Development of new mass spectrometry methods for the characterization of protein impurities in therapeutic antibodies

Noelia Milagros de Lama Valderrama

Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

École doctorale
Sciences chimiques | ED 222
Université de Strasbourg

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

IPHC – UMR 7178

THÈSE présentée par / DISSERTATION
presented by :

Noelia Milagros DE LAMA VALDERRAMA

soutenue le /defended on : **14 Mars 2025**

pour obtenir le grade de / to obtain the grade of :
Docteur de l'Université de Strasbourg / Strasbourg University Doctor

Discipline/Spécialité / Discipline/Specialty : Chimie Analytique

Development of new mass spectrometry methods for the characterization of protein impurities in therapeutic antibodies

THÈSE dirigée par / DISSERTATION supervisor :

Dr. CARAPITO Christine	Directrice de Recherche, Université de Strasbourg

RAPPORTEURS :

Dr. FERRO Myriam	Directrice de Recherche, CEA, Grenoble
Dr. PINEAU Charles	Directeur de Recherche, INSERM, Rennes

AUTRES MEMBRES DU JURY / OTHER MEMBERS OF THE JURY :

Dr. BERTACCINI Diego	Laboratory Manager, Merck, Vevey
Dr. BONES Jonathan	Professor, University College Dublin
Dr. GUERRERA Chiara	Ingénieur de Recherche, INSERM, Paris

To my family and close friends,

To my parents and my sister,

To my maternal grandparents,

for their love and unwavering support.

With all my gratitude and deep affection.

ACKNOWLEDGEMENTS

I would first like to thank my thesis directors, **Christine** and **Diego**, for believing in me and giving me the opportunity to work on such an important project. Your guidance and support, combined with just the right amount of pressure, were exactly what I needed to succeed. You both allowed me to grow, not just as a researcher but also as a professional, and for that, I am incredibly grateful. Thank you for trusting me with such responsibility and for always being there with advice, encouragement, or just a word of wisdom when I needed it.

A big thank you to **Sarah Cianférani** for welcoming me into the LSMBO and always offering her support. It's been a pleasure getting to know you, and your leadership made the lab an ideal place to work.

Huge thanks to everyone in **R2**: **Véronique**, **Laurence** and Valériane who always kept the bio room in order and made sure the Bravo ran smoothly. Your dedication made everything easier, and your care for the lab has not gone unnoticed. **François** and **Magali**, thank you both for always being there when I had questions or needed advice. I felt supported in every interaction we had. **Alex**, thank you so much for always being there when I needed help with those stubborn technical issues. Whether it was getting my session to work or sorting out unexpected problems, you were always ready to lend a hand. Your support meant a lot, and I'm really grateful for it. It's been a pleasure working with you.

To my colleagues in R2—**Charline**, **Pauline**, **Bastien**, **Adrien**, and **Hugo**—thank you for making the work environment so much more pleasant. **Pauline** and **Charline**, I wish you both the best of luck as you continue your thesis work. I've seen firsthand the effort and passion you both put into your projects, and I'm sure you'll achieve great success. **Bastien**, thank you for making our time together in the lab so enjoyable, for always listening, and for being a great colleague to share the space with. **Adrien**, I can't thank you enough for all your help with the R scripts! You always took the time to assist me, even when you were swamped, and your guidance was invaluable. **Hugo**, although we didn't chat as much as I would've liked, I appreciate every moment we had. I know you will do great things in your projects, and I hope your time at the lab is rewarding.

A huge thank you also to my office mates: **Jeewan**, **Farida**, and **Aurélie**. **Jeewan**, thank you for teaching me so much with such grace and patience. You were not just a great colleague but also a tutor and a friend who made the start of my PhD journey much smoother. **Aurélie**, your willingness to help whenever we needed it never went unnoticed. Thank you for being generous with your time and with all those snacks you shared. **Farida**, my almost little sister, I am so glad our paths crossed. You brought joy, laughter, and comfort to my final months at the lab, and we became close so quickly. Your energy made the hardest days more manageable, and you were always there for me, making the work environment so much livelier. I know you will excel in your PhD journey, and I will be there to support you whenever you need me. You have become a dear friend, and I hope our bond will last far beyond our time at the lab.

Now to **R5**, where so many people contributed to making my time in the lab so enjoyable: **Fabrice V.**, **Fabrice B.**, **Oscar**, **Delphine**, **Hélène**, **Christine S.**, **Stéphane**, and especially **Martine**, thank you all for being so welcoming and for always being ready to help out when needed. Working in R5 was always a pleasure, thanks to all of you. A special thank you to **Jean Marc**—your constant help with the Eclipse machine was a lifesaver for me. I knew I could always count on you to make sure everything was running smoothly, and your support made my work so much easier.

To **Shabnam**, **Reiko**, **Léa**, **Guillaume**, **Sarah N.**, **Sarah D.**, and **Turkan**, I wish you all the best with your projects and your PhD journeys. I know you'll all succeed and make the most of your time at the LSMBO. **Hugo**, my Eclipse partner, this journey wouldn't have been the same without you. You weren't just the one in charge of the Eclipse machine—you were always there to help me when things didn't go as planned. I could always count on you to make everything run smoothly, and I'm so grateful for that. It was a pleasure to share both the lab and the ASMS experience with you. I'm sure you'll do amazing things and finish your PhD with the same dedication you've shown throughout.

Lastly, **Sarahi**, muchas gracias por todo. Tu presencia fue esencial en estos años, no solo en el laboratorio, sino en mi vida personal. Siempre lista para escucharme o para tomarnos un café y despejarnos juntas. No sé cómo hubiera sobrellevado todo sin ti. Te extrañaré mucho y siempre estaré agradecida por haber compartido este tiempo contigo. Espero que sigamos conectadas, aunque estemos lejos.

A special thank you also goes to **Zinedine**, my now colleague and friend. It's been such a pleasure working with you on the HCP project and teaching you everything I could. But more than just being a project partner, you became someone I could always rely on, through both the highs and the lows. We became closer not just through our work, but through all the challenges and victories we shared. Your kindness, your constant willingness to help, and your calm demeanor always helped keep me grounded. I'm so glad you came back, and I have no doubt you have a bright future ahead of you.

A heartfelt thank you to **Alexandre**. Your help and support, especially in my last few months, meant more to me than I can express. I'm so glad we got to know each other better during the ASMS. You truly made the difficult times easier to handle. I didn't even have to say much because you always seemed to know when I needed someone to lean on. Your friendship and warm-heartedness made all the difference during those tough moments when I felt a little lost. I know you will succeed in your career and PhD because you truly deserve all the success that comes your way. Thanks for being such a solid friend and someone I could always count on.

I would like to extend a very special thanks to the Merck Serono Team: **Roberta, Elisa, Maura, Federica, Cristina, and Martina,** for welcoming me with open arms and allowing me to work on the premises for my project. A warm thanks to **Elisa** and **Maura** for teaching me so many things and for being so kind and patient with me. Federica, I truly couldn't have asked for a better mentor. Thank you for dedicating so much of your time to me and for being more than just a tutor; you were a friend when I needed one. I'll always cherish the memories we created together, and I'm deeply grateful for your kindness and the care with which you taught me. Cristina and Martina, thank you both for your friendship and

for making me feel like part of the team from the very start. I'm so grateful for the moments we shared, the laughs, and the great memories.

Finalmente, quiero dedicar un agradecimiento profundo y sincero a mi familia: **mi mami, mi pa, mi hermana Dani, Alfredo, Marujita, y Nico. Mami y Pa**, gracias por siempre confiar en mí y por estar siempre presentes, incluso a la distancia. Este logro es tan de ustedes como mío, porque sin su apoyo incondicional y el ejemplo que me han dado, no habría llegado hasta aquí. Comencé este doctorado con una enorme motivación, lista para enfrentarme a todo, pero cuando poco a poco esa motivación se fue apagando por diversos motivos, ustedes estuvieron ahí para ayudarme a encontrarla de nuevo, para recordarme por qué comencé este camino. Su apoyo constante me dio las fuerzas necesarias para seguir adelante, aun cuando sentía que todo se volvía más difícil.

Dani, hermana, gracias por siempre cuidarme, defenderme, y ser como una segunda mamá para mí. A pesar de que cada una de nosotras ha tomado su propio camino, siempre me siento acompañada por ti, seguras y unidas por el cariño tan grande que nos vincula. Tu apoyo, tus consejos, y ese amor inmenso que compartimos me han dado el impulso para continuar, incluso cuando quise desistir. **Alfredo**, gracias por siempre hacerme reír y por acogerme en su casa cuando no me sentía bien. Formamos un gran equipo los tres juntos, y siempre he sentido tu apoyo cercano.

Marujita, abuela, tus atenciones, tus mensajitos y tus cariños han hecho que mi estancia fuera de casa mucho más amena y llevadera. Aunque la distancia nos separe, tu presencia y amor siempre me han acompañado en estos años. **Nico**, aunque ya no estés físicamente conmigo, te siento en cada paso que doy, igual que el resto de los miembros de la familia. Sé que me has acompañado en los momentos más difíciles, y espero que estés orgulloso de lo que he logrado. Te extraño y siempre te llevo en mi corazón.

Table of contents

List of abbreviations

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
AIF	All-ion fragmentation
AQUA	Absolute quantification
BSA	Bovine serum albumin
CAPF	Protein A affinity Chromatography
CCS	Collision cross section
CE	Collision energy
CHO	Chinese hamster ovary
CID	Collision induced dissociation
CLA	Clarified Harvest
CV	Coefficient of variation
Da	Dalton
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DS	Drug substance
DSP	Downstream process
ELISA	Enzyme-linked immunosorbent assays
EMA	European medicines agency
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
iTRAQ	Isobaric tags for relative and absolute quantification

LC	Liquid chromatography
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
mAb	Monoclonal antibody
MALDI	Matrix-assisted laser desorption
MM	Mixed-mode chromatography
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MW	Molecular weight
m/z	Mass-to-charge ratio
NAF	Nano filtration
NeuCode	Neutron encoding
Ng/mg mAb	Nanogram of host cell proteins per milligram of monoclonal antibody
PAcIFIC	Precursor acquisition independent from ion count
PAI	Protein abundance index
pI	Isoelectric point
PPA	Post protein A affinity chromatography
PRM	Paralleled Reaction Monitoring
PSAQ	Protein standard absolute quantification
PSM	Peptide spectrum match
PTM	Post-translational modification
Q	Quadrupole analyser
QC	Quality control
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
SPTFF	Single-pass tangential flow filtration
SRM	Selected Reaction Monitoring
SWATH	Sequencial windowed acquisition of all theoretical fragment ion spectra
TMAE	Anionic exchange chromatography
TOF	Time-of-flight analyser
TPR	True positive rate
UPLC	Ultra performance liquid chromatography
USP	Upstream process
XDIA	Extended data-independent acquisition
XIC-MS1	Extracted ion chromatogram – MS1 filtering

Résumé en Français

Part I

Introduction

Les anticorps monoclonaux (mAbs) sont des produits biothérapeutiques largement utilisés, en effet, plus de 120 anticorps monoclonaux ont été mis sur le marché au cours des 20 dernières années pour traiter diverses pathologies. Ils sont fabriqués par des technologies recombinantes dans des cellules hôtes, principalement les cellules d'ovaires de hamster chinois (CHO)[1]. Ces cellules sont devenues un standard dans l'industrie biopharmaceutique en raison de leur efficacité à produire des mAbs de haute qualité. Toutefois, durant le processus de production, ces cellules hôtes introduisent des impuretés protéiques indésirables, appelées Host Cell Proteins (HCPs), qui doivent être éliminées afin de garantir la pureté et la sécurité des médicaments. Leur présence pourrait induire des conséquences graves sur les patients ainsi qu'une dégradation du médicament[2]. Par conséquent, il est crucial d'identifier et de quantifier les HCPs de manière précise et avec la meilleure sensibilité possible tout au long du processus de purification, détaillé dans la Figure 1.

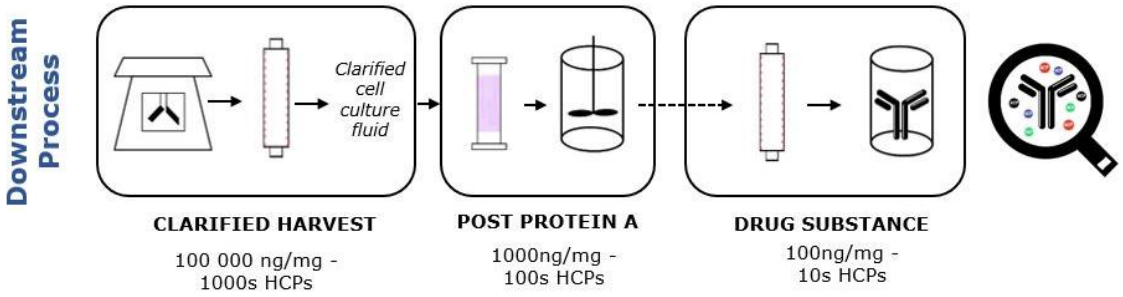


Figure 1: Élimination des HCPs durant le processus de purification. Illustration de la réduction en HCP à différentes étapes de la production des anticorps monoclonaux, tant en termes de quantité totale (ng de HCP par mg de mAb) que du nombre d'HCP individuels..

Traditionnellement, des tests ELISA utilisant des anticorps polyclonaux anti-HCPs sont utilisés pour quantifier le niveau total de HCPs dans les échantillons issus des différentes étapes de purification.[3] Bien que l'ELISA soit une méthode de référence, offrant une large gamme de détection et une haute sensibilité, elle présente des limitations très importantes. En effet, ce test repose sur la spécificité des anticorps et la couverture du protéome ciblé et est donc limitée par les interactions antigène-anticorps, pouvant affecter la précision des résultats. De plus, cette méthode ne permet pas l'identification spécifique des HCPs, limitant sa capacité à fournir une évaluation complète des risques.

Ces insuffisances ont conduit à l'émergence de nouvelles stratégies alternatives, parmi lesquelles la protéomique quantitative, fondée sur la combinaison de la chromatographie liquide et de la spectrométrie de masse en tandem (LC-MS/MS), s'impose comme une méthode privilégiée. La spectrométrie de masse (MS) permet une caractérisation exhaustive et sans biais des HCP détectables.[2]

Les avancées technologiques et bioinformatiques en protéomique ont permis le développement de multiples stratégies pour l'identification et la quantification des peptides et protéines (Figure 2). Le mode Data Dependent Acquisition (DDA) fournit une vue quantitative d'ensemble des peptides détectés, en isolant et en fragmentant les plus abondants. De ce fait, cette approche présente ses limites en termes de répétabilité et de gamme dynamique dans le cadre de l'analyse des échantillons complexes tel que les échantillons issus des étapes de purification du mAb. Ces contraintes ont conduit à la mise au point d'un nouveau mode d'acquisition, le mode Data Independent Acquisition (DIA). Cette approche repose sur la co-isolation et la co-fragmentation simultanée de l'ensemble des peptides présents dans des fenêtres d'isolation prédéfinies, couvrant l'ensemble de la gamme de masse. La quantification en DIA est réalisée à partir des spectres MS2, lui offrant une meilleure spécificité, une plus grande reproductibilité ainsi qu'une sensibilité et une précision accrues, faisant de la DIA une méthode de choix pour l'analyse des HCP présent à l'état de traces.

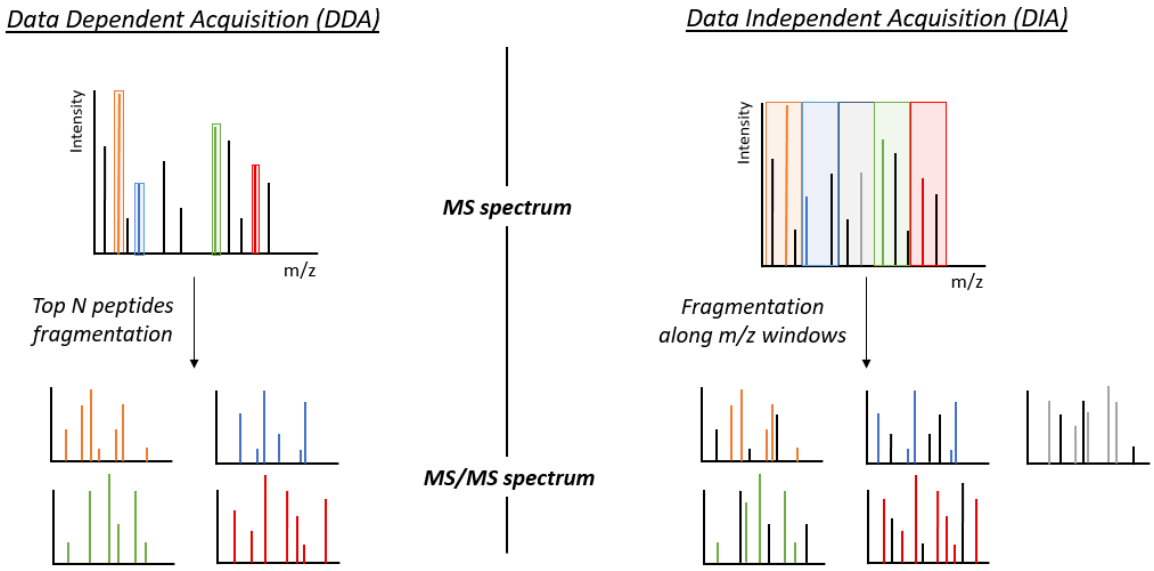


Figure 2: Comparaison des deux grandes modes d'acquisition utilisés lors d'une analyse par spectrométrie de masse.

Dans ce contexte, au cours des deux premières années de ma thèse, mes travaux se sont focalisés sur le développement d'une méthode d'immuno-capture couplée à la spectrométrie de masse ayant pour objectif de permettre d'évaluer/valider les résultats de tests ELISA. La nécessité de développer un test ELISA spécifique à chaque molécule thérapeutique en développement et à quel stade reste une importante question ouverte. Ainsi, disposer d'une méthode performante d'évaluation de tests ELISA génériques constitue un enjeu majeur pour l'industrie biopharmaceutique dans leur stratégie de développement de nouveaux produits.

Pendant la troisième année de thèse, des développements et optimisations sur le couplage LC-MS/MS ont été réalisés afin d'améliorer la couverture et la sensibilité ainsi que la précision et la justesse de quantification des impuretés protéiques de type HCPs dans des échantillons collectés à différentes étapes du process de purification dans le but de participer à l'amélioration des process.

Part II
Développement d'une méthode de validation des tests ELISA pour la détection des impuretés protéiques de type HCP

- **Processus d'optimisation des conditions d'immuno-capture**

L'identification et la quantification précise des HCPs dans les processus biopharmaceutiques sont cruciales pour garantir l'efficacité et la sécurité des médicaments. Malgré l'utilisation extensive des tests ELISA, certaines HCPs non immunogènes comportant tout de même des risques, restent invisibles et non détectées par ELISA[4]. L'objectif principal de mes premiers travaux a été de développer une méthode utilisant l'immunocapture et la spectrométrie de masse (MS) permettant d'évaluer la couverture de l'ELISA. La démarche utilisée est illustrée dans la Figure 3.

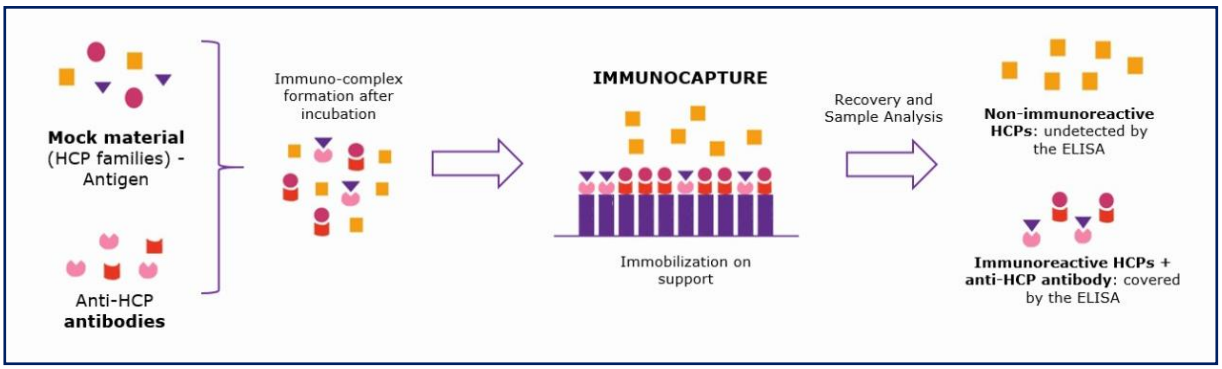


Figure 3: Processus d'immunocapture et échantillons obtenus.

Le processus d'immunocapture a été optimisé sur une période de neuf mois passés au sein de l'entreprise Merck (site de Guidonia en Italie), au cours de laquelle plusieurs paramètres ont été systématiquement évalués. Trois supports différents greffés avec de la protéine A ont été testés dans le but de capturer les immunocomplexes HCP/anticorps anti-HCP utilisées dans le kits ELISA: des billes magnétiques, des plaques, et des colonnes. L'optimisation des méthodes des différents supports s'est révélée complexe et a nécessité de nombreux ajustements, notamment en termes de quantité de matériel, de température et de temps et d'ordre successif ou simultané d'incubation. Finalement, après des mois d'essais, seul le support colonne a offert des résultats satisfaisants, permettant une immunocapture suffisamment efficace des HCPs et nous avons donc poursuivi avec ce support.

Le processus d'optimisation de l'immunocapture a été évalué chez Merck par SDS-PAGE, méthode choisie pour sa rapidité à fournir des résultats visuels immédiats. Cette approche nous a permis d'observer directement sur les gels si les antigènes avaient été capturés par les anticorps anti-HCP, grâce à la visualisation des bandes protéiques dans les fractions d'élution (Figure 4). En utilisant cette méthode, nous avons pu affiner les paramètres clés, tels que les ratios anticorps/antigènes, la température et le temps d'incubation. Une fois ces paramètres définis de manière satisfaisante, nous avons décidé de passer à une méthode plus précise pour évaluer la capture des HCPs : l'électrophorèse

différentielle bidimensionnelle (2D DIGE) encore largement utilisée en milieu industriel biopharmaceutique pour soutenir les demandes d'autorisation de mise sur le marché de nouveaux médicaments (FDA et EMA).[5]

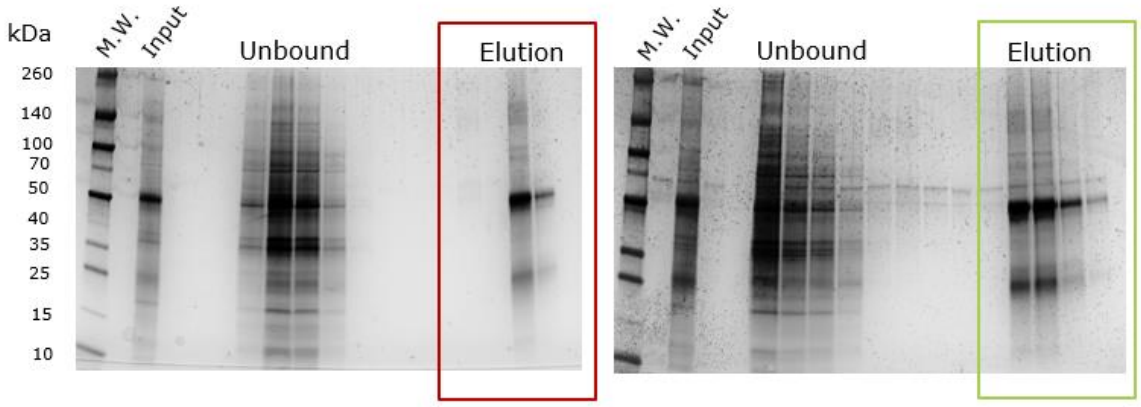


Figure 4 : Caractérisation par SDS-PAGE de l'efficacité de l'immuno-capture. La population d'HCPs présente dans l'élution révélée ici sert d'indicateur d'efficacité. Cette dernière est plus abondante dans le gel de droite, les conditions d'immuno-capture semble plus adaptées.

La technique 2D DIGE a permis une analyse plus approfondie de l'efficacité de l'immunocapture en séparant les protéines selon leur point isoélectrique et leur poids moléculaire. L'objectif était de mesurer le chevauchement entre l'échantillon Mock (antigènes) et la fraction unbound, représentant les HCPs non détectés par le pool d'anticorps du test ELISA. Un chevauchement réduit indique une capture efficace des HCPs, tandis qu'un chevauchement important représente une capture insuffisante.

Toutefois, l'analyse de la fraction d'élution a posé des difficultés, notamment en raison de la saturation du signal causée par une forte concentration d'anticorps anti-HCP, empêchant la visualisation des HCPs liés. Malgré cette limitation, la 2D DIGE a confirmé que le ratio anticorps-antigène de 1:1 offrait les meilleures performances, avec un chevauchement minimal entre Mock et unbound, indiquant une capture optimisée des HCPs, comme observé dans la Figure 5.

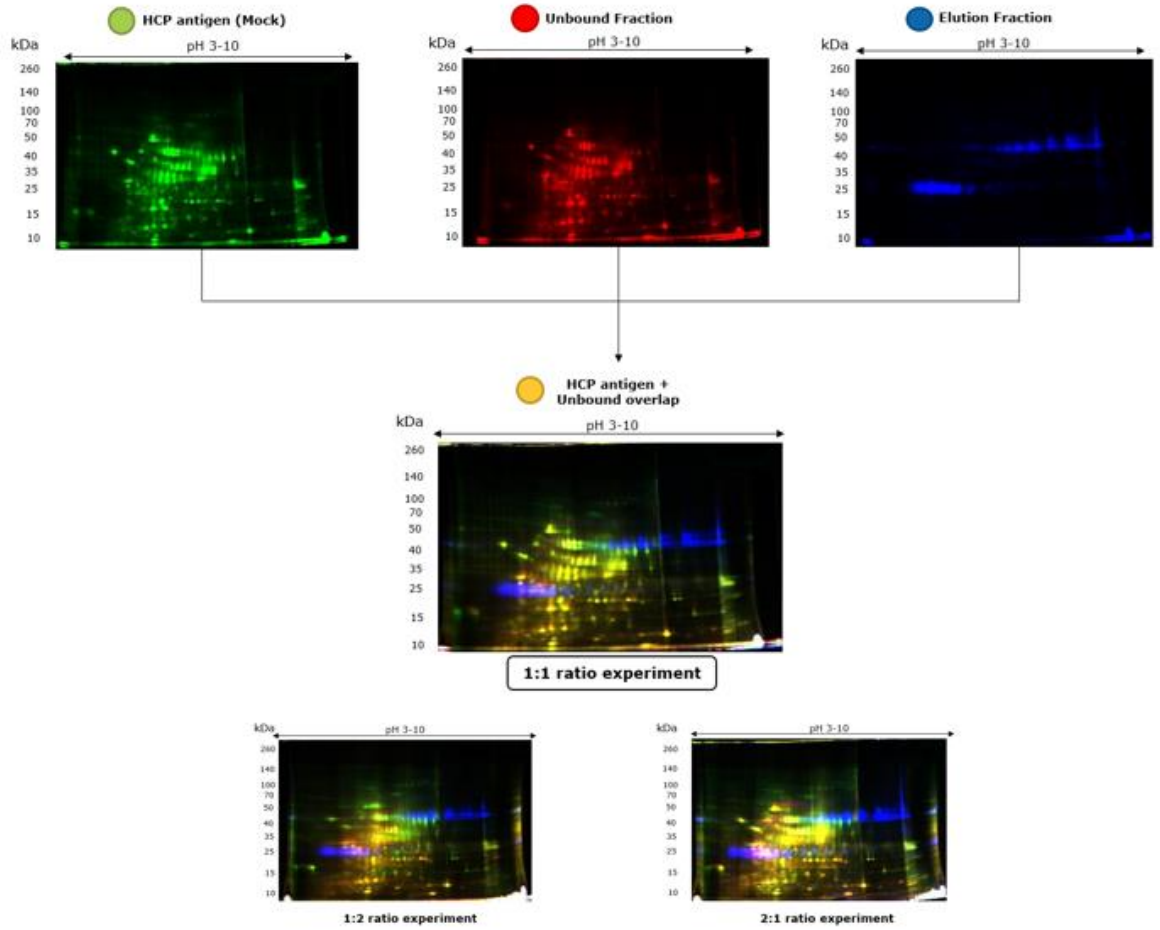


Figure 5 : Résultats 2D DIGE de l'optimisation du ratio avec une colonne de Protéine A, montrant le chevauchement entre les échantillons Mock et Unbound pour les expériences avec des ratios 1:1, 1:2, et 2:1.

Après avoir déterminé le ratio optimal, nous avons cherché à accélérer le processus de capture en augmentant la température de pré-incubation des complexes immunitaires. Ainsi, des températures d'incubation à température ambiante ou à 37°C pendant 2 heures ont été testées, car une température plus élevée est connue pour réduire le temps d'incubation requis. De plus, pour vérifier l'efficacité de l'optimisation, un contrôle positif a été réalisé en utilisant un couple commercial d'antigènes et d'anticorps anti-HCP (couple Cygnus), le plus communément employé en début des processus de développement de mAbs dans l'industrie pharmaceutique.

Les chromatogrammes dans la Figure 6 montrent clairement que le couple spécifique au processus dépasse le couple commercial, avec une fraction non liée beaucoup moins abondante, indiquant ainsi une meilleure efficacité du processus d'immuno-capture. La performance supérieure du couple antigènes/anticorps spécifiques au processus souligne l'importance de développer et d'utiliser des méthodes adaptées à chaque produit afin d'obtenir des résultats plus fiables par ELISA. Ce constat a des répercussions financières et temporelles importantes pour l'industrie du développement d'anticorps thérapeutiques mais est bien réel.

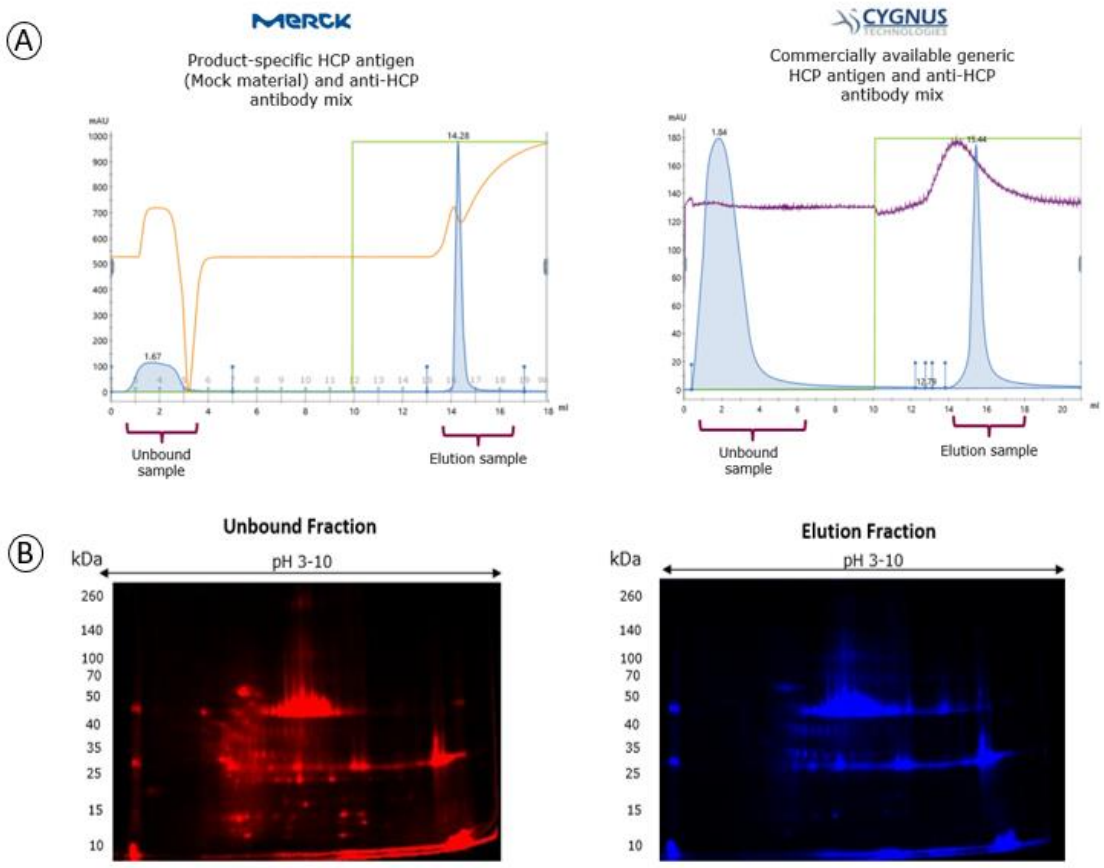


Figure 6 : Contrôle positif de la colonne protéine A. A) Comparaison des chromatogrammes entre le couple spécifique et le couple commercial**.** B) Mise en évidence d'une forte présence de l'anticorps anti-HCP dans la fraction Unbound et Elution pour le couple commercial.

Après avoir optimisé l'ensemble des paramètres de l'immunocapture et vérifié son efficacité par électrophorèse, des séries d'échantillons ont été préparées dans le but d'en faire une analyse beaucoup plus fine par MS à mon retour à Strasbourg.

L'objectif des méthodes de MS est de pouvoir identifier très précisément les HCPs non reconnues par les anticorps anti-HCP (ce qui n'est pas possible par un test ELISA), en plus d'affiner l'évaluation des conditions les plus efficaces pour la capture. Pour ce faire, nous avons construit une librairie spectrale la plus exhaustive possible à partir des échantillons Mock (correspondant à l'ensemble des HCPs possiblement présents), en employant deux méthodes distinctes. D'une part, nous avons effectué un fractionnement sur gel avant analyses LC-MS/MS, et d'autre part, nous avons exploré une alternative utilisant une méthode innovante de fractionnement en phase gazeuse (Gas Phase Fractionation GPF). Cette dernière consiste à injecter consécutivement l'échantillon d'antigènes, après digestion sur gel, en mode Data Independent Acquisition (DIA) sur des gammes de m/z consécutives (Figure 7).[6]

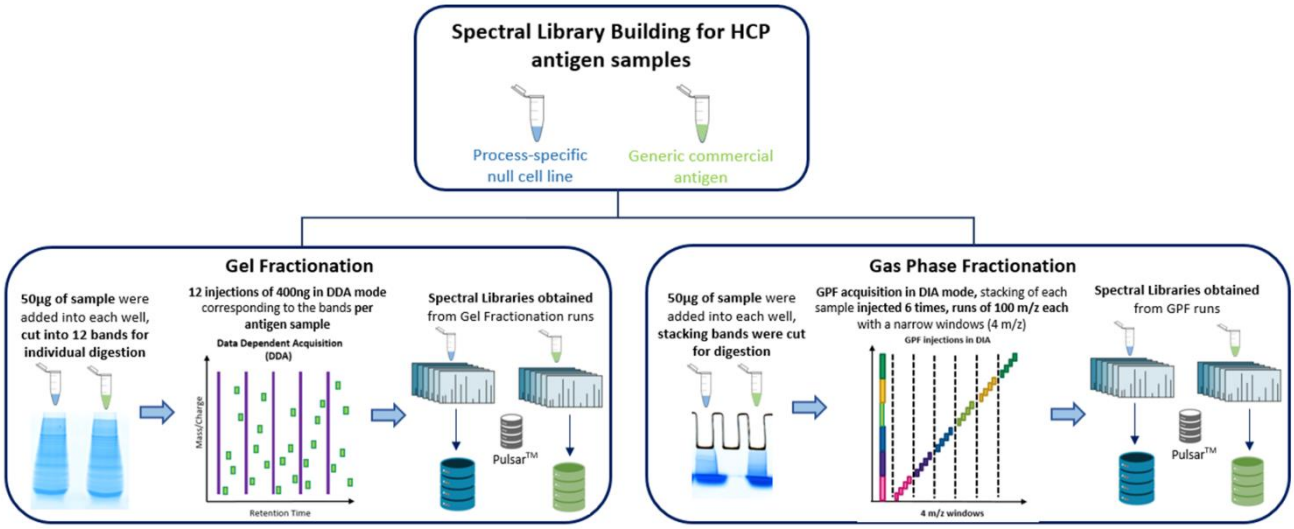


Figure 7 : Workflows utilisés pour la génération des librairies spectrales.

Les résultats ont montré que la GPF est une méthode robuste, comparable au fractionnement sur gel, les deux ayant permis d'identifier un nombre comparable de protéines. Afin d'enrichir les identifications, nous avons décidé de fusionner les deux librairies et d'utiliser cette nouvelle et plus exhaustive librairie spectrale pour analyser l'intégralité des échantillons obtenus à partir de l'immunocapture sur colonne.

Cette stratégie a confirmé que l'incubation à 37°C pendant 2 heures avec un ratio anticorps/antigènes de 1:1 constituait les conditions optimales, offrant les meilleures performances avec le plus faible nombre d'HCPs spécifiques dans la fraction non liée. Plus précisément, seulement 32 HCPs ont été détectées dans cette fraction, en contraste avec le nombre habituel, qui se situe généralement dans les centaines. Par ailleurs, nous avons pu identifier de manière exhaustive les HCPs non capturés par les anticorps anti-HCP et donc non détectés par l'ELISA, comblant ainsi les lacunes dans l'identification des HCPs jusqu'alors inconnus.

- **Implémentation de l'utilisation des PhyTips Protein A**

Dans une nouvelle étape, afin d'évaluer la performance et la reproductibilité des méthodes, nous avons exploré deux approches d'immunocapture avec les pointes de pipettes PhyTips Protein A. Ce support présente l'avantage majeur de la facilité de manipulation, ne nécessitant pas l'utilisation d'un système chromatographique et peut être aisément automatisé. Des expériences réalisées en triplicat nous ont permis d'évaluer l'efficacité d'une approche directe versus indirecte. Dans l'approche directe, les anticorps anti-HCP sont immobilisés directement sur le support de Protéine A avant l'ajout des antigènes, sans étape préalable d'incubation. À l'inverse, l'approche indirecte implique une incubation initiale des anticorps anti-HCP avec les antigènes pour former un complexe immunitaire, qui est ensuite capturé par la Protéine A.

Les résultats ont montré que l'approche indirecte était plus efficace, capturant environ 300 µg de protéines, tandis que l'approche directe présentait une capture plus limitée, probablement en raison des changements conformationnels des anticorps lorsqu'ils sont fixés directement à la Protéine A. De

plus, les trois réplicats de chaque méthode ont montré une bonne cohérence, confirmant ainsi la répétabilité de l'approche indirecte et sa supériorité en termes d'efficacité et de robustesse, comme observé dans la Figure 8. Ces observations soulignent l'intérêt d'utiliser cette méthode pour optimiser l'immunocapture grâce à un workflow simplifié et reproductible, ouvrant ainsi des perspectives prometteuses pour son adoption dans un environnement biopharmaceutique industriel.

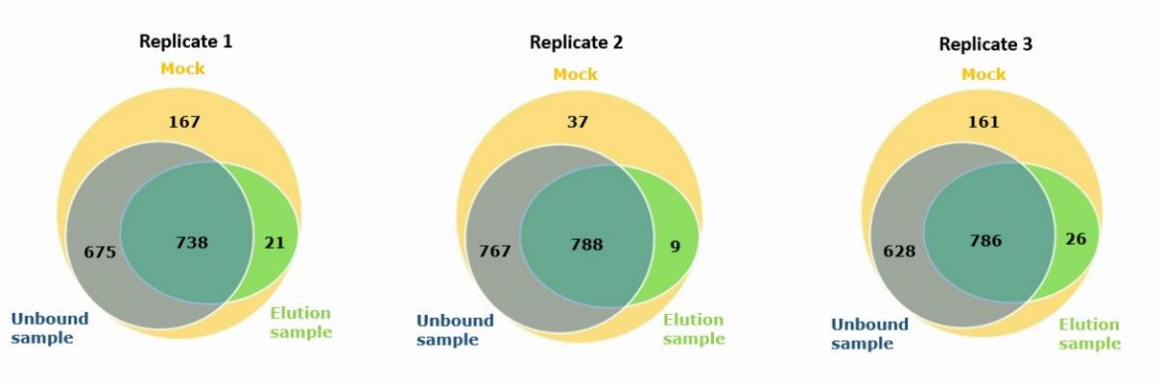


Figure 8 : Diagrammes de Venn montrant le chevauchement des groupes de protéines identifiées dans les échantillons Mock (antigènes), unbound et élution pour l'approche indirecte utilisant les pointes de pipette PhyTips.

Par ailleurs, nous avons également mené une analyse des HCP à haut risque sur les échantillons obtenus avec l'approche indirecte en Phytips, en se concentrant sur une liste de 29 protéines HCP jugées les plus critiques[7]. L'objectif était de déterminer si ces protéines étaient systématiquement présentes dans les différentes fractions des réplicats. Les résultats ont révélé qu'une protéine spécifique échappait systématiquement à la détection par les anticorps anti-HCP, et donc par la méthode ELISA. Cette observation met en lumière les limitations des méthodes classiques et souligne la nécessité d'utiliser une nouvelle méthode alternative comme la nôtre pour identifier ces protéines non détectées par les ELISA de routine.

Part III

Développement de méthodes LC-MS/MS pour une caractérisation fine et une quantification précise des impuretés protéiques de type HCP

Mon second objectif de thèse a été de développer un workflow LC-MS/MS permettant de renforcer la couverture et la sensibilité analytiques pour une quantification exhaustive et précise des HCPs dans des échantillons allant des récoltes brutes aux produits finaux. Ce travail s'inscrit dans la continuité de recherches menées lors de deux thèses précédentes au LSMBO, qui visaient à affiner les étapes du workflow de la préparation des échantillons jusqu'au traitement bioinformatique des données.[8,9]

Dans ce contexte, j'ai optimisé des méthodes LC-MS/MS incorporant des peptides standards dans le but d'améliorer la précision et la justesse de la quantification des HCPs. D'une part, j'ai utilisé un mélange de 54 peptides standards couvrant 6 points d'une large gamme de concentrations afin de réaliser une courbe de calibration interne (ReadyBeads ANAQUANT), permettant une quantification de l'ensemble des HCPs présentes dans les échantillons prélevés à différentes étapes du processus de purification. D'autre part, j'ai implémenté un mélange de 70 peptides marqués isotopiquement (standards AQUA), synthétisés et hautement purifiés pour la quantification ciblée et absolue de 29 HCPs à haut risque.[7] Cette double approche renforce la précision et la justesse de la quantification des HCPs jusqu'au niveau sub-ppm et améliore ainsi l'évaluation des risques tout au long du processus de purification (Figure 9).

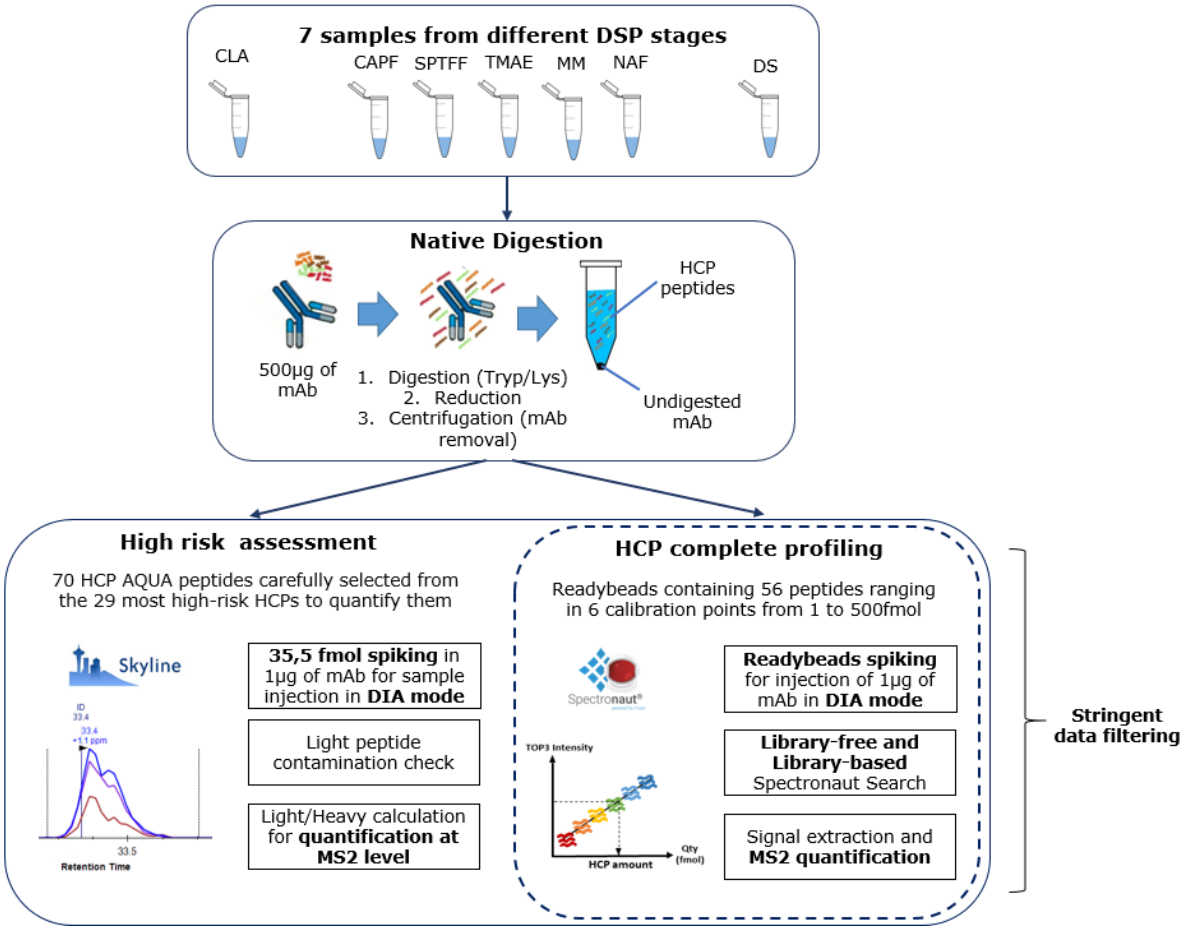


Figure 9 : Workflow utilisé pour la quantification globale des HCPs et l'évaluation des HCPs à haut risque.

- **Quantification globale des HCPs dans les échantillons issus des étapes de purification du mAb**

J'ai réalisé mes optimisations de méthodes sur une série de sept échantillons de l'anticorps monoclonal Avelumab, commercialisé sous le nom de Bavencio, correspondant à diverses étapes du processus de purification et fournis par Merck.

Ces échantillons incluent, dans l'ordre, le Harvest (CLA), la chromatographie d'affinité à la protéine A (CAPF), qui capture le mAb tout en éliminant les impuretés, l'étape d'inactivation virale à faible pH, la filtration tangentielle à passage unique (SPTFF), la chromatographie d'échange d'anions (TMAE), la chromatographie en mode mixte (MM) pour une réduction supplémentaire des impuretés, et enfin la nanofiltration (NAF) pour l'élimination des virus restants, menant à la substance finale (DS). Un échantillon de la lignée cellulaire nulle (Mock) a également été fourni, servant à la génération et à l'optimisation de la librairie spectrale, comme précédemment décrit.

Pour ces analyses, nous avons employé un protocole de digestion native adapté.[8] Les échantillons ont ensuite été enrichis avec les standards Readybeads, et les courbes d'étalonnage obtenues durant les sept étapes du processus ont toutes montré un coefficient de corrélation supérieur à 0,7, indiquant

des résultats prometteurs pour la quantification, malgré la complexité du milieu et l'augmentation de la gamme dynamique tout au long du processus de purification. Ces courbes ont montré une reproductibilité satisfaisante entre réplicas, démontrant la robustesse de notre approche pour cette application spécifique.

Qualitativement, une réduction significative du nombre de HCPs a été observée tout au long du processus de purification, comme illustré en Figure 10. Bien que cette tendance ait été attendue, elle a été confirmée ici par les résultats obtenus via la spectrométrie de masse. La réduction la plus marquée s'est produite entre les étapes SPTFF et TMAE, où le nombre de HCP identifiés a drastiquement chuté de 1143 à 23. J'ai comparé deux stratégies de traitement des données distinctes, avec et sans bibliothèques spectrales, et j'ai pu conclure que l'approche avec bibliothèque spectrale avait un impact limité mais négatif sur le nombre de protéines identifiées, ce avec ou sans filtre sur les coefficients de variation, entraînant une réduction modeste des nombres de groupes de protéines identifiés, en particulier aux étapes avancées de la purification.

Figure 10 : Histogramme représentant les nombres de protéines identifiées et quantifiées avec application d'un filtre CV, ainsi que l'utilisation ou non d'une librairie spectrale.

Quantitativement, comme illustré en Figure 11, une réduction substantielle de la concentration en HCP est observée tout au long des étapes de purification, exprimée en parties par million (ppm) par rapport au contenu en mAb, conformément aux exigences réglementaires.[10] Cette diminution est particulièrement notable entre les étapes CLA et CAPF, où plus de 60 % de la quantité totale des HCP est efficacement éliminée, marquant une purification significative. Une autre étape critique est observée entre la SPTFF et la chromatographie TMAE, révélant une réduction en quantité d'HCPs totaux d'un facteur 1000.

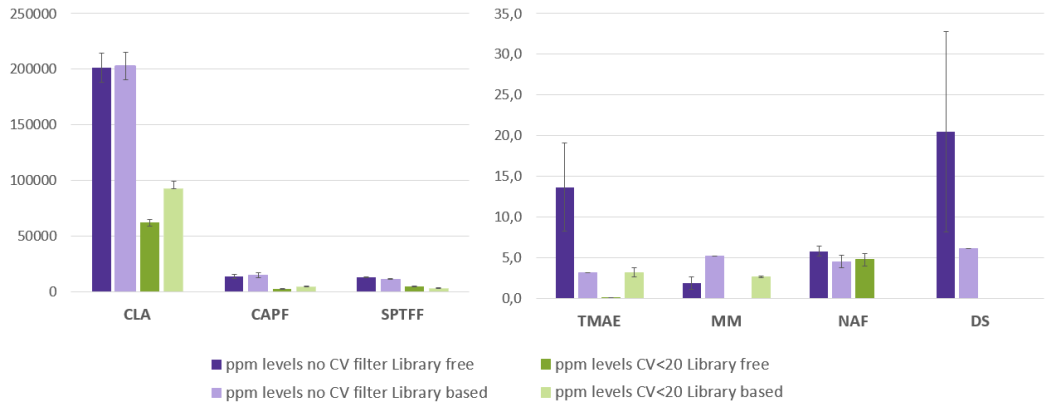


Figure 11 : Histogramme illustrant les niveaux de ppm dans les échantillons tout au long du processus de purification, de CLA à DS, avec ou sans application du filtre CV.

Bien que l'approche reposant sur une librairie spectrale offre un léger avantage dans les premières étapes du processus, où la concentration en HCP est plus élevée, l'approche sans librairie s'avère particulièrement efficace pour détecter les HCPs en faible abondance dans les phases de purification finales. Cette observation indique que la méthode sans librairie peut fournir une sensibilité accrue, malgré un nombre réduit de protéines identifiées, pour les HCPs à faibles abondances en fin de purification. Il est donc essentiel d'adapter les méthodes analytiques aux exigences spécifiques de chaque étape pour garantir une détection et une quantification précises des impuretés.

- **Quantification des HCPs à haut risque dans les échantillons issus des étapes de purification du mAb**

En plus des quantifications globales des HCPs réalisées grâce à l'approche Readybeads, nous avons souhaité évaluer la valeur ajoutée d'une quantification plus précise, voire absolue, d'une liste finie de HCP à haut risque pré-sélectionnés. Nous avons ainsi comparé, d'une part, les résultats de quantification obtenus avec les Readybeads aux résultats obtenus en mode ciblé sur 29 HCP à haut risque en utilisant 70 peptides AQUA marqués.

L'application de ces deux méthodes de quantification a permis de dresser un aperçu complet des HCP à haut risque tout au long de la chaîne de purification. Les résultats montrent que les Readybeads ont quantifié 4 HCP à haut risque, dont un détecté exclusivement par cette méthode, tandis que le mélange de peptides AQUA en a quantifié 12, avec 9 HCP détectés uniquement par cette méthode. Bien que complémentaires, ces résultats indiquent qu'il existe une réelle valeur ajoutée à utiliser des peptides AQUA spécifiques pour les HCP d'intérêt particulier. En effet, le mélange de peptides AQUA, en particulier, se distingue avec des niveaux de ppm plus élevés et donc une sensibilité accrue, notamment dans les étapes finales de purification, le rendant plus adapté à la détection des HCP à faible abondance dans ce contexte. Trois HCPs ont été communément quantifiés par ces deux approches, les résultats sont présentés dans la Figure 12.

Par ailleurs, les deux méthodes démontrent que tous les HCP à haut risque ont été éliminés lors de l'étape de chromatographie TMAE, soulignant son importance dans le processus de purification. Bien

que les Readybeads fournissent une quantification globale robuste, leur sensibilité limitée à détecter certains HCP critiques identifiés uniquement par le mélange AQUA suggère qu'une approche combinée permettrait une évaluation plus complète. En ciblant spécifiquement les HCP à haut risque, le mélange AQUA offre des informations détaillées sur leur réduction progressive au fil des étapes de purification, facilitant ainsi une gestion optimale des risques liés à ces impuretés.

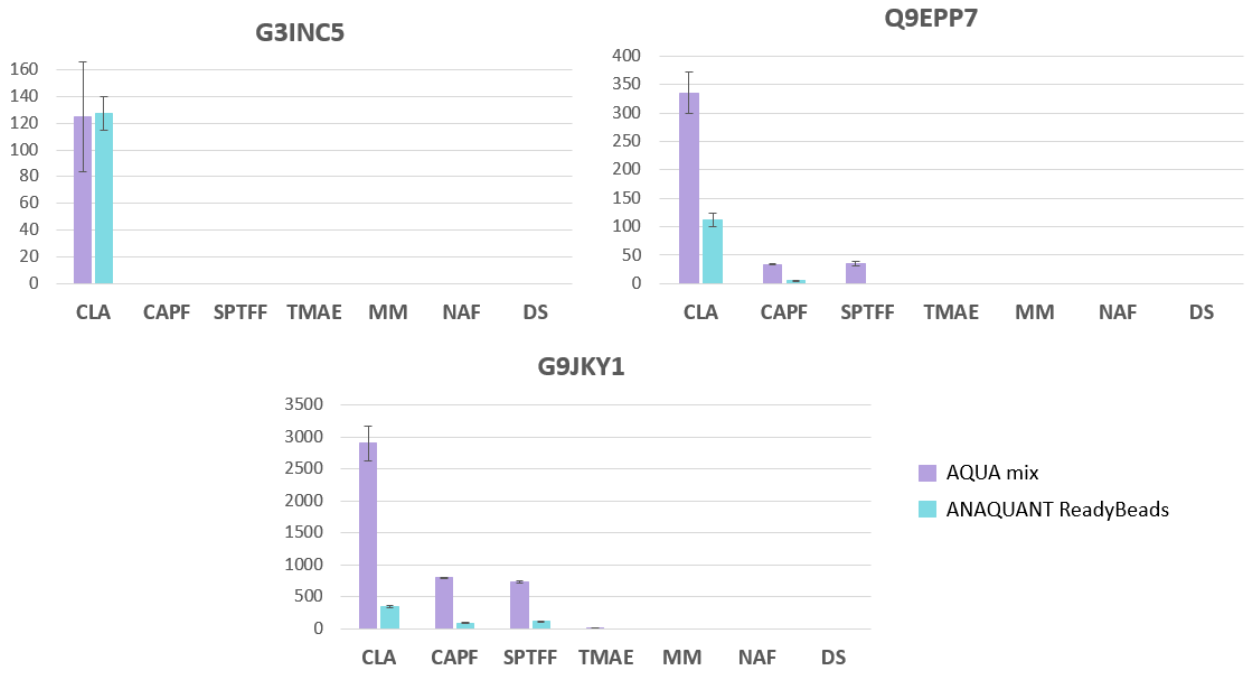


Figure 12: Quantification de trois HCPs à haut risque communément quantifiées dans les échantillons issus des étapes de purification du mAb par l'utilisation des méthodes AQUA et ANAQUANT. Cette figure présente les niveaux en ppm des trois HCPs à haut risques quantifiées. La baisse d'abondance de ces HCPs est clairement démontrée à chaque étape de la purification.

Conclusion

Au cours de ce projet de doctorat, nous avons développé avec succès une méthode d'immuno-capture conçue pour évaluer la performance des tests ELISA génériques et spécifiques pour la détection des impuretés HCP. En comparant divers supports d'immuno-capture, incluant les colonnes de Protéine A, les billes et les plaques, et en transposant finalement cette méthode à des pointes de pipettes PhyTips, facilement automatisables, et en employant une approche indirecte pour former l'immunocomplexe avant sa liaison correspondante à la Protéine A, nous avons identifié des conditions optimales pour capturer ces impuretés.

Nos optimisations ont mis en lumière les fortes limitations des tests ELISA génériques commercialisés, fréquemment utilisés aux premières étapes du développement des anticorps monoclonaux (mAb). En effet, nos résultats soulignent l'importance de développer un ELISA spécifique pour chaque produit, car un mélange d'anticorps anti-HCP spécifique au processus améliore considérablement la précision de la détection. De plus, contrairement aux méthodes ELISA standard, notre technique d'immunocapture garantit que les HCP potentiellement problématiques ne sont pas ignorées, renforçant ainsi la fiabilité des mesures de contrôle qualité et assurant la sécurité et l'efficacité des produits biopharmaceutiques.

Pour atteindre le deuxième objectif de ma thèse, l'incorporation de la spectrométrie de masse en tant que méthode analytique orthogonale a permis une compréhension plus approfondie de la distribution des HCP tout au long du processus de purification. La comparaison entre une quantification globale (Readybeads) et ciblée (mélange de peptides AQUA) nous a permis de démontrer une complémentarité entre les deux méthodes. Tandis que les Readybeads facilitent la quantification globale de l'ensemble des HCP présentes, le mélange de peptides AQUA, conçu spécifiquement pour cibler les HCP à haut risque, a montré une sensibilité accrue dans les étapes finales de purification.

Cette approche double a confirmé que des étapes critiques, en particulier la chromatographie TMAE, éliminaient efficacement les derniers HCP à haut risque pour une purification complète. En conclusion, ce projet met en avant les avantages de la spectrométrie de masse par rapport aux méthodes ELISA traditionnelles pour l'analyse des impuretés HCP, offrant non seulement une sensibilité accrue mais aussi une évaluation plus complète de ces contaminants biopharmaceutiques.

L'intégration de techniques analytiques avancées et la comparaison de diverses méthodes de quantification fournissent des informations essentielles sur l'efficacité des processus de purification, contribuant ainsi à la production de produits biothérapeutiques plus sûrs et de meilleure qualité.

Références

(1) Alejandra, W.-P.; Miriam Irene, J.-P.; Fabio Antonio, G.-S.; Patricia, R.-G. R.; Elizabeth, T.-A.; Aleman-Aguilar, J. P.; Rebeca, G.-V. Production of Monoclonal Antibodies for Therapeutic Purposes: A Review. *International Immunopharmacology* **2023**, *120*, 110376. https://doi.org/10.1016/j.intimp.2023.110376.

(2) Yang, F.; Li, D.; Kufer, R.; Cadang, L.; Zhang, J.; Dai, L.; Guo, J.; Wohlrab, S.; Greenwood-Goodwin, M.; Shen, A.; Duan, D.; Li, H.; Yuk, I. H. Versatile LC–MS-Based Workflow with Robust 0.1 Ppm Sensitivity for Identifying Residual HCPs in Biotherapeutic Products. *Anal. Chem.* **2022**, *94* (2), 723–731. https://doi.org/10.1021/acs.analchem.1c03095.

(3) Shukla, A. A.; Jiang, C.; Ma, J.; Rubacha, M.; Flansburg, L.; Lee, S. S. Demonstration of Robust Host Cell Protein Clearance in Biopharmaceutical Downstream Processes. *Biotechnology Progress* **2008**, *24* (3), 615–622. https://doi.org/10.1021/bp070396j.

(4) 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. *Biotech & Bioengineering* **2014**, *111* (12), 2367–2379. https://doi.org/10.1002/bit.25327.

(5) 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. *Biotech & Bioengineering* **2010**, *105* (2), 306–316. https://doi.org/10.1002/bit.22532.

(6) Searle, B. C.; Swearingen, K. E.; Barnes, C. A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating High Quality Libraries for DIA MS with Empirically Corrected Peptide Predictions. *Nat Commun* **2020**, *11* (1), 1548. https://doi.org/10.1038/s41467-020-15346-1.

(7) Jones, M.; Palackal, N.; Wang, F.; Gaza-Bulseco, G.; Hurkmans, K.; Zhao, Y.; Chitikila, C.; Clavier, S.; Liu, S.; Menesale, E.; Schonenbach, N. S.; Sharma, S.; Valax, P.; Waerner, T.; Zhang, L.; Connolly, T. "High-risk" Host Cell Proteins (HCPs): A Multi-company Collaborative View. *Biotech & Bioengineering* **2021**, *118* (8), 2870–2885. https://doi.org/10.1002/bit.27808.

(8) Pythoud, N.; Bons, J.; Mijola, G.; Beck, A.; Cianférani, 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. *J. Proteome Res.* **2021**, *20* (1), 923–931. https://doi.org/10.1021/acs.jproteome.0c00664.

(9) Beaumal, C.; Beck, A.; Hernandez-Alba, O.; Carapito, C. Advanced Mass Spectrometry Workflows for Accurate Quantification of Trace-level Host Cell Proteins in Drug Products: Benefits of FAIMS Separation and Gas-phase Fractionation DIA. *Proteomics* **2023**, *23* (16), 2300172. https://doi.org/10.1002/pmic.202300172.

(10) Chen, I.-H.; Xiao, H.; Daly, T.; Li, N. Improved Host Cell Protein Analysis in Monoclonal Antibody Products through Molecular Weight Cutoff Enrichment. *Anal. Chem.* **2020**, *92* (5), 3751–3757. https://doi.org/10.1021/acs.analchem.9b05081.

General Introduction

General Introduction

Monoclonal antibodies (mAbs) have revolutionized modern medicine, offering targeted therapeutic options for diseases such as cancer, autoimmune disorders, and infectious diseases. These biologics owe their success to their high specificity, efficacy, and versatility, establishing them as a fundamental component of the pharmaceutical industry. However, their production in mammalian cell systems, predominantly Chinese hamster ovary (CHO) cells, presents significant challenges. Among these challenges is the presence of host cell proteins (HCPs), impurities generated during cell culture that persist through downstream processing. Even trace levels of HCPs can compromise drug safety, stability, and efficacy, underscoring the critical importance of developing robust methods for their detection and quantification.

The complexity of HCPs lies in their vast heterogeneity and variability, which arise from differences in cell culture conditions, production scales, and purification processes. These impurities could potentially alter the mAb's therapeutic properties. Regulatory agencies such as the FDA and EMA impose stringent guidelines to minimize HCP content in final drug products, further emphasizing the need for highly sensitive and reliable analytical methods. Despite advancements in downstream processing technologies, including chromatographic and filtration techniques, achieving complete HCP removal remains an analytical challenge. Therefore, advanced analytical tools play an indispensable role in monitoring HCP levels and ensuring product quality throughout the biomanufacturing workflow.

Under this context, my thesis work is built upon two key objectives. The first objective focuses on the development and validation of an immune-capture mass spectrometry-based method as a complementary tool in order to assess the enzyme-linked immunosorbent assay results. While ELISA has been the industry standard for HCP quantification, its reliance on polyclonal antibodies introduces limitations in sensitivity, specificity, and impurity coverage. The method developed aims to overcome these constraints by integrating targeted immune-capture workflows with the analytical power of MS, thereby offering a more comprehensive HCP analysis and focuses on the biggest limitation that ELISA has, which is the undetection of those HCPs that are non-immuno-reactive. The second objective addresses the advancement of liquid chromatography-tandem mass spectrometry (LC-MS/MS) workflows for HCP quantification. Specifically, it explores a dual quantification approach leveraging peptide standards (ReadyBeads) and targeted AQUA peptides. This strategy seeks to achieve greater accuracy and flexibility in impurity profiling across mAb production and purification stages.

The manuscript is structured into three distinct parts, each aligned with the overarching goals of the research:

Part I provides an extensive bibliographic foundation, including an introduction to proteomics, the mechanisms of monoclonal antibody production, and the challenges associated with HCP removal in the downstream processing (DSP) cascade. This section also introduces the pivotal role of mass spectrometry in proteomic analysis and its transformative potential for quality control in mAb manufacturing. Additionally, this part delves into the analytical techniques employed for HCP

detection, tracing the transition from immunological methods to modern MS-based approaches. By situating the research within this context, Part I establishes the scientific groundwork for the methodological advancements discussed in subsequent sections.

Part II is dedicated to the first objective: the development of an immuno-capture MS method to complement and assess ELISA HCP quantification. This method addresses a crucial gap in traditional analytical workflows by providing a more sensitive and comprehensive means of detecting HCPs, especially those that are non-immunoreactive and typically undetectable by ELISA. The ability to identify such impurities is a key innovation of this approach, helping to overcome the limitations of ELISA, which is often hindered by issues of specificity and impurity coverage. The development of this method holds significant potential for biopharma companies, as it could streamline the validation of ELISA suitability, offering substantial savings in both time and resources. By enabling more efficient impurity profiling, this method supports the biopharmaceutical industry's ongoing efforts to improve product quality and regulatory compliance. This section also covers the optimization of immuno-capture protocols to enhance the recovery of non-immunoreactive HCPs, as well as the technical challenges and solutions encountered during method validation, including binding efficiency and reproducibility.

Part III centers on the second objective, which focuses on advancing LC-MS/MS workflows for comprehensive HCP quantification. This section begins by building on prior research, first addressing HCP quantification for the entire HCP population using a mixture of peptide standards covering a concentration range (ReadyBeads technology from Anaquant). The methodology applied here extends previous work, providing a broad overview of HCP content across the downstream processing cascade. After establishing this comprehensive quantification framework, the focus shifts toward a more detailed risk assessment. In this phase, both the ReadyBeads and AQUA peptide standards are employed, enabling a more targeted and refined analysis of specific HCPs that pose higher risks to drug safety and efficacy. This dual approach allows for a more nuanced evaluation of HCP profiles, which is critical for the identification of potential process-related impurities. The findings from this section highlight the interest of integrating both global and targeted strategies to ensure thorough impurity assessment, offering a comprehensive view of HCP dynamics throughout the purification process, while offering accurate and precise quantification of high risk candidates.

Through these three parts, the thesis aims to address critical gaps in HCP analysis while employing several MS applications. The research not only enhances existing analytical workflows but also underscores the potential of mass spectrometry in advancing the safety and efficacy of monoclonal antibody therapeutics.

PART I. Bibliographic Introduction: State of the Art in Quantitative Proteomics

Chapter 1
Mass Spectrometry in Bottom-up Proteomics

1. Overview of Mass Spectrometry-based Proteomics

The term "proteome," a combination of "protein" and "genome," was first introduced in 1994 by Mark Wilkins, a PhD student at Macquarie University in Sydney.[11] While the genome represents the full set of genes in an organism, the proteome encompasses all the proteins present in a biological system at a specific time and under specific conditions.[12] Unlike the static nature of the genome, the proteome is dynamic, reflecting the functional state of a cell, tissue, or organism.[13]

Building on this concept, Peter James coined the term "proteomics" a few years later.[14] Initially, proteomics was understood as the comprehensive study of all proteins in a cell, tissue, or organism.[13] Over time, this definition evolved to highlight the proteome's inherent dynamism. Today, proteomics focuses on the qualitative, quantitative, and functional analysis of proteins under specific conditions and at particular moments in time.[15] By directly characterizing a broad array of proteins simultaneously, proteomics provides an unbiased approach for studying biological processes and regulatory mechanisms.[16]

Mass spectrometry (MS) has become a critical tool in proteomics, owing to several key technological advancements. [15,17–19] Among these, the development of soft ionization techniques such as Matrix-Assisted Laser Desorption Ionization (MALDI) and Electro spray Ionization (ESI) by Koichi Tanaka and John B. Fenn, respectively, marked a major breakthrough in analyzing biological macromolecules like proteins.[20,21] This innovation earned both scientists the Nobel Prize in Chemistry in 2002. Progress in separation techniques, enhanced mass spectrometer sensitivity, increased resolution, and the refinement of data acquisition methods have further elevated MS capabilities. Additionally, improvements in curated protein databases and bioinformatics tools have streamlined the analysis and interpretation of complex proteomic data, making MS indispensable in modern proteomics.

Mass spectrometry-based proteomic analysis can be categorized into three primary approaches: bottom-up, middle-down, and top-down (Figure 1) below.

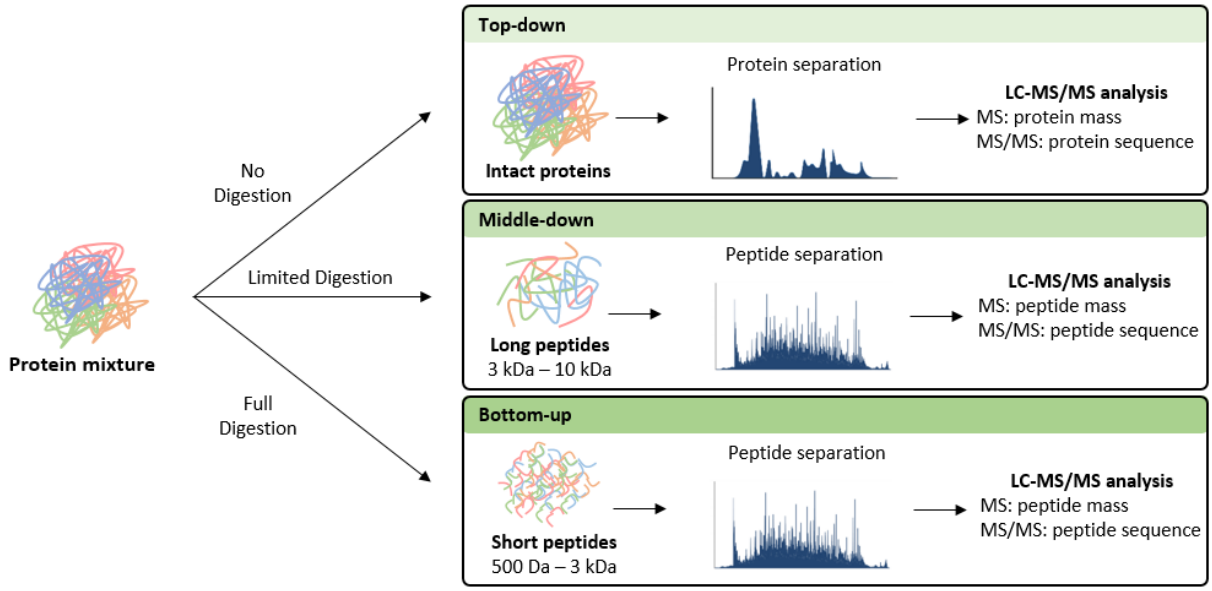


Figure 1: Overview of the three MS-based proteomic approaches: Bottom-up, Middle-down, and Top-down

The bottom-up approach is the most widely employed method due to its compatibility with large-scale and high-throughput proteomic studies. In this approach, proteins are enzymatically digested, typically using trypsin, to produce peptides smaller than 3 kDa. These peptides are then separated using liquid chromatography and analyzed by mass spectrometry (MS) to determine peptide mass and sequence (MS/MS). Protein identification is achieved by comparing the observed peptide masses with theoretical values derived from in silico digestion of protein sequences in curated databases. This process enables protein inference, where multiple peptide identifications are associated with their respective proteins. Challenges arise in cases where peptides are shared among proteins, requiring grouping and application of the parsimony principle to generate the smallest list of proteins explaining the observed data.[22] Advances in LC-MS/MS platforms now allow thousands of proteins to be identified within hours, solidifying the bottom-up approach as a cornerstone of modern proteomics.[23]

While conceptually straightforward, the bottom-up strategy involves inherent challenges, such as potential loss of information during protein digestion and preferential detection of abundant peptides, which may leave low-abundance proteins underrepresented.[24] Despite these limitations, this method remains the most reliable for analyzing complex biological samples due to its scalability and well-established workflows.

The top-down approach, in contrast, directly analyzes intact proteins without enzymatic digestion, offering advantages in proteoform differentiation and the analysis of post-translational modifications (PTMs). Proteins are ionized and fragmented within the mass spectrometer, allowing for comprehensive examination of their sequences. This method has proven effective for characterizing proteoforms and elucidating modifications by identifying discrepancies between measured protein masses and their predicted genomic sequences.[24] Recent advancements in multidimensional

separation techniques and high-resolution mass spectrometry have enabled the identification of thousands of proteoforms.[25,26] However, the top-down approach requires highly purified samples and specialized instrumentation to overcome challenges related to protein solubility, ionization, and fragmentation.[25–30] Despite its potential, these constraints currently limit its applicability to more complex biological samples.

Bridging the gap between bottom-up and top-down, the middle-down approach targets large peptides (3–10 kDa) produced through restricted enzymatic digestion. These peptides are analyzed by LC-MS/MS, offering insights into PTMs and proteoforms while reducing the complexity of peptide mixtures compared to bottom-up methods. However, this strategy requires precise optimization of proteolysis conditions and further advancements in data analysis tools (Han, 70, 71).[24,31,32]

While the top-down and middle-down approaches hold promise for specific applications, the bottom-up strategy remains the most robust, scalable, and widely adopted method for proteomic analysis. Consequently, the research presented in this manuscript exclusively employs the bottom-up approach, detailed in the following sections of this chapter.

2. Analytical workflow of Bottom-up Proteomics

The bottom-up proteomic workflow involves three fundamental steps: sample preparation**,** liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), and data processing and analysis (Figure 2).

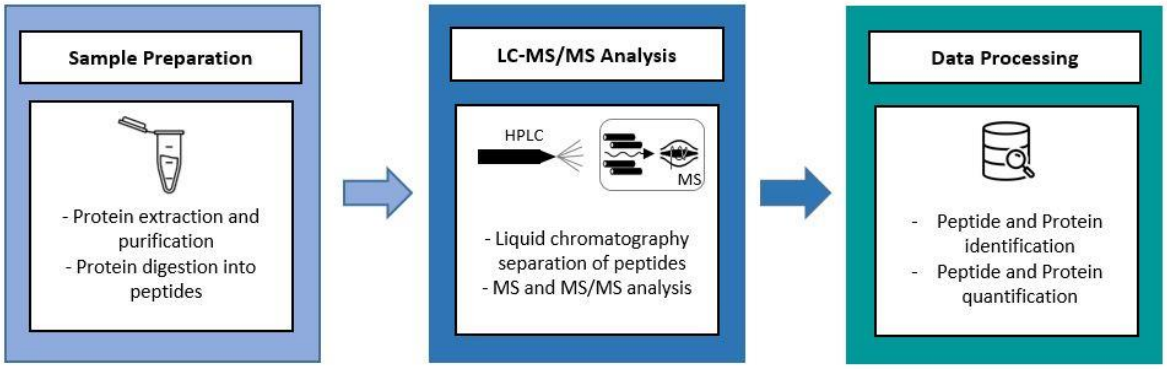


Figure 2: Overview of the key steps in the bottom-up proteomics workflow.

A. Sample preparation

Sample preparation is a critical step in the bottom-up proteomic workflow, as the quality and reproducibility of this process significantly influence the outcomes of the mass spectrometry (MS) analysis. Each sub-step in the preparation must be meticulously optimized and tailored to the specific characteristics of the sample being analyzed.[33,34]

i) <u>Protein Extraction techniques</u>

The first step in the bottom-up proteomics workflow is extracting proteins from biological samples such as tissues or fluids. This process aims to solubilize as many proteins as possible while avoiding

degradation or unwanted modifications. The efficiency of extraction is critical, as it determines how accessible proteins are to proteases, ultimately impacting the quality of the proteomic analysis. Therefore, this step must be carefully tailored to the sample type and quantity, the nature of the proteins, and the downstream analytical workflow.[35–37]

Proteins can be lysed and extracted through various approaches:[36,38]

1. **Mechanical disruption**, such as grinding or ultrasonication, physically breaks down the sample to release proteins.
2. **Chemical extraction**, involving:
    - Detergents (ionic, non-ionic, zwitterionic, or bile acid salts) to aid protein solubilization through micelle formation. Detergents compatible with protein digestion and LC-MS/MS analysis can be used for seamless workflows.
    - Chaotropic agents, like urea, which denature proteins and stabilize their unfolded states.
    - Organic solvents, such as acetonitrile, which disrupt protein structure and enhance solubility.

The most widely used method combines chemical lysis agents with mechanical techniques to achieve optimal protein recovery.[36]

Following lysis, additional purification steps may be necessary to eliminate contaminants, including lipids, nucleic acids, and residual detergents, which can interfere with LC-MS/MS analysis. These contaminants may cause chromatographic disruptions, suppress signals, or generate noisy spectra. Cleanup procedures often involve protein precipitation using reagents like cold acetone, trichloroacetic acid, or a chloroform-methanol mixture. Dialysis or ultrafiltration can also be employed. However, these steps require careful handling to minimize protein loss.[36]

ii) Preparation of Protein Mixture for LC/MS-MS Analysis

Samples used in proteomics research often present significant analytical challenges. These biological samples consist of thousands of proteins, which are found in various forms, adding to the complexity of analysis. The different forms of a protein, referred to as proteoforms, arise due to factors such as genetic sequence variations, splicing events, or post-translational modifications (PTMs) like phosphorylation, glycosylation, and oxidation. This variability within the proteome makes comprehensive analysis particularly challenging.

For example, the human genome contains approximately 20,000 protein-coding genes, as reported by the Human Genome Project.[39] When considering a single gene that produces a protein, the total number of proteoforms can increase to about 70,000 due to alternative splicing alone. When accounting for PTMs, the total number of proteoforms can reach several hundred thousand, adding considerable complexity to the analysis. Furthermore, the dynamic range of protein abundances within a sample significantly increases this complexity. Protein concentrations can vary by as much as 10 orders of magnitude, while current mass spectrometric techniques typically cover only up to 5 orders of magnitude.[15,40,41] To address these challenges, a variety of methods have been developed to reduce

the complexity of proteomics samples. One approach involves depleting highly abundant proteins, thus alleviating the dynamic range issue; however, this technique risks the loss of valuable biological material.[42] Another method involves fractionating proteins based on their physicochemical properties, such as molecular weight using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), size via size exclusion chromatography, isoelectric point using ion exchange chromatography, or polarity with reverse-phase chromatography.[33,34,43] These approaches are essential to simplify proteomic samples while retaining biological relevance.

In-Solution and In-Gel Protein Preparations for LC-MS/MS Analysis

Two primary protein preparation techniques are commonly used in the field of proteomics: in-solution digestion (ISD) and in-gel digestion protocols.[33,34,36] Both methods have their respective strengths and limitations, as outlined below:

- **In-Solution Preparations**: In this approach, proteins are denatured, disulfide bonds are reduced, and cysteine residues are alkylated before enzymatic digestion.[36] Buffers that are compatible with subsequent enzymatic digestion are essential for the success of this method. Over the years, several protocols have been developed to improve the efficiency of in-solution preparations. One of these protocols is Filter-Aided Sample Preparation (FASP), which facilitates protein extraction through filters or membranes. FASP also employs detergents to help solubilize proteins but carries a risk of material loss due to protein aggregation.[44] Another innovative method is the Single-Pot Solid-Phase Enhanced Sample Preparation (SP3), which utilizes paramagnetic beads to immobilize proteins or peptides, allowing for purification and digestion in a single step. Detergents can also be incorporated into this method to enhance protein solubilization.[44] For convenience, several commercial kits are available, such as the iST Preparation Kit (Preomics), S-Trap (ProtiFi), Pierce™ Mass Spec Sample Prep Kit (Thermo Fisher), and Sample Preparation Kit (Biognosys), all designed to streamline the sample preparation process for mass spectrometry.[45,46] Recent studies have also highlighted novel approaches, such as using on-bead purification during protein digestion[47], which presents a significant improvement over traditional methods by reducing sample loss during the isolation process.

- **In-Gel Preparations**: In contrast to in-solution methods, in-gel preparations are primarily used to eliminate contaminants such as detergents and salts. This approach is particularly useful for removing SDS when it is used in protein extraction protocols. In this study, two common in-gel protocols were employed:
    1. **1D SDS-PAGE**: This technique involves separating proteins by their migration through a polyacrylamide gel under the influence of an electric field. The application of an electric field allows proteins to migrate based on their charge-to-mass ratio.[48,49] The SDS in the loading buffer ensures that proteins adopt a linear conformation and carry a uniform negative charge, allowing them to be separated by molecular weight. The polyacrylamide gel consists of two distinct sections: a stacking gel with a low percentage of acrylamide (4

to 5%) where proteins are concentrated into a single band, and a separation gel with a higher percentage of acrylamide (8 to 15%) that resolves proteins based on their size.

2. **Stacking Gel**: This protocol is similar to 1D SDS-PAGE but differs in that the migration is stopped once the proteins are concentrated in a single band, just before entering the separation gel. This method permits the use of SDS for protein solubilization without requiring fractionation of the sample, making it especially suitable for accurate and precise quantification.

Additionally, there are more advanced gel-based methods such as Tube-Gel[50,51] and 2D SDS-PAGE[52]. The Tube-Gel method involves directly incorporating proteins into the polyacrylamide gel prior to polymerization, avoiding electrophoretic migration.

The 2D SDS-PAGE method consists of two separate stages: the first dimension separates proteins based on their isoelectric point, while the second stage uses SDS-PAGE to separate them based on molecular weight. These techniques are valuable for gaining deeper insights into protein profiles but are increasingly being replaced by more efficient and automated protocols for high-throughput analysis.

After gel separation, proteins are often fixed in the gel, visualized using Coomassie Blue staining[53], and then subjected to reduction and alkylation prior to in-gel enzymatic digestion. Although in-gel approaches are still widely used, especially in the stacking-gel form, they are gradually being replaced by more automated methods such as FASP and SP3, which provide better reproducibility and are more suited to large-scale proteomic studies. As Varnavides et al. (2025) emphasized, automated in-solution approaches offer higher throughput and have become increasingly popular for large-scale proteomic analyses due to their scalability and reproducibility.[47]

Another attractive alternative is two-dimensional difference gel electrophoresis (2D-DIGE), a powerful technique in proteomics that enhances efficiency and accuracy. By utilizing cyanine fluorescent dyes, this method enables the co-migration of multiple protein samples within the same gel, allowing for their simultaneous detection and significantly reducing both experimental and analytical time. Compared to traditional post-staining 2D-PAGE methods, such as colloidal Coomassie or silver nitrate staining, 2D-DIGE offers faster and more reliable gel matching while minimizing gel-to-gel variability. Additionally, it provides a broad dynamic range, making it well-suited for quantitative protein comparisons.[54] Once the samples have been run, they follow the same protocol as the 2D SDS-PAGE, where they are cut and submitted to reduction, alkylation and digestion.

iii) Enzymatic digestion

Enzymatic digestion, a crucial step in bottom-up proteomics, is most frequently carried out using trypsin. This enzyme owes its widespread use to its stringent cleavage specificity: it cleaves peptides at the carboxyl side of lysine and arginine residues, except when followed by proline.[55,56] This pattern of cleavage is particularly advantageous, as it results in an average of 61 peptides per protein in human proteins, as calculated by DBToolkit.[56] The positive charge at the C-terminal end of tryptic peptides facilitates ionization and fragmentation, making them ideally suited for LC-MS/MS analysis.

Additionally, trypsin-generated peptides typically fall within the 500–3000 Da range, which supports effective protein sequence coverage. Trypsin's versatility extends to in-gel digestion, as it efficiently penetrates polyacrylamide gels.

The combined use of multiple proteases in a single experiment has proven beneficial for improving sequence coverage. For instance, Lys-C is particularly effective when used alongside trypsin, as it complements trypsin's relatively lower efficiency with lysine residues compared to arginine.[57,58] Additionally, Lys-C exhibits resistance to denaturing conditions, further enhancing its utility in challenging proteomic workflows. Such combined enzymatic approaches underline the sophistication of modern proteomics workflows, which are designed to address complex biological questions.[59]

Despite its widespread adoption, trypsin digestion is not without limitations. The process typically requires 12 to 18 hours, which can be a bottleneck in high-throughput workflows. To address this, commercial solutions like the SMART Digest Kit (Thermo Fisher Scientific) and Rapid Digestion Kit (Promega) have been developed to drastically reduce digestion times.[60] These advancements highlight the diversity of available digestion protocols and the necessity of selecting methods tailored to specific sample types, protein properties, and research objectives.

B. Liquid Chromatography Coupled to Tandem Mass Spectrometry

i) Peptide separation techniques

Enzymatic digestion, which is essential for the bottom-up proteomics approach, inherently adds complexity to the sample.[15,19] To address this, liquid chromatography is typically employed as a means to reduce this complexity. By enhancing ionization efficiency and minimizing analyte competition, it offers improvements in sensitivity, selectivity, and proteome coverage.[61,62] In this study, reverse-phase liquid chromatography was utilized to achieve peptide separation based on hydrophobicity. This was accomplished by incrementally decreasing the polarity of the mobile phase, a mixture of water and acetonitrile. Prior to the analytical column, peptides were trapped in a trap column with low acetonitrile concentration (~1-3%) to eliminate contaminants.

For this project, an UltiMate™ 3000 RSLCnano system (Thermo Fisher Scientific) was employed. Further details regarding the chromatographic setup can be found in Table 1 below.

Properties	Trap Column	Analytical Column
Supplier	Thermo Fisher Scientific	IonOptics
Type	Acclaim™ PepMap™	Aurora™ Ultimate
Stationary phase	C18	C18
Column length (mm)	20	250
Internal diameter (µm)	75 µm	75 µm
Particle size (µm)	3	1.7
Pore size (Å)	100	120
Flow rate (µL/min)	10 nL/min	300 nL/min

Table 1: Technical specifications of the chromatographic system used in this study.

The separation efficiency of these systems is influenced by various column properties, such as length, internal diameter, pore size, and particle size.[19,63] Nanofluidic systems, operating under high pressure (>400 bar), are particularly advantageous for proteomic studies due to their heightened sensitivity, resolution, and peak capacity, even with minimal amounts of biological material. These attributes make them ideal for analyzing small-scale proteomic samples, where sample availability is often limited. However, they present challenges, including instability in ionization spray, difficulty in detecting leaks or dead volumes, and the need for specialized expertise to manage the system effectively.

In contrast, microflow chromatographic systems, while offering lower sensitivity, compensate for this drawback by allowing larger sample injections.[64] They are also more robust and easier to handle compared to nanoflow systems, making them a reliable alternative in specific scenarios.[13,64]

ii) Tandem Mass Spectrometry

Peptides separated through reverse-phase liquid chromatography are ionized directly using electrospray ionization (ESI) before being introduced into the mass spectrometer. This device initially records the mass-to-charge ratio (m/z) and the ion intensities to produce a mass spectrum, known as the MS or MS1 spectrum. Subsequently, specific ions are selected, isolated, and fragmented. The resulting fragments' m/z ratios and intensities are then measured to create an MS/MS (or MS2) spectrum. This process of sequentially producing MS1 and MS2 spectra, referred to as tandem mass spectrometry, is executed on hybrid mass spectrometers that integrate multiple mass analyzer technologies, including quadrupole (Q), time of flight (TOF), Orbitrap, ion trap, and Fourier-transform ion cyclotron resonance (FT-ICR) analyzers.[65]

All experiments described in this manuscript were performed using the Orbitrap Eclipse Tribrid (Thermo Fisher Scientific), a hybrid mass spectrometer that integrates quadrupole, ion trap, and Orbitrap analyzers. Key specifications are provided in Table 2 below.

Properties	Orbitrap Eclipse Tribrid
System vendor	Thermo Fisher Scientific
Analyzer	Quadrupole, Orbitral and Ion Trap (Tribrid instrument)
Resolution	Ion Trap: 0,3 – 3 m/z FWMH Orbitrap: 7,500 – 500 000 (at 200 m/z)
Mass Precision	5 ppm
Scan Rate	Ion Trap: up to 45Hz (in Turbo mode) Orbitrap: Up to 40 Hz (at 7,500 of resolution)
Mass Range	Up to 8,000 m/z
Ion mobility	FAIMS implementation option
Fragmentation Type	CID, HCD, ETD, EThcD, ETciD, UVPD
Year of implementation in the lab	2020

Table 2: Key specifications of the mass spectrometer used in this project.

a. Data Dependent Acquisition

Data Dependent Acquisition (DDA) is the most commonly employed method in bottom-up proteomics. During the chromatography gradient, the mass spectrometer, operating in DDA mode, continuously cycles through the acquisition of both MS1 and MS2 spectra. In each cycle, a MS1 spectrum is initially obtained, and the most intense precursor ions from this spectrum are selected for fragmentation in real-time. These precursor ions are isolated within a specific m/z window, generating N MS2 spectra for analysis (Figure 3). [66]

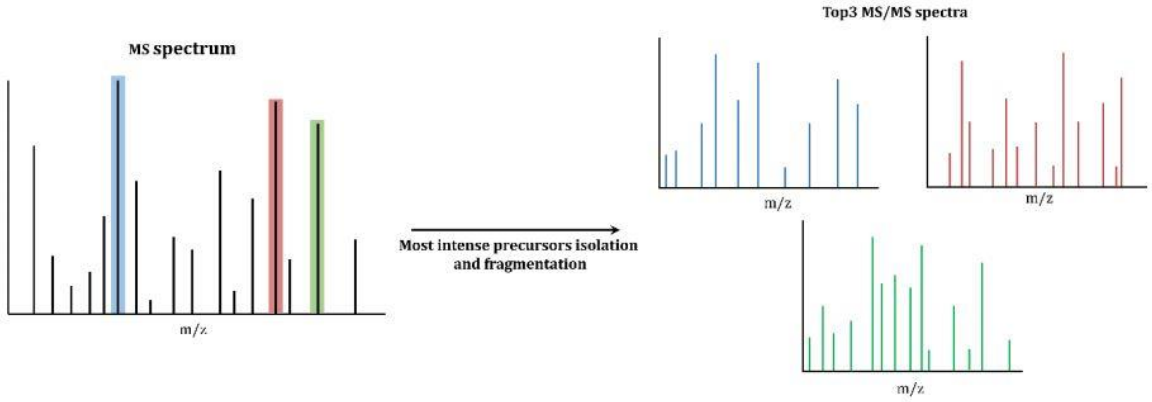


Figure 3: Illustration of the DDA approach, where the TopN most abundant precursors (MS1) are sequentially selected, isolated, and fragmented to generate MS2 spectra of the resulting fragment ions.

DDA enables the identification of thousands of proteins, offering substantial proteome coverage.[67,68] However, the reliance on precursor ion intensity for selection in the MS1 spectrum introduces variability, which can compromise reproducibility. This bias arises due to the stochastic nature of precursor selection, where highly abundant ions are favored over lower-intensity ones, leading to inconsistent detection across replicates[67]. As a result, less abundant peptides may be missed, particularly in complex samples.

Recent advances in instrument sensitivity and faster acquisition speeds have minimized the effects of undersampling.[67] Additionally, there are several technical strategies available to enhance the number of protein identifications. One such strategy is dynamic exclusion, which reduces the repetition of spectra by excluding previously selected precursor ions, thereby increasing the number of unique identifications.[69] Another approach is the use of inclusion or exclusion lists, which enable users to prioritize or filter out specific precursor ions, improving selection efficiency.[69,70] Despite these improvements, DDA remains inherently limited to relative quantification, as it does not provide absolute concentration measurements of target analytes.[71]

To overcome this limitation, targeted acquisition strategies such as Selected Reaction Monitoring (SRM)[72], Parallel Reaction Monitoring (PRM)[73], and Multiple Reaction Monitoring (MRM)[74] have been developed for absolute quantification, offering greater reproducibility and sensitivity by focusing on predefined peptide targets.

b. Peptide Fragmentation Methods

Peptides can be fragmented using various techniques on mass spectrometers, such as Collision Induced Dissociation (CID)[75], Higher Energy Collisional Dissociation (HCD)[76], Electron Transfer Dissociation (ETD)[77], Electron Capture Dissociation (ECD)[41], and a combination of ETD and HCD known as EThcD.

Among these, CID and HC fragmentations are the most widely employed fragmentation method in bottom-up proteomics[15]. In these cases, ions are isolated and accelerated to gain high kinetic energy. These ions then collide with neutral molecules, typically argon, helium, or nitrogen, within the collision cell. The collision converts kinetic energy into internal energy, causing peptide bonds to break, in line with the mobile proton model.[78,79] This process primarily generates b and y ions, following Biemann's nomenclature[80], as observed in Figure 4. These peptides typically carry at least two positive charges: one at the N-terminal ($NH3^+$) and another on the side chain of lysine or arginine residues[78]. These fragment ions enable precise peptide sequence determination.

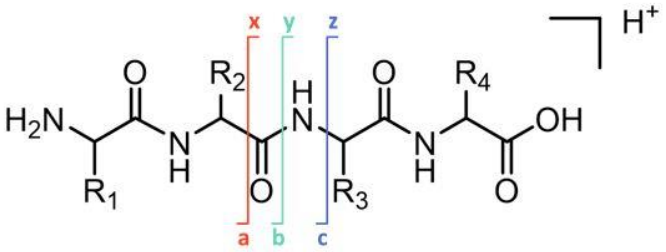


Figure 4: Ion types generated after fragmentation, according to the Biemann nomenclature. Six ion types can be produced: a, b, and c for N-terminal fragments, and x, y, and z for C-terminal fragments.

In the particular case of HCD, ions are subjected to greater acceleration, which results in higher energy fragmentation compared to CID. Unlike CID, which occurs within the ion trap's high-pressure cell on the Eclipse, HCD is typically performed in a collision cell[76]. On older instruments, ions would be accumulated in the C-trap, fragmented in the collision cell, and then transferred back to the C-trap for analysis. However, on the Eclipse, both ion accumulation and HCD fragmentation take place in the ion routing multipole (IRM), with the C-trap acting solely to guide the ions to either the Orbitrap or the IRM (Figure 5).

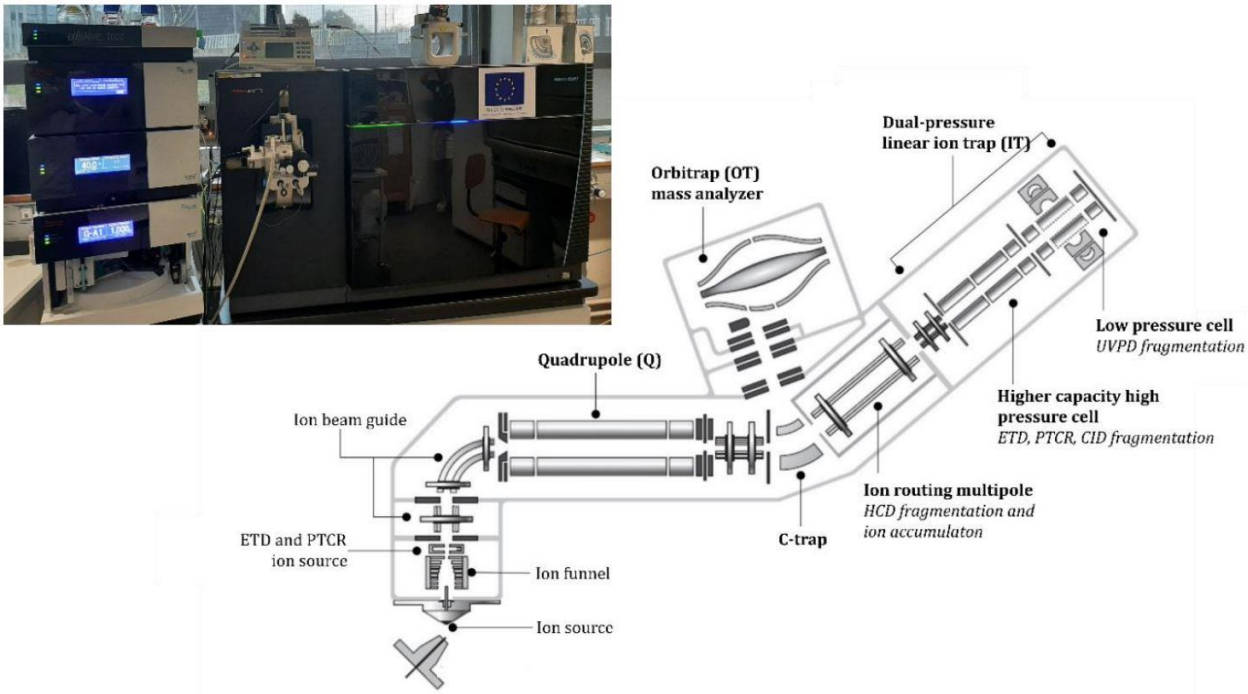


Figure 5: Diagram illustrating the LC-MS/MS setup and a schematic representation of the Orbitrap Eclipse Tribrid (Thermo Fisher Scientific).

C. Data Processing and Interpretation

The data acquired from the mass spectrometer, including precursor and fragment ion mass-to-charge ratios (MS1 and MS2), intensities, and retention times, are processed into a peak list. These experimental peaks are then matched against theoretical spectra generated through *in silico* digestion of protein sequences from the target organism. This methodology, referred to as peptide fragmentation fingerprinting (PFF)[81], serves as the foundation for peptide identification.

Protein identification is subsequently achieved by mapping the identified peptides to their corresponding proteins. While certain peptides uniquely match a single protein, others may align with multiple proteins, leading to the formation of protein groups. To simplify this complexity, protein inference employs the principle of parsimony, reducing the results to the smallest set of protein groups necessary to explain the observed data.[22]

i) Search Engine

The raw output from bottom-up proteomic analysis is converted into peak lists that contain both the mass-to-charge (m/z) ratios and intensities of precursor and fragment ions. These data are then used to identify peptides through a method known as peptide fragmentation fingerprinting. This process involves comparing the experimental mass lists to theoretical lists generated by *in-silico* digestion and fragmentation of protein sequence databases. Once peptides are identified, they are grouped together to make protein inferences. This step can be complex due to shared peptides between proteins or cases where proteins are identified from a single peptide[22]. Ultimately, this process results in the identification of proteins in the sample. The steps involved are typically automated using search

engines such as Andromeda[82], Mascot[83] (Matrix Science, London, UK), Pulsar (Biognosys, Schlieren, Switzerland), OMSSA[84], Sequest[85], Byonic™ (Protein Metrics, Cupertino, USA), and X!Tandem[86]. These search engines require specific information about the experimental conditions and the instruments used, such as:

- The protease employed and the maximum permissible number of missed cleavages.
- The protein sequence database being searched.
- The tolerance level for precursor and fragment ion m/z ratios.
- The charge state of precursor and fragment ions.
- Expected amino acid modifications, whether fixed or variable.
- The fragmentation method used.

For the work detailed in this manuscript, two search engines were employed: Mascot and Pulsar.

Mascot uses a probability-based scoring algorithm that calculates an "ion score" for each MS2 spectrum, reflecting the likelihood that the match between experimental and theoretical mass lists was made by chance. Higher ion scores indicate more confident peptide identifications. This score is based on the quality of the spectrum and does not consider the protein sequence database. Additionally, Mascot calculates identity and homology thresholds depending on the size of the database to determine whether the peptide identification was random. All peptide identifications are linked to the corresponding peptide spectrum matches (PSMs) and organized accordingly.

Pulsar, introduced by Biognosys in 2017, is incorporated into the Spectronaut™ software for generating spectral libraries used in the extraction of data from DIA experiments. As a commercial solution, the algorithm for Pulsar remains proprietary.

Despite improvements in both instrumentation and software, between 50 and 75% of MS2 spectra are still left unidentified.[87] The reasons for this are numerous:

- Some MS2 spectra lack sufficient quality.
- The co-isolation and co-fragmentation of peptides within narrow m/z windows (1–3) can negatively impact identification, especially if unassigned peaks are used in scoring calculations. Since version 2.5, Mascot has enabled the identification of peptides from chimeric spectra.
- Protein sequence databases may be incomplete or inadequate.
- A lack of consideration for modifications, which contribute to about one-third of unassigned spectra.[88,89]
- Errors during data processing, such as incorrect assignment of the mono-isotopic peak or charge state.

ii) Protein Sequence database

Peptide assignment via a protein sequence database limits the possible identifications to the content of this repository. Therefore, it is essential to use the most appropriate database for the biological context. Additionally, to extract accurate and high-quality information, it is vital to work with well-annotated databases that have undergone sequence annotation and data filtering. Several databases are available, varying in terms of annotation quality, completeness, and degree of redundancy.[22] These include:

- **RefSeq**[90] is produced by NCBI. This database is curated, verified, and free of redundancy. It includes information linking proteins to their corresponding genes and transcripts. As of January 8, 2025, RefSeq contained 391,903,900 proteins from 162,138 organisms. However, errors can still occur during the translation of nucleotide sequences into protein sequences, which may negatively impact MS data analysis.[91]

- **UniProtKB**[92] is the result of a collaborative effort between the European Bioinformatics Institute (EMBL-EBI), the Swiss Institute of Bioinformatics (SIB), and the PIR. It consists of two main components:
  - UniProtKB/TrEMBL contains protein sequences that are automatically translated, annotated, and classified from sources like GenBank, literature, and other databases, awaiting manual validation for inclusion in UniProtKB/SwissProt.
  - UniProtKB/SwissProt is the curated section of the database, considered the gold standard in proteomics. It contains high-quality, manually annotated protein sequences, offering detailed information about protein function, subcellular localization, interactions, associated pathologies, expression levels, structure, and post-translational modifications (PTMs). This section has been curated since 1986.

As of November 27, 2024, UniProtKB/TrEMBL contained 253,682,368 entries, while UniProtKB/SwissProt contained 572,619 entries.

Databases are frequently updated due to the discovery of new variants, sequences, and ongoing manual verification. Using up-to-date databases can significantly reduce the number of unassigned MS2 spectra. Additionally, combining results from genomic, transcriptomic, and proteomic studies enhances genome annotation, gene prediction algorithms, and peptide/protein identification.[93,94] This integration of data, first introduced in 2004, is referred to as proteogenomics.[95] Despite its potential, the task remains highly complex.[96]

In the context of this project, the CHO protein sequence database was specifically utilized. A major challenge remains the limited availability of high-quality, publicly accessible databases for CHO cell lines, despite their critical role in the production of most commercially available monoclonal antibodies. As of 2025, the CHO database on the UniProt website contains 83,419 entries, highlighting ongoing efforts to expand and refine its coverage. However, the completeness and accuracy of these

entries continue to be a subject of discussion, emphasizing the need for further curation and validation to support proteomics research.[97]

iii) <u>Validation Methodologies</u>

The identification of peptides and proteins in mass spectrometry (MS) is prone to errors due to the limitations of search algorithms. The scores assigned by search engines for peptide-spectrum matches (PSMs) are not always sufficient to determine the accuracy of identifications, and manually verifying thousands of PSMs is impractical. To address this issue, additional validation strategies are essential. The most widely employed method is the target-decoy strategy, which helps to differentiate correct identifications from false ones.[98] In this approach, both target protein sequences and decoy sequences—randomized or reversed versions of the target sequences—are included in the search database. The assumption is that decoy sequences do not occur in nature, so any match to these sequences is considered a false positive. The false discovery rate (FDR) is then estimated based on the number of decoy sequences assigned to MS2 spectra compared to the total number of assigned sequences, following the equation:

$$\boldsymbol{FDR}\ (\%) = \frac{\boldsymbol{Number\ of\ assigned\ decoy\ PSM}}{\boldsymbol{Number\ of\ assigned\ decoy\ PSM + Number\ of\ assigned\ target\ PSM}} \times \mathbf{100}$$

This method evaluates the global FDR by calculating the average proportion of false PSMs. Typically, in bottom-up proteomics, a target FDR of 1% is used to determine the threshold score above which PSMs are considered valid. However, this approach only considers a single score from the search engine, and multiple scores (such as C-Score or ΔC-Score for SEQUEST) can provide more nuanced information regarding the quality of the peptide assignments.

To enhance identification accuracy while controlling the FDR, the use of Percolator[99] has been proposed. Percolator is a semi-supervised machine learning algorithm that integrates multiple scores and features from the PSM scan, such as precursor mass error, matched ion fractions, peptide length, and XCorr. This algorithm assigns a statistically meaningful q-value to each PSM, which represents the minimal FDR at which the identification can be considered correct. By utilizing Percolator, researchers can significantly increase the number of correct peptide identifications compared to traditional search engine-based validation methods. It is compatible with any search engine output and is widely used in proteomics for improving data quality.

FDR estimation can be performed at various levels, PSMs, peptides, or proteins, and combining filters at multiple levels can further increase the confidence in peptide and protein identification. Nevertheless, no universal rules exist for which filters to apply, and the choice often depends on the specific study or dataset.

Chapter 2
Quantification Strategies in Proteomics

1. Global Quantification Strategies in Proteomics

Quantitative analysis of peptides and proteins in a sample provides critical insights into the functions and dynamics of biological systems. Variations in protein abundance can often serve as indicators of underlying biological processes or disturbances within the system.[15] Although mass spectrometry (MS) itself is not inherently quantitative, various strategies have been devised to leverage MS signals for comparative studies across multiple samples (Figure 6).

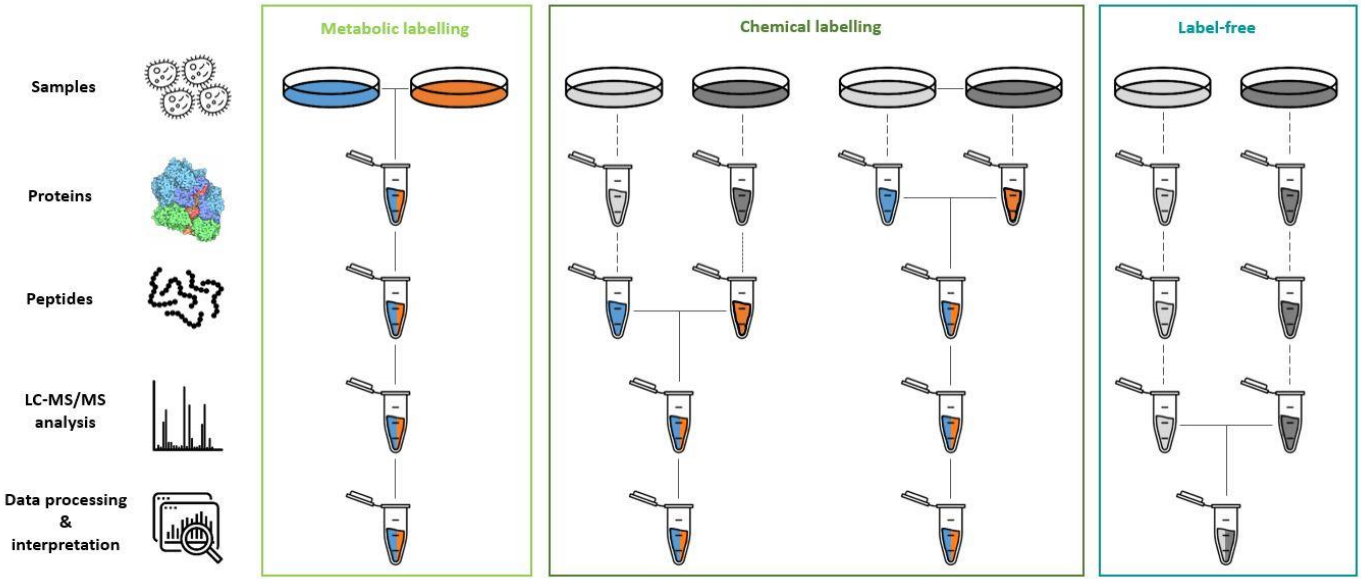


Figure 6: Overview of global quantification strategies in bottom-up proteomics (adapted from [100]). The light and dark gray Petri dishes and Eppendorf tubes represent the two conditions being compared without labeling, while orange and blue indicate labeled conditions. Tubes that are both orange and blue depict multiplexing of the two conditions. Dashed lines highlight steps where variations and errors in quantification may arise.

A. Label-free Relative Quantification Approaches

Quantitative proteomics has seen a growing reliance on label-free strategies in recent years, largely due to advances in mass spectrometry technology that have improved acquisition speed and sensitivity.[101] These methods bypass the need for stable isotope labeling and the associated complexities of evaluating labeling efficiency, offering a cost-effective and streamlined alternative. Furthermore, label-free quantification is compatible with diverse sample types, including cells, tissues, and biological fluids, and does not impose limits on the number of conditions that can be compared, as multiplexing is unnecessary.[102]

Since samples are processed and analyzed independently, variations at different stages of the workflow can influence the reliability of the results. To enhance precision and reproducibility, researchers often employ robust and stable LC-MS systems, analyzing all conditions within a single sequence.[103]

Two label-free quantification approaches are commonly used: spectral counting and extracted ion chromatogram MS1 (XIC-MS1).

i) Spectral Counting

The spectral counting approach operates on the premise that a protein's abundance is proportional to the number of MS2 spectra associated with it. One of its key advantages is its straightforward data processing, as it relies solely on the tools used for protein identification and validation.

Despite its simplicity, this method is hindered by the inherent variability and under sampling associated with the DDA acquisition mode, resulting in missing values and reduced reproducibility across samples. To address this, the dynamic exclusion setting, which prevents the repeated selection of precursor ions, should be minimized or disabled. This adjustment helps maintain the redundancy of spectra that reflects protein abundance; otherwise, it can introduce bias, particularly affecting small variations and low-abundance proteins.[102]

Another challenge lies in the relationship between protein length and the number of peptides, which influences the associated spectra. Small proteins (<20 kDa) are quantified less accurately than larger ones. To mitigate this issue, normalization strategies based on factors such as protein length, molecular weight, or the probability of detecting an MS2 spectrum have been proposed.[102] Additionally, the presence of shared peptides complicates quantification, especially if PSM spectra are attributed equally to all associated proteins. A more reliable approach is to distribute shared peptides proportionally among the proteins, accounting for the unique peptide distribution.[100,104] Lastly, to ensure reliable results, label-free quantification data should be normalized to account for variations introduced during the analytical workflow.

ii) Extracted Ion Chromatogram

The ion current extraction (XIC) technique relies on the assumption that peptide abundance is directly proportional to the chromatographic signal detected during mass spectrometry. XIC quantification is typically calculated using either the height or the area under the curve of the peptide peaks as they elute through the chromatographic column. These data are collected from the MS1 spectra, while the MS2 spectra are used primarily for peptide identification. This method typically uses data-dependent acquisition (DDA), which requires careful optimization of the method's parameters. The goal is to acquire sufficient MS1 spectra to define chromatographic peaks accurately, ensuring reliable quantification, while also obtaining high-quality MS2 spectra to maintain broad proteome coverage.

High-resolution and accurate mass (HR/AM) mass spectrometers are preferred for this approach, as they enable precise separation of precursor ion isotopes at specific m/z and retention times.[105]

Additionally, a stable and efficient chromatographic system is essential for accurate signal detection and quantification.[100]

Nevertheless, XIC requires more complex data analysis compared to spectral counting methods. To address potential variations introduced during sample preparation, chromatography, and signal instability, retention time alignment and data normalization are employed.[102,104,106,107] Several software tools have been developed to facilitate this process, such as Skyline[108], MaxQuant[109], Proline[110], and licensed solutions like SpectroMine™ and Proteome Discoverer.

Studies have compared the performance of these tools in processing quantitative proteomics data, and they play a crucial role in identifying and integrating chromatographic peaks for peptides, known as "features." These tools create LC-MS maps, providing details on m/z ratio, retention time, charge, and peak intensity. Following this, a second phase involves the alignment of retention times, peptide sequence identification, signal normalization, protein identification, and quantification.[100] The alignment process allows for the identification of peptides across different LC-maps from various experimental conditions, reducing missing values.[111] Alternatively, precursor ion lists can be generated through database searches of LC-MS/MS analyses. The resulting spectral library, which includes MS2 spectra of identified peptides, is then used to query their presence in the MS1 spectra, allowing for final XIC-based quantification.

B. Label-Based Relative Quantification Approaches

Label-based quantification methods rely on the principle that isotope-labeled (heavy) and unlabeled (light) peptides share similar physicochemical properties, such as chromatographic behavior (e.g., retention time) and mass spectrometric characteristics (e.g., ionization efficiency, fragmentation patterns, and MS signal response).

The only distinguishing feature between the two forms is their mass, with labeled peptides being heavier due to the isotopic labels. These strategies allow for the pooling of samples with different isotopes, enabling the comparison of multiple samples in a single acquisition for quantification. The labeling can be either metabolic or chemical, depending on the experimental setup. This approach simplifies the analysis and increases the throughput of quantification studies.

i) <u>Metabolic or Enzymatic Labelling Approaches</u>

Metabolic labeling is a method where stable isotopes are incorporated into proteins during cell culture, applied early in the proteomics workflow to reduce biases introduced by sample preparation, thus enhancing quantification accuracy. One widely adopted approach is SILAC (Stable Isotope Labeling with Amino Acids in Cell Culture)[112], which incorporates isotopically labeled amino acids like lysine and arginine (^{13}C or ^{15}N) into the cell culture. After harvesting, cell lysates from different conditions are combined in equal amounts before proceeding to the sample preparation for proteomics analysis. Quantification is achieved by comparing the MS signals of labeled and unlabeled peptides. However, SILAC has limitations, such as supporting only up to three conditions and being restricted to specific sample types, like cell cultures or SILAC mouse models.[113]

To address these limitations, new approaches have been introduced. Super-SILAC incorporates labeled internal standards to improve quantification across multiple sample[114], while NeuCode (Neutron enCoding) enhances multiplexing by using isotopic variants of lysine and combining them with other labeling techniques.[115] Despite these advancements, metabolic labeling strategies still face challenges in multiplexing, which can be expanded by adopting chemical labeling methods.

ii) Chemical Labelling Approaches

Unlike the previously discussed metabolic labeling strategies, chemical labeling methods rely on the stable isotope modification of specific reactive groups on peptides, such as the α- and ε-amino groups of lysine or the thiol group of cysteine. These approaches are particularly advantageous for their flexibility, as they can be applied to a wider range of sample types. Chemical labeling can be divided into two main categories: isotope tagging and isobaric labeling.

The first category, isotope tagging, was exemplified by the development of ICAT (Isotope Coded Affinity Tag) in 1999. ICAT uses reagents composed of a thiol-reactive group that targets cysteine residues, a stable isotope-labeled linker, and a biotin group. This setup enables enrichment of ICAT-labeled peptides via streptavidin affinity chromatography after the samples are combined.[116] While effective, ICAT is inherently limited to peptides containing cysteine and introduces potential issues such as altered chromatographic retention times due to deuterated labeling and MS signal interference from the biotin group. To overcome these limitations, modifications such as replacing deuterium with 13C for labeling or using cleavable ICAT reagents that remove biotin prior to MS analysis have been proposed.[102]

The second category involves isobaric labeling strategies, which target the N-terminal ends of peptides and proteins as well as the ε-amino groups of lysine side chains. These labels consist of a reactive group for primary amines, a cleavable group that produces reporter ions during fragmentation, and balance groups that ensure mass equivalence across labeled variants. Isobaric labeling enables multiplexed quantification by assigning distinct markers to different samples, with the relative abundance calculated from the intensity of reporter ions in the low m/z region of the MS2 spectrum. Meanwhile, peptide identification relies on the b- and y-ion fragments. Tandem Mass Tags (TMT) are one of the most widely adopted methods in this category, enabling the simultaneous analysis of up to 16 samples, with theoretical expansion to 18.[102,117] Another popular approach is iTRAQ (isobaric Tags for Relative and Absolute Quantification), which facilitates multiplexing of 4 to 8 samples while reducing spectrum complexity.[118,119] These technologies are commercially available through Thermo Fisher Scientific and Sciex, respectively.

Despite their benefits, isobaric labeling approaches are susceptible to challenges such as ratio compression caused by interference during co-isolation of precursor ions. To address this issue, advanced techniques like MS3-based isolation and fragmentation are employed to improve quantification accuracy.[102]

C. Label-Free “Absolute” Quantification Approaches

The goal of label-free "absolute" quantification strategies is to estimate protein quantities directly. These approaches offer the advantage of quantifying a large number of proteins while reducing costs and simplifying experimental workflows compared to labeling-based methods.[102]

One of the earlier methods developed for this purpose is emPAI (exponentially modified Protein Abundance Index). This technique builds upon the Protein Abundance Index (PAI)[120], which estimates protein abundance by calculating the ratio between the number of observed peptides and the number of theoretically observable peptides per protein.

In addition to emPAI, other strategies leveraging signal intensities have emerged. For instance, iBAQ ("intensity-Based Absolute Quantification") calculates protein abundance by dividing the sum of all peptide peak intensities by the number of theoretically observable peptides.[121]

Another widely used method is the Top3 approach, which incorporates an unlabeled reference protein spiked into the sample as a standard. Protein abundance is estimated by comparing the intensity of the three most abundant peptides of the target protein to the three most abundant peptides of the reference standard.[122] Notably, the Top3 strategy was applied in the work presented in this manuscript.

2. Targeted Quantification Strategies

Quantifying specific proteins with precision and reproducibility has become increasingly feasible through targeted proteomics strategies, which offer superior sensitivity and accuracy compared to global approaches.[123] Although data-dependent acquisition (DDA) methods can identify and quantify a broad spectrum of proteins, they are constrained by limited dynamic range, sensitivity, and reproducibility.[122] For this reason, targeted methods have emerged as the preferred choice for accurately measuring a limited number of proteins, typically in the range of a few dozen. The most well-established technique in this domain is Selected Reaction Monitoring (SRM), typically conducted using triple quadrupole (QQQ) mass spectrometers. Furthermore, recent innovations have enabled the successful adoption of targeted methods on high-resolution and accurate mass (HR/AM) instruments, such as Parallel Reaction Monitoring (PRM) performed on Q-Orbitrap systems.[124]

A. Selected Reaction Monitoring

The method known as SRM (Selected Reaction Monitoring), also referred to as MRM (Multiple Reaction Monitoring), is considered the gold standard in targeted proteomics.[102] This technique utilizes triple quadrupole (QQQ) mass spectrometers, where each quadrupole serves a distinct function: the first (Q1) isolates precursor ions, the second (Q2) acts as the collision cell for ion fragmentation, and the third (Q3) selects the fragment ions. For a successful SRM analysis, the analyst constructs an acquisition method based on a transition list, which specifies a set of precursor ions and their corresponding fragment ions. These paired ions are termed transitions. After isolating the precursor and fragment ions, ion chromatograms are extracted for each transition and then combined when they derive from the same precursor (Figure 7). By targeting both precursor and fragment ions, SRM delivers

highly sensitive and specific quantification. Additionally, increasing the number of transitions monitored enhances the overall specificity of the analysis.[123,125]

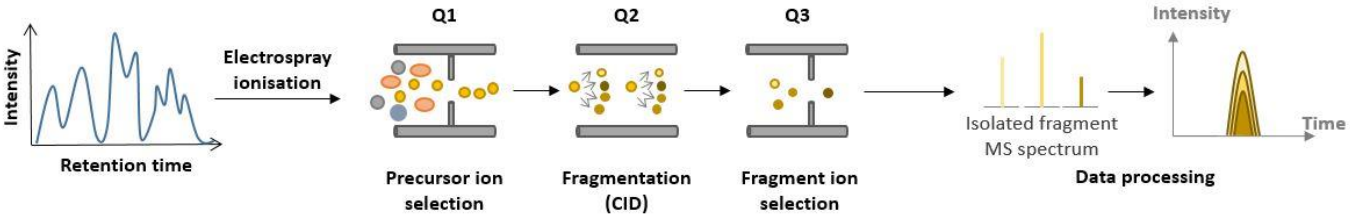


Figure 7: Illustration of the Selected Reaction Monitoring (SRM) principle.

B. Parallel Reaction Monitoring

The introduction of high-resolution, accurate-mass (HR/AM) mass spectrometers, such as the Q-Orbitrap and Q-TOF, has led to the development of an innovative targeted proteomics approach. This method, known as Parallel Reaction Monitoring (PRM) when used on a Q-Orbitrap instrument, offers an alternative to traditional SRM (Selected Reaction Monitoring).[124,125] In PRM, the predefined precursor ions are sequentially isolated using a quadrupole mass filter, which isolates ions based on their m/z ratio, and then fragmented in a collision cell. However, unlike SRM, PRM does not require the selection of specific fragment ions; instead, all fragment ions are measured simultaneously, generating full MS2 spectra. Ion chromatograms are then extracted for each precursor and fragment ion pair (transition) for analysis (Figure 8).

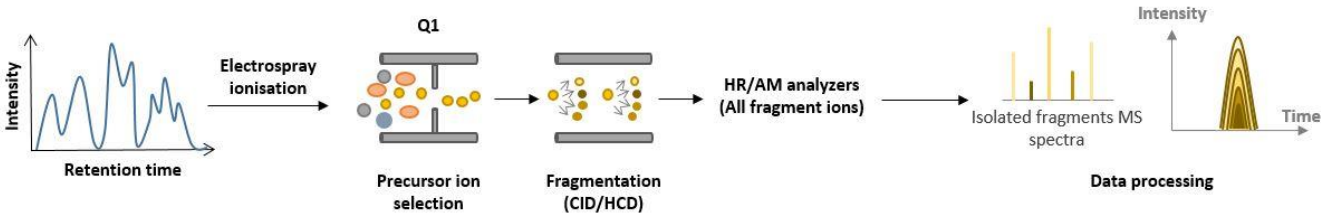


Figure 8: Illustration of the Parallel Reaction Monitoring (PRM) principle.

While both SRM and PRM share similar performance characteristics, such as reproducibility, accuracy, and quantification precision[18,95,124], PRM offers some distinct advantages. One of the main benefits of PRM is that it eliminates the need to predefine specific transitions to monitor, making the method easier to implement and optimize. Furthermore, the use of HR/AM instruments in PRM enhances selectivity, sensitivity, and dynamic range.[19,102,124] Another significant advantage of HR/AM mass spectrometers is their ability to perform both non-targeted discovery proteomics experiments (such as DDA) and targeted PRM analyses on the same instrument, facilitating method optimization and the transfer of key parameters across different experimental approaches.[126]

C. Absolute Quantification using Isotopic Dilution

i) Label-Based Absolute Quantification

To obtain absolute quantification of peptides and proteins, a known quantity of isotopically labeled peptide (AQUA[127], QconCAT[128]) or protein standards (PSAQ[129], FLEXIQuant[130], PrEST[131]) is added to the sample. These standards share the same sequence as the endogenous peptides or proteins, with the only difference being the incorporation of isotopic labels, which modify their mass. In proteomic studies using trypsin as a protease, peptides are commonly labeled at the C-terminal side of lysine and arginine with stable isotopes (^{13}C, ^{15}N), resulting in mass increases of 10.01 and 8.01 Da, respectively.[127,128]

More specifically, regarding the AQUA method, synthetic peptides are designed to mirror the native peptides, with stable isotopes incorporated at specific residues. These AQUA peptides behave identically to their endogenous counterparts, except for the mass difference due to the isotopic labeling. The labeled AQUA peptides are introduced into the sample alongside the target peptides, and both are analyzed in parallel using mass spectrometry. The intensity ratio of the labeled and unlabeled peptides allows for the accurate quantification of the target peptide. The AQUA method is widely used for its high accuracy and reproducibility in protein quantification, and was specifically employed for quantification in Part III of this manuscript.[132,133]

ii) Isotopic dilution

Isotopically labeled peptides are widely used for the precise detection and quantification of endogenous peptides of interest. There are two types of synthetic peptides available for this purpose, each serving different needs depending on the level of accuracy required for the experiment. Low-quality synthetic peptides, typically crude in nature, are often used for initial verification of the presence of target proteins or for relative quantification of peptides. These peptides usually have lower purity, and their quantities are not precisely known. However, they are relatively inexpensive, which makes them an accessible option for more basic applications.

On the other hand, high-quality synthetic peptides are employed when absolute quantification with high accuracy is required. These peptides are added at precisely known concentrations to the samples, ensuring accurate measurement of the target peptides. Due to the precision and higher quality of these peptides, they are significantly more expensive than their crude counterparts. As a result, their use is often reserved for studies where high accuracy and reproducibility are paramount, despite the increased cost.

For effective quantification, these isotopically labeled peptides are introduced into the sample alongside the endogenous peptides. The key to quantification lies in the intensity ratio between the labeled peptide and its endogenous counterpart, allowing researchers to calculate the concentration of the target peptide.

Quantification is typically performed through the extraction of ion chromatograms (XIC), as discussed in the previous section. To analyze and process the resulting data, specialized software tools have been

developed. Skyline, an open-access software, is one of the most widely used platforms due to its wide range of features and user-friendly interface, making data inspection and analysis much easier.

3. Data-Independent Acquisition Approaches

A. Principle of Data-Independent Acquisition

Proteomics aims to identify and quantify proteins in complex samples. While DDA-based approaches can detect and quantify thousands of proteins in a single experiment, they are limited by sensitivity, specificity, and dynamic range. In contrast, targeted approaches such as Selected Reaction Monitoring (SRM) and Parallel Reaction Monitoring (PRM) offer increased sensitivity, specificity, and dynamic range but are restricted to a limited number of target proteins. To overcome the limitations of both DDA and targeted methods, Data Independent Acquisition (DIA) has emerged as an approach that combines the broad proteome coverage of DDA with the high sensitivity, accuracy, reproducibility, and wide dynamic range characteristic of targeted methods.

DIA experiments are typically performed on high-resolution/accurate-mass (HR/AM) instruments, including Q-TOF or Q-Orbitrap systems. The key advantage of DIA is its ability to theoretically collect MS/MS spectra from all peptides present in the sample throughout the chromatographic gradient. This is achieved by dividing the mass range into specific windows, within which precursor ions are co-isolated and co-fragmented. These fragments are then co-analyzed, generating multiplexed MS2 spectra. From these multiplexed spectra, chromatographic peaks are extracted to facilitate the identification and quantification of peptides (and proteins), as observed in Figure 9 below.[19,134]

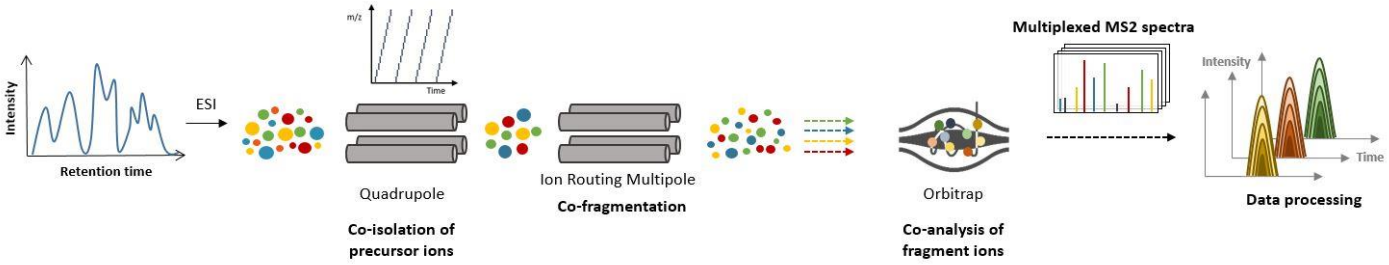


Figure 9: Principle of Data-Independent Acquisition (DIA) mode. Schematic representation of DIA strategies utilizing isolation windows.

The rise of DIA has been further fueled by significant advancements in mass spectrometry technology, particularly the development of high-resolution, fast-scanning instruments. In addition, the design of specialized bioinformatics tools for DIA data processing has played a key role in making DIA an increasingly viable option for quantitative proteomics. These technological and computational improvements have enhanced the ability to perform comprehensive proteome analysis. The success of DIA is evidenced by dedicated sessions hosted annually by the American Society for Mass Spectrometry (ASMS) and by the growing integration of DIA into diverse proteomics workflows.[96,98]

B. Evolution of DIA Approaches

The origins of Data Independent Acquisition (DIA) can be traced back to the early 2000s. In 2000, Masselon et al. conducted one of the first DIA-related experiments by characterizing several polypeptides using multiplexed MS/MS spectra generated on a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (FT-ICR). This approach leveraged the high mass measurement accuracy of the FT-ICR to analyze complex mixtures of peptides.[135]

Building on this foundation, Purvine et al. introduced the concept of "shotgun CID" in 2003. This method utilized a two-step in-source fragmentation process: the first step involved a low voltage setting to limit fragmentation and acquire precursor ion mass spectra, while the second step applied a higher voltage to induce peptide fragmentation, producing MS/MS spectra of the resulting fragment ions.

The term "data-independent acquisition" itself was formally introduced by Venable et al. in 2004. In their groundbreaking work, they employed a linear ion trap mass spectrometer to sequentially isolate and fragment precursors within large 10 m/z isolation windows. This approach was used to analyze yeast samples and marked a significant milestone in the evolution of DIA methodologies.[136]

These early studies laid the groundwork for the rapid development of diverse DIA-based strategies. Over the years, various methods have been implemented across different mass spectrometers, with approaches differing in MS acquisition parameters and data processing strategies. Broadly, these can be grouped into two main categories: strategies that fragment ions across the entire mass range and those that utilize defined isolation windows.

Table 3 provides a comprehensive overview of these methods, highlighting their unique features and historical progression. Some key approaches of particular relevance are detailed in subsequent sections.

DIA method	Year of Introduction	Precursor isolation window size (m/z)	Reference
Shotgun CID	2003	Full range	Purvine et al.[137]
DIA (original)	2004	10	Venable et al.[136]
MS^E	2005	Full range	Silva et al.[138]
PAcIFIC	2009	2,5	Panchaud et al.[139]
AIF	2010	/	Geiger et al.[140]
XDIA	2010	20	Caravalho et al.[141]
FT-ARM	2012	100	Weisbrod et al.[142]
SWATH	2012	25	Gillet et al.[143]
$HDMS^E$	2012	Full range	Geromanos et al.[144]
MSX	2013	4	Egertson et al.[145]
pSMART	2014	5-20	Prakash et al.[146]
WiSIM-DIA	2014	12	Martin et al.[147]
$UDMS^E$	2014	Full range	Distler et al.[148]
SWATH (variable windows)	2015	8-85	Zhang et al.[149]
HRM	2015	24-220	Bruderer et al.[150]
SONAR	2018	24	Moseley et al.[151]
BoxCar DIA	2018		Meier et al.[152]
Scanning SWATH	2019	5	Messner et al.[153]
diaPASEF	2019	25	Meier et al.[154]
PulseDIA	2019	4	Cai et al.[155]
DIA-FAIMS	2020	13.7	Bekker-Jansen et al.[156]

Table 3: Summary of the evolution of DIA approaches (adapted from [134,157]).

i) DIA Strategies performed over the entire mass range

In 2005, Waters introduced MS^E, a method designed to exploit the technological advancements of Q-TOF instruments.[138] This approach allowed simultaneous acquisition of MS1 and MS2 spectra over the full mass range during a chromatographic run. By alternating between low and high collision energies, MSE captured both precursor ion and fragment ion spectra, providing comprehensive data for proteomic analyses.

Expanding on the principles of MS^E, Thermo Fisher Scientific launched All-Ion Fragmentation (AIF) in 2010.[140] Designed for the Exactive instrument, AIF sequentially recorded MS1 spectra of precursor ions and MS2 spectra generated through HCD fragmentation. Unlike MSE, AIF did not isolate precursor ions before fragmentation, enabling a more inclusive analysis of complex samples.

Advancements in the MSE framework came with the introduction of High-Definition MS^E ($HDMS^E$) in 2012[144] and Ultra-Definition MS^E ($UDMS^E$) in 2014[148]. HDMSE incorporated ion mobility as a supplementary separation dimension, significantly reducing the complexity of fragment ion spectra. UDMSE further improved on this by employing variable collision energies tailored to ion mobility data,

optimizing fragmentation efficiency based on precursor ion characteristics and enhancing data quality across the board.

ii) DIA Strategies based on isolation windows

Most Data-Independent Acquisition (DIA) methods rely on the use of isolation windows, which can vary in width, being either fixed or adaptable (Figure 10).

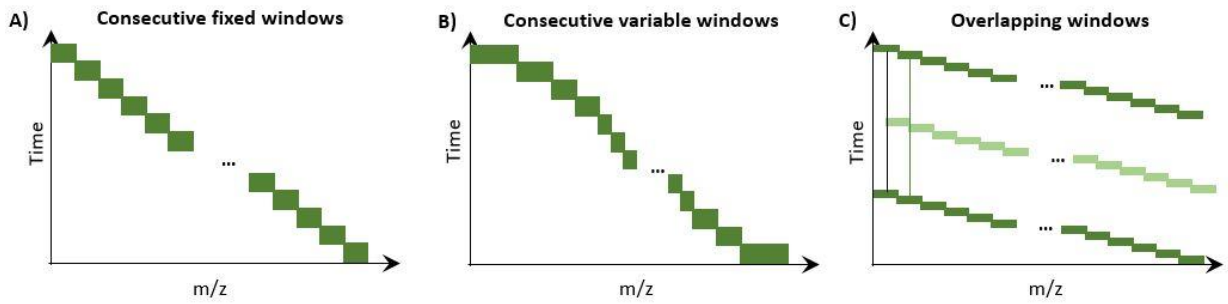


Figure 10: Overview of the main DIA approaches based on isolation windows. (A) Consecutive fixed windows, (B) consecutive variable windows, and (C) overlapping window isolation scheme.

Initially, fixed-width isolation windows were widely employed. The PAcIFIC approach[139], for example, utilized narrow 2.5Th isolation windows to fragment precursor ions across multiple LC-MS/MS runs, covering the entire mass range. Although this method provided extensive data, it required days of analysis to complete. More recently, PulseDIA[155] was introduced to significantly reduce analysis time by employing a reduced number of isolation windows with widths optimized to the ion density.

Another strategy involved the use of wider isolation windows, such as the 25Th windows implemented in the SWATH method[143] for Q-TOF instruments. Commercialized by Sciex, SWATH demonstrated that DIA could achieve quantification accuracy comparable to SRM, spanning a dynamic range of four orders of magnitude. This technique became one of the key innovations driving the adoption of large-scale DIA-based proteomics.

Further refinements to the SWATH approach included variable-width isolation windows, first proposed by Zhang et al. in 2015[149]. This adjustment addressed the uneven distribution of precursor ions across the chromatographic gradient, which could result in overcrowded MS2 spectra in certain windows and sparse spectra in others. By tailoring the window size to the ion density at different m/z ranges, the method achieved more consistent precursor coverage, leading to improved quantification accuracy and a higher rate of protein identification. A similar strategy, hyper-reaction monitoring (HRM)[151], was developed for Q-Orbitrap instruments.

Ion mobility technology has also been integrated as an additional separation dimension in high-resolution/accurate-mass (HR/AM) instruments, further enhancing DIA performance. Notable examples include FAIMS-DIA[158] for next-generation Q-Orbitrap and tribrid systems (Thermo Fisher Scientific) and diaPASEF[154] for Q-TOF platforms, specifically implemented on the TimsTOF Pro (Bruker).

Building on diaPASEF, advanced methods like Slice-PASEF[159] and midiaPASEF[148] are now being explored.

Another innovation, the BoxCar method[152], was developed for Q-Orbitrap instruments to boost the signal-to-noise (S/N) ratio and expand the dynamic range of precursor ions. By filling multiple narrow m/z windows during ion injection, this approach improved sensitivity and data quality. Its evolution, BoxCarmax[160], has been adapted for the analysis of SILAC isotope-labeled peptides, which are challenging to integrate with traditional DIA techniques.

Finally, overlapping window strategies have also been explored. These approaches utilize isolation windows that partially overlap with preceding windows, enhancing the selectivity of the analysis. One such method, SONAR[151], was introduced for Waters Q-TOF systems in 2018. Another approach, Scanning SWATH[153], replaced the acquisition of successive isolation windows with continuous scanning of the filtering quadrupole, thereby reducing the DIA cycle time.

C. Development of DIA Methods

Optimizing Data-Independent Acquisition (DIA) methods involves fine-tuning various parameters such as isolation window size, the total number of windows, cycle time, and the ion load for analysis. These adjustments are crucial as they directly impact the sensitivity, specificity, proteome coverage, and overall quantification precision.

The sensitivity of a DIA experiment depends heavily on the number of ions reaching the Orbitrap analyzer. This ion load is controlled by two factors: the automatic gain control (AGC) target and the maximum injection time. The AGC target determines the desired number of ions to accumulate in the C-trap before transferring them to the Orbitrap for mass measurement. If the target is not met within the specified timeframe, the transfer occurs after the maximum injection time elapses. Balancing these parameters is key to ensuring sufficient ion injection for sensitivity without unnecessarily extending acquisition cycles.

Cycle time, a pivotal parameter in DIA workflows, dictates the number of data points collected across chromatographic peaks. For accurate quantification, it is generally recommended to capture at least 7–8 data points per peak. For example, if a chromatographic peak lasts 20 seconds, the cycle time must not exceed 2.5 seconds to ensure adequate sampling. The calculation of the cycle time depends on the transient times of the MS1 and MS2 scans:

$$Cycle\ time = MS1\ transient\ time + (Number\ of\ isolation\ windows * MS2\ transient\ time)$$

Transient time, in turn, is directly linked to the instrument's resolving power. Higher resolving power, which enhances ion separation accuracy, also increases the transient time and consequently prolongs the cycle time. On the Orbitrap Eclipse Tribrid system, for instance, resolving powers of 15,000, 30,000, and 120,000 (at 200 m/z) correspond to transient times of 32 ms, 64 ms, and 256 ms, respectively.

The mass range analyzed during a DIA experiment and the size of the isolation windows together determine the number of windows needed in a given cycle. These parameters are often adjusted based

on prior DDA analysis to achieve the desired balance between proteome coverage and sensitivity. Additionally, overlapping isolation windows, typically by 0,5 – 1Da, can help mitigate reduced quadrupole transmission efficiency near window boundaries, improving signal consistency and sensitivity across the analyzed mass range (Figure 11).[143,161]

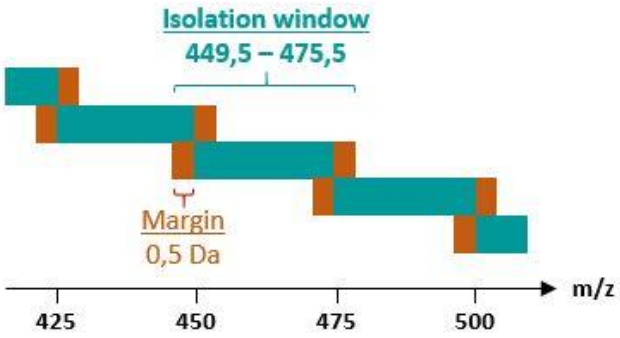


Figure 11: Illustration of the quadrupole isolation window overlap.

D. Challenges in Data Processing

DIA strategies offer high reproducibility in quantification with minimal missing values. However, the complexity of the multiplexed MS2 spectra, which contain fragments from co-isolated and co-fragmented precursor ions, requires the use of specialized bioinformatics tools for DIA data processing. Conventional tools designed for DDA are not suitable for this purpose.[134,162] There are two main approaches used for processing DIA data: the peptide-centric approach and the spectrum-centric approach (Figure 12).

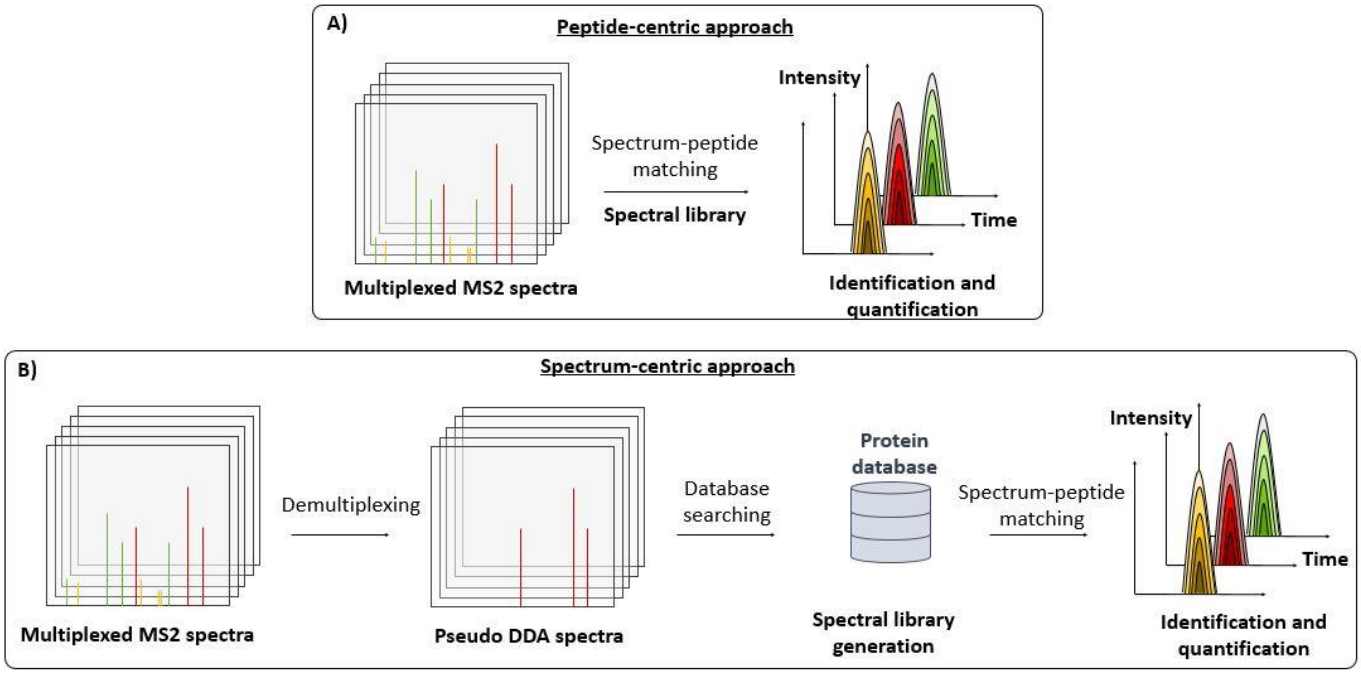


Figure 12: DIA data processing strategies (adapted from [163]). The peptide-centric approach utilizes a previously generated spectral library to search for peptide presence in DIA-MS2 spectra. In contrast, the spectrum-centric approach identifies peptides directly from DIA spectra by generating pseudo DDA spectra and performing a classical database search. The resulting assigned spectra are used to create a spectral library, which is then used for further signal extraction via the peptide-centric approach.

i) Peptide-Centric Approach

The peptide-centric approach relies on a spectral library to search and match DIA data. A spectral library contains various details, including the m/z of precursor and fragment ions, normalized retention time, precursor charge state, and fragment ion intensities.[163] Additional information, such as ion mobility drift time, can also be integrated into the library.[101]

In DIA peptide identification, the dataset is compared to peptides present in the library based on these features. A significant drawback of this approach is that only precursors within the library can be identified, meaning a comprehensive and exhaustive library is essential for optimal results. Building such a library typically involves fractionated DDA analysis of samples similar to those in the DIA experiment, using traditional search engines like Mascot, Andromeda, or Sequest. However, this process is time-consuming, requiring large sample volumes and matching experimental conditions (gradient length, fragmentation mode) to ensure library quality. An alternative to this laborious task is to use publicly available spectral libraries, such as SWATHAtlas[164], PeptideAtlas[165], or PRIDE[166], although these are often limited in scope and sample variety.

Numerous tools are available for searching MS2 DIA data against spectral libraries for peptide identification and scoring. These include OpenSWATH[167], PECAN[168], EncyclopeDIA[169], and Spectronaut[150]. These tools use library information to identify peptides and assign scores to the peak

groups based on the extracted ion chromatogram (XIC) at the MS2 level. Validation of peptide identifications typically uses the target-decoy strategy. More recently, algorithms integrated into Spectronaut and EncyclopeDIA have introduced methods for deconvoluting DIA data to generate hybrid libraries by combining DDA data with DIA data.

Advances in artificial intelligence (AI) and machine learning (ML) have further improved spectral library generation. Predictive algorithms such as DeepMass[170], Prosit[171], and pDeep[172], DeepLC[173] and MS^{2}PIP[174] are capable of predicting retention times and MS2 fragmentation patterns for peptides in a given database. This expands the library beyond the peptides identified through DDA, incorporating all theoretical peptides from the database[175]. ML-driven approaches, such as DIA-NN[176] or the Spectronaut software by Biognosys, also enhance peptide identification and improve DIA data processing. Regardless of the approach or tool used, the quality of the spectral library remains a pivotal factor in the success of any DIA experiment.

ii) <u>Spectrum-Centric Approach</u>

Spectrum-centric methods focus on deconvolving multiplexed MS2 spectra into pseudo-MS/MS spectra, which can then be directly searched against protein databases. These approaches are often more convenient, as they eliminate the need to generate spectral libraries. This approach was first proposed in 2003 by Purvine et al.[137], who reconstructed pseudo-DDA spectra from analyses performed at both low and high voltages. By leveraging the similar chromatographic properties of precursor and fragment ions, they were able to manually identify the peptides.

Since then, several tools have been developed to enable library-free DIA data processing. One notable advancement came with the introduction of DirectDIA™, integrated into the Spectronaut software (Biognosys), which leverages the Pulsar search engine to perform DIA data analysis without the need for a pre-existing spectral library. More recently, the Demichev group developed the DIA-NN[176] software, an algorithm that utilizes deep neural networks to facilitate both peptide-centric and spectrum-centric DIA data analysis, further enhancing the accuracy and efficiency of DIA workflows

Chapter 3
Monoclonal antibodies and Host Cell Proteins

1. Understanding Monoclonal antibodies (mAbs)

The defining characteristic of a monoclonal antibody (mAb) is its precise specificity, it binds exclusively to a single antigenic site, otherwise referred to as an epitope, on a particular molecule, making it monospecific.[177] The field of therapeutic mAbs has experienced remarkable advancements, establishing them as current vital components in the modern biopharmaceutical market. The concept of employing antibodies for targeted therapy, initially proposed by Paul Ehrlich over a century ago, gained recognition with the development of hybridoma technology in 1975, which enabled the efficient production of these agents.[178,179]

Early mAbs, primarily derived from murine sources, presented many challenges, including immunogenic responses in humans and limited activation of human immune mechanisms. However, innovations in antibody engineering have led to the creation of fully humanized mAbs, greatly enhancing their therapeutic applicability and even expanding their use in treating a range of diseases, more particularly, cancer, where they can effectively target tumor-associated antigens and initiate robust immune responses from patients.[180]

The therapeutic potential of mAbs has gained increasing recognition with over 400 mAbs currently in clinical development, with nearly half being dedicated to oncology.[181] The past decade has particularly seen a rise in interest, driven by the demand for therapies that specifically target disease processes while minimizing damage to healthy tissues.

A. Overview of mAb production processes

Manufacturing processes for mAbs have also evolved significantly since the approval of the first therapeutic mAb in 1986. The growing demand for these therapies has led to the establishment of large-scale production facilities and advancements in cell culture techniques, which have resulted in higher product yields and reduced production costs.[179] These improvements not only helped meet the current market needs but also contribute to making these important therapies more accessible to patients. As mAb production continues to develop, focusing on process improvements and facility design will be essential to adapt to technological advancements and the changing demands of the healthcare market.[180]

i) Expression systems and their implications

The production of mAbs is intrinsically linked to the selection of appropriate expression systems, crucial for eliciting robust immune responses against specific antigens. This process typically begins with the immunization of laboratory animals, predominantly mice and rats, using various antigens such as cellular components, purified proteins, peptides, carbohydrates, lipids, or nucleic acids. The purity

of these immunizing agents is critical; impure preparations can lead to contamination issues that may dominate the immune response, particularly when whole cells are used as immunogens. [182]

To enhance production levels, a range of expression systems is employed for mAbs, including yeast, bacteria, fungi, insect cells, mammalian cell lines, and even transgenic plants and animals. Nonetheless, mammalian cells are predominantly used for approved mAbs due to their ability to carry out post-translational modifications that closely mirror those occurring in human cells.[183,184] Chinese hamster ovary (CHO) cells, have emerged as the predominant system for producing recombinant therapeutic proteins. These cells excel in proper protein folding, assembly, and post-translational modifications, yielding superior protein quality and efficacy compared to other hosts. Notably, CHO cells account for approximately 70% of all recombinant proteins produced today, underscoring their significance in biopharmaceutical manufacturing. [185,186]Standard production processes involving CHO cells typically last between 7 and 14 days, with monoclonal antibody titers that can reach up to 10-13 g/L under optimized conditions.[187]

Despite historical challenges associated with mammalian cell cultures—such as low yields and medium complexity—recent advancements have significantly addressed these issues. The productivity of mammalian cell cultures in bioreactors has achieved impressive levels, with improved media compositions and process control leading to substantial enhancements. These advancements have enabled production cell lines to achieve specific productivities exceeding 20 picograms per cell per day and to maintain cell densities over 20 million cells/mL in fed-batch processes.[188]

As the demand for antibody therapies continues to escalate, the capacity for large-scale manufacturing becomes increasingly critical. Many companies have constructed large-scale manufacturing facilities equipped with multiple bioreactors, each capable of handling volumes of 10,000 liters or more, to meet the growing clinical and commercial needs.[177]

In summary, the reliance on CHO cells underscores the complexity of producing therapeutic mAbs. The continuous refinement of these systems reflects a commitment to improving productivity and quality, thereby ensuring that biopharmaceutical manufacturing evolves to meet the increasing demands of the industry.[189]

ii) Purification Process and Quality control

The production of mAbs is a highly complex and multi-stage process that involves careful optimization at both the upstream and downstream stages to ensure high yield, purity, and safety for biotherapeutic applications. Given the critical role of mAbs in clinical therapies, the production process must adhere to stringent regulatory requirements from health authorities, ensuring that every batch of antibodies is both safe and efficacious.[190]

The upstream process begins with the selection of a suitable mammalian cell line, which is essential for the efficient production of the target antibody. Cell line development, including cloning, transfection, and screening, ensures that the chosen host cells are capable of robust expression. Once the cell line is established, the process moves on to optimization of the growth media, feeding

strategies, and bioreactor conditions to support high-density cell cultures. These factors contribute directly to the overall yield of recombinant antibody produced in the system.[191] Additionally, controlling the accumulation of by-products and impurities like host cell proteins (HCPs), DNA, and endotoxins during the production phase is crucial for preventing unwanted interference during downstream purification.[186]

After the upstream phase, where the target protein is produced and secreted into the culture medium, the focus shifts to downstream process (DSP), which is essential for isolating and purifying the mAb from the host cell-derived contaminants that were mentioned previously (Figure 13). The first step in the DSP cascade is clarification, where the harvested cell culture fluid (HCCF) undergoes filtration or centrifugation to remove large particulates and cell debris, resulting in a clarified supernatant. This step is critical to ensure that the sample entering the subsequent purification stages is free of impurities that could interfere with the efficacy and safety of the mAb.[192] The clarified supernatant is then passed through the capture stage, which typically employs Protein A affinity chromatography to isolate the mAb. Protein A is usually favored due to its high specificity for binding the Fc region of antibodies, allowing for efficient separation from other components.[193] It also offers the advantage of high selectivity and yields a product with high purity after a single step, often exceeding 95% purity. However, despite its effectiveness, the high cost of Protein A resins remains a significant continuous challenge, and as such, biomanufacturers often strive to optimize the number of cycles performed per column to balance cost and efficiency.[194]

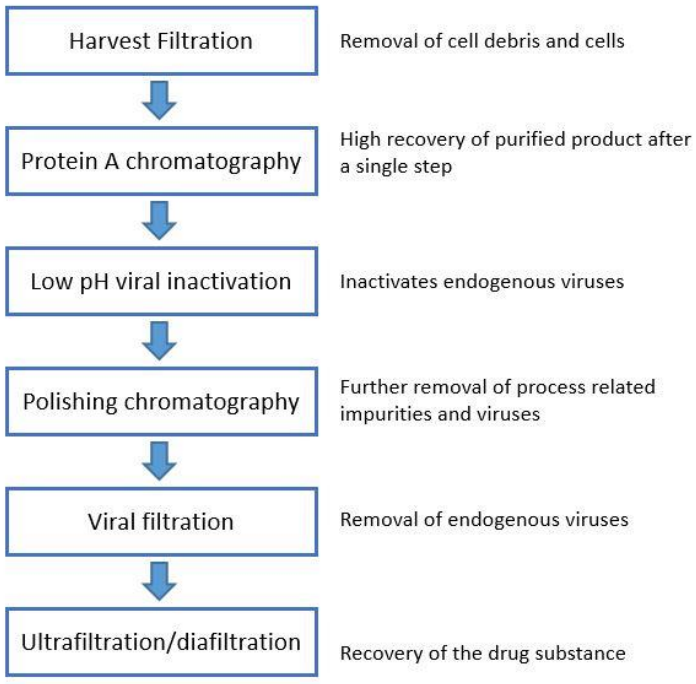


Figure 13: Downstream Process Cascade steps. This figure illustrates the key general stages involved in the DSP of mAbs, from the initial harvesting of the cell culture supernatant to the final purification of the mAb product. [193]

Following Protein A capture, the mAb undergoes multiple polishing steps to further refine its purity. Both cation and anion exchange are typically employed to achieve the desired separation, based on the mAb's charge properties and the characteristics of the contaminating species. In addition to ion exchange, techniques such as hydrophobic interaction chromatography (HIC) and mixed-mode chromatography may be employed to target specific impurities that may not be efficiently removed from prior steps.[187,195] These polishing steps are essential for ensuring that the final mAb product meets the high standards required for clinical use, the removal of aggregates, high molecular weight species, and product variants, such as glycosylation isoforms, which is crucial to ensuring the mAb's efficacy and minimizing potential immunogenicity.[196]

To strengthen even more the safety of the final product, viral clearance steps are usually further incorporated into the purification process. These typically involve the use of low pH treatment which serves to inactivate and remove any potential viruses, as well as viral filtration to physically remove viral particles from the final mAb preparation.[187] Viral clearance is a critical step, as any residual viral contaminants can pose serious safety risks to patients. After polishing and viral clearance steps, the final mAb product is subjected to concentration and diafiltration to remove residual small-molecule impurities and to adjust the formulation into an appropriate buffer for storage or final administration. This step ensures that the mAb is in its final, stable form, suitable for long-term storage and clinical application.[192,193]

At this stage, the purified mAb is typically stored either frozen or as a liquid, depending on the formulation requirements, which ensures the long-term stability of the product. The final formulation, which may include stabilizers, salts, and other excipients, is prepared to preserve the mAb's biological activity during storage and distribution. The mAb is then ready for distribution to pharmaceutical manufacturers for final packaging.[187]

The entire downstream process, from capture to final formulation, actually requires continuous tight control and careful monitoring to ensure that every batch meets the required specifications for purity, potency, and safety.[196] Rigorous quality control (QC) measures are implemented at each stage, in line with stringent regulatory guidelines. Regulatory authorities tend to set very tight limits on the impurities, with HCP levels typically required to be less than 100 ppm, residual DNA less than 10 ng/dose, and aggregates should not exceed 5% of the total product.[196] The use of advanced analytical methods such as liquid chromatography-mass spectrometry (LC-MS), size exclusion chromatography (SEC), and enzyme-linked immunosorbent assays (ELISA) are common in QC testing to monitor these parameters and verify the quality of the final mAb product.[197]

As the industry advances, the focus has now shifting toward improving scalability and cost-effectiveness in mAb production. Process optimization in late-stage production often involves implementing higher-capacity resins, continuous capture technologies, and novel buffer management strategies to accommodate higher titers and achieve more efficient large-scale production. The continued development of integrated and continuous production systems, which reduce processing

time and increase overall throughput, is expected to significantly enhance the efficiency and cost-effectiveness of mAb industry.[197]

2. Host Cell Protein impurities

Host Cell Proteins, otherwise referred to as HCPs, represent a significant type of process-related impurities in the production of mAbs. Produced by the cellular machinery of host organisms during recombinant protein expression, HCPs are endogenous proteins that can number in the hundreds to thousands within a single expression system, depending on the specific host and production conditions.[198,199] Their presence in therapeutic formulations poses potential risks to patients, including adverse immune reactions and reduced drug efficacy, underscoring the importance of comprehensive removal and monitoring strategies throughout the biomanufacturing process.[4,200]

A. Challenges in HCP quantification and analysis

The analysis and removal of HCPs during the production of recombinant therapeutic proteins is a complex and challenging task. HCPs are a diverse group of proteins that vary greatly in their physicochemical and immunological properties, making it difficult to predict their behavior and presence in the final product.[4,201] Despite extensive purification efforts, some HCPs can co-purify with the therapeutic protein due to similar physicochemical properties, or due to direct interactions with the therapeutic itself.[202,203] This adds a layer of complexity to the purification process, as it requires both precise analytical techniques and a deep understanding of the molecular interactions at play.

In addition to their low abundance in final products, the diversity of HCPs and their unpredictable behavior during production make it difficult to standardize methods for their detection and removal. The expression system used, whether it is mammalian or microbial, and the specific characteristics of the recombinant protein, can all influence the types and quantities of HCPs that are present at various stages of the process.[204] This variability means that HCP profiles are often dynamic and not fully understood until the process is tested in practice.[199] As a result, continuous monitoring is required throughout the manufacturing process, and the detection of HCPs must be sensitive enough to identify even the smallest trace amounts. Regulatory guidelines from organizations like the FDA require HCP levels to be reduced to trace amounts, and validated methods must be employed to ensure compliance.[205]

Achieving consistent removal of HCPs is not only essential for ensuring product safety but also serves as a benchmark for assessing the robustness of the bioprocess itself. A process that is capable of reliably removing HCPs reflects a well-controlled, efficient manufacturing system.[4] While the identification of individual HCPs can help assess potential risks, such as immunogenicity, and inform process adjustments, the dynamic nature of these contaminants means that monitoring must be ongoing, and adjustments must be made as necessary.[198] As biopharmaceutical manufacturing evolves, improvements in analytical techniques, such as more sensitive detection methods, continue to enhance the ability to monitor and control HCP levels. Nonetheless, the challenge of ensuring consistent HCP clearance remains an ongoing issue in bioprocessing, requiring both technical

innovation and careful management to guarantee the safety and efficacy of biotherapeutic products.[206]

B. Monitoring HCPs in Biopharmaceuticals

i) ELISA assays

ELISA (enzyme-linked immunosorbent assay) is one of the most widely used technique for detecting and quantifying HCPs in biopharmaceutical products. By using polyclonal antibodies, it can detect a broad range of proteins with high sensitivity and low cost, making it ideal for process monitoring and quality control.[207] This method is highly valued in the biopharmaceutical industry for its ability to measure HCP levels in a fast and high-throughput manner, ensuring that they are reduced to trace amounts in compliance with regulatory standards, thereby minimizing patient risk and enhancing drug stability.[208]

Under this context and application, ELISA's sandwich format is particularly advantageous, illustrated in Figure 14. This approach typically utilizes capture antibodies immobilized on a plate to bind HCPs, which are then detected by enzyme-conjugated antibodies, producing a colorimetric signal that is proportional to the HCP concentration.[4] Polyclonal antibodies are preferred because they offer broad reactivity against the diverse array of HCPs produced by host cells. The assay's high throughput capability, sensitivity, and specificity make it essential for tracking HCP impurities across various stages of production. [209]

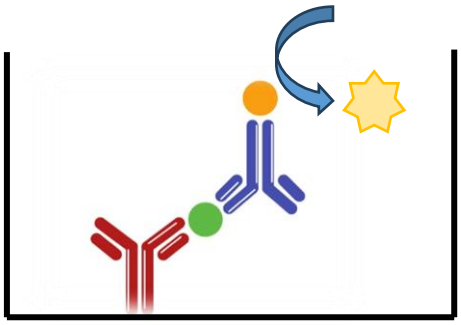


Figure 14: Principle of a Sandwich ELISA. Schematic representation of the Sandwich ELISA technique, where a capture antibody immobilized on a solid surface binds to the target antigen in the sample. After washing, a detection antibody conjugated to an enzyme binds to the antigen, forming a complex. The enzyme reacts with a substrate, generating a signal proportional to the antigen concentration.

Different types of ELISA kits are used depending on the clinical stage and also the pharmaceutical product being manufactured. Generic or commercial ELISA kits are frequently applied in early development and are designed to target HCPs. [210] These kits offer convenience and accessibility since they are "off-the-shelf" and provide a general assessment of HCP levels, but they do not account for the specific conditions or HCP profiles of individual production processes.[208] For more targeted monitoring, especially as development progresses, process-specific ELISAs are created to reflect the unique HCP profile of the particular cell line and production process used for the drug, providing more accurate results.[205]

Developing a process-specific ELISA involves producing anti-HCP antibodies using an HCP pool derived from the null cell line, which represents the host cells without the therapeutic gene. This approach allows for an antibody response specific to the most probable HCPs encountered during drug production, making it better suited for process monitoring and clearance validation.[211] The polyclonal anti-HCP antibodies are typically generated by immunizing animals, such as goats or rabbits, the resulting antibody mixture should ideally recognize a wide range of HCPs, taking into account the individuality of HCPs which can vary widely in antigenicity and concentration.[212]

Despite the advantages it offers, ELISA-based HCP detection and quantification also has major limitations. Indeed, a major challenge is the variability in antibody response; some HCPs may elicit a weak or no immune response, leading to under-detection. Conversely, highly immunogenic HCPs might provoke an exaggerated antibody response, resulting in over-detection of certain proteins.[213] Additionally, commercial kits, often used in the beginning of the drug development process, may lack the specificity needed to detect unique HCP profiles of different production processes, which can affect the assay's sensitivity and reliability.[209] What is more, while commercial kits are useful for broad applications, process-specific ELISAs are often required for more accurate quantification in later stages of product development, which also translates to a very important cost and time investment of the biopharmaceutical company.[205]

In addition, the accuracy of an HCP ELISA depends on multiple factors, including the choice of the host animal used for antibody production, the quality of the HCP pool used for immunization, and the ability of the antibodies to recognize the diverse array of proteins in the host cell proteome. These aspects must be optimized to ensure the polyclonal antibodies detect all relevant HCPs present in the product.[205,210,214] However, even with optimized anti-HCP antibodies, as mentioned previously, HCP ELISAs are inherently limited by their reliance on antibody-antigen interactions, which may not capture all HCPs present due to variable immunogenicity among the proteins.[4] As a result, ELISA may sometimes fail to detect low-abundance HCPs, especially when they co-purify with the therapeutic product due to similar physicochemical properties.[212] These limitations underscore the need for complementary analytical methods in cases where precise characterization of HCP content is crucial for regulatory submissions and in general to ensure product safety and efficacy.

ii) 2D Electrophoresis and Western Blotting

Under this context of HCP detection, as mentioned in the previous section, ELISA is widely considered the gold standard. However, the complexity and diversity of HCPs necessitate thorough validation to ensure the assay's specificity and coverage. In order to achieve this, ELISA is often paired with orthogonal techniques, more particularly 2D electrophoresis combined with Western blotting. This approach helps address the challenge of detecting a broad range of potential HCPs, ensuring that the polyclonal antibodies used in ELISA assays effectively cover a diverse spectrum of contaminants.[210,214]

The principle behind this coupling lies in the ability of 2D electrophoresis to separate proteins based on molecular weight and isoelectric point, generating a highly detailed protein profile.[215] In the subsequent Western blotting step, proteins are transferred onto a membrane, where the interaction

and binding between HCPs and polyclonal anti-HCP antibodies is revealed through chemiluminescence, as further illustrated in Figure 15. This process enables the calculation of a "coverage score," which reflects the ratio of proteins detected by the antibodies to the total proteins in the sample, offering a robust assessment of antibody coverage. Thus, this pairing enhances the reliability of ELISA by providing a qualitative measure of antibody reactivity, reinforcing its quantitative findings. [207,211,216]

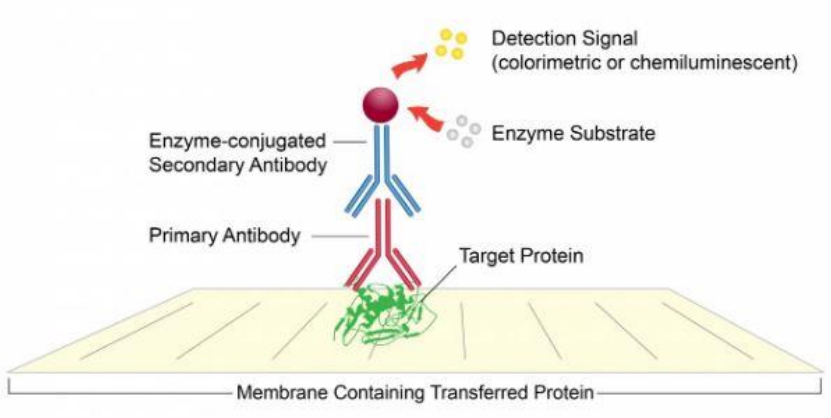


Figure 15: Schematic representation of a Western Blot technique (from leinco.com).

Despite its strengths, the combination of 2D electrophoresis and Western blotting is not without its limitations. The technique is prone to various technical limitations, such as potentially incomplete HCP resolution, complex spot patterns resulting from post-translational modifications of the same HCP producing several spots, protein loss during transfer, incomplete separation, and SDS-induced denaturation that may disrupt conformational epitopes.[217] Furthermore, issues like overlapping or co-migrating spots on the gel complicate the accurate visualization and quantification of individual HCPs. These challenges, coupled with the labor-intensive nature of the technique, make it unsuitable for routine, high-throughput applications or in-process monitoring. [4,205]

Recent advancements, however, have made significant improvements to the 2D electrophoresis-Western blotting workflow. Innovations such as employing stable pre-cast gels, next-generation imaging systems, and advanced image analysis software have helped reduce variability and increase reproducibility. Despite these enhancements, the method remains better suited as a validation tool and is actually requested by regulatory agencies to complement ELISA result in confirming the assay's reliability and coverage, rather than serving as the primary method for routine HCP analysis. [211,214,216]

iii) Mass Spectrometry

Mass spectrometry (MS) has emerged as an indispensable and valuable method for the precise identification and quantification of individual host cell proteins HCPs in therapeutic antibody production. Compared to traditional immunoassays like ELISA, which are limited by antibody specificity and coverage, MS offers a non-immunoreactive alternative that allows for a thorough analysis of HCP profiles, independent of the anti-HCP antibodies variability.[204,218] MS enables the detection and characterization of specific HCPs, providing critical data on each impurity's identity, abundance, and

potential risks, all of which are invaluable for evaluating manufacturing consistency and product safety.[198,217]

One of the major challenges MS faces in this context is the extreme dynamic range of HCPs in biotherapeutic samples. Indeed, to reach parts-per-million (ppm) levels of HCPs, by definition, a dynamic range of six orders of magnitude are needed, which is not reachable on a routine LC-MS/MS coupling setup (Figure 16.)[218,219] To meet this challenge, MS has so far relied on the implementation of innovative and robust sample preparation techniques such as antibody depletion, molecular weight cutoff, multi-dimensional fractionation. These strategies enhance sensitivity by reducing ion suppression and matrix effects, thereby improving the detection of low-abundance HCPs.[220,221] The goal to detect sub-ppm levels of HCPs is attainable and has already been achieved[204,222] but straightforward and robust methods compatible with a biopharmaceutical environment are still awaited and will certainly greatly benefit from further analytical innovations at all steps of the workflow.

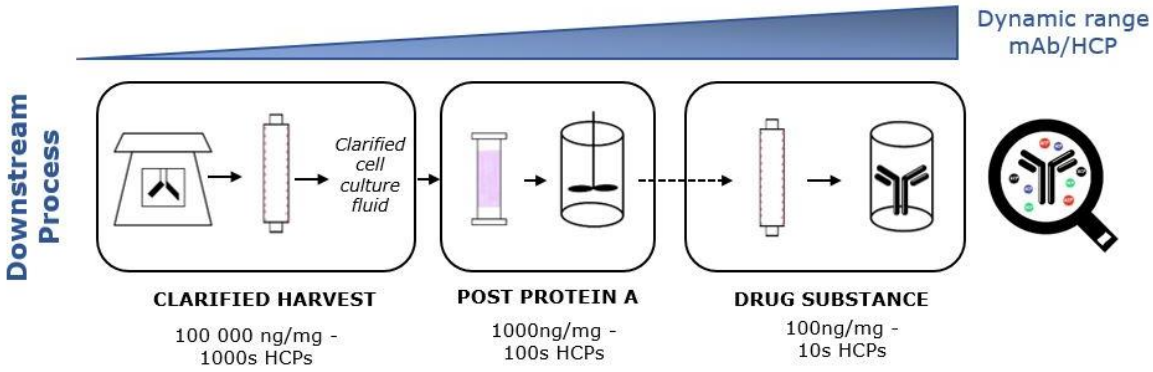


Figure 16: Host Cell Protein Clearance During the Purification Process. Representation of the reduction in both the total quantity (ng of HCP per mg of mAb) and the number of individual HCPs at various stages of the mAb production process. The image covers key steps such as Protein A affinity chromatography (Post Protein A), highlighting the extent of HCP removal.

What is more, advanced acquisition techniques further elevate MS's capabilities. Data-independent acquisition (DIA), for instance, has been shown to improve HCP detection across complex samples, as it captures data for all ions within a sample, thereby increasing both coverage and sensitivity .[218,221] Such methods actually address the limitations seen with data-dependent acquisition (DDA), which, while effective in high-quality spectra generation, can overlook low-abundance HCPs due to its selective ion sampling.[221] DIA's reliance on extensive spectral libraries also enhances the reproducibility of MS results, though this requires advanced data processing algorithms to manage the vast datasets and reduce potential false positives.[222]

In the context of risk assessment, MS's capacity to profile individual HCPs is invaluable for identifying contaminants that may affect product stability, immunogenicity, or even manufacturing efficiency. For example, knowing the identities and relative abundances of HCPs helps determine which impurities require stringent monitoring and removal, thus ensuring therapeutic safety.[217] Furthermore, MS-based

HCP analysis during process development supports bioprocess optimization by identifying problematic HCPs early and assessing the effectiveness of purification strategies at several stages of the downstream process cascade.[220–222]

Finally, continuous advances in MS instrumentation and proteomics methods promise to enhance HCP detection sensitivity and streamline workflow efficiency. Techniques such as native digestion protocols and innovative chromatographic separations address challenges related to matrix interference and assay robustness, enabling more reliable quantification of HCPs even in such complex samples.[218–220] These advancements position MS as a central component in impurity profiling and risk assessment, supporting both the rigorous quality standards required in therapeutic antibody production and the evolving regulatory expectations for HCP monitoring.[204]

PART II. Development of an Immuno-Capture Mass Spectrometry based method for ELISA assessment

Chapter 1
Development of Immuno-Capture Technique for Host Cell Protein identification

1. Context of the project: Challenges in Detecting HCPs with ELISA

In drug production, detecting host cell proteins (HCPs) is crucial for ensuring product quality and safety. Early in the drug development process, when the number of drug candidates is high and the risks of failure are very significant, companies commonly rely on commercially available ELISA kits. These kits typically contain broad-spectrum, polyclonal anti-HCP antibody mixture designed to detect a wide range of HCPs from various host cell lines. Such kits are useful for initial screenings but are limited in their ability to detect all HCPs, since their population may vary according to the conditions of the production process and some proteins.

As drug development progresses toward market release, the limitations of initial screening methods become more evident, underscoring the need for more specific, process-related anti-HCP antibodies. At this stage, companies often invest in immunization campaigns to generate antibodies that are tailored to the unique HCPs present in their production process. These process-specific antibodies can improve the sensitivity and coverage of ELISA, leading to better detection of target HCPs. However, ELISA still fundamentally relies on the affinity between the polyclonal anti-HCP antibodies and the HCPs they are designed to detect. While these antibodies are effective at capturing proteins that elicit a strong immune response, they may fail to recognize non-immunogenic or weakly immunogenic HCPs. This reliance on immune recognition means that proteins, which do not provoke a sufficient immune reaction, may remain undetected in the assay, leaving potential risks unaddressed.

The challenges of detecting these elusive HCPs are further illustrated in Figure 17, which depicts several possible outcomes of a Sandwich ELISA using process-specific anti-HCP antibodies from an immunization campaign. The first scenario represents the ideal outcome, where the primary antibody binds effectively to the HCP antigen, and the secondary antibody conjugate binds to form a detectable complex. However, as seen in the second scenario, HCPs that are not present in the null cell line used for immunization, yet appear in downstream processing (DSP) samples, will not be detected, as the antibodies used in the assay fail to bind these proteins. In the third case, the HCP did not create an immune response from the animal and therefore there is no antibody it can successfully bind to. The fourth scenario illustrates that even when the HCP is present in the null cell line, the secondary conjugated antibody may fail to bind the primary antibody–antigen complex, resulting in a lack of detection. Finally, in some cases, HCPs that have undergone structural modifications may not be recognized by the antibodies, further complicating detection. These outcomes demonstrate the persistent limitations of ELISA, even when employing a process-specific couple in order to enhance the sensitivity and accuracy of the results.

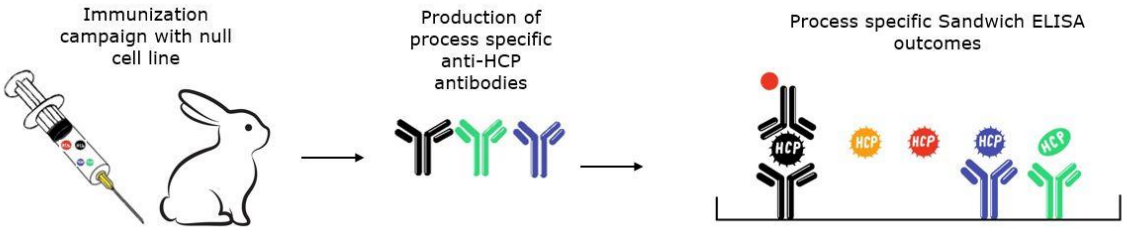


Figure 17: Illustration of possible outcomes from the Sandwich ELISA for a process-specific couple.

The inability of ELISA to detect all host cell proteins (HCPs), particularly those that are non-immunogenic presents a significant challenge in biopharmaceutical quality control. Given the regulatory and economic stakes, it is crucial to establish a reliable method for evaluating the coverage and limitations of generic or commercial ELISA assays for specific applications. To address this, the first key objective of this study is to develop an immune-affinity MS-based approach as a complementary tool to assess ELISA performance. By emulating the principle of ELISA, using the same anti-HCP antibodies to selectively capture immuno-reactive HCPs, this method not only evaluates ELISA coverage but also facilitates the recovery and identification of HCPs that are not recognized by the antibodies. This allows for a more comprehensive analysis of HCPs, enhancing the detection of low-abundance, non-immunogenic proteins that ELISA might miss.

Given that this project was carried out as a collaboration with Merck, they kindly lent me their premises at the Merck Guidonia site, where I was able to test and optimize the immune-affinity support system, which included evaluating three different Protein A supports: magnetic beads, plates, and columns. This thorough testing process, conducted over a nine-month period, ensured the reliability of the immune-capture procedure. Once optimized, I moved to Strasbourg to integrate the mass spectrometry MS component and further refined the immune-affinity process with the Phytips Protein A pipette tips which would allow for precise identification and quantification of HCPs that are missed or weakly detected by ELISA, addressing the key limitations of this traditional immunoassays. This ultimately provides a more comprehensive and reliable assessment of HCP contamination, contributing to improved safety and quality in biologic drug development.

2. Description of analytical methods used

In the exploratory phase of testing and optimizing the immune-affinity supports, we employed three main analytical methods: SDS-PAGE, Western blotting, and 2D DIGE. These techniques were specifically selected for their efficiency and speed in providing immediate results, which was essential given the preliminary nature of this stage. Since we aimed to evaluate the success of immune-capture quickly, these methods offered a practical alternative to mass spectrometry, which, although highly detailed, is time-intensive and better suited for later, confirmatory analysis.

Our primary goal was to first optimize and compare the performance of the three different supports: magnetic beads, plates, and columns. Testing involved a wide range of conditions to thoroughly assess each support's effectiveness in capturing HCPs. The use of SDS-PAGE, Western blotting, and 2D DIGE

allowed us to rapidly gauge the binding efficiency of the anti-HCP antibodies in capturing HCPs on each support. These methods provided clear, prompt feedback on the affinity and specificity of the immune-affinity interactions, enabling us to adjust our conditions swiftly.

Previous data from Merck's research had shown good affinity between polyclonal anti-HCP antibodies and HCP antigens in Western blotting, yet the antibody behavior in the new immune-affinity format required evaluation, given that they usually present different affinity when tested in different methods.[223] Immuno-affinity capture involves unique dynamics that differ from those in Western blotting; thus, by testing various conditions in a systematic manner, we were looking to establish optimal parameters for each support type, supports that had never been tested before for this specific application.

Ultimately, these analytical methods allowed us to conduct an efficient, iterative refinement of conditions for each support, providing a solid foundation for subsequent mass spectrometry analysis. Following several discussions, four parameters were identified as critical for optimization: the antibody-to-antigen ratio, incubation temperature, incubation time, and the total volume of the reaction mixture. These variables were selected to structure the experimental design and guide systematic evaluation. A summary table down below was included to present the tested conditions (Table 4). Once the most promising settings were established, we proceeded to mass spectrometry to finalize the identification and quantification of HCPs in the optimized samples. This phased approach ensured both rigorous evaluation and efficient progression through the exploratory stage of our immune-affinity support optimization.

Parameters	Condition Tested
antibody : antigen ratio	1:1 , 1:2 , 2:1
Incubation Temperature	4°C , 25°C, 37°C
Incubation Time	2h, 24h
Incubation Volume	300µL, 2mL

Table 4 : Overview of the four critical parameters selected for optimization during the immune-capture method development. The table summarizes the conditions tested for each parameter throughout the optimization process.

A. SDS-PAGE

SDS-PAGE analysis was performed on the sample obtained from each optimization using 4–20% Criterion™ TGX™ Precast Midi Protein Gels (Bio-Rad, 18-well comb, 30 µL, 1 mm). The gradient gels were selected for their ability to separate a broad range of molecular weights, facilitating optimal visualization of proteins across various size classes in the samples. Samples were loaded in 2 gels in parallel, one which would be dedicated to the silver staining and the second one which would be

dedicated to the Western Blot (Figure 18). The power supply was set to two stages: 180 Volts for 10 minutes, followed by 220 Volts for 45 minutes to ensure consistent protein separation. The running buffer used was Tris-Glycine-SDS (TGS, 10X from Bio-Rad), diluted to 1X with MilliQ water.

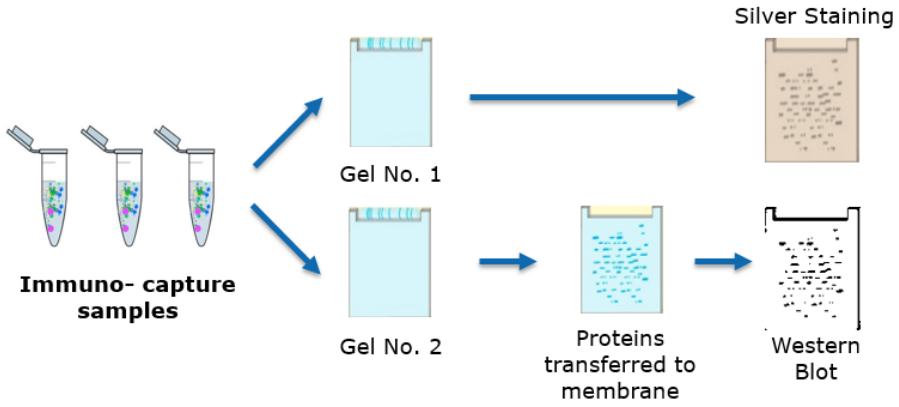


Figure 18: Workflow scheme for SDS-PAGE analysis, where immuno-capture samples were loaded onto two separate gels, one for silver staining and the other for western blotting.

After electrophoresis, the gels were stained with the ProteoSilver™ Silver Stain Kit (Sigma-Aldrich), chosen for its high sensitivity in detecting low-abundance proteins. This staining method provided a clear visualization of the protein profile in each fraction. Gel images were captured using the ChemiDoc imaging system (Bio-Rad), with imaging settings adjusted to facilitate subsequent analysis.

B. Western Blot

To further verify the immunoreactivity of the HCP population in the captured samples, Western blotting was employed using the Bio-Rad Trans-Blot Turbo System, following the Turbo protocol for fast protein transfer (7 minutes, 2.5 Amperes, and 25 Volts). Polyvinylidene fluoride (PVDF) membranes were used for protein transfer, which were initially blocked with the EveryBlot Blocking Buffer (Bio-Rad) to ensure maximum sensitivity. The blocking step lasted for 5 minutes, followed by primary antibody incubation using in-house anti-HCP antibody, the same one employed for the immune-capture assays, carried out under gentle stirring for 1.5 hours at room temperature.

After the primary incubation, the membranes were washed three times with 100 mL of 1X Phosphate-buffered saline (PBS) containing 0.1% Tween 20, under stirring conditions for 10 minutes each. Secondary antibody incubation was performed using chicken anti-rabbit horseradish peroxidase (HRP) from Invitrogen (diluted 1:5000) for 1 hour at room temperature, followed by additional washing steps.

For chemiluminescent detection of the antibody binding, Clarity Max Western ECL Blotting Substrates (Bio-Rad) were used. The membranes were imaged using the ChemiDoc system by Bio-Rad, with optimized acquisition settings for high-sensitivity detection of immune-reactive proteins.

C. 2D-DIGE

To gain a more comprehensive understanding of the HCP population present in the samples obtained, we transitioned from SDS-PAGE and Western blotting to 2D difference gel electrophoresis (DIGE). This

approach allowed for the simultaneous comparison of multiple conditions on a single gel, significantly improving the reproducibility of our results. The main advantage of DIGE is its ability to distinguish low-abundance proteins, offering enhanced sensitivity and precision compared to other methods. By using spectrally resolvable fluorescent dyes, Cy™2, Cy™3, and Cy™5, we were able to track proteins from different samples, even when they co-migrated in the gel, reducing technical variation and enabling more accurate comparative analysis.[224,225] The multiplexing capability of DIGE made it an ideal choice for analyzing complex biological samples and provided a deeper and more reliable insight into the HCP population captured by the immunoaffinity supports.

For each sample, 15 µg of protein underwent a clean-up step using the 2D Clean-up kit by Cytiva. The samples were then labeled with the fluorescent dyes, with the labeling reaction performed in the dark at 5°C for 10 minutes. The reaction was quenched using L-Lysine monohydrochloride, followed by a 15-minute incubation at room temperature. After labeling, the samples were pooled and their volume was adjusted to 200 µL with rehydration buffer composed of 7 M Urea, 0.032 M CHAPS, 2 M Thiourea, and 1% Dithiothreitol (DTT).

The samples were then subjected to bidimensional electrophoresis, starting with isoelectric focusing (IEF) using 11 cm IPG ReadyStrips (pH 3–10) in the PROTEAN® i12™ IEF system by Bio-Rad, following the pH gradient protocol. After focusing, the strips were reduced and alkylated before being transferred to 4–20% Criterion™ TGX™ Precast Midi Protein Gels for separation based on molecular weight. Finally, the gels were scanned using the GE Typhoon FLA 9500 detector, which captured the fluorescent signals at specific wavelengths for each dye: 473 nm for Cy™2, 532 nm for Cy™3, and 635 nm for Cy™5.

3. Experimental workflow and selection of Immuno-Capture supports

In terms of the immuno-capture experimental workflow, the process began with the incubation of anti-HCP antibodies alongside a null cell line, testing various conditions to optimize the capture process. Following this, the antibodies were immobilized onto the chosen support using Protein A or Protein A/G, as illustrated in Figure 19 below. This setup was designed to selectively capture immune-reactive HCPs, while non-immuno-reactive HCPs, which do not form complexes with the antibody, would pass through the support. These non-reactive HCPs were collected as the "unbound" fraction for further analysis.

On the other hand, the immune-reactive HCPs bound to the antibody on the support were eluted under controlled conditions, detaching from the surface to form the "elution" fraction. This two-step process—capturing the immune-reactive HCPs and isolating the non-immuno-reactive ones—allowed us to clearly distinguish between the two populations. Each fraction was then analyzed using the analytical methods described, providing a more refined understanding of the HCPs present in the samples.

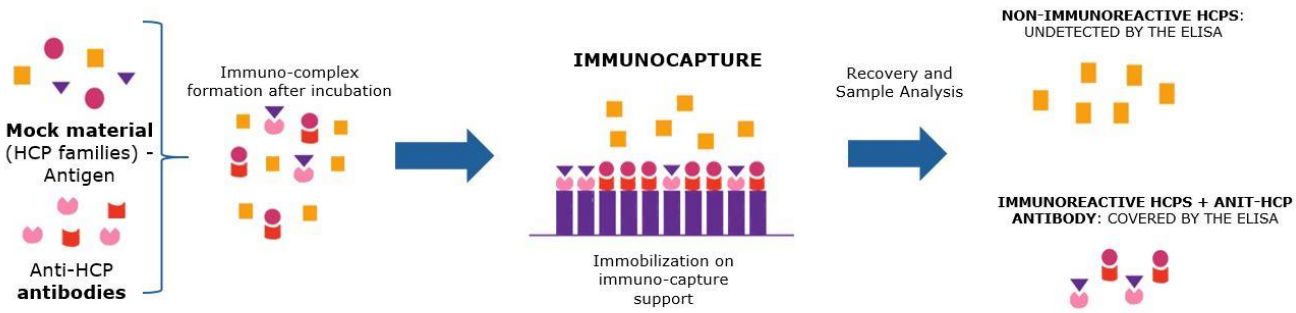


Figure 19: Illustrated scheme of an immuno-capture experiment. The input sample, after incubation, is loaded onto the support, resulting in two distinct fractions: the unbound fraction and the elution fraction.

To confidently identify HCPs with low or no immuno-reactivity, we compared the results of the immobilized ELISA reagent assays against several negative control conditions. These controls, outlined in Table 5, were designed to ensure the validity of our findings and to exclude any potential interference or contamination from the assay reagents. Importantly, the samples collected in these negative controls followed the same experimental workflow as the immuno-capture assay and were performed simultaneously.

The first negative control utilized empty support, meaning no antigen or anti-HCP antibody complex was present. This setup was essential to determine whether any assay reagents could contribute to non-specific binding or contamination. The second negative control involved the antigen alone, allowing us to investigate potential interactions between the antigen and the Protein A or Protein A/G on the support. This ensured that the antigen did not directly bind to the support in the absence of the anti-HCP antibody. Lastly, the third negative control was the antibody-only condition. In this scenario, the anti-HCP antibody was immobilized without the presence of the antigen. This control verified the proper capture and release of the antibody during the elution step, ensuring that the antibody behaved as expected throughout the workflow.

By incorporating these negative controls (Table 5), we ensured a robust assessment of the immuno-capture system and the reliability of the assay in distinguishing true immuno-reactive HCPs from those with low or no interaction with the anti-HCP antibody.

Assay performed	Antibody	Antigen	Beads
Input after incubation	X	X	
Immuno-capture	X	X	X
Empty Support			X
–ctrl Antigen		X	
–ctrol Antibody	X		

Table 5: Table containing the list of samples and negative controls performed simultaneously per immuno-capture test experiment.

Positive controls were carried out for each support following their respective optimizations. These controls served to confirm that the immuno-capture process performed consistently for both the process-specific HCP-antigen/anti-HCP antibody pair and the commercial generic HCP-antigen/anti-HCP antibody pair from Cygnus. By including these controls after optimization, we aimed to verify that the behavior of the support remained consistent for both systems and that no unintended variations occurred. Furthermore, the positive control validated the stability and suitability of the optimized method when applied to the commercial reagent.

The analytical workflow, depicted in Figure 20, began with the performance of the immuno-capture assay alongside its corresponding negative controls, ensuring the consistent recovery of input, unbound, and elution samples. These samples were first evaluated using SDS-PAGE coupled with Western blotting, allowing us to confirm the assay's success and examine the behavior of the controls. Following this preliminary assessment, the samples underwent a comprehensive 2D DIGE analysis, which provided a more detailed characterization of the HCP populations captured by the anti-HCP antibody on the chosen support.

Concerning, the choice of Protein A and Protein A/G as supports, this is driven by their high affinity for Rabbit IgG, the host species used to produce the anti-HCP antibodies. Protein A, in particular, effectively binds the Fc region of IgG, ensuring the proper orientation of the antibody for optimal antigen interaction.[226] This approach offers a reliable alternative to harsher methods such as biotinylation and streptavidin binding. Biotinylation, while commonly used, poses a risk of binding to unintended sites on the IgG molecule, including the antigen-binding epitope or close to that critical area, potentially interfering with the crucial interaction between the antibody and its target antigen.[227] By employing Protein A and Protein A/G supports, these limitations are mitigated, ensuring the integrity and functionality of the anti-HCP antibody during the assay.

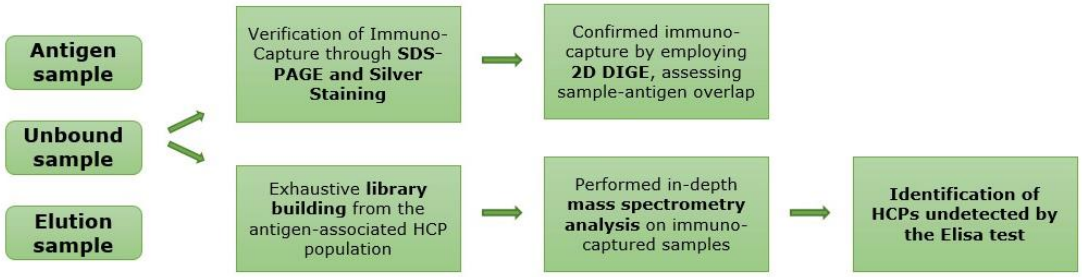


Figure 20: Analytical workflow for the analysis of immuno-capture samples. On one hand, a rapid assessment of SDS-PAGE, Western Blot, and 2D DIGE is performed, while on the other, MS analysis is conducted after the construction of an exhaustive spectral library.

A. Magnetic Protein A/G Beads:

The Pierce™ Classic Magnetic IP/Co-IP Kit (Thermo Scientific) was selected for immunoprecipitation (IP) of the HCPs due to its user-friendly protocol and the option for automation, which provided flexibility in our experimental workflow for future routine applications. The workflow followed the standard protocol provided by the supplier, but several parameters were optimized for our specific application. We adjusted reagent quantities, incubation times, and temperatures to fine-tune the immuno-complex formation, ensuring optimal capture of HCPs by the anti-HCP antibody from our samples.

The process began by incubating the anti-HCP antibody with the HCP antigen sample for 1-2 hours at room temperature to overnight at 4 degrees to allow for efficient antibody-antigen binding. The protein A/G magnetic beads were then added to the mixture, and the complex was incubated for 1 hour at room temperature. The unbound sample was recovered, with all the non-immunoreactive HCPs present and then, after washing to remove non-specific interactions, the bound immuno-complex was eluted using a denaturing Lane Marker Sample Buffer for subsequent analysis. The samples were loaded onto the gels for analysis via SDS-PAGE, Western blotting, and 2D DIGE. The flexibility of this immunoprecipitation technique, allowed us to test several key parameters in order to better evaluate the support.

B. Protein A/G coated plate

The Protein A/G-coated plate immunoprecipitation support by Cayman was selected due to its structural and operational similarity to ELISA, offering rapid execution and straightforward assessment.

According to the supplier's protocol, the process begins by diluting antibodies in the provided IP Buffer to a volume of 5–100 µg/µL per well, ensuring a total of 100 µL per well. Antigens are similarly prepared, diluted 1- to 100-fold in the same buffer. The wells are washed three times with 200 µL of Wash Buffer before adding 100 µL of the antibody solution. The plate is then covered, and the antibody is incubated. Following incubation, five additional washes are performed to remove any unbound material. Subsequently, 100 µL of the antigen solution is added to the wells, and the plate is incubated

again, with the supplier recommending either 4°C overnight or 1 hour at room temperature. After antigen incubation, the wells undergo another series of five washes. Elution is performed by adding 22 µL of 6X SDS Buffer and 12 µL of Neutralization Buffer into a new plate, followed by 100 µL of Elution Buffer to the original IP plate wells. After a brief incubation of 15–30 seconds, the eluted mixture is transferred to the new plate and heated at 100°C for 5 minutes to reduce the samples for downstream analysis, such as SDS-PAGE.

The plate system presents a unique aspect compared to other supports, as it introduces an additional sample for analysis: the unbound antibody. This sample, collected after the antibody incubation step, allows us to confirm whether the anti-HCP antibody successfully attaches to the Protein A/G support. As outlined in the supplier's workflow, see Figure 21, this extra step provides critical insight into the effectiveness of the antibody-support interaction.

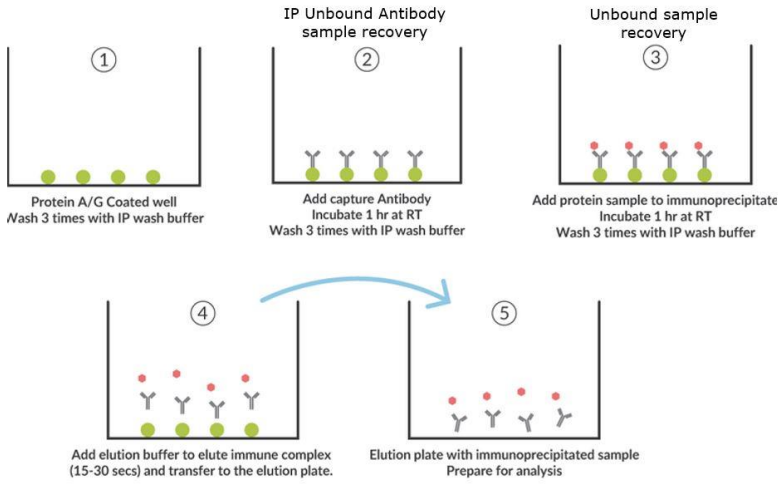


Figure 21: Illustrated Protein A/G coated plate workflow taken from the supplier's booklet. The image also depicts the two instances where critical samples are taken, IP Unbound Antibody and Unbound sample.

C. Protein A Column

For the immuno-capture experiments using the Protein A column, we chose to use the HiTrap™ Protein A HP columns prepacked with MabSelect PrismA protein A chromatography resin. This affinity resin, typically used for monoclonal antibody mAb purification in a clean-up context, is designed to have a high capacity due to its usual application in mAb processing. Given this large capacity, we decided to optimize the column's use for our specific needs, testing different reagent ratios to match the IgG-binding capacity of the column. The immuno-capture process was carried out using an ÄKTA pure™ chromatography system equipped with a HiTrap Protein A HP column of 1 mL capacity, both from Cytiva. The buffers used for this process included 20 mM sodium phosphate at pH 7 for the binding step and 0.1 M citric acid at pH 6 for elution. The flow rate was set at 1 mL/min, with a 10-minute column conditioning, 10 minutes for the binding step, and 5 minutes for elution. Samples were

collected manually based on the UV absorbance response, and concentrations were determined using the NanoDrop™ One/OneC by Thermo.

We tested three antibody-to-antigen ratios to determine the most optimal conditions for immuno-capture on the Protein A column. In addition to the standard overnight incubation at 4°C, we investigated the effect of varying incubation conditions to optimize the binding efficiency of the anti-HCP antibody-antigen interactions. Specifically, we explored two additional temperature settings, 37°C and room temperature, for shorter incubation periods of 2 hours.

After the immuno-complex incubation, the sample volume was adjusted to 2 mL, if needed, for injection into the chromatography system. The immuno-capture procedure allowed for the collection of the three distinct sample fractions: the input; the unbound fraction, where the non-immunoreactive HCPs, that did not form a complex with the antibody and thus did not bind to the column, are found; and the elution sample, where the immunoreactive HCPs coupled with the anti-HCP antibody are collected, eluted by low pH, which was subsequently neutralized with Tris-HCl pH 9.

These different incubation conditions were aimed at optimizing the immuno-capture process by maximizing the efficiency of the anti-HCP antibody binding while minimizing non-specific interactions. The combination of temperature, incubation time, and volume provided a comprehensive assessment of the Protein A column's ability to capture and isolate HCPs under different experimental conditions.[228]

4. Results of Immuno-capture method testing and optimization

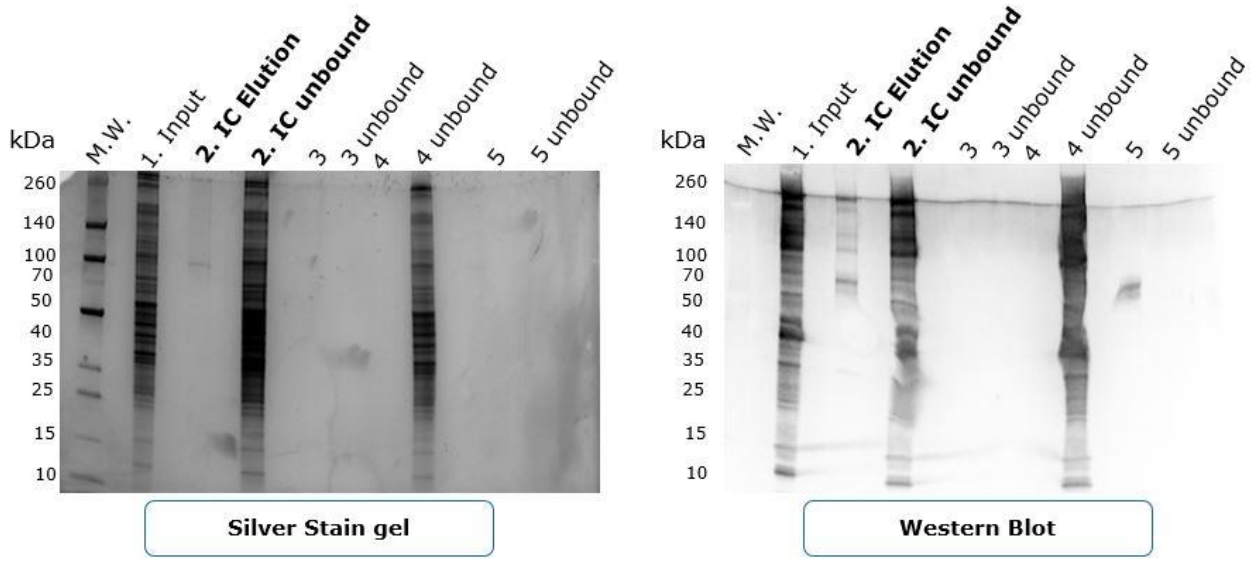
A. Magnetic Protein A/G beads

For the initial experiments with the Protein A/G magnetic beads, we followed the supplier's recommended parameters to familiarize ourselves with the support and establish a baseline for the immuno-capture procedure. We opted for a balanced approach, using the middle of the supplier’s suggested range for both reagent quantity and bead volume. Specifically, we used 25 µL of magnetic beads, coupled with 6 µg of anti-HCP antibody and 750 µg of antigen, which were incubated overnight at 4°C to enhance the antibody-antigen binding. Following this, the immuno-complex and bead incubation was carried out for 1 hour at room temperature. Negative controls, including antibody-only, antigen-only, and empty beads, were also included in the experimental setup to ensure the specificity of the binding and to check for any potential non-specific interactions. The samples were then recovered and loaded onto gels for subsequent SDS-PAGE and Western Blot analysis to evaluate the success of the immuno-capture procedure and the behavior of the negative controls.

The success of the immuno-capture procedure was evaluated by analyzing the gel bands present in the IC unbound and elution lanes. The IC unbound lane should reflect the non-immunoreactive HCPs that did not bind to the anti-HCP antibody and, therefore, passed through the support. A more populated IC unbound lane indicates an unsuccessful immuno-capture, as it suggests that many HCPs were not captured by the antibody. Conversely, an empty IC elution lane, which represents the immuno-complex that detached from the support during elution, suggests that the anti-HCP antibody

did not successfully bind to the immuno-reactive HCPs. In essence, a well-executed immuno-capture would result in minimal bands in the IC unbound lane and a strong signal in the IC elution lane, indicating efficient antibody binding and capture of the immunoreactive HCPs.

As observed in Figure 22, the first attempt at immuno-capture using this support was unfortunately highly unsuccessful, as evidenced by the crowded and populated IC unbound lane. This suggests that a significant amount of HCPs did not bind to the anti-HCP antibody and passed through the support. However, the negative control samples provided useful insights and helped confirm the performance of the reagents. The empty bead samples were particularly reassuring, as both the elution and unbound lanes were clear, indicating that there was no apparent contamination or non-specific binding of reagents.



Assay performed	Antibody	Antigen	Beads
1. Input sample	X	X	
2. Immuno-capture	X	X	X
3. Empty Support			X
4. –ctrl Antigen		X	
5. –ctrol Antibody	X		

Figure 22: Initial outcomes from the magnetic beads familiarization, including the Silver stain gel and corresponding Western Blot. The experiment utilized 25 µL of magnetic beads, 6 µg of anti-HCP antibody, and 750 µg of antigen. The antibody-antigen complex was incubated overnight at 4°C, followed by a 1-hour incubation at room temperature with the beads. The figure also includes a table detailing the samples and negative controls observed on the gels and membrane.

For the negative control with the antigen, it was also promising to see that the elution lane was clear, while the unbound lane contained the complete HCP population, confirming that the antigen does not

interact with the Protein A/G support in the absence of the anti-HCP antibody. Lastly, the negative control of the antibody further confirmed the correct functionality of the system. As seen in the image, the antibody was successfully detected in the elution sample, validating that it was properly released from the support during the elution step. These negative control results provided confidence in the overall integrity of the immuno-capture process, despite the issues observed with the initial antibody-antigen binding.

Given the initial unsuccessful immuno-capture experiment, it was first suspected that the main issue with the technique lay in the immuno-complex incubation step and the reagent ratios used. In light of this, we decided to adjust these parameters in a new experiment. For this iteration, we increased the bead volume to 50 µL, while the anti-HCP antibody was raised to 20 µg, and the HCP antigen quantity was set at 250 µg. The antibody-antigen incubation was maintained at 4°C overnight to enhance binding efficiency, while the immuno-complex and bead incubation continued for 1 hour at room temperature. These adjustments were made with the goal of improving the binding efficiency and ensuring better capture of the immunoreactive HCPs.

However, as shown in Figure 23, the lanes corresponding to the IC elution appeared almost completely absent of signal, further corroborating the failure of the experiment. Despite modifications to reagent ratios, the outcome indicated that the immuno-capture process was ineffective, with minimal binding of the anti-HCP antibody to the immuno-reactive HCPs. In this particular case, based on prior results where the IC unbound lane exhibited a high abundance of HCPs, the two subsequent washes of the beads were collected to assess whether any HCP population persisted after washing. As observed in the figure, these wash lanes also appeared empty, indicating that no residual HCPs were retained on the beads. This result suggests that further optimization of the immuno-capture parameters was still required to achieve the desired outcome.

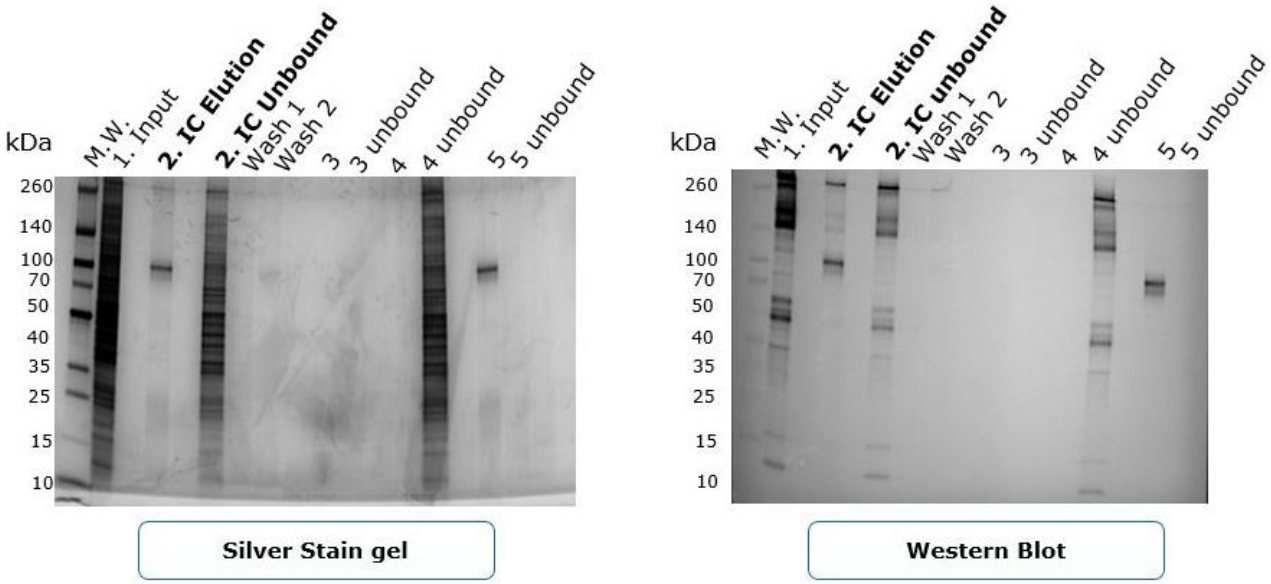


Figure 23: Images showing the first optimization of the magnetic beads, featuring both the silver stain and the corresponding Western blot. For this experiment, the bead volume was increased to 50

µL, with 20 µg of anti-HCP antibody and 250 µg of HCP antigen. The antibody-antigen incubation was carried out overnight at 4°C to enhance binding, followed by a 1-hour immuno-complex and bead incubation at room temperature.

Another attempt was carried out subsequently, where the reagent ratios were kept consistent with the previous trial, but the incubation times were extended. Both the immuno-complex incubation and the bead-incubation steps were prolonged from overnight to 24 hours, under the assumption that this extended incubation would help enhance the binding between the anti-HCP antibody and the immunoreactive HCPs and improve their attachment to the magnetic beads.

This change was made with the goal of allowing more time for the antigen and antibody to interact, potentially improving the efficiency of the immuno-capture process. Nonetheless, as illustrated in Figure 24, the IC elution lane appeared visibly empty, which indicates an unsuccessful outcome. What sets this result apart from previous experiments is the observation that the IC unbound lane also appeared empty. This can be attributed to the substantial dilution of the samples (1:10), which was implemented to prevent saturation of the gel due to the high concentration of HCPs in the unbound fraction. In prior experiments, the abundant HCPs in the IC unbound fraction led to signal saturation, making it challenging to detect the population present in the IC elution. Despite the dilution, the absence of any significant signal in the IC elution lane confirms that the immuno-capture process did not succeed in this instance.

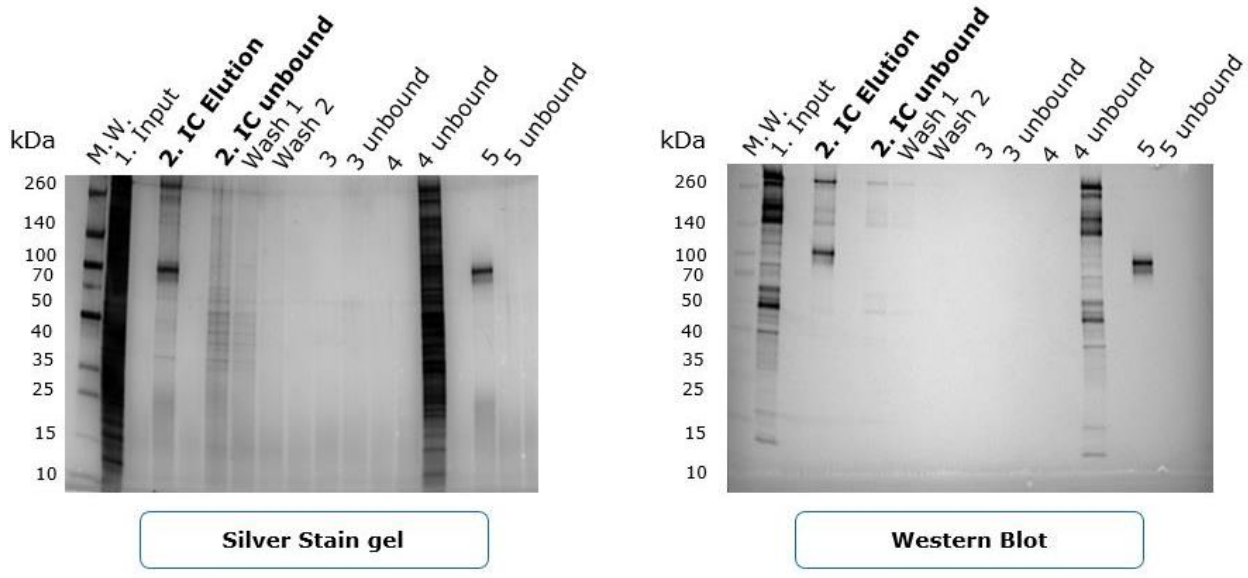


Figure 24: Images depicting the second optimization of the magnetic beads, including both silver stain and corresponding Western blot. To enhance binding efficiency, the incubation times were extended, with both the immuno-complex and bead incubations carried out for 24 hours, as opposed to the previous overnight protocol.

Still under the suspicion that the formation of the immuno-complex and its subsequent attachment to the Protein A/G on the beads was the key issue, another attempt was made by testing the temperature of both of the incubations. Rather than carrying out 24-hour incubations at 4°C, it was decided to try

incubating at room temperature for 24 hours to assess whether this might enhance the binding between the anti-HCP antibody and the immunoreactive HCPs. The rationale behind this change was that room temperature might facilitate stronger interactions and improve the efficiency of the immuno-capture process.

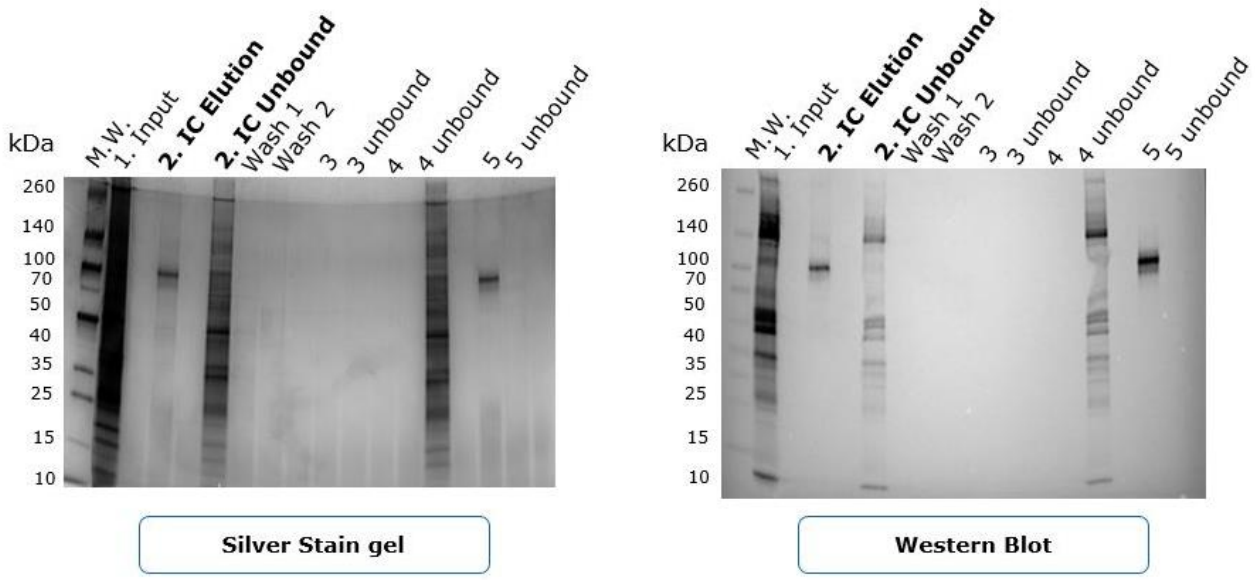


Figure 25: Images showing the third optimization of the magnetic beads, featuring both the silver stain and the corresponding Western blot. The immuno-complex was incubated for 24 hours at room temperature instead of 4°C, with the aim of assessing whether this adjustment would improve the binding efficiency between the anti-HCP antibody and the immuno-reactive HCPs.

Observing no significant improvement in the immuno-capture process despite the previous adjustments, as observed in Figure 25, we turned our attention to a deeper investigation of the bead binding capacity. Through thorough calculations, it was determined that the 50 µL of beads used in our experiments possessed a maximum binding capacity of 40 µg of anti-HCP antibody. Based on these findings, the experimental design was adjusted to use the calculated optimal anti-HCP antibody quantity, while maintaining the HCP antigen concentration at 250 µg. Both the immuno-complex incubation and bead-incubation steps were set to 24 hours at 4°C, with the aim of isolating the variable of binding capacity in this assessment.

As shown in Figure 26, the assay still failed to produce the desired outcome. The IC unbound lanes remain highly populated with the antigen, indicating that the binding efficiency between the anti-HCP antibody and the HCP antigen, as well as the subsequent attachment of the immuno-complex to the beads, continues to be inadequate.

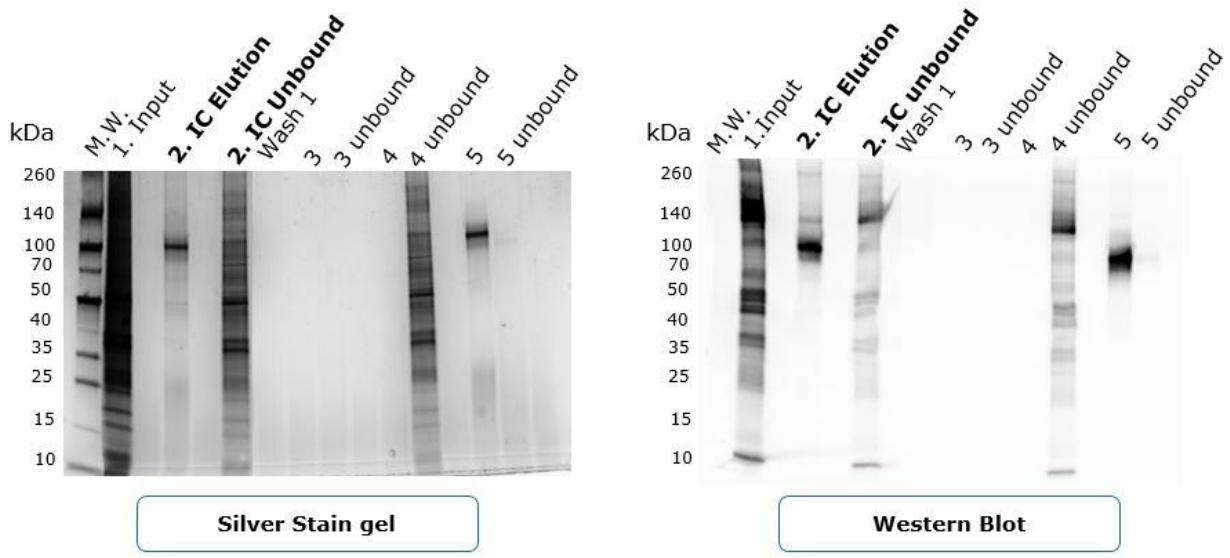


Figure 26: Images showing the fourth optimization of the magnetic beads, featuring both the silver stain and the corresponding Western blot. The anti-HCP antibody quantity was adjusted to 40 µg, based on the maximum binding capacity of 50 µL of beads. The HCP antigen concentration was maintained at 250 µg. Both the immuno-complex incubation and bead incubation were performed for 24 hours at 4°C.

To investigate the underlying reasons behind the beads' lack of success in achieving the desired results, we opted to explore an alternative approach by employing the Pierce Protein A IP Kit. This decision was driven by the possibility that using Protein A, which exhibits a particularly high affinity for rabbit IgG, might prove more effective than Protein A/G in facilitating the immuno-capture process. By leveraging the specific binding properties of Protein A, we aimed to assess whether this adjustment would enhance the attachment of the anti-HCP antibody to the support and improve the overall assay performance.

In this experiment, no negative controls were performed for the Protein A IP Kit, as this was intended to be a quick, exploratory test to evaluate the performance of the Protein A beads. The goal was to quickly assess whether the Protein A beads could improve the binding of the immuno-complex onto the support. As such, only the immuno-capture experiment was conducted to verify the outcome, without the inclusion of negative controls, which would not provide additional insights in this context.

Despite these adjustments, as illustrated in Figure 27, the immuno-capture process failed to meet expectations, with outcomes mirroring those of previous attempts. The IC Unbound lane remained heavily populated, while the IC Elution lane was devoid of any detectable signals. Based on these recurring challenges, it was concluded that the magnetic bead support is unsuitable for the scope of this project. This outcome highlights the complexity of working with the anti-HCP antibody mix, which involves polyclonal antibodies. Unlike monoclonal antibodies, polyclonal antibodies present significantly more complex affinity kinetics, making it exceedingly challenging to identify the parameters necessary for optimal binding.

After multiple rounds of optimization that failed to yield the desired outcome, we proceeded with a positive control experiment to further investigate the observed challenges. The aim was to verify whether the lack of success was specific to the process-specific HCP-antigen/anti-HCP antibody pair used during optimization or whether it extended to the commercial generic HCP-antigen/anti-HCP antibody pair from Cygnus.

In this positive control experiment, no negative controls were included, as negative controls are typically unnecessary when performing a positive control experiment. The purpose of a positive control is to confirm that the experimental conditions are functioning as expected, and as such, it is usually conducted without the inclusion of negative controls. The IC Unbound and IC Elution lanes were loaded twice on the gel to ensure clear visualization and comparison of the results.

As shown in Figure 27, on the right side of the image, the results confirmed that the immuno-capture outcome remained consistent regardless of the antibody-antigen pair used. Both the IC Unbound lane and IC Elution lane exhibited the same pattern as seen in previous experiments. This further supports the conclusion that the Protein A/G beads were unsuitable for this application. Additionally, the use of a polyclonal anti-HCP antibody introduces complexities in binding to both the antigen and the support, which may also contribute to the challenges observed.

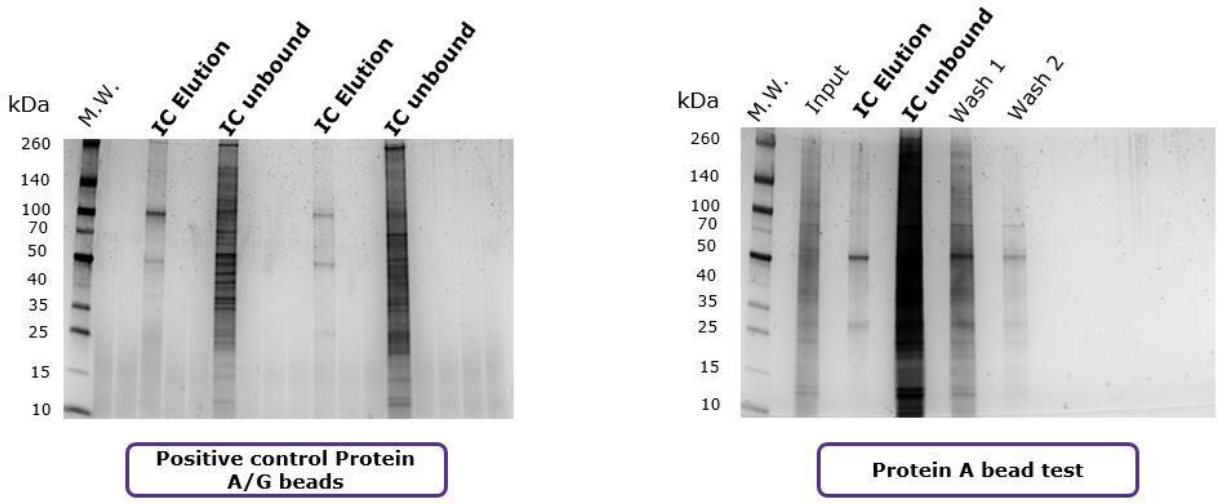


Figure 27: On the left, testing of Protein A/G magnetic beads with a generic commercial anti-HCP antibody mix and antigen, and on the right, testing of Protein A magnetic beads.

The difficulties stem from the nature of polyclonal antibodies, which are heterogeneous mixtures with varying affinities for their target antigen. Unlike monoclonal antibodies, whose binding affinity can be defined by a single equilibrium constant, polyclonal antibodies form multiple bonds with their antigens, making it impossible to assign a single affinity constant. This heterogeneity affects critical performance parameters, as the time required to reach binding equilibrium depends on the rate of diffusion, affinity, and environmental factors such as temperature, pH, and solvent. Given the intricate interplay of these factors, optimizing the immuno-capture process with polyclonal antibodies presents inherent challenges.[229]

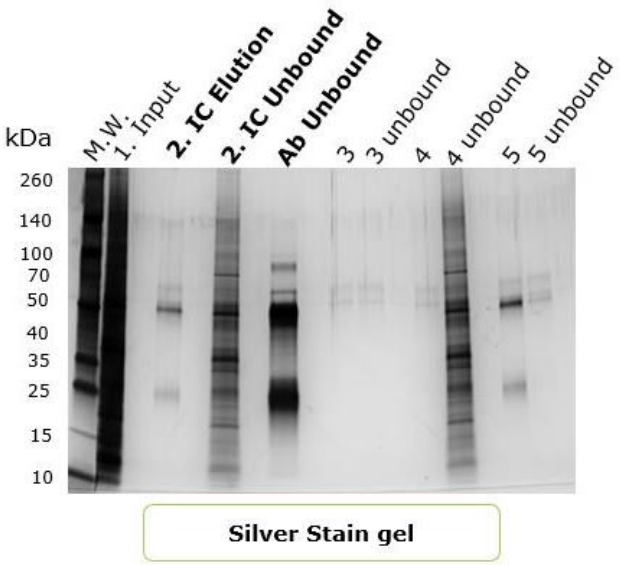
B. Protein A/G coated plates

Following the challenges encountered with the bead support and its limited suitability for our application, we decided to explore a plate support. This method offered rapid results, an easy-to-follow protocol, and a workflow closely resembling that of an ELISA assay. According to the supplier's protocol, key parameters, such as the concentrations of the antibody-antigen mixture and the incubation conditions (temperature and time), could be optimized to suit our needs. After several attempts that did not yield the desired results and considering time constraints, it was decided that only SDS-PAGE would serve as the initial analytical method of choice to assess the outcomes of the various supports. Should a successful immuno-capture be observed, we would then proceed to more detailed analysis to further confirm the results.

For an initial exploration of the plate support, we adhered closely to the supplier's guidelines, selecting parameters that positioned us midway within the recommended range. Specifically, the anti-HCP antibody solution was prepared at a concentration of 50 µg/mL and incubated for 1 hour at room temperature. Similarly, the antigen solution was diluted to a concentration of 30 µg/mL, following identical incubation conditions. After completing the incubations, the samples were carefully recovered and subjected to gel electrophoresis, followed by silver staining for analysis. To ensure the reliability of the outcomes, the same set of negative controls employed during the bead experiments was applied here as well, providing essential benchmarks for assessing the performance of the plate support.

Similar to the magnetic beads, the initial experiments with the plate support did not yield the desired outcomes. The IC Elution lane was nearly empty, while the IC Unbound lane appeared highly populated, indicating that the immuno-capture process was largely ineffective. This support also allowed us to analyze an additional sample: the unbound antibody fraction. After the initial incubation of the anti-HCP antibody, the unbound portion was also recovered and loaded onto the gel to evaluate the attachment efficiency. Unfortunately, as shown in Figure 28, this lane was heavily saturated, confirming poor binding of the anti-HCP antibody to the Protein A/G coating on the plate.

On a more positive note, the negative control samples produced reassuring results. The empty well control lanes were clean, showing no contamination or non-specific interactions. The antigen negative control confirmed that the HCPs do not bind to the Protein A/G surface in the absence of the anti-HCP antibody. Similarly, the antibody negative control demonstrated that the washes did not dislodge the antibody from the support, validating the integrity of the protocol in these aspects.



Assay performed	Antibody	Antigen	Plate
1. Input sample	X	X	
2. Immuno-capture	X	X	X
3. Empty Support			X
4. –ctrl Antigen		X	
5. –ctrol Antibody	X		

Figure 28: Silver stain of the initial outcomes from the plate support familiarization. The figure also includes a table listing the samples and negative controls with corresponding numbers. For this exploration, the anti-HCP antibody was used at 50 µg/mL and incubated for 1 hour at room temperature, while the antigen was prepared at 30 µg/mL with identical incubation conditions.

Given the initial outcome, it was decided to adjust the incubation temperature to 37°C for 1 hour, hypothesizing that this might enhance the affinity binding between the anti-HCP antibody and the antigen. Additionally, we explored different antigen solution concentrations to assess whether steric hindrance or antigen saturation could be interfering with effective binding. Specifically, concentrations of 30 µg/µL and 15 µg/µL were tested under these revised conditions.

However, as depicted in Figure 29, neither optimization strategy yielded meaningful progress. The IC Elution lane remained underwhelming, showing no significant improvement in the immuno-capture efficiency. This suggested that the changes to temperature and antigen concentration did not adequately address the underlying challenges.

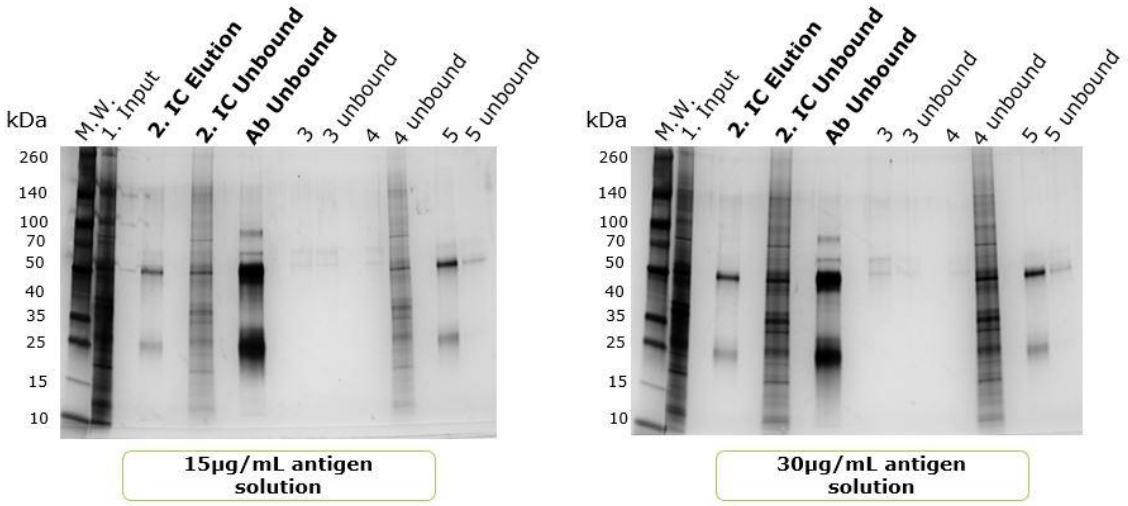


Figure 29: Silver stain gels showing the second optimization of the plate support, testing different antigen concentrations (30 µg/µL and 15 µg/µL). Both incubations were performed at 37°C for 1 hour to evaluate the effect of temperature and concentration on binding efficiency.

After observing no significant improvement in the previous optimization, another attempt was made by altering the incubation conditions. The incubations were carried out at 4°C for 24 hours, testing two different antigen concentrations: 30 µg/mL and 50 µg/mL, keeping the same anti-HCP antibody solution concentration at 50µg/mL. Despite these adjustments, the results remained consistent across both conditions. The IC elution lane showed minimal signal in Figure 30, suggesting that the changes in incubation temperature and duration did not enhance the immuno-capture efficiency.

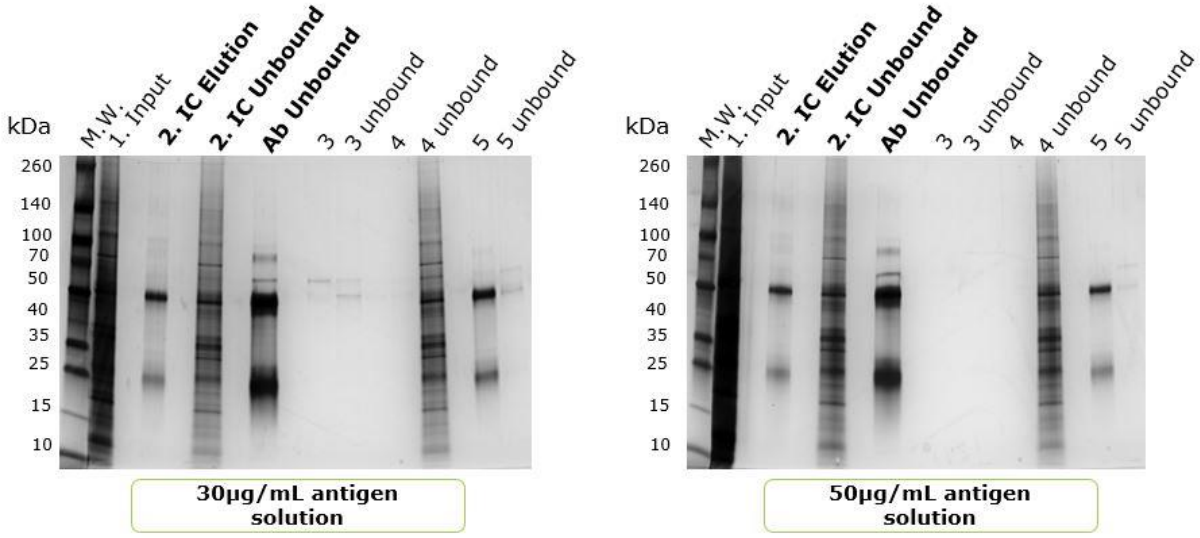


Figure 30: Silver stain gels showing third optimization, testing two antigen concentrations (30 µg/mL and 50 µg/mL) with a fixed anti-HCP antibody concentration of 50 µg/mL. Both immuno-complex and bead incubations were performed at 4°C for 24 hours.

In a final attempt to achieve a successful immuno-capture with this support, we decided to make one last adjustment to the incubation conditions. Both incubations were conducted for 1 hour with mild shaking, aiming to promote better interactions and assuming this might improve the immuno-capture process. Additionally, a much less concentrated antigen solution of 30 µg/mL was tested to minimize potential steric hindrance and enhance binding efficiency. However, as illustrated in Figure 31, the results were consistent with previous experiments. The immuno-capture process once again fell short of expectations, showing no noticeable improvement in binding.

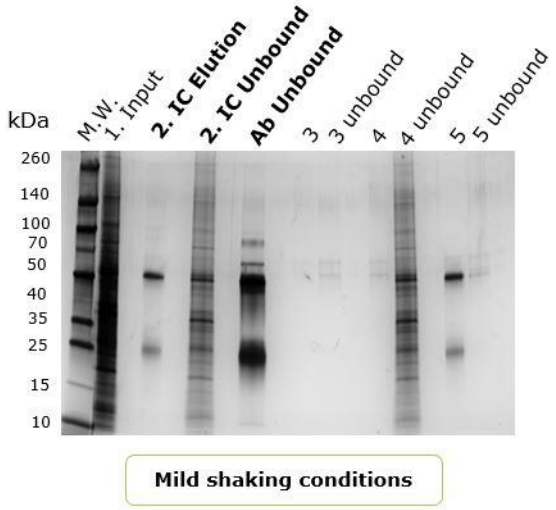


Figure 31: Silver stain gel from the final optimization of the plate support, incorporating mild shaking during incubations to enhance performance.

After numerous optimization attempts, we decided to return to the initial conditions, this time using the commercial generic couple from Cygnus. For this positive control, we also included the corresponding negative controls, which deviates from the standard practice described for the Protein A/G beads. Typically, the role of a positive control is to validate the performance of the support with an alternative antibody-antigen pair. However, due to the straightforward nature of the plate support system and the feasibility of conducting the controls alongside the primary experiment, we chose to incorporate them. This decision was made to provide a more thorough analysis of the results, even though it was not essential for confirming the support's behavior.

As illustrated in Figure 32, the results remained consistent with the earlier trials: the immuno-capture process showed no significant improvement. The commercial couple exhibited the same limitations as the process-specific couple, further highlighting the challenges of achieving effective binding under these conditions.

In light of these persistent issues, and much like with the magnetic beads, we concluded that the plate support was not suitable for the scope of our study. Despite adjustments to parameters such as incubation conditions and reagent concentrations, the support repeatedly failed to deliver the desired immuno-capture performance.

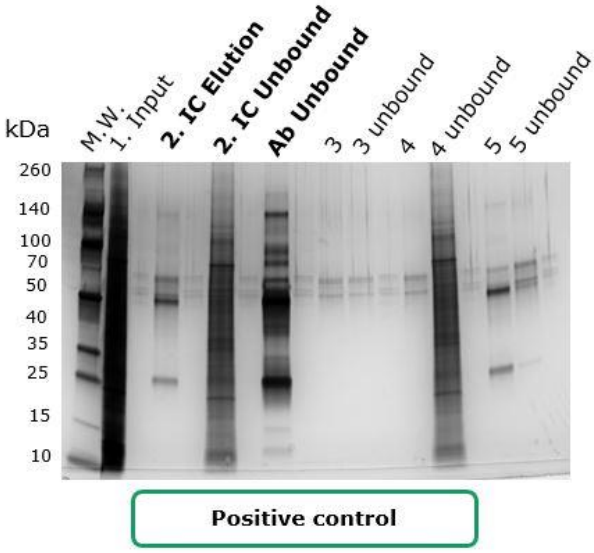


Figure 32: Silver stain gel of the positive control for the plate support, using the generic commercial anti-HCP antibody mix and antigen couple.

C. Protein A column

For the experiments involving the Protein A column, we decided to take a slightly different direction compared to the previous supports. Considering the substantial binding capacity that this type of column offers (20mg of IgG/mL), we chose to test much larger reagent ratios than before. This adjustment stemmed from the lack of success in earlier attempts and the need to explore a setup that could better align with the column's unique capabilities. By modifying the experimental dynamics in this way, we aimed to determine whether this alternative strategy would yield better outcomes in the immuno-capture process.

Regarding the negative controls for the Protein A column, these were approached slightly differently due to the consistent use of this support throughout the experiments. Specifically, the negative control for the antibody involved injecting 500 µg of the antibody into the column to verify its binding and elution. This was successfully confirmed during a dedicated run. Similarly, for the antigen, 250 µg were injected separately in another run. As expected, the entire antigen emerged during the washing steps, confirming that in the absence of the antibody, the antigen did not interact with the column support. For the negative control of the empty support, the column was subjected to washing steps, and the UV readings were monitored to ensure they remained at the baseline, confirming no non-specific interactions. Finally, the positive control was conducted at the end of the optimization process, consistent with the approach for the other supports, and the details of this control will be presented in the concluding section of this chapter.

i) Antibody to antigen ratio

For our initial optimization step, we focused on determining the ideal reagent ratios for the Protein A column. In this first trial, we utilized 1 mg of anti-HCP antibody and 1 mg of antigen, incubating the immuno-complex mixture at 4°C for 24 hours. A small sample of the input mixture was set aside for further analysis, and the remaining solution was injected into the ÄKTA chromatographic system for processing.

Figure 33 illustrates the chromatographic results of our initial experiment using the column support, along with a detailed workflow and the collected samples. The workflow begins with the injection of the incubated mixture into the column, where the unbound fraction, representing HCPs that did not interact with the anti-HCP antibodies and are therefore non-immunoreactive, is collected first. Subsequently, an elution step using a pH gradient releases the bound fraction, which includes the anti-HCP antibodies captured by the Protein A resin, along with the immunoreactive HCPs attached to them. The chromatogram clearly indicates a successful immuno-capture, with a significantly larger elution peak compared to the unbound fraction, providing a strong indication of efficient binding and separation under the tested conditions.

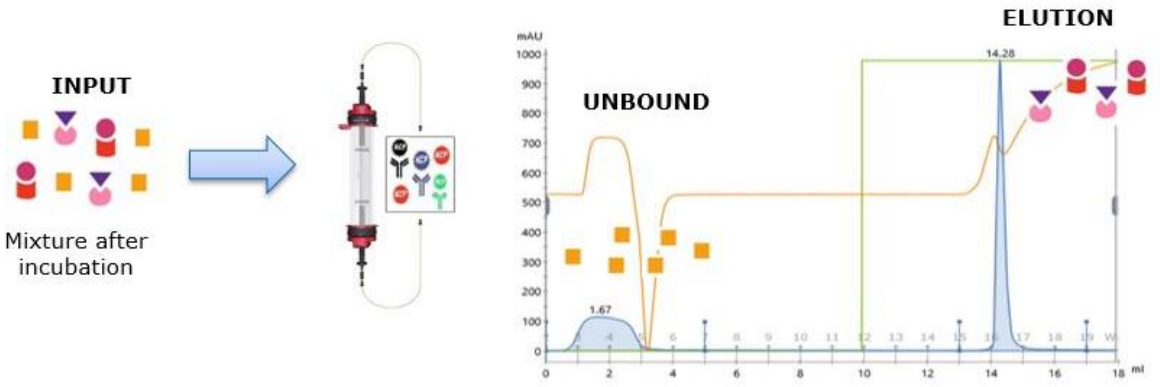


Figure 33: Illustration of sample recovery from the Protein A column support, accompanied by a representative chromatogram of a successful immuno-capture assay.

It is essential to emphasize that the unbound fraction holds particular significance for our study, as it contains the HCPs not detected by ELISA. Analyzing this fraction is crucial for identifying these undetected proteins and evaluating the comprehensiveness of the ELISA results. The promising success of this initial experiment with the column support, which yielded highly encouraging results, prompted us to explore different reagent ratios. Specifically, we tested two distinct ratios: 500 µg of anti-HCP antibody paired with 1 mg of antigen (a 1:2 ratio) and 1 mg of anti-HCP antibody with 500 µg of antigen (a 2:1 ratio). These experiments aimed to refine the balance between reagents and optimize the immuno-capture process further.

Unlike previous experiments, where samples were typically recovered in bulk, the column support experiment required the collection of each milliliter separately, as dictated by the chromatographic system's design. The fractions corresponding to the unbound and elution phases, as identified by the

UV signal, were loaded onto gels for further analysis. This allowed us to confirm the success of the immuno-capture process and assess the population of HCPs in each fraction through SDS-PAGE.

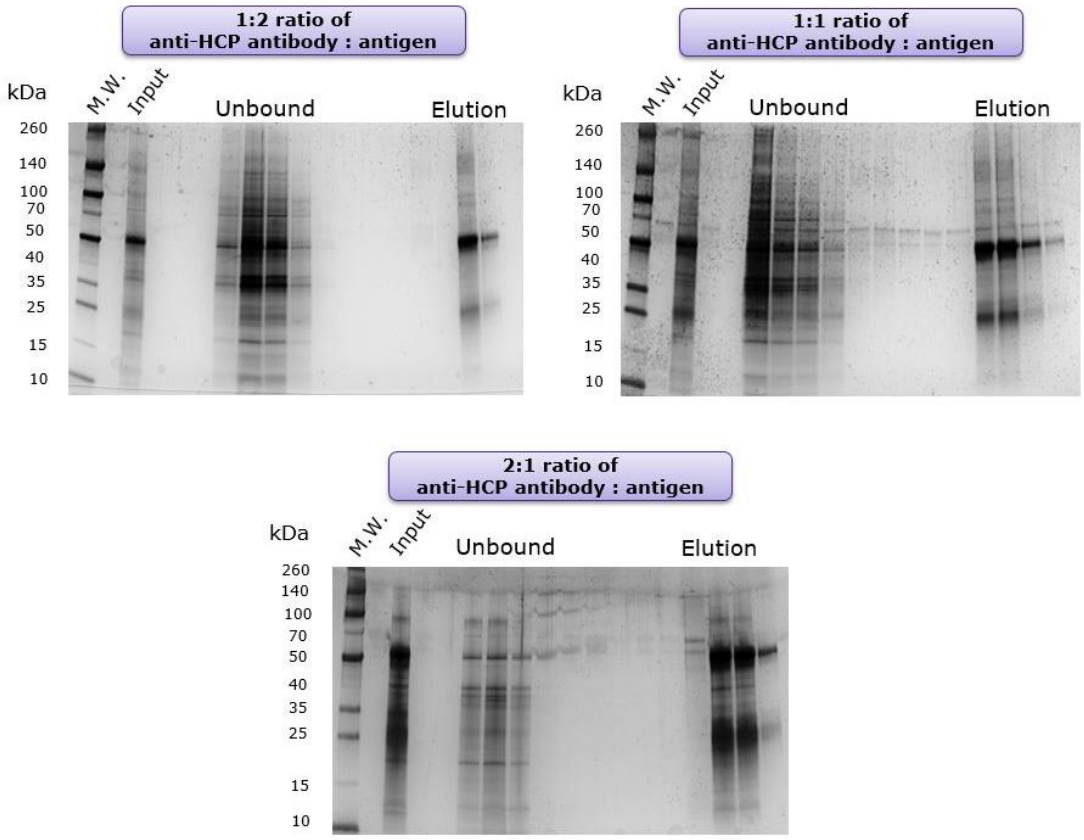


Figure 34: Silver stain gels showing the results of varying anti-HCP antibody-to-antigen ratios, with corresponding lanes for input sample, unbound fraction, and elution fraction following the assay.

As illustrated in Figure 34, the elution lanes clearly demonstrate the presence of HCPs, further validating the immuno-capture process. The antibody effectively captured the immuno-reactive HCPs from the antigen under the tested conditions. Encouraged by these promising results, it was decided to employ a more detailed analytical method to gain deeper insights into the samples obtained. To explore this, another 2:1 ratio experiment was conducted, using 500 µg of anti-HCP antibody and 250 µg of antigen. The chromatogram indicated a similar outcome to the previous experiments, and the fractions corresponding to the unbound and elution phases were recovered for analysis by 2D SDS-PAGE and subsequent Western blot.

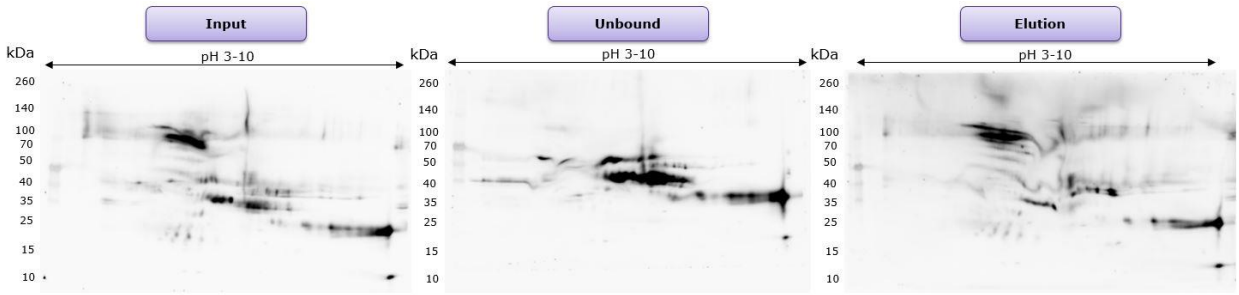


Figure 35: 2D Western blot of the input, unbound, and elution samples from the column Protein A support.

Unfortunately, these methods did not yield clear or satisfactory results, as the gels displayed no discernible spots. The Western blot, as shown in Figure 35, revealed inconsistencies in the results. While the input and elution samples produced comparable patterns, as expected, the unbound lane presented a markedly different profile. This divergence was surprising, as we anticipated that the combined profiles of the elution and unbound fractions would align with the input. The unexpected differences in the unbound fraction could be attributed to the complexity of using a polyclonal antibody mixture. Polyclonal antibodies target a diverse range of epitopes, and the HCPs themselves exhibit varying levels of immunogenicity, complicating the identification and assessment of HCPs present in the unbound fraction. Given these challenges and the lack of overlap between the fractions, it was decided to adopt a more sophisticated analytical approach, namely 2D-DIGE, to better evaluate the results and the samples obtained.

2D-DIGE offered the advantage of loading the antigen sample, unbound fraction, and elution fraction onto a single gel, enabling a comprehensive assessment of HCP overlap across these fractions. In Figure 36, the antigen sample is represented in green, the elution fraction in blue, and the unbound fraction in red. The primary goal of this analysis was to identify overlaps between the antigen (mock or null cell line) and the unbound fraction, indicated by yellow regions on the image. A minimal presence of yellow is desirable, as it would suggest that most HCPs were successfully captured by the anti-HCP antibody, leaving a low population in the unbound fraction.

While analyzing the elution fraction would have added valuable insights, this was hindered by the strong signal saturation caused by the anti-HCP antibody. This saturation obscured any HCPs present in the elution sample, emphasizing the challenges of working with polyclonal antibodies and their wide range of target specificities within a diverse HCP population.

From the image, it is evident that the 2:1 ratio yielded less favorable results, with a noticeable abundance of yellow overlap, signifying an excessive presence of unbound HCPs. Meanwhile, the 1:2 ratio showed relatively better results in terms of HCP capture but displayed smearing and irregular migration patterns, which made the overlap appear inconsistent and inconclusive. This inconsistency, coupled with the excess of antigen in the 1:2 ratio, suggested that this condition was not well-suited for achieving the desired outcome. On the other hand, the 1:1 ratio presented the most balanced and reliable results, with minimal yellow regions and proper migration patterns, indicating efficient

immuno-capture. Based on these observations, we chose to proceed with the 1:1 ratio as it provided a more reliable and appropriate condition for further optimization.

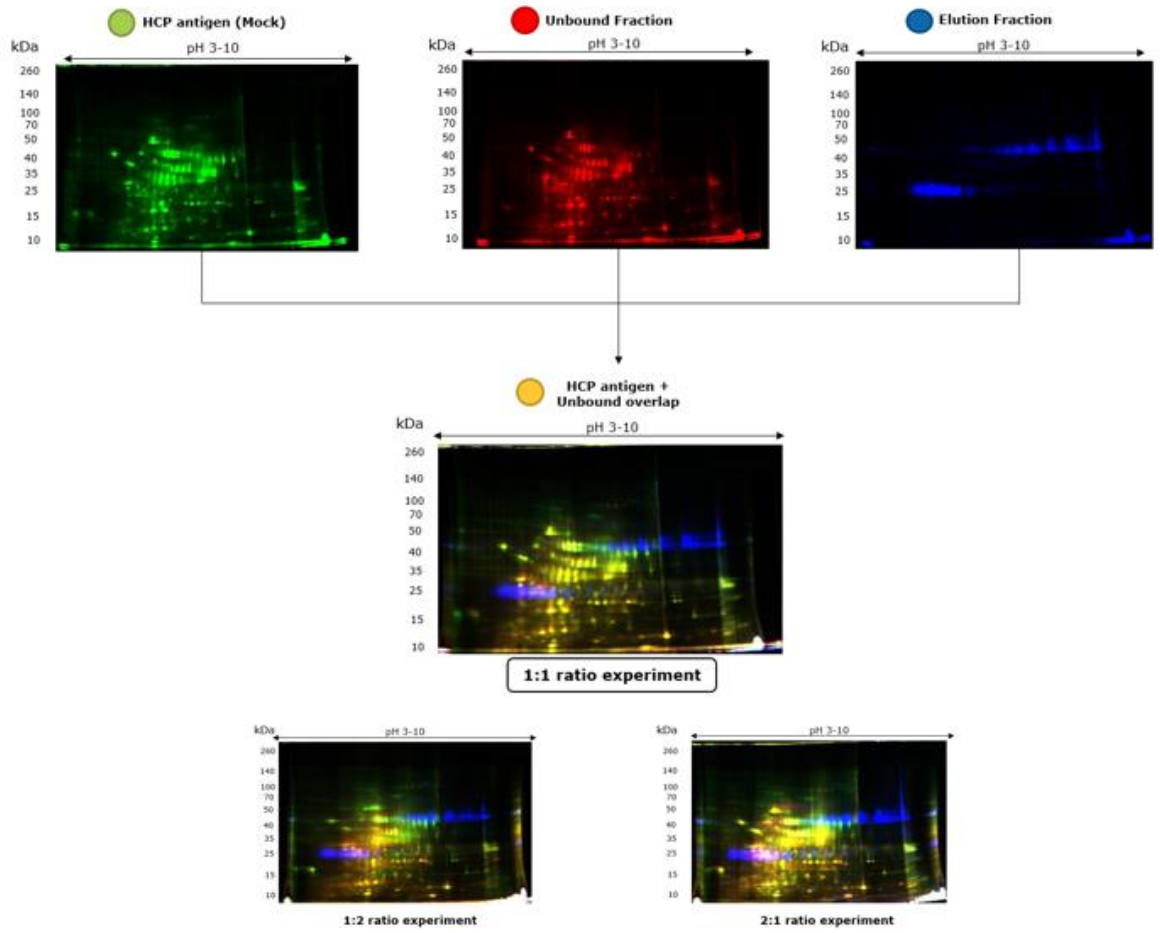


Figure 36: 2D DIGE analysis of the ratio optimization for the column support. At the top, the three separate channels display antigen (green), unbound (red), and elution (blue). The middle section presents the overlay of the three channels of the 1:1 ratio, highlighting the condition with the best outcome. At the bottom, results from two additional conditions at A:2 and 2:1 ratios are shown, which demonstrated less satisfactory outcomes.

ii) Temperature and incubation time

To optimize the immuno-complex formation, we explored adjustments to incubation temperature, time, and volume, integrating biologically relevant and practical considerations. The idea behind testing 37°C was to mimic physiological conditions similar to the host animal's temperature during the immunization campaign that produced the anti-HCP antibody mix. By reflecting this environment, we aimed to determine whether such conditions would enhance the affinity of the anti-HCP antibody mix for the antigen. In addition, increasing the temperature was hypothesized to accelerate molecular interactions, potentially reducing the required incubation time to 2 hours without compromising binding efficiency.

We also examined the effect of varying the incubation volume on the immuno-capture process. Two volumes were tested: 300 µL, a standard volume for antigen-antibody mixture with the reagents we employed, and 2 mL, which matches the injection volume required for the Protein A column. The rationale was that a larger volume might reduce steric hindrance, providing greater spatial freedom for antigen-antibody interactions and thereby improving binding efficiency. This comprehensive approach aimed to refine the conditions for immuno-complex formation, balancing biological relevance and operational practicality to maximize the efficiency of HCP capture.

While the outcomes of the experiments under the four tested conditions showed comparable patterns in the presence of yellow in the gel, as illustrated in Figure 37, this similarity in appearance made it challenging to draw definitive conclusions about the precise impact of variations in temperature and incubation volume on immuno-capture efficiency. Although differences may exist, the visual assessment provided by 2D DIGE was insufficient to distinguish these variations with clarity. While 2D DIGE has been invaluable in assessing the overall success of the immuno-capture process, its limitations in providing conclusive insights under these conditions highlight the need for a more advanced analytical approach. Consequently, transitioning to mass spectrometry was necessary to achieve a more thorough and detailed evaluation of the experimental results.

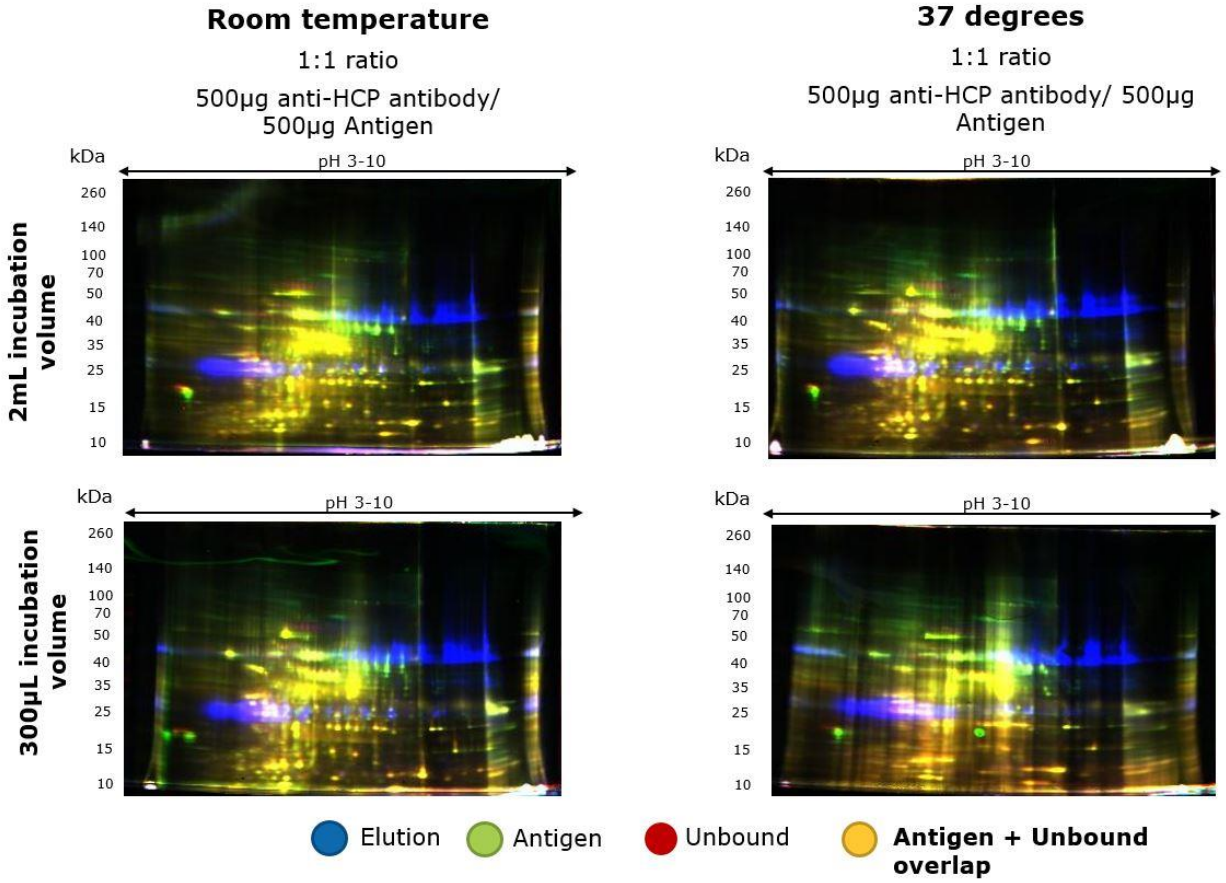


Figure 37: 2D DIGE analysis of temperature and volume optimization for the column. The top section shows results for a 2 mL mixture, while the bottom section displays results for a 300 µL mixture. Similarly, outcomes at room temperature are presented on the left side, whereas results at 37°C are displayed on the right side.

iii) Comparison of process-specific couple and a commercial antibody/antigen couple

After a lengthy journey of optimization, it was determined that the parameters established during the reagent ratio optimization, specifically the 1:1 ratio with overnight incubation at 4°C, would be used for testing the positive control. As shown in Figure 38, the product-specific antibody-antigen pair outperformed the generic one from Cygnus. The unbound fraction for the product-specific couple was significantly smaller, indicating that a higher proportion of the immuno-complex successfully bound to the column and was subsequently eluted, with much fewer non-immunoreactive proteins present in this fraction.

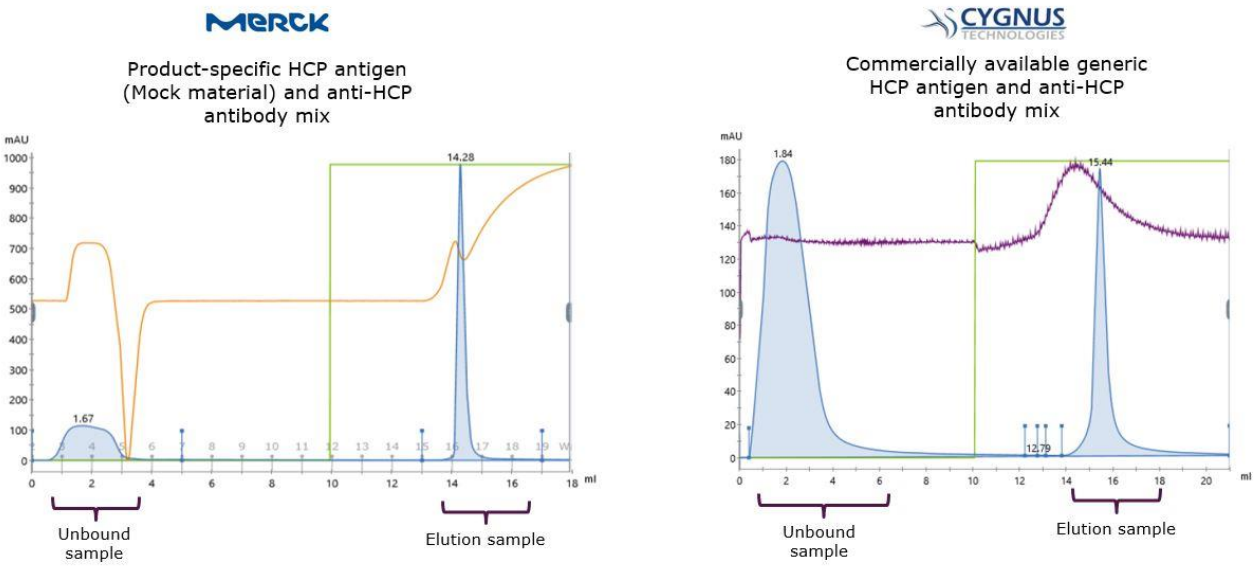


Figure 38: Negative control outcomes illustrated through chromatogram comparison. On the left, the chromatogram of a successful immuno-capture assay using the process-specific anti-HCP antibody and antigen is shown. On the right, the chromatogram from an immuno-capture process employing the generic commercial antibody-antigen couple is displayed, highlighting the superior performance of the process-specific couple over the latter.

In contrast, the generic couple demonstrated a considerable loss of material during the initial washing steps, suggesting that many HCPs failed to bind to either the anti-HCP antibody. These results strongly highlight the critical importance of developing a product-specific antibody pair, particularly in the later stages of drug development as the process moves closer to commercial release. Moreover, these findings emphasize the necessity of utilizing a method such as ours to validate and assess the accuracy of ELISA results, especially when using a commercially available generic anti-HCP antibody, where the affinity may not be as robust as that of a product-specific couple.

Chapter 2
Mass Spectrometry Method Developments for an in-depth characterization of the unbound/ELISA-blind HCP population

1. Mass Spectrometry method set-up

The optimization of an immuno-capture support was a critical step in this workflow. After evaluating several options, including magnetic beads, plates, and columns, the latter emerged as the most optimal and suitable for our objectives. This conclusion was drawn after extensive experimentation, which yielded multiple encouraging results. Initial assessments of the column's performance were conducted using SDS-PAGE, Western Blot, and 2D DIGE. However, these techniques posed limitations in pinpointing the most optimal experimental conditions due to their qualitative nature.

The transition to mass spectrometry marked a pivotal step toward achieving deeper insights. While the initial optimization of the immuno-capture column was indispensable, incorporating mass spectrometry was the next crucial phase. This advanced approach facilitates the identification and characterization of host cell proteins (HCPs) in the unbound fractions, specifically targeting those proteins that remain undetected by ELISA. By leveraging mass spectrometry, it became possible to overcome the limitations of this conventional immunoassay and gain a more comprehensive understanding of the HCP population.

The analytical workflow is summarized in Figure 39, which visually outlines the key steps of this process. A key component of this strategy was the development of an exhaustive spectral library. This library was built for both antigen samples: the null cell line provided by Merck and the antigen obtained from Cygnus. The CHO database, containing over 78,000 entries, was selected as the foundation for this library. However, due to the database's inherent redundancy and limited curation compared to more established resources like the human proteome, constructing a tailored spectral library was essential to ensure confident and reliable protein identification.

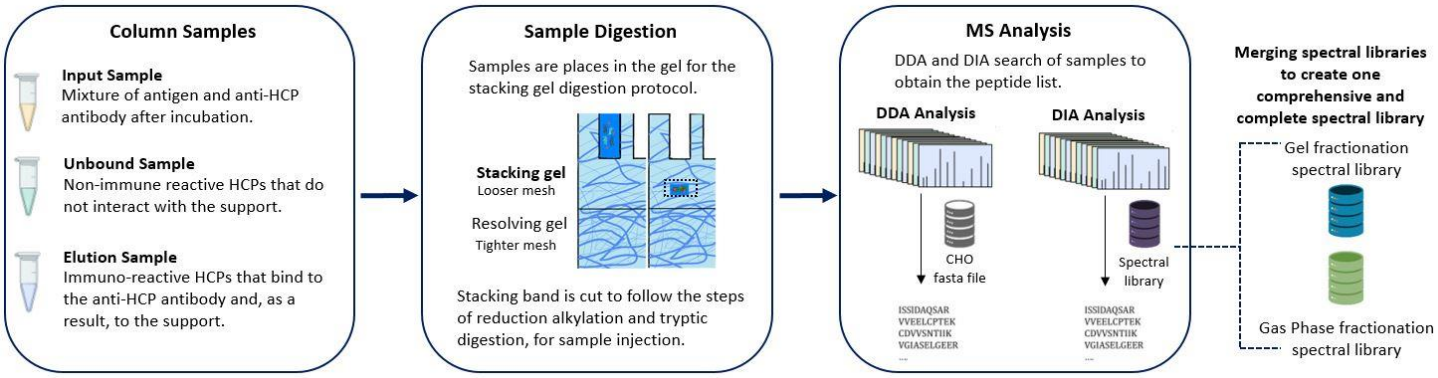


Figure 39: Analytical workflow employed for the analysis of samples retrieved from column support experiments. The workflow outlines the steps from sample preparation to the gel digestion protocol, followed by the analysis performed using DIA search with the complete antigen library derived from GPF and gel fractionation runs.

Two distinct methods were tested for building the spectral library, as further illustrated in Figure 40 (below), each offering unique advantages. The first approach involved the well-established gel fractionation method, which separates the sample into fractions before injection in Data-Dependent Acquisition (DDA) mode. This method leverages fractionation to reduce sample complexity, improving protein identification coverage. The second approach, gas-phase fractionation (GPF)[230], represents a more innovative technique. GPF involves consecutive Data-Independent Acquisition (DIA) injections of the same sample, each targeting different m/z windows. This approach bypasses the need for physical fractionation while achieving extensive proteomic coverage by systematically scanning the mass range. Comparing these two methods provided valuable insights into their respective performance and suitability for library building.

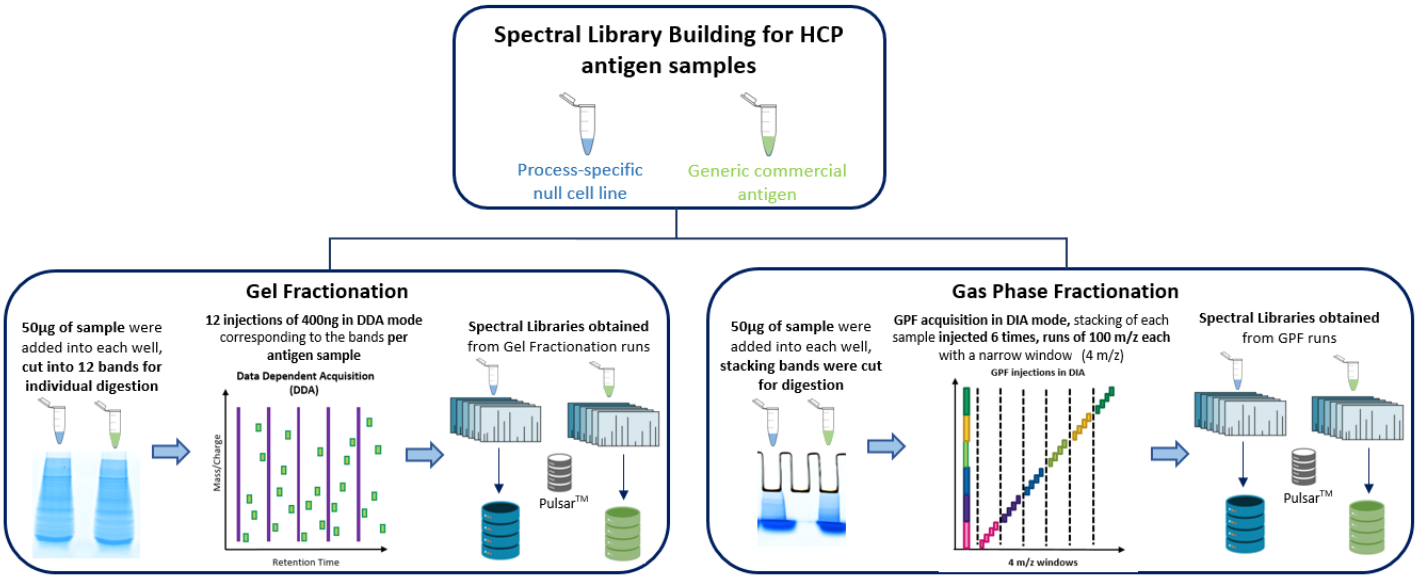


Figure 40: Spectral library building workflow. The illustration depicts both approaches used for spectral library construction: the more established gel fractionation on the left and the more recent gas phase fractionation on the right.

2. Sample Preparation and Injection

A. Column Protein A support samples

To ensure accurate and consistent protein amounts for enzymatic digestion, a bicinchoninic acid (BCA) assay was conducted using the Pierce™ BCA protein assay kit (Thermo Scientific™). This method was chosen not only for its sensitivity and reliability but also for its compatibility with the buffers used in the immuno-capture experiments. The BCA assay complemented the protein quantification previously performed using a NanoDrop™ spectrophotometer during my stay at Merck Guidonia, providing a confirmation of protein concentrations for the following digestion step.

An in-gel digestion approach was employed for these samples, as it effectively removes contaminants such as detergents, salts, and other interfering substances. A total of 10µg of protein from the immuno-capture experiments was loaded into a low-percentage polyacrylamide gel. The gel's concentrating zone compressed proteins into a narrow band, creating a unified starting point for subsequent processing. This ensured uniform positioning of the proteins within the gel prior to separation. Instead of allowing the proteins to migrate fully into the resolving gel, in-gel digestion was performed while they remained within the concentrating zone for subsequent recovery and digestion steps.

To prepare the proteins for digestion, the gel bands were subjected to reduction with 10 mM DTT for 30 minutes at 60°C to break disulfide bonds, followed by alkylation with 55 mM iodoacetamide (IAA) for 30 minutes in the dark to prevent reformation of these bonds. Trypsin (Promega) was added at a 1:25 enzyme-to-protein ratio, and the samples were incubated overnight at 37°C for approximately 15 to 15 hours to ensure complete digestion. The reaction was subsequently quenched, and peptides

were extracted from the gel bands using a sequential process with 0.1% formic acid (FA) and 80% acetonitrile (ACN) under gentle agitation at 110 rpm for 1 hour.

After extraction, peptides were vacuum-dried and reconstituted in a buffer of 2% ACN and 0.1% FA to achieve a final concentration of 100 ng/µL. For samples designated for DIA acquisition mode, internal retention time (iRT) standards (Biognosys) were added to the peptide solution at 10% of the final injection volume. These standards facilitated accurate alignment of retention times, further enhancing the reliability of protein identification and quantification.

For mass spectrometry analysis, samples obtained from the column optimization experiments were injected into a Dionex Ultimate 3000 RSLC nano system (Thermo Fisher Scientific). The analysis was performed using the Orbitrap Eclipse™ Tribrid™ mass spectrometer (Thermo Fisher Scientific), ensuring high sensitivity and resolution. The UPLC system featured an Acclaim PepMap 100 C18 pre-column (20 mm × 0.1 mm, 5 µm, 100 Å) from Thermo Fisher Scientific, coupled with an Aurora Series C18 UHPLC analytical column (250 mm × 75 µm, 1.6 µm, 120 Å) from IonOpticks.

The separation process utilized a binary mobile phase system. Mobile phase A consisted of water with 0.1% formic acid (FA), while mobile phase B was a mixture of acetonitrile (ACN), water, and 0.1% FA (79.9/20/0.1). Peptides were loaded at a steady flow rate of 300 nL/min and resolved over an 85-minute gradient. During the gradient, the proportion of mobile phase B gradually increased, reaching 43.8% by the 65-minute mark. At minute 66, the gradient reached 98% for column washing and was maintained until minute 71. Re-equilibration of the system was achieved by reducing the proportion of phase B to 2.5%, which was maintained from minute 72 to 85 to ensure system stability for subsequent analyses.

B. Library Building approaches

To prepare the samples for the spectral library building, two distinct fractionation approaches were considered: gel fractionation and gas phase fractionation (GPF). Each method offers unique advantages in terms of sample complexity reduction and efficiency. The gel fractionation method, although slightly more time-consuming, remains a reliable and widely used technique in proteomic workflows.

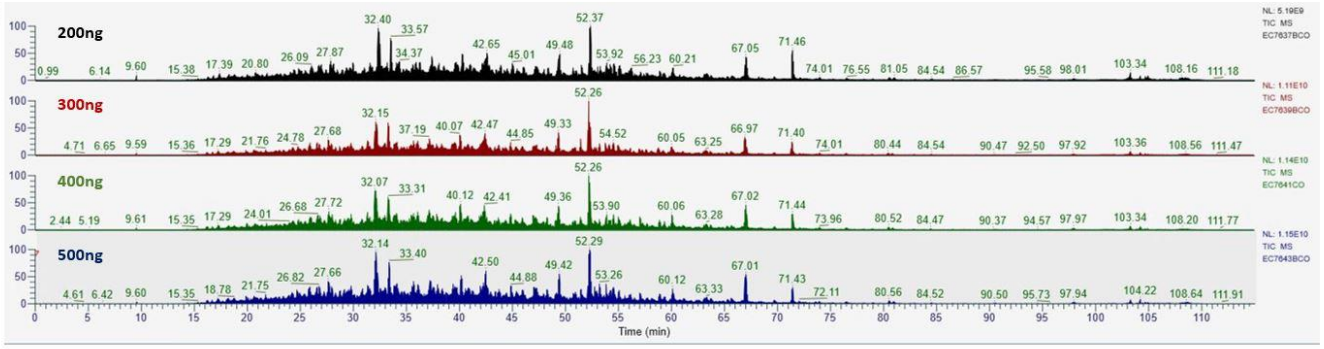
For the gel fractionation method, the samples were allowed to migrate through the polyacrylamide gel. After staining and visualizing the proteins, the lane was carefully sliced into 12 distinct bands. Each band was subjected to a standardized sample preparation protocol. Proteins were reduced in-gel with 10 mM DTT for 30 minutes at 60°C, followed by alkylation with 55 mM iodoacetamide (IAA) in the dark for 30 minutes. Digestion was carried out with trypsin at an enzyme-to-protein ratio of 1:50, with incubation at 37°C overnight for up to 15 hours. Peptides were extracted using 0.1% formic acid (FA) and 80% acetonitrile (ACN) with gentle agitation for 1 hour. Finally, the samples were vacuum dried and reconstituted in 2% ACN and 0.1% FA to achieve a final concentration of 100 ng/µL for injection.

Prior to the development of the spectral library, it was also essential to determine the saturation point at which the highest number of protein groups could be identified. This step was critical to optimize

the amount of sample injected for the spectral library runs. To achieve this, a 50/50 mixture of both antigens was prepared and analyzed using the mass spectrometer in DDA mode.

Various sample quantities, ranging from 200 ng to 500 ng, were tested in four separate injections. The analysis revealed that while the 500 ng injection resulted in the highest signal intensity, but it did not correspond to the highest number of identified protein groups. Interestingly, the 400 ng injection, which exhibited slightly lower intensity, yielded the maximum number of protein groups following Mascot search and Proline validation (assuring an FDR below 1%). Therefore, 400 ng was selected as the optimal quantity for the spectral library injections. Figure 41 presents the chromatograms and provides detailed insights into the quantities injected.

Ⓐ



Ⓑ

Quantity injected (ng)	Intensity	Number of Protein Groups (FDR<1%)
200	$5{,}19 \times 10^{9}$	1650
300	$1{,}11 \times 10^{10}$	1809
400	$\mathbf{1{,}14 \times 10^{10}}$	**2005**
500	$1{,}15 \times 10^{10}$	1983

Figure 41: A) Chromatograms corresponding to the 200 ng, 300 ng, 400 ng, and 500 ng injections. B) Table summarizing the signal intensity and the number of protein groups identified for each injection quantity (FDR<1%).

Gas phase fractionation followed a slightly modified workflow. Instead of allowing the proteins to migrate into the resolving gel, the samples were concentrated into a single band within the stacking layer of a low-percentage polyacrylamide gel. This avoided physical separation of the proteins while still enabling the removal of contaminants. From this point onward, the same sample preparation protocol used for gel fractionation, including reduction, alkylation, digestion, and peptide extraction, was applied. For GPF, the prepared peptides were sequentially injected into the mass spectrometer, with each run targeting a specific m/z range, enabling comprehensive peptide identification.

3. MS/MS Data Acquisition Method Optimizations

For the gel-band fractionation library built using data-dependent acquisition (DDA), full-scan MS spectra were collected across a broader range of 375–1500 m/z, at a resolution of 120,000 (at 200

m/z). The AGC target was set to 8×10^5, with a maximum injection time of 60 milliseconds. From each MS scan, the top 20 most abundant precursor ions with charge states between 2 and 6, and with intensities exceeding a threshold of 2×10^4 ions per second, were automatically selected for fragmentation. Higher-energy collisional dissociation (HCD) was performed at 30% normalized collision energy, and the resulting MS/MS spectra were recorded in the ion trap in Rapid mode. MS/MS scans spanned the same 375–1500 m/z range, with an AGC target of 5×10^4 and a maximum injection time of 54 milliseconds.

For the gas-phase fractionation (GPF) library built in data-independent acquisition (DIA) mode, six consecutive injections were performed to cover the mass range of 400–1000 m/z. Each full-scan MS spanned a 100 m/z window, utilizing the same parameters as those previously established for DIA acquisition. The MS/MS analysis was divided into 25 windows, each covering 4 m/z. These analyses were conducted at a resolution of 30,000 (at 200 m/z), with an automatic gain control (AGC) target of 5×10^5 and a maximum injection time of 54 milliseconds. This setup ensured high spectral coverage and reproducibility for peptide identification.

The goal of this comparative analysis was two-fold: first, to evaluate the efficacy of a library-free approach (DirectDIA), which eliminates the need for a pre-built spectral library by directly analyzing the raw data against the database; and second, to leverage the advantages of a library-based method, which relies on a spectral library to enhance the confidence and depth of protein identifications.

Consecutively to the spectral library building, mass spectrometry data were acquired across a mass range of 375–1200 m/z in full-scan mode, with a resolution of 120,000. The AGC target was set at 8×10^5, and the maximum injection time was 200 ms, ensuring a high level of sensitivity and accuracy. For MS/MS scans, spectra were recorded at a resolution of 15,000 with a maximum injection time of 22 ms. Peptide fragmentation was achieved using higher-energy collision-induced dissociation (HCD) with a normalized collision energy of 30%. Fragment analysis employed 18.4 m/z isolation windows with a 0.5 m/z overlap, enhancing the accuracy of the peptide identifications.

Protein identification were conducted using Spectronaut™ software (v.18), which processed raw data in combination with the CHO database (Cricetulus griseus entries from UniProtKB/TrEMBL). The analysis employed trypsin/P digestion, permitting one missed cleavage site per peptide. Additionally, methionine oxidation and N-terminal acetylation were accounted for as variable modifications. To ensure accurate identification, a Q-value cutoff of 0.01 was applied for both precursor ions and proteins.

4. Building a Comprehensive Mass Spectrometry Library

A. Library comparison for product-specific antigen

Initially, a 10µg quantity of starting material was used to create the spectral libraries for both gel fractionation and Gas Phase Fractionation (GPF). The results from this initial attempt showed that both methods identified a similar number of protein groups: gel fractionation identified 3290 protein

groups, and GPF identified 3286 protein groups. This was an encouraging outcome, particularly for GPF, which is generally less labor-intensive and reagent consuming compared to gel fractionation.

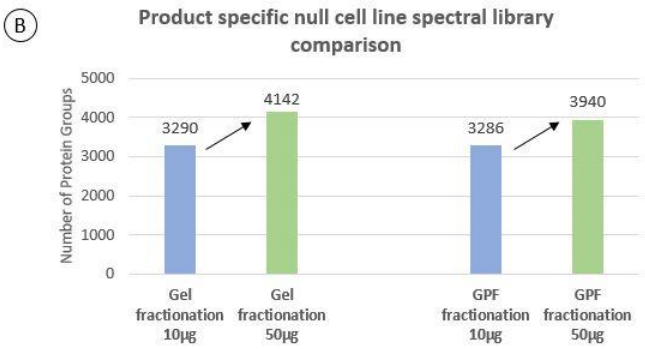
However, while the protein group counts were similar, a difference was observed in the number of peptides and precursors identified. Gel fractionation outperformed GPF with about 43% more peptides and precursors detected. This suggests that while both methods provide comparable protein group identifications, the peptide numbers were significantly lower in GPF compared to in-gel fractionation.

In response to the limitations observed with the 10 µg experiment, where the number of precursors, peptides, and protein groups appeared modest, the amount of input material was increased to 50 µg for both gel fractionation and GPF. This adjustment led to a notable improvement in the results, with gel fractionation identifying 4142 protein groups, corresponding to a 25.9% increase, and GPF identifying 3940 protein groups, for a 19.9% increase. These improvements highlight the significant impact of the starting material quantity on the depth of proteomic coverage. Interestingly, with the increased input, gel fractionation continued to identify more protein groups than GPF, and the difference between the two methods grew slightly. The increased peptide and precursor numbers observed with gel fractionation at 10 µg were also maintained at 50 µg.

To further visualize the overlap between the two libraries built from the 50 µg input experiment, we created a Venn diagram. A total of 2655 protein groups were shared between the two methods, highlighting the substantial overlap. Gel fractionation identified 4142 protein groups, with 1177 (26.4%) of them being unique, while GPF identified 3940 protein groups, with 1173 (26.3%) being unique. This Venn diagram, shown in Figure 42, illustrates these improvements, showing the trends in protein group identifications as the input material was increased. The substantial number of unique protein groups identified by each method further emphasizes the complementary strengths of gel fractionation and GPF, as both contribute distinct, valuable information to the overall proteomic analysis.

A

Spectral Library	Precursors	Peptides	Protein Groups
Gel fractionation 10µg	38660	32452	3290
Gel fractionation 50µg	56656	43706	4142
GPF fractionation 10µg	28252	23672	3286
GPF fractionation 50µg	46854	36346	3940



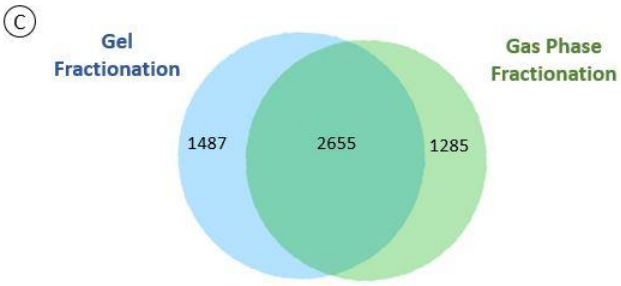


Figure 42: **A)** Table depicting the number of precursors, peptides, and protein groups identified according to the spectral library built and the quantity of antigen used. **B)** Histogram highlighting the increase in the number of protein groups influenced by the augmentation of initial material for library building. **C)** Venn diagram depicting the overlap between the gel fractionation and gas phase fractionation libraries for 50 µg initial material.

To fully leverage the strengths of both methods and maximize protein identifications, the libraries created with 50 µg were merged. This integration resulted in a comprehensive spectral library containing 5477 protein groups. By combining the unique advantages of both gel fractionation and GPF, the merged library provided a more exhaustive resource, enhancing both protein detection and characterization. This combined library was then used for further mass spectrometry analyses, specifically targeting the identification of HCPs in the unbound fractions of the column optimization samples. This strategy allowed us to achieve the highest possible proteomic coverage, benefiting from the complementary strengths of both techniques.

B. Library comparison for commercial antigen

For the commercial antigen, similar trends were observed as with the Merck antigen. Initially, 10 µg of starting material was used for both gel fractionation and GPF. As expected, the protein group identification results were comparable, with gel fractionation identifying 3969 protein groups and GPF identifying 3504. These numbers were again consistent with the findings for the process-specific antigen. Peptide and precursor numbers followed the same trends observed previously, with gel fractionation providing a higher yield of peptides and precursors compared to GPF.

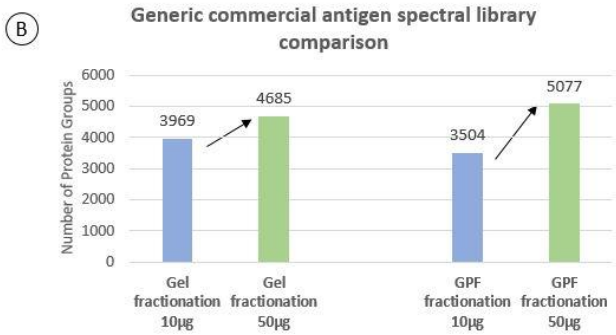
Given the modest results from the 10 µg test, where the number of protein groups identified was relatively low, the starting material was increased to 50 µg for both gel fractionation and GPF. This

adjustment led to a significant increase in protein group identifications. Interestingly, in this case, GPF outperformed gel fractionation, identifying 5077 protein groups compared to 4685 identified by gel fractionation. Moreover, GPF also showed superior peptide and precursor numbers, confirming its potential as a very good candidate for library building.

To further investigate the overlap and uniqueness of the protein groups identified by the two methods, we created a Venn diagram. A total of 3363 protein groups were shared between the two methods, highlighting the substantial overlap. Gel fractionation identified 4685 protein groups, with 1666 (29.5%) uniquely identified, while GPF identified 5077 protein groups, with 1201 (21.5%) uniquely identified. This Venn diagram, shown in Figure 43, illustrates these improvements, showing the trends in protein group identifications as the input material was increased. The substantial number of unique protein groups identified by each method emphasizes the complementary strengths of gel fractionation and GPF, with both contributing valuable information to the overall proteomic analysis.

A

Spectral Library	Precursors	Peptides	Protein Groups
Gel fractionation 10µg	40042	34000	3969
Gel fractionation 50µg	64460	49684	4685
GPF fractionation 10µg	25033	21532	3504
GPF fractionation 50µg	65086	50134	5077



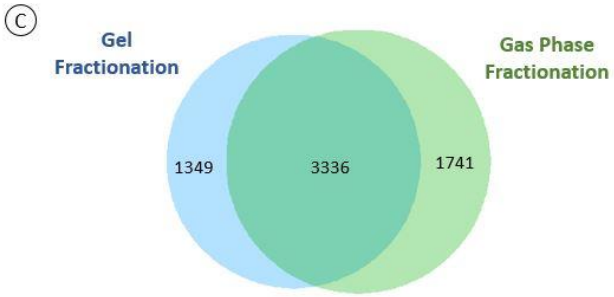
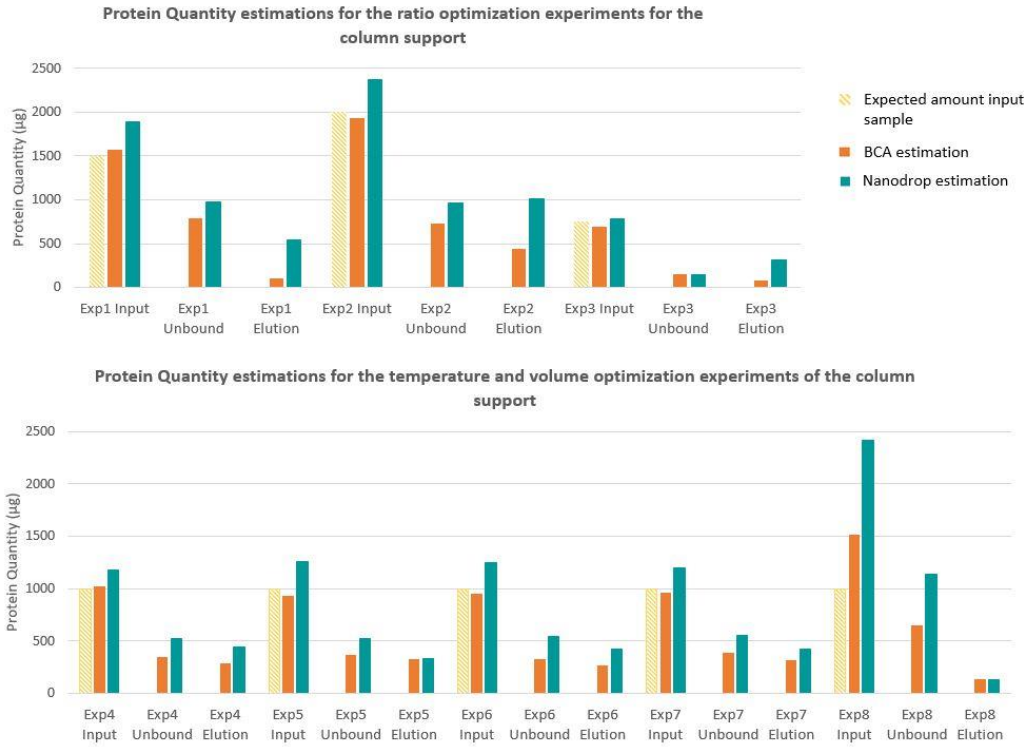


Figure 43: **A)** Table depicting the number of precursors, peptides, and protein groups identified according to the spectral library built and the quantity of antigen used. **B)** Histogram highlighting the increase in the number of protein groups influenced by the augmentation of initial material for library building. **C)** Venn diagram depicting the overlap between the gel fractionation and gas phase fractionation libraries for 50 µg initial material.

In light of these results, and to ensure the most comprehensive identification of protein groups, the 50µg libraries from both gel fractionation and GPF were merged. The combined library, containing protein groups identified by both methods, served as an exhaustive resource for subsequent sample analyses. Therefore, a final library of 5422 unique protein groups was then obtained, providing a robust foundation for further analysis and ensuring the most complete identification of proteins from the samples.

5. Mass Spectrometry Column Results and Discussion

As outlined previously, the initial step in processing the column samples involved quantifying the protein content to ensure accuracy and consistency in subsequent analyses. This was accomplished using a BCA assay, which was compared against results obtained from the Nanodrop spectrophotometer. While the Nanodrop provided a rapid estimation of protein concentration, it demonstrated a tendency to overestimate the protein quantity, as further illustrated in Figure 44. To aid in the comparison, the expected protein input was calculated, and illustrated in the histogram, from the known quantities of the anti-HCP antibody and antigen used in the immuno-complex formation.



Experiment number	Experiment condition
Exp1	1:2 ratio 500µg anti-HCP antibody/ 1mg antigen
Exp2	1:1 ratio 1mg anti-HCP antibody/ 1mg antigen
Exp3	2:1 ratio 500µg anti-HCP antibody/250µg antigen
Exp4	1:1 ratio 300µL volume at 25 degrees
Exp5	1:1 ratio 2mL volume at 25 degrees
Exp6	1:1 ratio 300µL volume at 37 degrees
Exp7	1:1 ratio 2mL volume at 37 degrees
Exp8	Positive control

Figure 44: Comparison of Nanodrop and BCA protein quantity estimations across column optimization experiments. The upper histogram highlights the discrepancies in protein quantification for the ratio optimization experiments, while the lower histogram compares estimations obtained from temperature and volume optimization experiments. A corresponding table provides a detailed explanation of the experimental abbreviations and outlines the specific conditions for each optimization.

The BCA assay provided more reliable and coherent results, aligning better with the expected protein quantities. While the Nanodrop estimations were deemed adequate enough for preliminary techniques such as SDS-PAGE and 2D DIGE for a first assessment, their tendency to overestimate necessitated the adoption of BCA assay results for MS analyses. This decision ensured that all MS experiments proceeded with accurately measured inputs.

A. Antigen: Antibody ratio optimization findings

For the column ratio optimization, samples from three distinct experiments were digested, injected, and processed to identify proteins. The focus of this analysis was the unbound fraction, specifically, the HCPs that do not bind to the anti-HCP antibody. This step was crucial in aligning with the overarching goal of the study: to characterize the HCPs present in the unbound fraction and therefore better assess and verify the ELISA test results.

To establish an analysis workflow for the subsequent experiments, the search assessment was first performed on the initial experiment of the series, which involved the 1:2 ratio of anti-HCP antibody and antigen (500 µg of antibody mix and 1 mg of the null cell line mix). This evaluation included a comparison of DDA, DIA using the gas-phase fractionation library, DIA with the gel fractionation library, and DIA with the comprehensive spectral library. The insights gained from this comparison informed the selection of the most effective search strategy for the remaining experiments.

The results for this assessment are detailed in Figure 45, where the mixed exhaustive library, built from the gel fractionation and GPF libraries, outperforms the other two libraries mentioned as well as the DDA search. To provide a more detailed assessment of the different search methods, the comparison between DDA and DIA revealed a striking improvement in protein group identifications when using DIA. Across all fractions, DIA consistently outperformed DDA, identifying roughly 3 times more protein groups on average. For instance, in the Input fraction, DDA identified 400 protein groups, whereas DIA searches using the gel fractionation, gas-phase fractionation, and mixed exhaustive libraries identified 900, 1050, and 1200 protein groups, respectively. This represents an improvement of 125–200% over DDA, further underscoring the comprehensive nature of DIA's data acquisition. Among the DIA searches, the mixed exhaustive library proved to be the most appropriate, consistently yielding the highest number of identifications in all fractions. This highlights the complementary value of combining the gel and gas-phase fractionation approaches for spectral library construction, which had already been discussed in previous chapters.

This behavior is due to DIA’s ability to collect data from all precursors within a defined window, resulting in a more comprehensive representation of the HCP population present in the samples. In contrast, DDA typically focuses on the most abundant precursors, which can lead to the loss of lower-abundance proteins. Given the superior performance of DIA, we decided to proceed with the search using the mixed exhaustive library for the rest of the samples.

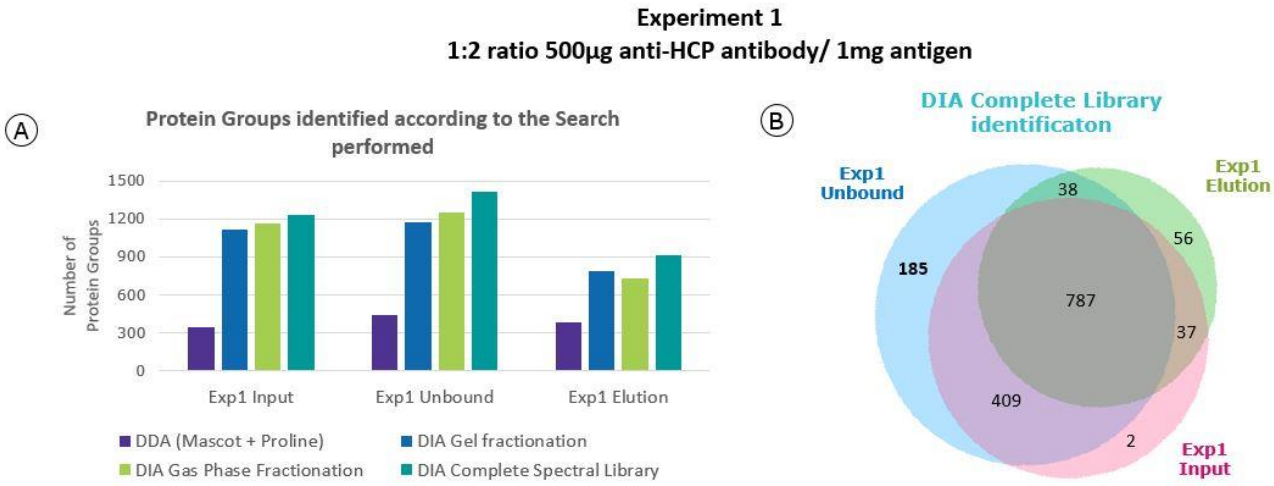


Figure 45: **A)** Histogram comparing the number of protein groups identified using different search strategies for the first ratio optimization of the column (1:2 ratio of anti-HCP antibody and antigen). The comparison includes DDA, DIA searches with gel fractionation, GPF fractionation, and the mixed complete spectral library. **B)** Venn diagram depicting the overlap in the protein groups identified in the experiment across the input, unbound, and elution samples.

Concerning the unbound fraction, as depicted in the illustration, a Venn diagram shows the overlap between the three fractions in terms of the protein groups identified with the mixed library search. There is a substantial overlap between the samples, with very few protein groups identified exclusively in the input and elution samples. However, the most notable finding was the presence of HCPs in the unbound fraction, which aligns with the objective of this study.

More specifically, we observed HCPs exclusively in the unbound fraction. While the HCPs found in the fraction shared with the input and excluded from the elution could be considered interesting, it is important to note that due to the strong presence of the anti-HCP antibody in the elution fraction, it would be challenging to conclude that these HCPs are actually immuno-reactive. The abundance of the antibody in this fraction could hinder the identification of low-abundance HCPs. In contrast, the unbound fraction, where the anti-HCP antibody is less prevalent, provided a clearer view of these low-abundance HCPs, which were undetected by the ELISA method. Therefore, we focused on the HCPs exclusively found in the unbound fraction, as these are considered to be truly undetected by ELISA, and their identities were revealed through our immuno-capture MS-based method.

The Venn diagram, as mentioned previously, highlights the overlaps among the samples, revealing key trends. Specifically, 63% of the protein groups identified in the Input fraction (787 protein groups out of 1235) were shared with both the Unbound and Elution fractions.

Indeed, since the Input fraction represents the anti-HCP antibody/antigen mixture prior to the immune-capture process, we would expect a complete overlap between the Unbound, Elution, and Input fractions if all proteins were detected by mass spectrometry. However, the strong presence of the anti-HCP antibody in both the Input and Elution fractions hinders the detection of the entire HCP

population across the different fractions. As a result, the overlap between the Unbound and Input fractions is much higher compared to the overlap between the Input and Elution fractions, and unbound and elution present unique protein groups.

The Unbound fraction yielded particularly interesting insights, containing 185 unique protein groups, representing 13% of the total identified in this fraction. These unique HCPs, absent from both the Input and Elution fractions, are likely of high relevance as they most likely represent proteins that evade capture by the anti-HCP antibody. Notably, the Unbound fraction contains much lower levels of the anti-HCP antibody, as it consists of HCPs that did not bind to the antibody mixture or attach to the Protein A column. This minimal antibody presence reduces interference and enhances the ability to detect HCPs within the sample. Interestingly, analysis using the mixed library revealed that the Unbound fraction contained more identified HCPs than the Input fraction, emphasizing how reduced antibody interference enables a broader exploration of the HCP population. This finding also highlights the challenges inherent in working with HCP samples, particularly under such complex conditions, where the anti-HCP antibody presence can significantly impact detection and identification.

Conversely, the Elution fraction showed fewer unique protein groups (56 protein groups, or 6% of the total identified in this fraction). This also emphasizes the limitations of the Elution fraction when detecting proteins of lower abundance, as the presence of the anti-HCP antibody in this fraction is significantly higher. Consequently, fewer HCPs are identified in the Elution fraction compared to the Unbound fraction or Input fraction, where the proportion of the anti-HCP antibody presence is smaller.

In Experiment 2, similar to Experiment 1, a 1:1 ratio of 1 mg of anti-HCP antibody to 1 mg of antigen was used. The results from this experiment, shown in the histogram in Figure 46, highlight again the advantages of using the mixed exhaustive library for DIA analysis against DDA. Across all fractions, the Unbound fraction consistently outperformed both the Input and Elution fractions in terms of protein identifications. Specifically, the Unbound fraction identified 1427 HCPs, significantly more than the 1085 identified in the Input fraction and the 548 in the Elution fraction. This trend mirrors the findings from Experiment 1, where reduced interference from the anti-HCP antibody in the Unbound fraction allowed for a broader detection of HCPs.

However, a slight deviation was observed in Experiment 2. The Elution fraction, which holds a higher ratio of antibody to antigen compared to Experiment 1, showed fewer HCPs (548). The stronger presence of the anti-HCP antibody in this fraction likely limited the detection of low-abundance HCPs, a limitation more pronounced here than in Experiment 1.

Moving to the Venn diagram, the overlap between the fractions supports these findings, with some notable differences compared to Experiment 1. Of the identified protein groups, 1039 were shared between the Input and Unbound fractions, and 490 were identified across all three fractions (Input, Unbound, and Elution). However, the overlap in Experiment 2 has reduced compared to Experiment 1, reflecting a more distinct separation between the fractions. This is further highlighted by the presence of 379 HCPs exclusively found in the Unbound fraction, a significant increase from Experiment 1.

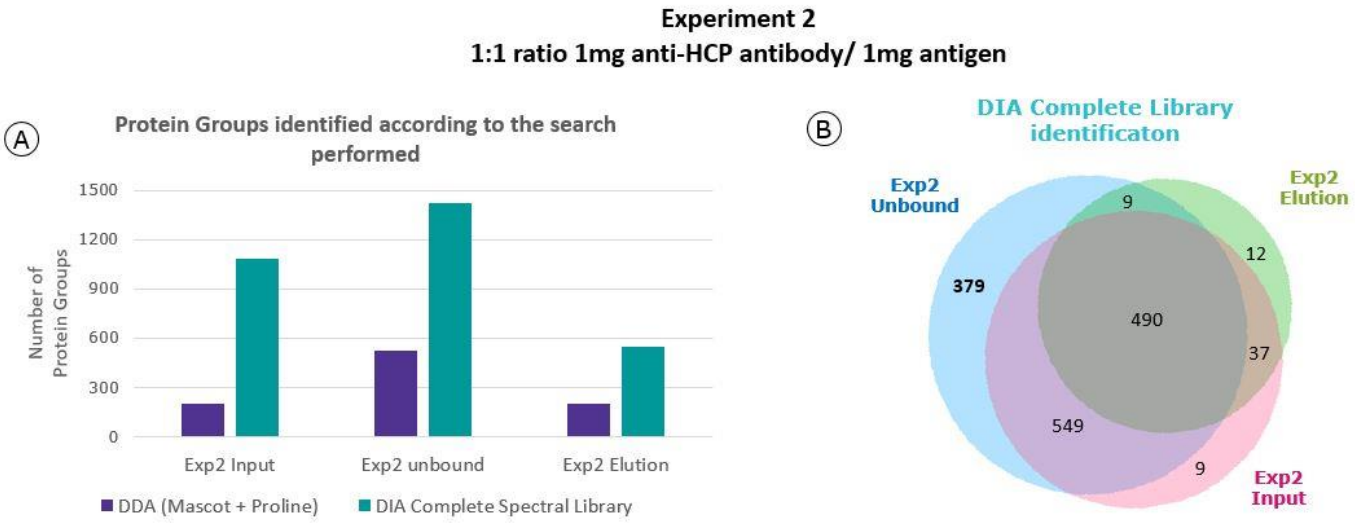


Figure 46: **A)** Histogram comparing the number of protein groups identified using different search strategies for the second ratio optimization of the column (1:1 ratio of anti-HCP antibody and antigen). The comparison includes DDA and the DIA mixed complete spectral library search. **B)** Venn diagram depicting the overlap in the protein groups identified in the experiment across the input, unbound, and elution samples.

For Experiment 3, conducted at a 2:1 ratio using 500 µg of anti-HCP antibody and 250 µg of antigen, similar trends to those observed in Experiments 1 and 2 were repeated. The Unbound fraction once again showed a significantly higher identification of protein groups, with 1214 identified compared to 750 in the Input fraction and 500 in the Elution fraction. This pattern likely reflects the same behavior observed in Experiment 2, as the higher ratio of antibody present in both the Input and Elution fractions results in a higher antibody signal saturation, which reduces the identification of HCPs present in the fraction.

As observed in Figure 47, the overlap between the Input, Unbound, and Elution fractions in Experiment 3 was reduced to 405, a smaller decrease than the drop observed from Experiment 1 to Experiment 2. This indicates that while the higher anti-HCP antibody ratio in the Input and Elution fractions continues to limit some protein group identification, the trend of a less significant overlap persists.

In terms of solely identified protein groups, the Unbound fraction showed a notable increase, with 475 unique HCPs identified, the highest among all the ratio experiments. This outcome reflects the experimental conditions, where the concentration of reagents in the Input and Elution fractions likely compromised their HCP identifications, resulting in a smaller overlap and thus a greater number of uniquely detected proteins in the Unbound fraction.

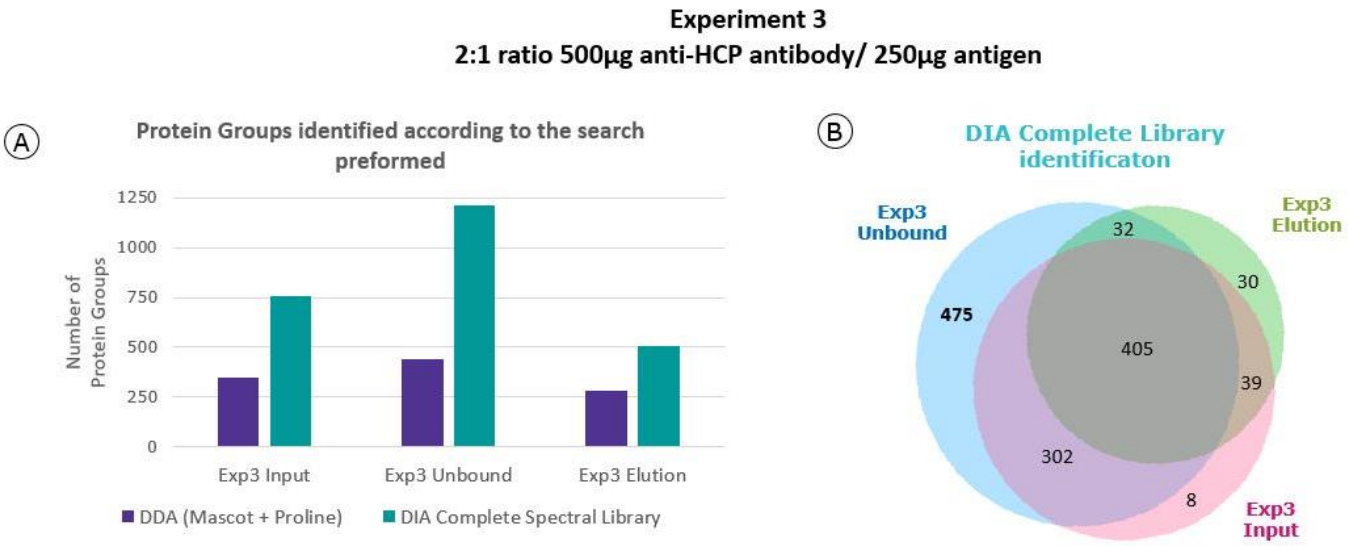


Figure 47: **A)** Histogram comparing the number of protein groups identified using different search strategies for the third ratio optimization of the column (2:1 ratio of anti-HCP antibody and antigen). The comparison includes DDA and the DIA mixed complete spectral library search. **B)** Venn diagram depicting the overlap in the protein groups identified in the experiment across the input, unbound, and elution samples.

B. Volume and temperature optimizations findings

Following the ratio optimization experiments detailed in a previous section of this manuscript, we proceeded to explore volume and temperature optimization. These parameters were assessed to better align with the objectives of the study by evaluating faster and more efficient conditions.

While the initial ratio experiments were conducted at 4°C with a 24-hour incubation period, we aimed to investigate the effects of increased temperature and shorter incubation times on the column's performance. Additionally, we examined whether a larger incubation volume could enhance interactions between the anti-HCP antibody and antigen, potentially improving steric accessibility.

For this series of experiments, we retained the 1:1 ratio of anti-HCP antibody to antigen, as determined to be the most appropriate based on the profiles analyzed through 2D DIGE and Mass Spectrometry, whose results reaffirmed the challenges associated with the analysis, as the strong presence of the anti-HCP antibody continued to interfere with HCP detection. Despite these challenges, the 1:1 ratio was deemed optimal, as neither excess antibody nor excess antigen provided significant improvement in identifications or experimental outcomes. Thus, the 1:1 ratio offered a balanced and practical approach for the subsequent optimizations.

Experiment 4, the first in the series of volume and temperature optimization experiments, aimed to evaluate the effect of a shorter incubation period at room temperature compared to the previous conditions. This experiment utilized a 300 µL immuno-complex incubation for 2 hours at room temperature, transitioning from the 24-hour incubation at 4°C used before.

As shown in Figure 48, the histogram indicates that both the Input and Unbound fractions achieved significantly higher protein group identifications, with 1803 and 1881 protein groups, respectively. These results mark a substantial improvement compared to Experiment 2, which identified 1085 protein groups in the Input fraction and 1427 in the Unbound fraction with a longer incubation. In contrast, the Elution fraction's protein group identifications remained relatively stable, with 543 protein groups in Experiment 4 versus 548 in Experiment 2.

The Venn diagram further supports these findings, showing a larger overlap between the fractions in Experiment 4. The Unbound fraction saw a reduction in unique protein groups compared to Experiment 2. This improvement, highlights the potential of these optimized conditions to balance rapidity with efficacy of immune-capture.

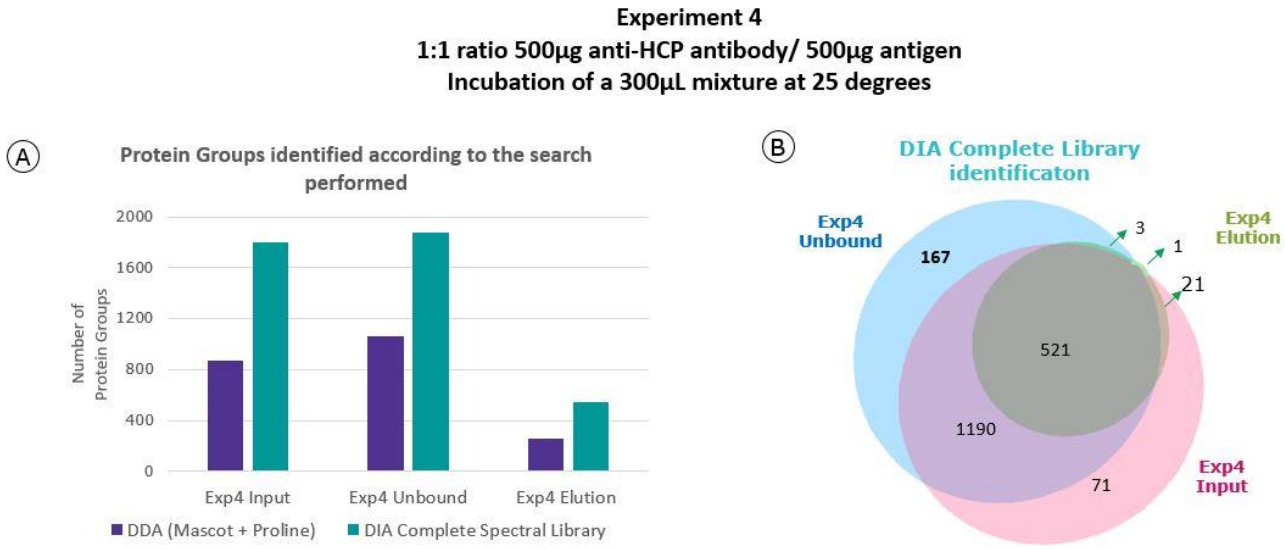


Figure 48: **A)** Histogram comparing the number of protein groups identified using different search strategies for the fourth ratio optimization of the column (1:1 ratio of anti-HCP antibody and antigen, with a 300µL mixture and 25 degree incubation). The comparison includes DDA and the DIA mixed complete spectral library search. **B)** Venn diagram depicting the overlap in the protein groups identified in the experiment across the input, unbound, and elution samples.

Experiment 5, conducted with a 2 mL sample volume, aimed to evaluate whether increased volume would enhance the affinity interactions between the anti-HCP antibody and antigen. As in previous experiments, the DIA method with the mixed library continued to provide superior protein group identification compared to DDA, consistently demonstrating still its advantage in capturing a broader range of HCPs.

The results did not show a substantial improvement compared to Experiment 4, which was carried out with a 300 µL sample volume, as further depicted in Figure 49. While the Input and Unbound fractions exhibited slightly higher protein group identifications, the overall trends observed in Experiment 4—such as the large overlap between the fractions and the reduced number of uniquely identified HCPs in the Unbound fraction—remained consistent.

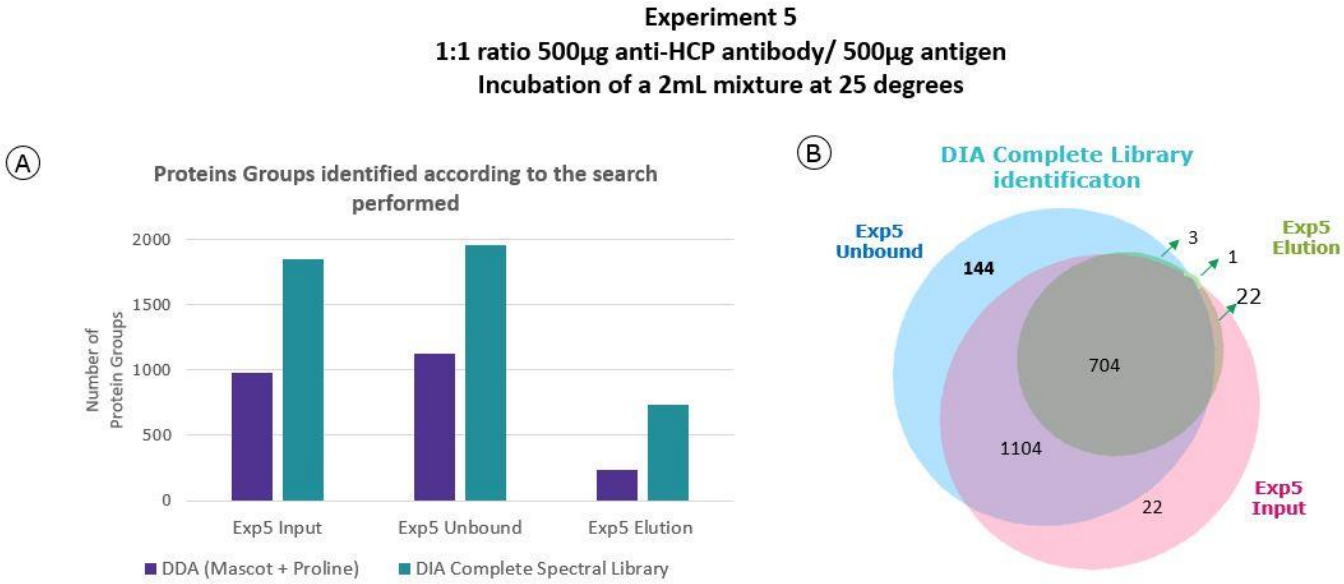


Figure 49: **A)** Histogram comparing the number of protein groups identified using different search strategies for the fifth ratio optimization of the column (1:1 ratio of anti-HCP antibody and antigen, with a 300µL mixture and 37 degree incubation). The comparison includes DDA and the DIA mixed complete spectral library search. **B)** Venn diagram depicting the overlap in the protein groups identified in the experiment across the input, unbound, and elution samples.

In Experiment 6, we tested the impact of increasing the incubation temperature to 37°C while keeping the immuno-complex volume at 300 µL. This change aimed to enhance the affinity between the anti-HCP antibody and antigen. This change aimed to enhance the affinity between the anti-HCP antibody and antigen, addressing a limitation observed in the earlier stages of the immuno-capture optimization process.

The results showed 1691 protein groups identified in the Input fraction, 1437 in the Unbound fraction, and 350 in the Elution fraction as observed in Figure 50. There was a slight improvement compared to the previous experiment in the number of exclusively identified protein groups in the Unbound fraction, which was the lowest so far in the series with 55 protein groups. This was also the first time the Elution fraction did not show any unique HCPs.

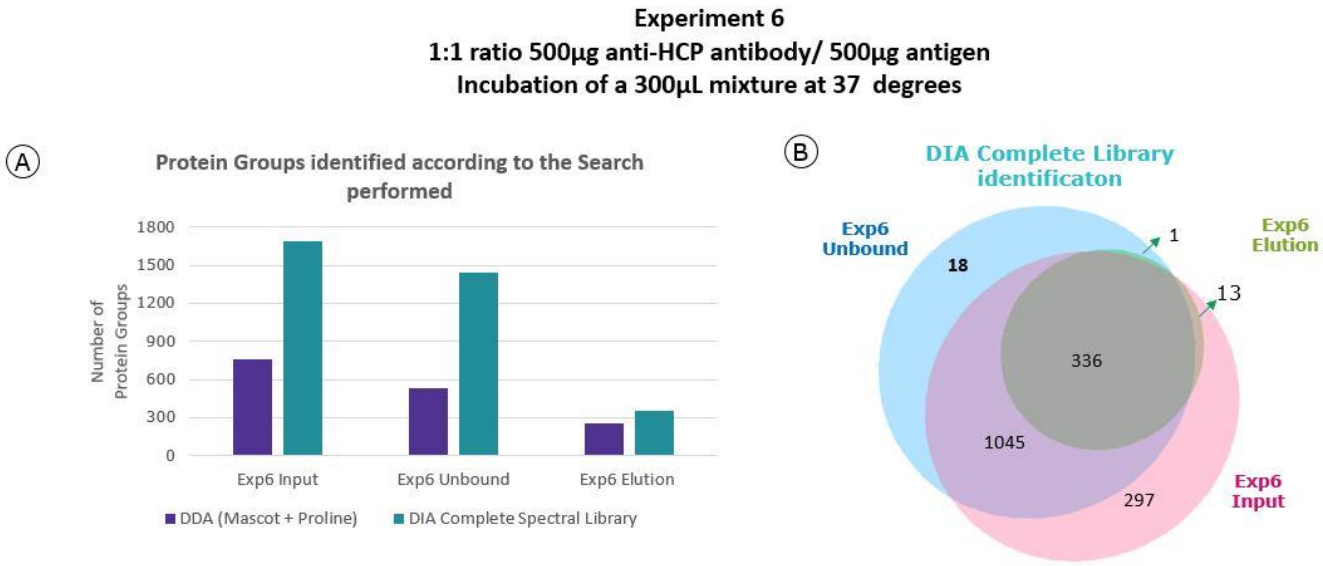


Figure 50: **A)** Histogram comparing the number of protein groups identified using different search strategies for the sixth ratio optimization of the column (1:1 ratio of anti-HCP antibody and antigen, with a 300µL mixture and 37 degree incubation). The comparison includes DDA and the DIA mixed complete spectral library search. **B)** Venn diagram depicting the overlap in the protein groups identified in the experiment across the input, unbound, and elution samples.

Experiment 7 aimed to assess the impact of increasing the volume of the immuno-complex mixture to 2 mL while maintaining the incubation temperature at 37°C. However, despite the changes in the experimental setup, the results did not show any significant improvements over previous experiments as further observed in Figure 51.

As observed in Experiment 6, the tendencies from Experiment 7 were largely similar. The Elution fraction still did not yield any unique protein groups, and both the Input and Unbound fractions continued to show a higher number of protein groups compared to those identified in the ratio optimization experiments.

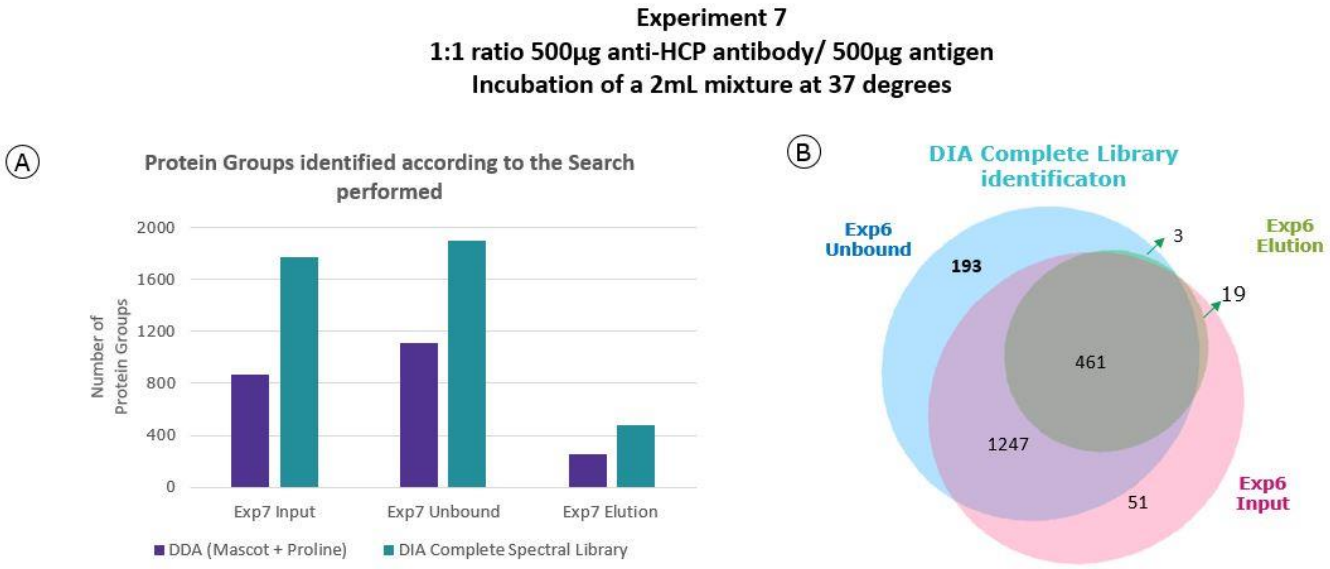


Figure 51: **A)** Histogram comparing the number of protein groups identified using different search strategies for the seventh ratio optimization of the column (1:1 ratio of anti-HCP antibody and antigen, with a 2mL mixture and 37 degree incubation). The comparison includes DDA and the DIA mixed complete spectral library search. **B)** Venn diagram depicting the overlap in the protein groups identified in the experiment across the input, unbound, and elution samples.

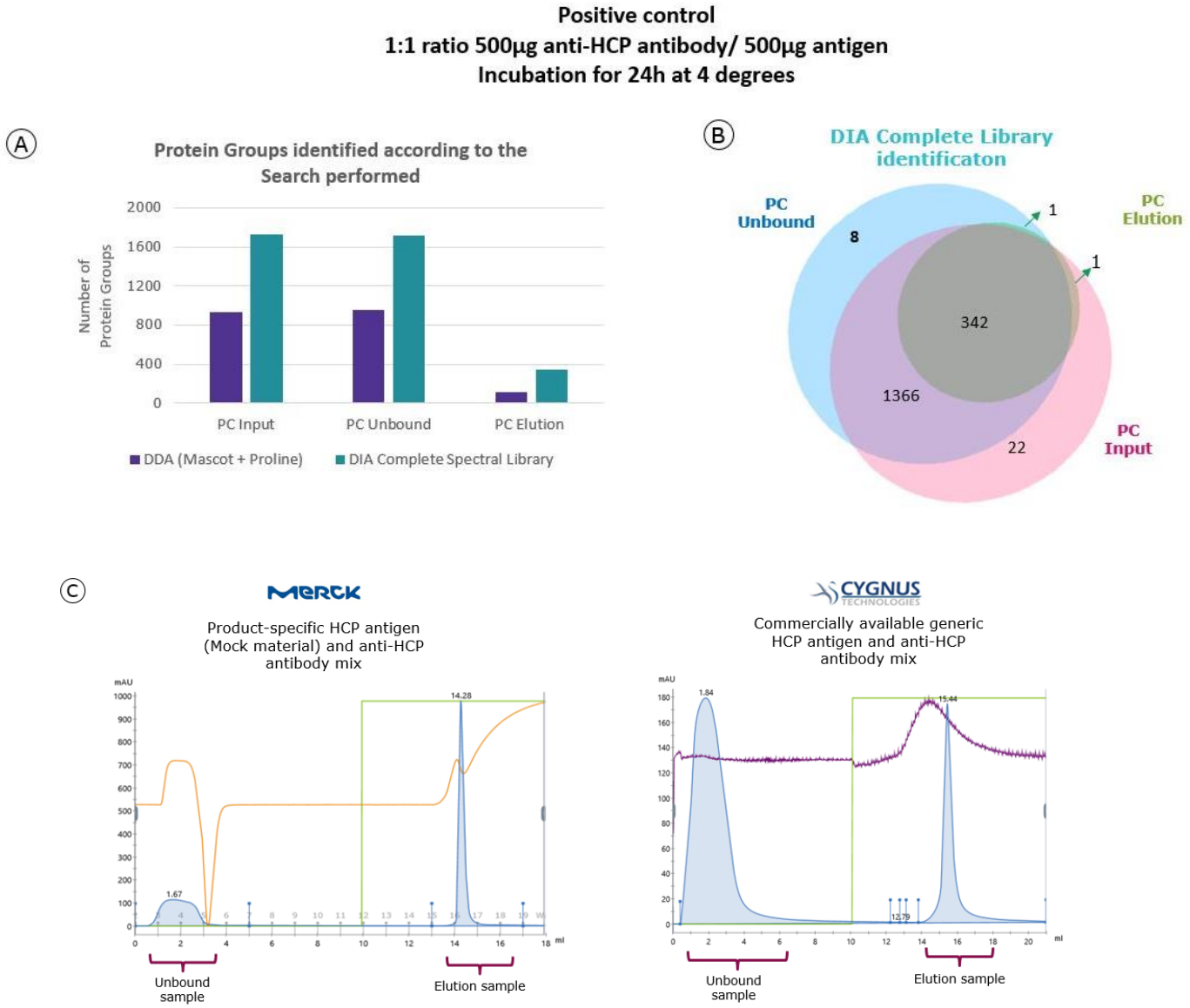
C. Positive control findings

For the positive control, the experiment was carried out using the 1:1 ratio optimization with the longest incubation time, 24 hours at 4°C, and employing a generic commercial anti-HCP antibody and antigen pair. The results indicate that the Input and Unbound fractions present similar numbers of protein groups, while the Elution fraction shows no uniquely identified protein groups. Notably, the overlap of protein groups across all three fractions appears high. However, this apparent similarity to the successful experiments is misleading due to the high proportion of anti-HCP antibody present in the unbound fraction.

To better illustrate these observations, additional analyses were included in Figure 52, such as the immuno-capture chromatogram and 2D DIGE of the fractions obtained. The chromatogram compares the positive control to a successful immuno-capture using the product-specific antibody-antigen pair. In the positive control, a significant signal was observed in the unbound fraction, reflecting the overwhelming presence of the anti-HCP antibody. This excessive abundance not only complicates the interpretation of the MS data obscuring the actual HCP population.

The 2D DIGE analysis reinforces these findings, with both the Unbound and Elution fractions dominated by the light and heavy chains of the anti-HCP antibody. These components saturate the signal, creating the false impression that fewer HCPs are present in the Unbound fraction. While the overlap between fractions appears high, and the number of uniquely identified HCPs in the Unbound fraction is low, this is a direct consequence of the failed immuno-capture process with the generic antibody.

These results underscore the difficulty of analyzing HCP samples and highlight the value of complementary analyses, such as 2D DIGE and chromatograms, in predicting and explaining the MS outcomes. The tailored product-specific antibody-antigen pair used in earlier experiments ensures a much more accurate assessment of the HCP population, emphasizing the limitations of generic antibody approaches in this context.



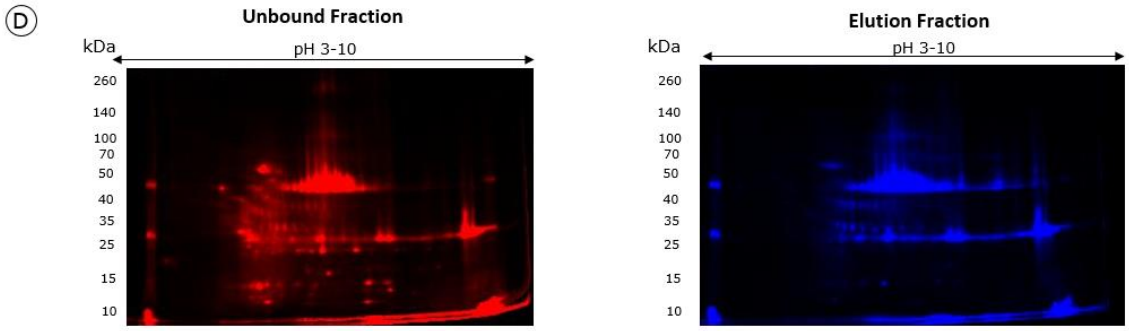


Figure 52: A) Histogram comparing the number of protein groups identified using different search strategies for the positive control of the column (1:1 ratio of anti-HCP antibody and antigen, with 24-hour incubation at 4°C). The comparison includes DDA and the DIA mixed complete spectral library search. **B)** Venn diagram depicting the overlap in the protein groups identified across the Input, Unbound, and Elution fractions. **C)** Immuno-capture chromatogram comparing a successful immuno-capture experiment using the product-specific antibody-antigen pair to the positive control performed with the generic commercial anti-HCP antibody. The chromatogram highlights the significant signal in the unbound fraction for the positive control, reflecting the overwhelming presence of the anti-HCP antibody. **D)** 2D DIGE analysis of the Unbound and Elution fractions from the positive control, illustrating the dominant presence of the light and heavy chains of the anti-HCP antibody, saturating the signal in both fractions.

Chapter 3
Transition to Phytips Protein A

1. Advantages for Switching to Phytips Protein A implementation

Following the comprehensive study and optimization of the column support as our immuno-capture platform of choice, alongside its corresponding MS analysis process, we sought to evaluate an alternative approach: Phytips Protein A. These pipette tips offer an appealing option due to their streamlined workflow and reduced preparation requirements. Unlike the column-based system, Phytips Protein A operates without the need for a chromatography setup, significantly easing experimental handling while maintaining similar immuno-capture principles. This combination of convenience and potential for high-throughput applications makes them a compelling alternative for our further exploration.

Phytips Protein A by Biotage were primarily designed for micro-volume antibody purification, featuring a unique dual-flow chromatography (DFC) technology. The DFC process uses a simple workflow of three steps: capture, wash, and elution, where the liquid is passed back and forth over the resin bed during each step, ensuring efficient antibody binding and purification. These tips are packed with ProPlus affinity resin, the same MabSelect™ resin used also in the column of choice for our optimization studies, reinforcing their compatibility with our research. Although their intended purpose is antibody purification, much like the column-based system, we adapted these pipette tips protocol for immuno-capture applications. This repurposing also demonstrates the versatility of Phytips Protein A in supporting alternative experimental objectives such as our own.

2. Application of Column Findings to Phytips Protein A

A. Experimental workflow and set-up

The aim of these experiments was to build upon the findings from our column-based immuno-capture studies. We decided to conduct triplicates for each experiment, which allowed for more robust and reproducible results. The versatility and faster workflow of the Phytips Protein A support enabled us to test two different immuno-capture strategies.

The first method, the direct approach, illustrated in Figure 53, involves directly attaching the antibody to the support. Following this, the antigen is added, and the immuno-complex is subsequently released from the Protein A resin. This straightforward approach takes advantage of the resin's affinity for antibodies, allowing us to capture the antigen-antibody complex with ease.

Figure 53: Workflow of the Direct Approach for the PhyTips Protein A. The illustration depicts the stepwise process starting with the capture of the anti-HCP antibody by the Protein A resin. This is followed by the binding of immuno-reactive HCPs to the immobilized antibody. Non-immunoreactive HCPs are removed in the unbound fraction and subsequent washes. Finally, the immuno-complex formed by the anti-HCP antibody and bound HCPs is released during the elution step.

In contrast, the indirect approach, as observed in Figure 54, mirrors the methodology used in the column experiments. Here, the antibody and antigen are first incubated together to form the immuno-complex. Afterward, the complex is added to the Protein A support, where it binds and is then released. The goal of using this approach was to compare it with the direct method and determine which would yield the best results for our immuno-capture application with Phytips Protein A. To maintain consistency with the column experiments and allow for meaningful comparison, we used the same 1:1 ratio of anti-HCP antibody to antigen, ensuring that the conditions were aligned across both types of support.

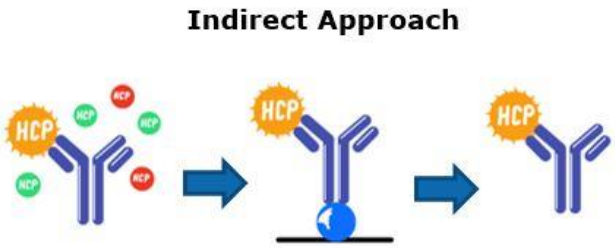


Figure 54: Workflow of the Indirect Approach for the PhyTips Protein A. The figure illustrates the formation of the immuno-complex following the incubation of the anti-HCP antibody with HCPs. This mixture is then passed through the Protein A resin, where the immuno-complex, comprising the anti-HCP antibody conjugated to immuno-reactive HCPs, is captured by the Protein A. Non-immunoreactive HCPs, which do not bind, are removed in the unbound fraction and subsequent washes. Finally, the immuno-complex is released during the elution step.

For the indirect approach, we adopted the conditions that had previously given the best results in our column experiments, ensuring consistency across our testing and providing a point of comparison. This approach was designed to verify which immuno-capture method, direct or indirect, would be most efficient when using the Phytips Protein A as the support.

To assess the success and efficiency of each approach, a BCA assay was performed on the collected samples. This assay allowed us to measure the protein content and determine which method provided

the most effective immuno-capture, offering valuable insights into the best workflow for Phytips Protein A.

i) Direct Approach

Given the specifications of the Phytips and their reported 50% recovery rate, we chose to focus first on the direct approach for immuno-capture. This decision was based on the practical advantages of the direct method, which involves binding the antibody directly to the Protein A resin before introducing the antigen. This simplified workflow leverages the affinity of Protein A for antibodies while avoiding the potential inefficiencies of pre-incubating an immuno-complex, as was performed in previous indirect approaches.

As mentioned previously, for the first time in the immuno-capture optimization studies, all experiments were performed in triplicate to ensure robust and reproducible results. To maintain consistency with previous column-based studies, a 1:1 ratio of anti-HCP antibody to antigen was employed, with 500 µg of each reagent used per replicate. These quantities were prepared by measuring the appropriate volumes of the stock solutions and then diluting them with equilibration buffer to a total volume of 1 mL. This approach ensured that experimental conditions were directly comparable across the two systems while accommodating the Phytips volume capacity of 1 mL.

The Phytips themselves feature an 80µL resin bed packed with Protein A, which is critical for antibody capture. This resin bed, while efficient, is associated with a recovery rate of approximately 50%, as reported by the supplier. Prior to their use, the tips were carefully handled to remove any glycerol present from storage, ensuring that the resin remained hydrated and free from contaminants. The aspiration and dispensing of liquids were performed manually and with great care, despite the supplier's recommendation for electronic pipettes. This manual handling provided precise control over the liquid movement, ensuring that the resin bed remained intact and preventing the introduction of air bubbles into the system.

The workflow began with the equilibration of the Phytips resin using equilibration buffer, passing 1 mL of the buffer through the resin bed in multiple cycles to prepare it for binding. Next, the anti-HCP antibody solution was introduced to the resin, allowing the antibodies to bind to the Protein A. This step was followed by a washing phase to remove any unbound material, which was collected for later analysis. Subsequently, the antigen solution was passed through the Phytips, allowing it to interact with the bound antibodies to form the desired immuno-complex.

After antigen capture, a second wash step was performed to remove unbound antigen and any remaining contaminants. Finally, the immuno-complex was eluted from the Protein A resin using an elution buffer, and a neutralization buffer was immediately added to stabilize the recovered sample. All eluates and unbound fractions from each step were carefully collected and stored for subsequent analysis to evaluate the efficiency of the immuno-capture.

ii) Indirect approach

Since initial results from the BCA assay revealed that while the anti-HCP antibody effectively bound to the Protein A resin, the antigen unbound fraction, representing the HCPs not captured by the antibody, remained abundant and retained nearly the full integrity of the original sample. To address this, the decision was made to explore the indirect approach. This method involves pre-forming the immune complex by incubating the antigen and antibody together prior to introducing the mixture to the Phytips Protein A resin.

The immuno-complex mixture was prepared by incubating antigen and anti-HCP antibody at a 1:1 ratio, corresponding to 500 µg of each, in a ThermoShaker at 37°C for two hours with gentle agitation set to 300 rpm. Following incubation, the reaction mixture, approximately 294 µL in volume, was diluted to a final volume of 1 mL with equilibration buffer. Careful pipetting techniques were employed to prevent introducing air or damaging the resin bed, thereby preserving the efficiency of the capture process.

Once the immune complex mixture was ready, the Phytips workflow began with equilibration of the resin bed using equilibration buffer, followed by the introduction of the pre-formed immune-complex. The HCPs that did not bind to the Protein A resin were collected as the unbound fraction, providing valuable insight into the immuno-capture efficiency. Subsequent wash steps ensured that any remaining unbound material was cleared from the resin bead. Finally, the immune-complexes that had successfully bound to the Protein A resin were eluted and immediately neutralized to prevent protein denaturation due to low pH.

iii) Sample preparation for MS analysis

The in-gel digestion approach was employed for these samples, similar to the method used for the column samples, as it effectively removes contaminants such as detergents, salts, and other interfering substances. A total of 10 µg of protein from the immuno-capture experiments was loaded into a low-percentage polyacrylamide gel (5%). The gel's concentrating zone compressed proteins into a narrow band, creating a unified starting point for subsequent processing and ensuring uniform positioning of the proteins within the gel before separation.

To prepare the proteins for digestion, the gel bands were subjected to reduction with 10 mM DTT for 30 minutes at 60°C to break disulfide bonds, followed by alkylation with 55 mM iodoacetamide (IAA) for 30 minutes in the dark to prevent the reformation of these bonds. Trypsin (Promega) was added at a 1:25 enzyme-to-protein ratio, and the samples were incubated overnight at 37°C for approximately 15 hours to ensure complete digestion. The reaction was subsequently quenched, and peptides were extracted from the gel bands using a sequential process with 0.1% formic acid (FA) and 80% acetonitrile (ACN) under gentle agitation at 110 rpm for 1 hour.

After extraction, peptides were vacuum-dried and reconstituted in a buffer containing 2% ACN and 0.1% FA to achieve a final concentration of 100 ng/µL, in order to inject 300ng . For samples designated

for DIA acquisition mode, internal retention time (iRT) standards (Biognosys) were added to the peptide solution at 10% of the final injection volume, allowing for accurate retention time alignment.

For mass spectrometry analysis, samples obtained from the column optimization experiments were injected into a Dionex Ultimate 3000 RSLC nano system (Thermo Fisher Scientific). The analysis was performed using the Orbitrap Eclipse™ Tribrid™ mass spectrometer (Thermo Fisher Scientific), ensuring high sensitivity and resolution. The UPLC system featured an Acclaim PepMap 100 C18 pre-column (20 mm × 0.1 mm, 5 µm, 100 Å) from Thermo Fisher Scientific, coupled with an Aurora Series C18 UHPLC analytical column (250 mm × 75 µm, 1.6 µm, 120 Å) from IonOpticks.

The separation process utilized a binary mobile phase system. Mobile phase A consisted of water with 0.1% formic acid (FA), while mobile phase B was a mixture of acetonitrile (ACN), water, and 0.1% FA (79.9/20/0.1). Peptides were loaded at a steady flow rate of 300 nL/min and resolved over an 85-minute gradient. During the gradient, the proportion of mobile phase B gradually increased, reaching 43.8% by the 65-minute mark. At minute 66, the gradient reached 98% for column washing and was maintained until minute 71. Re-equilibration of the system was achieved by reducing the proportion of phase B to 2.5%, which was maintained from minute 72 to 85 to ensure system stability for subsequent analyses.

iv) Quantification of high-risk HCPs

To achieve quantification, a set of 54 synthetic peptides covering a broad concentration range was used to create an internal calibration curve (ReadyBeads by Anaquant, see Part III for detailed description of ReadyBeads implementation).

Peptide intensity was summed from precursor ion intensities, and protein intensity was calculated using the top three peptides (Top3 strategy)[231]. HCP concentrations were quantified in ng HCP/mg injected, calculated under the ppm format using the ANAQUANT ReadyBeads calibration curve, allowing accurate determination of HCP contamination levels in the samples.

v) MS Acquisition Methods

For data-independent acquisition (DIA) MS data were collected over a mass range of 375–1200 m/z in full-scan mode, operating at a resolution of 120,000. An AGC target of 8×10^5 and a maximum injection time of 200 ms were utilized to ensure high sensitivity and precision. MS/MS scans were performed at a resolution of 15,000 with a maximum injection time of 22 ms. Peptide fragmentation was achieved using higher-energy collision-induced dissociation (HCD) at a normalized collision energy of 30%. The analysis employed isolation windows of 18.4 m/z with a 0.5 m/z overlap, improving the reliability of peptide identification.

Protein identification and quantification were performed using Spectronaut™ software (v.18), with raw data and the CHO database (Cricetulus griseus entries from UniProtKB/TrEMBL) as input. The analysis was based on trypsin/P digestion, allowing one missed cleavage per peptide. Two variable modifications, methionine oxidation and protein N-terminal acetylation, were included. Precise identification was ensured by applying a Q-value cutoff of 0.01 for both precursors and proteins.

For data-dependent acquisition (DDA), full-scan MS spectra were recorded over a broader range of 375–1500 m/z, with a resolution of 120,000 (at 200 m/z). The AGC target was configured at 8×10^5, and the maximum injection time was set to 60 milliseconds. During each MS scan, the 20 most abundant precursor ions, with charge states between 2 and 6 and intensities exceeding 2×10^4 ions per second, were automatically selected for fragmentation. Fragmentation was achieved using higher-energy collisional dissociation (HCD) at a normalized collision energy of 30%. The resulting MS/MS spectra were acquired in the ion trap in Rapid mode, covering the same mass range of 375–1500 m/z. For MS/MS scans, an AGC target of 5×10^4 and a maximum injection time of 54 milliseconds were employed.

Spectra were then searched using Mascot, with a mass tolerance of 5 ppm for MS scans and 0.05 Da for MS/MS scans. One missed trypsin cleavage was allowed. Carbamidomethylation of cysteine residues was considered a fixed modification, while oxidation of methionine residues and acetylation of protein N-termini were treated as variable modifications. The identification results were imported into Proline software for validation, applying a 1% false discovery rate at the peptide level using adjusted e-Values and at the protein level using Mascot Modified Mudpit scores.

B. Triplicate Analyses for Ensuring Repeatability

The results of the direct approach, as illustrated in Figure 55, reveal important insights into the efficiency of this strategy. The BCA assay demonstrated that the attachment of the anti-HCP antibody to the Protein A resin was highly effective, with only 20 to 50 µg of antibody detected in the unbound fraction. This indicates that nearly the entire population of anti-HCP antibodies, across all triplicates, successfully bound to the support. This outcome underscores the robust affinity of Protein A for the Fc region of antibodies, ensuring effective capture during this step.

However, the efficiency of antigen capture proved to be significantly lower. Examination of the unbound antigen fraction, which corresponds to HCPs that did not bind to the anti-HCP antibody, immobilized on the resin, revealed a substantial quantity of protein—approximately 500 µg, which matches the starting input. This finding indicates that the majority of the antigen remained unbound, suggesting limited interaction between the HCPs and the immobilized antibody under the chosen conditions. The results raise concerns about the accessibility or affinity of the anti-HCP antibody for its target when attached to the Protein A support. The challenges persisted during the elution phase, where only about 20 µg of the immuno-complex was recovered.

This low yield also highlights a potential limitation of the direct approach in this context. One plausible explanation lies in the structural consequences of antibody immobilization on Protein A. It is well-documented that the binding of antibodies to Protein A can induce subtle conformational changes, which may propagate to the antigen-binding Fab regions, diminishing the antibody's affinity for its target.[232,233] This structural alteration, combined with the inherent complexity of the HCP mixture, likely contributed to the reduced efficiency of antigen capture and subsequent elution.

Direct Approach

Samples obtained from assay	Protein quantity (µg)
Unbound Antibody 1	28,71
Unbound Antibody 2	51,07
Unbound Antibody 3	46,13
Unbound Antigen 1	499,05
Unbound Antigen 2	486,81
Unbound Antigen 3	504,70
Elution 1	24,00
Elution 2	22,35
Elution 3	23,29

Indirect Approach

Samples obtained from assay	Protein quantity (µg)
Unbound 1	698,70
Unbound 2	698,19
Unbound 3	661,78
Elution 1	179,65
Elution 2	164,47
Elution 3	21,10

Figure 55: **Comparative BCA assay results for the Direct and Indirect Approaches**. On the left, the Direct Approach shows the Unbound of Antibody, representing the amount of anti-HCP antibody that did not bind to the Protein A support; the Unbound of Antigen, reflecting the HCPs from the mock sample that did not bind to the anti-HCP antibody immobilized on the resin; and the Elution, representing the immuno-complex of anti-HCP antibody bound to immunoreactive HCPs released during the elution step. On the right, the Indirect Approach includes the Unbound Fraction, corresponding to non-immunoreactive HCPs that did not bind to the anti-HCP antibody immobilized on the resin, and the Elution, which reflects the immuno-complex of anti-HCP antibody and immunoreactive HCPs captured and released during the elution step.

After the underwhelming results observed with the direct approach, attention was turned to the outcomes of the indirect approach. This alternative method demonstrated a markedly improved performance. As shown in Figure 55, the protein concentration in the unbound fraction averaged around 700 µg. Considering the initial input of 1000 µg (500 µg each of antigen and anti-HCP antibody), this result indicates that approximately 30% of the mixture successfully attached to the Protein A resin in the Phytips support. This was a promising outcome, providing evidence of the improved efficiency of the indirect approach.

Further, the elution fraction yielded significantly higher amounts of immune-complex compared to the direct approach, with approximately 170 µg recovered. Notably, this consistency was observed across most triplicates, with one exception, reflecting the generally good reproducibility of the method. These findings highlight not only the limitations of the direct approach but also the clear advantages of the indirect method. By pre-forming the immune-complex before introducing it to the support, the interactions between the reagents were enhanced, resulting in greater efficiency in both binding and elution.

These results underscore the suitability of the indirect approach for our project's objectives, providing a more reliable pathway for achieving effective immuno-capture while addressing the challenges observed in earlier experiments.

To further evaluate the performance of the Phytips Protein A in immuno-capture experiments, we proceeded MS analysis using DirectDIA via Spectronaut, without employing MBR. This approach utilized the same DIA and DDA methods previously employed in the column experiments, ensuring consistency and enabling a direct comparison between the two supports. Given the stronger results obtained with the indirect approach, we opted to focus exclusively on samples from this method for the MS analysis, bypassing the less effective direct approach.

Figure 56 illustrates again the significant advantages offered by the DIA method over DDA in terms of protein group identification. This improvement is particularly evident in the unbound antigen and elution samples, where DIA demonstrates a marked increase in protein identification compared to DDA. These findings underscore the superior sensitivity and comprehensiveness of the DIA approach, especially for the analysis of complex samples, and reinforce the suitability of the indirect method for our immuno-capture workflow.

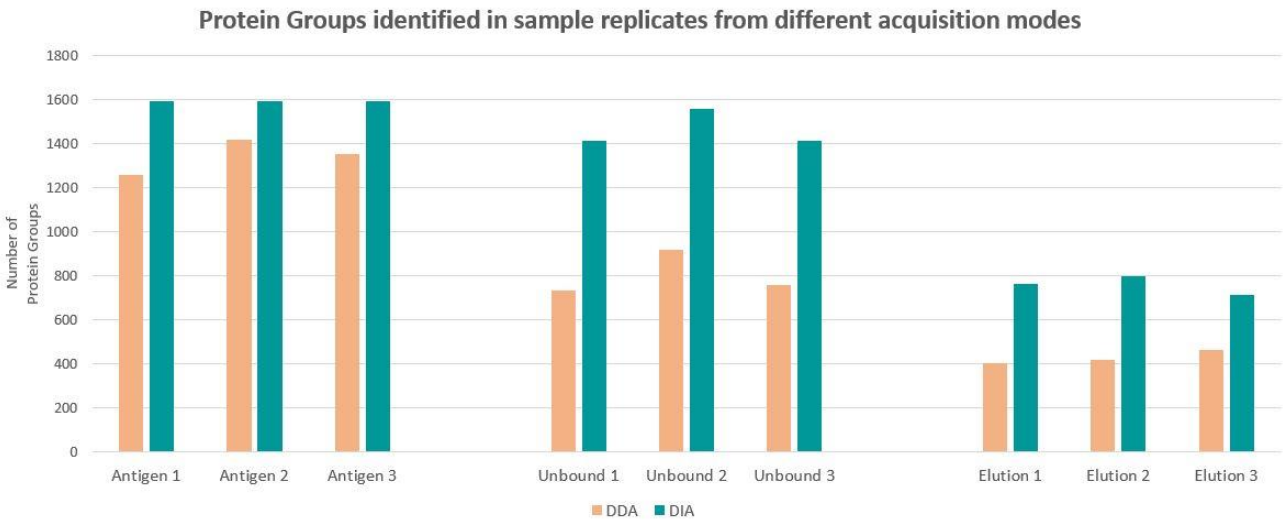


Figure 56: Histogram depicting the number of protein groups identified in the replicates of the antigen, unbound, and elution fractions following the indirect approach using Phytips Protein A. The histogram displays two bars for each sample, representing the protein group identification when employing either DDA or DIA for the analysis.

To gain deeper insights into the performance of the indirect approach and assess the reproducibility of the results, we extended our analysis to examine the overlap between the three key samples: the antigen (serving as the input), the unbound fraction, and the elution. This approach was chosen to explore the comprehensive coverage of the HCP population, as the antigen sample inherently contains the full spectrum of host cell proteins.

Figure 57 illustrates the overlap among the three replicates for these samples. Remarkably, we observed a near-complete overlap between the null cell line sample, the unbound fraction, and the

elution, demonstrating a high degree of consistency in the protein groups identified across the replicates. This overlap strongly supports the reproducibility and reliability of the indirect approach in capturing the target proteins.

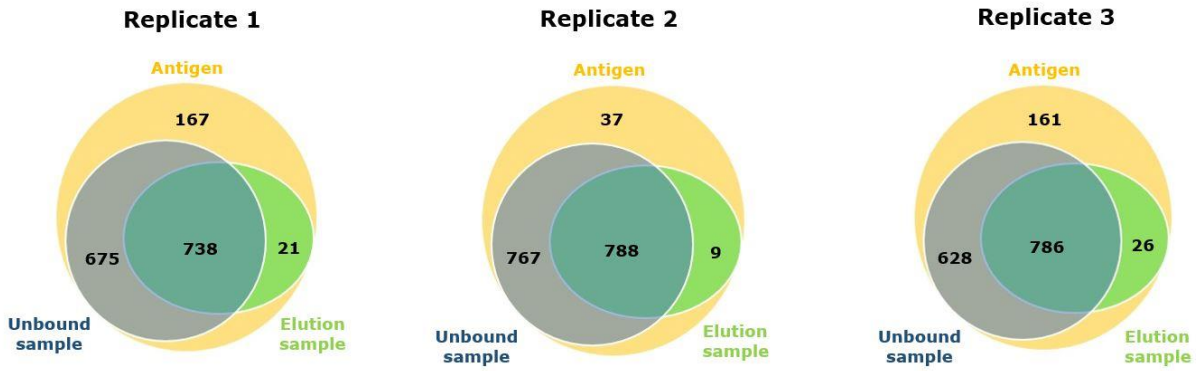


Figure 57: Venn diagrams showing the results of the triplicates for the indirect approach using Phytips Protein A. The diagrams highlight the complete overlap of protein groups identified across the antigen, unbound, and elution fractions, demonstrating consistency and reproducibility of the immuno-capture process under these conditions.

Under these conditions, the results are quite satisfactory. As mentioned previously, the overlap of all fractions is a first in this series of immuno-capture experiments, which is reassuring. Only a few protein groups in the antigen were not detected in the other 2 fractions. The number of protein groups found in the unbound fraction was consistent across replicates. With the overlap now complete, more HCPs that are considered non-immunoreactive were identified, given that when employing the antigen sample in the diagrams, there are no exclusively found HCPs in this fraction. However, due to the high presence of anti-HCP antibody in the elution, some of the HCPs in this fraction may not have been fully seen, likely leading to an underrepresentation of the complete HCP population in the elution.

For these types of samples, where HCPs tend to be a minority compared to the abundant anti-HCP antibody, the results are very satisfactory. These types of samples are notoriously difficult to handle, as previously discussed, and succeeding in achieving a successful immuno-capture procedure under these conditions is a notable achievement.

The temperature, closer to that of a biological system, allowed for a significant reduction in incubation time compared to the previous 24-hour incubation at 4°C. This optimization, alongside the column ratio, temperature, and volume adjustments made throughout the series, led us to a more effective support system for the best possible outcomes. Additionally, the use of pipette tips as the support for immuno-capture contributed to further improving the experimental timeline and allowed for easier automation of the process. The capture process was successful, and the experiments, conducted in triplicates for the first time in this series, exhibited repeatability both in protein amounts and the number of protein groups identified.

3. High-Risk Assessment of Host Cell Proteins

A. Identification of High-Risk HCPs

Following the completion of the experimental phase, we shifted our focus to a high-risk assessment of Host Cell Proteins (HCPs) across all samples: antigen, unbound, and elution. The goal was to identify HCPs that might evade detection by the anti-HCP antibody, particularly those present in the unbound fraction. As mentioned several times, this fraction holds special interest as it represents HCPs that failed to bind to the antibody and could potentially go undetected in ELISA testing.

The antigen sample, representing the complete HCP population, served as a reference point for comparison, while the elution contained HCPs successfully captured by the antibody and thus detectable by ELISA. By examining the unbound fraction, we sought to pinpoint high-risk HCPs that consistently escape interaction with the anti-HCP antibody, highlighting potential blind spots in the standard detection workflow. Addressing these gaps is crucial for improving ELISA's coverage and ensuring the reliability of biopharmaceutical quality assessments.

B. Findings Related to High-Risk HCPs.

After completing the MS protein group identification for the overall fractions, we decided to focus specifically on the high-risk HCPs and further investigate their presence and quantification. To do so, we employed the ANAQUANT ReadyBeads for quantification, as described in previous sections, and analyzed the results in detail.

As an initial assessment, the data is summarized in Figure 58, where the presence of high-risk HCPs is indicated based on either the identification or quantification analysis. Of the 29 high-risk HCPs analyzed, 17 were identified in the antigen fraction. However, only 8 of these HCPs were detected in the quantification analysis, which had more stringent filtering criteria. For the unbound fraction, 16 high-risk HCPs were identified, with 6 quantified, while for the elution, 15 were identified and only 4 quantified.

A striking observation from this comparison is the consistent presence of a specific high-risk HCP, G3HKV9, in the unbound fraction, where it was quantified at 4.8 ppm, while it was absent in the elution for both identification and quantification. Naturally, some high-risk HCPs are expected to overlap between the unbound and elution fractions. This overlap can be attributed to variations in protein concentrations across the fractions, which may necessitate further optimization in the immune-capture process. However, it’s important to note that when a high-risk HCP is detected in the elution fraction, it signifies that the protein has successfully bound to the anti-HCP antibody. This binding implies that the HCP is immune-reactive and, therefore, should be detectable using ELISA, as it is captured and released during the elution process.

In contrast, the consistent absence of G3HKV9 in the elution, despite its repeated presence in the unbound fraction, suggests that this particular HCP does not bind efficiently to the anti-HCP antibody. This lack of binding prevents its detection in the elution fraction and may explain its continued presence in the unbound fraction, where proteins that do not interact with the antibody are retained.

This finding is particularly intriguing, as it suggests that G3HKV9 might exhibit unique properties or characteristics that hinder its ability to bind effectively to the anti-HCP antibody, making it a potential target for further investigation in future experiments.

MS high-risk HCP identification

Accession	Mock	Unbound	Elution
G3I8R9	✓	✓	✓
G3I0W1	✗	✗	✗
A4URF0	✓	✗	✗
A0A061IFE2	✗	✗	✗
G3HR95	✗	✗	✗
G3H0L9	✓	✓	✓
G3INC5	✓	✓	✓
Q9EPP7	✓	✓	✓
G3HNJ3	✓	✓	✓
G3I3Y6	✗	✗	✗
G3H6V7	✓	✓	✓
G3HQY6	✓	✓	✓
G3HRK9	✓	✓	✓
G3GTT2	✓	✓	✓
Q9JKY1	✓	✓	✓
G3I6T1	✓	✓	✓
G3IIE7	✗	✗	✗
G3HC31	✗	✗	✗
G3H3Q1	✗	✗	✗
G3IBF4	✗	✗	✗
A0A3L7HS03	✗	✗	✗
G3IA12	✗	✗	✗
G3I5A4	✓	✓	✓
A0A061I7X9	✗	✗	✗
G3I4W7	✓	✓	✓
G3HAY9	✗	✗	✗
G3HKV9	✓	✓	✗
P14851	✓	✓	✓
Q8R4U2	✓	✓	✓

MS high-risk HCP Quantification

Accession	Mock	Unbound	Elution
G3I8R9	✗	✗	✗
G3I0W1	✗	✗	✗
A4URF0	✗	✗	✗
A0A061IFE2	✗	✗	✗
G3HR95	✗	✗	✗
G3H0L9	✗	✗	✗
G3INC5	✓	✓	✓
Q9EPP7	✓	✓	✓
G3HNJ3	✗	✗	✗
G3I3Y6	✗	✗	✗
G3H6V7	✗	✗	✗
G3HQY6	✓	✗	✗
G3HRK9	✓	✗	✗
G3GTT2	✓	✓	✓
Q9JKY1	✓	✓	✗
G3I6T1	✓	✓	✓
G3IIE7	✗	✗	✗
G3HC31	✗	✗	✗
G3H3Q1	✗	✗	✗
G3IBF4	✗	✗	✗
A0A3L7HS03	✗	✗	✗
G3IA12	✗	✗	✗
G3I5A4	✗	✗	✗
A0A061I7X9	✗	✗	✗
G3I4W7	✗	✗	✗
G3HAY9	✗	✗	✗
G3HKV9	✓	✓	✗
P14851	✗	✗	✗
Q8R4U2	✗	✗	✗

Figure 58: Tables illustrating the presence of high-risk HCPs during MS identification and quantification analysis. The presence of high-risk HCPs in either analysis is indicated by a check mark, while their absence is marked with a cross.

The successful development of this immuno-capture MS-based method has proven effective in assessing and verifying ELISA results, providing deeper insights into the presence of high-risk HCPs. Notably, it enabled the identification of G3HKV9, a high-risk HCP that eluded detection by the anti-HCP antibody. This protein corresponds to Phospholipase A2, an enzyme known to degrade polysorbates such as PS80 and PS20, commonly used surfactants added to monoclonal antibody formulations to enhance stability during storage.[234] The presence of such an enzyme poses a significant risk, as it may compromise the integrity of the drug product over time by indirectly accelerating mAb degradation. This finding underscores the method's ability to uncover functionally critical impurities that may remain undetected in traditional assays, and highlights its potential to support more comprehensive and accurate monitoring of HCP contamination throughout the drug development process.

PART III. LC-MS/MS Workflow Optimization for Host Cell Protein Characterization

Chapter 1
LC-MS/MS analyses as an efficient support for process improvements

The second primary objective of my PhD research focused on enhancing LC-MS/MS workflows to address the complexities of HCP profiling. This effort aimed to deliver improvements in analytical performance, particularly in terms of coverage, sensitivity, and the precision and accuracy of quantification. The overarching goal was to establish a robust methodology capable of reliably identifying and quantifying HCPs across diverse samples, ranging from crude harvests to final drug products, thus contributing to both fundamental understanding and practical biopharmaceutical applications, further illustrated in Figure 59.

This research extended and built upon foundational work carried out in two preceding PhD projects at LSMBO. These earlier efforts had refined various aspects of LC-MS/MS workflows, including optimization of sample preparation protocols and the development of sophisticated bioinformatics pipelines for data interpretation.[234,235] My contribution sought to push these advancements even further, specifically addressing the HCP quantification challenges, a critical aspect in ensuring the reliability of HCP assessments throughout bioprocessing workflows.

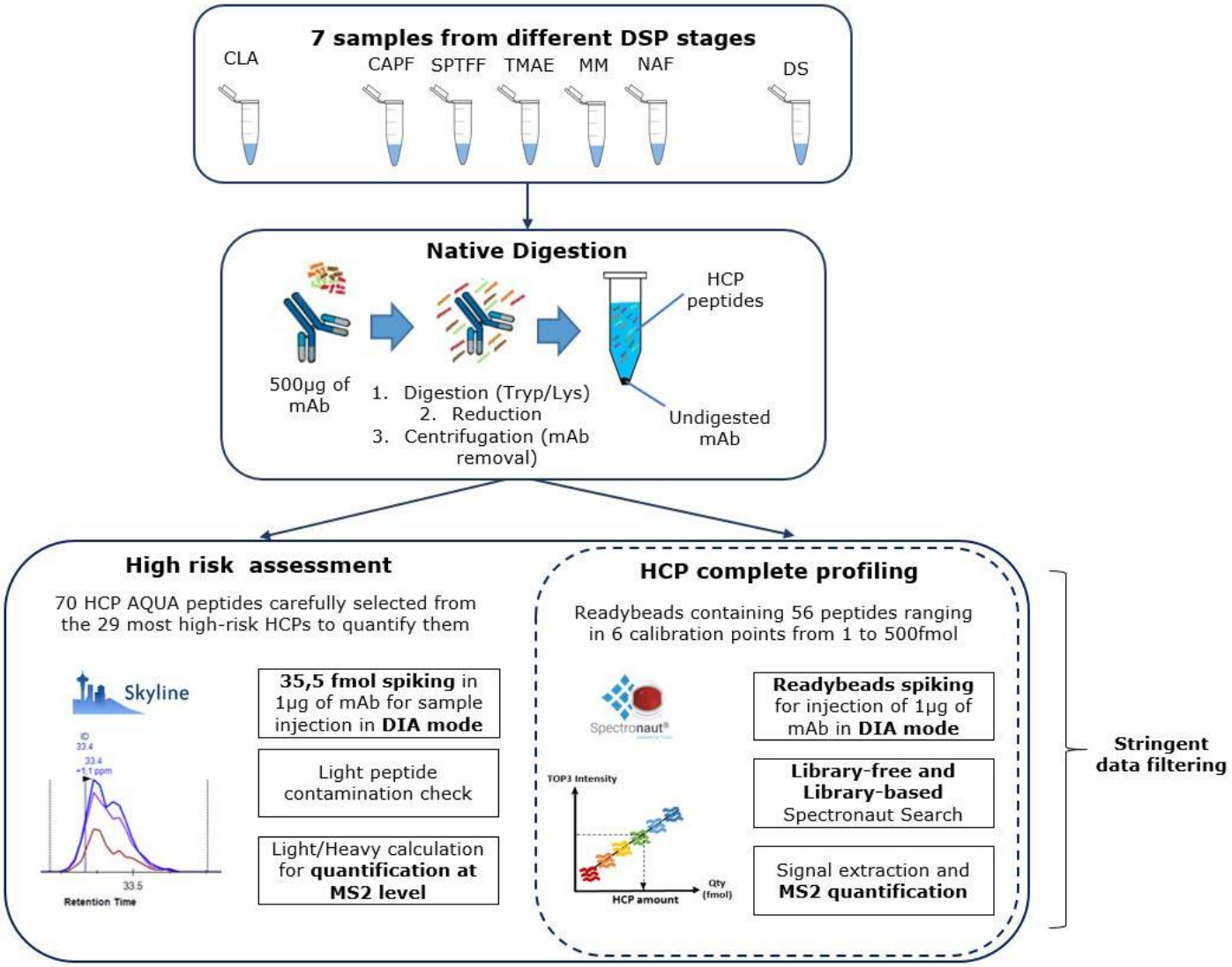


Figure 59: Analytical Workflow of DSP Cascade Samples. Illustration of the analytical workflow applied for DSP cascade samples HCP characterisation. The samples first undergo native digestion, as detailed in a subsequent section, where the digestion step is performed to favor HCP digestion favoring mAb's depletion. After this, the workflow branches into two distinct approaches. On the right, the workflow focuses on HCP complete population quantification. ReadyBeads are solubilized and spiked into the samples, which are then injected for analysis. Quantification is performed using a calibration curve, with stringent data filters applied to ensure confident and reliable results. On the left, the workflow outlines the high-risk HCP assessment, performed using two complementary methods. The first utilizes ReadyBeads, targeting a list of the 29 most high-risk proteins. The second employs the AQUA peptides approach, using a heavy-labeled peptide mix to specifically quantify these problematic HCPs. The workflow details the spiking process, light peptide assessment, and quantification, all conducted at the MS2 level for precise analysis.

To achieve these objectives, a dual quantification strategy was implemented that integrated innovative spike-in peptide standards to elevate the precision and accuracy of the quantification workflow. The first component of this strategy utilized a panel of 54 synthetic peptides spanning a wide dynamic range of concentrations to construct an internal calibration curve (ReadyBeads by Anaquant) as

detailed in Table 6. This approach enabled comprehensive and reproducible quantification of HCPs across all stages of the purification process, ensuring uniformity and consistency in profiling even when faced with highly complex sample matrices.

Protein Standards	Peptides	Quantity (fmol)
ANAQUANT-1	DGALLENVTVR	1
	EGAEPEIYNAIR	
	GDVAVFFGLSGTGGK	
ANAQUANT-2	EGCDLAGAIK	
	VGFEGNRPTNSILLR	
	VGNPETTLFLVASK	
ANAQUANT-3	LGAADVTGVPTLLVAR	
	VGLVPTQEAIQK	
	VGQQPEFAAAK	
ANAQUANT-4	AGVVEELAR	2.5
	VGQLLGSGSILR	
	VLGTDGFGGR	
ANAQUANT-5	AGFDFACLPNEGVGLAR	
	AGLLEFDDQEPQLQNEIR	
	GGVALSAGVQR	
ANAQUANT-6	GGSGPYFYLPK	
	VGIASELGEER	
	VGIDGQINLR	
ANAQUANT-7	AGLAEHGIVFGEPK	10
	EGVRPDAIICTGR	
	VGALLSHSNFGSSDCPSSSK	
ANAQUANT-8	CGADLGLETVIVER	
	FGTGANTLEVEGENGK	
	TGQVVVLGAGPAGYSAAFR	
ANAQUANT-9	TGYSGLDYPSLEAVIR	
	VGLSGPGLVNLYR	
	YGALVGDVGGTNAR	
ANAQUANT-10	GGLTDAAQQVVAAVEGK	50
	LGGADGNALFR	
	VGLEVTLR	
ANAQUANT-11	AGHPQLAEEFTR	
	TGVIGFGSPNK	
	YGINELQANPAK	
ANAQUANT-12	GGTLGQDVIDIR	
	IGTFIDGDEGILLHR	
	YGSIGQPFVYPR	
ANAQUANT-13	GGPLTTPVGGGIR	250
	TGVTYDFER	
	YGYQGTPSPVK	

ANAQUANT-14	AGLQAIAGPFSQVR	
	FGCPTGGISPANYR	
	TGSAESILTTGPVVPVIVVK	
ANAQUANT-15	FGFSQPLLLGK	
	GFGVTTLDIIR	
	SGDLFNVNAGIVK	
ANAQUANT-16	AGDAFAVIVK	500
	EGQGLTPVLCIGETEAENEAGK	
	IGYQLKPNPAVLICR	
ANAQUANT-17	EGSGLLGLTEVTSDCR	
	LGVLVLNCGSSSLK	
	YGTSSVVIDESVIQGIK	
ANAQUANT-18	EGLPLTESLALTIDR	
	VGIPYWNETILPR	
	YGYLGNADEIAAK	

Table 6: ANAQUANT peptide sequences and fmol quantities.

In parallel, a second, more targeted approach was developed to address high-risk HCPs, those deemed particularly challenging or potentially impactful to product quality[7]. This strategy incorporated, in addition to the Ready-beads, a custom-designed mixture of 70 heavy-labeled AQUA peptides by Thermo Fisher Scientific. These peptides were meticulously purified, accurately quantified, and selected to represent 29 high-risk HCPs previously identified as critical through earlier risk assessments. By enabling absolute quantification at exceptionally low concentrations, this method ensured that even trace amounts of these proteins could be reliably detected and measured, as described in the Chapter 2 of Part III.

Together, these approaches formed a highly refined LC-MS/MS workflow capable of delivering unparalleled quantification performance. The dual quantification strategy not only improved the precision and accuracy of HCP assessments down to the sub-ppm level but also offered invaluable insights into HCP dynamics across the purification process. This level of detail proved instrumental in risk assessment, particularly in identifying potential HCP carryover into the final drug product.

1. Sample Preparation for Global Downstream Analysis

A. Overview of Avelumab Sample Stages

Merck supplied seven samples of the Avelumab monoclonal antibody, commercially known as Bavencio, to facilitate the analysis of the downstream process cascade. These samples correspond to various stages of the process, beginning with the clarified harvest (CLA), which represents the sample taken prior to any purification steps. The next stage involves protein A affinity chromatography (CAPF), which captures the monoclonal antibody while removing process-related impurities. Following CAPF, the virus inactivation step is carried out at low pH to ensure viral safety.

Next, single pass tangential flow filtration (SPTFF) is used to concentrate the intermediate bulk before it is loaded onto trimethylaminoethyl (TMAE) anion exchange chromatography resin for further purification. The subsequent step involves mixed mode (MM) chromatography, which further reduces impurities.

Afterwards, nanofiltration (NAF) is performed to filter out any residual viruses, and ultrafiltration and diafiltration are carried out to achieve the final drug substance (DS). Additionally, Merck provided the corresponding null cell line (Mock) sample for use in the library-building process, which is described in detail in Part II, Chapter 2 of the manuscript, and for the optimization of the workflow.

B. DSP Cascade Sample Preparation Workflow for HCP Analysis

One of the major challenges posed by DSP cascade samples is their extreme dynamic range, particularly as the purification chain progresses closer to the drug substance. At these later stages, the concentration of the mAb becomes disproportionately high compared to the residual HCPs, complicating the analysis. To address this, a native digestion protocol adapted from Huang et al.[236] was employed, specifically designed to deplete the dominant mAb component and enable more effective detection and quantification of HCPs.

Each sample, containing 500 µg of mAb, was adjusted with ultrapure H_2O and Tris-HCl buffer (pH 8) to a final volume of 150,2 µL. Digestion was carried out overnight at 37°C for 15 hours using a Trypsin/Lys-C enzyme mixture at a 1:400 enzyme-to-protein ratio.

Following digestion, peptides were reduced with DTT for 10 minutes at 90°C to disrupt disulfide bonds, then centrifuged at 13,000 g for 2 minutes to separate insoluble components. Acidification with formic acid was performed to quench enzymatic activity. Samples were subsequently vacuum-dried and resuspended in a solution of water, acetonitrile, and formic acid (98:2:0.1), yielding a final peptide concentration of approximately 1 µg of mAb per µL.

For quantification, ANAQUANT ReadyBeads were used to provide an internal calibration curve. These beads, containing a mixture of 54 standard peptides, were spiked into the samples at volumes corresponding to 1 µg of mAb. This approach ensured a comprehensive calibration across six concentration points, improving the accuracy of HCP quantification in the complex DSP samples.

Samples prepared with the ANAQUANT ReadyBeads were analyzed using the Orbitrap Eclipse™ Tribrid™ mass spectrometer. Peptide separation was achieved over the 85-minute gradient and method employed during the analysis of the column support. The UPLC system was equipped with an Acclaim PepMap 100 C18 pre-column (20 mm × 0.1 mm, 5 µm, 100 Å) from Thermo Fisher Scientific, alongside an Aurora Series C18 UHPLC column (250 mm × 75 µm, 1.6 µm, 120 Å) from IonOpticks.

Mobile phase A consisted of H_2O with 0.1% formic acid (FA), while mobile phase B was a mixture of acetonitrile (ACN) and H_2O with 0.1% FA (79.9/20/0.1). Peptides were loaded at a flow rate of 300 nL/min and separated over an 85-minute gradient. The percentage of mobile phase B increased to 43.8% by minute 65, then ramped up to 98% by minute 66 for washing, maintaining this level until minute 71. The system was then re-equilibrated with 2.5% phase B from minute 72 to minute 85.

Full-scan MS spectra were acquired across a 375-1200 m/z range at a resolution of 120,000, with an AGC target of $8x10^5$ and a maximum injection time of 200 ms. MS/MS spectra were acquired at a resolution of 15,000, with a maximum injection time of 22 ms. HCD normalized fragmentation collision energy was set to 30%. Fragment analysis was performed using 18.4 m/z windows with a 0.5 m/z window overlap, ensuring comprehensive peptide fragmentation for accurate identification and quantification.

2. Data Processing and Adapted Filtering

For protein identification and quantification, we used Spectronaut™ software (v.18), uploading the raw files alongside the CHO database that included *Cricetulus griseus* entries (78,366 entries, 2022/03/15) extracted from UniProtKB/TrEMBL. The analysis was carried out with trypsin/P as the digestion enzyme, permitting one missed cleaved peptide. We also incorporated two variable modifications: methionine oxidation and N-terminal acetylation of the protein. To ensure accurate identification, we set a precursor and protein Q-value cutoff at 0.01. For quantification, we relied on MS^2 XIC peak areas, with no imputation, and required at least three fragments per precursor, applying interference correction where necessary. Depending on the approach, the data was processed either with DIA, leveraging the comprehensive spectral library previously decribed in Part II of the manuscript, or DirectDIA.

After the data acquisition, an in-house RScript was used for filtering, following the guidelines of Pythoud et al. The filtering steps included several key criteria, as observed in Figure 60.

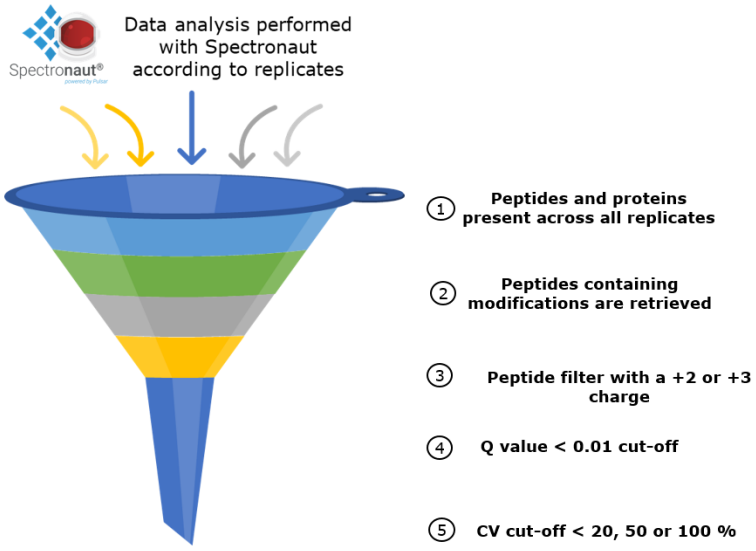


Figure 60: Data analysis filters applied post-Spectronaut analysis. The filters are numbered sequentially, corresponding to the steps outlined in the RScript. These filters ensure the reliability and robustness of the results by systematically refining the analysis.

First, peptides and proteins that were present in all replicates were retained to ensure consistency and reliability of the data. Peptides with oxidation or acetylation modifications were also specifically included, as these post-translational modifications are important for accurate protein characterization.

Next, only precursors with a charge state of +2 or +3 were kept, as these charge states are typically more stable and produce higher-quality data during analysis. A Q-value cutoff of less than 1% was applied to ensure that the FDR-adjusted P-values were within an acceptable range, minimizing false positives.

To assess data variability, three different coefficients of variation (CV) thresholds were applied: 20% and no filter. This allowed for flexible analysis based on the required stringency for each experiment.

Following filtering, peptide intensity was calculated by summing the precursor ion intensities. Protein intensity was then determined by adding the intensities of the top three most abundant peptides using the Top3 strategy. For HCP quantification, individual HCP concentrations (ng HCP/mg mAb or ppm levels) were estimated using the calibration curve derived from the ANAQUANT ReadyBeads profiler, ensuring precise measurement of HCP contamination levels in the samples.

3. Quantification Results of Downstream Process cascade Analysis

A. Impact of employing a comprehensive Spectral Library on HCP Quantification

The use of a comprehensive spectral library played a crucial role in HCP quantification across various stages of the downstream processing cascade. The seven purification stages were analyzed, with samples spiked with the ReadyBeads standards to generate calibration curves for accurate quantification, as mentioned in the previous section.

The calibration curves obtained at each stage exhibited strong correlations with R^2 values greater than 0.7, indicating the robustness and reliability of the quantification process, even in the face of increasingly complex sample matrices and expanding dynamic ranges throughout the purification process.

As expected, a significant reduction in HCP concentration was observed throughout the purification stages, particularly as the process advanced toward the final drug substance step (Figure 61). The most pronounced decrease in HCPs occurred between the SPTFF and TMAE stages, where the number of identified HCPs dropped sharply from 1143 to 23 proteins, a 98% reduction. This confirms the expected efficiency of the purification process at this critical step. This trend was anticipated based on prior knowledge of the purification process and was confirmed by the mass spectrometry results.

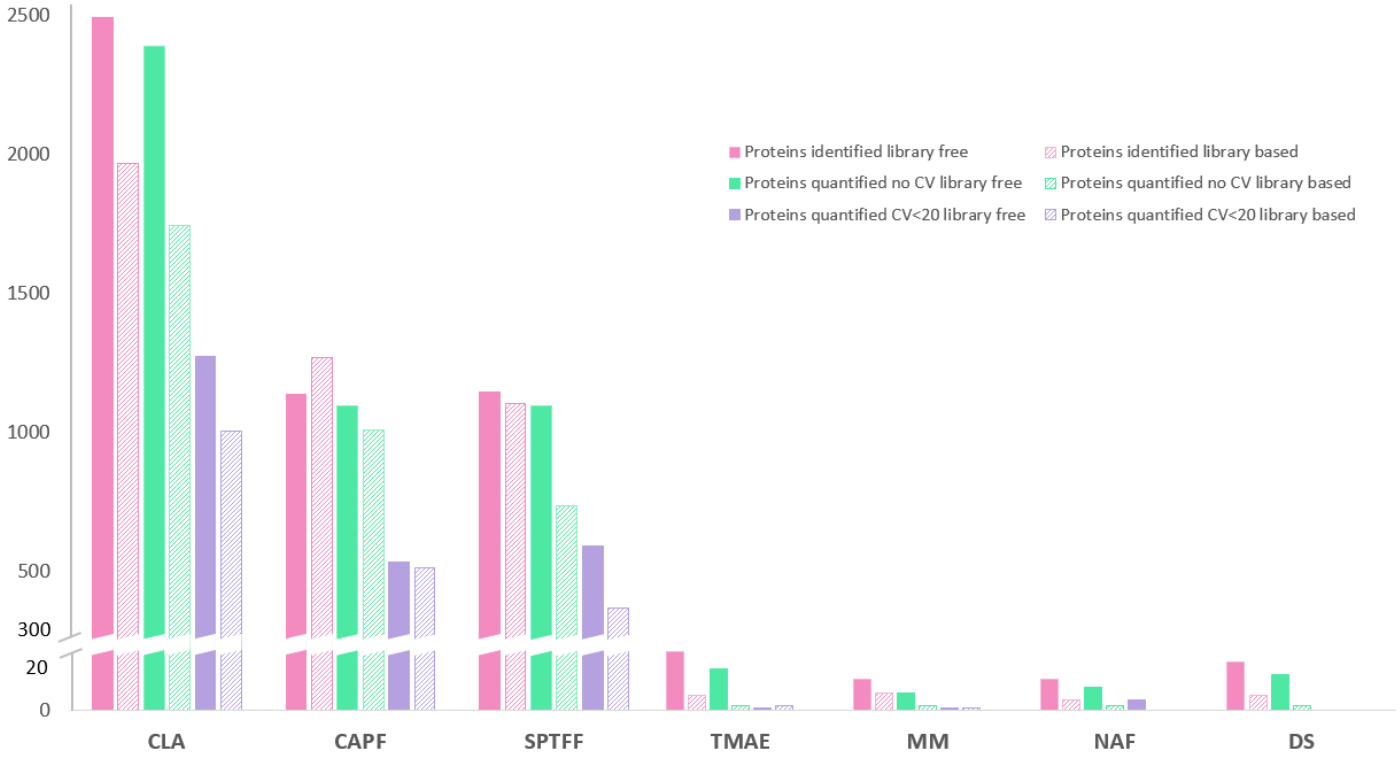


Figure 61: Histogram illustrating the number of protein groups identified and quantified using both the library-free and library-based approaches. The data is shown with and without the application of a CV cutoff at 20. The histogram highlights the differences in protein identification and quantification between the two methods, demonstrating the impact of the CV filter on the results across both approaches.

In the clarified harvest (CLA) stage, for instance, the library-free method identified 2493 proteins, which was a 27% increase over the 1970 proteins identified by the library-based method. Similarly, at the protein quantification stage, the library-free approach quantified 2390 proteins, marking a 37% increase compared to 1743 proteins quantified by the library-based method. This demonstrates the broad coverage and enhanced sensitivity of the library-free method, especially in the early stages of the purification chain.

However, in spite of this slight advantage when employing this approach, the application of the stringent CV<20 filter had a significant impact, particularly in the later stages, where HCP concentrations were lower, and the biotherapeutic drug concentration was higher. In these later stages, the CV<20 filter resulted in a sharp reduction in the number of proteins quantified, regardless of the used approach. For instance, in the TMAE stage, only 1 protein was quantified using the library-free approach, while the library-based approach quantified 2 proteins. This difference highlights the restrictive nature of the CV filter when applied to samples with low HCP abundance, where the filter's exclusivity may inadvertently exclude relevant proteins.

In these final stages of purification, the application of the CV filter led to a situation where only a limited number of proteins could be quantified. Some of these proteins were identified using Top 1 or Top 2 peptides, rather than the more robust Top 3 peptides. As a result, for the final data analysis

comparison for MS and ELISA, which will be discussed in a further section of the manuscript, the CV<20 filter was not applied, ensuring that the potential loss of low-abundance proteins due to excessive filtering was avoided.

The number of quantified HCPs became increasingly challenging as the purification process moved into the MM, NAF, and DS stages, where very few, if any, proteins were quantified. Despite this, the library-free approach demonstrated greater resilience to the CV<20 filter and gave an overall better performance, maintaining a larger proportion of identified and quantified proteins in these late stages. These observations suggest that the library-free approach, unconstrained by a pre-constructed spectral library, provides a more comprehensive view of the HCP landscape, particularly for low-abundance HCPs that may be overlooked in the library-based method.

Furthermore, the final downstream stage, corresponding to the Drug Substance (DS), did not exhibit the smallest number of identified or quantified HCPs. This was an unexpected outcome given the assumption that fewer proteins would be detected at this stage due to the advanced level of purification. However, this result may be attributed to the specific digestion method employed, namely native digestion. Unlike the traditional proteomic digestion workflow that uses denaturing conditions, native digestion was used to prioritize the digestion of HCPs by performing the digestion step in the beginning of the workflow and therefore manage to facilitate the depletion of the indigested mAb, which ,without any denaturation step, will indeed not be able to be digested properly given it's steric structure. Although native digestion reduces the dynamic range, it can introduce variability.

The quantitative outcomes from the two approaches also revealed notable differences, particularly when evaluating parts per million (ppm) levels of HCPs, as observed in Figure 62. The library-free method consistently showed higher ppm values compared to the library-based method, indicating greater sensitivity in detecting low-abundance HCPs in the final stages of purification. Specifically, in the DS stage, the library-based approach quantified only two HCPs, while the library-free method identified 13, with only one protein common to both methods. The ppm levels were 6,1 ± 4 for the library-based method versus 20,5 ± 12 for the library-free approach. These results suggest that the library-free method, with its broader scope and flexibility, was able to detect a more extensive range of low-abundance HCPs, which would otherwise have been missed using the library-based strategy.

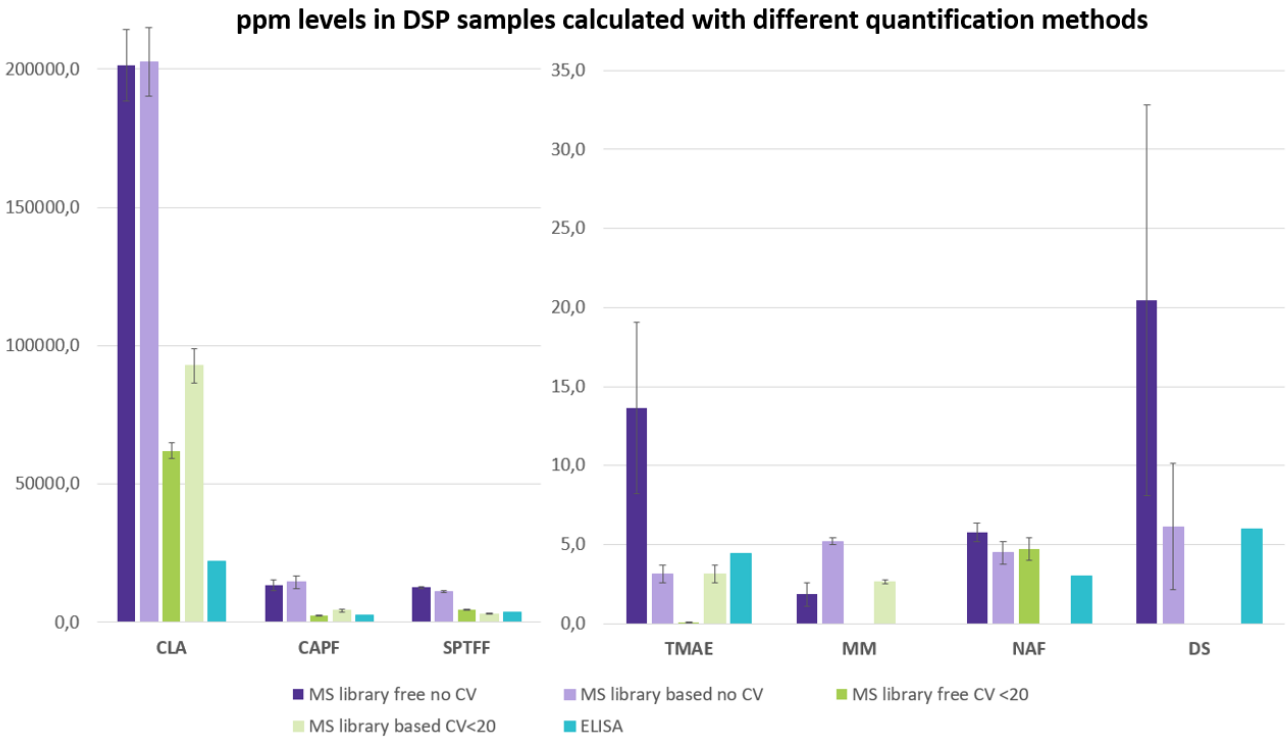


Figure 62: Comparison of library-free and library-based approaches, highlighting the impact of the CV filter on protein quantification, as shown by ppm levels. The histogram illustrates the changes in protein quantification before and after the critical purification step, with a cut on the axis to better visualize the differences in quantification. This figure emphasizes the effect of the CV filter's stringency on quantification results, comparing the two approaches throughout the purification process.

The discrepancies observed in the HCP profile between the two methods can be attributed to differences in how the spectral libraries are constructed. The library-based method relies on a curated set of proteins from a defined cell line, which may not capture the full spectrum of potential HCPs in the final stages of purification. In contrast, the library-free approach, while it may sacrifice some specificity, provides a broader snapshot of the protein landscape, enabling the identification of proteins that may be absent from a curated library. This highlights the inherent trade-off between accuracy and comprehensiveness, particularly in complex biological samples with a wide range of protein concentrations.

Ultimately, these findings underscore the importance of choosing the appropriate analytical strategy based on the specific needs of the study. While the library-based approach provides a more structured and targeted means of identifying HCPs, the library-free approach offers flexibility and broader coverage, especially useful in detecting low-abundance impurities at later stages of the purification process. The results also suggest that future strategies could benefit from a combination of both methods, depending on the nature of the sample and the objectives of the analysis.

B. Comparison of LC-MS/MS results with ELISA

Accurate identification and quantification of HCPs are critical in ensuring the safety and efficacy of biotherapeutics. As outlined in the bibliographic section of this manuscript, ELISA has traditionally

been the industry standard for HCP quantification. However, mass spectrometry has emerged as a powerful complementary tool, offering distinct advantages. Notably, MS can identify individual proteins and quantify them separately, whereas ELISA provides only a cumulative ppm value for total HCPs. This distinction highlights MS's transformative potential for characterizing HCPs, especially in samples with low-abundance proteins or complex impurity profiles.

In comparing these two methods, both MS and ELISA reveal a consistent trend of HCP reduction throughout the purification cascade, as depicted in Figure X (above). The CLA, which represents the earliest stage of the purification process, exhibits the highest ppm levels in both methods due to the abundance of impurities. As purification progresses, HCP ppm levels decrease substantially, with the most significant drop occurring between the SPTFF and TMAE chromatography steps. While both methods indicate this trend, MS offers greater sensitivity, consistently detecting higher ppm levels at each stage compared to ELISA. For example, at the DS stage, MS measured an average of 20.5 ± 12 ppm, whereas ELISA reported only 6.1 ppm. This discrepancy underscores MS's ability to detect low-abundance proteins that may elude ELISA.

Additionally, MS provides the unique capability to identify individual HCPs within each sample. In Figure 63, we present the exhaustive list of HCPs identified and quantified in the DS sample, representing the final stage of purification. This table includes proteins quantified using the Top 1, 2, or 3 peptides, their respective contributions to the total HCP content, and their coefficient of variation. The ability of MS to uncover individual HCPs, including those previously undetected by ELISA, demonstrates its enhanced utility for comprehensive HCP profiling.

HCPs quantified	Mean ppm	Top 1	Top 2	Top 3	Percetange
A0A061I5D1	1,06	✓			5,16
A0A061I8N8	0,42	✓			2,08
A0A061IB69	0,03	✓			0,17
A0A061IQJ9	0,03	✓			0,17
A0A3L7HB03	6,44			✓	31,47
A0A3L7HWC9	0,61			✓	2,96
A0A3L7IB04	10,05			✓	49,10
A0A3L7IB12	0,49	✓			2,38
A0A3L7IB55	0,99			✓	4,81
A0A3L7IL35	0,01	✓			0,04
A0A8C2LSP4	0,05	✓			0,24
A0A8C2MY60	0,19		✓		0,91
G3IBN9	0,10		✓		0,47

Figure 63: Table displaying the ppm values of HCPs identified in the Drug Substance (DS) stage, the final purification step. The table includes the percentage of each HCP relative to the total sample, and the quantification approach used (Top 3, Top 2, or Top 1).

Interestingly, despite the DS stage representing the final step of the purification process, both methods reveal higher ppm levels here than expected. This anomaly, observed consistently across MS and ELISA results, may stem from the specific cohort of samples analyzed or the complexity of the matrix itself. The broader range of HCPs detected by MS reinforces its critical role in providing a detailed and accurate understanding of the impurity profile, a key factor in ensuring the safety and regulatory compliance of biotherapeutic products.

MS's advantages over ELISA include its ability to quantify individual proteins, detect low-abundance HCPs, and provide more detailed insights into impurity profiles at every stage of the purification process. While ELISA remains a valuable tool for routine quality control due to its simplicity and cost-effectiveness, the depth and specificity of information offered by MS make it indispensable during early development and process optimization. Together, these methods form a robust analytical framework for HCP analysis, ensuring the consistent production of safe and effective biotherapeutics.

C. Conclusions on critical steps in the Downstream Process

The downstream process cascade of biotherapeutics plays a critical role in ensuring the safety, efficacy, and regulatory compliance of the final product. Throughout the research conducted in this section, we identified several key stages in the DSP cascade that are critical to achieving effective HCP clearance. Notably, the SPTFF and TMAE chromatography steps were found to be crucial for HCP depletion, consistently showing significant reductions in HCP levels. These stages were integral in reducing the overall HCP burden, reinforcing the importance of carefully selected purification steps in achieving high purity in the final drug substance.

A key observation from our study was also the initial substantial drop in both the identification and quantification of HCPs, particularly from the CLA to the CAPF step. This trend was consistent across both proteomic identification and quantification, where a large decrease in ppm levels was observed, reinforcing the effectiveness of Protein A affinity chromatography in the purification process. Although Protein A exhibits a strong affinity towards the mAb, this alone is not sufficient for complete HCP removal. The observed reduction in HCPs from CLA to CAPF highlighted the necessity of employing multiple other purification steps to achieve the desired level of purity, further justifying the use of complementary techniques like SPTFF and TMAE chromatography. This finding suggests that a single step cannot fully address the complexity of HCP removal, making it crucial to implement a multi-stage strategy to reach optimal product quality.

Interestingly, while the number of identified proteins decreased significantly across purification stages, the drop in ppm levels was even more pronounced. This highlights the importance of accurate quantification in understanding the full extent of HCP removal during the DSP cascade. While identification provides insight into the protein groups present, quantification offers a more detailed understanding of how much of each impurity remains at each step. The ability to quantify HCPs at each stage not only complements the protein identification process but also enables a more precise evaluation of the effectiveness of each purification step. This approach is crucial for obtaining a complete profile of HCP depletion, ensuring that potential low-abundance contaminants are not

overlooked. Quantification provides an additional layer of understanding that is vital for fine-tuning purification processes and achieving the required safety and efficacy standards for the final therapeutic product.

Moreover, the integration of mass spectrometry for HCP profiling throughout the DSP process provides a more comprehensive and robust analysis. This section highlighted that even though traditional methods like ELISA offer useful insights into the trends of HCP depletion, they fall short in providing detailed, individual protein quantification. MS, on the other hand, enables not only identification of HCPs but also precise quantification of each protein, shedding light on the exact contribution of each impurity at every purification stage. This nuanced approach significantly enhances our understanding of HCP behavior and clearance, providing a valuable tool for optimizing the DSP process. By integrating this advanced analytical technique as an orthogonal method to assess the ELISA results, we gained deeper insights into the purification process, reinforcing the importance of a multi-step purification strategy in the production of safer, higher-quality biotherapeutics.

Chapter 2
Focus on High-Risk HCP Assessment and Absolute Quantification Approaches

1. Methodology for high risk HCP assessment

A. Importance of studying high-risk HCPs in biopharmaceutical production

High-risk HCPs[7] are particularly critical to monitor because they can pose potential safety risks, including immunogenic reactions, if present in the final biopharmaceutical product. These proteins, even in low quantities, can affect the drug's safety and efficacy. Therefore, accurately identifying and quantifying these HCPs throughout the downstream process is essential for ensuring that the final product meets the required safety standards and is free from contaminants that could cause adverse effects in patients.

Given the challenges associated with removing all HCPs during purification, even trace amounts of high-risk proteins can remain in the final drug substance. Regulatory authorities demand comprehensive HCP assessments to confirm that these proteins are effectively cleared and do not pose a risk. By focusing on high-risk HCPs, manufacturers can better optimize their purification strategies and enhance the safety and quality of their biotherapeutic products.

Following the comprehensive HCP profiling of the DSP cascade, we shifted our focus to high-risk HCPs, utilizing a dual quantification strategy to enhance the precision and accuracy of our risk assessment. As outlined in the previous section, we initially employed a mixture of 54 standard peptides to construct an internal calibration curve from Anaquant, which allowed for detailed HCP profiling and consistent quantification across the purification stages.

In parallel, we implemented a targeted approach to address high-risk HCPs, using a custom mixture of 70 heavy-labeled AQUA peptides developed in collaboration with Thermo Fisher Scientific. These peptides were carefully selected to represent 29 high-risk HCPs identified as particularly challenging or potentially impactful to product quality[7]. By enabling absolute quantification of these specific proteins, even at trace levels, this strategy provided exceptional precision in tracking these critical impurities down to the sub-ppm level.

Together, these complementary approaches provided a more comprehensive and accurate risk assessment across the entire DSP cascade. The integration of broad-range profiling with targeted AQUA peptide quantification ensured that both general and high-risk HCPs were reliably detected and quantified, ultimately improving the ability to assess HCP dynamics and carryover into the final drug product. This dual strategy significantly enhanced the overall robustness and reliability of our HCP analysis, supporting the development of safer, higher-quality biotherapeutics.

B. Overview of absolute quantification techniques for High-Risk HCPs

i) Quantification using ANAQUANT ReadyBeads

The preparation of DSP cascade samples for HCP analysis was challenging due to the extreme dynamic range, particularly in the later stages of the purification process, where the mAb concentration is disproportionately higher than the residual HCPs. To address this, the native digestion protocol, as described in the previous section, was followed. This protocol, adapted from Huang et al., was specifically designed to deplete the dominant mAb component and enable the detection and quantification of HCPs with greater sensitivity.

Each sample, containing 500 µg of mAb, was processed with a Trypsin/Lys-C enzyme mixture at a 1:400 enzyme-to-protein ratio, digested overnight at 37°C. Following digestion, the peptides underwent reduction with DTT at 90°C to disrupt disulfide bonds, and the resulting solution was centrifuged to separate insoluble components. Acidification with formic acid halted the enzymatic activity and stabilized the peptides for subsequent analysis.

The prepared samples were then resuspended in a solution of water, acetonitrile, and formic acid (98:2:0.1), yielding a final peptide concentration of approximately 1 µg/µL. For accurate quantification, ReadyBeads were used, containing a mixture of 54 standard peptides that allowed for the generation of an internal calibration curve. This enabled precise HCP quantification across a wide concentration range, enhancing the consistency and reliability of the analysis, especially in complex sample matrices.

The peptides were analyzed using an Orbitrap Eclipse™ Tribrid™ mass spectrometer, with a UPLC system equipped with an Acclaim PepMap 100 C18 pre-column and an Aurora Series C18 UHPLC column. Peptide separation was achieved over an 85-minute gradient, and full-scan MS and MS/MS spectra were acquired for comprehensive peptide fragmentation and identification. These steps provided robust and reproducible HCP profiling, critical for assessing the purification stages in the DSP cascade and ensuring the accuracy of quantification in the presence of high mAb concentrations.

Full-scan MS spectra were acquired across a 375-1200 m/z range at a resolution of 120,000, with an AGC target of $8x10^5$ and a maximum injection time of 200 ms. MS/MS spectra were acquired at a resolution of 15,000, with a maximum injection time of 22 ms. HCD normalized fragmentation collision energy was set to 30%. Fragment analysis was performed using 18.4 m/z windows with a 0.5 m/z window overlap, ensuring comprehensive peptide fragmentation for accurate identification and quantification.

For protein identification and quantification, we utilized Spectronaut™ software (v.18) alongside a CHO database obtained from UniProtKB/TrEMBL, which included 78 366 Cricetulus griseus entries (as of 2022/03/15). The analysis employed trypsin/P as the digestion enzyme, allowing for up to one missed cleavage site. Methionine oxidation and protein N-terminal acetylation were set as variable modifications. To ensure high confidence in protein identification, we applied a Q-value cutoff of 0.01 at both the precursor and protein levels. Quantification relied on MS^2 XIC peak areas, without imputation, requiring at least three fragments per precursor, with interference correction applied as needed.

The data filtering and RScript-based analysis detailed in the previous section were also applied here, following the same processing workflow. However, for this dataset, the focus was uniquely on the high-risk HCPs identified in prior assessments.

ii) Quantification using AQUA Peptides

For the second quantification strategy, utilizing the AQUA heavy-labeled peptide mixture, 35.5 fmol of the peptide mix was spiked into each sample prior to injection. These peptides were meticulously designed to allow for absolute quantification of 29 high-risk HCPs. Additionally, in both AQUA peptide-based and ANAQUANT ReadyBeads workflows, iRT peptides were added to the samples to assist with retention time alignment. Specifically, iRT peptides were introduced at proportions of 10% and 5% of the final mixture for AQUA and ANAQUANT-based approaches, respectively, ensuring accurate timepoint correlation for all identified peptides.

The DSP samples containing the AQUA heavy-labeled peptides were processed using the same chromatographic system, mobile phase composition, and mass spectrometer as previously described. However, given that the Skyline library included predefined retention times for the AQUA peptides, parameters established during prior kit testing were used. These conditions employed a slightly modified gradient to maximize peptide separation efficiency. Samples were loaded at a flow rate of 300 nL/min, with peptide separation conducted over a 125-minute gradient. Mobile phase B was increased to 25% by minute 60, then to 37.5% by minute 90. It was subsequently ramped to 98% at minute 100 for column washing, maintained at this level until minute 105, and then re-equilibrated with 2.5% phase B from minute 110 to minute 125.

For the mass spectrometry analysis, the settings were consistent with those applied for the ANAQUANT ReadyBeads workflow, ensuring reproducibility and compatibility between the two strategies. The Orbitrap Eclipse™ Tribrid™ mass spectrometer provided high-resolution detection, enabling precise identification and quantification of peptides, including the spiked AQUA heavy-labeled standards.

Raw data from the AQUA peptide runs were imported into Skyline for targeted analysis. Each peptide was subjected to rigorous manual integration to confirm its presence, as illustrated in Figure 64. Two stringent criteria were applied: an isotopic dot product (idotp) value exceeding 0.8 and consistent retention times across multiple analyses. The establishment of the library was crucial for enabling the manual data processing, as it allowed for the verification of the retention time and identity of the AQUA peptides, ensuring that only validated signals were quantified.

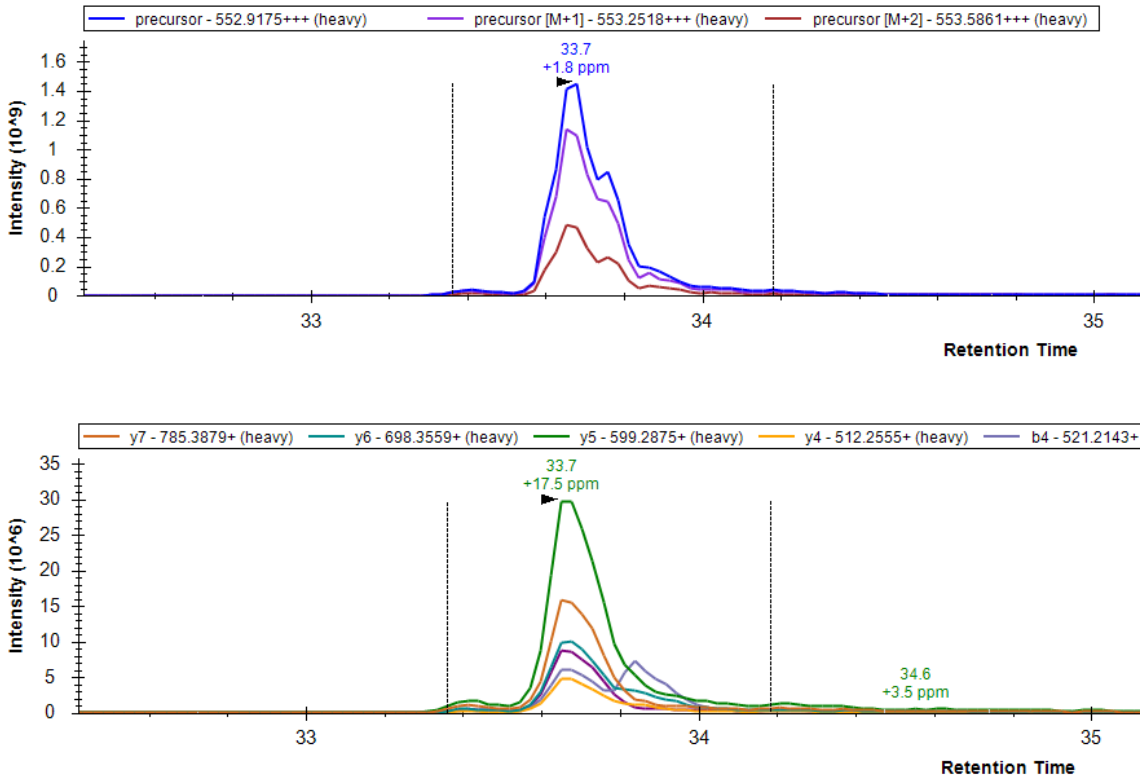


Figure 64: Example of an extracted ion chromatogram (XIC) corresponding to one of the heavy peptides that was manually integrated, with its corresponding fragment ion signals displayed below, based on Skyline data processing.

Before quantifying high-risk HCPs, a contamination assessment of the AQUA mix was conducted. The mixture was injected in replicates to evaluate the percentage of light peptides present as contaminants. This step confirmed the purity of the spiked heavy-labeled peptides and minimized the risk of interference in subsequent analyses. After manual integration of the peptide peaks, the mean area of each heavy and corresponding light peptide was calculated. The light/heavy response factor was derived from this data, enabling accurate estimation of high-risk HCP concentrations based on the known quantity of spiked peptides.

This workflow, combining rigorous data analysis in Skyline with the carefully designed AQUA peptide mixture, provided exceptional sensitivity and accuracy for detecting and quantifying high-risk HCPs. By employing a robust and validated methodology, this approach ensured that even trace levels of critical HCPs could be measured reliably, thereby enhancing the overall assessment of HCP risk throughout the downstream process.

2. Integration of LC-MS/MS Data for Risk Evaluation

A. Comparison of ANAQUANT ReadyBeads and AQUA Peptides

Our analysis focuses on a list of 29 HCPs, which have been previously identified in the literature as particularly critical due to their potential to elicit immune responses or affect the overall therapeutic efficacy[7]. These high-risk HCPs are of great concern, not only because of their potential safety risks but

also because their detection and removal can significantly influence the overall quality of the final product. To ensure the safety and efficacy of the therapeutic product, a comprehensive risk assessment is critical, relying on two specialized quantification techniques: the ANAQUANT ReadyBeads and the AQUA heavy-labeled peptide mixture by Thermo.

The ANAQUANT ReadyBeads method, as previously discussed, is designed to provide a broad overview of the entire HCP population present in the samples. By targeting a wide spectrum of HCPs, this method serves as an initial, generalized quantification tool, identifying a range of impurities across different stages of the purification process. While effective for broad quantification, it was put to test in terms of the specificity necessary for high-risk HCPs that may have more serious implications for product safety.

In contrast, the AQUA heavy-labeled peptide mixture method is specifically designed to focus on a narrower subset of HCPs, those known to be high-risk due to their immunogenicity or clinical relevance. This highly targeted approach allows for enhanced sensitivity, enabling the detection and quantification of these critical impurities even at low concentrations, which may otherwise remain undetected by broader methods.

To ensure the accuracy of the AQUA peptide-based approach, a thorough contamination assessment was performed before its application in the sample analysis. This crucial step validated the integrity of the peptide mixture used for quantification. Specifically, three replicates of the AQUA peptide mixture, containing the 70 distinct peptides labeled with heavy isotopes, were injected at a concentration of 35.5 fmol per injection, identical to the quantity used in the spiking process. This step aimed to establish the light-to-heavy peptide ratio and to calculate the percentage of light peptide contamination within the mixture. By determining the precise ratio of light to heavy peptides, we could create a reliable baseline for subsequent quantification in experimental samples.

This contamination assessment ensured that the peptides used for quantifying high-risk HCPs were accurately represented in terms of their concentration, accounting for any intrinsic contamination present in the peptide mixture. Employing DIA analysis, the contamination levels of the light peptides were found to be below 7%, confirming the reliability and integrity of the peptide mixture, as observed in Figure 65. The detailed percentage of contamination for each individual peptide can be found in the supplementary information section, providing a comprehensive overview of the contamination levels across the peptide set.

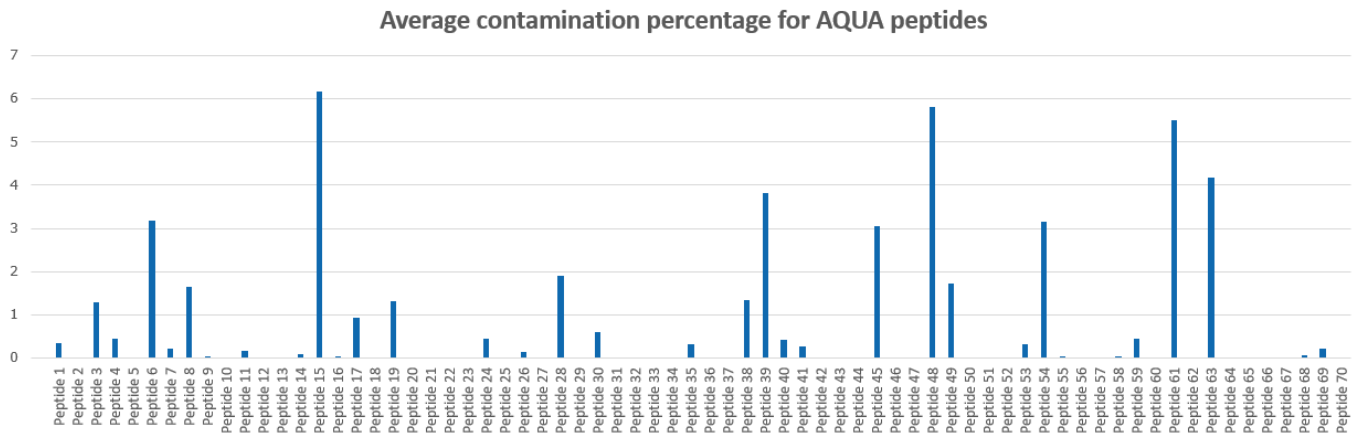


Figure 65: Histogram showing the average percentage of light peptide contamination in the AQUA heavy-labeled peptide mix across the 70 analyzed peptides.

Figure 66 illustrates the identification and quantification of high-risk HCPs in the downstream processing samples, employing both the ANAQUANT ReadyBeads and the AQUA heavy-labeled peptide mixture.

The ReadyBeads method detected four high-risk HCPs, with one being uniquely identified by this method. This reflects its capacity to capture a broad spectrum of HCPs present in the sample. In contrast, the AQUA heavy-labeled peptide mixture quantified a more targeted subset of high-risk HCPs, identifying twelve such contaminants, nine of which were detected exclusively by the AQUA method. This comparison highlights the complementary strengths of the two methods, while the ReadyBeads provide a general overview of the entire HCP population; the AQUA peptide-based method offers more focused insights into those specific HCPs that pose the greatest risk to product safety and efficacy.

High risk HCPs	Quantified with ANAQUANT ReadyBeads	Quantified with the AQUA peptide mix
G3I8R9		✓
G3I0W1		✓
G3I5A4		✓
G3H0L9		✓
G3INC5	✓	✓
Q9EPP7	✓	✓
G3I3Y6		✓
G3H6V7		✓
Q9JKY1	✓	✓
G3IIE7		✓
G3HC31		✓
G3H3Q1		✓
G3I6T1	✓	

Figure 66: High-risk HCPs quantified through different approaches in the downstream process cascade. On the left, a table lists the high-risk HCPs detected across different purification stages, detailing the quantification approach used for each (either AQUA heavy-labeled peptides or ANAQUANT

ReadyBeads). On the right, a Venn diagram visualizes the overlap between the high-risk HCPs quantified by the two methods, illustrating both the shared and unique protein groups identified by each technique.

The Venn diagram clearly further demonstrates the complementary nature of these two quantification strategies. It shows that, although both methods share some common high-risk HCPs, each provides unique information that the other does not. This overlap between the two techniques validates the use of both approaches, ensuring that both the broad population of HCPs and the high-risk contaminants are adequately monitored. By employing these methods in tandem, we would be able achieve a more comprehensive understanding of the HCP profile in the sample, which is crucial for ensuring the purity of the therapeutic product.

The analysis also revealed significant differences in the sensitivity of the two methods, particularly when examining the three HCPs that were quantified by both techniques. As illustrated in Figure 67, the AQUA heavy-labeled peptide mixture detected higher ppm concentrations of these HCPs across various stages of the purification process. This suggests that the AQUA approach is more sensitive, especially for detecting HCPs that may be present in trace amounts.

Additionally, the results indicated that the depletion of certain high-risk HCPs occurs at different stages of the process, depending on the quantification method used. The ReadyBeads method detected certain HCPs slightly earlier on in the purification cascade, while the AQUA method identified them at later stages, emphasizing the added value of using both methods to track the removal of these contaminants throughout the entire process.

One key example of the enhanced sensitivity of the AQUA peptide mixture is the detection of the high-risk HCP G9JKY1 in the TMAE step of the purification process. As shown in Figure 67, G9JKY1 was detected using the AQUA peptides mixture but was not identified in the TMAE step by the ReadyBeads method. This again, highlights the superior sensitivity of the AQUA heavy-labeled peptides in detecting low-abundance high-risk HCPs that may remain in the sample, even at later stages of purification, where the protein quantity is very scarce.

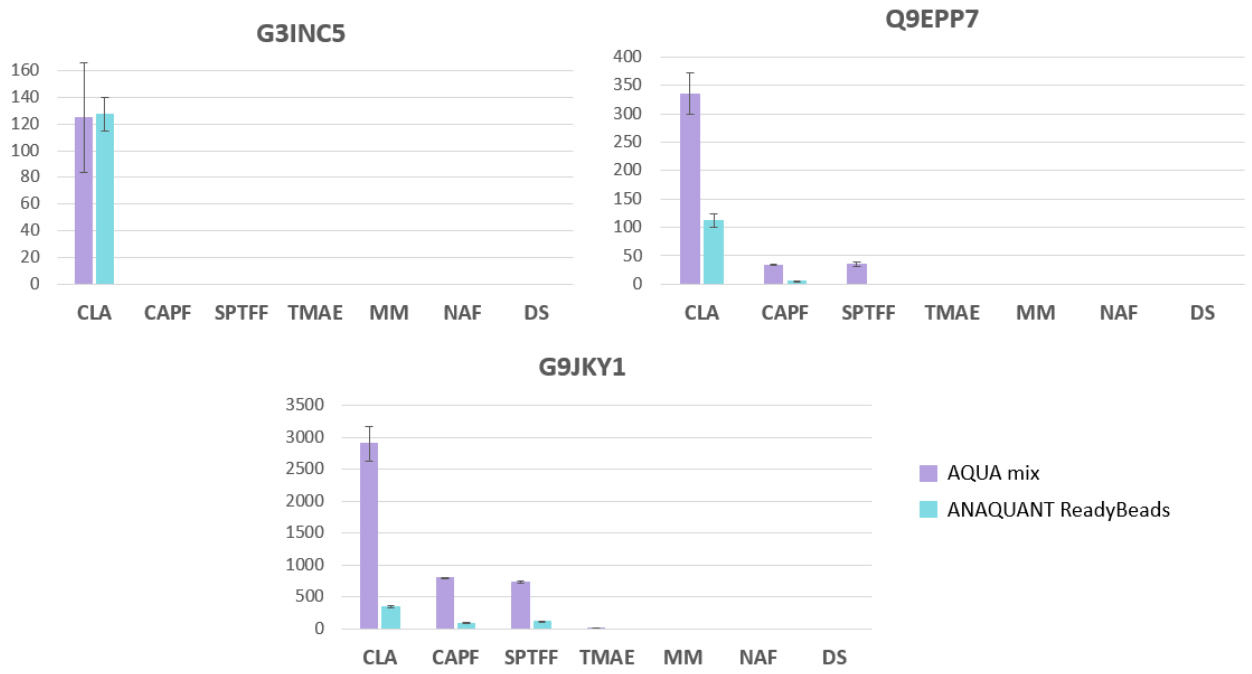


Figure 67: Quantification of three high-risk HCPs across the DP cascade using AQUA heavy-labeled peptide mixture and ANAQUANT ReadyBeads. This figure presents three histograms, each representing one of the three high-risk HCPs quantified using both AQUA heavy-labeled peptide mixture and ANAQUANT ReadyBeads. The diagrams show the ppm levels of each HCP across the downstream process cascade, from the clarified harveststage to the final drug substance stage. The reduction in the abundance of these high-risk HCPs at each stage of the purification process is clearly depicted, with a marked decrease observed as the process advances, culminating in the depletion of these HCPs by the final purification stage.

B. Identification of critical points in the Downstream Process

Identifying the critical steps in the purification process is essential for optimizing the removal of high-risk HCPs throughout the cascade. Both the ReadyBeads and AQUA heavy-labeled peptide mixture methods provide valuable insights, each highlighting different stages as particularly significant in ensuring the purity of the final product.

In the context of the ReadyBeads method, the TMAE chromatography step is identified as the critical stage for high-risk HCP depletion. This step is particularly significant as it leads to the complete elimination of the high-risk HCP population, making it a key process in ensuring the final product's purity.

On the other hand, the AQUA heavy-labeled peptide mixture method highlights two critical stages for high-risk HCP depletion: CAPF and TMAE. While the TMAE step in this method also contributes to the complete depletion of high-risk HCPs, it is the CAPF step that shows a significant initial reduction in the concentration of these contaminants. This reduction is notably observed as a large drop in ppm levels for the high-risk HCPs quantified by the AQUA peptides, underscoring the significant impact of CAPF in reducing HCPs at an early stage of the purification process.

Thus, while the ReadyBeads method emphasizes TMAE as the sole critical step for HCP depletion, the AQUA heavy-labeled peptide mixture method reveals the importance of both the CAPF and TMAE steps, with CAPF serving as a key stage for early reduction and TMAE ensuring the final elimination of high-risk HCPs.

Complete depletion bar diagrams for both the AQUA peptides and ReadyBeads methods, illustrating the extent of HCP removal at each step of the purification process, can be found in the supplementary information section.

C. Conclusions on the methods employed and their potential implications for Quality Control

The integration of both the ReadyBeads and AQUA heavy-labeled peptide mixture methods provides a powerful, multifaceted approach for the quantification of high-risk HCPs during the purification process. While the ReadyBeads method offers broad, unbiased insights into the entire spectrum of HCPs, the AQUA heavy-labeled peptide mixture enables precise targeting of specific high-risk contaminants.

This combination of broad-spectrum and targeted quantification tools has proven invaluable for achieving a comprehensive risk assessment. By leveraging the strengths of each technique, it is possible to ensure that the high-risk HCPs are effectively monitored and eliminated throughout the purification stages.

These advancements in HCP quantification are not just relevant for understanding the purification process but also for their significant implications for quality control in biopharmaceutical manufacturing. By enhancing sensitivity, the methods discussed enable the detection and management of even trace amounts of high-risk HCPs, ensuring that the final product adheres to stringent safety and regulatory standards. As regulatory bodies continue to emphasize the need for high quality therapeutic products, these techniques offer a promising path toward achieving both safety and compliance.

Moving forward, it is crucial to acknowledge that the list of high-risk HCPs and the detection methods used to quantify them will continue to evolve. As our understanding of these contaminants expands and detection technologies advance, the inclusion of additional high-risk HCPs in future assessments will be essential. Together, these strategies will continue to play a critical role in ensuring the production of biologics that meet the highest standards of quality, efficacy, and safety.

Conclusions

The primary objective of this project, developing an optimized immuno-capture process to enable an immuno-affinity mass spectrometry-based method for assessing and verifying ELISA test results, was successfully achieved. Through extensive experimentation and systematic optimization during my first PhD year, we identified PhyTips Protein A as the most suitable immuno-capture support, coupled with a detailed and robust workflow. This optimized approach enables a more efficient and effective capture of HCPs meeting the needs for reproducibility and adaptability to various stages of the drug development process.

The journey toward this outcome involved the exploration of multiple supports, including columns, magnetic beads, and plate-based formats. Among these, the column initially emerged as the most promising option due to its compatibility with the immuno-capture process and its capacity to yield successful outcomes. A comprehensive series of parameter optimization experiments followed, targeting the critical factors of reagent ratios, temperature, incubation time, and reaction volume. These experiments revealed that a 1:1 ratio of anti-HCP antibody to antigen, incubated at 37°C for 2 hours, provided the most favorable results, as validated by MS analysis.

While the column demonstrated suitability for immuno-capture experiments, its complexity and the considerable time, effort, and material demands made performing triplicates more challenging. This prompted the search for a more efficient and streamlined alternative, leading to the selection of PhyTips Protein A. PhyTips proved to be a more suitable protocol for performing triplicates, as it significantly reduced the experimental workflow and processing time, making it much easier to achieve reproducible results. Although triplicates could have been performed with the column-based method, the higher resource demands and complexity made it far less practical. With PhyTips, we were able to perform triplicates with greater ease, ensuring consistent and reliable results. Furthermore, we evaluated and compared two distinct workflows for PhyTips, the direct and indirect approaches, and found the indirect approach to be the most effective, yielding consistent and repeatable results across the triplicates.The mass spectrometry analysis of these experiments provided encouraging data. For the first time, we achieved complete overlap in the protein group identification across the antigen, unbound, and elution fractions. The consistency observed across replicates emphasized the repeatability and robustness of the optimized workflow. Additionally, the MS data facilitated a focused assessment of the unbound fraction, which represents the non-immunoreactive HCP population, critical for evaluating potential high-risk HCPs. Notably, one particular high-risk HCP (G3HKV9) was consistently detected in the unbound fraction across all replicates, while remaining absent from the elution fraction. This observation underscores the success of the method in isolating immuno-reactive and non-immunoreactive populations, allowing for a precise assessment of HCPs that may pose risks during drug development or the downstream process.

In summary, the first part of my PhD project successfully developed an immuno-capture MS-based method capable of verifying ELISA results with confidence. The method is adaptable for various stages of the drug development process, ranging from early candidate selection to downstream processing. Moreover, its potential for further automation, enabled by supports like PhyTips Protein A, positions it as a time-efficient and scalable alternative for modern biopharmaceutical workflows. These

advancements mark a significant milestone in the project, fulfilling its first objective and laying a strong foundation for subsequent studies in the biopharmaceutical environment at Merck.

The second objective of my PhD, centered on evaluating and improving HCP management during the downstream processing of a commercially available mAb, was also successfully achieved. By employing a range of complementary MS-based methodologies, this work provides a comprehensive and detailed characterization of HCP dynamics across the purification cascade, reinforcing the importance of orthogonal techniques in ensuring product safety and quality.

One of the critical achievements of this was the implementation of different types of internal standards with the objective to improve the precision and accuracy of HCPs' quantification. On the one hand, we have implemented a mixture of standard peptides offering the advantage of covering a large dynamic range of concentration, the ANAQUANT ReadyBeads. Those peptides were used for profiling the total HCP population across different DSP stages. This broad-spectrum method allowed for a quantitative assessment of HCP clearance, revealing distinct impurity profiles and identifying purification stages with significant reductions in HCP levels. Importantly, this approach demonstrated higher sensitivity than traditional ELISA, detecting elevated ppm levels of residual HCPs, particularly in the final drug substance. These findings underscore the limitations of the conventional method and highlight the need for advanced MS techniques to fully understand HCP behavior, particularly at critical purification steps.

In parallel, we conducted a comparative analysis of library-based and library-free MS strategies to further refine our understanding of HCP profiles. The library-free method proved advantageous for detecting low-abundance HCPs, especially in the later stages of the DSP, where sample complexity increases due to an exacerbated dynamic range between the mAb and remaining HCPs. This dual analysis offered a detailed perspective, highlighting the strengths of both approaches in achieving a comprehensive HCP assessment. The library-free method, in particular, proved invaluable for identifying contaminants that might evade detection through ELISA or even library-based MS, underscoring its importance in improving the quantification of impurities during monitoring.

As a second focal point, we investigated on the targeted quantification of high-risk HCPs, a subset of impurities with potential immunogenicity or adverse effects on drug stability and efficacy. This was accomplished by combining the ANAQUANT ReadyBeads method, which profiled the entire HCP population, with a mixture of well-chosen heavy-labeled AQUA peptides, which provided high accuracy in quantifying high-risk HCPs present in the DSP samples. The AQUA method revealed elevated ppm levels of high-risk HCPs compared to ANAQUANT, particularly in the intermediate stages of purification, and demonstrated its higher sensitivity in detecting these contaminants.

By leveraging these complementary approaches, we identified critical purification steps that significantly impacted both general and high-risk HCP levels. The Protein A chromatography (CAPF) step, for instance, was shown to reduce a substantial proportion of total HCPs early in the DSP cascade, while the TMAE chromatography step emerged as the definitive stage for achieving complete

depletion of high-risk HCPs. This two-step clearance process underscores the importance of integrating broad profiling and targeted quantification methods to fully assess the effectiveness of DSP workflows.

The results obtained through these advanced analytical techniques not only validated the effectiveness of the purification process but also provided key insights into improving impurity management strategies. By meticulously tracking the clearance of high-risk HCPs and general impurities throughout the DSP, we confirmed that the final DS met stringent safety and efficacy standards. Moreover, the ability to combine broad-spectrum and targeted HCP quantification methods strengthened the overall risk assessment framework, ensuring a more robust evaluation of the purification cascade.

This study section highlights the critical advantages of MS-based methods over traditional ELISA, particularly in their ability to offer both comprehensive and precise HCP assessments. It also demonstrates how combining different HCP quantification methods and complementary analytical approaches can improve HCP monitoring in biopharmaceutical manufacturing.

In conclusion, the findings of this study demonstrate the value of integrating ANAQUANT ReadyBeads and AQUA heavy-labeled peptides for both broad and targeted HCP quantification. These complementary methods not only improve our understanding of impurity clearance across the DSP thus contributing to the development of safer, higher-quality therapeutic products. As biopharmaceutical manufacturing continues to advance, the incorporation of such sophisticated analytical techniques could be critical in meeting evolving regulatory requirements and ensuring the production of effective and reliable biologics.

Perspectives

In exploring the immuno-capture approach, one potential avenue to further support the findings would be the implementation of a biotin-streptavidin system, a well-established technique that has demonstrated efficacy in several studies. This approach could provide a more detailed understanding of the interactions between the anti-HCP antibody mixture and the antigen-coupled components. Biotin-streptavidin systems have proven to enhance capture efficiency due to the strong affinity between biotin and streptavidin, which could potentially facilitate the identification of a wider range of HCPs. Although this experiment was considered, it was not executed within the scope of this study due to time limitations. However, the addition of this assay could yield valuable insights into the dynamics of HCP-antibody interactions, further validating the immuno-capture approach.

Another aspect that warrants further investigation is the application of DSP cascade analysis in conjunction with a comparative spectral library. A CLA harvest spectral library, constructed using a specific cell line, could provide a more comprehensive identification of HCPs in DSP samples. This library could then be compared to that of a null cell line, often considered a general representation of HCPs.[237–239] The use of the null cell line as a benchmark is common practice, but it is important to acknowledge that it may not capture the full spectrum of HCPs present in the DSP samples. The null cell line library tends to lack some of the specific proteins that might be unique to the DSP harvest, indicating that a broader spectral analysis could lead to a more accurate and complete profiling of HCPs.

Despite the widespread use of null cell line libraries, it is essential to recognize the limitations of this approach. The null cell line is only a surrogate representation of HCPs, and the overlap between the proteins present in the null cell line and those in DSP samples is not always complete. This discrepancy highlights the importance of considering the unique profile of each DSP sample when identifying HCPs. While the null cell line serves as a useful reference, it may fail to detect specific proteins that are uniquely expressed or enriched in the DSP process, suggesting that alternative approaches or additional libraries could enhance the sensitivity and accuracy of HCP detection.

Moreover, the presence of the mAb in the harvest introduces another layer of complexity. The mAb, while essential for the capture process, could potentially interfere with the identification of HCPs due to its high abundance and the possibility of cross-reactivity. The presence of mAb in the DSP samples must therefore be carefully considered when conducting HCP analysis. Special precautions may be needed to ensure that the presence of the mAb does not skew the results or mask the detection of low-abundance HCPs. This issue underscores the need for thoughtful experimental design and a thorough understanding of the potential confounding factors when analyzing HCPs in the presence of monoclonal antibodies.

Finally, as discussed throughout this manuscript, the analysis of HCPs is inherently challenging due to the diversity of host cell proteins, the extreme dynamic range aimed to be covered (6 orders to cover trace-level HCPs present in the ppm range) and the complexities of the DSP process. The strategies proposed, including the use of biotin-streptavidin systems and the expansion of spectral libraries, offer promising avenues for improving the detection and characterization of HCPs. However, it is important

to recognize the risks associated with such experiments, particularly when working with complex samples like DSP harvests. The decision-making process in HCP analysis must be informed by a comprehensive understanding of these variables, as well as by the inherent limitations of current methodologies. Future work in this area should focus on refining these techniques and addressing the challenges posed by mAb interference and the incomplete overlap of null cell line libraries with DSP samples.

List of Communications

List of oral communications and posters

Noelia de Lama, Diego Bertaccini, Christine Carapito, “Development of an orthogonal method for Host Cell Protein characterization via Immunoaffinity enrichment and MS Analysis”. ProteoAix 2023, the 3rd Joint Meeting of Spanish, French, and Portuguese Proteomics Societies, Aix-en-Provence, France, June 20-23, 2023. (Selected oral communication)

Noelia de Lama, Diego Bertaccini, Christine Carapito, “Innovative orthogonal method for Host Cell Protein characterization combining Immuno-affinity enrichment and MS Analysis”, International American Society of Mass Spectrometry (ASMS) conference, Houston, USA, June 3-7, 2024 (Selected poster communication).

Noelia de Lama, Corentin Beaumal, Jae Choi, Bhavin Patel, Nikki Jarrett, Kay Opperman, Christine Carapito, “Precise and accurate quantitation of critical host cell proteins (HCPs) using a targeted peptide mixture and LC-MS/MS analysis”, American Society of Mass Spectrometry (ASMS) conference, Houston, USA, June 3-7, 2024 (Selected poster communication).

Other presentations

Noelia de Lama, Diego Bertaccini, Christine Carapito, “Development of orthogonal methods for Host Cell Protein characterization via Immunoaffinity enrichment and MS Analysis”, Process Related Impurities virtual BioForum, Merck Guidonia, Rome, November 13-14, 2023. (Oral communication)

Annexe

Supporting Information

Supplementary Figure 1

Detailed Windows from the DIA method employed throughout the project

Windows	Start (m/z)	End (m/z)
1	399,4314	418,4401
2	418,4401	437,4487
3	436,4483	455,4569
4	454,4565	473,4651
5	473,4651	492,4737
6	491,4733	510,4819
7	510,4819	529,4906
8	528,4901	547,4987
9	546,4983	565,5069
10	565,5069	584,5156
11	583,5151	602,5238
12	602,5238	621,5324
13	620,5319	639,5406
14	638,5401	657,5488
15	657,5488	676,5574
16	675,557	694,5656
17	694,5656	713,5742
18	712,5738	731,5824
19	730,582	749,5906
20	749,5906	768,5992
21	767,5988	786,6074
22	786,6074	805,6161
23	804,6156	823,6243
24	822,6238	841,6324
25	841,6324	860,6411
26	859,6406	878,6493
27	878,6493	897,6579
28	896,6575	915,6661
29	914,6656	933,6743
30	933,6743	952,6829

Supplementary Figure 2

Table depicting the Light/Heavy ratios for the AQUA peptide during the contamination assessment, their average contamination percentage as well as the fmol quantity to deduct in order to obtain the correct quantity of peptide present in the sample

AQUA Peptide	Light/Heavy 1	Light/Heavy 2	Light/Heavy 3	Average Contamination percentage	fmol quantity to deduct
Peptide 1	0,19	0,86	0,00	0,35	0,12
Peptide 2	0,00	0,00	0,00	0,00	0,00
Peptide 3	1,23	1,22	1,44	1,30	0,46
Peptide 4	0,49	0,18	0,70	0,46	0,16
Peptide 5	0,00	0,00	0,00	0,00	0,00
Peptide 6	3,15	3,14	3,23	3,17	1,13
Peptide 7	0,21	0,22	0,20	0,21	0,07
Peptide 8	1,67	1,69	1,64	1,66	0,59
Peptide 9	0,03	0,04	0,03	0,03	0,01
Peptide 10	0,00	0,01	0,00	0,01	0,00
Peptide 11	0,21	0,15	0,18	0,18	0,06
Peptide 12	0,00	0,00	0,00	0,00	0,00
Peptide 13	0,00	0,00	0,00	0,00	0,00
Peptide 14	0,28	0,00	0,00	0,09	0,03
Peptide 15	6,23	6,31	5,93	6,16	2,19
Peptide 16	0,04	0,03	0,04	0,03	0,01
Peptide 17	0,97	0,95	0,90	0,94	0,33
Peptide 18	0,02	0,00	0,00	0,01	0,00
Peptide 19	1,20	1,42	1,33	1,32	0,47
Peptide 20	0,00	0,00	0,01	0,00	0,00

Peptide 21	0,00	0,00	0,00	0,00	0,00
Peptide 22	0,00	0,00	0,00	0,00	0,00
Peptide 23	0,00	0,00	0,00	0,00	0,00
Peptide 24	0,45	0,45	0,43	0,44	0,16
Peptide 25	0,00	0,00	0,00	0,00	0,00
Peptide 26	0,13	0,13	0,16	0,14	0,05
Peptide 27	0,00	0,00	0,00	0,00	0,00
Peptide 28	2,03	2,08	1,62	1,91	0,68
Peptide 29	0,00	0,00	0,00	0,00	0,00
Peptide 30	0,61	0,56	0,60	0,59	0,21
Peptide 31	0,00	0,00	0,00	0,00	0,00
Peptide 32	0,00	0,00	0,00	0,00	0,00
Peptide 33	0,00	0,00	0,00	0,00	0,00
Peptide 34	0,00	0,00	0,00	0,00	0,00
Peptide 35	0,49	0,07	0,37	0,31	0,11
Peptide 36	0,00	0,00	0,00	0,00	0,00
Peptide 37	0,01	0,02	0,00	0,01	0,00
Peptide 38	1,30	1,24	1,45	1,33	0,47
Peptide 39	6,87	3,52	1,07	3,82	1,36
Peptide 40	0,33	0,36	0,59	0,43	0,15
Peptide 41	0,24	0,32	0,26	0,28	0,10
Peptide 42	0,00	0,00	0,00	0,00	0,00
Peptide 43	0,00	0,00	0,00	0,00	0,00
Peptide 44	0,00	0,00	0,00	0,00	0,00

Peptide 45	2,91	2,93	3,27	3,04	1,08
Peptide 46	0,00	0,00	0,00	0,00	0,00
Peptide 47	0,00	0,00	0,00	0,00	0,00
Peptide 48	5,97	5,70	5,74	5,80	2,06
Peptide 49	1,46	1,75	1,95	1,72	0,61
Peptide 50	0,00	0,00	0,00	0,00	0,00
Peptide 51	0,00	0,00	0,00	0,00	0,00
Peptide 52	0,01	0,01	0,00	0,01	0,00
Peptide 53	0,36	0,28	0,36	0,33	0,12
Peptide 54	3,19	3,21	3,09	3,16	1,12
Peptide 55	0,07	0,04	0,00	0,04	0,01
Peptide 56	0,00	0,00	0,00	0,00	0,00
Peptide 57	0,00	0,00	0,00	0,00	0,00
Peptide 58	0,05	0,06	0,00	0,04	0,01
Peptide 59	0,46	0,45	0,47	0,46	0,16
Peptide 60	0,00	0,00	0,00	0,00	0,00
Peptide 61	5,48	5,48	5,58	5,51	1,96
Peptide 62	0,00	0,00	0,00	0,00	0,00
Peptide 63	6,05	6,47	0,00	4,17	1,48
Peptide 64	0,00	0,00	0,00	0,00	0,00
Peptide 65	0,01	0,01	0,01	0,01	0,00
Peptide 66	0,00	0,00	0,00	0,00	0,00
Peptide 67	0,00	0,00	0,00	0,00	0,00
Peptide 68	0,15	0,04	0,00	0,06	0,02

Peptide 69	0,41	0,22	0,00	0,21	0,08
Peptide 70	0,00	0,00	0,00	0,00	0,00

References

(1) Alejandra, W.-P.; Miriam Irene, J.-P.; Fabio Antonio, G.-S.; Patricia, R.-G. R.; Elizabeth, T.-A.; Aleman-Aguilar, J. P.; Rebeca, G.-V. Production of Monoclonal Antibodies for Therapeutic Purposes: A Review. *Int. Immunopharmacol.* **2023**, *120*, 110376. https://doi.org/10.1016/j.intimp.2023.110376.

(2) Yang, F.; Li, D.; Kufer, R.; Cadang, L.; Zhang, J.; Dai, L.; Guo, J.; Wohlrab, S.; Greenwood-Goodwin, M.; Shen, A.; Duan, D.; Li, H.; Yuk, I. H. Versatile LC–MS-Based Workflow with Robust 0.1 Ppm Sensitivity for Identifying Residual HCPs in Biotherapeutic Products. *Anal. Chem.* **2022**, *94* (2), 723–731. https://doi.org/10.1021/acs.analchem.1c03095.

(3) Shukla, A. A.; Jiang, C.; Ma, J.; Rubacha, M.; Flansburg, L.; Lee, S. S. Demonstration of Robust Host Cell Protein Clearance in Biopharmaceutical Downstream Processes. *Biotechnol. Prog.* **2008**, *24* (3), 615–622. https://doi.org/10.1021/bp070396j.

(4) 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. *Biotechnol. Bioeng.* **2014**, *111* (12), 2367–2379. https://doi.org/10.1002/bit.25327.

(5) 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. *Biotechnol. Bioeng.* **2010**, *105* (2), 306–316. https://doi.org/10.1002/bit.22532.

(6) Searle, B. C.; Swearingen, K. E.; Barnes, C. A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating High Quality Libraries for DIA MS with Empirically Corrected Peptide Predictions. *Nat. Commun.* **2020**, *11* (1), 1548. https://doi.org/10.1038/s41467-020-15346-1.

(7) Jones, M.; Palackal, N.; Wang, F.; Gaza-Bulseco, G.; Hurkmans, K.; Zhao, Y.; Chitikila, C.; Clavier, S.; Liu, S.; Menesale, E.; Schonenbach, N. S.; Sharma, S.; Valax, P.; Waerner, T.; Zhang, L.; Connolly, T. "High-risk" Host Cell Proteins (HCPs): A Multi-company Collaborative View. *Biotechnol. Bioeng.* **2021**, *118* (8), 2870–2885. https://doi.org/10.1002/bit.27808.

(8) Pythoud, N.; Bons, J.; Mijola, G.; Beck, A.; Cianférani, 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. *J. Proteome Res.* **2021**, *20* (1), 923–931. https://doi.org/10.1021/acs.jproteome.0c00664.

(9) Beaumal, C.; Beck, A.; Hernandez-Alba, O.; Carapito, C. Advanced Mass Spectrometry Workflows for Accurate Quantification of Trace-level Host Cell Proteins in Drug Products: Benefits of FAIMS Separation and Gas-phase Fractionation DIA. *PROTEOMICS* **2023**, *23* (16), 2300172. https://doi.org/10.1002/pmic.202300172.

(10) Chen, I.-H.; Xiao, H.; Daly, T.; Li, N. Improved Host Cell Protein Analysis in Monoclonal Antibody Products through Molecular Weight Cutoff Enrichment. *Anal. Chem.* **2020**, *92* (5), 3751–3757. https://doi.org/10.1021/acs.analchem.9b05081.

(11) Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J.-C.; Yan, J. X.; Gooley, Andrew. 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 Arnino Acid Analysis. *Nat. Biotechnol.* **1996**, *14* (1), 61–65. https://doi.org/10.1038/nbt0196-61.

(12) Anderson, N. L.; Anderson, N. G. Proteome and Proteomics: New Technologies, New Concepts, and New Words. *ELECTROPHORESIS* **1998**, *19* (11), 1853–1861. https://doi.org/10.1002/elps.1150191103.

(13) Graves, P. R.; Haystead, T. A. J. Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Biol. Rev.* **2002**, *66* (1), 39–63. https://doi.org/10.1128/MMBR.66.1.39-63.2002.

(14) James, P. Protein Identification in the Post-Genome Era: The Rapid Rise of Proteomics. *Q. Rev. Biophys.* **1997**, *30* (4), 279–331. https://doi.org/10.1017/S0033583597003399.

(15) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394. https://doi.org/10.1021/cr3003533.

(16) Sinha, A.; Mann, M. A Beginner's Guide to Mass Spectrometry–Based Proteomics. *The Biochemist* **2020**, *42* (5), 64–69. https://doi.org/10.1042/BIO20200057.

(17) Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537* (7620), 347–355. https://doi.org/10.1038/nature19949.

(18) 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. https://doi.org/10.1002/jms.3932.

(19) 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.* **2016**, *9* (1), 449–472. https://doi.org/10.1146/annurev-anchem-071015-041535.

(20) Karas, Michael.; Hillenkamp, Franz. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, *60* (20), 2299–2301. https://doi.org/10.1021/ac00171a028.

(21) 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. https://doi.org/10.1126/science.2675315.

(22) Nesvizhskii, A. I.; Aebersold, R. Interpretation of Shotgun Proteomic Data. *Mol. Cell. Proteomics* **2005**, *4* (10), 1419–1440. https://doi.org/10.1074/mcp.R500012-MCP200.

(23) 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. *Mol. Cell. Proteomics* **2014**, *13* (1), 339–347. https://doi.org/10.1074/mcp.M113.034769.

(24) Han, X.; Aslanian, A.; Yates, J. R. Mass Spectrometry for Proteomics. *Curr. Opin. Chem. Biol.* **2008**, *12* (5), 483–490. https://doi.org/10.1016/j.cbpa.2008.07.024.

(25) 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. *Mol. Cell. Proteomics* **2013**, *12* (12), 3465–3473. https://doi.org/10.1074/mcp.M113.030114.

(26) 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. https://doi.org/10.1021/acs.analchem.8b00693.

(27) Patrie, S. M. Top-Down Mass Spectrometry: Proteomics to Proteoforms. In *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*; Mirzaei, H., Carrasco, M., Eds.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, 2016; Vol. 919, pp 171–200. https://doi.org/10.1007/978-3-319-41448-5_8.

(28) Cupp-Sutton, K. A.; Wu, S. High-Throughput Quantitative Top-down Proteomics. *Mol. Omics* **2020**, *16* (2), 91–99. https://doi.org/10.1039/C9MO00154A.

(29) 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* (S1), 20–27. https://doi.org/10.1002/rcm.8255.

(30) 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. *Mol. Cell. Proteomics* **2019**, *18* (4), 796–805. https://doi.org/10.1074/mcp.RA118.000993.

(31) Cristobal, A.; Marino, F.; Post, H.; Van Den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89* (6), 3318–3325. https://doi.org/10.1021/acs.analchem.6b03756.

(32) Valkevich, E. M.; Sanchez, N. A.; Ge, Y.; Strieter, E. R. Middle-Down Mass Spectrometry Enables Characterization of Branched Ubiquitin Chains. *Biochemistry* **2014**, *53* (30), 4979–4989. https://doi.org/10.1021/bi5006305.

(33) Channaveerappa, D.; Ngounou Wetie, A. G.; Darie, C. C. Bottlenecks in Proteomics: An Update. In *Advancements of Mass Spectrometry in Biomedical Research*; Woods, A. G., Darie, C. C., Eds.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, 2019; Vol. 1140, pp 753–769. https://doi.org/10.1007/978-3-030-15950-4_45.

(34) Rogers, J. C.; Bomgarden, R. D. Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. In *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*; Mirzaei, H., Carrasco, M., Eds.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, 2016; Vol. 919, pp 43–62. https://doi.org/10.1007/978-3-319-41448-5_3.

(35) Bose, U.; Wijffels, G.; Howitt, C. A.; Colgrave, M. L. Proteomics: Tools of the Trade. In *Emerging Sample Treatments in Proteomics*; Capelo-Martínez, J.-L., Ed.; Advances in Experimental Medicine and

Biology; Springer International Publishing: Cham, 2019; Vol. 1073, pp 1–22. https://doi.org/10.1007/978-3-030-12298-0_1.

(36) Feist, P.; Hummon, A. Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples. *Int. J. Mol. Sci.* **2015**, *16* (2), 3537–3563. https://doi.org/10.3390/ijms16023537.

(37) 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), 1700011. https://doi.org/10.1002/pmic.201700011.

(38) 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), 2441–2451. https://doi.org/10.1021/pr201169q.

(39) International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431* (7011), 931–945. https://doi.org/10.1038/nature03001.

(40) Wang, D.; Eraslan, B.; Wieland, T.; Hallström, B.; Hopf, T.; Zolg, D. P.; Zecha, J.; Asplund, A.; Li, L.; 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. https://doi.org/10.15252/msb.20188503.

(41) Zubarev, R. A. The Challenge of the Proteome Dynamic Range and Its Implications for In-Depth Proteomics. *PROTEOMICS* **2013**, *13* (5), 723–726. https://doi.org/10.1002/pmic.201200451.

(42) 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. https://doi.org/10.1016/j.jprot.2016.04.002.

(43) Salvato, F.; Gallo De Carvalho, M. C. D. C.; Lima Leite, A. D. Strategies for Protein Separation. In *Integrative Proteomics*; Leung, H.-C., Ed.; InTech, 2012. https://doi.org/10.5772/29363.

(44) 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* (10), 757. https://doi.org/10.15252/msb.20145625.

(45) 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), 319–324. https://doi.org/10.1038/nmeth.2834.

(46) Zougman, A.; Selby, P. J.; Banks, R. E. Suspension Trapping (STrap) Sample Preparation Method for Bottom-up Proteomics Analysis. *PROTEOMICS* **2014**, *14* (9), 1006–1000. https://doi.org/10.1002/pmic.201300553.

(47) Varnavides, G.; Madern, M.; Anrather, D.; Hartl, N.; Reiter, W.; Hartl, M. In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods. *J. Proteome Res.* **2022**, *21* (10), 2397–2411. https://doi.org/10.1021/acs.jproteome.2c00265.

(48) Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **1970**, *227* (5259), 680–685. https://doi.org/10.1038/227680a0.

(49) 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), 816–843. https://doi.org/10.1002/mas.20204.

(50) Lu, X.; Zhu, H. Tube-Gel Digestion. *Mol. Cell. Proteomics* **2005**, *4* (12), 1948–1958. https://doi.org/10.1074/mcp.M500138-MCP200.

(51) Muller, L.; Fornecker, L.; Van Dorsselaer, A.; Cianférani, 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. https://doi.org/10.1002/pmic.201600288.

(52) Rabilloud, T.; Lelong, C. Two-Dimensional Gel Electrophoresis in Proteomics: A Tutorial. *J. Proteomics* **2011**, *74* (10), 1829–1841. https://doi.org/10.1016/j.jprot.2011.05.040.

(53) 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), 1327–1333. https://doi.org/10.1002/elps.200305844.

(54) Pasquali, M.; Serchi, T.; Planchon, S.; Renaut, J. 2D-DIGE in Proteomics. In *Functional Genomics*; Kaufmann, M., Klinger, C., Savelsbergh, A., Eds.; Methods in Molecular Biology; Springer New York: New York, NY, 2017; Vol. 1654, pp 245–254. https://doi.org/10.1007/978-1-4939-7231-9_17.

(55) Burkhart, J. M.; Schumbrutzki, 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), 1454–1462. https://doi.org/10.1016/j.jprot.2011.11.016.

(56) Vandermarliere, E.; Mueller, M.; Martens, L. Getting Intimate with Trypsin, the Leading Protease in Proteomics: TRYPSIN IN PROTEOMICS. *Mass Spectrom. Rev.* **2013**, *32* (6), 453–465. https://doi.org/10.1002/mas.21376.

(57) Glatter, T.; Ludwig, C.; Ahrné, E.; Aebersold, R.; Heck, A. J. R.; 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), 5145–5156. https://doi.org/10.1021/pr300273g.

(58) Hakobyan, A.; Schneider, M. B.; Liesack, W.; Glatter, T. Efficient Tandem LysC/Trypsin Digestion in Detergent Conditions. *PROTEOMICS* **2019**, *19* (20), 1900136. https://doi.org/10.1002/pmic.201900136.

(59) Olshina, M. A.; Sharon, M. Mass Spectrometry: A Technique of Many Faces. *Q. Rev. Biophys.* **2016**, *49*, e18. https://doi.org/10.1017/S0033583516000160.

(60) 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/.

(61) 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. https://doi.org/10.1016/j.chroma.2014.12.035.

(62) Xie, F.; Smith, R. D.; Shen, Y. Advanced Proteomic Liquid Chromatography. *J. Chromatogr. A* **2012**, *1261*, 78–90. https://doi.org/10.1016/j.chroma.2012.06.098.

(63) Zhang, Z.; Wu, S.; Stenoien, D. L.; Paša-Tolić, L. High-Throughput Proteomics. *Annu. Rev. Anal. Chem.* **2014**, *7* (1), 427–454. https://doi.org/10.1146/annurev-anchem-071213-020216.

(64) Moruz, L.; Käll, L. Peptide Retention Time Prediction. *Mass Spectrom. Rev.* **2017**, *36* (5), 615–623. https://doi.org/10.1002/mas.21488.

(65) Steen, H.; Mann, M. The Abc's (and Xyz's) of Peptide Sequencing. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (9), 699–711. https://doi.org/10.1038/nrm1468.

(66) 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. *J. Am. Soc. Mass Spectrom.* **1996**, *7* (6), 532–540. https://doi.org/10.1016/1044-0305(96)00057-8.

(67) 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. *J. Proteome Res.* **2011**, *10* (4), 1785–1793. https://doi.org/10.1021/pr101060v.

(68) 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), 701–714. https://doi.org/10.1038/nprot.2015.040.

(69) Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A. Fundamentals and Advances of Orbitrap Mass Spectrometry. In *Encyclopedia of Analytical Chemistry*; Meyers, R. A., Ed.; Wiley, 2019; pp 1–40. https://doi.org/10.1002/9780470027318.a9309.pub2.

(70) 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. https://doi.org/10.1016/j.jprot.2013.02.023.

(71) Marx, V. Targeted Proteomics. *Nat. Methods* **2013**, *10* (1), 19–22. https://doi.org/10.1038/nmeth.2285.

(72) Picotti, P.; Aebersold, R. Selected Reaction Monitoring–Based Proteomics: Workflows, Potential, Pitfalls and Future Directions. *Nat. Methods* **2012**, *9* (6), 555–566. https://doi.org/10.1038/nmeth.2015.

(73) Bourmaud, A.; Gallien, S.; Domon, B. Parallel Reaction Monitoring Using quadrupole-Orbitrap Mass Spectrometer: Principle and Applications. *PROTEOMICS* **2016**, *16* (15–16), 2146–2159. https://doi.org/10.1002/pmic.201500543.

(74) Zhang, H.; Liu, Q.; Zimmerman, L. J.; Ham, A.-J. L.; Slebos, R. J. C.; Rahman, J.; Kikuchi, T.; Massion, P. P.; Carbone, D. P.; Billheimer, D.; Liebler, D. C. Methods for Peptide and Protein Quantitation by Liquid Chromatography-Multiple Reaction Monitoring Mass Spectrometry. *Mol. Cell. Proteomics* **2011**, *10* (6), M110.006593. https://doi.org/10.1074/mcp.M110.006593.

(75) Sleno, L.; Volmer, D. A. Ion Activation Methods for Tandem Mass Spectrometry. *J. Mass Spectrom.* **2004**, *39* (10), 1091–1112. https://doi.org/10.1002/jms.703.

(76) 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), 709–712. https://doi.org/10.1038/nmeth1060.

(77) Syka, J. E. P.; 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.* **2004**, *101* (26), 9528–9533. https://doi.org/10.1073/pnas.0402700101.

(78) Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R. Influence of Basic Residue Content on Fragment Ion Peak Intensities in Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Anal. Chem.* **2004**, *76* (5), 1243–1248. https://doi.org/10.1021/ac0351163.

(79) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406. https://doi.org/10.1002/1096-9888(200012)35:12<1399::AID-JMS86>3.0.CO;2-R.

(80) Biemann, K. Appendix 5. Nomenclature for Peptide Fragment Ions (Positive Ions). In *Methods in Enzymology*; Elsevier, 1990; Vol. 193, pp 886–887. https://doi.org/10.1016/0076-6879(90)93460-3.

(81) Blueggel, M.; Chamrad, D.; Meyer, H. Bioinformatics in Proteomics. *Curr. Pharm. Biotechnol.* **2004**, *5* (1), 79–88. https://doi.org/10.2174/1389201043489648.

(82) 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), 1794–1805. https://doi.org/10.1021/pr101065j.

(83) 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. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

(84) 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), 958–964. https://doi.org/10.1021/pr0499491.

(85) 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. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2.

(86) Craig, R.; Cortens, J. P.; Beavis, R. C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J. Proteome Res.* **2004**, *3* (6), 1234–1242. https://doi.org/10.1021/pr049882h.

(87) Griss, J.; Perez-Riverol, Y.; Lewis, S.; Tabb, D. L.; Dianes, J. A.; del-Toro, N.; Rurik, M.; Walzer, M.; Kohlbacher, O.; Hermjakob, H.; Wang, R.; Vizcaíno, J. A. Recognizing Millions of Consistently Unidentified Spectra across Hundreds of Shotgun Proteomics Datasets. *Nat. Methods* **2016**, *13* (8), 651–656. https://doi.org/10.1038/nmeth.3902.

(88) 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), 743–749. https://doi.org/10.1038/nbt.3267.

(89) Skinner, O. S.; Kelleher, N. L. Illuminating the Dark Matter of Shotgun Proteomics. *Nat. Biotechnol.* **2015**, *33* (7), 717–718. https://doi.org/10.1038/nbt.3287.

(90) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, 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–D745. https://doi.org/10.1093/nar/gkv1189.

(91) Armengaud, J. A Perfect Genome Annotation Is within Reach with the Proteomics and Genomics Alliance. *Curr. Opin. Microbiol.* **2009**, *12* (3), 292–300. https://doi.org/10.1016/j.mib.2009.03.005.

(92) The UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506–D515. https://doi.org/10.1093/nar/gky1049.

(93) González-Gomariz, J.; Guruceaga, E.; López-Sánchez, M.; Segura, V. Proteogenomics in the Context of the Human Proteome Project (HPP). *Expert Rev. Proteomics* **2019**, *16* (3), 267–275. https://doi.org/10.1080/14789450.2019.1571916.

(94) Nesvizhskii, A. I. Proteogenomics: Concepts, Applications and Computational Strategies. *Nat. Methods* **2014**, *11* (11), 1114–1125. https://doi.org/10.1038/nmeth.3144.

(95) 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. https://doi.org/10.1002/pmic.200300511.

(96) 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* (S24), 669. https://doi.org/10.1186/s12859-019-3253-z.

(97) 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. *Biotechnol. Bioeng.* **2012**, *109* (6), 1386–1394. https://doi.org/10.1002/bit.24476.

(98) 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), 207–214. https://doi.org/10.1038/nmeth1019.

(99) Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-Supervised Learning for Peptide Identification from Shotgun Proteomics Datasets. *Nat. Methods* **2007**, *4* (11), 923–925. https://doi.org/10.1038/nmeth1113.

(100) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative Mass Spectrometry in Proteomics: A Critical Review. *Anal. Bioanal. Chem.* **2007**, *389* (4), 1017–1031. https://doi.org/10.1007/s00216-007-1486-6.

(101) Schubert, O. T.; Röst, 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. https://doi.org/10.1038/nprot.2017.040.

(102) Ankney, J. A.; Muneer, A.; Chen, X. Relative and Absolute Quantitation in Mass Spectrometry–Based Proteomics. *Annu. Rev. Anal. Chem.* **2018**, *11* (1), 49–77. https://doi.org/10.1146/annurev-anchem-061516-045357.

(103) Ong, S.-E.; Mann, M. Mass Spectrometry–Based Proteomics Turns Quantitative. *Nat. Chem. Biol.* **2005**, *1* (5), 252–262. https://doi.org/10.1038/nchembio736.

(104) Blein-Nicolas, M.; Zivy, M. Thousand and One Ways to Quantify and Compare Protein Abundances in Label-Free Bottom-up Proteomics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2016**, *1864* (8), 883–895. https://doi.org/10.1016/j.bbapap.2016.02.019.

(105) Ramus, C.; Hovasse, A.; Marcellin, M.; Hesse, A.-M.; Mouton-Barbosa, E.; Bouyssié, D.; Vaca, S.; Carapito, C.; Chaoui, K.; Bruley, C.; Garin, J.; Cianférani, S.; Ferro, M.; Van Dorssaeler, A.; Burlet-Schiltz, O.; Schaeffer, C.; Couté, 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. https://doi.org/10.1016/j.jprot.2015.11.011.

(106) Pino, L. K.; Just, S. C.; MacCoss, M. J.; Searle, B. C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol. Cell. Proteomics* **2020**, *19* (7), 1088–1103. https://doi.org/10.1074/mcp.P119.001913.

(107) Smith, R.; Tostengard, A. R. Quantitative Evaluation of Ion Chromatogram Extraction Algorithms. *J. Proteome Res.* **2020**, *19* (5), 1953–1964. https://doi.org/10.1021/acs.jproteome.9b00768.

(108) 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.* **2020**, *39* (3), 229–244. https://doi.org/10.1002/mas.21540.

(109) Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13* (9), 2513–2526. https://doi.org/10.1074/mcp.M113.031591.

(110) Bouyssié, D.; Hesse, A.-M.; Mouton-Barbosa, E.; Rompais, M.; Macron, C.; Carapito, C.; Gonzalez De Peredo, A.; Couté, Y.; Dupierris, V.; Burel, A.; Menetrey, J.-P.; Kalaitzakis, A.; Poisat, J.; Romdhani, A.; Burlet-Schiltz, O.; Cianférani, S.; Garin, J.; Bruley, C. Proline: An Efficient and User-Friendly Software Suite for Large-Scale Proteomics. *Bioinformatics* **2020**, *36* (10), 3148–3155. https://doi.org/10.1093/bioinformatics/btaa118.

(111) 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. https://doi.org/10.1002/mas.21595.

(112) 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. *Mol. Cell. Proteomics* **2002**, *1* (5), 376–386. https://doi.org/10.1074/mcp.M200025-MCP200.

(113) Emadali, A.; Gallagher-Gambarelli, M. La Protéomique Quantitative Par La Méthode SILAC: Technique et Perspectives. *médecine/sciences* **2009**, *25* (10), 835–842. https://doi.org/10.1051/medsci/20092510835.

(114) 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), 383–385. https://doi.org/10.1038/nmeth.1446.

(115) Merrill, A. E.; Hebert, A. S.; MacGilvray, 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. *Mol. Cell. Proteomics* **2014**, *13* (9), 2503–2512. https://doi.org/10.1074/mcp.M114.040287.

(116) 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), 994–999. https://doi.org/10.1038/13690.

(117) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895–1904. https://doi.org/10.1021/ac0262560.

(118) 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), 3651–3660. https://doi.org/10.1002/pmic.200700316.

(119) 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. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–1169. https://doi.org/10.1074/mcp.M400129-MCP200.

(120) Rappsilber, J.; Ryder, U.; Lamond, A. I.; Mann, M. Large-Scale Proteomic Analysis of the Human Spliceosome. *Genome Res.* **2002**, *12* (8), 1231–1245. https://doi.org/10.1101/gr.473902.

(121) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Correction: Corrigendum: Global Quantification of Mammalian Gene Expression Control. *Nature* **2013**, *495* (7439), 126–127. https://doi.org/10.1038/nature11848.

(122) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; 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), 761–776. https://doi.org/10.1021/pr9006365.

(123) Borràs, E.; Sabidó, E. What Is Targeted Proteomics? A Concise Revision of Targeted Acquisition and Targeted Data Analysis in Mass Spectrometry. *PROTEOMICS* **2017**, *17* (17–18), 1700180. https://doi.org/10.1002/pmic.201700180.

(124) 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. *Mol. Cell. Proteomics* **2012**, *11* (11), 1475–1488. https://doi.org/10.1074/mcp.O112.020131.

(125) Vidova, V.; Spacil, Z. A Review on Mass Spectrometry-Based Quantitative Proteomics: Targeted and Data Independent Acquisition. *Anal. Chim. Acta* **2017**, *964*, 7–23. https://doi.org/10.1016/j.aca.2017.01.059.

(126) 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), 28566–28581. https://doi.org/10.3390/ijms161226120.

(127) 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.* **2003**, *100* (12), 6940–6945. https://doi.org/10.1073/pnas.0832254100.

(128) 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), 587–589. https://doi.org/10.1038/nmeth774.

(129) Brun, V.; Dupuis, A.; Adrait, A.; Marcellin, M.; Thomas, D.; Court, M.; Vandenesch, F.; Garin, J. Isotope-Labeled Protein Standards. *Mol. Cell. Proteomics* **2007**, *6* (12), 2139–2149. https://doi.org/10.1074/mcp.M700163-MCP200.

(130) 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), 2201–2210. https://doi.org/10.1021/pr800654s.

(131) 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. *Mol. Cell. Proteomics* **2012**, *11* (3), O111.009613. https://doi.org/10.1074/mcp.O111.009613.

(132) Kettenbach, A. N.; Rush, J.; Gerber, S. A. Absolute Quantification of Protein and Post-Translational Modification Abundance with Stable Isotope–Labeled Synthetic Peptides. *Nat. Protoc.* **2011**, *6* (2), 175–186. https://doi.org/10.1038/nprot.2010.196.

(133) Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P. The Absolute Quantification Strategy: A General Procedure for the Quantification of Proteins and Post-Translational Modifications. *Methods* **2005**, *35* (3), 265–273. https://doi.org/10.1016/j.ymeth.2004.08.018.

(134) 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. https://doi.org/10.1002/pmic.201900276.

(135) 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), 1918–1924. https://doi.org/10.1021/ac991133+.

(136) 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. https://doi.org/10.1038/nmeth705.

(137) Purvine, S.; Eppel*, J.; Yi, E. C.; Goodlett, D. R. Shotgun Collision-induced Dissociation of Peptides Using a Time of Flight Mass Analyzer. *PROTEOMICS* **2003**, *3* (6), 847–850. https://doi.org/10.1002/pmic.200300362.

(138) 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), 2187–2200. https://doi.org/10.1021/ac048455k.

(139) 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), 6481–6488. https://doi.org/10.1021/ac900888s.

(140) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-Ion Fragmentation. *Mol. Cell. Proteomics* **2010**, *9* (10), 2252–2261. https://doi.org/10.1074/mcp.M110.001537.

(141) Carvalho, P. C.; Han, X.; Xu, T.; Cociorva, D.; Carvalho, M. D. G.; Barbosa, V. C.; Yates, J. R. XDIA: Improving on the Label-Free Data-Independent Analysis. *Bioinformatics* **2010**, *26* (6), 847–848. https://doi.org/10.1093/bioinformatics/btq031.

(142) 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), 1621–1632. https://doi.org/10.1021/pr2008175.

(143) 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. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717. https://doi.org/10.1074/mcp.O111.016717.

(144) Geromanos, S. J.; Hughes, C.; Ciavarini, S.; Vissers, J. P. C.; Langridge, J. I. Using Ion Purity Scores for Enhancing Quantitative Accuracy and Precision in Complex Proteomics Samples. *Anal. Bioanal. Chem.* **2012**, *404* (4), 1127–1139. https://doi.org/10.1007/s00216-012-6197-y.

(145) 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), 744–746. https://doi.org/10.1038/nmeth.2528.

(146) 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), 5415–5430. https://doi.org/10.1021/pr5003017.

(147) 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. https://doi.org/10.1002/pmic.201500450.

(148) 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), 167–170. https://doi.org/10.1038/nmeth.2767.

(149) 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), 4359–4371. https://doi.org/10.1021/acs.jproteome.5b00543.

(150) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Miladinović, 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. *Mol. Cell. Proteomics* **2015**, *14* (5), 1400–1410. https://doi.org/10.1074/mcp.M114.044305.

(151) 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. https://doi.org/10.1021/acs.jproteome.7b00464.

(152) 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. *Nat. Methods* **2018**, *15* (6), 440–448. https://doi.org/10.1038/s41592-018-0003-5.

(153) Messner, C. B.; Demichev, V.; Bloomfield, N.; Yu, J. S. L.; White, M.; Kreidl, M.; Egger, A.-S.; Freiwald, A.; Ivosev, G.; Wasim, F.; Zelezniak, A.; Jürgens, L.; Suttorp, N.; Sander, L. E.; Kurth, F.; Lilley, K. S.; Mülleder, M.; Tate, S.; Ralser, M. Ultra-Fast Proteomics with Scanning SWATH. *Nat. Biotechnol.* **2021**, *39* (7), 846–854. https://doi.org/10.1038/s41587-021-00860-4.

(154) 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. https://doi.org/10.1038/s41592-020-00998-0.

(155) 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. *J. Proteome Res.* **2021**, *20* (1), 279–288. https://doi.org/10.1021/acs.jproteome.0c00381.

(156) Bekker-Jensen, D. B.; Martínez-Val, A.; Steigerwald, S.; Rüther, 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. *Mol. Cell. Proteomics* **2020**, *19* (4), 716–729. https://doi.org/10.1074/mcp.TIR119.001906.

(157) 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. https://doi.org/10.15252/msb.20178126.

(158) Bekker-Jensen, D. B.; Bernhardt, O. M.; Hogrebe, A.; Martinez-Val, A.; Verbeke, L.; Gandhi, T.; Kelstrup, C. D.; Reiter, L.; Olsen, J. V. Rapid and Site-Specific Deep Phosphoproteome Profiling by Data-

Independent Acquisition without the Need for Spectral Libraries. *Nat. Commun.* **2020**, *11* (1), 787. https://doi.org/10.1038/s41467-020-14609-1.

(159) Szyrwiel, L.; Sinn, L.; Ralser, M.; Demichev, V. Slice-PASEF: Fragmenting All Ions for Maximum Sensitivity in Proteomics. Biochemistry October 31, 2022. https://doi.org/10.1101/2022.10.31.514544.

(160) Salovska, B.; Li, W.; Di, Y.; Liu, Y. BoxCarmax: A High-Selectivity Data-Independent Acquisition Mass Spectrometry Method for the Analysis of Protein Turnover and Complex Samples. *Anal. Chem.* **2021**, *93* (6), 3103–3111. https://doi.org/10.1021/acs.analchem.0c04293.

(161) Pino, L. K.; Just, S. C.; MacCoss, M. J.; Searle, B. C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol. Cell. Proteomics* **2020**, *19* (7), 1088–1103. https://doi.org/10.1074/mcp.P119.001913.

(162) Hu, A.; Noble, W. S.; Wolf-Yadlin, A. Technical Advances in Proteomics: New Developments in Data-Independent Acquisition. *F1000Research* **2016**, *5*, 419. https://doi.org/10.12688/f1000research.7042.1.

(163) 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. *Mol. Cell. Proteomics* **2015**, *14* (9), 2301–2307. https://doi.org/10.1074/mcp.O114.047035.

(164) Rosenberger, G.; Koh, C. C.; Guo, T.; Röst, 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* (1), 140031. https://doi.org/10.1038/sdata.2014.31.

(165) Desiere, F. The PeptideAtlas Project. *Nucleic Acids Res.* **2006**, *34* (90001), D655–D658. https://doi.org/10.1093/nar/gkj040.

(166) 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.; Vizcaíno, J. A. The ProteomeXchange Consortium in 2017: Supporting the Cultural Change in Proteomics Public Data Deposition. *Nucleic Acids Res.* **2017**, *45* (D1), D1100–D1106. https://doi.org/10.1093/nar/gkw936.

(167) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nat. Biotechnol.* **2014**, *32* (3), 219–223. https://doi.org/10.1038/nbt.2841.

(168) 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. https://doi.org/10.1038/nmeth.4390.

(169) Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry. *Nat. Commun.* **2018**, *9* (1), 5128. https://doi.org/10.1038/s41467-018-07454-w.

(170) Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndl, M.; Brant, A.; Cimermancic, P.; Cox, J. High-Quality MS/MS Spectrum Prediction for Data-Dependent and Data-Independent Acquisition Data Analysis. *Nat. Methods* **2019**, *16* (6), 519–525. https://doi.org/10.1038/s41592-019-0427-6.

(171) 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. Prosit: Proteome-Wide Prediction of Peptide Tandem Mass Spectra by Deep Learning. *Nat. Methods* **2019**, *16* (6), 509–518. https://doi.org/10.1038/s41592-019-0426-7.

(172) 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. https://doi.org/10.1021/acs.analchem.7b02566.

(173) Bouwmeester, R.; Gabriels, R.; Hulstaert, N.; Martens, L.; Degroeve, S. DeepLC Can Predict Retention Times for Peptides That Carry As-yet Unseen Modifications. *Nat. Methods* **2021**, *18* (11), 1363–1369. https://doi.org/10.1038/s41592-021-01301-5.

(174) Degroeve, S.; Martens, L. MS2PIP: A Tool for MS/MS Peak Intensity Prediction. *Bioinformatics* **2013**, *29* (24), 3199–3203. https://doi.org/10.1093/bioinformatics/btt544.

(175) Van Puyvelde, B.; Willems, S.; Gabriels, R.; Daled, S.; De Clerck, L.; Vande Casteele, S.; Staes, A.; Impens, F.; Deforce, D.; Martens, L.; Degroeve, S.; Dhaenens, M. Removing the Hidden Data Dependency of DIA with Predicted Spectral Libraries. *PROTEOMICS* **2020**, *20* (3–4), 1900306. https://doi.org/10.1002/pmic.201900306.

(176) 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. *Nat. Methods* **2020**, *17* (1), 41–44. https://doi.org/10.1038/s41592-019-0638-x.

(177) Dübel, S. Therapeutic Antibodies - From Past to Future. In *Handbook of Therapeutic Antibodies*; Dübel, S., Ed.; Wiley, 2007; pp 2–16. https://doi.org/10.1002/9783527619740.ch1.

(178) Freysd'ottir, J. Production of Monoclonal Antibodies. In *Diagnostic and Therapeutic Antibodies*; George, A. J. T., Urch, C. E., Eds.; Methods in Molecular Medicine; Humana Press: Totowa, NJ, 2000; Vol. 40, pp 267–279. https://doi.org/10.1385/1-59259-076-4:267.

(179) Weiner, L. M.; Surana, R.; Wang, S. Monoclonal Antibodies: Versatile Platforms for Cancer Immunotherapy. *Nat. Rev. Immunol.* **2010**, *10* (5), 317–327. https://doi.org/10.1038/nri2744.

(180) 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. https://doi.org/10.1293/tox.2015-0031.

(181) Reichert, J. M.; Valge-Archer, V. E. Development Trends for Monoclonal Antibody Cancer Therapeutics. *Nat. Rev. Drug Discov.* **2007**, *6* (5), 349–356. https://doi.org/10.1038/nrd2241.

(182) Moldenhauer, G. Selection Strategies I: Monoclonal Antibodies. In *Handbook of Therapeutic Antibodies*; Dübel, S., Ed.; Wiley, 2007; pp 18–44. https://doi.org/10.1002/9783527619740.ch2.

(183) Frenzel, A.; Hust, M.; Schirrmann, T. Expression of Recombinant Antibodies. *Front. Immunol.* **2013**, *4*. https://doi.org/10.3389/fimmu.2013.00217.

(184) Dumont, J.; Euwart, D.; Mei, B.; Estes, S.; Kshirsagar, R. Human Cell Lines for Biopharmaceutical Manufacturing: History, Status, and Future Perspectives. *Crit. Rev. Biotechnol.* **2016**, *36* (6), 1110–1122. https://doi.org/10.3109/07388551.2015.1084266.

(185) Li, F.; Vijayasankaran, N.; Shen, A. (Yijuan); Kiss, R.; Amanullah, A. Cell Culture Processes for Monoclonal Antibody Production. *mAbs* **2010**, *2* (5), 466–479. https://doi.org/10.4161/mabs.2.5.12720.

(186) Okamura, K.; Badr, S.; Murakami, S.; Sugiyama, H. Hybrid Modeling of CHO Cell Cultivation in Monoclonal Antibody Production with an Impurity Generation Module. *Ind. Eng. Chem. Res.* **2022**, *61* (40), 14898–14909. https://doi.org/10.1021/acs.iecr.2c00736.

(187) Kelley, B. Industrialization of mAb Production Technology: The Bioprocessing Industry at a Crossroads. *mAbs* **2009**, *1* (5), 443–452. https://doi.org/10.4161/mabs.1.5.9448.

(188) Wurm, F. M. Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells. *Nat. Biotechnol.* **2004**, *22* (11), 1393–1398. https://doi.org/10.1038/nbt1026.

(189) 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. *Biotechnol. J.* **2019**, *14* (3), 1700686. https://doi.org/10.1002/biot.201700686.

(190) Badr, S.; Sugiyama, H. A PSE Perspective for the Efficient Production of Monoclonal Antibodies: Integration of Process, Cell, and Product Design Aspects. *Curr. Opin. Chem. Eng.* **2020**, *27*, 121–128. https://doi.org/10.1016/j.coche.2020.01.003.

(191) Kunert, R.; Reinhart, D. Advances in Recombinant Antibody Manufacturing. *Appl. Microbiol. Biotechnol.* **2016**, *100* (8), 3451–3461. https://doi.org/10.1007/s00253-016-7388-9.

(192) Shukla, A. A.; Hubbard, B.; Tressel, T.; Guhan, S.; Low, D. Downstream Processing of Monoclonal Antibodies—Application of Platform Approaches. *J. Chromatogr. B* **2007**, *848* (1), 28–39. https://doi.org/10.1016/j.jchromb.2006.09.026.

(193) Liu, H. F.; Ma, J.; Winter, C.; Bayer, R. Recovery and Purification Process Development for Monoclonal Antibody Production. *mAbs* **2010**, *2* (5), 480–499. https://doi.org/10.4161/mabs.2.5.12645.

(194) Somasundaram, B.; Pleitt, K.; Shave, E.; Baker, K.; Lua, L. H. L. Progression of Continuous Downstream Processing of Monoclonal Antibodies: Current Trends and Challenges. *Biotechnol. Bioeng.* **2018**, *115* (12), 2893–2907. https://doi.org/10.1002/bit.26812.

(195) Chahar, D. S.; Ravindran, S.; Pisal, S. S. Monoclonal Antibody Purification and Its Progression to Commercial Scale. *Biologicals* **2020**, *63*, 1–13. https://doi.org/10.1016/j.biologicals.2019.09.007.

(196) Hogwood, C. E.; Bracewell, D. G.; Smales, C. M. Measurement and Control of Host Cell Proteins (HCPs) in CHO Cell Bioprocesses. *Curr. Opin. Biotechnol.* **2014**, *30*, 153–160. https://doi.org/10.1016/j.copbio.2014.06.017.

(197) Baumann, P.; Hubbuch, J. Downstream Process Development Strategies for Effective Bioprocesses: Trends, Progress, and Combinatorial Approaches. *Eng. Life Sci.* **2017**, *17* (11), 1142–1158. https://doi.org/10.1002/elsc.201600033.

(198) Molden, R.; Hu, M.; Yen E, S.; Saggese, D.; Reilly, J.; Mattila, J.; Qiu, H.; Chen, G.; Bak, H.; Li, N. Host Cell Protein Profiling of Commercial Therapeutic Protein Drugs as a Benchmark for Monoclonal Antibody-Based Therapeutic Protein Development. *mAbs* **2021**, *13* (1), 1955811. https://doi.org/10.1080/19420862.2021.1955811.

(199) Chen, I.-H.; Xiao, H.; Daly, T.; Li, N. Improved Host Cell Protein Analysis in Monoclonal Antibody Products through Molecular Weight Cutoff Enrichment. *Anal. Chem.* **2020**, *92* (5), 3751–3757. https://doi.org/10.1021/acs.analchem.9b05081.

(200) 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. *Anal. Chem.* **2020**, *92* (15), 10327–10335. https://doi.org/10.1021/acs.analchem.0c00360.

(201) 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), 10283–10291. https://doi.org/10.1021/acs.analchem.5b02103.

(202) Thomson, A. S.; Mai, S.; Byrne, M. P. A Novel Approach to Characterize Host Cell Proteins Associated with Therapeutic Monoclonal Antibodies. *Biotechnol. Bioeng.* **2017**, *114* (6), 1208–1214. https://doi.org/10.1002/bit.26256.

(203) 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. *Biotechnol. Prog.* **2020**, *36* (4), e2983. https://doi.org/10.1002/btpr.2983.

(204) Guo, J.; Kufer, R.; Li, D.; Wohlrab, S.; Greenwood-Goodwin, M.; Yang, F. Technical Advancement and Practical Considerations of LC-MS/MS-Based Methods for Host Cell Protein Identification and Quantitation to Support Process Development. *mAbs* **2023**, *15* (1), 2213365. https://doi.org/10.1080/19420862.2023.2213365.

(205) Gunawan, F.; Nishihara, J.; Liu, P.; Sandoval, W.; Vanderlaan, M.; Zhang, H.; Krawitz, D. Comparison of Platform Host Cell Protein ELISA to Process-specific Host Cell Protein ELISA. *Biotechnol. Bioeng.* **2018**, *115* (2), 382–389. https://doi.org/10.1002/bit.26466.

(206) Liu, X.; Chen, Y.; Zhao, Y.; Liu-Compton, V.; Chen, W.; Payne, G.; Lazar, A. C. Identification and Characterization of Co-Purifying CHO Host Cell Proteins in Monoclonal Antibody Purification Process. *J. Pharm. Biomed. Anal.* **2019**, *174*, 500–508. https://doi.org/10.1016/j.jpba.2019.06.021.

(207) Tscheliessnig, A. L.; Konrath, J.; Bates, R.; Jungbauer, A. Host Cell Protein Analysis in Therapeutic Protein Bioprocessing – Methods and Applications. *Biotechnol. J.* **2013**, *8* (6), 655–670. https://doi.org/10.1002/biot.201200018.

(208) Giordano, E.; Liori, B.; Cecchini, I.; Verani, R.; Leone, L. In-House CHO HCPs Platform: A Promising Approach for HCPs ELISA Monitoring. *Eur. J. Pharm. Sci.* **2024**, *192*, 106656. https://doi.org/10.1016/j.ejps.2023.106656.

(209) Goey, C. H.; Alhuthali, S.; Kontoravdi, C. Host Cell Protein Removal from Biopharmaceutical Preparations: Towards the Implementation of Quality by Design. *Biotechnol. Adv.* **2018**, *36* (4), 1223–1237. https://doi.org/10.1016/j.biotechadv.2018.03.021.

(210) Pilely, K.; Johansen, M. R.; Lund, R. R.; Kofoed, T.; Jørgensen, T. K.; Skriver, L.; Mørtz, E. Monitoring Process-Related Impurities in Biologics–Host Cell Protein Analysis. *Anal. Bioanal. Chem.* **2022**, *414* (2), 747–758. https://doi.org/10.1007/s00216-021-03648-2.

(211) Flatman, S.; Alam, I.; Gerard, J.; Mussa, N. Process Analytics for Purification of Monoclonal Antibodies. *J. Chromatogr. B* **2007**, *848* (1), 79–87. https://doi.org/10.1016/j.jchromb.2006.11.018.

(212) Eaton, L. C. Host Cell Contaminant Protein Assay Development for Recombinant Biopharmaceuticals. *J. Chromatogr. A* **1995**, *705* (1), 105–114. https://doi.org/10.1016/0021-9673(94)01249-E.

(213) Wan, M.; Wang, Y.; Rabideau, S.; Moreadith, R.; Schrimsher, J.; Conn, G. An Enzyme-Linked Immunosorbent Assay for Host Cell Protein Contaminants in Recombinant PEGylated Staphylokinase Mutant SY161. *J. Pharm. Biomed. Anal.* **2002**, *28* (5), 953–963. https://doi.org/10.1016/S0731-7085(01)00712-9.

(214) Berkelman, T.; Harbers, A.; Bandhakavi, S. 2-D Western Blotting for Evaluation of Antibodies Developed for Detection of Host Cell Protein. In *Proteomic Profiling*; Posch, A., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2015; Vol. 1295, pp 393–414. https://doi.org/10.1007/978-1-4939-2550-6_28.

(215) Yang, P.-C.; Mahmood, T. Western Blot: Technique, Theory, and Trouble Shooting. *North Am. J. Med. Sci.* **2012**, *4* (9), 429. https://doi.org/10.4103/1947-2714.100998.

(216) Posch, A.; Kollmann, F.; Berkelman, T.; Dreskin, E. Sample Preparation of Secreted Mammalian Host Cell Proteins and Their Characterization by Two-Dimensional Electrophoresis and Western Blotting. In *Proteomic Profiling*; Posch, A., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2021; Vol. 2261, pp 507–524. https://doi.org/10.1007/978-1-0716-1186-9_32.

(217) Waldera-Lupa, D. M.; Jasper, Y.; Köhne, P.; Schwichtenhövel, R.; Falkenberg, H.; Flad, T.; Happersberger, P.; Reisinger, B.; Dehghani, A.; Moussa, R.; Waerner, T. Host Cell Protein Detection Gap Risk Mitigation: Quantitative IAC-MS for ELISA Antibody Reagent Coverage Determination. *mAbs* **2021**, *13* (1), 1955432. https://doi.org/10.1080/19420862.2021.1955432.

(218) Kiyonami, R.; Melani, R.; Chen, Y.; Leon, A. D.; Du, M. Applying UHPLC-HRAM MS/MS Method to Assess Host Cell Protein Clearance during the Purification Process Development of Therapeutic mAbs. *Int. J. Mol. Sci.* **2024**, *25* (17), 9687. https://doi.org/10.3390/ijms25179687.

(219) Li, J.; Smith, L. S.; Zhu, H.-J. Data-Independent Acquisition (DIA): An Emerging Proteomics Technology for Analysis of Drug-Metabolizing Enzymes and Transporters. *Drug Discov. Today Technol.* **2021**, *39*, 49–56. https://doi.org/10.1016/j.ddtec.2021.06.006.

(220) Ji, Q.; Sokolowska, I.; Cao, R.; Jiang, Y.; Mo, J.; Hu, P. A Highly Sensitive and Robust LC-MS Platform for Host Cell Protein Characterization in Biotherapeutics. *Biologicals* **2023**, *82*, 101675. https://doi.org/10.1016/j.biologicals.2023.101675.

(221) Carvalho, S. B.; Profit, L.; Krishnan, S.; Gomes, R. A.; Alexandre, B. M.; Clavier, S.; Hoffman, M.; Brower, K.; Gomes-Alves, P. SWATH-MS as a Strategy for CHO Host Cell Protein Identification and Quantification Supporting the Characterization of mAb Purification Platforms. *J. Biotechnol.* **2024**, *384*, 1–11. https://doi.org/10.1016/j.jbiotec.2024.02.001.

(222) Panapitakkul, C.; Bulaon, C. J. I.; Pisuttinusart, N.; Phoolcharoen, W. Characterization of Host Cell Proteins in the Downstream Process of Plant-Based Biologics Using LC-MS Profiling. *Biotechnol. Rep.* **2024**, *44*, e00856. https://doi.org/10.1016/j.btre.2024.e00856.

(223) Pillai-Kastoori, L.; Heaton, S.; Shiflett, S. D.; Roberts, A. C.; Solache, A.; Schutz-Geschwender, A. R. Antibody Validation for Western Blot: By the User, for the User. *J. Biol. Chem.* **2020**, *295* (4), 926–939. https://doi.org/10.1016/S0021-9258(17)49905-4.

(224) Friedman, D. B. Quantitative Proteomics for Two-Dimensional Gels Using Difference Gel Electrophoresis. In *Mass Spectrometry Data Analysis in Proteomics*; Humana Press: New Jersey, 2006; Vol. 367, pp 219–240. https://doi.org/10.1385/1-59745-275-0:219.

(225) Marouga, R.; David, S.; Hawkins, E. The Development of the DIGE System: 2D Fluorescence Difference Gel Analysis Technology. *Anal. Bioanal. Chem.* **2005**, *382* (3), 669–678. https://doi.org/10.1007/s00216-005-3126-3.

(226) Hober, S.; Nord, K.; Linhult, M. Protein A Chromatography for Antibody Purification. *J. Chromatogr. B* **2007**, *848* (1), 40–47. https://doi.org/10.1016/j.jchromb.2006.09.030.

(227) Panyutich, A. V.; Baturevich, E. A.; Kolesnikova, T. S.; Ganz, T. The Effect of Biotinylation on the Antigenic Specificity of Anti-Defensin Monoclonal Antibodies. *J. Immunol. Methods* **1993**, *158* (2), 237–242. https://doi.org/10.1016/0022-1759(93)90219-W.

(228) Reverberi, R.; Reverberi, L. Factors Affecting the Antigen-Antibody Reaction. *Blood Transfus.* **2007**. https://doi.org/10.2450/2007.0047-07.

(229) Fischer, J.; Kaufmann, J. O.; Weller, M. G. Simple Determination of Affinity Constants of Antibodies by Competitive Immunoassays. *Methods Protoc.* **2024**, *7* (3), 49. https://doi.org/10.3390/mps7030049.

(230) Searle, B. C.; Swearingen, K. E.; Barnes, C. A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating High Quality Libraries for DIA MS with Empirically Corrected Peptide Predictions. *Nat. Commun.* **2020**, *11* (1), 1548. https://doi.org/10.1038/s41467-020-15346-1.

(231) Silva, J. C.; Gorenstein, M. V.; Li, G.-Z.; Vissers, J. P. C.; Geromanos, S. J. Absolute Quantification of Proteins by LCMSE. *Mol. Cell. Proteomics* **2006**, *5* (1), 144–156. https://doi.org/10.1074/mcp.M500230-MCP200.

(232) Al Qaraghuli, M. M.; Kubiak-Ossowska, K.; Ferro, V. A.; Mulheran, P. A. Antibody-Protein Binding and Conformational Changes: Identifying Allosteric Signalling Pathways to Engineer a Better Effector Response. *Sci. Rep.* **2020**, *10* (1), 13696. https://doi.org/10.1038/s41598-020-70680-0.

(233) Rabia, L. A.; Desai, A. A.; Jhajj, H. S.; Tessier, P. M. Understanding and Overcoming Trade-Offs between Antibody Affinity, Specificity, Stability and Solubility. *Biochem. Eng. J.* **2018**, *137*, 365–374. https://doi.org/10.1016/j.bej.2018.06.003.

(234) 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. *Anal. Chem.* **2020**, *92* (1), 1007–1015. https://doi.org/10.1021/acs.analchem.9b03952.

(235) Beaumal, C.; Beck, A.; Hernandez-Alba, O.; Carapito, C. Advanced Mass Spectrometry Workflows for Accurate Quantification of Trace-level Host Cell Proteins in Drug Products: Benefits of FAIMS Separation and Gas-phase Fractionation DIA. *PROTEOMICS* **2023**, *23* (16), 2300172. https://doi.org/10.1002/pmic.202300172.

(236) Pythoud, N.; Bons, J.; Mijola, G.; Beck, A.; Cianférani, 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. *J. Proteome Res.* **2021**, *20* (1), 923–931. https://doi.org/10.1021/acs.jproteome.0c00664.

(237) 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. https://doi.org/10.1021/acs.analchem.7b00304.

(238) Tait, A. S.; Hogwood, C. E. M.; Smales, C. M.; Bracewell, D. G. Host Cell Protein Dynamics in the Supernatant of a mAb Producing CHO Cell Line. *Biotechnol. Bioeng.* **2012**, *109* (4), 971–982. https://doi.org/10.1002/bit.24383.

(239) 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), 288–291. https://doi.org/10.4161/bioe.23382.

(240) 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. *Biotechnol. Bioeng.* **2010**, *105* (2), 306–316. https://doi.org/10.1002/bit.22532.

Noelia Milagros DE LAMA VALDERRAMA

Development of new mass spectrometry methods for the characterization of protein impurities in therapeutic antibodies

Résumé

Les protéines de cellule hôte (HCPs) sont des impuretés indésirables dans la production d'anticorps monoclonaux (mAbs), pouvant compromettre la sécurité, l'efficacité et la stabilité des traitements. Bien que l'ELISA soit couramment utilisée, elle présente des limites de couverture. Cette thèse explore des méthodes complémentaires basées sur la spectrométrie de masse. Une approche d'immuno-capture permet de détecter les HCPs non immune-réactifs, tandis que des workflows LC-MS/MS avancés avec des peptide standards offrent une quantification plus précise. Ces stratégies visent à améliorer le contrôle qualité dans la fabrication des mAbs.

Mots clés : Host Cell Proteins (HCP), Monoclonal Antibodies (mAbs), ELISA, LC-MS/MS, Data-Independent Acquisition (DIA).

Abstract

Host cell proteins (HCPs) are unwanted by-products in the production of monoclonal antibodies (mAbs), and even at low levels, they can affect the safety, efficacy, and stability of biopharmaceuticals. While ELISA is widely used for HCP detection, it lacks full impurity coverage. This work explores complementary mass spectrometry-based methods to address these limitations. An immune-capture MS approach targets non-immunoreactive HCPs missed by ELISA, while advanced LC-MS/MS workflows using peptide standards enable more accurate and flexible quantification. These tools aim to improve impurity profiling and strengthen quality control in mAb manufacturing.

Keywords : Host Cell Proteins (HCP), Monoclonal Antibodies (mAbs), ELISA, LC-MS/MS, Data-Independent Acquisition (DIA).