,2	0,0%	3,4	3,1	3,4	4,0

Injection										
Day 2	m/z=390.2 RT=2.68	ImpA m/z=436.3 RT=2.88 Hydroxy acid	ImpE m/z=404.3 RT=3.85 Lovastatin	ImpF m/z=404.3 RT=4.04 Etilovastatin	ImpG m/z=416.3 RT=4.80	Simvastatin m/z=418.3 RT=5.77	m/z=420.3 RT=8.08 Dihydro SVT	m/z=432.3 RT=9.84 SVT methyl ether	ImpC m/z=400.3 RT=10.07 Dehydro SVT	ImpD m/z=836.5 RT=15.35 SVT Dimer
1	0,1759%	0,2466%	2,2574%	0,8835%	0,4505%	100,0000%	0,4486%	0,6435%	0,6273%	0,0757%
2	0,1852%	0,2381%	2,1682%	0,8758%	0,4611%	100,0000%	0,4647%	0,6400%	0,6312%	0,0702%
3	0,1728%	0,2333%	2,1628%	0,8766%	0,4546%	100,0000%	0,4501%	0,6053%	0,6073%	0,0731%
4	0,1698%	0,2301%	2,1554%	0,8742%	0,4520%	100,0000%	0,4509%	0,6034%	0,6069%	0,0749%
5	0,1731%	0,2367%	2,1326%	0,8752%	0,4641%	100,0000%	0,4548%	0,6080%	0,5956%	0,0779%
6	0,1834%	0,2525%	2,2006%	0,8975%	0,4298%	100,0000%	0,4545%	0,6543%	0,6516%	0,0797%
Mean (counts)	0,1767%	0,2396%	2,1795%	0,8805%	0,4520%	100,0000%	0,4539%	0,6258%	0,6200%	0,0753%
SD (counts)	0,0062%	0,0085%	0,0440%	0,0090%	0,0121%	0,0000%	0,0058%	0,0227%	0,0205%	0,0034%
RSD (%)	3,5	3,5	2,0	1,0	2,7	0,0	1,3	3,6	3,3	4,5

Injection										
Day 3	m/z=390.2 RT=2.68	ImpA m/z=436.3 RT=2.88 Hydroxy acid	ImpE m/z=404.3 RT=3.85 Lovastatin	ImpF m/z=404.3 RT=4.04 Etilovastatin	ImpG m/z=416.3 RT=4.80	Simvastatin m/z=418.3 RT=5.77	m/z=420.3 RT=8.08 Dihydro SVT	m/z=432.3 RT=9.84 SVT methyl ether	ImpC m/z=400.3 RT=10.07 Dehydro SVT	ImpD m/z=836.5 RT=15.35 SVT Dimer
1	0,1841%	0,4076%	2,3209%	0,9234%	0,4715%	100,0000%	0,4817%	0,6422%	0,6802%	0,1036%
2	0,1831%	0,3812%	2,2844%	0,9113%	0,4785%	100,0000%	0,4805%	0,6491%	0,6577%	0,1040%
3	0,1871%	0,3858%	2,3075%	0,8917%	0,4779%	100,0000%	0,4718%	0,6361%	0,6445%	0,1036%
4	0,1728%	0,3954%	2,2463%	0,9037%	0,4602%	100,0000%	0,4660%	0,6194%	0,6506%	0,0964%
5	0,1738%	0,4127%	2,2306%	0,9031%	0,4853%	100,0000%	0,4735%	0,6142%	0,6391%	0,0944%
6	0,1805%	0,4116%	2,1893%	0,8968%	0,4765%	100,0000%	0,4562%	0,6239%	0,6315%	0,0951%
Mean (counts)	0,1802%	0,3990%	2,2632%	0,9050%	0,4750%	100,0000%	0,4716%	0,6308%	0,6506%	0,0995%
SD (counts)	0,0058%	0,0136%	0,0501%	0,0112%	0,0085%	0,0000%	0,0095%	0,0137%	0,0171%	0,0047%
RSD (%)	3,2	3,4	2,2	1,2	1,8	0,0	2,0	2,2	2,6	4,7

Inter-day repeatability										
Mean (n=18)	0,1833%	0,2824%	2,2528%	0,9170%	0,4723%	100,0000%	0,4691%	0,6410%	0,6453%	0,0842%
SD (n=18)	0,0090%	0,0865%	0,0780%	0,0405%	0,0198%	0,0000%	0,0159%	0,0261%	0,0272%	0,0117%
RSD (n=18)	4,9	30,6	3,5	4,4	4,2	0,0	3,4	4,1	4,2	13,9

## APPENDIX D

### Liquid chromatographic parameters

Parameter	Value																		
<b>Column</b>																			
<i>Type</i>	Kinetex™ C18, 50 mm x 2.1 mm, 2.6 μm																		
<i>Column temperature</i>	35°C																		
<b>Injector</b>																			
<i>Injection volume</i>	5 μL																		
<i>Autosampler temperature</i>	10°C																		
<i>Washing program</i>	Action																		
	<ol style="list-style-type: none"> <li>1. DRAW def. amount sample, def. speed,...</li> <li>2. NEEDLEWASH in flush port for 10.0 sec.</li> <li>3. INJECT</li> <li>4. WAIT 1.00 min</li> <li>5. VALVE bypass</li> <li>6. WAIT 14.5 min</li> <li>7. VALVE mainpass</li> <li>8. VALVE bypass</li> <li>9. VALVE mainpass</li> </ol>																		
<b>Pump</b>																			
<i>Flow rate</i>	0.5 mL.min <sup>-1</sup>																		
<i>Solvent A</i>	0.1% formic acid in water																		
<i>Solvent B</i>	0.1% formic acid in acetonitrile																		
<i>Gradient</i>	<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center; border-bottom: 1px solid black;">Time (min)</th> <th style="text-align: center; border-bottom: 1px solid black;">% mobile phase B</th> </tr> </thead> <tbody> <tr><td style="text-align: center;">0</td><td style="text-align: center;">42</td></tr> <tr><td style="text-align: center;">6.5</td><td style="text-align: center;">42</td></tr> <tr><td style="text-align: center;">7</td><td style="text-align: center;">53</td></tr> <tr><td style="text-align: center;">9.5</td><td style="text-align: center;">53</td></tr> <tr><td style="text-align: center;">14</td><td style="text-align: center;">87.5</td></tr> <tr><td style="text-align: center;">17</td><td style="text-align: center;">87.5</td></tr> <tr><td style="text-align: center;">17.2</td><td style="text-align: center;">42</td></tr> <tr><td style="text-align: center;">20</td><td style="text-align: center;">42</td></tr> </tbody> </table>	Time (min)	% mobile phase B	0	42	6.5	42	7	53	9.5	53	14	87.5	17	87.5	17.2	42	20	42
Time (min)	% mobile phase B																		
0	42																		
6.5	42																		
7	53																		
9.5	53																		
14	87.5																		
17	87.5																		
17.2	42																		
20	42																		
<b>UV detector</b>																			
<i>Channel A</i>	238 nm (4nm BW on DAD)																		
<i>Reference channel A</i>	none																		
<i>DAD range</i>	200 - 400 nm																		
<i>Step</i>	1 nm																		
<i>Peak width (response time)</i>	0.05 min (1.0s)																		
<i>Slit</i>	4 nm																		

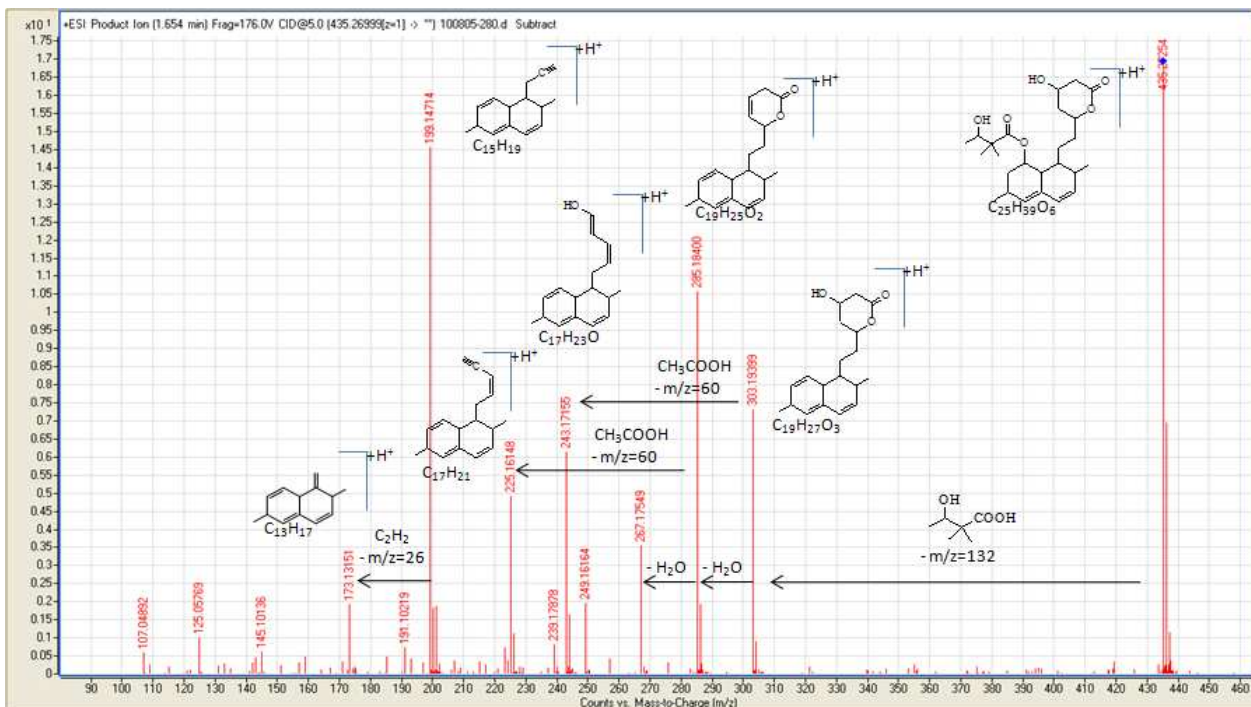
## APPENDIX E

### Mass spectrometer parameters

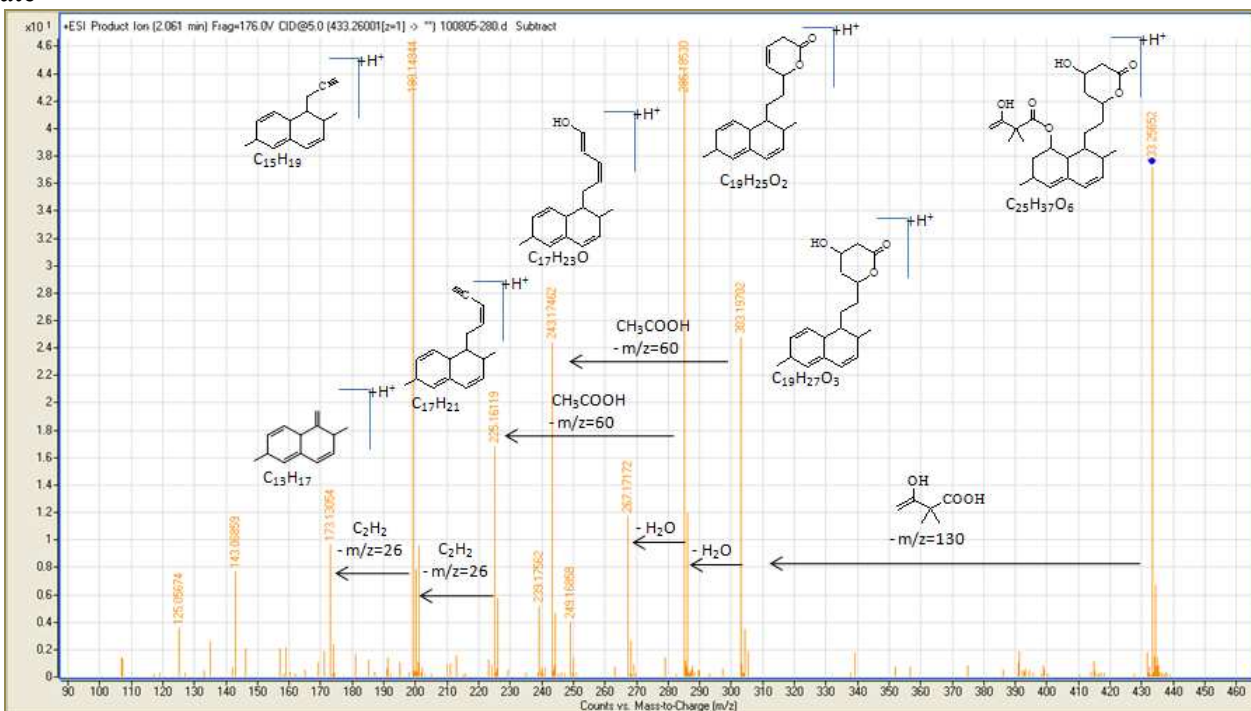
Parameter	Value	
<b>Ion source</b>		
<i>Type</i>	Dual Electrospray	
<i>Ionization mode</i>	positive	
<i>Nebulizer gas pressure</i>	35 psi	
<i>Drying gas flow rate</i>	11 mL.min <sup>-1</sup>	
<i>Drying gas temperature</i>	300 °C	
<b>Interface</b>		
<i>Fragmentor Voltage</i>	175 V	
<i>Capillary Voltage</i>	3100 V	
<i>Skimmer</i>	65 V	
<i>Octapole 1 Vpp</i>	750 V	
<b>Analyzer</b>		
<i>Acquisition mode</i>	Auto MS/MS	
<i>Acquisition rate</i>	1 spectra/s	
<i>Mass range</i>	110-1000 m/z	
<i>Isolation width</i>	Medium (~4m/z)	
<i>Collision energy</i>	5 V	
<b>Precursor selection</b>		
<i>Maximum number</i>	20	
<i>Absolute threshold</i>	1000 counts	
<i>Relative threshold</i>	0.005%	
<i>Active charge state</i>	1	
<i>Preferred ions</i>	Precursor m/z	Retention time (min)
	403.24790	3.1
	405.26355	3.8
	405.26355	4.0
	417.26355	4.8
	419.27920	5.8
	433.29485	9.8
	461.28977	9.9
Accepting range ( $\Delta$ )	20 ppm	0.2 min

## APPENDIX F

Fragment pathway and in-tandem mass spectra at 5eV collision energy of molecular ion located at 435.2725 m/z corresponding to (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl]ethyl]-3,7-dimethyl-1,2,3,7,8,8a-hexahydronaphtalen-1-yl-3-hydroxy-2,2-dimethyl-butanoate

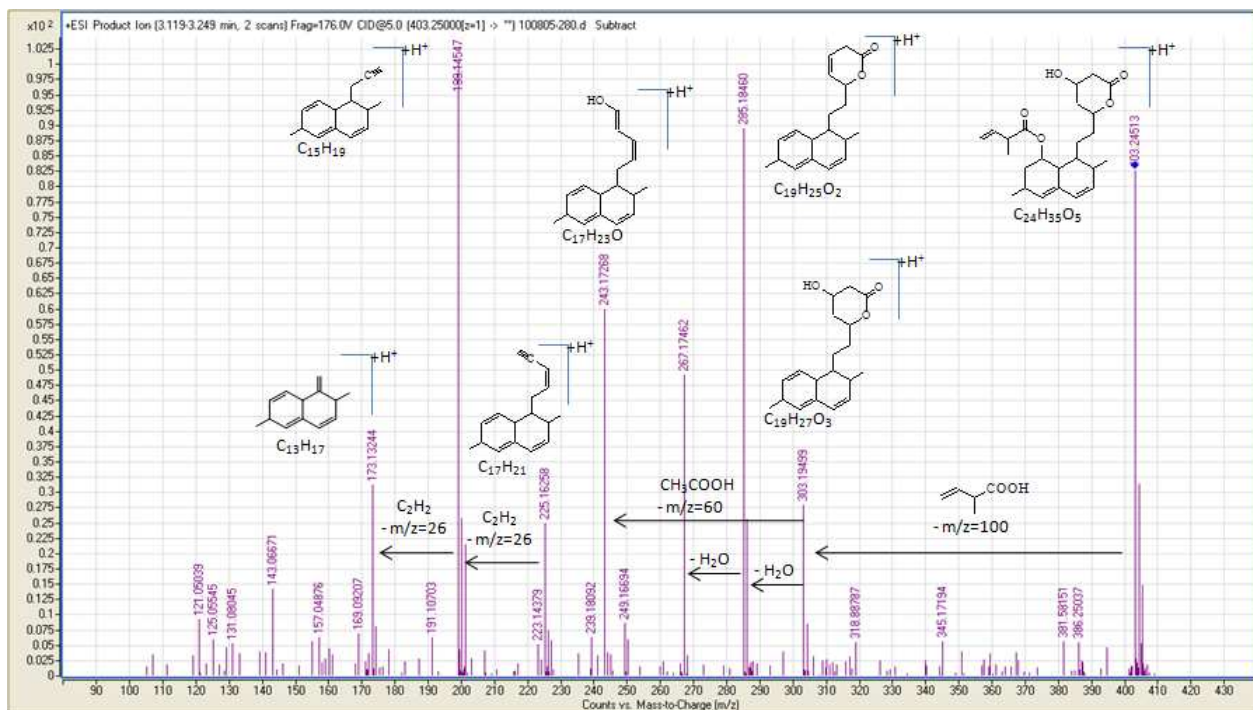


Fragment pathway and in-tandem mass spectra at 5eV collision energy of molecular ion located at 433.2565 m/z corresponding to (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl]ethyl]-3,7-dimethyl-1,2,3,7,8,8a-hexahydronaphtalen-1-yl-3-hydroxy-2,2-dimethyl-but-3-enoate

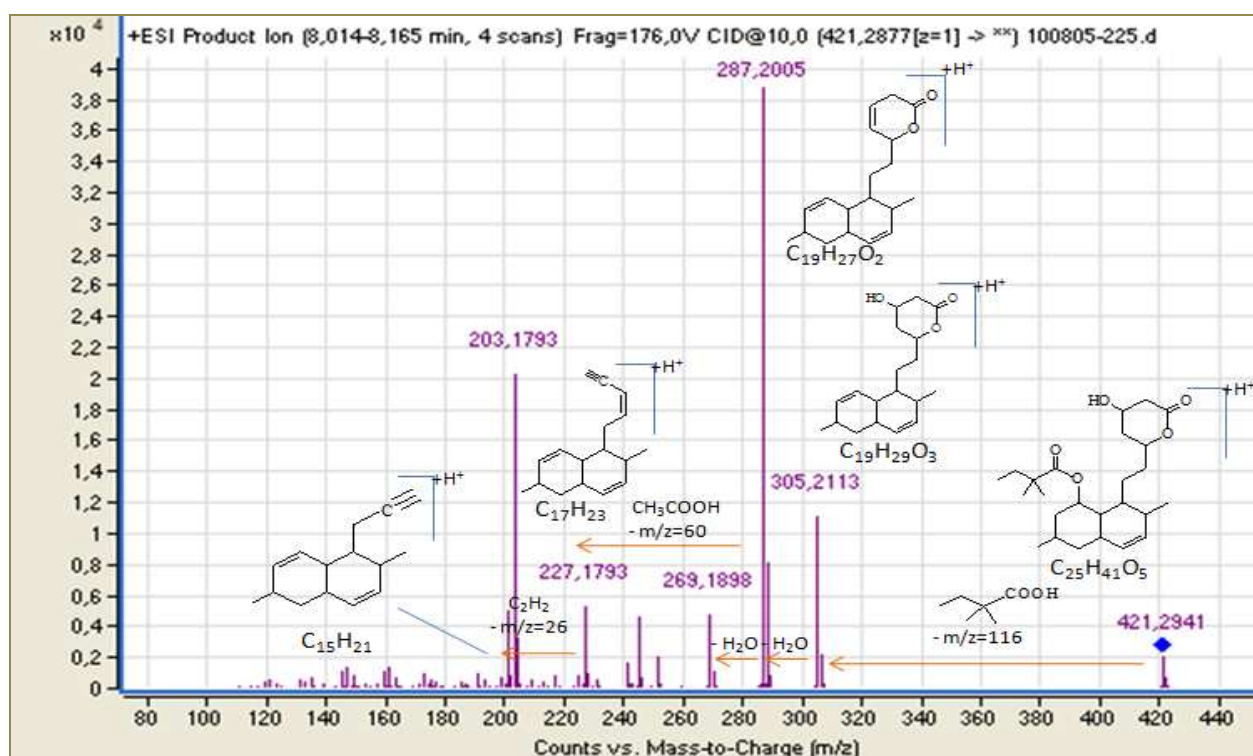


## APPENDIX G

Fragment pathway and in-tandem mass spectra at 5eV collision energy of molecular ion located at 403.2451 m/z and corresponding to (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl]ethyl]-3,7-dimethyl-1,2,3,7,8,8a-hexahydronaphtalen-1-yl-2-methyl-but-3-enoate.



Fragment pathway and in-tandem mass spectra at 10eV collision energy of molecular ion located at 421.2941 m/z and corresponding to (1S,3R,7S,8S,8aR)-8-[2-[(2R,4R)-4-hydroxy-6-oxo-tetra-hydro-2H-pyran-2-yl]ethyl]-3,7-dimethyl-1,2,3,7,8,8a-octahydronaphtalen-1-yl-2,2-dimethyl-butanoate.



## APPENDIX H

Reporting, identification and qualification thresholds of related substances in active substances according to the European Pharmacopoeia 7<sup>th</sup> edition general monograph “Substances for pharmaceutical use (2034)”.

Use	Maximum daily dose	Reporting threshold	Identification threshold	Qualification threshold
Human use or human and veterinary use	≤ 2 g/day	> 0.05 per cent	> 0.10 per cent or a daily intake of > 1.0 mg (whichever is the lower)	> 0.15 per cent or a daily intake of > 1.0 mg (whichever is the lower)
Human use or human and veterinary use	> 2 g/day	> 0.03 per cent	> 0.05 per cent	> 0.05 per cent
Veterinary use only	Not applicable	> 0.10 per cent	> 0.20 per cent	> 0.50 per cent

## REFERENCES

- [1] Swedish Medical Products Agency.  
[www.mpa.se](http://www.mpa.se)
- [2] French Health Products Safety Agency.  
[www.afssaps.fr](http://www.afssaps.fr)
- [3] European Directorate for the Quality of Medicines and HealthCare.  
[www.edqm.eu](http://www.edqm.eu)
- [4] Nature Reviews Drug Discovery – News & Analysis, 2010. Evolving R&D for emerging markets. NPG – Macmillan Publishers Ltd, **9** (6), 417-420.
- [5] Council of Europe. Council of Europe Convention on the counterfeiting of medical products and similar crimes involving threats to public health (December 2010).  
<http://www.coe.int/t/DGHL/StandardSetting/MediCrime/Medicrime-EdProv%20ENG.pdf>
- [6] WEIL D.A., TIMAR Z., ZUMWALT M., NAEGELE E., 2008. Detection and identification of impurities in pharmaceutical drugs. Computer-assisted extraction, profiling and analysis of Q-TOF data for determination of impurities using Agilent MassHunter software. Agilent application note, publication number 5989-8529EN.
- [7] KOERNER P. AND MATTHEWS T., 2009. Increased efficiency and resolution with Kinetex™ Core-Shell technology. Phenomenex applications, technical note TN-1058.
- [8] ELLISON D.K., MOORE W.D., PETTS C.R., 1993. *Analytical profiles of drug substances and excipients. Volume 22*. H.G. Brittain, New Jersey, USA, 591 pages.
- [9] COSTET P., 2010. Molecular pathways and agents for lowering LDL-cholesterol in addition to statins. *Pharmacol. Ther.*, **126**, 263-278.
- [10] MALIK A.K., BLASCO C., PICÓ Y., 2010. Liquid chromatography – mass spectrometry in food safety. *J. Chrom. A*, **1217**, 4018-4040.
- [11] MA C., KAVALIER A.R., JANG B., KENNELLY E.J., 2011. Metabolic profiling of *Actaea* species extracts using high performance liquid chromatography coupled with electrospray ionization time-of-flight mass spectrometry. *J. Chrom. A*, **1218**, 1461-1476.
- [12] DOOLEY K.C., 2003. Tandem mass spectrometry in the clinical chemistry laboratory. *Clin. Biochem.*, **36**, 471 - 481.
- [13] de CASTRO A., CONCHEIRO M., CHAKLEYA D.M., HUESTIS M.A., 2009. Development and validation of a liquid chromatography mass spectrometry assay for the simultaneous quantification of methadone, cocaine, opiates and metabolites in human umbilical cord. *J.Chrom. B*, **877** ( 27), 3065 - 3071.
- [14] LEWIS S.E., BRODIE J.E., BAINBRIDGE Z.T., ROHDE K.W., DAVIS A.M., MASTERS B.L., MAUGHAN M., DEVLIN M.J., MUELLER J.F., SCHAFFELKE B., 2009. Herbicides: a new threat to the Great Barrier Reef. *Environ. Poll.*, **157** (8-9), 2470 - 2484.



## REFERENCES (continued)

- [15] ROUESSAC F., ROUESSAC A., avec la collaboration de CRUCHE D., 2004. *Analyse Chimique. Méthodes et techniques instrumentales modernes. 6<sup>ème</sup> édition*. Dunod, Paris, 462 pages.
- [16] European Pharmacopoeia 7<sup>th</sup> edition. Chromatographic separation techniques. Chapter 2.2.46.
- [17] SUBRAMANIAN G., 2007. *Chiral separation techniques: a practical approach. Third completely revised and updated edition*. Wiley - VCH, Weinheim, Germany, 618 pages.
- [18] BOSCO G.L., 2010. The development of LC-MS. The marriage of the bird and the fish. *Trends Anal. Chem.*, **29** (38), 781-794.
- [19] SKOOG D.A., HOLLER F.J., NIEMAN T.A. 2003. *Principes d'analyse instrumentale. Traduction et révision scientifique de la 5<sup>ème</sup> édition américaine par C. Buess-Herman et F. Dumont*. De Boeck Université, Bruxelles, Belgique, 968 pages.
- [20] McLAFFERTY F.W., 1981. Tandem Mass Spectrometry. *Science*, **214** (4518), 280-287.
- [21] de HOFFMANN E., STROOBANT V., 2007. *Mass spectrometry, principles and applications. Third edition*. Wiley & Sons, Chichester, UK, 489 pages.
- [22] MENDHAM J., DENNEY R.C., BARNES J.D., THOMAS M.J.K., 2006. *Analyse Chimique Quantitative de Vogel. Traduction et révision scientifique de la 6<sup>ème</sup> édition anglaise par J. Toullec et M. Mottet*. De Boeck Université, Bruxelles, Belgique, 920 pages.
- [23] ARDREY R.E., 2003. Liquid chromatography – Mass spectrometry: an introduction. Wiley & Sons, Chichester, UK, 263 pages.
- [24] LC/GC chromatography online, Chester, UK.  
<http://chromatographyonline.findanalytichem.com/lcgc/article/articleDetail.jsp?id=504702>  
(connected in October 2010)
- [25] FENN J.B., MANN M., MENG C.K., WONG S.F., WHITEHOUSE C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, **246** (4926), 64–71.
- [26] University of Bristol – NERC Life Sciences Mass Spectrometry Facility, Bristol, UK.  
<http://www.bris.ac.uk/nerclsmsf/techniques/hplcms.html>  
(connected in October 2010)
- [27] New Objective, Woburn, USA  
<http://www.newobjective.com/electrospray/index.html>  
(connected in October 2010)

## REFERENCES (continued)

- [28] HORNING E.C., CARROLL D.I., DZIDIC I., HAEGELE K.D., HORNING M.G. AND STILLWELL R.N., 1974. Liquid chromatography – Mass spectrometer – Computer analytical systems: a continuous-flow system based on atmospheric pressure ionization mass spectrometry. *J.Chrom. A*, **99**, 13-21.
- [29] <http://www.chem.agilent.com/cag/other/appi%20source.gif>  
(connected in October 2010)
- [30] Agilent Technologies. Agilent 6510 Q-TOF LC/MS techniques and operation. Course number R1904A, Volume 1.
- [31] SREEKUMAR J., HOGAN T.J., TAYLOR S., TURNER Ph., KNOTT C., 2010. A quadrupole mass spectrometer for resolution of low mass isotopes. *J. Am. Soc. Mass Spectrom.*, **21**, 1364-1370.
- [32] PAUL W., STEINWEDEL H., 1960. Apparatus for separating charged particles of different specific charges. US Patent 2939952.
- [33] PEDDER R.E., 2001. Practical quadrupole theory: graphical theory. Extrel Application note RA\_2010A. Poster presented at the 49<sup>th</sup> ASMS Conference on Mass Spectrometry and allied topics, Chicago.
- [34] <http://commons.wikimedia.org/wiki/File:Quadrupole.gif>  
(connected in October 2010)
- [35] *Clarke's analysis of drugs and poisons in pharmaceuticals, body fluids and postmortem material. Third edition.* MOFFAT A.C., OSSELTON M.D., WIDDOP B., Eds, Pharmaceutical Press, London, UK, 2004, 1656 pages.
- [36] KOPPENAAL D.W., BARINAGA C.J., BONNER DENTON M., HIEFTJE G.M., SCHILLING G.D., ANDRADE F.J., BARNES J.H., 2005. MS Detectors. *Anal. Chem.*, **77** (21), 418A-427A.
- [37] MARSHALL A.G., HENDRICKSON C.L., 2008. High resolution mass spectrometers. *Annual Rev. Anal. Chem.*, **1**, 579-599.
- [38] University of Heidelberg. The little encyclopedia of mass spectrometry.  
<http://www.rzuser.uni-heidelberg.de/~bl5/encyclopedia.html>  
(connected in November 2010)
- [39] BALOGH M.P., 2004. Debating resolution and mass accuracy. *LCGC Europe*, **17** (3), 152-159.
- [40] MILLER J.N., MILLER J.C., 2005. *Statistics and chemometrics for analytical chemistry. Fifth edition.* Pearson Education Ltd, Harlow, UK, 325 pages.

## REFERENCES (continued)

- [41] Unscrambler® X version 10.01. CAMO software AS, Oslo, Norway.  
<http://www.camo.com>
- [42] Ondalys training session: “L’analyse de données multivariées appliqué aux methods analytiques sous le logiciel The Unscrambler® X” (December 2010). Ondalys, Prades-le-Lez, France.
- [43] Le dictionnaire Vidal 2006, 82<sup>ème</sup> édition. Issy-les-Moulineaux, France, ISBN 2-85091-139-9.
- [44] NOVÁKOVÁ L., VLČKOVÁ H., ŠATÍNSKÝ D., SADÍLEK P., SOLICHOVÁ D., BLÁHA M., BLÁHA V., SOLICH P., 2009. Ultra high performance liquid chromatography tandem mass spectrometric detection in clinical analysis of simvastatin and atorvastatin. *J. Chrom. B*, **877** (22), 2093-2103.
- [45] European Pharmacopoeia 7<sup>th</sup> edition. Monograph on simvastatin (04/2009:1563).
- [46] KOERNER P., MATHEWS T., 2009. Increased efficiency and resolution with Kinetex™ Core-Shell Technology. Phenomenex Applications TN-1058.
- [47] GRITTI F., LEONARDIS I., SHOCK D., STEVENSON P., SHALLIKER A., GUIOCHON G., 2010. Performance of columns packed with the new shell particles, Kinetex-C18. *J. Chrom. A*, **1217**, 1589-1603.
- [48] GRITTI F., GUIOCHON G., 2010. Performance of columns packed with the new shell Kinetex-C18 particles in gradient elution chromatography. *J. Chrom. A*, **1217**, 1604-1615.
- [49] CABOOTER D., FANIGLIULO A., BELLAZI G., ALLIERI B., ROTTIGNI A., DESMET G., 2010. Relationship between the particle size distribution of commercial fully porous and superficially porous high performance liquid chromatography column packings and their chromatographic performances. *J. Chrom. A*, **1217**, 7074-7081.
- [50] GUILLARME D., NGUYEN D.T.-T., RUDAZ S., VEUTHEY J.-L., 2007. Recent developments in liquid chromatography – Impact on qualitative and quantitative performance. *J. Chrom. A*, **1149**, 20-29.
- [51] GUILLARME D., HEINISH S., ROCCA J.-L., 2004. Effect of temperature in reversed phase liquid chromatography. *J. Chromatogr. A*, **1052**, 39-51.
- [52] SUBIRATS X., BOSCH E., ROSÉS M., 2009. Buffer considerations for LC and LC/MS. *Chromatography online*.
- [53] DOLAN J.W., 2010. Enhancing Signal-to-Noise. *LC/GC chromatography online*.
- [54] VULETIC M., CINDRIĆ M., KORUŽNJAK J.D., 2004. Identification of unknown impurities in simvastatin substance and tablets by liquid chromatography/tandem mass spectrometry. *J. Pharm. Biomed. Anal.*, **7**, 715-721.

## REFERENCES (continued)

- [55] PLUMB R.S., JONES M.D., RAINVILLE P., CASTRO-PEREZ J.M., 2007. The rapid detection and identification of the impurities of simvastatin using high resolution sub-2 $\mu$ m particle LC coupled to hybrid quadrupole time of flight MS operating with alternating high-low collision energy. *J. Sep. Sci.*, **30** (16), 2666-2675.
- [56] NOVÁKOVÁ L., ŠATÍNSKÝ D., SOLICH P., 2008. HPLC methods for the determination of simvastatin and atorvastatin. *Trends Anal. Chem.*, **27** (4), 352-367.
- [57] International Vocabulary of Metrology. Third edition. Basic and general concept and associated terms. JCGM 200:2008.
- [58] VACLAVIK L., LACINA O., HAJŠLOVA J., ZWEIGENBAUM J., 2011. The use of high performance liquid chromatography – quadrupole time-of-flight mass spectrometry coupled to advanced data mining and chemometric tools for discrimination and classification of red wines according to their variety. *Anal. Chim. Acta*, **685**, 45-51.
- [59] KUHNERT N., JAISWAL R., ERAVUCHIRA P., EL-ABASSY R.M., von der KAMMER B., MATERNY A., 2011. Scope and limitations of principal component analysis of high resolution LC-TOF-MS data: the analysis of the chlorogenic acid fraction in green coffee beans as a case study. *Anal. Methods*, **3** (1), 144-155.
- [60] XIAOHUI F., YI W., YIYU C., 2006. LC/MS fingerprinting of Shenmai injection: a novel approach to quality control of herbal medicines. *J. Pharm. Biomed. Anal.*, **40** (3), 591-597.

## RESUME

Le contrôle qualité des matières premières entrant dans la fabrication des produits finis pharmaceutiques est considéré comme primordial par l'industrie pharmaceutique et les agences de régulation. Ainsi, une méthode analytique permettant de déterminer l'origine des principes actifs dans les matières premières et produits finis est présentée dans ce document. Cette méthode, combinant analyse multivariée et profils d'impuretés obtenus par chromatographie liquide haute performance couplée à la spectrométrie de masse en tandem par analyseur hybride quadripôle – temps de vol (LC-MS/MS QTOF), a été mise en œuvre sur 49 échantillons de matières premières et produits finis contenant la substance active hypolipémiante simvastatin. L'extrême sensibilité de la technique LC-MS/MS QTOF a permis l'identification de 4 nouvelles substances apparentées non répertoriées dans la monographie correspondante. L'analyse en composantes principales, basée sur un modèle à 6 variables et 28 observations, exprimait 92,2% de la variance, après trois composantes, et affichait un coefficient de prédiction de 60%. Les résultats obtenus ont permis de discriminer sans ambiguïté entre 11 fournisseurs distincts, confirmant la capacité de la méthode combinant chimiométrie et profils d'impuretés LC-MS/MS QTOF à distinguer entre différentes origines de principe actif.

**Mots clés:** Simvastatin; Principes Actifs; Matières Premières; Profil d'Impuretés; Analyse en Composantes Principales; Chimiométrie; LC-MS/MS QTOF.

## ABSTRACT

Quality monitoring of active pharmaceutical ingredient (API) used in medicinal products is of highest interest for both the pharmaceutical industry and the regulatory agencies. Therefore, a new approach combining chemometrics with API impurity profiling using high performance liquid chromatography coupled to mass spectrometry in tandem equipped with a hybrid quadrupole time-of-flight analyzer (LC-MS/MS QTOF) was examined in order to discriminate between different origins of starting materials and finished drug products. Simvastatin, a hypolipidemic agent, was chosen as test molecule for the developed method. Impurity fingerprints of forty nine samples originated from eleven distinct providers were investigated. Firstly, the LC-MS/MS QTOF trace level sensitivity ( $4 \text{ ng.mL}^{-1}$  for simvastatin quantitation limit) enabled the identification of four new related substances. Secondly, principal component analysis, supported by hierarchical clustering analysis, was implemented to classify the various API sources. The training model, built with twenty eight observations and six variables, corresponding to six common extracted ion chromatogram peaks, explained cumulatively 92.2% of the variation, after three components, and presented a prediction coefficient of 60%. The results obtained demonstrated that the proposed approach consisting in combining singular LC-MS impurity fingerprinting with chemometric models led to unambiguously distinguish between different API suppliers.

**Key words:** Simvastatin; Active Pharmaceutical Ingredients; Fingerprinting; Impurities; Principal Component Analysis; Hierarchical Clustering Analysis; Chemometrics; LC-MS/MS QTOF.