

La protéomique par spectrométrie de masse

2.1 La spectrométrie de masse

La spectrométrie de masse (MS) est une technologie permettant de déterminer le rapport masse sur charge (m/z) d'une molécule chargée à l'état gazeux. Les spectromètres de masse sont composés de trois parties : une source d'ion, un analyseur et un détecteur. La source d'ion permet de faire passer les molécules d'un état condensé (liquide ou solide) à un état gazeux en les ionisant. Ces ions peuvent être chargés négativement ou positivement. Une fois produites au niveau de la source, les molécules chargées en phase gazeuse peuvent être analysées grâce à un champ électrique ou magnétique dans l'analyseur. L'analyseur permet ainsi de les séparer en fonction de leur rapport masse/charge (m/z) avant qu'elles ne soient comptées par le détecteur. Le résultat se présente ensuite sous forme d'un spectre de masse qui n'est rien d'autre qu'un graphique représentant les intensités des signaux détectés en fonction des différents m/z mesurés. Il est possible d'associer plusieurs analyseurs de masse pour faire de la spectrométrie de masse en mode tandem ou MS/MS. Dans cette configuration, il est possible d'isoler des molécules, de les fragmenter et d'analyser les fragments. Ce mode de fonctionnement de la MS permet d'obtenir des informations structurales supplémentaires sur les molécules pour en déduire par exemple leur séquence dans le cas des polypeptides.

L'utilisation de la spectrométrie de masse pour l'analyse des biomolécules a pris un essor important au début des années 1990 après le développement des méthodes d'ionisation douces l'ESI (ionisation par *electrospray*) et le MALDI (ionisation laser assistée par une matrice)¹²⁴. En effet, ces deux méthodes sont capables d'ioniser des biomolécules fragiles en minimisant leur dégradation sans les fragmenter. Grâce à ces avancées, la MS est devenue une technique de choix pour l'étude des protéines. Les échantillons protéiques étant très complexes, les performances de la MS dépendent énormément de la technique de séparation avec laquelle elle est associée¹⁵. Indépendamment de la méthode de séparation utilisée, l'analyse des protéines par MS se fait selon deux principales approches : le *bottom-up* et le *top-down* (Figure 3).

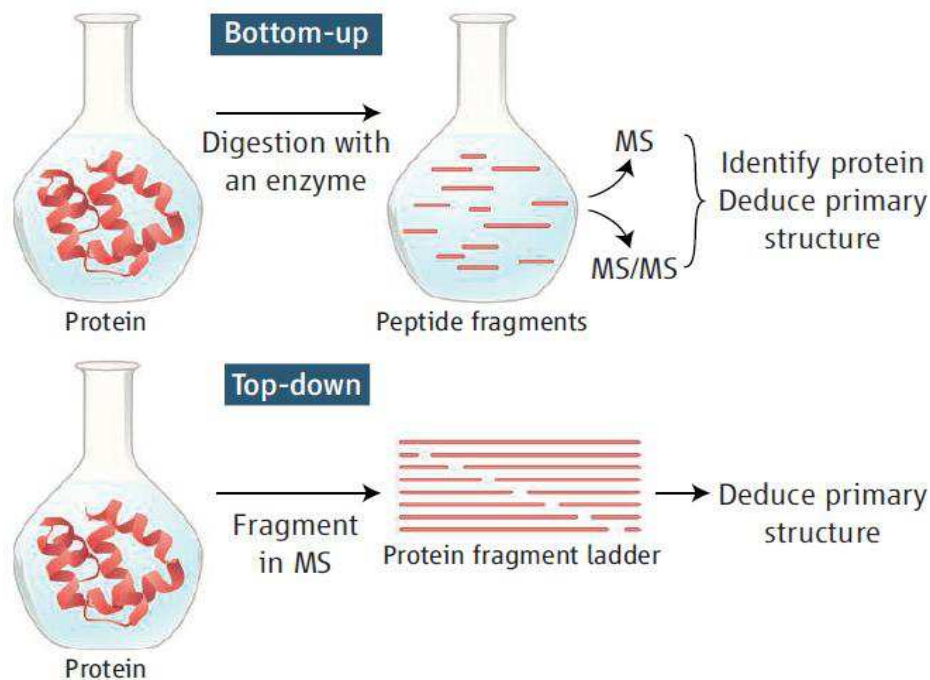


Figure 3 : Les deux principales approches en analyse des protéines par spectrométrie de masse

En protéomique *bottom-up*, les protéines sont digérées (protéolyse) en molécules plus petites appelées peptides protéolytiques qui seront analysés par MS. En *top-down*, par contre, la protéine entière est injectée dans le spectromètre de masse où elle est fragmentée et ses fragments analysés pour son identification. Le mode de fragmentation est d'une grande importance dans l'identification des protéines quelle que soit l'approche (*bottom-up* ou *top down*). Différentes techniques de fragmentation existent en MS. Le choix d'un mode de fragmentation dépend principalement de l'approche choisie et de l'appareil utilisé. L'analyse des fragments d'un peptide génère un spectre MS/MS qui va être utilisé pour son séquençage. En principe, la détermination de la séquence du peptide se fait en considérant la différence en masse entre 2 pics voisins dans le spectre MS/MS. Un système de nomenclature des fragments a été mis en place pour faciliter la lecture des spectres de fragmentation²⁹. Les fragments a, b, et c sont des fragments chargés comprenant le côté N-terminal du peptide fragmenté tandis que les fragments chargés contenant le côté C-terminal sont nommés x, y ou z. Les fragments N-ter sont nommés a, b ou c en fonction du type de la liaison rompue durant la fragmentation, pareil pour les fragments C-ter (x, y ou z) (Figure 4).

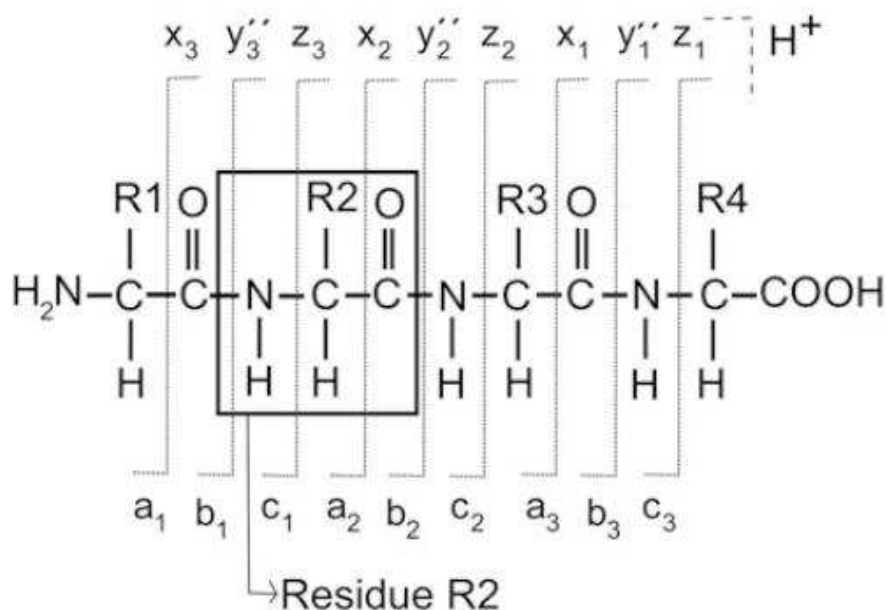


Figure 4 : Nomenclature des fragments ¹²⁴

Parmi les modes de fragmentation mis en œuvre en protéomique, le CID (*Collision Induced Dissociation*, dissociation induite par collision) est de loin le plus utilisé¹²⁸. La fragmentation CID commence par l'isolement d'un ion précurseur suivi par une série de collision de ce dernier avec un gaz neutre (Hélium) présent dans la cellule de fragmentation à une pression variant entre 1 et 5 mbar¹²⁴. Ces collisions conduisent à une augmentation de l'énergie vibrationnelle des ions précurseurs qui, une fois un certain seuil atteint, entraîne la rupture de liaisons covalentes. Le CID est utilisé dans des appareils de type triple-quadrupôle, piège quadripolaire et q-tof. Pour les appareils équipés d'un Orbitrap, c'est la fragmentation HCD *Higher-energy C-trap Dissociation*, dissociation induite par collision à plus haute énergie). La fragmentation HCD est similaire au CID, elle génère les mêmes types de fragments. Elle génère principalement des ions b et y. En dehors du CID, d'autres modes de fragmentation sont utilisés en protéomique comme l'ECD (*Electron Capture Dissociation*, dissociation par capture d'électron) ou l'ETD (*Electron Transfer Dissociation*, dissociation par transfert d'électron). Les modes de fragmentations diffèrent d'abord de par les appareils dans lesquels ils peuvent être mis en œuvre. La fragmentation ETD est mis en œuvre dans les appareils Q-TOF, Q-TRAPs ou LTQ-Orbitrap. On retrouve la fragmentation ECD dans des appareils de type FT-ICR (Analyse par Transformée de Fourier de la Résonance Cyclotronique des Ions). Ensuite, le CID diffère de l'ECD et de l'ETD de par le type de fragments générés. La fragmentation CID fournit des fragments b,y tandis que l'ECD ou l'ETD génère des fragments de type c, z¹²⁸.

Durant cette étude, le spectromètre de masse utilisé est un spectromètre de masse hybride associant un filtre quadripolaire, une cellule de collision et un analyseur de type Orbitrap (Q Exactive HF™, ThermoFisher Scientific) avec un mode de fragmentation de type HCD.

2.1.1 Approche *top-down*

Dans une expérience *top-down*, les protéines purifiées et intactes sont séparées avant analyse en MS, détectées puis fragmentées par ECD ou ETD (ou autres) apportant ainsi des informations sur la masse intacte de la protéine et sur sa séquence en acides aminés. L'approche *top-down* réduit considérablement la préparation d'échantillon et préserve ainsi des informations importantes qui seraient perdues avec une approche *bottom-up* comme la détection simultanée de plusieurs PTMs distantes sur la séquence primaire. Il est possible avec cette technique, d'avoir une description complète de la séquence de la protéine si assez de fragments informatifs sont observés²³. C'est une approche très utilisée dans l'étude des protéoformes et des systèmes biologiques ayant des combinaisons complexes de PTM²⁴. Les sources d'ionisation ESI et MALDI peuvent être utilisées pour une analyse *top-down*. L'ionisation ESI est plus avantageuse car elle produit des ions ayant des états de charge plus élevés facilitant ainsi la fragmentation²⁵, et parce que la gamme de masse de détection nécessaire est restreinte sur l'échelle des m/z . Du fait de la taille importante des protéines intactes et donc des états de charges atteints, une analyse des protéines par *top-down* se fait généralement avec des spectromètres de haute résolution du type FTICR ou Orbitrap. Cependant, des spectromètres de masse moins performants du type quadripôle associé à une fragmentation dans la source ESI ont été utilisés pour l'analyse en *top-down* d'une protéine intacte de plus de 60 kDa²⁶. Malgré tous ses avantages, l'approche *top-down* est beaucoup moins utilisée que la protéomique *bottom-up*. Ceci s'explique par plusieurs raisons. i) la solubilité des protéines est plus faible que celle des peptides. La solubilité des protéines, particulièrement les protéines membranaires est l'un des premiers challenges de la protéomique *top-down*. Il est en effet nécessaire, après leur extraction, de maintenir les protéines solubles avant, durant et après le fractionnement. Les détergents permettant de solubiliser ces protéines étant souvent incompatibles avec la MS, de nouveaux détergents sont proposés pour bien solubiliser les protéines sans gêner ensuite leur analyse en *top-down* mais s'avèrent souvent moins performants²⁴. ii) Les peptides se prêtent mieux à la séparation chromatographique, à l'ionisation et à la fragmentation que les protéines. La chromatographie en phase liquide à polarité de phase inversée est la plus fréquemment utilisée pour le fractionnement des échantillons complexes de protéines intactes car elle est directement compatible avec le mode d'ionisation ESI. Cependant, du fait d'une grande diversité des propriétés physico-chimiques d'un mélange de protéines intactes et de leur faible taux de récupération, la capacité de pics et la résolution de la séparation de protéines intactes est de loin plus faible que celles des peptides²⁷. iii) les spectres issus d'une analyse *top-down* sont beaucoup plus complexes. L'ionisation en ESI des protéines intactes produit une large gamme d'états de charge. Ainsi, une protéoforme détectée sous différents états de charge aura un profil de fragmentation qui apporte des informations redondantes contenues dans des spectres très complexes qu'il faut déconvoluer. Ceci représente un véritable challenge de traitement du signal pour l'analyse bio-informatique qui devient plus complexe lorsque sont intégrées les combinaisons complexes de PTMs inconnues. Récemment, plusieurs outils bio-informatiques ont été développés pour

l'analyse des données issus d'une analyse *top-down* mais le challenge demeure notamment pour l'identification des protéines ayant une grande masse moléculaire²⁸.

2.1.2 Approche *bottom-up*

La protéomique *bottom-up* vise à l'identification d'une protéine en analysant les peptides protéolytiques issus de sa digestion. Les protéines sont digérées en peptides ayant des tailles comprises entre 500 et 3000 Dalton. On parle de *shotgun proteomics* quand la protéomique *bottom-up* est appliquée à un échantillon complexe. Ce qui implique généralement, dans le cas du *shotgun proteomics*, le couplage d'une étape de séparation du type chromatographie liquide (LC) à une ou 2 dimensions à la spectrométrie de masse. Les peptides issus de la digestion des protéines sont ainsi fractionnés par chromatographie liquide avant d'être détectés par spectrométrie de masse d'abord en mode MS et puis en tandem (MS/MS) pour obtenir leur spectre de fragmentation. On parle ainsi de LC-MS/MS. L'identification des peptides se fait ensuite par la comparaison du spectre MS/MS issu de la fragmentation des peptides aux spectres MS/MS théoriques de peptides issus de la digestion *in silico* d'une base de données de protéines.

La digestion est l'étape la plus critique de la stratégie *bottom-up*. Elle peut être enzymatique ou chimique. Nous ne parlerons ici que la digestion enzymatique qui est de loin l'approche la plus utilisée. Un protocole typique d'une analyse *bottom-up* commence par l'extraction de protéines d'un échantillon biologique (cellules, tissus...). Cette étape consiste à extraire les protéines en détruisant la structure de la membrane cellulaire et en rompant les interactions physico-chimiques entre protéines ou avec d'autres biomolécules. Elle se fait souvent par dénaturation en présence d'un détergent comme le SDS (Sodium DodecylSulfate) dont le rôle est de solubiliser les protéines. Les protéines membranaires sont caractérisées par une faible solubilité. L'utilisation du SDS (ou autres détergents) est essentielle pour la solubilisation et l'extraction de ces dernières. Pour préparer les protéines extraites à la digestion, leurs ponts disulfures sont réduits par des agents réducteurs comme le dithiothréitol (DTT), le tris(2-carboxyethyl)phosphine (TCEP) ou le β -mercaptoéthanol. Ensuite, un agent alkylant comme l'iodoacétamide (IAM) ou le N-EthylMaléimide(NEM) est utilisé pour bloquer les cystéines et empêcher la reformation des ponts disulfures. Les réactifs en excès sont éliminés et un changement de tampon est opéré. Enfin, l'ajout d'une endoprotéase ayant des sites de coupure spécifiques (ou non) permettra de digérer les protéines en peptides. Arg-C, Lys-C, trypsine, chymotrypsine, pepsine, pronase... sont toutes des enzymes de digestion de protéines. Chacune possède des spécificités de sites de coupure et des conditions optimales de fonctionnement (température, pH...). Cependant, la trypsine est sans aucun doute l'enzyme qui est de loin la plus utilisée en *protéomique bottom-up*. La trypsine est une enzyme très spécifique qui coupe les liaisons peptidiques du côté C-terminal des résidus Arginine et Lysine si ces résidus ne sont pas suivis par la proline. En dehors de son coût abordable et de sa disponibilité en grande quantité, la large utilisation de la trypsine s'explique par l'abondance de la lysine et de l'arginine dans les séquences protéiques. De plus les peptides tryptiques portent au moins 2 charges positives, ce qui facilite leur ionisation et leur fragmentation en CID. Les spectres de fragmentation de ces peptides

sont d'une grande qualité, ce qui facilite leur identification par comparaison avec les spectres issus des bases de données de protéines.

La digestion augmente numériquement le nombre de molécules présent dans l'échantillon en transformant les protéines intactes en de nombreux peptides. Cependant, à la différence du mélange de protéines intactes, les peptides générés présentent une certaine homogénéité physico-chimique rendant ainsi la manipulation de cet échantillon moins complexe. Une conséquence directe de cette observation est que la séparation en chromatographie liquide d'un mélange de peptides est plus reproductible et résolue que celle d'un mélange de protéines intactes. L'autre avantage est qu'avec plusieurs peptides provenant d'une même protéine, la probabilité d'identifier cette protéine à partir de l'un de ses peptides en devient plus grande. L'approche *bottom-up* permet ainsi d'identifier un plus grand nombre de protéines par rapport au *top-down*. L'approche *bottom-up* a cependant quelques inconvénients : i) une importante partie du traitement d'échantillon entraîne une perte d'échantillon lors du changement de tampon ou de la suppression de réactifs en excès, ii) une perte d'information : la digestion simplifie l'échantillon mais entraîne la perte d'information notamment sur les positions et la coexistence des PTMs iii) risque de modification de l'échantillon : avec la manipulation de l'échantillon durant ce protocole long et fastidieux, il n'est pas exclu que quelques modifications de l'échantillon ne se produisent ou se dégradent (perte ou apparition de nouvelles PTMs).

2.2 La quantification en spectrométrie de masse

D'importantes avancées en appareillage en spectrométrie de masse ont été réalisées ces dernières décennies. Par ailleurs, les méthodes de séparation comme la chromatographie liquide ont aussi connu une importante amélioration de leurs performances. Des percées ont aussi été réalisées dans le traitement bioinformatique des spectres de masse. Ces progressions ont permis une couverture de plus en plus large du protéome. Cependant, l'analyse qualitative consistant à générer une liste de protéines identifiées d'un organisme donné n'est pas suffisante pour la compréhension de l'état physiologique de cet organisme³⁰. La quantification des protéines par l'approche *bottom-up* peut être absolue (estimation de la concentration en protéine dans un échantillon) ou relative (comparaison de la concentration des protéines entre 2 échantillons)³¹. La quantification en MS dépend largement de la méthode d'acquisition. Les 3 catégories de méthode d'acquisition existent aujourd'hui en MS. La *Data-Dependant Acquisition* (DDA, Acquisition dépendante des données) où les ions fragmentés et analysés en MS2 sont choisis en fonction de leur intensité. L'acquisition ciblée (PRM, SRM) : les ions à fragmenter sont choisis selon une liste définie à l'avance et incluse dans la méthode. Et enfin la *Data Independent Acquisition* (DIA, acquisition indépendante des données) où les ions précurseurs dans une certaine gamme de masse sont fragmentés sans discrimination. Au cours de cette thèse, toutes les acquisitions ont été faites en mode DDA (Annexe 1): elle est composée de deux étapes. La première, appelée MS1 sert à mesurer l'intensité des molécules séparées en fonction de leur rapport m/z . Ensuite, les ions correspondant aux intensités les plus élevées lors du MS1 sont sélectionnés et fragmentés lors de la seconde

étape appelée MS2 permettant d'avoir des informations structurales sur les ions. Les techniques de quantification en MS peuvent se baser sur les données recueillies lors du MS1 ou du MS2.

La chromatographie liquide couplée à la spectrométrie de masse est aujourd'hui la méthode la plus couramment utilisée pour l'identification des protéines. Cependant, la quantification par cette méthode est plus complexe. Ceci est dû d'abord à la préparation des échantillons. En effet, l'extraction des protéines n'est jamais complète, reproductible et uniforme pour tous les types de protéine. Pour l'approche *bottom-up*, caractérisée par une importante étape de préparation, les pertes d'échantillons par adsorption par exemple sont incompatibles avec la quantification. De plus, la spectrométrie de masse est une technique non quantitative entre molécules différentes. En effet, l'efficacité d'ionisation dépend des propriétés physico-chimiques des peptides ou protéines analysés. Les réponses de détection en MS de 2 peptides diffèrent en fonction de leurs propriétés physico-chimiques et non pas uniquement en fonction de leur concentration.

Des stratégies de quantification en MS ont été malgré tout développées pour profiter de la sensibilité et de la rapidité d'analyse de cette approche. Ces stratégies peuvent être divisées en 3 catégories : l'utilisation d'étalon interne, le marquage isotopique et enfin la quantification sans marquage appelé *Label Free*.

2.2.1 Etalonnages internes

Le recours à un étalon interne consiste à ajouter à l'échantillon une quantité connue d'une protéine ou peptide (selon l'étape à laquelle on souhaite utiliser l'étalon interne). Les peptides ou protéines étalons sont synthétisés avec des isotopes stables pour mimer les peptides ou protéines de l'échantillon. Ces étalons sont ensuite utilisés pour réaliser une quantification absolue en prenant en compte la différence entre la concentration obtenue et la concentration attendue des étalons ajoutés à l'échantillon. Cette méthode permet de s'affranchir des biais du protocole d'analyse des protéines. Tous les biais situés en amont de leur ajout dans le protocole d'analyse seront pris en compte. Il existe différentes approches de quantification absolue utilisant des étalons internes : on peut citer la méthode AQUA (*Absolute Quantification*) utilisant des peptides synthétiques marqués aux isotopes lourds (^{13}C , ^{15}N). Ces étalons sont ajoutés après la digestion et permettent donc de s'affranchir des biais situés en amont de cette étape dans le protocole d'analyse¹¹⁸. L'approche PSAQ (*Protein Standard Absolute Quantification*) consistant à utiliser des protéines recombinantes comme étalon interne permet de prendre en compte les biais liés à la digestion. Les étalons sont en effet introduits avant cette étape¹¹⁹.

La méthode est d'autant plus précise que l'étalon est introduit assez tôt dans le protocole d'analyse. L'utilisation de protéine étalon serait alors plus précise que l'utilisation de peptide. Cependant, les coûts des protéines ou peptides étalons sont élevés. Etant donné qu'il faudrait une protéine étalon pour chaque protéine à quantifier, l'utilisation de cette méthode n'est pas viable à l'échelle du protéome. L'utilisation d'étalons internes est donc une technique de quantification absolue ciblée. En effet, seuls

les protéines ou peptides dont les formes marquées (marquage isotopique) ont été ajoutées à l'échantillon pourront être quantifiées. Le marquage isotopique peut cependant être utilisée pour réaliser une quantification relative non ciblée entre 2 échantillons. Ces approches sont développées dans le paragraphe suivant.

2.2.2 Marquage isotopique

L'utilisation d'isotopes stables a pour but de créer un marqueur de masse (différence de quelques Da entre deux peptides identiques et issus de deux échantillons différents) mais de séquence identique). L'étiquette est spécifique à un échantillon et détectable en MS ou en MS/MS. Il est donc possible de mélanger des échantillons préalablement marqués à différentes étapes du protocole d'analyse. Il faut noter que les variations de toutes les étapes en aval du mélange sont prises en compte dans la quantification, d'où l'intérêt de faire le marquage le plus tôt possible. L'abondance relative peut être ensuite obtenue en mesurant les intensités des pics issus de chaque échantillon en les distinguant grâce aux marqueurs isotopiques. Les approches de marquage à l'isotope stable diffèrent par rapport aux stratégies de quantification, l'étape à laquelle le marquage est réalisé et le nombre de conditions comparable.

Isotope Coded Affinity Tag (ICAT) : spécifique aux résidus cystéines, ce marquage s'effectue avant la digestion enzymatique mais ne concerne que les peptides intégrant un résidu cystéine, séparés du reste de l'échantillon avant l'analyse en LC-MS¹²⁹. Le marqueur ICAT est composé d'une moitié IAM (partie réactive et spécifique aux résidus cystéines), d'une biotine clivable (pour l'enrichissement par affinité des peptides marqués) et enfin d'une chaîne carbonée (9 carbones) contenant des isotopes lourds (¹³C) ou légers (¹²C) : ce qui introduit une différence de masse 9 Da reconnaissable lors de l'analyse MS. On parle ainsi de ICAT léger portant une chaîne légère et de ICAT lourd portant une chaîne lourde. Ces 2 versions permettent de réaliser une quantification relative entre 2 échantillons. Cette quantification ne peut cependant concerner que les peptides intégrant un résidu cystéine. Ce qui en fait une approche adaptée à l'étude de l'oxydation des cystéines : elle a d'ailleurs été utilisée pour déterminer l'équilibre entre les cystéines oxydées (marquées avec un ICAT lourd) et les réduites (marquées avec un ICAT léger) du protéome au sein d'un même échantillon suite à une exposition à un stress oxydatif³². Cependant, si on dispose de deux échantillons ayant subi des conditions de stress oxydatif différentes, la comparaison de l'équilibre entre les cystéines réduites et oxydées entre les deux échantillons requiert plus de deux versions de ICAT. Cette méthode ne permet donc pas de réaliser la quantification relative de l'oxydation des cystéines entre deux échantillons.

Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) : est une méthode de marquage métabolique consistant à incorporer des acides aminés marqués aux isotopes lourds dans les protéines d'une culture cellulaire. Cette incorporation se fait durant la synthèse protéique et les divisions cellulaires. Le marquage SILAC, développé par Ong et al.,³³ en 2002 est considérée comme la moins biaisée de toutes les méthodes de marquage étant donné que le mélange des échantillons à comparer se fait avant même l'extraction des protéines. Cependant, le marquage SILAC ne peut être appliqué à tous les types

d'échantillons car elle nécessite une culture cellulaire. Le marquage métabolique qui utilise l'azote 15 est applicable à davantage de types d'échantillon mais elle est très peu utilisée car le traitement des spectres de masse issus de cette approche s'avère complexe¹⁶. Le marquage métabolique SILAC a été utilisé dans le protocole OcSILAC pour la quantification de l'oxydation des cystéines de 2 échantillons issus de différentes conditions de stress oxydant.

Les quantifications par ICAT et SILAC sont basées sur la MS1. En effet la détection de l'écart de masse et la comparaison des intensités MS1 permet de faire la quantification relative tandis la MS2 est utilisé pour déterminer la séquence. D'autres méthodes de marquage isotopique permettent de faire une quantification MS2 permettant ainsi une co-isolation et une co-fragmentation des peptides. L'*isobaric tag for relative and absolute quantification* (ITRAQ) et le *tandem mass tag* (TMT) sont dotés de marqueurs isobares (de la même masse) qui ne révèlent une différence qu'après la fragmentation du peptide en deux fragments de même structure et de masse inégale en fonction de l'étiquette utilisée et dont la somme des masses est constante. Les derniers résultats ont utilisé jusqu'à 8 marqueurs isotopiques différents en TMT par exemple³⁴.

Les méthodes de marquage présentent une grande précision surtout quand elles sont effectuées assez tôt dans le protocole de préparation d'échantillon. Elles permettent aussi de sauvegarder du temps machine car les échantillons à comparer sont analysés simultanément. Cependant, elles entraînent une dilution isotopique de l'échantillon et les étapes supplémentaires allonge les protocoles de préparation d'échantillon. Les marquages détectables en MS rendent plus complexe l'analyse du mélange de molécules. Les marquages détectables en MS/MS lissent le signal par risque de co-sélection d'espèces quasi-isobares marquées et sélectionnées pour leur fragmentation. Par ailleurs, le marquage isotopique présente un coût assez conséquent qui peut pousser à adopter des méthodes de quantification sans marquage.

2.2.3 La quantification sans marquage (*LabelFree*)

La quantification sans marquage est une alternative sérieuse face au coût et au temps de manipulation inhérents aux méthodes de quantification avec marquage. Elle profite aussi des avancées réalisées dans le traitement des spectres MS³⁵. C'est une méthode de quantification universelle (tous types d'échantillon) qui permet de comparer un nombre infini d'échantillons. Dans cette approche, les échantillons sont analysés séparément et la quantification peut être effectuée en se basant sur des données MS1 ou MS2. En effet, la quantification relative par la mesure de l'aire sous le pic (*AUC Area Under Curve*) se base sur les intensités MS1. La somme ou la moyenne des aires est ensuite comparée protéine par protéine. Le décompte des spectres MS/MS est basée sur les données MS2 et consiste à comparer le nombre de spectres MS/MS attribué à une protéine donnée. Cependant, la précision de la quantification *Label Free* repose en grande partie sur la reproductibilité des étapes de préparation d'échantillon. Les échantillons sont en général traités en parallèle pour limiter les variations au maximum, mais cela n'élimine pas les biais spécifiques du au traitement séparé de chaque tube.

2.3 Limites et challenges de la protéomique *bottom-up*

La protéomique *bottom-up* est la plus couramment utilisée pour l'identification des protéines car les méthodes de séparation des peptides tryptiques en chromatographie liquide, leur fragmentation et leur détection en spectrométrie de masse sont déjà éprouvées. L'identification d'un peptide grâce à un spectre MS/MS de grande qualité permet de remonter facilement à la protéine dont il provient. Cependant, l'ensemble des peptides d'une protéine est rarement identifié. En effet, tous les peptides tryptiques générés ne sont pas détectés et seulement une partie des peptides détectés fournissent un spectre de fragmentation exploitable. Dans l'étude des PTMs, il faut prendre en considération le fait que seule une portion de la protéine est modifiée et qu'il est difficile d'identifier le ou les peptides portant ces modifications. Il est cependant possible de marquer spécifiquement les peptides modifiés, de les enrichir et les analyser séparément. Cette méthode entraîne un allongement du protocole de préparation d'échantillon pour une analyse *bottom-up* qui est déjà long. Les pertes d'échantillons et les modifications possiblement induites durant le protocole sont aussi à prendre en considération.

2.3.1 L'étude des modifications post-traductionnelles en protéomique *bottom-up*

La transcription de l'ADN génère de l'ARNm qui est ensuite traduit en protéines. Les protéines peuvent ensuite faire l'objet de différentes modifications post-traductionnelles pouvant affecter leur structure et donc leur fonction. La caractérisation des PTMs est donc une partie importante de la description du protéome.

La digestion tryptique génère parfois des peptides de petite taille (4 à 5 acides aminés) qui ne peuvent pas être identifiés par chromatographie liquide couplée à la spectrométrie de masse. Ces peptides sont généralement trop hydrophiles pour être retenus en phase inverse et seront perdus dans le volume de tête d'élution. Si le système est assez perfectionné pour permettre leur identification, leur courte séquence est généralement partagée par plusieurs protéines. Il devient dès lors difficile de situer les PTMs qu'ils intègrent. Un peptide de taille moyenne portant une PTM peut être identifiée avec une séquence qui ne renvoie qu'à une seule protéine. Il est cependant impossible, avec l'approche *bottom-up*, de remonter à la combinaison de PTMs localisées dans différents peptides protéolytiques. Dans le cas, par exemple, d'un pont disulfure, il est impossible de savoir s'il était établi au niveau intra- ou inter-protéique.

L'utilisation de détergent ou d'agent chaotrope (urée, thiourée...) est nécessaire pour la solubilisation des protéines. Cependant ces composés ne sont pas toujours compatibles avec les conditions nécessaires à une digestion tryptique efficace. L'une des approches les plus utilisées pour l'élimination de ces composés avant la digestion est la précipitation acide des protéines, leur rinçage avec de l'acétone et leur reprise avec un tampon dont la composition et le pH sont adaptés à la digestion tryptique. Malheureusement, cette approche entraîne une perte d'échantillon non négligeable obligeant à travailler avec au moins 10 µg de protéine minimum au départ.

2.3.2 Etude des modifications post-traductionnelles par MS : Cas de OcSILAC

Plusieurs états physiologiques et pathologiques ont été associés au stress oxydant. Le stress oxydant fait référence à l'accumulation de dérivés réactifs de l'oxygène (ROS) dans l'organisme. Ils proviennent essentiellement de réactions successives de réduction du dioxygène¹²⁰. Les ROS réagissent avec l'ADN, les lipides et les protéines. La structure et la fonction de ces biomolécules peuvent être altérées en réagissant avec ces espèces³². Les protéines sont une des cibles majeures des oxydants du fait de leur grande abondance. Les ROS peuvent oxyder les protéines de manière réversible ou irréversible. Les modifications réversibles font partie du fonctionnement normal d'une protéine et sont parfois nécessaires. L'oxydation du groupement thiol des résidus cystéines en acide sulfénique (S-OH), en disulfure (-S-S) ou en S-nitrosylation (S-NO) ont déjà été identifiées comme étant des acteurs participant à la signalisation cellulaire³⁶. Cependant, les modifications irréversibles entraînent une perte définitive de la fonction d'une protéine. Le terme rédoxomique désigne l'étude de l'état redox des protéines. Ces modifications concernent les chaînes latérales des acides aminés. Parmi tous les acides aminés, la cystéine est la plus susceptible à l'oxydation du fait du groupement thiol présent dans sa chaîne latérale³⁷. La caractérisation des PTMs dans un échantillon complexe est un challenge du fait de leur faible abondance. Dans le cas particulier de la caractérisation de l'état redox de la cystéine, il faut prendre en compte la grande réactivité des thiols dans le développement d'un protocole d'analyse.

Pour l'étude de l'oxydation des cystéines, le protocole OcSILAC (Oc pour Oxidized *cysteine* et SILAC comme la méthode de marquage isotopique, Thèse Shakir SHAKIR, 2015) a été développé au sein de notre laboratoire. Il vise à caractériser l'oxydation des cystéines qualitativement et quantitativement dans le protéome. Le protocole OcSILAC combine le marquage métabolique SILAC (*Stable Isotope Labeling by Amino-acid in Cell culture*) et la technologie Biotin-switch pour l'enrichissement des cystéines oxydées vue leur faible abondance. Il comprend donc un marquage différentiel métabolique de cultures à comparer, un marquage différentiel chimique des résidus cystéine (selon qu'ils soient réduits ou oxydés) durant le protocole de préparation d'échantillon, une étape de protéolyse pour l'approche *bottom-up* et une étape d'enrichissement des cystéines initialement oxydées car elles sont généralement minoritaires, avant l'analyse en MS.

Contrôle de réactivité des thiols : Le protocole OcSILAC comprend plusieurs étapes de précipitation acide destinées à contrôler la réactivité des thiols et éviter les réactions croisées (Figure 5). Le pKa de la cystéine est de 8.3 mais peut varier entre 4 et 10 selon les acides aminés de son environnement. La forme la plus réactive du groupement thiol (-S⁻) est majoritaire quand le pH de la solution est supérieur au pKa de la cystéine. L'utilisation de l'acide trichloroacétique (TCA) permet de se placer à pH égal ou inférieur à 1. A un pH aussi bas, tous les groupements thiols des cystéines sont protonés et donc préservés d'une éventuelle oxydation³⁸.

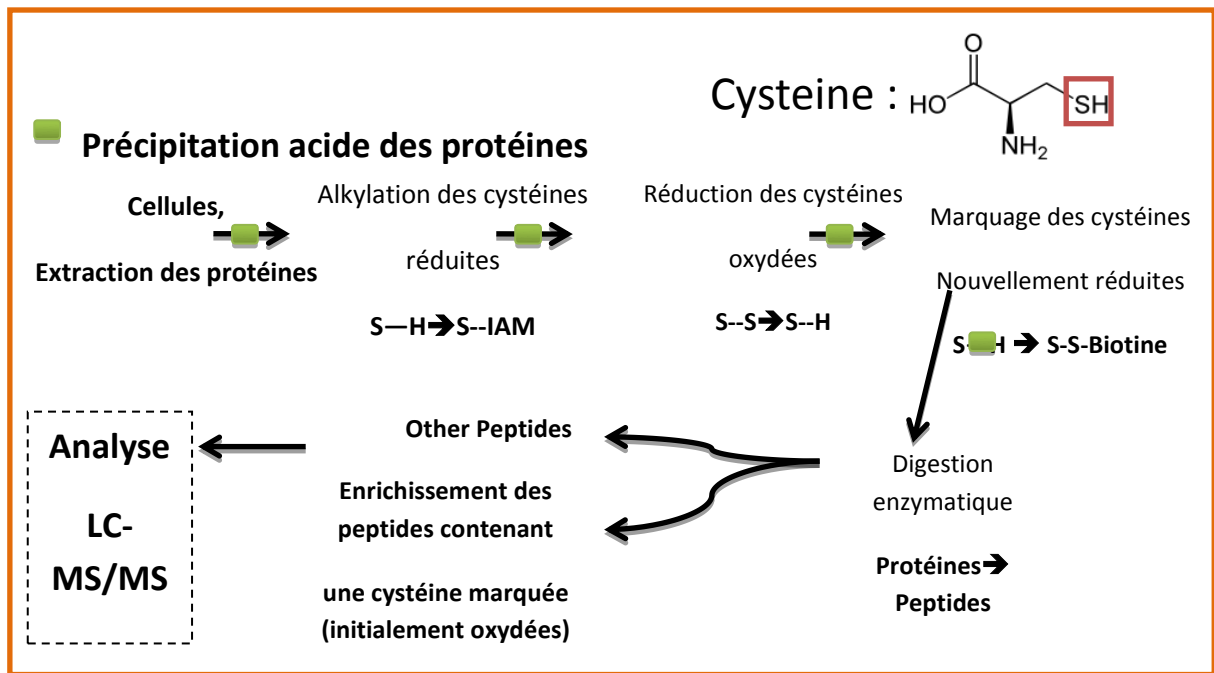


Figure 5 : Protocole OcsILAC

la précipitation acide des protéines est effectuée aux moments où les groupements thiols sont leur forme la plus réactive

Marquage métabolique : On dispose au départ de 2 cultures cellulaires. La première étape consiste à effectuer le marquage métabolique. Pour cela, l'une des 2 cultures est effectuée dans un milieu contenant de la lysine et de l'arginine marquées aux isotopes lourds : K8 : $^{13}\text{C}_6^{15}\text{N}_2$ et R10 : $^{13}\text{C}_6^{15}\text{N}_4$. Ce qui introduit un écart de masse de 8 et 10 Dalton respectivement par rapport à la lysine et l'arginine non marquées. Après ce marquage, on dispose de 2 cultures cellulaires dont la seule différence est la présence d'acides aminés lourds (Lysine, Arginine) dans l'une des 2. Chaque culture est exposée à l'une des 2 conditions à comparer (différents niveaux de stress oxydant par exemple).

Préparation de l'échantillon : Pour l'analyse des protéines par l'approche *bottom-up*, on commence par mélanger un nombre équivalent de cellules marquées aux isotopes lourds (H) et légers (L). On procède ensuite à la lyse de l'ensemble des cellules du mélange en présence de TCA. La lyse cellulaire dans un environnement permet de préserver l'échantillon d'éventuelles réactions d'oxydation ou de réduction artéfactuelles. L'extraction des protéines est effectuée avec une solution de TCA 20% (v/v). Les protéines extraites et précipitées sont rincées à l'acétone froid. Le rinçage à l'acétone froid permet d'éliminer l'acide (TCA) résiduel. Puis, les protéines sont reprises avec une solution tampon contenant un agent alkylant, l'iodoacétamide (IAM) pour bloquer les cystéines réduites. L'excès d'IAM est éliminé à la fin de réaction par précipitation acide des protéines suivi d'un rinçage à l'acétone froid. Les protéines sont de nouveau suspendues dans un tampon contenant un agent réducteur, le dithiothréitol (DTT), pour réduire les protéines réversiblement oxydées. Suite à cette réaction de réduction, l'excès de DTT est éliminé par le même procédé et les protéines sont suspendues dans un tampon contenant de la biotine-HPDP destinée au marquage des cystéines nouvellement réduites (donc oxydées au départ) par formation d'un pont disulfure cystéine-biotine. L'excès de biotine-

HPDP est éliminé à son tour par le même procédé. Après cette étape, les protéines sont reprises avec une solution d'ammonium bicarbonate (ABC) 50 mM compatible avec la digestion tryptique qui est l'étape suivante. Le Biotin-switch permet ensuite d'enrichir les peptides marqués avec la biotine-HPDP en passant l'ensemble des peptides générés par la digestion sur des billes d'agarose sur lesquelles est greffée de la streptavidine. Pour finir, la fraction non retenue est recueillie après une heure de contact avec les billes d'agarose tandis que les peptides retenus spécifiquement sont ensuite élués à l'aide d'une solution de DTT. La solution de DTT permet en effet de réduire les ponts disulfures formés entre la cystéine et la biotine.

Stratégie de quantification : Avec le marquage SILAC et le mélange de quantités équivalentes de cellules marquées et non marquées, les peptides seront présents en deux versions lourde et légère, chacune provenant de l'une des conditions à comparer. Les deux versions sont identiques en tout point sauf dans leur composition isotopique, elles vont être éluées et ionisées en même temps. Une étude quantitative peut alors se faire en comparant les intensités des pics relatifs en MS des deux versions de peptide. Le rapport des intensités MS des peptides (Lourd/léger) contenant une cystéine carbamidométhylée représente le rapport des peptides réduits tandis que celui des peptides portant des cystéines sous forme libre indique le rapport de la forme oxydée. Les autres peptides, à savoir ceux qui n'ont pas de cystéine dans leur séquence primaire permettront de déterminer le niveau d'expression de la protéine associée. Leur rapport d'intensité donne accès à la variation du niveau d'expression des protéines dans chaque condition et permettra de corriger la comparaison quantitative des peptides contenant une cystéine.

Le protocole OcSILAC est à la fois plus long et plus complexe qu'un protocole *shotgun* classique. Ceci s'explique par le fait qu'il s'intéresse à la caractérisation des PTMs en plus de l'identification des protéines. Les précipitations acides des protéines suivies du rinçage avec de l'acétone froid ont été intégrées dans le but de préserver les thiols des modifications artéfactuelles malgré leur grande réactivité. Cependant cette approche allonge encore le protocole et entraîne une perte supplémentaire d'échantillon par rapport à un protocole *Shotgun proteomics* classique. Trouver une alternative à la précipitation acide des protéines et procéder à la miniaturisation du protocole permettrait de gagner du temps, de réduire la consommation en échantillon et le préserver des oxydations artéfactuelles.

