

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES
IPHC – UMR 7178

THÈSE présentée par :

Corentin BEAUMAL

soutenue le : **20 octobre 2023**

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

Discipline/ Spécialité : Chimie analytique

**Développement de stratégies
analytiques et bioinformatiques en
spectrométrie de masse pour la
caractérisation des protéines
thérapeutiques et de leurs impuretés**

THÈSE dirigée par :

Dr. CARAPITO Christine

Directrice de recherche, CNRS, Strasbourg

RAPPORTEURS :

Pr. Jonathan BONES

Professor, University College Dublin

Dr. Julien MARCOUX

Chargé de recherche, CNRS, Toulouse

AUTRES MEMBRES DU JURY :

Dr. FERRO Myriam

Directrice de recherche, CEA, Grenoble

Dr. HERNANDEZ-ALBA Oscar

Chargé de recherche, CNRS, Strasbourg

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

IPHC – UMR 7178

THÈSE présentée par :

Corentin BEAUMAL

soutenue le : **20 octobre 2023**

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

Discipline/ Spécialité : Chimie analytique

**Development of analytical and
bioinformatics strategies in mass
spectrometry for the characterization
of therapeutic proteins and
their impurities**

THÈSE dirigée par :

Dr. CARAPITO Christine

Directrice de recherche, CNRS, Strasbourg

RAPPORTEURS :

Pr. Jonathan BONES

Professor, University College Dublin

Dr. Julien MARCOUX

Chargé de recherche, CNRS, Toulouse

AUTRES MEMBRES DU JURY :

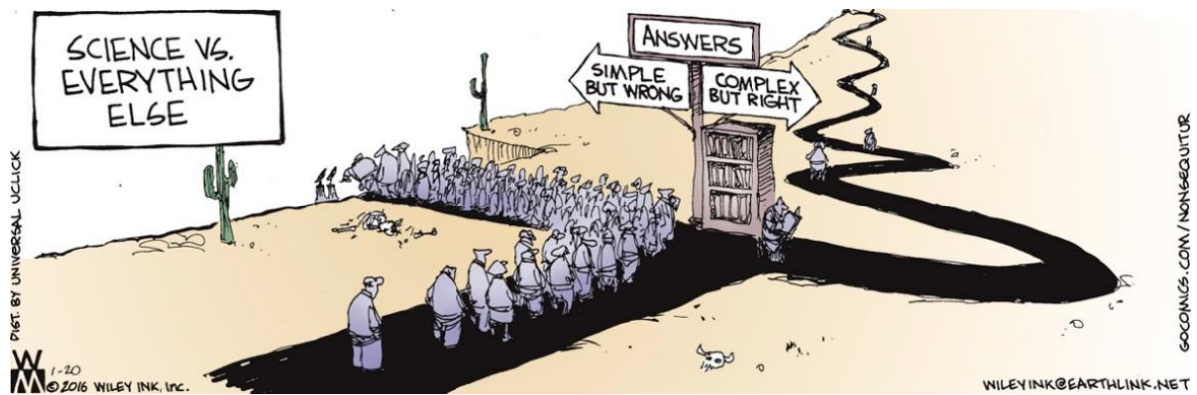
Dr. FERRO Myriam

Directrice de recherche, CEA, Grenoble

Dr. HERNANDEZ-ALBA Oscar

Chargé de recherche, CNRS, Strasbourg

*A mes parents,
A ma sœur,
A mes amis.*



NON SEQUITUR © Wiley Ink, Inc. Dist. By ANDREWS MCMEEL SYNDICATION.

REMERCIEMENTS

Je tiens tout d'abord à remercier Sarah Cianféroni de m'avoir accueilli au sein du Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO) de l'Institut Pluridisciplinaire Hubert Curien (IPHC, UMR 7178) au cours de ces trois années de thèse. Je remercie également l'Université de Strasbourg pour le financement de cette thèse.

Un grand merci à mes deux encadrants de thèse, Christine et Oscar, pour la confiance que vous m'avez accordée, pour votre encadrement et votre soutien. J'ai beaucoup appris à vos côtés et ce fût toujours un plaisir de travailler ensemble. Christine, je te remercie pour tes conseils toujours avisés, ta bienveillance et pour m'avoir encouragé à présenter mes travaux à de nombreuses reprises. Oscar, je suis fier d'avoir été un de tes premiers doctorants à encadrer et j'espère avoir été à la hauteur. Merci pour ton implication et ta bonne humeur.

I would like to thank Dr. Myriam Ferro, Dr. Jonathan Bones and Dr. Julien Marcoux for accepting to evaluate my work. Thank you also to Dr. Giovanni Chiappetta for being part of my PhD following committee.

Je voudrais remercier sincèrement l'ensemble des personnes avec qui j'ai pu collaborer. Merci à l'équipe de CompOmics à Gand, Lennart Martens, Sven Degroeve et plus particulièrement Arthur Declercq pour ton aide précieuse avec le traitement de données et la compréhension approfondie des outils bioinformatiques. Un grand merci à l'équipe de Thermo Fisher Scientific, Vincent Briaud, Sega N'Diaye et Shakir Shakir pour votre disponibilité, vos conseils et votre aide avec l'Eclipse. J'ai beaucoup appris avec vous. Merci également à Cyrille Chéry et Annick Gervais pour votre accueil à UCB, et à Somar Khalil pour le partage de ses connaissances et sa gentillesse.

Ensuite, je tiens à remercier l'ensemble des membres du LSMBO, actuels et passés, pour votre accueil et ces trois belles années passées à vos côtés. Ce fût toujours un plaisir d'échanger avec vous lors des séminaires ou autour d'un barbecue. Une mention spéciale à tous les U30 pour les soirées, les randos et tous les autres bons moments passés ensemble.

Merci Laurence de m'avoir initié (et converti !) à la spectrométrie de masse, ainsi que pour ton aide pour le concours de l'ED. Merci Christine S. pour les discussions diverses et variées autour de la DIA et de plein d'autres sujets. Merci également aux autres permanents, Agnès, Steph, Hélène, Delphine, Véronique et Fabrice B. Un merci particulier à Leslie, de m'avoir guidé pendant les premiers mois de ma thèse et de m'avoir aidé avec la mise en place de l'Eclipse. Merci aussi à Justine pour sa bonne humeur. Merci François pour tous nos échanges sur les différents logiciels, et de m'avoir fait découvrir des paramètres insoupçonnés. Je retiendrai ton goût du détail et ta passion pour l'analyse (très) approfondie des données. Merci Magali, pour ton aide avec le traitement de données et les statistiques, cela m'a été très utile.

Merci aux informaticiens Fabrice V. et Alex B. pour votre aide et les multiples dépannages !

Jean-Marc, le Mc Giver du labo, un grand merci pour ton dévouement. Tu as toujours une solution (plus ou moins artisanale) pour résoudre les problèmes. J'ai beaucoup appris grâce à toi, et je t'en remercie. Merci également pour tous les délicieux desserts et gâteaux dont tu nous fais profiter régulièrement.

Une mention spéciale à l'équipe des lève-tôt, Martine, Fabrice B., Alex et Alex, Rania, Jérôme et Hugo, pour les bons moments partagés autour du café/thé le matin. Martine, merci pour ton aide avec l'administratif et pour les bonbons dans ton bureau. Je te souhaite bon courage avec Notilus, mais je

suis sûr que d'ici quelques mois tu seras devenue une experte ! Fabrice B., je te souhaite plein de nouveaux voyages aux quatre coins du monde, pour aller étudier des espèces toujours plus exotiques.

Merci à vous, Steve et Nico, pour votre aide avec les HCPs, et tout le travail accompli avant mon arrivée. Ce fût un plaisir de prendre la suite. Noelia, c'est maintenant à ton tour de poursuivre le travail, je te souhaite le meilleur.

Thank you to all the people with whom I shared my office: Arnaud, Antonio, Zahraa, Delphine, Ilaria, Victoria, Cyrielle, Sarahi, Turkan, Charlotte et Alex. It was a nice place to work with all of you in this international environment. I hope I didn't disturb you too much when I started humming... Turkan, welcome back in the office in 2024, and I wish you good luck for the beginning of your Ph.D. Reiko, it was a pleasure to meet you and enjoy the rest of your stay in Strasbourg. Sarahi, thank you (and your husband) for feeding us quite often.

Merci à Léa et Marie R. pour leur participation à ce manuscrit. Léa, j'espère que tu n'as pas été trop traumatisée par la pompe (n'oublie pas, quand ça fait un bruit bizarre, c'est qu'il doit y avoir un problème ☺). Je te souhaite plein de bonnes choses pour le début de ta thèse (tu passeras le bonjour à Dani).

Merci à toute la team du R2. Pauline, Noelia et Charline, bon courage pour la suite (ou le début) de votre thèse. Merci Bastien pour ton investissement en tant que président du Club Jeunes de la FPS, et tu m'invites quand tu veux pour une soirée galettes ! Merci Adrien pour ton enthousiasme au quotidien, bonne continuation pour la suite (à Strasbourg ou ailleurs), et n'abuses pas trop de ChatGPT. Merci Valériane, je te souhaite bon courage avec le TMT. Jeewan, j'ai particulièrement apprécié travailler et échanger avec toi à propos de la DIA et de R. Félicitations pour tous tes progrès en français et je te souhaite le meilleur pour les quelques mois de thèse qu'il te reste.

Merci aussi à mes voisins du R5. Evolène, merci pour ton aide avec la native et le deruxtécane, et bonne continuation chez Albert. Marie L., encore bravo pour ta thèse et bon courage pour la suite. Sarah N., amuse-toi bien avec le Zeno et le top-down. Sarah D. from Boston, bon courage pour la suite de thèse. Chrystelle, garde ta bonne humeur et bonne continuation à Montpellier.

Un immense merci à ceux qui ont fait de cette thèse une superbe expérience.

Merci à mes protéomistes préférées, Aurélie, Charlotte et Marie G., pour tous les moments passés ensemble, que ce soit autour d'un verre (ou de plusieurs), d'un bon repas ou sur une piste de danse, à Strasbourg, à Prague ou ailleurs. J'espère que vous ne m'en voulez pas trop de vous avoir taquiné pendant ces trois ans, même si vous me l'avez bien rendu ! Charlotte, merci pour ton accueil à mon arrivée et de m'avoir supporté pendant 2 ans (c'est l'heure du point culture !). Garde ta bonne humeur et ta joie de vivre, je te souhaite le meilleur pour la suite. Marie G., merci pour tous nos échanges au cours de la thèse. Merci aussi d'être mon indic du R2 pour m'éviter quelques déplacements inutiles ☺. La fin de la thèse est toute proche maintenant pour toi, alors bon courage ! Aurélie, merci pour ton aide avec la préparation d'échantillons (j'ai la théorie, il faudra juste que je passe à la pratique maintenant). Je retiendrai nos longues discussions plus ou moins philosophiques et tous les bons moments passés ensemble. Même si j'aimerais bien gagner mon pari, je te souhaite de trouver ton bonheur (de préférence après janvier 2026 ☺).

Alex, j'ai été ravi de t'accueillir dans le bureau. Merci à toi pour toutes nos discussions et tous les bons moments passés, que ce soit tôt le matin ou tard le soir. Tu es un gars super, ne change pas ! Penses quand même à te reposer, la thèse est une épreuve d'endurance ! Bon courage pour la suite avec les ours, j'espère que tu pourras les approcher de près.

Un grand merci à la team Supramol, même si je garde le statut d'invité. Hugo, je te laisse la charge de notre magnifique Eclipse, prends en bien soin. Merci pour toutes nos discussions scientifiques et pour

le partage de tes connaissances. Bon courage pour la fin de la thèse ! Rania, merci pour tous ces débats animés, je pense qu'un jour on tombera d'accord ☺. Merci aussi pour ta bonne humeur, ta détermination et ton aide tout au long des trois ans. Bon courage pour les semaines à venir et vive le top-down ! Jérôme, merci pour ces trois ans, au labo ou à la coloc, ainsi que pour ton enthousiasme au quotidien. Je retiendrais tes talents de cuisinier, toujours bienvenus après une longue journée ! Bientôt la fin aussi pour toi, tout va bien se passer, courage !.

Un très grand merci à mes amis, qui m'ont soutenu pendant ces trois années, et plus.

Merci à mes anciens colocs, Baba, Titoms, Eva, Jojo et Dodo, pour les bons moments passés ensemble, votre soutien pendant cette thèse, et les WE sur un vélo ou en bord de mer pour partager le plaisir de se retrouver. Baba, Titoms et Jojo, bravo pour vos thèses respectives ! Dodo, courage pour la fin de la tienne.

Merci aux alsaciens adoptifs, Matho, Léa, Thomas, Loïc, pour les sorties aux quatre coins de l'Alsace. Merci Nathan, pour les nombreuses discussions autour d'une pinte et d'une tarte flambée.

Jakez, Valou, Lucien, Toinou, Clem et Célia, merci pour votre présence depuis de nombreuses années, et pour votre soutien malgré la distance.

Pour terminer, un merci tout particulier à ma famille, mais surtout à mes parents et ma sœur pour votre soutien sans faille. Vous m'avez toujours fait confiance et accompagné dans mes choix, quels qu'ils soient je vous en remercie. C'est grâce à vous si je suis devenu ce que je suis aujourd'hui.

RESUME EN FRANCAIS

Développement de stratégies analytiques et bio-informatiques en spectrométrie de masse pour la caractérisation des protéines thérapeutiques et de leurs impuretés

Thèse soutenue par Corentin BEAUMAL

Dirigée par Dr. Christine CARAPITO

Introduction

Les anticorps monoclonaux (mAbs) et leurs dérivés font partie des protéines thérapeutiques jouant un rôle majeur dans l'industrie pharmaceutique, avec la mise sur le marché de plus de 120 mAbs au cours des vingt dernières années, pour diverses applications¹. Le développement de ces mAbs est un processus long et fastidieux et nécessite une attention particulière pour s'assurer de l'efficacité du produit obtenu ainsi que de l'absence de risque pour le patient. Par conséquent, le produit médicamenteux doit être finement caractérisé afin de vérifier qu'il répond aux exigences réglementaires, définies notamment par ses attributs critiques de qualité (CQA)².

La présence d'impuretés protéiques issues de la cellule hôte (Host Cell Proteins, HCP) dans le produit médicamenteux est susceptible d'altérer l'efficacité du mAb ou de générer une réponse immunogène chez le patient³. De ce fait, les HCP doivent être quantifiées et la quantité globale recommandée par les agences sanitaires doit être inférieure à 100 ng par mg de mAb⁴. Aujourd'hui, le suivi de ces impuretés est très majoritairement réalisé via des méthodes immuno-enzymatiques telles que la méthode ELISA. Cependant, l'absence d'information sur l'identité et la quantité de chaque HCP ainsi qu'une couverture incomplète au sein du produit final sont des lacunes importantes de cette approche. Ces limitations ont donné lieu à l'émergence de nouvelles stratégies alternatives. Parmi elles, la protéomique quantitative, basée sur le couplage entre la chromatographie liquide et la spectrométrie de masse en tandem (LC-MS/MS), occupe une place de choix. De fait, la MS permet la caractérisation des HCP détectables sans aucun biais. Grâce aux nombreux développements technologiques et bio-informatiques en analyse protéomique, de multiples approches sont disponibles pour l'identification et la quantification des peptides/protéines (**Figure 1**). Le mode « Data Dependent Acquisition » (DDA) permet de donner une information quantitative globale sur l'ensemble des peptides détectés, mais présente une répétabilité et une gamme dynamique limitées. Depuis une dizaine d'années, le mode « Data Independent Acquisition » (DIA), basé sur la co-isolation et la co-fragmentation de l'ensemble des peptides présents dans des fenêtres d'isolation de largeurs prédéfinies sur toute la gamme de masse, s'est considérablement développée⁵. Avec une quantification basée sur la MS2, l'approche DIA démontre une spécificité, une répétabilité, une sensibilité et une précision de quantification importantes pour la caractérisation des HCP à l'état de traces. De plus, les avancées instrumentales récentes ont permis l'ajout de la mobilité ionique comme dimension de séparation supplémentaire, augmentant le nombre de protéines détectées grâce à la réduction du bruit de fond et l'accroissement de la gamme dynamique⁶. En complément de ces approches, une des clés pour améliorer l'identification des protéines réside dans le développement de nouveaux outils bio-informatiques, basés sur des algorithmes d'intelligence artificielle⁷.

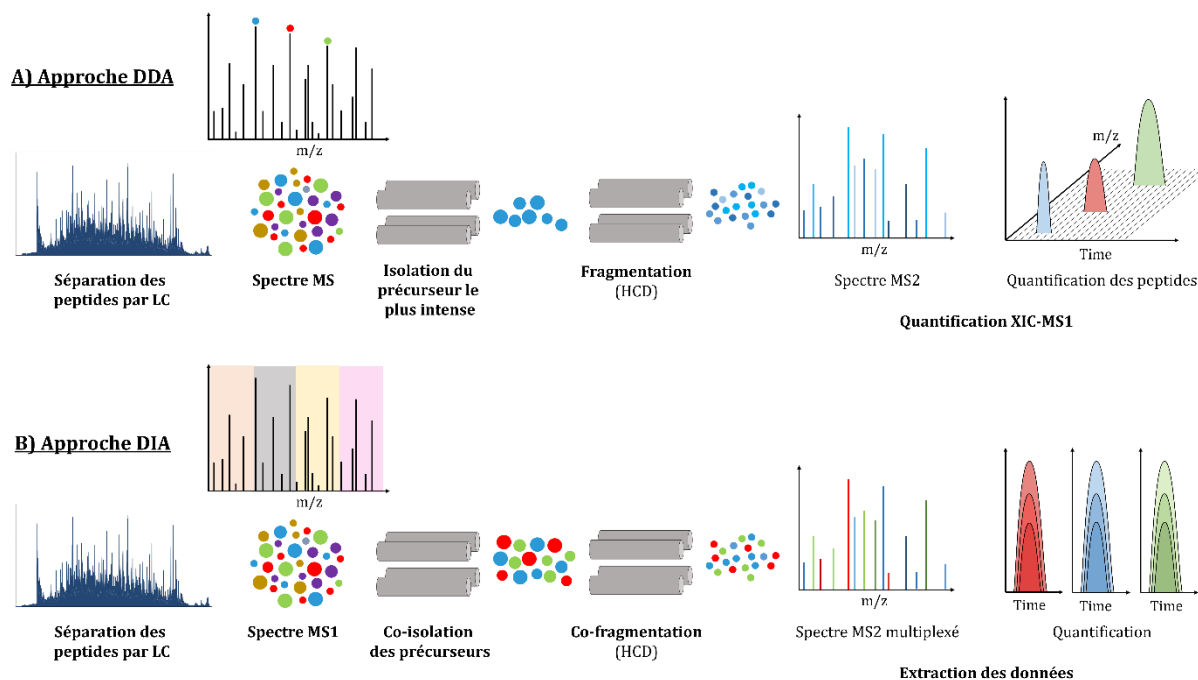


Figure 1: Aperçu des méthodes de protéomique quantitative. (A) Les quantités peptidiques peuvent être déterminées par des approches globales comme la DDA, avec l'extraction des courants d'ions des spectres MS1 acquis tout au long du gradient chromatographique. (B) L'approche DIA consiste à co-isoler et co-fragmenter des peptides co-éluant et contenus dans une même large fenêtre d'isolation, permettant l'extraction de l'information de manière ciblée à partir des spectres MS2.

Outre le suivi des impuretés de HCP présentes dans le produit médicamenteux, la caractérisation complète de la protéine thérapeutique revêt une importance fondamentale pour s'assurer de son efficacité. Pour cela, la spectrométrie de masse joue également un rôle crucial. Premièrement, l'analyse par MS en conditions natives, qui permet de conserver les interactions non-covalentes, donne accès à certaines informations structurales telles que l'homogénéité, la stœchiométrie ou bien la stabilité de la protéine thérapeutique⁸. Deuxièmement, l'utilisation de la cartographie peptidique en conditions dénaturantes permet l'analyse de la structure primaire de l'anticorps, qui a été préalablement digéré. En parallèle, les approches « top-down » (TD) et « middle-down » (MD) gagnent en popularité pour la caractérisation des modifications de l'anticorps et l'identification de protéoformes. Elles reposent sur l'analyse MS/MS des fragments générés à partir de l'anticorps entier (TD), ou partiellement digéré (MD)⁹ (**Figure 2**). Pour réaliser la fragmentation, de nombreuses techniques, basées sur des principes variés sont disponibles : dissociation par collision induite (CID) ou à haute énergie (HCD), dissociations activées par électrons (ExD) ou encore photodissociation par ultraviolet (UVPD). Des techniques additionnelles permettant de diminuer la complexité des spectres obtenus ont également été introduites, permettant d'augmenter la couverture de séquence, telles la réduction de charge par transfert de protons¹⁰ (PTCR) ou la considération des fragments internes¹¹. Une fois combinées, ces différentes approches complémentaires permettent une caractérisation précise et détaillée de la séquence d'acide aminés, et donne la possibilité de localiser des modifications, comme les agents cytotoxiques conjugués au mAb pour améliorer son efficacité. Là encore, le traitement bio-informatique des données reste une des clés pour l'obtention de couvertures de séquence plus importantes mais aussi pour assurer une confiance élevée dans les résultats obtenus.

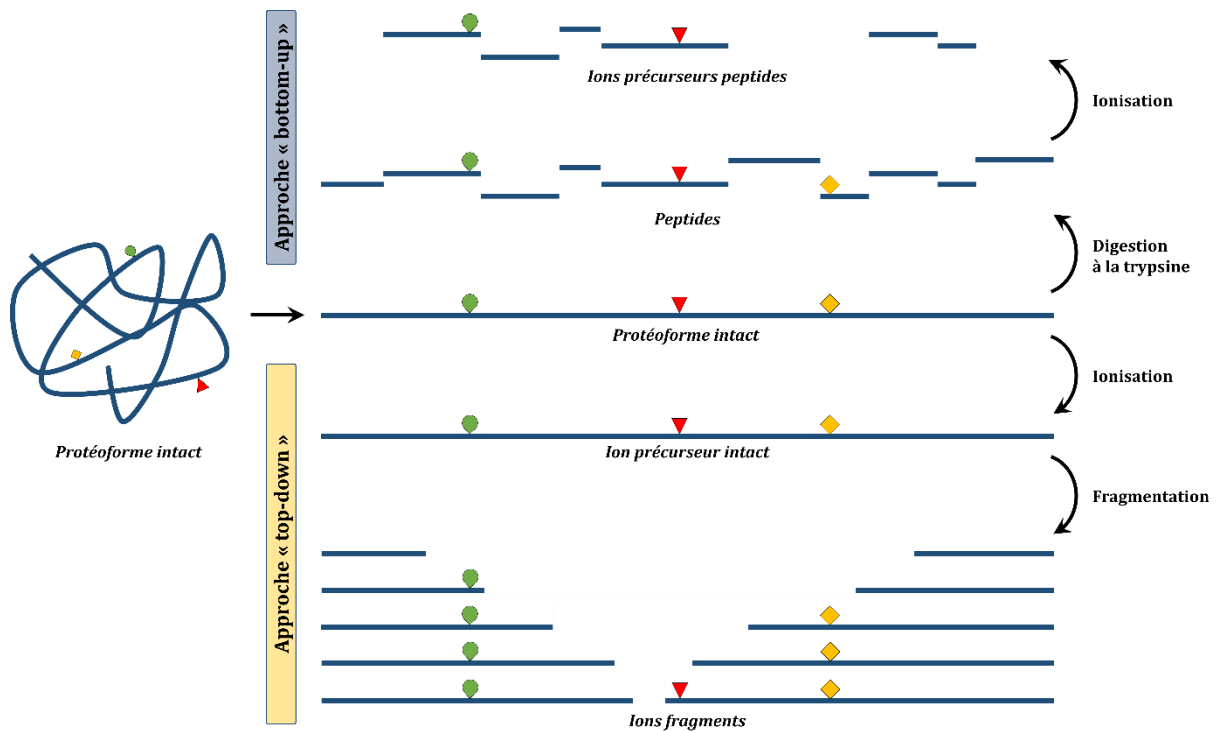


Figure 2: Identification de protéoformes par l'approche « bottom-up » et l'approche « top-down ».

Dans ce contexte, mon travail de thèse a porté sur le développement de méthodes basées sur la spectrométrie de masse, articulé autour de trois points principaux :

- L'évaluation et l'optimisation des stratégies instrumentales et bio-informatiques de protéomique quantitative
- Leur application pour la détection de HCP présents à l'état de traces dans des produits médicamenteux finaux.
- Le développement d'approches MD pour la caractérisation d'un anticorps conjugué (Antibody-Drug Conjugate, ADC), incluant la localisation précise des agents cytotoxiques.

Chapitre 1

Développements méthodologiques pour la protéomique quantitative

1. Optimisation de méthodes d'acquisition LC-MS/MS pour la protéomique quantitative

A. Evaluation des paramètres d'acquisition sur un instrument tribrid

L'ajout d'un nouvel analyseur, la trappe ionique (IT), sur les instruments de type Q-Orbitrap (OT) a donné naissance aux instruments tribrides. L'IT présente une vitesse d'acquisition plus rapide que l'OT (45 Hz vs. 40 Hz), permettant d'acquérir un plus grand nombre de spectres. De plus, cet instrument est doté de nombreuses techniques de fragmentation (CID, ETD, EThcD) qui viennent s'ajouter à l'HCD, généralement présent sur les instruments de type Orbitrap.

Ainsi, la première partie de ma thèse a consisté au développement et à l'optimisation de méthodes pour la protéomique quantitative sur un instrument de type Orbitrap Eclipse Tribrid (Thermo Fisher). Ces mises au point ont été réalisées sur un protéome total de cellules HeLa, couramment utilisé comme échantillon de référence en analyse protéomique. Les échantillons ont été analysés en mode DDA afin de comparer les modes de détection et les techniques de fragmentation (**Figure 3.A**). Les résultats de cette comparaison entre IT et OT ont montré les bénéfices de l'utilisation de l'IT pour la détection des spectres MS₂, avec un net gain en termes de peptides et de protéines identifiés et quantifiés comparé à la détection OT (**Figure 3.B**). Dans un second temps, l'analyse des différentes méthodes de fragmentation (**Figure 3.C**) a clairement fait ressortir la supériorité de la fragmentation HCD, comparé notamment aux techniques de dissociation basées sur le transfert d'électrons (ETD et EThcD).

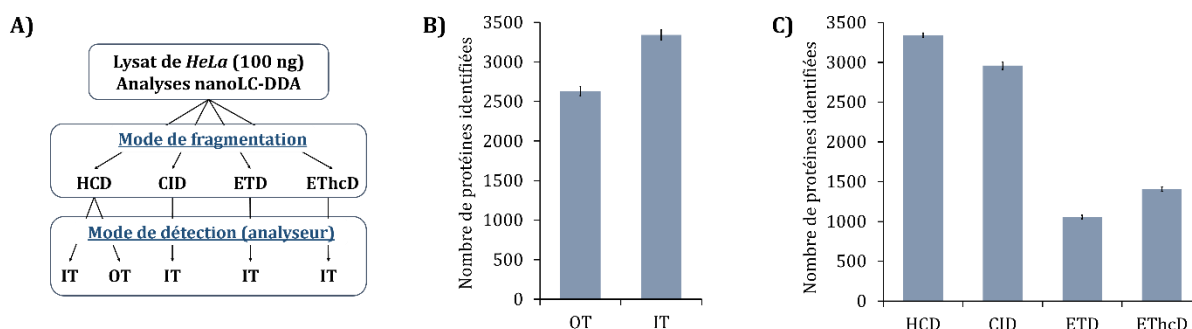


Figure 3: Evaluation des différents paramètres d'acquisition sur un instrument Orbitrap Eclipse Tribrid (Thermo Fisher). (A) Schéma analytique utilisé pour l'évaluation des modes de fragmentation (HCD, CID, ETD et EThcD) et des analyseurs (Orbitrap (OT) et trappe ionique (IT)). Nombre de protéines identifiées en comparant (B) les analyseurs et (C) les techniques de fragmentation.

Dans un second temps, différentes approches DIA ont été évaluées, en faisant varier le schéma d'isolation. Des fenêtres d'isolation fixes et variables ont été utilisées (**Figure 4.A**). Les résultats obtenus pour les fenêtres fixes ont montré que des fenêtres d'isolation de petite taille permettaient d'obtenir un nombre plus élevé de peptides et de protéines quantifiés (**Figure 4.B**). Cependant cela entraîne une augmentation du temps de cycle qui peut défavorablement impacter la précision de quantification, en limitant le nombre de points par pic chromatographique. Ainsi, l'utilisation de

fenêtres d'isolation de taille variable permet d'obtenir un nombre comparable de protéines et de peptides, tout en limitant le temps de cycle, ce qui augmente le nombre de points par pics.

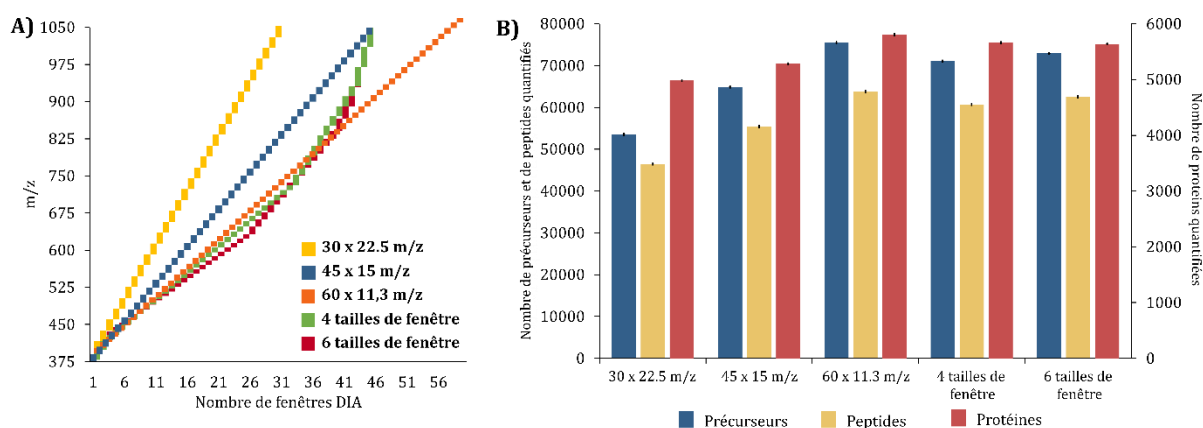


Figure 4: Influence du schéma d'isolation en DIA. (A) Différents schémas d'isolation testés avec des fenêtres fixes ou des fenêtres variables. (B) Nombre de précurseurs, de peptides et de protéines quantifiés avec chaque méthode.

Dans un troisième temps, les méthodes DDA et DIA optimisées ont été testées et comparées sur une gamme standard d'échantillons composés d'un mélange de 48 protéines humaines (Universal Protein Standard, UPS1, Merck) à 7 points de concentration différents ajouté à un lysat total d'*Arabidopsis thaliana* afin d'évaluer leurs performances (**Figure 5.A**). La DIA présente de meilleurs résultats que la DDA concernant les protéines quantifiées dans tous les réplicats pour les faibles quantités (25 et 125 amol). En effet, 43 des 48 protéines UPS1 sont identifiées à 25 amol en DIA contre 32 en DDA. De plus, nous avons quantifié toutes les protéines UPS1 dans tous les réplicats à partir de 250 amol en DIA contre 1 250 amol en DDA. La tendance à des résultats légèrement meilleurs en DIA est maintenue lorsque l'on filtre avec les protéines quantifiées dans les trois réplicats avec un CV <20%. De plus, les résultats en termes de précision de quantification, ainsi que le nombre d'identifications ont été comparés (**Figure 5.B**). Les deux approches présentent une dispersion des rapports expérimentaux autour des valeurs théoriques qui augmente avec le rapport théorique, révélant les limites de la précision de quantification des approches globales. Néanmoins, les rapports expérimentaux médians sont proches des valeurs théoriques, à l'exception du rapport le plus élevé (5 000 vs. 25 amol).

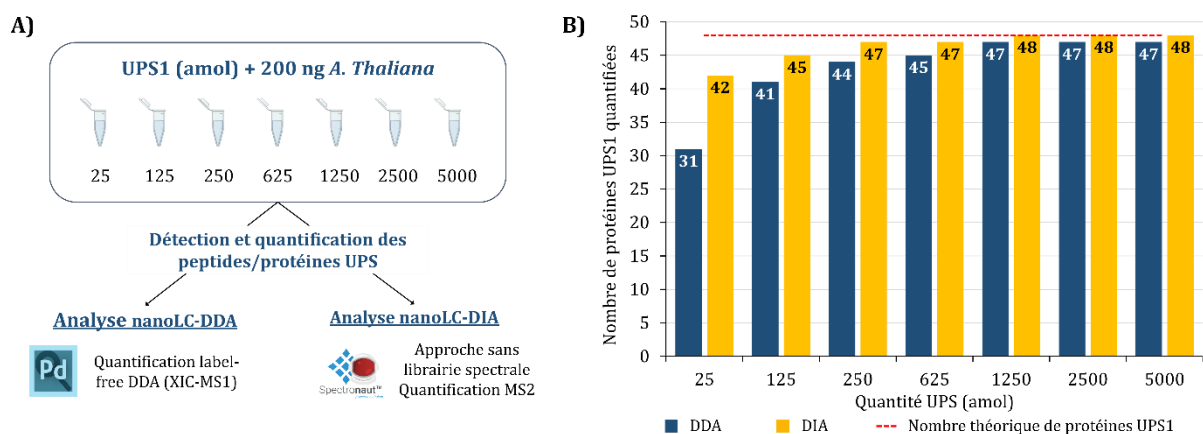


Figure 5: Evaluation des approches quantitatives en DDA et DIA sur une gamme d'échantillon parfaitement calibrée. (A) Schéma analytique utilisé. (B) Nombre de protéines UPS quantifiées dans les différents points de la gamme par les stratégies DDA et DIA.

L'ensemble des optimisations réalisées durant la première partie de ma thèse et présentées ici ont permis de démontrer les performances très prometteuses de la DDA et de la DIA pour la quantification de protéines.

B. Optimisation de la mobilité ionique sur un instrument de type Orbitrap Eclipse Tribrid

En complément des optimisations précédemment évoquées, j'ai également pu me concentrer sur la mise en place d'un module de mobilité ionique asymétrique à haut-champ (high-Field Asymmetric Ion Mobility Spectrometry, FAIMS, Thermo Fisher), agissant comme une dimension de séparation orthogonale, basée sur la charge et la mobilité ionique des ions en phase gazeuse. Suivant les voltages de compensation utilisés, les ions pénétrant dans l'instrument vont être différents, et une optimisation fine de ces paramètres permet de s'affranchir de la présence des ions monochargés, majoritairement responsables du bruit de fond⁶ (**Figure 6.A**).

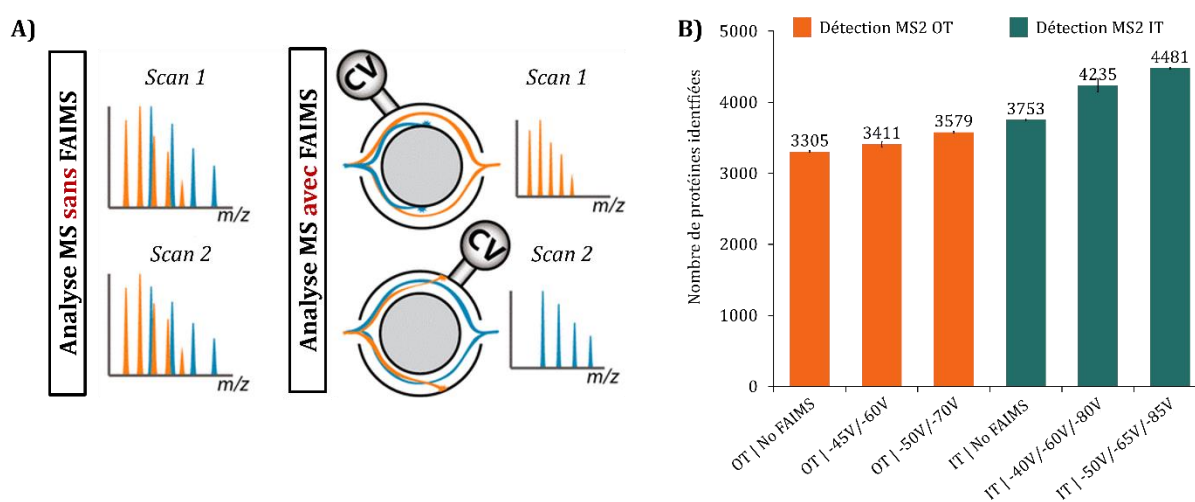


Figure 6: Analyses DDA avec l'ajout de la mobilité ionique FAIMS. (A) Principe de fonctionnement du FAIMS. (B) Nombre de protéines identifiées avec une détection Orbitrap et trappe ionique, sans le FAIMS et avec le FAIMS en variant les voltages de compensation.

Pour ces optimisations, les cellules HeLa précédemment mentionnées ont été utilisées et différentes combinaisons de voltages de compensation ont été testées, avec une détection en OT et IT, afin d'obtenir la comparaison la plus complète possible. Les résultats de l'ajout de la mobilité ionique, indépendamment de l'analyseur utilisé, permettent d'augmenter le nombre de protéines identifiées de près de 20% en IT et 8% en OT, avec environ 4 500 protéines identifiées dans les conditions optimales en IT (**Figure 6.B**). En revanche, le nombre de peptides observés est, lui, similaire avec et sans l'ajout du FAIMS, ce qui s'explique par le fait que certains peptides, par l'utilisation de différents voltages de compensation, ne peuvent entrer dans le spectromètre de masse. Cet outil reste néanmoins un avantage considérable pour augmenter le nombre de protéines identifiées.

C. Evaluation des stratégies d'acquisition dédiées à l'approche TMT

Une alternative à la protéomique quantitative sans marquage (« label-free ») est l'emploi d'approches utilisant un marquage chimique. Elles permettent de multiplexer différents échantillons grâce au greffage de marqueurs isotopique. Ici, le TMT¹² (« tandem mass tag ») a été utilisé pour marquer des cellules de HeLa. Deux paramètres d'acquisition ont été évalués : l'énergie d'activation, qui influence la fragmentation du marqueur isotopique et permet la quantification relative des différents échantillons, et les méthodes d'acquisition en MS2, en MS3 et en MS3 dirigée par la recherche en temps réel (« real time search », RTS).

L'étude des énergies de fragmentation a montré un nombre de protéines identifiées similaire, indépendamment de l'énergie utilisée, alors que le nombre de peptides présente un maximum pour une énergie normalisée de collision (NCE) de 35% (**Figure 7.A**). De plus, une haute énergie de collision aura pour effet d'augmenter le signal sur bruit des ions rapporteurs, améliorant ainsi la quantification. Ainsi, une valeur d'énergie de collision autour de 35% NCE semble un bon compromis pour obtenir un bon signal sur bruit tout en maximisant le nombre d'identifications.

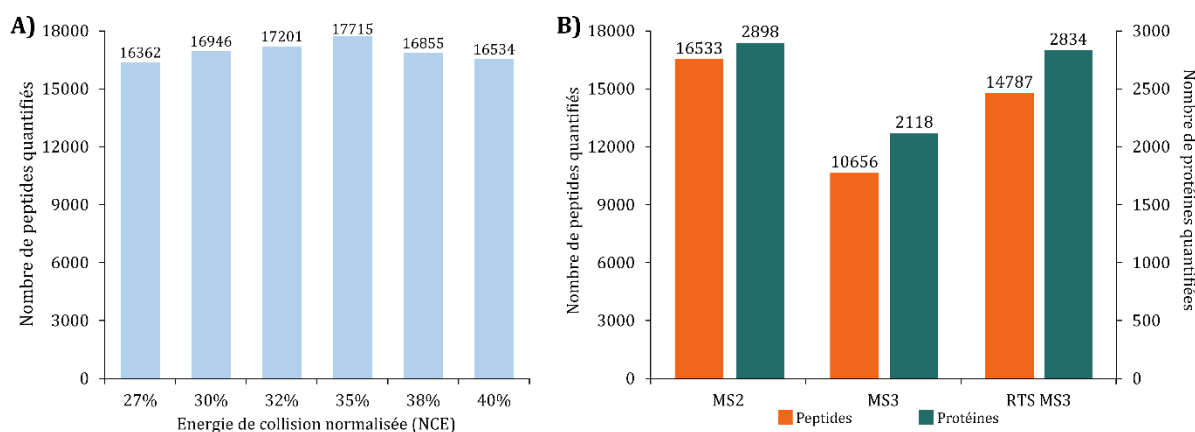


Figure 7: Evaluation des approches TMT. (A) Nombre de peptides quantifiés en fonction de l'énergie d'activation utilisée pour la fragmentation. (B) Nombre de protéines et de peptides quantifiés en fonction de la méthode d'acquisition employée (MS2, MS3 ou MS3 RTS).

Dans un second temps, nous nous sommes concentrés sur les méthodes d'acquisition dédiées au TMT. Une des limitations du TMT est la compression des ratios, due à la co-isolation et la co-fragmentation d'ions interférents. Pour cela, des stratégies comme l'emploi de fenêtres d'isolation réduites ou l'ajout d'une étape de fragmentation supplémentaire ont été proposées. Notamment, la quantification MS3^{13, 14} permet de diminuer significativement la compression des ratios mais induit un temps de cycle beaucoup plus long, impactant négativement le nombre de peptides détectés. Pour cela, la recherche en temps réel a été introduite, permettant de fragmenter en MS3 uniquement les ions donnant lieu à une identification. Les résultats de cette évaluation sont présentés en **Figure 7.B**. Il apparaît clairement que l'utilisation de la MS3 RTS permet de pallier au long temps de cycle induit par l'utilisation de la MS3. En terme de quantification, des résultats similaires sont obtenus avec les trois techniques. Ainsi, l'approche MS2 avec des fenêtres d'isolation étroites et l'approche MS3 RTS semblent de bons compromis pour la quantification d'échantillons marqués au TMT.

2. Evaluation d'outils bio-informatiques pour le traitement des données

En parallèle de ces développements méthodologiques, une partie de ces travaux de thèse ont été consacrés à l'évaluation de différents outils de traitement de données. Au cours des dernières années, de nombreux outils, utilisant des algorithmes basés sur l'intelligence artificielle et le « deep learning », ont été développés afin d'améliorer les identifications peptidiques et protéiques. Ces nouveaux algorithmes permettent notamment d'attribuer un nouveau score au peptide identifié, en prenant en compte de multiples éléments dans le calcul de ce score. Parmi ces éléments, la comparaison entre le spectre expérimental et le spectre prédit par apprentissage, grâce aux nombreuses données expérimentales disponibles, joue un rôle important¹⁵. Ces outils ont permis d'augmenter le nombre d'identification de l'ordre de 20% à 30% (**Figure 8.A**). De plus, ils permettent de mieux différencier les séquences cibles (« target ») des leurres (« decoys ») (**Figure 8.B**), renforçant la confiance dans les identifications obtenues.

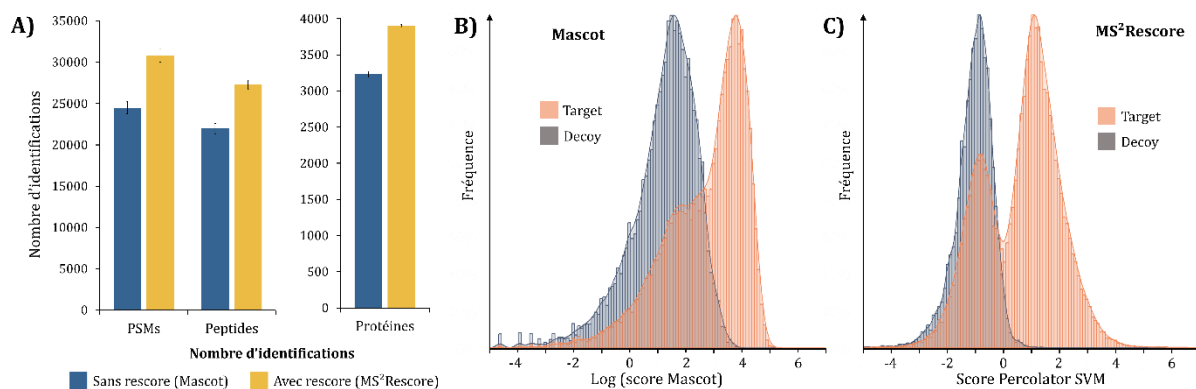


Figure 8: Influence des algorithmes de « rescoring ». (A) Nombre de PSMs, peptides et protéines identifiés sans et avec « rescoring ». Séparation entre les séquences « targets » et « decoys » en fonction du score (B) avant « rescoring » et (C) après « rescoring ».

D'autres outils, se basant directement sur la prédiction des spectres (intensité des fragments, allure du spectre, etc.) pour réaliser la recherche ont également révélé d'excellentes performances, permettant de doubler le nombre d'identifications obtenues en détectant des peptides issus d'un même spectre de fragmentation. Enfin, les algorithmes autorisant une recherche ouverte des modifications peuvent aussi être utilisés¹⁶. Contrairement aux logiciels classiques, avec lesquels la recherche contre la base de données est réalisée en autorisant un nombre très limité de modifications, ces algorithmes permettent de rechercher n'importe quelle modification sans aucune restriction. Ainsi, de nouveaux peptides qui n'étaient pas identifiés auparavant, puisqu'ils n'étaient pas recherchés, peuvent être détectés. Ces approches, bien que récentes et à leurs prémices, se révèlent très prometteuses. Dans ce manuscrit, le logiciel Ionbot a été testé et a mis en évidence l'utilité de la recherche ouverte grâce à l'identification de nombreuses nouvelles PTMs.

Enfin, deux algorithmes permettant l'analyse de données DIA en mode spectre-centré, Spectronaut et DIA-NN¹⁷ ont été comparés et plusieurs paramètres ont été modifiés. Les résultats obtenus sont comparables entre les deux logiciels, et montrent que les paramètres par défaut semblent bien adaptés pour des recherches de protéomique quantitative classiques. La précision de quantification obtenue avec différents paramètres souligne également que les paramètres proposés apparaissent comme les plus adéquats.

Chapitre 2

Application des méthodes développées pour la quantification d'impuretés d'HCPs présentes à l'état de traces

Les méthodes développées dans la première partie de cette thèse ont ensuite été employées et adaptées pour la quantification d'impuretés d'HCPs présentes à l'état de traces dans les produits médicamenteux. Un produit médicamenteux de référence, le NIST mAb (Sigma), a été utilisé pour réaliser les optimisations. Les méthodes optimisées ont finalement été éprouvées sur deux produits commerciaux, le trastuzumab et le nivolumab, tous deux utilisés en oncologie. L'objectif principal de cette étude était de pouvoir identifier ces HCP mais également de pouvoir les quantifier avec précision de manière répétable, en utilisant une courbe de calibration interne. Fort des résultats obtenus précédemment avec la gamme d'échantillons parfaitement calibrée et de l'ajout de la mobilité ionique, 4 méthodes d'acquisition différentes ont été comparées (**Figure 9.A**) : une méthode DDA, une méthode DIA, une méthode DDA avec ajout du FAIMS et une méthode DIA avec fractionnement en phase gazeuse¹⁸ (GPF).

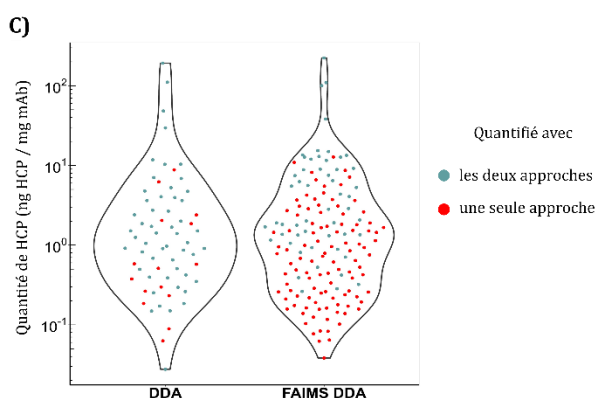
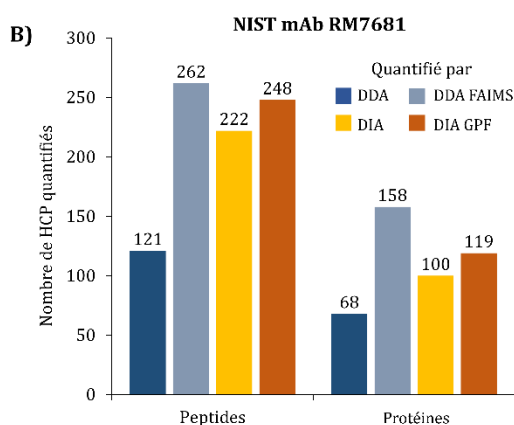
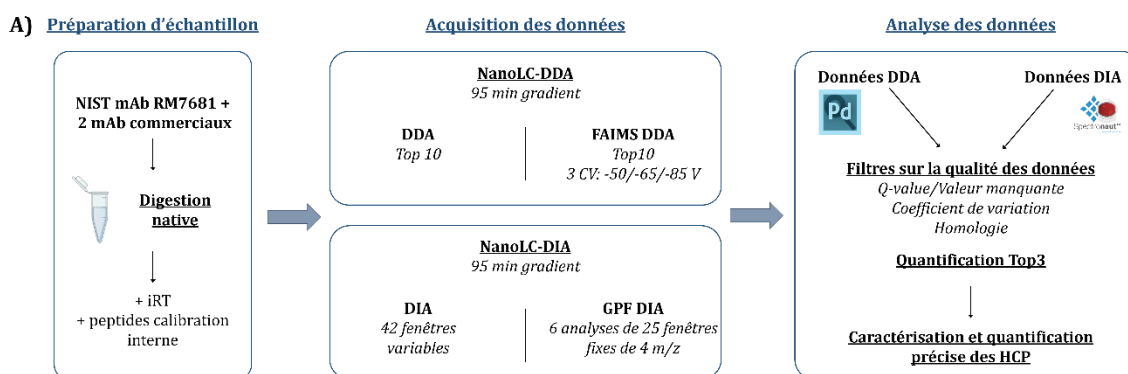


Figure 9: Evaluation des différentes méthodes pour la quantification précise d'HCP présentes à l'état de traces dans des produits médicamenteux. A) Schéma analytique utilisé pour l'évaluation des différentes stratégies DDA (combinée ou non au FAIMS) et DIA (avec ou sans fractionnement en phase gazeuse). B) Nombre de peptides et de protéines HCP quantifiés après analyse des données dans l'échantillon de NIST mAb RM7681. C) Graphique en violon représentant la distribution de la quantité individuelle de chaque HCP pour les stratégies DDA avec et sans FAIMS.

Plusieurs méthodes DIA ont été testées et ont montré des résultats comparables en termes d'identifications. Cependant, après application de filtres stricts tenant compte de la qualité du signal et de la reproductibilité de l'analyse, la méthode avec un schéma d'isolation de 4 largeurs variables de fenêtre apparaît comme la plus adaptée pour l'étude des HCPs. Comme attendu, la méthode DIA a montré un avantage important comparé à la méthode DDA, avec la quantification de 100 HCP contre 68 en DDA sur l'échantillon de NIST mAb (**Figure 9.B**). Ensuite, l'ajout du FAIMS a été évalué en mode DDA et différents voltages de compensation ont été testés. L'utilisation de valeurs optimisées avec le FAIMS a permis de multiplier par 2 le nombre de protéines quantifiées dans cet échantillon par rapport à une analyse sans FAIMS, permettant la quantification de plus de 130 HCP.

De même, pour le mode DIA, le GPF a lui aussi permis une hausse du nombre de protéines quantifiées comparé à la méthode DIA classique, avec respectivement 119 et 100 HCP (**Figure 9.B**). Malgré ce gain, l'approche GPF DIA reste moins performante que l'approche DDA FAIMS. En complément de ce gain, il peut être mentionné que les HCP supplémentaires quantifiées par ces deux approches (FAIMS DDA et DIA avec GPF) sont présentes à des concentrations très faibles (**Figure 9.C**), démontrant la capacité de ces méthodes à accroître la profondeur d'analyse. En considérant ces résultats positifs et très prometteurs sur l'échantillon de référence NIST mAb, ces mêmes méthodes ont été appliquées avec succès aux deux produits médicamenteux précédemment mentionnés. Là encore, la méthode FAIMS DDA permet d'identifier le plus grand nombre de HCP. Au final, une trentaine de HCP pour une quantité totale d'environ 55 ng of HCP/mg of mAb ont pu être quantifiés dans le trastuzumab et le nivolumab (**Figure 10**). Cela montre le réel potentiel des approches innovantes en spectrométrie de masse dans le contexte des HCP.

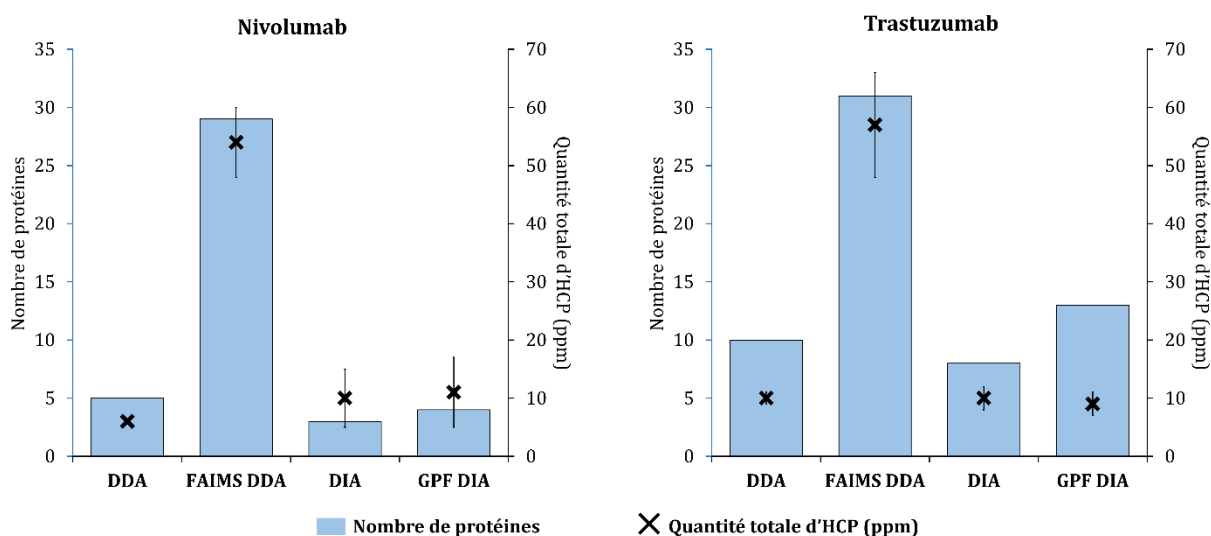


Figure 10: Nombre de HCP quantifiées dans le trastuzumab et la nivolumab pour chaque approche, et quantité total d'HCP associée. Les barres bleues représentent le nombre de HCP et les croix noires la quantité totale de HCP (ppm).

Chapitre 3

Caractérisation d'anticorps monoclonaux thérapeutiques par spectrométrie de masse

Après avoir développé différentes stratégies pour le suivi et la quantification des HCP dans les produits médicamenteux, le deuxième axe de recherche de mon travail de thèse s'est focalisé sur le développement des approches de type « top-down » pour la caractérisation de protéines thérapeutiques.

1. Optimisation des paramètres d'acquisition pour les stratégies « top-down » et « middle-down »

A. Mise au point des paramètres clés d'acquisition

Dans un premier temps, ce travail de thèse s'est porté sur l'optimisation des multiples méthodes de fragmentation disponibles sur l'instrument (HCD, CID, ETD, EThcD et UVPD) afin d'obtenir la meilleure couverture de séquence possible. Différents paramètres ont été évalués, tels que l'énergie d'activation, le choix des ions précurseurs, ou encore le temps d'acquisition. Ces optimisations ont tout d'abord été réalisées sur des protéines de référence (myoglobine et anhydrase carbonique) en infusion directe. Avec les résultats obtenus, plusieurs points importants ont été mis en évidence. Premièrement, l'influence de l'état de charge sur l'efficacité de fragmentation et la couverture de séquence a été montrée. En fonction de la méthode de fragmentation utilisée, le précurseur donnant la meilleure fragmentation n'est pas le même. Pour les techniques de dissociation basées sur le transfert d'électrons, un état de charge plus important permet d'augmenter la couverture de séquence, alors que la charge du précurseur n'a que peu d'influence sur la fragmentation par photodissociation. Deuxièmement, l'énergie d'activation employée pour la fragmentation des protéines a été optimisée (**Figure 11**).

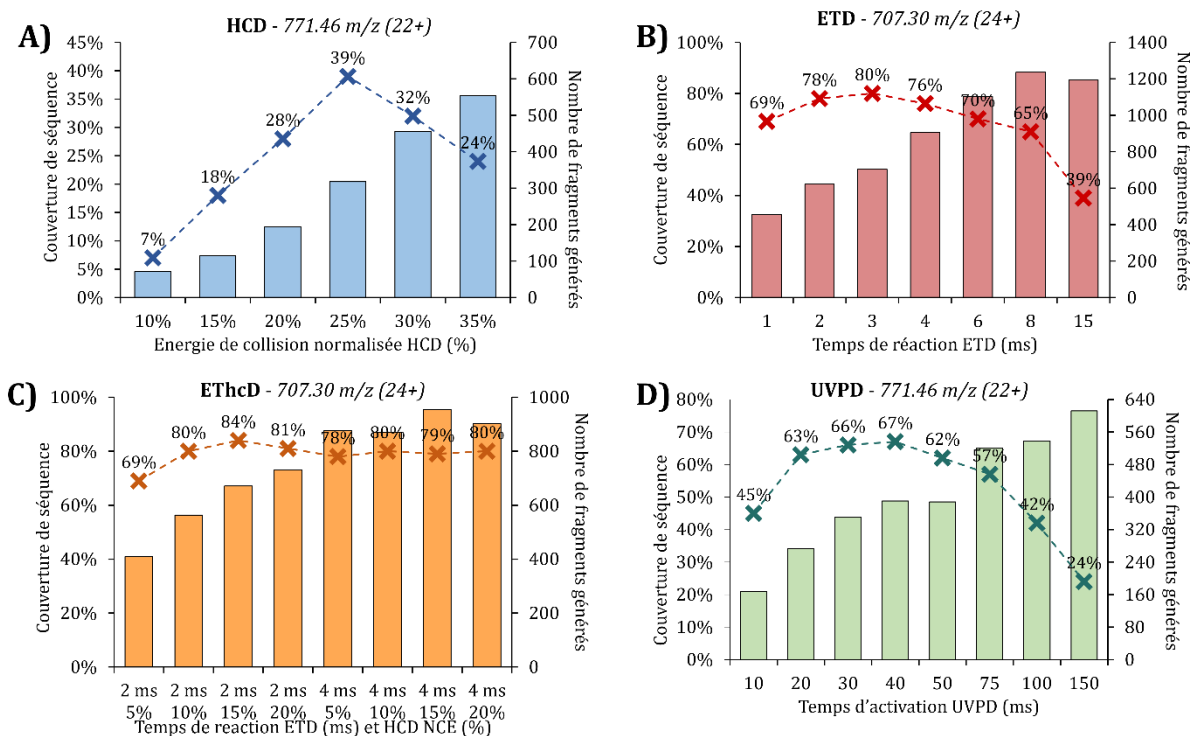


Figure 11: Résultats de fragmentation de la myoglobine en fonction de l'énergie d'activation. Couverture de séquence et nombre de fragments générés après (A) HCD sur l'ion 22+, (B) ETD sur l'ion 24+, (C) EThcD sur l'ion 24+, et (D) UVPD sur l'ion 22+.

Alors que le nombre de fragments générés est directement lié à l'augmentation de l'énergie d'activation, ce n'est pas le cas pour la couverture de séquence. En effet, une énergie d'activation trop importante entraîne une sur-fragmentation de l'ion précurseur, générant de petits fragments qui ne sont pas identifiés. Au contraire, une énergie de fragmentation trop basse donne une fragmentation limitée de la protéine. Par conséquent, les énergies d'activation optimales ont été déterminées pour chaque technique de fragmentation.

Les deux paramètres mentionnés plus haut sont dépendant l'un de l'autre. Par conséquent, l'énergie d'activation optimale ne sera pas la même pour un ion précurseur avec un état de charge différent. Ces résultats soulignent l'importance de l'optimisation des paramètres expérimentaux pour chaque protéine avec l'approche « top-down »¹⁹.

Dans un second temps, une dimension de séparation chromatographique a été ajoutée en amont de la MS, impliquant une restriction du temps d'analyse disponible, limité à la largeur du pic chromatographique. Cette approche a été utilisée pour la caractérisation des sous-unités d'un anticorps monoclonal. Là encore, le choix de l'ion précurseur joue un rôle important pour maximiser la couverture de séquence. De plus, le choix des paramètres chromatographiques est primordial. En effet, il est important d'assurer une bonne séparation des sous-unités ainsi que des pics d'une largeur suffisante afin de maximiser le nombre de spectres acquis pour chaque espèce. Une fois ces paramètres optimisés, les sous-unités ont été fragmentées avec chaque technique. La **Figure 12** illustre les résultats obtenus.



Figure 12: Couverture de séquence obtenue avec les différentes techniques de fragmentation sur la chaîne légère du trastuzumab. (A) HCD, (B) CID, (C) ETD, (D) EThcD, (E) UVPD et (F) combinaison de l'ensemble des techniques.

Les techniques de fragmentation par collision présentent le plus faible recouvrement de séquence, entre 22% et 38%, comparable à l'UVPD (38% à 42%). En revanche, l'ETD et l'EThcD apparaissent comme des techniques de fragmentation plus efficace avec des couvertures de séquence de l'ordre de 50% à 60%. La combinaison de l'ensemble des techniques de fragmentation donne une couverture de séquence de l'ordre de 80% à 90%, démontrant leur complémentarité pour obtenir la caractérisation la plus complète possible de la séquence de la protéine.

B. Evaluation de l'apport de la réduction de charge par transfert de proton (PTCR)

Les spectres générés par la fragmentation de précurseurs multichargés sont généralement très complexes, pouvant donner lieu à des difficultés lors du traitement des données. Ainsi, différentes approches ont été introduites pour pallier cette limitation. L'une d'entre elles est la réduction de charge par transfert de proton^{10, 20} (PTCR). Initialement, la PTCR était utilisée pour réduire la charge des ions précurseurs au niveau de la MS1, afin d'améliorer la qualité du spectre, notamment pour les espèces de haut poids moléculaire. Ensuite, différentes applications ont montré l'intérêt de la PTCR pour décomplexifier des spectres de fragmentation MS2 ou MS3, en réduisant le chevauchement entre les pics. Une réaction entre les ions fragments et le réactif anionique a lieu, permettant de réduire la charge des ions fragments, et ainsi de les distribuer sur une gamme de m/z plus importante. La PTCR

a été appliquée sur une grande fenêtre d'isolation de 1000 m/z ainsi que sur cinq fenêtres d'isolation plus étroites de 200 m/z (**Figure 13**).

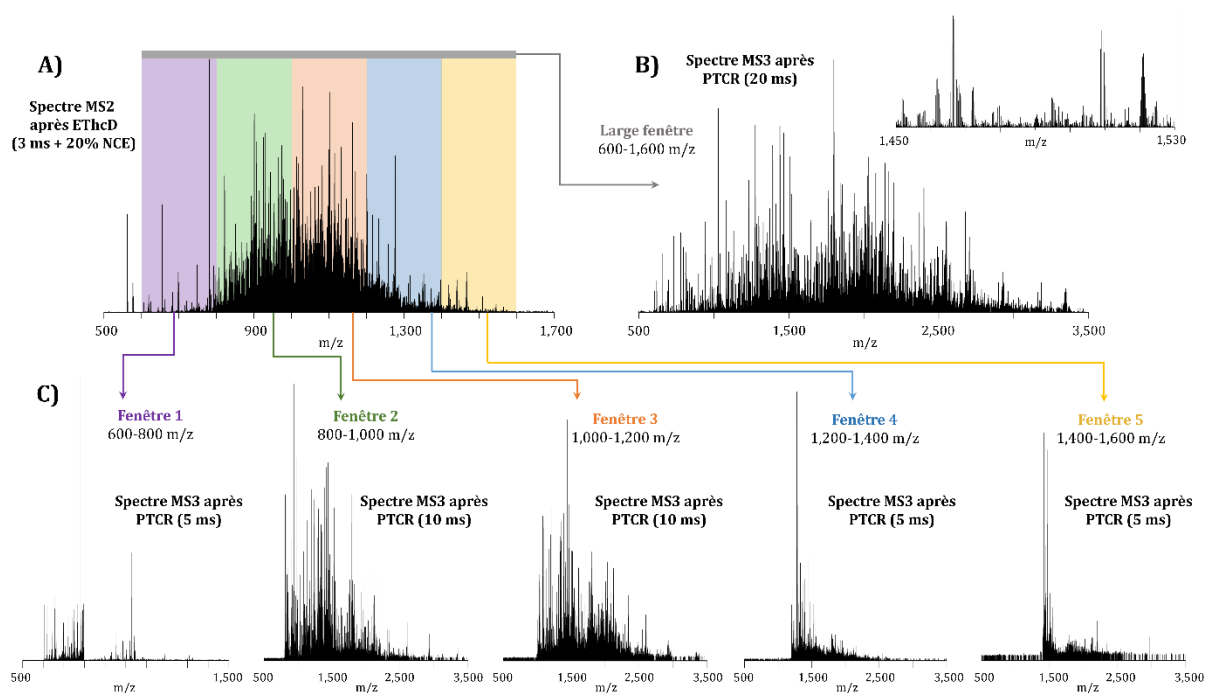


Figure 13: Représentation schématique de l'utilisation de la PTCR. (A) Spectre MS2 standard obtenu après fragmentation EThcD. (B) Isolation des ions produits par la fragmentation sur une fenêtre 1000 m/z et application de la PTCR. (C) Isolation de sous-populations d'ions dans des fenêtres restreintes de 200 m/z et application de la PTCR.

De ce fait, le spectre est beaucoup moins complexe, facilitant l'étape de déconvolution des ions fragments lors du traitement de données. L'emploi de la PTCR sur des fenêtres d'isolation larges et étroites après la fragmentation ETD et EThcD a permis d'obtenir une couverture de séquence plus importante de 5% à 15%, selon les sous-unités et la taille de fenêtre utilisée.

2. Importance du traitement de données pour les approches « top-down » et « middle-down »

Le traitement de données est une des clés des stratégies « top-down » et « middle-down ». Les spectres de fragmentation obtenus sont déconvolués afin de pouvoir identifier les fragments appartenant à la protéine. Dans un premier temps, le risque de faux positifs a été évalué en utilisant des séquences mélangées et une séquence inversée. Les résultats obtenus montrent qu'un certain nombre de fragments sont quand même identifiés avec ces fausses séquences. Cependant, le P-Score (défini comme la probabilité qu'une identification soit liée au hasard²¹) observé pour ces fausses séquences est très haut, pointant le fort risque de faux positifs. Ainsi, ce score apparaît comme un bon indicateur de la fiabilité de l'identification des ions fragments.

Lors de la déconvolution, deux paramètres cruciaux ont une influence importante sur la couverture de séquence : le seuil de signal sur bruit (S/N) et le facteur de ressemblance (« fit factor »). L'impact de la variation de ces paramètres a été évalué dans le cadre de cette partie. Il a été observé que des paramètres de déconvolution trop stricts diminuaient drastiquement la couverture de séquence obtenue. En revanche, des paramètres trop souples, bien qu'augmentant le nombre de fragments déconvolués, ne participent pas nécessairement à un gain significatif de la couverture de séquence et entraîne un risque plus important d'identifier un faux positif (**Figure 14**).

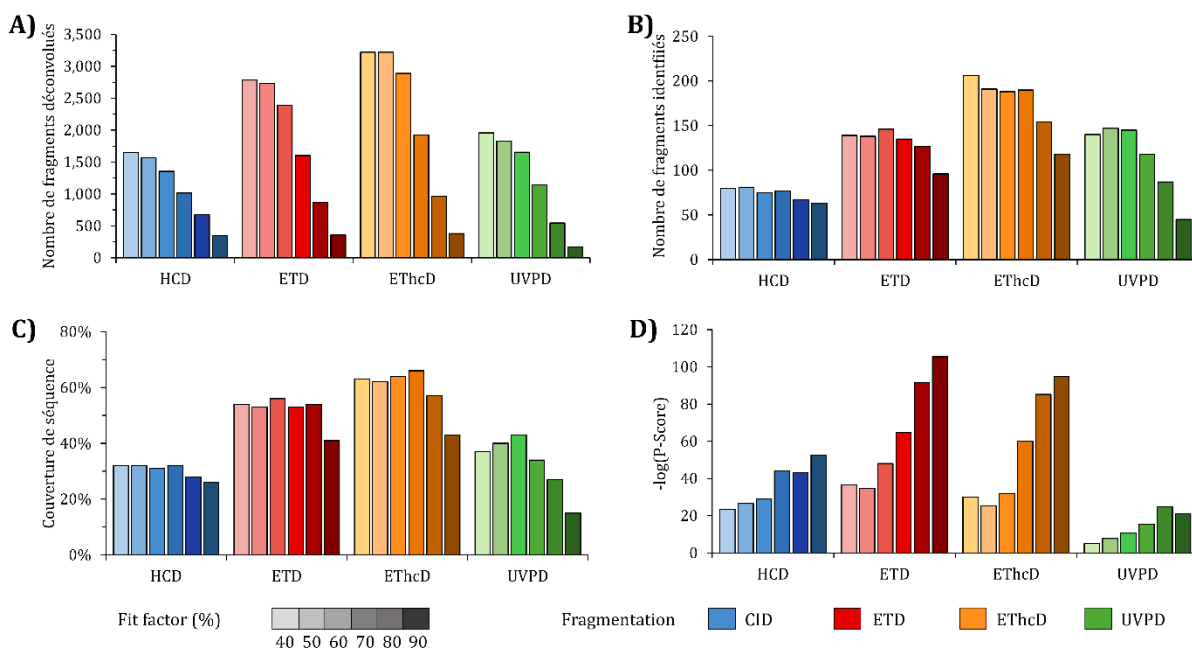


Figure 14: Résultats d'identification de la chaîne légère du trastuzumab en fonction de la valeur de « fit factor » utilisé. (A) Nombre de fragments déconvolués, (B) nombre de fragments identifiés, (C) couverture de séquence (%), et (D) valeurs de $-\log(\text{P-Score})$.

Les résultats présentés ont permis d'établir des valeurs qui peuvent servir de base pour la déconvolution : un ratio de S/N entre 5 et 7 et un facteur de ressemblance entre 70% et 80% semblent des compromis acceptables pour assurer une couverture de séquence importante tout en limitant le risque de faux positifs. De manière similaire, la tolérance de masse utilisée pour l'identification des fragments a été évaluée. Là encore, une tolérance de masse trop stricte a significativement réduit la couverture de séquence, alors qu'un seuil trop souple n'augmentait pas le nombre de fragments identifiés et accroissait le risque de faux positifs.

De manière générale, les logiciels de traitement de données ne prennent que les fragments terminaux en compte. Cependant, parmi le nombre très important de fragments générés lors de la fragmentation, une partie des ions non attribués peut être assigné à des fragments internes¹¹. Ainsi, de nouveaux algorithmes incluant la recherche des fragments internes pour l'identification ont été proposés. Pour cela, le logiciel ClipsMS a été utilisé afin d'apprécier le gain en couverture de séquence considérant les fragments internes. Les couvertures de séquences obtenues ont bondi de 20% à 60%, suivant la méthode de fragmentation utilisée, pour atteindre un pourcentage de clivages inter-résidus observés entre 65% (HCD) et 95% (EThcD et UVPD). Cependant, une des limites de l'utilisation des fragments internes est le risque important de faux positifs. En effet, des couvertures de séquence comparables ont été observées entre de fausses séquences (séquence inversée ou mélangée), et la vraie séquence, montrant le fort risque d'erreurs induit par les fragments internes, ainsi que la nécessité de considérer ces résultats avec précaution.

3. Caractérisation complète d'un anticorps conjugué thérapeutique par l'approche « middle-down »

Dans un second temps, les méthodes optimisées ont été adaptées pour la caractérisation d'un ADC, le trastuzumab deruxtecan²² (T-DXd, **Figure 15.A**). En amont de la caractérisation de la séquence et de la localisation des molécules cytotoxiques par « middle-down », la protéine a aussi été étudiée en conditions natives et dénaturantes. Par cette approche, la masse de la protéine ainsi que le nombre théorique de molécules cytotoxiques conjuguées à l'anticorps ont été confirmés. Il a ainsi été établi que

le T-DXd présentait bien un nombre moyen de molécules cytotoxiques homogène de 8 (**Figure 15.B**). Ensuite, le T-DXd a été partiellement digéré et les sous-unités ont été séparées par chromatographie liquide avant d'être fragmentées. Le recours aux différentes techniques de fragmentation a rendu possible l'obtention d'une couverture de séquence comprise entre 15% et 60% suivant les sous-unités et les techniques utilisées. Une fois combinés, ces résultats ont permis d'obtenir un recouvrement de séquence compris entre 60% et 80% (**Figure 15.C**).

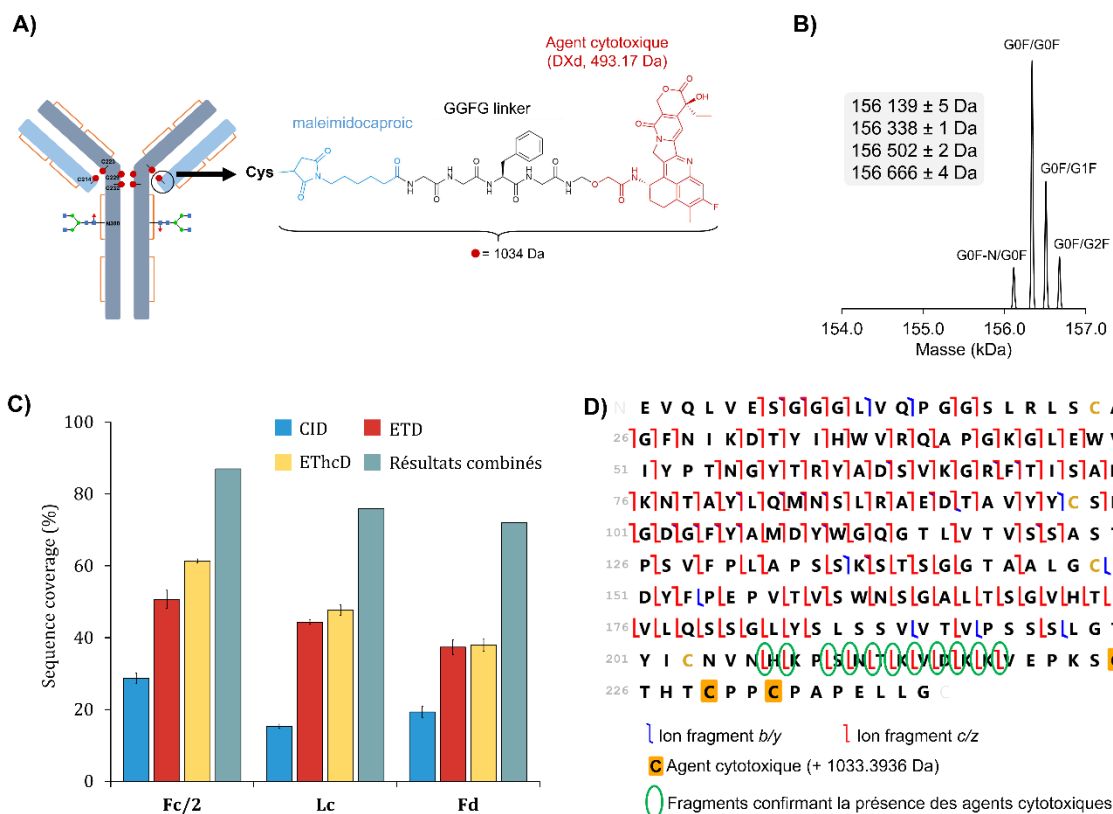


Figure 15: Caractérisation du trastuzumab deruxtecan. (A) Molécule de trastuzumab deruxtecan (T-DXd). (B) Spectre déconvolué de T-DXd avec la présence de 4 glycoformes. (C) Couvertures de séquence des sous-unités (Fc/2, chaîne légère (LC) et Fd) de T-DXd obtenues avec les différentes méthodes de fragmentation, et après combinaison de ces techniques. (D) Carte de fragmentation représentant l'ensemble des fragments obtenus pour la sous-unité Fd contenant 3 charges utiles après combinaison de l'ensemble des méthodes de fragmentations et de la PTCR. Les fragments confirmant la présence des agents cytotoxiques sont entourés en vert.

De plus, certains fragments caractéristiques ont été observés venant confirmer la localisation des molécules cytotoxiques sur les cystéines impliquées dans les ponts disulfures inter chaînes (**Figure 15.D**). Enfin, des approches innovantes ont été éprouvées sur cet échantillon, pour augmenter la couverture de séquence et conforter la localisation des agents cytotoxiques. Grâce à la réduction de charge par transfert de proton (PTCR), les spectres obtenus sont moins complexes, limitant le risque de chevauchement. Ainsi, la déconvolution est plus efficace, permettant une meilleure identification des fragments. Dans le cas du T-DXd, 3 fragments diagnostiques additionnels ont été identifiés sur la chaîne légère et 4 sur la sous-unité Fd, venant ainsi confirmer la position des agents cytotoxiques (**Figure 15.D**). La recherche des fragments internes en plus des fragments terminaux rend également possible l'attribution d'un nombre beaucoup plus important d'ions fragments, engendrant de ce fait une augmentation de la couverture de séquence. Dans cette étude de cas, les fragments internes ont permis d'obtenir un recouvrement de séquence de 100% pour chacune des sous-unités. Cependant, ils doivent être considérés avec précaution en raison des multiples possibilités d'attribution et du risque de faux positifs²³.

La totalité des approches utilisées dans cette sous-partie ont démontré les capacités de la spectrométrie de masse pour la caractérisation exhaustive d'un ADC, allant de sa masse intacte et du nombre moyen de molécules cytotoxiques conjuguées à un séquençage complet de la structure primaire et une localisation précise des sites de conjugaison des agents cytotoxiques.

Conclusion

Ce travail de thèse m'a permis d'obtenir une expertise dans le domaine de la spectrométrie de masse appliquée à la caractérisation de protéines thérapeutiques et de leurs impuretés.

La première partie de ce travail comprend de nombreux développements méthodologiques en protéomique quantitative. Des optimisations des paramètres d'acquisition spécifiques à la géométrie tribride ont été menées. Différentes méthodes DIA basées sur des schémas d'isolation fixes et variables ont été testées. Les méthodes optimisées ont été évaluées par l'analyse d'une gamme d'échantillons standards maîtrisés ajoutés dans un mélange complexe à différentes concentrations. De plus, l'apport de la mobilité ionique via l'utilisation du FAIMS a été démontré, avec une hausse significative du nombre de protéines identifiées. De même, l'intérêt de méthodes MS3 pour des approches de quantification par marquage a été apprécié. Dans un second temps, les avancées bioinformatiques liées au traitement des données ont été discutées, montrant le potentiel bénéfique de ces approches pour augmenter le nombre d'identifications et la confiance en celles-ci.

La seconde partie de ce manuscrit est dédiée à l'application de méthodes de protéomique quantitative dans le contexte exigeant de la caractérisation des impuretés de type HCP dans des produits médicaments finaux. Le potentiel de la spectrométrie de masse comme nouvelle méthode de référence pour l'identification et la quantification précise des HCP a été souligné.

La troisième partie de ce manuscrit s'articule autour du développement d'approches « top-down » et « middle-down » pour la caractérisation exhaustive de la structure primaire des protéines thérapeutiques. Les premières optimisations ont été réalisées sur des protéines de référence, permettant de mettre en avant les paramètres clés pour ces stratégies. Ensuite, des optimisations additionnelles ont été réalisées sur un mAb. L'ajout de la PTCLR pour la simplification des spectres de fragmentation a également été examiné. Dans un second temps, l'influence des paramètres de déconvolution et d'identification des fragments a été déterminée, permettant d'établir un compromis entre haut recouvrement de séquence et faible risque de faux positifs. De plus, le potentiel gain de couverture de séquence par l'utilisation des fragments internes a été mis en avant. Pour terminer, l'ensemble des optimisations réalisées a été appliqué à l'étude du trastuzumab deruxtecan. Combinées à des approches en conditions natives et dénaturantes, ces stratégies ont permis la caractérisation complète de l'anticorps conjugué, incluant sa masse intacte, le nombre moyen de charges utiles, la séquence primaire et les sites de conjugaison.

Références

- (1) Kaplon, H.; Chenoweth, A.; Crescioli, S.; Reichert, J. M. Antibodies to watch in 2022. *MAbs* **2022**, *14* (1), 2014296. DOI: 10.1080/19420862.2021.2014296
- (2) Guidance for Industry Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products. (ICH), I. C. f. H. o. T. R. f. P. f. H. U., Ed.; 1999.
- (3) 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. DOI: 10.1002/bit.27808
- (4) <1132> Residual Host Cell Protein Measurement in Biopharmaceuticals. Pharmacopoeia, U. S., Ed.; 2016; Vol. 39.
- (5) 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), 0111.016717. DOI: 10.1074/mcp.0111.016717
- (6) Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J. J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem* **2018**, *90* (15), 9529-9537. DOI: 10.1021/acs.analchem.8b02233
- (7) Silva AS C.; Bouwmeester, R.; Martens, L.; Degroeve, S. Accurate peptide fragmentation predictions allow data driven approaches to replace and improve upon proteomics search engine scoring functions. *Bioinformatics* **2019**, *35* (24), 5243-5248. DOI: 10.1093/bioinformatics/btz383
- (8) Deslignière, E.; Diemer, H.; Erb, S.; Coliat, P.; Pivot, X.; Detappe, A.; Hernandez-Alba, O.; Cianférani, S. A Combination of Native LC-MS Approaches for the Comprehensive Characterization of the Antibody-Drug Conjugate Trastuzumab Deruxtecan. *Front Biosci (Landmark Ed)* **2022**, *27* (10), 290. DOI: 10.31083/j.fbl2710290
- (9) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal Chem* **2018**, *90* (14), 8421-8429. DOI: 10.1021/acs.analchem.8b00984
- (10) Stephenson, J. L., Jr.; McLuckey, S. A. Simplification of product ion spectra derived from multiply charged parent ions via ion/ion chemistry. *Anal Chem* **1998**, *70* (17), 3533-3544. DOI: 10.1021/ac9802832
- (11) Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A. ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J Proteome Res* **2021**, *20* (4), 1928-1935. DOI: 10.1021/acs.jproteome.0c00952
- (12) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003**, *75* (8), 1895-1904. DOI: 10.1021/ac0262560
- (13) McAlister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wühr, M.; Huttlin, E. L.; Erickson, B. K.; Rad, R.; Haas, W.; Gygi, S. P. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem* **2014**, *86* (14), 7150-7158. DOI: 10.1021/ac502040v
- (14) Furtwängler, B.; Üresin, N.; Motamedchaboki, K.; Huguet, R.; Lopez-Ferrer, D.; Zabrouskov, V.; Porse, B. T.; Schoof, E. M. Real-Time Search-Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. *Mol Cell Proteomics* **2022**, *21* (4), 100219. DOI: 10.1016/j.mcpro.2022.100219
- (15) Declercq, A.; Bouwmeester, R.; Hirschler, A.; Carapito, C.; Degroeve, S.; Martens, L.; Gabriels, R. MS(2)Rescore: Data-driven rescoring dramatically boosts immunopeptide identification rates. *Mol Cell Proteomics* **2022**, 100266. DOI: 10.1016/j.mcpro.2022.100266

- (16) Degroeve, S.; Gabriels, R.; Velghe, K.; Bouwmeester, R.; Tichshenko, N.; Martens, L. ionbot: a novel, innovative and sensitive machine learning approach to LC-MS/MS peptide identification. *bioRxiv* **2022**, 2021.2007.2002.450686. DOI: 10.1101/2021.07.02.450686
- (17) 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. DOI: 10.1038/s41592-019-0638-x
- (18) 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. DOI: 10.1074/mcp.P119.001913
- (19) Dhenin, J.; Dupré, M.; Druart, K.; Krick, A.; Mauriac, C.; Chamot-Rooke, J. A multiparameter optimization in middle-down analysis of monoclonal antibodies by LC-MS/MS. *J Mass Spectrom* **2023**, *58* (3), e4909. DOI: 10.1002/jms.4909
- (20) Kline, J. T.; Mullen, C.; Durbin, K. R.; Oates, R. N.; Huguet, R.; Syka, J. E. P.; Fornelli, L. Sequential Ion-Ion Reactions for Enhanced Gas-Phase Sequencing of Large Intact Proteins in a Tribrid Orbitrap Mass Spectrometer. *J Am Soc Mass Spectrom* **2021**, *32* (9), 2334-2345. DOI: 10.1021/jasms.1c00062
- (21) Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat Biotechnol* **2001**, *19* (10), 952-957. DOI: 10.1038/nbt1001-952
- (22) Mishima, S.; Shitara, K. Trastuzumab deruxtecan for the treatment of HER2-positive gastric cancer. *Expert Opin Biol Ther* **2021**, *21* (7), 825-830. DOI: 10.1080/14712598.2021.1912007
- (23) Dunham, S. D.; Wei, B.; Lantz, C.; Loo, J. A.; Brodbelt, J. S. Impact of Internal Fragments on Top-Down Analysis of Intact Proteins by 193 nm UVPD. *J Proteome Res* **2022**. DOI: 10.1021/acs.jproteome.2c00583

TABLE OF CONTENTS

TABLE OF CONTENTS	15
LIST OF COMMUNICATIONS	21
MAIN ABBREVIATIONS	23
GENERAL INTRODUCTION	27
GENERAL INTRODUCTION	29
BIBLIOGRAPHIC INTRODUCTION	31
CHAPTER 1: INTRODUCTION TO QUANTITATIVE BOTTOM-UP PROTEOMICS	33
1. <i>Mass-spectrometry-based proteomic approaches</i>	33
2. <i>Bottom-up strategy analytical workflow</i>	34
A. Sample preparation.....	35
i) Protein extraction.....	35
ii) Mixture preparation and enzymatic digestion	35
B. Liquid chromatography coupled to tandem mass spectrometry	37
i) Peptide separation using reverse phase liquid chromatography	37
ii) Tandem mass spectrometry.....	37
a) Data-dependent acquisition.....	38
b) Fragmentation techniques.....	39
c) Ion mobility	40
C. Data processing.....	42
i) Protein sequence databases.....	42
ii) Search engines.....	42
iii) Validation of the results.....	44
iv) Machine learning-based strategies in proteomics	44
3. <i>Global quantification strategies in bottom-up proteomics</i>	45
A. Label-based relative quantification approaches.....	46
i) Metabolic labeling.....	46
ii) Chemical labeling.....	46
B. Label-free relative quantification approaches	48
i) Spectral counting	48
ii) Extracted ion chromatogram (XIC-MS1).....	48
C. Label-free “absolute” quantification.....	48
4. <i>Data-independent acquisition approaches</i>	49
A. Principle of data independent acquisition	49
B. Evolution of DIA-based approaches	49
i) DIA strategies performed over the entire mass range	50
ii) DIA strategies based on isolation windows	51
C. Development of DIA methods	52
D. The challenge of DIA data processing.....	53
i) Peptide-centric approach.....	53
ii) Spectrum-centric approaches	54
CHAPTER 2: THE ROLE OF TOP-DOWN MASS SPECTROMETRY IN THE CONTEXT OF STRUCTURAL MS	55
1. <i>Protein-centric approaches in structural MS</i>	56
A. Native MS.....	56
B. Other approaches in structural mass spectrometry.....	57
2. <i>Peptide-centric approaches</i>	58
A. Hydrogen-Deuterium Exchange (HDX-MS).....	58

B.	Cross-link MS (XL-MS)	58
C.	Limited proteolysis	59
3.	<i>Protein primary structure characterization: a peptide- or a protein-centric approach?</i>	59
A.	Bottom-up versus top-down	59
B.	Definition of top-down and middle-down MS	60
4.	<i>Instrumentation and methods for top-down and middle-down MS</i>	62
A.	MS instrumentation	62
B.	Fragmentation techniques	62
i)	Collision-based dissociation	63
ii)	Electron-based dissociation	64
iii)	Photodissociation	65
iv)	Hybrid fragmentation techniques	66
v)	Proton transfer charge reduction	66
C.	Data processing	67
i)	MS/MS spectra deconvolution	67
ii)	Fragments assignment to a candidate sequence	67
5.	<i>Main applications of top-down MS</i>	68
A.	Proteoform identification in complex mixtures	68
B.	Purified intact proteins analysis	69
C.	Native top-down approaches	69
CHAPTER 3: INTRODUCTION TO MONOCLONAL ANTIBODY-RELATED THERAPEUTIC PROTEINS		71
1.	<i>Monoclonal antibodies and derivative biotherapeutics</i>	71
A.	Structure of immunoglobulins	71
B.	Diversity of mAb-related therapeutic proteins	72
i)	Evolution of mAb-related products	72
ii)	Antibody-drug conjugates	73
C.	Characterization of mAbs and mAbs derivatives	73
i)	Intact mass measurement and middle-up experiments	75
ii)	Top-down and middle-down analysis in the characterization of mAbs and ADCs	76
2.	<i>The concern of host cell proteins in monoclonal antibodies</i>	78
A.	Monoclonal antibodies manufacturing process	78
i)	Expression system	78
ii)	Manufacturing process	78
a)	Upstream process	79
b)	Downstream process	79
B.	Host cell proteins monitoring	80
i)	Immuno-specific approaches	80
a)	ELISA	81
b)	Western blot	82
ii)	Non-specific approaches	82
a)	Gel electrophoresis	83
b)	Mass spectrometry	83
iii)	Comparison of HCP monitoring approaches	84
PART I METHODOLOGICAL DEVELOPMENTS IN QUANTITATIVE PROTEOMICS		87
CHAPTER 1: OPTIMIZATION OF LC-MS/MS ACQUISITION METHODS FOR QUANTITATIVE PROTEOMICS		89
1.	<i>Context of the project</i>	89
2.	<i>Analytical strategy applied for method optimizations</i>	89
3.	<i>Evaluation of the acquisition parameters on a tribrid instrument</i>	91
A.	Comparison of Orbitrap and ion trap analyzers for data dependent acquisition	91
B.	Comparison of HCD, CID, ETD and EThcD fragmentation techniques	93

C.	Optimization of the isolation windows scheme for DIA	95
D.	Application of the optimized methods to a mixture of well calibrated samples	97
4.	<i>Optimization of high-field asymmetric waveform ion mobility spectrometry (FAIMS) compensation voltages for quantitative proteomics</i>	100
5.	<i>Evaluation of acquisition strategies dedicated to TMT-labelled samples</i>	104
A.	Optimization of the fragmentation energy	104
B.	Comparison of MS2 and MS3 methods for TMT quantification	106
6.	<i>Conclusion and perspectives</i>	109
CHAPTER 2: EVALUATION OF BIOINFORMATICS TOOLS FOR DATA PROCESSING		111
1.	<i>Context of the project</i>	111
2.	<i>Analytical strategy applied for data processing optimizations</i>	112
3.	<i>Evaluation of innovative tools for DDA data processing</i>	113
A.	Evaluation of rescoring algorithms for peptides/proteins identification	114
i)	MS ² Rescore: an open-source algorithm for peptide spectrum matches rescoring.....	114
ii)	Inferys rescoring for Sequest HT data	116
B.	Chimerys, a new artificial intelligence-based search engine.....	118
C.	Evaluation of ionbot: an open modification search algorithm.....	121
4.	<i>Evaluation of library-free approaches in DIA</i>	123
5.	<i>Conclusions and perspectives</i>	127
PART II METHODOLOGICAL DEVELOPMENTS FOR THE IDENTIFICATION AND QUANTIFICATION OF TRACE LEVEL HOST CELL PROTEIN IMPURITIES IN DRUG PRODUCTS		129
1.	<i>Context of the project</i>	131
2.	<i>Analytical search</i>	131
3.	<i>Implementation of a stringent data filtering workflow</i>	133
4.	<i>Internal calibration curve-based quantification of host cell proteins</i>	135
5.	<i>Evaluation of FAIMS DDA methods to dig deeper into the HCP landscape</i>	135
6.	<i>Evaluation of DIA strategies for HCP characterization in drug products</i>	138
A.	Optimizations of acquisition methods.....	138
B.	Influence of the quantity injected and comparison to DDA	141
C.	Use of a chromatogram library from gas phase fractionation for DIA data processing	143
7.	<i>Evaluation of open-modification search tools for HCPs characterization</i>	146
8.	<i>Conclusion and perspectives</i>	147
PART III STRUCTURAL MASS SPECTROMETRY TO CHARACTERIZE MAB-BASED BIOTHERAPEUTICS		149
CHAPTER 1: WORKFLOW OPTIMIZATION OF TOP-DOWN AND MIDDLE-DOWN MASS SPECTROMETRY APPROACHES		151
1.	<i>Context of the project and analytical strategy</i>	151
2.	<i>Optimization of the key parameters for top-down data acquisition by infusion of reference proteins</i>	153
A.	Influence of the fragmentation parameters	153
i)	Optimization of the precursor ion charge and intensity	154
ii)	Optimization of the fragmentation parameters	156
iii)	Parallel optimization between precursor ion charge and energy reaction for an enhanced sequence coverage.....	158
B.	Development of MS3 approaches to increase protein sequence coverage	160
3.	<i>Optimization of the acquisition parameters for middle-down LC-MS/MS analysis of monoclonal antibody</i>	162
A.	Implementation of LC separation before MD-MS for mAb analysis	162
i)	Influence of precursor ion multiplexing.....	163

ii)	Influence of the isolation window width.....	164
iii)	Optimization of the LC separation to improve sequence coverage.....	165
B.	Complementarity and repeatability of mAb subunits fragmentation by LC-MD/MS.....	165
4.	Implementation of proton transfer charge reduction (PTCR) for mAb analysis.....	168
5.	Conclusion and perspectives.....	171
CHAPTER 2: IMPACT OF DECONVOLUTION PARAMETERS AND FRAGMENT ION MATCHING IN MIDDLE-DOWN MS WORKFLOWS.....		173
1.	Context of the project and analytical strategy.....	173
2.	Evaluation of the risk of false positives.....	174
3.	Evaluation of the deconvolution parameters.....	176
A.	Influence of the S/N ratio.....	177
B.	Influence of the fit factor threshold.....	178
4.	Evaluation of the mass tolerance threshold for fragments assignment.....	180
5.	Consideration of internal fragments to increase sequence coverage.....	181
A.	Evaluation of the mass tolerance threshold for internal fragment matching.....	181
B.	Evaluation of internal fragments for sequence modification localization.....	184
6.	Conclusion and perspectives.....	186
CHAPTER 3: CASE STUDY: CHARACTERIZATION OF TRASTUZUMAB DERUXTECAN ANTIBODY-DRUG CONJUGATE BY MD-MS.....		189
1.	Context of the project and analytical strategy.....	189
2.	Intact and middle-level analysis of T-DXd for drug-load-distribution (DL D) determination and payload location assessment at the subunit level.....	190
A.	Intact and middle-up level analysis of T-DXd in SEC-nMS.....	190
B.	Intact and middle-up level analysis of T-DXd in denaturing conditions.....	192
3.	MD-MS analysis for sequence characterization and precise drug location.....	193
A.	T-DXd Lc and Hc subunits MD-MS analysis.....	194
B.	T-DXd Fc/2, Lc and Fd subunits MD-MS analysis.....	195
4.	Contribution of internal fragments for T-DXd subunits sequence characterization.....	198
5.	Strengthening drug location using PTCR.....	200
6.	Conclusion and perspectives.....	202
GENERAL CONCLUSION		203
GENERAL CONCLUSION		205
EXPERIMENTAL SECTION.....		209
EXPERIMENTAL SECTION.....		211
1.	Methodological development in quantitative proteomics.....	211
A.	Optimization of LC-MS/MS acquisition methods.....	211
i)	Evaluation of the acquisition parameters.....	211
a)	Comparison of Orbitrap and ion trap analyzers.....	211
b)	Sample preparation.....	211
c)	Comparison of HCD, CID, ETD and EThcD fragmentation techniques.....	212
d)	Optimization of the isolation windows scheme for DIA.....	212
e)	Application to a mixture of well calibrated samples.....	213
ii)	Optimization of FAIMS compensation voltages.....	214
a)	Sample preparation.....	214
b)	LC-MS/MS analyses.....	214
c)	Data analysis.....	215
iii)	Evaluation of acquisition strategies dedicated to TMT samples.....	215
a)	Sample preparation.....	215
b)	LC-MS/MS analysis for collision energy evaluation.....	215

c)	LC-MS/MS analysis for TMT acquisition method evaluation.....	215
d)	Data processing.....	216
B.	Evaluation of bioinformatics tools for data processing.....	216
i)	Evaluation of innovative tools for DDA data processing.....	216
a)	MS ² Rescore.....	216
b)	Inferys.....	217
c)	Chimerys.....	217
d)	Ionbot.....	218
ii)	Comparison of library-free approaches in DIA.....	218
a)	Spectronaut.....	218
b)	DIA-NN.....	218
2.	<i>Methodological developments for HCP quantification in drug products.....</i>	219
A.	HCP-PROFILER for internal calibration.....	219
B.	Evaluation of FAIMS DDA methods.....	219
i)	Sample preparation.....	219
ii)	LC-MS/MS analysis.....	220
iii)	Data analysis.....	220
C.	Evaluation of DIA strategies for HCP quantification.....	220
i)	Sample preparation.....	220
ii)	LC-MS/MS analysis.....	221
iii)	Data analysis.....	221
3.	<i>Top-down and middle-down mass spectrometry to characterize mAb-based therapeutics ...</i>	222
A.	Optimization of top-down and middle-down acquisition parameters.....	222
i)	Optimization of key parameters for top-down acquisition by infusion.....	222
a)	Sample preparation.....	222
b)	MS/MS analysis.....	222
c)	Data analysis.....	222
ii)	Optimization of the acquisition parameters for MD-MS analysis of mAbs.....	222
a)	Sample preparation.....	222
b)	LC-MS/MS analysis.....	223
c)	Data analysis.....	224
iii)	Assessment of data processing for TD-MS and MD-MS experiments.....	224
a)	Sample preparation.....	224
b)	LC-MS/MS analysis.....	224
c)	Data analysis.....	224
B.	Characterization of trastuzumab deruxtecan ADC by MD-MS.....	225
i)	Sample preparation.....	225
a)	Middle-up analysis sample preparation.....	225
b)	Middle-down analysis sample preparation.....	225
ii)	LC-MS/MS analysis.....	225
a)	SEC-nMS analysis experiments.....	225
b)	Middle-level LC-MS and LC-MS/MS analysis.....	225
iii)	Data analysis.....	226
a)	SEC-nMS analyses.....	226
b)	MD-MS analysis.....	226
	ANNEXES.....	227
	REFERENCES.....	230
	REFERENCES.....	232

LIST OF COMMUNICATIONS

Publications

Beaumal, C., Beck, A., Hernandez-Alba, O., & Carapito, C. (2023). 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*, 23, e2300172. DOI: 10.1002/pmic.202300172

Beaumal, C., Deslignière, E., Diemer, H., Carapito, C., Cianférani, S., & Hernandez-Alba, O. (2023). Improved characterization of trastuzumab deruxtecan with PTCR and internal fragments implemented in middle-down MS workflows. *Analytical and Bioanalytical Chemistry*. DOI: 10.1007/s00216-023-05059-x.

Selected oral communications

70th American Society for Mass Spectrometry (ASMS) annual Conference, June 5th – 9th 2022, Minneapolis, United States.

Accurate quantification of trace-level Host Cell Protein impurities in biotherapeutics using FAIMS separation and new data independent acquisition approaches.

Corentin Beaumal, Alain Beck, Oscar Hernandez-Alba, Christine Carapito.

PhD Students Day, November 7th 2022, Strasbourg, France

Accurate quantification of trace-level Host Cell Protein impurities in biotherapeutics using FAIMS separation and new data independent acquisition approaches.

Corentin Beaumal, Oscar Hernandez-Alba, Christine Carapito.

Selected oral communications

6th Annual Bioprocessing Summit, March 14th – 15th 2023, Barcelona, Spain.

Advanced LC-MS/MS Workflows for HCP Quantification Including Optimized Standards, Ion Mobility Separation and Data Independent Acquisition (DIA) MS.

Corentin Beaumal, Christine Carapito.

Thermo Fisher Scientific Customer Event, October 20th 2022, Illkirch-Graffenstaden, France

Bottom-up and top-down method developments on an Orbitrap Eclipse Tribrid instrument for characterization of biotherapeutics and their impurities.

Corentin Beaumal, Oscar Hernandez-Alba, Christine Carapito.

Poster communications

69th American Society for Mass Spectrometry (ASMS) annual Conference, November 17th 2021, online. *Host Cell Protein impurities characterization in biotherapeutics using finely tuned mass spectrometry-based workflows.*

Corentin Beaumal, Segal Ndiaye, Claire Dauly, Oscar Hernandez-Alba, Christine Carapito.

Proteomic Forum – XIVth Annual Congress of the European Proteomics Association (EuPA), April 3rd – 7th 2022, Leipzig, Germany.

Finely tuned mass spectrometry-based workflows for trace-level Host Cell Protein impurities characterization in biotherapeutics.

Corentin Beaumal, Oscar Hernandez-Alba, Christine Carapito.

Analytics 2022, September 5th – 8th 2022, Nantes, France.

Complete characterization and accurate drug conjugation site's location of trastuzumab deruxtecan using middle middle-down mass spectrometry.

Corentin Beaumal, Evolène Deslignière, Christine Carapito, Sarah Cianférani, Oscar Hernandez-Alba.

Journées de la Société Française de Spectrométrie de Masse (JSFSM) 2023, September 5th – 8th 2023, Marseille, France.

Complete characterization and accurate drug conjugation site's location of trastuzumab deruxtecan using middle middle-down mass spectrometry.

Corentin Beaumal, Evolène Deslignière, Christine Carapito, Sarah Cianférani, Oscar Hernandez-Alba.

MAIN ABBREVIATIONS

ACN	Acetonitrile
ADC	Antibody-drug conjugate
AGC	Automatic gain control
AI	Artificial intelligence
Astral	Asymmetric Track Lossless
avDAR	Average drug-to-antibody ratio
BLAST	Basic local alignment search tool
CID	Collision induced dissociation
CQA	Critical quality attribute
CRIGR	<i>Cricetulus griseus</i>
CV	Compensation voltage
Da	Dalton
DAR	Drug-to-antibody ratio
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DLD	Drug load distribution
DP	Drug product
DTT	Dithiothreitol
ECD	Electron capture dissociation
ELISA	Enzyme linked immunosorbent assay
EMA	European Medicines Agency
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FA	Formic acid
Fab'	Fragment antigen-binding
FAIMS	Field asymmetric ion mobility spectrometry
FC	Fold change
Fc	Fragment crystallizable region
FDA	U.S. Food and drug administration

FDR	False discovery rate
GPF	Gas-phase fractionation
Hc	Heavy chain
HCD	Higher-energy collision dissociation
HCP	Host cell proteins
HMWS	High molecular weight species
HRAM	High-resolution accurate mass
IAM	Iodoacetamide
IM-MS	Ion Mobility-Mass spectrometry
IT	Ion trap
Lc	Light chain
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
LFQ	Label-free quantification
LWMS	Low molecular weight species
m/z	Mass over charge ratio
mAb	Monoclonal antibody
MALDI	Matrix-Assisted Laser Desorption Ionization
MD	Middle-down
ML	Machine learning
MS	Mass spectrometry
MW	Molecular weight
NCE	Normalized collision energy
NIST	National institute of standards and technology
nMs	Native mass spectrometry
OMS	Open modification search
OT	Orbitrap
PASEF	Parallel accumulation - serial fragmentation
PD	Proteome Discoverer
PMF	Peptide mapping fingerprint
ppm	Parts per million
PRM	Parallel reaction monitoring
PSM	Peptide-spectrum match

PTCR	Proton transfer charge reduction
PTM	Post-translational modification
Q	Quadrupole analyzer
RF	Radiofrequency
RT	Retention time
RTS	Real-time search
S/N	Signal over noise ratio
SEC	Size exclusion chromatography
SPS	Simultaneous precursor selection
SVM	Support vector machine
SWATH	Sequential windowed acquisition of all theoretical fragment ion spectra
TCEP	Tris(2-carboxyethyl)phosphine
TD	Top-down
T-DXd	Trastuzumab deruxtecan
TIMS	Trapped ion mobility spectrometry
TMT	Tandem mass tag
ToF	Time-of-flight
UPS	Universal proteomics standard
UVPD	Ultraviolet photodissociation
V	Volt
XIC-MS1	Extracted ion chromatogram - MS1 filtering
HDX	Hydrogen-deuterium exchange
XL-MS	Cross-link mass spectrometry
UHPLC	Ultra high-pressure liquid chromatography
TFA	Trifluoroacetic acid

GENERAL INTRODUCTION

General introduction

Proteins are biological macromolecules composed of a succession of amino acids coded by genes. They are key players in many biological processes and perform various essential functions. The proteome is a complex and dynamic entity, defined as all the proteins in a cell, a tissue, or an organism at a particular time and under defined conditions. Although around 20,000 protein-coding genes have been identified in humans, hundreds of thousands of protein forms have been observed¹. The multiple molecular forms of an expressed protein, resulting from various processes, including genomic sequence variants, splice variants, mistranslation events, or post-translational modifications, are called proteoforms². In addition, the protein content is dynamic as it varies depending on the time, the environmental conditions, or the location. Among the wide variety of proteins, monoclonal antibodies (mAb) therapeutic proteins and the related products are particularly interesting for their applications in various fields, including oncology or autoimmune disorders. It is the fastest-growing class of therapeutics, with more than 120 mAb-related products approved³ and a market size of around \$200 billion in 2022. Before being released, critical quality attributes (CQAs) such as the mAb primary sequence, its structure, its physico-chemical properties, and the impurities of the mAbs drug product – including host cell protein (HCP) impurities from the manufacturing process – should be thoroughly characterized.

In this context, mass spectrometry (MS) has become the method of choice for the characterization of proteins since the introduction of electrospray ionization⁴ (ESI) and matrix-assisted laser desorption ionization⁵ (MALDI) in combination to advances in sample preparation and data processing. Recent years have seen significant improvements in the resolution, precision, sensitivity, and acquisition speed of mass spectrometers, the fragmentation techniques available, and the development of numerous increasingly powerful bioinformatics tools dedicated to data processing. Multiple strategies have been proposed for protein analysis using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). Quantitative bottom-up proteomics approaches – which rely on analyzing peptides from the proteolytic digestion of proteins – are widely used. The development of acquisition modes such as data-independent acquisition⁶ (DIA) and the improvement of algorithms for signal extraction allow the quantification of thousands of proteins in a few hours⁷. Nevertheless, this approach is not well suited for analyzing PTMs and proteoforms identification, as co-existing PTMs on the same protein cannot be evidenced. Alternatively, top-down approaches are based on analyzing intact proteins that are directly fragmented into the mass spectrometer. With this strategy, the intact mass of the protein is conserved, allowing the identification of the proteoforms⁸ thanks to specific fragment ions. This strategy is particularly adapted for mAb analysis, enabling the extensive determination of the primary sequence and the precise location of PTMs^{9, 10}.

In this context, my PhD work has focused on (i) methodological developments and evaluation of analytical and bioinformatics strategies in mass spectrometry-based quantitative proteomics, (ii) their application for the identification and the quantification of host-cell proteins in drug products, and (iii) the development of top-down and middle-down approaches for the characterization of therapeutic proteins.

This manuscript is therefore structured in five parts summarized below:

- The **Bibliographic introduction** overviews state-of-the-art quantitative proteomics and intact protein characterization using mass spectrometry. It includes a description of the general bottom-up proteomics workflow, from sample preparation to data processing, through the different acquisition modes. It introduces protein characterization at the intact level with a special focus on top-down and middle-down strategies.

Then, the importance of therapeutic proteins and the challenge of host protein impurities monitoring are detailed.

- **Part I** combines methodological developments in quantitative proteomics to quantify proteins in complex mixtures.
 - **Chapter 1** evaluates the acquisition parameters for DDA and DIA analysis on a tribrid instrument for the quantification of proteins in calibrated samples. The implementation of the ion mobility using the FAIMS is optimized and dedicated acquisition strategies for TMT samples are assessed.
 - **Chapter 2** compares innovative tools based on machine learning and artificial intelligence for DDA and DIA data processing. Two DDA strategies relying on rescoring algorithms and open modification search are compared, and library-free approaches using machine learning for DIA data analysis are examined.
- **Part II** presents the optimizations of the previously developed acquisition parameters in the context of trace-level host-cell protein impurities quantification drug products, and the performances of FAIMS and DIA approaches are underlined. Data processing using open modification search is also evaluated in this context, for identifying additional HCPs.
- **Part III** focuses on developing top-down and middle-down approaches to characterize intact proteins.
 - **Chapter 1** details optimizations of top-down and middle-down key acquisition parameters, including precursor ion selection and fragmentation mode. Fragmentation of intact protein using infusion or after chromatographic separation was tested on reference proteins and mAbs. PTCR reaction is also evaluated to simplify complex MS/MS spectra and help identify more fragments.
 - In **Chapter 2**, the influence of the data processing parameters during the deconvolution and the fragments identification steps is described on therapeutic proteins. Additionally, internal fragments' contribution to improving the sequence coverage in top-down and middle-down strategies is appraised.
 - In **Chapter 3**, the optimized acquisition parameters and data processing strategies are applied to characterize an ADC, trastuzumab deruxtecan. After the characterization of the global drug-to-antibody ratio in native conditions, middle-down strategies are employed for precise drug conjugation site location with the help of PTCR.
- The last part corresponds to the **Experimental section**, in which experimental details on the different works discussed in the manuscript are reported.

This Ph.D. work has been funded by the French Ministry of Higher Education and Research.

BIBLIOGRAPHIC INTRODUCTION

Chapter 1:

Introduction to quantitative bottom-up proteomics

The term *proteome* was first introduced in 1994 by Mark Wilkins¹¹. It comes from the combination of the notions of *protein* and *genome*, as it is defined as a set of proteins encoded by a genome. Whereas the genome represents all the genetic information of an organism, the proteome designates the entire set of proteins contained in a living entity under given conditions and at a given time, implying in this case the notion of dynamism¹². By analogy to *genomics* – the study of the genome – the term *proteomics* was coined in 1997 by Peter James¹³. It can be defined as the qualitative, quantitative and functional study of all proteins contained in a cell or a tissue at a specific time and under specific conditions.

Mass spectrometry (MS) has become a tool of choice for proteomics analysis thanks to numerous major breakthrough¹⁴⁻¹⁷. First, the development in the late 1980's of soft ionization sources represents a cornerstone for the study of biological macromolecules, such as proteins, by MS. Namely, MALDI (Matrix-Assisted Laser Desorption Ionization)⁵ and ESI (Electrospray ionization)⁴ have been proposed by Koichi Tanaka (Japan) and John B. Fenn (USA), respectively, who received the Nobel Prize in Chemistry in 2002, for these innovations. Second, the advances for protein and peptide separation prior to MS analysis combined to continuous improvements of mass spectrometers with enhanced sensitivity, dynamic range, resolution, mass accuracy and acquisition speed performances, leading to the development of new acquisition modes has drastically expanded the possibilities of MS in proteomics. Third, the production of large-scale protein databases of high quality, i.e., well-annotated and curated, have enabled the confident identification of proteins. Finally, the constant progress and breakthrough of bioinformatics tools for the analysis of large amounts of MS data aiming toward the identification, quantification functional interpretation of proteins bring new perspectives for MS-based proteomics.

1. Mass-spectrometry-based proteomic approaches

Mass spectrometry-based proteomics can be divided into three approaches: bottom-up, middle-down and top-down (**Figure 1**).

Bottom-up proteomics aims at identifying and quantifying thousands of proteins in complex mixtures via their peptides (< 3 kDa). Proteins are enzymatically digested into peptides and subsequently separated by liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS)¹⁸. Conversely, top-down proteomics refers to the characterization of intact protein directly analyzed using LC-MS/MS without any prior digestion step. It is particularly well suited for PTM analysis, alternative RNA splicing, and proteoforms identification and quantification^{19, 20}. In middle-down proteomics, proteins undergo limited proteolysis to generate large peptides (3-10 kDa) that are separated and analyzed in the same way as the bottom-up approach. Identification of longer peptides allows an increased sequence coverage and provides more information about co-occurring neighboring PTMs^{21, 22}.

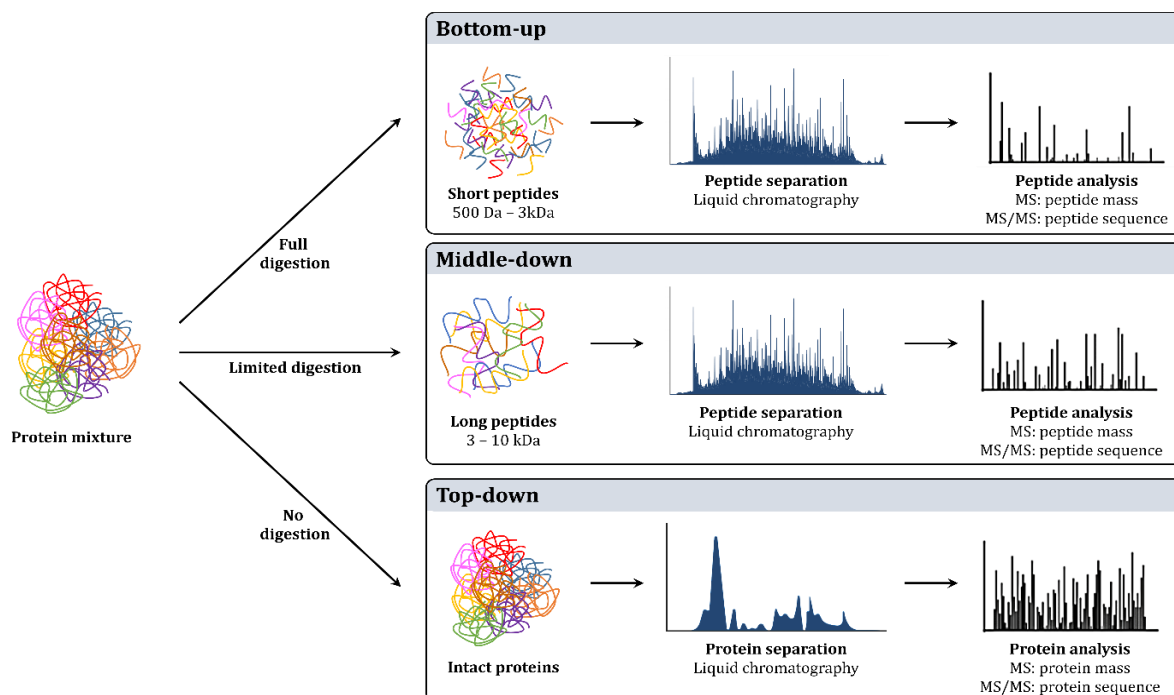


Figure 1: Global representation of the MS-based approaches in proteomics: bottom-up, middle-down and top-down.

Despite the growing popularity of top-down and middle-down proteomics approaches, driven by the numerous advances allowing complex mixture analysis, they are still mainly used for specific applications, like PTMs location and proteoform detection and characterization. The bottom-up approach remains the method of choice for large-scale and high-throughput proteomics studies, mainly thanks to the maturity of MS workflows and data processing tools dedicated to this strategy.

A part of the work related to the characterization of proteins in complex mixtures presented in this manuscript is based on bottom-up proteomics, detailed in this chapter. Another part is related to the characterization of purified proteins, using top-down and middle-down MS approaches. These techniques are described in **Chapter 2** of the bibliographic introduction.

2. Bottom-up strategy analytical workflow

The bottom-up approach relies on characterizing proteins via their peptides. Proteins are extracted from the sample and undergo enzymatic digestion, usually with trypsin²³, to give peptides below 3 kDa. These peptides are separated by LC, ionized, and analyzed by MS to obtain the peptide mass. These ions are selected, undergo fragmentation, and the mass of the fragments is measured (MS/MS). The experimental mass list is searched against a theoretical one, generated from *in silico* digestion of the protein database of the studied organism for peptide identification. Peptides identified are assigned to the proteins they originate from by peptide-to-protein inference, a nontrivial strategy since peptides can be unique to a protein or shared between several²⁴.

The bottom-up proteomics workflow comprises three main steps summarized in **Figure 2**: the sample preparation, the LC-MS/MS analysis, and the data processing.

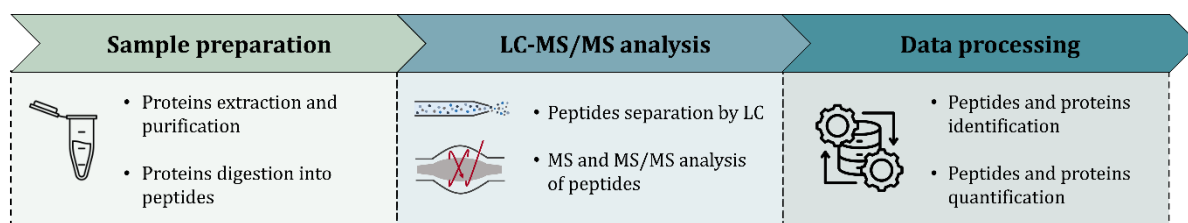


Figure 2: Main steps of a bottom-up proteomics experiment.

Currently, the bottom-up approach is the method of choice for identifying proteins in complex mixtures, as recent LC-MS/MS systems enable the characterization of thousands of proteins in a few hours²⁵. It has been used for a significant part of the work presented in this manuscript and will therefore be detailed below.

A. Sample preparation

Sample preparation is essential in the bottom-up approach but remains a primary challenge with several laborious steps. The preparation's quality and repeatability directly and significantly affect the overall proteomics results. Hence, each step, namely protein extraction, mixture preparation and enzymatic digestion, should be carefully optimized according to the sample and the aim of the analysis²⁶⁻²⁸.

i) Protein extraction

One strength of the bottom-up approach is its ability to be applied to a wide variety of samples, from purified proteins, such as therapeutic proteins²⁹, to whole organisms³⁰, through biofluids^{31, 32}, cells³³ or tissues³⁴. During protein extraction, the objective is to solubilize the maximum of proteins without altering, modifying, or degrading them and ensure their accessibility to proteases for digestion. It requires careful optimizations and must be adapted to the sample type, the protein type, and the following analytical workflow.

Sample lysis and protein extraction can be performed by two approaches: mechanically or chemically²⁷. Based on grinding or ultrasonication, the mechanical approach is generally used for tissues and cells and commonly combined with the chemical approach to maximize protein extraction. The chemical approach uses detergents (e.g., ionic, zwitterionic, non-ionic, salts) to improve protein extraction and solubilization through micelle formation³⁵, as well as chaotropic agents (e.g., urea, thiourea), combined or not to organic solvents, for protein denaturation and unfolding.

Many protein extraction protocols introduce detergents or contaminants to the preparation, in addition to the impurities already present (salts, lipids, etc.). These potentially interfering species must be removed as they can affect the rest of the sample preparation and alter the analysis through MS signal suppression or column capacity and performance reduction, for example. To this end, common strategies employed are protein precipitation using organic solvents (cold acetone, water/methanol/chloroform mixture) with or without organic acid (e.g., trichloroacetic acid), or ultrafiltration or dialysis, to a lesser extent. While necessary for the following of the analysis, this step can induce sample loss³⁶.

ii) Mixture preparation and enzymatic digestion

The protein mixture obtained is complex because of thousands of proteins, proteoforms, and their high-dynamic range. Proteins can be present under several molecular forms, called proteoforms, increasing drastically the number of species in the mixture (e.g., for human, from ~ 20,000 gene-predicted proteins to hundreds of thousands of proteoforms¹). One way to tackle this limitation is to fractionate the mixture based on the physicochemical properties of the proteins³⁷: isoelectric point

(e.g., ion exchange chromatography (IEX) or isoelectric focusing (IEF)), molecular weight (e.g., sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE)), polarity (reverse phase liquid chromatography (rpLC)), hydrodynamic volume (size exclusion chromatography (SEC)) or by combining several of these techniques.

Proteins are also expressed over a wide dynamic range, from 6-9 orders of magnitude in mammalian cells up to 12 orders of magnitude in specific samples like plasma³⁸. Multiple strategies based on the depletion of highly abundant proteins or the enrichment of proteins of interest can be used to enhance the detection of low abundant proteins.

Prior to enzymatic digestion, disulfide bridges – covalent bonds between two cysteine residues and implied in the regulation of the tertiary structure of proteins³⁹ – are reduced (e.g., using dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP)) followed by thiol alkylation (generally by iodoacetamide (IAM)) to avoid disulfide bridges reformation and thiol oxidation. The reduction and alkylation are necessary to ensure the accessibility of the cleavage sites to proteases for the digestion of the proteins.

After reduction and alkylation, proteins are digested into peptides. Trypsin is the most widely used enzyme, available in large quantities and cheaply. It displays a very-high specificity by cleaving peptide bonds after arginine and lysine basic residues (except when followed by proline)⁴⁰. Since these residues are well distributed throughout the protein, tryptic peptides obtained are from 500 to 3,000 Da (5-30 amino acids length) and carry at least two positive charges, ideally suited for MS analysis by favoring ionization and fragmentation⁴¹. Other enzymes with different preferential cleavage sites are also available to obtain a more comprehensive sequencing: ArgC, chymotrypsin, pepsin, GluC, or AspN⁴². In addition, using multiple enzymes can enhance digestion efficiency, such as combining Lys-C with trypsin⁴³.

Various approaches exist for protein digestion: in-gel digestion, in-solution digestion, on-membranes digestion, or bead-based digestion. A brief overview of each protocol is presented below:

- In-solution digestion is straightforward, consisting of adding the enzyme directly to the protein mixture after reduction and alkylation. Buffers compatible with digestion are required but they should be eliminated by solid-phase extraction before MS analysis³⁶.
- In-gel digestion is typically employed in the presence of SDS from the extraction step, known to interfere with chromatographic and MS analysis. It also eliminates detergents, salts, or other contaminants from the samples⁴⁴. Proteins are loaded onto polyacrylamide gels and separated by migration through their molecular weight. Gel bands containing proteins are revealed, collected, and washed. Then, proteins undergo reduction/alkylation and digestion directly in the gel. In-gel digestion has been the method of choice for years, but it is time-consuming and laborious.
- On-membrane digestion methods utilize a membrane to separate proteins from detergents and impurities. After removal of detergents and contaminants, proteins are digested on-membrane, and peptides are eluted for clean-up and MS analysis. This approach is of growing interest, with the development of multiple solutions such as filter-aid sample preparation (FASP)⁴⁵ or suspension trapping (S-trap)⁴⁶, for example.
- On-bead digestion, including single-pot solid-phase-enhanced sample preparation (SP3)⁴⁷, uses hydrophilic carboxylate-coated paramagnetic beads. Reduced and alkylated proteins are trapped on SP3 beads and separated from contaminants using a magnetic rack. Subsequently, proteins are washed to remove contaminants, digested, and peptides are eluted.

B. Liquid chromatography coupled to tandem mass spectrometry

i) Peptide separation using reverse phase liquid chromatography

The mixture obtained after protein digestion is highly complex and contains tens of thousands of peptides. For this reason, liquid chromatography is systematically used to simplify the mixture before MS analysis and, thus, decrease the ionization competition between species. It allows for increasing sensitivity, selectivity, and proteome coverage.

The widely used system for bottom-up proteomics is reverse-phase liquid chromatography in acidic conditions (pH < 2). Before entering the analytical column for separation, peptides are trapped in a trap column for a few minutes using a low acetonitrile (ACN) proportion (~1-3%) to remove the remaining contaminants. Peptides are then separated according to their hydrophobicities by progressively decreasing the mobile phase polarity, composed of an acidified mixture of water and ACN.

For this Ph.D. work, an UltiMate™ 3000 RSLCnano (Thermo Fisher Scientific) system was used. Additional details about the chromatographic setup are described in **Table 1**:

Table 1: Specifications of the chromatographic system used for the work presented in this manuscript.

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

Various parameters influence the chromatographic separation of peptides, such as solvent composition, the flow rate, or the length of the chromatographic gradient. Properties of the column – including its length, internal diameter, particle and pore sizes, and the stationary phase, also play a critical role in the quality of peptide separation. Multiple technical advances have been developed on chromatographic systems to improve separation⁴⁸. Particularly, ultra-high pressure liquid chromatography (UHPLC) enabled the use of nano-columns to work at a shallow flow rate (typically 200-500 nL/min) and high pressure (> 400 bars). Although these nanoflow systems are more prone to leaks, spray instability, or dead volume issues, they drastically increase the peak capacity, resolution, and sensibility of the analyses, leading to more peptide identifications. Moreover, only low amounts of samples (100-400 ng of peptides) are required for nano-LC, making this approach well suited for proteomics due to the limited quantity of samples used.

ii) Tandem mass spectrometry

Along the separation through the nanoLC, peptides are ionized by ESI, a soft ionization source preserving the sample integrity, before entering the mass spectrometer. Charged liquid droplets containing peptides are sprayed from the emitter using a high-voltage (~2 kV). Solvent evaporation results in the formation of desolvated ions (ions in the gas phase) entering the mass spectrometer and guided through the instrument by electromagnetic fields.

Instruments determine the mass-to-charge (m/z) ratio and the intensity of the precursor ion – referred to as a peptide with a specific charge state – in the first step (MS or MS1) and the related fragment ions in the second step (MS2 or MS/MS). This so-called tandem mass spectrometry strategy is enabled by using hybrid or tribrid instruments, which combine different analyzers⁴⁹: quadrupole (Q), time-of-flight (TOF), Orbitrap (OT), ion trap (IT), Astral and Fourier transform-ion cyclotron resonance (FT-ICR). While instruments from various vendors and combining multiple mass analyzers are available on the market, Q-TOF and Q-Orbitrap hybrid geometries are the two most popular for bottom-up proteomics.

All the work presented in this manuscript was carried out on the Orbitrap Eclipse Tribrid (Thermo Fisher Scientific), a tribrid geometry instrument combining quadrupole, ion trap, and Orbitrap mass analyzers. Its main specifications are detailed in **Table 2**.

Table 2: Specifications of the mass spectrometer used for the work presented in this manuscript.

Properties	Orbitrap Eclipse Tribrid
System vendor	Thermo Fisher Scientific
Analyzers	Quadrupole, Orbitrap and ion trap (Tribrid instrument)
Resolution	Orbitrap: 7,500 – 500,000 (at 200 m/z) Ion trap: 0.3-3 m/z FWHM
Mass precision	5 ppm
Scan rate	Orbitrap: up to 40 Hz (at 7,500 resolution) Ion trap: up to 45 Hz (in Turbo mode)
Mass range	Up to 8,000 m/z
Fragmentation type	HCD, CID, ETD, EThcD, ETciD, UVPD
Ion mobility	FAIMS option
Installation year in the lab	2020

a) Data-dependent acquisition

Data-dependent acquisition (DDA) is currently the most used acquisition mode in bottom-up proteomics. It relies on the iterative acquisition cycles of precursor ion MS (or MS1) and fragment ions MS/MS (or MS2) spectra along the chromatographic gradient. In the DDA approach, precursor ion selection is based on their intensity. As detailed in **Figure 3**, the N most intense precursor ions (TopN strategy) of an MS1 spectrum will be isolated in the quadrupole and fragmented sequentially to obtain N fragment ions MS2 spectra. In the case of the Eclipse and with tribrid instruments in general, it is possible to parallelize MS1 and MS2 spectra acquisitions by running precursor detection in the Orbitrap, while MS2 fragmentation spectra are acquired simultaneously in the ion trap but at a lower resolution. It reduces the cycle time and allows fragmenting more precursor ions.

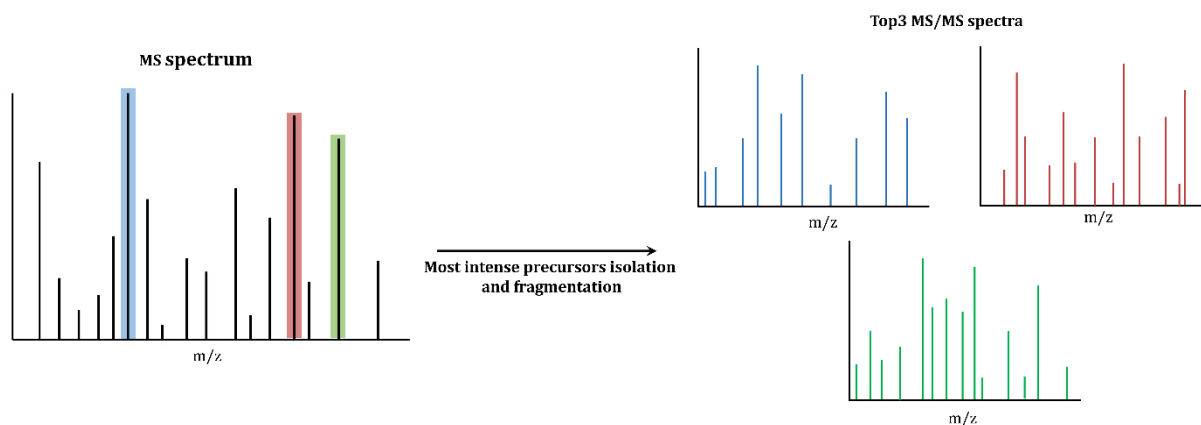


Figure 3: Schematic representation of the DDA approach. The TopN most abundant precursors (MS1) are successively isolated and fragmented to obtain fragmentation spectra (MS2) of fragment ions.

In DDA mode, precursors selection based on their intensity on the MS1 spectrum induces bias and lack of reproducibility due to stochasticity, as intense precursor ions are favored compared to less abundant ones⁵⁰. Unless faster and more sensitive instruments are available now, a non-negligible part of the precursors is still undetected in DDA mode. For example, three injection replicates analyzed on the UltiMate 3000 coupled to the Orbitrap Eclipse Tribrid resulted in an 80% overlap at the protein level and 62% at the peptide level.

To limit this effect, precursors already fragmented once are excluded for a certain period using dynamic exclusion⁵¹. It reduces the redundancy and increases the number of identifications. Similarly, it is possible to set inclusion or exclusion lists that favor or prevent the selection of user-defined specific precursors⁵².

While DDA allows only relative quantification, other acquisition modes⁵³, using targeted methods, have been developed for absolute quantification, such as Selected Reaction Monitoring (SRM)⁵⁴, Parallel Reaction Monitoring (PRM)⁵⁵ or Multiple Reaction Monitoring (MRM)⁵⁶.

b) Fragmentation techniques

In bottom-up proteomics, peptides are fragmented to identify the sequence by breaking the peptide bond between each amino acid. Fragmentation can generate several ion types, classified by Klaus Biemann⁵⁷ and depicted in **Figure 4**.

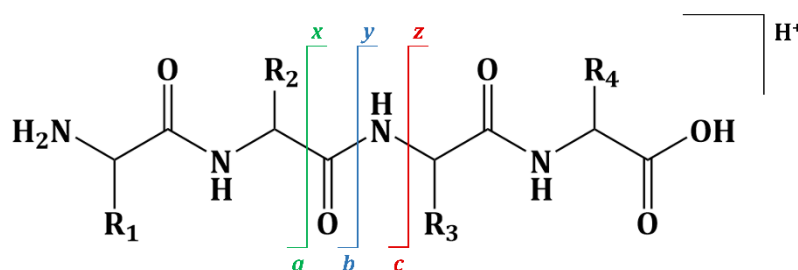


Figure 4: Ion types obtained after fragmentation as defined by the Biemann nomenclature. Six types of ions can be generated: *a*, *b*, and *c* for N-terminal fragments, and *x*, *y* and *z* for C-terminal fragments.

Various fragmentation modes have been implemented for peptide fragmentation. Collision-induced dissociation (CID) and higher-energy collision dissociation (HCD) are bottom-up approaches' most commonly used fragmentation techniques. In both cases, isolated ions are accelerated to increase their kinetic energy and then allowed to collide with neutral molecules (generally helium, argon, or nitrogen). Due to the collision, some of the kinetic energy is converted into internal energy, and the

migration of a mobile proton through the amide bonds results in the breakage of the peptide bond (mobile proton model)⁵⁸, and consequently, the peptide fragmentation⁵⁸⁻⁶⁰. Especially in the case of HCD, ions are more accelerated, resulting in higher energy for this type of fragmentation than CID. HCD is typical of Orbitrap instruments and, contrary to CID fragmentation, performed in a trap (high-pressure cell of the ion trap on the Eclipse), HCD occurs in a collision cell⁶¹. On previous instruments, ions were accumulated in the C-trap, fragmented in the collision cell, and transferred back to the C-trap before being analyzed. For the Eclipse, ion accumulation and HCD fragmentation are performed in the ion routing multipole (IRM), and the C-trap only guides the ions to either the Orbitrap or the IRM (**Figure 5**). Both CID and HCD fragmentations produce *b*- and *y*-type ions, as defined in the Biemann nomenclature, and are well suited for tryptic peptides, which carry at least two positive charges – a mobile one at the N-terminal extremity (NH_3^+) and a second one on the side chain of the lysine or arginine residues⁶⁰. Thanks to these fragment ions, it is possible to determine the peptide sequence.

Fragmentation using electron transfer dissociation (ETD) is obtained by the reaction between a peptide ion and an anion (usually fluoranthene generated by a chemical ionization source) via electron transfer leading to fragmentation along the peptide backbone^{62, 63}. ETD generates *c*- and *z*-type ions and presents the advantage of being a soft fragmentation, preserving labile PTMs that are generally lost using collision-induced dissociation techniques⁶⁴. ETD can be combined with HCD through electron transfer/higher-energy collision dissociation (ETHcD) to fragment peptide for specific applications⁶⁵⁻⁶⁷ thanks to the generation of *b*-, *c*-, *y*- and *z*-type ions.

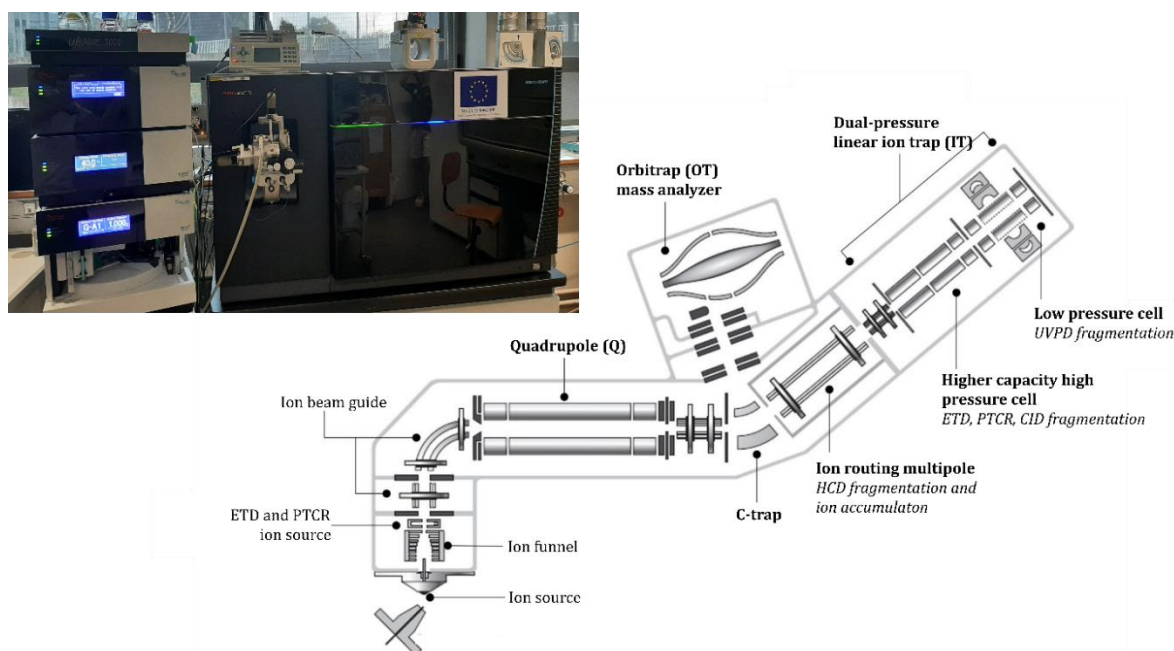


Figure 5: Picture of the LC-MS/MS coupling and schematic representation of the Orbitrap Eclipse Tribid (Thermo Fisher Scientific).

Other fragmentation techniques, including electron capture dissociation (ECD)⁶⁸ or ultraviolet photodissociation (UVPD)⁶⁹, can also be employed in bottom-up proteomics, but they will not be used in this manuscript for this purpose.

c) Ion mobility

Ion mobility spectrometry (IMS) coupled with mass spectrometry (MS) is of growing interest in proteomics, as it provides an additional dimension of separation. It relies on the separation of ions in a buffer gas under the influence of an electric field based on their mobility in the gas phase, which depends on their charge, size, and shape⁷⁰. Multiple IMS technologies have been developed for MS analysis (**Figure 6**). The implementation of high-field asymmetric waveform ion mobility

spectrometry (FAIMS)^{71, 72} and trapped ion mobility spectrometry (TIMS)⁷³⁻⁷⁵ technologies developed by Thermo Fisher Scientific and Bruker, respectively, on their instrument has boosted the use of IMS for bottom-up proteomics.

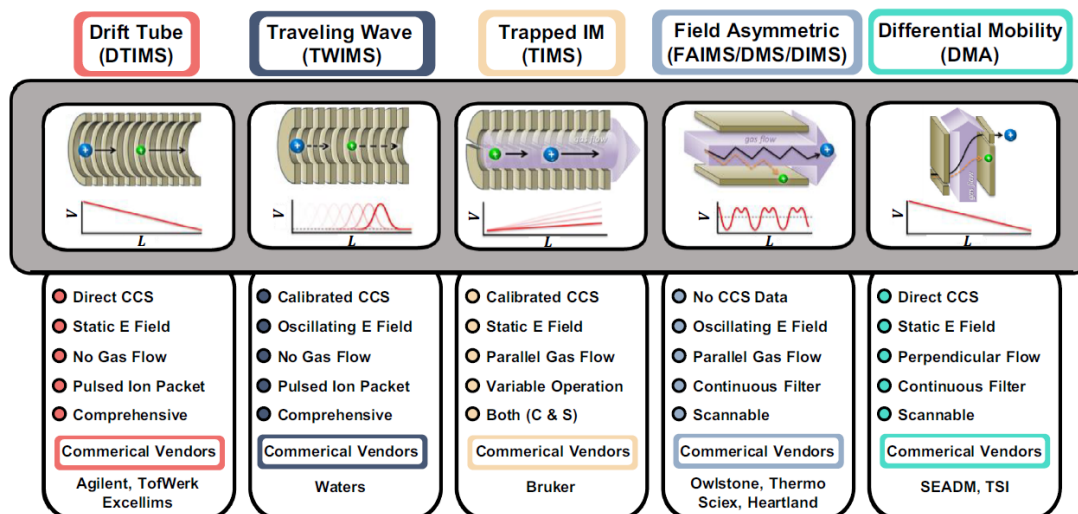


Figure 6: Summary of the various IMS devices with their specificities and vendors (from ⁷⁰)

The combination of TIMS with Parallel Accumulation-Serial Fragmentation (PASEF)^{75, 76}, implemented in the TimsTOF Pro, improves the number of identified peptides through the additional separation of co-eluting with the ion mobility. This technology was not used for the work described in this manuscript and will not be detailed here.

FAIMS technology is an optional feature that can be easily added or removed from the mass spectrometer. It is an atmospheric pressure device interfaced between the electrospray emitter and the entrance of the mass spectrometer, which performs efficient and fast gas-phase separation of peptides based on their mobility^{77, 78}. Briefly, ions are transmitted by alternating between high and low electric fields between inner and outer cylindrical electrodes. Ions with limited differences in mobility will migrate through the device, while ions with a significant enough difference in mobility collide against the electrodes (**Figure 7.A**). The addition of a dc voltage, named compensation voltage (CV), influences the trajectory of the ions between the electrodes. Thus, selecting an appropriate CV will compensate for the drift of a specific ion or group of ions, allowing them to pass through the device. By changing the CV, different ion populations can enter the mass spectrometer as the CV value, determined empirically, controls the ion population transmitted through the device. In addition, the quick change between CV values (~25 ms per transition) enables to switch between multiple current values during an MS cycle and thus independently analyze ions of different charges (**Figure 7.B**).

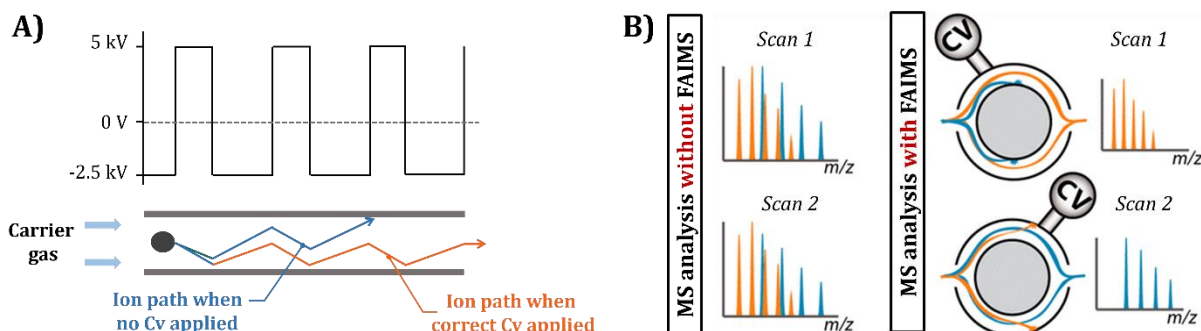


Figure 7: Schematic representation of the FAIMS device (adapted from Hebert *et al.*⁷⁷). (A) Ion paths under alternating low and high electric fields without CV (blue) and with a correct CV applied (orange), allowing ions to enter the MS. (B) Compensation voltages switching representation.

FAIMS cannot accumulate ions but is of particular interest in removing singly charged species and consequently improving the signal-to-noise ratio (S/N) by reducing the background noise.

The FAIMS device (Thermo Fisher Scientific) was the only device used in this manuscript.

C. Data processing

The vast amount of information acquired by the mass spectrometer, including the mass-to-charge ratio of the precursor (from MS1 spectra) and fragment (from MS2 spectra) ions, their measured intensities, and their retention time, are compiled in a peak list. These experimental data are compared to a theoretical database obtained from the *in silico* digestion – using the enzyme selected for the sample preparation – of the protein sequences of the studied organism. This approach is named peptide fragmentation fingerprinting (PFF)⁷⁹. Peptide identifications lead to protein identifications by inference. Peptides are unique to a protein in the database or shared by multiple proteins; the latter can be merged into the same protein group. Protein inference is a complex process following the parsimony principle to obtain the smallest possible list of protein groups²⁴.

i) Protein sequence databases

The choice of the protein database is of utmost importance to ensure the correct identification of peptides and proteins as it directly determines the search space. Databases containing redundant or false sequences will hamper the results of the data analysis.

Multiple databases are available:

- UniProtKB⁸⁰ was created through the collaboration between the European Bioinformatics Institute (EMBL-EBI), the Swiss Institute of Bioinformatics (SIB), and the consortium Protein Information Resource. It is the most widely used database and is composed of two different sections:
 - SwissProt⁸¹ is considered the reference database in proteomics when enough proteins are available for the studied organism. It comprises manually annotated and curated entries by biologists to keep only high-quality information. Additionally to the sequence, this section provides information about the known PTMs, sequence variations, function and localization of the protein, its structure, or its interactions with other proteins. On August 13, 2023, it contained 569,793 entries.
 - TrEMBL⁸² contains unreviewed entries computationally annotated by automated systems from GenBank or the literature. They are waiting to be manually validated and added to SwissProt. As this database is not manually curated, it is of poor quality, containing multiple redundant sequences and false annotations. On August 13, 2023, it contained 248,272,897 entries.
- RefSeq⁸³ is generated by the National Center for Biotechnology Information (NCBI). The protein sequences are retrieved from the automated annotation of the genomic data, and some of them were manually validated. The link between the gene, transcript, and protein is available for each protein. Nevertheless, errors from gene sequencing or nucleic sequences translation into protein sequences are possible and can affect the search quality⁸⁴. The last release (RefSeq release 219, July 18, 2023) contains 271,985,047 protein sequences.

ii) Search engines

Many search engines are available to search experimental data against databases and perform peptide and protein identifications. Among them, Mascot⁸⁵ (Matrix Science), Andromeda⁸⁶, Sequest (Thermo Fisher Scientific)^{87, 88}, Comet⁸⁹ or X!Tandem⁹⁰.

In the work presented in this manuscript, only Sequest HT (implemented in Proteome Discoverer 2.5) and Mascot, which are both non-open-source algorithms, were used. They are described below:

- Mascot is a search engine commercialized by Matrix Science (London, UK). For each MS2 spectra, a so-called ion score is calculated based on the probability that the spectra attribution to a sequence from the database is a false positive. Higher the score, the more confident the identification is. The score is given at the PSM level, corresponding to all the identifications associated with a spectrum.
- Sequest was the first fully automated peptide identification software^{87, 88}. For each spectrum, predicted peptides are filtered to keep only precursors that fit the m/z values of the isolation window of the experimental spectra, and the theoretical spectra related to these predicted peptides are generated. Then, theoretical spectra are compared to the experimental spectrum, and cross-correlation is computed to give a C-Score (also called XCorr). Then, the experimental spectrum matches the theoretical spectra with the highest C-Score.

To perform a search, various user-defined parameters are required to consider the experimental conditions:

- The database, which defines the search space;
- The enzyme(s) used for proteolytic digestion, its cleavage sites, and the number of missed cleavages allowed;
- The modifications and the modification sites to look for, either fixed (e.g., carbamidomethylation of cysteines) or variable (e.g., methionine oxidation);
- The mass tolerances for matching a precursor (MS1) or fragment (MS2) ion to its theoretical mass;
- The fragmentation employed to search for suitable ion types.

Despite the efforts to have the best search engines, 50% to 75% of the MS/MS spectra remain unassigned in a single run⁹¹. There are several reasons for this:

- The quality of the database is of utmost importance, as mentioned in the previous section (**Chapter 1 - 2.C.i**).
- Poor quality spectra due to an insufficient number of fragments or too low overall intensity.
- Errors in the data treatment, such as wrong monoisotopic peak or charge state attribution, or bad peak extraction, for example.
- Presence of chimeric spectra – i.e., co-isolation and co-fragmentation of two or more precursors leading to MS2 spectra containing fragments from different peptides – even using narrow isolation windows (generally 1-3 m/z). Since version 2.5, Mascot can search all the possible peptides from chimeric spectra. Similarly, since version 2.4 (2019), a precursor detector node has been implemented in the Proteome Discoverer (PD) workflow solution before a Sequest search. It detects other isotopic clusters in the isolation window and duplicates the MS/MS spectrum for each precursor detected.
- Presence of many unexpected and varied PTMs inducing a mass shift from the unmodified peptide sequence. If this mass shift is not specified for the search, modified peptides will not be identified. In the literature, it is estimated that one-third of the unassigned can be attributed to unexpected modifications^{92, 93}. Indeed, searching for numerous PTMs drastically increases the search space, requiring enormous computing resources and exploding computational time for conventional search engines.

Open modification search (OMS) allows considering all the possible modifications found on proteins for the search. It reduces the bias when using user-defined PTMs that prevent identifying peptides with unexpected modifications. Fast and efficient tools dedicated to OMS have been developed, such as MSFragger⁹⁴, Open-pFind⁹⁵, or ionbot⁹⁶.

For the work detailed in this manuscript, ionbot (CompOmics, Ghent, Belgium) is the only OMS tool used. It is a recent algorithm performing OMS by combining multiple features, including retention time prediction from DeepLC⁹⁷ (2021) and fragment intensities predictions from MS²PIP^{98, 99}, to improve the identification of unexpectedly modified peptides.

iii) Validation of the results

Algorithms used for MS data searching and processing are not faultless and can induce peptides and/or proteins misattributions in the results. Scores assigned to PSMs by each search engine do not allow judging the exactness of the assignment, and manual verification of thousands of spectra is not amenable. Automated validation methods have been developed to ensure the quality of the PSMs identified. The most widely used in proteomics is the target/decoy strategy, allowing evaluation of the proportion of false positives in a dataset^{100, 101}. Thus, decoy sequences – generated by random scrambling or reversing the target sequences¹⁰⁰ – are added to the database to be searched simultaneously with target sequences. Two assumptions must be considered: (i) decoy sequences are not present in real life, and then any match of a spectrum to a decoy sequence is presumed to be a random event and considered as a false identification; (ii) random non-peptide spectrum is equally likely to be assigned to a decoy or a target peptide. With this approach, it can be considered that for a given score threshold, the number of decoy PSMs above the threshold is similar to the number of false target PSMs. Hence, the false discovery rate (FDR) is calculated as described in the following equation:

$$FDR (\%) = \frac{\text{number of decoy PSMs}}{\text{number of decoy PSMs} + \text{number of target PSMs}} \times 100$$

In bottom-up proteomics, the target FDR is usually set at 1%. Thus, the algorithm will determine the lowest score threshold that allows reaching this FDR value, and each target PSM above this score will be accepted as valid.

With the approach described above, only the global FDR is controlled, i.e., the average proportion of accepted false PSMs. In addition, it only considers one score from the search engine to control the FDR. In contrast, multiple scores (e.g., C-Score, Δ C-Score for SEQUEST) are provided and have different abilities to distinguish target from decoy PSMs. To improve the identification of true peptides while controlling the FDR, Percolator¹⁰², a semi-supervised machine learning algorithm combining multiple scores and features from the PSM scan, has been developed. Percolator employs twenty features, such as the precursor mass error, the fraction of matched *b* and *y* ions, the peptide length, or the XCorr, to assign a statistically meaningful q-value to each PSM, defined as the minimal FDR at which the identification is considered correct¹⁰³⁻¹⁰⁵. Data validation using Percolator increases the number of target peptide identifications compared to the classical strategy based only on the score provided by the search engine. It is compatible with any search engine output and is widely used in proteomics.

FDR estimation can be performed at various levels: PSMs, peptides, and proteins. Combining FDR filtering at multiple levels is possible to increase the identification's confidence. Nevertheless, no rules exist regarding which filters to use.

iv) Machine learning-based strategies in proteomics

The widespread use of machine learning (ML) has made it possible to develop new strategies and tools to improve peptide identifications and confidence in their attribution. Machine learning refers to the ability of a machine to learn from a vast amount of data, determine patterns and then apply what has

been learned to new datasets^{106, 107}. In proteomics, machine learning and deep learning (a subset of machine learning) are used for multiple applications, including retention time (RT) predictions, MS2 spectra feature predictions, PTM predictions, or protein structure predictions.

Accurate retention time prediction is of particular interest in LC-MS/MS-based proteomics for improving the sensitivity of the database searching, confirming the identification of peptides bearing unexpected PTMs, as a quality metric for peptide identification or spectral library generation. Multiple ML algorithms are available to predict retention time, such as Elude^{108, 109}, DeepRT¹¹⁰, or DeepLC⁹⁷. Particularly, DeepLC is the only model able to predict RT for peptides with modifications that have not been seen in the training dataset, as the prediction is based on the atomic composition of the peptides⁹⁷. Nevertheless, prediction accuracy is slightly lower for extensively modified peptides than for unmodified ones.

MS/MS spectrum predictions are another application of ML. Matching experimental spectra to theoretical ones mainly relies on comparing the m/z values and the intensities of the fragment ions. These experimental MS/MS spectra obtained after fragmentation of a precursor ion display patterns, defined by several key factors¹¹¹: (i) the fragmentation technique and the energy applied; (ii) the type of MS instrument; (iii) the peptide sequence and the associated charge state. While the mechanisms inherent to the fragmentation of each peptide are intricate to demonstrate and still not fully explained and understood, it has been shown that fragmentation patterns are reproducible and predictable^{87, 112}. Thus, many tools have been engineered to predict MS/MS fragmentation spectra using ML, including MS²PIP^{98, 99}, MS²CNN¹¹³, or pDeep^{114, 115}.

Other algorithms combine both RT and MS/MS spectra predictions to perform peptide identifications, such as Prosi¹¹⁶, DeepMass¹¹⁷, or the tool detailed in Guan *et al.*¹¹⁸.

Using such bioinformatics tools provides access to a wide range of LC-based or MS-based metrics not available in classical search engines. These metrics can be merged into post-processing tools, such as Percolator, to improve scoring functions that discriminate target from decoy PSMs (**Figure 8**). Indeed, ML-based data validation has demonstrated significant improvements in the number of confidently identified peptides, using MS²Rescore¹¹⁹, Inferys¹²⁰, or DeepRescore¹²¹, for example. Regarding the promising results obtained, the interest in these ML algorithms is growing fast in proteomics.

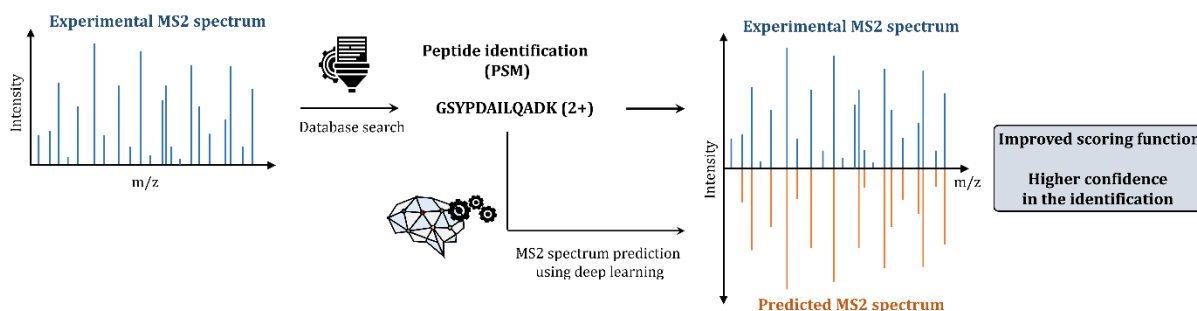


Figure 8: Schematic representation of the PSM rescoring using deep learning approaches. Experimental MS2 spectra are searched against a database, and then, each PSM is compared to its predicted spectra. It allows adding new features in the scoring functions, to increase the identification confidence.

3. Global quantification strategies in bottom-up proteomics

In many studies, the identification of peptides and proteins is insufficient to answer the biological questions addressed, and quantitative information is required to get a deeper understanding of the studied biological system's dynamics and variability²³. Although MS is not primarily designed for

quantitative analysis, approaches have been implemented to compare quantitatively between samples using MS signal (**Figure 9**).

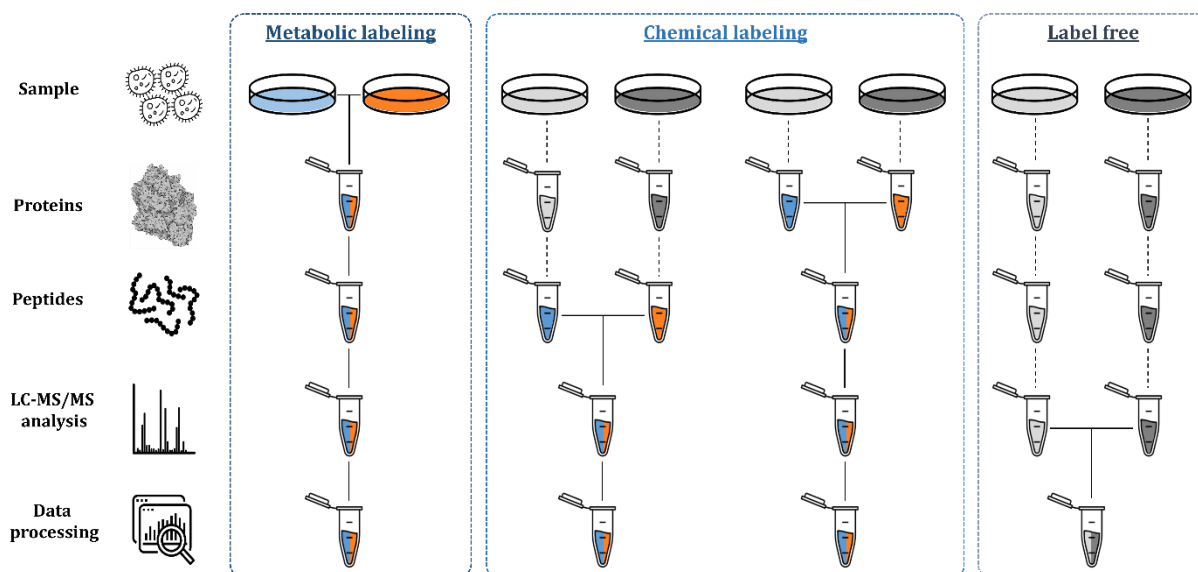


Figure 9: Overview of the global quantification strategies in bottom-up proteomics (adapted from ¹²²). The light and dark gray Petri plates and tubes represent the two conditions to be compared without labeling. When these are orange and blue, labeling has been performed. The tubes that are both orange and blue illustrate the multiplexing of the two conditions. The dashed lines indicate the steps in which variations and errors in quantification between samples from different conditions can occur.

A. Label-based relative quantification approaches

Label-based relative quantification approaches use isotope labels to quantify peptides and by-end proteins. It assumes that isotope-labeled (heavy) and unlabeled (light) peptides have the same physicochemical properties and, consequently, the same behavior during the chromatographic separation (retention time) and mass spectrometric (fragmentation pattern, ionization efficiency, and MS signal response) steps. Their mass is the only property allowing distinguishing between unlabeled and labeled peptides, as the latter is heavier due to the isotope labeling. In label-based quantification experiments, samples with different isotopes are pooled to quantify multiple samples in the same spectrum. Two different types of labeling can be used: metabolic or chemical.

i) Metabolic labeling

In metabolic labeling strategies, stable heavy isotope labels are incorporated into proteins through labeled amino acids during the cell culture. Labeled (heavy) and unlabeled (light) samples are then pooled before undergoing classical proteomics sample preparation, followed by LC-MS/MS analysis. MS1 signals comparison of heavy and light peptides allows the relative quantification between the samples. Stable isotope labeling with amino acids in cell culture (SILAC)^{123, 124} is the most used approach for metabolic labeling, where heavy arginine and lysine residues containing isotopes (²H, ¹³C, and ¹⁵N) are incorporated in the culture medium to label the cells. Nevertheless, this strategy restricts multiplexing (i.e., three samples maximum) and sample type (cell culture mainly). Other metabolic labeling strategies have also been developed, such as NeuCode¹²⁵ or super-SILAC¹²⁶ but they still suffer from a lack of multiplexing.

ii) Chemical labeling

In contrast to metabolic labeling, chemical labeling enables multiplexing, and two strategies can be employed. The first one, introduced in 1999 and called isotope-coded affinity tag (ICAT)¹²⁷, is based

on isotope tagging on the cysteine residues of the proteins prior to digestion (protein level). The second one relies on isobaric labeling of the peptides and is used in different approaches, including isobaric tags relative and absolute quantification (iTRAQ)^{128, 129} and tandem mass tag (TMT)¹³⁰. As the TMT strategy was used in the work presented in this manuscript, it is detailed below and in **Figure 10**.

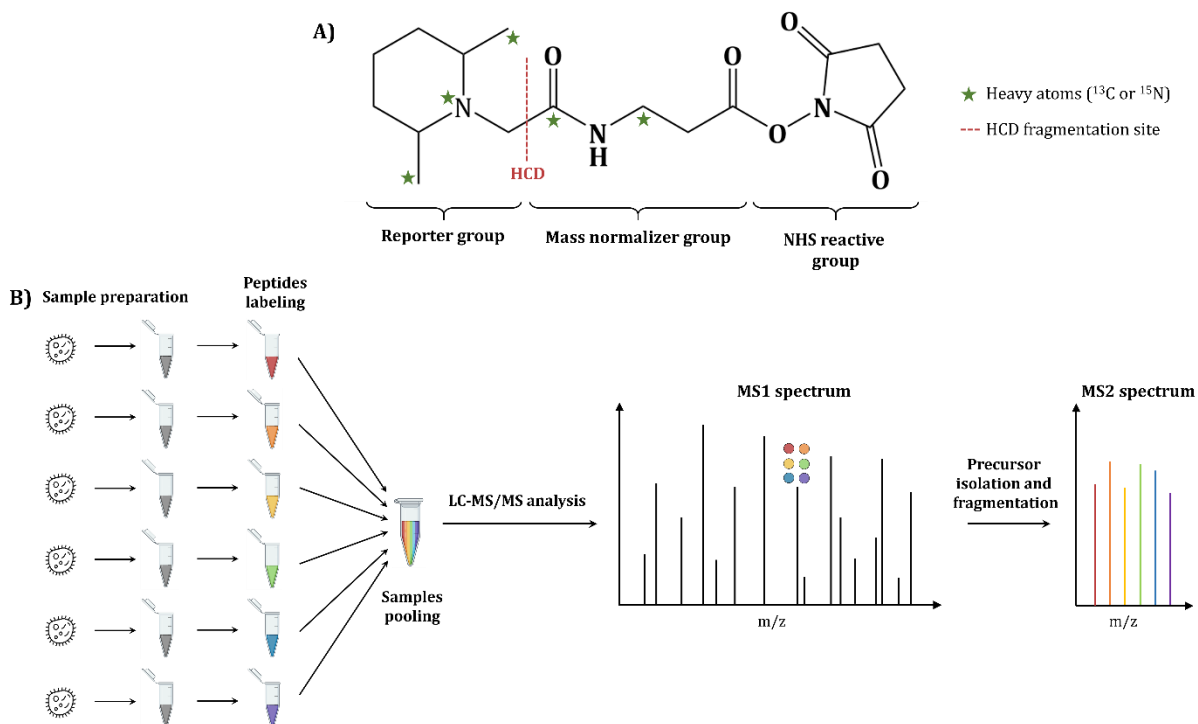


Figure 10: Overview of the TMT approach. (A) Representation of the 6plex TMT reagent composed of a reporter group, a mass normalizer group and a NHS reactive group. The dashed red line represents the cleavage during HCD fragmentation. Green stars represent the position of the heavy atoms for TMT reagent 130 from 6plex kit. (B) Schematic representation of the TMT workflow, including sample preparation and peptides labeling. Labeled peptides are pooled and analyzed in a single run. They are indistinguishable at precursor level (MS1) but after fragmentation, it is possible to perform relative quantification based on the reporter groups.

In the TMT approach, peptides are labeled at the end of the sample preparation process (i.e., after the enzymatic digestion). TMT reagent reacts with the N-terminal and ϵ lysine side chain amino groups of peptides through an amine-reactive NHS-ester group. In addition to the reactive group, the TMT reagent comprises a mass reporter group and a mass normalizer group (**Figure 10.A**). Thanks to the mass normalizer group, all the tags have the same global mass, making it impossible to distinguish between peptides from one sample or another. During fragmentation, the bond between the reporter and the balance groups is broken, enabling the relative quantification of the different samples using the relative intensity of each reporter ion located in the low m/z region of the MS/MS spectra (**Figure 10.B**). Other fragment ions are employed for peptide identification. The TMT strategy enables multiplexing up to 18 samples, with the development of the TMTpro-18plex (Thermo Fisher Scientific)¹³¹. It allows for reducing the instrument time and directly comparing multiple samples. Moreover, since the signals from multiple samples are combined, the MS1 signal intensity of each peptide is higher, allowing injecting less sample than what would be done for individual injections. Nevertheless, ratio compression issues due to interferences are observed but can be overcome by using additional isolation and fragmentation step to generate MS3 spectra that will be used for the relative quantification of the samples^{132, 133}.

B. Label-free relative quantification approaches

Label-free quantification approaches have rapidly grown in interest thanks to the improvements in sensitivity and scan speed of modern mass spectrometers¹³⁴. It allows coping with the limitations of label-based quantification, such as additional sample preparation steps, cost of the labeling reagents, labeling efficiency, or lack of multiplexing. Here, label-free methods can be used for any samples and are not limited in number of samples as there is no multiplexing¹²², making these approaches widely used in proteomics. However, as samples are prepared and analyzed independently, each step can affect the precision of quantification. Two strategies are available for label-free quantification: spectral counting and extracted ion chromatogram MS1 (XIC-MS1).

i) Spectral counting

For spectral counting, it is assumed that the number of MS2 spectra acquired for a protein is correlated to its abundance. Its main asset is the ease of quantification that only requires tools used for identification. However, stochasticity of the DDA acquisition mode and under-sampling related to the use of dynamic exclusion to maximize the number of identified peptides are major drawbacks, induce many missing values, and limit the ability to quantify small abundance changes¹³².

ii) Extracted ion chromatogram (XIC-MS1)

Just as spectral counting assumes that the number of MS2 spectra is correlated to the protein abundance, XIC-MS1 assumes that precursor signal intensity is proportionally linked to precursor abundance¹³⁵. The abundance can be calculated using either the area under the MS1 chromatographic peak or the apex of this peak. In this approach, acquisition parameters must be optimized to balance between a high enough number of MS1 spectra to define the chromatographic peak and a high enough number of MS2 spectra used for peptide identification¹³⁶. Hence, it requires robust and reproducible chromatographic systems¹²² coupled with high-resolution and accurate mass (HR/AM) MS instruments.

In contrast to spectral counting, XIC-MS1 needs dedicated and more advanced data processing bioinformatics tools. Numerous software has been developed to mitigate the variability induced along the experiment, from the sample preparation to the LC-MS/MS analysis, using retention time recalibration and precursor intensity normalization^{132, 136, 137}. Bioinformatics solutions comprise open-source tools – such as Proline (developed by the French Proteomics Infrastructure, ProFI)¹³⁸, MaxQuant¹³⁹ (Max Plank Institute) or Skyline¹⁴⁰ (MacCoss lab) – and vendor software such as Proteome Discoverer (Thermo Fisher Scientific) or SpectroMine™ (Biognosys). The retention time and chromatographic peak intensity, m/z ratio, and precursor charge state for each precursor are used to generate LC-MS maps that are then employed to align identification results with other MS1 signals, even if the peptide has not been isolated and fragmented to generate an MS2 spectrum (called “match between runs” in MaxQuant). It allows to obtain a quantitative value in samples where peptides were not correctly identified, limiting the number of missing values. The XIC signals finally obtained are used to perform the quantification. The performances of these tools for XIC-MS1 quantification in proteomics have been widely compared in multiple studies available in the literature^{138, 141, 142}.

C. Label-free “absolute” quantification

Label-free approaches can be used to perform “absolute” quantification to obtain an estimation of the protein quantity in the sample without the use of isotopes. Similarly to label-free relative quantification, the low cost and the easy experimental setup make label-free “absolute” quantification attractive¹³². For example, emPAI¹⁴³ (exponentially modified protein abundance index) derived from PAI¹⁴⁴ is based on the same principle as spectrum counting, estimating the protein abundance by counting the number of observed peptides from a protein.

Intensity-based methods, such as intensity-based absolute quantification (iBAQ), use the sum of all peptide intensities divided by the total number of peptides theoretically observable in this protein¹⁴⁵. Another approach is Top3 quantification (or Hi-3), which uses unlabeled proteins standard spike in the sample of interest¹⁴⁶. The intensity of the three most intense peptides of the proteins is compared to the intensity of the three most abundant peptides of the standard to calculate the abundance. This method was used in Part II of this manuscript.

4. Data-independent acquisition approaches

Identifying and quantifying proteins in complex samples is the essence of bottom-up proteomics. DDA methods can face this challenge by identifying and quantifying thousands of proteins in a single run but suffer from a limited dynamic range and lack of sensitivity and specificity. Other approaches, such as selected reaction monitoring (SRM) or parallel reaction monitoring (PRM), together known as targeted approaches, can overcome the limitations of DDA but are only applicable to a limited number of proteins of interest. Hence, data-independent acquisition (DIA) has been developed as a new approach combining the advantages of DDA (comprehensive proteome coverage) and targeted methods (high sensibility, sensitivity, reproducibility and accuracy, wide dynamic range). DIA's emergence and rapid growth have been possible thanks to recent instrumental advances in MS, particularly regarding their high-resolution and fast scan rate. It also implies the design of dedicated bioinformatics tools, which have primarily contributed to the widespread use of DIA for quantitative proteomics^{147, 148}.

A. Principle of data independent acquisition

DIA-MS experiments are performed on HR/AM instruments, including Q-TOF or Q-Orbitrap. DIA allows, theoretically, the collection of MS/MS spectra from all the peptides present in the sample along the chromatographic gradient. In detail, the mass range is divided into windows, and precursor ions included in each isolation window are co-isolated and co-fragmented. The fragments are co-analyzed, generating multiplex MS2 spectra. Finally, chromatographic peaks are extracted from MS2 data to allow peptides (and proteins) identification and quantification (**Figure 11**).

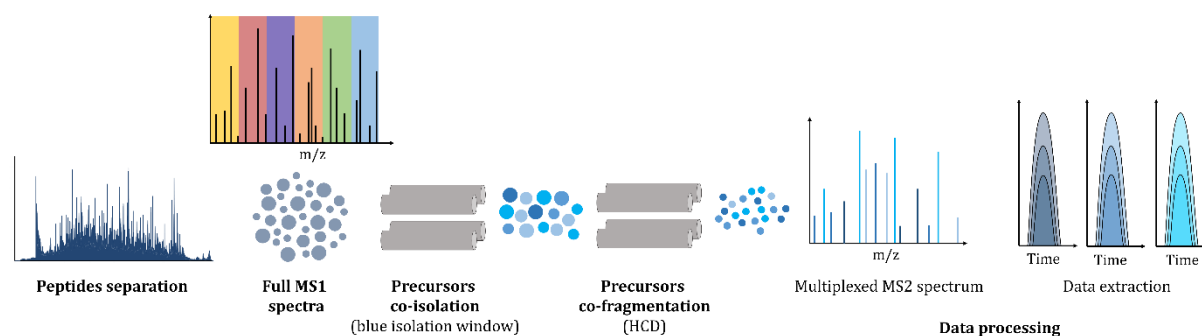


Figure 11: Overview of the DIA workflow using consecutive isolation windows. Eluted peptides are analyzed by MS and all the precursors from a window are co-isolated and co-fragmented to obtain multiplexed MS2 spectra. MS2 spectra are then processed to identify and quantify peptides.

B. Evolution of DIA-based approaches

The first DIA-related experiments were carried out in the early 2000s. Masselon *et al.*¹⁴⁹ characterized several polypeptides from multiplexed spectra on an FT-ICR in 2000. A few years later, Purvine *et al.*¹⁵⁰ presented shotgun CID, using a two steps analysis with low voltage in-source CID for precursor mass measurement (MS) followed by high voltage in-source CID to fragment peptides and obtain the MS/MS spectra of fragment ions.

The term “data independent acquisition” was first coined in 2004 in a study by Venable and coworkers, where they used sequential 10 m/z isolation windows to fragment precursor ions from yeast samples on an LTQ Orbitrap instrument¹⁵¹.

These three studies paved the way for developing various DIA-based strategies on different MS instruments using multiple approaches, from MS acquisition parameters to data processing. A summary of the methods developed since 2004 to date is displayed in **Table 3** divided into two categories, those using the full mass range for fragmentation and others using defined isolation windows. Some approaches of particular interest are detailed below.

Table 3: Summary of the DIA approaches since 2003 (adapted from ^{147,148}).

Method	Instrument	DIA isolation window (m/z)	Year of introduction	Reference
Shotgun CID	Q-TOF	Full range	2003	Purvine <i>et al.</i> ¹⁵⁰
Original DIA	LTQ Orbitrap	10	2004	Venable <i>et al.</i> ¹⁵¹
MS ^E	Q-TOF	Full range	2006	Silva <i>et al.</i> ¹⁵²
PAcIFIC	LTQ Orbitrap	2.5	2009	Panchaud <i>et al.</i> ¹⁵³
AIF	Exactive Orbitrap	Full range	2010	Geiger <i>et al.</i> ¹⁵⁴
XDIA	Ion Tap-ETD-CAD	20	2010	Carvalho <i>et al.</i> ¹⁵⁵
SWATH	TripleTOF	25	2012	Gillet <i>et al.</i> ⁶
MSX	Q-Exactive Orbitrap	4	2013	Egertson <i>et al.</i> ¹⁵⁶
WiSIM	Orbitrap Fusion	200	2014	Martin <i>et al.</i> ¹⁵⁷
pSMART	Q-Exactive	5-20	2014	Prakash <i>et al.</i> ¹⁵⁸
UDMS ^E	Synapt G2-S HDMS	Full range	2014	Distler <i>et al.</i> ¹⁵⁹
HRM	Q-Exactive	24-222	2015	Bruderer <i>et al.</i> ¹⁶⁰
SONAR	Xevo G2-XS QTOF	24	2018	Moseley <i>et al.</i> ¹⁶¹
BoxCar	Q Exactive HF Orbitrap		2018	Meier <i>et al.</i> ²⁵
Scanning SWATH	TripleTOF	5	2019	Messner <i>et al.</i> ¹⁶²
diaPASEF	TimsTOF Pro	25	2019	Meier <i>et al.</i> ¹⁶³
PASS-DIA	Q Exactive HF Orbitrap	2	2020	Mun <i>et al.</i> ¹⁶⁴
PulseDIA	Q Exactive HF-X Orbitrap, TripleTOF	20/60-140	2020	Cai <i>et al.</i> ¹⁶⁵
DIA-FAIMS	Orbitrap Exploris 480	13.7	2020	Bekker-Jensen <i>et al.</i> ¹⁶⁶
BoxCarMax	Orbitrap Fusion Lumos		2021	Salovska <i>et al.</i> ¹⁶⁷
Slice-PASEF	TimsTOF Pro, TimsTOF Pro 2	3-215	2022	Szyrwił <i>et al.</i> ¹⁶⁸
Hybrid-DIA	Orbitrap Exploris 480	61.1	2023	Martínez-Val <i>et al.</i> ¹⁶⁹
midiaPASEF	TimsTOF Pro 2	36	2023	Distler <i>et al.</i> ¹⁷⁰

i) DIA strategies performed over the entire mass range

Full mass range DIA approaches were first introduced in 2005 by Silva *et al.*¹⁵² on Q-TOF instruments, inspired by the work of Purvine and coworkers¹⁵⁰. In this method, named MS^E, a first full scan is acquired for precursor ions mass measurement followed by co-fragmentation of all those precursor ions to obtain fragment ion masses. In 2014, Ultra-definition MS^E (UDMS^E)¹⁵⁹ was proposed with the implementation of ion mobility and variable collision energies to the original MS^E approach.

A similar approach performed on Q-Exactive Orbitrap instruments was developed by Thermo Fisher Scientific in 2010, named all-ion fragmentation¹⁵⁴ (AIF).

ii) DIA strategies based on isolation windows

Most of the DIA strategies are based on isolation windows. These isolation windows can be fixed or variable widths (**Figure 12**).

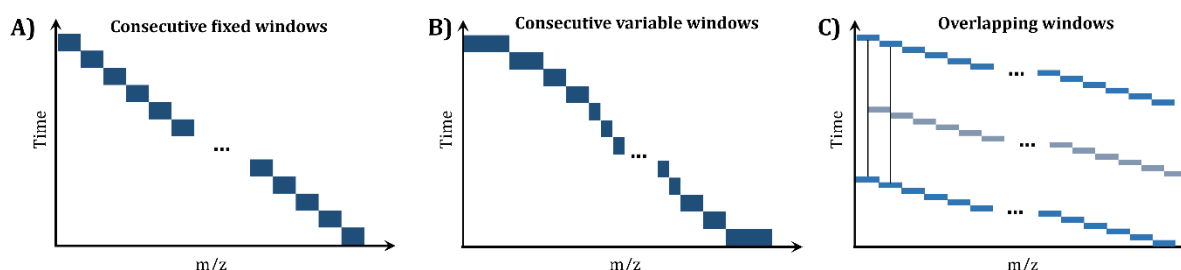


Figure 12: Summary of the main DIA approaches based on isolation windows. (A) Consecutive fixed windows, (B) consecutive variable windows, and (C) overlapping windows isolation scheme.

First, fixed-width isolation window methods were employed by many groups. The PACIFIC¹⁵³ method was used to perform fragmentation of precursors from narrow isolation windows of 2.5 Th across multiple LC-MS/MS runs to cover the entire mass range, resulting in several days of analysis. More recently, PulseDIA¹⁶⁵ was introduced to decrease the analysis time thanks to the reduced number of isolation windows, having width adapted to the ion density.

Another approach using wider isolation windows of 25 Th was proposed for Q-TOF instruments, named sequential windowed acquisition of all theoretical fragment ion spectra⁶ (SWATH), now commercialized by the Sciex company. It demonstrated the ability of DIA to obtain comparable quantification accuracy as SRM over a 4-orders dynamic range. The SWATH approach is considered one of the strategies that led to widespread DIA-based proteomics at a large scale.

One improvement of the SWATH approach is using variable windows, proposed by Zhang *et al.*¹⁷¹ in 2015. Indeed, precursor ions are not homogeneously distributed over the m/z along the chromatographic gradient, resulting in very complex MS2 spectra from isolation windows containing multiple precursors. At the same time, only a few fragments are observed in MS2 spectra containing only a couple of precursors. Hence, the size of the windows is adapted to match the ion density depending on the m/z to have comparable numbers of precursor ions in each isolation window. It leads to better quantification accuracy and an increase in identification. Similarly, hyper-reaction monitoring¹⁶⁰ (HRM) has been developed on Q-Orbitrap instruments.

Another cornerstone of DIA-based methods development is the implementation of ion mobility as an additional separation dimension on HR/AM instruments. Two methods were developed, DIA-FAIMS¹⁶⁶ for last-generation Q-Orbitrap and tribrid instruments (Thermo Fisher Scientific), and diaPASEF¹⁶³ for Q-TOF instruments and implemented on the TimsTOF Pro (Bruker). Following diaPASEF, new strategies such as Slice-PASEF¹⁶⁸ and midiaPASEF¹⁷⁰ are being developed.

BoxCar²⁵ is another alternative developed for Q-Orbitrap instruments. It aims to enhance the signal-to-noise (S/N) ratio and the dynamic range of precursor ions by filling multiple narrow m/z windows to increase the ion injection. An evolution of BoxCar is BoxCarmax¹⁶⁷, adapted for the analysis of SILAC isotope-labeled peptides, which is intricate with conventional DIA methods.

Finally, methods have been designed using overlapping windows (**Figure 12.C**). In these approaches, the isolation scheme is composed of windows covering half of the mass range analyzed with the previous window, increasing the selectivity of the analysis. It was introduced on Waters Q-TOF

instruments by Moseley *et al.* in 2018, using the SONAR¹⁶¹ approach. Another method, the Scanning SWATH¹⁶², used continuous scanning of the filtering quadrupole instead of successive isolation windows acquisition, allowing to reduce the DIA cycle time.

C. Development of DIA methods

The development of DIA methods requires optimizations of various acquisition parameters, including the number and the size of isolation windows, the cycle time, and the number of precursor ions. These parameters directly influence the sensibility, sensitivity, quantification accuracy, and proteome coverage of the analysis.

The method's sensitivity is essentially dependent on the number of ions entering the Orbitrap analyzer. This number is determined by either the automatic gain control target (AGC target) or the maximum injection time. The AGC target is the number of ions to reach before injecting ions from the C-trap to the Orbitrap for mass measurement. If this number is not reached, ions will be transferred from the C-trap to the Orbitrap after a certain amount of time, called the maximum injection time. These parameters allow maximizing the number of ions entering the Orbitrap and thus the sensitivity while avoiding increasing the cycle time.

The cycle time is a crucial parameter in DIA experiments, directly affecting the quantification quality. Indeed, a high enough number of points over the chromatographic peak (around 7-8 in DIA experiments) is required to ensure an accurate quantification. For example, a maximum cycle time of 2.5 s is allowed to obtain 8 points over a chromatographic peak of 20 s. The cycle is calculated as follows when the maximum injection time is lower than the transient time:

$$\text{Cycle time} = \text{MS1 transient time} + (\text{number of isolation windows} \times \text{MS2 transient time})$$

In Orbitrap-based instruments, the transient time is directly related to the resolving power and corresponds to the capacity of the Orbitrap to discriminate ions. Consequently, the higher the resolution, the longer the cycle time. On the Orbitrap Eclipse Tribrid, transient times of 32 ms, 64 ms, and 256 ms are observed for acquisition using resolving power (at 200 m/z) of 15,000, 30,000, and 120,000, respectively.

The number of isolation windows is determined by the mass range covered by the DIA method during the cycle and the size of the windows. These two parameters can be set empirically based on a DDA analysis of the sample of interest to obtain a balance between proteome coverage, sensitivity, and sensibility.

Finally, minor overlapping (e.g., 0.5-1 Da) between isolation windows must be considered in DIA methods to counterbalance the loss of sensitivity due to lower transmission efficiency of the quadrupole observed for precursors at each window edge^{6, 172} (**Figure 13**).

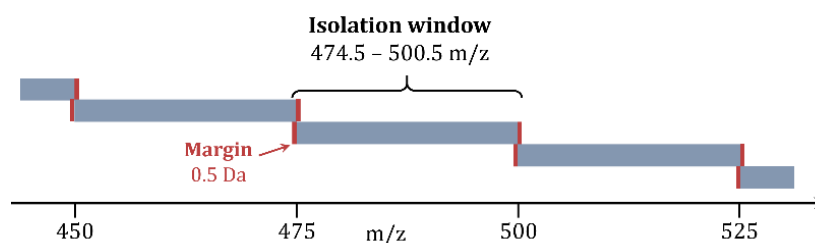


Figure 13: Representation of the quadrupole isolation overlap.

D. The challenge of DIA data processing

DIA strategies have shown a high quantification reproducibility with low missing values. Nevertheless, the complexity of the multiplexed MS2 spectra containing fragments from co-isolated and co-fragmented precursor ions requires the use of dedicated bioinformatics tools for DIA data processing, as the ones used for DDA are not amenable^{148, 173}. Two approaches can be used for DIA data processing¹⁷⁴: the peptide-centric approach and the spectrum-centric approach (**Figure 14**).

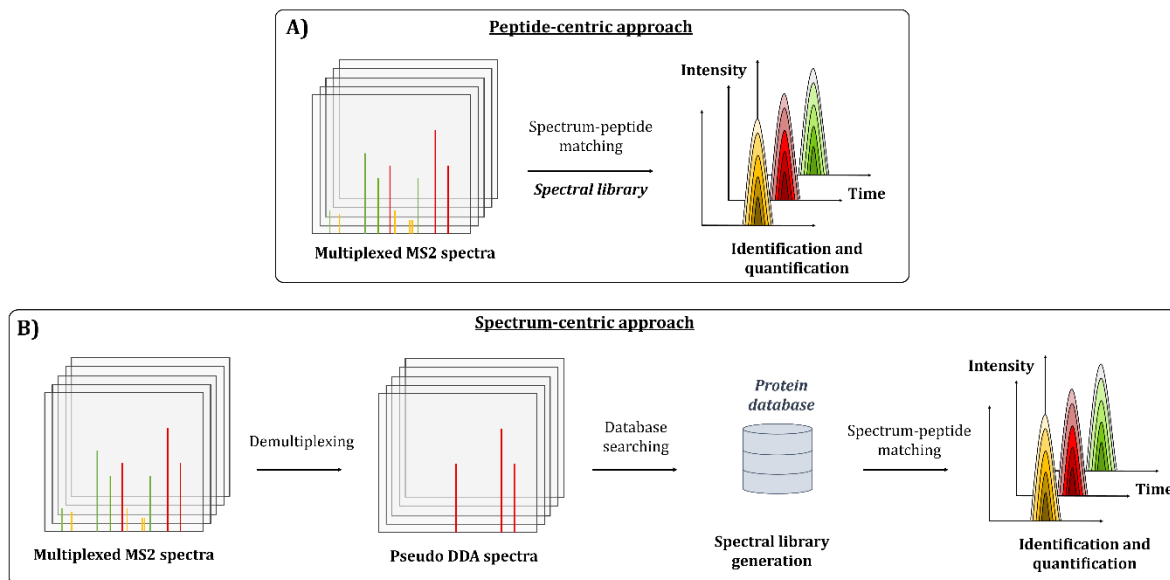


Figure 14: DIA data processing strategies (adapted from ¹⁷⁴). (A) Peptide-centric approach, involving the use of a spectral library to match peptides from multiplexed MS2 spectra. (B) Spectrum-centric approach identifies peptides directly from MS2 spectra using pseudo-DDA spectra obtained after deconvolution of multiplexed MS2 spectra. Then, signals are extracted to perform peptides identification and quantification using the peptide-centric approach.

i) Peptide-centric approach

The peptide-centric approach uses a spectral library to search DIA data. A spectral library includes features such as precursor and fragment ions m/z and normalized retention time, precursors charge state, and relative intensity of fragment ions¹⁷⁴. Other information, including ion mobility drift time, can also be added to this library¹³⁴.

Peptide identification in DIA is based on matching the DIA dataset to peptides present in the library using the features above. Hence, a significant limitation of the peptide-centric approach is that only precursors present in the library can be identified, resulting in considerable efforts to obtain the most exhaustive spectral library possible.

Different manners can be employed to generate spectral libraries. In general, identification results from fractionated DDA analysis of similar samples than those used for DIA and searched with classical DDA search engines (e.g., Mascot, Andromeda, Sequest). Nevertheless, spectral library building using fractionated DDA analyses is a tedious and time-consuming task, requiring large amounts of samples¹⁷⁵. Moreover, DDA analyses should be performed on the same system (LC and MS) and under the same conditions (gradient length, fragmentation mode, and energy) as the DIA one to ensure the high quality of the library. Another alternative is to use public spectral libraries available online, such as SWATHAtlas¹⁷⁶, PeptideAtlas¹⁷⁷, or PRIDE¹⁷⁸. However, they are restricted to a limited number of samples and organisms.

Numerous tools have been developed to search MS2 DIA data against spectral libraries and perform peptide identification and scoring. It includes OpenSWATH¹⁷⁹, PECAN¹⁸⁰, EncyclopeDIA¹⁸¹ or Spectronaut¹⁶⁰. They use the information in the spectral library to identify peptides in the DIA analysis. Then, the XIC is collected at the MS2 level for all the transitions of identified precursors, and a score is attributed for each peak group. Peptide identifications are finally validated using the target decoy approach described in **Chapter 1 – 2.C.iii**.

Recently, new algorithms implemented in Spectronaut and EncyclopeDIA have been introduced for the deconvolution of DIA data to generate DIA-based libraries or so-called hybrid libraries when combined with DDA data.

Advances in artificial intelligence and machine learning have also paved the way to predicted library generation. Predictions algorithms (DeepMass¹¹⁷, ProsiT¹¹⁶, pDeep¹¹⁵) can predict retention time as well as MS2 fragmentation spectra pattern of peptides contained in a database. Indeed, in this case, the library is not limited to what have been identified in DDA runs, but it contains all the theoretical peptides that are present in the database¹⁸². Other algorithms, such as DIA-NN¹⁸³ or Spectronaut (Biognosys), also use ML to improve the identification of peptides in DIA data processing.

Independently of the method used to generate the library and perform the DIA data processing, the quality of the spectral library remain a crucial parameter for the success of a DA experiment.

ii) Spectrum-centric approaches

Spectrum-centric approaches are based on the deconvolution of multiplexed MS2 spectra into pseudo-MS/MS spectra that can be directly searched against the protein database. They are more convenient approaches as they do not require the generation of spectral libraries.

Several tools have been developed to perform library-free DIA data processing. DIA-Umpire¹⁸⁴, developed in 2015, was one of the first software to perform the non-targeted identification and quantification of peptides. As with many other solutions, it can also generate spectral libraries from DIA data and perform a peptic-centric search using this library. A few years later, DirectDIA™ was added to the Spectronaut software (Biognosys), allowing it to perform library-free searches thanks to the implementation of the Pulsar search engine. More recently, Demichev group has developed DIA-NN¹⁸³, an algorithm based on deep neural networks to perform both peptide-centric and spectrum-centric DIA data analysis.

This spectrum-centric strategy was the most used approach in the work described in this manuscript.

Chapter 2:

The role of top-down mass spectrometry in the context of structural MS

Historically, the characterization of proteins by mass spectrometry (MS) was performed at the primary structure level, i.e., the linear sequence of amino acids, using bottom-up approaches. The technological boom in MS over the last twenty years has enabled new approaches to provide new insights into the higher-order structure of proteins (secondary, tertiary, and quaternary structures), grouped under the term structural MS¹⁸⁵⁻¹⁸⁷. Knowledge of the higher-order structure of proteins is of utmost interest as the protein structure directly drives their biochemical functions. These structural MS approaches have been widely applied to a vast array of biomolecules such as multimeric proteins¹⁸⁸, multiprotein complexes¹⁸⁹⁻¹⁹¹, membrane proteins^{192, 193}, intrinsically disordered proteins¹⁹⁴, and oligonucleotides¹⁹⁵. The strategies encompassed within the term structural MS can be mainly categorized into two subclasses: peptide-centric and protein-centric approaches (**Figure 15**). Although these techniques cannot provide atomistic details, they can contribute to the characterization of biomolecules and their complexes with structural information about stoichiometry, residue level proximity, protein dynamics, conformation, and protein post-translational modifications (PTMs). Thereby, structural MS techniques have largely proved their utility to complement and support other high-resolution techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), small angle scattering (SAS), and cryo-electron microscopy (cryo-EM). Currently, the studies in the literature aim to characterize more challenging and complex systems. Therefore, combining several structural MS techniques is needed to provide a more comprehensive picture of a protein structure. In this context, a brief introduction will be provided about the most significant structural MS techniques, highlighting the main contributions and advantages, with special attention to the role of the TD-MS approach in the characterization of different biological systems.

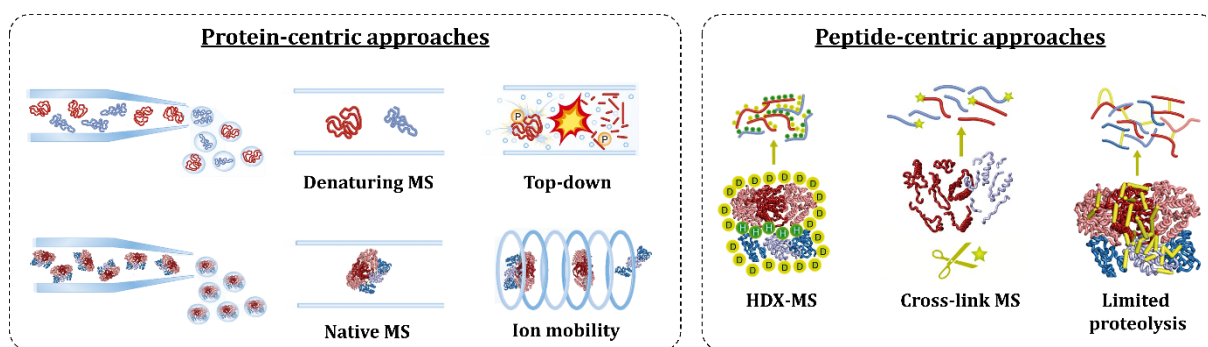


Figure 15: Summary of the protein-centric and peptide-centric approaches used in structural MS (adapted from ¹⁹⁶).

The work described in this manuscript is based chiefly on the characterization of purified proteins using top-down MS-based approaches, aiming for complete protein backbone sequencing. Hence, this bibliographic introduction mainly describes top-down and middle-down approaches for purified proteins. First, other approaches for higher-order characterization are briefly discussed, with a particular focus on native conditions analysis.

1. Protein-centric approaches in structural MS

A. Native MS

Native mass spectrometry (nMS) refers to analyzing intact proteins or biological complexes in their native or “near-native” state. This technique, introduced by Chait and Henion in the early 1990s, enables maintaining the non-covalent interactions of the biomolecules^{197, 198}. Under these experimental conditions, the protein’s folded structure is maintained in the gas phase, as well as the interactions with their partners (quaternary structure). The term “native MS” was formally coined by Heck and coworkers in 2004^{199, 200}.

In nMS, the pH and the ionic strength of the solution are controlled to keep the folded structure of the intact protein or protein complex. Thereby, the number of charges accommodated in the protein structure is lower compared to denaturing MS analysis, resulting in higher m/z MS signals (**Figure 16**). The analysis of large intact proteins or protein complexes in nMS is not trivial from an instrumental point of view. Indeed, the energy provided to the ions must be finely optimized to ensure their complete desolvation and transmission along the instrument while keeping the integrity of the non-covalent interactions in the protein (or complex). To this end, the pressure at the interface, the acceleration voltages, and radio frequencies must be modified. Historically, TOF and Q-TOF instruments have been the most widely used MS platforms for nMS experiments due to their (in principle) absence of upper m/z scan limit, resolution, and sensitivity. However, modifications on other high-resolution MS instruments such as Orbitrap^{51, 201}, and FT-ICR^{202, 203} have paved the way to adapt those instruments to conduct nMS experiments, providing higher sensitivity, resolution, and better desolvation²⁰⁴.

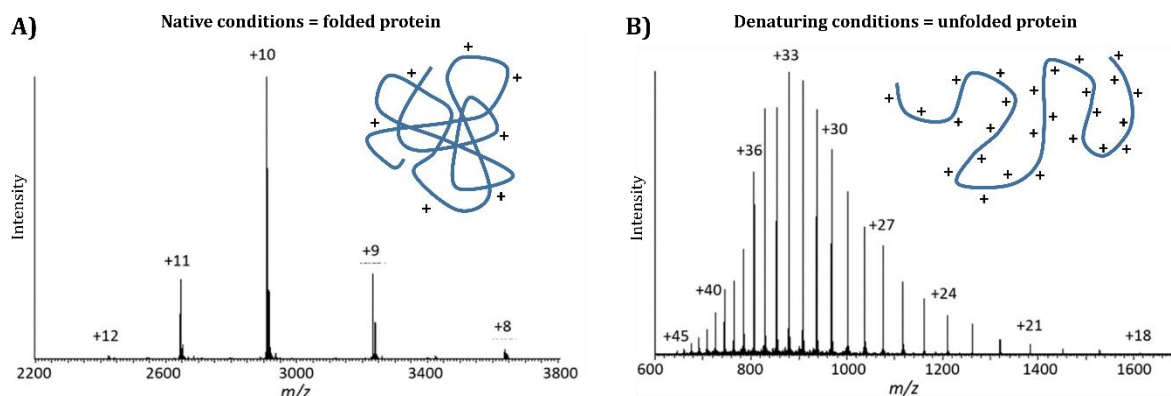


Figure 16: Mass spectra of carbonic anhydrase under (A) native and (B) denaturing conditions (adapted from ²⁰⁵).

Through the last years, technical performances in MS have improved these instruments' sensitivity, resolution, and ability to transmit ions in the high m/z mass range, enabling the analysis of large intact proteins or multiprotein complexes in their native state with masses of several MDa. Since its inception, nMS has been used to characterize a wide range of intact proteins and biological complexes of varying complexity and heterogeneity.

As mentioned previously, purified intact proteins are generally contained in buffers with non-volatile salts incompatible with MS analysis. Then, buffer exchange is generally required. This process can be performed manually with techniques such as ultracentrifugation or gel filtration, but they are extremely laborious and time-consuming. One alternative to overcome this sample preparation step is to couple nMS with non-denaturing LC methods^{206, 207}. The benefit of this combination is twofold: i) salts and excipients are removed within the LC dimension; ii) the characterization and relative quantitation of different populations of protein samples are improved.

Multiple information can be obtained thanks to nMS. The most prominent application of nMS is the mass measurement of the protein or complex and determining the stoichiometry of the interactions²⁰⁸⁻²¹¹, complexes' gas-phase stability^{212, 213}, or binding affinities^{214, 215}. Regarding the applications, nMS has been proven to be a versatile technique with contributions to the characterization of different proteins and biomolecules such as chaperonin²¹⁶, mAb-related products²¹⁷, hemocyanins²¹⁸, virus capsids²¹⁹ or membrane proteins²²⁰.

B. Other approaches in structural mass spectrometry

Another strategy to obtain information about the structure of a protein is the use of ion mobility spectrometry (IMS). In combination with nMS (nIM-MS), the conformation of the detected ions can be assessed thanks to the additional characterization dimension. IMS relies on measuring a drift time, corresponding to the time necessary for an ion to go through the ion mobility cell filled with background gas under the influence of a constant or time-dependent electric field⁷⁰. The drift time depends on the charge and mass of the ions and their shape²²¹. Ions with a compact and folded structure will move faster through the cell. In contrast ions with an extended or unfolded conformation will drift slower (higher drift time) because of increased collisions with the buffer gas molecules²²² (**Figure 17**). IM drift times of each ion can be converted into structural information (rotationally-averaged collision cross section (CCS)) to afford direct information about the conformation of the ions.

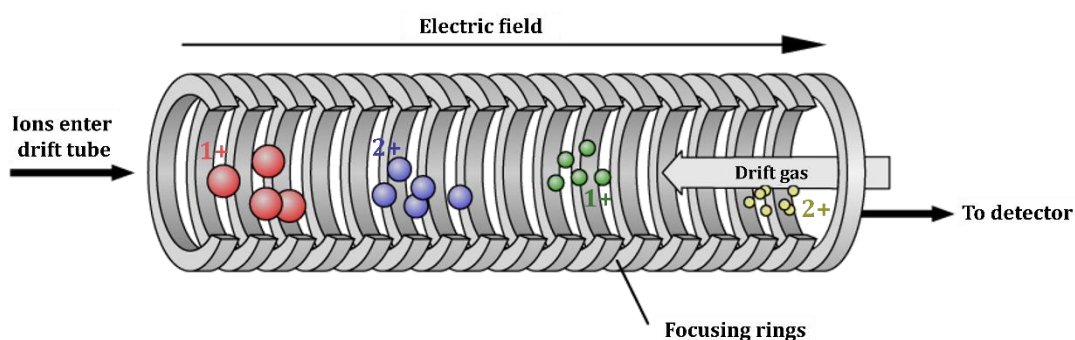


Figure 17: Schematic representation of drift tube ion mobility spectrometry (adapted from Jeff Dhal, CC BY-SA 3.0).

Several IMS devices have been introduced over the years⁷⁰. For IMS-MS, the drift tube²²³ applies a linear and homogeneous electric field in the tube, and the traveling wave (TWIMS)²²⁴, using continuous propagation of potential waves, are the two most used approaches. The former is considered the reference IM system since it is the sole IMS device with the ability to measure the mobility of ions as a primary method and calculate CCS from the Mason-Schamp equation. Otherwise, instrument calibration is mandatory for other IM devices to convert ions mobilities into CCS. IM-MS approaches have afforded valuable structural information to differentiate conformations of proteins and characterize their conformational landscapes²²⁵⁻²²⁸. More recently, an IMS-derived technique, so-called collision-induced unfolding (CIU), has been developed as an elegant option to circumvent the technical limitations of classical IM experiments. In this case, molecular ions are activated by means of energetic collisions with a background gas prior to the IM separation, allowing the analysis of macromolecular ions beyond their native state²²⁹. The internal energy of the ions increases due to the collisions, leading to a progressively unfolding of their conformation. Thereby, changes in conformation can be monitored as a function of collision energy, providing information about the number of transient unfolding states and the gas-phase stability of ions^{226, 229-233}.

2. Peptide-centric approaches

Peptide-centric strategies use a bottom-up-based approach to probe protein structures, dynamics, interactions, and conformations. Several methods can be employed, among which hydrogen-deuterium exchange (HDX), chemical cross-linking mass spectrometry (XL-MS), or limited proteolysis mass spectrometry (LiP-MS).

A. Hydrogen-Deuterium Exchange (HDX-MS)

HDX-MS, like other surface labeling techniques, assumes that the parts directly accessible by the solvent are more likely to be modified than those of the folded or bound regions are. This approach is mainly used as a relative technique, comparing the obtained results between two different states of an analyte, i.e., between the apo- and holo-state of a protein. Proteins, or protein complexes, are incubated in a deuterated solvent. Under these experimental conditions, labile amide backbone hydrogens are exchanged with deuterium, and the reaction is quenched using low pH (2.5) and temperature (0°C) to avoid back-exchange²³⁴ (**Figure 18**). Multiple deuteration times are commonly used to generate a whole HDX-MS data set. Then, the protein is enzymatically digested and analyzed by LC-MS/MS. The main idea is to assess the deuterium uptake of specific parts of the analyte as a function of the incubation time in a deuterated solvent to identify different relative deuterium uptakes and relay them to specific parts of the proteins. HDX-MS has been employed for many applications²³⁵ such as studying membrane proteins or comparing biosimilar products^{236, 237}, for instance. This technique also provides information regarding contact regions between multi-protein complexes, identifies the binding sites between proteins and ligands,²³⁸ investigates conformation of proteins or protein complexes, evidence for allosteric effects²³⁹ and is widely used for epitope mapping on biotherapeutics²⁴⁰.

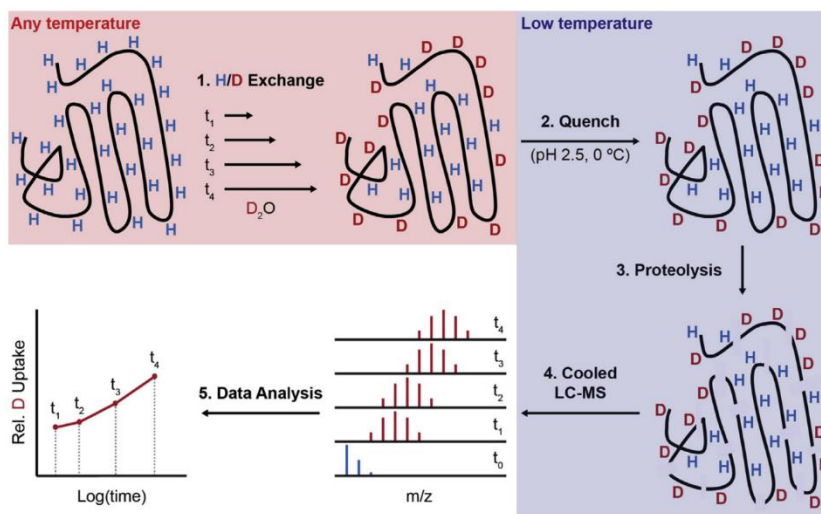


Figure 18: Schematic representation of workflow used for HDX-MS (from Trabjerg *et al.*²⁴¹).

B. Cross-link MS (XL-MS)

XL-MS is employed to evidence interactions in protein complexes or to study protein conformation¹⁸⁹. It relies on forming new covalent bonds, in the native state, between spatially close residues, using a cross-linker of a defined length²⁴². Cross-linkers are composed of a spacer indicating the distance between the two residues involved in the bonds and a reactive end group to target specific amino acid residues²⁴³. After the cross-linking reaction, proteins are enzymatically digested into peptides and analyzed by LC-MS/MS (**Figure 19**). The study of the cross-linked peptides identified gives information about the spatial proximity between two residues, either from the same protein (intra-distance) or from two different proteins (inter-distance)²⁴⁴. In purified systems, the structural information arising from XL-MS experiments can significantly contribute to speeding up the modeling

process or complementing protein or protein complexes characterization with other biophysical techniques¹⁸⁵. However, it has also been used for specific applications such as proteome-wide XL-MS to investigate protein-protein interactions at a system-wide scale²⁴⁵.

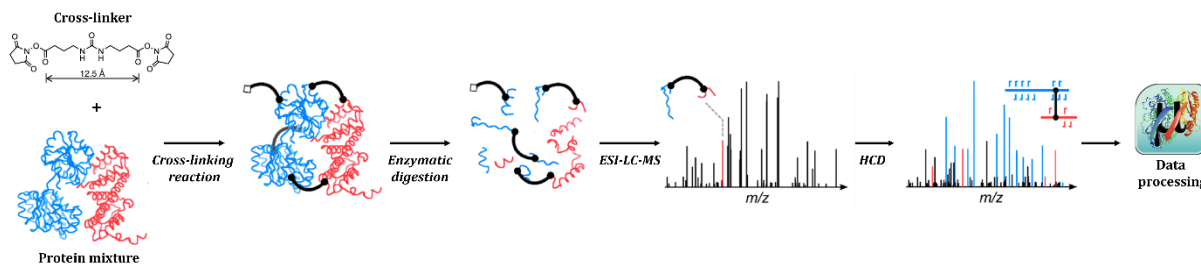


Figure 19: Schematic representation of workflow used for cross-link MS (adapted from ²⁴⁴).

C. Limited proteolysis

Limited proteolysis is built on the principle that a protein's digestion depends not only on its primary amino acid sequence but also on its higher-order structure²⁴⁶. Consequently, in native conditions, parts of the primary sequence on the surface or solvent-accessible regions of the protein are more likely to be digested into peptides than those in the core of the protein. LiP-MS typically provide information on the higher-order structures of proteins and on conformational changes of the protein^{247, 248}. Recent developments of LiP-MS on whole lysate demonstrate the ability of this approach to identify ligand targets (and off-target) at the cellular level^{249, 250}. Moreover, this technique opens the way to new applications, such as thermal proteome profiling, bridging structural MS and bottom-up proteomics together, as applied by Picotti and coworkers, with the determination of the thermal stability of proteins on a proteome-wide scale at domain-level resolution, and underlying the role of specific types of proteins²⁵¹.

3. Protein primary structure characterization: a peptide- or a protein-centric approach?

Intact protein analysis by structural MS techniques allows obtaining information about the primary structure or the higher order structure of proteins and protein complexes. Primary structure information includes the intact mass of the protein and aims at fully characterizing the amino acid sequence. It allows the identification of proteoforms – referred to as each individual molecular form of an expressed protein² – and the precise localization of post-translational modifications (PTMs) or amino acid variations, using the top-down approach^{252, 253}.

A. Bottom-up versus top-down

The bottom-up (BU) approach is the gold-standard method to perform protein primary structure characterization. This strategy is based on the reduction, alkylation, and enzymatic (using trypsin, pepsin, or a combination of proteases) digestions of the proteins, followed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of the generated peptides. Even if routinely performed, this approach has a series of limitations. First, digestion often occurs at basic pH and can lead to artifactual modifications of the protein (e.g., deamidation, oxidation²⁵⁴⁻²⁵⁶). In some cases, peptide mapping approaches can also lead to incomplete sequence coverage and, thus an absence of information on the uncovered regions. BU approaches are mainly performed using collision-based fragmentation techniques, which can induce the fragmentation of relatively labile modifications such as phosphorylation or glycosylation (in the particular cases of ADCs, these techniques induce the fragmentation of the payload), which complicates the correct location and characterization of these PTMs. More importantly, since proteins are previously digested, and the structural information is

inferred from the resulting peptides, the correlation between the different protein modifications is lost (**Figure 20**). For all the aforementioned reasons, the correct characterization of protein proteoforms is hampered when using BU approaches. Nevertheless, correctly characterizing proteoforms or primary sequence variants of a protein of interest is crucial since those different forms have been related to disease mechanisms or associated with an altered biological or therapeutic activity.

Instrumental development, along with the evolution of data processing tools have paved the way for alternative strategies to perform protein sequencing. During the last two decades, top-down (TD) MS has emerged as a complementary approach to bottom-up, as the former analyzes proteins at the intact level. Briefly, intact proteins are ionized and fragmented in the gas phase (MS/MS) normally using different activation techniques, allowing fragmentation of protein backbone and providing information about proteoforms through PTMs assessment²⁵⁷, and sequence variants. Compared to BU approaches, TD presents several main advantages. The low number of steps in the sample preparation allows to drastically decreasing the risk of introducing chemical artifacts commonly observed in bottom-up. In addition, the intact mass of the protein and its fragmentation spectrum are correlated, which facilitates the identification and characterization of proteoforms, especially those with several co-occurring modifications (**Figure 20**).

Furthermore, TD-MS also appears as an orthogonal and complementary technique in the structural MS field for several reasons: i) the structural information provided with this approach is directly linked with the intact structure of the protein, ii) it provides structural information that cannot be characterized with the different approaches encompassed within the structural MS toolbox.

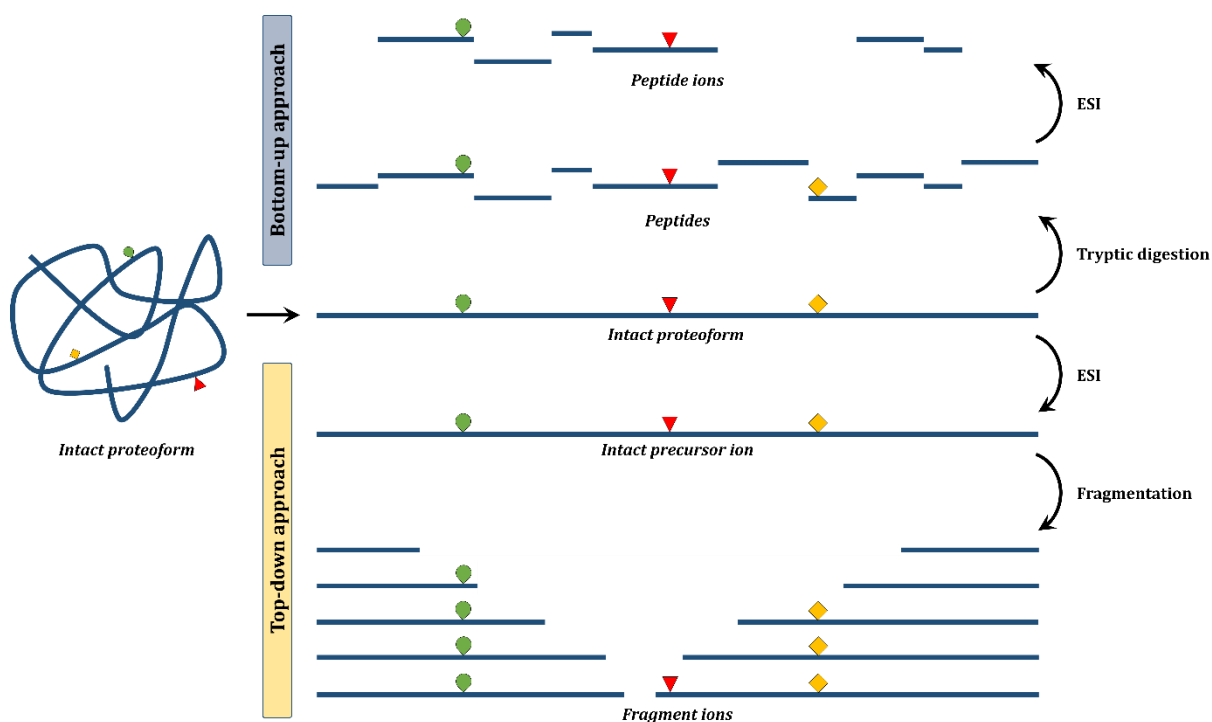


Figure 20: Proteoform identification using the bottom-up (top) or the top-down approach (bottom).

B. Definition of top-down and middle-down MS

Fragmentation of intact proteins can be performed in many conditions and for many purposes. Since the first papers in the literature using the “top-down” approach in the early 2000s²⁵⁸⁻²⁶², many studies have been released using this terminology. Nevertheless, the term “top-down” has often been misused as no standardized nomenclature was defined. To this end, Lermyte *et al.*²⁶³ proposed a standardized lexicon in 2019 to clarify TD-MS approaches.

This paper describes multiple strategies, with a specific terminology employed depending on the input sample and the resulting information obtained after analysis. These approaches are summarized in **Figure 21**.

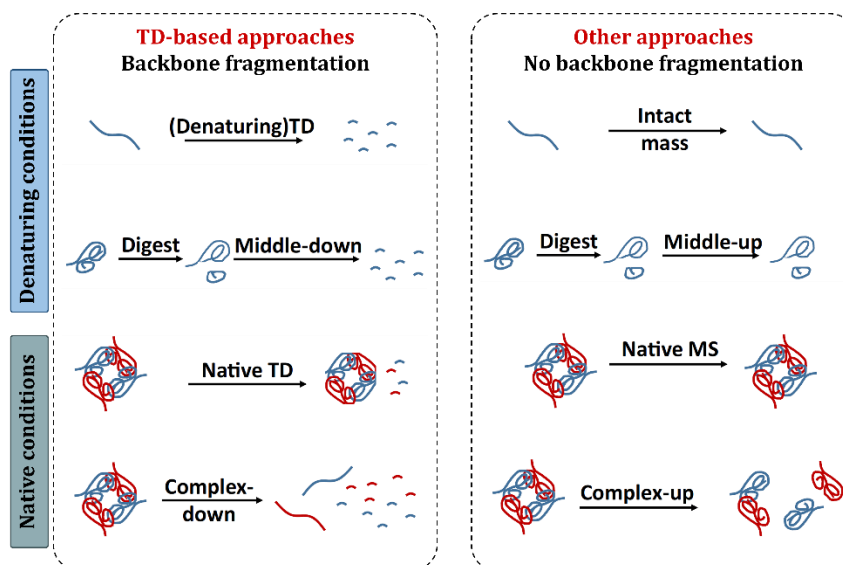


Figure 21: Illustration of top-down related approaches, inducing or not backbone fragmentation (adapted from ²⁶³).

Before any top-down experiments, the mass of the intact protein is generally required to ascertain the protein integrity and provide information about the potential presence of modifications or sequence variants. If the analysis is performed under native or denaturing conditions, it is referred to as native MS and intact mass measurement, respectively. These approaches are not considered top-down experiments as such.

The term top-down refers to experiments aiming to identify a protein and fully characterize its amino acid sequence. Classical TD experiments are performed in denaturing conditions (i.e. H₂O/ACN/HCOOH) and fragmented in the gas phase using one or multiple dissociation techniques. If this experiment is performed under native conditions, for instance with ammonium acetate at pH 6.8, the term native TD should be employed.

The analysis of protein complexes is also possible following this approach. These experiments are always performed under native conditions, as protein complexes are not maintained in a denaturing environment. In this case, specific terminology is used. If the fragmentation of the whole complex induces backbone cleavage, this experiment is called native TD, as for intact protein analysis. However, softer fragmentation can be used to dissociate the complex to obtain the mass of each subunit, named complex-up, and additional fragmentation of one (or multiple) subunits of the complex can be performed to obtain information about their specific sequence, referred to as complex-down experiments.

In addition, progresses in sample preparation have made possible the analysis of partially digested proteins, leading to large peptides/subunits above 5-7 kDa^{21, 264}. This approach is mainly used to analyze monoclonal antibodies (mAbs)^{10, 22, 265}. Similarly to protein complexes approaches, mass measurement of the intact subunits from limited proteolysis – usually separated through liquid chromatography prior to MS analysis – is coined middle-up, while the fragmentation of the amino acid backbone of these subunits is called middle-down (MD).

In the work described in this manuscript, only the top and middle-down strategies were used for protein sequencing.

4. Instrumentation and methods for top-down and middle-down MS

Classic top-down and middle-down experiments consist of three main steps: sample preparation, protein analysis and fragmentation, and data processing. Before any fragmentation in TD/MD-MS analysis, intact mass measurement of the protein is usually performed to ensure its integrity and identify potential proteoforms that are present. It also allows the selection of the precursors used for fragmentation. Once the precursor is selected, it undergoes fragmentation, and fragment ions with different charges and sizes will be generated. Taking into account the different stages that compose a typical TD/MD-MS experiment, several considerations have to be taken into account to perform this approach. These experiments require a high accuracy mass measurement of both the intact proteins, and the fragment ions, while high-resolution MS measurement is mandatory to differentiate and distinguish the isotopic profile of a plethora of fragment ions. Finally, high efficient and complementary fragmentation techniques have to be included to induce the fragmentation (if possible) of the protein's backbone at each protein residue (if possible) and afford the most complete primary structure characterization.

In this context, to conduct TD- and MD-MS experiments, the MS platforms have to fulfil specific requirements in terms of mass accuracy, resolution, and versatility to be perfectly adapted to provide the most complete characterization of different proteins. Thereby, the most suited MS instruments, and fragmentation techniques will be presented in the next sections, highlighting the evolution of this approach during the last years.

A. MS instrumentation

TD-MS experiments have historically been performed on triple quadrupole and quadrupole-ion trap instruments, but the low resolving power of these analyzers rapidly appeared as a clear limitation²⁶⁶. Hence, the introduction of FTICR instruments enables to ensure the need for high-resolution accurate-mass analyzers to ensure the correct assignment of the ion charge state and elucidate between overlapping peaks, making them widely used for TD-MS experiments²⁶⁷.

Nowadays, Orbitrap-based and FTICR instruments are used for most of the top-down experiments mentioned in the literature. Indeed, Tribrid Orbitrap instrument design allows the combination of multiple fragmentation techniques, which is particularly useful to provide extensive sequence coverage in TD-MS experiments. Some ESI-Q-TOF or MALDI-TOF-TOF instruments are also used, but to a lesser extent²⁶⁸.

TD/MD-MS experiments require MS/MS acquisition parameters to be optimized. The vast amount of information generated by the fragmentation of multiply charged protein precursors results in complex spectra, and a key factor to consider is the signal-to-noise (S/N) ratio. The S/N ratio directly affects the deconvolution quality and the fragment ions' assignment. Several possibilities are available to maximize the S/N ratio. First, performing multiple micro scans per spectrum (usually 2-5) to eliminate the background noise is possible. Microscans correspond to the number of transients acquired to obtain one (recorded) scan. In addition, a substantial number of recorded scans from the same precursor and fragmentation technique are averaged to decrease the S/N ratio.

B. Fragmentation techniques

A plethora of fragmentation techniques has been developed to analyze intact proteins. They rely on different fragmentation mechanisms by increasing the internal energy of the ions with different principles, prior to the dissociation of the protein backbone. Fragmentation modes can be mainly grouped into different categories: collision-based dissociation, electron-based dissociation, and resonant photodissociation. Although not a dissociation technique, ion/ion reactions such as proton transfer charge reduction (PTCR) can be used for charge reduction, and MS/MS spectrum decluttering.

This section describes the different types of fragmentation, with a particular focus on the fragmentation techniques used for the work presented in this manuscript.

i) Collision-based dissociation

Collision-based dissociation covers two techniques: collision-induced dissociation (CID) and higher-energy collision dissociation (HCD). CID was first described in the late 1960s by McLafferty and Bryce²⁶⁹ and Jennings²⁷⁰. This technique is also called collision-activated dissociation (CAD). Isolated precursor ions are accelerated to increase their kinetic energy and then allowed to collide with neutral gas molecules (generally helium, argon, or nitrogen) in an ion trap. After the collision, some of the kinetic energy from the ion is converted into internal vibrational energy, leading to peptide bond cleavage. The protein fragmentation depends on the activation period, as dissociation is led by repeated multiple collisions (discrete activation), classing CID as a slow activation method.

Using CID, low-energy pathways are favored²⁷¹. Hence, *b* and *y* ions tend to be observed, resulting from the fragmentation of the peptide bond (N-C(O) bond), as described in **Figure 22**. The adopted explanation of the protein backbone fragmentation is CID is the ‘mobile proton’ model^{58, 59}. Briefly, a proton from a basic site (i.e., N-terminal or basic residues such as arginine or lysine) is mobilized by collisional activation of the ion. This proton moves to a backbone heteroatom (carbonyl oxygen or amide nitrogen) to trigger the activation reaction. Hence, CID is a charge-directed fragmentation, demonstrating preferential activation pathways depending on the amino acid sequence^{58, 59}.

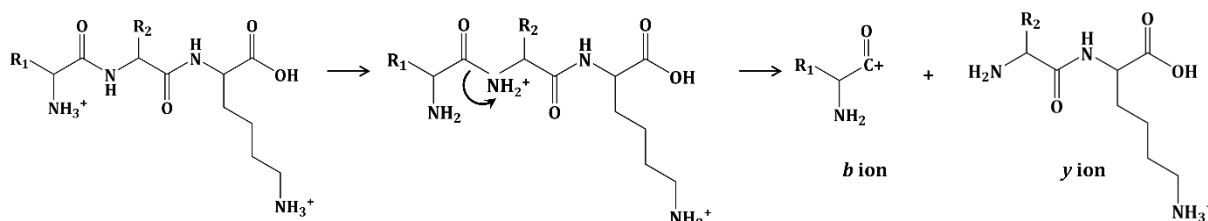


Figure 22: Fragmentation scheme for production of *b* and *y* ion types by collision-induced dissociation (from Syka *et al.*⁶²).

As CID tends to break weak bonds preferentially, labile PTMs are generally not retained and cannot be observed in the MS/MS spectrum after CID fragmentation. Thus, this method is not optimal for the identification of labile PTMs. Another limitation of CID when applied to intact proteins is the high number of degrees of freedom into which the energy can potentially dissipate^{272, 273}, hampering the comprehensive fragmentation of the protein backbone²⁷⁴.

HCD – sometimes referred to as beam-type CID – has been developed to increase intact protein fragmentation. It is similar to CID, but ions accumulate higher internal energy, which opens the door to additional fragmentation pathways that cannot occur using CID. HCD is typical of Orbitrap instruments and is performed in a collision cell⁶¹. On the Orbitrap Eclipse (Thermo Fisher Scientific), the tribrid geometry of the instrument allows to perform both CID (in the linear ion trap) and HCD (in the ion routing multipole).

Moreover, an additional collision-based activation called in-source dissociation (ISD) can be mentioned. In a few words, ions exiting the source region are activated during the transfer to the ion-focusing region, similarly to CID but without any specific precursor selection²⁷⁵. While ISD is not of substantial interest for intact protein fragmentation in TD experiments, it has been widely used to improve desolvation and remove adducts^{200, 276}. Another application of ISD is disassembling protein complexes before the fragmentation of the subunits²⁶⁸.

Both CID and HCD fragmentation have been widely used in top-down and middle-down MS as they are implemented in most of the MS instruments available. Their use is nevertheless limited for large proteins due to the impossibility to obtain exhaustive amino acid cleavage and to retain labile PTMs²⁷⁴.

ii) Electron-based dissociation

Electron-based dissociation techniques (ExD) are alternatives to CID and HCD. The first ExD method was proposed by Zubarev and coworkers²⁷⁷ with electron capture dissociation (ECD) for the characterization of intact proteins of several kDa, demonstrating the substantial superiority of ECD to obtaining extensive sequence coverage. ECD experiments rely on the irradiation of multiply charged precursor ions with low energy electrons (<0.2 eV) from a heated metal filament (tungsten or barium), resulting in charge-reduced radical ions that have captured the electron but are not dissociated. Then, protein backbone fragmentation is performed by radical-driven reaction leading to *c* and *z*[•] ions (**Figure 23**), corresponding to the N- α bond cleavage. Several mechanisms trying to explain this reaction have been proposed, known as the Cornell mechanism^{277, 278} and Utah-Washington^{279, 280}. However, no consensus has yet been found²⁸¹. In addition to *c* and *z*[•] ions, *c*[•] and *z* fragment ions are likely to be observed^{282, 283}. A few additional fragments ions can be found from minor pathways, but they are generally not considered for fragment identification after ECD in top-down²⁷². In addition, disulfide bond breakage has also been reported using ECD²⁸², but with a higher efficiency in peptides compared to proteins²⁸⁴.

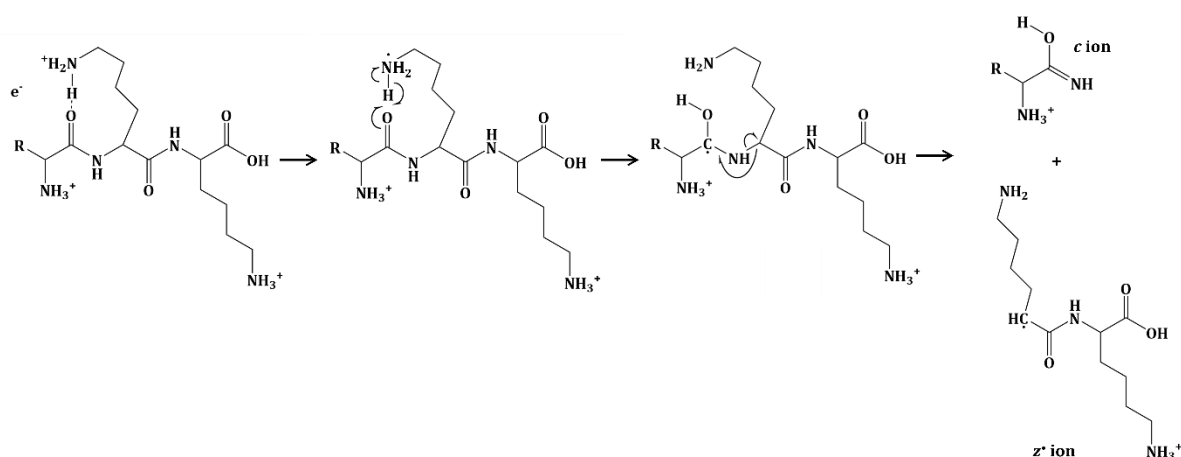


Figure 23: Fragmentation scheme for production of *c* and *z*[•] ion types of a low energy electron with a multiply charge peptide (from Syka *et al.*⁶²).

One limitation of ECD is the necessity to trap low-energy electrons and precursor ions. While this is easily possible on FT-ICR instruments, it has been much more complex to implement on other instruments, restricting its use. Then, Syka and coworkers developed a similar ion-ion reaction technique, named electron transfer dissociation (ETD), based on the electron transfer to the precursor cation from a radical anion⁶². The pathways involved in the fragmentation are the same as those of ECD, producing *c* and *z* ions and their radical counterparts (*c*[•] and *z*[•]). Radical anion used for ETD is usually fluoranthene or anthracene, generated from negative chemical ionization^{285, 286}. For ETD reaction, precursor cations, and ETD reagent anions are stored at the extremities of the ionic trap in potential wells and collide when the potential wells are removed.

ECD and ETD are beneficial in top-down MS for two reasons: they fragment the sequence at random cleavage sites (except proline N-terminal bonds due to its ring structure^{282, 287}), allowing higher sequence coverage compared to CID/HCD²⁷⁷, and can preserve labile PTMs²⁸⁸.

Because ETD (and ECD) are ion-ion reactions, the higher the charge state (i.e., charge density), the better the fragmentation will be, highlighting the high dependency of the technique on the charge state

(Lermyte 2018). One shortcoming of these electron-based dissociation techniques is the possibility for charge reduction by proton transfer (PT) or non-dissociative reactions (ETnoD/ECnoD)²⁸¹. Charge reduction by PT generally occurs for precursor ions with low charge density, where a proton is removed from the precursor cation (**Figure 24**)^{281, 289-291}. It can also happen when the reaction time between the ETD reagent and the precursor ion is not long enough. For ETnoD, backbone cleavage occurs, producing *c* and *z* ions, but they are held together by non-covalent interactions, preventing their dissociation^{281, 289, 290}.

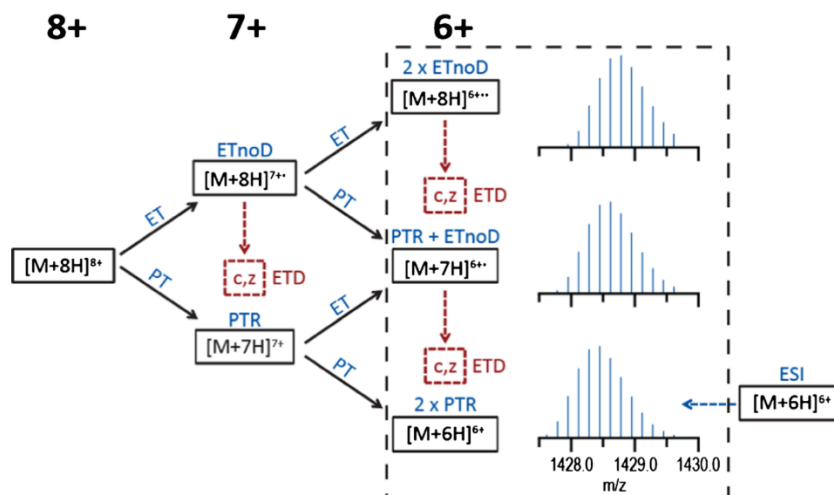


Figure 24: Reaction pathways of $[M+8H]^{8+}$ ubiquitin. Fragmentation of the protein into *c* and *z* ions is depicted in red. Reactions where no fragments are observed (ETnoD and PTR) are represented in blue. Figure from Lermyte *et al.*²⁸⁹.

To circumvent this limitation and the charge state dependency of ETD/ECD and improve the fragmentation, activated ion ETD/ECD (AI-ETD/ECD) have been proposed by Horn *et al.*²⁹². It relies on the slow-heating precursor cations simultaneously to ETD/ECD fragmentation, using infrared (IR) photons, for example²⁹³. This technique allows the destruction of the higher-order structure of the protein and better detection of *c* and *z* ions. It demonstrates more extensive sequence coverage than ETD^{294, 295}.

Multiple other electron-based dissociation methods have been derived from ETD and ECD, but they are not widely used in top-down MS. It includes “hot” ECD (hECD), electronic excitation dissociation (EED), negative electron transfer dissociation (NETD), electron detachment dissociation (EDD) or electron capture-induced dissociation (ECID)²⁷¹. Some devices, such as the Omnitrap²⁹⁶ (Fasmatech) or the electron-activated dissociation (EAD)²⁹⁷ device, allow performing different fragmentation techniques based on electron dissociation.

iii) Photodissociation

Photodissociation is the fragmentation of a protein (or a peptide) due to the absorption of photons²⁹⁸, which can be performed by infrared multiphoton photodissociation (IRMPD) or ultraviolet photodissociation (UVPD). These techniques use lasers to produce photons, either in the IR or the UV domains, and specific instrumental setups (mirrors) allow the beam of photons to intersect with the precursor ions²⁹⁸. Numerous photon sources have been used for photodissociation, but they will only be detailed briefly here. It includes – but is not limited to – CO_2 lasers, excimer lasers, Nd:YAG lasers, gas discharge lamps, or femtosecond titanium sapphire lasers^{69, 298, 299}. Critical parameters of photodissociation are the energy of the photons, the activation time, and the collision cross-section of the precursor ion.

For IRMPD, photons are generally produced using a CO₂ laser with a wavelength of 10.6 μm. Regarding the low energy of the photons (~ 0.1 eV per photon), protein backbone dissociation with this technique is a stepwise process that requires the accumulation of many photons to fragment the protein. Similarly to CID, as this process needs energy accumulation, low energy fragmentation pathways are favored, mostly leading to the formation of *b* and *y* ions^{298, 300, 301}.

On the opposite of IRMPD, UVPD photons have high energy resulting in the direct excitation of ions to excited electronic states. It allows protein backbone fragmentation via additional high-energy dissociation pathways not observed with stepwise dissociation. Hypotheses have been made to explain the fragmentation pathways involved in UVPD dissociation, but no evidence has been found^{302, 303}. In a few words, two relevant mechanisms have been identified. On the one hand, direct dissociation, induced by the excitation of an electron to a dissociative orbital, can be observed and lead to fragmentation pathways unique to UVPD^{302, 304}. On the other hand, internal energy conversion to vibrational energy can induce the fragmentation of low-energy bonds (i.e., *b* and *y* ions). For UVPD, all the ion types are likely to be observed (i.e., *a*, *b*, *c*, *x*, *y*, and *z*).

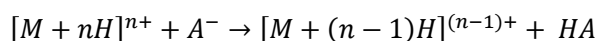
Most TD-MS experiments have been carried out using 193 nm wavelength lasers⁶⁹ that are well suited for intact protein analysis and demonstrate high sequence coverage^{257, 305-307}. Nevertheless, these lasers are not implemented on commercial instruments and must be installed by the user. To broaden the use of UVPD, particularly for TD-MS, a 213 nm laser (solid state Nd:YAG) has been released on tribrid instruments by Thermo Fisher Scientific, allowing to perform UVPD straightforwardly and providing similar sequence coverages^{308, 309}. Comparison of the 213 nm and 193 nm by Lanzillotti and Brodbelt did not reveal any apparent differences in the fragmentation pathways and efficiency³¹⁰. Still, the lower energy of the 213 nm laser requires a longer activation time compared to the 193 nm to obtain a similar fragmentation efficiency.

iv) Hybrid fragmentation techniques

Hybrid activation methods have been designed to increase the sequence coverage and circumvent the limitations inherent to each fragmentation mode. They consist of the combination of two fragmentation modes in simultaneous or sequential activation of the precursors. Most of them are based on electron-based fragmentation and aim to bring additional energy to get rid of the ETnoD process³¹¹. All fragment ions from ETD (same for ECD) are submitted to broadband HCD (ETHcD)³¹², CID (ETciD), or UVPD (ETuVPD)³¹³, allowing the fragmentation of the charge-reduced or the non-dissociated species from the first ETD step. These methods demonstrate a considerable increase of the sequence coverage thanks to both a higher number of *c* and *z* ions from ETD but also because of the complementary ion types generated by HCD/CID (*b* and *y*) or UVPD (*a*, *b*, *x* and *y*)^{10, 312, 314}.

v) Proton transfer charge reduction

Proton transfer charge reduction (PTCR) is an ion-ion reaction proposed in 1998 by McLuckey and Stephenson³¹⁵. PTCR, while not a fragmentation technique, aims at simplifying product ion spectra from multiply charged precursor ions. Indeed, MS/MS spectra obtained after fragmentation contain a vast amount of information, resulting in many overlapping peaks that can prevent the identification of the fragments. To tackle this issue, multiply charged fragment ions are submitted to an additional reaction with a perfluorinated anion to give a charge-reduced cation and a neutral molecule, as described here:



This additional ion-ion reaction step reduces the complexity of the MS/MS spectra by distributing the fragment ions over a broader *m/z* range, limiting the overlap of the isotopic profiles. This approach is usually used after ETD-based fragmentation – which gives the more complex spectra – and has

demonstrated a significant increase in the sequence coverage^{316,317}. PTCR has found some applications recently in TD-MS, but it is not widely used for now^{307, 316, 318}.

C. Data processing

Data processing of the MS/MS spectra acquired consists of two main steps: deconvolution of the fragmentation spectra to obtain the zero-charge mass of the fragment ions and the assignment of those masses to the specific fragments of a candidate sequence.

i) MS/MS spectra deconvolution

Deconvolution of the MS/MS spectra is critical to TD/MD-MS workflows. Different strategies can be used, depending on the resolution used for the data acquisition. High-resolution acquisition at both MS1 and MS2 levels is the most used method for purified protein as it provides high-quality data. Nevertheless, for proteins >30 kDa or in complex mixtures, MS1 spectra are commonly acquired at low resolution (typically 17.5k), and for complex mixtures analysis, LC-MS/MS is required instead of direct infusion. It can also happen that both MS1 and MS2 spectra are acquired at low resolution. In these cases, other algorithms based on the charge state envelopes rather than the isotopic envelopes are used. Raw data are converted into a list of peaks using either the monoisotopic mass (M) or the single charge mass (M+H⁺). The accuracy of the deconvolution significantly affects the results of the assignment of the fragments. As mentioned before, S/N is a crucial factor that influences deconvolution. Hence, MS/MS spectra are usually averaged before deconvolution to limit the background noise.

The most widely used method for deconvolution is peak assignment using dedicated algorithms³¹⁹. The deconvolution process can be divided into four steps (i) candidate isotopomer enveloped generation; (ii) theoretical isotopomer enveloped generation; (iii) candidate and theoretical isotopomer envelopes comparison using a scoring function; (iv) monoisotopic mass extraction. These four steps have been extensively described in a review by Zhong *et al.*³¹⁹ and will only be summarized in this section.

The moving window method is employed to determine the monoisotopic peak and the charge state of candidate isotopomer envelopes. Once identified, theoretical isotopomer envelopes are generated based on the monoisotopic peak and the charge state of each candidate sequence. The theoretical and candidate envelopes are then compared using a scoring function. All candidate envelopes passing the scoring function threshold are selected, and the monoisotopic mass and the charge state are added to the list of deconvoluted masses.

Multiple deconvolution algorithms have been developed and can be used to process data, such as TRASH³²⁰, MSDeconv³²¹, UniDec³²², or MASHSuite³²³. Among them, Xtract³²⁴, derived from TRASH, has been used for the work presented in this manuscript. A comparison of some deconvolution tools has been made recently by Tabb and coworkers, pointing out the pros and the cons for each of them³²⁵.

ii) Fragments assignment to a candidate sequence

Once the monoisotopic peak list has been generated, fragments are matched against a candidate sequence. For the case of complex protein mixtures, identification is performed using a database containing all the potential proteins present in the sample, as for conventional bottom-up approaches. However, dealing with purified intact protein only requires the sequence of interest, making the search much more convenient. The algorithms calculate the theoretical mass of each possible sequence fragment, depending on the fragmentation technique used. Then, monoisotopic masses from the deconvolution under the mass tolerance threshold (typically 10 ppm) are matched, allowing obtaining the sequence coverage.

The most employed software for matching the fragment of purified proteins is ProSight Lite, developed by the Kelleher group³²⁶, and incorporated in the Proteome Discoverer suite commercialized by Thermo. Other tools, including MASHSuite Pro³²⁷, TopMG³²⁸, or TOPPic³²⁹ can also be used for purified proteins but are more dedicated to the analyses of complex mixtures.

In the work described in this manuscript, only ProSight Lite was used. This software also includes a score to estimate the confidence of the identification, referred to as P-Score and defined as follows³³⁰:

$$P_{f,n} = \frac{(xf)^n \times e^{-xf}}{n!}$$

In this equation, $P_{f,n}$ is the probability of random protein matching, f is the number of monoisotopic masses in the input list, n is the number of random fragment ion hits, and x is equal to the average probability of the fragment ion mass of random protein matching³³⁰. It can be described as the probability that a random sequence could account for the matching ions.

In addition, the vast majority of the software consider only the theoretical terminal fragments to match experimental fragment ions. However, it has been demonstrated that fragmentation of proteins also generates internal fragments that accounts for a substantial part of the fragment ions observed on a MS/MS spectrum³³¹⁻³³³. These fragment ions are not identified by conventional algorithms resulting in a potential loss of information. Hence, several groups proposed algorithms allowing considering internal fragments^{334, 335}, such as ClipsMS, developed by Loo and coworkers³³⁵. Several studies reported the benefits of internal fragments to improve sequence coverage in top-down experiments^{332, 336}. Nevertheless, some limitations regarding the increased risk of false positives have also been pointed out³³⁷.

5. Main applications of top-down MS

Top-down can be used for different purposes: identification of proteoforms in complex mixtures, extensive protein backbone sequencing of purified proteins, and structural characterization of native proteins, and protein complexes.

A. Proteoform identification in complex mixtures

As mentioned, bottom-up is not the most suitable approach for identifying proteoforms. However, these molecular species are key to understanding biological processes. Information about proteoform variations will contribute to a better understanding of many diseases or biological disorders and are of particular interest in translational and clinical research³³⁸. Different studies on identifying proteoforms to understand better their role in diseases have been published in the literature^{339, 340}. It includes studies about neurodegenerative diseases^{341, 342} or heart diseases^{339, 343, 344}. Other studies focused on histones, which are known to be complex mixtures of multiple proteoforms^{306, 345}, or the identification of proteoforms in HeLa cells³⁴⁶. To date, few thousands of proteoforms have been identified in complex samples such as human fibroblast³⁴⁷ (up to 2,000 proteoforms) or human H1299 cells³⁴⁸ (more than 5,000 proteoforms).

Some challenges to improve the identification of new proteoforms in complex mixtures still need to be addressed. They include limited protein solubility, high dynamic range, need for improved separation prior to MS, and data processing and linking proteoforms to their biological function³⁴⁹.

With this approach, the identification of proteoforms relies only on the detection of a few specific diagnostic ions, but obtaining an extensive sequence coverage is not the main objective.

B. Purified intact proteins analysis

In this context, all the efforts are focused on characterizing one protein (or family of proteins) of interest. This approach aims to provide the most comprehensive characterization of the primary sequence by combining different fragmentation techniques and selecting different precursor ions. Then, the different results are gathered to improve the overall sequence coverage and assess PTM location. Typically, this methodology has been applied to reference proteins (e.g., myoglobin, ubiquitin, carbonic anhydrase), mainly as proof of concepts for new techniques developments^{310, 350, 351}. However, it has been extensively used to characterize therapeutic proteins, as they can be obtained in an easy way¹⁰. Their comprehensive characterization is also of utmost importance to ensure the quality of the biotherapeutics. These applications are detailed in **Chapter 3** of the bibliographic introduction. Moreover, few other applications of purified proteins have been reported for endogenous phospholamban proteoforms identification³⁵², apolipoprotein proteoform analysis³⁵³, biomarker discovery³⁵⁴, phosphorylated proteins³⁵⁵, or proteasome subunits quantification³⁵⁶.

While extensive sequence coverage can now be attained thanks to the combination of multiple fragmentation techniques for small proteins^{296, 310, 357}, the comprehensive dissociation of large biomolecules remains a challenge, with sequence coverage <70%^{310, 350, 351}.

C. Native top-down approaches

Finally, top-down has been used for higher-order structure characterization in native conditions³⁵⁸. It is most commonly employed for complex-up and complex-down experiments, allowing obtaining information about the quaternary structure of the complex and primary sequence. A complex is introduced in the mass spectrometer under native conditions to measure the mass of the complex. Then, complex subunits are released by low-energy fragmentation, allowing the identification of each subunit. Finally, subunits are subsequently fragmented to obtain information about the primary sequence¹⁹¹. Native top-down has been applied to various samples, such as membrane proteins^{359, 360}, hemoglobin³⁶¹, tau/CRL01 complex³⁶², or SARS-CoV-2 spike protein³⁶³. For instance, Ruotolo and coworkers performed native top-down on transmembrane protein using infrared photoactivation, probing the potential of infrared photoactivation for micelles dissociation and enabling sequence coverage of membrane proteins up to 60%³⁵⁹.

Native top-down has also been employed to probe protein structure, using electron-based fragmentation techniques as they can generate backbone fragmentation while maintaining the higher-order structure^{271, 364, 365}. Other groups focused on protein-metal complexes characterization by TD-MS, such as Wongkongkathep *et al.*, who employed native top-down for cobalt and manganese binding site determination to the C-terminal region of α -synuclein³⁶⁶. In another study, Nshanian and coworkers identified the binding of CLR01 to the tau protein thanks to TD-MS³⁶².

These are some of the examples of different applications of TD/MD-MS showcasing the versatility of these techniques to provide not only information about the primary structure of proteins but also to afford further insights about higher order structure of protein and protein complexes, emerging as an important technique in the structural MS toolbox.

Chapter 3:

Introduction to monoclonal antibody-related therapeutic proteins

Antibodies, or immunoglobulins (Igs), are produced by differentiated B-lymphocytes, playing a vital role in the humoral immune response³⁶⁷. Their role is to recognize and neutralize antigens (Ag). Among IgGs, monoclonal antibodies (mAbs) are specific for one unique target epitope on an Ag, leading to the binding of the two structures. Then, the mAb/Ag immune complex is eliminated through different processes, including antibody-dependent cell toxicity³⁶⁸ (ADCC), antibody-dependent cell phagocytosis³⁶⁹ (ADCP), and complement-dependent cytotoxicity³⁷⁰ (CDC).

Because of their high specificity and efficiency, mAbs and related therapeutic proteins are among the fastest-growing biopharmaceutical products^{3, 371, 372}. Since the approval of the first mAb in 1986³⁷³ (Muromonab-CD3), the United States Food and Drug Administration (FDA) and/or the European Medicines Agency (EMA) have approved the release of more than 130 mAbs, and about 20 are currently under review^{3, 372}. In 2022, the market size of mAb-related products was valued at more than \$200 billion³⁷⁴. These biotherapeutics proteins have been designed for various indications, such as inflammatory diseases, autoimmune disorders, or cancers³⁷¹.

1. Monoclonal antibodies and derivative biotherapeutics

A. Structure of immunoglobulins

Among the five classes of Igs in mammals, most of the clinically available mAbs are immunoglobulins G (IgGs). IgGs are Y-shaped glycoproteins with a total molecular weight of ~ 150 kDa. They are constituted of two identical heavy chains (Hc) of ~50 kDa and two identical light chains (Lc) of ~25 kDa held together via covalent inter-chain disulfide bonds (**Figure 25.A**). The Lc is composed of one variable domain (VL) and one constant domain (CL), while the Hc comprises one variable domain (VH) and three constant domains (CH1, CH2, and CH3). The Lc and the VH and CH1 domains of the HC constitute the fragment antigen binding (Fab) region (~50 kDa), forming a flexible arm and linked in the hinge region by disulfide bridges to the other Fab to establish the dimeric structure F(ab')₂ (~100 kDa). The antigen-binding site specificity of the Fab is determined by the complementary-determining regions (CDRs) through three hypervariable loops located in the variable domains of the Lc and the Hc.

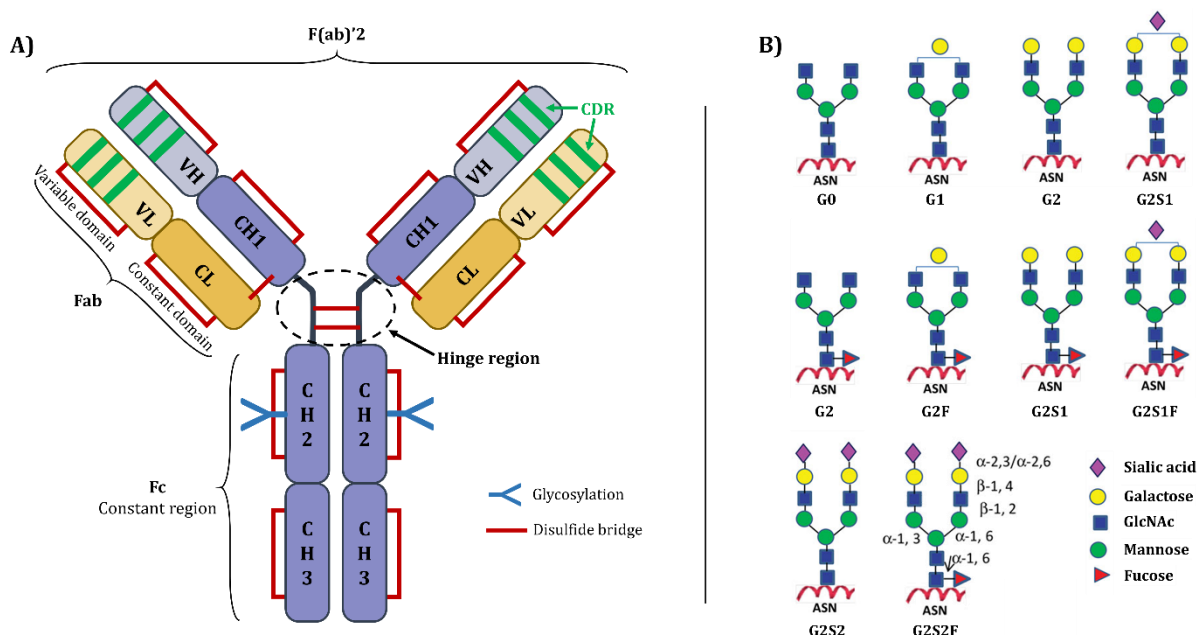


Figure 25: (A) General structure of a monoclonal antibody, exemplified for an IgG1. (B) Examples of the main glycans of IgG, attached on the asparagine residue 297 (adapted from ³⁷⁵).

The fragment crystallizable (Fc) consists of the remaining CH2 and CH3 domains from the Hc and is responsible for effector functions³⁷⁶. Interactions between mAbs and Fc receptors are directly affected by the presence of a conserved glycosylation site³⁷⁷ located on the CH2 domain (on the asparagine residue 297 for IgG1)³⁷⁸. Various complex and heterogeneous glycosylations can take place in mAbs structures. The core structure of glycans is usually composed of N-acetyl glucosamine (GlcNAc), mannose, galactose, fucose, or sialic acid moieties³⁷⁷ (**Figure 25.B**). These glycosylations are crucial for mAb structure and stability, and play a critical role in its biological activities, particularly for the Fc receptor-mediated effector functions³⁷⁹.

IgGs can be divided into four subclasses – IgG1, IgG2, IgG3, and IgG4 – demonstrating a high amino acid sequence similarity but differing from one another by the number and the position of the inter-chain linkages³⁷⁸. The high stability and the higher affinity to Fc receptors of IgG1 compared to other subclasses make it widely used in biotherapeutics³⁸⁰. Thereby, around 70% of approved mAbs belong to the IgG1 subclass, but some IgG2 and IgG4 formats have been developed³⁷².

B. Diversity of mAb-related therapeutic proteins

Numerous mAb-related biotherapeutics have been specifically developed for various applications³⁸¹, such as cancers^{382, 383}, cardiovascular disorders³⁸⁴, infectious and autoimmune diseases^{385, 386} or organ transplantation³⁸⁷.

i) Evolution of mAb-related products

The first mAb with murine origin was produced in 1975 by Köhler and Milstein³⁸⁸ using the hybridoma technology. Nevertheless, the side effects induced by the increased response of the human anti-mouse antibodies restricted the use of murine mAbs for therapeutic applications³⁸⁹. Hence, multiple sources have been used for the production of highly specific antibodies, which can be differentiated by the mAb suffix nomenclature (defined by the international nonproprietary names): murine (“-omab”), chimeric (“-ximab”), humanized (“-zumab”) and human (“-umab”)³⁸¹.

However, in several cases, resistance mechanisms have still been observed. For this reason, multiple companies and groups have gathered their efforts to develop new mAb-derived formats to increase the therapeutic efficiency of mAbs. It includes, but is not limited to, bispecific antibodies and their

derivative products³⁹⁰⁻³⁹³, truncated forms³⁹⁴ (e.g., BiTE, VHH, nanobodies, HSABody), Fc fusion proteins³⁹⁵ or antibody-drug conjugates³⁹⁶ (ADCs).

ii) Antibody-drug conjugates

Antibody-drug conjugates (ADCs) are a class of biotherapeutics designed for the targeted treatment of cancers³⁹⁷, and they can be qualified as “magic bullets” as envisioned by Paul Ehrlich³⁹⁸. They consist of a cytotoxic payload attached to a recombinant mAb by a synthetic linker (cleavable or not)³⁹⁹ (**Figure 26**), hence combining the high-specificity of the mAb and the efficient killing capabilities of the cytotoxic molecule.

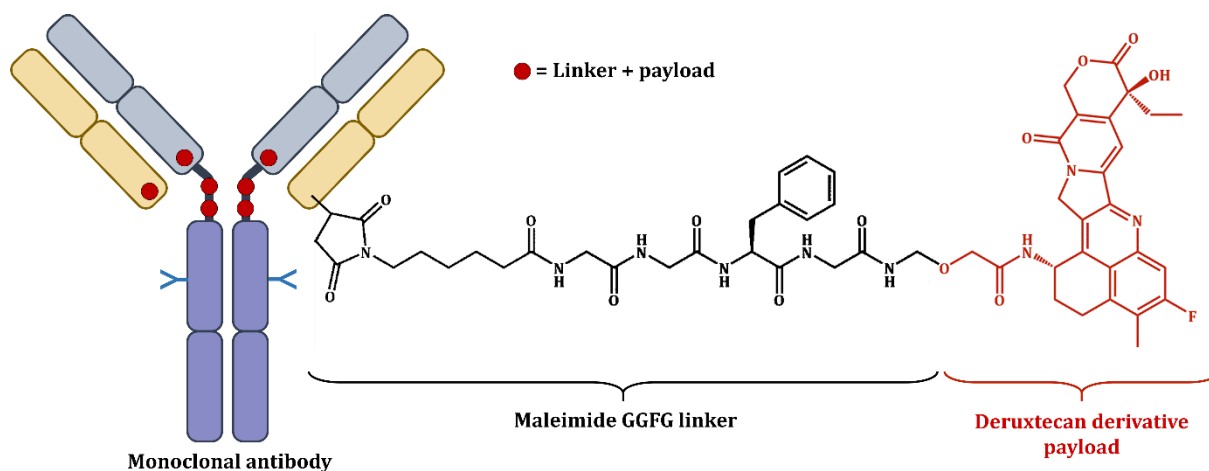


Figure 26: Structure of ADC through the example of trastuzumab deruxtecan. Here, trastuzumab deruxtecan is composed of the mAb (trastuzumab), a maleimide GGFG linker and a payload (deruxtecan derivative).

Briefly, the antibody, designed to target a specific tumor antigen, binds to the antigen on the surface of cancerous cells. After ADC internalization, the payload is released into the cytoplasm and interacts with its targets to engender tumor cell apoptosis⁴⁰⁰.

Multiple parameters must be carefully considered in the design of ADCs⁴⁰¹:

- The mAb should be specific to a targeted antigen over-expressed on the surface of the tumor cells with a high-binding affinity and demonstrate minimal immunogenic effects.
- The selection of the linker directly affects the safety (e.g., ADC aggregation) and the efficiency (e.g., premature release⁴⁰²). The linker can be cleavable or non-cleavable⁴⁰¹.
- The cytotoxic agent is responsible for the efficiency of the ADC. They are highly toxic for cells and usually increase the hydrophobicity of the therapeutic compared to the unconjugated counterpart.

Depending on the conjugation strategy (i.e., lysine conjugation, cysteine conjugation, or site-specific conjugation), the number of attached payloads, also referred to as drug-to-antibody ratio(DAR), differs³⁹⁹. In addition, the conjugation strategy also influences the distribution of the drugs (drug load distribution, DLD) and the average DAR (avDAR).

C. Characterization of mAbs and mAbs derivatives

Due to their inherent structural complexity and heterogeneity, comprehensive characterization of biotherapeutics is required along the manufacturing process. Moreover, many critical quality attributes (CQAs) must be monitored before the commercial release of a new product, as they may affect the product's stability, efficiency, purity, and immunogenicity. These CQAs are defined in the

guideline Q6B from the International Conference for Harmonization⁴⁰³ (ICH) and include various physicochemical properties, such as the presence of charge or size variants, glycosylation, disulfide pairing, or other PTMs (e.g., oxidation, deamidation, etc.). In the specific case of ADC, the DAR/avDAR and the DLD must also be provided.

Regarding the structural complexity of mAbs and ADCs, mainly due to their size, micro variability, and dynamics, several techniques envisaging the analysis of these proteins at different levels are required to provide a thorough characterization. In this context, mass spectrometry has primarily contributed to the characterization of these proteins thanks to its versatility, resolution sensitivity, and selectivity. It allows the specific designing of different approaches to provide complementary insights on their structure at different levels (e.g., quaternary structure, tertiary structure, amino acid sequence).

Generally, a multi-level approach is used (**Figure 27**) to obtain the required information^{404, 405}:

- Higher order structure analysis using HDX-MS, XL-MS, or nMS with IMS-MS is used for conformational characterization of the mAb, aggregates analysis, epitope mapping, and measurement of the dynamics of interaction^{406, 407}.
- Top-level analysis (intact mAb) investigates the intact mass, the DAR profile and the avDAR, the DLD, and the CCS by LC-MS, LC-nMS, and IMS-MS. It also provides information about the major modifications on the mAb.
- Middle-level analysis is performed on the mAb subunits that can be obtained by disulfide bridges reduction (Lc and Hc), IdeS cleavage (Fc/2 and F(ab)'2), papain digestion (Fab and Fc) or combining reduction and enzymatic cleavage (Fd, Lc and Fc/2). Subunit analysis by LC-MS is used for positional isomers determination or subunit identification and with their major modifications. Top-down and middle-down approaches enable amino acid sequence elucidation and modification of conjugation site location^{9, 408}.
- Bottom-level analysis of the mAb by LC-MS/MS after enzymatic digestion using peptide mapping that gives detailed information about the amino acid primary sequence, the localization of the modifications, the potential conjugation sites of the payloads for ADCs^{409, 410}, or and multi-attributes methods (MAM) that allows the monitoring of product quality attributes including PTMs identification and quantification^{411, 412}.

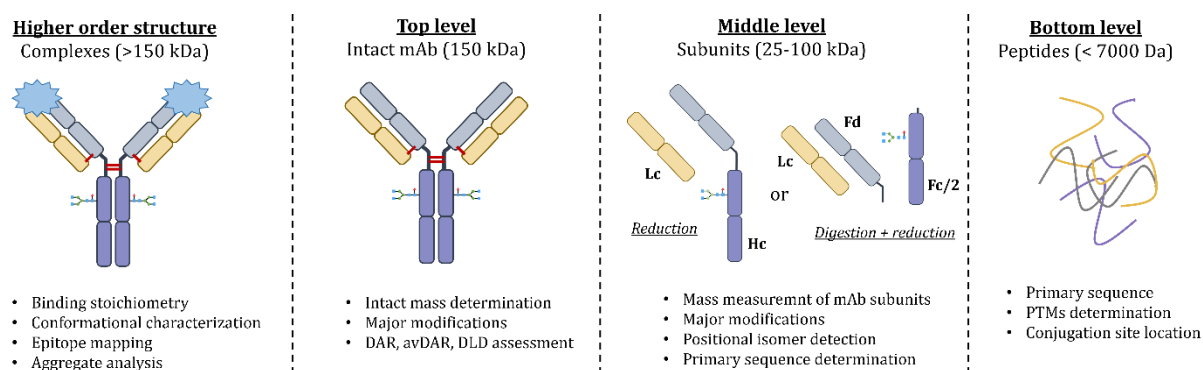


Figure 27: Description of the multi-level workflow for comprehensive characterization of mAb-based products (adapted from ³⁹⁶).

As the work presented in this manuscript mainly focuses on top and middle-level analysis, the contributions of the different MS-based approaches at these levels will be discussed in the following sections. On the one hand, experiments can be conducted without fragmentation (intact mass and middle-up) to obtain intact mass or subunits masses, information about the DAR, avDAR, and DLD, or identify major modifications (e.g., glycoforms). On the other hand, fragmentation of the intact mAb or

its subunits can be performed (i.e., top-down and middle-down) to characterize the protein's primary sequence and identify modification sites. These two strategies (intact mass/middle-up and TD/MD) are detailed below.

i) Intact mass measurement and middle-up experiments

Intact mass and middle-up experiments investigate the protein mass, the DAR profile and the avDAR, the DLD, and the CCS using LC-MS, LC-nMS, and IMS-MS. It can also provide information about the glycoforms^{396, 413, 414}. Analysis at the middle-up level is used to have additional information at the subunit level.

Native MS is widely used for this kind of experiment as it preserves the mAb quaternary and tertiary structures in the gas phase. It provides information on the oligomerization state of the mAb⁴¹⁴. Thanks to preserving the non-covalent interactions, nMS is also of utmost importance for ADC characterization, allowing the determination of the number of drugs conjugated⁴¹³. Indeed, for ADCs, covalent bonds between light and heavy chains can be reduced (e.g., Cys conjugated ADC⁴⁰¹) to accommodate the cytotoxic payloads, hampering the characterization of the intact structure of the ADC in denaturing conditions. Native MS is also beneficial for CQAs assessment – such as size variants, hydrophobic variants, charge variants, and modification identification – when combined with non-denaturing liquid chromatography separation. Non-denaturing LC methods are well established and routinely used in research and development or quality laboratories, particularly when combined with UV detection. Implementing of MS after the LC separation is more challenging as it requires using MS-compatible buffers (volatile salts)⁴¹⁵ and necessitates a thorough knowledge of MS analysis. They are based on the separation of the species considering their differing physicochemical properties, including hydrophobicity, size, affinity, or charge. An overview of the four main non-denaturing LC methods for mAbs and ADCs analysis is provided below:

- Size exclusion chromatography (SEC), where species are separated based on their differences in hydrodynamic volumes. It is the method of choice for size variants – high and low molecular weight species (HMWS and LMWS, respectively) – separation and detection⁴¹⁶.
- Hydrophobic interaction chromatography (HIC) relies on separating hydrophobic species through their differences in hydrophobicity using a decreased gradient of salts⁴¹⁷. This technique is the gold standard approach for DAR species separation in ADCs.
- Ion exchange chromatography (IEX) separates acidic and basic charge variants based on their difference in net surface charge⁴¹⁸. Elution is performed by applying a pH gradient or varying the ionic strength. IEX uses different resin compositions with negatively charged functional groups to elute cations, called cation exchange chromatography (CEX)^{419, 420}. Conversely, anion exchange chromatography (AEX) separates anions with resins bearing positively charged functional groups⁴²¹.
- Affinity chromatography is mainly employed for protein purification. Protein A and Protein G affinity chromatography are the two most used methods for mAbs because of the specific and reversible binding of mAb Fc regions to these proteins⁴²². They are eluted by decreasing the pH and/or increasing the ionic strength.

However, non-denaturing conditions are not always mandatory for mAbs and ADCs analysis. These molecules can also be characterized in denaturing conditions, using more classical approaches regarding LC and MS. Hence, denaturing LC is generally implemented to characterize mAbs at the top and middle levels. It is performed in acidic conditions using organic solvents that denature the mAb. These denaturing LC methods use MS-compatible mobile phases, simplifying their hyphenation to MS. At the top level, the two main denaturing LC methods employed are:

- Reverse phase liquid chromatography (rpLC) separates species based on their difference in hydrophobicity by increasing the proportion of organic solvent in the mobile phase, reducing the interactions between the species and the stationary phase. It is the method of choice for intact mass measurement, primary sequence determination, and PTM identification^{396, 414}.
- Hydrophilic interaction chromatography (HILIC), where the separation is performed according to the difference of hydrophilicity using an increasing water gradient in organic solvents, resulting in fewer interactions between the mAb species and the stationary phase (amide ligand). It is mainly employed for released mAb glycans analysis^{423, 424}.

Some strategies combining two chromatographic separation techniques (2D-LC) have also been introduced to take advantage of the orthogonal separation of the compounds based on different properties^{425, 426}. However, these strategies are well suited to characterize the structure of mAb-based formats at the intact or the subunit levels, but further insights are needed to localize the modifications observed within the LC dimension at the residue level. For this purpose, and similar to other protein analysis, fragmentation of the different populations have to be conducted to give access to this information. To this end, TD/MD-MS approaches have played an interesting role in characterizing therapeutic proteins.

ii) Top-down and middle-down analysis in the characterization of mAbs and ADCs

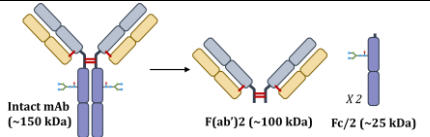
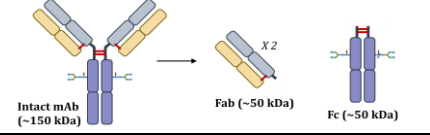
The method of choice for primary structure analysis is the bottom-up strategy, but it presents some limitations. Digestion of the intact mAb into peptides requires multiple sample preparation steps that can generate artefact²⁶. Information about the intact mass of the protein is not available, making it impossible to identify the entire proteoform. In addition, fragmentation techniques usually employed for peptide mapping experiments can fragment the payload conjugated on the ADC, hindering its identification and precise location by this approach⁴²⁷. On the contrary, TD/MD-MS first allows the intact measurement and then the fragmentation of the protein backbone, providing information about the whole protein and the amino acid sequence composition.

Primarily, TD-MS experiments have been scarcely used for intact mAb fragmentation. However, the fragmentation of the intact mAb does not allow extensive protein backbone dissociation due to its size (~150 kDa), limiting the sequence coverage observed. A few studies reported sequence coverage between 20% and 50% for the analysis of intact mAb, mainly using ETD or ECD fragmentation modes⁴²⁸⁻⁴³¹. Another critical point for the fragmentation of intact mAb is the presence of multiple inter-chain and intra-chain disulfide bonds, preventing efficient energy redistribution and dissociation but also being a challenge for data processing.

Then, several groups reduced the disulfide bonds, giving rise to two subunits: the light chain (Lc) of ~25 kDa and the heavy chain (Hc) of 50 kDa. Reducing the size of the molecule to fragment increased the sequence coverage, particularly for the Lc, with residue cleavages up to 80%^{432, 433}. Despite the benefits observed for the Lc, Hc fragmentation remains a pitfall with sequence coverages lower than 50%.

In this context, the MD strategy allows obtaining mAb subunits of lower MW (between 25 kDa and 100 kDa), favoring their analysis by MS. Subunits are obtained by using enzymes specifically designed for mAb-based proteins. They are highly specific, generally targeting a single cleavage site⁴³⁴. Regarding their structure and the specificity of the cleavage sites, denaturation and disulfide bond reduction of mAbs are not required prior to digestion. Most enzymes cleave the structure of the mAbs in the vicinity of the hinge region, either above or below, providing different subunits. Digestion reactions are mostly performed at 37°C during a short period (from 30 min to 2h hours, except for papain and pepsin)⁴³⁵. Some of the most widely used proteases, their cleavage sites, and the generated subunits are listed in **Table 4**.

Table 4: Most used enzymes for mAb-related product digestion.

Protease	Experimental conditions	Cleavage site	Subunits
Ides	30 min at 37°C, pH 6-8	Under hinge region	
IdeZ	120 min at 37°C, pH 6-8		
Pepsin	1-48 h at 37°C, pH 4-4.5		
KGP	60-120 min at 37°C, pH 8	Above hinge region	
SpeB	60-120 min at 37°C, pH 6.5-8		
Papain	6-12 h at 37°C, pH 5-8		

In addition to digestion, an optional additional step of reduction is possible. Indeed, subunits obtained using digestion only remain relatively large (~100 kDa for F(ab')₂) depending on the enzyme used and can still not be fragmented efficiently. Therefore, a reducing agent is also included during the sample preparation. Subunits are generally denatured using Guanidine-HCl combined with TCEP or DTT at 50-100 mM reducing agents, which are principally used for disulfide bond reduction⁴³⁶. In the specific case of antibodies, disulfide bond reduction allows the release of subunits covalently bonded by inter-chain disulfide bridges^{434, 435}. After digestion and reduction, three subunits of ~25 kDa are obtained, i.e., Lc, Fd, and Fc/2.

MD-MS is the most employed method to fragment comprehensively mAb-based proteins. Fragmentation of mAb subunits ~25 kDa has demonstrated very promising, and plenty of studies describe this approach in the literature. Various fragmentation techniques have been used to characterize the subunits. Pipes *et al.* used CID to fragment an IgG1 molecule and identified site-specific methionine oxidation⁴³⁷. Another study by Cotham and Brodbelt used 193 nm UVPD fragmentation to identify unambiguously glycosylation site location and to obtain 60% sequence coverage on a single LC-MS/MS run⁴³⁸

The combination of multiple fragmentation techniques – enabled by the introduction of cutting-edge MS platforms – is now regularly used to maximize residue cleavages. In 2020, an interlaboratory study involving 14 laboratories demonstrated the complementarity of the fragmentation techniques to achieve extensive sequence coverage⁴³⁹. Fornelli and coworkers also applied ETD, EThcD, and UVPD in different LC-MS/MS runs, reaching sequence coverage of 90% on average¹⁰.

Regarding the results achieved for mAbs, the combination of fragmentation has been applied to ADCs, enabling high sequence coverage and precise drug conjugation site location^{9, 265, 440, 441}. For instance, Hernandez Alba *et al.* evidence the payloads conjugation sites on cysteines of an avDAR 4 ADC using ETD, HCD, and UVPD⁹.

MD-MS also found applications for investigating or confirming disulfide bonds^{428, 430, 431, 442} in mAbs, and PTM identification and localization⁴⁴³⁻⁴⁴⁶. For instance, Zhu and coworkers recently effectively characterized trastuzumab and adalimumab combining MD-MS and intact mass⁴⁴⁶. They confirmed the glycosylation site on the Asn61 of the Fc/2 fragment and pointed out the presence of C-terminal lysine-modified variants.

Despite the multiple activation techniques available, achieving complete sequencing of mAb subunits is intricate. Several strategies already introduced in the literature for other types of proteins can be used to increase the sequence coverage. First, PTCLR appears as a promising approach, allowing simplifying fragmentation spectra. Indeed, mAb subunits are 25 kDa molecules, leading to very complex spectra, particularly after ETD, EThcD, or UVPD. Hence, PTCLR can distribute overlapping fragments over a broader mass range. It already demonstrated some benefits for proteins of comparable size or larger^{305, 318}, but has never been applied to therapeutic proteins. Second, internal

fragments consideration already demonstrated significant improvements in the sequence coverage^{332, 335, 442}. A recent study by Wei *et al.* shows the potential of internal fragments for sequence coverage increase (up to 75% for a TD-MS experiment), and the ability to determine disulfide bonds and conjugation sites on ADCs³³⁶.

Some examples in the literature highlight the advantages of TD/MD-MS for therapeutic protein analysis. In this thesis work, MD-MS strategies have been optimized, including cutting-edge fragmentation techniques, combined with ion/ion reactions and internal fragment ion search to improve the sequence coverage and the localization of cytotoxic payloads of recent ADCs formats.

2. The concern of host cell proteins in monoclonal antibodies

A. Monoclonal antibodies manufacturing process

i) Expression system

The introduction of hybridoma technology for the *in vitro* production of antibodies by Köhler and Milstein³⁸⁸, for which they were awarded the Nobel Prize in Physiology and Medicine in 1984, has led to the widespread use of mAbs for diagnostic and therapeutic applications. The hybridoma technology relies on cell fusion to produce monoclonal antibodies specific to an antigen of interest. An immune response specific to an antigen of interest is induced by introducing this antigen into a mouse. The resulting B-lymphocytes are isolated and fused with myeloma cells (immortal cancerous cell lines), creating a hybrid cell called hybridoma³⁸⁸ that takes advantage of the ability to secrete a specific antibody from B-lymphocytes and the immortality from myeloma cells. Nevertheless, the low amounts of mAb produced, their short half-life in humans, and the side effects induced by the increased response of the human anti-mouse antibodies restricted the use of these murine mAbs for therapeutic applications^{389, 447, 448}. Indeed, these parameters are vital for the efficiency of the mAb. Through the years, chimeric⁴⁴⁹, humanized⁴⁵⁰, and more recently, human⁴⁵¹ mAbs have been introduced following the advances in molecular biology and genetic modification techniques. They present a reduced immunogenicity in patients and an extended half-life in humans. To improve their production yield, mAbs are produced in diverse expression systems, including bacteria, yeast, fungi, insects, mammalian cell lines, and transgenic plants or animals⁴⁵². Mammalian host cells are the most common expression systems for the commercial production of therapeutic mAbs due to their ability to introduce post-translational modifications consistent with those seen in humans⁴⁵³. Today, the majority of the recombinant mAbs are manufactured in Chinese hamster ovary (CHO) cell lines for many reasons: (i) CHO cells are robust and can grow in suspension culture and using serum-free chemically defined media, suitable for large-scale culture in bioreactors; (ii) robust gene amplification systems that increase their productivity exist for CHO cells; (iii) the CHO expression system is now well known and has been demonstrated to be a safe host, which could facilitate the approval from the regulatory agencies^{454, 455}.

ii) Manufacturing process

The mAb manufacturing process follows two steps (**Figure 28**): the upstream process (USP), consisting of the cell culture in bioreactors, and the downstream process (DSP), where mAbs are purified.

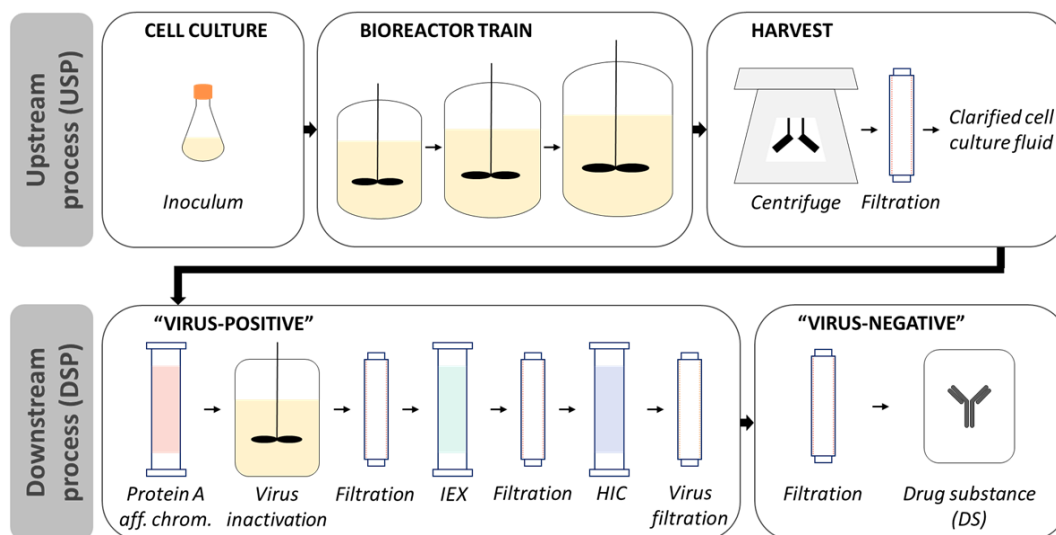


Figure 28: Monoclonal Antibodies manufacturing process (adapted from ⁴⁵⁶). HIC: Hydrophobic Interaction Chromatography. IEX: Ion Exchange Chromatography. Protein A aff. chrom.: Protein A affinity chromatography.

a) Upstream process

Before mAb manufacturing, a suitable and robust cell clone with high stability and high yield is selected, then expanded and stored in hundreds of vials at -180°C in different locations. It constitutes a cell stock called the Master Cell Bank (MCB) that will be used only if necessary. The MCB is comprehensively characterized by stability, purity, and identity. Then, cells originating from a vial of MCB will be grown for multiple passages and again aliquoted in hundreds of vials stored at -180°C . These vials constitute the Working Cell Bank (WCB) and are used for mAb production.

The USP consists of several steps. First, a vial of cells from the WCB is thawed, and cells are grown in a small volume (typically 50 mL) of cell culture and then expanded to larger volumes corresponding to bioreactor capacities (5,000 L to 20,000 L). After the expansion phase, the mAb is secreted by the mammalian cells during the production phase. It accumulates in the culture medium to reach typical titers of around 1 g/l in batch and 1-10 g/l in fed-batch processes⁴⁵⁷. Finally, the cell culture fluid (CCF) is submitted to centrifugation and filtration by microfiltration to remove cells and cellular debris⁴⁵⁸, leading to clarified cell culture fluid. The microfiltration step is the last step of the USP.

b) Downstream process

The DSP is composed of a succession of purification steps, including chromatography, filtration, and viral elimination, to obtain a pure and concentrated mAb solution^{459, 460} by eliminating process-related impurities such as nucleic acids, lipids, host cell proteins (HCP) or other product-related impurities⁴⁶¹.

The first purification step is usually performed using Protein A affinity chromatography and is named the capture step. It allows for removing most impurities from the raw harvest materials in a single step⁴⁶². Protein A was initially found in the cell wall of *Staphylococcus aureus*, and the first description of its high affinity to the Fc region of IgG from multiple species was made in 1958⁴⁶³. Through the years, various modified versions of protein A have been introduced and they demonstrate increased stability and binding capacities^{464, 465}.

After Protein A purification, the mAbs solution is subjected to three chromatographic steps (or polishing steps), along with viral clearance (inactivation and filtration), product concentration using ultrafiltration, and buffer components removal using diafiltration^{459, 466}. Finally, the mAb is

concentrated during the last filtration step into a buffer suitable to its formulation (i.e., addition of an excipient) and conditioning (e.g., lyophilization).

To allow the commercialization of the mAb product, the level of impurities, particularly DNA and HCPs, must be monitored throughout the whole process⁴⁰³ to reach purity goals, typically <100 ng of HCP/mg of mAb for HCPS and < 10 ng/dose for DNA⁴⁶¹.

B. Host cell proteins monitoring

Host cell proteins (HCPs) are bioprocess-related impurities; their removal is critical in mAb manufacturing. These HCPs may degrade the antibody, potentially leading to reduced stability and efficiency⁴⁶⁷⁻⁴⁷⁰, or endanger patients' safety by causing a damaging immune response weakening the therapeutic protein efficiency^{468, 471, 472}. Hence, they are considered critical quality attributes (CQAs) by the regulatory agencies. Their amount should be minimized to trace levels and carefully monitored by highly sensitive methods^{403, 473, 474}. Despite general guidelines not indicating a strict limit on the HCP level in drug products (DPs), it is classically considered that their global amount should not exceed 100 ppm (100 ng of HCP/mg of mAb)⁴⁷⁵⁻⁴⁷⁷. More than from the global HCP quantity, the immunogenicity or stability issues are more likely to originate from only a few specific HCPs, even at low levels^{470, 478, 479}. Additionally to these issues, variations of the HCP profile can be impacted by manufacturing changes in the USP (e.g., cell culture duration, culture temperature, feeding strategies)⁴⁸⁰, production scale-up⁴⁶⁰ or by the HCPs co-elution with the mAb during the DSP^{475, 481}.

All these parameters underline the need for careful monitoring of HCPs along the manufacturing process to assess the effects of bioprocessing developments. To this end, analytical techniques that cover the wide dynamic range between the mAb and the HCP impurities (up to 5-6 orders of magnitude) are required⁴⁸² (**Figure 29**). Several analytical methods are available and can be grouped as immune-specific (ELISA and Western-Blot) and non-specific (electrophoresis and mass spectrometry) approaches^{477, 483, 484}.

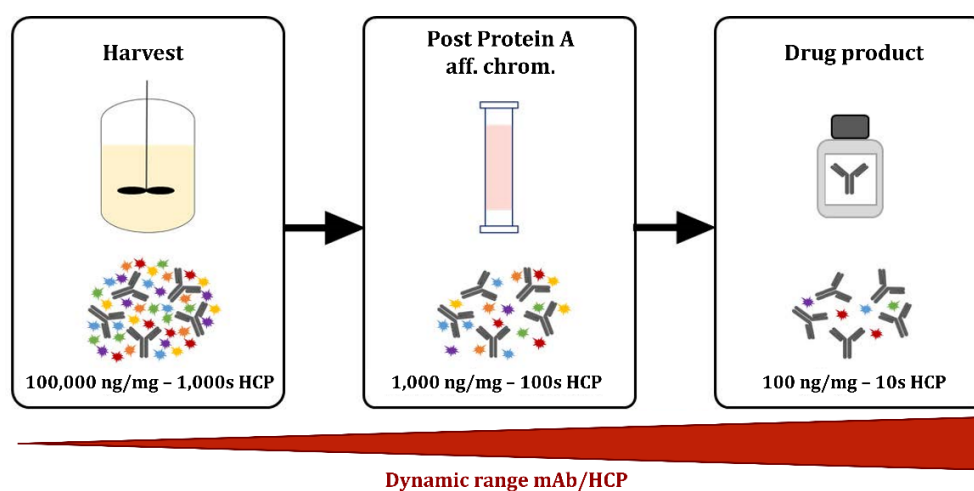


Figure 29: Host Cell proteins clearance during purification process (adapted from ^{456, 482}). Order of magnitude of the overall impurities quantity (ng of HCP per mg of mAb) and number of HCPs at different step of the mAb manufacturing process. Protein A aff. Chrom.: Protein A affinity chromatography.

i) Immuno-specific approaches

Immuno-specific methods use polyclonal anti-HCP antibodies to detect HCPs. They are usually produced in rabbits or goats after repeated injections of the HCP mixture of interest. The selection of the HCP mixture is critical, as it will determine the range of HCPs that the polyclonal antibodies will detect. The anti-HCP antibodies are usually generated from a null cell line, i.e., a mock-transfected cell line, to prevent the production of anti-mAb polyclonal antibodies. With this approach, it is assumed

that the null cell line and the mAb-producing version of this cell line have similar HCP content^{477, 483, 484}.

Wide varieties of commercial anti-HCP antibodies, usually produced from multiple null cell strains, have been designed in response to the strong interest in immuno-specific methods. Assays employing these commercial anti-HCP antibodies to detect a broad spectrum of HCPs from various cell strains and bioprocessing are called generic assays. While quick and easy to implement, they exhibit low specificity in purified products and crude samples^{484, 485}. Process-specific assays can be raised in-house using specific material from the cell line used (USP specific) or the manufacturing process (DSP specific) to increase the assay's specificity. However, specific assays are still limited to the material used for their production, and the detection of new HCPs introduced by changes in the manufacturing process will not be possible.

Lastly, neither generic nor specific assays can detect the entire spectrum of HCPs resulting from a specific manufacturing process. They are restricted only to the HCPs, inducing an immune response in the organism that produces the anti-HCP antibodies. In addition, the development of process-specific assays is time-consuming and costly⁴⁸⁶, and anti-HCPs antibodies are a limited reagent that will need to be reproduced.

a) ELISA

Enzyme-linked immunosorbent assay (ELISA) is the gold standard method to monitor HCPs along the manufacturing process, from the mAb production to the final product formulation^{477, 487}. ELISA is an easy-to-use, high-throughput method presenting a high sensitivity (ng of HCP/mg of mAb) and a high specificity widely used in biopharmaceutical companies for HCP monitoring.

Among the several types of ELISAs developed, sandwich ELISA – where two antibodies are used to target and detect antigens in a sandwich manner – is the most commonly used method (**Figure 30**).

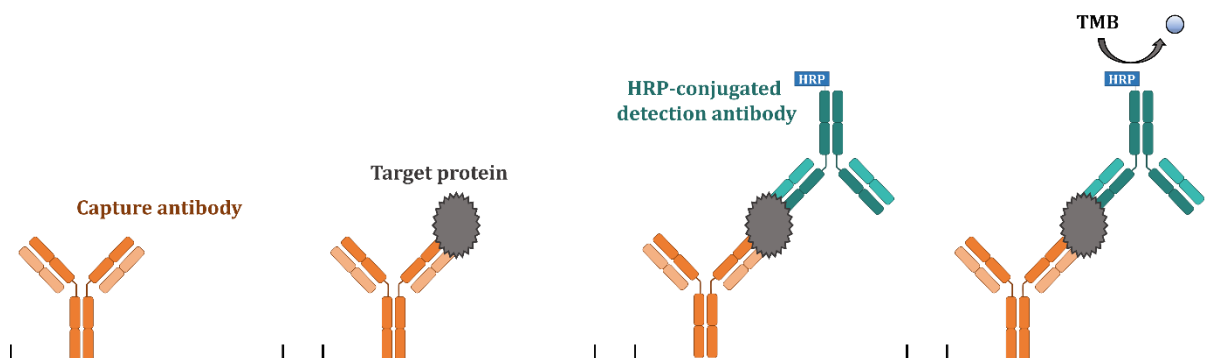


Figure 30: Principle of the sandwich ELISA.

First, capture antibodies specific to the targeted protein are coated on a multiwell plate, and the unbound antibodies are removed after incubation. The remaining unbound sites are blocked by incubation with bovine serum albumin (BSA) or non-fat dry milk, for example, in order to avoid subsequent non-specific binding onto the wells. After washing, the samples are added to the wells, and the target protein binds to the immobilized capture antibodies. Again, a washing step is performed to remove the unbound target proteins. Then, detection antibodies conjugated to an enzyme (horseradish peroxidase (HRP), for example) will bind another specific epitope of the protein of interest. Combining these two antibodies allows the simultaneous capture and detection of the protein of interest. Antibodies are incubated, and unbound materials are washed away. Finally, the substrate (e.g., TMB) is added and converted by the enzyme, allowing the quantification of the product by absorbance measurement using a spectrophotometer. In the end, a standard calibration from standard samples of known concentration is used to estimate the antigen concentration.

Sandwich ELISA aims at quantifying a large number of HCPs and generally uses the same anti-HCP antibodies for capture and detection. On the one hand, capture antibodies are directly coated to the well. Alternatively, they can be conjugated to biotin to improve their binding onto plates coated with streptavidin for an increased assay sensitivity⁴⁸⁸. On the other hand, detection antibodies are conjugated to an enzyme, usually HRP^{489, 490}. Both capture and detection antibodies must bind simultaneously to the HCP to be detected. However, steric hindrance or lack of and limited number of accessible binding epitopes will result in the absence of detection of these HCPs⁴⁸⁶.

ELISA for the detection of HCPs suffers from two main limitations. First, there is a lack of coverage for low or non-immunogenic species that the anti-HCP mixture cannot identify. Second, ELISA only gives a global estimation of the total HCP amount but cannot provide either HCP identification or individual quantification^{468, 483}.

b) Western blot

Another immune-specific method available for HCP detection is the Western blot. This approach is routinely employed in biomedical research to detect specific proteins in a complex sample^{491, 492}. It consists of three steps: separation by size using gel-electrophoresis, transfer of the protein onto a membrane (western blot step), and detection of the target protein by specific antibodies (**Figure 31**).

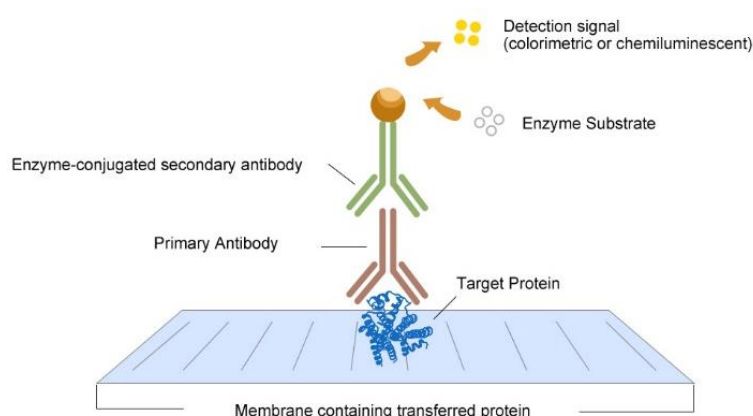


Figure 31: Principle of protein detection by Western blot (from www.elabscience.com).

In a few words, the membrane is incubated with a blocking solution (e.g., BSA or non-fat dry milk) to avoid nonspecific antibody binding, washed and incubated with specific primary antibodies designed to target the protein of interest. After washing out the unbound primary antibodies, secondary antibodies conjugated to an enzyme specifically target the primary antibodies. The membrane is washed and incubated. Then, the substrate is added, and the enzyme will produce a compound that will be detected using a spectrophotometer.

Western blot, like ELISA, aims to detect many HCPs. Nevertheless, it is generally used as a supplement supporting the development of ELISA by assessing the coverage of the anti-HCP antibodies during the animal immunization process after western blot results comparison to the global HCP profile obtained from 2D gels. Finally, as the western blot relies on the same principle as ELISA, it suffers from the same limitations, providing an incomplete HCP coverage.

ii) Non-specific approaches

Non-specific methods are orthogonal and complementary to immuno-specific methods as they allow the detection of both immunogenic and non-immunogenic proteins comprehensive HCP monitoring.

a) Gel electrophoresis

Polyacrylamide gel electrophoresis (PAGE) is based on separating proteins according to their size. For 2D-PAGE experiments, an additional separation step using isoelectric focusing (IEF) is added to differentiate the proteins based on their isoelectric point in a polyacrylamide gel strip containing an immobilized pH gradient⁴⁹³. After migration, proteins are fixed into the gel. They can be stained using global dyes such as Coomassie blue⁴⁹⁴ or silver staining⁴⁹⁵ to allow global protein profiling.

Most of the biopharmaceutical companies use 2D-PAGE and differential gel electrophoresis (DIGE) to monitor HCPs during the manufacturing process, as they are robust and provide visualization of the HCPs profile along two dimensions, i.e., their molecular weight and isoelectric point^{496, 497}. Moreover, these techniques allow the visualization of PTMs⁴⁹⁸, such as phosphorylation. Nevertheless, identification of HCPs with these techniques is not possible, and they need to be combined with mass spectrometry to provide information^{493, 498}.

The limited dynamic range is the major drawback of these approaches, as only the most abundant proteins will be displayed. It can be particularly problematic for identifying HCPs in the presence of overwhelming mAb heavy and light chains that can hide the presence of low abundant HCPs (**Figure 32**). In addition, analyzing proteins exhibiting small or huge molecular weights, extreme basicity or acidity, or high hydrophobicity is generally tricky^{493, 499}. The limited dynamic range is a significant issue.

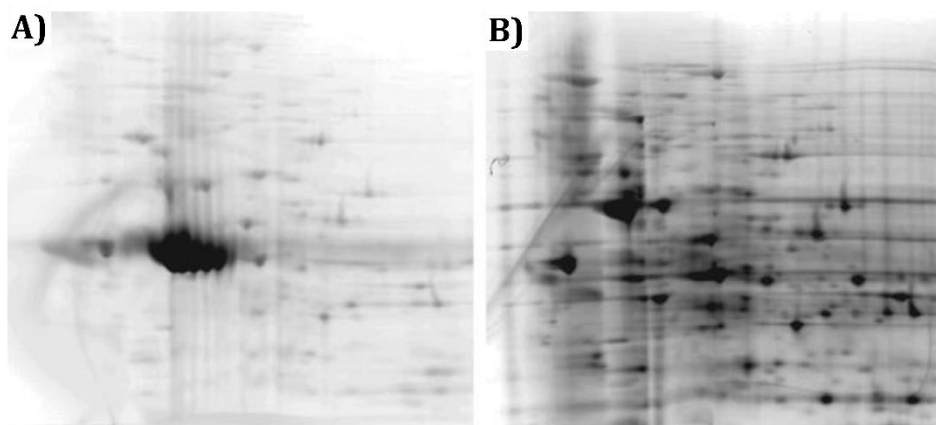


Figure 32: Limitation of the 2D-PAGE method for the detection of HCP (from ⁴⁹⁷). 2D gel of a CCCF fraction (A) containing the mAb product and (B) without the mAb product that was removed by affinity chromatography.

Other limitations of 2D-PAGE include low throughput, labor-intensiveness, and lack of reproducibility^{484, 493}.

b) Mass spectrometry

Following the technical advances in instrumentation, mass spectrometry (MS) has emerged as a promising alternative to the previously mentioned methods. MS allows the individual identification and unbiased quantification of HCPs, mainly due to the improvements in MS2-based quantification using DIA and targeted methods. Indeed, these strategies allowed a 2-8-fold gain in sensitivity⁶ and a significant increase in specificity and dynamic range compared to MS1-based quantification methods. Hence, MS enables a more comprehensive risk assessment by considering the nature of the HCP (e.g., proteolytic activity or immunogenicity)^{477, 500}.

The targeted approaches combining isotope dilution and SRM or MRM have been the method of choice for MS-based quantification of HCPs, presenting the highest sensitivity, accuracy, and robustness^{54, 478}.

However, designing targeted methods is time-consuming, and its application is limited to only a few hundred candidates. Alternatively, DIA allows the collection of MS/MS spectra for all detectable species, enabling the complete proteome map to be obtained. Moreover, DIA reaches similar sensitivity, quantification accuracy, and robustness as targeted methods^{6, 7, 501}, and is now commonly used for HCP analysis⁵⁰²⁻⁵⁰⁵.

Numerous studies used LC-MS/MS methods for various purposes, including validation of HCP coverage of ELISA assays⁵⁰⁶, HCP risk assessment^{472, 507}, or support of process development^{476, 508}. Despite the advances in MS2 signal extraction and processing, the wide dynamic range between the HCPs and the mAb remains a big challenge for accurately quantifying low-abundant proteins. To tackle this challenge, several strategies have been proposed, such as mAb depletion using native digestion⁵⁰⁹, HCPs enrichment⁵¹⁰⁻⁵¹², during the sample preparation step or multidimensional chromatography⁵¹³⁻⁵¹⁶ and implementation of ion mobility^{517, 518} for the LC-MS/MS step. Moreover, some of these techniques allow the quantification of HCPs down to the sub-ppm level⁵¹⁹⁻⁵²¹.

Other limitations to the development of HCP analysis by MS include the cost and the need for trained and skilled people to run mass spectrometers. Moreover, the lack of publicly available high-quality protein sequence databases for CHO cell lines is another pitfall⁵²², as the major part of the commercially available mAbs are produced from CHO.

iii) Comparison of HCP monitoring approaches

Today, ELISA is the method of choice used by biopharmaceutical companies for HCPs monitoring along the mAb bioprocessing and for product purity evaluation. It is a highly sensitive and specific method, allowing high-throughput analysis from crude harvest to drug product. Moreover, it is accepted by the regulatory agencies for HCP monitoring. While ELISA gives the global HCP quantity, it cannot provide individual identification and quantification of the HCP present in the sample. It is a significant limitation as the immunogenicity or the risk of product degradation is known to be linked to the presence of specific HCPs and not to the global amount of HCPs. Consequently, the need to develop and use alternative orthogonal techniques is crucial. The available methods previously described in this chapter for the monitoring of HCPs are summarized in **Table 5**.

Table 5: Summary of the available methods for HCP monitoring (from ^{475, 484, 508}).

Method		Sensitivity	Advantages	Limitations
Immuno-specific	ELISA	Total HCP 1-100 ppm	High-throughput, sensitivity, sensibility	Detects only immunogenic HCPs, total HCP amount only, costly and time consuming development
	Western blot	Individual HCP 20-200 ppm	MW and pI separation, PTM detection	Detects only immunogenic HCPs, labor-intensive, costly and time consuming development
Non-specific	2D-PAGE	Individual HCP 8 ppm	MW and pI separation, PTM detection, MS compatible	Low dynamic range, interferences from the mAb product
	Mass spectrometry	Individual HCP 0.1-500 ppm	Individual HCP identification and quantification	Lack of high quality CHO databases, expensive equipment and high skilled operators

Therefore, the emergence of LC-MS/MS-based strategies for HCP monitoring arises as a promising alternative. They provide individual identification and quantification of HCPs present in the sample, enabling them to perform risk assessment. In this bibliographic introduction, all the capabilities of quantitative proteomics have been shown, each having strengths and weaknesses in the context of

HCP characterization⁵⁰⁸. The most appropriate strategy should be selected depending on the step of the manufacturing process and the objective of the analysis (e.g., precise quantification of a known HCP in the drug product, overall identification of HCPs).

Another fact underlying the growing interest in MS-based methods for HCP monitoring is the current revision of the US Pharmacopeia chapter related to HCP monitoring⁴⁷³. Indeed, in its precedent version, this document almost exclusively focused on ELISA and immuno-specific methods, while MS approaches were only discussed to a limited extent. This new chapter aims to define general guidance and best practices for identifying and quantifying HCPs. As the US Pharmacopeia is one of the most comprehensive sources of guidelines for the biopharmaceutical industry, even if not a regulatory authority, this new chapter would reinforce the role of mass spectrometry for HCP characterization.

In this manuscript, quantitative LC-MS/MS-based methods were developed and optimized for the quantification of HCPs in drug products.

PART I
METHODOLOGICAL DEVELOPMENTS IN
QUANTITATIVE PROTEOMICS

Chapter 1: Optimization of LC-MS/MS acquisition methods for quantitative proteomics

1. Context of the project

Since the beginning of proteomics, instrumental innovations have been one of the cornerstones of progress in protein identification and quantification. Numerous improvements are continuously being made in liquid chromatography and mass spectrometry by the manufacturers. These technological advances allow today's LC-MS/MS systems to operate with high speed and resolution while providing high sensitivity. These systems have become an essential tool for characterizing and quantifying proteins in complex biological mixtures with limited sample fractionation and reduced chromatographic gradient time. Critical parameters related to the analysis of the peptide mixture must be evaluated and optimized to take advantage of these advances to increase proteome coverage.

In the first step, we evaluated key parameters related to the acquisition on a last-generation nanoLC-tribrid instrument (Eclipse from Thermo Fisher Scientific):

- The analyzer used for MS2 spectra acquisition (ion trap or Orbitrap) in order to find a best compromise between acquisition speed and resolution.
- The fragmentation technique used to fragment the peptides, which must be adapted to the type of samples and the goal of the analysis.
- For data-independent acquisition (DIA), the isolation window scheme by optimizing the number of precursors per window while keeping a reasonable cycle time to keep the quantification precision.

All optimizations were performed using a total proteome extracted from a human cancer line, HeLa cells. Then, optimized data-dependent (DDA) and independent acquisition (DIA) methods were applied to a set of well calibrated reference samples composed of a mixture of 48 human proteins (Universal Protein Standard, UPS1, Merck, Darmstadt, Germany) spiked into a total lysate of *Arabidopsis thaliana*.

In a second step, we optimized the compensation voltages on the high-field asymmetric ion mobility spectrometry (FAIMS) device in front of the mass spectrometer to ascertain the influence of ion mobility on the detection and quantification of peptides and proteins.

In a third step, we appraised collision energies and different strategies based on either MS2 or MS3 to quantify a HeLa Pierce™ (Thermo Fisher Scientific) digest labeled with TMT and TMTpro™ (Thermo Fisher Scientific).

2. Analytical strategy applied for method optimizations

The overall analytical scheme is shown in **Figure 33**.

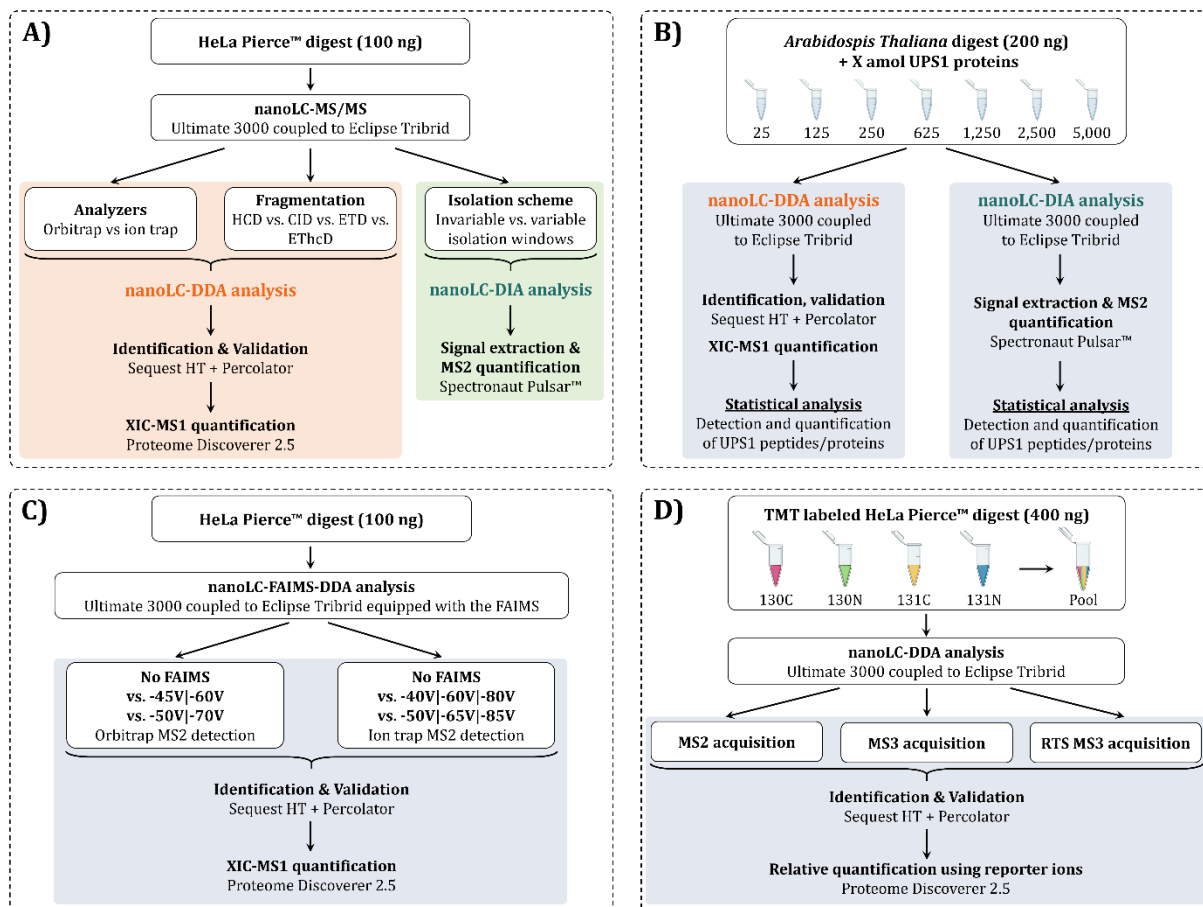


Figure 33: Analytical scheme used for nanoLC-MS/MS acquisition methods optimizations for quantitative proteomics. (A) HeLa Pierce™ digest was analyzed in triplicates on a nanoLC-Orbitrap Eclipse Tribrid coupling in DDA mode (orange) for the evaluation of the analyzer for MS2 detection and the fragmentation technique, and in DIA mode (green) for the comparison of different isolation schemes. The number of peptides and proteins identified as well as their quantification were used for the evaluation. Identification, validation and quantification results were obtained from Sequest and Percolator, both included in the Proteome Discoverer 2.5 software for DDA data, and from Spectronaut Pulsar™ using the library-free DirectDIA™ approach for DIA data. (B) A range of perfectly calibrated samples composed of UPS1 proteins spiked in different quantities in an *Arabidopsis thaliana* digest were analyzed on a nanoLC-Orbitrap Eclipse Tribrid coupling in both DDA and DIA modes using optimized methods. Identification, validation and quantification results were obtained from Sequest and Percolator, both included in the Proteome Discoverer 2.5 software for DDA data, and from Spectronaut Pulsar™ using the library-free DirectDIA™ approach for DIA data. (C) HeLa Pierce™ digest was analyzed in triplicates on a nanoLC-Orbitrap Eclipse Tribrid coupling equipped with FAIMS in DDA mode for the comparison of compensation voltages combinations. The number of peptides and proteins identified as well as their quantification were used for the evaluation. Identification, validation and quantification results were obtained from Sequest and Percolator, both included in the Proteome Discoverer 2.5 software. (D) HeLa Pierce™ digest labeled with 4 channels of TMTpro™ 16plex were analyzed on a nanoLC-Orbitrap Eclipse Tribrid coupling in DDA mode using multiple methods for precursor ions' quantification at MS2 or MS3 level. Identification, validation and quantification results were obtained from Sequest and Percolator, both included in the Proteome Discoverer 2.5 software.

First, when evaluating MS acquisition parameters of the tribrid instrument, HeLa Pierce™ cell digests were used (**Figure 33.A**). For DDA, the detection of MS2 spectra using the ion trap working in Rapid mode and the Orbitrap at a resolution of 30K (at 200 m/z) analyzers were compared. Most of the fragmentation techniques available on the instrument, i.e., HCD, CID, ETD, and ETHcD, were also tested,

with MS2 spectra detection performed in the ion trap. Samples were analyzed in triplicates; peptide and protein identifications were done using the Sequest search engine, and validation using the Percolator algorithm, both implemented in Proteome Discoverer 2.5 software, by applying a 1% FDR at the PSM, peptide, and protein levels. XIC-MS1 quantification was performed by summing the peptides' intensities for each protein.

For DIA, samples were also analyzed in triplicates by nanoLC-MS/MS on the Orbitrap Eclipse Tribrid instrument, following various isolation schemes detailed later in this chapter. DIA data were processed using the default parameters of the Spectronaut Pulsar™ software (Biognosys, Schlieren, Switzerland) working with the DirectDIA™ library-free mode. Then, a perfectly calibrated set of samples consisting of a mixture of 48 human UPS1 proteins (Merck) spiked into a total *Arabidopsis thaliana* lysate was used to evaluate the previously optimized DDA and DIA acquisition parameters (**Figure 33.B**).

Second, the HeLa Pierce™ cell digest was used to determine the optimal compensation voltages to apply when the FAIMS device is set upfront the mass spectrometer (**Figure 33.C**). Samples were analyzed in triplicates by nanoLC-FAIMS-MS/MS on the Orbitrap Eclipse Tribrid instrument, with peptide and protein identifications performed using the Sequest search engine and validation using the Percolator algorithm, both implemented in the Proteome Discoverer 2.5 software, by applying a 1% FDR at the PSM, peptide, and protein levels. XIC-MS1 quantification was performed by summing the peptide intensities of each peptide.

Third, the aforementioned HeLa Pierce™ cell digest was labeled with TMT reagent to evaluate the label-based quantification approach on a tribrid instrument (**Figure 33.D**). Several collision energies used for peptide fragmentation and acquisition methods allowing reporter ion quantification at either MS2 or MS3 level were tested and are detailed in this chapter. Samples were analyzed in triplicates by nanoLC-FAIMS-MS/MS on the Orbitrap Eclipse Tribrid instrument, and peptide and protein identifications were made using the Sequest search engine and validation using the Percolator algorithm, both implemented in the Proteome Discoverer software, by applying a 1% FDR at the PSM, peptide, and protein levels. Quantification was performed using the intensities of the reporter ions of each protein.

Additional details on sample preparation, LC-MS/MS methods and data processing are listed in the **Experimental Section**.

3. Evaluation of the acquisition parameters on a tribrid instrument

The evaluation of the acquisition parameters was initiated during the installation of the Orbitrap Eclipse Tribrid instrument from Thermo Fisher Scientific in the lab. We first started optimizing the acquisition parameters for shotgun bottom-up proteomics analysis of complex samples. Indeed, tribrid instruments offer a multitude of options regarding the acquisition method design, with two analyzers used for signal acquisition, i.e., an Orbitrap and an ion trap, and the possibility to perform peptide fragmentation with multiple techniques. In addition, the chromatographic setup (liquid chromatography, trap column, and column, emitter) coupled to the mass spectrometer was also completely new. This study aimed at determining, evaluating and optimizing the critical parameters of DDA and DIA acquisition methods to ensure the best peptide and protein identification and quantification results with excellent accuracy and reproducibility are obtained.

A. Comparison of Orbitrap and ion trap analyzers for data dependent acquisition

Scan rate and acquisition speed are significant concerns in quantitative proteomics, as they directly influence the number of points per chromatographic peak conditioning the quantification precision. An increase in resolution will allow better differentiation of ions with close m/z ratios, but the acquisition time required to record the spectra will rise accordingly, affecting the quantification

precision. With tribrid instruments that include both a linear ion trap (scan rate of 45 Hz) and an Orbitrap (scan rate of 40 Hz), we can reduce the cycle time by performing the MS/MS analysis in the ion trap at a lower resolution while recording the MS spectra, used for the XIC-MS1 quantification, at high resolution in the Orbitrap. Also, because of the lower resolution of the ion trap MS/MS, the fragment match tolerance parameter needs to be adapted during data treatment. While the recommended value for the Orbitrap is 0.02 Da, it can be extended to 0.5 Da when the ion trap is used.

Both analyzers were evaluated using a 100 ng HeLa Pierce™ (Thermo Fisher Scientific) digest injected in triplicates along a 60 min gradient and analyzed using a Top20 DDA method using HCD for fragmentation. Considering its faster acquisition speed, ion trap analyses contained around 30% more MS/MS spectra than those done with the Orbitrap (97,000 MS/MS versus 73,000 MS/MS). After a database search, ion trap data lead to the identification of 3,753 proteins and 28,069 peptides, representing 450 (+ 14%) additional proteins (**Figure 34.A**) and 1,000 peptides (+4%) compared to the Orbitrap data (**Figure 34.B**). Similarly, we quantified the same proportion of extra proteins (+13%) and peptides (+4%) with the ion trap compared to the Orbitrap, and both analyzers allowed to quantify more than 98% and 95% of the identified proteins and peptides, respectively. To evaluate the precision of quantification, we kept only the features quantified in all three replicates with a coefficient of variation (CV) lower than 20%. Consequently, the number of quantified proteins decreased by 35% for the Orbitrap detection and 30% for the ion trap detection, with an average of 2,122 and 2,253 proteins, respectively.

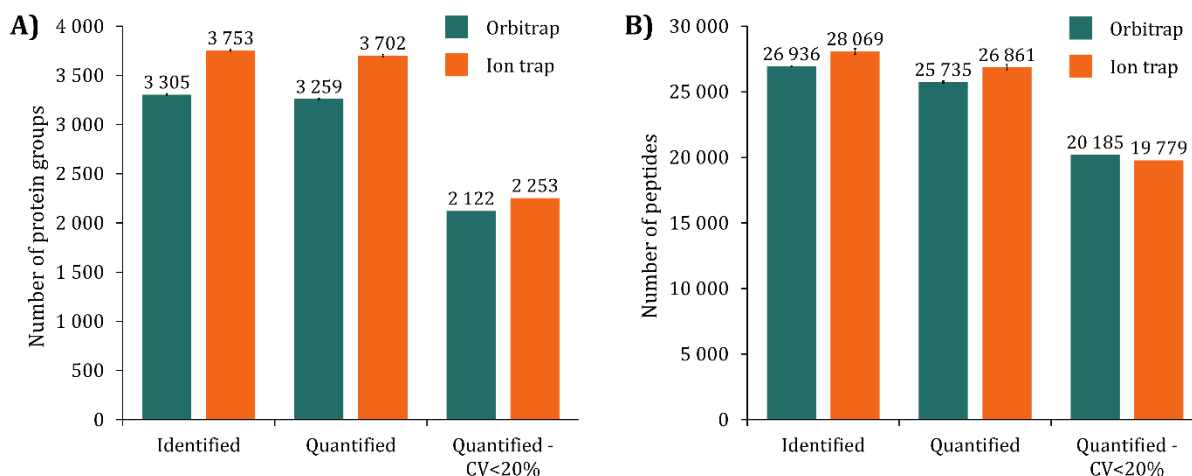


Figure 34: Results obtained with the Orbitrap and ion trap analyzers. (A) Number of proteins identified, quantified and quantified with a CV < 20% using Orbitrap (green) and ion trap (orange) MS2 detection. (B) Number of peptides identified, quantified and quantified with a CV < 20% using Orbitrap (green) and ion trap (orange) MS2 detection.

As mentioned before, the faster scan rate of the ion trap allows for acquiring more spectra, resulting in a median number of 22 points per chromatographic peak against 20 with the Orbitrap. Even so, quantification repeatability is close, with median CVs of 10.3% and 12.3% for the Orbitrap and the ion trap, respectively (**Figure 35**). Although a few more points per peak are available using the ion trap, the number obtained using the Orbitrap is already high enough to ensure repeatable quantification with a Top20 method. Notwithstanding, the ion trap's fast acquisition speed can increase the number of MS/MS spectra acquired during each cycle (e.g., Top50 method) or be useful for short gradients analyses giving narrower elution peaks.

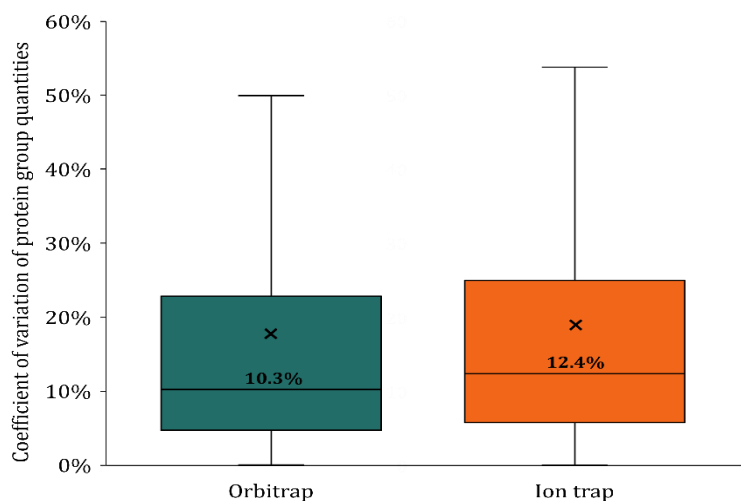


Figure 35: Boxplot representing the distribution of the coefficient of variation calculated from the average abundance of the proteins quantified in all the replicates using Orbitrap (green) and ion trap (orange) MS2 detection. Values displayed correspond to the median CV and crosses to the mean of the CVs. For clarity purpose, outliers were removed.

To summarize, even if the ion trap is slightly faster, the results achieved with the Orbitrap are very close in terms of coverage and acquiring high resolution data opens many more perspectives, especially in terms of data interpretation options. High resolution is for instance needed to use innovative AI-based prediction algorithms.

B. Comparison of HCD, CID, ETD and EThcD fragmentation techniques

After evaluating the influence of the analyzer on the number of identified and quantified features, performances of the various fragmentation techniques available on the Orbitrap Eclipse Tribrid instrument were compared. We decided to appraise techniques based on collision-based (CID, HCD) and electron-based (ETD, EThcD) dissociation. Ultraviolet photodissociation (UVPD), also available on the instrument, was not included in the study because of the high activation time needed and its limited use for shotgun proteomics. While HCD, allowed by a collision cell, is the most used fragmentation technique on Q-Orbitrap instruments, CID remains widely used on Q-ToF instruments⁵²³. In addition, electron-based fragmentation techniques are also used for specific applications, including the analysis of labile PTMs^{524, 525} (phosphorylation⁵²⁶, N- and O-glycosylation^{527, 528}) using either ETD or EThcD.

We evaluated the four fragmentation techniques on a 100 ng HeLa Pierce™ (Thermo Fisher Scientific) digest injected in triplicates. MS/MS detection was done in low resolution in the ion trap for all fragmentation modes, and the type of fragment ions generated by each technique was considered for the database search (see **Section Materials and Methods**). As expected, collision-based dissociation techniques (i.e., CID and HCD) give a higher number of identifications, with 22,120 peptides from 3,342 protein groups for HCD and 20,019 peptides from 2,959 protein groups for CID (**Figure 36**). In opposition, only 1,057 and 1,405 protein groups (5,077 and 8,828 peptides) were identified with ETD and EThcD, respectively. We observed the same trend for quantification, with around 2.5 times more proteins quantified in HCD/CID compared to ETD and 2 times compared to EThcD.

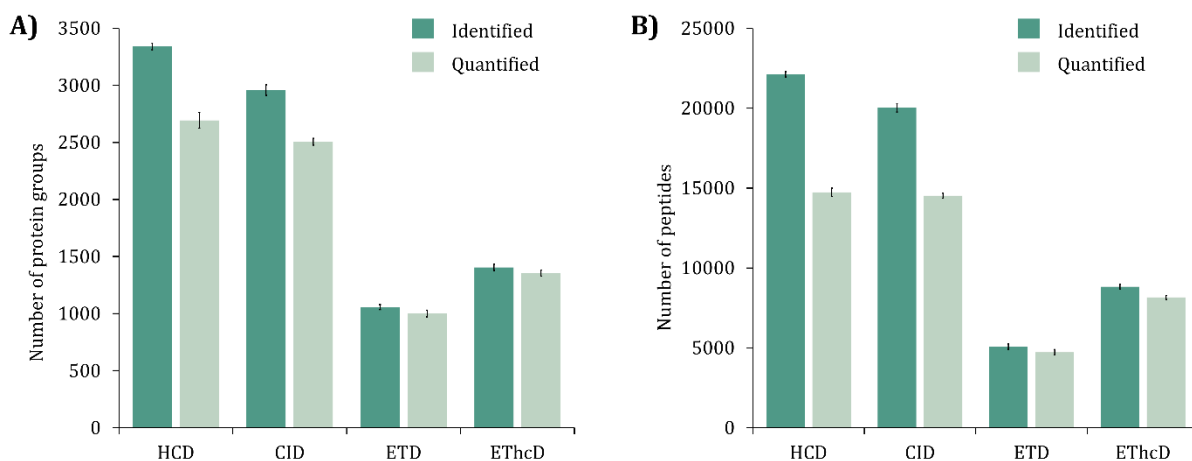


Figure 36: Results obtained from HCD, CID, ETD and EThcD fragmentation techniques. (A) Number of proteins identified and quantified. (B) Number of peptides identified and quantified.

While these results point out the limits of electron-based fragmentation for identifying proteins and peptides in “classical” samples (i.e., without labile PTMs), some valuable observations can be highlighted. With this intention, we investigated the precursor charge states for each fragmentation technique (**Figure 37**). It shows that 50% of the precursors identified with ETD were 3+ ions, while this proportion goes down to 25% to 29% with other techniques. Similarly, the proportion of 4+ precursors identified is 4 times higher with ETD than with HCD/CID. This pattern had already been observed previously in the literature^{285, 529}. Indeed, the precursor charge directly influences the peptide dissociation for ETD. Firstly, a higher charge state will lead to a better dissociation of the peptide, generating more intense fragments, while a lower charge state will poorly dissociate the peptide. Secondly, a non-dissociative process can occur when the charge state is too low (e.g., 2+), leading to an absence of fragmentation (ETnoD)³¹¹. Multiple studies have proposed applying supplemental collision energy after ETD to counterbalance the lack of fragmentation observed on 2+ precursors due to potential non-covalent interactions. As already demonstrated in the literature⁵³⁰, we can observe that 2+ precursors are more efficiently fragmented using EThcD (**Figure 37, right**). Regardless, we can underline that using extra energy does not allow us to obtain the same number of identifications as the collision-induced-based techniques (**Figure 36**). In this case, considering the ETD/EThcD reaction time (50 ms), a lower number of MS/MS spectra can be recorded (around 25,000 for ETD/EThcD against more than 85,000 for HCD/CID for a run of 90 min), consequently leading to a poor number of identifications.

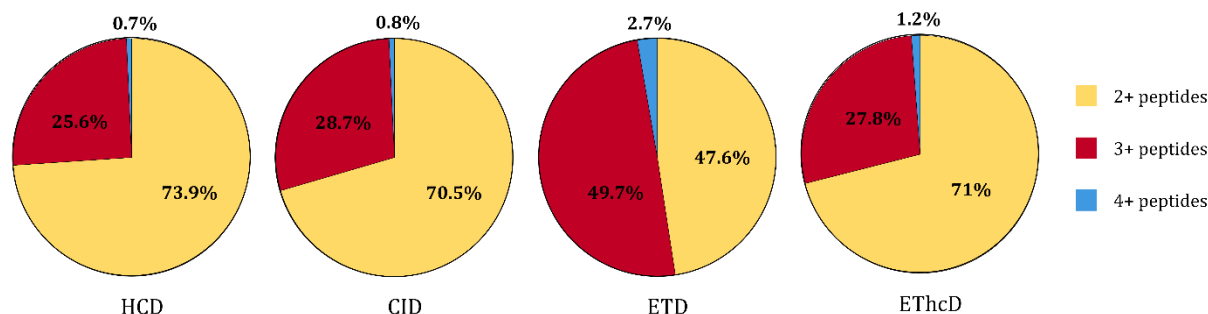


Figure 37: Analysis of the identified peptides with respect to their charge state distribution.

To summarize, we underline that collision-based dissociations, namely CID and HCD, are the most suitable for tryptic peptide fragmentation as they lead to more identifications. Because of the long reaction time required and the low fragmentation of doubly charged precursors, electron-based dissociation is not amenable for analyzing “conventional” bottom-up samples. However, its ability to

preserve traditionally labile PTMs and to cleave randomly the precursor backbone bonds make ETD appropriate for the analysis of modified peptides, such as phosphopeptides, that are generally highly charged. To ensure good coverage of the sample proteome while maximizing the fragmentation and the identification of modified peptides, methods based on decision trees can also be set up, with an online selection of the fragmentation method depending on the charge and/or the m/z ratio of the peptide^{526, 530}.

C. Optimization of the isolation windows scheme for DIA

The development of DIA methods requires some parameters to be finely tuned to optimize the performances of detection and quantification⁵³¹⁻⁵³³. First, DIA needs resolving power to distinguish fragments from co-isolated precursors. For this reason, the ion trap cannot be used to perform the MS/MS acquisition. Then, the number of points per peak is a critical point. To obtain a reliable quantification in DIA, it is admitted that at least 6 to 8 points per chromatographic peak are needed. Hence, the cycle time is key to ensuring precise quantification. Several factors influence the cycle time, such as the transition time (directly linked to the resolution) and the number of isolation windows, as described in the following equation:

$$\text{Cycle time} = \text{Transient time}_{MS1} + (\text{Number of isolation windows} \times \text{Transient time}_{MS2})$$

Multiple evaluations of the acquisition parameters, like the MS2 resolution and the AGC target, had already been carried out in the lab previously. Here, we focused on the isolation windows scheme, which includes the number of isolation windows affecting the cycle time and the size of these windows, defining the number of precursors simultaneously fragmented.

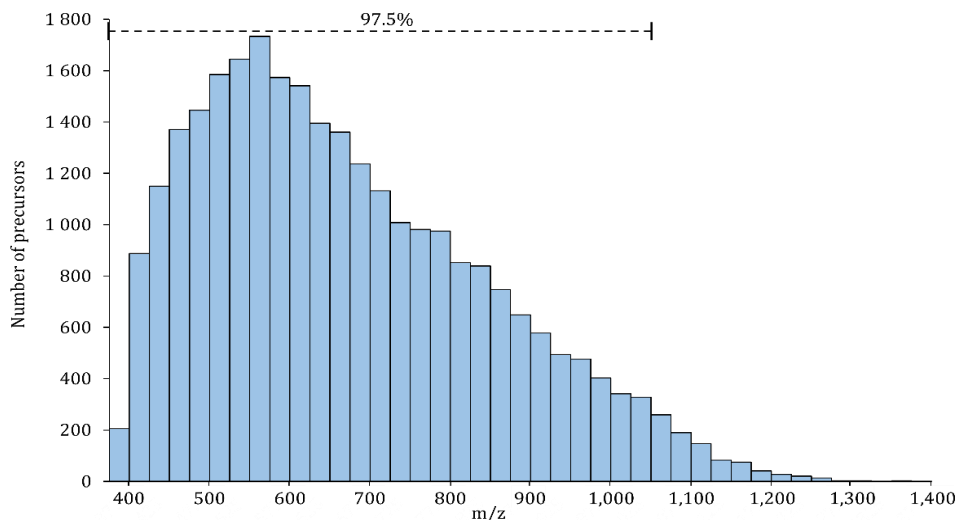


Figure 38: Distribution of the precursors identified from the analysis of a HeLa Pierce™ digest in nanoLC-DDA as a function of mass to charge ratio (m/z).

The isolation windows scheme was evaluated on a 100 ng HeLa Pierce™ (Thermo Fisher Scientific) digest. First, a DDA analysis was run on the same sample with the same gradient length. It allows appraising the repartition of the precursors among the m/z range and then defining accordingly the mass range to be used for DIA. Based on DDA results, we could notice that 97.5% of the precursors identified were between 375 and 1,050 m/z (**Figure 38**). Then, we decided to use these values as the mass range for DIA acquisition. Extending the mass range to 1,400 m/z will significantly increase the cycle time while only a few additional precursors could be detected. In this sense, three isolation schemes with fixed windows were evaluated (**Figure 39.A**): 30 windows of 22.5 Da, 45 windows of 15 Da, and 60 windows of 11.3 Da. Two isolation schemes with variable window widths were also tested

(**Figure 39.B**). The first one with 4 DIA methods composed of 45 windows from 10.5 to 40 Da, and the second one with 6 DIA methods composed of 45 windows from 9 to 41 Da. The goal of using variable windows is to limit the number of co-isolated and co-fragmented precursors in the same window by adapting its size as a function of the precursor density along the mass range (**Figure 38**).

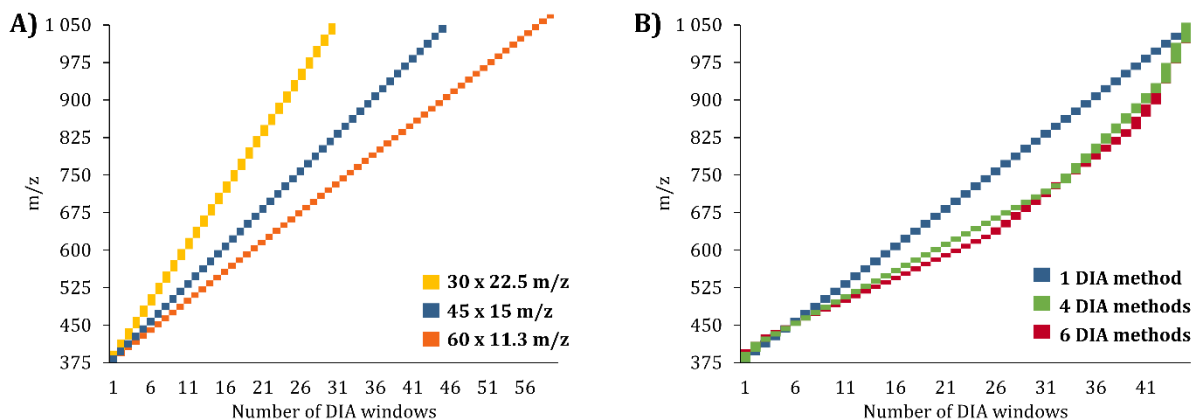


Figure 39: DIA isolation schemes. (A) Fixed isolation windows schemes: 30 x 22.5 m/z (yellow), 45 x 15 m/z (blue), and 60 x 11.3 m/z (orange). (B) Variable isolation windows schemes: 1 isolation width (blue), 4 isolation widths (green) and 6 isolation widths (red).

For the fixed isolation windows, the width reduction increases peptide and protein identifications. (**Figure 40**) However, it causes the number of points per chromatographic peak to decrease due to the expansion of the cycle time (**Figure 41, left**). Whereas we quantified similar proportions of precursors with a CV <20% for the three schemes, isolation windows of 11.3 Da allow us to quantify up to 15% more (**Figure 41, right**).

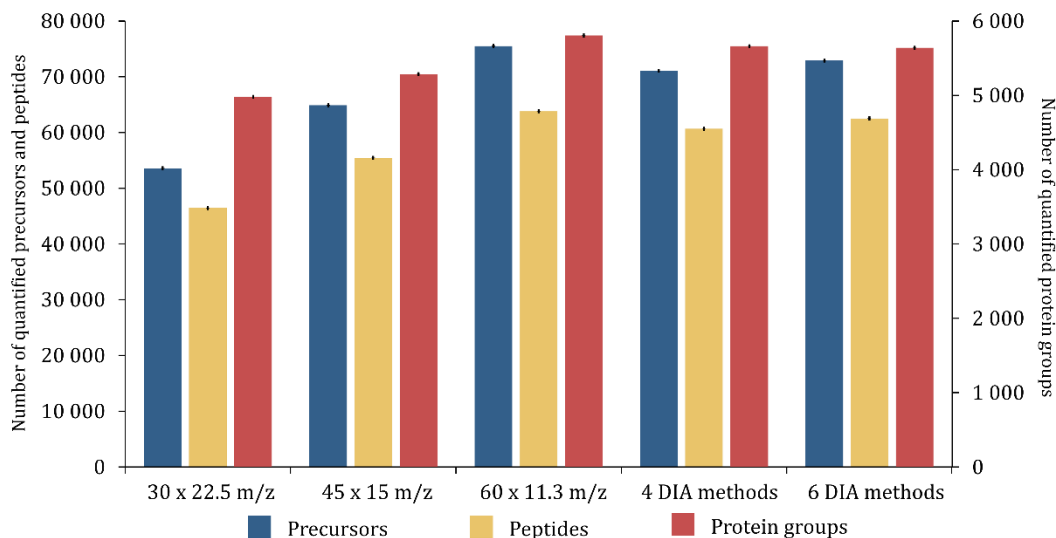


Figure 40: Results from HeLa Pierce™ DIA acquisition. Number of precursors (blue), peptides (yellow) and protein groups (red) quantified according to the different isolation schemes. Bar heights represent the number of quantified features and error bars the standard deviation.

In addition to this, variable windows provide peptide and protein identifications close to what is obtained by the 11.3 Da fixed windows (**Figure 40**) while maintaining the cycle time and, thus, the number of points per peak to reasonable values (**Figure 41**). Even so, no apparent gain in quantification reproducibility can be observed, with similar proportions of precursors quantified with a CV <20%.

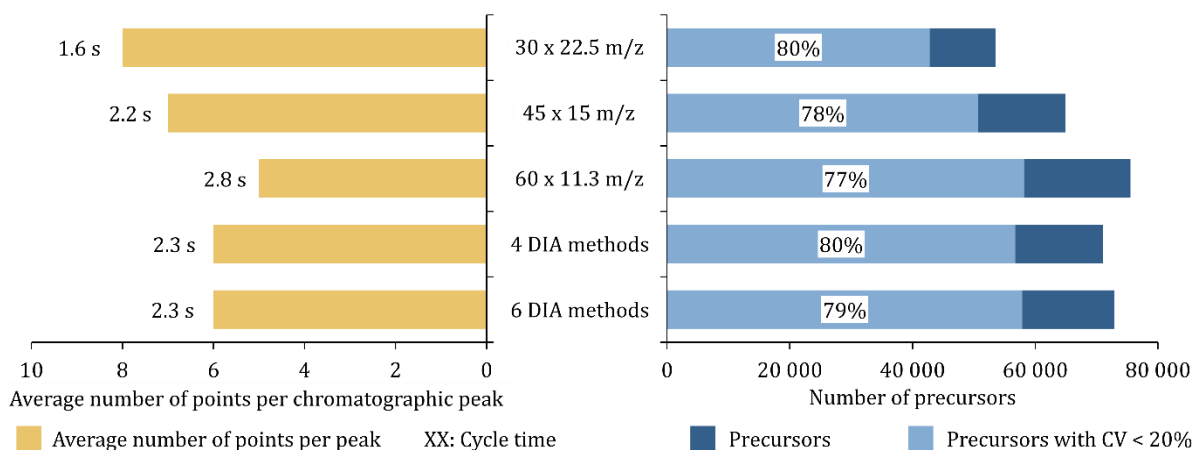


Figure 41: Results obtained according to the different isolation schemes in DIA for a HeLa Pierce™ digest. Average number of points per chromatographic peak and the associated cycle time are illustrated for each condition (left) and proportion of quantified precursors with a CV < 20% (light blue, right)

Given these points, we show that using small windows increases the number of identifications by restricting competing precursors, thus limiting interference and improving sensitivity⁵³⁴. However, we obtain this gain at the expense of the number of points per chromatographic peak impacting the reproducibility of the signal. Using variable isolation windows allows counterbalancing the low number of points per peak while identifying a comparable number of peptides and proteins to the 11.3 Da isolation windows. It demonstrates the importance of optimizing the isolation scheme to get the best results from our samples.

D. Application of the optimized methods to a mixture of well calibrated samples

In quantitative proteomics, the development of new methods, and especially their validation, is generally performed with reference samples. In this context, model proteomes are commonly used as standard samples. Here, we used *Arabidopsis thaliana* as a reference proteome because of its availability and complexity, allowing us to mimic real complex biological samples. Hence, we spiked a mixture of 48 UPS1 human proteins in known quantities to the reference matrix to obtain a range of perfectly calibrated samples. The objective was to test the previously optimized quantitative proteomics methods, i.e., DDA and DIA, and to evaluate their performance in terms of the numbers of quantified features, quantification precision and accuracy.

The strength of global proteomics approaches lies in their ability to identify, in a single analysis, several thousands of proteins in complex biological mixtures. However, limitations regarding reproducibility and quantification precision have already been reported¹³⁴. The recent advancements in chromatography, mass spectrometry, and bioinformatics have considerably improved their performances, particularly since the development of DIA. It appears very promising, providing deep proteome coverage and robust and accurate quantification^{535, 536}.

First, the number of quantified proteins in all replicates with a CV < 20% from the background proteome, i.e., *Arabidopsis thaliana*, is represented in **Figure 42.A**. Globally, DIA shows comparable results to those obtained with DDA; low variability is observed between the runs for both approaches (**Figure 42.B**).

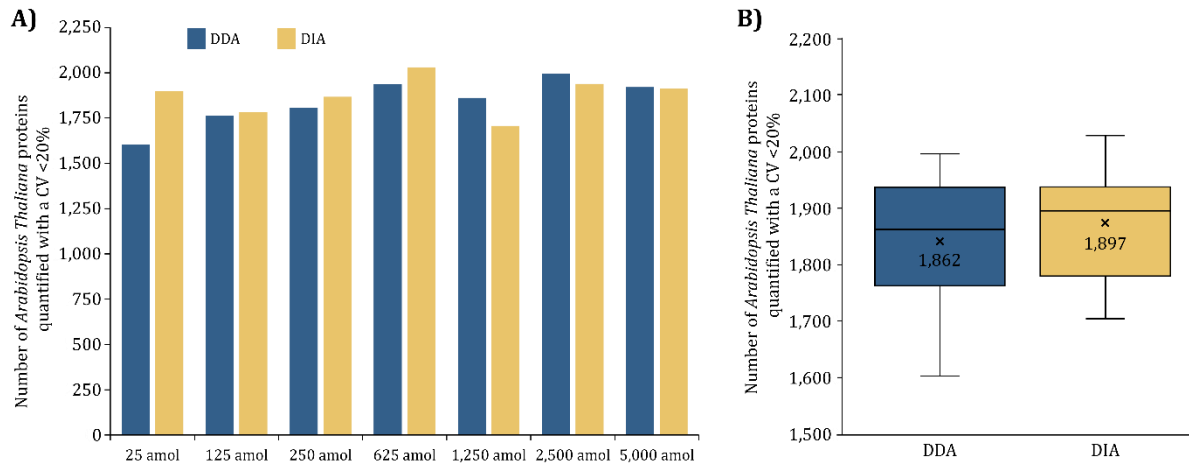


Figure 42: Results obtained for *Arabidopsis thaliana* proteins quantified in all the replicates with a CV < 20%. (A) Number of *Arabidopsis thaliana* proteins quantified in all replicates with a CV < 20% for each condition using DDA (blue) and DIA (yellow). (B) Boxplot representing the distribution of the number of *Arabidopsis thaliana* proteins quantified in all replicates with a CV < 20%. Number above the boxplot represent the median and crosses the mean numbers of proteins.

The number of quantified UPS1 proteins at each point of the range is represented in **Figure 43**. DIA demonstrates better results regarding the proteins quantified in all replicates for the low quantities, i.e., 25 and 125 amol, than DDA. Indeed, 43 out of the 48 UPS1 proteins are identified at 25 amol for DIA against 32 for DDA. Additionally, we quantified all UPS1 proteins in all replicates from 250 amol in DIA against 1,250 amol in DDA. The trends of slightly better results in DIA is maintained when filtering with proteins quantified in the three replicates with a CV < 20%.

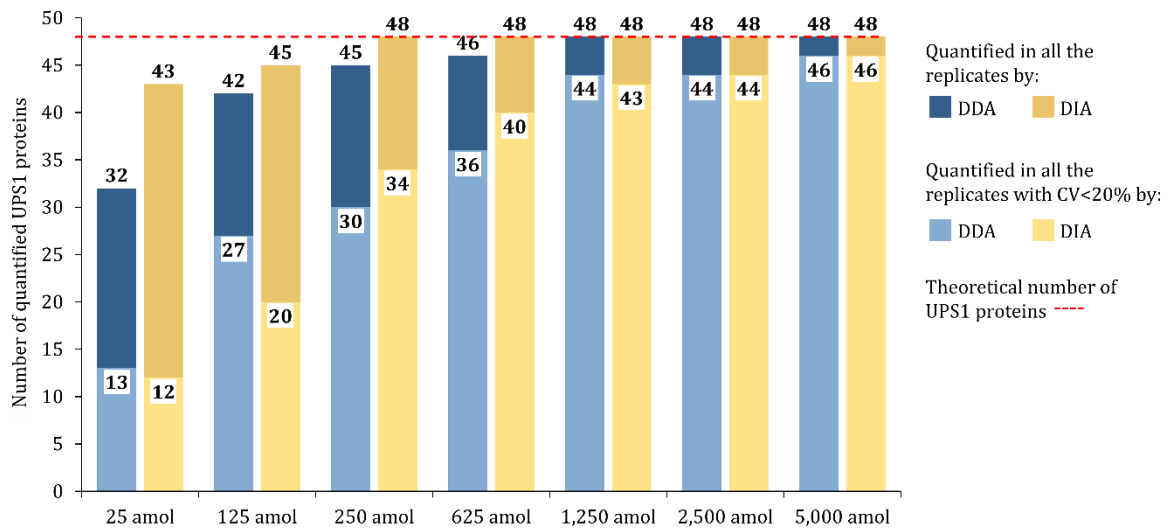


Figure 43: Results obtained from UPS1 protein quantification. The number of UPS1 proteins quantified in all the replicates is represented in blue for DDA and yellow for DIA. The number of UPS1 proteins quantified in all the replicates with a CV < 20% is represented in hatched blue for DDA and hatched yellow for DIA. The theoretical number of UPS1 proteins (48) is represented by the red dashed line.

Second, we evaluated the quantification accuracy of UPS1 peptides for DDA and DIA (**Figure 44**). Both approaches present a dispersion of the experimental ratios around the theoretical values, which increases with the theoretical ratio, revealing the limits of quantification accuracy for global approaches. Nonetheless, median experimental ratios are close to the theoretical values, except for the highest ratio (5,000 vs. 25 amol).

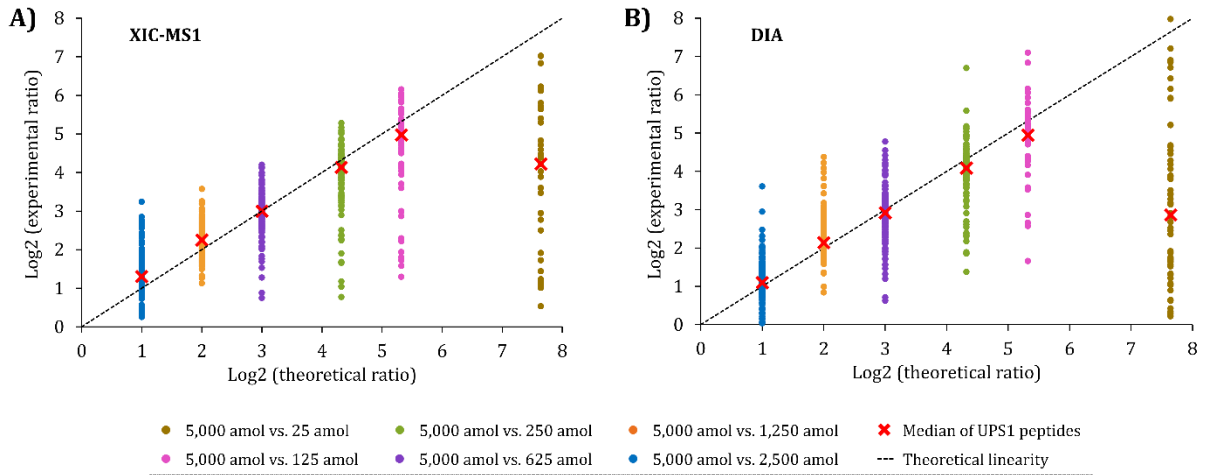


Figure 44: Evaluation of the precision of quantification of XIC-MS1 and DIA as global quantification approaches. Experimental ratios, calculated from the highest point of the range, were represented as a function of the theoretical ratios of UPS1 peptides for XIC-MS1 (A) and DIA (B) approaches.

In addition, volcano plots represented in **Figure 45** attest for the differential expression of UPS1 proteins. We observed a good accuracy for a \log_2 (theoretical ratio) up to 5.3 for DIA and 4.3 for DDA. Even if proteins cluster together, the experimental fold change of the 5,000 vs. 2,500 amol ratio is not aligned with the theoretical fold change, undoubtedly because of the slight difference between both concentrations.

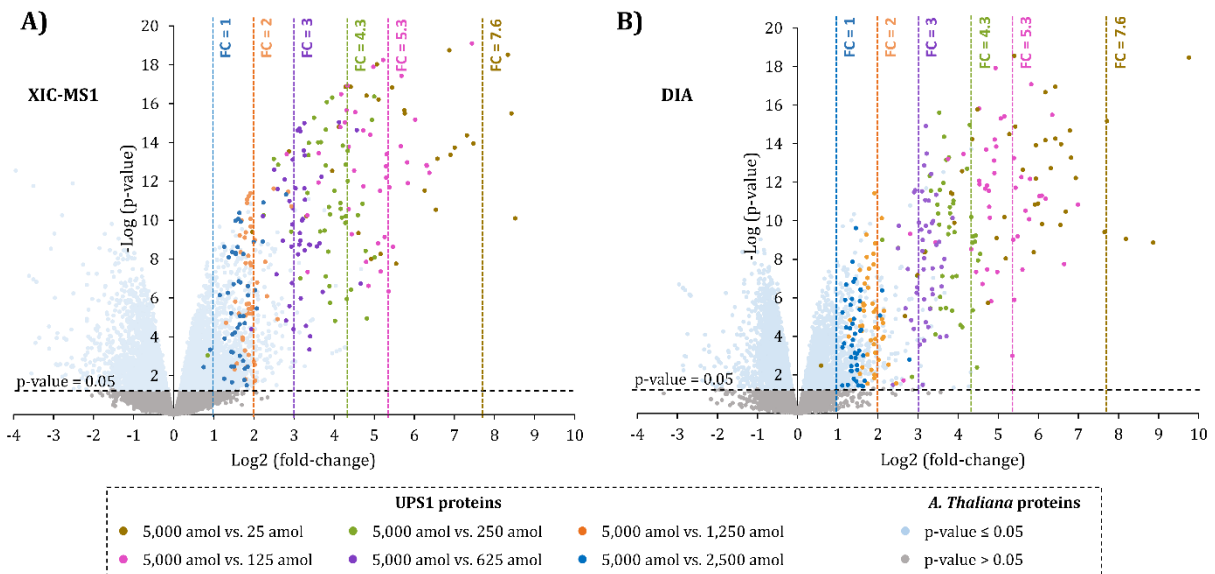


Figure 45: Differential analysis of the UPS1 proteins quantified in an *Arabidopsis thaliana* mixture. Volcano plot of the experimental fold-change as a function of the p-value for each protein for XIC-MS1 (A) and DIA (B). Light blue dots stand for *Arabidopsis thaliana* proteins and colored dots for UPS1 proteins. Theoretical ratios were calculated from the highest point of the range, represented by vertical dashed lines.

These results confirm the interest of DIA as a global approach for quantitative proteomics. Indeed, the spectrum-centric approach displays comparable identification results to XIC-MS1 from DDA for UPS1 peptides and protein groups. Furthermore, the accuracy of UPS1 peptides quantification is comparable for both strategies, with even an improvement in the number of quantified UPS1 proteins at low concentrations, i.e., 25 and 125 amol, with DIA. Finally, we demonstrate the need to optimize the isolation windows scheme to get the best results from the samples. Although encouraging results, some limitations regarding the quantification accuracy and signal extraction still need to be mentioned

for low-level proteins. Altogether, this work reveals that evaluating optimized methods with well calibrated samples is very useful to assess overall performances of an LC-MS/MS system.

4. Optimization of high-field asymmetric waveform ion mobility spectrometry (FAIMS) compensation voltages for quantitative proteomics

Today, mass spectrometers can collect tandem mass spectra at great speed and sensitivity. Their combination to cutting-edge chromatographic systems and advanced precursor characterization⁷⁷ increases the number of precursors and, consequently, the depth of analysis. Still, a single-shot DDA analysis does not identify more than half of the expressed proteome⁹¹. Indeed, the mixture complexity, the large dynamic range and the inability to correctly separate precursors are limiting the identifications. To circumvent these limitations, approaches based on mixture complexity reduction using online (e.g., 2D LC-MS/SM) or offline (e.g., 2D-gel) fractionation had been developed. Although improving the number of identified features, they also require high sample amounts and considerable effort during the sample preparation step. Hence, ion mobility mass spectrometry (IM-MS) revealed to be a promising alternative to increase peak separation capacity without additional efforts during sample preparation.

The high-field asymmetric waveform ion mobility spectrometry (FAIMS) technology allows a fast and efficient gas phase separation of precursor ions as they leave the electrospray emitter before entering the mass spectrometer. Ions are transmitted between the inner and the outer electrodes of the FAIMS source as a function of their mobility in a high or low electric field generated from an asymmetric waveform⁷⁷. While ions with no or limited difference in mobility are transmitted, ions with a significant enough difference in mobility between the high and low electric fields collide onto the electrodes. The ion path may be modified by adding a dc voltage, named compensation voltage (CV). The selection of an appropriate CV will compensate for the trajectory of a specific ion or group of ions, allowing them to pass through the device and enter the mass spectrometer. By switching the CV, alternate groups of ions will pass (**Figure 46**). Thus, the choice of the compensation voltage value, determined empirically, allows for controlling which population of ions is passing through the FAIMS device. Thanks to the quick transition between two CVs values (around 25 ms per transition), switching between several CVs in a single MS analysis is possible. Hence, ions with different charges can be analyzed separately.

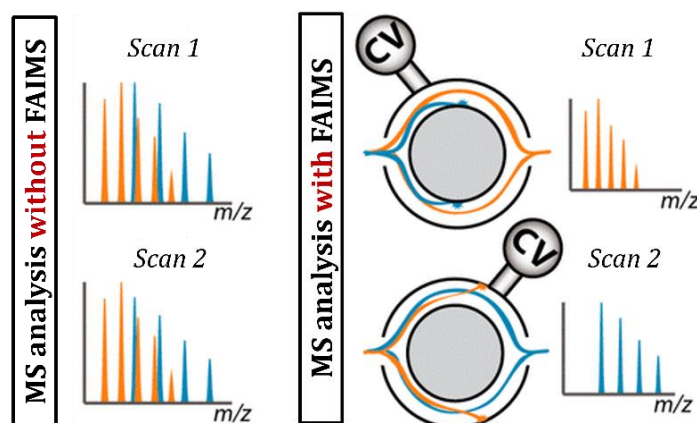


Figure 46: Scheme representing the capacity of the FAIMS to switch between compensation voltages during a MS analysis⁷⁷.

Given these points, we used a 100 ng HeLa Pierce™ digest (Thermo Fisher Scientific) to optimize the compensation voltage values of the FAIMS device and compared the results to no FAIMS analysis. The study was performed in DDA mode, and HCD fragmentation using both ion trap and Orbitrap MS2 detection were tested. Recommendations from the literature and the manufacturer state that a maximum of two CVs must be used for runs between 1 and 2h when using Orbitrap detection. Because

of its higher speed of analysis, ion trap-based MS2 detection methods would allow the use of an additional CV, i.e., 3 compensation voltages. Values tested were selected from the literature and are displayed in **Table 6**.

Table 6: List of compensation voltages evaluated

Compensation voltages evaluated	Orbitrap MS2 detection	Ion trap MS2 detection
	No FAIMS -45V -60V -50V -70V	No FAIMS -40V -60V -80V -50V -65V -85V

As explained above, depending on their mobility, each CV will favor a different group of ions. Hence, base peak chromatograms extracted from the analysis without FAIMS and runs with different CV values are entirely different (**Figure 47**). In addition, the total ion chromatogram intensity is 2 to 9 times lower, depending on the CV value, when the FAIMS is used. Indeed, some ions, particularly singly charged ones, are not entering the mass spectrometer, reducing background noise and signal intensity.

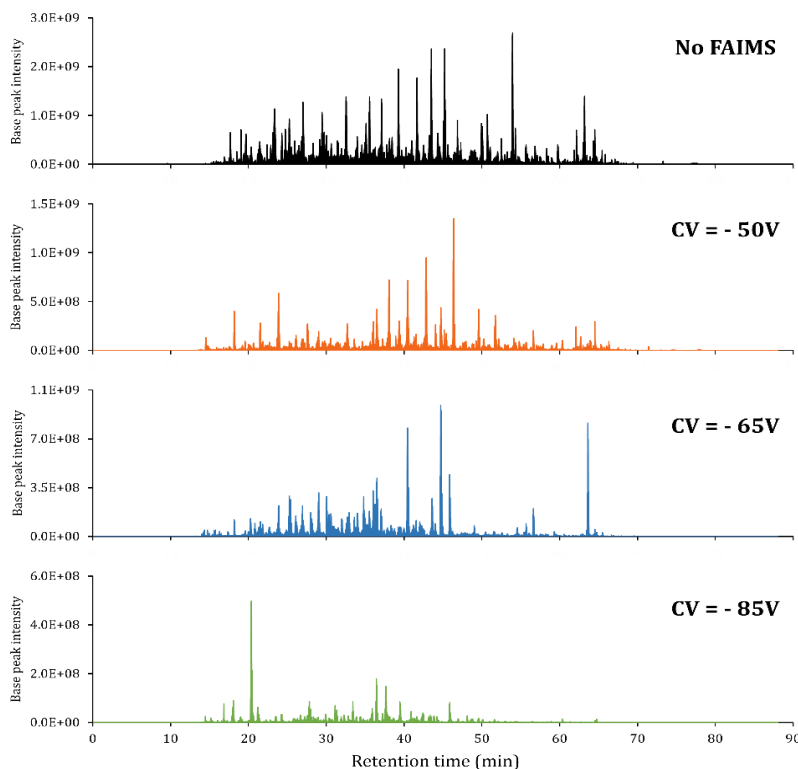


Figure 47: Base peak chromatogram of a 100 ng HeLa Pierce™ digest obtained without the FAIMS (black) and with the FAIMS for CV of -50 V (orange), -65 V (blue) and -85 V (green).

The results from the CV optimizations are represented in **Figure 48**. First and foremost, they confirm what has been previously observed (**Part I, Chapter 1.3.A**) with more identified features when the MS2 detection is performed in the ion trap compared to the Orbitrap. On average, we identified 13% and 25% more proteins with the ion trap compared to the Orbitrap without and with the FAIMS, respectively (**Figure 48.B**). The best CV combinations allow identifying up to 3,600 proteins (-50 V and -70 V) with Orbitrap detection and 4,500 proteins (-50 V, -65 V, and -85 V) with ion trap detection. It represents an 8% and 20% increase compared to results without the FAIMS, respectively. In opposition, we detected fewer peptides when the FAIMS device was used compared to its absence. Indeed, the FAIMS prevents some precursors from entering the mass spectrometer because of their mobility. Nevertheless, precursors that could not be detected in the absence of FAIMS because of their low abundance or co-elution with other precursors can now be observed, thanks to the cleaner signal

obtained once the FAIMS is up. Background noise reduction and selection of precursors with specific charges and/or ion mobility increase the diversity of precursors, resulting in a gain in the number of protein groups identified.

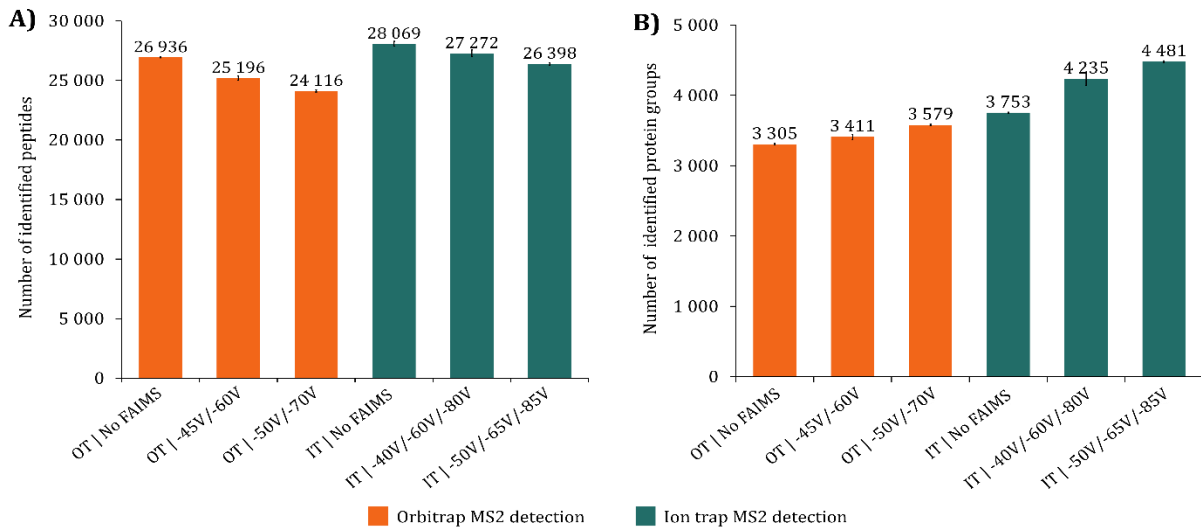


Figure 48: Identification results from 100 ng HeLa Pierce™ digest. Number of (A) peptides and (B) protein groups identified according to different acquisition methods: using Orbitrap (OT) MS2 detection (orange) without the FAIMS and with CV -45V/-60V and -50V/-70V, and ion trap (IT) MS2 detection (green) without the FAIMS and with CV -40V/-60V/-80V and -50V/-65V/-85V.

When selected cautiously, CVs maximize the number of precursor ions passing through the device (**Figure 49**). To clarify, we saw that the more different the chosen CVs values are in an analysis, the lower the overlap between precursors selected by multiple CVs and the higher the number of proteins identified. As an example, the analysis with CVs of -50 V/-70 V, where only 4% of the peptides are detected with both, leads to the identification of 170 additional proteins compared to the analysis with CVs -45 V/-60 V, where 11.9% of the peptides are shared between both CVs values.

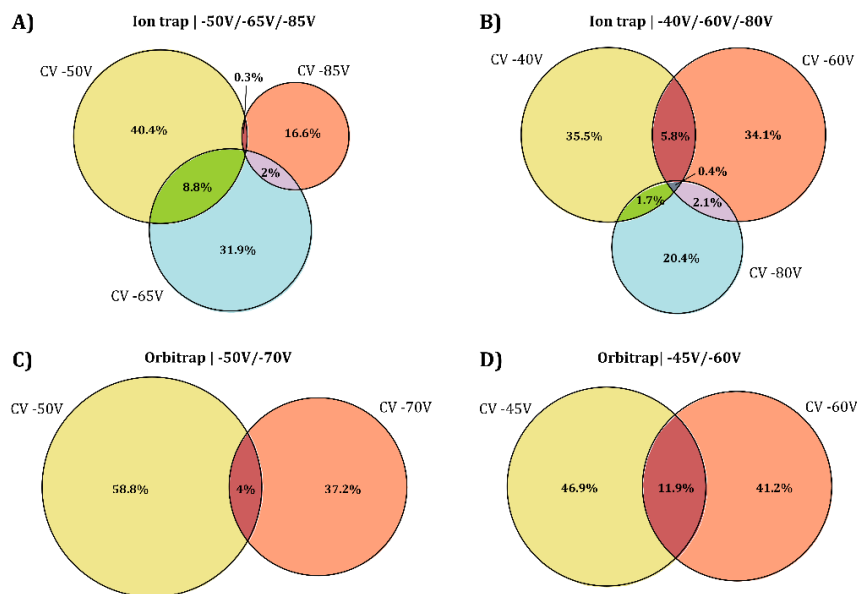


Figure 49: Venn diagrams representing the proportion of peptides identified by each CV in FAIMS MS analysis. (A) Ion trap MS2 detection with CV -50V, -65V and -85V. (B) Ion trap MS2 detection with CV -40V, -60V and -80V. (C) Orbitrap MS2 detection with CV -50V and -70V. (D) Orbitrap MS2 detection with CV -45V and -60V.

To correlate what have been stated, the proportion of each precursor charge state depending on the FAIMS compensation voltage is represented in **Figure 50**. Higher CVs (i.e., -40 V – -50V) favor the transmission of doubly charged ions while lower values (i.e., -70 V – -85 V) favor triply charged ions. Higher charge state precursor (i.e., $\geq 4+$) are mostly transmitted for CVs around -50 V – -60 V. Comparisons with the charge state repartition for analysis without FAIMS point out the benefits of low CVs values to transmit additional peptide ions, as depicted in **Figure 49**. Indeed, CVs of -40 V and -45 V present the same charge state repartition, contrary to lower values, explaining the low overlap between two distant CVs.

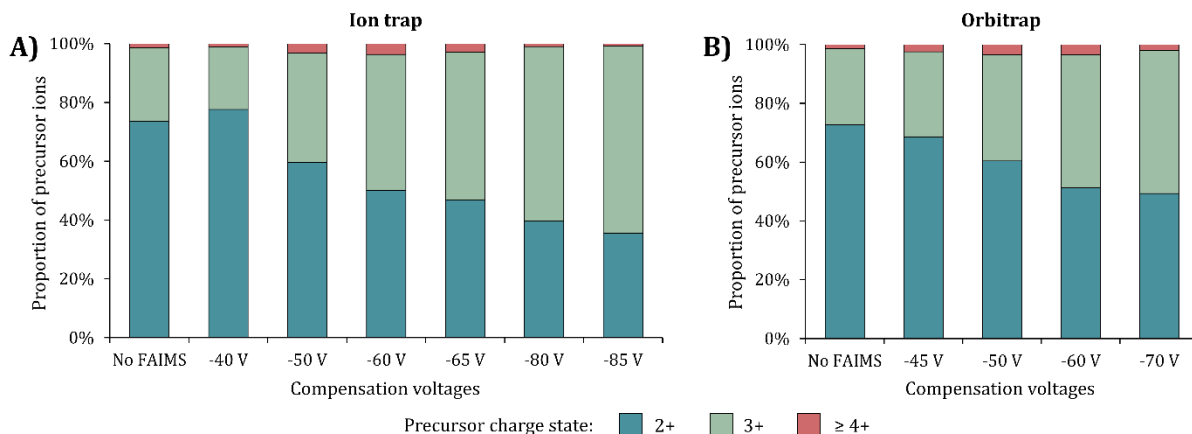


Figure 50: Repartition of precursors charge state depending on the CV. (A) Ion trap analyses. (B) Orbitrap analyses.

We also looked at the quantification results along with these observations on the identified features. The same trends can be noticed for all analyses, in which 97%-99% of the identified proteins and 94%-96% of the identified peptides were quantified (**Figure 51**). Notably, the distribution of the coefficient of variations is comparable between analyses with and without the FAIMS (**Figure 52**), and a comparable proportion of features is quantified with a coefficient of variation $<20\%$ (**Figure 51**). However, multiple CVs induce a drop in the median number of points per peak when FAIMS is used, from 22 to 11 and from 20 to 13 with IT and OT detection, respectively.

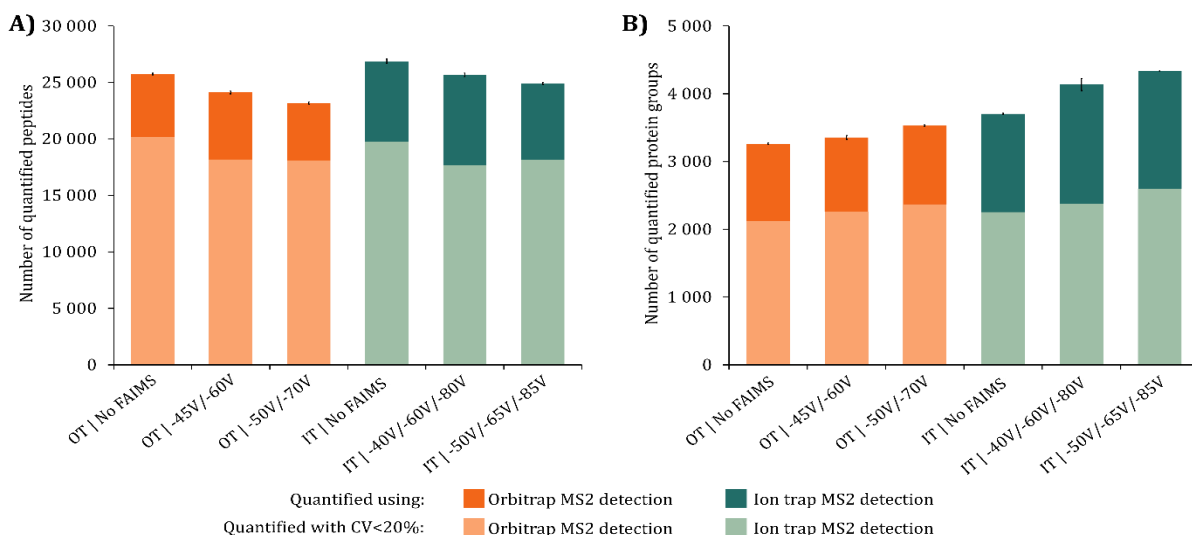


Figure 51: Quantification results from 100 ng HeLa Pierce™ digest. Number of quantified peptides (orange) and protein groups (green) and quantified with a coefficient of variation $<20\%$ (light colors) from 100 ng HeLa Pierce™ digest analyzed according to different acquisition methods: Orbitrap (OT) MS2 detection in the absence of the FAIMS and its presence with CV -45V/-60V and -50V/-70V, and ion trap (IT) MS2 detection in the absence of the FAIMS and its presence with CV -40V/-60V/-80V and -50V/-65V/-85V.

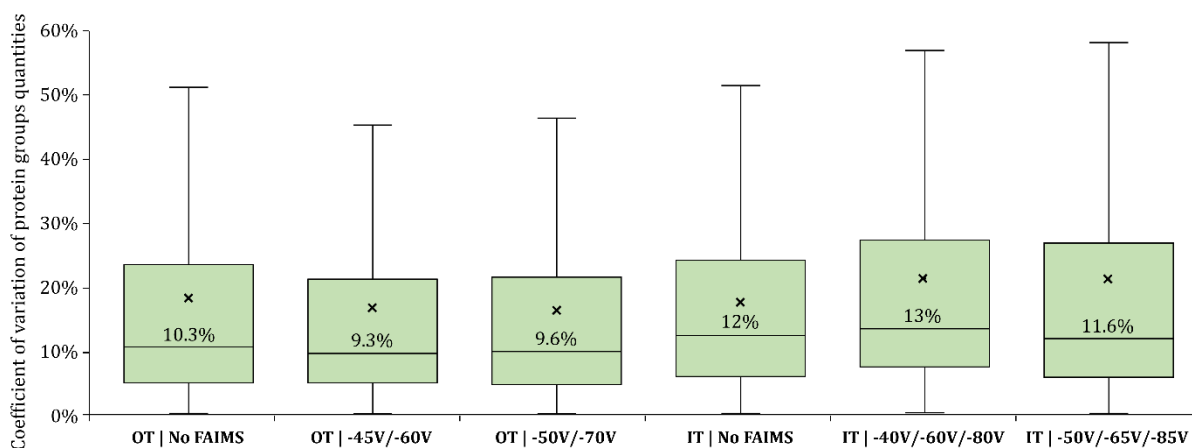


Figure 52: Boxplot representing the distribution of the coefficients of variation calculated from the average abundance of the proteins quantified in all the replicates from analysis with and without the FAIMS using Orbitrap (OT) or ion trap (IT) for MS2 detection. Values displayed correspond to the median CV and crosses to the mean of the CVs. For clarity purpose, outliers were removed.

In summary, implementing the FAIMS in front of the mass spectrometer increases the number of identified and quantified proteins, allowing for the detection of more than 4,500 protein groups in a single-shot analysis. When using FAIMS, selecting suitable compensation voltages is the key to perform a good analysis, and these values need to be determined empirically. Even if the use of multiple compensation voltages causes a decrease in the number of points per peak, their careful optimization allows obtaining reproducible quantification results with a higher proteome depth.

5. Evaluation of acquisition strategies dedicated to TMT-labelled samples

Tandem Mass Tag (TMT) is a chemical labeling quantification strategy based on isobaric labels. These isobaric labels are composed of a group that reacts with the primary amines of the peptides (i.e., N-terminus of the peptides and lysine side chain) and another group that cleaves during fragmentation. The latter includes so-called reporter and balance groups, which will present several variants but keep an identical global mass. Thus, these different isobaric variants are only differentiated by the masses of the fragments produced. Multiple TMT labels are available today, from TMT0 containing a unique label to TMTpro™ 18plex¹³¹ with 18 quantification channels. The mass difference between two neighbor channels in TMTpro™ kits (e.g., 130N and 130C) is only 0.00632 Da. Hence, their analysis requires extreme resolving power.

Considering the benefits of multiplex approaches for analyzing large cohorts or for specific applications such as thermal proteome profiling⁵³⁷, we decided to optimize acquisition methods dedicated to TMT. On the one hand, we fine-tuned the collision energy required for TMT analysis. On the other hand, we evaluated the potential of the tribrid geometry that allows MS3 and real-time search (RTS) MS3 analyses to improve quantification.

A. Optimization of the fragmentation energy

Peptide quantification using TMT labeling relies on accurately measuring low-mass reporter ions generated during the MS2 fragmentation process. Regarding the absence of low-mass cutoff, high-resolution detection, and high-quality MS/MS spectra generated, the combination of HCD fragmentation with Orbitrap detection is most commonly used for isobaric-based quantification^{523, 538}. Even so, a low fragmentation energy will result in low reporter ion intensity, decreasing the quantification precision, while high HCD fragmentation energy can lead to less confident

identifications⁵³⁹. Hence, optimizing the collision energy is crucial to ensure precise and accurate peptide quantification without compromising identifications^{540, 541}.

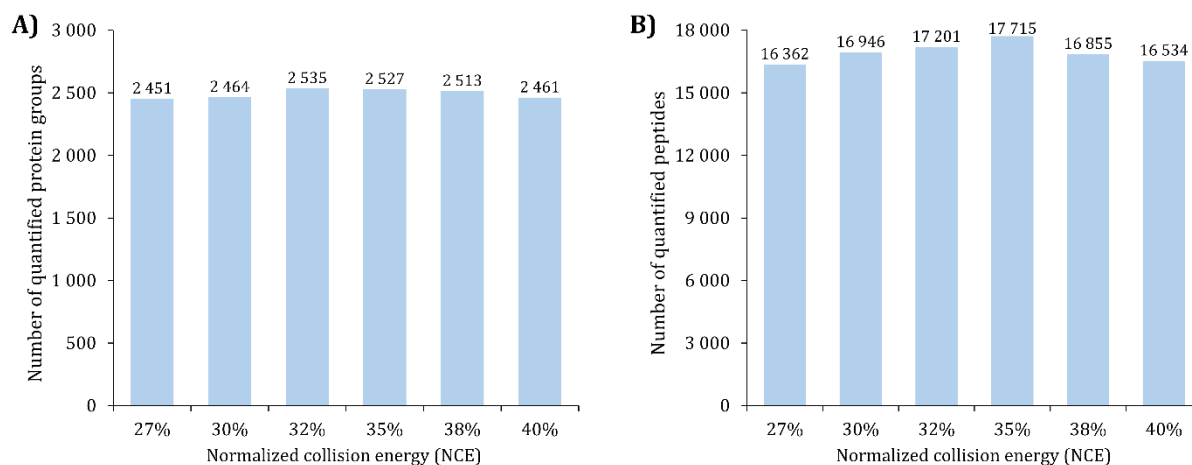


Figure 53: Number of TMT-labeled quantified (A) protein groups and (B) peptides quantified in all the replicates depending on normalized collision energy.

We decided to test several HCD normalized collision energy values (NCE) from 27% to 40% on a digest of HeLa cells labeled with TMT0. We quantified a comparable number of protein groups, between 2,451 and 2,535, independently of the NCE (**Figure 53.A**). Differing from protein groups, a maximum of 17,715 peptides were quantified for a NCE of 35%, representing an 8% increase compared to peptides quantified with a NCE of 27% (16,362 peptides), as represented in **Figure 53.B**. A too-high NCE induces a drop in quantified features, with up to 1,200 fewer peptides quantified at NCE 40%.

In addition, the distribution of the S/N ratio is displayed in **Figure 54**. As expected, we see the correlation between the rise of the S/N ratio and a higher NCE. We observed a 3-fold increase in the S/N ratio between NCE 27% and NCE 40%. Similarly, values are more centered on the median for higher NCEs than lower ones.

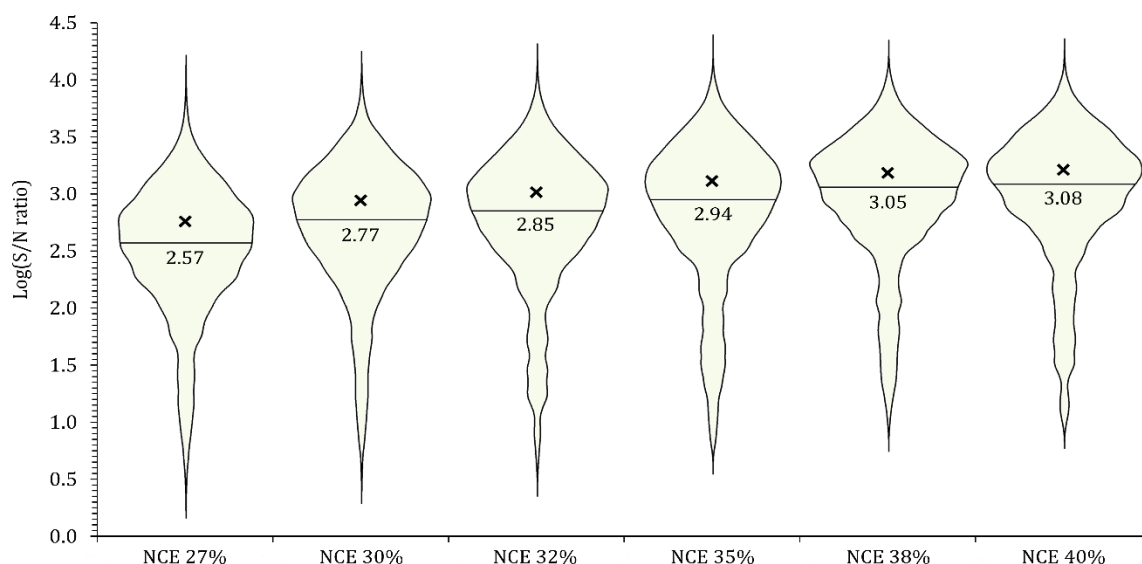


Figure 54: Distribution of the S/N ratio of the TMT-labeled peptides depending on the collision energy. Values below horizontal lines are the median, represented by the lines and crosses represent the average S/N ratios.

Therefore, a normalized collision energy of 35% appears to be the optimal value, leading to the highest number of quantified peptides with a median S/N ratio of 870. This corresponds to the values found in the literature where the optimal normalized collision energy is generally between 35% and 45%^{540, 542}, depending on the instrument's geometry.

Indeed, an increase in the normalized collision energy leads to a higher S/N ratio, but a too-high NCE also decreases the number of identifications. Hence, a reasonable compromise should be found between reporter ion intensity and peptide identification to give the highest number of quantified features. Some groups also discussed concomitant HCD and CID MS/MS acquisition modes as a valuable alternative for improved identification and quantification of TMT-labeled peptides showing comparable results to what is obtained with optimal HCD fragmentation^{540, 541}.

B. Comparison of MS2 and MS3 methods for TMT quantification

Multiplexing quantitative analyses using isobaric channel labeling demonstrates excellent results for high throughput, increasing coverage by avoiding missing values, and deepening analysis. Nevertheless, TMT ratio compression resulting from the co-isolation and co-fragmentation of interfering ions remains one of the main limitations of TMT quantification⁵⁴³. Several approaches, including the reduction of the precursor isolation window and the use of MS3-based quantification, were developed¹³³ to mitigate the negative impact of interfering signals. Precursor ions are fragmented using CID and analyzed in the ion trap. Then, a subsequent HCD fragmentation of the MS2 fragment ions leads to the release of the isobaric tag for quantification in a MS3 scan (**Figure 55.A**). While drastically lowering the ratio interference, MS3 methods cause a decrease in the number of quantified peptides due to sensitivity loss and significant reduction of the acquisition speed. Consequently, several alternatives have been devised to preserve the benefits of MS3-based quantification while obtaining as many quantified peptides as with MS2 approaches. First, synchronous precursor selection (SPS) using isolation waveform with multiple frequency notches increases the number of reporter ions selected for MS3, hence increasing their signal¹³³. Second, real-time search-assisted acquisition (RTS)⁵⁴⁴ reduces the cycle time by selecting only MS1 precursors leading to peptide identification for the time-consuming MS3 quantification step (**Figure 55.B**).

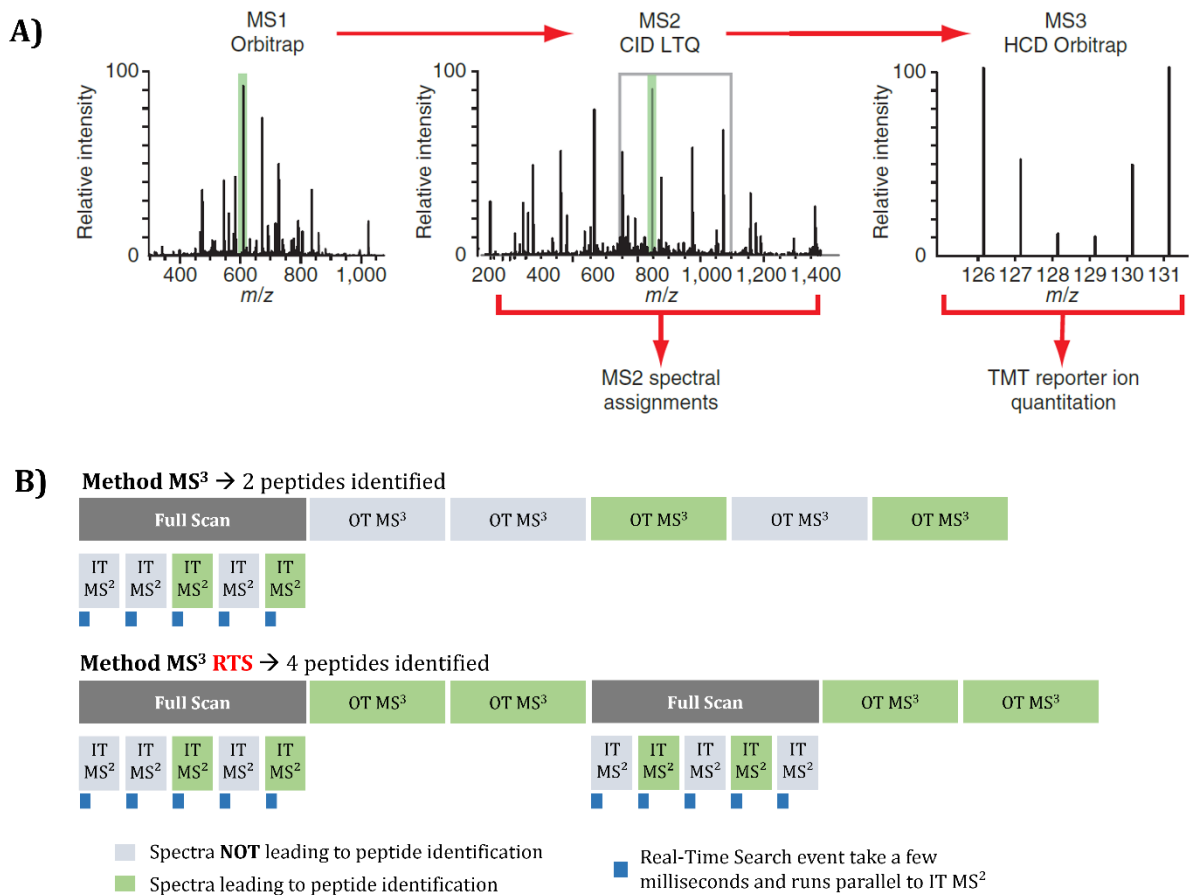


Figure 55: Acquisition strategies for TMT experiments. (A) Precursor ions for MS2 spectra were selected from a full MS scan acquired on the Orbitrap. Fast CID-MS2 experiments were used for peptide fragmentation and assignment, followed by the selection of the most intense fragment ions from MS2 (green) for HCD-MS3, where TMT reporter ion intensities are measured in the Orbitrap¹³³. (B) For classical MS3 acquisition, all precursors selected for MS2 are submitted to further MS3 fragmentation while using the real-time search (RTS) option, only precursors assigned to a peptide from the database (green) are further fragmented in HCD-MS3 to obtain a quantification spectrum.

We labeled HeLa cells digest with TMTpro™ 16plex and used quantification channels 130N, 130C, 131N, and 131C in an equimolar ratio for our experiment. First, the MS2 acquisition method considered as the reference method quantified up to 2,900 protein groups and 16,500 peptides (**Figure 56**). Compared to MS2, the MS3 approach quantified 35% fewer peptides and 27% fewer proteins (i.e., 10,656 peptides and 2,118 protein groups). As mentioned before, the reduced acquisition speed can explain this difference. Indeed, only 40,000 MS2 spectra (used for peptide identification) were acquired in the MS3 method, against more than 60,000 with the MS2 method (**Figure 56**). In addition, the proportion of quantifying spectra (i.e., MS2 spectra for the MS2 method and MS3 for the MS3 method) is lower using the MS3 method, 20% compared to 28%, meaning that 80% of the time used to acquire MS3 spectra does not provide any quantitative information. On the contrary, more than 2,800 protein groups and 14,700 peptides were quantified by adding the real-time search-assisted acquisition MS3 (RTS MS3). These results are similar to the MS2 approach and represent a 39% and 34% gain at the peptide and protein group levels, respectively, compared to the MS3 approach. Thanks to the RTS option, more than 80,000 MS2 spectra were acquired, and the proportion of quantifying spectra leading to a quantification result rose to 60% (**Figure 56.B**). The remaining proportion of MS3 spectra (i.e., not leading to a quantification result) can be explained by the difference between the search engine used for RTS (Comet⁸⁹) and the post-analysis search (Sequest HT), as well as the 1%

FDR filtering performed after the analysis. Nevertheless, these results demonstrate the ability of RTS MS3 to circumvent the limitations of the MS3 method by increasing the acquisition speed, resulting in a higher number of quantified features.

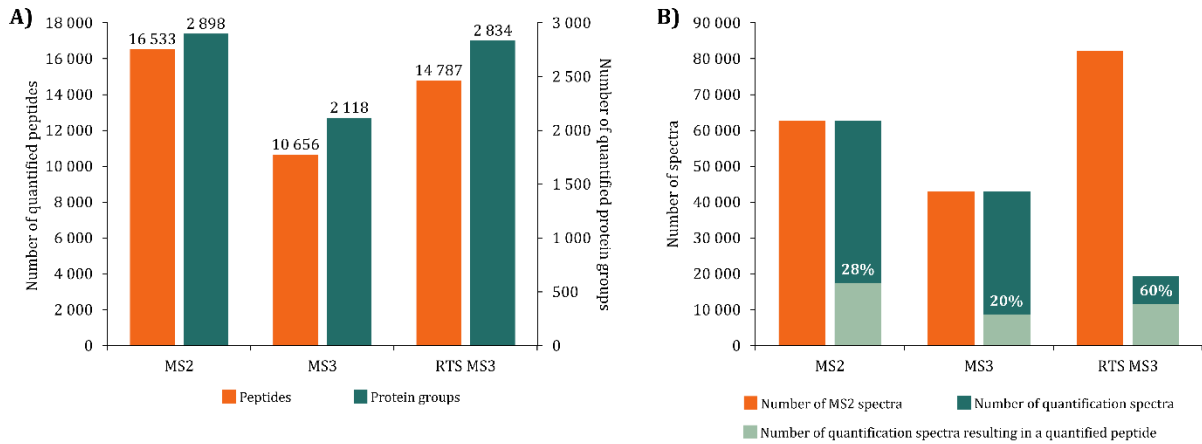


Figure 56: Results obtained from HeLa cells digest labelled with TMTpro™ 16plex. (A) Number of quantified peptides (orange) and protein groups (green). (B) Number of MS2 spectra (orange) and quantification spectra, i.e., MS2 spectra for MS2 methods and MS3 spectra for MS3-based methods, (green). Percentages indicated stand for quantification spectra leading to a peptide quantification value in the results (hatched green).

In the same fashion, a comparison of the quantification accuracy of each method, i.e., MS2, MS3, and RTS MS3 was conducted. Therefore, we represented the distribution of the experimental ratios 130N vs. 130C, 131N vs. 130C, and 131C vs. 130C for each method (**Figure 57**). For all methods, the median of the experimental ratio is close to the theoretical value (ratio 1:1), and the same trend is observed, with a slightly higher median value for the 131C vs. 130C ratio with all methods, probably due to a sample preparation bias. Nevertheless, both MS3-based approaches reveal a slightly higher dispersion around the median than MS2. Of note is also that we used a 0.7 m/z precursor isolation window for all methods, while 2 m/z windows are generally used in the literature. Using such a narrower isolation window helps limiting ratio compression from interfering ions.

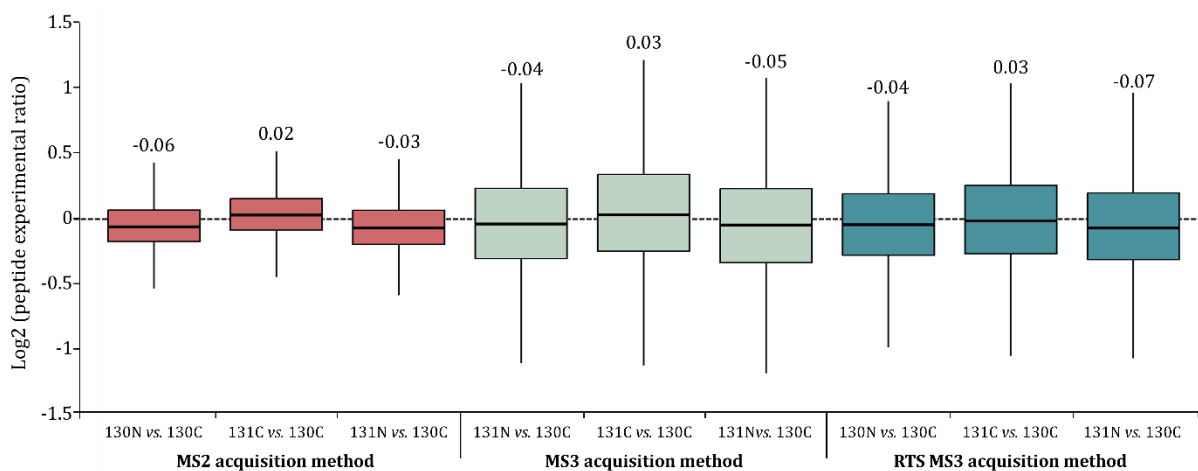


Figure 57: Results obtained from HeLa cells digest labelled with TMTpro™ 16plex using channels 130C, 130N, 131C and 131N. Distribution of the peptide experimental ratios 130N vs. 130C, 131C vs. 130C and 131N vs. 130C for MS2 (red), MS3 (green) and RTS MS3 (blue) acquisition methods. Channel 130C was selected as reference channel and the theoretical ratio is represented by the dashed black line. Values displayed on top of the boxplots are median of the experimental ratio.

In summary, MS2 quantification appears to remain the best approach for TMT quantification, despite the good performances of RTS MS3. While the MS2 method can suffer from ratio compression from interfering ions, a narrow isolation window helps limiting this phenomenon. At the same time, the addition of real-time search to a MS3 quantification-based method increases the acquisition speed (compared to MS3 only methods) by fragmenting only precursor ions that can lead to a PSM. This strategy allows obtaining comparable numbers of quantified features as an MS2-based quantification. The study's goal should direct the choice of the method. Today, new features for TMT quantification methods are being developed. Papers demonstrated the benefits of adding FAIMS in front of the MS⁵⁴⁵, ⁵⁴⁶ and it is now possible to perform filtering based on the FDR or the score during the real-time search, further reducing the chance of fragmenting ions that would not lead to useful quantitative information.

6. Conclusion and perspectives

This chapter evaluates several factors in LC-MS/MS for quantitative proteomics. The first part showed that the choice of the analyzer and the fragmentation method are critical to ensure the accurate quantification of thousands of proteins and peptides. In addition, the cycle time and the isolation window scheme were identified as critical parameters of DIA method optimizations. Subsequently, the benefits of these optimizations for quantitative proteomics were proven on a set of well-calibrated samples. In the second part, it was exposed that the choice of the compensation voltages is of utmost importance when FAIMS is implemented to maximize its advantages. In the last part, collision energy and quantification strategy (i.e., MS2 or MS3) were identified as the main elements to consider for TMT-based quantification.

Many improvements are continuously achieved at the experimental design or the MS level. Developing commercial multiplexing approaches, such as TMTpro™ 18plex¹³¹ or DiLeu labeling⁵⁴⁷ with up to 18 and 21 quantification channels, allows for performing relative quantification at very high throughput. In addition, some studies in the literature reported the combination of two TMT kits to increase the multiplexing capabilities to 27plex⁵⁴⁸ or 29plex⁵⁴⁹. Regarding MS, new acquisition modes are continuously developed to increase proteome coverage with high sensitivity and repeatability. It is particularly true for DIA mode, now available in various ways. Some use large, continuously sliding precursor windows for fragmentation to enhance the correlation of precursor and fragment ions, like SONAR^{161, 550} or Scanning SWATH¹⁶². Recently, the sequential release of ions from high to low *m/z* using an ion trap before the TOF analyzer, the ZenoSWATH technology⁵⁵¹, produced cleaner spectra improving the sensibility. Others combined the benefits of ion mobility as an additional separation dimension to decrease the spectral complexity, using FAIMS⁵³⁵ or trapped ion mobility separation (TIMS⁷³) for the DIA-PASEF technology¹⁶³ to improve the depth of analysis. In addition, a study brought together sample multiplexing and DIA MS as a new alternative⁵⁵².

All these new technologies bear witness to the constant progress in quantitative proteomics. Nevertheless, the bioinformatics processing of the data after MS acquisition remains one of the most critical steps for the precise and accurate quantification of proteomes. It will be developed in **Part I, Chapter 2**.

Chapter 2:

Evaluation of bioinformatics tools for data processing

1. Context of the project

Data processing is a major step in the success of quantitative proteomics studies. In this chapter, we focused on analyzing DDA and DIA data. While the former approach focuses on the fragmentation of an intense precursor ion at a time, isolated in a small isolation window, the latter co-isolates and co-fragments all precursors in a wide isolation window. MS2 spectra generated are then subjected to data analysis for peptide identification and quantification. For DDA, experimental precursor and fragment ions masses are generally compared to theoretical ones obtained from the *in-silico* digestion and fragmentation of a specific protein sequence database. However, many limitations exist for DDA data analysis, including, but not limited to, the stochasticity of MS/MS acquisition, the presence of missing values, the co-isolation and co-fragmentation of precursor ions in small isolation windows leading to chimeric spectra or the underestimation of the number of modifications carried by the peptides⁹². To circumvent these limitations, multiple strategies were developed, taking advantage of the power of artificial intelligence and machine learning. Numerous tools are available to date, drastically increasing the number of features identified. Rescoring tools use additional predicted features, such as theoretical ion fragmentation patterns and retention time predictions, to calculate and assign more accurate scores, resulting in a gain of identifications at the same confidence level. Nevertheless, many spectra remain unidentified in a standard DDA proteomics experiment^{91,93}. One main reason identified for this is the presence of unexpected modifications of peptides, resulting either from post-translational modifications (PTMs)⁵⁵³ or chemical modifications induced during sample preparation⁵⁵⁴. Indeed, classical identification search engines restrict the search space to something manageable by only considering a limited number of modifications. Consequently, many relevant modifications may be missed and result in false negatives. Several open modification search (OMS) tools have been developed to identify these peptides, such as Open-pFind⁹⁵, MSFragger⁹⁴, and SpecOMS⁵⁵⁵. The CompOmics group (Ghent, Belgium) recently developed an OMS tool named ionbot⁹⁶. It combines all the relevant matching information, including retention time and fragmentation predictions, in a machine learning-based score to confidently identify peptides with unexpected modifications.

Given that the link between fragment ions and their corresponding precursors is lost, conventional data processing strategies cannot be employed for DIA data^{173,556}. As a result, two main approaches have been developed to overcome the complexity of these data: the peptide-centric and the spectrum-centric approaches. The first one, proposed in 2012 by Gillet *et al.*⁶, enables the identification and quantification of peptides using information contained in a spectral library generated from MS2 spectra acquired in DDA and for which peptide sequences have been assigned with confidence⁵⁵⁷. Finally, the quality of the signals is controlled by the use of scores to validate peptide detection. The second one, known as spectrum-centric, is designed to eliminate the limitations induced by using a spectral library. Indeed, it is important to emphasize that only peptides and proteins present in the spectral library can be identified during DIA analysis. To this end, MS2 pseudo-spectra are typically generated from a fasta file. From there, conventional searches are performed using DDA search engines and protein inference tools.

Since its introduction in 2003¹⁵⁰, many bioinformatics advances have been made, with the development of new algorithms based on the use of deep neural networks and machine learning for

the spectrum-centric approach^{148, 180, 183, 184, 558, 559}. They predict fragmentation and retention time more accurately, increasing the number of identified and quantified peptides.

In this context, we appraised recent DDA and DIA data processing tools. In the first step, we compared classical DDA database search strategies (Mascot and Sequest HT) to approaches including commercial (Inferys¹²⁰) and open-source (MS²Rescore¹¹⁹) rescoring tools as well as innovative search engines for the improved identifications from deconvolution of chimeric spectra (Chimerys, MSAID) or the ionbot algorithm for open modification search. In the second step, we assessed the capabilities of spectrum-centric approaches for DIA data processing. We compared the library-free approaches search parameters of commercial and open-source solutions, namely Spectronaut Pulsar™ and DIA-NN¹⁸³, for proteins and peptides identification and quantification.

2. Analytical strategy applied for data processing optimizations

The overall analytical scheme is shown in **Figure 58**.

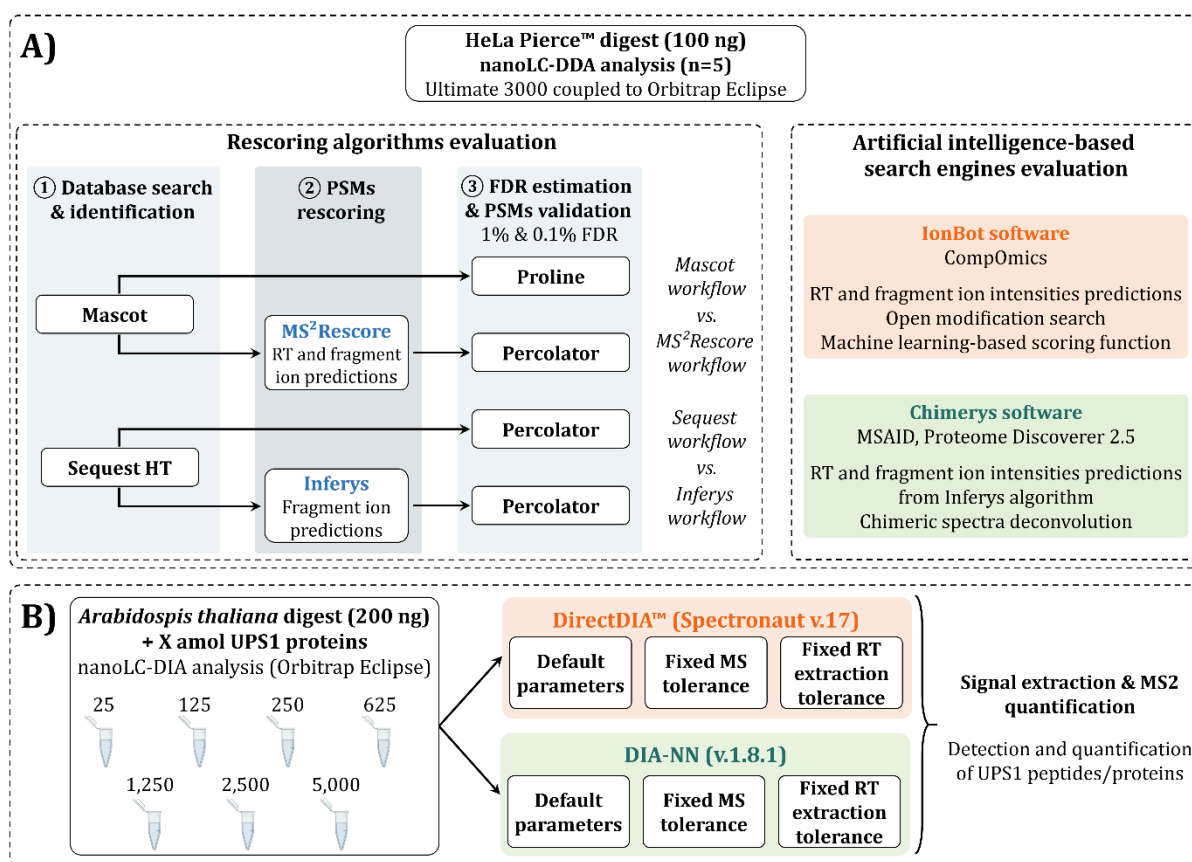


Figure 58: Analytical scheme used for the evaluation of innovative bioinformatics tools for DDA and DIA data processing. (A) HeLa Pierce™ digest was analyzed in five replicates using a Top20 DDA acquisition strategy. First, we evaluated rescoring algorithms. Data were searched against classical search engines, Mascot and Sequest HT, and then PSMs were rescored using MS²Rescore (for Mascot output) and Inferys (for Sequest HT output). Data were then validated at 1% and 0.1% FDR to evaluate the performances of the search engines. Second, the raw data were processed with ionbot, allowing open modification searching, and Chimerys (MSAID, Proteome Discoverer 2.5), and enabling chimeric spectra deconvolution, to evaluate their performances. (B) Raw data from *Arabidopsis thaliana* digest spiked with UPS1 proteins analyzed in DIA mode (see **Part 1, Chapter 1**) were processed by two library-free search engines, DirectDIA™ (Spectronaut v.17) and DIA-NN¹⁸³ (v1.8.1). For the searches, three approaches were tested, either using the default parameters implemented in the software or by setting the MS tolerances or retention time (RT) extraction windows. Finally, the quantification of UPS1 proteins and peptides was appraised.

First, we used a HeLa Pierce™ (Thermo Fisher Scientific) cells digest injected in five replicates to compare the different DDA data processing strategies. NanoLC-MS/MS analyses were performed on an Orbitrap Eclipse Tribrid instrument operated in Top20 DDA mode with both MS1 and MS2 spectra detection using the Orbitrap to ensure high-quality data. Data were searched with classical identification search engines, and the results were compared with those obtained after the addition of a rescoring step, performed by multiple algorithms. As the combination of all rescoring algorithms tested is not possible with the same search engine, we used two different associations (**Figure 58.A**). On the one hand, we used the search results from Mascot followed by the validation using Proline¹³⁸ as the standard before rescoring. Then, output from Mascot were rescored using the open-source MS²Rescore (CompOmics, Ghent, Belgium) algorithm. On the other hand, results from Sequest HT followed by Percolator validation, both included in Proteome Discoverer 2.5 software (Thermo Fisher Scientific), were used as reference for Inferys (MSAID) commercial rescoring tool evaluation. Inferys was added to the workflow to perform rescoring of the data obtained from Sequest HT. To ensure a clear understanding of the workflows used, they will be named in this chapter as follows:

- Mascot workflow: raw data were searched using the Mascot search engine and FDR estimation was performed using the target/decoy strategy based on the score in Proline.
- MS²Rescore workflow: raw data were searched using the Mascot search engine, and all the PSMs from Mascot were rescored by MS²Rescore, before FDR estimation using Percolator.
- Sequest workflow: raw data were searched using the Sequest HT search engine and using Percolator for FDR estimation, both implemented in Proteome Discoverer 2.5.
- Inferys workflow: raw data were searched using the Sequest HT search engine, and all the PSMs from the search were rescored by Inferys (also implemented in Proteome Discoverer 2.5), before FDR estimation using Percolator.

In addition, we compared Chimerys, a commercial solution from MSAID (Munich, Germany) implemented in Proteome Discoverer 3.0 (Thermo Fisher Scientific) using an algorithm taking advantage of artificial intelligence-based predictions from Inferys directly during the search, to the aforementioned strategies. Moreover, we appraised the capability of ionbot developed by the CompOmics group (Ghent University, VIB, Belgium) to perform open modification searches. Searches were performed with open modification searching enabled and results were compared in terms of identifications with those obtained with classical search engines (**Figure 58.A**).

Second, raw data acquired on a mixture of the 48 UPS1 proteins spiked into a total *Arabidopsis thaliana* lysate were used to appraise DIA data analysis tools in terms of identification and precision of quantification. Two software were compared: Spectronaut Pulsar™ (Biognosys, Schlieren, Switzerland) and DIA-NN (Demichev lab, Munich, Germany). Searches were performed using the default parameters defined directly by the algorithm based on the data and with defined parameters regarding the mass tolerances (5 ppm for MS search and 20 ppm for MS/MS) or the retention time extraction window (10 min with DirectDIA™ and 5 scan windows in DIA-NN). All data were validated using a 1% FDR at precursor level. Numbers of identified and quantified features, as well as the precision of quantification, were evaluated (**Figure 58.B**).

Additional details on sample preparation, LC-MS/MS methods and data processing are provided in the **Experimental Section**.

3. Evaluation of innovative tools for DDA data processing

For the evaluation of DDA data processing tools, we used five replicates of HeLa Pierce™ samples acquired on an Orbitrap Eclipse Tribrid instrument using a Top20 strategy. For all the data processing approaches, samples were processed independently and only identification results were considered, as the quantification is not amenable for some of the tools.

A. Evaluation of rescoring algorithms for peptides/proteins identification

Classical identification search engines try to match experimental tandem mass spectra to theoretical ones *in silico* calculated for all sequences contained in the searched database. However, they generally do not consider other features including the fragment ion intensities or the expected retention time, to calculate the scores used for FDR discrimination. As the only way to have access to this information before was the use of spectral libraries, major improvements in machine learning and publication of high-quality datasets have enabled the development of multiple rescoring tools. Fragment ion intensities predictions from deep learning algorithms today allow the generation of *in-silico* spectral libraries that can be used to perform rescoring.

i) MS²Rescore: an open-source algorithm for peptide spectrum matches rescoring

Different groups are developing multiple algorithms to bring out the best of the tens of thousands of MS/MS data acquired during a run. In the frame of a collaboration with the CompOmics group (Ghent, Belgium), we used MS²Rescore to increase the number of identifications after a Mascot search with validation based on the score in Proline. MS²Rescore performs PSMs rescoring by combining predicted spectra properties using MS²PIP⁹⁹ for MS/MS peak intensities predictions and DeepLC⁹⁷, a retention time predictor for peptides that uses deep learning. All features from these two algorithms (more than 80 features in total) were aggregated to features from the Mascot search engine in Percolator^{102, 560} to discriminate between target and decoy PSMs.

The results before and after rescoring of Mascot identifications are presented in **Figure 59**. After rescoring, the number of identified protein groups rises by 21%, and peptides and PSMs by around 25% at 1% FDR validation, for a total of around 3,900 protein groups, 27,000 peptides, and 30,800 PSMs detected using the MS²Rescore workflow. Two main reasons explain this huge increase in identifications. First, Mascot data were validated using a classical score-based target-decoy strategy without Percolator, which would have slightly improved the number of identifications. Second, MS²Rescore compares experimental PSMs to their theoretical counterpart on multiple features, including retention time and fragment ion intensities predictions. It strongly improves the confidence compared to the Mascot score, based only on the probability that the observed match between the experimental data and the theoretical sequence is a random event.

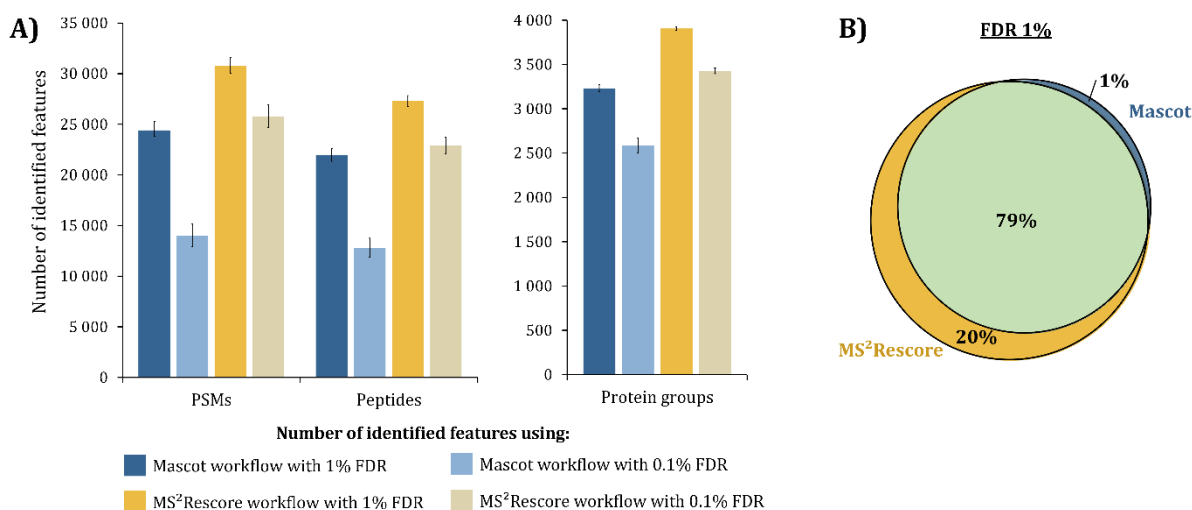


Figure 59: Results obtained after data processing using Mascot and MS²Rescore workflows. (A) Number of identified PSMs, peptides and protein groups for Mascot (blue) and MS²Rescore workflows (yellow) at 1% FDR (dark colors) and 0.1% FDR (light colors). (B) Venn diagram representing the overlap between peptides identified at 1% FDR with Mascot and MS²Rescore.

The benefits of MS²Rescore are even higher when the FDR cutoff is lowered to 0.1% (**Figure 59.A**). Higher numbers of identifications are still achieved compared to the Mascot workflow at 1% FDR, demonstrating the power of the algorithm. Conversely, a 0.1% FDR threshold applied to Mascot results validated in Proline leads to a significant drop in identifications, with 40% fewer PSMs and peptides and 20% fewer protein groups. The drop is limited to 15% for peptides and PSMs and 10% for protein groups with MS²Rescore.

Thanks to numerous features, MS²Rescore can better discriminate between targets and decoys, as presented in **Figure 60**. It illustrates that a better separation is performed, reducing the risk of false positives and increasing the number of PSMs passing the FDR cutoff.

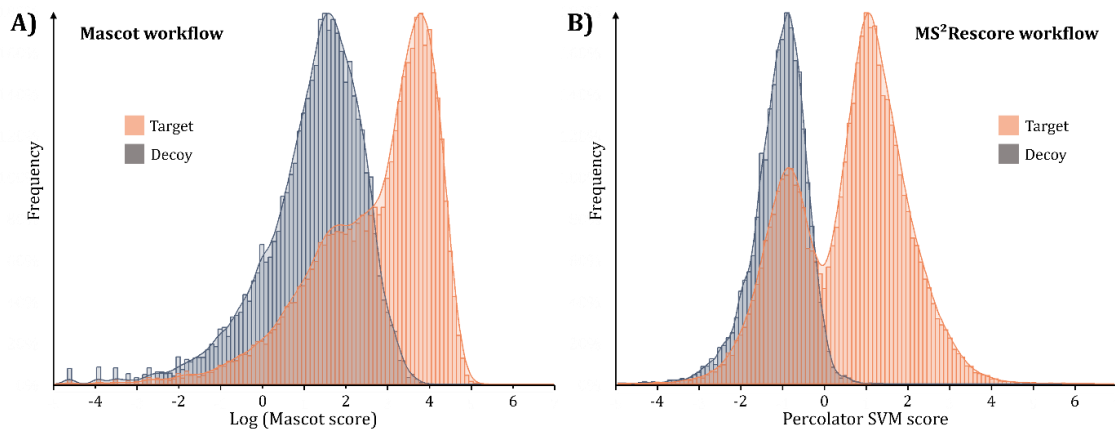


Figure 60: Target and decoy PSMs distribution. (A) Distribution of the target (orange) and decoy (grey) PSMs as a function of the log(Mascot score) for the Mascot workflow. (B) Distribution of the target (orange) and decoy (grey) PSMs as a function of the Percolator SVM score for the MS²Rescore workflow.

In addition, identified peptides show a very high overlap, with 79% of the peptides common to both workflows (**Figure 59.B**), as expected because the input for MS²Rescore is based on the output of the Mascot search.

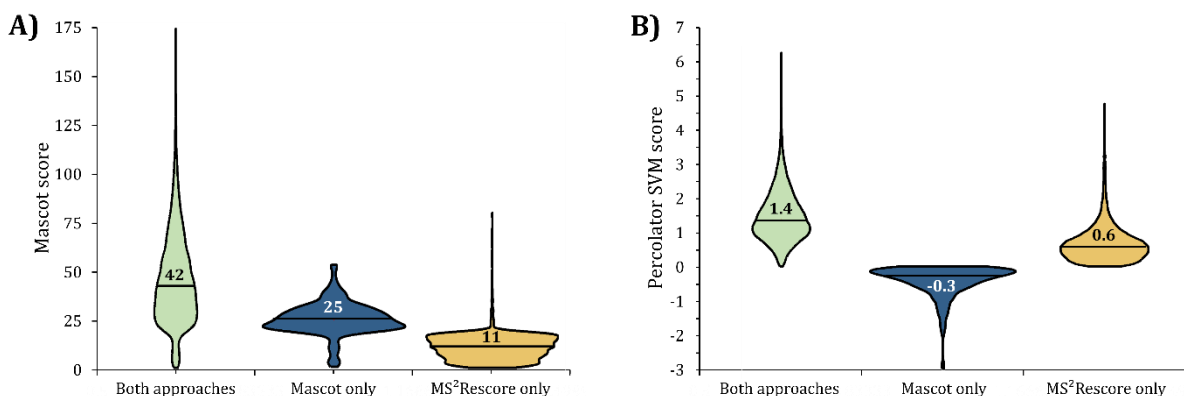


Figure 61: Characteristics of PSMs identified in both workflows (green), Mascot workflow only (blue) or MS²Rescore workflow only (yellow). (A) Violin plots representing the distribution of the Mascot scores. (B) Violin plot representing the distribution of the Percolator SVM scores. Horizontal black lines represent the median values.

While PSMs identified in both workflows present high scores with a median score of 42, PSMs unique to one workflow display a completely different distribution having lower median scores of 25 and 11 for Mascot and MS²Rescore workflows, respectively (**Figure 61.A**). This observation is consistent with the hypothesis that high-score PSMs are more likely expected to be detected by both workflows, as the

score reflects the confidence of the identification. Lowest scores are obtained for PSMs only identified in MS²Rescore, in accordance to their rejection during the validation in Proline, based on the score. Then, we looked at the Percolator SVM scores obtained after rescoring by MS²Rescore: PSMs identified by both approaches demonstrate similar distributions as the ones identified by MS²Rescore only, with median scores of 1.4 and 0.6, respectively. In opposition, PSMs identified in Mascot only keep low scores (median score of -0.3) and present an opposite distribution (**Figure 61.B**). The SVM Percolator scores distribution underlines the potential of MS²Rescore to increase confidence in identifications and remove potentially bad hits. Score refinement using additional features enables us to either strengthen the reliance on the Mascot identifications or remove low score matches that are not corroborated by the retention time and fragment ion intensities predictions. In the end, MS²Rescore increases the number of identifications and the confidence in the matches, thanks to additional features combined with finer scoring functions. Theoretically, output from MS²Rescore can be combined to a quantification algorithm, such as FlashLFQ⁵⁶¹ for example, to perform quantification of the peptides and proteins identified.

ii) *Inferys rescoring for Sequest HT data*

Inferys¹²⁰ is another rescoring algorithm developed by MSAID, and it was implemented in PD 2.5, commercialized by Thermo Fisher Scientific. This algorithm rescoring Sequest search identifications, and Percolator is applied to calculate the FDR. It was compared to the classical Sequest search, followed by Percolator for FDR estimation to assess its performances (**Figure 62**). First, comparing the results at 1% FDR demonstrates a minor gain of identified features, with an increase of the identified features only between 2% to 3%, with a total of more than 33,000 PSMs accounting for 28,500 peptides and 3,500 protein groups.

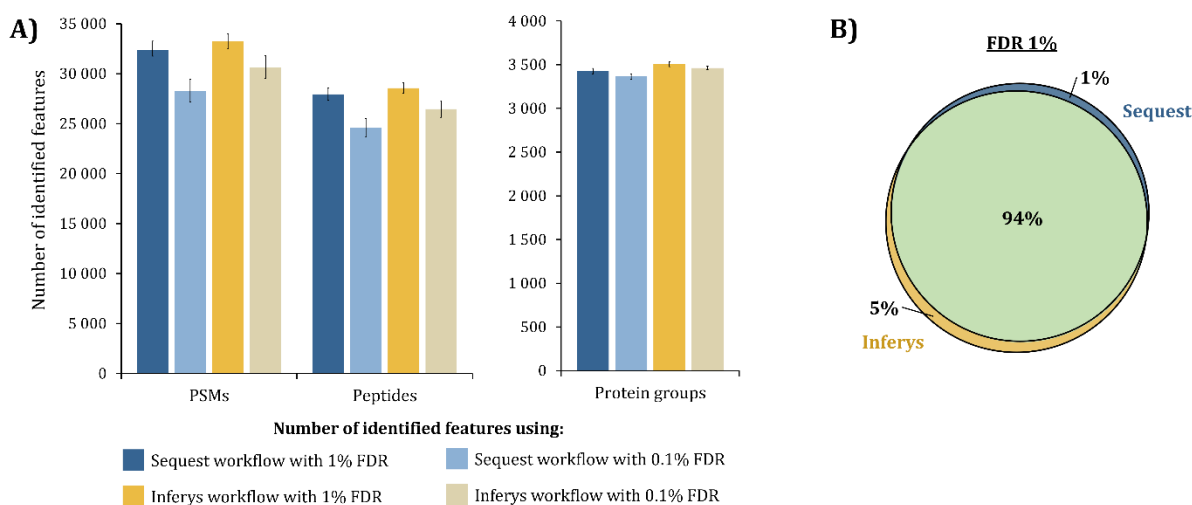


Figure 62: Results obtained after data processing using Sequest and Inferys workflows. (A) Number of identified PSMs, peptides and protein groups for Sequest (blue) and Inferys workflows (yellow) at 1% FDR (dark colors) and 0.1% FDR (light colors). (B) Venn diagram representing the overlap between peptides identified at 1% FDR with Sequest and Inferys.

Second, reducing the FDR threshold to 0.1% logically leads to a decrease in identifications. While the number of protein groups remains almost the same for both workflows (less than a 2% decrease compared to 1% FDR), the number of peptides and PSMs was reduced by 13% for Sequest and 8% for Inferys (**Figure 62.A**). In the end, Inferys identified more than 30,000 PSMs and 26,400 peptides from more than 3,450 protein groups at 0.1% FDR. For both FDR thresholds, the slight gain observed is explained by the minor differences between the workflows. FDR estimation of Sequest search results is already performed by Percolator, which considers multiple features for the score calculation. With

Inferys, only a few pieces of information regarding the fragment ion intensities are added, with a limited influence on the global score and, consequently, not enabling to increase the number of identified features drastically. Inferys algorithm does not use retention time predictions (used in MS²Rescore, for example), which would have helped to rescore PSMs and increase the identification numbers. Nevertheless, the few additional features added by Inferys identified up to 9% peptides and PSMs at 0.1% FDR.

The comparison of the peptides identified using a 1% FDR reveals that 94% are identified by both strategies (**Figure 62.B**). As previously explained for MS²Rescore and Mascot, Inferys uses the Sequest search results without filtering as input. Consequently, we expect the results to be similar to the hits from the same search. As mentioned above, Inferys rescoring benefits for identifications are limited because of the few differences in the features used during the FDR calculation step with Percolator. Nevertheless, as illustrated in **Figure 63**, it still allows for better discrimination between target and decoy PSMs.

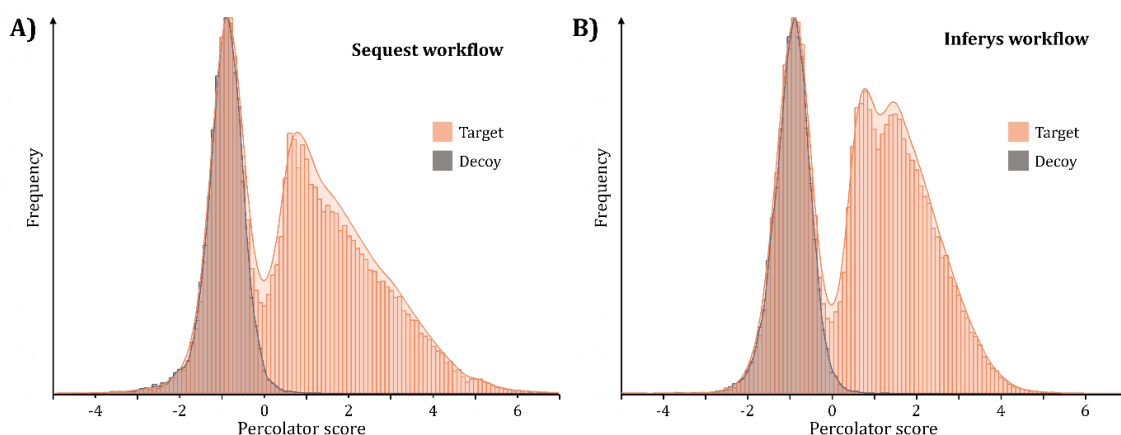


Figure 63: Target and decoy PSMs distribution. (A) Distribution of the target (orange) and decoy (grey) PSMs as a function of the Percolator score for the Sequest workflow. (B) Distribution of the target (orange) and decoy (grey) PSMs as a function of the Percolator score for the Inferys workflow.

In the PD 2.5 output, multiple features are available. We selected three of them to characterize PSMs depending if they are identified by both workflows or only one or the other. First, the study of the isolation interference – referring to the relative amount of ion current within the isolation window that is not attributed to the precursor itself, in percent – reveals that PSMs identified by both approaches have a lower isolation interference (median value of 38%) compared to those detected only with Sequest or Inferys workflows only (median values of 67% and 61%, respectively, **Figure 64.A**). Similarly, the same trend is observed for the Sequest score distribution (i.e., score attributed by the Sequest search engine only), with comparable distribution and median scores, 1.2 and 1.1, for PSMs identified only by Sequest or Inferys, respectively (**Figure 64.B**). In comparison, PSMs identified in both approaches present a median score of 2.5. In short, the study of these two parameters reveals that PSMs identified by only one approach are low-quality spectra intricate to validate using a classical search engine and that are potentially false positives. Hence, we looked at the spectral angle⁵⁶², reflecting the spectral similarity between experimental and theoretical intensities of fragment ions (**Figure 64.C**). Contrary to what was observed for the two previous features, PSMs identified by both approaches and by Inferys only depict similar trends, with relatively high median spectral angles of 0.87 and 0.63, respectively. In opposition, PSMs identified only with Sequest have low spectral angle values (median of 0.25). It underlines those fragment ion intensities predictions, and their comparison with the experimental spectra enables us to differentiate between PSMs having similar properties using the isolation interference and the Sequest score. Hence, the Percolator score of the PSMs with

good spectral correlation will be increased, allowing them to pass the threshold, while those with bad spectral angles will have their score reduced and then not be identified after the FDR cutoff.

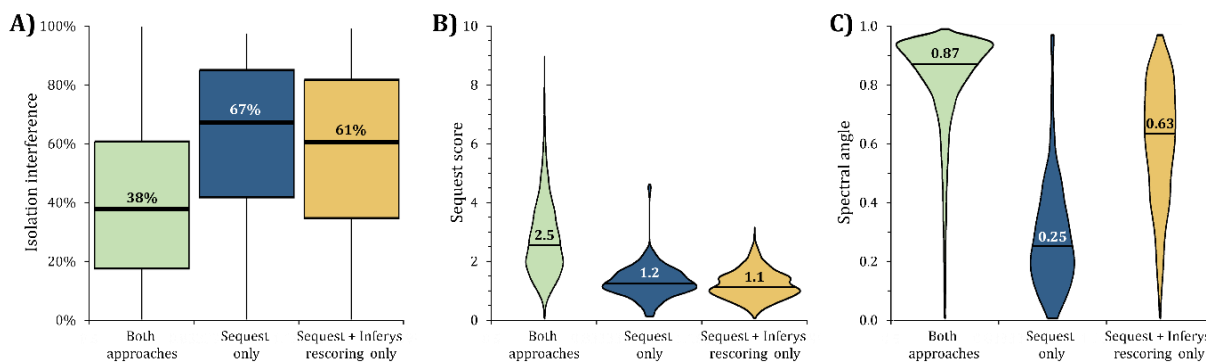


Figure 64: Characteristics of PSMs identified in both workflows (green), Mascot workflow only (blue) or MS²Rescore workflow only (yellow). (A) Boxplot of the distribution of the isolation interference. (B) Violin plot representing the distribution of the Sequest scores. (C) Violin plot representing the distribution of the spectral angles. Horizontal black lines represent the median values.

Briefly, Inferys does not drastically increase the number of identifications compared to Sequest, followed by Percolator. Indeed, Percolator is already a powerful semi-supervised machine learning algorithm that uses twenty features to discriminate target from decoy PSMs^{102, 560}. Nevertheless, it increases the confidence in the identified PSMs thanks to considering the similarities between experimental and predicted spectra. In addition, quantification can be directly done in the PD 2.5 using the quantification nodes provided. Nevertheless, Inferys prediction model presents limitations regarding the data that can be processed. Only a few PTMs can be searched, namely carbamidomethylation on cysteine residues, TMT/TMTpro-labeled peptides, and oxidized methionines, even if models are currently under development for the analysis of other common modifications, such as phosphorylation, for example. Similarly, only CID and HCD data can be processed with this algorithm, reducing its potential when looking for specific PTMs.

B. Chimerys, a new artificial intelligence-based search engine

To take advantage of the ion intensities predictions from the Inferys algorithm, MSAID developed an artificial intelligence-based search engine capable of identifying multiple PSMs per spectrum. First, it runs an initial coarse search. It uses the Inferys algorithm to perform recalibration and refinement of the prediction model to maximize the prediction accuracy⁵⁶³. Then, MS/MS spectra are analyzed without any pre-processing or precursor selection from MS spectra. All candidates in the isolation window of a given MS/MS spectrum are searched simultaneously and compete for measured fragment ion intensity in a concerted step. Chimerys aims to explain as many measured intensities with as few candidate peptides as possible by doing the deconvolution of chimeric spectra. Finally, the FDR control is performed using Percolator. This strategy enables the identification of co-isolated precursors. Indeed, from a previous analysis, it has been found that around 45% of the MS/MS spectra acquired display an isolation interference higher than 50%, i.e., that the selected precursor intensity accounts for less than half of the total intensity in the isolation window. Moreover, previous work demonstrated that 50% of all spectra contain more than one peptide⁵⁶⁴.

We evaluated this software using the same five raw files used for previous data analysis in this chapter. Chimerys identified 66% and 54% more protein groups and peptides, respectively, compared to the Inferys workflow and increased by 2.6 times the number of PSMs identified (**Figure 65.A**). At 1% FDR, more than 86,000 PSMs were detected, leading to 44,000 peptides and 5,800 protein groups.

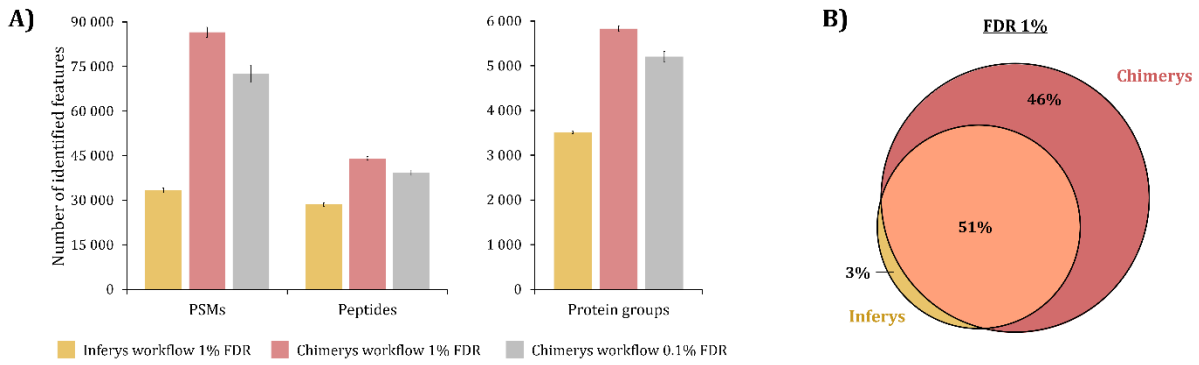


Figure 65: Results obtained after data processing using Inferys and Chimerys workflows. (A) Number of identified PSMs, peptides and protein groups for Inferys at 1% FDR (yellow), Chimerys at 1% FDR (red) and Chimerys at 0.1% FDR (grey). (B) Venn diagram representing the overlap between peptides identified at 1% FDR with Inferys and Chimerys.

By reducing the FDR cutoff to 0.1%, the contribution of Chimerys remains significant as around 85% of the identified features are still below the validation threshold, for a total of 5,200 protein groups, 39,000 peptides, and 72,000 PSMs detected (**Figure 65.A**), outperforming Inferys results. It demonstrates the strength of Chimerys to increase the number of identifications drastically. The results from Inferys and Chimerys reveal a huge overlap of the detected features, with up to 95% of the peptides identified using Inferys also detected using Chimerys (**Figure 65.B**).

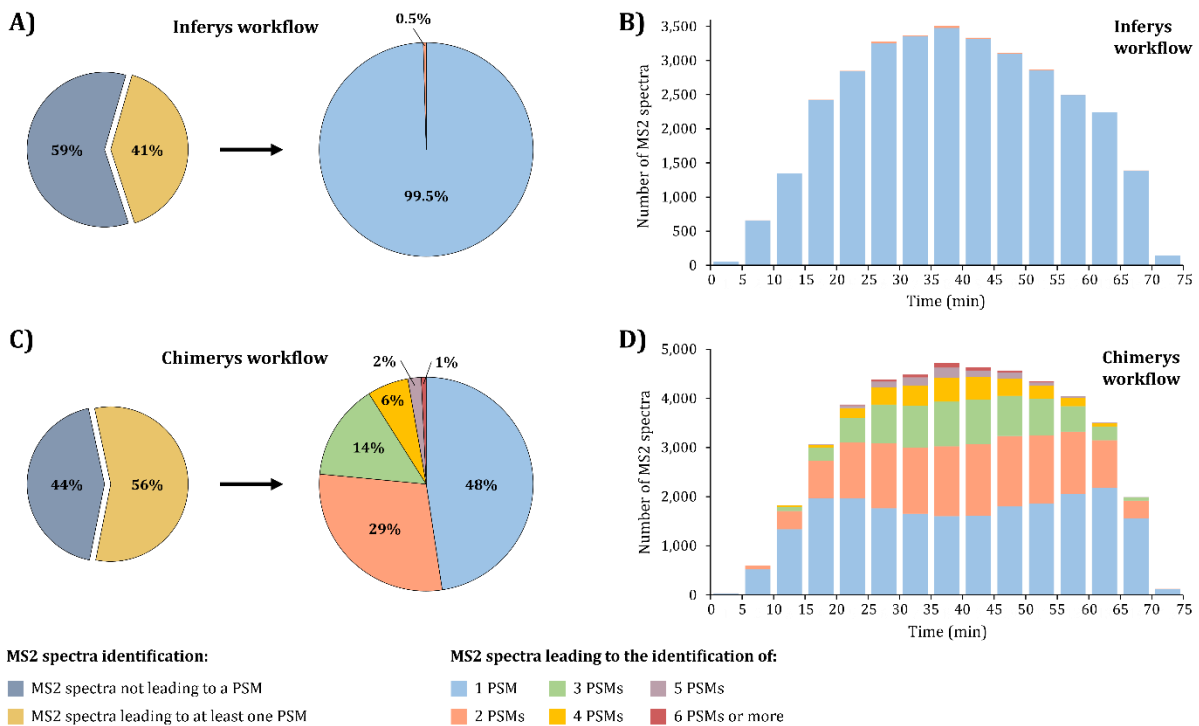


Figure 66: Number of MS/MS spectra leading to a PSM. (A) Pie charts of the number of MS/MS spectra leading or not to at least one PSM (left) and number of PSMs identified per MS/MS spectra (right) for the Inferys workflow. (B) Distribution of the number of MS/MS spectra leading at least one PSM over the gradient time for Inferys data analysis. (C) Pie charts of the number of MS/MS spectra leading or not to at least one PSM (left) and number of PSMs identified per MS/MS spectra (right) for the Chimerys workflow. (D) Distribution of the number of MS/MS spectra leading at least one PSM over the gradient time for Chimerys data analysis.

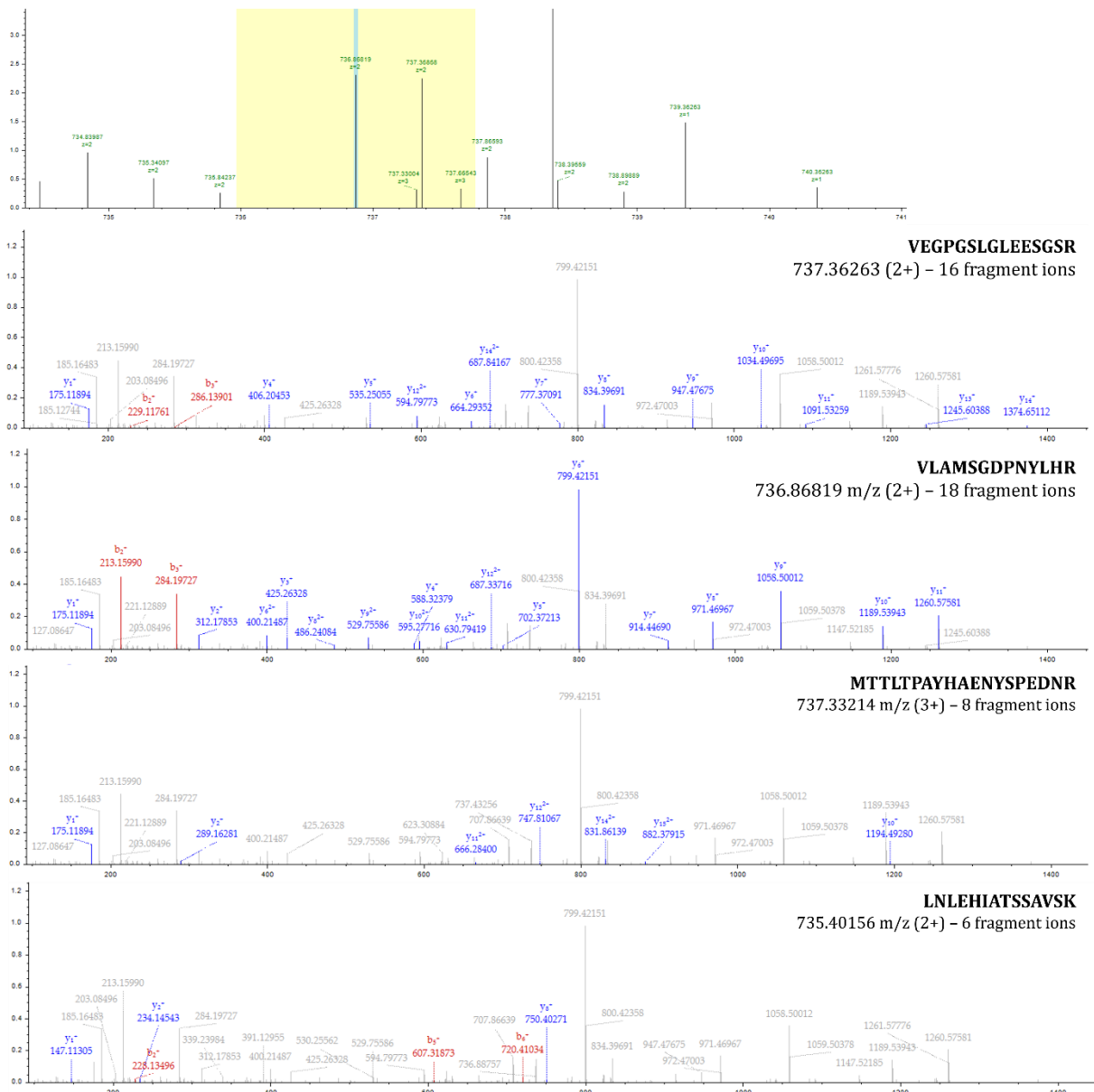


Figure 67: MS/MS spectra leading to the identification of five peptides (bottom) and its corresponding MS spectrum (top). Attribution of the fragment ions from five different peptides on a unique MS/MS spectrum from the precursor selected in the top panel.

Moreover, the number of PSMs attributed to each spectrum (**Figure 66**) highlights the importance of being able to analyze chimeric spectra. For the Inferys workflow, 41% of the MS/MS spectra lead to at least one PSM, against 56% using Chimerys. In detail, almost all (99.5%) of the MS/MS spectra leading to a PSM are matched to only one peptide with Inferys. With Chimerys, more than half of the MS/MS spectra lead to more than one PSM, with up to eight peptides identified in a unique spectrum. We plot the number of MS/MS spectra leading to PSMs as a function of the gradient (**Figure 66.D**). As expected, the highest numbers of PSMs per spectrum are obtained in the middle of the gradient, when co-elution and co-isolation are more probable. Thanks to its ability to deconvolute chimeric spectra, Chimerys can identify multiple peptides with high confidence from the same MS/MS spectrum. For example, **Figure 67** shows the MS spectra, with the precursor ion and the isolation window, and the MS/MS spectra, leading to the confident identification (FDR < 1%) of four peptides, with multiple fragment ions corroborating the identification of the peptides.

In essence, Chimerys drastically increases the identified features, leading to more than 5,800 protein groups and 44,000 peptides detected in a one-hour gradient. It uses machine learning-based prediction of peptide properties (e.g., retention time, fragment ion intensities) combined with the deconvolution of the MS/MS spectra to identify a maximum number of peptides in each. Thanks to the deconvolution algorithm, the presence of the precursor ion is not required anymore to identify a peptide. Indeed, all peptides theoretically included in the isolation window, or even close to the isolation window, will be searched against the MS/MS spectra to identify fragment ions. It can result on the identification of peptides out of the isolation window and without precursor signal (e.g., peptide LNLEHIATSSAVSK, **Figure 67, bottom panel**). Even if this peptide pass the FDR threshold, there is a risk of over interpretation of the MS/MS spectrum, potentially leading to false positives.

To summarize, wide windows acquisition schemes can be used, to increase the potential number of peptides detected, as already shown in few studies^{565, 566}. However, peptides identified without detecting their precursor ion will not be quantified, as the quantification is still performed at the MS1 level. Indeed, for more than half of the peptides identified, no intensity value of the precursor can be detected, limiting its use for quantitative proteomics. Moreover, as it uses the prediction model from Inferys, Chimerys is limited in the number of PTMs that can be searched, and only CID and HCD data can be processed.

C. Evaluation of ionbot: an open modification search algorithm

The numerous algorithms developed demonstrate their ability to increase identifications. However, they present several limiting factors, including restrictions on the fragmentation methods used and the PTMs available for the search, but also from the initial search performed by classical search engines. In the frame of the collaboration with CompOmics, we tested ionbot, an open modification search engine incorporating all relevant matching information to calculate machine learning-based scores. It uses the previously mentioned algorithms developed in CompOmics, DeepLC and MS²PIP, for retention time and fragment intensities predictions, respectively. Thanks to machine learning from multiple datasets, these algorithms can predict peptide properties for any sequence provided, no matter the PTMs potentially present.

Ionbot was compared to the MS²Rescore workflow, which uses the same algorithms for rescoring (**Figure 68.A**). Ionbot identified, at 1% FDR, 14% more protein groups, 10% more peptide groups (i.e., same raw sequence), and 50% more PSMs. In order to have comparable results, we used the peptide sequence without any modification as the number of peptide groups identified, explaining why the increase in peptide groups is much lower than the one observed for PSMs. At 0.1% FDR, ionbot still detects 25% more PSMs than MS²Rescore at 1% FDR, while the number of peptides and PSMs are similar. The comparison of the peptide sequences identified shows an overlap of 66% (**Figure 68.B**).

On all the 30,000 peptides identified by ionbot, 5,830 are bearing an unexpected modification, i.e., another modification than cysteine carbamidomethylation and methionine oxidation. Among these peptides bearing an unexpected modification, 54% also have the unmodified peptide counterpart identified, and 46% are only identified with their unexpected PTM. Nevertheless, they present median q-values of 0.0003, allowing us to be confident in the hits identified. In addition, ionbot enables the detection of additional peptides that are detected only under a modified form. Indeed, for 89% of the peptides with unexpected PTMs and identified only by ionbot, no unmodified counterpart was detected. This proportion drops to 7% for peptides identified by ionbot and MS²Rescore.

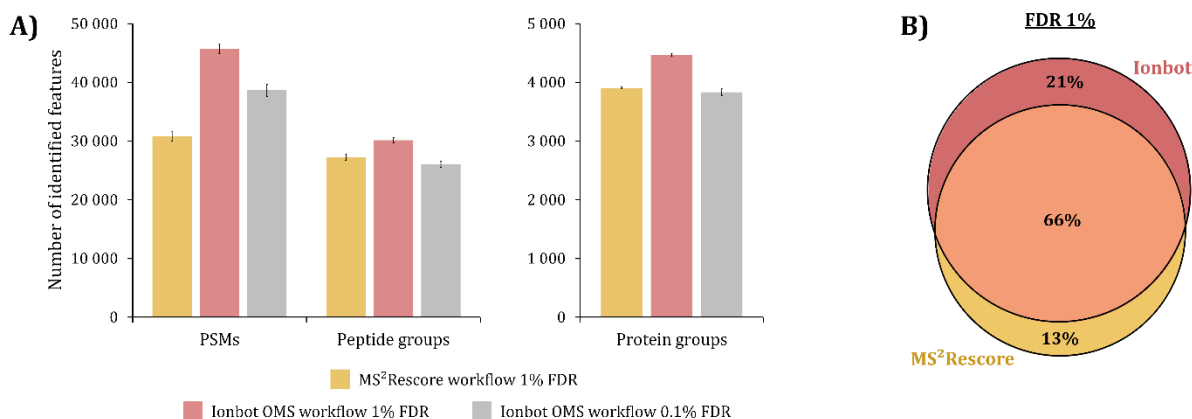


Figure 68: Results obtained after data processing using MS²Rescore and ionbot workflows. (A) Number of identified PSMs, peptides and protein groups for MS²Rescore at 1% FDR (yellow), ionbot at 1% FDR (red) and ionbot at 0.1% FDR (grey). (B) Venn diagram representing the overlap between peptides identified at 1% FDR with MS²Rescore and ionbot.

Then, we investigated the PTMs detected by ionbot to determine the relevance of the results obtained. More than 200 modifications were found, but most of them are only identified in one or a few PSMs. **Figure 69** represents the number of PSMs identified for the ten most frequent modifications identified. Among them, we noticed the cysteine carbamidomethylation and oxidation, usually added to classical searches. Other PTMs identified are mainly chemical artifacts coming from the sample preparation. A quick description of the ten most abundant PTMs is provided in **Table 7**. To confirm the presence of these unexpected, we performed a Mascot search by setting carbamylation as variable modification. After validation, more than 85% of the peptides bearing a carbamylation identified with ionbot were also identified by Mascot, corroborating that these matches are not false positives. A comparable results has been found for mesityloxyde, but in lower extent (60%).

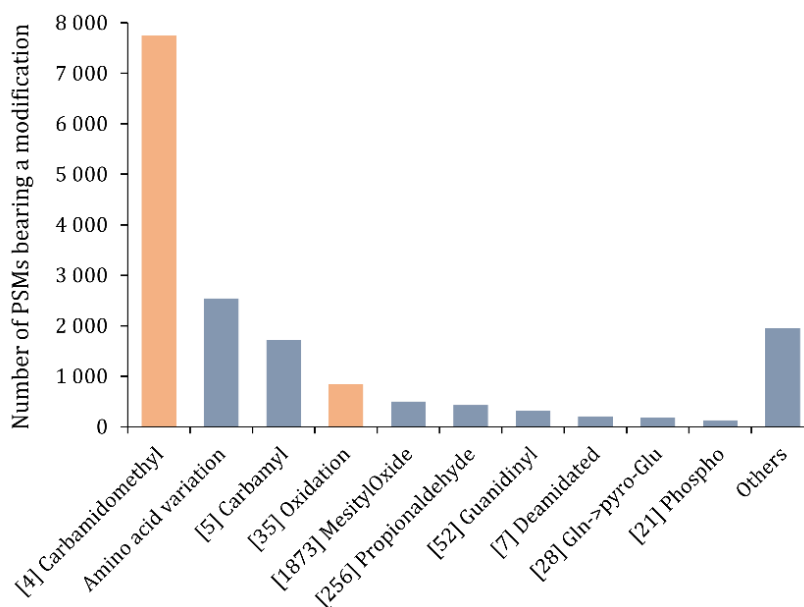


Figure 69: Number of PSMs bearing one of the ten most frequently identified modifications. Unimod numbers are between square brackets. Orange bars represent the modifications that we look at with classical search engines.

Table 7: List of the ten most frequently found modifications by ionbot, their Unimod identifier, their mass shift and a brief description of the modification.

Name	Unimod	Mass shift	Description	Ref.
Carbamidomethylation	4	+57.02 Da	Iodoacetamide derivative, from the alkylation process.	567
Amino acid variation	/	/	Induced by single nucleotide variation in the protein-coding region of the genome.	568
Carbamylation	5	+43.01 Da	Non-enzymatic PTM related to protein ageing. It can be potentially introduced during sample preparation with urea.	569
Oxidation	35	+15.99 Da	Methionine oxidation from reactive species (oxidative stress) or sample preparation, commonly observed PTM.	570-572
Mesityl oxide	1873	+98.07 Da	Acetone chemical artefact, from acetone use for precipitation step during sample preparation.	573
Propionaldehyde	256	+40.03 Da	Acetone chemical artefact, from acetone use for precipitation step during sample preparation.	574
Guanidyl	52	+42.02 Da	Potentially coming from guanidination of lysine by O-methylisourea.	575
Deamidation	7	+0.98 Da	Nonenzymatic and spontaneous PTM, associated with proteome ageing and degenerative disease.	576, 577
Gln → pyro-Glu	28	-17.03 Da	Spontaneous rearrangement of N-term glutamine into pyroglutamate, frequently found in many proteins.	578, 579
Phosphorylation	21	+79.97 Da	Common PTM, implicated in multiple processes at protein level.	580

To conclude, ionbot uses machine learning-based predictions to search all possible PTMs. It allows us to increase the number of identifications and detect modified peptide forms that we are not looking at when performing classical searches. The peptides identified with unexpected modifications appear as confident matches as most are kept, even at 0.1% FDR. This tool is interesting for discovery analysis in shotgun proteomics. It allows screening of all possible PTMs, and from that, the identification of some modifications that can be overexpressed in some specific samples. In addition, the team developing this software is currently working on the implementation of quantification from the ionbot output to have a quantitative overview of these modified peptides.

4. Evaluation of library-free approaches in DIA

The past and current developments of new algorithms make library-free approaches very powerful for DIA data analysis. In this context, we compared the performances of two bioinformatics tools working in a library-free mode, DirectDIA™ (v.17.5), a licensed algorithm from Spectronaut, and DIA-NN¹⁸³, an open-source algorithm. They both use artificial intelligence-based algorithms to maximize the number of identifications, and the algorithm sets many parameters automatically after a first coarse search. Hence, we evaluated two key parameters of these algorithms: the MS and MS/MS tolerances and the retention time extraction window. Results using optimal parameters found by the algorithm were compared to results with user-defined parameters. For this purpose, the optimized method for analyzing a total lysate of *Arabidopsis thaliana* spiked with 48 UPS1 proteins at different concentrations (see **Part I, Chapter 1**) was used to evaluate the identification results and the quantification accuracy. Only precursor ions that satisfied a q-value < 0.01 were kept. Finally, peptide areas were obtained by summing precursor ion areas and protein areas by summing peptide areas.

First, we looked at the background proteome results obtained using the default parameters (i.e., automatically set mass tolerances and retention time extraction windows), fixed MS tolerances (5 ppm for MS and 20 ppm for MS/MS), and fixed retention time extraction windows (10 min for DirectDIA™

and scan window radius to 5). The average number of peptides and proteins identified for each data processing strategy is presented in **Figure 70**. Using default parameters, DIA-NN demonstrates better performances at the protein level, with around 10% more protein groups identified, whereas DirectDIA™ detected an average of 30% more peptides. The standard deviation observed is also slightly higher for DIA-NN results compared to DirectDIA™.

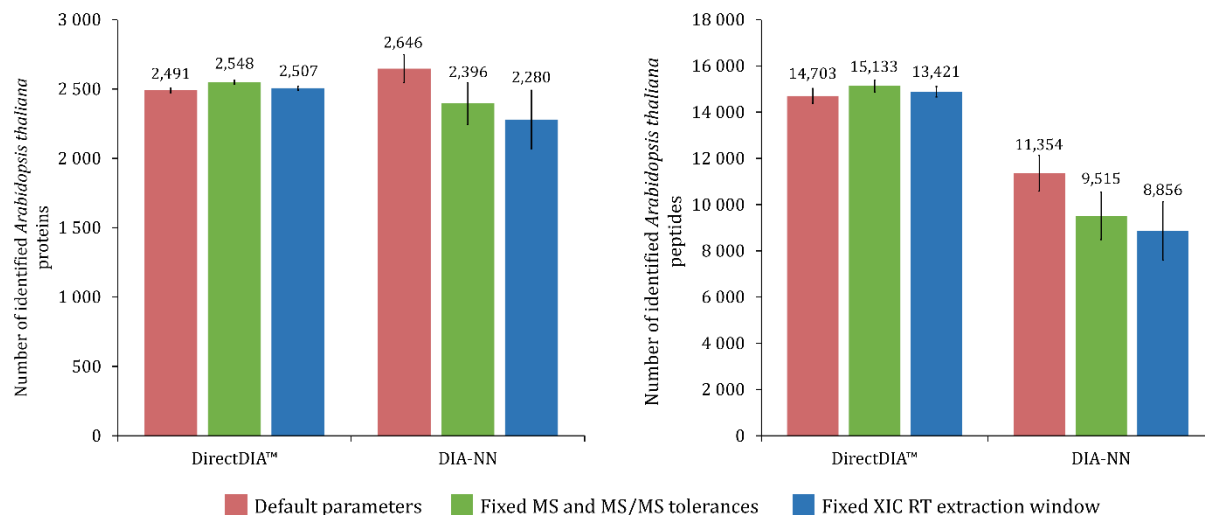


Figure 70: Results obtained with the DirectDIA™ and DIA-NN software. Number of identified *Arabidopsis thaliana* proteins and peptides with the default parameters (red), fixed MS tolerances (green) and fixed RT extraction windows (blue).

Quantification using DIA-NN shows lower CVs, with a median CV of 5.5% against 9.0% using DirectDIA™, and consequently a higher number of proteins with a CV <20% (2,106 vs. 1,917, **Figure 71**). When MS tolerances and retention time extraction windows are fixed manually, two different trends are observed. For DirectDIA™, it resulted in a similar number of peptide and protein group identifications, with even slightly better results using fixed MS tolerances compared to fixed RT extraction window and default parameters. For DIA-NN, manually set parameters compared to default parameters led to a drop of identifications: 10% and 14% less protein group and peptide identifications, respectively, using fixed MS tolerances, and 17% and 28%, respectively, using fixed RT extraction windows. Regarding quantification, between 73% and 80% of the identified proteins were quantified with a CV <20% using DIA-NN, and quantified proteins demonstrate low CVs, around 5 to 6%, no matter the strategy employed. For DirectDIA™, 76% of the identified protein groups are conserved after CV <20% filtering for fixed MS tolerances, comparable to default parameters, while 70% of the protein groups pass the filter for fixed RT extraction windows. The distribution of the CVs corroborates these results, with a slightly higher median CV (9.5%) and a larger dispersion. To summarize, DirectDIA™ generally identified and quantified a higher number of proteins and peptides compared to DIA-NN. Only the default parameters of DIA-NN allow us to obtain better results than DirectDIA™ regarding identified and quantified protein groups. Nevertheless, lower CVs are obtained using DIA-NN, increasing the confidence in the quantification.

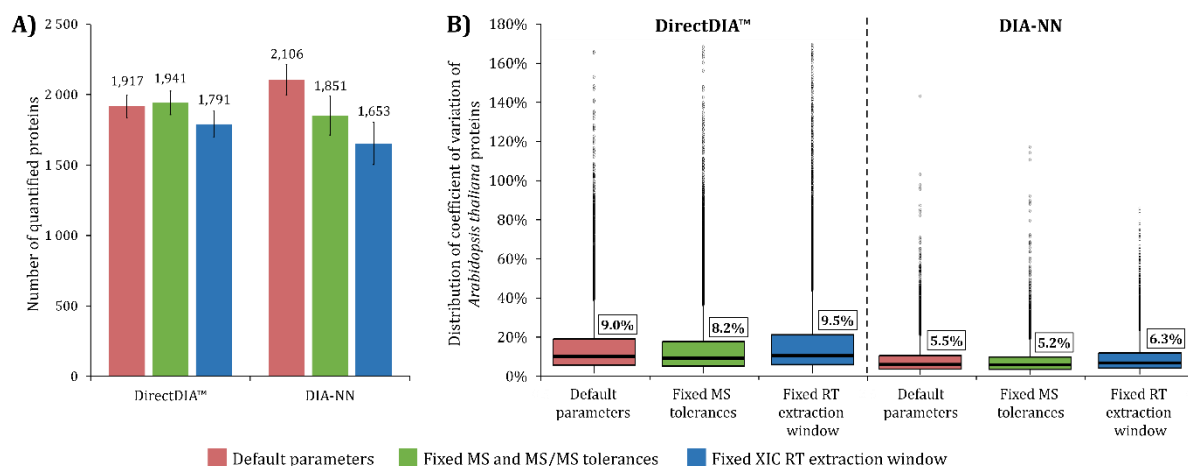


Figure 71: *Arabidopsis thaliana* quantification results using DirectDIA™ and DIA-NN. (A) Number of quantified proteins using DirectDIA™ and DIA-NN with default parameters (red), fixed MS tolerances (green) and fixed retention time extraction windows (blue). (B) Box plot of the coefficient of variations for each results.

To go deeper into the quantification assessment with these two algorithms, we looked at the UPS1 proteins spiked in calibrated quantities. The number of UPS1 proteins quantified in all replicates and quantified with a CV <20% are shown in **Figure 72.A** for DirectDIA™ and **Figure 72.B** for DIA-NN. Taking the results as a whole, DIA-NN shows gaps in the quantification of low-abundant proteins, with around 10 proteins quantified at 25 amol and between 14 and 25 at 125 amol, depending on the strategy. In comparison, at least 90% of the 48 UPS1 proteins expected are quantified in all the replicates regardless of the concentration and the parameters used with DirectDIA™.

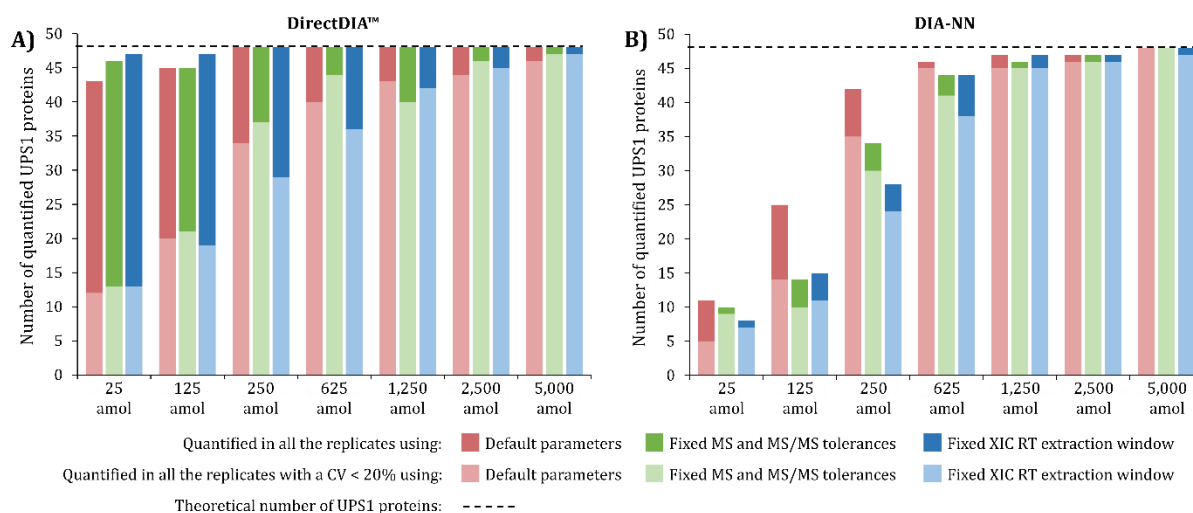


Figure 72: Quantification results obtained for UPS1 proteins with DirectDIA™ and DIA-NN. Number of quantified UPS1 proteins in all the replicates (dark colors) and with a CV <20% (light colors) using default parameters (red), fixed MS tolerances (green) and fixed RT extraction windows (blue) for (A) DirectDIA™ and (B) DIA-NN.

In detail, working with default parameters in DIA-NN allows us to quantify more UPS1 proteins at low concentration levels (i.e., 25 to 250 amol) compared to searches with fixed parameters. Filtering out of the proteins with a CV >20% gives the same trend, except for 25 amol. From 625 amol, all search strategies used in DIA-NN present similar results, with the quantification of almost all proteins expected with a CV <20%. In contrast, DirectDIA™ quantified almost all UPS1 proteins, even at low

concentrations. However, CV filtering removes an important part of the quantified proteins, showing a lack of quantification reproducibility for low concentrations. From 250 amol, all the proteins are quantified whatever the search parameters. Even if, results obtained from fixed MS tolerances lead to more proteins quantified with a CV <20%, while those from fixed RT extraction windows are lower, particularly for 250 and 625 amol. DIA-NN and DirectDIA™ present similar performances for high concentrations (i.e., 1,250 amol to 5,000 amol), but DirectDIA™ is the most suitable for low quantities, with no clear evidence of higher performances using manually set tolerances.

Finally, the accuracy of UPS1 peptide quantification was also appraised. For this purpose, the experimental ratios between the highest point in the range (5,000 amol) and the other lower points were plotted against the theoretical ratios (**Figure 73**). Overall, the dispersion of the ratios around the theoretical values and the linearity range obtained with the two software packages and parameters used are similar. Indeed, a coefficient of determination greater than 0.99 was achieved up to 125 amol of UPS1 injected for all searches, demonstrating the precision of quantification, even for low signal abundance. Very low concentrations still suffer from an underestimation of their quantity, with signals too close to the background noise to be extracted correctly. Last, it is clear that the dispersion of the ratios around the theoretical values underlines the fact that quantification accuracy could be better, especially for these low-abundance peptides. It is even more noticeable for results obtained with DirectDIA™, particularly when the RT extraction window was fixed manually.

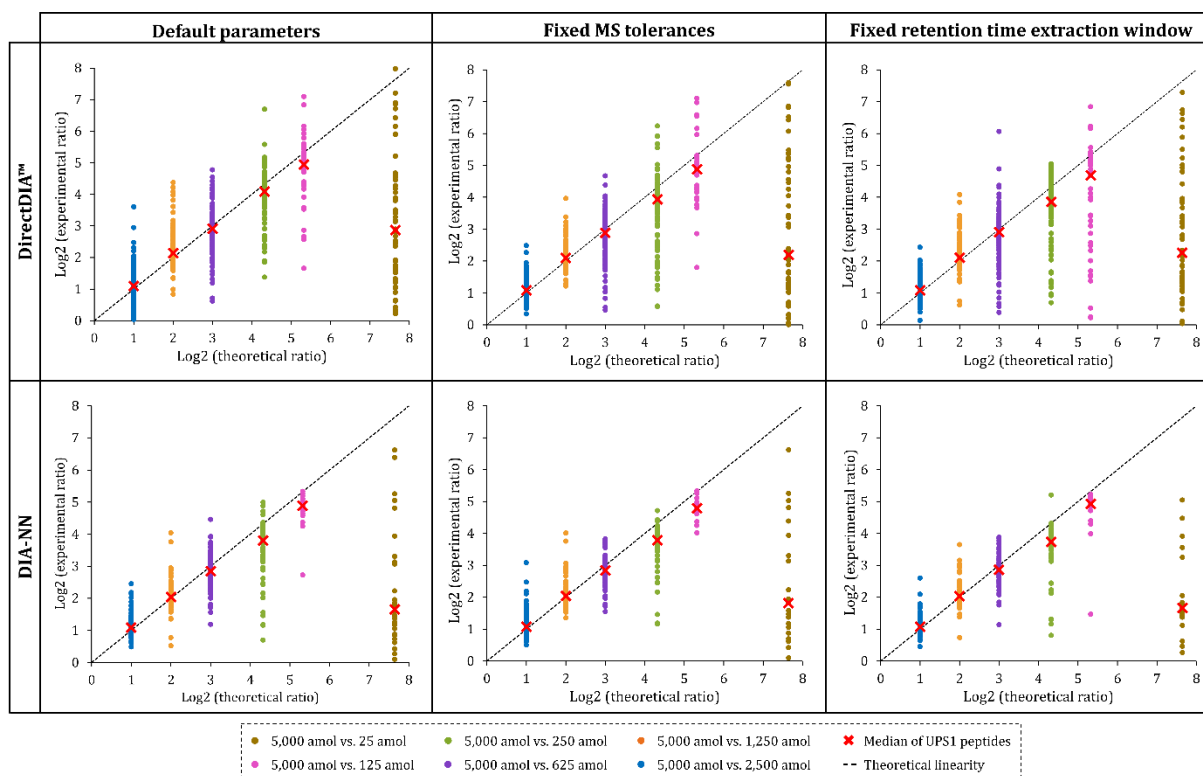


Figure 73: Evaluation of the precision of quantification of DirectDIA™ and DIA-NN on UPS1 peptides. Experimental ratios, calculated from the highest point of the range, were represented as a function of the theoretical ratios of UPS1 peptides for DirectDIA™ and DIA-NN using default parameters, fixed MS tolerances and fixed RT extraction window.

To conclude on DIA data processing using library-free algorithms, both DIA-NN and DirectDIA™ demonstrate good overall performances using the default parameters that are not improved with user-set parameters and even worsen for DIA-NN. DirectDIA™ is the solution providing the highest number of identified and quantified peptides and protein groups, including with a CV <20%, but it displays higher CVs. On the contrary, DIA-NN while not providing the maximum number of quantified features,

but the results obtained are very reproducible, with a median CV of 5%. The high coefficients of determination observed evidence of the accuracy of quantification up to 125 amol, even if imperfect, as established by the dispersion around the theoretical value. These results clarify the potential of library-free approaches to perform accurate quantification in complex mixtures, as well as their limits. Despite the energy employed by the community to develop more efficient tools, allowing to quantify higher numbers of protein, library-based approaches remain more sensitive and accurate than library-free approaches, as demonstrated in several studies^{148, 172, 181, 581}.

5. Conclusions and perspectives

This chapter assessed several new data processing algorithms and tools for improved identification and quantification of proteins in complex mixtures. The first part showed the benefits of artificial intelligence and machine learning to increase the number of identifications in DDA. Prediction models built from already published datasets or synthetic peptides analyses allow to extend the number of features considered for the scoring functions of the PSMs. Hence, after rescoring with Inferys or MS²Rescore, a more accurate score is obtained allowing to better differentiating between true and false positives. These prediction models combined to artificial intelligence also demonstrate their performances to identify multiple peptides in chimeric spectra. Moreover, once the models are trained, they can be applied for many features, including peptides bearing uncommon modifications. The consideration of these unexpected modifications drastically improves the number of PSMs detected, by identifying modified peptides. Nevertheless, if peptide and protein quantification is available after rescoring, XIC-MS1 quantification cannot be performed after the use of Chimerys for identified peptides that have no precursors detected. Regarding ionbot, quantification using FlashLFQ⁵⁶¹ after open modification is normally available but still presents some issues that need to be fixed.

The second part of the chapter highlights the high capabilities of library-free search DIA data processing in terms of identification and quantification. Default parameters appear the best to obtain optimal results. Indeed, they are calculated independently for each run after a first search, and supposed to fit best with the data. By setting some parameters manually, one induces eventual biases in the algorithms that most of the time lead to a loss of identifications. Furthermore, both software tested, DirectDIA™ and DIA-NN, present high quality and precision quantification results. Nevertheless, while DIA-NN shows lower CVs than DirectDIA™, the latter is performing better for low abundant proteins.

The colossal quantity of information acquired in a single shotgun proteomics run by state-of-the-art mass spectrometers requires the data processing step to be adapted in order to get the best from what has been acquired. This chapter underlines the potential of artificial intelligence and deep learning to tackle this challenge, using the amount of data generated to train the models. The development of innovative tools is growing fast, with the release of multiple algorithms per year for peptide properties predictions. Since the first example for retention time predictions by Petritis *et al.*^{582, 583}, artificial neural networks have been employed by many groups for the development of methods demonstrating good accuracy, such as an algorithm from Guan *et al.*¹¹⁸, DeepRT¹¹⁰ and DeepLC⁹⁷. Other algorithms have been developed for peptide fragmentation predictions, which are key features of rescoring strategies. While retention time predictions can be based on the peptides chemical composition, tools for fragmentation pattern prediction, including pDeep¹¹⁵, MS²PIP⁹⁹ or DeepMass¹¹⁷, generally require high-quality data to be trained on. In addition, some solutions (e.g., Prosit¹¹⁶, AlphaPeptDeep⁵⁸⁴) are also combining the benefits of both retention time and peptide fragment intensities to increase identifications. Deep learning can also be used to predict other properties, such as collisional cross-sections⁵⁸⁵. All these solutions can be implemented either for DDA data processing to improve the scoring functions, but also in DIA for high-quality spectral library generation. They appear very promising to take the best of the data acquired by the mass spectrometers.

Despite this, keeping a critical eye on these new tools is crucial. They are not infallible, and the biological conclusions drawn from these results should always be verified. In addition, these algorithms are developed by people from bioinformatics who apply these technologies to proteomics. On the other hand, proteomists only sometimes have the background to understand how these deep-learning solutions work. From my point of view, it is essential to include bioinformatics in their training to ensure that they will use these tools correctly and understand how they are built. Close collaborations between bioinformatics specialists and proteomists is always the best configuration to ultimately implement fit-for-purpose solutions that answer the community's needs.

PART II

**METHODOLOGICAL DEVELOPMENTS FOR
THE IDENTIFICATION AND QUANTIFICATION
OF TRACE LEVEL HOST CELL PROTEIN
IMPURITIES IN DRUG PRODUCTS**

Part II

Methodological developments for the identification and quantification of trace level host cell protein impurities in drug products

1. Context of the project

Therapeutic monoclonal antibodies (mAbs) constitute a significant biopharmaceutical market with more than 125 approved products today^{3,372}. Despite the multiple purification steps performed during the downstream process (i.e., Protein A affinity chromatography, ion exchange chromatography, hydrophobic interaction chromatography, and multiple filtration steps), low levels of HCPs may co-purify with the therapeutic protein and thus remain in the drug products (DPs). Because of their potential adverse action on patient safety or mAb stability, HCPs have been considered a critical quality attribute (CQA) in DPs, and their total amount should be minimized to trace levels and monitored by highly sensitive analytical methods^{403, 473}. Currently, enzyme-linked immunosorbent assays (ELISA) are commonly used for HCP impurities analysis for their ease-of-use and high precision⁴⁸³. However, ELISA has some limitations: certain HCP species are not targeted by the assay and thus cannot be detected, and only global quantification is possible, with neither individual HCP identification nor quantification possible^{482, 586}. In addition, developing fit-for-purpose product-specific immunoassays to measure HCPs in DP is a long, expensive, and fastidious process, as they must satisfy numerous criteria (accuracy, precision, range, linearity, etc.) to be validated⁵⁸⁷. Therefore, LC-MS/MS has emerged as a promising alternative, orthogonal and complementary method to precisely identify and individually quantify trace-level HCP impurities in final drug products, and multiple studies point out its benefits^{521, 588}. In addition, previous work performed in the lab by Dr. Gauthier Husson, Dr. Nicolas Pythoud, and Dr. Steve Hessmann demonstrated the performances of DIA compared to DDA for HCPs quantification across the downstream process⁵⁰³⁻⁵⁰⁵. Nevertheless, quantifying low-abundance HCPs in drug products remains challenging due to the extreme dynamic range observed. While the latest generation LC-MS/MS couplings reach 3 to 4 orders of magnitude, quantifying ppm-level HCPs in DPs requires six orders of dynamics. A few studies have been devoted to optimizing sample preparation^{509, 511} or LC-MS/MS steps^{516, 589}.

In this context, we focused on optimizing LC-MS/MS workflows to quantify HCPs in drug products accurately. First, considering the work already presented in **Part I, Chapter 1.4**, we assessed the implementation of the FAIMS DDA approach in the context of HCPs and compared its results to classical DDA. Second, DIA methods were finely optimized to characterize HCPs in drug products. Third, we appraised gas phase fractionation (GPF) DIA for the generation of a chromatogram library from DIA runs with very narrow isolation windows, in terms of depth of analysis compared to classical DIA approaches. All developments mentioned above were performed on the NIST mAb Reference Material (Sigma Aldrich), and the optimized methods were then applied to two FDA/EMA-approved mAbs, namely trastuzumab, and nivolumab. Finally, the potential benefits of open-modification searches for enhanced impurities detection was discussed.

2. Analytical search

The overall analytical scheme is shown in **Figure 74**.

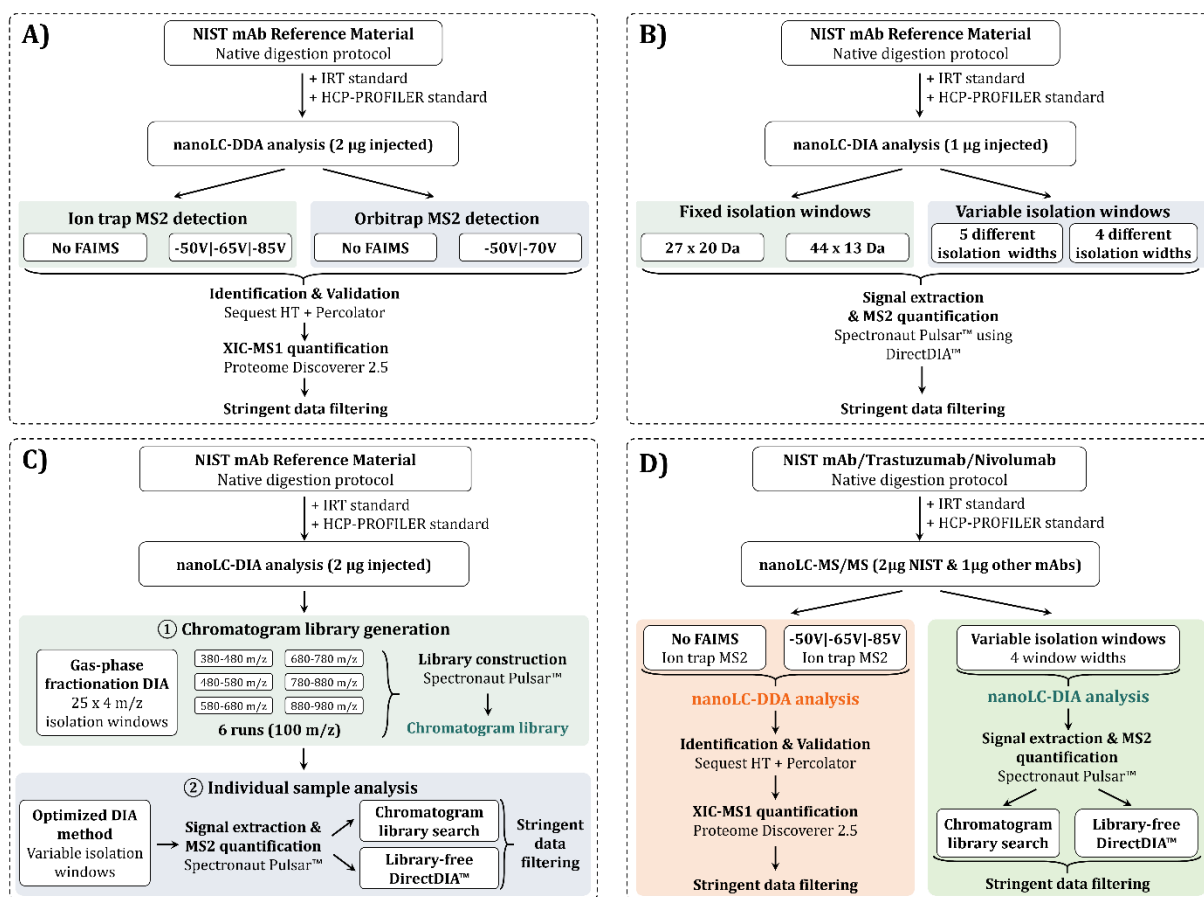


Figure 74: Analytical scheme used for nanoLC-MS/MS methods optimization on an Ultimate 3000 coupled to an Orbitrap Eclipse for HCPs profiling. NIST mAb Reference Material was prepared in triplicate following a native digestion protocol for all four workflows. Quantities injected represent the equivalent protein quantity before the mAb digestion. (A) In DDA mode with or without the FAIMS device and using ion trap (green) or Orbitrap (blue) MS2 detection for the evaluation of the acquisition methods. Identification, validation and quantification results were obtained from Sequest and Percolator, both implemented in the Proteome Discoverer 2.5 software and then filtered using stringent validation criteria. (B) In DIA mode with fixed isolation windows (green) or variable isolation windows (blue) for the evaluation of the acquisition methods. For variable isolation windows, isolation schemes using a combination of 4 and 5 different windows widths were used. Identification, validation and quantification results were obtained from Spectronaut Pulsar™ using the library-free DirectDIA™ approach and then filtered using stringent validation criteria. (C) in DIA mode with six runs of a 100 m/z mass range and 25 x 4 Da isolation windows to generate a chromatogram library using Spectronaut Pulsar™ (green). Data from DIA variable windows isolation scheme were processed using the library-free approach or the chromatogram library using Spectronaut Pulsar™ (blue) before being filtered with stringent validation criteria. (D) NIST mAb Reference Material, trastuzumab and nivolumab were prepared in triplicates following the native digestion protocol⁵⁰⁹ and analyzed on a nanoLC-Orbitrap Eclipse Tribrid coupling in both DDA (green) mode with and without the FAIMS and DIA mode (orange) using optimized methods. Identification, validation and quantification results were obtained from Sequest and Percolator, both included in the Proteome Discoverer 2.5 software for DDA and obtained from Spectronaut Pulsar™ using the either the chromatogram library-based or the library-approaches. All results were stringently validated.

For MS methods optimizations on the Orbitrap Eclipse Tribrid (Thermo Fisher Scientific), the NIST mAb Reference Material (Sigma Aldrich) was used. Samples were prepared in triplicates following the native digestion protocol⁵⁰⁹, followed by the addition of HCP-PROFILER peptides (Anaquant, Lyon, France) and iRT peptides (Biognosys, Schlieren, Switzerland). All the injected quantities indicated in this part refers to the theoretical quantities considering the absence of mAb depletion during the

native digestion, they do not represent the real quantity injected on the column. Data analysis was performed against a database containing all *Mus musculus* entries (17,050 entries, TaxID = 10090, 2021/05/26) extracted from UniProtKB/SwissProt, as well as the common contaminants, iRT retention time standards, HCP-PROFILER standard sequences, and NIST mAb heavy and light chain sequences. Finally, we used the number of identified/quantified HCP features and the R^2 determination coefficient to compare the acquisition methods.

In the first step, we evaluated the potential benefits of FAIMS implementation in DDA to increase the number of quantified HCPs (**Figure 74.A**). For that, we compared 4 MS methods: two without FAIMS (one using ion trap MS2 spectra detection and the other using Orbitrap detection) and two with FAIMS (-50V|-65V|-85V for ion trap and -50V|-70V for Orbitrap). Indeed, the fastest acquisition speed of the ion trap analyzer allow to use an additional CV compared to Orbitrap. The three technical replicates were analyzed with each method. Peptide and protein identifications were performed using the Sequest search engine, and validation using the Percolator algorithm, both implemented in Proteome Discoverer software (Thermo Fisher Scientific), applying a 1% FDR threshold at PSM and peptide levels. XIC-MS1 quantification was performed by summing the peptide intensities of each protein.

In the second step, the same samples were used to compare 4 DIA-based MS methods (**Figure 74.B****Figure 74**): two isolation schemes with fixed isolation windows (27 x 20 Da and 44 x 13 Da) and two methods with variable isolation windows (4 and 5 isolation widths per method). These methods will be detailed later in this chapter. Data were processed using the default parameters of the Spectronaut Pulsar™ software (Biognosys, Schlieren, Switzerland) working in DirectDIA™ library-free mode. A 1% FDR threshold was applied at the PSM and peptide levels.

In the third step, we pooled the three technical replicates of NIST mAb to appraise the capabilities of chromatogram library-based DIA data processing after gas phase fractionation (**Figure 74.C**). A total of 6 DIA methods were used with an MS1 range of 100 m/z followed by 25 narrow isolation windows of 4 m/z covering the mass range. Data obtained were searched with Pulsar™ implemented in Spectronaut (Biognosys, Schlieren, Switzerland) to generate a chromatogram library. Then, DIA data obtained from the optimized method in the previous section were searched against this chromatogram library. The optimized methods were applied on the NIST mAb Reference Material as well as on two FDA/EMA-approved mAbs, trastuzumab, and nivolumab, allowing the characterization and the quantification of trace-level HCPs (**Figure 74.D**).

Finally, we appraised open-modification search tools for the detection of new HCPs bearing unexpected modifications.

Additional details concerning the sample preparation, nanoLC-MS/MS methods and data processing are listed in the **Experimental Section**.

3. Implementation of a stringent data filtering workflow

The precise characterization of HCP impurities in drug products requires stringent validation. Considering the potential adverse effects of some HCPs with their eventual clinical consequences and potential economic impact, strict guidelines and criteria must be followed before drug product release and it is necessary to ensure that detected impurities are true positives. However, the manual checking of all the extracted signals is not feasible and gave rise to the implementation of rigorous validation criteria to end up with only confident and reproducible signals. Therefore, the following validation workflow was implemented, including five stages of data filtering, adapted from a workflow developed in the lab by Dr. Nicolas Pythoud (**Figure 75.A**).

First, we removed oxidized and acetylated precursors with their unmodified counterparts and kept only +2 and +3 precursors. These modifications and +4 precursors generally show poor-quality signals

regarding intensity and peak shape. Second, we add a filter on the signal quality with the removal of precursors without values for quantification or that are shared between protein groups. Precursors with more than one q-value >0.01 and/or “profiled” (DIA)/“peak found” (DDA) were also eliminated. Third, precursors with a CV above 20% were filtered out to ensure signal reproducibility. Fourth, we applied a homology filter using BLASTP⁵⁹⁰ (v.2. 10.0+) against mAb light and heavy chain sequences to remove HCP peptides that can potentially arise from unspecific cleavage or mAb degradation. HCP peptide sequences containing a minimum of 6 amino acids, 80% coverage, and 100% identity were eliminated.

To appraise the influence of stringent data filtering on our data, we tested it on two sets of three technical replicate runs of 2 μg (starting quantity before native digestion) NIST mAb Reference Material injected: data from DDA runs with ion trap MS2 detection and from DIA runs with 45 variable isolation windows. **Figure 75.B and C** present the multiple filtering step results. In detail, while the DDA approach demonstrates a continuous diminution of the number of features at each step of the filtering process, some filters do not affect this number for DIA data. It is particularly true for signal quality and blast filters. Although we are not expecting more than a few sequences to be removed due to poor signal quality using the DIA strategy – thanks to its ability to, in theory, detect every precursor that is present in a selected mass range – the stochasticity of precursor selection in DDA induces missing values between replicates, responsible for the removal of multiple sequences (from 278 to 175 peptides and 120 to 89 protein groups, **Figure 75.B and C**). As expected, the blast filter has no effect, no matter the strategy used. Indeed, we use it as a security filter to prevent the quantification of a highly abundant peptide from the mAb as an impurity and the probability of having an HCP sequence meeting this criterion is very low. The reproducibility criterion, filtering out precursors with a CV $<20\%$, is the most impactful filter. It removes between 25% and 35% of the features in DDA and 45% of the protein groups, and 65% of the peptides in DIA, compared to the previous filtering step.

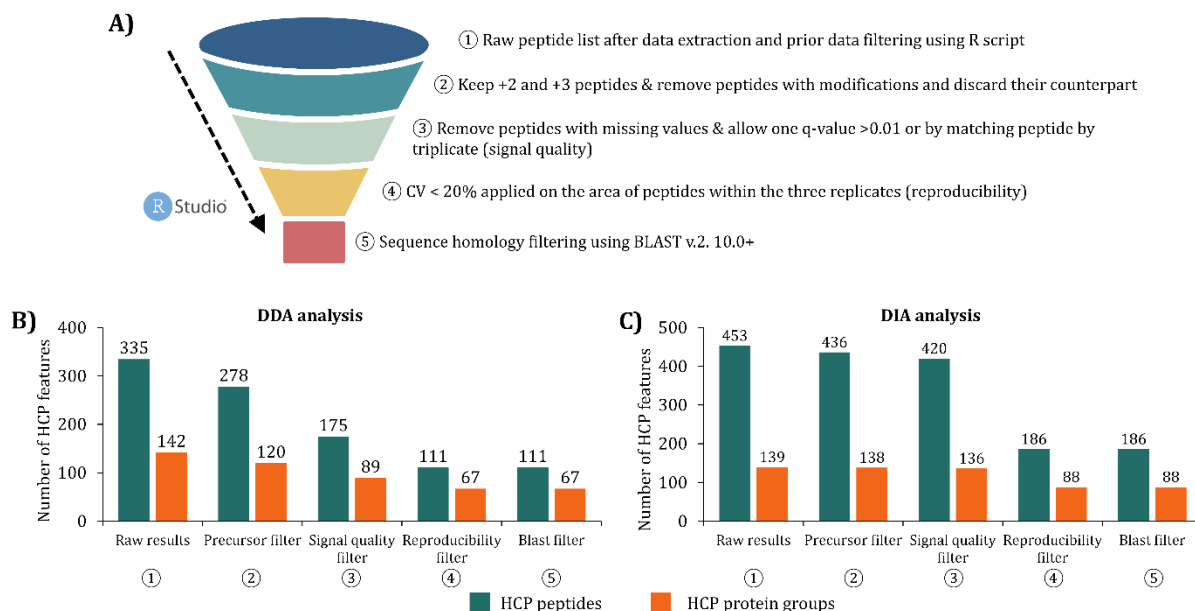


Figure 75: Impact of consecutive validation criteria applied on the number of HCP protein groups and peptides quantified using DDA and DIA acquisition methods. (A) Description of each step used to filter the raw list of HCP peptides. Number of HCP peptides and proteins groups remaining after each filter applied to the data obtained from (B) a DDA method, and (C) a DIA method.

In the end, the number of peptides and protein groups decreases by 67% and 50%, respectively, after all the filtering steps compared to the raw results in DDA, and by 60% for peptides and 37% for protein groups in DIA. It drastically reduces the number of HCP impurities but allows us to be confident in

their characterization and assume that the ones remaining are true positives. This high level of confidence is imperative when working on a drug product as the detection/quantification of a high-risk HCP, even at a low amount, can prevent the drug product from being released and generate considerable time and money costs for the company to get rid of this HCP.

4. Internal calibration curve-based quantification of host cell proteins

Residual HCPs passing all the validation filters are considered confidently characterized and require precise quantification. With this objective, multiple strategies have been employed in the lab. During his Ph.D., Dr. Gauthier Husson developed an innovative Top3-ID-DIA⁵⁰³ method combining a Top3 quantification strategy¹⁴⁶ for global HCP monitoring and isotopic dilution (ID) for key HCPs accurate quantification. The Top3 strategy assumes that the average MS signal response for the three most intense peptides per mole of protein is constant within a coefficient of variation of less than 10%. An internal standard is used to calculate a signal response factor (Top3 peptides signal/mol) and to perform HCP quantity estimation (HCP Top3 peptides signal/signal response factor). This approach has also been employed by Dr. Nicolas Pythoud, using a mix of four standard proteins (i.e., PYGM, ADH, ENL, BSA), previously reported for HCP quantification in several studies^{513, 514}, considering the PYGM protein as a reference and the three other proteins to calculate ratios, as an internal control⁵⁰⁴. Recently, a growing interest in standard proteins use has been observed, either using a single reference protein or based on an average amount calculated from the quantity obtained from each standard protein^{505, 509}. In this regard, Dr. Steve Hessmann evaluated the HCP-PROFILER standard, developed by the French company ANAQUANT, and demonstrated its reproducibility and ability for accurate HCP quantification. It is composed of 54 synthetic peptides (deriving from the sequences of 18 *Escherichia coli* proteins) that are grafted on a water-soluble bead (READY-BEADS technology)⁵⁹¹. Once added to the sample, they are released at six different concentration points (9 peptides per concentration point), from 1 to 500 fmol. It allows the building of an internal calibration curve of the log₂ (Top3 abundance) as a function of the log₂ (HCP-PROFILER theoretical quantity (fmol)), used to quantify HCPs.

The experimental details for HCP-PROFILER sample preparation and the list of peptides, sequences, and concentration points are mentioned in **Experimental Section**.

Compared to other approaches, ANAQUANT provides a ready-to-use kit, limiting bias on reproducibility potentially introduced when performing homemade standard mixture sample preparation. In addition, it ensures confidence and quantification accuracy over 2.7 orders of magnitude, a reasonable range for HCPs. With this in mind, we used the HCP-PROFILER standard to quantify HCP impurities in drug products. The influence of acquisition methods on the quantification accuracy will be discussed in the following sections of this chapter.

5. Evaluation of FAIMS DDA methods to dig deeper into the HCP landscape

The use of DDA for trace-level HCPs detection in drug products is limited by the bias induced by selecting a fixed number of abundant precursors. Indeed, HCPs are present in very low amounts, and in the case of co-elution of precursors from mAb and HCP, only the mAb precursor will be selected for further fragmentation. In **Part I, Chapter 1**, we demonstrated the benefits of FAIMS implementation to act as an orthogonal separation technique for co-eluting precursors and reduce the background noise to enhance low-abundant peptides detection. Furthermore, the addition of this ion mobility device does not reveal any loss in terms of quantification compared to analyses without the FAIMS. Hence, it is a promising approach to reduce the extreme dynamic range of drug product samples while allowing precise quantification of trace-level HCP impurities.

First, we appraised the precision of quantification, performed thanks to the HCP-PROFILER standard. We quantified more than 90% (50 over 54) of the HCP-PROFILER peptides in each method, except for

the one combining FAIMS to Orbitrap MS2 detection, with the quantification of only 46 over 54 peptides. For low abundant HCP-PROFILER standards (i.e., 1 and 2.5 fmol), only 1 or 2 peptides instead of the 3 expected are observed. However, the XIC-MS1 signal extraction of missing peptides corroborates their presence as chromatographic peaks are observed at their expected retention time. We saw that some are co-eluting with superabundant mAb peptides. Then, we supposed that their low abundance prevents them from being selected for fragmentation in the presence of mAb peptides at the same retention time. In addition, no apparent effect of FAIMS implementation can be observed on the detection of the peptides. Internal calibration curves built using HCP-PROFILER standards are presented in **Figure 76**. Two main patterns are observed, based on the presence or the absence of the FAIMS: for analyses without the FAIMS, ANAQUANT-4 standard (2.5 fmol) is less intense than the two other standards at this quantity, while this is the case for ANAQUANT-9 (10 fmol) and ANAQUANT-17 (500 fmol) for analyses with the FAIMS. For ANAQUANT-9 and ANAQUANT-17, FAIMS implementation can be responsible for a lower MS signal response, as these two proteins got their three peptides quantified and presented good signal response without the FAIMS. Overall, the R^2 determination coefficient calculated is between 0.86 and 0.93, with lower values for methods with the FAIMS, probably related to the low response signal of standards ANAQUANT-9 and ANAQUANT-17.

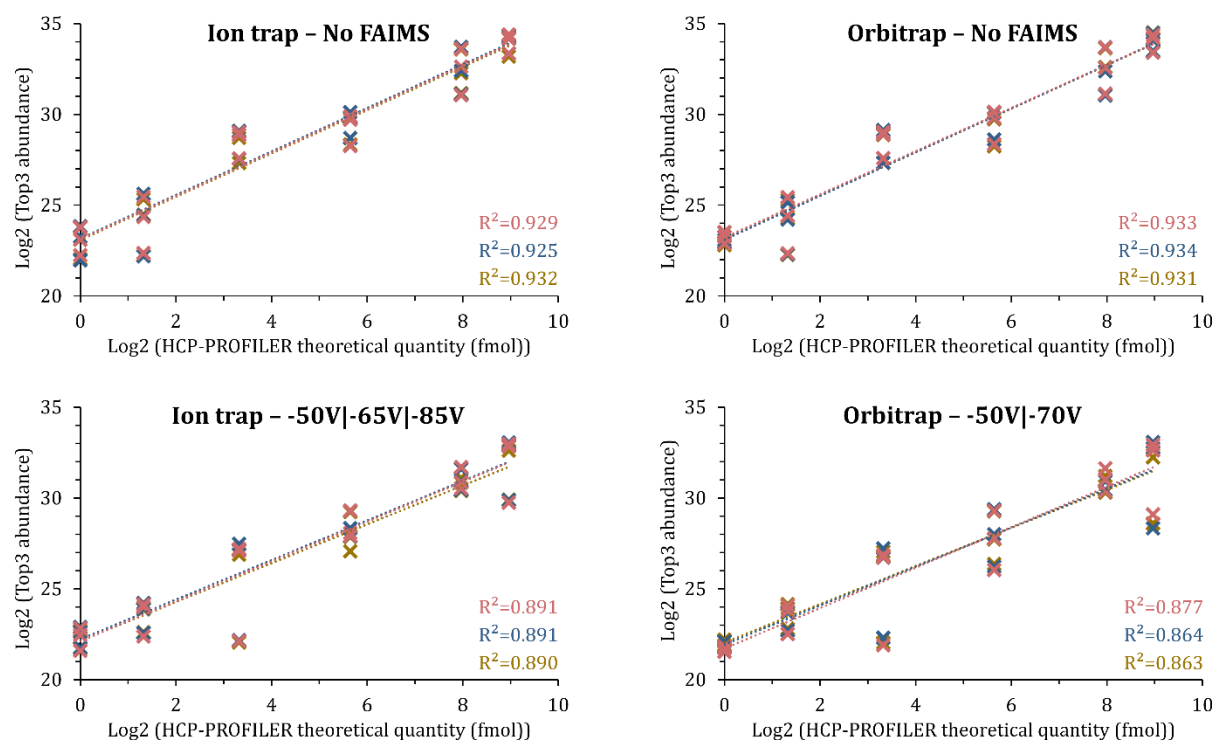


Figure 76: Representation of the HCP-PROFILER-based internal calibration curves obtained from DDA optimizations using ion trap and Orbitrap MS2 detection combined or not with the FAIMS. Internal calibration curve of the log2 of Top3 HCP-PROFILER abundance as a function of the log2 of the HCP-PROFILER theoretical quantity (fmol) for replicate 1 (red), 2 (blue) and 3 (yellow). For each replicate, R^2 determination coefficient are displayed at the bottom-right of the plot.

The assessment of the DDA methods with and without the FAIMS was realized with ion trap and Orbitrap MS2 detection. As expected, the number of HCP features detected is higher using the ion trap MS2 detection than the Orbitrap. In detail, Orbitrap MS2 detection identified 148 and 205 peptides from 82 and 89 HCP protein groups, with and without the FAIMS, respectively (**Figure 77**). The higher number of features identified without the FAIMS indicates that using two compensation voltages is not optimal for HCP detection. Contrarily, using ion trap MS2 detection, FAIMS implementation increases the number of identified HCP protein groups by 40% and peptides by 26% (**Figure 77**). The best conditions, using three compensation voltages (-50V, -65V, and -85V), result in the identification of

more than 420 peptides originating from up to 200 HCPs. Then, we applied confidence filters mentioned in **Section 3**. These filters eliminate a non-negligible part of the features identified, as previously observed and discussed. In essence, less than 50% of the proteins and around 30% of the peptides from analyses without FAIMS were conserved only, in contrast to 65% and 50%, respectively, for analyses with the FAIMS, highlighting the ability of FAIMS methods to perform good quantification. The same trends as before filtering are observed, and FAIMS combined to ion trap MS2 detection appears as the best approach, with the confident characterization of 221 peptides and 130 HCP protein groups. We used the internal calibration curve obtained for each method to quantify HCPs that have passed stringent validation criteria. A total quantity between 330 and 570 ng of HCPs per mg of mAb (ppm) was found, with lowest quantities detected with methods without the FAIMS and higher with those with the FAIMS implemented (**Figure 77. A**). On the one hand, it can be explained by the considerable increase in the number of HCPs between analyses in the absence and presence of the FAIMS with ion trap MS2 detection. On the other hand, as the number of HCPs quantified with the FAIMS is not significantly higher than without, only one HCP, representing more than 15% of the total amount and not found by any other approach, appears to be responsible for the global quantity rise with Orbitrap MS2 detection (**Figure 78**).

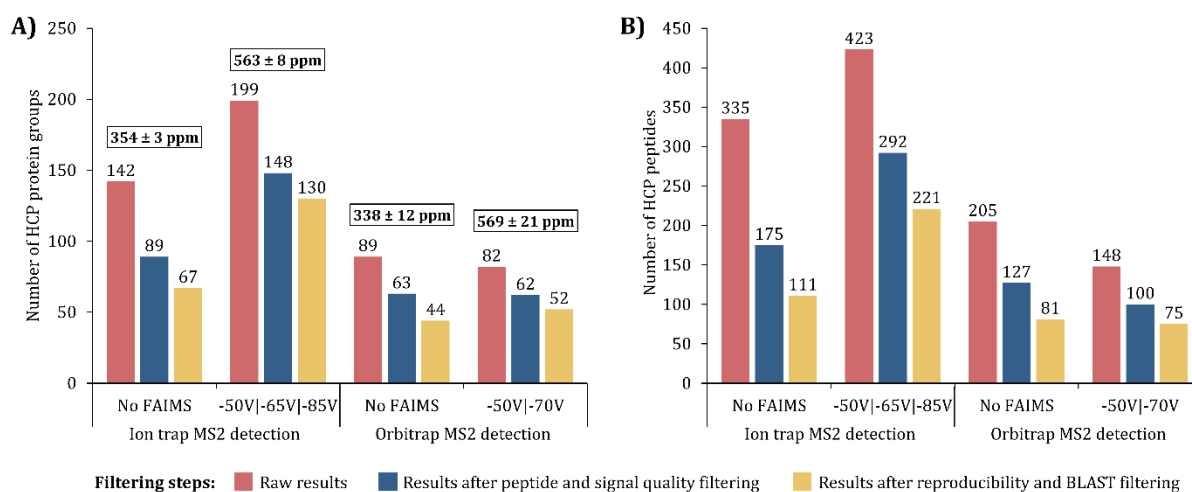


Figure 77: Results from NIST mAb Reference Material DDA methods optimizations. (A) Number of HCP protein groups and (B) number of HCP peptides before filtering (red), after precursors and signal quality filtering (blue) and after reproducibility and blast filtering (yellow) for MS2 spectra detection performed in the ion trap and the Orbitrap, without FAIMS and with the FAIMS (-50V|-65V|-85V for ion trap, -50V|-70V for Orbitrap). Values in the boxes represent the average quantity of ng of HCP per mg of mAb (ppm) calculated for the three technical replicates after all the filtering steps.

Above all, FAIMS implementation allows us to dig deeper into the HCP landscape, as we can quantify more HCPs, predominantly detected at trace-level quantities. In **Figure 78**, most of the HCPs quantified only with the FAIMS are low abundant, with up to one order of magnitude gain compared to analyses without FAIMS.

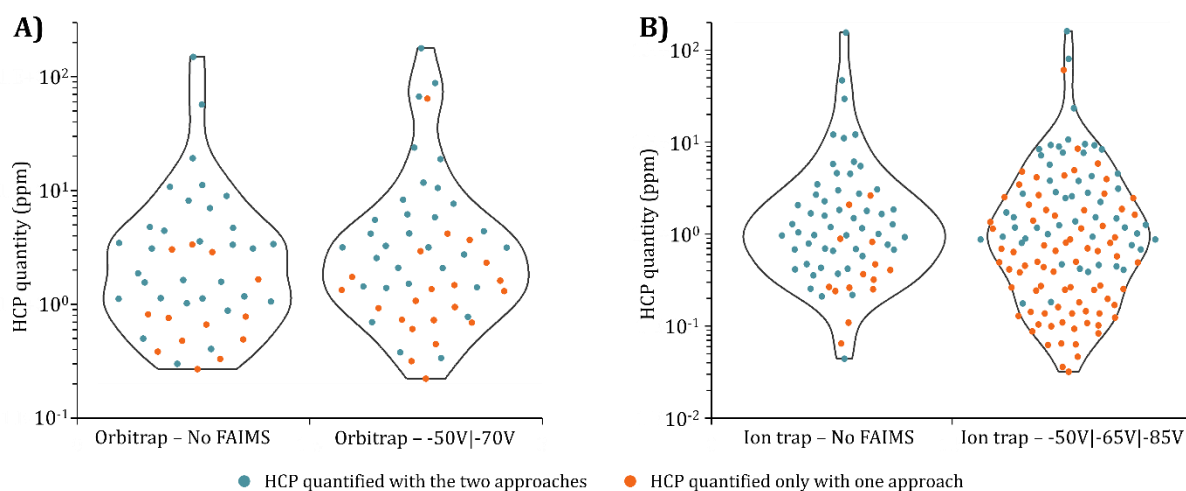


Figure 78: Violin plot representing the density of HCPs as a function of their concentration in drug products. (A) Comparison between acquisition methods with and without the FAIMS using Orbitrap MS2 detection. (B) Comparison between acquisition methods with and without the FAIMS using ion trap MS2 detection. Green dots stand for proteins quantified in both approaches (without and with the FAIMS) and orange dots for proteins quantified only with one approach or the other.

In summary, ion trap MS2 detection remains the best approach for HCPs characterization. Its combination with the FAIMS allows drastically increasing the identified and quantified HCPs features. It demonstrates the benefits of digging deeper into the HCP landscape by quantifying HCPs down to the sub-ppm level. Even if its use slightly decreases the precision of quantification, lowering the response signal of some standards used for the calibration curve building, it remains the most powerful approach to accurately quantify trace-level impurities in drug products and increase the depth of analysis, thanks to background noise reduction and orthogonal separation of the co-eluting precursor ions.

6. Evaluation of DIA strategies for HCP characterization in drug products

A. Optimizations of acquisition methods

Characterization of HCP impurities in a drug product represents a significant analytical challenge, mainly because of the extreme dynamic range present in this sample between the superabundant mAb and the trace-level impurities. In the section above, we saw the DDA approach's limits to detecting low-abundance peptides. With this in mind, DIA-MS appears as a promising approach for HCP study in these samples as every precursor present in the selected mass range is fragmented, theoretically allowing the detection of all the species, in contrast to DDA, where only the most abundant precursors, are selected for further fragmentation. Moreover, in **Part I, Chapter 1**, DIA demonstrated its capability for quantifying low abundant UPS1 proteins spiked in complex samples. In this context, we assessed different DIA methods based either on fixed or variable window widths isolation schemes.

First and foremost, we used a preliminary DDA run to set up the mass range of the DIA methods and adapt the isolation windows placement for variable isolation windows schemes. Only the precursors from HCPs (i.e., *Mus musculus* precursors) were kept to focus the DIA approaches on impurities precursors and not introduce bias with mAb precursors. The repartition of the precursors over the mass range is represented in **Figure 79**. It shows that most of the HCP precursor ions m/z are in the 450-750 m/z mass range and that its extension to 380-975 m/z contains about 95% of the *Mus musculus* precursors. Then, we selected this interval to set up our DIA methods. Expanding the mass range to 1,100 or 1,200 m/z would have significantly increased the cycle time while only a few more precursors could be detected and we therefore decided to not expend the mass range over 975 m/z.

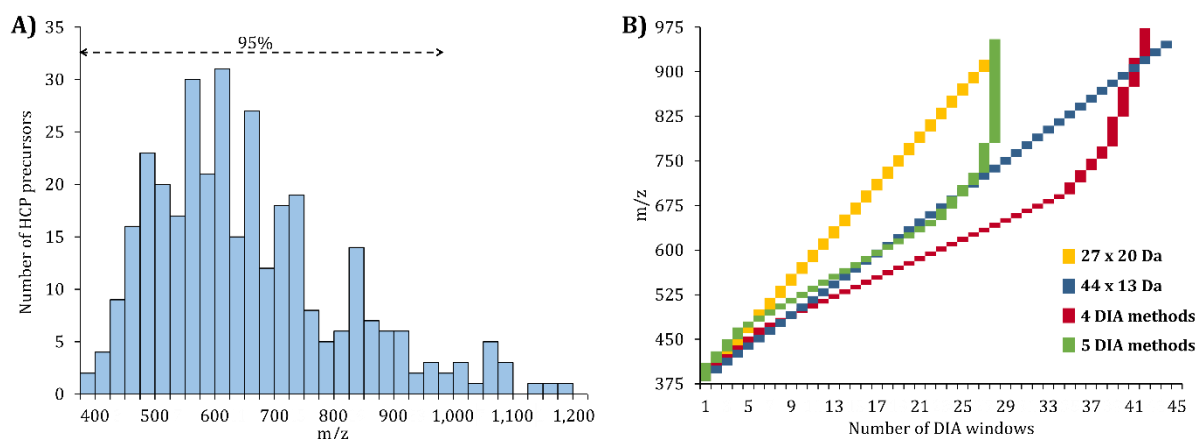


Figure 79: Optimizations of DIA acquisition methods in accordance with sample complexity. (A) Distribution of the HCP precursors among the mass range for a DDA analysis. (B) DIA isolation schemes including fixed isolation windows (27 x 20 Da (yellow) and 44 x 13 Da (blue)) and variable isolation windows (using 28 windows of 5 different widths (green) and 45 windows of 4 different widths (red)).

Provided that two methods with fixed isolation windows were designed: 27 windows of 20 Da and 44 windows of 13 Da, covering the mass range of 380-920 m/z and 380-952 m/z, respectively. In addition, two isolation schemes with variable isolation widths were also evaluated. The first one, in the 380 and 955 m/z mass range composed of 28 windows from 10 to 175 Da (5 different widths), and the second, in the 380-974 m/z mass range, composed of 45 windows from 8 to 50 Da (4 different widths). The four isolation schemes are displayed in **Figure 79**.

First, we had a look at the HCP-PROFILER peptides that are used for quantification. Depending on the isolation scheme, 50 to 54 over 54 peptides were quantified, but at least one peptide was quantified per standard in each method. By and large, the isolation schemes with 44 fixed and 45 variable windows display more HCP-PROFILER peptides quantified. Then, we verified the quantification accuracy by building internal calibration curves for each isolation scheme (**Figure 80**). First, a similar pattern is observed with a low dispersion between the three proteins of the same quantity, except for the quantities of 10 fmol and 500 fmol for the approach with 27 fixed windows and the one with 45 variable windows, respectively. For each replicate, the R^2 determination coefficient of the internal calibration curve is between 0.92 and 0.95, with the approaches mentioned above demonstrating a broader dispersion in some calibration points, showing the lowest R^2 coefficients. These values are slightly higher than those observed in DDA, highlighting DIA capabilities in the context of HCP quantification.

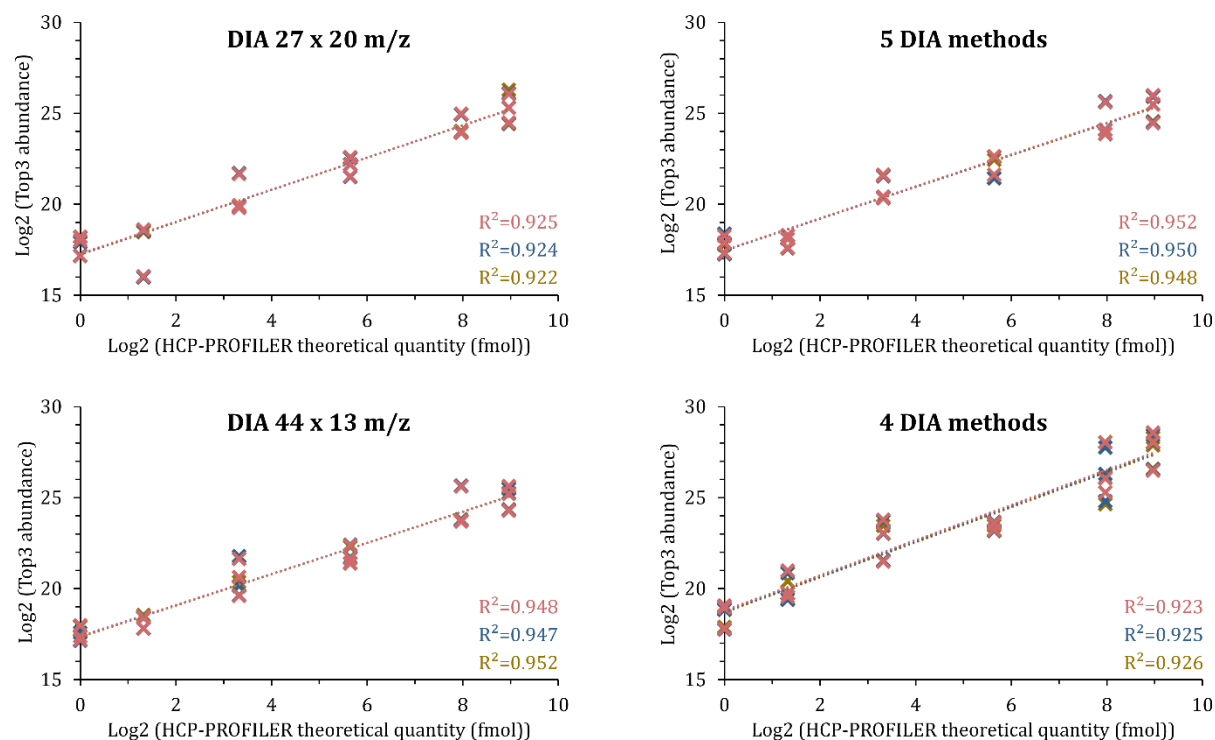


Figure 80: Representation of the HCP-PROFILER-based internal calibration curves obtained from DIA optimizations using fixed and variable windows isolation schemes. Internal calibration curve of the log₂ of Top3 HCP-PROFILER abundance as a function of the log₂ of the HCP-PROFILER theoretical quantity (fmol) for replicate 1 (red), 2 (blue) and 3 (yellow). For each replicate, R² determination coefficient are displayed at the bottom-right of the plot.

Then, the comparison of the results obtained for fixed isolation windows points out a gain of 6% and 15% of HCP peptides and protein groups identified, respectively, using the 20 Da windows compared to the 13 Da ones, for a total of 93 HCPs (**Figure 81. A and B**). Isolation schemes with variable isolation windows present two different trends. While the scheme with 28 windows of 5 different widths resulted in the detection of fewer than 150 peptides and 71 protein groups, the strategy with 45 windows of 4 different widths allowed the identification of 177 peptides from 88 HCPs, close to the results obtained with fixed isolation windows. After data filtering using stringent signal quality and reproducibility criteria, we noticed that the proportion of features that passed the filters is low, between 20% and 35%, for most isolation schemes. Only the isolation scheme with 45 variable windows from 4 DIA methods presents slightly better results with 45% of the protein groups and 40% of the peptides that are kept after stringent filtering (**Figure 81. A and B**). In light of these results, the cycle time of 2 s used for the method with 45 isolation windows, resulting in 7 points per chromatographic peak on average, does not seem to significantly impact the reproducibility of the quantification (**Figure 81. C**). Further, HCP amounts between 180 and 310 ppm were calculated, with the highest quantity retrieved from the method with 45 variables windows, in line with the high number of features quantified with this approach (**Figure 81. A**). Thus, this method emerges as the best compromise to confidently characterize HCPs in drug products, with 69 peptides from 39 HCP protein groups quantified.

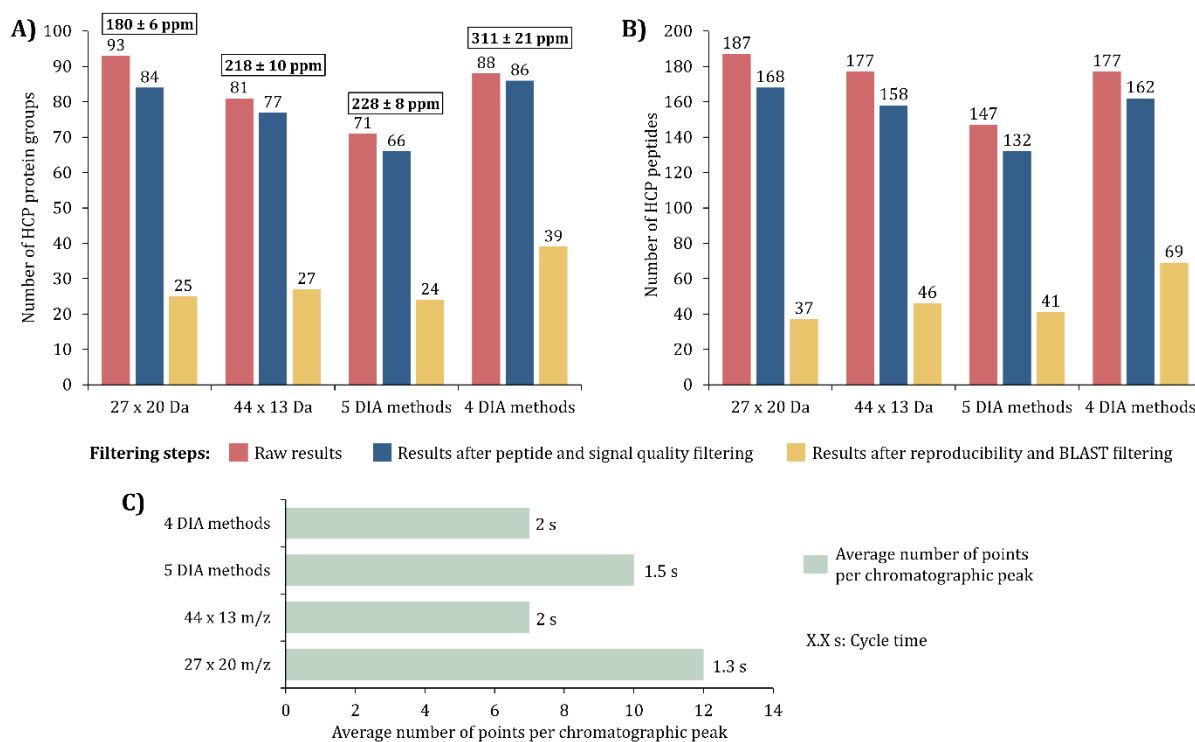


Figure 81: Results from NIST mAb Reference Material DIA methods optimizations. (A) Number of HCP protein groups and (B) number of HCP peptides before filtering (red), after precursors and signal quality filtering (blue) and after reproducibility and blast filtering (yellow) for DIA isolation schemes with fixed window widths and variable window widths. Values in the boxes represent the average quantity of ng of HCP per mg of mAb (ppm) calculated for the three technical replicates after all the filtering steps. (C) Average number of points per chromatographic peak and the associated cycle time are illustrated for each acquisition method.

The DIA isolation scheme needs to be optimized to keep a compromise between the number of windows, their placement, and the cycle time. Using DIA, we can also accurately quantify the HCPs detected, as internal calibration curves show good linearity over the whole range. Then, the isolation scheme with 45 variable windows allowing a cycle time of around 2 s is the best approach to quantify HCP impurities, and it can be used for their monitoring in FDA/EMA-approved drug products.

B. Influence of the quantity injected and comparison to DDA

Globally, results obtained from the DIA mode present fewer HCPs than DDA (**Part II-5**). This observation can be explained by the lower quantity injected in DIA (1 µg in DIA vs. 2 µg in DDA, starting quantities before native digestion). Hence, to have a fair comparison, we injected the same amount of NIST mAb Reference Material (i.e., 2 µg) in DIA. Briefly, the method with ion trap MS2 detection without the FAIMS was used as a reference for DDA and with 45 variable isolation windows for DIA. The quantities indicated are theoretical quantities considering the absence of mAb depletion during the native digestion, they do not represent the real quantity injected on the column.

The results of the comparison are presented in **Figure 82**.

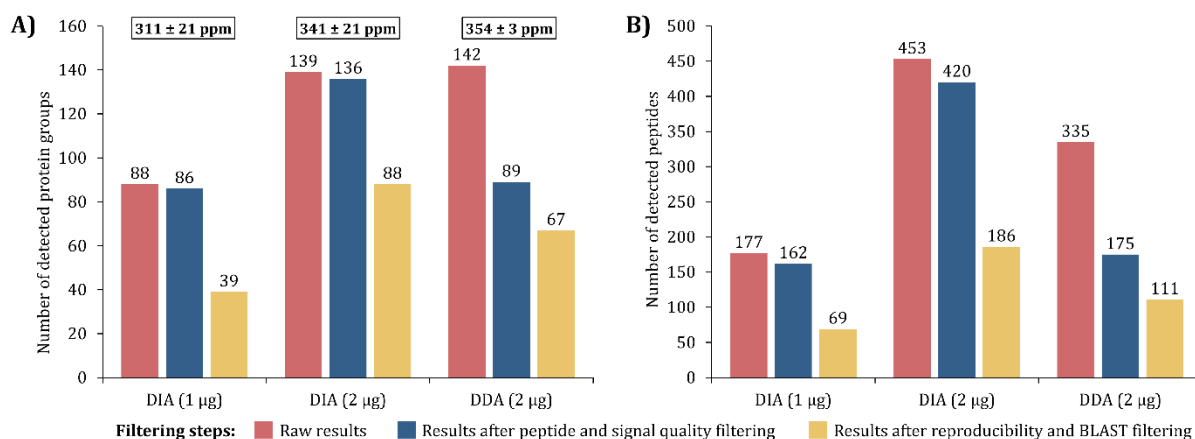


Figure 82: Results from NIST mAb Reference Material for DIA and DDA. (A) Number of HCP protein groups and (B) number of HCP peptides before filtering (red), after precursors and signal quality filtering (blue) and after reproducibility and blast filtering (yellow) for the optimized DIA method (45 variable windows) for the injection of 1 µg and 2 µg of NIST mAb Reference Material and for DDA using ion trap MS2 detection without the FAIMS for the injection of 2 µg. Values in the boxes represent the average quantity of ng of HCP per mg of mAb (ppm) calculated for the three technical replicates after all the filtering steps.

First and foremost, the injection of 2 µg instead of 1 µg in DIA resulted in a gain in the number of features detected. Regarding the data before any filtering, the number of detected protein groups raised by 60%, and the peptides were increased by 2.6 times by injecting 2 µg. After stringent filtering, a significant gain is still observed, with 2.5 and 2.7 times more protein groups and peptides confidently quantified, respectively. A similar proportion, around 60%, of the features was removed after stringent filtering, independently of the quantity injected. Although the increase of HCPs impurities quantified when 2 µg was injected, the total concentration (ng of HCP per mg of mAb, ppm) increased only by 10%. Indeed, the four most abundant HCPs, which are the same for 1 µg and 2 µg, account for around 75% of the total quantity (**Table 8**). Consequently, even if more proteins are quantified, their low amount does not significantly affect the total HCP quantity.

Table 8: Concentration of the four most abundant HCPs found for each method and their contribution on the total HCP quantity calculated.

	ALDOA_MOUSE (P05063)	PDIA6_MOUSE (Q922R8)	ALDOC_MOUSE (P05064)	G6PI_MOUSE (P06745)	ROA2_MOUSE (O88569)	Total 4 most abundant proteins	Proportion over the total HCP quantity
DIA 1 µg	111 ppm	60 ppm	52 ppm	21 ppm	X	244 ppm	78%
DIA 2 µg	119 ppm	92 ppm	58 ppm	12 ppm	X	281 ppm	82%
DDA 2 µg	155 ppm	X	47 ppm	30 ppm	12 ppm	244 ppm	69%

Results from 2 µg NIST mAb Reference Material DDA and DIA injections were compared (**Figure 82**). A comparable number of HCP protein groups was detected (142 for DDA vs. 139 for DIA), but DIA detected up to 120 additional peptides compared to DDA (453 vs. 335). After stringent filtering, DIA allows us to confidently quantify 186 peptides from 88 HCPs against 111 from 67 HCPs for DDA. Specifically, the criteria on the precursor modifications, charge states, and missing values filtered out 48% of the peptides and 38% of the protein groups in DDA. In comparison, less than 10% were removed in DIA. This trend, already observed before in this chapter, highlights the limitation of DDA to identify low abundant peptides in a reproducible way between runs because of the stochasticity of the approach and, consequently, the advantage of DIA to circumvent this issue. Similarly, DIA demonstrates a better reproducibility of the identified features, as 41% of the peptides and 63% of the proteins detected before any filters were confidently quantified compared to 33% and 47% in DDA,

respectively. Ultimately, a similar concentration of HCP, around 350 ppm, was found for both DDA and DIA strategies. They show the four most abundant proteins, accounting for 79% in DIA and 66% in DDA of the total amount (**Table 8**), explaining that additional proteins are not increasing the total HCP amount considerably.

C. Use of a chromatogram library from gas phase fractionation for DIA data processing

Several strategies are available to process DIA data: the “peptide centric” approach, based on a spectral library generated from DDA runs, or the “spectrum centric” approach, where DIA data are directly searched against in-silico predicted libraries. In the context of HCP characterization in drug products, the spectral library-based approach is hard to implement. Indeed, spectral library generation is time and sample-consuming and requires access to the null cell line to build the most comprehensive spectral library possible in DDA before DIA data analysis. To circumvent these limitations, the use of gas phase fractionation to generate chromatogram libraries was recently described by Pino *et al.*¹⁷². The workflow is composed of the following steps and detailed in **Figure 83**: (i) pooling of the samples of interest to be the most representative of the samples; (ii) analysis of the pool in six independent DIA runs, covering each one a 100 m/z range and composed of 25 windows of 4 m/z; (iii) data analysis of these six runs using the Pulsar™ search engine and generation of the chromatogram library; (iv) analysis of the previously acquired DIA data from each sample against the chromatogram library.

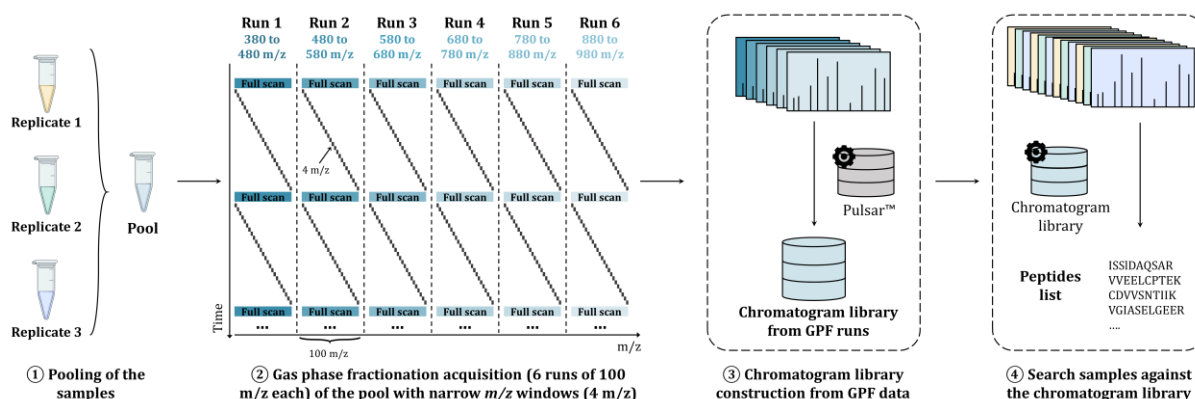


Figure 83: Chromatogram library generation using gas phase fractionation DIA. First, samples are pooled together (①) and analyzed in six independent DIA runs of 100 m/z using narrow isolation windows of 4 m/z (②). Results from these runs are merged together and search and a library using the Pulsar™ search engine to generate a chromatogram library (③). Finally, DIA data previously obtained are searched against the chromatogram library to give the list of peptides (④).

To generate our chromatogram library, we pooled the three technical replicates and analyzed them using six runs of 100 m/z mass range, from 380-980 m/z. The chromatogram library generated contains 1,380 precursors, 1,227 peptides, and 398 protein groups, including the 18 HCP-PROFILER standards and the iRT and mAb heavy and light chains sequences.

We assessed HCP-PROFILER standards detection and the resulting internal calibration curves for both strategies. As expected – as we are using the same dataset – we detected all the peptides when searching the data against the chromatogram library. Moreover, the calibration curves in **Figure 84** illustrate the similar results obtained for both data processing using the library-free and chromatogram library-based approaches, with an R^2 determination coefficient between 0.92 and 0.93.

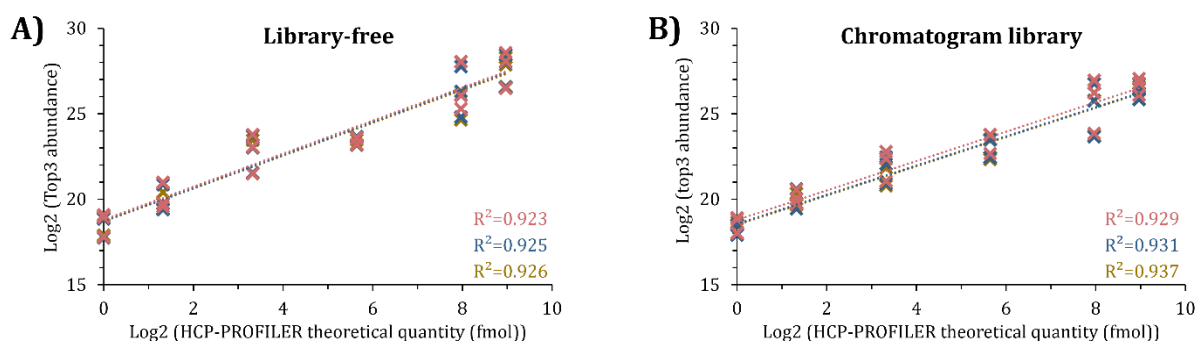


Figure 84: Representation of the HCP-PROFILER-based internal calibration curves obtained from DIA acquisition method with variable windows widths. (A) Data using the library-free approach. (B) Data using the chromatogram library-based approach. Internal calibration curve of the log₂ of Top3 HCP-PROFILER abundance as a function of the log₂ of the HCP-PROFILER theoretical quantity (fmol) for replicate 1 (red), 2 (blue) and 3 (yellow). For each replicate, the R² determination coefficient are displayed at the bottom-right of the plot.

We used the data obtained from 2 µg of NIST mAb Reference Material injected using the 45 variable windows already presented in the previous section. The results obtained from the chromatogram library and the library-free approach are displayed in **Figure 85.A and B**. Data processing using the chromatogram library doubles the number of HCP protein groups and peptides detected. Overall, we detected up to 875 peptides from 281 HCP impurities. In contrast, after stringent filtering following the steps mentioned previously in this chapter, the benefits of the chromatogram library approach are slightly less significant, with 45% more HCP peptides and 24% more HCP protein groups quantified using this approach compared to DirectDIA™ (270 vs. 186 for peptides and 109 vs. 88 for protein groups, respectively). Whereas a similar proportion (around 10%) of features is removed after precursor and signal quality filtering using both approaches, the reproducibility filter eliminates up to 65% and 60% of the remaining peptides and protein groups, respectively, for data obtained after chromatogram library data treatment. In comparison, solely 35% of the protein groups and 55% of the peptides are filtered out because of the reproducibility criteria when using DirectDIA™. The variability of very low abundant precursor intensities detected and the difficulty distinguishing between precursor signal and background noise result in a lower proportion of validated features at these trace level amounts. In addition, the global HCP amount is around 340 ppm for the library-free approach, while it increases to 360 ppm using the chromatogram library (**Figure 85.A**). Importantly, as with the FAIMS implementation in DDA, data treatment using a chromatogram library increases the depth of analysis, with most of the HCP quantified uniquely with this approach present at concentrations down to the sub-ppm level (**Figure 85.C**). Using the chromatogram library gives obvious evidence of its benefits for detecting trace-level impurities with the accurate quantification of proteins below 1 ppm that are not found using the library-free strategy.

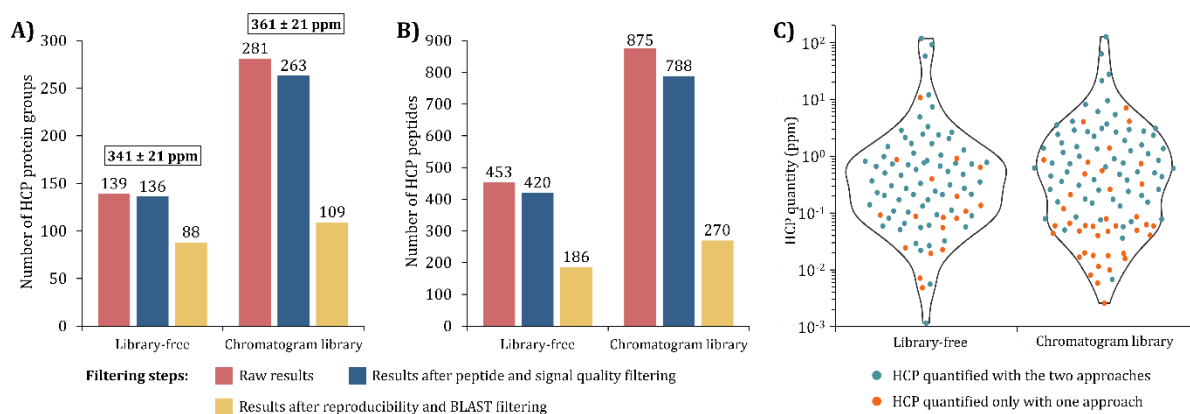


Figure 85: Results from NIST mAb Reference Material DIA data treatment. (A) Number of HCP protein groups and (B) number of HCP peptides before filtering (red), after precursors and signal quality filtering (blue) and after reproducibility and blast filtering (yellow) for DIA data treatment using the library-free approach (DirectDIA™) or the chromatogram library-based approach. Values in the boxes represent the average quantity of ng of HCP per mg of mAb (ppm) calculated for the three technical replicates after all the filtering steps. (C) Violin plot representing the density of HCPs as a function of their concentration in drug products for HCPs quantified using the library-free or the chromatogram library-based approach. Green dots stand for proteins quantified in both approaches (without and with the FAIMS) and orange dots for proteins quantified only with one approach or the other.

As demonstrated above, using a chromatogram library presents some advantages, such as the reduced time and sample amounts required to generate the library compared to sample fractionation followed by DDA analysis. It significantly increases the number of detected HCP features, among which up to 120 impurities are accurately and confidently quantified. In addition, it proves its ability to dig deeper by quantifying proteins level not found in the library-free approach down to the sub-ppm. Nevertheless, an essential part of the signals is of poor quality or at too low levels, causing poor quantification reproducibility of the detected species and demonstrating the importance of the stringent validation criteria.

The methods optimized in this chapter were applied to the NIST mAb Reference Material and two drug products, trastuzumab, and nivolumab, to characterize and quantify HCPs. The results have been published this year in *Proteomics* (see publication in **Annexes**). Briefly, the same trends were observed for the two commercial drug products as the NIST, with the FAIMS DDA approach displaying the highest number HCP impurities quantified, around 30, for a total amount around 55 ppm (**Figure 86**).

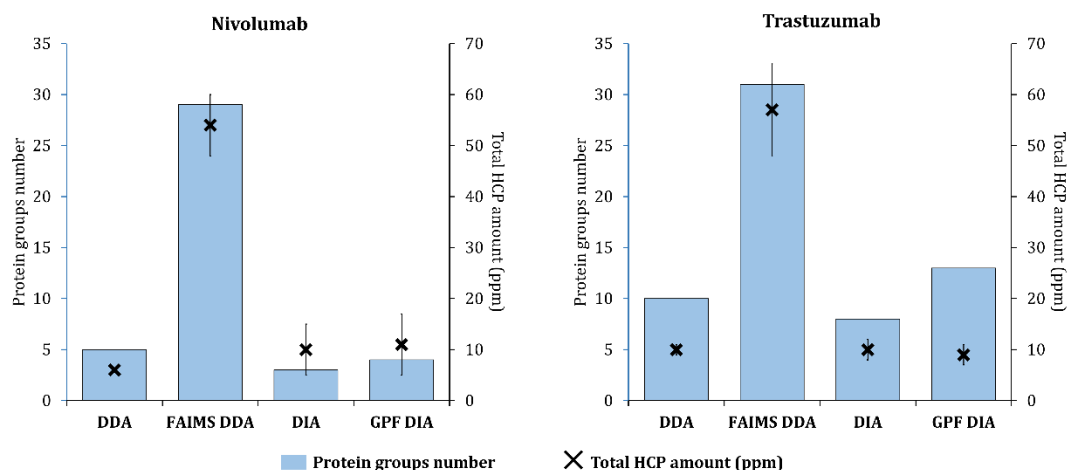


Figure 86: HCPs quantified by each method for trastuzumab and nivolumab and total HCP amount. Blue bars represent the number of HCPs quantified and black crosses the total HCP amount, in ppm.

7. Evaluation of open-modification search tools for HCPs characterization

The experimental conditions applied throughout the manufacturing process (i.e., upstream and downstream) can affect the integrity of the mAb and co-eluting HCPs. Multiple PTMs, part of the CQAs, are usually monitored on the mAb. Some of these common PTMs, such as acetylation or oxidation, are known to be occurring and are generally included in the database search. However, protein sequences can also exhibit other, less known modifications that people are generally not seeking, as these complicate general database searches because of the vast search space expansion. Nonetheless, the presence of these modifications is non-negligible and thus should be considered when monitoring HCP impurities. Hence, we used ionbot to perform an open modification search (OMS) on triplicate DDA runs from the NIST mAb sample prepared by native digestion. These data have been obtained from Orbitrap detection as the ionbot requires high-resolution data.

The results from OMS led to the detection of more than 200 PTMs, among which more than half are amino acid variations, considering all the replicates. An average number of 190 modified HCP peptides were identified in each replicate. The comparison of the list of these modified peptides between replicates underlines that 60 of them (~30%) are common to the three replicates. In addition, the major part of the peptides identified with a modification is not identified in their unmodified form. Hence, it highlights the benefits of using OMS to identify new peptides.

After that, we had a look at the modifications detected after OMS. **Figure 87** illustrates the average number of modified peptides identified in each replicates for the most common modifications found with ionbot. Two significant modifications are found: oxidation and deamidation. These modifications are part of the most expected PTMs for the HCPs and can be induced during the manufacturing process⁵⁹². Oxidation and deamidation are CQAs that should be monitored for the mAb, as they can reduce the efficiency of the mAb.

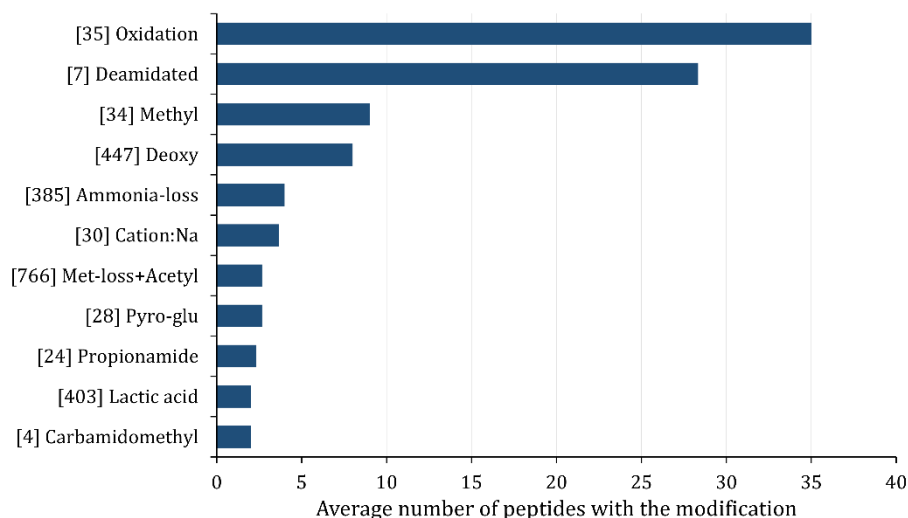


Figure 87: Average number of peptides bearing one of the most detected modifications. Modifications are indicated using their Unimod code and name.

Other modifications such as methylation, deoxydation, sodium adduct, or pyro-Glu are also detected. These modifications are either part of the common modifications occurring on the mAb during the manufacturing process⁵⁹³, such as pyro-Glu or deoxydation, or are common modifications found on proteins, such as methylation, ammonia-loss, or sodium adducts. These results evidence the ability of OMS to detect biologically relevant modifications, allowing the identification of additional peptides. It could be of great interest to HCPs, as these proteins undergo multiple experimental conditions during the mAb manufacturing process that can affect their integrity and lead to PTMs. Hence, considering all

the possible PTMs opens the door to identifying these modified HCPs and providing a more comprehensive overview of the HCP content in the sample.

Nevertheless, it should be noted that the identification of false positives is still possible with OMS. Indeed, for this study, it has been found that carbamidomethylated peptides are present in the sample. However, samples were prepared using the native digestion protocol, theoretically avoiding the presence of carbamidomethylation (i.e., because no alkylation step is performed).

To conclude, OMS is a promising tool in the context of HCPs. It allows the detection of new HCPs identified only thanks to modified peptides but can also point out the presence of unexpected modifications. Suppose a specific unexpected modification is found on many HCPs. In that case, it might suggest some issues regarding the bioprocessing and imply a closer look at these modifications on the therapeutic protein. A paper further investigating the potential of OMS for HCP detection is in preparation in collaboration with the CompOmics group (Ghent, Belgium).

8. Conclusion and perspectives

After demonstrating the value of FAIMS implementation and DIA for quantifying proteins in complex mixtures in Part I, this chapter aimed to test these approaches for the characterization of HCP impurities in drug products. Considering the potential consequences of high-risk HCPs in final drug products, we adopted a strategy combining stringent validation criteria and internal calibration curve-based quantification. Hence, only highly-confident features are kept to be accurately quantified thanks to the internal calibration curve, allowing us to provide reliable results.

Afterwards, the addition of a further separation dimension was considered in order to increase HCP detection sensitivity. The combined use of ion mobility and MS was therefore evaluated on the Orbitrap Eclipse Tribrid coupled to the FAIMS device. Compared with analyses without ion mobility, the implementation of the FAIMS demonstrates very encouraging performances in this context, with a significant increase in the number of HCP quantified. In addition, it shows the ability to overcome the extreme dynamic range of drug product samples and dig deeper into the HCP landscape, with the quantification of HCPs down to 1 ppm.

Then, DIA methods were optimized to provide the best parameters for HCP detection, and two data treatment strategies were employed. Compared to the library-free approach, using a chromatogram library drastically increased the number of detected impurities, pointing out its potential benefits for trace-level analysis. Nevertheless, the reproducibility of the MS2 signals extracted with this approach remains a downside to ensuring the confident and accurate quantification of such trace-level amounts of impurities.

Finally, OMS was used for the detection of HCP peptides. A few hundred peptides were identified, with a significant part identified only under a modified form (i.e., bearing a PTM). Modifications found on these peptides are, for the most part, related to the mAb manufacturing process. Hence, OMS appears to be a promising approach to identifying additional HCPs.

This section has highlighted the potential of MS for monitoring HCPs in highly challenging samples. Given the importance of these impurities in drug substances, other alternatives and improvements are continuously being considered. Numerous recent studies have focused on optimizing innovative sample preparation to reduce the dynamic range between the HCP impurities and the mAb, including HCP enrichment strategies with Protein A affinity chromatography⁵¹⁷ or molecular weight cutoff⁵¹⁰. In addition every ongoing innovation in MS has quickly been evaluated for its HCP profiling capabilities. As for instance, diaPASEF¹⁶³, in which the addition of ion mobility measurements to a high-resolution accurate mass (HRAM) instrument (TimsTOF Pro, Bruker) operating in DIA mode was leveraged. Unlike FAIMS, which only filters ions, this instrument features an ion mobility tube, giving access to

measurements at this level. The capability of the novel Astral™ – for Asymmetric Track Lossless – mass analyzer coupled to an Orbitrap (Thermo Fisher Scientific) allowing a scan rate of up to 200 Hz⁵⁹⁴ was also evaluated for HCP profiling⁵⁹⁵.

Ultimately, the implementation of MS-based methods is becoming widespread for HCP characterization, with the submission of MS data to health authorities as part of the approval procedure for new biotherapeutics. For instance, while the previous chapter of the United States Pharmacopeia (USP) dedicated to HCP measurement in biopharmaceuticals⁴⁷³ was almost only focused on immunoassay methods with little discussion about orthogonal methods, a new chapter named <1132.1> *Residual Host Cell Proteins Measurement in Biopharmaceuticals by Mass Spectrometry*, containing guidance and best practices for MS-based HCP identification and quantification is currently being written. This certainly highlights the importance of mass spectrometry for HCP monitoring.

PART III
STRUCTURAL MASS SPECTROMETRY TO
CHARACTERIZE mAb-BASED
BIOTHERAPEUTICS

Chapter 1:

Workflow optimization of top-down and middle-down mass spectrometry approaches

1. Context of the project and analytical strategy

Mass spectrometry is a core tool in biomolecule identification and characterization. While MS analysis of intact proteins gives access to their intact mass, the use of peptide mapping fingerprinting (PMF) enables the determination of the primary structure of the proteins. It has been considered the gold standard method for protein sequencing, PTM identification, and sequence variant characterization. However, since PMF is a peptide-centric technique, the information obtained from the LC-MS/MS analysis cannot be directly correlated with the intact mass of the protein, which precludes the characterization of proteoforms²⁵³. Thus, top-down and middle-down mass spectrometry (TD/MD-MS) have emerged as complementary techniques since these strategies envision the fragmentation of the proteins at the intact level (TD-MS) or after a limited proteolysis process (MD-MS)^{9, 10, 439, 596}. These strategies allow correlating the information derived from the fragmentation spectrum with the intact mass of either the protein or the different subunits. However, the fragmentation of intact proteins is still considered a challenging task, especially when dealing with proteins with molecular masses greater than 30 kDa, thus requiring the use of complementary fragmentation techniques in order to enhance the characterization of the primary sequence.

The introduction of various fragmentation techniques relying on different principles, such as collision-based dissociation, electron-based dissociation or photodissociation played a key role in the development of TD/MD-MS methods. While their fragmentation performances vary depending on the activation time, the experimental conditions or the sample, combination of results obtained from different dissociation modes demonstrates a substantial increase of the sequence coverage^{10, 441}. It illustrates the complementary nature of the fragmentation pathways specific to each technique for protein fragmentation, and its benefits for extensive protein characterization.

For all these reasons, it is of utmost importance to dispose of versatile MS platforms accommodating multiple and complementary fragmentation techniques. In our particular case, the geometry of the Orbitrap Eclipse Tribrid (Thermo Fisher Scientific) instrument allows to use multiple activation techniques, namely HCD, CID, ETD, EThcD/ETciD, UVPD, and PTCR, which can be combined for comprehensive protein characterization (**Figure 88**).

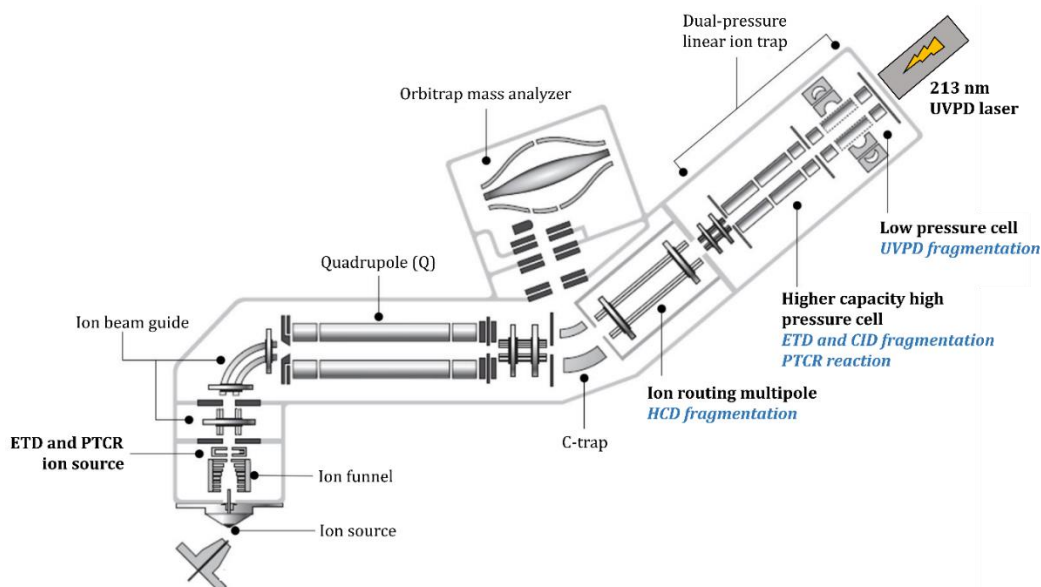


Figure 88: Schematic representation of the Orbitrap Eclipse Tribrid instrument (Thermo Fisher Scientific).

HCD is performed in the ion routing multipole (IRM) while other activations take place in the dual-pressure linear ion trap. For ETD fragmentation, fluoranthene radical anion (202 m/z), generated right after the source glow discharge, is used as electron carrier/donor. Similarly, PTCR reactant (perfluoroperhydrophenanthrene, PFPP, 623.9 m/z) is also located in the source and ions are produced using the same discharge ion source as ETD. The instrument is also equipped with a UVPD source employing a 213 nm laser with 2.5 kHz rate delivering $>1.2 \mu\text{J}/\text{pulse}$, located at the back of the instrument. The Orbitrap Eclipse Tribrid also enables to select fragment ions from MS2 to apply additional fragmentation steps (MS^n). Therefore, the Orbitrap Eclipse platform was chosen as the reference system for TD-MD/MS experiments, taking into account all the aforementioned fragmentation capabilities.

Extensive protein sequence characterization requires numerous parameters to be finely tuned. We optimized some key parameters to maximize the sequence coverage in this context. First, the activation time and/or energy required for each fragmentation technique and the influence of precursor ions selection were carefully optimized on the analysis of reference proteins by infusion. In addition, we selected potentially informative fragment ions obtained after the first fragmentation of the intact protein to undergo an additional fragmentation step (**Figure 89.A**). Second, similar optimizations were performed for MD-MS experiments using LC separation prior to MS analysis (**Figure 89.B**). In addition, we assessed the complementarity of the various fragmentation techniques as well as the repeatability between technical replicates of TD-MD/MS experiments. Finally, we implemented PTCR after fragmentation and evaluated its benefits to reduce spectral complexity (**Figure 89.C**).

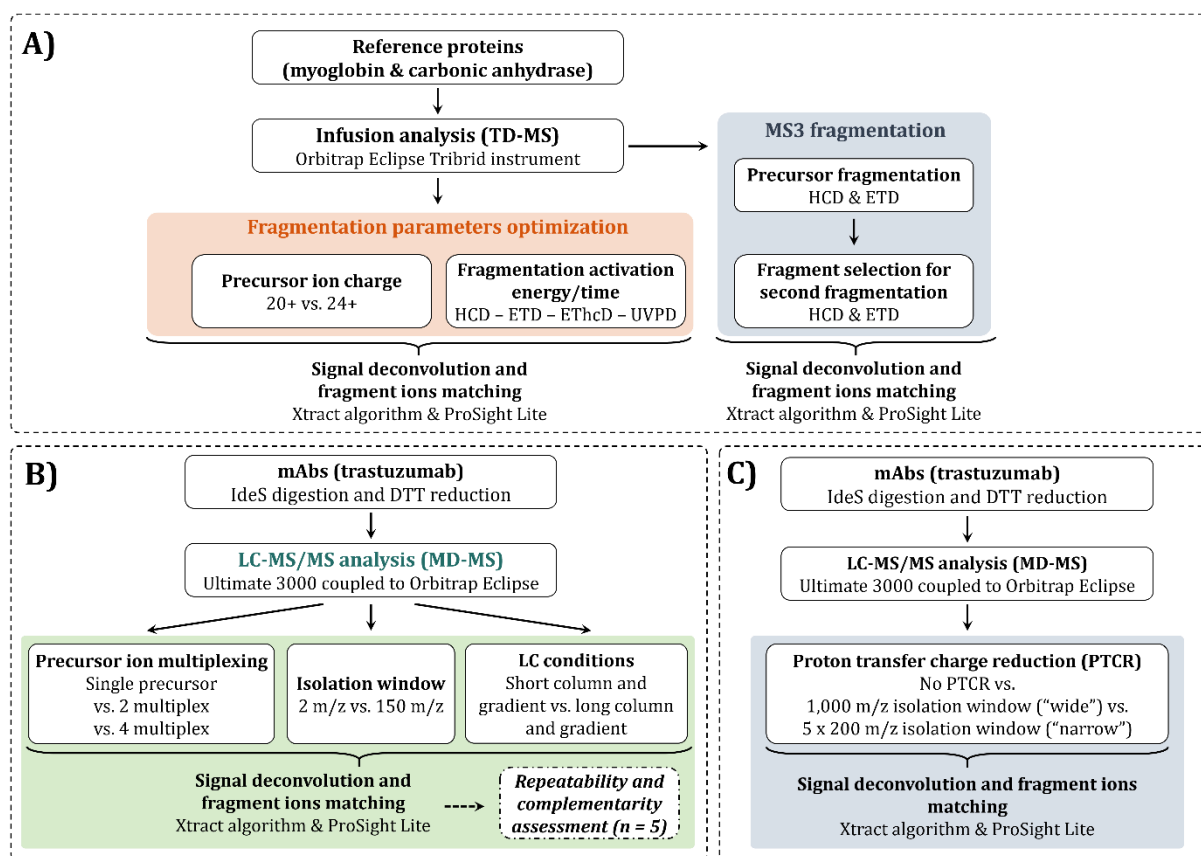


Figure 89: Analytical strategy used for the optimization of top-down and middle-down approaches. (A) Myoglobin and carbonic anhydrase analyzed by infusion to optimize the choice of the precursor ion and the fragmentation energy. In addition, MS3 fragmentation was evaluated on carbonic anhydrase to see the potential benefits of an additional fragmentation step. (B) Trastuzumab subunits were analyzed by LC-MS/MS to assess precursor ion multiplexing and evaluate the influence of liquid chromatography prior to MS analysis. Repeatability and complementarity of fragmentation methods were also investigated. (C) Trastuzumab subunits were submitted to PTCT using wide and narrow isolation windows. All MS/MS spectra were deconvoluted using Xtract and fragment ion matching was performed with ProSight Lite.

2. Optimization of the key parameters for top-down data acquisition by infusion of reference proteins

A. Influence of the fragmentation parameters

Many parameters must be tuned to obtain a broad sequence coverage. First, we focused our efforts on specific parameters related to each fragmentation technique, i.e., the activation energy (for collision-based fragmentations), the reaction time (for electron-based fragmentations), or the activation time (for UVPD). We also evaluated the influence of the charge of the precursor ion on the fragmentation techniques. All these parameters are crucial to obtain optimal fragmentation and directly influence the number of fragments generated as well as the quality of the signal^{308, 351, 432}. We first performed these optimizations on myoglobin, a reference protein usually used for method development. It was analyzed by infusion to ensure a constant signal over a long enough period, thus maximizing the S/N ratio.

i) Optimization of the precursor ion charge and intensity

Myoglobin from *Equus caballus* comprises 153 amino acids and has a monoisotopic molecular weight of 16,940.96 Da. **Figure 90** depicts the MS spectra of myoglobin in denaturing conditions. Peaks corresponding to the myoglobin charge state envelope from 9+ to 26+ are observed. Typically, a good trade-off between ion intensity and charge is well suited for TD fragmentation. On one hand, this selection relies on the fact that some fragmentation techniques are sensitive to the net charge state of the precursor ion, with generally slightly better sequence coverage associated to ions with higher charge states. On the other hand, intense precursor ions are preferred in order to obtain fragment ions with enough intensity to be precisely deconvoluted and then identified. For this reason, low abundant high and low charged states ions are generally excluded

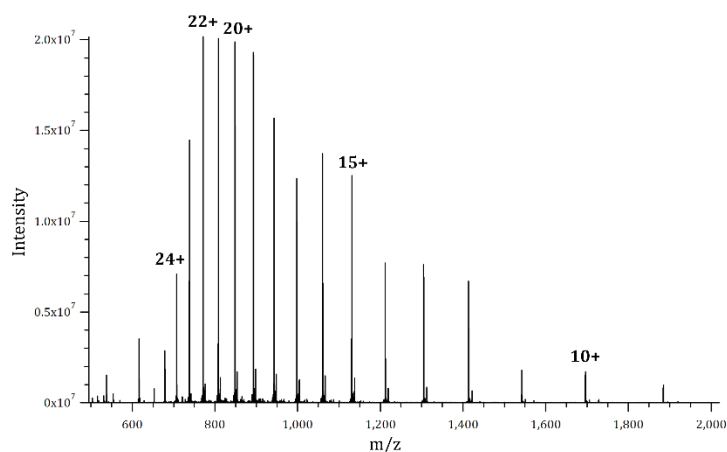


Figure 90: Full MS spectrum of myoglobin.

Here, we selected three precursors (24+, 22+ and 20+), with a significant intensity to ensure fragment ions with an optimal S/N ratio, in order to evaluate the influence of the charge state. Very different fragmentation patterns were observed for the three charge states selected and a different fragment ions distribution after deconvolution using HCD and ETD fragmentation (**Figure 91**). HCD activation of the 24+ precursor generates more low-mass fragment ions than the 20+ and 22+ precursors using the same collision energy. Similarly to HCD, ETD fragmentation of the 24+ charge state provides a very rich MS/MS spectrum with intense fragment ions distributed all over the mass range, while the lower charge states fragmentation spectra (20+ and 22+) are dominated by charge reduced species and less efficient fragmentation (i.e., non-dissociative ETD (ETnoD)) (particularly charge state 20+). It results in low-intensity fragment ions compared to the remaining amount of intact precursor ion, as depicted in **Figure 91**. Contrary to ETD and HCD, UVPD fragmentation spectra are similar independently of the charge state, leading to similar deconvolution results.

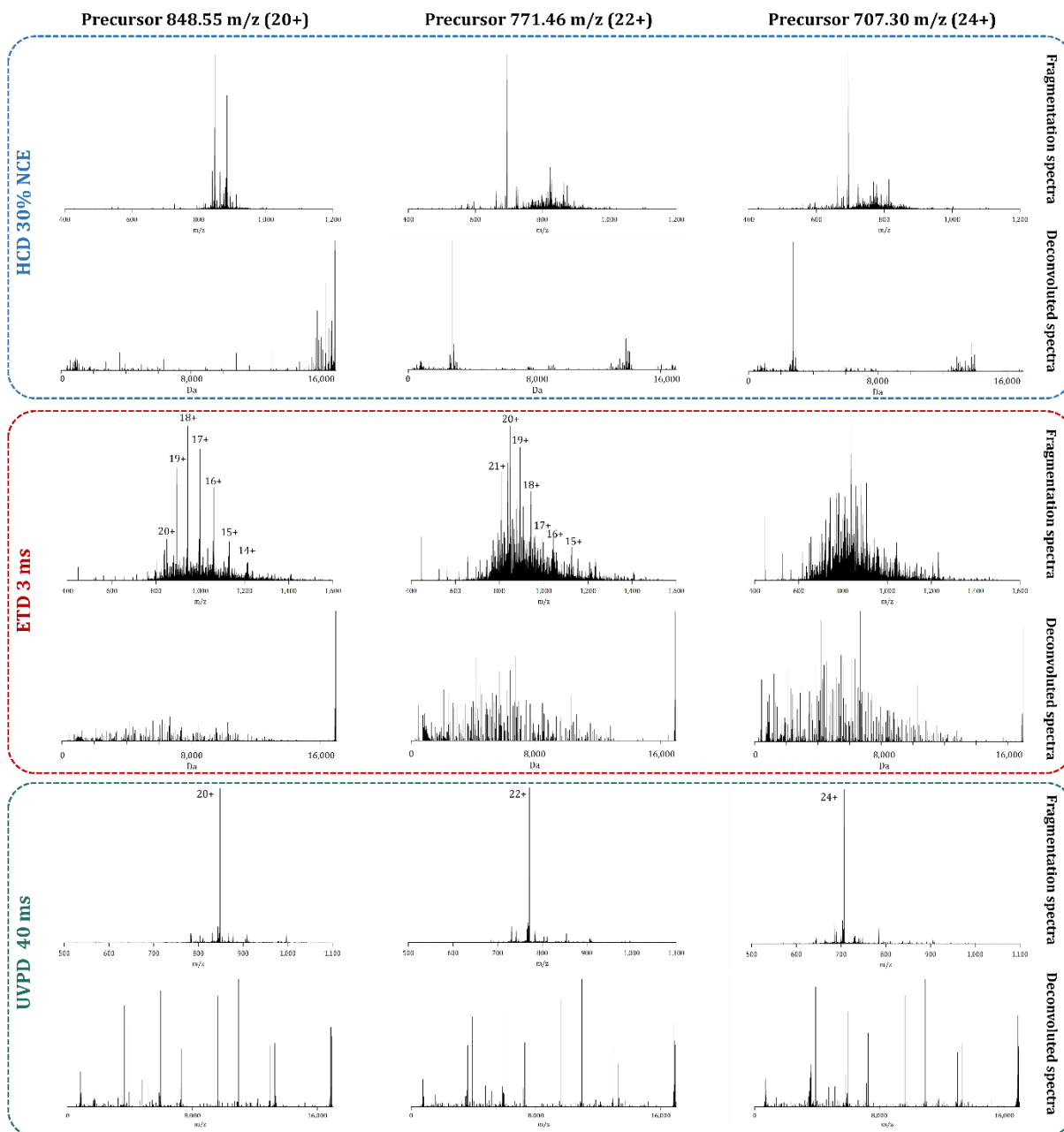


Figure 91: Fragmentation spectra and deconvoluted spectra obtained for charge states 20+, 22+ and 24+ of myoglobin using HCD 30% NCE (upper panel), ETD 3 ms (middle panel) and UVPD 40 ms (lower panel).

Because of the differences observed in the fragmentation patterns, we compared the sequence coverage obtained for each charge state and the cleavage site distribution (**Figure 92**). The sequence coverage is improved with the fragmentation of a higher charge state precursor ion for ETD, while more intense ions give the maximum backbone dissociation for HCD. Using UVPD, comparable sequence coverages are obtained independently of the charge state. Low mass fragments observed in HCD with precursors 22+ and 24+ are helpful to improve sequence coverage near the N- and C-termini, whereas the 20+ precursor give higher mass fragments along the protein (**Figure 92**). Nevertheless, the unidentified low mass fragments obtained from deconvolution can also result from the over fragmentation of some ions, and do not bring additional information.

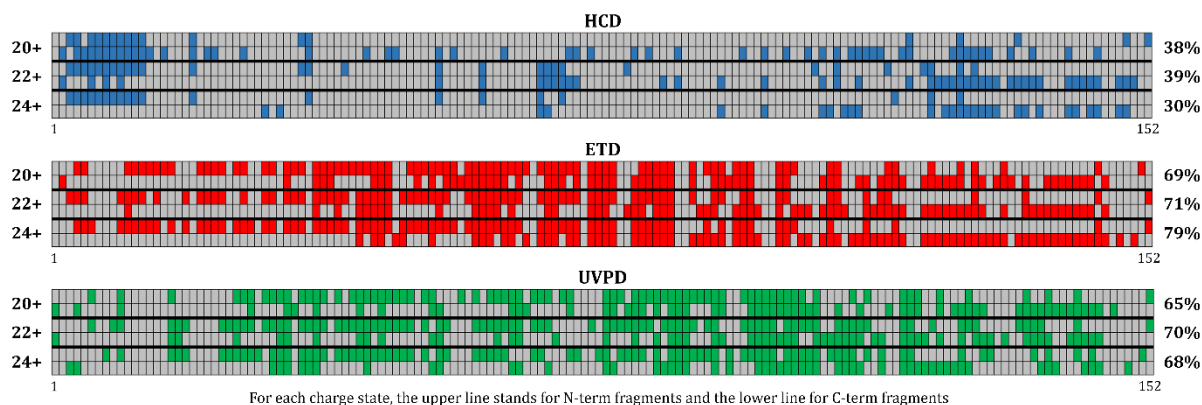


Figure 92: Cleavage sites map of each charge state using HCD (upper panel), ETD (middle panel) and UVPD (lower panel). For each charge state, the upper line represent N-terminal fragment ions (i.e., *a*, *b* and *c*) and the lower line C-terminal fragment ions (i.e., *x*, *y* and *z*). Total sequence coverage for each charge state is displayed on the right hand side of the figure.

The complementarity of the cleavage sites for HCD depending on the charge state allow increasing the sequence coverage to 57% when results from each precursor ion are combined. Indeed, different fragmentation channels are induced as a function of the charge state, giving complementary information^{259, 271, 285, 597}. Similarly, combination of results from each charge state in ETD and UVPD results leads to a total sequence coverage of 85%. The gain in sequence coverage observed for these two techniques is lower considering that both techniques provide an extensive backbone fragmentation obtained with an unique charge state.

According to the results, the selection of highly charged precursors provides a high sequence coverage, particularly for ETD, according to the fact that the electron transfer cross section is proportional to the square of the charge^{277, 288}. For HCD, a too-highly charged precursor submitted to high energy will undergo over-fragmentation, leading to non-informative fragment ions and decreasing the sequence coverage. Thus, the charge state precursor is not the only parameter to play with to increase the sequence coverage, as it should be combined with fine-tuning the activation energy.

ii) Optimization of the fragmentation parameters

As mentioned, various fragmentation techniques are available on the Orbitrap Eclipse Tribrid. After selecting the best precursor ions for fragmentation, we optimized the activation parameters for HCD, ETD, EThcD, and UVPD fragmentations.

For each fragmentation technique, multiple collision energies (HCD and EThcD), reaction time (ETD and EThcD), and activation time (UVPD) have been tested. **Figure 93** displays the sequence coverage and the number of fragment ions generated by each fragmentation method and for each reaction conditions. First and foremost, we observe that all the techniques behave mainly in the same way. At the beginning, when low amounts of energy are used, the sequence coverage correlates almost linearly with the number of fragment ions. Then, the sequence coverage starts decreasing upon reaching the maximum value while the number of fragment ions continuously rises. Indeed, large amounts of energy will lead to an over-fragmentation of the precursor, concomitantly generating shorter terminal fragments along with the generation of internal fragments that will not contribute to the characterization of protein sequence (taking into account that only terminal fragments are used in this case). In contrast, too-low energy conditions will result in limited fragmentation, which will lead to the generation of the most favored fragmentation pathways.

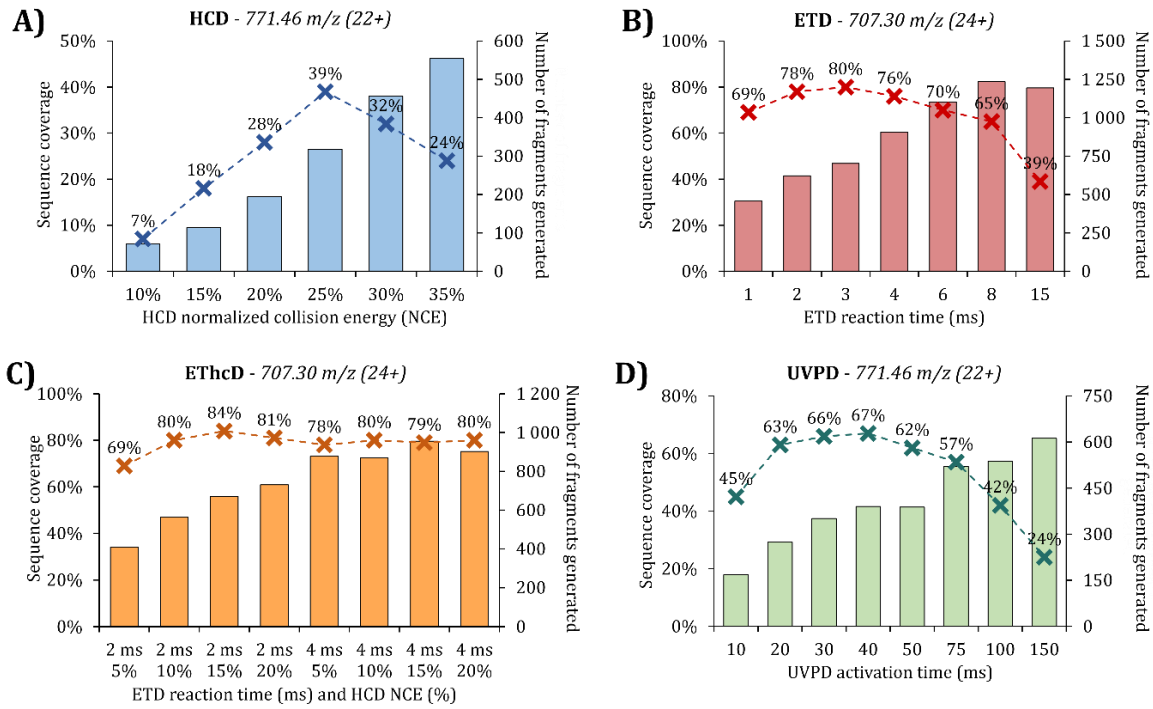


Figure 93: Myoglobin fragmentation energy optimizations results. Sequence coverage obtained (A) after HCD (10%-35% NCE) on precursor 22+, (B) after ETD (1-15 ms) on precursor 24+, (C) after EThcD (2-4 ms, 5%-20% NCE) on precursor 24+, and (D) after UVPD (10-150 ms) on precursor 22+.

To further illustrate this phenomenon, the fragmentation spectra of myoglobin upon fragmentation with several ETD reaction times are displayed in **Figure 94**. Only very low abundant fragment ions are observed for the lowest reaction time (1 ms), with principally charge reduction occurring. With the increase of the reaction time, more fragments are generated, making the spectra more complex. The results for the longest reaction times (8 and 15 ms) show smaller fragments with an increase in the number of unmatched fragment ions due to over-fragmentation, hampering the identification of cleavage sites on the mid-region of the sequence (**Figure 94**).

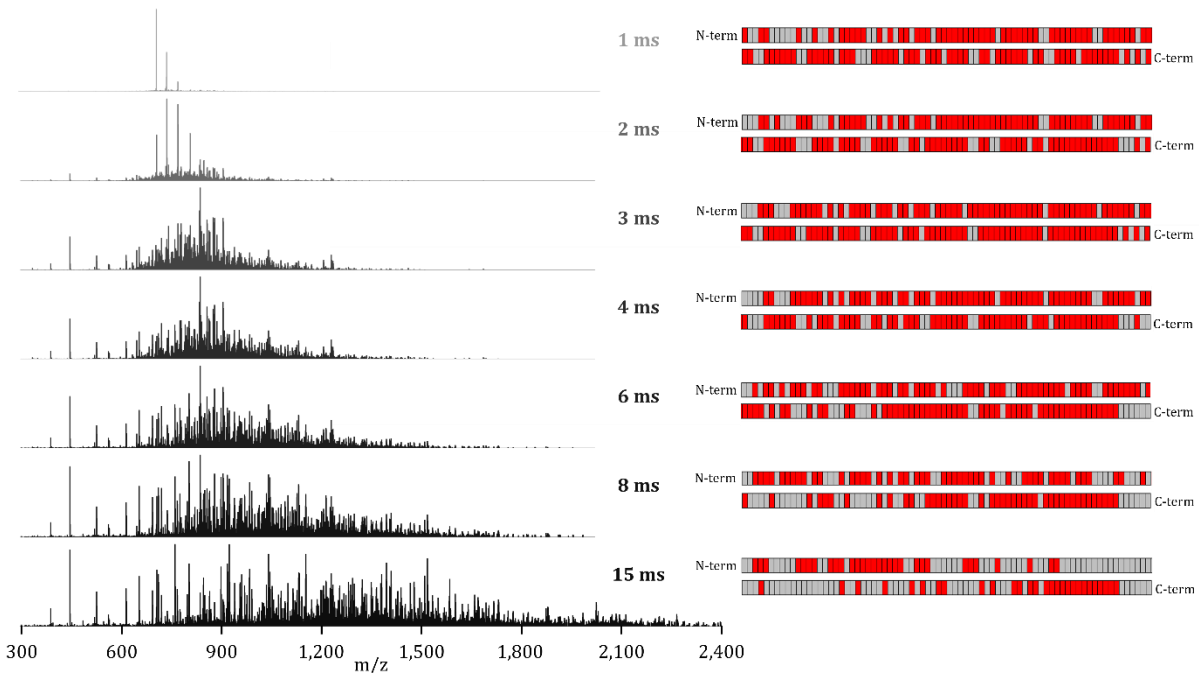


Figure 94: ETD fragmentation spectra of charge state 24+ of myoglobin using ETD reaction time from 1 to 15 ms (left) and the associated map of the cleavage sites (right).

Considering these results presented above, we can identify the optimal parameters for each technique (**Figure 93**). HCD leads to the lowest sequence coverage, with a maximum of up to 39%. In contrast, UVPD provides more than 65% of backbone cleavage, while ETD and EThcD allow the characterization of more than 80% of the sequence. These findings have been also corroborated in previous publications, highlighting the suitability of electron-driven and photo-fragmentation to provide extensive sequence characterization^{257, 310, 350, 598}.

It is worth noticing that the most efficient fragmentation technique in terms of depletion of the precursor ion, does not provide the most informative fragmentation spectrum as illustrated in **Figure 95**. HCD is a very efficient fragmentation technique, giving rise to the complete depletion of the selected precursor ion, whereas UVPD is much less efficient. Conversely, this approach is not the most informative as few and most favored fragmentation channels are observed, resulting in low sequence coverages. On the contrary, the UVPD MS/MS spectrum displays a highly abundant peak corresponding to the precursor, revealing the difficulty of fully fragmenting the precursor with this technique. However, this method is very informative as it more closely reflects statistical fragmentation. Spectra obtained from electron-based techniques (i.e., ETD and EThcD) are rich, exhibiting many intense fragment ions over a wide mass range. These methods combine a relatively high efficiency while being very informative.

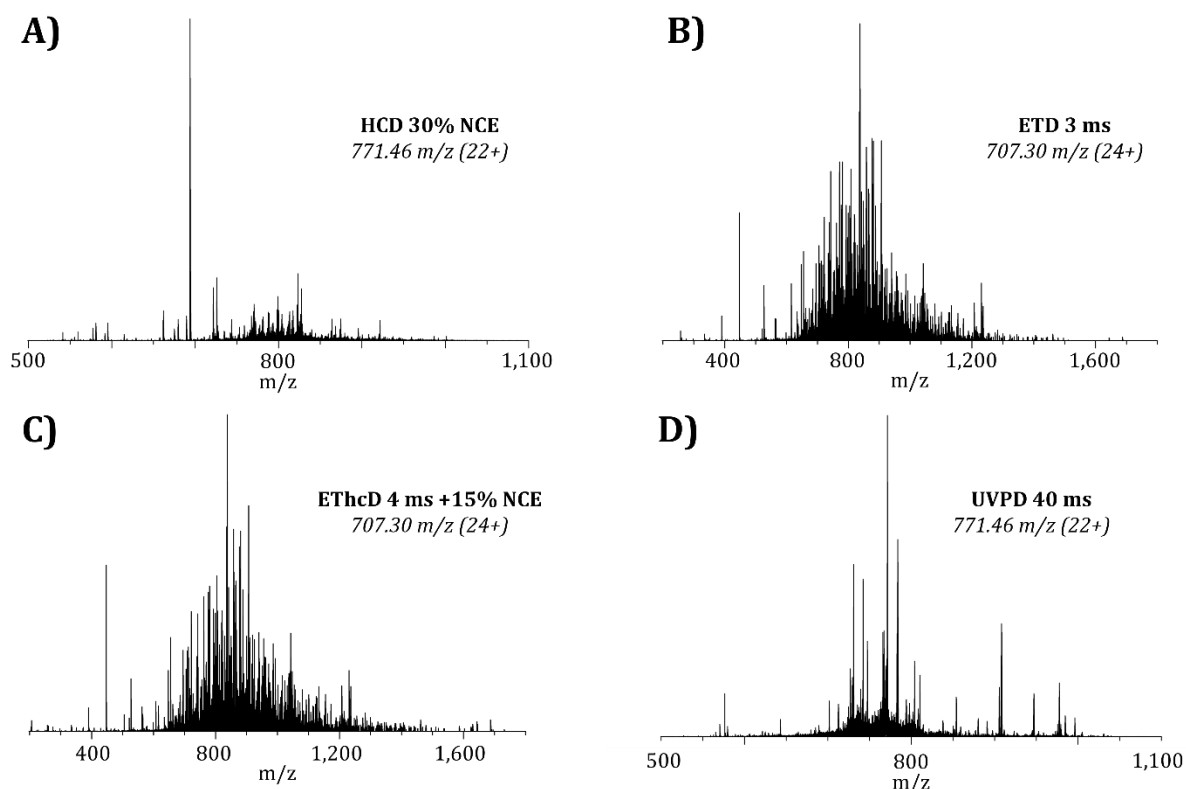


Figure 95: Fragmentation spectra of myoglobin after (A) HCD NCE 30% (ion 22+), (B) ETD 3 ms (ion 24+), (C) EThcD 4 ms + 15% NCE (ion 24+) and (D) UVPD 40 ms (ion 22+).

iii) Parallel optimization between precursor ion charge and energy reaction for an enhanced sequence coverage

The results presented in this section demonstrate the importance of the choice of the precursor ion and the fragmentation parameters optimizations to obtain the maximum sequence coverage. However, the choice of fragmentation to obtain a comprehensive backbone fragmentation is directly linked to the precursor selected, and vice versa. None of the parameters can be optimized before the other.

Hence, we monitored the sequence coverage evolution by varying the precursor charge state and the activation energy simultaneously, as displayed in **Figure 96**.

First, the best sequence coverage is obtained with charge state 22+ for HCD (39%) and UVPD (67%) and with charge state 24+ for ETD (80%) and EThcD (84%). The result from HCD corroborates the previous assumption that the numerous small fragment ions (< 800 Da) observed with the high charge state are not bringing extensive information on the sequence and mostly come from the re-fragmentation of larger fragments³³². Second, all fragmentation techniques underline that a higher charge state precursor requires less energy to reach a maximum sequence coverage, except for UVPD.

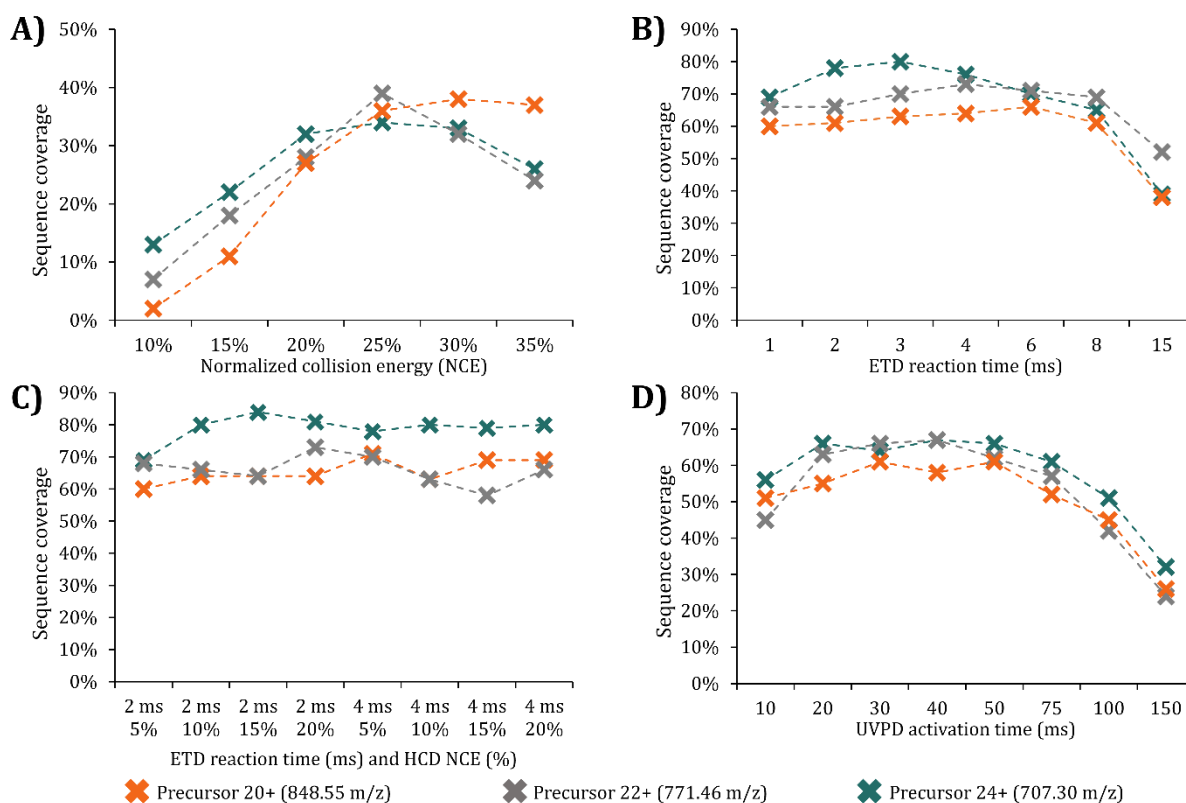


Figure 96: Myoglobin sequence coverage of charge states 20+, 22+ and 24+ depending on the dissociation energy. Sequence coverages obtained using (A) HCD, (B) ETD, (C) EThcD and (D) UVPD.

Indeed, for HCD, with higher charge states, more mobile protons are available to allow charge-directed cleavages, and consequently, less acceleration voltage will be required to promote fragmentation⁵⁹. Similarly, as electron-based dissociation relies on an ion/ion reaction, the number of charges within the protein molecular ion are crucial to undergo fragmentation. The higher charge states result in an increased energy surface on which the reaction can occur, and consequently, a lower reaction time is required⁵⁹⁷. Nevertheless, a too-highly charged precursor submitted to high energy will undergo over-fragmentation, leading to non-informative fragment ions and decreasing the sequence coverage. Conversely, UVPD fragmentation is not a charge-dependent reaction, explaining the similar sequence coverages obtained throughout all the different precursor charge states, with variation of the percentage of cleavages mostly related to the activation time.

However, myoglobin is a small protein containing neither cysteines nor PTMs. Hence, the same strategy was used to characterize a larger protein, carbonic anhydrase (*Bos taurus*, UniProt ID P00921), a 29 kDa protein bearing an acetylation in the N-terminal site. Therefore, two precursors with different charge states (42+ and 34+) were fragmented at multiple energies. A summary of the results is presented in **Figure 97**. Other charge states and activation energies have been tested but are not detailed in this manuscript.

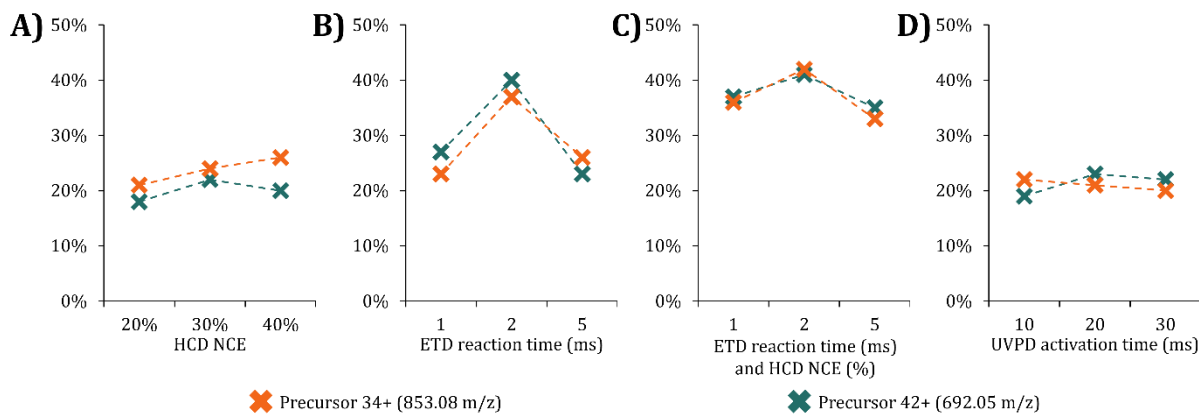


Figure 97: Carbonic anhydrase sequence coverage of charge states 34+ and 42+ depending on the dissociation energy. Sequence coverages obtained using (A) HCD, (B) ETD, (C) EThcD and (D) UVPD.

Globally, carbonic anhydrase data present the same trend observed for myoglobin. Lower charge states require more energy to fragment efficiently. In addition, the high charge states selected (42+ and 34+) directly influence the reaction time for electron-based dissociation, with optimal fragmentation obtained with 2 ms of reaction, against 3-5 ms for myoglobin. In addition, sequence coverage obtained by UVPD is similar for the two charge states. Even if electron-based methods provide the best results, the global sequence coverage observed is relatively low (< 50%)³⁵¹; pointing out the difficulty of fragmenting comprehensively the protein backbone for large molecules⁵⁹⁹.

B. Development of MS3 approaches to increase protein sequence coverage

Extensive characterization of large proteins by TD-MS cannot be achieved straightforwardly using conventional TD-MS strategies. To tackle this challenge, an additional fragmentation step on ions of interest generated by the fragmentation of a precursor appears as a promising approach to increase the sequence coverage. It consists of the selection of a specific ion of interest to undergo an additional fragmentation step and then increase the backbone cleavage of the protein. The whole process is briefly illustrated in **Figure 98**. Full MS spectra are acquired to verify the mass of the protein. Then, a precursor charge state is selected for fragmentation (e.g., HCD or ETD in this case). It is essential to mention that the selected precursor for MS3 analysis must be highly charged for a better fragmentation, and relatively intense to mitigate the fragment ion loss during the subsequent isolation and fragmentation. After the deconvolution of the MS2, we obtain a fragmentation map. Finally, one fragment ion that could be one potential interest – because it bears a PTM or it contains a bunch of the sequence that is not well characterized- is selected and fragmented again, using either the same activation technique or an alternative one.

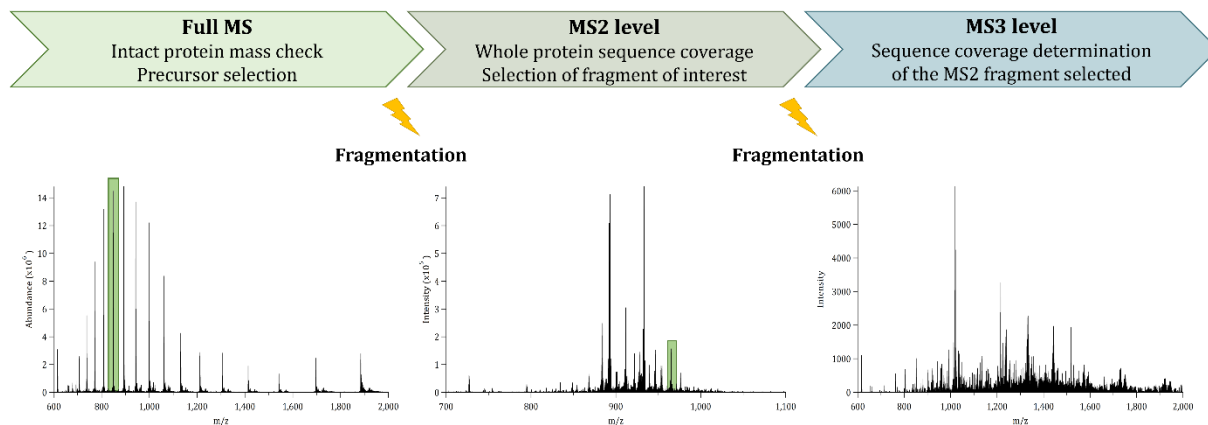


Figure 98: Schematic representation of the MS3 fragmentation approach.

For the proof of concept, we used bovine carbonic anhydrase II. We selected charge state 36+ (807.22 m/z) for fragmentation using HCD (30% NCE) and ETD (1.5 ms), resulting in a sequence coverage of 21% and 40%, respectively. UVPD was not selected to fragment the protein ion precursor for different reasons including its long reaction time and the low intensity of the generated fragments. After combination of HCD and ETD results, the sequence coverage increases to 52% (**Figure 99.A and B**). Nevertheless, some regions of the sequence remain poorly characterized (e.g., residues 1-13, 100-130, or 150-190). We selected a fragment ion from the MS2 spectra to be fragmented again and generate residue cleavages in these regions. Several criteria drove the choice of the MS2 ion that will undergo further fragmentation. It should be a terminal fragment containing the poorly sequenced region, which is intense and charged enough to allow further fragmentation and does not overlap with other ions. Indeed, precursor's co-isolation would give a more complex spectrum containing fragment ions from multiple precursors that could potentially increase the risk of false positive. In our case, numerous ions from the mid-region of the sequence (i.e., amino acids 110-160) are potentially available, as exhibited in **Figure 99.B**.

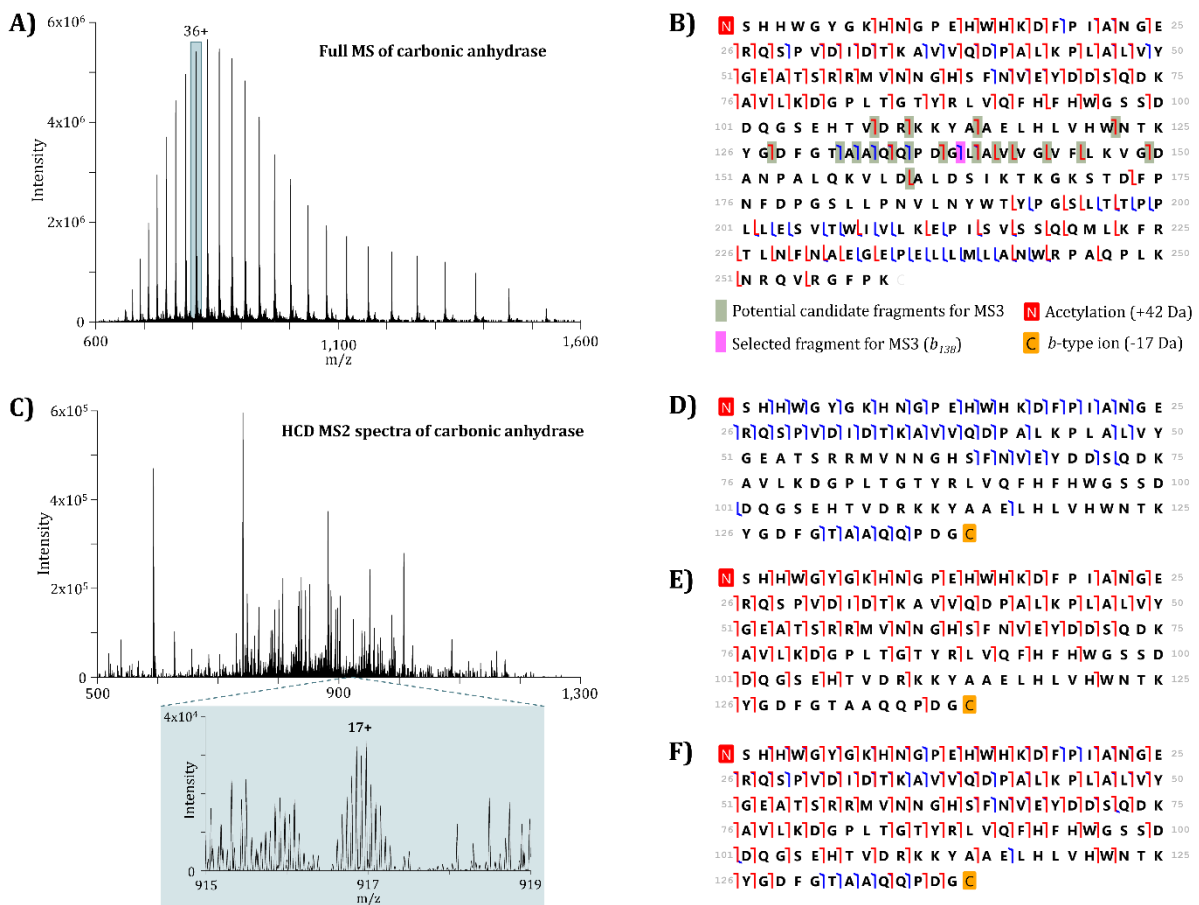


Figure 99: Carbonic anhydrase fragmentation using the MS3 approach. (A) Full MS spectrum of carbonic anhydrase. (B) Fragmentation map of the combination of HCD and ETD fragmentation of the 36+ precursor. (C) HCD MS2 fragmentation spectrum of the 36+ precursor and zoom on the *b*₁₃₈ fragment (917 m/z, 17+) selected to undergo further fragmentation (inset). (D) HCD and (E) ETD MS3 fragmentation map of *b*₁₃₈ fragment. (F) Fragmentation map combining results of ETD/HCD MS2 and MS3 fragmentation on the residues 1-138 of carbonic anhydrase.

However, only a few of them meet the requirements for a good MS3 precursor candidate. Ions from ETD fragmentation (i.e., *c* and *z* ions) are generally low abundant and overlapping with other fragment ions (as already described in **Part III, Chapter 1.4**), making them unsuitable for further fragmentation. In contrast, some fragments ions from HCD have an intense signal, with no interference from other

fragment ions, and highly charged, making them suitable for MS3 analysis. Here, we selected the intense and isotopically resolved 17+ ion at 917.91 m/z, corresponding to the b_{138} fragment (15,578.52 Da) to perform MS3 using HCD and ETD (**Figure 99.C**). Deconvolution of the MS3 spectra obtained results on sequence coverage of 37% for HCD and 60% for ETD on the region of the sequence selected (**Figure 99.D and E**). Once combined, 80% of backbone cleavage is observed on this part of the sequence (residues 1-138), while only 60% of sequence coverage was achieved using MS2 (**Figure 99.F**). Considering the whole sequence, residue cleavages were improved by 11%, from 52% to 63%. In addition, MS3 allows for characterizing previously poorly fragmented regions, particularly for the fifteen amino acids on the N-terminal part and for residues 100 to 110. It also evidences the position of the acetylation on the N-terminal side of the protein. The choice of two complementary activation techniques for MS2 and MS3 fragmentation is also beneficial, given the differences in the fragmentation pathways.

Regarding these results, the potential of MS3 to improve the sequence coverage of specific regions of the sequence of a protein, and thus having an impact overall sequence coverage ratio is clearly illustrated. Furthermore, additional fragmentation steps can be of great interest to confirm the presence of sequence variants modifications and strengthen the confidence to elucidate their precise location. MSn workflows could find their utility when dealing with the fragmentation of large intact proteins (> 30 kDa) since the sequence coverage normally decreases with the mass of the proteins^{257, 338, 600}. Nevertheless, the use of MS3 is restricted to certain conditions, particularly regarding the choice of the precursor that will undergo the second fragmentation step. Finally, while suitable for working in infusion conditions, it is intricate to implement this approach using liquid chromatography separation. Under the latter conditions, the fragmentation spectra can be recorded during the elution time of the analytes, which reduce the acquisition time and the S/N ratio, will be lower after MS3. More particularly, depending on the activation technique and the experimental parameters, the number of spectra recorded during a LC-MS/MS/MS could be reduced between 2-5 times.

3. Optimization of the acquisition parameters for middle-down LC-MS/MS analysis of monoclonal antibody

A. Implementation of LC separation before MD-MS for mAb analysis

While analysis of reference intact proteins by infusion is reasonably straightforward, mAb analysis is considerably more challenging for many reasons. Samples are often provided in specific buffers containing salts or detergents, to maintain their structure and ensure their stability, that are incompatible with MS analysis. Moreover, they are biomolecules around 150 kDa, and many studies have already demonstrated the limitations of TD-MS for exhaustive sequencing of intact mAbs^{428, 433, 601}. They also contain disulfide bridges between cysteines, preventing the unfolding of the molecule and, consequently, an efficient fragmentation.

Several strategies can be employed to tackle these challenges and allow mAbs analysis. Middle-down MS, consisting on the analysis of mAb subunits of 25-50 kDa is the most widely used approach. Here, we performed limited proteolysis of our mAb, trastuzumab, followed by the reduction of the disulfide bridges using DTT (**Figure 100.A**). This workflow allows the release of three subunits around 25 kDa, namely light chain (Lc), fragment crystallizable region (Fc), and Fd (N-terminal part of the heavy chain) subunits. Thus, we used reverse-phase LC (HPLC) with a short gradient of 12 min on a 2.1 x 50 mm column (Zorbax 300 SB-C8, 1.8 μ m diameter particles, Agilent) with a flow rate of 100 μ L/min to separate the three subunits (**Figure 100.B**). The different masses corresponding to each chromatographic peak allowed to identify the three subunits (**Figure 100.C-E**). For the Fc/2 subunit, 4 glycoforms were observed (G0F-N/G0F, G0F/G0F, G0F/G1F and G0F/G2F, **Figure 100.C**), but only the most intense (G0F/G0F) was selected for further fragmentation.

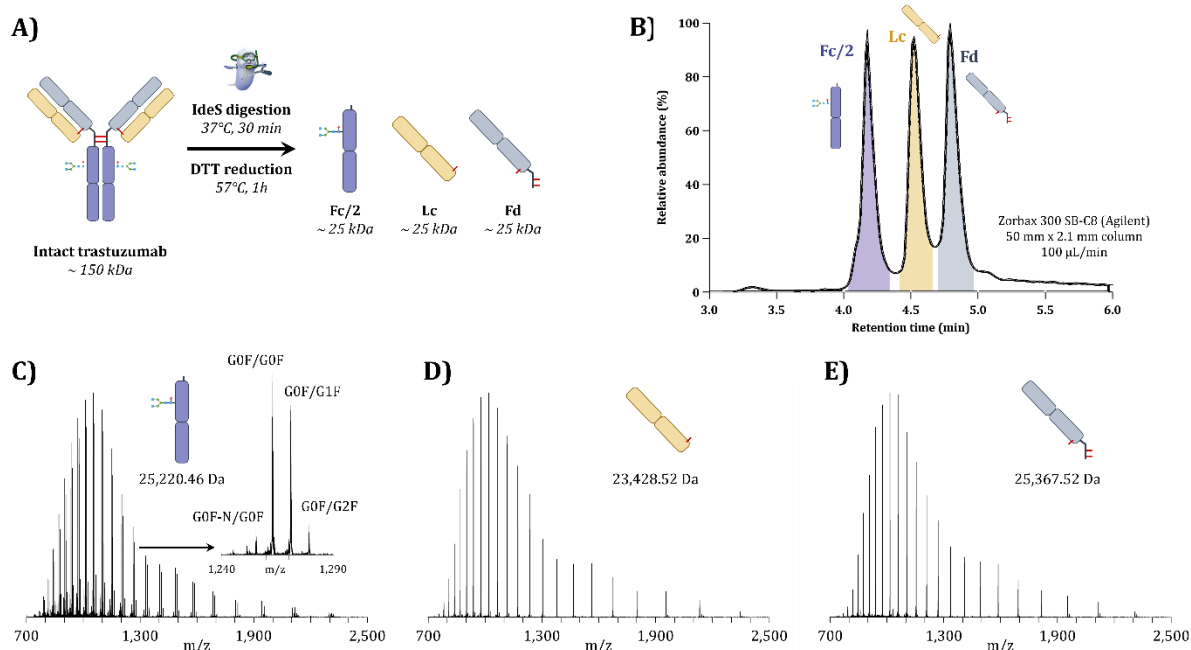


Figure 100: Trastuzumab analysis by LC-MS. (A) Intact trastuzumab IdeS digestion and DTT reduction to obtain mAb subunits, i.e., Fc/2, Lc and Fd. (B) TIC chromatogram of the subunits analysis. (C) Fc/2, (D) Lc and (E) Fd intact subunits MS spectra.

Chromatographic separation before MD-MS limits the time available for fragmentation, and consequently, fewer MS/MS spectra can be averaged, potentially leading to a lower sequence coverage. In this context, optimizations regarding the isolation window width and the number of precursor ions selected (multiplexing) were performed to take full advantage of the different fragmentation propensity of precursor ions with different charge states (**Part III, Chapter 1.A.1**). Moreover, the chromatographic conditions were also modified to extend the time available for fragmentation.

i) Influence of precursor ion multiplexing

Considering that the fragmentation of several precursors using the same activation method can increase the sequence coverage⁴³², we evaluated the influence of performing ion multiplexing. For this purpose, 2 and 4 charge states were selected prior to MD-MS analysis and compared the results obtained with the fragmentation of one single charge state (**Table 9**). The best activation conditions were selected for each fragmentation technique.

Table 9: Precursor charge states selected for multiplexing experiments.

	Fc/2	Lc	Fd
Single charge state	23+	22+	21+
2 charge states multiplexing	23+ & 22+	22+ & 21+	21+ & 20+
4 charge states multiplexing	23+, 22+, 21+ & 20+	22+, 21+, 20+ & 19+	21+, 20+, 19+ & 18+

Independently of the subunit, it was observed that multiplexing of precursor ions does not improve the sequence coverage (**Figure 101**). When two different charge states are selected, the sequence coverage of the different subunits decreases of 2%, 6%, and 7% on average with HCD, ETD, and UVPD respectively. The outcome is similar for HCD, and ETD fragmentations when four different precursor ions are selected. However, in the case of UVPD this effect is even more marked since 13% of sequence coverage decrease was obtained.

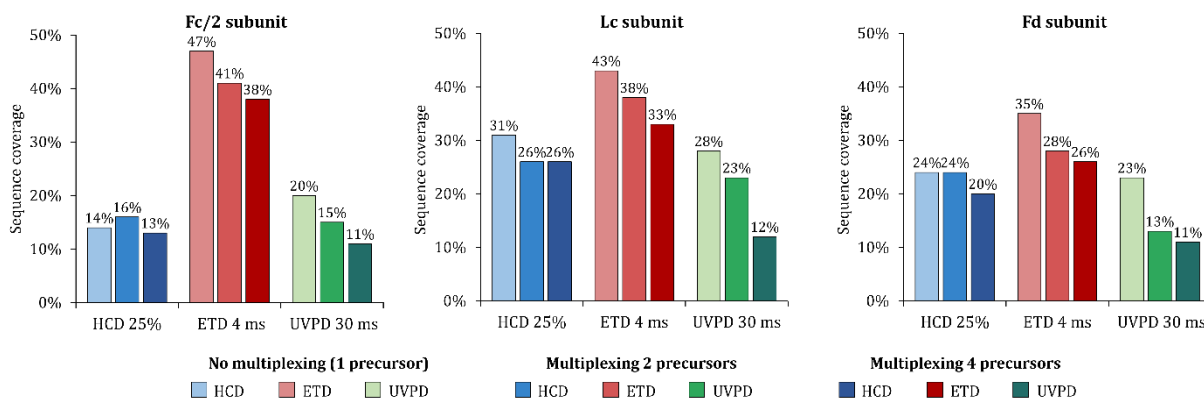


Figure 101: Sequence coverage results for the Fc/2, Lc and Fd subunits after single charge state, 2 charge states and 4 charge states multiplexing fragmentation.

The sequence coverage drop observed, mainly when using ETD and UVPD, results from implementing the multiplexing on a chromatographic time scale. Therefore, the chromatographic peak width limits the number of scans that can be accumulated. In detail, ETD, and UVPD even more, require greater activation times to induce the fragmentation of the precursor ions. Consequently, they lead to fewer spectra averaged over the chromatographic peak than HCD. For instance, 20 MS/MS spectra from HCD, 18 from ETD, and 11 from UVPD are acquired over the Fc/2 chromatographic peak. Moreover, ETD and UVPD techniques are less efficient than HCD (in terms of depletion of presence of the precursor ion in the MS/MS spectra) and generate more fragment ions (more fragmentation channels) that can differ from one precursor to another. Consequently, the resulting fragment ions are of low intensity. The combination of low-intensity ions and a low number of averaged spectra induce a lower S/N ratio for the fragment ions from UVPD and ETD, directly affecting the sequence coverage. Even so, some studies already showed that multiplexing of precursor ions can improve the sequence coverage under specific conditions⁴³².

ii) Influence of the isolation window width

To circumvent the limitation of multiplexing, i.e., the lower number of MS/MS spectra performed for each precursor, we decided to use a wider isolation window⁴²⁹. In this case an isolation window of 150 m/z instead of 2 m/z was selected, thus including multiple charge states simultaneously.

Results comparing the use of 2 m/z and 150 m/z isolation windows are shown in **Figure 102**. Globally, slightly lower sequence coverages are observed with wide isolation windows. HCD and UVPD demonstrate a lower drop in the percentage of residue cleavages, between 2% and 5% (except for the LC using HCD, 8%), compared to ETD, with a drop of 8% and 7% for Fc/2 and Lc subunits, respectively. Indeed, using 150 m/z isolation windows, we also increase the overall background noise and thus affecting the S/N ratio of the fragment ions. These results show that wide isolation can slightly compensate for the limitations of multiplexing (more MS2 spectra from the same precursors), but it is still insufficient to improve the results obtained by selecting a single precursor in a narrow isolation window, which seems to remain the method of choice.

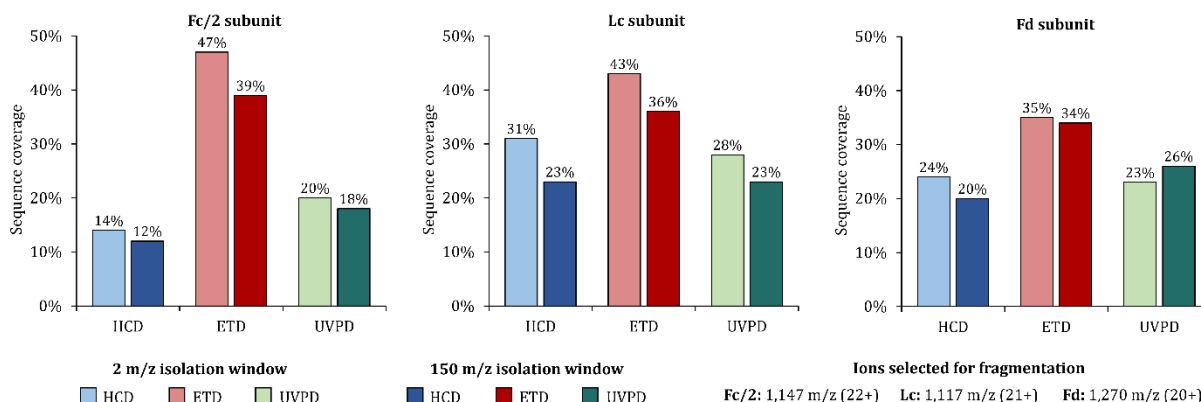


Figure 102: Sequence coverage results for the Fc/2, Lc and Fd subunits after single charge state fragmentation using 2 m/z and 150 m/z isolation windows.

iii) Optimization of the LC separation to improve sequence coverage

Fragmentation of mAb subunits with molecular masses around 25 kDa requires averaging the largest number of scans to extract low-intensity fragments from the background noise. One strategy to cope with this accumulation issue is using rpLC conditions to obtain broader peaks, and thus increase the acquisition time without affecting the subunit separation. With this in mind, we used a longer column (150 mm instead of 50 mm) and a longer gradient (40 min instead of 10 min), enabling us to average the spectrum over up to 35 s (FWMH ~12 sec) as shown in **Figure 103.A**. With the previous conditions, the chromatographic peaks were narrower, allowing signal averaging over 12 s (FWMH ~6 s) only (**Figure 100.B**). Consequently, more than 45 spectra are acquired over the whole chromatographic peak of the Fc/2 using HCD, for example, compared to 20 with the former chromatographic setup. It results in an improvement of the sequence coverage, as illustrated by **Figure 103.B**. UVPD results are not shown because the laser was replaced between the two sets of experiments, inducing bias in the comparison. As expected, the increase of residue cleavages is higher with ETD (between 5% and 12%) compared to HCD (between 1% and 11%), as the ETD fragmentation is more sensitive to acquisition time increase. These results confirmed the choice of these chromatographic parameters, which allow averaging many more spectra, improving the subunits sequencing.

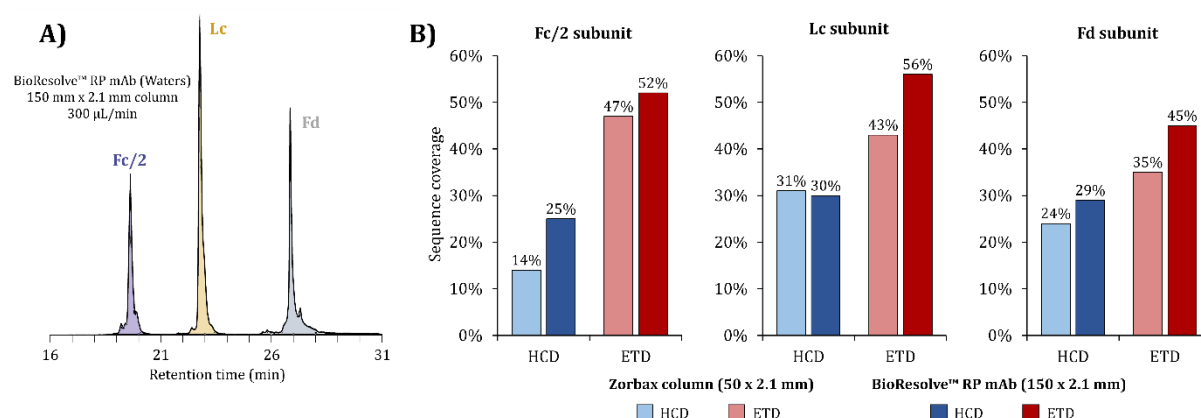


Figure 103: Trastuzumab subunit analysis using the BioResolve™ RP mAb column. (A) Fc/2, Lc and Fd subunits separation at 300 µL/min. (B) Sequence coverage results for Fc/2, Lc and Fd using the Zorbax column and the BioResolve™ column.

B. Complementarity and repeatability of mAb subunits fragmentation by LC-MD/MS

The optimization of the previously described parameters allow us to increase the sequence coverage. Hence, we used the optimal parameters to compare the benefits of using all the fragmentation

techniques to obtain the most extensive sequence coverage of each subunit. Each mAb subunit was fragmented with HCD, ETD, EThcD and UVPD to assess the ability of each technique to cleave residues. CID was also used to evaluate its ability for intact protein fragmentation. Five replicates of injection were performed for each technique.

Best sequence coverage obtained from each fragmentation is presented in **Figure 104**. Independently of the subunit, electron-based dissociation techniques (i.e., ETD and EThcD) are the ones providing the highest sequence coverage while collision-based fragmentations lead to the lowest cleavage residue. In detail, EThcD enable the sequencing of 49% of the Fd, 60% of the Fc/2 and 63% of the Lc subunits, corresponding to an improvement between 6% and 10% compared to ETD, depending of the subunit. This increase of the backbone cleavage with EThcD stems from the additional energy provided by HCD. It contributes to dissociate ETnoD products that are already fragmented in the backbone of the protein but held together with non-covalent interactions after ETD activation³¹⁴. For HCD and CID, average sequence coverages between 22% and 38% are obtained, with higher average residue cleavages for the Lc subunit. UVPD fragmentation gave an average sequence coverage around 40% for all the subunits. This sequence coverage observed for UVPD is quite lower than what have been found in the literature (around 50%-55%)^{9, 10, 441}. It can be explained by the lower energy of our laser, which generates 1.2 $\mu\text{J}/\text{pulse}$, while lasers generally used in other studies provide higher energies^{308, 346}. Another explanation is that the parameters we selected for the deconvolution were more stringent (minimum fit factor of 70% and S/N ratio of 3) than those usually used in the literature (minimum fit factor of 44% and S/N ratio of 3). When using relaxed deconvolution parameters, the sequence coverage observed is around 45%-50%, depending on the subunits, closer to what has been reported in the literature. The influence of these deconvolution parameters is discussed in more detail in the next chapter (**Part III, Chapter 2**).

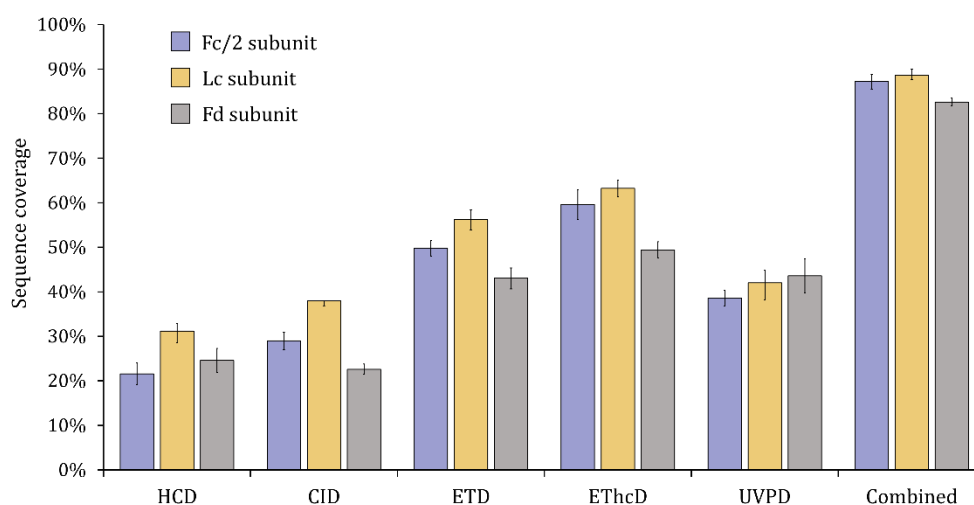


Figure 104: Trastuzumab subunits fragmentation results using HCD, CID, ETD, EThcD and UVPD and the combination of one run from each fragmentation technique (“combined”). Error bars are calculated using the standard deviation of the sequence coverage, with $n = 5$.

In addition, fragmentation maps displayed in **Figure 105** revealed that specific cleavage patterns occur for each technique. In particular, HCD and CID demonstrate preferential regions for residue cleavages. Hence, we combined the results obtained by individual fragmentation technique, rising the average sequence coverage to 83% for the Fd, 87% for the Fc/2 and 89% for the Lc subunits. Moreover, we compared the proportion of residue cleavages common to each approach. For instance, for the Fc/2 subunit, a high similarity is observed between fragments obtained from dissociation relying on the same principle (i.e., ETD and EThcD for electron-based or HCD and CID for collision-based). Of note, no specific cleavages from HCD bring additional sequence coverage. Additionally, EThcD and UVPD afford the highest complementary since both of them bring 17 unique cleavages, representing 8% of

combination of multiple replicates allow taking the advantage of the fragmentation variability of these techniques to increase the sequence coverage. Hence, the maximum backbone coverage obtained from EThcD combining the five replicates is 83% for the Fc/2, 82% for the Lc and 75% for the Fd, close to what was found by cumulating all the fragmentation techniques (87%, 89% and 83%, respectively). However, it is important to keep in mind that fragments identified in only one technical replicate over five are more likely to be false positives than ones identified in multiple replicates.

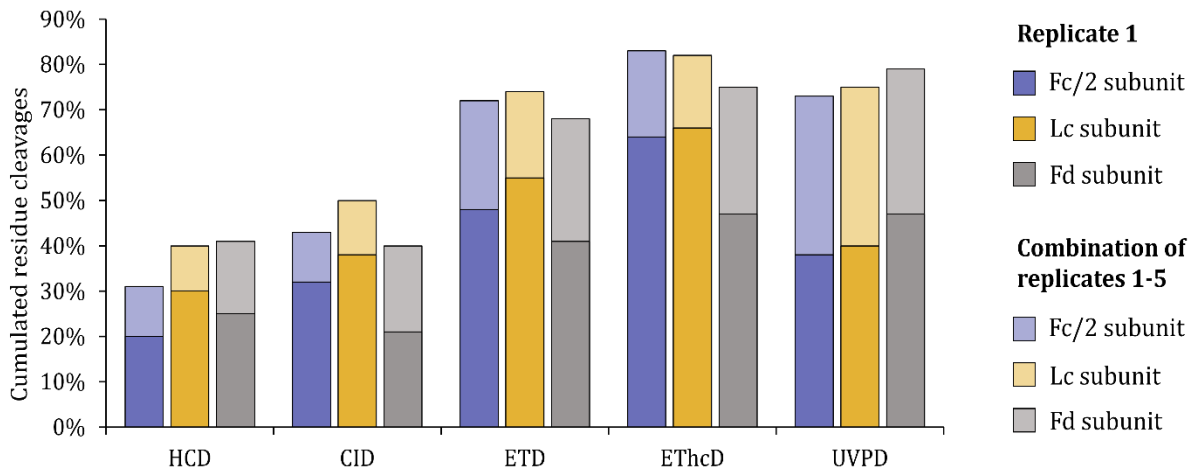


Figure 107: Summed residue cleavages over five replicates for each fragmentation technique of the trastuzumab Fc/2, Lc and Fd subunits.

Multiple parameters must be optimized when performing TD/MD-MS experiments. The choice of the precursor ion and the activation energy are two critical parameters, as they will directly affect the fragmentation efficiency and sequence coverage. They should be carefully optimized. Using multiple precursor ions could improve the sequence coverage regarding the dependence of the fragmentation channels on the charge state. However, its implementation on an LC time scale is intricate due to the need for multiple spectra accumulation. Combining multiple fragmentation techniques should always be considered to obtain extensive backbone fragmentation regarding the complementarity of the cleavage pathways of each technique. Finally, using technical replicates is another way to improve the sequence coverage due to the signal variability. After combination of different fragmentation techniques, up to 90% of residue cleavages can be observed after mAb subunit analysis in LC-MD-MS experiments, being closer to sequence coverages typically provided in peptide mapping workflows.

4. Implementation of proton transfer charge reduction (PTCR) for mAb analysis

Despite the high sequence coverage obtained, especially with electron-based fragmentation (i.e., ETD and EThcD), spectra generated by these techniques are generally overcrowded with the possibility of different overlapping isotopic profiles from different fragments. It can hinder the deconvolution process of the resulting ions and thus lose valuable information to provide a complete sequencing of the protein in MD-MS workflows. To circumvent this issue and identify the maximum number of fragment ions to increase sequence coverage, proton transfer charge reduction (PTCR) has been implemented for mAb subunit characterization. Here, we appraised two different PTCR methods to assign the fragment ions better: “wide window” (1,000 m/z isolation window) and “narrow window” (5 x 200 m/z isolation window) modes. Isolation windows were selected based on the fragment ions distribution over the m/z range after ETD and EThcD, as presented in **Figure 108**. Similarly, ion population in each isolation was roughly estimated to adapt the PTCR reaction time: highly populated windows were submitted to a more prolonged PTCR reaction (see **Experimental Section**).

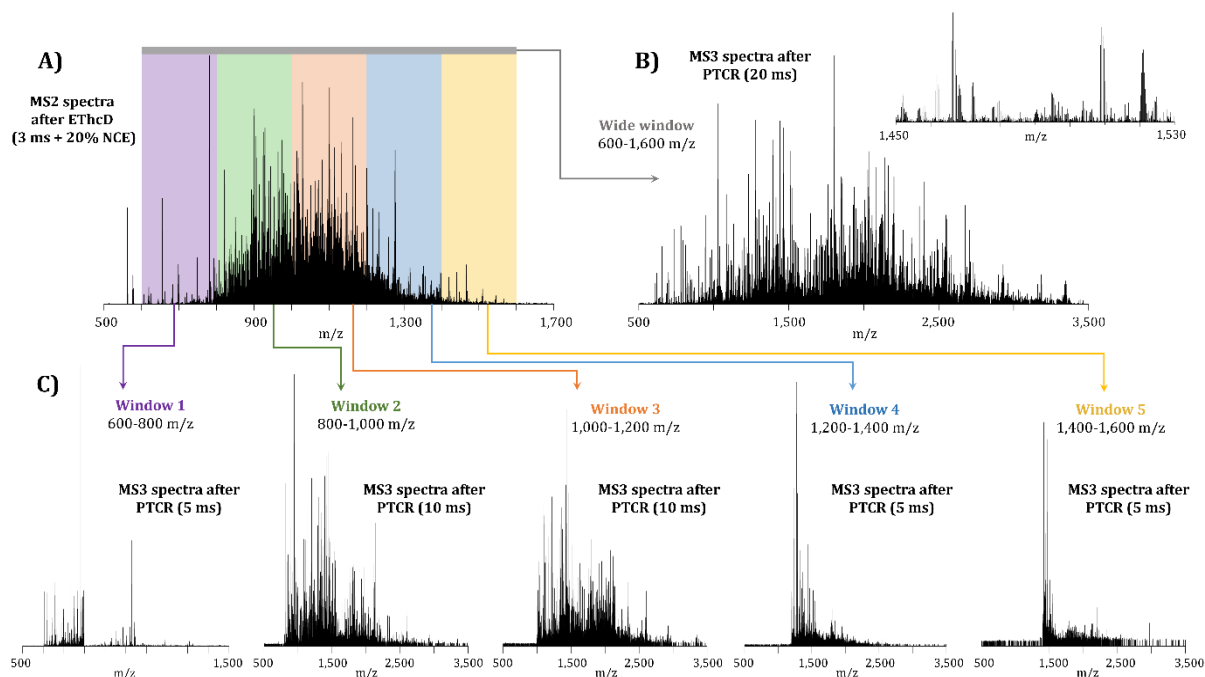


Figure 108: Schematic representation of applied data acquisition schemes. (A) Standard MS2 spectra generated from EThcD fragmentation. (B) Product ions isolation using a single, 1,000 m/z-wide isolation window and subjected to PTCR prior to their detection. (C) Product ion subpopulations isolation via narrower isolation windows (5 x 200 m/z) and subjected one by one to PTCR prior to their detection.

After ETD/EThcD fragmentation, a complex MS2 spectrum is generated, with many fragments distributed only over a relatively small m/z range (600 – 1,600 m/z, **Figure 109**). For instance, using PTCR for 20 ms on this wide window drastically reduced the spectral congestion of this zone of the MS2 spectra obtained after ETD 3 ms, spreading the signal over a broader m/z range, up to 3,500 m/z, and reducing overlapping signals and interferences.

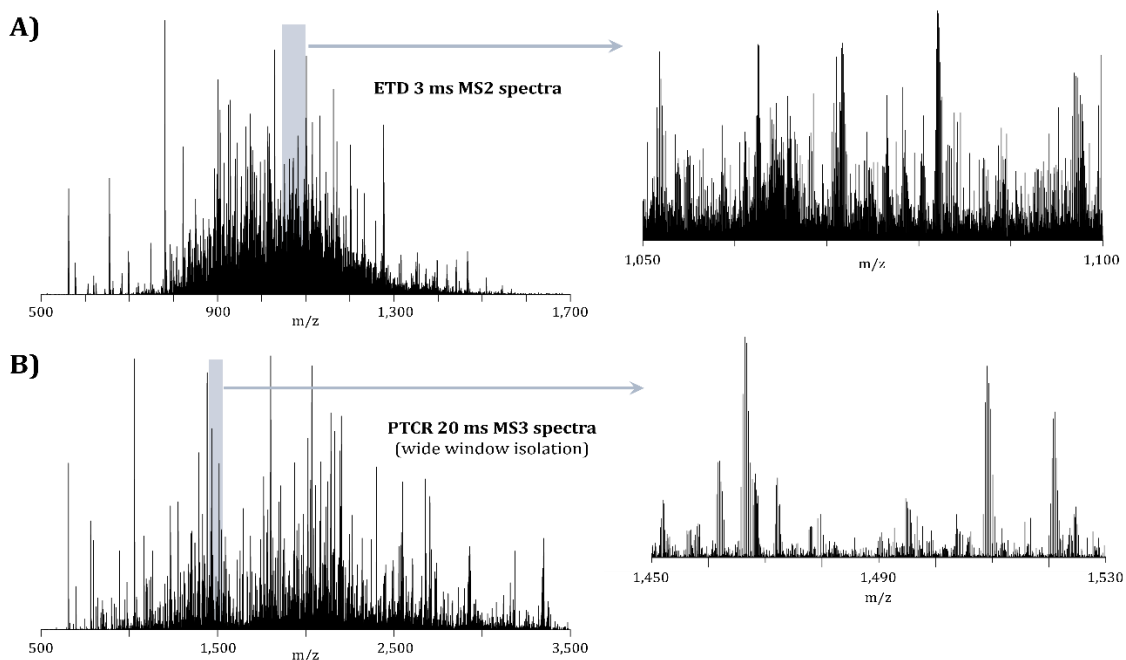


Figure 109: Fragmentation spectra of trastuzumab Lc subunit. (A) MS2 spectra obtained after ETD 3 ms fragmentation and zoom on the 1,050-1,100 m/z region (inset). (B) MS3 spectra obtained after PTCR 20 ms performed on the 600-1,600 m/z range (wide isolation window) after ETD 3 ms fragmentation, and zoom on the 1,450-1,530 m/z region (inset).

Results from PTCR implementation after ETD/ETHcD are displayed in **Figure 110**. After ETD, wide isolation windows improve the sequence coverage by 5%-8% for ETD, and the combination of narrow isolation windows by 12%-14%. In contrast, PTCR using a single 1,000 m/z isolation window leads to the highest sequence coverage after ETHcD compared to narrow ones, with an increase of 10%-16% against 5%-12%, respectively. One hypothesis of the counter-intuitive results obtained from PTCR after ETHcD for narrow windows is that lower charge states (due to additional HCD fragmentation) should require lower PTCR reaction time for each window. Indeed, a more extended reaction time would result in the neutralization of the fragment ions with low charge states. This is particularly the case for windows containing only a few fragments. Moreover, this behavior has already been observed in the literature³¹⁸, and the authors found that results from PTCR on small windows lead to lower mass fragments compared to wide windows. An in-depth investigation of our data reveals that a comparable trend is observed. Hence, small windows will only improve the sequence coverage on the terminal part of the sequence, while using a wide window provides additional fragments all over the protein backbone, leading to a higher increase in the sequence coverage.

Globally, implementation of PTCR after ETD fragmentation leads to an average rise of the sequence coverage from 50% to 64%, 56% to 70%, and 43% to 55% for the Fc/2, Lc, and Fd subunits, respectively. Similarly, the average percentage of residue cleavages rise from 60% to 73%, 63% to 76%, and 49% to 65% for the Fc/2, Lc, and Fd subunits using PTCR after ETHcD (**Figure 110**). In addition to an increased sequence coverage, implementation of PTCR allows the identification of new fragment ions in specific regions of the sequence. For instance, 9 additional cleavages were detected between amino acids 176 and 200 of the Fd subunit using PTCR after ETHcD (i.e., 10 without PTCR and 19 after PTCR implementation).

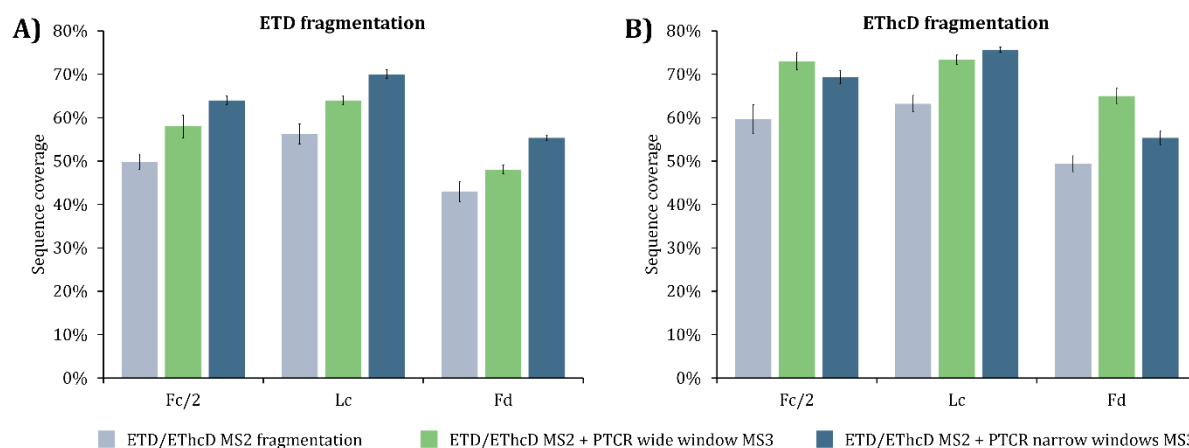


Figure 110: Average sequence coverage (n = 3) obtained after (A) ETD and (B) ETHcD and the addition of PTCR with wide and narrow isolation windows for trastuzumab subunits.

In summary, PTCR implementation after fragmentation inducing congested MS2 spectra, i.e., ETD and ETHcD, improves the average percentage of residue cleavages between 5% and 15%, depending on the subunit. Notably, it can enhance the sequence coverage on poorly characterized regions. It can be of great interest when dealing with mAbs (or other proteins) bearing modifications (e.g., ADC) localized in regions of the sequence that are not extensively characterized because of fragment ion overlapping. However, PTCR will not help for regions that are not fragmented, as it will not induce additional fragmentation. Nevertheless, PTCR induces a lower-intensity signal due to the repeated fragmentation and analysis of the same precursor ion. Similarly, reaction times for PTCR are between 5 ms to 30 ms, increasing the cycle time and consequently lowering the number of spectra acquired. Hence, it is critical to keep in mind to average a maximum number of spectra to have a high enough S/N ratio for fragments deconvolution.

5. Conclusion and perspectives

This chapter assessed various acquisition strategies for characterizing proteins and biomolecules using TD/MD-MS. The first part underlined the influence of the precursor charge state in the fragmentation process to obtain the best sequence coverage, as well as the critical need for activation time/energy fine-tuning. It also illustrated the ability to perform MD-MS after chromatographic separation and highlighted the challenge of working on a chromatographic time scale. In addition, it was exposed that combining fragmentation techniques and injection replicates is beneficial to take advantage of their complementarity for sequence coverage improvement. The second part demonstrated that additional PTCR ion-ion reaction after ETD/ETHcD is of remarkable value in improving the sequence coverage and identifying additional residue cleavages. In the last part, supplementary fragmentation of specific fragment ions was identified as a promising approach to obtain information on the region of interest of proteins.

In top-down and middle-down approaches, fragmentation is crucial to obtain the maximum sequence coverage and information from the spectrum. Even if some global trends guiding the choice of the experimental parameters can be found in the literature, the high dependency on the fragmentation results in the protein of interest (size, amino acid composition, etc.) highlights the need for careful optimization of the fragmentation to ensure a maximum sequence coverage.

As we saw in this chapter, multiple fragmentation techniques are available and implemented in state-of-the-art MS instruments. Nevertheless, additional fragmentation strategies can be used to perform TD/MD-MS experiments. Electron capture dissociation (ECD)^{271, 277, 602}, mainly implemented on FT-ICR instruments, is based on the reaction between free electrons and multiply protonated molecules to obtain fragment ions. It is commonly used in TD/MD-MS to characterize intact proteins and identify their proteoforms^{357, 603-605}. Supplemental activation of proteins using infrared irradiation was also combined with electron-based dissociation approaches (i.e., ETD and ECD) to enhance protein fragmentation (AI-ETD and AI-ECD)^{292, 294, 428}. In addition, hybrid activation techniques (including ETuvpD, ETHcD, and ETciD) have been developed as they offer a more balanced distribution of ion types and decrease the spectral congestion via fragmentation of multiple charge-reduced precursors and charge reduction of fragment ions^{312, 313}. Nevertheless, they do not simplify the spectra nor improve the sensitivity. Recently, a new platform named Omnitrap (Fasmatech) was released, incorporating an extended range of various activation techniques, including external injection of reagent ions, radical neutral species, photons, electrons, and collisions with neutrals²⁹⁶. It allows multiple-stage tandem processing of ions for in-depth top-down characterization of proteins.

Information provided by TD-MS experiments is generally very complex, with many fragments observed and often overlapping. Consequently, several approaches have been used to “simplify” information obtained after TD-MS. We saw previously in this chapter the benefits of PTCR to increase sequence coverage by spreading the signal over a broad m/z range. It has been used by multiple groups to improve the sequencing of large proteins^{316, 318} or for more confident characterization of histone proteoforms³³⁷. Other strategies take advantage of cyclic ion mobility spectrometry (cIMS) to simplify the product ions population arriving in the mass analyzer^{603, 605} or FAIMS to separate proteoforms before fragmentation to get rid of signal from multiple proteoforms on the same spectra⁶⁰⁶.

All these technological advances substantiate the interest of TD/MD-MS for the characterization of intact proteins and their proteoforms. Nonetheless, processing the vast amount of data obtained by these experiments is currently a critical point and a challenge that needs to be addressed to ensure the identification of the maximum of fragments while limiting the number of false positives. This part will be developed in **Part III, Chapter 2**.

Chapter 2:

Impact of deconvolution parameters and fragment ion matching in middle-down MS workflows

1. Context of the project and analytical strategy

In top-down and middle-down experiments, the vast amount of information generated after fragmentation needs to be processed to obtain information about the primary structure of the proteins. Data processing follows two main steps: the deconvolution of the MS/MS data to obtain the zero-charge mass of the fragment ions, and the assignment of those masses to the specific ions of a candidate sequence. Even though some specific TD/MD-MS software can perform fragment ion matching from the raw spectrum (without deconvolution) such as TDValidator, the vast majority of this informatics tools require a zero-charge fragment ion mass list to identify the fragment ions. This step is the main bottleneck for MS/MS data treatment since it allows obtaining the precise monoisotopic mass of the ions from complex and overcrowded MS/MS spectra, and thus, it will directly affect the number of fragments identified and the overall sequence coverage. For all these reasons, TD/MD-MS experiments require high-resolution data (i.e., at least 120,000 resolution at 200 m/z) to differentiate neighboring fragment ions, and obtain a detailed isotopic profile along with a precise experimental mass to facilitate the correct identification. Nevertheless, accurate extraction of monoisotopic peaks by the algorithms is challenging for many reasons: background noise is generally hampering the overall MS/MS quality, fragments of interest with different charge states can potentially overlap, and a vast number of fragment ions (depending on the activation technique) are present in low abundance. Various bioinformatics tools exist to deconvolute MS/MS spectra, employing different deconvolution methods, all having pros and cons^{319, 325}. Then, the list of deconvoluted spectra is matched to the theoretical sequence to identify residue cleavages. Many bioinformatics tools have been developed for purified proteins or high-throughput TD proteomics^{325-327, 329}. Depending on the TD experiment, the level of characterization of the species varies. On one hand, high-throughput TD proteomics aims at identifying the maximum number of proteoforms in a complex sample along a chromatographic timescale. Hence, the presence of several fragments specific to the proteoforms is considered enough to validate the presence of proteoforms, and discard any spurious identifications³³⁰. To this end, deconvoluted masses are searched against a database by setting potential PTMs as fixed or variable modifications, like in bottom-up proteomics. Hence, all proteoforms of the proteins present in the sample can potentially be identified. Currently, these approaches still provide a lower number of identified proteins; however, compared to bottom-up approaches, the number of identified proteoforms is clearly higher, showing the synergy between both methods. On the other hand, TD experiments on purified proteins seeks to characterize comprehensively the protein, i.e., maximizing the sequence coverage by combining different fragmentation techniques and conditions, and obtaining information on the precise location of PTM. Thus, fragment assignment and signal quality are generally checked manually to ensure a correct identification.

While the number of tools is growing fast, up to today, no clear guidelines have been established neither for parameters affecting the deconvolution of fragment ions, nor for the type of fragment ions to consider with each individual activation technique. Indeed, some parameters are experiment dependent and must be adapted accordingly, such as ion types^{297, 439, 601} (e.g., *b* and *y* ion types for CID and HCD, *c* and *z* ion types for ETD, or *a*, *b*, *c*, *x*, *y* and *z* for UVPD) or deconvolution parameters

depending on the fragmentation^{10, 350, 439}. In addition, regardless of the fragmentation used, the potential for fragmenting the same protein multiple times must be considered, as sufficiently collisional energy can result in secondary fragmentation of product ions or as multiple electrons or photons can initiate independent fragmentation events^{333, 337}. Most bioinformatics tools ignore this information in the MS/MS spectra during the analysis of fragment ions, as they only consider N- and C-terminal ions. Nevertheless, previous studies show their potential interest not only to increase the sequence coverage^{332, 334, 607, 608} but also to improve the confidence on PTMs localisation⁴⁴². To account for these internal fragments, which are growing in number with the size of the protein³³¹, Loo and coworkers developed a Python algorithm named ClipsMS³³⁵, allowing the assignment of both terminal and internal fragments.

Considering the lack of guidelines regarding the deconvolution of MS/MS spectra, we decided to assess the influence of some deconvolution and fragment ion identification parameters on the sequence coverage of proteins, and mAbs. For this evaluation, we used the Xtract³²⁴ (Thermo Fisher Scientific) deconvolution software based on the TRASH algorithm³²³ and the ProSight Lite³²⁶ software for fragment identification (**Figure 111**). We focused on critical parameters, namely signal-to-noise ratio (S/N), fit factor threshold for deconvolution, and fragment tolerance for identification. In the second part, we used ClipsMS to appraise the potential of using internal fragments to increase the number of residue cleavages and localize PTMs (**Figure 111**).

Additional details on sample preparation, LC-MS/MS methods and data processing are listed in the **Experimental Section**.

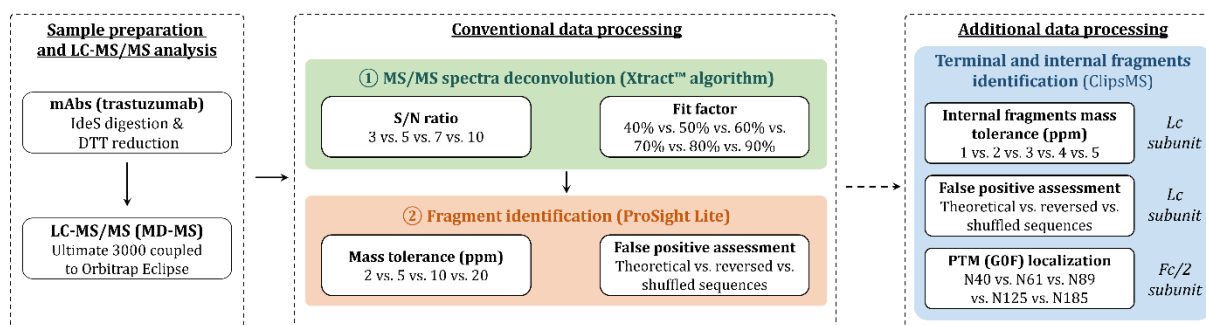


Figure 111: Analytical strategy used for the evaluation of top-down and middle-down data treatment strategies. Evaluation of the influence of S/N ratio and fit factor value on MS/MS spectra deconvolution of trastuzumab Lc subunit using the Xtract™ algorithm, followed by the assessment of the influence of the mass tolerance used for fragments matching in ProSight Lite³²⁶, as well as the risk of false positives, using reversed and shuffled sequences. Then, additional data processing considering internal fragments for sequence coverage improvements using the ClipsMS³³⁵ software on the trastuzumab Lc subunit, as well as assessment of the risk of false positives. The potential of internal fragments for PTM localization was performed using the GOF species of the trastuzumab Fc/2 subunit.

2. Evaluation of the risk of false positives

Before the evaluation of the deconvolution and matching parameters, we assessed the risk of false positive when performing fragment ion matching upon deconvolution, using the reverse sequence and five shuffled sequences generated from the original one. We used different metrics to interpret the data, such as the number of identified fragments, the sequence coverage and the P-Score. This score represent the probability that a random sequence could account for the matching ions³³⁰, thus a lower P-Score will correspond to a lower probability of false matching. Regarding the P-Score, the $-\log(\text{P-Score})$ was used for representation convenience, meaning that the lower the P-Score, the higher will be $-\log(\text{P-Score})$ value.

To evaluate the risk of false positives, the fragmentation of the Lc subunit upon fragmentation with HCD, ETD, EThcD and UVPD were used. In the case of the data associated to the shuffled sequences, the average of the five shuffled sequences is displayed with the corresponding standard deviation of the five sequences.

Results of this evaluation are presented in **Figure 112**. As expected, the theoretical sequence logically leads to the highest sequence coverage and number of fragments identified. In comparison, sequence coverage is 2 to 6 times and 3 to 18 times lower using the reversed and shuffled sequences, respectively. In particular, EThcD and UVPD show the highest number of identified fragments using false sequences. Indeed, with these fragmentations techniques, we are looking at 4 (*b*, *c*, *y*, and *z* ions) and 6 (*a*, *b*, *c*, *x*, *y*, and *z* ions) types of ions, respectively, increasing the probability of matching a fragment by chance. A significant part of the fragments assigned to reverse/shuffled could account for false positive assignment. Nevertheless, some of these ions calculated from shuffled or reversed sequences can be also isobaric or near isobaric fragments with some fragments of the original sequence. In this case, these ions cannot be discarded based on the mass tolerance error used to match the fragments.

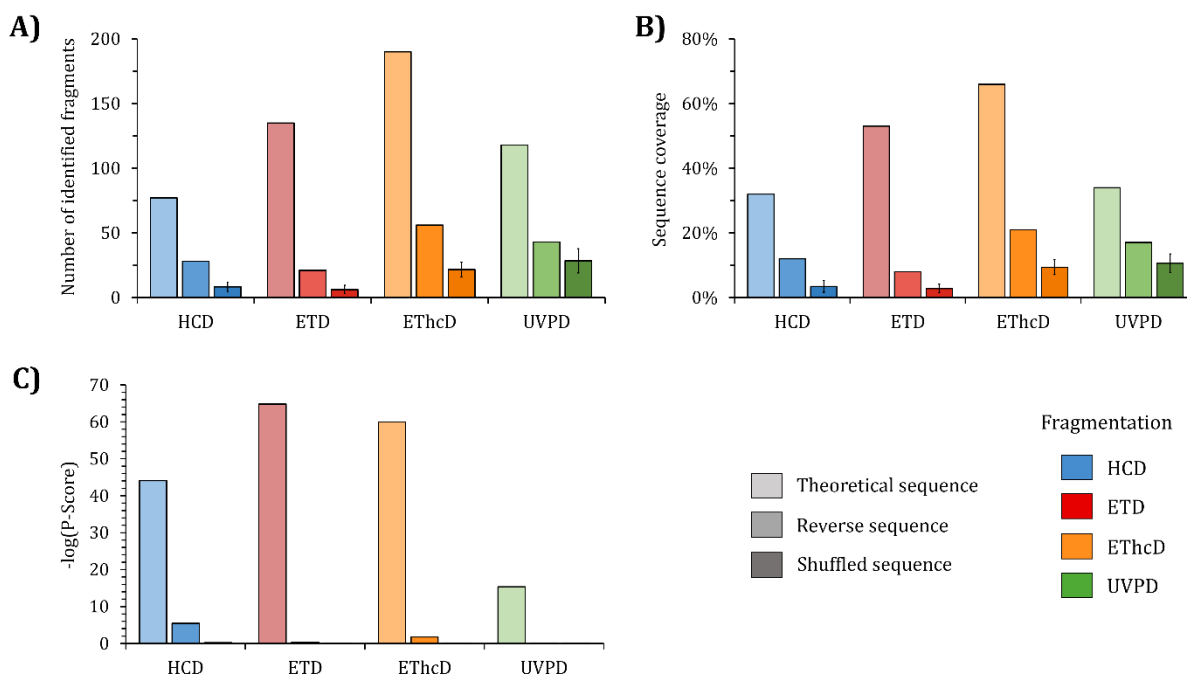


Figure 112: Identification results of the trastuzumab Lc subunit using the theoretical sequence, the reverse sequence or randomly shuffled sequences. (A) Number of identified fragments, (B) sequence coverage (%), and (C) $-\log(\text{P-Score})$ values. Results from HCD are displayed in blue shades, red for ETD, orange for EThcD and green for UVPD.

The analysis of false positives matching is relevant upon comparisons of the data regarding the sequence coverage and the number of identified fragments from the original sequence and the reversed/shuffled ones (**Figure 112.B**). For instance, 3% and 12% sequence coverage are obtained with HCD fragments searched against the scrambled and reversed sequences, respectively, compared to 32% for the theoretical ones. However, studying the P-Score values for these “false” sequences undoubtedly allows us to discriminate between original and reversed/shuffled sequences. Fragment matching against theoretical sequences results in $-\log(\text{P-Score})$ values 8 to 200 times higher than reversed sequence and 145 to 25,000 times compared to shuffled sequences, depending on the fragmentation technique (**Figure 112.C**).

To conclude, shuffling or reversing the theoretical sequence shows that it is still possible to match some fragments underlying the presence of false positives when analyzing TD/MD-MS data. Nevertheless, the $-\log(\text{P-Score})$ is an appropriate and suited indicator for false positive evaluation, as the probability that the sequence matching is wrong was always close to 1 when dealing with reversed and shuffled sequences. Consequently, this metric was used along with the sequence coverage and the number of identified fragments to evaluate the relevance of the deconvolution and ion matching assignment values used in the different sections.

3. Evaluation of the deconvolution parameters

Plenty of parameters are available when performing mass deconvolution. In the FreeStyle™ (Thermo Fisher Scientific) interface, hosting the Xtract algorithm, all parameters can be defined (an exhaustive list is provided in **Table 10**). The definition of these parameters depends on multiple factors, including the analyzer, the fragmentation method, the spectral quality, and the intrinsic characteristic of the precursor ion.

Table 10: Deconvolution parameters available using the Xtract™ algorithm and values applied for the deconvolution of the trastuzumab Lc subunit.

Parameter	Default value	Description
<i>Global parameters</i>		
Fit factor (%)	40%-90%	Quality of the match between a measured isotope pattern and an averaging distribution of the same mass
Remainder threshold (%)	25%	Height of the smaller overlapping isotopic cluster with respect to the height of the most abundant isotopic cluster
Consider overlaps	Yes	If selected, the algorithm is more tolerant of errors when the spectrum intensity is significantly higher than expected for the theoretical isotopic cluster.
Minimum intensity	1	Minimum intensity threshold to filter out possible background noise
Expected intensity error	3	Permissible percentage of error allowed in calculating ratio of the most abundant isotope to the next isotope that is higher in mass in the isotope series
Resolution	120,000	Resolution of the source spectrum
S/N threshold	3-10	S/N threshold above which the Xtract algorithm considers a measured peak to be a real accepted peak
<i>Spectrum specific parameters</i>		
m/z range	Spectra m/z range	Portion of the input spectrum to process (automatically defined by default)
Output mass	MH+	Whether the Xtract algorithm returns a single peak at either the monoisotopic mass (M) or the monoisotopic with adduct mass (MH+) for each detected component
Adduct element	H+ (1.00728)	Adduct ions used during ESI processing
Charge range	1-24 or 1-27	Lowest and highest charge state to be deconvoluted
Minimum number of detected charge	1	Minimum number of charge states required to produce a component
Analyzer type	Orbitrap	Type of mass analyzer that was used to obtain the mass spectral data
Relative abundance threshold	0%	Threshold below which the Xtract algorithm filters out data for data reporting. This threshold is applied on the deconvoluted spectrum.

The influence of the deconvolution parameters was assessed with the fragmentation of the Lc subunit of trastuzumab. In this case, different precursor charge states were chosen depending on the activation technique, i.e. the 24+ charge state was selected for HCD and UVPD fragmentations, while the 27+ charge state was isolated to perform ETD and EThcD. MS/MS spectra were recorded at a resolution of 120,000. Some parameters such as the minimum intensity, the inclusion of isotope ion overlap, the

intensity error, and the relative abundance threshold were fixed according to the manufacturer default values since no significant differences were observed when they were modified within a rational range of values. Then, charge range, analyzer type and resolution have been set according to the experimental conditions and were maintained through all the data treatment. Similarly, the minimum number of detected charge has been set to 1, as most of the fragment ions exhibit a single charge state population. Alternatively, parameters dependent of the precursor ion or the fragmentation method, including the maximum charge state and the m/z range were defined accordingly for each spectrum. Values used for each parameters are detailed in **Table 10**. These parameters are usually dependent of the user perception and habits and are usually not changed across the experiments. Finally, two key parameters were modified along this evaluation: the S/N ratio and the fit factor threshold. Their influence on the number of deconvoluted masses and identified fragments, the sequence coverage, and the $-\log(\text{P-Score})$ was assessed (see **Experimental Section**).

A. Influence of the S/N ratio

In TD/MD-MS experiments, the signal-to-noise ratio is directly linked to the spectrum quality. The presence of high background noise can prevent the correct detection, and hence deconvolution, of the isotopic profile of fragment ions, particularly for those having a low abundance signal. To improve S/N factor of the MS/MS spectra, fragmentation of the Lc subunit was averaged during the elution time of the subunit. Even though the number of transient that can be accumulated/averaged in an liquid-chromatography time scale is limited to the width of the chromatographic peaks, the resulting fragmentation spectra showed well-shaped isotopic profiles with S/N values up to 5,000, which led to the detection of several hundreds of ions under each fragmentation conditions.

First, the number of fragments obtained after deconvolution is displayed in **Figure 113.A**. As expected, it can be seen that this number decreases when the S/N value is raised. It is particularly marked for UVPD, where the number of fragments is divided by two between 3 S/N and 10 S/N, while this number drops by 30% for ETD. For methods including collision (i.e., HCD and EThcD), around 20% fewer fragments are observed between the lowest and the highest S/N values. This trend might be explained by the lower intensity of the fragments generated upon UVPD fragmentation. Overall, the photo-fragmentation exhibits the lowest fragmentation efficiency compared to radical- and collision-driven dissociation, along with the highest number of fragmentation channels, which lead to the production of an array of low-intense fragments. Despite the drop in the number of deconvoluted masses, the number of identified fragments and the sequence coverage are less sensitive to the S/N threshold compared to the number of deconvoluted fragments (**Figure 113.B and C**). Thus only limited sequence coverage variations (lower than 10%) are observed between 3 S/N and 10 S/N. One probable explanation to this is that the S/N ratio of the fragment ions of interest are enough intense to be not significantly impacted by the increase of the threshold. For instance, HCD fragments that are known to usually be abundant, are almost not affected by a stringent threshold, while this is the case for UVPD, which tends to give rise to low-intensity fragments.

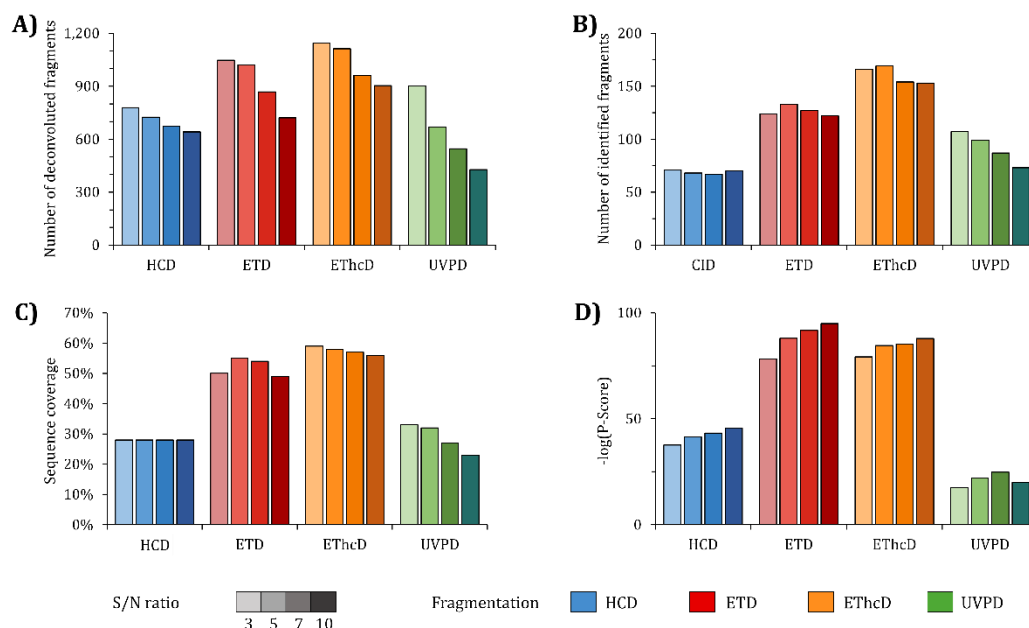


Figure 113: Identification results of the trastuzumab Lc subunit depending on the S/N ratio used. (A) Number of deconvoluted fragments, (B) number of identified fragments, (C) sequence coverage (%) and (D) $-\log(\text{P-Score})$ values. Results from HCD are displayed in blue shades, red for ETD, orange for EThcD and green for UVPD. A darker shade means a higher S/N ratio.

For ETD, we observe a lower sequence coverage and number of fragments identified with 3 S/N compared to 5 S/N and 7 S/N. No clear evidence explaining this behavior was found. Nevertheless, upon a more detailed analysis of the data, it was found that some ions were wrongly deconvoluted when using low values of S/N threshold since low abundant fragment ions hinder the deconvolution process leading to wrong fragment ion masses. Therefore, isotopomer envelopes will then directly influence the number of ions identified, and hence the overall sequence coverage. Similar observation have already been mentioned in the literature^{10, 307}. Above that, considering the $-\log(\text{P-Score})$ also corroborates the excellent quality of the results, independently of the S/N ratio value used. We observed $-\log(\text{P-Score})$ values of 18-25 for UVPD, 38-46 for HCD, 78-95 for ETD, and 79-88 for EThcD (**Figure 113.D**). The lower $-\log(\text{P-Scores})$ obtained for UVPD are due to considering multiple ion types (i.e., a/x , b/y , and c/z), increasing the probability of mismatching and consequently decreasing the $-\log(\text{P-Score})$.

These results illustrate that the S/N ratio parameter only slightly influences the sequence coverage, the number of ions identified, and the probability of mismatching. However, by setting a reasonable value, i.e., higher than 5 S/N, the confidence of fragment ion matching is enhanced, allowing to discard lousy-quality ions that could impair the accuracy of the deconvolution process of fragment ions or could have been identified as false positive matching if they were included. In contrast, a too-high value (i.e., 10 S/N) will eliminate a significant number of well-identified fragment ions that can improve the overall sequence coverage of the proteins, or even provide valuable information about the precise position of important PTMs. Hence, S/N value of 5 to 7 is recommended for collisional and electron-driven fragmentations while S/N threshold of 5 is more suited for UVPD fragmentation since the sequence coverage substantially decreases with values above 5.

B. Influence of the fit factor threshold

After considering the S/N value as a parameter that can influence deconvolution, we focused on the fit factor, representing the match quality between a measured isotope pattern and an averaging distribution of the same mass. As said above, we selected a S/N value of 7 to test the influence of the

fit factor, and we evaluated values from 40% to 90% to assess its influence on the final outcome in a wide range of values.

Results are presented in **Figure 114**. Globally, every item studied (i.e., number of deconvoluted masses and identified fragments, sequence coverage, and P-Score) demonstrate considerable changes with the variation of the fit factor value. In detail, the number of deconvoluted masses drops with the increase of the fit factor, especially for values higher than 60% (**Figure 114.A**). For instance, while more than 3,000 masses were obtained after deconvolution of EThcD data using a fit factor of 40% or 50%, this number dropped lower than 2,000 for 70% and less than 400 fragments for 90%. This point indicates the strong effect of the fit factor on the deconvolution process. Indeed, low abundant ions may exhibit a wrong isotopic pattern and consequently not be selected when increasing the fit factor cutoff.

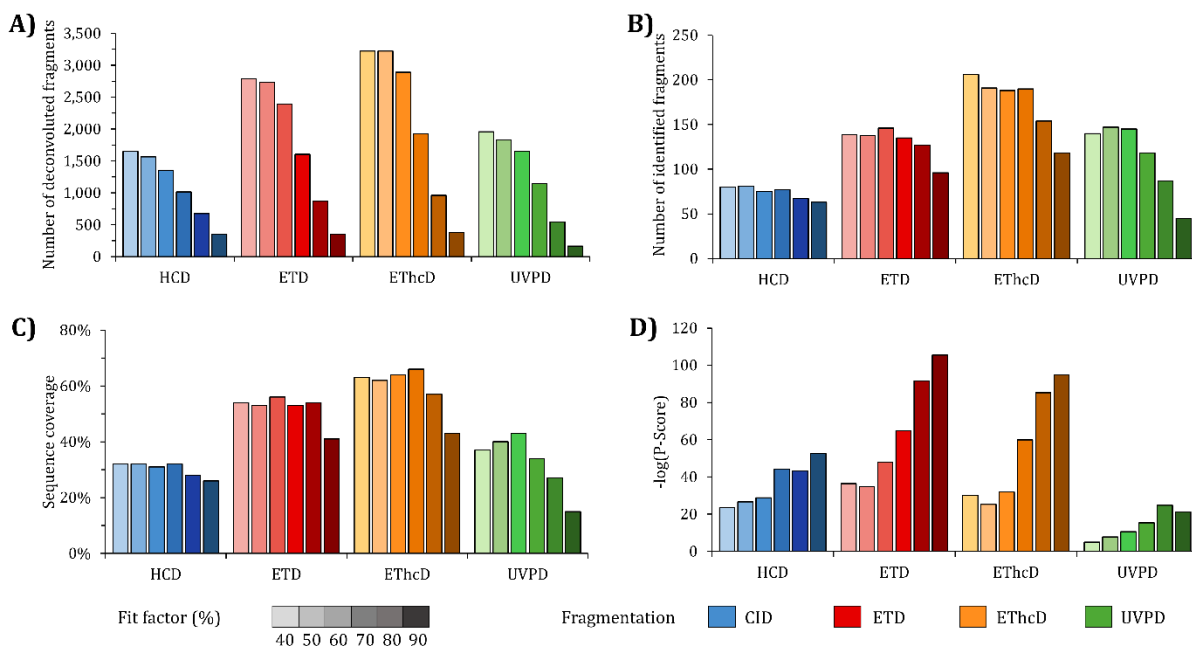


Figure 114: Identification results of the trastuzumab Lc subunit depending on the fit factor used. (A) Number of deconvoluted fragments, (B) number of identified fragments, (C) sequence coverage (%), and (D) $-\log(\text{P-Score})$ values. Results from HCD are displayed in blue shades, red for ETD, orange for EThcD and green for UVPD. A darker shade means a higher fit factor.

Despite this decrease in the number of masses available for identification, the sequence coverage and the number of identified fragments remain stable for fit factors up to 70%-80% (except for UVPD, 60%) while dropping when a 90% threshold is used (**Figure 114.B and C**). In addition, low fit factor threshold (i.e., 40% and 50%) result in higher P-Score (i.e., inferior $-\log(\text{P-Score})$ values), highlighting that these fit factor values are not convenient to limit the risk of false positive matching (**Figure 114.D**). It is particularly important for UVPD fragmentation, as the $-\log(\text{P-Score})$ is already low due to the multiple ion types that can be matched. Nevertheless, P-Scores obtained with a fit factor threshold of 70% or higher are already low enough to ensure that most matched fragments are true positives.

To summarize, using low fit factor values (i.e., 60% or lower) for deconvolution boosts the number of deconvoluted fragments, however, the additional fragments do not have a straight impact on the sequence coverage of the protein. Consequently, the balance between benefit and potential risk of false positive identification is low for low fit factor values as attested by the variation of the $-\log(\text{P-Score})$. In contrast, a too stringent fit factor (i.e., 90%) decreases the number of deconvoluted fragment ions between four and ten-fold factor, reducing drastically the overall sequence coverage (10 to 20% loss). Thus, we recommend using fit factor values between 60% and 80% to ensure reliable deconvolution, and characterization of the sequence of proteins in TD/MD-MS workflows, even if slightly lower values

can be used for UVPD fragmentation, as a more significant impact of the fit factor is observed for this fragmentation method.

In conclusion, the S/N ratio and fit factor are essential parameters to optimize for deconvolution as they directly influence the number of deconvoluted fragments and, consequently, the sequence coverage. Stringent parameters will increase the confidence in the fragments identified but concomitantly will lead to the loss of valuable information contained within the MS/MS spectra. Conversely, loose conditions of these parameters will increase the risk of false positive matching. However, the additional increase in sequence coverage afforded with low deconvolution thresholds can be particularly appealing for TD-MD/MS practitioners. For example, deconvolution of EThcD spectra with low deconvolution thresholds (i.e., S/N 3 and fit factor 40%) gives a sequence coverage of 69% while using stringent thresholds (i.e., S/N 7 and fit factor 80%) leads to a drop of the sequence coverage to 57%. Therefore, this sequence coverage decrease can be a compelling reason to favor the choice of loose ion fragment deconvolution parameters when the relevance and suitability of data treatment process is relying solely on the final sequence.

4. Evaluation of the mass tolerance threshold for fragments assignment

The fragments generated from deconvolution are searched against the theoretical sequence using ProSight Lite. In this software solution, only a few parameters can be set, including the type of input masses (either neutral (M) or protonated (MH⁺), depending on how ions were deconvoluted, the type of fragmentation, and the mass tolerance. Here, we used MH⁺ mass as input, and the ion types for matching were selected depending on the fragmentation methods (see **Experimental section**). We focused on the fragment tolerance for matching fragments and set tolerances from 2 ppm to 20 ppm.

As before, we used the number of identified fragments, the sequence coverage, and the P-Score as metrics to evaluate the influence of mass tolerance (**Figure 115**). For all fragmentation techniques, we observed a rise in the number of identified fragments and, consequently, in the sequence coverage when the mass tolerance is set to higher values. In detail, a stringent mass tolerance of 2 ppm significantly affects the sequence coverage, with residue cleavages of 7% for UVPD, 9% for HCD and ETD, and 15% for EThcD (**Figure 115.A**). Extending the fragment tolerance to 5 ppm increases the sequence coverage by 4-5.5 times compared to 2 ppm. However, no big differences are observed in terms of number of identified fragments and overall sequence coverage between 10, and 20 ppm of mass tolerance. The highest tolerance gave rise to an increase of sequence coverage between 4% and 15% compared to 5 ppm, depending on the activation technique. Conversely to the sequence coverage and the number of identified fragments, the $-\log(\text{P-Score})$ exhibited different trends. While $-\log(\text{P-Score})$ decreases (except for ETD) when the fragment tolerance increase between 5 ppm and 20 ppm, as expected, the lowest score, by far, is obtained for a mass tolerance of 2 ppm with values two- to five-fold lower than the one observed for 20 ppm for example (**Figure 115.C**). This result is a direct consequence of the low proportion of identified fragments.

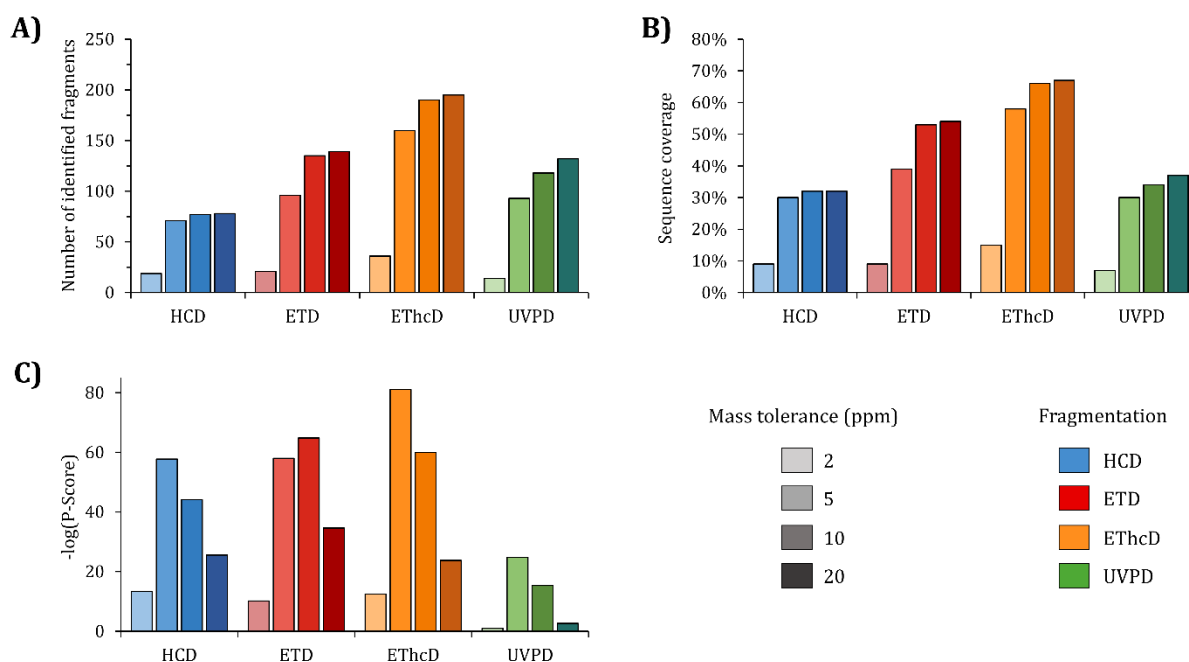


Figure 115: Identification results of the trastuzumab Lc subunit depending on the mass tolerance used for identification in ProSight Lite. (A) Number of identified fragments, (B) sequence coverage (%), and (C) $-\log(\text{P-Score})$ values. Results from HCD are displayed in blue shades, red for ETD, orange for EThcD and green for UVPD. A darker shade means a higher mass tolerance.

In summary, a too-stringent fragment tolerance (i.e., 2 ppm) drastically reduced the sequence coverage and the number of identified fragments, negatively affecting the P-Score. Conversely, a 20 ppm threshold for identification did not improve the percentage of residue cleavages and increased the risk of false positives. Hence, 5 and 10 ppm mass tolerance provide the best trade-off between maximizing sequence coverage while limiting the number or at least the risk of false positives.

5. Consideration of internal fragments to increase sequence coverage

One of the main concerns of TD/MD-MS is the extensive exploitation of the vast amount of information generated during the fragmentation of intact proteins or their subunits. Classical algorithms match the deconvoluted masses to theoretical sequences by only considering fragments that contain either the N- or the C-terminal sites, called terminal fragments. Nevertheless, there is still a significant amount of fragments that remains unmatched upon deconvolution using the terminal fragments. It has been found that a substantial portion of these unidentified masses can be attributed to internal fragments^{334, 336, 337, 607}. Here, we evaluated the benefits of using internal fragments through the ClipsMS software to improve the sequence coverage of trastuzumab subunits, along with the particular considerations that have to be taken when including these ion fragments.

A. Evaluation of the mass tolerance threshold for internal fragment matching

Using ClipsMS, various parameters can be adapted, including the mass tolerance for both terminal and internal fragments, the minimum length of the internal fragments, and modifications and ion types to consider. The terminal fragments were retrieved from the internal fragment searching to reduce the possibility of false positives.

Searching for internal fragments increases the search space, from 426 possible terminal fragments to 22,162 terminal and internal fragments for HCD on the Lc subunit for example. Hence, the risk of false positive attributions. Hence, a stringent mass tolerance is generally used (i.e., 1 or 2 ppm)^{333, 335, 337, 608}. Here, we assessed the variation of the sequence coverage by varying the mass tolerance from 1 to 5

ppm, as illustrated in **Figure 116**. The consideration of internal fragments boosted the percentage of inter-residue cleavages from 32% to 48%, 53% to 64%, 66% to 87% and 34% to 78% for HCD, ETD, EThcD and UVPD, respectively, when comparing results with terminal fragments only and the addition of internal fragments at 1 ppm, in line with a previous study from Loo and coworkers³³². Setting the mass tolerance to 5 ppm results in a sequence coverage of around 85% for HCD and ETD and higher than 95% for EThcD and UVPD. The additional gain in sequence coverage is more limited when the mass tolerance is modified from 3 to 5 ppm (particularly for EThcD and UVPD).

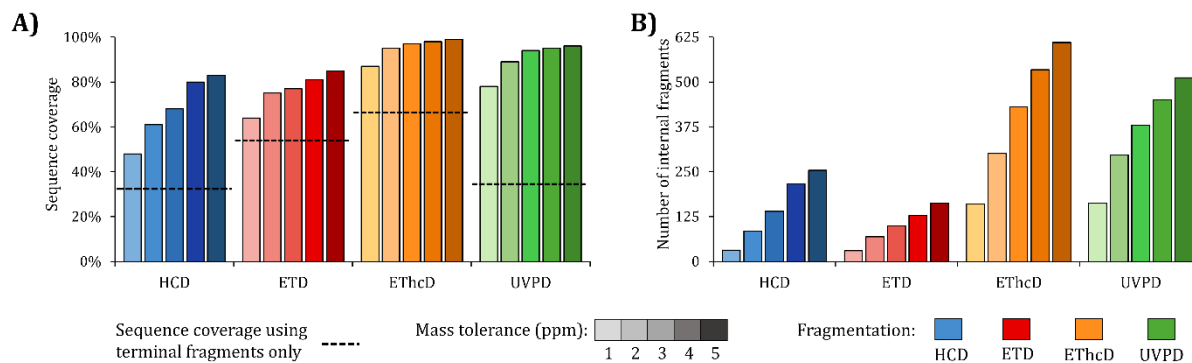


Figure 116: Identification results of the trastuzumab Lc subunit when varying the mass tolerance for internal fragments attribution. (A) Sequence coverage (%) and (B) number of internal fragments identified. Results from HCD are displayed in blue shades, red for ETD, orange for EThcD and green for UVPD. Darker shades indicate a higher mass tolerance. Dashed lines represent the sequence coverage when only terminal fragments are considered.

This improvement of the sequence coverage is directly related to the higher number of internal fragments identified when the mass tolerance was increased. Notably, it was multiplied by around 2-2.5 times when the threshold was raised from 1 ppm to 2 ppm, while the increase was limited to 1.3-1.6 times switching from 2 ppm to 3 ppm and from 3 ppm to 5 ppm (**Figure 116.B**). Additionally, the number of internal fragments that are detected is widely higher using EThcD and UVPD (160 and 163 at 1 ppm, respectively) compared to HCD and ETD (32 and 30 at 1 ppm, respectively). Indeed, EThcD and UVPD have demonstrated their ability to provide additional fragmentation pathways that are not accessible with collision-based dissociation; consequently, the probability of generating internal fragments is increased using this technique. Particularly, using HCD as a second activation method right after ETD during EThcD increase the internal energy of the protein and can lead to additional fragmentation of the backbone. Because these two fragmentation methods have different preferential cleavage sites, the potential number of internal fragments generated is also increased. Moreover, only one type of internal fragments is considered for HCD and ETD (*by* ions and *cz* ions, respectively), whereas four and nine ion types were queried for EThcD (*by*, *bz*, *cy*, and *cz*) and UVPD (*ax*, *ay*, *az*, *bx*, *by*, *bz*, *cx*, *cy*, and *cz*), respectively. It results in a more significant number of matching possibilities, estimated up to 21,000 for HCD/ETD (one internal ion type (e.g., *by* or *cz*)), up to 84,000 for EThcD (four ion types), and 188,000 for (UVPD) nine ion types. The huge number of theoretical internal fragments also increase the risk of false positives for these fragmentations.

Notably, matching one internal fragment results in the identification of two inter-residue cleavage sites, while only one site is characterized when considering terminal fragments. For this reason, the contribution of the internal fragments to the characterization of the sequence coverage is higher even with a low number of internal fragments matched (**Figure 116**). In addition, the presence of multiple isobaric internal fragments can also increase the sequence coverage from a single deconvoluted mass. For instance, if two isobaric fragments are matched, they will result in four inter-residue cleavages (instead of two for a single internal fragment). The proportion of these isobaric fragments is not negligible. For instance, the Lc subunit of trastuzumab exhibits 3,510 *by*-type internal fragments with

at least one isobaric counterpart (1,675 isobaric masses in total), representing 16% of all the possible *by*-type internal fragments for this sequence (21,736 internal fragments). These isobaric fragments are found for multiple reasons (**Figure 117**). First, *ax*, *by*, and *cz* internal fragments of the same amino acid sequence are isobaric (complementary ion series) and can be found when using techniques providing different multiple ion types (i.e., EThcD and UVPD). Nevertheless, the sequence coverage will not be increased, as all the ions correspond to the same inter-residue cleavage. Second, multiple internal fragments can have the same amino acid composition, leading to isobaric fragments. Third, multiple fragments with completely different amino acid sequences can have the same mass (considering or not the same ion type, e.g., *ax* and *bz*), resulting in the identification of both fragments. For the two last cases, the identification of multiple isobaric fragments lead to an higher increase of the sequence coverage, as four inter-residue cleavages are considered instead of two (if there are two isobaric fragments).

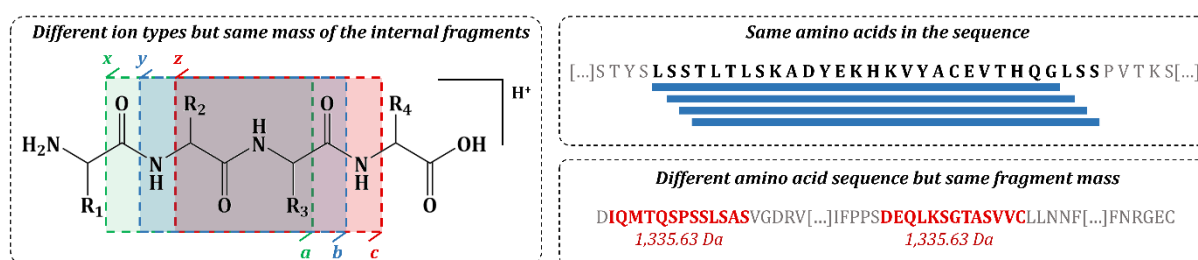


Figure 117: Types of theoretical isobaric fragments found in the trastuzumab Lc sequence.

In our example, around 17% of the internal fragments identified were isobaric using ETD (representing 9% of the deconvoluted masses detected), independently of the mass tolerance. In comparison, this proportion increases to 26% for HCD (14% of the total ions), 49% for UVPD (27%), and around 33% for EThcD (36%). Although, for the two latter methods mentioned, a significant part of the isobaric fragments identified are from the matching of the *ax* (UVPD only), *by* and *cz* fragments of the same amino acid sequence, as expected. Hence, they do not bring any additional sequence coverage.

In addition, as performed for terminal fragments, we searched internal fragments in the MS/MS spectra against reversed and scrambled sequences to evaluate the risk of false positive assignment with the different parameters. Nevertheless, no score reflecting the probability of mismatching is provided by the software. Hence, only the sequence coverage and the number of identified fragments are used. Slightly lower sequence coverages are observed than the theoretical sequence using the reverse amino acid backbone, with 20% less for ETD, 10% for HCD, 4% for EThcD, and the same coverage for UVPD (**Figure 118.A**). Nevertheless, this decrease was only caused by the reduced number of terminal fragments identified, as the same or slightly higher internal fragments were detected using the reverse sequence. Indeed, reversing the amino acid series order is not supposed to impact the detection of internal fragments because the mass will remain the same, independently of the order of the amino acids, as observed for HCD (**Figure 118.B**). Differing from HCD, other techniques depict more internal fragments identified. This slight difference comes from *c* and *z* fragmentation at proline residues, which the algorithm considers improbable and thus removed. Hence, some internal fragments are eliminated depending on the order of the amino acids. In contrast, all other internal fragments with no proline at their cleavage site were identified using theoretical and reversed sequences. For the scrambled sequences (the identical five shuffled sequences used previously), the average sequence coverage is lower for ETD (-28%), HCD (-12%), and EThcD (-7%) while comparable for UVPD (-1%) compared to the theoretical sequence. Here again, the decrease of the sequence coverage is due to the absence of terminal fragments identifications as comparable (for HCD and ETD) or higher (for EThcD and UVPD) number of internal fragments detected. Among these matches, it should be noted that some correspond to isobaric masses of the theoretical sequence, as

between 150 and 300 internal fragments from the scrambled sequences are isobaric to those of the theoretical one.

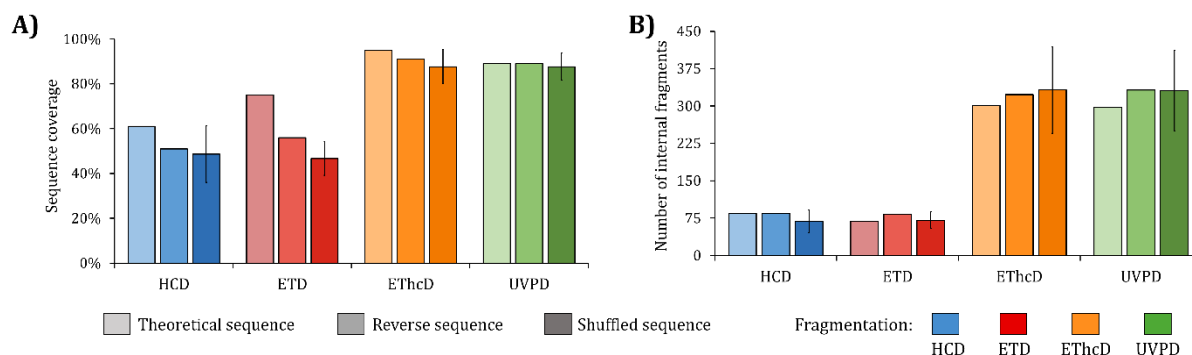


Figure 118: Identification results of the trastuzumab Lc subunit using the theoretical sequence, the reverse sequence or shuffled sequences. (A) Sequence coverage and (B) number of internal fragments identified. Results from HCD are displayed in blue shades, red for ETD, orange for EThcD and green for UVPD.

To summarize, considering internal fragments significantly improved the sequence coverage independently of the fragmentation technique used, as observed for mAb subunits in Wei *et al.*³³⁶. The assessment of the mass tolerance variation corroborates the low threshold generally used (i.e., 1 or 2 ppm) to mitigate the risk of false positives. Increasing the mass tolerance would drastically raise the risk of false positives, as demonstrated by the highest number of ions matched. For instance, analysis of the shuffled sequences with a 2 ppm mass tolerance on HCD data gives a sequence coverage of 50% for the shuffled sequences, against 41% at 1 ppm, 74% at 3 ppm and 85% at 5 ppm. Similarly, the number of internal fragments from the shuffled sequences identified increases with the mass tolerance, from 47 at 1 ppm, to 68 at 2 ppm, 132 at 3 ppm and more than 200 at 5 ppm, evidencing the increase risk of false positive matching. Because they are fit to isobaric fragments, some deconvoluted masses have a non-negligible influence on the sequence coverage. Nevertheless, it is impossible to corroborate if the ion detected in the MS/MS spectra is the fragmentation result of either both isobaric amino acid sequences or one of them. In addition, scrambling the sequence to evaluate the risk of false positives demonstrates that the probability of random mass matching cannot be ignored, with many internal fragments identified in these scrambled sequences. In this regard, internal fragments should be used with extreme caution for protein identification or sequence confirmation, even when setting the mass tolerance to a relatively low level (i.e., ≤ 2 ppm).

B. Evaluation of internal fragments for sequence modification localization

One main objective of TD/MD-MS experiments is the identification of proteoforms. However, the contribution of terminal fragments sometimes is not enough to identify the position of the modification. This task is particularly challenging when the modification is located in the mid-region of the proteins. Thus, we assessed the use of internal fragments to pinpoint the specific position of some PTMs. For this task, the use of internal fragment identification were included to determine the position of a well-characterized and common PTM in mAbs structure, i.e. the G0F glycoform of the Fc/2 subunit of trastuzumab.

The position of glycosylation on mAbs is well known, and this modification has been extensively studied in the literature^{375,609}. The G0F glycan is located on the residue Asn61 of the trastuzumab Fc/2 subunit, with a monoisotopic mass of 1,444.534 Da. We studied the sequence coverage variation upon modification of the position of the glycan moiety to other Asn residues in the sequence, namely Asn40, Asn89, Asn125, and Asn185. **Figure 119.A** shows high and comparable sequence coverage with internal fragments, regardless of the G0F modification position, except for ETD, where N61 and N89

demonstrate the highest sequence coverage. For all fragmentation techniques, the PTM location at N125 or N185 slightly lower the sequence coverage observed. Thus, it is impossible to provide clear-cut evidences of the PTM location at Asn61 based on internal fragments consideration. On the opposite, results from terminal fragments only corroborate the position at Asn61, with the highest sequence coverage obtained at this position for EThcD (59% for N61 versus <53% for other locations), ETD (48% for N61 versus <44% for other positions) and UVPD (30% for N61 versus <28% for other locations). At the same time, HCD does not allow to discriminate the position of the modification (Figure 119.A).

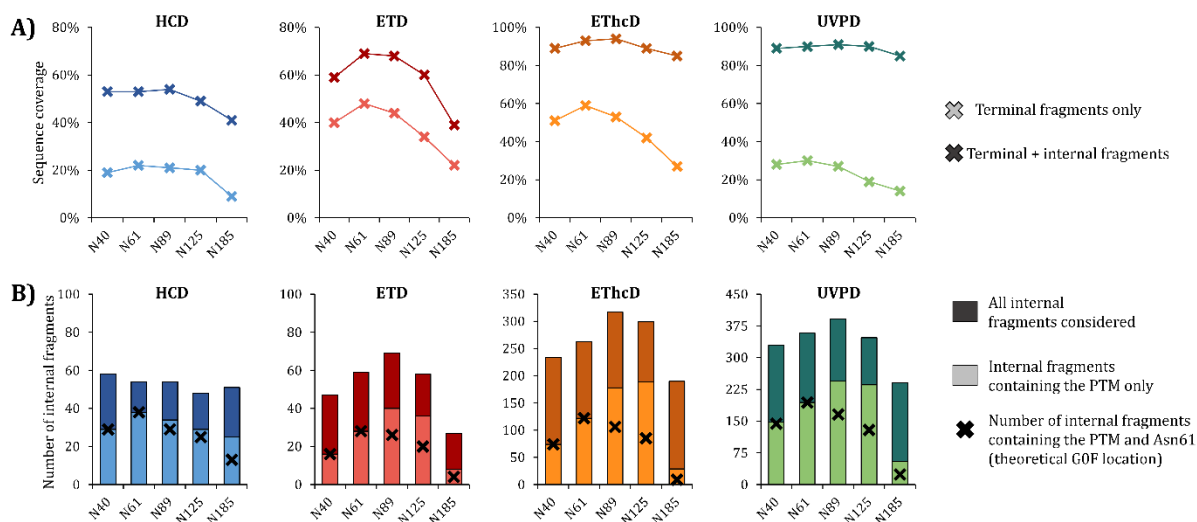


Figure 119: Identification results of the trastuzumab Fc/2 subunit when changing the GOF modification position among the protein backbone. (A) Sequence coverage obtained for HCD, ETD, EThcD and UVPD depending on the position of the modification on Asn40, Asn61, Asn89, Asn125, and Asn185 residues. (B) Number of internal fragments identified for each position (dark colors) and number of internal fragments containing the PTM (light colors) for HCD, ETD, EThcD and UVPD. Black crosses represent the number of internal fragments containing the PTM as well as Asn61, its theoretical position.

Hence, we looked into details to explain why the position of the GOF glycan is not confirmed using internal fragments. **Figure 119.B** illustrates that the total number of identified internal fragments raises when the modification is located close to the middle of the sequence (i.e., Asn89 and/or Asn125). In contrast, this number is lower for the N- and C-terminal positions (i.e., Asn40 and/or Asn185), except for HCD. It could be explained by the higher number of possibilities for amino acids on the mid-region of the sequence to be involved in internal fragments than those close to the N- or C-terminus of the protein, independently of the position of the GOF moiety. Then, we investigated the proportion of internal fragments containing the modification. For Asn40, 35%-50% of the internal fragments bear the GOF moiety, against 45%-70% for Asn61, Asn89, and Asn125 and 15%-50% for Asn185 (**Figure 119.B**), depending on the fragmentation technique used. Additionally, the presence of the Asn61 residue in most of the internal fragments with the modification hinders to certify the position of the GOF moiety on a specific Asn residue (**Figure 119.B**). For example, between 95% and 100% of the internal fragments bearing the PTM located at Asn40 also contain the residue Asn61. This percentage logically decreases with the distance between the putative Asn residues that can contain the PTM, rising from Asn61 to 60%-85% when the PTM is located at Asn89, to 45%-85% at Asn125, and 30%-50% at Asn185. On top of that, the intensity and size (90 amino acids on average) of the internal fragments bearing the PTM are comparable (within a fragmentation technique) independently of the location of the GOF modification, as well as isotopic resolution, disabling to use these criteria to discriminate between true and false positives.

These results point out that internal fragments alone can provide ambiguous results to decipher the position of PTM's. Internal fragments bearing the modification are generally large (>50 amino acids) and contain multiple amino acids that could potentially contain the modification (e.g., 20 Asn residues in the Fc/2 sequence distributed all over the protein backbone). On the contrary, terminal fragments are much more informative in locating the PTM, as the sequence coverage and the number of fragments containing the modification vary significantly with the correct position of the modification when using terminal fragments, and the trend of these parameters is not well defined when internal fragments are included in the study.

6. Conclusion and perspectives

This chapter highlights the importance of data processing for characterizing proteins in TD/MD-MS experiments. In each step of analyzing data, crucial parameters significantly influencing the sequence coverage and the number of fragments identified must be optimized to mitigate the risk of false positives. The first part showed the need for a relatively strict S/N ratio and fit factor values to ensure the proper deconvolution of fragment ions, allowing obtaining a high sequence coverage and number of fragments identified, but also a low P-Score, evidencing a low probability of ion mismatching. The second part underlined that a too-stringent mass tolerance threshold for deconvoluted-ion matching should be avoided, as it would discard a significant amount of valuable fragment ions, while a too-large cutoff would not provide a significant gain in sequence coverage but increases the risk of false positive identification. In addition, after trying to match masses against reverse and shuffled sequences, it was observed that the risk associated to false positives identification was relatively low, especially when looking at the scores obtained for matching wrong protein backbones. Finally, the last part depicted the potential benefits and pitfalls of internal fragments consideration, showing that those ions provides a non-negligible benefit regarding the sequence coverage characterization. However, internal fragments implies also an increased probability for false positive identification, making them more appropriate to boost the sequence coverage and confirm the primary structure of the proteins instead of using them for protein identification or PTM localization. In addition, no score is available to evaluate the probability of false positives when dealing with internal fragments, which remains a drawback of this approach. Consequently, more weight should be given to terminal fragments than internal ones to assess the identification confidence. It is in line with the statistical fragmentation of proteins, demonstrating that the production of terminal ions is favored over internal ions and that they account for at least 50% of the total ion current (because of the signal dilution of internal ions for large proteins) increasing the reliability of the matching³³¹. Concomitantly, other strategies such as additional MS step (MS3 or PTCR) or the use of more stringent filters, regarding the intensity or the type of ions selected to search internal fragments, are potential alternatives to simplify the spectra and improve the confidence in terminal fragments identifications.

In TD/MS-MS experiments, no clear guidelines exist for data processing, making comparing studies intricate, even if some consensus ranges of values are commonly accepted regarding some parameters. Another point for TD/MD-MS data treatment is the multitude of algorithms available from the deconvolution step to proteoform identification. All of them have specific parameters and work differently. For deconvolution, noise detection and isotopic signal detection and matching have been addressed differently depending on the algorithms. TRASH³²³, developed in 2000 by Horn and coworkers, has been one of the core algorithms for TD-MS data preprocessing. It calculates the S/N ratio using noise filtering based on 4 Da windows. Other algorithms, such as AID-MS⁶¹⁰ and MasSPIKE⁶¹¹, identify the noise on the whole spectra to speed up the process. Another solution, MASH Suite^{327, 612}, allows users to adapt the S/N ratio for some specific window to detect low-intensity peaks often considered noise and removed them. These tools also employ different strategies to obtain the experimental isotopomer envelopes and identify the charge states, as described by Zhong and

coworkers³¹⁹. All these strategies have pros and cons, and none is perfect, as extracting monoisotopic peaks in complex MS/MS spectra containing multiple overlapping peaks remains hard.

Many algorithms have also been developed for experimental deconvoluted masses matching and proteoform identification, some more dedicated to identifying proteoforms in complex mixtures (i.e., identifying proteoforms is the key, with no need for extensive sequence coverage). In contrast, others are more suitable for protein characterization (i.e., working with a unique sequence and trying to obtain the maximum coverage). For proteoform identification, multiple tools, including ProSightPD⁶¹³, TopPIC^{329, 614, 615}, pTop^{616, 617} or MASH Suite^{327, 612} are generally used and have been recently compared³²⁵. These tools also include scores that are beneficial to discriminate between true and false positives and increase confidence in the identification result⁶¹⁸⁻⁶²¹. For purified protein characterization, ProSight Lite³²⁶ and related tools^{622, 623} are the most used as they allow a rapid analysis of the data.

In addition to the existing tools, new bioinformatics strategies to tackle the challenges of TD/MD-MS approaches are developed, as demonstrated by ClipsMS for identifying internal fragments. For example, MASH Native⁶²⁴ or ProSight Native⁶²⁵ can be used for native top-down, as it can process both isotopically resolved and unresolved MS spectra for deconvolution and protein identification. Several nodes available in the Proteome Discoverer software for the processing of low-resolution data also allow native TD-MS data analysis. In addition, an R package named TopPICR⁶²⁶ was also recently released, enabling to perform label-free proteoform quantification. Other tools are also available to visualize better TD/MD-MS-based data, including TopMSV⁶²⁷, MS-TAFI⁶²⁸, VisioProt-MS⁶²⁹, and PSpcterR⁶³⁰ or to evaluate and compare experimental conditions, using TDFragMapper⁶³¹.

All these tools demonstrate the plethora of solution available for top-down data analysis and the multiple possibilities to get the best from the data. One example illustrating the abundance of workflows available and settings used to analyze TD/MD-MS is depicted in an interlaboratory study from Srzentić and coworkers⁴³⁹, where each group participating in the study had its workflow, resulting in more than 20 strategies for characterizing purified proteins. Regarding this, newcomers in the field might feel lost by all the solutions at their disposal. For this purpose, it should be noted that efforts are being made by the community, and especially by the Consortium for Top-Down Proteomics⁶³², to set guidelines and standardized approaches, from the sample preparation to data analysis, to make it more accessible for people out of the field⁵⁹⁶.

Chapter 3:

Case study: Characterization of trastuzumab deruxtecan antibody-drug conjugate by MD-MS

1. Context of the project and analytical strategy

Antibody-drug conjugates (ADC) is a class of therapeutic proteins that is widely employed in oncology. Trastuzumab deruxtecan (T-DXd) is an antibody-drug conjugate (ADC) composed of the humanized anti-HER2 trastuzumab antibody with a cleavable peptidyl-based linker (GGFG) and a potent topoisomerase I inhibitor payload (DXd, 1,033.39 Da)^{633, 634} (**Figure 120**). T-DXd is used as monotherapy in third-line treatment for patients with metastatic HER2-positive breast cancer who have already received at least two previous anti-HER2 antibody-based therapies (trastuzumab/pertuzumab and trastuzumab emtansine (T-DM1))⁶³⁵. T-DXd production is performed using a conventional conjugated cysteine (Cys) strategy with a maleimide precursor targeting Cys residues of trastuzumab⁶³³. In a word, cysteines located in the hinge region and involved in interchain disulfide bridges of trastuzumab are reduced using tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl), followed by the addition of the precursor drug linkers to the reduced mAb. It leads to a high drug-to-antibody ratio (DAR) of 8 with homogeneous conjugation, enabling the efficient delivery of the payload to HER2-overexpressing tumor cells^{633, 636}.

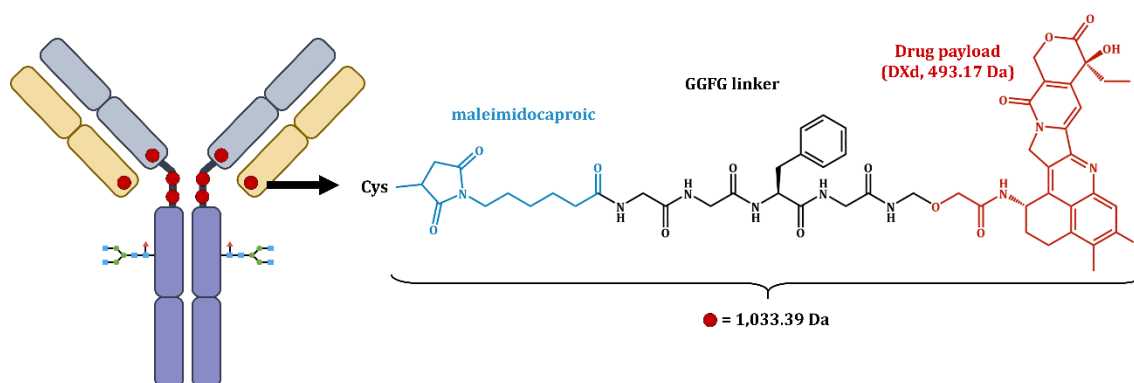


Figure 120: Trastuzumab deruxtecan (T-DXd) structure and zoom on the drug composed of a linker and the payload.

In addition to the intrinsic microvariability of mAbs⁵⁹³, the heterogeneity of the conjugates makes the characterization of ADCs much more challenging. Indeed, for these biomolecules, the drug-load distribution (DLD), average drug-to-antibody ratio (avDAR), and the amount of unconjugated mAb (DAR₀), all considered critical quality attributes (CQA) have to be thoroughly monitored³⁹⁷. In this context, native mass spectrometry (nMS), either as a standalone technique or in combination with liquid chromatography (LC), has significantly contributed to the characterization of ADCs with increased complexity giving valuable information about the DLD, DAR₀, avDAR⁴⁰⁵. However, this technique only provides a global picture limited to the number and distribution of the payloads at ADC intact or middle-level.

Peptide-mapping fingerprint (PMF) has been considered the gold standard for protein sequencing, PTM identification, and sequence variant characterization, but it suffers from some limitations, particularly for ADCs analysis, as the information between the DAR value and the identification of the conjugation site is lost⁴¹⁰. PMF also has evidenced some pitfalls to deciphering the residues where the

payload is conjugated. This problem mainly stems from the fact that most often, collisional activation techniques are used for peptide sequencing, which leads to the fragmentation of the payload (the ester bond between the linker and the drug), preventing the precise localization of the conjugation site^{441, 637}.

Consequently, TD/MD-MS have emerged as complementary approaches to PMF since these strategies envision the fragmentation at intact (TD-MS) or subunit level (MD-MS) with complementary activation techniques⁵⁹⁶. As mentioned previously in this manuscript, TD/MD-MS approaches allow obtaining significant sequence coverages on mAb subunits. Moreover, these techniques have also been recently applied successfully to ADCs using different instruments and fragmentation techniques to provide a thorough characterization of the primary structure^{9, 444, 638}. In addition, we already evidenced in this manuscript the benefits of additional strategies to improve the characterization mAb subunits. Among them, using PTRC after ETD and EThcD fragmentation with wide and narrow isolation windows increases the sequence coverage up to 15% (see **Part III, Chapter 1. 4**). Data processing considering internal fragments also boosts the number of cleavage residues observed and allow us to characterize completely the sequences of trastuzumab subunits after combination of multiple fragmentation techniques (see **Part III, Chapter 2.5**).

Regarding these optimizations and the results obtained previously, we decided to combine several MS-based techniques on the Orbitrap Eclipse to achieve the complete characterization of T-DXd, including the precise location of the drugs. First, LC-nMS at intact and middle-up levels was used to identify the different T-DXd species and thus determine the DLD and the avDAR values of the ADC. Moreover, analyses were performed in denaturing conditions to corroborate the position of the payloads on the different subunits. Second, TD-MS and MD-MS were performed respectively on the intact and digested T-DXd, to obtain comprehensive subunit sequencing and identify payload locations. Finally, we evaluated the benefit of internal fragments for sequence coverage improvement and the potential of PTRC reaction on subunits bearing payloads to reinforce the conjugation sites' location.

2. Intact and middle-level analysis of T-DXd for drug-load-distribution (DLD) determination and payload location assessment at the subunit level

A. Intact and middle-up level analysis of T-DXd in SEC-nMS

The analysis of T-DXd was firstly envisaged at the intact level with the use of the size-exclusion chromatography (SEC)-nMS coupling to provide a precise mass measurement to identify all the different T-DXd species and thus allow determining the DLD, and the avDAR values of the ADC. The combination of SEC with nMS has proven several benefits on the analysis of therapeutic proteins among which increased nMS throughput with the possibility to characterize and relatively quantify the high (HMWS) and low molecular weight species (LMWS)^{206, 639}. Similar to standalone nMS analysis, SEC-nMS coupling has demonstrated its ability to maintain the non-covalent interactions between the different chains, being of utmost importance during the characterization of cysteine-linked ADCs, like T-DXd⁶³⁹. These experiments were conducted on the Eclipse Orbitrap Tribrid instrument, enabling to assess its performances for native MS approaches compared to other instruments including the Orbitrap Exactive Plus EMR (Thermo Fisher Scientific, Bremen, Germany) and Synapt G2 HDMS (Waters, Manchester, UK), that were already used in the lab by Dr. Evolène Deslignière to perform T-DXd analysis in native conditions.

The chromatogram of intact T-DXd depicted in **Figure 121.A** exhibits only one peak corresponding to the monomeric T-DXd (**Figure 121.B**). The experimental mass of the most intense peak is $156\,338 \pm 1$ Da and can be assigned to the G0F/G0F T-DXd glycoform bearing 8 conjugated drugs. Three other glycoforms were also detected in different proportions: G0F/G1F, G0F/G2F or/and G1F/G1F and G0F-N/G0F (**Figure 121.C**). This result confirms the homogeneity of the T-DXd, with a unique DLD and

with the absence of other DAR populations. Moreover, it pinpoints the ability of the experimental parameters used throughout the SEC-nMS analysis to maintain non-covalent interactions of intact T-DXd structure, although the drugs are conjugated on the cysteines usually involved in the inter-chain bonds, being consistent with the literature⁴²⁷. Of note, only minor proportions of conjugated Lc is observed in the low m/z region of the spectrum, confirming the suitability of the experimental parameters to limit the disruption of the non-covalent interactions of the ADC (**Figure 121.D**).

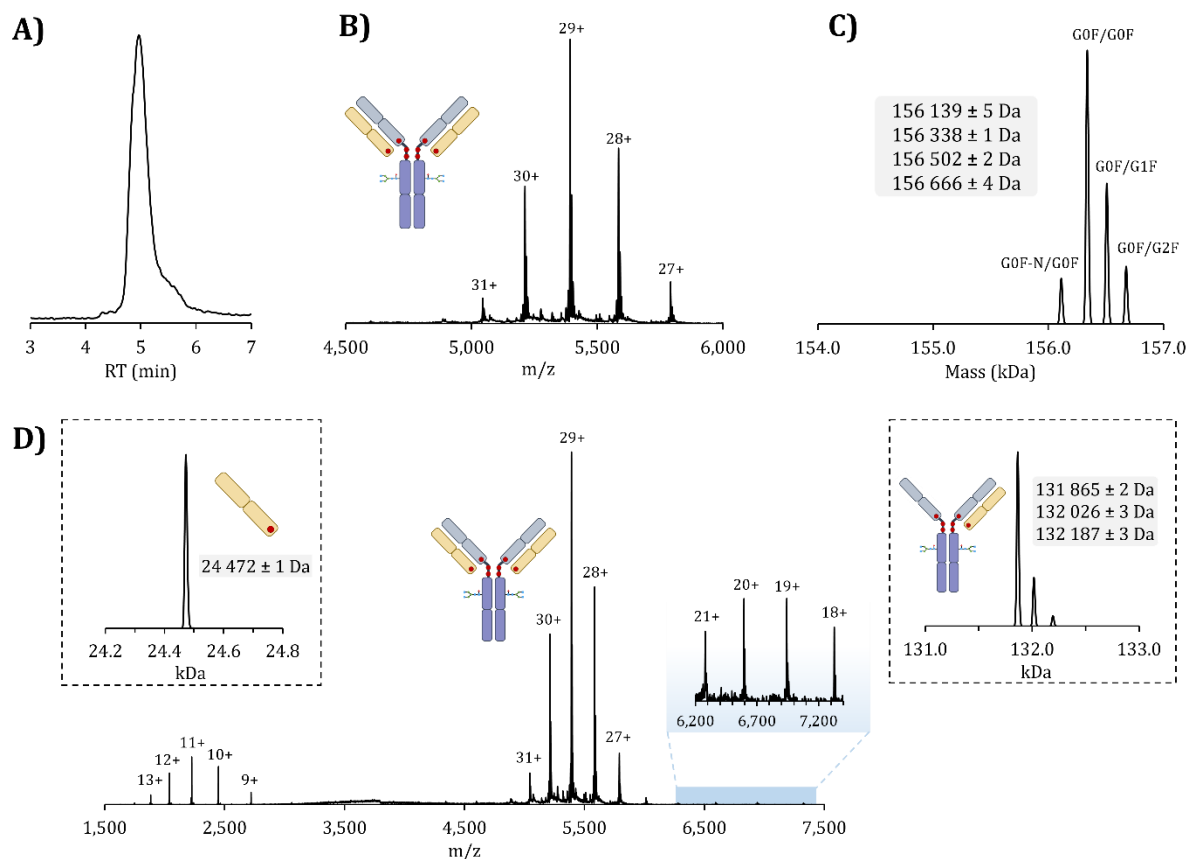


Figure 121: SEC-nMS analysis of intact T-DXd. (A) SEC-nMS chromatogram of intact T-DXd. (B) nMS spectra of the intact T-DXd peak detected. (C) Mass deconvolution of the intact monomer. (D) nMS spectra of intact T-DXd including Lc release (around 2,000 m/z) and intact T-DXd with missing Lc, and their mass deconvolution (insets).

Subsequently, middle-up level analysis of T-DXd was performed to provide a more precise location of the drugs on the different subunits of the ADC. In this particular case, Fc and Fab' domains were obtained after IdeS enzymatic digestion, degrading the scaffold of the ADC below the hinge region. The corresponding chromatogram is composed of two main peaks standing for the Fc and Fab', respectively (**Figure 122.A**). The deconvolution profile of the Fc domain points out four glycoforms with no conjugated drug ($50,466 \pm 1$ Da for the main glycoform G0F/G0F, **Figure 122.B**) as demonstrated with the previous LC-MS data. For the peak at 8.5 min, a single compound was identified, corresponding to the Fab' carrying 4 drugs ($52,955 \pm 1$ Da, **Figure 122.C**). Observed retention time difference between Fc and Fab' species (similar molecular weight) is explained by the interactions between the column stationary phase and the Fab' compound bearing four drugs, as already observed⁴²⁷. These results confirm that the cytotoxic payloads are conjugated on the upper part of the mAb ($F(ab')_2$ domain). Since the bioconjugation strategy provides a homogeneous conjugation of the ADC, the avDAR can also be calculated based on the middle-level analysis leading to an avDAR value of 8 (two Fab' domains per mAb).

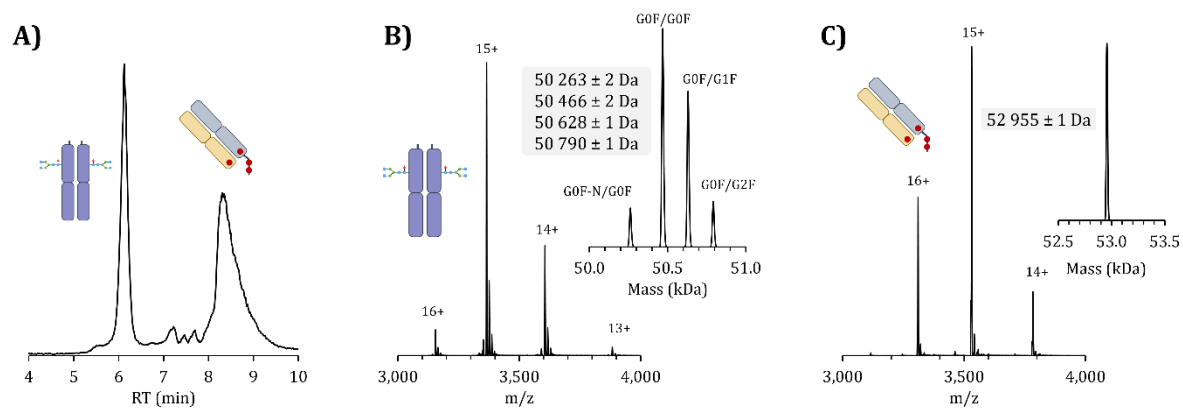


Figure 122: SEC-nMS analysis of IdeS digested T-DXd. (A) SEC-nMS chromatogram of IdeS digested T-DXd. (B) nMS spectra of the T-DXd Fc subunit (peak at 6 min), and the mass deconvolution (inset). (C) nMS spectra of the T-DXd Fab' subunit (peak at 8.2 min), and the mass deconvolution (inset).

To summarize, SEC-nMS strategies (i.e., intact and middle-level analyses) employed confirmed the avDAR and afforded the conjugation location on the F(ab')₂ domain of T-DXd.

B. Intact and middle-up level analysis of T-DXd in denaturing conditions

After performing DAR determination and intact as well as middle-up mass analysis in native conditions, denaturing conditions were used to separate and characterize the subunits of T-DXd (rpLC-MS) either after reduction of disulfide bridges (heavy (Hc) and light chains (Lc)) or upon IdeS digestion followed by DTT reduction (Lc, Fd, Fc/2 subunits). Indeed, denaturing conditions allow to disrupt non-covalent interactions and provide additional information regarding the conjugated subunits.

Reduced T-DXd analysis reveals two main species eluting at 23.6 min and 30.5 min (**Figure 123.A**). The first one corresponds to the Lc carrying one drug (24,476.9 ± 0.2 Da, **Figure 123.B** and **Table 11**) and the second one to the heavy chain (Hc) with three conjugated drugs (53,703.6 ± 0.4 Da for the main glycoform G0F, **Figure 123.C** and **Table 11**).

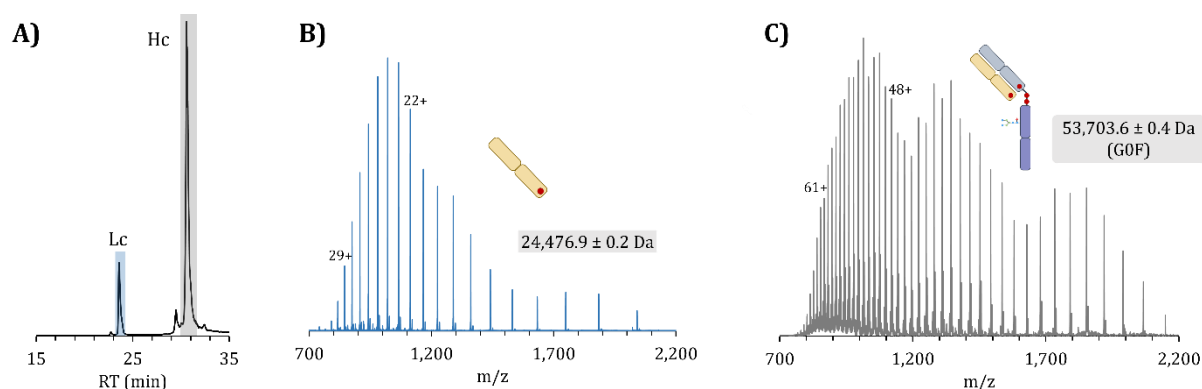


Figure 123: rpLC-MS analysis of DTT reduced T-DXd. (A) rpLC-MS chromatogram of reduced T-DXd with two peaks (23.5 min, Lc and 30.5 min, Hc). Mass spectra of the (B) Lc, and (C) Hc (G0F) subunits.

Both peaks exhibit additional populations at slightly lower retention times corresponding to partial payload scaffolds (-475 Da)⁴²⁷. In addition, the deconvoluted masses obtained for the heavy chain glycoforms validate the presence of the G0F/G2F species with the identification of the G2F glycoform (**Table 11**).

Table 11: Experimental and theoretical masses of the reduced and IdeS digested + reduced T-DXd.

Subunit	Experimental mass	Theoretical mass	Error (ppm)
<i>DTT reduced T-DXd</i>			
Light chain + 1 drug	24 476.9 ± 0.2 Da	24 477.1 Da	8
Heavy chain G0F-N + 3 drugs	53 500.0 ± 0.9 Da	53 500.8 Da	15
Heavy chain G0F + 3 drugs	53 703.4 ± 0.4 Da	53 703.9 Da	9
Heavy chain G1F + 3 drugs	53 865.3 ± 0.8 Da	53 866.1 Da	15
Heavy chain G2F + 3 drugs	54 028.1 ± 0.9 Da	54 028.2 Da	2
<i>IdeS digested + DTT reduced T-DXd</i>			
Fc/2 G0F-N	25 032.8 ± 0.3 Da	25 033.1 Da	12
Fc/2 G0F	25 235.9 ± 0.3 Da	25 236.2 Da	12
Fc/2 G1F	25 398.1 ± 0.2 Da	25 398.4 Da	12
Fc/2 G2F	25 560.2 ± 0.5 Da	25 560.5 Da	12
Lc + 1 drug	24 476.9 ± 0.2 Da	24 477.1 Da	8
Fd + 3 drugs	28 485.4 ± 0.2 Da	28 485.7 Da	10

Then, the analysis of IdeS digested and reduced T-DXd showed three main peaks (**Figure 124.A**), corresponding to the Fc/2 (23,235.9 ± 0.3 Da for the main glycoform G0F, **Figure 124.B**), Lc with one drug (24,476.9 ± 0.2 Da, **Figure 124.C**) and Fd bearing three drugs (28,485.5 ± 0.2 Da, **Figure 124.D**) subunits, respectively. In line with previous results for reduced T-DXd, minor chromatographic peaks resulting from Lc and Fd bearing partial drug(s) (-475 Da) are observed prior to the major chromatographic peaks of these subunits, but no drug-free species were detected (**Figure 124.A**).

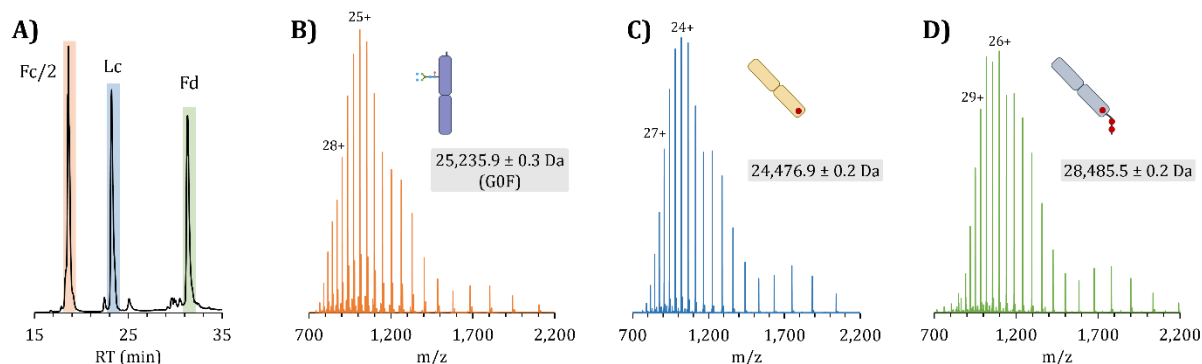


Figure 124: rpLC-MS analysis of IdeS digested and DTT reduced T-DXd. (A) rpLC-MS chromatogram of IdeS digested and reduced T-DXd with three peaks (18.5 min for Fc/2, 23.5 min for Lc and 31.5 min for Fd). Mass spectra of the (B) Fc/2, (C) Lc, and (D) Fd subunits.

For both experiments, no drug-free species were detected at intact and middle-level, in agreement with SEC-nMS, corroborating the homogeneous avDAR 8.0 previously calculated. Moreover, they pinpoint the conjugation of one and three drugs on the Lc and Fd subunits, respectively.

3. MD-MS analysis for sequence characterization and precise drug location

The combination of SEC-nMS and rpLC-MS allowed us to set the global picture of T-DXd by confirming its mass and identifying the number of conjugation per subunit. Subsequently, we performed MD-MS analysis of T-DXd at the subunit level to locate the drug conjugation sites precisely. The precursor ions selected for MS/MS fragmentation are detailed in the **Experimental section** for each subunit considered. In general, abundant precursor (with a significant charge state) was selected for CID. In

contrast, highly charged precursors were preferred for electron-based fragmentations (ETD/ETHcD), as already evidenced in **Part III, Chapter I**. Conversely to classical MD-MS strategies where sequence coverages are optimized, usually leading to accurate identification of drug location, we decided here to first address the position of the payloads by monitoring the sequence coverage as a function of drug bound to different cysteine residues. Once the “most probable” conjugation site is determined, the MS/MS data obtained from multiple fragmentation techniques were combined to increase the overall sequence coverage of the subunits. Of note, activation energy optimizations will not be displayed in this section, as they are similar to what has been performed in **Part III, Chapter 1** for trastuzumab. Moreover, UVPD was not used for T-DXd because of the low fragmentation efficiency observed during the optimizations; CID was used instead of HCD as it presented higher sequence coverage.

A. T-DXd Lc and Hc subunits MD-MS analysis

For reduced T-DXd, Lc+1Dxd or Hc+3Dxd species were selected for further CID, ETD, and ETHcD fragmentation. In this context, MS/MS spectra of the Lc were matched with 5 different positional isomer sequences (5 Cys residues in the Lc sequence). Similarly, those of the Hc were compared to the sequences of 165 different isomers (combinations of 3 conjugated Cys residues out of 11, **Figure 125**).

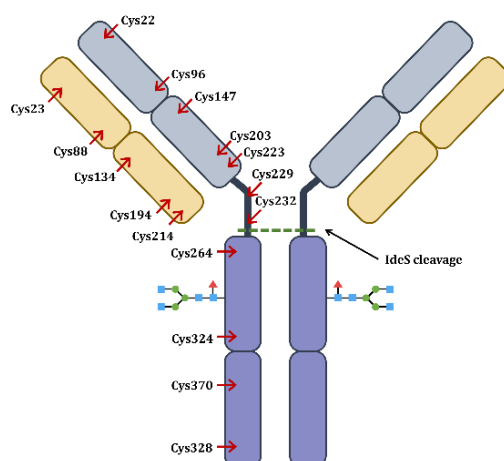


Figure 125: Putative conjugation sites of the payload on the heavy and light chains.

Upon fragment ion assignment with the different proteoforms, the conjugation site of the Lc subunit can be easily drawn based on the variation of the overall sequence coverage of the subunit when the conjugation site is hypothesized at each cysteine residue. From Lc MS/MS data interpretation, it can be concluded that the conjugation site is located on the C-terminal part (Cys194 or Cys214, **Figure 126**), as highest sequence coverages (14%/38%/42% for CID/ETD/ETHcD for Cys194-DXd versus 14%/39%/42% for Cys214-DXd) were achieved when considering DXd bound to one of those Cys residues. However, the z_{12} fragment obtained in ETD and ETHcD replicates, and the z_{16} obtained in ETD allows determining that the payload is precisely located on Cys214 residue (**Figure 127.A**).

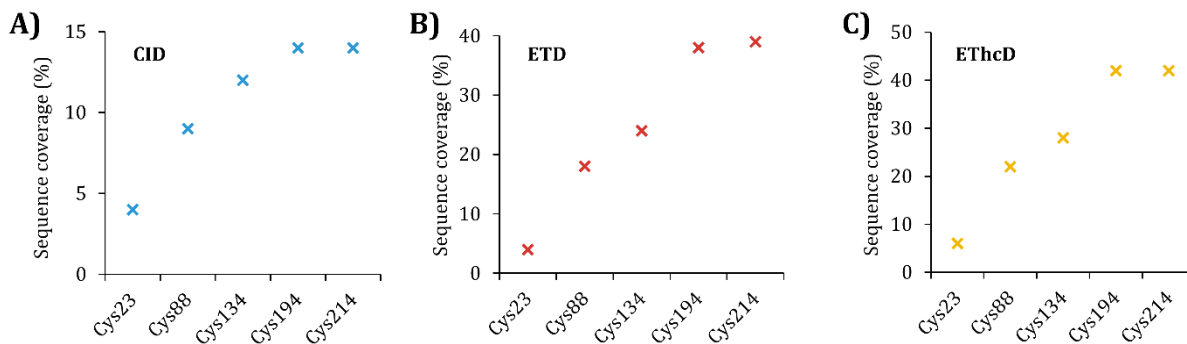


Figure 126: Sequence coverage for the T-DXd Lc depending on the conjugated cysteine position. Results obtained using (A) CID, (B) ETD and (C) ETHcD fragmentation.

Data interpretation is even more challenging for Hc as the sequence coverage obtained is comparable (8%-11%) when considering the theoretical positions (i.e., Cys223, Cys229, and Cys232) of the payloads, regardless of the activation technique used to induce the fragmentation. The fragmentation map of the combined results for the Hc points out the absence of fragmentation in the middle of the Hc sequence, which impairs the reliable location of the conjugation sites in this chain (**Figure 127.B**).

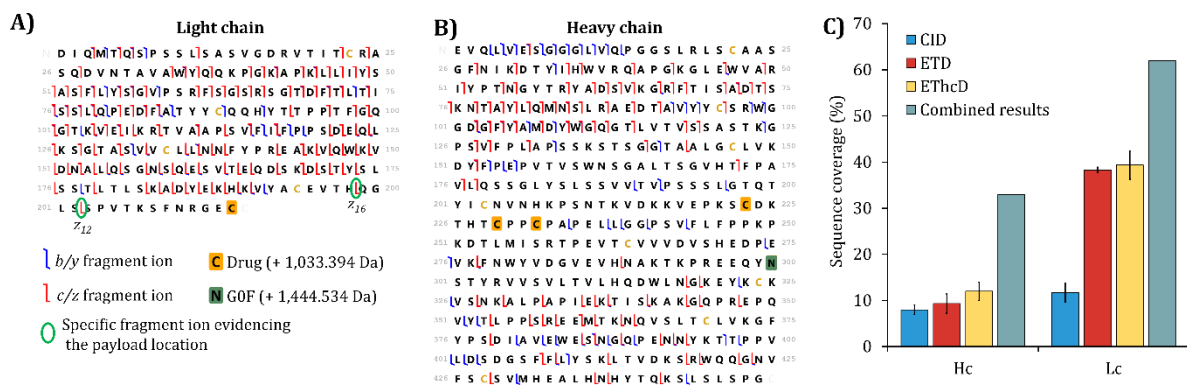


Figure 127: Fragmentation results for the T-DXd Hc and Lc subunits. Fragmentation maps of the (A) Lc and (B) Hc subunits, combining all the replicates. (C) Hc and Lc average sequence coverages obtained by CID, ETD and ETHcD for three injection replicates, and the combined results from all runs.

As expected and previously demonstrated, electron-based fragmentations (ETD/ETHcD) afforded higher sequence coverage for both chains (39% for Lc and 12% for Hc). Finally, sequence coverage upon the combination of CID, ETD, and ETHcD fragmentation replicates raised to 62% for the Lc and 33% for the Hc (**Figure 127.C**), which are still relatively low. Our results align with sequence coverages previously reported^{10, 432}, which can be explained by the difficulty of providing a complete sequence coverage characterization when fragmenting proteins larger than 30 kDa⁵⁹⁹. In the particular case of the Hc, this limited fragmentation hampers the precise localization of the conjugation sites.

B. T-DXd Fc/2, Lc and Fd subunits MD-MS analysis

Regarding the impossibility of obtaining comprehensive sequence coverage and information about the precise drug location on the Hc, T-DXd was IdeS digested and reduced to release the three ~25 kDa subunits (Fc/2, Lc, and Fd) and subsequently fragmented using the activation modes mentioned above. Similarly to results obtained on Lc generated after DTT treatment, MD-MS results obtained after IdeS digestion allow concluding that DXd is located on the Cys214 of Lc (**Figure 128.A**), as the two same diagnostic fragments specific to the drug conjugation on Cys214 were identified (z_{12} and z_{16} , **Figure 128.B**) upon ETD/ETHcD fragmentation.

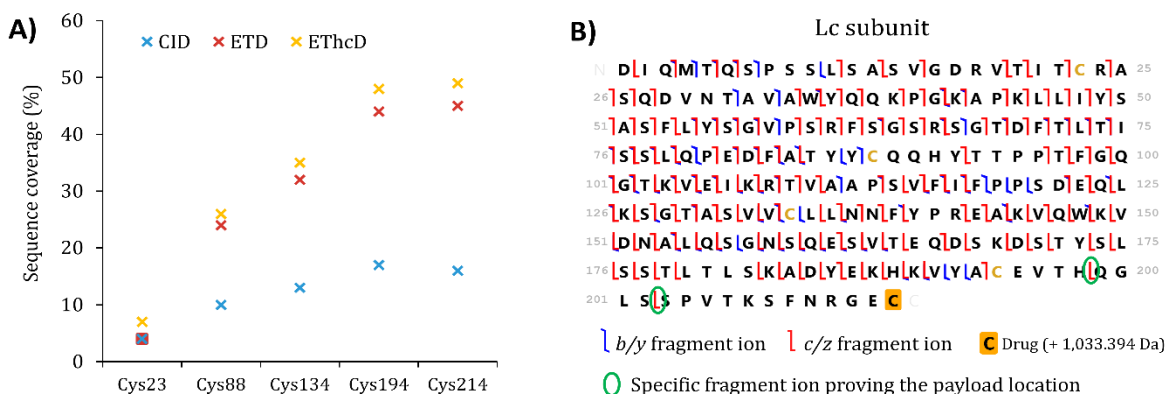


Figure 128: Fragmentation results for the T-DXd Lc subunit. (A) Sequence coverage depending on the conjugated cysteine positions for CID (blue), ETD (red), EThcD (yellow). (B) Fragmentation map combining all the replicates. Fragments evidencing the position of the drugs are circled in green.

As IdeS digestion efficiently downsized the Hc chain, we next analyzed the results from the 35 possible Fd isomers (3 conjugated Cys residues out of 7 Cys residues in the Fd, **Figure 125**). It came out that proteoforms including conjugated C-terminal Cys (i.e., Cys203, Cys223, Cys229, and Cys 232) led to the highest sequence coverage (**Figure 129.A**), suggesting possible DXd location on abovementioned Cys. Nevertheless, 7 fragments (all from ETD/EThcD) corroborated the presence of the payloads on the 3 C-terminal Cys residues initially involved in the inter-chain disulfide bridge of the mAb counterpart (**Figure 129.B**). Here, we observed fragment ions confirming the drug location for both Lc (Cys214) and Fd subunits (Cys223, Cys229, and Cys232). Moreover, after manual validation, it was confirmed that those ions exhibited resolved isotope profiles with no interfering signals and with a significant S/N ratio (> 150), thus increasing the confidence in their identification (**Figure 129.C**).

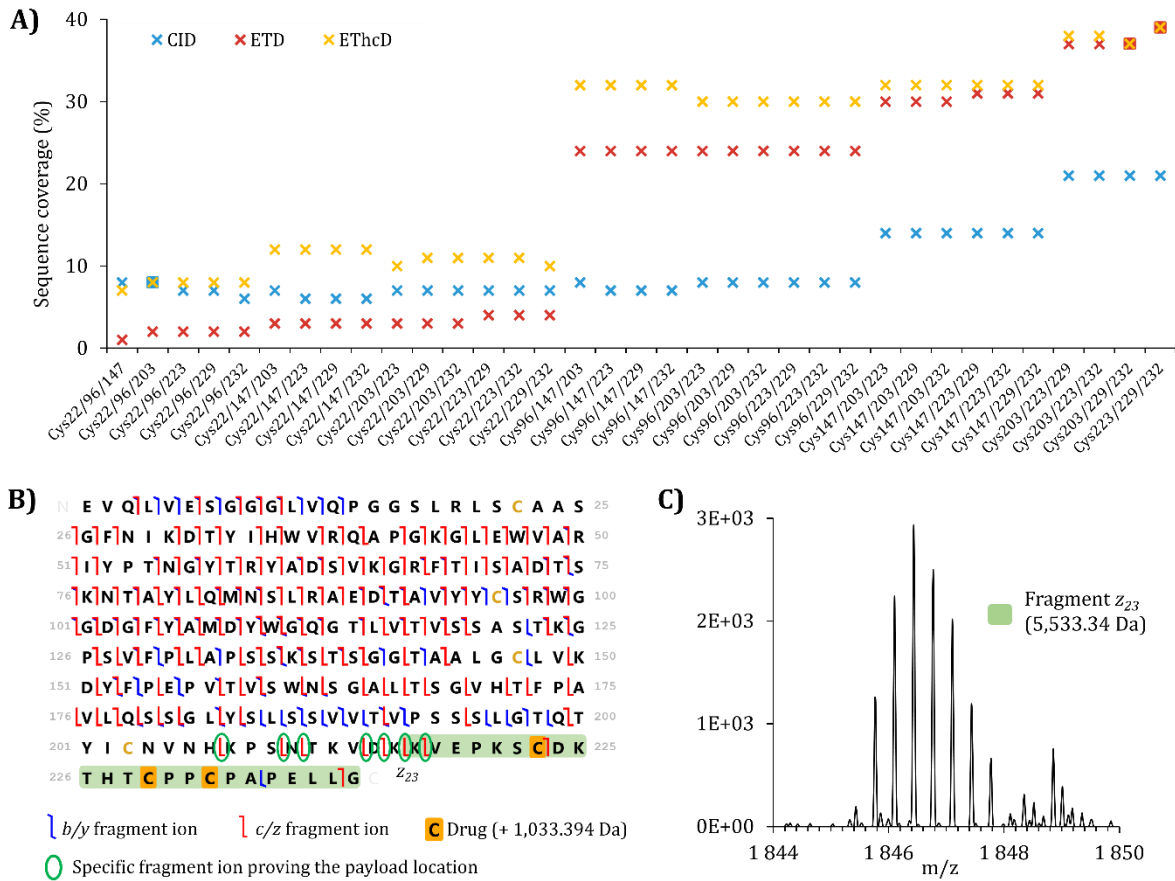


Figure 129: Fragmentation results for the T-DXd Fd subunit. (A) Sequence coverage depending on the conjugated cysteine positions for CID, ETD, and EThcD. (B) Fragmentation map combining all the replicates. Fragments evidencing the position of the drugs are circled in green. (C) Zoom on the z₂₃ fragment MS/MS spectra of the Fd subunit (1,846.5 m/z (3+)).

In addition, it should be mentioned that CID leads to the fragmentation of the drug, resulting in the release of the payload, which is less abundant in electron-based fragmentation (**Figure 130**).

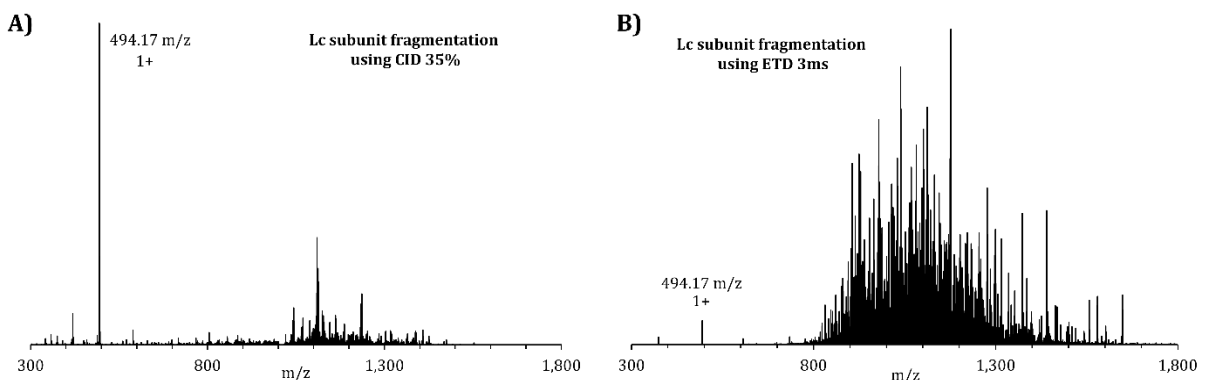


Figure 130: MS/MS spectra of T-DXd Lc subunit bearing one drug. Fragmentation performed using (A) CID 35% and (B) ETD 3 ms. Peak at 494.17 m/z stand for the released payload of the drug.

This side fragmentation is a critical point to consider for sequence-matching drug location, as it will influence the identification of all C-terminal fragments. It appears that fragmentation of the drug is not binary, as multiple species bearing the intact drug (i.e., linker + payload), with the release of one or various payloads, can be observed. However, the maximum sequence coverage was found without the payload (i.e., linker only, +540.233 Da) for Lc and with all the drugs intact (i.e., linker + payload, +

1,033.394 Da) for Fd (**Table 11**), even if an abundant signal corroborates the payload release at 494.17 m/z, illustrating the coexistence of multiple species.

Compound	Sequence coverage*	Number of fragments*
Lc + one intact drug (+ 1,033.394 Da)	10 %	26
Lc + one fragmented drug (+ 540.234 Da)	15.3 %	37
Fd + three intact drugs (3 x 1,033.394 Da)	19.3 %	52
Fd + two intact and one fragmented drugs (2 x 1,033.394 Da + 540.234 Da)	18 %	48
Fd + one intact and two fragmented drugs (1,033.394 Da + 2 x 540.234 Da)	15.3 %	26
Fd + three fragmented drugs (3 x 540.234 Da)	14.3 %	37

*average of three replicates

Table 12: Sequence coverage and number of fragments depending on the drug fragmentation.

Under those assumptions, the best sequence coverages achieved for T-DXd subunits for each fragmentation technique ranged between 29%-61% for Fc/2, 15%-48% for Lc, and 19%-37% for Fd, with best results originating from EThcD fragmentation (61%, 48% and 37% with EThcD for the Fc/2, Lc, and Fd parts, respectively). Interestingly, combining replicates from all fragmentation modes resulted in remarkably high sequence coverage for the Fc/2, Lc, and Fd subunits with 87%, 76%, and 72%, respectively (**Figure 131.A**), covering not only both termini of the sequences but also the internal regions as expected when fragmenting 25 kDa subunits (**Figure 128.B**, **Figure 129.B** and **Figure 131.B**).

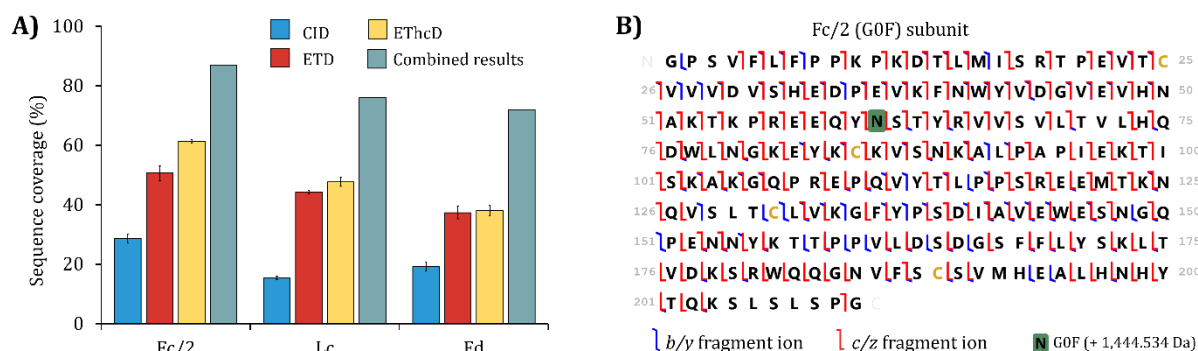


Figure 131: Fragmentation results of IdeS digested and DTT reduced T-DXd. (A) Fragmentation map of the Fc/2 subunit combining all the replicates. (B) Fc/2, Lc and Fd average sequence coverage obtained by CID, ETD and EThcD for three injection replicates, and the combined results of all runs.

In summary, MD-MS investigations of T-DXd subunits confirm the payloads conjugation on the Cys214 for the Lc (with Z_{12} and/or Z_{16} specific diagnostic ions). At the same time, DXd binding can be restricted to the C-terminal area of the Fd subunit only (Cys223, Cys229, or Cys232) with the detection of 7 specific fragment ions. As expected and previously observed, MD-MS results regarding sequence coverage were better with EThcD and after IdeS digestion, indicating that smaller-sized subunits (~25-30 kDa) are more adapted to MD-MS analysis.

4. Contribution of internal fragments for T-DXd subunits sequence characterization

In classical TD/MD-MS experiments, MS/MS spectra contain a vast amount of information. It has been shown previously that a significant part of the ions generated corresponds to protein backbone cleavages at different sites, giving rise to internal fragments. Regarding the benefits provided to increase the sequence coverage on trastuzumab subunits in **Part III, Chapter 2**, we used ClipsMS³³⁵ to tackle internal fragments generated after T-DXd fragmentation. Nevertheless, we demonstrated in this previous chapter that consideration of internal fragments increases the risk of ion mismatching, even

when applying low mass tolerance. We highlighted the lack of precision in identifying modification positions using those fragments. In the case of T-DXd subunits, several additional reasons make internal fragments not particularly suited to decipher the payloads' locations. Conjugation sites are mainly located in the C-terminal region of the Lc and Fd subunits, and hence, a meager amount of internal fragments will contain the payload moiety. Conversely, to terminal fragments, the sequence coverage variation obtained upon consideration of internal fragment ions is less sensitive to the payload's position. For instance, the sequence coverage of the Lc subunit upon ETD fragmentation increases from 4% to 45% when the conjugation site is assumed either at Cys23 or Cys214, respectively, clearly pinpointing that the conjugation site is located in the latter residue. However, upon internal fragment matching, the sequence coverage variation between these two conjugated positions is more attenuated (from 51% to 67%), leading to a more ambiguous result. Finally, internal fragment consideration was observed to have an increased probability of ion mismatching³³⁷. For example, the search for internal fragments after ETD by positioning the drug on the Cys23 residue of the Lc leads to the detection of more than 20 internal fragments containing the drug (**Figure 132.A**). For instance, one identified internal ion exhibits a clear isotopic profile with a precise mass measurement that could correspond with the 21-52 internal fragment bearing the payload (4,523.22 Da, **Figure 132.B**). Nevertheless, the Cys23 conjugation site was undoubtedly excluded from our previous MD-MS results based on terminal fragment identification because of the low sequence coverage assigned to this position. For all these reasons and from what has been previously observed on trastuzumab, the characterization of the conjugation sites relied on the identification of terminal fragments. In contrast, internal fragment ions were used with the primary purpose of boosting the sequence coverage of the different subunits, assuming the theoretical position of the payloads.

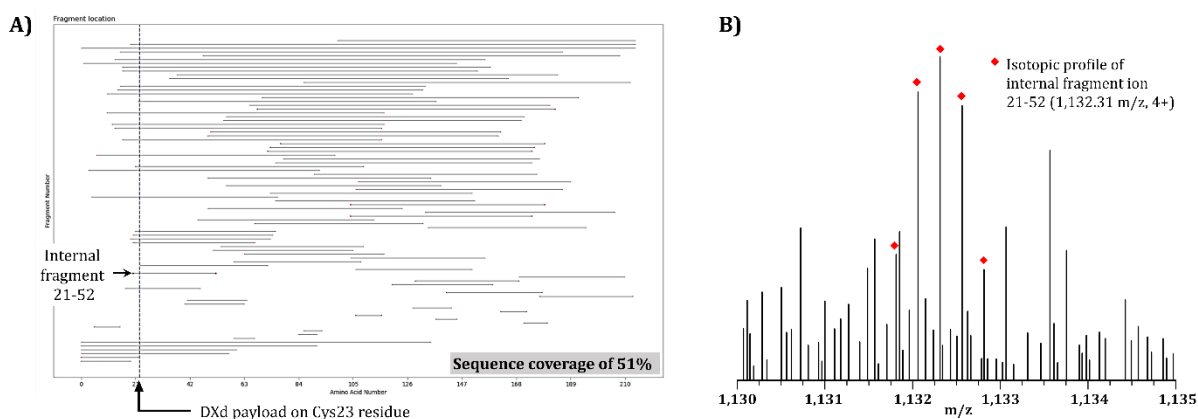


Figure 132: Lc subunit with DXd payload located on Cys23. (A) Localization map of internal and terminal fragments of the subunit. Dashed line represent the modification (+ 1,033.394 Da) on the Cys23 residue. (B) Isotopic profile of the internal fragment ion 21-52.

The results showed that the proportion of internal fragments identified differs from the fragmentation method. Depending on the subunit, the average number of internal fragments identified varies between 72 and 111 for HCD, 62 and 94 for ETD, and 214 and 327 for EThcD. As expected, EThcD provides more internal fragments, as *bz* and *cy* ions can be formed, in addition to *by* and *cz*. In the end, a total number of fragments (terminal and internal) around 130, 200, and 400 for CID, ETD, and EThcD, respectively, were detected (**Figure 133.A-C, right**). Consequently, the sequence coverage drastically increased when including internal fragments, reaching 50% to 60% with CID, 70% with ETD, and 90% for EThcD (**Figure 133.A-C, left**). Combining the sequence coverage from replicates of all the fragmentation techniques (i.e., CID, ETD, and EThcD) resulted in 100% subunits sequence confirmation. Concomitantly, the relative proportion of assigned fragment ions increased by 1.6 to 3 times when using internal fragment ion searching.

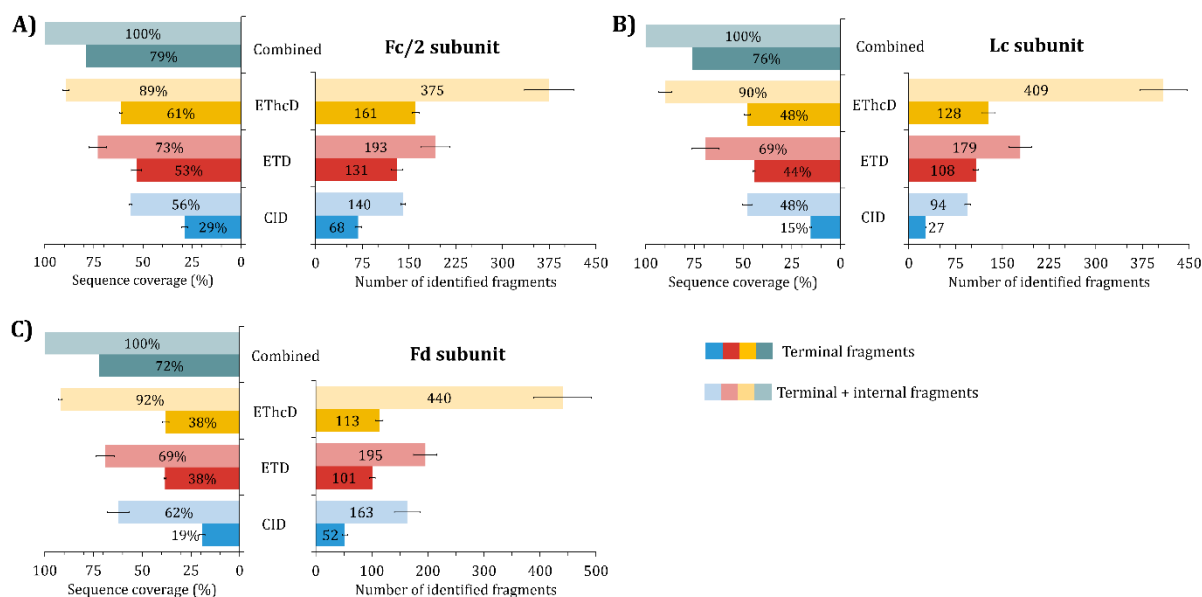


Figure 133: Average sequence coverage and number of fragment identified for T-DXd subunits. Average sequence coverage (left) and number of identified fragments (right) for (A) Fc/2, (B) Lc and (C) Fd subunits. Terminal fragments consideration only is represented in dark colors and terminal + internal fragments in light colors, for CID, ETD, EThcD triplicates and combined results from all the runs.

In summary, these results clearly show that considering internal fragment ion search during MD-MS workflows allows to take full advantage of all the information contained in the MS/MS spectra, highlighting the contribution of internal fragments to increase the sequence coverage for an improved subunit sequence confirmation of T-DXd. Despite this, internal fragments have a limited contribution to identifying conjugation sites.

5. Strengthening drug location using PTCR

The combination of MD-MS experiments and attribution of internal fragments allows us to obtain a complete sequence coverage of T-DXd subunits and find specific ions confirming the presence of the payloads at their theoretical positions. While internal fragments are beneficial for sequence coverage and allow identifying a significant part of the vast amount of information in MS/MS spectra, more is needed to decipher complex MS/MS spectra with numerous overlapping peaks. Thus, we decided to combine PTCR with the most efficient fragmentation techniques, i.e., ETD and EThcD, to declutter MS2 spectra and confirm drug localization. Indeed, PTCR has already proven its benefits to simplify MS/MS spectra and make possible the identification of additional fragment ions (see **Part III, Chapter 1**).

Here, two different PTCR methods were used: i) the “wide window” (1,000 m/z isolation window) method and ii) the “narrow window” (5 x 200 m/z isolation window) mode. As highlighted in **Figure 134.A and B**, the combination of PTCR results lead to an average raise of the sequence coverage by 10%-13% after ETD fragmentation and 2%-7% after EThcD fragmentation, respectively. The practical effect of this ion/ion reaction is further illustrated in **Figure 134.C and D**. After ETD fragmentation, a complex MS2 spectrum is generated, with many fragments distributed only over a small m/z range (700 – 1,700 m/z, **Figure 134.C**). Using PTCR for 10 ms on the 1,100 – 1,300 m/z window drastically reduced the spectral congestion of this zone of the MS2 spectrum, decreasing the signal interference in the lower m/z, where the major part of the fragment ions of interest are observed. For instance, overlapping signals and interferences are observed in the ETD MS2 spectrum of the diagnostic fragment ion z_{12} (2,341.00 Da), which allows the precise localization of the DXd on the Lc (**Figure 134.D**). Thanks to the addition of PTCR, the same ion could be assigned with a resolved isotope profile and a low interfering signal, leading to a more reliable identification of the fragment.

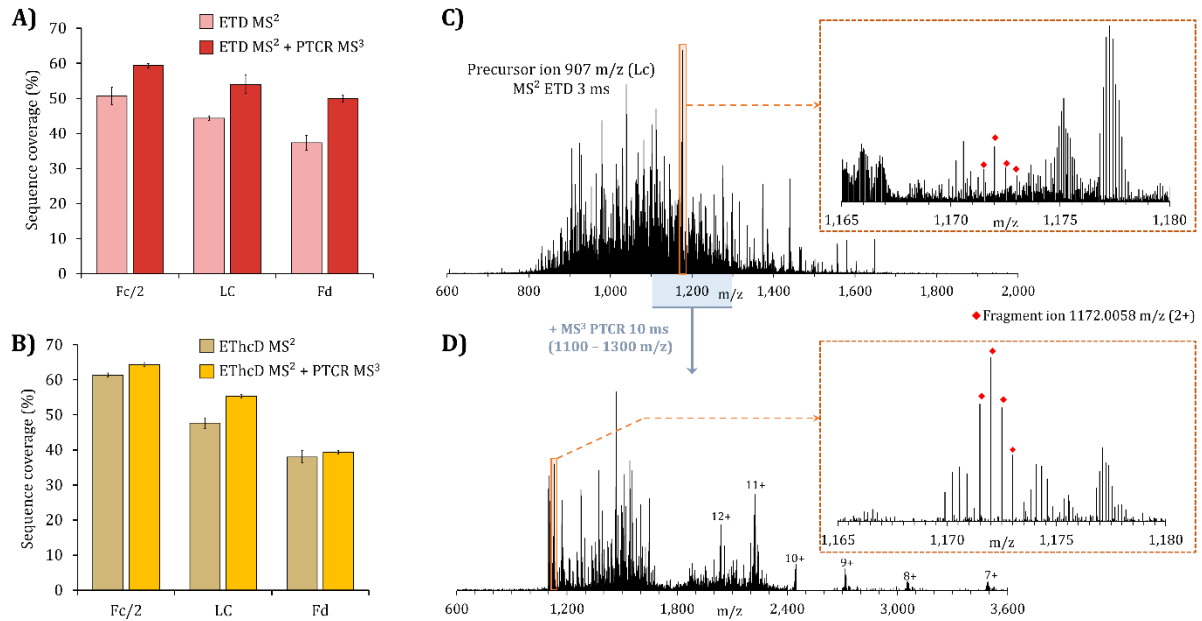


Figure 134: PTZR fragmentation results for the T-DXd Fc/2, Lc and Fd subunits. Fc/2, Lc and Fd average sequence coverage obtained after (A) ETD only (light red), ETD + PTZR (dark red) and (B) EThcD only (light yellow), EThcD + PTZR (dark yellow) for the three injection replicates. (C) Lc ETD (3 ms) fragmentation spectra (907 m/z (27+) precursor ion) and zoom on the 1,172.0058 m/z fragment ion (inset). (D) Lc PTZR (10 ms, performed on the 1,100-1,300 m/z isolation window of B) fragmentation spectra of the Lc and zoom to the 1,172.0058 m/z fragment ion (inset).

However, the PTZR MS2 spectra decluttering capabilities afforded more confident identification of diagnostic ions detected without PTZR and, more interestingly, allowed identification of additional specific diagnostic fragment ions on the different subunits. Upon addition of PTZR, a total of 5 fragment ions specific of the Cys214 conjugation on the Lc (2 fragments found with ETD/EThcD and 3 additional found after PTZR implementation (Z_{13} , Y_{19} and Y_{20} , **Figure 135.A**) and 11 signature fragments specific of the conjugation of the Fd subunit on the Cys223, Cys229, and Cys232 residues (7 fragments found with ETD/EThcD and 4 additional fragments detected after PTZR implementation (Z_{26} , Z_{27} , Z_{30} , and Z_{32} , **Figure 135.B**) were finally seen.

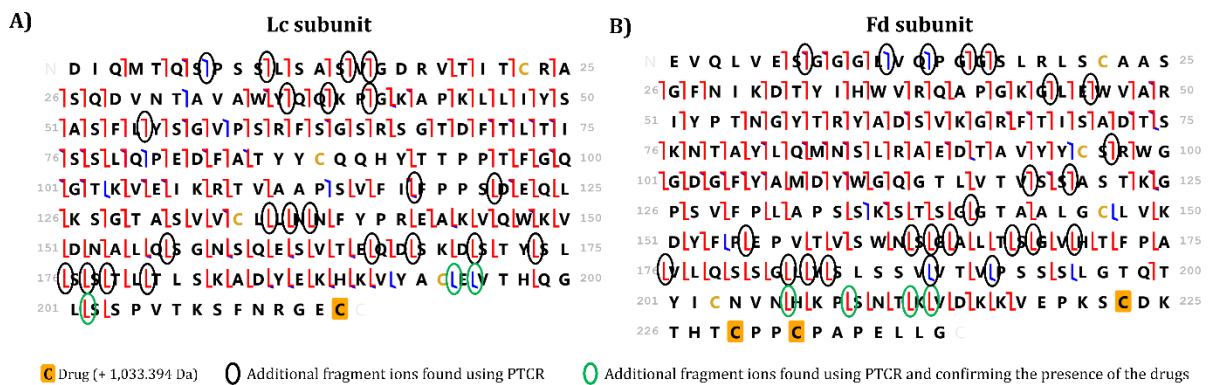


Figure 135: Fragmentation maps obtained after PTZR for Lc and Fd subunits. Fragmentation map combining ETD/EThcD MS2 and PTZR MS3 results of the (A) Lc and (B) Fd subunits. Fragments found uniquely thanks to PTZR are circled in black and those confirming the position of the drugs are circled in green.

These results highlight that PTZR improves the overall sequence coverage to a lesser extent than internal fragment ion search but enhances the number and confidence of identified diagnostic

fragment ions of the conjugation site, providing additional experimental evidence to localize the sequence modifications of interest.

6. Conclusion and perspectives

The case study depicted in this chapter fully illustrates the potential of complementary MS-based strategies to characterize trastuzumab deruxtecan antibody-drug conjugate. In the first part, SEC-nMS experiments evidenced the theoretical homogeneous avDAR 8 expected for T-DXd, pinpointing the drug conjugation on the Fab domain exclusively. These findings were corroborated by analysis of DTT reduced, IdeS digested, and DTT reduced T-DXd in denaturing conditions, elucidating the position of one and three drugs on the Lc and Fd subunits, respectively. In the second part, MD-MS experiments combining multiple techniques to fragment the Fc/2, Lc, and Fd subunits have been highly informative. They allow comprehensive sequence coverage of ADC subunits, has already experienced with trastuzumab previously. Moreover, they also demonstrate great interest for ADC characterization, since they enable the detection of several diagnostic fragments certifying precisely the conjugation site's location. Then, the attribution of internal fragments drastically improved subunit sequencing to reach a 100% sequence coverage of each subunit, as it has already been observed. In the last part, the contribution of PTCR implementation to declutter complex MS2 spectra generated after ETD or EThcD slightly increases the number of terminal fragments identified. Along with increasing the sequence coverage, it significantly improved the confidence in the location of the conjugation sites thanks to the identification of additional diagnostic fragment ions, highlighting the substantial interest of PTCR in the case of ADCs.

With this case study, we evidenced that MD-MS strategies, which have already proved their efficiency for the characterization of mAbs, are also particularly suitable in the context of ADC analysis.

In particular, cutting-edge MD-MS workflows enable to provide a complete characterization of complex therapeutic proteins such as ADCs. These strategies can also be employed for other types of biotherapeutics including Fc-fusion proteins, bispecific antibodies or therapeutic glycoproteins^{445, 640-642}. MD-MS is able to give additional information by identifying diagnostic fragment ions containing the whole payload thanks to the use of fragmentation techniques preserving labile modifications. In contrast, HCD is usually used to fragment peptides in classical bottom-up experiments and can lead to drug fragmentation, resulting in the impossibility to identify clearly the conjugation site at peptide level.

In the end, the continuous development of tailored software and the implementation of novel MS strategies in last generation instruments are expected to reinforce the use of TD/MD-MS strategies for the complete characterization of biotherapeutics, in combination to other classical methods including SEC-nMS or peptide mapping fingerprint.

The results presented in this chapter have been published this year in *Analytical and Bioanalytical Chemistry*.

GENERAL CONCLUSION

General conclusion

My Ph.D. work intended to improve the quantification of trace-level impurities in complex samples by developing analytical and bioinformatic strategies in mass spectrometry-based quantitative proteomics and to develop top-down and middle-down strategies to characterize purified therapeutic proteins comprehensively.

The first part of this manuscript consisted of a bibliographic study on the current state of mass spectrometry-based approaches for protein characterization. The classical bottom-up quantitative proteomics workflow, including the sample preparation step, the data acquisition by mass spectrometry, and the data analysis, have been described, focusing on data-independent approaches and innovative bioinformatics tools for data processing. Multiple strategies for analyzing intact proteins by mass spectrometry have been detailed, and special attention was paid to describing the top-down and middle-down approaches. Finally, monoclonal antibodies and product-related host cell protein (HCP) impurities have been introduced, and strategies available for their characterization have been addressed.

In this context, the objectives of my Ph.D. work were the following:

- The optimization of acquisition parameters on an Orbitrap tribrid instrument for quantitative bottom-up proteomics,
- The evaluation and comparison of various innovative algorithms using machine learning for data processing,
- The development of quantification strategies for trace-level HCP impurities characterization in drug products,
- The optimization of acquisition parameters for intact and subunit protein analysis by top-down and middle-down and the assessment of the dedicated data processing strategies,
- The development of analytical strategies for the comprehensive characterization of mAb-related products.

The first part of the results gathered several methodological developments for quantitative proteomics. First, acquisition parameters specific to the tribrid geometry have been optimized for both DDA and DIA acquisition modes, and their performances were evaluated using a range of perfectly calibrated samples. This comparison evidenced the interest of DIA as a global approach for quantitative proteomics, with accurate quantification of proteins in complex mixtures even at low levels. Moreover, the contribution of implementing the FAIMS ion mobility device for DDA analysis was discussed. Then, acquisition strategies dedicated to TMT samples available on tribrid instruments were assessed. In particular, quantification approaches at the MS2 or the MS3 levels were compared, showing that the real-time search for MS3 acquisition is a reasonable alternative to MS2 quantification. Second, innovative bioinformatic tools for DDA and DIA data processing were evaluated. The performance of several rescoring tools using machine-learning algorithms was examined for DDA. It showed that these tools increase identifications and boost confidence in the results obtained by considering multiple parameters linked to the LC-MS/MS experiment. In addition, two innovative search engines were assessed. Chimerys showed impressive identification numbers by matching multiple peptides on a unique spectrum, but the absence of precursor ions for some matches prevents these peptides' quantification at MS1 level. Open modification searches with Ionbot led to the identification of various, including unexpected, modifications. For DIA, the influence of search parameters on library-free approaches was discussed. At the time of the study, both software (Spectronaut and DIA-NN) showed comparable good overall performances using the default parameters. In this comparison, the advances of library-free strategies regarding identifications and quantification reproducibility were underlined.

The second part of the results was dedicated to optimizing DDA and DIA strategies in the context of HCP impurities characterization. The optimized methods developed in the first part of the results were adapted to quantify trace-level HCPs in drug products accurately. This application of proteomics methods represent a significant challenge due to the wide dynamic range in these samples. Efforts were concentrated on implementing new MS strategies to reduce the intra-sample dynamic range. The addition of FAIMS before DDA analysis and the use of gas phase fractionation DIA have significantly improved the number of quantified HCPs. Moreover, robustness and quantification accuracy were increased by the use of internal calibration curves and stringent data filtering. As a result, a few hundred of HCPs were quantified in the NIST mAb sample and tens in FDA/EMA-approved drug products, allowing MS to be considered a reliable alternative for HCP monitoring. In addition, recent advances in MS, combining ion mobility and HR/AM instruments, represent good added value in this demanding analytical context.

The third and last part of the results was dedicated to analyzing intact proteins using the top-down and middle-down strategies.

First, key parameters influencing the fragmentation, including the precursor ion charge state and the activation energy, were evaluated and optimized to maximize the sequence coverage of reference proteins, underlying their effect on the fragmentation efficiency, and hence in the final sequence coverage. The specificities of each fragmentation technique and their complementarity were shown and discussed as well. Furthermore, the potential benefits of performing an additional fragmentation step (MS3) to increase the sequence coverage were estimated. In addition, the implementation of LC prior to MS for the analysis of mAb subunits was assessed, and the need for further optimizations regarding the chromatographic method and the MS acquisition was highlighted. The reproducibility of the fragmentation was also appraised using multiple replicates. With these optimizations, high sequence coverage of each mAb subunit was achieved. Then, PTCR was implemented to increase the sequence coverage by simplifying fragmentation spectra. Substantial improvements in the sequence coverage were highlighted, but limitations regarding its implementation on a chromatographic scale and the intensity of the fragment ions after PTCR were also evidenced.

Second, data processing parameters were evaluated, including deconvolution and fragment ion matching steps. The significant influence of the deconvolution parameters on the sequence coverage was underlined, and ranges of values maximizing the sequence coverage and limiting the risk of false positives were recommended. The risk of false positives was also assessed using reversed and shuffled sequences. The P-Score was identified as the most reliable metric to estimate the confidence in ion identification in TD/MD-MS experiments. Finally, the consideration of internal fragments was investigated. Significant improvements allowing the complete sequencing of the mAb subunits were observed. However, the increased risk of misattribution and the impossibility of confidently identifying modification sites were pinpointed as the main limitations of internal fragments consideration.

Third, a combination of different MS-based approaches, i.e. intact and subunit analysis in both native, and denaturing conditions in combination with cutting-edge TD/MD-MS workflows were performed to characterize a recently approved antibody-drug conjugate (ADC), trastuzumab deruxtecan (T-DXd). The average drug-to-antibody ratio, the drug load distribution, and the number of locations of the payloads at the subunit levels were confirmed using native MS and middle-up strategies. Then, the precise payload locations at amino acid level were investigated using MD-MS methods combining complementary fragmentation techniques (CID, ETD, and ETHcD). Thus, the location of the payloads on the cysteines involved in inter-chain disulfide bonds was evidenced by several diagnostic fragment ions, confirmed by the high sequence coverage observed. Subsequently, the complete sequence coverage of each subunit was obtained by looking at internal fragments. Finally, payload locations were confirmed by additional diagnostic fragment ions identification afforded by PTCR. In this case study on trastuzumab deruxtecan, the promising potential of combining multi-level intact protein-

based approaches (i.e., native MS, middle-up, and middle-down) for comprehensive characterization of an ADC has been established.

The work described in this manuscript highlights the importance of analytical developments in mass spectrometry, from data acquisition to data processing. The contribution of mass spectrometry to the characterization of therapeutic proteins and their impurities has been discussed throughout the manuscript, and some potential perspectives in this field have been suggested. To conclude, I would like to emphasize some points.

This manuscript has used two main approaches, bottom-up and top-down, facing comparable challenges.

For bottom-up proteomics, continuous improvements in acquisition speed, sensitivity, and sensibility drive DIA's democratization. This approach is now becoming more and more popular for various proteomic applications. It brings together the advantages of global and targeted methods in a single analysis, allowing for reproducible and accurate quantification. Introducing new strategies combining ion mobility and DIA boosts the number of identified and quantified features and should undoubtedly contribute to the widespread use of this approach in proteomics studies. In addition to data acquisition, data processing has seen significant advances in recent years.

The advent of deep learning and artificial intelligence significantly improved the data analysis step, one of the bottlenecks of the bottom-up proteomic workflow. Promising performances of these bioinformatics tools have been demonstrated in this manuscript. Predictions of peptide properties, such as retention time or theoretical fragmentation spectra, from the vast amount of data available in public repositories, bring additional features that increase identifications and overall improve confidence. These prediction tools are particularly interesting for DIA, allowing the generation of extensive predicted libraries without the need for time-consuming sample preparation and MS analysis. While I am convinced that the work from people in informatics or bioinformatics will make the current data processing practices evolve, I still believe that biologists or mass spectrometry people require a thorough comprehension of these tools. Indeed, these algorithms are not error-free, and experts should keep a critical eye on the field to evaluate the results' relevance.

For top-down analysis of purified proteins, fragmentation techniques providing extensive fragmentation of the protein backbone and high-resolution accurate mass instruments allowing the detection of hundreds of fragment ions in each spectrum are required. The commercial release of instruments enabling multiple fragmentation techniques or the development of devices to upgrade existing instruments allows the combination of complementary techniques to obtain significant sequence coverages. With these improvements, and regarding the considerable efforts made by the community to ensure the spread of this approach, top-down strategies are becoming more accessible to people out of the field. In addition, the need for intact proteoforms identification and quantification in understanding biological systems would open the door to widen use of top-down in the coming years. Nevertheless, additional work is still necessary for the development of multi-stage fragmentations (e.g., MS3, MS4), that would improve both the sequence coverage and the confidence in the identified fragments.

A last point I would like to discuss concerns the lack of guidelines for top-down data processing. The vast amount of data generated by the fragmentation of intact protein results in complex MS/MS spectra deconvoluted to obtain a list of fragments that will be matched to the theoretical sequence. However, key steps of the process, including deconvolution and fragment matching, suffer from a total absence of standardization, making the comparison between studies difficult. Even if data processing parameters should be adapted to the experimental conditions, using too relaxed parameters can sometimes raise questions about the proportion of misattributions. A compromise between too-stringent criteria filtering out half of the identifications and too-relaxed criteria leading to a substantial proportion of false positives should be found. I think that the community should propose

standardization initiatives, or at least general guidelines, about the key parameters directly affecting the results. It would undoubtedly improve the reliability of the results provided and could benefit beginners in the field by helping them set up their first experiments.

EXPERIMENTAL SECTION

Experimental section

Unless otherwise indicated, all products were purchased from Sigma Aldrich (St. Louis, MO, USA). Solvents (acetonitrile, methanol) and acids (formic acid (FA), trifluoroacetic acid (TFA)) were supplied by Fisher Scientific (Hampton, NH, USA).

1. Methodological development in quantitative proteomics

All the LC-MS/MS analyses were performed on a Dionex UltiMate 3000 RSLC nano system (Thermo Fisher Scientific) coupled to an Orbitrap Eclipse™ Tribrid™ (Thermo Fisher Scientific) mass spectrometer,

A. Optimization of LC-MS/MS acquisition methods

i) Evaluation of the acquisition parameters

a) Comparison of Orbitrap and ion trap analyzers

Sample preparation

Commercially available Pierce™ HeLa Protein Digest Standard was obtained from Thermo Fisher Scientific. A 20 µg aliquot was resuspended with 200 µL of 0.1% FA in water to obtain a final concentration of 100 ng/µL. A total of 1 µL (100 ng) was injected in triplicates for each method.

LC-MS/MS analysis

Data-dependent acquisitions (DDA) were performed on a Dionex UltiMate 3000 RSLC nano system (Thermo Fisher Scientific) coupled to an Orbitrap Eclipse™ Tribrid™ (Thermo Fisher Scientific) mass spectrometer, equipped with a FAIMS Pro interface (Thermo Fisher Scientific). Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in ACN. Peptides were loaded onto an Acclaim PepMap 100 C18 20 mm x 0.1 mm, 5 µm diameter particles precolumn (Thermo Fisher Scientific) for 3 min at 10 µL/min, and eluted on an Aurora Series C18 UHPLC (250 mm x 75µm, 1.6 µm diameter particles, IonOpticks) at a 300 nL/min flow rate following a linear gradient: 2% B during 3 min, 25% B at 53 min, 40% B at 63 min and 90% B from 65 to 70 min. The column was finally re-equilibrated in 2% B for 19 min.

Full-scan MS spectra were acquired over a 375-1,500 m/z range at a resolution of 120,000 (at 200 m/z), with an automatic gain control (AGC) target of 4.105 and a maximum injection time set to Automatic. Top20 precursor ions with an intensity exceeding 5×10^3 ions per second and charge states between 2 and 7 were automatically selected from each MS spectrum for higher-energy collisional dissociation (HCD) fragmentation at 30% normalized collision energy. MS/MS spectra were collected either in the Orbitrap or the ion trap, depending on the experiment. For Orbitrap MS/MS detection, fragmentation spectra were collected at 15,000 resolution (at 200 m/z) using an automatic scan range, with an AGC target of 5×10^4 and a maximum injection time set to 22 ms. For ion trap MS/MS detection, fragmentation spectra were collected from in Rapid mode using an automatic scan range, with an AGC target of 1×10^4 and a maximum injection time set to 35 ms.

Data analysis

DDA data analysis was performed with Proteome Discoverer v. 2.5 (Thermo Fisher Scientific) using Sequest HT (Thermo Fisher Scientific) against *Homo sapiens* entries (20,324 entries, 2021/06/03) extracted from SwissProt and common contaminants were included. Trypsin (full) was used as

digestion enzyme and one missed cleavage was allowed. Cysteine carbamidomethylation was set as fixed modification; protein N-term acetylation and methionine oxidation were set as variable modifications. Mass tolerance was set to 5 ppm for precursor ion masses and 0.5 Da or 0.02 Da for fragment ion masses analyzed in the ion trap or the Orbitrap, respectively. The false discovery rate (FDR) was calculated using the Percolator node set to 1% at PSM, peptide and protein levels. XIC-MS1 quantification was performed using unique peptides only and the chromatographic alignment was set to 10 min and 10 ppm between the 3 technical replicates.

b) Comparison of HCD, CID, ETD and EThcD fragmentation techniques

Sample preparation

The same samples as those described in **Experimental Section 1.A.1.i.** were used.

LC-MS/MS analysis

LC-MS/MS analyses were performed using the setup described in **Experimental Section 1.A.1.i.** with MS/MS detection performed in the ion trap. Full-scan MS spectra were acquired over a 300-1,800 m/z range at a resolution of 120,000 (at 200 m/z), with an automatic gain control (AGC) target of 4×10^5 and a maximum injection time set to Automatic. Precursors ions with an intensity exceeding 5×10^3 ions per second and charge states between 2 and 7 were automatically selected from each MS spectrum using a 1.6 m/z isolation window for fragmentation and the global cycle time was set to 3 s. MS/MS spectra were collected either in the ion trap in Rapid mode with a maximum injection time of 35 ms and AGC target of 5×10^4 . For ETD and EThcD fragmentation, ETD target was set to 1×10^7 with a maximum ETD reagent injection time of 200 ms.

Fragmentation methods used and their corresponding energies are described in **Table 13**.

Table 13: Activation energy used for each method.

Fragmentation technique	HCD	CID	ETD	EThcD
Activation time/energy	30% NCE	35%	50 ms	50 ms + 35% NCE

Data analysis

The same data analysis parameters as those described in **Experimental Section 1.A.1.i.** were used.

c) Optimization of the isolation windows scheme for DIA

Sample preparation

The same samples as those described in **Experimental Section 1.A.1.i.** were used.

LC-MS/MS analyses

LC-MS/MS analyses were performed using the setup described in **Experimental Section 1.A.1.i.** Chromatographic conditions are also the same as already described in **Experimental Section 1.A.1.i.** Full-scan MS were acquired over a 375-1,050 m/z range at a resolution of 120,000 (at 200 m/z) with an AGC target of 1×10^6 and a maximum injection time of 50 ms. Fragment analysis (MS/MS) was subdivided into numerous windows, described in **Figure 136**. A resolution of 15,000 (at 200 m/z), an AGC target of 1×10^6 and a maximum injection time of 22 ms were use. HCD fragmentation normalized collision energy was set to 30%.

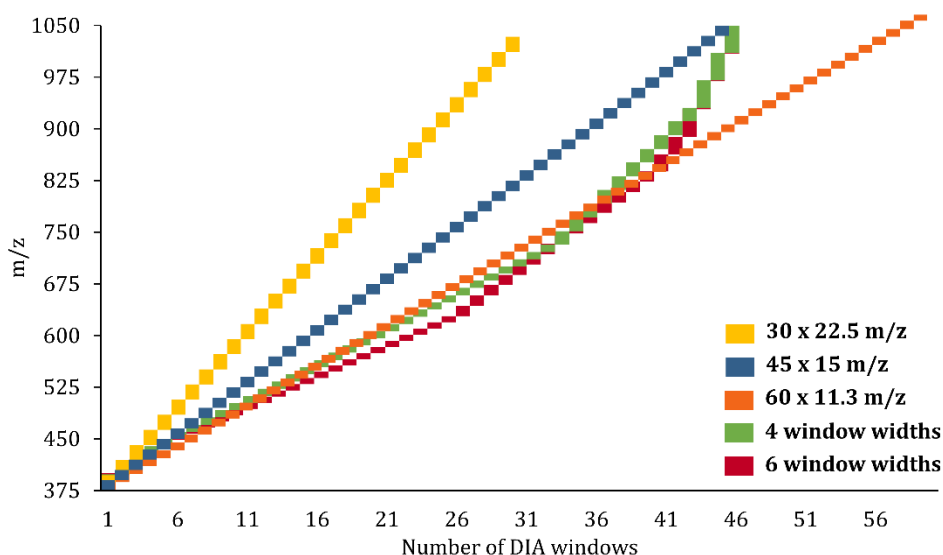


Figure 136: DIA isolation schemes used for optimization.

Data analysis

DIA data analysis was performed with Spectronaut v. 17.5 (Biognosys) using the *Homo sapiens* fasta file (20,324 entries, 2021/06/03) and the following settings. Briefly, trypsin/P was set as digestion enzyme, cysteine carbamidomethylation was set as fixed modification, methionine oxidation and protein N-term acetylation were set as variable modifications and one missed cleavage was allowed. Other parameters were set as default. Identification was performed using 0.01 precursor and protein Q-value cutoffs. Quantification was performed using the default parameters.

d) Application to a mixture of well calibrated samples

Sample preparation

Arabidopsis thaliana cells were solubilized in 500 μL of a 8 M urea and 0.1 M NH_4HCO_3 solution and sonicated for 5 min in an ice bath. The protein concentration of the solution was measured using the RC-DC protein assay, and was adjusted to 0.6 $\mu\text{g}/\mu\text{L}$ with a 25 mM Tris-HCl (pH 7) buffer. UPS1 protein range was prepared to obtain final concentration of 0.05, 0.25, 0.50, 1.25, 2.50, 5 and 10 fmol/ μL in 0.2 $\mu\text{g}/\mu\text{L}$ *Arabidopsis thaliana*.

Protein mixture was reduced with TCEP at a final concentration of 2 mM for 30 min at 37°C and alkylated with 10 μL of 100 mM IAM for 1 h at 25°C in the dark. Reduced and alkylated proteins were digested using trypsin/Lys-C (Promega) using a 1:10 enzyme:protein ratio for 1 h at 70°C. Digestion was quenched with 2 μL of FA. Peptides were purified by solid-phase extraction (SPE), dried vacuum and solubilized in $\text{H}_2\text{O}/\text{ACN}$ 98/2 with 0.1% FA to a final concentration of 200 $\mu\text{g}/\mu\text{L}$ and iRT peptides (Biognosys) were added to the mixture.

LC-MS/MS analysis

Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in ACN. Peptides were loaded onto an Acclaim PepMap 100 C18 20 mm x 0.1 mm, 5 μm diameter particles precolumn (Thermo Fisher Scientific) for 3 min at 10 $\mu\text{L}/\text{min}$, and eluted on an Aurora Series C18 UHPLC (250 mm x 75 μm , 1.6 μm diameter particles, IonOpticks) at a 300 nL/min flow rate following a linear gradient: 2% B during 3 min, 20% B at 63 min, 40% B at 82 min and 90% B from 83 to 85 min. The column was finally re-equilibrated in 2% B for 13 min.

For DDA acquisition, full-scan MS spectra were acquired over a 375-1,500 m/z range at a resolution of 120,000 (at 200 m/z), with an AGC target of 4×10^5 and a maximum injection time set to Automatic. Precursor ions with an intensity exceeding 5×10^3 ions per second and charge states between 2 and 6 were automatically selected from each MS spectrum for HCD fragmentation at 30% NCE. MS/MS spectra were collected in the ion trap in Rapid mode for a total cycle time of 2 s. MS/MS fragmentation spectra were collected over an automatic scan range, with an AGC target of 1×10^4 and a maximum injection time set to 35 ms.

For DIA acquisition, full-scan MS were acquired over a 375-1,250 m/z range at a resolution of 120,000 (at 200 m/z) with an AGC target of 4×10^5 and a maximum injection time of 50 ms. Fragment analysis (MS/MS) was subdivided into numerous variable width windows, described in **Figure 137**. MS/MS spectra were collected in the Orbitrap at a resolution of 30,000 (at 200 m/z), an AGC target of 5×10^5 and a maximum injection time of 54 ms were used. HCD fragmentation normalized collision energy was set to 30%.

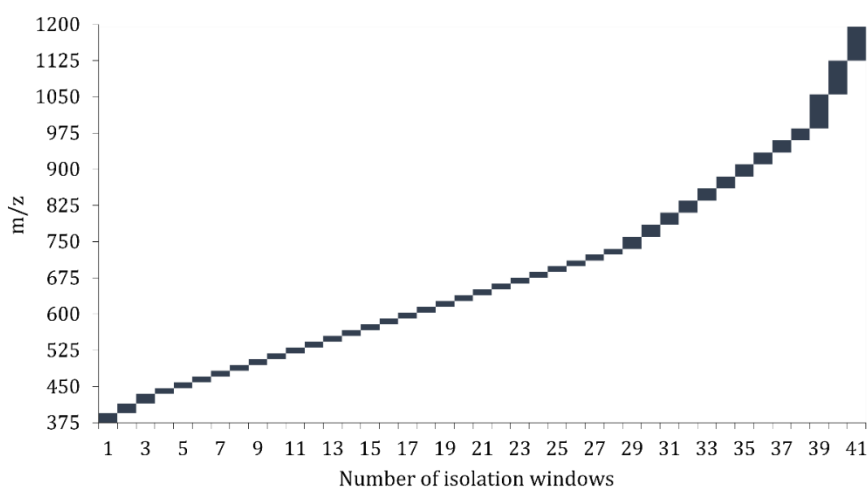


Figure 137: DIA isolation scheme used for well-calibrated samples analysis

Data analysis

DDA data analysis was performed using parameters described in **Experimental Section 1.A.1.i.** and DIA data analysis using Spectronaut (v. 15.7) with the parameters described in **Experimental Section 1.A.1.iii.** Searches were performed against a fasta file combining *Arabidopsis thaliana* (SwissProt, 16,121 entries, 2021/03/01), UPS1 proteins and common contaminants entries. Differential analysis was performed using ProStar v.1.28.0.

ii) Optimization of FAIMS compensation voltages

a) Sample preparation

The same samples as those described in **Experimental Section 1.A.1.i.** were used.

b) LC-MS/MS analyses

The same LC-MS/MS setup and parameters as those described in **Experimental Section 1.A.1.i.** were used. Experiments with the FAIMS Pro interface were conducted as described in Hebert *et al.*⁷⁷. Briefly, FAIMS carrier gas flow was 4.7 L/min N₂, the FAIMS electrodes were set to 100°C, entrance plate voltage was 250 V, asymmetric waveform divert voltage was -5000 V, compensation voltage (CV) settling time was 25 ms and CVs values were applied successively in the same run. Two CVs were used for Orbitrap experiments (-45V|-60V and -50V|-70V) and three for ion trap ones (-45V|-60V|-80V and -50V|-65V|-85V).

c) Data analysis

Data analysis was performed as described in **Experimental Section 1.A.1.i**.

iii) Evaluation of acquisition strategies dedicated to TMT samples

a) Sample preparation

HeLa cells from human were solubilized in 800 μL of a 8 M urea and 0.1 M NH_4HCO_3 solution and sonicated for 10 min in an ice bath. The protein concentration of the solution was measured using the RC-DC protein assay. Protein mixture was reduced with DTT at a final concentration of 5 mM for 30 min at 37°C and alkylated with 15 mM (final concentration) of IAM for 1 h at 25°C in the dark. Reduced and alkylated proteins were diluted to 1 M urea with 100 mM NH_4HCO_3 and digested using trypsin/Lys-C (Promega) using a 1:25 enzyme:protein ratio at 37°C overnight. Digestion was quenched with 2 μL of FA and peptides were purified by solid-phase extraction (SPE) using a C18 cartridge (Sep-Pack).

Peptides were diluted with 50 μL of 100 mM 3-[4-(2-Hydroxyethyl)-1-piperazine]propane sulfonic acid (EPPS) to obtain a pH of 8.5. TMT0 reagent was equilibrated at room temperature and dissolved in 41 μL of ACN. Then, 4 μL of TMT0 reagent were added to the sample for a protein:TMT ratio of 1:8 and were incubated for 1 h at room temperature. The labeling reaction was quenched by the addition of 3 μL of 5% hydroxylamine. Samples were incubated during 15 min and were diluted to 1 $\mu\text{g}/\mu\text{L}$ with 100 mM EPPS. Labeled peptides were purified by PSE using RPS cartridges, vacuum dried and resuspended in $\text{H}_2\text{O}/\text{ACN}$ 98/2 with 0.1% FA to a final concentration of 400 $\mu\text{g}/\mu\text{L}$. The same protocol was used for TMTpro -16plex.

b) LC-MS/MS analysis for collision energy evaluation

Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in 80% ACN. Peptides were loaded onto an Acclaim PepMap 100 C18 20 mm x 0.1 mm, 5 μm diameter particles precolumn (Thermo Fisher Scientific) for 3 min at 10 $\mu\text{L}/\text{min}$, and eluted on an Aurora Series C18 UHPLC (250 mm x 75 μm , 1.6 μm diameter particles, IonOpticks) at a 300 nL/min flow rate following a linear gradient: 2.5% B during 3 min, 31.2% B at 53 min, 50% B at 63 min and 98% B from 65 to 70 min. The column was finally re-equilibrated in 2.5% B for 19 min.

Full-scan MS spectra were acquired over a 400-1,600 m/z range at a resolution of 120,000 (at 200 m/z), with an AGC target of 4×10^5 and a maximum injection time set to Automatic. Precursor ions with an intensity exceeding 2.5×10^4 ions per second and charge states between 2 and 6 were automatically selected from each MS spectrum for HCD fragmentation using 27%-40% NCE. Dynamic exclusion was set to 60 s. MS/MS spectra were collected in the Orbitrap for a total cycle time of 3 s. MS/MS fragmentation spectra were collected over an automatic scan range at a resolution of 30,000 (at 200 m/z), with an AGC target of 1.25×10^5 and a maximum injection time set to 54 ms.

c) LC-MS/MS analysis for TMT acquisition method evaluation

Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in 80% ACN. Peptides were loaded onto an Acclaim PepMap 100 C18 20 mm x 0.1 mm, 5 μm diameter particles precolumn (Thermo Fisher Scientific) for 3 min at 10 $\mu\text{L}/\text{min}$, and eluted on an Aurora Series C18 UHPLC (250 mm x 75 μm , 1.6 μm diameter particles, IonOpticks) at a 300 nL/min flow rate following a linear gradient: 2.5% B during 3 min, 31.2% B at 163 min, 50% B at 183 min and 98% B from 185 to 190 min. The column was finally re-equilibrated in 2.5% B for 19 min.

Full-scan MS spectra were acquired over a 400-1,600 m/z range at a resolution of 120,000 (at 200 m/z), with an AGC target of 4×10^5 and a maximum injection time set to Automatic. For TMT MS2 experiments, precursor ions with an intensity exceeding 5×10^4 ions per second and charge states

between 2 and 6 were automatically selected using 0.7 m/z isolation window from each MS spectrum for HCD fragmentation at 35% NCE. Dynamic exclusion was set to 60 s. MS/MS spectra were collected in the Orbitrap at a resolution of 50,000 (at 200 m/z) for a total cycle time of 3 s. Scan range was set to automatic, AGC target to 1.25×10^5 and a maximum injection time set to 120 ms. For TMT MS3 experiments, precursor ions with an intensity exceeding 5×10^3 ions per second and charge states between 2 and 6 were automatically selected using 0.7 m/z isolation window from each MS spectrum for CID fragmentation at 35%. MS/MS spectra were collected on the ion trap in Rapid mode, using an AGC target of 1×10^4 and a maximum injection time of 50 ms. Then, the ten most abundant precursors were selected using a 2 m/z isolation window for HCD fragmentation at 55% NCE. MS3 spectra were recorded in the Orbitrap at a resolution of 50,000 K (at 200 m/z) using a 110-500 m/z scan range. AGC target was set to 1.2×10^5 and maximum injection time to 120 ms. For MS3 RTS, the same parameters were used for data acquisition. In addition, real-time search was performed using the Comet search engines with the following parameters: trypsin (full) was used as enzyme with one maximum missed cleavage, carbamidomethylation of cysteine and TMTpro (K and N-term) were set as fixed modifications, methionine oxidation as variable modification and a tolerance of 25 ppm was used for precursor matching. Only precursors that have been identified were submitted to MS3 analysis.

d) Data processing

Data analysis was performed with Proteome Discoverer v. 2.5 (Thermo Fisher Scientific) using Sequest HT (Thermo Fisher Scientific) against *Homo sapiens* entries (20,324 entries, 2021/06/03) extracted from SwissProt and common contaminants were included. Trypsin (full) was used as digestion enzyme and one missed cleavage was allowed. Cysteine carbamidomethylation and TMT (or TMTpro) on lysine and peptide N-term were set as fixed modifications; methionine oxidation was set as variable modification. Mass tolerance was set to 5 ppm for precursor ion masses and 0.02 Da for fragment ion masses (0.5 Da for TMT MS3 experiments). The false discovery rate (FDR) was calculated using the Percolator node set to 1% at PSM, peptide and protein levels. Quantification was performed using the Reporter ions quantifier node, using unique peptides only. Co-isolation threshold of precursors was set to 50% and no filter on the minimum channel occupancy were set.

B. Evaluation of bioinformatics tools for data processing

i) Evaluation of innovative tools for DDA data processing

a) MS²Rescore

Sample preparation

Commercially available Pierce™ HeLa Protein Digest Standard was obtained from Thermo Fisher Scientific. A 20 µg aliquot was resuspended with 200 µL of 0.1% FA in water to obtain a final concentration of 100 ng/µL. For each of the five replicates, 200 ng were injected.

LC-MS/MS analysis

Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in 80% ACN. Peptides were directly eluted on an Aurora Series C18 UHPLC (250 mm x 75µm, 1.6 µm diameter particles, IonOpticks) at a 300 nL/min flow rate following a linear gradient: 2.5% B during 3 min, 31.2% B at 53 min, 50% B at 63 min and 98% B from 65 to 70 min. The column was finally re-equilibrated in 2.5% B for 19 min.

Full-scan MS spectra were acquired over a 300-1,800 m/z range at a resolution of 60,000 (at 200 m/z), with an AGC target of 1×10^6 and a maximum injection time set to 50 ms. The 20 most intense precursor ions with an intensity exceeding 5×10^3 ions per second and charge states between 2 and 7 were automatically selected using 1.8 m/z isolation window for HCD fragmentation at 30% NCE and

Dynamic exclusion was set to 60 s. MS/MS spectra were collected in the Orbitrap at a resolution of 15,000 (at 200 m/z). MS/MS fragmentation spectra were collected over an automatic scan range at a resolution of 30,000 (at 200 m/z), with an AGC target of 1.25×10^5 and a maximum injection time set to 22 ms.

Data analysis

Each replicate was analyzed separately. Raw files were converted to MGF files using MS Convert (v. 3.0). Data were analyzed using Mascot (Matrix Science) and a fasta file containing *Homo sapiens* entries (SwissProt, 20,324 entries, 2021/06/03) and common contaminants. Trypsin/P was used as the digestion enzyme and one missed cleavage was allowed. Oxidation of methionine was set as variable modification and cysteine carbamidomethylation as fixed modification. Data were extracted using peptide mass tolerance of 5 ppm and MS/MS mass tolerance at 0.02 Da. A FDR of 1% was set at the PSM level. For MS²Rescore, data obtained after Mascot search at 100% FDR were used as input, and the same modifications and mass tolerances as for the initial Mascot search were used. HCD2021 model was selected for spectrum prediction and other parameters were set as default. PSMs with q-value > 0.01 were filtered out.

b) Inferys

Sample preparation and LC-MS/MS analysis

The same samples and LC-MS/MS acquisition methods as those described in **Experimental Section 1.B.i.a.** were used.

Data analysis

Each replicate was analyzed separately. Data analysis was performed with Proteome Discoverer v. 2.5 (Thermo Fisher Scientific) using Sequest HT (Thermo Fisher Scientific) against *Homo sapiens* entries (20,324 entries, 2021/06/03) extracted from SwissProt and common contaminants were included. Trypsin (full) was used as digestion enzyme and one missed cleavage was allowed. Cysteine carbamidomethylation was set as fixed modification; and methionine oxidation was set as variable modification. Mass tolerance was set to 5 ppm for precursor ion masses and 0.02 Da for fragment ion masses. The false discovery rate (FDR) was calculated using the Percolator node set to 1% at PSM level. For Inferys search, an additional node (Inferys node) was added before Percolator node and the collision energy value was set to 30% for spectra prediction.

c) Chimerys

Sample preparation and LC-MS/MS analysis

The same samples and LC-MS/MS acquisition methods as those described in **Experimental Section 1.B.i.a.** were used.

Data analysis

Each replicate was analyzed separately. Data analysis was performed with Proteome Discoverer v. 3.0 (Thermo Fisher Scientific) using Chimerys (MSAID) against *Homo sapiens* entries (20,324 entries, 2021/06/03) extracted from SwissProt and common contaminants were included. Chimerys used the Inferys_2.1_fragmentation model for spectra prediction. Trypsin (full) was used as digestion enzyme and one missed cleavage was allowed. Cysteine carbamidomethylation was set as fixed modification; and methionine oxidation was set as variable modification. Mass tolerance was set to 20 ppm for fragment ion masses. The false discovery rate (FDR) was calculated using the Percolator node set to 1% at PSM level.

d) IonbotSample preparation and LC-MS/MS analysis

The same samples and LC-MS/MS acquisition methods as those described in **Experimental Section 1.B.i.a.** were used.

Data analysis

Each replicate was analyzed separately. Data analysis was performed with ionbot v.0.10.0 (CompOmics, Ghent, Belgium) against *Homo sapiens* entries (20,324 entries, 2021/06/03) extracted from SwissProt and common contaminants were included. Trypsin was used as digestion enzyme and one missed cleavage was allowed. Cysteine carbamidomethylation was set as fixed modification, methionine oxidation as variable modification and open modification search was enabled. Mass tolerance was set to 20 ppm for precursor ion and 0.02 Da for fragment ion masses. Output results were filter using q-value < 0.01.

*ii) Comparison of library-free approaches in DIA***a) Spectronaut**Sample preparation

The same samples as those described in **Experimental Section 1.B.i.a.** were used.

LC-MS/MS analysis

The same LC-MS/MS acquisition parameters as those described in **Experimental section 1.A.i.d** were used.

Data analysis

DIA data analysis was performed with Spectronaut v. 17.5 (Biognosys) against a fasta file combining *Arabidopsis thaliana* proteins (SwissProt, 16,121 entries, 2021/03/01), UPS1 proteins and common contaminants entries. Briefly, trypsin/P was set as digestion enzyme, cysteine carbamidomethylation was set as fixed modification, methionine oxidation was set as variable modification and one missed cleavage was allowed. Parameters studied during this evaluation (i.e., MS/MS mass tolerance and retention time extraction window) were set as described in the main text. Other parameters were set as default. Identification was performed using 0.01 precursor and protein Q-value cutoffs. Quantification was performed using the default parameters. Differential analysis was performed using ProStar v. 1.28.0.

b) DIA-NNSample preparation

The same samples as those described in **Experimental Section 1.B.i.a.** were used.

LC-MS/MS analysis

The same LC-MS/MS acquisition parameters as those described in **Experimental section 1.A.i.d.** were used.

Data analysis

DIA data analysis was performed with DIA-NN v.1.8.1 against a fasta file combining *Arabidopsis thaliana* proteins (SwissProt, 16,121 entries, 2021/03/01), UPS1 proteins and common contaminants entries. Briefly, trypsin was set as digestion enzyme, cysteine carbamidomethylation was set as fixed

modification, methionine oxidation was set as variable modification and one missed cleavage was allowed. Parameters studied during this evaluation (i.e., MS/MS mass tolerance and retention time extraction window) were set as described in the main text. Other parameters were set as default. Identification was performed using 0.01 precursor and protein Q-value cutoffs. Quantification was performed using the default parameters. Differential analysis was performed using ProStar v. 1.28.0.

2. Methodological developments for HCP quantification in drug products

A. HCP-PROFILER for internal calibration

The list of peptides used for the building of the calibration curves is provided in **Table 14**. For each quantity, three proteins with three peptides each are used, from 1 fmol to 500 fmol.

Table 14: List of HCP-PROFILER (Anaquant, Lyon) peptides and proteins.

Quantity	Standard	Peptide	Quantity	Standard	Peptide
1 fmol	ANAQUANT-1	DGALLENVTVR EGAPEIYNAIR GDVAVFFGLSGTGGK	50 fmol	ANAQUANT-10	GGLTDAAQVVAAVEGK LGGADGNALFR VGLEVTLR
	ANAQUANT-2	EGCDLAGAIK VGFEGNRPTNSILLR VGNPETTLFLVASK		ANAQUANT-11	AGHPQLAEFTR TGVIGFGSPNK YGINELQANPAK
	ANAQUANT-3	LGAADVTVGPTLLVAR VGLVPTQEAIQK VGQQPEFAAAK		ANAQUANT-12	GGTLGQDVIDIR IGTFIDGDEGILLHR YGSIGQPFVYPR
2.5 fmol	ANAQUANT-4	AGVVEELAR VGQLLGSGSILR VLGTDGFGGR	250 fmol	ANAQUANT-13	GGPLTTPVGGGIR TGVTYDFER YGYQGTSPVK
	ANAQUANT-5	AGFDFACLPNEGVLAR AGLLEFDDQEPQLQNEIR GGVALSAGVQR		ANAQUANT-14	AGLQAIAGPFSQVR FGCPTGGISPANYR TGSAESILTTGPVVPVIVVK
	ANAQUANT-6	GGSGPYFYLPK VGIASELGEER VGIDGQINLR		ANAQUANT-15	FGFSQPLLLGK LFGVTTLDIIR SGDLFNVNAGIVK
10 fmol	ANAQUANT-7	AGLAEHGIVFGPEK EGVRPDAIICTGR VGALLSHSNFGSSDCPSSSK	500 fmol	ANAQUANT-16	AGDAFAVIVK EGQGLTPVLCIGETEAEENEAGK IGYQLKPNPAVLICR
	ANAQUANT-8	CGADLGLETVIVER FGTGANTLEVEGENGK TGQVVVLGAGPAGYSAFR		ANAQUANT-17	EGSGLLGLTEVTSDCR LGLVVLNCGSSSLK YGTSSVVIDESVIQGIK
	ANAQUANT-9	TGYSGLDYPSEAVIR VGLSGPGLVNLIR YGALVGDVGGTNR		ANAQUANT-18	EGLPLTESLALTIDR VGIPYWNTELIPR YGYLGNADIEIAAK

B. Evaluation of FAIMS DDA methods

i) Sample preparation

The NIST mAb standard reference material RM8671 was purchased from Merck. HCP-PROFILER standard peptides were acquired from Anaquant (Villeurbanne, France). Mass spectrometry grade trypsin/Lys-C enzymes were obtained from Promega (Madison, WI, USA).

The digestion protocol was adapted from Huang *et al.*⁵⁰⁹. Briefly, a 100 μ L aliquot of 10 μ g/ μ L mAb was supplemented with 90 μ L of ultrapure H₂O and 5 μ L of 1 M Tris-HCl, pH 8. Proteins were digested overnight at 37°C using a solution of Trypsin/Lys C enzymes at a 1:400 enzyme/protein ratio. Digests were reduced with 1.5 μ L of 303 mM DTT for 10 min at 90°C and centrifuged at 13 000g for 2 min. Supernatant was finally acidified with 0.5 μ L of FA. Then, after vacuum drying, peptides were solubilized in 2% ACN and 0.1% FA to obtain a final protein concentration of 1 μ g/ μ L. One HCP-

PROFILER bead was dissolved in 150 μL of 5% ACN and 0.1% FA, to obtain standard peptides in a concentration range from 0.5 to 250 fmol/ μL . Retention time iRT standards (Biognosys, Schlieren, Switzerland) and 2 μL of HCP-PROFILER solution were mixed to 1 or 2 μL of NIST mAb Reference Material, respectively, before injection.

ii) LC-MS/MS analysis

DDA analyses were performed on a Dionex UltiMate 3000 RSLC nano system (Thermo Fisher Scientific) coupled to an Orbitrap Eclipse™ Tribrid™ (Thermo Fisher Scientific) mass spectrometer, equipped with a FAIMS Pro interface (Thermo Fisher Scientific). Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in 80% ACN/20% water. Peptides were loaded onto an Acclaim PepMap 100 C18 20 mm x 0.1 mm, 5 μm diameter particles precolumn (Thermo Fisher Scientific) for 3 min at 10 $\mu\text{L}/\text{min}$, and eluted on an Aurora Series C18 UHPLC (250 mm x 75 μm , 1.6 μm diameter particles, IonOpticks) at a 300 nL/min flow rate following a linear gradient: 2.5% B at 0 min, 43.8% B at 95 min and 98% B from 96 to 101 min. The column was finally re-equilibrated in 2.5% B for 15 min.

Full-scan MS spectra were acquired over a 375-1,500 m/z range at a resolution of 60,000 (at 200 m/z), with an automatic gain control (AGC) target of 1×10^6 and a maximum injection time of 60 ms. The top 10 most intense precursor ions with an intensity exceeding 5×10^3 ions per second and charge states between 2 and 6 were automatically selected from each MS spectrum for higher-energy collisional dissociation (HCD) fragmentation at 30% normalized collision energy. MS/MS spectra were collected from 120-1200 m/z either in the ion trap or in the Orbitrap. For ion trap detection, Rapid mode was used, with an AGC target of 1×10^4 and a maximum injection time set to Automatic. For Orbitrap detection, resolution was set to 15,000 (at 200 m/z), with an AGC target of 1×10^5 and a maximum injection time set to 22 ms. Experiments with the FAIMS Pro interface were conducted as described in Hebert *et al.*⁷⁷. Briefly, FAIMS carrier gas flow was 4.7 L/min N_2 , the FAIMS electrodes were set to 100°C, entrance plate voltage was 250 V, asymmetric waveform divert voltage was -5000 V, compensation voltage (CV) settling time was 25 ms and CVs values of -50, -65 and -85 V were applied successively in the same run for ion trap detection; -50 V and -70 V for Orbitrap detection.

iii) Data analysis

DDA data analysis was performed with Proteome Discoverer v. 2.5 (Thermo Fisher Scientific) using Sequest HT (Thermo Fisher Scientific) against a database containing all *Mus musculus* entries (17,050 entries, TaxID=10090, 2021/05/26) extracted from UniProtKB/SwissProt for NIST mAb samples. Each fasta file included common contaminants, iRT retention time standards, HCP-PROFILER standard sequences and mAbs heavy and light chain sequences. Trypsin/P was used as digestion enzyme and one missed cleavage was allowed. Methionine oxidation and protein N-term acetylation were set as variable modifications. Mass tolerance was set to 5 ppm for precursor ion masses and 0.5 Da for fragment ion masses analyzed in the ion trap 0.02 for analysis in the Orbitrap. The FDR was calculated using the Percolator node set to 1% at PSM level. XIC-MS1 quantification was performed using unique peptides only and the chromatographic alignment was set to 10 min and 10 ppm between the 3 technical replicates

C. Evaluation of DIA strategies for HCP quantification

i) Sample preparation

The same samples as those described in **Experimental Section 1.B.i.a.** were used. For gas-phase fractionation (GPF) chromatogram library generation, technical replicates were pooled together.

ii) LC-MS/MS analysis

Same chromatographic conditions as those described in **Experimental section 2.A.ii.** were used. For DIA analyses, full-scan MS was acquired over a 375-1250 m/z range at a resolution of 60,000 (at 200 m/z) with an AGC target of 1×10^6 and a maximum injection time of 60 ms. Fragment analysis (MS/MS) was subdivided into different isolation schemes, described in **Figure 138**. A resolution of 15,000 (at 200 m/z), an AGC target of 1×10^6 and a maximum injection time of 22 ms were used for a total cycle time of 1.5 s. HCD fragmentation normalized collision energy was set to 30%.

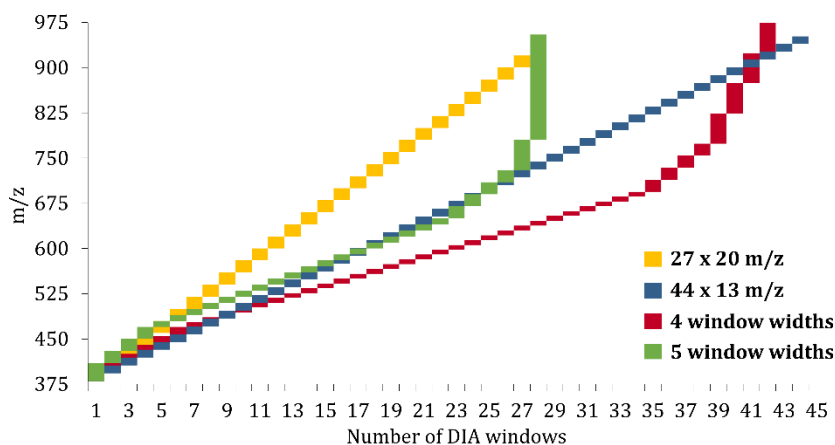


Figure 138: DIA isolation schemes used for HCP analysis.

For DIA-GPF, 6 individual injections were performed, covering the 380-980 m/z mass range. A 100 m/z full-scan MS was acquired with the parameters described above for DIA, and the MS/MS analysis was divided into 25 windows of 4 m/z using a 30,000 (at 200 m/z) resolution, an AGC target of 1×10^6 and maximum injection time of 60 ms.

iii) Data analysis

DIA data analysis was performed with Spectronaut v. 15.7 (Biognosys) using the fasta files described above and the following settings. Briefly, trypsin/P was set as digestion enzyme, methionine oxidation and protein N-term acetylation were set as variable modifications and one missed cleavage was allowed. Data was extracted using dynamic mass tolerances and a 10 min XIC RT extraction window. Identification was performed using 0.01 precursor q-value cutoffs. Quantification was performed using interference correction and at least three fragments per precursor, without imputation. Quantity is based on MS² XIC peak areas. Non-identified precursors in rows with at least one q-value below 0.01 were selected for iRT profiling, by enabling carrying over the average template peak position.

For GPF data, a chromatogram library was generated in Spectronaut (v. 15.7). The six raw files were processed altogether with the Pulsar search engine with similar parameters than the ones used for DIA analysis: trypsin/P as digestion enzyme, methionine oxidation and protein N-term acetylation were as variable modifications, one missed cleavage authorized. Additional filters were applied for library generation, including a 1% FDR threshold at PSM, peptide and protein levels and a number of fragments between 3 and 8 per precursor. Then, data acquired in DIA were searched against the chromatogram library with the same parameters as for library-free DIA data analysis described above.

3. Top-down and middle-down mass spectrometry to characterize mAb-based therapeutics

A. Optimization of top-down and middle-down acquisition parameters

i) Optimization of key parameters for top-down acquisition by infusion

a) Sample preparation

Myoglobin from equine heart (Sigma-Aldrich) and carbonic anhydrase from bovine erythrocytes (Sigma-Aldrich) were diluted to 2 μ M in 49.5/49.5/1 H₂O/ACN/FA

b) MS/MS analysis

Proteins were analyzed by infusion at 5 μ L/min using the H-ESI source (Thermo Fisher Scientific) heated at 275°C hyphenated to an Orbitrap Eclipse Tribrid instrument (Thermo Fisher Scientific) in Intact Protein mode. Spray voltage was set to 3,800 V, sheath gas to 4.0 arbitrary units (a. u.) and auxiliary gas to 2.0 a. u. Ion routing multipole pressure was set to 2 mTorr. Fragmentation was carried out using HCD, ETD, EThcD and UVPD. Activation energies and precursor ions charge state are described in the main text (see **Part II, Chapter 1**). For ETD and EThcD fragmentations, the anionic reagent target was set at 7×10^5 with a maximum reagent injection time of 200 ms. Fragmentation MS/MS spectra were collected in the Orbitrap at a resolution of 120,000 (at 200 m/z) using 10 micro scans and averaged for 1 min. For MS3, precursor product ion selection and MS collection were performed as for MS2.

c) Data analysis

Spectra were averaged over 1 min and deconvoluted using Xtract™ algorithm included in the FreeStyle software (Thermo Fisher Scientific) using the following parameters: S/N was set to 3, fit factor to 80%, remainder threshold to 25, maximum charge state to the charge state of the precursor ion and minimum number of charges to 1. Deconvoluted [MH⁺] masses were exported as an Excel file. Fragment identification was performed using ProSight Lite with a mass tolerance of 10 ppm, and ion types were searched accordingly to the fragmentation method. For UVPD data, UVPD9 was used to match the masses. Protein sequences are provided in **Table 15**.

Table 15: Protein sequences of myoglobin and carbonic anhydrase II.

Protein	Monoisotopic Mass	Sequence
Myoglobin (P68082)	16,940.96 Da	GLSDGEWQQVFLNVWGKVEADIAGHGQEVLRIRLFTGHPETLEKFDKFKHLKTEAE MKASEDLKKHGTVVLTALGGILKKKGHHEAEKPLAQSHATKHKIKPIKYLEFISD AIIHVLHSHKHPGDFGADAQGMATKALELFRNDIAAKYKELGFQG
Carbonic anhydrase II (P00921)	29,005.70 Da	[acetylation]SHHWGYGKHNGPEHWHKDFPIANGERQSPVDIDTKAVVQDPALK PLALVYGEATSRMNVNNGHSFNVEYDSDQKAVLKDGPLTGTYRLVQFHFHWG SSDDQGSEHTVDRKKYAAELHLVHWNTKYGDFGTAAQQPDGLAVGVFLKVG ANPALQKVLDAALDSIKTKGKSTDFPNFDPGSLLPNVLNYWTPGSLTTPPLESV TWIVLKEPISVSSQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK

ii) Optimization of the acquisition parameters for MD-MS analysis of mAbs

a) Sample preparation

Trastuzumab was provided by the Institut de Cancérologie de Strasbourg (ICANS, Strasbourg, France). FabRICATOR® enzyme (Genovis, Lund, Sweden) was prepared by diluting 300 units in 75 μ L of water (4 units/ μ L). In an Eppendorf, 300 μ g of trastuzumab (~15 μ L) were mixed to 75 μ L of the IdeS solution and 10 μ L of phosphate buffer at pH 6.8. Sample was incubated for 30 min at 37°C and 500 rpm. After

digestion, 33 mg of 6 M guanidine-HCl and DTT 10 μ L of DTT at 1 M (final concentration of 100 mM) were added to the sample. Reduction was performed at 57°C for 1 h. Sample was diluted to a final concentration of 0.5 μ g/ μ L in 49.5/49.5/1 H₂O/ACN/FA. Around 1 μ g of sample were loaded on the column.

b) LC-MS/MS analysis

Short chromatographic gradient

LC separation of the subunits was performed on a Dionex UltiMate 3000 RSLC nano system (Thermo Fisher Scientific). Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in ACN. Antibody subunits were eluted on a Zorbax 300 SB-C8 column (50 mm x 2.1 mm, 1.8 μ m, Agilent) at a flow rate of 100 μ L/min following a linear gradient of 12 min: from 20% to 50% B in 5.5 min, 90% B for 2 min and equilibration at 20% B for 4 min.

MS and MS/MS analyses were performed on an Orbitrap Eclipse™ Tribrid™ (Thermo Fisher Scientific) mass spectrometer using the H-ESI source (Thermo Fisher Scientific) and working in Intact Protein mode with low pressure (2 mTorr). Source parameters were set as follows: ion transfer tube to 320°C, vaporizer to 300°C, spray voltage to 3,500 V, sheath gas to 25 a. u., auxiliary gas to 10 a. u. and in-source fragmentation of 10 V. Selected precursors ions were fragmented in a defined time scale corresponding to the elution peak of each subunit. The list of precursors used is mentioned in the main text (**Part III, Chapter 1**). Precursors were fragmented either using HCD at 25% NCE, ETD with 3 ms reaction time or UVPD with 30 ms activation time. Same ETD parameters were used as those described in **Experimental section 3.A.i.b**). MS/MS spectra were collected in the Orbitrap at a resolution of 120,000 (at 200 m/z) using two micro scans per scan. AGC target was set to 7.5×10^5 , maximum injection time to 200 ms and isolation windows to 3 m/z .

Long chromatographic gradient

LC separation of the subunits was performed on a Dionex UltiMate 3000 RSLC nano system (Thermo Fisher Scientific). Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in ACN. Antibody subunits were eluted on a BioResolve™ RP mAb column (150 mm x 2.1 mm, 2.7 μ m, Waters) at a flow rate of 300 μ L/min following a linear gradient of 60 min: from 25% to 44% B in 40 min, 75% B for 5 min and equilibration at 25% B for 14 min.

MS and MS/MS analyses were performed on an Orbitrap Eclipse™ Tribrid™ (Thermo Fisher Scientific) mass spectrometer using the H-ESI source (Thermo Fisher Scientific) and working in Intact Protein mode with low pressure (2 mTorr). Source parameters were set as follows: ion transfer tube to 275°C, vaporizer to 200°C, spray voltage to 3,800 V, sheath gas to 15 a. u., auxiliary gas to 5 a. u. and in-source fragmentation of 10 V. Selected precursors ions were fragmented in a defined time scale corresponding to the elution peak of each subunit. The list of precursors used is mentioned in the main text (**Part III, Chapter 1**). Precursors were fragmented either using HCD at 25% NCE, CID at 35%, ETD with 3 ms reaction time, UVPD with 15 ms activation time and EThcD with 3 ms and 20% NCE. Same ETD parameters were used as those described in **Experimental section 3.A.i.b**). MS/MS spectra were collected in the Orbitrap at a resolution of 120,000 (at 200 m/z) using two micro scans per scan. AGC target was set to 1×10^6 , maximum injection time to 200 ms and isolation windows to 3 m/z .

PTCR was performed after ETD fragmentation on either a wide isolation window of 1 000 m/z (700–1 700 m/z) or five isolation windows of 200 m/z between 700 and 1 700 m/z . The PTCR reaction time was set to 20 ms for the wide isolation windows. For narrow windows, reaction times of either 5 ms (for 700–900 m/z , 1 300–1 500 m/z and 1 500–1 700 m/z isolation windows) or 10 ms (for 900–1 100 m/z and 1 100–1 300 m/z isolation windows) were used, depending on the ion population. Reagent target was set to 5×10^5 and the maximum reagent injection time to 200 ms. MS³ spectra were

recorded using a mass range of 500–4 000 m/z , a resolving power of 120 000 (at 200 m/z) and 2 microscans per spectra were selected.

c) Data analysis

Spectra were averaged over the chromatographic peak and deconvoluted using Xtract™ algorithm included in the FreeStyle software (Thermo Fisher Scientific) using the following parameters: S/N was set to 3, fit factor to 80%, remainder threshold to 25, maximum charge state to the charge state of the precursor ion and minimum number of charges to 1. Deconvoluted [MH⁺] masses were exported as an Excel file. Fragment identification was performed using ProSight Lite with a mass tolerance of 10 ppm, and ion types were searched accordingly to the fragmentation method. For UVPD data, UVPD9 was used to match the masses. Trastuzumab subunit sequences are provided in **Table 16**.

Table 16: Trastuzumab subunits sequences.

Subunit	Monoisotopic Mass	Sequence
Lc	23,196.47 Da	DIQMTQSPSSLSASVGDRTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLY SGVPSRFSGSRSGTDFLTITSLQPEDFATYYCQQHYTTPPTFGGQTKVEIKRTVA APSVFIFPPSDEQLKSGTASVCLLNFPYFREAKVQWKVDNALQSGNSQESVTEQ DSKDSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
Fc/2 (G0F)	25,220.46 Da	GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTK PREEQYN[G0F]STYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKG QPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPP PVLDSGDGFFLYSKLTVDKSRWQQGNVFCFSVMHEALHNHYTQKSLSLSPG
Fd	25,367.52 Da	EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLVWVARIYPTN GYTRYADSVKGRFTISADTSKNTAYLQMNSLRRAEDTAVYYCSRWGGDGFYAMD YWGQGLTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNS GALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKEVE PKSCDKTHTCPPCPAPELLG

For internal fragments analysis, ClipsMS³³⁵ was used. Internal fragments were searched with an error tolerance of 2 ppm (10 ppm was kept for the terminal fragments) and a smallest internal fragment size of 5 amino acids. The following internal fragment ions were searched depending on the fragmentation method: *by* ions for CID, *cz* ions for ETD, *by*, *cy*, *bz* and *cz* ions for EThcD and *ax*, *bx*, *cx*, *ay*, *by*, *cy*, *az*, *bz* and *cz*. Other parameters (modifications, terminal fragments tolerance) were set as mentioned above

iii) Assessment of data processing for TD-MS and MD-MS experiments

a) Sample preparation

The same samples as those described in **Experimental Section 3.A.ii.a)** were used.

b) LC-MS/MS analysis

The same analysis parameters as those described in **Experimental Section 3.A.ii.b)** were used.

c) Data analysis

Spectra were averaged over the chromatographic peak and deconvoluted using Xtract™ algorithm included in the FreeStyle software (Thermo Fisher Scientific) using the following parameters: remainder threshold to 25, maximum charge state to the charge state of the precursor ion and minimum number of charges to 1. Different S/N ratio and fit factor values were tested, as mentioned in the main text. Deconvoluted [MH⁺] masses were exported as an Excel file. Fragment identification was performed using ProSight Lite with a mass tolerance from 2-10 ppm, and ion types were searched accordingly to the fragmentation method. For UVPD data, UVPD9 was used to match the masses. Trastuzumab subunit sequences are provided in **Experimental section 3.A.ii.c)**.

B. Characterization of trastuzumab deruxtecan ADC by MD-MS

i) Sample preparation

Trastuzumab Deruxtecan (T-DXd) samples were provided by the Institut de Cancérologie Strasbourg Europe (ICANS, Strasbourg, France). All chromatographic solvents are LC-MS grade.

a) Middle-up analysis sample preparation

For middle-up level experiments, IdeS digestion was performed by incubating one unit of FabRICATOR® enzyme (Genovis, Lund, Sweden) per microgram of ADC for 30 min at 37°C. For denaturing conditions analysis, TFA was added to the digested ADC to a final concentration of 1%. The final concentration of the ADC in solution was 0.8 µg/µL.

b) Middle-down analysis sample preparation

For middle-down experiments, sample IdeS digestion was performed as mentioned for middle-up experiments. Disulfide bonds were subsequently reduced during 60 min at 37°C in strong denaturing conditions (6M guanidine hydrochloride,) using DTT (Sigma) as a reducing agent (100 mM final DTT concentration). Finally, trifluoroacetic acid (TFA) was added to the solution at a final concentration of 1%. The final concentration of the ADC in solution was 0.7 µg/µL.

ii) LC-MS/MS analysis

a) SEC-nMS analysis experiments

A Dionex UltiMate 3000 RSLC system (Thermo Fisher Scientific) was coupled to an Orbitrap Eclipse™ Tribrid™ (Thermo Fisher Scientific) mass spectrometer equipped with an EASY Max NG ion source with a heated electrospray ionization H-ESI probe. 15 and 12 µg were injected for intact and middle-up level analyses, respectively. The separation was carried out in isocratic mode with a 100 mM AcONH₄ mobile phase at pH 7 using an ACQUITY Protein BEH SEC (200 Å, 1.7 µm, 4.6 × 150 mm, Waters) SEC column. The flow rate was set to 250 µL/min for both intact and middle-up analyses. The source parameters were set as follows: sheath gas flowrate to 25 a. u., auxiliary gas flowrate to 5 a. u., ion transfer capillary temperature to 275°C, vaporizer temperature to 200°C, and capillary voltage 3.8 kV. In-source fragmentation energy was set to 150 V and 50 V for the intact and middle-up analysis, respectively, and the RF Lens to 150%. Intact protein mode was activated, with a pressure of 20 mTorr in the ion routing multipole (high pressure mode). Acquisitions were performed in the *m/z* range 500–8,000 with a 3 ms scan and at a resolution of 15,000. The AGC target was set to 2×10⁶, and the maximum injection time was 50 ms.

b) Middle-level LC-MS and LC-MS/MS analysis

Separation of the Fc/2, Fd, and Lc subunits was performed with a BioResolve™ RP mAb Polyphenyl column (450 Å pore size, 2.7 µm particle size, 2.1 × 150 mm, Waters) preheated at 60 °C, using a Dionex Ultimate 3000 RSLC system (Thermo Fisher Scientific). Mobile phase A consisted of 0.1% TFA in water, and mobile phase B was 0.1% TFA in ACN. For LC-MS and LC-MS/MS analyses, 3.5 µg and 5.5 µg of digested ADC were respectively loaded onto the column using 25% of mobile phase B at constant flow rate of 300 µL/min during 2 minutes. Digested ADC was eluted following a linear gradient: 25% at 2 min, 44% B at 40 min and 75% B from 40.5 to 45.5 min. The column was finally re-equilibrated at 25% B during 15 min. The LC system was hyphenated to an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) in Intact Protein mode. RF-Lens was set to 60%, and auxiliary and sheath gas were set to 5 a. u. and 15 a. u., respectively. For all experiments, the spray voltage was set to 3.8 kV, the ion transfer tube temperature to 275°C, the vaporizer to 200°C and the advanced

peak determination was activated. MS spectra were recorded using a mass range of 500–4,000 m/z and resolving power of 120,000 (at 200 m/z) for denaturing middle-up analysis.

For middle-down analysis, four different fragmentation modes were used: CID, ETD, PTCR and EThcD. MS² spectra were recorded using a mass range of 300–2 000 m/z, a resolving power of 120 000 (at 200 m/z) and 2 microscans per spectra were selected. Precursor ions were selected over each specific elution window with an isolation window of 3 m/z and subsequently fragmented. For each subunit, the most intense, for CID, or the highest charge state, for ETD/EThcD, precursor ions were selected and submitted to fragmentation. The AGC was set to 1×10^6 and the maximum injection time to 200 ms for precursor ions. CID collision energy was set to 35%, with a CID activation time of 10 ms and an activation Q parameter of 0.25. For ETD fragmentation, the reaction time between precursor ions and the anionic fluoranthene reagent was set to 3 ms. To ensure proper precursor ions fragmentation, the anionic reagent target was set at 7×10^5 with a maximum reagent injection time of 200 ms. For EThcD fragmentation, the same parameters as those described for ETD were used, and a supplemental collision energy of 20% was used. PTCR was performed after ETD and EThcD fragmentation on either a wide isolation window of 1,000 m/z (700–1,700 m/z) or five isolation windows of 200 m/z between 700 and 1,700 m/z. The PTCR reaction time was set to 20 ms for the wide isolation windows. For narrow windows, reaction times of either 5 ms (for 700–900 m/z, 1,300–1,500 m/z and 1,500–1,700 m/z isolation windows) or 10 ms (for 900–1,100 m/z and 1,100–1,300 m/z isolation windows) were used, depending on the ion population. Reagent target was set to 5×10^5 and the maximum reagent injection time to 200 ms. MS³ spectra were recorded using a mass range of 500–4,000 m/z, a resolving power of 120,000 (at 200 m/z) and 2 microscans per spectra were selected.

iii) Data analysis

a) SEC-nMS analyses

Data were interpreted using BioPharma Finder v3.2 (Thermo Fisher Scientific, San Jose, CA, USA). The deconvolution mass tolerance was set to 10 ppm.

b) MD-MS analysis

MS/MS spectra were deconvoluted in FreeStyle 1.8 using the Xtract algorithm and deconvoluted masses were matched to the sequence using ProSight Lite v1.4. MS/MS spectra were averaged through the different subunit elution windows (0.5 min, 0.8 min and 0.6 min for the Fc/2, Lc and Fd subunits, respectively) and then deconvoluted using a S/N ratio of 7, a fit factor of 80% and a remainder threshold of 25%. Deconvoluted masses were exported to csv file and matched to the sequence with a 10 ppm tolerance (see **Experimental section 3.A.ii.c**). Different fragment ion series were considered depending on the fragmentation technique: *b/y* for CID, *c/z* for ETD, and *b/y* and *c/z* in the case of EThcD. Because of DTT reduction performed during the sample preparation, all cysteines involved in intra-chain disulfide bridges were considered in their reduced form. For the heavy chain and Fc/2 subunits, ions corresponding to the G0F glycoform were targeted so G0F (+ 1,444.5339 Da) on Asp300 of the Hc (corresponding to Asp61 when Fc/2 is considered) was defined as a fixed modification to search the data. Drug conjugation (+ 1,033.394 Da) on the cysteines involved in inter-chain disulfide bridges (Cys223, Cys229 and Cys232 on the Fd, Cys214 on the Lc) were set as fixed modifications.

For internal fragments analysis, ClipsMS³³⁵ was used. Internal fragments were searched with an error tolerance of 2 ppm (10 ppm was kept for the terminal fragments) and a smallest internal fragment size of 5 amino acids. The following internal fragment ions were searched depending on the fragmentation method: *by* ions for CID, *cz* ions for ETD and *by*, *cy*, *bz* and *cz* ions for EThcD. Other parameters (modifications and terminal fragments tolerance) were set as mentioned above.

ANNEXES

RESEARCH ARTICLE

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

 Corentin Beaumal^{1,2}  | Alain Beck³ | Oscar Hernandez-Alba^{1,2} | Christine Carapito^{1,2}

¹Laboratoire de Spectrométrie de Masse BioOrganique, IPHC UMR 7178, CNRS, Université de Strasbourg, Strasbourg, France

²Infrastructure Nationale de Protéomique ProFI - FR2048, Strasbourg, France

³IRPF, Centre d'Immunologie Pierre-Fabre (CIPF), Saint-Julien-en-Genevois, France

Correspondence

Christine Carapito, Laboratoire de Spectrométrie de Masse BioOrganique, IPHC UMR 7178, CNRS, Université de Strasbourg, 25 rue Becquerel, F-67087 Strasbourg, France.
 Email: ccarapito@unistra.fr

Funding information

Agence Nationale de la Recherche, Grant/Award Numbers: ANR-10-INBS-08-03, ProFI FR2048; Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche

Abstract

Therapeutic monoclonal antibodies (mAb) production relies on multiple purification steps before release as a drug product (DP). A few host cell proteins (HCPs) may co-purify with the mAb. Their monitoring is crucial due to the considerable risk they represent for mAb stability, integrity, and efficacy and their potential immunogenicity. Enzyme-linked immunosorbent assays (ELISA) commonly used for global HCP monitoring present limitations in terms of identification and quantification of individual HCPs. Therefore, liquid chromatography tandem mass spectrometry (LC-MS/MS) has emerged as a promising alternative. Challenging DP samples show an extreme dynamic range requiring high performing methods to detect and reliably quantify trace-level HCPs. Here, we investigated the benefits of adding high-field asymmetric ion mobility spectrometry (FAIMS) separation and gas phase fractionation (GPF) prior to data independent acquisition (DIA). FAIMS LC-MS/MS analysis allowed the identification of 221 HCPs among which 158 were reliably quantified for a global amount of 880 ng/mg of NIST mAb Reference Material. Our methods have also been successfully applied to two FDA/EMA approved DPs and allowed digging deeper into the HCP landscape with the identification and quantification of a few tens of HCPs with sensitivity down to the sub-ng/mg of mAb level.

KEYWORDS

data-independent acquisition mass spectrometry (dia-ms), drug product (dp), high-field asymmetric ion mobility spectrometry (faims), host cell protein (hcp) impurities

1 | INTRODUCTION

Since the approval of the first monoclonal antibody (mAb) in 1986 [1], the Muromonab-CD3, an immunoglobulin reducing acute rejection in patients with organ transplants, interest for this class of biopharmaceutical products has been growing fast [2, 3]. Today, the United States Food and Drug Administration (FDA) and/or the European Medicines

Agency (EMA) have accepted more than 125 mAbs and about twenty are currently under review by these regulatory agencies [4, 5]. Production of recombinant therapeutic proteins using expression cell lines technology results in the presence of bioprocess-related host cell protein (HCP) impurities in addition to the desired mAb. During the downstream process (i.e., Protein A affinity chromatography, ion exchange chromatography, hydrophobic interaction chromatography, and multiple filtration steps), the total amount of HCPs and the mixture complexity are drastically reduced [6, 7]. However, low levels of HCPs may co-purify with the therapeutic protein and thus remain in the drug products (DPs). Some of these residual HCPs may degrade the drug or

Abbreviations: CQA, critical quality attribute; CV, compensation voltage; DDA, data-dependent acquisition; DIA, data-independent acquisition; DP, drug product; FAIMS, high-field asymmetric waveform ion mobility spectrometry; GPF, gas phase fractionation; HCP, host cell proteins; NIST, National Institute of Standards and Technology.

excipients in the DP, potentially leading to reduced stability or inactivation of the mAb [8–11], while other impurities may endanger patients' safety by causing an unexpected and damaging immune response weakening the therapeutic protein efficiency [10, 12–14]. The potential risks associated with HCPs have conducted the regulatory agencies to consider them as a critical quality attribute (CQA) in DPs, and their total amount should be minimized to trace levels and monitored by highly sensitive analytical methods during downstream process [15–17]. Even if general guidelines indicate that there is no strict limit on the HCPs level in DPs, many biopharmaceutical companies consider that their global amount should not exceed 100 ppm (100 ng/mg of mAb) in the DP [18–20]. Besides, it cannot be excluded that the presence of a specific HCP, even at a very low level, could jeopardize the action of the therapeutic protein [21]. Currently, enzyme-linked immunosorbent assays (ELISA) are commonly used for HCP impurities analysis for their easy-of-use and high precision [22–24]. However, ELISA has some limitations: certain HCP species are not targeted by the assay and thus cannot be detected and only global quantification is possible, with neither individual HCP identification nor quantification possible [7, 22, 25]. In addition, the development of fit for purpose product-specific immunoassays to measure HCPs in DP is a long, expensive, and fastidious process, as they must satisfy numerous criteria (accuracy, precision, range, linearity, etc.) to be validated [26].

Therefore, liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) has emerged as a promising orthogonal and complementary method to characterize HCP impurities over the bioprocess and ultimately in DPs [27–29]. Multiple studies point out LC-MS/MS benefits for HCP monitoring, using targeted approaches (selected reaction monitoring (SRM) and parallel reaction monitoring (PRM)) [30–34], data-dependent acquisition (DDA) [35–37] and data-independent acquisition (DIA) [33, 38, 39]. DIA presents the advantage of allowing the extraction of quantitative information on all detectable species through MS2 data [40]. Until recently, the use of spectral-libraries and so-called “peptide centric” approaches was the most efficient way to interpret DIA data, while time and sample consuming due to the necessity to generate the most comprehensive spectral library possible in DDA prior to DIA analysis [41, 42]. Today, “spectrum centric” approaches are growing in interest, thanks to continuous improvements of acquisition strategies [43–46] and the development of innovative tailored tools, allowing identification and precise quantification without spectral libraries [47–50] or using artificial intelligence (AI) in silico-predicted libraries [51–53]. While the dynamic range achievable by latest generation LC-MS/MS couplings reaches three to four orders of magnitude, quantifying ppm-level HCP in DPs by definition requests six orders of dynamics. To achieve this, a few studies have focused on optimizing sample preparation via mAb depletion using native digestion [54] or reversely via HCPs enrichment using several approaches: molecular weight cutoff enrichment [55], protein A affinity chromatography [56, 57] or hydrophilic interaction chromatography (HILIC) [58]. Other studies relied on improvements of the LC-MS/MS part including, among others, the multiple chromatography separations [31, 37, 59, 60] or implementing an ion mobility separation step [30, 61, 62].

Significance Statement

Host cell protein (HCP) impurities monitoring in biotherapeutics is a critical concern because of their potential adverse actions. In this study, we developed innovative mass spectrometry (MS)-based approaches on a tribrid instrument to overcome the limitations of standard methods for the characterization of trace level HCP impurities in DPs. We demonstrate the benefits of implementing a High-Field Asymmetric Ion Mobility Spectrometry (FAIMS) separation step in the workflow as well as of the use of Data Independent Acquisition (DIA) combined with a Gas-Phase Fractionation approach to address the challenges of accurate and precise quantification of very low-level HCPs in samples presenting an extreme dynamic range. Major improvements towards robust quantification of trace-level HCPs in DPs have been demonstrated both on the NIST mAb Reference Material standard and two FDA/EMA approved DPs, down to the sub-ppm level.

In this study, we focused on the most challenging samples, namely DPs. We have evaluated several LC-MS/MS settings for their potential to identify the most exhaustive list of HCPs but also to reliably quantify the largest possible fraction of them. First, we analyzed the NIST mAb reference material standard using DDA and DIA acquisition methods, leading to the identification of hundreds of HCP impurities and the accurate quantification of tens of them. Then, the high field asymmetric ion mobility spectrometry (FAIMS) device allowing gas-phase separation of co-eluting peptides through fast internal compensation voltages (CV) switching, was implemented in front of the mass spectrometer to evaluate its benefits in terms of depth and coverage of HCP impurities detection. In parallel, a gas-phase fractionation DIA strategy was evaluated. Both workflows were benchmarked against results described in the literature. Finally, we applied our methods to monitor HCPs in two FDA/EMA approved mAbs with a very high medical benefits in oncology, trastuzumab and nivolumab, enabling the identification and quantification of several tens of HCPs and demonstrating that these approaches can be used to routinely analyze DPs in a rapid and accurate way.

2 | MATERIALS AND METHODS

2.1 | Materials

The NIST mAb reference material standard RM8671, dithiothreitol (DTT) and Tris-HCl were purchased from Merck. Commercially available mAbs, trastuzumab and nivolumab, were obtained as European Union pharmaceutical-grade DPs from the Pierre Fabre Immunologic Center (CIPF, Saint-Julien-en-Genevois, France). HCP-PROFILER

standard peptides were acquired from Anaquant (Villeurbanne, France). Mass spectrometry grade trypsin/Lys-C enzymes were obtained from Promega (Madison, WI, USA).

2.2 | Sample preparation

The digestion protocol was adapted from Huang et al. [54]. Briefly, a 100 μL aliquot of 10 $\mu\text{g}/\mu\text{L}$ mAb was supplemented with 90 μL of ultrapure H_2O and 5 μL of 1 M Tris-HCl, pH 8. Proteins were digested overnight at 37°C using a solution of Trypsin/Lys C enzymes at a 1:400 enzyme/protein ratio. Digests were reduced with 1.5 μL of 303 mM DTT for 10 min at 90°C and centrifuged at 13 000 $\times g$ for 2 min. Supernatant was finally acidified with 0.5 μL of formic acid (FA). Then, after *vacuum* drying, peptides were solubilized in 2% acetonitrile (ACN) and 0.1% FA to obtain a final protein concentration of 1 $\mu\text{g}/\mu\text{L}$. For gas-phase fractionation (GPF) chromatogram library generation, technical replicates were pooled together. One HCP-PROFILER bead was dissolved in 150 μL of 5% ACN and 0.1% FA, to obtain standard peptides in a concentration range from 0.5 to 250 fmol/ μL . Retention time iRT standards (Biognosys, Schlieren, Switzerland) and 2 μL of HCP-PROFILER solution were mixed to 1 or 2 μL of FDA/EMA approved mAb or NIST mAb Reference Material, respectively, before injection.

2.3 | Nano LC-MS/MS analysis

Data-dependent acquisition (DDA) and data-independent acquisition (DIA) analyses were performed on a Dionex UltiMate 3000 RSLC nano system (Thermo Scientific) coupled to an Orbitrap Eclipse Tribrid (Thermo Scientific) mass spectrometer, equipped with a FAIMS Pro interface (Thermo Scientific). Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in 80% ACN/20% water. Peptides were loaded onto an Acclaim PepMap 100 C18 20 mm \times 0.1 mm, 5 μm diameter particles precolumn (Thermo Scientific) for 3 min at 10 $\mu\text{L}/\text{min}$, and eluted on an Aurora Series C18 UHPLC (250 mm \times 75 μm , 1.6 μm diameter particles, IonOpticks) at a 300 nL/min flow rate following a linear gradient: 2.5% B at 0 min, 43.8% B at 95 min and 98% B from 96 to 101 min. The column was finally re-equilibrated in 2.5% B for 15 min.

For DDA analyses, full-scan MS spectra were acquired over a 375–1500 m/z range at a resolution of 60,000 (at 200 m/z), with an automatic gain control (AGC) target of 1.10^6 and a maximum injection time of 60 ms. The top 10 most intense precursor ions with an intensity exceeding 5.10^3 ions per second and charge states between 2 and 6 were automatically selected from each MS spectrum for higher-energy collisional dissociation (HCD) fragmentation at 30% normalized collision energy. MS/MS spectra were collected from 120 to 1200 m/z in the ion trap in Rapid mode, with an AGC target of 1.10^4 and a maximum injection time set to Automatic. Experiments with the FAIMS Pro interface were conducted as described in Hebert et al. [62]. Briefly, FAIMS carrier gas flow was 4.7 L/min N_2 , the FAIMS electrodes were set to 100°C, entrance plate voltage was 250 V, asymmetric waveform

divert voltage was –5000 V, CV settling time was 25 ms and CVs values of –50, –65, and –85 V were applied successively in the same run.

For DIA analyses, full-scan MS was acquired over a 375–1250 m/z range at a resolution of 60,000 (at 200 m/z) with an AGC target of 1.10^6 and a maximum injection time of 60 ms. Fragment analysis (MS/MS) was subdivided into 42 windows of variable widths, described in Table S1. A resolution of 15,000 (at 200 m/z), an AGC target of 1.10^6 and a maximum injection time of 22 ms were used for a total cycle time of 1.5 s. HCD fragmentation normalized collision energy was set to 30%. For DIA-GPF, 6 individual injections were performed, covering the 380–980 m/z mass range. A 100 m/z full-scan MS was acquired with the parameters described above for DIA, and the MS/MS analysis was divided into 25 windows of 4 m/z using a 30,000 (at 200 m/z) resolution, an AGC target of 1.10^6 and maximum injection time of 60 ms.

2.4 | DDA data analysis

DDA data analysis was performed with Proteome Discoverer v. 2.5 (Thermo Fisher Scientific) using Sequest HT (Thermo Fisher Scientific) against a database containing all *Mus musculus* entries (17,050 entries, TaxID = 10090, 2021/05/26) extracted from UniProtKB/SwissProt for NIST mAb samples and a database containing all *Cricetulus Griseus* entries (78,366 entries, Tax ID = 10029, 2022/03/15) extracted from UniProtKB/TrEMBL for trastuzumab and nivolumab samples. Each fasta file included common contaminants, iRT retention time standards, HCP-PROFILER standard sequences and mAbs heavy and light chain sequences. Trypsin/P was used as digestion enzyme and one missed cleavage was allowed. Methionine oxidation and protein N-term acetylation were set as variable modifications. Mass tolerance was set to 5 ppm for precursor ion masses and 0.5 Da for fragment ion masses analyzed in the ion trap. The false discovery rate (FDR) was calculated using the Percolator node set to 1% at PSM, peptide and protein levels. XIC-MS1 quantification was performed using unique peptides only and the chromatographic alignment was set to 10 min and 10 ppm between the three technical replicates.

2.5 | DIA data analysis

DIA data analysis was performed with Spectronaut v. 15.7 (Biognosys) using the fasta files described above and the following settings. Briefly, trypsin/P was set as digestion enzyme, methionine oxidation and protein N-term acetylation were set as variable modifications and one missed cleavage was allowed. Data was extracted using dynamic mass tolerances and a 10 min XIC RT extraction window. Identification was performed using 0.01 precursor and protein Qvalue cutoffs. Quantification was performed using interference correction and at least three fragments per precursor, without imputation. Quantity is based on MS^2 XIC peak areas. Non-identified precursors in rows with at least one q-value below 0.01 were selected for iRT profiling, by enabling carrying over the average template peak position.

For GPF data, a chromatogram library was generated in Spectronaut (v. 15.7). The six raw files were processed altogether with the Pulsar search engine with similar parameters than the ones used for DIA analysis: trypsin/P as digestion enzyme, methionine oxidation and protein N-term acetylation were as variable modifications, one missed cleavage authorized. Additional filters were applied for library generation, including a 1% FDR threshold at PSM, peptide and protein levels and a number of fragments between 3 and 8 per precursor. Then, data acquired in DIA were searched against the chromatogram library with the same parameters as for library-free DIA data analysis described above.

2.6 | HCP quantification

Validation filters described in Pythoud et al.[39] were applied prior to HCP quantification. Briefly, oxidized and acetylated precursors and their counterparts were removed and only host organism or standard protein precursors with charge states of two or three were kept. For DDA analysis, shared and not quantified precursors as well as precursors with rank > 1 and with |RT Sequest–RT Top Apex| > 4 min were removed. One missing value and/or one by matching value at maximum were authorized over the three replicates and a CV on the triplicate precursor intensity values below 20% must be achieved to consider the precursor for quantification. Finally, a sequence homology filter was applied to remove HCP peptides that can potentially arise from unspecific cleavage or mAb degradation.

After filtering, peptide intensity was obtained by summing all precursor ion intensities and protein intensity by summing the three most intense peptides, based on the Top3 strategy described by Silva et al. [63]. Protein mole quantities were estimated using the equation of the HCP Profiler calibration curve. Individual HCP ng/mg mAb (ppm) amounts were estimated using the molecular weight and injected mAb quantities.

3 | RESULTS AND DISCUSSION

3.1 | LC-MS/MS methods to precisely and reliably monitor trace-level HCPs

Three technical replicates of NIST mAb reference material were prepared and analyzed in DDA mode on a nanoLC-Orbitrap Tribrid coupling, with fragment detection (MS2) performed in the ion trap. These analyses led to the detection and identification of 325 peptides originating from 133 distinct HCP protein groups (Figure 1A). We compared the list of protein impurities identified in our study to the lists of HCPs identified and/or quantified in recently published papers from literature, applying comparable sample preparation (based on Huang native digestion protocol) and DDA LC-MS/MS acquisition methods [36, 54, 64]. The results obtained point out a fairly good overlap between the studies, with 43 proteins (32% of the proteins detected in our study) identified in common in all four studies, and 21 additional

HCPs (16% of the proteins detected in our study) shared by at least two papers and our results (Figure 1B). Overall, this demonstrates the reproducibility of the native digestion protocol, independently of the LC-MS/MS system used and the location of experiments. This is a critical point regarding the perspective of transferring this protocol in a biopharmaceutical environment. Additionally, Figure 1B also highlights that we have uniquely identified 54 HCPs thus slightly outperforming the other methods. Then, quantification of these trace-level proteins was performed on the precursor ion intensities extracted at MS1 level. To ensure accurate and robust quantification of the impurities detected, stringent filters based on the signal quality and repeatability were applied as described in Pythoud et al. [39]. Briefly, only 2+ and 3+ precursors were kept, shared peptides, peptides without any quantification value and modified peptides (i.e., carrying a methionine oxidation or acetylation) were removed, as well as precursors with a difference between their apex RT and their elution RT higher than 4 min. Then, peptides with missing values in more than one of the three replicates and peptides identified using match between run in more than one over three replicates were removed. Finally, a last filter on the signal repeatability was applied, based on the removal of peptides with a coefficient of variation higher than 20% on measured intensities over the three replicates. Ultimately, to prevent any misidentification of the remaining peptides, a sequence homology filter was applied to remove any peptides that could potentially originate from unspecific mAb cleavage or degradation. Altogether, these stringent filters reduced the number of peptides by 63% (from 325 identified to 121 quantified) and protein groups by 49% (from 133 identified to 68 quantified), as illustrated in Figure 1A. This decrease between identified and quantified features demonstrates that a major part of the identified HCPs is detected at trace-levels, between the detection and quantification limits. The good overlap between our 68 quantified HCPs and previously cited papers with 65% also found in two or three other papers (29 HCPs found in our study and the three papers, 14 in our study and two of the three papers, Figure S1) further strengthens the confidence and reliability on the reported results. Accurate and reproducible characterization and quantification of trace-level HCPs in matrices showing a very high dynamic range with a superabundant mAb such as DPs needs and deserves the use of serious and stringent analytical filters for high-confidence results. Thus, only individual peptides that passed the filters from the replicate associated HCP-PROFILER standard internal calibration curve were calculated. Trauchessec et al. [65] already described the internal calibration benefits for quantification reproducibility and robustness (example of calibration curve in Figure S2). Subsequently, protein group quantities were retrieved based on the Top3 approach, [63] affording an average total amount of HCP impurities of 518 ± 20 ng per mg of mAb or ppm. The quantification accuracy of our results was evaluated against published data in the literature. Figure 1C represents the correlation between the quantities obtained for each HCP in common to our results and selected papers [54, 64]. The Pearson coefficients obtained are 0.97 and 0.99 for the Molden et al. and Huang et al. papers respectively, demonstrating the strong correlation of the absolute quantities calculated. Altogether these results clearly show that advanced LC-MS/MS workflows are perfectly

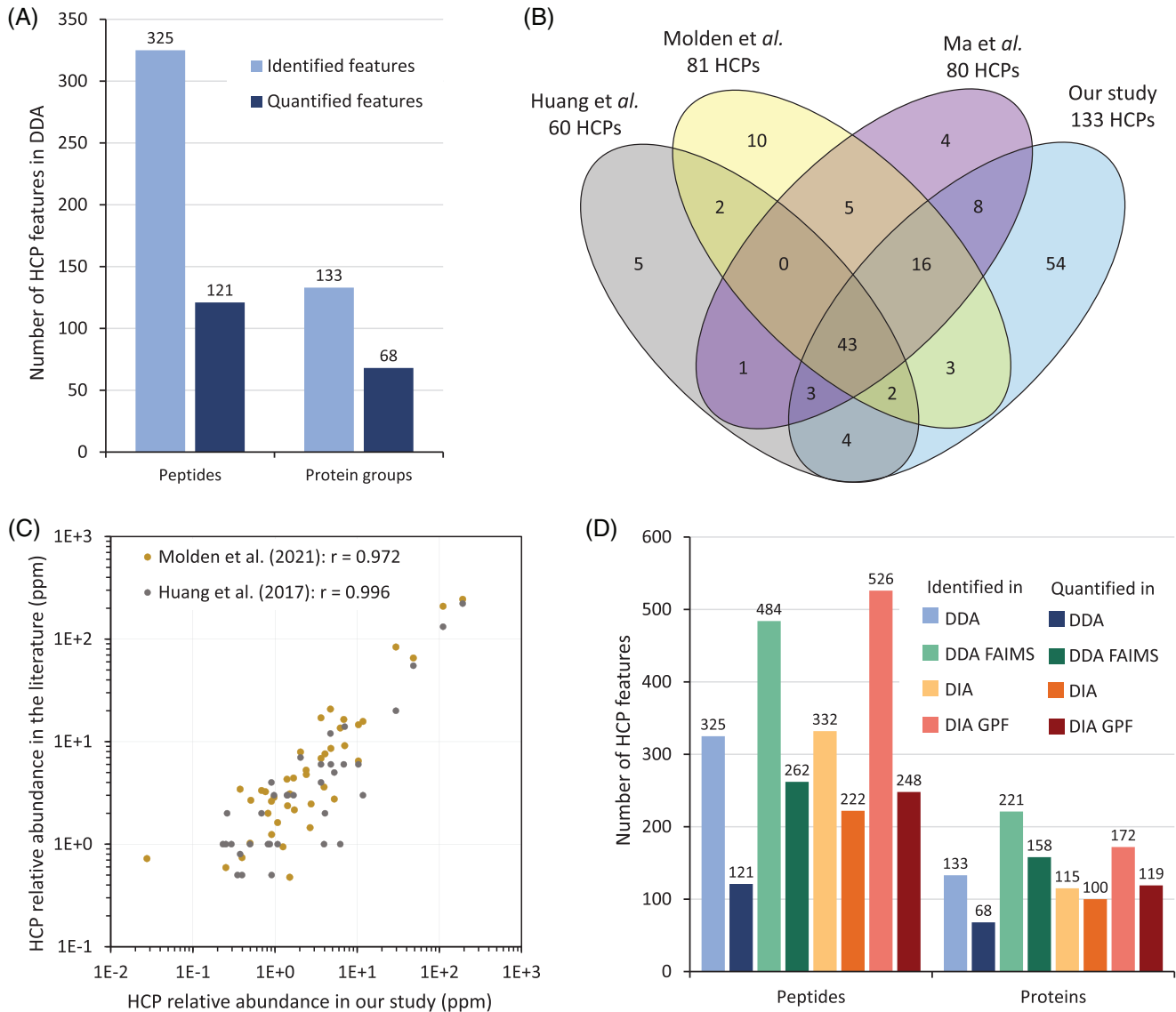


FIGURE 1 (A) Number of HCP peptides and protein groups identified (light blue) and quantified after applying validation filters (navy blue) in DDA mode. (B) Overlap of host cell proteins identified in Huang et al. (2017) [54] (grey), Molden et al. (2021) [64] (yellow), Ma et al. (2020) [36] (purple), and in the current study (blue). All studies used a native digestion protocol adapted from Huang et al. [54]. Data from Molden et al. were acquired in LC-MS/MS, versus nanoLC-MS/MS for Huang et al., Ma et al. (44 min gradient), and our study. (C) Pearson coefficient and correlation between the individual amounts of host cell proteins in ng of HCP per mg of NIST mAb (ppm) in our study using nanoLC-MS/MS DDA analysis after native digestion versus previously reported quantities using native digestion followed by nano-LC-MS/MS obtained by Huang et al. (grey) and Molden et al. (yellow). For representation convenience, results from Molden, originally in micromoles to moles ratio, were converted into ng of HCP to mg of mAb ratio. (D) Number of HCP peptides and protein groups identified (light colors) and quantified after applying validation filters (dark colors) using DDA (blue), FAIMS DDA (green), DIA (orange), and GPF DIA (red).

suites to reliably monitor trace-level HCPs in DPs and thus eventually become the method of choice for quality control of released DPs.

3.2 | Digging deeper into the HCP landscape with the implementation of an additional FAIMS separation

We explored high-field asymmetric waveform ion mobility spectrometry (FAIMS) benefits for detection of trace-level HCPs in DPs. Indeed,

while the native digestion protocol enables the removal of a major part (estimated between 70% and 80% depletion, data not shown) of the mAb and thus reduces the dynamic range in the samples, mAb-specific peptides are still being detected and abundant in comparison to HCP impurities. The use of CVs applied to FAIMS electrodes enables a better resolution of co-eluting peptides, based on the precursor ions gas phase mobility and charge, by filtering the ion population entering the mass spectrometer. To investigate the potential of applying multiple CVs in our LC-MS/MS workflows to separate HCP peptides from co-eluting mAb peptides, the three technical replicates

previously analyzed using nanoLC-DDA analysis were analyzed using an optimized FAIMS method. Chromatograms observed without FAIMS and with different FAIMS CVs applied during the experiment are provided in Figure S3 and clearly illustrate the benefits of ion mobility to clean the signal by removing highly abundant peptides depending on the CV applied (Figure S3A). Indeed, as an example, co-elution of the mAb heavy chain peptide STSGGTAALGCLVK and the mouse PPAC (low molecular weight phosphotyrosine protein phosphatase, accession number: Q9D358) HCP peptide SPIAEAVFR prevented the identification of peptide SPIAEAVFR without FAIMS due to the wide dynamic range between those two peptides. On the contrary, when a CV of -85 V was applied, a clean signal for peptide SPIAEAVFR allowed its identification and quantification while the mAb heavy chain peptide STSGGTAALGCLVK was filtered out at this CV value (Figure S3B and C). Overall, the use of the optimized FAIMS method led to an increase of 49% (from 325 to 484) and 66% (from 133 to 221) of identified peptides and protein groups, respectively (Figure 1D). The validation filters earlier described were applied to quantify HCPs. Figure 1D illustrates that signal quality and repeatability filters resulted in a drop of 46% (from 484 to 262) and 28% (from 221 to 158) between identified and quantified peptides and protein groups, respectively. Of note is that the latter decrease in identified and quantified peptides with FAIMS is less significant compared to the DDA workflow without the ion mobility dimension in the front-end of the MS (63% and 49%, respectively). The signal repeatability filtering step (coefficient of variation $< 20\%$ filter) is the one excluding the highest numbers of peptides and protein groups. Again, the drop is much more pronounced in DDA (26% for proteins and 38% for peptides) compared to FAIMS DDA (10% for proteins and 24% for peptides), demonstrating that data acquired with the FAIMS interface seem to be overall more reproducible. Furthermore, FAIMS addition allowed increasing the number of both peptides and protein groups quantified by more than twofold (from 121 to 262, and from 68 to 158, respectively). Among the 158 quantified protein groups, 53 had also been quantified in DDA without FAIMS, representing 78% of the 68 proteins quantified with DDA only. This significant gain in terms of quantified HCPs logically results in 70% increase of total HCP amount quantified in the NIST mAb, from 518 ± 20 ppm to 880 ± 15 ppm without and with the FAIMS, respectively, thus demonstrating the real asset of adding FAIMS in the workflow. To further describe the benefits of implementing FAIMS, we compared in detail the amounts of each HCPs quantified either with or without the FAIMS. The quantities derived by both methods highly correlate with a Pearson coefficient of 0.97. Moreover, the amount of each individual protein was investigated to show the benefits of FAIMS to afford a more detailed characterization of the trace-level HCP landscape. Figure 2A represents the density of HCP protein groups as a function of the concentration range. It points out the contribution of FAIMS separation in the quantification of low-abundant impurities since more than 75% of HCPs quantified (82 out of 105) exhibited concentrations lower than 2-ppm. Considering these results, FAIMS appeared to be very useful to increase the number of quantified HCP, thanks to dynamic range and sensitivity enhancements. The removal of singly charged peptides and better separation of some highly abundant mAb peptides

improved the signal to noise ratio, and consequently pushing the detection and quantification limits of the instrument for remaining HCPs. Notwithstanding, the CV values applied also influenced the HCP peptides, resulting in a slight decrease in the average number of peptides per protein, from 1.8 without the FAIMS to 1.6 with the FAIMS. As a conclusion, the implementation of FAIMS in the workflow appears to be highly beneficial to allow digging deeper into the HCP landscape, with quantification of up to 158 protein impurities down to sub-ppm level after stringent validation filters.

3.3 | Digging deeper into the HCP landscape with data independent acquisition supplemented with gas phase fractionation

We [38, 39] and others [33, 66] have previously demonstrated the benefits of using DIA methods and MS²-based quantification to accurately monitor HCP impurities, and it is clearly confirmed in the current study. While DDA allowed the identification of the highest number of protein groups, namely 135, against 115 in DIA (Figure 1D), DIA allowed quantifying 100 HCPs (222 peptides) against 68 in DDA (121 peptides). This can be explained by the stochasticity of DDA causing missing values and CV filtering steps to remove more peptides and proteins. Contrary to DDA, the high reproducibility of DIA, with the systematic fragmentation of all peptides included in a defined mass range, significantly reduces the rate of missing values and the abundance variability. DIA increased by 45% and 83%, respectively, the number of protein groups and peptides quantified, compared to DDA. Consequently, an average of 2.2 peptides per protein was determined using DIA against 1.8 using DDA strengthening the confidence of the identifications. Fifty out of the 68 (74%) HCPs quantified in DDA are also quantified in DIA, representing a good overlap between both approaches (Table S3). The retrieved quantities of these 50 common proteins also attest the strong correlation between DDA and DIA, with a Pearson coefficient of 0.96. Overall, total HCP amounts retrieved by both acquisition modes are comparable, with a total HCP amount of 518 ± 20 ppm in DDA and 465 ± 28 ppm in DIA.

As an alternative to sample-fractionation to generate comprehensive spectral libraries, gas-phase fractionation DIA using chromatogram libraries can be applied with significantly reduced sample consumption and analysis time [46]. To evaluate this approach, we have performed six runs of 100 m/z mass range, from 480 to 1080 m/z, to generate a chromatogram library containing 844 precursors from 226 protein groups. Figure 1D displays the results achieved by processing DIA data previously acquired using this chromatogram library: 526 peptides and 172 protein groups, corresponding to a 59% and 50% increase respectively, compared to the DIA "spectrum-centric" data interpretation. After applying our stringent validation filtering, we quantified 119 HCP protein groups and 248 peptides. In comparison to library-free processing, chromatogram library DIA data processing enables to quantify 20% more protein groups. The total amount of quantified HCPs by GPF was 457 ± 20 ppm, a very similar value compared to the library-free DIA strategy. Furthermore, 88 HCPs were

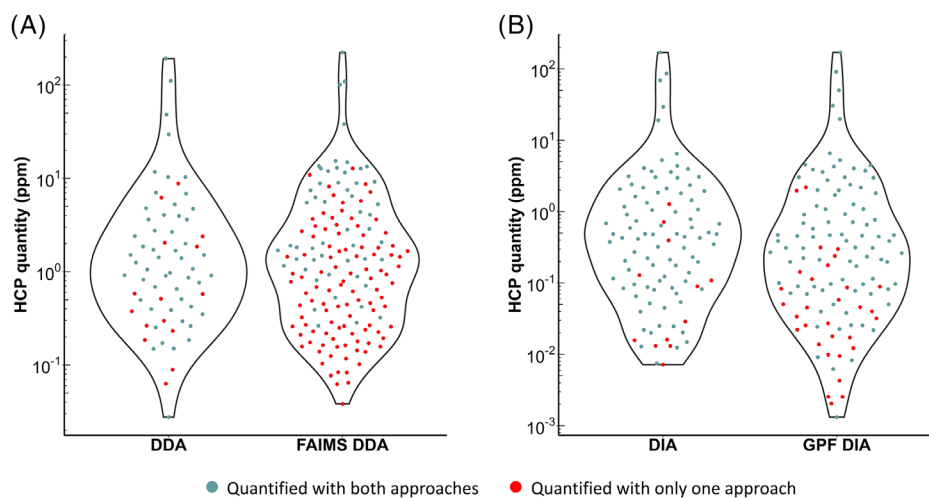


FIGURE 2 Violin plots representing the density of host cell proteins as a function of their concentration in ng of HCP per mg of NIST mAb (ppm). Blue dots represent proteins quantified by both approaches and red dots proteins quantified only in one approach, with (A) the comparison of DDA and FAIMS DDA and (B) the comparison of DIA and GPF DIA.

simultaneously identified and quantified using both DIA approaches, representing 88% and 73% of the proteins quantified using the library-free and chromatogram library based approaches, respectively. In addition to this high overlap, the Pearson coefficient of 0.99 also highlights the strong correlation between the two strategies. This result demonstrates that chromatogram library based approach increases the number of quantified proteins, without impairing the individual quantification of the proteins, as evidenced by the similar total HCP amounts reported with both methods. On top of that, we investigated the capabilities of the GPF DIA approach to quantify low-level proteins and digging deeper in the HCP landscape. By looking at the HCPs only quantified using the GPF DIA approach, we highlighted that these proteins are mostly under the ppm level, as depicted in Figure 2B. The shape of the violin plot clearly underlines that the density of HCPs is more important for low concentration levels using GPF DIA compared to DIA, mainly because of the presence of proteins quantified only with this approach. These elements clearly highlight that the use of GPF DIA allows increasing the number of identified and quantified HCPs, without the current drawbacks of generating exhaustive spectral library from DDA data after fractionation, requiring huge amounts of sample along with time-consuming sample preparation, and data acquisition of fractionated samples. In addition, FAIMS implementation in DIA can be performed, but data interpretation is still the limiting step. In summary, DIA-based strategies appear to improve classical DDA, by increasing the number of identified and quantified proteins and digging deeper into the HCP landscape.

3.4 | Combined coverage achieved on the NIST and application of the methods to FDA/EMA approved DPs

The merging of the results obtained using the four methods evaluated in previous sections (i.e., DDA, FAIMS DDA, DIA, and GPF DIA) allows

reaching a total of 278 HCP protein groups identified in the NIST mAb Reference Material, among which 80 are common to the four methods. The list of identified and quantified HCPs by each method is available in Table S2. We matched these identification results with those obtained in two recent papers showing high numbers of identified HCPs thanks to the combination of innovative technologies, including protein A depletion, native digestion, FAIMS, and 2h30 gradient in Johnson et al. [61] or ultralow trypsin concentration, long column and gradient (50 cm, 4 h) and BoxCar acquisition in Nie et al. (2021) [67]. We discovered 213 common HCPs with the literature (i.e., Johnson et al. and Nie et al.), representing 77% of the 278 proteins we identified in this work, and we uniquely identified 32 proteins (Figure S4).

The optimizations performed on the NIST mAb Reference Material demonstrate the enhanced performances achieved when implementing innovative methods such as FAIMS DDA or GPF DIA compared to classical DDA and DIA. However, the NIST mAb has not been designed for clinical use and its purification process is not at the level of the ones employed for FDA/EMA approved DPs. Therefore, we analyzed two high-value approved DPs, trastuzumab and nivolumab, with our advanced MS methods (Figure 3A and B, peptides results are presented in Figure S5). Overall, our results confirm the trends observed with the NIST mAb, with an increase of identified features using the FAIMS DDA and GPF DIA compared to DDA and DIA, respectively. Indeed, 67 HCPs in trastuzumab and 60 HCPs in nivolumab using DDA with FAIMS were detected, while only 37 and 18 HCP impurities could be reported without FAIMS. In terms of quantification, 31 and 29 protein groups in trastuzumab and nivolumab, respectively, were successfully quantified with our FAIMS DDA method. Thus, FAIMS hyphenation in DDA increased by threefold the number of quantified trace level proteins compared to DDA without FAIMS in trastuzumab (from 10 to 31). Similarly, the use of GPF quantified 4 more HCPs, from 8 in DIA to 12 in GPF DIA.

These few proteins quantified account for a total HCP amount between 6 and 57 ppm, depending on the methods. Briefly, quantities

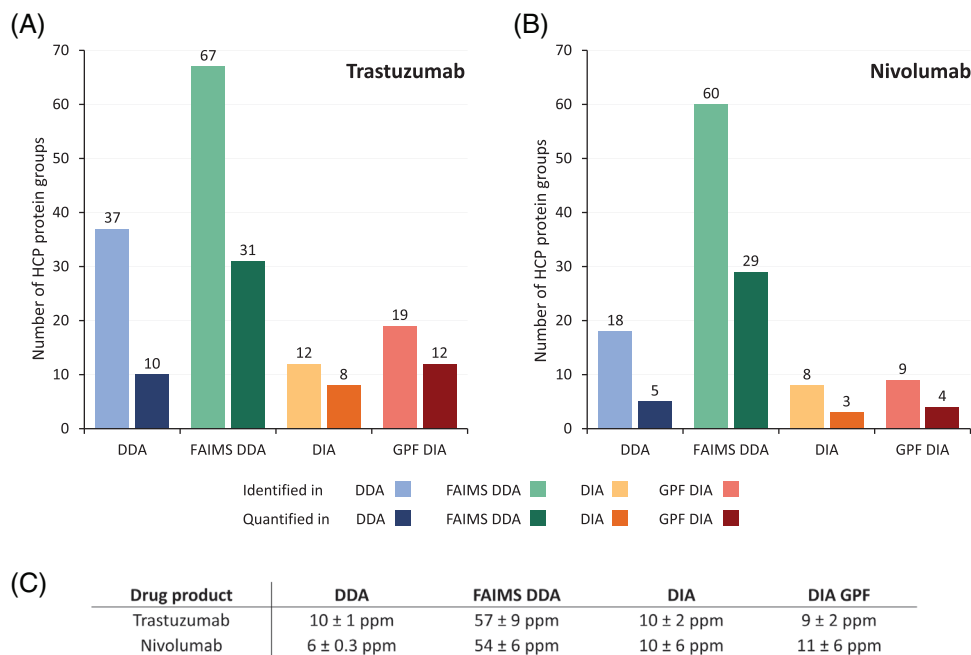


FIGURE 3 Number of HCP peptides identified (light colors) and quantified after applying validation filters (dark colors) using DDA (blue), FAIMS DDA (green), DIA (orange), and GPF DIA (red) (A) in trastuzumab and (B) in nivolumab. (C) Average quantity of HCPs quantified in nivolumab and trastuzumab drug products using DDA, DDA FAIMS, DIA, and DIA GPF.

are around 10 ppm in trastuzumab and nivolumab using DDA, DIA and GPF DIA methods, and up to 57 ppm using DDA FAIMS (Figure 3C, Table S4 and S5). In a previous study by Pythoud et al. [39], comparable numbers and quantities of HCPs were measured for nivolumab and trastuzumab in both DDA and DIA mode, with 3 and 10 HCPs quantified representing a total HCP amount between 20 and 80 ppm. Additionally, a study by Molden et al. [64] about HCPs profiling in various mAb-based therapeutic proteins, including nivolumab and trastuzumab, reported comparable results. Indeed, among the 28 therapeutic proteins analyzed, 7 HCPs on average were identified for an average total amount of 20 ppm in each DP. These findings are consistent with the current results, confirming the low number and abundance of HCP impurities in approved DPs and being in line with the current target of many biopharmaceuticals companies and authorities recommendations, with admitted quantities of HCPs in DP tolerated between 1 and 100 ppm [18]. Our work highlights the capabilities of LC-MS/MS to quantify routinely and accurately individual HCPs in challenging FDA/EMA approved DPs demonstrating trace-level impurities. Based on their individual identification and quantification, investigations on the potential immunogenic effect of the monitored HCPs can be undertaken.

4 | CONCLUDING REMARKS

To summarize, we compared LC-MS/MS acquisition methods to characterize and monitor HCPs in DPs and achieved individual identification and quantification of hundreds of these impurities. First, using DDA analysis on the NIST mAb Reference Material, we proved the extreme

necessity to apply stringent validation filtering steps in order to precisely and accurately quantify individual trace-level HCPs, by removing low-confidence and non-reproducible signals. Then, we evaluated optimized LC-MS/MS acquisition methods including DDA with and without FAIMS implementation, DIA, and gas-phase fractionation DIA. The benchmarking of DIA against DDA confirmed its benefits for HCPs monitoring in DP, with an increase of 47% in the number of impurities quantified. The addition of the chromatogram library-based approach for DIA data analysis also revealed very promising results, allowing quantification of 119 HCPs. Finally, the FAIMS implementation in DDA provided enhanced capabilities with the quantification of 158 protein impurities for a total amount of 880 ppm. Finally, we applied the aforementioned methods to trastuzumab and nivolumab, two FDA/EMA approved DPs known and presumably highly-pure products. FAIMS implementation to DDA reveals to be a powerful approach, with the quantification of around 30 HCPs for a total amount around 55 ppm in both trastuzumab and nivolumab. This work demonstrated that LC-MS/MS based approaches followed by stringent validation filters can be employed for precise individual HCP impurities characterization in very challenging DPs.

ACKNOWLEDGMENTS

This work was supported by the “Agence Nationale de la Recherche” via the French Proteomic Infrastructure (ProFI FR2048; ANR-10-INBS-08-03) and by the French Ministry of Higher Education and Research for the PhD fellowship of C.B.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

Additional supporting information may be found online in the Supporting Information section at the end of the article. The dataset was deposited with the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD039582 (for DIA data) and PXD039585 (for DDA data) [68].

ORCID

Corentin Beaumal  <https://orcid.org/0000-0003-4231-780X>

REFERENCES

- Reichert, J. M. (2012). Marketed therapeutic antibodies compendium. *mAbs*, 4(3), 413–415. <https://doi.org/10.4161/mabs.19931>
- Todd, P. A., & Brogden, R. N. (1989). Muromonab CD3. *Drugs*, 37(6), 871–899. <https://doi.org/10.2165/00003495-198937060-00004>
- Emmons, C., & Hunsicker, L. G. (1987). Muromonab-CD3 (Orthoclone OKT3): The first monoclonal antibody approved for therapeutic use. *Iowa Medicine*, 77(2), 78–82.
- The Antibody Society. Therapeutic monoclonal antibodies approved or in review in the EU or US. Retrieved from www.antibodysociety.org/resources/approved-antibodies
- Kaplon, H., Crescioli, S., Chenoweth, A., Visweswaraiiah, J., & Reichert, J. M. (2023). Antibodies to watch in 2023. *mAbs*, 15(1), 2153410. <https://doi.org/10.1080/19420862.2022.2153410>
- Liu, H. F., Ma, J., Winter, C., & Bayer, R. (2010). Recovery and purification process development for monoclonal antibody production. *mAbs*, 2(5), 480–499. <https://doi.org/10.4161/mabs.2.5.12645>
- Walker, D. E., Yang, F., Carver, J., Joe, K., Michels, D. A., & Yu, X. C. (2017). A modular and adaptive mass spectrometry-based platform for support of bioprocess development toward optimal host cell protein clearance. *mAbs*, 9(4), 654–663. <https://doi.org/10.1080/19420862.2017.1303023>
- Bierau, H., Rossi, M., Agugiario, D., Soranzo, T., Broly, H., & Mitchell-Logean, C. (2009). Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of a CHO cathepsin D protease. *Biotechnology and Bioengineering*, 104(6), 1132–1141. <https://doi.org/10.1002/bit.22494>
- Dixit, N., Salamat-Miller, N., Salinas, P. A., Taylor, K. D., & Basu, S. K. (2016). Residual host cell protein promotes polysorbate 20 degradation in a sulfatase drug product leading to free fatty acid particles. *Journal of Pharmaceutical Sciences*, 105(5), 1657–1666. <https://doi.org/10.1016/j.xphs.2016.02.029>
- Vanderlaan, M., Zhu-Shimoni, J., Lin, S., Gunawan, F., Waerner, T., & Van Cott, K. E. (2018). Experience with host cell protein impurities in biopharmaceuticals. *Biotechnology Progress*, 34(4), 828–837. <https://doi.org/10.1002/btpr.2640>
- Li, X., An, Y., Liao, J., Xiao, L., Swanson, M., Martinez-Fonts, K., Pavon, J. A., Sherer, E. C., Jawa, V., Wang, F., Gao, X., Letarte, S., & Richardson, D. D. (2021). Identification and characterization of a residual host cell protein hexosaminidase B associated with N-glycan degradation during the stability study of a therapeutic recombinant monoclonal antibody product. *Biotechnology Progress*, 37(3), e3128. <https://doi.org/10.1002/btpr.3128>
- Jawa, V., Joubert, M. K., Zhang, Q., Deshpande, M., Hapuarachchi, S., Hall, M. P., & Flynn, G. C. (2016). Evaluating immunogenicity risk due to host cell protein impurities in antibody-based biotherapeutics. *The Aaps Journal [Electronic Resource]*, 18(6), 1439–1452. <https://doi.org/10.1208/s12248-016-9948-4>
- Fischer, S. K., Cheu, M., Peng, K., Lowe, J., Araujo, J., Murray, E., McClintock, D., Matthews, J., Siguenza, P., & Song, A. (2017). Specific immune response to phospholipase B-Like 2 protein, a host cell impurity in lebrizumab clinical material. *The Aaps Journal [Electronic Resource]*, 19(1), 254–263. <https://doi.org/10.1208/s12248-016-9998-7>
- Jones, M., Palackal, N., Wang, F., Gaza-Bulsecu, 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. (2021). High-risk host cell proteins (HCPs): A multi-company collaborative view. *Biotechnology and Bioengineering*, 118(8), 2870–2885. <https://doi.org/10.1002/bit.27808>
- Guidance for Industry Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products (1999).
- <1132>Residual Host Cell Protein Measurement in Biopharmaceuticals (2016).
- DNA and host cell protein impurities, routine testing versus validation studies (1997). (CPMP/BWP/382/97).
- Gilgunn, S., El-Sabbahy, H., Albrecht, S., Gaikwad, M., Corrigan, K., Deakin, L., Jellum, G., & Bones, J. (2019). Identification and tracking of problematic host cell proteins removed by a synthetic, highly functionalized nonwoven media in downstream bioprocessing of monoclonal antibodies. *Journal of Chromatography A*, 1595, 28–38. <https://doi.org/10.1016/j.chroma.2019.02.056>
- Wang, X., Hunter, A. K., & Mozier, N. M. (2009). Host cell proteins in biologics development: Identification, quantitation and risk assessment. *Biotechnology and Bioengineering*, 103(3), 446–458. <https://doi.org/10.1002/bit.22304>
- Champion, K. M., Madden, H., Dougherty, J., & Shacter, E. (2005). Defining your product profile and maintaining control over it, Part 2. Challenges of monitoring host cell protein impurities. *Bioprocess International*.
- Chon, J. H., & Zarbis-Papastoitis, G. (2011). Advances in the production and downstream processing of antibodies. *New Biotechnology*, 28(5), 458–463. <https://doi.org/10.1016/j.nbt.2011.03.015>
- Zhu-Shimoni, J., Yu, C., Nishihara, J., Wong, R. M., Gunawan, F., Lin, M., & Vanderlaan, M. (2014). Host cell protein testing by ELISAs and the use of orthogonal methods. *Biotechnology and Bioengineering*, 111(12), 2367–2379. <https://doi.org/10.1002/bit.25327>
- Tscheliessnig, A. L., Konrath, J., Bates, R., & Jungbauer, A. (2013). Host cell protein analysis in therapeutic protein bioprocessing—methods and applications. *Biotechnology Journal*, 8(6), 655–670. <https://doi.org/10.1002/biot.201200018>
- Bracewell, D. G., Francis, R., & Smales, C. M. (2015). The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnology and Bioengineering*, 112(9), 1727–1737. <https://doi.org/10.1002/bit.25628>
- Pilely, K., Johansen, M. R., Lund, R. R., Kofoed, T., Jørgensen, T. K., Skriver, L., & Mørtz, E. (2021). Monitoring process-related impurities in biologics-host cell protein analysis. *Analytical and Bioanalytical Chemistry*, 1–12. <https://doi.org/10.1007/s00216-021-03648-2>
- European Pharmacopeia Monograph 2.6.34 Host-Cell Protein Assays. 9. (2017). *European Pharmacopeia Monograph 2.6.34 Host-Cell Protein Assays*, 9, 2017.
- Valente, K. N., Levy, N. E., Lee, K. H., & Lenhoff, A. M. (2018). Applications of proteomic methods for CHO host cell protein characterization in biopharmaceutical manufacturing. *Current Opinion in Biotechnology*, 53, 144–150. <https://doi.org/10.1016/j.copbio.2018.01.004>
- Hogwood, C. E., Bracewell, D. G., & Smales, C. M. (2014). Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. *Current Opinion in Biotechnology*, 30, 153–160. <https://doi.org/10.1016/j.copbio.2014.06.017>
- Chen, Y., Xu, C. F., Stanley, B., Evangelist, G., Brinkmann, A., Liu, S., & Yeung, B. (2021). A highly sensitive LC-MS/MS method for targeted quantitation of lipase host cell proteins in biotherapeutics. *Journal of Pharmaceutical Sciences*, 110(12), 3811–3818. <https://doi.org/10.1016/j.xphs.2021.08.024>

30. Doneanu, C., Fang, J., Alelyunas, Y., Yu, Y. Q., Wrona, M., & Chen, W. (2018). An HS-MRM assay for the quantification of host-cell proteins in protein biopharmaceuticals by liquid chromatography ion mobility QTOF mass spectrometry. *Journal of Visualized Experiments: JoVE*, 17(134), 55325. <https://doi.org/10.3791/55325>
31. Doneanu, C. E., Xenopoulos, A., Fadgen, K., Murphy, J., Skilton, S. J., Prentice, H., & Chen, W. (2012). Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *mAbs*, 4(1), 24–44. <https://doi.org/10.4161/mabs.4.1.18748>
32. Gao, X., Rawal, B., Wang, Y., Li, X., Wylie, D., Liu, Y. H., & Richardson, D. D. (2020). Targeted host cell protein quantification by LC-MRM enables biologics processing and product characterization. *Analytical Chemistry*, 92(1), 1007–1015. <https://doi.org/10.1021/acs.analchem.9b03952>
33. Kreimer, S., Gao, Y., Ray, S., Jin, M., Tan, Z., Mussa, N. A., & Karger, B. L. (2017). Host cell protein profiling by targeted and untargeted analysis of data independent acquisition mass spectrometry data with parallel reaction monitoring verification. *Analytical Chemistry*, 89(10), 5294–5302. <https://doi.org/10.1021/acs.analchem.6b04892>
34. E, S. Y., Hu, Y., Molden, R., Qiu, H., & Li, N. (2022). Identification and quantification of a problematic host cell protein to support therapeutic protein development. *Journal of Pharmaceutical Sciences*, 112(3), 673–679. <https://doi.org/10.1016/j.xphs.2022.10.008>
35. Krawitz, D. C., Forrest, W., Moreno, G. T., Kittleson, J., & Champion, K. M. (2006). Proteomic studies support the use of multi-product immunoassays to monitor host cell protein impurities. *Proteomics*, 6(1), 94–110. <https://doi.org/10.1002/pmic.200500225>
36. Ma, J., & Kilby, G. W. (2020). Sensitive, rapid, robust, and reproducible workflow for host cell protein profiling in biopharmaceutical process development. *Journal of Proteome Research*, 19(8), 3396–3404. <https://doi.org/10.1021/acs.jproteome.0c00252>
37. Yang, F., Walker, D. E., Schoenfelder, J., Carver, J., Zhang, A., Li, D., & Michels, D. A. (2018). A 2D LC-MS/MS strategy for reliable detection of 10-ppm level residual host cell proteins in therapeutic antibodies. *Analytical Chemistry*, 90(22), 13365–13372. <https://doi.org/10.1021/acs.analchem.8b03044>
38. Husson, G., Delangle, A., O'Hara, J., Cianferani, S., Gervais, A., Van Dorselaer, A., & Carapito, C. (2018). Dual data-independent acquisition approach combining global HCP profiling and absolute quantification of key impurities during bioprocess development. *Analytical Chemistry*, 90(2), 1241–1247. <https://doi.org/10.1021/acs.analchem.7b03965>
39. Pythoud, N., Bons, J., Mijola, G., Beck, A., Cianferani, S., & Carapito, C. (2021). Optimized sample preparation and data processing of data-independent acquisition methods for the robust quantification of trace-level host cell protein impurities in antibody drug products. *Journal of Proteome Research*, 20(1), 923–931. <https://doi.org/10.1021/acs.jproteome.0c00664>
40. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., & Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics*, 11(6), 016717. <https://doi.org/10.1074/mcp.O111.016717>
41. Zhang, F., Ge, W., Ruan, G., Cai, X., & Guo, T. (2020). Data-independent acquisition mass spectrometry-based proteomics and software tools: A glimpse in 2020. *Proteomics*, 20(17–18), e1900276. <https://doi.org/10.1002/pmic.201900276>
42. Barkovits, K., Pacharra, S., Pfeiffer, K., Steinbach, S., Eisenacher, M., Marcus, K., & Uszkoreit, J. (2020). Reproducibility, specificity and accuracy of relative quantification using spectral library-based data-independent acquisition. *Molecular & Cellular Proteomics*, 19(1), 181–197. <https://doi.org/10.1074/mcp.RA119.001714>
43. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., & Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular Systems Biology*, 14(8), e8126. doi: [10.15252/msb.20178126](https://doi.org/10.15252/msb.20178126)
44. Cai, X., Ge, W., Yi, X., Sun, R., Zhu, J., Lu, C., & Guo, T. (2021). PulseDIA: Data-independent acquisition mass spectrometry using multi-injection pulsed gas-phase fractionation. *Journal of Proteome Research*, 20(1), 279–288. <https://doi.org/10.1021/acs.jproteome.0c00381>
45. Amodei, D., Egertson, J., MacLean, B. X., Johnson, R., Merrihew, G. E., Keller, A., & MacCoss, M. J. (2019). Improving precursor selectivity in data-independent acquisition using overlapping windows. *Journal of the American Society for Mass Spectrometry*, 30(4), 669–684. <https://doi.org/10.1007/s13361-018-2122-8>
46. Pino, L. K., Just, S. C., MacCoss, M. J., & Searle, B. C. (2020). Acquiring and analyzing data independent acquisition proteomics experiments without spectrum libraries. *Molecular & Cellular Proteomics*, 19(7), 1088–1103. <https://doi.org/10.1074/mcp.P119.001913>
47. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., & Ralser, M. (2020). DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods*, 17(1), 41–44. <https://doi.org/10.1038/s41592-019-0638-x>
48. Gotti, C., Roux-Dalvai, F., Joly-Beauparlant, C., Mangnier, L., Leclercq, M., & Droit, A. (2021). Extensive and accurate benchmarking of DIA acquisition methods and software tools using a complex proteomic standard. *Journal of Proteome Research*, 20(10), 4801–4814. <https://doi.org/10.1021/acs.jproteome.1c00490>
49. Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A. C., & Nesvizhskii, A. I. (2015). DIA-Umpire: Comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods*, 12(3), 258–264. 257 p following 264. <https://doi.org/10.1038/nmeth.3255>
50. van der Spek, S. J. F., Gonzalez-Lozano, M. A., Koopmans, F., Miedema, S. S. M., Palukhovich, I., Smit, A. B., & Li, K. W. (2021). Age-dependent hippocampal proteomics in the APP/PS1 Alzheimer Mouse Model: A comparative analysis with classical SWATH/DIA and directDIA Approaches. *Cells*, 10(7). <https://doi.org/10.3390/cells10071588>
51. Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., & Wilhelm, M. (2019). Prosit: Proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature Methods*, 16(6), 509–518. <https://doi.org/10.1038/s41592-019-0426-7>
52. Searle, B. C., Swearingen, K. E., Barnes, C. A., Schmidt, T., Gessulat, S., Küster, B., & Wilhelm, M. (2020). Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nature Communications*, 11(1), 1548. <https://doi.org/10.1038/s41467-020-15346-1>
53. Van Puyvelde, B., Willems, S., Gabriels, R., Daled, S., De Clerck, L., Vande Castele, S., & Dhaenens, M. (2020). Removing the hidden data dependency of DIA with predicted spectral libraries. *Proteomics*, 20(3–4), e1900306. <https://doi.org/10.1002/pmic.201900306>
54. Huang, L., Wang, N., Mitchell, C. E., Brownlee, T., Maple, S. R., & De Felippis, M. R. (2017). A novel sample preparation for shotgun proteomics characterization of HCPs in antibodies. *Analytical Chemistry*, 89(10), 5436–5444. <https://doi.org/10.1021/acs.analchem.7b00304>
55. Chen, I. H., Xiao, H., Daly, T., & Li, N. (2020). Improved host cell protein analysis in monoclonal antibody products through molecular weight cutoff enrichment. *Analytical Chemistry*, 92(5), 3751–3757. <https://doi.org/10.1021/acs.analchem.9b05081>
56. Madsen, J. A., Farutin, V., Carbeau, T., Wudyka, S., Yin, Y., Smith, S., & Capila, I. (2015). Toward the complete characterization of host cell proteins in biotherapeutics via affinity depletions, LC-MS/MS, and multivariate analysis. *mAbs*, 7(6), 1128–1137. <https://doi.org/10.1080/19420862.2015.1082017>
57. Thompson, J. H., Chung, W. K., Zhu, M., Tie, L., Lu, Y., Aboulaich, N., & Mo, W. D. (2014). Improved detection of host cell proteins (HCPs) in a mammalian cell-derived antibody drug using liquid chromatography/mass spectrometry in conjunction with an HCP-enrichment

- strategy. *Rapid Communications in Mass Spectrometry*, 28(8), 855–860. <https://doi.org/10.1002/rcm.6854>
58. Wang, Q., Slaney, T. R., Wu, W., Ludwig, R., Tao, L., & Leone, A. (2020). Enhancing host-cell protein detection in protein therapeutics using HILIC enrichment and proteomic analysis. *Analytical Chemistry*, 92(15), 10327–10335. <https://doi.org/10.1021/acs.analchem.0c00360>
59. Doneanu, C. E., Anderson, M., Williams, B. J., Lauber, M. A., Chakraborty, A., & Chen, W. (2015). 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. *Analytical Chemistry*, 87(20), 10283–10291. <https://doi.org/10.1021/acs.analchem.5b02103>
60. Farrell, A., Mittermayr, S., Morrissey, B., Mc Loughlin, N., Navas Iglesias, N., Marison, I. W., & Bones, J. (2015). Quantitative host cell protein analysis using two dimensional data independent LC-MS(E). *Analytical Chemistry*, 87(18), 9186–9193. <https://doi.org/10.1021/acs.analchem.5b01377>
61. Johnson, R. O., Greer, T., Cejkov, M., Zheng, X., & Li, N. (2020). Combination of FAIMS, Protein A depletion, and native digest conditions enables deep proteomic profiling of host cell proteins in monoclonal antibodies. *Analytical Chemistry*, 92(15), 10478–10484. <https://doi.org/10.1021/acs.analchem.0c01175>
62. Hebert, A. S., Prasad, S., Belford, M. W., Bailey, D. J., McAlister, G. C., Abbatiello, S. E., & Coon, J. J. (2018). Comprehensive single-shot proteomics with FAIMS on a hybrid orbitrap mass spectrometer. *Analytical Chemistry*, 90(15), 9529–9537. <https://doi.org/10.1021/acs.analchem.8b02233>
63. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P., & Geromanos, S. J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & Cellular Proteomics*, 5(1), 144–156. <https://doi.org/10.1074/mcp.M500230-MCP200>
64. Molden, R., Hu, M., Yen, E. S., Saggese, D., Reilly, J., Mattila, J., & Li, N. (2021). Host cell protein profiling of commercial therapeutic protein drugs as a benchmark for monoclonal antibody-based therapeutic protein development. *mAbs*, 13(1), 1955811. <https://doi.org/10.1080/19420862.2021.1955811>
65. Trauchessec, M., Hesse, A. M., Kraut, A., Berard, Y., Herment, L., Fortin, T., & Manin, C. (2021). An innovative standard for LC-MS-based HCP profiling and accurate quantity assessment: Application to batch consistency in viral vaccine samples. *Proteomics*, 21(5), e2000152. <https://doi.org/10.1002/pmic.202000152>
66. Strasser, L., Oliviero, G., Jakes, C., Zaborowska, I., Floris, P., Ribeiro da Silva, M., & Bones, J. (2021). Detection and quantitation of host cell proteins in monoclonal antibody drug products using automated sample preparation and data-independent acquisition LC-MS/MS. *Journal of Pharmaceutical Analysis*, 11(6), 726–731. <https://doi.org/10.1016/j.jpha.2021.05.002>
67. Nie, S., Greer, T., O'Brien Johnson, R., Zheng, X., Torri, A., & Li, N. (2021). Simple and sensitive method for deep profiling of host cell proteins in therapeutic antibodies by combining ultra-low trypsin concentration digestion, long chromatographic gradients, and BoxCar mass spectrometry acquisition. *Analytical Chemistry*, 93(10), 4383–4390. <https://doi.org/10.1021/acs.analchem.0c03931>
68. Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., & Vizcaino, J. A. (2017). The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition. *Nucleic Acids Research*, 45(D1), D1100–d1106. <https://doi.org/10.1093/nar/gkw936>

SUPPORTING INFORMATION

Additional supporting information may be found online <https://doi.org/10.1002/pmic.202300172> in the Supporting Information section at the end of the article.

How to cite this article: Beaumal, C., Beck, A., Hernandez-Alba, O., & Carapito, C. (2023). 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*, e2300172. <https://doi.org/10.1002/pmic.202300172>



Improved characterization of trastuzumab deruxtecan with PTCR and internal fragments implemented in middle-down MS workflows

Corentin Beaumal^{1,2} · Evolène Deslignière^{1,2} · H  l  ne Diemer^{1,2} · Christine Carapito^{1,2} · Sarah Cianf  rani^{1,2} · Oscar Hernandez-Alba^{1,2}

Received: 12 September 2023 / Revised: 13 November 2023 / Accepted: 14 November 2023
  The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2023

Abstract

Antibody-drug conjugates (ADCs) are highly complex proteins mainly due to the structural microvariability of the mAb, along with the additional heterogeneity afforded by the bioconjugation process. Top-down (TD) and middle-down (MD) strategies allow the straightforward fragmentation of proteins to elucidate the conjugated amino acid residues. Nevertheless, these spectra are very crowded with multiple overlapping and unassigned ion fragments. Here we report on the use of dedicated software (ClipsMS) and application of proton transfer charge reduction (PTCR), to respectively expand the fragment ion search space to internal fragments and improve the separation of overlapping fragment ions for a more comprehensive characterization of a recently approved ADC, trastuzumab deruxtecan (T-DXd). Subunit fragmentation allowed between 70 and 90% of sequence coverage to be obtained. Upon addition of internal fragment assignment, the three subunits were fully sequenced, although internal fragments did not contribute significantly to the localization of the payloads. Finally, the use of PTCR after subunit fragmentation provided a moderate sequence coverage increase between 2 and 13%. The reaction efficiently decluttered the fragmentation spectra allowing increasing the number of fragment ions characteristic of the conjugation site by 1.5- to 2.5-fold. Altogether, these results show the interest in the implementation of internal fragment ion searches and more particularly the use of PTCR reactions to increase the number of signature ions to elucidate the conjugation sites and enhance the overall sequence coverage of ADCs, making this approach particularly appealing for its implementation in R&D laboratories.

Keywords Trastuzumab deruxtecan · Middle-down (MD) · Electron transfer dissociation (ETD) · Proton transfer charge reduction (PTCR) · Internal fragments

Abbreviations

ADC	Antibody-drug conjugate	Hc	Heavy chain
AGC	Automatic gain control	Lc	Light chain
avDAR	Average drug-to-antibody ratio	LC	Liquid chromatography
CID	Collision-induced dissociation	mAb	Monoclonal antibody
ETD	Electron transfer dissociation	MS	Mass spectrometry
EThcD	Electron transfer/higher-energy collision dissociation	nMS	Native mass spectrometry
Fc/2	Half crystallizable fragment	PTCR	Proton transfer charge reduction
HCD	Higher-energy collision dissociation	PTM	Post-translational modification
		RPLC	Reverse-phase liquid chromatography
		SEC	Size exclusion chromatography
		T-DXd	Trastuzumab deruxtecan
		TFA	Trifluoroacetic acid
		TD/MD-MS	Top-down/middle-down mass spectrometry

  Oscar Hernandez-Alba
ahernandez@unistra.fr

¹ Laboratoire de Spectrom  trie de Masse Bio Organique, IPHC UMR 7178, CNRS, Universit   de Strasbourg, 67087 Strasbourg, France

² Infrastructure Nationale de Prot  omique ProFI - FR2048, Strasbourg, France

Introduction

From the first approval until today, monoclonal antibodies (mAbs) have experienced a continuous evolution that has changed the treatment of an array of diseases. This kind of therapeutics has proven its efficiency not only in oncology but also in immune-mediated disorders, infectious diseases, and cardiovascular disorders, among others, showcasing the therapeutic versatility of these proteins [1]. Currently, more than 170 mAbs are approved or under review by regulatory agencies, and more than 1200 mAbs are estimated to be in clinical trials [2]. Concomitantly, this outstanding evolution has motivated the development of other mAb-related therapeutic proteins, among which antibody-drug conjugates (ADCs)—constituted of a mAb, a linker, and a payload—are one of the most prominent formats [3, 4]. Different classes of conjugated mAbs have been reported with different bioconjugation strategies (cysteine [5], lysine [6], site [7], engineered [8], and unnatural [9] residue-specific ADCs), and different types of cargo molecules such as tubulin inhibitors, DNA damaging agents, and immunomodulators [10]. The size (around 150 kDa), the intrinsic primary sequence micro-variability of mAbs (with tens of post-translational modifications, PTMs) [11], and the additional complexity arising from the bioconjugation process of different payloads make the characterization of these molecules particularly challenging.

In this context, native mass spectrometry (nMS), either as a standalone technique or in combination with liquid chromatography (LC), has significantly contributed to the characterization of ADCs of increased complexity [12–14]. However, the conjugation location afforded by these techniques is limited to the subunit level, and the combination of complementary MS-based methods is crucial to determining modified amino acid residues. Top-down and middle-down mass spectrometry (TD/MD-MS) have emerged as complementary approaches consisting of the fragmentation of the proteins at the intact level (TD-MS) or the subunit level after a limited proteolysis step (MD-MS) with several activation techniques [15], including electron-based dissociations (ExD), collision-induced dissociation (CID), electron-transfer/higher-energy dissociation (ET_hcD), and ultraviolet photodissociation (UVPD). This combination allows correlating the information derived from the fragmentation spectrum with the intact mass of the protein, leading to a better characterization of protein proteoforms while limiting the introduction of primary sequence artifacts during sample preparation [16]. Outstanding outcomes have been reported for an array of proteins of different sizes and complexity, providing high sequence coverages (up

to 90%) while determining the position of relevant PTMs [17, 18]. Different TD/MD-MS strategies have also been applied to mAbs [19–24] and even ADCs [25–28] using different instruments and fragmentation techniques in order to provide a thorough characterization of the primary structure. Furthermore, numerous fragment ions bearing the intact scaffold of the glycoforms and payloads are detected, allowing a precise location of these modifications that might be fragmented in the bottom-up analysis [25, 28, 29].

One of the main concerns of TD/MD-MS is the extensive exploitation of the huge amount of information generated during the fragmentation of the intact proteins. Hence, several strategies have been developed to overcome this issue, including the design of experimental workflows aiming at decluttering the highly congested fragmentation spectra [30–34] or the development of tailored bioinformatics solutions [35, 36] to identify better and assign all the different fragment ions. Proton transfer charge reduction (PTCR), an ion/ion reaction developed in the 1990s by McLuckey and coworkers [37], has been designed to reduce the charge state of gas-phase ions and spread the MS signals to a broader m/z range. More particularly, the ions of interest undergo reaction with an anion (usually a perfluorinated compound) to lower the charge state of the ions while the anionic reactant is neutralized. Several applications of this reaction have been recently described to improve the identification of large proteoforms [32] (more than 30 kDa) or to enhance the sequence coverage of reference proteins [34] and histones [31] in TD-MS workflows. In the latter examples, the implementation of PTCR has been reported to reduce the spectral congestion in fragmentation spectra, improving the signal-to-noise ratio of fragment ions and thus providing between 5 and 20% of additional sequence coverage depending on the protein of interest and the activation technique used for sequencing [31, 33, 34]. Although quite promising, this technique has been reported in a limited number of studies and is scarcely implemented in the case of mAb-based therapeutic proteins. An additional way to enhance TD-MS protein sequencing is to take into account internal fragments [38, 39]. Despite the formation of internal fragments (secondary fragmentation of protein backbone) during TD/MD-MS analysis of intact proteins [40–43], depending on the protein size and fragmentation type, in most studies, only terminal fragment ions (containing either N- or C-terminal sides) are considered for TD/MD-MS data interpretation. Therefore, several bioinformatics tools have been developed to boost TD/MD-MS data interpretation by taking into account internal fragment ions [35, 36]. When applying these tools, not only drastic sequence coverage increase is observed [33, 38, 42, 43] (in some cases more than 40–50% of protein sequence coverage), but also more accurate determination of the specific position of modifications located in the

internal region of proteins [39, 44]. However, some previous publications reported on the increased risk of false positive fragment ion assignment when internal fragment ions are included. This risk is even more marked when several types of fragment ions are considered during ion searching [41], or when complex fragmentation techniques are used to induce the fragmentation of the backbone (such as UVPD [45]). For all those reasons, the use of internal fragments to enhance the characterization of the primary structure of proteins should be considered with caution to increase the reliance on fragment ion assignment [45].

Here, we report on the thorough characterization of a recently approved therapeutic site-specific ADC, trastuzumab deruxtecan (T-DXd), using a combination of intact MS approaches with cutting-edge MD-MS techniques encompassing the application of PTCR and internal fragment analysis. This ADC is composed of the humanized anti-HER2 trastuzumab antibody with a cleavable peptidyl-based linker (GGFG) and a potent topoisomerase I inhibitor payload (DXd) [46, 47]. The cysteines involved in the inter-chain disulfide bridges of the mAb are targeted by the bioconjugation strategy to produce a homogeneous average drug-to-antibody ratio (avDAR) of 8 [13, 46]. The fragmentation of the ADC subunits led to the characterization of the conjugation sites with sequence coverages ranging from 70 to 90%. Both conjugation sites and sequence coverage were significantly improved with the addition of internal fragment searching and PTCR reaction, increasing the number of conjugation-diagnostic fragment ions by 30–40%, along with the full sequence coverage of each T-DXd subunit. Altogether, our results pinpoint the synergy between intact MS analyses and cutting-edge MD-MS strategies to offer an in-depth characterization of a homogeneous ADC without extensive sample preparation.

Materials and methods

Materials

Trastuzumab deruxtecan (T-DXd) samples were provided by the Institut de Cancérologie Strasbourg Europe (ICANS, Strasbourg, France). All chromatographic solvents are LC-MS grade.

Middle-up analysis sample preparation

For middle-up level experiments, IdeS digestion was performed by incubating one unit of FabRICATOR enzyme (Genovis, Lund, Sweden) per microgram of ADC for 30 min at 37 °C. For denaturing conditions analysis, trifluoroacetic acid (TFA) was added to the digested ADC to a final

concentration of 1%. The final concentration of the ADC in solution was 0.8 µg/µL.

Middle-down analysis sample preparation

For middle-down experiments, sample IdeS digestion was performed as mentioned for middle-up experiments. Disulfide bonds were subsequently reduced during 60 min at 37 °C in strong denaturing conditions (6 M guanidine hydrochloride) using DTT (Sigma) as a reducing agent (100 mM final DTT concentration). Finally, TFA was added to the solution at a final concentration of 1%. The final concentration of the ADC in solution was 0.7 µg/µL.

Bottom-up peptide mapping analysis

Sample preparation Ten micrograms of untreated T-DXd was solubilized in 0.1% RapiGest (Waters). Disulfide reduction was performed with 5 mM TCEP for 30 min at 57 °C. Ten millimolar of iodoacetamide for 40 min in the dark was added to alkylate-free thiol groups on cysteine residues and prevent the reformation of disulfide bridges. Digestion was performed by adding 5 µL of the prior prepared trypsin solution, i.e., 20 µg of trypsin (Promega) suspended in 100 µL of H₂O which corresponds to 1:50 enzyme to substrate ratio, at 37 °C for 5 h. The reaction was stopped by adding 1 µL of TFA. RapiGest was eliminated by incubation for 30 min at 37 °C and centrifugation at 13,000 g for 10 min.

LC-MS/MS analysis NanoLC-MS/MS analyses were performed on a Dionex UltiMate 3000 RSLC nanosystem (Thermo Scientific) coupled to an Orbitrap Eclipse™ Tribrid™ (Thermo Fisher Scientific) mass spectrometer. Mobile phase A contained 0.1% FA in water and mobile phase B contained 0.1% FA in 80% ACN/20% water. Peptides were loaded onto an Acclaim PepMap 100 C18 20 mm × 0.1 mm, 5-µm-diameter particles precolumn (Thermo Scientific) for 3 min at 10 µL/min and eluted on an Aurora Series C18 UHPLC (250 mm × 75 µm, 1.6-µm-diameter particles, IonOpticks) at a 300 nL/min flow rate following a linear gradient: 2.5% B at 0 min, 2.5% at 3 min, 7.5% at 6 min, 50% B at 43 min, and 98% B from 44 to 48 min. The column was finally re-equilibrated in 2.5% B for 15 min. Full-scan MS spectra were acquired over a 300–1800 m/z range at a resolution of 120,000 (at 200 m/z), with an automatic gain control (AGC) target of 1×10^6 and a maximum injection time of 50 ms. The top 10 most intense precursor ions with an intensity exceeding 1.10^4 ions per second and charge states between 2 and 7 were automatically selected from each MS spectrum for higher-energy collisional dissociation (HCD) fragmentation at 30% normalized collision energy. MS/MS spectra were collected in the Orbitrap at a resolution of 15,000, with an AGC target of 1×10^5 and a maximum

injection time set to 22 ms. The complete system was fully controlled by Thermo Scientific™ Xcalibur™ software. Raw data collected were processed and converted with MSConvert into.mgf peak list format.

Data processing Search engine MASCOT 2.6.2 algorithm (Matrix Science) was used. The search was performed against the sequence of the light and heavy chains of the ADC. Spectra were searched with a mass tolerance of 10 ppm for MS and 0.05 Da for MS/MS data. The search was made without enzyme specified, in order to allow the identification of any non-specific cleavage peptide. Variable modifications were specified: oxidation of methionine residues, pyro-glutamylolation of the N-termini, deamidation of asparagine, isomerization of aspartic acid residues, and drug-linker conjugation ($C_{52}FH_{56}N_9O_{13}$ and $C_{26}H_{32}O_8N_6$) on cysteine residues. Peptide identifications were validated with a minimal Mascot ion score of 30 using Proline Studio.

SEC-nMS analysis experiments

A Dionex UltiMate 3000 RSLC system (Thermo Scientific) was coupled to an Orbitrap Eclipse™ Tribrid™ (Thermo Fisher Scientific) mass spectrometer equipped with an EASY Max NG source with a heated electrospray ionization HESI probe. Fifteen micrograms and 12 µg were injected for intact and middle-up level analyses, respectively. The separation was carried out in isocratic mode with a 100 mM AcONH₄ mobile phase at pH 7 using an ACQUITY Protein BEH SEC 200 Å, 1.7 µm, 4.6 × 150 mm (Waters) SEC column. The flow rate was set to 250 µL/min for intact and middle-up analyses. The source parameters were set as follows: sheath gas flow rate to 25 a.u., auxiliary gas flow rate to 5 a.u., ion transfer capillary temperature to 275 °C, vaporizer temperature to 200 °C, and capillary voltage 3.8 kV. In-source fragmentation energy was set to 150 V and 50 V for the intact and middle-up analyses, respectively, and the RF Lens to 150%. Intact protein mode was activated, with a pressure of 20 mTorr in the ion routing multipole (high-pressure mode). Acquisitions were performed in the *m/z* range 500–8000 with a 3-ms scan and resolution of 15,000. The AGC target was set to 2×10^6 , and the maximum injection time was 50 ms. Data were interpreted using BioPharma Finder v3.2 (Thermo Fisher Scientific, San Jose, CA, USA). The deconvolution mass tolerance was set to 10 ppm.

Middle-level LC-MS and LC-MS/MS analysis

Separation of the Fc/2, Fd, and LC subunits was performed with a BioResolve™ RP mAb polyphenyl column (450 Å pore size, 2.7 µm particle size, 2.1 × 150 mm, Waters) preheated at 60 °C, using a Dionex UltiMate 3000 RSLC system (Thermo Scientific). Mobile phase A consisted of

0.1% TFA in water, and mobile phase B was 0.1% TFA in ACN. For LC-MS and LC-MS/MS analyses, 3.5 µg and 5.5 µg of digested ADC were respectively loaded onto the column using 25% of mobile phase B at constant flow rate of 300 µL/min for 2 min. Digested ADC was eluted following a linear gradient: 25% at 2 min, 44% B at 40 min, and 75% B from 40.5 to 45.5 min. The column was finally re-equilibrated at 25% B for 15 min. The LC system was hyphenated to an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) in intact protein mode. RF Lens was set to 60%, and auxiliary and sheath gases were set to 5 a.u. and 15 a.u., respectively. For all experiments, the spray voltage was set to 3.8 kV, the ion transfer tube temperature to 275 °C, the vaporizer to 200 °C, and the advanced peak determination was activated. MS spectra were recorded using a mass range of 500–4000 *m/z* and a resolving power of 120,000 (at 200 *m/z*) for denaturing middle-up analysis.

For middle-down analysis, three different fragmentation modes were used: CID, ETD, and EThcD; as well as the PTCR ion-ion reaction. MS² spectra were recorded using a mass range of 300–2000 *m/z*, a resolving power of 120,000 (at 200 *m/z*) and 2 micro scans per spectra were selected. Precursor ions were selected over each specific elution window with an isolation window of 3 *m/z* following subsequent fragmentation. For each subunit, the most intense, for CID, or the highest charge state, for ETD/EThcD, precursor ions were selected and submitted to fragmentation. The AGC was set to 1×10^6 and the maximum injection time to 200 ms for precursor ions. CID collision energy was set to 35%, with a CID activation time of 10 ms and an activation Q parameter of 0.25. For ETD fragmentation, the reaction time between precursor ions and the anionic fluoranthene reagent was 3 ms. To ensure proper precursor ions fragmentation, the anionic reagent target was set to 7×10^5 with a maximum reagent injection time of 200 ms. For EThcD fragmentation, the same parameters as those described for ETD were used, and a supplemental collision energy of 20% was used. PTCR was performed after ETD and EThcD fragmentation on either a wide isolation window of 1000 *m/z* (700–1700 *m/z*) or five isolation windows of 200 *m/z* between 700 and 1700 *m/z*. The PTCR reaction time was set to 20 ms for the wide isolation windows. For narrow windows, reaction times of either 5 ms (for 700–900 *m/z*, 1300–1500 *m/z*, and 1500–1700 *m/z* isolation windows) or 10 ms (for 900–1100 *m/z* and 1100–1300 *m/z* isolation windows) were used, depending on the ion population. The reagent target was set to 5×10^5 and the maximum reagent injection time to 200 ms. MS³ spectra were recorded using a mass range of 500–4000 *m/z* and a resolving power of 120,000 (at 200 *m/z*), and 2 micro scans per spectra were selected.

Data analysis

MS/MS spectra were deconvoluted in FreeStyle 1.8 using the Xtract algorithm and deconvoluted masses were matched to the sequence using ProSight Lite v1.4. MS/MS spectra were averaged through the different subunit elution windows (0.5 min, 0.8 min, and 0.6 min for the Fc/2, Lc, and Fd subunits, respectively) and then deconvoluted using an S/N ratio of 7, a fit factor of 80%, and a remainder threshold of 25%. Deconvoluted masses were exported to.csv file and matched to the sequence with a 10-ppm tolerance. Different fragment ion series were considered depending on the fragmentation technique: *b/y* for CID, *c/z* for ETD, and *b/y* and *c/z* in the case of EThcD. Because of the DTT reduction performed during the sample preparation, all cysteines involved in intra-chain disulfide bridges were considered in their reduced form. In the case of the heavy chain and Fc/2 subunits, ions corresponding to the G0F glycoform were targeted, so G0F (+ 1444.5339 Da) at the Asp300 of the heavy chain (corresponding to Asp61 when Fc/2 is considered) was defined as a fixed modification to search the data. Drug conjugation (+ 1033.394 Da) on the cysteines involved in inter-chain disulfide bridges (Cys223, Cys229, and Cys232 on the Fd, Cys214 on the Lc) were set as fixed modifications.

For internal fragments analysis, ClipsMS [35] was used. Internal fragments were searched with an error tolerance of 2 ppm (10 ppm was kept for the terminal fragments) and the smallest internal fragment size of 5 amino acids. The following internal fragment ions were searched depending on the fragmentation method: *by* ions for CID, *cz* ions for ETD, and *by*, *cy*, *bz*, and *cz* ions for EThcD. Other parameters (modifications and terminal fragments tolerance) were set as mentioned above. For ions attributed to two different internal fragments, the assignment with the lower mass difference is kept and the other one is removed. If two (or more) fragments have exactly the same theoretical mass and mass difference with the assigned ion, then both identifications are kept.

Results and discussion

Intact and middle-level analysis of T-DXd for drug load distribution (DLD) determination and payload location assessment at the subunit level

The analysis of untreated T-DXd was first performed at the intact level with the use of the SEC-nMS coupling to confirm the drug load distribution (DLD) and the avDAR values of the ADC. Intact T-DXd analysis results in the identification of homogeneous T-DXd (and its glycoforms) bearing 8 conjugated drugs (avDAR of 8.0) in agreement with previously reported data [13] (Supplementary Figure S1).

We next performed classical RPLC-MS to separate and characterize the subunits of T-DXd obtained either after reduction (heavy (Hc) and light chains (Lc)) or upon IdeS digestion followed by DTT reduction (Lc, Fd, Fc/2 subunits, Supplementary Figure S2). Reduced T-DXd analysis reveals two main species eluting at 23.6 min and 30.5 min (Supplementary Figure S2.A). The first one corresponds to the Lc carrying one drug ($24,476.9 \pm 0.2$ Da, Supplementary Figure S2.B and Supplementary Table S1) and the second one to the Hc with three conjugated drugs ($53,703.6 \pm 0.4$ Da for the main glycoform G0F, Supplementary Figure S2.C and Supplementary Table S1). Both peaks exhibit additional populations at slightly lower retention times corresponding to the binding of partial payload scaffolds (-475 Da) [13]. The analysis of IdeS digested and reduced T-DXd showed three main peaks (Supplementary Figure S2.D), corresponding to the Fc/2 ($23,235.9 \pm 0.3$ Da for the main glycoform G0F, Supplementary Figure S2.E), Lc with one drug ($24,476.9 \pm 0.2$ Da, Supplementary Figure S2.F) and Fd bearing three drugs ($28,485.5 \pm 0.2$ Da, Supplementary Figure S2.G) subunits, respectively. In line with previous results for reduced T-DXd, minor chromatographic peaks resulting from Lc and Fd bearing partial drug(s) (-475 Da) are observed prior to the major chromatographic peaks of these subunits, but no drug-free species were detected. In addition, no signal corresponding to free Lc or Hc was detected. All these results regarding the absence of drug-free and/or partially conjugated species underline the high efficiency of the conjugation process for T-DXd. Middle-up level RPLC-MS analysis corroborates the location of 3 drugs on the Hc/Fd domain and one on the Lc domain, correlating with a homogeneous avDAR 8.0 calculated previously.

Middle-down mass spectrometry (MD-MS) analysis for sequence characterization and precise drug location

We next performed MD-MS analyses of T-DXd at the subunit level. The precursor ions selected for MS/MS fragmentation are displayed in Supplementary Figure S2 and detailed in the “Materials and methods” for each considered subunit. The most abundant charge state was selected for CID, while highly charged precursors were preferred for electron-based fragmentations (ETD and EThcD). Conversely to classical bottom-up strategies where sequence coverages are optimized, usually leading to accurate identification of drug location, we decided here to first address the position of the payloads by monitoring the sequence coverage as a function of payloads bound to different cysteine residues. Once the “most probable” conjugation site was determined, the MS/MS data obtained from each replicate of each fragmentation technique were combined to increase the overall sequence coverage of the subunits (9 runs combined).

For reduced T-DXd, Lc+1DXd or Hc+3DXd species were selected for further CID, ETD, and EThcD fragmentation. In this context, the MS/MS spectra of the Lc were matched with 5 different positional isomer sequences (5 Cys residues in the Lc sequence), while those of the Hc were compared with the sequence of 165 different isomers (combinations of 3 conjugated Cys residues out of 11 Cys,

Supplementary Figure S3). From Lc MS/MS data interpretation, it can be concluded that the conjugation site is located on the C-terminal part (Cys194 or Cys214, Fig. 1A–C), as higher sequence coverages (14%/38%/42% for CID/ETD/EThcD for Cys194-DXd, and 14%/39%/42% for CID/ETD/EThcD for Cys214-DXd) were achieved when considering DXd bound to one of those Cys residues. However, the z_{12}

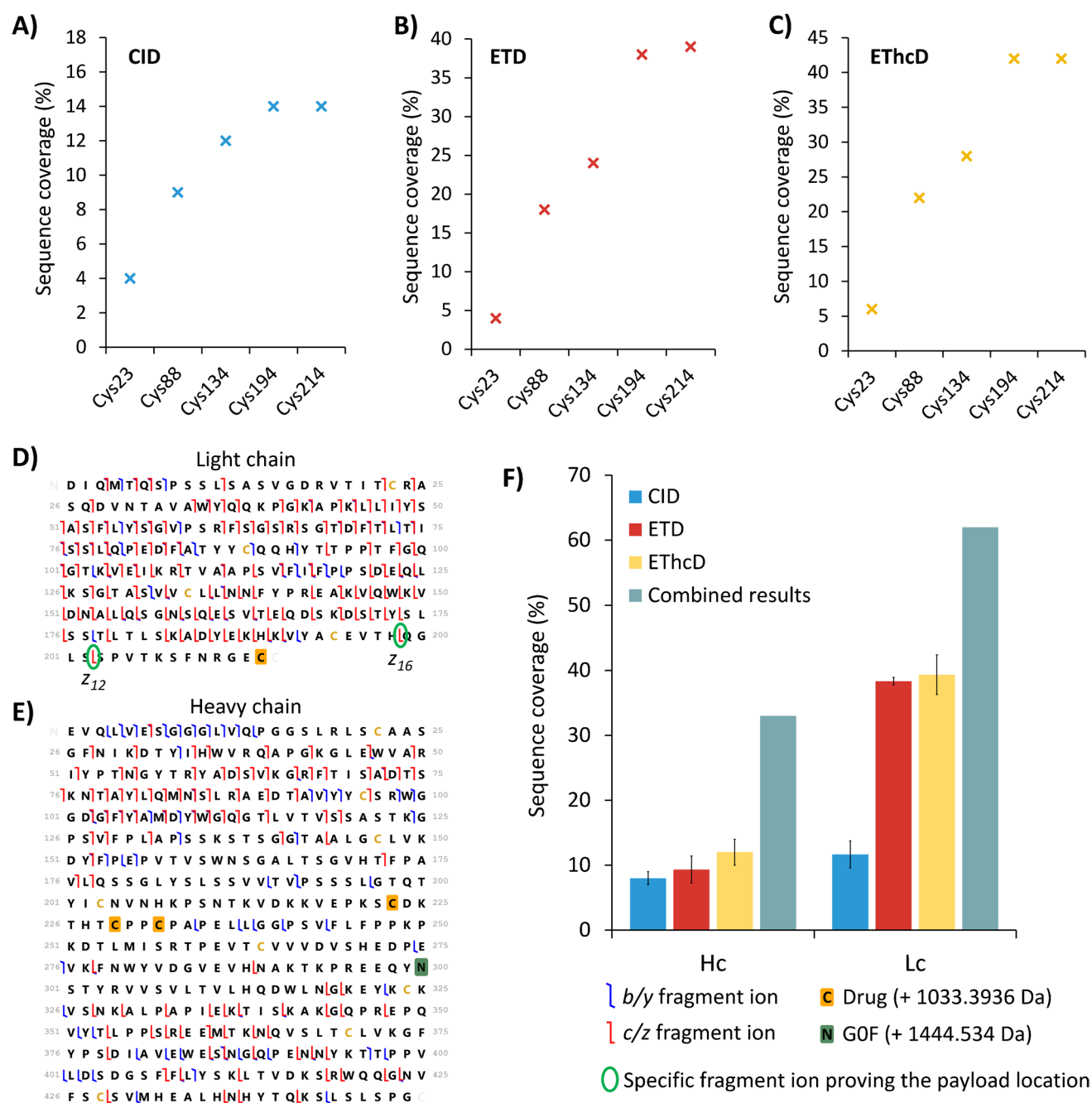


Fig. 1 Fragmentation results for the T-DXd heavy (Hc) and light chains (Lc). CID (A), ETD (B), and EThcD (C) light chain sequence coverage depending on the conjugated cysteine position. (D) Light chain fragmentation map of the combined results from all replicates.

(E) Heavy chain fragmentation map of the combined results from all replicates. (F) Hc and Lc average sequence coverage obtained by CID (blue), ETD (red), and EThcD (yellow) for three injection replicates, and the combined results from all runs (green)

fragment ion obtained in 3 over 3 ETD and EThcD replicates and the z_{16} obtained in 3 over 3 ETD replicates both allow determining that the payload is precisely located on Cys214 residue (Fig. 1D). Data interpretation is even more challenging for Hc (Fig. 1E) as the sequence coverage obtained is similar (8–11%) when considering the theoretical positions (i.e., Cys223, Cys229, and Cys232) of the payloads (Supplementary Figure S4), regardless of the activation technique used to induce the fragmentation. The fragmentation map of the combined results (9 runs) for the Hc clearly points out the absence of fragmentation in the middle of the Hc sequence, which impairs the reliable location of the conjugation sites in this chain (Fig. 1E). As expected, electron-based fragmentations (ETD/EThcD) afforded the most complete sequence coverage for both chains (39% for Lc and 12% for Hc), as already reported on other systems [28]. Finally, sequence coverage upon combining CID, ETD, and EThcD fragmentation replicates raised to 62% for the Lc and 33% for the Hc (Fig. 1F), which is still relatively low. Our results are in line with sequence coverages previously reported [48, 49], which can be explained by the difficulty of providing a complete sequence coverage characterization when fragmenting proteins larger than 30 kDa [50]. In the particular case of the Hc, this limited fragmentation hampers the precise location of the conjugation sites.

In order to increase the sequence coverage and locate more precisely the conjugation sites, T-DXd was IdeS digested and reduced to release the three ~ 25 kDa subunits (Fc/2, Lc, and Fd) and subsequently fragmented using the aforementioned activation modes (Supplementary Figure S5). Similar to results obtained on Lc generated after DTT treatment, MD-MS results obtained after IdeS digestion allow to conclude that DXd is located on the Cys214 of Lc (Fig. 2A), as the two same diagnostic fragments specific of the drug conjugation on Cys214 were identified (z_{12} and z_{16} , Supplementary Figure S6.A) upon ETD/EThcD fragmentation (in 3/3 and 2/3 replicates, respectively). As IdeS digestion efficiently downsized the Hc chain, we next analyzed the results from the 35 possible Fd isomers (3 conjugated Cys residues out of 7 Cys residues in the Fd, see Supplementary Figure S3). It came out that proteoforms including conjugated C-terminal Cys (i.e., Cys203, Cys223, Cys229, and Cys 232) led to the highest sequence coverage (Fig. 2B), suggesting possible DXd location on abovementioned Cys. Nevertheless, 7 fragments (all from ETD/EThcD and assigned in at least 2/3 replicates for each fragmentation technique) corroborated the presence of the payloads on the 3 C-terminal Cys residues involved in inter-chain disulfide bridges of the mAb counterpart (Supplementary Figure S6.B). Here, we observed fragment ions confirming the drug location for both Lc (Cys214) and Fd subunits (Cys223, Cys229, and Cys232). Moreover, after manual validation, it was confirmed that those ions exhibited resolved isotope

profiles with no interfering signals and with a significant S/N ratio (> 150), thus increasing the confidence in their identification (Supplementary Figure S6.C and S7).

The conjugation sites inferred from MD-MS strategies were corroborated with bottom-up analysis (Supplementary Figure S8), with no detection of additional conjugation sites by the latter technique, highlighting the complementarity between both strategies. However, peptide mapping results still evidence the difficulties of this approach to decipher the drug conjugation of an ADC, as only a few fragments contain the whole payload scaffold and the location of the conjugation site is performed by detecting payload-bound peptide ions, with the detection of a specific signature fragment ion coming from the fragmentation of the cargo molecule (Supplementary Figure S6).

Best sequence coverages achieved for T-DXd subunits for each individual fragmentation technique ranged between 29 and 61% for Fc/2, 15–48% for Lc, and 19–37% for Fd, with best results originating from EThcD fragmentation (61%, 48%, and 37% with EThcD for the Fc/2, Lc, and Fd parts, respectively). Between 40 and 55% of all the identified terminal ions were observed in all the replicates of the same fragmentation technique, and the individual p-scores associated to the different subunits with the precise conjugation site were lower than 10^{-10} for CID, 10^{-20} for EThcD, and 10^{-37} for ETD. Interestingly, combining replicates from all fragmentation modes resulted in a high sequence coverage for the Fc/2, Lc, and Fd subunits with 87%, 76%, and 72%, respectively (Fig. 2C) covering not only both termini of the sequences but also the internal regions as expected when fragmenting 25-kDa subunits.

Altogether, MD-MS investigations of T-DXd subunits confirm the payloads conjugation on the Cys214 for the Lc (with z_{12} and/or z_{16} specific diagnostic ions), while DXd binding could only be restricted to the C-terminal area of the Fd subunit (Cys223, Cys229, or Cys232) with the detection of 7 specific fragment ions. As expected, MD-MS results in terms of sequence coverage were better with EThcD and after IdeS digestion, in which smaller subunits (~ 25–30 kDa) were more adapted to MD-MS analysis.

Contribution of internal fragments in the context of T-DXd subunits sequence characterization

In classical TD/MD-MS experiments, only N- and C-terminal fragments (originating from one single cleavage site in the protein backbone) are matched to determine the sequence coverage. However, protein backbone can undergo subsequent residue cleavages leading to the formation of internal fragments that are not containing neither the N- nor the C-terminus of the protein. ClipsMS [35] was used to tackle internal fragments generated in our MD-MS workflow. Terminal fragments were assigned first using a 10-ppm

to 45% when the conjugation site is assumed either at Cys23 or Cys214, respectively, clearly pinpointing that the conjugation site is located in the latter residue. However, upon

internal fragment matching, the sequence coverage variation between these two conjugated positions is more attenuated (from 51 to 67%), leading to a more ambiguous result.

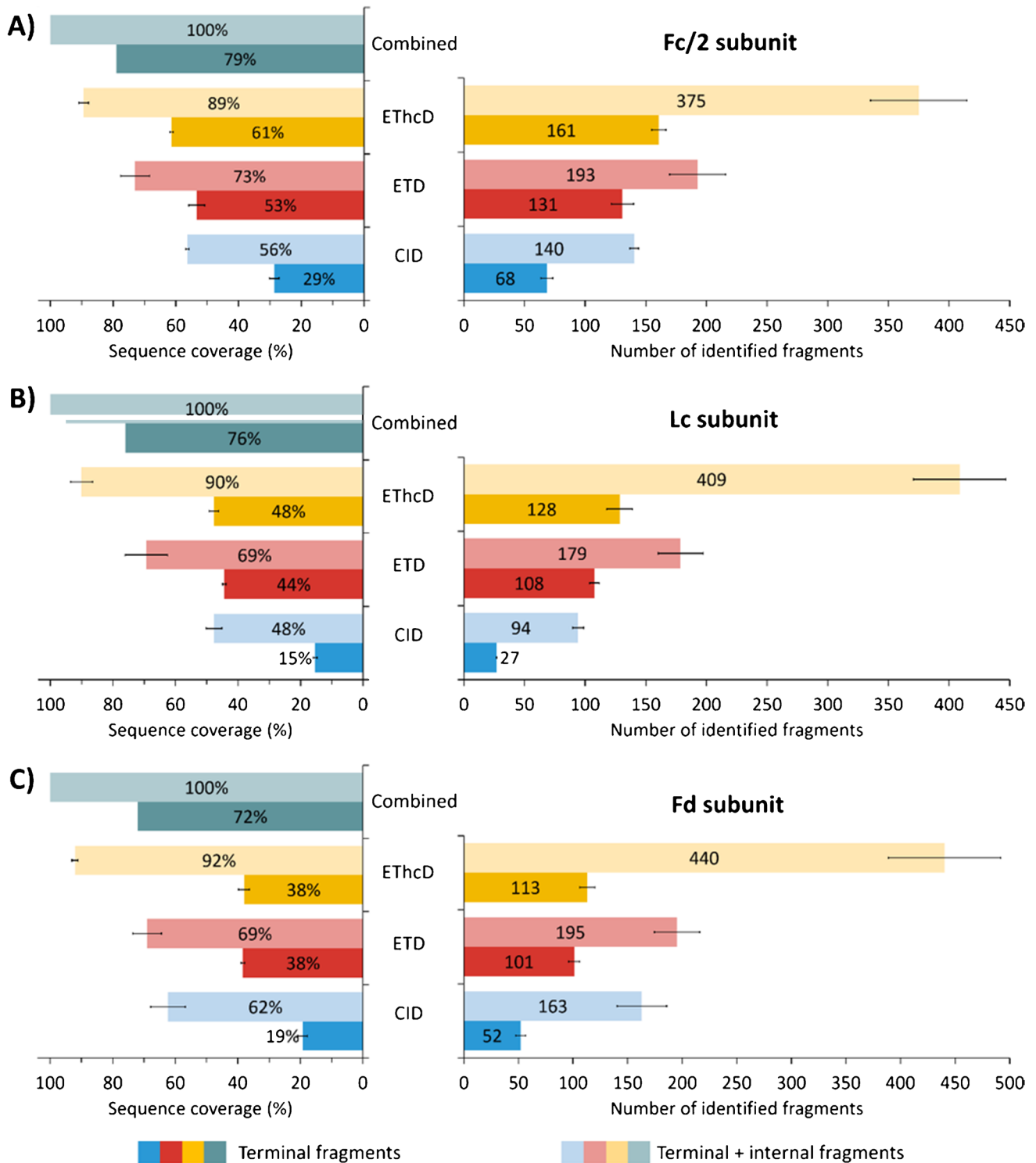


Fig. 3 Average sequence coverage and number of fragments identified for T-DXd Fc/2, Fd, and Lc subunits using internal fragments. Average sequence coverage (left) and number of identified fragments (right) considering the terminal fragments only (dark colors) and ter-

terminal+internal fragments (light colors) for CID (blue), ETD (red), and EThcD (yellow) of Fc/2 (A), LC (B), and Fd (C) subunits. The results of the individual fragmentation techniques were run in triplicate and the combined results are depicted in green

Finally, it was observed that internal fragment consideration entails an increased probability of ion mismatching [45]. For example, the search for internal fragments after ETD by positioning the drug on the Cys23 residue of the Lc leads to the detection of more than 20 internal fragments containing the drug. More precisely, one identified internal fragment exhibits a clear isotopic profile with a very precise mass measurement that could correspond with the 21–52 internal fragment bearing the payload (4523.22 Da, Supplementary Figure S9). Nevertheless, the Cys23 conjugation site was excluded from our previous MD-MS results based on terminal fragment identification because of the low sequence coverage assigned to this position. The golden-standard bottom-up method did not provide any peptide containing the drug at position 23, clearly suggesting that all those ions could not correspond to the fragmentation of a bunch of the Lc backbone with conjugated Cys23. For all these reasons, the characterization of the conjugation sites relied on the identification of terminal fragments, while internal fragment ions were used with the main purpose of boosting the sequence coverage of the different subunits.

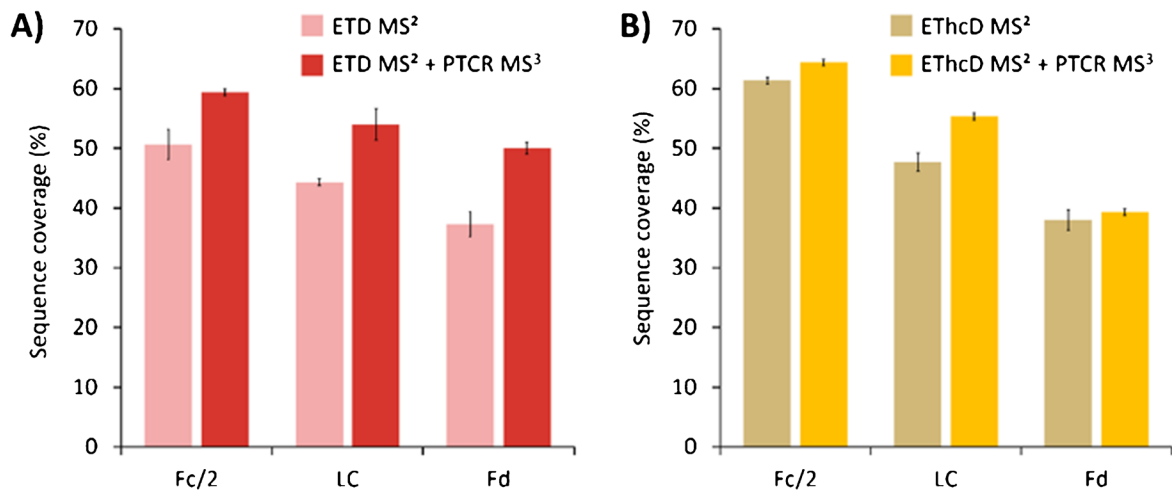
The proportion of internal fragments identified differs with the fragmentation method. Depending on the subunit, the average number of internal fragments identified varies between 72 and 111 for CID, 62 and 94 for ETD, and 214 and 327 for EThcD. As expected, EThcD provides the higher number of internal fragments resulting from the addition of a collisional-activation step of ions that have been previously activated/dissociated upon electron transfer. In the end, a total number of fragments (terminal and internal) around 130, 200, and 400 for CID, ETD, and EThcD, respectively, were detected (Fig. 3A–C, right). Consequently, the sequence coverage drastically increased when including internal fragments to reach 50 to 60% with CID (15–30% considering terminal fragments only), 70% with ETD (38–53% considering terminal fragments only), and 90% for EThcD (38–61% considering terminal fragments only) (Fig. 3A–C, left). Manual validation of the internal fragments identified can help to ascertain the confidence in the match, and it was found that internal fragments were relatively well resolved (Supplementary Figure S10). Combining the sequence coverage from replicates of all the fragmentation techniques (i.e., CID, ETD, and EThcD) resulted in 100% subunit sequence confirmation (Fig. 3). Concomitantly, the relative proportion of assigned fragment ions increased by 1.6 to 3 times when using internal fragment ion searching. However, poor reproducibility between different replicates was observed regarding the internal fragment identification since only around 20% of the total number of the internal fragments were detected in at least 2 over 3 replicates (and less than 5% in 3 over 3 replicates). These results clearly show that considering internal fragment ion search during MD-MS workflows allows taking full advantage of

all the information contained in the MS/MS spectra, highlighting the contribution of internal fragments to increase the sequence coverage although care must be taken to reduce the number of false or ambiguous assignments.

Strengthening drug location using proton transfer charge reduction (PTCR)

While the attribution of internal fragments allows obtaining the complete sequence coverage by exploiting the huge amount of information available from MS² spectra, the risk of false positives and the position of the payloads on the C-termini of the sequences make its contribution limited to the identification of the conjugation sites. With this in mind, PTCR was combined with the most efficient fragmentation, i.e., ETD and EThcD, to declutter MS² spectra and confirm drug location. To do so, two different PTCR methods were used: (i) the “wide window” (1000 *m/z* isolation window) method and (ii) the “narrow window” (5 × 200 *m/z* isolation window) mode (Supplementary Figure S11). As highlighted in Fig. 4A and B, the combination of PTCR results from all the windows leads to an average rise of the sequence coverage by 10–13% after ETD fragmentation and by 2–7% after EThcD fragmentation, respectively. The practical effect of this ion/ion reaction is further illustrated in Supplementary Figure S8. After ETD fragmentation, a complex MS² spectrum is generated, with a very large number of fragments distributed only over a small *m/z* range (700–1700 *m/z*, Supplementary Figure S8.A). Using PTCR for 10 ms on the 1100–1300 *m/z* window drastically reduced the spectral congestion of this zone of the MS² spectrum, decreasing the signal interference in the lower *m/z*, where the major part of the fragment ions of interest are observed. For instance, overlapping signals and interferences are observed in the ETD MS² spectrum of the diagnostic fragment ion *z*₁₂ (2341.00 Da), which allows the precise location of the DXd on the Lc (Supplementary Figure S12.A). Thanks to the addition of PTCR, the same ion could be assigned with a resolved isotope profile and a low interfering signal, leading to a more reliable identification of the fragment (Supplementary Figure S8.B). The PTCR MS² spectra decluttering capabilities not only afforded more confident identification of diagnostic ions detected without PTCR but, more interestingly, allowed

Fig. 4 PTCR fragmentation results for the T-DXd Fc/2, Lc, and Fd subunits. **(A)** Fc/2, Lc, and Fd average sequence coverage obtained by ETD only (light red) and ETD+PTCR (dark red) for the three injection replicates. **(B)** Fc/2, Lc, and Fd average sequence coverage obtained by EThcD only (light yellow) and EThcD+PTCR (dark yellow) for the three injection replicates. Fragmentation maps combining ETD/EThcD MS² and PTCR MS³ results of the Lc **(C)** and Fd **(D)** subunits. Fragments found uniquely thanks to PTCR are circled in black and those confirming the position of the drugs are circled in green



C) Lc subunit

N D I Q M T Q S P S S L L S A S M G D R V T I T C R A 25
 26 S Q D V N T A V A W Y Q Q K P G K A P K L L I Y S 50
 51 A S F L Y S G V P S R F S G S R S G T D F T L T I 75
 76 S S L Q P E D F A T Y Y C Q Q H Y T T P P T F G Q 100
 101 G T K V E I K R T V A A P S V F I F P P S D E Q L 125
 126 K S G T A S V V C L L L N F Y P R E A K V Q W K V 150
 151 D N A L L Q S G N S Q E S V T E Q D S K D S T Y S L 175
 176 S S L T L L T L S K A D Y E K H K V Y A C L E V T H Q G 200
 201 L S S P V T K S F N R G E C C

D) Fd subunit

N E V Q L V E S G G G L V Q P G S L R L S C A A S 25
 26 G F N I K D T Y I H W V R Q A P G K G L E W V A R 50
 51 I Y P T N G Y T R Y A D S V K G R F T I S A D T S 75
 76 K N T A Y L Q M N S L R A E D T A V Y Y C S R W G 100
 101 G D G F Y A M D Y W G Q G T L V T V S S A S T K G 125
 126 P S V F P L L A P S S K S T S G G T A A L G C L V K 150
 151 D Y F F E P V T V S W N S E A L L T S G V H T F P A 175
 176 Y L L Q S S G L L S L S S V L V T V P S S S L G T Q T 200
 201 Y I C N V N H K P S N T K V D K K V E P K S C D K 225
 226 T H T C P P C P A P E L L G C

- Additional fragment ions found using PTZR
- Additional fragment ions found using PTZR and confirming the presence of the drugs

identification of additional specific diagnostic fragment ions on the different subunits. Upon addition of PTCR, a total of 5 fragment ions specific of the Cys214 conjugation on the Lc (2 fragments found with ETD/ETHcD and 3 additional found after PTCR implementation, z_{13} , y_{19} , and y_{20} , all in at least 2/3 replicates, Fig. 4C) and 11 signature fragments specific of the conjugation of the Fd subunit on the Cys223, Cys229, and Cys232 residues (7 fragments found with ETD/ETHcD and 4 additional fragments detected after PTCR implementation, z_{26} , z_{27} , z_{30} , and z_{32} , all identified in at least 2/3 replicates, Fig. 4D) were finally detected. For instance, the fragment z_{26} that was not observed at any charge states using ETD only is assigned after the use of PTCR for 10 ms on the 1300–1500 m/z mass range, as the 4+ charge state has been detected and presents a well-resolved isotopic profile (Supplementary Figure S13). These results highlight that PTCR contributes to the improvement of the overall sequence coverage to a lesser extent compared to internal fragment ion search but enhances the number and confidence of identified diagnostic fragment ions of the conjugation site, providing additional experimental evidence to locate the sequence modifications of interest.

Conclusions

In this study, several MS-based techniques have been combined to characterize a homogeneous DAR8 ADC, T-DXd.

The different subunits bearing the cargo molecules were subjected to fragmentation to elucidate the conjugation sites. Results evidenced that the precise location of the three payloads from the fragmentation of the entire Hc was still challenging due to the lack of fragmentation in the mid-region of the backbone. In this case, the fragmentation of the Fd subunit was more informative, providing a more comprehensive sequence coverage along with a significant number of signature fragment ions characteristic of the three conjugated cysteine residues.

PTCR reaction and tailored TD/MD-MS software were used to maximize the number of identified fragment ions. The identification of internal fragments with the use of ClipsMS helped to confirm the whole sequence of the ADC, giving rise to the characterization of the full sequence of each subunit. Although very informative to improve the overall sequence coverage, special care should be taken when considering internal fragment ions since the probabilities for false positive ion assignment increase, providing ambiguous results about the precise location of the cytotoxic payload in the sequence of the ADC.

PTCR emerged as an elegant alternative to explore more deeply the fragmentation spectra information. This technique led to an overall sequence coverage increase between

2 and 13% for each ADC subunit. In spite of the moderate sequence coverage contribution of PTCR, the enhanced separation of charge-reduced fragment ions allowed for increasing the confidence of the location of the conjugation sites by increasing the number of diagnostic ions by 1.5- to 2.5-fold. According to these results, the combination of a single activation technique (i.e., ETD or ETHcD) with PTCR reaction can be an appealing alternative to the combination of multiple activation techniques (CID, HCD, ETD, ETHcD, UVPD) that will afford additional information to characterize the sequence modification of interest.

Altogether, these results highlight the potential of combining different MS-based approaches, with particular interest in cutting-edge MD-MS strategies, in the characterization of approved therapeutic proteins to evaluate the homogeneity of the bioconjugation with limited sample preparation. More particularly, the contribution of MD-MS workflows in the conjugation site characterization of ADCs has been clearly shown in comparison with classical bottom-up approaches, facilitating the identification of fragment ions containing the whole payload, which is a significant limitation when ADCs are characterized at the peptide level. Hopefully, the continuous development of tailored software and the implementation of novel strategies in last-generation MS platforms will foster the use of intact or partially digested protein fragmentation in different R&D laboratories, especially (but not exclusively) those dedicated to the analysis of therapeutic proteins.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00216-023-05059-x>.

Acknowledgements The authors acknowledge Alexandre Detappe (ICANS, Strasbourg, France) for providing the samples.

Author contribution CB: methodology, investigation, formal analysis, data curation, writing—original draft preparation, reviewing and editing

CC: writing—reviewing and editing

ED: investigation, writing—reviewing and editing

HD: investigation, formal analysis, data curation, writing—reviewing and editing

OHA: conceptualization, supervision, writing—original draft preparation, reviewing, and editing

SC: resources, writing—reviewing and editing

Funding This work was supported by the CNRS, and the “Agence Nationale de la Recherche” via the French Proteomic Infrastructure (ProFI FR2048; ANR-10-INBS-08-03), the ConformAbs project (ANR-21-CE29-0009-01), by the Région Grand-Est (Fonds Régional de la Coopération pour la Recherche, HRMS-INFECT project) and by the French Ministry of Higher Education and Research for the PhD fellowship of C.B.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Kaplon H, Chenoweth A, Crescioli S, Reichert JM. Antibodies to watch in 2022. *mAbs*. 2022;14(1):2014296.
- Kaplon H, Crescioli S, Chenoweth A, Visweswarajah J, Reichert JM. Antibodies to watch in 2023. *mAbs*. 2023;15(1):2153410.
- Fu Z, Li S, Han S, Shi C, Zhang Y. Antibody drug conjugate: the “biological missile” for targeted cancer therapy. *Signal Transduct Target Ther*. 2022;7(1):93.
- Joubert N, Beck A, Dumontet C, Denevault-Sabourin C. Antibody-drug conjugates: the last decade. *Pharmaceuticals (Basel)*. 2020;13(9):245.
- Nadkarni DV. Conjugations to endogenous cysteine residues. *Methods Mol Biol (Clifton, NJ)*. 2020;2078:37–49.
- Haque M, Forte N, Baker JR. Site-selective lysine conjugation methods and applications towards antibody-drug conjugates. *Chem Commun (Camb)*. 2021;57(82):10689–702.
- Sadiki A, Vaidya SR, Abdollahi M, Bhardwaj G, Dolan ME, Turna H, et al. Site-specific conjugation of native antibody. *Antib Ther*. 2020;3(4):271–84.
- Adhikari P, Zacharias N, Ohri R, Sadowsky J. Site-specific conjugation to Cys-engineered TH10MAB™ antibodies. *Methods Mol Biol (Clifton, NJ)*. 2020;2078:51–69.
- Hallam TJ, Wold E, Wahl A, Smider VV. Antibody conjugates with unnatural amino acids. *Mol Pharm*. 2015;12(6):1848–62.
- Diamantis N, Banerji U. Antibody-drug conjugates—an emerging class of cancer treatment. *Br J Cancer*. 2016;114(4):362–7.
- Xu Y, Wang D, Mason B, Rossomando T, Li N, Liu D, et al. Structure, heterogeneity and developability assessment of therapeutic antibodies. *mAbs*. 2019;11(2):239–64.
- Beck A, D’Atri V, Ehkirch A, Fekete S, Hernandez-Alba O, Gahoual R, et al. Cutting-edge multi-level analytical and structural characterization of antibody-drug conjugates: present and future. *Expert Rev Proteomics*. 2019;16(4):337–62.
- Deslignière E, Diemer H, Erb S, Coliat P, Pivot X, Detappe A, et al. A combination of native LC-MS approaches for the comprehensive characterization of the antibody-drug conjugate trastuzumab deruxtecan. *Front Biosci (Landmark Ed)*. 2022;27(10):290.
- van Schaick G, Haselberg R, Somsen GW, Wuhler M, Domínguez-Vega E. Studying protein structure and function by native separation-mass spectrometry. *Nat Rev Chem*. 2022;6(3):215–31.
- Donnelly DP, Rawlins CM, DeHart CJ, Fornelli L, Schachner LF, Lin Z, et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods*. 2019;16(7):587–94.
- Broadbelt JS. Deciphering combinatorial post-translational modifications by top-down mass spectrometry. *Curr Opin Chem Biol*. 2022;70: 102180.
- Chapman EA, Aballo TJ, Melby JA, Zhou T, Price SJ, Rossler KJ, et al. Defining the sarcomeric proteoform landscape in ischemic cardiomyopathy by top-down proteomics. *J Proteome Res*. 2023;22(3):931–41.
- Jeanne Dit Fouque K, Miller SA, Pham K, Bhanu NV, Cintron-Diaz YL, Leyva D, et al. Top-“double-down” mass spectrometry of histone H4 proteoforms: tandem ultraviolet-photon and mobility/mass-selected electron capture dissociations. *Anal Chem*. 2022;94(44):15377–85.
- Lodge JM, Schauer KL, Brademan DR, Riley NM, Shishkova E, Westphall MS, et al. Top-down characterization of an intact monoclonal antibody using activated ion electron transfer dissociation. *Anal Chem*. 2020;92(15):10246–51.
- Jin Y, Lin Z, Xu Q, Fu C, Zhang Z, Zhang Q, et al. Comprehensive characterization of monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. *mAbs*. 2019;11(1):106–15.
- He L, Anderson LC, Barnidge DR, Murray DL, Hendrickson CL, Marshall AG. Analysis of monoclonal antibodies in human serum as a model for clinical monoclonal gammopathy by use of 21 Tesla FT-ICR top-down and middle-down MS/MS. *J Am Soc Mass Spectrom*. 2017;28(5):827–38.
- van der Burgt YEM, Kilgour DPA, Tsybin YO, Srzentić K, Fornelli L, Beck A, et al. Structural analysis of monoclonal antibodies by ultrahigh resolution MALDI in-source decay FT-ICR mass spectrometry. *Anal Chem*. 2019;91(3):2079–85.
- Shaw JB, Liu W, Vasil Ev YV, Bracken CC, Malhan N, Guthals A, et al. Direct determination of antibody chain pairing by top-down and middle-down mass spectrometry using electron capture dissociation and ultraviolet photodissociation. *Anal Chem*. 2020;92(1):766–73.
- Kellie JF, Schneck NA, Causon JC, Baba T, Mehl JT, Pohl KI. Top-down characterization and intact mass quantitation of a monoclonal antibody drug from serum by use of a quadrupole TOF MS system equipped with electron-activated dissociation. *J Am Soc Mass Spectrom*. 2023;34(1):17–26.
- Watts E, Williams JD, Miesbauer LJ, Bruncko M, Brodbelt JS. Comprehensive middle-down mass spectrometry characterization of an antibody-drug conjugate by combined ion activation methods. *Anal Chem*. 2020;92(14):9790–8.
- Larson EJ, Roberts DS, Melby JA, Buck KM, Zhu Y, Zhou S, et al. High-throughput multi-attribute analysis of antibody-drug conjugates enabled by trapped ion mobility spectrometry and top-down mass spectrometry. *Anal Chem*. 2021;93(29):10013–21.
- Chen B, Lin Z, Zhu Y, Jin Y, Larson E, Xu Q, et al. Middle-down multi-attribute analysis of antibody-drug conjugates with electron transfer dissociation. *Anal Chem*. 2019;91(18):11661–9.
- Hernandez-Alba O, Houel S, Hessmann S, Erb S, Rabuka D, Huguet R, et al. A case study to identify the drug conjugation site of a site-specific antibody-drug-conjugate using middle-down mass spectrometry. *J Am Soc Mass Spectrom*. 2019;30(11):2419–29.
- Sang H, Lu G, Liu Y, Hu Q, Xing W, Cui D, et al. Conjugation site analysis of antibody-drug-conjugates (ADCs) by signature ion fingerprinting and normalized area quantitation approach using nano-liquid chromatography coupled to high resolution mass spectrometry. *Anal Chim Acta*. 2017;955:67–78.
- Bashyal A, Dunham SD, Broadbelt JS. Characterization of unbranched ubiquitin tetramers by combining ultraviolet photodissociation with proton transfer charge reduction reactions. *Anal Chem*. 2023;95(37):14001–8.
- Walker JN, Lam R, Broadbelt JS. Enhanced characterization of histones using 193 nm ultraviolet photodissociation and proton transfer charge reduction. *Anal Chem*. 2023;95(14):5985–93.
- Huguet R, Mullen C, Srzentić K, Greer JB, Fellers RT, Zabrouskov V, et al. Proton transfer charge reduction enables high-throughput top-down analysis of large proteoforms. *Anal Chem*. 2019;91(24):15732–9.
- Dunham SD, Wei B, Lantz C, Loo JA, Broadbelt JS. Impact of internal fragments on top-down analysis of intact proteins by 193 nm UVPD. *J Proteome Res*. 2023;22(1):170–81.
- Kline JT, Mullen C, Durbin KR, Oates RN, Huguet R, Syka JEP, et al. Sequential ion-ion reactions for enhanced gas-phase sequencing of large intact proteins in a tribrid Orbitrap mass spectrometer. *J Am Soc Mass Spectrom*. 2021;32(9):2334–45.
- Lantz C, Zenaidee MA, Wei B, Hemminger Z, Ogorzalek Loo RR, Loo JA. ClipsMS: an algorithm for analyzing internal fragments resulting from top-down mass spectrometry. *J Proteome Res*. 2021;20(4):1928–35.
- Schmitt ND, Berger JM, Conway JB, Agar JN. Increasing top-down mass spectrometry sequence coverage by an order of magnitude through optimized internal fragment generation and assignment. *Anal Chem*. 2021;93(16):6355–62.

37. Stephenson JL Jr, McLuckey SA. Simplification of product ion spectra derived from multiply charged parent ions via ion/ion chemistry. *Anal Chem.* 1998;70(17):3533–44.
38. Rolfs Z, Smith LM. Internal fragment ions disambiguate and increase identifications in top-down proteomics. *J Proteome Res.* 2021;20(12):5412–8.
39. Wei B, Lantz C, Liu W, Viner R, Ogorzalek Loo RR, Campuzano IDG, et al. Added value of internal fragments for top-down mass spectrometry of intact monoclonal antibodies and antibody-drug conjugates. *Anal Chem.* 2023;95(24):9347–56.
40. Lyon YA, Riggs D, Fornelli L, Compton PD, Julian RR. The ups and downs of repeated cleavage and internal fragment production in top-down proteomics. *J Am Soc Mass Spectrom.* 2018;29(1):150–7.
41. Zenaidee MA, Lantz C, Perkins T, Jung W, Loo RRO, Loo JA. Internal fragments generated by electron ionization dissociation enhance protein top-down mass spectrometry. *J Am Soc Mass Spectrom.* 2020;31(9):1896–902.
42. Zenaidee MA, Wei B, Lantz C, Wu HT, Lambeth TR, Diedrich JK, et al. Internal fragments generated from different top-down mass spectrometry fragmentation methods extend protein sequence coverage. *J Am Soc Mass Spectrom.* 2021;32(7):1752–8.
43. Wei B, Zenaidee MA, Lantz C, Ogorzalek Loo RR, Loo JA. Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry. *Anal Chim Acta.* 2022;1194: 339400.
44. Wei B, Zenaidee MA, Lantz C, Williams BJ, Totten S, Ogorzalek Loo RR, et al. Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins. *Analyst.* 2022;148(1):26–37.
45. Dunham SD, Wei B, Lantz C, Loo JA, Brodbelt JS. Impact of internal fragments on top-down analysis of intact proteins by 193 nm UVPD. *J Proteome Res.* 2022;22(1):170–81.
46. Nakada T, Masuda T, Naito H, Yoshida M, Ashida S, Morita K, et al. Novel antibody drug conjugates containing exatecan derivative-based cytotoxic payloads. *Bioorg Med Chem Lett.* 2016;26(6):1542–5.
47. Ogitani Y, Aida T, Hagihara K, Yamaguchi J, Ishii C, Harada N, et al. DS-8201a, A novel HER2-targeting ADC with a novel DNA topoisomerase I inhibitor, demonstrates a promising antitumor efficacy with differentiation from T-DM1. *Clin Cancer Res.* 2016;22(20):5097–108.
48. Dhenin J, Dupré M, Druart K, Krick A, Mauriac C, Chamot-Rooke J. A multiparameter optimization in middle-down analysis of monoclonal antibodies by LC-MS/MS. *J Mass Spectrom JMS.* 2023;58(3): e4909.
49. Melani RD, Srzentić K, Gerbasi VR, McGee JP, Huguet R, Fornelli L, et al. Direct measurement of light and heavy antibody chains using ion mobility and middle-down mass spectrometry. *mAbs.* 2019;11(8):1351–7.
50. Fornelli L, Parra J, Hartmer R, Stoermer C, Lubeck M, Tsybin YO. Top-down analysis of 30–80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry. *Anal Bioanal Chem.* 2013;405(26):8505–14.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

REFERENCES

References

- (1) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschläger, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B. How many human proteoforms are there? *Nat Chem Biol* **2018**, *14* (3), 206-214. DOI: 10.1038/nchembio.2576
- (2) Smith, L. M.; Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat Methods* **2013**, *10* (3), 186-187. DOI: 10.1038/nmeth.2369
- (3) Kaplon, H.; Crescioli, S.; Chenoweth, A.; Visweswarajah, J.; Reichert, J. M. Antibodies to watch in 2023. *MAbs* **2023**, *15* (1), 2153410. DOI: 10.1080/19420862.2022.2153410
- (4) 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. DOI: 10.1126/science.2675315
- (5) Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988**, *60* (20), 2299-2301. DOI: 10.1021/ac00171a028
- (6) 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), 0111.016717. DOI: 10.1074/mcp.0111.016717
- (7) Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol Omics* **2019**, *15* (5), 348-360. DOI: 10.1039/c9mo00082h
- (8) Holt, M. V.; Wang, T.; Young, N. L. High-Throughput Quantitative Top-Down Proteomics: Histone H4. *J Am Soc Mass Spectrom* **2019**, *30* (12), 2548-2560. DOI: 10.1007/s13361-019-02350-z
- (9) Hernandez-Alba, O.; Houel, S.; Hessmann, S.; Erb, S.; Rabuka, D.; Hugué, R.; Josephs, J.; Beck, A.; Drake, P. M.; Cianféroni, S. A Case Study to Identify the Drug Conjugation Site of a Site-Specific Antibody-Drug-Conjugate Using Middle-Down Mass Spectrometry. *J Am Soc Mass Spectrom* **2019**, *30* (11), 2419-2429. DOI: 10.1007/s13361-019-02296-2
- (10) Fornelli, L.; Srzentić, K.; Hugué, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal Chem* **2018**, *90* (14), 8421-8429. DOI: 10.1021/acs.analchem.8b00984
- (11) Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **1996**, *14* (1), 61-65. DOI: 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. DOI: 10.1002/elps.1150191103
- (13) James, P. Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* **1997**, *30* (4), 279-331. DOI: 10.1017/s0033583597003399
- (14) Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Darie, C. C. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field. *Proteomes* **2020**, *8* (3). DOI: 10.3390/proteomes8030014
- (15) Nesvizhskii, A. I.; Vitek, O.; Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **2007**, *4* (10), 787-797. DOI: 10.1038/nmeth1088
- (16) Graves, P. R.; Haystead, T. A. Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev* **2002**, *66* (1), 39-63; table of contents. DOI: 10.1128/mmb.66.1.39-63.2002
- (17) Gillet, L. C.; Leitner, A.; Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu Rev Anal Chem (Palo Alto Calif)* **2016**, *9* (1), 449-472. DOI: 10.1146/annurev-anchem-071015-041535
- (18) Miller, R. M.; Smith, L. M. Overview and considerations in bottom-up proteomics. *Analyst* **2023**, *148* (3), 475-486. DOI: 10.1039/d2an01246d
- (19) 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. DOI: 10.1021/acs.analchem.8b00693

- (20) Brodbelt, J. S. Deciphering combinatorial post-translational modifications by top-down mass spectrometry. *Curr Opin Chem Biol* **2022**, *70*, 102180. DOI: 10.1016/j.cbpa.2022.102180
- (21) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W.; Mohammed, S.; Heck, A. J. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal Chem* **2017**, *89* (6), 3318-3325. DOI: 10.1021/acs.analchem.6b03756
- (22) Pandeswari, P. B.; Sabareesh, V. Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry. *RSC Adv* **2018**, *9* (1), 313-344. DOI: 10.1039/c8ra07200k
- (23) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013**, *113* (4), 2343-2394. DOI: 10.1021/cr3003533
- (24) Nesvizhskii, A. I.; Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **2005**, *4* (10), 1419-1440. DOI: 10.1074/mcp.R500012-MCP200
- (25) 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. DOI: 10.1038/s41592-018-0003-5
- (26) Duong, V. A.; Lee, H. Bottom-Up Proteomics: Advancements in Sample Preparation. *Int J Mol Sci* **2023**, *24* (6). DOI: 10.3390/ijms24065350
- (27) Rogers, J. C.; Bomgarden, R. D. Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. *Adv Exp Med Biol* **2016**, *919*, 43-62. DOI: 10.1007/978-3-319-41448-5_3
- (28) Channaveerappa, D.; Ngounou Wetie, A. G.; Darie, C. C. Bottlenecks in Proteomics: An Update. *Adv Exp Med Biol* **2019**, *1140*, 753-769. DOI: 10.1007/978-3-030-15950-4_45
- (29) Millán-Martín, S.; Jakes, C.; Carillo, S.; Buchanan, T.; Guender, M.; Kristensen, D. B.; Sloth, T. M.; Ørgaard, M.; Cook, K.; Bones, J. Inter-laboratory study of an optimised peptide mapping workflow using automated trypsin digestion for monitoring monoclonal antibody product quality attributes. *Anal Bioanal Chem* **2020**, *412* (25), 6833-6848. DOI: 10.1007/s00216-020-02809-z
- (30) Quque, M.; Benhaim-Delarbre, M.; Deneubourg, J. L.; Sueur, C.; Criscuolo, F.; Bertile, F. Division of labour in the black garden ant (*Lasius niger*) leads to three distinct proteomes. *J Insect Physiol* **2019**, *117*, 103907. DOI: 10.1016/j.jinsphys.2019.103907
- (31) Macron, C.; Núñez Galindo, A.; Cominetti, O.; Dayon, L. A Versatile Workflow for Cerebrospinal Fluid Proteomic Analysis with Mass Spectrometry: A Matter of Choice between Deep Coverage and Sample Throughput. *Methods Mol Biol* **2019**, *2044*, 129-154. DOI: 10.1007/978-1-4939-9706-0_9
- (32) Olszowy, P.; Buszewski, B. Urine sample preparation for proteomic analysis. *J Sep Sci* **2014**, *37* (20), 2920-2928. DOI: 10.1002/jssc.201400331
- (33) Tascher, G.; Burbán, A.; Camus, S.; Plumel, M.; Chanon, S.; Le Guevel, R.; Shevchenko, V.; Van Dorselaer, A.; Lefai, E.; Guguen-Guillouzo, C.; Bertile, F. In-Depth Proteome Analysis Highlights HepaRG Cells as a Versatile Cell System Surrogate for Primary Human Hepatocytes. *Cells* **2019**, *8* (2). DOI: 10.3390/cells8020192
- (34) Smith, K. M.; Xu, Y. Tissue sample preparation in bioanalytical assays. *Bioanalysis* **2012**, *4* (6), 741-749. DOI: 10.4155/bio.12.19
- (35) Tanca, A.; Biosa, G.; Pagnozzi, D.; Addis, M. F.; Uzzau, S. Comparison of detergent-based sample preparation workflows for LTQ-Orbitrap analysis of the *Escherichia coli* proteome. *Proteomics* **2013**, *13* (17), 2597-2607. DOI: 10.1002/pmic.201200478
- (36) Feist, P.; Hummon, A. B. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **2015**, *16* (2), 3537-3563. DOI: 10.3390/ijms16023537
- (37) Issaq, H. J.; Conrads, T. P.; Janini, G. M.; Veenstra, T. D. Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis* **2002**, *23* (17), 3048-3061. DOI: 10.1002/1522-2683(200209)23:17<3048::Aid-elps3048>3.0.Co;2-1
- (38) Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **2002**, *1* (11), 845-867. DOI: 10.1074/mcp.r200007-mcp200
- (39) Ishikawa, H.; Kim, S.; Kwak, K.; Wakasugi, K.; Fayer, M. D. Disulfide bond influence on protein structural dynamics probed with 2D-IR vibrational echo spectroscopy. *Proc Natl Acad Sci U S A* **2007**, *104* (49), 19309-19314. DOI: 10.1073/pnas.0709760104
- (40) Vandermarliere, E.; Mueller, M.; Martens, L. Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom Rev* **2013**, *32* (6), 453-465. DOI: 10.1002/mas.21376
- (41) Switzar, L.; Giera, M.; Niessen, W. M. Protein digestion: an overview of the available techniques and recent developments. *J Proteome Res* **2013**, *12* (3), 1067-1077. DOI: 10.1021/pr301201x
- (42) Dau, T.; Bartolomucci, G.; Rappsilber, J. Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Anal Chem* **2020**, *92* (14), 9523-9527. DOI: 10.1021/acs.analchem.0c00478

- (43) Saveliev, S.; Bratz, M.; Zubarev, R.; Szapacs, M.; Budamgunta, H.; Urh, M. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nat Methods* **2013**, *10* (11), i-ii. DOI: 10.1038/nmeth.f.371
- (44) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **2006**, *1* (6), 2856-2860. DOI: 10.1038/nprot.2006.468
- (45) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **2009**, *6* (5), 359-362. DOI: 10.1038/nmeth.1322
- (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. DOI: 10.1002/pmic.201300553
- (47) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **2019**, *14* (1), 68-85. DOI: 10.1038/s41596-018-0082-x
- (48) Howard, J. W.; Kay, R. G.; Pleasance, S.; Creaser, C. S. UHPLC for the separation of proteins and peptides. *Bioanalysis* **2012**, *4* (24), 2971-2988. DOI: 10.4155/bio.12.283
- (49) Li, C.; Chu, S.; Tan, S.; Yin, X.; Jiang, Y.; Dai, X.; Gong, X.; Fang, X.; Tian, D. Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Front Chem* **2021**, *9*, 813359. DOI: 10.3389/fchem.2021.813359
- (50) 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. DOI: 10.1021/pr101060v
- (51) Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A. Fundamentals and Advances of Orbitrap Mass Spectrometry. In *Encyclopedia of Analytical Chemistry*, 2019; pp 1-40.
- (52) 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. DOI: 10.1016/j.jprot.2013.02.023
- (53) Marx, V. Targeted proteomics. *Nat Methods* **2013**, *10* (1), 19-22. DOI: 10.1038/nmeth.2285
- (54) Picotti, P.; Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **2012**, *9* (6), 555-566. DOI: 10.1038/nmeth.2015
- (55) Bourmaud, A.; Gallien, S.; Domon, B. Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. *Proteomics* **2016**, *16* (15-16), 2146-2159. DOI: 10.1002/pmic.201500543
- (56) Zhang, H.; Liu, Q.; Zimmerman, L. J.; Ham, A. J.; Slebos, R. J.; Rahman, J.; Kikuchi, T.; Massion, P. P.; Carbone, D. P.; Billheimer, D.; Liebner, D. C. Methods for peptide and protein quantitation by liquid chromatography-multiple reaction monitoring mass spectrometry. *Mol Cell Proteomics* **2011**, *10* (6), M110.006593. DOI: 10.1074/mcp.M110.006593
- (57) Biemann, K. Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol* **1990**, *193*, 886-887. DOI: 10.1016/0076-6879(90)93460-3
- (58) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Brechi, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **2000**, *35* (12), 1399-1406. DOI: 10.1002/1096-9888(200012)35:12<1399::Aid-jms86>3.0.Co;2-r
- (59) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J Am Chem Soc* **1996**, *118* (35), 8365-8374. DOI: 10.1021/ja9542193
- (60) Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R., 3rd. Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Anal Chem* **2004**, *76* (5), 1243-1248. DOI: 10.1021/ac0351163
- (61) 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. DOI: 10.1038/nmeth1060
- (62) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004**, *101* (26), 9528-9533. DOI: 10.1073/pnas.0402700101
- (63) McAlister, G. C.; Russell, J. D.; Rumachik, N. G.; Hebert, A. S.; Syka, J. E.; Geer, L. Y.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. Analysis of the acidic proteome with negative electron-transfer dissociation mass spectrometry. *Anal Chem* **2012**, *84* (6), 2875-2882. DOI: 10.1021/ac203430u
- (64) Wiesner, J.; Premisler, T.; Sickmann, A. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* **2008**, *8* (21), 4466-4483. DOI: 10.1002/pmic.200800329
- (65) Mommen, G. P.; Frese, C. K.; Meiring, H. D.; van Gaans-van den Brink, J.; de Jong, A. P.; van Els, C. A.; Heck, A. J. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc Natl Acad Sci U S A* **2014**, *111* (12), 4507-4512. DOI: 10.1073/pnas.1321458111

- (66) Zhokhov, S. S.; Kovalyov, S. V.; Samgina, T. Y.; Lebedev, A. T. An EThcD-Based Method for Discrimination of Leucine and Isoleucine Residues in Tryptic Peptides. *J Am Soc Mass Spectrom* **2017**, *28* (8), 1600-1611. DOI: 10.1007/s13361-017-1674-3
- (67) Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. F.; Mechtler, K.; Heck, A. J.; Mohammed, S. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res* **2013**, *12* (3), 1520-1525. DOI: 10.1021/pr301130k
- (68) Zubarev, R. A. Electron capture dissociation LC/MS/MS for bottom-up proteomics. *Methods Mol Biol* **2009**, *492*, 413-416. DOI: 10.1007/978-1-59745-493-3_25
- (69) Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem Rev* **2020**, *120* (7), 3328-3380. DOI: 10.1021/acs.chemrev.9b00440
- (70) Dodds, J. N.; Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J Am Soc Mass Spectrom* **2019**, *30* (11), 2185-2195. DOI: 10.1007/s13361-019-02288-2
- (71) Bonneil, E.; Pfammatter, S.; Thibault, P. Enhancement of mass spectrometry performance for proteomic analyses using high-field asymmetric waveform ion mobility spectrometry (FAIMS). *J Mass Spectrom* **2015**, *50* (11), 1181-1195. DOI: 10.1002/jms.3646
- (72) Cooper, H. J. To What Extent is FAIMS Beneficial in the Analysis of Proteins? *J Am Soc Mass Spectrom* **2016**, *27* (4), 566-577. DOI: 10.1007/s13361-015-1326-4
- (73) Fernandez-Lima, F.; Kaplan, D. A.; Suetering, J.; Park, M. A. Gas-phase separation using a trapped ion mobility spectrometer. *Int J Ion Mobil Spectrom* **2011**, *14* (2-3). DOI: 10.1007/s12127-011-0067-8
- (74) Fernandez-Lima, F. A.; Kaplan, D. A.; Park, M. A. Note: Integration of trapped ion mobility spectrometry with mass spectrometry. *Rev Sci Instrum* **2011**, *82* (12), 126106. DOI: 10.1063/1.3665933
- (75) Meier, F.; Park, M. A.; Mann, M. Trapped Ion Mobility Spectrometry and Parallel Accumulation-Serial Fragmentation in Proteomics. *Mol Cell Proteomics* **2021**, *20*, 100138. DOI: 10.1016/j.mcpro.2021.100138
- (76) Meier, F.; Brunner, A. D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; Bache, N.; Hoerning, O.; Cox, J.; Räther, O.; Mann, M. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* **2018**, *17* (12), 2534-2545. DOI: 10.1074/mcp.TIR118.000900
- (77) Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J. J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem* **2018**, *90* (15), 9529-9537. DOI: 10.1021/acs.analchem.8b02233
- (78) Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; Belford, M.; Dunyach, J. J.; Thibault, P. A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Mol Cell Proteomics* **2018**, *17* (10), 2051-2067. DOI: 10.1074/mcp.TIR118.000862
- (79) Blueggel, M.; Chamrad, D.; Meyer, H. E. Bioinformatics in proteomics. *Curr Pharm Biotechnol* **2004**, *5* (1), 79-88. DOI: 10.2174/1389201043489648
- (80) Consortium, T. U. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **2023**, *51* (D1), D523-d531. DOI: 10.1093/nar/gkac1052
- (81) Boutet, E.; Lieberherr, D.; Tognolli, M.; Schneider, M.; Bairoch, A. UniProtKB/Swiss-Prot. *Methods Mol Biol* **2007**, *406*, 89-112. DOI: 10.1007/978-1-59745-535-0_4
- (82) Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M. C.; Estreicher, A.; Gasteiger, E.; Martin, M. J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **2003**, *31* (1), 365-370. DOI: 10.1093/nar/gkg095
- (83) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretin, 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), D733-745. DOI: 10.1093/nar/gkv1189
- (84) Armengaud, J. A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr Opin Microbiol* **2009**, *12* (3), 292-300. DOI: 10.1016/j.mib.2009.03.005
- (85) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551-3567. DOI: 10.1002/(sici)1522-2683(19991201)20:18<3551::Aid-elps3551>3.0.Co;2-2

- (86) 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. DOI: 10.1021/pr101065j
- (87) 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. DOI: 10.1016/1044-0305(94)80016-2
- (88) Tabb, D. L. The SEQUEST family tree. *J Am Soc Mass Spectrom* **2015**, *26* (11), 1814-1819. DOI: 10.1007/s13361-015-1201-3
- (89) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **2013**, *13* (1), 22-24. DOI: 10.1002/pmic.201200439
- (90) 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. DOI: 10.1021/pr049882h
- (91) Griss, J.; Perez-Riverol, Y.; Lewis, S.; Tabb, D. L.; Dianes, J. A.; Del-Toro, N.; Rurik, M.; Walzer, M. W.; Kohlbacher, O.; Hermjakob, H.; Wang, R.; Vizcaíno, J. A. Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat Methods* **2016**, *13* (8), 651-656. DOI: 10.1038/nmeth.3902
- (92) Skinner, O. S.; Kelleher, N. L. Illuminating the dark matter of shotgun proteomics. *Nat Biotechnol* **2015**, *33* (7), 717-718. DOI: 10.1038/nbt.3287
- (93) 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. DOI: 10.1038/nbt.3267
- (94) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* **2017**, *14* (5), 513-520. DOI: 10.1038/nmeth.4256
- (95) Chi, H.; Liu, C.; Yang, H.; Zeng, W. F.; Wu, L.; Zhou, W. J.; Wang, R. M.; Niu, X. N.; Ding, Y. H.; Zhang, Y.; Wang, Z. W.; Chen, Z. L.; Sun, R. X.; Liu, T.; Tan, G. M.; Dong, M. Q.; Xu, P.; Zhang, P. H.; He, S. M. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat Biotechnol* **2018**. DOI: 10.1038/nbt.4236
- (96) Degroeve, S.; Gabriels, R.; Velghe, K.; Bouwmeester, R.; Tichshenko, N.; Martens, L. ionbot: a novel, innovative and sensitive machine learning approach to LC-MS/MS peptide identification. *bioRxiv* **2022**, 2021.2007.2002.450686. DOI: 10.1101/2021.07.02.450686
- (97) 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. DOI: 10.1038/s41592-021-01301-5
- (98) Degroeve, S.; Martens, L. MS2PIP: a tool for MS/MS peak intensity prediction. *Bioinformatics* **2013**, *29* (24), 3199-3203. DOI: 10.1093/bioinformatics/btt544
- (99) Gabriels, R.; Martens, L.; Degroeve, S. Updated MS²PIP web server delivers fast and accurate MS² peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques. *Nucleic Acids Res* **2019**, *47* (W1), W295-w299. DOI: 10.1093/nar/gkz299
- (100) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol Biol* **2010**, *604*, 55-71. DOI: 10.1007/978-1-60761-444-9_5
- (101) 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. DOI: 10.1038/nmeth1019
- (102) 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. DOI: 10.1038/nmeth1113
- (103) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J Proteome Res* **2008**, *7* (1), 40-44. DOI: 10.1021/pr700739d
- (104) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res* **2008**, *7* (1), 29-34. DOI: 10.1021/pr700600n
- (105) Storey, J. D.; Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **2003**, *100* (16), 9440-9445. DOI: 10.1073/pnas.1530509100
- (106) Desaire, H.; Go, E. P.; Hua, D. Advances, obstacles, and opportunities for machine learning in proteomics. *Cell Rep Phys Sci* **2022**, *3* (10). DOI: 10.1016/j.xcrp.2022.101069
- (107) Wen, B.; Zeng, W. F.; Liao, Y.; Shi, Z.; Savage, S. R.; Jiang, W.; Zhang, B. Deep Learning in Proteomics. *Proteomics* **2020**, *20* (21-22), e1900335. DOI: 10.1002/pmic.201900335
- (108) Moruz, L.; Staes, A.; Foster, J. M.; Hatzou, M.; Timmerman, E.; Martens, L.; Käll, L. Chromatographic retention time prediction for posttranslationally modified peptides. *Proteomics* **2012**, *12* (8), 1151-1159. DOI: 10.1002/pmic.201100386

- (109) Moruz, L.; Tomazela, D.; Käll, L. Training, selection, and robust calibration of retention time models for targeted proteomics. *J Proteome Res* **2010**, *9* (10), 5209-5216. DOI: 10.1021/pr1005058
- (110) Ma, C.; Ren, Y.; Yang, J.; Ren, Z.; Yang, H.; Liu, S. Improved Peptide Retention Time Prediction in Liquid Chromatography through Deep Learning. *Anal Chem* **2018**, *90* (18), 10881-10888. DOI: 10.1021/acs.analchem.8b02386
- (111) Barton, S. J.; Whittaker, J. C. Review of factors that influence the abundance of ions produced in a tandem mass spectrometer and statistical methods for discovering these factors. *Mass Spectrom Rev* **2009**, *28* (1), 177-187. DOI: 10.1002/mas.20188
- (112) Liu, K.; Li, S.; Wang, L.; Ye, Y.; Tang, H. Full-Spectrum Prediction of Peptides Tandem Mass Spectra using Deep Neural Network. *Anal Chem* **2020**, *92* (6), 4275-4283. DOI: 10.1021/acs.analchem.9b04867
- (113) Lin, Y. M.; Chen, C. T.; Chang, J. M. MS2CNN: predicting MS/MS spectrum based on protein sequence using deep convolutional neural networks. *BMC Genomics* **2019**, *20* (Suppl 9), 906. DOI: 10.1186/s12864-019-6297-6
- (114) Zeng, W. F.; Zhou, X. X.; Zhou, W. J.; Chi, H.; Zhan, J.; He, S. M. MS/MS Spectrum Prediction for Modified Peptides Using pDeep2 Trained by Transfer Learning. *Anal Chem* **2019**, *91* (15), 9724-9731. DOI: 10.1021/acs.analchem.9b01262
- (115) 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. DOI: 10.1021/acs.analchem.7b02566
- (116) 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. DOI: 10.1038/s41592-019-0426-7
- (117) Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndt, 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. DOI: 10.1038/s41592-019-0427-6
- (118) Guan, S.; Moran, M. F.; Ma, B. Prediction of LC-MS/MS Properties of Peptides from Sequence by Deep Learning. *Mol Cell Proteomics* **2019**, *18* (10), 2099-2107. DOI: 10.1074/mcp.TIR119.001412
- (119) Declercq, A.; Bouwmeester, R.; Hirschler, A.; Carapito, C.; Degroeve, S.; Martens, L.; Gabriels, R. MS(2)Rescore: Data-driven rescoring dramatically boosts immunopeptide identification rates. *Mol Cell Proteomics* **2022**, 100266. DOI: 10.1016/j.mcpro.2022.100266
- (120) Zolg, D. P.; Gessulat, S.; Paschke, C.; Graber, M.; Rathke-Kuhnert, M.; Seefried, F.; Fitzemeier, K.; Berg, F.; Lopez-Ferrer, D.; Horn, D.; Henrich, C.; Huhmer, A.; Delanghe, B.; Frejno, M. INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results. *Rapid Commun Mass Spectrom* **2021**, e9128. DOI: 10.1002/rcm.9128
- (121) Li, K.; Jain, A.; Malovannaya, A.; Wen, B.; Zhang, B. DeepRescore: Leveraging Deep Learning to Improve Peptide Identification in Immunopeptidomics. *Proteomics* **2020**, *20* (21-22), e1900334. DOI: 10.1002/pmic.201900334
- (122) Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **2012**, *404* (4), 939-965. DOI: 10.1007/s00216-012-6203-4
- (123) Ong, S. E. The expanding field of SILAC. *Anal Bioanal Chem* **2012**, *404* (4), 967-976. DOI: 10.1007/s00216-012-5998-3
- (124) 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. DOI: 10.1074/mcp.m200025-mcp200
- (125) 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. DOI: 10.1074/mcp.M114.040287
- (126) 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. DOI: 10.1038/nmeth.1446
- (127) 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. DOI: 10.1038/13690
- (128) 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. DOI: 10.1002/pmic.200700316

- (129) 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.; Bartlett-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. DOI: 10.1074/mcp.M400129-MCP200
- (130) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003**, *75* (8), 1895-1904. DOI: 10.1021/ac0262560
- (131) Li, J.; Cai, Z.; Bomgardner, R. D.; Pike, I.; Kuhn, K.; Rogers, J. C.; Roberts, T. M.; Gygi, S. P.; Paulo, J. A. TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. *J Proteome Res* **2021**, *20* (5), 2964-2972. DOI: 10.1021/acs.jproteome.1c00168
- (132) Ankney, J. A.; Muneer, A.; Chen, X. Relative and Absolute Quantitation in Mass Spectrometry-Based Proteomics. *Annu Rev Anal Chem (Palo Alto Calif)* **2018**, *11* (1), 49-77. DOI: 10.1146/annurev-anchem-061516-045357
- (133) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* **2011**, *8* (11), 937-940. DOI: 10.1038/nmeth.1714
- (134) 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. DOI: 10.1038/nprot.2017.040
- (135) Chelius, D.; Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **2002**, *1* (4), 317-323. DOI: 10.1021/pr025517j
- (136) Blein-Nicolas, M.; Zivy, M. Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochim Biophys Acta* **2016**, *1864* (8), 883-895. DOI: 10.1016/j.bbapap.2016.02.019
- (137) Pino, L. K.; Searle, B. C.; Huang, E. L.; Noble, W. S.; Hoofnagle, A. N.; MacCoss, M. J. Calibration Using a Single-Point External Reference Material Harmonizes Quantitative Mass Spectrometry Proteomics Data between Platforms and Laboratories. *Anal Chem* **2018**, *90* (21), 13112-13117. DOI: 10.1021/acs.analchem.8b04581
- (138) 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.; Cianféroni, S.; Garin, J.; Bruley, C. Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **2020**, *36* (10), 3148-3155. DOI: 10.1093/bioinformatics/btaa118
- (139) Cox, J.; Hein, M. Y.; Lubner, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **2014**, *13* (9), 2513-2526. DOI: 10.1074/mcp.M113.031591
- (140) 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. DOI: 10.1002/mas.21540
- (141) Ramus, C.; Hovasse, A.; Marcellin, M.; Hesse, A. M.; Mouton-Barbosa, E.; Bouyssié, D.; Vaca, S.; Carapito, C.; Chaoui, K.; Bruley, C.; Garin, J.; Cianféroni, 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. DOI: 10.1016/j.jprote.2015.11.011
- (142) Välikangas, T.; Suomi, T.; Elo, L. L. A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief Bioinform* **2018**, *19* (6), 1344-1355. DOI: 10.1093/bib/bbx054
- (143) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **2005**, *4* (9), 1265-1272. DOI: 10.1074/mcp.M500061-MCP200
- (144) Rappsilber, J.; Ryder, U.; Lamond, A. I.; Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **2002**, *12* (8), 1231-1245. DOI: 10.1101/gr.473902
- (145) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473* (7347), 337-342. DOI: 10.1038/nature10098
- (146) Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **2006**, *5* (1), 144-156. DOI: 10.1074/mcp.M500230-MCP200
- (147) Kitata, R. B.; Yang, J. C.; Chen, Y. J. Advances in data-independent acquisition mass spectrometry towards comprehensive digital proteome landscape. *Mass Spectrom Rev* **2022**, e21781. DOI: 10.1002/mas.21781

- (148) 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), e1900276. DOI: 10.1002/pmic.201900276
- (149) 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. DOI: 10.1021/ac991133+
- (150) Purvine, S.; Eppel, J. T.; Yi, E. C.; Goodlett, D. R. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **2003**, *3* (6), 847-850. DOI: 10.1002/pmic.200300362
- (151) 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. DOI: 10.1038/nmeth705
- (152) 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. DOI: 10.1021/ac048455k
- (153) 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. DOI: 10.1021/ac900888s
- (154) 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. DOI: 10.1074/mcp.M110.001537
- (155) Carvalho, P. C.; Han, X.; Xu, T.; Cociorva, D.; Carvalho Mda, G.; Barbosa, V. C.; Yates, J. R., 3rd. XDIA: improving on the label-free data-independent analysis. *Bioinformatics* **2010**, *26* (6), 847-848. DOI: 10.1093/bioinformatics/btq031
- (156) 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. DOI: 10.1038/nmeth.2528
- (157) Martin, L. B.; Sherwood, R. W.; Nicklay, J. J.; Yang, Y.; Muratore-Schroeder, T. L.; Anderson, E. T.; Thannhauser, T. W.; Rose, J. K.; 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. DOI: 10.1002/pmic.201500450
- (158) 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. DOI: 10.1021/pr5003017
- (159) 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. DOI: 10.1038/nmeth.2767
- (160) 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. DOI: 10.1074/mcp.M114.044305
- (161) 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. DOI: 10.1021/acs.jproteome.7b00464
- (162) 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.; Müllleder, M.; Tate, S.; Ralser, M. Ultra-fast proteomics with Scanning SWATH. *Nat Biotechnol* **2021**, *39* (7), 846-854. DOI: 10.1038/s41587-021-00860-4
- (163) 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. DOI: 10.1038/s41592-020-00998-0
- (164) Mun, D. G.; Renuse, S.; Saraswat, M.; Madugundu, A.; Udainiya, S.; Kim, H.; Park, S. R.; Zhao, H.; Nirujogi, R. S.; Na, C. H.; Kannan, N.; Yates, J. R., 3rd; Lee, S. W.; Pandey, A. PASS-DIA: A Data-Independent Acquisition Approach for Discovery Studies. *Anal Chem* **2020**, *92* (21), 14466-14475. DOI: 10.1021/acs.analchem.0c02513
- (165) 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. DOI: 10.1021/acs.jproteome.0c00381

- (166) Bekker-Jensen, D. B.; Bernhardt, O. M.; Högbe, A.; Martínez-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. DOI: 10.1038/s41467-020-14609-1
- (167) 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. DOI: 10.1021/acs.analchem.0c04293
- (168) Szyrwił, L.; Sinn, L.; Ralser, M.; Demichev, V. Slice-PASEF: fragmenting all ions for maximum sensitivity in proteomics. *bioRxiv* **2022**, 2022.2010.2031.514544. DOI: 10.1101/2022.10.31.514544
- (169) Martínez-Val, A.; Fort, K.; Koenig, C.; Van der Hoeven, L.; Franciosa, G.; Moehring, T.; Ishihama, Y.; Chen, Y. J.; Makarov, A.; Xuan, Y.; Olsen, J. V. Hybrid-DIA: intelligent data acquisition integrates targeted and discovery proteomics to analyze phospho-signaling in single spheroids. *Nat Commun* **2023**, *14* (1), 3599. DOI: 10.1038/s41467-023-39347-y
- (170) Distler, U.; Łacki, M. K.; Startek, M. P.; Teschner, D.; Brehmer, S.; Decker, J.; Schild, T.; Krieger, J.; Krohs, F.; Raether, O.; Hildebrandt, A.; Tenzer, S. midiaPASEF maximizes information content in data-independent acquisition proteomics. *bioRxiv* **2023**, 2023.2001.2030.526204. DOI: 10.1101/2023.01.30.526204
- (171) Zhang, Y.; Bilbao, A.; Bruderer, T.; Luban, J.; Strambio-De-Castilla, 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. DOI: 10.1021/acs.jproteome.5b00543
- (172) 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. DOI: 10.1074/mcp.P119.001913
- (173) Hu, A.; Noble, W. S.; Wolf-Yadlin, A. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res* **2016**, *5*. DOI: 10.12688/f1000research.7042.1
- (174) 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. DOI: 10.1074/mcp.O114.047035
- (175) Wu, J. X.; Song, X.; Pascovici, D.; Zaw, T.; Care, N.; Krisp, C.; Molloy, M. P. SWATH Mass Spectrometry Performance Using Extended Peptide MS/MS Assay Libraries. *Mol Cell Proteomics* **2016**, *15* (7), 2501-2514. DOI: 10.1074/mcp.M115.055558
- (176) 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*, 140031. DOI: 10.1038/sdata.2014.31
- (177) Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Edes, J.; Loevenich, S. N.; Aebersold, R. The PeptideAtlas project. *Nucleic Acids Res* **2006**, *34* (Database issue), D655-658. DOI: 10.1093/nar/gkj040
- (178) 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. DOI: 10.1093/nar/gkw936
- (179) 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. DOI: 10.1038/nbt.2841
- (180) 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. DOI: 10.1038/nmeth.4390
- (181) 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. DOI: 10.1038/s41467-018-07454-w
- (182) Van Puyvelde, B.; Willems, S.; Gabriels, R.; Daled, S.; De Clerck, L.; Vande Casteele, S.; Staes, A.; Impens, F.; Deforce, D.; Martens, L.; Degroove, S.; Dhaenens, M. Removing the Hidden Data Dependency of DIA with Predicted Spectral Libraries. *Proteomics* **2020**, *20* (3-4), e1900306. DOI: 10.1002/pmic.201900306
- (183) 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. DOI: 10.1038/s41592-019-0638-x
- (184) Tsou, C. C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A. C.; Nesvizhskii, A. I. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* **2015**, *12* (3), 258-264, 257 p following 264. DOI: 10.1038/nmeth.3255

- (185) Britt, H. M.; Cragolini, T.; Thalassinou, K. Integration of Mass Spectrometry Data for Structural Biology. *Chem Rev* **2022**, *122* (8), 7952-7986. DOI: 10.1021/acs.chemrev.1c00356
- (186) Dafun, A. S.; Marcoux, J. Structural mass spectrometry of membrane proteins. *Biochim Biophys Acta Proteins Proteom* **2022**, *1870* (8), 140813. DOI: 10.1016/j.bbapap.2022.140813
- (187) Wörner, T. P.; Shamorkina, T. M.; Snijder, J.; Heck, A. J. R. Mass Spectrometry-Based Structural Virology. *Anal Chem* **2021**, *93* (1), 620-640. DOI: 10.1021/acs.analchem.0c04339
- (188) Živković, D.; Sanchez Dafun, A.; Menneteau, T.; Schahl, A.; Lise, S.; Kervarrec, C.; Toste Rêgo, A.; da Fonseca, P. C. A.; Chavent, M.; Pineau, C.; Burret-Schiltz, O.; Marcoux, J.; Bousquet, M. P. Proteasome complexes experience profound structural and functional rearrangements throughout mammalian spermatogenesis. *Proc Natl Acad Sci U S A* **2022**, *119* (15), e2116826119. DOI: 10.1073/pnas.2116826119
- (189) Piersimoni, L.; Sinz, A. Cross-linking/mass spectrometry at the crossroads. *Anal Bioanal Chem* **2020**, *412* (24), 5981-5987. DOI: 10.1007/s00216-020-02700-x
- (190) Vimer, S.; Ben-Nissan, G.; Morgenstern, D.; Kumar-Deshmukh, F.; Polkinghorn, C.; Quintyn, R. S.; Vasil'ev, Y. V.; Beckman, J. S.; Elad, N.; Wysocki, V. H.; Sharon, M. Comparative Structural Analysis of 20S Proteasome Ortholog Protein Complexes by Native Mass Spectrometry. *ACS Cent Sci* **2020**, *6* (4), 573-588. DOI: 10.1021/acscentsci.0c00080
- (191) Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; Compton, P. D. Top-down characterization of endogenous protein complexes with native proteomics. *Nat Chem Biol* **2018**, *14* (1), 36-41. DOI: 10.1038/nchembio.2515
- (192) Bolla, J. R.; Agasid, M. T.; Mehmood, S.; Robinson, C. V. Membrane Protein-Lipid Interactions Probed Using Mass Spectrometry. *Annu Rev Biochem* **2019**, *88*, 85-111. DOI: 10.1146/annurev-biochem-013118-111508
- (193) Gault, J.; Liko, I.; Landreh, M.; Shutin, D.; Bolla, J. R.; Jefferies, D.; Agasid, M.; Yen, H. Y.; Ladds, M.; Lane, D. P.; Khalid, S.; Mullen, C.; Remes, P. M.; Huguette, R.; McAlister, G.; Goodwin, M.; Viner, R.; Syka, J. E. P.; Robinson, C. V. Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nat Methods* **2020**, *17* (5), 505-508. DOI: 10.1038/s41592-020-0821-0
- (194) Beveridge, R.; Calabrese, A. N. Structural Proteomics Methods to Interrogate the Conformations and Dynamics of Intrinsically Disordered Proteins. *Front Chem* **2021**, *9*, 603639. DOI: 10.3389/fchem.2021.603639
- (195) Largy, E.; Gabelica, V. Native Hydrogen/Deuterium Exchange Mass Spectrometry of Structured DNA Oligonucleotides. *Anal Chem* **2020**, *92* (6), 4402-4410. DOI: 10.1021/acs.analchem.9b05298
- (196) Lössl, P.; van de Waterbeemd, M.; Heck, A. J. The diverse and expanding role of mass spectrometry in structural and molecular biology. *Embo j* **2016**, *35* (24), 2634-2657. DOI: 10.15252/embj.201694818
- (197) Ganem, B.; Li, Y. T.; Henion, J. D. Detection of noncovalent receptor-ligand complexes by mass spectrometry. *J Am Chem Soc* **1991**, *113* (16), 6294-6296. DOI: 10.1021/ja00016a069
- (198) Katta, V.; Chait, B. T. Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J Am Chem Soc* **1991**, *113* (22), 8534-8535. DOI: 10.1021/ja00022a058
- (199) Heck, A. J.; Van Den Heuvel, R. H. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev* **2004**, *23* (5), 368-389. DOI: 10.1002/mas.10081
- (200) Leney, A. C.; Heck, A. J. Native Mass Spectrometry: What is in the Name? *J Am Soc Mass Spectrom* **2017**, *28* (1), 5-13. DOI: 10.1007/s13361-016-1545-3
- (201) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat Methods* **2012**, *9* (11), 1084-1086. DOI: 10.1038/nmeth.2208
- (202) Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G. Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J Am Soc Mass Spectrom* **2018**, *29* (1), 183-193. DOI: 10.1007/s13361-017-1799-4
- (203) Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A. Native top-down electrospray ionization-mass spectrometry of 158 kDa protein complex by high-resolution Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* **2014**, *86* (1), 317-320. DOI: 10.1021/ac4033214
- (204) Fort, K. L.; van de Waterbeemd, M.; Boll, D.; Reinhardt-Szyba, M.; Belov, M. E.; Sasaki, E.; Zschoche, R.; Hilvert, D.; Makarov, A. A.; Heck, A. J. R. Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. *Analyst* **2017**, *143* (1), 100-105. DOI: 10.1039/c7an01629h
- (205) Kafader, J. O.; Melani, R. D.; Schachner, L. F.; Ives, A. N.; Patrie, S. M.; Kelleher, N. L.; Compton, P. D. Native vs Denatured: An in Depth Investigation of Charge State and Isotope Distributions. *J Am Soc Mass Spectrom* **2020**, *31* (3), 574-581. DOI: 10.1021/jasms.9b00040
- (206) Deslignière, E.; Ley, M.; Bourguet, M.; Ehkirch, A.; Botzanowski, T.; Erb, S.; Hernandez-Alba, O.; Cianférani, S. Pushing the limits of native MS: Online SEC-native MS for structural biology applications. *International Journal of Mass Spectrometry* **2021**, *461*, 116502. DOI: 10.1016/j.ijms.2020.116502

- (207) van Schaick, G.; Haselberg, R.; Somsen, G. W.; Wührer, M.; Domínguez-Vega, E. Studying protein structure and function by native separation-mass spectrometry. *Nat Rev Chem* **2022**, *6* (3), 215-231. DOI: 10.1038/s41570-021-00353-7
- (208) Loo, J. A.; Berhane, B.; Kaddis, C. S.; Wooding, K. M.; Xie, Y.; Kaufman, S. L.; Chernushevich, I. V. Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *J Am Soc Mass Spectrom* **2005**, *16* (7), 998-1008. DOI: 10.1016/j.jasms.2005.02.017
- (209) Shepherd, D. A.; Ariza, A.; Edwards, T. A.; Barr, J. N.; Stonehouse, N. J.; Ashcroft, A. E. Probing bunyavirus N protein oligomerisation using mass spectrometry. *Rapid Commun Mass Spectrom* **2014**, *28* (7), 793-800. DOI: 10.1002/rcm.6841
- (210) Karch, K. R.; Snyder, D. T.; Harvey, S. R.; Wysocki, V. H. Native Mass Spectrometry: Recent Progress and Remaining Challenges. *Annu Rev Biophys* **2022**, *51*, 157-179. DOI: 10.1146/annurev-biophys-092721-085421
- (211) Harvey, S. R.; Liu, Y.; Liu, W.; Wysocki, V. H.; Laganowsky, A. Surface induced dissociation as a tool to study membrane protein complexes. *Chem Commun (Camb)* **2017**, *53* (21), 3106-3109. DOI: 10.1039/c6cc09606a
- (212) Rogniaux, H.; Van Dorsselaer, A.; Barth, P.; Biellmann, J. F.; Barbanton, J.; van Zandt, M.; Chevrier, B.; Howard, E.; Mitschler, A.; Potier, N.; Urzhumtseva, L.; Moras, D.; Podjarny, A. Binding of aldose reductase inhibitors: correlation of crystallographic and mass spectrometric studies. *J Am Soc Mass Spectrom* **1999**, *10* (7), 635-647. DOI: 10.1016/s1044-0305(99)00030-6
- (213) Ren, D.; Pipes, G. D.; Liu, D.; Shih, L. Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V. An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Anal Biochem* **2009**, *392* (1), 12-21. DOI: 10.1016/j.ab.2009.05.018
- (214) Tjernberg, A.; Carnö, S.; Oliv, F.; Benkestock, K.; Edlund, P. O.; Griffiths, W. J.; Hallén, D. Determination of dissociation constants for protein-ligand complexes by electrospray ionization mass spectrometry. *Anal Chem* **2004**, *76* (15), 4325-4331. DOI: 10.1021/ac0497914
- (215) El-Hawiet, A.; Kitova, E. N.; Klassen, J. S. Quantifying carbohydrate-protein interactions by electrospray ionization mass spectrometry analysis. *Biochemistry* **2012**, *51* (21), 4244-4253. DOI: 10.1021/bi300436x
- (216) Rostom, A. A.; Robinson, C. V. Detection of the Intact GroEL Chaperonin Assembly by Mass Spectrometry. *J Am Chem Soc* **1999**, *121* (19), 4718-4719. DOI: 10.1021/ja990238r
- (217) Rose, R. J.; Labrijn, A. F.; van den Bremer, E. T.; Loverix, S.; Lasters, I.; van Berkel, P. H.; van de Winkel, J. G.; Schuurman, J.; Parren, P. W.; Heck, A. J. Quantitative analysis of the interaction strength and dynamics of human IgG4 half molecules by native mass spectrometry. *Structure* **2011**, *19* (9), 1274-1282. DOI: 10.1016/j.str.2011.06.016
- (218) Sanglier, S.; Leize, E.; Van Dorsselaer, A.; Zal, F. Comparative ESI-MS study of approximately 2.2 MDa native hemocyanins from deep-sea and shore crabs: from protein oligomeric state to biotope. *J Am Soc Mass Spectrom* **2003**, *14* (5), 419-429. DOI: 10.1016/s1044-0305(03)00131-4
- (219) Shoemaker, G. K.; van Duijn, E.; Crawford, S. E.; Uetrecht, C.; Baclayon, M.; Roos, W. H.; Wuite, G. J.; Estes, M. K.; Prasad, B. V.; Heck, A. J. Norwalk virus assembly and stability monitored by mass spectrometry. *Mol Cell Proteomics* **2010**, *9* (8), 1742-1751. DOI: 10.1074/mcp.M900620-MCP200
- (220) Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V. Micelles protect membrane complexes from solution to vacuum. *Science* **2008**, *321* (5886), 243-246. DOI: 10.1126/science.1159292
- (221) Konijnenberg, A.; Butterer, A.; Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim Biophys Acta* **2013**, *1834* (6), 1239-1256. DOI: 10.1016/j.bbapap.2012.11.013
- (222) Clemmer, D. E.; Jarrold, M. F. Ion Mobility Measurements and their Applications to Clusters and Biomolecules. *Journal of Mass Spectrometry* **1997**, *32*, 577-592. DOI: 10.1002/(SICI)1096-9888(199706)32:6<577::AID-JMS530>3.0.CO;2-4
- (223) Mason, E. A.; Schamp, H. W. Mobility of gaseous ions in weak electric fields. *Annals of Physics* **1958**, *4* (3), 233-270. DOI: 10.1016/0003-4916(58)90049-6
- (224) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* **2004**, *18* (20), 2401-2414. DOI: 10.1002/rcm.1641
- (225) Lössl, P.; Brunner, A. M.; Liu, F.; Leney, A. C.; Yamashita, M.; Scheltema, R. A.; Heck, A. J. Deciphering the Interplay among Multisite Phosphorylation, Interaction Dynamics, and Conformational Transitions in a Tripartite Protein System. *ACS Cent Sci* **2016**, *2* (7), 445-455. DOI: 10.1021/acscentsci.6b00053
- (226) McCabe, J. W.; Hebert, M. J.; Shirzadeh, M.; Mallis, C. S.; Denton, J. K.; Walker, T. E.; Russell, D. H. THE IMS PARADOX: A PERSPECTIVE ON STRUCTURAL ION MOBILITY-MASS SPECTROMETRY. *Mass Spectrom Rev* **2021**, *40* (3), 280-305. DOI: 10.1002/mas.21642

- (227) Ben-Nissan, G.; Chotiner, A.; Tarnavsky, M.; Sharon, M. Structural Characterization of Missense Mutations Using High Resolution Mass Spectrometry: A Case Study of the Parkinson's-Related Protein, DJ-1. *J Am Soc Mass Spectrom* **2016**, *27* (6), 1062-1070. DOI: 10.1007/s13361-016-1379-z
- (228) Ciudad, S.; Puig, E.; Botzanowski, T.; Meigooni, M.; Arango, A. S.; Do, J.; Mayzel, M.; Bayoumi, M.; Chaignepain, S.; Maglia, G.; Cianferani, S.; Orekhov, V.; Tajkhorshid, E.; Bardiaux, B.; Carulla, N. A β (1-42) tetramer and octamer structures reveal edge conductivity pores as a mechanism for membrane damage. *Nat Commun* **2020**, *11* (1), 3014. DOI: 10.1038/s41467-020-16566-1
- (229) Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T. Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Curr Opin Chem Biol* **2018**, *42*, 93-100. DOI: 10.1016/j.cbpa.2017.11.010
- (230) Botzanowski, T.; Hernandez-Alba, O.; Malissard, M.; Wagner-Rousset, E.; Deslignière, E.; Colas, O.; Haeuw, J. F.; Beck, A.; Cianfèrani, S. Middle Level IM-MS and CIU Experiments for Improved Therapeutic Immunoglobulin Subclass Fingerprinting. *Anal Chem* **2020**, *92* (13), 8827-8835. DOI: 10.1021/acs.analchem.0c00293
- (231) Deslignière, E.; Botzanowski, T.; Diemer, H.; Cooper-Shepherd, D. A.; Wagner-Rousset, E.; Colas, O.; Béchade, G.; Giles, K.; Hernandez-Alba, O.; Beck, A.; Cianfèrani, S. High-Resolution IMS-MS to Assign Additional Disulfide Bridge Pairing in Complementarity-Determining Regions of an IgG4 Monoclonal Antibody. *J Am Soc Mass Spectrom* **2021**, *32* (10), 2505-2512. DOI: 10.1021/jasms.1c00151
- (232) Deslignière, E.; Ollivier, S.; Ehkirch, A.; Martelet, A.; Ropartz, D.; Lechat, N.; Hernandez-Alba, O.; Menet, J. M.; Clavier, S.; Rogniaux, H.; Genet, B.; Cianfèrani, S. Combination of IM-Based Approaches to Unravel the Coexistence of Two Conformers on a Therapeutic Multispecific mAb. *Anal Chem* **2022**, *94* (22), 7981-7989. DOI: 10.1021/acs.analchem.2c00928
- (233) Vallejo, D. D.; Rojas Ramírez, C.; Parson, K. F.; Han, Y.; Gadkari, V. V.; Ruotolo, B. T. Mass Spectrometry Methods for Measuring Protein Stability. *Chem Rev* **2022**, *122* (8), 7690-7719. DOI: 10.1021/acs.chemrev.1c00857
- (234) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M.; Hamuro, Y.; Heck, A. J. R.; Houde, D.; Jacob, R. E.; Jørgensen, T. J. D.; Kaltashov, I. A.; Klinman, J. P.; Konermann, L.; Man, P.; Mayne, L.; Pascal, B. D.; Reichmann, D.; Skehel, M.; Snijder, J.; Strutzenberg, T. S.; Underbakke, E. S.; Wagner, C.; Wales, T. E.; Walters, B. T.; Weis, D. D.; Wilson, D. J.; Wintrode, P. L.; Zhang, Z.; Zheng, J.; Schriemer, D. C.; Rand, K. D. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat Methods* **2019**, *16* (7), 595-602. DOI: 10.1038/s41592-019-0459-y
- (235) Engen, J. R.; Botzanowski, T.; Peterle, D.; Georgescauld, F.; Wales, T. E. Developments in Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal Chem* **2021**, *93* (1), 567-582. DOI: 10.1021/acs.analchem.0c04281
- (236) Pan, L. Y.; Salas-Solano, O.; Valliere-Douglass, J. F. Antibody structural integrity of site-specific antibody-drug conjugates investigated by hydrogen/deuterium exchange mass spectrometry. *Anal Chem* **2015**, *87* (11), 5669-5676. DOI: 10.1021/acs.analchem.5b00764
- (237) Ozohanics, O.; Ambrus, A. Hydrogen-Deuterium Exchange Mass Spectrometry: A Novel Structural Biology Approach to Structure, Dynamics and Interactions of Proteins and Their Complexes. *Life (Basel)* **2020**, *10* (11). DOI: 10.3390/life10110286
- (238) Masson, G. R.; Maslen, S. L.; Williams, R. L. Analysis of phosphoinositide 3-kinase inhibitors by bottom-up electron-transfer dissociation hydrogen/deuterium exchange mass spectrometry. *Biochem J* **2017**, *474* (11), 1867-1877. DOI: 10.1042/bcj20170127
- (239) Lee, S.; Wales, T. E.; Escudero, S.; Cohen, D. T.; Luccarelli, J.; Gallagher, C. G.; Cohen, N. A.; Huhn, A. J.; Bird, G. H.; Engen, J. R.; Walensky, L. D. Allosteric inhibition of antiapoptotic MCL-1. *Nat Struct Mol Biol* **2016**, *23* (6), 600-607. DOI: 10.1038/nsmb.3223
- (240) Deng, B.; Lento, C.; Wilson, D. J. Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development - A review. *Anal Chim Acta* **2016**, *940*, 8-20. DOI: 10.1016/j.aca.2016.08.006
- (241) Trabjerg, E.; Nazari, Z. E.; Rand, K. D. Conformational analysis of complex protein states by hydrogen/deuterium exchange mass spectrometry (HDX-MS): Challenges and emerging solutions. *TrAC Trends in Analytical Chemistry* **2018**, *106*, 125-138. DOI: 10.1016/j.trac.2018.06.008
- (242) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nat Protoc* **2018**, *13* (12), 2864-2889. DOI: 10.1038/s41596-018-0068-8
- (243) Belsom, A.; Rappsilber, J. Anatomy of a crosslinker. *Curr Opin Chem Biol* **2021**, *60*, 39-46. DOI: 10.1016/j.cbpa.2020.07.008

- (244) Götze, M.; Iacobucci, C.; Ihling, C. H.; Sinz, A. A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions. *Anal Chem* **2019**, *91* (15), 10236-10244. DOI: 10.1021/acs.analchem.9b02372
- (245) Liu, F.; Lössl, P.; Rabbitts, B. M.; Balaban, R. S.; Heck, A. J. R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Mol Cell Proteomics* **2018**, *17* (2), 216-232. DOI: 10.1074/mcp.RA117.000470
- (246) Linderstrøm-Lang, K. Structure and enzymatic break-down of proteins. *Cold Spring Harb Symp Quant Biol* **1950**, *14*, 117-126. DOI: 10.1101/sqb.1950.014.01.016
- (247) Hubbard, S. J. The structural aspects of limited proteolysis of native proteins. *Biochim Biophys Acta* **1998**, *1382* (2), 191-206. DOI: 10.1016/s0167-4838(97)00175-1
- (248) Bothner, B.; Dong, X. F.; Bibbs, L.; Johnson, J. E.; Siuzdak, G. Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J Biol Chem* **1998**, *273* (2), 673-676. DOI: 10.1074/jbc.273.2.673
- (249) Cappelletti, V.; Hauser, T.; Piazza, I.; Pepelnjak, M.; Malinowska, L.; Fuhrer, T.; Li, Y.; Dörig, C.; Boersema, P.; Gillet, L.; Grossbach, J.; Dugourd, A.; Saez-Rodriguez, J.; Beyer, A.; Zamboni, N.; Caflisch, A.; de Souza, N.; Picotti, P. Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **2021**, *184* (2), 545-559.e522. DOI: 10.1016/j.cell.2020.12.021
- (250) Piazza, I.; Kochanowski, K.; Cappelletti, V.; Fuhrer, T.; Noor, E.; Sauer, U.; Picotti, P. A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **2018**, *172* (1-2), 358-372.e323. DOI: 10.1016/j.cell.2017.12.006
- (251) Leuenberger, P.; Gansch, S.; Kahraman, A.; Cappelletti, V.; Boersema, P. J.; von Mering, C.; Claassen, M.; Picotti, P. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **2017**, *355* (6327). DOI: 10.1126/science.aai7825
- (252) Patrie, S. M. Top-Down Mass Spectrometry: Proteomics to Proteoforms. *Adv Exp Med Biol* **2016**, *919*, 171-200. DOI: 10.1007/978-3-319-41448-5_8
- (253) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun* **2014**, *445* (4), 683-693. DOI: 10.1016/j.bbrc.2014.02.041
- (254) Zhang, Z. Q.; Pan, H.; Chen, X. Y. MASS SPECTROMETRY FOR STRUCTURAL CHARACTERIZATION OF THERAPEUTIC ANTIBODIES. *Mass Spectrometry Reviews* **2009**, *28* (1), 147-176. DOI: 10.1002/mas.20190
- (255) Ren, D.; Pipes, G. D.; Liu, D. J.; Shih, L. Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V. An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Analytical Biochemistry* **2009**, *392* (1), 12-21, Article. DOI: 10.1016/j.ab.2009.05.018
- (256) Krokhin, O. V.; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G. Deamidation of -Asn-Gly-sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Analytical Chemistry* **2006**, *78* (18), 6645-6650, Article. DOI: 10.1021/ac061017o
- (257) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc* **2013**, *135* (34), 12646-12651. DOI: 10.1021/ja4029654
- (258) Fridriksson, E. K.; Beavil, A.; Holowka, D.; Gould, H. J.; Baird, B.; McLafferty, F. W. Heterogeneous glycosylation of immunoglobulin E constructs characterized by top-down high-resolution 2-D mass spectrometry. *Biochemistry* **2000**, *39* (12), 3369-3376. DOI: 10.1021/bi9919091
- (259) Reid, G. E.; Wu, J.; Chrisman, P. A.; Wells, J. M.; McLuckey, S. A. Charge-state-dependent sequence analysis of protonated ubiquitin ions via ion trap tandem mass spectrometry. *Anal Chem* **2001**, *73* (14), 3274-3281. DOI: 10.1021/ac0101095
- (260) Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W. Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J Am Chem Soc* **2002**, *124* (4), 672-678. DOI: 10.1021/ja011335z
- (261) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc Natl Acad Sci U S A* **2002**, *99* (4), 1774-1779. DOI: 10.1073/pnas.251691898
- (262) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J Am Chem Soc* **1999**, *121* (4), 806-812. DOI: 10.1021/ja973655h
- (263) Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A. Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *J Am Soc Mass Spectrom* **2019**, *30* (7), 1149-1157. DOI: 10.1007/s13361-019-02201-x
- (264) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* **2016**, *11* (5), 993-1006. DOI: 10.1038/nprot.2016.057
- (265) Todoroki, K.; Mizuno, H.; Sugiyama, E.; Toyo'oka, T. Bioanalytical methods for therapeutic monoclonal antibodies and antibody-drug conjugates: A review of recent advances and future perspectives. *J Pharm Biomed Anal* **2020**, *179*, 112991. DOI: 10.1016/j.jpba.2019.112991

- (266) Reid, G. E.; McLuckey, S. A. 'Top down' protein characterization via tandem mass spectrometry. *J Mass Spectrom* **2002**, *37* (7), 663-675. DOI: 10.1002/jms.346
- (267) Loo, J. A.; Quinn, J. P.; Ryu, S. I.; Henry, K. D.; Senko, M. W.; McLafferty, F. W. High-resolution tandem mass spectrometry of large biomolecules. *Proc Natl Acad Sci U S A* **1992**, *89* (1), 286-289. DOI: 10.1073/pnas.89.1.286
- (268) Vincent, D.; Mertens, D.; Rochfort, S. Optimisation of Milk Protein Top-Down Sequencing Using In-Source Collision-Induced Dissociation in the Maxis Quadrupole Time-of-Flight Mass Spectrometer. *Molecules* **2018**, *23* (11). DOI: 10.3390/molecules23112777
- (269) McLafferty, F. W.; Bryce, T. A. Metastable-ion characteristics: characterization of isomeric molecules. *Chemical Communications (London)* **1967**, (23), 1215-1217, 10.1039/C19670001215. DOI: 10.1039/C19670001215
- (270) Jennings, K. R. Collision-induced decompositions of aromatic molecular ions. *International Journal of Mass Spectrometry and Ion Physics* **1968**, *1* (3), 227-235. DOI: 10.1016/0020-7381(68)85002-8
- (271) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom Rev* **2018**, *37* (6), 750-771. DOI: 10.1002/mas.21560
- (272) Jones, A. W.; Cooper, H. J. Dissociation techniques in mass spectrometry-based proteomics. *Analyst* **2011**, *136* (17), 3419-3429. DOI: 10.1039/c0an01011a
- (273) Feng, Y.; De Franceschi, G.; Kahraman, A.; Soste, M.; Melnik, A.; Boersema, P. J.; de Laureto, P. P.; Nikolaev, Y.; Oliveira, A. P.; Picotti, P. Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol* **2014**, *32* (10), 1036-1044. DOI: 10.1038/nbt.2999
- (274) Sleno, L.; Volmer, D. A. Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* **2004**, *39* (10), 1091-1112. DOI: 10.1002/jms.703
- (275) Macias, L. A.; Santos, I. C.; Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal Chem* **2020**, *92* (1), 227-251. DOI: 10.1021/acs.analchem.9b04859
- (276) Leney, A. C. Subunit pI Can Influence Protein Complex Dissociation Characteristics. *J Am Soc Mass Spectrom* **2019**, *30* (8), 1389-1395. DOI: 10.1007/s13361-019-02198-3
- (277) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J Am Chem Soc* **1998**, *120* (13), 3265-3266. DOI: 10.1021/ja973478k
- (278) Breuker, K.; Oh, H.; Lin, C.; Carpenter, B. K.; McLafferty, F. W. Nonergodic and conformational control of the electron capture dissociation of protein cations. *Proc Natl Acad Sci U S A* **2004**, *101* (39), 14011-14016. DOI: 10.1073/pnas.0406095101
- (279) Syrstad, E. A.; Turecek, F. Toward a general mechanism of electron capture dissociation. *J Am Soc Mass Spectrom* **2005**, *16* (2), 208-224. DOI: 10.1016/j.jasms.2004.11.001
- (280) Sobczyk, M.; Anusiewicz, I.; Berdys-Kochanska, J.; Sawicka, A.; Skurski, P.; Simons, J. Coulomb-assisted dissociative electron attachment: application to a model peptide. *J Phys Chem A* **2005**, *109* (1), 250-258. DOI: 10.1021/jp0463114
- (281) Lermyte, F.; Łacki, M. K.; Valkenborg, D.; Baggerman, G.; Gambin, A.; Sobott, F. Understanding reaction pathways in top-down ETD by dissecting isotope distributions: A mammoth task. *International Journal of Mass Spectrometry* **2015**, *390*, 146-154. DOI: 10.1016/j.ijms.2015.08.008
- (282) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* **2000**, *72* (3), 563-573. DOI: 10.1021/ac990811p
- (283) Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A. Hydrogen rearrangement to and from radical z fragments in electron capture dissociation of peptides. *J Am Soc Mass Spectrom* **2007**, *18* (1), 113-120. DOI: 10.1016/j.jasms.2006.09.008
- (284) Ganisl, B.; Breuker, K. Does Electron Capture Dissociation Cleave Protein Disulfide Bonds? *ChemistryOpen* **2012**, *1* (6), 260-268. DOI: 10.1002/open.201200038
- (285) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **2007**, *6* (11), 1942-1951. DOI: 10.1074/mcp.M700073-MCP200
- (286) Martens, J.; Berden, G.; Oomens, J. Structures of Fluoranthene Reagent Anions Used in Electron Transfer Dissociation and Proton Transfer Reaction Tandem Mass Spectrometry. *Anal Chem* **2016**, *88* (12), 6126-6129. DOI: 10.1021/acs.analchem.6b01483
- (287) Hayakawa, S.; Hashimoto, M.; Matsubara, H.; Turecek, F. Dissecting the proline effect: dissociations of proline radicals formed by electron transfer to protonated Pro-Gly and Gly-Pro dipeptides in the gas phase. *J Am Chem Soc* **2007**, *129* (25), 7936-7949. DOI: 10.1021/ja0712571
- (288) McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K. Electron capture dissociation of gaseous multiply charged ions by fourier-transform ion cyclotron resonance. *J Am Soc Mass Spectrom* **2001**, *12* (3), 245-249. DOI: 10.1016/S1044-0305(00)00223-3

- (289) Lermyte, F.; Łacki, M. K.; Valkenborg, D.; Gambin, A.; Sobott, F. Conformational Space and Stability of ETD Charge Reduction Products of Ubiquitin. *J Am Soc Mass Spectrom* **2017**, *28* (1), 69-76. DOI: 10.1007/s13361-016-1444-7
- (290) Lermyte, F.; Williams, J. P.; Brown, J. M.; Martin, E. M.; Sobott, F. Extensive Charge Reduction and Dissociation of Intact Protein Complexes Following Electron Transfer on a Quadrupole-Ion Mobility-Time-of-Flight MS. *J Am Soc Mass Spectrom* **2015**, *26* (7), 1068-1076. DOI: 10.1007/s13361-015-1124-z
- (291) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal Chem* **2016**, *88* (1), 30-51. DOI: 10.1021/acs.analchem.5b04563
- (292) Horn, D. M.; Ge, Y.; McLafferty, F. W. Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal Chem* **2000**, *72* (20), 4778-4784. DOI: 10.1021/ac000494i
- (293) Ledvina, A. R.; Beauchene, N. A.; McAlister, G. C.; Syka, J. E. P.; Schwartz, J. C.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J. Activated-ion electron transfer dissociation improves the ability of electron transfer dissociation to identify peptides in a complex mixture. *Anal Chem* **2010**, *82* (24), 10068-10074. DOI: 10.1021/ac1020358
- (294) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal Chem* **2015**, *87* (14), 7109-7116. DOI: 10.1021/acs.analchem.5b00881
- (295) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. *J Proteome Res* **2017**, *16* (7), 2653-2659. DOI: 10.1021/acs.jproteome.7b00249
- (296) Papanastasiou, D.; Kounadis, D.; Lekkas, A.; Orfanopoulos, I.; Mpozatzidis, A.; Smyrnakis, A.; Panagiotopoulos, E.; Kosmopoulou, M.; Reinhardt-Szyba, M.; Fort, K.; Makarov, A.; Zubarev, R. A. The Omnitrap Platform: A Versatile Segmented Linear Ion Trap for Multidimensional Multiple-Stage Tandem Mass Spectrometry. *J Am Soc Mass Spectrom* **2022**, *33* (10), 1990-2007. DOI: 10.1021/jasms.2c00214
- (297) Baba, T.; Ryumin, P.; Duchoslav, E.; Chen, K.; Chelur, A.; Loyd, B.; Chernushevich, I. Dissociation of Biomolecules by an Intense Low-Energy Electron Beam in a High Sensitivity Time-of-Flight Mass Spectrometer. *J Am Soc Mass Spectrom* **2021**, *32* (8), 1964-1975. DOI: 10.1021/jasms.0c00425
- (298) Brodbelt, J. S. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem Soc Rev* **2014**, *43* (8), 2757-2783. DOI: 10.1039/c3cs60444f
- (299) Reilly, J. P. Ultraviolet photofragmentation of biomolecular ions. *Mass Spectrom Rev* **2009**, *28* (3), 425-447. DOI: 10.1002/mas.20214
- (300) Raspopov, S. A.; El-Faramawy, A.; Thomson, B. A.; Siu, K. W. Infrared multiphoton dissociation in quadrupole time-of-flight mass spectrometry: top-down characterization of proteins. *Anal Chem* **2006**, *78* (13), 4572-4577. DOI: 10.1021/ac052248i
- (301) Madsen, J. A.; Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C., Jr.; Brodbelt, J. S. Top-down protein fragmentation by infrared multiphoton dissociation in a dual pressure linear ion trap. *Anal Chem* **2009**, *81* (21), 8677-8686. DOI: 10.1021/ac901554z
- (302) Cui, W.; Thompson, M. S.; Reilly, J. P. Pathways of peptide ion fragmentation induced by vacuum ultraviolet light. *J Am Soc Mass Spectrom* **2005**, *16* (8), 1384-1398. DOI: 10.1016/j.jasms.2005.03.050
- (303) Julian, R. R. The Mechanism Behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions. *J Am Soc Mass Spectrom* **2017**, *28* (9), 1823-1826. DOI: 10.1007/s13361-017-1721-0
- (304) Antoine, R.; Dugourd, P. Visible and ultraviolet spectroscopy of gas phase protein ions. *Phys Chem Chem Phys* **2011**, *13* (37), 16494-16509. DOI: 10.1039/c1cp21531k
- (305) Sanders, J. D.; Mullen, C.; Watts, E.; Holden, D. D.; Syka, J. E. P.; Schwartz, J. C.; Brodbelt, J. S. Enhanced Sequence Coverage of Large Proteins by Combining Ultraviolet Photodissociation with Proton Transfer Reactions. *Anal Chem* **2020**, *92* (1), 1041-1049. DOI: 10.1021/acs.analchem.9b04026
- (306) Greer, S. M.; Sidoli, S.; Coradin, M.; Schack Jespersen, M.; Schwämmle, V.; Jensen, O. N.; Garcia, B. A.; Brodbelt, J. S. Extensive Characterization of Heavily Modified Histone Tails by 193 nm Ultraviolet Photodissociation Mass Spectrometry via a Middle-Down Strategy. *Anal Chem* **2018**, *90* (17), 10425-10433. DOI: 10.1021/acs.analchem.8b02320
- (307) Walker, J. N.; Lam, R.; Brodbelt, J. S. Enhanced Characterization of Histones Using 193 nm Ultraviolet Photodissociation and Proton Transfer Charge Reduction. *Anal Chem* **2023**, *95* (14), 5985-5993. DOI: 10.1021/acs.analchem.2c05765
- (308) Fornelli, L.; Srzentić, K.; Toby, T. K.; Doubleday, P. F.; Huguet, R.; Mullen, C.; Melani, R. D.; Dos Santos Seckler, H.; DeHart, C. J.; Weisbrod, C. R.; Durbin, K. R.; Greer, J. B.; Early, B. P.; Fellers, R. T.; Zabrouskov, V.; Thomas, P. M.; Compton, P. D.; Kelleher, N. L. Thorough Performance Evaluation of 213 nm Ultraviolet Photodissociation for Top-down Proteomics. *Mol Cell Proteomics* **2020**, *19* (2), 405-420. DOI: 10.1074/mcp.TIR119.001638

- (309) Greisch, J. F.; Tamara, S.; Scheltema, R. A.; Maxwell, H. W. R.; Fagerlund, R. D.; Fineran, P. C.; Tetter, S.; Hilvert, D.; Heck, A. J. R. Expanding the mass range for UVPD-based native top-down mass spectrometry. *Chem Sci* **2019**, *10* (30), 7163-7171. DOI: 10.1039/c9sc01857c
- (310) Lanzillotti, M.; Brodbelt, J. S. Comparison of Top-Down Protein Fragmentation Induced by 213 and 193 nm UVPD. *J Am Soc Mass Spectrom* **2023**. DOI: 10.1021/jasms.2c00288
- (311) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E.; Coon, J. J. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* **2007**, *79* (2), 477-485. DOI: 10.1021/ac061457f
- (312) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F.; Heck, A. J. Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization. *Anal Chem* **2015**, *87* (8), 4152-4158. DOI: 10.1021/acs.analchem.5b00162
- (313) Cannon, J. R.; Holden, D. D.; Brodbelt, J. S. Hybridizing ultraviolet photodissociation with electron transfer dissociation for intact protein characterization. *Anal Chem* **2014**, *86* (21), 10970-10977. DOI: 10.1021/ac5036082
- (314) Frese, C. K.; Altelaar, A. F.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem* **2012**, *84* (22), 9668-9673. DOI: 10.1021/ac3025366
- (315) Stephenson, J. L., Jr.; McLuckey, S. A. Simplification of product ion spectra derived from multiply charged parent ions via ion/ion chemistry. *Anal Chem* **1998**, *70* (17), 3533-3544. DOI: 10.1021/ac9802832
- (316) Huguet, R.; Mullen, C.; Srzentić, K.; Greer, J. B.; Fellers, R. T.; Zabrouskov, V.; Syka, J. E. P.; Kelleher, N. L.; Fornelli, L. Proton Transfer Charge Reduction Enables High-Throughput Top-Down Analysis of Large Proteoforms. *Anal Chem* **2019**, *91* (24), 15732-15739. DOI: 10.1021/acs.analchem.9b03925
- (317) Anderson, L. C.; English, A. M.; Wang, W.; Bai, D. L.; Shabanowitz, J.; Hunt, D. F. Protein derivatization and sequential ion/ion reactions to enhance sequence coverage produced by electron transfer dissociation mass spectrometry. *International Journal of Mass Spectrometry* **2015**, *377*, 617-624. DOI: 10.1016/j.ijms.2014.06.023
- (318) Kline, J. T.; Mullen, C.; Durbin, K. R.; Oates, R. N.; Huguet, R.; Syka, J. E. P.; Fornelli, L. Sequential Ion-Ion Reactions for Enhanced Gas-Phase Sequencing of Large Intact Proteins in a Tribrid Orbitrap Mass Spectrometer. *J Am Soc Mass Spectrom* **2021**, *32* (9), 2334-2345. DOI: 10.1021/jasms.1c00062
- (319) Zhong, J.; Sun, Y.; Xie, M.; Peng, W.; Zhang, C.; Wu, F. X.; Wang, J. Proteoform characterization based on top-down mass spectrometry. *Brief Bioinform* **2021**, *22* (2), 1729-1750. DOI: 10.1093/bib/bbaa015
- (320) Morgner, N.; Robinson, C. V. Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal Chem* **2012**, *84* (6), 2939-2948. DOI: 10.1021/ac300056a
- (321) Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A. Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol Cell Proteomics* **2010**, *9* (12), 2772-2782. DOI: 10.1074/mcp.M110.002766
- (322) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K.; Benesch, J. L.; Robinson, C. V. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal Chem* **2015**, *87* (8), 4370-4376. DOI: 10.1021/acs.analchem.5b00140
- (323) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J Am Soc Mass Spectrom* **2000**, *11* (4), 320-332. DOI: 10.1016/s1044-0305(99)00157-9
- (324) Avtonomov, D. M.; Polasky, D. A.; Ruotolo, B. T.; Nesvizhskii, A. I. IMTBX and Grppr: Software for Top-Down Proteomics Utilizing Ion Mobility-Mass Spectrometry. *Anal Chem* **2018**, *90* (3), 2369-2375. DOI: 10.1021/acs.analchem.7b04999
- (325) Tabb, D. L.; Jeong, K.; Druart, K.; Gant, M. S.; Brown, K. A.; Nicora, C.; Zhou, M.; Couvillion, S.; Nakayasu, E.; Williams, J. E.; Peterson, H. K.; McGuire, M. K.; McGuire, M. A.; Metz, T. O.; Chamot-Rooke, J. Comparing Top-Down Proteoform Identification: Deconvolution, PrSM Overlap, and PTM Detection. *J Proteome Res* **2023**, *22* (7), 2199-2217. DOI: 10.1021/acs.jproteome.2c00673
- (326) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: graphical software to analyze top-down mass spectrometry data. *Proteomics* **2015**, *15* (7), 1235-1238. DOI: 10.1002/pmic.201570050
- (327) Cai, W.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X.; Ge, Y. MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Mol Cell Proteomics* **2016**, *15* (2), 703-714. DOI: 10.1074/mcp.O115.054387

- (328) Kou, Q.; Wu, S.; Tolic, N.; Paša-Tolic, L.; Liu, Y.; Liu, X. A mass graph-based approach for the identification of modified proteoforms using top-down tandem mass spectra. *Bioinformatics* **2017**, *33* (9), 1309-1316. DOI: 10.1093/bioinformatics/btw806
- (329) Kou, Q.; Xun, L.; Liu, X. TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016**, *32* (22), 3495-3497. DOI: 10.1093/bioinformatics/btw398
- (330) Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat Biotechnol* **2001**, *19* (10), 952-957. DOI: 10.1038/nbt1001-952
- (331) Lyon, Y. A.; Riggs, D.; Fornelli, L.; Compton, P. D.; Julian, R. R. The Ups and Downs of Repeated Cleavage and Internal Fragment Production in Top-Down Proteomics. *J Am Soc Mass Spectrom* **2018**, *29* (1), 150-157. DOI: 10.1007/s13361-017-1823-8
- (332) Zenaidee, M. A.; Wei, B.; Lantz, C.; Wu, H. T.; Lambeth, T. R.; Diedrich, J. K.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A. Internal Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J Am Soc Mass Spectrom* **2021**, *32* (7), 1752-1758. DOI: 10.1021/jasms.1c00113
- (333) Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A. Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry. *Anal Chim Acta* **2022**, *1194*, 339400. DOI: 10.1016/j.aca.2021.339400
- (334) Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N. Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal Chem* **2021**, *93* (16), 6355-6362. DOI: 10.1021/acs.analchem.0c04670
- (335) Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A. ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J Proteome Res* **2021**, *20* (4), 1928-1935. DOI: 10.1021/acs.jproteome.0c00952
- (336) Wei, B.; Lantz, C.; Liu, W.; Viner, R.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A. Added Value of Internal Fragments for Top-Down Mass Spectrometry of Intact Monoclonal Antibodies and Antibody-Drug Conjugates. *Anal Chem* **2023**. DOI: 10.1021/acs.analchem.3c01426
- (337) Dunham, S. D.; Wei, B.; Lantz, C.; Loo, J. A.; Brodbelt, J. S. Impact of Internal Fragments on Top-Down Analysis of Intact Proteins by 193 nm UVPD. *J Proteome Res* **2022**, *22* (1), 170-181. DOI: 10.1021/acs.jproteome.2c00583
- (338) Toby, T. K.; Fornelli, L.; Srzentić, K.; DeHart, C. J.; Levitsky, J.; Friedewald, J.; Kelleher, N. L. A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nat Protoc* **2019**, *14* (1), 119-152. DOI: 10.1038/s41596-018-0085-7
- (339) Cai, W.; Tucholski, T. M.; Gregorich, Z. R.; Ge, Y. Top-down Proteomics: Technology Advancements and Applications to Heart Diseases. *Expert Rev Proteomics* **2016**, *13* (8), 717-730. DOI: 10.1080/14789450.2016.1209414
- (340) Gregorich, Z. R.; Ge, Y. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics* **2014**, *14* (10), 1195-1210. DOI: 10.1002/pmic.201300432
- (341) Contini, C.; Fadda, L.; Lai, G.; Masala, C.; Olianias, A.; Castagnola, M.; Messana, I.; Iavarone, F.; Bizzarro, A.; Masullo, C.; Solla, P.; Defazio, G.; Manconi, B.; Diaz, G.; Cabras, T. A top-down proteomic approach reveals a salivary protein profile able to classify Parkinson's disease with respect to Alzheimer's disease patients and to healthy controls. *Proteomics* **2023**, e2300202. DOI: 10.1002/pmic.202300202
- (342) Contini, C.; Olianias, A.; Serrao, S.; Deriu, C.; Iavarone, F.; Boroumand, M.; Bizzarro, A.; Lauria, A.; Faa, G.; Castagnola, M.; Messana, I.; Manconi, B.; Masullo, C.; Cabras, T. Top-Down Proteomics of Human Saliva Highlights Anti-inflammatory, Antioxidant, and Antimicrobial Defense Responses in Alzheimer Disease. *Front Neurosci* **2021**, *15*, 668852. DOI: 10.3389/fnins.2021.668852
- (343) Chapman, E. A.; Aballo, T. J.; Melby, J. A.; Zhou, T.; Price, S. J.; Rossler, K. J.; Lei, I.; Tang, P. C.; Ge, Y. Defining the Sarcomeric Proteoform Landscape in Ischemic Cardiomyopathy by Top-Down Proteomics. *J Proteome Res* **2023**, *22* (3), 931-941. DOI: 10.1021/acs.jproteome.2c00729
- (344) Tiambeng, T. N.; Tucholski, T.; Wu, Z.; Zhu, Y.; Mitchell, S. D.; Roberts, D. S.; Jin, Y.; Ge, Y. Analysis of cardiac troponin proteoforms by top-down mass spectrometry. *Methods Enzymol* **2019**, *626*, 347-374. DOI: 10.1016/bs.mie.2019.07.029
- (345) Zhou, M.; Wu, S.; Stenoién, D. L.; Zhang, Z.; Connolly, L.; Freitag, M.; Paša-Tolić, L. Profiling Changes in Histone Post-translational Modifications by Top-Down Mass Spectrometry. *Methods Mol Biol* **2017**, *1507*, 153-168. DOI: 10.1007/978-1-4939-6518-2_12
- (346) Cleland, T. P.; DeHart, C. J.; Fellers, R. T.; VanNispen, A. J.; Greer, J. B.; LeDuc, R. D.; Parker, W. R.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J Proteome Res* **2017**, *16* (5), 2072-2079. DOI: 10.1021/acs.jproteome.7b00043

- (347) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J Proteome Res* **2017**, *16* (2), 609-618. DOI: 10.1021/acs.jproteome.6b00698
- (348) 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. DOI: 10.1074/mcp.M113.030114
- (349) Melby, J. A.; Roberts, D. S.; Larson, E. J.; Brown, K. A.; Bayne, E. F.; Jin, S.; Ge, Y. Novel Strategies to Address the Challenges in Top-Down Proteomics. *J Am Soc Mass Spectrom* **2021**, *32* (6), 1278-1294. DOI: 10.1021/jasms.1c00099
- (350) Juetten, K. J.; Brodbelt, J. S. Top-Down Analysis of Supercharged Proteins Using Collision-, Electron-, and Photon-Based Activation Methods. *J Am Soc Mass Spectrom* **2023**, *34* (7), 1467-1476. DOI: 10.1021/jasms.3c00138
- (351) Riley, N. M.; Mullen, C.; Weisbrod, C. R.; Sharma, S.; Senko, M. W.; Zabrouskov, V.; Westphall, M. S.; Syka, J. E.; Coon, J. J. Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. *J Am Soc Mass Spectrom* **2016**, *27* (3), 520-531. DOI: 10.1007/s13361-015-1306-8
- (352) Rogers, H. T.; Roberts, D. S.; Larson, E. J.; Melby, J. A.; Rosslor, K. J.; Carr, A. V.; Brown, K. A.; Ge, Y. Comprehensive Characterization of Endogenous Phospholamban Proteoforms Enabled by Photocleavable Surfactant and Top-down Proteomics. *Anal Chem* **2023**. DOI: 10.1021/acs.analchem.3c01618
- (353) Lloyd-Jones, C.; Dos Santos Seckler, H.; DiStefano, N.; Sniderman, A.; Compton, P. D.; Kelleher, N. L.; Wilkins, J. T. Preparative Electrophoresis for HDL Particle Size Separation and Intact-Mass Apolipoprotein Proteoform Analysis. *J Proteome Res* **2023**, *22* (5), 1455-1465. DOI: 10.1021/acs.jproteome.2c00804
- (354) Dubois, C.; Payen, D.; Simon, S.; Junot, C.; Fenaille, F.; Morel, N.; Becher, F. Top-Down and Bottom-Up Proteomics of Circulating S100A8/S100A9 in Plasma of Septic Shock Patients. *J Proteome Res* **2020**, *19* (2), 914-925. DOI: 10.1021/acs.jproteome.9b00690
- (355) Dong, X.; Sumandea, C. A.; Chen, Y. C.; Garcia-Cazarin, M. L.; Zhang, J.; Balke, C. W.; Sumandea, M. P.; Ge, Y. Augmented phosphorylation of cardiac troponin I in hypertensive heart failure. *J Biol Chem* **2012**, *287* (2), 848-857. DOI: 10.1074/jbc.M111.293258
- (356) Dafun, A. S.; Živković, D.; Leon-Icaza, S. A.; Möller, S.; Froment, C.; Bonnet, D.; de Jesus, A. A.; Alric, L.; Quaranta-Nicaise, M.; Ferrand, A.; Cougoule, C.; Meunier, E.; Burlet-Schiltz, O.; Ebstein, F.; Goldbach-Mansky, R.; Krüger, E.; Bousquet, M. P.; Marcoux, J. Establishing 20S Proteasome Genetic, Translational and Post-Translational Status from Precious Biological and Patient Samples with Top-Down MS. *Cells* **2023**, *12* (6). DOI: 10.3390/cells12060844
- (357) Fort, K. L.; Cramer, C. N.; Voinov, V. G.; Vasil'ev, Y. V.; Lopez, N. I.; Beckman, J. S.; Heck, A. J. R. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J Proteome Res* **2018**, *17* (2), 926-933. DOI: 10.1021/acs.jproteome.7b00622
- (358) Zhou, M.; Lantz, C.; Brown, K. A.; Ge, Y.; Paša-Tolić, L.; Loo, J. A.; Lermyte, F. Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry. *Chem Sci* **2020**, *11* (48), 12918-12936. DOI: 10.1039/d0sc04392c
- (359) Juliano, B. R.; Keating, J. W.; Ruotolo, B. T. Infrared Photoactivation Enables Improved Native Top-Down Mass Spectrometry of Transmembrane Proteins. *Anal Chem* **2023**. DOI: 10.1021/acs.analchem.3c02788
- (360) Konijnenberg, A.; Bannwarth, L.; Yilmaz, D.; Koçer, A.; Venien-Bryan, C.; Sobott, F. Top-down mass spectrometry of intact membrane protein complexes reveals oligomeric state and sequence information in a single experiment. *Protein Sci* **2015**, *24* (8), 1292-1300. DOI: 10.1002/pro.2703
- (361) Zhang, J.; Reza Malmirchegini, G.; Clubb, R. T.; Loo, J. A. Native top-down mass spectrometry for the structural characterization of human hemoglobin. *Eur J Mass Spectrom (Chichester)* **2015**, *21* (3), 221-231. DOI: 10.1255/ejms.1340
- (362) Nshanian, M.; Lantz, C.; Wongkongkathep, P.; Schrader, T.; Klärner, F. G.; Blümke, A.; Despres, C.; Ehrmann, M.; Smet-Nocca, C.; Bitan, G.; Loo, J. A. Native Top-Down Mass Spectrometry and Ion Mobility Spectrometry of the Interaction of Tau Protein with a Molecular Tweezer Assembly Modulator. *J Am Soc Mass Spectrom* **2019**, *30* (1), 16-23. DOI: 10.1007/s13361-018-2027-6
- (363) Roberts, D. S.; Mann, M.; Melby, J. A.; Larson, E. J.; Zhu, Y.; Brasier, A. R.; Jin, S.; Ge, Y. Structural O-Glycoform Heterogeneity of the SARS-CoV-2 Spike Protein Receptor-Binding Domain Revealed by Top-Down Mass Spectrometry. *J Am Chem Soc* **2021**, *143* (31), 12014-12024. DOI: 10.1021/jacs.1c02713
- (364) Haverland, N. A.; Skinner, O. S.; Fellers, R. T.; Tariq, A. A.; Early, B. P.; LeDuc, R. D.; Fornelli, L.; Compton, P. D.; Kelleher, N. L. Defining Gas-Phase Fragmentation Propensities of Intact Proteins During Native Top-Down Mass Spectrometry. *J Am Soc Mass Spectrom* **2017**, *28* (6), 1203-1215. DOI: 10.1007/s13361-017-1635-x

- (365) Lermyte, F.; Sobott, F. Electron transfer dissociation provides higher-order structural information of native and partially unfolded protein complexes. *Proteomics* **2015**, *15* (16), 2813-2822. DOI: 10.1002/pmic.201400516
- (366) Wongkongkathep, P.; Han, J. Y.; Choi, T. S.; Yin, S.; Kim, H. I.; Loo, J. A. Native Top-Down Mass Spectrometry and Ion Mobility MS for Characterizing the Cobalt and Manganese Metal Binding of α -Synuclein Protein. *J Am Soc Mass Spectrom* **2018**, *29* (9), 1870-1880. DOI: 10.1007/s13361-018-2002-2
- (367) Litman, G. W.; Rast, J. P.; Fugmann, S. D. The origins of vertebrate adaptive immunity. *Nat Rev Immunol* **2010**, *10* (8), 543-553. DOI: 10.1038/nri2807
- (368) Hubert, P.; Amigorena, S. Antibody-dependent cell cytotoxicity in monoclonal antibody-mediated tumor immunotherapy. *Oncoimmunology* **2012**, *1* (1), 103-105. DOI: 10.4161/onci.1.1.17963
- (369) Weiskopf, K.; Weissman, I. L. Macrophages are critical effectors of antibody therapies for cancer. *MAbs* **2015**, *7* (2), 303-310. DOI: 10.1080/19420862.2015.1011450
- (370) Wang, B.; Yang, C.; Jin, X.; Du, Q.; Wu, H.; Dall'Acqua, W.; Mazor, Y. Regulation of antibody-mediated complement-dependent cytotoxicity by modulating the intrinsic affinity and binding valency of IgG for target antigen. *MAbs* **2020**, *12* (1), 1690959. DOI: 10.1080/19420862.2019.1690959
- (371) Kaplon, H.; Chenoweth, A.; Crescioli, S.; Reichert, J. M. Antibodies to watch in 2022. *MAbs* **2022**, *14* (1), 2014296. DOI: 10.1080/19420862.2021.2014296
- (372) *The Antibody Society. Therapeutic monoclonal antibodies approved or in review in the EU or US.* www.antibodysociety.org/resources/approved-antibodies (accessed 23/12/2022).
- (373) Reichert, J. M. Marketed therapeutic antibodies compendium. *MAbs* **2012**, *4* (3), 413-415. DOI: 10.4161/mabs.19931
- (374) www.grandviewresearch.com. *Monoclonal Antibodies Market Size, Share & Trends Analysis Report By Source Type (Chimeric, Murine, Humanized, Human), By Production Type (In Vivo, In Vitro), By Application, By End-use, By Region, And Segment Forecasts, 2023 - 2030.* 2023. <https://www.grandviewresearch.com/industry-analysis/monoclonal-antibodies-market> (accessed 2023/08/05).
- (375) Liu, L. Antibody glycosylation and its impact on the pharmacokinetics and pharmacodynamics of monoclonal antibodies and Fc-fusion proteins. *J Pharm Sci* **2015**, *104* (6), 1866-1884. DOI: 10.1002/jps.24444
- (376) Kang, T. H.; Jung, S. T. Boosting therapeutic potency of antibodies by taming Fc domain functions. *Exp Mol Med* **2019**, *51* (11), 1-9. DOI: 10.1038/s12276-019-0345-9
- (377) Shade, K.-T. C.; Anthony, R. M. Antibody Glycosylation and Inflammation. In *Antibodies*, 2013; Vol. 2, pp 392-414.
- (378) Vidarsson, G.; Dekkers, G.; Rispen, T. IgG subclasses and allotypes: from structure to effector functions. *Front Immunol* **2014**, *5*, 520. DOI: 10.3389/fimmu.2014.00520
- (379) Zheng, K.; Yarmarkovich, M.; Bantog, C.; Bayer, R.; Patapoff, T. W. Influence of glycosylation pattern on the molecular properties of monoclonal antibodies. *MAbs* **2014**, *6* (3), 649-658. DOI: 10.4161/mabs.28588
- (380) Irani, V.; Guy, A. J.; Andrew, D.; Beeson, J. G.; Ramsland, P. A.; Richards, J. S. Molecular properties of human IgG subclasses and their implications for designing therapeutic monoclonal antibodies against infectious diseases. *Mol Immunol* **2015**, *67* (2 Pt A), 171-182. DOI: 10.1016/j.molimm.2015.03.255
- (381) Singh, S.; Kumar, N. K.; Dwiwedi, P.; Charan, J.; Kaur, R.; Sidhu, P.; Chugh, V. K. Monoclonal Antibodies: A Review. *Curr Clin Pharmacol* **2018**, *13* (2), 85-99. DOI: 10.2174/1574884712666170809124728
- (382) Costa, R. L. B.; Czerniecki, B. J. Clinical development of immunotherapies for HER2(+) breast cancer: a review of HER2-directed monoclonal antibodies and beyond. *NPJ Breast Cancer* **2020**, *6*, 10. DOI: 10.1038/s41523-020-0153-3
- (383) Martinelli, E.; De Palma, R.; Orbitura, M.; De Vita, F.; Ciardiello, F. Anti-epidermal growth factor receptor monoclonal antibodies in cancer therapy. *Clin Exp Immunol* **2009**, *158* (1), 1-9. DOI: 10.1111/j.1365-2249.2009.03992.x
- (384) de Faire, U.; Frostegård, J. Natural antibodies against phosphorylcholine in cardiovascular disease. *Ann N Y Acad Sci* **2009**, *1173*, 292-300. DOI: 10.1111/j.1749-6632.2009.04748.x
- (385) Tsumoto, K.; Isozaki, Y.; Yagami, H.; Tomita, M. Future perspectives of therapeutic monoclonal antibodies. *Immunotherapy* **2019**, *11* (2), 119-127. DOI: 10.2217/imt-2018-0130
- (386) Atzeni, F.; Doria, A.; Carrabba, M.; Turiel, M.; Sarzi-Puttini, P. Potential target of infliximab in autoimmune and inflammatory diseases. *Autoimmun Rev* **2007**, *6* (8), 529-536. DOI: 10.1016/j.autrev.2007.03.009
- (387) Chang, A. T.; Platt, J. L. The role of antibodies in transplantation. *Transplant Rev (Orlando)* **2009**, *23* (4), 191-198. DOI: 10.1016/j.trre.2009.06.002
- (388) Köhler, G.; Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **1975**, *256* (5517), 495-497. DOI: 10.1038/256495a0

- (389) Hwang, W. Y.; Foote, J. Immunogenicity of engineered antibodies. *Methods* **2005**, *36* (1), 3-10. DOI: 10.1016/j.ymeth.2005.01.001
- (390) Brinkmann, U.; Kontermann, R. E. The making of bispecific antibodies. *MAbs* **2017**, *9* (2), 182-212. DOI: 10.1080/19420862.2016.1268307
- (391) Brinkmann, U.; Kontermann, R. E. Bispecific antibodies. *Science* **2021**, *372* (6545), 916-917. DOI: 10.1126/science.abg1209
- (392) Kontermann, R. E.; Brinkmann, U. Bispecific antibodies. *Drug Discov Today* **2015**, *20* (7), 838-847. DOI: 10.1016/j.drudis.2015.02.008
- (393) Spiess, C.; Zhai, Q.; Carter, P. J. Alternative molecular formats and therapeutic applications for bispecific antibodies. *Mol Immunol* **2015**, *67* (2 Pt A), 95-106. DOI: 10.1016/j.molimm.2015.01.003
- (394) Goswami, S.; Wang, W.; Arakawa, T.; Ohtake, S. Developments and Challenges for mAb-Based Therapeutics. In *Antibodies*, 2013; Vol. 2, pp 452-500.
- (395) Duivelshof, B. L.; Murisier, A.; Camperi, J.; Fekete, S.; Beck, A.; Guillarme, D.; D'Atri, V. Therapeutic Fc-fusion proteins: Current analytical strategies. *J Sep Sci* **2021**, *44* (1), 35-62. DOI: 10.1002/jssc.202000765
- (396) Beck, A.; Terral, G.; Debaene, F.; Wagner-Rousset, E.; Marcoux, J.; Janin-Bussat, M. C.; Colas, O.; Van Dorsselaer, A.; Cianféroni, S. Cutting-edge mass spectrometry methods for the multi-level structural characterization of antibody-drug conjugates. *Expert Rev Proteomics* **2016**, *13* (2), 157-183. DOI: 10.1586/14789450.2016.1132167
- (397) Hamilton, G. S. Antibody-drug conjugates for cancer therapy: The technological and regulatory challenges of developing drug-biologic hybrids. *Biologicals* **2015**, *43* (5), 318-332. DOI: 10.1016/j.biologicals.2015.05.006
- (398) Strebhardt, K.; Ullrich, A. Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat Rev Cancer* **2008**, *8* (6), 473-480. DOI: 10.1038/nrc2394
- (399) Beck, A.; Goetsch, L.; Dumontet, C.; Corvaia, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nat Rev Drug Discov* **2017**, *16* (5), 315-337. DOI: 10.1038/nrd.2016.268
- (400) Peters, C.; Brown, S. Antibody-drug conjugates as novel anti-cancer chemotherapeutics. *Biosci Rep* **2015**, *35* (4). DOI: 10.1042/bsr20150089
- (401) Baah, S.; Laws, M.; Rahman, K. M. Antibody-Drug Conjugates-A Tutorial Review. *Molecules* **2021**, *26* (10). DOI: 10.3390/molecules26102943
- (402) Han, T. H.; Zhao, B. Absorption, distribution, metabolism, and excretion considerations for the development of antibody-drug conjugates. *Drug Metab Dispos* **2014**, *42* (11), 1914-1920. DOI: 10.1124/dmd.114.058586
- (403) Guidance for Industry Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products. (ICH), I. C. f. H. o. T. R. f. P. f. H. U., Ed.; 1999.
- (404) Camperi, J.; Goyon, A.; Guillarme, D.; Zhang, K.; Stella, C. Multi-dimensional LC-MS: the next generation characterization of antibody-based therapeutics by unified online bottom-up, middle-up and intact approaches. *Analyst* **2021**, *146* (3), 747-769. DOI: 10.1039/d0an01963a
- (405) Beck, A.; D'Atri, V.; Ehkirch, A.; Fekete, S.; Hernandez-Alba, O.; Gahoual, R.; Leize-Wagner, E.; François, Y.; Guillarme, D.; Cianféroni, S. Cutting-edge multi-level analytical and structural characterization of antibody-drug conjugates: present and future. *Expert Rev Proteomics* **2019**, *16* (4), 337-362. DOI: 10.1080/14789450.2019.1578215
- (406) Kaltashov, I. A.; Bobst, C. E.; Pawlowski, J.; Wang, G. Mass spectrometry-based methods in characterization of the higher order structure of protein therapeutics. *J Pharm Biomed Anal* **2020**, *184*, 113169. DOI: 10.1016/j.jpba.2020.113169
- (407) Debaene, F.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaia, N.; Van Dorsselaer, A.; Beck, A.; Cianféroni, S. Time resolved native ion-mobility mass spectrometry to monitor dynamics of IgG4 Fab arm exchange and "bispecific" monoclonal antibody formation. *Anal Chem* **2013**, *85* (20), 9785-9792. DOI: 10.1021/ac402237v
- (408) Wagner-Rousset, E.; Janin-Bussat, M. C.; Colas, O.; Excoffier, M.; Ayoub, D.; Haeuw, J. F.; Rilatt, I.; Perez, M.; Corvaia, N.; Beck, A. Antibody-drug conjugate model fast characterization by LC-MS following IdeS proteolytic digestion. *MAbs* **2014**, *6* (1), 273-285. DOI: 10.4161/mabs.26773
- (409) Bongers, J.; Cummings, J. J.; Ebert, M. B.; Federici, M. M.; Gledhill, L.; Gulati, D.; Hilliard, G. M.; Jones, B. H.; Lee, K. R.; Mozdzanowski, J.; Naimoli, M.; Burman, S. Validation of a peptide mapping method for a therapeutic monoclonal antibody: what could we possibly learn about a method we have run 100 times? *J Pharm Biomed Anal* **2000**, *21* (6), 1099-1128. DOI: 10.1016/s0731-7085(99)00181-8
- (410) Janin-Bussat, M. C.; Dillenbourg, M.; Corvaia, N.; Beck, A.; Klinguer-Hamour, C. Characterization of antibody drug conjugate positional isomers at cysteine residues by peptide mapping LC-MS analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **2015**, *981-982*, 9-13. DOI: 10.1016/j.jchromb.2014.12.017

- (411) Millán-Martín, S.; Jakes, C.; Carillo, S.; Bones, J. Multi-attribute method (MAM) to assess analytical comparability of adalimumab biosimilars. *J Pharm Biomed Anal* **2023**, *234*, 115543. DOI: 10.1016/j.jpba.2023.115543
- (412) Butré, C. I.; D'Atri, V.; Diemer, H.; Colas, O.; Wagner, E.; Beck, A.; Cianferani, S.; Guillarme, D.; Delobel, A. Interlaboratory Evaluation of a User-Friendly Benchtop Mass Spectrometer for Multiple-Attribute Monitoring Studies of a Monoclonal Antibody. *Molecules* **2023**, *28* (6). DOI: 10.3390/molecules28062855
- (413) Debaene, F.; Boeuf, A.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaia, N.; Van Dorsselaer, A.; Beck, A.; Cianferani, S. Innovative native MS methodologies for antibody drug conjugate characterization: High resolution native MS and IM-MS for average DAR and DAR distribution assessment. *Anal Chem* **2014**, *86* (21), 10674-10683. DOI: 10.1021/ac502593n
- (414) Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S. Characterization of therapeutic antibodies and related products. *Anal Chem* **2013**, *85* (2), 715-736. DOI: 10.1021/ac3032355
- (415) Farsang, E.; Guillarme, D.; Veuthey, J. L.; Beck, A.; Lauber, M.; Schmudlach, A.; Fekete, S. Coupling non-denaturing chromatography to mass spectrometry for the characterization of monoclonal antibodies and related products. *J Pharm Biomed Anal* **2020**, *185*, 113207. DOI: 10.1016/j.jpba.2020.113207
- (416) Goyon, A.; Fekete, S.; Beck, A.; Veuthey, J. L.; Guillarme, D. Unraveling the mysteries of modern size exclusion chromatography - the way to achieve confident characterization of therapeutic proteins. *J Chromatogr B Analyt Technol Biomed Life Sci* **2018**, *1092*, 368-378. DOI: 10.1016/j.jchromb.2018.06.029
- (417) Fekete, S.; Veuthey, J. L.; Beck, A.; Guillarme, D. Hydrophobic interaction chromatography for the characterization of monoclonal antibodies and related products. *J Pharm Biomed Anal* **2016**, *130*, 3-18. DOI: 10.1016/j.jpba.2016.04.004
- (418) Fekete, S.; Beck, A.; Veuthey, J. L.; Guillarme, D. Ion-exchange chromatography for the characterization of biopharmaceuticals. *J Pharm Biomed Anal* **2015**, *113*, 43-55. DOI: 10.1016/j.jpba.2015.02.037
- (419) Füssl, F.; Trappe, A.; Carillo, S.; Jakes, C.; Bones, J. Comparative Elucidation of Cetuximab Heterogeneity on the Intact Protein Level by Cation Exchange Chromatography and Capillary Electrophoresis Coupled to Mass Spectrometry. *Anal Chem* **2020**, *92* (7), 5431-5438. DOI: 10.1021/acs.analchem.0c00185
- (420) Trappe, A.; Füssl, F.; Carillo, S.; Zaborowska, I.; Meleady, P.; Bones, J. Rapid charge variant analysis of monoclonal antibodies to support lead candidate biopharmaceutical development. *J Chromatogr B Analyt Technol Biomed Life Sci* **2018**, *1095*, 166-176. DOI: 10.1016/j.jchromb.2018.07.037
- (421) Liu, A. P.; Yan, Y.; Wang, S.; Li, N. Coupling Anion Exchange Chromatography with Native Mass Spectrometry for Charge Heterogeneity Characterization of Monoclonal Antibodies. *Anal Chem* **2022**, *94* (16), 6355-6362. DOI: 10.1021/acs.analchem.2c00707
- (422) Grodzki, A. C.; Berenstein, E. Antibody purification: affinity chromatography - protein A and protein G Sepharose. *Methods Mol Biol* **2010**, *588*, 33-41. DOI: 10.1007/978-1-59745-324-0_5
- (423) D'Atri, V.; Fekete, S.; Beck, A.; Lauber, M.; Guillarme, D. Hydrophilic Interaction Chromatography Hyphenated with Mass Spectrometry: A Powerful Analytical Tool for the Comparison of Originator and Biosimilar Therapeutic Monoclonal Antibodies at the Middle-up Level of Analysis. *Anal Chem* **2017**, *89* (3), 2086-2092. DOI: 10.1021/acs.analchem.6b04726
- (424) Singh, S. K.; Lee, K. H. Characterization of Monoclonal Antibody Glycan Heterogeneity Using Hydrophilic Interaction Liquid Chromatography-Mass Spectrometry. *Front Bioeng Biotechnol* **2021**, *9*, 805788. DOI: 10.3389/fbioe.2021.805788
- (425) Ehkirch, A.; Goyon, A.; Hernandez-Alba, O.; Rouviere, F.; D'Atri, V.; Dreyfus, C.; Haeuw, J. F.; Diemer, H.; Beck, A.; Heinisch, S.; Guillarme, D.; Cianferani, S. A Novel Online Four-Dimensional SEC×SEC-IM×MS Methodology for Characterization of Monoclonal Antibody Size Variants. *Anal Chem* **2018**, *90* (23), 13929-13937. DOI: 10.1021/acs.analchem.8b03333
- (426) Stoll, D. R. Recent advances in 2D-LC for bioanalysis. *Bioanalysis* **2015**, *7* (24), 3125-3142. DOI: 10.4155/bio.15.223
- (427) Deslignière, E.; Diemer, H.; Erb, S.; Coliat, P.; Pivot, X.; Detappe, A.; Hernandez-Alba, O.; Cianferani, S. A Combination of Native LC-MS Approaches for the Comprehensive Characterization of the Antibody-Drug Conjugate Trastuzumab Deruxtecan. *Front Biosci (Landmark Ed)* **2022**, *27* (10), 290. DOI: 10.31083/j.fbl2710290
- (428) Lodge, J. M.; Schauer, K. L.; Brademan, D. R.; Riley, N. M.; Shishkova, E.; Westphall, M. S.; Coon, J. J. Top-Down Characterization of an Intact Monoclonal Antibody Using Activated Ion Electron Transfer Dissociation. *Anal Chem* **2020**, *92* (15), 10246-10251. DOI: 10.1021/acs.analchem.0c00705
- (429) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol Cell Proteomics* **2012**, *11* (12), 1758-1767. DOI: 10.1074/mcp.M112.019620

- (430) Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; Wurm, F. M.; Hartmer, R. Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Anal Chem* **2011**, *83* (23), 8919-8927. DOI: 10.1021/ac201293m
- (431) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. Top-down structural analysis of an intact monoclonal antibody by electron capture dissociation-Fourier transform ion cyclotron resonance-mass spectrometry. *Anal Chem* **2013**, *85* (9), 4239-4246. DOI: 10.1021/ac303525n
- (432) Dhenin, J.; Dupré, M.; Druart, K.; Krick, A.; Mauriac, C.; Chamot-Rooke, J. A multiparameter optimization in middle-down analysis of monoclonal antibodies by LC-MS/MS. *J Mass Spectrom* **2023**, *58* (3), e4909. DOI: 10.1002/jms.4909
- (433) Melani, R. D.; Srzentić, K.; Gerbasi, V. R.; McGee, J. P.; Huguet, R.; Fornelli, L.; Kelleher, N. L. Direct measurement of light and heavy antibody chains using ion mobility and middle-down mass spectrometry. *MAbs* **2019**, *11* (8), 1351-1357. DOI: 10.1080/19420862.2019.1668226
- (434) Dadouch, M.; Ladner, Y.; Perrin, C. Analysis of Monoclonal Antibodies by Capillary Electrophoresis: Sample Preparation, Separation, and Detection. In *Separations*, 2021; Vol. 8.
- (435) Bobaly, B.; D'Atri, V.; Goyon, A.; Colas, O.; Beck, A.; Fekete, S.; Guillarme, D. Protocols for the analytical characterization of therapeutic monoclonal antibodies. II - Enzymatic and chemical sample preparation. *J Chromatogr B Analyt Technol Biomed Life Sci* **2017**, *1060*, 325-335. DOI: 10.1016/j.jchromb.2017.06.036
- (436) Formolo, T.; Ly, M.; Levy, M.; Kilpatrick, L.; Lute, S.; Phinney, K.; Marzilli, L.; Brorson, K.; Boyne, M.; Davis, D.; Schiel, J. Determination of the NISTmAb Primary Structure. In *State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 2. Biopharmaceutical Characterization: The NISTmAb Case Study*, ACS Symposium Series, Vol. 1201; American Chemical Society, 2015; pp 1-62.
- (437) Pipes, G. D.; Campbell, P.; Bondarenko, P. V.; Kerwin, B. A.; Treuheit, M. J.; Gadgil, H. S. Middle-down fragmentation for the identification and quantitation of site-specific methionine oxidation in an IgG1 molecule. *J Pharm Sci* **2010**, *99* (11), 4469-4476. DOI: 10.1002/jps.22158
- (438) Cotham, V. C.; Brodbelt, J. S. Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level using Middle-Down 193 nm Ultraviolet Photodissociation. *Anal Chem* **2016**, *88* (7), 4004-4013. DOI: 10.1021/acs.analchem.6b00302
- (439) Srzentić, K.; Fornelli, L.; Tsybin, Y. O.; Loo, J. A.; Seckler, H.; Agar, J. N.; Anderson, L. C.; Bai, D. L.; Beck, A.; Brodbelt, J. S.; van der Burgt, Y. E. M.; Chamot-Rooke, J.; Chatterjee, S.; Chen, Y.; Clarke, D. J.; Danis, P. O.; Diedrich, J. K.; D'Ippolito, R. A.; Dupré, M.; Gasilova, N.; Ge, Y.; Goo, Y. A.; Goodlett, D. R.; Greer, S.; Haselmann, K. F.; He, L.; Hendrickson, C. L.; Hinkle, J. D.; Holt, M. V.; Hughes, S.; Hunt, D. F.; Kelleher, N. L.; Kozhinov, A. N.; Lin, Z.; Malosse, C.; Marshall, A. G.; Menin, L.; Millikin, R. J.; Nagornov, K. O.; Nicolardi, S.; Paša-Tolić, L.; Pengelley, S.; Quebbemann, N. R.; Resemann, A.; Sandoval, W.; Sarin, R.; Schmitt, N. D.; Shabanowitz, J.; Shaw, J. B.; Shortreed, M. R.; Smith, L. M.; Sobott, F.; Suckau, D.; Toby, T.; Weisbrod, C. R.; Wildburger, N. C.; Yates, J. R., 3rd; Yoon, S. H.; Young, N. L.; Zhou, M. Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *J Am Soc Mass Spectrom* **2020**, *31* (9), 1783-1802. DOI: 10.1021/jasms.0c00036
- (440) Song, Y.; Gao, J.; Meng, Q.; Tang, F.; Wang, Y.; Zeng, Y.; Huang, W.; Shao, H.; Zhou, H. Conjugation site characterization of antibody-drug conjugates using electron-transfer/higher-energy collision dissociation (EThcD). *Anal Chim Acta* **2023**, *1251*, 340978. DOI: 10.1016/j.aca.2023.340978
- (441) Watts, E.; Williams, J. D.; Miesbauer, L. J.; Bruncko, M.; Brodbelt, J. S. Comprehensive Middle-Down Mass Spectrometry Characterization of an Antibody-Drug Conjugate by Combined Ion Activation Methods. *Anal Chem* **2020**, *92* (14), 9790-9798. DOI: 10.1021/acs.analchem.0c01232
- (442) Wei, B.; Zenaidee, M. A.; Lantz, C.; Williams, B. J.; Totten, S.; Ogorzalek Loo, R. R.; Loo, J. A. Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins. *Analyst* **2022**, *148* (1), 26-37. DOI: 10.1039/d2an01517j
- (443) Gstöttner, C.; Reusch, D.; Haberger, M.; Dragan, I.; Van Veelen, P.; Kilgour, D. P. A.; Tsybin, Y. O.; van der Burgt, Y. E. M.; Wuhrer, M.; Nicolardi, S. Monitoring glycation levels of a bispecific monoclonal antibody at subunit level by ultrahigh-resolution MALDI FT-ICR mass spectrometry. *MAbs* **2020**, *12* (1), 1682403. DOI: 10.1080/19420862.2019.1682403
- (444) Chen, B.; Lin, Z.; Zhu, Y.; Jin, Y.; Larson, E.; Xu, Q.; Fu, C.; Zhang, Z.; Zhang, Q.; Pritts, W. A.; Ge, Y. Middle-Down Multi-Attribute Analysis of Antibody-Drug Conjugates with Electron Transfer Dissociation. *Anal Chem* **2019**, *91* (18), 11661-11669. DOI: 10.1021/acs.analchem.9b02194
- (445) Tran, B. Q.; Barton, C.; Feng, J.; Sandjong, A.; Yoon, S. H.; Awasthi, S.; Liang, T.; Khan, M. M.; Kilgour, D. P. A.; Goodlett, D. R.; Goo, Y. A. Comprehensive glycosylation profiling of IgG and IgG-fusion proteins by top-down MS with multiple fragmentation techniques. *J Proteomics* **2016**, *134*, 93-101. DOI: 10.1016/j.jprot.2015.10.021

- (446) Zhu, W.; Li, M.; Zhang, J. Integrating Intact Mass Analysis and Middle-Down Mass Spectrometry Approaches to Effectively Characterize Trastuzumab and Adalimumab Structural Heterogeneity. *J Proteome Res* **2021**, *20* (1), 270-278. DOI: 10.1021/acs.jproteome.0c00373
- (447) Sgro, C. Side-effects of a monoclonal antibody, muromonab CD3/orthoclone OKT3: bibliographic review. *Toxicology* **1995**, *105* (1), 23-29. DOI: 10.1016/0300-483x(95)03123-w
- (448) Li, F.; Vijayasankaran, N.; Shen, A. Y.; Kiss, R.; Amanullah, A. Cell culture processes for monoclonal antibody production. *MAbs* **2010**, *2* (5), 466-479. DOI: 10.4161/mabs.2.5.12720
- (449) Morrison, S. L.; Johnson, M. J.; Herzenberg, L. A.; Oi, V. T. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc Natl Acad Sci U S A* **1984**, *81* (21), 6851-6855. DOI: 10.1073/pnas.81.21.6851
- (450) Jones, P. T.; Dear, P. H.; Foote, J.; Neuberger, M. S.; Winter, G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **1986**, *321* (6069), 522-525. DOI: 10.1038/321522a0
- (451) Kempeni, J. Preliminary results of early clinical trials with the fully human anti-TNFalpha monoclonal antibody D2E7. *Ann Rheum Dis* **1999**, *58 Suppl 1* (Suppl 1), I70-72. DOI: 10.1136/ard.58.2008.i70
- (452) Frenzel, A.; Hust, M.; Schirrmann, T. Expression of recombinant antibodies. *Front Immunol* **2013**, *4*, 217. DOI: 10.3389/fimmu.2013.00217
- (453) 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. DOI: 10.3109/07388551.2015.1084266
- (454) Kim, J. Y.; Kim, Y. G.; Lee, G. M. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl Microbiol Biotechnol* **2012**, *93* (3), 917-930. DOI: 10.1007/s00253-011-3758-5
- (455) Fischer, S.; Handrick, R.; Otte, K. The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnol Adv* **2015**, *33* (8), 1878-1896. DOI: 10.1016/j.biotechadv.2015.10.015
- (456) Hessmann, S. Development of analytical strategies in quantitative proteomic: quantitation of host cell proteins by mass spectrometry as a quality control tool for the biopharmaceutical industry. University of Strasbourg, 2021.
- (457) Kunert, R.; Reinhart, D. Advances in recombinant antibody manufacturing. *Appl Microbiol Biotechnol* **2016**, *100* (8), 3451-3461. DOI: 10.1007/s00253-016-7388-9
- (458) Liu, H. F.; Ma, J.; Winter, C.; Bayer, R. Recovery and purification process development for monoclonal antibody production. *MAbs* **2010**, *2* (5), 480-499. DOI: 10.4161/mabs.2.5.12645
- (459) 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. DOI: 10.1002/bit.26812
- (460) Chahar, D. S.; Ravindran, S.; Pisal, S. S. Monoclonal antibody purification and its progression to commercial scale. *Biologicals* **2020**, *63*, 1-13. DOI: 10.1016/j.biologicals.2019.09.007
- (461) Chon, J. H.; Zerbis-Papastoitis, G. Advances in the production and downstream processing of antibodies. *N Biotechnol* **2011**, *28* (5), 458-463. DOI: 10.1016/j.nbt.2011.03.015
- (462) Follman, D. K.; Fahrner, R. L. Factorial screening of antibody purification processes using three chromatography steps without protein A. *J Chromatogr A* **2004**, *1024* (1-2), 79-85. DOI: 10.1016/j.chroma.2003.10.060
- (463) Jensen, K. A NORMALLY OCCURRING STAPHYLOCOCCUS ANTIBODY IN HUMAN SERUM. *Acta Pathologica Microbiologica Scandinavica* **1958**, *44* (4), 421-428. DOI: 10.1111/j.1699-0463.1958.tb01093.x (accessed 2023/08/26).
- (464) Ramos-de-la-Peña, A. M.; González-Valdez, J.; Aguilar, O. Protein A chromatography: Challenges and progress in the purification of monoclonal antibodies. *J Sep Sci* **2019**, *42* (9), 1816-1827. DOI: 10.1002/jssc.201800963
- (465) Hober, S.; Nord, K.; Linhult, M. Protein A chromatography for antibody purification. *J Chromatogr B Analyt Technol Biomed Life Sci* **2007**, *848* (1), 40-47. DOI: 10.1016/j.jchromb.2006.09.030
- (466) Gronemeyer, P.; Ditz, R.; Strube, J. Trends in Upstream and Downstream Process Development for Antibody Manufacturing. *Bioengineering (Basel)* **2014**, *1* (4), 188-212. DOI: 10.3390/bioengineering1040188
- (467) Robert, F.; Bierau, H.; Rossi, M.; Agugiaro, D.; Soranzo, T.; Broly, H.; Mitchell-Logean, C. Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of a CHO cathepsin D protease. *Biotechnol Bioeng* **2009**, *104* (6), 1132-1141. DOI: 10.1002/bit.22494
- (468) Vanderlaan, M.; Zhu-Shimoni, J.; Lin, S.; Gunawan, F.; Waerner, T.; Van Cott, K. E. Experience with host cell protein impurities in biopharmaceuticals. *Biotechnol Prog* **2018**, *34* (4), 828-837. DOI: 10.1002/btpr.2640

- (469) Dixit, N.; Salamat-Miller, N.; Salinas, P. A.; Taylor, K. D.; Basu, S. K. Residual Host Cell Protein Promotes Polysorbate 20 Degradation in a Sulfatase Drug Product Leading to Free Fatty Acid Particles. *J Pharm Sci* **2016**, *105* (5), 1657-1666. DOI: 10.1016/j.xphs.2016.02.029
- (470) Li, X.; An, Y.; Liao, J.; Xiao, L.; Swanson, M.; Martinez-Fonts, K.; Pavon, J. A.; Sherer, E. C.; Jawa, V.; Wang, F.; Gao, X.; Letarte, S.; Richardson, D. D. Identification and characterization of a residual host cell protein hexosaminidase B associated with N-glycan degradation during the stability study of a therapeutic recombinant monoclonal antibody product. *Biotechnol Prog* **2021**, *37* (3), e3128. DOI: 10.1002/btpr.3128
- (471) Jones, M.; Palackal, N.; Wang, F.; Gaza-Bulsecu, 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. DOI: 10.1002/bit.27808
- (472) Jawa, V.; Joubert, M. K.; Zhang, Q.; Deshpande, M.; Hapuarachchi, S.; Hall, M. P.; Flynn, G. C. Evaluating Immunogenicity Risk Due to Host Cell Protein Impurities in Antibody-Based Biotherapeutics. *Aaps j* **2016**, *18* (6), 1439-1452. DOI: 10.1208/s12248-016-9948-4
- (473) <1132> Residual Host Cell Protein Measurement in Biopharmaceuticals. Pharmacopoeia, U. S., Ed.; 2016; Vol. 39.
- (474) DNA and host cell protein impurities, routine testing versus validation studies. **1997**, (CPMP/BWP/382/97).
- (475) 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. DOI: 10.1016/j.copbio.2014.06.017
- (476) Gilgunn, S.; El-Sabbahy, H.; Albrecht, S.; Gaikwad, M.; Corrigan, K.; Deakin, L.; Jellum, G.; Bones, J. Identification and tracking of problematic host cell proteins removed by a synthetic, highly functionalized nonwoven media in downstream bioprocessing of monoclonal antibodies. *J Chromatogr A* **2019**, *1595*, 28-38. DOI: 10.1016/j.chroma.2019.02.056
- (477) Bracewell, D. G.; Francis, R.; Smales, C. M. The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnol. Bioeng.* **2015**, *112* (9), 1727-1737. DOI: 10.1002/bit.25628
- (478) 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. DOI: 10.1021/acs.analchem.9b03952
- (479) Rane, S. S.; Dearman, R. J.; Kimber, I.; Uddin, S.; Bishop, S.; Shah, M.; Podmore, A.; Pluen, A.; Derrick, J. P. Impact of a Heat Shock Protein Impurity on the Immunogenicity of Biotherapeutic Monoclonal Antibodies. *Pharm Res* **2019**, *36* (4), 51. DOI: 10.1007/s11095-019-2586-7
- (480) 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. DOI: 10.4161/bioe.23382
- (481) Falkenberg, H.; Waldera-Lupa, D. M.; Vanderlaan, M.; Schwab, T.; Krapfenbauer, K.; Studts, J. M.; Flad, T.; Waerner, T. Mass spectrometric evaluation of upstream and downstream process influences on host cell protein patterns in biopharmaceutical products. *Biotechnol Prog* **2019**, *35* (3), e2788. DOI: 10.1002/btpr.2788
- (482) Valente, K. N.; Levy, N. E.; Lee, K. H.; Lenhoff, A. M. Applications of proteomic methods for CHO host cell protein characterization in biopharmaceutical manufacturing. *Curr Opin Biotechnol* **2018**, *53*, 144-150. DOI: 10.1016/j.copbio.2018.01.004
- (483) 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. DOI: 10.1002/bit.25327
- (484) 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. DOI: 10.1002/biot.201200018
- (485) Henry, S. M.; Sutlief, E.; Salas-Solano, O.; Valliere-Douglass, J. ELISA reagent coverage evaluation by affinity purification tandem mass spectrometry. *MAbs* **2017**, *9* (7), 1065-1075. DOI: 10.1080/19420862.2017.1349586
- (486) Wang, X.; Hunter, A. K.; Mozier, N. M. Host cell proteins in biologics development: Identification, quantitation and risk assessment. *Biotechnol Bioeng* **2009**, *103* (3), 446-458. DOI: 10.1002/bit.22304
- (487) 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. DOI: 10.1002/bit.26466
- (488) Zhu, M.; Gong, X.; Hu, Y.; Ou, W.; Wan, Y. Streptavidin-biotin-based directional double Nanobody sandwich ELISA for clinical rapid and sensitive detection of influenza H5N1. *J Transl Med* **2014**, *12*, 352. DOI: 10.1186/s12967-014-0352-5

- (489) Peng, J.; Song, S.; Xu, L.; Ma, W.; Liu, L.; Kuang, H.; Xu, C. Development of a monoclonal antibody-based sandwich ELISA for peanut allergen Ara h 1 in food. *Int J Environ Res Public Health* **2013**, *10* (7), 2897-2905. DOI: 10.3390/ijerph10072897
- (490) Wang, S. Y.; Li, Z.; Wang, X. J.; Lv, S.; Yang, Y.; Zeng, L. Q.; Luo, F. H.; Yan, J. H.; Liang, D. F. Development of monoclonal antibody-based sandwich ELISA for detection of dextran. *Monoclon Antib Immunodiagn Immunother* **2014**, *33* (5), 334-339. DOI: 10.1089/mab.2014.0014
- (491) MacPhee, D. J. Methodological considerations for improving Western blot analysis. *J Pharmacol Toxicol Methods* **2010**, *61* (2), 171-177. DOI: 10.1016/j.vascn.2009.12.001
- (492) Mahmood, T.; Yang, P. C. Western blot: technique, theory, and trouble shooting. *N Am J Med Sci* **2012**, *4* (9), 429-434. DOI: 10.4103/1947-2714.100998
- (493) Magdeldin, S.; Enany, S.; Yoshida, Y.; Xu, B.; Zhang, Y.; Zureena, Z.; Lokamani, I.; Yaoita, E.; Yamamoto, T. Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. *Clin Proteomics* **2014**, *11* (1), 16. DOI: 10.1186/1559-0275-11-16
- (494) Simpson, R. J. Rapid coomassie blue staining of protein gels. *Cold Spring Harb Protoc* **2010**, *2010* (4), pdb.prot5413. DOI: 10.1101/pdb.prot5413
- (495) Chevallet, M.; Luche, S.; Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nat Protoc* **2006**, *1* (4), 1852-1858. DOI: 10.1038/nprot.2006.288
- (496) Grzeskowiak, J. K.; Tscheliessnig, A.; Toh, P. C.; Chusainow, J.; Lee, Y. Y.; Wong, N.; Jungbauer, A. 2-D DIGE to expedite downstream process development for human monoclonal antibody purification. *Protein Expr Purif* **2009**, *66* (1), 58-65. DOI: 10.1016/j.pep.2009.01.007
- (497) 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. DOI: 10.1002/bit.22532
- (498) Hayduk, E. J.; Choe, L. H.; Lee, K. H. A two-dimensional electrophoresis map of Chinese hamster ovary cell proteins based on fluorescence staining. *Electrophoresis* **2004**, *25* (15), 2545-2556. DOI: 10.1002/elps.200406010
- (499) Beranova-Giorgianni, S. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. *TrAC Trends in Analytical Chemistry* **2003**, *22* (5), 273-281. DOI: 10.1016/S0165-9936(03)00508-9
- (500) Bailey-Kellogg, C.; Gutiérrez, A. H.; Moise, L.; Terry, F.; Martin, W. D.; De Groot, A. S. CHOPPI: a web tool for the analysis of immunogenicity risk from host cell proteins in CHO-based protein production. *Biotechnol Bioeng* **2014**, *111* (11), 2170-2182. DOI: 10.1002/bit.25286
- (501) Kreimer, S.; Gao, Y.; Ray, S.; Jin, M.; Tan, Z.; Mussa, N. A.; Tao, L.; Li, Z.; Ivanov, A. R.; Karger, B. L. Host Cell Protein Profiling by Targeted and Untargeted Analysis of Data Independent Acquisition Mass Spectrometry Data with Parallel Reaction Monitoring Verification. *Anal Chem* **2017**, *89* (10), 5294-5302. DOI: 10.1021/acs.analchem.6b04892
- (502) Strasser, L.; Oliviero, G.; Jakes, C.; Zaborowska, I.; Floris, P.; Ribeiro da Silva, M.; Füssl, F.; Carillo, S.; Bones, J. Detection and quantitation of host cell proteins in monoclonal antibody drug products using automated sample preparation and data-independent acquisition LC-MS/MS. *J Pharm Anal* **2021**, *11* (6), 726-731. DOI: 10.1016/j.jpha.2021.05.002
- (503) Husson, G.; Delangle, A.; O'Hara, J.; Cianferani, S.; Gervais, A.; Van Dorsselaer, A.; Bracewell, D.; Carapito, C. Dual Data-Independent Acquisition Approach Combining Global HCP Profiling and Absolute Quantification of Key Impurities during Bioprocess Development. *Anal Chem* **2018**, *90* (2), 1241-1247. DOI: 10.1021/acs.analchem.7b03965
- (504) Pythoud, N.; Bons, J.; Mijola, G.; Beck, A.; Cianfèrani, 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. DOI: 10.1021/acs.jproteome.0c00664
- (505) Hessmann, S.; Chéry, C.; Sikora, A.-S.; Gervais, A.; Carapito, C. Host Cell Protein Quantification Workflow Using Optimized Standards combined with Data-Independent Acquisition Mass Spectrometry. *J Pharm Anal* **2023**.
- (506) 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. DOI: 10.1002/btpr.2983
- (507) Gao, S. X.; Zhang, Y.; Stansberry-Perkins, K.; Buko, A.; Bai, S.; Nguyen, V.; Brader, M. L. Fragmentation of a highly purified monoclonal antibody attributed to residual CHO cell protease activity. *Biotechnol Bioeng* **2011**, *108* (4), 977-982. DOI: 10.1002/bit.22982
- (508) 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. DOI: 10.1080/19420862.2023.2213365

- (509) 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. DOI: 10.1021/acs.analchem.7b00304
- (510) 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. DOI: 10.1021/acs.analchem.9b05081
- (511) Madsen, J. A.; Farutin, V.; Carbeau, T.; Wudyka, S.; Yin, Y.; Smith, S.; Anderson, J.; Capila, I. Toward the complete characterization of host cell proteins in biotherapeutics via affinity depletions, LC-MS/MS, and multivariate analysis. *MAbs* **2015**, *7* (6), 1128-1137. DOI: 10.1080/19420862.2015.1082017
- (512) Thompson, J. H.; Chung, W. K.; Zhu, M.; Tie, L.; Lu, Y.; Aboulaich, N.; Strouse, R.; Mo, W. D. Improved detection of host cell proteins (HCPs) in a mammalian cell-derived antibody drug using liquid chromatography/mass spectrometry in conjunction with an HCP-enrichment strategy. *Rapid Commun Mass Spectrom* **2014**, *28* (8), 855-860. DOI: 10.1002/rcm.6854
- (513) 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. DOI: 10.1021/acs.analchem.5b02103
- (514) Doneanu, C. E.; Xenopoulos, A.; Fadgen, K.; Murphy, J.; Skilton, S. J.; Prentice, H.; Stapels, M.; Chen, W. Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *MAbs* **2012**, *4* (1), 24-44. DOI: 10.4161/mabs.4.1.18748
- (515) Yang, F.; Walker, D. E.; Schoenfelder, J.; Carver, J.; Zhang, A.; Li, D.; Harris, R.; Stults, J. T.; Yu, X. C.; Michels, D. A. A 2D LC-MS/MS Strategy for Reliable Detection of 10-ppm Level Residual Host Cell Proteins in Therapeutic Antibodies. *Anal Chem* **2018**, *90* (22), 13365-13372. DOI: 10.1021/acs.analchem.8b03044
- (516) 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. DOI: 10.1021/acs.analchem.0c00360
- (517) Johnson, R. O.; Greer, T.; Cejko, M.; Zheng, X.; Li, N. Combination of FAIMS, Protein A Depletion, and Native Digest Conditions Enables Deep Proteomic Profiling of Host Cell Proteins in Monoclonal Antibodies. *Anal Chem* **2020**, *92* (15), 10478-10484. DOI: 10.1021/acs.analchem.0c01175
- (518) Doneanu, C.; Fang, J.; Alelyunas, Y.; Yu, Y. Q.; Wrona, M.; Chen, W. An HS-MRM Assay for the Quantification of Host-cell Proteins in Protein Biopharmaceuticals by Liquid Chromatography Ion Mobility QTOF Mass Spectrometry. *J Vis Exp* **2018**, (134). DOI: 10.3791/55325
- (519) 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. DOI: 10.1021/acs.analchem.1c03095
- (520) Nie, S.; Greer, T.; O'Brien Johnson, R.; Zheng, X.; Torri, A.; Li, N. Simple and Sensitive Method for Deep Profiling of Host Cell Proteins in Therapeutic Antibodies by Combining Ultra-Low Trypsin Concentration Digestion, Long Chromatographic Gradients, and BoxCar Mass Spectrometry Acquisition. *Anal Chem* **2021**, *93* (10), 4383-4390. DOI: 10.1021/acs.analchem.0c03931
- (521) Ma, J.; Kilby, G. W. Sensitive, Rapid, Robust, and Reproducible Workflow for Host Cell Protein Profiling in Biopharmaceutical Process Development. *J Proteome Res* **2020**, *19* (8), 3396-3404. DOI: 10.1021/acs.jproteome.0c00252
- (522) 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. DOI: 10.1002/bit.24476
- (523) Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol Cell Proteomics* **2011**, *10* (12), M111.009910. DOI: 10.1074/mcp.M111.009910
- (524) Coon, J. J. Collisions or electrons? Protein sequence analysis in the 21st century. *Anal Chem* **2009**, *81* (9), 3208-3215. DOI: 10.1021/ac802330b
- (525) Mikesch, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* **2006**, *1764* (12), 1811-1822. DOI: 10.1016/j.bbapap.2006.10.003
- (526) Kim, M. S.; Zhong, J.; Kandasamy, K.; Delanghe, B.; Pandey, A. Systematic evaluation of alternating CID and ETD fragmentation for phosphorylated peptides. *Proteomics* **2011**, *11* (12), 2568-2572. DOI: 10.1002/pmic.201000547
- (527) Polanco, G.; Scott, N. E.; Lye, L. F.; Beverley, S. M. Expanded Proteomic Survey of the Human Parasite *Leishmania major* Focusing on Changes in Null Mutants of the Golgi GDP-

- Mannose/Fucose/Arabinopyranose Transporter LPG2 and of the Mitochondrial Fucosyltransferase FUT1. *Microbiol Spectr* **2022**, *10* (6), e0305222. DOI: 10.1128/spectrum.03052-22
- (528) Sanda, M.; Yang, Q.; Zong, G.; Chen, H.; Zheng, Z.; Dhani, H.; Khan, K.; Kroemer, A.; Wang, L. X.; Goldman, R. LC-MS/MS-PRM Quantification of IgG Glycoforms Using Stable Isotope Labeled IgG1 Fc Glycopeptide Standard. *J Proteome Res* **2023**, *22* (4), 1138-1147. DOI: 10.1021/acs.jproteome.2c00475
- (529) Molina, H.; Matthiesen, R.; Kandasamy, K.; Pandey, A. Comprehensive comparison of collision induced dissociation and electron transfer dissociation. *Anal Chem* **2008**, *80* (13), 4825-4835. DOI: 10.1021/ac8007785
- (530) Swaney, D. L.; McAlister, G. C.; Coon, J. J. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* **2008**, *5* (11), 959-964. DOI: 10.1038/nmeth.1260
- (531) Ishikawa, M.; Konno, R.; Nakajima, D.; Gotoh, M.; Fukasawa, K.; Sato, H.; Nakamura, R.; Ohara, O.; Kawashima, Y. Optimization of Ultrafast Proteomics Using an LC-Quadrupole-Orbitrap Mass Spectrometer with Data-Independent Acquisition. *J Proteome Res* **2022**, *21* (9), 2085-2093. DOI: 10.1021/acs.jproteome.2c00121
- (532) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Xuan, Y.; Sondermann, J.; Schmidt, M.; Gomez-Varela, D.; Reiter, L. Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol Cell Proteomics* **2017**, *16* (12), 2296-2309. DOI: 10.1074/mcp.RA117.000314
- (533) Reubsæet, L.; Sweredoski, M. J.; Moradian, A. Data-Independent Acquisition for the Orbitrap Q Exactive HF: A Tutorial. *J Proteome Res* **2019**, *18* (3), 803-813. DOI: 10.1021/acs.jproteome.8b00845
- (534) Gotti, C.; Roux-Dalvai, F.; Joly-Beauparlant, C.; Mangnier, L.; Leclercq, M.; Droit, A. Extensive and Accurate Benchmarking of DIA Acquisition Methods and Software Tools Using a Complex Proteomic Standard. *J Proteome Res* **2021**, *20* (10), 4801-4814. DOI: 10.1021/acs.jproteome.1c00490
- (535) Bekker-Jensen, D. B.; Martínez-Val, A.; Steigerwald, S.; Rütther, 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. DOI: 10.1074/mcp.TIR119.001906
- (536) Bracewell, D. G.; Francis, R.; Smales, C. M. The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnol Bioeng* **2015**, *112* (9), 1727-1737. DOI: 10.1002/bit.25628
- (537) Mateus, A.; Kurzawa, N.; Becher, I.; Sridharan, S.; Helm, D.; Stein, F.; Typas, A.; Savitski, M. M. Thermal proteome profiling for interrogating protein interactions. *Mol Syst Biol* **2020**, *16* (3), e9232. DOI: 10.15252/msb.20199232
- (538) Guo, Y.; Chowdhury, T.; Seshadri, M.; Cupp-Sutton, K. A.; Wang, Q.; Yu, D.; Wu, S. Optimization of Higher-Energy Collisional Dissociation Fragmentation Energy for Intact Protein-Level Tandem Mass Tag Labeling. *J Proteome Res* **2023**, *22* (5), 1406-1418. DOI: 10.1021/acs.jproteome.2c00549
- (539) Köcher, T.; Pichler, P.; Schutzbier, M.; Stingl, C.; Kaul, A.; Teucher, N.; Hasenfuss, G.; Penninger, J. M.; Mechtler, K. High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all. *J Proteome Res* **2009**, *8* (10), 4743-4752. DOI: 10.1021/pr900451u
- (540) Chiva, C.; Sabidó, E. HCD-only fragmentation method balances peptide identification and quantitation of TMT-labeled samples in hybrid linear ion trap/orbitrap mass spectrometers. *J Proteomics* **2014**, *96*, 263-270. DOI: 10.1016/j.jprot.2013.11.013
- (541) Dayon, L.; Pasquarello, C.; Hoogland, C.; Sanchez, J. C.; Scherl, A. Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags. *J Proteomics* **2010**, *73* (4), 769-777. DOI: 10.1016/j.jprot.2009.10.015
- (542) Altelaar, A. F.; Frese, C. K.; Preisinger, C.; Hennrich, M. L.; Schram, A. W.; Timmers, H. T.; Heck, A. J.; Mohammed, S. Benchmarking stable isotope labeling based quantitative proteomics. *J Proteomics* **2013**, *88*, 14-26. DOI: 10.1016/j.jprot.2012.10.009
- (543) McAlister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wühr, M.; Huttlin, E. L.; Erickson, B. K.; Rad, R.; Haas, W.; Gygi, S. P. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem* **2014**, *86* (14), 7150-7158. DOI: 10.1021/ac502040v
- (544) Furtwängler, B.; Üresin, N.; Motamedchaboki, K.; Huguet, R.; Lopez-Ferrer, D.; Zabrouskov, V.; Porse, B. T.; Schoof, E. M. Real-Time Search-Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. *Mol Cell Proteomics* **2022**, *21* (4), 100219. DOI: 10.1016/j.mcpro.2022.100219
- (545) Stingl, C.; Lau, S. P.; van der Burg, S. H.; Aerts, J. G.; van Eijck, C. H. J.; Luider, T. M. Dataset from a proteomics analysis of tumor antigens shared between an allogenic tumor cell lysate vaccine and pancreatic tumor tissue. *Data Brief* **2022**, *44*, 108490. DOI: 10.1016/j.dib.2022.108490

- (546) Johnson, A.; Stadlmeier, M.; Wühr, M. TMTpro Complementary Ion Quantification Increases Plexing and Sensitivity for Accurate Multiplexed Proteomics at the MS2 Level. *J Proteome Res* **2021**, *20* (6), 3043-3052. DOI: 10.1021/acs.jproteome.0c00813
- (547) Frost, D. C.; Feng, Y.; Li, L. 21-plex DiLeu Isobaric Tags for High-Throughput Quantitative Proteomics. *Anal Chem* **2020**, *92* (12), 8228-8234. DOI: 10.1021/acs.analchem.0c00473
- (548) Wang, Z.; Yu, K.; Tan, H.; Wu, Z.; Cho, J. H.; Han, X.; Sun, H.; Beach, T. G.; Peng, J. 27-Plex Tandem Mass Tag Mass Spectrometry for Profiling Brain Proteome in Alzheimer's Disease. *Anal Chem* **2020**, *92* (10), 7162-7170. DOI: 10.1021/acs.analchem.0c00655
- (549) Sun, H.; Poudel, S.; Vanderwall, D.; Lee, D. G.; Li, Y.; Peng, J. 29-Plex tandem mass tag mass spectrometry enabling accurate quantification by interference correction. *Proteomics* **2022**, *22* (19-20), e2100243. DOI: 10.1002/pmic.202100243
- (550) Juvvadi, P. R.; Moseley, M. A.; Hughes, C. J.; Soderblom, E. J.; Lennon, S.; Perkins, S. R.; Thompson, J. W.; Geromanos, S. J.; Wildgoose, J.; Richardson, K.; Langridge, J. I.; Vissers, J. P. C.; Steinbach, W. J. Scanning Quadrupole Data-Independent Acquisition, Part B: Application to the Analysis of the Calcineurin-Interacting Proteins during Treatment of *Aspergillus fumigatus* with Azole and Echinocandin Antifungal Drugs. *J Proteome Res* **2018**, *17* (2), 780-793. DOI: 10.1021/acs.jproteome.7b00499
- (551) Wang, Z.; Mülleder, M.; Batruch, I.; Chelur, A.; Textoris-Taube, K.; Schwecke, T.; Hartl, J.; Causon, J.; Castro-Perez, J.; Demichev, V.; Tate, S.; Ralser, M. High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. *Elife* **2022**, *11*. DOI: 10.7554/eLife.83947
- (552) Derks, J.; Leduc, A.; Wallmann, G.; Huffman, R. G.; Willetts, M.; Khan, S.; Specht, H.; Ralser, M.; Demichev, V.; Slavov, N. Increasing the throughput of sensitive proteomics by plexDIA. *Nat Biotechnol* **2022**. DOI: 10.1038/s41587-022-01389-w
- (553) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Mol Cell Proteomics* **2006**, *5* (12), 2384-2391. DOI: 10.1074/mcp.M600248-MCP200
- (554) Hains, P. G.; Robinson, P. J. The Impact of Commonly Used Alkylating Agents on Artifactual Peptide Modification. *J Proteome Res* **2017**, *16* (9), 3443-3447. DOI: 10.1021/acs.jproteome.7b00022
- (555) David, M.; Fertin, G.; Rogniaux, H.; Tessier, D. SpecOMS: A Full Open Modification Search Method Performing All-to-All Spectra Comparisons within Minutes. *J Proteome Res* **2017**, *16* (8), 3030-3038. DOI: 10.1021/acs.jproteome.7b00308
- (556) Amodè, D.; Egertson, J.; MacLean, B. X.; Johnson, R.; Merrihew, G. E.; Keller, A.; Marsh, D.; Vitek, O.; Mallick, P.; MacCoss, M. J. Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. *J Am Soc Mass Spectrom* **2019**, *30* (4), 669-684. DOI: 10.1007/s13361-018-2122-8
- (557) 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. DOI: 10.15252/msb.20178126
- (558) Mehta, D.; Scandola, S.; Uhrig, R. G. BoxCar and Library-Free Data-Independent Acquisition Substantially Improve the Depth, Range, and Completeness of Label-Free Quantitative Proteomics. *Anal Chem* **2022**, *94* (2), 793-802. DOI: 10.1021/acs.analchem.1c03338
- (559) Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; Prianichnikov, N.; Yilmaz, Ş.; Rudolph, J. D.; Tenzer, S.; Perez-Riverol, Y.; Nagaraj, N.; Humphrey, S. J.; Cox, J. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nat Biotechnol* **2021**, *39* (12), 1563-1573. DOI: 10.1038/s41587-021-00968-7
- (560) The, M.; MacCoss, M. J.; Noble, W. S.; Käll, L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J Am Soc Mass Spectrom* **2016**, *27* (11), 1719-1727. DOI: 10.1007/s13361-016-1460-7
- (561) Millikin, R. J.; Solntsev, S. K.; Shortreed, M. R.; Smith, L. M. Ultrafast Peptide Label-Free Quantification with FlashLFQ. *J Proteome Res* **2018**, *17* (1), 386-391. DOI: 10.1021/acs.jproteome.7b00608
- (562) Toprak, U. H.; Gillet, L. C.; Maiolica, A.; Navarro, P.; Leitner, A.; Aebersold, R. Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Mol Cell Proteomics* **2014**, *13* (8), 2056-2071. DOI: 10.1074/mcp.0113.036475
- (563) Frejno, M.; Zolg, D. P.; Schmidt, T.; Gessulat, S.; Graber, M.; Seefried, F.; Rathke-Kuhnert, M.; Ben Fredj, S.; Premnadh, S.; Samaras, P.; Fritzmeier, K.; Berg, F.; Nasir, W.; Horn, D.; Delanghe, B.; Henrich, C.; Kuster, B.; Wilhelm, M. Chimerys: An AI-Driven Leap Forward in Peptide Identification. In *69th ASMS Conference on Mass Spectrometry and Allied Topics*, Philadelphia, PA, 2021.
- (564) Wang, J.; Bourne, P. E.; Bandeira, N. MixGF: spectral probabilities for mixture spectra from more than one peptide. *Mol Cell Proteomics* **2014**, *13* (12), 3688-3697. DOI: 10.1074/mcp.0113.037218
- (565) Truong, T.; Webber, K. G. I.; Madisyn Johnston, S.; Boekweg, H.; Lindgren, C. M.; Liang, Y.; Nydegger, A.; Xie, X.; Tsang, T. M.; Jayatunge, D.; Andersen, J. L.; Payne, S. H.; Kelly, R. T. Data-Dependent Acquisition

- with Precursor Coisolation Improves Proteome Coverage and Measurement Throughput for Label-Free Single-Cell Proteomics. *Angew Chem Int Ed Engl* **2023**, e202303415. DOI: 10.1002/anie.202303415
- (566) Mayer, R. L.; Matzinger, M.; Schmücker, A.; Stejskal, K.; Krššáková, G.; Berger, F.; Mechtler, K. Wide Window Acquisition and AI-based data analysis to reach deep proteome coverage for a wide sample range, including single cell proteomic inputs. *bioRxiv* **2022**, 2022.2009.2001.506203. DOI: 10.1101/2022.09.01.506203
- (567) Sechi, S.; Chait, B. T. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal Chem* **1998**, *70* (24), 5150-5158. DOI: 10.1021/ac9806005
- (568) Iqbal, S.; Pérez-Palma, E.; Jespersen, J. B.; May, P.; Hoksza, D.; Heyne, H. O.; Ahmed, S. S.; Rifat, Z. T.; Rahman, M. S.; Lage, K.; Palotie, A.; Cottrell, J. R.; Wagner, F. F.; Daly, M. J.; Campbell, A. J.; Lal, D. Comprehensive characterization of amino acid positions in protein structures reveals molecular effect of missense variants. *Proc Natl Acad Sci U S A* **2020**, *117* (45), 28201-28211. DOI: 10.1073/pnas.2002660117
- (569) Kollipara, L.; Zahedi, R. P. Protein carbamylation: in vivo modification or in vitro artefact? *Proteomics* **2013**, *13* (6), 941-944. DOI: 10.1002/pmic.201200452
- (570) Kuznetsova, K. G.; Levitsky, L. I.; Pyatnitskiy, M. A.; Ilina, I. Y.; Bubis, J. A.; Solovyeva, E. M.; Zgoda, V. G.; Gorshkov, M. V.; Moshkovskii, S. A. Cysteine alkylation methods in shotgun proteomics and their possible effects on methionine residues. *J Proteomics* **2021**, *231*, 104022. DOI: 10.1016/j.jprot.2020.104022
- (571) Stadtman, E. R.; Van Remmen, H.; Richardson, A.; Wehr, N. B.; Levine, R. L. Methionine oxidation and aging. *Biochim Biophys Acta* **2005**, *1703* (2), 135-140. DOI: 10.1016/j.bbapap.2004.08.010
- (572) Ghesquière, B.; Gevaert, K. Proteomics methods to study methionine oxidation. *Mass Spectrom Rev* **2014**, *33* (2), 147-156. DOI: 10.1002/mas.21386
- (573) Güray, M. Z.; Zheng, S.; Doucette, A. A. Mass Spectrometry of Intact Proteins Reveals +98 u Chemical Artifacts Following Precipitation in Acetone. *J Proteome Res* **2017**, *16* (2), 889-897. DOI: 10.1021/acs.jproteome.6b00841
- (574) Simpson, D. M.; Beynon, R. J. Acetone precipitation of proteins and the modification of peptides. *J Proteome Res* **2010**, *9* (1), 444-450. DOI: 10.1021/pr900806x
- (575) Hulshof, T. G.; Rutherford, S. M.; Sforza, S.; Bikker, P.; van der Poel, A. F.; Hendriks, W. H. O-Methylisourea Can React with the α -Amino Group of Lysine: Implications for the Analysis of Reactive Lysine. *J Agric Food Chem* **2017**, *65* (4), 964-972. DOI: 10.1021/acs.jafc.6b03096
- (576) Ying, Y.; Li, H. Recent progress in the analysis of protein deamidation using mass spectrometry. *Methods* **2022**, *200*, 42-57. DOI: 10.1016/j.ymeth.2020.06.009
- (577) Hao, P.; Ren, Y.; Alpert, A. J.; Sze, S. K. Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Mol Cell Proteomics* **2011**, *10* (10), O111.009381. DOI: 10.1074/mcp.O111.009381
- (578) Purwaha, P.; Silva, L. P.; Hawke, D. H.; Weinstein, J. N.; Lorenzi, P. L. An artifact in LC-MS/MS measurement of glutamine and glutamic acid: in-source cyclization to pyroglutamic acid. *Anal Chem* **2014**, *86* (12), 5633-5637. DOI: 10.1021/ac501451v
- (579) Schilling, S.; Wasternack, C.; Demuth, H. U. Glutaminyl cyclases from animals and plants: a case of functionally convergent protein evolution. *Biol Chem* **2008**, *389* (8), 983-991. DOI: 10.1515/bc.2008.111
- (580) Oliveira, A. P.; Sauer, U. The importance of post-translational modifications in regulating *Saccharomyces cerevisiae* metabolism. *FEMS Yeast Res* **2012**, *12* (2), 104-117. DOI: 10.1111/j.1567-1364.2011.00765.x
- (581) Barkovits, K.; Pacharra, S.; Pfeiffer, K.; Steinbach, S.; Eisenacher, M.; Marcus, K.; Uszkoreit, J. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Mol Cell Proteomics* **2020**, *19* (1), 181-197. DOI: 10.1074/mcp.RA119.001714
- (582) Petritis, K.; Kangas, L. J.; Ferguson, P. L.; Anderson, G. A.; Pasa-Tolić, L.; Lipton, M. S.; Auberry, K. J.; Strittmatter, E. F.; Shen, Y.; Zhao, R.; Smith, R. D. Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. *Anal Chem* **2003**, *75* (5), 1039-1048. DOI: 10.1021/ac0205154
- (583) Petritis, K.; Kangas, L. J.; Yan, B.; Monroe, M. E.; Strittmatter, E. F.; Qian, W. J.; Adkins, J. N.; Moore, R. J.; Xu, Y.; Lipton, M. S.; Camp, D. G., 2nd; Smith, R. D. Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information. *Anal Chem* **2006**, *78* (14), 5026-5039. DOI: 10.1021/ac060143p
- (584) Zeng, W. F.; Zhou, X. X.; Willems, S.; Ammar, C.; Wahle, M.; Bludau, I.; Voytik, E.; Strauss, M. T.; Mann, M. AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. *Nat Commun* **2022**, *13* (1), 7238. DOI: 10.1038/s41467-022-34904-3
- (585) Meier, F.; Köhler, N. D.; Brunner, A. D.; Wanka, J. H.; Voytik, E.; Strauss, M. T.; Theis, F. J.; Mann, M. Deep learning the collisional cross sections of the peptide universe from a million experimental values. *Nat Commun* **2021**, *12* (1), 1185. DOI: 10.1038/s41467-021-21352-8

- (586) Walker, D. E.; Yang, F.; Carver, J.; Joe, K.; Michels, D. A.; Yu, X. C. A modular and adaptive mass spectrometry-based platform for support of bioprocess development toward optimal host cell protein clearance. *MAbs* **2017**, *9* (4), 654-663. DOI: 10.1080/19420862.2017.1303023
- (587) European Pharmacopeia Monograph 2.6.34 Host-Cell Protein Assays. 9. European Pharmacopeia Monograph 2.6.34 Host-Cell Protein Assays. 9, 2017. 2017; Vol. 9.
- (588) 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. DOI: 10.1080/19420862.2021.1955811
- (589) Farrell, A.; Mittermayr, S.; Morrissey, B.; Mc Loughlin, N.; Navas Iglesias, N.; Marison, I. W.; Bones, J. Quantitative host cell protein analysis using two dimensional data independent LC-MS(E). *Anal Chem* **2015**, *87* (18), 9186-9193. DOI: 10.1021/acs.analchem.5b01377
- (590) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421. DOI: 10.1186/1471-2105-10-421
- (591) Trauchessec, M.; Hesse, A. M.; Kraut, A.; Berard, Y.; Herment, L.; Fortin, T.; Bruley, C.; Ferro, M.; Manin, C. An innovative standard for LC-MS-based HCP profiling and accurate quantity assessment: Application to batch consistency in viral vaccine samples. *Proteomics* **2021**, *21* (5), e2000152. DOI: 10.1002/pmic.202000152
- (592) Spanov, B.; Govorukhina, N.; Merbel, N. C. v. d.; Bischoff, R. Analytical tools for the characterization of deamidation in monoclonal antibodies. *Journal of Chromatography Open* **2022**, *2*, 100025. DOI: 10.1016/j.jcoa.2021.100025
- (593) Xu, Y.; Wang, D.; Mason, B.; Rossomando, T.; Li, N.; Liu, D.; Cheung, J. K.; Xu, W.; Raghava, S.; Katiyar, A.; Nowak, C.; Xiang, T.; Dong, D. D.; Sun, J.; Beck, A.; Liu, H. Structure, heterogeneity and developability assessment of therapeutic antibodies. *MAbs* **2019**, *11* (2), 239-264. DOI: 10.1080/19420862.2018.1553476
- (594) Stewart, H.; Grinfeld, D.; Giannakopoulos, A.; Petzoldt, J.; Shanley, T.; Garland, M.; Denisov, E.; Peterson, A.; Damoc, E.; Zeller, M.; Arrey, T. N.; Pashkova, A.; Renuse, S.; Hakimi, A.; Kühn, A.; Biel, M.; Kreutzmann, A.; Hagedorn, B.; Colonius, I.; Schütz, A.; Stefes, A.; Dwivedi, A.; Mourad, D.; Hoek, M.; Reitemeier, B.; Cochems, P.; Kholomeev, A.; Ostermann, R.; Quiring, G.; Ochmann, M.; Möhring, S.; Wagner, A.; Petker, A.; Kanngiesser, S.; Wiedemeyer, M.; Balschun, W.; Hermanson, D.; Zabrouskov, V.; Makarov, A.; Hock, C. Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis. *bioRxiv* **2023**, 2023.2006.2002.543408. DOI: 10.1101/2023.06.02.543408
- (595) Smith, J.; Carillo, S.; Srzentić, K.; Arrey, T. N.; Pashkova, A.; Scheffler, K.; Broster, K.; Damoc, N. E.; Bones, J. Host cell protein (HCP) profiling and quantitation in gene therapy products on a novel high-resolution accurate mass platform. In *71st ASMS Conference on Mass Spectrometry and Allied Topics*, Houston, TX, USA, 2023.
- (596) Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Paša-Tolić, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods* **2019**, *16* (7), 587-594. DOI: 10.1038/s41592-019-0457-0
- (597) Liu, J.; McLuckey, S. A. Electron Transfer Dissociation: Effects of Cation Charge State on Product Partitioning in Ion/Ion Electron Transfer to Multiply Protonated Polypeptides. *International Journal of Mass Spectrometry* **2012**, *330-332*, 174-181. DOI: 10.1016/j.ijms.2012.07.013
- (598) Weisbrod, C. R.; Kaiser, N. K.; Syka, J. E. P.; Early, L.; Mullen, C.; Dunyach, J. J.; English, A. M.; Anderson, L. C.; Blakney, G. T.; Shabanowitz, J.; Hendrickson, C. L.; Marshall, A. G.; Hunt, D. F. Front-End Electron Transfer Dissociation Coupled to a 21 Tesla FT-ICR Mass Spectrometer for Intact Protein Sequence Analysis. *J Am Soc Mass Spectrom* **2017**, *28* (9), 1787-1795. DOI: 10.1007/s13361-017-1702-3
- (599) Fornelli, L.; Parra, J.; Hartmer, R.; Stoermer, C.; Lubeck, M.; Tsybin, Y. O. Top-down analysis of 30-80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry. *Anal Bioanal Chem* **2013**, *405* (26), 8505-8514. DOI: 10.1007/s00216-013-7267-5
- (600) Graham, K. A.; Lawlor, C. F.; Borotto, N. B. Characterizing the top-down sequencing of protein ions prior to mobility separation in a timsTOF. *Analyst* **2023**, *148* (7), 1534-1542. DOI: 10.1039/d2an01682f
- (601) Kellie, J. F.; Schneck, N. A.; Causon, J. C.; Baba, T.; Mehl, J. T.; Pohl, K. I. Top-Down Characterization and Intact Mass Quantitation of a Monoclonal Antibody Drug from Serum by Use of a Quadrupole TOF MS System Equipped with Electron-Activated Dissociation. *J Am Soc Mass Spectrom* **2023**, *34* (1), 17-26. DOI: 10.1021/jasms.2c00206
- (602) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay Ü, A.; Tsybin, Y. O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem Soc Rev* **2013**, *42* (12), 5014-5030. DOI: 10.1039/c3cs35477f

- (603) Berthias, F.; Cooper-Shepherd, D. A.; Holck, F. H. V.; Langridge, J. I.; Jensen, O. N. Full Separation and Sequencing of Isomeric Proteoforms in the Middle-Down Mass Range Using Cyclic Ion Mobility and Electron Capture Dissociation. *Anal Chem* **2023**, *95* (29), 11141-11148. DOI: 10.1021/acs.analchem.3c02120
- (604) Chen, B.; Guo, X.; Tucholski, T.; Lin, Z.; McIlwain, S.; Ge, Y. The Impact of Phosphorylation on Electron Capture Dissociation of Proteins: A Top-Down Perspective. *J Am Soc Mass Spectrom* **2017**, *28* (9), 1805-1814. DOI: 10.1007/s13361-017-1710-3
- (605) Shaw, J. B.; Cooper-Shepherd, D. A.; Hewitt, D.; Wildgoose, J. L.; Beckman, J. S.; Langridge, J. I.; Voinov, V. G. Enhanced Top-Down Protein Characterization with Electron Capture Dissociation and Cyclic Ion Mobility Spectrometry. *Anal Chem* **2022**, *94* (9), 3888-3896. DOI: 10.1021/acs.analchem.1c04870
- (606) Gerbasi, V. R.; Melani, R. D.; Abbatiello, S. E.; Belford, M. W.; Huguet, R.; McGee, J. P.; Dayhoff, D.; Thomas, P. M.; Kelleher, N. L. Deeper Protein Identification Using Field Asymmetric Ion Mobility Spectrometry in Top-Down Proteomics. *Anal Chem* **2021**, *93* (16), 6323-6328. DOI: 10.1021/acs.analchem.1c00402
- (607) Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L. Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *J Am Soc Mass Spectrom* **2015**, *26* (5), 782-787. DOI: 10.1007/s13361-015-1078-1
- (608) Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Loo, R. R. O.; Loo, J. A. Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry. *J Am Soc Mass Spectrom* **2020**, *31* (9), 1896-1902. DOI: 10.1021/jasms.0c00160
- (609) Shrivastava, A.; Joshi, S.; Guttman, A.; Rathore, A. S. N-Glycosylation of monoclonal antibody therapeutics: A comprehensive review on significance and characterization. *Anal Chim Acta* **2022**, *1209*, 339828. DOI: 10.1016/j.aca.2022.339828
- (610) Chen, L.; Sze, S. K.; Yang, H. Automated intensity descent algorithm for interpretation of complex high-resolution mass spectra. *Anal Chem* **2006**, *78* (14), 5006-5018. DOI: 10.1021/ac060099d
- (611) Kaur, P.; O'Connor, P. B. Algorithms for automatic interpretation of high resolution mass spectra. *J Am Soc Mass Spectrom* **2006**, *17* (3), 459-468. DOI: 10.1016/j.jasms.2005.11.024
- (612) Guner, H.; Close, P. L.; Cai, W.; Zhang, H.; Peng, Y.; Gregorich, Z. R.; Ge, Y. MASH Suite: a user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization. *J Am Soc Mass Spectrom* **2014**, *25* (3), 464-470. DOI: 10.1007/s13361-013-0789-4
- (613) Scheffler, K.; Viner, R.; Damoc, E. High resolution top-down experimental strategies on the Orbitrap platform. *J Proteomics* **2018**, *175*, 42-55. DOI: 10.1016/j.jprot.2017.03.028
- (614) Choi, I. K.; Abeysinghe, E.; Coulter, E.; Marru, S.; Pierce, M.; Liu, X. TopPIC Gateway: A Web Gateway for Top-Down Mass Spectrometry Data Interpretation. *Pearc20 (2020)* **2020**, *2020*, 461-464. DOI: 10.1145/3311790.3400853
- (615) Choi, I. K.; Liu, X. Top-Down Mass Spectrometry Data Analysis Using TopPIC Suite. *Methods Mol Biol* **2022**, *2500*, 83-103. DOI: 10.1007/978-1-0716-2325-1_8
- (616) Sun, R. X.; Luo, L.; Wu, L.; Wang, R. M.; Zeng, W. F.; Chi, H.; Liu, C.; He, S. M. pTop 1.0: A High-Accuracy and High-Efficiency Search Engine for Intact Protein Identification. *Anal Chem* **2016**, *88* (6), 3082-3090. DOI: 10.1021/acs.analchem.5b03963
- (617) Sun, R. X.; Wang, R. M.; Luo, L.; Liu, C.; Chi, H.; Zeng, W. F.; He, S. M. Accurate Proteoform Identification and Quantitation Using pTop 2.0. *Methods Mol Biol* **2022**, *2500*, 105-129. DOI: 10.1007/978-1-0716-2325-1_9
- (618) 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. DOI: 10.1074/mcp.RA118.000993
- (619) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top down proteomics. *J Proteome Res* **2014**, *13* (7), 3231-3240. DOI: 10.1021/pr401277r
- (620) Kou, Q.; Wang, Z.; Lubeckyj, R. A.; Wu, S.; Sun, L.; Liu, X. A Markov Chain Monte Carlo Method for Estimating the Statistical Significance of Proteoform Identifications by Top-Down Mass Spectrometry. *J Proteome Res* **2019**, *18* (3), 878-889. DOI: 10.1021/acs.jproteome.8b00562
- (621) Li, Z.; He, B.; Kou, Q.; Wang, Z.; Wu, S.; Liu, Y.; Feng, W.; Liu, X. Evaluation of top-down mass spectral identification with homologous protein sequences. *BMC Bioinformatics* **2018**, *19* (Suppl 17), 494. DOI: 10.1186/s12859-018-2462-1
- (622) LeDuc, R. D.; Taylor, G. K.; Kim, Y. B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res* **2004**, *32* (Web Server issue), W340-345. DOI: 10.1093/nar/gkh447
- (623) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: improved protein identification and characterization

- for top down mass spectrometry. *Nucleic Acids Res* **2007**, *35* (Web Server issue), W701-706. DOI: 10.1093/nar/gkm371
- (624) Larson, E. J.; Pergande, M. R.; Moss, M. E.; Rossler, K. J.; Wenger, R. K.; Krichel, B.; Josyer, H.; Melby, J. A.; Roberts, D. S.; Pike, K.; Shi, Z.; Chan, H. J.; Knight, B.; Rogers, H. T.; Brown, K. A.; Ong, I. M.; Jeong, K.; Marty, M. T.; McIlwain, S. J.; Ge, Y. MASH Native: a unified solution for native top-down proteomics data processing. *Bioinformatics* **2023**, *39* (6). DOI: 10.1093/bioinformatics/btad359
- (625) Durbin, K. R.; Robey, M. T.; Voong, L. N.; Fellers, R. T.; Lutomski, C. A.; El-Baba, T. J.; Robinson, C. V.; Kelleher, N. L. ProSight Native: Defining Protein Complex Composition from Native Top-Down Mass Spectrometry Data. *J Proteome Res* **2023**, *22* (8), 2660-2668. DOI: 10.1021/acs.jproteome.3c00171
- (626) Martin, E. A.; Fulcher, J. M.; Zhou, M.; Monroe, M. E.; Petyuk, V. A. TopPICR: A Companion R Package for Top-Down Proteomics Data Analysis. *J Proteome Res* **2023**, *22* (2), 399-409. DOI: 10.1021/acs.jproteome.2c00570
- (627) Choi, I. K.; Jiang, T.; Kankara, S. R.; Wu, S.; Liu, X. TopMSV: A Web-Based Tool for Top-Down Mass Spectrometry Data Visualization. *J Am Soc Mass Spectrom* **2021**, *32* (6), 1312-1318. DOI: 10.1021/jasms.0c00460
- (628) Juetten, K. J.; Brodbelt, J. S. MS-TAFI: A Tool for the Analysis of Fragment Ions Generated from Intact Proteins. *J Proteome Res* **2023**, *22* (2), 546-550. DOI: 10.1021/acs.jproteome.2c00594
- (629) Locard-Paulet, M.; Parra, J.; Albigot, R.; Mouton-Barbosa, E.; Bardi, L.; Burlet-Schiltz, O.; Marcoux, J. VisioProt-MS: interactive 2D maps from intact protein mass spectrometry. *Bioinformatics* **2019**, *35* (4), 679-681. DOI: 10.1093/bioinformatics/bty680
- (630) Degnan, D. J.; Bramer, L. M.; White, A. M.; Zhou, M.; Bilbao, A.; McCue, L. A. PSpectreR: A User-Friendly and Interactive Application for Visualizing Top-Down and Bottom-Up Proteomics Data in R. *J Proteome Res* **2021**, *20* (4), 2014-2020. DOI: 10.1021/acs.jproteome.0c00857
- (631) Dhenin, J.; Borges Lima, D.; Dupré, M.; Chamot-Rooke, J. TDFragMapper: a visualization tool for evaluating experimental parameters in top-down proteomics. *Bioinformatics* **2022**, *38* (4), 1136-1138. DOI: 10.1093/bioinformatics/btab784
- (632) Consortium for Top-Down Proteomics. <https://www.topdownproteomics.org/> (accessed 07/08/2023).
- (633) Nakada, T.; Masuda, T.; Naito, H.; Yoshida, M.; Ashida, S.; Morita, K.; Miyazaki, H.; Kasuya, Y.; Ogitani, Y.; Yamaguchi, J.; Abe, Y.; Honda, T. Novel antibody drug conjugates containing exatecan derivative-based cytotoxic payloads. *Bioorg Med Chem Lett* **2016**, *26* (6), 1542-1545. DOI: 10.1016/j.bmcl.2016.02.020
- (634) Ogitani, Y.; Aida, T.; Hagihara, K.; Yamaguchi, J.; Ishii, C.; Harada, N.; Soma, M.; Okamoto, H.; Oitate, M.; Arakawa, S.; Hirai, T.; Atsumi, R.; Nakada, T.; Hayakawa, I.; Abe, Y.; Agatsuma, T. DS-8201a, A Novel HER2-Targeting ADC with a Novel DNA Topoisomerase I Inhibitor, Demonstrates a Promising Antitumor Efficacy with Differentiation from T-DM1. *Clin Cancer Res* **2016**, *22* (20), 5097-5108. DOI: 10.1158/1078-0432.ccr-15-2822
- (635) Cortés, J.; Kim, S. B.; Chung, W. P.; Im, S. A.; Park, Y. H.; Hegg, R.; Kim, M. H.; Tseng, L. M.; Petry, V.; Chung, C. F.; Iwata, H.; Hamilton, E.; Curigliano, G.; Xu, B.; Huang, C. S.; Kim, J. H.; Chiu, J. W. Y.; Pedrini, J. L.; Lee, C.; Liu, Y.; Cathcart, J.; Bako, E.; Verma, S.; Hurvitz, S. A. Trastuzumab Deruxtecan versus Trastuzumab Emtansine for Breast Cancer. *N Engl J Med* **2022**, *386* (12), 1143-1154. DOI: 10.1056/NEJMoa2115022
- (636) Lewis Phillips, G. D.; Li, G.; Dugger, D. L.; Crocker, L. M.; Parsons, K. L.; Mai, E.; Blättler, W. A.; Lambert, J. M.; Chari, R. V.; Lutz, R. J.; Wong, W. L.; Jacobson, F. S.; Koeppen, H.; Schwall, R. H.; Kenkare-Mitra, S. R.; Spencer, S. D.; Sliwkowski, M. X. Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody-cytotoxic drug conjugate. *Cancer Res* **2008**, *68* (22), 9280-9290. DOI: 10.1158/0008-5472.can-08-1776
- (637) Sang, H.; Lu, G.; Liu, Y.; Hu, Q.; Xing, W.; Cui, D.; Zhou, F.; Zhang, J.; Hao, H.; Wang, G.; Ye, H. Conjugation site analysis of antibody-drug-conjugates (ADCs) by signature ion fingerprinting and normalized area quantitation approach using nano-liquid chromatography coupled to high resolution mass spectrometry. *Anal Chim Acta* **2017**, *955*, 67-78. DOI: 10.1016/j.aca.2016.11.073
- (638) Larson, E. J.; Roberts, D. S.; Melby, J. A.; Buck, K. M.; Zhu, Y.; Zhou, S.; Han, L.; Zhang, Q.; Ge, Y. High-Throughput Multi-attribute Analysis of Antibody-Drug Conjugates Enabled by Trapped Ion Mobility Spectrometry and Top-Down Mass Spectrometry. *Anal Chem* **2021**, *93* (29), 10013-10021. DOI: 10.1021/acs.analchem.1c00150
- (639) Etkirch, A.; Hernandez-Alba, O.; Colas, O.; Beck, A.; Guilleme, D.; Cianféroni, S. Hyphenation of size exclusion chromatography to native ion mobility mass spectrometry for the analytical characterization of therapeutic antibodies and related products. *J Chromatogr B Analyt Technol Biomed Life Sci* **2018**, *1086*, 176-183. DOI: 10.1016/j.jchromb.2018.04.010
- (640) Shi, R. L.; Xiao, G.; Dillon, T. M.; Ricci, M. S.; Bondarenko, P. V. Characterization of therapeutic proteins by cation exchange chromatography-mass spectrometry and top-down analysis. *MAbs* **2020**, *12* (1), 1739825. DOI: 10.1080/19420862.2020.1739825

- (641) Wu, D.; Robinson, C. V. Native Top-Down Mass Spectrometry Reveals a Role for Interfacial Glycans on Therapeutic Cytokine and Hormone Assemblies. *Angew Chem Int Ed Engl* **2022**, *61* (49), e202213170. DOI: 10.1002/anie.202213170
- (642) Babović, M.; Shliha, P. V.; Gibb, S.; Jensen, O. N. Effective Amino Acid Sequencing of Intact Filgrastim by Multimodal Mass Spectrometry with Topdownr. *J Am Soc Mass Spectrom* **2022**, *33* (11), 2087-2093. DOI: 10.1021/jasms.2c00193

Développement de stratégies analytiques et bioinformatiques en spectrométrie de masse pour la caractérisation des protéines thérapeutiques et de leurs impuretés

Résumé

La spectrométrie de masse est une technique de choix pour la caractérisation des protéines. L'objectif de cette thèse a été de développer des stratégies analytiques et bioinformatiques en spectrométrie de masse pour l'analyse des protéines thérapeutiques et de leurs impuretés.

Dans un premier temps, de ce travail a été axé sur l'optimisation de méthodes MS en protéomique quantitative et sur l'évaluation de stratégies bioinformatiques innovantes pour le traitement des données. Dans une deuxième partie, ces approches ont été appliquées pour la quantification précise de protéines de la cellule hôte présentes à l'état de traces dans des produits médicamenteux. La troisième partie de cette thèse a porté sur le développement de stratégies « top-down » et « middle-down » combinant de multiples techniques de fragmentation pour l'analyse de protéines intactes, et a souligné l'importance d'employer un traitement de données adapté pour assurer la pertinence des résultats obtenus. Enfin, ces stratégies ont été utilisées pour la caractérisation complète d'un anticorps conjugué thérapeutique.

Ce travail a démontré le fort intérêt de la spectrométrie de masse pour l'analyse de protéines thérapeutiques et de leurs impuretés.

Mots clés : Spectrométrie de masse (MS), analyse protéomique quantitative, protéine intacte, protéines de la cellule hôte (HCP), anticorps monoclonaux thérapeutiques.

Résumé en anglais

Mass spectrometry has been a technique of choice for the characterization of proteins. This Ph.D. work aims to develop analytical and bioinformatics strategies in mass spectrometry to analyze therapeutic proteins and their impurities.

The first part of this work focused on optimizing mass spectrometry methods for quantitative proteomics and evaluating innovative bioinformatics strategies for data processing. The second part applied these approaches to quantifying host cell proteins present at trace levels in drug products precisely. The third part of this thesis focused on developing "top-down" and "middle-down" strategies combining multiple fragmentation techniques to analyze intact proteins. It also highlighted the importance of employing suitable data processing to ensure the relevance of the results obtained. Finally, these strategies were used to characterize a therapeutic antibody-drug conjugate comprehensively.

This work has demonstrated a strong interest in mass spectrometry for analyzing therapeutic proteins and their impurities.

Keywords: Mass spectrometry (MS), quantitative proteomics analysis, intact protein, host cell proteins (HCPs), therapeutic monoclonal antibodies.