

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

IPHC – UMR 7178

THÈSE présentée par :

Steve HESSMANN

soutenue le : 6 décembre 2021

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Chimie Analytique

Développement de stratégies analytiques en protéomique quantitative: quantification des protéines de la cellule hôte par spectrométrie de masse comme outil de contrôle qualité pour l'industrie biopharmaceutique

THÈSE dirigée par :

Dr. CARAPITO Christine

Chargée de recherche, Université de Strasbourg

RAPPORTEURS :

Dr. FERRO Myriam

Pr. BRACEWELL Daniel

Directrice de recherche, CEA, Grenoble

Professor, University College London

AUTRES MEMBRES DU JURY :

Dr. CHERY Cyrille

Dr. CIANFERANI Sarah

Dr. O'HARA John

Directeur PCMD, UCB, Braine l'Alleud

Directrice de recherche, Université de Strasbourg

Director BFCS, UCB, Slough

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

IPHC – UMR 7178

THÈSE présentée par :

Steve HESSMANN

soutenue le : 6 décembre 2021

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Chimie Analytique

**Development of analytical strategies in
quantitative proteomic: quantitation of host cell
proteins by mass spectrometry as a quality
control tool for the biopharmaceutical industry**

THÈSE dirigée par :

Dr. CARAPITO Christine

Chargée de recherche, Université de Strasbourg

RAPPORTEURS :

Dr. FERRO Myriam

Pr. BRACEWELL Daniel

Directrice de recherche, CEA, Grenoble

Professor, University College London

AUTRES MEMBRES DU JURY :

Dr. CHERY Cyrille

Dr. CIANFERANI Sarah

Dr. O'HARA John

Directeur PCMD, UCB, Braine l'Alleud

Directrice de recherche, Université de Strasbourg

Director BFCS, UCB, Slough

A ma famille et mes amis,

A mon frère et mes sœurs,

A mes parents,

A Umray,

A ma grand-mère,

« Il n'y a qu'une façon d'échouer, c'est d'abandonner avant d'avoir réussi. »

Georges Clemenceau

«Sometimes science is more art than science»

Rick Sanchez

Acknowledgments

Ce travail de thèse a été réalisé au sein du Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO) de l'Institut Pluridisciplinaire Hubert Curien (IPHC, UMR7178).

Tout d'abord, je tiens à remercier Sarah Cianféran pour m'avoir accueilli au sein de ce laboratoire de qualité. Merci pour vos conseils, votre encadrement et votre confiance qui m'ont permis d'étoffer mes connaissances et compétences en spectrométrie de masse.

Je tiens bien évidemment à remercier chaleureusement ma directrice de thèse, Christine Carapito. Merci pour ton encadrement, ta bienveillance et pour les "coups de boost/pression" dont tu as le secret. J'ai beaucoup appris lors de cette aventure et c'est en grande partie grâce à toi.

Je remercie l'équipe PCMD d'UCB. Merci pour votre accueil lors de ma mission. Merci Somar pour ton aide sur le BioAccord. Je te remercie Anne-sophie pour tes conseils et tes idées. Merci à Cyrille Chéry pour ton accueil et ton encadrement tout au long de cette thèse. Tu as toujours été de bon conseils et de bonne critiques. Je tiens également à remercier Annick Gervais. Un grand merci pour ta gentillesse, tes conseils et ta disponibilité.

I would like to thank Dr. Myriam Ferro, Pr. Daniel Bracewell, Dr. John O'Hara, Dr. Cyrille Chéry, and Dr. Sarah Cianferani for having accepted to review and evaluate my PhD work.

Je souhaite également remercier les collaborateurs avec lesquels j'ai pu travailler lors de ces trois années : Guillaume Béchade, Nicolas Autret et Tanguy Fortin.

Un immense merci à l'ensemble du LSMBO pour votre accueil, votre sens du partage et votre bonne humeur !

Merci à Laurence, Christine S., Fabrice B. et Alain pour vos conseils et pour avoir pris le temps de répondre à mes questions.

Merci à la team info/bioinfo : Fabrice V., Alex pour votre aide précieuse ! Merci pour votre patience, votre gentillesse et bien sûr vos compétences.

Merci aux permanents du R5 : Hélène, Stella, Martine, Delphine, Agnès et Jean-Marc pour votre disponibilité. Merci Hélène et Agnès pour votre aide sur les instruments à mon arrivée au labo et dès que j'avais une question. Merci Martine pour ton aide lors des démarches administratives et pour les petites discussions le matin.

Un immense merci à Jean-Marc, JMS, le beau gosse... Merci pour tes conseils précieux, ton humour et ta disponibilité. J'ai beaucoup appris à tes côtés. Cela a toujours été un plaisir d'entendre: « T'en veux? » une fois un problème résolu.

Merci aux permanents du R2 : Magalie, François et Véronique pour votre aide et conseils. Magali, François, je vous souhaite de bien vous amuser avec la DIA.

Acknowledgments

Merci aux non-permanents R5: Corentin et Charlotte, je vous souhaite bon courage pour la fin. Corentin amuse toi bien avec la DIA/top-down/HCP, etc. Charlotte merci pour ta bonne humeur et pour les discussions peu importe le sujet (science, film, etc.). Lâche un peu la blouse quand même. Merci Justine pour ta bonne humeur, ton rire et les débats au coin café du R5.

Merci aux non-permanents R2: Tout d'abord merci à Leslie et Papo pour les discussions et conseils. Merci Valérieane pour ton aide sur les dernières minutes du manuscrit. Jeewan, have fun with the TimsTOF ;) Chloé merci pour les discussions et pour les expériences TimsTOF-HCP. Marie, la présidente, madame R, merci pour le soutien. Reste calme, boulard modéré et bonne continuation. Marie, la colloque de bureau, j'espère un jour pouvoir m'installer sur une chaise comme toi. Cette cohabitation a été courte mais agréable.

Merci à mon premier bureau de thésard, alias le bureau XOXO, pour les bons moments. On peut citer le calendrier (je joue plus...) et les goûters. Merci Jojo et Nico pour votre accueil au sein de la Team HCP. Jojo, malgré ton statut tu étais toujours disponible pour discuter et je t'en remercie. N'oublie pas de brancher tes écouteurs si jamais. Nico, merci pour ton partage de connaissance et ton soutien. Au final, on a réussi à prendre le pouvoir au bureau. Aurélie, Merci pour ta gentillesse, ton aide, ton soutien et pour les discussions/débats ;). Ne change pas. J'ai adoré participer aux soirées labo, aux discussion sur le meilleur pâtissier et aux séances de Sissy Mua ;).

Merci à la team supramol. Oscar, CR...7, fait pas le bordel ! Merci pour ton aide et pour les discussions. N'oublie pas : hala Madrid ! Merci Stéphane, Mr Qualité, Mr bigB. On reste en contact pour les scores et vidéos intéressantes que tu as à partager (#PauseCafé) ;) Botza, ne change pas shy commandant. Profite de One fitness pour moi.

Merci Marie pour ton soutien et ton aide sur cette fin de manuscrit. Amuse-toi bien avec le XL. Je te souhaite une bonne continuation et plein de réussite. Tu le mérites. Arrête un peu de radoter quand même :p. Evo, Evoshy, on a eu une bonne idée de bureau. C'était un bon bureau pour un moment peu agréable. Mets un peu à jour tes playlists quand même et surtout n'oublie pas MA MA MA.

Bon courage aux petits derniers, Jérôme et Rania, plein de belles choses sont à venir.

Je vais finir ce paragraphe sur le groupe Nose/Ankle/Belly. Il est facile de savoir qui est qui. Merci Max. Ne change rien sur le plan professionnel et personnel. Merci pour tes formations et ton aide au quotidien. On se capte sur Warzone. Anthony, merci pour tout et tout résume beaucoup de chose. On peut dire que je suis un follower. Ce fût un plaisir de te faire découvrir Kendrick, Damso et travis/The last kingdom et surtout choisis ton exo ! Ne m'en veut pas je fais court. On aura plein d'occasion de discuter de tout cela tous les trois.

Merci à ma famille, belle famille et mes amis. Vous m'avez toujours soutenu. Merci à mon frère, mes sœurs et mes parents pour votre aide au quotidien. Vous êtes toujours là pour moi même si ma présence s'est faite rare ces derniers temps. Pour cela, je ne vous en remercierai jamais assez. Je tiens également à remercier ma grand-mère sans qui je ne serais pas là.

Umray, je fini par te remercier. Sans ton soutien au quotidien dans les bons et mauvais moments, je n'aurai pas réussi. Merci d'avoir relevé ce défi avec moi et de m'avoir poussé dans les moments difficiles. Une nouvelle vie commence maintenant !

Table of content

Abbreviations.....	11
Résumé en français	15
Chapitre 1 : La problématique des protéines de la cellule hôte dans l'industrie pharmaceutique	17
Chapitre 2 : Développement de stratégies analytiques de quantification des HCPs pour assister la chaîne de production d'anticorps thérapeutiques	21
1. Développement d'une stratégie innovante de quantification Top3 dédiée au suivi des HCPs	21
2. Évaluation des stratégies DIA pour le suivi global des HCPs sur un instrument de type Q-orbitrap.....	24
3. Étude de cas : Impact de la capacité du bioréacteur sur le profil des HCPs	25
4. Avantage de la mobilité ionique pour le suivi des HCPs	28
Chapitre 3 : Le défi analytique des produits pharmaceutiques purifiés.....	31
1. Amélioration de la préparation des échantillons pour l'analyse des substances médicamenteuses	31
2. Étude de cas : Enquête sur le changement de procédé de production d'un anticorps monoclonal.....	34
Chapitre 4 : Développement d'une méthode de suivi des HCPs sur une plateforme MS pour l'analyse de routine	37
Chapitre 5 : Conclusion générale	41
General introduction	43
Bibliographic introduction.....	49
Chapter 1: Bottom-up proteomic strategies	51
1. Mass Spectrometry-based proteomic approaches	52
2. Analytical workflow of the bottom-up strategy	53
A. Sample preparation.....	53
<i>i. Protein extraction</i>	<i>53</i>
<i>ii. Protein preparation for LC-MS/MS analysis.....</i>	<i>54</i>
<i>iii. Enzymatic digestion</i>	<i>55</i>
B. Liquid chromatography coupled to tandem mass spectrometry analysis	56

i. Peptide separation by reversed phased-liquid chromatography	56
ii. Tandem mass spectrometry.....	57
C. Ion mobility spectrometry for bottom-up proteomics	59
i. Field Asymmetric Ion Mobility Spectrometry (FAIMS)	60
ii. Trapped Ion Mobility Spectrometry (TIMS).....	61
D. Data processing and interpretation	62
i. Protein identification strategies.....	62
ii. Validation of the identification results.....	65
Chapter 2 : Global quantification strategies.....	67
1. Label-based relative strategies.....	67
A. Metabolic or enzymatic labelling strategies	67
B. Chemical labelling strategies.....	68
2. Label-free relative quantification approaches	68
A. Spectral counting	69
B. Extracted ion chromatogram	69
3. Label-free “absolute” quantification	70
Chapter 3: Targeted quantification strategies.....	71
1. Selected reaction monitoring.....	71
2. Parallel Reaction Monitoring.....	72
A. The principle of Parallel Reaction Monitoring	72
B. The development of a PRM method	72
i. Hypothesis and protein selection.....	73
ii. Peptide selection.....	73
C. Absolute quantification using isotopic dilution.....	74
i. Label-based absolute quantification.....	74
ii. Isotopic dilution.....	74
iii. The determination of the quantification limits	75
Chapter 4: Data-independent acquisition	77
1. Principle of data-independent acquisition.....	78
2. Evolution of the DIA-based approaches.....	78
A. DIA strategies performed over the entire mass range	79
B. DIA strategies based on isolation windows.....	80
i. Consecutive fixed width windows	80
ii. Consecutive variable width windows	81
iii. Overlapping windows	81
iv. Multiplexed strategies	82
3. DIA assay development	82

A. Hypothesis.....	83
B. LC-DIA parameter optimization.....	83
C. LC-DIA data processing.....	85
D. New hypothesis.....	85
4. The challenge of data processing	85
A. Peptide-centric approach.....	86
<i>i. Extraction via a spectral library</i>	<i>86</i>
<i>ii. Targeted data extraction</i>	<i>87</i>
<i>iii. Direct spectrum matches</i>	<i>87</i>
B. Spectrum-centric approach.....	88
Chapter 5: The concern of Host Cell Proteins in biopharmaceutical companies	89
1. Monoclonal antibody	89
A. Expression system	89
B. Manufacturing process.....	90
<i>i. Upstream process</i>	<i>90</i>
<i>ii. Downstream process.....</i>	<i>90</i>
2. Host cell proteins monitoring.....	91
A. Immuno-specific methods	92
<i>i. ELISA</i>	<i>92</i>
<i>ii. Western-blot</i>	<i>94</i>
B. Non-specific methods	94
<i>i. Gel electrophoresis</i>	<i>95</i>
<i>ii. Mass spectrometry</i>	<i>96</i>
Conclusion	97
 Part I: Development of an accurate HCP quantification method to support mAb manufacturing	99
Chapter 1: Development of an innovative Top3 quantification strategy dedicated to HCP profiling.....	101
1. Context of the project	101
2. Experimental design	103
3. Inter-laboratory study	104
4. HCP Profiler standard to support process development	106
5. Conclusion	107
Chapter 2: Evaluation of DIA strategies for global HCP profiling on a Q-orbitrap instrument	109

1. The promises of DIA-based quantification	109
2. Context of the project	110
3. Experimental design	111
4. Data acquisition and processing optimisations	112
A. Sample-dependent data acquisition optimizations	112
B. Evaluation of the HCP Profiler standard combined with DIA	113
C. Implementation of a stringent data filtering workflow	114
5. Results	115
A. Benchmarking DIA versus MS1-XIC DDA	115
B. Spectral library free-based approach investigation	117
C. Where do we stand in regards to ELISA and the Sciex Triple-TOF instrument (TTOF 6600)?	120
6. Conclusion & Perspectives	121
Chapter 3: Case study: Impact of the bioreactor capacity on the HCPs profile	123
1. Context of the project and samples description	123
2. Data analysis strategy	124
3. Results	125
A. Optimization of a project-specific spectral library	125
B. Evaluation of a spectral library free-based strategy	128
C. HCP profiles characterization	130
4. Conclusion	132
Chapter 4: Benefit of ion mobility in the field of HCPs monitoring	135
1. Presentation of the TimsTOF PRO instrument	135
2. Analytical strategy	138
3. Results	139
A. DIA-PASEF for in-depth characterization of HCPs	139
B. diaPASEF combined with spectrum-centric data extraction	142
C. Benchmarking of diaPASEF against DIA on a Q-Orbitrap instrument	143
4. Conclusion & Perspectives	144
Part II: The analytical challenge of drug products	145
Chapter 1: Sample preparation improvements for the analysis of drug substances	147
1. Challenge of the analysis of final drug substances	147
2. Context of the project	148
3. Experimental design	150
4. Results	152

A. Optimization of a stringent data validation workflow	152
B. Development of an optimised native liquid digestion protocol	154
C. Implementation of a focused ultrasonication step	155
6. Conclusion & Perspectives	157
Chapter 2: Case study: Manufacturing process investigation	159
1. Context of the project	159
2. Experimental design	159
3. Results	160
A. Quantification results.....	160
B. Process related impurities study.....	161
4. Conclusions.....	162
 Part III: Development of a HCP profiling method on a MS platform for routine analysis	163
Chapter 1: Evaluation of the BioAccord LC-MS system	165
1. Context of the project	165
2. Experimental design	166
3. Results	168
A. HCP monitoring simulation	168
B. Targeted profiling of NIST mAb HCPs.....	171
4. Conclusion	173
 General conclusion.....	175
 Experimental section.....	181
1. Development of an accurate HCP quantification method to support mAb manufacturing ...	183
A. Development of an innovative Top3 quantification strategy dedicated to HCP profiling..	183
<i>i. Sample preparation protocol</i>	<i>183</i>
<i>ii. NanoLC-MS/MS analysis</i>	<i>185</i>
<i>iii. DDA data treatment</i>	<i>186</i>
B. Evaluation of DIA strategies for global HCP profiling on a Q-Orbitrap instrument	186
<i>i. Sample preparation protocol</i>	<i>186</i>
<i>ii. nanoLC-MS/MS analysis</i>	<i>187</i>
<i>iii. Data treatment</i>	<i>187</i>
C. Case study: Impact of the bioreactor capacity on HCPs profile	189
<i>i. Sample preparation protocol</i>	<i>189</i>

ii. NanoLC-DIA analysis	189
iii. DIA data treatment	190
D. Benefit of ion mobility in the field of HCPs monitoring	190
i. Sample preparation protocol	190
ii. NanoLC-IM-MS/MS analysis	192
iii. Data treatment	192
2. The analytical challenge of drug product	193
A. Sample preparation improvements for drug substances	193
i. Sample preparation protocol	193
ii. nanoLC-MS/MS analysis	194
iii. DIA data treatment	195
B. Case study: Manufacturing process investigation	196
i. Sample preparation protocol	196
ii. NanoLC-MS/MS analysis	196
iii. DIA data treatment	196
3. Development of a HCP profiling method on a MS platform for routine analysis	197
A. Evaluation of the BioAccord LC-MS system	197
i. Sample preparation protocol	197
ii. LC-MSe analysis	197
iii. Targeted data treatment	197
List of communications	203
Annexe.....	205
References	237

Abbreviations

1D SDS-PAGE	One-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis
2D DIGE	Two-dimensional differential gel electrophoresis
2D PAGE	Two-dimensional polyacrylamide gel electrophoresis
ACN	Acetonitrile
ADH	Alcohol dehydrogenase
AFA	Adapted focused acoustic
AIF	All-ion fragmentation
AQUA	Absolute quantification
AUC	Area under the curve
BSA	Bovine serum albumin
CCCF	Clarified cell culture fluid
CCS	Collision cross section
CE	Collision energy
CHO	Chinese hamster ovary
CID	Collision induced dissociation
CQA	Critical quality attribute
CV	Coefficient of variation
Da	Dalton
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DS	Drug substance
DSP	Downstream process
ECD	Electron capture dissociation
ELISA	Enzyme-linked immunosorbent assays
EMA	European medicines agency
ENL	Enolase
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FA	Formic acid
FAIMS	Field asymmetric ion mobility spectrometry
FDA	U.S. food and drug administration
FDP	False discovery proportion
FDR	False positive rate
FLEXIQuant	Full-length expressed stable isotope-labelled proteins for quantification
FT-ARM	Fourier transform-all reaction monitoring
FWHM	Full width at half maximum
HCD	Higher collision dissociation
HCP	Host Cell Protein
HR/AM	High-resolution/accurate mass
HRM	Hyper reaction monitoring
iBAQ	Intensity-based absolute quantification

ICAT	Isotope-coded affinity tag
ID	Isotope dilution
IgG	Immunoglobulins G
IMS	Ion mobility spectrometry
iTRAQ	Isobaric tags for relative and absolute quantification
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
LLOQ	Lower limit of quantification
LOD	Limit of detection
LOQ	Limit of quantification
mAb	Monoclonal antibody
MALDI	Matrix-assisted laser desorption
Mix 4P	Mix of four standard proteins
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MW	Molecular weight
m/z	Mass-to-charge ratio
NeuCode	Neutron encoding
NIST	National institute of standards and technology
Ng/mg mAb	Nanogram of host cell proteins per milligram of monoclonal antibody
PACIFIC	Precursor acquisition independent from ion count
PAI	Protein abundance index
PASEF	Parallel accumulation – serial fragmentation
pI	Isoelectric point
PPA	Post protein A affinity chromatography
PrEST	Protein epitope signature tag
PRM	Paralleled Reaction Monitoring
PSAQ	Protein standard absolute quantification
PSM	Peptide spectrum match
PTM	Post-translational modification
PYGM	Phosphorylase B
Q	Quadrupole analyser
QC	Quality control
QconCAT	Quantification concatamer
RP-HPLC	Reverse-phase high performance liquid chromatography
RT	Retention time
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SIB	Swiss Institute of Bioinformatics
SILAC	Stable isotope labelling with amino acids in cell culture
SPE	Solid-phase extraction
SRM	Selected Reaction Monitoring
SWATH	Sequential windowed acquisition of all theoretical fragment ion spectra
TIMS	Trapped ion mobility spectrometry
TMPP	Trimethoxyphenyl phosphonium

TMT	Tandem mass tag
TOF	Time-of-flight analyser
TPR	True positive rate
ULOQ	Upper limit of quantification
UPLC	Ultra performance liquid chromatography
USP	Upstream process
WiSIM	Wide selected-ion monitoring
XDIA	Extended data-independent acquisition
XIC-MS1	Extracted ion chromatogram – MS1 filtering

Résumé en français

Chapitre 1

La problématique des protéines de la cellule hôte dans l'industrie pharmaceutique

Les anticorps monoclonaux (mAbs) et produits dérivés font partie des protéines thérapeutiques qui jouent un rôle très important dans l'industrie pharmaceutique¹. Au cours des dernières années, les agences américaines et européennes (FDA et EMA), responsables de la pharmacovigilance ont autorisé la mise sur le marché de plus de 90 mAbs pour des applications variées²⁻³. Ces protéines thérapeutiques sont produites de manière recombinante dans des systèmes cellulaires. Le procédé de production du médicament se divise en deux étapes : un processus de culture cellulaire en bioréacteurs (Upstream Process, USP) suivi d'étapes successives de purification (Downstream Process, DSP). Dans des conditions optimisées et contrôlées, les cellules sécrètent les mAbs dans le milieu de culture. Le fluide de culture cellulaire clarifié (CCCF) est ensuite recueilli. Il va notamment contenir les protéines d'intérêt et diverses impuretés dont les protéines résiduelles de la cellule hôte (Host Cell Proteins, HCP). Plusieurs étapes de purification vont permettre d'obtenir un produit concentré et le plus pur possible lors du DSP (Figure 1).

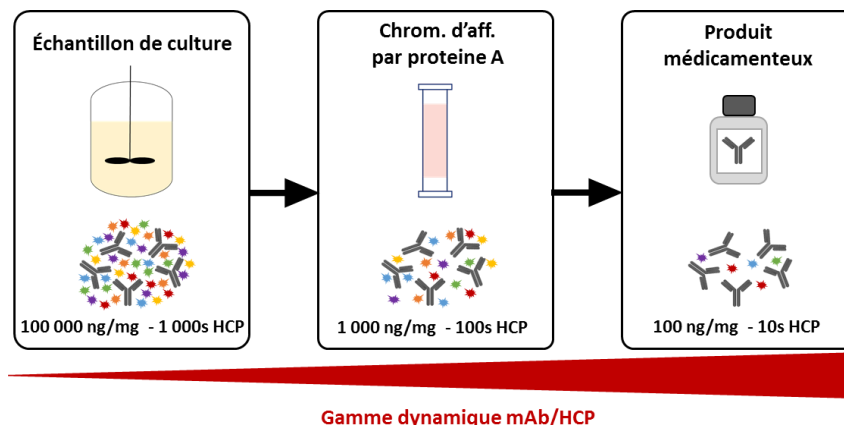


Figure 1: Élimination des protéines de la cellule hôte pendant le processus de purification (adapté de ⁴). Ordre de grandeur de la quantité globale d'impuretés (ng de HCPs par mg de mAb) et nombre de HCPs observés aux différentes étapes du processus de production des mAbs. Chrom. d'aff. par protéine A : Chromatographie d'affinité par protéine A.

Les HCPs font l'objet d'attentions particulières car elles peuvent réduire la quantité de biomolécules actives ou induire une réponse immunogène chez le patient⁵⁻⁶. C'est pourquoi, les autorités réglementaires recommandent la quantification de ces impuretés et requièrent une quantité globale inférieure à 100 ng d'HCPs par mg de mAb⁷. L'ELISA (« Enzyme Linked Immunosorbent Assay ») est la méthode la plus couramment utilisée pour quantifier les HCPs. Malheureusement, cette technique souffre de plusieurs facteurs limitants. Parmi ces facteurs, on peut citer l'obtention d'une quantité globale sans information individuelle sur l'identité des HCPs ou encore une couverture incomplète de ces impuretés au sein du produit final⁸. Toutes ces limitations ont encouragé le développement de méthodes alternatives en vue d'améliorer la qualité du produit médicamenteux et potentiellement réduire le coût de production des mAbs.

Dans ce contexte, la protéomique quantitative, basée sur le couplage entre la chromatographie liquide et la spectrométrie de masse en tandem (LC-MS/MS), est devenue une méthode incontournable. En effet, la détection par spectrométrie de masse (MS) va permettre l'identification et la quantification non biaisées de chaque HCPs détectable. Les avancées technologiques et bio-informatiques en analyse protéomique offrent un grand nombre de stratégies permettant l'identification et la quantification des protéines (figure 2).

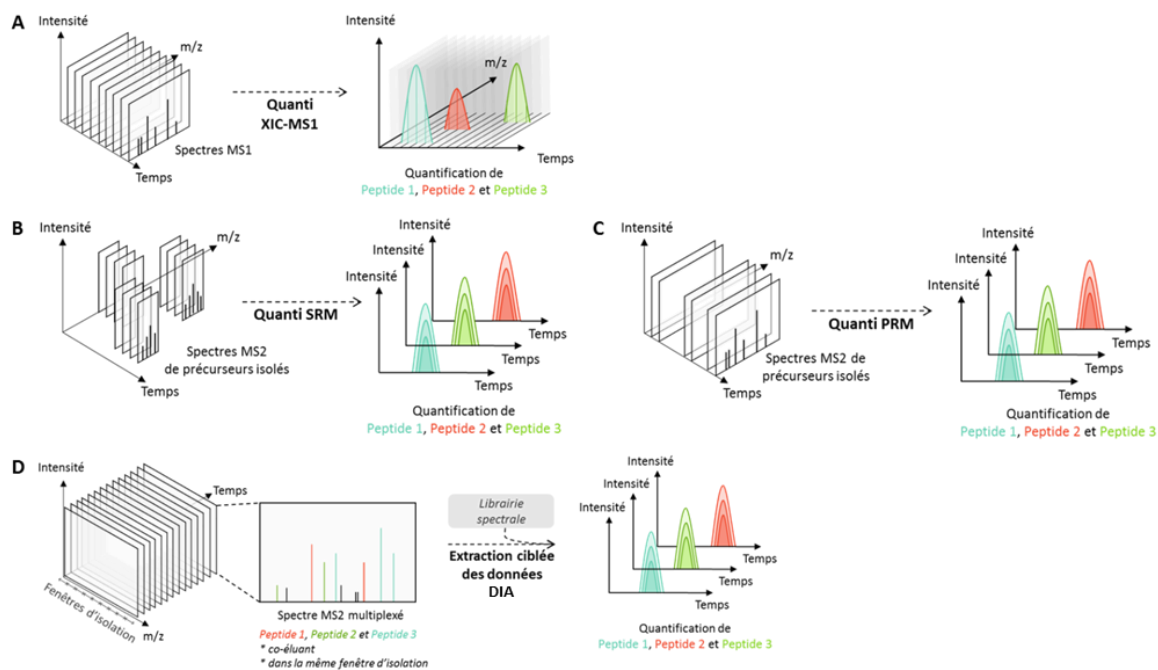


Figure 2: Représentation schématique des méthodes de protéomique quantitative. Les quantités peptidiques peuvent être déterminées par des approches globales telles que l'extraction des courants d'ions à partir des spectres MS1 acquis tout au long du gradient chromatographique (A). Les approches ciblées, de type SRM (B) et PRM (C), permettent de quantifier des peptides préalablement sélectionnés et séquentiellement isolés pour générer, après fragmentation, des spectres MS2 à partir desquels l'information quantitative est extraite. L'approche DIA consiste à co-isoler et co-fragmenter des peptides co-élutés et contenus dans une même large fenêtre d'isolation (généralement 25 Da) (D). L'information quantitative peut ensuite être extraite de manière ciblée à partir des spectres MS2 et à l'aide d'une librairie spectrale.

L'utilisation du mode « Data Dependent Acquisition » (DDA) va permettre d'obtenir une image globale du profil des HCPs et de leur quantité grâce à l'identification par MS/MS et l'extraction du courant d'ion des peptides présents (Figure 2.A). Cette méthode offre la possibilité de quantifier des milliers de protéines, avec une justesse et précision de quantification dépendant fortement de la nature des standards internes utilisés. Cependant, l'acquisition des données DDA est basée sur la sélection des ions les plus intenses pour fragmentation. En conséquence, la gamme dynamique ainsi que la répétabilité de la méthode s'en retrouvent limitées⁹. Les approches ciblées par « Selected Reaction Monitoring » (SRM, Figure 2.B) sur des instruments de type triple quadripôle ou « Parallel Reaction Monitoring » (PRM, Figure 2.C) sur des instruments à haute résolution vont permettre de cibler l'acquisition sur des peptides préalablement sélectionnés. La combinaison de ces méthodes ciblées avec la dilution isotopique permet une quantification des HCPs hautement précise, répétable et

spécifique et a été l'approche initialement préconisée pour le dosage d'HCPs^{7, 10-11}. Néanmoins, le développement d'une méthode de quantification ciblée est long et fastidieux comparé à la mise en œuvre d'une approche globale de type DDA. De plus, ces stratégies sont toujours limitées à la sélection d'une centaine de candidats¹²⁻¹³.

Au cours de la dernière décennie, les développements en MS ont permis l'émergence d'un nouveau mode d'acquisition de type « Data Independent Acquisition » (DIA, Figure 2.D). Ce dernier repose sur le principe d'une co-isolation et co-fragmentation de l'ensemble des ions contenus dans des fenêtres de largeurs variables de m/z prédéfinies, pour couvrir l'ensemble de la gamme de masse. L'acquisition de tous les signaux MS2 des espèces détectables confère une spécificité, sensibilité, répétabilité et une précision de quantification comparables aux méthodes ciblées tout en maximisant le recouvrement des protéines présentes et donc des HCPs présents dans notre cas. Le facteur limitant de cette approche réside à ce jour dans les stratégies d'extraction des signaux quantitatifs fiables. Initialement, l'utilisation d'une librairie spectrale spécifique, générée à partir de données DDA pour extraire les informations quantitatives à partir des données DIA, a montré son efficacité. Néanmoins, la génération de cette librairie spectrale nécessite du temps et idéalement la mise en œuvre de stratégies de fractionnement du protéome étudié. Le développement récent d'algorithmes ne nécessitant pas de librairie spectrale accroîtra certainement encore l'intérêt et l'applicabilité des stratégies DIA, en particulier pour le suivi d'HCPs¹⁴⁻¹⁵.

C'est dans ce contexte que mes travaux de thèse ont été réalisés avec comme objectif de développer et d'évaluer des méthodes de préparation d'échantillons, de MS et de traitement des données, adaptées pour la détection et quantification globale des HCPs présentes à l'état de traces dans des préparations de molécules thérapeutiques.

Chapitre 2

Développement de stratégies analytiques de quantification des HCPs pour assister la chaîne de production d'anticorps thérapeutiques

1. Développement d'une stratégie innovante de quantification Top3 dédiée au suivi des HCPs

Parmi les approches de protéomique quantitative globale, la stratégie Top3 permet de quantifier les espèces détectées sans avoir recours à des standards marqués coûteux ou à une préparation supplémentaire des échantillons pour marquer les HCPs. Elle repose sur l'hypothèse que la somme de la réponse MS des trois peptides les plus intenses par mole de protéine est constante, avec un coefficient de variation (CV) inférieur à 10 %, pour estimer les quantités absolues de HCPs¹⁶. Un standard interne est utilisé pour calculer un facteur de réponse universel (signal des Top3 peptides /mol) et pour réaliser l'estimation de la quantité des HCPs (signal des Top3 peptides HCPs/facteur de réponse). L'utilisation d'un mélange de quatre protéines standard (Mix 4P) a été précédemment rapportée à cette fin, en considérant la protéine PYGM comme référence et les trois autres protéines (ADH, BSA et ENL) pour calculer les ratios, comme des contrôles internes¹⁷. Les stratégies développées dans ce manuscrit pour la quantification globale des HCPs sont destinées à être utilisées dans un environnement biopharmaceutique réglementé. Par conséquent, il convient d'envisager une standardisation de l'approche de quantification Top3, ce qui nécessite le développement de standards dédiés. Dans ce contexte, nous avons évalué les performances du standard HCP Profiler¹⁸, un kit prêt à l'emploi pour le suivi global des HCPs (Technologie READYBEADS, Anaquant).

Le standard HCP Profiler est une bille hydrosoluble qui libère des peptides non marqués en quantité connue. L'estimation de la quantité des HCPs est obtenue à partir de la réponse MS de 54 peptides rassemblés en 6 points de calibration dans une gamme de 1 à 500 fmol. Les 54 peptides proviennent de 18 protéines (arrangements de tripeptides synthétisés à partir de E. Coli) et chaque point de calibration contient 3 protéines standards. Ainsi, les 54 peptides permettent d'obtenir une courbe de calibration du $\text{Log}_2(\text{somme des aires des 3 peptides par protéines standards})$ en fonction du $\text{Log}_2(\text{quantité des protéines standards (fmol)})$.

Dans un premier temps, en collaboration avec Anaquant, nous avons réalisé une étude inter-laboratoire de la solution HCP Profiler à l'aide d'échantillons de protéines de cellule HeLa, digérés à la trypsine, et injectés sur des instruments de type Quadripôle-Orbitrap (Q-Orbitrap) (figure 3). Ainsi, trois réplicats techniques ont été injectés sur un Q Exactive HF (Anaquant) et Q Exactive HF-X (LSMBO) avec le standard HCP Profiler ajouté dans une gamme de 1 à 500 fmol. Après extraction des courants d'ions (XIC), la quantification Top3 des protéines a été réalisée sur le logiciel HCP Profiler développé par Anaquant.

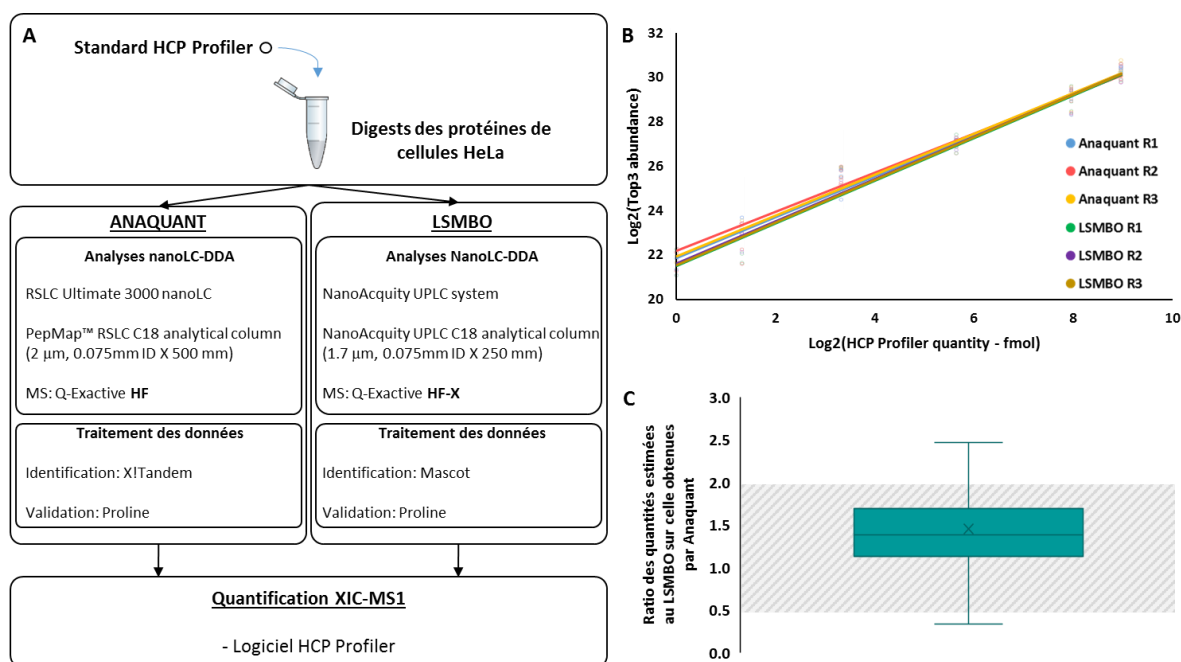


Figure 3: Etude inter-laboratoires du standard HCP Profiler. Schéma analytique utilisé pour quantifier, à l'aide du standard HCP Profiler combiné à une acquisition DDA, les protéines de cellules HeLa aux laboratoires d'Anaquant (Lyon) et du LSMBO (Strasbourg) (A). Superposition des courbes obtenues à partir de la quantification DDA XIC-MS1 sur un Q Exactive HF (Anaquant) et un Q Exactive HF-X (LSMBO) (B). Comparaison inter-laboratoires des quantités individuelles estimées pour les 500 protéines les plus abondantes (C). Au total, 88% des 500 protéines communes ont été observées dans un facteur 2 (zone grise).

Malgré les différences entre les systèmes chromatographiques, les spectromètres de masse et les moteurs de recherche utilisés pour les traitements de données, les courbes de calibration ont montrées une répétabilité significative (Figure 3.B). Les coefficients de variation (CVs) calculés sur les pentes, les ordonnées à l'origine et les coefficients de régression sont inférieurs à 3 % pour les six courbes obtenues. Ainsi, lorsque l'on compare les protéines quantifiées sur les deux plateformes MS, on remarque que 88% des estimations fournies par le standard HCP Profiler sont cohérentes dans un facteur 2 (Figure 3.C). Ces résultats affichent donc la reproductibilité inter-laboratoires du standard.

Dans un second temps, nous avons confronté notre méthode de quantification actuelle (Mix 4P) au standard HCP Profiler pour la quantification des HCPs présentes dans des échantillons d'anticorps mAb A33 produits à partir de cellules ovariennes de hamster chinois (CHO) : quatre fluides de culture cellulaire clarifiés (CCCF, Harvest) et sept éluats de chromatographie d'affinité par protéine A (PPA). Ces derniers possèdent différents degrés de complexité avec une gamme dynamique mAbs/HCPs croissante.

Les échantillons ont été digérés sur gel. Ensuite, le standard iRT (Biognosys), pour la normalisation des temps de rétention, et une bille du HCP Profiler ont été ajoutés dans chaque échantillon. Trois réplicats d'injections de 400 ng de protéines ont été réalisés sur le Q Exactive HF-X, opéré en mode DDA. Finalement, l'identification des peptides a été réalisée sur Mascot et la validation et les XICs ont été réalisés par Proline¹⁹. Après l'extraction des signaux MS1, un script R a été utilisé afin de filtrer les peptides puis quantifier les HCPs à l'aide de la courbe de calibration du HCP Profiler (Figure 4).

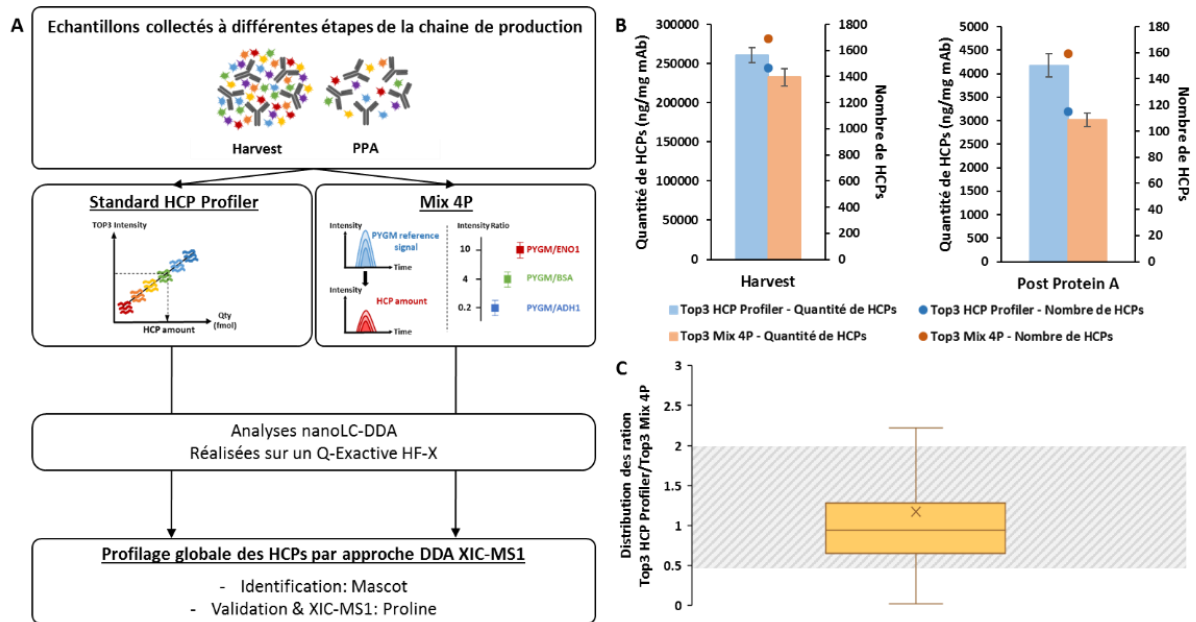


Figure 4: Le standard HCP Profiler versus Mix 4P pour le suivi des HCPs. Schéma analytique utilisé pour quantifier les HCPs restants dans des échantillons collectés après culture (Harvest) et après chromatographie d'affinité par protéine A (PPA) (A). Quantités globales de HCPs et nombres de HCPs quantifiées moyens obtenus pour les quatre échantillons Harvest et pour les sept fractions PPA (B). La hauteur des barres représente la moyenne de la quantité globale de HCPs dans les triplicats d'injection de chaque échantillon. Les barres d'erreur représentent l'écart-type. Représentation des ratios calculés entre les quantités obtenues par les approches Top3 HCP Profiler-DDA et de Top3 Mix 4P-DDA pour les 5305 HCPs communes aux deux stratégies (C). Les rapports hors limites ont été supprimés. Au total, 78% des 5305 HCPs communes ont été observées dans un facteur 2 (zone grise).

En moyenne, la méthode Top3 HCP Profiler-DDA a permis de quantifier 1464 HCPs dans les fractions CCCF avec des quantités globales comprises entre 222 646 et 365 145 ng/mg, et 115 HCPs dans les fractions PPA représentant 569 à 19 153 ng/mg. Ces résultats sont en accord avec les quantités globales obtenues en utilisant le facteur de réponse du signal de la protéine standard PYGM (Figure 4.B). Ces résultats se sont également montrés cohérents au niveau individuel des HCPs. En effet, 78% des 5305 HCPs quantifiés avec les deux approches se trouvent dans un facteur 2 (Figure 4.C). Ce qui est conforme aux études précédentes qui ont démontrées la variabilité de la stratégie Top3^{9, 20}.

En conclusion, le standard HCP Profiler est apparu comme une solution efficace. L'utilisation d'une bille hydrosoluble, qui libère des quantités contrôlées de peptides non marqués, permet d'éviter les biais analytiques induits par l'utilisateur. De plus, les courbes de calibration en 6 points obtenues, couvrant 2,7 ordres de grandeur, assurent la confiance et la précision de la quantification dans une gamme raisonnable. En conséquence, ce kit prêt à l'emploi présente une reproductibilité intra et inter-laboratoires qui permettra la comparaison des lots de production et le suivi des impuretés tout au long de la chaîne de production.

2. Évaluation des stratégies DIA pour le suivi global des HCPs sur un instrument de type Q-orbitrap

L'amélioration de la vitesse d'acquisition sur les instruments de géométrie Quadripôle-Orbitrap (Q-Orbitrap) (40Hz) a permis d'envisager le développement de stratégies DIA, au début de ma thèse, sur un instrument de type Q Exactive HF-X²¹ (Thermo Fischer Scientific). Les développements de méthodes ont été réalisés sur les quatre fractions Harvest et sept échantillons PPA du mAb A33 discutés au sein de la section 1 de ce chapitre. Dans un premier temps, une méthode DIA a dû être développée pour chaque type d'échantillon, en se basant sur la distribution des précurseurs sur la plage d'acquisition des ratios masse sur charge (m/z) observée en DDA et les temps de scan MS et MS/MS de l'instrument. Deux méthodes composées de 40 fenêtres d'isolation variables ont pu être développées permettant d'obtenir une médiane de 6 cycles MS par pic chromatographique. Dans un second temps, le fractionnement par électrophorèse sur gel de polyacrylamide en présence de dodécylsulfate de sodium (1D SDS-PAGE) d'un échantillon de lignée cellulaire CHO n'exprimant pas de mAb (null cell line), a permis la création d'une librairie spectrale la plus exhaustive possible des HCPs de ce type de cellule. Les échantillons ont été digérés sur gel. Ensuite, le standard iRT (Biognosys), pour la normalisation des temps de rétention, et une bille du HCP Profiler ont été ajoutés dans chaque échantillon. Trois réplicats d'injections de 400 ng de protéines ont été réalisés sur le Q Exactive HF-X opéré en mode DIA. Après l'acquisition des données et l'extraction des signaux peptidiques par le logiciel Spectronaut, nous avons pu comparer les résultats des méthodes DIA, avec et sans l'utilisation d'une librairie spectrale, aux données DDA acquises sur les mêmes échantillons (Figure 5).

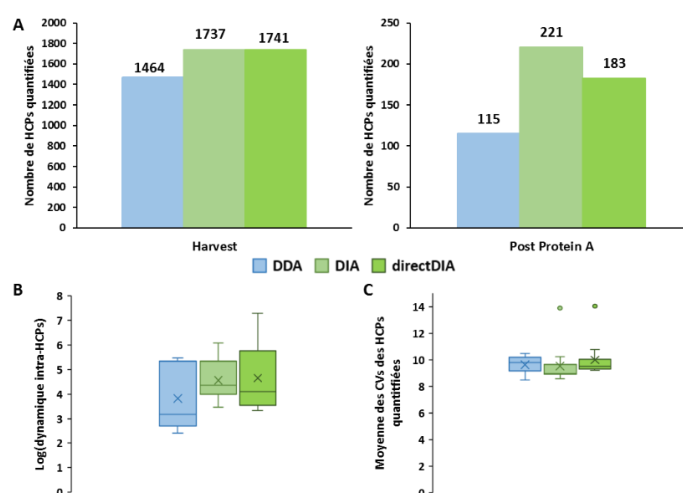


Figure 5: Evaluation des approches DIA, avec et sans librairie spectrale, contre la stratégie DDA XIC-MS1. Moyenne des nombres de HCPs quantifiées au sein des quatre échantillons Harvest et des sept fractions PPA (A). Représentation de la gamme dynamique intra-HCPs obtenue pour les trois méthodes DDA, DIA et directDIA (B). Comparaison des CVs moyens calculés sur les intensités des peptides HCPs dans les trois réplicats des analyses DDA, DIA et directDIA (C). La DIA correspond à l'approche centrée sur les peptides (avec librairie) et la directDIA correspond à l'approche centrée sur les spectres (sans librairie), toutes deux implémentées au sein de Spectronaut.

Le gain obtenu avec les méthodes DIA par rapport à la DDA est significatif quel que soit le type d'échantillons CCCF ou PPA. Une augmentation moyenne du nombre de HCPs quantifiées de 19% pour

les échantillons Harvest et de 75% pour les PPA a été observée (Figure 5.A). De plus, les méthodes DIA ont montrées une sensibilité supérieure à la DDA avec des résultats de quantification plus robustes et précis (Figure 5.B et Figure 5.C). Lorsque l'on compare les méthodes d'extraction des données DIA, on remarque que 62% des HCPs quantifiées sont communes aux deux méthodes avec une extraction des signaux similaire. En effet, 82% des 5833 quantités de HCPs estimées par DIA et directDIA sont cohérentes dans un facteur 2. Ces observations sont encourageantes pour l'approche sans librairie spectrale. Néanmoins, ces protéines communes ont été quantifiées à l'aide de 11 436 peptides pour la DIA et 11 039 peptides pour la directDIA. Ainsi, les 346 peptides en moins utilisés pour quantifier les HCPs communes par directDIA soulignent les difficultés qui existent encore pour extraire les impuretés à l'échelle de trace dans l'échantillon.

En conclusion, l'approche DIA, avec extraction des données via librairie spectrale, a montré une amélioration majeure de la couverture des HCPs par rapport à la méthode DDA. De ce fait, la combinaison avec le standard HCP Profiler permet une quantification plus robuste et précise des HCPs tout en atteignant une sensibilité inférieure au ng/mg de mAb. La méthode ainsi développée peut-être implémentée au sein d'un environnement biopharmaceutique pour surveiller la teneur en HCPs des échantillons tout au long de la chaîne de production des anticorps monoclonaux.

3. Étude de cas : Impact de la capacité du bioréacteur sur le profil des HCPs

La méthode DIA précédemment développée a été appliquée pour répondre à une question spécifique adressée par UCB Pharma : l'intensification de la production des mAbs par l'utilisation de bioréacteurs de plus grande capacité a-t-elle un impact sur le profil des HCPs ? Dans la littérature, des modifications telles que la composition du milieu de culture, des changements de température ou des modifications génétiques ont montré un impact sur le profil des HCPs^{4, 22}. Or, lors du développement des mAbs de la phase recherche jusqu'à la phase de commercialisation, des quantités toujours plus importantes du produit médicamenteux sont nécessaires, requérant l'utilisation de bioréacteurs de volumes croissants. De manière générale, la capacité de ces bioréacteurs va de 80 litres jusqu'à 25 000 litres. Afin de réaliser cette étude, neuf échantillons Harvest et neuf échantillons PPA provenant de quatre anticorps produits à partir des cellules CHO au sein de bioréacteurs allant de 80 à 15 000 litres ont été fournis par UCB (Tableau 1).





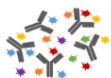
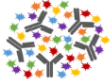
Type d'échantillon	Capacité du bioréacteur (L)	mAb 1 	mAb 2 	mAb 3 	mAb 4 
 PPA	80			X	X
	200	X			X
	2 000	X	X	X	X
	15 000			X	
 Harvest	80	X		X	
	200	X	X		
	2 000	X	X	X	X
	15 000			X	

Tableau 1: Série d'échantillons étudiée afin d'observer l'impact de l'augmentation des quantités de mAb produites. La série est composée de neuf fractions Harvest et neuf PPA provenant de quatre mAbs produits à partir de cellules CHO DG44. Les échantillons ont été cultivés dans des bioréacteurs de 80 L à 15 000 L.

De nombreuses études ont montré l'impact du mAb produit sur les protéines co-exprimées et co-éluées au cours de l'USP et DSP²³⁻²⁴. Par conséquent, nous pouvons questionner l'utilisation de notre librairie spectrale spécifique au projet IgG4 mAb A33 développée au sein de la section 2 de ce chapitre. Les HCPs qui n'ont pas été identifiées dans les analyses DDA de la "null cell line" et qui co-exprimeraient préférentiellement avec l'un des quatre mAbs plutôt qu'avec le mAb A33 seront absentes de la librairie spectrale et ne seront pas ciblées lors de l'extraction des données DIA. Par ailleurs, la création d'une librairie spécifique aux quatre anticorps nécessiterait une ou deux semaines afin de réaliser le fractionnement sur gel d'échantillons Harvest de ces mAbs et l'acquisition DDA de tous les échantillons. Dans le but de développer une méthode directe, nous avons décidé d'évaluer l'utilisation d'une librairie spectrale hybride, générée à partir des données DDA et DIA¹⁵. Afin de générer cette librairie hybride, nous avons combiné les analyses DIA des échantillons Harvest et PPA des quatre mAbs avec l'archive de recherche de la librairie DDA spécifique au mAb A33. Après la génération d'une librairie spectrale par Spectronaut, le résultat de la recherche Pulsar est sauvegardé en tant qu'archive avant l'application de tout filtre FDR (False Discovery Rate) visant à contrôler le taux de faux-positifs dans la librairie. Cela permet de combiner cette archive avec d'autres analyses ou archives de recherche pour enrichir une librairie tout en ayant un contrôle adéquat du FDR à l'échelle de la librairie spectrale. Comparé à la librairie DDA de la section 2, un nombre de peptides et de groupes de protéines similaires ont été obtenus avec un recouvrement de 88% et 84%, respectivement. On peut noter 274 HCPs supplémentaires identifiées au sein de la librairie hybride.

Les échantillons ont été digérés sur gel puis le standard iRT ainsi que le mix de quatre protéines standards (Mix 4P) ont été ajoutés. Les échantillons ont été injectés en triplicat sur le Q Exactive HF-X opéré en mode DIA. Après l'acquisition des données et l'extraction des signaux peptidiques par le logiciel Spectronaut, les HCPs ont été quantifiées par l'approche Top3 à l'aide du facteur de réponse de PYGM. Ainsi, les résultats de quantification obtenus par l'extraction via librairie hybride ont été comparés aux résultats obtenus avec la librairie DDA développée en section 2 de ce chapitre (Figure 6).

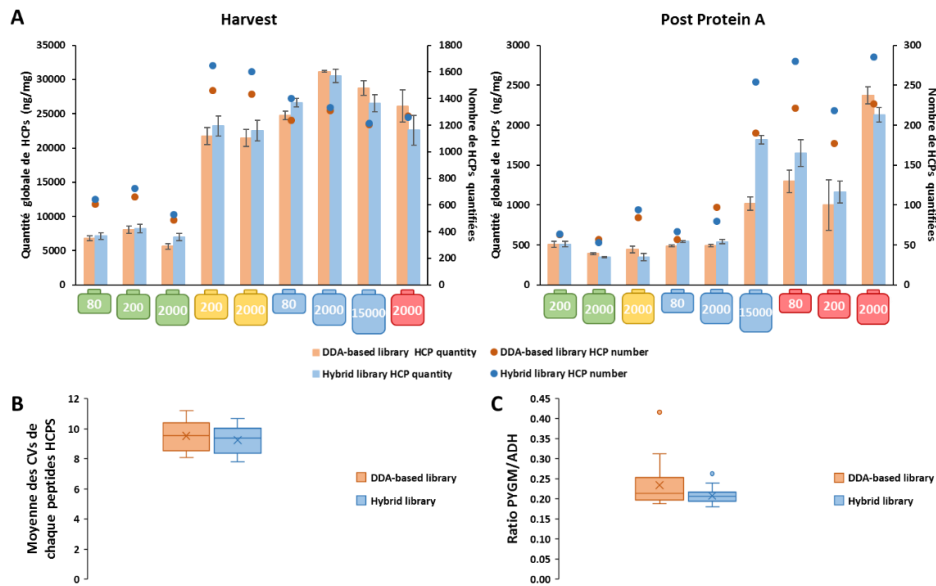


Figure 6: Apport de l'utilisation d'une librairie spectrale hybride pour extraire les données DIA. Représentation des quantités globales de HCP obtenues et du nombre de HCPs quantifiées à l'aide des librairies spectrales hybrides et DDA (A). Les points représentent les nombres de HCPs quantifiées. Les hauteurs des barres représentent la moyenne de la quantité globale de HCPs dans les triplicats d'injection. Les barres d'erreur représentent l'écart-type. Distribution des CVs moyens calculés sur les intensités des peptides HCPs dans les trois réplicats de chaque échantillon en utilisant la librairie hybride ou DDA (B). Distribution des ratios entre les aires de la somme des trois peptides les plus intenses de PYGM sur les aires du Top3 de la protéines standard ADH (C). La valeur attendue est de 0.2 (2 fmol injecté de PYGM / 10 fmol injecté de ADH).

En général, les résultats obtenus sont cohérents en termes de quantités globales et de nombres de HCP quantifiées (Figure 6.A). Il est à noter que la librairie hybride a permis de quantifier un plus grand nombre d'impuretés protéiques pour presque tous les échantillons. En examinant plus en détail les résultats, nous avons observé une légère amélioration de la reproductibilité de l'extraction des signaux avec une médiane des CVs calculés sur les intensités des peptides dans les réplicats de 9,4 pour l'approche avec librairie hybride et de 9,6 pour l'approche avec librairie DDA (Figure 6.B). Les rapports PYGM/ADH obtenus avec la librairie spectrale hybride sont également plus précis autour de la valeur attendue de 0,2 (Figure 6.C). L'ajout des données DIA à l'archive de recherche de la librairie DDA du mAb A33 a fourni des valeurs iRT d'une autre source de données, celle des analyses d'intérêt elles-mêmes. Ensuite, Spectronaut utilise la meilleure source iRT disponible afin de normaliser les temps de rétention et d'extraire les signaux. Par conséquent, nous pouvons conclure que la librairie spectrale hybride, générée avec les analyses DIA d'intérêt, a amélioré l'extraction des signaux peptidiques.

Finalement, nous avons utilisé les résultats de la stratégie basée sur la librairie hybride pour comparer le contenu en HCPs des échantillons des quatre mAbs (Figure 7).

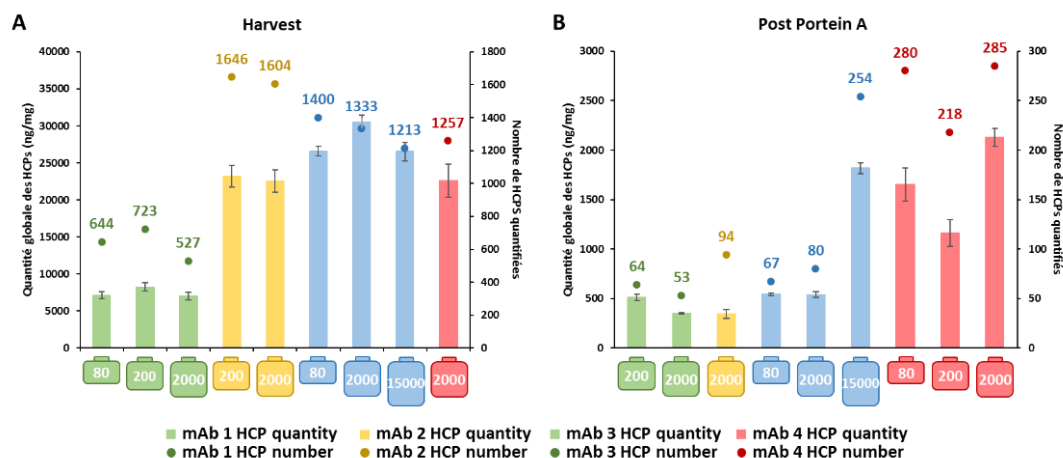


Figure 7: Évaluation de l'impact de la capacité du bioréacteur sur le profil des HCPs. Quantités globales de HCP et nombre de HCP quantifiés obtenus pour les fractions Harvest (A) et PPA (B) en utilisant la stratégie Top3 Mix 4P-DIA. La hauteur des barres représente la moyenne des quantités globales de HCPs dans les triplicats d'injection. Les barres d'erreur représentent l'écart-type. Les points et les valeurs représentent le nombre de HCP quantifiés.

À la fin de l'USP, nous avons observé des résultats de quantification similaires pour les échantillons Harvest provenant de la production du même mAb (Figure 7.A). En se focalisant sur l'impact de la capacité du bioréacteur, aucune tendance n'a émergé. Les faibles variations obtenues dans les fractions d'une même production de mAb ne permettent pas de conclure sur l'impact de la capacité du bioréacteur. Ainsi, au lieu d'un impact du bioréacteur, nos résultats suggèrent plutôt un effet du biothérapeutique produit sur les HCPs co-exprimées. De plus, les variations observées pour les fractions PPA des anticorps mAb 3 et mAb 4 montrent l'importance de suivre la teneur en HCPs tout au long de la chaîne de production (Figure 7.B). Cela permettra une meilleure compréhension de l'impact de chaque étape de culture ou de purification et ainsi d'éviter la plupart des problèmes liés aux HCPs.

4. Avantage de la mobilité ionique pour le suivi des HCPs

Au cours des dernières années, les spectromètres de masse ont connu des améliorations de leur vitesse d'acquisition qui ont permis l'implémentation de la mobilité ionique comme méthode de séparation supplémentaire²⁵. Dans ce contexte, j'ai pu évaluer l'apport d'un instrument de type TimsTOF PRO (Bruker Daltonics) pour la caractérisation des HCPs. Cet instrument de type Quadripôle-Temps de vol (Q-TOF) intègre une cellule de spectrométrie de mobilité des ions piégés très innovante (TIMS). Cette cellule va permettre de séparer des ions qui co-éluent en sortie de LC par rapport à leur forme et charge. En combinaison avec le mode de scan appelé "parallèle accumulation – fragmentation en série" (PASEF), le TimsTOF PRO permet d'obtenir des centaines de scan MS/MS par seconde sans perte de sensibilité²⁶. Dans le cadre des HCPs, l'apport d'une dimension de séparation supplémentaire permettant de séparer des peptides d'HCPs de ceux du mAb avec lesquels ils co-

éluent, paraît évident. Les développements de méthode ont été réalisés sur les quatre fractions Harvest et sept échantillons PPA du mAb A33 discutés au sein de la section 1 de ce chapitre.

Les échantillons ont été digérés sur gel. Ensuite, le standard iRT (Biognosys), pour la normalisation des temps de rétention, et une bille du HCP Profiler ont été ajoutés dans chaque échantillon. Trois réplicats d'injections de 273 ng de protéines ont été réalisés sur le TimsTOF PRO opéré en mode ddaPASEF et diaPASEF. Après l'acquisition des données et l'extraction des signaux peptidiques par les logiciels MaxQuant (ddaPASEF) et Spectronaut (diaPASEF), nous avons pu comparer les performances des deux méthodes d'acquisitions (Figure 8).

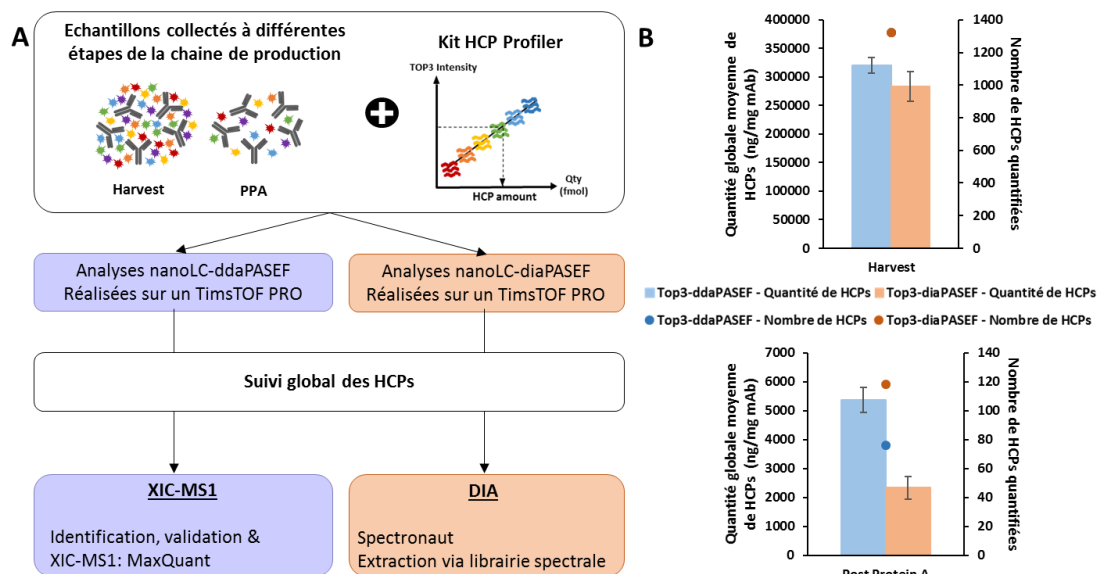


Figure 8: Evaluation des approches ddaPASEF et diaPASEF du TimsTOF PRO pour le suivi des HCPs. Schéma analytique utilisé pour quantifier les HCPs restants dans des échantillons collectés après culture et après chromatographie d'affinité par protéine A (A). Quantités globales de HCPs et nombres de HCPs quantifiés moyens obtenus pour les quatre échantillons Harvest et pour les sept fractions PPA (B). La hauteur des barres représente la moyenne de la quantité globale de HCPs dans les triplicats d'injection de chaque échantillon. Les barres d'erreur représentent l'écart-type.

Pour commencer, à cause d'une trop grande sensibilité du TimsTOF PRO, nous n'avons pas pu injecter les 400 ng de protéines précédemment injectées sur le Q Exactive HF-X. Nous avons injecté 273 ng de protéines avec le standard HCP Profiler dans une gamme de 0,68 à 341 fmol.

Ensuite, les résultats de quantification présentés en Figure 8.B ont montrés une plus grande couverture des HCPs avec l'approche diaPASEF. En accord avec une couverture plus élevée, la stratégie diaPASEF a présenté une gamme dynamique intra-HCPs plus élevée ainsi qu'une sensibilité accrue. Cette sensibilité accrue a également été observée pour les peptides standard du kit HCP Profiler. En effet, les 54 peptides standards ont été identifiés et quantifiés par diaPASEF alors que l'approche ddaPASEF n'a pas permis l'extraction du signal d'un peptide standard du point de calibration le plus bas à 0.68 fmol pour les quatre fractions Harvest.

Néanmoins, malgré un nombre de HCPs quantifiées supérieur, la stratégie diaPASEF à afficher une quantité globale d'impuretés inférieure pour presque tous les échantillons. De plus, une médiane des CVs calculés sur l'intensité des peptides HCPs dans les répliquats de 9,2 a été obtenue par diaPASEF comparé à 7,8 par ddaPASEF (Figure 9.A). Ceci montre une précision de quantification inférieure de l'approche diaPASEF. En examinant de plus près les signaux des peptides standard du HCP Profiler, nous avons observé des problèmes d'extraction des signaux des données diaPASEF. En effet, environ 25% des peptides standards ont montré une extraction des signaux incomplète ou partielle dans un des répliquats. Par conséquent, les courbes de calibration obtenues en diaPASEF se trouvent bien inférieure à celles de la ddaPASEF et les quantités globales de HCPs sont sous-estimées (Figure 9.B).

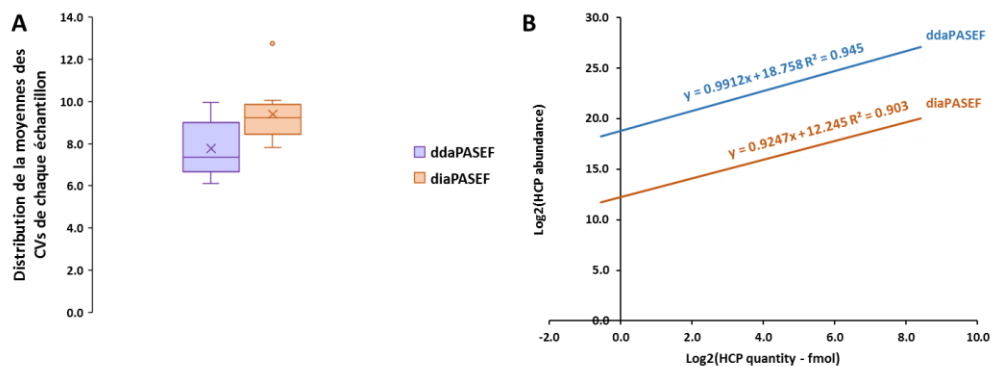


Figure 9: Limite de l'approche diaPASEF. Distribution des CVs moyens calculés sur les intensités des peptides HCPs dans les trois répliquats de chaque échantillon par ddaPASEF et diaPASEF (A). Superposition des courbes de calibration obtenues par les méthodes ddaPASEF et diaPASEF (B). Les courbes de calibration présentées ont été obtenues en faisant la moyenne de la pente, de l'ordonnée à l'origine et des coefficients de régression des 33 courbes de calibration obtenues par ddaPASEF et diaPASEF.

En conclusion, l'intégration du dispositif TIMS à l'entrée d'un instrument Q-TOF fournit une dimension de séparation supplémentaire après l'élution de la LC. Par conséquent, les méthodes d'acquisition ddaPASEF et diaPASEF laissent entrevoir une couverture plus importante des impuretés protéiques dans nos échantillons de mAb sans préparation supplémentaire de l'échantillon. Malheureusement, le TimsTOF PRO dans sa version actuelle présente une sensibilité accrue qui limite la quantité de matière pouvant être injectée. Cette limitation représente un inconvénient dans le contexte des HCPs. De plus, malgré des résultats d'identification supérieurs à ceux d'autres instruments sans mobilité ionique, les outils de traitement des données ne sont pas assez matures pour effectuer une quantification fiable de tous les signaux. En effet, ils rencontrent encore certaines difficultés qui conduisent à des extractions de signaux incorrectes.

Dans les années à venir, la méthode diaPASEF réalisée avec le TimsTOF PRO, associée à un logiciel perfectionné, devrait pouvoir être utilisée dans un environnement biopharmaceutique pour le suivi des HCPs pendant les développements de bioprocédés, la libération de lots ou la validation de kits ELISA.

Chapitre 3

Le défi analytique des produits pharmaceutiques purifiés

1. Amélioration de la préparation des échantillons pour l'analyse des substances médicamenteuses

Le produit final correspond au médicament prêt à être injecté au patient ayant subi une multitude d'étapes de purification afin de garantir des quantités d'impuretés à des niveaux tolérables⁷. Une gamme dynamique extrême entre le mAb et les HCPs présentes à l'état de traces caractérise donc ces échantillons. Au cours des dernières années, plusieurs stratégies ont été appliquées afin de contourner ce défi analytique majeur. On peut citer des méthodes de chromatographie 2D ou encore de déplétion du mAb²⁷⁻²⁹. C'est dans cette seconde direction que nous avons choisi d'aller afin de pouvoir proposer un protocole compatible avec une analyse de routine en milieu industriel. Le protocole développé met en œuvre une digestion en condition native utilisant une très faible quantité d'enzyme pour favoriser la digestion des HCPs tout en limitant la digestion du mAb fortement structuré pour pouvoir dans un second temps faire précipiter ce dernier à l'aide d'une étape de centrifugation (Figure 10)²⁹.

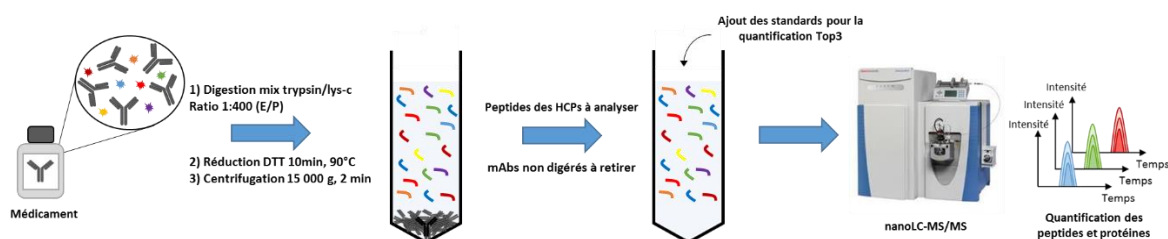


Figure 10: Développement d'une stratégie dédiée à l'étude des produits médicaments. Schéma du protocole de digestion en condition native pour la déplétion du mAb.

Sur la base du protocole de digestion en condition native proposé par Huang et al.²⁹, nous avons implémenté l'utilisation d'une double digestion trypsine/Lys-C destinée à améliorer l'efficacité protéolytique en réduisant les clivages manqués au niveau des résidus lysines³⁰. Ce protocole a été appliqué à des échantillons de NIST mAb, un anticorps monoclonal de référence produit par le National Institute of Standards and Technology (NIST). Ensuite, le standard iRT a été ajouté avec le mix de quatre protéines standard (Mix 4P : ADH, PYGM, BSA, ENL) dans chaque réplicat technique. Les échantillons ont été injectés sur le Q-Exactive HF-X en DDA et DIA. Les analyses DDA ont été utilisées afin d'optimiser une méthode DIA de 60 fenêtres d'isolation variables spécifique aux substances médicamenteuses. Ensuite, les signaux MS2 ont été extraits des données DIA à partir d'une librairie spectrale générée avec les analyses DDA du fractionnement sur gel d'un échantillon de NIST mAb. Il faut noter que les HCPs identifiées par Spectronaut ont été validées par l'approche target-decoy avec un FDR (taux de faux positifs) de 1% au niveau des peptides. En effet, le nombre de protéines identifiées étant faible, l'utilisation de cette approche au niveau des protéines conduirait à une estimation inexacte du taux de faux positifs.

A partir des résultats bruts d'identification, quatre niveaux de filtres ont été appliqués avant de procéder à la quantification des peptides et des protéines. Tout d'abord, nous avons conservé les peptides doublement et triplement chargés et nous avons éliminé les peptides oxydés et acétylés avec leurs homologues non modifiés. La plupart du temps, ces modifications et les peptides quatre fois chargés présentent des signaux de mauvaise qualité en termes d'intensité et de forme de pic. Le deuxième filtre peut être décrit comme un filtre sur la qualité des signaux. Les ions précurseurs présentant plus d'une Q value > 0,01 et/ou "Profiled" sur les trois répliquats d'injections ont été éliminés. Le troisième critère est lié à la reproductibilité du signal car les précurseurs ayant un CV supérieur à 20% sont filtrés. Ensuite, un filtre d'homologie des séquences est appliqué en utilisant la fonction BLASTP³¹ (v.2. 10.0+) contre les séquences des chaînes lourdes et légères du mAb. Ce dernier filtre n'élimine pas beaucoup de peptides mais c'est un filtre important car le mAb est la protéine la plus abondante dans l'échantillon. Un signal peptidique de l'anticorps attribué à tort à une HCP entraînerait alors inévitablement une surestimation de sa quantité. Finalement, la sélection des trois peptides les plus intenses a été effectuée et les quantités absolues de HCPs ont été estimées en utilisant le facteur de réponse de la protéine standard PYGM (Figure 11).

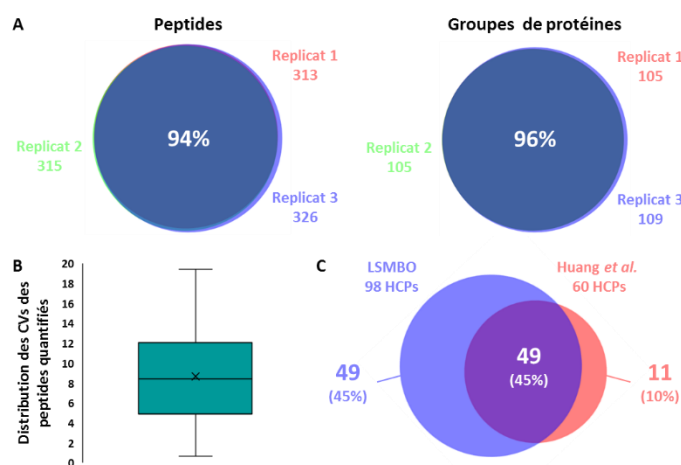


Figure 11: Performance du workflow HCPs dédié aux substances médicamenteuses. Recouvrement des peptides et des groupes de protéines identifiés dans trois répliques techniques d'échantillons de NIST mAb soumis à notre protocole optimisé de digestion native et analysés par nanoLC-DIA sur un Q Exactive HF-X (Thermo Fischer Scientific) (A). Distribution des CV calculés sur les intensités des peptides HCP dans les trois répliquats techniques (B). Recouvrement des HCP quantifiées dans le NIST mAb par l'approche Top3 Mix 4P-DIA avec la liste obtenue par Huang et al., sur un Synapt G2-S (Waters) (C).

Lorsque l'on applique ce schéma analytique à des répliquats techniques de NIST mAb, on remarque de suite la grande reproductibilité du protocole avec 94% et 96% de recouvrement entre les listes de peptides et groupes de protéines identifiés au sein des trois répliquats (Figure 11.A). Après l'application de nos filtres, nous avons pu quantifier 98 HCPs pour une quantité globale de 247 ng/mg de mAb. De plus, une médiane des CVs calculés sur les intensités des peptides HCPs de 8,4% a été obtenue, démontrant la précision de l'extraction DIA sur les trois répliquats techniques de ces échantillons purifiés (Figure 11.B). Finalement, en comparant nos résultats avec ceux de Huang et al., nous avons remarqué que 82% des HCPs identifiées sur leur instrument Q-TOF ont été couvertes par notre méthode DIA, réalisée sur le Q Exactive HF-X, avec 49 HCPs supplémentaires (Figure 11.C). Des résultats intéressants, étant donné que l'on peut trouver différents spectromètres de masse sur différents sites de production

de médicaments. Finalement, ce protocole a montré une reproductibilité et une précision de quantification similaires aux méthodes développées dans le chapitre 2, en utilisant les fractions Harvest et PPA, tout en diminuant la gamme dynamique mAb/HCP qui représente le goulot d'étranglement de l'analyse des substances médicamenteuses.

Pour améliorer encore le protocole de digestion native, nous avons implémenté une étape de micro-mélange sur le temps de digestion au sein d'un ultrasonicateur Covaris. Covaris fabrique des appareils qui délivrent des ultrasons focalisés, ou acoustique focalisée adaptée (AFA), qui permettent de focaliser l'énergie acoustique directement sur l'échantillon, améliorant ainsi l'efficacité de l'énergie délivrée avec moins de chaleur globale. Ainsi, la double digestion trypsine/Lys-C a été réalisée dans le bain du système Covaris à 30°C avec une énergie AFA de 2.5 W appliquée en continu. Deux temps d'incubation ont été testés : 10 et 30 min.

Une fois digérés, le standard iRT a été ajouté avec le mix de quatre protéines standard (Mix 4P : ADH, PYGM, BSA, ENL) dans chaque échantillon. Les trois réplicats techniques ont été injectés sur le Q Exactive HF-X opéré en mode DIA et les résultats de quantification ont été comparés à ceux de l'échantillon contrôle présenté en figure 11 (Figure 12).

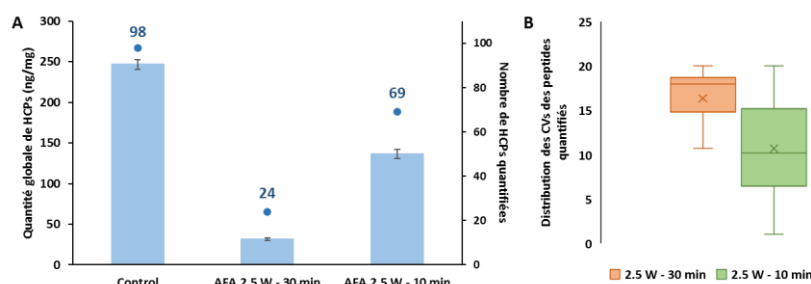


Figure 12: La technologie AFA de Covaris pour améliorer la digestion native. Quantité globale de HCPs et nombre de HCPs quantifiées dans l'échantillon témoin digéré pendant la nuit et dans deux échantillons de NIST mAb soumis à une digestion de 10 et 30 min sur le système Covaris (A). Les hauteurs des barres représentent les moyennes de la quantité globale de HCPs dans les réplicats techniques. Les barres d'erreur représentent l'écart type des quantités globales de HCPs dans les réplicats techniques. Les points représentent le nombre de HCPs quantifiées. Distribution des CVs calculés sur les intensités des peptides HCPs dans les trois réplicats techniques (B).

Il est intéressant de noter que 69 HCPs ont été quantifiées dans l'échantillon digéré pendant 10 minutes avec une ultrasonication continue à 2,5 W (Figure 12.A). Comparé aux 98 HCPs quantifiées dans l'échantillon contrôle qui a été soumis à une digestion pendant la nuit (14 heures), ces résultats sont prometteurs. En effet, cela permettrait de diminuer le temps de préparation des échantillons. Cependant, nous avons quantifié 29 HCPs de moins avec cette digestion ultrarapide. Un nombre non négligeable dans notre contexte. L'échantillon de mAb du NIST digéré pendant 30 minutes a montré 24 HCPs quantifiés. Il a donc montré une couverture des HCPs plus faible que l'échantillon contrôle et celui digéré en 10 min. Ce résultat s'explique par une faible répétabilité de l'extraction des signaux au sein des trois réplicats de l'échantillon digéré en 30 min (Figure 12.B). En effet, une perte d'échantillon au sein des deux premier réplicats de l'échantillon digéré en 30 min a été observée. Le Covaris S220 fonctionne pour un seul échantillon à la fois et nécessite l'utilisation de tubes spécifiques. Il est nécessaire de transférer l'échantillon dans le tube Covaris. Ensuite, pour l'étape finale de centrifugation, l'échantillon doit être transféré dans un Eppendorf. Ces transferts, notamment le

transfert du tube Covaris à l'Eppendorf, entraînent des pertes d'échantillon car le tube Covaris contient une fibre qui gêne la récupération de l'échantillon. En conclusion, ces résultats prometteurs nécessitent une nouvelle expérience dédiée pour confirmer l'intérêt de l'étape d'ultrasonication pour améliorer le protocole de digestion native.

2. Étude de cas : Enquête sur le changement de procédé de production d'un anticorps monoclonal

Au cours du développement d'un anticorps, UCB Pharma a dû revoir le processus de purification afin de diminuer la quantité de HCPs dans la substance médicamenteuse. Alors que les quantités mesurées par un kit ELISA commercial ont montrées une diminution de la teneur en impuretés, l'utilisation d'un kit ELISA interne n'a pas confirmé cette conclusion. Ces résultats contradictoires ont conduit à une enquête supplémentaire afin de déterminer si le problème venait du kit ELISA interne, qui est censé être plus spécifique aux échantillons, ou si le nouveau procédé de production ne répondait pas aux attentes. Ainsi, l'équipe de caractérisation des protéines d'UCB Pharma basée à Slough a effectué des analyses LC-MS/MS en mode DDA sur un instrument de type Q Exactive. De plus, nous avons été impliqué dans cette enquête et cinq échantillons provenant des deux procédés de production ont été envoyés à Strasbourg, ce qui nous a permis d'appliquer notre stratégie DIA sur un cas concret.

Les deux échantillons du procédé 1 et les trois échantillons du procédé 2 ont été digérés en condition native, comme décrit en section 1 de ce chapitre. Ensuite, Le mélange de quatre protéines standard du kit de digestion MassPREP de Waters a été ajouté aux échantillons avec le standard iRT. Finalement, les échantillons ont été injectés en triplicat sur le Q Exactive HF-X opéré en mode DIA. Le facteur de réponse de la protéine PYGM a été utilisé pour estimer la quantité des HCPs (Figure 13).

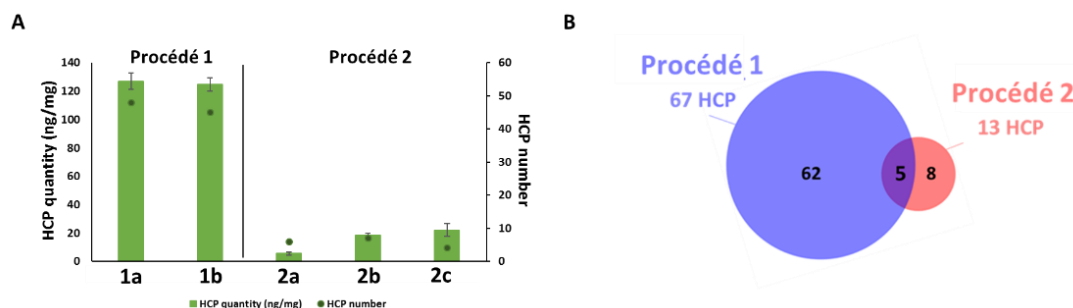


Figure 13: Évaluation de la teneur en HCPs dans des substances médicamenteuses provenant de deux procédés de production différents. Quantification globale et nombre de HCPs quantifiées en DIA sur les échantillons des procédés de production 1 et 2 (A). Diagramme de Venn représentant la corrélation des listes de HCPs obtenues pour les deux procédés de production (B).

L'approche DIA basée sur la quantification des signaux MS2 a été capable de quantifier une moyenne de 47 HCPs dans les échantillons du procédé 1 avec des quantités globales entre 125 et 127 ng/mg mAb. Pour le procédé 2, une moyenne de 6 HCPs ont été quantifiés représentant 6 à 22 ng/mg mAb (Figure 13.A). Ainsi, les résultats obtenus ne laissent aucun doute sur la conclusion. En moyenne, le procédé 2 a permis de diminuer la quantité de HCPs d'un facteur 8 avec plus de 40 HCPs en moins. L'objectif de réduction de la teneur en impuretés protéiques a donc été atteint grâce au développement de ce nouveau protocole de purification.

De plus, UCB Pharma avait déjà réalisé, en interne, une évaluation des risques liés aux HCPs identifiées au sein du procédé 1. Comme le procédé de production lui-même peut influencer le profil des HCPs, nous avons comparé en détail les HCP identifiées et quantifiées dans les deux processus (Figure 13.B). Les HCPs quantifiées dans les échantillons 1.A et 1.B ont été combinées en une seule liste pour le procédé 1 et la même chose a été faite pour le procédé 2 avec les échantillons 2.A, 2.B et 2.C. Le diagramme de Venn a révélé 6 HCPs quantifiées seulement dans les échantillons du procédé 2. Finalement, sur la base de ces résultats, UCB Pharma doit maintenant réaliser une nouvelle évaluation des risques en se concentrant sur les 6 protéines identifiées uniquement dans le procédé 2, ce qui permettra de voir si le procédé de production doit encore être amélioré ou non.

En conclusion, la méthode Top3-DIA a pu être appliquée à un cas concret. Elle a permis de quantifier des dizaines de HCPs dans cinq substances médicamenteuses issues de deux procédés de production. La méthode s'est avérée robuste, avec une moyenne des coefficients de variation calculés sur les intensités peptidiques inférieure à 13%, et sensible avec une quantification d'espèces inférieure au ng/mg de mAb. Cette stratégie représente une alternative prometteuse à l'ELISA comme méthode de contrôle de qualité pour une utilisation en routine. Les résultats discutés dans ce chapitre ont été intégrés dans la demande de commercialisation d'un anticorps auprès des organismes de réglementation.

Chapitre 4

Développement d'une méthode de suivi des HCPs sur une plateforme MS pour l'analyse de routine

Le projet a été mené dans le cadre d'une collaboration entre UCB Pharma et Waters, tous deux intéressés par l'évaluation du système LC-MS BioAccord pour la caractérisation des HCPs. Cet instrument, fabriqué par Waters, vise à fournir une solution simple d'utilisation pour les compagnies biopharmaceutiques. Le système est entièrement contrôlé par le logiciel UNIFI, qui permet une utilisation efficace du BioAccord avec la création de workflow analytique de l'acquisition au traitement des données. En outre, le logiciel tout-en-un comprend un système complet de traçabilité pour l'acquisition, le traitement et le rendu de rapport, ce qui limite le risque d'erreurs et facilite la préparation des audits.

Le BioAccord fonctionne avec une source d'ionisation ESI et un détecteur ACQUITY RDa. Deux modes d'acquisition sont proposés : full scan et full scan avec fragmentation. La deuxième option procède à la fragmentation en source des peptides ionisés avec l'application d'une rampe d'énergie au niveau de la tension de cône. Le système alternera un scan avec une faible tension de cône (peptides) et avec une rampe d'énergie élevée (fragments). Nous nous sommes évidemment tournés vers ce second mode d'acquisition pour caractériser les HCPs et avons travaillé sur l'échantillon de référence NIST mAb pour l'évaluation de l'instrument.

Le système BioAccord est uniquement composé d'une source d'ionisation, de lentilles pour guider les ions et d'un analyseur TOF. Il ne permet donc pas de sélectionner les ions pendant l'analyse. Ainsi, les scans MS et MS/MS contiendront tous les ions précurseurs et fragments contenus dans la gamme d'acquisition de 50 à 2000 m/z. De plus, le logiciel UNIFI n'est pas un logiciel de protéomique permettant l'analyse de centaines de milliers de protéines. Par exemple, il n'est possible d'ajouter qu'une seule séquence de protéines à la fois dans le logiciel. Ceci est une limitation pour les mAbs produits à partir de cellules CHO car l'organisme CHO contient actuellement 56 569 entrées protéiques. C'est pourquoi, nous avons utilisé une méthode UNIFI, appelée "*Accurate mass screening on MSe data*", qui permet d'identifier les ions précurseurs et de réaliser l'extraction des courants d'ions (XIC) des peptides ciblés contenus dans une librairie.

Deux expériences ont été réalisées pour évaluer le système LC-MS BioAccord pour la caractérisation des HCP dans le NIST mAb. Pour les deux expériences, le NIST mAb a été soumis au protocole de digestion native développé au sein de la section 1 du chapitre 3.

La première expérience a été conçue pour simuler la quantification des HCP en utilisant les quatre protéines standard du kit de digestion MassPREP. Ainsi, dans cinq échantillons de NIST mAb digérés, les protéines ADH, PYGM et ENL ont été ajoutées à des quantités croissantes de 20, 40, 80, 200 et 400 ng/mg de mAb. En parallèle, de la BSA a été ajoutée à une concentration constante de 100 ng/mg de mAb comme protéine de contrôle. Un échantillon de contrôle sans protéines standard dopées a également été préparé. Enfin, pour les six échantillons, cinq réplicats d'injections ont été effectués sur le BioAccord avec une quantité constante de 50 µg de mAb chargée sur la colonne. Ensuite, une

librairie de peptides a été créée sur UNIFI comprenant 27 peptides des protéines standards et 4 peptides du NIST mAb. Pour ces 31 peptides, les masses de 3 fragments ont été ajoutées à la librairie. Tous ces peptides ont été sélectionnés sur la base de leur détection et profil de fragmentation observés au sein d'analyses préliminaires effectuées sur le BioAccord. Ensuite, un traitement ciblé des données a été effectué en utilisant la méthode *Accurate Mass Screening on MSe data* avec des tolérances de masse sur les précurseurs et des ions fragments de 15 et 20 ppm, respectivement, et une tolérance sur les temps de rétention de 0,3 min. Les signaux identifiés et extraits ont été validés manuellement sur la base de leur profil isotopique, de la forme des pics chromatographiques et de l'identification des fragments. Enfin, l'abondance des peptides a été utilisée pour évaluer la performance du BioAccord pour extraire les signaux peptidiques de protéines dopées jusqu'à 20 ng/mg de mAb (Figure 14)

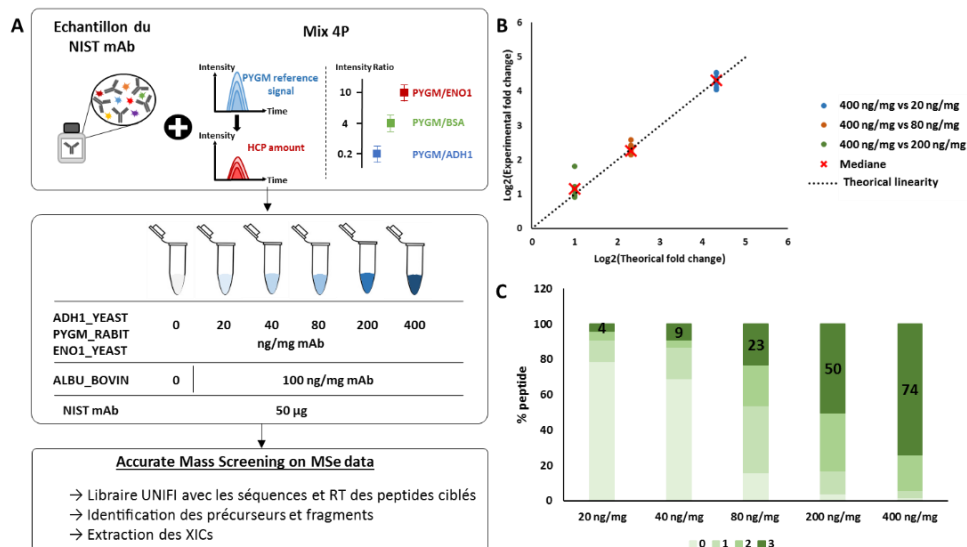


Figure 14: Simulation du suivi des HCPs. Schéma analytique suivi afin d'observer l'extraction du signal d'espèces jusqu'à 20 ng/mg de mAb (A) Le principe de l'expérience a été décrit dans le paragraphe ci-dessus. Les ratios expérimentaux de la somme des Top3 des protéines standard, calculés à partir du point le plus haut (400 ng/mg de mAb), ont été représentés en fonction des ratios théoriques (B). Le point à 40 ng/mg de mAb a été retiré à cause d'une erreur lors de la préparation d'échantillon. Représentation du pourcentage de peptides identifiés avec 0, 1, 2 ou 3 fragments sur les 3 attendus (C). Les nombres représentent le pourcentage de peptides identifiés avec 3/3 fragments.

La Figure 14.B démontre la linéarité de la réponse MS obtenue avec le BioAccord. On remarque que la médiane des ratios expérimental sur les ratios théoriques est proche de la courbe linéaire attendue et ce jusqu'à 20 ng/mg de mAb. Cette expérience démontre que le BioAccord et la méthode développée permettent la quantification des HCPs jusqu'à 20 ng/mg de mAb avec une bonne précision. Il est très probable que la limite de quantification soit inférieure, mais une expérience supplémentaire avec des protéines standard dopées jusqu'à 1 ng/mg de mAb serait nécessaire pour appuyer cette affirmation. De plus, l'étude des fragments a montré qu'à 20 ng/mg, aucun fragment n'a été identifié pour la majorité des peptides (Figure 14.C). Le nombre limité de fragments par protéine utilisé et leur saisie manuelle fastidieuse dans le logiciel UNIFI rendent vraisemblablement trop stricte l'utilisation des fragments comme critère de validation. Par conséquent, lors de la seconde expérience, les peptides ont été validés sur la base de leurs profils isotopiques et de leurs erreurs de masse.

La deuxième expérience a été conçue pour quantifier 115 peptides provenant de 30 HCPs ciblées au sein du NIST mAb. Ces impuretés ont été sélectionnées à partir d'une approche globale de suivi des HCPs précédemment réalisées sur notre Q Exactive HF-X opéré en DIA (section 1 du chapitre 3). Cinq répliquats d'injection ont été effectués en utilisant le BioAccord. Ainsi, 50, 75 et 100 µg de NIST mAb ont été injectés avec les mêmes quantités de protéines standard ajoutées aux échantillons, à savoir 2 pmol d'ADH, 1 pmol de PYGM, 600 fmol de BSA et 400 fmol de ENL. Ensuite, une librairie de peptides a été créée sur UNIFI comprenant les 115 peptides des 30 HCPs et 16 peptides provenant des quatre protéines standards. La création d'une telle librairie nécessite la saisie manuelle de la séquence et du temps de rétention de chaque peptide ciblé. Comme précédemment, la librairie créée a été utilisée pour cibler les peptides HCPs en utilisant la méthode *Accurate Mass Screening on MSe data* avec une tolérance de masse sur les ions précurseurs de 15 ppm. Pour cette expérience, aucune restriction concernant le temps de rétention attendu n'a été fixée car les peptides à cibler ont été sélectionnés sur une plateforme MS utilisée à un débit différent, ce qui entraîne un décalage des temps de rétention. Les signaux identifiés et extraits ont été validés manuellement sur la base de leur profil isotopique et de la forme du pic chromatographique. Ensuite, nous n'avons gardé que les peptides qui ne présentaient pas de valeurs manquantes dans les cinq répliquats par quantité injectée et qui présentaient un CV<20% calculé sur les intensités des peptides parmi les cinq répliquats. Enfin, les 2 pmol d'ADH ont été utilisés pour estimer la quantité des HCPs ciblées via une stratégie de quantification Top3 (Figure 15).

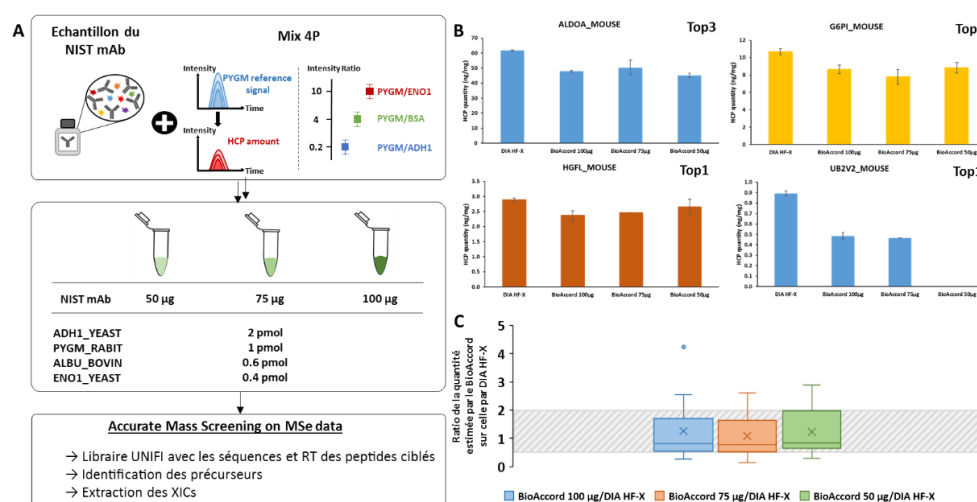


Figure 15: Quantification des HCPs ciblées du NIST mAb sur le système LC-MS BioAccord. Schéma analytique suivi afin de quantifier, en utilisant la stratégie Top3, les HCPs restant dans les échantillons de NIST mAb injectés à différentes quantités : 50, 75 et 100 µg sur le BioAccord (A). Cinq répliquats d'injections ont été réalisés. Représentation des quantités estimées de quatre HCPs quantifiées dans les échantillons de NIST mAb injectés à 100, 75 et 500 µg, respectivement, comparées aux quantités estimées en utilisant notre méthode DIA sur le Q Exactive HF-X (B). Les hauteurs des barres représentent les moyennes des quantités de HCPs dans les cinq répliquats d'injection en utilisant le BioAccord et dans les triplicats d'injection en utilisant le Q Exactive HF-X. Les barres d'erreur représentent l'écart type sur les répliquats. Les annotations Top1 ou Top3 sur chaque graphique représentent le nombre de peptides utilisés pour la quantification dans les données BioAccord. Distribution des ratios entre la quantité de HCPs estimée par le BioAccord et celle obtenue par l'approche DIA HF-X pour les trois quantités injectées d'échantillons de NIST mAb (C).

A première vue, on remarque qu'avec trois quantités différentes de matériel injecté, les quantités estimées de HCPs sont similaires (Figure 15.B). Pour la plupart des HCPs, la barre d'erreur démontre une quantification hautement reproductible en utilisant la stratégie de traitement des données ciblées sur le BioAccord. En outre, les quantités estimées sont du même ordre de grandeur que celles estimées sur le Q Exactive HF-X. Nous avons observé que 70% des 40 valeurs comparées pour les trois quantités de matériel injecté sont cohérentes dans un facteur 2 (Figure 15.C). Ces résultats mettent en évidence les performances prometteuses du BioAccord tout en montrant certaines limites en ce qui concerne la quantité de HCPs présente et la quantité de matériel injecté. En effet, en dessous de 4 ng/mg de mAb, aucune HCP n'a été quantifiée avec plus d'un peptide. De plus, les protéines EF1A1_MOUSE et UB2V2_MOUSE n'ont pas été quantifiées dans les injections de 50 µg de matériel de départ du NIST mAb. Ainsi, la quantité injectée doit être finement optimisée afin d'avoir les HCPs en quantité suffisante pour pouvoir les quantifier.

En conclusion, le BioAccord peut être utilisé en analyse de routine pour la quantification ciblée des HCPs d'intérêts. Cependant, une expérience de suivi global des impuretés doit être réalisée en amont sur un instrument dédié à l'analyse protéomique de type "shotgun", de préférence sur un système chromatographique similaire, afin de sélectionner les peptides protéotypiques des HCPs ciblées. Une fois ces peptides sélectionnés, ils doivent être idéalement synthétisés avec un marquage isotopique afin d'effectuer une quantification précise/absolue par dilution isotopique. Dans ces conditions, le BioAccord a un réel potentiel pour être appliqué à la quantification précise et robuste des HCPs ciblées.

Chapitre 5

Conclusion générale

Ce travail de thèse m'a permis d'acquérir une expertise dans le domaine de la protéomique basée sur la spectrométrie de masse.

La première partie de ma thèse s'articule autour des développements méthodologiques pour l'identification et la quantification des protéines de la cellule hôte par spectrométrie de masse en support de la chaîne de production. Tout d'abord, j'ai réalisé l'implémentation du HCP Profiler, un standard de quantification innovant, au protocole en place afin de standardiser l'approche de quantification globale Top3. Aussi, j'ai évalué plusieurs méthodes d'extraction des données DIA grâce à des échantillons provenant de différentes étapes de la chaîne de production. J'ai également observé l'influence de plusieurs paramètres en amont de cette étape, à savoir la manière de générer la librairie spectrale et des paramètres d'acquisition de la méthode DIA-MS sur des instruments de type Q Exactive HF-X (Q-Orbitrap) et TimsTOF PRO (TIMS-Q-TOF). Finalement, j'ai pu évaluer ces différentes approches afin de démontrer que notre méthode combinant le HCP Profiler avec une acquisition DIA et une extraction basée sur une librairie spectrale réalisées sur un instrument de type Q-Orbitrap peut-être implémenté dans un environnement biopharmaceutique.

La deuxième partie a été consacrée à l'optimisation de la méthode de quantification globale pour le suivi des impuretés protéiques présentes dans les substances médicamenteuses. Pour ces échantillons purifiés, la détection des HCPs représente un véritable défi analytique en raison de la large gamme dynamique entre la protéine thérapeutique surabondante et les impuretés protéiques. Ainsi, nous avons optimisé un schéma analytique dédié comprenant un protocole de préparation des échantillons, l'utilisation de la DIA avec une approche centrée sur les peptides, ainsi que l'implémentation de filtres spécifiques pour le traitement des données, afin d'obtenir des résultats robustes. Grâce à la méthode développée, il a été possible d'identifier et de quantifier plusieurs HCPs dans diverses substances médicamenteuses finales et de soutenir l'investigation d'un changement de procédé de production. La spectrométrie de masse représente donc une méthode orthogonale efficace pour le contrôle qualité de ces produits sensibles.

La troisième partie était consacrée au développement d'une stratégie analytique ciblée sur une plateforme MS dédiée à l'analyse de routine. Lors de la mise en œuvre d'une méthode MS dans un environnement réglementé, il est nécessaire de disposer d'un système LC-MS robuste et d'un logiciel répondant aux attentes en matière d'intégrité des données. C'est le cas du système LC-MS BioAccord piloté par le logiciel UNIFI. Nous avons évalué la performance d'une approche ciblée par le traitement des données pour l'identification et la quantification d'HCPs présélectionnées. Ainsi, le système BioAccord a démontré sa robustesse et sa capacité à quantifier les HCPs de faible niveau présélectionnées à partir d'une étude de suivi global des HCPs réalisée sur un instrument dédié, tel que les instruments HR/AM. Ainsi, le système LC-MS BioAccord, simple d'utilisation, peut être envisagé pour l'analyse de routine d'un ensemble de HCPs problématiques.

General introduction

General introduction

During the last two decades, the monoclonal antibody (mAb) market has remarkably grown up with a plethora of approved antibodies by the FDA and EMA with a current sales market of over \$100 billion^{2, 32}. The high specificity of mAbs to target diverse molecules or antigens and their various modes of action allow them to be used as pharmaceuticals for a wide range of applications³³⁻³⁴. The high demands of mAbs require the production of well-characterized drug products, both in terms of the mAb sequence and structure itself and its impurities, namely Host Cell Proteins (HCPs), remaining from the production process. These impurities are included in the Critical Quality Attributes (CQAs) risk assessment as they can affect the product efficacy due to eventual protease activities and the patient's safety by inducing immunogenic reactions⁵⁻⁶. Guidelines state classically that the acceptable HCPs amount in the final drug product should be below 100 ng/mg mAb^{7, 35}. Ultimately, the level of impurities should be as low as possible as issues related to HCPs arise from specific proteins rather than from overall impurities amounts³⁶⁻³⁸.

Enzyme-linked immunosorbent assays (ELISA) are commonly used for this purpose as they provide the sensitivity and throughput requested^{8, 39}. However, ELISA has several limitations as it provides a global amount as an output without individual identification of the HCPs present and its coverage is incomplete^{8, 40}. Since immunogenic risk or mAb degradation are related to specific HCPs and not necessarily to their amount, these drawbacks raise an urgent need for alternative methods.

In this context, Mass Spectrometry (MS) became the most promising alternative to monitor HCPs allowing risk assessment with individual HCP identification and unbiased quantification. In recent years, Liquid Chromatography – tandem MS (LC-MS/MS) based studies have been conducted. Data-dependent acquisition (DDA) strategies were successfully applied^{9, 27, 41-42} but the stochasticity, the presence of missing values and a discrimination towards the quantification of most abundant proteins are significant issues when the HCPs impurities are present at trace levels compared to the biotherapeutic. Targeted strategies (Selected Reaction Monitoring (SRM) or Paralleled Reaction Monitoring (PRM)) have been developed enabling robust and accurate quantification of targeted HCPs down to the sub ng/mg mAb level^{12, 37, 43}. Nevertheless, the development of a targeted quantification assay is time consuming and limited to the selection of about hundred candidates.

Over the past decade, advances in mass spectrometry have highlighted the potential of data-independent acquisition (DIA) on high-resolution/accurate mass (HR/AM) instruments. DIA is based on the co-isolation and co-fragmentation of all ions contained in predefined m/z windows of variable widths to cover the entire mass range. The acquisition of MS2 signals from all detectable species allows recording complete digital proteome maps while reaching the sensitivity, quantification accuracy and robustness of targeted methods¹⁴⁻¹⁵. These advantages make DIA approaches attractive in the context of HCP monitoring. The bottleneck of DIA analysis today resides in the extraction of reliable quantitative signals. Each MS2 scan contains the fragment information of all co-isolated precursors, rendering peptide identification difficult. Initially, the use of a specific spectral library generated from DDA runs⁴⁴ to extract quantitative information from DIA data was shown to be effective. However, the generation of this spectral library requires time and ideally the implementation of fractionation strategies of the studied proteome. The recent development of algorithms that do not require a spectral library will certainly further increase the interest and applicability of DIA strategies, in particular for the monitoring of HCPs⁴⁵⁻⁴⁷.

Unfortunately, despite the advances in data acquisition and extraction, the analytical challenge remains in the dynamic range mAb/HCP. The ubiquity of the antibody may interfere with the reliable extraction of HCP peptides and lead to biases in the MS-based quantification of these impurities. Few studies have attempted to decrease this dynamic range by using multi-dimensional chromatography^{11, 27-28, 48-49} or mAb depletion^{29, 50-51}.

My PhD work was part of this context. It focused on the development and evaluation of analytical and bioinformatics strategies in mass spectrometry-based proteomics for the accurate quantification of trace-level proteins. The main objective was to develop a method allowing the identification and accurate quantification of a large number of proteins while being straightforward and robust enough to be implemented in a biopharmaceutical environment.

This manuscript is organized into five parts, which are briefly presented below:

- The **bibliographic introduction** is a state of the art of quantitative proteomics and an introduction to host cell proteins. It includes a description of the different steps of the analytical scheme of the "bottom-up" proteomic strategy, from the sample preparation to the data processing through the different existing acquisition modes. The different quantification strategies will also be detailed in this section:
 - Global quantification approaches: relative quantification with and without labeling as well as absolute quantification without labeling.
 - Targeted quantification approaches such as SRM ("Selected Reaction Monitoring") or PRM ("Parallel Reaction Monitoring").
 - The more recent DIA (Data Independent Acquisition) strategy.Then, the importance of controlling HCPs in therapeutic proteins as well as the different methods to monitor them are detailed.
- **Part I** describes the development of a DIA method enabling the monitoring of HCPs throughout the mAb manufacturing process:
 - In **Chapter 1**, the performance of the HCP Profiler solution, an innovative quantification standard, is benchmarked against more conventional standard protein spikes using DDA mode on a Q-Orbitrap instrument for the characterization of HCPs.
 - In **Chapter 2**, the HCP Profiler standard is combined to a DIA approach developed on the same Q-Orbitrap instrument. Two DIA data extraction strategies, relying or not on a spectral library, are evaluated and further benchmarked against a classical MS1-XIC DDA method. The objective of this study is to evaluate the potential of MS2-based DIA quantification to support the monitoring of HCPs during the mAb manufacturing process.
 - In **Chapter 3**, the use of a hybrid spectral library, which combines DDA and DIA analyses, is evaluated against the use of a classical DDA library or the spectral library-free approach. The purpose of this chapter is to optimize the DIA method to make it applicable to different antibodies. Furthermore, the MS2-based DIA strategy is applied to a concrete case aiming at answering the question of whether the capacity of the culture bioreactor impacts the HCPs profile.
 - In **Chapter 4**, the performance of the TimsTOF PRO instrument is evaluated to investigate the benefits of ion mobility as an additional separation dimension for bottom-up analysis. First, the ddaPASEF and diaPASEF approaches are developed.

Then, the HCP coverage of the TimsTOF PRO is benchmarked against the Q-Orbitrap instrument on the same sample set.

- **Part II** focuses on the development of a specific analytical workflow dedicated to final drug products' analysis, which represents a real analytical challenge:
 - In **chapter 1**, various optimizations are described to improve analytical performances for HCPs characterization in final drug products. It includes optimizations at the sample preparation level with an improved native digestion protocol and the implementation of a focused ultrasonication step, at the data acquisition level with a dedicated DIA method and at the data interpretation level with the implementation of a stringent data validation pipeline.
 - In **Chapter 2**, we apply our optimized workflow to support the investigation of a mAb manufacturing process change.
- **Part III** presents the development of a method for HCP profiling on a MS platform aimed at enabling routine analysis. The performance of the BioAccord LC-MS system is evaluated for the targeted identification and quantification of selected HCPs using the NIST reference mAb and standard protein spikes. The objective of this chapter is to evaluate the applicability of this instrument platform for the targeted quantification of problematic HCPs in a routine manner and in a quality-compliant pharmaceutical environment.
- The last part corresponds to the **experimental section**, in which experimental details on the different works discussed in this manuscript are described.

Bibliographic introduction

Chapter 1

Bottom-up proteomic strategies

The term *proteome* was originally coined in 1994 by Mark Wilkins, a PhD student at Macquaire University in Sydney⁵². It combines the terms *protein* and *genome*. While the genome represents the set of genes that characterize an organism, the proteome designates all proteins contained in a living entity at a given time and under given conditions⁵³. Unlike the genome, the proteome implies a notion of dynamism⁵⁴.

The term *proteomics* was introduced a few years later by Peter James⁵⁵. By analogy to genomics, the study of genomes, proteomics was defined as the characterization of all proteins in a cell, tissue or organism⁵⁴. However, this definition has evolved to include the notion of dynamism inherent to the proteome. Now, proteomics represents the qualitative, quantitative and functional study of proteins contained in a cell, tissue or organism at a specific time and under specific conditions⁵⁶.

Mass spectrometry has become a tool of choice for proteomic analysis, thanks to significant developments⁵⁶⁻⁵⁹:

- The discovery of two soft ionization sources: MALDI (Matrix-Assisted Laser Desorption Ionization)⁶⁰ and ESI (Electrospray ionization)⁶¹ by Koichi Tanaka (Japan) and John B. Fenn (USA), respectively. These innovations represent a major breakthrough for the MS analysis of biological macromolecules such as proteins. They were awarded the Nobel Prize in Chemistry in 2002;
- The developments of separative methods for peptides and proteins prior to MS analysis;
- The constant instrumental progress of mass spectrometers allowing to improve sensitivity, dynamic range, resolution, acquisition speed and mass accuracy performances;
- The development of new acquisition modes;
- The improvement of high quality, i.e. curated and well-annotated protein databases;
- The development and constant improvement of bioinformatics tools for efficient and automated data analysis and interpretation.

1. Mass Spectrometry-based proteomic approaches

Mass spectrometry-based proteomic analysis can be divided into three approaches: Bottom-up, Middle-down and Top-down (Figure 16).

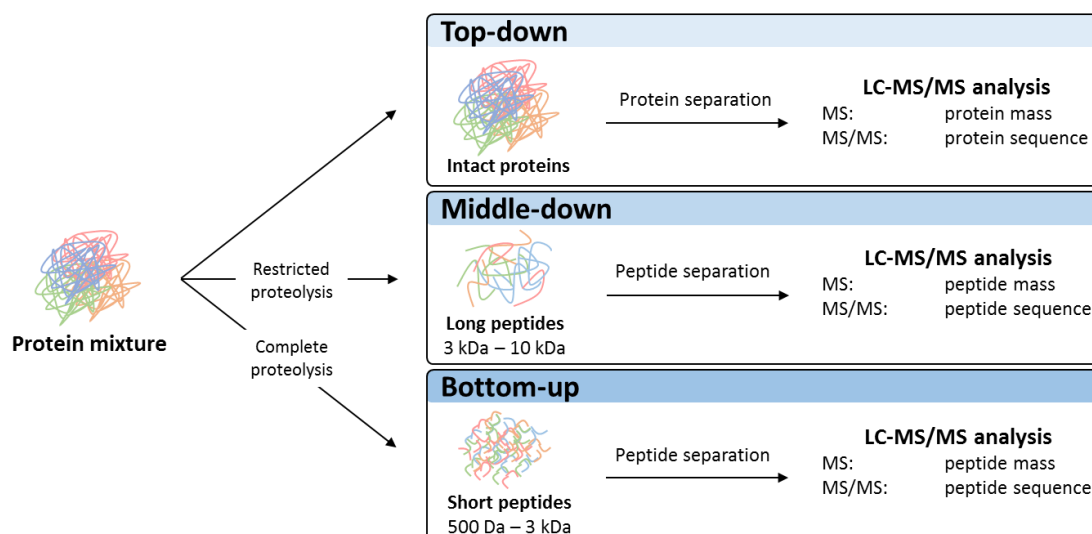


Figure 16: Schematic representation of the three MS-based proteomic approaches: Bottom-up, Middle-down and Top-down.

The bottom-up approach consists in characterizing proteins through their peptides after enzymatic digestion, mainly with trypsin. These peptides, generally lower than 3 kDa, are then separated by liquid chromatography and analyzed in MS (peptide mass) and MS/MS (peptide sequence). The comparison of the measured mass list to a theoretical one, obtained by *in silico* digestion of protein sequences from a given database, allows the identification of peptides and proteins by inference. This protein inference is achieved by associating multiple peptide identifications to a protein. The inference step is challenging since a peptide can be unique to a protein or shared between several, the latter can be grouped. Therefore, the result of the bottom-up approach corresponds to the smallest list of proteins satisfying the parsimony principle⁶². This approach has become the method of choice for the identification and quantification of thousands of proteins. On recent LC-MS/MS platforms, results can be obtained in a few hours⁶³.

The Top-down approach enables the characterization of intact proteins by MS (protein mass) and MS/MS (protein sequence) without enzymatic digestion. The objective of the approach is to provide high coverage and a complete characterization of a targeted protein. It has some advantages over the bottom-up approach, especially for proteoforms differentiation and post-translational modifications (PTM) analysis⁶⁴. Recently, the Top-down approach succeeded the identifications of more than a thousand proteins and thousands of proteoforms using multidimensional separation⁶⁵⁻⁶⁶. However, this approach has some limitations⁶⁷⁻⁶⁹. First, the difficulties related to the solubility, separation, ionisation and fragmentation of proteins limit the sensitivity of the approach. Instruments with high resolution, mass accuracy and scan speed are required to finely resolve the isotopic envelopes of the multiple charge states proteins analyzed. This also implies that the approach is better suited for highly purified samples. Finally, bioinformatics tools still need to be improved.

The middle-down approach is a hybrid of bottom-up and top-down approaches^{56, 70-71}. It allows the characterization of proteins from large peptides (3 to 10 kDa), obtained after a restricted enzymatic digestion. Similar to the bottom-up approach, the peptides are fractionated and analyzed in MS (peptide mass) and MS/MS (peptide sequence). Since the number of peptides obtained with the restricted proteolysis is lower than with the bottom-up approach, it limits the proportion of shared peptides and facilitates protein inference. Furthermore, these large peptides provide access to the same information as the top-down approach while overcoming the challenge of analyzing intact proteins. However, restricted proteolysis requires special attention. It must be finely optimized in terms of enzyme used (specific or non-specific), protein/enzyme ratio and proteolysis conditions. Moreover, as for the top-down approach, data analysis represents a major challenge for the middle-down strategy.

The top-down and middle-down approaches still require optimization and development to make them compatible with the analysis of complex biological samples. Consequently, the bottom-up strategy remains the most straightforward, large-scale and high-throughput proteomic method. The work done in this manuscript is only based on this approach, which is described in the next section.

2. Analytical workflow of the bottom-up strategy

The bottom-up proteomic workflow consists of three main parts: sample preparation, liquid chromatography coupled to tandem mass spectrometry and data treatment (Figure 17).

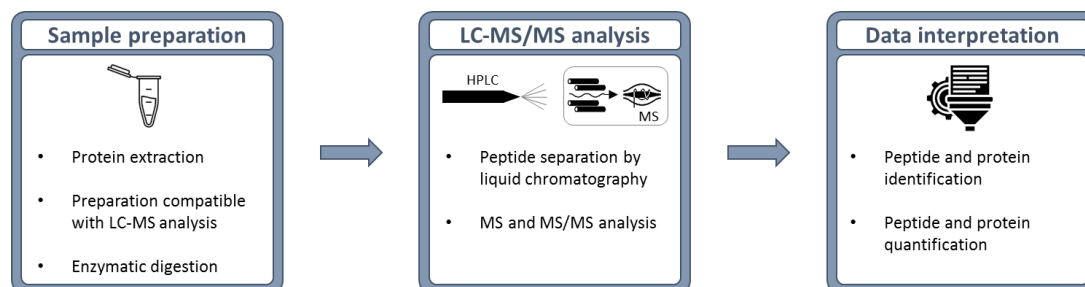


Figure 17: Bottom-up proteomic workflow description.

A. Sample preparation

Sample preparation is a crucial step in the bottom-up approach. The quality and repeatability of the preparation will have a major impact on the MS results. Thus, the preparation sub-steps should be carefully chosen according to the type of samples⁷²⁻⁷³.

i. Protein extraction

The extraction of proteins from the original biological sample (fluids, tissues ...) represents the first step of the bottom-up approach. It aims at solubilizing the maximum of proteins without degrading or modifying them. The extraction will influence the accessibility of the proteins to proteases and thus affect the quality of the proteomic analysis. Therefore, this step must be optimized according to the quantity and type of sample, the protein type and the analytical workflow used afterwards⁷⁴⁻⁷⁶.

Lysis and extraction of proteins can be performed^{75, 77}:

- Mechanically by grinding or ultrasonication for instance.
- Chemically by using:
 - Detergents (ionic, non-ionic, zwitterionic, salts of bile acids or detergent directly compatible with protein digestion and LC-MS/MS analysis) to enhance protein extraction and solubilisation *via* micelle formation.
 - Chaotropic agents (e.g. urea) to denature and unfold proteins by stabilizing them in their unfolded structure.
 - Organic solvents, such as acetonitrile, to facilitate the denaturation of proteins by altering their conformation.

Chemical lysis and extraction buffer combined with a mechanical stimulus is the most common strategy⁷⁵.

After extraction, additional steps may be necessary to remove detergents and contaminants such as lipids or nucleic acids that negatively affect the LC-MS/MS analysis: interference during liquid chromatography, signal suppression or noisy spectra. This can be achieved *via* protein precipitation with cold acetone, trichloroacetic acid, chloroform-methanol mix or using dialysis or ultrafiltration steps. However, these extra steps are tricky and can lead to material loss⁷⁵.

ii. Protein preparation for LC-MS/MS analysis

Samples subjected to proteomics analysis often represent a real analytical challenge. Indeed, those samples are composed of thousands of proteins that are present under several forms. The diverse forms of a protein, called proteoforms, result from splicing variants, genetic sequence variants or PTMs (oxidation, phosphorylation, glycosylation ...) increases the complexity of the proteome. For instance, around 20 000 protein-coding genes were reported by the human genome project⁷⁸. If we consider a gene for a protein, then we go from 20 000 proteins to about 70 000 proteoforms if we only take into account splice variants, and even hundreds of thousands with the eventual PTMs. Furthermore, the dynamic range of protein abundances accentuates the complexity of these samples. Indeed, it can reach up to 10 orders of magnitude^{56, 79-81}, while mass spectrometers cover up to 5 orders of magnitude^{72, 81-82}. Various solutions are possible to reduce the complexity of the samples. It is possible to deplete abundant proteins in order to overcome the dynamic range issue, but with the risk of losing biological material⁸³. It is also possible to fractionate proteins based on their physico-chemical properties: molecular weight (sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE)), protein size (size exclusion chromatography (SEC)), pI (ion exchange chromatography, isoelectric focusing), polarity (reverse phase chromatography) or by combining them^{72, 75, 84}.

In general, two main protein preparations for LC-MS/MS analysis have been mostly employed: in-solution or in-gel preparations^{72-73, 75}. In this work, both preparations were used and are detailed below:

- **In-solution preparations** consist of denaturing proteins, reducing disulfide bridges and alkylating cysteine residues before protein digestion⁷⁵. For this purpose, buffers compatible with enzymatic digestion are required. To facilitate the use of this approach, a series of innovative protocols was recently developed. Among them, FASP (*Filter-Aided Sample*

preparation) allows the use of detergents to extract proteins *via* filters or membranes that remove these unwanted compounds before proteolysis, but with a risk of material loss. The single-pot, solid-phase-enhanced sample preparation (SP3) protocol, another alternative, uses paramagnetic beads to immobilize proteins or peptides. Detergents can also be used with this approach⁸⁵. Finally, commercial kits are available as ready to use kits: *Ist Preparation kit* (Preomics)⁸⁶, *S-Trap* (ProtiFi)⁸⁷, *Pierce™ Mass Spec Sample Prep Kit* (Thermo Fischer) and *Sample preparation kit* (Biognosys).

- **In-gel preparations** aim to eliminate detergents, salts and contaminants. In particular, they allow the removal of SDS when used for protein extraction. In this work, two gel protocols were used:
 - 1D SDS-PAGE: it consists in separating proteins by migration through a polyacrylamide gel while subjecting them to an electric field. Electrophoresis separates charged analytes according to their charge-to-size ratio under the application of an electric field⁸⁸⁻⁸⁹. SDS, contained in the loading buffer, gives the proteins a linear conformation and a uniform negative charge allowing them to be separated on the basis of their molecular weight. The polyacrylamide gel used is composed of two distinct parts. First, in the stacking gel, composed of a low percentage of acrylamide/bis-acrylamide (4 to 5%), proteins will be concentrated in a single band before entering the separation gel. This second part, containing a higher percentage of acrylamide/bis-acrylamide (8 to 15%), allows the proteins to be separated thanks to its tighter network.
 - Stacking gel: It is based on the same principle as 1D SDS-PAGE. By contrast, migration is stopped once the proteins are concentrated in a single band, just before they enter the separation gel. This protocol allows the use of SDS for protein solubilisation without fractionation of the sample. It has demonstrated to be well-suited for accurate and precise quantification.

In addition to these protocols, tube-gel⁹⁰⁻⁹² and 2D SDS-PAGE⁹³ approaches can be mentioned. The former consists of the direct incorporation of proteins into the polyacrylamide gel prior polymerization, thus avoiding electrophoretic migration. The tube-gel combines the advantages of stacking gel with a short sample preparation time. The latter involves a first dimension of protein separation according to their isoelectric points and a second *via* a 1D SDS-PAGE gel.

Finally, the proteins are fixed in the gel, revealed by Coomassie Blue staining⁹⁴, reduced and alkylated prior in-gel enzymatic digestion. However, in-gel approaches are less used and are progressively replaced by more automated protocols that are well suited to high-throughput analysis.

iii. Enzymatic digestion

Enzymatic digestion for bottom up analysis is often performed using trypsin. This is mainly due to its low cost, robustness and specific cleavage at the C-terminus of lysines and arginines, except when followed by proline⁹⁵⁻⁹⁶. The presence of a positive charge at the C-terminal basic end of the peptide facilitates ionization and fragmentation of the peptide, which makes tryptic peptides ideally suited for LC-MS/MS analysis. Furthermore, the natural abundance of these two residues in proteins leads to the

generation of peptides between 500 and 3000 Da, allowing for proper coverage of protein sequences. Besides, this enzyme is also able to penetrate the polyacrylamide gels enabling in-gel digestion.

The combined use of different proteases in the same experiment has been shown to provide better sequence coverage. In the particular case of the bottom-up proteomic approach, the combination with Lys-C seems particularly effective⁹⁷⁻⁹⁸. Indeed, this enzyme is resistant to denaturing conditions and complements trypsin that is less efficient towards lysine than arginine residues.

Finally, another limitation linked to the use of "classical" trypsin is the digestion step duration, i.e. between 12 and 18h. Commercial solutions based on modified enzymes, such as the *SMART Digest Kit* from Thermo Fisher Scientific or the *Rapid Digestion Kit* from Promega⁹⁹ are available and drastically reduce the time required for this step. Nevertheless, due to the diversity of digestion methods, there is no universal protocol. It depends on the sample type, the properties of the proteins to be analysed and the purpose of the study.

B. Liquid chromatography coupled to tandem mass spectrometry analysis

i. Peptide separation by reversed phased-liquid chromatography

Essential for the bottom-up approach, enzymatic digestion increases the complexity of the sample^{56, 59}. For this reason, liquid chromatography is commonly used to reduce it. As a result, it improves ionization efficiency by limiting competition between analytes, but also increases sensitivity, selectivity and proteome coverage¹⁰⁰⁻¹⁰¹. In this manuscript, reverse phase liquid chromatography was used to separate peptides according to their hydrophobicity by progressively decreasing the polarity of the mobile phase (mixture of water and acetonitrile). Two different chromatographic systems were used and are detailed in Table 2.

LC system	NanoAcquity UPLC at nanoflow	NanoElute at nanoflow	Acquity UPLC at microflow
System vendor	Waters	Bruker	Waters
Column vendor	Waters	IonOpticks	Waters
Stationary phase	C18	C18	C18
Column length (mm)	250	250	150
Internal diameter	75 µm	75 µm	2.1 mm
Particle size (µm)	1.7	1.6	1.7
Pore size (Å)	130	120	130
Flow rate	400 nL/min	400 nL/min	200 µL/min

Table 2: Description of the LC system used in my PhD work.

The separation efficiency of chromatographic systems depends on the properties of the column, i.e. its length, internal diameter, pore and particle sizes^{59, 102}. Nanofluidic systems operating at high pressure (>500 bar), provide improved sensitivity, resolution and peak capacity, while requiring small amounts of biological material. These systems are particularly suitable for proteomic analysis where samples are often available in small quantities. However, they are subject to recurrent problems (ionization spray instability, difficult-to-detect leaks or dead volumes) and require real expertise.

On the other hand, microflow systems display a reduced sensitivity, which can be compensated by increasing the amount of material injected¹⁰³. They also show a higher robustness than nanoflow systems^{54, 103}.

ii. Tandem mass spectrometry

Peptides eluted from reverse phase liquid chromatography are directly ionized, via electrospray ionization (ESI), before entering the mass spectrometer. This instrument first measures the mass-to-charge ratio (m/z) and intensity of each ion to generate an MS (or MS1) spectrum. These ions are then isolated and fragmented. The m/z ratios and intensities of the fragments are measured, generating an MS/MS (or MS2) spectrum¹⁰⁴. The successive generation of MS1 and MS2 spectra, so-called tandem mass spectrometry, is accomplished on hybrid instruments coupling different mass analyzers (quadrupole (Q), time of flight (TOF), Orbitrap, ion trap and ion cyclotron resonance (FT-ICR)).

In my work presented in this manuscript, three different mass spectrometers were used: Q-Orbitrap, Q-TOF and TOF instrument types. Their specifications are detailed in Table 3.

MS system	Q Exactive HF-X	TimsTOF PRO	BioAccord
System vendor	Thermo Fisher Scientific	Bruker Daltonics	Waters
Analyzer	Q-Orbitrap	Q-TOF	TOF
Resolution	From 7500 to 24000 (at 200 m/z)	40 000 at 622 m/z	>10 000 (at 550 m/z)
Mass accuracy	5 ppm	10 ppm	5 ppm
Acquisition speed	Up to 40 Hz	> 100 Hz	Up to 20 Hz
Fragmentation	HCD	CID	In-source
Ion mobility	-	TIMS	-

Table 3: Description of the MS systems used during my PhD work.

a. Data Dependent Acquisition

Data Dependent Acquisition represents the most used in the field of bottom-up proteomics. Throughout the gradient, the mass spectrometer operated in DDA mode will perform the successive acquisition of MS1 and MS2 spectra in a cyclic manner. Over a cycle, an MS1 spectrum will first be acquired, and the N most intense precursor ions of this spectrum will be sequentially and in real time selected in a narrow m/z window for fragmentation before being analyzed to generate N MS2 spectra¹⁰⁵ (Figure 18).

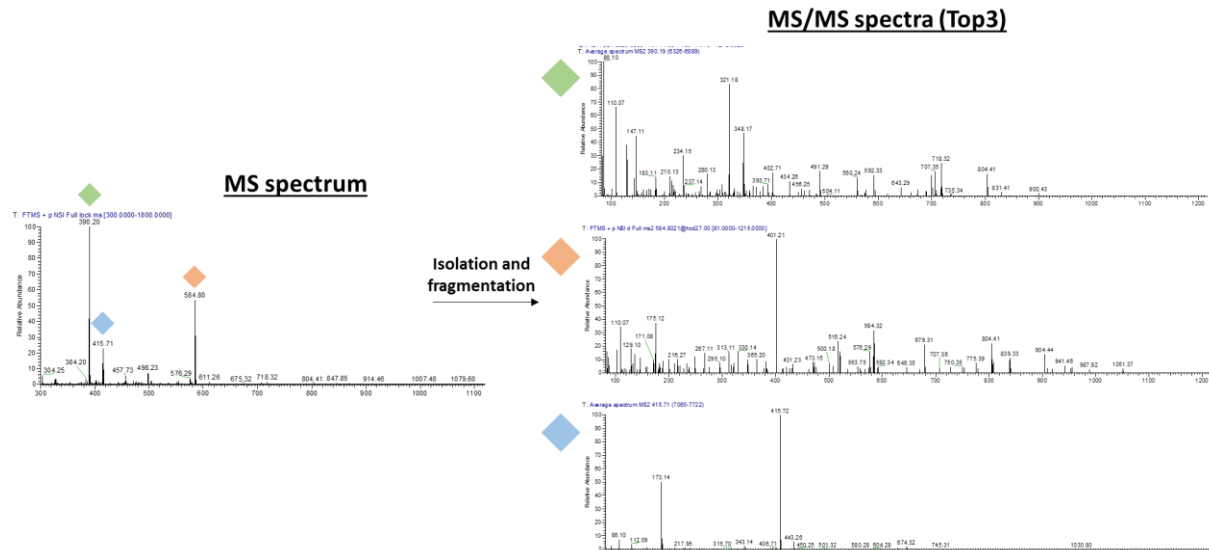


Figure 18: Data Dependent acquisition scheme. In the present example, the three most intense precursor ions on the MS1 spectrum (390.20 m/z in green, 584.80 m/z in orange, and 415.71 m/z in blue) were sequentially isolated and fragmented, generating three MS2 spectra that combine their fragments.

This acquisition mode provides identification of several thousand proteins, thus offering a good coverage of proteomes¹⁰⁶⁻¹⁰⁷. However, the selection principle based on precursor ions intensities on the MS1 spectrum leads to stochasticity and a lack of reproducibility. Instrumental advances in sensitivity and acquisition speed have reduced the undersampling effect¹⁰⁶. Furthermore, technical solutions are available to increase the number of identifications. The discrimination towards the selection of the most intense precursor ions can be overcome by dynamic exclusion which allows to reduce the redundancy of the collected spectra or by the definition of inclusion or exclusion lists¹⁰⁸⁻¹⁰⁹.

b. peptide fragmentation

Several approaches have been implemented on mass spectrometers to fragment peptides: Collision Induced Dissociation¹¹⁰ (CID), Higher energy C-trap Dissociation¹¹¹ (HCD) Electron Transfer Dissociation¹¹² (ETD), Electron Capture Dissociation¹¹³ (ECD) and even a combination of ETD and HCD called ETHCD. Among these fragmentation modes, CID is the most commonly used for bottom-up proteomics analysis⁵⁶. Ions are isolated and accelerated to induce high kinetic energy. Then, collision with neutral molecules (argon, helium or nitrogen) present in the collision cell results in the conversion of kinetic energy into internal energy inducing the rupture of the peptide bond according to the mobile proton model¹¹⁴⁻¹¹⁵. CID fragmentation generates γ and b ions, as defined by Biemann's nomenclature¹¹⁶ (Figure 19). It is well suited for tryptic peptides that have at least two positive charges, one mobile at the N-terminus and one on the lysine or arginine side chain.

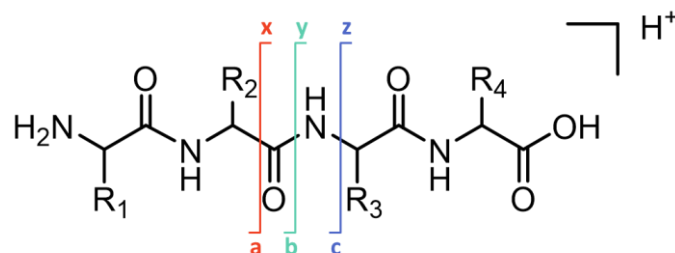


Figure 19: Biemann nomenclature for peptide fragmentation. a-, b- and c-ions carry the positive charge at the N-terminal extremity, while x-, y- and z-ions on the C-terminal extremity. CID and HCD fragmentation favour b- and y-ions, while ETD and ECD c- and z-ions.

Finally, the fragment ions produced by the CID fragmentation are analyzed simultaneously to generate an MS2 spectrum, from which the peptide sequence can be determined (Figure 20).

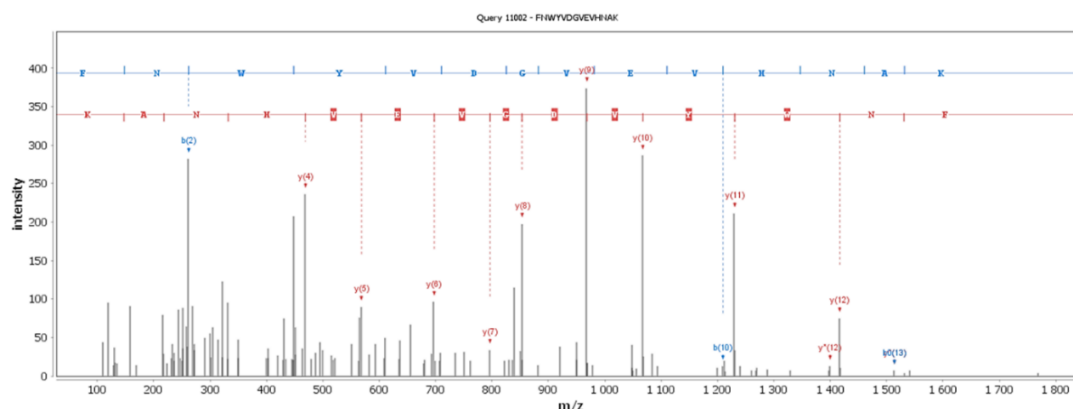


Figure 20: Annotated MS2 spectrum generated using CID fragmentation. The protein sequence is read from left to right for b-ions and right to left for y-ions.

C. Ion mobility spectrometry for bottom-up proteomics

Ion mobility spectrometry (IMS) in combination with mass spectrometry (MS) has grown significantly in the last 20 years¹¹⁷⁻¹¹⁸. IMS allows the separation of ions in a buffer gas under the influence of an electric field based on their mobility. The mobility of the ions depends on their size, shape and charge. Thus, the information obtained by IMS can be considered as complementary to MS. For instance, the IMS-MS combination has been used to separate isomers, filter signal and annotate untargeted features *via* Collision Cross Section (CCS) database matching^{117, 119-121}.

Several ion mobility technologies have been developed with their own applied field and gas dynamics. Therefore, it has become common to convert the mobility into CCS values that are related to the ion shape for a given compound in a specific gas environment. A summary of the various type of IMS devices commercialized is presented in Figure 21.

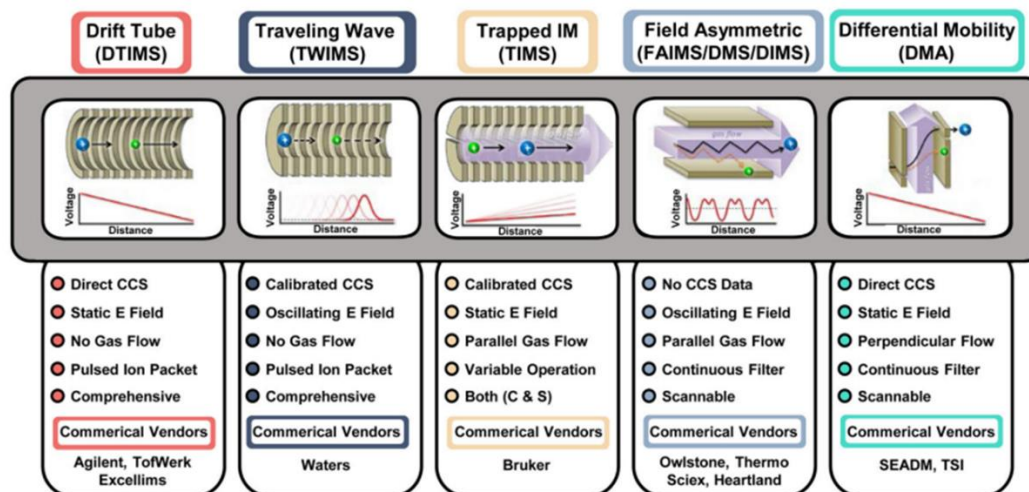


Figure 21: Summary of the various IMS devices with their specificities and vendors (from ¹¹⁸).

Among these technologies, two IMS devices are the most used in combination with MS for bottom-up proteomic analysis: the Field Asymmetric Ion Mobility Spectrometry device (FAIMS) and Trapped Ion Mobility Spectrometry (TIMS). Their implementation acts as an additional separation dimension and greatly impacts data acquisition and processing. Therefore, we will describe the two principles in this section.

i. Field Asymmetric Ion Mobility Spectrometry (FAIMS)

FAIMS technology performs fast and efficient gas phase separation of precursor ions as they leave the electrospray emitter and before they enter the mass spectrometer¹²²⁻¹²⁴. Indeed, the FAIMS source transmits ions between the inner and outer electrodes according to their mobility difference when they are in a high or low electric field. Ions with a large mobility difference between the high and low field migrate toward the electrodes, while ions with no or small mobility difference are transmitted. The ion trajectory can be modified by adding a DC voltage, so-called compensation voltage (CV). Thus, selecting an appropriate DC level will compensate the drift of a specific ion or ion group, allowing them to pass through the device. This CV value, determined empirically, controls which ion population passes through the FAIMS device. Moreover, because the CV can be changed rapidly (~25 milliseconds per transition), it is possible to switch between multiple current values during a MS cycle, and thus independently analyze ions of different charges (Figure 22).

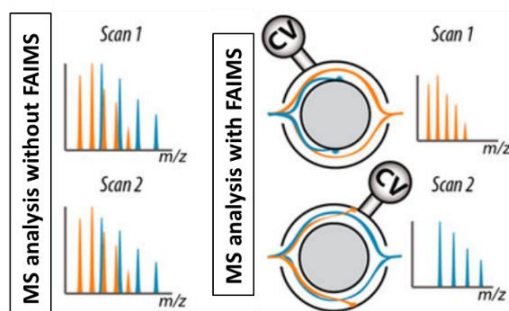


Figure 22: Schematic representation of the FAIMS permutation capacity between several current values during a MS cycle¹²³.

Finally, FAIMS does not use ion accumulation. Thus, its duty cycle is about 100% for the ions passing the filter allowing an improved signal to noise ratio. Due to the application of this waveform electrical field, those devices are not able to provide CCS values.

ii. Trapped Ion Mobility Spectrometry (TIMS)

Trapped Ion Mobility Spectrometry (TIMS) is one of the newest IMS technologies. It was developed by Bruker and equips the TimsTOF instruments. It is necessary to differentiate between the classical TIMS device and the TIMS device included in the TimsTOF instruments, which use a 9.7 cm dual TIMS cell. Here, only the basic principle of TIMS will be described. The operation of the dual TIMS cell and the resulting PASEF acquisition mode, which achieves a duty cycle of about 100%, will be detailed in the results section of this manuscript (Part I, Chapter 4, section 1.A).

A TIMS device is divided into three parts, two ion funnels at the entrance and exit of a TIMS tunnel as shown in Figure 23.A. The funnels accumulate the ions entering and exiting the TIMS tunnel. In this technology, ions are driven into the TIMS cell by a constant buffer gas flow and are retained by the application of a static electric field. Three steps are performed in the TIMS tunnel as described in Figure 23.B. First, the ions from the ion source are accumulated. Then they are trapped at their equilibrium position in the tunnel, as shown in Figure 23.C. Their position in the tunnel depends on their shape for the same charge state. Larger ions will be dragged by the gas flow further into the tunnel and will be closer to its exit. The ions are then released in decreasing order of CCS by a slow decrease of the electric field.

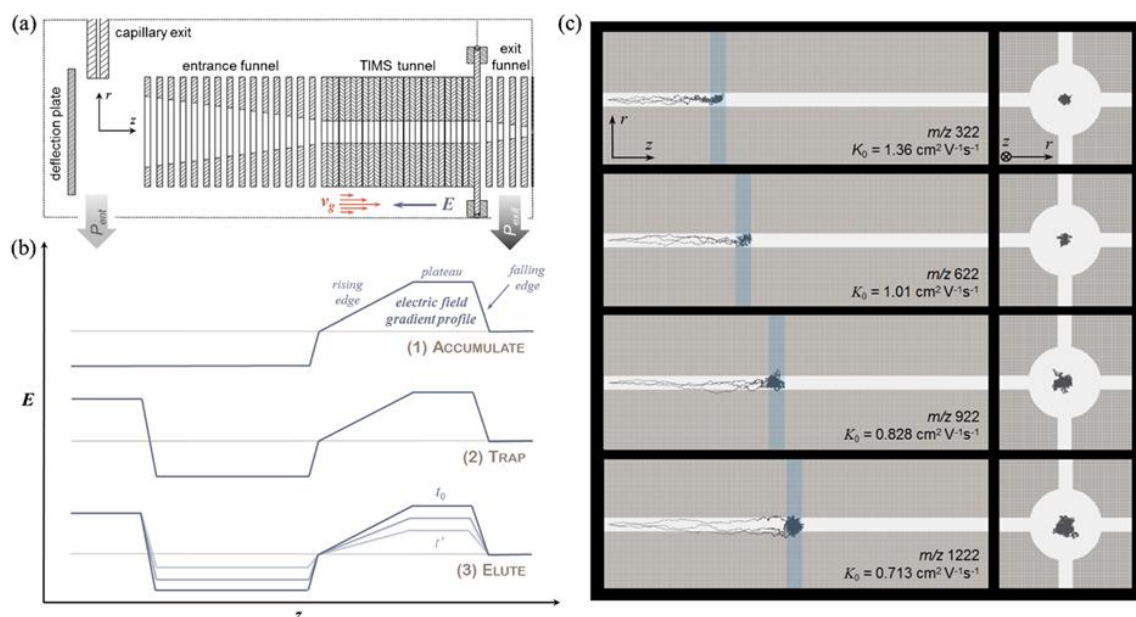


Figure 23: Principle of the TIMS device (from ¹²⁵). Schematic representation of the TIMS component (A). Diagram of the voltage applied during the three steps of one ion packet separation (B). Illustration of the ions position in the TIMS tunnel (C).

The TIMS device can achieve high resolution. The principle of the TIMS also increase the signal-to-noise ratio as the noise is diluted on the complete ion mobility range whereas signal ions are packed at the same position during the trapping step. Another advantage of TIMS is its size between 5 and 10 cm allowing to easily coupling it with other instruments and making upgrading possible.

In this manuscript, the TimsTOF PRO (Bruker Daltonic) was the only MS system with an IMS device.

D. Data processing and interpretation

i. Protein identification strategies

Multiple strategies have been developed to automatically assign peptide sequences to MS2 spectra generated via LC-MS/MS analysis¹²⁶. Among them, spectrum-centric database searching is the most commonly used approach. De novo sequencing, an alternative approach, is based on the peptide sequence extraction directly from the MS/MS spectra. This strategy is of particular interest for the study of non-sequenced organisms. The work done during my PhD involved the study of sequenced organisms. Thus, only the spectrum-centric approach, detailed below, was used.

a. Search engine

The raw data generated by bottom-up analysis are converted into peak lists containing the information on the mass of precursor and fragment ions as well as their intensities. Then, peptides are identified by peptide fragmentation fingerprinting¹²⁷. It consists in comparing experimental mass lists to theoretical ones generated *via in-silico* digestion and fragmentation of a specific protein sequence databank. The identified peptides are then gathered for protein inference. This step can be complex because of peptides shared between proteins or proteins identified with a unique peptide⁶². Finally, the list of identified proteins in the sample is obtained. All these steps are performed automatically by

search engines such as Andromeda¹²⁸, Mascot¹²⁹ (Matrix Science, London, UK), Pulsar (Biognosys, Schlieren, Switzerland), OMSSA¹³⁰, Sequest¹³¹, Byonic™ (Protein Metrics, Cupertino, USA) and X!Tandem¹³². Regardless of the search engine used, they will all require information about the experimental and instrumental conditions under which the data were generated:

- The protease used and the maximum number of missed cleavages allowed.
- The protein sequences database.
- The tolerance on the precursor and fragment ions m/z ratio.
- The charge of the precursor and fragment ions.
- The expected modifications (fixed or variable) of particular amino acids.
- The fragmentation type used.

In the work described in this manuscript, three search engines were used: Mascot, Andromeda and Pulsar.

Mascot is a commercial search engine, which use a probability-based scoring algorithm that is not accessible. For each MS2 spectrum, a score named "ion score" is calculated, equivalent to the probability that the match between the experimental and theoretical mass lists, obtained *in silico*, occurs by chance. The higher the score, the higher the confidence in the peptide identification. This score only considers the quality of the spectrum and is independent of the protein sequence database. In addition, identity and homology thresholds are determined by taking into account the size of the database in order to determine if the identification of a peptide is random. Finally, all identifications associated with a peptide spectrum, named "Peptide Spectrum Matches" (PSM), are classified.

In line with Mascot, a probability-based scoring algorithm also drives **Andromeda**. It can be used on its own or integrated with Maxquant software, which enables the recalibration of spectra. Despite different scoring scales, the comparison of the two search engines showed comparable results¹²⁸.

Finally, **Pulsar** was introduced by Biognosys in 2017 and has been implemented in the Spectronaut™ software. It is dedicated to the generation of spectral libraries, which are then used for the extraction of DIA data via Spectronaut™. It is a commercial solution, as such the algorithm is not accessible.

Despite instrumental and software developments, between 60 and 75% of MS2 spectra remain unidentified¹³³. The reasons include:

- The insufficient quality of some MS2 spectra.
- The co-isolation and co-fragmentation of peptides in the small isolation windows used (1 to 3 m/z). It could have a negative effect when search engine use unassigned peaks on MS2 spectra for score calculation. Since the 2.5 version, Mascot enables the identification of peptides from chimeric spectra.
- The use of incomplete and inadequate protein sequence database.
- The underestimation of the number of modifications, which represents one third of the unassigned spectra¹³⁴⁻¹³⁵.
- Errors in data processing resulting to wrong assignment of the mono-isotopic peak or charge state.

b. Protein sequence database

Peptide assignment via a protein sequence database limits the possible identifications to the content of this repository. Thus, it is crucial to work with the most adequate database for the biological context. Moreover, in order to extract relevant and quality information, it is necessary to use high quality databases obtained with sequence annotation and data filtering. Several databases are available, differing in annotation quality, completeness and degree of redundancy⁶². Among which:

- **NCBI Entrez**¹³⁶ is created by the National Center for Biotechnology Information. This large database combines protein sequences from the Protein Databank (PDB)¹³⁷, Protein Research Foundation (PRF), Protein Information Resource (PIR)¹³⁸, RefSeq¹³⁹ and SwissProt¹⁴⁰, together with protein sequences derived from the translation of nucleotide sequence banks found in EMBL, DDBJ¹⁴¹ and GenBank¹⁴². This library of sequences has varying levels of annotation with high redundancy.
- **RefSeq**¹³⁹ is also produced by NCBI. Nevertheless, the content of this database is verified, annotated and does not contain redundancy. For each protein, the link between the protein, the gene and the transcript is available. The 16 July 2021 (Release 207), it contained 209,035,492 proteins representing 112,462 species.

Nevertheless, errors in the translation of nucleotide sequences into protein sequences can be observed in both databases and further negatively affect the analysis of MS data¹⁴³.

- **UniProtKB** results from the collaboration between the European Bioinformatics Institute (EMBL-EBI), the Swiss Institute of Bioinformatics (SIB) and the PIR. This bank is composed of :
 - *UniProtKB/TrEMBL* that contains automatically translated, annotated and classified protein sequences from GenBank, sequences from the literature and others requiring validation for insertion in UniProtKB/SwissProt.
 - *UniProtKB/SwissProt*, which is the result of extensive efforts to manually annotate and sort sequences from the literature in order to conserve only high quality information. This work, started in 1986, provides many levels of information: function, subcellular localization, interactions, associated pathologies, expression level, structure and PTMs, for instance.

Both databases are continually updated. The 2nd of September 2021, UniProtKB/TrEMBL and UniProtKB/SwissProt contained 219,174,961 and 565,254 entries, respectively.

Databases are updated frequently due to the discovery of new variants, new sequences and manual verification of entries. Thus, the issue of a high proportion of unassigned MS2 spectra can be reduced by using updated databases. Moreover, combining results from genomic, transcriptomic, and proteomic studies enables improved genome annotation and gene prediction algorithms, as well as peptide and protein identifications¹⁴⁴⁻¹⁴⁵. This combination of data, first introduced in 2004, was defined under the term proteogenomics¹⁴⁶. Nevertheless, the task is highly complex¹⁴⁷. During this work, high quality databases were preferentially chosen when available. The aim is to limit the search space and to limit false identifications.

ii. Validation of the identification results

The automatic identification of peptides and proteins can be subject to errors. Search engines have their limits. Since the score associated with each identification is not sufficient to assess the correct assignment and it is not feasible to manually check several thousands of identifications, it is therefore essential to validate the identification results using an additional criterion.

For this purpose, the "target-decoy" strategy is the most commonly used, allowing to distinguish correct identifications from wrong assignments¹⁴⁸. It consists in performing searches using a database including the target protein sequences and decoy protein sequences. The latter corresponds to randomized or inverted protein sequences. These decoy sequences preserve amino acid frequencies as well as protein and peptide sizes and masses identical to those of the corresponding target proteins and peptides. The false positive rate (FDR) is then estimated as the proportion of decoy sequences assigned to MS2 spectra among the total number of assigned sequences. It is calculated as follows¹⁴⁹:

$$FDR = 2 \times \frac{\text{Number of assigned decoy sequences}}{\text{Number of assigned decoy sequences} + \text{Number of assigned target sequences}} \times 100$$

Although some journals have established strict validation guidelines for publication, no consensus has been established⁷².

Chapter 2

Global quantification strategies

Beyond protein identification, quantitative information on the peptides and proteins present in the samples allows a better understanding of the functions and dynamics of a biological system. Indeed, abundance variations may reflect biological processes or perturbations in the system⁵⁶. Even if MS, as such, is not a quantitative approach, strategies have been developed to use these signals for comparative quantitative studies between several samples (Figure 24).

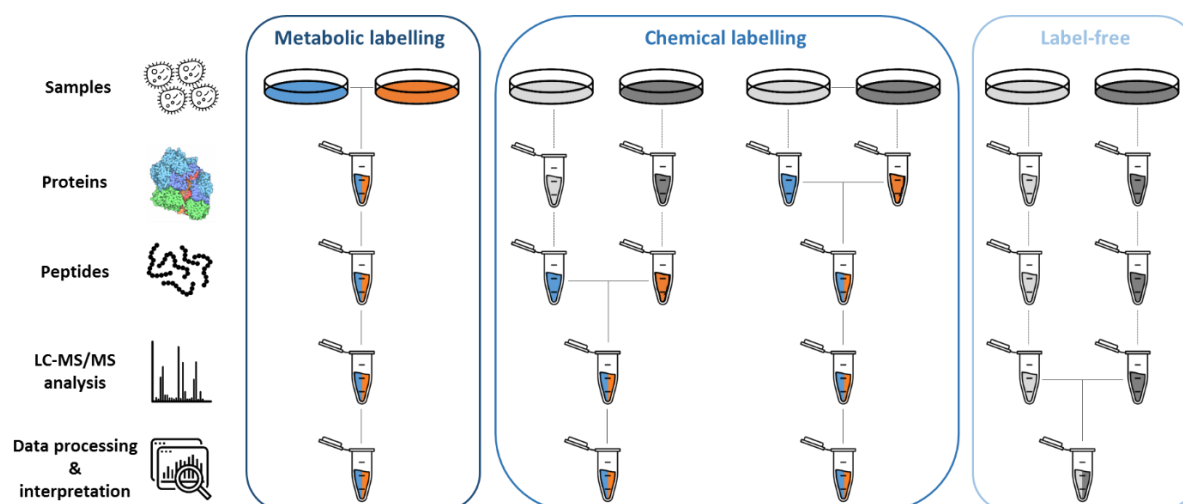


Figure 24: Overview of the global quantification strategies in bottom-up proteomic (adapted from ¹⁵⁰). The light and dark gray Petri plates and Eppendorf tubes represent the two conditions to be compared without labeling. When these are orange and blue, labeling has been performed. The tubes that are both orange and blue illustrate the multiplexing of the two conditions. The dashed lines indicate the steps in which variations and errors in quantification can occur.

1. Label-based relative strategies

Labeling quantification strategies are based on the principle that isotope-labeled and unlabeled peptides have similar physicochemical properties, e.g. chromatographic (elution time) and mass spectrometric conditions (ionization efficiency, MS signal response, fragmentation pattern ...). The labeled forms being heavier than the unlabeled ones, their mass will be the only characteristic to differentiate them. These strategies have the advantage of being able to gather samples of interest in order to analyze them in a single acquisition and to compare the MS measurement for quantification.

A. Metabolic or enzymatic labelling strategies

Metabolic labeling of proteins consists in the incorporation of stable isotopic labels during cell culture. This labeling step is performed at an early stage of the proteomics analytical scheme allowing to limit the biases related to the sample preparation protocol for precise and accurate quantification. **SILAC**¹⁵¹ (*Stable Isotope Labelling with Amino acids in Cell culture*) is the most common metabolic labelling strategy. Labeled lysine and arginine residues (¹³C, ¹⁵N) are incorporated into the cell culture. Once the

samples are harvested, the cell lysates are mixed in equivalent amounts before the sample preparation protocol for proteomic analysis. Finally, the comparison of labeled and unlabeled peptide MS signals enables their relative quantification. However, this approach is limited to three conditions and is restricted to a particular type of sample, i.e. cell cultures or SILAC mice¹⁵². To overcome these limitations, new strategies have been developed such as **super-SILAC**¹⁵³ which proposes to incorporate labeled internal standards for relative quantification control, or **NeuCode**¹⁵⁴ (*Neutron enCoding*) which allows to increase multiplexing by the incorporation of lysine isotopologues and by the combination with other labeling strategies. However, these metabolic labeling strategies display a lack of multiplexing, which can be improved via approaches including chemical labeling.

B. Chemical labelling strategies

These strategies involve a labeling with stable isotope to modify reactive groups of peptides, i.e. the α and ϵ amine groups of lysine or the thiol group of cysteine. A first category is based on isotope tagging. **ICAT**¹⁵⁵ (*Isotope Coded Affinity Tag*), introduced in 1999, uses light and heavy reagents composed of a thiol group that will react with the cysteine residues of the proteins, an isotope-labeled linker and a biotin group to enrich the ICAT-labeled peptides *via* streptavidin affinity chromatography after the mixing of the samples to compare. However, this method is limited to cysteine-containing peptides. Modification of retention times with the deuterated form and MS signal interferences, caused by biotin, are other limitations. Alternative approach have been developed to overcome these issues. Among them, the use of ¹³C for labeling or the development of cleavable ICAT agents to remove biotin before MS analysis¹⁵⁶.

The second category uses isobaric labeling. These methods target the N-terminal ends of peptides and proteins as well as the ϵ -amino group of the lysine side chain. These isobaric labels are composed of a group reactive with the primary amines of peptides and a cleavable group upon fragmentation. The latter includes so-called reporter and balance groups that will present several variants. These isobaric variants produce different fragments allowing them to be differentiated. Thus, the relative quantification is obtained by labeling each sample with different isobaric markers at the peptide level. The intensity ratio of the reporter ions, located in the low m/z part of the MS2 spectrum, is used for relative quantification while the b- and y-ions from the peptide are used for identification. One of these approaches, **TMT**¹⁵⁷ (*Tandem Mass Tags*), is commonly used in proteomics. It enables the multiplexing of up to 16 samples and even 18 theoretically¹⁵⁶. Another, named **iTRAQ** (*isobaric Tags Relative and Absolute Quantification*), also allows to generate labeled forms of the same peptide and to multiplex from 4 to 8 samples¹⁵⁸⁻¹⁵⁹, reducing the complexity of the spectra. Both solutions are marketed by Thermo Fisher Scientific and Sciex, respectively. Nevertheless, these approaches suffer from ratio compression due to interferences, which can be alleviated by adding isolation and fragmentation step to generate MS3 spectra¹⁵⁶.

2. Label-free relative quantification approaches

In recent years, label-free quantification strategies have gained popularity in the field of quantitative proteomics, thanks to increasing acquisition speed and sensitivity of mass spectrometers¹⁶⁰. Without the need of stable label or evaluation of the labelling efficiency, label-free method are cost-effective as well as easier and faster to set up. Moreover, they are compatible with all types of samples (cells,

fluids, tissues, etc.) and are not restricted in the number of conditions to be compared since they do not include multiplexing¹⁵⁶. However, samples are therefore prepared individually and analyzed separately under similar conditions. Thus, each step in the analytical scheme can have a significant impact on the precision and accuracy of the quantification. To reduce the variability between samples, all conditions are often analyzed in the same sequence on the same robust and stable LC-MS system¹⁶¹. The DDA acquisition mode is the most commonly used for label-free quantification strategies, providing information on the identity and quantity of peptides and proteins present in samples.

A. Spectral counting

The spectral counting strategy is based on the assumption that the abundance of a protein is correlated with the number of MS2 spectra acquired for it. A major advantage of this approach is its simple data processing, since it requires only the tools used for protein identification and validation.

However, this strategy suffers from the stochasticity and undersampling of the DDA acquisition mode. As a consequence, it generates missing values and affects the repeatability between samples. In addition, the dynamic exclusion parameter, which excludes previously selected precursor ions, must be removed or minimized to preserve the redundancy of the spectra that correlates with protein abundance. Otherwise, a bias towards small variations and low abundance proteins is observed¹⁵⁶.

Furthermore, another limitation of this approach comes from the fact that the number of peptides, link to number of associated spectra, is dependent on protein length. Quantification of small proteins (<20 kDa) would be less accurate than for large proteins. To overcome this issue, normalization methods based on protein length, protein mass or detection probability of an MS2 spectrum were proposed¹⁵⁶. Shared peptides would also affect the quantification if the PSM spectra is assigned to all proteins. It is rather recommended to proportionally distribute them between the proteins involved by considering the distribution of the unique peptides¹⁶²⁻¹⁶³. Finally, label-free quantification data should be normalized to minimize the variations that may occur during the analytical workflow.

B. Extracted ion chromatogram

The ion current extraction method is based on the assumption that the abundance of a peptide is linearly correlated with its MS-acquired chromatographic signal, named XIC ("eXtracted Ion Chromatogram")¹⁶⁴. The XIC quantification can be obtained with the height or area under the curve of the peptides as they elute on the chromatographic column. It is collected at the MS1 signals while the MS2 signals are used for peptide identification. Since the DDA acquisition mode is usually used for this strategy, the parameters of the method must be balanced. The aim is to collect enough MS1 spectra to correctly define the chromatographic peaks and obtain an accurate quantification. But at the same time, it is also essential to generate enough high-quality MS2 spectra to achieve a good coverage of the proteome. Mass spectrometers with good mass accuracy, also called HR/AM ("High-Resolution/Accurate-Mass") are used, allowing good separation of the precursor ion isotopes at defined m/z and retention times¹⁶⁵. It is also recommended to use a robust chromatographic system for good signal discrimination and to ensure an accurate quantification¹⁶².

However, this approach requires a more sophisticated data processing algorithm than the spectral counting strategy. Indeed, to compensate for possible variations induced during the steps of the

analytical workflow (sample preparation, chromatographic variations and signal instability), an alignment of the retention times and a normalization of the data can be performed^{156, 163, 166-167}. Several software were developed, including open source solutions such as Skyline¹⁶⁸, Maxquant¹⁶⁹ or even Proline¹⁹, developed by the French proteomics infrastructure ProFI, and those under license such as SpectroMine™ (Biognosys) or Proteome Discoverer (Thermo Fisher Scientific). Comparative studies have already investigated the performance of the different software solutions for processing quantitative proteomics data^{19, 165, 170}. The role of these tools is to detect and integrate the chromatographic peaks of ions with a peptide isotope profile, called "features", in order to generate LC-MS maps gathering information on the m/z ratio, retention time, charge and peak intensity. Then, a second step consists of RT alignment, peptide sequence identification, signal normalization, protein identification and quantification¹⁶². The alignment performed enables the peptide identification between LC-maps of different condition, thus limiting the number of missing value^{162, 171}. Alternatively, a list of precursor ions can be defined from database searches of LC-MS/MS analyses of interest. The generated spectral library, including the MS2 spectra of the identified peptides, is used to query their presence in the MS1 spectra. Finally, the XIC is obtained and used for quantification.

3. Label-free "absolute" quantification

The aim of label-free "absolute" quantification strategies is to provide an estimation of the protein quantity. They enable the quantification of a large number of proteins with a reduced cost and a simplified experimental protocol compared to labeled-based approaches¹⁵⁶.

A first approach, named emPAI¹⁷² (*exponentially modified Protein Abundance Index*), is defined as $\text{emPAI} = 10^{\text{PAI}} - 1$. It is the extension of PAI¹⁷³, which provides an estimation of the protein abundance by calculating the ratio of the number of observed peptides per protein to the number of observable peptides. Other approaches using intensities have been developed. This is the case of iBAQ ("intensity-Based Absolute Quantification"), which divides the sum of the intensities of all peptide peaks by the number of theoretically observed peptides¹⁷⁴. There is also the Top3¹⁶ approach, which uses an unlabeled protein spiked in the sample as a standard protein. The abundance is then calculated based on the three most intense peptides of a protein by comparing them with the three most abundant peptides of the standard protein. This latter strategy was used in the work presented in this manuscript.

Chapter 3

Targeted quantification strategies

Targeted proteomics strategies became popular for the quantification of proteins of interest with high sensitivity, quantification accuracy and reproducibility¹⁷⁵. Indeed, even though global proteomics approaches allow the identification and quantification of a significant number of proteins, the DDA acquisition mode remains limited in terms of sensitivity, reproducibility and dynamic range¹⁷⁶. Thus, targeted approaches are preferred for the precise quantification of a few dozen target proteins. SRM (Selected Reaction Monitoring), the reference method in targeted proteomics, is performed on triple quadrupole (QQQ) mass spectrometers. However, these approaches have been successfully implemented on HR/AM instruments, such as PRM (Parallel Reaction Monitoring) on Q-Orbitrap instruments¹⁷⁷.

During this work, a targeted approach was developed on the BioAccord system, a user-friendly instrument. However, this system is only composed of a TOF analyzer without a quadrupole for ion selection. The targeted approach was therefore optimized at the data processing level in order to extract the signals of interest. Besides, only HR/AM instruments were used. Therefore, this chapter will focus on the PRM approach that can be implemented on such instruments and more particularly to the Q-Orbitrap HF-X system.

1. Selected reaction monitoring

SRM, also called Multiple Reaction Monitoring (MRM), is the reference method for targeted proteomic approaches¹⁵⁶. It is conducted on QQQ mass spectrometers, composed of three quadrupoles: the first (Q1) and the third (Q3) are used to filter precursor and fragments ions according to their m/z ratios, while the second (Q2) is used as a collision-induced dissociation cell. In order to perform a SRM study, the analyst creates an acquisition method based on a transition list, i.e. a predefined list of precursor and fragment ions to be sequentially isolated. The precursor and a fragment ion pair is called a transition. For each of these transitions, ion chromatograms are extracted and then combined when they come from the same precursor ion (Figure 25).

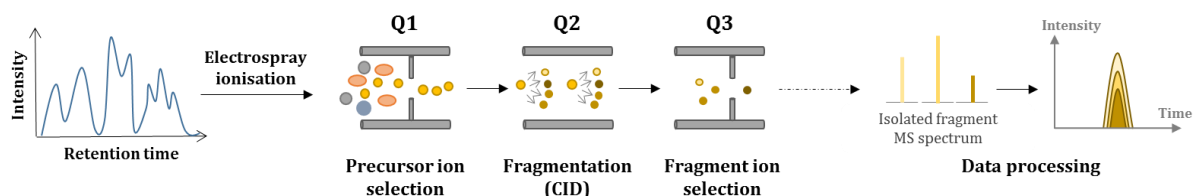


Figure 25: Selected Reaction Monitoring (SRM) principle.

The selection of a precursor ion and a specific fragment ion during the SRM mode results in a quantification with high specificity and sensitivity. In addition, the higher the number of transitions followed, the higher the specificity^{175, 178}.

2. Parallel Reaction Monitoring

A. The principle of Parallel Reaction Monitoring

The development of HR/AM mass spectrometers, such as the Q-Orbitrap or Q-TOF, has allowed the emergence of a new targeted approach. It is known as Parallel Reaction Monitoring (PRM) when performed on a Q-Orbitrap instrument^{177, 179}, but also as MRM-High Resolution (MRM-HR) or Targeted MS/MS. Like SRM, the predefined precursor ions are sequentially isolated using a quadrupole, which acts as a mass filter (1-2 m/z), and then fragmented in a collision cell. However, unlike SRM, no selection of fragment ions is performed. Thus, all these ions are analyzed simultaneously and full MS2 spectra are generated. Finally, ion chromatograms are extracted for each precursor and fragment ion pair (transition) (Figure 26).

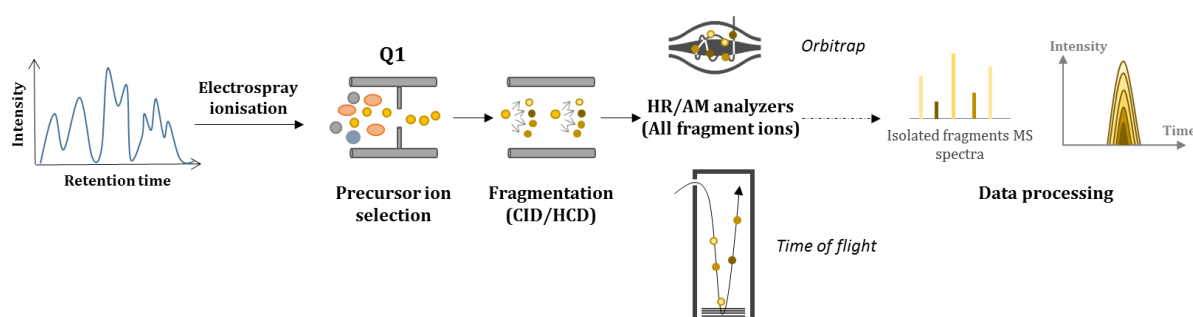


Figure 26: Parallel Reaction Monitoring (PRM) principle.

The SRM and PRM approaches demonstrate similar performances in terms of reproducibility, accuracy and precision of quantification^{58, 156, 180}. The PRM approach is an easier method to implement, since there is no need to predefine the transitions to monitor. This advantage greatly facilitates the optimization of the method. Another advantage of PRM is the use HR/AM instruments, which improves selectivity, sensitivity and dynamic range^{59, 156, 177}. Moreover, they offer the possibility to perform both non-targeted (discovery proteomics experiments such as DDA) and targeted analyses (PRM experiments) on the same instrumental coupling facilitating method optimization through the transfer of key parameters¹⁸¹.

B. The development of a PRM method

The steps required to develop a PRM method are presented in Figure 27. The first two steps will be detailed in this section. The PRM approach allows the quantification of a limited number of peptides of interest with a high specificity and an increased sensitivity. However, the optimization of the method can be complex and time-consuming even if it is simplified compared to an SRM method.

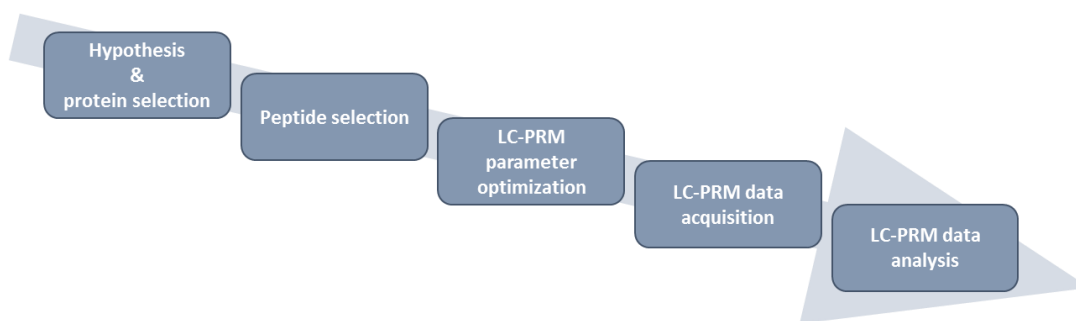


Figure 27: Workflow of the PRM assay development.

i. Hypothesis and protein selection

PRM, like SRM, is a hypothesis-driven approach. The targeted proteins must be selected beforehand according to the biological context and the results of the proteomic discovery study. This approach enables the monitoring of hundreds of proteins with optimized acquisition parameters to maximize the multiplexing capabilities.

ii. Peptide selection

To infer protein abundance, a representative set of peptides is used. For a robust quantification, a minimum of two peptides per protein is recommended. The selection of these peptides is based on several criteria¹⁸¹:

- **Unicity:** peptides must be unique to the proteins of interest to ensure high specificity. This information can be obtained by performing a BLAST³¹ (Basic Local Alignment Search Tool) search against the proteome of interest to determine sequence homologies.
- **Peptide sequence:** between 7 and 25 amino acids are recommended in order to obtain m/z ratio values within the mass range of the instrument. Sequences should not contain any enzyme missed cleavage sites (unless a proline is at the carboxyl end of the arginine or lysine), nor should there be any series of arginine or lysine at the ends of the sequences, e.g. KK, KR, RK or RR. Finally, it is recommended to avoid peptides subject to chemical modifications (oxidation, deamidation, N-terminal cyclization).
- **MS response:** peptides already detected during LC-MS/MS analyses are preferred in order to obtain a high sensitivity. The use of previous experiments or publicly available databases (PRIDE¹⁸², PeptideAtlas¹⁸³, Panorama¹⁸⁴, SRMAtlas¹⁸⁵, NeXtProt¹⁸⁶ or Human Proteinpedia¹⁸⁷) can help to select peptides that display well-defined chromatographic peak and efficient ionization. If experimental data are not available, prediction tools exist to predict LC-MS behaviors such as PeptideSieve¹⁸⁸, ESP¹⁸⁹, PREGO¹⁹⁰ or even AP3¹⁹¹.

Peptides meeting the criteria of unicity and MS-response are called proteotypic, while those that meet the three criteria are described as quantotypic¹⁸¹.

C. Absolute quantification using isotopic dilution

i. Label-based absolute quantification

In order to obtain an absolute quantification of peptides and proteins, it is possible to add to the samples a known quantity of isotopically labelled peptide (AQUA¹⁹², QconCAT¹⁹³, ...) or protein standards (PSAQ¹⁹⁴, FLEXIQuant¹⁹⁵, PrEST¹⁹⁶, ...). These standards are designed with the same amino acid sequence as the endogenous peptide or protein to maintain the same physicochemical properties. However, the overall mass is modified due to the labeling. In proteomic studies including trypsin as protease, peptides and proteins are usually labeled on the C-terminal side of the amino acids arginine and lysine with ¹³C and ¹⁵N, resulting in a mass increase of 10.01 and 8.01 Da, respectively.

ii. Isotopic dilution

Isotopically labeled peptides are commonly used for the accurate detection and quantification of endogenous peptides of interest. However, synthetic standards are commercially available with different degrees of quality:

- Low-quality synthetic crude peptides are used to verify the presence of endogenous proteins in samples, or for relative quantifications of peptides or proteins of interest. These standards have a low purity and are added at unknown quantities in samples. However, they have the advantage of being cheap (about 20€/peptide).
- High quality synthetic peptides are preferred when the objective of the study is to achieve a precise and accurate quantification, since these standards are spiked at known amounts in the samples. However, these standards are expensive (about 300€/peptide) and can therefore represent an economic constraint.

The amount of these isotopically labeled standard peptides must be optimized according to the samples. The aim is to obtain a balanced concentration between the light form (endogenous peptide) and the heavy form (labeled peptide). A ratio between the two forms close to 1 maximizes the sensitivity of the method. Peptide quantity is estimated by summing the areas under the curves of the different transitions. Thus, the ratio of the areas of the endogenous peptide and its isotopically labeled form determines the amount of endogenous peptide in the sample (Figure 28).

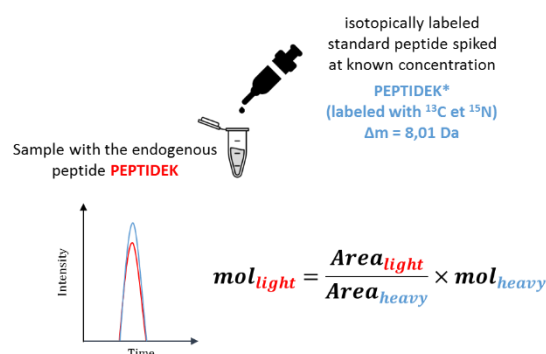


Figure 28: The principle of isotopic dilution. Isotopically labeled peptides are spiked into the samples at equivalent and known amounts. Then, the area ratios of light and heavy peptides are used to determine the amount of endogenous peptide.

Peptide quantification is therefore performed from the XIC (extracted ion chromatogram) that was discussed in the previous chapter (Chapter 2, part 2.B). Dedicated software for the processing of these data has been developed. Among them, Skyline¹⁶⁸ is the most widely used because of its free access but also because of the numerous functionalities and representations it offers, facilitating the work of the user when inspecting the data. We can also mention the SpectroDive™ software, commercialized by Biognosys, which includes a high quality automated algorithm for peak detection.

iii. The determination of the quantification limits

In order to determine the absolute amount of peptides, it is essential to determine the conditions for which the area under the chromatographic signal curve is directly correlated to the peptide amount. Therefore, calibration curves are generated from the MS response of the standard peptide in order to determine the limits of quantification (LOQ). In targeted proteomics, these curves are obtained by adding different amounts of isotopically labeled peptides to a mixture representative of the sample matrix with the endogenous peptides. Thereafter, strict data validation criteria are applied: $CV \leq 20\%$ between replicates, a difference of less than 20% from the expected value, and a coefficient of curve determination (R^2) greater than 0.99. Finally, the lowest amount of peptide allowing accurate quantification will be defined as the low limit of quantification (LLOQ) and the highest amount as the upper limit of quantification (ULOQ) (Figure 29).

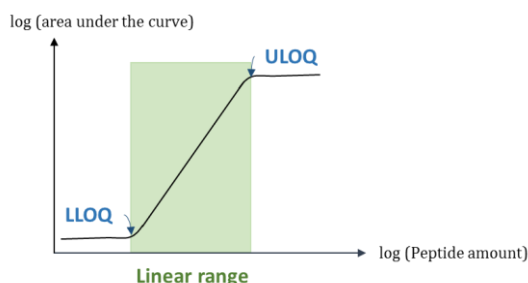


Figure 29: Limits of quantification of peptides. Peptide chromatographic signal is correlated to the amount of peptide in a linear range (green area). This range is determined by calibration curves for each peptide. Thus, the lower limit of quantification (LLOQ) and the upper limit of quantification (ULOQ) can be obtained.

Chapter 4

Data-independent acquisition

Proteomics aims to provide the identification and quantification of several proteins in multiple samples. However, although DDA-based approaches allow the quantification of several thousands of proteins, they display a limited sensitivity and specificity and a restricted dynamic range. Targeted approaches (SRM and PRM), on the other hand, offer increased sensitivity, specificity and dynamic range in a multitude of samples but for a limited number of target proteins. Over the last decade, developments in MS have allowed the emergence of a new mode of acquisition: Data Independent Acquisition (DIA). This mode of acquisition promises to combine the advantages of both approaches: the proteome coverage of DDA on the one hand, and the high sensitivity, specificity, accuracy and dynamic range of targeted methods on the other.

Indeed, the instrumental developments in the early 2000s have resulted in very fast mass spectrometers with high resolution. Moreover, the interest in this approach has been intensified by the numerous algorithms dedicated to the processing of DIA data that were developed by the scientific community^{45, 59} (Figure 30). As further evidence of this success, the American Society of Mass Spectrometry (ASMS) hosts a dedicated session each year.

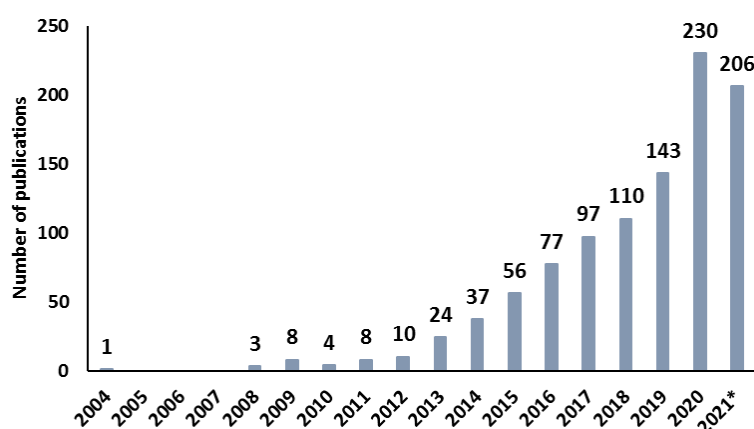


Figure 30: Number of publications for which the term "data-independent acquisition" is included in the abstract. *For 2021, the number of publications was collected on 07/09/2021.

1. Principle of data-independent acquisition

The DIA acquisition mode allows the collection of MS/MS spectra of all peptides along the chromatographic gradient in an untargeted and unbiased manner. It is performed on HR/AM instruments, such as Q-TOF or Q-Orbitrap. Briefly, DIA relies on the generation of MS₂ spectra of all precursor ions isolated in predefined isolation window, or possibly contained within the entire mass range. The acquisition of the fragments from all co-isolated and co-fragmented peptides results in multiplexed MS₂ spectra. Finally, chromatographic peaks are extracted for all detected fragment ions in order to perform their quantification (Figure 31).

Thus, DIA mode enables the collection of information on all peptides contained in complex samples, with the only constraint being the detection limit of the instrument⁵⁹.

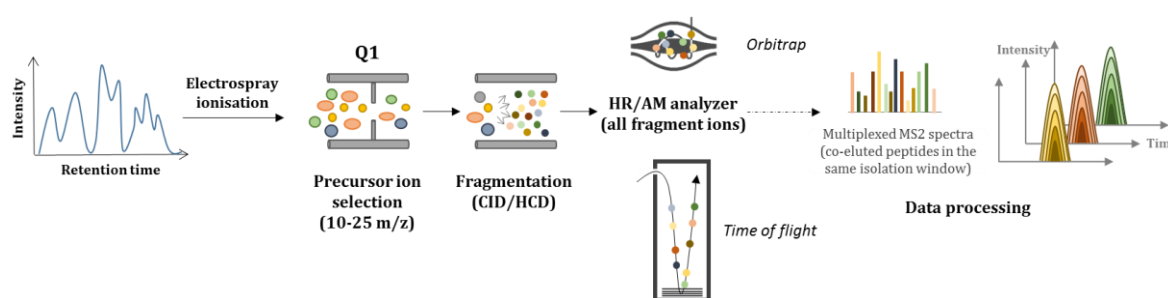


Figure 31: The principle of Data-Independent Acquisition mode. Schematic representation of DIA strategies based on isolation windows.

2. Evolution of the DIA-based approaches

DIA was introduced in 2000 by Masselon *et al.*¹⁹⁷. In this experiment, several polypeptides were characterized from multiplexed MS/MS spectra generated on an instrument with high mass measurement accuracy, a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR).

In 2003, **shotgun CID** was proposed by Purvine *et al.*¹⁹⁸. This alternative approach is based on in-source fragmentation of peptides by CID. A first analysis is performed at low source voltage to limit the CID fragmentation and thus obtain the MS spectra of the precursor ions. Then, a second analysis is conducted at high source voltage in order to induce the fragmentation and generate the MS spectra of the fragment ions.

However, the term "**data-independent acquisition**" did not appear until the work done by Venable *et al.*¹⁹⁹. In 2004, they proposed a new alternative strategy based on sequential isolation and fragmentation of precursors included in large isolation windows (10 m/z) *via* a linear ion trap mass spectrometer.

Since then, several DIA-based approaches have been developed on different instruments, using different strategies for MS acquisition parameters but also for data processing. Table summarizes these methods. The strategies can be gathered in two main categories: those performed on the whole mass range and those based on the use of isolation windows.

DIA method	Year of Introduction	Precursor isolation window size (m/z)	Reference
Shotgun CID	2003	Full range	Purvine et al. ¹⁹⁸
DIA (original)	2004	10	Venable et al. ¹⁹⁹
MS ^E	2005	Full range	Silva et al. ²⁰⁰
PAcIFIC	2009	2,5	Panchaud et al. ²⁰¹
AIF	2010	/	Geiger et al. ²⁰²
XDIA	2010	20	Carvalho et al. ²⁰³
FT-ARM	2012	100	Weisbrod et al. ²⁰⁴
SWATH	2012	25	Gillet et al. ¹⁴
HDMS ^E	2012	Full range	Geromanos et al. ²⁰⁵
MSX	2013	4	Egertson et al. ²⁰⁶
pSMART	2014	5-20	Prakash et al. ²⁰⁷
WiSIM-DIA	2014	12	Martin et al. ²⁰⁸
UDMS ^E	2014	Full range	Distler et al. ²⁰⁹
SWATH (variable windows)	2015	8-85	Zhang et al. ²¹⁰
HRM	2015	24-220	Bruderer et al. ²¹¹
SONAR	2018	24	Moseley et al. ²¹²
BoxCar DIA	2018		Meier et al. ²¹³
Scanning SWATH	2019	5	Messner et al. ²¹⁴
diaPASEF	2019	25	Meier et al. ²¹⁵
PulseDIA	2019	4	Cai et al. ²¹⁶
DIA-FAIMS	2020	13.7	Bekker-Jansen et al. ²¹⁷

Table 4: Evolution of DIA approaches (adapted from ⁴⁴⁻⁴⁵).

A. DIA strategies performed over the entire mass range

A first alternative to the CID shotgun, called **MS^E**, was proposed in 2005 by Waters²⁰⁰. This strategy takes advantage of technological developments on the Q-TOF instruments allowing the acquisition of MS1 and MS2 spectra over the entire mass range along the chromatographic gradient. For this purpose, an alternation between low and high collision energy is performed in order to generate MS spectra of precursor ions and fragment ions respectively. This approach was used in the development of an approach on the BioAccord system (Part III, chapter 1).

Thermo Fisher Scientific introduced a similar approach in 2010 called **All-Ion Fragmentation**²⁰² (AIF). It is based on the sequenced acquisition of MS1 spectra of precursor ions and MS2 spectra generated after HCD fragmentation of all precursor ions without prior selection. This method was developed on an Exactive instrument.

In 2012 and 2014, the **High-Definition MS^E** (HDMS^E)²⁰⁵ and **Ultra-Definition MS^E** (UDMS^E)²⁰⁹ approaches were proposed. These methods represent two successive advances of the MS^E strategy. First, the addition of ion mobility, as an additional separation dimension, has reduced the complexity of MS2 spectra. Second, the application of variable collision energies as a function of ion mobility data allowed the optimization of precursor ion fragmentation based on their masses.

B. DIA strategies based on isolation windows

i. Consecutive fixed width windows

The **Precursor Acquisition Independent From Ion Count** (PACIFIC) approach was proposed in 2009 by Panchaud et al.²⁰¹. It is based on the use of narrow isolation windows (2.5 Da) to reduce the MS2 spectra complexity. However, it requires multiple injection of the sample over several days to cover the entire mass range. An improved version of this method was published in 2011, reducing the required acquisition time thanks to instrumental developments²¹⁸. Furthermore, a similar strategy, called **PulseDIA**, was proposed more recently by Cai et al.²¹⁶. The PulseDIA strategy aims to decrease the number of windows and thus the number of injections required with the use of isolation window sizes adapted to the ion density.

In 2010, Carvalho et al.²⁰³ proposed an alternative strategy, named **eXtended Data-Independent Acquisition** (XDIA). It relies on the acquisition of a high-resolution MS1 spectrum at the beginning of each cycle followed by a combination of CID and ETD to fragment the precursor ions.

The **Fourier Transform-All Reaction Monitoring** (FT-ARM) approach was introduced by Weisbrod et al.²⁰⁴ in 2012. Precursor ions are co-isolated and co-fragmented using 12 or 100 m/z windows on LTQ-FT or LTQ-Orbitrap instruments. The same year, Gillet et al.¹⁴ proposed a similar strategy on a Q-TOF instrument. This approach is called **Sequential Windowed Acquisition of All Theoretical fragment ion spectra** (SWATH) and uses isolation windows of 25 m/z. It is now commercialized by the Sciex company.

Next, Prakash et al. and Martin et al. used the parallelization capabilities of some Q-Orbitrap instruments to propose the **pSMART**²⁰⁷ and **Wide Selected-Ion Monitoring**²⁰⁸ (WiSIM) approaches, respectively. Therefore, while high-resolution MS spectra are acquired, MS2 spectra are generated independently after isolation of precursors in restricted mass windows. The MS1 spectra are then used for quantification while the MS2 spectra are used for identification.

In 2019 and 2020, the addition of ion mobility in front of HR/AM instruments, as an additional separation dimension has led to improvements in SWATH. Thus, **diaPASEF**²¹⁵ has been implemented on the TimsTOF Pro (Bruker, Billerica, MA, USA) while **DIA-FAIMS**²¹⁷ has been developed on the latest generation Q-Orbitrap instruments.

In addition to these approaches, the **BoxCar** acquisition method seems interesting. The method was implemented in 2018 by Meier et al.²¹³. It relies on MS1 acquisition of narrow mass windows to improve the ion filling time compared to a standard method. Thus, an increased dynamic range is obtained from MS1 signals. This approach seems promising, if combined with DIA, as it would result in much better precursor information²¹⁹.

ii. Consecutive variable width windows

In LC-MS/MS analysis, tryptic peptides are not homogeneously distributed over the m/z acquisition range along the chromatographic gradient. Therefore, the use of different size isolation windows based on ion density would allow the ion population to be evenly distributed. As a result, better selectivity and overall quantification performance would be achieved. To this end, the **SWATH approach with variable windows** was proposed by Zhang et al.²¹⁰. Small isolation windows are used when the ion density is high and larger windows are defined when the number of competing ions is lower. Software tools have been developed, such as swathTUNER, to automate the optimization of the isolation window size to uniformly distribute the precursor ion density. This strategy is implemented by Sciex as SWATH 2.0.

The **Hyper Reaction Monitoring**²¹¹ (HRM) approach is a similar strategy but conducted on Q-Orbitrap instruments. It was implemented by Bruderer et al. and this term is now the property of the company Biognosys. However, it should be mentioned that unlike SWATH methods developed on Q-TOF instruments, the isolation scheme of an HRM method must be written manually for acquisition on a Q-Orbitrap instrument.

iii. Overlapping windows

Other strategies are based on the use of overlapping isolation windows (Figure 32.A). Thus, the isolation scheme is composed of windows that cover half of the mass range already covered by the previous window, resulting in increased selectivity. Information from overlapping MS2 spectra is used to generate demultiplexed MS2 spectra (Figure 32.B).

In 2018, Moseley et al.²¹² developed the **SONAR** approach on a Waters Q-TOF instrument. The MS2 spectra are acquired continuously over the 400-900 m/z mass range using 24 Da isolation windows. Similarly, the **Scanning SWATH** strategy has been developed by Messner et al.²¹⁴ for restricted chromatographic gradients. It reduces cycle time compared to conventional DIA methods because successive window acquisition is replaced by continuous scanning with the first quadrupole.

iv. Multiplexed strategies

In 2013, Egertson et al.²⁰⁶ proposed a new acquisition method termed as MSX. It consists of sequentially co-isolating the precursor ions contained in randomly selected isolation windows (Figure 32.A). The mass range 500-900 m/z is thus divided into 100 windows of 4Da each. The MS2 spectra are then computationally demultiplexed to increase the selectivity and the signal-to-noise ratio. This approach is based on the multiplexing capabilities of the Q-Orbitrap instruments but it may still suffer from a loss of sensitivity due to the limited time for ion trapping.

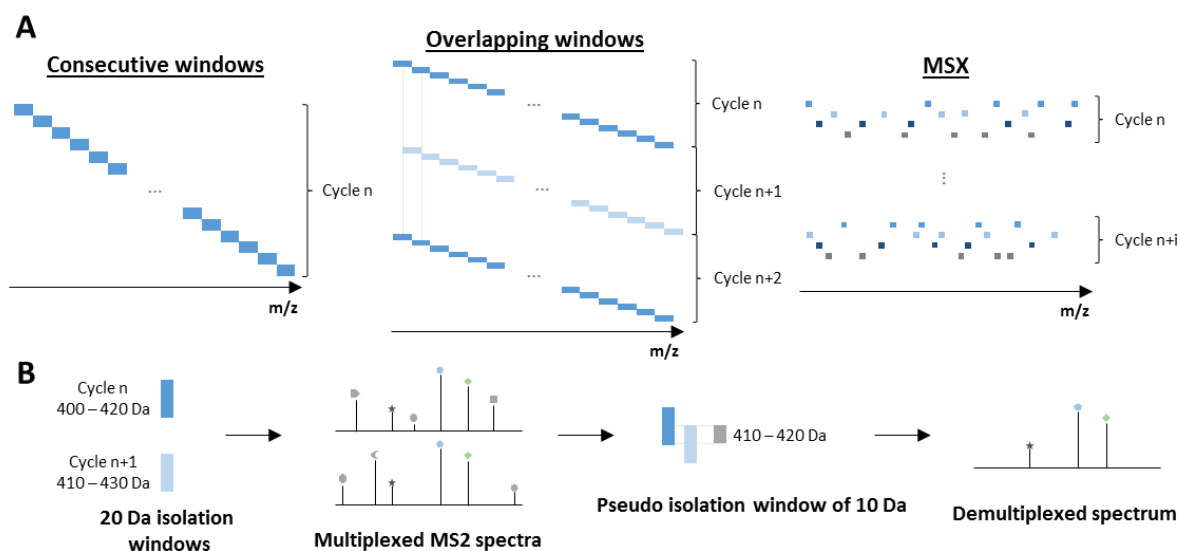


Figure 32: Description of multiplexed approaches in DIA. Comparison of isolation schemes of two multiplexing strategies (overlapping isolation windows and the other with random windows) with the isolation scheme constitutes of consecutive windows (A). Example of MS2 spectra demultiplexing when using an overlapping windows strategy (B).

3. DIA assay development

The process to develop a DIA method is shown in Figure 33. In this manuscript, the DIA method was optimized on a Q-Orbitrap instrument. Therefore, the following section will focus on this type of instrument.

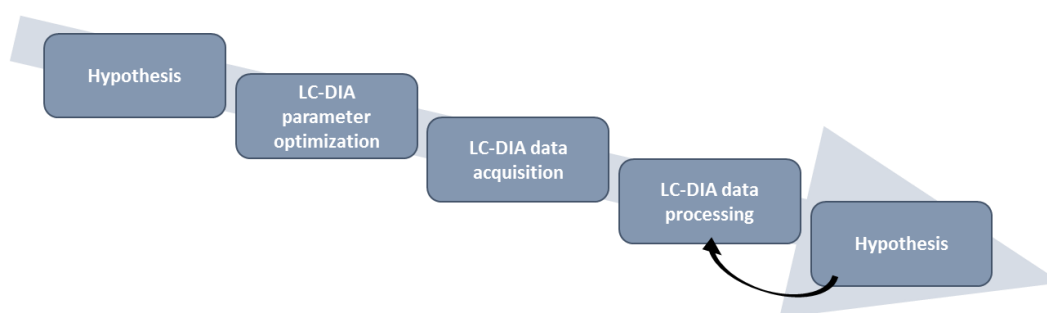


Figure 33: Workflow of the DIA method development.

A. Hypothesis

The development of all proteomics studies begins with a hypothesis. Unlike targeted proteomics approaches, this hypothesis does not condition data acquisition.

B. LC-DIA parameter optimization

The optimization of instrumental parameters remains crucial to achieve high sensitivity, specificity and quantification accuracy, while ensuring a good coverage of the proteome. Several parameters need to be optimized: the number of targeted precursor ions, the cycle time, the number and the window size (Figure 34).

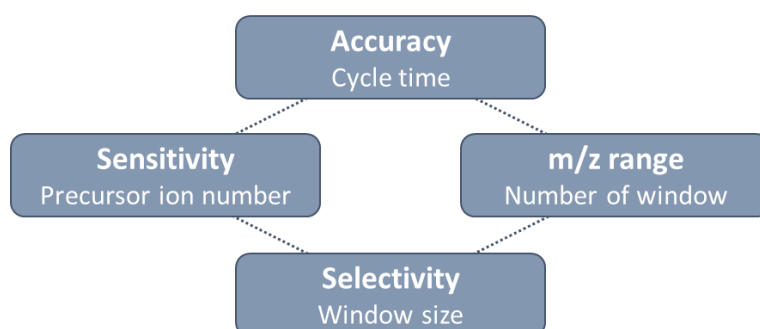


Figure 34: Instrumental parameters to optimize for variable windows DIA analysis.

The **number of ions** accumulated and analyzed is critical for the sensitivity of the method. Several parameters will influence this number, such as the maximum injection time and the automatic gain control target (AGC target) value. The first refers to the maximum time allowed for the accumulation of ions in a trapping cell, called C-trap, before injection into the Orbitrap. The second is set to limit the maximum number of ions to be accumulated and thus limit the duration of this step when a large amount of ions is available. This accumulation step is performed in parallel with the detection of ions in the Orbitrap to avoid the loss of analysis time. Thus, the higher the number of accumulated ions, the higher the sensitivity. However, even if the accumulation of the ions is carried out in parallel of the detection, it is essential to define these parameters carefully in order to avoid increasing the cycle time.

The **cycle time** represents the time required to perform the acquisitions in all isolation windows that cover the mass range. This parameter is crucial to achieve an accurate XIC-based quantification but also depends on the chromatographic conditions influencing peak width. Between 8 and 10 MS scans are recommended to obtain a good chromatographic peak resolution. This means that for a 30 seconds chromatographic peak, the cycle time should not exceed 3 seconds to have 10 MS acquisition points (10 cycles). The cycle time is calculated according to the following equation:

$$\begin{aligned}
 &\text{If maximum injection time} < \text{Transient time} \\
 &\text{Cycle time} = \text{MS1 accumulation time} + \\
 &(\text{Number of isolation windows} * \text{MS2 accumulation time})
 \end{aligned}$$

The transient time or acquisition time is linked to the resolving power, which corresponds to the capacity of the Orbitrap to discriminate ions according to their m/z ratios. Thus, the higher the value of the resolution, the greater the transient time. For instance, for resolution values of 15 000, 30 000 and 45 000 (at 200 m/z), the respective transient times are equal to 32, 64 and 128 ms on the Q Exactive HF-X.

The **window size** refers to the mass range over which the precursor ions are co-isolated and co-fragmented. It will directly impact the selectivity of the method, since the MS2 spectra complexity increases with the isolation window size, but also the sensitivity with a reduced dynamic range.

The succession of **isolation windows** will allow to cover a certain m/z range during a cycle. Thus, the higher the number of windows, the longer the cycle time will be. It must be defined carefully in order to ensure a good coverage of the proteome while maintaining a high sensitivity and a high accuracy of quantification. The distribution of tryptic peptides as a function of the m/z ratio depends on the sample but is generally between 350 and 1200 m/z , with an increased density between 400 and 800 m/z (Figure 35). Thus, a balance between all these parameters must be established according to the instrument speed in order to cover the largest mass range while adjusting isolation windows.

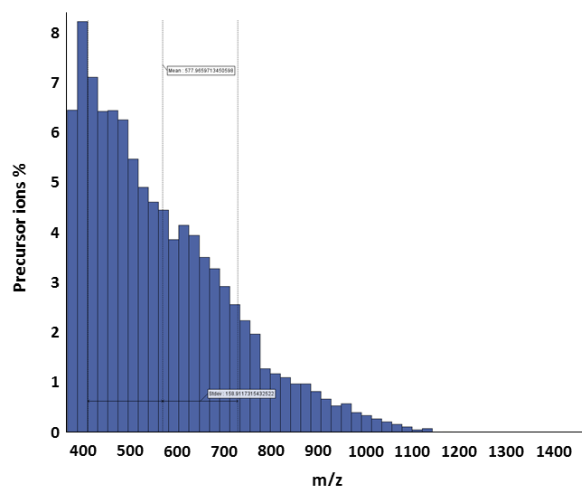


Figure 35: Precursor ion percentage as a function of m/z ratio for analysis of a HeLa digest.

Finally, a last factor to optimize for a DIA analysis is related to the ion transmission of the quadrupole mass analyzer. Indeed, even if it is almost square, the transmission efficiency is not optimal at the limits of the mass range. To overcome this constraint, it was proposed to set a 1 Da overlap between two consecutive windows¹⁴ (Figure 36). Thus, fragment ions are extracted over the mass range covered by the isolation window excluding the margins on both sides.

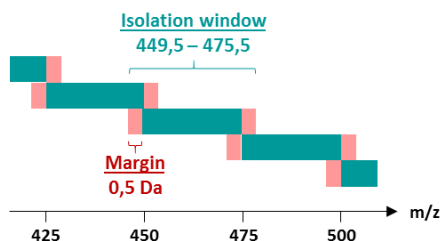


Figure 36: Q1 isolation window slight overlap.

In Part I, Chapter 2, Section 3 of this manuscript, the optimization of several key acquisition parameters of a DIA method performed on the Q-Exactive HF-X is detailed.

C. LC-DIA data processing

The processing of DIA data requires dedicated bioinformatics tools and strategies due to the MS2 spectra complexity. This step is described in the following section.

D. New hypothesis

Since the DIA approach theoretically collects the MS2 spectra of all detectable peptides contained in a sample, the DIA data can be reprocessed endlessly based on new hypothesis.

4. The challenge of data processing

Data processing remains the bottleneck of the DIA strategy. Indeed, the multiplexed MS2 spectra generated in cannot be interpreted with the conventional tools used for DDA data^{45, 220}. Therefore, dedicated bioinformatics tools have been developed. They are based on two different approaches: peptide-centric and spectra-centric (Figure 37).

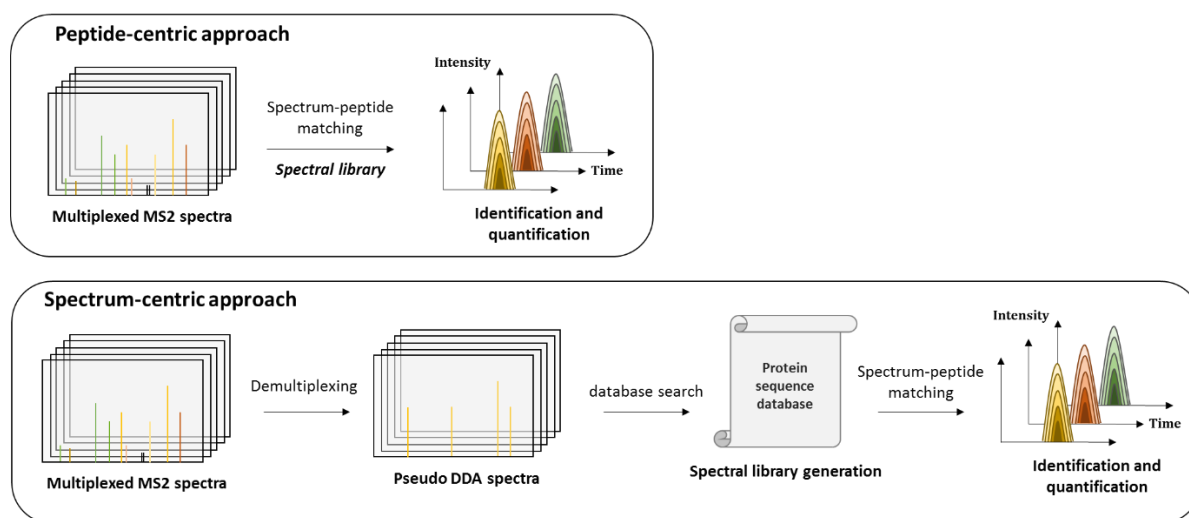


Figure 37: DIA data processing strategies (adapted from ²²¹). The peptide-centric approach involves the use of a previously generated spectral library to query for the presence of peptides in the DIA-MS2 spectra. The spectrum-centric approach aims at identifying peptides directly from DIA spectra. For this purpose, pseudo DDA spectra are generated from the DIA-MS2 spectra and a classical database search is performed. The assigned spectra are used to generate a spectral library and further extract the signals *via* a peptide-centric approach.

A. Peptide-centric approach

The analysis of DIA data based on a peptide-centric approach can be performed by targeted data extraction or by direct comparison of the spectra. It can be performed using a spectral library or possibly a protein database.

i. Extraction via a spectral library

The spectral library is built from MS2 spectra that have been assigned to a peptide sequence with a high level of confidence^{44, 222}. These spectra can be collected from previously acquired data in DDA mode or from the deconvolution of spectra acquired in DIA mode. The use of a spectral library for DIA data extraction implies that only the peptides contained in the library are targeted. It is therefore essential to ensure the generation of a complete spectral library. The limitations of the DDA acquisition mode do not allow the use of the same sample acquired with the DDA mode. Therefore, it is required to perform sample fractionation or peptide/protein enrichment beforehand. Unfortunately, these steps can be time-consuming, tedious, and expensive²²³. Alternatively, recombinant proteins can be expressed or peptides synthesized to ensure the presence of the peptides and proteins of interest in the spectral library²²⁴. "Hybrid" spectral libraries, resulting from a combination of endogenous and synthetic peptide analyses, have already been performed²²⁵.

To avoid these additional steps related to the generation of the sample-specific spectral library, public spectral libraries are available. This information can come from dedicated platforms that collect spectral libraries for DIA-SWATH data extraction (PeptideAtlas¹⁸³, MassIVE, PRIDE²²⁶ or SWATHAtlas²²⁷). Moreover, it is also possible to access files that are available on various public platforms. They were used to generate spectral libraries in a specific context and can be reinvested in other studies.

The use of a spectral library generated from data acquired on the same LC-MS coupling and under the same chromatographic conditions remains the optimal method for DIA analysis^{223, 228}. Nevertheless, studies have shown that MS2 spectra acquired on instruments using the CID fragmentation are sufficiently comparable. Moreover, the normalization of retention times through standard peptides, e.g. iRT peptides (Biognosys), allows the transfer of chromatographic information between different system. The only requirement being that the peptide elution order is maintained⁴⁴⁻⁴⁵. An approach named Multiple Characteristic Intensity Pattern (MCIP) has been introduced and allows a better understanding of the spectral variability during the generation of libraries²²⁹. Thus, multiple publicly available LC-MS/MS data sets can be used, as long as standard peptides for retention time normalization are included in the analyses.

More recently, artificial intelligence-based tools were developed for the prediction of MS2 spectra (DeepMass²³⁰, Prosit²³¹, pDeep²³²). Depending on the solution used, they can predict fragmentation patterns, fragment ion intensity and/or retention time. Several software packages, such as Spectronaut (Biognosys) or DIA-NN²³³, now include artificial intelligence in their algorithm to improve the analysis of DIA data. However, the predicted fragment intensities remain instrument dependent²³⁴. Regardless of the strategy used, the quality of the spectral library remains essential. Indeed, all the precursor ions contained in the library are considered as correctly identified, even if it has been

validated with an FDR of less than 1%. This means that the remaining false positives are also considered as true identifications^{222, 229}. Since spectral libraries can represent large search space, it is necessary to control the error rate through extensive statistical tests^{44, 235}.

ii. Targeted data extraction

Targeted extraction of MS2 spectra was proposed by Gillet et al.¹⁴ for processing SWATH data. This approach exploits the information contained in the spectral library to query for the presence of peptides in DIA analyses. It uses the information:

- At the peptide level: sequence and normalized retention time,
- At the precursor ion level: m/z ratio and charge,
- At the fragment ion level: fragment type, m/z ratios, charges and relative intensities.

Next, extracted ion chromatograms (XIC) are collected at the MS2 level for all transitions and a score is computed to each peak group. Statistical validation of peptide identifications is performed *via* the "target-decoy" approach through the calculation of the false positive rate (FDR). This strategy was explained in the bibliographic introduction, chapter 1, section 2.D.ii. Several software packages use this strategy such as OpenSWATH²³⁶, PeakView (Sciex), DIA-NN²³³, Skyline¹⁶⁸ and Spectronaut (Biognosys). In the latter two, the mProphet²³⁷ algorithm is implemented and the scoring model is trained for each data set.

Furthermore, algorithms have been developed, such as Transfer of Identification Confidence²³⁸ (TRIC) or DIALignR²³⁹, to normalize the retention time between analyses. These algorithms are used to improve the robustness of identifications and limit the proportion of incorrect identifications. Others, such as Mobi-DIK²¹⁵ or Spectronaut (Biognosys), exploit technological advances in mass spectrometry, e.g. the use of ion mobility as an additional separation dimension on HR/AM instruments.

iii. Direct spectrum matches

Another strategy for extracting DIA data consists in directly comparing the multiplexed MS2 spectra acquired in DIA with those theoretically assigned and contained in a spectral library or in a data bank. Thus, contrary to the previous approach, this extraction is performed in a non-targeted way.

A first example, the ProbiDtree²⁴⁰ software includes an algorithm that identifies all potential precursor ions contained in an isolation window using the corresponding MS1 spectrum. The program calculates a probability score on the identification of each peptide, and then these are distributed in a peptide tree. At each step, a new DIA-MS2 pseudo-spectrum is generated by removing the already matched fragments. A second software, MSPLIT-DIA²⁴¹, tries to demultiplex the DIA-MS2 spectra by evaluating the similarities between these spectra and the MS2 spectra contained in the spectral library. It includes the removal of spectra too similar for targeted extraction, a combination of scores as a function of retention time and a FDR control. As final example, the EncyclopeDIA²⁴² software is based on the PECAN⁴⁶ algorithm, which uses a spectral library generated from multiple injections of the sample in DIA *via* narrow isolation windows (4 m/z).

B. Spectrum-centric approach

Spectrum-centric approaches are based on the generation of pseudo MS2 spectra from DIA-MS2 spectra. These pseudo-spectra are then subjected to a database search via classical search engines. This approach was introduced in 2003 by Purvine et al.¹⁹⁸. From analyses performed at low and high voltages, they reconstructed pseudo DDA spectra based on the similar chromatographic characteristics of precursor and fragment ions to identify them manually.

Since then, several spectrum-centric algorithms have been developed in parallel with the creation of new DIA acquisition methods^{44-45, 220}. Among them, the DIA-Umpire⁴⁷ software, allows the non-targeted identification and quantification of peptides. It also offers the possibility to generate a spectral library from the identification results and then perform a peptide-centric extraction of the initial DIA data. More recently, with the implementation of the Pulsar search engine in the Spectronaut software, Biognosys has integrated a spectrum-centric algorithm, called directDIA™. Finally, the DIA-NN²³³ software also offers the possibility to perform DIA data extraction with both peptide- and spectra-centric approaches.

Chapter 5

The concern of Host Cell Proteins in biopharmaceutical companies

1. Monoclonal antibody

Monoclonal antibodies (mAbs) became a major class of human therapeutics mainly because of their high specificity to target diverse molecules or antigens and their various modes of action³³⁻³⁴. The therapeutic effect of mAbs includes the neutralization of soluble antigen, blocking or stimulation of intracellular signal pathway, activation of cellular and complement-mediated mechanisms (ADCC, ADCP and CDC) and target delivery of various components^{34, 243}. Muromonab-CD3 (trade name: Orthoclone OKT3) was the first marketed therapeutic antibody in 1986. It was used as immunosuppressive agent during organ transplantation. Since then, the US and European agencies (FDA and EMA) responsible for pharmacovigilance have approved more than 90 mAbs with a current sales market of over \$100 billion^{2, 32}.

A. Expression system

In 1975, Köhler and Milstein were the first to report the *in vitro* production of murine mAbs using hybridoma technology²⁴⁴. Nine years later, they received the Noble Prize of Physiology and Medicine for their work that revolutionized the diagnostic and therapeutic use of antibodies. Hybridoma technology, for the continuous production of mAbs, is based on cell fusion. First, an antigen is injected into a mouse to induce an immune response. B-lymphocytes are then isolated and fused with myeloma cells. The resulting hybrid cell, called hybridoma, combines the ability to secrete specific antibody from B cells with the immortality from the myeloma cells. However, murine mAbs have limited therapeutic use due to side effects reported by increasing the human anti-mouse antibody response²⁴⁵. In addition, they display a relatively short half-life in humans compared to human Immunoglobulins G (IgG)²⁴⁶. Two parameters that may be critical to their efficacy, especially in oncology. Advances in molecular biology and genetic manipulation techniques enabled the production of chimeric²⁴⁷, then humanized²⁴⁸, or even human mAbs, in order to extend the half-life in humans and to reduce the immunogenicity towards the mAb product. In order to improve their yield, mAbs are produced in a variety of expression systems, ranging from yeast, bacteria, fungi, insects, mammalian cell lines to transgenic plants and animals²⁴⁹. However, the large majority of currently approved mAbs are produced in mammalian host cells due to their abilities to introduce post-translational modifications analogous to those in humans²⁵⁰. Chinese hamster ovary (CHO) cells are the most frequently used host cell system for the industrial manufacture of recombinant therapeutic proteins due to several key advantages over other cell types, including: (i) robust growth in chemically-defined and serum-free suspension cultures, (ii) powerful gene amplification systems that increase their productivity, (iii) the ability to express mAbs with human post-translational modifications. CHO cells have been demonstrated to be a safe host, which facilitates approval by regulatory agencies²⁵¹⁻²⁵².

B. Manufacturing process

The mAb manufacturing process is divided in two steps: a cell culture process in bioreactors (Upstream Process, USP) followed by cells removal and protein purification (Downstream Process, DSP) (Figure 38).

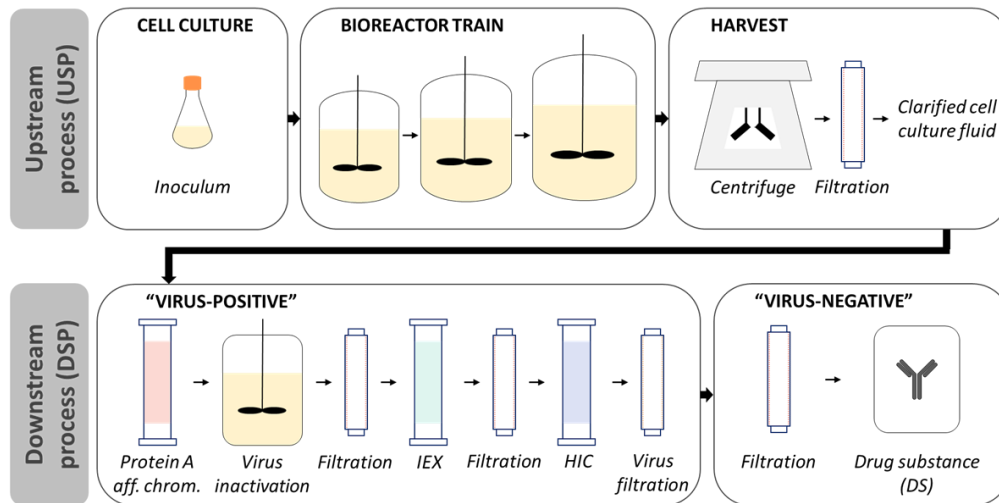


Figure 38: Monoclonal Antibodies manufacturing process (adapted from ²⁵³⁻²⁵⁴). HIC: Hydrophobic Interaction Chromatography. IEX: Ion Exchange Chromatography. Protein A aff. Chrom.: Protein A affinity chromatography.

i. Upstream process

First a stable, robust and high yielding cell clone is selected. Then, it is expanded and several hundred cell vials are stored at -180°C in various locations to form a stock of cells to be used only when needed. This stock is called the Master Cell Bank (MCB). The MCB is thoroughly characterized for identity, purity and stability. Cells from a vial of MCB will be grown for multiple passages and hundreds of aliquots will again be stored at -180°C to form the Working Cell Bank (WCB). The WCB cells are used for production. The first step of the mAb manufacturing process is the thawing of a vial of cells from the WCB. The cells are then grown in a small volume (50 ml) and expanded to the volume of a production bioreactor (500 to 20 000 L). This expansion phase is followed by a production phase in which the mammalian cells secrete mAb in the culture media until typical titers of 1 g/L in batch and 1-10 g/L in fed-batch processes²⁵⁵. Next, the cell culture fluid (CCF) is centrifuged and filtered (microfiltration) to remove cells and cell debris²⁵⁶. The resulting clarified cell culture fluid (CCF) is the final step in USP.

ii. Downstream process

During the DSP, a succession of purification steps using chromatographic systems, viral elimination and multiple filtrations will allow to obtain a concentrated product that reaches the desired specifications²⁵⁷⁻²⁵⁸.

DSP commonly begins with a protein A affinity chromatography step. This step, also called capture step, aims to remove the majority of impurities from the raw harvest material²⁵⁹. Protein A was initially found in the cell wall of the bacterium *Staphylococcus aureus*. Its natural high affinity to the Fc region

of immunoglobulin G (IgG) from various species was first described in 1958²⁶⁰. Since then, modified versions of protein A have shown increased stability and binding capacity²⁶¹. Thereafter, mAbs undergo up to three chromatographic steps, also known as polishing steps. These include viral clearance (inactivation and filtration), and filtration steps to concentrate the product (ultrafiltration) or remove buffer components (diafiltration)^{254, 257}. The final filtration step is designed to concentrate the mAb product in a buffer to allow formulation (i.e., addition of an excipient) and packaging (e.g., lyophilization). Impurities, particularly HCP and DNA, must be monitored throughout the process²⁶², and typical purity goals are <100 ppm for HCP (<100 ng HCP/mg mAb), and <10 ng/dose for DNA²⁶³.

2. Host cell proteins monitoring

A crucial step in the biopharmaceuticals manufacturing process involves the removal of Host Cell Proteins (HCPs). These impurities are included in the Critical Quality Attributes (CQAs) risk assessment as they can affect the product efficacy with protease activity and patient's safety by inducing immunogenic reactions⁵⁻⁶. Guidelines state classically that the acceptable HCP amount in the final drug product should be below 100 ng/mg^{7, 35}. However, this level is only a recommendation because problems with HCPs arise from specific proteins rather than the overall amount³⁶⁻³⁸. On top of these issues, the HCP profile can be affected by various USP decisions (cell culture duration, feeding strategies or culture temperature)²⁶⁴, the increasing production to commercial scale²⁵⁸ or the biopharmaceutical itself which can impact the co-elution of HCPs during DSP^{7, 265}. All these factors highlight the need to monitor HCPs throughout the manufacturing process in order to understand the impact of a given bioprocess development. Therefore, analytical methods capable of covering the wide dynamic range between mAb and impurities (about 5 to 6 order of magnitude) are needed (Figure 39)^{4, 11}. Several detection methods for HCPs are available. They can be classified as immuno-specific (Western-Blot and ELISA) or non-specific (electrophoresis and mass spectrometry) methods^{35, 39}.

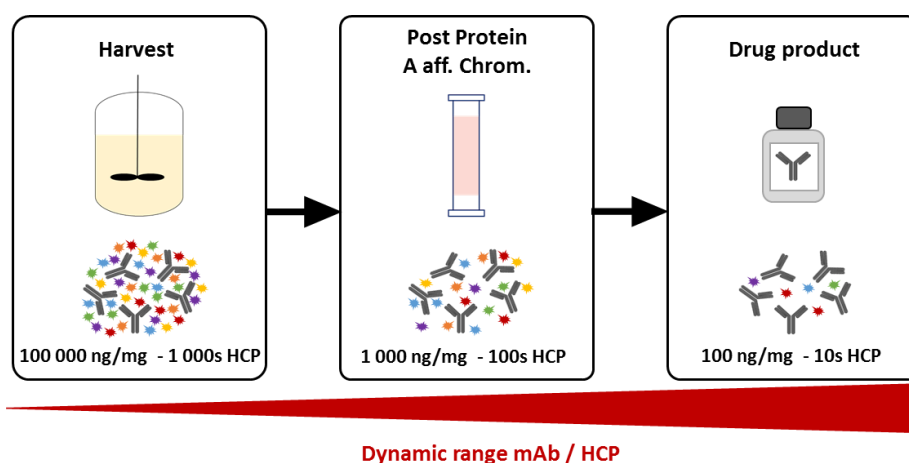


Figure 39: Host Cell proteins clearance during purification process (adapted from ⁴). Order of magnitude of the overall impurities quantity (ng of HCP per mg of mAb) and number of HCPs at different step of the mAb manufacturing process. Protein A aff. Chrom.: Protein A affinity chromatography.

A. Immuno-specific methods

Immuno-specific methods detect HCPs using polyclonal anti-HCP antibodies. These antibodies are usually produced in rabbits or goats by repeated injections of the HCP mixtures of interest. The choice of these HCP pools is crucial, as it will determine the spectrum of HCPs that will be detected by the anti-HCP antibodies. In order to avoid the production of anti-mAb antibodies, HCP mixtures are usually produced using a null version of the host cell line, i.e., a mock transfected cell line. This approach is based on the assumption that the HCP contents of the null cell line and the mAb-producing cell line are similar^{8, 35, 39}.

With the strong interest in immunospecific strategies, anti-HCP antibodies are commercially available. They are generally produced from cell lysates or culture supernatants of multiple null cell strains. Assays relying on these anti-HCP antibodies are called generic assays. Easy and quick to implement, they are capable of detecting a broad spectrum of HCPs from a variety of cell strains and processing conditions. However, they display a low specificity in purified products and even in crude samples (30%)^{39, 266}. To overcome this limitation, process-specific anti-HCP antibodies can be generated in-house using specific material from the cell line employed (upstream process specific), or from the manufacturing process (downstream process specific). It will increase the specificity of the assays but still face issues because new HCPs resulting from a change in the manufacturing process will not be detected.

Finally, no anti-HCP antibodies reagent can cover the entire spectrum of HCPs resulting from a specific manufacturing process. The method will be limited to the HCPs that induce an immune reaction in the organism used to generate the anti-HCP antibodies. Moreover, the development of a process-specific immunoassay is time consuming and costly²⁶⁷. The anti-HCP antibodies are limited reagents and will need to be reproduced.

i. ELISA

The enzyme-linked immunosorbent assay (ELISA) is routinely used as a diagnostic tool in the fields of medicine and biotechnology²⁶⁸⁻²⁷⁰. It is currently the gold standard method to measure the amount of HCPs during the manufacturing process of mAbs, from process development to final product formulation. This method displays a high sensitivity (down to ng HCP/mg mAb), high specificity and high throughput that help to monitor HCP amounts in biopharmaceutical companies^{8, 39}.

There are several types of ELISAs, but the most widely used for HCP monitoring is the sandwich ELISA. Its name comes from the use of two antibodies to target and detect antigens in a sandwich manner, as described in Figure 40.

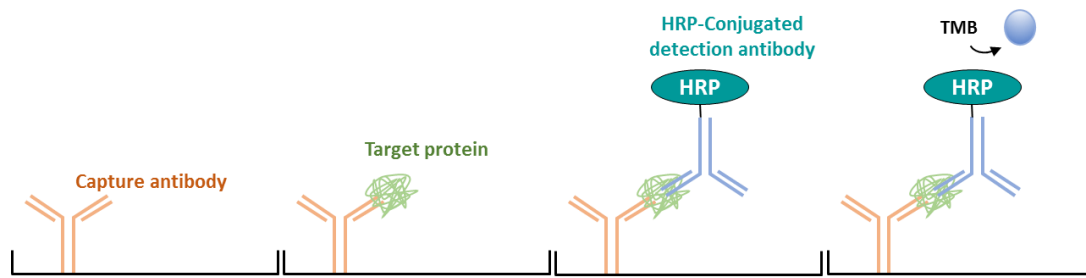


Figure 40: Principle of the sandwich ELISA assay

First, capture antibodies specific to the target protein are coated onto a multi-well plate. After incubation and removal of the unbound antibodies, the remaining protein binding sites are blocked by incubation with bovine serum albumin (BSA) or non-fat dry milk, for instance. The purpose is to prevent further non-specific binding onto the wells. Following washing, the samples are added and the target protein binds to the immobilized capture antibodies. After incubation, unbound target proteins are washed away. The detection antibodies are conjugated to an enzyme (e.g. horseradish peroxidase (HRP)) and are specific to another epitope of the target protein. The use of these antibodies allows simultaneous binding of the capture and detection antibodies to the target protein, which is necessary to detect the protein of interest. After addition and incubation of the detection antibodies, unbound reagents are removed by washing. Finally, the substrate (e.g., TMB) is added and converted by the enzyme. The product is then quantified by measuring its absorbance with a spectrophotometer. The concentration of the antigen is estimated using calibration curves derived from standard sample of known concentration.

The ELISA for HCP is designed to quantify a large number of proteins. Usually, the same anti-HCP antibodies are used for binding and detection of HCPs. However, the binding antibodies are either applied directly to the well or conjugated to biotin to improve their binding to streptavidin-coated plates and enhance the sensitivity of the assay²⁷¹⁻²⁷². On the other hand, detection antibodies are conjugated to an enzyme, most often HRP²⁷³⁻²⁷⁴. In order to be detected, the simultaneous binding of both antibodies to the HCP must be sterically possible. Steric hindrance, limited number or lack of accessible binding epitope will lead to the non-detection of HCPs²⁶⁷.

However, the limitation of ELISA come from low or non-immunogenic species that will not be covered by the anti-HCP antibody pool. Another limitation, which highlights the need for orthogonal methods, is that ELISA assay give a global HCP amount without individual identification^{8,40}.

ii. Western-blot

Western blot is another immuno-specific approach that is routinely used in various fields of scientific research to detect specific proteins from a complex sample²⁷⁵⁻²⁷⁶. The method is based on three steps described in Figure 41. First, separation of the proteins according to their size by gel electrophoresis. Second, the transfer of the proteins onto a membrane, the actual western blot step. Finally, the detection of the target protein using specific antibodies.

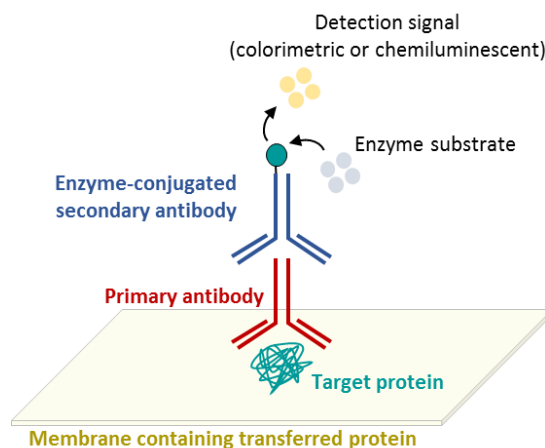


Figure 41: Principle of protein detection by Western blot

To summarize, the membrane is incubated with a blocking solution (e.g., BSA or dry nonfat milk) to prevent nonspecific binding of the antibodies to the membrane. After a washing step, the membrane is incubated with primary antibodies that are specific to the target protein. Following the incubation, the unbound primary antibodies are washed away. Secondary antibodies, which are conjugated to an enzyme, specifically target the primary antibodies. After incubation and washing, the substrate is added and the enzyme generates a compound that is detected using a spectrophotometer.

As well as ELISA, western blot is used to detect a large number of HCPs but it is usually used to support the development of ELISA. Indeed, this technique is used to assess the coverage of anti-HCP antibodies during the animal immunization process. The HCP fraction detected after the western blot will be compared to the global HCP profile visualized using 2D gels. This comparison can be performed at different stages of the manufacturing process to assess the coverage of anti-HCP antibodies in purified samples. Finally, the western blot remains an immune-specific method that suffers from the same limitation of ELISA by providing incomplete coverage of HCPs.

B. Non-specific methods

Non-specific methods are also available to characterize HCP profiles. It is recommended to use these orthogonal methods as they allow detection of non-immunogenic proteins and complement immuno-specific methods for a comprehensive HCP monitoring.

i. Gel electrophoresis

Polyacrylamide gel electrophoresis (PAGE) was described in the sample preparation steps to follow prior a LC-MS/MS analysis (Chapter 1, Section 2.A.ii). The technique consists in the separation of proteins according to their size. The 2D-PAGE method adds an isoelectric focusing (IEF) step before protein separation by molecular weight. IEF enables the migration of proteins according to their isoelectric point in a polyacrylamide gel strip containing an immobilized pH gradient²⁷⁷. Once they have migrated, proteins are fixed in the gel and stained with global dyes (Coomassie blue²⁷⁸ or silver stain²⁷⁹) to allow global protein profiling.

Nowadays, 2D PAGE and differential gel electrophoresis (2D-DIGE) are widely used to monitor HCPs during manufacturing process development²⁸⁰⁻²⁸¹. Both techniques are robust and provide visual mapping of the HCP population along with their molecular weight and isoelectric point. In addition, they also allow visualization of several post-translational modifications (PTMs)²⁸², such as phosphorylation. 2D PAGE does not allow the identification of HCPs and needs to be combined with mass spectrometry (MS) to provide the information^{277, 282}.

The main drawback of these methods is their limited dynamic range. Indeed, 2D-PAGE will display the most abundant proteins of the studied sample. Moreover, some proteins are difficult to analyze. This is the case for small or very large proteins, extremely basic or acidic proteins or hydrophobic proteins²⁷⁷. In the field of HCP, the limited dynamic range is a major issue. For instance, the presence of the overwhelming mAb heavy and light chains can hide the presence of low abundance HCPs (Figure 42).

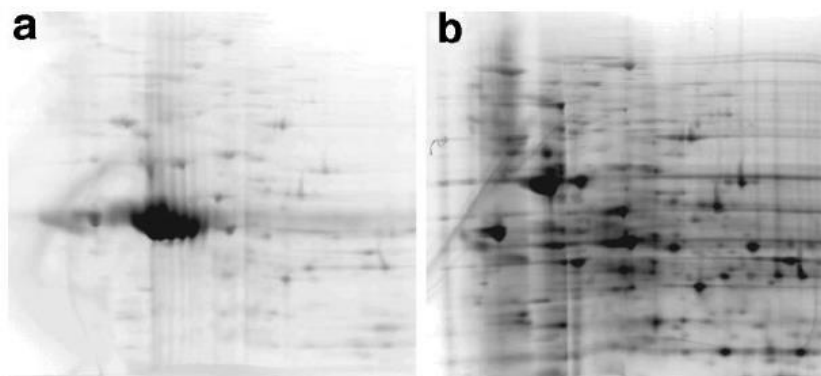


Figure 42: Limitation of the 2D-PAGE method for the detection of HCP (from ²⁸¹). 2D gel of a CCCF fraction containing the mAb product (A). 2D gel of a CCCF fraction without the mAb product that was removed by affinity chromatography.

Among the analytical toolbox to characterize HCPs, the non-specific methods based on polyacrylamide gel can be used conjointly with ELISA to evaluate the dynamics of the HCP content^{281, 283}. However, they are not suitable for high throughput and display a lack of reproducibility^{39, 277}. To overcome this limitation, 2D-DIGE approach can be used to analyze up to five samples on the same gel using different fluorescent dyes (cyanine dyes). However, the labeling step of 2D-DIGE slightly alters the physical properties of the proteins such as their solubility, hydrophobicity and size³⁹.

ii. Mass spectrometry

In this context, Mass Spectrometry (MS) became the most promising alternative to characterize HCP allowing risk assessment with individual HCP identification and unbiased quantification. Recent advances in MS, in particular the use of MS2 signals for quantification by targeted or DIA methods, have resulted in a 2- to 8-fold gain in sensitivity¹⁴ as well as a significant gain in specificity and dynamic range compared to the use of MS1 signals. These features are crucial in the field of HCP.

The targeted strategy using SRM coupled to isotopic dilution has been the gold standard MS-based technique enabling to achieve the highest sensitivity, quantification robustness and accuracy^{12, 37, 43}. However, the development of a targeted quantification method is time consuming and still limited to the selection of about hundred candidates. Besides, DIA performs the acquisition of MS2 signals from all detectable species allowing to obtain the complete proteome map while reaching the sensitivity, quantification accuracy and robustness of targeted methods¹⁴⁻¹⁵.

Tandem MS coupled to liquid chromatography (LC-MS/MS) studies have been conducted to validate the HCP coverage of ELISA assays²⁸⁴, to evaluate risk due to HCPs^{37, 48} or to support process developments^{4, 285}. Unfortunately, despite the advances in data acquisition and extraction, the analytical challenge remains in the dynamic range mAb/HCP. The ubiquity of the antibody may interfere with the reliable extraction of HCP peptides and lead to biases in the MS-based quantification of these impurities. Few studies have attempted to decrease this dynamic range by using multi-dimensional chromatography^{11, 27-28, 48-49} or mAb depletion^{29, 50-51}.

The main limitations of MS are its cost and the need for a highly trained operator. In addition, the lack of a high quality, publicly available CHO protein sequence database is another issue²⁸⁶, as most of the approved mAbs are produced from CHO cells.

Conclusion

Biopharmaceutical companies rely on the ELISA method to monitor HCP during the mAb manufacturing process. This method is accepted by the regulatory agencies. It has a high sensitivity (up to ng HCP/mg mAb), high specificity and high throughput that allow the assessment of the HCP content in samples with different purity levels (from crude harvest to the final drug product)^{8, 39}. Unfortunately, this technique has several limitations as a global quantification result without individual identification or an incomplete coverage of the HCPs in the drug product^{8, 40}. As the immunogenic risk or product degradation are linked to particular HCPs and not necessarily to their quantity, these drawbacks raised an important need of alternative orthogonal analytical methods. Table 5 summarizes the available methods for HCP monitoring with their detection limits, pros and cons.

Method		Sensitivity	Advantages	Drawbacks
Immuno-specific	ELISA	Total HCP 1-100 ppm	High throughput, sensitivity, specificity	Detects only immunogenic HCP ≥ 2 antibodies / HCP Total HCP amount Development costly and time consuming
	Western blot	Individual HCP 20-200 ppm	MW and pI Visible PTM	Detects only immunogenic HCP Development costly and time consuming Labor intensive
Non-specific	2D-PAGE (cyanide dye)	Individual HCP 8 ppm	MW and pI Visible PTM MS-compatible	Low dynamic range HCP hidden by the mAb product
	MS	Individual HCP 1-10 ppm	HCP identification and quantification	No high quality CHO protein database High skilled operator Expensive equipment

Table 5: Summary of the HCP monitoring approaches (adapted from ^{7, 39}).

In this context, mass spectrometry, and more particularly the bottom-up proteomic analysis, has emerged as a key tool. From a LC-MS/MS analysis, it is possible to perform a risk assessment of HCPs through the identification and quantification of protein impurities present in the sample. Through this bibliographic introduction, it is easy to realize the analytical arsenal available in MS-based quantitative proteomics. All of these MS-based approaches will have strengths and weaknesses (Table 6). It is of course essential to consider these in order to choose the appropriate strategy to answer the question at hand. Indeed, different methods will be used if the goal of the analysis is to quantify a small number of problematic HCPs or if the overall HCP content needs to be assessed.

Criteria	DDA	SRM/PRM	DIA
Method development	++	-	+
Data analysis	++	+	-
Hypothesis-driven?	-	+	+/-
Proteome coverage	+	-	+
Sensitivity, dynamic range and selectivity	-	++	+
Quantification precision	+/-	++	+
Reproducibility	-	++	+
Retrospective hypothesis?	+	+	++

Table 6: Summary of the Data-Dependent Acquisition (DDA), Selected/Parallel Reaction Monitoring (SRM/PRM) and Data-Independent Acquisition (DIA) strengths and weaknesses (adapted from ^{220, 287}).

The main objective of my PhD was to develop MS-based quantification methods for HCPs profiling in order to support process development, manufacturing and final purity assessment of the mAbs. The developed approaches should therefore allow an accurate and robust quantification of HCPs while covering the largest possible spectrum of HCPs present.

**Part I: Development of an accurate HCP
quantification method to support mAb
manufacturing**

Chapter 1

Development of an innovative Top3 quantification strategy dedicated to HCP profiling

1. Context of the project

During his PhD, Dr. Gauthier Husson developed an innovative method based on mass spectrometry to characterize Host Cell Proteins impurities in biopharmaceuticals²⁸⁸. The Top3-ID-DIA method combines a Top3 quantification strategy¹⁶ for global HCP profiling and isotopic dilution (ID) to accurately quantify key HCPs. Among the quantitative proteomics approaches, the Top3 strategy allows the quantification of detected species without the use of costly labelled standards or supplementary sample preparation to labelled the HCPs. It relies on the assumption that the sum of the MS response of the three best responding peptides per mole of protein is constant within a coefficient of variation of less than 10% to estimate absolute amounts of HCPs¹⁶. An internal standard is used to calculate a signal response factor (TOP3 peptides signal/mol) and to perform the HCP quantity estimation (HCP TOP3 peptides signal/signal response factor). The use of a mixture of four standard proteins (Mix 4P) has been previously reported for this purpose^{10, 17}, considering the PYGM protein as a reference and the three other proteins (ADH, BSA, and ENL) to calculate ratios, as internal controls. In recent years, this approach has been used in different ways. The use of three, five or seven standard proteins has been reported^{29, 48, 289}. Some methods have been developed using a single reference protein while others are based on an average amount calculated from the quantity obtained by each standard proteins. The strategies developed in this manuscript for the quantification of HCPs are intended to be used in regulated biopharmaceutical environments. Therefore, a standardization of the Top3 quantification approach should be considered, which requires the development of dedicated standards.

In this context, we choose to confront our current quantification workflow against the HCP Profiler standard, an original mixture of peptides coated on a water-soluble bead releasing controlled amounts of the peptides after solubilization (READYBEADS technology, from Anaquant)¹⁸. ANAQUANT, the French company that developed the standard, aimed to provide a ready-to-use kit for global profiling of HCPs. An objective perfectly in line with the need for standardization in the field.

First, we started a collaboration with ANAQUANT. Together we conducted an inter-laboratory study in order to assess the reproducibility of the solution. The HCP Profiler standard was solubilized in a HeLa (Human cervical cancer cells) cells total proteins digest prior to injections run in DDA mode on two Q-Orbitrap instruments, one in their laboratory in Lyon and one in our laboratory in Strasbourg. Thereafter, the ready-to-use kit was evaluated in a matrix containing HCPs.

We therefore used the sample set produced by Gauthier Husson in collaboration with UCB Pharma and the University College London (UCL), including two levels of HCP complexity. Two cell culture durations (7 and 10 days), three harvest procedures (no shear, low shear or high shear) and two protein A purification protocols were investigated resulting in four HCP-rich Clarified Cell Culture Fluids (CCCF) and seven purified Post Protein A (PPA) fractions (Figure 43).

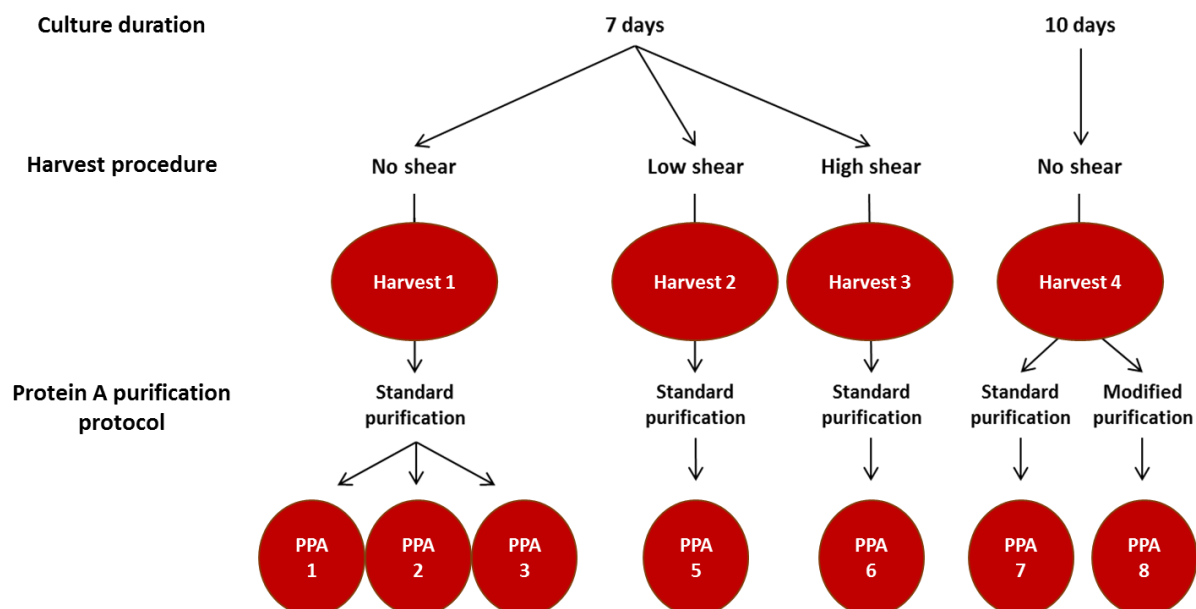


Figure 43: mAb samples production workflow (adapted from ²⁸⁸). Four IgG4 A33 mAb samples were collected after 7 or 10 days of cell culture followed by different shear stress conditions during harvest using an Ultra Scale-Down (USD) shear device²⁹⁰⁻²⁹¹. Finally, two protein A purification protocols were applied, a standard and a modified protocol, resulting in seven Post Protein A (PPA) fractions.

The CCCF and PPA samples, produced from CHO cells, were spiked with the water-soluble bead or the mix of four standard proteins classically used in our laboratory. Both methods were benchmarked using DDA acquisitions on a nanoLC-Q Exactive HF-X instrument (Thermo Fischer Scientific). The global and individual HCP quantities were compared in order to assess the accuracy and reproducibility of the HCP Profiler standard versus the use of the Mix 4P.

2. Experimental design

The analytical workflow presented in Figure 44 shows the two experiments used to assess the HCP Profiler kit's performances.

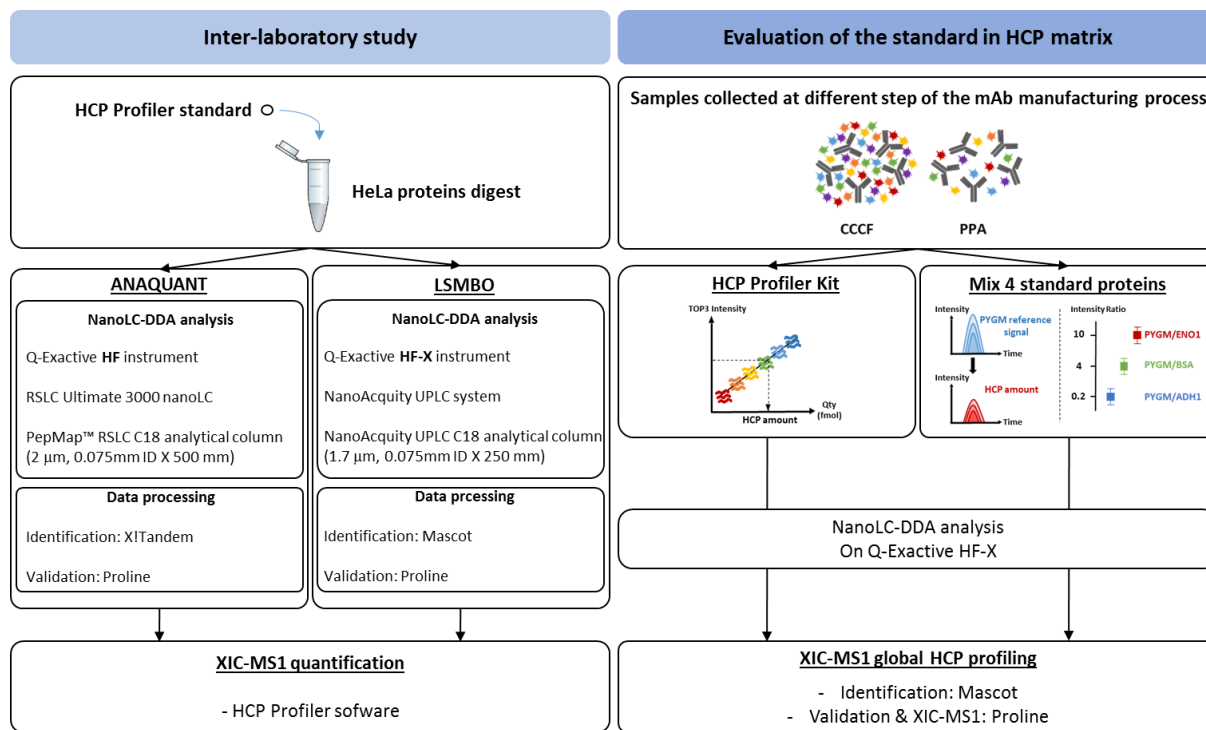


Figure 44: Analytical workflow used to evaluate the HCP Profiler standard. An inter-laboratory study of the HCP Profiler standard was performed on a HeLa digest. A sample containing the HCP Profiler peptides in a range from 1 to 500 fmol was injected in triplicate using DDA on two different nanoLC-MS/MS platforms in two different labs. Both datasets were compared using XIC-MS1 quantification on Proline software (Left). Four CCCF and eleven PPA fractions were in-gel digested and spiked with the HCP profiler standard or the mix of 4 standard proteins (ADH, PYGM, BSA, ENL from MassPREP digestion standard kit, Waters). Samples were injected in triplicate using DDA acquisition mode on a nanoLC-Q Exactive HF-X (Thermo fischer Scientific) system. Both accurate MS-based TOP3 quantification methods were benchmarked using XIC-MS1 DDA quantification on Proline software (Right).

We started the evaluation of the standard with an inter-laboratory study. One bead was solubilized in HeLa proteins digest and the samples were analyzed using DDA on two different nanoLC-MS/MS platforms. Three injection replicates of 400 ng of HeLa proteins with the HCP Profiler peptides in a range from 1 to 500 fmol were performed on Q-Orbitrap instruments at the Anaquant laboratory (Lyon) and the LSMBO (Strasbourg). Peptides and proteins were identified using X!Tandem (Anaquant) or Mascot (LSMBO) database searches and further validated using Proline¹⁹. After XIC-MS1 quantification by Proline, the data was loaded on the HCP Profiler software that performs the selection of the HCP Profiler peptides to determine the calibration curve. Finally, the software developed by Anaquant calculates each HCP's amount.

The second experiment involves the samples produced by Dr. Gauthier Husson, collected after a harvest procedure (4 samples) and after Protein A affinity chromatography (7 samples) as described in [288]. CCCF and PPA samples were stacked in a single band. After in-gel trypsin digestion, samples were split and spiked either with a water-soluble bead or with the mix of 4 standard proteins (ADH, PYGM, BSA and ENL from Waters MassPREP Digestion Standard Kit) for HCP quantification. In all samples, retention time standards (iRT kit, Biognosys) were spiked. Finally, Samples were analyzed in triplicate on a nanoLC-Q Exactive HF-X (Thermo Fisher Scientific) in DDA mode.

Peptides and proteins were identified using Mascot database search engine. The validation of the dataset as well as the XIC-MS1 quantification were performed by Proline¹⁹. Then, a Top3 strategy¹⁶ was applied on precursors showing a coefficient of variation below 20%. For the HCP Profiler kit, a calibration curve of the $\log_2(\text{TOP3 standard peptides abundance})$ in function of $\log_2(\text{standard proteins quantity})$ is obtained and allowed us to estimate protein mol quantities. Using the Mix 4P standard, the 2 fmol of PYGM were used as a universal response factor (Top3 peptides signal/mol) to perform the HCP quantity estimation (HCP Top3 peptides signal/signal response factor).

Sample preparation, nanoLC-MS/MS methods and data treatment are detailed in the experimental part, section 1.A.

3. Inter-laboratory study

In collaboration with ANAQUANT, we conducted an inter-laboratory study to evaluate the performance of their innovative standard. The HCP Profiler standard is a water-soluble bead, which releases unlabeled peptides at known amount. The quantity estimation of the HCPs is obtained from the MS response of 54 peptides gathered in 6 calibration points in a range from 1 to 500 fmol. The 54 peptides come from 18 proteins (tripeptides arrangements synthesized from *E.Coli*) and each calibration point contains 3 HCP Profiler proteins. Thus, the 54 standard peptides allows to obtain a calibration curve of the $\log_2(\text{TOP3 HCP Profiler peptides area})$ as function of the $\log_2(\text{HCP Profiler proteins quantity (fmol)})$. First, we focused on the calibration curves obtained in a mixture composed of thousands of proteins (Figure 45).

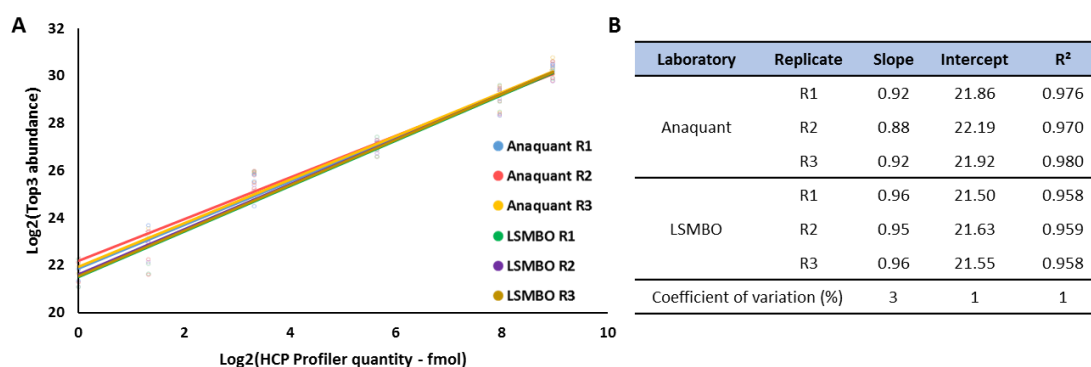


Figure 45: Representation of the HCP Profiler-based calibration curves obtained by the two laboratories. Superposition of the curves obtained from XIC-MS1 DDA quantification on a Q Exactive HF (Anaquant) and a Q Exactive HF-X (LSMBO) (A). Calibration curve equations for each replicate and CVs calculated on the slope, intercept and regression coefficient (B).

Even with differences in chromatographic systems, mass spectrometers and database search engines, the calibration curves showed significant repeatability (Figure 45.A). The coefficients of variation (CVs) calculated on the slopes, intercepts and regression coefficients are below 3% (Figure 45.B). This reproducibility is an important criterion that will allow batch-to-batch comparisons or support inter-production sites investigations.

Following these positive results, we compared the overall and individual protein amounts provided by the HCP Profiler standard (Figure 46).

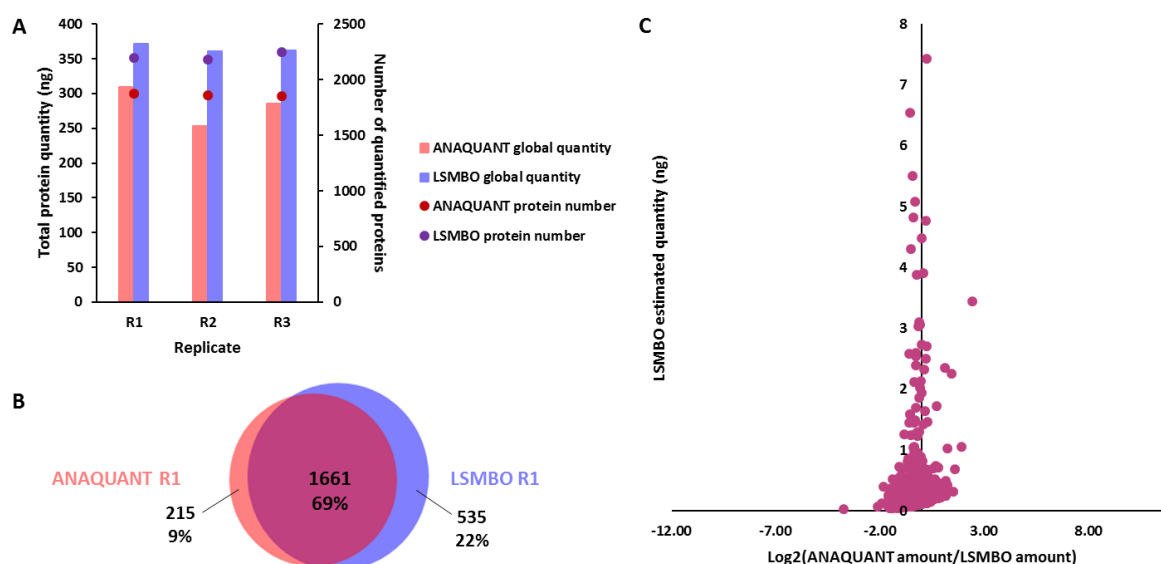


Figure 46: Inter-laboratory global and individual quantification results obtained using the HCP Profiler standard. Representation of the global amount and number of quantified HeLa proteins obtained per replicate (A) Venn diagram of HeLa proteins quantified in replicate 1 for both laboratories (B) Inter-laboratory comparison between the individual quantification of the 500 most abundant proteins (C).

First, the overall amounts were obtained by excluding proteins detected out of the range (1 to 500 fmol) of the HCP Profiler proteins. These out-of-range proteins represent a total of 25 to 50 ng and around 40 ng, respectively in Anaquant and LSMBO data. Even without these proteins, the global estimations are close to the 400 ng of HeLa proteins injected (Figure 46.A). It can be noticed that higher amounts and number of quantified proteins were obtained by LSMBO. These differences might be due to the use of the Q Exactive HF-X, which is more sensitive than its predecessor²¹, the Q Exactive HF, used in the Anaquant laboratory. On average, 344 additional HeLa proteins could be quantified with the Q Exactive HF-X. An increase that still displays a great correlation with about 69% of common proteins (1661 proteins) between Anaquant and LSMBO data (Figure 46.B).

Finally, we compared the individual amounts estimated for the 500 most abundant proteins common to both datasets as these proteins represent about 78% of the global quantity. Figure 46.C shows the ratios calculated between Anaquant and LSMBO amounts. The log of a ratio at 0 means that the same amount was obtained by both laboratories. We can observe that the variation between quantities is low for the most abundant proteins and is more diffused for the less abundant ones. At least 88% of the Top3 estimations provided by the HCP Profiler standard are consistent within a factor 2. These results are in line with previous studies that demonstrated the variability of the Top3 strategy^{9, 20}.

As a conclusion, the inter-laboratory study revealed high reproducibility cross laboratory usability of the HCP Profiler standard. The use of different nanoLC-MS/MS platforms and database search engines even strengthens the investigation results. It demonstrates the reproducibility of the standard regardless of the analytical conditions. Therefore, the HCP Profiler kit can be an efficient solution to characterize HCPs on different production sites. In order to investigate the suitability of the standard to support process developments, we performed a second experiment using real HCP matrices.

4. HCP Profiler standard to support process development

The HCP sample set used includes two levels of HCP complexity and had been produced in 2015 to investigate bioprocess developments by MS. We re-used the four CCCF and seven PPA samples to benchmark the HCP Profiler standard against our current workflow that involves a mixture of 4 standard proteins PYGM, ADH, BSA and ENL (Mix 4P), from the MassPREP digestion standards (Waters). Both MS-based Top3 quantification standards were spiked after digestion and peptides extraction from the gel bands. Samples were injected in triplicate in DDA mode on a Q-Exactive HF-X instrument.

First, the 33 HCP Profiler-based calibration curves showed coefficients of variation (CVs) calculated on the slopes, intercepts and regression coefficients below 3%. The curves obtained in this experiment support the reproducibility of the standard observed during the inter-laboratory study. Without filters applied to the standard peptide signals, the reproducibility was observed among CCCF and PPA fractions without differences related to the matrix complexity. Next, the overall HCP amounts as well as the number of HCPs quantified were benchmarked for both Top3 strategies (Figure 47).

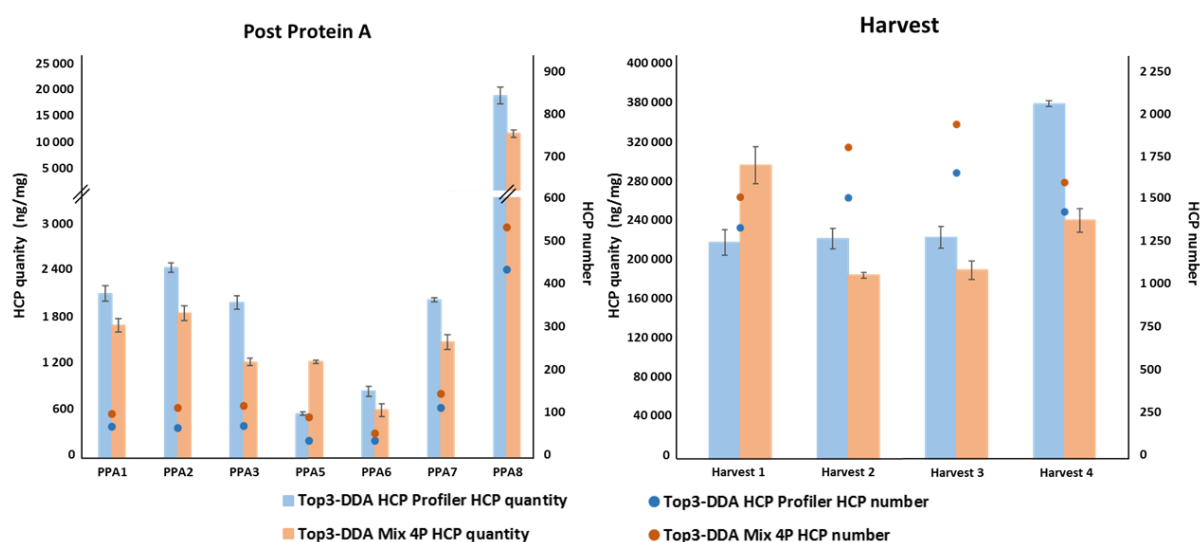


Figure 47: Benchmarking of the HCP Profiler standard against the Mix 4P. Comparison of the XIC-MS1 DDA global quantification and numbers of HCPs quantified obtained on a Q-Exactive HF-X using the HCP Profiler standard and the Mix 4P. Error bars represent the standard deviation.

On average, 1464 HCPs were quantified in the CCCF fractions with global quantities between 222 646 and 365 145 ng/mg, and 115 HCPs in PPA fractions representing 569 to 19 153 ng/mg. These results are in accordance with the global amounts obtained using the signal response factor of PYGM (Figure 47). We noticed that for almost all samples, the global HCP amounts are higher using the 6 points

calibration curves, except for PPA 5 and Harvest 1 samples. In contrast, a lower number of HCPs was quantified, an average of 34% less for PPA samples and 13% for CCCF fractions. In order to understand this observation, we compared the individual HCP amounts obtained for the protein impurities quantified using both quantification methods (Figure 48).

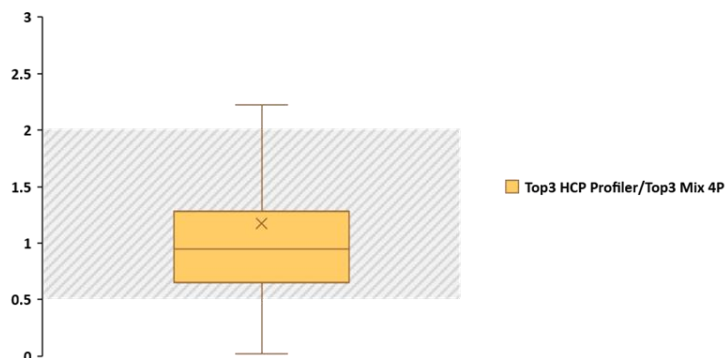


Figure 48: Comparison of the individual HCP amounts obtained using the Top3 HCP Profiler-DDA and Top3 Mix 4P-DDA strategies. Representation of the ratios calculated between the Top3 HCP Profiler-DDA and Top3 Mix 4P-DDA amounts obtained for the 5305 common HCPs identified with both acquisitions methods. Out of range ratios were removed. In total 78% of the 5305 common HCPs were observed within a factor 2 (grey area).

The ratios between the Top3 HCP Profiler over the Top3 Mix 4P were overall very consistent (Figure 48). Indeed, 78% of the 5305 ratios were in accordance within a factor 2. The median value of all ratios is 1.3 in favor of higher amounts estimated with the water-soluble bead. Indeed, the sum of increased quantities estimated at the individual level results in higher global HCP amounts. As we observed similar data extraction of common HCPs in our samples, the differences are most explainable by the way quantities are calculated. Since the ionization efficiency of peptides can vary, taking into account the MS response of 54 peptides spiked at different amounts leads to an increased quantity estimated alongside with an increased confidence about the results.

As a conclusion, the evaluation of the HCP Profiler standard in HCP matrices supports the high reproducibility observed in our previous inter-laboratory study. The difference in HCP content between PPA and CCCF fractions did not impact the signal extraction of the HCP Profiler peptides. It demonstrates the potential of this promising standard to monitor HCPs at different stages of the mAb manufacturing process and therefore support process developments. Moreover, the HCP Profiler-based Top3 quantification showed consistent results in line with previous studies^{9, 20} which increases the interest in using this innovative standard for HCP profiling.

5. Conclusion

Mass spectrometry has emerged as the major orthogonal method to ELISA for HCPs monitoring. The combination with a Top3 quantification strategy allows, in addition to individual identifications, to quantify all detectable species. Due to these advantages, a multitude of studies have been conducted highlighting a lack of standardization of the internal standards used for Top3 quantification. Most of the standards involve the preparation of a homemade mixture of proteins. A direct impact on the reproducibility can be observed as the analyst can induce biases during the sample preparation.

Consequently, The HCP Profiler standard appeared to be an efficient solution. The use of a water-soluble bead, which releases controlled amounts of unlabeled peptides, allows to avoid user-induced analytical biases. Moreover, the 6 points calibration curves obtained, covering 2.7 orders of magnitude, ensures confidence and quantification accuracy over a reasonable range. As a result, the ready-to-use kit displays an intra- and inter-laboratory reproducibility that allows batch-to-batch comparison and process development support.

The HCP Profiler standard demonstrated its ability to accurately quantify HCPs. To continue the standardization of the field, Anaquant also developed a software dedicated to the data treatment of the HCP Profiler-based XIC-MS1 quantification. The software will perform the selection of the three most intense peptides per proteins. Then, it will calculate the calibration curve and estimate HCP amounts. The advantages of the software solution are the representation of the data available (histograms, pie plots, 2D gel-like representations) and the highlight of risky HCPs. Unfortunately, the user has no control over the Top3 selection and the list of risky HCPs, which can only be modified by Anaquant for the moment. Prior to the Top3 selection, HCP peptides identified have to be filtered in order to keep highly confident and reproducible signals. This is why we choose to develop an R script dedicated to the treatment of the HCP profiler-based quantification allowing a more straightforward application of the method. The applied filters will be detailed in the next chapter.

Chapter 2

Evaluation of DIA strategies for global HCP profiling on a Q-orbitrap instrument

1. The promises of DIA-based quantification

In recent years, nanoLC-MS/MS-based studies have been conducted in the field of HCPs. Data-dependant acquisition (DDA) strategies were successfully applied allowing global HCP profiling and reliable individual quantifications down to 10 ng/mg mAb^{9, 27, 41-42}. Most of the time, Top 10 to 15 ions in a MS scan are isolated for fragmentation resulting in a limited dynamic and reproducibility. The presence of missing values and a discrimination towards the quantification of abundant proteins are major issues when the HCP impurities are present at trace levels compared to the biopharmaceutical itself. Targeted strategies (Selected Reaction Monitoring (SRM) or Parallel Reaction Monitoring (PRM)) have been developed enabling robust and accurate quantification of targeted HCPs down to the sub ng/mg mAb level^{12, 37, 43}. Nevertheless, the development of a targeted quantification method is time consuming compared to the implementation of a global DDA-type approach, and it is still limited to the selection of about hundred candidates. Consequently, the lack of throughput of these strategies does not allow for routine quantification of the maximum number of HCPs.

Over the past decade, advances in mass spectrometry have shown the potential of data-independent acquisition (DIA) on high-resolution/accurate mass (HR/AM) instruments. DIA relies on the co-isolation and co-fragmentation of all ions contained in predefined m/z windows of variable widths, to cover the entire mass range. The acquisition of MS2 signals from all detectable species allows to obtain the complete proteome map while reaching the sensitivity, quantification accuracy and robustness of targeted methods¹⁴⁻¹⁵. These advantages make DIA approaches attractive in the context of HCP monitoring. The bottleneck of the DIA strategy today is still the reliable extraction of quantitative signals. The generation of multiplexed MS2 spectra leads to the loss of link between the fragment ions and the precursor, making peptide identification challenging. Consequently, software generally used for the processing of DDA analyses are not adapted to the extraction of peptides/fragments information from DIA data. Several approaches have been developed to overcome the high complexity of DIA data. They can be divided in two categories: **peptide-centric and spectrum-centric approaches**. Both categories can also be described as library-based or library-free strategies.

The first **peptide-centric strategy** became popular in 2012 with the development by *Gillet et al.* of Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (SWATH-MS)^{14, 45}. SWATH-MS allowed the identification and quantification of peptides through targeted extraction using a spectral library. This approach requires the generation of a spectral library that will contain MS2 spectra from DDA analyses and/or deconvoluted pseudo-MS2 spectra from DIA data, with high-confidence peptide sequence assignments⁴⁴. In general, the library is composed of the precursor m/z , fragment ion m/z , relative intensity of fragment ions and normalised retention time (RT) for each precursor and its fragments. The library can also contain ion mobility information on most recent instruments²¹⁵. Then, the library is used to screen for the presence of specific targeted peptides in the DIA MS2 spectra, as it is done in targeted approaches¹⁶. Finally, the signal quality is controlled using computed scores allowing to validate peptide identification. This peptide-centric approach has been

implemented in several software, such as OpenSWATH²³⁶, Skyline²⁹², Spectronaut^{15, 211, 293} (Biognosys), PeakView (Sciex), DIA-NN²³³.

The second approach, called **spectrum-centric**, aims to overcome the limitation of the peptide-centric approach. Indeed, the generation of the spectral library requires time and ideally, the implementation of fractionation strategies of the studied proteome. Moreover, peptides absent from the library cannot be targeted by the library-based search, even if they are present in detectable amounts and will thus be missed. Alternatively, the spectrum-centric approach consists of an untargeted analysis of the DIA data. It relies on the generation of pseudo-MS2 spectra from co-eluting precursor and fragment ions which will be used to query a research using conventional search engines and protein inference tools as for DDA analysis. Since its introduction in 2003¹⁹⁸, many bioinformatics advances have allowed the development of new algorithms based on the spectrum-centric approach^{46-47, 201, 203} in line with new DIA methods^{44, 220}. The library-free approach will certainly increase the interest and applicability of DIA strategies in the future, in particular for the monitoring of HCPs⁴⁵⁻⁴⁷.

2. Context of the project

The implementation of DIA methods on a Q-TOF instrument was obvious due to its intrinsic capabilities in terms of acquisition speed and dynamic of detection. Indeed, Sciex TripleTOF instruments were the first MS able to scan at up to 100 Hz. In this context, Dr. Gauthier Husson developed the Top3-ID-DIA method that benefited from the fast acquisition speed and wide dynamic range of the TripleTOF 6600 (Sciex) to characterize HCPs.

In the meantime, Thermo has conducted major developments on their Q-Orbitrap instruments to improve both acquisition speed and intra-spectral dynamics. These improvements are featured on the Q Exactive HF-X that can achieve a scan speed of up to 40 Hz, thanks to its brighter ion source and optimized ion transfers²¹. In comparison with its predecessor, the Q Exactive HF could reach a maximum scan speed of 23 Hz. Based on these characteristics, we considered the development of DIA strategies at the beginning of my thesis on this other type of HR/AM instrument.

Using the sample series of IgG4 mAb A33 CCCF and PPA samples (Part I, Chapter 1), we applied the HCP Profiler Top3 quantification strategy combined with DIA on a Q-Exactive HF-X instrument (Thermo Fischer Scientific). First, UCB Pharma provided a CHO-DG44 null cell line harvest sample to generate a spectral library dedicated to the organism used for the culture. We generated a comprehensive spectral library including the fractionation of the null cell line harvest sample and DDA runs of the mAb A33 CCCF and PPA fractions. In order to process the DIA data, we used the SpectronautTM software which offers the possibility to extract the data by peptide- and spectrum-centric approaches with its own Pulsar search engine and the development of directDIATM. Finally, the DIA peptide-centric approach was benchmarked against the MS1-XIC DDA method developed in the previous chapter and the directDIA approach. The data acquisition and extraction methods were evaluated through the comparison of the number of HCPs quantified as well as the quantification accuracy obtained.

3. Experimental design

The analytical workflow is presented in Figure 49.

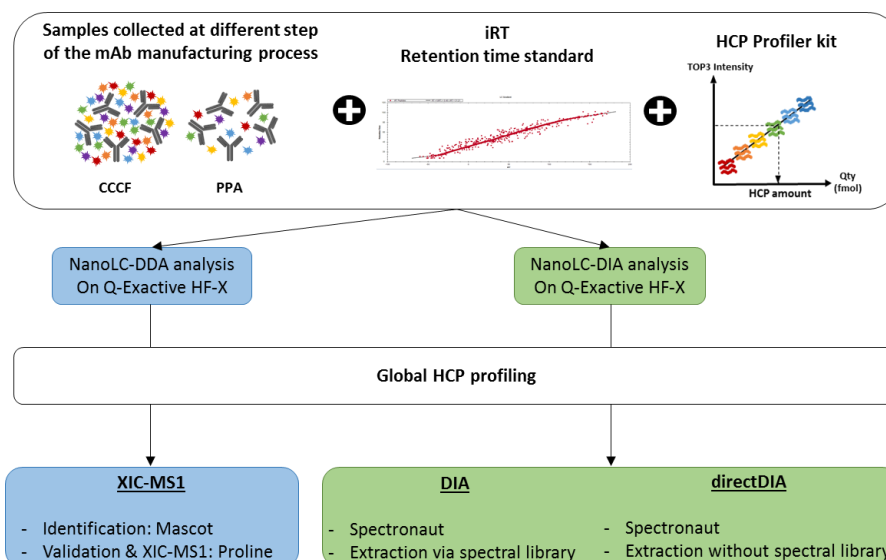


Figure 49: Analytical workflow used to assess the performance of the DIA acquisition mode on a Q-Orbitrap instrument. Four CCCF and eleven PPA fractions were in-gel digested and spiked with iRT retention time standards (Biognosys) and a bead of the HCP Profiler standard. Samples were injected in triplicate using DDA and DIA acquisition modes on a nanoLC-Q Exactive HF-X (Thermo Fischer Scientific) system. MS2-based DIA quantification was benchmarked against a classical XIC-MS1 DDA quantification. Furthermore, the spectrum-centric approach developed by Spectronaut, called directDIA, was evaluated against the spectral library-based approach (DIA) on the same software.

Samples produced by Dr. Gauthier Husson, collected after harvest procedure (4 samples) and after Protein A affinity chromatography (7 samples) as described in [288], were stacked in a single band. After in-gel trypsin digestion, retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) for RT normalization during DIA data processing and a bead of the HCP Profiler standard were spiked. Samples were analyzed in triplicate on a nanoLC-Q Exactive HF-X (Thermo Fisher Scientific) in DDA and DIA modes.

First, peptides and proteins were identified from DDA analyses using Mascot. The validation of the dataset as well as the XIC-MS1 quantification were performed by Proline¹⁹. Then, a Top3 strategy¹⁶ was applied to derive HCP amounts using the 6 points calibration curve of the HCP Profiler standard.

Then, a dedicated spectral library was generated using DDA runs of a CCCF CHO-DG44 null cell line fractionation (24 analyses including iRT retention time standards from Biognosys) combined with all DDA analyses of the current CCCF and PPA mAb A33 samples spiked with the HCP Profiler kit as well. The software SpectronautTM and its Pulsar algorithm were used to generate a CHO-specific spectral library containing 40 281 peptides from 3 978 protein groups. Moreover, SpectronautTM was used in order to extract MS2 signals from the DIA data using either the spectral library or the directDIATM

algorithm. For both extraction methods, a Q-value filter of 0.01 was applied and iRT profiling was set on. Finally, HCP amounts were estimated using the Top3 HCP Profiler strategy.

Sample preparation, LC-MS/MS methods and data treatment are detailed in the experimental part, section 1.B.

4. Data acquisition and processing optimisations

A. Sample-dependent data acquisition optimizations

HR/AM instruments achieve acquisition speeds allowing a survey MS scan followed by MS2 scans of co-eluting precursors' fragments within small m/z windows in a few seconds. The Q Exactive HF-X is slower than the TTOF 6600. Therefore, the DIA method had to be optimized according to the MS system used but not only. As shown in Figure 50.B, the distribution of precursors over the m/z acquisition range is not uniform. Consequently, the use of windows of variable widths is necessary to reduce the MS2 spectra complexity and increase the specificity towards the sample studied. Samples collected at the same purification step tend to have similar HCP profiles in terms of numbers and quantities, respectively around 100 000s ng/mg mAb for 1 000s of HCPs in CCCF fractions and 1 000s ng/mg mAb for 100s HCPs in PPA samples. As we aim to develop a method that can support the manufacturing process of mAbs, we started to optimize DIA methods dedicated to HCP-rich harvest samples on the one side and less complex PPA fractions on the other side (Figure 50).

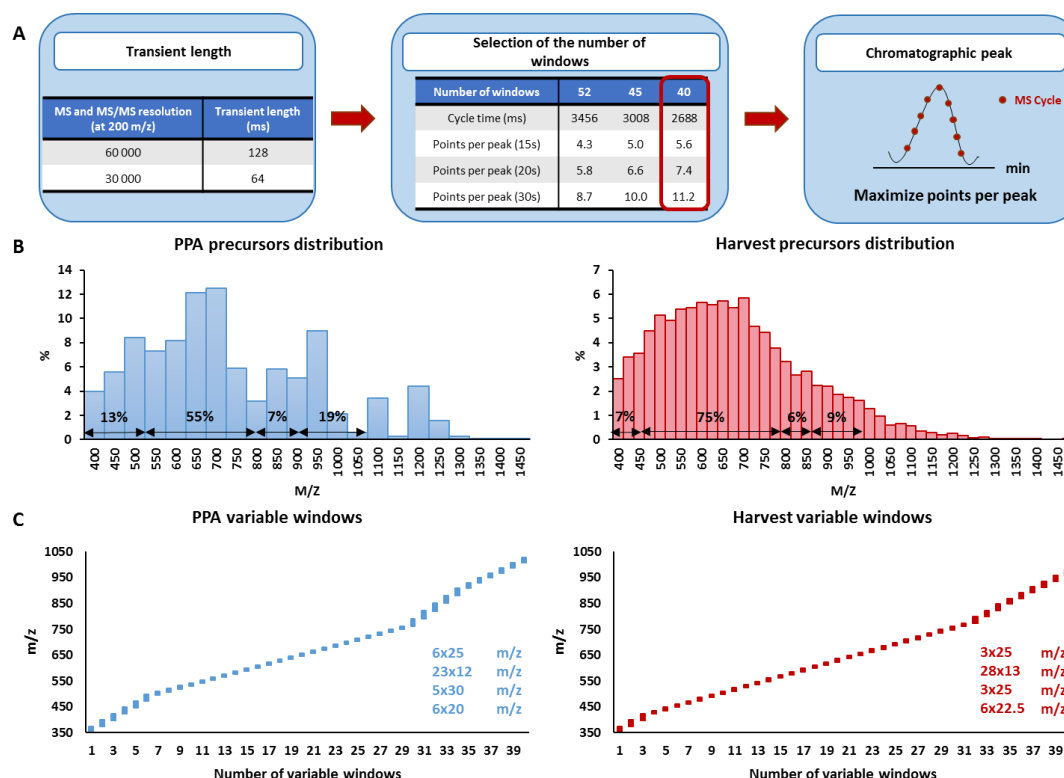


Figure 50: Optimization of variable windows DIA methods in accordance with sample complexity. Based on the transient length at our working MS resolutions, a 40 variable windows DIA method was used in order to maximize the MS cycle per chromatographic peak (A). Then, by looking at the distribution of precursors over the m/z acquisition range obtained in DDA runs (B), the 40 variable windows were split from 350 to 1006 and 350 to 957 m/z , respectively for PPA and CCCF samples (C).

First, we selected the number of isolation windows based on the MS and MS/MS scan times at our working resolutions. The aim is to maximize the number of MS cycles per elution peak, i.e. the number of times the instrument will perform a MS scan followed by the MS/MS acquisitions within a chromatographic peak. A 40 variable windows DIA method that will theoretically allow 5 to 11 MS cycles per peak was selected (Figure 50.A). Then, based on the distributions of precursors over the m/z acquisition range in DDA runs (Figure 50.B), a method was optimized for each type of sample (Figure 50.C). Finally, we were able to reach a median of 6 data points per peak for PPA fractions and 8 for CCCF samples. The value was calculated in Spectronaut based on all peptide precursors identified in the RT calibration. As the number of peptides is higher in CCCF fractions, we observed a higher median peak width resulting in an increased data points per peak number.

B. Evaluation of the HCP Profiler standard combined with DIA

In the previous chapter, we demonstrated the strength of the HCP Profiler standard for the global profiling of HCPs. The standard was initially developed using a XIC-MS1 DDA approach and the use of MS2-based DIA quantification combined with the HCP Profiler had never been investigated.

At first, we performed the peptide-centric approach using the dedicated spectral library including the IgG4 mAb A33, the HCPs and the HCP Profiler peptides information. This search lead to wrong assignments for the profiler peptides such as acetylated and oxidized peptides or peptides showing one missed cleavage. The resulting calibration curves showed regression coefficients between 0.7 and 0.9 (Figure 51.A). As the HCP Profiler standard is well characterized by ANAQUANT, we know exactly which peptides should be observed in our data. Consequently, a second search was performed using a spectral library that contained only the HCP Profiler peptides information. Optimized settings, such as no missed cleavage allowed and only carbamidomethylation of cysteine as fixed modification, were used to generate the HCP Profiler specific library with the Pulsar search engine. Similarly, the parameters used to extract HCP peptide signals using the spectrum-centric approach were not suitable. The most accurate calibration curves were obtained in a second search using the same optimized settings and a database composed of the 18 HCP Profiler proteins. After the improvements, we benchmarked the DIA HCP Profiler calibration curves against DDA curves (Figure 51.B).

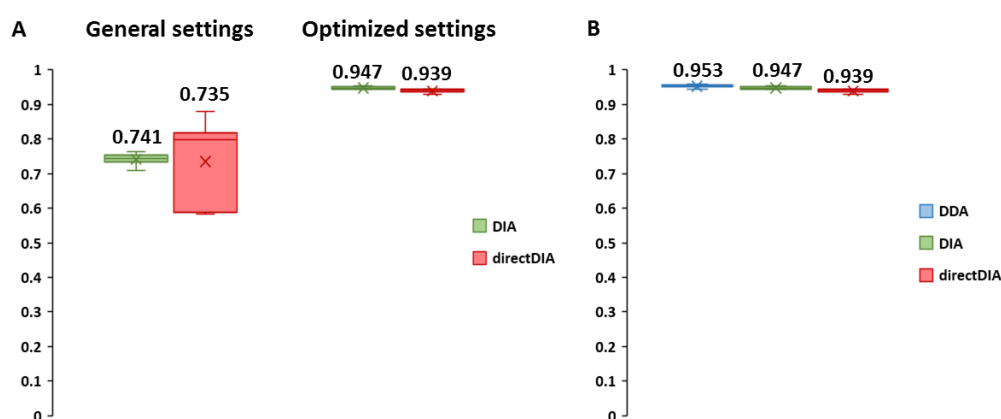


Figure 51: Regression coefficients observed for the HCP Profiler standard calibration curves from DDA, DIA and directDIA approaches. Improved calibration curves using DIA or directDIA extraction methods with a search specific to the HCP Profiler standard (A). No missed cleavage allowed and carbamidomethylation of cysteines as fixed modification were used for the optimized settings. General settings are detailed in the experimental part, section 1.A. DDA, DIA and directDIA regression coefficients dispersion observed for all PPA and CCCF samples (B).

The same reproducibility was observed for both acquisition modes as CVs calculated on the slopes, intercepts and regression coefficients were all below 3%. Furthermore, DIA calibration curves were obtained using the signal of the 54 standard peptides for PPA samples and 53 out of the 54 peptides for CCCF samples. A difference could be noted using the directDIA approach with CCCF calibration curves obtained from 52 out of the 54 peptides. The missing peptides in both DIA strategies are from the lowest calibration point (1 fmol). In comparison, no DDA calibration curves were obtained with all standard peptides. They were obtained using 51 to 53 peptides depending on the sample. As for DIA, the missing standard peptides were from low calibration points (1 or 2.5 fmol), the increased sensitivity of the DIA approach was already observed at the standard peptides level with a better coverage of the low calibration points. This observation will be further investigated with the HCP impurities.

C. Implementation of a stringent data filtering workflow

After the optimization of the HCP Profiler peptides extraction, we focused our efforts on HCP peptides. Following data acquisition and peptides signals extractions, the manual checking of many extracted signals rise the need to implement stringent validation criteria in order to end up with only highly confident and reproducible signals. Therefore, the following validation workflow including 5 stages of data filtering was implemented (Figure 52).

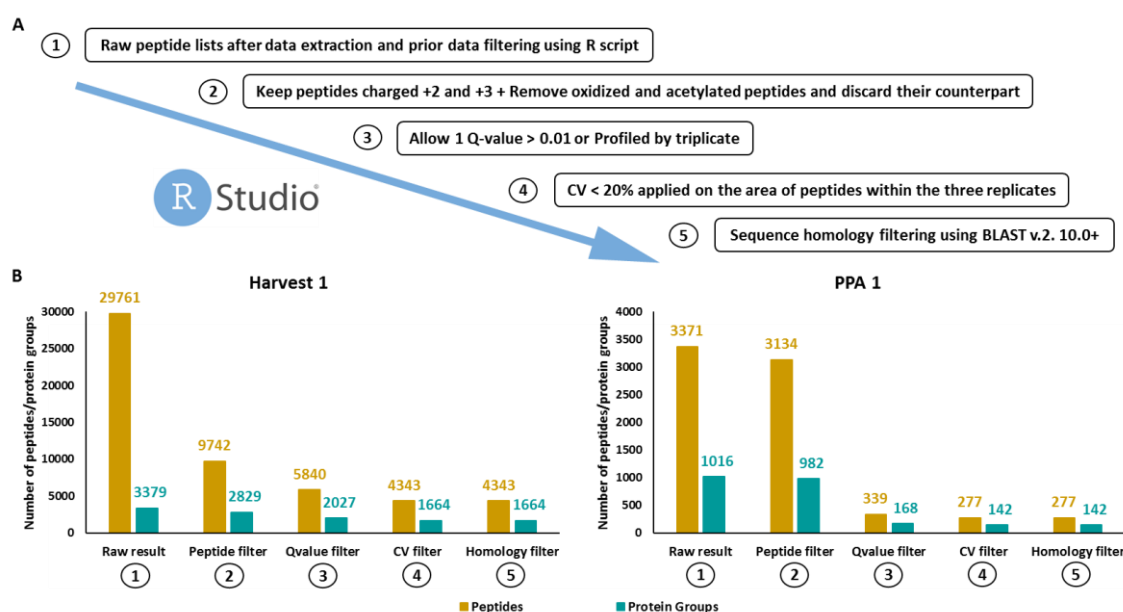


Figure 52: Impact of consecutive validation criteria applied on the number of peptides and protein groups quantified using the spectral library-based DIA approach. Description of each step used to filter the raw lists of HCP peptides (A). The filters are detailed in the text. Numbers of peptides and protein groups remaining after each filter applied to the Harvest 1 and PPA 1 list of HCP peptides identified using Spectronaut are shown (B).

First, we kept doubly and triply charged peptides and we removed oxidized and acetylated peptides with their unmodified counterparts (2). Most of the time, these modifications and the four times charged peptides display bad quality signal in terms of intensity and peak shape. The second filter can be described as signal quality filter (3). Precursors with more than one Q-value > 0.01 and/or profiled were removed. This filter is the most impactful as it removes several peptides and protein groups (Figure 52.B). The 89% less peptides for PPA 1 also demonstrates the difficulty of the software to extract signals of low abundant proteins. The third criteria is linked to the reproducibility of the signal

as precursors with CV above 20% were filtered out (4). Next, a homology filter is applied using BLASTP³¹ (v.2. 10.0+) against mAb heavy and light chain sequences (5). HCP peptide sequences containing a minimum of 6 amino acids, 80% coverage and 100% identity were removed. This last filter does not remove many peptides but it is an important filter as the mAb is the most abundant protein in the sample. A mAb peptide signal wrongfully attributed to a HCP would then inevitably result in an over-estimation of its quantity. Finally, the selection of the 3 most intense peptides was performed and absolute HCP amounts were estimated using the HCP Profiler standard.

5. Results

A. Benchmarking DIA versus MS1-XIC DDA

To dig deeper into the global profiling and individual quantification of HCP impurities, we decided to investigate DIA on the Q Exactive HF-X. DIA strategy promises gains in reproducibility, sensitivity and quantification accuracy over DDA acquisition mode. The performance of the peptide-centric DIA strategy was evaluated against the standard MS1-XIC DDA approach on the same samples. The number of protein groups quantified and the global HCP amounts estimated are presented in Figure 53.

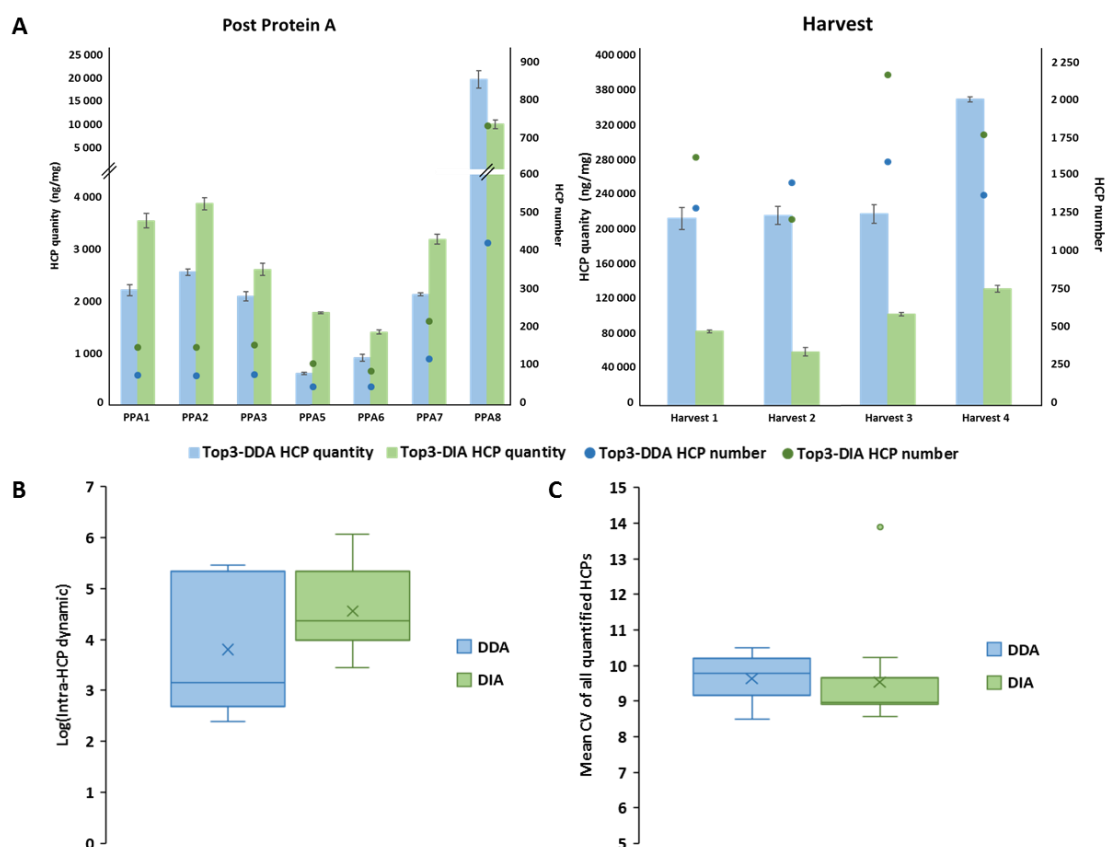


Figure 53: Evaluation of the spectral library-based DIA strategy against MS1-XIC DDA for global HCP profiling using the HCP Profiler standard. Global HCP amounts and numbers of HCPs quantified obtained for PPA and CCCF fractions using MS1-XIC DDA and MS2-based DIA strategies (A). Bar heights represent the mean of the global HCP amount in injection triplicates. Error bars represent the standard deviation. Representation of the intra-HCPs dynamic range observed using DDA and DIA acquisition modes (B). Comparison of the mean CVs calculated on HCP peptides intensities within replicates for DDA and DIA analyses (C).

The MS1-XIC DDA approach was able to quantify an average of 1464 HCPs in the CCCF fractions with global quantities between 222 646 and 365 145 ng/mg mAb, and 115 HCPs in PPA fractions representing 569 to 19 153 ng/mg mAb. Through the use of MS2 signals acquired in DIA and extracted via a spectral library, we quantified an average of 1737 HCPs with a global estimation between 62 792 and 138 297 ng/mg mAb for CCCF samples and 159 HCPs in PPA fractions with a quantity between 1 339 and 11 992 ng/mg mAb (Figure 53.A).

At first sight, we observed the improvement achieved by the DIA peptide-centric strategy on the HCP coverage as around 19% and 111% more HCPs were quantified for CCCF and PPA samples respectively. The Harvest 2 sample was the only one to show a lower number of HCP quantified in DIA. This is largely due to a problem with one of the injection replicates resulting in the loss of several HCP peptides after the CV filter. Unfortunately, the sample could not be re-injected. For PPA samples, the overall estimated HCP amounts were consistent with the increased numbers of HCPs quantified. On the contrary, even with a higher number of quantified HCPs, the CCCF fractions showed lower overall amounts compared to the DDA strategy. This observation shows the potential presence of false positives or interferences during the extraction of MS1 signals resulting in an overestimation of the HCP amounts.

As previously observed with the calibration curves, the DIA acquisition displayed an increased dynamic compared to DDA. While MS1-XIC DDA was able to achieve 2.4 to 5.5 orders of magnitude between the least and most abundant HCP, MS2-based DIA allowed to reach a dynamic between 3.5 and 6.1 (Figure 53.B). We can highlight that 7 orders of magnitude between the mAb and the HCPs were covered with the DDA and DIA acquisition modes. The precision of the DIA data extraction was observed by looking at CVs calculated on peptide intensities within injection replicates for all HCPs quantified, a median of 9.0% was obtained for DIA compared to 9.8% for DDA (Figure 53.C). Our DIA approach showed an average variation of less than 10% despite a higher number of identified peptides. Then, we compared the ratios of the quantities obtained by Top3-DIA over Top3-DDA (Figure 54).

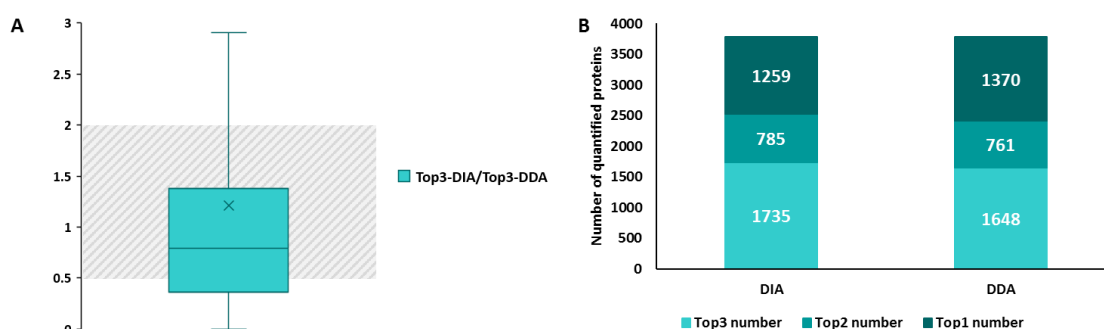


Figure 54: Benchmarking of consistency and coverage of the DDA and DIA acquisitions modes. Representation of the ratios calculated between the Top3-DIA and Top3-DDA amounts obtained for the common HCPs between both acquisition methods (A) Out of range ratios were removed to observe the 52% of 3779 common HCPs within a factor 2. Grey area represent the ratios comprised within a factor 2. Representation of the number of peptides used to quantify each HCP (B).

Within the DDA and DIA results of the 11 samples, 3779 HCPs were quantified by both methods. We notice that only 52% of the ratios are within a factor 2 (Figure 54.A). This is due to the number of peptides used to quantify each protein (Figure 54.B). A larger number of peptides was used in DIA compared to DDA, 8034 and 7836 peptides, respectively. Among the 3779 proteins compared, 1735 HCPs were quantified with a Top3 in DIA while this number drops to 1648 in DDA. As we sum the area

of the Top3 peptides, the increased number of peptides per HCP obtained in DIA has a direct impact on the HCPs quantification. Thus, the Top3 strategy combined with DIA improves the accuracy of quantification results.

Our results demonstrate a great reproducibility and accurate quantification using the Top3-DIA approach and a sensitivity down to sub-ng/mg mAb. The DIA acquisition mode has shown its ability to extract signals close to the background noise. This is a significant benefit when the impurities of interest are at the level of traces. Moreover, the combination with the Top3 quantification method provides an overview of the HCPs content and thus to follow manufacturing process developments.

B. Spectral library free-based approach investigation

A crucial point of the DIA approach is the way to extract the signals from each MS² scan, which contains the fragmentation information of several precursors. We already demonstrated the use of a specific spectral library to extract peptide signals from DIA data. To continue the development of a straightforward method using DIA acquisition, we focused on the spectrum-centric approach called directDIA in Spectronaut (Biognosys AG). This approach relies on the generation of “pseudo MS/MS spectra” from the DIA runs which will be used to query a search against the organism’s proteome database as it is done using classical DDA search engine. The directDIA algorithm was evaluated against the spectral library-based approach on the same sample set. The overall HCP quantities and numbers of HCPs quantified by the peptide- and spectrum-centric approaches were benchmarked (Figure 55).

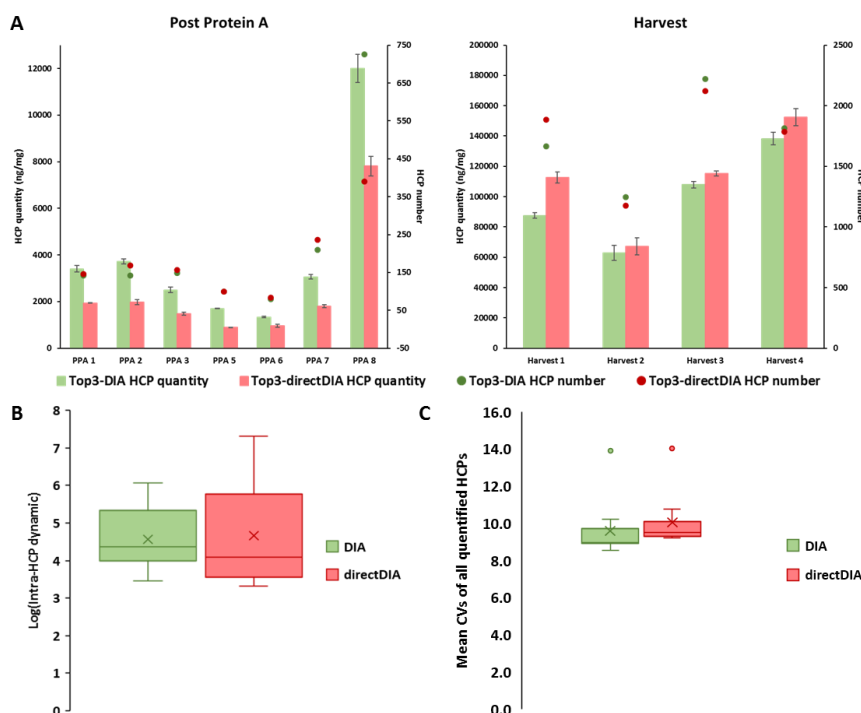


Figure 55: Evaluation of DIA peptide- and spectrum-centric strategies for global HCP profiling using the HCP Profiler standard. Global HCP amounts and numbers of HCPs quantified obtained for PPA and CCCF fractions using DIA and directDIA strategies (A). Bar heights represent the mean of the global HCP amount in injection triplicates. Error bars represent the standard deviation. Representation of the intra-HCPs dynamic range observed using DIA and directDIA extraction methods (B). Representation of the log ratios between the DIA and directDIA quantities of the 500 most abundant HCPs of the harvest 1 sample (C).

Similar quantification results and close numbers of quantified HCPs were obtained. The directDIA strategy was able to quantify an average of 1741 HCPs in the CCCF fractions with global quantities between 67 305 and 152 458 ng/mg mAb, and 183 HCPs in PPA fractions representing 896 to 14 918 ng/mg mAb (Figure 55.A). The sensitivity of the DIA approach was not affected by the directDIA extraction mode with quantification of HCPs down to the sub-ng/mg mAb level. Both extraction strategies were able to achieve 3 to 6 orders of magnitude between the least and most abundant HCPs (Figure 55.B). When we relate this to the overwhelming mAb, we notice that nearly 7 orders of magnitude are covered by both methods. In addition, the precision of extraction of the two approaches is also very close with an average of CVs calculated on peptide intensities within replicates about 9 and 9.5, respectively for DIA and directDIA (Figure 55.C).

Looking further at individual HCP amounts, we compared the quantities obtained using Top3-DIA over Top3-directDIA (Figure 56).

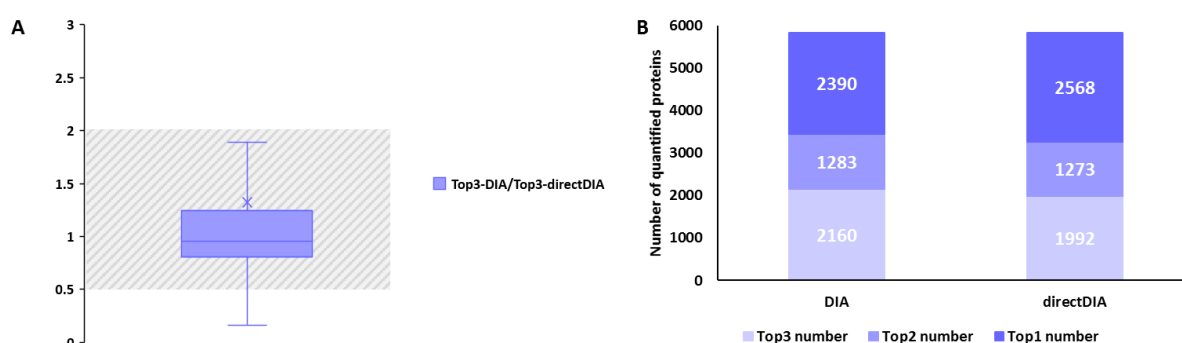


Figure 56: Benchmarking of accuracy of the DIA and directDIA extraction strategies. Representation of the ratios calculated between the Top3-DIA and Top3-directDIA amounts obtained for the common HCPs between both extraction methods (A) Out of range ratios were removed to observe the 82% of 5833 common HCPs within a factor 2 (grey area). Representation of the numbers of peptides used to quantify each HCPs (B).

For the 5833 HCPs commonly quantified with both methods. A median of 0.96 was obtained and the amounts estimated were in accordance for 82% of the common HCPs within a factor 2 (Figure 56.A). Similarly, close numbers of peptides were used to quantify those common HCPs, respectively 11 436 and 11 090 for DIA and directDIA (Figure 56.B). These observations showed confidence towards the spectral library free approach. Nevertheless, the lower numbers of HCPs quantified using directDIA results still highlight the difficulties that still exist to extract low-level impurities using directDIA.

To continue the comparison between both extraction strategies, we focused on the list of quantified HCPs. An average of 62% of the HCPs quantified in DIA were also extracted in directDIA. On the one hand, about 38% of the HCPs quantified using the spectral library could not be extracted by directDIA. This observation thus concerns hundreds of unidentified HCPs highlighting the need for optimization of the search algorithm as these HCPs were identified in the spectral library. On the other hand also, hundreds of HCPs were not quantified using the spectral library-based approach highlighting the incompleteness of the spectral library (Figure 57).

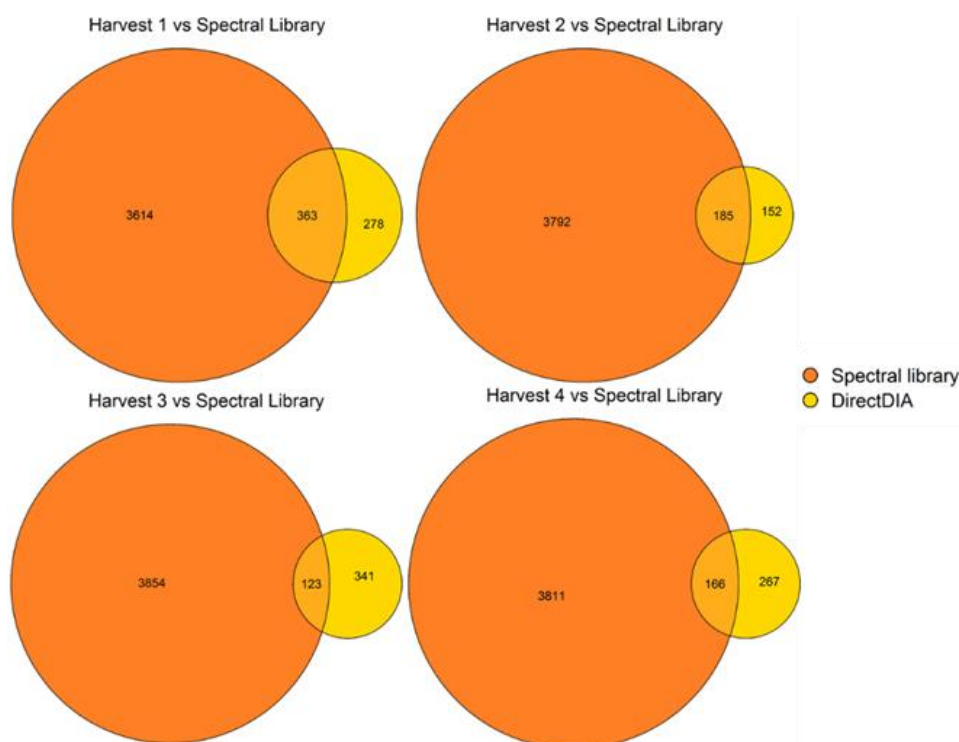


Figure 57: Venn diagrams of the HCPs that compose the generated in-house spectral library and the HCPs only quantified using directDIA in CCCF fractions. Around 56% of the HCPs only quantified using directDIA are not covered by the in-house CHO specific spectral library.

Looking further into these lists of HCPs, we observed that on average 260 HCPs for each CCCF sample were not identified in the spectral library. These HCPs cover 5 orders of magnitude between the least and most abundant HCP with a maximum reaching thousands of ng/mg mAb. The number and quantities of those HCPs are not negligible and bring to light the limitations of the spectral library extraction approach as HCPs not present in the library will not be identified and further quantified. However, we cannot be completely confident in these results. We could suppose that HCPs with an estimated amount around thousands of ng/mg mAb should be identified in the spectral library generated, at least with a few peptides. The number of false positives must therefore still be significant in directDIA while current methods still fail to properly estimate the error rates. Compared to the targeted DIA analysis, the spectral library free-based approach relies only on the use of the CHO protein sequence database. Unfortunately, this database is composed of 56 565 protein sequences (UniProtKB/TrEMBL) in which 99% are annotated without any manual verification nor curation. As a result, the redundancy and high number of proteins complicates the extraction of peptides unique to a protein or even protein group. At least, these are encouraging results to a move towards a universally applicable method for an accurate and reproducible HCP profiling. Overall the directDIA approach supported the increased HCPs coverage enabled by the DIA acquisition in line with previous studies^{45-46, 294}.

C. Where do we stand in regards to ELISA and the Sciex Triple-TOF instrument (TTOF 6600)?

As mentioned in the presentation of the project context, our HCP Profiler-DIA method developed on the Q Exactive HF-X follows the first implementation of a DIA-SWATH method on a Triple TOF 6600 by Dr. Gauthier Husson. It is interesting to see where we stand in comparison to the fast scanning TTOF and the gold standard ELISA. As the ELISA assays were performed on the PPA samples, we compared the global quantities and numbers of HCPs quantified using both MS methods and the overall HCP amounts obtained for the three strategies on the seven PPA fractions (Figure 58).

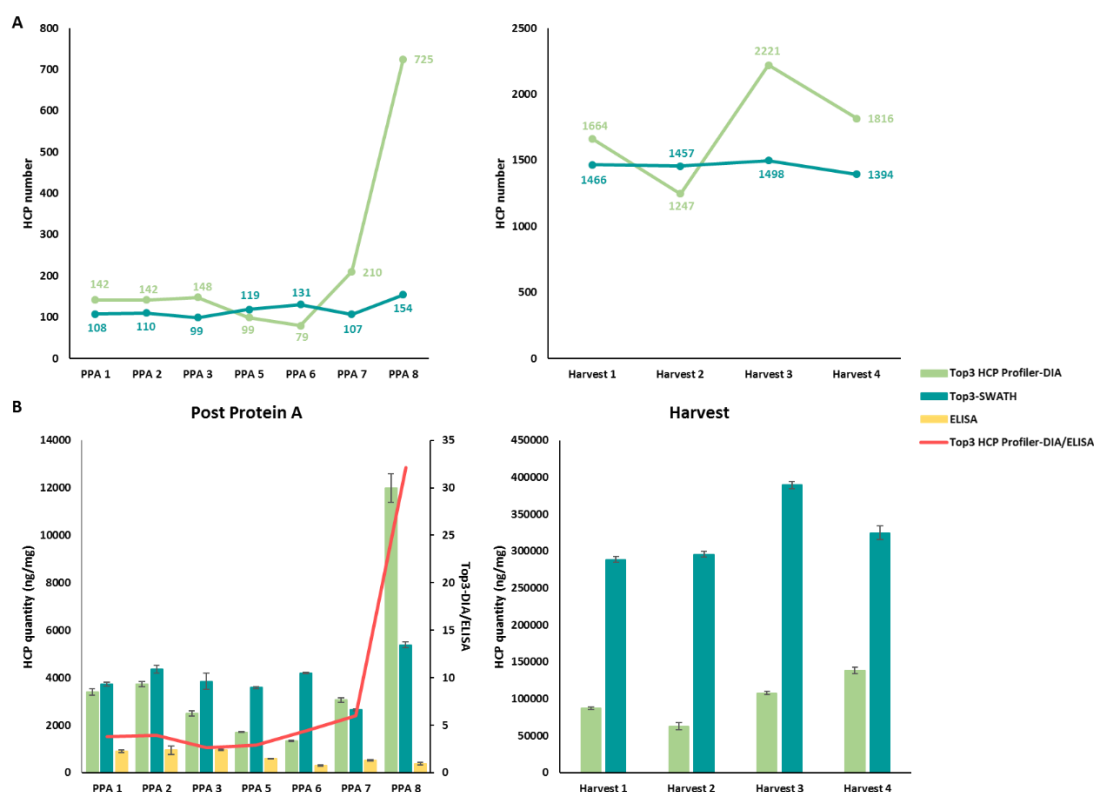


Figure 58: Benchmarking of the HCP Profiler-DIA method against the gold standard ELISA and the DIA-SWATH method. Representation of the overall HCP amounts obtained for the three analytical methods and superposition of the ratios between the HCP Profiler-DIA approach and ELISA assay (A). Bar heights represent the means of the global HCP amounts in injection triplicates. Error bars represent the standard deviation. Number of HCPs quantified by the HCP Profiler-DIA method on the Q Exactive HF-X and the DIA-SWATH method performed on the TTOF 6600 (B).

Several differences between both MS methods made the comparison of the quantification results challenging. First, the samples spent five years in the -80°C freezer since their first injections on the TTOF 6600. The LC system of the Q Exactive HF-X was operated in nano-flow while the TTOF instrument was run in micro-flow mode. As a result, the nano-LC system offers an increased sensitivity at the expense of robustness compared to micro-LC system. Therefore, we compared injections of 400 ng of peptides on the Q Exactive HF-X to injections of 8 μg on the TTOF 6600. Then, the lower acquisition speed of the Q-Orbitrap instrument (40 Hz vs 100 Hz) made us decrease the number of variable windows in order to increase the number of MS cycles per elution peak thus ensuring a minimal number of 6 points per peak (40 vs 75 variable windows methods). Finally yet importantly, we moved from Skyline to Spectronaut for DIA data processing. Indeed, Dr. Joanna Bons and Dr. Nicolas Pythoud

demonstrated the advantages of Spectronaut that allows non-linear regression for RT alignment²⁹³ and the dynamic adjustment of the RT window width. The first setting offers an increased sensitivity compared to a linear regression that does not take into account local unexpected variations and non-linear chromatographic gradients, while the second allows to best fit the peptide elution profile.

Despite all these changes, the optimizations of the spectral library and the MS method in accordance to the sample series allowed us to reach equivalent or even better HCPs coverage (Figure 58.A). We were able to increase the numbers of HCP quantified for most of the samples. The lower number of HCPs quantified in the Harvest 2 sample is largely due to a problem with one of the injection replicates resulting in the loss of several HCP peptides after the CV filter. For PPA 5 and PPA 6 samples, no explanation could be found as a lower number of HCPs was identified in each replicates for both samples. In contrast, the overall HCP amounts estimated using the DIA-SWATH strategy were higher (Figure 58.B). This result was not surprising as the data extraction and data treatment have been largely optimized since the implementation of the initial DIA-SWATH method. Overall, thanks to Spectronaut and our post-data extraction filters (Part 4.C), we have increased confidence in the results while reducing the presence of false positives or the use of interfered signals for the quantification.

Compared to the gold standard ELISA assay, the HCP Profiler-DIA method showed global quantities higher by a factor 4 for standard samples and 32 for the PPA 8 sample that was obtained with a modified protocol. These results highlighted the limits of ELISA that were previously mentioned such as the incomplete coverage of HCPs. In parallel, this has further shown the importance of MS as an orthogonal method. Indeed, the increased HCP map coverage alongside with the accurate and reproducible quantification make DIA a method of choice to reach an overview of the HCP contents of samples from various steps of the mAb manufacturing process.

6. Conclusion & Perspectives

HR/AM instruments operated in DIA mode became powerful tools for HCP impurities identification and quantification. These instruments have sufficient acquisition speeds, resolutions and dynamic ranges to really benefit from the DIA mode. Indeed, DIA combines the advantages of DDA for the global profiling of HCPs and the ones of targeted strategies for the precise quantification.

The evaluation of DIA against the standard DDA-based method displayed a major improvement towards the HCP coverage as an average of 1.8 more HCPs were quantified. The combination with a Top3 strategy enables the quantification of all identified HCPs. The use of the HCP Profiler standard has shown to benefit from the DIA acquisition. Its use improves the robustness of the sample preparation. Moreover, the multi-level peptides allow more precise quantification and internal controls of the analysis. Consequently, the method is ready to be implemented in several production sites.

Our HCP Profiler-DIA approach allows to obtain the identity and the quantity of HCPs in a single analysis. Therefore, the method enables the monitoring of HCP impurities throughout the manufacturing process to support the release of safer biotherapeutics. In the next chapter, we have challenged our developed Top3-DIA method in order to increase our understanding of the manufacturing process.

One of the main objectives of the developed method was to support the analyses of USP and DSP samples. The harvest and PPA fractions used for this purpose were collected at steps where the HCP content is still accessible, even though hardly, by MS. On the contrary, the final drug products display the lowest HCP contents but they are the most complex to analyze due to the extreme dynamic range between the mAb and the HCPs requiring the detection and quantification of impurities at the amol-level. The next part of this manuscript will focus on the optimization of the sample preparation dedicated to these challenging drug products. Several methods can be applied to decrease the dynamic range^{27, 29, 49, 51} but we will continue to progress with the intention of developing a robust and straightforward method to achieve a maximal HCPs coverage.

Chapter 3

Case study: Impact of the bioreactor capacity on the HCPs profile

1. Context of the project and samples description

Following the development of the Top3-DIA method on the Q Exactive HF-X, we wanted to challenge our approach and use it to increase the knowledge of the mAb manufacturing process from UCB production cell line. UCB Pharma had the idea to study a potential impact of the production scale up during the development of a new therapeutic until its commercialization. From early research stages to the commercialization of the drug product, the demand for mAbs grows from grams to kilograms. In order to meet the need, the culture during the USP is conducted in larger capacity bioreactors. In general, the volume of these bioreactors ranges from 80 L to 25 000 L at the latest stages of this scale up process.

Recent studies have explored the impact of USP conditions on the HCP Profiles^{4, 22, 295}. It was shown that mAb type, cell line, culture viability, temperature and duration could have an effect on HCPs and other CQAs. In addition to the work done, we wanted to observe any tendency or impact of the production scale up. Therefore, we performed the study of four mAb CCCF and PPA samples produced from a CHO DG44 cell line and cultured in bioreactors of varying capacity from 80 L to 15 000 L (Table 7). The sample series was prepared and injected on the Q Exactive HF-X using DIA mode. Unfortunately, we could not get provided with a sufficient number of HCP Profiler beads to use them for the study of the 18 samples. This is why we decided to use the mix of four standard proteins from the MassPREP digestion standard kit (Waters). The use of the mix also relies on the Top3 strategy considering the PYGM protein as a reference and the three other proteins (ADH, BSA, and ENL) to calculate ratios, as internal controls.





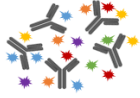
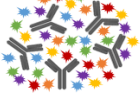
Sample type	Bioreactor capacity (L)	mAb 1 	mAb 2 	mAb 3 	mAb 4 
 PPA	80			X	X
	200	X			X
	2 000	X	X	X	X
	15 000			X	
 CCCF	80	X		X	
	200	X	X		
	2 000	X	X	X	X
	15 000			X	

Table 7: Sample series studied in order to observe an impact of the production scale up. A series of nine CCCF and nine PPA fractions produced from four CHO DG44-derived mAbs. Samples were cultivated in bioreactors from 80 L to 15 000 L.

First, samples produced by UCB Pharma, collected after harvest procedure (9 samples) and after Protein A affinity chromatography (9 samples), were stacked in a single band. After in-gel trypsin digestion, retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) for RT normalization and the mix of 4 standard proteins (ADH, PYGM, BSA and ENL from Waters MassPREP Digestion Standard Kit) were spiked. Samples were analyzed in triplicate on a nanoLC-Q Exactive HF-X (Thermo Fisher Scientific) in DIA mode. Next, we investigated the use of a hybrid spectral library composed of DDA and DIA analyses to increase the specificity of our library-based DIA approach toward the analysis of these four mAbs fractions. This hybrid library was benchmarked against the spectral library generated in the previous chapter and further against the directDIA approach. Finally, the evaluation of the process scale up was performed starting from the overall HCP amounts and the numbers of HCPs quantified. Then, we focused on intra- and inter-mAb samples comparisons in terms of HCP profiles and individual quantities.

2. Data analysis strategy

The general workflow of the data analysis is presented in Figure 59.

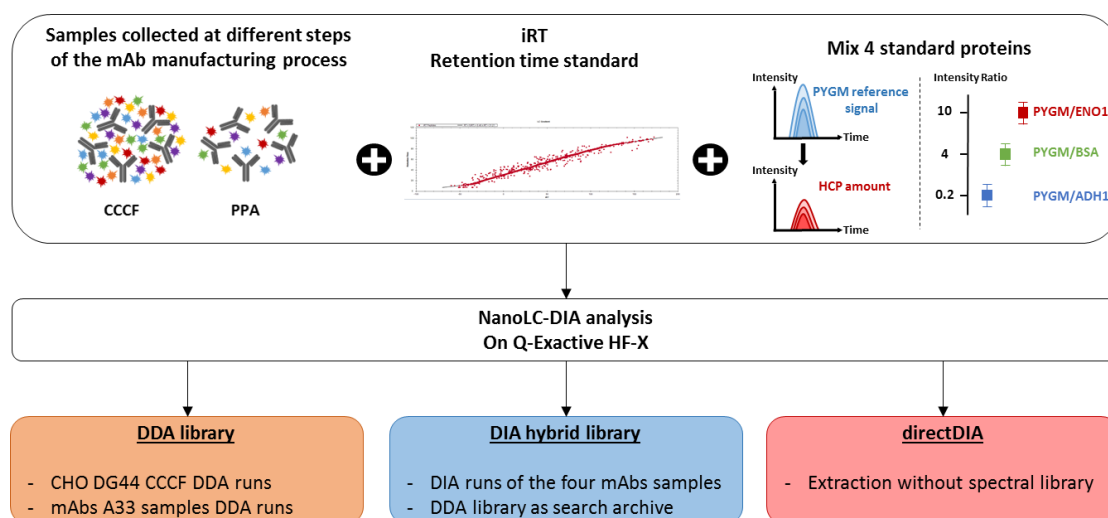


Figure 59: Data processing workflow used to generate a project-specific spectral library. Four mAbs CCCF and PPA fractions were in-gel digested and spiked with iRT retention time standards (Biognosys) and the mix of four standard proteins from the MassPREP digestion kit. Samples were injected in triplicate using DIA acquisition modes on a nanoLC-Q Exactive HF-X (Thermo fischer Scientific) system. Hybrid spectral library-based search was benchmarked against a classical DDA library-based analysis and directDIA using Spectronaut software.

The possibility to use hybrid spectral libraries combining both DDA and DIA runs was very recently implemented into Spectronaut™ with the aim to improve the accuracy of signals extraction and the coverage of the proteome using spectral library DIA data extraction. We wanted to evaluate the benefits of this strategy on our sample series and have therefore set up a hybrid spectral library using the DIA runs of the four mAbs PPA and CCCF samples added to the search archive of the IgG4 mAb A33 project-specific spectral library (DDA library). The use of the hybrid library generated containing 46 330 peptides and 4139 protein groups was benchmarked against the mAb A33 project-specific spectral library, including 44 726 peptides and 4308 protein groups, and directDIA approach. Then, the filters

developed in the previous chapter were applied after data extraction. HCP amounts were estimated using the universal response factor calculated from the PYGM Top3 peptide area. Finally, the overall and individual amounts of HCP were investigated through intra- and inter-mAb type samples comparisons.

Sample preparation, LC-MS/MS methods and data treatment are detailed in the experimental part, section 1.C.

3. Results

A. Optimization of a project-specific spectral library

Several studies have shown the impact of the mAb type on co-expressed and co-eluted HCPs during USP and DSP²³⁻²⁴. Therefore, we can raise questions regarding the use of our IgG4 mAb A33 project-specific spectral library (DDA library). HCPs that were not identified in the mock cell line DDA analyses and that would preferentially co-express with any of the four mAbs rather the mAb A33 will be missed in the spectral library and not targeted during DIA data extraction. Besides, the generation of a library specific to the four mAbs would require additional sample preparation and acquisition time. Gel fractionation of a CCCF sample from the four mAbs added to the acquisition time of each band would take a total of one to two weeks. With the aim to develop a straightforward method, we rather decided to evaluate the use of a hybrid spectral library generated from DDA and DIA data¹⁵. First, we compared the identification results of the hybrid library against the DDA library (Figure 60).

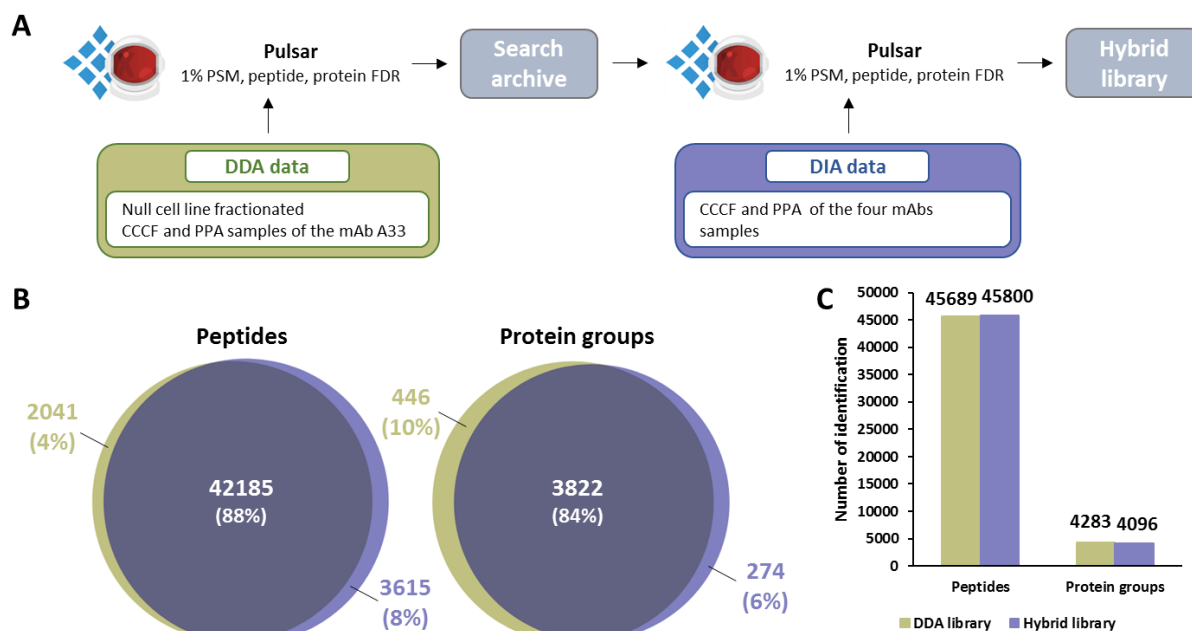


Figure 60: Hybrid spectral library generation. The DDA-based spectral library search archive combined with the 54 DIA runs of the CCCF and PPA samples enables the generation of a project specific hybrid library (adapted from ¹⁵) (A). Overlap of the protein groups identified in both spectral libraries (B). Representation of the number of peptides and protein groups identified in the DDA and hybrid library (C).

In order to generate this hybrid library, we combined the DIA analyses of the CCCF and PPA samples of the four mAbs with the search archive of the mAb A33-specific DDA library (Figure 60.A). After the generation of a spectral library, the result from the Pulsar search is saved as an archive before applying any FDR filter. It allows combining this archive with other analyses or search archives to enhance a library while having an adequate FDR control at the spectral library scale. First, the majority of the HCPs identified are common to both libraries and we were able to observe 274 new HCPs (Figure 60.B). Looking at the overall peptides and protein groups identification number, the generation of this hybrid library allowed us to identify 111 additional peptides compared to the DDA library (Figure 60.C). Interestingly, we identified 182 fewer protein groups. Since our samples are from the same cell line culture, our results showed significant overlap with small differences that can be explained by the application of the FDR filter on two different data sets. Next, we compared the quantification results to evaluate the capabilities of the hybrid library (Figure 61).

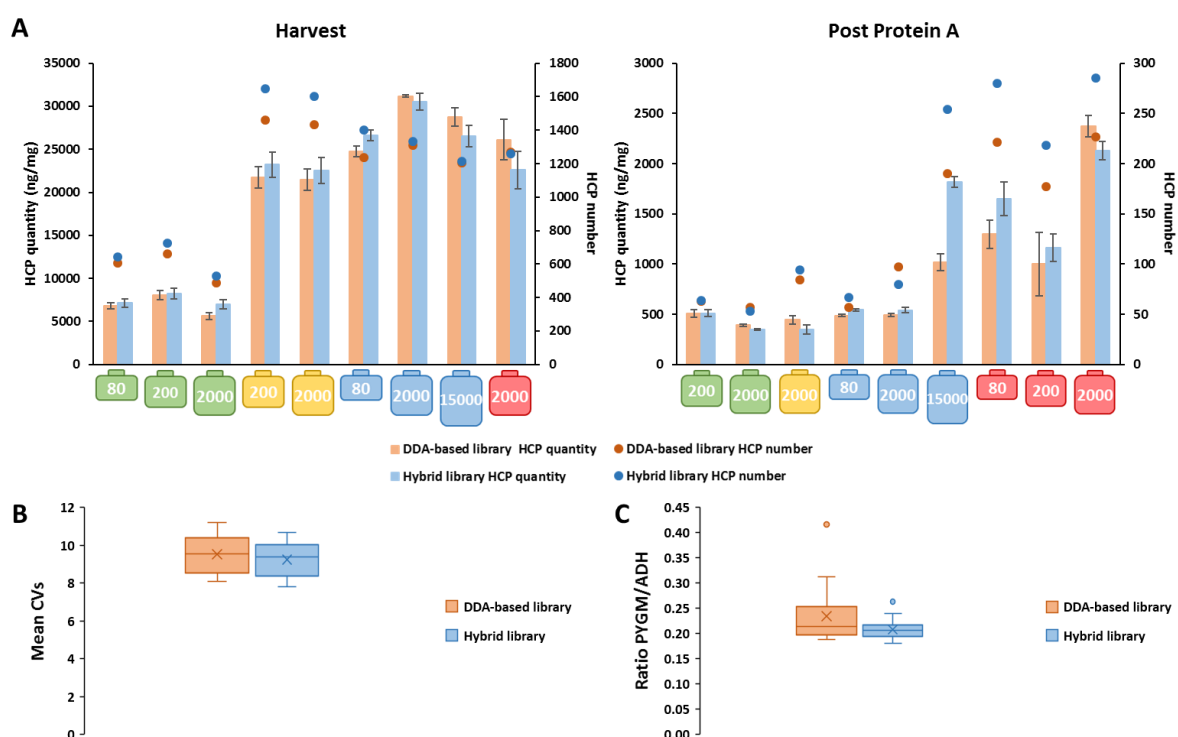


Figure 61: Benchmarking of the hybrid spectral library against DDA-based library for DIA data extraction. Representation of the overall HCP quantities obtained and numbers of HCPs quantified by the hybrid and DDA spectral libraries (A). The dots represent the numbers of HCPs quantified. Bar heights represent the mean of the global HCP amount in injection triplicates. Error bars represent the standard deviation. Distribution of the average CVs calculated on the HCP peptides intensities among replicates for the 18 samples (B). Dispersion of PYGM/ADH ratios obtained with both spectral libraries (C). mAb 1 (green tank), mAb 2 (yellow tank), mAb 3 (blue tank) and mAb 4 samples (red tank).

In general, the results obtained are consistent in terms of overall amounts and numbers of HCPs quantified (Figure 61.A). It is noteworthy that despite a less comprehensive library, due to a lower number of protein identifications, the hybrid library allowed the quantification of more HCPs for almost all samples. Looking further into the results, we observed a slight improvement of the extraction reproducibility with a median of the CVs calculated on peptides intensities within replicates of 9.4 for the hybrid library and 9.6 for the mAb A33 spectral library (Figure 61.B). The PYGM/ADH

ratios obtained with the hybrid library are also more accurate around the expected value of 0.2 (Figure 61.C). The addition of the DIA data to the mAb A33 DDA search archive provided iRT values from another source, including the analyses of interest themselves. The source-specific iRT calibration, implemented in Spectronaut thus allows the use of the iRT calibration from best source possible. As a result, we can conclude that the hybrid spectral library implemented with the DIA runs of interest has improved the extraction of peptide signals.

However, the presence of iRT values from different sources may also generate signal extraction issues in some cases in a hybrid approach. This phenomenon has been observed with the CCCF fractions of mAb 3 from the 2000L and 15 000L bioreactors and the CCCF fraction of mAb 4 from the 2000L bioreactor in our dataset (Figure 62). Those three samples showed a lower than normal PYGM/ADH ratio. A closer look at the PYGM peptides revealed partially extracted signals, while the use of the DDA-based library only allowed for complete signals extraction (Figure 62.A and Figure 62.B). The same issue was observed for the three samples that showed lower retention times and lower reproducibility between replicates. The use of a correction factor of 1.5 applied to the XIC extraction window dynamically chosen by Spectronaut allowed us to overcome the problem. This led to the same signals extractions as using the DDA-based library (Figure 62.C). Also, the application of the correction factor allowed the quantification of approximately 200 additional HCPs. The use of this parameter on the other six samples gave random results with a gain or loss in the numbers of HCPs quantified. Therefore, we decided to apply this correction factor only to the three samples mentioned. Further investigation is needed to determine if this parameter should be systematically applied to all DIA analyses.

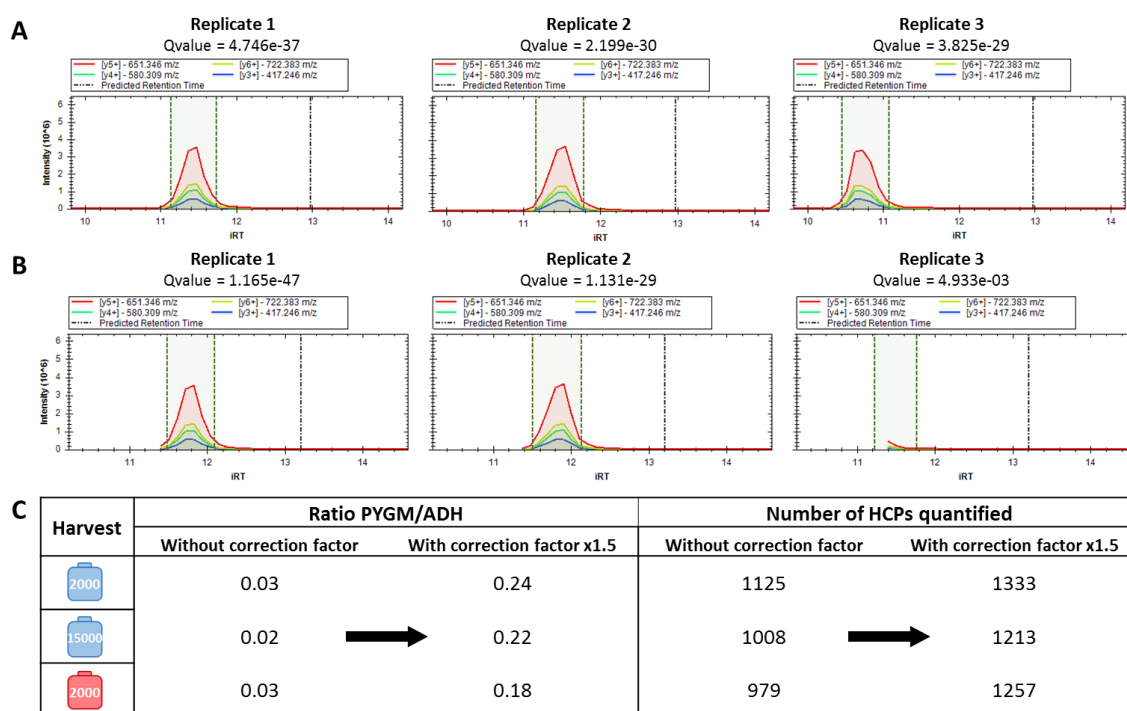


Figure 62: Signal extraction issues observed with the hybrid spectral library. Doubly charged PYGM peptide 'FAAYLER' MS2 signals extraction for the three replicates of the mAb 4 CCCF fraction from the 2000 L tank obtained using the DDA-based spectral library (A) and the hybrid library (B). Impact of the correction factor of 1.5 applied to the XIC extraction window dynamically chosen by Spectronaut for the mAb 3 CCCF fractions from 2000L and 15 000L tanks and mAb 4 CCCF fraction from a 2000 L bioreactor (C).

In conclusion, the use of the hybrid spectral library to target HCPs in our analyses improved reproducibility and accuracy of quantification without tedious experimental labor and extra acquisition time. In our case, the addition of iRT values from the DIA runs of interest in the hybrid library allowed the correction of the XIC extraction windows to best fit peptide elution. After the benchmarking of the hybrid library against a classical DDA-based library, we compared our results against the spectral library free-based approach of Spectronaut, namely directDIA.

B. Evaluation of a spectral library free-based strategy

When the intention is to implement a new method in a biopharmaceutical environment, the objective remains the same: the method must be straightforward and allow a maximum recovery and an accurate quantification. A first step was reached with the use of a hybrid spectral library, which showed an improved accuracy of quantification as well as a higher number of HCPs quantified. However, an approach that would not need the use of a spectral library still remains in theory the most appropriate method to extract data from our four antibody samples as it prevents any concerns about the absence of HCPs from the spectral library. Therefore, the directDIA approach was evaluated against the use of the hybrid library. The overall amounts, numbers of HCPs quantified, CVs and PYGM/ADH ratio were benchmarked (Figure 63).

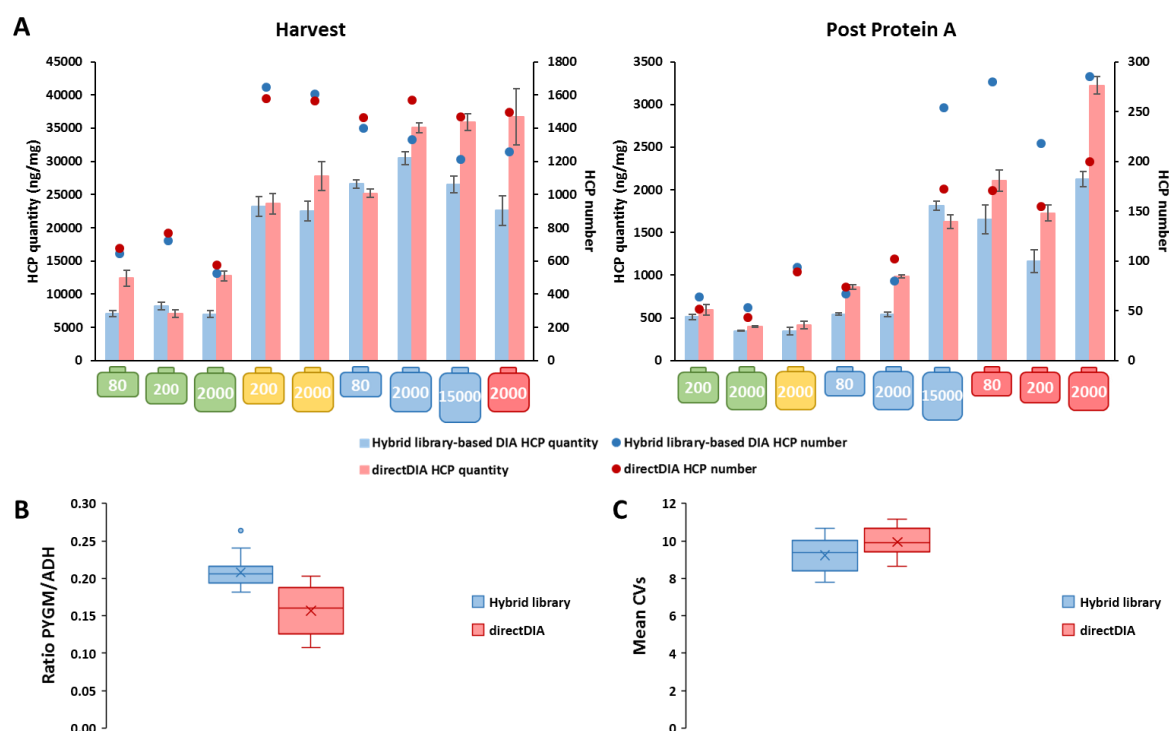


Figure 63: Benchmarking of the hybrid spectral library-based DIA against directDIA. Representation of the overall HCP quantities obtained and numbers of HCPs quantified by the hybrid and directDIA methods (A). The dots represent the numbers of HCPs quantified. Bar heights represent the mean of the global HCP amount in injection triplicates. Error bars represent the standard deviation. Distribution of the average CVs calculated on the HCP peptides intensities among replicates for the 18 samples (B). Dispersion of PYGM/ADH ratios obtained with both extraction strategies (C). mAb 1 (green tank), mAb 2 (yellow tank), mAb 3 (blue tank) and mAb 4 samples (red tank).

When we look at the nine harvest samples, we notice that directDIA allowed us to quantify 9 to 179 more HCPs than using the hybrid spectral library (Figure 63.A). On the contrary, a higher number of HCPs were quantified within PPA samples using the hybrid library-based approach with the exception of the mAb 3 PPA sample from 80L tank. These differences of HCPs contents can also be seen at the PYGM Top3 peptides level (Table 8).

Harvest	Hybrid library - Top3 PYGM	directDIA - Top3 PYGM
80	AWEVTVK	AWEVTVK
	FAAYLER	FAAYLER
	VLVDLER	VLVDLER
200	FAAYLER	FAAYLER
	GLAGVENVTELKK	GLAGVENVTELKK
	VLVDLER	VLVDLER
2000	FAAYLER	FAAYLER
	GLAGVENVTELKK	GLAGVENVTELKK
	VLVDLER	VLVDLER
200	AWEVTVK	AWEVTVK
	FAAYLER	FAAYLER
	VLVDLER	VLVDLER
2000	AWEVTVK	FAAYLER
	FAAYLER	GLAGVENVTELKK
	VLVDLER	VLVDLER
80	FAAYLER	FAAYLER
	GLAGVENVTELKK	GLAGVENVTELKK
	VLVDLER	VLVDLER
2000	FAAYLER	FAAYLER
	GLAGVENVTELKK	GLAGVENVTELKK
	VLVDLER	VLVDLER
15000	FAAYLER	FAAYLER
	GLAGVENVTELKK	GLAGVENVTELKK
	VLVDLER	VLVDLER
2000	AWEVTVK	AWEVTVK
	FAAYLER	FAAYLER
	VLVDLER	VLVDLER

Sample	Hybrid library - Top3 PYGM	directDIA - Top3 PYGM
200	AWEVTVK	AWEVTVK
	FAAYLER	FAAYLER
	VLVDLER	GLAGVENVTELKK
2000	AWEVTVK	AWEVTVK
	FAAYLER	FAAYLER
	VLVDLER	VFADYEEYVK
2000	AWEVTVK	AWEVTVK
	FAAYLER	FAAYLER
	VLVDLER	LITAIGDVVNHDPPVVGDR
80	AWEVTVK	AWEVTVK
	FAAYLER	FAAYLER
	VLVDLER	VFADYEEYVK
2000	FAAYLER	AWEVTVK
	GLAGVENVTELKK	FAAYLER
	VLVDLER	GLAGVENVTELKK
15000	FAAYLER	FAAYLER
	GLAGVENVTELKK	GLAGVENVTELKK
	VLVDLER	VFADYEEYVK
80	FAAYLER	FAAYLER
	GLAGVENVTELKK	GLAGVENVTELKK
	VLVDLER	VFADYEEYVK
200	FAAYLER	FAAYLER
	VFADYEEYVK	LITAIGDVVNHDPPVVGDR
	VLVDLER	VFADYEEYVK
2000	FAAYLER	FAAYLER
	VFADYEEYVK	LITAIGDVVNHDPPVVGDR
	VLVDLER	VFADYEEYVK

Table 8: PYGM Top3 peptides obtained using the hybrid library-based DIA and directDIA methods. List of the three most intense peptides of PYGM extracted in each samples by hybrid library-based DIA or directDIA approaches. Peptide sequences in red represent peptides that differ between the two Top3 methods. mAb 1 (green tank), mAb 2 (yellow tank), mAb 3 (blue tank) and mAb 4 samples (red tank).

The peptides used within the harvest samples were identical for the majority of samples. An average of 1.02 was obtained when comparing the response factors obtained by the hybrid library and the directDIA strategies. However, the Top3 peptides of PYGM used within the PPA samples were not the same with at least one peptide differing. One can highlight the absence of the peptide "VLVDLER" which was not identified by the directDIA method while it was used for all PYGM Top3 of the hybrid library approach. As a result, the PYGM/ADH ratios of the PPA samples are less accurate with a median of 0.13 obtained by directDIA and 0.20 with the hybrid library (Figure 63.B). Our results confirm the limitations of the approach without spectral library for the extraction of low abundant signals that was observed in the previous chapter. In addition, directDIA did not improve the reproducibility of the extraction with a median of CVs calculated on peptide intensities within replicates of 9.9 compared to the median of 9.4 obtained for the hybrid library (Figure 63.C). Finally, we kept the results of the hybrid library-based strategy to compare the HCPs contents of the samples from the four mAbs.

C. HCP profiles characterization

To begin the study, we compared the overall HCPs quantities and the numbers of HCPs quantified in the CCCF and PPA fractions from the production of four mAbs (Figure 64).

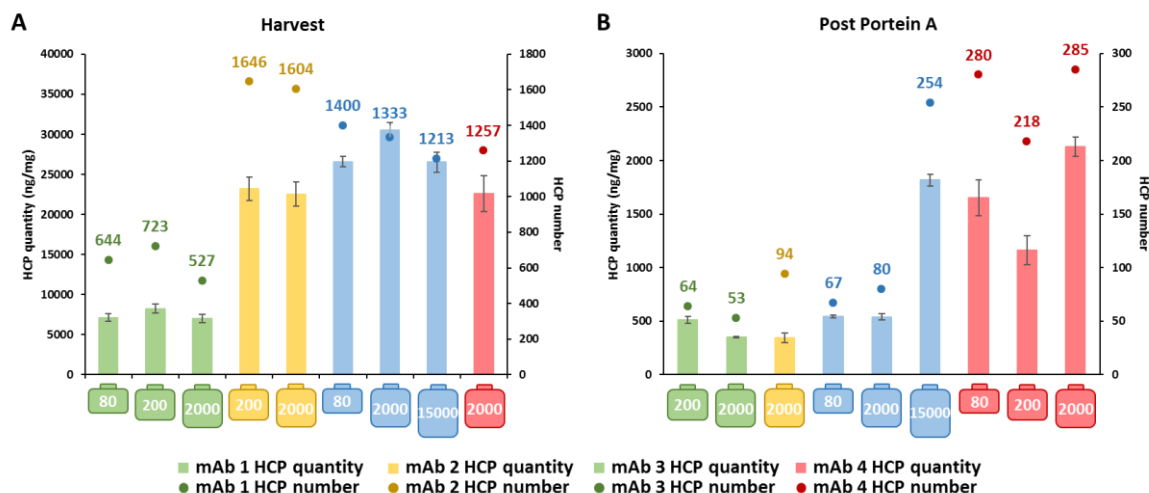


Figure 64: Evaluation of the bioreactor capacity's impact on the HCP profile. Global HCP amounts and numbers of HCPs quantified obtained for CCCF (A) and PPA (B) fractions using Top3-DIA strategy. Bar heights represent the mean of the global HCP amounts in injection triplicates. Error bars represent the standard deviation. The dots and values represent the numbers of HCPs quantified.

At the end of the USP, we observed similar quantification results for harvest samples from the same mAb production (Figure 64.A). Focusing on the impact of the bioreactor capacity, no trend emerged. Samples from mAb 1 and 2 (green and yellow) displayed a decrease of HCP impurities when the culture was performed in a 2000L bioreactor. In contrast, the mAb 3 CCCF fraction (blue) produced in the 2000L bioreactor showed the highest amount of impurities. The low variation obtained in fractions from the same mAb production does not allow to conclude on the impact of the bioreactor capacity. In addition to the overall quantification results, we performed the comparison of the quantified HCPs lists to study the profile of the co-expressed HCPs in CCCF fractions (Figure 65).

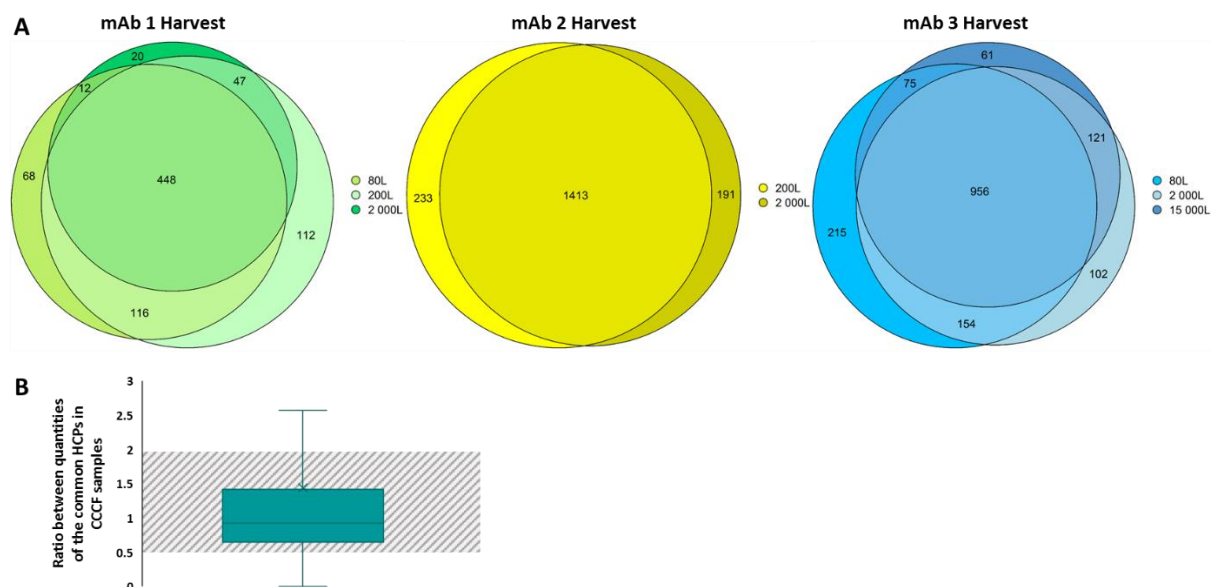


Figure 65: HCPs profiles in fractions from the same mAb production. Venn diagrams of the HCPs quantified in CCCF fractions from the same mAb (A). Distribution of the ratios between the estimated amounts of common HCPs in samples from the same mAb using the Top3-DIA method (B). Out of range ratios were removed to observe the 74% of 5357 common HCPs within a factor 2 (grey area).

Venn diagrams showed that around 64% of the co-expressed HCPs are common in samples from the same mAb production (Figure 65.A). Looking further at the individual HCP quantities, we saw that 74% of the 5357 ratios between common HCP amounts are within a factor 2 (Figure 65.B). The results are consistent within the same mAb project and support that the majority of co-expressed HCPs are common with similar amounts. Therefore, the process scale up has a low impact on the HCP profile. The variations are more likely to come from small changes in the culture media, the incubation temperature or a concentration effect.

In a second step, we compared all samples together. The mAb 1 fractions showed about twice less quantified HCPs compared to the three other mAbs that are in the same range of contaminants. This trend was supported by the results obtained for the PPA fractions (Figure 64.B). Co-eluted HCPs during the DSP are also impacted by the mAb of interest. While the mAb 1, mAb 2 and mAb 3 showed similar HCP amounts, the PPA fractions of the mAb 4 (red) displayed about 3-fold more HCPs. Thus, instead of an impact of the bioreactor, our results rather suggest an effect of the biotherapeutic of interest on co-expressed HCPs. The venn diagrams of the lists of quantified HCPs within each mAb-project also point in the same direction (Figure 66).

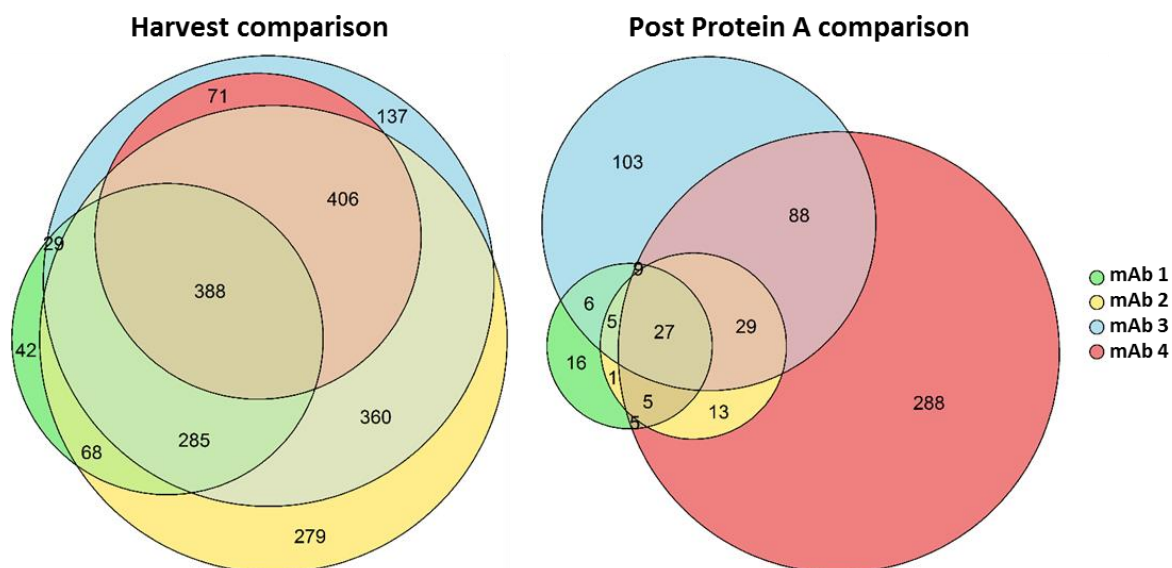


Figure 66: Inter-mAb type HCPs profile comparison. Overlay of the HCPs quantified in CCCF fractions from the same mAb.

With the overlap of CCCF fractions, we notice 54 to 199 HCPs unique to a mAb project. The PPA samples confirm this observation with 7 to 174 HCPs unique to a mAb production. This highlights the importance of using a spectral library specific to each culture process knowing that it will lead to the expression of a specific set of proteins.

In addition, the mAb 3 fraction from a 15 000L bioreactor showed a HCPs content of 1016 ng/mg while the other two fractions from 80 and 200L tanks have an overall HCPs amount of 488 and 495 ng/mg, respectively (Figure 66.B). Twice the amount of other fractions with more than 90 additional HCPs can have serious consequences. Although no trend was observed in relation to an impact of the culture process scale up, the HCPs content of the samples should be monitored at different stages of the mAb manufacturing process in order to understand the impact of each culture or purification steps. Over time, this will help to understand and avoid most of the issues related to HCPs.

4. Conclusion

The study of the fractions from the production of the four antibodies allowed us to address one of the crucial points of the DIA acquisition mode. Indeed, the peptide-centric approach requires the use of a complete and specific spectral library in order to take full advantage of the DIA strategy. Interestingly, the same trend was observed between the overall amounts of the different samples whether it was estimated using a DDA-based library, a hybrid library or the directDIA approach to extract peptide signals.

The use of a hybrid spectral library that combines the identification results of DDA and DIA analyses has shown great potential. It allowed deep coverage of HCPs while maintaining iRT-precision and FDR control at different stages of the DIA search. Although it is always recommended to fractionate a non-transfected mock cell line and project-specific CCCF samples to generate the most comprehensive spectral library, it will be possible to implement this library with the runs of interest to obtain a project-specific library allowing an improved iRT precision. This can be done until an optimization of the

directDIA approach that still showed limitation in sensitivity compared to the approach with spectral library.

The ultimate aim of the study was to observe an impact of the production scale up during the development of a new therapeutic. However, the small variations between samples did not bring any conclusion. Our results, in agreement with previous studies^{22-23, 296}, showed the impact of the mAb type produced on the HCPs profile. Even if the samples are from the same CHO cell line, special attention must be given to generate a specific spectral library for each biotherapeutic. This library combined with the accumulation of MS analyses will create a database allowing a better understanding of each culture and purification steps leading to the safest final product.

Chapter 4

Benefit of ion mobility in the field of HCPs monitoring

1. Presentation of the TimsTOF PRO instrument

Following the arrival of the TimsTOF PRO in the laboratory in 2019, we were interested to see the performance of this system for the identification and quantification of HCPs. More precisely, two acquisition modes specific to the TimsTOF PRO were evaluated: DDA and DIA acquisition combined with the Parallel Accumulation-Serial Fragmentation approach (ddaPASEF and diaPASEF).

As mentioned in the bibliographic introduction, the TimsTOF PRO is equipped with a dual ion mobility cell. The constant buffer gas flow will drive the ions into the first TIMS cell. They will be accumulated and retained by the application of a static electric field. Once the ions are accumulated, they are transferred to the second cell where they are separated according to their ionic mobility by a slow decrease of the electric field. With the same charge, the ions with a higher mass and surface will be located near the exit of the mobility cell while the more mobile ions will remain near the entrance (Figure 67).

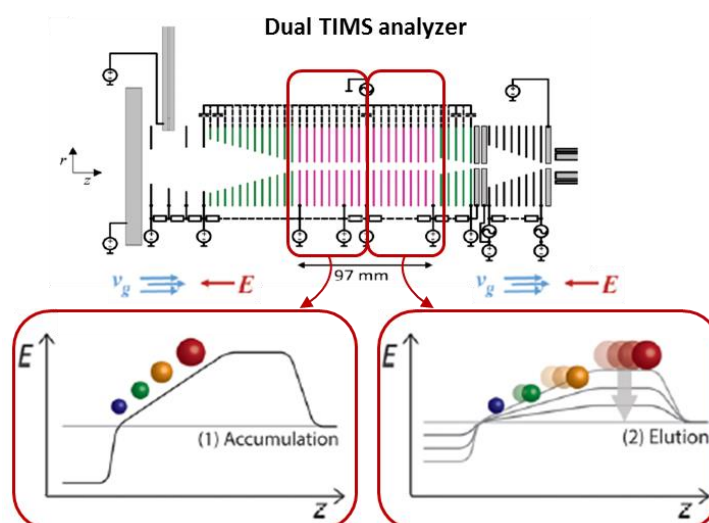


Figure 67: Principle of Parallel Accumulation occurring in the dual TIMS cell (adapted from ²⁶).

The time during the accumulation of ions in the first TIMS cell is called accumulation time. It can be fixed or variable. The time during which the ions are separated and eluted is called the ramp time. It can be set to be less or equal to the accumulation time. Thus, a duty cycle of 100% can be achieved ensuring no ions loss.

The mobility cell of the TimsTOF PRO will therefore act as an additional separation dimension for co-eluted ions from the nanoLC. In addition, the TimsTOF PRO features a quadrupole synchronized with the ion elution from the mobility cell. Thus, a rapid switch of the quadrupole mass position allows the selection of different precursor ions at different m/z during their TIMS elution time. This acquisition mode, which combines the dual mobility cell and quadrupole synchronization, is called PASEF²⁶ (Parallel Accumulation - Serial Fragmentation). It allows full use of the target ions for fragmentation (Figure 68).

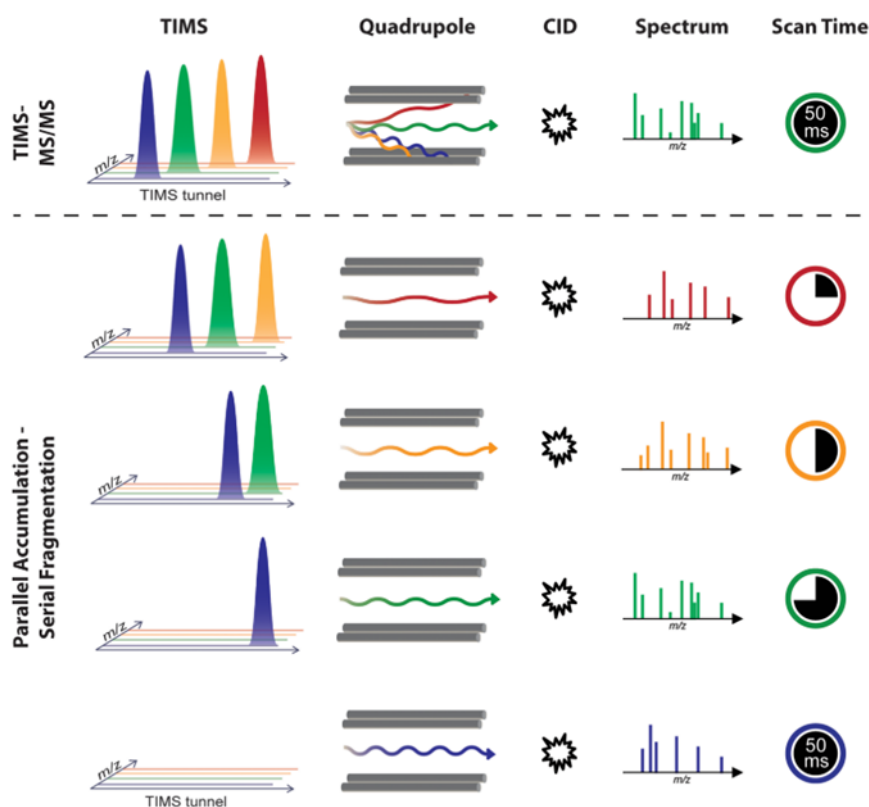


Figure 68: Principle of the PASEF mode (from ²⁶). The upper panel represents the standard TIMS-MS/MS operation mode. One precursor ion is selected from a single TIMS scan, while the others are discarded. The PASEF mode allows full use of target ions for fragmentation by rapid switching of the quadrupole mass position to select the precursor during its TIMS elution time.

PASEF allows to exploit the speed capacity of the Q-TOF analyzer and to reach frequencies higher than 100Hz. It can be coupled with DDA or DIA acquisition modes. Figure 69 represents the progress of a ddaPASEF analysis with the time scale of each step.

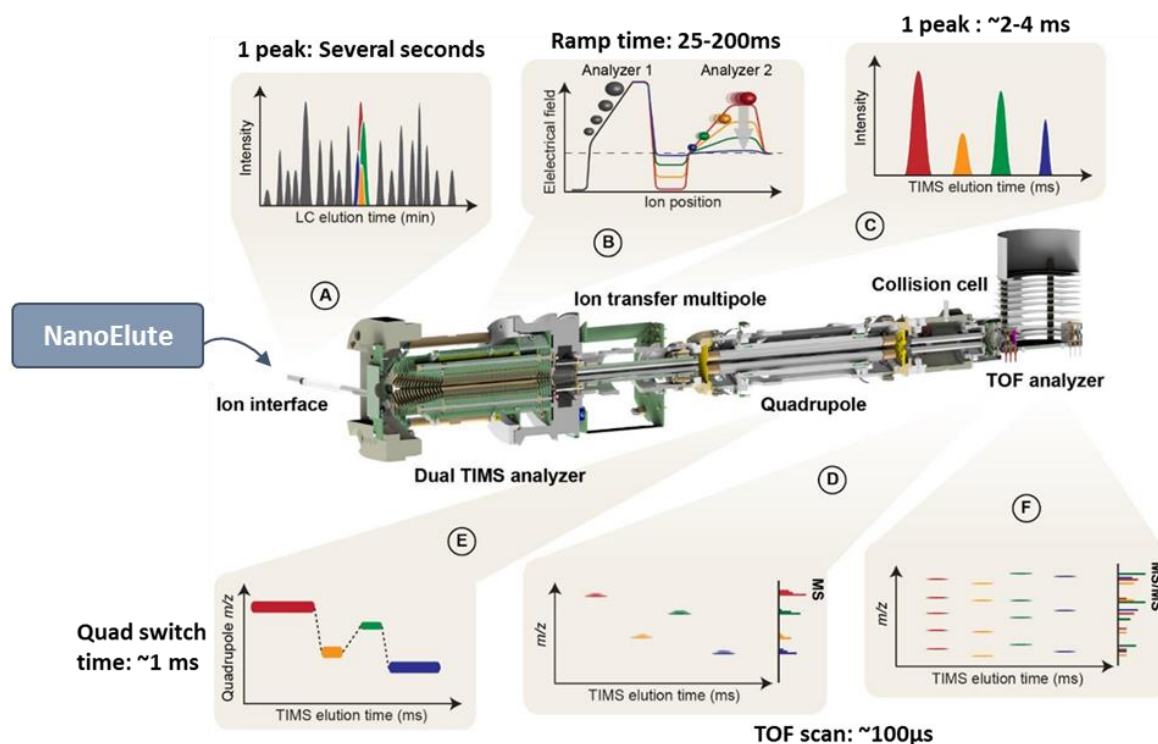


Figure 69: Scheme of the ddaPASEF acquisition mode on the TimsTOF PRO instrument (adapted from ²⁵).

Nevertheless, ddaPASEF mode will still suffer from the limitations of DDA acquisition. Therefore, a DIA approach specific to TimsTOF, called diaPASEF²¹⁵, was developed to overcome the undersampling and lack of reproducibility inherent to DDA. As with a conventional DIA approach, the precursor ion groups eluted from Tims will be co-isolated and co-fragmented in small m/z windows covering the entire mass range. The main difference is that the m/z isolation windows are defined in two dimensions, ion mobility and m/z (Figure 70). Finally, the DIA mode will benefit from the acquisition speed of the instrument, the noise reduction and the separation of co-eluted ions from the LC thanks to ion mobility.

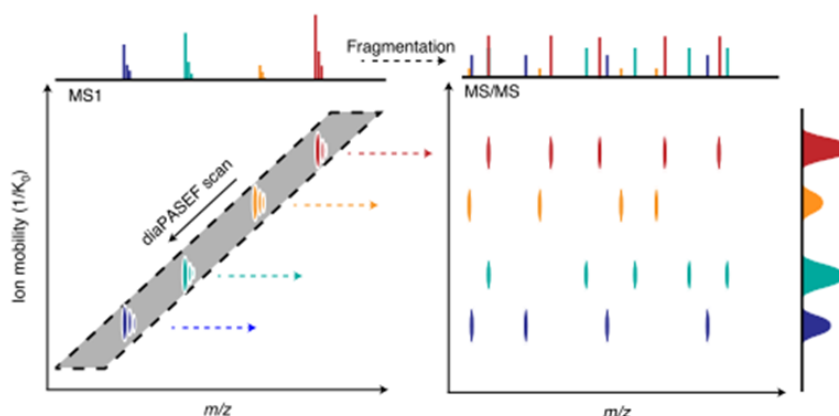


Figure 70: Principle of diaPASEF (from ²¹⁵). The quadrupole isolation window (gray) is dynamically positioned according to ion mobility (arrow). In a given Tims scan, ions from the selected mass ranges are fragmented to record ion mobility resolved MS/MS spectra of all precursor ions.

In order to evaluate the performance of TimsTOF PRO in the context of HCPs, we analyzed the sample series produced by Dr. Gauthier Husson supplemented with our previously introduced HCP Profiler standard. These 11 samples were used in the development of the DIA method on the Orbitrap Q Exactive HF-X (Part I - Chapter 2). First, the ddaPASEF and diaPASEF methods were developed to investigate the identification and quantification capabilities of the TIMS-QTOF instrument. The HCPs quantification results were compared to evaluate its sensitivity, reproducibility and accuracy of quantification. Then, the performance of the TimsTOF PRO and the Q Exactive HF-X were benchmarked to assess the contribution of TIMS to the characterization of HCP impurities.

2. Analytical strategy

The analytical workflow is presented in Figure 71.

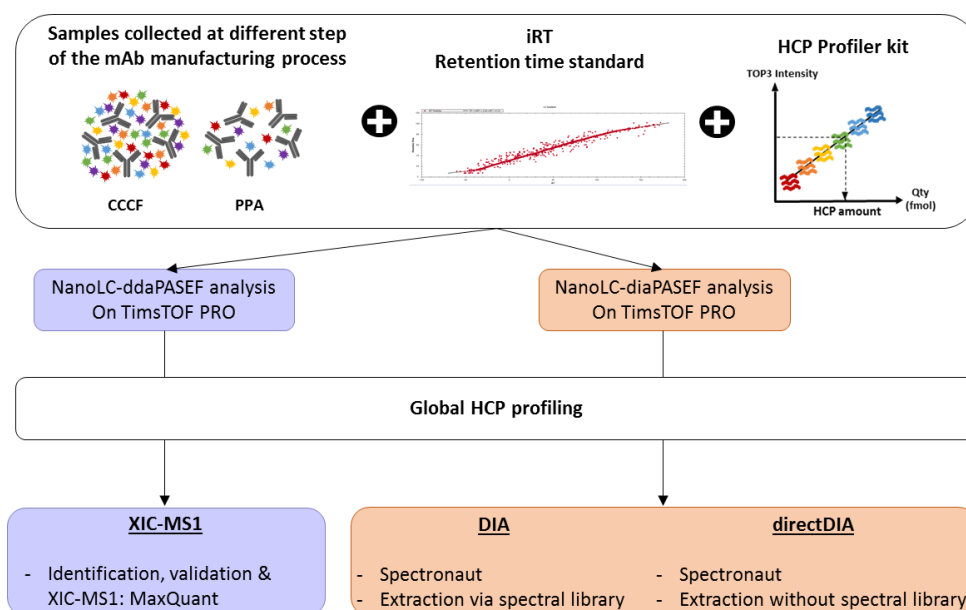


Figure 71: Evaluation of the TimsTOF PRO instrument for HCPs characterization. Four mAbs CCCF and PPA fractions were in-gel digested and spiked with iRT retention time standards (Biognosys) and a bead of the HCP Profiler standard. Samples were injected in triplicate using ddaPASEF and diaPASEF acquisition modes on a nanoElute-TimsTOF PRO (Bruker) system. The ddaPASEF, diaPASEF and directDIA-PASEF quantification results were benchmarked to assess the performances of the Tims-QTOF instrument.

Samples produced by Dr. Gauthier Husson, collected after harvest procedure (4 samples) and after Protein A affinity chromatography (7 samples) as described in [288], were in-gel digested using trypsin. Retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) and a bead of the HCP Profiler standard were spiked. Samples were analyzed in triplicate on a nanoElute-TimsTOF PRO (Bruker) system in ddaPASEF and diaPASEF modes.

First, peptides and proteins were identified from ddaPASEF analyses using Andromeda search engine integrated in MaxQuant software. Dataset validation and XIC-MS1 quantification were also performed in MaxQuant. Next, the Top3 strategy¹⁶ was applied to derive HCP amounts using the 6 points calibration curve of the HCP Profiler standard.

Then, Pulsar search engine, integrated in Spectronaut™, was used to generate a mAb A33 project-specific spectral library using ddaPASEF runs of a CCCF CHO-DG44 null cell line sample fractionated (24 analyses including iRT retention time standards from Biognosys) combined with all ddaPASEF analyses of the CCCF and PPA mAb A33 samples spiked with the iRT standard and HCP Profiler kit. The generated spectral library contains 56 818 peptides from 5826 protein groups. Moreover, Spectronaut™ was used to extract MS2 signals from the diaPASEF data using the project-specific spectral library or the directDIA™ algorithm. For both extraction methods, IM and RT extraction windows were adjusted dynamically based on a large sample set during calibration. A Q-value filter of 0.01 was applied and iRT profiling was set on. Finally, HCP amounts were estimated using the Top3 HCP Profiler strategy.

Sample preparation, LC-MS/MS methods and data treatment are detailed in the experimental part, section 1.D.

3. Results

A. DIA-PASEF for in-depth characterization of HCPs

First, we used the ddaPASEF parameters optimized by Chloé Moritz, a PhD student in the laboratory, using human cervical cancer cells (HeLa) as reference sample. The accumulation and ramp times were set to 100 ms. Thus, one full frame (MS1 spectrum) followed by 10 PASEF frames, containing an average of 12 MS/MS spectra, were acquired resulting in a cycle time of 1.1 s. This means that in 1.1 s, the TimsTOF PRO can acquire one MS spectrum followed by up to 120 MS/MS spectra (Figure 72). MaxQuant software was used for data processing of the ddaPASEF analysis thanks to its improvements, which exploit the added data dimension for feature detection, mass recalibration, alignment and matching between runs²⁹⁷.

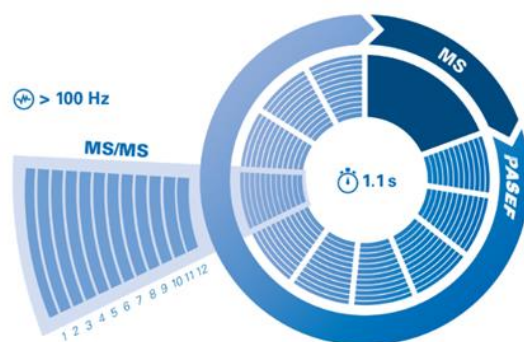


Figure 72: Schematic representation of a PASEF frame.

Next, based on the initial publication of the diaPASEF²¹⁵ mode on complex proteomes, we decided to use a DIA method composed of 64 windows of 25 m/z. Those windows are covered in 16 TIMS cycle of 100 ms. Concretely, four m/z windows are selected and fragmented in a TIMS elution. The windows overlap in the ion mobility dimension to reduce potential artifacts linked to reduced ion transmission at the edges of the diaPASEF windows. After data acquisition, Spectronaut software was used to process the data from the diaPASEF analysis.

As it was done on the Q Exactive HF-X, we applied our HCP workflow on the TimsTOF PRO. We started with 400 ng protein injections. However, the total ion current (TIC) of the first injections showed signs of detector and/or column saturation. Thus, we had to reduce the amount of material to be injected. Finally, 273 ng of protein were loaded onto the column for the 11 samples studied and the HCP Profiler standard peptides were in a range between 0.68 and 341 fmol. Next, the overall amounts, the number of HCPs quantified, the intra-sample dynamic range and the extraction repeatability were investigated (Figure 73).

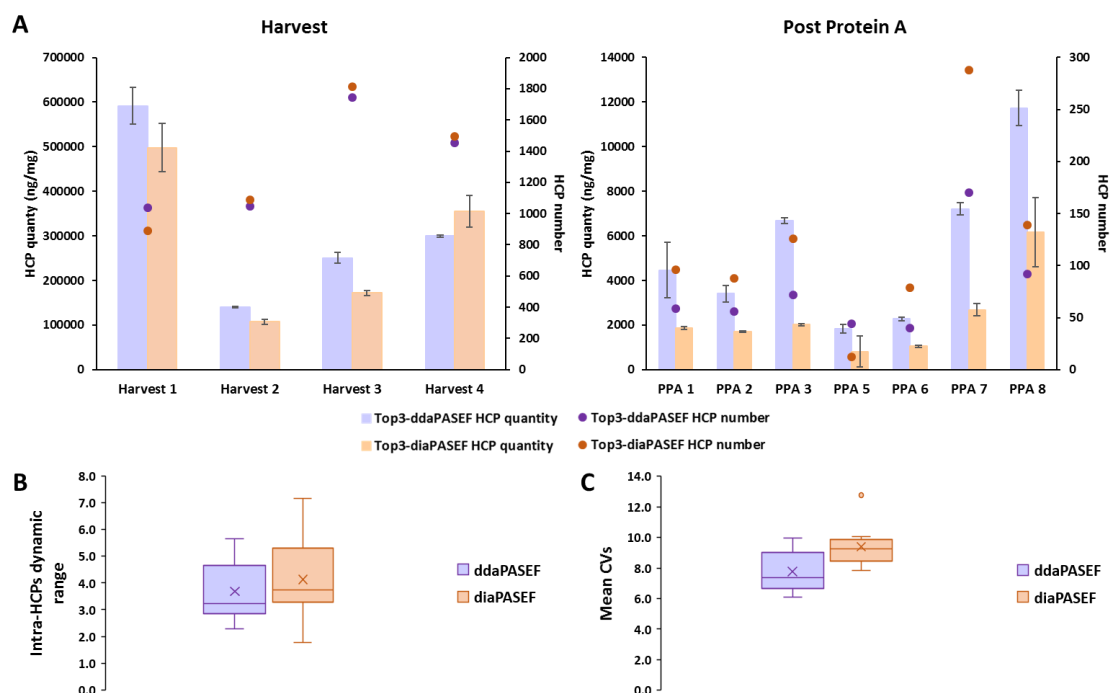


Figure 73: Evaluation of the spectral library-based diaPASEF strategy against MS1-XIC ddaPASEF for global HCP profiling using the HCP Profiler standard. Global HCP amounts and numbers of HCPs quantified obtained for PPA and CCCF fractions using MS1-XIC ddaPASEF and MS2-based diaPASEF strategy (A). Bar heights represent the mean of the global HCP amount in injection triplicates. Error bars represent the standard deviation. Representation of the intra-HCPs dynamic range observed using DDA and DIA acquisition modes (B). Comparison of the mean CVs calculated on HCP peptides intensities within replicates for ddaPASEF and diaPASEF (C).

The quantification results, presented in Figure 73.A, displayed a higher coverage of HCPs with the diaPASEF approach for most samples. Between 32 and 118 additional HCPs were obtained with diaPASEF. Two exceptions were noted, PPA 5 and Harvest 1 fractions showed a lack of repeatability within replicates. As a result, application of the CV filter on the repeatability of MS2 signals led to the loss of several peptides. In accordance with a higher coverage, the diaPASEF strategy achieved between 1.8 and 7.2 orders of magnitude between the least and most abundant HCP compared to 2.3 and 5.7 for the ddaPASEF method (Figure 73.B). A higher dynamic range as well as an increased sensitivity were obtained with the diaPASEF approach. This increased sensitivity was also observed for the standard peptides of the HCP Profiler kit. Indeed, the 54 standard peptides were identified and quantified by diaPASEF while the ddaPASEF approach did not allow signal extraction of one standard peptide from the lowest calibration point at 0.68 fmol for the four CCCF fractions.

Nevertheless, despite a slight increased sensitivity, the diaPASEF strategy displayed a lower overall amount of impurities for almost all samples. Moreover, the quantification accuracy obtained with a median of 9.2 was lower than that of the ddaPASEF method which was about 7.8 (Figure 73.C). Looking further at the signals of the standard HCP Profiler peptides, we observed issues with the extracted signals from diaPASEF data (Figure 74).

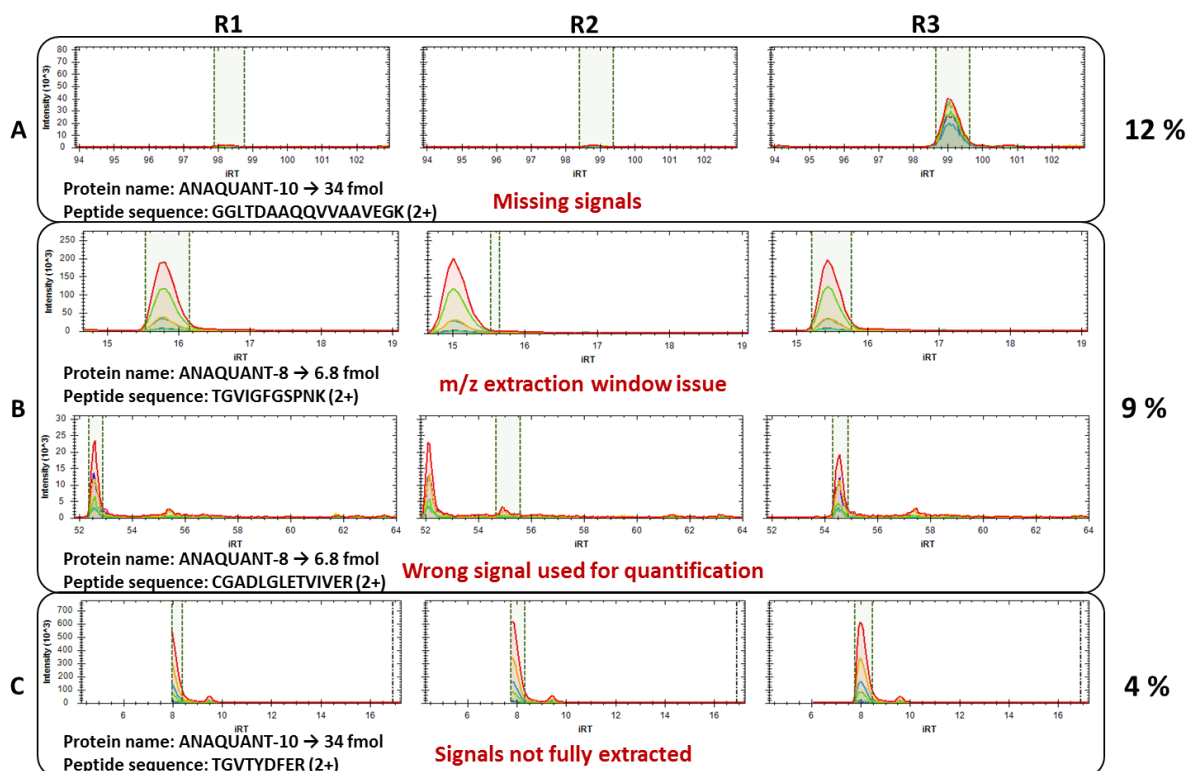


Figure 74: Example of the issues observed on the HCP Profiler standard peptide extracted signals from diaPASEF data. The problems can be classified into three categories: (A) Presence of unextracted signals, (B) Misplaced m/z extraction windows, and (C) Presence of partially extracted signals. Each spectrum represents the extracted MS2 signals of the y- and b- ions of a given peptides.

DIA data processing led to the signal extraction of 78 precursor ions from the standard peptides in each sample. Among these precursor ions checked, 25% showed problems related to data extraction. They can be classified in three categories: 12% with missing signals for one or two replicates (Figure 74.A), 9% with a wrong signal used for quantification (Figure 74.B) and 4% with the presence of partially extracted signals (Figure 74.C). These issues have a direct impact on the calibration curves and, in our case, lead to the underestimation of the HCPs amounts. For the second category (Figure 74.B), the extraction window can be manually adjusted to use the desired peptide signals. On the contrary, when the signals are missing (Figure 74.A) or partially extracted (Figure 74.C) a new search must be performed. The use of a correction factor for the m/z extraction window can help extracting the signals, as discussed in the previous chapter (Part I, chapter 3, section 3.A). However, this is not always the case and on this dataset, it did not allow recovering all missing signals.

Finally, the additional separation dimension complicates the signal extraction and requires the use of dedicated algorithms. In addition, the link between a precursor ion and its fragments is lost with the DIA approach. Thus, minor variations in LC or TIMS elution times between replicates can lead to incorrect signal extraction. Current software solutions do not yet allow to overcome some of these problems and they continue to be optimized on a case by case basis. In the future, they will allow the reliable extraction of signals while limiting the number of missing values.

B. diaPASEF combined with spectrum-centric data extraction

Following the evaluation of the diaPASEF method combined with an extraction method based on a spectral library, we investigated the directDIA approach implemented in Spectronaut (Figure 75).

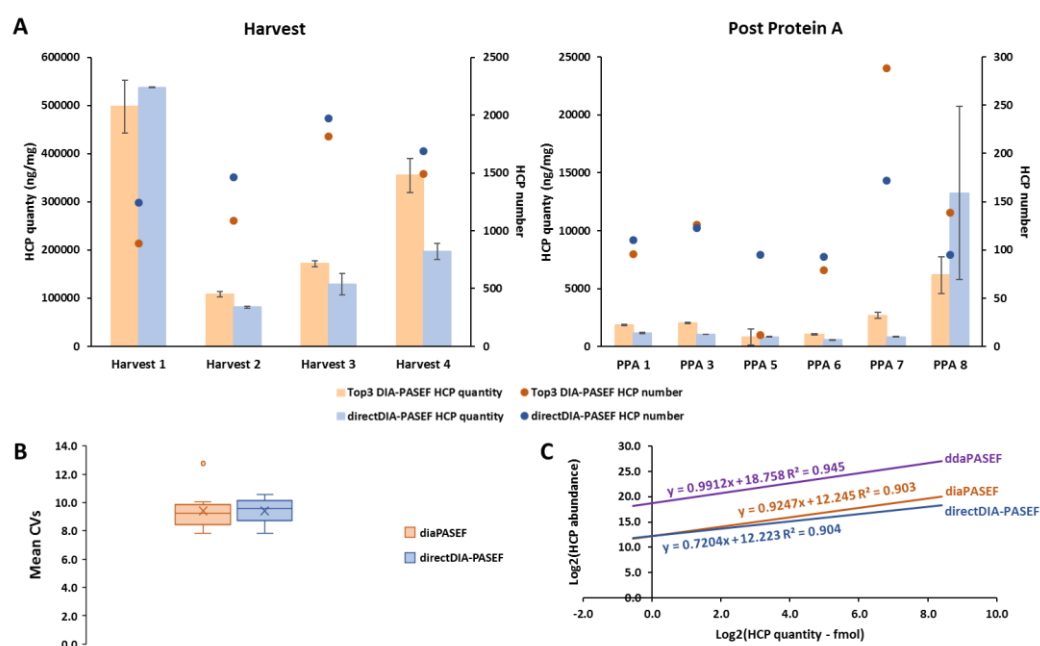


Figure 75: Benchmarking of the diaPASEF against directDIA-PASEF extraction strategies. Global HCP amounts and numbers of HCPs quantified obtained for PPA and CCCF fractions using spectral library-based diaPASEF and directDIA-PASEF strategy (A). Bar heights represent the means of the global HCP amounts in injection triplicates. Error bars represent the standard deviation. Comparison of the mean CVs calculated on HCP peptides intensities within replicates for diaPASEF and directDIA-PASEF (B). Superposition of the calibration curves obtained using ddaPASEF, diaPASEF and directDIA-PASEF methods (C). The calibration curves presented were obtained by averaging the slope and intercept of the 33 curves of each method.

Whether in terms of the numbers of HCPs quantified, the overall HCPs amounts, or the repeatability of the signal extractions, similar results were obtained with both approaches (Figure 75.A and Figure 75.B). However, as shown in Figure 75.C, comparable calibration curves were obtained by both approaches. Thus, the directDIA approach displays the same difficulties regarding the signal extraction and more particularly at the level of standard peptides. Finally, our results also highlight the need for optimization of the search algorithm to improve reliable signal extraction of the PASEF mode combined with DIA acquisition.

C. Benchmarking of diaPASEF against DIA on a Q-Orbitrap instrument

In the previous sections, we have seen that the diaPASEF approach allows to increase the coverage of HCP compared to a ddaPASEF method. Even though the search algorithm for the extraction of reliable quantitative signals is not yet mature, we wanted to see if the performance of the TimsTOF PRO in terms of peptide and protein identifications surpasses the results obtained on the Q Exactive HF-X (Figure 76).

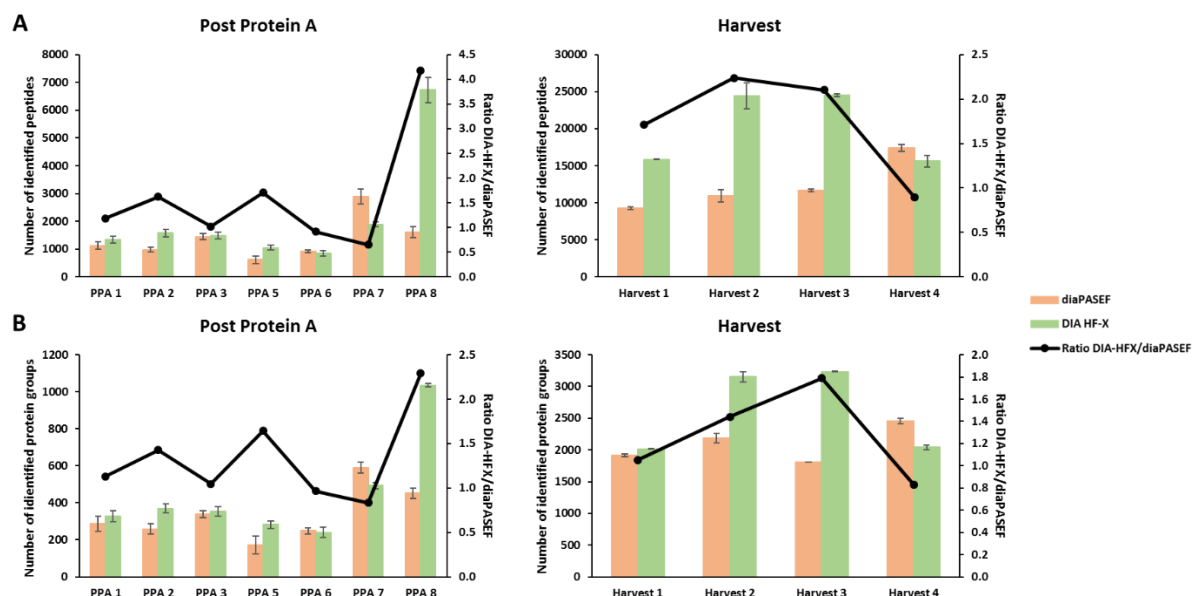


Figure 76: Benchmarking of the identification performance of the diaPASEF approach against the DIA method performed on the Q Exactive HF-X. Numbers of peptides identified in PPA and CCCF fractions using spectral library-based diaPASEF and DIA HF-X strategies (A). Number of protein groups identified in PPA and CCCF fractions using spectral library-based diaPASEF and DIA HF-X strategies (B). Bar heights represent the mean of the numbers of identifications in injection triplicates. Error bars represent the standard deviation. The straight lines with marker represent the ratio of the values obtained with the Q Exactive HF-X over the one obtained with the TimsTOF PRO.

The comparison illustrated in Figure 76 is actually not really fair. Indeed, the amount of proteins injected is not the same on the TimsTOF PRO and the Q Exactive HF-X, as 273 ng and 400 ng were injected, respectively. In addition, the samples injected on both instruments are from different aliquots of the same samples frozen in 2015. This may explain the large difference in peptides and protein groups identified for the PPA 8 sample. Nevertheless, the rest of the samples show results close to those of the Q Exactive HF-X with 1.5 times less material injected. It is reasonable to assume that the TimsTOF PRO can exceed the performance of the Orbitrap HF-X due to its acquisition speed. Indeed, the diaPASEF method allows the acquisition of one MS spectrum followed by 64 MS/MS spectra in about 2 seconds while the DIA method developed on the Q Exactive HF-X allows to perform one MS scan followed by 40 MS/MS in 3 seconds. These promising results demonstrate the power of the TimsTOF PRO that, combined with an adapted and optimized extraction tool, will allow, in addition to a deeper coverage, to obtain a reliable quantification of the HCP content in samples from various steps of the mAb manufacturing process.

4. Conclusion & Perspectives

The integration of the TIMS device at the entrance of a Q-TOF instrument provides an additional separation dimension following the elution from the LC. Consequently, the ddaPASEF and diaPASEF acquisition methods hint at a deeper coverage of protein impurities in our mAb samples without additional sample preparation.

Unfortunately, the TimsTOF PRO in its current version shows an increased sensitivity that limits the amount of material that can be injected. This limitation represents a drawback in the context of HCPs, i.e., the identification and quantification of impurities at trace level compared to the mAb present in large amounts in the samples. Since our injections, a new ion mobility cell, called SRIG (Stacked Ring Ion Guide), was installed on our instrument. It has two major advantages: it can be dismantled for cleaning and it has a greater ion capacity. With this improvement, we can hope to inject more samples and thus quantify HCPs that were below the detection limit in our analyses. Furthermore, despite identification results that are superior to those of other instruments without ion mobility, the data processing tools are not mature enough to perform a reliable quantification of all signals. Indeed, they still encounter some difficulties that lead to incorrect signal extractions.

In the years to come, the diaPASEF method performed with the TimsTOF PRO, combined with a perfected software solution, should be suitable for use in a biopharmaceutical environment for HCP monitoring during process development, batch release or validation of ELISA kits.

Part II: The analytical challenge of drug products

Chapter 1

Sample preparation improvements for the analysis of drug substances

1. Challenge of the analysis of final drug substances

In the previous parts, we addressed the critical need to monitor protein impurities from the host cell during the manufacturing process of biotherapeutics. Indeed, the monitoring of HCP is considered as a critical quality attribute (CQA) and the impurity content in the final product must be as low as possible^{35, 298}. Among the analytical arsenal allowing the control of HCPs, we have demonstrated the power of mass spectrometry for the identification and quantification of impurities during the manufacturing process. For instance, mass spectrometry has been used to validate the HCP coverage of ELISA assays²⁸⁴, to evaluate risks due to HCPs^{37, 48} or to support process developments^{4, 285}. However, only a handful of publications have shown the use of mass spectrometry to study the HCP content in final products. Most of them concern the study of the NIST mAb^{17, 29, 50, 299-300} (Merck, Darmstadt, Germany), a reference mAb produced by the National Institute of Standards and Technology (NIST), and very few are related to the study of drug substances administered to humans^{17, 301-303}.

Despite the high potential of MS, the analysis of these high purity samples remains a major analytical challenge. Indeed, protein impurities are progressively removed during the purification steps and identifying HCPs in final drug products is like looking for needles in a haystack^{4, 35, 283}. While the concentration of mAb increases, the amount of HCPs decreases. As a result, the monitoring of these impurities becomes more complicated during the purification process and requires extreme sensitivity and covered dynamic range for their detection at the end of the process (Figure 77).

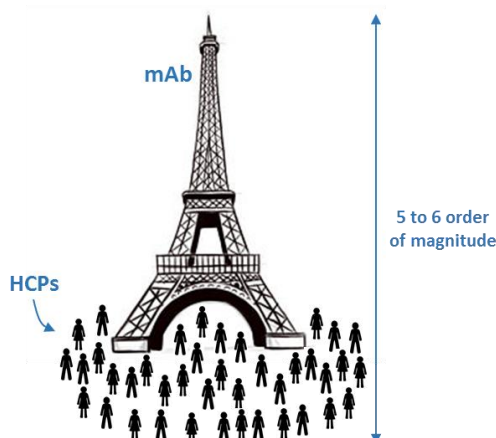


Figure 77: Dynamic range between antibody and HCPs in a drug substance.

In this context, interferences due to the ubiquitous mAb represent an obstacle to the accurate detection and quantification of HCPs by mass spectrometry. To overcome this problem, few studies have attempted to decrease the amount of therapeutic proteins using multidimensional chromatography^{10-11, 27, 48-49} or mAb depletion^{29, 50-51}. Both approaches aim at reducing the dynamic range of mAb versus HCPs in the sample. However, despite the various couplings that can be

considered, the use of a multidimensional chromatography step inevitably reduces the overall robustness of the strategy. For this reason, the native digestion protocol introduced in 2017 by Huang *et al.*, has gained popularity over past few years²⁹. This protocol involves a liquid digestion in native conditions with a small enough amount of trypsin to keep the mAb almost undigested while targeting and digesting the HCPs. A heat stress followed by a centrifugation step is then used to precipitate the antibody, decrease its amount in the sample and concentrate HCPs digested peptides in the supernatant. This strategy has shown improved coverage of HCPs in drug substances (DS) compared to traditional digestion^{50-51, 303}.

2. Context of the project

For samples analytically as challenging as DS, every step becomes crucial, from the sample preparation to the MS acquisition instrumentation and scanning methods down to the data processing workflow. We have been investing significant effort in the laboratory over the past few years in this context and have attempted optimizing several steps of the overall workflow³⁰³.

In this chapter, we will first detail the native digestion protocol that we have adapted and optimized from the Huang *et al.*²⁹ protocol by implementing a double trypsin/Lys-C digestion intended to improve the proteolytic efficiency by reducing missed cleavages at lysine residues³⁰. Then, we optimized a DIA acquisition method and a data processing workflow dedicated to the study of final drug products. Finally, we describe our attempt to further improve the sample preparation of DS by implementing a micro-mixing step via focused ultrasonication. To conduct this latter evaluation, we benefited from the technical support of the Covaris company and especially of Dr. Nicolas Autret. Covaris manufactures devices that deliver focused ultrasonication, or adaptive focused acoustics (AFA), to samples for DNA/RNA shearing, FFPE extraction, chromatin shearing or cell lysis (Figure 78).

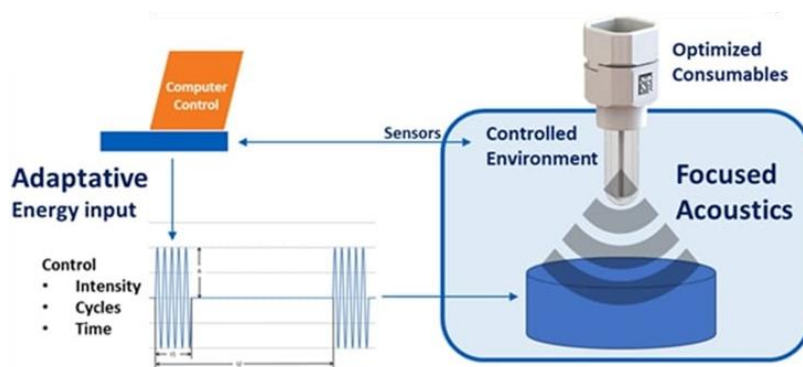


Figure 78: The principle of Adapted Focalised Acoustic (AFA) technology.

The electronics and ultra-high frequency transducer produce a wavelength of only a few millimeters, allowing the ultrasonic acoustic energy to converge into a focal area inside a sample vessel immersed in the water bath. The acoustic energy passing through the water bath induces localized pressure fluctuations that form the cavitation bubbles. Then, the generated cavitation bubbles oscillate or grow to a critical size to finally collapse, creating a hydrodynamic shear stress in the sample. The oscillation and collapse also generate a micro acoustic current, which generates fluid flow in the sample. As a result, the AFA technology enables to focalize the acoustic energy directly to the sample thus improving the efficiency of the energy delivered with less overall heat.

When seeing very preliminary, though promising results, on the analysis of blood samples using an ultrasonication device, we wanted to explore this idea further for the analysis of DS. Indeed, those preliminary results highlighted the potential improved recovery of blood proteins by allowing the identification of proteins interacting with albumin, which are thus generally removed during the albumin depletion step. This triggered our idea to attempt applying this approach to drug substance samples in which the HCPs remaining after extensive USP and DSP purification tend to be the ones that interact with the highly predominant antibody. We thus wanted to evaluate the benefits that such a micro-mixing step could bring to the coverage of identified HCPs. We could get access to a local Covaris S220 ultrasonicator (Figure 79) system located at the Research Center of Immunology and Hematology in Strasbourg (UMR_S INSERM U1109).

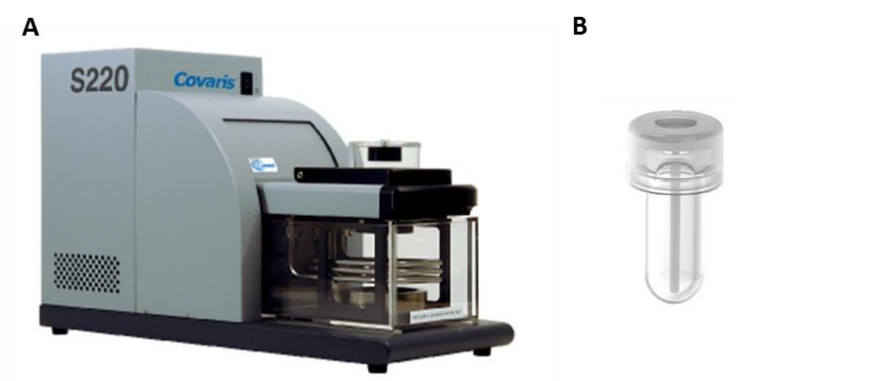


Figure 79: Presentation of the system used for the experiments. Covaris S220 Focused-ultrasonicator (A). MicroTUBE AFA Fiber Pre-Slit Snap-Cap 6x16mm (B).

This system delivers a precise and controlled acoustic energy while controlling the temperature of the sample. It allows to process one sample at a time using a microTUBE AFA Fiber Pre-split-Cap from Covaris.

3. Experimental design

The general analytical scheme is presented in Figure 80.

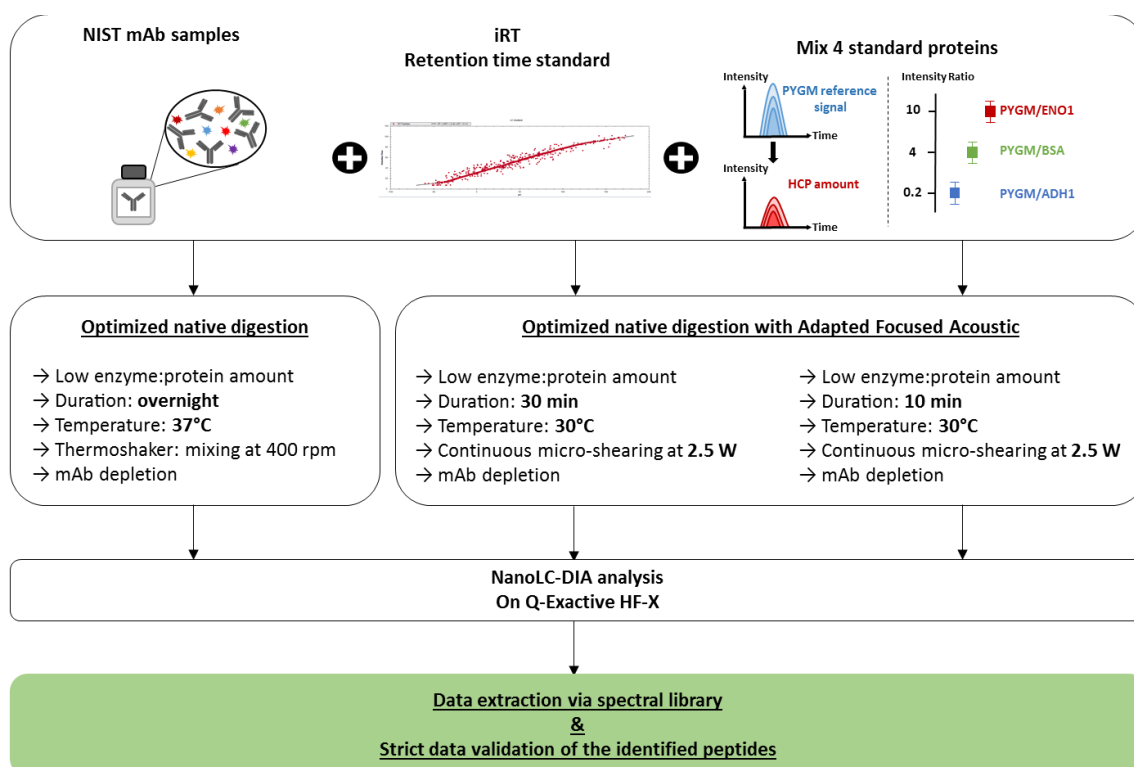


Figure 80: Optimization of a DS analysis workflow. NIST mAb samples were subjected to an optimized native digestion protocol and spiked with iRT retention time standards (Biognosys) and the mixture of 4 standard proteins (ADH, PYGM, BSA, ENL from the MassPREP digestion standard kit, Waters). Technical replicates were injected using a DIA variable isolation windows method dedicated to DS on a nanoLC-Q Exactive HF-X system (Thermo fischer Scientific). Data extraction was performed via a spectral library and filters applied to the list of identified peptides, using an R script, were finely tuned for this type of challenging sample.

Prior to any sample preparation and acquisition method optimizations, we have developed a stringent data validation workflow including a series of filters. This is of utmost importance when conclusions about HCP amounts have to be drawn on final drug products. More than ever, the identification and quantification results need to be reliable and robust.

With the data validation workflow in hand, we have first developed a native liquid digestion protocol inspired and adapted from the protocol of Huang *et al.* The NIST mAb was digested overnight in its buffer at an adjusted pH with a trypsin/Lys-C mixture. The use of a low protease:protein ratio of 1:400 aims at preferentially digesting HCPs of smaller size than the mAb, which under native conditions remains highly structured and is thus less accessible to enzymatic digestion. Then, a heat stress of the sample at 90°C followed by a centrifugation step at 13 000 g allows the precipitation of the antibody. Finally, the supernatant is collected after addition of the retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) for RT normalization during DIA data processing and the mixture of 4 standard proteins (ADH, PYGM, BSA and ENL from Waters MassPREP Digestion Standard Kit) for HCP quantification.

Second, we optimized a DIA method of 60 isolation windows, dedicated to drug substances, based on the NIST mAbs DDA analyses performed on the nanoLC-Q Exactive HF-X (Thermo Fisher Scientific). Technical triplicates of the NIST mAb samples subjected to the native digestion were injected on the same instrumental coupling in DIA mode to assess the robustness of the sample preparation protocol. A spectral library-based approach was applied to extract the MS2 signals from the DIA data. The Spectronaut™ software and a mouse spectral library, generated from 24 DDA runs of a NIST mAb fractionated sample (including iRT retention time standard from Biognosys) combined with 16 DDA analyses of the NIST mAb samples spiked with the mix of four standard proteins, were used for this purpose. This library includes 771 peptides from 143 protein groups. Then, a Q-value filter of 0.01 was applied and iRT profiling was set on. The 2 fmol of PYGM was used as a universal response factor (Top3 peptides signal/mol) to perform the HCP quantity estimation (HCP Top3 peptides signal/signal response factor). Finally, we improved the R script dedicated to the validation and quantification of HCP peptides.

Following the development of the HCP workflow dedicated to drug substances, we implemented a micro-mixing step to try to further enhance the native liquid digestion. Thus, we evaluated the use of an AFA step during a 10 min and 30 min trypsin/Lys-C double digestion with a continuous energy of 2.5W applied. The addition of this ultrasonication step involves transferring the sample into the microTUBE AFA Fiber Pre-split-Cap (AFA tube). The samples are then transferred to an eppendorf for final sample preparation. Finally, technical triplicates were injected on the nanoLC-Q Exactive HF-X and were subjected to the MS2-based DIA quantification workflow discussed above.

Sample preparation, nanoLC-MS/MS methods and data treatment are detailed in the experimental part, section 2.A.

4. Results

A. Optimization of a stringent data validation workflow

We have optimized a stringent data validation workflow including a series of filters. First, it should be noted that the identified HCPs were validated by the target-decoy approach with a FDR of 1% at the peptide level. However, the number of identified proteins being low, the use of this approach at the protein level would lead to an inaccurate estimation of the false positive rate. Thus, we have implemented four stringent levels of filtering that were applied on the raw identification results before proceeding to peptides and proteins quantification (Figure 81).

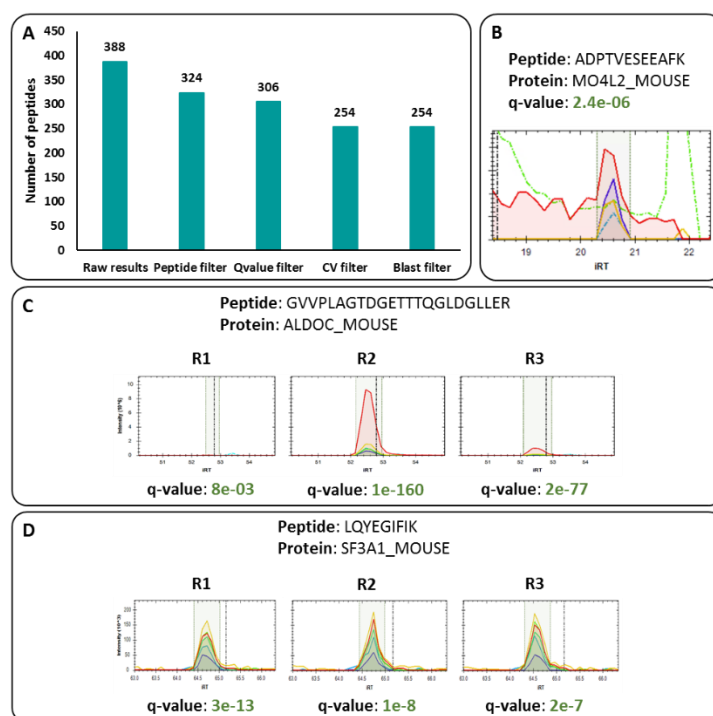


Figure 81: Filters applied for the validation of identified and quantified HCPs in drug substances. Representation of the effect of each filter applied to validate HCPs peptides prior to quantification (A). Example of a peptide with an interference leading to a biased quantification (B). HCP peptide with non-reproducible extracted signals (C). MS2 signals extracted from a peptide that has passed all filters leading to a robust and accurate quantification of the HCP (D).

From signal extraction to HCP quantification, 134 peptides were removed (Figure 81.A). Indeed, the excess of antibody can cause signal suppression or lead to identification and quantification biases (Figure 81.B). These biases can significantly influence the conclusions and lead to time-consuming and costly investigations. Thus, we filter the identified peptides on four levels. The first filter level concerns the type of peptide. Its purpose is to:

- Eliminate peptides carrying modifications commonly observed during a bottom-up proteomic analysis, such as acetylation of proteins n-termini and oxidation of methionine. Their unmodified counterparts are also removed because the modifications are not complete and represent a portion of the peptide signal. Therefore, quantifying the unmodified counterpart will lead to biases in the quantification step as its signals are not representative of the entire peptide signal.

- Keep doubly and triply charged peptides only.
- Keep peptide specific to a protein group.

Modified peptides and four times charged peptides display bad quality signal in terms of intensity and peak shape.

The second level of filtering can be described as a signal quality filter. Only precursors with at least two q-values ≤ 0.01 among the three replicates are kept.

Then, the reproducibility of the precursor ion abundances extracted at the MS2 level is controlled. Retention time profiling between replicates is activated. This option proposed by several software including Spectronaut allows, if a precursor ion is not identified in one of the replicates, to use the information collected in the other samples to extract the signals of the missing precursor. As a result, no missing values are tolerated allowing to retain only precursor ions that can be consistently quantified. In addition, only precursor ions with a CV $\leq 20\%$ across the replicate areas are conserved to ensure reproducibility of the quantification. This avoids the use of peptides, such as the one in Figure 81.C, which would lead to a non-robust quantification.

For the fourth filtering level, the similarity between the quantified peptides and the mAb sequences is studied. After removal of protein groups including either mAb chains or common contaminants included in the database, an alignment of the identified HCP sequences with those of the mAb chains is performed using the BLASTp algorithm (v.2.10.0+). Thus, HCP sequences of at least six amino acids with more than 80% overlap with the targeted mAb sequence, and without sequence variation, were excluded from the identification results. The purpose of this filter is to avoid false identifications of HCPs resulting from uncontrolled mAb degradation³⁰⁴ or aspecific cleavages caused by protease impurities³⁰⁵⁻³⁰⁷.

The application of all these filters allows the use of reliable identifications with reproducible abundances in order to perform the HCPs quantification by the Top3 strategy¹⁶ (Figure 81.D). However, a manual verification of the most abundant HCPs is still recommended to avoid any surprises. This was the case for a drug substance produced from CHO cells. The most abundant impurity was quantified at 479 ng/mg with the second most abundant HCP at 18 ng/mg. Investigations showed that the peptide used for quantification had the same mass as a mAb peptide, i.e. 402.7449 m/z for the doubly charged peptide. A similarity in terms of mass up to the fourth decimals was observed. This example illustrates how important the manual checking of the results is, especially when dealing with HCPs characterization in final drug substances.

B. Development of an optimised native liquid digestion protocol

Based on the native digestion protocol proposed by Huang et al., we implemented the use of a double trypsin/Lys-C digestion instead of using trypsin alone. Then, we applied our developed MS2-based DIA quantification workflow to the NIST mAb samples. The results are presented in **Figure 82**.

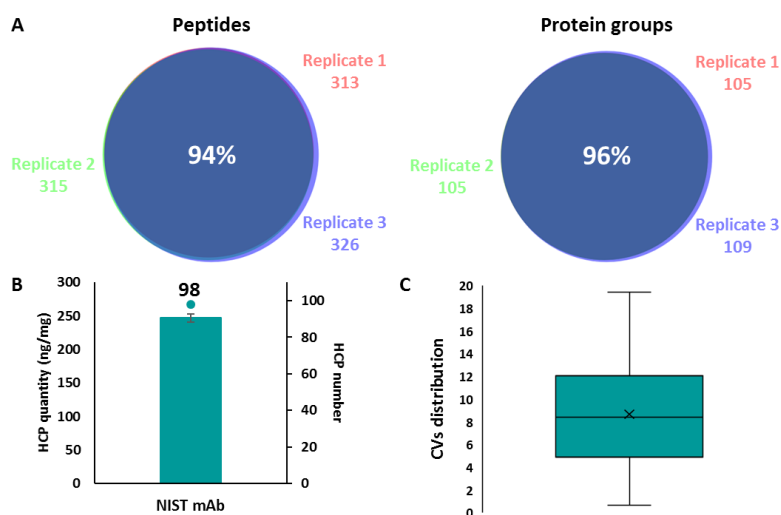


Figure 82: Performance of the HCP workflow dedicated to drug substances. Overlap of the peptides and protein groups identified in three technical replicates of NIST mAb samples subjected to our optimized native digestion protocol and analyzed using a nanoLC-DIA Q Exactive HF-X (Thermo Fischer Scientific) system (A). NIST mAb overall HCPs amount and number of HCPs quantified using the developed MS2-based DIA approach (B). Distribution of the CVs calculated on HCPs peptides intensities within the three technical replicates (C).

Venn diagrams presented in Figure 82.A display the 313, 315 and 326 peptides and 105, 105 and 109 protein groups identified in the three technical replicates, both showing 94 and 96% similarity, respectively. These results demonstrate a high reproducibility of our optimized digestion protocol, which uses about 8 times less amount of enzyme than a classical digestion protocol. After application of our stringent data validation workflow presented in section 4.A., we were able to quantify 98 HCPs for an overall amount of 247 ng/mg mAb (Figure 82.B). A median of CVs calculated on HCPs peptides intensities of 8.4% was obtained, demonstrating the accuracy of DIA extraction on the three technical replicates of these challenging samples (Figure 82.C). Furthermore, when comparing our results with those of Huang et al., we noted that 82% of the HCPs identified on their Q-TOF instrument could be covered by our DIA method, performed on the Q Exactive HF-X, with 49 additional HCPs (Figure 83). Interesting results, given that various mass spectrometers can be found at different drug production sites.

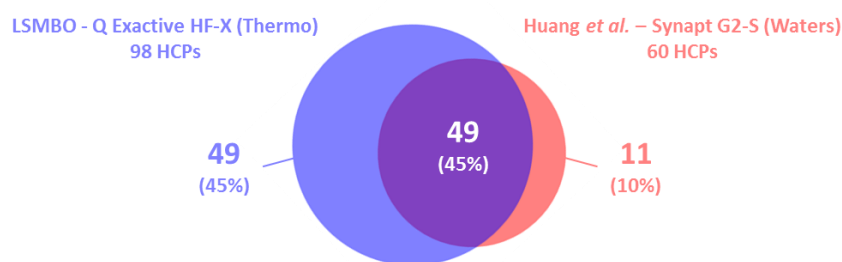


Figure 83: HCPs coverage in the NIST mAb using the optimized native liquid digestion protocol. Overlap of the HCPs quantified in the NIST mAb using the optimized native digestion combined with MS2-based DIA quantification with the list obtained by Huang et al.

Finally, this protocol showed a reproducibility and accuracy of quantification similar to the workflow developed in Part I using CCCF and PPA fractions while decreasing the dynamic range mAb/HCP which represent the bottleneck of drug substance analysis.

C. Implementation of a focused ultrasonication step

Before using the Covaris ultrasonicator as a micro-mixing step for the duration of the digestion, we had considered using an AFA step before or after addition of the trypsin/Lys-C mix to detach HCP peptides that would be in interaction with the antibody before the depletion of the latter. First, we aimed at investigating the effect of power of the AFA step. We applied AFA energy to the samples ranging from 5 to 30 W for a duration of 4 minutes. However, the results obtained showed such a low reproducibility that for most samples, no HCPs could be quantified (filtered out at the CV level). In addition, the identification results did not display an increased HCPs coverage but rather a decrease in the number of HCPs identified compared to the control sample. We consider robustness as a crucial criterion for our method and as no improvement was observed in terms of HCPs coverage, we did not continue the experiments with this starting hypothesis.

In addition, the chromatograms of these analyses showed maximum TIC intensities 3 to 7 times higher than the maximum expected intensity around $1E+10$. The most intense peptides being those of the antibody, it is very likely that the addition of the described ultrasonication step tends to destabilize the antibody and makes it more prone to digestion. Thus, the amount of injected material is underestimated, which leads to TIC chromatograms with low repeatability and high intensity.

Following these experiments, we considered using the micro-mixing step during the digestion time. Instead of using AFA energy to potentially detach HCPs interacting with the mAb, we wanted to use it to enhance the digestion of HCPs. Thus, the trypsin/Lys-C double digestion was performed in the bath of the Covaris system at 30°C with a continuous AFA energy applied of 2.5 W. Two incubation times were tested: 10 and 30 min. The samples were injected in triplicates on the Q Exactive HF-X and the quantification results were compared to those of the control sample (Figure 84).

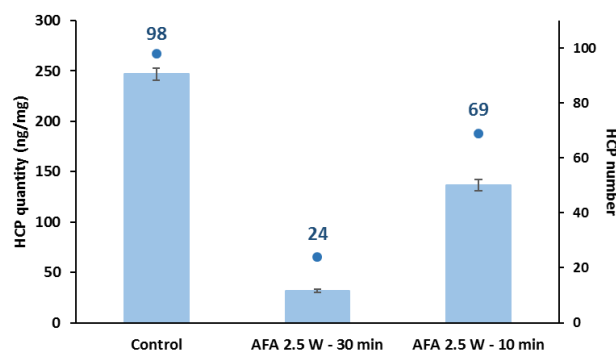


Figure 84: Covaris AFA to enhance the native digestion. Overall HCPs amounts and numbers of HCPs quantified in the control sample digested overnight and two NIST mAb samples subjected to a 10 and 30 min digestion on the Covaris system (A). Bar heights represent the means of the global HCPs amount in technical replicates. Error bars represent the standard deviation of the overall HCPs amounts in technical replicates.

Interestingly, 69 HCPs were quantified in the sample digested during 10 minutes with continuous ultrasonication at 2.5 W. Compared to the 98 HCPs quantified in the control sample that was subjected to an overnight digestion, we note promising results that would decrease the sample preparation time below the 20-23 hours currently required. However, we quantified 29 fewer HCPs with this ultrafast digestion. A non-negligible number in our context. The NIST mAb sample digested for 30 min displayed 24 quantified HCPs. It therefore showed a lower HCPs coverage than the control sample and the one digested in 10 min. To understand these results, we investigated the identification results and the CVs of the HCPs peptides used for quantification (Figure 85).

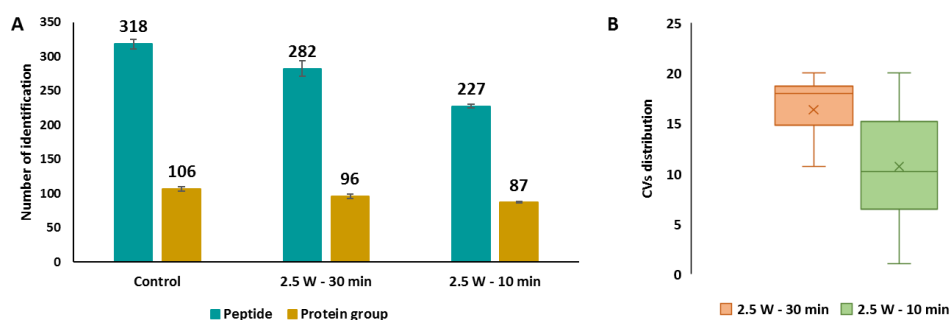


Figure 85: Quantification results investigation. Numbers of identified peptides and protein groups in the control sample digested overnight and two NIST mAb samples subjected to a 10 and 30 min digestion on the Covaris system (A). Bar heights represent the means of the numbers of identifications in technical replicates. Error bars represent the standard of the numbers of identifications in technical replicates (A). Distribution of the CVs calculated on HCPs peptides intensities within the three technical replicates (B).

Looking at the numbers of peptides and protein groups identified, we observed a decrease in the numbers of identifications in line with the reduction in digestion time (Figure 85.A). These results highlighted that the application of our strict filters led to the loss of a higher number of HCPs for the samples digested in 30 min compared to those digested in 10 min. This assumption was confirmed by Figure 85.B. Indeed, we observe a median of CVs calculated on HCPs peptides intensities of 17.9% for the 40 HCPs peptides quantified in the 30 min digestion sample compared to 10.2% for the 114 peptides quantified in the 10 min digestion sample. This low repeatability observed for the samples digested in 30 min is explained by the use of the Covaris S220 system. It works for one sample at a time and requires the use of specific tubes. It is necessary to transfer the sample into the Covaris tube. Then for the final centrifugation step, the sample must be transferred into an Eppendorf. These transfers, especially the Covaris-Eppendorf tube transfer, lead to material losses because the Covaris tube contains a fiber that hinders the recovery of the sample. This problem was particularly observed for the first two replicates of the 30 min digestion samples. In conclusion, these results, while promising, would require a new dedicated experiment to confirm the interest of the ultrasonication step to improve the native digestion protocol.

6. Conclusion & Perspectives

In this chapter, we have seen that drug substances represent the most complex sample type from an analytical point of view. Indeed, the extreme dynamic range of mAb/HCP and the overabundance of antibody make it difficult to analyze these samples by MS. This is why we have developed a workflow dedicated to final drug products. The native digestion protocol was found to be the most adequate to increase the coverage of HCPs while keeping a fast and repeatable sample preparation.

Several approaches can be considered to further optimize the protocol proposed by Huang et al. We have investigated the use of an ultrasonication step to enhance HCP coverage. Our first hypothesis was not confirmed as the use of AFA before or after addition of the trypsin/Lys-C mix tends to destabilize the mAb and makes the depletion of the latter less efficient. Our second approach gave interesting results. The use of a Covaris system during the digestion would help to enhance the proteolysis step. Therefore, lower digestion duration can be considered, which will help to develop a more straightforward sample preparation even more suitable to routine analysis. A new experiment needs to be done. This experiment would investigate the use of a Covaris system at low energy (2.5W) during the digestion. Thus, an overnight digestion as well as proteolysis duration of 30 min, 1 hours and 5 hours can be considered and the results can be compared to an overnight native digestion without ultrasonication.

Finally, regardless of the method, strict data validation is always required to ensure confident quantification of HCP impurities.

Chapter 2

Case study: Manufacturing process investigation

1. Context of the project

During the development of an antibody, UCB Pharma had to redesign the purification process in order to decrease the amount of HCPs in the drug substance. While the amounts measured via a commercial ELISA kit showed a decrease in the impurity content, the use of an in-house ELISA kit did not support this conclusion. These conflicting results led to an additional investigation to determine if the problem was with the in-house ELISA kit, which is supposed to be more specific to the samples, or if the new bioprocess did not meet expectations. Thus, the protein characterization team of UCB Pharma based in Slough performed LC-MS/MS analyses using DDA analysis on a Q-Exactive instrument. In addition, we got involved in this investigation and five samples from both manufacturing processes were sent to Strasbourg, allowing us to apply our DIA strategy on a concrete case.

The two samples from process 1 and the three samples from process 2 were analyzed on the Q Exactive HF-X with the Top3-DIA approach. The four standard proteins mix of the MassPREP digestion kit from Waters was used. In the interest of confidentiality, only the results obtained in Strasbourg will be discussed. Moreover, the purification process and the samples will be anonymized.

2. Experimental design

The experimental design is detailed in Figure 86.

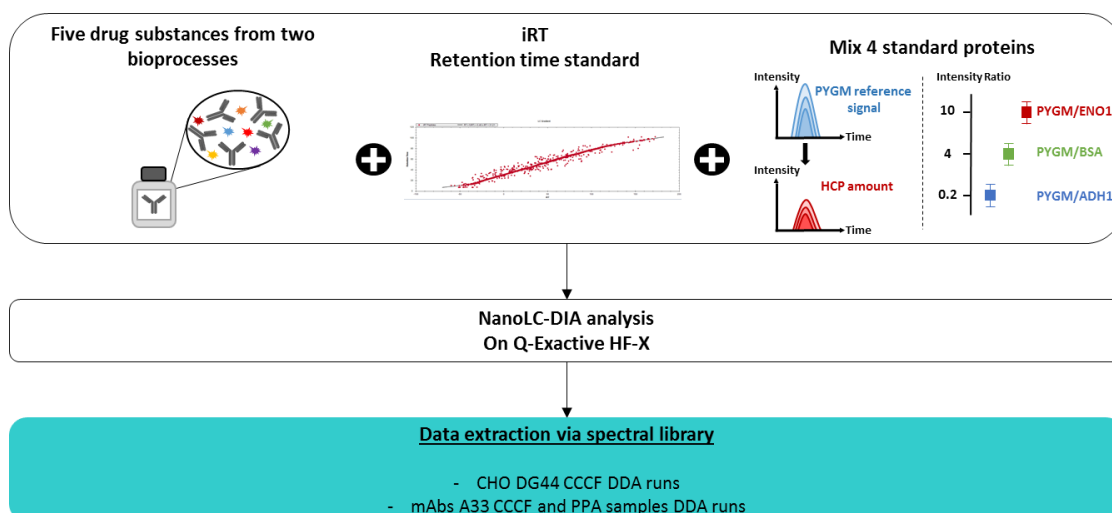


Figure 86: Analytical workflow used to assess the impact of two bioprocesses on the HCP content.

Five drug substance samples were subjected to the native digestion protocol and spiked with iRT retention time standards (Biognosys) and the mix of 4 standard proteins (ADH, PYGM, BSA, ENL from MassPREP digestion standard kit, Waters). Samples were injected in injection triplicates using DIA acquisition on a nanoLC-Q Exactive HF-X (Thermo fischer Scientific) system. MS2-based quantification results were used to evaluate the effects of process 2 on the HCPs content.

The five samples were digested overnight in their buffer at an adjusted pH with a trypsin/Lys-C mix at a protease:protein ratio of 1:400. Then, a heat stress of the sample at 90°C followed by a centrifugation step at 13 000 g was performed in order to precipitate the mAb. Finally, the supernatant was collected and ready to be injected after addition of the retention time standard (iRT kit, Biognosys, Schlieren, Switzerland) for RT normalization during DIA data processing and the mix of 4 standard proteins (ADH, PYGM, BSA and ENL from Waters MassPREP Digestion Standard Kit) for HCP quantification. Injection triplicates were performed on the nanoLC-Q Exactive HF-X (Thermo Fisher Scientific) in DIA mode.

A spectral library based approach was applied to extract the MS2 signals from the DIA data. The Spectronaut™ software and a CHO-specific spectral library, generated from 24 DDA runs of a CCCF CHO-DG44 null cell line fractionated sample (including iRT retention time standards from Biognosys) combined with all DDA analyses of the CCCF and PPA mAb A33 samples spiked with the mix of four standard proteins, were used for this purpose. This library includes 44 726 peptides from 4 308 protein groups. Then, a Q-value filter of 0.01 was applied and iRT profiling was set on. Finally, the 2 fmol of PYGM was used as a universal response factor (Top3 peptides signal/mol) to perform the HCP quantity estimation (HCP Top3 peptides signal/signal response factor).

Sample preparation, nanoLC-MS/MS methods and data treatment are detailed in the experimental part, section 2.B.

3. Results

A. Quantification results

The overall amounts and numbers of HCPs quantified thanks to our DIA strategy are presented in Figure 87.

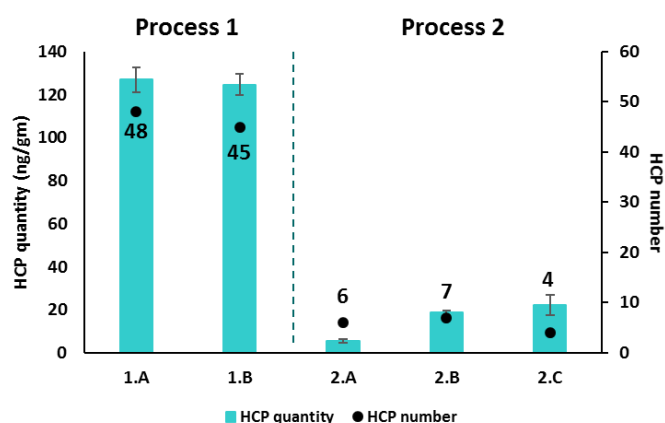


Figure 87: Evaluation of HCPs contents in drug substances produced from two bioprocesses by our DIA approach. Global HCP amounts and numbers of HCPs quantified obtained for the process 1 samples (samples 1.A and 1.B) and the process 2 samples (samples 2.A, 2.B and 2.C) using MS2-based DIA quantification. Bar heights represent the means of the global HCPs amounts in injection triplicates. Error bars represent the standard deviation on injection triplicates.

The MS2-based DIA approach was able to quantify an average of 47 HCPs within the Process 1 samples with overall amounts between 125 and 127 ng/mg mAb. For process 2, an average of 6 HCPs were quantified representing 6 to 22 ng/mg mAb. Thus, the results obtained leave no doubt about the conclusion. On average, process 2 allowed to decrease the amount of HCPs by a factor 8 with more than 40 fewer HCPs identified. The objective of reducing the protein impurity content was thus reached thanks to the development of this new purification protocol.

B. Process related impurities study

When developing a biopharmaceutical, in addition to the quantity, it is necessary to obtain the identity of process-related HCPs to perform risk assessment. Indeed, the presence of residual proteases has been reported to cause product degradation^{305-306, 308-310}. Some contaminants, such as proteins with reactive cysteine residues, have been linked to protein aggregation³¹¹⁻³¹². Alternatively, other enzymes have been found to degrade stabilizing surfactants³¹³⁻³¹⁴. Various issues that can lead to reduce shelf-life of mAb, diminish its efficiency or promote immune reactions.

For our case study, UCB Pharma had already performed a risk assessment study on the internally identified HCPs from process 1. As the manufacturing process itself may influence the HCP profile, we have compared in detail the HCPs identified and quantified in both processes (Figure 88).

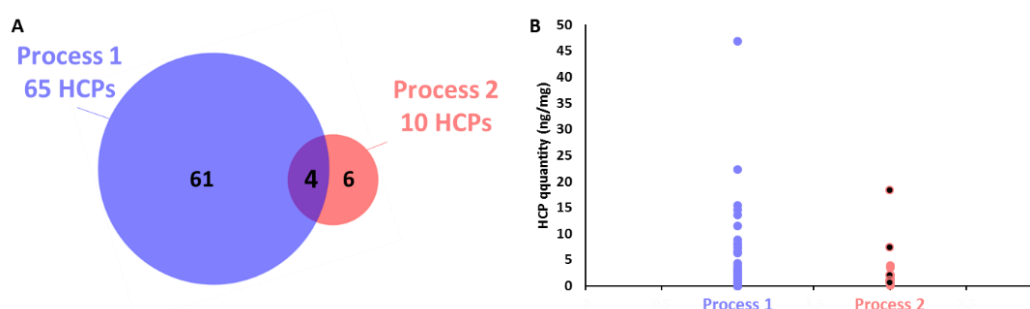


Figure 88: Process-related HCPs comparison. Overlap of the quantified HCPs in each process (A). Representation of the HCPs quantities determined in the two process 1 samples and in the three process 2 samples (B). Black dots represent the HCPs uniquely quantified in process 2.

The quantified HCPs in samples 1.A and 1.B were combined into a single list for process 1 and the same was done for process 2 with samples 2.A, 2.B and 2.C. The Venn diagram revealed 6 HCPs uniquely quantified in process 2 samples (Figure 88.A). These proteins represent between 0.7 and 18 ng/mg mAb (Figure 88.B). Noteworthy is that three of these HCPs proteins were also identified in process 1 samples but they were not quantified because they did not pass the CV < 20% filter or the Q-value filter, i.e., at least two Q values < 0.01 out of the three replicates. Finally, based on these results, UCB Pharma must now perform a new risk assessment study focusing on the 6 proteins uniquely identified in Process 2 which will then lead to the decision of whether the purification process should still be improved or not.

4. Conclusions

The Top3-DIA method could be applied to a concrete case. It allowed the quantification of 10s of HCPs in five drug substances from two bioprocesses. The method was shown to be robust, with an average of the coefficients of variation calculated on peptide intensities lower than 13%, and sensitive with a quantification down to the sub-ng/mg mAb level.

As a conclusion, we were able to assist UCB Pharma in its investigation. The results obtained revealed a decrease of the HCPs content with the new purification process, in line with the conclusions obtained internally by the UCB Pharma protein characterization team. Moreover, a new risk assessment study can be performed on the HCPs uniquely identified in Process 2. The results obtained by ELISA can be explained by the presence of some HCPs that manifested high immunoreactivity to anti-HCP reagents, which could explain the high response observed in the second process samples. Indeed, the development of a specific ELISA kit is performed with a mock cell line sample. Thus, if the antibody production conditions change, compared to this standard, the HCP population may vary and bias the quantification. It is in this context that we demonstrated the strength of mass spectrometry as an orthogonal method.

Finally, the spectral library-based DIA approach, combined with a Top3 quantification strategy and a suitable sample preparation, provides a comprehensive picture of HCPs remaining in drug substances. This strategy represents a promising alternative to ELISA as a quality control method for daily use. The results discussed in this chapter were incorporated into the commercialization request of an antibody to regulatory agencies.

Part III: Development of a HCP profiling method on a MS platform for routine analysis

Chapter 1

Evaluation of the BioAccord LC-MS system

1. Context of the project

The project was conducted in the frame of a collaboration between UCB Pharma and Waters, both interested in evaluating the BioAccord LC-MS system for the characterization of HCPs. This instrument, manufactured by Waters, aims at providing an easy-to-use solution for biopharmaceutical companies (Figure 89). The BioAccord is a self-calibration tool that will assist the analyst if a problem occurs during analysis. For a simple problem, manageable by inexperienced users, the system will indicate the source of the problem using icons displayed on the front panel of the MS, e.g. for problems with electronics, solvent bottles, etc. If the problem is too serious to be handled by an inexperienced analyst, the system will advise the action of the Waters service team. Thus, the operation of this instrument should be mastered with reduced training time.

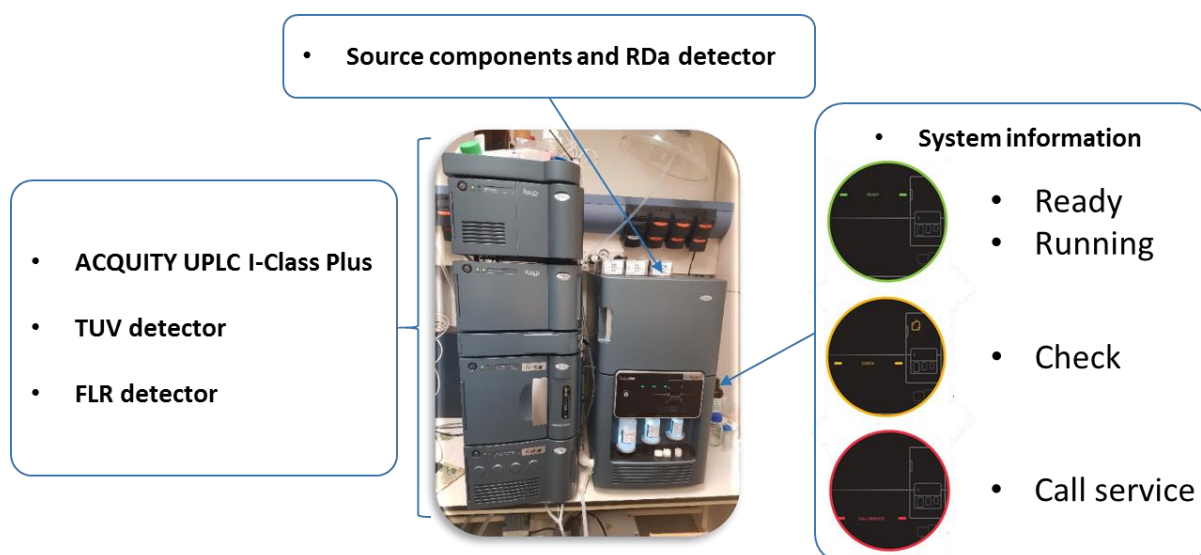


Figure 89: Presentation of the BioAccord LC-MS system.

The system is fully controlled by UNIFI software, which enables efficient use of BioAccord with the creation of analytical workflows that allow for parallel data acquisition and processing. In addition, the all-in-one software includes a complete audit trail system for data acquisition, processing and reporting, which limits the risk of errors and facilitates audit preparation.

Concerning the system itself, it is composed of an ACQUITY UPLC I-Class Plus system, which includes a UV and fluorescent detector, coupled to a mass spectrometer that operates with an ESI ionization source and an ACQUITY RDa detector. Two acquisition modes are offered: Full scan and full scan with fragmentation. The second option proceeds to the in-source fragmentation of the ionized peptides with the application of a cone voltage energy ramp. The system will alternate a scan with low energy voltage (peptides) and with high energy voltage ramping (fragments).

We obviously turned to this second acquisition mode to characterize HCPs and have worked on the reference NIST mAb sample for the evaluation of the instrument. The BioAccord system is only composed of an ionization source, lenses to guide the ions and a TOF analyzer and it thus does not allow to select the ions during the analysis. Thus, MS and MS/MS scans will contain all precursors and fragment ions contained in the 50 to 2000 m/z acquisition range. Furthermore, the UNIFI software is not a proteomics software *per se* enabling the analysis of hundreds of thousands of proteins. For example, it is only possible to add one protein sequence at a time in the software. This is a limitation for CHO cell based products as the CHO organism currently contains 56,569 entries.

To conduct this instrumental platform's evaluation, I have spent three months at UCB Pharma in Braine l'Alleud within the Physico-Chemistry Method Development (PCMD) team. We performed the evaluation of the BioAccord LC-MS system for the characterization of dozens of HCPs using a targeted approach via manual data processing. For this purpose, we used the NIST mAb standard prepared according to the native liquid digestion protocol developed in Part II in which we spiked standard proteins. Once the data were acquired on the BioAccord, a UNIFI method, called "accurate mass screening on MSe data", was used to identify the precursor ions and perform the ion current extraction (XIC) of the peptides selected via a discovery approach earlier conducted on the Q-Orbitrap HF-X.

2. Experimental design

The general workflow is presented in Figure 90.

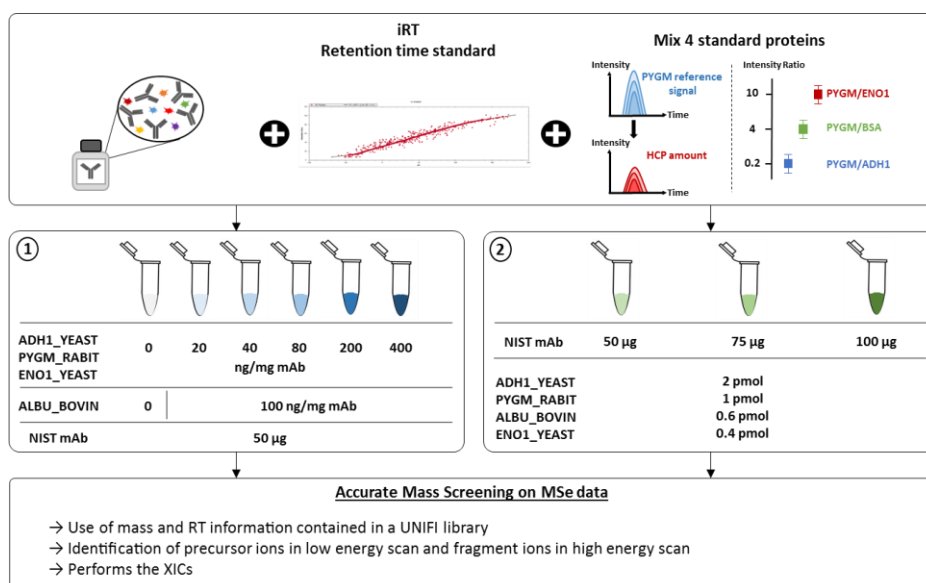


Figure 90: Analytical workflow used to assess the performances of the BioAccord LC-MS instrument for the characterization of HCP impurities. Two experiments were performed. Experiment 1 consisted in spiking standard proteins (ADH1_YEAST, PYGM_RABIT, ENO1_YEAST) at different concentrations from 0 to 400 ng/mg mAb to observe the signal extraction of species down to 20 ng/mg mAb. The ALBU_BOVIN (BSA) protein was added at a constant concentration of 100 ng/mg mAb in order to have a reference that should show the same behavior within the different samples. All these standard proteins were spiked in a constant background of NIST mAb at 50 µg. The aim of experiment 2 was to quantify, using the Top3 strategy, HCPs remaining in NIST mAb samples injected at different amounts: 50, 75 and 100 µg of NIST mAb were injected.

Two experiments were performed to evaluate the BioAccord solution for the characterization of HCPs in the NIST mAb using the workflow optimized for Drug Substances analysis in Part II.

The first experiment was designed to simulate the quantification of HCPs using the four standard proteins of the MassPREP digestion kit. Therefore, in five digested NIST mAb samples, ADH, PYGM and ENL proteins were spiked at increasing amounts of 20, 40, 80, 200 and 400 ng/mg mAb. In parallel, BSA was spiked at a constant concentration of 100 ng/mg mAb as a control protein. A control sample without spiked standard proteins was also prepared. Finally, for the six samples, five injection replicates were performed using the BioAccord system with a constant amount of 50 µg mAb loaded on the column. Then, a peptide library was created on UNIFI including 27 peptides of the standard proteins and 4 peptides of the NIST mAb. For those 31 peptides, the masses of 3 fragments per peptide were added to the library. All these peptides were selected based on their detection in preliminary analyses performed on the BioAccord. Next, targeted data processing was performed using the *Accurate Mass Screening on MSe data* method with precursor and fragment ion mass tolerances of 15 and 20 ppm respectively, and a retention time tolerance of 0.3 min. The identified and extracted signals were manually validated on the basis of their isotopic profile, chromatographic peak shape and fragments identifications. Finally, peptides abundances were used to assess the performance of the BioAccord to extract peptides signals of proteins spiked down to 20 ng/mg mAb.

The second experiment was designed to quantify 30 targeted HCPs of the NIST mAb, selected from global analyses previously conducted on our Q Exactive instrument. Five injection replicates were performed using the BioAccord system. Fifty, 75, and 100 µg of NIST mAb were loaded on column with the same amounts of standard proteins, namely 2 pmol of ADH, 1 pmol of PYGM, 600 fmol of BSA, and 400 fmol of ENL spiked in all 3 samples. Then, a peptide library was created on UNIFI including 115 peptides uniquely belonging to the 30 targeted HCPs and 16 peptides from the four standard proteins. Creating such a library requires the manual entry of every single targeted peptide's sequence. As the number of peptides was quite high, we did not add information about the fragments at this stage. This step would have been very tedious and time consuming as it would have required manually entering all fragments' masses with a precision to the fourth decimal. As previously, the created library was used to target the HCPs peptides using the *Accurate Mass Screening on MSe data* method with a precursor ions mass tolerance of 15 ppm. For this experiment, no restriction about the predicted retention time were set as the peptides to target were selected on a MS platform used at a different flow rate, which leads to a shift in retention times. The identified and extracted signals were manually validated on the basis of their isotopic profile and chromatographic peak shape. Then, we kept only the peptides that did not present missing values within the five replicates per injected amount and that showed CV<20% calculated on peptide intensities among the five replicates. Finally, the 2 pmol of ADH was used to estimate the amount of the targeted HCPs via a Top3 quantification strategy.

Sample preparation, nanoLC-MS/MS methods and data treatment are detailed in the experimental part, section 3.

3. Results

A. HCP monitoring simulation

The BioAccord system is not adapted to a discovery approach aiming at the identification of all detectable HCPs in the sample. We therefore opted for a targeted approach, called *Accurate Mass Screening on MSe data*, available in the UNIFI software (Figure 91).

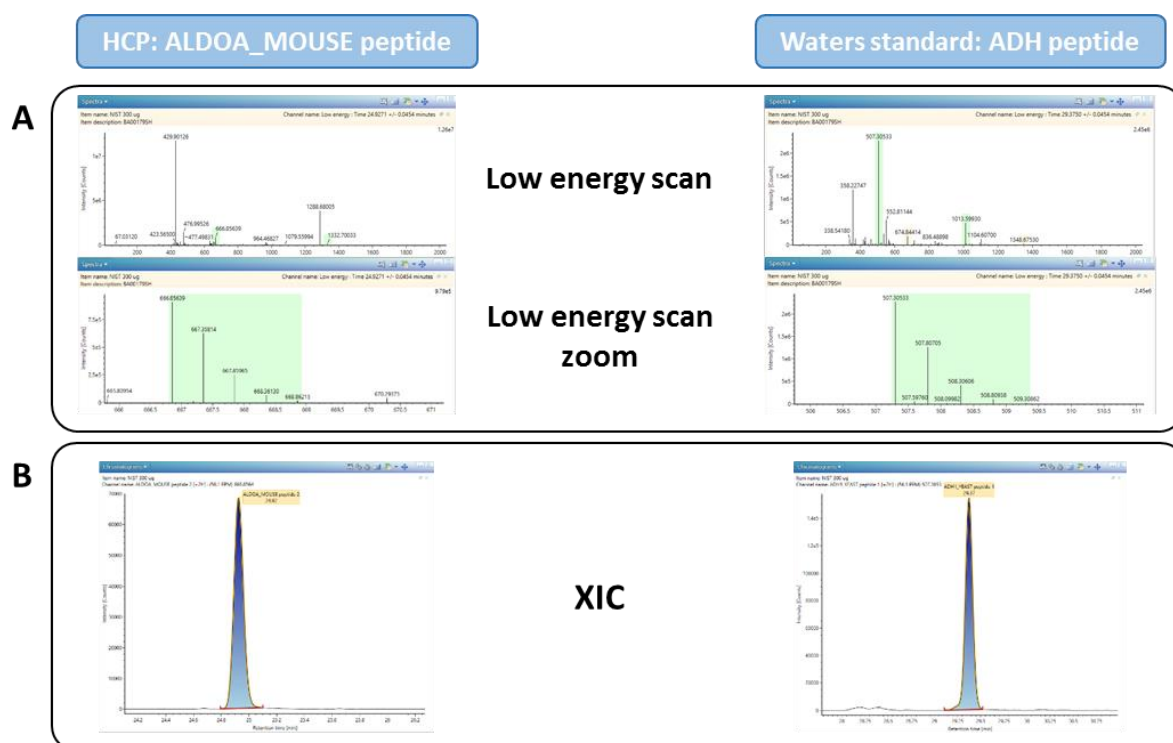


Figure 91: Targeted data processing to investigate the amounts of specific HCPs. Low scan energy observed for a HCP peptide (ALDOA_MOUSE) and a standard peptide (ADH) (A). Extracted Ion Current (XIC) for both peptides (B).

The first step in this approach is to select representative peptides and fragments of the protein of interest. This step is detailed in chapter 3 section 2.B.ii of the bibliographic introduction. As a reminder, the selection is based on unicity, peptide sequence and MS response criteria. Once the peptides are selected, a library can be created in UNIFI. It will contain the peptide sequences, their masses and the expected retention times. Finally, with the library information, precursor and fragment ion mass tolerances (15 and 20 ppm respectively) and retention time tolerance (0.3 min), the UNIFI software performs a search of the selected peptides and fragments in the low energy and high energy scans (Figure 91.A). In addition, it allows the XIC extraction at the MS1 level, which will enable their quantification (Figure 91.B).

We mentioned the fact that UNIFI is not a proteome characterization software. However, it is possible to perform an analysis called *Pep Map MS*, which allows identifying the peptides of a protein in the manner of DDA search engines. Thus, a peptide identified in a low energy scan will be validated by the presence of its fragments in the high energy scan. It is worth noting that we were impressed by the fragmentation of the NIST mAb peptides using the in-source fragmentation of the BioAccord (Figure 92).

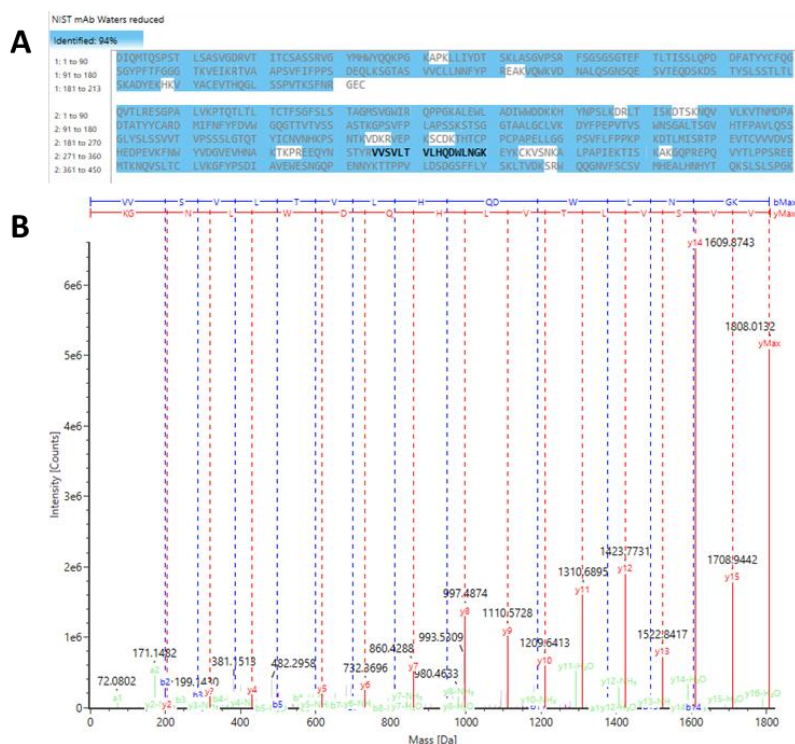


Figure 92: NIST mAb peptide mapping experiment. NIST mAb sequence coverage obtained after an overnight digestion using trypsin and injection on the BioAccord instrument (A). In-source fragmentation spectrum of the black bold peptide observed on the sequence coverage displayed in A: VVSVLTLVHQDWLNGK (B). Red peaks represent the y-ion series, blue peaks display b-ion series and green peaks show fragment ions with a neutral of water-loss.

Repeated injections of 1 µg of NIST mAb digested using trypsin allowed us to evaluate the reproducibility of the BioAccord system. Indeed, more than 90% of the NIST mAb peptides could be covered in each *Pep Map MS* analysis (Figure 92.A). Based on five injection replicates, we observed a relative standard deviation percentage <5% on peptide intensities and sequence coverage and <0.05 min on peptide elution. These results highlight the high reproducibility offered by the LC-MS instrument. In addition, we also performed *Pep Map MS* analyses of NIST mAb samples spiked with the four standard proteins (Mix 4P) to get hands-on experience with BioAccord. Thus, it was on these analyses that we selected the peptides to follow the four standard proteins and NIST mAb for our HCP monitoring simulation experiment. Since we did not have isotopically labeled peptides for an isotopic dilution quantification approach, we also selected the most intense fragments of each peptide as validation criteria and to study their behavior when the amount of injected standards was decreased.

First, we selected a total of 27 peptides and 3 fragments per peptides for the four standard proteins from the analyses performed on the BioAccord instrument. Directly on the result file of the Pep Map MS analysis in UNIFI, it is possible to select a peptide and send it to the desired peptide library. Thus, the sequence, mass and RT information will be automatically updated but the fragment information must be added manually with a precision to the fourth decimal. This step is tedious and time consuming when you have to manually add the masses of 108 fragments. Finally, we performed the *Accurate Mass Screening on MSe data* analysis using the created library. The ion abundances were extracted and we investigated the MS response of species spiked at different amounts with BioAccord (Figure 93).

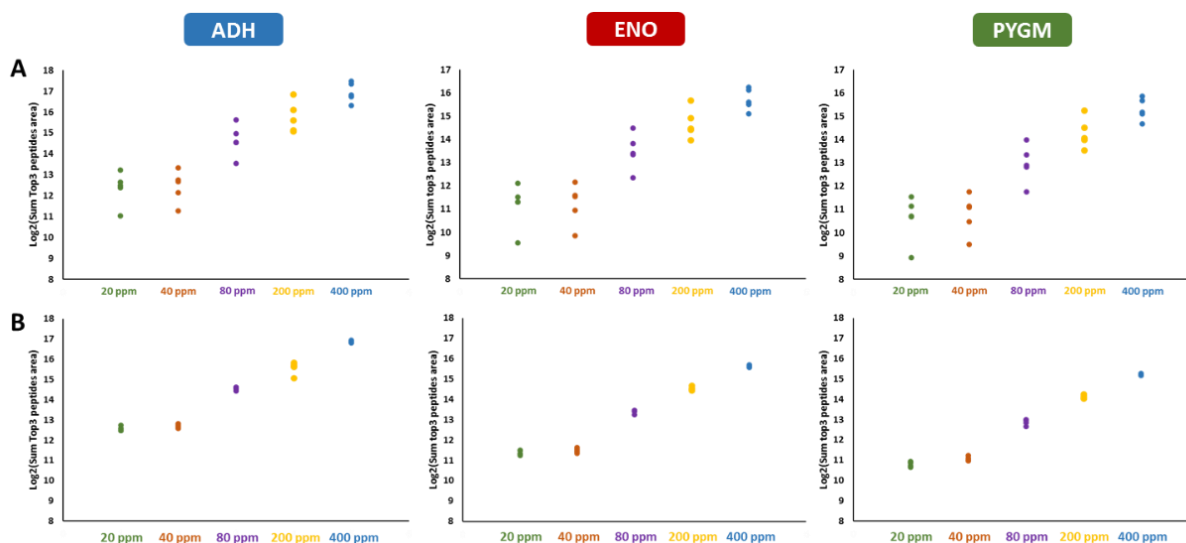


Figure 93: BioAccord MS response study. Log₂ representation of the sum of the areas of the 3 most intense peptides of ADH, ENL and PYGM calculated in each replicate of the 20, 40, 80, 200 and 400 ng/mg mAb samples (A). Log₂ representation using the peptide abundances normalized using BSA signals (B).

For each replicate, the sum of the areas of the three most intense peptides was plotted on a log₂ scale (Figure 93.A). An increasing peptide MS response was observed with an increasing amount of standard injected. However, the results were dispersed due to some instrument instabilities resulting in less reproducible TIC chromatogram intensities. Therefore, we used the median of the BSA Top3 areas as a factor to normalize the signals of the peptides of interest (Figure 93.B). This resulted in much more homogenous and linear results. Noteworthy is that the 40 ng/mg mAb spike point showed the same behavior as the 20 ng/mg mAb spike point. This can be explained by an error during sample preparation that led to the presence of the 20 ng/mg sample twice. We therefore removed the 40 ng/mg sample from the rest of the study. To further document our investigation, we have represented the measured fold changes between the different spike points (Figure 94). The MS responses of the 20, 80 and 200 ng/mg mAb samples were compared to the 400 ng/mg sample and plotted on a log₂ scale of the experimental fold change in function of the theoretical/calculated fold change. For example, we divided the Top3 ADH peptide area obtained in the 400 ng/mg sample by the Top3 ADH peptide area in the same replicate of the 200 ng/mg sample, a fold change of 2 and a log₂(fold change) of 1 are expected.

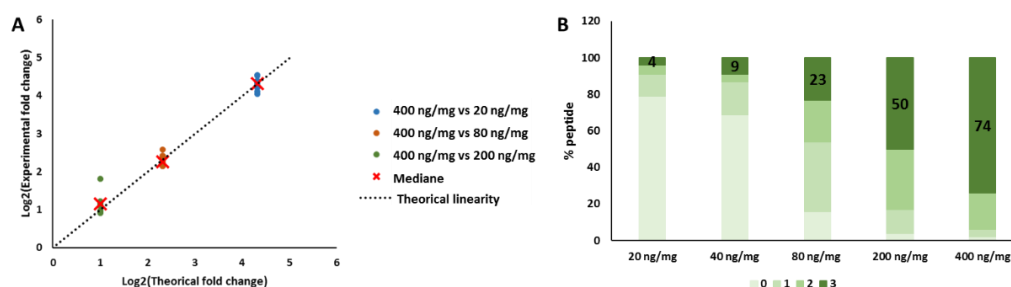


Figure 94: Measured fold changes with the BioAccord RDa detector. Representation of the measured fold changes between ADH, ENL and PYGM Top3 peptide areas of the 20, 80 and 200 ng/mg samples compared to the 400 ng/mg sample used as reference (A). Representation of the percentage of peptides identified with 0, 1, 2 or 3 fragments on the 3 expected (B). Black numbers display the percentage of peptides with 3/3 fragments identified.

Figure 94.A confirmed the linearity of the MS response observed in Figure 93. It can be noticed that the median of the experimental fold change is close to the expected linear curve and this, down to 20 ng/mg mAb. This experiment demonstrates that the BioAccord and the developed method allows the quantification of HCPs down to 20 ng/mg with good accuracy. It is very likely that the quantification limit is lower, but an additional experiment with spiked standard proteins down to at least 1 ng/mg mAb would be requested to support this statement. Furthermore, the fragments study showed that at 20 ng/mg, no fragment was identified for the majority of the peptides (Figure 94.B). The limited numbers of fragments per protein used and the tedious manual entry of fragments in the UNIFI software make it likely too stringent and fastidious to use fragments as validation criteria. Therefore, peptides were validated on the basis of their isotopic profiles and mass errors.

B. Targeted profiling of NIST mAb HCPs

In Part II-Chapter 1, we quantified 98 HCPs from a NIST mAb sample on our Q Exactive HF-X instrument run in DIA mode. Based on these data, we selected HCPs that have been quantified using a minimum of 3 unique peptides. From those, 115 peptides from 30 HCPs ranging from 2 to 100 ng/mg mAb were added to a UNIFI library. As these peptides were not identified on BioAccord data, the sequence, mass and RT information have been manually added to the library. In addition, as discussed in the previous section we did not add the mass information of 345 fragments for obvious reasons. This step would be too tedious and the fragmentation method differs between both instruments (HCD vs. in-source), which may have an impact on the MS response of these fragments anyway. The fragment information is not a requirement, as the software will only target precursor ions if no fragments are added to the library.

Then, we performed a first targeted search without restricting the predicted retention times. It is of note that the Q Exactive HF-X was operated in nanoflow compared to the use of the BioAccord operated at microflow. Therefore, even though the gradient is the same, a shift in retention times is expected. Since iRT had been spiked in all samples and followed in all analyses, RT alignment was possible. Next, the identified and extracted peptides signals were manually validated based on their isotopic profile and chromatographic peak shape. This manual verification of the signals was necessary as we observed the presence of signals falsely attributed to our peptides of interest. We observed peptides identified with a single isotope for instance. These peptides were discarded as they are not

sufficiently informative and we do not have a fragmentation spectrum to validate the identification of the peptide in question. Then, doubly or triply charged peptides were identified from an isotopic profile corresponding to a peptide with a different charge. The third example of error corresponds to a peptide identified from the N+1 or N+2 isotopes of another peptide.

After manual verification, we obtained 65 peptides from 21 HCPs of sufficient quality. Then, we kept only the peptides that did not present missing values within the five replicates per injected amount and that showed CV<20% calculated on peptide intensities among the five replicates. This final filtration led to the selection of 23 peptides corresponding to 14 HCPs that were confidently quantified in NIST mAb samples injected at 50, 75 and 100 µg of starting material (Figure 95).

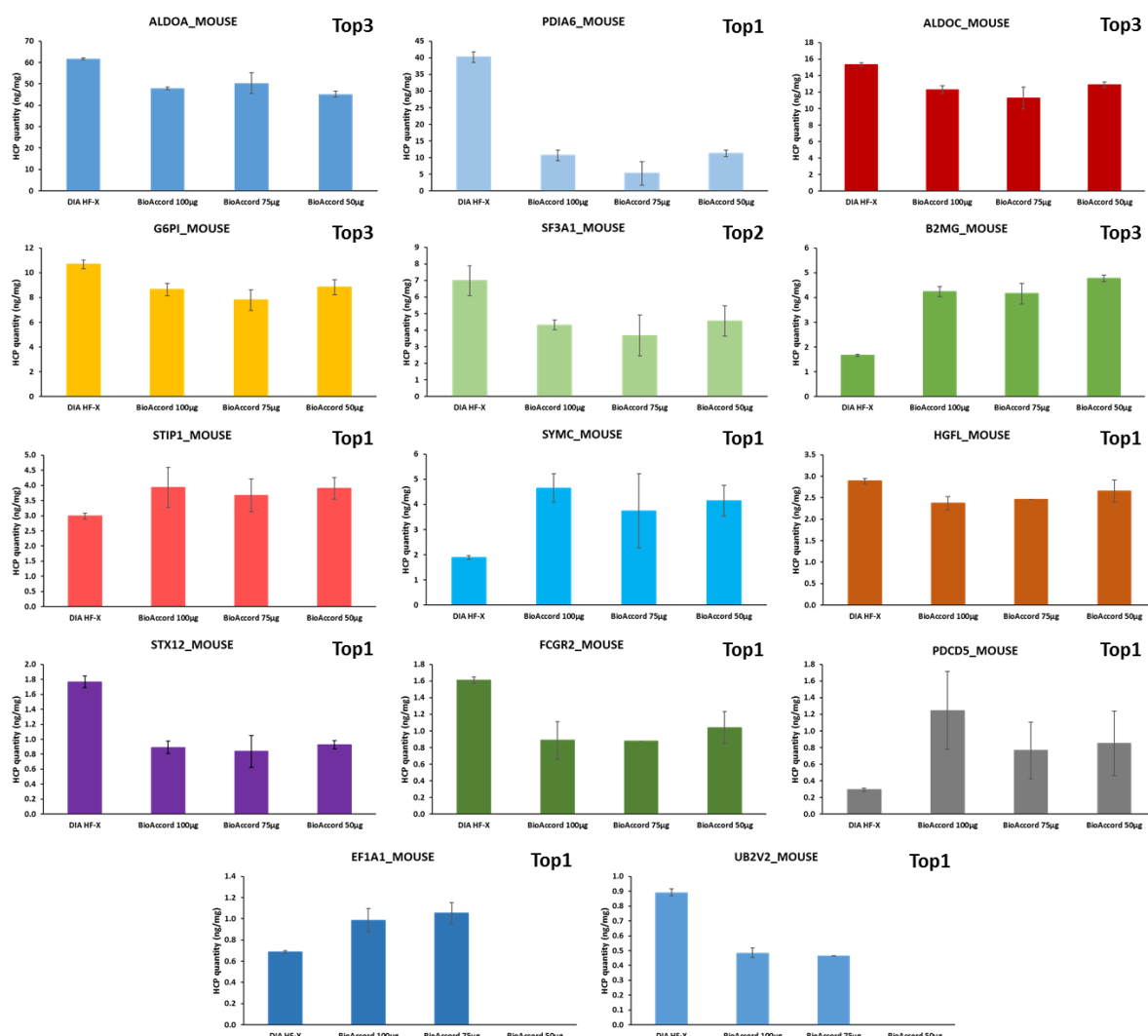


Figure 95: Individual amounts of the 14 HCPs quantified in the NIST mAb sample using the BioAccord LC-MS instrument. Representation of the estimated amounts of the 14 HCPs quantified in the NIST mAb samples injected at 50, 75 and 100 µg, respectively compared to the amounts estimated using our DIA method on the Q Exactive HF-X. Bar heights represent the means of the global HCP amounts in the five injection replicates using the BioAccord and in injection triplicates using the Q Exactive HF-X. Error bars represent the standard deviation over the replicates. The Top1, Top2 or Top3 annotations on each graph represent the numbers of peptides used for quantification within the BioAccord data. Proteins are arranged in decreasing order of quantities.

At first sight, it was noticed that with three different amounts of injected material, the estimated quantities of HCPs were similar. For most of the HCPs, the small error bar demonstrates a highly reproducible quantification using the targeted data processing strategy on the BioAccord. In addition, amounts in the same order of magnitude were estimated when compared to the amounts estimated on the Q Exactive HF-X. We have represented the comparison of these values on Figure 96.

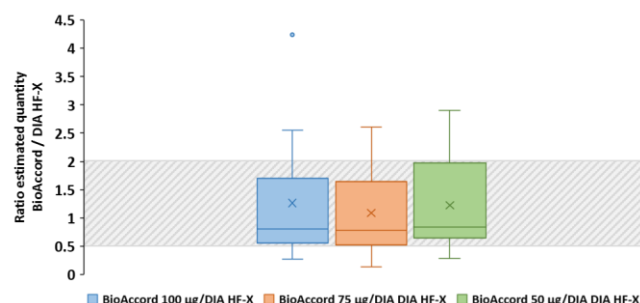


Figure 96: Evaluation of the BioAccord quantification performances over the Top3-DIA approach on the Q Exactive HF-X. Representation of the ratios between the BioAccord HCP estimated amount over the DIA HF-X for the three injected quantity of NIST mAb samples.

We observed that 70% of the 40 values compared for the three amounts of injected material are consistent within a factor 2. These results highlight the promising performance of BioAccord while showing some limitations with respect to the amount of HCP present and the amount of injected material. Indeed, below 4 ng/mg mAb, no HCPs were quantified with more than one peptide (Figure 95). Moreover, the EF1A1_MOUSE and UB2V2_MOUSE proteins were not quantified in the injections of 50 µg of NIST mAb starting material. We noted that the quantity injected needs to be optimized in order to have the HCPs in sufficient amount to be able to quantify them.

4. Conclusion

During the handling of the system, we could observe the easy-to-use feature of the instrument that will allow even unexperienced MS analysts to run injections on the system. In addition, the robustness of the instrument combined with the compliant UNIFI software will allow running the instrument in quality-regulated environments of biopharmaceutical companies.

Our evaluation of the BioAccord LC-MS system highlighted its performance for the quantification of HCPs in a targeted mode on preselected lists of HCPs of interest. The first experiment showed that the BioAccord allows to obtain proportional MS responses and thus to confidently quantify species down to 20 ng/mg mAb. Unfortunately, we could not extend the experiment to the ng/mg mAb or sub-ng/mg mAb levels. It will be interesting to perform this experiment to get a clear idea of the lower quantification limit of the system. The second experiment demonstrated the possibility of applying a Top3 approach to quantify a set of pre-selected HCPs. Indeed, the estimated quantities were comparable to those obtained on a Q-Orbitrap instrument. The main drawback of the developed approach is the data processing. I spent several weeks selecting the HCPs of interest, creating the UNIFI peptide library, and manually verifying all identified and extracted signals.

Finally, the BioAccord instrument can be used in routine analysis for the targeted quantification of HCPs of interest. However, a prior discovery experiment has to be conducted on an instrument dedicated to shotgun proteomics, preferably on a similar chromatographic system, in order to select proteotypic peptides of the targeted HCPs. Once those peptides have been selected, they should be ideally synthesized with an isotopic labeling in order to conduct accurate/absolute quantification using isotope dilution. In these conditions, the BioAccord has a real potential to be applied for the accurate and robust quantification of targeted HCPs.

General conclusion

General conclusion

The objective of my PhD work was to improve the quantification of trace level proteins in complex biological mixtures, namely host cell proteins in therapeutic antibodies, through the development of analytical and bioinformatic strategies using quantitative proteomics based on mass spectrometry.

The first part of this manuscript focused on a bibliographic summary of quantitative proteomics. In this part, the different steps of the "bottom-up" proteomic approach, aiming at identifying proteins by LC-MS/MS, have been described. They include sample preparation, analysis by liquid chromatography coupled to mass spectrometry and protein identification *via* the use of protein sequence databases. Then, the different quantification approaches available in proteomics were detailed with, on the one hand, global quantification strategies (relative quantification based on labelling and relative or "absolute" quantification without labelling), and on the other hand, targeted acquisition methods (SRM, PRM). In addition, the data-independent acquisition approaches, which combine the advantages of global and targeted strategies, and which are becoming increasingly popular, were also presented. Finally, the topic of host cell proteins was addressed. These protein impurities require careful monitoring during the production of monoclonal antibodies. For this purpose, immunospecific and non-immunospecific methods have been presented while paying particular attention to MS, which is gradually becoming an essential method to characterize these impurities.

In this context, the objectives of my PhD work were the following:

- Development of a DIA-based global quantification workflow for the characterization of trace level impurities at all stages of the mAb manufacturing process.
- Improvement of dedicated data processing parameters and strict validation criteria in accordance with the sample complexity.
- Development of a targeted quantification workflow on a MS platform dedicated to routine analysis.

The first part of the results involved several methodological developments for the identification and quantification of host cell proteins by mass spectrometry. First, an innovative quantification standard was evaluated using crude harvest and post protein A fractions and a MS1-XIC data-dependent acquisition approach. This study, conducted in collaboration with ANAQUANT, demonstrated the high intra- and inter-laboratory reproducibility of the HCP Profiler standard. Thus, the HCP profiler allows batch-to-batch comparison and process development support while addressing the need for standardization of the Top3 quantification approach for an implementation in a biopharmaceutical environment.

Then, I developed DIA methods on a Q-Orbitrap instrument using the same sample set. For global protein quantification, the peptide-centric approach of DIA seems particularly interesting compared to DDA as it offers better performance in terms of proteome coverage and quantification accuracy. On the other hand, the spectrum-centric approach of DIA, although very promising in principle, still shows some limitations compared to the peptide-centric approaches. Therefore, I demonstrated that the combination of the HCP Profiler with a peptide-centric DIA strategy enables the monitoring of HCPs impurities throughout the manufacturing process to support the release of safer biotherapeutics.

Next, I also evaluated the influence of a hybrid spectral library-based strategy, which combines DDA and DIA analyses, on samples from the production of four mAbs. The use of the DIA analyses of interest to generate the hybrid spectral library allows the addition of the IRT sources of the peptides to target in the peptide-centric DIA approach. Then, signal extraction is improved with the use of the best available IRT source, especially for standard peptides, and thus more accurate quantification of HCPs is obtained. Furthermore, the application of the hybrid spectral library-based DIA method to various CHO-derived mAbs highlighted the need to monitor HCPs contents at each change in the manufacturing process.

Finally, we have shown that recent advances in MS, combining ion mobility and HR/AM instruments, represent a promising added value in the analytical context of HCPs. Indeed, the performance of the TimsTOF PRO suggests increased coverage due to the additional ion mobility cell. Unfortunately, the increased sensitivity of the instrument, which drastically limits the total proteins/peptides injection amount capacity and the fact that the data extraction software are not yet mature do not allow reliable quantification of protein impurities yet.

The second part of the results was dedicated to the adaptation of the global quantification method for the monitoring of protein impurities present in drug substances. For these purified samples, the detection of HCPs represents a real analytical challenge due to the extreme dynamic range between the overabundant therapeutic protein and the protein impurities. Thus, we have optimized a dedicated analytical workflow including a sample preparation protocol, the use of DIA with a peptide-centric approach, as well as the implementation of specific filters for data processing, in order to obtain robust results. Thanks to the developed workflow, it was possible to identify and quantify several HCPs in various final drug substances and to support bioprocess investigation. Therefore, mass spectrometry represents an efficient orthogonal method for the quality control of these sensitive products.

The third and final part of the results was dedicated to the development of a targeted analytical strategy on a MS platform designed and dedicated to routine analysis. When implementing a MS method in the regulated environment of biopharmaceutical companies, a robust LC-MS system and software that meets data integrity expectations is required. This is the case of the BioAccord LC-MS system driven by the UNIFI software. We evaluated the performance of a targeted data processing approach on this system for the identification and quantification of problematic HCPs. Thus, the BioAccord system demonstrated its robustness and capacity to quantify low level HCPs pre-selected from a discovery study performed on a dedicated instrument, such as HR/AM instruments. Thus, the user-friendly BioAccord LC-MS system can be considered for targeted routine analysis of a set of problematic HCP.

My PhD work highlights the importance of analytical developments, at all levels of the workflow, for proteomic analysis and in particular on samples as challenging as therapeutic proteins with their trace-level impurities.

Throughout this manuscript, I have discussed the contribution of proteomics in the context of therapeutic proteins and more particularly the monitoring of protein impurities. I have also mentioned some potential perspectives in this field. Thus, to conclude, I would like to emphasize some of these points.

A significant part of this work has been focused on the optimization of the DIA approach. This strategy is very popular in the scientific community and I think it has a great future in the field of proteomics and HCPs monitoring in particular. Indeed, the possibility to combine in a single analysis the performances of the global and targeted approaches for the identification and quantification of proteins is continuously demonstrated in different biological contexts. In addition, it represents a real added value compared to the ELISA assay that gives total proteins amount without identification.

Furthermore, since this method theoretically provides information on all detectable peptides in a sample, it offers the possibility of reprocessing the data and combining the results to improve the HCPs coverage of the production organism. Thus, it will be possible to consider reprocessing the data with an improved CHO database, which is still the bottleneck in the search for impurities in CHO-derived mAbs.

However, I have also shown that the processing of these data remains the Achilles' heel of this approach. Considerable efforts have been made in this respect but I am convinced that the implementation of artificial intelligence in bioinformatics tools can further improve this critical step with, for instance, the prediction of fragmentation spectra for the generation of spectral libraries or a better discrimination of real signals from background noise.

In addition to bioinformatics innovations, technical advances on mass spectrometers are continuously carried out. Besides the fact that these instruments are more and more sensitive, fast and accurate, the addition of an extra dimension *via* ion mobility on the very latest generation instruments (timsTOF Pro (Bruker), Exploris 480 with FAIMS (Thermo Fisher Scientific) as an example) promises to further increase their performance in terms of specificity and sensitivity.

Finally, the last point I would like to mention concerns the implementation of MS-based methods in quality control (QC) environments. The use of the HCP Profiler standard represents a first step to standardize the global quantification approach. Indeed, this type of standard will improve the sample preparation step with reduced user-induced analytical biases. In the meantime, mass spectrometers became more and more robust allowing routine analysis. Nevertheless, the main challenge concerns the software that will run the acquisition, data treatment and report the results. These software solutions require to meet data integrity expectations in terms of creation, maintenance, transmission, storage and modification of electronic records. Currently, software like UNIFI that meet the criteria cannot handle the data treatment of complex mixtures of proteins. Even if they run the system from the acquisition to the release of reports, they are more dedicated to intact protein or peptide mapping analysis of a single protein. Thus, an effort is needed to make DIA software like Spectronaut compliant. This action will undoubtedly facilitate the implementation of MS systems in this type of environment.

Experimental section

Experimental section

1. Development of an accurate HCP quantification method to support mAb manufacturing

A. Development of an innovative Top3 quantification strategy dedicated to HCP profiling

i. Sample preparation protocol

Four crude harvest and seven Post Protein A affinity chromatography (PPA) samples of an IgG4 mAb A33 were obtained from a CHO DG44 cell culture as described in Husson *et al.*²⁸⁸ work. Dr. Husson performed the culture, collection and cold acetone precipitation before freezing the samples.

a. Protein quantification

Protein pellets were resuspended in gel loading buffer (10 mM Tris, 1 mM EDTA, 5% β -Mercaptoethanol, 5% SDS, 10% glycerol, pH 6.8). The total protein concentration was determined using the RC DC Protein Assay kit (Bio-Rad laboratories, Hercules, CA, USA) following manufacturer's instructions.

b. Gel stacking

CCCF and PPA samples were stacked in a single band for HCP quantification. Proteins were in-gel reduced with 10 mM DTT for 30 min at 60°C. Alkylation was performed with 55 mM IAA for 30 min in the dark. Then trypsin (Promega, Madison, USA) was added to a 1:50 enzyme:substrate ratio). Samples were incubated overnight at 37°C (14 hours). Peptides were extracted from gel band using 60% ACN, 0.1% FA for 1 hour under agitation and a second step with 100% ACN for 1 hour. After vacuum drying, samples were resolubilized in 2% ACN, 0.1% FA to obtain a final protein concentration of 0.4 $\mu\text{g}/\mu\text{L}$. In all samples, retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) were spiked. For the HCP Profiler solution (Anaquant, Villeurbanne, France) based quantification, one bead was spiked in 150 μL of 0.2 ng/ μL protein solution. For the mix of standard protein, four accurately quantified standard proteins (on column 10 fmol of ADH (yeast alcohol dehydrogenase, P00330), 2 fmol of PYGM (rabbit phosphorylase b, P00489), 0.5 fmol of BSA (bovine serum albumin, P02769) and 0.2 fmol of ENL (yeast enolase, P00924)) from the MassPREP Digestion Standard Kit (Waters, Milford, USA) were spiked.

c. HCP Profiler standard

The HCP Profiler standard is a water-soluble bead, which releases unlabeled peptides at known amounts. Eighteen tripeptides distributed over six concentration points ranging from 1 to 500 fmol, thus a total 54 peptides, are coated on a bead (Table 9). Those 54 peptides areas extraction allows building a calibration curve that is further used to derive each individual HCP amount.

Protein name	Peptide sequence	Injected quantity (fmol)
ANAQUANT-1	DGALLENVTVR	1
	EGAEPEIYNAIR	
	GDVAVFFGLSGTGGK	
ANAQUANT-2	EGCDLAGAIK	
	VGFEENRPTNSILLR	
	VGNPETTLFLVASK	
ANAQUANT-3	LGAADVTVPTLLVAR	
	VGLVPTQEAQK	
	VGQQPEFAAAK	
ANAQUANT-4	AGVVEELAR	2.5
	VGQLLGSGSILR	
	VLGTDGFGGR	
ANAQUANT-5	AGFDFACLPNEGVLAR	
	AGLLEFDDQEPQLQNEIR	
	GGVALSAGVQR	
ANAQUANT-6	GGSGPYFYLK	
	VGIASELGEER	
	VGIDGQINLR	
ANAQUANT-7	AGLAEHGIVFGEPK	10
	EGVRPDIICTGR	
	VGALLSHSNFGSSDCPSSK	
ANAQUANT-8	CGADLGLETIVIER	
	FGTGANTLEVEGENK	
	TGQVVVLGAGPAGYSAFR	
ANAQUANT-9	TGYSGLDYPSEAVIR	
	VGLSGPGLVNLIR	
	YGALVGDVGGTNAR	
ANAQUANT-10	GGLTDAAQQVVAAVEGK	50
	LGGADGNALFR	
	VGLEVTLR	
ANAQUANT-11	AGHPQLAEFTR	
	TGVIGFGSPNK	
	YGINELQANPAK	
ANAQUANT-12	GGTLGQDVIDIR	
	IGTFIDGDEGILLHR	
	YGSIGQPFVYPR	
ANAQUANT-13	GGPLTPVGGGIR	250
	TGVTYDFER	
	YGYQGTPSPVK	
ANAQUANT-14	AGLQAIAGPFSQVR	
	FGCPTGGISPANYR	
	TGSAESILTTGPVVPVIVK	
ANAQUANT-15	FGFSQPLLLGK	
	GFGVTTLDIIR	
	SGDLFNVNAGIVK	
ANAQUANT-16	AGDAFAVIVK	500
	EGQGLTPVLCIGETEAENEAGK	

	IGYQLKPNPAVLICR	
ANQUANT-17	EGSGLLGLTEVTSDCR	
	LGVLVLNCGSSSLK	
	YGTSSVVIDESVIQGIK	
ANQUANT-18	EGLPLTESLALTIDR	
	VGIPYWNITILPR	
	YGYLGNADIEIAAK	

Table 9: Information on the 54 peptides that constitute the HCP Profiler standard.

ii. NanoLC-MS/MS analysis

Analyses were performed on a NanoAcquity UPLC device (Waters) coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Mobile phase A was 0.1% (v/v) formic acid in water and mobile phase B was 0.1% (v/v) formic acid in acetonitrile. The equivalent of 400ng of proteins was trapped onto a Symmetry C18 precolumn (20 mm × 180 µm, 5 µm diameter particles; Waters) and eluted on an Acquity UPLC BEH130 C18 column (250 mm × 75 µm, 1.70 µm particles; Waters).

a. Chromatographic conditions

The analyses were performed on the nanoAcquity UPLC-Q Exactive HF-X coupling described in the Experimental section, section 1.A.ii., with the 115 min gradient detailed Table 10.

Percentage of solvent B	Time (min)
2	0
35	95
80	96
80	101
2	102
2	115

Table 10: Chromatographic gradient used for the analysis of biotherapeutic samples containing HCPs.

b. NanoLC-DDA analysis

The Q Exactive HF-X source temperature was set at 250°C and spray voltage to 2 kV. The system was fully controlled by XCalibur software v4.0.27.19, 2013 (Thermo Scientific) and NanoAcquity UPLC console v1.51.3347 (Waters). DDA injections were performed in a randomized injection sequence. Full scan MS spectra (375-1500 m/z) were acquired in positive mode at a resolution of 120 000 at 200 m/z, a maximum injection time of 60 ms and an AGC target value of 3×10^6 charges. The 10 most intense multiply charged peptides per full scan (charge states ≥ 2) were isolated using a 2 m/z window and fragmented using higher energy collisional dissociation (normalized collision energy set at 27). MS/MS spectra were acquired with a resolution of 15 000 at 200 m/z, a maximum injection time of 60 ms and an AGC target value of 1×10^5 , and dynamic exclusion was set to 40 sec.

iii. DDA data treatment

Raw DDA files were converted to .mgf peaklists using MsConvert and were submitted to Mascot database searches on a local server (version 2.5.1, MatrixScience, London, UK) against a .fasta database including all *Critecutulus griseus* entries extracted from UniProtKB/TrEMBL (56 566 protein entries, February 15, 2021) together with their reversed sequences, as well as the iRT retention time standards (Biognosys), the four standard proteins of the MassPREP Digestion Standard Kit (Waters) or HCP Profiler kit proteins, the mAb heavy and light chains and common contaminants. Spectra were searched with a mass tolerance of 5 ppm in MS mode and 0.05 Da in MS/MS mode. One trypsin missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. Identification results were imported into Proline software version 1.6 (<http://proline.profiroteomics.fr/>; Bouyssié *et al.*, *Bioinformatics* 2020¹⁹) for validation. A false discovery rate of 1% was set at peptide levels using Adjusted e-Value and protein level using Mascot Modified Mudpit scores. Peptides abundances were extracted with Proline software using an extraction m/z tolerance and PSM/Peake1 matching m/z tolerance of 5 ppm. Alignment of the LC-MS runs was performed using Loess smoothing, Peptide Identity method and with a time tolerance of 300 seconds. Cross assignment of peptide ions abundances was performed among PPA or harvest samples using a m/z tolerance of 5 ppm and a retention time tolerance of 40 seconds.

B. Evaluation of DIA strategies for global HCP profiling on a Q-Orbitrap instrument

i. Sample preparation protocol

The same four harvest and seven PPA samples that were prepared in section 1.A.i were analyzed with the DIA mode on the nanoLC-MS/MS system in the Experimental section, section 1.A.ii. In addition, a CHO-DG44 mock cell line sample was provided by UCB Pharma (Braine l'Alleud, Belgium) to generate the spectral library further used for DIA data extraction. Thus, the sample was fractionated using a 1D SDS-PAGE gel and analyzed using DDA mode.

a. Protein precipitation

The mock cell line sample was mixed with 4 volumes of cold acetone. After 1h30 incubation at -20°C, the sample was centrifuged at 14 000 g for 10 min. The supernatant was discarded and the protein pellet was kept.

b. Protein quantification

The protein pellet was resuspended in gel loading buffer (10 mM Tris, 1 mM EDTA, 5% β-Mercaptoethanol, 5% SDS, 10% glycerol, pH 6.8). The total protein concentration was determined using the RC DC Protein Assay kit (Bio-Rad laboratories, Hercules, CA, USA) following manufacturer's instructions.

c. Gel separation

CHO-DG44 mock cell line sample was fractionated onto 12% acrylamide sodium dodecyl sulfate-polyacrylamide gel electrophoresis SDS-PAGE for spectral library generation. Gel bands of the fractionated mock cell line sample were cut into small pieces. Proteins were in-gel reduced with 10 mM DTT for 30 min at 60°C. Alkylation was performed with 55 mM IAA for 30 min in the dark. Then trypsin (Promega, Madison, USA) was added to a 1:50 enzyme:substrate ratio (we estimated 1 µg of

proteins in each band of the mock cell line fractionation). Samples were incubated overnight at 37°C (14 hours). Peptides were extracted from gel band using 60% ACN, 0.1% FA for 1 hour under agitation and a second step with 100% ACN for 1 hour. After vacuum drying, samples were resolubilized in 2% ACN, 0.1% FA to obtain a final protein concentration of 0.4 µg/µL. In all 24 samples, retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) were spiked.

ii. nanoLC-MS/MS analysis

The DDA analyses of the fractionated mock cell line and DIA analyses of CCCF and PPA samples were performed on the nanoAcquity UPLC-Q Exactive HF-X coupling described in the Experimental section, section 1.A.ii., with the 115 min gradient detailed Table 10. The Q-Exactive HF-X source temperature was set at 250°C and spray voltage to 2 kV. The system was fully controlled by XCalibur software v4.0.27.19, 2013 (Thermo Scientific) and NanoAcquity UPLC console v1.51.3347 (Waters).

a. DDA acquisition

DDA analyses of the fractionated mock cell line were performed successively, without a blank between each fraction. The data were acquired using the MS parameters described in the Experimental section, section 1.A.ii.b.

b. DIA acquisition

DIA injections of CCCF and PPA samples were performed in a randomized injection sequence. Full-scan MS spectra were collected from 350–1 250 m/z at a resolution of 60 000 at 200 m/z with an AGC target fixed at $3 \cdot 10^6$ ions and a maximum injection time of 60 ms. Fragments analysis (MS/MS) was subdivided into 40 windows of variable widths. Two acquisitions methods were developed for CCCF and PPA samples (Figure 97). Resolution was set to 30 000 at 200 m/z, AGC target fixed at $1 \cdot 10^6$ ions and with an automatic maximum injection time.

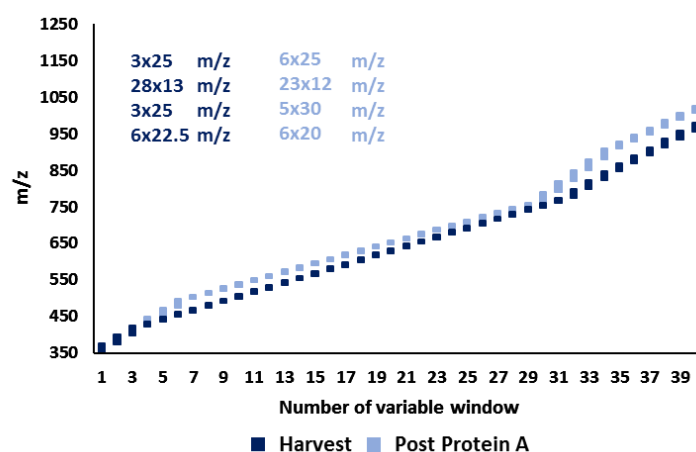


Figure 97: Representation of the 40 variable isolation windows optimized for harvest and PPA samples.

iii. Data treatment

a. DDA data treatment

The data treatment of the DDA analysis is described in the Experimental section, section 1.A.iii.

b. CHO (*Cricetulus griseus*) reference spectral library generation

A reference spectral library combining a series of analyses conducted on different samples in DDA mode was generated using the Spectronaut and Pulsar algorithm (v.14.5; Biognosys). This series of analyses included:

- 24 gel bands obtained by SDS PAGE fractionation of the CHO DG44 mock cell line sample.
- All DDA analyses of current CCCF and PPA samples, including iRT retention time standards (Biognosys) and the 18 proteins from HCP Profiler kit.

Raw DDA files were uploaded into Spectronaut and searched with the Pulsar algorithm against a fasta database containing all *Cricetulus griseus* entries extracted from UniProtKB/TrEMBL (56 687 protein entries, 2021/02/15), as well as the iRT retention time standards (Biognosys), the 18 proteins from HCP Profiler kit, the reference sequence of the mAb and common contaminants).

Trypsin/P enzyme was used and one missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. MS and MS/MS mass tolerances were set in dynamic mode. The spectral library was validated as follows: a False Discovery Rate (FDR) of 0.01 was set at Peptide Spectrum Matches (PSM), Peptides and Proteins levels. Fragment ions window was set between 300 and 1800 m/z with a minimum of 4 fragments per precursor and up to 6.

A second spectral library specific to the HCP Profiler standard was generated using all DDA analysis of the CCCF and PPA samples, including iRT retention time standards (Biognosys) and the 18 proteins from HCP Profiler kit.

Raw DDA files were uploaded into Spectronaut and searched with the Pulsar algorithm against a fasta database containing the 18 proteins from the HCP Profiler standard. Trypsin/P enzyme was used and no missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. MS and MS/MS mass tolerances were set in dynamic mode. The spectral library was validated as follows: a False Discovery Rate (FDR) of 1 was set at Peptide Spectrum Matches (PSM), Peptides and Proteins levels. Fragment ions window was set between 300 and 1800 m/z with a minimum of 3 fragments per precursor and up to 6.

c. DIA data extraction

DIA data of the harvest and PPA samples were analysed in Spectronaut software (v.14.5; Biognosys) with a peptide-centric approach using the upper described in house generated spectral library and with a spectrum-centric approach called directDIA™. For both approaches, trypsin/P was used as digestion enzyme with one missed cleavage allowed. Carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. For quantitative data extraction, MS and MS/MS mass tolerances, Extracted Ion Chromatogram (XIC) and retention time windows were all set as dynamic. iRT regression type was set to local (non-linear) regression. A false discovery rate of 1% was set at precursors and proteins levels. At this extraction stage, a sparse q-value filter was applied. Peptide quantities corresponding to the sum of a minimum of 4 fragments XIC areas and up to 6 fragments (interference correction parameter was turned on) were calculated. Precursors with a Qvalue below 0.01 were used for IRT profiling.

C. Case study: Impact of the bioreactor capacity on HCPs profile

i. Sample preparation protocol

UCB Pharma provided nine crude harvest and nine PPA fractions from the production of four mAbs obtained after a CHO DG44 cell culture in bioreactors of different capacities.

a. Protein precipitation

The 18 samples were mixed with 4 volumes of cold acetone. After 1h30 incubation at -20°C, samples were centrifuged at 14 000 g for 10 min. The supernatant was discarded and the protein pellets were kept.

b. Protein quantification

Protein pellets were resuspended in gel loading buffer (10 mM Tris, 1 mM EDTA, 5% β-Mercaptoethanol, 5% SDS, 10% glycerol, pH 6.8). The total protein concentration was determined using the RC DC Protein Assay kit (Bio-Rad laboratories, Hercules, CA, USA) following manufacturer's instructions.

c. Gel stacking

CCCF and PPA samples were stacked in a single band for HCP quantification. Proteins were in-gel reduced with 10 mM DTT for 30 min at 60°C. Alkylation was performed with 55 mM IAA for 30 min in the dark. Then trypsin (Promega, Madison, USA) was added to a 1:50 enzyme:substrate ratio. Samples were incubated overnight at 37°C (14 hours). Peptides were extracted from gel bands using 60% ACN, 0.1% FA for 1 hour under agitation and a second step with 100% ACN for 1 hour. After vacuum drying, samples were resolubilized in 2% ACN, 0.1% FA to obtain a final protein concentration of 0.4 µg/µL. In all samples, retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) were spiked. In addition, four accurately quantified standard proteins (on column 10 fmol of ADH (yeast alcohol dehydrogenase, P00330), 2 fmol of PYGM (rabbit phosphorylase b, P00489), 0.5 fmol of BSA (bovine serum albumin, P02769) and 0.2 fmol of ENL (yeast enolase, P00924)) from the MassPREP Digestion Standard Kit (Waters, Milford, USA) were spiked.

ii. NanoLC-DIA analysis

DIA analyses of CCCF and PPA samples were performed on the nanoAcquity UPLC-Q Exactive HF-X coupling described in the Experimental section, section 1.A.ii., with the 115 min gradient detailed Table 10. The Q Exactive HF-X source temperature was set at 250°C and spray voltage to 2 kV. The system was fully controlled by XCalibur software v4.0.27.19, 2013 (Thermo Scientific) and NanoAcquity UPLC console v1.51.3347 (Waters). DIA injections were performed in a randomized injection sequence with the injection of all PPA samples followed by all harvest sample. Full-scan MS spectra were collected from 350–1 250 m/z at a resolution of 60 000 at 200 m/z with an AGC target fixed at 3.10^6 ions and a maximum injection time of 60 ms. Fragments analysis (MS/MS) was subdivided into 40 windows of variable widths. Two acquisitions methods were used for CCCF and PPA samples. Both methods are described in the Experimental section, section 1.B.ii.b. Resolution was set to 30 000 at 200 m/z, AGC target fixed at 1.10^6 ions and with an automatic maximum injection time.

iii. DIA data treatment

a. CHO (*Cricetulus griseus*) spectral library generation

Two spectral libraries, DDA and hybrid, combining a series of analyses conducted on different samples in DDA and DIA mode were generated using the Spectronaut and Pulsar algorithm (v.14.5; Biognosys). The series of analyses used for the DDA spectral library included:

- 24 gel bands obtained by SDS PAGE fractionation of the CHO DG44 mock cell line sample.
- All DDA analyses of the mAb A33 CCCF and PPA samples (section 1.A), including iRT retention time standards (Biognosys) and the 4 standard proteins from the MassPREP Digestion Standard Kit (Waters, Milford, USA).

The hybrid library was generated from the combination of the DIA analyses of the 18 CCCF and PPA samples of the four mAbs with the search archive of the DDA library described above.

Raw DDA files were uploaded into Spectronaut and searched with the Pulsar algorithm against a fasta database containing all *Critecutulus griseus* entries extracted from UniProtKB/TrEMBL (56 687 protein entries, 2021/02/15), as well as the iRT retention time standards (Biognosys), the 18 proteins from HCP Profiler kit, the reference sequence of the mAb and common contaminants).

Trypsin/P enzyme was used and one missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. MS and MS/MS mass tolerances were set in dynamic mode. The spectral library was validated as follows: a False Discovery Rate (FDR) of 0.01 was set at Peptide Spectrum Matches (PSM), Peptides and Proteins levels. Fragment ions window was set between 300 and 1800 m/z with a minimum of 4 fragments per precursor and up to 6.

b. DIA data extraction

DIA data of the harvest and PPA samples were analysed in Spectronaut software (v.14.5; Biognosys) with a peptide-centric approach using the upper described DDA and hybrid spectral libraries and with a spectrum-centric approach called directDIA™. For both approaches, trypsin/P was used as digestion enzyme with one missed cleavage allowed. Carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. For quantitative data extraction, MS and MS/MS mass tolerances, Extracted Ion Chromatogram (XIC) and retention time windows were all set as dynamic. iRT regression type was set to local (non-linear) regression. A false discovery rate of 1% was set at precursors and proteins levels. At this extraction stage, a sparse q-value filter was applied. Peptide quantities corresponding to the sum of a minimum of 4 fragments XIC areas and up to 6 fragments (interference correction parameter was turned on) were calculated. Precursors with a Qvalue below 0.01 were used for IRT profiling.

D. Benefit of ion mobility in the field of HCPs monitoring

i. Sample preparation protocol

A new sample preparation was performed for the four crude harvest and seven PPA samples of an IgG4 mAb A33, discussed in the Experimental section, section 1.A et 1.B. In addition, the mock cell line sample prepared as described in Experimental section, section 1.B.i. was also injected on the TimsTOF PRO instrument using ddaPASEF acquisition mode.

a. Protein quantification

Protein pellets were resuspended in gel loading buffer (10 mM Tris, 1 mM EDTA, 5% β -Mercaptoethanol, 5% SDS, 10% glycerol, pH 6.8). The total protein concentration was determined using the RC DC Protein Assay kit (Bio-Rad laboratories, Hercules, CA, USA) following manufacturer's instructions.

b. Gel stacking

CCCF and PPA samples were stacked in a single band for HCP quantification. Proteins were in-gel reduced with 10 mM DTT for 30 min at 60°C. Alkylation was performed with 55 mM IAA for 30 min in the dark. Then trypsin (Promega, Madison, USA) was added to a 1:50 enzyme:substrate ratio). Samples were incubated overnight at 37°C (14 hours). Peptides were extracted from gel band using 60% ACN, 0.1% FA for 1 hour under agitation and a second step with 100% ACN for 1 hour. After vacuum drying, samples were resolubilized in 2% ACN, 0.1% FA to obtain a final protein concentration of 0.4 $\mu\text{g}/\mu\text{L}$. In all samples, retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) were spiked. In addition, one bead of the HCP Profiler standard (section 1.A.i.c, Table 1) was spiked in 150 μL of 0.2 $\text{ng}/\mu\text{L}$ protein solution. Due to the increased sensitivity of the TimsTOF PRO mass spectrometer, we had to inject less than 400 ng of protein amount. Thus, we decreased the injection volume to load 273 ng of protein onto the column. As a result, the range of the HCP Profiler standard was also reduced to a range between 0.68 to 341 fmol (Table 11).

Calibration point	Protein name	Injected quantity (fmol)
1	ANAQUANT-1	0.68
	ANAQUANT-2	
	ANAQUANT-3	
2	ANAQUANT-4	1.7
	ANAQUANT-5	
	ANAQUANT-6	
3	ANAQUANT-7	6.8
	ANAQUANT-8	
	ANAQUANT-9	
4	ANAQUANT-10	34
	ANAQUANT-11	
	ANAQUANT-12	
5	ANAQUANT-13	171
	ANAQUANT-14	
	ANAQUANT-15	
6	ANAQUANT-16	341
	ANAQUANT-17	
	ANAQUANT-18	

Table 11: Dynamic range covered by the HCP Profiler standard proteins.

ii. NanoLC-IM-MS/MS analysis

Analyses were performed on a NanoElute UPLC device coupled to a TimsTOF PRO mass spectrometer (Bruker, Bremen, Germany). Mobile phase A was 0.1% (v/v) formic acid in water and mobile phase B was 0.1% (v/v) formic acid in acetonitrile. The equivalent of 273 ng of proteins was trapped onto a nanoEase M/Z Symmetry C18 precolumn (20 mm × 180 µm, 5 µm diameter particles; Waters) and eluted on an Aurora C18 column (250 mm × 75 µm, 1.6 µm particles; IonOptics). Peptide elution was performed using the 115 min gradient previously described in the Experimental section, section 1.A.ii.a (Table 10).

The TimsTOF PRO source temperature was set at 180°C and spray voltage to 1.6 kV. The system was fully controlled by otofControl software v6.0 and HyStar v5.0 (Bruker). For ddaPASEF and diaPASEF, the collision energy was ramped linearly as a function of the mobility from 52 eV at $1/K0 = 1.6 \text{ Vs cm}^{-2}$ to 20 eV at $1/K0 = 0.6 \text{ Vs cm}^{-2}$.

a. ddaPASEF

The dual TIMS analyzer was operated at a fixed duty cycle close to 100% using equal accumulation and ramp times of 100 ms each. We performed DDA in PASEF mode with 10 PASEF scans per topN acquisition cycle. Singly charged precursors were excluded by their position in the m/z -ion mobility plane, and precursors that reached a target value of 17 000 arbitrary units were dynamically excluded for 0.4 min. The quadrupole isolation width was set to 2 m/z for $m/z < 700$ and to 3 m/z for $m/z > 700$ and MS spectra were collected from 100–1 700 m/z .

b. diaPASEF

In DIA experiment, we defined 32×25 Th isolation windows from m/z 100 to 1 700. To adapt the MS1 cycle time in diaPASEF, we set the repetitions to 2. Thus, 64 windows of 25 m/z are covered in 16 TIMS cycle of 100 ms. Concretely, four m/z windows are selected and fragmented in a TIMS elution. The windows overlap in the ion mobility dimension to reduce potential artifacts linked to reduced ion transmission at the edges of the diaPASEF windows.

iii. Data treatment

a. DDA data treatment

Raw DDA files were converted to .mgf peaklists using Data analysis (v5.3, Bruker) and were submitted to Andromeda database searches included in MaxQuant software (version 1.6.10.43, Max Planck Institute of Biochemistry, Germany) against a .fasta database including all *Critecutulus griseus* entries extracted from UniProtKB/TrEMBL (56 566 protein entries, February 15, 2021) together with their reversed sequences, as well as the iRT retention time standards (Biognosys), the HCP Profiler kit proteins, the mAb heavy and light chains and common contaminants. MS and MS/MS tolerances as well as other TimsTOF PRO parameters were set using the feature Bruker TIMS in the instrument type section of MaxQuant. Trypsin/P was used as digestion enzyme with one missed cleavage allowed. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. A false discovery rate of 1% was set at peptide, PSM and protein levels. Peptides abundances were extracted and match between runs feature was set on.

b. CHO (*Cricetulus griseus*) reference spectral library generation

A reference spectral library combining a series of analyses conducted on different samples in DDA mode was generated using the Spectronaut and Pulsar algorithm (v.14.5; Biognosys). This series of analyses comprised:

- 24 gel bands obtained by SDS PAGE fractionation of the CHO DG44 mock cell line sample.
- All DDA analyses of current CCCF and PPA samples., including iRT retention time standards (Biognosys) and the 18 proteins from HCP Profiler kit.

Raw DDA files were uploaded into Spectronaut and searched with the Pulsar algorithm against a fasta database containing all *Critecutulus griseus* entries extracted from UniProtKB/TrEMBL (56 687 protein entries, 2021/02/15), as well as the iRT retention time standards (Biognosys), the 18 proteins from HCP Profiler kit, the reference sequence of the mAb and common contaminants).

Trypsin/P enzyme was used and one missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. MS and MS/MS mass tolerances were set in dynamic mode. The spectral library was validated as follows: a False Discovery Rate (FDR) of 0.01 was set at Peptide Spectrum Matches (PSM), Peptides and Proteins levels. Fragment ions window was set between 300 and 1800 m/z with a minimum of 4 fragments per precursor and up to 6.

c. DIA data extraction

DIA data of the harvest and PPA samples were analysed in Spectronaut software (v.14.5; Biognosys) with a peptide-centric approach using the upper described in house generated spectral library and with a spectrum-centric approach called directDIA™. For both approaches, trypsin/P was used as digestion enzyme with one missed cleavage allowed. Carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. For quantitative data extraction, MS and MS/MS mass tolerances, Extracted Ion Chromatogram (XIC), retention time and ion mobility windows were all set as dynamic. iRT regression type was set to local (non-linear) regression. A false discovery rate of 1% was set at precursors and proteins levels. At this extraction stage, a sparse q-value filter was applied. Peptide quantities corresponding to the sum of a minimum of 4 fragments XIC areas and up to 6 fragments (interference correction parameter was turned on) were calculated. Precursors with a Qvalue below 0.01 were used for IRT profiling.

2. The analytical challenge of drug product

A. Sample preparation improvements for drug substances

i. Sample preparation protocol

Optimizations of the native liquid digestion protocol, proposed by Huang et al.²⁹, were performed using the humanized IgG1 RM8671 monoclonal antibody (NIST mAb) from Sigma-Aldrich (Merck, Darmstadt, Germany). Three technical replicate of each condition were prepared.

a. Optimized native liquid digestion

One milligram of NIST mAb starting material, in 25 mM Tris-HCL buffer (pH 8), was digested overnight at 37°C (14 hours) with Trypsin/Lys-C enzymes solution (Promega, Madison, USA) in a 1:400 enzyme-protein ratio condition. Digests were reduced with 3 mM Dithiothreitol (DTT) for 10 min at 90°C, then centrifuged at 13 000g for 2 min. Finally, the supernatants were finally acidified with 0.5 µL of pure Formic Acid (FA). Samples were then dried upon speed vacuum concentrator and peptides were resolubilized in 2% ACN, 0.1% FA to obtain a final protein concentration of 0.4 µg/µL. Retention time standards iRT (Biognosys, Schlieren, Switzerland) and four accurately quantified standard proteins (on-column 10 fmol of ADH (yeast alcohol dehydrogenase, P00330), 2 fmol of PYGM (phosphorylase b, P00489), 0.5 fmol of BSA (bovin serum albumin, P02769), and 0.2 fmol of ENL (yeast enolase, P00924) from the MassPREP Digestion Standard Kit (Waters, Milford, MA, USA) were spiked in each sample prior to injection.

b. Covaris focused ultrasonication to enhance digestion

One milligram of NIST mAb starting material in 25 mM Tris-HCL buffer (pH 8) was used for the two conditions of the experiment. Trypsin/Lys-C enzyme solution (Promega, Madison, USA) was added to each sample to perform digestion in the Covaris S220 ultrasonicator in a 1:400 enzyme to protein ratio condition. After the addition of the proteases, the samples were transferred to a microTUBE AFA Fiber Pre-split-Cap (Covaris) and were placed in the Covaris S220 bath at **30°C**. An Adapted Focused Acoustic (AFA) step was performed with acoustic energy delivered at **2.5W continuously** for the duration of the digestion step. Thus, **two samples** were subjected to a **10 min** digestion step and **two samples** underwent a **30 min** digestion step (Table 12).

Covaris settings	10 min digestion	30 min digestion
Peak Incidence Power (W)	10	10
Duty factor (%)	25	25
Bath temperature	30°C	30°C
Duration	10 min - continuous	30 min - continuous

Table 12: Covaris focused ultrasonication to enhance the digestion step.

After the focused ultrasonication/digestion step, samples were transferred in an Eppendorf tube. Digests were reduced with 3 mM Dithiothreitol (DTT) for 10 min at 90°C, then centrifuged at 13 000g for 2 min. Finally, the supernatants were finally acidified with 0.5 µL of pure Formic Acid (FA). Samples were then dried upon speed vacuum concentrator and peptides were resolubilized in 2% ACN, 0.1% FA to obtain a final protein concentration of 0.4 µg/µL. Retention time standards iRT (Biognosys, Schlieren, Switzerland) and four accurately quantified standard proteins (on-column 10 fmol of ADH (yeast alcohol dehydrogenase, P00330), 2 fmol of PYGM (phosphorylase b, P00489), 0.5 fmol of BSA (bovin serum albumin, P02769), and 0.2 fmol of ENL (yeast enolase, P00924) from the MassPREP Digestion Standard Kit (Waters, Milford, MA, USA) were spiked in each sample prior to injection.

ii. nanoLC-MS/MS analysis

DIA analyses of NIST mAb samples were performed on the nanoAcquity UPLC-Q Exactive HF-X coupling described in the Experimental section, section 1.A.ii., with the 115 min gradient detailed Table 10. The Q Exactive HF-X source temperature was set at 250°C and spray voltage to 2 kV. The system was fully

controlled by XCalibur software v4.0.27.19, 2013 (Thermo Scientific) and NanoAcquity UPLC console v1.51.3347 (Waters). DIA injections were performed in a randomized injection sequence. Full-scan MS spectra were collected from 350–1 250 m/z at a resolution of 60 000 at 200 m/z with an AGC target fixed at 3.10^6 ions and a maximum injection time of 60 ms. Fragments analysis (MS/MS) was subdivided into 60 windows of variable widths (Figure 98). Resolution was set to 30 000 at 200 m/z, AGC target fixed at 1.10^6 ions and with an automatic maximum injection time.

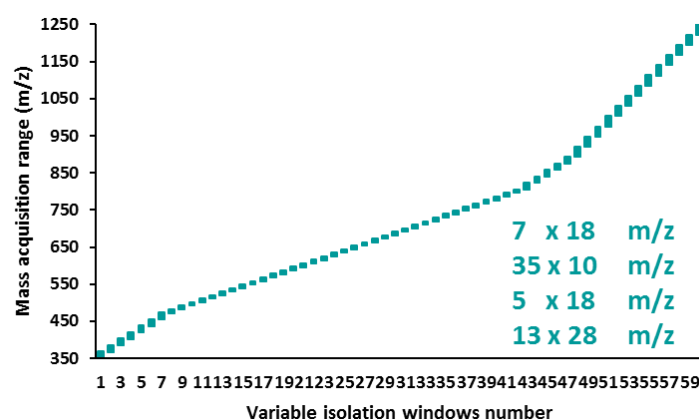


Figure 98: Representation of the 60 variable isolation windows optimized for drug substances.

iii. DIA data treatment

a. CHO (*Cricetulus griseus*) reference spectral library generation

A spectral library combining a series of analyses conducted on different samples in DDA mode was generated using the Spectronaut and Pulsar algorithm (v.14.5; Biognosys). This series of analyses comprised:

- 24 gel bands obtained by SDS PAGE fractionation of the NIST mAb sample.
- The 16 DDA analyses of current NIST mAb samples, including iRT retention time standards (Biognosys) and the 4 standard proteins from the MassPREP Digestion Standard Kit (Waters, Milford, USA).

Raw DDA files were uploaded into Spectronaut and searched with the Pulsar algorithm against a fasta database containing all *Mouse* entries extracted from UniProtKB/Swiss-Prot (17 082 protein entries, 2019/03/22), as well as the iRT retention time standards (Biognosys), the 4 standard proteins from the MassPREP Digestion Standard Kit, the reference sequence of the mAb and common contaminants).

Trypsin/P enzyme was used and one missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. MS and MS/MS mass tolerances were set in dynamic mode. The spectral library was validated as follows: a False Discovery Rate (FDR) of 0.01 was set at Peptide Spectrum Matches (PSM), Peptides and Proteins levels. Fragment ions window was set between 300 and 1800 m/z with a minimum of 4 fragments per precursor and up to 6.

b. DIA data extraction

DIA data of the NIST mAb samples were analysed in Spectronaut software (v.14.5; Biognosys) with a peptide-centric approach using the upper described in house generated spectral library. For Trypsin/P was used as digestion enzyme with one missed cleavage allowed. Carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues and acetylation of proteins

n-termini were set as variable modifications. For quantitative data extraction, MS and MS/MS mass tolerances, Extracted Ion Chromatogram (XIC) and retention time windows were all set as dynamic. iRT regression type was set to local (non-linear) regression. A false discovery rate of 1% was set at the precursors level. At this extraction stage, a sparse q-value filter was applied. Peptide quantities corresponding to the sum of a minimum of 4 fragments XIC areas and up to 6 fragments (interference correction parameter was turned on) were calculated. Precursors with a Qvalue below 0.01 were used for IRT profiling.

B. Case study: Manufacturing process investigation

i. Sample preparation protocol

Five drug substances, provided by UCB Pharma, were subjected to the native digestion protocol described in the experimental section, section 2.A.i.a.

ii. NanoLC-MS/MS analysis

DIA analyses of the five samples were performed on the nanoAcquity UPLC-Q Exactive HF-X coupling described in the Experimental section, section 1.A.ii., with the 115 min gradient detailed Table 10. The Q Exactive HF-X source temperature was set at 250°C and spray voltage to 2 kV. The system was fully controlled by XCalibur software v4.0.27.19, 2013 (Thermo Scientific) and NanoAcquity UPLC console v1.51.3347 (Waters). DIA injections were performed in a randomized injection sequence. Full-scan MS spectra were collected from 350–1 250 m/z at a resolution of 60 000 at 200 m/z with an AGC target fixed at 3.10^6 ions and a maximum injection time of 60 ms. Fragments analysis (MS/MS) was subdivided into 60 windows of variable widths, as described in the experimental section, section 2.A.ii (Figure 2). Resolution was set to 30 000 at 200 m/z, AGC target fixed at 1.10^6 ions and with an automatic maximum injection time.

iii. DIA data treatment

DIA data of the five DS samples were analysed in Spectronaut software (v.14.5; Biognosys) with a peptide-centric approach using the in house generated spectral library described in the Experimental section, section 1.B.iii.b. For Trypsin/P was used as digestion enzyme with one missed cleavage allowed. Carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. For quantitative data extraction, MS and MS/MS mass tolerances, Extracted Ion Chromatogram (XIC) and retention time windows were all set as dynamic. iRT regression type was set to local (non-linear) regression. A false discovery rate of 1% was set at the precursors level. At this extraction stage, a sparse q-value filter was applied. Peptide quantities corresponding to the sum of a minimum of 4 fragments XIC areas and up to 6 fragments (interference correction parameter was turned on) were calculated. Precursors with a Qvalue below 0.01 were used for IRT profiling.

3. Development of a HCP profiling method on a MS platform for routine analysis

A. Evaluation of the BioAccord LC-MS system

i. Sample preparation protocol

The evaluation of the BioAccord LC-MS System was performed using the humanized IgG1 RM8671 monoclonal antibody (NIST mAb) from Sigma-Aldrich (Merck, Darmstadt, Germany). The NIST mAb samples were subjected to the native liquid digestion protocol described in the experimental section, section 2.B.i.

ii. LC-MSe analysis

Analyses were performed on a ACQUITY UPLC I-Class Plus system coupled to a BioAccord mass spectrometer (Waters). Mobile phase A was 0.1% (v/v) formic acid in water and mobile phase B was 0.1% (v/v) formic acid in acetonitrile. The injected peptides were eluted on an ACQUITY PREMIER CSH (C18, 130 Å, 1.7 µm, 2.1 x 150 mm) column with the 115 min gradient detailed in the Experimental section, section 1.A.ii.a (Table 10). The BioAccord source temperature was set at 550°C and spray voltage to 1.5 kV. The system was fully controlled by UNIFI software (v1.9.13.9, Waters). Full-scan MS spectra followed by fragments analysis were performed in positive mode from 50–2 000 m/z at a scan speed of 5 Hz. MS scan of the precursor were acquired with a cone voltage of 30 V and fragments were acquired with a cone voltage ramp from 60 to 120 V.

iii. Targeted data treatment

a. HCP monitoring simulation

Data acquisition and data treatment were conducted using UNIFI software. Previous data of the NIST mAb samples spiked with the four standard proteins of the MassPREP digestion kit were acquired on the BioAccord LC-MS system. A *Pep Map MS* processing method with a library composed of the sequence of the NIST mAb, the iRT retention time standards (Biognosys), the 4 standard was applied to observe the sequence coverage of each proteins. Spectra were searched with a mass tolerance of 15 ppm in MS mode and 20 ppm in MS/MS mode. One trypsin missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. From this search, 27 standard protein peptides were selected with 3 fragments per peptides based on their intensity and MS response (Table 13).

Protein Name	Peptide sequence	Expected m/z	Retention time (min)
ADH1	YVVDTSK	406.216	8
ADH1	IGDYAGIK	418.730	16
ADH1	ANELLINVK	507.306	29
ADH1	SISIVGSYVGNR	626.341	30
ADH1	DIVGAVLK	407.758	32
ADH1	VLGIDGGEGKEELFR	809.932	33
ADH1	EALDFFAR	484.749	37

ADH1	VVGLSTLPEIYEK	724.412	42
PYGM	NNVVNTMR	474.244	9
PYGM	NLAENISR	458.748	12
PYGM	MSLVEEGAVK	531.779	22
PYGM	VAAAFPGDVDR	559.289	23
PYGM	VLVDLER	422.252	24
PYGM	EIWGVEPSR	536.781	26
PYGM	VIFLENYR	527.294	33
PYGM	LLSYVDDEAFIR	720.876	42
BSA	CCTESLVNR	569.756	12
BSA	YICDNQDTISSK	722.325	14
BSA	LVTDLTK	395.240	17
BSA	EYEATLEECCA	751.814	18
BSA	EACFAVEGPK	554.262	20
BSA	YLYEIAR	464.252	24
ENO1	IATAIEK	373.227	9
ENO1	LNQLLR	378.741	16
ENO1	GNPTVEVELTTEK	708.857	23
ENO1	AADALLK	407.755	25
ENO1	NVNDVIAPAFVK	643.858	37
NIST mAb	SLSLSPGK	394.732	15
NIST mAb	ALPAPIEK	419.759	19
NIST mAb	DTLMISR	418.225	20
NIST mAb	FNWYVDGVEVHNAK	839.411	32

Table 13: List of peptides selected from the Pep Map MS search.

Then, the analyses of the HCP monitoring simulation experiment were processed using the *Accurate Mass Screening on MSe data* method with a library composed of the sequence, precursor and fragment masses and retention time of the selected peptides. Finally, identified and extracted signals were manually validated based on their isotopic profile and chromatogram peak shape.

b. NIST mAb HCP profiling

For this experiment, the analyses were processed using the *Accurate Mass Screening on MSe data* method with a library composed of the sequence of the 115 selected peptides from 30 HCPs observed on the NIST mAb analysis detailed in the Experiment section, section 2.A (Table 14). Spectra were searched with a mass tolerance of 15 ppm in MS mode and 20 ppm in MS/MS mode. No restriction about the peptide retention times were set as the selected peptides were observed on a nanoLC-Q Exactive HF-X instrument operated in nanoflow compared to the BioAccord at microflow.

Protein name	Peptide sequence
ALDOA_MOUSE	GILAADESTGSIK
	PHPYPALTPEQK
	QLLLTADDR
	ADDGRFPQVIK
	ALANSLACQGK
PDIA6_MOUSE	HQSLGGQYGVQGFPTIK
	TGEAIVDAALSALR
	AATALKDVK
ALDOC_MOUSE	ALQASALNAWR
	LSQIGVENTEENR
	QVLFSADDR
	LSQIGVENTEENRR
G6PI_MOUSE	HFVALSTNTAK
	TFTTQETITNAETAK
	VFEGNRPTNSIVFTK
	EWFLAAK
SF3A1_MOUSE	IHEATGMPAGK
	LQYEGIFIK
	VQVPNMQDK
	TQQAQANITLQEQIEAIHK
	TEDSLMPEEEFLR
PPN_MOUSE	DAVVDGTPCEPGQR
	GYNQIFIIPAGATSIR
	WLPYYAAPNK
	GDEGSPVHAAACLLK
STIP1_MOUSE	DAIHFYNK
	DPQALSEHLK
	LAYINPDLEEK
	DPQALSEHLK
HGFL_MOUSE	GPWCYTTNR
	SPLNDFQLFR
	TAGGLPCQAWSR
	VVGGHGPNPWTVSLR
SYMC_MOUSE	NQVAAEVAK
	QLALAEGKPIETPK
	PAAVEAVTAAGSQHIQTLTDEVTK
B2MG_MOUSE	TPQIQVYSR
	TVYWDRDM
	VEMSDMSFSK
FCGR2_MOUSE	LEPPWIQVLK
	SQVQASYTFK
	TLHQSKPVTITVQGPK

	PVTITVQGPK
TKT_MOUSE	ILATPPQEDAPSVDIANIR
	NSTFSELFK
	VLDPFTIKPLDR
	LGQSDPAPLQHQQVDIYQK
	TSRPENAIISNNEDFQVGQAK
	SVPMSTVFYPSDGVATEK
KAD2_MOUSE	DDITGEPLIR
	LEAYHTQTTPLVEYYR
	NGFLLDGFPR
	TRLEAYHTQTTPLVEYYR
UB2V2_MOUSE	VECGSKYPEAPPSVR
	VILQELR
	YPEAPPSVR
	LLEELEEGQK
PCBP1_MOUSE	IITLTGPTNAIFK
	INISEGNCPER
	LVVPATQCGSLIGK
	LEEDINSSMTNSTAASRPPVTLR
EF1A1_MOUSE	IGGIGTVPVGR
	QTVAVGVK
	YYVTIIDAPGHR
AIFM1_MOUSE	AIASATEGGSVPQIR
	LAGENMTGAAK
	LNDGSQITFEK
	LNDGSQITFEK
	ILPQYLSNWTMEK
FUMH_MOUSE	AIEMLGGEGLGSK
	IYELAAGGTAVGTGLNTR
	VAALTGLPFVTAPNK
NPS3B_MOUSE	PGGPALWGNAFK
	TNEFLENFK
	VHVLWWNESADSR
	FLIPNLAFIDK
ITIH5_MOUSE	LVGAPEEYGK
	SYLEITPSR
	LIDGVYK
MDHM_MOUSE	MIAEAIPELK
	TIIP LISQCTPK
	VNVPVIGGHAGK
LIPR2_MOUSE	IFPWSPEDIDTR
	TEYTQASYNTR
	VINLFRPTMGASQITVQR

PRDX5_MOUSE	LLADPTGAFGK
	THLPGFVEQAGALK
	VNLAELFK
	VGDAIPSVFEGEPGKK
PDCD5_MOUSE	AVENYLIQMAR
	LSNLALVKPEK
	NSILAQVLDQSAR
	LAELQAK
TAGL2_MOUSE	LINSLYPEGQAPVK
	NVIGLQMGTR
	TLMNLGGLAVAR
IRF4_MOUSE	SNDFEELVER
	QWLIDQIDSGK
	YPGLVWENEEK
NDKB_MOUSE	NIIHGSDSVESAEK
	TFIAIKPDGVQR
	VMLGETNPADSK
	VMLGETNPADSKPGTIR
BLVRB_MOUSE	HDLGHFMLR
	IAIFGATGR
	ILQESGLK
	LPSEGPQPAHVVG DVR
	LQDVTDDHIR
CAP1_MOUSE	LEAVSHTSDMHCGYGDSPSK
	SALFAQINQGESITHALK
	EMNDAAMFYTR
STX12_MOUSE	ELGSLPLPLSASEQR
	NLMSQLGTK
	LQENLQQHQHSTNQLAK
	DFNSIIQTCSGNIQR

Table 14: List of HCP peptides selected on the nanoLC-Q Exactive HF-X analyses.

Identified and extracted signals were manually validated based on their isotopic profile and chromatogram peak shape. Then, we kept only the peptides that did not present missing values within the five replicates per injected amount and that showed CV<20% calculated on peptide intensities among the five replicates. Finally, the 2 pmol of ADH was used to estimate the amount of the targeted HCPs via a Top3 quantification strategy.

List of communications

Publications

Hernandez-Alba, O.; Houel, S.; Hessmann, S.; Erb, S.; Rabuka, D.; Huguet, R.; Josephs, J.; Beck, A.; Drake, P. M.; Cianférani, S., **A Case Study to Identify the Drug Conjugation Site of a Site-Specific Antibody-Drug-Conjugate Using Middle-Down Mass Spectrometry.** *Journal of the American Society for Mass Spectrometry* **2019**, *30* (11), 2419-2429.

Dovgan, I.; Hentz, A.; Koniev, O.; Ehkirch, A.; Hessmann, S.; Ursuegui, S.; Delacroix, S.; Riomet, M.; Taran, F.; Cianférani, S.; Kolodych, S.; Wagner, A., **Automated linkage of proteins and payloads producing monodisperse conjugates.** *Chemical Science* **2020**, *11* (5), 1210-1215.

Sornay, C.; Hessmann, S.; Erb, S.; Dovgan, I.; Ehkirch, A.; Botzanowski, T.; Cianférani, S.; Wagner, A.; Chaubet, G., **Investigating Ugi/Passerini Multicomponent Reactions for the Site-Selective Conjugation of Native Trastuzumab.** *Chemistry – A European Journal* **2020**, *26* (61), 13797-13805.

Belorusova, A. Y.; Bourguet, M.; Hessmann, S.; Chalhoub, S.; Kieffer, B.; Cianférani, S.; Rochel, N., **Molecular determinants of MED1 interaction with the DNA bound VDR-RXR heterodimer.** *Nucleic Acids Research* **2020**, *48* (19), 11199-11213.

Bragantini, B.; Charron, C.; Bourguet, M.; Paul, A.; Tiotiu, D.; Rothé, B.; Marty, H.; Terral, G.; Hessmann, S.; Decourty, L.; Chagot, M.-E.; Strub, J.-M.; Massenet, S.; Bertrand, E.; Quinternet, M.; Saveanu, C.; Cianférani, S.; Labialle, S.; Manival, X.; Charpentier, B., **The box C/D snoRNP assembly factor Bcd1 interacts with the histone chaperone Rtt106 and controls its transcription dependent activity.** *Nature Communications* **2021**, *12* (1), 1859.

Hessmann, S.; Chéry, C.; Sikora, A.-S.; Gervais, A.; Carapito, C., **Improved Mass Spectrometry based Host Cell Proteins Quantification Workflow Using Optimized Standards combined with Data-Independent Acquisition On a Quadrupole-Orbitrap Instrument.** Submitted to Analytical Chemistry (Annexe).

Oral communication

Speaker at the “BioTech Science Forum – UCB Pharma”, 17 septembre 2020,
Hessmann, S.; Chéry, C.; Sikora, A.-S.; Gervais, A.; Carapito, C., **Development of mass spectrometry based methods for the characterization of host cell proteins impurities.**

Annexe

Improved Mass Spectrometry based Host Cell Proteins Quantification Workflow Using Optimized Standards combined with Data-Independent Acquisition On a Quadrupole-Orbitrap Instrument

Steve Hessmann¹, Cyrille Chéry², Anne-Sophie Sikora², Annick Gervais², Christine Carapito¹

¹Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, UMR 7178, Université de Strasbourg, CNRS, 25 rue Becquerel, 67087 Strasbourg, France

²Department of Analytical Development Sciences for Biologicals, UCB, Chemin du Foriest, Braine L'Alleud, Belgium

▪ Abstract

Monitoring of host cell proteins (HCPs) during the manufacture of monoclonal antibodies (mAb) has become a critical requirement to provide effective and safe drug product. ELISA assays are still the current gold standard for the quantification of protein impurities. However, this technique has several limitations and does, among others, not enable the precise identification of proteins. In this context, mass spectrometry (MS) has emerged as an alternative and orthogonal method that delivers qualitative and quantitative information on all identified HCPs. However, in order to be routinely implemented in biopharmaceutical companies, liquid chromatography (LC)-MS based methods still need to be standardized to provide highest sensitivity and robust and accurate quantification. In this study, we developed a promising MS-based analytical workflow coupling the use of an innovative quantification standard, the HCP Profiler solution, with a spectral library-based data-independent acquisition (DIA) method and strict data validation criteria. The performance of the HCP Profiler solution was compared to more conventional standard protein spikes and the DIA approach was benchmarked against classical data-dependent acquisition (DDA) using samples collected at various stages of the manufacturing process. Finally, we further explored the possibility to use spectral library-free DIA. As a result, our method showed accurate and reproducible (coefficients of variation (CVs) < 10%) quantification of HCPs while reaching a sensitivity down to the sub-ng/mg mAb level.

▪ Introduction

For 20 years now, the monoclonal antibody (mAb) market has remarkably grown up with a plethora of approved antibodies by the FDA and EMA with a current sales market of over \$100 billion¹⁻². The high specificity of mAbs to target diverse molecules or antigens and their various modes of action allow them to be used as pharmaceuticals for a wide range of applications³⁻⁴. The high demands of mAbs require the production of well-characterized drug products, both in terms of the mAb sequence and structure itself and its impurities, namely Host Cell Proteins (HCP), remaining from the production process. These impurities are included in the Critical Quality Attributes (CQAs) risk assessment as they can affect the product efficacy due to eventual protease

activities and the patient's safety by inducing immunogenic reactions⁵⁻⁶. Guidelines state classically that the acceptable HCP amount in the final drug product should be below 100 ng/mg mAb⁷⁻⁸. Ultimately, the level of impurities should be as low as possible as issues related to HCPs arise from specific proteins rather than from overall impurities amounts⁹⁻¹¹. Of note is that the HCP profile can be affected by numerous UpStream Process (USP) decisions¹²⁻¹³ (cell culture duration, feeding strategies or culture temperature) or by the production upscale for commercialisation¹⁴, which highlights the need to be able to finely monitor HCPs throughout all steps the manufacturing process. Indeed, specific and sensitive analytical methods allowing reaching 5 to 6 orders of magnitude

dynamics are needed to detect trace level HCPs in the presence of the mAb¹⁵⁻¹⁶. Enzyme-linked immunosorbent assays (ELISA) are commonly used for this purpose as they provide the sensitivity and throughput requested¹⁷⁻¹⁸. However, ELISA has several limitations as it provides a global amount as an output without individual identification of the HCPs present and its coverage is incomplete¹⁸⁻¹⁹. Since immunogenic risk or mAb degradation are related to specific HCPs and not necessarily to their amount, these drawbacks raise an urgent need for alternative methods.

In this context, Mass Spectrometry (MS) became the most promising alternative to monitor HCP allowing risk assessment with individual HCP identification and unbiased quantification. In recent years, Liquid Chromatography – tandem MS (LC-MS/MS) based studies have been conducted. Data-dependent acquisition (DDA) strategies were successfully applied allowing global HCP profiling and reliable individual quantifications down to 10 ng/mg mAb²⁰⁻²³. DDA analysis still suffers from stochasticity, the presence of missing values and a discrimination towards the quantification of most abundant proteins, which are significant issues when the HCP impurities are present at trace levels compared to the biotherapeutic. Targeted strategies (Selected Reaction Monitoring (SRM) or Paralleled Reaction Monitoring (PRM)) have been developed enabling robust and accurate quantification of targeted HCPs down to the sub ng/mg mAb level^{10, 24-25}. Nevertheless, the development of a targeted quantification assay is time consuming, compared to the implementation of a global DDA method, and it is still limited to the selection of about hundred candidates.

Over the past decade, advances in mass spectrometry have highlighted the potential of data-independent acquisition (DIA) on high-resolution/accurate mass (HR/AM) instruments. DIA is based on the co-isolation and co-fragmentation of all ions contained in predefined m/z windows of variable widths to cover the entire mass range. The acquisition of

MS2 signals from all detectable species allows recording complete digital proteome maps while reaching the sensitivity, quantification accuracy and robustness of targeted methods²⁶⁻²⁷. These advantages make DIA approaches attractive in the context of HCP monitoring. The bottleneck of DIA analysis today resides in the extraction of reliable quantitative signals. Each MS2 scan contains the fragment information of all co-isolated precursors, rendering peptide identification difficult. Initially, the use of a specific spectral library generated from DDA runs²⁸ to extract quantitative information from DIA data was shown to be effective. However, the generation of this spectral library requires time and ideally the implementation of fractionation strategies of the studied proteome. The recent development of algorithms that do not require a spectral library will certainly further increase the interest and applicability of DIA strategies, in particular for the monitoring of HCPs²⁹⁻³¹. Unfortunately, despite the advances in data acquisition and extraction, the analytical challenge remains in the dynamic range mAb/HCP. The ubiquity of the antibody may interfere with the reliable extraction of HCP peptides and lead to biases in the MS-based quantification of these impurities. Few studies have attempted to decrease this dynamic range by using multi-dimensional chromatography^{16, 23, 32-34} or mAb depletion³⁵⁻³⁷. In addition, a method to estimate absolute amounts of the HCPs is needed. The Top3 strategy introduced by Silva et al.³⁸ in 2006 has been successfully applied in a few studies using different standard proteins^{22, 35, 39-40}.

In this context, we developed an original MS-based quantitative method to characterize HCP contents in a sample series collected at different purification levels. It combines the use of a robust and accurate quantification using the HCP Profiler standard⁴¹ with the benefit of DIA methods on a fast-scanning Q-Orbitrap instrument.

■ Experimental procedure

Reagents and material

Crude harvest and Post Protein A affinity chromatography (PPA) samples of an IgG4 mAb A33 were obtained from a CHO DG44 cell culture as described in the Supporting Information (Figure S1). A CHO-DG44 mock cell line sample was also provided by UCB Pharma (Braine l'Alleud, Belgium) to generate the spectral library further used for DIA data extraction.

Protein quantification

Protein pellets were resuspended in gel loading buffer (10 mM Tris, 1 mM EDTA, 5% β -Mercaptoethanol, 5% SDS, 10% glycerol, pH 6.8). The total protein concentration was determined using the RC DC Protein Assay kit (Bio-Rad laboratories, Hercules, CA, USA) following manufacturer's instructions.

Sample preparation

CHO-DG44 null cell line sample was fractionated onto 12% acrylamide sodium dodecyl sulfate-polyacrylamide gel electrophoresis SDS-PAGE for spectral library generation. CCCF and PPA samples were stacked in a single band for HCP quantification. Gel bands of the fractionated null cell line sample and stacked bands were cut into small pieces. Proteins were in-gel reduced with 10 mM DTT for 30 min at 60°C. Alkylation was performed with 55 mM IAA for 30 min in the dark. Then trypsin (Promega, Madison, USA) was added to a 1:50 enzyme:substrate ratio (we estimated 1 μ g of proteins in each band of the null cell line fractionation). Samples were incubated overnight at 37°C (14 hours). Peptides were extracted from gel band using 60% ACN, 0.1% FA for 1 hour under agitation and a second step with 100% ACN for 1 hour. After vacuum drying, samples were resolubilized in 2% ACN, 0.1% FA to obtain a final protein concentration of 0.4 μ g/ μ L. In all samples, retention time standards (iRT kit, Biognosys, Schlieren, Switzerland) were spiked. For the HCP Profiler solution (Anaquant, Villeurbanne, France) based quantification, one bead was spiked in 150 μ L of 0.2 ng/ μ L

protein solution. For the mix of standard protein, four accurately quantified standard proteins (on column 10 fmol of ADH (yeast alcohol dehydrogenase, P00330), 2 fmol of PYGM (rabbit phosphorylase b, P00489), 0.5 fmol of BSA (bovine serum albumin, P02769) and 0.2 fmol of ENL (yeast enolase, P00924)) from the MassPREP Digestion Standard Kit (Waters, Milford, USA) were spiked.

nanoLC-MS/MS acquisitions

Data Dependent Acquisition and Data Independent Acquisition were performed on a NanoAcquity UPLC device (Waters) coupled to a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Mobile phase A was 0.1% (v/v) formic acid in water and mobile phase B was 0.1% (v/v) formic acid in acetonitrile. The equivalent of 400ng of proteins was trapped onto a Symmetry C18 precolumn (20 mm \times 180 μ m, 5 μ m diameter particles; Waters) and eluted on an Acquity UPLC BEH130 C18 column (250 mm \times 75 μ m, 1.70 μ m particles; Waters). A 115 minutes chromatographic gradient (2-35% B in 95 min, 35-80% B in 1 min, 80% B for 5 min, 80-2% B in 1 min and maintained 2% B for 13 min) was applied at 400nl/min, with a column temperature set at 60°C. The Q-Exactive HF-X source temperature was set at 250°C and spray voltage to 2 kV. The system was fully controlled by XCalibur software v4.0.27.19, 2013 (Thermo Scientific) and NanoAcquity UPLC console v1.51.3347 (Waters). DIA and DDA injections were performed in a randomized injection sequence.

DDA acquisition

Full scan MS spectra (375-1500 m/z) were acquired in positive mode at a resolution of 120 000 at 200 m/z, a maximum injection time of 60 ms and an AGC target value of 3×10^6 charges. The 10 most intense multiply charged peptides per full scan (charge states ≥ 2) were isolated using a 2 m/z window and fragmented using higher energy collisional dissociation (normalized collision energy set at 27). MS/MS spectra were acquired with a resolution of

15 000 at 200 m/z, a maximum injection time of 60 ms and an AGC target value of 1×10^5 , and dynamic exclusion was set to 40 sec.

DIA acquisition

Full-scan MS spectra were collected from 350–1 250 m/z at a resolution of 60 000 at 200 m/z with an AGC target fixed at 3.10^6 ions and a maximum injection time of 60 ms. Fragments analysis (MS/MS) was subdivided into 40 windows of variable widths. Two acquisitions methods were developed for CCCF and PPA samples (Table S1 and S2). Resolution was set to 30 000 at 200 m/z, AGC target fixed at 1.10^6 ions and with an automatic maximum injection time.

DDA data treatment

Raw DDA files were converted to .mgf peaklists using MsConvert and were submitted to Mascot database searches on a local server (version 2.5.1, MatrixScience, London, UK) against a .fasta database including all *Cricetulus griseus* entries extracted from UniProtKB/TrEMBL (56 566 protein entries, February 15, 2021) together with their reversed sequences, as well as the iRT retention time standards (Biognosys), the four standard proteins of the MassPREP Digestion Standard Kit (Waters), HCP Profiler kit proteins, the mAb heavy and light chains and common contaminants. Spectra were searched with a mass tolerance of 5 ppm in MS mode and 0.05 Da in MS/MS mode. One trypsin missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. Identification results were imported into Proline software version 1.6 (<http://proline.profiroteomics.fr/>; Bouyssie *et al.*, *Bioinformatics* 2020⁴²) for validation. A false discovery rate of 1% was set at peptide levels using Adjusted e-Value and protein level using Mascot Modified Mudpit scores. Peptides abundances were extracted with Proline software using an extraction m/z

tolerance and PSM/Peake1 matching m/z tolerance of 5 ppm. Alignment of the LC-MS runs was performed using Loess smoothing, Peptide Identity method and with a time tolerance of 300 seconds. Cross assignment of peptide ions abundances was performed among PPA or harvest samples using a m/z tolerance of 5 ppm and a retention time tolerance of 40 seconds.

Cricetulus griseus, CHO reference spectral library generation

A reference spectral library combining a series of analyses conducted on different samples in DDA mode was generated using the Spectronaut and Pulsar algorithm (v.14.5; Biognosys). This series of analyses comprised:

- 24 gel bands obtained by SDS PAGE fractionation of the CHO DG44 null cell line sample.
- All DDA analyses of current CCCF and PPA samples., including iRT retention time standards (Biognosys) and the 18 proteins from HCP Profiler kit.

Raw DDA files were uploaded into Spectronaut and searched with the Pulsar algorithm against a fasta database containing all *Cricetulus griseus* entries extracted from UniProtKB/TrEMBL (56 687 protein entries, 2021/02/15), as well as the iRT retention time standards (Biognosys), the 18 proteins from HCP Profiler kit, the reference sequence of the mAb and common contaminants).

Trypsin/P enzyme was used and one missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. MS and MS/MS mass tolerances were set in dynamic mode. The spectral library was validated as follows: a False Discovery Rate (FDR) of 0.01 was set at Peptide Spectrum Matches (PSM), Peptides and Proteins levels. Fragment ions window was set between 300 and 1800 m/z with a minimum of 4 fragments per precursor and up to 6.

DIA data treatment

DIA data was analysed with a peptide-centric approach using the Spectronaut algorithm and the upper described in house generated spectral library (v.14.5; Biognosys). Trypsin/P was used as digestion enzyme with one missed cleavage allowed. Carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues and acetylation of proteins n-termini were set as variable modifications. For quantitative data extraction, MS and MS/MS mass tolerances, Extracted Ion Chromatogram (XIC) and retention time windows were all set as dynamic. iRT regression type was set to local (non-linear) regression. A false discovery rate of 1% was set at precursors and proteins levels. At this extraction stage, a sparse q-value filter was applied. Peptide quantities corresponding to the sum of a minimum of 4 fragments XIC areas and up to 6 fragments (interference correction parameter was turned on) were calculated. Precursors with a Qvalue below 0.01 were used for IRT profiling.

HCP TOP3 quantification

After data extraction, a list of identified peptides with their corresponding intensities is exported in excel format for both DDA and DIA data. Prior the TOP3 quantification, filters are applied to remove oxidized and acetylated peptides alongside with their counterparts. Precursors inferred to host organism proteins, standard proteins and precursors with charge states 2 and 3 are kept. For DDA data, a maximum of one precursor validated by cross-assignment was allowed. Precursors with more than one Q-value > 0.01 or profiled were removed for DIA data. For both acquisition methods, quantity estimation was performed using precursors intensities showing a coefficient of variation below 20% within technical triplicates. Finally, HCP peptides showing 100% sequence identity with a semi-trypsinic or non-trypsinic mAb peptide were removed. After applying these stringent validation filters in DDA and DIA modes, peptide intensities were obtained by summing

all precursor intensities and protein intensities by summing the three most intense peptides intensities. For HCP Profiler kit, a calibration curve of the $\log_2(\text{TOP3 standard peptides abundance})$ in function of $\log_2(\text{standard proteins quantity})$ is obtained and allowed us to estimate protein mol quantities. For the four standard proteins mix, the universal signal response factor⁴³ (MS signal/mol of protein) was calculated using PYGM as a reference, and allowed estimating protein mol quantities. Finally, for both quantification method, using protein molecular weight and injected mAb quantities, individual HCP ng/mg mAb amounts were estimated.

■ Results and discussion

A multi-stage MS-based HCP profiling workflow optimization

A sample set including two levels of HCP complexity was produced to investigate bioprocess developments by MS. Two cell culture durations (7 and 10 days), three harvest procedures (no shear, low shear or high shear) and two protein A purification protocols were investigated resulting in four HCP-rich Clarified Cell Culture Fluids (CCCF) and seven purified Post Protein A (PPA) fractions (Figure S1). The work done by Husson et al.⁴⁴ on those samples led to the development of the Top3-ID-DIA method on a Q-TOF instrument. The combination of a Top3 quantification strategy for global HCP profiling and isotopic dilution (ID) to accurately quantify key HCPs, has proven to be a method of choice to obtain a complete picture of the HCP content. In order to continue the development of a robust, accurate and sensitive DIA method, we used the previously mentioned sample series to support the optimizations conducted at different levels of the general workflow presented in Figure 1.

First, two accurate MS-based quantification methods were benchmarked. On the one hand, samples were spiked with an original mixture of peptides coated on a water-soluble bead releasing controlled amounts of a peptide range after solubilization (READYBEADS

technology, from Anaquant)⁴¹. On the other hand, samples were spiked with a mixture of four standard proteins PYGM, ADH, BSA and ENL, from Waters, more commonly applied to derive HCP amounts with MS methods^{39, 45}. Second, Data Independent Acquisition (DIA) methods were finely tuned on a Quadruple-Orbitrap instrument to thoroughly compare the performances achievable with DIA

methods against more classical Data Dependent Acquisitions (DDA) approaches on the same instrument. Then, two DIA data extraction and interpretation strategies were evaluated, a peptide-centric approach requiring the acquisition of a spectral library and a spectrum-centric approach without spectral library.

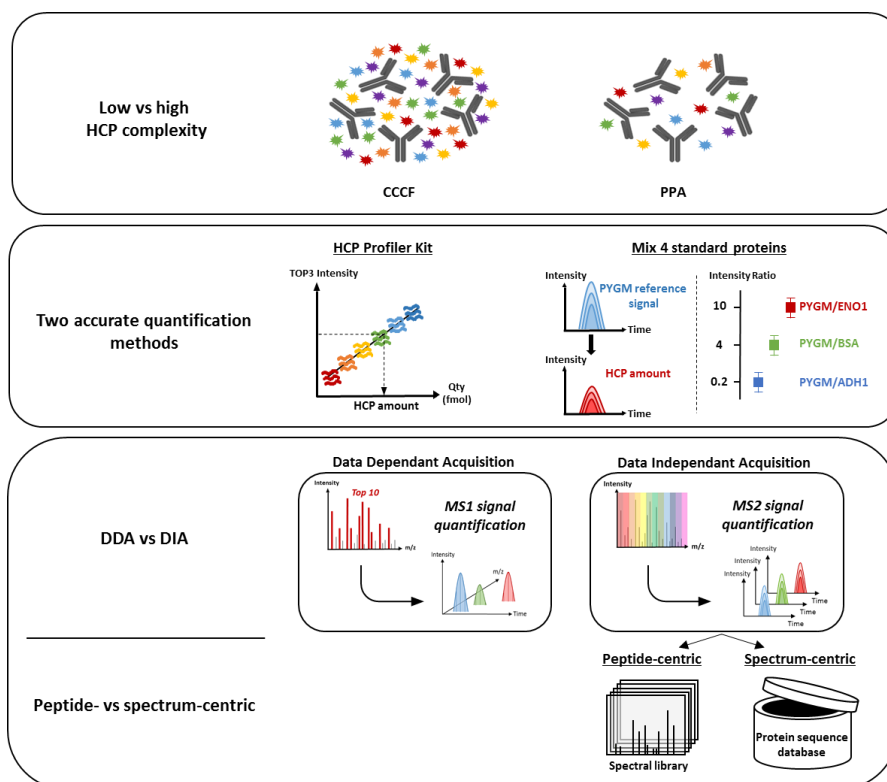


Figure 1: Experimental design for the optimization of an MS-based quantification strategy for HCP monitoring.

Implementation of the HCP Profiler standard for quantification

Top-3 quantification methods based on the assumption that the sum of the MS response of the three best responding peptides per mole of protein is constant within a coefficient of variation of less than 10% were used to estimate absolute amounts of HCP³⁸. Therefore, an internal standard must be used to calculate a signal response factor (TOP3 peptides signal/mol) and to perform the HCP quantity estimation (HCP TOP3 peptides signal/signal response factor). The use of a mixture of four standard proteins has been previously reported for this purpose^{39, 45-47},

considering the PYGM protein as a reference and the three other proteins (ADH, BSA, ENL) to calculate ratios, as internal controls. However, the development of more adapted and dedicated standards is a valuable challenge. This is why we choose to confront our current quantification workflow against the HCP Profiler standard⁴¹. This new standard is composed of a water-soluble bead, which releases unlabeled peptides at known amounts. Eighteen tripeptides distributed over six concentration points ranging from 1 to 500 fmol, thus a total 54 peptides, are coated on a bead. Those 54 peptides areas extraction thus allows building a calibration curve that is

further used to derive each individual HCP amount.

The robustness and reproducibility of the HCP Profiler standard were assessed by coefficients of variations (CVs) calculated on the slopes, intercepts and R^2 of the 11 calibration curves obtained on the 11 samples, all below 3% (Table S3). On average, 1464 HCPs were quantified in CCCF fractions with global quantities between 222 646 and 365 145 ng/mg mAb, and 115 HCPs in PPA fractions representing 569 to 19 153 ng/mg mAb. These results are coherent with the global amounts obtained using the signal response factor of PYGM (Figure 2). Of note is that for almost all samples, the global HCP amounts are higher using the HCP Profiler standard, except for PPA 5 and Harvest 1 samples. However, a lower number of HCPs was quantified with an

average of 34% less identified HCPs for PPA samples and 13% for CCCF fractions. In order to understand this observation, we compared the individual amounts obtained for all HCPs quantified using both methods (Figure S2). The ratios between both methods were consistent as 78% of the 5305 ratios span within a factor 2. In addition, it is of note that the median value of all ratios is at 1.3 in favor of higher amounts estimated with the water-soluble bead. Since individual peptide's ionization efficiencies and response factors vary, taking into account the MS response of 54 peptides spiked at different amounts rather than only of 3 peptides from a single protein leads to a more reliable amount estimation. Thus the sum of individually increased quantities ultimately results in higher global HCP amounts even when less HCPs are identified.

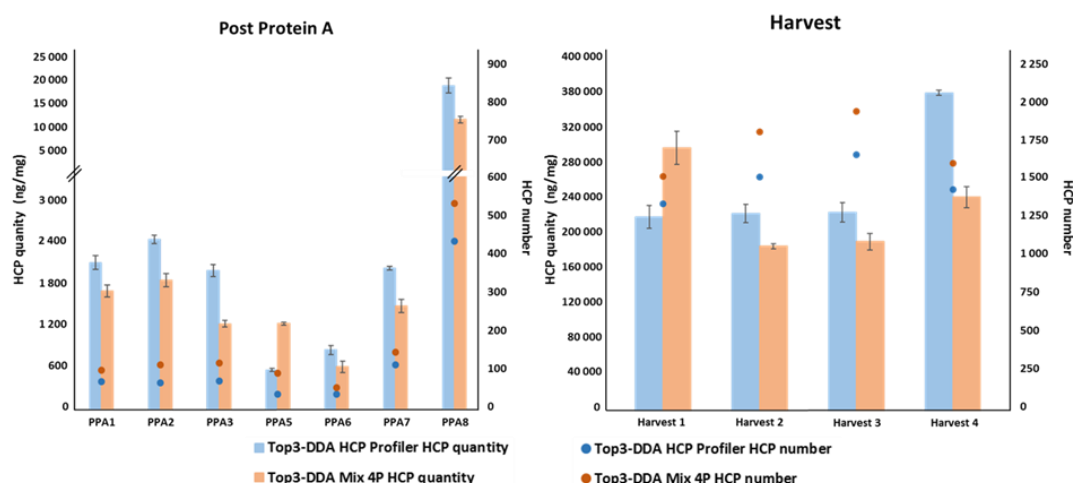


Figure 2: Benchmarking of the HCP Profiler standard against the Mix 4P using a MS1-XIC DDA approach. Comparison of global HCPs numbers and amounts obtained on a Q-Exactive HF-X using the HCP Profiler standard and the Mix 4P. Bar height represents the mean of the global HCP amount in injection triplicates. Error bars represent the standard deviation. Dots indicate the numbers of quantified HCPs.

As a conclusion, we were able to observe the quantification accuracy and robustness of the HCP Profiler standard using its multiple peptides at different concentration levels. Those peptides ensure confidence and quality control of the experiment. The high reproducibility of the calibration curves

independently of the sample complexity will allow a monitoring at the different steps of the manufacturing process of the mAb. Due to its easy and ready-to-use characteristics, the standard will allow avoiding user-induced analytical biases that may occur while preparing the mixture of proteins at known

amounts. With ID-SRM as the gold standard method to derive absolute quantities of problematic HCPs^{10-11, 33}, the use of the HCP Profiler standard to estimate absolute amounts for all other HCPs, in parallel or as an alternative, is a great compromise to get the best overview of the overall HCP content. Thus, considering the advantages of the HCP Profiler bead, we continued optimizations using this standard.

Benchmarking of DIA versus DDA methods

To dig deeper into the global profiling and individual quantification of the HCPs, we developed a dedicated MS2-based DIA approach and evaluated its performances against DDA data acquired in parallel on the same samples and instrument.

In order to get full advantages of the peptide-centric DIA approach, we first generated the most comprehensive spectral library from a fractionated CHO DG44 null cell line sample combined with DDA runs acquired on all CCCF and PPA fractions. The combination of these DDA data allowed us to generate a project-specific spectral library containing 40 281 peptides from 3 978 protein groups. Then, a DIA variable isolation windows method was developed for both types of samples based on the distribution of precursors over the *m/z* acquisition range observed in DDA, the MS and MS/MS scan times of the instrument and the theoretical number of MS cycles per chromatographic peak. Two methods composed of 40 variable isolation windows, one dedicated to HCP-rich harvest samples and a second one dedicated to PPA samples, were developed allowing to obtain a median of 5 to 8 MS cycles per chromatographic peak (Figure S3).

Following data acquisition and peptides signals extractions, we applied stringent filters as validation criteria. The first filter acts as a signal quality filter: precursors with more than one Q-value > 0.01 and/or profiled were removed from DIA results and those with more than one cross-assigned attribution were removed from DDA data. The second quality filter refers to the

reproducibility of the signals as precursors with CVs above 20% were excluded. Then, a sequence homology filter was applied using the BLASTP⁴⁸ (v.2. 10.0+) algorithm run against the mAb heavy and light chain sequences. HCP peptides showing 100% sequence identity with a semi-trypsin or non-trypsin mAb peptide were removed. This last filter does not have a major impact but it is important due to the large excess of the mAb compared to trace-level HCPs. Indeed, a mAb peptide, resulting from an unspecific trypsin cleavage, wrongly attributed to an HCP sequence, would lead to a significant overestimation of this HCP amount and even of the overall HCP amount. Finally, the selection of the 3 most intense peptides was performed and absolute HCP amounts were estimated using the HCP Profiler standard.

An average of 1737 HCPs with a global estimation between 62 792 and 138 297 ng/mg mAb for CCCF samples and 221 HCPs in PPA fractions with a quantity between 1 339 and 11 992 ng/mg mAb were obtained (Figure 3). At first sight, we observe a significant benefit of DIA on the HCP coverage, as around 19% more HCPs were quantified for CCCF and 111% more for PPA samples when compared to DDA. The Venn diagrams in Figure S4 show a great correlation between DIA and DDA lists of quantified HCPs. For PPA samples, the overall estimated HCP amounts were consistent with the increased number of HCPs quantified. On the contrary, even with a higher number of quantified HCPs, the CCCF fractions showed lower overall amounts when compared to DDA results. This observation shows the potential presence of false positives or interferences during MS1 signals extraction eventually resulting in an overestimation of the HCP amounts. While MS1-XIC DDA was able to achieve 2.4 to 5.5 orders of magnitude within the least and most abundant HCP, MS2-based DIA allowed to reach a dynamic between 3.5 and 6.1 (Table S4). The precision of the DIA data extraction was assessed by the CVs calculated on peptides intensities within injection replicates for all HCPs quantified. A median of 9.0% was obtained for DIA

compared to 9.8% for DDA results (Figure S5). When comparing the ratios of the quantities obtained by DIA over DDA for the 3779 HCPs common to both methods, we notice that only 52% of the ratios are within a factor 2 (Figure S6.A). This is likely due to the number of peptides used to quantify each protein (Figure S6.B). A larger number of peptides was used in DIA compared to DDA, 8034 and 7836 peptides, respectively. Among the 3779 proteins compared, 1735 HCPs were quantified with a Top3 in DIA while this number drops to 1648 in DDA. As we sum the areas of the Top3 peptides, the increased number of peptides per HCP obtained in DIA has a direct impact on the HCPs quantification.

Overall, our results demonstrate a high reproducibility and accurate quantification using a DIA method and a sensitivity down to

sub-ng/mg mAb. The DIA acquisition mode has shown its ability to extract signals close to the background noise. This is a significant benefit when the impurities of interest are present at trace levels. Compared to the gold standard ELISA assay, our HCP Profiler-DIA method shows global quantities higher by a factor 4 on average for PPA samples except for PPA 8 (obtained with a modified protocol) in which it raised to a factor 32 (Figure S7). These results clearly highlight the previously mentioned limitations of ELISA assays. Therefore, the increased HCP map coverage alongside with the accurate and reproducible quantification make our HCP-Profiler DIA method valuable to achieve an overview of the HCP content of samples from various steps of the mAb manufacturing process.

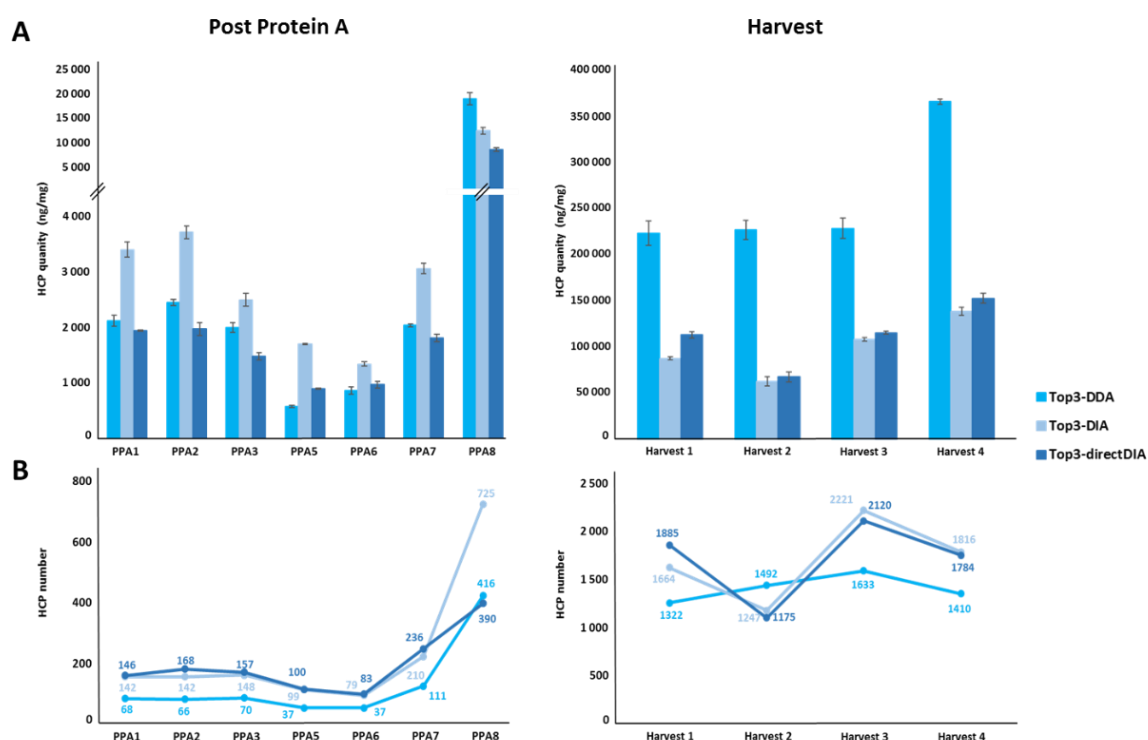


Figure 3: Evaluation of DIA strategies against DDA for global HCP profiling using the HCP Profiler standard. Quantification results obtained for PPA and CCCF fractions (A). Bar height represents the mean of the global HCP amount in injection triplicates. Error bars represent the standard deviation. Number of HCPs quantified per samples for PPA and CCCF fractions (B).

Implementation of a spectral library-free DIA approach

The generation of a spectral library requests sample preparation and instrument acquisition time. Even though this takes less time than the development of a specific ELISA kit, a new dedicated library must be generated for each drug product. In addition, HCPs not covered by the library will not be searched for even if they are present in detectable amounts. For all those reasons, alternative approaches for DIA data extraction have been recently introduced avoiding the use of a spectral library²⁹⁻³¹. With the aim to make our DIA method even more straightforward, we evaluated the spectrum-centric approach called directDIA in Spectronaut (Biognosys AG). This approach relies on the generation of “pseudo MS/MS spectra” from the DIA runs which will be used to query a search against the organism database as it is done in classical DDA search engines.

Comparable reproducibility of the signal extraction was achieved as CV values calculated on the slopes, intercepts and R^2 of the 33 calibration curves were all below 2% (Table S3). The results were also very consistent regarding HCPs contents as comparable quantification results and close numbers of quantified HCPs were obtained (Figure 3). Equivalent sensitivity down to the sub-ng/mg mAb level was achieved as with the spectral library-based data extraction (Table S4). Both extraction strategies were able to achieve 3 to 6 orders of magnitude between the least and most abundant HCP. Looking further at individual HCP amounts, we compared the quantities obtained using DIA over directDIA for the 5833 HCPs commonly detected with both methods. A median of 0.96 was obtained and the amounts estimated were in accordance for 82% of the common HCPs within a factor 2 (Figure S8.A). Similarly, close numbers of peptides were used to quantify those common HCPs, respectively 11 436 and 11 090 for DIA and directDIA (Figure S8.B). These observations showed confidence towards the spectral library free approach.

Nevertheless, the lower numbers of HCPs quantified using directDIA highlight the difficulties that still exist to extract low-level impurities using directDIA.

Having a close look to the individual HCPs quantified by both DIA strategies, Venn diagrams show great correlation of both lists (Figure S4). On average, 62% of the HCPs quantified in DIA were extracted in directDIA as well. On the one hand, about 38% of the HCPs quantified using the spectral library could not be extracted by directDIA. This percentage represents hundreds of unidentified HCPs highlighting the room of improvement that still exists for spectrum-centric search algorithms. On the other hand, hundreds of HCPs were missed using the spectral library-based approach. On average 260 HCPs for each CCCF sample were not identified in the spectral library (Figure S9). These HCPs cover about 5 orders of magnitude between the least and most abundant HCP with a maximum reaching thousands of ng/mg mAb (Table S5). The number and quantities of these HCPs are not negligible and thus also bring to light the limitations of the spectral library extraction approach as HCPs not present in the library will not be identified and thus further quantified. However, spectral library free results should still be taken with caution, as one could seriously argue that HCPs with an estimated amount around thousands of ng/mg mAb should be identified in the spectral library with at least a few peptides. The number of false positives must be significant in directDIA although difficult to estimate yet. Compared to the targeted DIA analysis, the spectral library free-based approach relies on the use of the CHO protein sequence database, that is yet poorly annotated and curated (56 565 entries in UniProtKB/TrEMBL). As a result, the high redundancy significantly hinders the extraction of specific peptides. Overall, the directDIA approach supported the increased HCPs coverage enabled by the DIA acquisition in line with previous studies^{29-30, 49}.

To conclude, as previously shown using a spectral library-based strategy, the directDIA

approach highlights benefits of DIA acquisition. Our developed method showed high sensitivity as well as a robust and accurate quantification that will allow the monitoring of HCPs in samples with different purity levels. The directDIA method displayed a significant gain with the quantification of HCPs not identified in the spectral library. Unfortunately, the current limitations of the spectral library-free approach make it difficult to implement it in a regulated environment working with CHO cell-based bioproducts.

■ Conclusion

We have demonstrated that the HCP Profiler standard offers a ready-to-use solution for HCP quantification with its internal 6 points calibration curves representing a move towards the standardization needed for the implementation of MS-based methods in a biopharmaceutical environment. Initially applied with MS1-XIC DDA analysis, we prove here that the standard undergoes no loss of performance when combined with a DIA method on a Q-Orbitrap instrument. Our study also demonstrates again the advantages of MS-methods over ELISA assays. On average, our MS-based approach identifies each HCP with 5 peptides containing 7 to 28 amino acids, which could be considered the equivalent of 5 to 14 epitopes. Therefore, a finer granularity of results is obviously achieved with the main difference being the detection principle. MS allows to avoid the inherent gaps in the ELISA tests such as the absence of specific polyclonal antibodies (pAbs) or impaired binding due to loss of conformational epitopes⁵⁰. Similarly, we could argue on the gap related to the spectral library-based DIA approach, which is restricted to identifying HCPs that are present in the library. However, the development of a new spectral library takes less time than the months requested to generate a specific ELISA kit. In addition, anti-HCP antibodies are perishable and have to be reproduced by a new immunization campaign whenever needed. By contrast, within a week, it is possible to generate a comprehensive spectral library

specific to the mAb produced, i.e., a library generated from the analysis of a mock cell line and/or HCP-rich harvest samples. This library can be updated endlessly with new analyses in case of manufacturing process changes suspected to lead to the presence of new HCPs. Thus, once generated, the spectral library can be unlimitedly used to extract signals from DIA data. Altogether, the developed MS-based method presents a sufficient robustness to consider its implementation within a biopharmaceutical environment to support process development or batch-to-batch consistency. In a short-term perspective, the directDIA approach that is straightforward and non dependent on any prior information, will likely become the best suited way to support the release of safer bioproducts.

■ Acknowledgments

This project was supported by the Association Nationale de la Recherche et de la Technologie and UCB Pharma via the CIFRE fellowship of Steve Hessmann. This work was supported by the “Agence Nationale de la Recherche” and the French Proteomic Infrastructure (ProFI FR2048; ANR-10-INBS-08-03). The authors thank Laura Herment and Tanguy Fortin from Anaquant for their support in the use of the HCP Profiler standard.

■ Reference

1. Grilo, A. L.; Mantalaris, A., The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends in Biotechnology* **2019**, *37* (1), 9-16.
2. Kaplon, H.; Reichert, J. M., Antibodies to watch in 2021. *MAbs* **2021**, *13* (1), 1860476.
3. Reichert, J. M.; Valge-Archer, V. E., Development trends for monoclonal antibody cancer therapeutics. *Nature reviews. Drug discovery* **2007**, *6* (5), 349-56.
4. Suzuki, M.; Kato, C.; Kato, A., Therapeutic antibodies: their mechanisms of action and the pathological findings they induce in toxicity studies. *J Toxicol Pathol* **2015**, *28* (3), 133-139.
5. Cui, T.; Chi, B.; Heidbrink Thompson, J.; Kasali, T.; Sellick, C.; Turner, R., Cathepsin D: Removal strategy on protein A chromatography, near real time monitoring and characterisation

during monoclonal antibody production. *Journal of Biotechnology* **2019**, *305*, 51-60.

6. Yang, B.; Li, W.; Zhao, H.; Wang, A.; Lei, Y.; Xie, Q.; Xiong, S., Discovery and characterization of CHO host cell protease-induced fragmentation of a recombinant monoclonal antibody during production process development. *Journal of Chromatography B* **2019**, *1112*, 1-10.

7. Hogwood, C. E.; Bracewell, D. G.; Smales, C. M., Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. *Current opinion in biotechnology* **2014**, *30*, 153-60.

8. Bracewell, D. G.; Francis, R.; Smales, C. M., The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnology and bioengineering* **2015**, *112* (9), 1727-37.

9. Rane, S. S.; Dearman, R. J.; Kimber, I.; Uddin, S.; Bishop, S.; Shah, M.; Podmore, A.; Pluen, A.; Derrick, J. P., Impact of a Heat Shock Protein Impurity on the Immunogenicity of Biotherapeutic Monoclonal Antibodies. *Pharmaceutical research* **2019**, *36* (4), 51.

10. Gao, X.; Rawal, B.; Wang, Y.; Li, X.; Wylie, D.; Liu, Y.-H.; Breunig, L.; Driscoll, D.; Wang, F.; Richardson, D. D., Targeted Host Cell Protein Quantification by LC-MRM Enables Biologics Processing and Product Characterization. *Analytical Chemistry* **2020**, *92* (1), 1007-1015.

11. Li, X.; An, Y.; Liao, J.; Xiao, L.; Swanson, M.; Martinez-Fonts, K.; Pavon, J. A.; Sherer, E. C.; Jawa, V.; Wang, F.; Gao, X.; Letarte, S.; Richardson, D. D., Identification and characterization of a residual host cell protein hexosaminidase B associated with N-glycan degradation during the stability study of a therapeutic recombinant monoclonal antibody product. *Biotechnology progress* **2021**, e3128.

12. Falkenberg, H.; Waldera-Lupa, D. M.; Vanderlaan, M.; Schwab, T.; Krapfenbauer, K.; Studts, J. M.; Flad, T.; Waerner, T., Mass spectrometric evaluation of upstream and downstream process influences on host cell protein patterns in biopharmaceutical products. *Biotechnology progress* **2019**, *35* (3), e2788.

13. Hogwood, C. E.; Bracewell, D. G.; Smales, C. M., Host cell protein dynamics in recombinant CHO cells: impacts from harvest to purification and beyond. *Bioengineered* **2013**, *4* (5), 288-91.

14. Chahar, D. S.; Ravindran, S.; Pisal, S. S., Monoclonal antibody purification and its progression to commercial scale. *Biologicals : journal of the International Association of Biological Standardization* **2020**, *63*, 1-13.

15. Valente, K. N.; Levy, N. E.; Lee, K. H.; Lenhoff, A. M., Applications of proteomic methods for CHO host cell protein characterization in

biopharmaceutical manufacturing. *Current opinion in biotechnology* **2018**, *53*, 144-150.

16. Zhang, Q.; Goetze, A. M.; Cui, H.; Wylie, J.; Trimble, S.; Hewig, A.; Flynn, G. C., Comprehensive tracking of host cell proteins during monoclonal antibody purifications using mass spectrometry. *MAbs* **2014**, *6* (3), 659-70.

17. Tscheliessnig, A. L.; Konrath, J.; Bates, R.; Jungbauer, A., Host cell protein analysis in therapeutic protein bioprocessing - methods and applications. *Biotechnology journal* **2013**, *8* (6), 655-70.

18. Zhu-Shimoni, J.; Yu, C.; Nishihara, J.; Wong, R. M.; Gunawan, F.; Lin, M.; Krawitz, D.; Liu, P.; Sandoval, W.; Vanderlaan, M., Host cell protein testing by ELISAs and the use of orthogonal methods. *Biotechnology and bioengineering* **2014**, *111* (12), 2367-79.

19. Vanderlaan, M.; Zhu-Shimoni, J.; Lin, S.; Gunawan, F.; Waerner, T.; Van Cott, K. E., Experience with host cell protein impurities in biopharmaceuticals. *Biotechnology progress* **2018**, *34* (4), 828-837.

20. Bomans, K.; Lang, A.; Roedel, V.; Adolf, L.; Kyrioglou, K.; Diepold, K.; Eberl, G.; Mølhøj, M.; Strauss, U.; Schmalz, C.; Vogel, R.; Reusch, D.; Wegele, H.; Wiedmann, M.; Bulau, P., Identification and monitoring of host cell proteins by mass spectrometry combined with high performance immunochemistry testing. *PloS one* **2013**, *8* (11), e81639-e81639.

21. Madsen, J. A.; Farutin, V.; Carbeau, T.; Wudyka, S.; Yin, Y.; Smith, S.; Anderson, J.; Capila, I., Toward the complete characterization of host cell proteins in biotherapeutics via affinity depletions, LC-MS/MS, and multivariate analysis. *MAbs* **2015**, *7* (6), 1128-37.

22. Walker, D. E.; Yang, F.; Carver, J.; Joe, K.; Michels, D. A.; Yu, X. C., A modular and adaptive mass spectrometry-based platform for support of bioprocess development toward optimal host cell protein clearance. *mAbs* **2017**, *9* (4), 654-663.

23. Yang, F.; Walker, D. E.; Schoenfelder, J.; Carver, J.; Zhang, A.; Li, D.; Harris, R.; Stults, J. T.; Yu, X. C.; Michels, D. A., A 2D LC-MS/MS Strategy for Reliable Detection of 10-ppm Level Residual Host Cell Proteins in Therapeutic Antibodies. *Analytical Chemistry* **2018**, *90* (22), 13365-13372.

24. Picotti, P.; Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **2012**, *9* (6), 555-566.

25. Lavoie, R. A.; di Fazio, A.; Williams, T. I.; Carbonell, R.; Menegatti, S., Targeted capture of Chinese hamster ovary host cell proteins: Peptide ligand binding by proteomic analysis.

- Biotechnology and bioengineering* **2020**, *117* (2), 438-452.
26. Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & cellular proteomics : MCP* **2012**, *11* (6), O111.016717.
 27. Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Treiber, T.; Bruderer, R.; Reiter, L., Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Molecular omics* **2019**, *15* (5), 348-360.
 28. Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R., Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol* **2018**, *14* (8), e8126.
 29. Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T., Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020**, *20* (17-18), 1900276.
 30. Ting, Y. S.; Egertson, J. D.; Bollinger, J. G.; Searle, B. C.; Payne, S. H.; Noble, W. S.; MacCoss, M. J., PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nat Methods* **2017**, *14* (9), 903-908.
 31. Tsou, C. C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A. C.; Nesvizhskii, A. I., DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* **2015**, *12* (3), 258-64, 7 p following 264.
 32. Jawa, V.; Joubert, M. K.; Zhang, Q.; Deshpande, M.; Hapuarachchi, S.; Hall, M. P.; Flynn, G. C., Evaluating Immunogenicity Risk Due to Host Cell Protein Impurities in Antibody-Based Biotherapeutics. *The AAPS journal* **2016**, *18* (6), 1439-1452.
 33. Chen, I. H.; Xiao, H.; Li, N., Improved host cell protein analysis in monoclonal antibody products through ProteoMiner. *Analytical Biochemistry* **2020**, *610*, 113972.
 34. Wang, Q.; Slaney, T. R.; Wu, W.; Ludwig, R.; Tao, L.; Leone, A., Enhancing Host-Cell Protein Detection in Protein Therapeutics Using HILIC Enrichment and Proteomic Analysis. *Analytical Chemistry* **2020**, *92* (15), 10327-10335.
 35. Huang, L.; Wang, N.; Mitchell, C. E.; Brownlee, T.; Maple, S. R.; De Felippis, M. R., A Novel Sample Preparation for Shotgun Proteomics Characterization of HCPs in Antibodies. *Anal Chem* **2017**, *89* (10), 5436-5444.
 36. Johnson, R. O. B.; Greer, T.; Cejckov, M.; Zheng, X.; Li, N., Combination of FAIMS, Protein A Depletion, and Native Digest Conditions Enables Deep Proteomic Profiling of Host Cell Proteins in Monoclonal Antibodies. *Analytical Chemistry* **2020**, *92* (15), 10478-10484.
 37. Nie, S.; Greer, T.; O'Brien Johnson, R.; Zheng, X.; Torri, A.; Li, N., Simple and Sensitive Method for Deep Profiling of Host Cell Proteins in Therapeutic Antibodies by Combining Ultra-Low Trypsin Concentration Digestion, Long Chromatographic Gradients, and BoxCar Mass Spectrometry Acquisition. *Analytical Chemistry* **2021**, *93* (10), 4383-4390.
 38. Silva, J. C.; Gorenstein, M. V.; Li, G.-Z.; Vissers, J. P. C.; Geromanos, S. J., Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms Acquisition *S. *Molecular & Cellular Proteomics* **2006**, *5* (1), 144-156.
 39. Doneanu, C. E.; Anderson, M.; Williams, B. J.; Lauber, M. A.; Chakraborty, A.; Chen, W., Enhanced Detection of Low-Abundance Host Cell Protein Impurities in High-Purity Monoclonal Antibodies Down to 1 ppm Using Ion Mobility Mass Spectrometry Coupled with Multidimensional Liquid Chromatography. *Anal Chem* **2015**, *87* (20), 10283-91.
 40. Reiter, K.; Suzuki, M.; Olano, L. R.; Narum, D. L., Host cell protein quantification of an optimized purification method by mass spectrometry. *Journal of pharmaceutical and biomedical analysis* **2019**, *174*, 650-654.
 41. Trauchessec, M.; Hesse, A. M.; Kraut, A.; Berard, Y.; Herment, L.; Fortin, T.; Bruley, C.; Ferro, M.; Manin, C., An innovative standard for LC-MS-based HCP profiling and accurate quantity assessment: Application to batch consistency in viral vaccine samples. *Proteomics* **2021**, *21* (5), e2000152.
 42. Bouyssie, D.; Hesse, A. M.; Mouton-Barbosa, E.; Rompais, M.; Macron, C.; Carapito, C.; Gonzalez de Peredo, A.; Coute, Y.; Dupierris, V.; Burel, A.; Menetrey, J. P.; Kalaitzakis, A.; Poisat, J.; Romdhani, A.; Burlet-Schiltz, O.; Cianferani, S.; Garin, J.; Bruley, C., Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **2020**, *36* (10), 3148-3155.
 43. Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **2006**, *5* (1), 144-56.
 44. Husson, G.; Delangle, A.; O'Hara, J.; Cianferani, S.; Gervais, A.; Van Dorsselaer, A.; Bracewell, D.; Carapito, C., Dual Data-Independent Acquisition Approach Combining Global HCP Profiling and Absolute Quantification of Key

Impurities during Bioprocess Development. *Anal Chem* **2018**, 90 (2), 1241-1247.

45. Doneanu, C. E.; Xenopoulos, A.; Fadgen, K.; Murphy, J.; Skilton, S. J.; Prentice, H.; Stapels, M.; Chen, W., Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *MAbs* **2012**, 4 (1), 24-44.

46. Zhang, Q.; Goetze, A. M.; Cui, H.; Wylie, J.; Tillotson, B.; Hewig, A.; Hall, M. P.; Flynn, G. C., Characterization of the co-elution of host cell proteins with monoclonal antibodies during protein A purification. *Biotechnology progress* **2016**, 32 (3), 708-17.

47. Li, D.; Farchone, A.; Zhu, Q.; Macchi, F.; Walker, D. E.; Michels, D. A.; Yang, F., Fast, Robust, and Sensitive Identification of Residual Host Cell Proteins in Recombinant Monoclonal Antibodies Using Sodium Deoxycholate Assisted Digestion. *Analytical Chemistry* **2020**, 92 (17), 11888-11894.

48. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L., BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, 10, 421.

49. Mehta, D.; Scandola, S.; Uhrig, R. G., Direct data-independent acquisition (direct DIA) enables substantially improved label-free quantitative proteomics in Arabidopsis. *bioRxiv* **2020**, 2020.11.07.372276.

50. Seisenberger, C.; Graf, T.; Haindl, M.; Wegele, H.; Wiedmann, M.; Wohlrab, S., Questioning coverage values determined by 2D western blots: A critical study on the characterization of anti-HCP ELISA reagents. *Biotechnology and bioengineering* **2021**, 118 (3), 1116-1126.

Supporting Information

Improved Mass Spectrometry based Host Cell Proteins Quantification Workflow Using Optimized Standards combined with Data-Independent Acquisition On a Quadrupole-Orbitrap Instrument

Steve Hessmann¹, Anne-Sophie Sikora², Cyrille Chéry², Annick Gervais², Christine Carapito¹

¹ Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, UMR 7178, Université de Strasbourg, CNRS, 25 rue Becquerel, 67087 Strasbourg, France

² Department of Analytical Development Sciences for Biologicals, UCB, Chemin du Foriest, Braine L'Alleud, Belgium

Table of Contents

Supplementary Table S1: DIA variable windows acquisition method for Post Protein A samples.

Supplementary Table S2: DIA variable windows acquisition method for harvest samples.

Supplementary Table S3: DDA, DIA and directDIA calibration curves information. The DIA method represents the spectral library-based method and directDIA corresponds to the spectral library free-based method developed in Spectronaut (Biognosys AG).

Supplementary Table S4: Order of magnitude between the least and the most abundant HCP with a Top3 covered by DDA, DIA and directDIA strategies.

Supplementary Table S5: Order of magnitude covered by the HCPs only identified using the spectral library free-based approach.

Supplementary Figure S1: mAb samples production workflow. Four IgG4 A33 mAb samples were collected after 7 or 10 days of CHO DG44 cell culture duration followed by different shear stress conditions during harvest using an Ultra Scale-Down (USD) shear device. Finally, two protein A purification protocols were applied, a standard and a modified protocol, resulting in seven Post Protein A (PPA) fractions.

Supplementary Figure S2: Comparison of the individual HCP amounts obtained using the HCP Profiler versus the Mix 4P strategies. Ratios distributions calculated between the HCP Profiler versus the Mix 4P amounts on all HCPs identified with both methods. 78% of the 5305 common HCPs span within a factor 2 as represented in the greyed area.

Supplementary Figure S3: Representation of the 40 variable windows DIA methods. Histogram representing precursor distribution over the m/z acquisition range observed in DDA runs for PPA and CCCF samples (A) Presentation of the variable windows selected based on the precursor distribution and acquisition speed of the Q-Orbitrap HF-X instrument (B).

Supplementary Figure S4: Venn diagram of the quantified HCP lists obtained by the DDA, DIA and directDIA approaches.

Supplementary Figure S5: Benchmarking of the precision achieved by DDA, DIA and directDIA. Comparison of CVs calculated on HCP peptide intensities within replicates.

Supplementary Figure S6: Benchmarking of accuracy of the DDA and DIA acquisitions modes. Representation of the ratios calculated between the Top3-DIA and Top3-DDA amounts obtained for the common HCPs between both acquisitions methods (A) Out of range ratio were removed to observe the 52% of 3779 common HCPs within a factor 2. Grey area represents the ratios within a factor 2. Representation of the number of peptides used to quantify each HCPs (B).

Supplementary Figure S7: Benchmarking of our HCP Profiler-DIA method against ELISA. Comparison of the global HCP quantity obtained using DIA and ELISA assay. Bar height represents the mean of the global HCP amount in injection triplicates. Error bars represent the standard deviation. Superposition of the ratio between the HCP Profiler-DIA approach and ELISA assay.

Supplementary Figure S8: Benchmarking of accuracy of the DIA and directDIA extraction strategies. Representation of the ratios calculated between the Top3-DIA and Top3-directDIA amounts obtained for the common HCPs between both extractions methods (A) Out of range ratio were removed to observe the 82% of 5833 common HCPs within a factor 2. Grey area represents the ratios within a factor 2. Representation of the number of peptides used to quantify each HCPs (B).

Supplementary Figure S9: Venn diagram of the HCPs quantified using directDIA and identified in the spectral library.

Supplementary Table S1

Windows	Start (m/z)	End (m/z)	Center (m/z)	Width (m/z)
1	350	375	362.5	25
2	374.5	399.5	387	25
3	399	424	411.5	25
4	423.5	448.5	436	25
5	448	473	460.5	25
6	472.5	497.5	485	25
7	497	509	503	12
8	508.5	520.5	514.5	12
9	520	532	526	12
10	531.5	543.5	537.5	12
11	543	555	549	12
12	554.5	566.5	560.5	12
13	566	578	572	12
14	577.5	589.5	583.5	12
15	589	601	595	12
16	600.5	612.5	606.5	12
17	612	624	618	12
18	623.5	635.5	629.5	12
19	635	647	641	12
20	646.5	658.5	652.5	12
21	658	670	664	12
22	669.5	681.5	675.5	12
23	681	693	687	12
24	692.5	704.5	698.5	12
25	704	716	710	12
26	715.5	727.5	721.5	12
27	727	739	733	12
28	738.5	750.5	744.5	12

29	750	762	756	12
30	761.5	791.5	776.5	30
31	791	821	806	30
32	820.5	850.5	835.5	30
33	850	880	865	30
34	879.5	909.5	894.5	30
35	909	929	919	20
36	928.5	948.5	938.5	20
37	948	968	958	20
38	967.5	987.5	977.5	20
39	987	1007	997	20
40	1006.5	1026.5	1016.5	20

Supplementary Table S2

Windows	Start (m/z)	End (m/z)	Center (m/z)	Width (m/z)
1	350	375	362.5	25
2	374.5	399.5	387	25
3	399	424	411.5	25
4	423.5	436.5	430	13
5	436	449	442.5	13
6	448.5	461.5	455	13
7	461	474	467.5	13
8	473.5	486.5	480	13
9	486	499	492.5	13
10	498.5	511.5	505	13
11	511	524	517.5	13
12	523.5	536.5	530	13
13	536	549	542.5	13
14	548.5	561.5	555	13
15	561	574	567.5	13
16	573.5	586.5	580	13
17	586	599	592.5	13
18	598.5	611.5	605	13
19	611	624	617.5	13
20	623.5	636.5	630	13
21	636	649	642.5	13
22	648.5	661.5	655	13
23	661	674	667.5	13
24	673.5	686.5	680	13
25	686	699	692.5	13
26	698.5	711.5	705	13
27	711	724	717.5	13
28	723.5	736.5	730	13

29	736	749	742.5	13
30	748.5	761.5	755	13
31	761	774	767.5	13
32	773.5	798.5	786	25
33	798	823	810.5	25
34	822.5	847.5	835	25
35	847	869.5	858.25	22.5
36	869	891.5	880.25	22.5
37	891	913.5	902.25	22.5
38	913	935.5	924.25	22.5
39	935	957.5	946.25	22.5
40	957	979.5	968.25	22.5

Supplementary Table S3

<i>Acquisition mode</i>	<i>Sample</i>	<i>Replicate</i>	<i>slope</i>	<i>intercept</i>	<i>R²</i>
DDA	PPA 1	1	0.938	24.157	0.953
		2	0.931	24.287	0.950
		3	0.935	24.148	0.955
	PPA 2	1	0.939	24.084	0.954
		2	0.940	24.119	0.952
		3	0.935	24.196	0.948
	PPA 3	1	0.912	24.192	0.951
		2	0.922	24.099	0.952
		3	0.922	24.061	0.952
	PPA 5	1	0.917	24.386	0.951
		2	0.900	24.495	0.946
		3	0.921	24.382	0.948
	PPA 6	1	0.932	23.871	0.958
		2	0.912	24.074	0.958
		3	0.918	24.073	0.951
	PPA 7	1	0.930	23.866	0.955
		2	0.935	23.935	0.951
		3	0.940	23.876	0.954
	PPA 8	1	0.956	23.839	0.958
		2	0.940	23.999	0.949
		3	0.935	24.063	0.948
	Harvest 1	1	0.964	23.859	0.948
		2	0.934	23.931	0.954
		3	0.943	23.935	0.955
	Harvest 2	1	0.938	24.044	0.956
		2	0.935	24.049	0.957
		3	0.942	23.948	0.966
	Harvest 3	1	0.963	23.669	0.955
		2	0.948	23.764	0.967
		3	0.945	23.827	0.955
	Harvest 4	1	0.981	23.248	0.944
		2	0.984	23.274	0.943
		3	0.993	23.259	0.943
DIA	PPA 1	1	1.080	19.761	0.949
		2	1.087	19.809	0.944
		3	1.081	19.745	0.948
	PPA 2	1	1.100	19.535	0.945
		2	1.079	19.783	0.949
		3	1.089	19.782	0.942
	PPA 3	1	1.064	19.888	0.946
		2	1.056	19.839	0.943
		3	1.077	19.792	0.945

	PPA 5	1	1.060	20.127	0.947
		2	1.068	20.139	0.948
		3	1.066	20.118	0.942
	PPA 6	1	1.060	19.674	0.943
		2	1.070	19.715	0.945
		3	1.079	19.728	0.942
	PPA 7	1	1.073	19.497	0.951
		2	1.060	19.497	0.949
		3	1.072	19.48	0.951
	PPA 8	1	1.088	19.621	0.944
		2	1.099	19.538	0.943
		3	1.110	19.527	0.941
	Harvest 1	1	1.040	19.71	0.946
		2	1.041	19.815	0.943
		3	1.054	19.814	0.949
	Harvest 2	1	1.041	19.966	0.953
		2	1.041	20.061	0.951
		3	1.027	19.491	0.950
	Harvest 3	1	1.050	19.69	0.958
		2	1.050	19.83	0.953
		3	1.049	19.736	0.953
	Harvest 4	1	1.034	19.33	0.952
		2	1.023	19.419	0.952
		3	1.044	19.566	0.946
DirectDIA	PPA 1	1	1.064	19.865	0.945
		2	1.087	19.87	0.945
		3	1.083	19.845	0.942
	PPA 2	1	1.086	19.634	0.944
		2	1.084	19.743	0.943
		3	1.075	19.873	0.941
	PPA 3	1	1.047	19.947	0.936
		2	1.036	19.912	0.932
		3	1.060	19.86	0.937
	PPA 5	1	1.066	20.136	0.949
		2	1.054	20.194	0.942
		3	1.052	20.179	0.937
	PPA 6	1	1.053	19.711	0.935
		2	1.062	19.74	0.938
		3	1.070	19.76	0.936
	PPA 7	1	1.065	19.517	0.944
		2	1.058	19.524	0.943
		3	1.066	19.494	0.943
	PPA 8	1	1.065	19.517	0.944
		2	1.058	19.524	0.943
		3	1.066	19.494	0.943

	Harvest 1	1	1.083	19.45	0.941
		2	1.079	19.586	0.936
		3	1.073	19.61	0.931
	Harvest 2	1	1.077	19.712	0.943
		2	1.058	19.87	0.932
		3	1.048	19.251	0.931
	Harvest 3	1	1.065	19.503	0.942
		2	1.082	19.617	0.944
		3	1.082	19.522	0.945
	Harvest 4	1	1.049	19.11	0.932
		2	1.039	19.207	0.930
		3	1.060	19.333	0.922
	SD DDA		2.2	1.2	0.6
	SD DIA		2.0	1.0	0.4
	SD directDIA		1.3	1.3	0.6

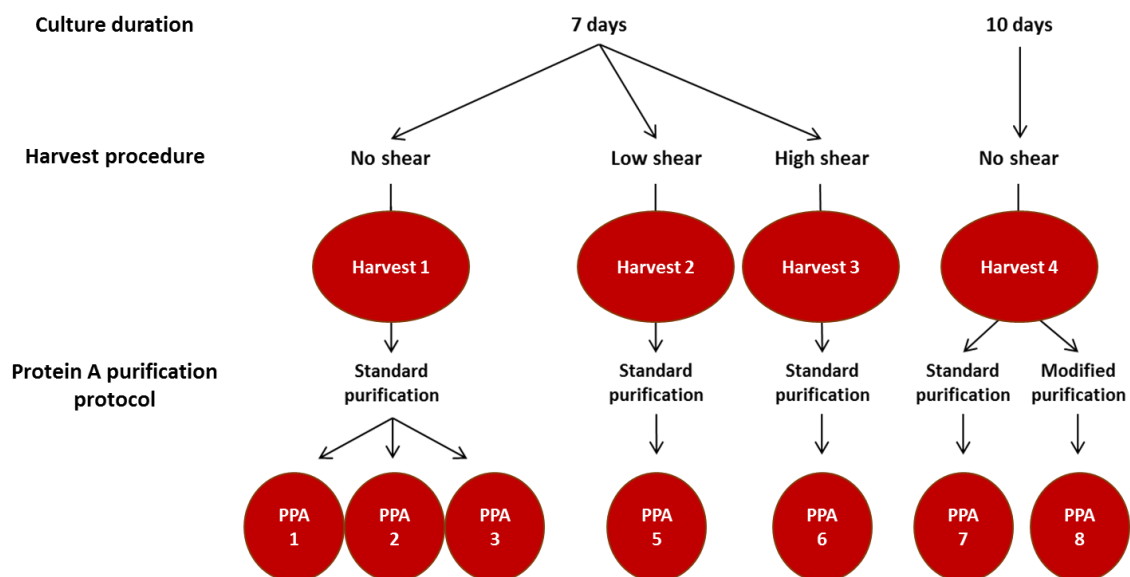
Supplementary Table S4

Acquisition mode	Sample	Minimum quantity (ng/mg)	sd	Maximum quantity (ng/mg)	sd	log
DDA	PPA 1	0.33	0.02	206	19	2.8
	PPA 2	0.45	0.05	223	8	2.7
	PPA 3	0.31	0.04	182	17	2.8
	PPA 5	0.15	0.01	66	6	2.7
	PPA 6	0.61	0.02	152	6	2.4
	PPA 7	0.19	0.03	279	12	3.2
	PPA 8	0.22	0.04	1782	160	3.9
	Harvest 1	0.03	0.003	9603	605	5.5
	Harvest 2	0.05	0.01	14112	364	5.5
	Harvest 3	0.07	0.01	14356	631	5.3
	Harvest 4	0.11	0.01	25377	699	5.3
DIA	PPA 1	0.05	0.01	573	57	4.1
	PPA 2	0.05	0.002	666	77	4.1
	PPA 3	0.13	0.003	541	27	3.6
	PPA 5	0.06	0.01	559	1	4.0
	PPA 6	0.14	0.02	416	24	3.5
	PPA 7	0.03	0.003	594	9	4.4
	PPA 8	0.02	0.004	511	25	4.5
	Harvest 1	0.004	0.001	5179	326	6.1
	Harvest 2	0.02	0.003	2517	187	5.1
	Harvest 3	0.02	0.001	4140	176	5.3
	Harvest 4	0.05	0.01	16233	651	5.5
DirectDIA	PPA 1	0.04	0.01	224	19	3.7
	PPA 2	0.08	0.02	182	16	3.4
	PPA 3	0.07	0.01	147	11	3.3
	PPA 5	0.02	0.002	185	30	4.1
	PPA 6	0.03	0.005	125	13	3.6
	PPA 7	0.03	0.005	125	13	3.6
	PPA 8	0.001	0.0001	459	21	5.8
	Harvest 1	0.01	0.001	5532	317	5.6
	Harvest 2	0.02	0.01	2636	202	5.1
	Harvest 3	0.0002	0.00002	4540	160	7.3
	Harvest 4	0.03	0.004	18175	1578	5.8

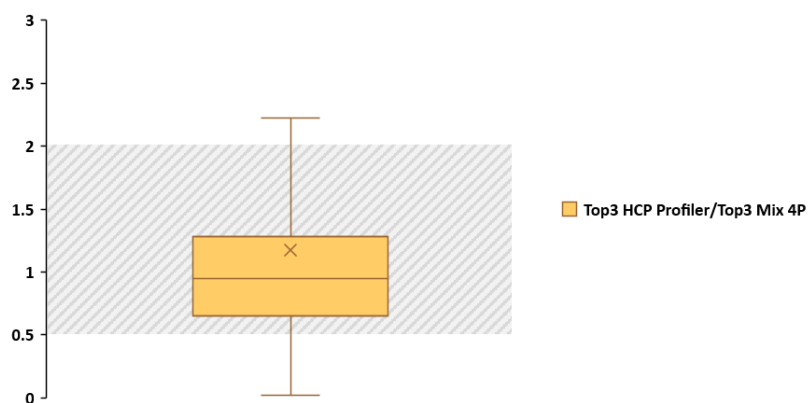
Supplementary Table S5

Sample	Minimum quantity (ng/mg)	sd	Maximum quantity (ng/mg)	sd	log
Harvest 1	0.01	0.001	993	74	4.9
Harvest 2	0.02	0.01	1423	79	4.8
Harvest 3	0.0002	0.00002	2777	270	7.1
Harvest 4	0.04	0.01	2123	308	4.7

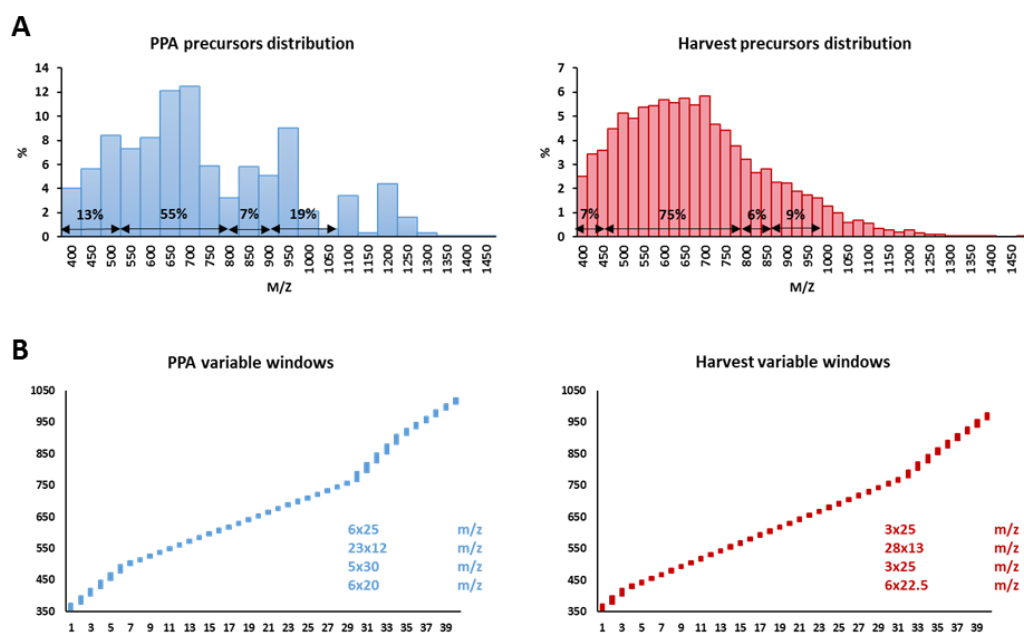
Supplementary Figure S1



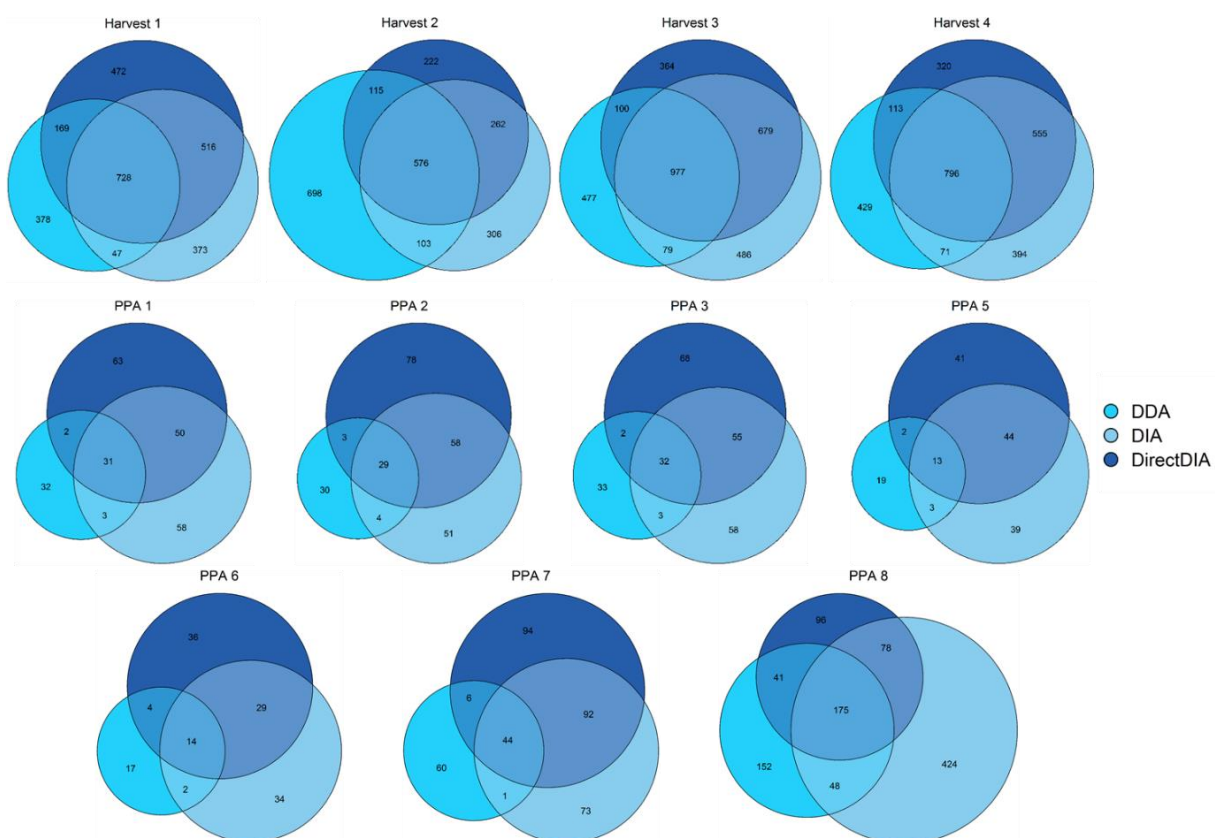
Supplementary Figure S2



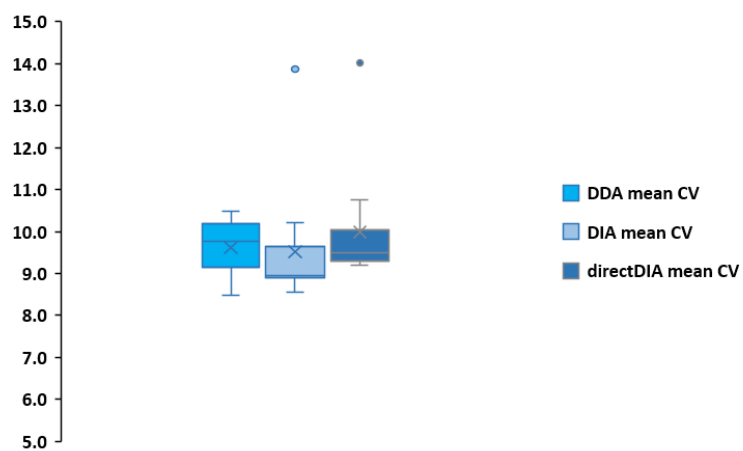
Supplementary Figure S3



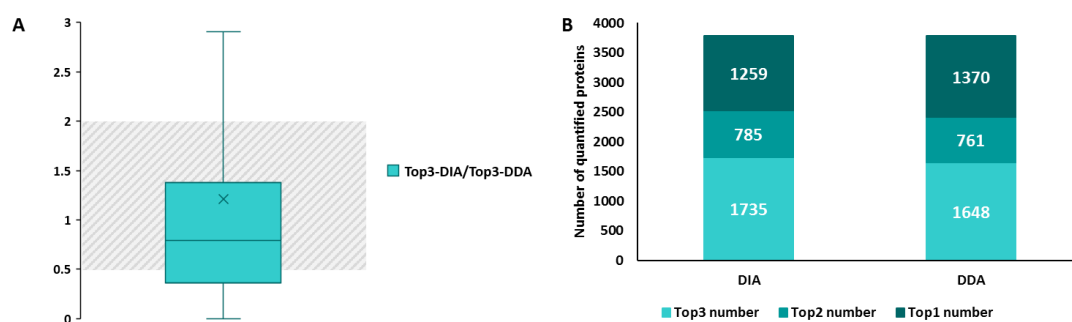
Supplementary Figure S4



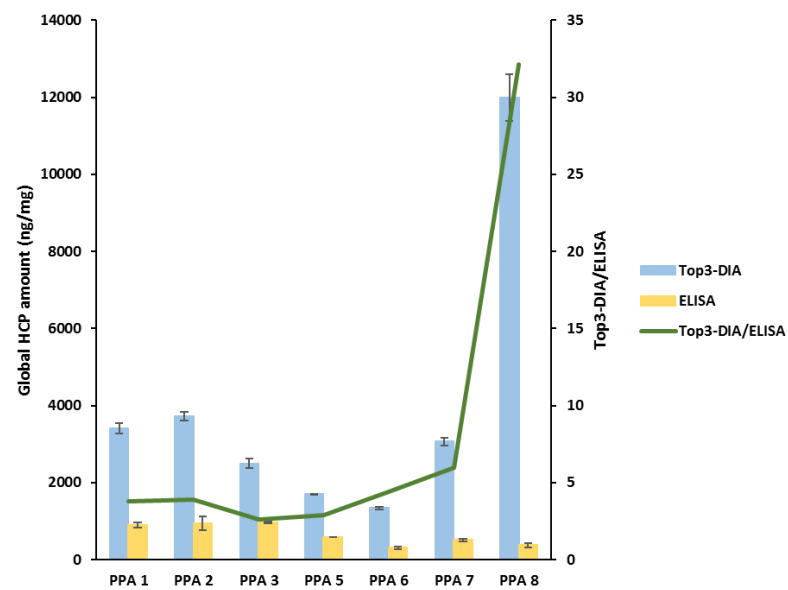
Supplementary Figure S5



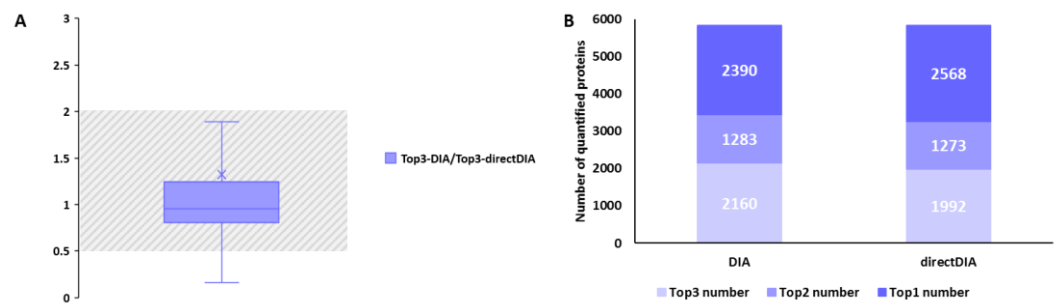
Supplementary Figure S6



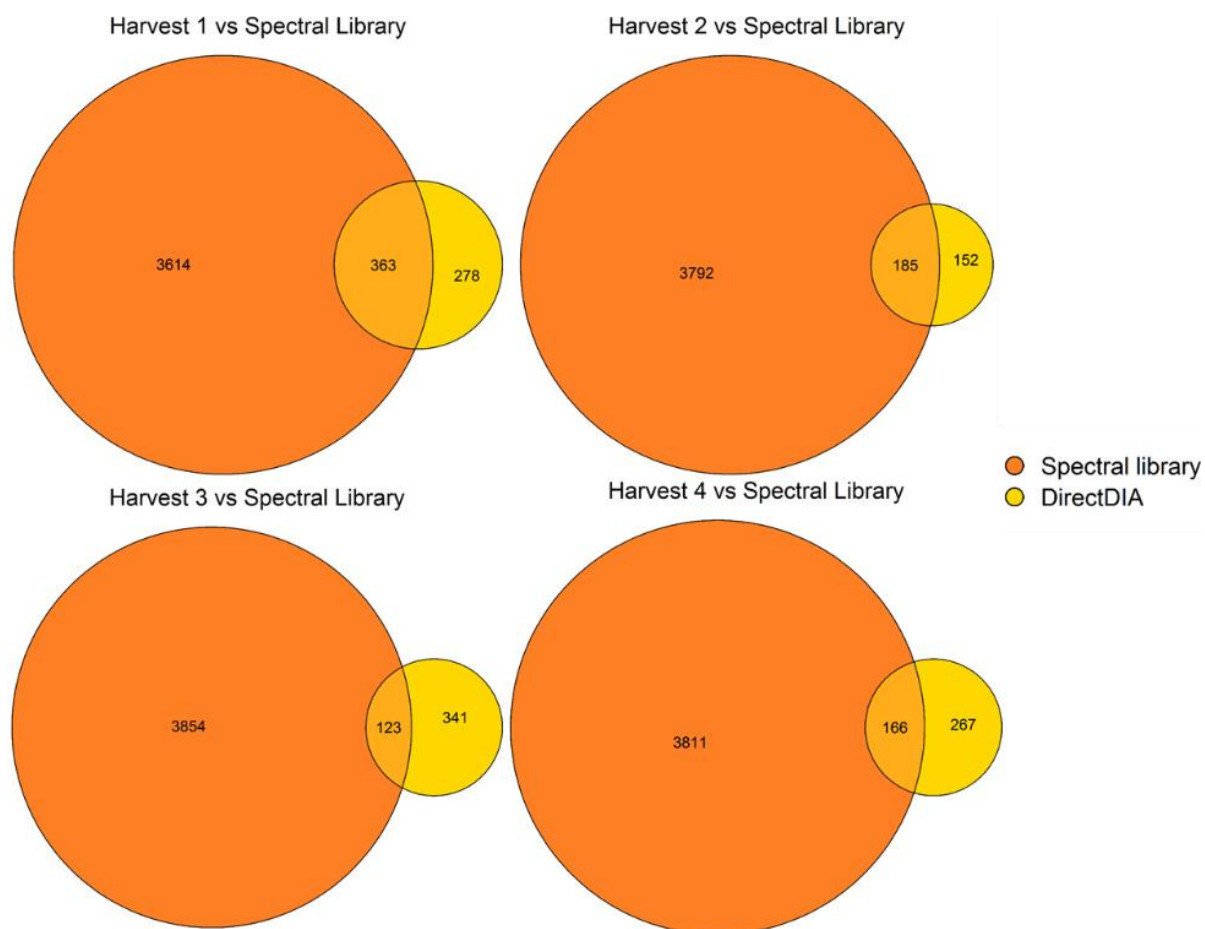
Supplementary Figure S7



Supplementary Figure S8



Supplementary Figure S9



References

1. Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N., Strategies and challenges for the next generation of therapeutic antibodies. *Nature Reviews Immunology* **2010**, *10* (5), 345-352.
2. Grilo, A. L.; Mantalaris, A., The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends in Biotechnology* **2019**, *37* (1), 9-16.
3. Kaplon, H.; Muralidharan, M.; Schneider, Z.; Reichert, J. M., Antibodies to watch in 2020. *mAbs* **2020**, *12* (1), 1703531-1703531.
4. Valente, K. N.; Levy, N. E.; Lee, K. H.; Lenhoff, A. M., Applications of proteomic methods for CHO host cell protein characterization in biopharmaceutical manufacturing. *Current opinion in biotechnology* **2018**, *53*, 144-150.
5. Yang, B.; Li, W.; Zhao, H.; Wang, A.; Lei, Y.; Xie, Q.; Xiong, S., Discovery and characterization of CHO host cell protease-induced fragmentation of a recombinant monoclonal antibody during production process development. *Journal of Chromatography B* **2019**, *1112*, 1-10.
6. Cui, T.; Chi, B.; Heidbrink Thompson, J.; Kasali, T.; Sellick, C.; Turner, R., Cathepsin D: Removal strategy on protein A chromatography, near real time monitoring and characterisation during monoclonal antibody production. *Journal of Biotechnology* **2019**, *305*, 51-60.
7. Hogwood, C. E.; Bracewell, D. G.; Smales, C. M., Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. *Current opinion in biotechnology* **2014**, *30*, 153-60.
8. Zhu-Shimoni, J.; Yu, C.; Nishihara, J.; Wong, R. M.; Gunawan, F.; Lin, M.; Krawitz, D.; Liu, P.; Sandoval, W.; Vanderlaan, M., Host cell protein testing by ELISAs and the use of orthogonal methods. *Biotechnology and bioengineering* **2014**, *111* (12), 2367-79.
9. Walker, D. E.; Yang, F.; Carver, J.; Joe, K.; Michels, D. A.; Yu, X. C., A modular and adaptive mass spectrometry-based platform for support of bioprocess development toward optimal host cell protein clearance. *mAbs* **2017**, *9* (4), 654-663.
10. Doneanu, C. E.; Xenopoulos, A.; Fadgen, K.; Murphy, J.; Skilton, S. J.; Prentice, H.; Stapels, M.; Chen, W., Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *MAbs* **2012**, *4* (1), 24-44.
11. Zhang, Q.; Goetze, A. M.; Cui, H.; Wylie, J.; Trimble, S.; Hewig, A.; Flynn, G. C., Comprehensive tracking of host cell proteins during monoclonal antibody purifications using mass spectrometry. *MAbs* **2014**, *6* (3), 659-70.
12. Picotti, P.; Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **2012**, *9* (6), 555-566.
13. Clavier, S.; Fougeron, D.; Petrovic, S.; Elmaleh, H.; Fourneaux, C.; Bugnazet, D.; Duffieux, F.; Masiero, A.; Mitra-Kaushik, S.; Genet, B.; Fromentin, Y.; Kreiss, P.; Laborderie, B.; Brault, D.; Menet, J. M., Improving the analytical toolbox to investigate copurifying host cell proteins presence: N-(4)-(β-acetylglucosaminyl)- L-asparaginase case study. *Biotechnology and bioengineering* **2020**, *117* (11), 3368-3378.
14. Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & cellular proteomics : MCP* **2012**, *11* (6), O111.016717.
15. Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Treiber, T.; Bruderer, R.; Reiter, L., Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Molecular omics* **2019**, *15* (5), 348-360.
16. Silva, J. C.; Gorenstein, M. V.; Li, G.-Z.; Vissers, J. P. C.; Geromanos, S. J., Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms Acquisition *S. *Molecular & Cellular Proteomics* **2006**, *5* (1), 144-156.
17. Doneanu, C. E.; Anderson, M.; Williams, B. J.; Lauber, M. A.; Chakraborty, A.; Chen, W., Enhanced Detection of Low-Abundance Host Cell Protein Impurities in High-Purity Monoclonal Antibodies Down to 1 ppm Using Ion Mobility Mass Spectrometry Coupled with Multidimensional Liquid Chromatography. *Anal Chem* **2015**, *87* (20), 10283-91.
18. Trauchessec, M.; Hesse, A. M.; Kraut, A.; Berard, Y.; Herment, L.; Fortin, T.; Bruley, C.; Ferro, M.; Manin, C., An innovative standard for LC-MS-based HCP profiling and accurate quantity assessment: Application to batch consistency in viral vaccine samples. *Proteomics* **2021**, *21* (5), e2000152.

19. Bouyssie, D.; Hesse, A. M.; Mouton-Barbosa, E.; Rompais, M.; Macron, C.; Carapito, C.; Gonzalez de Peredo, A.; Coute, Y.; Dupierris, V.; Burel, A.; Menetrey, J. P.; Kalaitzakis, A.; Poisat, J.; Romdhani, A.; Burlet-Schiltz, O.; Cianferani, S.; Garin, J.; Bruley, C., Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **2020**, *36* (10), 3148-3155.
20. Schenauer, M. R.; Flynn, G. C.; Goetze, A. M., Identification and quantification of host cell protein impurities in biotherapeutics using mass spectrometry. *Anal Biochem* **2012**, *428* (2), 150-7.
21. Kelstrup, C. D.; Bekker-Jensen, D. B.; Arrey, T. N.; Hoglebe, A.; Harder, A.; Olsen, J. V., Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J Proteome Res* **2018**, *17* (1), 727-738.
22. Goey, C. H.; Alhuthali, S.; Kontoravdi, C., Host cell protein removal from biopharmaceutical preparations: Towards the implementation of quality by design. *Biotechnology advances* **2018**, *36* (4), 1223-1237.
23. Aboulaich, N.; Chung, W. K.; Thompson, J. H.; Larkin, C.; Robbins, D.; Zhu, M., A novel approach to monitor clearance of host cell proteins associated with monoclonal antibodies. *Biotechnology progress* **2014**, *30* (5), 1114-24.
24. Levy, N. E.; Valente, K. N.; Choe, L. H.; Lee, K. H.; Lenhoff, A. M., Identification and characterization of host cell protein product-associated impurities in monoclonal antibody bioprocessing. *Biotechnology and bioengineering* **2014**, *111* (5), 904-12.
25. Meier, F.; Brunner, A. D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; Bache, N.; Hoerning, O.; Cox, J.; Räther, O.; Mann, M., Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Molecular & cellular proteomics : MCP* **2018**, *17* (12), 2534-2545.
26. Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M. A.; Raether, O.; Mann, M., Parallel Accumulation-Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J Proteome Res* **2015**, *14* (12), 5378-87.
27. Yang, F.; Walker, D. E.; Schoenfelder, J.; Carver, J.; Zhang, A.; Li, D.; Harris, R.; Stults, J. T.; Yu, X. C.; Michels, D. A., A 2D LC-MS/MS Strategy for Reliable Detection of 10-ppm Level Residual Host Cell Proteins in Therapeutic Antibodies. *Analytical Chemistry* **2018**, *90* (22), 13365-13372.
28. Chen, I. H.; Xiao, H.; Li, N., Improved host cell protein analysis in monoclonal antibody products through ProteoMiner. *Analytical Biochemistry* **2020**, *610*, 113972.
29. Huang, L.; Wang, N.; Mitchell, C. E.; Brownlee, T.; Maple, S. R.; De Felippis, M. R., A Novel Sample Preparation for Shotgun Proteomics Characterization of HCPs in Antibodies. *Anal Chem* **2017**, *89* (10), 5436-5444.
30. Li, Q.; Feng, Y.; Tan, M.-J.; Zhai, L.-H., Evaluation of Endoproteinase Lys-C/Trypsin Sequential Digestion Used in Proteomics Sample Preparation. *Chinese Journal of Analytical Chemistry* **2017**, *45* (3), 316-321.
31. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L., BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421.
32. Kaplon, H.; Reichert, J. M., Antibodies to watch in 2021. *MAbs* **2021**, *13* (1), 1860476.
33. Reichert, J. M.; Valge-Archer, V. E., Development trends for monoclonal antibody cancer therapeutics. *Nature reviews. Drug discovery* **2007**, *6* (5), 349-56.
34. Suzuki, M.; Kato, C.; Kato, A., Therapeutic antibodies: their mechanisms of action and the pathological findings they induce in toxicity studies. *J Toxicol Pathol* **2015**, *28* (3), 133-139.
35. Bracewell, D. G.; Francis, R.; Smales, C. M., The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnology and bioengineering* **2015**, *112* (9), 1727-37.
36. Rane, S. S.; Dearman, R. J.; Kimber, I.; Uddin, S.; Bishop, S.; Shah, M.; Podmore, A.; Pluen, A.; Derrick, J. P., Impact of a Heat Shock Protein Impurity on the Immunogenicity of Biotherapeutic Monoclonal Antibodies. *Pharmaceutical research* **2019**, *36* (4), 51.
37. Gao, X.; Rawal, B.; Wang, Y.; Li, X.; Wylie, D.; Liu, Y.-H.; Breunig, L.; Driscoll, D.; Wang, F.; Richardson, D. D., Targeted Host Cell Protein Quantification by LC-MRM Enables Biologics Processing and Product Characterization. *Analytical Chemistry* **2020**, *92* (1), 1007-1015.
38. Li, X.; An, Y.; Liao, J.; Xiao, L.; Swanson, M.; Martinez-Fonts, K.; Pavon, J. A.; Sherer, E. C.; Jawa, V.; Wang, F.; Gao, X.; Letarte, S.; Richardson, D. D., Identification and characterization of a residual host cell protein hexosaminidase B associated with N-glycan degradation during the stability study of a therapeutic recombinant monoclonal antibody product. *Biotechnology progress* **2021**, e3128.
39. Tscheliessnig, A. L.; Konrath, J.; Bates, R.; Jungbauer, A., Host cell protein analysis in therapeutic protein bioprocessing - methods and applications. *Biotechnology journal* **2013**, *8* (6), 655-70.
40. Vanderlaan, M.; Zhu-Shimoni, J.; Lin, S.; Gunawan, F.; Waerner, T.; Van Cott, K. E., Experience with host cell protein impurities in biopharmaceuticals. *Biotechnology progress* **2018**, *34* (4), 828-837.

41. Bomans, K.; Lang, A.; Roedl, V.; Adolf, L.; Kyriosoglou, K.; Diepold, K.; Eberl, G.; Mølhøj, M.; Strauss, U.; Schmalz, C.; Vogel, R.; Reusch, D.; Wegele, H.; Wiedmann, M.; Bulau, P., Identification and monitoring of host cell proteins by mass spectrometry combined with high performance immunochemistry testing. *PLoS one* **2013**, *8* (11), e81639-e81639.
42. Madsen, J. A.; Farutin, V.; Carbeau, T.; Wudyka, S.; Yin, Y.; Smith, S.; Anderson, J.; Capila, I., Toward the complete characterization of host cell proteins in biotherapeutics via affinity depletions, LC-MS/MS, and multivariate analysis. *MAbs* **2015**, *7* (6), 1128-37.
43. Lavoie, R. A.; di Fazio, A.; Williams, T. I.; Carbonell, R.; Menegatti, S., Targeted capture of Chinese hamster ovary host cell proteins: Peptide ligand binding by proteomic analysis. *Biotechnology and bioengineering* **2020**, *117* (2), 438-452.
44. Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R., Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol* **2018**, *14* (8), e8126.
45. Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T., Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020**, *20* (17-18), 1900276.
46. Ting, Y. S.; Egertson, J. D.; Bollinger, J. G.; Searle, B. C.; Payne, S. H.; Noble, W. S.; MacCoss, M. J., PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nat Methods* **2017**, *14* (9), 903-908.
47. Tsou, C. C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A. C.; Nesvizhskii, A. I., DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* **2015**, *12* (3), 258-64, 7 p following 264.
48. Jawa, V.; Joubert, M. K.; Zhang, Q.; Deshpande, M.; Hapuarachchi, S.; Hall, M. P.; Flynn, G. C., Evaluating Immunogenicity Risk Due to Host Cell Protein Impurities in Antibody-Based Biotherapeutics. *The AAPS journal* **2016**, *18* (6), 1439-1452.
49. Wang, Q.; Slaney, T. R.; Wu, W.; Ludwig, R.; Tao, L.; Leone, A., Enhancing Host-Cell Protein Detection in Protein Therapeutics Using HILIC Enrichment and Proteomic Analysis. *Analytical Chemistry* **2020**, *92* (15), 10327-10335.
50. Johnson, R. O. B.; Greer, T.; Cejckov, M.; Zheng, X.; Li, N., Combination of FAIMS, Protein A Depletion, and Native Digest Conditions Enables Deep Proteomic Profiling of Host Cell Proteins in Monoclonal Antibodies. *Analytical Chemistry* **2020**, *92* (15), 10478-10484.
51. Nie, S.; Greer, T.; O'Brien Johnson, R.; Zheng, X.; Torri, A.; Li, N., Simple and Sensitive Method for Deep Profiling of Host Cell Proteins in Therapeutic Antibodies by Combining Ultra-Low Trypsin Concentration Digestion, Long Chromatographic Gradients, and BoxCar Mass Spectrometry Acquisition. *Analytical Chemistry* **2021**, *93* (10), 4383-4390.
52. Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **1996**, *14* (1), 61-5.
53. Anderson, N. L.; Anderson, N. G., Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **1998**, *19* (11), 1853-61.
54. Graves, P. R.; Haystead, T. A., Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev* **2002**, *66* (1), 39-63.
55. James, P., Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* **1997**, *30* (4), 279-331.
56. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013**, *113* (4), 2343-94.
57. Aebersold, R.; Mann, M., Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537* (7620), 347-55.
58. Li, H.; Han, J.; Pan, J.; Liu, T.; Parker, C. E.; Borchers, C. H., Current trends in quantitative proteomics - an update. *J Mass Spectrom* **2017**, *52* (5), 319-341.
59. Gillet, L. C.; Leitner, A.; Aebersold, R., Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu Rev Anal Chem (Palo Alto Calif)* **2016**, *9* (1), 449-72.
60. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988**, *60* (20), 2299-301.
61. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246* (4926), 64-71.
62. Nesvizhskii, A. I.; Aebersold, R., Interpretation of Shotgun Proteomic Data. *Molecular & Cellular Proteomics* **2005**, *4* (10), 1419-1440.

63. Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J., The one hour yeast proteome. *Molecular & cellular proteomics : MCP* **2014**, *13* (1), 339-47.
64. Patrie, S. M., Top-Down Mass Spectrometry: Proteomics to Proteoforms. *Adv Exp Med Biol* **2016**, *919*, 171-200.
65. Catherman, A. D.; Durbin, K. R.; Ahlf, D. R.; Early, B. P.; Fellers, R. T.; Tran, J. C.; Thomas, P. M.; Kelleher, N. L., Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Molecular & cellular proteomics : MCP* **2013**, *12* (12), 3465-73.
66. McCool, E. N.; Lubeckyj, R. A.; Shen, X.; Chen, D.; Kou, Q.; Liu, X.; Sun, L., Deep Top-Down Proteomics Using Capillary Zone Electrophoresis-Tandem Mass Spectrometry: Identification of 5700 Proteoforms from the Escherichia coli Proteome. *Anal Chem* **2018**, *90* (9), 5529-5533.
67. Cupp-Sutton, K. A.; Wu, S., High-throughput quantitative top-down proteomics. *Molecular omics* **2020**, *16* (2), 91-99.
68. Ghezellou, P.; Garikapati, V.; Kazemi, S. M.; Strupat, K.; Ghassempour, A.; Spengler, B., A perspective view of top-down proteomics in snake venom research. *Rapid Commun Mass Spectrom* **2019**, *33* Suppl 1, 20-27.
69. LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Shams, D. P.; Thomas, P. M.; Kelleher, N. L., Accurate Estimation of Context-Dependent False Discovery Rates in Top-Down Proteomics. *Molecular & cellular proteomics : MCP* **2019**, *18* (4), 796-805.
70. Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W.; Mohammed, S.; Heck, A. J., Toward an Optimized Workflow for Middle-Down Proteomics. *Anal Chem* **2017**, *89* (6), 3318-3325.
71. Pandeswari, P. B.; Sabareesh, V., Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry. *RSC Advances* **2019**, *9* (1), 313-344.
72. Channaveerappa, D.; Ngounou Wetie, A. G.; Darie, C. C., Bottlenecks in Proteomics: An Update. *Adv Exp Med Biol* **2019**, *1140*, 753-769.
73. Rogers, J. C.; Bomgarden, R. D., Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. *Adv Exp Med Biol* **2016**, *919*, 43-62.
74. Bose, U.; Wijffels, G.; Howitt, C. A.; Colgrave, M. L., Proteomics: Tools of the Trade. *Adv Exp Med Biol* **2019**, *1073*, 1-22.
75. Feist, P.; Hummon, A. B., Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **2015**, *16* (2), 3537-63.
76. Tubaon, R. M.; Haddad, P. R.; Quirino, J. P., Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016. *Proteomics* **2017**, *17* (20).
77. Shevchenko, G.; Musunuri, S.; Wetterhall, M.; Bergquist, J., Comparison of extraction methods for the comprehensive analysis of mouse brain proteome using shotgun-based mass spectrometry. *J Proteome Res* **2012**, *11* (4), 2441-51.
78. International Human Genome Sequencing, C., Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431* (7011), 931-945.
79. Wang, D.; Eraslan, B.; Wieland, T.; Hallstrom, B.; Hopf, T.; Zolg, D. P.; Zecha, J.; Asplund, A.; Li, L. H.; Meng, C.; Frejno, M.; Schmidt, T.; Schnatbaum, K.; Wilhelm, M.; Ponten, F.; Uhlen, M.; Gagneur, J.; Hahne, H.; Kuster, B., A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol* **2019**, *15* (2), e8503.
80. Wu, L.; Han, D. K., Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert Rev Proteomics* **2006**, *3* (6), 611-9.
81. Zubarev, R. A., The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* **2013**, *13* (5), 723-6.
82. Gstaiger, M.; Aebersold, R., Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet* **2009**, *10* (9), 617-27.
83. Gianazza, E.; Miller, I.; Palazzolo, L.; Parravicini, C.; Eberini, I., With or without you - Proteomics with or without major plasma/serum proteins. *J Proteomics* **2016**, *140*, 62-80.
84. Salvato, F.; Gallo de Carvalho, M. C. d. C.; Lima Leite, A. d., Strategies for Protein Separation. **2012**.
85. Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J., Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol* **2014**, *10*, 757.
86. Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M., Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **2014**, *11* (3), 319-24.
87. Zougman, A.; Selby, P. J.; Banks, R. E., Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics* **2014**, *14* (9), 1006-0.
88. Laemmli, U. K., Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **1970**, *227* (5259), 680-685.

89. Rabilloud, T.; Vaezzadeh, A. R.; Potier, N.; Lelong, C.; Leize-Wagner, E.; Chevallet, M., Power and limitations of electrophoretic separations in proteomics strategies. *Mass Spectrom Rev* **2009**, *28* (5), 816-43.
90. Lu, X.; Zhu, H., Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. *Molecular & cellular proteomics : MCP* **2005**, *4* (12), 1948-58.
91. Muller, L.; Fornecker, L.; Chion, M.; Van Dorsselaer, A.; Cianferani, S.; Rabilloud, T.; Carapito, C., Extended investigation of tube-gel sample preparation: a versatile and simple choice for high throughput quantitative proteomics. *Sci Rep* **2018**, *8* (1), 8260.
92. Muller, L.; Fornecker, L.; Van Dorsselaer, A.; Cianferani, S.; Carapito, C., Benchmarking sample preparation/digestion protocols reveals tube-gel being a fast and repeatable method for quantitative proteomics. *Proteomics* **2016**, *16* (23), 2953-2961.
93. Rabilloud, T.; Lelong, C., Two-dimensional gel electrophoresis in proteomics: a tutorial. *J Proteomics* **2011**, *74* (10), 1829-41.
94. Candiano, G.; Bruschi, M.; Musante, L.; Santucci, L.; Ghiggeri, G. M.; Carnemolla, B.; Orecchia, P.; Zardi, L.; Righetti, P. G., Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **2004**, *25* (9), 1327-33.
95. Burkhart, J. M.; Schumbrutski, C.; Wortelkamp, S.; Sickmann, A.; Zahedi, R. P., Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J Proteomics* **2012**, *75* (4), 1454-62.
96. Vandermarliere, E.; Mueller, M.; Martens, L., Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom Rev* **2013**, *32* (6), 453-65.
97. Glatter, T.; Ludwig, C.; Ahrne, E.; Aebersold, R.; Heck, A. J.; Schmidt, A., Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* **2012**, *11* (11), 5145-56.
98. Hakobyan, A.; Schneider, M. B.; Liesack, W.; Glatter, T., Efficient Tandem LysC/Trypsin Digestion in Detergent Conditions. *Proteomics* **2019**, *19* (20), e1900136.
99. Kobs, G. Same-Day Mass Spec Sample Prep Now a Reality: Shorten Digestion Time to as Little as 60 Minutes. <https://france.promega.com/resources/pubhub/2017/same-day-mass-spectrometry-sample-prep-is-now-a-reality/> (accessed 08/06/2020).
100. Camerini, S.; Mauri, P., The role of protein and peptide separation before mass spectrometry analysis in clinical proteomics. *J Chromatogr A* **2015**, *1381*, 1-12.
101. Xie, F.; Smith, R. D.; Shen, Y., Advanced proteomic liquid chromatography. *J Chromatogr A* **2012**, *1261*, 78-90.
102. Zhang, Z.; Wu, S.; Stenoien, D. L.; Pasa-Tolic, L., High-throughput proteomics. *Annu Rev Anal Chem (Palo Alto Calif)* **2014**, *7*, 427-54.
103. Moruz, L.; Kall, L., Peptide retention time prediction. *Mass Spectrom Rev* **2017**, *36* (5), 615-623.
104. Steen, H.; Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nature reviews. Molecular cell biology* **2004**, *5* (9), 699-711.
105. Stahl, D. C.; Swiderek, K. M.; Davis, M. T.; Lee, T. D., Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. *Journal of the American Society for Mass Spectrometry* **1996**, *7* (6), 532-540.
106. Michalski, A.; Cox, J.; Mann, M., More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC-MS/MS. *Journal of Proteome Research* **2011**, *10* (4), 1785-1793.
107. Richards, A. L.; Hebert, A. S.; Ulbrich, A.; Bailey, D. J.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J., One-hour proteome analysis in yeast. *Nat Protoc* **2015**, *10* (5), 701-14.
108. Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A., Fundamentals and Advances of Orbitrap Mass Spectrometry. **2019**, 1-40.
109. Hodge, K.; Have, S. T.; Hutton, L.; Lamond, A. I., Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *J Proteomics* **2013**, *88*, 92-103.
110. Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* **2004**, *39* (10), 1091-112.
111. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **2007**, *4* (9), 709-12.
112. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004**, *101* (26), 9528-33.

113. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W., Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* **2000**, *72* (3), 563-73.
114. Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R., 3rd, Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Anal Chem* **2004**, *76* (5), 1243-8.
115. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Brei, L. A., Mobile and localized protons: a framework for understanding peptide dissociation. *Journal of Mass Spectrometry* **2000**, *35* (12), 1399-1406.
116. Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). **1990**, *193*, 886-887.
117. D'Atri, V.; Causon, T.; Hernandez-Alba, O.; Mutabazi, A.; Veuthey, J. L.; Cianferani, S.; Guillaume, D., Adding a new separation dimension to MS and LC-MS: What is the utility of ion mobility spectrometry? *Journal of separation science* **2018**, *41* (1), 20-67.
118. Dodds, J. N.; Baker, E. S., Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *Journal of the American Society for Mass Spectrometry* **2019**, *30* (11), 2185-2195.
119. Landreh, M.; Sahin, C.; Gault, J.; Sadeghi, S.; Drum, C. L.; Uzdaviny, P.; Drew, D.; Allison, T. M.; Degiacomi, M. T.; Marklund, E. G., Predicting the Shapes of Protein Complexes through Collision Cross Section Measurements and Database Searches. *Analytical Chemistry* **2020**, *92* (18), 12297-12303.
120. Stiving, A. Q.; Jones, B. J.; Ujma, J.; Giles, K.; Wysocki, V. H., Collision Cross Sections of Charge-Reduced Proteins and Protein Complexes: A Database for Collision Cross Section Calibration. *Analytical Chemistry* **2020**, *92* (6), 4475-4483.
121. Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z.-J., Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nature Communications* **2020**, *11* (1), 4334.
122. Guevremont, R., High-field asymmetric waveform ion mobility spectrometry: A new tool for mass spectrometry. *Journal of Chromatography A* **2004**, *1058* (1-2), 3-19.
123. Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J. J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J., Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem* **2018**, *90* (15), 9529-9537.
124. Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; Belford, M.; Dunyach, J. J.; Thibault, P., A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Molecular & cellular proteomics : MCP* **2018**, *17* (10), 2051-2067.
125. Michelmann, K.; Silveira, J. A.; Ridgeway, M. E.; Park, M. A., Fundamentals of Trapped Ion Mobility Spectrometry. *Journal of the American Society for Mass Spectrometry* **2015**, *26* (1), 14-24.
126. Nesvizhskii, A. I., A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics* **2010**, *73* (11), 2092-123.
127. Blueggel, M.; Chamrad, D.; Meyer, H. E., Bioinformatics in proteomics. *Curr Pharm Biotechnol* **2004**, *5* (1), 79-88.
128. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M., Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **2011**, *10* (4), 1794-805.
129. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551-3567.
130. Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H., Open mass spectrometry search algorithm. *J Proteome Res* **2004**, *3* (5), 958-64.
131. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **1994**, *5* (11), 976-989.
132. Craig, R.; Cortens, J. P.; Beavis, R. C., Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *Journal of Proteome Research* **2004**, *3* (6), 1234-1242.
133. Griss, J.; Perez-Riverol, Y.; Lewis, S.; Tabb, D. L.; Dianes, J. A.; Del-Toro, N.; Rurik, M.; Walzer, M. W.; Kohlbacher, O.; Hermjakob, H.; Wang, R.; Vizcaino, J. A., Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat Methods* **2016**, *13* (8), 651-656.
134. Chick, J. M.; Kolippakkam, D.; Nusinow, D. P.; Zhai, B.; Rad, R.; Huttlin, E. L.; Gygi, S. P., A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat Biotechnol* **2015**, *33* (7), 743-9.
135. Skinner, O. S.; Kelleher, N. L., Illuminating the dark matter of shotgun proteomics. *Nat Biotechnol* **2015**, *33* (7), 717-8.

136. Coordinators, N. R., Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **2018**, *46* (D1), D8-D13.
137. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000**, *28* (1), 235-42.
138. Wu, C. H.; Yeh, L. S.; Huang, H.; Arminski, L.; Castro-Alvaredo, J.; Chen, Y.; Hu, Z.; Kourtesis, P.; Ledley, R. S.; Suzek, B. E.; Vinayaka, C. R.; Zhang, J.; Barker, W. C., The Protein Information Resource. *Nucleic Acids Res* **2003**, *31* (1), 345-7.
139. O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V. S.; Kodali, V. K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K. M.; Murphy, M. R.; O'Neill, K.; Pujar, S.; Rangwala, S. H.; Rausch, D.; Riddick, L. D.; Schoch, C.; Shkeda, A.; Storz, S. S.; Sun, H.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R. E.; Vatsan, A. R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M. J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T. D.; Pruitt, K. D., Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **2016**, *44* (D1), D733-45.
140. UniProt, C., UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **2019**, *47* (D1), D506-D515.
141. Ogasawara, O.; Kodama, Y.; Mashima, J.; Kosuge, T.; Fujisawa, T., DDBJ Database updates and computational infrastructure enhancement. *Nucleic Acids Res* **2020**, *48* (D1), D45-d50.
142. Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W., GenBank. *Nucleic Acids Res* **2016**, *44* (D1), D67-72.
143. Armengaud, J., A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr Opin Microbiol* **2009**, *12* (3), 292-300.
144. González-Gomariz, J.; Guruceaga, E.; López-Sánchez, M.; Segura, V., Proteogenomics in the context of the Human Proteome Project (HPP). *Expert Review of Proteomics* **2019**, *16* (3), 267-275.
145. Nesvizhskii, A. I., Proteogenomics: concepts, applications and computational strategies. *Nat Methods* **2014**, *11* (11), 1114-25.
146. Jaffe, J. D.; Berg, H. C.; Church, G. M., Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **2004**, *4* (1), 59-77.
147. Eicher, T.; Patt, A.; Kautto, E.; Machiraju, R.; Mathé, E.; Zhang, Y., Challenges in proteogenomics: a comparison of analysis methods with the case study of the DREAM proteogenomics sub-challenge. *BMC Bioinformatics* **2019**, *20* (24), 669.
148. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007**, *4* (3), 207-14.
149. Navarro, P.; Vázquez, J. S., A Refined Method To Calculate False Discovery Rates for Peptide Identification Using Decoy Databases. *Journal of Proteome Research* **2009**, *8* (4), 1792-1796.
150. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* **2007**, *389* (4), 1017-31.
151. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* **2002**, *1* (5), 376-86.
152. Emadali, A.; Gallagher-Gambarelli, M., [Quantitative proteomics by SILAC: practicalities and perspectives for an evolving approach]. *Med Sci (Paris)* **2009**, *25* (10), 835-42.
153. Geiger, T.; Cox, J.; Ostasiewicz, P.; Wisniewski, J. R.; Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* **2010**, *7* (5), 383-5.
154. Merrill, A. E.; Hebert, A. S.; MacGillivray, M. E.; Rose, C. M.; Bailey, D. J.; Bradley, J. C.; Wood, W. W.; El Masri, M.; Westphall, M. S.; Gasch, A. P.; Coon, J. J., NeuCode labels for relative protein quantification. *Molecular & cellular proteomics : MCP* **2014**, *13* (9), 2503-12.
155. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999**, *17* (10), 994-9.
156. Ankney, J. A.; Muneer, A.; Chen, X., Relative and Absolute Quantitation in Mass Spectrometry-Based Proteomics. *Annu Rev Anal Chem (Palo Alto Calif)* **2018**, *11* (1), 49-77.
157. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003**, *75* (8), 1895-904.

158. Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; Pribil, P.; Lee, K. H., 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* **2007**, *7* (20), 3651-60.
159. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics : MCP* **2004**, *3* (12), 1154-69.
160. Schubert, O. T.; Rost, H. L.; Collins, B. C.; Rosenberger, G.; Aebersold, R., Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc* **2017**, *12* (7), 1289-1294.
161. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **2005**, *1* (5), 252-62.
162. Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry* **2012**, *404* (4), 939-65.
163. Blein-Nicolas, M.; Zivy, M., Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochim Biophys Acta* **2016**, *1864* (8), 883-95.
164. Nogueira, F. C.; Domont, G. B., Survey of shotgun proteomics. *Methods Mol Biol* **2014**, *1156*, 3-23.
165. Ramus, C.; Hovasse, A.; Marcellin, M.; Hesse, A. M.; Mouton-Barbosa, E.; Bouyssie, D.; Vaca, S.; Carapito, C.; Chaoui, K.; Bruley, C.; Garin, J.; Cianferani, S.; Ferro, M.; Van Dorssaeler, A.; Burlet-Schiltz, O.; Schaeffer, C.; Coute, Y.; Gonzalez de Peredo, A., Benchmarking quantitative label-free LC-MS data processing workflows using a complex spiked proteomic standard dataset. *J Proteomics* **2016**, *132*, 51-62.
166. Pino, L. K.; Searle, B. C.; Huang, E. L.; Noble, W. S.; Hoofnagle, A. N.; MacCoss, M. J., Calibration Using a Single-Point External Reference Material Harmonizes Quantitative Mass Spectrometry Proteomics Data between Platforms and Laboratories. *Anal Chem* **2018**, *90* (21), 13112-13117.
167. Smith, R.; Tostengard, A. R., Quantitative Evaluation of Ion Chromatogram Extraction Algorithms. *J Proteome Res* **2020**, *19* (5), 1953-1964.
168. Pino, L. K.; Searle, B. C.; Bollinger, J. G.; Nunn, B.; MacLean, B.; MacCoss, M. J., The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev* **2017**, *39* (3), 229-244.
169. Cox, J.; Hein, M. Y.; Lubner, C. A.; Paron, I.; Nagaraj, N.; Mann, M., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* **2014**, *13* (9), 2513-26.
170. Valikangas, T.; Suomi, T.; Elo, L. L., A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief Bioinform* **2018**, *19* (6), 1344-1355.
171. Wang, X.; Shen, S.; Rasam, S. S.; Qu, J., MS1 ion current-based quantitative proteomics: A promising solution for reliable analysis of large biological cohorts. *Mass Spectrom Rev* **2019**, *38* (6), 461-482.
172. Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M., Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & cellular proteomics : MCP* **2005**, *4* (9), 1265-72.
173. Rappsilber, J.; Ryder, U.; Lamond, A. I.; Mann, M., Large-scale proteomic analysis of the human spliceosome. *Genome Res* **2002**, *12* (8), 1231-45.
174. Schwanhauss, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M., Global quantification of mammalian gene expression control. *Nature* **2011**, *473* (7347), 337-42.
175. Borrás, E.; Sabido, E., What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics* **2017**, *17* (17-18).
176. Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A. J.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C., Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res* **2010**, *9* (2), 761-76.
177. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J., Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & cellular proteomics : MCP* **2012**, *11* (11), 1475-88.
178. Vidova, V.; Spacil, Z., A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Anal Chim Acta* **2017**, *964*, 7-23.
179. Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B., Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP* **2012**, *11* (12), 1709-23.

180. Kockmann, T.; Trachsel, C.; Panse, C.; Wahlander, A.; Selevsek, N.; Grossmann, J.; Wolski, W. E.; Schlapbach, R., Targeted proteomics coming of age - SRM, PRM and DIA performance evaluated from a core facility perspective. *Proteomics* **2016**, *16* (15-16), 2183-92.
181. Rauniyar, N., Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int J Mol Sci* **2015**, *16* (12), 28566-81.
182. Fan, H.-F.; Cheng, Y.-S.; Ma, C.-H.; Jayaram, M., Single molecule TPM analysis of the catalytic pentad mutants of Cre and Flp site-specific recombinases: contributions of the pentad residues to the pre-chemical steps of recombination. *Nucleic Acids Res* **2015**, *43* (6), 3237-3255.
183. Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Edes, J.; Loevenich, S. N.; Aebersold, R., The PeptideAtlas project. *Nucleic Acids Res* **2006**, *34* (Database issue), D655-8.
184. Sharma, V.; Eckels, J.; Taylor, G. K.; Shulman, N. J.; Stergachis, A. B.; Joyner, S. A.; Yan, P.; Whiteaker, J. R.; Halusa, G. N.; Schilling, B.; Gibson, B. W.; Colangelo, C. M.; Paulovich, A. G.; Carr, S. A.; Jaffe, J. D.; MacCoss, M. J.; MacLean, B., Panorama: a targeted proteomics knowledge base. *J Proteome Res* **2014**, *13* (9), 4205-10.
185. Picotti, P.; Lam, H.; Campbell, D.; Deutsch, E. W.; Mirzaei, H.; Ranish, J.; Domon, B.; Aebersold, R., A database of mass spectrometric assays for the yeast proteome. *Nat Methods* **2008**, *5* (11), 913-4.
186. Zahn-Zabal, M.; Lane, L., What will neXtProt help us achieve in 2020 and beyond? *Expert Rev Proteomics* **2020**, *17* (2), 95-98.
187. Prasad, T. S.; Kandasamy, K.; Pandey, A., Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods Mol Biol* **2009**, *577*, 67-79.
188. Mallick, P.; Schirle, M.; Chen, S. S.; Flory, M. R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; Kuster, B.; Aebersold, R., Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* **2007**, *25* (1), 125-31.
189. Fusaro, V. A.; Mani, D. R.; Mesirov, J. P.; Carr, S. A., Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat Biotechnol* **2009**, *27* (2), 190-8.
190. Searle, B. C.; Egerton, J. D.; Bollinger, J. G.; Stergachis, A. B.; MacCoss, M. J., Using Data Independent Acquisition (DIA) to Model High-responding Peptides for Targeted Proteomics Experiments. *Molecular & cellular proteomics : MCP* **2015**, *14* (9), 2331-40.
191. Gao, Z.; Chang, C.; Yang, J.; Zhu, Y.; Fu, Y., AP3: An Advanced Proteotypic Peptide Predictor for Targeted Proteomics by Incorporating Peptide Digestibility. *Anal Chem* **2019**, *91* (13), 8705-8711.
192. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **2003**, *100* (12), 6940-5.
193. Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J., Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* **2005**, *2* (8), 587-9.
194. Brun, V.; Dupuis, A.; Adrait, A.; Marcellin, M.; Thomas, D.; Court, M.; Vandenesch, F.; Garin, J., Isotope-labeled protein standards: toward absolute quantitative proteomics. *Molecular & cellular proteomics : MCP* **2007**, *6* (12), 2139-49.
195. Singh, S.; Springer, M.; Steen, J.; Kirschner, M. W.; Steen, H., FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J Proteome Res* **2009**, *8* (5), 2201-10.
196. Zeiler, M.; Straube, W. L.; Lundberg, E.; Uhlen, M.; Mann, M., A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & cellular proteomics : MCP* **2012**, *11* (3), O111 009613.
197. Masselon, C.; Anderson, G. A.; Harkewicz, R.; Bruce, J. E.; Pasa-Tolic, L.; Smith, R. D., Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures. *Anal Chem* **2000**, *72* (8), 1918-24.
198. Purvine, S.; Eppel, J. T.; Yi, E. C.; Goodlett, D. R., Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **2003**, *3* (6), 847-50.
199. Venable, J. D.; Dong, M. Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* **2004**, *1* (1), 39-45.
200. Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S., Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **2005**, *77* (7), 2187-200.
201. Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R., Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Anal Chem* **2009**, *81* (15), 6481-8.
202. Geiger, T.; Cox, J.; Mann, M., Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Molecular & cellular proteomics : MCP* **2010**, *9* (10), 2252-61.

203. Carvalho, P. C.; Han, X.; Xu, T.; Cociorva, D.; Carvalho Mda, G.; Barbosa, V. C.; Yates, J. R., 3rd, XDIA: improving on the label-free data-independent analysis. *Bioinformatics* **2010**, *26* (6), 847-8.
204. Weisbrod, C. R.; Eng, J. K.; Hoopmann, M. R.; Baker, T.; Bruce, J. E., Accurate peptide fragment mass analysis: multiplexed peptide identification and quantification. *J Proteome Res* **2012**, *11* (3), 1621-32.
205. Geromanos, S. J.; Hughes, C.; Ciavarini, S.; Vissers, J. P.; Langridge, J. I., Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples. *Analytical and bioanalytical chemistry* **2012**, *404* (4), 1127-39.
206. Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; Wu, C. C.; MacCoss, M. J., Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods* **2013**, *10* (8), 744-6.
207. Prakash, A.; Peterman, S.; Ahmad, S.; Sarracino, D.; Frewen, B.; Vogelsang, M.; Byram, G.; Krastins, B.; Vadali, G.; Lopez, M., Hybrid data acquisition and processing strategies with increased throughput and selectivity: pSMART analysis for global qualitative and quantitative analysis. *J Proteome Res* **2014**, *13* (12), 5415-30.
208. Martin, L. B. B.; Sherwood, R. W.; Nicklay, J. J.; Yang, Y.; Muratore-Schroeder, T. L.; Anderson, E. T.; Thannhauser, T. W.; Rose, J. K. C.; Zhang, S., Application of wide selected-ion monitoring data-independent acquisition to identify tomato fruit proteins regulated by the CUTIN DEFICIENT2 transcription factor. *Proteomics* **2016**, *16* (15-16), 2081-2094.
209. Distler, U.; Kuharev, J.; Navarro, P.; Levin, Y.; Schild, H.; Tenzer, S., Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods* **2014**, *11* (2), 167-70.
210. Zhang, Y.; Bilbao, A.; Bruderer, T.; Luban, J.; Strambio-De-Castillia, C.; Lisacek, F.; Hopfgartner, G.; Varesio, E., The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J Proteome Res* **2015**, *14* (10), 4359-71.
211. Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Miladinovic, S. M.; Cheng, L. Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; Vitek, O.; Rinner, O.; Reiter, L., Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Molecular & cellular proteomics : MCP* **2015**, *14* (5), 1400-10.
212. Moseley, M. A.; Hughes, C. J.; Juvvadi, P. R.; Soderblom, E. J.; Lennon, S.; Perkins, S. R.; Thompson, J. W.; Steinbach, W. J.; Geromanos, S. J.; Wildgoose, J.; Langridge, J. I.; Richardson, K.; Vissers, J. P. C., Scanning Quadrupole Data-Independent Acquisition, Part A: Qualitative and Quantitative Characterization. *J Proteome Res* **2018**, *17* (2), 770-779.
213. Meier, F.; Geyer, P. E.; Virreira Winter, S.; Cox, J.; Mann, M., BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nature Methods* **2018**, *15* (6), 440-448.
214. Messner, C. B.; Demichev, V.; Bloomfield, N.; Yu, J. S. L.; White, M.; Kreidl, M.; Egger, A.-S.; Freiwald, A.; Ivoisev, G.; Wasim, F.; Zeleznik, A.; Jürgens, L.; Suttrop, N.; Sander, L. E.; Kurth, F.; Lilley, K. S.; Müllender, M.; Tate, S.; Ralser, M., Ultra-fast proteomics with Scanning SWATH. *Nature Biotechnology* **2021**, *39* (7), 846-854.
215. Meier, F.; Brunner, A. D.; Frank, M.; Ha, A.; Bludau, I.; Voytik, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; Aebersold, R.; Collins, B. C.; Röst, H. L.; Mann, M., diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat Methods* **2020**, *17* (12), 1229-1236.
216. Cai, X.; Ge, W.; Yi, X.; Sun, R.; Zhu, J.; Lu, C.; Sun, P.; Zhu, T.; Ruan, G.; Yuan, C.; Liang, S.; Lyu, M.; Huang, S.; Zhu, Y.; Guo, T., PulseDIA: Data-Independent Acquisition Mass Spectrometry Using Multi-Injection Pulsed Gas-Phase Fractionation. *Journal of Proteome Research* **2021**, *20* (1), 279-288.
217. Bekker-Jensen, D. B.; Martinez-Val, A.; Steigerwald, S.; Ruther, P.; Fort, K. L.; Arrey, T. N.; Harder, A.; Makarov, A.; Olsen, J. V., A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Molecular & cellular proteomics : MCP* **2020**, *19* (4), 716-729.
218. Panchaud, A.; Jung, S.; Shaffer, S. A.; Aitchison, J. D.; Goodlett, D. R., Faster, quantitative, and accurate precursor acquisition independent from ion count. *Anal Chem* **2011**, *83* (6), 2250-7.
219. Huang, T.; Bruderer, R.; Muntel, J.; Xuan, Y.; Vitek, O.; Reiter, L., Combining Precursor and Fragment Information for Improved Detection of Differential Abundance in Data Independent Acquisition. *Molecular & cellular proteomics : MCP* **2020**, *19* (2), 421-430.
220. Hu, A.; Noble, W. S.; Wolf-Yadlin, A., Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res* **2016**, *5*.
221. Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.; Smith, R. D.; Noble, W. S.; MacCoss, M. J., Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Molecular & cellular proteomics : MCP* **2015**, *14* (9), 2301-7.
222. Schubert, O. T.; Gillet, L. C.; Collins, B. C.; Navarro, P.; Rosenberger, G.; Wolski, W. E.; Lam, H.; Amodei, D.; Mallick, P.; MacLean, B.; Aebersold, R., Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat Protoc* **2015**, *10* (3), 426-41.

223. Wu, J. X.; Song, X.; Pascovici, D.; Zaw, T.; Care, N.; Krisp, C.; Molloy, M. P., SWATH Mass Spectrometry Performance Using Extended Peptide MS/MS Assay Libraries. *Molecular & cellular proteomics : MCP* **2016**, *15* (7), 2501-14.
224. Matsumoto, M.; Matsuzaki, F.; Oshikawa, K.; Goshima, N.; Mori, M.; Kawamura, Y.; Ogawa, K.; Fukuda, E.; Nakatsumi, H.; Natsume, T.; Fukui, K.; Horimoto, K.; Nagashima, T.; Funayama, R.; Nakayama, K.; Nakayama, K. I., A large-scale targeted proteomics assay resource based on an in vitro human proteome. *Nat Methods* **2017**, *14* (3), 251-258.
225. Schubert, O. T.; Mouritsen, J.; Ludwig, C.; Rost, H. L.; Rosenberger, G.; Arthur, P. K.; Claassen, M.; Campbell, D. S.; Sun, Z.; Farrah, T.; Gengenbacher, M.; Maiolica, A.; Kaufmann, S. H. E.; Moritz, R. L.; Aebersold, R., The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis. *Cell Host Microbe* **2013**, *13* (5), 602-612.
226. Deutsch, E. W.; Csordas, A.; Sun, Z.; Jarnuczak, A.; Perez-Riverol, Y.; Ternent, T.; Campbell, D. S.; Bernal-Llinares, M.; Okuda, S.; Kawano, S.; Moritz, R. L.; Carver, J. J.; Wang, M.; Ishihama, Y.; Bandeira, N.; Hermjakob, H.; Vizcaino, J. A., The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res* **2017**, *45* (D1), D1100-D1106.
227. Rosenberger, G.; Koh, C. C.; Guo, T.; Rost, H. L.; Kouvonen, P.; Collins, B. C.; Heusel, M.; Liu, Y.; Caron, E.; Vichalkovski, A.; Faini, M.; Schubert, O. T.; Faridi, P.; Ebhardt, H. A.; Matondo, M.; Lam, H.; Bader, S. L.; Campbell, D. S.; Deutsch, E. W.; Moritz, R. L.; Tate, S.; Aebersold, R., A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci Data* **2014**, *1*, 140031.
228. Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Xuan, Y.; Sondermann, J.; Schmidt, M.; Gomez-Varela, D.; Reiter, L., Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Molecular & cellular proteomics : MCP* **2017**, *16* (12), 2296-2309.
229. Ammar, C.; Berchtold, E.; Csaba, G.; Schmidt, A.; Imhof, A.; Zimmer, R., Multi-Reference Spectral Library Yields Almost Complete Coverage of Heterogeneous LC-MS/MS Data Sets. *Journal of Proteome Research* **2019**, *18* (4), 1553-1566.
230. Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndt, M.; Brant, A.; Cimermanic, P.; Cox, J., High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nature Methods* **2019**, *16* (6), 519-525.
231. Gessulat, S.; Schmidt, T.; Zolg, D. P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; Reimer, U.; Ehrlich, H. C.; Aiche, S.; Kuster, B.; Wilhelm, M.; Prosser, proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods* **2019**, *16* (6), 509-518.
232. Zhou, X. X.; Zeng, W. F.; Chi, H.; Luo, C.; Liu, C.; Zhan, J.; He, S. M.; Zhang, Z., pDeep: Predicting MS/MS Spectra of Peptides with Deep Learning. *Anal Chem* **2017**, *89* (23), 12690-12697.
233. Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M., DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020**, *17* (1), 41-44.
234. Xu, L. L.; Young, A.; Zhou, A.; Rost, H. L., Machine Learning in Mass Spectrometric Analysis of DIA Data. *Proteomics* **2020**, e1900352.
235. Rosenberger, G.; Bludau, I.; Schmitt, U.; Heusel, M.; Hunter, C. L.; Liu, Y.; MacCoss, M. J.; MacLean, B. X.; Nesvizhskii, A. I.; Pedrioli, P. G. A.; Reiter, L.; Rost, H. L.; Tate, S.; Ting, Y. S.; Collins, B. C.; Aebersold, R., Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat Methods* **2017**, *14* (9), 921-927.
236. Rost, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinovic, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmstrom, J.; Malmstrom, L.; Aebersold, R., OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* **2014**, *32* (3), 219-23.
237. Reiter, L.; Rinner, O.; Picotti, P.; Huttenhain, R.; Beck, M.; Brusniak, M. Y.; Hengartner, M. O.; Aebersold, R., mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* **2011**, *8* (5), 430-5.
238. Rost, H. L.; Liu, Y.; D'Agostino, G.; Zanella, M.; Navarro, P.; Rosenberger, G.; Collins, B. C.; Gillet, L.; Testa, G.; Malmstrom, L.; Aebersold, R., TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. *Nat Methods* **2016**, *13* (9), 777-83.
239. Gupta, S.; Ahadi, S.; Zhou, W.; Rost, H., DIALignR Provides Precise Retention Time Alignment Across Distant Runs in DIA and Targeted Proteomics. *Molecular & cellular proteomics : MCP* **2019**, *18* (4), 806-817.
240. Zhang, N.; Li, X. J.; Ye, M.; Pan, S.; Schwikowski, B.; Aebersold, R., ProbiDtree: an automated software program capable of identifying multiple peptides from a single collision-induced dissociation spectrum collected by a tandem mass spectrometer. *Proteomics* **2005**, *5* (16), 4096-106.

241. Wang, J.; Tucholska, M.; Knight, J. D.; Lambert, J. P.; Tate, S.; Larsen, B.; Gingras, A. C.; Bandeira, N., MSPLIT-DIA: sensitive peptide identification for data-independent acquisition. *Nat Methods* **2015**, *12* (12), 1106-8.
242. Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villen, J.; MacCoss, M. J., Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat Commun* **2018**, *9* (1), 5128.
243. Golay, J.; Introna, M., Mechanism of action of therapeutic monoclonal antibodies: Promises and pitfalls of in vitro and in vivo assays. *Archives of Biochemistry and Biophysics* **2012**, *526* (2), 146-153.
244. KÖHLER, G.; Milstein, C., Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **1975**, *256* (5517), 495-497.
245. Hwang, W. Y.; Foote, J., Immunogenicity of engineered antibodies. *Methods* **2005**, *36* (1), 3-10.
246. Ober, R. J.; Radu, C. G.; Ghetie, V.; Ward, E. S., Differences in promiscuity for antibody-FcRn interactions across species: implications for therapeutic antibodies. *International Immunology* **2001**, *13* (12), 1551-1559.
247. Morrison, S. L.; Johnson, M. J.; Herzenberg, L. A.; Oi, V. T., Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proceedings of the National Academy of Sciences* **1984**, *81* (21), 6851.
248. Jones, P. T.; Dear, P. H.; Foote, J.; Neuberger, M. S.; Winter, G., Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **1986**, *321* (6069), 522-5.
249. Frenzel, A.; Hust, M.; Schirrmann, T., Expression of Recombinant Antibodies. *Frontiers in Immunology* **2013**, *4* (217).
250. Dumont, J.; Euwart, D.; Mei, B.; Estes, S.; Kshirsagar, R., Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical reviews in biotechnology* **2016**, *36* (6), 1110-1122.
251. Fischer, S.; Handrick, R.; Otte, K., The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology advances* **2015**, *33* (8), 1878-1896.
252. Kim, J. Y.; Kim, Y. G.; Lee, G. M., CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl Microbiol Biotechnol* **2012**, *93* (3), 917-30.
253. Minow, B.; Rogge, P.; Thompson, K., Implementing a Fully Disposable MAb Manufacturing Facility: Solutions and challenges. *BioProcess International* **2012**, *10*, 48-57.
254. Gronemeyer, P.; Ditz, R.; Strube, J., Trends in Upstream and Downstream Process Development for Antibody Manufacturing. *Bioengineering (Basel)* **2014**, *1* (4), 188-212.
255. Kunert, R.; Reinhart, D., Advances in recombinant antibody manufacturing. *Appl Microbiol Biotechnol* **2016**, *100* (8), 3451-61.
256. Liu, H. F.; Ma, J.; Winter, C.; Bayer, R., Recovery and purification process development for monoclonal antibody production. *mAbs* **2010**, *2* (5), 480-499.
257. Somasundaram, B.; Pleitt, K.; Shave, E.; Baker, K.; Lua, L. H. L., Progression of continuous downstream processing of monoclonal antibodies: Current trends and challenges. *Biotechnology and bioengineering* **2018**, *115* (12), 2893-2907.
258. Chahar, D. S.; Ravindran, S.; Pisal, S. S., Monoclonal antibody purification and its progression to commercial scale. *Biologicals : journal of the International Association of Biological Standardization* **2020**, *63*, 1-13.
259. Follman, D. K.; Fahrner, R. L., Factorial screening of antibody purification processes using three chromatography steps without protein A. *Journal of Chromatography A* **2004**, *1024* (1), 79-85.
260. Jensen, K., A NORMALLY OCCURRING STAPHYLOCOCCUS ANTIBODY IN HUMAN SERUM. *Acta Pathologica Microbiologica Scandinavica* **1958**, *44* (4), 421-428.
261. Hober, S.; Nord, K.; Linhult, M., Protein A chromatography for antibody purification. *J Chromatogr B Analyt Technol Biomed Life Sci* **2007**, *848* (1), 40-7.
262. ICH Guidance for Industry Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological products; 1999.
263. Chon, J. H.; Zerbis-Papastoitsis, G., Advances in the production and downstream processing of antibodies. *N Biotechnol* **2011**, *28* (5), 458-63.
264. Hogwood, C. E.; Bracewell, D. G.; Smales, C. M., Host cell protein dynamics in recombinant CHO cells: impacts from harvest to purification and beyond. *Bioengineered* **2013**, *4* (5), 288-91.
265. Falkenberg, H.; Waldera-Lupa, D. M.; Vanderlaan, M.; Schwab, T.; Krapfenbauer, K.; Studts, J. M.; Flad, T.; Waerner, T., Mass spectrometric evaluation of upstream and downstream process influences on host cell protein patterns in biopharmaceutical products. *Biotechnology progress* **2019**, *35* (3), e2788.
266. Henry, S. M.; Sutlief, E.; Salas-Solano, O.; Valliere-Douglass, J., ELISA reagent coverage evaluation by affinity purification tandem mass spectrometry. *MABs* **2017**, *9* (7), 1065-1075.

267. Wang, X.; Hunter, A. K.; Mozier, N. M., Host cell proteins in biologics development: Identification, quantitation and risk assessment. *Biotechnology and bioengineering* **2009**, *103* (3), 446-58.
268. Griffin, J. F. T.; Spittle, E.; Rodgers, C. R.; Liggett, S.; Cooper, M.; Bakker, D.; Bannantine, J. P., Immunoglobulin G1 enzyme-linked immunosorbent assay for diagnosis of Johne's Disease in red deer (*Cervus elaphus*). *Clin Diagn Lab Immunol* **2005**, *12* (12), 1401-1409.
269. Porcelli, B.; Ferretti, F.; Vindigni, C.; Terzuoli, L., Assessment of a Test for the Screening and Diagnosis of Celiac Disease. *Journal of clinical laboratory analysis* **2016**, *30* (1), 65-70.
270. Sblattero, D.; Berti, I.; Trevisiol, C.; Marzari, R.; Tommasini, A.; Bradbury, A.; Fasano, A.; Ventura, A.; Not, T., Human recombinant tissue transglutaminase ELISA: an innovative diagnostic assay for celiac disease. *The American journal of gastroenterology* **2000**, *95* (5), 1253-7.
271. Winkler, I. G.; Löchel, M.; Levesque, J. P.; Bodem, J.; Flügel, R. M.; Flower, R. L., A rapid streptavidin-capture ELISA specific for the detection of antibodies to feline foamy virus. *Journal of immunological methods* **1997**, *207* (1), 69-77.
272. Zhu, M.; Gong, X.; Hu, Y.; Ou, W.; Wan, Y., Streptavidin-biotin-based directional double Nanobody sandwich ELISA for clinical rapid and sensitive detection of influenza H5N1. *Journal of translational medicine* **2014**, *12*, 352.
273. Peng, J.; Song, S.; Xu, L.; Ma, W.; Liu, L.; Kuang, H.; Xu, C., Development of a monoclonal antibody-based sandwich ELISA for peanut allergen Ara h 1 in food. *Int J Environ Res Public Health* **2013**, *10* (7), 2897-2905.
274. Wang, S. Y.; Li, Z.; Wang, X. J.; Lv, S.; Yang, Y.; Zeng, L. Q.; Luo, F. H.; Yan, J. H.; Liang, D. F., Development of monoclonal antibody-based sandwich ELISA for detection of dextran. *Monoclonal antibodies in immunodiagnosis and immunotherapy* **2014**, *33* (5), 334-9.
275. MacPhee, D. J., Methodological considerations for improving Western blot analysis. *Journal of pharmacological and toxicological methods* **2010**, *61* (2), 171-7.
276. Mahmood, T.; Yang, P. C., Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences* **2012**, *4* (9), 429-34.
277. Magdeldin, S.; Enany, S.; Yoshida, Y.; Xu, B.; Zhang, Y.; Zureena, Z.; Lokamani, I.; Yaoita, E.; Yamamoto, T., Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. *Clinical Proteomics* **2014**, *11* (1), 16.
278. Simpson, R. J., Rapid coomassie blue staining of protein gels. *Cold Spring Harbor protocols* **2010**, *2010* (4), pdb.prot5413.
279. Chevallet, M.; Luche, S.; Rabilloud, T., Silver staining of proteins in polyacrylamide gels. *Nat Protoc* **2006**, *1* (4), 1852-8.
280. Grzeskowiak, J. K.; Tscheliessnig, A.; Toh, P. C.; Chusainow, J.; Lee, Y. Y.; Wong, N.; Jungbauer, A., 2-D DIGE to expedite downstream process development for human monoclonal antibody purification. *Protein expression and purification* **2009**, *66* (1), 58-65.
281. Jin, M.; Szapiel, N.; Zhang, J.; Hickey, J.; Ghose, S., Profiling of host cell proteins by two-dimensional difference gel electrophoresis (2D-DIGE): Implications for downstream process development. *Biotechnology and bioengineering* **2010**, *105* (2), 306-316.
282. Hayduk, E. J.; Choe, L. H.; Lee, K. H., A two-dimensional electrophoresis map of Chinese hamster ovary cell proteins based on fluorescence staining. *Electrophoresis* **2004**, *25* (15), 2545-56.
283. Hogwood, C. E.; Tait, A. S.; Koloteva-Levine, N.; Bracewell, D. G.; Smales, C. M., The dynamics of the CHO host cell protein profile during clarification and protein A capture in a platform antibody purification process. *Biotechnology and bioengineering* **2013**, *110* (1), 240-51.
284. Pilely, K.; Nielsen, S. B.; Draborg, A.; Henriksen, M. L.; Hansen, S. W. K.; Skriver, L.; Mørtz, E.; Lund, R. R., A novel approach to evaluate ELISA antibody coverage of host cell proteins-combining ELISA-based immunocapture and mass spectrometry. *Biotechnology progress* **2020**, *36* (4), e2983.
285. Gilgunn, S.; El-Sabbahy, H.; Albrecht, S.; Gaikwad, M.; Corrigan, K.; Deakin, L.; Jellum, G.; Bones, J., Identification and tracking of problematic host cell proteins removed by a synthetic, highly functionalized nonwoven media in downstream bioprocessing of monoclonal antibodies. *Journal of Chromatography A* **2019**, *1595*, 28-38.
286. Meleady, P.; Hoffrogge, R.; Henry, M.; Rupp, O.; Bort, J. H.; Clarke, C.; Brinkrolf, K.; Kelly, S.; Müller, B.; Doolan, P.; Hackl, M.; Beckmann, T. F.; Noll, T.; Grillari, J.; Barron, N.; Pühler, A.; Clynes, M.; Borth, N., Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. *Biotechnology and bioengineering* **2012**, *109* (6), 1386-1394.
287. Ludwig, K. R.; Schroll, M. M.; Hummon, A. B., Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. *J Proteome Res* **2018**, *17* (7), 2480-2490.

288. Husson, G.; Delangle, A.; O'Hara, J.; Cianferani, S.; Gervais, A.; Van Dorsselaer, A.; Bracewell, D.; Carapito, C., Dual Data-Independent Acquisition Approach Combining Global HCP Profiling and Absolute Quantification of Key Impurities during Bioprocess Development. *Anal Chem* **2018**, *90* (2), 1241-1247.
289. Li, D.; Farchone, A.; Zhu, Q.; Macchi, F.; Walker, D. E.; Michels, D. A.; Yang, F., Fast, Robust, and Sensitive Identification of Residual Host Cell Proteins in Recombinant Monoclonal Antibodies Using Sodium Deoxycholate Assisted Digestion. *Analytical Chemistry* **2020**, *92* (17), 11888-11894.
290. Boychyn, M.; Yim, S. S.; Bulmer, M.; More, J.; Bracewell, D. G.; Hoare, M., Performance prediction of industrial centrifuges using scale-down models. *Bioprocess and biosystems engineering* **2004**, *26* (6), 385-91.
291. Lau, E. C.; Kong, S.; McNulty, S.; Entwisle, C.; McIlgorm, A.; Dalton, K. A.; Hoare, M., An ultra scale-down characterization of low shear stress primary recovery stages to enhance selectivity of fusion protein recovery from its molecular variants. *Biotechnology and bioengineering* **2013**, *110* (7), 1973-1983.
292. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966-8.
293. Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Reiter, L., High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. *Proteomics* **2016**, *16* (15-16), 2246-56.
294. Mehta, D.; Scandola, S.; Uhrig, R. G., Direct data-independent acquisition (direct DIA) enables substantially improved label-free quantitative proteomics in Arabidopsis. *bioRxiv* **2020**, 2020.11.07.372276.
295. Reinhart, D.; Damjanovic, L.; Kaisermayer, C.; Sommeregger, W.; Gili, A.; Gasselhuber, B.; Castan, A.; Mayrhofer, P.; Grünwald-Gruber, C.; Kunert, R., Bioprocessing of Recombinant CHO-K1, CHO-DG44, and CHO-S: CHO Expression Hosts Favor Either mAb Production or Biomass Synthesis. *Biotechnology journal* **2019**, *14* (3), e1700686.
296. Zhang, Q.; Goetze, A. M.; Cui, H.; Wylie, J.; Tillotson, B.; Hewig, A.; Hall, M. P.; Flynn, G. C., Characterization of the co-elution of host cell proteins with monoclonal antibodies during protein A purification. *Biotechnology progress* **2016**, *32* (3), 708-17.
297. Prianichnikov, N.; Koch, H.; Koch, S.; Lubeck, M.; Heilig, R.; Brehmer, S.; Fischer, R.; Cox, J., MaxQuant Software for Ion Mobility Enhanced Shotgun Proteomics. *Molecular & cellular proteomics : MCP* **2020**, *19* (6), 1058-1069.
298. Boé, J.-F.; Beck, A.; Carrie, A.; Duhau, L.; Laloux, O.; Mouz, N.; Savoy, L.-A.; G., V., Analysis of impurities in biological active substances : a case study of monoclonal antibodies. *STP PHARMA PRATIQUES* **2014**, *24*.
299. Chen, W.; Doneanu, C. E.; Lauber, M.; Koza, S.; Prakash, K.; Stapels, M.; Fountain, K. J., Improved Identification and Quantification of Host Cell Proteins (HCPs) in Biotherapeutics Using Liquid Chromatography-Mass Spectrometry. In *State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques*, American Chemical Society: 2015; Vol. 1202, pp 357-393.
300. Ma, J.; Kilby, G. W., Sensitive, Rapid, Robust, and Reproducible Workflow for Host Cell Protein Profiling in Biopharmaceutical Process Development. *J Proteome Res* **2020**.
301. Ahluwalia, D.; Dhillon, H.; Slaney, T.; Song, H.; Boux, H.; Mehta, S.; Zhang, L.; Valdez, A.; Krishnamurthy, G., Identification of a host cell protein impurity in therapeutic protein, P1. *Journal of pharmaceutical and biomedical analysis* **2017**, *141*, 32-38.
302. Fussl, F.; Trappe, A.; Cook, K.; Scheffler, K.; Fitzgerald, O.; Bones, J., Comprehensive characterisation of the heterogeneity of adalimumab via charge variant analysis hyphenated on-line to native high resolution Orbitrap mass spectrometry. *MABs* **2019**, *11* (1), 116-128.
303. Pythoud, N.; Bons, J.; Mijola, G.; Beck, A.; Cianféran, S.; Carapito, C., Optimized Sample Preparation and Data Processing of Data-Independent Acquisition Methods for the Robust Quantification of Trace-Level Host Cell Protein Impurities in Antibody Drug Products. *Journal of Proteome Research* **2021**, *20* (1), 923-931.
304. Nowak, C.; J, K. C.; S, M. D.; Katiyar, A.; Bhat, R.; Sun, J.; Ponniah, G.; Neill, A.; Mason, B.; Beck, A.; Liu, H., Forced degradation of recombinant monoclonal antibodies: A practical guide. *MABs* **2017**, *9* (8), 1217-1230.
305. Bee, J. S.; Tie, L.; Johnson, D.; Dimitrova, M. N.; Jusino, K. C.; Afdahl, C. D., Trace levels of the CHO host cell protease cathepsin D caused particle formation in a monoclonal antibody product. *Biotechnology progress* **2015**, *31* (5), 1360-9.
306. Gao, S. X.; Zhang, Y.; Stansberry-Perkins, K.; Buko, A.; Bai, S.; Nguyen, V.; Brader, M. L., Fragmentation of a highly purified monoclonal antibody attributed to residual CHO cell protease activity. *Biotechnology and bioengineering* **2011**, *108* (4), 977-82.

307. Robert, F.; Bierau, H.; Rossi, M.; Agugiaro, D.; Soranzo, T.; Broly, H.; Mitchell-Logean, C., Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of a CHO cathepsin D protease. *Biotechnology and bioengineering* **2009**, *104* (6), 1132-41.
308. Bee, J. S.; Machiesky, L. M.; Peng, L.; Jusino, K. C.; Dickson, M.; Gill, J.; Johnson, D.; Lin, H.-Y.; Miller, K.; Heidbrink Thompson, J.; Remmele Jr, R. L., Identification of an IgG CDR sequence contributing to co-purification of the host cell protease cathepsin D. *Biotechnology progress* **2017**, *33* (1), 140-145.
309. Richter, W.; Hermsdorf, T.; Kronbach, T.; Dettmer, D., Refolding and Purification of Recombinant Human PDE7A Expressed in Escherichia coli as Inclusion Bodies. *Protein expression and purification* **2002**, *25* (1), 138-148.
310. Wang, W.; Ignatius, A. A.; Thakkar, S. V., Impact of Residual Impurities and Contaminants on Protein Stability. *Journal of Pharmaceutical Sciences* **2014**, *103* (5), 1315-1330.
311. Giese, A.; Bader, B.; Bieschke, J.; Schaffar, G.; Odoy, S.; Kahle, P. J.; Haass, C.; Kretzschmar, H., Single particle detection and characterization of synuclein co-aggregation. *Biochemical and Biophysical Research Communications* **2005**, *333* (4), 1202-1210.
312. Schokker, E. P.; Singh, H.; Pinder, D. N.; Creamer, L. K., Heat-induced aggregation of β -lactoglobulin AB at pH 2.5 as influenced by ionic strength and protein concentration. *International Dairy Journal* **2000**, *10* (4), 233-240.
313. Chiu, J.; Valente, K. N.; Levy, N. E.; Min, L.; Lenhoff, A. M.; Lee, K. H., Knockout of a difficult-to-remove CHO host cell protein, lipoprotein lipase, for improved polysorbate stability in monoclonal antibody formulations. *Biotechnology and bioengineering* **2017**, *114* (5), 1006-1015.
314. Zhang, S.; Xiao, H.; Molden, R.; Qiu, H.; Li, N., Rapid Polysorbate 80 Degradation by Liver Carboxylesterase in a Monoclonal Antibody Formulated Drug Substance at Early Stage Development. *Journal of Pharmaceutical Sciences* **2020**, *109* (11), 3300-3307.

Développement de stratégies analytiques en protéomique quantitative: quantification des protéines de la cellule hôte par spectrométrie de masse comme outil de contrôle qualité pour l'industrie biopharmaceutique

Résumé

Les stratégies de protéomique quantitative par spectrométrie de masse sont très attractives pour le suivi des protéines de la cellule hôte lors de la production d'anticorps thérapeutiques. Elles peuvent être divisées en trois catégories : les stratégies globales, les stratégies ciblées et les stratégies DIA qui combinent les avantages des deux précédentes.

Ce travail de thèse porte sur le développement de stratégies analytiques et bioinformatiques en protéomique quantitative pour l'identification et la quantification d'impuretés protéiques à l'état de traces. La première partie de mon travail implique l'optimisation de différentes étapes du protocole analytique en me focalisant sur un standard de quantification innovant, le mode d'acquisition DIA et le traitement de ces données complexes. Ces développements ont été réalisés afin d'assister les différentes étapes de la chaîne de production des anticorps. Par la suite, une méthode de quantification globale a été optimisée pour le suivi des impuretés protéiques présentes dans les substances médicamenteuses finales qui représentent un véritable défi analytique. Enfin, la troisième partie a consisté à développer une stratégie analytique ciblée sur une plateforme MS dédiée à l'analyse de routine.

Mots-clés : Protéomique quantitative, Spectrométrie de masse, Acquisition indépendante des données (DIA), Protéines de la cellule hôte (HCPs)

Résumé en anglais

Quantitative proteomics strategies based on mass spectrometry are very attractive for the monitoring of host cell proteins during the manufacture of therapeutic antibodies. They can be divided into three categories: global strategies, targeted strategies, and DIA strategies that combine the advantages of both.

This PhD work focuses on the development of analytical and bioinformatics strategies in quantitative proteomics for the identification and quantification of trace-level proteins. My first objective was to optimize different steps of the analytical workflow by focusing on an innovative quantification standard, the DIA acquisition mode, and the processing of these complex data. These developments aim at assisting the different steps of the manufacturing process of antibodies. Then, a global quantification method was optimized for the monitoring of protein impurities present in the final drug substances, which represent a real analytical challenge. Finally, the third part consisted in developing a targeted analytical strategy on a MS platform dedicated to routine analysis.

Keywords : Quantitative proteomics, Mass spectrometry, Data-independent acquisition (DIA), Host cell proteins (HCPs)