



HAL
open science

Exploration des modifications post-traductionnelles des protéines : nouvelles approches et nouveaux modèles biologiques

Alain Dedieu

► **To cite this version:**

Alain Dedieu. Exploration des modifications post-traductionnelles des protéines : nouvelles approches et nouveaux modèles biologiques. Biochimie, Biologie Moléculaire. Université Montpellier I, 2014. Français. NNT : 2014MON13516 . tel-01132156

HAL Id: tel-01132156

<https://theses.hal.science/tel-01132156>

Submitted on 16 Mar 2015

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

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

THÈSE

Pour obtenir le grade de
Docteur

Délivrée par **UNIVERSITE de MONTPELLIER I**

Préparée au sein de l'école doctorale : Sciences Chimiques et
Biologiques pour la Santé (ED168)
N° DE SECTION DU CFNU : 64
Et de l'unité de recherche Laboratoire de Biochimie des Systèmes
perturbés (CEA Marcoule – DSV)

Spécialité : **Biochimie et Biologie moléculaire**

Présentée par **Alain DEDIEU**

**EXPLORATION DES MODIFICATIONS POST-
TRADUCTIONNELLES DES PROTEINES :**
Nouvelles approches et nouveaux modèles biologiques

Soutenue le 26 novembre 2014 devant le jury composé de :

Mr **Jean ARMENGAUD**, Directeur de recherche, CEA-DSV
Mme **Julie HARDOUIN**, Maître de conférences, Université de Rouen
Mr **Luc CAMOIN**, Ingénieur de Recherche, INSERM Marseille
Mr **Fabrice CONFALIONERI**, Professeur, Paris Sud
Mr **Christian ROUMESTAND**, Professeur, Montpellier I

Président du jury
Rapporteur
Rapporteur
Examineur
Examineur

UNIVERSITE de MONTPELLIER I

ECOLE DOCTORALE: Sciences Chimiques et Biologiques pour la Santé (ED168)

SPECIALITE: Biochimie et biologie moléculaire

N° DE SECTION DU CNU: 64

THESE de DOCTORAT
pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE MONTPELLIER I
présentée par

Alain DEDIEU

EXPLORATION DES MODIFICATIONS POST- TRADUCTIONNELLES DES PROTEINES :

Nouvelles approches et nouveaux modèles biologiques

Thèse dirigée par Christian ROUMESTAND

soutenue publiquement le 26 novembre 2014 devant le jury composé de:

Dr Jean ARMENGAUD

Président du jury

Dr Julie HARDOUIN

Rapporteur

Dr Luc CAMOIN

Rapporteur

Dr Fabrice CONFALONIERI

Examineur

Dr Christian ROUMESTAND

Directeur de thèse

Citations

Solidaire ou solitaire

Une belle journée commençait, mais Jonas ne s'en apercevait pas. Il avait retourné la toile contre le mur. Épuisé, il attendait, assis, les mains offertes sur ses genoux. Il se disait que maintenant il ne travaillerait plus jamais, il était heureux. Il entendait les grognements de ses enfants, des bruits d'eau, les tintements de la vaisselle. Louise parlait. Les grandes vitres vibraient au passage d'un camion sur le boulevard. Le monde était encore là, jeune, adorable : Jonas écoutait la belle rumeur que font les hommes. De si loin, elle ne contrariait pas cette force joyeuse en lui, son art, ces pensées qu'il ne pouvait pas dire, à jamais silencieuses, mais qui le mettaient au-dessus de toutes choses, dans un air libre et vif. Les enfants couraient à travers les pièces, la fillette riait, Louise aussi maintenant, dont il n'avait pas entendu le rire depuis longtemps. Il les aimait ! Comme il les aimait ! Il éteignit la lampe et, dans l'obscurité revenue, là, n'était-ce pas son étoile qui brillait toujours ? C'était elle, il la reconnaissait, le cœur plein de gratitude, et il la regardait encore lorsqu'il tomba, sans bruit.

« Ce n'est rien, déclarait un peu plus tard le médecin qu'on avait appelé. Il travaille trop. Dans une semaine, il sera debout. - Il guérira, vous en êtes sûr ? disait Louise, le visage défait. - Il guérira. » Dans l'autre pièce, Rateau regardait la toile, entièrement blanche, au centre de laquelle Jonas avait seulement écrit, en très petits caractères, un mot qu'on pouvait déchiffrer, mais dont on ne savait s'il fallait y lire **solitaire ou solidaire**.

Albert Camus – L'exil et le royaume – Jonas ou l'artiste au travail

Pour sourire :

« Comme le citoyen Laplace présentait au général Bonaparte la première édition de son *Exposition du Système du monde*, le général lui dit : « Newton a parlé de Dieu dans son livre. J'ai déjà parcouru le vôtre et je n'y ai pas trouvé ce nom une seule fois. À quoi Laplace aurait répondu : « Citoyen premier Consul, je n'ai pas eu besoin de cette hypothèse. »

Et puis pour me faire plaisir une citation d'Einstein que certain ont voulu faire passer pour croyant !

« Le mot Dieu n'est pour moi rien de plus que l'expression et le produit des faiblesses humaines, la Bible un recueil de légendes, certes honorables mais primitives qui sont néanmoins assez puérides. Aucune interprétation, aussi subtile soit-elle, ne peut selon moi changer cela. »

Albert Einstein - lettre à Eric Gutkind 3 janvier 1954

*A Marcelle et Roger,
mes parents.....
excusez-moi pour le retard!*

*Cette thèse est aussi dédiée à toutes les personnes
qui m'ont supporté,
dans tous les sens du terme et
d'une façon ou
d'une autre, au cours de ces dernières années.*

REMERCIEMENTS

REMERCIEMENTS

Je tiens à remercier l'ensemble des membres du jury qui ont accepté d'évaluer mon travail.

Je remercie M. Jean Armengaud qui a accepté la présidence de ce jury, ainsi que Mme Julie Hardouin et M. Luc Camoin pour avoir bien voulu être rapporteur, ainsi que M. Fabrice Confaloneri d'être examinateur. A tous, merci d'avoir pris de votre temps pour juger ce travail.

Je tiens à remercier en premier lieu, le Professeur Christian Roumestand, d'avoir accepté pendant plus de trois ans d'être mon directeur de thèse et d'avoir ainsi établi le lien avec l'école doctorale de l'Université de Montpellier I.

Mes remerciements les plus vifs vont vers Jean Armengaud qui m'a donné l'opportunité d'effectuer cette thèse au sein de son laboratoire. En effet voilà près de quatre ans lorsque j'ai abordé avec lui la question d'effectuer un travail de thèse à « mon âge », il a répondu sans hésiter par l'affirmative. Il m'a de plus, soutenu sans faille pendant tout ce temps, me permettant de me spécialiser au sein de son laboratoire dans le domaine des Modifications Post-Traductionnelles. Gageons que nos « origines presque communes », le Sud-ouest, et notre passage par cette ville si attachante de Toulouse nous ont certainement rapprochés. Je tiens à souligner ses qualités de scientifiques et de manager.

Cette thèse a débuté et s'est déroulée au sein du Service de Biochimie et Toxicologie Nucléaire dirigé alors par Eric Quéméneur qui lui aussi a soutenu sans hésiter mon initiative, qu'il en soit remercié.

Je me dois aussi de remercier Catherine Duport qui a accepté de suivre l'avancement de mes travaux lors des comités de thèse. Egalement Christine Enjalbal (Professeur à Montpellier II) qui nous a toujours apporté de précieux conseils en spectrométrie de masse.

Bien sûr un travail de thèse s'insère dans la dynamique de recherche d'un laboratoire et profite de tous les efforts que ce laboratoire consent pour faire avancer cette thématique. Je dois donc remercier tous les stagiaires de Master que j'ai formés au cours de leur séjour et qui en retour ont contribué à ce travail. Leur nom évidemment figure sur les posters et publications que nous avons réalisés et qu'ils soient ici remerciés. En réussissant tous les trois brillamment leur examen de Master, ils ont contribué à ce travail. Tout d'abord Jimmy Mevaere, venu du master de génomique et protéomique de Lille et ensuite Jean-Baptiste Boyer et Damien Joulié du master « Stratégies de découverte de molécules bioactives (SDMB) dirigé à Montpellier par Christine Enjalbal. Sans oublier en ce qui concerne une période plus ancienne, Guillaume Gabant, avec qui dès 2007, j'avais travaillé sur les phosphorylations de ChK2 et qui avait brillamment passé son DRT (Diplôme de Recherche Technologique). Je n'oublie pas Mathieu Baudet, qui en 2009 avait lui aussi passé avec succès un DRT en amorçant un travail de protéogénomique sur l'identification de peptides N-terminaux chez *Deinococcus deserti*. Travailler avec tous ces jeunes gens (qui avaient l'âge de mon fils) enthousiastes et motivés fut pour moi un vrai plaisir.

Au sein du laboratoire LBSP (Laboratoire de Biologie des Systèmes Perturbés) que toute l'équipe avec laquelle je travaille depuis des années soit remerciée et plus particulièrement Jean-Charles

Gaillard notre spécialiste passionné du système nano-HPLC Orbitrap XL et qui s'est investi dans le traitement des données des modifications post-traductionnelles de la thyroglobuline chez la souris, ainsi que Philippe Guérin qui a contribué au travail fait sur nos modèles de référence *Deinococcus deserti* et *Thermococcus gammatolerans*. Egalement Bernard Fernandez, notre spécialiste des purifications sur colonne, jamais avare de conseils et qui a contribué en particulier au travail sur la protéomique structurale. Enfin Olivier Pible bio-informaticien passionné qui a su apporter à ce travail des petits et grands coups de pouces.

La relecture du document a été consciencieusement réalisée par deux de mes collègues féminines, Béatrice Alpha Bazin et Agnès Hagège (notre collègue berrichonne du CNRS). Cette dernière me fit une relecture express qui révéla de si nombreuses coquilles.....

Enfin un travail sur plusieurs années s'effectue aussi avec des collègues avec qui l'on ne partage pas nécessairement de longues discussions scientifiques et techniques mais aussi les petites interrogations de chaque jour et les grandes interrogations et rigolades du moment.

Que soient remerciées Fabienne ma collègue de bureau qui m'a trop longtemps supporté, Frédéric au courant de tout et du reste, Gilles ou Odette, témoins des grandes interrogations, Mickaël témoin de quelques facéties....

RESUME

Exploration des modifications post-traductionnelles des protéines : Nouvelles approches et nouveaux modèles biologiques

RESUME

L'étude des modifications post-traductionnelles a connu au cours des dernières années un regain d'intérêt notable. Tout d'abord car elle s'effectue aujourd'hui au travers d'approches basées sur la spectrométrie de masse, technique qui pendant cette période a connu de profonds bouleversements, conduisant à des études plus aisées et systématiques.

Mais aussi car tant par leur variété que par le rôle qu'elles jouent dans la vie et la régulation cellulaire, ces modifications ne peuvent plus être négligées. Par ailleurs au cours de ces quinze dernières années, nous avons assisté concernant les procaryotes à un changement total de paradigme. En effet à la fin des années 90, l'idée dominante était que ces modifications pouvaient exister chez ceux-ci mais de façon très partielle et/ou très particulière.

Dans ce travail, les divers degrés d'iodation de la tyrosine ont été sondés par une approche de type «shotgun» sur un organe entier, la thyroïde de souris. L'efficacité de ce type d'approche démontrée, les modifications post-traductionnelles potentiellement présentes dans des organismes modèles radiorésistants, la bactérie *Deinococcus deserti* et l'archée *Thermococcus gammatolerans* ont été analysées. Dans le premier cas, les données de protéomique montrent que de nombreuses acétylations N-terminales portent sur un motif spécifique (essentiellement des thréonines et sérines), cas très atypique pour une bactérie. Chez *Thermococcus gammatolerans* les acétylations N-terminales sont rares, mais la présence d'acétylations sur les chaînes latérales des lysines est notable. La présence de phosphorylations sur ces mêmes protéines, laisse entrevoir un possible phénomène de «cross talk» entre les lysines acétylées et les sérines, thréonines phosphorylées.

Ici, nous démontrons que la complexité du protéome chez les procaryotes par le biais des MPT est bien réelle et que de possibles interdépendances entre MPT mériteraient un regard nouveau.

Mots clés: Modifications post-traductionnelles (MPT), spectrométrie de masse, procaryote, radio-résistant, acétylations N-terminale, acétylations des lysines, acétylome.

Exploration of protein post-translational modifications: New approaches and novel biological models

ABSTRACT

Recently, the study of post-translational modifications has greatly evolved, mainly because of crucial progresses in mass spectrometry methodology which have allowed high-throughput, high resolution analysis. Their variety and their role in the regulation of key molecular mechanisms are increasingly documented. In this work, the different degrees of iodination of tyrosine were probed with a "shotgun" approach carried out from an entire organ, the mice thyroid. Post-translational modifications present in two radioresistant organism models, the bacterium *Deinococcus deserti* and the archaeon *Thermococcus gammatolerans*, were analyzed. The large scale exploration of N-terminal acetylation in *D. deserti* indicates a specific pattern of this modification on serine and threonine, as well as an atypical, high propension to acetylation with 50% of modified N-termini.

In *T. gammatolerans*, N-terminal acetylation is rare, but the presence of acetylation on lysine side chains is significant. The presence of phosphorylation on these proteins suggests a potential "cross talk" between the acetylated lysine and phosphorylated serine or threonine residues. This work demonstrates that the complexity of the proteome in prokaryotes through post-translational modifications is higher than expected when extremophiles are scrutinized compared to classical prokaryote models. Interdependencies between post-translational modifications definitively deserve a fresher look.

Key words: Post Translational Modification (PTM), mass spectrometry, prokaryote, radio-resistant, N-terminal acetylation, lysine acetylation, acetylome.

TABLE DES MATIERES

TABLE DES MATIERES SIMPLIFIEE

| | |
|---|------------|
| 1. INTRODUCTION BIBLIOGRAPHIQUE | 24 |
| 1.1 LES MODIFICATIONS POST-TRADUCTIONNELLES DES PROTEINES | 24 |
| 1.2 LA SPECTROMETRIE DE MASSE, UN OUTIL ANALYTIQUE EN TRES FORTE EVOLUTION, TECHNIQUE CLEF DE LA PROTEOMIQUE | 54 |
| 1.3 LA PREPARATION DE L'ECHANTILLON POUR ANALYSE DES MPTS | 68 |
| 1.4 LES APPROCHES PROTEOMIQUES, DU TRADITIONNEL GEL 2D AUX ANALYSES SHOTGUN GLOBALES ET LES CONSEQUENCES POUR LA RECHERCHE DES MPT. | 81 |
| 1.5 PRESENTATION DU SUJET DE THESE | 93 |
| 2. MATERIELS ET METHODES | 98 |
| 2.1. DESCRIPTIF DE L'ORBITRAP XL | 98 |
| 2.2 OPTIMISATION DE LA METHODE D'ACQUISITION | 99 |
| 2.3 PROTOCOLE DE DIGESTION | 101 |
| 3. RESULTATS | 105 |
| 3.1 -L'APPROCHE PROTEOMIQUE SHOTGUN GLOBALE, UN OUTIL DE CHOIX POUR L'ETUDE DES MODIFICATIONS POST-TRADUCTIONNELLES | 105 |
| 3.2 - DECOUVERTE D'UNE MODIFICATION POST-TRADUCTIONNELLE CHEZ DEINOCOCCUS DESERTI EN CONDITION DE STRESS | 117 |
| 3.3 - DEINOCOCCUS DESERTI UNE BACTERIE AUX ACETYLATIONS N-TERMINALES ABONDANTES ET CARACTERISTIQUES | 133 |
| 3.4 - N- AND O- ACETYLATION LES PIEGES A EVITER ! | 146 |
| 3.5 - THERMOCOCCUS GAMMA TOLERANS, UNE ARCHEA AUX LYSINES E-ACETYLEES | 152 |
| 3.6 – DEFINIR DES OBJECTIFS EN BIOLOGIE STRUCTURALE A TRAVERS LE PRISME DE LA PROTEOGENOMIQUE, L'EXEMPLE DE LA PROTEINE TGAM_1934 DE L'ARCHEE THERMOCOCCU GAMMATOLERANS | 170 |
| 4. DISCUSSION | 183 |
| 4.1 – MAITRISER LA PROTEOMIQUE ET L'OUTIL SPECTROMETRIE DE MASSE POUR ALLER PLUS LOIN.... | 183 |
| 4.2 – ETABLIR LA PRESENCE DE MODIFICATIONS POST-TRADUCTIONNELLES DANS DIVERS ORGANISMES | 184 |
| 4.3 – ROLE DES MODIFICATIONS POST-TRADUCTIONNELLES, UTILISER LE MODELE PROCARYOTE | 185 |
| 4.4 – UN NOUVEAU NIVEAU DE COMPLEXITE LE « DIALOGUE » DES MPT | 187 |

TABLE DES MATIERES COMPLETE

| | |
|---|-----------|
| 1. INTRODUCTION BIBLIOGRAPHIQUE | 24 |
| 1.1 LES MODIFICATIONS POST-TRADUCTIONNELLES DES PROTEINES | 24 |
| 1.1.1 AU-DELA DE LA SYNTHÈSE ET DE LA MATURATION DES PROTEINES, LES MODIFICATIONS POST-TRADUCTIONNELLES | 24 |
| 1.1.2 LES MODIFICATIONS POST-TRADUCTIONNELLES (MPT) DEFINITION : | 25 |
| 1.1.3 LISTE DES PRINCIPALES MPT ET DE LEURS CARACTERISTIQUES | 27 |
| 1.1.3.1 Au commencement était la méthionine.... | 28 |
| 1.1.3.2 Les phosphorylations | 28 |
| 1.1.3.2.1 Une modification réversible | 29 |
| 1.1.3.2.2 Une estérification des résidus alcool (plus rarement amine) des chaînes latérales de quelques acides aminés | 29 |
| 1.1.3.2.3 Une modification chimique aux multiples aspects | 30 |
| 1.1.3.2.4 Une modification chimique jouant un rôle essentiel dans les diverses régulations cellulaires | 30 |
| 1.1.3.2.5 Une modification post traductionnelles très répandue | 30 |
| 1.1.3.3 Les acétylations, une MPT type Janus | 31 |
| 1.1.3.3.1 Différentes acétylations | 31 |
| 1.1.3.3.2 Conséquences physico-chimiques de cette modification | 32 |
| 1.1.3.4 Les glycosylations ou le monde complexe des sucres ¹ | 34 |
| 1.1.3.4.1 Deux grands types de glycosylation | 35 |
| 1.1.3.4.2 Rôle des glycosylations | 37 |
| 1.1.3.4.3 Comment caractériser les glycosylations par spectrométrie de masse | 37 |
| 1.1.3.5 La méthylation, une MPT modeste qui joue dans la cour des grands ! | 45 |
| 1.1.3.6 L'ubiquitination (ou ubiquitylation) et la sumoylation – les sœurs Garnier des MPT ! | 46 |
| 1.1.3.6.1 L'ubiquitination est une modification | 46 |
| 1.1.3.6.2 La sumoylation | 49 |
| 1.1.4 QUELQUES MPTS SPECIFIQUES ET LEURS CARACTERISTIQUES | 50 |
| 1.1.4.1 L'iodination des tyrosines | 51 |
| 1.1.4.2 La bromation | 53 |
| 1.2 LA SPECTROMETRIE DE MASSE, UN OUTIL ANALYTIQUE EN TRES FORTE EVOLUTION, TECHNIQUE CLEF DE LA PROTEOMIQUE | 54 |
| 1.2.1 LES DIFFERENTES TECHNIQUES D'IONISATION | 55 |
| 1.2.2 RESOLUTION, EXACTITUDE, PRECISION DE MASSE ET HAUT DEBIT | 56 |
| 1.2.2.1 Résolution | 56 |
| 1.2.2.2 Précision et exactitude | 57 |
| 1.2.3 DES MODES DE FRAGMENTATION VARIES | 58 |
| 1.2.3.1 CID | 60 |
| 1.2.3.2 HCD | 60 |
| 1.2.3.3 ETD –Fragmentation par transfert d'électrons | 61 |
| 1.2.3.4 CID, HCD, ETD des modes de fragmentation complémentaires pouvant être utilisés en synergie. | 62 |
| 1.2.4 DES SPECTROMETRES HYBRIDES PERMETTENT DES APPROCHES DIFFERENTES. | 63 |
| 1.2.4.1 Mode de fonctionnement courant | 63 |
| 1.2.4.2 Balayage de l'ion précurseur (« Precursor ion Scanning ») | 63 |

| | |
|--|-----------|
| 1.2.4.3 Balayage par perte de neutre (« Neutral loss scan ») | 64 |
| 1.2.4.4 Détection d'ions multiples (essai de type SRM/MRM pour « Selected Reaction Monitoring » et « MRM pour Multi Reaction Monitoring ») | 65 |
| 1.2.4.5 Détection d'ions multiples en parallèle (PRM pour « Parallel Reaction Monitoring ») | 66 |
| 1.2.4.6 « SWATH-MS » méthode fonctionnant lors de l'acquisition des données en mode « data-independent ». | 66 |
| 1.3 LA PREPARATION DE L'ECHANTILLON POUR ANALYSE DES MPTS | 68 |
| 1.3.1 LA COMPLEXITE DU PROTEOME. | 68 |
| 1.3.2 DES PRECAUTIONS ELEMENTAIRES : « FIGER L'ECHANTILLON ». | 69 |
| 1.3.3 LE FRACTIONNEMENT DU PROTEOME | 70 |
| 1.3.3.1 Séparation sur gels mono ou bidimensionnels (Gels 1D et 2D) | 70 |
| 1.3.3.2 Séparation de type « Off-Gel ». | 73 |
| 1.3.3.3. Approches chromatographiques | 75 |
| 1.3.3.3.1 Hydrophobicité et charge | 75 |
| 1.3.3.3.2 Couplage des techniques– chromatographie liquide multidimensionnelle. | 75 |
| 1.3.3.3.3 Intérêt pour les MPT. | 76 |
| 1.3.3.4 Stratégies de préparation de l'échantillon pour capturer des MPT peu abondantes | 77 |
| 1.3.3.4.1 Immunocapture par anticorps. | 77 |
| 1.3.3.4.2 Affinité pour un métal immobilisé. | 78 |
| 1.3.3.4.3 Les lectines. | 79 |
| 1.4 LES APPROCHES PROTEOMIQUES, DU TRADITIONNEL GEL 2D AUX ANALYSES SHOTGUN GLOBALES ET LES CONSEQUENCES POUR LA RECHERCHE DES MPT. | 81 |
| 1.4.1 LES APPROCHES ASCENDANTES DITES « BOTTOM-UP ». | 81 |
| 1.4.1.1 Digestion trypsique. | 82 |
| 1.4.1.2 Analyse traditionnelle de protéomique par spectrométrie de masse – gels 2-D : empreinte peptidique. | 84 |
| 1.4.1.3. L'analyse des peptides par spectrométrie de masse en tandem. | 84 |
| 1.4.2 LES APPROCHES TOP-DOWN. | 86 |
| 1.4.3 LES APPROCHES MIDDLE-DOWN | 86 |
| 1.4.4 LES STRATEGIES DE QUANTIFICATION DES PROTEINES ET MPTS | 86 |
| 1.4.4.1 SILAC (stable isotope labeling by amino acids in cell culture) | 87 |
| 1.4.4.2 iTRAQ (Isobaric tags for relative and absolute quantitation) | 87 |
| 1.4.4.3 Label free (spectral count, XIC) | 87 |
| 1.4.5 INTERPRETATION DES SPECTRES ET STATISTIQUES ASSOCIEES AUX ANALYSES ETENDUES ET AUX MPT | 88 |
| 1.4.5.1 Les algorithmes | 88 |
| 1.4.5.2 Paramétrages des moteurs de recherche, contraintes apportées par la recherche des MPT | 89 |
| 1.4.6 INTERPRETATION DE SPECTRES, EVITER LES FAUX-POSITIFS OU LES MPT « IN VITRO » | 89 |
| 1.4.7 BASE DE DONNEES ET MODIFICATIONS POST-TRADUCTIONNELLES | 91 |
| I.5 PRESENTATION DU SUJET DE THESE | 93 |
| I.5.1 CONTEXTE GENERAL DU LABORATOIRE D'ACCUEIL | 93 |
| I.5.2 POSITIONNEMENT DU SUJET DE LA THESE | 93 |
| 1.5.2.1 Un changement de paradigme chez les procaryotes ! | 94 |
| 1.5.2.2 Maîtriser l'outil spectrométrie de masse | 96 |

| | |
|--|------------|
| 2. MATERIELS ET METHODES | 98 |
| 2.1. DESCRIPTIF DE L'ORBITRAP XL | 98 |
| 2.2 OPTIMISATION DE LA METHODE D'ACQUISITION | 99 |
| 2.3 PROTOCOLE DE DIGESTION | 101 |
| 2.3.1 - AMELIORATION D'UNE DIGESTION PAR LA TRYPSINE GRACE AU SURFACTANT « PROTEASE MAX™ » | 101 |
| 2.3.2 – DIGESTION DE PROTEINES THERMORESISTANTES DE THERMOCOCCUS GAMMA TOLERANS | 101 |
| 3. RESULTATS | 105 |
| 3.1 -L'APPROCHE PROTEOMIQUE SHOTGUN GLOBALE, UN OUTIL DE CHOIX POUR L'ETUDE DES MODIFICATIONS POST-TRADUCTIONNELLES | 105 |
| 3.1.1 RE-ANALYSE DES SITES D'IODATION DE LA THYROGLOBULINE AU TRAVERS D'UNE STRATEGIE DE TYPE « SHOTGUN » PORTANT SUR L'ANALYSE D'UN ORGANE ENTIER LA THYROÏDE DE SOURIS | 105 |
| 3.1.2 CONCLUSIONS ET PERSPECTIVES | 116 |
| 3.2 - DECOUVERTE D'UNE MODIFICATION POST-TRADUCTIONNELLE CHEZ DEINOCOCCUS DESERTI EN CONDITION DE STRESS | 117 |
| 3.2.1 CHANGEMENTS MAJEURS INTERVENANT DANS LE PROTEOME SOLUBLE DE DEINOCOCCUS DESERTI DANS LES PREMIERES HEURES SUIVANT UNE FORTE IRRADIATION PAR RAYONS GAMMA | 117 |
| 3.2.2 CONCLUSIONS ET PERSPECTIVES | 132 |
| 3.3 - DEINOCOCCUS DESERTI UNE BACTERIE AUX ACETYLATIONS N-TERMINALES ABONDANTES ET CARACTERISTIQUES | 133 |
| 3.3.1 CHANGEMENTS MAJEURS INTERVENANT DANS LE PROTEOME SOLUBLE DE DEINOCOCCUS DESERTI AUX PREMIERES HEURES SUIVANT UNE FORTE IRRADIATION PAR RAYONS GAMMA | 133 |
| 3.3.2 CONCLUSIONS ET PERSPECTIVES | 145 |
| 3.4 - N- AND O- ACETYLATION LES PIEGES A EVITER ! | 146 |
| 3.4.1 N ET O ACETYLATION DES RESIDUS THREONINE DANS LE CONTEXTE DE LA PROTEOMIQUE | 146 |
| SUPPLEMENTARY DATA AVAILABLE AT : | 151 |
| 3.4.2 CONCLUSIONS ET PERSPECTIVES | 151 |
| 3.5 - THERMOCOCCUS GAMMA TOLERANS, UNE ARCHEA AUX LYSINES E-ACETYLEES | 152 |
| 3.5.1 ANALYSE DES LYSINES ACETYLEES DU PROTEOME DE L'ARCHAEA THERMOCOCCUS GAMMATOLERANS | 152 |
| 3.5.2 CONCLUSIONS ET PERSPECTIVES | 169 |
| 3.6 – DEFINIR DES OBJECTIFS EN BIOLOGIE STRUCTURALE A TRAVERS LE PRISME DE LA PROTEOGENOMIQUE, L'EXEMPLE DE LA PROTEINE TGAM_1934 DE L'ARCHEE THERMOCOCCU GAMMATOLERANS | 170 |
| 3.6.1 ANALYSE STRUCTURALE D'UNE PROTEINE PARTICULIERE DE L'ARCHAEA THERMOCOCCUS GAMMATOLERANS (TGAM_1934) DONT LE CHOIX A ETE GUIDE PAR DES TRAVAUX DE PROTEOGENOMIQUE | 170 |
| 3.6.2 CONCLUSIONS ET PERSPECTIVES | 180 |
| 4. DISCUSSION | 183 |
| 4.1 – MAITRISER LA PROTEOMIQUE ET L'OUTIL SPECTROMETRIE DE MASSE POUR ALLER PLUS LOIN.... | 183 |
| 4.2 – ETABLIR LA PRESENCE DE MODIFICATIONS POST-TRADUCTIONNELLES DANS DIVERS ORGANISMES | 184 |
| 4.3 – ROLE DES MODIFICATIONS POST-TRADUCTIONNELLES, UTILISER LE MODELE PROCARYOTE | 185 |
| 4.4 – UN NOUVEAU NIVEAU DE COMPLEXITE LE « DIALOGUE » DES MPT | 187 |

LISTE DES FIGURES

| | |
|---|----|
| <i>Figure 1 : Principaux résidus d'acide aminé phosphorylés.</i> | 29 |
| <i>Figure 2 : Représentation du dieu Janus.</i> | 31 |
| <i>Figure 3 : Différents modes d'acétylation de fonctions amine ou alcool décrits à ce jour (Yang and Hang 2011).</i> | 31 |
| <i>Figure 4 : L'acétylation conduit à une modification de la charge au niveau de la fonction amine.</i> | 33 |
| <i>Figure 5 : Spectre de fragmentation en mode CID du peptide LDLDEIDGELKR portant une acétylation sur la lysine.</i> | 34 |
| <i>Figure 6 : Représentation schématique d'un motif de N-glycosylation.</i> | 35 |
| <i>Figure 7 : Représentation schématique des trois groupes de structures de N-glycanes.</i> | 36 |
| <i>Figure 8 : Représentation schématique d'une O-glycosylation.</i> | 37 |
| <i>Figure 9 : Stratégies possibles pour étudier en spectrométrie de masse des protéines glycosylées.</i> | 38 |
| <i>Figure 10 : Nomenclature de fragmentation des oligosaccharides proposée par Domon et Costello.</i> | 41 |
| <i>Figure 11: Spectre de fragmentation MS/MS en mode positif de l'ion à m/z 779.</i> | 43 |
| <i>Figure 12: Spectre de fragmentation MS/MS en mode positif de l'ion à m/z 1041.</i> | 44 |
| <i>Figure 13: L'extrémité N-terminale d'un résidu Lysine, susceptible d'être acétylée e.</i> | 45 |
| <i>Figure 14: A gauche Protéines SUMO humaine (PDB 1A5R) et Ubiquitine humaine (PDB 1UBQ).</i> | 46 |
| <i>Figure 15 : La cascade d'ubiquitination. D'après (Ravid & Hochstrasser, 2008).</i> | 47 |
| <i>Figure 16 : L'ubiquitine, deux faces de la molécule d'ubiquitine.</i> | 48 |
| <i>Figure 17 : Principaux types d'ubiquitination.</i> | 48 |
| <i>Figure 18 : Sumoylation, la mise en parallèle du cycle de l'ubiquitylation et de la sumoylation souligne les fortes similitudes et l'intervention d'enzymes apparentées (Miteva, Keusekotten et al. 2010).</i> | 50 |
| <i>Figure 19 : Iodation de la thyroglobuline dans les cellules folliculaires de la thyroïde.</i> | 51 |
| <i>Figure 20 : Mécanisme d'iodation de la thyroglobuline.</i> | 52 |
| <i>Figure 21 : Bromation du tryptophane en 6-bromotryptophane.</i> | 53 |
| <i>Figure 22 : La protéomique au point de rencontre de trois groupes de techniques.</i> | 54 |
| <i>Figure 23 : La résolution traduit la capacité à distinguer deux évènements différents s.</i> | 56 |
| <i>Figure 24 : Conséquence directe de différents pouvoirs de résolution d'un spectromètre de masse.</i> | 57 |
| <i>Figure 25 : Nomenclature officielle des fragments pouvant être obtenus au cours d'une analyse par MS/MS.</i> | 59 |
| <i>Figure 26 : Les ions complémentaires obtenus par fragmentations de basse énergie.</i> | 59 |

| | |
|--|-----|
| <i>Figure 27: Fragmentation de type HCD.</i> | 61 |
| <i>Figure 28 : Schéma d'un spectromètre de masse en fonctionnement courant.</i> | 63 |
| <i>Figure 29 : Schéma d'un spectromètre de masse fonctionnant en balayage de l'ion précurseur.</i> | 63 |
| <i>Figure 30 : Schéma d'un spectromètre de masse fonctionnant en balayage par perte de neutre.</i> | 64 |
| <i>Figure 31 : Schéma d'un spectromètre de masse fonctionnant en détection d'ions multiples.</i> | 65 |
| <i>Figure 32 : Schéma d'un spectromètre de masse fonctionnant en détection d'ions multiples en parallèle).</i> | 66 |
| <i>Figure 33 : Inhibiteurs de phosphatases proposés par Sigma Aldrich.</i> | 69 |
| <i>Figure 34 : Gels 1-D.</i> | 70 |
| <i>Figure 35 : Différence de migration entre une caséine phosphorylée et déphosphorylée.</i> | 71 |
| <i>Figure 36 : Gel 2-D du protéome soluble de la bactérie radio résistante Deinococcus deserti.</i> | 72 |
| <i>Figure 37 : Gel d'électrophorèse bidimensionnelle de la protéine recombinante 6His-Chk2 exprimée chez Baculovirus).</i> | 73 |
| <i>Figure 38 : Principe de la séparation « Off Gel » (Document Agilent).</i> | 74 |
| <i>Figure 39 : Représentation schématiques des différentes stratégies mises en œuvre en protéomique.</i> | 81 |
| <i>Figure 40 : Recouvrement de séquence plus complet de la protéine recombinante ChK2.</i> | 83 |
| <i>Figure 41 : Principales enzymes utilisées pour les digestions enzymatiques des protéomes solubles.</i> | 83 |
| <i>Figure 42 : Titres de diverses publications indiquant ce changement de paradigme.</i> | 95 |
| <i>Figure 43 : Vue d'ensemble d'une stratégie optimisée.</i> | 96 |
| <i>Figure 44 : Schéma d'une Orbitrap de type XL (ci-dessus) avec son module ETD (à droite).</i> | 98 |
| <i>Figure 45 : Résultats présentés pour la colonne de 50 cm.</i> | 100 |
| <i>Figure 46 : Résultats présentés pour la colonne de 15 cm.</i> | 100 |
| <i>Figure 47 : Optimisation d'un protocole de digestion pour l'étude du protéome thermorésistant.</i> | 103 |
| <i>Figure 48 : Détection par Western blot de protéines contenant des résidus lysine acétylée.</i> | 159 |
| <i>Figure 49 : Détection par Western blot de protéines contenant des résidus lysine acétylée.</i> | 160 |
| <i>Figure 50 : Diagramme de Venn montrant la complémentarité de la reconnaissance des peptides acétyl-lysine).</i> | 162 |
| <i>Figure 51 : Identification de l'acétylation de peptides par spectrométrie de masse.</i> | 163 |
| <i>Figure 52 : Nombre de sites d'acétylation par protéine.</i> | 164 |
| <i>Figure 53 : L'analyse de la fréquence de la séquence d'acides aminés autour des sites d'acétylation.</i> | 165 |
| <i>Figure 54 : Annotation fonctionnelle des protéines acétylées chez Thermococcus gammatolerans.</i> | 166 |

Figure 55 : Illustration schématique de la présence des protéines acétylées sur les principales voies du métabolisme central chez Thermococcus gammatolerans. 167

Figure 56 : Recouvrement des protéines modifiés post-traductionnellement)...... 169

Figure 57 : Classification des dialogues des MPT (d'après A.S. Venne et al, 2014)...... 188

LISTE DES TABLEAUX

| | |
|---|-----|
| <i>Tableau 1 : Différentes approches pour l'étude des phosphorylations des protéines selon M. Mann et al</i> | 26 |
| <i>Tableau 2 : Données statistiques extraites de « Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database».....</i> | 27 |
| <i>Tableau 3 : Principaux glucides présents chez les glycanes.</i> | 35 |
| <i>Tableau 4 : Ions diagnostics observés en mode négatif sur les spectres de fragmentation pour la détermination de la séquence des O-glycanes de type mucine</i> | 42 |
| <i>Tableau 5 : Incréments de masse, caractéristiques des différents degrés d'iodation de la tyrosine et des divers résidus produits au cours de ce processus.</i> | 52 |
| <i>Tableau 6 : Spectromètres de masse et résolution.</i> | 57 |
| <i>Tableau 7 : Comparaison des différentes approches MS pour l'analyse et la quantification des protéines).</i> | 67 |
| <i>Tableau 8 : Paramètres d'optimisation pour les méthodes d'acquisition testées.....</i> | 99 |
| <i>Tableau 9 : Protéines phosphorylées chez Thermococcus gammatolerans et peptides phosphorylés associés.</i> | 168 |

LISTE DES PRINCIPALES ABREVIATIONS

| | |
|--|--|
| amu atomic mass unit | MRM-MS Multiple Reaction Monitoring-Mass Spectrometry |
| BLAST Basic Local Alignment Search Tool | MS Mass Spectrometry Spectrométrie de masse |
| CEA Commissariat à l'Energie Atomique et aux énergies alternatives | MS/MS Tandem Mass Spectrometry |
| CID Collision Induced Dissociation | MudPIT : Multidimensional Protein Identification Technology |
| Da Dalton | NCBI : National Center for Biotechnology Information |
| DSV Direction des Sciences du Vivant | p-value probability value |
| DTT Dithiothréitol | PBS Phosphate Buffered Saline |
| E Value Expectation value | pI point Isoélectrique |
| ESI ElectroSpray Ionization | PM Poids Moléculaire |
| ETD Electron Transfer Dissociation | Q Quadripôle |
| FDR False Discovery Rate | RPLC Reversed-Phase Liquid Chromatography |
| FTMS Fourier Transform Mass Spectrometry – Spectrométrie de masse à transformée de Fourier | SCX Strong Cation eXchange chromatography |
| Gel 1D ou Gel 2D à 1 ou 2 dimensions | SDS Sodium DodecylSulfate |
| HCD Higher-energy Collisional Dissociation | SDS-PAGE Sodium DodecylSulfate PolyAcrylamide Gel Electrophoresis |
| HPLC High Performance Liquid Chromatography | TFA TriFluoro acetic Acid |
| HRMS High Resolution Mass Spectrometry | TOF Time of Flight |
| IAA IodoAcétAmide | TRIS Tris(hydroxyméthyl)aminométhane ou Trizma® base |
| IPG strips Immobilized pH Gradient strips | v/v Volume/Volume |
| LBSP Laboratoire de Biochimie des Systèmes Perturbés | |
| LC Liquid Chromatography | |
| LC-MS/MS Liquid Chromatography coupled to tandem Mass Spectrometry | |
| LTQ Linear ion Trap Quadrupole <i>m/z</i> rapport masse (<i>m</i>) sur charge (<i>z</i>) | |
| MALDI Matrix-Assisted Laser Desorption/Ionisation | |
| MALDI-TOF/MS Spectrométrie de masse avec ionisation MALDI couplée à un analyseur TOF | |
| MPT Modifications post-traductionnelles | |

**INTRODUCTION
BIBLIOGRAPHIQUE**

1. INTRODUCTION BIBLIOGRAPHIQUE

1.1 Les modifications post-traductionnelles des protéines

1.1.1 Au-delà de la synthèse et de la maturation des protéines, les Modifications Post-Traductionnelles

La synthèse des protéines s'effectue dans les cellules au niveau des ribosomes où le concert des ARN messagers, des ARN de transfert amenant les acides aminés spécifiques est réglé pour donner naissance au long « ruban » d'acides aminés « attachés » par les liaisons peptidiques. Cet enchaînement d'acides aminés constitue la structure primaire de la protéine. Les structures secondaires (feuillettes, hélices) constituent des repliements locaux de cette chaîne tandis que la structure tertiaire correspond au repliement de la chaîne peptidique dans l'espace. Enfin la structure quaternaire sera l'assemblage d'au moins deux chaînes polypeptidiques pour former une nouvelle entité fonctionnelle.

Si certaines protéines sortent du ribosome prêtes à fonctionner, d'autres vont au cours de la synthèse ou à la fin de celle-ci, subir des modifications. Ces modifications seront dites co-traductionnelles ou post-traductionnelles.

Acquisition d'une structure tertiaire : spontanément, certaines protéines peuvent acquérir une structure tridimensionnelle fonctionnelle, comme la ribonucléase pancréatique capable sous certaines conditions de passer d'une forme dénaturée à une structure tertiaire avec création de ponts disulfures. D'autres vont avoir besoin de l'apport de protéines particulières les chaperonnes (Hartl 1996) (Martin and Hartl 1997) pour obtenir leur conformation définitive ou leur permettre des passages transmembranaires. Ces étapes sont capitales dans la maturation des protéines. Si elles ne peuvent être réalisées au sein de la cellule, l'accumulation de protéines mal repliées entraîne l'apparition par exemple de maladies neurologiques gravissimes telles que les maladies de Creutzfeld-Jacob, Alzheimer, d'Huntington ou l'encéphalopathie spongiforme bovine (ESB dite maladie de la vache folle).

Les protéines destinées à l'exportation (voies sécrétoires) vont être synthétisées au sein des ribosomes liés à la membrane du réticulum endoplasmique rugueux avec à leur extrémité N-terminale un peptide signal (ou séquence signal). Ce peptide n'est pas unique pour toutes les protéines (Matlack, Mothes et al. 1998), il possède des caractères structuraux particuliers à savoir une charge positive en N-terminal, un corps hydrophobe constitué de 8 à 12 acides aminés hydrophobes et une partie C-terminale plus polaire. C'est ce dernier site qui servira de site de coupure, lui permettant d'être reconnu par un complexe de protéines (6) et d'ARN (de 300 nucléotides) appelé, particule de reconnaissance du signal (« SRP » pour Signal

Recognition Particle). Ce processus permettra la relocalisation de la protéine fraîchement synthétisée. Une « signal peptidase » détachera ce peptide signal du reste de la protéine au niveau du site de coupure C-terminal précédemment évoqué.

Ces deux exemples montrent qu'au cours de la synthèse dans les ribosomes de la chaîne peptidique, la maturation des protéines n'est pas terminée. Ces ultimes étapes ne font pas partie à proprement parlé des modifications post-traductionnelles, mais sont bien des modifications capitales à la maturation des protéines.

1.1.2 Les modifications post-traductionnelles (MPT) définition :

Ces modifications chimiques sont de nature covalente et peuvent se traduire par de simples coupures protéolytiques ou par l'addition d'un groupe chimique sur un ou plusieurs résidus de la chaîne peptidique. Ces modifications provoquent un changement de masse de la protéine ainsi modifiée. Ces modifications peuvent être totales ou partielles, tout ou partie du site étant affecté par la modification.

Ces modifications ne sont pas des transformations accessoires (Mann and Jensen 2003). Elles vont fortement contribuer à augmenter la diversité fonctionnelle des protéines qui les portent. Elles vont se traduire sur la protéine par une modification physico-chimique, localisée ou étendue, affectant l'hydrophobicité, la conformation, le poids moléculaire ou la stabilité de la protéine. Souvent elles vont entraîner une activation ou un changement d'activité de la protéine ou bien sa localisation dans ou hors de la cellule qui l'a produite (Pawson and Scott 2005; Levine, Vannier et al. 2006). Certaines modifications post-traductionnelles vont jouer des rôles majeurs dans la régulation fine d'activités cellulaires essentielles (phosphorylation, acétylation) (Grangeasse, Cozzone et al. 2007) (Thao and Escalante-Semerena 2011).

Certaines d'entre elles, telles les phosphorylations, ont été détectées depuis longtemps sur quelques protéines (Burnett and Kennedy 1954). Leur étude systématique a été freinée par l'absence de techniques adaptées. En 2002 Matthias Mann et al (Mann, Ong et al. 2002) comparaient les techniques alors disponibles pour l'étude du phosphoprotéome tel que cela est indiqué dans le tableau 1.

La spectrométrie de masse y apparaissait comme la technique la mieux adaptée, ceci par le simple fait que la modification chimique à la base de la modification post-traductionnelle, entraîne un changement de masse du peptide considéré, aisément mesurable par spectrométrie de masse. Au cours des années suivantes, ceci fut pleinement confirmé tant les approches de protéomique par spectrométrie de masse ont gagné en diversité, sensibilité et exactitude au-delà de l'étude du phosphoprotéome (Witze, Old et al. 2007).

Ainsi ce domaine d'étude s'est fortement développé. Une recherche par le moteur de recherche « Pubmed » avec comme seul mot clef « post-translational modification » permet de retrouver 47445 articles traitant du sujet, dont 32321 depuis janvier 2000 (à la date du 20/05/2014).

Tableau 1 : Différentes approches pour l'étude des phosphorylations des protéines selon M. Mann et al (Mann, M et al 2002)

Différentes techniques, marquage au phosphore 32, séquençage d'Edman, spectrométrie de masse peuvent être mise en œuvre pour étudier les phosphorylations présentes sur les protéines. La combinaison des principaux critères choisis pour les comparer donne un net avantage à la spectrométrie de masse. Depuis plus de dix ans les faits expérimentaux ont largement confirmé cette tendance.

| | Marquage au Phosphore-32 | Séquençage d'Edman | Spectrométrie de masse |
|--|---|---|--|
| Utilisation de la radioactivité | Grande quantité nécessaire | Peut être utilisée | Pas nécessaire |
| Sensibilité | La plus sensible | La moins sensible (pmol) | Très sensible (fmol) |
| Localisation des sites | Difficile sans expérience de mutagenèse | Possible, les résidus tyrosine posent problème | Plus aisée |
| Couverture des protéines | Couverture totale difficile | Possible mais nécessite une grande quantité de matériel | Couverture totale difficile, plus abordable avec la FTMS |
| Possibilité de travailler à haut débit | Non | Difficile | Oui |
| Nécessité d'utiliser des protéines purifiées | Oui | Oui | Non |

Certaines de ces modifications sont très répandues et présentes donc sur de nombreuses protéines (phosphorylations, acétylations, méthylation, glycosylations...). D'autres ont une présence plus restreinte, telles que :

- l'hydroxylation (ajout d'un groupement hydroxyle sur la proline ou la lysine. L'exemple le plus connu est la présence en grande quantité d'hydroxy proline (14%) dans les fibres de collagène (Myllyharju 2005).
- La biotinylation qui consiste à fixer la biotine (ou vitamine H ou B8), groupement volumineux de façon covalente sur des résidus lysine par l'enzyme biotine protéine ligase. Ce coenzyme est ainsi rajouté à quelques enzymes de transfert du groupement CO₂ (Ingaramo and Beckett 2011) (Kuroishi, Rios-Avila et al. 2011).
- La sulfonation peut apparaître comme une modification encore moins répandue car présente à ce jour sur deux hormones peptidiques gastro-intestinales, la CCK (Cholécystokinine) et la gastrine (Medzihradzsky, Darula et al. 2004).
- L'iodination des groupements tyrosine, largement décrite pour la thyroglobuline (Dunn and Dunn 1999) (Dedieu, Gaillard et al. 2011), mais pouvant exister sur d'autres protéines, constituant un iodo-protéome encore à explorer.

Ceci ne peut qu'illustrer un monde complexe puisque plus de 300 modifications post-traductionnelles ont été décrites (<http://dbptm.mbc.nctu.edu.tw/statistics.php>).

1.1.3 Liste des principales MPT et de leurs caractéristiques

Il n'est pas ici possible de donner une vision complète de l'ensemble des modifications post-traductionnelles décrites à ce jour. Nous nous contenterons d'évoquer ici les principales modifications étudiées. Khoury, G.A. et al ont créé une ressource bio-informatique (<http://selene.princeton.edu/PTMCuration>) tenue à jour et qui à partir de la base Swiss-Prot indique les modifications les plus étudiées (Khoury, Baliban et al. 2011) (Voir figure 2).

Tableau 2 : Données statistiques extraites de « Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database» (au 15/04/2014) et présentant d'un point de vue expérimental et non expérimental, les principales MPT.

On entend par « Expérimental » les modifications réellement détectées, « Non-expérimental » des modifications qui dans les annotations de la base de données Swiss-Prot sont qualifiées de « potentielle », « probable » ou par « similarité ».

| Experimental | | Non expérimental | |
|--------------|------------------------|------------------|------------------------|
| Fréquence | Modifications | Fréquence | Modifications |
| 40241 | Phosphorylation | 98410 | N-linked glycosylation |
| 7840 | Acétylation | 67021 | Phosphorylation |
| 5947 | N-linked glycosylation | 21681 | Acétylation |
| 3178 | Amidation | 8420 | Méthylation |
| 1752 | Méthylation | 5726 | Palmitoylation |
| 1695 | Hydroxylation | 4498 | Citrullination |
| 1333 | O-linked glycosylation | 3227 | Sulfatation |
| 1122 | Ubiquitylation | 2632 | O-linked glycosylation |

C'est donc en se basant sur ce tableau que nous parlerons de modifications principales en ce sens qu'elles sont à ce jour très étudiées ou qu'elles existent potentiellement à la lumière des connaissances stockées à jour dans les bases de données.

De 2003 à 2012 le nombre de modifications post-traductionnelles déposés dans la base de données UniProt/SwissProt a cru de façon exponentielle. En janvier 2013, 72320 modifications post-traductionnelles ont été reportées dans cette base de données. Elles se rapportaient à 22538 protéines. Ce chiffre bien que conséquent est cependant à comparer aux 5 millions de séquences présentes dans cette base de données (on parle là de séquences prédites, trouvées par homologie, au niveau transcriptomique ou au niveau protéomique). Or de façon surprenante entre 2010 à 2013 le nombre de protéines modifiées post-traductionnellement et déposées dans cette base de données s'est effondré !

Ceci traduit peut être un effet de mode ou des difficultés dans l'étude des MPT car au-delà de leur seule description il convient maintenant de comprendre leur rôle. Remarquons enfin

qu'au cours de ces trois dernières années un tiers des travaux publiés l'ont été avec des approches « shotgun », et les deux tiers restant par des approches ciblées certainement plus efficaces.

1.1.3.1 Au commencement était la méthionine....

La méthionine est le résidu invariant du début de toutes les chaînes polypeptidiques. Une des premières transformations subies par certaines protéines sera l'enlèvement de ce résidu méthionine au moyen de la Met-aminopeptidase (MAP), enzyme liée au ribosome. Cette coupure n'est pas systématique et se fera de préférence lorsque l'acide aminé de la chaîne polypeptidique qui suit la méthionine est de petite taille (alanine, cystéine, glycine, proline, sérine, thréonine ou valine (Hirel, Schmitter et al. 1989; Meinnel and Giglione 2008).

Cette coupure s'effectue dans plus de la moitié des cas tant chez les eucaryotes que chez les procaryotes.

Chez les procaryotes notons que la méthionine initiale est formylée, nécessitant l'action d'une peptide deformylase pour éliminer le groupement formyle avant l'action de MAP.

Il semble étonnant qu'aussitôt synthétisé un résidu soit presque immédiatement retiré. Plusieurs hypothèses ont été formulées (Bonissone, Gupta et al. 2013), comme, l'introduction d'une diversité à l'extrémité N-terminale, après son enlèvement pour « préparer » la protéine à d'autres modifications post-traductionnelles (acétylations). Certains résidus seraient plus particulièrement concernés alanine, sérine, thréonine. Une autre hypothèse consiste à évoquer le recyclage rapide de cet acide aminé, constamment utilisé pour l'initiation du ruban peptidique et ailleurs évidemment dans la chaîne peptidique. Enfin dernière hypothèse, le lien qui pourrait exister entre l'excision de la méthionine et les processus de stabilisation ou de dégradation des protéines. Ce processus est basé sur la reconnaissance du résidu N-terminal.

Savoir si l'enlèvement de la méthionine est une modification co-traductionnelle (s'effectuant au moment de la biosynthèse des protéines au niveau du ribosome) ou une modification post-traductionnelle (s'effectuant une fois cette biosynthèse achevée) peut aussi être discuté. Notons avant tout que toutes ces transformations (comme le clivage du peptide signal) indiquent bien que la biosynthèse des protéines ne se résume pas à la formation de la seule chaîne polypeptidique, mais que les processus de maturation et de modification chimique des protéines qui s'effectuent pendant et après cette biosynthèse illustrent un remodelage constant subi par les protéines au cours de leur vie.

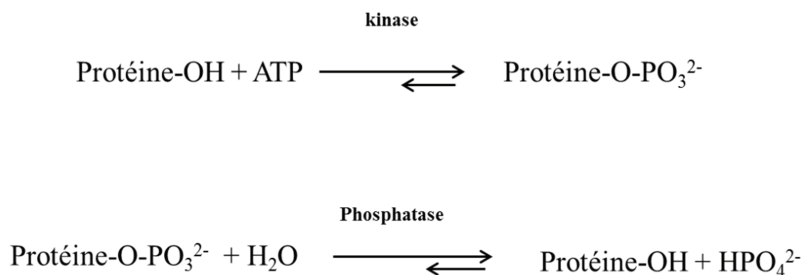
1.1.3.2 Les phosphorylations

Il s'agit de la modification post-traductionnelle la plus étudiée. Le processus de phosphorylation connu depuis plus d'un siècle (Levene and Beatty 1906) a été suivi de la mise en évidence d'une phosphosérine dans la vitelline (Lipmann & Levene, 1932) puis de la première description de la phosphorylation d'une protéine (Burnett and Kennedy 1954). L'étude plus approfondie du mécanisme de phosphorylation réversible des protéines dans les années 50, valut à Edmond H. Fischer, biologiste suisse naturalisé américain ainsi qu'à son collègue, Edwin G. Krebs, le prix Nobel de médecine en 1992. Ce prix leur fut attribué pour

« leurs découvertes concernant les phosphorylations réversibles des protéines comme un mécanisme de régulation biologique ».

1.1.3.2.1 Une modification réversible

La phosphorylation consiste en l'addition d'un groupe phosphoryle PO_3^{2-} sur certaines chaînes latérales d'acides aminés de la chaîne peptidique. C'est une MPT réversible, la phosphorylation des protéines étant catalysée par des protéines kinases, alors que la déphosphorylation se fera avec l'aide de phosphatases.



1.1.3.2.2 Une estérification des résidus alcool (plus rarement amine) des chaînes latérales de quelques acides aminés

Les résidus alcool des chaînes latérales de la sérine, thréonine ou tyrosine sont estérifiées dans des proportions bien différentes, les phosphorylations étant plus fréquentes sur l'alcool primaire de la sérine (1800) puis sur l'alcool secondaire de la thréonine (200), les phosphorylations les plus rares étant celles de la tyrosine (1)

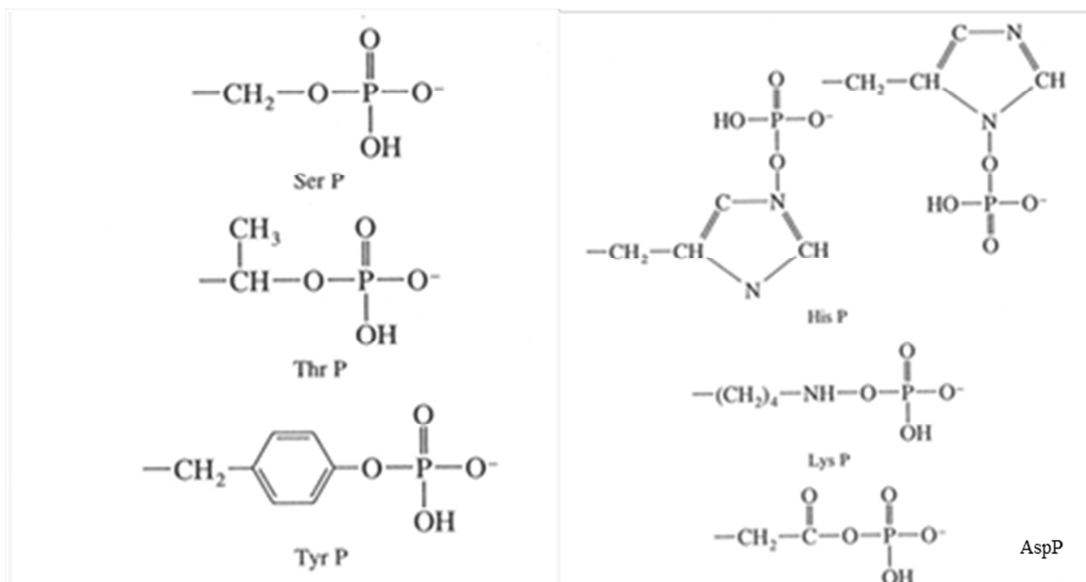


Figure 1 : Principaux résidus d'acide aminé phosphorylés.

Colonne de gauche, les trois résidus hydroxylés sont les plus fréquemment phosphorylés.

Colonne de droite, dans des cas plus rares la fonction amine des lysines ou de l'histidine ainsi que la fonction carboxyle de l'acide aspartique peuvent être phosphorylés.

D'autres résidus peuvent être phosphorylés au niveau des histidines, des lysines, des arginines, des acides aspartiques, ou des cystéines (Matthews 1995; Pannifer, Flint et al. 1998; Wagner and Vu 2000).

1.1.3.2.3 Une modification chimique aux multiples aspects

L'addition de ce groupement phosphoryle présente des caractéristiques intéressantes à plusieurs niveaux:

- Au niveau structural : L'apport des charges négatives peut modifier les interactions électrostatiques au sein de la protéine, entraînant des changements structuraux locaux et éventuellement une modification de l'activité biologique de la protéine. Par ailleurs la structure du groupement phosphorylé, de forme tétraédrique permet des interactions très dirigées,
- Au niveau énergétique, une molécule d'ATP est consommée et une partie de l'énergie utilisée est conservée au sein de la molécule,
- Au niveau cinétique, ces phénomènes sont rapides (moins d'une seconde) et peuvent perdurer plusieurs heures, avec un effet de propagation et donc d'amplification très grand.

1.1.3.2.4 Une modification chimique jouant un rôle essentiel dans les diverses régulations cellulaires

Les phosphorylations des protéines vont leur permettre de jouer un rôle dans de nombreux mécanismes de régulations des cellules (cycle de division cellulaire, déclenchement des processus d'apoptose, processus de différenciation...) et très souvent une petite fraction de la protéine sera phosphorylée lui permettant d'exercer une activité de régulation mais rendant encore plus difficile et délicate l'étude des phosphorylations. Notons que dans de rares cas la phosphorylation massive d'une protéine va jouer un rôle différent, celui de stockage d'ions phosphate (caséine dans le lait, phosvitine du jaune d'œuf).

1.1.3.2.5 Une modification post traductionnelles très répandue

Le phénomène de phosphorylation joue un rôle clef dans les cellules eucaryotes. On estime au tiers le nombre des protéines de ces cellules susceptibles d'être phosphorylées. Il a été estimé que 2% du génome humain code pour des kinases ou des phosphatases (Cohen 2000; Delom and Chevet 2006; Morandell, Stasyk et al. 2006). Chez l'homme plus de 500 kinases et une centaine de phosphatases ont à ce jour été identifiées. Notons que les kinases peuvent être très spécifiques ou multifonctionnelles (Pidoux and Tasken 2010).

1.1.3.3 Les acétylations, une MPT type Janus



Figure 2 : Représentation du dieu Janus.

Buste romain de Janus, Musée du Vatican. (Janus, un dieu antique à double face !)

1.1.3.3.1 Différentes acétylations

C'est une modification chimique covalente au moyen d'un groupement acétyle. On doit distinguer ici les acétylations des groupements ϵ -amine des chaînes latérales de la lysine, et les acétylations N-terminales des protéines. Tout récemment, des acétylations des groupements hydroxyles des serines et thréonines ont été mises en évidence. La figure 5 ci-dessous résume les trois principales possibilités d'acétylation décrites à ce jour.

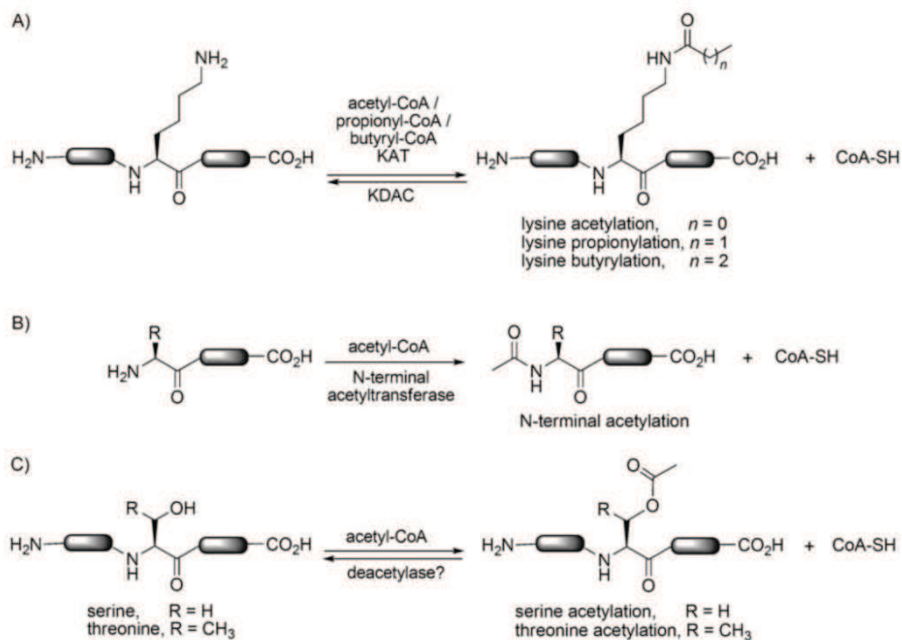


Figure 3 : Différents modes d'acétylation de fonctions amine ou alcool décrits à ce jour (Yang and Hang 2011).

A) Acétylation des groupements ϵ -amine des chaînes latérales de la lysine, cette première modification décrite voilà plus de cinquante ans (Phillips 1963) a d'abord

été observée sur les histones. Elle a été décrite comme étant réversible. Ainsi des lysine acétyl-transferases (KATs autrefois nommées HATs pour histone acétyl transferases) et des déacétylases (plus connues sous le nom d' Histone déacétylases) ont été décrites et contribuent à l'acétylation réversible de la fonction ε-amine des lysines. Un des premiers rôles reconnu à ces acétylations a été la régulation de l'expression des gènes (Phillips 1963; Allfrey, Faulkner et al. 1964). Des études récentes ont démontré que ces acétylations jouent un rôle central dans la régulation de l'ensemble du métabolisme cellulaire (Choudhary, Kumar et al. 2009; Kim, Yu et al. 2013).

- B) Acétylation N-terminale des protéines, cette deuxième modification est considérée comme irréversible. Cette modification fait partie du processus de maturation des protéines et intervient massivement après le départ de la méthionine dans les cellules eucaryotes ou de la formyl-méthionine dans les cellules procaryotes. Ce processus est co-translationnel chez les eucaryotes. Il est activé par des N-acétyltransférases dont trois familles ont été identifiées chez les eucaryotes NatA, NatB, NatC (Polevoda and Sherman 2002). Quant à la fonction de cette modification, elle n'est pas encore clairement établie (Arnesen T., 2011).
- C) Acétylation des groupements hydroxyle des sérines et thréonines. Cette dernière modification intervient dans le contexte biologique particulier de l'infection bactérienne (Mukherjee, Keitany et al. 2006) au cours de laquelle un effecteur bactérien (YopJ) semble acétyler les résidus sérine et thréonine des protéines de l'hôte, empêchant leur phosphorylation et perturbant la réponse inflammatoire de l'hôte.

1.1.3.3.2 Conséquences physico-chimiques de cette modification

- Dans tous les cas, elle s'accompagne d'un incrément de masse de 42,0105 Da (Zhang, Yau et al. 2004) qui ne peut plus être confondu avec une tri-méthylation (incrément 42,0468 Da) avec les appareils de masse à haute résolution disponibles aujourd'hui.
- Dans le cas des acétylations du groupement lysine, une modification de la charge du peptide portant la modification peut être constatée.

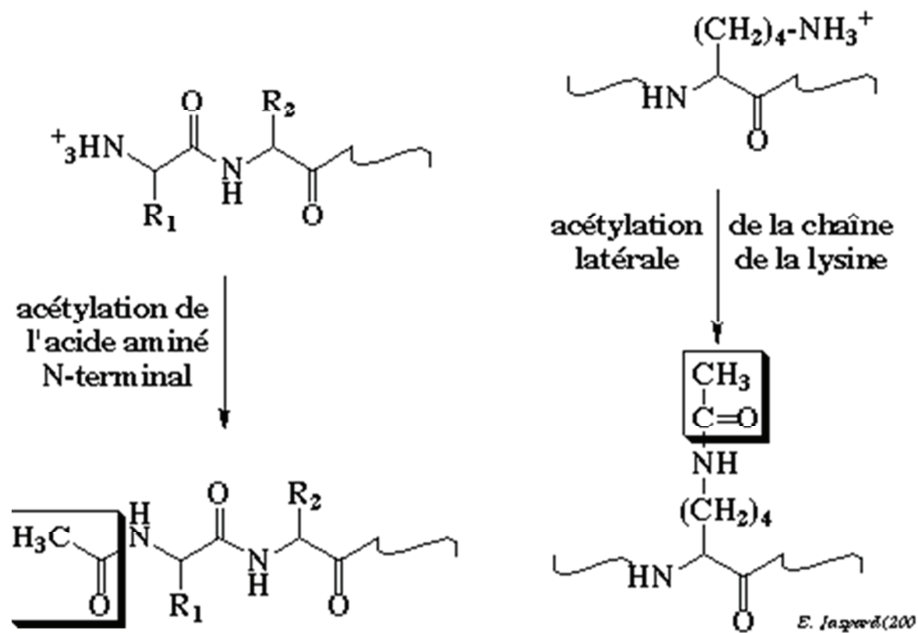


Figure 4 : L'acétylation conduit à une modification de la charge au niveau de la fonction amine.

Cette modification peut être mise à profit pour enrichir, par chromatographie d'échange ionique avec un échangeur cationique fort, des fractions en peptides acétylés, comme celle utilisée pour le pré-enrichissement des peptides phosphorylés (Mohammed and Heck 2011). Après digestion trypsique les peptides N-terminaux vont perdre une charge et se retrouver présents dans les premières fractions de la chromatographie.

Dans une analyse faite par spectrométrie de masse en cas de mono acétylation le parent verra sa masse augmenter comme indiqué ci-dessus. Lors de la fragmentation du peptide (MS/MS), il n'y a pas de rupture particulière et l'on observera bien le résidu Lysine augmenter de la masse portée par l'acétylation.

Une fragmentation de type CID d'un résidu Lysine acétylé, produira un ion immonium à $m/z = 126,1$. Notons d'ailleurs qu'une lysine tri-méthylée entraînera une perte de neutre du groupement tri-méthyl amine [$\text{N}(\text{CH}_3)_3$]. En fragmentation HCD l'ion immonium formé serait visible, établissant une preuve supplémentaire de l'acétylation de la lysine. La figure 5 présente un spectre MS/MS relativement complet obtenu par fragmentation CID.

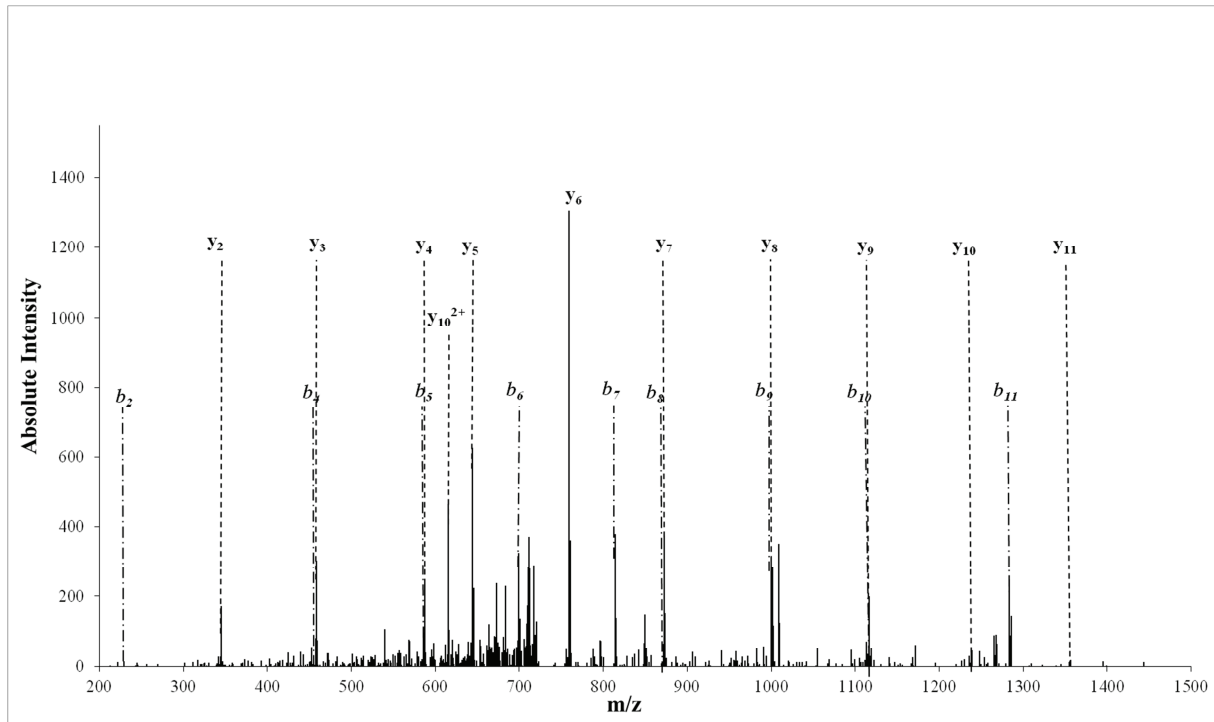


Figure 5 : Spectre de fragmentation en mode CID du peptide LDLDEIDGELKR portant une acétylation sur la lysine.

Cette modification contrairement à la phosphorylation n'est pas labile au cours de la fragmentation. On peut obtenir des séries complètes d'ions y et b permettant de reconnaître parfaitement l'acétylation portée par la lysine.

1.1.3.4 Les glycosylations ou le monde complexe des sucres ¹









La glycosylation est une autre MPT majeure. Par l'intermédiaire d'une réaction enzymatique, un glycane va être lié de façon covalente à une chaîne peptidique (notons que cette fixation de glycanes pourra se faire sur d'autres molécules telles les lipides). Ici encore ce sont les chaînes latérales des acides aminés qui vont intervenir au travers de leur fonction amine ou alcool.

Ces glycanes sont composés d'unités de monosaccharides (glucides) formant une structure polysaccharide plus ou moins complexe. Les principaux monosaccharides présents chez les glycanes sont décrits dans le tableau ci-dessous. La glycosylation se retrouve autant chez les eucaryotes que chez les procaryotes (Upreti, Kumar et al. 2003).

¹ *Ce paragraphe a été inspiré par le travail présenté par Catherine Masselot (Université de Lille), dans le document de travail en préparation portant sur les MPT et commandité par la SFSM (non publié). Merci pour son autorisation de reproduire quelques schémas et tables.*

Tableau 3 : Principaux glucides présents chez les glycanes.

(D'après Burlingame A.L. and Carr S.A. Mass Spectrometry in the Biological Science – Humana Press – 1996 p 542-545)

| Glucides | Abréviations | Symboles | Masse isotopique des résidus (Da) |
|----------------------------|--------------|---|-----------------------------------|
| Xylose | Xyl |  | 132,04 |
| Fucose | Fuc |  | 146,06 |
| Galactose | Gal |  | 162,05 |
| Glucose | Glc |  | 162,05 |
| Mannose | Man |  | 162,05 |
| N-acétylglucosamine | GlcNAc |  | 203,08 |
| N-acétylgalactosamine | GalNAc |  | 203,08 |
| Acide-N-acétylneuraminique | NeuNAc |  | 291,10 |

1.1.3.4.1 Deux grands types de glycosylation

On les distingue selon le type de liaison entre le glycanes et l'acide aminé d'ancrage.

La N-glycosylation se fait sur les résidus asparagine suivis d'une sérine ou d'une thréonine deux résidus plus loin (séquence consensus N-X-S/T). La N-glycosylation est une modification majeure des protéines dans les cellules eucaryotes (Spiro Robert G et al ; 2002). Elle implique la fonction hémi-acétalique d'un résidu de N-acétylglucosamine (GlcNAc) et la fonction amide d'un résidu d'asparagine (voir figure 6).

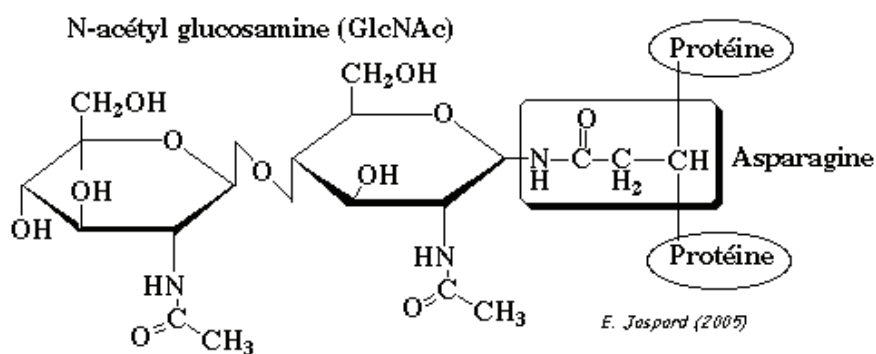


Figure 6 : Représentation schématique d'un motif de N-glycosylation.

Le résidu N-acétyl glucosamine est fixé à la protéine par la fonction amide d'un résidu asparagine

Ces modifications, qui s'effectuent au niveau du réticulum endoplasmique, forment des arborescences particulièrement complexes (voir figure 7)

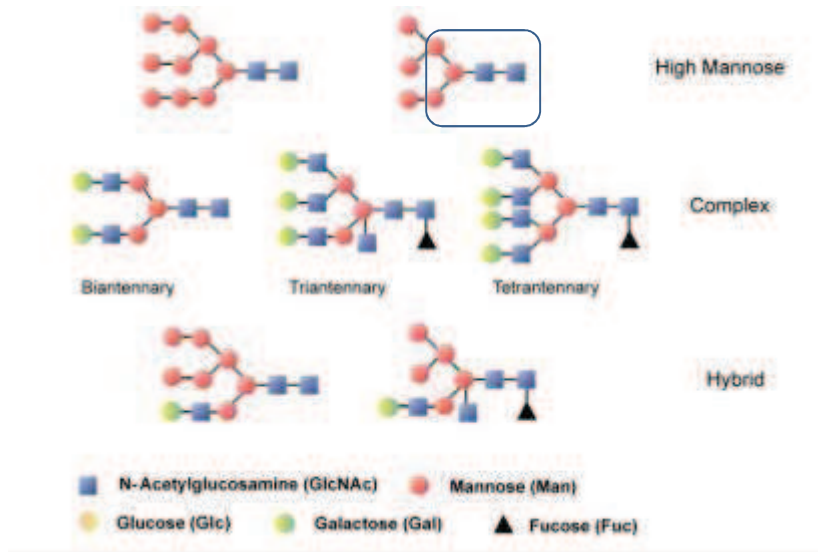


Figure 7 : Représentation schématique des trois groupes de structures de N-glycanes (d'après Dalpathado DS et Desaire H, Analyst, 2008, 133, 731–738 | 731).

A partir d'une structure de base conservée (voir encadré bleuté) composée de trois résidus de mannose embranchés sur le résidu N-acétylglucosamine trois types de structures peuvent être décrites.

Le terme O-glycosylation regroupe un ensemble de modifications post-traductionnelles qui ont en commun la condensation de la fonction hémi-acétalique d'un monosaccharide avec la fonction alcool primaire d'un acide aminé hydroxylé sérine ou thréonine, apparemment sans séquence consensus). Les chaînes formées seront plus courtes (3 à 4 résidus d'oses), mais plus variables que les chaînes des N-glycosylations. Il existe 7 types de O-glycosylations chez les mammifères : O-N-acétylgalactosamine (O-GalNAc de type mucine) la plus répandue, O-Mannose (O-Man), O-acétylglucosamine (O-GlcNAc), O-Fucose (O-Fuc), O-Xylose (O-Xyl), O-Glucose (O-Glc) et O-Galactose (O-Gal). Elles s'effectuent au niveau de l'appareil de Golgi. Ici les chaînes de sucres seront plus courtes mais plus variables, les sucres sont ajoutés directement sur les acides aminés cibles de façon séquentielle. Les sucres sont ajoutés sous une forme « activée », c'est-à-dire liés à un nucléotide (voir figure 8).

Le processus de glycosylation se produit surtout sur les protéines membranaires ou les protéines sécrétées (anticorps). Il débute pour les N-glycosylations dans le réticulum endoplasmique rugueux pour se terminer dans l'appareil de Golgi, tandis que la O-glycosylation se fait exclusivement dans l'appareil de Golgi.

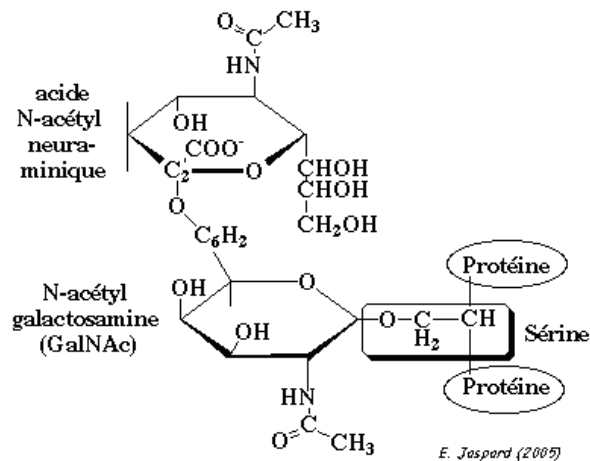


Figure 8 : Représentation schématique d'une O-glycosylation.

Reliée par le groupement hydroxyle d'une sérine ou d'une thréonine, des chaînes plus courtes (3 à 4 résidus) mais plus variées caractérisent les O-glycosylations.

La glycosylation des protéines implique de nombreuses glycosyltransférases. Dans les cellules eucaryotes, plus de 100 différentes ont été identifiées. Elles vont assurer le transfert d'un sucre d'un groupement donneur vers la protéine. La spécificité de ces transférases est souvent liée aux divers sucres transférés parmi les plus courants : Mannose, Galactose, Glucose, Fucose, N-acétyl galactosamine, N-acétylglucosamine ou l'acide N-acétylneuraminique.

1.1.3.4.2 Rôle des glycosylations

La glycosylation rend les protéines concernées plus résistantes à la dégradation enzymatique, elle renforce leur stabilité en rendant difficilement accessibles les sites de coupures enzymatiques.

La glycosylation va affecter l'activité biologique de la protéine concernée en orientant aussi sa localisation. Elle est impliquée dans la reconnaissance des signaux, la liaison de ligands, jouant ainsi un rôle important dans la communication intercellulaire (Dwek 1996), la prolifération et la différenciation cellulaire ainsi que les infections virales. Presque toutes les molécules clés impliquées dans la réponse immunitaire innée et adaptative sont des glycoprotéines (Rudd, Elliott et al. 2001). Les antigènes ABO des groupes sanguins sont des glycanes liés à des protéines.

1.1.3.4.3 Comment caractériser les glycosylations par spectrométrie de masse

Nous abordons ici cette étude de par la spécificité et complexité de cette modification. Alors que toutes les autres modifications post-traductionnelles s'étudient directement à partir des peptides modifiés, ici de par la complexité de la modification portée par ce peptide modifié, on se voit dans l'obligation de détacher la modification (l'oligosaccharide) pour l'étudier, la caractériser.

La glycosylation est une modification particulière des protéines de par sa complexité. En effet contrairement à la plupart des autres modifications post-traductionnelles, ici les modifications sont constituées de motifs variables et complexes.

En glycoprotéomique on cherchera : d'abord à déterminer le site, l'acide aminé sur lequel la chaîne glucidique est branchée, ensuite à déterminer le type de sucres et leur enchaînement dans la chaîne glucidique. Ces études se feront en gardant en tête, la notion de macro et micro hétérogénéité et en sachant que ces molécules complexes sont au cœur d'un processus biologique changeant au cours duquel la structure de ces chaînes latérales est appelée à se modifier. La macro hétérogénéité se rencontre dans le cas d'une N-glycosylation pour laquelle la modification sera obligatoirement portée par une asparagine présente dans une séquence consensus de type (Asn – X –Ser/Thr). Cette asparagine appelée site de glycosylation sera ou non effectivement glycosylée, ceci pouvant changer d'une protéine à l'autre, générant ce que l'on appelle une macro hétérogénéité.

Ensuite au niveau de la glycosylation elle-même, la nature des glycanes, la longueur de la chaîne, sa ramification, seront différentes d'un site à l'autre créant ce que l'on appelle une micro hétérogénéité.

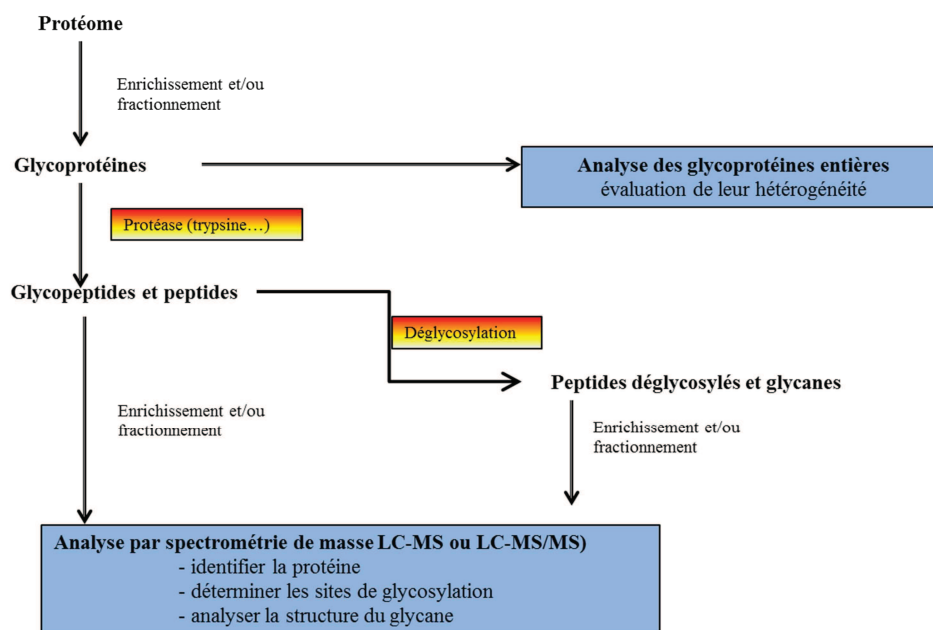


Figure 9 : Stratégies possibles pour étudier en spectrométrie de masse des protéines glycosylées.

Les trois grandes étapes de cette stratégie :

1. Analyse des glycoprotéines entières qui va permettre l'évaluation de leur hétérogénéité.

La chromatographie d'affinité avec des lectines pourra être utilisée. Les différentes lectines, telles la Concavaline A (ConA) ou l'agglutinine de germe de blé (WGA Wheat germ agglutinin), avec leur spécificité différente vont permettre d'isoler des glycoprotéines ayant des glycoformes différentes (Nawarak, Phutrakul et al. 2004). Mais si la protéine possède plus d'un site de glycosylation, l'hétérogénéité des molécules étudiées est telle que l'analyse par spectrométrie de masse (tant Maldi qu'ESI) se révèle limitée pour apporter des informations significatives.

2. Digestion des glycoprotéines.

Le but est ici d'obtenir des glycopeptides possédant un seul site de glycosylation, de déterminer ce site de glycosylation (Dalpathado and Desaire 2008). Si la trypsine reste dans ce cas encore la protéase la plus couramment utilisée, d'autres protéases (GluC, AspN, chymotrypsine) pourront aussi être mises en œuvre.

L'enrichissement en un glycopeptide particulier pourra alors se faire par affinité (lectine) ou par chromatographie d'interaction hydrophile (HILIC), autre approche utilisant l'hydrophilie des peptides glycosylés.

3. Analyse des glycanes

3.1 – Libération des glycanes

On doit au préalable déglycosyler le peptide, soit chimiquement (Morelle, Canis et al. 2006) soit au moyen d'une enzyme. Cependant il n'existe pas d'enzyme à large spectre d'action.

Pour les N-glycanes on utilisera des enzymes de la famille des endo-N-acétyl- β -glucosamidases ou des peptidyl-N-glycosidases (PNGases) et en particulier la PNGase F et la PNGaseA (Geyer and Geyer 2006). Il faut au préalable dénaturer les glycoprotéines (agents réducteurs), dessaler, concentrer et marquer les oligosaccharides.

Pour les O-glycanes, ici il n'y a pas de méthode enzymatique possible. On aura recours à des méthodes chimiques plus délicates à mettre en œuvre. Ces méthodes sont surtout basées sur la β -élimination en milieu réducteur, en présence d'hydroxyde de sodium et de borohydrure de sodium, pour éviter la dégradation des sucres. D'autres méthodes de libération des oligosaccharides en milieu non-réducteur permettent en même temps de marquer les O-glycanes, telle l'hydrazinolyse qui à 60°C libère les O-Glycanes (Patel, Bruce et al. 1993) améliorée par Kozak et al (Kozak, Royle et al. 2012)

3.2– Modification chimique des glycanes

Les N- et O-glycanes peuvent être analysés par spectrométrie de masse (pour en déterminer leur structure) sous forme native quand ils ont été libérés du site de fixation les rattachant à la chaîne peptidique.

Il peut être intéressant de les modifier chimiquement pour : augmenter la sensibilité de détection en spectrométrie de masse, permettre la quantification des oligosaccharides réducteurs, stabiliser les résidus acides (acides sialiques et groupements sulfates) pour l'analyse en spectrométrie de masse.

Dans les deux premiers cas, on pourra utiliser quand elle est disponible une fonction aldéhyde libre qui réagira avec un fluorophore possédant une fonction amine libre ; la base de Schiff formée sera stabilisée par amination réductive.

Dans le dernier cas, pour stabiliser les résidus acides on perméthylera les glycanes c'est-à-dire que l'on remplacera les hydrogènes des groupes hydroxyle, amine et carboxyle libres par des groupements méthyle (Ciucanu and Costello 2003).

3.3– Analyse des glycanes par spectrométrie de masse

La spectrométrie de masse (de type CID-MS/MS) apparaît ici aussi comme l'instrument le plus adapté pour caractériser structurellement les glycanes. L'ionisation se fera en mode positif pour les glycanes neutres, en mode négatif pour les glycanes acides. La fragmentation des oligosaccharides s'effectue selon des chemins préférentiels et la nomenclature proposée par Domon et Castello (Domon and Costello 1988) est représentée dans la figure 10.

Des ions de type B et Y apparaissent majoritairement, permettant de reconstituer la séquence primaire en monosaccharides.

« Pour les O-glycanes, outre la séquence primaire obtenue à partir des ions fragments Y ou B, en mode négatif des ions fragments de type A ou D permettront de déterminer le type de liaisons glycosidiques, ce qui permet par exemple de définir les déterminants de groupes sanguins. De façon générale l'interprétation des spectres de fragmentation des O-glycanes est facilitée par la présence d'ions diagnostics (Robbe, Capon et al. 2004) (Voir Figure 12)

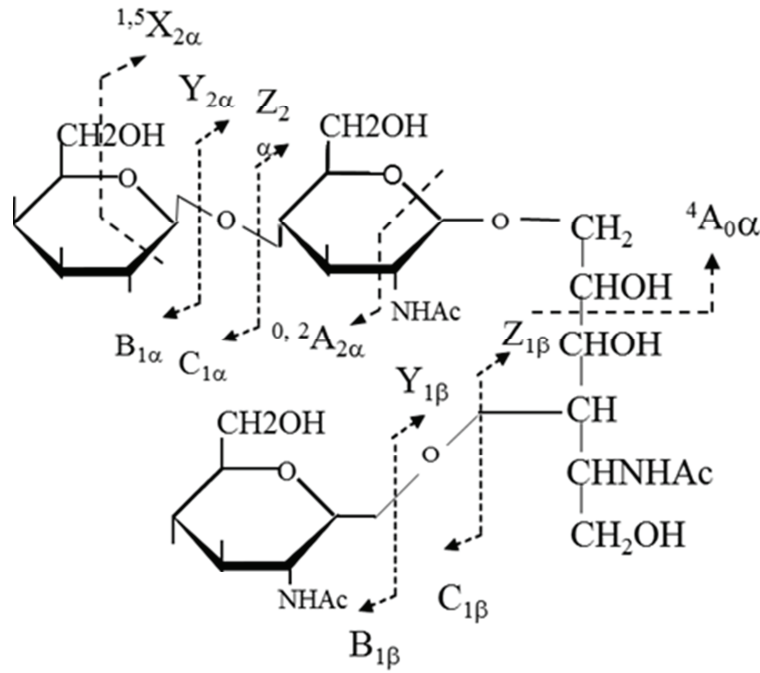


Figure 10 : Nomenclature de fragmentation des oligosaccharides proposée par Domon et Costello.

Le clivage de la liaison interglycosidique donne naissance aux fragments B_j , C_j , Y_j et Z_j tandis que le clivage intracyclique donne les fragments de type A_j et X_j

Les figures 11 et 12 illustrent le séquençage nano-ESI-Q-TOF-MS/MS de O-glycanes perméthylés provenant de mucines intestinales. En se basant sur le schéma de biosynthèse des mucines, les hexoses ont été attribués à des galactoses et les N-acétylhexosamines des chaînes lactosaminiques à des N-acétylglucosamines. Les ions fragments ainsi que la structure précise des glycanes sont indiqués sur ces figures.

Par exemple, l'analyse du spectre de fragmentation de l'ion à m/z 779 (Figure 11) nous permet de déterminer que cet ion correspond à la masse d'un trisaccharide composé de la N-acétylgalactosamine (GalNAc) du point d'attache, d'un résidu de galactose (Gal) et d'un résidu de N-acétylglucosamine (GlcNAc). Deux structures isomériques ont été identifiées en se basant sur les ions fragments. Les ions fragments B_1 et C_1 à m/z 282 et 300 respectivement sont spécifiques d'une GlcNAc en position terminale (Figure 11A) alors que les ions fragments B_1 et C_1 à m/z 241 et 259 respectivement sont spécifiques d'un Gal en position terminale. Les ions Y_2 et Z_2 à m/z 520 et 502 indiquent qu'un Gal est relié à la GalNAc alors que les ions Y_2 et Z_2 à m/z 561 et 543 indiquent que c'est une GlcNAc reliée à la GalNAc.

Tableau 4 : Ions diagnostics observés en mode négatif sur les spectres de fragmentation pour la détermination de la séquence des O-glycanes de type mucine d'après Robbe et al

| Epitope | m/z | Fragmentation spécifique | Information sur la séquence |
|----------------------------------|--|---|---|
| Core 1 | 384, 366 [M-H] ⁻ - 223 [M-H] ⁻ - 108 | | Galβ1-3GalNAcol perte de la GalNAcol perte de C ₃ H ₅ O ₄ |
| Core 3 | 425, 407 [M-H] ⁻ - 223 [M-H] ⁻ - 108 | | GlcNAcβ1-3GalNAcol perte de la GalNAcol perte de C ₃ H ₅ O ₄ |
| Core 2 | 384, 366, 348 425, 407, 389 | | |
| | Masse moléculaire de branche en C6 - H + 60 | ⁴ A _{0α} clivage de GalNAcol | Composition de la branche C-6 |
| Core 4 | 425, 407, 389 628, 610 | | |
| | Masse moléculaire de branche en C6 - H + 60 | ⁴ A _{0α} clivage de GalNAcol | Composition de la branche C-6 |
| Chaîne de type 2 | | | |
| | Masse moléculaire du fragment + 102 | ^{0,2} A _i clivage de GlcNAc | |
| | 263, 281 409 466 | | Galβ1-4GlcNAc Fucα1-2Galβ1-4GlcNAc GalNAcβ1-4Galβ1-4GlcNAc (groupe sanguin Sda/Cad) |
| Chaîne de type 1 | | Clivage prédominant de type Z _i | |
| Oligosaccharides sulfatés | | | |
| | 241 | Ion de type B _i | Gal-SO ₃ |
| | 282 | Ion de type B _i | GlcNAc-SO ₃ |
| | 97 | | HSO ₄ ⁻ |
| | 139 | ^{1,3} A _i , ^{0,4} A _i , ^{2,4} A _i | |
| | 199 | ^{0,2} A _i , ^{2,4} X _i | Résidu sulfate sur C-4 ou C-6 de Gal, GlcNAc ou GalNAc |
| Oligosaccharides sialylés | | | |
| | 290 | Clivage type B _i | NeuAc |
| | - 221 | ^{0,2} X _i ("ion j") | |
| | - 44 | perte de CO ₂ | |
| | 306 | ⁴ A _{0α} - CO ₂ | NeuAc6 liés |
| | 408, 611 | B _i - CO ₂ | NeuAc3 liés |
| | 513 | Y _{1β} | NeuAc6GalNAcol |
| Groupes sanguins | | | |
| H | 205 | Clivage ^{1,3} A ₂ du Gal | |
| | 247 | B ₂ - C ₂ H ₄ O ₂ | Fuc2Gal |
| | 307, 325 | Clivages B ₂ , C ₂ | |
| Lewis a/x | | | |
| | [M-H] ⁻ - 344 | Clivage Z _i /Z _i | Gal(Fuc3/4)GlcNAc ^a |
| | [M-H] ⁻ - 374 | Z _i /Z _i - CH ₂ O | |
| Lewis b/y | | | |
| | [M-H] ⁻ - 490 | Clivage Z _i /Z _i | Fuc2Gal(Fuc3/4)GlcNAc ^a |
| | [M-H] ⁻ - 520 | Z _i /Z _i - CH ₂ O | |
| | | Ions fragments caractéristiques Z_i/Z_i | |
| A Lewis a/x | [M-H] ⁻ - 547, - 577 | | (GalNAcα1-3)Gal(Fuc3/4)GlcNAc ^a |
| B Lewis a/x | [M-H] ⁻ - 506, - 536 | | (Galα1-3)Gal(Fuc3/4)GlcNAc ^a |
| A Lewis b/y | [M-H] ⁻ - 693, - 723 | | Fuc2(GalNAcα1-3)Gal(Fuc3/4)GlcNAc ^a |
| B Lewis b/y | [M-H] ⁻ - 652, - 682 | | Fuc2(Gal α1-3)Gal(Fuc3/4)GlcNAc ^a |

^aPerte de la branche substituant la GlcNAc

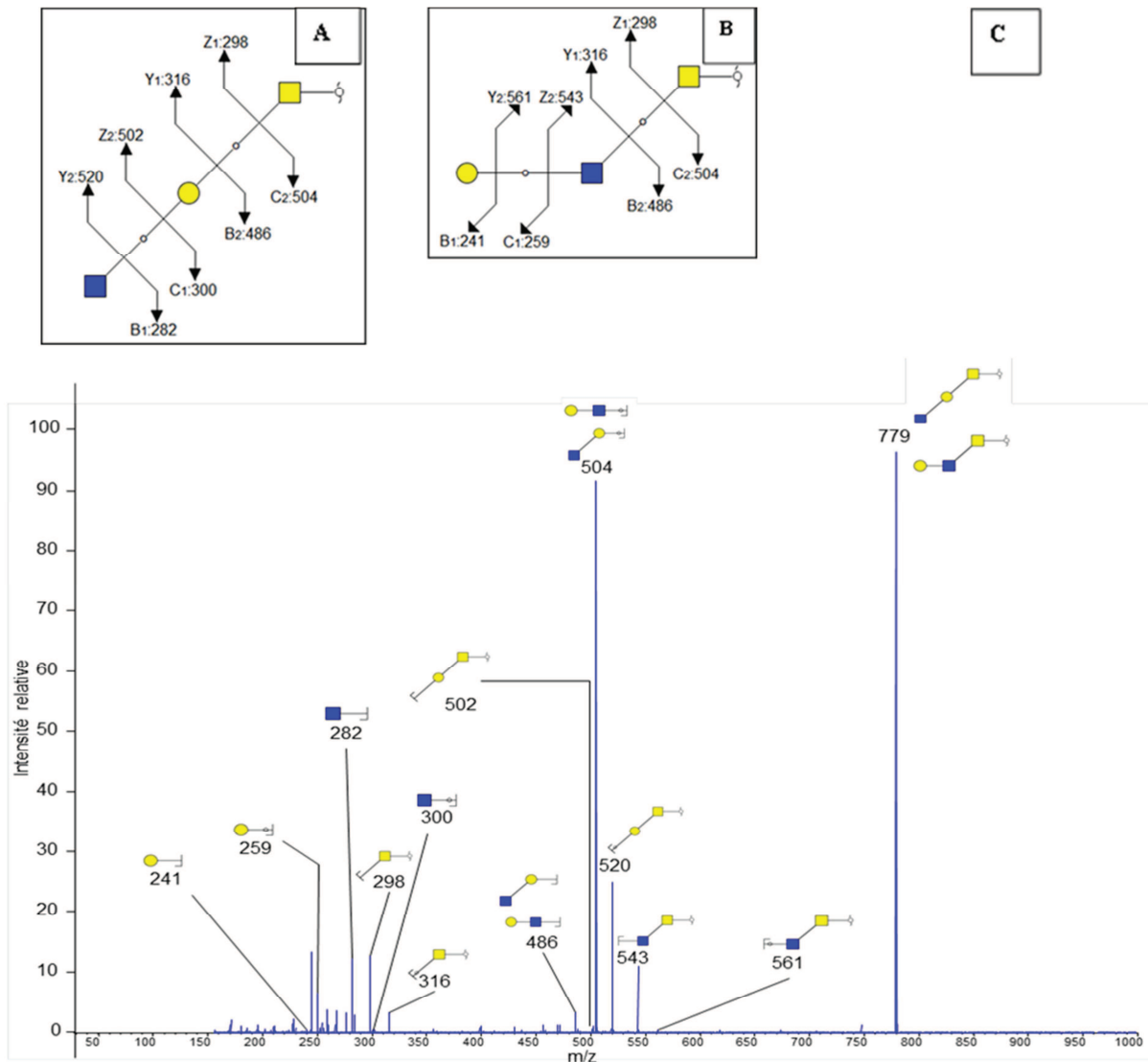


Figure 11: Spectre de fragmentation MS/MS en mode positif de l'ion à m/z 779 correspondant à un trisaccharide porté par les mucines intestinales. (A-B) Schémas de fragmentation du trisaccharide. (C) Spectre de fragmentation.

L'analyse du spectre de fragmentation de l'ion à m/z 1041 (Figure 12) montre que cet ion correspond à la masse d'un tétra saccharide sulfaté formé de la GalNAc du point d'attache, d'un résidu de Gal, d'un résidu de GlcNAc, d'un résidu de fucose et d'un groupement sulfate. L'identification et l'interprétation des différents ions fragments nous permettent de déterminer sa structure, annotée sur la figure 12A. Par convention, les ions fragments correspondant à la chaîne liée en 6 sur la GalNAc sont notés α alors que ceux correspondant à la chaîne liée en 3 sont notés β . L'ion $B_{1\alpha}$ à m/z 370 indique que c'est la GlcNAc qui porte le groupement sulfate. Les ions $Y_{1\beta}$ et $Z_{1\beta}$ à m/z 649 et 631 nous montrent que cette GlcNAc sulfatée est liée sur la GalNAc. Enfin, la série d'ions à m/z 415, 433, 506 et 694 nous permet de déterminer que la branche liée en 3 sur la GalNAc est composée d'un résidu Gal lié à un résidu Fuc.

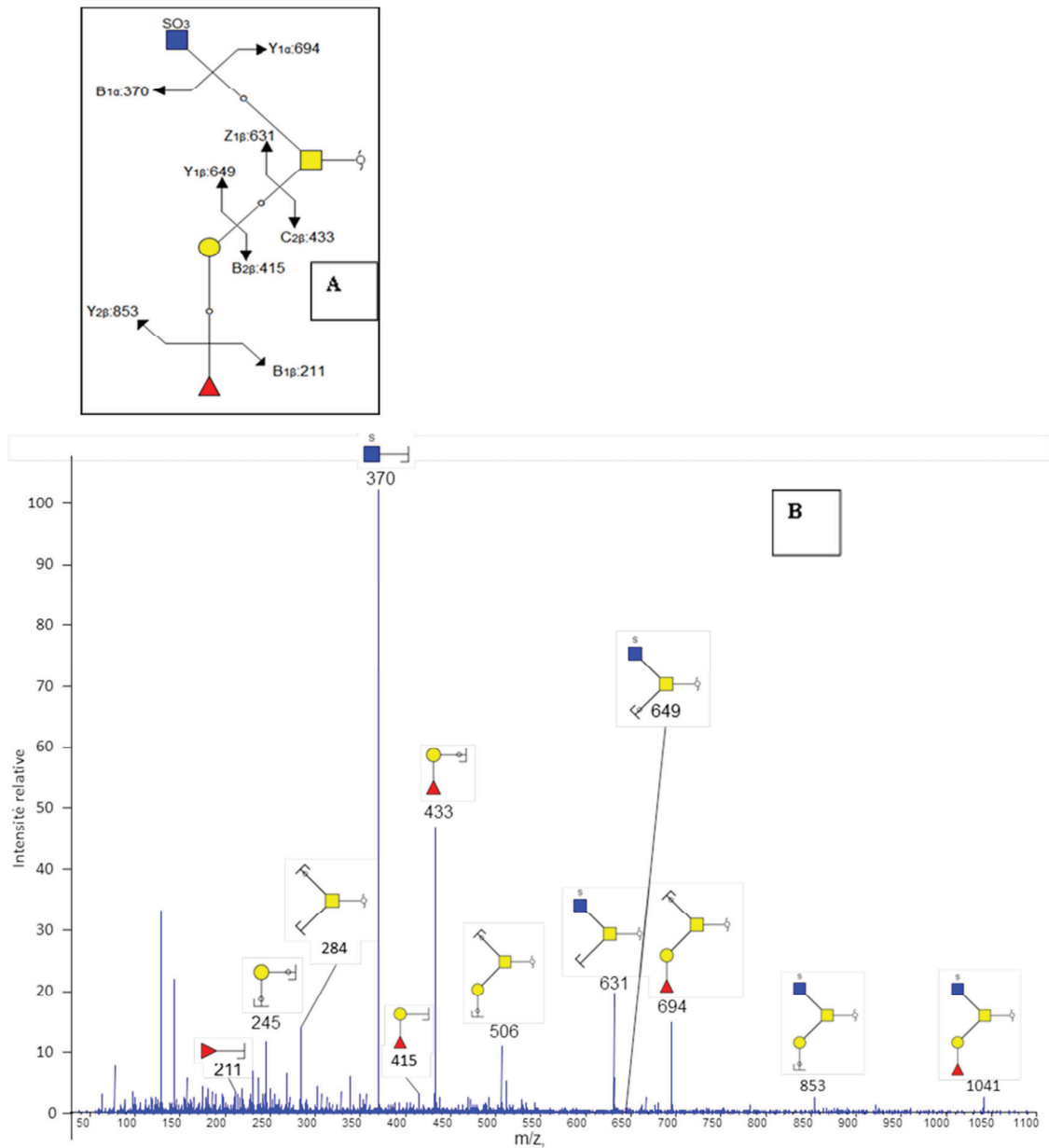


Figure 12: Spectre de fragmentation MS/MS en mode positif de l'ion à m/z 1041 correspondant à un tétra saccharide sulfaté porté par les mucines intestinales. (A) Schéma de fragmentation du tétra saccharide. (B) Spectre de fragmentation.

1.1.3.5 La méthylation, une MPT modeste qui joue dans la cour des grands !

Dès le début des années soixante les premières publications apparurent sur la méthylation de groupements lysine des protéines flagellaires chez des bactéries (Ambler and Rees 1959) puis des histones (Murray 1964).

Les principaux acides aminés subissant cette modification sont la lysine, l'arginine, l'acide aspartique, l'acide glutamique. D'autres résidus comme l'histidine, la cystéine, la sérine et la thréonine ont aussi été décrits comme pouvant l'être (Paik, Paik et al. 2007). La substitution d'un atome d'hydrogène par un groupement méthyl s'accompagnera d'un incrément de masse de 14,0156 Da, 28,0312 Da et 42,0468 Da dans le cas respectivement d'une mono, di ou tri-méthylation (dans le cas d'un résidu lysine).

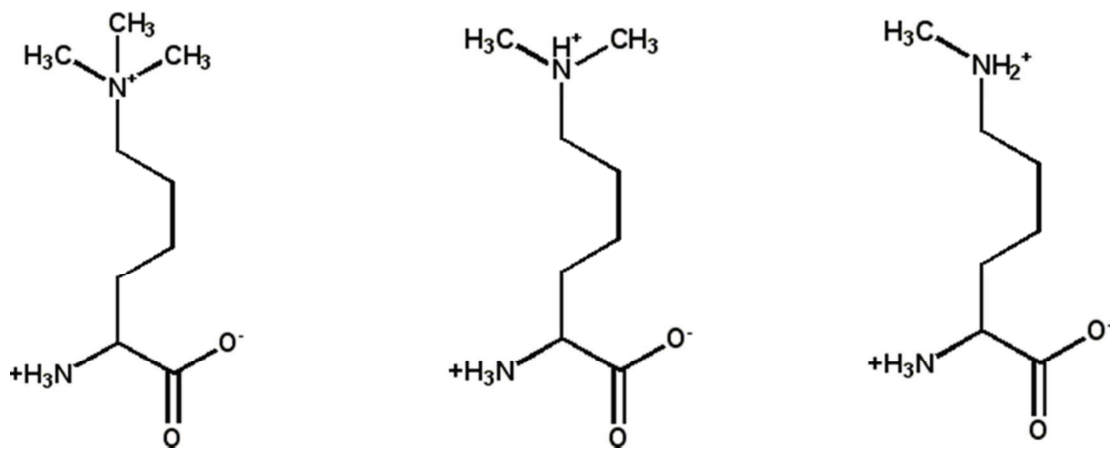


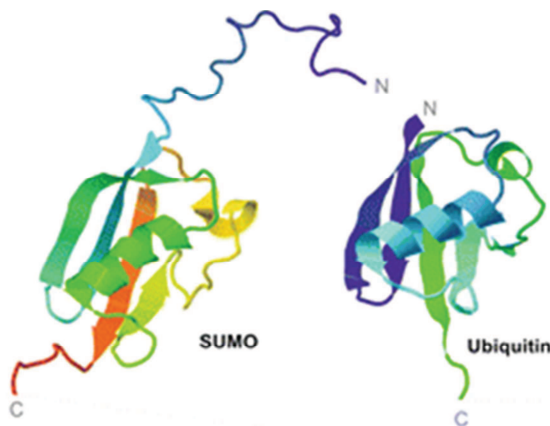
Figure 13: L'extrémité N-terminale d'un résidu Lysine, susceptible d'être acétylée peut aussi être mono, di ou tri-méthylée.

Notons que la différence de masse entre une tri-méthylation et une acétylation est de 0,0364 Da soit inférieure à 40 ppm pour un peptide de taille moyenne (soit $m = 700$ Da). Elle peut être désormais facilement distinguée par des appareils à haute résolution de type FT-ICR-MS ou Orbitrap. La recherche d'ions marqueurs ou de perte de neutre dans les spectres MS/MS sera un moyen supplémentaire pour distinguer tri-méthylation et acétylation mais aussi confirmer la présence d'une mono-méthylation, di-méthylation ou tri-méthylation sur une lysine ou arginine (Afjehi-Sadat and Garcia 2013).

A partir des années 80, l'intérêt s'est accru pour ces méthylations possédant différentes fonctions biologiques telles la régulation des gènes et la transduction du signal. Les histones ont longtemps été la famille de protéines sur lesquelles les méthylations ont été le plus activement recherchées (Greer and Shi 2012). Cette modification constitue avec les acétylations, les phosphorylations et les ubiquitinations le « code histone » qui permet la régulation de l'ADN au niveau épigénétique. Aujourd'hui les méthylations sont aussi activement recherchées sur d'autres protéines n'appartenant pas à la famille des histones. (Zhang, Wen et al. 2012).

1.1.3.6 L'ubiquitination (ou ubiquitylation) et la sumoylation – les sœurs Garnier des MPT !

Les protéines SUMO (pour Single Small Ubiquitin-like modifier) et l'ubiquitine sont des protéines homologues présentant 20% de similarité de séquence en acides aminés. Comme on peut l'observer ci-dessous, leurs structures tridimensionnelles présentent de fortes homologies.



Source : Castro et al Cell. Mol. Life Sci. (2012) 69:3269-3283



Figure 14: A gauche Protéines SUMO humaine (PDB 1A5R) et Ubiquitine humaine (PDB 1UBQ) à la structure tridimensionnelle présentant de fortes homologies (Castro, Tavares et al. 2012). A gauche les sœurs Garniers héroïnes du film « Les Demoiselles de Rochefort » de Jacques Demy (1967).

1.1.3.6.1 L'ubiquitination est une modification post-traductionnelle des protéines qui consiste en un greffage d'une chaîne polypeptidique d'ubiquitine sur les lysines des protéines (Ciechanover 1994). L'ubiquitine est une petite protéine très stable de 76 acides aminés de poids moléculaire 8,4 kDa. Présente dans toutes les cellules eucaryotes, elle est très conservée au cours de l'évolution, l'ubiquitine humaine et celle de la levure présentant 96% d'identité sur leur séquence protéique au cours de l'évolution.

L'ubiquitination des protéines est réalisée par des enzymes en trois étapes.

1. Activation : L'enzyme activatrice E1 active l'ubiquitine par carboxylation terminale nécessitant de l'ATP.
2. Conjugaison : L'ubiquitine est transférée sur la cystéine active d'E1 par une liaison thioester puis de la cystéine active de E1 vers la cystéine active de l'enzyme de conjugaison E2.
3. Transfert : E2 se complexe alors avec l'ubiquitine ligase E3 qui catalyse le transfert de l'ubiquitine sur le groupement amine de la chaîne latérale de la lysine de la protéine cible.

E2 apporte la spécificité de cible protéique en jouant le rôle de module de reconnaissance de ce complexe.

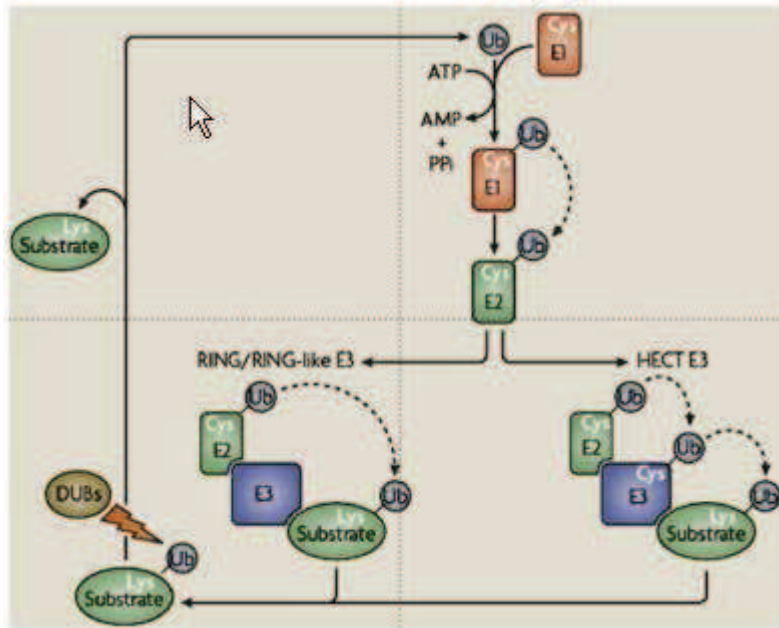


Figure 15 : La cascade d'ubiquitination. D'après (Ravid & Hochstrasser, 2008).

Il existe de nombreuses enzymes E1, E2 ou E3. Dans les cellules eucaryotes si l'on ne compte que deux gènes codant pour E1, 50 codent pour E2 et plus de 600 pour E3 (Terzic and Dikic 2011). En outre, il existe plus de 60 DUBs (Deubiquitinating enzymes) (Wilkinson 1997). Cela reflète la complexité et la dynamique de cette modification.

Mono- et poly- ubiquitination. Une des difficultés rencontrées dans l'étude de cette modification est le fait que l'ubiquitine soit capable de se greffer sur elle-même en 7 positions car elle contient elle-même 7 lysines (K6, K11, K27, K29, K33, K48 et K63) (Voir la Figure 16) d'après (Hicke, Schubert et al. 2005).

Suivant le type de greffage dans les chaînes d'ubiquitine, la conformation donnée à la protéine diffère. Les liaisons K48 conduisent généralement à la dégradation des protéines par le protéasome alors que les liaisons K63 servent plutôt à réguler la localisation de la protéine. Les mono ubiquitinations favorisent les interactions entre protéines.

En dehors de l'ubiquitination il existe d'autres types de modifications des protéines par greffage peptidique. Ces modifications utilisent d'autres systèmes enzymatiques qui sont reconnus par d'autres types de protéines adaptatrices mais par analogie se nomment Ubl pour Ubiquitin like. Elles comprennent les « small Ub-like modifier » (SUMO), les « neural precursor expressed, developmentally downregulated 8 » (NEDD8) et « l'interferon induced 15 kDa protein » (ISG15) (Kerscher, Felberbaum et al. 2006).

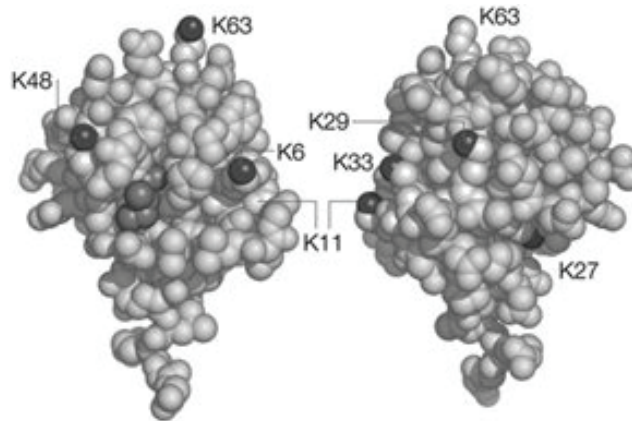


Figure 16 : L'ubiquitine, deux faces de la molécule d'ubiquitine sont ici représentées permettant de mettre en évidence les principales lysines qui vont être sollicitées dans les processus d'ubiquitylation.

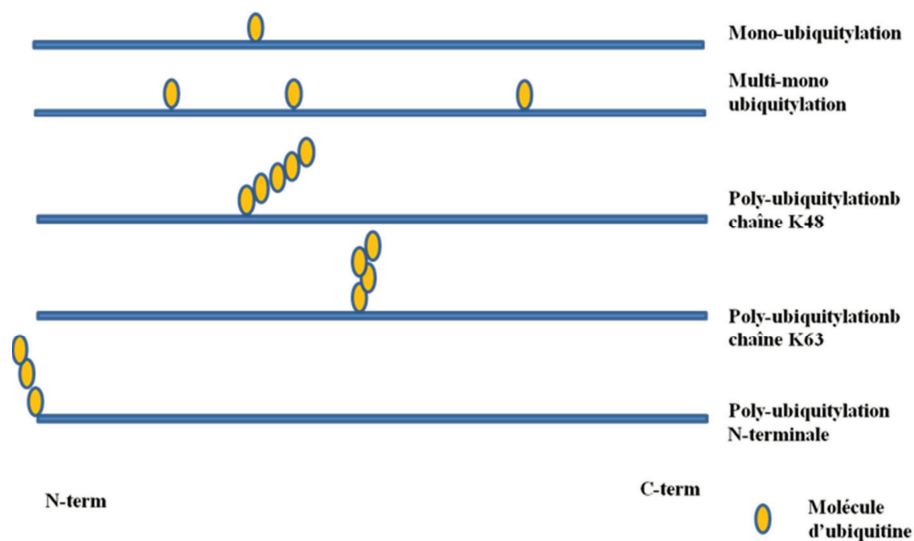


Figure 17 : Principaux types d'ubiquitination.

Les chaînes d'ubiquitine dites typiques ne contiennent que des branchements en K48. Les autres types de chaîne sont considérés comme atypiques. Elles peuvent être homotypiques si elles ne contiennent qu'un seul type de branchement ou des chaînes à liaisons mixtes (K6/11, 27/29, 29/48, 29/33 ((Kim, Kim et al. 2007). Des chaînes hétérologues avec des branchements Ubiquitine et SUMO ou NEDD8 ont également été publiées.

1.1.3.6.2 La sumoylation

A l'instar de l'ubiquitination, la sumoylation est la fixation de protéines de type SUMO (« Small Ubiquitin-related **M**odifier ») via encore une fois une lysine sur des protéines cibles. On parlera de famille concernant les protéines de type SUMO, d'un poids moléculaire d'environ 10 kDa. Les protéines SUMO sont exprimées de façon ubiquitaire chez les eucaryotes, le génome humain code pour quatre protéines SUMO différentes (Geiss-Friedlander and Melchior 2007).

Cette modification réversible va s'effectuer via une cascade enzymatique en partie analogue au processus d'ubiquitination, que l'on décrit ici en quatre étapes :

- Activation des protéines SUMO qui ont été synthétisées sous forme de précurseur inactif. Sous l'action d'une protéase spécifique des ubiquitines (« Ubiquitin-like Specific Protease »), les quatre derniers acides aminés sont clivés faisant apparaître une extrémité Gly-Gly,
- Une autre enzyme (« E1-activating enzyme ») composée de deux sous-unités va poursuivre l'activation des protéines SUMO par une réaction d'adénylation et de thio-estérification,
- Trans-estérification du résidu catalytique Cys de « E2-conjugating enzyme » UBC9,
- C'est cette nouvelle entité UBC9 qui se lie à la protéine SUMO et va permettre à la protéine cible d'être reconnue par une séquence consensus du type [I/L/V]K,X[D/E] où [I/L/V] représente un résidu d'acide aminé hydrophobe, X un acide aminé quelconque, quant à [D/E] un des acides aminés acide. C'est au cours de cette quatrième étape que se formera une liaison isopeptidique entre le groupement Glycine de la partie C-terminale de la protéine SUMO et l'amine de la chaîne latérale de la lysine présente dans la séquence consensus de la protéine cible.

Enfin notons que le processus de sumoylation est réversible et la desumoylation s'effectuera au moyen de la même protéine Ulp ou SENP qui catalyse à la fois l'hydrolyse du précurseur SUMO (première étape) et en fin de cycle la désomoylation de la protéine cible.

Comme pour l'ubiquitination, la sumoylation concernera un ou plusieurs sites (lysine) sur une même protéine, entraînant des états de monosumoylation, lorsqu'un seul résidu SUMO est attaché à la protéine, de multisumoylation lorsque plusieurs résidus lysine de la protéine sont concernés par une sumoylation ou enfin de polysumoylation lorsqu'une chaîne de résidus SUMO se fixe sur un résidu lysine de la protéine.

La sumoylation est un processus très dynamique soumis à des stimuli tant internes qu'externes, tels qu'un moment du cycle cellulaire, un état nutritionnel, un choc thermique, ou des dommages à l'ADN.

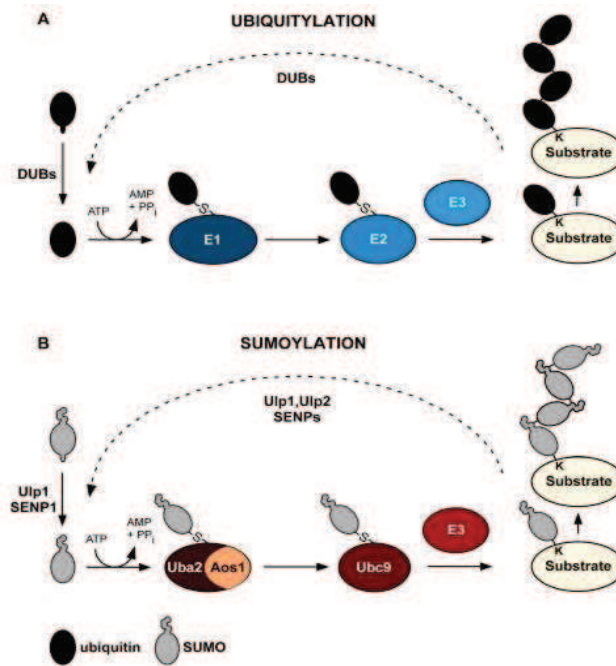


Figure 18 : Sumoylation, la mise en parallèle du cycle de l'ubiquitylation et de la sumoylation souligne les fortes similitudes et l'intervention d'enzymes apparentées (Miteva, Keusekotten et al. 2010).

Le rôle de la sumoylation est cependant très différent de celui de l'ubiquitination essentiellement destinée à la dégradation des protéines. L'ajout d'un groupement protéique relativement volumineux peut entraîner des effets mécanistiques évidents :

- Par encombrement stérique, empêcher l'interaction entre la protéine cible (sumoylée) et une autre protéine.
- Au contraire offrir une nouvelle zone d'interaction à la protéine cible nouvellement sumoylée et provoquer une modification de la localisation de la protéine ainsi sumoylée.
- Modifier la protéine cible d'un point de vue conformationnel et modifier ainsi son activité.

Des centaines de protéines cibles du motif SUMO sont connues, la majorité d'entre elles sont des protéines nucléaires (Geiss-Friedlander and Melchior 2007), mais on trouve aussi des protéines sumoylées au niveau du cytoplasme, de la membrane du réticulum endoplasmique ou de la membrane plasmique.

1.1.4 Quelques MPTs spécifiques et leurs caractéristiques

Il existe des modifications post-traductionnelles plus particulières. Le choix présenté ici de deux MPT spécifiques est purement subjectif et en partie seulement relié au travail expérimental présenté plus loin.

1.1.4.1 L'iodation des tyrosines

La thyroglobuline est produite par les cellules épithéliales de la thyroïde appelées thyrocytes qui forment des follicules sphériques.

La Thyroglobuline (Tg) est une glycoprotéine (PM=670 kDa) synthétisée par les cellules folliculaires de la thyroïde (Figure 19). Elle est sécrétée et stockée dans le colloïde de la lumière folliculaire. Elle est la protéine précurseur des hormones thyroïdiennes, qui diffusent ensuite dans la circulation capillaire.

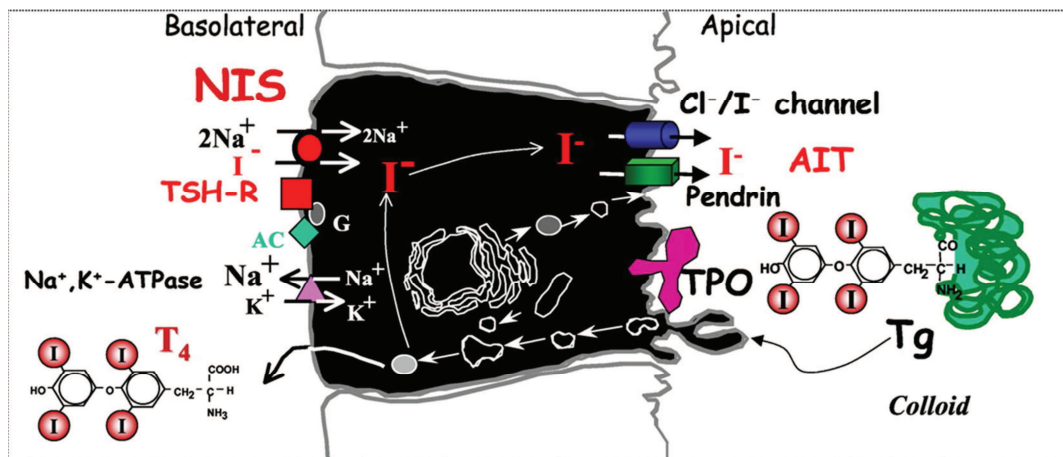


Figure 19 : Iodation de la thyroglobuline dans les cellules folliculaires de la thyroïde.

La Tg subit une iodation dans le colloïde. L'ensemble formé par la Tg et molécules T3, T4, MIT et DIT, est alors stocké dans le colloïde. Après son passage par microendocytose du colloïde dans la cellule épithéliale, la thyroglobuline est hydrolysée par des enzymes protéolytiques libérant ainsi les hormones thyroïdiennes T3 et T4 qui sont ensuite sécrétées dans le plasma. Ces étapes, notamment la libération, sont activées par la TSH dont la sécrétion est freinée par les hormones thyroïdiennes.

L'iodure, qui arrive dans le colloïde, partie centrale du follicule thyroïdien est activé par la peroxydase thyroïdienne ou thyroperoxydase, enzyme à sélénium, en I° ou I^{+} qui se fixe sur les noyaux tyrosine de la thyroglobuline pour former des résidus de monoiodotyrosine (MIT) et de diiodotyrosine (DIT) (Figure 20-A) (Taurog, Dorris et al. 1996; Dunn and Dunn 1999).

Un résidu de monoiodotyrosine et un résidu de diiodotyrosine se combinent pour former la triiodothyronine, T3, et deux résidus de diiodotyrosine pour former la tétraiodothyronine ou thyroxine, T4. T3 et T4 sont fixées à la thyroglobuline (Figure 20-B).

Une partie de la T3 libérée par les thyrocytes provient de la transformation de T4 en T3 sous l'influence de la 5'-désiodase dans les tissus périphériques (foie) (Figure 20-C).

Ces différentes modifications que subit la thyroglobuline vont conduire à des incréments de masse conséquents (Tableau 5), qui vont permettre de déterminer sans ambiguïté le degré d'iodation des tyrosines par spectrométrie de masse et ceci d'autant plus que cette modification est totalement résistante aux processus d'ionisation utilisés par cette technique.

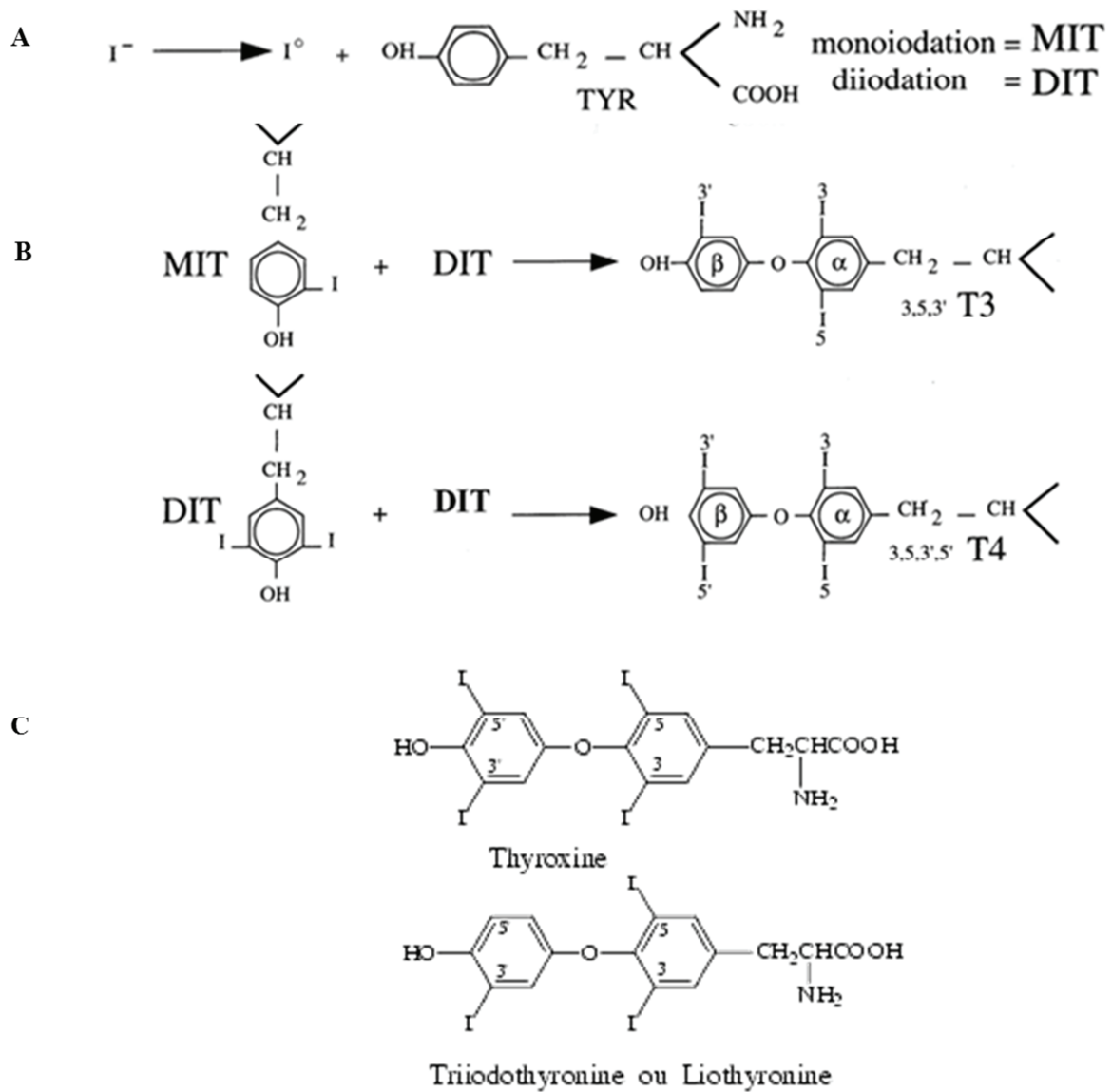


Figure 20 : Mécanisme d'iodation de la thyroglobuline.

Tableau 5 : Incréments de masse, caractéristiques des différents degrés d'iodation de la tyrosine et des divers résidus produits au cours de ce processus.

| | |
|-----------------------------|-------------|
| Tyrosine | 163,176 |
| Sulfo (Y) | + 79,946815 |
| Iodo (Y) | +125,896648 |
| Diiodo (Y) | +251,793296 |
| Triiodothyronine (Y) | +469,716159 |
| Thyroxine | +595,612807 |
| Tyr → Dha (Y) | -94,041865 |

Dha : Di-hydro-alanine

1.1.4.2 La bromation

Nous avons choisi de décrire cette modification en raison de sa rareté et son originalité. Il s'agit d'une bromation s'effectuant sur un résidu tryptophane décrite pour la première fois sur un heptapeptide d'un escargot cône marin (Craig, Jimenez et al. 1997).

La modification transforme le résidu tryptophane en 6-bromotryptophane. Cette modification a été mise en évidence par spectrométrie de masse (77,9105), co-élution en comparaison avec un peptide synthétique. Ici aussi comme dans le cas des peptides iodés, l'aspect hydrophobe retarde sur une colonne en phase inverse l'élution du peptide.

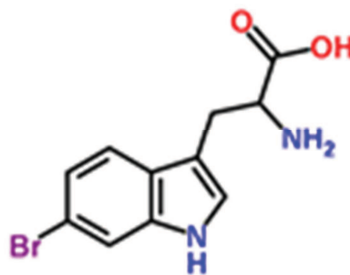


Figure 21 : Bromation du tryptophane en 6-bromotryptophane.

Les peptides produits par ces escargots présentent des structures très contraintes de par les très nombreux ponts disulfures qu'ils contiennent et les nombreuses modifications post-traductionnelles de ses acides aminés, hydroxy-proline, acide gamma-carboxyglutamique et bien sûr bromotryptophane. Tout cela contribue à donner la capacité, à ces peptides présents dans le venin de ces cônes, de bloquer divers récepteurs avec une haute affinité et une grande sélectivité.

De façon similaire d'autres peptides ainsi modifiés ont été détectés chez *Conus textile*, et conduisent à la réduction de la libération de neuromédiateur au niveau de la synapse cholinergique, ceci en réduisant l'afflux pré-synaptique de calcium, de façon lente et réversible (Rigby, Lucas-Meunier et al. 1999).

Plus récemment (Aguilar, Lopez-Vera et al. 2005) découvraient chez *Conus delessertii* un peptide plus gros (13 aa, masse monoisotopique 3487,09 Da), comprenant cette modification et d'autres encore comme 3 (4-hydroxyproline) et 2 (5-hydroxylysine).

1.2 La spectrométrie de masse, un outil analytique en très forte évolution, technique clef de la protéomique

La protéomique s'est développée à partir des années 90. Depuis sa « naissance », l'essor de la protéomique repose sur trois piliers (Patterson 2003; Patterson and Aebersold 2003) (voir Figure 22)

- Les moyens de subdiviser un protéome très complexe (Gels 1D, 2D, immuno-séparation, chromatographie SCX...), des modes de séparation des peptides ou des protéines en nombre élevé ; toute une biochimie séparative ou extractive dont le défi est de permettre d'extraire ou/et de séparer des protéines par centaines ou milliers, des peptides par milliers ou dizaine de milliers.
- La spectrométrie de masse qui nécessite des appareils n'ayant cessé d'améliorer leur résolution, leur précision de mesure, leur rapidité. Ils permettent d'analyser un nombre toujours plus grand de peptides, disposant de modes de fragmentation divers (CID, HCD, ETD...).
- Enfin les bases de données et d'analyses toujours plus nombreuses, mieux annotées dotées de logiciels permettant d'effectuer une analyse et une validation des données toujours plus performante.

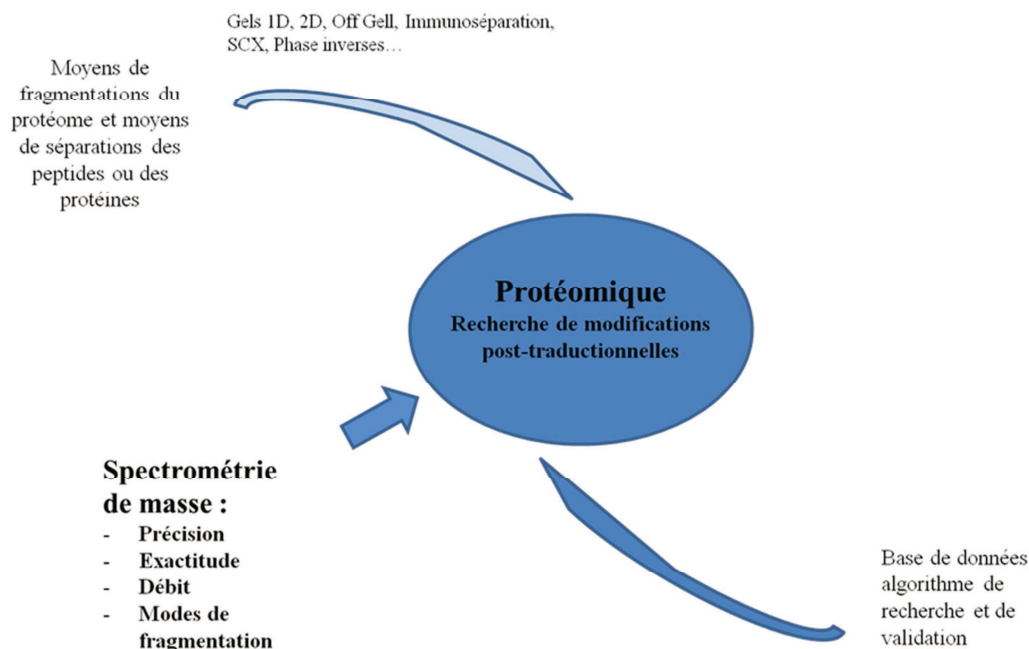


Figure 22 : La protéomique au point de rencontre de trois groupes de techniques.

De ces trois groupes de techniques, la spectrométrie de masse a plus particulièrement connu des évolutions, essentielles et remarquables qui ont permis de grandement contribuer à l'étude des MPT. C'est ce que nous développerons dans les chapitres suivants.

De la chimie des protéines à la protéomique :

Jusqu'au début des années 90 la chimie des protéines a été étudiée au travers des techniques traditionnelles de biochimie et de biophysique (extraction, centrifugation, purification, caractérisation structurale par RMN ou cristallographie, caractérisation fonctionnelle...).

Au tournant des années 90, la protéomique a durablement transformé cette approche.

Mark Wilkins a proposé le terme protéome en 1994 lors d'un colloque organisé à Sienne en Italie. Le terme protéomique a lui été introduit en 1997 par Peter James pour faire pendant au terme génomique. Le protéome est l'ensemble des protéines présentes dans un organisme, une cellule, un tissu, un organe à un moment donné dans des conditions données. La protéomique est donc l'étude du protéome et se décline sous différentes formes :

- La cartographie des protéines au niveau cellulaire ou sub-cellulaire,
- La protéomique différentielle qui étudiera l'état comparatif du protéome entre un état normal et stressé (Monteoliva and Albar 2004) et qui est renforcée par des approches de protéomique quantitative, dimension supplémentaire qui permet de quantifier les protéines présentes
- L'étude des modifications post-traductionnelles, permettant de cerner des sous-protéomes fonctionnels, tels que le phosphoprotéome, l'acétylome....
- Protéomique fonctionnelle: cartographie des voies de signalisation s'attachant à l'analyse des complexes de protéines, création de réseaux de protéines qui interagissent et analyse dynamique des réseaux d'interactions (Colland, Jacq et al. 2004).

Nous allons ici aborder les principales caractéristiques des spectromètres de masses pour comprendre comment cette technique a pu bouleverser l'étude en masse des protéines et de leurs modifications post-traductionnelles.

1.2.1 Les différentes techniques d'ionisation

La spectrométrie de masse s'est d'abord développée sur des composés organiques de faible poids moléculaire. Les techniques d'ionisation utilisées, relativement énergétiques n'étaient pas adaptées à l'ionisation des peptides ou des protéines.

C'est avec l'introduction de méthodes d'ionisation plus douces, permettant l'ionisation de composés non-volatils et relativement fragiles que l'analyse des peptides a pu se développer.

Les ionisations de type ESI (ElectroSpray Ionisation) (Whitehouse, Dreyer et al. 1985) et MALDI (Matrice Assisted Laser Dissociation and Ionisation) (Tanaka K. et al. 1988), permettent l'ionisation douce de composés organiques tels les peptides ou les protéines entières. Elles ont été optimisées et se trouvent être aujourd'hui les techniques qui se sont imposées en protéomique.

Le processus mis en œuvre dans ces deux cas est très différent, les peptides ne seront pas ionisés de la même façon, avec le même rendement. Ainsi elles peuvent avoir une certaine complémentarité qui peut être mise à profit dans la recherche intensive de modifications post-traductionnelles. On attribue à l'ionisation ESI une meilleure ionisation des peptides hydrophobes tandis que l'ionisation de type MALDI permettra une meilleure observation des peptides basiques et aromatiques (Stapels and Barofsky 2004).

1.2.2 Résolution, exactitude, précision de masse et haut débit

1.2.2.1 Résolution

En spectrométrie de masse, la résolution est la capacité pour un appareil de mesurer deux signaux distincts et de pouvoir les considérer comme deux événements distincts (Figure 23). Une résolution élevée va se traduire sur un spectre de masse par la finesse des pics mesurés. On peut définir la résolution comme étant le rapport $m/\Delta m$ où m sera la masse de l'ion mesuré et Δm la différence entre deux pics voisins. On peut aussi prendre en compte la largeur du pic par exemple à mi-hauteur ;

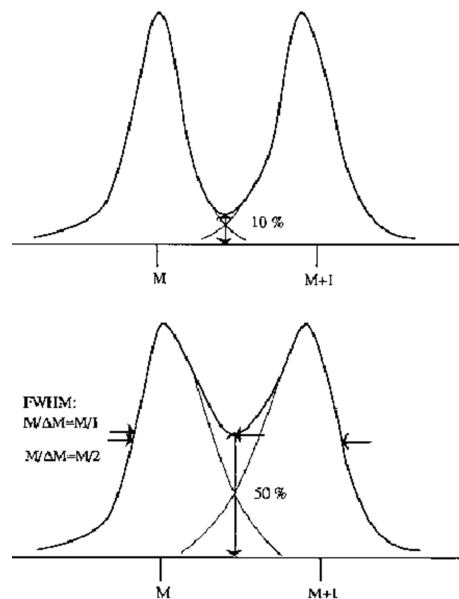


Figure 23 : La résolution traduit la capacité à distinguer deux événements différents ici à mesurer deux rapports (masse/charge) voisins.

Dans le premier cas, être capable de distinguer un peptide ionisé de masse 1000 et un autre de masse 1000,1 suppose un appareil ayant une résolution de $1000/(1000,1-1000) = 10\ 000$. Dans le deuxième cas, une résolution de 10 000 sera atteinte pour un ion de masse 1000 dont la largeur du pic à mi-hauteur sera de $0,1Th^*$

- *Le thompson (Th), l'unité utilisée en spectrométrie de masse et qui mesure le rapport m/z (masse sur charge du peptide ionisé).*

Selon les appareils des différences considérables existent concernant ce pouvoir de résolution ce qui conduira à des résultats sensiblement différents (Figure 24).

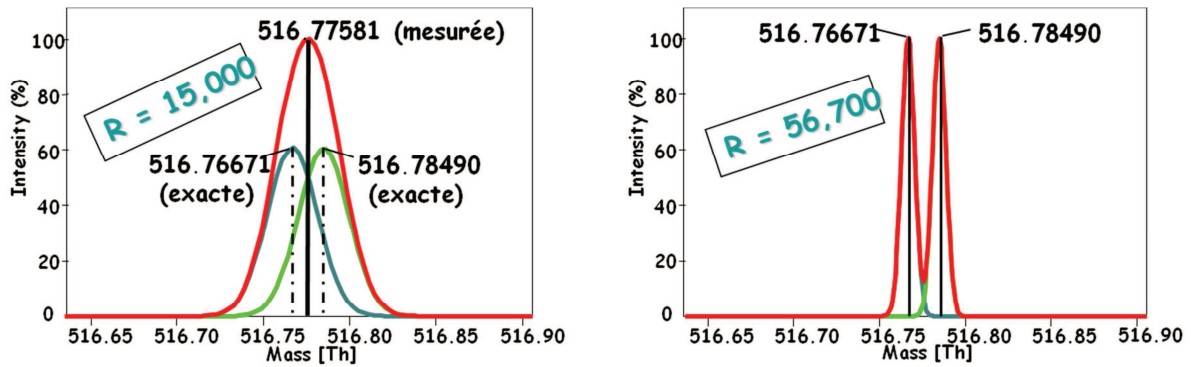


Figure 24 : Conséquence directe de différents pouvoirs de résolution d'un spectromètre de masse sur la détermination de deux masses voisines.

Un appareil (figure de gauche) ayant un pouvoir de résolution de 15 000, ne permet pas de distinguer entre deux peptides de masses moléculaires proches (moins de 35 ppm). Au contraire on peut distinguer sans ambiguïté ces deux peptides (figure de droite) avec un appareil dont le pouvoir de résolution est supérieur à 50 000. Concrètement, on pourra alors distinguer sans ambiguïté une acétylation d'un groupement lysine (incrément de masse, 42,0105) d'une tri-méthylation de ce même groupement (incrément de masse (42,0468)).

On dispose aujourd'hui d'appareils de masse aux résolutions très différentes (Tableau 6)

Tableau 6 : Spectromètres de masse et résolution. Les spectromètres de masse ont considérablement évolué au cours des dernières années, on peut aujourd'hui travailler couramment avec des appareils à haute résolution (30 à 60 000).

| Pouvoir de résolution des spectromètres de masse | |
|--|-----------------------------|
| | Valeurs types |
| • Quadrupole | 1 000 (50% du maximum) |
| • TOF (linéaire) | 2 000 (largeur mi-hauteur) |
| • TOF (reflectron) | 30 000 (largeur mi-hauteur) |
| • Instruments à champs magnétique | 150 000 (10% du maximum) |
| • ORBITRAP (haute résolution) | 150 000 (10% du maximum) |
| • FT-ICR (haute résolution) | 1 000 000 (10% du maximum) |

1.2.2.2 Précision et exactitude

La haute résolution permet d'obtenir des mesures précises c'est-à-dire obtenues avec un grand nombre de chiffres significatifs après la virgule. Pour autant ces mesures précises seront exactes si l'étalonnage du spectromètre de masse a été correctement effectué.

Ces évolutions ont été très sensibles au cours des quinze dernières années.

Au début des années 2000 la mesure des peptides par les approches par empreintes peptidiques s'effectuait couramment avec une tolérance de 50 ppm et au mieux à 25 ppm.

Aujourd'hui avec un appareil de masse travaillant à une résolution de 30 à 60 000 les mesures de masse de ces mêmes peptides s'effectueront à 5 ppm d'incertitude près, voire dans les meilleurs cas à moins de 1 ppm !

La conséquence immédiate est que la détermination d'une masse exacte avec une précision plus grande va réduire considérablement le nombre de peptides candidats potentiels dont la masse se situera dans l'intervalle de précision. Le risque de mauvaise interprétation est ainsi minimisé. Typiquement l'exactitude apportée par les spectromètres de masse les plus récents permet de distinguer aujourd'hui sans ambiguïté une tri-méthylation d'une acétylation portant sur l'amine d'un résidu lysine.

Cependant la question de pouvoir distinguer des aminoacides isobares (leucine et isoleucine) reste posée.

1.2.3 Des modes de fragmentation variés

La capacité d'un spectromètre de masse ne consiste pas seulement à mesurer avec précision et exactitude la masse d'un peptide, mais à être capable de fragmenter au mieux toutes sortes de peptides afin d'en déterminer la séquence en acides aminés. Aussi a-t-on vu au cours des dernières années l'amélioration des modes de fragmentation ou même, l'apparition de modes de fragmentation nouveaux.

Ces modes de fragmentation constituent l'originalité de la spectrométrie de masse en tandem MS/MS.

Ces fragmentations vont se produire essentiellement au niveau de la chaîne peptidique et le type d'ions fragments observés va dépendre :

- de la séquence primaire du peptide à fragmenter,
- de l'énergie interne des liaisons,
- de l'état de charge du peptide,
- de l'énergie fournie pour la fragmentation.

Une nomenclature commune a été proposée dès 1984 par Roepstorff et Fohlman (Roepstorff and Fohlman 1984), modifiée par (Johnson, Martin et al. 1987) et (Biemann 1988). Cette dernière version proposée par Biemann est la version officielle aujourd'hui utilisée et présentée dans la figure 25.

Les fragments ne seront détectés que s'ils portent au moins une charge. Si cette charge est retenue sur le fragment N-terminal, l'ion est classé comme étant soit a, b ou c. Si la charge au contraire est retenue sur la partie C- terminale, le type d'ions est soit x, y ou z. Un indice indique le nombre de résidus dans le fragment. Ainsi en mode de fragmentation CID les ions prédominants seront de type y et b.

Pouvoir identifier les ions fragments d'une même série (les séries y ou b) conduira à la détermination de la séquence du peptide car la différence de masse entre deux pics de la même série correspond à la masse caractéristique d'un résidu.

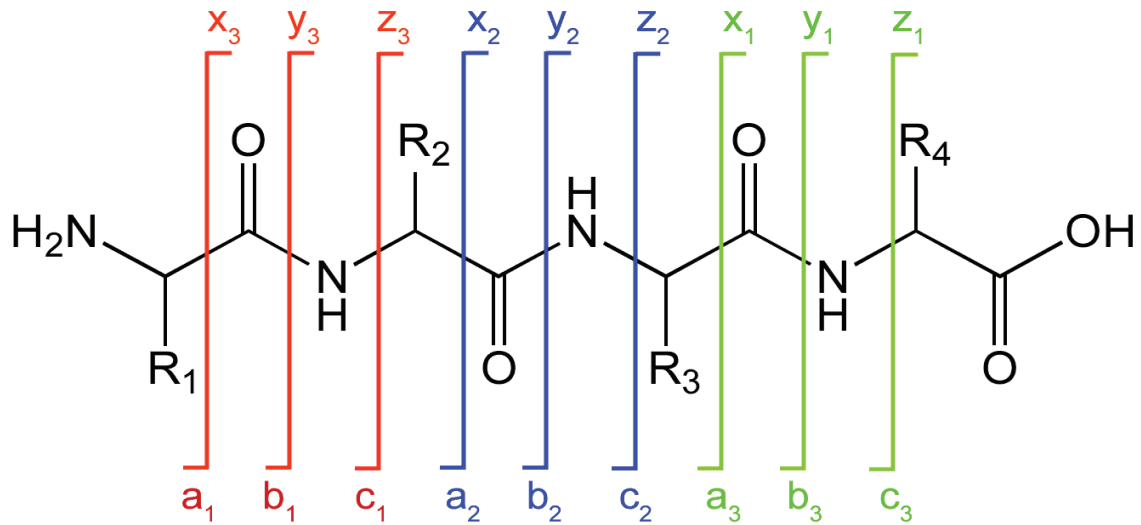


Figure 25 : Nomenclature officielle des fragments pouvant être obtenus au cours d'une analyse par MS/MS.

Sur ce tri-peptide théorique, les ions fragments retenant la charge positive vont pouvoir apparaître de façon complémentaire, a_2 , b_2 , c_2 pour la partie N-terminale et x_2 , y_2 , z_2 seront respectivement les ions fragments complémentaires pour la partie C-terminale de cette chaîne.

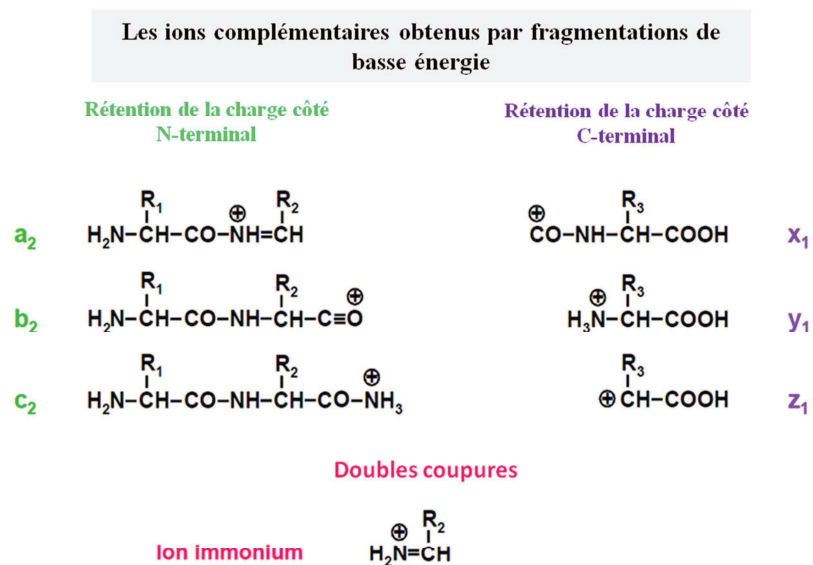


Figure 26 : Les ions complémentaires obtenus par fragmentations de basse énergie.

Des séries d'ions complémentaires vont être formées retenant des charges (selon la nature des chaînes latérales des acides aminés qu'ils contiennent). Ainsi les charges pourront être retenues d'un seul côté. Notons la possibilité par des doubles coupures de voir apparaître des ions immonium qui peuvent signer la présence d'un acide aminé particulier (Figure 26).

1.2.3.1 CID

C'est le mode de fragmentation le plus courant qui se décline sous les termes de CAD (Collisionally Activated Dissociation) ou CID (Collision-Induced Dissociation). C'est un mode de fragmentation considéré comme doux car à faible niveau d'énergie (Wells J.M. et al « Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol.* 2005, 402, 148-185).

Le bombardement par un gaz neutre (Ar, He, N₂, CO₂) d'ions peptidiques (ions parents) va entraîner

- l'excitation qui conduira une partie de l'énergie cinétique de l'ion à se convertir en énergie interne produisant ainsi un ion « excité ».
- puis la fragmentation en ions complémentaires (ions fils) qui seront ensuite séparés selon le rapport Masse/Charge.

C'est la liaison covalente entre le carbonyle et l'amine de la liaison peptidique qui va être coupée, donnant naissance aux ions fragments complémentaires b et y.

Des réarrangements ultérieurs peuvent couramment se produire : perte d'eau, perte d'ammoniaque.

1.2.3.2 HCD

C'est un mode de dissociation plus énergétique (**H**igher-energy **C**ollisional **D**issociation). Ce mode de fragmentation a été introduit dans les analyseurs de type Orbitrap. Ces analyseurs, outre la trappe LTQ où s'effectue en particulier la fragmentation de type CID, sont pourvus d'une nouvelle cellule de fragmentation sous forme d'un octopole dédié. Les peptides ionisés sont accélérés (1kV ou plus) pour entrer dans cette cellule de collision où se trouve un gaz, par exemple de l'azote avec lequel ils vont rentrer en collision et se fragmenter.

Les types de spectres obtenus pour un peptide donné présenteront le même profil que pour un spectre de type CID mais avec, dans la partie basse de la gamme de masse, des données nouvelles qui apparaissent clairement, les paires d'ions a₂, b₂ ainsi que les ions y₁ et y₂. Certains ions immonium seront parfaitement vus permettant de confirmer une séquence, par exemple les ions de la phosphotyrosine. Les ions rapporteurs de certaines méthodes de quantification telle que la méthode SILAC (Stable Isotope Labelling by amino acid in Cell Culture) pourront également être facilement détectés dans ces conditions (Olsen, Macek et al. 2007).

Dans l'exemple donné Figure 27, la fragmentation de type HCD va permettre une collecte d'information à partir d'une gamme de masse débutant à m/z = 100 Th. Pour un peptide de masse inférieure à 1250 Da, nous pouvons obtenir une série d'ions y presque complète de y₁ à y₆, détecter en outre entre y₃ et y₄ une différence de masse de 632,78 Da correspondant à la modification tri-iodo thyronine sur la tyrosine (différence de masse théorique 632,892). Tout ceci est confirmé par la présence de l'ion immonium tri-iodo thyronine à 605,792.

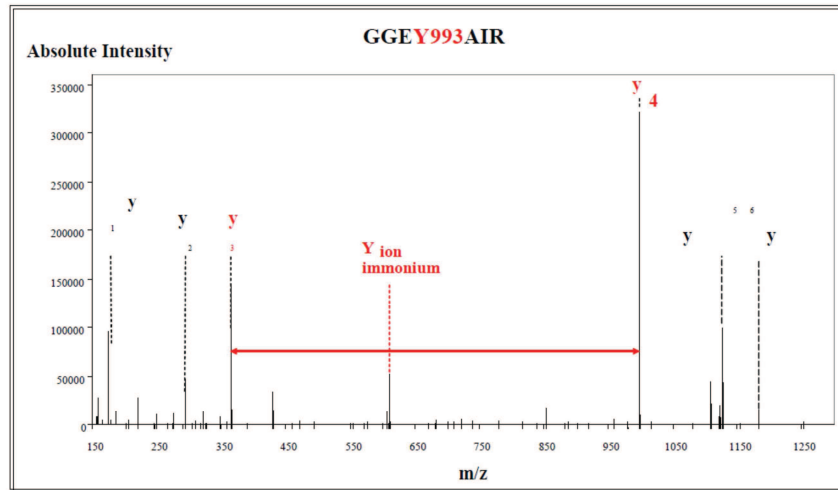


Figure 27: Fragmentation de type HCD.

Spectre HCD obtenu sur le peptide GGEYAIR et montrant une attribution sans ambiguïté d'une modification de type tri-iodo thyronine sur la Y993 par la présence des ions Y4 et Y5 mais aussi de l'ion immonium de la tyrosine.

1.2.3.3 ETD –Fragmentation par transfert d'électrons

Les fragmentations de type CID surtout et HCD plus récemment se sont imposées comme des méthodes de fragmentation de choix des peptides. Elles permettent de séquencer avec succès des peptides de petite taille (500 m/z) ou de taille intermédiaire (1500 m/z), obtenus par digestion par diverses protéases. Elles ne s'avèrent plus performantes lorsque ces peptides portent des modifications post-traductionnelles labiles telles que les phosphorylations sur des résidus serine, thréonine ou tyrosine ou encore des glycosylations de type O-GlcNAc.

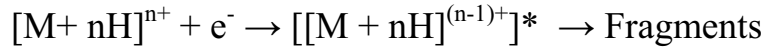
McLafferty dès 1998 proposait une nouvelle méthode de fragmentation fort différente : (Zubarev, Horn et al. 2000) complétée en 2004 par une méthode similaire proposée par Syka « Electron Transfer Dissociation (ETD) (Syka, Coon et al. 2004). L'ETD comme l'ECD est basée sur une fragmentation de cations (peptides) suite à un transfert d'électrons (dans le cas de l'ECD) ou d'un anion (dans le cas de l'ETD). Plusieurs étapes seront nécessaires:

- la génération d'un analyte (le peptide à fragmenter) sous forme d'un cation, et d'un réactif sous forme anion.
- l'apport de ces deux éléments en un même espace et un même temps,
- l'analyse massique des ions résultants.

Dans le cas de l'ETD, l'anion est généré par un ion auxiliaire (2-(fluoranthene-8-carbonyl) acide benzoïque) et transféré sur un ion peptidique multichargé. Les cassures vont se produire dans la chaîne peptidique là où l'énergie a été « déposée ». Ce n'est pas la liaison la plus fragile qui sera rompue, contrairement à la fragmentation de type CID. Les ions générés seront ici essentiellement de type c ou z (Nomenclature de Roepstorff et Fohlman).



Dans le cas de l'ECD, un électron est capté par un peptide multichargé.



La méthode par fragmentation ETD est apparue d'abord comme une méthode de choix pour détecter certaines modifications post-traductionnelles (phosphorylations et certaines glycosylations). En évitant les pertes de neutres c'est-à-dire le départ du groupement phosphate avant toute fragmentation de la chaîne peptidique, la méthode ETD apportera une information sur la localisation de la modification sur le peptide étudié (Kim and Pandey 2012) (Liu 2013).

1.2.3.4 CID, HCD, ETD des modes de fragmentation complémentaires pouvant être utilisés en synergie.

Si la méthode de fragmentation de type CID s'est imposée en spectrométrie de masse appliquée aux protéines, les deux autres méthodes ici décrites ne valent pas que pour leur originalité et capacité à générer de nouveaux fragments peptidiques préservant certaines liaisons fragiles porteuses de MPT. Elles peuvent servir à obtenir un meilleur recouvrement des protéines détectées.

Par ailleurs ce meilleur recouvrement peut justement permettre en explorant plus à fond chaque protéine identifiée de découvrir sur de nouveaux peptides jusque-là ignorés des modifications post-traductionnelles.

Les méthodes de fragmentation ETD seront plus efficaces sur des peptides de grande taille > 1000 m/z de charge +3 ou au-delà. Alors que le CID et HCD seront plus efficaces et complémentaires pour étudier des peptides de taille intermédiaire < 1000 m/z ou plus petite < 750 m/z. Là encore les améliorations apportées dans les instruments plus récents Orbitrap Velos ou Q-Exactive montrent même que les fragmentations de type HCD peuvent donner de meilleurs résultats en terme de spectre de fragmentation se traduisant par un score peptidique plus élevé. (Frese, Altelaar et al. 2011).

Par ailleurs l'ensemble de données obtenues par ces trois méthodes de fragmentation comme l'indiquent Guthals et al (Guthals, Clauser et al. 2013), permet par l'analyse de l'ensemble des ions obtenus b/y/c/z d'améliorer l'interprétation des peptides longs et chargés et d'améliorer grandement les analyses de novo d'identification de protéines.

1.2.4 Des spectromètres hybrides permettent des approches différentes.

Les spectromètres de masse en tandem permettent de mettre en œuvre une spectrométrie de masse type MS/MS. Ils peuvent être hybrides et associent dans ce cas plusieurs types d'analyseurs. Ces approches pourront parfois être mises à profit pour la détection de modifications post-traductionnelles.

1.2.4.1 Mode de fonctionnement courant

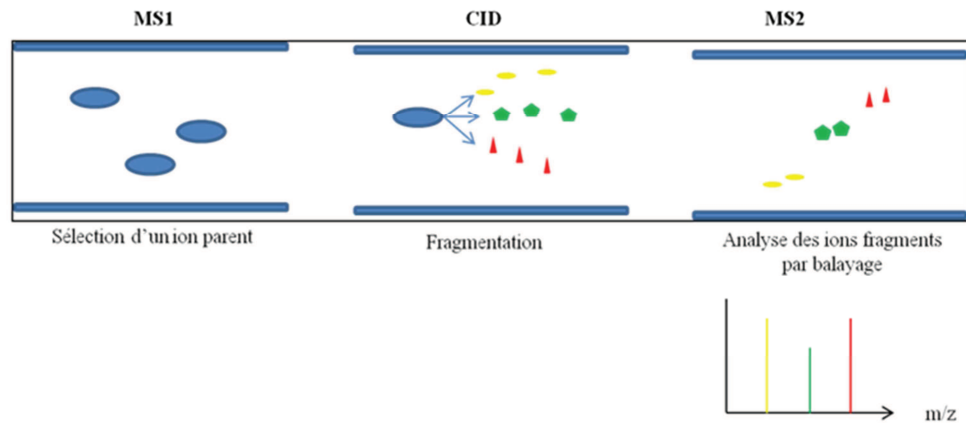


Figure 28 : Schéma d'un spectromètre de masse en fonctionnement courant.

Dans le fonctionnement le plus habituel d'un spectromètre de masse de type tandem, les ions produits dans la source, sont sélectionnés au cours d'un balayage, le peptide sélectionné et isolé est fragmenté par collision et les produits de cette fragmentation sont analysés afin de tenter de reconstituer la séquence du peptide parent.

1.2.4.2 Balayage de l'ion précurseur (« Precursor ion Scanning »)

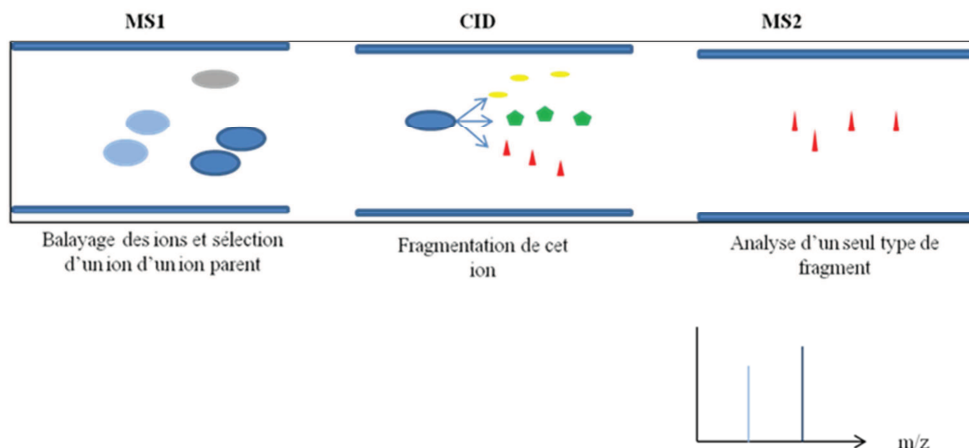


Figure 29 : Schéma d'un spectromètre de masse fonctionnant en balayage de l'ion précurseur.

Dans ce type d'approche le second analyseur MS2 ne transmettra qu'un fragment particulier attendu à une certaine masse tandis que dans le premier analyseur, un balayage va détecter tous les ions parents qui pourraient contenir ce fragment.

Typiquement ce mode de fragmentation s'applique sur la perte d'un ion phosphite à -79 Da spécifique des phosphopeptides lors de leur fragmentation par CID en mode négatif (Carr, Huddleston et al. 1996). Ce scan est généralement utilisé sur des machines de type triple quadripôle. Un premier quadripôle filtre les précurseurs un à un de façon très rapide. Un second quadripôle sert de chambre de collision de ce précurseur avec une faible énergie de collision en rampe. Enfin un troisième quadripôle ne mesure les signaux MS que dans une petite fenêtre de masse autour de -79 Da pour gagner en rapidité. Lorsqu'aucun précurseur ne génère pas d'ion phosphite, le cycle reprend en mode négatif. Lorsqu'un précurseur génère un ion phosphite, le spectromètre de masse passe en mode positif et acquiert un spectre de fragmentation pour le peptide concerné en recalculant sa masse m/z correspondante en mode positif.

Pour les phosphotyrosines, la procédure est identique mais l'ion signature recherché est l'immonium de la phosphotyrosine à 216 Da (Steen, Kuster et al. 2001). Tout se fait donc en mode positif mais cet immonium est souvent un peu faible et les peptides porteurs de phosphotyrosines très minoritaires dans les digestions tryptiques.

1.2.4.3 Balayage par perte de neutre (« Neutral loss scan »)

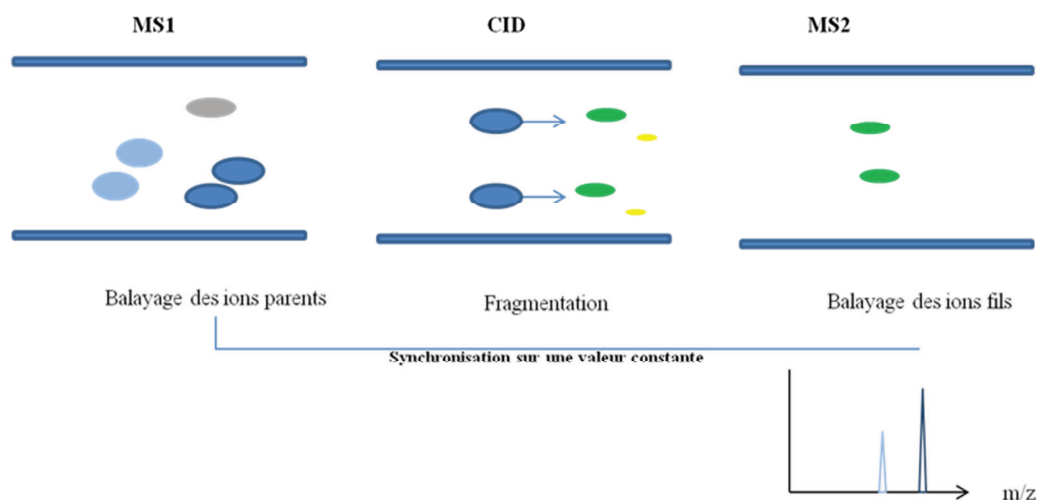


Figure 30 : Schéma d'un spectromètre de masse fonctionnant en balayage par perte de neutre.

Dans ce mode de fragmentation le balayage de perte de neutre s'accompagne d'un balayage synchronisé des deux analyseurs, de sorte que la différence de masse des ions passant à travers MS1 et MS2 reste constante. Ici aussi ce mode sera utilisé pour la recherche de phosphopeptides.

Ce balayage ne détecte pas un ion signature à une masse unique mais la perte de neutre de -98 Da du précurseur phosphorylé lorsqu'il commence à se fragmenter (Hunter and Games 1994). Il peut être utilisé pour déclencher une MS3 sur le précurseur - 98 (Beausoleil, Jedrychowski et al. 2004) et améliorer l'information de séquence des phosphopeptides. Sa véritable utilité dans le cadre d'une approche globale a cependant été critiquée (Villen and Gygi 2008).

1.2.4.4 Détection d'ions multiples (essai de type SRM/MRM pour « Selected Reaction Monitoring » et « MRM pour Multi Reaction Monitoring »)

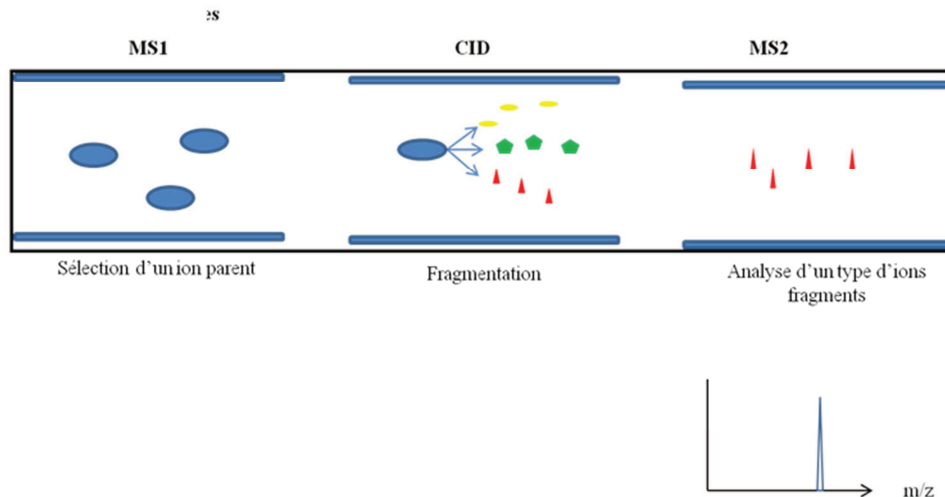


Figure 31 : Schéma d'un spectromètre de masse fonctionnant en détection d'ions multiples.

Ici l'analyseur MS1 va sélectionner un type d'ion parent et l'analyseur MS2 un type d'ions fragments de ce parent. Ainsi plusieurs balayages peuvent être mis en boucle au cours d'une expérience pour détecter la présence de plusieurs ions spécifiques dans un mélange complexe.

Certains spectromètres du type LTQ-Orbitrap (ThermoFinnigan) sont capables d'activer plusieurs masses pour fragmentation après isolation d'un précurseur (multistage activation, MSA, (Schroeder, Shabanowitz et al. 2004). Dans le cas des phosphopeptides, un nombre non négligeable de phosphopeptides donne un spectre de fragmentation où la perte de neutre -98 Da à partir du précurseur est prédominante et de faible valeur quant à la localisation du site de phosphorylation. Le spectre de fragmentation peut être enrichi lorsque la masse du précurseur moins 98 Da est co-activée avec celle du précurseur. Ne connaissant pas la charge du précurseur à l'avance, il convient de réaliser la MSA sur une perte de neutre de -98 Th (monochargé), -49 Th (dichargé) et -32.66 Th (tri-chargé).

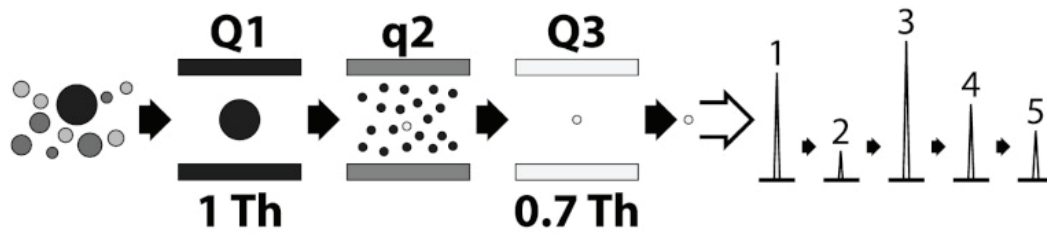
En règle générale les approches de type SRM/MRM apporteront la reproductibilité et la sensibilité manquante aux approches protéomiques « classiques » par « shotgun » (Peterson, Russell et al. 2012).

Dans le domaine de la protéomique quantitative, de très nombreuses études utilisant des approches QqQ SRM se sont développées pour des études quantitatives tant relatives qu'absolues.

1.2.4.5 Détection d'ions multiples en parallèle (PRM pour « Parallel Reaction Monitoring »)

Ici le schéma de sélection, fragmentation et analyse est semblable au précédent mais le troisième quadrupole est remplacé par un module à haute résolution et haute précision de type Orbitrap ou Q-ToF. Le but est de mettre à profit la haute précision et l'exactitude de masse pour des approches ciblées de type SRM (Peterson, Russell et al. 2012).

A SRM



B PRM

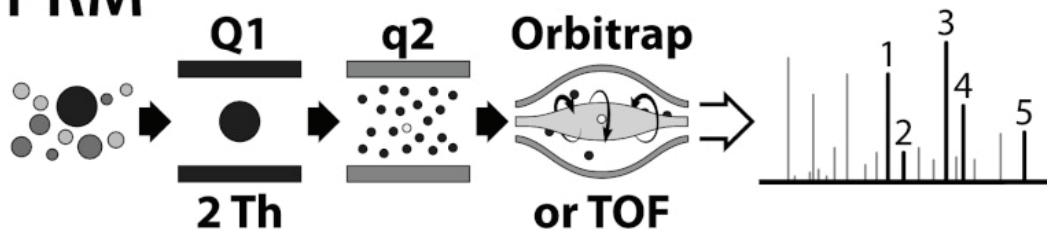


Figure 32 : Schéma d'un spectromètre de masse fonctionnant en détection d'ions multiples en parallèle (Peterson, Russell et al. 2012).

Cette technique présente plusieurs avantages potentiels sur l'approche traditionnelle SRM.

D'abord l'identité du peptide se fera à partir d'un plus grand nombre d'ions fragments déterminés avec une haute précision.

Par ailleurs cette approche présente une plus grande tolérance à des peptides présents dans le bruit de fond et qui ont été co-isolés avec le peptide d'intérêt. De par le fait que le peptide est identifié et quantifié par de nombreux ions, l'influence de ces « contaminants » est minimisée.

Enfin, l'approche PRM permettant de suivre toutes les transitions, il n'est pas nécessaire d'avoir une connaissance à priori de la transition recherchée et donc du travail préalable nécessaire dans les essais PRM pour rechercher cette cible (Peterson, Russell et al. 2012).

1.2.4.6 « SWATH-MS » méthode fonctionnant lors de l'acquisition des données en mode « data-independent ».

Les approches protéomiques de type « top-down » largement majoritaires ont en commun la préparation de l'échantillon à étudier. Cette préparation consiste toujours en la transformation des protéines de l'échantillon en ses peptides constitutifs par une digestion protéolytique.

Cette digestion sera souvent suivie d'une « décomplexification » de l'échantillon par diverses chromatographies liquides à haute pression.

Pour l'acquisition des données au niveau du spectromètre de masse, deux approches sont alors possibles :

- l'une dominante à ce jour, utilisée dans les stratégies de type « shotgun » est dite « data-dependent ». Les spectres de fragmentation MS2 obtenus le sont à partir d'ions précurseurs détectés et sélectionnés au cours d'un scan préalable (MS1). Ainsi lors du traitement des données les spectres MS1 et MS2 sont nécessairement assignés l'un à l'autre.
- Dans cette nouvelle approche SWATHTM-MS développée par ABSciex, le but est de dépasser les limitations des approches SRM. Ces dernières ne permettent de réaliser qu'un nombre limité (quelques milliers) d'analyses par run LC-MS/MS. On abandonne ici l'idée de provoquer une fragmentation soit à partir de la détection d'un peptide parent (méthode shotgun) soit de sa connaissance préalable (méthode SRM). Ici tout au long du cycle de chromatographie on fragmentera tous les peptides « sortants » sur des fenêtres d'isolation large (25Da) mais à grande vitesse (100 ms). L'exploitation de ces spectres de fragmentations complexes se fera par des algorithmes particuliers et en utilisant des bases de données dédiées de type SRM Atlas (Gillet, Navarro et al. 2012)

Tableau 7 : Comparaison des différentes approches MS pour l'analyse et la quantification des protéines (d'après Aebersold et Schubert 2014).

| | Protéomique de type Shotgun | Selected Reaction Monitoring (SRM) | SWATH-MS |
|--|------------------------------------|---|------------------------|
| Reproductibilité | + | ++++ | ++++ |
| Sensibilité | ++ | ++++ | +++ |
| Couverture du protéome | Large | Faible | Très large |
| Gamme dynamique (quantification) | 3 à 4 logs | 5 logs | 4 logs |
| Mycobacterium tuberculosis (4019 ORFs théo.) | 2100 protéines (1 run) | 2900 protéines (300 runs !!!) | 2400 protéines (1 run) |

1.3 La préparation de l'échantillon pour analyse des MPTs

Dans les études protéomiques, disposer d'appareils de masse performants aux techniques variées ne suffit pas face à la complexité du protéome. La recherche de modifications post-traductionnelles au sein de ce protéome requiert une maîtrise de la préparation de l'échantillon.

1.3.1 La complexité du protéome.

La complexité du protéome se retrouve dans le nombre de protéines différentes présentes dans une cellule à un moment donné mais surtout dans les différences de concentrations entre ces protéines.

Des protéines par milliers.... Pour un organisme entier de type eucaryote contenant 25 à 30 000 gènes, on peut estimer, en prenant en compte l'épissage alternatif et les modifications post-traductionnelles, que les protéines potentiellement exprimées seront comprises entre 500 000 et 1 000 000. Dans une cellule eucaryote on estime à 20 000 le nombre de protéines pouvant être exprimées. Alors que l'on peut estimer à un maximum de 2000 le nombre de protéines visibles sur un gel 2-D.

Des protéines si différentes... Ces protéines se distinguent en outre par des propriétés physico-chimiques très différentes, des poids moléculaires de 5 à 1000 kDa, des points isoélectriques de 2,5 à 12,5, des solubilités très variables (problème des protéines membranaires). Leur concentration enfin varie de façon extrême. Dans le sérum la différence de concentration entre l'albumine ou la Troponine T peut varier d'un facteur 10^{12} . Dans une cellule, cette différence reste encore très élevée : 10^9 entre l'actine et certains facteurs de transcription, alors que la dynamique de mesure des meilleures méthodes analytiques se situe autour de 10^4 . L'étude du protéome et du protéome humain en particulier reste un vrai défi.

Les modifications post-traductionnelles n'affectent souvent que partiellement une protéine. C'est très souvent le cas des phosphorylations, de certaines acétylations... Ces modifications peuvent se produire en diverses positions de la protéine ou n'être que transitoires. Voilà donc un défi qui ne pourra pas être résolu sans une approche particulière du protéome à étudier.

1.3.2 Des précautions élémentaires : « figer l'échantillon »

En biochimie clinique il est connu depuis longtemps que les échantillons biologiques sur lesquels les mesures se font, sont des échantillons susceptibles d'évolution, de profondes modifications. Aussi des dispositions toutes particulières doivent être prises pour « figer » l'état de ces échantillons (manipulation rapide des échantillons, congélation...). Il doit en être de même en protéomique.

La première démarche va consister à « figer » le protéome dans sa globalité et d'empêcher l'action des protéases par ajout d'inhibiteurs génériques de protéases et bien sûr une manipulation rapide des échantillons, suivie le plus souvent d'une congélation.

| Phosphatase | pH spectrum | Inhibitors |
|--|-------------|---|
| Alkaline phosphatases ⁶⁻⁸ | 8-11 | Bromotetramisole, ⁴ levamisole, vanadate, ⁵ imidazole ¹⁶ |
| Ser/Thr phosphatases ^{9,10} (i.e., PP1 and PP2A) | 7-8 | Okadaic acid; ⁷ Micro- cystin LR Cantharidin, Inhibitor-2 |
| Ser/Thr phosphatases ^{9,10} - Ion dependent (i.e., PP2B and PP2C) | 7-8 | EDTA, Okadaic acid, cyclosporin, FK-506 ¹⁸ |
| Tyrosine phosphatases ¹² | 7-8 | Vanadate, ⁸ peroxovana- dium compounds ¹⁹ |
| Dual specificity phosphatases ³ | | Vanadate ¹⁸ |

Figure 33 : Inhibiteurs de phosphatases proposés par Sigma Aldrich.

En ce sens, la recherche d'une MPT doit susciter d'autres questions. Cette MPT est-elle ou non réversible? Dans l'affirmative il convient de bloquer toute évolution de ce protéome. L'étude du phosphoprotéome est un exemple type. Comme décrit précédemment la

phosphorylation est une MPT réversible aussi pour étudier le phosphoprotéome, il conviendra dès l'extraction des protéines du protéome de bloquer l'action des phosphatases par le biais d'inhibiteurs tels ceux présentés dans la Figure33.

1.3.3 Le fractionnement du protéome

La complexité du protéome telle que décrite dans le paragraphe précédent va nécessiter une action de simplification.

Ce fractionnement a pour but soit d'identifier le plus de protéines possibles mais aussi et surtout dans le cas de la recherche de MPT d'obtenir un recouvrement le plus complet possible des protéines identifiées.

Pour cela, il est fait appel aux techniques classiques de fractionnement ou de séparations des protéines ou des peptides.

1.3.3.1 Séparation sur gels mono ou bidimensionnels (Gels 1D et 2D)

L'électrophorèse en conditions dénaturantes sur gel de polyacrylamide contenant du dodécylsulfate de sodium (**gel SDS Page 1-D**) est certainement la méthode la plus simple à mettre en œuvre pour réduire la complexité de l'échantillon. La séparation qui va se faire selon le poids moléculaire, est de faible résolution.

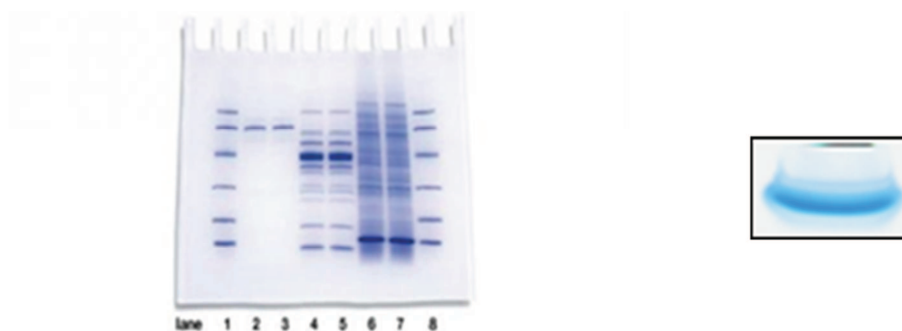


Figure 34 : Gels 1-D.

Permettant une simplification du protéome qui sur la présentation, image de gauche peut être séparé en une dizaine de fractions ou image de droite en une seule fraction pour laquelle on se débarrasse de beaucoup d'impuretés.

Le mélange de protéines est dissous dans un tampon contenant un réducteur des groupements thiol (mercapto-éthanol ou DTT) et du SDS (Sodium dodecyl sulfate) qui « chargera » la protéine négativement. Sous l'effet d'un champ électrique les protéines ainsi chargées migreront au travers des mailles se resserrant du gel de polyacrylamide et seront ralenties en fonction de leur taille et donc par voie de conséquence de leur masse. La séparation selon le poids moléculaire, de faible résolution, permet cependant une séparation simple d'un mélange de protéines moyennement complexe (Figure 34A). Le gel est ensuite découpé en bandes et les protéines contenues dans ces bandes seront digérées par une protéase avant d'être analysées par un spectromètre de masse de type MS/MS.

Dans son expression la plus simple, le gel 1-D peut simplement servir de moyen de séparation d'impuretés pour l'analyse protéomique par spectrométrie de masse. On se contentera alors, d'une migration très courte sur gel 1-D (Figure 34B).

Il est plus rare d'obtenir des informations à ce niveau concernant les modifications post-traductionnelles. Constatons cependant qu'une protéine phosphorylée peut voir sa migration perturbée en gel 1-D (Figure 35).

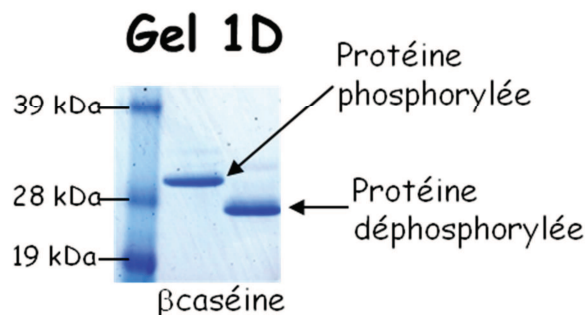


Figure 35 : Différence de migration entre une caséine phosphorylée et déphosphorylée.

L'électrophorèse bidimensionnelle ou gel 2-D. Ce fut dans les années 90, avec l'essor de la protéomique, la technique de fractionnement de choix du protéome.

Ici deux propriétés physico-chimiques des protéines sont utilisées : migration, dans la première dimension en fonction du point isoélectrique (pI), puis dans la seconde dimension en fonction de leur poids moléculaire (PM). Cette technique décrite par O'Farrell dès 1975 (O'Farrell 1975), n'a connu sa maturité technique qu'à la fin des années 90 grâce à la

commercialisation de bandes IPG de différents formats et des techniques de coloration et de détection mieux maîtrisées. Le pouvoir résolutif d'une telle technique est bien plus élevé que pour la simple approche par gel 1D. La position occupée par une protéine sur ce type de gel renseigne à la fois sur le poids moléculaire et le point isoélectrique de la protéine, information souvent précieuse pour identifier avec plus de certitude une protéine, Figure 36. Les spots visualisés par différentes techniques de coloration seront excisés, traités avec une protéase et les peptides extraits seront analysés par MS/MS afin d'identifier la ou les protéines présentes dans ce spot.

Dans les meilleurs cas, 1000 à 2000 protéines sont identifiables sur un gel 2D. Il est possible ainsi de séparer, identifier et quantifier plusieurs milliers de protéines à moins de 1 ng par spot (Klose 1999; Gorg, Weiss et al. 2004). Pour autant, ceci ne constitue qu'une faible fraction pour de nombreux protéomes. Par ailleurs cette technique a des limitations liées à la gamme de pI de 3 à 10, excluant de ce fait une bonne résolution pour les protéines basiques telles que les histones. Des limitations liées au PM des protéines existent également. Les protéines ayant un PM inférieur à 10 kDa ne sont pas présentes sur le gel et au-delà de 200kDa la résolution est très faible. Enfin les protéines membranaires difficilement solubles n'apparaîtront pas sur de tels gels (Santoni, Molloy et al. 2000).

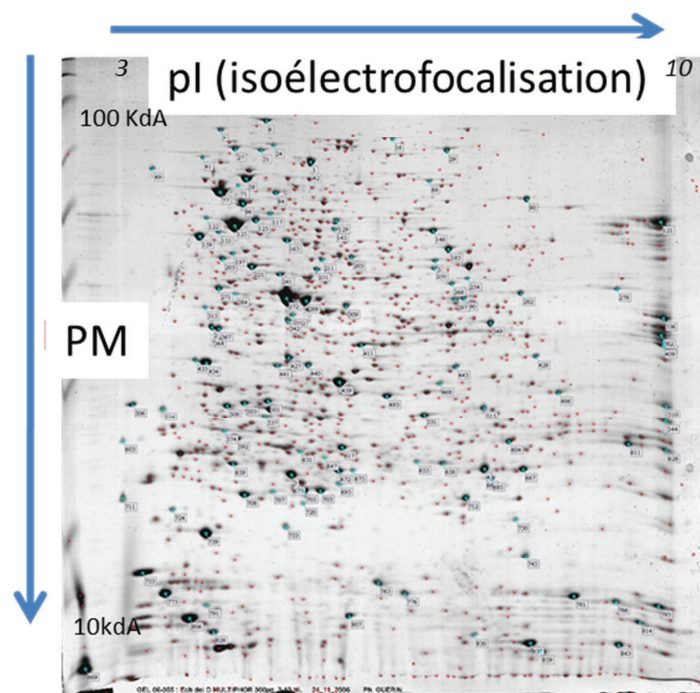


Figure 36 : Gel 2-D du protéome soluble de la bactérie radio résistante *Deinococcus deserti*. Coloration au bleu de coomassie

Ces limitations ont fait que vers les années 2005 les approches protéomiques par gel 2-D ont été supplantées par les approches shotgun.

Cependant bien plus que pour les gels 1-D, cette décomplexification peut nous renseigner sur la présence de certaines modifications post-traductionnelles au travers de cette image du protéome. En effet la séparation de la première dimension est basée sur le pI de la protéine qui dépend de la charge de la protéine à un pH donné. Certaines modifications post-traductionnelles (phosphorylations sur les sérines, thréonines et tyrosines ou acétylations du résidu N-terminal de la chaîne peptidique des protéines vont modifier cette charge (Rabilloud, Chevallet et al. 2010). Ainsi certaines protéines hyper phosphorylées apparaîtront sur un gel 2-D sous forme d'un chapelet.

Un autre exemple est donné dans la Figure 37 ci-dessous. Ce profil en chapelet est typique d'une protéine phosphorylée. La présence de groupements phosphate modifie la charge nette de la protéine et on observe un décalage des protéines phosphorylées vers les pI les moins élevés.

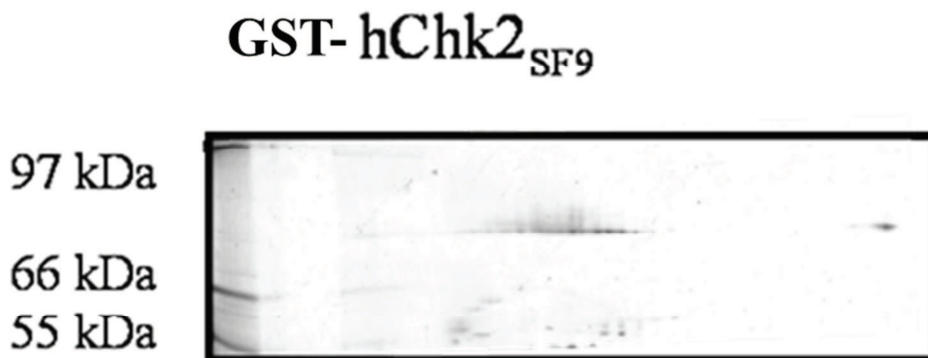


Figure 37 : Gel d'électrophorèse bidimensionnelle de la protéine recombinante 6His-Chk2 exprimée chez Baculovirus (Gabant, Lorphelin et al. 2008).

La quantité déposée est de 1 μg , coloration au nitrate d'argent.

D'une façon similaire sur un gel 2-D les changements importants de PM d'une protéine après glycosylation peuvent être observés, permettant ainsi de repérer ce type de MPT.

1.3.2.2 Séparation de type « Off-Gel »

Dans les deux cas décrits précédemment, la simplification du protéome se fait par un changement de phase. Les protéines sont extraites d'un protéome sous forme liquide. Le dépôt sur un gel 1-D ou 2-D pour simplifier ce protéome, oblige donc à extraire ensuite du gel les protéines ou le plus souvent les peptides de ces protéines générés par une digestion

protéique « in gel ». Le processus d'extraction se fait avec un rendement qui n'est pas total et entraîne donc une perte possible d'information et de matériel.

Pour pallier cet inconvénient, les protéines dans les approches « Off-gel » peuvent être séparées selon leur seul point isoélectrique tout en restant en solution. Le système se compose d'une chambre divisée en 12 ou 24 compartiments et reposant sur un gel contenant des ampholytes créant un gradient de pH. L'échantillon est déposé en quantité égale dans les compartiments de la chambre. Un couvercle est placé au-dessus de la chambre pour éviter toute évaporation. Une tension est appliquée aux extrémités du gel. Le système doit être refroidi car la résistance liée aux déplacements des charges électriques que sont les protéines ou les peptides dégage de la chaleur. Le déplacement s'effectue dans la zone de pH du gel équivalente au pI du peptide ou de la protéine (Ros, Faupel et al. 2002; Michel, Reymond et al. 2003).

L'avantage de cette méthode est que l'on peut ainsi faire migrer des protéines ou une solution de peptides qui n'aura subi qu'une unique digestion enzymatique (gage de reproductibilité). Les intervalles de pI peuvent être de seulement 0,1 unité de pH, conduisant à une séparation bien plus fine que sur un gel 2-D. Les quantités de protéines à mettre en œuvre sont bien plus faibles que pour un gel 2-D.

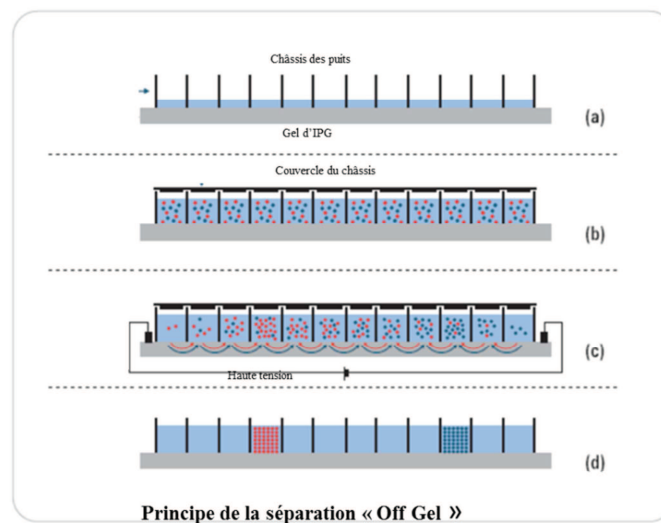


Figure 38 : Principe de la séparation « Off Gel » (Document Agilent).

Pour séparer les protéines suivant leur point isoélectrique, une enceinte contenant une bande de gel avec un gradient de pH est utilisée. Le gel est d'abord réhydraté avant que l'échantillon dilué dans la solution tampon soit distribué équitablement dans les puits situés au-dessus du gel. Une tension est alors appliquée et les protéines de l'échantillon migrent à travers le gel jusqu'à ce que le pH soit égal à leur pI . Les protéines sont alors séparées et en solution, ce qui facilite leur analyse ultérieure.

1.3.2.3. Approches chromatographiques

Le développement des techniques chromatographiques à haute pression au cours des vingt dernières années a permis de développer des méthodes de séparation d'un mélange complexe de protéines mais surtout de peptides qui ont pu être adoptées pour certaines approches protéomiques. On parlera alors de protéomique sans gel. Ces approches de protéomique sans gel, ont supplanté peu à peu les approches basées sur les gels 2-D.

Les techniques mises en œuvre qui sont le plus couramment utilisées, seront présentées.

1.3.2.3.1 Hydrophobicité et charge

Deux propriétés des peptides sont mises en avant pour pouvoir les séparer de façon résolutive. Pour des peptides il est évident que la chromatographie d'exclusion basée sur la taille des molécules est inadaptée, par contre les techniques basées sur l'hydrophobicité des peptides ou leur charge nette (positive ou négative) seront couramment mise à profit :

- chromatographie en phase inverse RP-LC (Reverse Phase Liquid Chromatography),
- chromatographie basée sur l'échange de cations, SCX (Strong Cation Exchange Chromatography), donc la charge positive nette des peptides,
- chromatographie basée sur l'échange d'anions, SAX (Strong Anion Exchange Chromatography), donc basée sur la charge négative nette des peptides.

Dans tous les cas, ces propriétés physico-chimiques seront mises en œuvre avec les technologies actuelles, c'est-à-dire des supports sous forme de microbilles de très faibles dimensions de quelques μm , offrant une surface d'échange très importante, obligeant à travailler à haute pression (quelques centaines de bars), mais avec de très faibles quantités de produit.

1.3.2.3.2 Couplage des techniques– chromatographie liquide multidimensionnelle

Les couplages entre des techniques chromatographiques et d'autres sont parfois nécessaires pour résoudre la complexité des protéomes. Aussi, plusieurs méthodes de séparation ont été ajoutées, créant le concept de chromatographie multidimensionnelle.

- **Séparation continue**, dès 2001 M. Washburn et al publient (Washburn, Wolters et al. 2001) une méthode de chromatographie multidimensionnelle en ligne, se plaçant en concurrence directe avec les approches d'étude du protéome par gel 2-D. Cette approche nommée MudPIT pour (Multidimensional Protein Identification Technology) va consister à séparer un mélange complexe de peptides préalablement digérés par une double séparation en ligne basée d'abord sur un échange cationique

fort (SCX) suivi d'une séparation en phase inverse. Ce système couplé à un appareil de masse en tandem, a permis d'identifier plus de 1484 protéines de *S. cerevisiae* (Washburn, Wolters et al. 2001). Même si les deux méthodes de séparation mises en ligne font appel à des solvants compatibles avec l'ionisation de type ESI, l'optimisation concomitante des deux méthodes de séparation très différentes est difficile à régler. Ceci a pour conséquence que les peptides, encore trop « nombreux » pour les capacités d'analyse des spectromètres de masse, ne sont pas nécessairement représentatifs de l'ensemble du protéome et les protéines les plus abondantes sont « surreprésentées ». Ce phénomène appelé échantillonnage au hasard ou sous-échantillonnage est analysé dans la publication de Liu et al (Liu, Sadygov et al. 2004). Pour des mélanges complexes de protéines, la reproductibilité de telles approches est moindre puisqu'il faut dans le cas de l'analyse du protéome de la levure ci-dessus indiqué, jusqu'à dix analyses du même échantillon sur le spectromètre de masse pour corriger partiellement ce biais. (Gevaert, Van Damme et al. 2007).

- **Séparation discontinue** : en parallèle des types de séparation en discontinu ont été proposées. Plus simples à mettre en œuvre, ils permettent une optimisation de chaque étape de façon indépendante. Demandant plus de travail dans la collecte de chaque échantillon ils apportent cependant une plus grande souplesse dans leur utilisation. Certaines fractions collectées au cours de la première dimension, où se trouveront concentrés des peptides d'intérêt, retiendront notre attention. Ces approches ne permettent pas toutefois une automatisation complète (Delahunty and Yates 2005).

1.3.2.3.3 Intérêt pour les MPT

Ces séparations discontinues ont trouvé un intérêt tout particulier dans la recherche de modifications post-traductionnelles.

Généralement, pour ces séparations orthogonales discontinues, la deuxième étape sera une séparation basée sur une colonne de type phase inverse, qui mettra à profit les différences d'hydrophobicité des peptides pour mieux les séparer mais qui permettra en outre de se débarrasser des sels présents encore dans la solution à analyser.

Pour la première étape, la chromatographie d'échange cationique (SCX) a été mise à profit dans les cas où la modification va provoquer un changement de charge du peptide modifié. Ce sera le cas des phosphorylations des sérines, thréonines et tyrosines pour lesquelles l'estérification par un groupement phosphoryle va apporter une charge négative qui va

modifier le pI du peptide et de la protéine à laquelle il appartient. Le fractionnement par chromatographie d'échange de cations (SCX) a été utilisé pour un pré-enrichissement des phosphopeptides (Beausoleil, Jedrychowski et al. 2004). A pH acide les phosphopeptides rendus moins « cationiques » par l'estérification par le groupement phosphoryle sortiront dans les premières fractions d'un gradient croissant en sel d'une telle chromatographie. Malgré l'usage de sel volatil, comme le formate d'ammonium, un dessalage sur un support de type phase inverse sera nécessaire pour les étapes ultérieures.

De même dans le cas des acétylations N-terminales, lors de l'acétylation de l'amine N-terminale chargée positivement, le peptide va perdre sa charge par ce processus d'acétylation. L'utilisation d'un fractionnement par SCX enrichira fortement les premières fractions en peptides acétylés sur le N-terminal.

Pour remplacer cette première étape, il a été proposé plus récemment une chromatographie d'interaction hydrophile (HILIC) à pH acide (McNulty and Annan 2008). Dans ce cas l'interaction entre les peptides et la phase stationnaire hydrophile neutre se fait via des liaisons hydrogène. Cette chromatographie exploite la forte polarité (hydrophilie) du groupement phosphate, les peptides phosphorylés seront ainsi « retardés », puis élués lorsque augmentera la polarité de la phase mobile. Cette méthode est présentée comme très complémentaire de la chromatographie en phase inverse, qui précède généralement l'analyse des mélanges par spectromètre de masse. Pour la recherche des phosphopeptides cette méthode est présentée par ses auteurs (McNulty and Annan 2008) comme équivalente au fractionnement SCX décrit ci-dessus.

1.3.2.4 Stratégies de préparation de l'échantillon pour capturer des MPT peu abondantes

L'étude de MPT présentes sur des protéines peu abondantes, protéines qui seront modifiées partiellement ou transitoirement, peut se révéler une entreprise complexe qui requiert une stratégie adaptée. Aussi les méthodes de fractionnement du protéome décrites ci-dessus peuvent ne pas être suffisantes pour avoir le panorama le plus complet du protéome modifié. Les peptides modifiés du protéome vont être extraits par des moyens plus spécifiques qui le plus souvent vont se rajouter aux méthodes précédentes.

1.3.2.4.1 Immunocapture par anticorps

L'utilisation des anticorps polyclonaux et plus encore monoclonaux a permis de reconnaître et de doser, dès les années 80, des molécules présentes en quantités infinitésimales, par des dosages de type ELISA (Voller, Bartlett et al. 1978). L'utilisation d'anticorps fixés à une

membrane solide pour capturer des molécules spécifiques dans un mélange complexe et ainsi les extraire pour mieux les identifier sera privilégiée.

En protéomique cette approche a été tentée pour la recherche de modifications post-traductionnelles. Les phosphorylations ont fait l'objet de nombreux essais (Gronborg, Kristiansen et al. 2002). Cependant les anticorps antiphosphotyrosine se sont montrés plus performants et spécifiques pour capturer les phosphotyrosines des peptides modifiés que les anticorps antiphospho-sérine ou -thréonine (Zoumaro-Djayoon, Heck et al. 2012).

Ceci est probablement dû au fait que le groupement phosphoryle fixé sur un résidu tyrosine est plus volumineux et immunogène quelle que soit la nature des acides aminés entourant la tyrosine. Ceci n'est pas le cas pour ce même groupement phosphoryle présent à l'extrémité d'une chaîne aliphatique courte (cas de la sérine et thréonine), conférant aux anticorps qui en sont issus une moindre spécificité.

De la même façon, les peptides dont le groupement amine ϵ -terminal des chaînes latérales des lysines est acétylé, peuvent être immunocapturés par des anticorps polyclonaux ou monoclonaux spécifiques (Choudhary, Kumar et al. 2009). Même si le problème de l'immunocapture n'est pas aussi délicat que pour les peptides phosphorylés sur une sérine et thréonine, l'utilisation de cocktails d'anticorps monoclonaux (Shaw, Chaerkady et al. 2011) peut permettre d'obtenir des peptides acétylés différents et complémentaires soulignant les spécificités différentes de ces divers anticorps. Une étude menée en 2009 (Zhang, Sprung et al. 2009) portant sur l'acétylome d'*Escherichia coli* mettait en évidence 138 sites d'acétylation sur 91 protéines. Quatre ans plus tard, une autre étude réalisée avec des anticorps différents (Zhang, Zheng et al. 2013) indiquait la découverte de 1070 sites d'acétylation sur 349 protéines différentes, retrouvant 75% (68) des protéines de l'étude précédente. Beaucoup d'acétylomes publiés au cours de ces dernières années y compris sur des organismes de type procaryote sont ainsi certainement loin d'être exhaustifs.

1.3.2.4.2 Affinité pour un métal immobilisé

L'extraction des peptides phosphorylés a donné lieu à de nombreux développements technologiques de par les limitations de l'extraction des peptides phosphorylés sur la sérine ou thréonine par des anticorps monoclonaux.

Basée sur l'affinité du groupement phosphorylé chargé négativement pour des cations métalliques, une chromatographie d'affinité a été mise au point. L'IMAC (Immobilized Metal Affinity Chromatography) met en œuvre des cations métalliques (Ga^{3+} , Fe^{3+} , Zr^{4+}) chélatés via des fonctions chimiques greffées sur des résines (Thingholm and Jensen 2009). Evidemment tous les peptides portant des charges négatives auront tendance à se fixer sur ce type de support (ceux contenant des acides carboxyliques portés par les acides aminés aspartiques ou glutamiques).

D'autres supports peuvent être utilisés comme des billes recouvertes d'oxyde de titane qui permettent d'absorber de façon spécifique des phosphates organiques solubles dans l'eau (Pinkse, Uitto et al. 2004). La méthode peut être optimisée si les colonnes d'oxyde de titane sont au préalable saturées par un composé acide de type dihydrobenzoïque (DHB) (Larsen, Thingholm et al. 2005), évitant ainsi la fixation des peptides non-phosphorylés.

Dans tous les cas les peptides phosphorylés retenus sur ces différents supports seront élués par des tampons alcalins (pH 10-11) de type hydroxyde d'ammonium (NH_4OH), de l'EDTA ou des tampons phosphates.

Il semble qu'il y ait une complémentarité entre ces deux variantes, l'IMAC retenant plus facilement les peptides polyphosphorylés et l'oxyde de titane les peptides mono phosphorylés, d'où une utilisation préconisée par certains des deux méthodes (Thingholm, Larsen et al. 2008).

1.3.2.4.3 Les lectines

Ces glycoprotéines, ayant la faculté de se lier de façon spécifique et réversible à certains glucides, sont utilisées pour mettre au point des stratégies d'affinité permettant d'isoler des glucides et donc par extension des glycopeptides présents dans un mélange complexe.

Dans le monde très complexe des glycoprotéines, les lectines sont un outil pour la capture et la caractérisation des O-glycoprotéines ou N-glycoprotéines. Si la plupart des lectines interagissent avec la partie terminale non-réductrice d'une chaîne d'oligosaccharides, certaines de ces lectines telle la concavalline A (ConA) ou l'agglutinine de germes de blé peuvent se lier aux motifs internes de cette chaîne d'oligosaccharides. Pour l'agglutinine de blé il s'agira des motifs α 2,3-liés à l'acide sialique (Vosseller, Trinidad et al. 2006). Pour la ConA le spectre d'affinité est plus large et concerne les résidus mannosyl et glycosyl de ces

chaînes d'oligosaccharides ayant cependant des groupements hydroxyles libres en position C3, C4 et C6 (Tang, Liu et al. 2010).

De plus la capacité des lectines à être couplées à différents supports et en particulier à des billes magnétiques rend leur usage aisé lorsqu'il s'agit de capturer des glycopeptides dans des mélanges complexes (Madera, Mann et al. 2008; Loo, Jones et al. 2010; Ferreira, Daniel-da-Silva et al. 2011).

1.4 Les approches protéomiques, du traditionnel gel 2D aux analyses shotgun globales et les conséquences pour la recherche des MPT

1.4.1 Les approches ascendantes dites « bottom-up »

Quelles que soient les variantes utilisées, le développement de la protéomique s'est fait essentiellement à partir d'approches ascendantes. Dans la grande majorité des cas on passera par une simplification du protéome et toujours par une digestion protéolytique des protéines constitutives de ce protéome. Ces deux étapes pouvant se faire dans l'ordre choisi par l'expérimentateur. Enfin l'identification des protéines se fera à partir de la connaissance de quelques-uns de leurs peptides constitutifs (Figure 39).

Les stratégies possibles

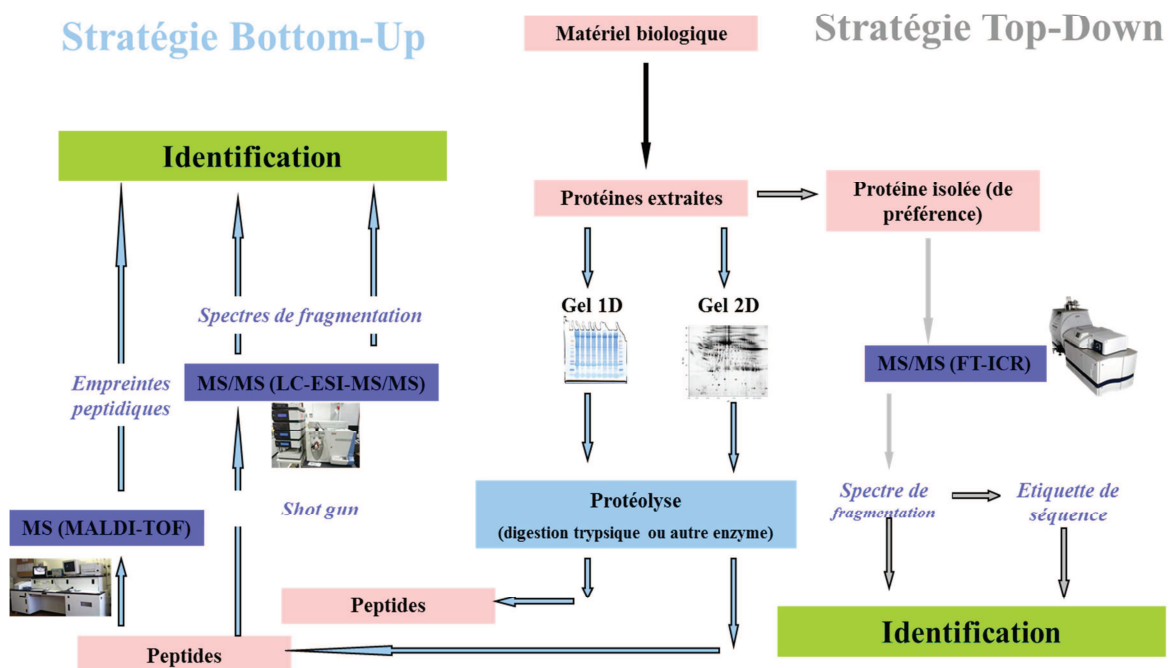


Figure 39 : Représentation schématiques des différentes stratégies mises en œuvre en protéomique.

Les stratégies mise en œuvre évoluent très rapidement en fonction des capacités des spectromètres de masse et des techniques de décomplexification du protéome. A ce jour les stratégies « Bottom-Up » ont été le plus développées.

1.4.1.1 Digestion trypsique.

Dans le cas d'approches ascendantes la digestion enzymatique est l'étape clef. La trypsine s'est imposée en protéomique comme l'enzyme protéolytique de choix, de par certaines de ses qualités propres :

- son faible poids moléculaire (24kDa), qui lui permet de pénétrer aisément dans les gels,
- sa spécificité (coupure après les résidus K et R), qui lui permet de générer de nombreux peptides de taille comprise entre 0,5 kDa et 4 kDa facilement analysables sur la plupart des spectromètres de masse,
- au cours de la digestion, l'autoprotéolyse de l'enzyme peut générer des pics sur lesquels on peut s'appuyer pour une calibration interne en Maldi-Tof. (842,51 – 1046 – et 2211 Th).
- Les résidus basiques, lysine et arginine, sont avec l'histidine des aminoacides susceptibles d'être protonés dans les conditions d'un électrospray (Covey, Huang et al. 1991). Ce qui conduira souvent à des peptides di-chargés facilement fragmentables et identifiables en spectrométrie de masse.

Cependant tant pour les besoins de la protéomique courante (besoin de mieux couvrir l'ensemble d'un protéome) que celui des recherches de modifications post-traductionnelles, il peut être nécessaire de digérer par d'autres enzymes le protéome ou la protéine isolée sur laquelle des MPT sont recherchées (Figure 40).

Dans le cas de la recherche de phosphorylations sur une protéine recombinante (ChK2), afin de couvrir le plus complètement possible la séquence primaire, deux digestions enzymatiques ont été réalisées par la trypsine et l'endoprotéinase GluC. Le cumul de ces digestions a permis de couvrir au total 81% de la séquence de la protéine et d'être renseigné sur le statut d'un plus grand nombre de S, T et Y pouvant être phosphorylées (Gabant, Lorphelin et al. 2008) (Figure : 39).

Il faut noter que le recours à une protéase autre que la trypsine peut être nécessaire lorsque la lysine sera sujette à une modification post-traductionnelle pouvant empêcher le clivage par la trypsine après cette lysine chimiquement modifiée. C'est notamment le cas pour les acétylations ou les glycosylations affectant une lysine.

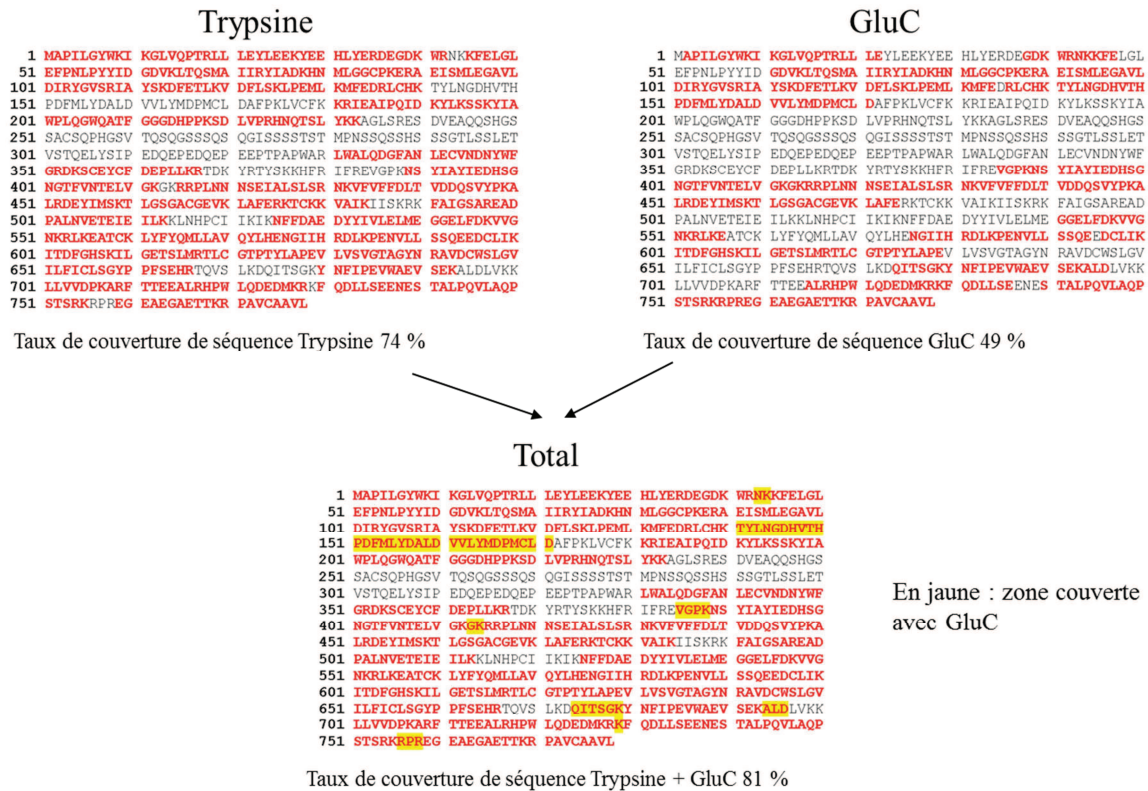


Figure 40 : Recouvrement de séquence plus complet de la protéine recombinante ChK2 par la combinaison de deux digestions enzymatiques.

| Enzyme | origine | Site de clivage | Exception | Gamme de pH | Type d'endo-peptidase |
|--|-----------------------------------|---|-----------|-------------|-----------------------|
| Trypsin (EC 3.4.21.4) | Pancréas de boeuf | Côté C-terminal de R-X, K-X | Si X=P | 7-9 | Sérine |
| Chymotrypsin (EC 3.4.21.1) | Pancréas de boeuf (Spéc Roche) | Côté C-terminal de F-X, Y-X, W-X, et à un taux plus faible L-X, I, M-X, A-X, D-X, E-X | Si X=P | 7-9 | Sérine |
| Endoprotéinase Glu-C ou protéase V-8 (V8-DE) (EC 3.4.21.19) | <i>Staphylococcus aureus</i> V8 | Côté C-terminal de E-X, D-X | Si X=P | 4-8 | Sérine |
| Endoprotéinase Lys-C (EC 3.4.21.50) | <i>Lysobacter enzymogenes</i> | Côté C-terminal de K-X | Si X=P | 8.5-8.8 | Sérine |
| Endoprotéinase Arg-C (EC 3.4.22.8) | <i>Clostridium histolyticum</i> | Côté C-terminal de R-X | Si X=P | 7-8 | Cystéine |

Figure 41 : Principales enzymes utilisées pour les digestions enzymatiques des protéomes solubles.

1.4.1.2 Analyse traditionnelle de protéomique par spectrométrie de masse – gels 2-D : empreinte peptidique

La recherche systématique de modifications post-traductionnelles de cette façon n'est pas la plus appropriée. En effet le principe de reconnaissance par empreinte peptidique (Henzel et al, 1993) suppose que sous le spot de gel 2-D se trouve une protéine unique ou à la rigueur un mélange simple avec une protéine en plus forte concentration. En outre convient-il que cette protéine soit répertoriée dans une banque de données !

Le spot d'intérêt est découpé, la protéine y sera digérée dans le gel le plus souvent avec de la trypsine. L'ensemble des m/z des peptides ainsi générés est stocké et comparé à l'ensemble des m/z générés (in silico) à partir de chaque protéine de la base de données. Un algorithme (Mascot Sequest...) va comparer cette liste de m/z expérimentales à celles obtenues « in silico » à partir de la base de données et établir un score statistique pour valider la combinaison la plus vraisemblable. Cette méthode simple et facile à mettre en œuvre suppose l'obtention d'un nombre suffisant de m/z de peptides correspondant à une même protéine. Dans le cas de modifications post-traductionnelles encore faut-il que dans la base de données se trouve répertorié le m/z correspondant au peptide modifié (ce qui était rarement le cas, dans les bases de données au début des années 2000).

Pour autant cette approche classique n'était pas sans intérêt pour la recherche de modifications post-traductionnelles. L'examen minutieux d'un gel 2-D peut laisser apparaître des chapelets de spots qui traduisent la présence de protéines modifiées selon la charge (donc potentiellement phosphorylées ou acétylées).

Par ailleurs des colorations spécifiques peuvent être utilisées pour mettre en évidence certaines de ces MPT. Des colorants fluorescents disponibles commercialement tels le ProQ Diamond (Molecular Probes) permettent de mettre en évidence des protéines phosphorylées préalablement séparées sur gel 2-D. Ce colorant qui se lie de façon spécifique aux groupements phosphate des sérines, thréonines et tyrosines, peut détecter des protéines présentes à moins de 4 ng par spot. L'intensité de la coloration est corrélée au nombre de groupements phosphate présents sur la molécule (Steinberg, Agnew et al. 2003).

De la même façon certaines glycoprotéines peuvent être détectées sur gel avec le colorant Pro-Q Emerald 300 (Steinberg, Pretty On Top et al. 2001). Ce composé peut réagir avec les glycols présents dans les glycoprotéines pour donner un conjugué fluorescent. Ces glycols sont au préalable oxydés en aldéhyde en utilisant l'acide periodique.

1.4.1.3. L'analyse des peptides par spectrométrie de masse en tandem

Dès 2001 cependant, Yates et ses collaborateurs (Washburn, Wolters et al. 2001; Wolters, Washburn et al. 2001) proposent une approche radicalement nouvelle basée sur une séparation multidimensionnelle d'un mélange complexe de peptides (provenant d'un protéome). Les peptides sont identifiés par spectrométrie de masse en tandem et la recherche s'effectue dans les bases de données par l'algorithme SEQUEST. Cette approche est appelée MudPIT pour (Multidimensional protein identification technology).

Cette stratégie qui s'est répandue dans les laboratoires au cours des années 2000 a totalement supplanté les approches de type empreintes peptidiques. Ceci d'autant plus que des spectromètres de masse à haute résolution et haut débit (de type Orbitrap) sont apparus sur le marché.

Ici la démarche est différente de celle visant à déterminer une empreinte peptidique puisque l'on travaille au niveau d'un peptide individuel et de sa fragmentation. Ce sont les travaux préliminaires de J. Yates et J. Eng (Eng, McCormack et al. 1994) qui ont ouvert la voie. L'algorithme SEQUEST qui fut le premier utilisé compare un spectre MS/MS expérimental aux spectres théoriques (in silico) MS/MS des peptides tryptiques (si la digestion a été effectuée par la trypsine) d'une base de données.

Le spectromètre de masse fonctionnera en mode d'acquisition dépendant des données (mode « data-dépendant »). Les ions précurseurs dont on mesure la masse de façon la plus précise possible sont sélectionnés et fragmentés. On peut au cours de ces acquisitions choisir les ions les plus intenses du scan MS en cours. Ces ions sont alors sélectionnés, isolés et fragmentés ; Une fois ce cycle terminé, grâce au paramétrage d'exclusion dynamique, on peut éventuellement, ne plus sélectionner ces ions, pour s'intéresser à d'autres ions.

Les moteurs de recherche de type Mascot qui vont permettre d'interpréter ces spectres sont eux aussi paramétrables. On peut typiquement choisir, la base de données, l'enzyme de digestion et éventuellement les modifications qu'auraient subies ces peptides.

On comprend là tout l'intérêt de cette approche pour la recherche de modifications post-traductionnelles.

Cette approche aléatoire « shotgun » a été appliquée dès 2001 par l'équipe de Yates sur le protéome de la levure (Washburn, Wolters et al. 2001) et plus récemment complétée par l'équipe de M. Mann (Thakur, Geiger et al. 2011). Dans cette dernière étude faite sans préfractionnement préalable on est passé de l'identification de 1484 protéines en 2001 à 2990 protéines en 2011 correspondant à 35 155 séquences uniques de peptides et un taux de couverture moyen des séquences de 18%. Cette étude reste remarquable par la gamme dynamique mise en évidence entre les protéines les plus faiblement et les plus fortement concentrées. Le faible taux de couverture des séquences protéiques montre toutefois le chemin encore à parcourir pour avoir une vision complète du protéome sur ces protéines détectées et en particulier des modifications post-traductionnelles.

Aussi les approches ciblées, telles que décrites dans les paragraphes précédents restent nécessaires pour avoir une vision plus complète des protéomes modifiés. Ainsi l'étude du phosphoprotéome doit faire appel à des stratégies où il vaut mieux combiner toutes les approches que nous avons précédemment décrites, comme l'enrichissement préalable des phosphopeptides (avec l'oxyde de titane par exemple), l'utilisation de modes de fragmentation complémentaires CID/ETD/HCD (Boersema, Mohammed et al. 2009).

1.4.2 Les approches top-down

Ces approches radicalement différentes explorées par les équipes de Ge et al (Ge, Lawhorn et al. 2002) dès la fin des années 90, sont encore peu répandues même si elles paraissent à certains comme très prometteuses.

Ici c'est la protéine entière qui est ionisée selon des modes ECD ou ETD et sa masse analysée. Des fragments peuvent aussi être générés et analysés à l'aide de HRMS sur des FT-ICR ou des Orbitrap.

Outre le fait de devoir disposer d'appareils onéreux et de manquer de logiciels d'interprétation des spectres obtenus, la difficulté de l'entreprise réside ici dans le fait que la protéine avant d'être analysée doit être relativement pure et en quantité suffisante, de l'ordre de 150 à 250 fmol. Quelques études encore rares sont disponibles (Forbes, Patrie et al. 2004; Tran, Zamdborg et al. 2011). Pour autant dans un article récent (Gault, Malosse et al. 2014) il apparaît que la stratégie top-down (couplée à une « stratégie bottom-up » semble être l'approche de choix pour faire une étude exhaustive des modifications post-traductionnelles de la protéine Pile de *Neisseria meningitidis*. Par ailleurs de telles approches sont de plus en plus utilisées pour étudier les MPT chez les histones (Britton, Gonzales-Cope et al. 2011).

1.4.3 Les approches middle-down

Un compromis peut être trouvé en utilisant les approches intermédiaires dites « middle-down ». Dans ce cas les deux approches précédemment décrites sont combinées, la protéine partiellement (60%) ou plus complètement (>90%) purifiée est partiellement dégradée avec de la LysC (AspN ou GluC peuvent aussi être utilisées) générant des peptides de 3kDa à 10-20 kDa. Les peptides ainsi générés sont analysés sur des appareils de masse MS/MS à très haute résolution. Dès 2001 l'équipe de Kelleher (Forbes, Mazur et al. 2001) parvenait à couvrir 100% de la séquence d'une protéine de 150kDa, la Pyochelin synthetase PchE, obtenant ainsi de précieuses informations sur les modifications post-traductionnelles qu'elle porte. Plus récemment (Young, DiMaggio et al. 2009) ont proposé une analyse fouillée des MPT sur les histones humaines.

1.4.4 Les stratégies de quantification des protéines et MPTs

Détecter, identifier les modifications post-traductionnelles est une étape nécessaire mais non suffisante pour tenter de cerner le rôle que jouent les MPT dans le fonctionnement des cellules. Savoir si ces MPT sont induites ou réprimées par un stress, des conditions de culture différentes, savoir si dans des conditions différentes de culture, telle voie métabolique est

concernée par une stimulation ou une répression des MPT des protéines la caractérisant, peut donner d'importantes indications sur leurs rôles.

De façon très classique toutes les techniques de marquage utilisées en protéomique différentielle peuvent ici aussi être mises en œuvre (Aebersold and Mann 2003; Yan and Chen 2005).

1.4.4.1 SILAC (stable isotope labeling by amino acids in cell culture)

La technique SILAC est une méthode d'incorporation d'isotopes stables au cours de culture cellulaire. Cette incorporation peut se faire par l'addition de l'isotope ^{15}N dans le milieu de culture, ou bien l'addition d'isotopes stables insérés dans certain acides aminés (^3D -Leu ou $^{13}\text{C}_6$ -Arg). Dès 2002, Ong SE et al de l'équipe de Matthias Mann proposent un marquage différentiel pour deux conditions de marquage de cellules en culture (Ong, Kratchmarova et al. 2003), l'une se faisant avec la ^{12}C -Arg, l'autre avec la $^{13}\text{C}_6$ -Arg, permettant ainsi à partir des spectres de masse (MS) obtenus sur les peptides, d'évaluer de façon différentielle les quantités de protéines en présence. Il est ainsi possible de différencier de 2 à 5 conditions de culture.

1.4.4.2 iTRAQ (Isobaric tags for relative and absolute quantitation)

Il s'agit aussi d'une technique de marquage hors gel. Le marquage s'effectuera ici après l'extraction des protéines et leur digestion protéolytique. Nous avons alors un marquage avec des étiquettes isobariques, marquage qui se fera sur les extrémités N-terminale ou C-terminales des peptides générés. Les tags strictement isobariques permettent au cours de la séparation par nano-HPLC sur phase inverse de minimiser fortement les différences entre les peptides ainsi marqués de façon différentielle. Cependant, au cours de leur fragmentation (MS/MS), ils vont présenter des produits de fragmentation différents (m/z à 114, 115, 116 et 117). Des peptides identiques mais obtenus dans des conditions différentes seront sélectionnés comme un peptide unique en MS (du fait de leur même masse). Ils se distingueront par leur produit de fragmentation. Les ratios combinés des intensités des peptides de protéines permettront une quantification relative de cette protéine. La technique permet de comparer de 4 à 8 échantillons différents (Gan, Chong et al. 2007).

1.4.4.3 Label free (spectral count, XIC)

Même si les deux approches précédentes ont connu un certain succès et des développements complémentaires, elles peuvent être prises en défaut. Concernant l'approche SILAC, il peut être délicat d'introduire ce marquage d'isotopes stables par la culture cellulaire sur des cellules eucaryotes animales même si certaines avancées ont été proposées (Krüger M., et al 2008) (Zanivan, Krueger et al. 2012). Concernant l'approche iTRAQ, elle peut s'appliquer sans problème sur des cellules de mammifères, mais sera défailante, lorsque la recherche des MPT nécessitera l'utilisation d'enrichissements par des méthodes d'affinité (anticorps, IMAC) (Mueller, Brusniak et al. 2008).

L'alternative à ces limitations est l'utilisation des approches de quantification sans marquage (« label-free »). Ces approches seront soit basées sur le comptage (« spectral counting ») des spectres MS/MS obtenus pour un ion parent et censés refléter l'abondance de cet ion parent (Liu, Sadygov et al. 2004), ou bien sur l'intensité de l'ion parent mesuré au travers des spectres MS accumulés au cours de l'enregistrement (Bondarenko, Chelius et al. 2002) (Wang, Zhou et al. 2003). Différents algorithmes sont proposés permettant ces acquisitions (Nahnsen, Bielow et al. 2013). Des exemples récents de comptage spectral sont donnés dans la publication de Schilling B et al (Schilling, Rardin et al. 2012) appliqués à des protéines acétylées ou phosphorylées ou dans le travail de Tan F et al (Tan, Zhang et al. 2011) sur les modifications portées par des histones.

1.4.5 Interprétation des spectres et statistiques associées aux analyses étendues et aux MPT

Les études par analyse protéomique de type « shotgun » utilisées pour la recherche des modifications post-traductionnelles, génèrent des milliers de spectres. L'interprétation et la validation de ces spectres ne peut se faire sans outils bio-informatiques, les mêmes que pour l'identification des protéines. La partie la plus délicate est l'attribution correcte des spectres de fragmentation à une séquence peptidique contenue dans la base de données.

1.4.5.1 Les algorithmes permettant d'interpréter les spectres MS/MS. Deux séries de données sont essentielles, la masse du parent et le spectre de fragmentation de celui-ci stocké sous forme d'une liste des masses (m/z) et des intensités afférentes. L'algorithme ira chercher dans la base de données choisie pour l'interprétation des spectres, les parents correspondant à la masse obtenue à l'imprécision de mesure près. Aussi il est aisé de comprendre que la détermination de la masse du parent avec la plus grande précision sera un avantage décisif car elle va réduire le nombre des peptides candidats et donc lever des ambiguïtés (tri-méthylation et acétylation). Dans notre cas la séquence peptidique contenant la modification recherchée n'est pas nécessairement stockée dans la base de données. Seul le peptide non modifié sera présent. Aussi convient-il d'obtenir la masse du peptide modifié avec la plus grande précision possible (analyseur à haute résolution) et de renseigner le plus exactement possible l'incrément de masse apporté par la modification.

Chaque algorithme donne une évaluation de l'attribution proposée, au travers d'un score qui traduira la qualité de cette attribution. Deux types d'approches existent :

- D'une part, les algorithmes heuristiques qui cherchent à établir une corrélation entre le spectre expérimental et le spectre théorique. Le score est alors calculé à partir de la similarité entre les deux spectres comme **Sequest** (Eng, McCormack et al. 1994) ou bien **X!Tandem** (Craig and Beavis 2004).
- D'autre part les programmes probabilistes, comme par exemple **Mascot** (Perkins, Pappin et al. 1999) où la fonction déterminant le score prend en considération la probabilité que la séquence peptidique identifiée puisse être

obtenue par hasard. Une comparaison entre certains de ces algorithmes a été présentée par Kapp et al (Kapp, Schutz et al. 2005).

1.4.5.2 Paramétrages des moteurs de recherche, contraintes apportées par la recherche des MPT

Si lors du paramétrage de la base de données utilisée, le type d'ions fragments n'impose aucune contrainte particulière, la masse exacte du peptide supposé modifié devra être connue car de cette estimation va dépendre la levée de certaines ambiguïtés.

Les contraintes liées à la digestion enzymatique doivent être examinées avec plus d'attention. La trypsine largement utilisée clive spécifiquement la chaîne peptidique après une lysine ou une arginine, sauf si un résidu proline suit ces acides aminés. Par ailleurs des MPT courantes telles des acétylations ou des glycosylations sont portées par le groupement ϵ -amine des chaînes latérales des lysines. L'action de la trypsine peut en être altérée et il conviendra d'adapter le nombre de coupures défailtantes (« miss-cleavage ») au cas par cas. Il est bon cependant de ne pas augmenter ce nombre de façon inconsidérée, car ces degrés de liberté accordés peuvent produire des attributions douteuses.

1.4.6 Interprétation de spectres, éviter les faux-positifs ou les MPT « in vitro »

Si la spectrométrie de masse reste l'outil de choix pour mettre en évidence les MPT ou en découvrir de nouvelles, il convient cependant de rester prudent pour éviter les mauvaises attributions de spectres.

La protéomique haut débit qui consiste à accumuler des dizaines de milliers de spectres au cours d'une expérience a pour corollaire une interprétation automatique des spectres par des algorithmes. Pour autant il faut rester prudent car les mauvaises interprétations ou attributions ont des origines multiples.

La première source d'erreur est la manipulation de l'échantillon. Au cours de la préparation préalable à l'injection dans un spectromètre de masse, l'échantillon peut subir une modification chimique non voulue qui peut être interprétée comme une modification post-traductionnelle.

Pendant la préparation de l'échantillon, il convient d'éviter de modifier artificiellement les protéines afin de ne pas attribuer à une modification « in vivo » une modification créée in vitro et de se livrer à des interprétations erronées. A cet effet Gang Xing et al (Xing, Zhang et al. 2008) indiquent 4 modifications pouvant survenir couramment :

- éthylation des résidus aspartate et glutamate (+28 Da) par de l'éthanol,
- estérification des résidus aspartate et glutamate (+74 Da) par du glycérol,
- une perte de 19 Da sur les résidus de lysine (réarrangement moléculaire),
- une addition de 108 Da à cystéine (nature inconnue).

Ces modifications viennent s'ajouter à une liste déjà longue de modifications pouvant survenir pendant la manipulation des échantillons comme :

- les adduits d'acrylamide qui peuvent se former lors de l'utilisation de gels SDS-PAGE. Les cystéines peuvent via leur groupement thiol réagir lors de la migration dans le gel avec de l'acrylamide libre (non polymérisé) provoquant un incrément de masse de 71 Da (Haebel, Jensen et al. 1995).
- les adduits de β -mercaptoéthanol, toujours sur les chaînes latérales de cystéine, amenant à une augmentation de masse de 76 Da (Klarskov, Roecklin et al. 1994).
- l'oxydation sur quelques résidus tels que la méthionine mais aussi le tryptophane ou la cystéine (Stadtman and Levine 2003; Perdivara, Deterding et al. 2010).
- des produits de dérivatisation apparaissant lors de l'utilisation de l'acide iodoacétique ou de l'iodoacétamine, dans le but de bloquer les résidus cystéine. Une dérivatisation de la méthionine a ainsi été observée (Lapko, Smith et al. 2000).
- la désamination des résidus asparagine et glutamine respectivement en acide aspartique et acide glutamique qui est très marquée à pH basique (>10). (Sarioglu, Lottspeich et al. 2000).
- la méthylation ou l'éthylation sur les chaînes latérales des acides aspartique et glutamique qui survient lorsque le méthanol ou l'éthanol sont utilisés dans les tampons pour colorer ou décolorer après une migration sur gel SDS-PAGE (Sprung, Chen et al. 2008).

Mais ces modifications artéfactuelles peuvent aussi être générées dans la cellule de collision d'un spectromètre de masse. Des réarrangements en phase gazeuse peuvent se produire conduisant à des produits nouveaux ne devant pas être confondus avec des MPT (Boys, Kuprowski et al. 2009).

Mais au-delà des précautions à prendre lors de la manipulation de l'échantillon, un score ionique élevé obtenu par un algorithme d'interprétation comme Mascot ne signe pas pour autant une identification juste d'une modification post-traductionnelle. (Chen, Zhang et al. 2009). Ces auteurs nous indiquent que de faux positifs peuvent apparaître suite à :

- une coupure enzymatique sur un site « anormal »,
- une mauvaise interprétation d'un état de charge,
- une mauvaise interprétation de la modification de la protéine,
- la mauvaise attribution d'un massif isotopique.

Si certains de ces cas peuvent se résoudre par l'utilisation d'un spectromètre de masse à haute résolution, d'autres n'échapperont pas à l'examen attentif du spectre MS/MS. Notons qu'en général la présence de pics notables non assignés doit conduire à s'interroger sur l'annotation proposée par le logiciel. Dans certains cas les logiciels de détection de massif isotopique de

type peptidique peuvent ne pas tenir compte de la présence d'un ion plus rare (par exemple l'iode) et conduire à de mauvaises interprétations de spectres, comme établi dans cette thèse, (Dedieu, Gaillard et al. 2011),

La confirmation par Western blot, sous réserve de posséder les anticorps spécifiques de la modification suspectée sera un élément supplémentaire pour prouver la présence de la modification. Cependant cela n'est pas toujours possible. Dans tous les cas outre l'analyse détaillée du spectre MS/MS, la confirmation d'une nouvelle modification post-traductionnelle se fera par l'analyse à haute résolution du peptide synthétique correspondant. Cependant dans quelques cas une telle approche peut être insuffisante et certains auteurs Zhang et al (Zhang, Chen et al. 2010) et Lee et al (Lee, Tan et al. 2013) propose de suivre une procédure de validation plus stricte qui outre l'analyse MS/MS du peptide synthétique correspondant, propose de bien vérifier la co-élution en HPLC des deux types de peptides et en cas de doute persistant d'effectuer des MS/MS/MS.

1.4.7 Base de données et modifications post-traductionnelles

La mise en évidence d'un nombre toujours plus grand de modifications post-traductionnelles a largement contribué au cours des vingt dernières années à l'explosion des données de protéomique (Khoury, Baliban et al. 2011). Aussi l'intégration de ces nouvelles connaissances s'est faite soit dans des bases de données généralistes, soit des bases de données spécialisées.

Concernant les bases de données généralistes, SWISS-PROT est certainement une base de référence de par la qualité de son annotation et des corrections apportées à ces annotations. En effet comme nous l'avons déjà souligné par ailleurs, l'affirmation de la présence d'une MPT sur un peptide reste malgré les progrès de la spectrométrie de masse une opération délicate et certaines données stockées peuvent être biaisées. dbMPT a été créée en 2006 (Lee, Huang et al. 2006) (<http://dbPTM.mbc.nctu.edu.tw/>.) dans le but de créer une base données contenant la vision la plus large possible des MPT reportées avec un certain niveau de confiance sur les protéines de cette base. dbPTM a été mise à jour en 2007.

Pour une protéine donnée on retrouve dans cette base la présence des modifications post-traductionnelles dûment prouvées (avec généralement une référence bibliographique), ou supposées être présentes, dans ce cas un commentaire (« potential », « probable » ou « by similarity ») est appliqué indiquant la probabilité de présence plus ou moins grande de cette modification.

Les initiatives sont nombreuses pour référencer dans des bases particulières les données des MPT. On retiendra cependant l'adresse suivante : http://www.expasy.org/proteomics/post-translational_modification d'Expasy (Expert Protein Analysis System) serveur de l'Institut Suisse de Bioinformatique.

En 2011 constatant que cette base n'avait pas été mise à jour, Khoury et al (Khoury, Baliban et al. 2011) proposent un outil (<http://selene.princeton.edu/PTMCuration>) qui permet une vérification et mise à jour des données de MPT dans la base SWISSProt.

On trouve des outils de prédictions concernant les MPT dans SWISS-PROT mais aussi sur d'autres sites comme le CBS (Center for Biological Sequence Analysis), qui propose à l'adresse suivante : <http://www.cbs.dtu.dk/databases/PTMpredictions/>) des outils de prédiction concernant les phosphorylations, glycosylations, la sulfatation, la myristoylation, la sumoylation ainsi que des outils prédictifs pour des sites de coupure spécifique.

La prédiction automatique des sites de MPT prend de plus en plus d'importance dans la communauté de la bioinformatique (Basu and Plewczynski 2010). Les outils actuellement disponibles sont basés :

- soit sur des recherches à partir d'approches textuelles. Dans ce premier groupe, se trouvent des outils tels ELM pour (« Eukaryotic Linear Motifs ») qui permet la recherche de motifs particuliers dans les séquences de protéines, soit des outils proposés par PROSITE qui cherche à prédire la présence de MPT à partir de séquences consensus.

- soit sur des recherches à partir de méthodes statistiques qui ont donné naissance à différents outils. Notons, les réseaux neuronaux artificiels (Adya et al, 1998) basés sur des modèles matriciels avec différents degrés de liberté, avec un programme très utilisé NetPhosK (Blom, Sicheritz-Ponten et al. 2004; Hjerrild, Stensballe et al. 2004) qui permet à l'utilisateur de choisir des niveaux de seuil pour valider la prédiction. On utilise aussi les « Support Vector Machine » (SVM), basées sur des statistiques linéaires ou polynomiales tel PredPhospho (Kim, Lee et al. 2004) qui permet la prédiction des sites de phosphorylation ainsi que le type de kinase qui agit sur chaque site.

I.5 Présentation du sujet de thèse

I.5.1 Contexte général du laboratoire d'accueil

Le laboratoire de Biochimie des Systèmes Perturbés (LBSP) au sein duquel ce travail de thèse a été conduit a développé une expertise en protéomique lors de l'analyse d'organismes radiotolérants (Zivanovic, Armengaud et al. 2009) (de Groot, Dulermo et al. 2009) (Baudet, Ortet et al. 2010) et un savoir-faire unique en annotation protéogénomique de microorganismes (Christie-Oleza, Pina-Villalonga et al. 2012; Armengaud, Hartmann et al. 2013). Des études de type shotgun ont été menées sur divers micro-organismes tels *Deinococcus deserti* (de Groot, Dulermo et al. 2009) ou *Ruegeria pomeroyi* (Christie-Oleza, Miotello et al. 2012) afin d'estimer le taux d'erreurs d'annotation de leur génome. Par ailleurs des travaux sur l'exo-protéome (Armengaud, Christie-Oleza et al. 2012) ont permis de caractériser les facteurs de virulence sécrétés par *Bacillus cereus* (Clair, Roussi et al. 2010) ou de révéler des stratégies d'adaptation différentes parmi douze bactéries marines du clade des Roseobacter (Christie-Oleza, Pina-Villalonga et al. 2012). Plus récemment, un savoir-faire a été acquis en N-terminomique dans le cadre plus général de la protéogénomique (Hartmann and Armengaud 2014). L'étude spécifique du N-terminome (par protéomique) permet de vérifier expérimentalement l'identification du codon d'initiation de la traduction et de certifier au niveau du génome, les données obtenues (Bland, Bellanger et al. 2014; Bland, Hartmann et al. 2014).

L'expertise du laboratoire repose sur la chimie des protéines, (marquage des extrémités N-terminales), la spectrométrie de masse et l'intégration de données (génomique comparée, carte protéogénomique, expression différentielle de protéines). L'analyse des modifications post-traductionnelles a été initiée dans le laboratoire par l'étude de la protéine phosphorylée, ChK2 (Gabant, Lorphelin et al. 2008) qui a révélé l'hyper phosphorylation de cette protéine et l'existence de 5 sites de phosphorylations communs aux protéines recombinantes produites par les cellules bactériennes et celles d'insectes. Ces sites apparaissent comme de potentiels régulateurs positifs de l'activité kinase de la protéine. Plus récemment, l'analyse de la phosphorylation de la NP protéine du virus Ebola a également été conduite (Peyrol, Thizon et al. 2013). Au-delà de ces études de cas spécifiques nous avons réfléchi à l'opportunité d'étude des MPT plus globales et proposé ce sujet de thèse pour l'analyse de nouveaux modèles biologiques.

I.5.2 Positionnement du sujet de la thèse

Comme nous l'avons souligné dans l'introduction les modifications post-traductionnelles que subissent les protéines après leur synthèse sont variées et fréquentes, elles ne se limitent pas comme il a longtemps été pensé aux seuls eucaryotes et confèrent des fonctions très particulières aux protéines qu'elles modifient. Au-delà des modifications post-traductionnelles, la coupure de la méthionine N-terminale, des peptides signaux, voire des coupures protéolytiques se produisant in vivo, vont nous renseigner sur le métabolisme de ces

organismes. Mais au-delà nous assistons depuis dix ans au moins à un changement de paradigme (Figure 42).

1.5.2.1 Un changement de paradigme chez les procaryotes !

Au cours des quinze dernières années notre vision des choses concernant les MPT chez les organismes vivants a été totalement bouleversée, notamment par l'étude de nouveaux modèles bactériens.

Certes dans un article paru dès 1984 Alain Cozzone indiquait qu'après une décade de controverses il semblait alors clair que non seulement les cellules eucaryotes mais aussi les procaryotes possédaient des protéines kinases capables de phosphoryler des protéines cytoplasmiques ou membranaires (Cozzone 1984).

Par la suite la problématique prit un tour différent et comme le soulignent Kennelly et Potts, la dichotomie avoir/ne pas avoir de phosphorylations entre eucaryotes/procaryotes s'est transformée en une interrogation sur leur localisation : pour les eucaryotes des phosphorylations sur les résidus alcools des sérines, thréonines et tyrosines et pour les procaryotes sur les histidines ou sur les groupements carboxyle des acides aminés acides (Kennelly and Potts 1996).

Cependant au milieu des années 90 lors de l'émergence de la protéomique, il n'était pas complètement admis que les modifications existaient en nombre et variété chez les procaryotes.

C'est donc grâce à la protéomique basée sur la spectrométrie de masse et aux énormes progrès techniques réalisés dès le début des années 2000 avec cette approche, que les protéomes et en particulier les protéomes d'organismes procaryotes ont pu être étudiés de façon plus approfondie. Ces travaux ont bouleversé la connaissance que l'on avait des modifications post-traductionnelles chez ces organismes. D'ailleurs les titres de certaines publications plein d'humour, reportés Figure 42, soulignaient ce changement de paradigme I. Benz et A. Schmidt dans une revue intitulée « Never say never again : protein glycosylation in pathogenic bacteria » parue dans *Molecular Microbiology* en 2002 (Benz and Schmidt 2002) dénonçaient « an almost dogmatic belief that prokaryotes are not able to synthesize glycoproteins » (« une croyance quasi dogmatique que les procaryotes ne sont pas capables de synthétiser des glycoprotéines ») et cependant cinq ans auparavant S. Moens et J. Vanderleyden soulignaient « rather recently it has become clear that prokaryotes (Archea and Bacteria) are able to glycosylate proteins » (Moens and Vanderleyden 1997). Il aura fallu dans ce domaine cinq années et l'apport de la spectrométrie de masse pour mieux asseoir cette vérité.

Dans une autre revue « fancy Meeting You Here ! a fresh Look at "Prokaryotic" Protein Phosphorylation » PJ Kennelly and M.Potts soulignaient dès 1996 que, si dès la fin des années 80 l'on reconnaissait que la phosphorylation était un phénomène universel présent aussi chez les procaryotes, la problématique fut déplacée sur le fait que seuls les eucaryotes étaient capables de phosphoryler des sérines, thréonines et tyrosines alors que les phosphorylations chez les procaryotes étaient censées se produire sur des histidines ou acides carboxyliques !

La phosphorylation, MPT emblématique car la plus étudiée, joue un rôle capital dans les régulations fines de maints processus biologiques des cellules. Désormais, il est acquis que les sérines et thréonines sont phosphorylées chez les procaryotes, que les protéines kinases et les

phosphatases des eucaryotes et procaryotes présentent de fortes similarités (Cozzone 1998). Au début des années 2000 il devient évident que la phosphorylation sur les tyrosines se rencontre chez un grand nombre de bactéries et apparaît même comme un processus ubiquitaire chez les procaryotes (Cozzone, Grangeasse et al. 2004; Grangeasse, Cozzone et al. 2007).

Dans une revue en 2005 J. Eichler et M. Adams (Eichler and Adams 2005) font un point sur les modifications post-traductionnelles alors mises en évidence chez les Archées. De nombreuses MPT existant chez les eucaryotes sont aujourd'hui répertoriées chez les procaryotes aussi :

- O- et N-glycosylation,
- Des phosphorylations sur les sérines, thréonines, tyrosines et histidine,
- Les lipides sont aussi présents sur les protéines des Archées, lipoprotéines portant en position N-terminale des motifs lipidiques, mais aussi acylation des chaînes latérales de certains acides aminés par des groupements myristoyle ou palmitoyle, des groupements isoprényles de différentes longueurs, voire des groupements aminoglycans liés à des phospholipides, des ancres GPI (GlycosylPhosphatidylInositol).
- Des méthylations,
- Des ponts disulfures.
- Des coupures au niveau des méthionines N-terminales.



Arch Microbiol (2004) 181 : 171–181
DOI 10.1007/s00203-003-0640-6

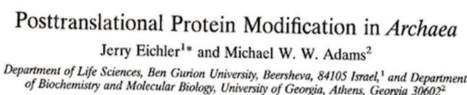
MINI-REVIEW

Alain J. Cozzone · Christophe Grangeasse
Patricia Doublet · Bertrand Duclos

Protein phosphorylation on tyrosine in bacteria



MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Sept. 2005, p. 393–425
1092-2172/05/090393-0 doi:10.1128/MEMBR.09.3.393-425.2005
Copyright © 2005, American Society for Microbiology. All Rights Reserved.



Cellular Microbiology (2009) 11(10), 1425–1434

doi:10.1111/j.1462-2020.01346.x
First published online 1 July 2009

Microreview

How microbes utilize host ubiquitination



Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology

Figure 42 : Titres de diverses publications indiquant ce changement de paradigme.

1.5.2.2 Maîtriser l'outil spectrométrie de masse

Etudier ces modifications suppose comme nous l'avons indiqué la maîtrise des outils de protéomique et au premier rang desquels le spectromètre de masse (de type MS/MS) à haut débit et haute précision. C'est ce que nous avons réalisé dans notre première approche pour l'étude sur organe (thyroïde de souris) des iodations portant sur la thyroglobuline, protéine largement majoritaire au sein de cet organe. Nous avons dû mettre à profit toutes les ressources du spectromètre de masse (Orbitrap XL) pour étudier de la façon la plus exhaustive qui soit les iodations des tyrosines ainsi que la formation des hormones T3 et T4.

Dans un deuxième temps, nous avons voulu appliquer ce savoir-faire à l'étude de modifications post-traductionnelles sur des micro-organismes radio-résistants, une bactérie *Deinococcus deserti* et une archée *Thermococcus gammatolerans*, pour y rechercher des modifications post-traductionnelles importantes (phosphorylations, acétylations N-terminales ou acétylation des amines des chaînes latérales des lysines).

Cette approche peut nous permettre de nous poser la question d'un rôle éventuel de ces modifications dans la radio-résistance mais aussi nous apporter des informations complémentaires dans l'approche protéogénomique de l'étude de ces procaryotes.

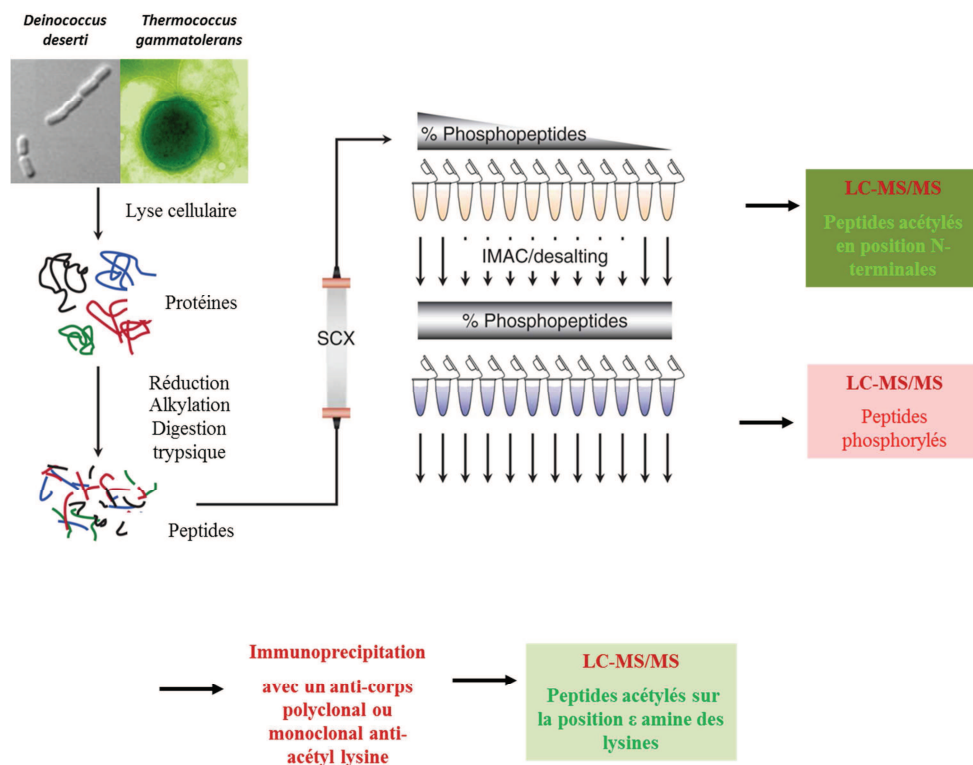


Figure 43 : Vue d'ensemble d'une stratégie optimisée.

Une telle stratégie optimisée peut permettre d'étudier ces trois modifications essentielles à partir d'une seule préparation de protéines solubles obtenues par lyse cellulaire et être systématiquement appliquée à divers microorganismes d'intérêt.

MATERIELS ET METHODES

2. MATERIELS ET METHODES

La plupart des informations relatives aux matériels et méthodes utilisés dans nos approches expérimentales ont été décrits dans les paragraphes correspondants des publications.

Cependant il nous paraît nécessaire de donner des informations complémentaires sur des matériels et méthodes utilisés dans l'ensemble de nos approches expérimentales.

2.1. Descriptif de l'Orbitrap XL

Le spectromètre de masse que nous avons utilisé pour la plus grande partie de nos études de protéomique est une LTQ-Orbitrap XL™ Thermo, livrée en 2008 dans notre laboratoire. C'est un appareil hybride constitué de plusieurs analyseurs. Des descriptions précises en ont été faites par son concepteur Alexander Makarov (Makarov, Denisov et al. 2009; Zubarev and Makarov 2013). Les deux parties essentielles sont une trappe ionique linéaire et un piège d'ions, appelé trappe Orbitale à transformées de Fourier. Ici l'analyseur à transformées de Fourier (basé sur l'utilisation d'un champ électrostatique, va apporter la haute précision en masse et la haute résolution sans utiliser d'aimant supraconducteur comme pour les appareils de type FTICR.

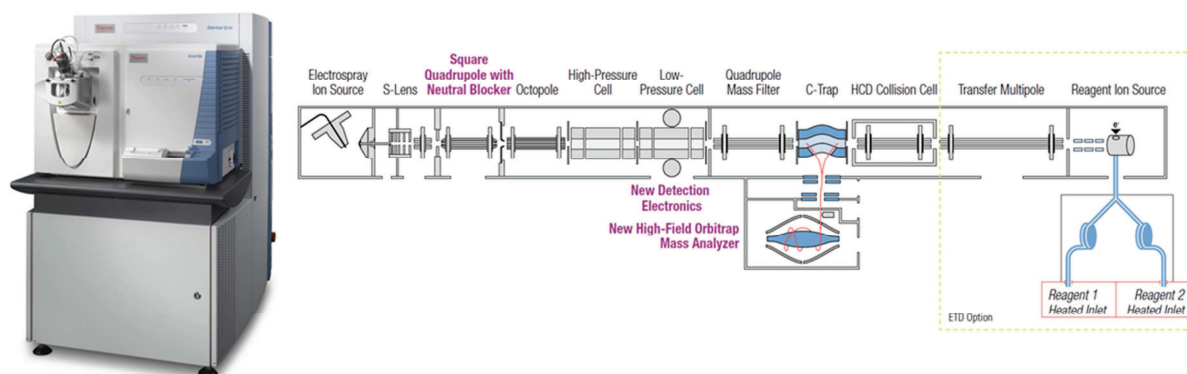


Figure 44 : Schéma d'une Orbitrap de type XL (ci-dessus) avec son module ETD (à droite). Photo d'une Orbitrap XL (Documentation Thermo Fisher)

La source ESI est montée sur une interface à pression atmosphérique, réglée en température, elle travaille à des débits de $1\mu\text{L}$ à $1000\mu\text{L}/\text{min}$ (débit optimal $20\text{-}400\mu\text{L}/\text{min}$), elle est ajustable sur deux axes X, Y, Z. Une source nanoESI, peut travailler soit en mode statique (infusion) soit en mode dynamique (couplage avec un système nano-LC). Sa gamme de débit va de $1\text{nL}/\text{min}$ à $50\mu\text{L}/\text{min}$.

Le premier analyseur de masse (LTQ) est une trappe ionique linéaire qui apporte une haute sensibilité, une gamme dynamique élevée, un temps de cycle rapide et la capacité de réaliser des analyses de types MS_n. La gamme de masse va de 15 à 4000 m/z (15-200 m/z, 50-2000

m/z, et 100-4000m/z). Le LTQ possède différents modes et vitesses de balayage permettant de faire un spectre MS complet, une analyse d'un ion spécifique (SIM) et des spectres MS/MS. La fragmentation se fait selon le mode CID (Dissociation induite par collision) grâce à un gaz (azote ou argon). L'acquisition automatique de spectres MS/MS, va se faire par l'exclusion dynamique qui permet de sélectionner les ions précurseurs et de choisir des ions de faible intensité. Au moyen de listes d'exclusion, on pourra exclure ou retenir des ions en fonction de leur intensité, leur état de charge, leur profil isotopique.

Le second analyseur, l'OrbitrapXL™ est un piège à champs électrostatique, chaque ion entrant dans ce piège aura autour de l'électrode chargée un mouvement de rotation et axial. Ce mouvement axial génère un courant induit enregistré par deux électrodes. Un calcul de Transformée de Fourier permet de calculer la fréquence axiale des ions, ω , dépendante du rapport m/z. On peut ainsi atteindre des résolutions très élevées et une excellente précision de masse.

2.2 Optimisation de la méthode d'acquisition

Les méthodes d'acquisition sont les méthodes paramétrables qui vont nous permettre d'acquérir le signal au sein du spectromètre de masse pour le mélange peptidique après sa décomplexification dans la colonne nano-HPLC.

Nous pouvons jouer sur le nombre de MS que nous souhaitons effectuer sur un même peptide, le seuil d'intensité à partir duquel cette mesure est effectuée, le temps d'exclusion pendant lequel ce peptide est exclu des mesures avant d'être à nouveau analysé, le nombre de MS/MS effectuées pour chaque mesure de ce peptide « parent » (MS). Mais également la durée du gradient sur des colonnes de 15 ou 50 cm sera un paramètre décisif dans la décomplexification du protéome. Enfin pour simuler un protéome plus ou moins complexe, deux niveaux de concentration seront choisis.

L'expérience d'optimisation a porté sur un digestat de Salmonella. *Ce travail a été effectué en interne par Oliver Pible, Jean Charles Gaillard, Guylaine Miotello.*

Tableau 8 : Paramètres d'optimisation pour les méthodes d'acquisition testées.

| Nom de la méthode | Réplicats de MS (peptide parent) | Seuil | Temps d'exclusion | Nombre de MS2 (sur peptide parent) | Gradient |
|------------------------|----------------------------------|-------|-------------------|------------------------------------|----------|
| Ref | 2 | 15000 | 60 | 3 | 90 |
| Excldynr | 1 | 10000 | 10 | 3 | 90 |
| Excldynr+7MS2 | 1 | 10000 | 10 | 7 | 90 |
| Excldynr+7MS2 +Grd 180 | 1 | 10000 | 10 | 7 | 180 |

Les résultats sont résumés dans les figures 45 et 46 pour la colonne de 50 cm et celle de 15 cm respectivement.

Le gain le plus significatif est apporté par la réduction du temps d'exclusion dynamique qui peut passer de 60s à 10s. On peut alors observer 10 à 20% de protéines supplémentaires identifiées avec la colonne de 50 cm et ce gain étant même porté à 37% avec la colonne de 15 cm.

Le gain le plus spectaculaire est obtenu avec un gradient porté à 180 min (au lieu de 90 mn). Le gain est ici porté à 66% pour les protéines identifiées avec deux peptides pour la colonne de 50 cm et à 75% avec la colonne de 15 cm, le nombre de spectres de peptides étant doublé quel que soit le type de colonne.

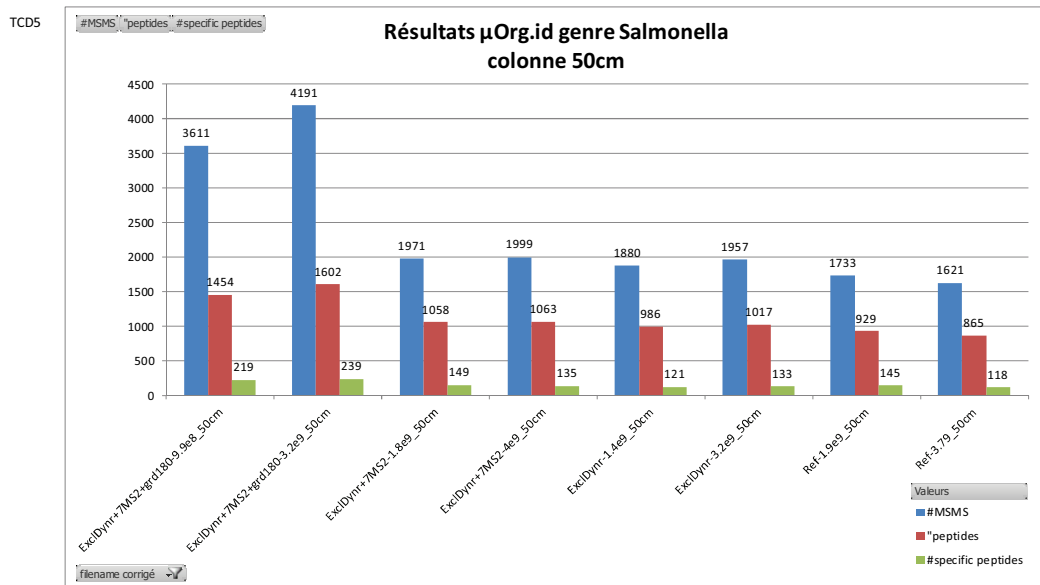


Figure 45 : Résultats présentés pour la colonne de 50 cm, nombre de MS/MS, de peptides et de peptides spécifiques.

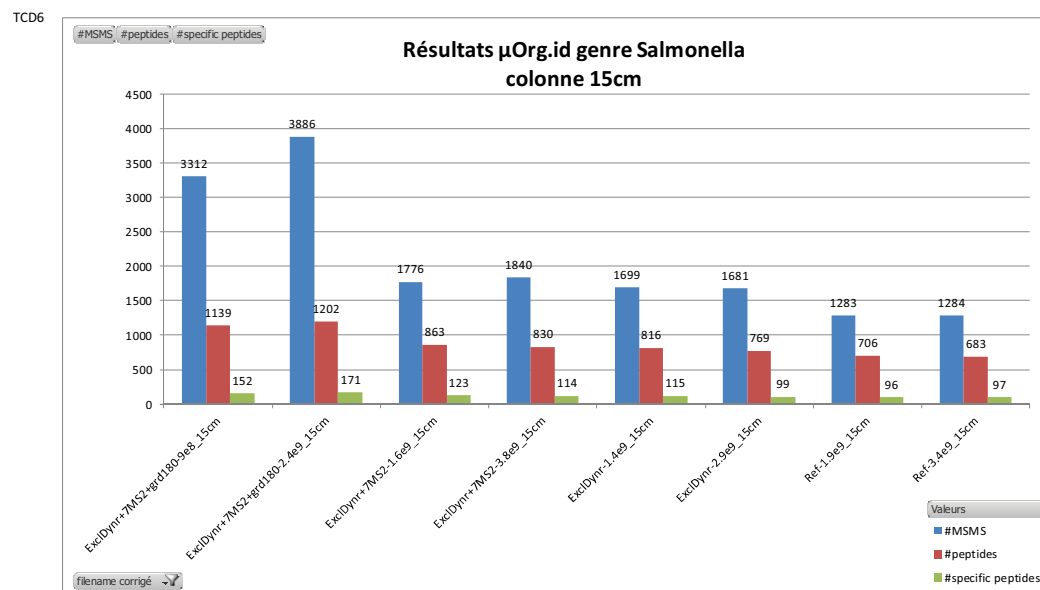


Figure 46 : Résultats présentés pour la colonne de 15 cm, nombre de MS/MS, de peptides et de peptides spécifiques.

2.3 Protocole de digestion

Dans les approches de protéomique descendante de type «Top Down», l'extraction du protéome que l'on veut étudier est toujours suivie d'une digestion enzymatique. Il s'agit là d'une étape clef qui doit être accomplie avec le maximum d'efficacité pour pouvoir identifier les protéines extraites. Aussi avons-nous prêté une attention particulière à ces protocoles de digestion et soit adopté les optimisations proposées par les industriels, soit optimisé nous-mêmes des protocoles pour l'étude de protéomes particuliers.

2.3.1 - Amélioration d'une digestion par la trypsine grâce au surfactant « Protease Max™ »

Comme nous l'avons indiqué précédemment, une décomplexification du protéome peut s'effectuer sur gel-1D ou gel-2D. La digestion enzymatique va alors se faire dans le gel. Le surfactant utilisé va permettre une meilleure digestion des protéines dans le gel (temps de digestion généralement raccourci) mais aussi une meilleure extraction des peptides du gel, étape cruciale, pouvant être cause de pertes de peptides (selon Promega, les peptides de grande taille sont plus facilement extraits).

Dans le protocole adopté au laboratoire, les étapes de décoloration, réduction, alkylation, ne sont pas modifiées.

Le détergent intervient dans la préparation de la solution enzymatique à hauteur de 1% en final en ayant pris soin de conserver au cours de la préparation la solution de ProteaseMAX™ à une température inférieure à 4°C (sur la glace). La solution d'enzyme utilisée sera au final à 10 µg/µL de trypsine dans 0,01% de ProteaseMAX™. Le temps d'incubation est réduit à 4 heures à 37°C.

L'extraction des peptides se fera de façon conventionnelle après arrêt de la réaction enzymatique avec une solution de TFA 5% (concentration finale 0,5%). La solution d'extraction le surfactant ProteaseMAX™ devra être impérativement rajoutée. Le temps total d'obtention des peptides est réduit à 6 heures (au lieu de 24h dans le protocole de digestion/extraction classique).

2.3.2 – Digestion de protéines thermorésistantes de *Thermococcus gamma tolerans*

D'autres optimisations peuvent se révéler nécessaires lorsque l'on veut digérer non plus en gel mais en solution un protéome extrait d'un organisme thermorésistant. C'est le cas pour *Thermococcus gammatolerans*, archaéa hyper thermophile et radiorésistante. Cette archaéa a une température de croissance optimale de 88°C et comme on peut le voir dans la figure 47, une digestion classique en milieu liquide se révèle très incomplète.

Nous avons optimisé le processus classique de digestion protéolytique en milieu liquide en jouant sur une double incubation soit avec de la trypsine, soit de la Lys-C et en utilisant des temps d'incubation longs. Dans tous les cas la digestion de ce protéome reste fort incomplète (Figure 47, gel A).

Nous avons dû nous inspirer des travaux de Lee AM et al (Lee, AM et al, 2009) pour digérer nos protéines thermorésistantes par une digestion trypsique (toute la nuit) précédée d'une hydrolyse en milieu acide assistée par micro-ondes. Les résultats représentés sur le gel B Figure 47, indiquent que l'utilisation du four micro-onde pour la dégradation en milieu acide d'un protéome thermo résistant est nécessaire. Seule l'acidification ne donne pas des résultats aussi complets.

Le protéome non thermorésistant de la bactérie *Deinococcus deserti*, bien que très particulier ne présente pas la même résistance à la digestion trypsique.

Protocole : Pour chaque échantillon, un volume correspondant à 8 mg a été prélevé et distribué dans un tube à centrifuger. Une solution de DTT (0,1M) a été ajoutée, pour obtenir une concentration finale de 10 mM. Les tubes ont été ensuite incubés 1 h à 37 ° C. Une solution aqueuse d'acide formique à 50% a été ajoutée à la concentration finale 6% (vol/vol). Les tubes ont été scellés et placés sur un flotteur dans un bain d'eau, puis, exposés dans un micro-onde domestique à 750 W d'énergie pendant 2,5 minutes. Les tubes sont ensuite retirés du bain-marie et placés sur de la glace. Un volume de 160 µL de tampon Tris (0,5M) a été ajouté dans chaque tube afin d'obtenir un pH situé entre 8,0 à 8,5. L'iodoacétamide (0,2 M) a été ajouté à raison de 10 µL par tube. L'incubation se poursuit 30 min à l'obscurité à température ambiante. La protéolyse est effectuée avec une solution de trypsine à 0,5 µg / µL dans du TFA 0,01%, (rapport de 2: 100 en poids). Après une nuit d'incubation à 37 ° C, la protéolyse est arrêtée en ajoutant une solution d'acide formique à 50%, ramenant le milieu à un pH de 4. Tous les tubes d'un même échantillon sont alors poolés. Le résultat de la protéolyse a été vérifié par SDS-PAGE réalisée sur un gel de gradient 4-12% Bis-Tris NuPAGE. Le digestat trypsique (360 mg de matière de départ) a été dessalé en utilisant un système Sep-Pak plus (cartouche C18) selon les préconisations du fabricant.

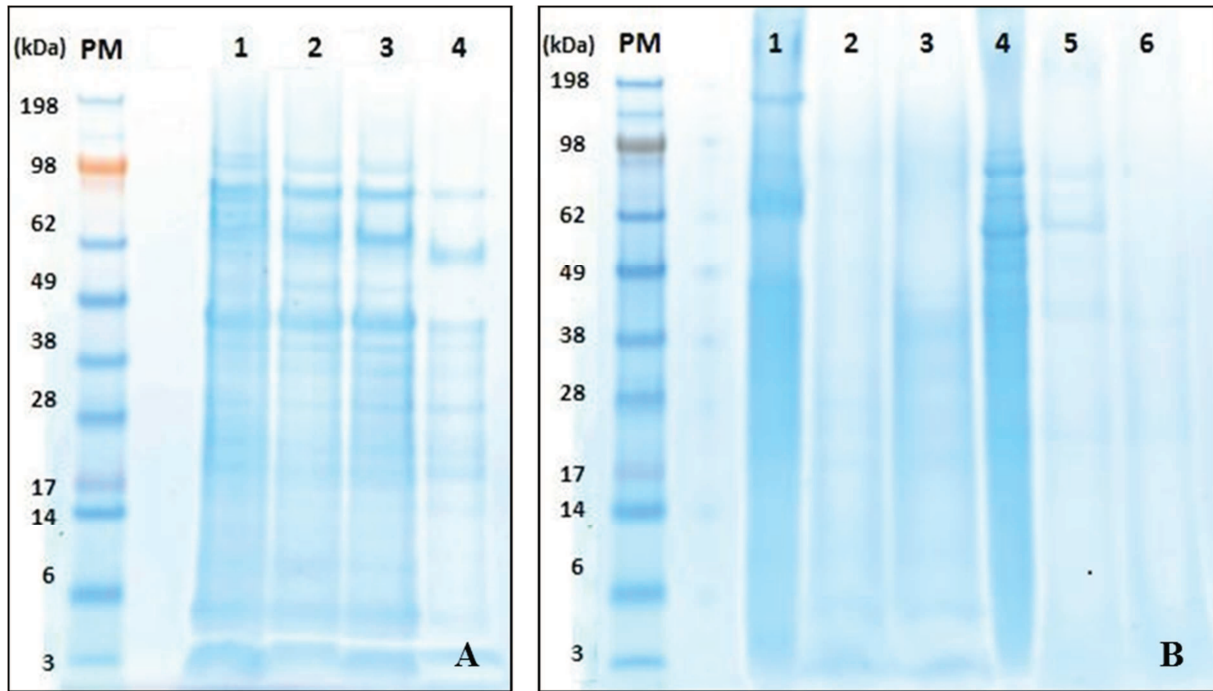


Figure 47 : Optimisation d'un protocole de digestion pour l'étude du protéome thermorésistant de *Thermococcus gammatolerans*. Les différents résultats de ces optimisations, sont évalués sur gel NuPage.

Gel A : Protocole de digestion double : (1 Protéome non digéré *Thermococcus gammatolerans*, (2) Digestion Lys-C 1h à 37°C, (3) Digestion par la trypsine 18h à 37°C puis ajout de 5µg de LysC pour 1h à 37°C (4) Digestion Trypsine 2% 5H à 37°C puis ajout de trypsine fraîche 2% à 37°C pour 5h.

Gel B : Protocole de digestion « assistée par micro-onde » (1) Protéome non digéré *Thermococcus gammatolerans*, (2) Digestion « assistée par micro-onde » en milieu acide (acide formique 6%) (3) Digestion classique avec traitement préalable à l'acide formique (6%) (4) Protéome non digéré *Deinococcus deserti*, (5) Digestion classique trypsique de *Deinococcus deserti*, (6) Digestion trypsique assistée par micro-onde de *Deinococcus deserti*

RESULTATS

3. RESULTATS

3.1 -L'approche protéomique SHOTGUN globale, un outil de choix pour l'étude des modifications post-traductionnelles

3.1.1 Ré-analyse des sites d'iodation de la thyroglobuline au travers d'une stratégie de type « shotgun » portant sur l'analyse d'un organe entier la thyroïde de souris

Dans ce premier travail, nous avons mis en œuvre toutes les ressources de l'approche protéomique par spectrométrie de masse pour obtenir la vision la plus complète des différents degrés d'iodation présents sur une thyroglobuline de souris. L'originalité de l'approche à consister à travailler directement sur l'organe (la thyroïde de souris) sans une purification préalable poussée de cette protéine largement majoritaire dans cet organe.

Les différentes digestions enzymatiques, la décomplexification partielle du protéome sur gel 1-D, des listes de travail dédiées, différents modes de fragmentations...ont été mis en œuvre pour obtenir une vision panoramique de cette protéine composée de deux sous-unités de 330 kDa.

Revisiting Iodination Sites in Thyroglobulin with an Organ-oriented Shotgun Strategy^[S]

Received for publication, June 30, 2010, and in revised form, October 26, 2010. Published, JBC Papers in Press, October 26, 2010, DOI 10.1074/jbc.M110.159483

Alain Dedieu^{†1}, Jean-Charles Gaillard[‡], Thierry Pourcher[§], Elisabeth Darrouzet^{¶2}, and Jean Armengaud^{‡2}

From the Commissariat à l'Énergie Atomique, DSV, iBEB, [‡]Laboratoire de Biochimie des Systèmes Perturbés and [§]Laboratoire des Transporteurs en Imagerie et Radiothérapie en Oncologie, Bagnols-sur-Cèze F-30207, France

Thyroglobulin (Tg) is secreted by thyroid epithelial cells. It is essential for thyroid hormonogenesis and iodine storage. Although studied for many years, only indirect and partial surveys of its post-translational modifications were reported. Here, we present a direct proteomic approach, used to study the degree of iodination of mouse Tg without any preliminary purification. A comprehensive coverage of Tg was obtained using a combination of different proteases, MS/MS fragmentation procedures with inclusion lists and a hybrid mass high-resolution LTQ-Orbitrap XL mass spectrometer. Although only 16 iodinated sites are currently known for human Tg, we uncovered 37 iodinated tyrosine residues, most of them being mono- or diiodinated. We report the specific isotopic pattern of thyroxine modification, not recognized as a normal peptide pattern. Four hormonogenic sites were detected. Two donor sites were identified through the detection of a pyruvic acid residue in place of the initial tyrosine. Evidence for polypeptide cleavages sites due to the action of cathepsins and dipeptidyl proteases in the thyroid were also detected. This work shows that semi-quantitation of Tg iodination states is feasible for human biopsies and should be of significant medical interest for further characterization of human thyroid pathologies.

Thyroglobulin (Tg)³ is one of the most abundant proteins produced by the thyroid gland and comprises two identical subunits of ~330 kDa. This prohormonal glycoprotein is secreted by thyroid epithelial cells. It is stored in the lumen of thyroid follicles in a highly condensed and covalently cross-linked form with numerous disulfide bridges. There, Tg is used as a scaffold for thyroid hormonogenesis and as an iodine reservoir. Many post-translational modifications are known to occur on this protein such as glycosylation, sulfation, and iodination (1). A signal peptide is processed during its export to the lumen. The most specific Tg post-translational modification is certainly its iodination and 3,5,3'-tri-

iodothyronine (T3) and thyroxine (T4) thyroid hormone synthesis. Indeed, through the combined action of NADPH oxidase and thyroperoxidase on the outer surface of the thyrocyte apical membrane, iodide ions that are also transported to the follicular lumen are covalently linked to some Tg tyrosyl residues to form mono- or diiodotyrosines (MIT or DIT, respectively). Coupling of two of these DIT residues then leads to the formation of T4 at an acceptor site, whereas coupling one MIT and one DIT moiety essentially forms T3 (2). In this reaction, the release of an iodotyrosyl moiety at the donor site leaves an "empty" side chain and the fate of this unusual residue is controversial. Some authors claim that a dehydroalanine is left instead of the tyrosine, whereas others observed pyruvic acid associated with cleavage of the polypeptide chain (3–7). Finally, the protein is proteolytically processed in the lumen, followed by endocytosis in lysosomes, to release the hormones. The principal proteases involved include the cathepsins (B, D, K, L, and S, with more or less specific roles ranging from solubilizing the highly cross-linked forms of Tg to releasing the hormone moiety), aminopeptidase N and dipeptidylpeptidases (8–12).

Studies of the post-translational modification of Tg began in the 1950s using radioactive iodide. Researchers attributed the radioactivity to different Tg fractions according to their sedimentation coefficient or molecular weight on SDS-PAGE gel. Later, Tg fragments obtained by proteolysis with trypsin or other proteases were separated by reversed-phase HPLC. The amount of radioactivity and the chemical form of the iodinated residues (MIT, DIT, T3, or T4) were then analyzed (13–15). Most of the studies were performed on purified Tg or Tg fragments from bovine, rat, guinea pig, rabbit, or human tissues, after *in vitro* chemical iodination or not. Significant progress was made in the 1980s when the first polypeptide sequence was reported for *Bos taurus* (16). The tyrosine residues involved were then identified more accurately using Edman amino acid sequencing and, later, mass spectrometry (MS) to identify the peptides. These experiments led to the establishment of a list of a number of iodinated tyrosine residues. Nonetheless, except for peptide identification, little has been done concerning MS-based characterization of tyrosine iodination. An initial attempt was made using purified (1237–1610) bovine Tg tryptic fragment (full sequence numbering). After endoproteinase Asp-N digestion, the resulting peptides were resolved by reversed-phase liquid chromatography and analyzed by electrospray and fast atom bombardment MS (6). The post-translational modifications of all seven tyrosine residues on this fragment were characterized at an unprece-

^[S] The on line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1–S2 and Tables S1–S8.

¹ To whom correspondence should be addressed: CEA Marcoule, DSV, iBEB, SBTN, LBSP, F-30207 Bagnols-sur-Cèze, France. Tel.: 00-33-4-66-79-68-13; Fax: 00-33-4-66-79-19-05; E-mail: alain.dedieu@cea.fr.

² Both authors contributed equally to this work as senior authors and should be considered co-last authors.

³ The abbreviations used are: Tg, thyroglobulin; MIT, monoiodotyrosine; DIT, diiodotyrosine; T3, triiodothyronine; T4, thyroxine; endo-GluC, endoproteinase GluC; HCD, higher energy collision dissociation; CID, collision-induced dissociation; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol.

Iodination Sites of Thyroglobulin

dented level of resolution, highlighting Tyr¹³¹⁰ as a new T3 and T4 acceptor site, and unveiling a dehydroalanine at position 1394 as a donor site residue. After proteolysis and reversed-phase separation as well, Dunn and co-workers (3) used MS to identify another donor site (Tyr¹⁴⁹) on bovine Tg with the presence of pyruvic acid. Finally, in a more technical study in 2005, Salek and Lehmann (17) tackled the problem of iodotyrosine identification. They showed that peptides containing a mono- or diiodotyrosine residue generated specific markers during collisional-induced dissociation (CID) fragmentation.

To study post-translational modifications of Tg in the thyroid and their heterogeneity, we proposed to analyze them using the most direct strategy possible, avoiding all purification steps and their inherent bias, and without the need for radioactivity. We dissected thyroid glands from mice and directly analyzed Tg using a shotgun-based MS/MS approach. We revealed as many tyrosine positions as possible by using a combination of different proteases and different MS/MS fragmentation procedures, involving analysis with inclusion lists, and using a high-resolution LTQ-Orbitrap XL hybrid mass spectrometer. We identified 37 modified tyrosine residues in mouse tissue, whereas only 16 sites are known for human Tg. Of these, we clearly identified 4 hormonogenic sites and two donor sites by identifying a pyruvic residue instead of the initial tyrosine. The analysis of cathepsin and dipeptidyl cleavage sites provided information of what happens in the cells. Our panoramic results bring new challenging data relative to Tg post-translational modifications.

EXPERIMENTAL PROCEDURES

Obtaining Tg from Mouse Thyroid Lobes—The study protocol was approved by the Committee for Animal Studies and was in accordance with the guidelines of the National Institute of Health principles of animal laboratory care (NIH publication 86-23, revised 1995). Three mice (mixed C57BL6/129Sv background) were sacrificed. Per animal, a thyroid lobe was removed and quickly cut into 6–8 small pieces. NuPage lithium dodecyl sulfate sample buffer (Invitrogen) supplemented with 15% β -mercaptoethanol was immediately added resulting in direct protein solubilization. The samples were incubated for 5 min at 99 °C. After 30 s of centrifugation, the soluble material was used for SDS-PAGE.

One-dimensional SDS-PAGE and In-gel Proteolysis—Proteins were resolved on NuPAGE Novex 4–12% BisTris gels (Invitrogen) run with MES buffer (Invitrogen). The gels were stained with Coomassie Blue Safe stain (Invitrogen). Each lane was cut into 2.5-mm slices from top to the bottom. Each sample was deposited in triplicate for subsequent proteolysis with trypsin, chymotrypsin, or GluC. For proteolysis, the protein bands were treated as previously described (18). For trypsin digestion, dry gel pieces were rehydrated with 12 ng/ml of trypsin (Roche Applied Science) in 25 mM NH_4HCO_3 (pH 8.5) containing 1% CaCl_2 and incubated overnight at 37 °C. Sequencing grade chymotrypsin (Roche Applied Science) was used as previously described (19). GluC (Roche Applied Science) was reconstituted in distilled water. After proteolysis, the peptides were extracted, first with 100% HCOOH, then

with 56% CH_3CN , 1% HCOOH, and finally with 100% CH_3CN . The resulting pools were dried completely in a vacuum and stored at -20 °C until needed for MS analysis.

Nano-LC-MS/MS Analysis—LC-MS/MS experiments were performed on a LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings). The peptide mixtures (0.1–2 pmol) were loaded and desalted online in an LC Packings Acclaim Pep-map 100 C18 reversed-phase precolumn (5 μm bead size, 100-Å pore size, 5 mm \times 300 μm). They were resolved on an LC Packings nanoscale Acclaim Pepmap 100 C18 (3 μm bead size, 100-Å pore size, 15 cm \times 75 μm) at a flow rate of 0.3 $\mu\text{l}/\text{min}$. The peptides were separated using a 90-min gradient (5 to 60% solvent B) with aqueous solvent A (0.1% HCOOH) and solvent B (0.1% HCOOH, 80% CH_3CN). The column was then washed for 10 min with 100% B and re-equilibrated for 20 min with 5% B. Full-scan mass spectra were measured from m/z 300 to 1800. The LTQ-Orbitrap XL mass spectrometer was run in data-dependent mode using TOP5 strategy with a Fourier transform mass spectrometer resolution set at 30,000 as previously reported (18, 20). In brief, a scan cycle was started with a full scan of high mass accuracy in the Orbitrap, followed by MS/MS scans in the linear ion trap on the 5 most abundant precursor ions with dynamic exclusion of previously selected ions. This dynamic exclusion consisted in two acquisitions of MS/MS spectra of the most abundant ion over 30 s and then excluding this ion for the fragmentations over the next 60 s. The activation type normally used was CID with a standard normalized collision energy set at 30. Alternatively the Higher Energy Collision Dissociation (HCD) mode was used. In this case, a TOP3 strategy was used with Fourier transform mass spectrometer resolution set at 30,000. MS/MS scans in the linear ion trap were then performed at 7,500 resolution and a normalized energy set at 35. In some experiments a specific parent mass list set-up was used to detect all the expected iodine post-translational modifications including dehydroalanine or pyruvic acid at donor sites. This parent mass list comprised 1335 values.

Polypeptide Database Mining—Peak lists were generated using Matrix Science MASCOT DAEMON software (version 2.2.2) from the ThermoFisher Xcalibur FT (version 2.0.7, extract_msn.exe) data import filter. Data import filter options were set at 400 (minimum mass), 5000 (maximum mass), 0 (grouping tolerance), 0 (intermediate scans), and 1000 (threshold). Using the MASCOT (2.2.06) search engine (Matrix Science), we searched all MS/MS spectra against a home-made mouse database built using the VARSPLIC algorithm (21). The VARSPLIC tool generates all natural variants or sequence conflicts described in the SwissProt data base (version 57.7). This database comprises 925,162 polypeptide sequences, totaling 1,244,593,373 amino acids and was used for producing supplemental Tables S1–S6. A restricted database was used to analyze post-translational modifications as explained under “Results.” It contains 1294 peptide sequences, totaling 630,319 amino acids. The list of proteins in this sub-database is shown in supplemental Table S7. In all cases, peptide searches were performed with the following parameters: a mass tolerance of 7 ppm on the parent ion and 0.5 Da for

Iodination Sites of Thyroglobulin

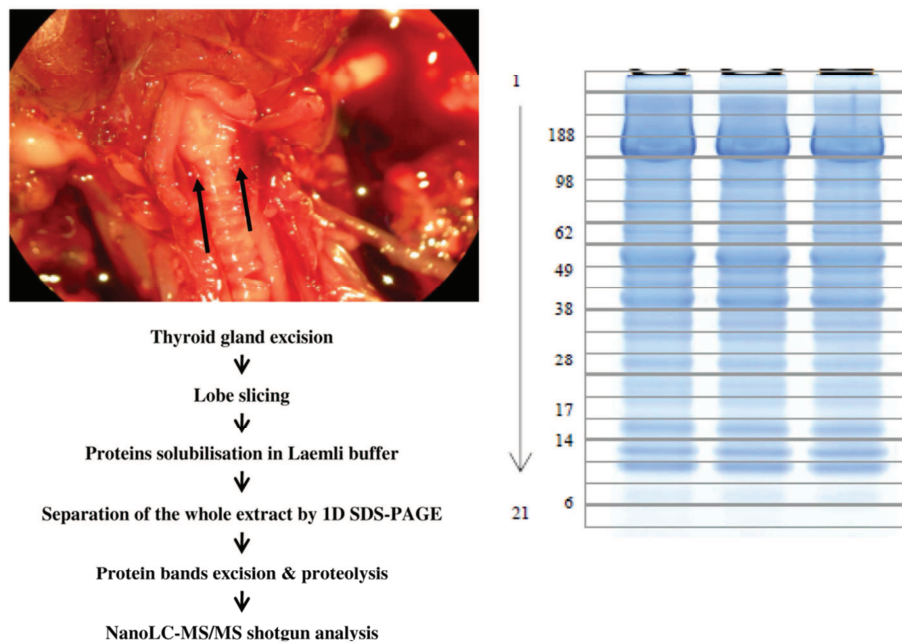


FIGURE 1. **Schematic illustration of analytical flow.** Proteomic analysis consists in direct extraction of thyroid lobes and high-throughput nano-LC ESI-MS/MS. In the dissected mouse shown on the left, the thyroid lobes are indicated with black arrows. The SDS-PAGE gel stained with Coomassie Blue Safe stain is shown on the right. The molecular masses in the first well corresponds to 188, 98, 62, 38, 28, 17, 14, and 6 kDa. The numbering of the gel slices (1–21) is also indicated.

the MS/MS, static modifications of carbamidomethylated Cys (+57.0215), dynamic modification of oxidized Met (+15.9949), tyrosine post-translational modification: +125.8966 for monoiodotyrosine; +251.7933 for diiodotyrosine; +469.7162 for triiodothyronine; +595.6128 for thyroxine; and –94.0419 for dehydroalanine. The maximum number of missed cleavages was set at 3 for the semi-chymotrypsin search and 2 for all other searches. All peptide matches with a Peptide Score above the peptide identity threshold set at $p < 0.01$ for the selected data base, and rank 1, were filtered using IRMa 1.22.4 software (22). The proteins were validated (supplemental Tables S2, S4, and S6) when at least two peptides were identified above the identity threshold. The false-positive rate for protein identification was estimated using the appropriate decoy data base below 0.7% in the worse case with these parameters.

MS Data Deposition—MS data were deposited in the PRIDE PRoteomics IDentifications data base (23) under accession numbers 12019 to 12032 (inclusive). They are freely available online.

RESULTS

Tg Is the Most Abundant Protein in Thyroid Extract—Fig. 1 shows the analytical flow chart we used to obtain the most comprehensive characterization of Tg iodinated sites based on MS. We removed a thyroid lobe from three mice. Each lobe was quickly cut into 6–8 pieces and dissolved directly in lithium dodecyl sulfate buffer. Then, total proteins were then resolved by one-dimensional SDS-PAGE and fractionated into 21 different slices on the basis of their molecular weight. In these samples, we first focused our attention on the four upper bands corresponding to proteins with molecular

masses above 250 kDa, where the intact Tg polypeptide was expected. Following trypsin proteolysis of these bands from the three thyroid samples, the proteins were identified using nano-LC-MS/MS shotgun by means of an LTQ-Orbitrap XL mass spectrometer. The 12 runs produced a dataset comprising 96,569 recorded MS/MS spectra. Besides the mouse O08710 Tg sequence, several variants are currently reported in the SwissProt data base. One natural variant was described with an L2283P mutation (24) causing congenital goiter in mice. The Q2NKY1 (isoform CRA_a) with the T1327A and I1721T double modification is reported in the NCBI data base (25). A third variant was described with a similar sequence to Q2NKY1 but with 13 additional amino acids at its N-terminal (26). Moreover, Tg encompasses a 20-amino acid peptide signal. We built a specific mouse protein database taking all the currently described variants as well as the 15 sequence conflicts indicated in the SwissProt data base into account for Tg. Using this in-house data base, we used the MASCOT search engine to assign a set of 1382 unique peptides in these fractions (supplemental Table S1), corresponding to 107 different proteins (supplemental Table S2). The presence of Tg in these fractions was demonstrated with 172 confident unique peptides, covering 59% of the protein sequence. The peptide(1305–1328), which is proteotypic of the Q2NKY1 sequence was detected in abundance in all three animals, whereas no proteotypic peptides for the other variants were found (supplemental Table S1). On this basis, we assumed that the three mice with mixed C57BL6/129Sv genetic background have the Q2NKY1 Tg. Through spectral counting, we estimated that Tg was the most abundant protein. Of the 21,044 assigned MS/MS spectra, three-quarters (15,935) are

Downloaded from http://www.jbc.org/ by guest on September 29, 2014

Iodination Sites of Thyroglobulin

Tg peptide signatures. This proportion was observed for the three animals analyzed (83, 65, and 79%). As Tg is quite an unusually huge protein, we further analyzed its abundance after normalizing the spectral count using the molecular weight of each polypeptide (supplemental Table S2). In this case, Tg is also far from the other identified polypeptides in terms of quantity (the second one is Myosin Q5SX39 identified with 148 unique peptides and 1265 MS/MS). Such abundance allowed us to further investigate Tg without the need for purification.

Following iodination in the thyroid lumen, the Tg was processed by cathepsins and exopeptidases. The various resulting fragments were resolved from intact Tg onto our SDS-PAGE. We analyzed the 17 other SDS-PAGE bands comprising proteins with molecular masses lower than 250 kDa for each of the three samples. Combining the results for all 21 slices, a total of 9935 unique peptides were confidently identified (supplemental Table S3). They correspond to the 1364 proteins listed in supplemental Table S4. In this dataset, we identified 174 different peptides corresponding to Tg fragments (supplemental Table S3). Once again, we did not detect any tryptic peptides proteotyping the other sequences apart from Q2NKY1 (supplemental Table S3).

It should be noted that all the bands corresponding to molecular masses below 250 kDa were found to contain Tg peptides. Regarding spectral counts, Tg fragments account for 29% of total proteins, whereas hemoglobin (HBB1), the second in order of importance, represents only about 8%. So Tg is really the most abundant protein in thyroid extract.

Comprehensive Coverage of Tg—To get a comprehensive coverage of intact mouse Tg, we used a multiplexed proteolytic digestion strategy. Samples from the three animals were re-applied to SDS-PAGE in duplicate to process the four upper bands containing intact and heavy fragments of Tg with chymotrypsin, and endo-GluC proteases. The 588 unique peptides in this MS/MS dataset are listed in supplemental Table S5. The proteins identified with these peptides are listed in supplemental Table S6. 172 different Tg peptides were previously found with trypsin, 250 were found with chymotrypsin, and 28 with endo-GluC. Chymotrypsin was found to be quite helpful in unveiling additional tyrosine residues because of its cleavage specificity (*i.e.* after aromatic and hydrophobic residues). On collecting all the results for the three proteolytic enzymes, we obtained 83% coverage of the Tg sequence. No tryptic peptides matching the Tg peptide signal described earlier (27) could be detected. This was expected, because the exported processed form accumulated in the thyroid follicle is present in significantly excessive quantities compared with the immature form. However, we confirmed that the mature sequence starts at N21, as we detected the 21–32 semi-chymotryptic, 21–39 semi-tryptic, and 21–40 semi-GluC peptides.

To search for Tg-specific post-translational modifications, as well as lumen or intracellular proteolytic events, we restricted the protein data base to the proteins detected in the extract (28). Supplemental Table S7 shows the compilation of proteins that were inserted into this subdatabase.

Identification of Iodination Sites by Pseudo-CID Fragmentation—Using the raw data sets obtained for all bands using all types of enzymes, we searched the protein subdatabase for specific tyrosine modifications in the Tg polypeptide: monoiodination, diiodination, triiodothyronine, and thyroxine modification. The 1356 unique Tg peptides identified in this search are listed in supplemental Table S8. Identification was highly redundant as these peptides account for 52285 MS/MS spectra. Of the 1356 peptides, 881 were unmodified and 184 comprised at least one iodination. Tyrosine iodination was found to be a very stable post-translational modification. Pseudo-CID fragmentation of these iodinated peptides obtained in the LTQ linear trap followed the usual fragmentation pathway of protonated peptides leading to informative *y* or *b* ion series. Remarkably, we observed the four expected iodinations on Tyr²⁵⁷² for the large peptide(2555–2577), whose sequence is ILAAAVWYYSLEHSTDDYAFSR. Fig. 2 shows four MS/MS spectra corresponding to these modifications at this specific position: monoiodotyrosine, diiodotyrosine, triiodotyrosine, and thyroxine. These spectra were observed, as expected, at four different retention times (43, 59, 67, and 69 min, respectively) because of the increase in hydrophobicity due to the successive addition of iodine atoms. As shown in these spectra, the *y* and *b* series are complete enough for unambiguously assigning the modifications to Tyr²⁵⁷² and not to one of the two other tyrosine residues at its N terminus. The *y*₅ and *y*₆ (red labels) in the first three spectra are clearly visible. Sometimes in the type D spectrum, the *y*₆ ion is not present but the presence of the doubly charged *y*₀₍₆₎²⁺ is noted. Furthermore, the *b*_s, *b*_o, or doubly charged *y*₍₁₄₎²⁺, *y*₍₁₅₎²⁺, and *y*₍₁₆₎²⁺ ions confirmed that Tyr²⁵⁶² and Tyr²⁵⁶³ are not modified in these peptides.

Regarding the thyroxine modification, we pointed out that the specificities of its isotopic pattern should be carefully taken into consideration when analyzing the spectra. With a high resolution mass spectrometer such as the LTQ-Orbitrap XL, monoisotopic mass, *i.e.* nomination of the correct peak of an isotopically resolved group of peptides peak as a monoisotopic peak (29), is determined using an algorithm based on the study reported by Senko *et al.* (30). These authors proposed the concept of a virtual amino acid, Averagine, constructed using the statistical occurrences of amino acids in the PIR protein data base (31). Such an Averagine contains the normal isotope variants of carbon (C), hydrogen (H), nitrogen (N), oxygen (O), and sulfur (S). This Averagine can then be used to obtain an isotopic pattern for any peptide, with precise masses and typical relative heights. The iodine atom (53, its atomic mass is 126.904) is inadequately present in the PIR data base and was simply neglected in this algorithm. In our study, the four iodine atoms in the thyroxine modification particularly influence the relative heights of the peaks in the isotopic pattern. Fig. 3 shows the experimental isotopic pattern recorded for peptide(2555–2577) containing a thyroxine modification at Tyr²⁵⁷² exhibiting a theoretical monoisotopic molecular mass of 1087.6808, as well as the relative heights of the first isotopes of an “Averagine peptide” (*black line*) with a closely related molecular mass of 1087.5869. This comparison shows a clear difference between both profiles. We therefore

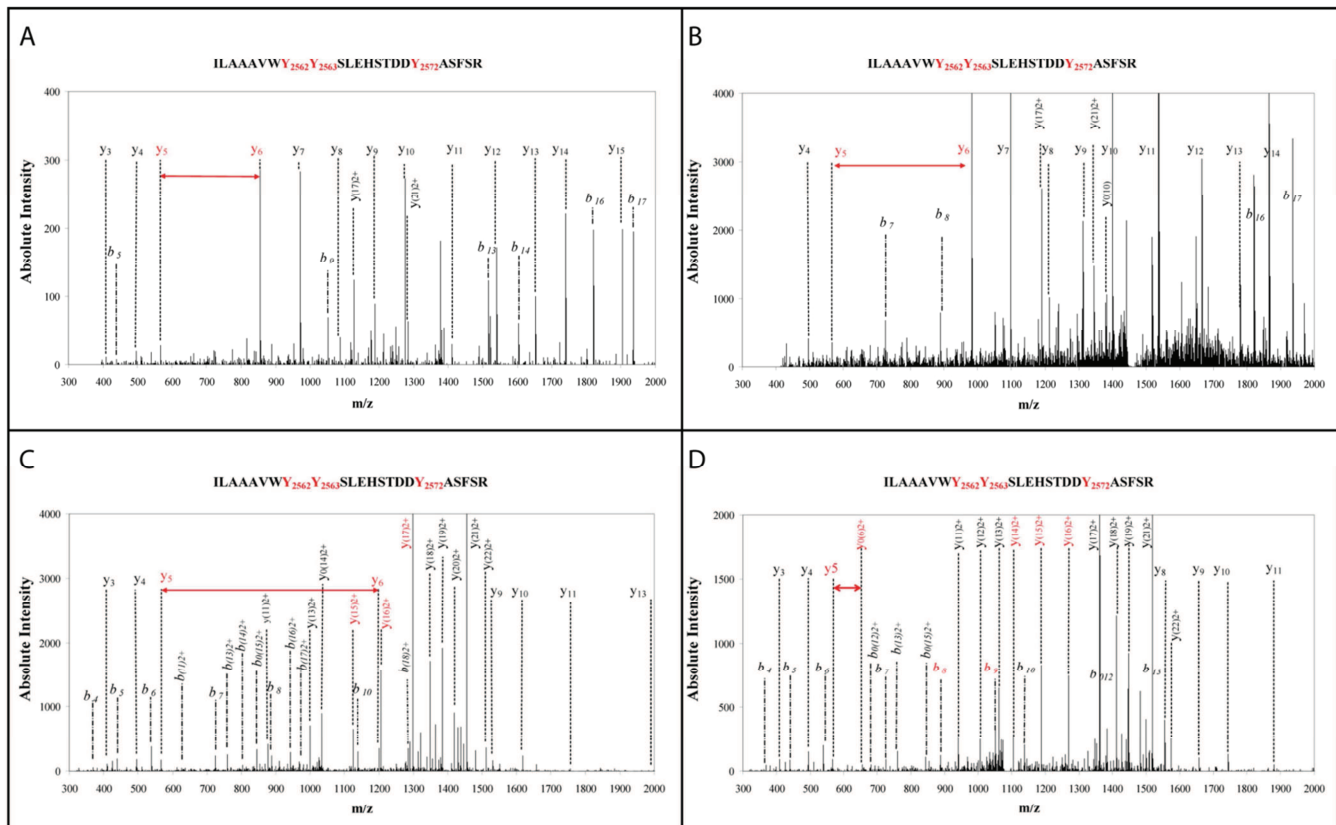


FIGURE 2. Unambiguous assignment of four different iodination patterns on Tyr²⁵⁷². Representative ESI CID spectra obtained for Tg peptide(2555–2577) (ILAAAVWY²⁵⁶²Y²⁵⁶⁶SLEHSTDDY²⁵⁷²ASF²⁵⁷⁷R) are presented. Monoiodotyrosine (panel A), diiodotyrosine (panel B), triiodothyronine (panel C), and thyroxine (panel D) modifications on Tyr²⁵⁷² were observed. In these spectra, information from the y and b series was complete enough to assign the modifications to Tyr²⁵⁷² unambiguously. Note the presence of y₅ and y₆ ions (red labels) in all spectra.

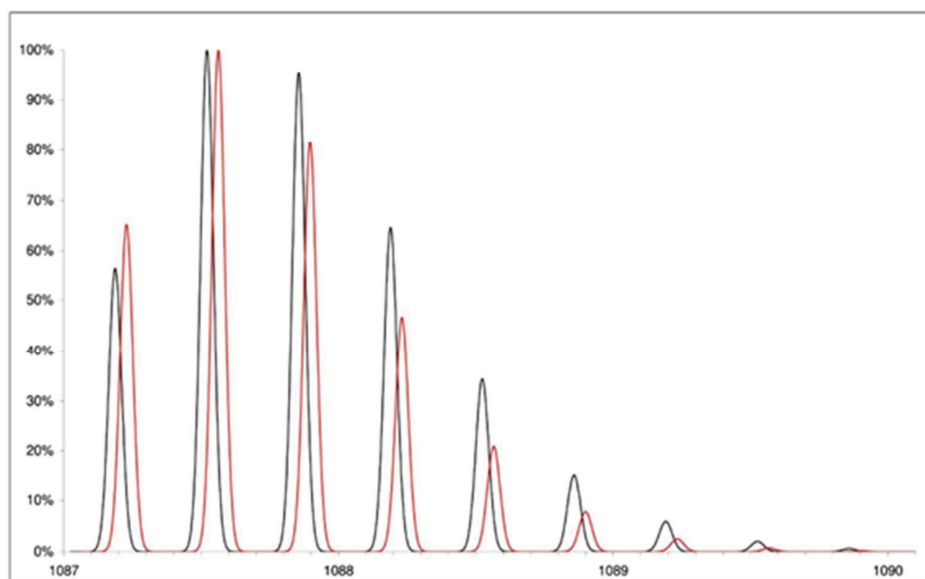
concluded that the effect of four iodine atoms should not be ignored or this type of iodinated pattern would not be recognized as a normal peptide pattern. Being aware of this problem, we found a total of 3 different thyroxine modifications as described below.

Table 1 shows a summary of all the iodinated tyrosines detected using our methodology. The corresponding peptide characteristics (supplemental Table S8) and mapping of unmodified and modified peptides onto the mouse Tg sequence (supplemental Fig. S1) are given as supplementary data. A total of 36 and 24 tyrosines were found to be monoiodinated (MIT) and diiodinated (DIT), respectively. Thirteen tyrosines were only monoiodinated. Of the diiodinated residues, 23 were also found to be monoiodinated. Thus, one position, namely Tyr⁸¹⁰, was only seen as a diiodinated tyrosine. T3 acceptor sites are sites where 3',3,5-triiodothyronine is formed by coupling one DIT and one MIT. The 3',3,5-triiodotyrosine modification was detected on three tyrosine residues: 993, 1310, and 2572. T4 acceptor sites are sites where thyroxine is formed by coupling two DITs. This type of modification was detected on three tyrosines: 25, 1310, and 2572. It is worth noting that Tyr¹³¹⁰ and Tyr²⁵⁷² positions were also found to be triiodinated. We found that all these T3 and T4 acceptor positions were also found as modified mono and diiodo forms. Interestingly, Tyr¹³¹⁰ was found to be modified (monoiodo, diiodotyrosine, 3',3,5-triiodothyronine, and thy-

roxine) on the peptide that is proteotypic of Q2NKY1 (supplemental Fig. S2), confirming once more that this isoform is encoded in the genome of the mice used here. This could be another illustration of the proteogenomics concept (32). Taking these iodinations into account, tyrosine sites had 91% Tg coverage when trypsin, chymotrypsin, and endo-GluC proteolysis results were merged. Our data show a reproducible iodination pattern as we found 1484, 1047, and 1596 iodinated peptides for each of the 3 mice. As currently recommended in shotgun analysis (33), the standardized values obtained by dividing these numbers by the total number of peptides observed in each mouse are quite consistent: 7.8, 7.5, and 8.3% of the total peptides in each case. 80% of mono- and diiodotyrosine modifications were observed in all three mice.

The Search for Immonium Ions Recorded Poor Results—Precursor ion scanning is used to recognize modified peptides on the basis of residue-specific fragment ions. Salek and Lehman (17) noticed that peptides containing a mono- or diiodotyrosine residue generate abundant immonium ions under CID at *m/z* 261.97 and 387.87, respectively. These residue-specific marker ions were suggested as possibly helpful in discovering new iodination sites. These immonium ions are generated at rather “high” *m/z* and can sometimes be observed after CID fragmentation, depending on the size and charge of the fragmented peptide. Besides CID, we used HCD mode to obtain more information in the lower *m/z* part of the spectra. This

Iodination Sites of Thyroglobulin



| | | |
|-----------------------|----------------------|----------------------|
| Distribution Type: | Averagine | Elemental |
| Resolution: | 10000 | 10000 |
| Parent Charge: | 3 | 3 |
| Elemental Composition | C145 H239 N40 O43 S1 | C129 H174 N29 O39 I4 |
| Monoisotopic M/Z: | 1087.5869 | 1087.62808 |

| Isotope Number | m/z | Percent maximum |
|----------------|---------------------|-----------------|
| 0 | 1087.5869 1087.6281 | 54.64 62.84 |
| 1 | 1087.9212 1087.9624 | 100 100 |
| 2 | 1088.2553 1088.2967 | 98.21 84.05 |
| 3 | 1088.5895 1088.631 | 68.07 49.34 |
| 4 | 1088.9235 1088.9652 | 37.1 22.62 |
| 5 | 1089.2576 1089.2995 | 16.83 8.6 |
| 6 | 1089.5916 1089.6337 | 6.58 2.82 |
| 7 | 1089.9257 1089.9679 | 2.27 0.81 |
| 8 | 1090.2597 1090.3021 | 0.71 0.21 |
| 9 | 1090.5938 1090.6363 | 0.2 0.05 |
| 10 | 1090.9278 1090.9705 | 0.05 0.01 |

FIGURE 3. Specific ion pattern for peptide containing the thyroxine modification. This figure was constructed using Protein Prospector software with default resolution set to 10,000. This parameter has no effect on the ratios of the two isotopic patterns shown here. The relative height is shown for the first isotopes of an "Averagine peptide" (black line) with a molecular mass of 1087.5869 and the iodinated peptide (red line) (ILAAVWYYSLEHSTDDYAFSR) containing a thyroxine modification at Tyr²⁵⁷² with a molecular mass of 1087.6808. The elemental composition and characteristics of these two chemical entities are given below, as well as the calculated values for isotopic distribution. The isotopic distribution of iodinated peptide is different from that of noniodinated peptide with a similar mass. The effect of 4 iodine atoms should not be ignored, otherwise this type of iodinated pattern would not be recognized as a peptide pattern.

HCD fragmentation takes place in a dedicated octopole collision cell in the LTQ-Orbitrap XL hybrid mass spectrometer. Fig. 4 shows the spectrum obtained after HCD fragmentation for peptide(990–996), the sequence of which is GGEYAIR, modified with a triiodothyronine. For such a small peptide we obtained an almost complete y series, particularly y_3 and y_4 ions with a mass difference of 632.780 Da (for a theoretical mass difference of 632.892). The triiodothyronine immonium ion, which is clearly present on the spectra (m/z at 605.792) is a residue-specific marker ion that confirms the assignment. These immonium ions have been observed in some of our MS/MS spectra for mono- or diiodotyrosines. However, in most HCD and CID spectra, these signature ions have not been found. Because mono- and diiodotyrosine, triiodothyronine, and thyroxine were identified in the set of HCD data

with a complete series of y (and sometimes b) ions, we wonder if these immonium ions appear only under specific experimental conditions.

Fate of the Lost Side Chain during Thyroid Hormonogenesis—During thyroid hormone synthesis, the iodophenyl moiety of the iodotyrosine residue (donor site) is cleaved and transferred to the other tyrosine residue (acceptor site). As shown in Table 1, many potential donor sites exist, as we found numerous monoiodinated and diiodinated tyrosines. Whether dehydroalanine or pyruvic acid (with a polypeptide break) replace Tyr residues at these sites was investigated using our large experimental dataset. We did not detect any peptides with a dehydroalanine residue replacing a tyrosine, even including the specific MS/MS records we obtained by including the parent mass list for such events (data not shown). How-

Iodination Sites of Thyroglobulin

TABLE 1
Iodinated tyrosines in mouse and human Tg

| Residues ^a | Mouse ^b | Human ^b |
|-----------------------|---------------------------|---|
| | Our results | Literature |
| Y25 | M, D, Thyr. | M, Tri. ² ,Thyr. ^{1,2} (24) |
| Y150 | M, D | M, D ² (149) |
| Y235 | M | Not detected |
| Y237 | M | Not conserved |
| Y259 | M, D | M ² (258) |
| Y377 | M | Not detected |
| Y383 | M, D | Not detected |
| Y430 | M, D | Not conserved |
| Y632 | M, D | Not conserved |
| Y692 | M, D | Not conserved |
| Y704 | M, D | M, D,Tri. ² ,Thyr. ^{1,2} (704) |
| Y785 | M, D | M ² (785) |
| Y810 | D | Not detected |
| Y840 | M | Not detected |
| Y867 | M | M, D ² (866) |
| Y884 | M | D ² (883) |
| Y993 | M, D, Tri. | M, D ² (991) |
| Y1006 | M, D | Not detected |
| Y1115 | M | Not detected |
| Y1166 | M, D | Not detected |
| Y1310 | M, D, Tri., Thyr. | M, Tri. ² ,Thyr. ^{1,2} (1309) |
| Y1384 | M, D | Not conserved |
| Y1424 | M, D | Not conserved |
| Y1446 | M, D | Not detected |
| Y1466 | Not conserved | M, D ² (1466) |
| Y1480 | M | Not detected |
| Y1681 | M | Not conserved |
| Y1913 | M | Not detected |
| Y2180 | M, D | Not conserved |
| Y2182 | M | M ² (2183) |
| Y2477 | M, D | Not detected |
| Y2486 | M, D | Not conserved |
| Y2539 | M, D | Iodinated? ² |
| Y2572 | M, D, Tri., Thyr. | M, D, Tri. ² ,Thyr. ^{1,2} (2572) |
| Y2586 | M, D | M ² ,Thyr. ¹ (2586) |
| Y2616 | Not detected | M ² (2616) |
| Y2629 | M | Not conserved |
| Y2657 | M | Not detected |
| Y2669 | M, D | Not detected |
| Y2696 | Not detected | D ² (2696) |
| Y2764 | Not detected ^c | M, D ² ,Tri. ^{1,2} ,Thyr. ² (2765) |

^a The residue positions correspond to the mouse Tg sequence. Only tyrosine residues iodinated in at least one species are shown.

^b Information on iodinated positions for human Tg was collected from (1) the SwissProt database (2) (49). Not detected means tyrosine was present at this position but not iodinated. Not conserved means tyrosine was not present at this position, M stands for monoiodination, D for diiodination, Tri for triiodo-thyronine modification, and Thyr for thyroxine.

^c This position was only detected with 3 unmodified peptides and no definitive conclusion should be drawn.

ever, we did find evidence of the existence of at least two donor sites (Tyr²⁵⁹ and Tyr²⁵³⁹) in which the tyrosine is replaced by pyruvic acid and the polypeptide chain is broken. For example, Tyr²⁵⁹ is usually encompassed by a large Tg peptide(245–279) with the sequence ELAETGLELLLEIY²⁵⁹-DTIFAGLDQASTFTQSTMYR (detected unmodified and with a monoiodo or diiodo modification at Tyr²⁵⁹). Fig. 5 shows the spectrum assigned to a shorter peptide(259–279) corresponding to a subset of this sequence, Y*DTIFAGLDQASTFTQSTMYR, where the first tyrosine residue had been replaced with pyruvic acid (detected 55 times with a maximum confidence score of 75.64). Moreover, we also specifically detected the complementary peptide(245–258) with the sequence ELAETGLELLLEI in our dataset (19 times with a maximum confidence score of 100.76). For Tyr²⁵³⁹, the Y*QALQNSLGGEDSDAR peptide with the first tyrosine residue replaced by pyruvic acid was detected 25 times with a maximum confidence score of 108.76. In that case the complementary TAF peptide(2536–2538) was too

short to be detected. It is worth mentioning that the two tyrosine residues found as donor sites (Tyr²⁵⁹ and Tyr²⁵³⁹) were both detected as mono- or diiodotyrosine residues, a logical prerequisite to be a donor tyrosyl residue.

Evidence for Cathepsins and Other Proteases in Processed Tg—Many proteolytic events take place in the lumen and intracellular compartments during the processing of iodinated Tg. We analyzed our MS/MS data to list all proteolytic events detected. Supplemental Table S8 reports all semi-trypsin, semi-chymotrypsin, and semi-GluC peptides belonging to Tg. Of these, those arising from nonspecific MASCOT protease patterns (758) were extracted. In terms of spectral count, these represented ~20% of all the Tg peptides detected during the whole study. The abundant semi-tryptic peptide(426–432) (AVEHYQR) was found to be diiodinated. This peptide is a signature of a cellular proteolytic event at Ile⁴²⁵. Note that some proteolytic events were found after two different proteolytic digestive processes: Lys⁵²⁰, Cys¹⁷⁴⁸, Cys¹⁸⁹⁴, Ser²⁰⁷⁵, and Ile²⁶⁸⁸. Other major proteolytic events were detected at Tyr²³⁷, Ile⁴²⁵, Gln⁶⁷⁰, Cys⁷³⁰, Arg¹¹⁶⁰, Ser¹²⁵⁹, Gln¹⁶⁰¹, Ile¹⁷⁴⁷, Gln²⁰⁷⁹, His²¹⁶³, Tyr²⁶³⁹, and Lys²⁷⁶². Some Tg peptide motifs seem very sensitive to proteolytic attack: Gln¹⁰⁹⁶-Arg¹⁰⁹⁷-Asn¹⁰⁹⁸-Pro¹⁰⁹⁹, Ala²³⁸⁸-Asp²³⁸⁹-Val²³⁹⁰-Ala²³⁹¹, and Gln²⁴³⁰-Gln²⁴³¹ are worth mentioning. The resulting peptides are evidence of sites of Tg processing by cathepsins and aminopeptidase/dipeptidylpeptidase trimming.

DISCUSSION

Tg is a polypeptide matrix necessary for the synthesis of thyroid hormones. It is a very large glycoprotein consisting of two identical chains of 330 kDa. The numerous studies carried out so far have only led to a partial and indirect survey of its degree of iodination. Table 1 shows the positions of the iodinated tyrosines in both mouse and human Tg. A more general view of the status of Tg modification has been attempted through review work and many results have been deduced similarity with other established sequences, but were not always demonstrated (6, 34–36). Palumbo *et al.* (37) noted that of the 140 tyrosines in the Tg dimer in rats, only 25–30 are normally iodinated and a much smaller number undergo coupling to form thyroid hormones (T3 and T4). Dunn *et al.* (38, 39) mentioned the presence of four essential hormonogenic sites in rabbit Tg. Those sites, designated A, B, C, and D, were found either experimentally or by homology when the cDNA sequences were established (16, 40, 41). These sites correspond to tyrosyl numbers 5, 2553, 2746, and 1290, respectively (this numbering applies to the mature Tg rabbit molecule, omits the peptide signal, and corresponds to positions 25, 2572, 2764, and 1310 in the mouse sequence with its presequence). Sites A and B are T4 and T3 sites; D is exclusively T4, and C mainly T3. As summarized in Table 1, our proteomic approach provides the broadest panoramic survey ever achieved using native Tg. The sequence coverage for Tg is very high (89% when iodination modifications are included). In the present study, we detected mouse equivalent hormonogenic sites A, B, and D. In our case, sites B and D carry T4 and T3, whereas site A carries only T4. By analogy with other Tg mammals, site C (Tyr²⁷⁶⁴, mouse numbering) is

Downloaded from http://www.jbc.org/ by guest on September 29, 2014

Iodination Sites of Thyroglobulin

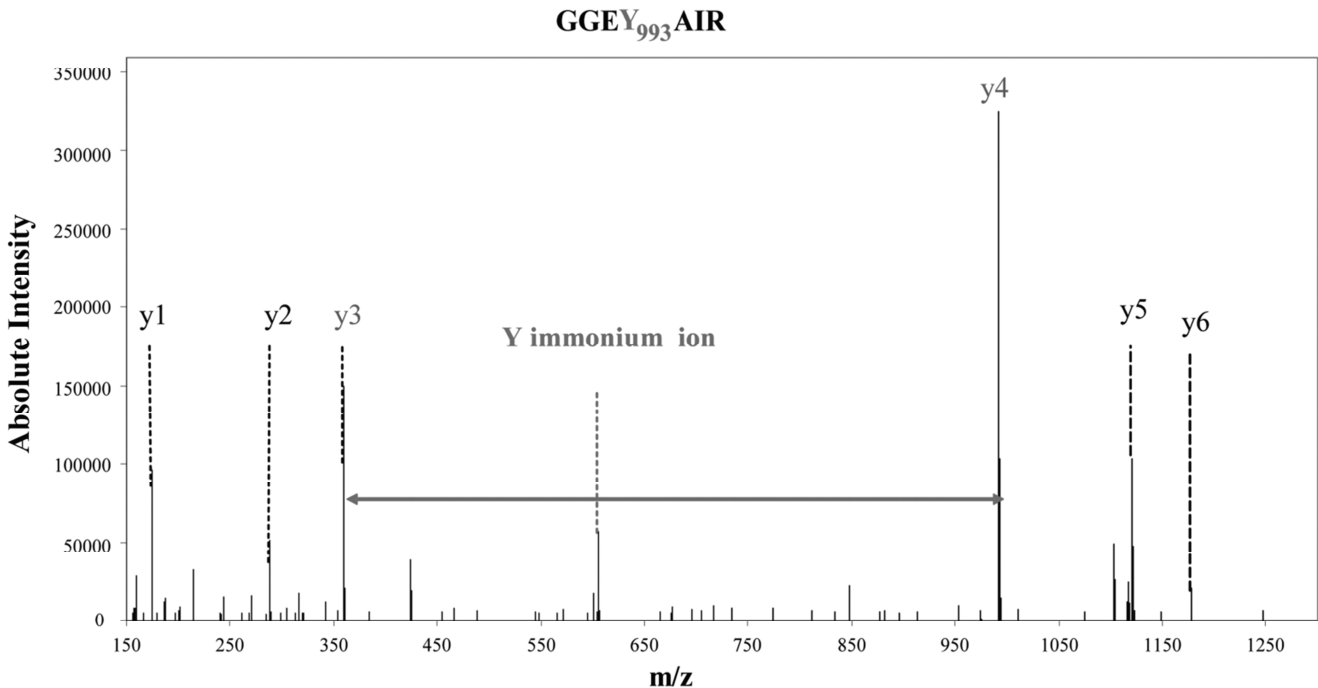


FIGURE 4. HCD fragmentation spectrum showing unambiguous attribution of peptide(990–996) with a triiodothyronine modification on Tyr⁹⁹³. HCD fragmentation provides information from $m/z = 100$ Thomson. For such a small peptide, peptide(990–996), we obtained an almost complete y series. The y_3 and y_4 ions with a mass difference of 632.780 Da (for a theoretical mass difference of 632.892) indicated the presence of the modification. The triiodothyronine immonium ion, which is clearly present on the spectra (605.792), is a residue-specific marker ion that confirms the complete y series.

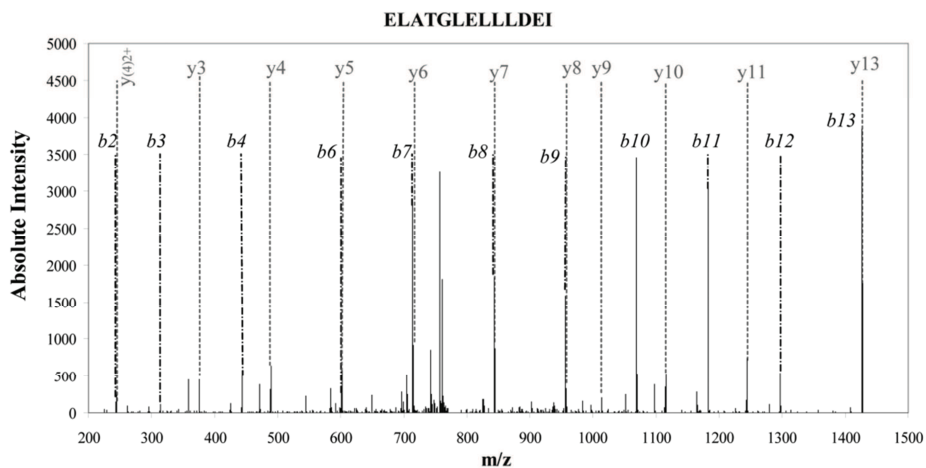


FIGURE 5. CID fragmentation spectra of peptide(245–258) (ELATGLELLLDEI). The almost complete b and y series obtained from a non-noisy spectrum (35 fragment ions using 42 most intense peaks) gave a Mascot ion score of 101. This spectrum assignment indicates the break on the peptide bond after an isoleucine, which cannot be explained enzymatically but by the transformation of a monoiodo (or diiodo) tyrosine into a pyruvic acid residue.

located at the C terminus. In our proteomic strategy, the SYSK peptide containing Tyr²⁷⁶⁴ generated by trypsin was too small to be detected by the mass spectrometer unless a limited proteolysis strategy was used. As far as triiodothyronine is concerned, we demonstrated its presence at positions Tyr²⁵⁷² (site B), Tyr¹³¹⁰ (site D), and also at Tyr⁹⁹³ (mouse numbering) as mentioned in Table 1. This latter position has not previously been described as a hormonogenic site. It should be noted that we did not detect triiodothyronine at position 25 (site A).

We have already mentioned that the sequence coverage obtained for Tg was very high. We wondered why some re-

gions were not covered and noted that most of the undetected peptides carry a potential glycosylation site as predicted by bioinformatic tools based on homology with other species. The number of tyrosyl residues detected (71 of 76), particularly the number of mono- or diiodinated tyrosines, 36 and 24, respectively, is higher than had ever been found before (37). If we consider that only a few positions (4 in our study) are involved in hormonogenesis we might think that the coupling process is highly regioselective and probably driven by constraints because of the native three-dimensional structure of Tg. The abundant iodination found in mouse Tg is probably a storage process for iodine, which can be mobilized when

Downloaded from <http://www.jbc.org/> by guest on September 29, 2014

Iodination Sites of Thyroglobulin

iodine intake becomes limited. More iodination sites were found in mouse Tg compared with those previously identified in the thyroids of larger mammals. This difference may be related to the more active thyroid metabolism in smaller animals or to the greater sensitivity of our methodology.

Dunn *et al.* (39) proposed a classification of iodinated tyrosyl groups into three consensus families: ((D/E)Y), ((S/T)YS)), and (EXY) motives. (i) The ((D/E)Y) di-amino acid consensus was associated with T4 at sites A, B, and D. This is also the case in mouse Tg as supported by our results. It should be noted that the novel T3 site characterized here at residue 993 belongs to this EY consensus. Moreover, this consensus sequence is found iodinated but with apparently no hormone formation at four other locations Tyr³⁸³ (DY), Tyr¹¹¹⁵ (EY), Tyr²⁴⁸⁶ (DY), and Tyr²⁵⁸⁶ (DY). (ii) The ((S/T)YS) tri-amino acid consensus was associated with triiodothyronine synthesis at site C. In the present study, we noticed that of the 36 mono- or 24 diiodotyrosines none could be linked to this consensus group because the sequence ((S/T)YS) is only present at position 2764 (mouse numbering). (iii) The consensus sequence (EXY) appeared to favor iodination. Of 8 possibilities of this consensus in mouse Tg, 6 are iodinated sites as shown here. We can increase this figure to 9 of 12 if we include an extended ((E/D)XY) consensus sequence. This ratio is significantly higher than the mean ratio of 36 iodinated positions of 71 detectable tyrosine residues. Our results confirmed that the presence of an acidic residue located two amino acids upstream of a tyrosine promotes iodination.

Identifying donor tyrosyl residues has been a difficult challenge for years. At donor sites, the coupling process should leave dehydroalanine, which may be converted to alanine, pyruvic acid, or acetic acid during subsequent isolation steps (5, 42). Marriq *et al.* (36, 43) first reported a cleavage in the peptide bond with the appearance of pyruvate instead of tyrosine. In 1997, Gentile *et al.* (6) noted that the identification of donor tyrosyl residues was indirect in all the cases reported so far. In their study, they claimed to provide the first direct identification of tyrosine 1375 in mature bovine Tg (not conserved in humans or mice) as a donor residue. This result was obtained by taking into account the fact that tyrosine was converted into dehydroalanine. They used electrospray MS with 250–500 ppm precision. In 1998, Dunn *et al.* (3) using also electrospray MS demonstrated a cleavage in the polypeptidic chain in bovine Tg, and identified a peptide in which Tyr¹³⁰ (Tyr¹⁵⁰ in mouse numbering) was replaced by pyruvate. In our large dataset, we were unable to detect any peptide displaying a dehydroalanine modification, whereas working to 5 ppm precision, even when the parent mass list was included. We identified two donor tyrosine residues at N-terminal and C-terminal fragments of the Tg molecule with peptide bond cleavage, and demonstrated the presence of pyruvic acid replacing the tyrosine residue. Our thyroid-wide proteomic analysis suggests that the residual side chain is converted to pyruvic acid *in vivo* or *in vitro* after iodotyrosyl transfer. We do not have any evidence that this cleavage occurs in the thyroid *in vivo* but we wonder whether the acidic conditions used during subsequent experimental manipula-

tions were severe enough to break the peptide bond. In any case, detection of this cleavage event provides clear evidence for donor site identification.

Although MS is widely used in the analysis of post-translational protein modifications (44–46), it has rarely been used to identify iodinated tyrosyl residues in Tg and only on fragmented Tg (3, 6, 17). Here, we report the first detailed analysis of the entire Tg molecule from a direct, fresh extract of the mouse thyroid gland. Using this strategy, we obtained unambiguous characterization of 71 tyrosyl residues of this huge protein, confirmed the presence of three thyroxine sites and three triiodothyronine sites, and detected one new T3 site (Tyr⁹⁹³). From the same dataset, we were able to characterize two hormonogenic donor tyrosines at each end of the molecule. Based on spectral count, the ratio of the number of detected spectra for modified to unmodified peptides can be calculated for each modified tyrosine site. Because each peptide has its own ionization characteristics, this ratio is not very informative *per se* but it could become very enlightening when comparing Tg iodination levels under different conditions such as pathophysiological states. The use of isotope synthetic peptides carrying the different iodinated forms using a stable isotope dilution strategy (47) could also be used to record the absolute quantity of each major modification.

The release of thyroid hormones from the Tg prohormone requires the presence of different cathepsins, B, D, and L (9). Friedrichs *et al.* (11) noted that this rather complex process involves sequential proteolytic events with some interaction with different cathepsins, B, K, and L (48). Covalently cross-linked Tg is stored in the thyroid follicular lumen where limited proteolysis occurs. Tg processing continues under intracellular conditions after Tg fragment internalization (12). Our organ-oriented data were obtained from a complex mixture of partially or fully degraded Tg. Novel proteomic strategies, such as those developed to label N-terminal extremities for degradosome or proteogenomic studies (19), could be developed to follow the *in vivo* degradation process in thyroid follicles more accurately.

In conclusion, in this study, we established the iodination status of Tg in mice. Our direct organ to mass spectrometer shotgun approach revealed a more comprehensive set of iodination sites in Tg. Our methodology may be applied to analyze whether potential alteration in the process of Tg iodination could be detected. This study could be carried out on thyroid extracts from various mouse models, such as Tg processing or iodine uptake using genetically modified mutants. Drugs that modify the Tg iodination status could also be investigated with this type of monitoring. The approach does not require large samples. It should be straightforward in application to human thyroid punctures, which could be important for clinical studies on iodination dysfunction.

Acknowledgments—We thank C. Bruley (CEA-Grenoble, iRTSV) for kindly providing the IRMa 1.21.0 parser, G. Imbert and O. Pible (CEA-Marcoule, iBEB) for help with data processing, and E. Quéméneur (CEA-Marcoule, iBEB) for continuous support.

Iodination Sites of Thyroglobulin

REFERENCES

1. Deshpande, V., and Venkatesh, S. G. (1999) *Biochim. Biophys. Acta* **1430**, 157–178
2. Taurog, A., Dorris, M. L., and Doerge, D. R. (1996) *Arch. Biochem. Biophys.* **330**, 24–32
3. Dunn, A. D., Corsi, C. M., Myers, H. E., and Dunn, J. T. (1998) *J. Biol. Chem.* **273**, 25223–25229
4. Gavaret, J. M., Cahnmann, H. J., and Nunez, J. (1981) *J. Biol. Chem.* **256**, 9167–9173
5. Gavaret, J. M., Nunez, J., and Cahnmann, H. J. (1980) *J. Biol. Chem.* **255**, 5281–5285
6. Gentile, F., Ferranti, P., Mamone, G., Malorni, A., and Salvatore, G. (1997) *J. Biol. Chem.* **272**, 639–646
7. Ohmiya, Y., Hayashi, H., Kondo, T., and Kondo, Y. (1990) *J. Biol. Chem.* **265**, 9066–9071
8. Brix, K., Linke, M., Tepel, C., and Herzog, V. (2001) *Biol. Chem.* **382**, 717–725
9. Dunn, A. D., Crutchfield, H. E., and Dunn, J. T. (1991) *J. Biol. Chem.* **266**, 20198–20204
10. Dunn, A. D., Crutchfield, H. E., and Dunn, J. T. (1991) *Endocrinology* **128**, 3073–3080
11. Friedrichs, B., Tepel, C., Reinheckel, T., Deussing, J., von Figura, K., Herzog, V., Peters, C., Saftig, P., and Brix, K. (2003) *J. Clin. Invest.* **111**, 1733–1745
12. Jordans, S., Jenko-Kokalj, S., Kühl, N. M., Tedelind, S., Sendt, W., Brömme, D., Turk, D., and Brix, K. (2009) *BMC Biochem.* **10**, 23
13. Gavaret, J. M., Dème, D., Nunez, J., and Salvatore, G. (1977) *J. Biol. Chem.* **252**, 3281–3285
14. Lejeune, P. J., Marriq, C., Rolland, M., and Lissitzky, S. (1983) *FEBS Lett.* **156**, 77–82
15. Turner, C. D., Chernoff, S. B., Taurog, A., and Rawitch, A. B. (1983) *Arch. Biochem. Biophys.* **222**, 245–258
16. Mercken, L., Simons, M. J., Swillens, S., Massaer, M., and Vassart, G. (1985) *Nature* **316**, 647–651
17. Salek, M., and Lehmann, W. D. (2005) *Proteomics* **5**, 351–353
18. de Groot, A., Dulermo, R., Ortet, P., Blanchard, L., Guérin, P., Fernandez, B., Vacherie, B., Dossat, C., Jolivet, E., Siguier, P., Chandler, M., Barakat, M., Dedieu, A., Barbe, V., Heulin, T., Sommer, S., Achouak, W., and Armengaud, J. (2009) *PLoS Genet.* **5**, e1000434
19. Baudet, M., Ortet, P., Gaillard, J. C., Fernandez, B., Guerin, P., Enjalbal, C., Subra, G., de Groot, A., Barakat, M., Dedieu, A., and Armengaud, J. (2010) *Mol. Cell Proteomics* **9**, 415–426
20. Zivanovic, Y., Armengaud, J., Lagorce, A., Leplat, C., Guérin, P., Duterre, M., Anthouard, V., Forterre, P., Wincker, P., and Confalonieri, F. (2009) *Genome Biol.* **10**, R70
21. Kersey, P., Hermjakob, H., and Apweiler, R. (2000) *Bioinformatics* **16**, 1048–1049
22. Dupierriis, V., Masselon, C., Court, M., Kieffer-Jaquinod, S., and Bruley, C. (2009) *Bioinformatics* **25**, 1980–1981
23. Vizcaíno, J. A., Côté, R., Reisinger, F., Barsnes, H., Foster, J. M., Rameseder, J., Hermjakob, H., and Martens, L. (2010) *Nucleic Acids Res.* **38**, D736–742
24. Kim, P. S., Hossain, S. A., Park, Y. N., Lee, I., Yoo, S. E., and Arvan, P. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9909–9913
25. Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaefer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Usdin, T. B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullahy, S. J., Bosak, S. A., McEwan, P. J., McKernan, K. J., Malek, J. A., Gunaratne, P. H., Richards, S., Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W., Villalon, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butterfield, Y. S., Krzywinski, M. I., Skalska, U., Smailus, D. E., Schnerch, A., Schein, J. E., Jones, S. J., and Marra, M. A. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16899–16903
26. Mural, R. J., Adams, M. D., Myers, E. W., Smith, H. O., Miklos, G. L., Wides, R., Halpern, A., Li, P. W., Sutton, G. G., Nadeau, J., Salzberg, S. L., Holt, R. A., Kodira, C. D., Lu, F., Chen, L., Deng, Z., Evangelista, C. C., Gan, W., Heiman, T. J., Li, J., Li, Z., Merkulov, G. V., Milshina, N. V., Naik, A. K., Qi, R., Shue, B. C., Wang, A., Wang, J., Wang, X., Yan, X., Ye, J., Yooseph, S., Zhao, Q., Zheng, L., Zhu, S. C., Biddick, K., Bolanos, R., Delcher, A. L., Dew, I. M., Fasulo, D., Flanigan, M. J., Huson, D. H., Kravitz, S. A., Miller, J. R., Mobarry, C. M., Reinert, K., Remington, K. A., Zhang, Q., Zheng, X. H., Nusskern, D. R., Lai, Z., Lei, Y., Zhong, W., Yao, A., Guan, P., Ji, R. R., Gu, Z., Wang, Z. Y., Zhong, F., Xiao, C., Chiang, C. C., Yandell, M., Wortman, J. R., Amanatides, P. G., Hladun, S. L., Pratts, E. C., Johnson, J. E., Dodson, K. L., Woodford, K. J., Evans, C. A., Gropman, B., Rusch, D. B., Venter, E., Wang, M., Smith, T. J., Houck, J. T., Tompkins, D. E., Haynes, C., Jacob, D., Chin, S. H., Allen, D. R., Dahlke, C. E., Sanders, R., Li, K., Liu, X., Levitsky, A. A., Majoros, W. H., Chen, Q., Xia, A. C., Lopez, J. R., Donnelly, M. T., Newman, M. H., Glodek, A., Kraft, C. L., Nodell, M., Ali, F., An, H. J., Baldwin-Pitts, D., Beeson, K. Y., Cai, S., Carnes, M., Carver, A., Caulk, P. M., Center, A., Chen, Y. H., Cheng, M. L., Coyne, M. D., Crowder, M., Danaher, S., Davenport, L. B., Desilets, R., Dietz, S. M., Doup, L., Dulaghan, P., Ferreira, S., Fosler, C. R., Gire, H. C., Gluecksmann, A., Gocayne, J. D., Gray, J., Hart, B., Haynes, J., Hoover, J., Howland, T., Ibegwam, C., Jalali, M., Johns, D., Kline, L., Ma, D. S., MacCawley, S., Magoon, A., Mann, F., May, D., McIntosh, T. C., Mehta, S., Moy, L., Moy, M. C., Murphy, B. J., Murphy, S. D., Nelson, K. A., Nuri, Z., Parker, K. A., Prudhomme, A. C., Puri, V. N., Qureshi, H., Raley, J. C., Reardon, M. S., Regier, M. A., Rogers, Y. H., Romblad, D. L., Schutz, J., Scott, J. L., Scott, R., Sitter, C. D., Smallwood, M., Sprague, A. C., Stewart, E., Strong, R. V., Suh, E., Sylvester, K., Thomas, R., Tint, N. N., Tsonis, C., Wang, G., Wang, G., Williams, M. S., Williams, S. M., Windsor, S. M., Wolfe, K., Wu, M. M., Zaveri, J., Chaturvedi, K., Gabrielian, A. E., Ke, Z., Sun, J., Subramanian, G., Venter, J. C., Pfannkoch, C. M., Barnstead, M., and Stephenson, L. D. (2002) *Science* **296**, 1661–1671
27. Malthiery, Y., Marriq, C., Bergé-Lefranc, J. L., Franc, J. L., Henry, M., Lejeune, P. J., Ruf, J., and Lissitzky, S. (1989) *Biochimie* **71**, 195–209
28. Tharakan, R., Edwards, N., and Graham, D. R. (2010) *Proteomics* **10**, 1160–1171
29. Valkenborg, D., Jansen, I., and Burzykowski, T. (2008) *J. Am. Soc. Mass Spectrom.* **19**, 703–712
30. Senko, M. W., Beu, S. C., and McLafferty, F. W. (1994) *Anal. Chem.* **66**, 415–418
31. Barker, W. C., Garavelli, J. S., McGarvey, P. B., Marzec, C. R., Orcutt, B. C., Srinivasarao, G. Y., Yeh, L. S., Ledley, R. S., Mewes, H. W., Pfeiffer, F., Tsugita, A., and Wu, C. (1999) *Nucleic Acids Res.* **27**, 39–43
32. Armengaud, J. (2010) *Expert Rev. Proteomics* **7**, 65–77
33. Liu, H., Sadygov, R. G., and Yates, J. R., 3rd (2004) *Anal. Chem.* **76**, 4193–4201
34. Dunn, J. T., Kim, P. S., Dunn, A. D., Heppner, D. G., Jr., and Moore, R. C. (1983) *J. Biol. Chem.* **258**, 9093–9099
35. Kim, P. S., Dunn, J. T., and Kaiser, D. L. (1984) *Endocrinology* **114**, 369–374
36. Marriq, C., Lejeune, P. J., Venot, N., and Vinet, L. (1991) *Mol. Cell. Endocrinol.* **81**, 155–164
37. Palumbo, G., Gentile, F., Condorelli, G. L., and Salvatore, G. (1990) *J. Biol. Chem.* **265**, 8887–8892
38. Dunn, J. T., Anderson, P. C., Fox, J. W., Fassler, C. A., Dunn, A. D., Hite, L. A., and Moore, R. C. (1987) *J. Biol. Chem.* **262**, 16948–16952
39. Dunn, J. T., and Dunn, A. D. (1999) *Biochimie* **81**, 505–509
40. Malthiery, Y., and Lissitzky, S. (1987) *Eur. J. Biochem.* **165**, 491–498
41. Musti, A. M., Avvedimento, E. V., Polistina, C., Ursini, V. M., Obici, S., Nitsch, L., Cocozza, S., and Di Lauro, R. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 323–327
42. Dunn, J. T. (1995) *Eur. J. Endocrinol.* **132**, 603–604

Downloaded from <http://www.jbc.org/> by guest on September 29, 2014

Supplementary data available at:

<http://www.jbc.org/content/286/1/259/suppl/DC1>

3.1.2 Conclusions et perspectives

Cette publication illustre la puissance des approches protéomiques basées sur la spectrométrie de masse pour étudier les modifications post-traductionnelles de la thyroglobuline. Dans les années 70 et pendant plus de vingt ans la formation des hormones thyroïdiens à partir de la thyroglobuline sur divers modèles a fait l'objet d'intenses travaux. Les approches alors utilisées et basées sur des approches biochimiques classiques ont permis après plusieurs années d'effort d'avoir une vision presque complète (en regroupant divers travaux partiels) de l'évolution de la thyroglobuline.

Ici par une approche unique, sans purification poussée de la thyroglobuline (protéine largement majoritaire de la thyroïde) nous arrivons à avoir une vision presque exhaustive de toutes les modifications subies par la thyroglobuline dans cet organe. Nous mettons en évidence de nouveaux sites d'iodation et apportons des éléments nouveaux pour trancher une question restée en suspens à la fin des années 90.

Cette approche devrait permettre d'étudier plus facilement les degrés d'iodation de la Tg dans divers cas pathologiques ou sur des mutants modifiés génétiquement.

3.2 - Découverte d'une modification post-traductionnelle chez *Deinococcus deserti* en condition de stress

3.2.1 Changements majeurs intervenant dans le protéome soluble de *Deinococcus deserti* dans les premières heures suivant une forte irradiation par rayons gamma

Deinococcus deserti est une bactérie modèle étudiée dans notre institut (iBEB). Nous avons étudié la dynamique du protéome de cette bactérie après lui avoir fait subir une irradiation gamma de 3 kGy. Dans les heures qui suit cette irradiation, la bactérie est capable de reconstituer entièrement son génome, détruit par l'irradiation. Nous avons analysés à partir de gels 2-D, la reconstitution du protéome qui va contribuer à la reconstitution du génome. L'analyse des protéines a été effectuée par la technique du « clustering hiérarchique ». Couplée à une analyse attentive des gels 2-D nous avons mis en évidence les protéines dont la synthèse est fortement augmentée après irradiation et découvert sur l'une d'elle une MPT remarquable (N-terminal acétylase) ainsi que la surexpression d'une acétyltransférase.

RESEARCH

Open Access

Major soluble proteome changes in *Deinococcus deserti* over the earliest stages following gamma-ray irradiation

Alain Dedieu^{1*}, Elodie Sahinovic¹, Philippe Guérin¹, Laurence Blanchard^{2,3,4}, Sylvain Fochesato^{2,3,4}, Bruno Meunier⁵, Arjan de Groot^{2,3,4} and Jean Armengaud¹

Abstract

Background: *Deinococcus deserti* VCD115 has been isolated from Sahara surface sand. This radiotolerant bacterium represents an experimental model of choice to understand adaptation to harsh conditions encountered in hot arid deserts. We analysed the soluble proteome dynamics in this environmentally relevant model after exposure to 3 kGy gamma radiation, a non-lethal dose that generates massive DNA damages. For this, cells were harvested at different time lapses after irradiation and their soluble proteome contents have been analysed by 2-DE and mass spectrometry.

Results: In the first stage of the time course we observed accumulation of DNA damage response protein DdrB (that shows the highest fold change ~11), SSB, and two different RecA proteins (RecA_P and RecA_C). Induction of DNA repair protein PprA, DNA damage response protein DdrD and the two gyrase subunits (GyrA and GyrB) was also detected. A response regulator of the SarP family, a type II site-specific deoxyribonuclease and a putative N-acetyltransferase are three new proteins found to be induced. In a more delayed stage, we observed accumulation of several proteins related to central metabolism and protein turn-over, as well as helicase UvrD and novel forms of both gyrase subunits differing in terms of isoelectric point and molecular weight.

Conclusions: Post-translational modifications of GyrA (N-terminal methionine removal and acetylation) have been evidenced and their significance discussed. We found that the Deide_02842 restriction enzyme, which is specifically found in *D. deserti*, is a new potential member of the radiation/desiccation response regulon, highlighting the specificities of *D. deserti* compared to the *D. radiodurans* model.

Keyword: Proteome, Post-translational modification, Irradiation, Early response, Reference 2D map, Kinetics, Hierarchical clustering

Introduction

Deinococcus deserti belongs to the *Deinococcaceae*, a group of extremely radiotolerant bacteria. *D. deserti* VCD115 has been isolated from upper sand layers of the Sahara desert [1]. In this environment, cells are exposed to very changing life conditions with cycles of high temperatures, high UV irradiation, desiccating conditions, low temperatures and rehydrating conditions. Probably as a consequence of adaptation to such harsh, DNA-

damaging conditions, *D. deserti* and other *Deinococcaceae* [2,3] exhibit an extraordinary ability to withstand ionizing radiation. Chromosomes with numerous radiation- or desiccation-induced double-strand breaks can be repaired in a few hours in *D. deserti* [4]. To learn about the specificities of this bacterium compared with other *Deinococcaceae*, its entire genome sequence has been determined [4]. It consists of a 2.8-Mb chromosome and three large plasmids called P1 (324 kb), P2 (314 kb), and P3 (396 kb). The comparative analysis showed some interesting differences between *D. deserti* and other sequenced *Deinococcus* species. For example, *D. deserti* possesses supplementary genes involved in DNA repair, such as

* Correspondence: alain.dedieu@cea.fr

¹Laboratoire de Biochimie des Systèmes Perturbés, CEA Marcoule, DSV, iBEB, SBTN, BAGNOLS-SUR-CEZE F-30207, France
Full list of author information is available at the end of the article

three *recA* that code for two different RecA proteins, whereas it lacks homologs of several radiation-induced genes in *Deinococcus radiodurans* (e.g., *ddrP* encoding a putative DNA ligase).

The extreme radiotolerance of *Deinococcaceae* was the object of intense investigations during the last years using *D. radiodurans* as model. In cells subjected to irradiation, DNA recombinase RecA was the first protein that was found strongly induced [5]. RecA is essential for radiotolerance [6] and for the fidelity of DNA repair and genome stability in *D. radiodurans* [7]. The molecular mechanisms underlying DNA repair were also examined by transcriptomics [8-10] leading to the description of a repertoire of genes responding to acute gamma irradiation, including genes involved in DNA replication, repair and recombination, cell wall metabolism, cellular transport and many with uncharacterized functions. Another transcriptional study has shown that 72 genes were up-regulated three fold or higher in *D. radiodurans* following gamma irradiation [11]. In this study, besides genes with already known assigned function linked to DNA repair (*gyrA*, *gyrB*, *uvrA*, *uvrB*, *ruvB* and *recA*), five novel genes with unknown or hypothetical assigned functions were highlighted: *ddrA*, *ddrB*, *ddrC*, *ddrD* and *pprA* (*ddr*, DNA damage response; *ppr*, pleiotropic protein promoting DNA repair). Genetic analyses demonstrated a role of these five genes in radiotolerance [11], and further studies have reported that *DdrA*, *DdrB* and *PprA* are involved in DNA repair [12-19]. Radiotolerance of *D. radiodurans* probably results from a combination of different highly coordinated physiological pathways [20]. Besides enzymatic pathways such as DNA repair, passive contributions to radiotolerance have been described such as the nucleoid structure where DNA molecules are highly condensed, which would limit diffusion of DNA fragments [21], and the high intracellular Mn^{2+}/Fe^{2+} ratio that limits protein oxidation [20,22-24]. Today, the survival kit components of *Deinococcaceae* are far to be exhaustive and some possible factors remain controversial [25].

To further analyze these mechanisms of radioresistance, proteomic approaches are promising as both mass spectrometry (MS) tools and protein separation techniques are ultimately quickly improving [26]. As gene expression and presence of the corresponding protein in the cell are not always strictly correlated, it is worth to compile multi-omics data on such subject [27]. Zhang et al. [28] identified by a classical 2D electrophoresis (2DE) gel approach 21 proteins whose cellular levels significantly changed following γ -irradiation of *D. radiodurans*. Surprisingly, among them few proteins are directly related to DNA metabolism. A comparative 2DE gel-based proteomic study, using the wild-type and a *pprI* knock-out strain, has shown that 31 proteins are

up-regulated after exposure to low dose of γ radiation in a *PprI*-dependent manner in *D. radiodurans* [29]. *PprI*, also called *IrrE*, is a *Deinococcus*-specific transcriptional factor essential for radiotolerance [30-32]. Moreover, phosphorylation of two proteins (*DR_A0283* which encodes a serine protease and *DR_1343*, a glyceraldehyde-3-phosphate dehydrogenase) was shown to be dependent upon the presence of *PprI*. More recently, a 2D-proteome reference map has been reported for *Deinococcus geothermalis* grown in standard conditions [33], as well as a shotgun analysis of its major membrane components [34].

Comparisons of the genome sequences of various *Deinococcaceae* shed new light on their potential arsenal related to DNA repair and radiotolerance [4,35,36]. Convinced that comparison of the proteome dynamics of several species will also be fruitful, we performed a time course analysis of the proteome of *D. deserti* after exposure to 3 kGy γ -radiations. The proteome from cells harvested at different time lapses after irradiation were analyzed with a 2-D gel approach. A hierarchical clustering analysis pointed at accumulated proteins at the earliest stages after irradiation. Among these are several important DNA repair proteins. Interestingly, post-translational modifications of *GyrA* were detected. We also observed the up-regulation of novel proteins such as the conserved *Deide_20140* acetyltransferase, *Deide_19260* response regulator, *Deide_21840* PilT protein, as well as specific *D. deserti* proteins such as the *Deide_02842* restriction enzyme. The physiological significance of these novel results is discussed.

Results

Time course analysis of *D. deserti* cells after drastic irradiation

We established a reference proteome map for *D. deserti* cells harvested at the exponential phase (Additional file 1: Figure S1 & Additional file 2: Table S1). We observed that several stress response and DNA repair proteins are already found in standard condition, for example two desiccation tolerance-associated proteins (*Deide_07540*, *Deide_08510*) and *RecA_C* (*Deide_19450*). We investigated the time-dependent changes in the proteome of *D. deserti* after drastic irradiation conditions. Two independent cultures of *D. deserti* VCD115 cells were grown overnight to reach an optical density of 0.4. Each culture was exposed to 3 kGy of gamma rays. After irradiation, each culture was incubated without change of growth medium to avoid any additional stress. Sampling took place immediately before irradiation and 0, 2, 4, 6, 8, 22 and 30 hours post-irradiation. We performed 2DE-gels from our two biological replicates with a specific effort in terms of technical replicates for time-points before irradiation and right after (0, 2 and 4 hours) in order to detect tiny changes in

protein accumulation. In parallel to the 2DE-gel analysis, cell growth as well as DNA damage and repair was followed at the different time points. Figure 1 shows that while DNA is broken in small pieces by irradiation, DNA repair is slow in our experimental set-up with completion visible at 22 hours. Therefore, the early stages of DNA repair can be analyzed following the 2, 4, 6, and 8-hour time-points. No growth was observed for the first 22h (Additional file 3: Figure S2). The cells have started to grow only 30 h post-irradiation. This is in agreement with the PFGE data (Figure 1) that show an increased amount of DNA due to replication at 30 h.

Proteins involved in response to drastic irradiation conditions

We determined the concentration of total proteins for each sample and deposited the same amount of proteins for each gel. We matched with the ImageMaster software the gels belonging to the same class, *i.e.* same time point. When comparing the different classes a significant trend was observed in terms of number of spots. The number of spots reported here is a mean value of the replicates. While before irradiation, roughly 488 (\pm 29) spots were delineated, only 467 (\pm 65) were detected at 0h, 458 (\pm 112) at 2h, 425 (\pm 98) at 4h and 411 (\pm 3) at 6h and 410 (\pm 90) at 8 hours. An increase to 526 (\pm 14) and 546 (\pm 100) spots was recorded for 22 and 30 hours, respectively. The minimum for both biological duplicates was reached 4–6 hours after irradiation. As it will be discussed below, the presence in large quantities of at least

two proteases up-regulated in the first stages after irradiation could explain this phenomenon.

After comparing gel images with the statistical tools implemented in the ImageMaster software, we detected five spots that were significantly increased in intensity after irradiation, with fold changes above 1.5 ($p < 0.05$). Figure 2 shows image captures of three of these spots and fold changes associated along the time courses. They are clearly more intense at 2, 4 and 6 hours and remain high at 8 and 22 h, which is in agreement with the kinetics of the genome recovery shown in Figure 1. We analyzed by MS the polypeptide content of each of these spots throughout the whole time course. The three series of spots shown in Figure 2 correspond to: Deide_1p01260/Deide_3p00210 (RecA_P) and Deide_19450 (RecA_C) in spot sp_411, Deide_02990 (DdrB; spot sp_683), and Deide_15490 (GyrB, subunit B of DNA gyrase; spot sp_101). The two others are Deide_00120 (SSB; spot sp_494) and Deide_2p01380 (PprA; spot sp_523) (data not shown). The two *recA* genes on plasmids P1 and P3 code for exactly the same protein named RecA_P, which shares 81% sequence identity with the chromosome-encoded RecA_C. They can be easily distinguished from each other with several proteotypic peptides (Additional file 4: Figure S3). Under the spot sp_411, RecA_P is detected sooner along the time course and always in larger quantity compared with RecA_C. These six proteins are all involved in DNA protection or repair as implied by several studies performed with the *D. radiodurans* model [8,11,37], or, for the different RecA proteins, with the *D. deserti* model [38]. Deide_02990 (DdrB) is clearly accumulating from 2 hours after irradiation (fold change \sim 3). Its presence is maximal at 6 hours (fold change \sim 11) in the time course whatever the biological replicate. The accumulation of Deide_00120 (SSB) is also detected early, as soon as 2 hours after irradiation. In this case, the fold change remains moderate (\sim 2.5) till 8 hours. Abundance of RecA_P, RecA_C, GyrB, PprA is significantly increased only after 4 hours in our experimental set-up. We did not detect with confidence down-regulated proteins with such fold changes.

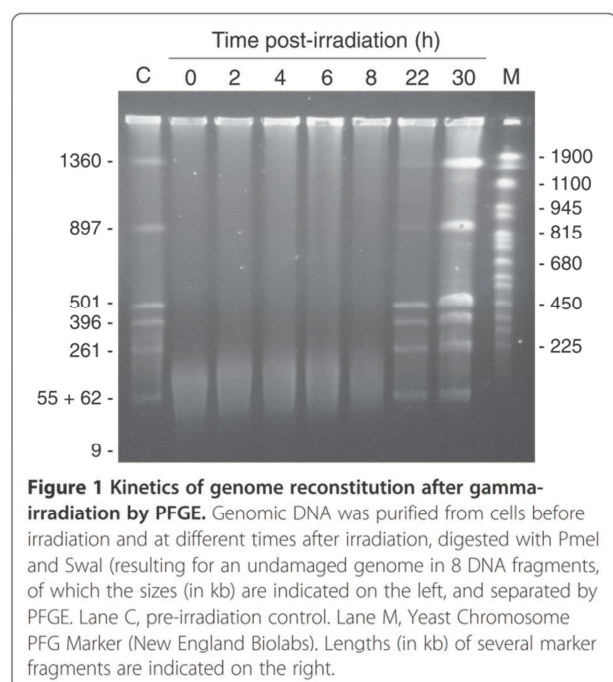
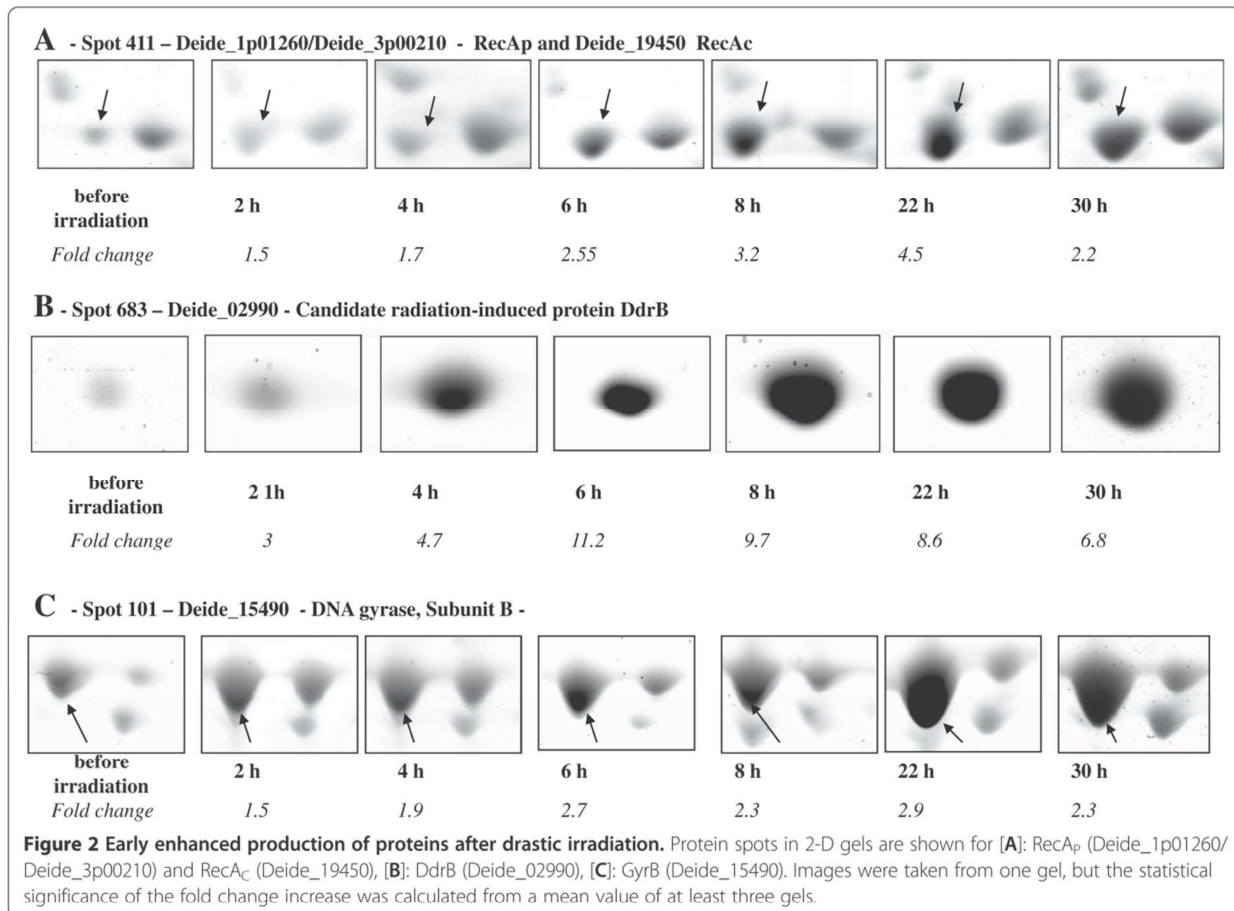


Figure 1 Kinetics of genome reconstitution after gamma-irradiation by PFGE. Genomic DNA was purified from cells before irradiation and at different times after irradiation, digested with Pml and SwaI (resulting for an undamaged genome in 8 DNA fragments, of which the sizes (in kb) are indicated on the left, and separated by PFGE. Lane C, pre-irradiation control. Lane M, Yeast Chromosome PFG Marker (New England Biolabs). Lengths (in kb) of several marker fragments are indicated on the right.

Hierarchical clustering methodologies revealed novel key proteins in the early and delayed responses

Clustering and heat map visualization was performed to identify groups of co-regulated proteins whose expression change over time. Figure 3 shows three clusters, namely red, orange and green clusters, corresponding to proteins accumulating at early stages: 2 h (high abundance increase), 2 h (less pronounced abundance increase), and 4 h after irradiation, respectively. We identified the polypeptides present in these spots and confirmed their presence along the whole time course if



relevant. Table 1 presents the main characteristics of these proteins and the fold changes along the time course after irradiation. The “red” cluster comprises seven spots. Among these, five spots containing six proteins were already detected earlier in our direct gel *t*-test comparison. In the two additional spots, four new proteins were revealed: Deide_12520 (DNA gyrase subunit A) on the first hand, Deide_19260 (response regulator of the SarP family), Deide_20140 (GCN5-related N-acetyltransferase) and Deide_21840 (predicted twitching motility protein) on the other.

The “orange” cluster comprises five spots slightly more intense from 2 hours after irradiation, but not as much as the former cluster. In these five spots we could identify six up-regulated proteins (Table 1): Deide_12520 (DNA gyrase subunit A) which appeared previously in the “red” cluster in a clearly distinguishable spot, Deide_19450 (chromosome-encoded RecA_C), Deide_02842 (a protein related to Type II site-specific deoxyribonuclease), Deide_14090 (a kinase involved in LAO/AO transport system), Deide_13740 (signal recognition particle-docking protein FtsY/cell division protein FtsY), and Deide_01160 (homologue of the *D. radiodurans* radiation-induced

protein DdrD). In spot sp_400 (orange cluster), RecA_C is detected by many proteotypic peptides.

The “green” cluster comprises nine spots whose intensities slightly increase from 4 hours. From these spots we could identify 9 different up-regulated proteins (Table 1). Among these, several proteins are related to central metabolism (Deide_16440, Deide_1p00780, Deide_15960) and protein turn-over (the Deide_19590 La protease and Deide_02310 dipeptidyl aminopeptidase). Remarkably, we found within this cluster three novel spots corresponding to the gyrase subunits, Deide_12520 and Deide_15490. We also found up-regulation of DNA helicase UvrD (Deide_12100). UvrD plays a major role in Double-Strand Break repair through Extended Synthesis-Dependent Strand Annealing [39]. Last but not least, we found the EngA GTPase (Deide_23290) that was shown essential for cell viability in *E. coli* and required for ribosome assembly and stability [40].

Evidences for post-translational modifications of GyrA N-terminus

We detected several distinguishable spots (spots sp_58, sp_59, sp_60, and sp_61) belonging to the three different

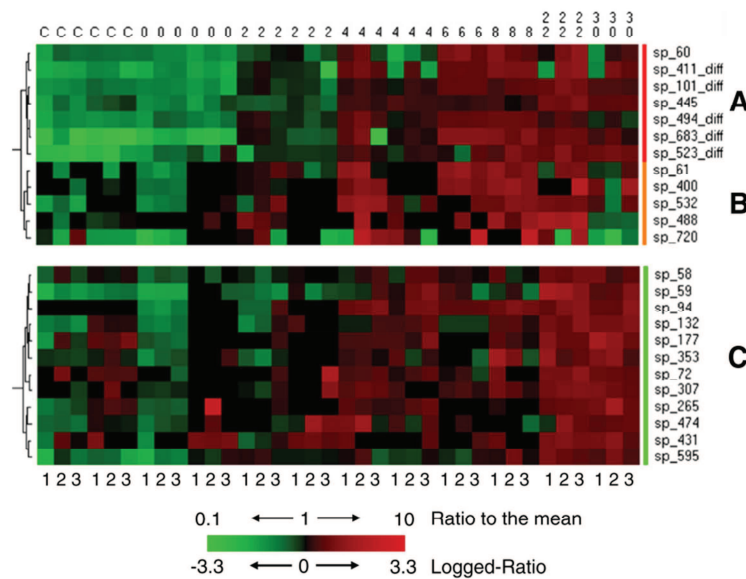


Figure 3 Hierarchical clustering to identify early accumulating proteins. Three clusters were selected from hierarchical clustering results. **A** (red cluster) and **B** (orange cluster) look very similar with up-regulated proteins at an early stage (2 hours after irradiation) but are statistically different. The red cluster contains 5 spots already detected by the statistic method implemented in the ImageMaster software (indicated with diff). The third one **C** (green cluster) gathers proteins up-regulated from 4 hours and beyond. The dendrogram on the left indicates the order of the protein grouping. The intensities (protein accumulation) range from bright green (underproduced in comparison with the mean value) to bright red (overproduced in comparison with the mean value) according to the color scale at the bottom of the figure. The three successive columns for each biological sample correspond to the corresponding 3 technical replicates recorded and they are numbered in the bottom of the figure. Red and orange clusters are statistically different partly because of (i) missing values and (ii) inability of the statistical method available in the software program to analyze gel images (here, ImageMaster 2D) to detect these spots in some gels.

HCA clusters for the subunit A of gyrase (Deide_12520). Whether a post-translational modification such as Ser-Thr-Tyr phosphorylation, Lys or N-terminus acetylation, methylation, or maturation by proteases could explain this behaviour was investigated. In these spots, we found that the GyrA polypeptide is modified with the removal of its first residue (methionine) as can be expected because the second residue in the unmodified polypeptidic chain has a small lateral chain. Figure 4 shows the MS/MS spectrum of the N-terminal most peptide [TGIHPV-DITSEVK] resulting from this specific modification. Moreover, for the spectrum of this peptide obtained with a LTQ-Orbitrap-XL mass spectrometer, the b_3 ion with a m/z value of 314.17 indicated that an acetylation occurs on the TGI N-terminal tripeptide while the parent mass is unambiguously determined with a precision below 1 ppm. We noted that the first threonine residue can be partially acetylated as another MS/MS spectrum corresponded to the unacetylated [TGIHPVDITSEVK] form, with a m/z of the parent at 632.33 and a charge of 2^+ . A pI shift for acidic or neutral proteins due to acetylation has been observed [41,42]. While a mixture of acetylated peptide and non-modified peptide is seen in spots sp_59, sp_60, and sp_61, only the acetylated form is detected in spot sp_58. While this post-translational modification could explain the shift for spot sp_58, other differences

between the three other spots should come from additional post-translational modifications. Interestingly, the up-regulation of Deide_20140 (a putative N-acetyl transferase) as presented in Table 1 could explain differences in acetylation pattern.

Discussion

Following the proteome dynamics of *D. deserti* after irradiation by a 2DE-gel approach highlighted the up-regulation of 21 proteins. Among these, ten proteins either have no homolog in *D. radiodurans* or were not found to be radiation-induced at the protein or RNA level in this organism (see Additional file 5: Table S2 and below), whereas eight others are known to be important players of the irradiation response in *D. radiodurans*: DdrB, SSB, PprA, RecA, GyrA/B, UvrD and DdrD. Concerning RecA, a notable difference is that two functionally different RecA proteins are up-regulated in *D. deserti*. The other sequenced *Deinococcus* species possess only one *recA* gene. Both RecA_C and RecA_P contribute to radiotolerance, but only RecA_C is able to induce expression of DNA translesion polymerases in *D. deserti* [38]. The proteomic results obtained by Zhang *et al.* [28] from the 2D-gel based analysis of *D. radiodurans* after irradiation did not reveal the presence of GyrB and RecA while transcriptomics revealed the up-regulation of

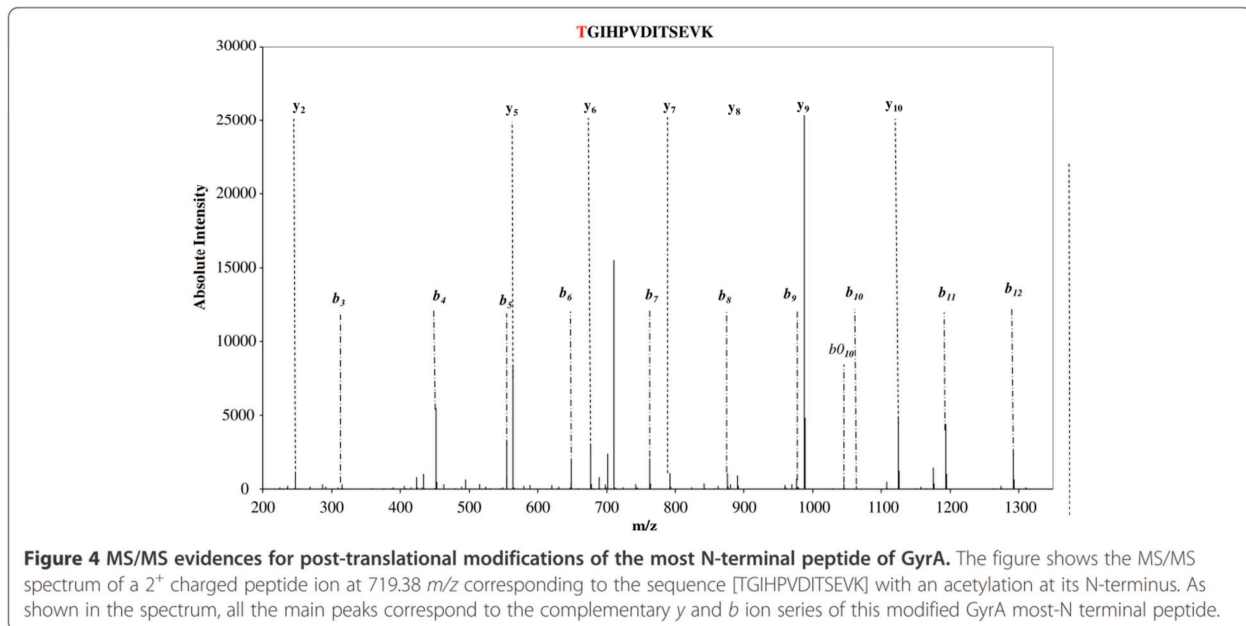
Table 1 Radiation-induced proteins identified by hierarchical clustering analysis

| Spot | Protein identification | Protein description | Time (h) post irradiation | | | | | | | | | | | | | | | | | |
|-------------|---------------------------------|--|---------------------------|------------|---------|-----------|------------|---------|-----------|------------|---------|-----------|------------|---------|------------|------------|---------|------------|------------|---------|
| | | | 2 | | 4 | | 6 | | 8 | | 22 | | 30 | | | | | | | |
| | | | FC (2/ct) | % VOL test | t- test | FC (4/ct) | % VOL test | t- test | FC (6/ct) | % VOL test | t- test | FC (8/ct) | % VOL test | t- test | FC (22/ct) | % VOL test | t- test | FC (30/ct) | % VOL test | t- test |
| sp_60 | Deide_12520 | DNA gyrase, subunit A | 1,64 | 0,14 | <0,01 | 1,84 | 0,16 | 0,02 | 3,89 | 0,34 | <0,01 | 3,59 | 0,31 | <0,01 | 3,68 | 0,32 | <0,01 | 2,00 | 0,18 | <0,01 |
| sp_101_diff | Deide_15490 | DNA gyrase, subunit B | 1,82 | 0,39 | <0,01 | 2,44 | 0,53 | <0,01 | 3,36 | 0,72 | <0,01 | 3,07 | 0,66 | <0,01 | 3,93 | 0,85 | <0,01 | 2,85 | 0,62 | <0,01 |
| sp_411_diff | Deide_1p01260/ Deide_3p00210 | RecA _p | 2,28 | 0,16 | <0,01 | 3,31 | 0,23 | <0,01 | 4,51 | 0,31 | <0,01 | 5,62 | 0,39 | <0,01 | 4,91 | 0,34 | <0,01 | 3,74 | 0,26 | <0,01 |
| sp_445 | Deide_20140 | GCN5-related N-acetyltransferase | 1,29 | 0,08 | 0,09 | 2,21 | 0,14 | <0,01 | 2,06 | 0,13 | <0,01 | 2,03 | 0,13 | <0,01 | 3,35 | 0,21 | <0,01 | 2,67 | 0,17 | <0,01 |
| sp_445 | Deide_19260 | response regulator, SaP | 1,29 | 0,08 | 0,09 | 2,21 | 0,14 | <0,01 | 2,06 | 0,13 | <0,01 | 2,03 | 0,13 | <0,01 | 3,35 | 0,21 | <0,01 | 2,67 | 0,17 | <0,01 |
| sp_445 | Deide_21840 | twitching motility protein | 1,29 | 0,08 | 0,09 | 2,21 | 0,14 | <0,01 | 2,06 | 0,13 | <0,01 | 2,03 | 0,13 | <0,01 | 3,35 | 0,21 | <0,01 | 2,67 | 0,17 | <0,01 |
| sp_494_diff | Deide_00120 | single-stranded DNA-binding protein | 2,05 | 0,37 | <0,01 | 2,85 | 0,52 | <0,01 | 3,40 | 0,62 | <0,01 | 3,63 | 0,66 | <0,01 | 2,84 | 0,52 | <0,01 | 2,07 | 0,38 | <0,01 |
| sp_523_diff | Deide_2p01380 | DNA repair protein PprA | 1,81 | 0,14 | 0,05 | 2,66 | 0,21 | <0,01 | 2,50 | 0,20 | 0,04 | 4,05 | 0,32 | <0,01 | 4,32 | 0,34 | <0,01 | 4,13 | 0,33 | <0,01 |
| sp_683_diff | Deide_02990 | DNA damage response protein DdrB | 4,50 | 0,57 | <0,01 | 5,99 | 0,76 | <0,01 | 11,98 | 1,52 | <0,01 | 10,51 | 1,33 | <0,01 | 8,86 | 1,12 | <0,01 | 6,83 | 0,87 | <0,01 |
| sp_61 | Deide_12520 | DNA gyrase, subunit A | 2,51 | 0,07 | <0,01 | 4,77 | 0,13 | 0,01 | 5,49 | 0,15 | <0,01 | 5,10 | 0,14 | <0,01 | 3,66 | 0,10 | 0,01 | 2,42 | 0,07 | 0,01 |
| sp_400 | Deide_19450 | RecA _c | 1,55 | 0,28 | 0,06 | 3,05 | 0,55 | <0,01 | 3,12 | 0,57 | <0,01 | 3,69 | 0,67 | <0,01 | 5,92 | 1,07 | <0,01 | 2,64 | 0,48 | ns |
| sp_488 | Deide_13740 | signal recognition particle-docking protein FtsY | 1,69 | 0,03 | 0,12 | 2,66 | 0,05 | 0,06 | 2,45 | 0,04 | 0,01 | 2,58 | 0,04 | 0,02 | 3,89 | 0,07 | <0,01 | 0,75 | 0,01 | ns |
| sp_532 | Deide_02842 | Type II site-specific deoxyribonuclease | 2,13 | 0,06 | 0,06 | 3,22 | 0,09 | 0,01 | 3,10 | 0,09 | <0,01 | 3,54 | 0,10 | <0,01 | 3,26 | 0,09 | <0,01 | 2,04 | 0,06 | 0,06 |
| sp_532 | Deide_14090 | LAO/AO transport system kinase | 2,13 | 0,06 | 0,06 | 3,22 | 0,09 | 0,01 | 3,10 | 0,09 | <0,01 | 3,54 | 0,10 | <0,01 | 3,26 | 0,09 | <0,01 | 2,04 | 0,06 | 0,06 |
| sp_720 | Deide_01160 | DNA damage response protein DdrD | 1,48 | 0,07 | ns | 5,00 | 0,23 | 0,03 | 10,23 | 0,46 | <0,01 | 10,24 | 0,46 | <0,01 | 3,29 | 0,15 | 0,19 | 0,67 | 0,03 | ns |
| sp_58 | Deide_19590 | ATP-dependent protease La | 1,06 | 0,07 | ns | 1,43 | 0,09 | 0,01 | 1,28 | 0,09 | 0,14 | 1,34 | 0,09 | 0,08 | 2,06 | 0,14 | <0,01 | 1,85 | 0,12 | <0,01 |
| sp_58 | Deide_12520 | DNA gyrase, subunit A | 1,06 | 0,07 | ns | 1,43 | 0,09 | 0,01 | 1,28 | 0,09 | 0,14 | 1,34 | 0,09 | 0,08 | 2,06 | 0,14 | <0,01 | 1,85 | 0,12 | <0,01 |
| sp_59 | Deide_12520 | DNA gyrase, subunit A | 1,45 | 0,06 | ns | 2,96 | 0,12 | <0,01 | 2,28 | 0,09 | 0,01 | 4,60 | 0,18 | 0,04 | 6,50 | 0,26 | <0,01 | 3,67 | 0,14 | <0,01 |
| sp_72 | Deide_12100 | DNA helicase II (UvrD) | 1,16 | 0,03 | ns | 1,20 | 0,03 | 0,25 | 0,77 | 0,02 | ns | 7,62 | 0,18 | 0,10 | 1,56 | 0,04 | 0,04 | 1,23 | 0,03 | ns |
| sp_94 | Deide_15490 | DNA gyrase, subunit B | 1,95 | 0,04 | ns | 3,80 | 0,07 | <0,01 | 3,28 | 0,06 | 0,01 | 4,11 | 0,08 | <0,01 | 4,45 | 0,09 | 0,01 | 4,07 | 0,08 | 0,03 |
| sp_132 | Deide_02310 | dipeptidyl aminopeptidase | 1,19 | 0,04 | ns | 1,63 | 0,06 | 0,01 | 1,12 | 0,04 | ns | 1,73 | 0,06 | 0,03 | 2,62 | 0,09 | <0,01 | 2,61 | 0,09 | <0,01 |
| sp_307 | Deide_23290 | GTP-binding protein | 1,30 | 0,07 | ns | 1,34 | 0,07 | ns | 1,12 | 0,06 | ns | 1,64 | 0,09 | 0,10 | 1,83 | 0,10 | 0,05 | 1,64 | 0,09 | 0,12 |

Table 1 Radiation-induced proteins identified by hierarchical clustering analysis (Continued)

| | | | | | | | | | | | | | | | | | | | | |
|--------|---------------|--|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-------|------|------|-------|
| sp_353 | Deide_16440 | biotin synthase/thiamine biosynthesis enzyme | 0,79 | 0,02 | ns | 1,29 | 0,04 | 0,07 | 5,14 | 0,15 | 0,03 | 2,21 | 0,07 | 0,06 | 2,12 | 0,06 | <0,01 | 2,17 | 0,06 | <0,01 |
| sp_474 | Deide_1p00780 | flavin dependant oxidoreductase | 1,78 | 0,04 | 0,10 | 1,70 | 0,04 | 0,06 | 1,83 | 0,04 | 0,02 | 1,63 | 0,04 | 0,07 | 1,88 | 0,04 | 0,05 | 1,78 | 0,04 | 0,01 |
| sp_595 | Deide_15960 | NADH dehydrogenase | 1,72 | 0,06 | 0,02 | 1,60 | 0,06 | 0,04 | 1,55 | 0,05 | ns | 2,18 | 0,08 | 0,01 | 2,82 | 0,10 | <0,01 | 2,58 | 0,09 | <0,01 |

FC, fold change; %VOL, averaged spot intensity calculated as the relative volume; ns, not significant; ct, control (average of intensity at time 0 and before irradiation). When several proteins were identified in the same gel spot, the number of peptides, the percentage of sequence coverage, and the emPAI score of each protein were analysed. If the variation of these parameters were identical between the identified abundant proteins, the global fold change of the spot was assigned to these proteins.



their genes [11] (Additional file 5: Table S2). Lu *et al.* [29] did see RecA induction, but apparently not GyrA, GyrB, DdrB, and DdrD accumulation. Here, we found a good correlation with the transcriptomics data obtained with *D. radiodurans* because GyrA, GyrB, RecA, DdrB and DdrD polypeptides are found accumulating in large quantities from 2 hours after irradiation. While this work was in progress, a new proteomic study with *D. radiodurans* was published, also showing up-regulation of GyrA, GyrB and DdrB, but not of DdrD (Additional file 5: Table S2). Using a 2DE-gel approach, Basu & Apte [37] identified ten and two different forms of induced SSB and DdrB, respectively, with each form in a different spot. These different forms are likely resulting from proteolytic processing of the C-terminal end of these proteins [37]. In our study with *D. deserti*, we identified up-regulated SSB and DdrB in only one spot each. Interestingly, we found different spots for both GyrA and GyrB (see below). The use of different bacterial species and experimental conditions are likely explanations for the differences between our results and those previously published on *D. radiodurans* [37].

Remarkably, we found the presence of several spots assigned to the same protein: four (spots sp_58, sp_59, sp_60, sp_61) and two (spots sp_101, sp_94) spots correspond to the GyrA and GyrB subunits, respectively. The different forms of Deide_12520 (GyrA) detected in the three HCA clusters are produced in higher amount after irradiation. The GyrA spots (spots sp_58, sp_59, sp_60, and sp_61) are only separated by small pI shifts. As reported previously by Zhu *et al.* [42], shifts in pI often correlate to protein modifications. These modifications include truncations or deletions, but also post-

translational modifications of amino acid lateral chains such as phosphorylation or acetylation. They can also be due in a limited number of cases to alternative translation starts as revealed in an extensive survey of N-termini of proteins from *D. deserti* [43], or N-terminal trimming by exo-peptidases. At the N-terminus and side chains of amino groups, post-translational modifications will lead generally to small (<0.2) isoelectric point shift. Here, we observed a sequence coverage of ~30% with the GyrA peptides in the four spots. We detected the [2-TGIHPVDITSEV-13] and the [790-VINIAERDSVISAFPIRR-807] peptides that are very close from the expected N-terminal and C-terminal extremities of the 811 amino acid-long GyrA. For this reason, a small truncation of the protein is unlikely, except maturation of the N-terminus (removal of the initial methionine) as evidenced by the most-N-terminal peptide. Regarding post-translational modification of GyrA, we could not detect any phosphorylation for serine, threonine or tyrosine residues, nor acetylation of the lysine side chains within the detected peptides. However, we found that GyrA N-terminus is acetylated. While a mixture of acetylated peptide and non-modified peptide is seen in spots sp_59, sp_60, and sp_61, only the acetylated form is detected in spot sp_58. For eukaryotic proteins acetylation is one of the most common covalent modifications with phosphorylation [44] while it is not frequent for prokaryotic proteins [45,46] even if prokaryotic proteins are also extensively post-translationally modified [47]. Amino-terminal acetylation occurs co-translationally with eukaryotic proteins [45] and post-translationally in the case of prokaryotic proteins [48]. Acetylation may affect protein functions as their

stability and DNA binding activity may be modified; the hyperacetylation of histones being the most illustrative example in this respect as it influences drastically transcription [49]. Hwang *et al.* [50] noted that N-terminal acetylation of proteins could act as a degradation signal in yeast. Interestingly, we found the Deide_20140 GCN5-related acetyltransferase being up-regulated in the earliest time-points after irradiation. Whether this enzyme is responsible for the acetylation observed on GyrA is an interesting hypothesis that deserves further investigation. However, at least under our experimental conditions, a potential Deide_20140-specific function is not essential for radiation resistance, since a *Deide_20140* deletion mutant appeared to be as radiotolerant as the wild-type strain (data not shown). The *D. deserti* genome encodes various putative N-acetyltransferases, and some of these might have redundant activities.

We observed two up-regulated proteases: Deide_19590 (protease La, also called Lon protease) and Deide_02310 (related to dipeptidyl-aminopeptidase/acylaminoacyl-peptidase). Of these, Deide_02310 is not conserved in other sequenced *Deinococcus* species. The presence of these proteases in significant amounts is delayed compared to most DNA-repair related proteins. Their up-regulation, together with the presence of several other proteases (Additional file 2: Table S1) [4], could explain the significant trend observed in terms of number of spots detected in our 2DE-gels. A very active protein turn-over occurs because of protein damages and intense metabolic changes. Protein degradation after exposure to radiation was also reported for *D. radiodurans* [51], which also encodes a high number of proteases. Other data indicated that Lon proteases play a key role in degradation of damaged proteins in *D. radiodurans* [52]. Proteolysis may also control the levels of radiation-induced proteins, as shown for *E. coli* where numerous SOS response proteins, including RecA and UvrA, are substrates for ClpPX, Lon and other proteases [53,54]. Accumulated levels of some DNA repair proteins can be deleterious, and their activity must be restricted to regions of DNA damage. Translesion DNA polymerases are other examples of induced proteins that are rapidly degraded *in vivo* [54]. *D. deserti* encodes functional translesion DNA polymerases that are induced upon DNA damage [38], and their activity should be strictly controlled to prevent elevated levels of mutagenesis. Remarkably, the regulation of the dipeptidyl-aminopeptidase/acylaminoacyl-peptidase (Deide_02310) could explain the heterogeneity observed for the GyrA/B subunits.

Besides the already known proteins related to DNA-repair or metabolic changes described above, we detected the presence of several novel proteins that probably fulfill key roles in the radiation responses in *D. deserti*. Deide_20140 presents some far-related similarities with

MshD (E value: 3.01e-09), the acetyl-transferase that catalyzes the final step of mycothiol biosynthesis in various members of the Actinomycetes. Mycothiol replaces glutathione in these species. Glutathione is a well-known antioxidant that helps protecting cells from reactive oxygen species such as free radicals and peroxides. Together with other antioxidants, such as Mn²⁺ complexes [5,22], the Deide_20140 acetyl-transferase in *D. deserti* could contribute to tolerance to oxidative stress, which is generated by ionizing radiation-induced water radiolysis [55]. Oxidative stress may also occur during dehydration as described by Fredrickson and co-workers [56]. Interestingly, we observed the up-regulation at the earliest stage after irradiation of the Deide_19260 protein (COG3947) that shows strong similarities with response regulators from two-component systems. COG3947 members usually comprise two structural domains: i) at their N-terminus, a CheY-like receiver with a phosphoacceptor site (Asp52 in Deide_19260) that is phosphorylated by histidine kinase homologs and ii) a DNA-binding transcriptional activator of the SARP family at their C-terminus. We tried to check the phosphorylation status of the CheY domain, but a peptide covering residue 52 was not detected in our experiments. Such aspartate-phosphorylation should be stable enough to be detected in our experimental conditions as previously found [57,58].

It is worth to note that this putative two-component regulator is highly conserved in all radiotolerant *Deinococcus* species (70–84% identity) and in the thermophilic and radiotolerant *Truepera radiovictrix* (41% identity). However, the genetic context for *Deide_19260* and its homologs is not conserved. None of the genes flanking *Deide_19260* or its homologs encodes a histidine kinase in these species. The putative cognate histidine kinase for *Deide_19260* is unknown. The *Deide_19260* protein could be an important regulator for stress and/or DNA-damage response in *D. deserti*, besides the previously identified IrrE transcriptional activator. It will be of interest to study the targets of the SARP regulator. However, in contrast to the radiation-sensitive *irrE* mutant [32], deletion of *Deide_19260* did not result in loss of radiation tolerance (data not shown). This could mean that *Deide_19260* is not involved in radiation tolerance or that its function is redundant with another response regulator. The Deide_21840 up-regulated protein is related to PilT (COG2805), a nucleotide binding protein responsible for the retraction of type IV pili, likely by pili disassembly. This retraction provides the force required for travel of bacteria in low water environments. This protein is also required for DNA uptake in several bacteria [59]. Three proposed roles for DNA uptake are genetic transformation, DNA repair, and to provide a source of nutrient [60]. PilT may thus be an important element for survival at the population level of *D. deserti* upon adverse conditions. Another up-regulated protein, Deide_02842,

presents far-related similarities with BglI restriction enzyme of *Bacillus atrophaeus*, previously known as *Bacillus globigii* [61,62]. We can predict that, like BglI, Deide_02842 is also a type-II site-specific ribonuclease, *i.e.* it cleaves specifically within or close to the recognition sequence in DNA. The Deide_02841 gene predicted to encode a DNA methylase is adjacent to Deide_02842. DNA methylation by this enzyme would protect *D. deserti*'s own DNA from cleavage by Deide_02842. Homologs of both genes are absent from other sequenced *Deinococcus* species. *D. deserti* may have acquired these genes by horizontal gene transfer, as suggested by their low GC percentage (46 and 50% versus an average of 63% for the total genome) and the nearby located transposase and integrase genes. Such restriction enzyme and methylase duo is known to be involved in the protection of bacterial cells by limiting incorporation of incoming foreign DNA, such as from bacteriophages, into the host genome. The Deide_13740 up-regulated protein (FtsY signal recognition particle GTPase) constitutes a universally conserved protein targeting pathway that ensures the co-translational delivery of substrates to the membrane-bound Sec translocon [63]. Logically, the delivery of proteins synthesized in the cytosol to their correct cellular compartment is of utmost importance for the cell following gamma-irradiation. Deide_14090 shows strong sequence similarities with the ArgK kinase that phosphorylates periplasmic binding proteins involved in the LAO (lysine, arginine, ornithine)/AO transport systems [64].

To investigate if the proteins identified in this study may be up-regulated by a common mechanism, the upstream regions of the corresponding 22 genes were analyzed using MEME [65]. A 17-bp palindromic motif was found upstream of 11 genes: *gyrA*, *gyrB*, *ssb*, *pprA*, *ddrB*, *ddrD*, *uvrD*, *recA_C*, *recA_{P1}*, *recA_{P3}* and Deide_02842 (Type II restriction enzyme). This motif corresponds to the radiation/desiccation response motif (RDRM) first identified after analysis of radiation-induced genes in *D. radiodurans* and *D. geothermalis* [36]. In a previous study we scanned the entire genome of *D. deserti* with this motif and a match with the RDRM was found in the upstream region of 25 genes, including 10 of the 11 genes mentioned above [4]. Here, using another method (MEME), Deide_02842 is found as a new potential member of the RDRM regulon [4,36]. It has been shown that radiation-induced transcription of at least some of these RDRM-containing genes is dependent on PprI (IrrE) [30,31,38]. Besides the RDRM upstream of 11 genes, no other motifs were found for the up-regulated proteins identified here, and transcription of the other 11 genes may be regulated in different manners. Alternatively, accumulation of the corresponding proteins may not be related to up-regulated transcription but rather to an

increase of their translation and/or to protein-protein or protein-DNA interactions that increase their stability.

We have shown that DdrB and SSB are clearly accumulating in large quantities in the very earliest stages after irradiation, at least in our experimental conditions. These two single-stranded DNA binding proteins are probably of high importance to protect ssDNA that is formed after massive DNA damage and/or to initiate genome reconstitution by recruitment of other DNA repair proteins. That they are both up-regulated at the same early stage may indicate that they work concomitantly. Interestingly, Xu et al. [19] have recently shown that both proteins interact *in vitro*. Another recent study reported that DdrB is involved in DNA repair through a single-strand annealing (SSA) process that precedes the Extended Synthesis-Dependent Strand Annealing (ESDSA) [12]. Finally, the specific post-translational modifications of GyrA detected from irradiated samples raises the importance of post-translational modifications of the *Deinococcus* proteome upon DNA damages. Whether the resulting heterogeneity impacts the cellular response is an open question.

Materials and methods

D. deserti cellular growth and irradiation conditions

D. deserti strain VCD115 [1] was grown in 10-fold diluted tryptic soy broth (TSB/10) supplemented with trace elements [32]. To prevent other fluctuations than the irradiation, pre- and post-irradiation incubation was performed under conditions similar to the irradiation ones, *i.e.* 20°C and without vigorous shaking. Two independent 1L cultures were grown to exponential phase (OD₆₀₀ 0.4) and exposed to 3 kGy of gamma rays (39 Gy/min, ⁶⁰Co source; CEA Cadarache, France). This resulted in 68% survival as determined by plating of serial dilutions and colony-forming unit counting. Samples of 100–150 ml for proteome analysis were taken immediately before irradiation and at different time points after irradiation. Cells were harvested by centrifugation, washed twice with 50 mM Tris-HCl pH 8.0, frozen in liquid nitrogen and stored at -80°C. Culture samples (5 ml) were also taken for pulsed-field gel electrophoresis (PFGE) analysis. PFGE analysis of genome reconstitution was performed as described [4].

D. deserti protein extracts

For each condition, cells (0.2 g of wet material) were resuspended in 3 mL of lysis buffer consisting of 8.75 M urea, 2.5 M thiourea, 50 mM DTT, 25 mM spermine, and 5% CHAPS, and containing a protease inhibitor cocktail (Complete Mini) from Roche (1 tablet per sample). Cells were disrupted with an ultrasound Vibracell 75042 sonicator (Fisher Bioblock Scientific) for 35 sec (7 cycles of 5 sec pulse and 5 sec pause). The extracts were

then ultracentrifuged for 1h at 100,000 *g* at 4°C to remove cellular debris. Protein concentration from the supernatants was determined by a Bradford assay (Biorad) using Bovine Serum Albumin as a standard after diluting the samples to perform the assay below the limit of 4M urea.

2-D electrophoresis and image analysis

Because we followed several points along the time course after irradiation, only two independent biological replicates were performed. Because numerous biological replicates are difficult to perform regarding the radiation facilities used, specific effort in term of technical replicates has been done especially for time-points before irradiation and right after (0, 2 and 4 hours) with 4 gels done per biological replicate for these time-points while 2 gels were done per biological replicate at 6, 8, 22 and 30 hours, resulting in a total of 48 gels for the whole analysis. Immobilines 18 cm Drystrips pH 3–10 (GE Healthcare) were rehydrated overnight with 300 µg of protein extract. Isoelectric focusing was performed with a Multiphor system (GE Healthcare) up to 60 kVh. After focusing, separation strips were equilibrated for 15 min in 50 mM Tris–HCl buffered at pH 8.8 and containing 6 M urea, 30% glycerol, and 10 mg/ml dithiothreitol to disrupt disulphide bridges. The reduced thiols were then alkylated with 25 mg/ml iodoacetamide in the same buffer for 15 min. The second dimension was performed using 12 % acrylamide gels in Protean II xi 2-D cell (Biorad) at 25V for 1h then at 12.5 W/gel. The gels were first fixed in 5% acetic acid and 30% ethanol overnight, washed 3 times in H₂O for 10 min, treated with 0.02% sodium thiosulphate for 1 min, and then rinsed twice with H₂O for 1 min. The gels were stained overnight with Coomassie Brilliant Blue G250 (Biorad) and then rinsed twice with distilled water. These gels were scanned using an Umax Scanner (Pharmacia) at 600 dpi. Image analysis was performed using ImageMaster 2D Platinum v5.0 (GE Healthcare). The protein content of each spot was determined by its relative volume compared with the sum of all spot volumes in the gel, and expressed as a percentage of volume (%Vol). The replicated gels corresponding to the same time point were grouped in the same class. Statistical analysis was performed on individual sample classes, the control class comprising gels where unirradiated samples were resolved. Class ratios above 1.5 (meaning 1.5 fold change compared with control) were selected, t-test controlled ($p < 0.05$), and the differences were examined visually prior to identification by MS. For the 2DE reference map, protein spots were identified by peptide mass fingerprint using a Biflex IV mass spectrometer (Bruker Daltonics). Some faint spots were identified by tandem mass spectrometry done with an Esquire ion trap

(Bruker Daltonics) coupled to an Ultimate™ NanoLC system (Dionex-LC Packings). For the time-course experiment, protein spots were identified by tandem mass spectrometry done with an LTQ-Orbitrap XL (Thermo) mass spectrometer.

Hierarchical clustering analysis (HCA)

HCA was used to group proteins with similar expression profile and help for visual selection of interesting clusters as described [66]. To this end, the complete %Vol matrix from image analysis was logged-ratio transformed (i.e. each protein profile centered on its mean value after being logged) and imported in the PermutMatrix software [67]. HCA was then processed according to the Pearson distance and the Ward aggregation procedure to construct the resulting dendrogram and to allow heat map visualization of the clustered data matrix [66]. Only clusters with interesting profile regarding reconstitution of intact DNA molecules, i.e. those with spots intensity clearly increasing during the early time-course experiment (time 2, 4, 6 and 8 h post irradiation), were selected. The corresponding spots were manually controlled for missing values and/or bad alignments so that false positive detection was reduced.

In-gel trypsin digestion

2-DE protein spots were excised from gels and treated as previously described [43]. They were trypsin digested using the Montage In-Gel DigestZP Kit (Millipore). The resulting peptides were analysed either by Matrix-Assisted Laser Desorption Ionisation - Time Of Flight (MALDI-TOF) MS or nano-liquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS).

Maldi-tof MS

A Biflex IV mass spectrometer (Bruker Daltonics) was used in reflectron mode for MALDI-TOF MS analysis of peptides and mass fingerprint identification of proteins. A saturated solution of α -cyano-4 hydroxycinnamic acid prepared in acetonitrile:water 1:1 with 0.1% trifluoroacetic acid was ¼ diluted in the same solvent, and used as a matrix. Peptide and matrix solution were spotted on a stainless steel plate and air-dried. Calibration was done using a standard peptide mixture (Peptide Calibration Standard) from Bruker Daltonics as previously described [68]. Spectra were acquired over the 500–3500 *m/z* range by accumulating spectra from at least 100 shots, with ion extraction mode using an accelerating voltage of 19 kV and an extraction delay of 0.2 ns. Peptide mass fingerprints were analysed using the Mascot software version 2.2 (Matrix Science) and a local database comprising 3,455 *D. deserti* protein sequences totalling 980,690 amino acids [4]. The maximum number of miss-cleavages was set at one. Carbamidomethylation of

cysteines was set as fixed modification and oxidation of methionines as variable modification. Protein identification reliability was evaluated on the basis of the probability-based Mowse score ($p < 0.05$), mass error, number of peptides matches and similarity of experimental and theoretical protein molecular masses and isoelectric points.

NanoLC-MS/MS

NanoLC-MS/MS was done either with an Esquire ion trap (Bruker Daltonics) coupled to an Ultimate™ NanoLC system (Dionex-LC Packings) or an LTQ-Orbitrap XL instrument (Thermo) coupled to an Ultimate 3000 LC-system (Dionex-LC Packings). In the former case, the instrument was controlled by HyStar™ software and on-line connected with a PepMap C18 nanocolumn (75 $\mu\text{m} \times 150$ LC Packings). Samples (20 μL) were injected and separated with a 30 min linear gradient. The gradient was achieved from 5% solvent A (95:5 H₂O/ACN 0.1%FA) to solvent 50% B (20:80 H₂O/ACN 0.1% FA) at a flow rate of 20 nl/min. Eluted peptides were electro-sprayed into the nano-ESI source of an Esquire 3000 Plus (Bruker). Spray voltage was set at 2000 V and capillary temperature at 170°C. The ion trap was operated in data-dependent MS to MS/MS switching mode using various precursors detected in the 100–2000 m/z unit window, scan resolution 13 000 m/z per second, selected using a 3 m/z unit ion isolation window and excluding single charged ions. Data were processed with the DataAnalysis software (Bruker Daltonics). Mgf files were searched against the *D. deserti* protein sequence local database using MASCOT 2.2 software (MatrixScience). MASCOT search parameters used with MS/MS data were similar as for those described above except: peptide tolerance at 0.4 Da, MS/MS tolerance at 0.4 Da, peptide charge at +1/+2/+3. Only peptides on rank 1 with an ion score above 30 were selected. Proteins identified by at least 2 different peptides were validated. Otherwise, LC-MS/MS experiments were performed on a LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to a UltiMate 3000 LC system (Dionex-LC Packings) in conditions similar as those previously described [69]. Briefly, peptide mixtures (0.1–2 pmol) were loaded and desalted online in a reverse phase precolumn Acclaim Pepmap 100 C18 (5 μm bead size, 100 Å pore size, 5mm \times 300 μm) from LC Packings. They were resolved on a nano-scale Acclaim Pepmap 100 C18 (3 μm bead size, 100Å pore size, 15 cm \times 75 μm) from LC Packings at a flow rate of 0.3 $\mu\text{L}/\text{min}$. Peptides were separated using a 30 min-gradient (5 to 60% solvent B) with aqueous solvent A (0.1% HCOOH) and solvent B (0.1% HCOOH/80% CH₃CN). The column was then washed for 10 min with 100% B and re-equilibrated for 20 min with 5% B. Full-scan mass spectra were measured from m/z 300 to 1800. The LTQ-orbitrap XL mass spectrometer was operated in the data-dependent mode using the TOP5 strategy with a Fourier Transform

Mass Spectrometer (FTMS) resolution sets at 30,000 as previously reported [4]. In brief, a scan cycle was initiated with a full scan of high mass accuracy in the Orbitrap, which was followed by MS/MS scans in the linear ion trap on the 5 most abundant precursor ions with dynamic exclusion of previously selected ions. This dynamic exclusion consisted in two acquisitions of MS/MS spectra of the most abundant ion during a period of 30 sec and then excluding this ion for the followed fragmentations during the next 60 sec. The activation type was collisional-induced dissociation with a standard normalized collision energy sets at 30.

Construction and radiation survival of *D. deserti* mutant strains

The *Deide_19260* deletion mutant was constructed using recently developed genetic tools for *D. deserti* [38]. Briefly, the mutant was obtained by allelic replacement via double homologous recombination after transformation of *D. deserti* wild-type with a plasmid construct derived from *E. coli* vector pUC19, which does not replicate in *D. deserti*. In this plasmid, one-kb fragments corresponding to DNA upstream and downstream of *Deide_19260* were cloned in the correct orientation respectively upstream and downstream of a kanamycin resistance cassette. The latter was cloned as a BamHI-PstI fragment in pUC19. Expression of the kanamycin resistance gene in this cassette is driven by the constitutive promoter of the *tuf* gene (*Deide_18970*, encoding elongation factor EF-Tu). After three rounds of selection on plates containing 10 $\mu\text{g}/\text{ml}$ kanamycin, diagnostic PCR to confirm double homologous recombination at the correct locus and complete absence of *Deide_19260* was performed using appropriate primers. The *Deide_19260* deletion mutant was called SF7. The *Deide_20140* mutant, called SF8, was constructed in a similar way (but here the kanamycin cassette was present in pUC19 as a BamHI-XbaI fragment). All primer sequences are listed in Additional file 6: Table S3. Survival of mutants and wild-type after exposure to increasing doses of gamma (up to 10 kGy) or UV (up to 750 J/m²) irradiation was determined as described [32].

Additional files

Additional file 1: Figure S1. Reference 2D electrophoresis gel map for exponential *Deinococcus deserti* VCD115 cells. This map was established for cells grown at 30°C and harvested at the exponential phase. From 863 clearly distinguishable Coomassie-stained spots, we selected 131 spots among the most intense that could serve as landmarks. They are labelled on the figure and their characteristics reported in Additional file 2: Table S1.

Additional file 2: Table S1. List of proteins of the reference 2DE-gel map. This table lists the 171 soluble polypeptides unambiguously identified from 137 spots selected among the most intense. We can note the presence of 20 hypothetical conserved proteins and two of them

highly expressed (Deide_06180 and Deide_11730). The most abundant proteins appear in bold face.

Additional file 3: Figure S2. Post-irradiation growth of *D. deserti*.

Additional file 4: Figure S3. Comparison between *Deide_1p01260* & *Deide_3p00210* genes, and RecA_C & RecA_P polypeptides. Panel A shows the sequence of *Deide_1p01260* and *Deide_3p00210* open reading frames with identical nucleotides in black and differing nucleotides in red. These different nucleotides are all in the third position of the codon and are silent mutations. Panel B shows the amino acid sequences of RecA_C (upper line) and RecA_P (bottom line). Amino acids that are different in these sequences are in blue and red respectively. Proteotypic tryptic peptides detected by mass spectrometry are indicated with blue and red lines.

Additional file 5: Table S2. Comparison of radio-induced proteins of *D. deserti* with transcriptomics and proteomics data from *D. radiodurans*.

Additional file 6: Table S3. Primers.

Abbreviations

PFGE: Pulsed-Field Gel Electrophoresis; HCA: Hierarchical Clustering Analysis.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AD designed the experiments, analysed and interpreted the data, and drafted the manuscript. ES and PG recorded the proteomic and mass spectrometry data, and participated in the data interpretation. LB, SF and AdG prepared the microbial samples and performed the genetic experiments, BM performed the HCA analysis. AdG and JA conceived the study, participated in the data interpretation, drafted and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We gratefully acknowledge Eric Quéméneur and Thierry Heulin for constant support, as well as Véronique Malard for advices regarding preparation of protein extracts. We thank the Commissariat à l'Energie Atomique et aux Energies Alternatives and the Agence Nationale de la Recherche (ANR-07-BLAN-0106-02) for financial support.

Author details

¹Laboratoire de Biochimie des Systèmes Perturbés, CEA Marcoule, DSV, iBEB, SBTN, LBSP, BAGNOLS-SUR-CEZE F-30207, France. ²CEA, DSV, IBEB, Lab Ecol Microb Rhizosphere & Environ Extrem (LEMIRE), Saint-Paul-lez-Durance F-13108, France. ³CNRS, UMR 6191, Biol Veget & Microbiol Environ, Saint-Paul-lez-Durance F-13108, France. ⁴Aix-Marseille Université, Saint-Paul-lez-Durance F-13108, France. ⁵INRA, UR1213 Herbivores, Saint Genès Champanelle F-63122, France.

Received: 3 August 2012 Accepted: 23 December 2012

Published: 15 January 2013

References

- De Groot A, Chapon V, Servant P, Christen R, Saux MF, Sommer S, Heulin T: *Deinococcus deserti* sp. nov., a gamma-radiation-tolerant bacterium isolated from the Sahara Desert. *Int J Syst Evol Microbiol* 2005, **55**:2441–2446.
- Mattimore V, Battista JR: Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *J Bacteriol* 1996, **178**:633–637.
- Rainey FA, Ray K, Ferreira M, Gatz BZ, Nobre MF, Bagaley D, Rash BA, Park MJ, Earl AM, Shank NC, *et al*: Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample. *Appl Environ Microbiol* 2005, **71**:5225–5235.
- de Groot A, Dulermeo R, Ortet P, Blanchard L, Guerin P, Fernandez B, Vacherie B, Dossat C, Jolivet E, Siguier P, *et al*: Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. *PLoS Genet* 2009, **5**:e1000434.
- Carroll JD, Daly MJ, Minton KW: Expression of recA in *Deinococcus radiodurans*. *J Bacteriol* 1996, **178**:130–135.
- Gutman PD, Carroll JD, Masters CI, Minton KW: Sequencing, targeted mutagenesis and expression of a recA gene required for the extreme radioresistance of *Deinococcus radiodurans*. *Gene* 1994, **141**:31–37.
- Repar J, Cvjetan S, Slade D, Radman M, Zahradka D, Zahradka K: RecA protein assures fidelity of DNA repair and genome stability in *Deinococcus radiodurans*. *DNA Repair (Amst)* 2010, **9**:1151–1161.
- Liu Y, Zhou J, Omelchenko MV, Beliaev AS, Venkateswaran A, Stair J, Wu L, Thompson DK, Xu D, Rogozin IB, *et al*: Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc Natl Acad Sci USA* 2003, **100**:4191–4196.
- Edwards JS, Battista JR: Using DNA microarray data to understand the ionizing radiation resistance of *Deinococcus radiodurans*. *Trends Biotechnol* 2003, **21**:381–382.
- Narumi I: Unlocking radiation resistance mechanisms: still a long way to go. *Trends Microbiol* 2003, **11**:422–425.
- Tanaka M, Earl AM, Howell HA, Park MJ, Eisen JA, Peterson SN, Battista JR: Analysis of *Deinococcus radiodurans*'s transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance. *Genetics* 2004, **168**:21–33.
- Bouthier De La Tour C, Boissard S, Norais C, Touelle M, Bentchikou E, Vannier F, Cox MM, Sommer S, Servant P: The deinococcal DdrB protein is involved in an early step of DNA double strand break repair and in plasmid transformation through its single-strand annealing activity. *DNA Repair (Amst)* 2011, **10**:1223–1231.
- Gutsche I, Vujcic-Zagar A, Siebert X, Servant P, Vannier F, Castaing B, Gallet B, Heulin T, de Groot A, Sommer S, Serre L: Complex oligomeric structure of a truncated form of DdrA: a protein required for the extreme radiotolerance of *Deinococcus*. *Biochim Biophys Acta* 2008, **1784**:1050–1058.
- Harris DR, Tanaka M, Saveliev SV, Jolivet E, Earl AM, Cox MM, Battista JR: Preserving genome integrity: the DdrA protein of *Deinococcus radiodurans* R1. *PLoS Biol* 2004, **2**:e304.
- Jolivet E, Lecointe F, Coste G, Satoh K, Narumi I, Bailone A, Sommer S: Limited concentration of RecA delays DNA double-strand break repair in *Deinococcus radiodurans* R1. *Mol Microbiol* 2006, **59**:338–349.
- Narumi I, Satoh K, Cui S, Funayama T, Kitayama S, Watanabe H: PprA: a novel protein from *Deinococcus radiodurans* that stimulates DNA ligation. *Mol Microbiol* 2004, **54**:278–285.
- Norais CA, Chitteni-Pattu S, Wood EA, Inman RB, Cox MM: DdrB protein, an alternative *Deinococcus radiodurans* SSB induced by ionizing radiation. *J Biol Chem* 2009, **284**:21402–21411.
- Sugiman-Marangos S, Junop MS: The structure of DdrB from *Deinococcus*: a new fold for single-stranded DNA binding proteins. *Nucleic Acids Res* 2010, **38**:3432–3440.
- Xu G, Lu H, Wang L, Chen H, Xu Z, Hu Y, Tian B, Hua Y: DdrB stimulates single-stranded DNA annealing and facilitates RecA-independent DNA repair in *Deinococcus radiodurans*. *DNA Repair (Amst)* 2010, **9**:805–812.
- Cox MM, Battista JR: *Deinococcus radiodurans* - the consummate survivor. *Nat Rev Microbiol* 2005, **3**:882–892.
- Levin-Zaidman S, Englander J, Shimoni E, Sharma AK, Minton KW, Minsky A: Ringlike structure of the *Deinococcus radiodurans* genome: a key to radioresistance? *Science* 2003, **299**:254–256.
- Daly MJ, Gaidamakova EK, Matrosova VY, Kiang JG, Fukumoto R, Lee DY, Wehr NB, Viteri GA, Berlett BS, Levine RL: Small-molecule antioxidant proteome-shields in *Deinococcus radiodurans*. *PLoS One* 2010, **5**:e12570.
- Daly MJ, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Leapman RD, Lai B, Ravel B, Li SM, Kemner KM, Fredrickson JK: Protein oxidation implicated as the primary determinant of bacterial radioresistance. *PLoS Biol* 2007, **5**:e92.
- Daly MJ, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Venkateswaran A, Hess M, Omelchenko MV, Kostandaritshes HM, Makarova KS, *et al*: Accumulation of Mn(II) in *Deinococcus radiodurans* facilitates gamma-radiation resistance. *Science* 2004, **306**:1025–1028.
- Blasius M, Sommer S, Hubscher U: *Deinococcus radiodurans*: what belongs to the survival kit? *Crit Rev Biochem Mol Biol* 2008, **43**:221–238.
- Armengaud: Systems biology and synthetic biology: understanding biological complexity on the critical path to personalized medicine. *Curr Pharmacogenomics & Personalized Medicine* 2010, **8**:257–267.

27. Armengaud J: Proteogenomics and systems biology: quest for the ultimate missing parts. *Expert Rev Proteomics* 2010, **7**:65–77.
28. Zhang C, Wei J, Zheng Z, Ying N, Sheng D, Hua Y: Proteomic analysis of *Deinococcus radiodurans* recovering from gamma-irradiation. *Proteomics* 2005, **5**:138–143.
29. Lu H, Gao G, Xu G, Fan L, Yin L, Shen B, Hua Y: *Deinococcus radiodurans* Ppr1 switches on DNA damage response and cellular survival networks after radiation damage. *Mol Cell Proteomics* 2009, **8**:481–494.
30. Earl AM, Mohundro MM, Mian IS, Battista JR: The IrrE protein of *Deinococcus radiodurans* R1 is a novel regulator of recA expression. *J Bacteriol* 2002, **184**:6216–6224.
31. Hua Y, Narumi I, Gao G, Tian B, Satoh K, Kitayama S, Shen B: Ppr1: a general switch responsible for extreme radioresistance of *Deinococcus radiodurans*. *Biochem Biophys Res Commun* 2003, **306**:354–360.
32. Vujicic-Zagar A, Dulermo R, Le Gorrec M, Vannier F, Servant P, Sommer S, de Groot A, Serre L: Crystal structure of the IrrE protein, a central regulator of DNA damage repair in deinococcaeae. *J Mol Biol* 2009, **386**:704–716.
33. Liedert C, Bernhardt J, Albrecht D, Voigt B, Hecker M, Salkinoja-Salonen M, Neubauer P: Two-dimensional proteome reference map for the radiation-resistant bacterium *Deinococcus geothermalis*. *Proteomics* 2010, **10**:555–563.
34. Tian B, Wang H, Ma X, Hu Y, Sun Z, Shen S, Wang F, Hua Y: Proteomic analysis of membrane proteins from a radioresistant and moderate thermophilic bacterium *Deinococcus geothermalis*. *Mol Biosyst* 2010, **6**:2068–2077.
35. Griffiths E, Gupta RS: Identification of signature proteins that are distinctive of the *Deinococcus-Thermus* phylum. *Int Microbiol* 2007, **10**:201–208.
36. Makarova KS, Ormelchenko MV, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Lapidus A, Copeland A, Kim E, Land M, et al: *Deinococcus geothermalis*: the pool of extreme radiation resistance genes shrinks. *PLoS One* 2007, **2**:e955.
37. Basu B, Apte SK: Gamma radiation-induced proteome of *Deinococcus radiodurans* primarily targets DNA repair and oxidative stress alleviation. *Mol Cell Proteomics* 2012, **11**:M111–011734.
38. Dulermo R, Fochesato S, Blanchard L, de Groot A: Mutagenic lesion bypass and two functionally different RecA proteins in *Deinococcus deserti*. *Mol Microbiol* 2009, **74**:194–208.
39. Bentchikou E, Servant P, Coste G, Sommer S: A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in *Deinococcus radiodurans*. *PLoS Genet* 2010, **6**:e1000774.
40. Bharat A, Jiang M, Sullivan SM, Maddock JR, Brown ED: Cooperative and critical roles for both G domains in the GTPase activity and cellular function of ribosome-associated *Escherichia coli* EngA. *J Bacteriol* 2006, **188**:7992–7996.
41. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D: The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 1993, **14**:1023–1031.
42. Zhu K, Zhao J, Lubman DM, Miller FR, Barder TJ: Protein pI shifts due to posttranslational modifications in the separation and characterization of proteins. *Anal Chem* 2005, **77**:2745–2755.
43. Baudet M, Ortel P, Gaillard JC, Fernandez B, Guerin P, Enjalbal C, Subra G, de Groot A, Barakat M, Dedieu A, Armengaud J: Proteomics-based refinement of *Deinococcus deserti* genome annotation reveals an unwanted use of non-canonical translation initiation codons. *Mol Cell Proteomics* 2010, **9**:415–426.
44. Plevoda B, Sherman F: The diversity of acetylated proteins. *Genome Biol* 2002, **3**:reviews0006.
45. Plevoda B, Sherman F: Nalpha-terminal acetylation of eukaryotic proteins. *J Biol Chem* 2000, **275**:36479–36482.
46. Persson B, Flinta C, von Heijne G, Jornvall H: Structures of N-terminally acetylated proteins. *Eur J Biochem* 1985, **152**:523–527.
47. Bradshaw RA, Brickey WW, Walker KW: N-terminal processing: the methionine aminopeptidase and N-alpha-acetyl transferase families. *Trends Biochem Sci* 1998, **23**:263–267.
48. Tanaka S, Matsushita Y, Yoshikawa A, Isono K: Cloning and molecular characterization of the gene rim1 which encodes an enzyme acetylating ribosomal protein L12 of *Escherichia coli* K12. *Mol Gen Genet* 1989, **217**:289–293.
49. Kuo MH, Allis CD: Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* 1998, **20**:615–626.
50. Hwang CS, Shemorry A, Varshavsky A: N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* 2010, **327**:973–977.
51. Joshi B, Schmid R, Altendorf K, Apte SK: Protein recycling is a major component of post-irradiation recovery in *Deinococcus radiodurans* strain R1. *Biochem Biophys Res Commun* 2004, **320**:1112–1117.
52. Servant P, Jolivet E, Bentchikou E, Mennecier S, Bailone A, Sommer S: The ClpXP protease is required for radioresistance and regulates cell division after gamma-irradiation in *Deinococcus radiodurans*. *Mol Microbiol* 2007, **66**:1231–1239.
53. Neher SB, Villen J, Oakes EC, Bakalarski CE, Sauer RT, Gygi SP, Baker TA: Proteomic profiling of ClpXP substrates after DNA damage reveals extensive instability within SOS regulon. *Mol Cell* 2006, **22**:193–204.
54. Pruteanu M, Baker TA: Controlled degradation by ClpXP protease tunes the levels of the excision repair protein UvrA to the extent of DNA damage. *Mol Microbiol* 2009, **71**:912–924.
55. Burns WG, Sims HE: Effect of radiation type in water radiolysis. *J Chem Soc, Faraday Trans 1* 1981, **77**:2803–2813.
56. Fredrickson JK, Li SM, Gaidamakova EK, Matrosova VY, Zhai M, Sulloway HM, Scholten JC, Brown MG, Balkwill DL, Daly MJ: Protein oxidation: key to bacterial desiccation resistance? *ISME J* 2008, **2**:393–403.
57. Barak R, Eisenbach M: Co-regulation of acetylation and phosphorylation of CheY, a response regulator in chemotaxis of *Escherichia coli*. *J Mol Biol* 2004, **342**:375–381.
58. Rudolph J, Tolliday N, Schmitt C, Schuster SC, Oesterheld D: Phosphorylation in halobacterial signal transduction. *EMBO J* 1995, **14**:4249–4257.
59. Wolfgang M, Lauer P, Park HS, Brossay L, Hebert J, Koomey M: Pili mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. *Mol Microbiol* 1998, **29**:321–330.
60. Palchevskiy V, Finkel SE: *Escherichia coli* competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *J Bacteriol* 2006, **188**:3902–3910.
61. Lee YH, Chirikjian JG: Sequence-specific endonuclease Bgl I. Modification of lysine and arginine residues of the homogeneous enzyme. *J Biol Chem* 1979, **254**:6838–6841.
62. Newman M, Lunnen K, Wilson G, Greci J, Schildkraut I, Phillips SE: Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. *EMBO J* 1998, **17**:5466–5476.
63. Koch HG, Moser M, Muller M: Signal recognition particle-dependent protein targeting, universal to all kingdoms of life. *Rev Physiol Biochem Pharmacol* 2003, **146**:55–94.
64. Celis RT, Leadlay PF, Roy I, Hansen A: Phosphorylation of the periplasmic binding protein in two transport systems for arginine incorporation in *Escherichia coli* K-12 is unrelated to the function of the transport system. *J Bacteriol* 1998, **180**:4828–4833.
65. Bailey TL, Elkan C: Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994, **2**:28–36.
66. Meunier B, Dumas E, Pic I, Bechet D, Hebraud M, Hocquette JF: Assessment of hierarchical clustering methodologies for proteomic data mining. *J Proteome Res* 2007, **6**:358–366.
67. Caraux G, Pinloche S: PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 2005, **21**:1280–1281.
68. Gabant G, Augier J, Armengaud J: Assessment of solvent residues accessibility using three Sulfo-NHS-biotin reagents in parallel: application to footprint changes of a methyltransferase upon binding its substrate. *J Mass Spectrom* 2008, **43**:360–370.
69. Clair G, Roussi S, Armengaud J, Duport C: Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions. *Mol Cell Proteomics* 2010, **9**:1486–1498.

doi:10.1186/1477-5956-11-3
Cite this article as: Dedieu et al.: Major soluble proteome changes in *Deinococcus deserti* over the earliest stages following gamma-ray irradiation. *Proteome Science* 2013 **11**:3.

Supplementary data available at:

<http://www.proteomesci.com/content/11/1/3>

3.2.2 Conclusions et perspectives

De nombreux articles ont tenté de comprendre quels étaient les mécanismes qui permettaient aux organismes radio-resistants de reconstituer leurs ADN totalement détruites après une irradiation intense (tant pour *Deinococcus radiodurans* que *Deinococcus deserti*). Comme attendu nous retrouvons dans notre étude des protéines capables de réparer l'ADN, telles DdrB, SSB, deux RecA différentes (RecAp et RecAc), PprA, DdrD et deux sous-unités de gyrase (GyrA et GyrB).

Mais de façon moins attendue nous retrouvons parmi les protéines induites une N-acétyltransferase putative et des gyrases modifiées pour lesquelles la méthionine N-terminale a disparu et qui ont été acétylées sur le nouveau résidu N-terminal. On peut se demander quelle signification exacte porte cette acétylation mais de façon plus générale s'interroger sur un éventuel rôle de modifications post-traductionnelles dans l'adaptation à des stress extrêmes sur des organismes modèles tel *Deinococcus deserti*.

3.3 - *Deinococcus deserti* une bactérie aux acétylations N-terminales abondantes et caractéristiques

3.3.1 Changements majeurs intervenant dans le protéome soluble de *Deinococcus deserti* aux premières heures suivant une forte irradiation par rayons gamma

Dans la publication précédente nous avons découvert au cours d'une étude portant sur les adaptations du protéome soluble de *Deinococcus deserti*, aux fortes irradiations, la présence d'une protéine N-terminal acétylée (Gyrase) et d'une N-acétyltransférase putative. Les acétylation N-terminales se rencontrent très fréquemment chez les eucaryotes (80-90% des protéines cytosoliques des mammifères sont acétylés, jusqu'à 50% des protéines solubles de la levure. Par ailleurs chez les Archées dès 2006, une première publication a indiqué une présence notable de cette modification alors que jusqu'à présent elle a rarement été décrite chez les bactéries.

Nous avons exploré cette modification post-traductionnelle dans une ancienne lignée bactérienne, le phylum *Deinococcus-Thermus*, afin d'en démontrer sa présence et suivre sa trace dans un des phylums bactériens les plus anciens.

The frequent N-terminal protein acetylation in *Deinococcus deserti* highlights striking differences for protein maturation between *Deinococcales* and *Thermales* orders.

Alain Dedieu¹, Jimmy Mewaere¹, Jean-Charles Gaillard, Bernard Fernandez¹, Arjan de Groot^{2,3,4}, Christian Roumestand⁵, Jean Armengaud^{1#}

¹CEA, DSV, IBEB, Lab Biochim System Perturb, Bagnols-sur-Cèze, F-30207, France.

²CEA, DSV, IBEB, Lab Ecol Microb Rhizosphere & Environ Extrem (LEMIRE), Saint-Paul-lez-Durance, F-13108, France.

³CNRS, UMR 6191, Biol Veget & Microbiol Environ, Saint-Paul-lez-Durance, F-13108, France. ⁴Aix-Marseille Université, Saint-Paul-lez-Durance, F-13108, France.

⁵Centre de Biochimie Structurale, INSERM U1054, CNRS UMR 5048, Universités de Montpellier, France.

Corresponding author. Jean Armengaud, Laboratoire de Biochimie des Systèmes Perturbés, CEA Marcoule, DSV, iBEB, SBTN, LBSP, F-30207 BAGNOLS-SUR-CEZE, France. Email : jean.armengaud@cea.fr; Tel: +33 (0)4 66 79 68 02; Fax: +33 (0)4 66 79 19 05.

Running title: *D. deserti* N-terminal protein acetylation

Key words: proteome, post-translational modification, acetylation, protein translation, initiation of translation

Total number of words: 3514

Abstract

N-terminal acetylation is common in eukaryotes and also found in archaea and bacteria, but its exact role remains elusive. Here, we explored this post-translational modification in an ancient bacterial lineage, the *Deinococcus-Thermus* phylum, in order to trace it back during Life evolution. A large-scale survey of N-terminal protein maturation in *Deinococcus deserti*, a gram-negative bacterium which has been isolated from the surface sands sampled from Sahara desert, led to the detection of a total of 1835 proteins, representing 53 % of its theoretical annotated proteome. We found that a remarkably high percentage of N-terminal peptides, *i.e.* 50 %, are acetylated on their N-terminal free amine. Strikingly, 84% of the corresponding acetylated amino acids are threonine, 10% serine, and 5% methionine. Specific traits of this ancient bacterial order in terms of protein N-terminus and acetylation were evaluated by comparative genomics. Pattern for the first residues in the whole proteome of 12 sequenced representatives of *Deinococcus-Thermus* phylum indicates that this specific pattern of acetylation in *D. deserti* is probably conserved among all *Deinococcales*, but not *Thermales*.

Introduction

Maturation of newly synthesized proteins is common to all living organisms (Hartmann and Armengaud 2014). Co-translational cleavage of the initiator methionine by the methionine aminopeptidase (MAP) is universal as seen through its occurrence in all prokaryotic and eukaryotic cells. Acetylation of the N-terminal residue of proteins by N-terminal acetyl transferases (NATs) is another well described post-translational modification (Polevoda and Sherman 2000) (Bradshaw, Brickey et al. 1998). It may occur either on the initial methionine or on the penultimate residue when the methionine is removed. This modification is frequent in yeasts and mammals, as well as other eukaryotes, occurring in 80-90% of cytosolic mammalian proteins and 50% of proteins in *Saccharomyces cerevisiae* (Polevoda and Sherman 2003). In *S. cerevisiae* three proteins are involved in N-terminal acetylation: NatA which acetylates N-termini with sequences beginning with Ser, Ala, Gly, and Thr (Driessen, de Jong et al. 1985; Polevoda and Sherman 2003); NatB which acetylates Met residues followed by a bulky hydrophobic amino-acid; and NatC specific for the N-terminal Met-Glu and Met-Asp sequences (Polevoda and Sherman 2003). N-terminal acetylation in eukaryotes occurs cotranslationally, *i.e.* when the N-terminus of the protein leaves the ribosome and is considered completed when the first 40-50 first amino acids of the polypeptide have been synthesised (Polevoda and Sherman 2002; Spange, Wagner et al. 2009). In some archaea, such as *Halobacterium salinarum* and *Natronomonas pharaonis*, this modification occurs in about 15% of the proteins, with a specific pattern of N-terminal acetylation of serine and alanine residues (Falb, Aivaliotis et al. 2006). In bacteria, N-terminal acetylation was considered as rare (Soppa 2010), but recent reports have shown their occurrence in some bacteria (Bonissone, Gupta et al. 2013). Bonissone et al. reported an interesting link between N-terminal methionine excision of Ala and Ser residues and their specific acetylation through a survey of the proteomes of 57 species (Bonissone, Gupta et al. 2013).

Deinococcus deserti is a gram-negative bacteria which was isolated from surface sands sampled in Sahara (de Groot, Chapon et al. 2005). This bacterium was shown to be particularly tolerant to gamma and UV radiation, as well as desiccation. Its genome was determined and carefully annotated by means of proteogenomics studies (de Groot, Dulermo et al. 2009; Baudet, Ortet et al. 2010). It belongs to the *Deinococcus-Thermus* phylum, which is considered as the most ancient bacterial lineage with no clear affinities to any of the other known bacterial lineages (Makarova, Aravind et al. 2001). *Deinococcus* and *Thermus* bacteria are extremophiles, the former being well known as radiation resistant while the later are thermophiles. A comparative genomics study has shown the divergent routes of adaptation to thermophily and radiation resistance of these two types of bacteria (Omelchenko, Wolf et al. 2005).

Here, we report a large-scale experimental survey of N-terminal protein acetylation in *Deinococcus deserti*. Fractionation of peptides generated by trypsin proteolysis of soluble proteins from cells harvested at two different growth phases was carried out by means of OFFGEL electrophoresis and strong cation chromatographic exchange, followed by high-throughput tandem mass spectrometry. Amongst the 1835 proteins identified, we found that a remarkably high percentage of N-terminal peptides, *i.e.* 50%, are N-terminally acetylated. N-terminus acetylated peptides are mainly located on threonine (84 %), serine (10 %), and methionine (5 %). We

discussed this specific pattern in the *Deinococcus* phylum by means of comparative genomics in the light of the early evolution of the translation machinery.

Materials and Methods

***D. deserti* cellular growth conditions and preparation of protein extracts.** *D. deserti* was grown in 10-fold diluted tryptic soy broth (TSB/10) supplemented with trace elements [de Groot et al., 2009]. Cells were harvested at the exponential and stationary phases by centrifugation, washed twice with 50 mM TRIS-HCl pH 8.0, frozen in liquid nitrogen and stored at -80°C . For each condition, 3 g of wet cellular material were resuspended in 15 mL of 100 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ buffer containing a protease inhibitor cocktail Complete EDTA-free (Roche, 1 tablet per 50 ml), phosphatase inhibitor cocktails 2 and 3 (Sigma-Aldrich, 1/100), benzonase nuclease (Sigma-aldrich, 1/1000), and 1mM MgCl_2 . Cells were disrupted with an ultrasound Vibracell 75042 sonicator (Fisher Bioblock Scientific) for 35 s (7 cycles of 5 sec pulse and 5 sec pause). The extracts were then centrifuged for 20 min at 20,000 g and 4°C to remove cellular debris. Protein concentration from the supernatants was determined by a Bradford assay (Biorad) using bovine-globulin as standard. Proteins were precipitated using a 4:1 methanol/chloroform solution (vol/vol). Typically, 100 μL of protein sample containing 1 mg of proteins were mixed with 500 μL of the precipitant solution, thoroughly mixed, and centrifuged at 20,000 g for 2 min at room temperature. The resulting supernatant was then discarded, while the pellet was washed with 300 μL methanol, and centrifuged. Finally, proteins from dried pellets were dissolved in 200 μL of 50 mM NH_4HCO_3 .

Trypsin proteolysis. Proteins (1mg) were reduced for 45 min at 56°C by adding DTT (10 mM final). Then, alkylation was carried out at room temperature for 15 min in the dark with the addition of iodoacetamide (55 mM final). Sequence grade trypsin (Roche) was diluted in 0.01% TFA to get a solution at 0.5 $\mu\text{g}/\mu\text{L}$. The soluble protein sample (200 μL) was overnight proteolyzed at 37°C in presence of 1M urea with 40 μL of this solution. Proteolysis was stopped by adding 20 μL of 10% TFA aqueous solution. The digest was desalted by using ZipTip C₁₈ (Millipore) and then dried under vacuum (Speed Vac, Savant Thermo Fisher Scientific).

Peptide fractionation by OFFGEL electrophoresis. Peptides (1 mg) were resolved on a 3100 OFFGEL Fractionator (Agilent Technologies) with a pH 3–10 OFFGEL kit and a 24-well tray. The 24 cm long IPG gel strip was rehydrated for 15 min with 20 μL of rehydration solution per well. The desalted tryptic digest was diluted to 3.6 mL with the addition of OFFGEL peptide fractionation buffer. Volumes of 150 μL were introduced into each well. Electrophoresis was carried out for approximately 24 h until 50 kVh was reached. Solutions (100 to 200 μL per well) were removed from each well, resulting in 24 complex peptide fractions. These were evaporated to 50 μL and then, desalted with SpinC18 cartridges (Harvard).

Peptide fractionation by SCX chromatography. Peptides (1 mg) were dissolved in 60 μL of 5 mM ammonium formate containing 30 % acetonitrile (can). They were fractionated using a polysulfoethyl-A column (3.2 x 200 mm; 5 μm , 200 \AA) from PolyLC (Columbia, MD, USA) on a Agilent 1100 HPLC system. Buffer A (5 mM ammonium formate in 30% ACN, pH 2.6) and Buffer B (500 mM ammonium formate in 30% ACN, pH 3.0) were used to establish a 55 min gradient of 0% buffer B for 5 min, 0%–100% buffer B for 50 min, followed by 15 min at 100%

buffer B at a flow rate of 1 mL/min. A total of 20 fractions were collected and then evaporated to 50 μ L via vacuum centrifuge. The samples were desalted with SpinC18 cartridges (Harvard).

Tandem mass spectrometry. Peptide fractions were dried under vacuum (Speed Vac) and dissolved in 100 μ L of 0.1% TFA. NanoLC-MS/MS analysis of each peptide fraction was carried out with a LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings). The peptide mixtures (0.1–2 pmol) were loaded and desalted online in an LC Packings Acclaim Pepmap 100 C18 reversed-phase precolumn (5 μ m bead size, 100-Å pore size, 5 mm \times 300 μ m). They were resolved on an LC Packings nanoscale Acclaim Pepmap 100 C18 (3 μ m bead size, 100-Å pore size, 15 cm \times 75 μ m) at a flow rate of 0.3 μ L/min. The peptides were separated using a 60-min gradient (5 to 60% solvent B) with aqueous solvent A (0.1% HCOOH) and solvent B (0.1% HCOOH, 80% CH₃CN). The column was then washed for 10 min with 100% B and re-equilibrated for 20 min with 5% B. Full-scan mass spectra were measured from m/z 300 to 1800. The LTQ-Orbitrap XL mass spectrometer was run in data-dependent mode using TOP5 strategy with a Fourier transform mass spectrometer resolution set at 30,000 as previously reported (Hartmann, Allain et al. 2014). A scan cycle was started with a full scan of high mass accuracy in the Orbitrap, followed by MS/MS scans in the linear ion trap on the 5 most abundant precursor ions with dynamic exclusion of previously selected ions. This dynamic exclusion consisted in two acquisitions of MS/MS spectra of the most abundant ion over 30 s and then excluding this ion for the fragmentations over the next 60 s. The activation type normally used was CID with a standard normalized collision energy set at 30.

Database Mining. Peak lists were generated using version 2.3.2 MASCOT DAEMON (Matrix Science) and the data import filter version 2.0.7 extract_msn.exe from Xcalibur FT (Thermo). Filtering options were set at: 0 (grouping tolerance), 0 (intermediate scans), 400 (minimum mass), 5000 (maximum mass), and 1000 (threshold). MS/MS spectra were searched with MASCOT (2.2.06) search engine (Matrix Science) against a homemade *D. deserti* protein sequence database which has been previously described (Touelle, Mirabella et al. 2012). The following parameters were used for this search: 5 ppm and 0.5 Da for mass tolerances on the parent ion and the MS/MS, respectively, static modifications of carbamidomethylated Cys (+57.0215), dynamic modification of oxidized Met (+15.9949), N-terminal protein acetylation (+42.0105), and a maximum of two missed cleavages. Peptide matches were filtered with IRMa 1.28.0 software (Dupierris, Masselon et al. 2009) with a stringent peptide identity threshold set at $p < 0.001$. Protein identifications were validated when at least two different peptides were observed. The false-positive rate for protein identification was estimated with the corresponding decoy database to be below 0.02% with these parameters.

Comparative genomics of *Deinococcus-Thermus*. The theoretical proteomes of *Deinococcus deserti* VCD115 (NC_012526; NC_012527; NC_012528; NC_012529), *Deinococcus geothermalis* DSM 11300, (NC_008010; NC_008025; NC_009939), *Deinococcus maricopensis* DSM 21211 (NC_014958), *Deinococcus proteolyticus* MRP (NC_015161; NC_015162; NC_015163; NC_015169; NC_015170), *Deinococcus radiodurans* R1 (NC_000958; NC_000959; NC_001263; NC_001264), *Marinithermus hydrothermalis* DSM 14884 (NC_015387), *Meiothermus ruber* DSM 1279 (NC_013946), *Meiothermus silvanus* DSM 9946

(NC_014212; NC_014213; NC_014214), *Oceanithermus profundus* DSM 14977 (NC_014753; NC_014761), *Thermus scotoductus* SA-01 (NC_014974; NC_014975), *Thermus thermophilus* HB27 (NC_005835, NC_005838), *Thermus thermophilus* HB8 (NC_006461; NC_006462; NC_006463), and *Truepera radiovictrix* DSM 17093 (NC_014221) were download from NCBI (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/). The whole set of presently annotated proteins in these 12 microorganisms comprised 36,536 polypeptide sequences. Search for type of the second residue and count were done with Excel.

Results & Discussion

Strategy for an extensive analysis of the acetylated N-terminome of *D. deserti*. In order to extensively cover the *D. deserti* N-terminome, we analyzed by tandem mass spectrometry the soluble protein content of cells harvested at the exponential and stationary phases. As shown in **Figure 1**, the peptide mixtures resulting from trypsin digestion of these proteomes were fractionated using either OFFGEL electrophoresis or SCX chromatography. OFFGEL electrophoresis allows resolving acetylated peptides along their pI, which can vary over a

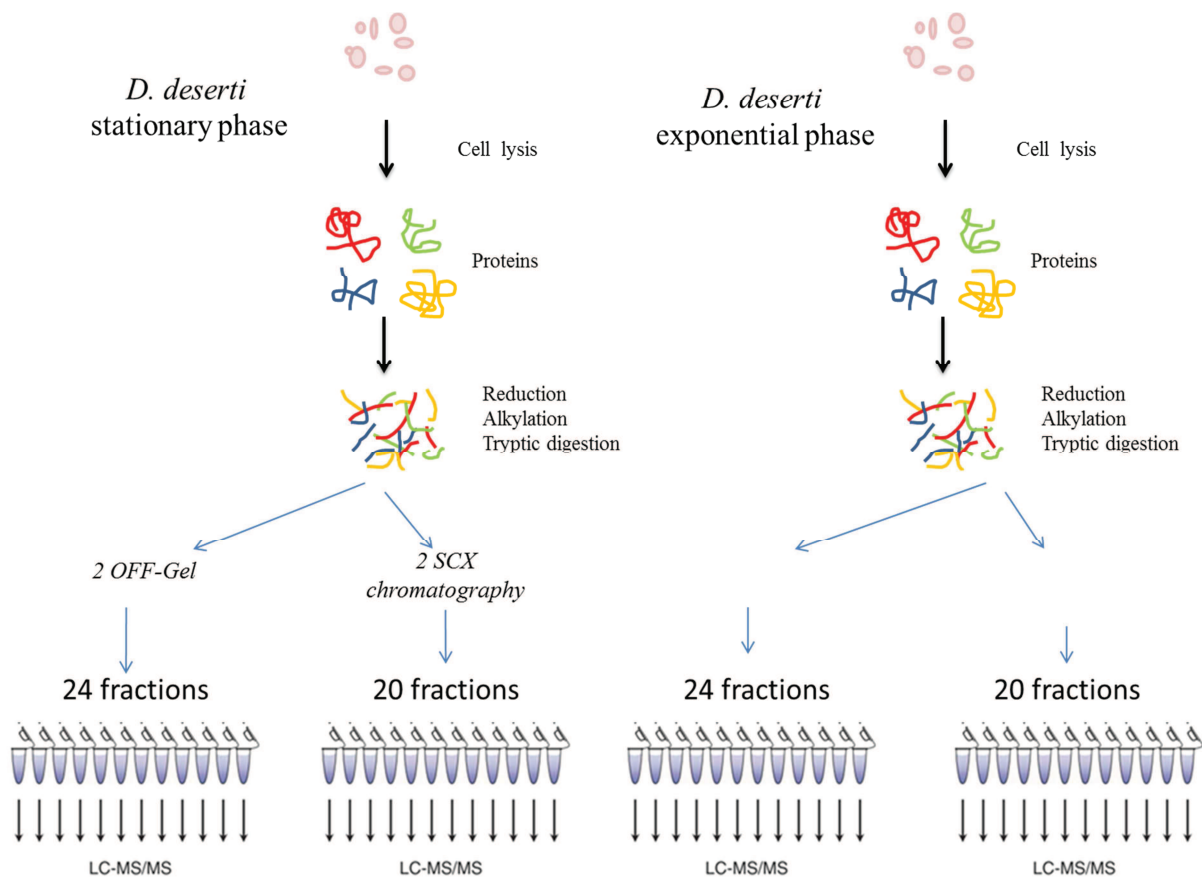


Figure 1 : Strategy for the extensive analysis of post-translational modifications of *D. deserti* N-terminome. Soluble proteins were extracted from cells harvested at the exponential and stationary phases. Peptide mixtures were fractionated by means of OFFGEL electrophoresis and SCX chromatography, resulting in 24 and 20 fractions, respectively. Technical duplicates were carried out for each methodological approach, resulting in a total of 176 peptide fractions.

wide range, but does not specifically enrich them compared to non-acetylated peptides. On the contrary, SCX chromatography allows a significant enrichment of N-acetylated peptides based on the charge of the peptides when eluted at low pH. The dual fractionation was performed systematically twice (technical duplicates) for each of the cellular condition assayed, resulting in a total of 176 peptide fractions that were all resolved by nanoHPLC and analyzed by high-throughput tandem mass spectrometry. Amongst the 1160479 MS/MS spectra recorded, 293779 (25%) could be assigned with stringent parameters (p value below 0.001), giving a list of 17461 unique peptide sequences (**Supplementary data, Table S1**). The choice of the stringent threshold used here was dictated by the high level of confidence required for certifying the presence of an acetylation as post-translational modification. A total of 1835 proteins could be validated with at least 2 peptides uncovering 52% of the theoretical proteome of *D. deserti* (**Supplementary data, Table S2**). Most proteins (1481, 81%) were identified in fractions from both fractionation methods. A total of 64 proteins were only identified in the SCX peptide fractions while 290 were specific from the OFFGEL peptide fractions. The global sequence coverage obtained in the OFFGEL fractionation is slightly higher compared to the SCX fractionation: 30% instead of 26%.

Search of acetylated N-terminal peptides. Figure 2 shows the MS/MS spectra of a representative peptide [TATATQPLTAAQAIETR] belonging to the Deide_15960 protein (gi|226356574), which is a NADH dehydrogenase. This peptide is detected as either acetylated (**Figure 2, Panel A**) or unacetylated (**Figure 2, Panel B**). Because with our experimental set-up, valuable information within the MS/MS spectra below m/z 150 is not obtained; it is not possible to detect directly the N-terminal residue with or without the acetylated modification. However, the excellent fragmentation obtained, resulting in the production of complete series of y ions, complemented by many b ions, as well as the mass of the parent ion determined in both cases within less than 2 ppm, allow certifying unambiguously their assignments (**Figure 2**). For a long while, differentiation between peptides containing N-terminal acetylation or tri-methylation {Zhang, 2004 #31} was difficult to distinguish because the corresponding masses were too close: 42.010565 Da and 42.04695 Da, respectively. For the doubly charged peptide with an observed m/z at 893.47192 shown in Figure 2 (Panel A) which has a theoretical mass of 1784.926819 (mass tolerance of 1.38 ppm), the mass difference between an acetylated or a tri-methylated peptide will be 0.03639 m/z , resulting in a relative mass difference of 36 ppm far above the mass tolerance limit of 5 ppm used here.

NanoLC-MS/MS analyses uncover a large number of acetylated N-terminal peptides. Amongst the 17461 unique peptide sequences (**Supplementary data, Table S1**), we detected 531 N-terminal peptide sequences (8744 MS/MS spectra). A minority of them, *i.e.* 144, start with a Methionine at position 1, while most, *i.e.* 387, start from the second residue of the theoretical polypeptide sequence as a consequence of the removal of the initiator methionine (**Supplementary data, Table S3**). The high prevalence of methionine excised peptide is remarkable as this was not previously highlighted for any organisms by Bonissone et al. in their large survey of 57 bacterial species (Bonissone, Gupta et al. 2013). The whole set of N-terminal peptides correspond to a total of 490 unique proteins. Only for 20 proteins, both intact methionine and processed methionine N-terminal peptides were observed. We correlated this partial

processing to the higher quantity of these proteins compared to the others assessed by their respective normalized spectral abundance factors (NSAF) calculated with the total spectral counts per protein (**Supplementary data, Table S2**). Indeed, the mean NSAF calculated for these 20 proteins (11.65) is higher than the mean NSAF for the others (5.28). Among the 369 proteins with a processed methionine, a large majority (80 %) starts with either a threonine (220, 60%) or a serine (75, 20%) residues. Other significant amino-acids found as starts are as expected small lateral chain residues: proline (30, 8%), alanine (21, 6%), and leucine (4, 1%). Regarding acetylation, we detected 271 modified N-terminal peptides (51% of the global set of N-terminal peptides) and these were accounting for 49% of the total spectral counts of N-terminal peptides. A total of 19 peptides are acetylated and start with a methionine residue, but the great majority of acetylated peptides are methionine processed. In other words, 14% of the detected peptides with unprocessed initial methionine are acetylated.

Peptides with their initial methionine removed and acetylated on the resulting N-terminus are numerous, *i.e.* 250. The ratio of acetylation for these matured N-terminal peptides is higher (65 %). Among these, 214 start with a threonine (89 %), 31 with a serine (11 %), and only three with an alanine (**Supplementary data, Table S3**). In terms of abundance, the whole set of N-terminal acetylated polypeptides are accounting for 14% of the total proteome based on the NSAF parameter (15% in terms of spectral count).

Specific N-terminal acetylation pattern. The high percentage of acetylated N-termini in *D. deserti* proteome is remarkable, but the quite specific pattern of acetylation on Threonine and

Serine residues are also intriguing. A total of 257 proteins were found N-terminally acetylated, 19 starting with an unprocessed methionine, 202 with a threonine and 31 with a serine after methionine excision. We investigated whether a specific trend in terms of protein N-termini could be observed in *D. deserti* genome and could explain the acetylation strong bias on threonine residues.

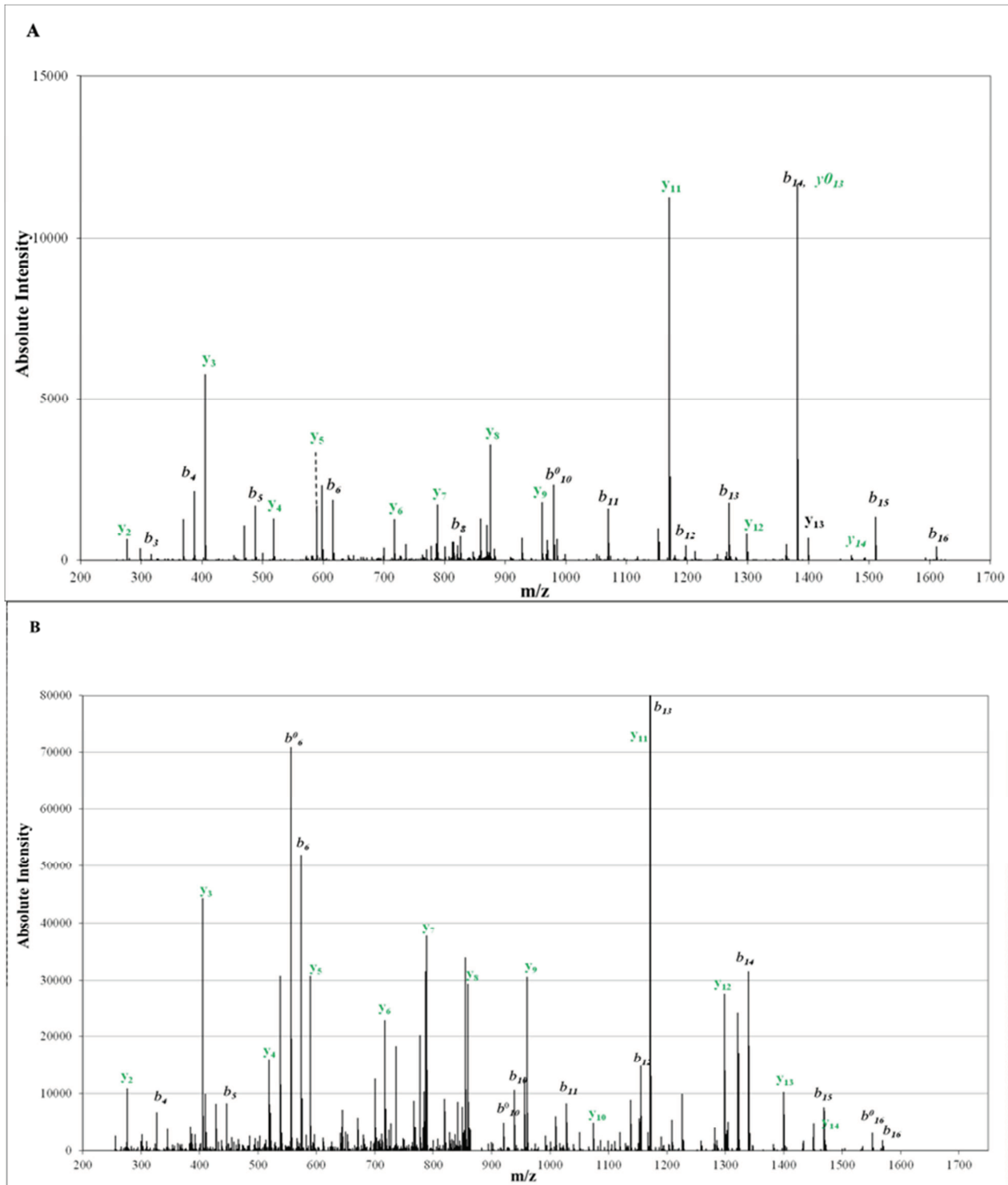


Figure 2: MS/MS spectra of a representative N-terminal peptide in its acetylated (Panel A) and unacetylated (Panel B) forms.

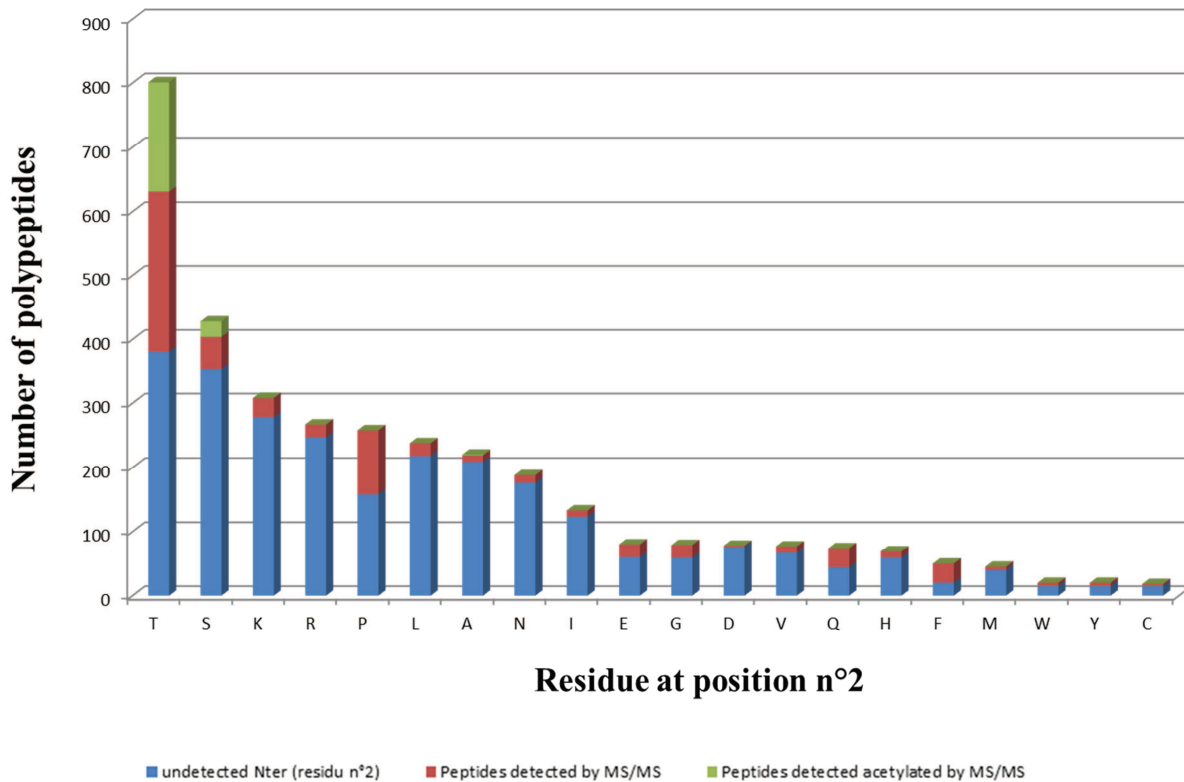


Figure 3: Analysis regarding the second residue found in *D. deserti* annotated polypeptides. The distribution of polypeptides per residue is shown from the most abundant to the least abundant. The corresponding fractions of experimentally detected unacetylated and acetylated peptides are indicated.

Figure 3 shows a whole genome analysis of theoretical N-terminal protein sequence starts, more precisely through the lens of the second residue found after the initial methionine. Amongst the 3459 annotated CDS in *D. deserti* cells, a large majority starts with a threonine (801) or a serine (428), as these polypeptides will have their initial methionine processed. This amounts to 36% of the total number of CDS. Roughly, all of the corresponding N-terminal peptides are detected in our large shotgun survey. **Figure 3** indicates that a specific bias in terms of N-termini sequence exists in favor of threonine and serine in *D. deserti*, but also that the larger ratio of acetylated threonine and serine residues is not directly due to the larger amount of these residues as initial residue but rather to a specificity of the acetylase towards these residues. We checked whether such a trend could be observed for the whole *Deinococcus-Thermus* phylum.

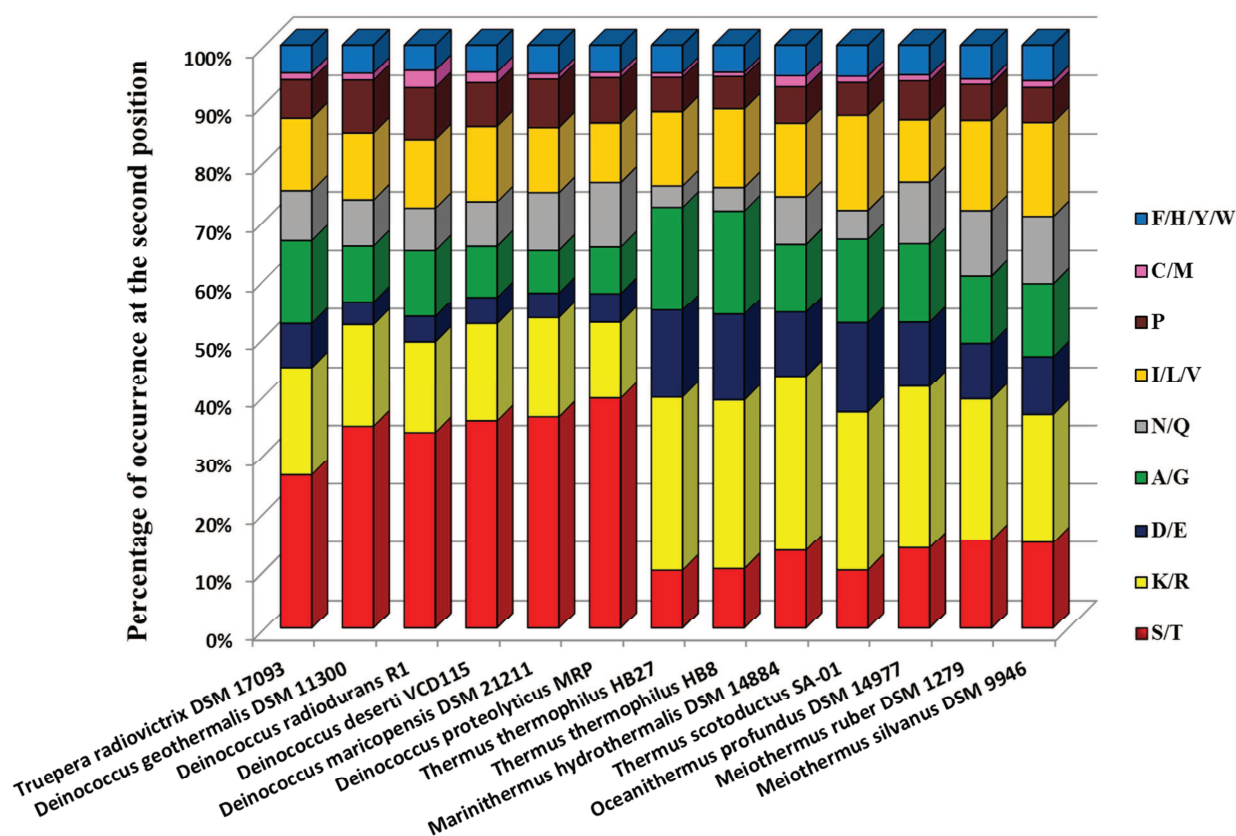


Figure 4: Pattern for the second residue found in annotated polypeptides of 13 sequenced representatives of *Deinococcus-Thermus* phylum. Residues with the same physico-chemical properties are grouped. The species are ranked along their threonine/serine in second position ratio.

Figure 4 shows the ratio of polypeptides with a given amino acid in position 2 in the annotated CDSs of 13 sequenced strains from this phylum. Remarkably, we found that a larger number of polypeptides are starting with a threonine or serine residues amongst *Deinococcus*-related species, ranging between 26% for *Truepera radiovictrix* DSM17093 to 39% for *Deinococcus proteolyticus* MRP, compared to *Thermus*-related species (between 12% for *Marinithermus hydrothermalis* DSM 14884 to 15% for *Meiothermus ruber* DSM 1279). As a comparison, the ratio found for *Escherichia coli* str. K-12 substr. DH10B is 11%. Concomitantly, the ratio of polypeptides starting with alanine or glycine residues, as well as the ratio of polypeptides with the methionine not excised, *i.e.* polypeptides with a charged residue found in second position, are anti-correlated to the ratio of polypeptides starting with threonine or serine residues, whatever the species. As a consequence, proteins with an excised methionine are more frequent in *Deinococcales* than in *Thermales*.

Conclusive remarks. While N-terminal acetylation was considered until recently as rare in bacteria (Soppa 2010), our proteomic data show that acetylated N termini are common in *Deinococcus deserti*, the acetylation affecting about 14% of the total proteome. The significance and function of acetylation in bacteria still remain obscure (Jones and O'Connor 2011), but should be important because of its conservation amongst bacterial phyla. Furthermore, a specific pattern

of N-terminal acetylation was defined in this study with a privileged modification of threonine and serine residues after methionine excision. We have also shown that serine and threonine residues are significantly overrepresented in second position of *D. deserti* annotated polypeptides. This trend is conserved amongst *Deinococcales* but not observed amongst *Thermales*. Because this specific trend is for the first time discovered in an ancient bacterial lineage, its origin and function are interesting open questions.

Acknowledgements

We thank the Commissariat à l’Energie Atomique et aux Energies Alternatives and the Agence Nationale de la Recherche (ANR-07-BLAN-0106-02) for financial support.

REFERENCES

- Baudet, M., P. Ortet, et al. (2010). "Proteomics-based refinement of *Deinococcus deserti* genome annotation reveals an unwonted use of non-canonical translation initiation codons." *Mol Cell Proteomics* **9**(2): 415-426.
- Bonissone, S., N. Gupta, et al. (2013). "N-terminal protein processing: a comparative proteogenomic analysis." *Mol Cell Proteomics* **12**(1): 14-28.
- Bradshaw, R. A., W. W. Brickey, et al. (1998). "N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families." *Trends in Biochemical Sciences* **23**(7): 263-267.
- de Groot, A., V. Chapon, et al. (2005). "*Deinococcus deserti* sp. nov., a gamma-radiation-tolerant bacterium isolated from the Sahara Desert." *Int J Syst Evol Microbiol* **55**(Pt 6): 2441-2446.
- de Groot, A., R. Dulermo, et al. (2009). "Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*." *PLoS Genet* **5**(3): e1000434.
- Driessen, H. P., W. W. de Jong, et al. (1985). "The mechanism of N-terminal acetylation of proteins." *CRC Crit Rev Biochem* **18**(4): 281-325.
- Dupierris, V., C. Masselon, et al. (2009). "A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa." *Bioinformatics* **25**(15): 1980-1981.
- Falb, M., M. Aivaliotis, et al. (2006). "Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey." *J Mol Biol* **362**(5): 915-924.
- Hartmann, E. M., F. Allain, et al. (2014). "Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria." *Methods Mol Biol* **1197**: 275-285.
- Hartmann, E. M. and J. Armengaud (2014). "N-terminomics and proteogenomics, getting off to a good start." *Proteomics*.
- Jones, J. D. and C. D. O'Connor (2011). "Protein acetylation in prokaryotes." *Proteomics* **11**(15): 3012-3022.
- Makarova, K. S., L. Aravind, et al. (2001). "Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics." *Microbiol Mol Biol Rev* **65**(1): 44-79.
- Omelchenko, M. V., Y. I. Wolf, et al. (2005). "Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance." *BMC Evol Biol* **5**: 57.
- Polevoda, B. and F. Sherman (2000). "N-alpha -terminal acetylation of eukaryotic proteins." *J Biol Chem* **275**(47): 36479-36482.
- Polevoda, B. and F. Sherman (2002). "The diversity of acetylated proteins." *Genome Biol* **3**(5): reviews0006.

- Polevoda, B. and F. Sherman (2003). "Composition and function of the eukaryotic N-terminal acetyltransferase subunits." *Biochem Biophys Res Commun* **308**(1): 1-11.
- Polevoda, B. and F. Sherman (2003). "N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins." *J Mol Biol* **325**(4): 595-622.
- Soppa, J. (2010). "Protein acetylation in archaea, bacteria, and eukaryotes." *Archaea* **2010**.
- Spange, S., T. Wagner, et al. (2009). "Acetylation of non-histone proteins modulates cellular signalling at multiple levels." *Int J Biochem Cell Biol* **41**(1): 185-198.
- Touille, M., B. Mirabella, et al. (2012). "A comparative proteomic approach to better define *Deinococcus* nucleoid specificities." *J Proteomics* **75**(9): 2588-2600.

SUPPLEMENTARY DATA*

Table S1. List of MS/MS spectra assigned.

Table S2. List of identified proteins and their respective spectral counts.

Table S3. List of acetylated and unacetylated N-terminal peptides.

*Comme l'article n'est pas encore publié, les tables supplémentaires fichiers volumineux seront remises aux membres du jury par l'intermédiaire d'une clé USB.

3.3.2 Conclusions et perspectives

Notre étude montre que les acétylations N-terminales peu décrites ou considérées comme rares chez les bactéries, sont bien présentes chez *Deinococcus deserti* où elles représentent 14% du protéome total. Le rôle de cette modification si fréquente chez les eucaryotes n'a pas été élucidé à ce jour. Pourtant sa présence dans un phylum si ancien montre qu'elle a été conservée au cours de l'évolution. On peut supposer que ce rôle est important et utiliser cette bactérie si particulière qu'est *Deinococcus deserti* pour le comprendre.

On peut aussi s'interroger sur un rôle éventuel de cette modification lors du stress radiatif.

Enfin, le motif spécifique que l'on met ici pour la première fois en évidence dans cette lignée bactérienne, sur la partie N-terminale des peptides acétylés, doit aussi nous interroger sur son rôle et sa fonction.

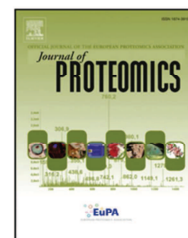
3.4 - N- and O- acétylation les pièges à éviter !

3.4.1 N et O acétylation des résidus thréonine dans le contexte de la protéomique

L'étude précédente met en évidence les nombreuses acétylations des résidus N-terminaux des protéines de *Deinococcus deserti*. Ces acétylations se produisent essentiellement sur des résidus sérine et thréonine qui possèdent à la fois des groupements amine et alcool. L'instabilité de résidus O-acétylés est marquée lorsqu'une fonction amine libre se trouve à proximité. La question que nous avons posé était : les acétylations observées étaient-elles liées à la fonction amine ou bien alcool ? Question d'autant plus pertinente que très récemment des O-acétylations ont été décrites.

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Technical note

N- and O-acetylation of threonine residues in the context of proteomics



Jean-Baptiste Boyer^a, Alain Dedieu^a, Jean Armengaud^a, Pascal Verdié^b, Gilles Subra^b, Jean Martinez^b, Christine Enjalbal^{b,*}

^aCEA, DSV, IBEB, Lab Biochim System Perturb, Bagnols-sur-Cèze F-30207, France

^bInstitut des Biomolécules Max Mousseron (IBMM), UMR 5247, Universités Montpellier 1 et 2, CNRS, Bâtiment Chimie (17), Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

ARTICLE INFO

Article history:

Received 31 March 2014

Accepted 9 June 2014

Keywords:

Protein post-translational acetylation

Threonine acetylation

O-acetyl threonine

O–N acetyl transfer

O–N acyl shift

ABSTRACT

The detection of post-translational modifications (PTMs) of proteins is a matter of intensive research. Among all possible pitfalls that may lead to misidentifications, the chemical stability of modified peptides is scarcely questioned. Global proteomic studies devoted to protein acetylation are becoming popular. Thus, we were concerned about the intrinsic stability of O-acetylated peptides because of the O–N acyl transfer reactivity occurring when an amino moiety is present in the vicinity of the acylated hydroxyl group. Here, the behavior of isomeric O- and N-acetylated, N-terminal threonine-containing peptides was explored in a standard proteomic workflow. We demonstrated a strong chemical instability of O-acetylation, which prevents its detection.

© 2014 Elsevier B.V. All rights reserved.

We were concerned by the chemical stability of peptides issued from O-acetylation of hydroxyl-containing amino acids (mainly serine and threonine) [1–4], because of a well-known reactivity – the O–N acyl transfer – occurring when an amino moiety is present in the vicinity of the acylated hydroxyl group [5]. While this reactivity is an important question in organic chemistry [6–10], it has not yet been discussed in proteomics. Numerous acetylation studies have been reported recently, highlighting the importance of such post-translational modifications in eukaryotes and bacteria [3,11]. As a general viewpoint, O-acetylated serines or threonines could be located at the N-terminal extremity of a polypeptide chain and could also be located at the N-terminus of internal peptides generated by the proteolysis of polypeptides. In particular, any threonine or serine following a

residue of arginine or lysine (four possible (R/K)–(S/T) dipeptide combinations) will provide, upon trypsin digestion, a proteolytic peptide presenting the hydroxyl function next to a free amino group (Fig. 1). Typically, in the human proteome (Uniprot 3A9606), the RT and KT combinations are found 65,640 and 71,894 times, respectively. In these cases, the intrinsic liability of the C–O bond makes the O-acetylation product far less stable than the N-acetylated one, driving the O–N acetyl transfer equilibrium towards the N-acetylation. So far, to our knowledge there is no actual proof from the published analytical data that the site of acetylation under such circumstances is perfectly controlled. As highlighted by recent data [12–15], extreme caution must be undertaken when assigning post-translational modification, which prompted us to carry out further investigations to probe

* Corresponding author. Tel.: +33 467 14 38 19; fax: +33 467 14 48 66.
E-mail address: enjalbal@univ-montp2.fr (C. Enjalbal).

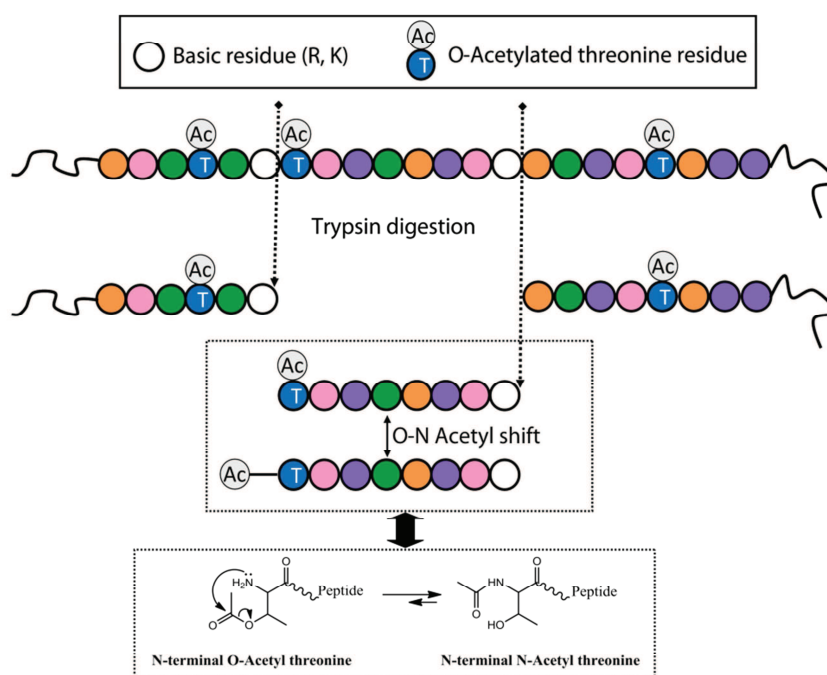


Fig. 1 – O–N acetyl transfer from N-terminal, O-acetylated, threonine-containing tryptic peptides.

whether such acetyl transfer could occur along a proteomic workflow.

We took as a model compound a peptide that was discovered in the laboratory in a large survey of N-terminal-acetylated peptides from a specific bacterium, namely *Deinococcus deserti*. This model sequence (TQLEAVER) exhibited an N-terminal threonine residue. Whether the residue was N- or O-acetylated was open to question. We prepared the N-acetylated peptide (Ac-TQLEAVER) together with its O-acetylated counterpart (T(Ac)QLEAVER) in order to investigate their stabilities, i.e. the intra-molecular inter-conversion of both species, in a classical proteomic protocol. Details of both peptide syntheses are provided in the supplemental information section, sample characterization being displayed in Figs. S1 and S2. Indeed, a proteomic sample is subjected to multiple treatments (protein digestion, sample fractionation, chromatographic separation and mass analysis) with a pH that varies from 8.2 down to 2.7. The two model compounds were thus first conditioned in an aqueous solution of acetonitrile (H₂O at pH 8/CH₃CN, 75/25, v/v). This solution was then acidified by the addition of 0.1% aqueous trifluoroacetic acid to reach a pH of around 2, mimicking the pH range of any proteomic workflow. In addition, short- and middle-term sample stabilities were also evaluated by leaving the peptide solutions at 4 °C, or at room temperature, for several hours or days prior to analysis. These experiments performed on two mass spectrometers (QTOF and LTQ-Orbitrap XL) are described in the supplemental information section. The results are presented in Tables 1 and 2 and in Figs. S3 to S7. NanoLC/ESI(+)-MS/MS conducted on an LTQ-Orbitrap XL mass spectrometer was chosen to probe the stability of the prepared N- and O-acetylated, threonine-containing peptides since the retention times and the MS/MS data of both compounds are known. In accordance with published data [16], only one specific fragment ion featuring O-acetylation and coming from the loss of the

side-chain substituent as a neutral carboxylic acid can be observed provided that the singly protonated ion has been selected as the parent ion in LC/MS/MS, which is not a standard parameter in proteomics. Actually, only di- or tri-charged peptides are systematically selected in current electrospray-based shotgun analysis in order to avoid mass spectrometry time on parent ions that will not generate assignable MS/MS spectra. The dissociation of the two singly charged protonated ions produced from MS/MS analyses of the two regioisomers should give different fragmentation patterns, with the exclusive loss of acetic acid (–60 Da) from the O-acetylated parent ion. The detection in the upper mass range of the MS/MS spectrum of such neutral loss unambiguously indicates the presence of an acetylated threonine side-chain. We have thus implemented this specific analytical parameter (loss of acetic acid from the singly charged molecular ion) to be able to reveal any O- versus N-acetylation, relying also on the known retention times of both isomers. The results obtained in the present study showed that the N-acetylated peptide remained stable throughout (Fig. 2). This was not the case for the O-acetylated peptide, which evolved quickly to the N-acetylated form in blank samples. Finally, the two synthetic peptides were engaged in proteomic experiments by spiking each compound in the *D. deserti* proteome, and processed in parallel without the proteome. The four samples were digested with trypsin following a classical *in-solution* proteolytic protocol. The resulting peptide digests were subjected to either SCX chromatography or OFFGEL electrophoresis to simulate standard fractionation protocols. This resulted in eight samples that were then analyzed by nanoLC/ESI-MS/MS by means of an LTQ-Orbitrap XL mass spectrometer. In all cases, the N-acetylated peptide was always found whereas the O-acetylated peptide was not detected, thus indicating that N-terminal O-acetylated threonine is not stable in the proteomics workflow.

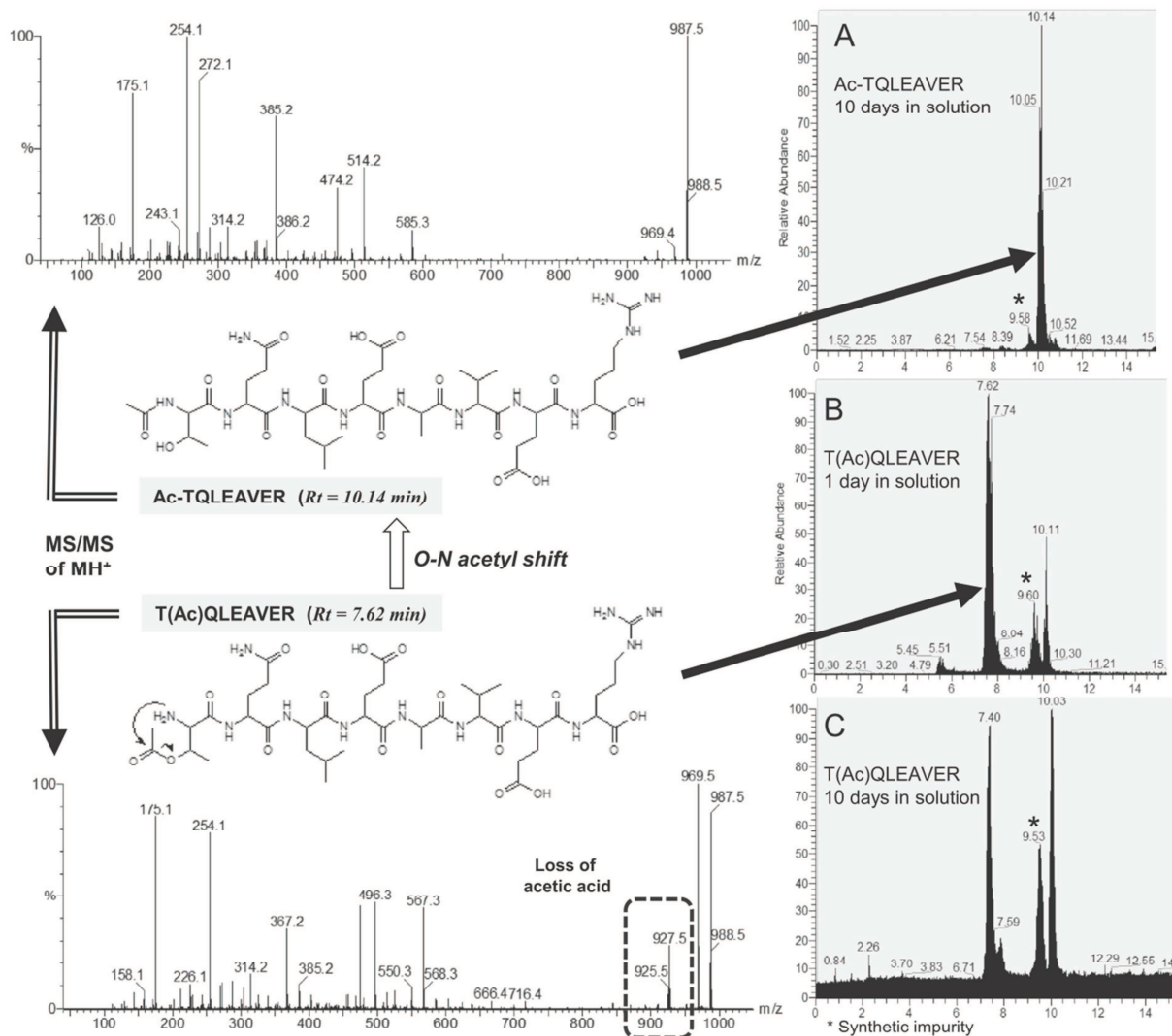


Fig. 2 – Synthetic samples of Ac-TQLEAVER and T(Ac)QLEAVER left in solution before nanoLC/MS/MS analyses.

Since, in our hands, the N-terminal O-acetyl threonine was too labile to be detected in proteomic samples, the description of N-acetylated peptides that bear a threonine at the N-terminal position of a peptide sequence must be taken with caution. One cannot be sure that the original acetylation location was on the N-terminal amino group. It could also originate from an O-acetylation that is completely converted during the experiment into the N-acetylated peptide.

Conflict of interest

All authors declare that there is no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jprot.2014.06.005>.

REFERENCES

- [1] Mukherjee S, Keitany G, Li Y, Wang Y, Ball HL, Goldsmith EJ, et al. *Yersinia YopJ* acetylates and inhibits kinase activation by blocking phosphorylation. *Science* 2006;312:1211–4.
- [2] Mittal R, Peak-Chew S-Y, McMahon HT. Acetylation of MEK2 and I κ B kinase (IKK) activation loop residues by YopJ inhibits signaling. *Proc Natl Acad Sci U S A* 2006;103:18574–9.
- [3] Mukherjee S, Hao YH, Orth K. A newly discovered post-translational modification—the acetylation of serine and threonine residues. *Trends Biochem Sci* 2007;32:210–6.
- [4] Li Y, Ball HL. Study of acetylation on Ser/Thr/Tyr/Lys and trimethylation on Lys using electrospray tandem mass spectrometry. *Int J Mass Spectrom* 2009;281:24–30.
- [5] Skwarczynski M, Kiso Y. Application of the O–N intramolecular acyl migration reaction in medicinal chemistry. *Curr Med Chem* 2007;14:2813–23.
- [6] Mimna R, Camus MS, Schmid A, Tuchscherer G, Lashuel HA, Mutter M. Switch peptides: disruption of amyloid fibrils through controlled induction of β -sheet to α -helix transformation. *Angew Chem Int Ed* 2007;46:2681–4.

- [7] Lecaillon J, Gilles P, Subra G, Martinez J, Amblard M. Synthesis of cyclic peptides via O–N-acyl migration. *Tetrahedron Lett* 2008;49:4674–6.
- [8] Tailhades J, Gidel MA, Grossi B, Lecaillon J, Brunel L, Subra G, et al. Synthesis of peptide alcohols on the basis of an O–N acyl-transfer reaction. *Angew Chem Int Ed* 2010;49:117–20.
- [9] Tailhades J, Blanquer S, Nottelet B, Coudane J, Subra G, Verdie P, et al. From polyesters to polyamides via O–N acyl migration: an original multi-transfer reaction. *Macromol Rapid Commun* 2011;32:876e880.
- [10] Jebors S, Enjalbal C, Amblard M, Subra G, Mehdi A, Martinez J. Switchable polymer-grafted mesoporous silica's: from polyesters to polyamides biosilica hybrid materials. *Tetrahedron* 2013;69:7670–4.
- [11] Falb M, Aivaliotis M, Garcia-Rizo C, Bisle B, Tebbe A, Klein C, et al. Archeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey. *J Mol Biol* 2006;362:915–24.
- [12] Bradshaw RA, Medzihradzky KF, Chalkley RJ. Protein PTMs: post-translational modifications or pesky trouble makers? *J Mass Spectrom* 2010;45:1095–7.
- [13] Lee S, Tan M, Dai L, Kwon OK, Yang JS, Zhao Y, et al. MS/MS of synthetic peptide is not sufficient to confirm new types of protein modifications. *J Proteome Res* 2013;12:1007–13.
- [14] Zhang J, Chen Y, Zhang Z, Xing G, Wysocka J, Zhao Y. MS/MS reveals false positive identification of histone serine methylation. *J Proteome Res* 2010;9:585–94.
- [15] Lapko VN, Smith DL, Smith JB. Identification of an artifact in the mass spectrometry of proteins derivatized with iodoacetamide. *J Mass Spectrom* 2000;35:572–5.
- [16] Mouis L, Subra G, Enjalbal C, Martinez J, Aubagnac J-L. O–N-acyl migration in N-terminal serine containing peptides: mass spectrometric elucidation and subsequent development of site-directed acylation protocols. *Tetrahedron Lett* 2004;45:1173–8.

Supplementary data available at :

<http://www.sciencedirect.com/science/article/pii/S187439191400308X>

3.4.2 Conclusions et perspectives

Comme nous l'avons indiqué dans notre revue bibliographique, si la spectrométrie de masse est un outil indispensable à la recherche des modifications post-traductionnelles (page 23, 54), elle ne peut pas permettre à elle seule d'établir l'existence d'une MPT, aussi précise et exacte que soit la mesure de la masse du peptide portant cette MPT. Les nombreux travaux présentés par Y. Zhao (Page 90), indiquent clairement que des réarrangements chimiques peuvent se produire, au cours de l'extraction, le stockage ou l'ionisation du peptide sensé porter une modification post-traductionnelle. L'exemple documenté par l'article précédent montre une fois encore que la recherche et l'analyse de modifications post-traductionnelles ne peut faire l'économie du contexte biologique et technique dans lequel s'effectue l'analyse.

3.5 - Thermococcus Gamma tolerans, une Archea aux lysines ϵ -acétylées

3.5.1 Analyse des lysines acétylées du protéome de l'Archaea Thermococcus Gammatolerans

Thermococcus gamma tolerans est une archeae isolée à partir d'échantillons prélevés dans des fonds marins près de cheminées hydrothermales. C'est un organisme hétérotrophe, anaérobie strict, radiotolérant, se développant à la température optimum de 88°C. Une étude protéomique basée sur la recherche de modifications post-traductionnelles a été entreprise sur cet organisme en utilisant l'expérience accumulée sur *Deinococcus deserti*. Contrairement à ce dernier organisme les acétylations N-terminales ne semblent pas présentes en nombre chez cet archeae. Par contre des acétylations sur les chaînes latérales des lysines sont présentes et ont pu être isolées par immunopurification par deux anticorps différents. Egalement des phosphorylations isolées par immuno-capture (IMAC) ont été retrouvées. Certaines protéines portent les deux MPT.

Le travail présenté dans ce chapitre pourrait donner lieu à un article dont le titre et les auteurs sont présentés ci-dessous.

Survey of the lysine-specific acetylated proteome from the archaeon *Thermococcus Gammatolerans*

Alain Dedieu¹, Jean-Baptiste Boyer¹, Damien Joullié¹, Murielle Dutertre², Arnaud Lagorce², Yvan Zivanovic², Christian Roumestand³, Fabrice Confalonieri², Jean Armengaud¹

¹CEA, DSV, IBEB, Lab Biochim System Perturb, Bagnols-sur-Cèze, F-30207, France.

² Laboratoire de Génomique des Archae, Université Paris-Sud 11, CNRS, UMR8621, Bât400 F-91405 Orsay, France

³ Centre de Biochimie Structurale, INSERM U1054, CNRS UMR 5048, Universités de Montpellier, France.

Introduction

L'acétylation réversible sur les résidus lysine des protéines est une modification post-traductionnelle (MPT) qui fut d'abord mise en évidence sur des histones du nucléosome (Phillips 1963; Allfrey, Faulkner et al. 1964). Par la suite, il devint évident que l'acétylation n'était pas l'unique MPT affectant les histones. Méthylation, phosphorylation, ubiquitination et ribosylation de l'ADP peuvent aussi modifier la structure de la chromatine et donc l'expression des gènes (Bartova, Krejci et al. 2008).

L'acétylation des lysines fut également mise en évidence sur d'autres protéines différentes des histones et aux fonctions biologiques variées, la régulation de la transcription, la traduction et les voies du métabolisme central (Lin, Qi et al. 2008; Finkel, Deng et al. 2009; Zhao, Xu et al. 2010).

Les acétylations des lysines ne se produisent pas seulement sur les chaînes latérales des lysines. L'extrémité amino-terminale des protéines (Bonissone, Gupta et al. 2013) peut être concernée, il s'agit là d'un tout autre type d'acétylation, considéré comme non réversible.

Ces deux types d'acétylation sont très répandus chez les eucaryotes. Jusqu'à récemment, seules quelques protéines ont été décrites comme acétylées chez les micro-organismes (Soppa 2010), telles la protéine CheY et l'acétylCoA synthétase. Cependant certaines familles de déacétylase, comme la Sir2, sont très bien conservés et présentes non seulement chez les eucaryotes mais aussi chez les bactéries et les archées. En outre, la présence de nombreuses protéines acétylées dans les mitochondries, issues d' α -protobacteria (Gray, Burger et al. 1999; Kim, Yu et al. 2013) a été une autre indication permettant de supputer l'existence d'acétylomes conséquents chez les procaryotes.

Les nouvelles approches protéomiques développées au cours des dernières années, ont permis de mettre en évidence et d'explorer l'acétylome de plusieurs organismes bactériens. Il est aujourd'hui établi que cette modification est bien présente dans différents types de bactéries, les bactéries Gram négatives telles que *Escherichia coli* (Zhang, Sprung et al. 2009; Zhang, Zheng et al. 2013), *Salmonella enterica* (Wang, Zhang et al. 2010), mais aussi des bactéries gram-positives telles que *Bacillus subtilis* (Kim, Yu et al. 2013), également des thermophiles Gram positif comme *Geobacillus kaustophilus* (Lee, Kim et al. 2013) ou dans une autre bactérie extrémophile appartenant à l'ancien phylum Thermus-Deinococcus, comme *Thermus thermophilus* HB8 (Okanishi, Kim et al. 2013). Tous ces résultats indiquent clairement que l'acétylation de protéines est largement répandue chez les bactéries. Chez *Erwinia amylovora* (Wu, Vellaichamy et al. 2013), il a été démontré que, parmi les 96 protéines acétylées, près de la moitié d'entre elles était impliquée dans un large éventail de processus métaboliques, mais de façon plus remarquable dans les processus de virulence.

A ce jour, nous n'avons trouvé dans la littérature aucune étude globale portant sur un acétylome d'archée. Le but de ce travail est de vérifier l'existence d'un acétylome conséquent chez *Thermococcus gammatolerans*, et d'établir si le profil de cet acétylome est plus proche de celui des eucaryotes ou des bactéries. *Thermococcus gammatolerans* a été isolé à partir d'échantillons prélevés dans les cheminées hydrothermales situées dans la dorsale médio-atlantique et le bassin de Guayamas (Golfe du Mexique) (Jolivet, L'Haridon et al. 2003; Jolivet, Corre et al. 2004). Il est

l'un des organismes les plus radiorésistants connus sur Terre, pouvant supporter des irradiations atteignant 5 kGy, sans létalité (Tapias, Leplat et al. 2009). *T. gammatolerans* possède un chromosome circulaire de 2.045 Mbps codant pour 2157 protéines (Zivanovic, Armengaud et al. 2009). *T. gammatolerans* EJ3 a été décrit comme un organisme hétérotrophe anaérobie obligatoire qui se développe de manière optimale à 88 ° C en présence de soufre ou de cystine sur extrait de levure, la tryptone peptone et produisant de l'hydrogène sulfuré (H₂S). Finalement, nous avons mis en évidence un acétylome significatif de 154 protéines chez *T. gammatolerans*, dans deux conditions physiologiques différentes. Ce travail est la première mise en évidence massive chez les archées de cette MPT.

Matériels and Méthodes

Souches, milieu et conditions de croissance. *T. gammatolerans* EJ3 a été cultivé dans des flacons, dans des conditions anaérobies à 85°C, soit en milieu organique complexe (VSM) additionné de soufre (2 g/L) ou en milieu Mayt comme décrit dans (Kanai, Imanaka et al. 2005) et supplémenté avec 5 g/L de pyruvate (milieu de MAYTP). Des dosages de croissance ont été effectués dans des milieux synthétiques (ASW: eau de mer artificielle) complétés avec un mélange de 20 acides aminés (Sato, Fukui et al. 2003) ou supplémentés avec des substrats protéiques spécifiques (acides Casamino, extrait de levure, de la tryptone et de la peptone) à une concentration de 5 g/L en présence ou en absence de S °. Des dosages de croissance ont été également réalisés dans le milieu complexe riche en éléments nutritifs (Mayt) supplémenté avec des sucres comme sources de carbone (5 g/L) en présence ou en absence de S °. VSM est composé de 20 g/L de NaCl, 0,25 g/L de KCl, 0,05 g/L de NaBr, 0,02 g/L d'acide borique, 0,01 g/L SrCl₂ • 6H₂O, 0,5 g/L de citrate trisodique, 3 g/L PIPES (pipérazine-1,4-bis (acide 2-éthanesulfonique)), 1 g/L d'extrait de levure, 4 g/L Bactotryptone, 5 ml de MgSO₄ 20%, 1 ml de CaCl₂ 5%, 1 ml de KH₂PO₄ à 5%. Le pH a été ajusté à 6,8 par addition de NaOH. Les milieux ont été stérilisés par autoclavage et transférés dans des flacons à sérum. L'air contenu dans les bouteilles a été éliminé par pompage, puis remplacé par N₂. Afin de réduire l'oxygène dissous dans le milieu, il a été ajouté du Na₂S • 9H₂O à une concentration finale de 0,1%, jusqu'à ce que la couleur de la résazurine (NaCl) (1 mg/L) apparaisse. Typiquement, des flacons ont étéensemencés à une densité cellulaire de 5 x 10⁵ cellules/ml et incubés à 85°C. Les tests de croissance ont été effectués en triple. La croissance a été enregistrée durant 3 jours.

Conditions d'irradiation.

Deux types d'irradiation ont été conduits, l'irradiation gamma et radiolyse pulsée

Des volumes importants (600mL dans des bouteilles 1L Schott, 7x10⁷ cellules / ml) de cellules en croissance exponentielle ont été filtrés (MIRACLOTH Calbiochem, La Jolla, CA, USA) pour éliminer les composés soufrés. Dans des conditions anaérobies, les cellules ont été concentrées 10 fois (dans des tubes de Hungate spéciaux pour γ -irradiation ou en quartz cellules Hellma pour la radiolyse pulsée). Les cellules ont ensuite été exposées à une dose de 2500 Gy ou 5000 Gy assurant 100% de survie (Tapias, Leplat et al. 2009). Pour cela on utilise soit une source de γ -rayons ¹³⁷Cs (40 Gy / min, IBL 637 CisBio International, Institut Curie, Orsay, France) soit un accélérateur linéaire d'électrons de 4,5 MeV pour la radiolyse pulsée (Kinetron Linac, Institut Curie, Orsay, France). L'accélérateur linéaire d'électrons délivre des doses en ~ 1 sec alors que

pour la source de ^{137}Cs de rayons gamma l'irradiation dure 1 ou 2 h. A titre de témoin, le même nombre de cellules non irradiées a été maintenu sur de la glace pendant l'irradiation. Les cellules irradiées et non irradiées ont été immédiatement prélevées et incubées sous agitation pendant 60 min à 85°C dans un milieu riche contenant du soufre élémentaire (VSM-S $^\circ$) avant centrifugation (4000g, 20 min à 4°C). Les cellules ont été immédiatement congelées et conservées à -80°C jusqu'à leur traitement ultérieur. Chaque culture a été dupliquée.

L'extraction des protéines. Un tampon de lyse a été préparé, contenant une solution sans EDTA, avec un cocktail d'inhibiteurs de protéase (Roche, 1 comprimé par 50 ml), des inhibiteurs de phosphatases 2 et 3 (1: 100, v: v, chacun), de l'urée 8M, du NaCl 75 mM, le tout dans un tampon Tris 50 mM (pH 8,2) contenant en outre 50 mM de β -glycérophosphate, 1 mM d'orthovanadate de sodium, du pyrophosphate de sodium 10 mM, 1 mM de MgCl_2 , et 1: 10 000 (v: v) Benzomase 25 kU. Les cellules de *T. gammatolerans* ont été remises en suspension dans ce tampon de lyse à raison de 3,5 ml de tampon par gramme de cellules. La suspension a été soniquée et centrifugée à 16000 g pendant 10 min. Les surnageants ont été collectés, et les concentrations de protéines ont été mesurées par la méthode de Bradford. Six échantillons de protéines ont été lysés composés de trois réplicats biologiques, C1, C2, C3, en tant que témoins et I1, I2, I3, en tant qu'échantillons stressés.

Immunodétection par Western Blot des lysines acétylées. Chaque lysat de protéine (20 μg) a été analysé par SDS-PAGE (gel Nu-PAGE à 4-12%) dans un tampon MOPS 5% (Novex) et transféré sur une membrane de PVDF de 0,45 μm (Thermo Scientific Pierce) par procédé électrophorétique au moyen d'un MiniTrans -Blot (Bio-Rad). Le blocage des liaisons non spécifiques a été réalisé en plaçant la membrane pendant 6 h dans un tampon TBST (Trizma base 25 mM, glycine 192 nM, EtOH à 10% (vol/vol), pH 8,5, Tween 20 0,1%) supplémenté en BSA (4% poids/vol). L'acétylation des lysines a été détectée en utilisant alternativement un anticorps polyclonal de lapin anti-acétyl-lysine (ICP380, ImmuneChem) dilué au 1: 1000 ou un anticorps monoclonal (pendant une nuit, 4°C) et un anticorps secondaire anti-lapin HRP à 1: 2 000, ou un anti-acétyl-lysine anticorps monoclonal de lapin (Cell Signal) pendant 1 h à température ambiante. Dans tous les cas, les anticorps ont été dilués dans du tampon TBST. Le système Western Immobilon (Millipore) a été utilisé pour la révélation (5 min). Enfin la détection a été effectuée par une exposition de 45 s dans un automate de révélation le SMXD-101A (Amersham).

Protéolyse en solution assistée par micro-ondes. Pour chaque échantillon de *T.gammatolerans*, un volume correspondant à 8 mg a été prélevé et déposé dans un tube à centrifuger. Une solution de DTT (0,1M) a été ajoutée pour obtenir une concentration finale en DTT de 10 mM. Les tubes ont ensuite été incubés 1 h à 37°C . Une solution aqueuse d'acide formique à 50% a été ajoutée pour obtenir une solution finale à la concentration de 6% (vol / vol). Les tubes ont été scellés et placés sur un flotteur dans un bain d'eau, et ensuite, exposés à un rayonnement micro-onde de 750 W fourni par un appareil micro-onde domestique et ceci pendant 2,5 minutes. Après dénaturation, les tubes ont été retirés du bain-marie et placés sur de la glace fondante. Un volume de 160 μL de tampon Tris-base à 0,5 M a été ajouté dans chaque tube afin d'obtenir un pH situé entre 8,0 à 8,5. Une solution d'iodoacétamide (0,2 M) a été ajoutée (21 μL par tube). Les tubes ont été ensuite incubés 30 min à l'obscurité et à température ambiante. La protéolyse a été effectuée avec une solution de trypsine à 0,5 $\mu\text{g}/\mu\text{L}$ dans une solution de TFA (0,01%). La quantité de trypsine

apportée l'a été dans un rapport de 2: 100 (en poids) de la masse de la protéine totale dans l'échantillon. Après une nuit d'incubation à 37 ° C, la protéolyse est arrêtée en ajoutant une solution d'acide formique à 50%, conduisant à ramener le pH de la solution à 4. Tous les tubes d'un même échantillon ont été recueillis. L'efficacité de la protéolyse a été vérifiée par gel SDS-PAGE (Bis-Tris gel gradient de NuPAGE 4-12% comme décrit précédemment). Le produit de la digestion trypsique (360 mg de matière initiale) a été dessalé en utilisant les cartouches C18 Sep-Pak Plus et élué selon les instructions du fabricant.

Enrichissement de peptides acétyl-lysine par immunocapture. Nous avons utilisé deux anticorps différents pour enrichir notre protéome en peptides acétyl-lysine.

Un anticorps anti-acétyl-lysine purifié par affinité (ImmunoChem Pharmaceuticals Inc., Burnaby, Colombie-Britannique, Canada) a été utilisé à partir d'une préparation commerciale d'ImmuneChem (ICP0388) constituée de billes d'agarose déjà liées de façon covalente avec cet anticorps. Le digestat trypsique de notre protéome extrait a été ajouté à 40 µL de billes d'agarose-anticorps et incubé pendant une nuit à 4 ° C. Les billes ont été lavées trois fois avec du tampon ETN. Les peptides liés ont été élués à partir des billes avec 2 x 30 µL de tampon (pH 2,3) glycine-HCl (50 mM).

Un anticorps monoclonal reconnaissant également les peptides acétyl-lysine (Cell Signaling, # 9814) a été incubé avec des billes d'agarose revêtues de protéine A (ImmuneChem, ICP1001) pendant 4 h à 4 ° C. Le surnageant a été éliminé et les billes ont été lavées deux fois avec du tampon ETN. Le digestat trypsique de notre protéome extrait a été ensuite ajouté à ces billes et incubé à 4 ° C pendant une nuit. Ensuite, les billes ont été lavées trois fois avec du tampon d'ETN et les peptides liés ont été élués à partir des billes avec 2 x 30 µL de tampon glycine-HCl (pH 2,5).

Dans les deux cas on procède à deux enrichissements puis les éluats sont dessalés sur ZipTip C18 (Millipore, Bedford, Massachusetts) selon les instructions du fabricant et séchés par SpeedVac (Savant Thermo Scientific). Les peptides récupérés ont été solubilisés dans du TFA 0,1% et analysés par nano-HPLC MS/MS.

Enrichissement de phosphopeptides par chromatographie d'affinité IMAC

A titre exploratoire et sur un seul réplicat biologique C1, une recherche de peptides phosphorylés a été tentée sur un extrait de protéines solubles de *Thermococcus gamatolerans*.

Préalablement à un enrichissement par chromatographie d'affinité de type IMAC, un fractionnement peptidique par chromatographie SCX est réalisé pour préenrichir le digestat trypsique de la fraction C1 en peptides phosphorylés. On utilise une colonne SCX PolySULFOETHYL AspartamideTM (300 x 3,2mm, 5µm, 200Å, PolyLC) (« The Nest Group - Southborough, MA»). Les peptides trypsiques dessalés et séchés sont repris par 120µL de tampon SCX A (Tampon 5 mM NH₄HCO₂ ; 30% ACN ; pH ajusté à 2,7 par de l'acide formique). L'échantillon est centrifugé et 100µL sont injectés sur la colonne (HPLC Agilent 1100) à un débit de 0,4mL/min. L'élution des peptides retenus est effectuée à l'aide d'un gradient linéaire 0 – 100% de tampon SCX B (Tampon 500mM NH₄HCO₂ ; 30% ACN ; pH ajusté à 3,5 par de l'acide formique) sur 70 min, après 5 minutes à 0% de tampon SCX B.

Des fractions de 1 mL, correspondant à 2,30 min d'élution sont récoltées. Ces fractions sont évaporées pour éliminer l'acétonitrile, dessalées SPE Macro SpinColumn (Harvard Apparatus). Les fractions sont reprises par 100µL de TFA 0,1% et peuvent être stockées à -80°C.

Les fractions sont collectées à raison de 80µL pour chaque fraction de 1 à 22, 80µL pour chacune des fractions 23 à 25 qui sont regroupées, et 80µL pour chacune des fractions 26 à 28 qui sont aussi regroupées. Toutes ces fractions sont évaporées au speedvac puis reprises avec 100µL d'une solution R (acide acétique 250 mM et 30% ACN).

Pour chaque fraction, 20µL de billes PHOS-SelectTM (Sigma Aldrich) sont placées dans un support pour filtration de type SigmaPrepTM Spin Columns (Sigma Aldrich). Les billes sont lavées deux fois avec la solution R et chaque fois centrifugées sur leur support (2 min, 4000g). L'échantillon est alors mis en contact avec les billes laissées dans leur support (1h sous agitation) en utilisant un Thermomixer (22°C, 600rpm) (Eppendorf). Les tubes sont centrifugés, les billes y sont lavées une fois avec la solution R. puis on procède à une élution en milieu basique avec 100µL d'une solution E1 (NH₄OH 400mM) par mise en contact 2min avec les billes (sur Thermomixer), la solution est alors éluée par centrifugation (2 min, 4000g). Une élution en milieu basique avec 100 µL d'une solution E2 (H₃PO₄ 5%) est alors réalisée avec le même traitement que l'élution précédente. Les 2 éluats sont récupérés et mélangés. On procède ainsi pour les 24 fractions récoltées qui sont ensuite dessalées SPE Macro SpinColumn (Harvard Apparatus) en éluant avec une solution (50% ACN et 0,1% TFA). Les éluats sont évaporés (speedvac) avant d'être repris par 30 µL de TFA 0,1% et analysés en MS/MS.

Analyse NanoLC-MS / MS. La détection des peptides a été effectuée par nanoLC-MS / MS réalisée en utilisant un spectromètre de masse hybride LTQ-Orbitrap XL (ThermoFisher) couplé à un système de nano-LC 3000 UltiMate (Dionex-LC Packings). Le protocole détaillé a été décrit dans une publication précédente (Hartmann, Allain et al. 2014). Chaque échantillon a été reconstitué avec 40 μ L de TFA à 0,1%. Un volume de 10 μ L de cette solution a été utilisé et préconcentré sur pré colonne en phase inverse C-18 Acclaim Pepmap (LC Packings). Les peptides ont ensuite été séparés sur colonne en phase inverse Acclaim Pepmap 100 C-18 (LC Packings) fonctionnant à un débit de 0,3 μ L / min pour un gradient de 120 min de 0 à 45% de solvant B (80% de CH_3CN , 0,1% de HCO_2H), solvant A est de 0,1% HCO_2H . Les spectres de masse ont été acquis de 5 à 120 min avec un balayage allant de 300 à 1800 m/z. Une stratégie d'acquisition TOP 7 a été utilisée, consistant en une mesure très précise dans l'Orbitrap (résolution 30000) des 7 peptides les plus intenses suivie de leur fragmentation dans la trappe linéaire. Les fragmentations sont provoquées par des ions d'hélium et une énergie de collision normalisée, réglée à 35% de l'amplitude des radiofréquences appliquées à l'ion sélectionné.

Exploitation des données. Les listes de spectres ont été créés en utilisant le logiciel DAEMON MASCOT (version 2.3.2 Matrice Sciences) et le filtre d'importation proposé par Thermo Fisher Xcalibur FT (version 2.0.7, extract_msn.exe). Les options de filtrage ont été fixées à 400 (masse minimale), 5000 (masse maximale), 0 (tolérance de regroupement), 0 (analyses intermédiaires), et 1000 (seuil). En utilisant le moteur de recherche Mascott (2.2.06 de Matrix Science), nous avons cherché tous les spectres MS / MS contre une base de données interne. Dans tous les cas, la recherche de peptides a été réalisée avec les paramètres suivants: une tolérance de masse de 5 ppm de l'ion parent et 0,5 Da pour les spectres de fragmentation, la modification statique retenue était la carbamidométhylation de la Cystéine (57,0215), les modifications dynamiques l'oxydation de la Méthionine oxydée (15,9949) et l'acétylation du groupe amino de la chaîne latérale des lysines (42,0105) (Zhang, Yau et al. 2004) Dans le cas des recherches de phosphorylation la modification dynamique portera toujours sur la méthionine oxydée mais aussi sur la recherche de phosphorylation sur les sérines, thréonines et tyrosines (S, T, Y). Le nombre maximum de clivages défailants a été fixé à 2 pour le résultat final, mais une analyse avec 3 coupures défailantes a aussi été réalisée. Les peptides retenus doivent avoir un score peptidique supérieur au seuil d'identité calculé pour une p value fixée à $p < 0,01$ avec la base de données CDS de *T.gammatolerans* (téléchargée à partir du NCBI 2013). En outre, ces peptides doivent être de rang1. Les données sont filtrées en utilisant ces critères avec le logiciel IRMa 1.28.0 (Dupierris, Masselon et al. 2009). Le taux de faux positifs pour l'identification des peptides a été estimé en utilisant la base de données inversée à moins de 0,7%.

Résultats

1. Mise en évidence par Western Blot d'une présence importante de protéines portant des lysines acétylées.

Nous avons d'abord étudié la présence de protéines acétylées par Western blot en utilisant les deux types d'anticorps (voir matériels et méthodes) dirigés contre les motifs acétyl-lysine. La figure 48 montre le western-blot avec révélation par chimioluminescence de protéines extraites à partir de *Thermococcus gammatolerans* (puits 1,2 et 4,5) et de *Saccharomyces cerevisiae* (utilisé ici à titre de témoin, puits 3 et 6). Les protéines acétylées ont été révélées avec un anticorps polyclonal anti acétyl-lysine de lapin (figure 48, les puits 1, 2 et 3) ou un anticorps anti-acétyl-lysine monoclonal de lapin (Figure 48, puits 4, 5 et 6). Dans les deux cas nous avons pu détecter des bandes intenses couvrant une large gamme en masse de protéines, signant la présence de motifs acétyl-lysine. Il faut noter que l'anticorps polyclonal a révélé la présence de quatre bandes de protéines, abondamment détectées (puits 1 et 2), tandis que l'anticorps monoclonal montre des motifs de révélation bien moins intenses et légèrement différentes. La figure 49 montre l'immunodétection de protéines portant des motifs acétyl-lysine sur six échantillons biologiques, trois provenant de cellules non stressées (A, C et E) et trois de cellules irradiées (B, D et F). Il est difficile à partir de ces résultats de voir clairement une tendance sur une induction ou répression des protéines acétylées après irradiation.

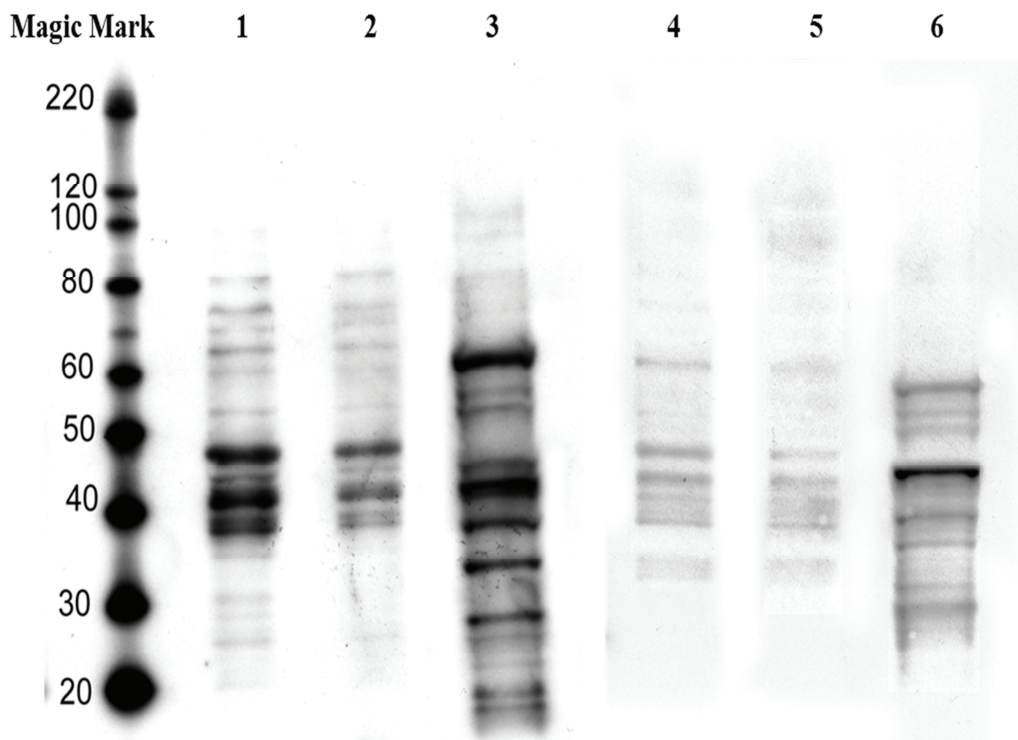


Figure 48 : Détection par Western blot de protéines contenant des résidus lysine acétylée en utilisant un anticorps polyclonal (puits 1,2 et 3) ou monoclonal (puits 4 ,5 et 6).

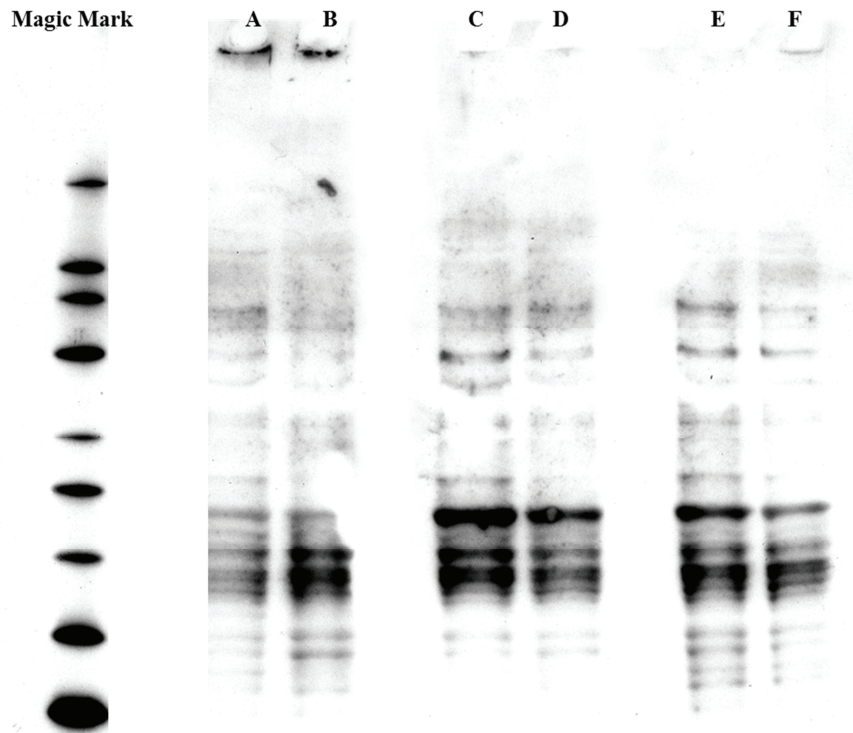


Figure 49 : Détection par Western blot de protéines contenant des résidus lysine acétylée, en condition normale (puits A, C et D) ou après un stress radiatif (puits B, D et F)

2. Recherche et mise en évidence de peptides et protéines acétylés chez *Thermococcus gammatolerans*.

Une stratégie protéomique a été mise en œuvre afin de rechercher et mettre en évidence le plus grand catalogue possible de peptides acétyl-lysine chez *Thermococcus gammatolerans*. Cette approche comprend quatre étapes principales.

Tout d'abord, nous avons préparé trois cultures biologiques différentes et distinctes de *T. gammatolerans* C1, C2, C3 qui n'ont pas été exposées, et trois autres qui ont été soumises à un stress radiatif important (5 kGy) produisant ainsi trois échantillons différents nommés I1, I2, I3. Les lysats de protéines ont été obtenus à partir de ces six échantillons protéolysée à la trypsine.

La deuxième étape a été une purification par affinité des peptides acétyl-lysine à partir de ces six échantillons biologiques, au moyen de deux anticorps anti-acétyl-lysine différents.

La troisième étape était l'identification par nanoLC-MS / MS à l'aide d'un spectromètre de masse à haute résolution du plus grand nombre de peptides présents dans nos préparations immunopurifiées.

Enfin, la quatrième étape a été l'identification de sites d'acétylation de la lysine par l'interrogation de l'ensemble de données MS / MS.

Il a déjà été constaté dans la littérature (Zhang, Sprung et al. 2009; Guan, Yu et al. 2010) que la recherche de motifs acétyl-lysine est plus efficace sur un protéome digéré à l'état de peptides que sur le mélange de protéines entières de ce même protéome. En effet, les résidus acétyl-lysine peuvent dans ce dernier cas être moins accessibles et moins bien reconnus par l'anticorps utilisé. Nous avons donc privilégié une immunocapture des peptides plutôt que celle des protéines. Cependant, concernant les protéines de *Thermococcus gammatolerans*, nous nous sommes heurtés à leur grande stabilité et un procédé classique de protéolyse trypsique ne donne qu'une digestion incomplète. Cette résistance s'explique par la nature de ces organismes thermophiles qui conduisent à la sélection de protéines possédant à leur surface des nombreuses interactions ioniques qui permettent de rigidifier leur structure. (Gabant, Lorphelin et al. 2008). Aussi avons-nous dû développer un protocole de digestion en solution avec de la trypsine, avec dégradation préalable des protéines sous l'action d'un rayonnement micro-ondes.

Nous avons exploité l'idée que la spécificité de deux anticorps peut être légèrement différente et amener à une meilleure couverture de l'acétylome (Shaw, Chaerkady et al. 2011) que nous voulons mettre en évidence. Nous avons donc utilisé deux anticorps pour l'enrichissement de peptides acétylés-lysine :

- un anticorps polyclonal anti-acétyl-lysine, purifiés par affinité à partir d'une préparation commerciale dans laquelle cet anticorps était déjà fixé de façon covalente sur des billes d'agarose.

- un anticorps monoclonal dont nous avons optimisé la fixation sur des billes d'agarose revêtue de protéine. La Figure 50 représente un diagramme de Venn indiquant la reconnaissance par chacun des anticorps de site d'acétylation uniques. Ce résultat a été obtenu en compilant toutes les expériences effectuées et validées pour chacun des deux anticorps dans toutes les conditions décrites précédemment.

Comme il était attendu, nous constatons une claire complémentarité entre les deux anticorps permettant une meilleure couverture de l'acétylome.

Les données générées par l'Orbitap XL ont été traitées en recherchant systématiquement un incrément de masse de 42,0105 Da correspondant à celui d'une acétylation sur le groupement amine d'une chaîne latérale de lysine. Les spectres MS/MS ainsi obtenus, comme ceux présentés Figure 51, sont interprétables sans ambiguïté. En effet, sur des appareils de haute précision comme celui utilisé dans cette expérience cet incrément ne peut plus être confondu avec celui correspondant à une tri-méthylation (42,0468 Da). En effet cette différence est de 36 ppm, or toutes nos recherches se font à moins de 5 ppm et la plupart de nos mesures sont effectivement réalisées à moins de 1 ppm. C'est donc sans ambiguïté que nos peptides ainsi identifiés ne peuvent correspondre qu'à des peptides acétylés. La digestion par la trypsine se caractérise par une coupure de la liaison peptidique après une arginine ou une lysine. Comme il peut être possible qu'une lysine acétylée gêne l'action de la trypsine nous avons exploité nos résultats en introduisant la possibilité d'identifier des peptides jusqu'à trois coupures non réalisées (miscleavage). Les résultats non probants nous ont conduits à analyser l'ensemble de nos données avec deux « misscleavages ». Cette recherche a été effectuée avec une « p-value » ($p < 0,01$) conduisant à un taux de faux positifs inférieur à 0,5% estimé en utilisant une base de données « inversée ».

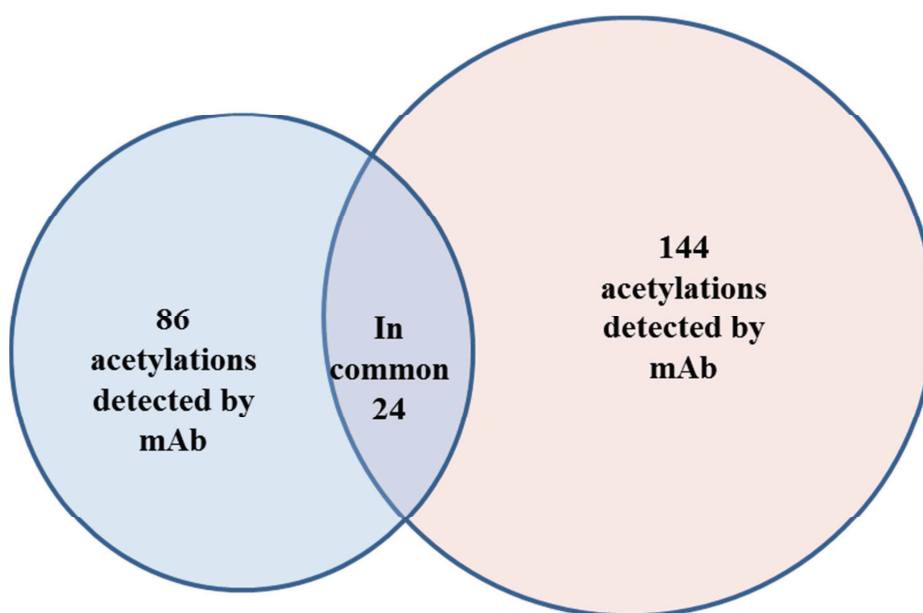


Figure 50 : Diagramme de Venn montrant la complémentarité de la reconnaissance des peptides acétyl-lysine par l'anticorps monoclonal (disque bleu et l'anticorps polyclonal disque brun).

L'exploitation des données obtenues à partir des six échantillons biologiques (chacun traité par les deux protocoles d'immuno-purification) nous a conduits à traiter 50 882 spectres MS / MS. Nous avons pu affecter 7035 spectres à 771 séquences peptidiques (données supplémentaires, tableau S1). Parmi ceux-ci, 1 798 spectres MS / MS correspondaient à des peptides acétylés. Un total de 258 peptides acétylés uniques ont ainsi pu être détectés. Parmi les 275 protéines différentes détectées dans cette étude, 154 (56%) ont été trouvées acétylées (données supplémentaires, tableau S2).

Ces résultats indiquent que l'acétylation de la lysine est assez importante dans *T. gammatolerans*. Lorsque le nombre de protéines acétylées est comparé au nombre théorique de protéines actuellement annotées dans cet organisme (2160), le rapport des protéines acétylées est de 7,1%. Ce ratio est nettement plus élevé que ce qui a été rapporté pour la plupart des autres procaryotes (bactéries) pour lesquelles un acétylome a été publiées en 2013 (*Erwinia amylovora* 2,6% (Wu, Vellaichamy et al. 2013), *Thermus thermophilus* 4,4% (Okanishi, Kim et al. 2013), *Geobacillus kaustophilus* 3,1% (Lee, Kim et al. 2013) , *Bacillus subtilis* 4,4% (Kim, Yu et al. 2013), et presque équivalent à *E. coli* où ce rapport est compris entre 2.3 et 8.7% selon les études (Zhang, Zheng et al. 2013).

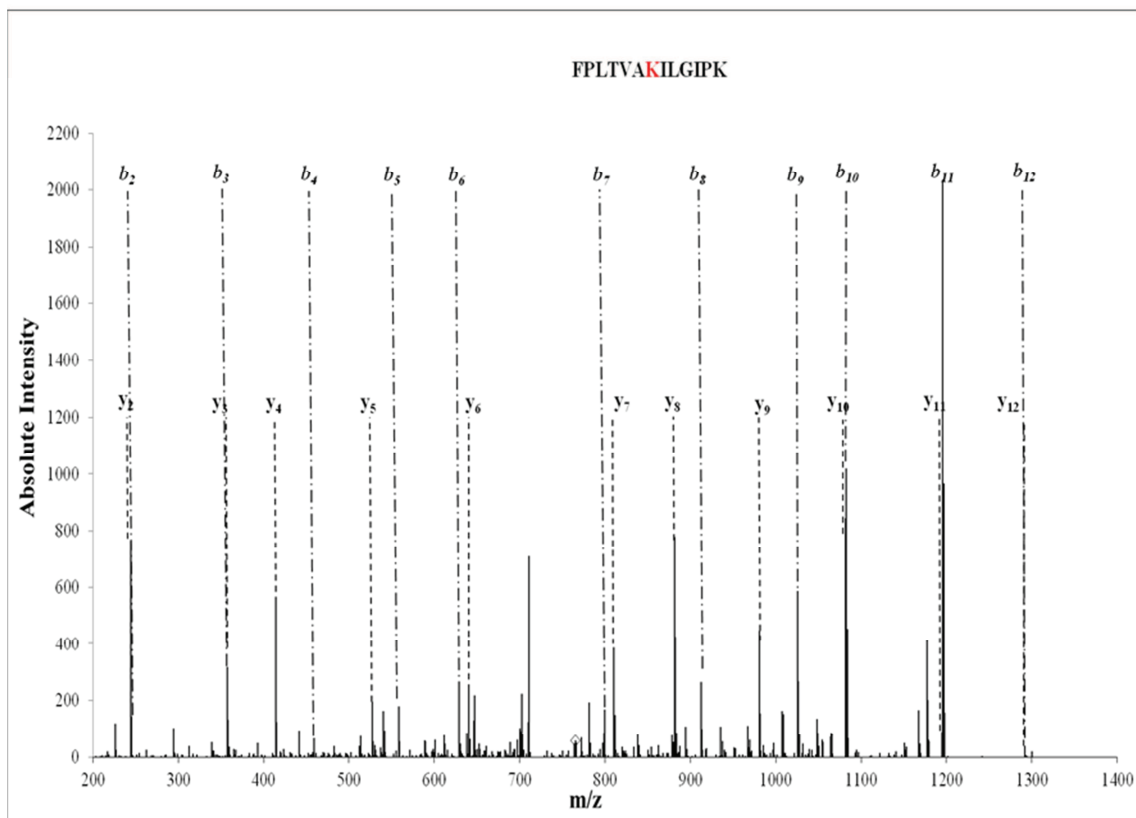
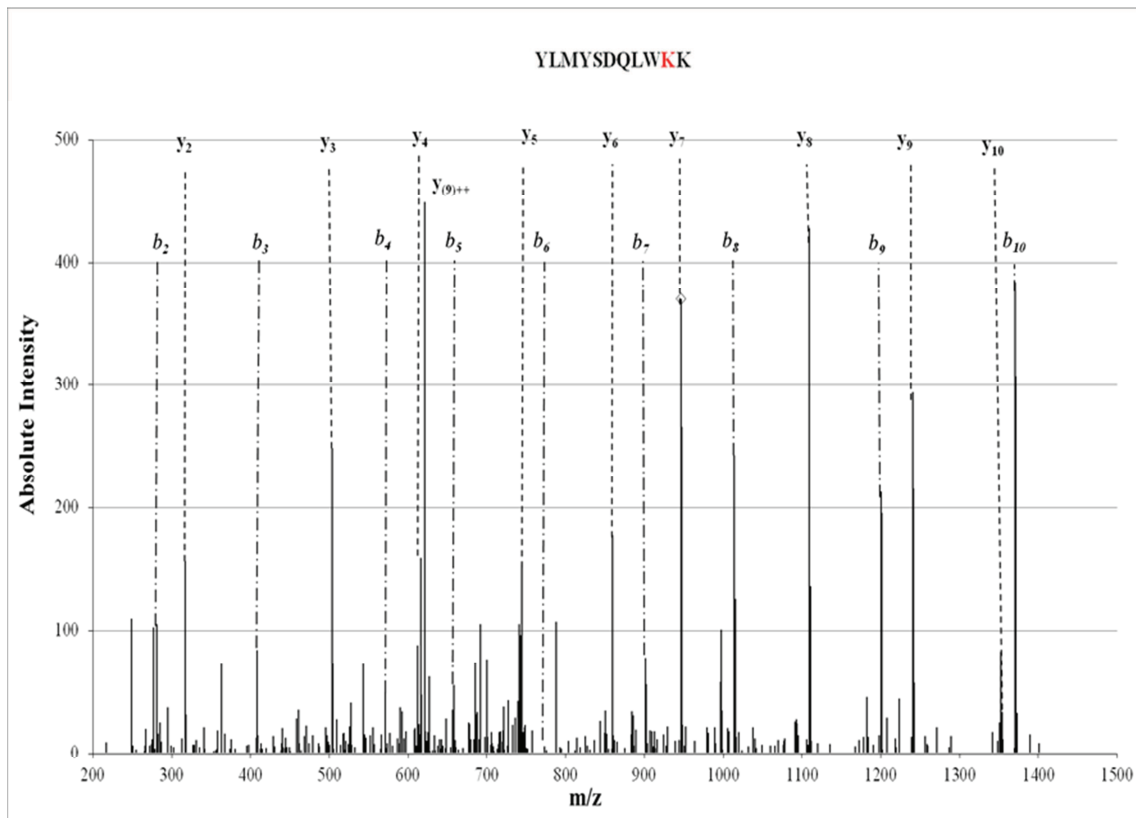


Figure 51 : Identification de l'acétylation de peptides par spectrométrie de masse.

Dans la figure 52 est mis en évidence le nombre de sites d'acétylation détectés dans cette étude par protéine. La grande majorité ne comprend qu'un site acétylé. Par contre, un nombre

relativement important de protéines peut contenir deux sites d'acétylation et quelques protéines ont de multiples sites. Nous avons trouvé quarante-deux protéines contenant 2 sites d'acétylation, douze en contiennent 3, six contiennent 4 sites, deux ont été détectés avec 5 sites d'acétylation. Certaines protéines sont hyper acétylées comme le facteur d'élongation EF-2 (Fusa) et l'oxydoréductase pyruvate / cétoisovalérate (sous-unité gamma) avec chacune 6 sites d'acétylation. Enfin le facteur 1-alpha d'élongation (EF-1-alpha) (facteur élongation Tu) et la synthase phosphoenolpyruvate contiennent 7 sites acétylés.

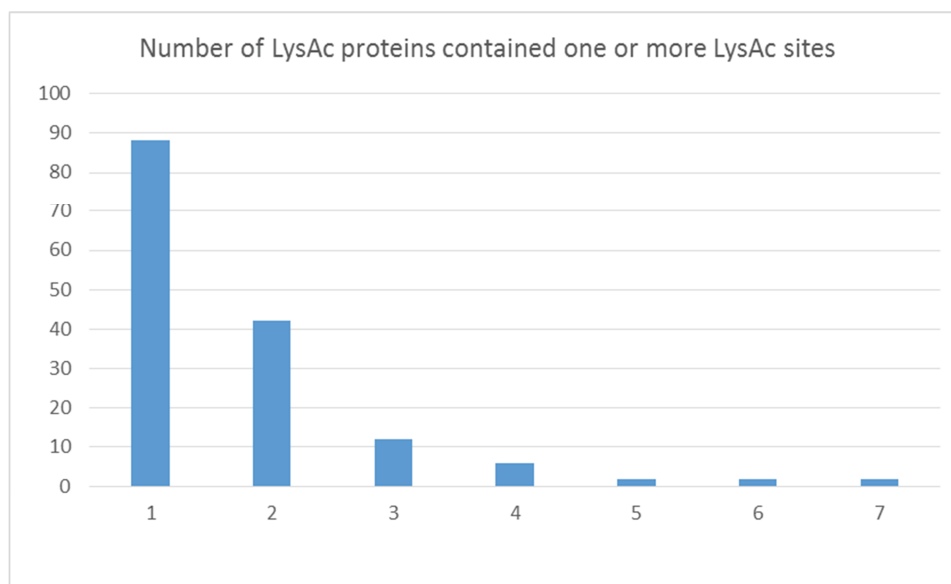


Figure 52 : Nombre de sites d'acétylation par protéine.

Dans un travail précédent publié en 2009 (Zivanovic, Armengaud et al. 2009), 951 protéines du protéome soluble de *T. gammatolerans* (croissance dans des conditions standards) avaient été identifiées. L'abondance relative de chaque protéine a été estimée par le calcul du facteur normalisé d'abondance par comptage spectral, en anglais NASF (pour Normalized Abundance Spectral count Factor). Nous avons comparé la liste des protéines acétylées obtenues dans notre étude avec ces données précédentes afin de déterminer si nous n'avions pas un biais majeur en ne détectant ici que les protéines acétylées parmi les plus abondantes. Or parmi les 154 protéines acétylées trouvées, 50% ont un NASF global de plus de 0,09%. D'autre part, certaines protéines acétylées se retrouvent parmi les protéines les moins abondantes du protéome global, telles l'exonucléase RecJ simple brin et l'ADN topoisomérase A qui ne représentent respectivement que 0,003% et 0,0008% du protéome.

3. Recherche et mise en évidence de motifs d'acétylation remarquables autour de la lysine concernée.

Nous avons cherché à déterminer si une séquence consensus ou au moins un profil particulier pouvait exister autour du site d'acétylation du résidu lysine. Pour cela, nous avons comparé la séquence des six acides aminés en aval et amont de ce résidu et par le logiciel Weblogo fait apparaître la fréquence de présence des acides aminés sur ces douze positions. Les résultats sont représentés dans la figure 53.

Nous avons distingué le jeu de peptides acétylés obtenu avec l'anticorps polyclonal (Fig 53 A) de celui obtenu avec l'anticorps monoclonal (Fig 53 B). La figure 50 indiquait déjà une grande complémentarité au niveau de ces deux jeux de données. Dans les deux cas ce sont les résidus acides ou hydrophobes qui dominent du moins sur les positions en amont de la lysine acétylée, de façon plus marquée pour l'anticorps monoclonal. Ce caractère acide disparaît totalement en position -1 pour cet anticorps polyclonal. Au regard des résidus reconnus par l'anticorps polyclonal, l'élément le plus remarquable se situe cependant sur la position +1 ou dans près de 70% des cas il est retrouvé un résidu R ou K après la lysine acétylée. Pour l'anticorps monoclonal, le caractère acide/hydrophobe est lui plus clairement marqué de part et d'autre (position -1 et +1) de la lysine acétylée. Ces différences de reconnaissance entre les deux types d'anticorps se retrouvent dans les résultats de Shaw et al (Shaw, Chaerkady et al. 2011). Dans cette étude aussi le recours à des anticorps polyclonaux et monoclonaux permet d'obtenir un acétylome plus large mais montre un profil de reconnaissance différent des peptides acétylés par les deux anticorps. La nécessité, pour obtenir un acétylome le plus complet possible, de recourir à des anticorps polyclonaux ou monoclonaux semble se confirmer. Par ailleurs, ce caractère acide en position +1 se retrouve dans l'étude de l'acétylome de *Bacillus subtilis* (Kim, Yu et al. 2013), et plus nettement encore dans l'étude réalisée sur *Thermus thermophilus* (Okanishi, Kim et al. 2013). Dans cette dernière étude notons la présence très marquée des résidus acides.

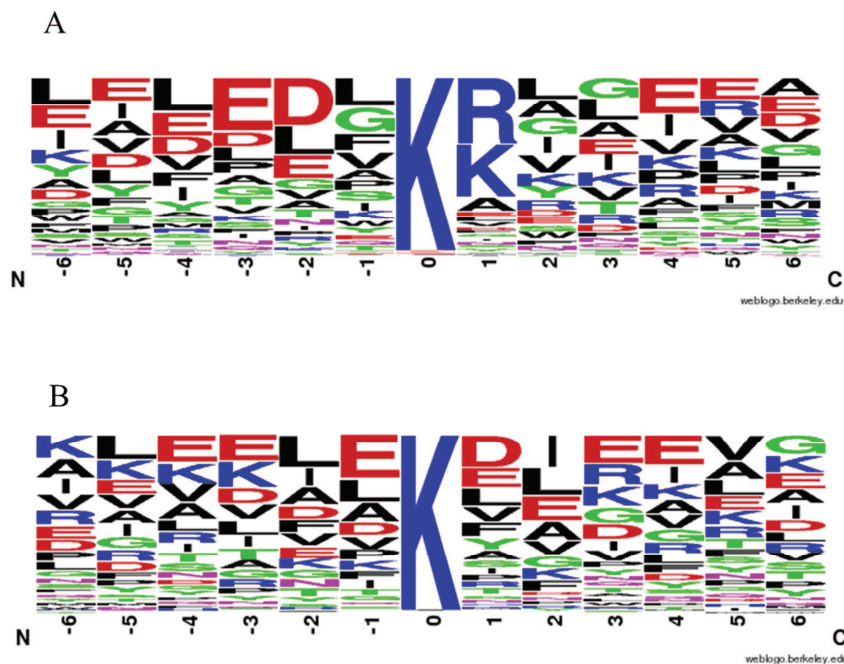


Figure 53 : L'analyse de la fréquence de la séquence d'acides aminés autour des sites d'acétylation.

Ce travail a été effectuée avec l'outil weblogo (<http://weblogo.berkeley.edu/>). Les acides aminés sont colorés en fonction de leurs propriétés chimiques: acides aminés polaires (G, S, T, Y, C, A, N) sont de couleur verte, basiques (K, R, H) bleu, acides (D, E) rouge et hydrophobes (A, V, L, I, P, W, F, M) sont noirs la grandeur de chaque motif est proportionnel à l'abondance relative des (+/-)6 acides aminés à partir du résidu Lysine (K) acétylé. Le pattern A est obtenu en utilisant l'anticorps polyclonal, le B l'anticorps monoclonal. De façon tout à fait intéressante, ce profil d'acidité /hydrophobie se retrouve dans un étude proposée par Jeffers et al (Jeffers and Sullivan 2012), sur les seules protéines acétylées des mitochondries.

4. Classification fonctionnelle des protéines acétylées sur les lysines chez *Thermococcus gammatolerans*.

Nous avons classé les protéines acétylées en fonction de leur fonction biologique, classification COG (Figure 54). Comme chez les bactéries (Yu, Kim et al. 2008; Zhang, Sprung et al. 2009; Wang, Zhang et al. 2010; Zhang, Zheng et al. 2013) de nombreuses protéines acétylées (42) se retrouvent au niveau du métabolisme central (Glycolyse, TCA, PPP...). Mais cela concerne également la réplication, la transcription et la traduction avec 55 protéines dont 27 au niveau de la traduction. Enfin une proportion notable de protéines acétylées (25) se retrouvent au niveau de la production d'énergie. Dans les principaux cycles du métabolisme central, toutes les voies à l'exception du cycle de l'acide tri-carboxylique (cycle TCA Figure 55D) possèdent plusieurs protéines acétylées, avec la présence, au bout de la voie de la glycolyse, de la phospho-énol pyruvate syntase qui avec 7 sites d'acétylation est hyperacétylée. Les acétylations doivent contribuer à la régulation de ces voies.

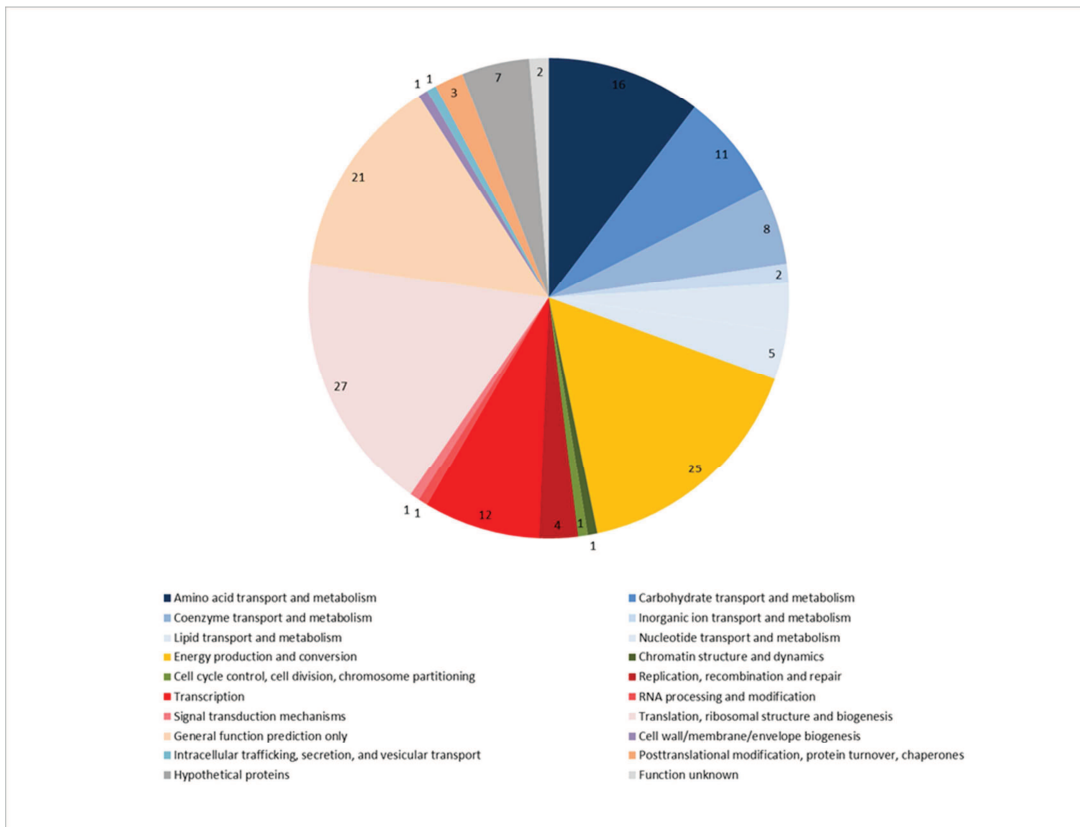


Figure 54 : Annotation fonctionnelle des protéines acétylées chez *Thermococcus gammatolerans*. Les 154 protéines sont ici représentées et sont classées par principales fonctions (classification COG).

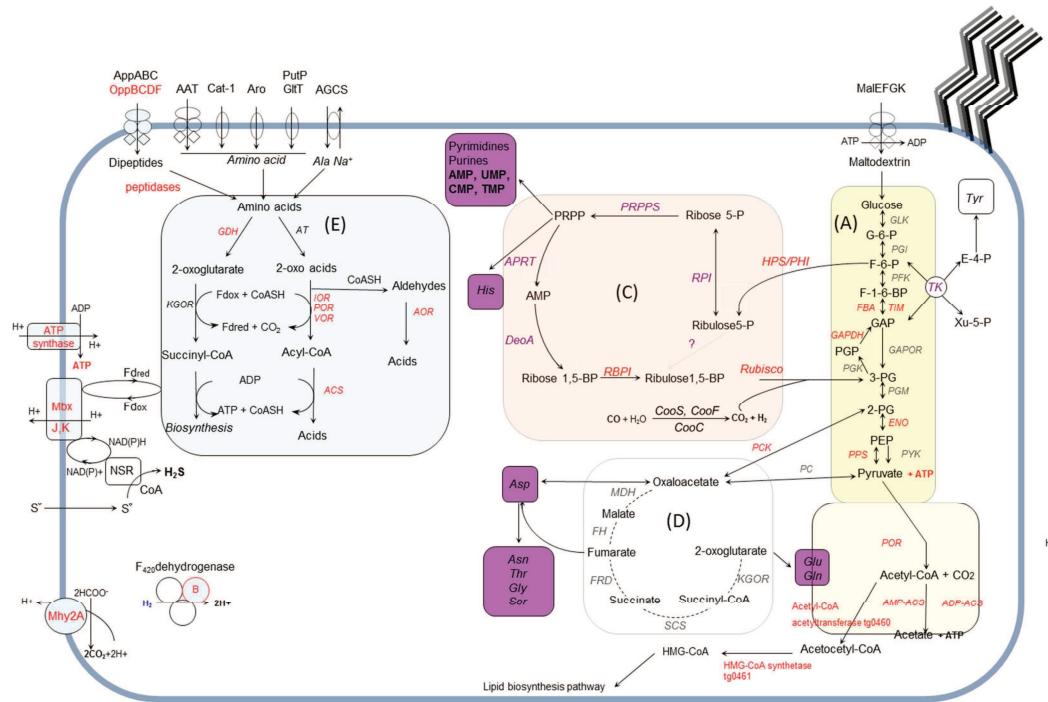


Figure 55 : Illustration schématique de la présence des protéines acétylées sur les principales voies du métabolisme central chez *Thermococcus gammatolerans*.

5. Phosphorylation chez *Thermococcus gammatolerans*.

Ce premier coup de sonde met en évidence 46 peptides validés et phosphorylés sur des sérines (24), thréonines (8) ou tyrosines (14). In fine on note la présence de 39 protéines phosphorylées. 34 d'entre elles sont mono-phosphorylées, 4 di-phosphorylées (Thermosome alpha subunit (Chaperonin alpha subunit), Threonyl-tRNA synthetase, Phosphoenolpyruvate carboxykinase [GTP], Conserved hypothetical protein), et une porte quatre phosphorylations, Phosphoenolpyruvate synthase (PEP synthase) Tableau9. Beaucoup de ces protéines sont impliquées dans la régulation du métabolisme central, d'autres dans des domaines très différents tels celui de la translation ou de la transcription comme, Translation elongation factor EF-2, RecJ-like phosphoesterase family protein, RecA family AAA ATPase, Predicted DNA-binding protein containing OB-fold nucleic acid binding domain, RNA-binding protein FAU-1, ribonuclease E homolog.

Cependant une optimisation plus poussée du protocole expérimental et son application à tous les échantillons biologiques disponibles, permettrait d'avoir une évaluation plus précise du phosphoprotéome de *Thermococcus gammatolerans* qui ne se résume pas certainement à ces 39 premières protéines.

Tableau 9 : Protéines phosphorylées chez *Thermococcus gammatolerans* et peptides phosphorylés associés.

| Accession | Description | sequence | score | modifications |
|-----------|--|--|----------------------------------|--|
| tgam08585 | Phosphoenolpyruvate synthase (pyruvate,water dikinase) (PEP synthase) | ASAIITDEGGRTSHAAIVSR VNVMPEVAER TLDAPTFEFK VNVMPEVAER | 52.77 43.65 27.48 27.33 | Phospho (ST) [12] Phospho (ST) [4], Oxidation (M) [5] Phospho (ST) [6] Phospho (ST) [4] |
| tgam01002 | Conserved hypothetical protein | LVGPDVDFDDPAFDYAR ELGLYGAGQDLLK | 76.10 31.18 | Phospho (Y) [15] Phospho (Y) [5] |
| tgam01365 | Thermosome alpha subunit (Thermosome subunit 1) (Chaperonin alpha subunit) | GGSEDFSSDLLD DVSPEDVETLKK | 63.69 26.93 | Phospho (ST) [8] Phospho (ST) [9] |
| tgam08252 | Threonyl-tRNA synthetase | TVPEEGVSKKEER FVNYEIAK | 45.64 24.81 | Phospho (ST) [9] Phospho (Y) [4] |
| tgam14714 | Phosphoenolpyruvate carboxykinase [GTP] | KTYFTGAYPSMCGK TYFTGAYPSMCGK | 29.24 25.97 | Phospho (ST) [10], Oxidation (M) [11] Phospho (ST) [9], Oxidation (M) [10] |
| tgam01153 | Methylthioadenosine phosphorylase | IGIGGSGVYGVFEPK | 58.81 | Phospho (ST) [7] |
| tgam01418 | Nucleoside triphosphate phosphohydrolase | IDGPFLLDDSGLFVEALK | 70.13 | Phospho (ST) [10] |
| tgam01825 | Transcriptional regulatory protein, arsR family | ILSLLR | 21.43 | Phospho (ST) [3] |
| tgam02117 | [NiFe] hydrogenase maturation protein | VYPSQYIGNTGK | 28.19 | Phospho (Y) [6] |
| tgam02180 | Ketoisovalerate oxidoreductase subunit vorA | VVSGNYAAAYAVK | 37.32 | Phospho (Y) [6] |
| tgam02189 | Pyruvate/ketoisovalerate oxidoreductases common gamma subunit | GGQGAVTAANILASAAFK | 111.53 | Phospho (ST) [7] |
| tgam02556 | Triosephosphate isomerase | ALELGSVGVLLASGVTK | 77.22 | Phospho (ST) [13] |
| tgam02608 | D-3-phosphoglycerate dehydrogenase | AGFEILYEEYPDEER | 37.54 | Phospho (Y) [10] |
| tgam02947 | Undylate kinase | AGSGGVVDALAAK | 44.59 | Phospho (ST) [4] |
| tgam03117 | Nicotinate-nucleotide pyrophosphorylase | TIDVYFIR | 51.41 | Phospho (Y) [5] |
| tgam03347 | Glutamate synthase (NADPH), homotetrameric | LGCDVTIAYR | 59.42 | Phospho (Y) [9] |
| tgam04102 | LSU ribosomal protein L12A | KVLEAAGVSPDEAR | 58.83 | Phospho (ST) [9] |
| tgam04716 | RecJ-like phosphoesterase family protein | ILILGGSGIGR | 49.07 | Phospho (ST) [8] |
| tgam05007 | Phosphopentomutase | EHGDAGVMITASHNPPTDNGIK | 57.24 | Phospho (ST) [12] |
| tgam05098 | RecA family AAA ATPase | YGVVEEYVAR | 48.57 | Phospho (Y) [6] |
| tgam06027 | 2,3-bisphosphoglycerate-independent phosphoglycerate mutase | LGILGQQDPIKPGQAPAGSDTAHLSIFGYDPYK | 35.81 | Phospho (ST) [18] |
| tgam07352 | L-asparaginase/archaeal Glu-tRNA ^{Gln} amidotransferase subunit D | VNLNVYSNGR | 35.51 | Phospho (ST) [7] |
| tgam07463 | Peptidase, prolyl oligopeptidase family | EIPFFWFNGVGVVYGK | 52.92 | Phospho (Y) [13] |
| tgam08012 | Serine/threonine protein kinase | HKTLIYSSPEQIVNPELNDAR | 49.68 | Phospho (ST) [3] |
| tgam08356 | Deoxyhypusine synthase | VFLGYTSNIISGLR | 41.37 | Phospho (ST) [7] |
| tgam08448 | Predicted DNA-binding protein containing OB-fold nucleic acid binding domain | LNPNIPIYKTR | 35.69 | Phospho (ST) [9] |
| tgam08692 | Elongation factor 1-alpha (EF-1-alpha) (Elongation factor Tu) (EF-Tu) | IPIQDVYSIK | 33.99 | Phospho (ST) [8] |
| tgam09364 | 3-hydroxy-3-methylglutaryl-coenzyme A reductase | GCSALTEAGGVK | 50.76 | Phospho (ST) [3] |
| tgam09509 | Nicotinamide-nucleotide adenyllyltransferase | DILSATEIR | 44.18 | Phospho (ST) [4] |
| tgam10949 | Peptidase U62, modulator of DNA gyrase. PmbA/TldD protein homolog | VFVNSEGAYVK | 37.81 | Phospho (Y) [9] |
| tgam11302 | Translation elongation factor EF-2 | TGQEVYLINSK | 58.12 | Phospho (Y) [6] |
| tgam13410 | Phosphoribosyltransferase | VLIVDDISDTGESLTLAK | 37.64 | Phospho (ST) [13] |
| tgam14541 | Ribulose 1,5-bisphosphate carboxylase large subunit (RuBisCO) | NFSGPAFGIEGVR | 29.43 | Phospho (ST) [3] |
| tgam14635 | Replication factor A complex, RPA41 subunit | LVLWDSQVQVK | 26.65 | Phospho (ST) [6] |
| tgam15099 | Glutamate dehydrogenase C-terminal domain | VVAVSDSKGGIYNPDGLDPDEVLK | 52.89 | Phospho (ST) [7] |
| tgam15193 | Leucyl-tRNA synthetase | LTREFFDYIFR | 21.77 | Phospho (Y) [8] |
| tgam15242 | Aspartate aminotransferase | GAFYVWAK | 44.32 | Phospho (Y) [4] |
| tgam16987 | RNA-binding protein FAU-1, ribonuclease E homolog | GIYSTALTK | 45.53 | Phospho (ST) [5] |
| tgam17570 | UTP-glucose-1-phosphate uridylyltransferase | AVIPAAGLGTTR | 23.70 | Phospho (ST) [10] |

On peut cependant déjà comparer la liste des 154 protéines acétylées et cette première liste de protéines phosphorylées. On constate que 17 protéines (Figure 56) sont à la fois acétylées et phosphorylées. Ce sont : **tgam13410** (Phosphoribosyltransferase), **tgam 02608** (D-3-phosphoglycerate dehydrogenase), **tgam08692** (Elongation factor 1-alpha), **tgam15099** (Glutamate dehydrogenase C-terminal domain), **tgam03347** (Glutamate synthase (NADPH), homotetrameric), **tgam02180** (Ketoisovalerate oxidoreductase subunit vorA), **tgam15193** (Leucyl-tRNA synthetase), **tgam04102** (LSU ribosomal protein L12A), **tgam10949** (Peptidase U62, modulator of DNA gyrase), **tgam07463** (Peptidase, prolyl oligopeptidase family), **tgam08585** (Phosphoenolpyruvate synthase (pyruvate,water dikinase)), **tgam05098** (RecA-superfamily ATPase, Rad55-like protein), **tgam14635** (Replication factor A complex, RPA41 subunit), **tgam14541** (Ribulose 1,5-bisphosphate carboxylase large subunit (RuBisCO) (rbcL)), **tgam01365** (Thermosome alpha subunit (Thermosome subunit 1) (Chaperonin alpha subunit) (thsA)), **tgam011302** (Translation elongation factor EF-2 (fusA)), **tgam02556** (Triosephosphate isomerase (tpiA))

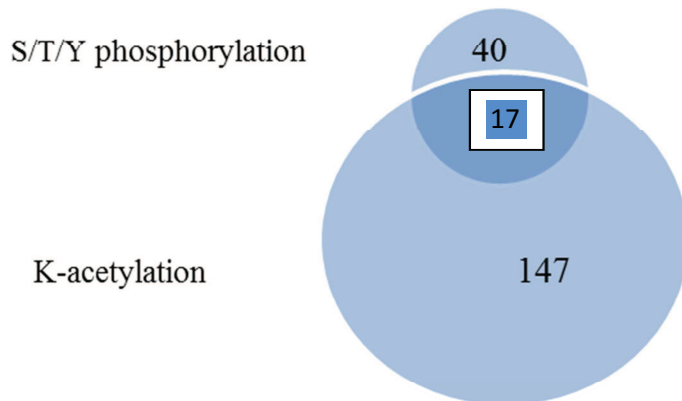


Figure 56 : Recouvrement des protéines modifiées post-traductionnellement chez *Thermococcus Gammatolerans* (acétylation et phosphorylation).

Dès 2012, des publications ont généralisé aux bactéries le phénomène de MPTs communicantes entre phosphorylations et acétylations (van Noort, Seebacher et al. 2012) (Soufi, Soares et al. 2012). Ce phénomène suppose la présence possible sur une même protéine de deux types de MPT. Cette première condition est réunie ici. Restera à démontrer le type d'interférence qui chez cet archaee peut exister entre phosphorylations et acétylations

3.5.2 Conclusions et perspectives

C'est bien au cas par cas que la recherche de MPT doit se faire chez les procaryotes. Acétylations et phosphorylations sont bien présentes chez cet organisme. Les acétylations à l'instar des phosphorylations sont réputées moduler l'activité des protéines qui les portent.

Il serait intéressant de savoir si ces MPT jouent un rôle dans la résistance aux stress extrêmes que peuvent supporter ces organismes. Mais aussi d'étudier plus en détail les relations de dépendance qui doivent exister entre ces deux MPT portées parfois par les mêmes protéines.

3.6 – Définir des objectifs en biologie structurale à travers le prisme de la protéogénomique, l'exemple de la protéine TGAM_1934 de l'archée *Thermococcus gammatolerans*

3.6.1 Analyse structurale d'une protéine particulière de l'Archaea *Thermococcus gammatolerans* (TGAM_1934) dont le choix a été guidé par des travaux de protéogénomique

Cette étude ne traite pas de MPT, mais a été réalisée sur *Thermococcus gammatolerans*, elle illustre la capacité de la protéogénomique à sélectionner une protéine hypothétique pour laquelle il n'existe aucune similarité de séquence. En choisissant une protéine dont l'abondance et l'absence de similarité de séquence est révélé par la protéomique, on introduit le concept de protéomique structurale. Appliquée sur une petite protéine abondante, TGAM_1934, on met en évidence une structure originale.

RESEARCH ARTICLE

Prioritizing targets for structural biology through the lens of proteomics: The archaeal protein TGAM_1934 from *Thermococcus gammatolerans*

Yin-Shan Yang¹, Bernard Fernandez², Arnaud Lagorce³, Valérie Aloit², Karine Montet De Guillen¹, Jean-Baptiste Boyer², Alain Dedieu², Fabrice Confalonieri³, Jean Armengaud² and Christian Roumestand¹

¹ Centre de Biochimie Structurale, Universités de Montpellier, Montpellier, France

² Laboratoire de Biochimie des Systèmes Perturbés, CEA, DSV, IBEB, Bagnols-sur-Cèze, France

³ Laboratoire de Génomique des Archaea, Institut de Génétique et de Microbiologie, Université Paris-Sud 11, Orsay, France

ORFans are hypothetical proteins lacking any significant sequence similarity with other proteins. Here, we highlighted by quantitative proteomics the TGAM_1934 ORFan from the hyperthermophilic *Thermococcus gammatolerans* archaeon as one of the most abundant hypothetical proteins. This protein has been selected as a priority target for structure determination on the basis of its abundance in three cellular conditions. Its solution structure has been determined using multidimensional heteronuclear NMR spectroscopy. TGAM_1934 displays an original fold, although sharing some similarities with the 3D structure of the bacterial ortholog of frataxin, CyaY, a protein conserved in bacteria and eukaryotes and involved in iron-sulfur cluster biogenesis. These results highlight the potential of structural proteomics in prioritizing ORFan targets for structure determination based on quantitative proteomics data. The proteomic data and structure coordinates have been deposited to the ProteomeXchange with identifier PXD000402 (<http://proteomecentral.proteomexchange.org/dataset/PXD000402>) and Protein Data Bank under the accession number 2mcf, respectively.

Received: December 3, 2013

Revised: October 1, 2014

Accepted: October 24, 2014

Keywords:

Genome annotation / High-throughput proteomics / Microbiology / NMR structure / ORFan / Protein evolution



Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

The current avalanche of genome sequencing data has resulted in an impressive number of proteins in the NCBI nonredundant database with around 20 million accession numbers. Large genome sequencing projects have indicated

the presence of many hypothetical proteins that are not related to any previously characterized proteins. Efforts of structural biologists to cover the protein structure diversity have resulted in a large number of structures (more than 80 000 to date) deposited in the Protein Data Bank (PDB), but only a small fraction of all structures solved in recent years have been identified as new folds. Structural genomics involves deciphering the 3D structure of proteins on a genome-wide scale. It has been previously reported that structural genomics of hypothetical proteins has revealed an enrichment of novel folds compared with the general set of proteins registered in the PDB [1]. Indeed, such "new structure" oriented programs are considered to be the largest contributors of novel structural leverage [2].

Correspondence: Dr. Jean Armengaud, Laboratoire de Biochimie des Systèmes Perturbés, CEA Marcoule, DSV, IBEB, SBTN, LBSP, F-30207 BAGNOLS-SUR-CEZE, France

E-mail: jean.armengaud@cea.fr

Fax: +33 0 4 66 79 19 05

Abbreviations: NOE, Nuclear Overhauser effect; PDB, Protein Data Bank

Hypothetical proteins lacking any significant sequence similarity with other proteins are called ORFans, standing for “Orphan ORFs”, and by extension may comprise poorly conserved ORFs found in a very restricted number of species. Mini-proteins, defined as polypeptides containing no more than 100 amino acids, are mostly species-specific and contribute in large part to ORFans. Compared with hypothetical conserved proteins, species-specific ORFans have been largely neglected [3–10]. Whether ORFans may or may not encode proteins has been the subject of many discussions, as their annotation may be due to “overcalling” ORFs with automatic annotation software [11]. Definitively, such a question requires experimental data, as illustrated in [8, 12–15], but obtaining results is time consuming. Proteogenomics, the use of high-throughput proteomic data to better annotate genomes, could help in clarifying the true existence of ORFans at the genome-wide scale [16].

The archaeal ancestor has been proposed as being most ancient of the ancestors of bacteria and eukarya based on the evolutionary history of proteinfold families and proteomes [17]. Thus, archaeal targets have been widely exploited in structural genomics as possible relics of primitive proteins. Extremophiles are also popular, due to their specific characteristics. Here, we have focused on the proteins of *Thermococcus gammatolerans* EJ3, an archaeon that was recently characterized as the most radiotolerant archaeal organism, demonstrating recovery from gamma-irradiation at a dose of 30 kGy [18]. The growth conditions and radiotolerance characteristics of this anaerobic sulfur-reducing organism make *T. gammatolerans* a fascinating cellular model when defining the extreme edges of conditions for life on Earth. *Thermococcus gammatolerans* has been isolated from samples collected from hydrothermal chimneys located in the Guaymas basin, Gulf of California [18]. Its genome has been sequenced and annotated based on a large proteomic dataset [19]. The organism was recently shown to exhibit strong resistance to various metals, that is, cobalt, cadmium, and zinc, as well as tolerance to copper, nickel, and arsenate [20]. In the present study, we selected an interesting target for structure determination based on its cellular abundance determined by proteomics. We identified the ORFan, TGAM_1934, as a key cellular player, based on its cellular abundance assayed by proteomics. We solved the structure of TGAM_1934 by NMR and found a fold that presents some similarities with the 3D structure of the CyaY protein.

2 Materials and methods

2.1 Strain and growth conditions

Thermococcus gammatolerans EJ3 strain was cultivated in serum bottles or in 1 L Schott bottles, under anaerobic conditions at 85°C in complex organic medium (VSM), as previously described [19, 21]. Briefly, the VSM medium comprised 1 g/L yeast extract, 4 g/L Bacto Tryptone, 3 g/L PIPES, 20 g/L NaCl, 0.5 g/L trisodium citrate, 0.05 g/L NaBr, 0.25 g/L KCl,

0.02 g/L boric acid, 0.01 g/L SrCl₂·6H₂O, 5 mL 20% MgSO₄, 1 mL 5% CaCl₂, 1 mL 5% KH₂PO₄. The pH was adjusted to 6.8 by addition of HCl. Inorganic sulfur (S⁰) was added at a concentration of 2 g/L. Na₂S₉H₂O was added at final concentration of 0.1% to reduce the oxygen dissolved in the medium, until the color of the resazurin sodium salt (1 mg/L) became clear. Microbial culture densities were monitored by optical microscopy using a Thoma cell. For γ-ray irradiation assays, exponentially growing cells (7 × 10⁷ cells/mL) were filtered to eliminate sulfur compounds and then, concentrated ten times by centrifugation in Hungate tubes [21]. Cells were subjected to 2500 Gy (100% survival [21]) using a ¹³⁷Cs γ-ray source (40 Gy/min). As a control, nonirradiated cells were kept on ice. Irradiated and nonirradiated cells were then cultivated for 120 min at 85°C for recovery in fresh VSM-S⁰ medium, on a reciprocal shaker, before centrifugation.

2.2 RT-PCR analyses

Total RNAs were extracted with TriReagent (Sigma). Residual DNA was removed with DNA-free DNase (Ambion). RNA quality and quantity were estimated with Experion Automated Electrophoresis System (Bio-Rad Laboratories) and Nanodrop BD-1000 spectrophotometer (Nanodrop Technologies), respectively. The absence of genomic DNA in RNA samples was verified by PCR with several gene-specific couples of primers using 2 μg total RNA as template. Reverse transcriptions were carried out on 5 μg total RNA with gene-specific reverse primers and SuperScriptIII as the RT (Invitrogen, Carlsbad, CA) as described in [20]. Coexpression of TGAM_1933, TGAM_1934, and TGAM_1935 genes was analyzed on RT products by PCR using forward and reverse primers. As a control, the same PCR reactions were performed on genomic DNA. Standard conditions were used with Dream Taq DNA Polymerase (Thermo Scientific), that is, DNA denaturation at 95°C for 5 min followed by 25 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s for small PCR products (164–204 bp) or 90 s for the 881–1631 bp PCR products, followed by a final extended elongation step of 5 min. Reaction products were analyzed by electrophoresis on 1% agarose gel followed by ethidium bromide staining.

2.3 Protein extracts and *in solution* proteolysis

For protein extraction from *T. gammatolerans* cells, the lysis buffer comprised 50 mM Tris-HCl buffered at pH 8.2, and containing complete, EDTA-free, protease inhibitor cocktail (Roche, 1 tablet per 50 mL), phosphatase inhibitors 2 and 3 (1:100, v/v, each), 8 M urea, 75 mM NaCl, 50 mM β-glycerophosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1 mM MgCl₂, and 1:10 000 v/v Benzamide 25 KU. Cells were resuspended in 3.5 mL lysis buffer per gram of cells (wet weight) and then disrupted by sonication. The resulting sample was centrifuged and the supernatant collected. Protein concentration was measured

using the Bradford method with BSA as the standard. The proteins were trypsinized in solution with microwave assistance. The samples were first supplemented with 0.1 M DTT to reach a final concentration of 10 mM, incubated for 1 h at 37°C, and then further supplemented with 50% formic acid solution to obtain a 6% (v/v) final concentration. Samples were dispensed into safe seal microcentrifuge tubes (30 µL of sample per tube) and exposed to 750 W for 2 min 30 s in a domestic microwave. After denaturation, the samples were placed on ice and supplemented with 0.5 M Tris-base solution to reach pH 8–8.5 (about 160 µL per tube). Following this, 0.2 M iodoacetamide was added to a tenfold dilution. After an incubation of 30 min in the dark at room temperature, proteolysis was triggered with a solution of 0.5 µg/µL trypsin in 0.01% TFA in water, to obtain a 2:100 w/w final ratio of trypsin per total protein amount. After an overnight incubation at 37°C, proteolysis was stopped with 20 µL 50% formic acid solution. The extent of proteolysis was evaluated by SDS-PAGE on 4–12% Bis-Tris gradient NuPAGE gels (Invitrogen).

2.4 OFFGEL 3100 electrophoresis

A sample comprising 1 mg tryptic digest was divided into four equal fractions, each of which was desalted using a C-18 Macro Spin Column (Harvard), the binding capacity of these columns being 300 µg peptide. The desalted peptides were dried under speed vacuum and resuspended in 1.25 × Peptide OFFGEL stock solution (Agilent) as recommend by the supplier. Electrophoresis was performed for 30 h on an OFFGEL 3100 device (Agilent) using a 12-well, pH 3–10 IPG strip. The 12 resolved fractions were collected, dried under speed vacuum, desalted using a C-18 Macro Spin Column (Harvard), dried, and finally resuspended in 100 µL 0.1% TFA. Each OFFGEL fraction was diluted fivefold in 0.1% TFA immediately prior to nano-LC-MS/MS analysis.

2.5 Shotgun nano-LC-MS/MS analysis

Nano-LC-MS/MS experiments were carried out with an LTQ-Orbitrap XL instrument (ThermoFisher) coupled to an Ultimate3000 LC system (Dionex-LC Packings), as previously described [22]. Samples (10 µL) were loaded and concentrated online on an Acclaim Pepmap 100 C-18 reverse-phase precolumn (LC Packings). They were then resolved on an Acclaim Pepmap 100 C-18 15 cm reverse-phase column (LC Packings) operated with a 0.3 µL/min flow rate for a 100-min gradient from 5 to 95% solvent B (80% CH₃CN, 0.1% HCO₂H), solvent A being 0.1% HCO₂H. MS/MS spectra were acquired in 5–90 min with a full scan from *m/z* 300 to 1800 in data-dependent mode following a TOP3 strategy consisting of a high mass accuracy scan in the Orbitrap analyzer (30 000 resolution; internal calibration enabled) followed by MS/MS scans in the LTQ on the three most abundant parent ions (minimum signal required set at 15 000 and potential charge states of 2⁺ and 3⁺), with

dynamic exclusion of previously selected ions (exclusion duration of 60 s), and normalized collision energy set at 35% of the amplitude of the radio frequencies applied to the selected ion. The lock mass option on the LTQ Orbitrap XL mass spectrometer was enabled in MS mode. Polydimethylcyclosiloxane ions generated in the electrospray process from ambient air (monoprotonated [(CH₃)₂SiO]₆ with *m/z* at 445.120024) were used for internal recalibration in real time, as recommended by the mass spectrometer manufacturer.

2.6 Proteomic data interpretation and repository

Peak lists were produced as described previously [34]. MS/MS spectra were searched with MASCOT (version 2.2.04, Matrix Science) against an in-house polypeptide sequence database comprising the 2157 *T. gammatolerans* annotated protein sequences [19] and totaling 636 507 amino acids. Search parameters were full-trypsin specificity, static modification of carboxyamidomethylated Cys, and dynamic modification of oxidized Met, and one maximal missed cleavage. Mass tolerances were set at 10 ppm and 0.5 Da for the parent ion and MS/MS, respectively. Peptide matches with a stringent *p*-value of 0.05 were parsed with IRMA 1.18.1 [23]. A protein was validated with at least two different detected peptides. Peptides matching multiple proteins were assigned to the polypeptide with the most specific peptides following the principle of parsimony. In terms of protein identification, the false-positive rate was calculated to be lower than 0.1% against a reversed decoy database. The data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [24] with the dataset identifier PXD000402.

2.7 Cloning and overexpression of the *TGAM_1934* gene

Thermococcus gammatolerans total DNA was extracted from cells with the DNeasy Tissue Kit (Qiagen). The synthetic oligonucleotide primers, oBj42 (5'-caccATGAAGTACGACCGTCG-3') and oBj43 (5'-AGCGAGGCACTCCTCAATG-3'), were used for PCR amplification of *TGAM_1934* (nucleotides not present in the original sequence are shown in lowercase). The 370 bp fragment obtained after amplification was electrophoretically separated on a 1.5% SeaKem GTG-agarose gel and extracted with the Qiaquick Gel Extraction kit (Qiagen) prior to cloning into pET101/D-TOPO (Invitrogen). The resulting plasmid, pSBTN-AK10, was DNA sequenced to ascertain the integrity of the nucleotide sequence. Hyperexpression of the recombinant *TGAM_1934* construct with a 6HIS-tag at its C-terminus was achieved with *Escherichia coli* BL21 Star (DE3) One Shot (Invitrogen) transformed by heat shock with pSBTN-AK10. Preliminary assays for selecting the best overexpressing clone and evaluating the solubility of the recombinant product were carried out as previously described [25, 26].

2.8 Protein expression and purification

The construct allowing the expression of TGAM_1934 was transformed into *E. coli* BL21(DE3). After protein expression (3 h, 30°C, 0.2 mM IPTG), cells were suspended in lysis buffer comprising 50 mM Tris-HCl buffered at pH 8.0 and containing 300 mM NaCl. Lysozyme at 1 µg/µL was added to aid cell disruption. Cells were disrupted with a Vibra cell 72405 sonicator (bursts of 2 s for 3 min with the probe at 60% amplitude). Cell debris was removed by centrifugation (Avanti J20XP centrifuge from Beckman Coulter with a 25.50 rotor, at 6°C, and 18 000 rpm, that is, average relative centrifugal field 26 600 × g). The supernatant was filtered with a 0.45 µm PVDF membrane (Millipore). The soluble proteins were loaded onto a 1 mL HisTrap HP column (GE Healthcare) equilibrated with the same buffer as for the lysis and operated with an AKTA system. After imidazole gradient elution, fractions containing the TGAM_1934 protein were pooled, and finally injected into a Superdex S75 16/60 (GE Healthcare) column, equilibrated with 50 mM sodium phosphate buffered at pH 7.2 and containing 50 mM NaCl. The fractions containing the pure protein were pooled, concentrated with a Vivaspin 20 centrifugal concentrator (Sartorius Stedim Biotech) to obtain 0.5 mM protein as determined by Bradford assay, and stored at -80°C until NMR analysis. Uniform ¹⁵N/¹³C labeling was obtained by growing cells in minimal M9 medium with ¹³C₆-glucose and ¹⁵NH₄Cl as the sole carbon and nitrogen sources, respectively (Cambridge Isotope Laboratories, Inc.).

2.9 NMR spectroscopy

NMR ¹H,¹⁵N double-resonance and ¹H,¹⁵N,¹³C triple-resonance experiments were performed at 20°C on Bruker AVANCE III 700 and 500 MHz spectrometers, respectively, equipped with cryogenic 5 mm Z-gradient ¹H-¹³C-¹⁵N probe heads. Backbone and C β resonance assignments were obtained using HNCA, HNCACB, CACB(CO)NH, HNCO, and HN(CO)CA standard experiments [27] carried out on the [U-¹⁵N,¹³C] TGAM_1934 sample. HN and H α assignments were established using 3D [¹H,¹⁵N] NOESY-HSQC (mixing time 100 ms) and TOCSY-HSQC (isotropic mixing: 60 ms) experiments. Water suppression was achieved with the WATERGATE sequence [28]. ¹H chemical shifts were assigned from the methyl resonance of DSS, while ¹³C and ¹⁵N chemical shifts were assigned indirectly from the absolute frequency ratios ¹⁵N/¹H = 0.101329118 and ¹³C/¹H = 0.251449530.

2.10 Structure modelization

The 1737 assigned NOE (Nuclear Overhauser Effect) peaks were classified into five categories (very strong [≤ 2.4 Å], strong [≤ 2.8 Å], medium [≤ 3.6 Å], weak [≤ 4.4 Å], or very weak [≤ 4.8 Å]). This set was completed with 202 ϕ and ψ backbone dihedral angle restraints taken out from TALOS predictions [29], and 32 χ_1 side chain dihedral angle restraints obtained from stereospecific assignments. The initial structures

of TGAM_1934 were calculated using the CYANA program [30]. A total of 200 structures were calculated at this stage. The 15 best structures, as determined from the values of the final target penalty function, were minimized according to RECOORD [31] with CNS 1.2 and analyzed with PROCHECK [32]. The rms deviations were calculated with MOLMOL [33].

3 Results and discussion

3.1 Quantitative proteomics revealed TGAM_1934 as an abundant small protein

We previously analyzed the *T. gammatolerans* proteome by a 1D SDS-PAGE shotgun procedure where 463 840 MS/MS spectra were recorded (PRIDE accession numbers #9212 to #9218). A total of 951 proteins were detected with at least two tryptic peptides (Supporting Information Table 1). These were ranked by their normalized spectral abundance factor comprising their spectral counts normalized by their polypeptide molecular weight [34]. In this catalogue, proteins annotated as “conserved hypothetical proteins” are by far the most numerous, with 177 items. A restricted set of 20 proteins are annotated as “hypothetical” but are not conserved in a majority of *Thermococci*, thus could be considered as ORFans. Table 1 lists the most abundant of these proteins and their characteristics. Among the most abundant proteins, we observed that TGAM_1934 (bold in Table 1) was well covered in terms of sequence with nine different peptides, as shown in Fig. 1. Remarkably, the N-terminus of this protein is well defined with the peptide [1–14] detected as a tricharged ion at m/z 583.9721 (error tolerance: 1.4 ppm). Because it has been noted through large-scale N-terminome analyses that a relatively large number of translational start codons are wrongly annotated [35], this piece of evidence is important. We conducted another shotgun approach, including in-solution proteolysis, after microwave denaturation of the hyperthermophilic and trypsin-recalcitrant proteins from nonirradiated and 5.0 kGy irradiated cells, OFFGEL electrophoresis, and nano-LC-MS/MS. We performed this experiment on duplicated biological samples. These proteomic data have been deposited in the ProteomeXchange repository with identifier PXD000402. A total of 221 368 MS/MS spectra were recorded and 63 108 were assigned with confidence to peptide sequences (Supporting Information Table 2). A total of 6907 peptide sequences could be observed, validating the detection of 852 proteins based on the “two peptides” rule. We detected in both conditions the TGAM_1934 protein (Supporting Information Table 3), that is, for the four analyzed samples.

3.2 TGAM_1934 is found in a highly conserved *Thermococcus* genomic context

TGAM_1934 was considered throughout this study as an ORFAN because until recently no related sequence was

Table 1. List of ORFans identified in *Thermococcus gammatolerans* proteome^{a)}

| Protein accession | Protein functional description | MASCOT score | MW (Da) | Sequence coverage (%) | No. of peptides | No. of MS/MS | NSAF | Global NSAF (%) | Low MW (<20 kDa) NSAF (%) ^{b)} |
|-------------------|--|--------------|---------------|-----------------------|-----------------|--------------|--------------|-----------------|---|
| TGAM_1885 | Hypothetical protein | 415 | 18 233 | 44 | 12 | 104 | 5.704 | 0.13% | 0.76% |
| TGAM_1934 | Hypothetical protein^{c)} | 384 | 14 666 | 47 | 8 | 22 | 1.500 | 0.03% | 0.20% |
| TGAM_1886 | Hypothetical protein | 222 | 14 331 | 27 | 5 | 19 | 1.326 | 0.03% | 0.18% |
| TGAM_1714 | hypothetical protein ^{c)} | 187 | 19 379 | 15 | 3 | 16 | 0.826 | 0.02% | 0.11% |
| TGAM_1366 | Hypothetical protein ^{c)} | 177 | 53 994 | 12 | 4 | 40 | 0.741 | 0.02% | - |
| TGAM_0210 | Hypothetical protein ^{c)} | 232 | 34 559 | 15 | 5 | 23 | 0.666 | 0.02% | - |
| TGAM_1831 | Hypothetical protein | 137 | 22 280 | 14 | 4 | 12 | 0.539 | 0.01% | - |
| TGAM_0173 | Hypothetical protein ^{c)} | 222 | 31 686 | 11 | 3 | 15 | 0.473 | 0.01% | - |
| TGAM_0987 | Hypothetical protein ^{c)} | 178 | 73 135 | 5 | 4 | 32 | 0.438 | 0.01% | - |
| TGAM_0908 | Hypothetical protein | 182 | 54 938 | 6 | 6 | 23 | 0.419 | 0.01% | - |
| TGAM_2140 | Hypothetical protein | 91 | 37 326 | 8 | 2 | 14 | 0.375 | 0.01% | - |
| TGAM_1371 | Hypothetical protein | 214 | 38 659 | 11 | 5 | 14 | 0.362 | 0.01% | - |
| TGAM_1343 | Hypothetical protein ^{c)} | 245 | 55 267 | 13 | 7 | 20 | 0.362 | 0.01% | - |
| TGAM_0681 | Hypothetical protein | 175 | 21 433 | 21 | 2 | 4 | 0.187 | 0.00% | - |
| TGAM_1202 | Hypothetical protein ^{c)} | 88 | 29 483 | 5 | 2 | 5 | 0.170 | 0.00% | - |
| TGAM_1896 | Hypothetical protein ^{c)} | 132 | 36 058 | 7 | 2 | 6 | 0.166 | 0.00% | - |
| TGAM_0924 | Hypothetical protein ^{c)} | 103 | 90 655 | 3 | 3 | 11 | 0.121 | 0.00% | - |
| TGAM_2058 | Hypothetical protein | 80 | 19 988 | 14 | 2 | 2 | 0.100 | 0.00% | 0.01% |
| TGAM_1730 | Hypothetical protein | 122 | 29 686 | 10 | 2 | 2 | 0.067 | 0.00% | - |
| TGAM_1206 | Hypothetical protein ^{c)} | 79 | 76 804 | 1 | 2 | 4 | 0.052 | 0.00% | - |

Only proteins validated with at least two distinct peptides and annotated as hypothetical, poorly conserved proteins are shown.

a) Compiled from raw data available in the PRIDE PRoteomics IDentifications database (#9212–#9218).

b) The low MW NSAF factors of five other proteins are indicated as references: TGAM_1552 (DNA-binding protein Alba): 6.11%; TGAM_0477 (archaeal histone B): 1.85%; TGAM_2001 (LSU ribosomal protein L23P): 0.67%; TGAM_1422 (eIF1A translation initiation factor): 0.55%; TGAM_0489 (LSU ribosomal protein L12A): 0.13%; TGAM_1216 (superoxide reductase): 0.12%.

c) Proteins for which a unique orthologue was detected in the very closely related *Thermococcus* sp. AM4.

known. A blast search with the TGAM_1934 sequence against the NCBI nr database resulted in the identification of a homolog, TAM4_143, from *Thermococcus* sp. AM4 and, when revising this manuscript, BD01_0811 from *Thermococcus nautili*. Both species are phylogenetically close to *T. gammatolerans* [36]. Remarkably, they were all isolated from hydrothermal vents and are hyperthermophilic and anaerobic species. Thermococci are mainly obligate heterotrophs, and share a lot of metabolic pathways used to grow on complex carbon sources. Some specific features and pathways may

be encoded in only a subset of these Thermococci, but differences are also found in very close *Thermococcus* species, such as *T. gammatolerans*, *Thermococcus* sp. AM4, and *T. nautili*. Therefore, to our knowledge, we cannot explain the presence of TGAM_1934 and homologs in *Thermococcus* sp. AM4 and *T. nautili* by known physiological differences between this subgroup and the other Thermococci. These three protein sequences share 105 of 122 identical amino acids (86%). Figure 2 shows the genomic context of these genes and the related genomic loci in other *Thermococcus* sequenced

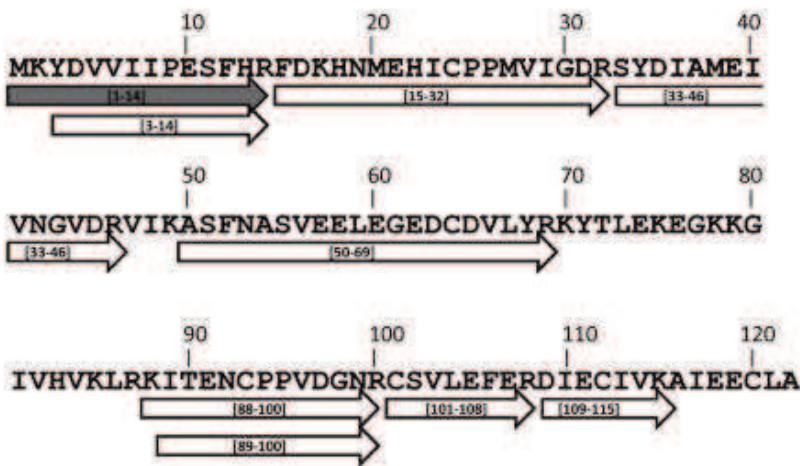


Figure 1. Sequence coverage of TGAM_1934 by proteomics. The data were extracted from the compilation of proteins identified for the proteogenomic annotation of *Thermococcus gammatolerans* [19]. Raw data are available in the PRIDE under accession #9212-#9218. The peptides assigned to MS/MS spectra are shown with arrows. The most N-terminal peptide is in gray.

species. Besides the TGAM_1934 gene, other genes identified include those encoding a conserved GPN-GTPase that was recently shown to be involved in sister chromatid maintenance in eukaryotes [37–39], the cell division Mind-2 ATPase [40], the PilT ATPase possibly involved in cell polarity [41], the Maf septum-formation protein [42], and the PhP_PolX,

a protein containing the polymerase and histidinol phosphatase domain of bacterial polymerase X. The genomic organization of the three genes, *mind2*, *pilT*, and *maf*, is strictly conserved among the *Thermococci* with *mind2* and *pilT* in the opposite orientation. Remarkably, TGAM_1934 is inserted between *pilT* and *maf*. In the opposite direction,



Figure 2. Genomic context of TGAM_1934 in *Thermococcus gammatolerans* EJ3 and comparison with closely related archaea. Annotated ORFs and their orientation for transcription are shown with arrows. TGAM_1934 and its closely related homolog are shown in yellow. The GPN-GTPase, *mind2*, *pilT*, *maf*, and *php_PolX* homologs are indicated with color arrows. The gene OCC1_06433 from *Thermococcus litoralis* is longer in size, because of the presence of an insertion sequence, encoding probably an intein.

mind2 and GPN-GTPase are conserved in almost all *Thermococcus* species. The intergenic region between TGAM_1933 and TGAM_1935 may be a hot spot for integration in the *Thermococcales*, as in seven species shown in Fig. 2 several independent integration events occurred. The presence of TGAM_1934 homologs in only the two closest *Thermococcus* species, that is, *T. sp AM4* and *T. nautili*, and their conserved genomic context, may indicate a quite recent appearance of this new protein in the *Thermococcus* lineage. Alternatively, the gene could be encoded by a common ancestor of these *Thermococci* and was subsequently lost by the other species.

3.3 TGAM_1934 and TGAM_1935 are part of the same transcription unit

In order to determine if TGAM_1933 (*maf*), TGAM_1934, TGAM_1935 (*pilT*) genes are expressed or cotranscribed, we analyzed their expression in *T. gammatolerans* exponentially growing cells by RT-PCR assays. The results presented in Supporting Information Fig. 1 indicate that these three genes are expressed in normal growth conditions, but only TGAM_1934 and TGAM_1935 encoding the PilT ATPase appeared to be coexpressed and, thus, are part of the same transcription unit. Remarkably, we detected these two proteins in our proteomic datasets with relatively similar global normalized spectral abundance factor, that is, 1.29 and 1.53, respectively (Supporting Information Tables 1 and 3).

3.4 TGAM_1934 NMR structure

The [¹H, ¹⁵N] HSQC spectrum of TGAM_1934 is presented in Supporting Information Fig. 2. From the heteronuclear experiments, we assigned more than 97.5% of the amide group resonances for the nonproline residues (eight prolines over 148 residues), and 97% of the other backbone resonances (C α , C', and H α) and of the C β resonances. The chemical shifts are available with the accession number BMRB-19438 at the BioMagResBank.

Figure 3A shows the NMR bundle representation of the 15 minimized NMR conformers from CNS 1.2 calculation. The bundle has good covalent geometry and nonbonded contacts that are in agreement with the experimental restraints. None of the selected conformers has NOE violations greater than 0.2 Å or dihedral angle violations greater than 5°. Based on PROCHECK [32], 99.6% of nonproline and nonglycine residues are within the allowed region of the Ramachandran plot. The residues in the disallowed part of the plot correspond to residues located in the C-terminal disordered segment of the protein (residues 126–148). The average backbone (heavy atoms) atomic root-mean-square deviation (rmsd) of 0.98 ± 0.12 Å (residues 1–125) indicates a good atomic precision. The statistics are summarized in Table 2. The structure

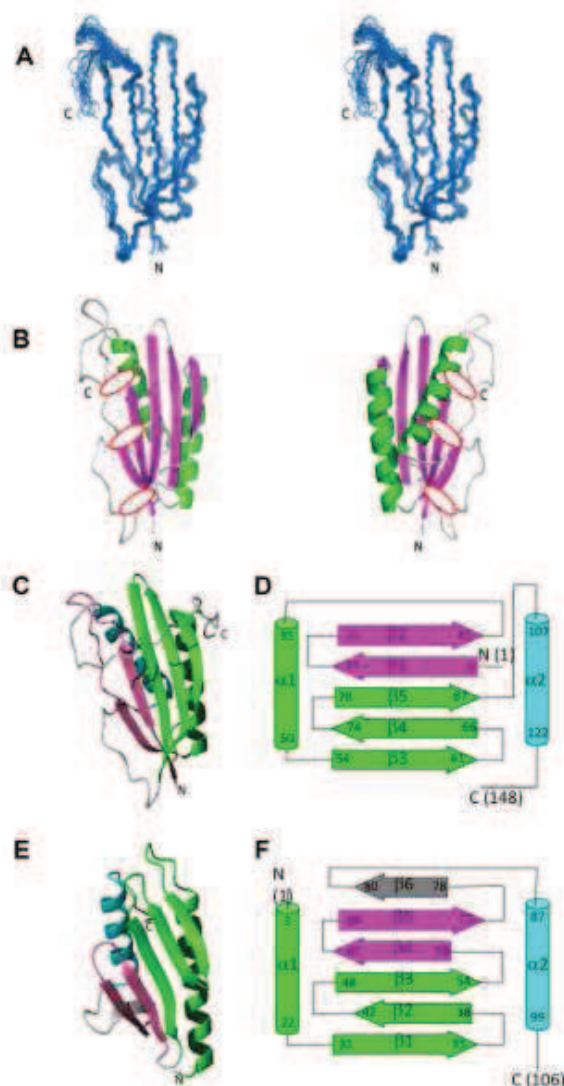


Figure 3. Solution structure of TGAM_1934 (residues 1–125) from *Thermococcus gammatolerans* and structural similarities with CyaY from *Escherichia coli*. (A) Stereo view of an overlay of the 15 deposited NMR structures with lowest energy from CNS 1.2 calculation. (B) Ribbon representation of two 180° rotated views of TGAM_1934. The three disulfide bridges are displayed in yellow and highlighted with red ellipses. In the two representations, the disordered C-terminal segment (residues 126–148) has been removed. (C) Ribbon representation of the structure of TGAM_1934 and (E) CyaY (PDB: 1EW4). The same color is used for secondary structure elements located at similar positions in the 3D structure of the two proteins. Topological diagrams of TGAM_1934 (D) and CyaY (F). The same color code has been used as in (C) and (E). The numbers indicate the secondary structure element limits. In addition to a supplementary β -strand, the β -sheet of CyaY has a topology of +1, +1, +1, +1, and +1, whereas the β -sheet of TGAM_1934 has a topology of +1, –3, –2, and –1 [43]. The α 1 helix is N-terminal in the structure of CyaY, whereas it is inserted between the β 2 and β 3 strands in the structure of TGAM_1934.

Table 2. Structure parameters for TGAM_1934

| NMR distance and dihedral constraints | |
|---|---------------------|
| Distance restraints | |
| Total NOE | 1737 |
| Intraresidue | 426 |
| Interresidue | |
| Sequential ($ i - j = 1$) | 593 |
| Medium range ($1 < i - j \leq 4$) | 223 |
| Long range ($ i - j \geq 5$) | 495 |
| Hydrogen bonds | 40 |
| Dihedral angle restraints | |
| ϕ | 99 |
| ψ | 103 |
| χ_1 | 32 |
| Structure statistics | |
| Violations | |
| Distance constraint violation ($>0.2 \text{ \AA}$) | 0 |
| Dihedral angle violation ($>5^\circ$) | 0 |
| Deviations from idealized geometry | |
| Bond lengths (\AA) | 0.0122 ± 0.0002 |
| Bond angles ($^\circ$) | 1.2870 ± 0.0258 |
| Impropers ($^\circ$) | 1.6254 ± 0.0813 |
| Ramachandran plot (%) | |
| Most favored region | 80.5 ± 1.9 |
| Additionally allowed region | 16.7 ± 2.3 |
| Generously allowed region | 2.4 ± 1.2 |
| Disallowed region | 0.4 ± 0.5 |
| Average pairwise rms deviation ^{a)} (\AA) | |
| Backbone | 0.98 ± 0.12 |
| Heavy | 1.72 ± 0.13 |

a) Pairwise rms deviation was calculated among 15 refined structures.

coordinates have been deposited in the PDB under the accession number 2mcf.

The structure of the TGAM_1934 protein from *T. gammatolerans* comprises a twisted antiparallel β -sheet formed by five β -strands ($\beta 1$: residues 3–13, $\beta 2$: residues 21–31, $\beta 3$: residues 54–61, $\beta 4$: residues 66–74, $\beta 5$: residues 78–87), flanked on one side by two amphipathic α -helices ($\alpha 1$: residues 35–50, and the C-terminal helix $\alpha 2$: residues 107–122), and by the disordered C-terminal segment (residues 126–148) on the opposite side (Fig. 3B). Aliphatic and aromatic side chains of the following residues contribute to the formation of a hydrophobic core: Val-5 and Ile-7 on $\beta 1$; Met-27 and Ile-29 on $\beta 2$; Met-38, Ile-40, Val-41, Val-44, Val-47, and Ile-48 on $\alpha 1$; Phe-52 on the loop connecting $\alpha 1$ to $\beta 3$; Val-56 on $\beta 3$; Leu-67, Tyr-71, and Leu-73 on $\beta 4$; Val-82, Val-84, and Leu-86 on $\beta 5$; Val-103, Ile-110, Ile-113, Val-114, Ile-117, and Leu-121 on $\alpha 2$. Further contributions to fold stability might arise from salt bridges between Asp-4 ($\beta 1$) and Arg-87 ($\beta 5$), Arg-32 (on the loop connecting $\beta 2$ to $\alpha 1$) and Asp-109 ($\alpha 2$), Asp-45 ($\alpha 1$) and Arg-69 ($\beta 4$), Arg-46 and/or Lys-49 ($\alpha 1$) and Asp-143 (C-terminal), Glu-57 ($\beta 3$) and Lys-70 ($\beta 4$), Asp-65 (on the turn connecting $\beta 3$ and $\beta 4$) and Lys-85 ($\beta 5$), and Lys-75 and/or Lys-77 (on the turn connecting $\beta 4$ and $\beta 5$) and Glu-118 ($\alpha 2$). Finally, three disulfide bridges strap the secondary structure elements in

the tridimensional structure of TGAM_1934: Cys-24-Cys-120 connects $\beta 2$ and $\alpha 2$, while the loop between $\beta 5$ and $\alpha 2$ is connected to $\alpha 2$ by the bridge Cys-101-Cys-112 and to the turn between $\beta 3$ and $\beta 4$ by the bridge Cys-64-Cys-93.

A search for overall structural similarities against the DALI [44] database revealed few significant topological similarities: a partial superposition was achieved with a region from the 3D structure of the β -lactamase protein family (Z-score: 5.0; rmsd: 4.3 \AA). This region is far from the active site of these proteins, and irrelevant for enzymatic activity. More interestingly, a significant match (Z-score: 3.8; rmsd: 3 \AA) was achieved with the 3D structure of CyaY [45, 46] from *E. coli*. The 12 kDa CyaY protein is the bacterial ortholog of the mitochondrial iron-binding protein frataxin, the loss of which leads to the neurodegenerative disorder Friedreich's ataxia [47, 48]. If the 3D structure of the two proteins are strikingly similar (Fig. 3C and E), their topology [50] is markedly different (Fig. 3D and F), and the iron-binding site of CyaY, formed by several glutamates and aspartates clustered in a negatively charged ridge [46], is absent in TGAM_1934, suggesting that these two proteins have unrelated functions.

4 Concluding remarks

Major concerns regarding the real existence of ORFans have been raised, as most genome annotation procedures tend to "overcall" ORFs [11, 47]. Because ORFans lack any significant sequence similarity with other proteins, experimental proof of their existence is thus necessary. Because of their hypothetical nature, such ORFans represent difficult but valuable targets for structure determination. Here, we have shown that structural analysis of the TGAM_1934 ORFan from the radioresistant *T. gammatolerans* archaeon is worthy of being carried out on the basis of quantitative proteomics. To achieve an insight into its function, one may suggest inactivating the TGAM_1934 gene in *T. gammatolerans*, but genetic tools are not yet available for this species. The structure of TGAM_1934 in solution, solved by multidimensional heteronuclear NMR spectroscopy, indicates a global fold with similarities to the bacterial ortholog of frataxin, CyaY, a protein conserved in bacteria and eukaryotes and involved in iron-sulfur cluster biogenesis [45, 46]. Whether an ancestral cognate protein could explain the structural similarities between TGAM_1934 and CyaY is an open question. Alternatively, the creation of a new sequence and convergent evolution toward the same fold appears to be more probable, due to the low molecular weight of the polypeptide and the tremendous prokaryote and viral diversity [48–50]. Further studies are needed to better document the origin and evolution of ORFans [3, 10, 51–54]. The concept of "structural proteomics," that is, prioritizing ORFan targets for structure determination based on quantitative proteomics data, could in the near future be highly efficient in focusing the structural efforts of the scientific community on the most crucial protein targets.

This study was supported by the Commissariat à l'Énergie Atomique et aux Énergies Alternatives, the Centre National de la Recherche Scientifique, the University Paris-Sud 11, the Agence Nationale de la Recherche (ANR-12-BSV6-0012-01), and the French Infrastructure for Integrated Structural Biology (FRISBI, ANR-10-INSB-05-01).

The authors have declared no conflict of interest.

5 References

- Galkin, A., Sarikaya, E., Lehmann, C., Howard, A., Herzberg, O., X-ray structure of HI0817 from Haemophilus influenzae: protein of unknown function with a novel fold. *Proteins* 2004, 57, 874–877.
- Nair, R., Liu, J., Soong, T. T., Acton, T. B. et al., Structural genomics is the largest contributor of novel structural leverage. *J. Struct. Funct. Genomics* 2009, 10, 181–191.
- Daubin, V., Ochman, H., Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* 2004, 14, 1036–1042.
- Fischer, D., Rational structural genomics: affirmative action for ORFans and the growth in our structural knowledge. *Protein Eng.* 1999, 12, 1029–1030.
- Hanson, A. D., Pribat, A., Waller, J. C., de Crecy-Lagard, V., "Unknown" proteins and "orphan" enzymes: the missing half of the engineering parts list—and how to find it. *Biochem. J.* 2010, 425, 1–11.
- Siew, N., Azaria, Y., Fischer, D., The ORFanage: an ORFan database. *Nucleic Acids Res.* 2004, 32, D281–D283.
- Siew, N., Fischer, D., Analysis of singleton ORFans in fully sequenced microbial genomes. *Proteins* 2003, 53, 241–251.
- Siew, N., Fischer, D., Structural biology sheds light on the puzzle of genomic ORFans. *J. Mol. Biol.* 2004, 342, 369–373.
- Yin, Y., Fischer, D., On the origin of microbial ORFans: quantifying the strength of the evidence for viral lateral transfer. *BMC Evol. Biol.* 2006, 6, 63.
- Yin, Y., Fischer, D., Identification and investigation of ORFans in the viral world. *BMC Genomics* 2008, 9, 24.
- Armengaud, J., A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr. Opin. Microbiol.* 2009, 12, 292–300.
- Abergel, C., Monchois, V., Chenivresse, S., Jeudy, S., Claverie, J. M., Crystallization and preliminary crystallographic study of b0220, an "ORFan" protein of unknown function from *Escherichia coli*. *Acta Crystallogr. D Biol. Crystallogr.* 2000, 56, 1694–1695.
- Ang, D., Georgopoulos, C., An ORFan no more: the bacteriophage T4 39.2 gene product, Nwgl, modulates GroEL chaperone function. *Genetics* 2012, 190, 989–1000.
- Hakim, M., Ezerina, D., Alon, A., Vonshak, O., Fass, D., Exploring ORFan domains in giant viruses: structure of mimivirus sulfhydryl oxidase R596. *PLoS One* 2012, 7, e50649.
- Shmueli, H., Dinitz, E., Dahan, I., Eichler, J. et al., Poorly conserved ORFs in the genome of the archaea Halobacterium sp. NRC-1 correspond to expressed proteins. *Bioinformatics* 2004, 20, 1248–1253.
- Armengaud, J., Microbiology and proteomics, getting the best of both worlds! *Environ. Microbiol.* 2013, 15, 12–23.
- Kim, K. M., Caetano-Anolles, G., The evolutionary history of protein fold families and proteomes confirms that the archaeal ancestor is more ancient than the ancestors of other superkingdoms. *BMC Evol. Biol.* 2012, 12, 13.
- Jolivet, E., L'Haridon, S., Corre, E., Forterre, P., Prieur, D., Thermococcus gammatolerans sp. nov., a hyperthermophilic archaeon from a deep-sea hydrothermal vent that resists ionizing radiation. *Int. J. Syst. Evol. Microbiol.* 2003, 53, 847–851.
- Zivanovic, Y., Armengaud, J., Lagorce, A., Leplat, C. et al., Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea. *Genome Biol.* 2009, 10, R70.
- Lagorce, A., Fourcans, A., Dutertre, M., Bouyssiere, B. et al., Genome-wide transcriptional response of the archaeon *Thermococcus gammatolerans* to cadmium. *PLoS One* 2012, 7, e41935.
- Tapias, A., Leplat, C., Confalonieri, F., Recovery of ionizing-radiation damage after high doses of gamma ray in the hyperthermophilic archaeon *Thermococcus gammatolerans*. *Extremophiles* 2009, 13, 333–343.
- de Groot, A., Dulerio, R., Ortet, P., Blanchard, L. et al., Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. *PLoS Genet.* 2009, 5, e1000434.
- Dupierris, V., Masselon, C., Court, M., Kieffer-Jaquinod, S., Bruley, C., A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMA. *Bioinformatics* 2009, 25, 1980–1981.
- Vizcaino, J. A., Cote, R. G., Csordas, A., Dianas, J. A. et al., The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013, 41, D1063–D1069.
- Armengaud, J., Fernandez, B., Chaumont, V., Rollin-Genetet, F. et al., Identification, purification, and characterization of an eukaryotic-like phosphopantetheine adenylyltransferase (coenzyme A biosynthetic pathway) in the hyperthermophilic archaeon *Pyrococcus abyssi*. *J. Biol. Chem.* 2003, 278, 31078–31087.
- Gabant, G., Auxilien, S., Tuszyńska, I., Locard, M. et al., THUMP from archaeal tRNA:m22G10 methyltransferase, a genuine autonomously folding domain. *Nucleic Acids Res.* 2006, 34, 2483–2494.
- Sattler, M., Schleucher, J., Griesinger, C., Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Mag. Reson. Spectrosc.* 1999, 34, 93–158.
- Piotto, M., Saudek, V., Sklener, V., Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 1992, 2, 661–665.
- Cornilescu, G., Delaglio, F., Bax, A., Protein backbone angle restraints from searching a database for chemical shift

- and sequence homology. *J. Biomol. NMR* 1999, **13**, 289–302.
- [30] Guntert, P., Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* 2004, **278**, 353–378.
- [31] Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z. et al., RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagRes-Bank. *Proteins* 2005, **59**, 662–672.
- [32] Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., Thornton, J. M., AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 1996, **8**, 477–486.
- [33] Koradi, R., Billeter, M., Wuthrich, K., MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 1996, **14**, 51–55, 29–32.
- [34] Christie-Oleza, J. A., Fernandez, B., Nogales, B., Bosch, R., Armengaud, J., Proteomic insights into the lifestyle of an environmentally relevant marine bacterium. *ISME J.* 2012, **6**, 124–135.
- [35] Baudet, M., Orset, P., Gaillard, J. C., Fernandez, B. et al., Proteomics-based refinement of *Deinococcus deserti* genome annotation reveals an unwonted use of non-canonical translation initiation codons. *Mol. Cell. Proteomics* 2010, **9**, 415–426.
- [36] Gorlas, A., Croce, O., Oberto, J., Gauiliard, E. et al., *Thermococcus nautili* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal deep-sea vent. *Int. J. Syst. Evol. Microbiol.* 2014, **64**, 1802–1810.
- [37] Alonso, B., Beraud, C., Meguellati, S., Chen, S. W. et al., Eukaryotic GPN-loop GTPases paralogs use a dimeric assembly reminiscent of archeal GPN. *Cell Cycle* 2013, **12**, 463–472.
- [38] Alonso, B., Chaussinand, G., Armengaud, J., Godon, C., A role for GPN-loop GTPase γ GPN1 in sister chromatid cohesion. *Cell Cycle* 2011, **10**, 1828–1837.
- [39] Gras, S., Chaumont, V., Fernandez, B., Carpentier, P. et al., Structural insights into a new homodimeric self-activated GTPase family. *EMBO Rep.* 2007, **8**, 569–575.
- [40] Sakai, N., Yao, M., Itou, H., Watanabe, N. et al., The three-dimensional structure of septum site-determining protein MinD from *Pyrococcus horikoshii* OT3 in complex with Mg-ADP. *Structure* 2001, **9**, 817–826.
- [41] Bulyha, I., Lindow, S., Lin, L., Bolte, K. et al., Two small GTPases act in concert with the bactofilin cytoskeleton to regulate dynamic bacterial cell polarity. *Dev. Cell.* 2013, **25**, 119–131.
- [42] Wirth, R., Bellack, A., Bertl, M., Bilek, Y. et al., The mode of cell wall growth in selected archaea is similar to the general mode of cell wall growth in bacteria as revealed by fluorescent dye analysis. *Appl. Environ. Microbiol.* 2011, **77**, 1556–1562.
- [43] Richardson, J. S., The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 1981, **34**, 167–339.
- [44] Holm, L., Sander, C., Protein folds and families: sequence and structure alignments. *Nucleic Acids Res.* 1999, **27**, 244–247.
- [45] Cho, S. J., Lee, M. G., Yang, J. K., Lee, J. Y. et al., Crystal structure of *Escherichia coli* CyaY protein reveals a previously unidentified fold for the evolutionarily conserved frataxin family. *Proc. Natl. Acad. Sci. USA* 2000, **97**, 8932–8937.
- [46] Nair, M., Adinolfi, S., Pastore, C., Kelly, G. et al., Solution structure of the bacterial frataxin ortholog, CyaY: mapping the iron binding sites. *Structure* 2004, **12**, 2037–2048.
- [47] Armengaud, J., Marie Hartmann, E., Bland, C., Proteogenomics for environmental microbiology. *Proteomics* 2013, **13**, 2731–2742.
- [48] Brodie, E. D., 3rd, Convergent evolution: pick your poison carefully. *Curr. Biol.* 2010, **20**, R152–R154.
- [49] Schaeffer, R. D., Daggett, V., Protein folds and protein folding. *Protein Eng. Des. Sel.* 2011, **24**, 11–19.
- [50] Te Velthuis, A. J., Bagowski, C. P., Linking fold, function and phylogeny: a comparative genomics view on protein (domain) evolution. *Curr. Genomics* 2008, **9**, 88–96.
- [51] Boyer, M., Gimenez, G., Suzan-Monti, M., Raoult, D., Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses. *Intervirology* 2010, **53**, 310–320.
- [52] Cortez, D., Forterre, P., Gribaldo, S., A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol.* 2009, **10**, R65.
- [53] Yomtovian, I., Teerakulkittipong, N., Lee, B., Moulf, J., Unger, R., Composition bias and the origin of ORFan genes. *Bioinformatics* 2010, **26**, 996–999.
- [54] Yu, G., Stoltzfus, A., Population diversity of ORFan genes in *Escherichia coli*. *Genome Biol. Evol.* 2012, **4**, 1176–1187.

3.6.2 Conclusions et perspectives

Appliquée à notre problématique, la recherche de MPT, la protéogénomique structurale pourrait nous permettre de sélectionner une protéine portant deux sortes de MPT et en élucidant sa structure connaître l'éventuelle proximité spatiale de sites portant ces MPT.

DISCUSSION

4. DISCUSSION

4.1 – Maitriser la protéomique et l’outil spectrométrie de masse pour aller plus loin....

L’étude des modifications post-traductionnelles connaît depuis plus de dix ans un essor considérable. Bien que connues depuis longtemps leur étude a été aujourd’hui grandement facilitée par le développement des techniques autour des approches protéomiques et du raffinement de l’outil spectrométrie de masse. Maîtriser les approches protéomiques nous paraît être aujourd’hui l’étape incontournable pour aborder ces problématiques. C’est ce qui est illustré dans notre premier travail sur le nouveau regard porté sur l’iodination de la thyroglobuline obtenue à partir d’une glande thyroïde de souris. Ce travail a permis d’obtenir une vue panoramique quasi-complète de l’iodination de la thyroglobuline et de répondre même à une question restée en suspens depuis plus de quinze ans concernant le mécanisme qui se met en place au niveau des tyrosines « donneuses » et qui semble confirmer l’apparition d’acide pyruvique accompagnée d’une rupture de la chaîne polypeptidique. Cette vue panoramique n’était pas possible par les approches de biochimie classique qui étaient mises en œuvre dans les années quatre-vingt et quatre-vingt-dix. Aujourd’hui, l’approche que nous avons mise au point et qui ne peut être que plus complète et rapide à mettre en œuvre avec les spectromètres de masse les plus récents pourrait permettre d’autres ouvertures. D’un point de vue fondamental, la poursuite de ce travail devrait permettre de mieux comprendre la régulation de la fonction thyroïdienne en connaissant mieux comment la thyroïde contrôle l’iodination de la thyroglobuline. Aujourd’hui l’étude de la régulation de cette étape de la synthèse des hormones thyroïdiennes se fait par la mesure du niveau d’expression ou d’activité des enzymes impliquées et notamment la thyropéroxydase (TPO) et les oxydases (Yoshihara, Hara et al. 2012).

Du point de vue des applications cliniques, la cartographie de l’iodination de la thyroglobuline pourrait permettre de caractériser des anomalies de cette iodination et de certaines pathologies. De plus, le diagnostic de métastases thyroïdiennes après prise de la tumeur est basé sur la détection de thyroglobulines circulantes. La connaissance fine de leur degré d’iodination pourrait permettre de mesurer la capacité des cellules à organiser l’iode et donc permettre d’améliorer le pronostic d’efficacité d’un traitement à l’iode 131.

4.2 – Etablir la présence de modifications post-traductionnelles dans divers organismes

Nos travaux portant sur *Deinococcus deserti* et *Thermococcus gammatolerans* avaient pour objectif d'établir la présence de modifications post-traductionnelles dans divers organismes. La question peut se poser de l'intérêt d'en rester à un simple but descriptif et d'établir une liste qui se voudra la plus exhaustive possible d'une ou de plusieurs modifications post-traductionnelles que l'on peut retrouver dans le protéome de ces organismes. Voilà plus de dix ans la simple description du protéome d'un organisme a très vite présenté un intérêt très limité. En effet tous les organismes biologiques vivants sont censés avoir des protéines et en donner la seule liste s'est vite avéré insuffisant.

Pour les modifications post-traductionnelles la situation est quelque peu différente tant il a été longtemps considéré que ces modifications post-traductionnelles étaient l'apanage quasi exclusif des organismes eucaryotes. Chez *Deinococcus deserti* montrer une présence notable d'acétylations N-terminales reste intéressant puisque il s'agit de la première étude conséquente chez une bactérie. De même mettre en évidence un profil particulier concernant l'avant dernière position des protéines ainsi acétylées permet de s'interroger sur la signification de ce profil et de se demander si chez les Thermus qui ont un profil notablement différent ces acétylations sont présentes. Il est d'ailleurs remarquable que ces acétylations bien présentes chez cette bactérie radiorésistante, semblent ne pas être présentes (données non publiées) chez l'archée *Thermococcus gammatolerans* lui aussi radio-résistant. Par contre, les acétylations des résidus des chaînes latérales des lysines clairement acétylées chez cet organisme ne semblent pas l'être au travers d'un test Western Blot effectué avec les mêmes anticorps anti-acétyl lysine (données non publiées). Ces remarques nous montrent que l'étude des MPT chez les procaryotes au moins doit s'effectuer au cas par cas.

Mais rechercher des MPT dans un organisme quel qu'il soit doit se faire en gardant en mémoire que le monde des MPT est en constante évolution. Il ne suffit pas de rechercher les modifications les plus connues. Comme nous l'avions indiqué précédemment, plus de 300 modifications post-traductionnelles ont été décrites à ce jour (<http://dbptm.mbc.nctu.edu.tw/statistics.php>). Mais pour autant la liste est loin d'être close. Au fil des publications de nouvelles MPT sont mises en évidence. Il serait trop long de les énumérer, mais voici quelques exemples :

- Succinylation des lysines (Zhang, Tan et al. 2011) chez *E. coli* strain K-12,
- AMPylation consistant à fixer sur les résidus tyrosine des résidus AMP en provenance d'ATP, sous l'influence d'un facteur de virulence bactérien, cette addition se faisant sur des domaines conservés au cours de l'évolution d'une classe de protéines. Ces domaines sont conservés chez les eucaryotes et les procaryotes (Yarbrough and Orth 2009) et on doit pouvoir légitimement se poser la question de la présence de cette modification chez les procaryotes.

- Dans une revue datant de 2013, Arnaudo et al constataient qu'au cours des cinq dernières années un nombre significatif de nouvelles modifications post-traductionnelles (entre autres, crotonylation, N-formylation, succinylation, malonylation, butyrylation des lysines) avait été observé chez les histones augmentant singulièrement la complexité du code histone (Arnaudo and Garcia 2013).

De façon plus spécifique à l'instar des brominations décrites chez les escargots cônes marins (page 52), des modifications post-traductionnelles plus « particulières » sont toujours décrites, telles les O-mycoloylations, consistant en une acylation de sérines par un acide mycolique (acides gras à longue chaîne alpha-ramifiés et bêta-hydroxylés) provenant de la membrane de bactéries à gram négative, les *Corynebacterineae*. Cette modification affecte des porines. « Les *Corynebactéries* sont des bactéries caractérisées par la présence d'une paroi atypique pour des bactéries Gram+ avec notamment la présence d'une membrane externe appelée mycomembrane », elles appartiennent au phylum des actinobactéries. (Huc E, Thèse, O-mycoloylation : caractérisation d'une nouvelle modification post-traductionnelle de petites protéines d'enveloppe chez les *Corynebacterineae*, 2011, IPBS, Université Paul Sabatier, Toulouse) (Rath, Saurel et al. 2013).

4.3 – Rôle des modifications post-traductionnelles, utiliser le modèle procaryote

Si certaines modifications post-traductionnelles ont un rôle bien établi, pour beaucoup leur rôle est ignoré ou flou. Aussi ne convient-il pas de simplement établir la présence de modifications post-traductionnelles au sein d'un protéome mais aussi de savoir, quelle est la signification biologique de ces modifications.

Les phosphorylations jouent clairement un rôle dans la régulation fine de beaucoup de fonctions biologiques de cellules où ces modifications sont présentes. Le fait que cette MPT soit transitoire, réversible (Système Kinase/Phosphatase) permet d'appréhender son mécanisme d'action. L'arrivée ou le départ d'un groupement phosphate va par sa charge modifier localement la structure de la protéine et activer ou inhiber un site actif de cette protéine, site actif qui peut être celui d'une activité catalytique ou de site de reconnaissance pour des liaisons avec d'autres protéines. Ainsi existe-t-il des cascades de kinases qui par ce jeu de phosphorylations/déphosphorylations vont permettre des régulations fines qui seront perturbées dans certaines pathologies (cancer). Il en va de même des acétylations des fonctions amines des chaînes latérales des lysines qui à l'instar des phosphorylations, sont réversibles (acétylase/déacétylase) et participent à la régulation de nombreuses fonctions biologiques.

Concernant *Thermococcus gammatolerans*, nous avons mis en évidence de nombreuses acétylations sur les chaînes latérales des lysines, particulièrement sur des protéines impliquées dans presque toutes les voies du métabolisme central. Nous pourrions étudier en fonction des

stress extrêmes (radiatifs) que peut subir cet organisme le rôle des régulations particulières qui apparaissent et étudier comment ces acétylations sont éventuellement modifiées et les conséquences sur ces voies métaboliques.

L'autre type d'acétylation dont nous avons parlé dans notre travail, les acétylations N-terminales des protéines, n'a pas une fonction décrite qui fasse consensus au sein de la communauté scientifique. Bien que cette MPT soit extrêmement répandue chez les protéines des eucaryotes (80% des protéines humaines et 50% des protéines de *S. cerevisiae* sont acétylées) et très présente chez les procaryotes, jusqu'en 2010 peu de travail avait été réalisé sur le rôle de cette modification. Longtemps l'acétylation N-terminale des protéines a été considérée comme un élément de protection de leur dégradation (Persson, Flinta et al. 1985) et des données expérimentales indiquaient que les protéines acétylées étaient plus stables *in vivo* que les mêmes protéines non-acétylées. Par ailleurs en 2004, le processus d'ubiquitination des protéines qui consiste à attacher l'ubiquitine sur l'acide aminé N-terminal est lui décrit (Ben-Saadon, Fajerman et al. 2004) comme un vrai signal de dégradation et l'on comprend que les protéines acétylées sur cette position ne peuvent pas être ubiquitinylées. En 2010 l'équipe de Varshavsky propose (Hwang, Shemorry et al. 2010) de prêter au contraire à cette modification le rôle de signal de dégradation pour les protéines concernées. En effet une ubiquitine ligase Doa10 reconnaît les protéines acétylées en N-terminal et les marque avec l'ubiquitine en vue de leur dégradation. L'histoire se complique quand Yi et al (Yi, Pan et al. 2011) suggèrent en 2011 que les acétylations N-terminales peuvent remplir des fonctions biologiques plus nuancées et variées. Travaillant sur des cellules HeLa, ils mettent en évidence un lien entre le niveau d'acétyl CoA et l'acétylation N-terminale. A un faible niveau d'acétyl CoA correspond un faible niveau d'acétylation et vice versa. La protéine Bcl-xL, qui a un rôle fonction anti-apoptotique et est connue par ailleurs pour influencer le métabolisme mitochondrial, fournit selon eux un lien entre le niveau d'acétylation des protéines en N-terminal et l'apoptose. L'augmentation du niveau d'acétyl CoA dans les cellules exprimant Bcl-xL restaure l'acétylation N-terminale des protéines et confère à ces cellules une sensibilité accrue à des stimuli apoptotiques.

On voit clairement que l'histoire n'est pas close. Pour prolonger ce travail de thèse ne serait-il pas pertinent d'étudier le rôle des acétylations N-terminales chez *Deinococcus deserti*, organisme procaryote bien particulier, pour tenter de jeter quelques éléments de plus dans le tableau de connaissances présenté ci-dessus ? On pourrait également s'interroger sur un rôle éventuel de cette modification dans l'extraordinaire capacité de ces organismes à reconstituer leur ADN. N'oublions pas que nos travaux ont montré qu'une ADN gyrase A est surexprimée après irradiation et acétylée et qu'une GCN5-related N-acétyltransferase est également surexprimée.

4.4 – Un nouveau niveau de complexité le « dialogue » des MPT

Au cours des quinze dernières années l'accumulation des connaissances portant sur les MPT a radicalement changé le paysage tant par les nouvelles MPT découvertes que par leur découverte dans des types cellulaires différents.

L'étude des phosphorylations comme il a été souligné a dans le passé particulièrement retenu l'attention des équipes. De plus dans les années 2000 l'acétylation des lysines a connu un succès grandissant ainsi que l'étude des méthylations, sumoylations, ubiquitinations et O et N-glycosylations. Mais un nouvel élément est venu apporter un regard novateur sur les phénomènes biologiques : les MPT « communiquent », « dialoguent », « interagissent », sont dépendantes les unes des autres (notion de « crosstalk » en anglais). Dès 2004 un premier papier de Barak et Eisenbach montrait que sur la protéine CheY d'*Escherichia coli*, phosphorylations des sérines, thréonines et tyrosines sont présentes ainsi que des acétylations sur les chaînes latérales des lysines. Mais, fait plus intéressant, ces MPT sont liés et co-régulés (Barak and Eisenbach 2004). Les histones ont joué dans l'émergence de ce nouveau concept un rôle clef. Car outre les acétylations des lysines qui y ont été identifiées voilà plus de quarante ans, elles sont le siège de nombreuses autres MPT comme les méthylations, les ubiquitinations et les phosphorylations.... et bien d'autres encore (Kouzarides 2007). Dans cette revue parue dans Cell en 2007, Kouzarides indique que plus de 60 résidus différents sur une même histone peuvent porter ces modifications chimiques. Ce chiffre est très certainement sous-estimé pour certaines histones (H3), mais il permet cependant d'avoir un aperçu de la complexité chez certaines protéines de l'impact des modifications post-traductionnelles. Ceci est d'autant plus remarquable que ces modifications ne se retrouvent pas sur la même histone en même temps. Leur présence va dépendre des conditions de « stress » de la cellule et bien entendu peuvent être considérées comme des modifications épigénétiques importantes qui dans les cellules filles après division pourra expliquer certains niveaux de régulation que seul le génome ne peut expliquer. Selon le même auteur on a utilisé de façon certainement un peu abusive le terme de « code histone » pour décrire le rôle de ces modifications pour « activer » les fonctions de l'ADN. Bien qu'utile pour décrire un ensemble de MPT présentes sur une histone pour lui permettre d'accomplir une tâche donnée vis-à-vis de l'ADN, il ne s'agit pas selon Liu et al (Liu, Kaplan et al. 2005) d'un code au sens « prévisible » du terme.

Mais le fait pour une protéine de porter plusieurs modifications ne se réduit pas aux seules histones. Une revue même partielle de la littérature de ces dernières années indique que beaucoup de protéines sont décrites comme portant des MPT différentes qui « dialoguent ». Yang et al dès 2008 dans une revue parue dans Molecular Cell soulignent l'importance des acétylations des chaînes latérales des lysines qui sont en communication avec d'autres MPT présentes sur ces mêmes molécules, des phosphorylations, méthylations, ubiquitinations, sumoylations, celles-ci se produisant sur des protéines non-histones. Ils proposent même de reprendre le concept « discuté » du « code histone » pour expliquer ces régulations

complexes. Zeidan et al en 2010 (Zeidan and Hart 2010) décrivent des O-GlcNacylations communicant avec des phosphorylations dans tous les compartiments de cellules eucaryotes. En 2012, des publications décrivent également ce phénomène de MPT communicantes entre phosphorylations et acétylations chez des bactéries (van Noort, Seebacher et al. 2012) (Soufi, Soares et al. 2012). Enfin en 2013, Nguyen LK et al (Nguyen, Kolch et al. 2013) décrivent les interactions existant entre les ubiquitinations et les phosphorylations entre les protéines EGFR (Epidermal Growth Factor Receptor) et MAPK (Mitogen Activated Protein Kinase), jouant un rôle capital dans la régulation cellulaire.

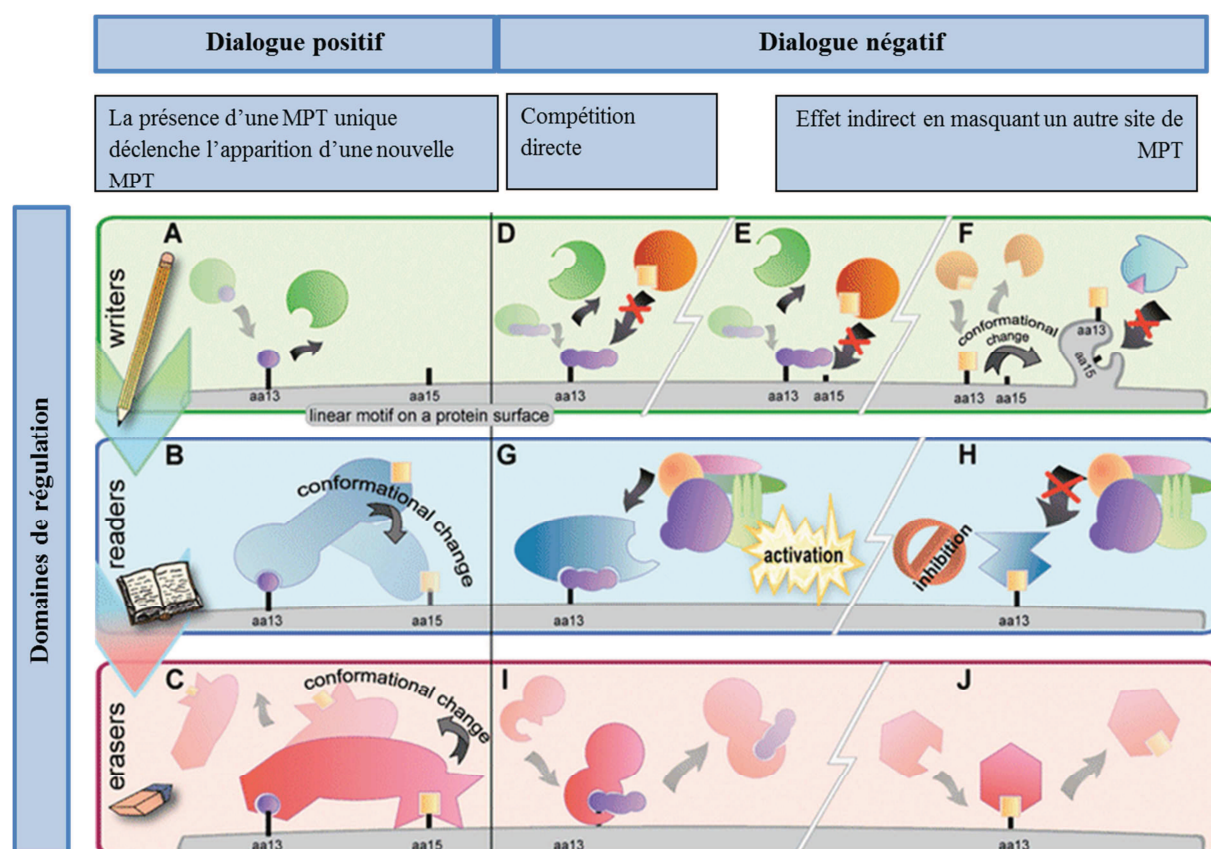


Figure 57 : Classification des dialogues des MPT (d'après A.S. Venne et al, 2014).

Les auteurs appellent « Domaine de régulation » les courts motifs linéaires à la surface des protéines permettant une interaction protéine-protéine. Selon leur fonction, ils sont classés comme « écrivains » (writers), lecteurs (readers) et « effaceurs » (erasers) (rectangles vert, bleu, rouges). On parlera pour décrire les interactions entre MPT de « Dialogue positif » ou de « dialogue négatif ».

Dialogue positif :

(A) Une MPT se fixe sur un acide aminé (aa) d'une protéine cible. (B) Cette MPT peut attirer un autre ensemble protéique qui va déclencher l'addition d'une seconde MPT, par exemple, ou induire par un changement conformationnel qui va faciliter la fixation d'une autre MPT. (C) et au contraire, une MPT peut être détectée et éliminée par une protéine « effaceuse ».

Dialogue négatif peut être subdivisé en compétition directe ou indirecte.

(D) Deux MPT différentes rentrent en concurrence pour se fixer au même résidu. (E) deux MPT possédant des sites de liaison différents mais proches, de telle sorte que lors de la fixation initiale de la première MPT, le deuxième site de liaison est indirectement masqué. (F) La première MPT conduit à un changement de conformation, qui cache le deuxième site de fixation de la seconde MPT. (G) Une MPT est fixée et recrute un complexe protéique qui par son activation va empêcher la fixation d'une autre MPT. (H) Au contraire une MPT présente peut inhiber une voie d'activation d'une autre MPT. (I, J) Une MPT déjà présente peut être éliminée par un complexe protéique « effaceur ».

Ce qui fait en 2014 écrire à Venne et al dans *Proteomics* : « The next level of complexity ; Crosstalk of posttranslational modifications » (Venne, Kollipara et al. 2014). Dans cet article, un panorama des mécanismes possibles par lesquels les modifications post-traductionnelles peuvent communiquer, est décrit (Figure 57). Aussi nous pourrions dans une suite possible de notre travail, étudier plus en détail ces « dialogues » qui doivent exister chez notre archée, *Thermococcus gammatolerans*, puisque nous avons mis en évidence sur les mêmes protéines phosphorylations et acétylations et peut-être mieux comprendre le rôle de ces modifications chez ce procaryote.

Enfin dans une perspective beaucoup plus lointaine une connaissance plus approfondie des MPT chez les procaryotes ne permettrait-elle pas d'apporter quelques éléments pour alimenter le débat scientifique relatif à la phylogénie des Eubacteria, la place de la racine de l'arbre universel ?

REFERENCES BIBLIOGRAPHIQUE DE L'INTRODUCTION

A

- Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." *Nature* **422**(6928): 198-207.
- Afjehi-Sadat, L. and B. A. Garcia (2013). "Comprehending dynamic protein methylation with mass spectrometry." *Curr Opin Chem Biol* **17**(1): 12-19.
- Aguilar, M. B., E. Lopez-Vera, et al. (2005). "A novel conotoxin from *Conus delessertii* with posttranslationally modified lysine residues." *Biochemistry* **44**(33): 11130-11136.
- Allfrey, V. G., R. Faulkner, et al. (1964). "Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis." *Proc Natl Acad Sci U S A* **51**: 786-794.
- Ambler, R. P. and M. W. Rees (1959). "Epsilon-N-Methyl-lysine in bacterial flagellar protein." *Nature* **184**: 56-57.
- Armengaud, J., J. A. Christie-Oleza, et al. (2012). "Exoproteomics: exploring the world around biological systems." *Expert Rev Proteomics* **9**(5): 561-575.
- Armengaud, J., E. M. Hartmann, et al. (2013). "Proteogenomics for environmental microbiology." *Proteomics* **13**(18-19): 2731-2742.
- Arnaudo, A. M. and B. A. Garcia (2013). "Proteomic characterization of novel histone post-translational modifications." *Epigenetics Chromatin* **6**(1): 24.
- Barak, R. and M. Eisenbach (2004). "Co-regulation of acetylation and phosphorylation of CheY, a response regulator in chemotaxis of *Escherichia coli*." *J Mol Biol* **342**(2): 375-381.
- Bartova, E., J. Krejci, et al. (2008). "Histone modifications and nuclear architecture: a review." *J Histochem Cytochem* **56**(8): 711-721.
- Basu, S. and D. Plewczynski (2010). "AMS 3.0: prediction of post-translational modifications." *BMC Bioinformatics* **11**: 210.
- Baudet, M., P. Ortet, et al. (2010). "Proteomics-based refinement of *Deinococcus deserti* genome annotation reveals an unwonted use of non-canonical translation initiation codons." *Mol Cell Proteomics* **9**(2): 415-426.
- Beausoleil, S. A., M. Jedrychowski, et al. (2004). "Large-scale characterization of HeLa cell nuclear phosphoproteins." *Proc Natl Acad Sci U S A* **101**(33): 12130-12135.
- Ben-Saadon, R., I. Fajerman, et al. (2004). "The tumor suppressor protein p16(INK4a) and the human papillomavirus oncoprotein-58 E7 are naturally occurring lysine-less proteins that are degraded by the ubiquitin system. Direct evidence for ubiquitination at the N-terminal residue." *J Biol Chem* **279**(40): 41414-41421.
- Benz, I. and M. A. Schmidt (2002). "Never say never again: protein glycosylation in pathogenic bacteria." *Mol Microbiol* **45**(2): 267-276.
- Biemann, K. (1988). "Contributions of mass spectrometry to peptide and protein structure." *Biomed Environ Mass Spectrom* **16**(1-12): 99-111.
- Bland, C., L. Bellanger, et al. (2014). "Magnetic immunoaffinity enrichment for selective capture and MS/MS analysis of N-terminal-TMPP-labeled peptides." *J Proteome Res* **13**(2): 668-680.
- Bland, C., E. M. Hartmann, et al. (2014). "N-Terminal-oriented proteogenomics of the marine bacterium *roseobacter denitrificans* Och114 using N-Succinimidylloxycarbonylmethyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP) labeling and diagonal chromatography." *Mol Cell Proteomics* **13**(5): 1369-1381.
- Blom, N., T. Sicheritz-Ponten, et al. (2004). "Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence." *Proteomics* **4**(6): 1633-1649.
- Boersema, P. J., S. Mohammed, et al. (2009). "Phosphopeptide fragmentation and analysis by mass spectrometry." *J Mass Spectrom* **44**(6): 861-878.

- Bondarenko, P. V., D. Chelius, et al. (2002). "Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry." *Anal Chem* **74**(18): 4741-4749.
- Bonissone, S., N. Gupta, et al. (2013). "N-terminal protein processing: a comparative proteogenomic analysis." *Mol Cell Proteomics* **12**(1): 14-28.
- Boys, B. L., M. C. Kuprowski, et al. (2009). "Protein oxidative modifications during electrospray ionization: solution phase electrochemistry or corona discharge-induced radical attack?" *Anal Chem* **81**(10): 4027-4034.
- Bradshaw, R. A., W. W. Brickey, et al. (1998). "N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families." *Trends in Biochemical Sciences* **23**(7): 263-267.
- Britton, L. M., M. Gonzales-Cope, et al. (2011). "Breaking the histone code with quantitative mass spectrometry." *Expert Rev Proteomics* **8**(5): 631-643.
- Burnett, G. and E. P. Kennedy (1954). "The enzymatic phosphorylation of proteins." *J Biol Chem* **211**(2): 969-980.
- Carr, S. A., M. J. Huddleston, et al. (1996). "Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry." *Anal Biochem* **239**(2): 180-192.
- Castro, P. H., R. M. Tavares, et al. (2012). "SUMO, a heavyweight player in plant abiotic stress responses." *Cell Mol Life Sci* **69**(19): 3269-3283.
- Chen, Y., J. Zhang, et al. (2009). "Mascot-derived false positive peptide identifications revealed by manual analysis of tandem mass spectra." *J Proteome Res* **8**(6): 3141-3147.
- Choudhary, C., C. Kumar, et al. (2009). "Lysine acetylation targets protein complexes and co-regulates major cellular functions." *Science* **325**(5942): 834-840.
- Christie-Oleza, J. A., G. Miotello, et al. (2012). "High-throughput proteogenomics of *Ruegeria pomeroyi*: seeding a better genomic annotation for the whole marine Roseobacter clade." *BMC Genomics* **13**: 73.
- Christie-Oleza, J. A., J. M. Pina-Villalonga, et al. (2012). "Comparative proteogenomics of twelve *Roseobacter* exoproteomes reveals different adaptive strategies among these marine bacteria." *Mol Cell Proteomics* **11**(2): M111 013110.
- Ciechanover, A. (1994). "The ubiquitin-proteasome proteolytic pathway." *Cell* **79**(1): 13-21.
- Ciucanu, I. and C. E. Costello (2003). "Elimination of oxidative degradation during the per-O-methylation of carbohydrates." *J Am Chem Soc* **125**(52): 16213-16219.
- Clair, G., S. Roussi, et al. (2010). "Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions." *Mol Cell Proteomics* **9**(7): 1486-1498.
- Cohen, P. (2000). "The regulation of protein function by multisite phosphorylation--a 25 year update." *Trends in Biochemical Sciences* **25**(12): 596-601.
- Colland, F., X. Jacq, et al. (2004). "Functional proteomics mapping of a human signaling pathway." *Genome Res* **14**(7): 1324-1332.
- Covey, T. R., E. C. Huang, et al. (1991). "Structural characterization of protein tryptic peptides via liquid chromatography/mass spectrometry and collision-induced dissociation of their doubly charged molecular ions." *Anal Chem* **63**(13): 1193-1200.
- Cozzone, A. J. (1984). "Protein-Phosphorylation in Bacteria." *Trends in Biochemical Sciences* **9**(9): 400-403.
- Cozzone, A. J. (1998). "Post-translational modification of proteins by reversible phosphorylation in prokaryotes." *Biochimie* **80**(1): 43-48.
- Cozzone, A. J., C. Grangeasse, et al. (2004). "Protein phosphorylation on tyrosine in bacteria." *Arch Microbiol* **181**(3): 171-181.
- Craig, A. G., E. C. Jimenez, et al. (1997). "A novel post-translational modification involving bromination of tryptophan. Identification of the residue, L-6-bromotryptophan, in peptides from *Conus imperialis* and *Conus radiatus* venom." *J Biol Chem* **272**(8): 4689-4698.

- Craig, R. and R. C. Beavis (2004). "TANDEM: matching proteins with tandem mass spectra." *Bioinformatics* **20**(9): 1466-1467.
- Dalpathado, D. S. and H. Desaire (2008). "Glycopeptide analysis by mass spectrometry." *Analyst* **133**(6): 731-738.
- de Groot, A., V. Chapon, et al. (2005). "Deinococcus deserti sp. nov., a gamma-radiation-tolerant bacterium isolated from the Sahara Desert." *Int J Syst Evol Microbiol* **55**(Pt 6): 2441-2446.
- de Groot, A., R. Dulermo, et al. (2009). "Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*." *PLoS Genet* **5**(3): e1000434.
- Dedieu, A., J. C. Gaillard, et al. (2011). "Revisiting iodination sites in thyroglobulin with an organ-oriented shotgun strategy." *J Biol Chem* **286**(1): 259-269.
- Delahunty, C. and J. R. Yates, 3rd (2005). "Protein identification using 2D-LC-MS/MS." *Methods* **35**(3): 248-255.
- Delom, F. and E. Chevet (2006). "Phosphoprotein analysis: from proteins to proteomes." *Proteome Sci* **4**: 15.
- Domon, B. and C. E. Costello (1988). "Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry." *Biochemistry* **27**(5): 1534-1543.
- Driessen, H. P., W. W. de Jong, et al. (1985). "The mechanism of N-terminal acetylation of proteins." *CRC Crit Rev Biochem* **18**(4): 281-325.
- Dunn, J. T. and A. D. Dunn (1999). "The importance of thyroglobulin structure for thyroid hormone biosynthesis." *Biochimie* **81**(5): 505-509.
- Dupierris, V., C. Masselon, et al. (2009). "A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa." *Bioinformatics* **25**(15): 1980-1981.
- Dwek, R. A. (1996). "Glycobiology: Toward Understanding the Function of Sugars." *Chem Rev* **96**(2): 683-720.
- Eichler, J. and M. W. Adams (2005). "Posttranslational protein modification in Archaea." *Microbiol Mol Biol Rev* **69**(3): 393-425.
- Eng, J. K., A. L. McCormack, et al. (1994). "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database." *J Am Soc Mass Spectrom* **5**(11): 976-989.
- Falb, M., M. Aivaliotis, et al. (2006). "Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey." *J Mol Biol* **362**(5): 915-924.
- Ferreira, J. A., A. L. Daniel-da-Silva, et al. (2011). "Synthesis and optimization of lectin functionalized nanoprobe for the selective recovery of glycoproteins from human body fluids." *Anal Chem* **83**(18): 7035-7043.
- Finkel, T., C. X. Deng, et al. (2009). "Recent progress in the biology and physiology of sirtuins." *Nature* **460**(7255): 587-591.
- Forbes, A. J., M. T. Mazur, et al. (2001). "Toward efficient analysis of >70 kDa proteins with 100% sequence coverage." *Proteomics* **1**(8): 927-933.
- Forbes, A. J., S. M. Patrie, et al. (2004). "Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS." *Proc Natl Acad Sci U S A* **101**(9): 2678-2683.
- Frese, C. K., A. F. Altelaar, et al. (2011). "Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos." *J Proteome Res* **10**(5): 2377-2388.
- Gabant, G., A. Lorphelin, et al. (2008). "Autophosphorylated residues involved in the regulation of human chk2 in vitro." *J Mol Biol* **380**(3): 489-503.
- Gan, C. S., P. K. Chong, et al. (2007). "Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ)." *J Proteome Res* **6**(2): 821-827.
- Gault, J., C. Malosse, et al. (2014). "Complete posttranslational modification mapping of pathogenic *Neisseria meningitidis* pilins requires top-down mass spectrometry." *Proteomics* **14**(10): 1141-1151.
- Ge, Y., B. G. Lawhorn, et al. (2002). "Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry." *J Am Chem Soc* **124**(4): 672-678.

- Geiss-Friedlander, R. and F. Melchior (2007). "Concepts in sumoylation: a decade on." Nat Rev Mol Cell Biol **8**(12): 947-956.
- Gevaert, K., P. Van Damme, et al. (2007). "A la carte proteomics with an emphasis on gel-free techniques." Proteomics **7**(16): 2698-2718.
- Geyer, H. and R. Geyer (2006). "Strategies for analysis of glycoprotein glycosylation." Biochim Biophys Acta **1764**(12): 1853-1869.
- Gillet, L. C., P. Navarro, et al. (2012). "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 **11**(6): O111 016717.
- Gorg, A., W. Weiss, et al. (2004). "Current two-dimensional electrophoresis technology for proteomics." Proteomics **4**(12): 3665-3685.
- Grangeasse, C., A. J. Cozzone, et al. (2007). "Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology." Trends in Biochemical Sciences **32**(2): 86-94.
- Gray, M. W., G. Burger, et al. (1999). "Mitochondrial evolution." Science **283**(5407): 1476-1481.
- Greer, E. L. and Y. Shi (2012). "Histone methylation: a dynamic mark in health, disease and inheritance." Nat Rev Genet **13**(5): 343-357.
- Gronborg, M., T. Z. Kristiansen, et al. (2002). "A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate." Mol Cell Proteomics **1**(7): 517-527.
- Guan, K. L., W. Yu, et al. (2010). "Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides." Nat Protoc **5**(9): 1583-1595.
- Guthals, A., K. R. Clauser, et al. (2013). "Sequencing-grade de novo analysis of MS/MS triplets (CID/HCD/ETD) from overlapping peptides." J Proteome Res **12**(6): 2846-2857.
- Haebel, S., C. Jensen, et al. (1995). "Isoforms of a cuticular protein from larvae of the meal beetle, *Tenebrio molitor*, studied by mass spectrometry in combination with Edman degradation and two-dimensional polyacrylamide gel electrophoresis." Protein Sci **4**(3): 394-404.
- Hartl, F. U. (1996). "Molecular chaperones in cellular protein folding." Nature **381**(6583): 571-579.
- Hartmann, E. M., F. Allain, et al. (2014). "Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria." Methods Mol Biol **1197**: 275-285.
- Hartmann, E. M. and J. Armengaud (2014). "N-terminomics and proteogenomics, getting off to a good start." Proteomics.
- Hicke, L., H. L. Schubert, et al. (2005). "Ubiquitin-binding domains." Nat Rev Mol Cell Biol **6**(8): 610-621.
- Hirel, P. H., M. J. Schmitter, et al. (1989). "Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid." Proc Natl Acad Sci U S A **86**(21): 8247-8251.
- Hjerrild, M., A. Stensballe, et al. (2004). "Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry." J Proteome Res **3**(3): 426-433.
- Hunter, A. P. and D. E. Games (1994). "Chromatographic and mass spectrometric methods for the identification of phosphorylation sites in phosphoproteins." Rapid Commun Mass Spectrom **8**(7): 559-570.
- Hwang, C. S., A. Shemorry, et al. (2010). "N-terminal acetylation of cellular proteins creates specific degradation signals." Science **327**(5968): 973-977.
- Ingaramo, M. and D. Beckett (2011). "Biotinylation, a post-translational modification controlled by the rate of protein-protein association." J Biol Chem **286**(15): 13071-13078.
- Jeffers, V. and W. J. Sullivan, Jr. (2012). "Lysine acetylation is widespread on proteins of diverse function and localization in the protozoan parasite *Toxoplasma gondii*." Eukaryot Cell **11**(6): 735-742.

- Johnson, R. S., S. A. Martin, et al. (1987). "Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine." *Anal Chem* **59**(21): 2621-2625.
- Jolivet, E., E. Corre, et al. (2004). "Thermococcus marinus sp. nov. and Thermococcus radiotolerans sp. nov., two hyperthermophilic archaea from deep-sea hydrothermal vents that resist ionizing radiation." *Extremophiles* **8**(3): 219-227.
- Jolivet, E., S. L'Haridon, et al. (2003). "Thermococcus gammatolerans sp. nov., a hyperthermophilic archaeon from a deep-sea hydrothermal vent that resists ionizing radiation." *Int J Syst Evol Microbiol* **53**(Pt 3): 847-851.
- Jones, J. D. and C. D. O'Connor (2011). "Protein acetylation in prokaryotes." *Proteomics* **11**(15): 3012-3022.
- Kanai, T., H. Imanaka, et al. (2005). "Continuous hydrogen production by the hyperthermophilic archaeon, Thermococcus kodakaraensis KOD1." *J Biotechnol* **116**(3): 271-282.
- Kapp, E. A., F. Schutz, et al. (2005). "An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis." *Proteomics* **5**(13): 3475-3490.
- Kennelly, P. J. and M. Potts (1996). "Fancy meeting you here! A fresh look at "prokaryotic" protein phosphorylation." *J Bacteriol* **178**(16): 4759-4764.
- Kerscher, O., R. Felberbaum, et al. (2006). "Modification of proteins by ubiquitin and ubiquitin-like proteins." *Annu Rev Cell Dev Biol* **22**: 159-180.
- Khoury, G. A., R. C. Baliban, et al. (2011). "Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database." *Sci Rep* **1**.
- Kim, D., B. J. Yu, et al. (2013). "The acetylproteome of Gram-positive model bacterium Bacillus subtilis." *Proteomics* **13**(10-11): 1726-1736.
- Kim, H. J., S. H. Kim, et al. (2007). "Drosophila homolog of APP-BP1 (dAPP-BP1) interacts antagonistically with APPL during Drosophila development." *Cell Death Differ* **14**(1): 103-115.
- Kim, J. H., J. Lee, et al. (2004). "Prediction of phosphorylation sites using SVMs." *Bioinformatics* **20**(17): 3179-3184.
- Kim, M. S. and A. Pandey (2012). "Electron transfer dissociation mass spectrometry in proteomics." *Proteomics* **12**(4-5): 530-542.
- Klarskov, K., D. Roecklin, et al. (1994). "Analysis of recombinant Schistosoma mansoni antigen rSmp28 by on-line liquid chromatography-mass spectrometry combined with sodium dodecyl sulfate polyacrylamide gel electrophoresis." *Anal Biochem* **216**(1): 127-134.
- Klose, J. (1999). "Large-gel 2-D electrophoresis." *Methods Mol Biol* **112**: 147-172.
- Kouzarides, T. (2007). "Chromatin modifications and their function." *Cell* **128**(4): 693-705.
- Kozak, R. P., L. Royle, et al. (2012). "Suppression of peeling during the release of O-glycans by hydrazinolysis." *Anal Biochem* **423**(1): 119-128.
- Kuroishi, T., L. Rios-Avila, et al. (2011). "Biotinylation is a natural, albeit rare, modification of human histones." *Molecular Genetics and Metabolism* **104**(4): 537-545.
- Lapko, V. N., D. L. Smith, et al. (2000). "Identification of an artifact in the mass spectrometry of proteins derivatized with iodoacetamide." *J Mass Spectrom* **35**(4): 572-575.
- Larsen, M. R., T. E. Thingholm, et al. (2005). "Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns." *Mol Cell Proteomics* **4**(7): 873-886.
- Lee, D. W., D. Kim, et al. (2013). "Proteomic analysis of acetylation in thermophilic Geobacillus kaustophilus." *Proteomics* **13**(15): 2278-2282.
- Lee, S., M. Tan, et al. (2013). "MS/MS of synthetic peptide is not sufficient to confirm new types of protein modifications." *J Proteome Res* **12**(2): 1007-1013.
- Lee, T. Y., H. D. Huang, et al. (2006). "dbPTM: an information repository of protein post-translational modification." *Nucleic Acids Res* **34**(Database issue): D622-627.
- Levene, P. A. and W. A. Beatty (1906). "On the Phosphotungstates of Certain Aminoacids." *J Exp Med* **8**(3): 463-466.

- Levine, A., F. Vannier, et al. (2006). "Analysis of the dynamic *Bacillus subtilis* Ser/Thr/Tyr phosphoproteome implicated in a wide variety of cellular processes." *Proteomics* **6**(7): 2157-2173.
- Lin, Y. Y., Y. Qi, et al. (2008). "A comprehensive synthetic genetic interaction network governing yeast histone acetylation and deacetylation." *Genes Dev* **22**(15): 2062-2074.
- Liu, C. L., T. Kaplan, et al. (2005). "Single-nucleosome mapping of histone modifications in *S. cerevisiae*." *PLoS Biol* **3**(10): e328.
- Liu, H., R. G. Sadygov, et al. (2004). "A model for random sampling and estimation of relative protein abundance in shotgun proteomics." *Anal Chem* **76**(14): 4193-4201.
- Liu, L. Q. a. M. (2013). "CID, ETD and HCD Fragmentation to Study Protein Post-Translational Modifications." *Modern Chemistry and Applications* **1**(1).
- Loo, D., A. Jones, et al. (2010). "Lectin magnetic bead array for biomarker discovery." *J Proteome Res* **9**(10): 5496-5500.
- Madera, M., B. Mann, et al. (2008). "Efficacy of glycoprotein enrichment by microscale lectin affinity chromatography." *J Sep Sci* **31**(14): 2722-2732.
- Makarov, A., E. Denisov, et al. (2009). "Performance evaluation of a high-field Orbitrap mass analyzer." *J Am Soc Mass Spectrom* **20**(8): 1391-1396.
- Makarova, K. S., L. Aravind, et al. (2001). "Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics." *Microbiol Mol Biol Rev* **65**(1): 44-79.
- Mann, M. and O. N. Jensen (2003). "Proteomic analysis of post-translational modifications." *Nat Biotechnol* **21**(3): 255-261.
- Mann, M., S. E. Ong, et al. (2002). "Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome." *Trends Biotechnol* **20**(6): 261-268.
- Martin, J. and F. U. Hartl (1997). "Chaperone-assisted protein folding." *Curr Opin Struct Biol* **7**(1): 41-52.
- Matlack, K. E., W. Mothes, et al. (1998). "Protein translocation: tunnel vision." *Cell* **92**(3): 381-390.
- Matthews, H. R. (1995). "Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: a possible regulator of the mitogen-activated protein kinase cascade." *Pharmacol Ther* **67**(3): 323-350.
- McNulty, D. E. and R. S. Annan (2008). "Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection." *Mol Cell Proteomics* **7**(5): 971-980.
- Medzihradzsky, K. F., Z. Darula, et al. (2004). "O-sulfonation of serine and threonine: mass spectrometric detection and characterization of a new posttranslational modification in diverse proteins throughout the eukaryotes." *Mol Cell Proteomics* **3**(5): 429-440.
- Meinzel, T. and C. Giglione (2008). "Tools for analyzing and predicting N-terminal protein modifications." *Proteomics* **8**(4): 626-649.
- Michel, P. E., F. Reymond, et al. (2003). "Protein fractionation in a multicompartiment device using Off-Gel isoelectric focusing." *Electrophoresis* **24**(1-2): 3-11.
- Miteva, M., K. Keusekotten, et al. (2010). "Sumoylation as a signal for polyubiquitylation and proteasomal degradation." *Subcell Biochem* **54**: 195-214.
- Moens, S. and J. Vanderleyden (1997). "Glycoproteins in prokaryotes." *Arch Microbiol* **168**(3): 169-175.
- Mohammed, S. and A. Heck, Jr. (2011). "Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications." *Curr Opin Biotechnol* **22**(1): 9-16.
- Monteoliva, L. and J. P. Albar (2004). "Differential proteomics: an overview of gel and non-gel based approaches." *Brief Funct Genomic Proteomic* **3**(3): 220-239.
- Morandell, S., T. Stasyk, et al. (2006). "Phosphoproteomics strategies for the functional analysis of signal transduction." *Proteomics* **6**(14): 4047-4056.

- Morelle, W., K. Canis, et al. (2006). "The use of mass spectrometry for the proteomic analysis of glycosylation." *Proteomics* **6**(14): 3993-4015.
- Mueller, L. N., M. Y. Brusniak, et al. (2008). "An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data." *J Proteome Res* **7**(1): 51-61.
- Mukherjee, S., G. Keitany, et al. (2006). "Yersinia YopJ acetylates and inhibits kinase activation by blocking phosphorylation." *Science* **312**(5777): 1211-1214.
- Murray, K. (1964). "The Occurrence of Epsilon-N-Methyl Lysine in Histones." *Biochemistry* **3**: 10-15.
- Myllyharju, J. (2005). "Intracellular post-translational modifications of collagens." *Collagen* **247**: 115-147.
- Nahnsen, S., C. Bielow, et al. (2013). "Tools for label-free peptide quantification." *Mol Cell Proteomics* **12**(3): 549-556.
- Nawarak, J., S. Phutrakul, et al. (2004). "Analysis of lectin-bound glycoproteins in snake venom from the Elapidae and Viperidae families." *J Proteome Res* **3**(3): 383-392.
- Nguyen, L. K., W. Kolch, et al. (2013). "When ubiquitination meets phosphorylation: a systems biology perspective of EGFR/MAPK signalling." *Cell Commun Signal* **11**: 52.
- O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." *J Biol Chem* **250**(10): 4007-4021.
- Okanishi, H., K. Kim, et al. (2013). "Acetylome with structural mapping reveals the significance of lysine acetylation in *Thermus thermophilus*." *J Proteome Res* **12**(9): 3952-3968.
- Olsen, J. V., B. Macek, et al. (2007). "Higher-energy C-trap dissociation for peptide modification analysis." *Nat Methods* **4**(9): 709-712.
- Omelchenko, M. V., Y. I. Wolf, et al. (2005). "Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance." *BMC Evol Biol* **5**: 57.
- Ong, S. E., I. Kratchmarova, et al. (2003). "Properties of ¹³C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC)." *J Proteome Res* **2**(2): 173-181.
- Paik, W. K., D. C. Paik, et al. (2007). "Historical review: the field of protein methylation." *Trends in Biochemical Sciences* **32**(3): 146-152.
- Pannifer, A. D., A. J. Flint, et al. (1998). "Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by x-ray crystallography." *J Biol Chem* **273**(17): 10454-10462.
- Patel, T., J. Bruce, et al. (1993). "Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins." *Biochemistry* **32**(2): 679-693.
- Patterson, S. D. (2003). "Proteomics: evolution of the technology." *Biotechniques* **35**(3): 440-444.
- Patterson, S. D. and R. H. Aebersold (2003). "Proteomics: the first decade and beyond." *Nat Genet* **33** **Suppl**: 311-323.
- Pawson, T. and J. D. Scott (2005). "Protein phosphorylation in signaling--50 years and counting." *Trends in Biochemical Sciences* **30**(6): 286-290.
- Perdivara, I., L. J. Deterding, et al. (2010). "Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: chemical artifact or post-translational modification?" *J Am Soc Mass Spectrom* **21**(7): 1114-1117.
- Perkins, D. N., D. J. Pappin, et al. (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." *Electrophoresis* **20**(18): 3551-3567.
- Persson, B., C. Flinta, et al. (1985). "Structures of N-terminally acetylated proteins." *Eur J Biochem* **152**(3): 523-527.
- Peterson, A. C., J. D. Russell, et al. (2012). "Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics." *Mol Cell Proteomics* **11**(11): 1475-1488.
- Peyrol, J., C. Thizon, et al. (2013). "Multiple phosphorylable sites in the Zaire Ebolavirus nucleoprotein evidenced by high resolution tandem mass spectrometry." *J Virol Methods* **187**(1): 159-165.
- Phillips, D. M. (1963). "The presence of acetyl groups of histones." *Biochem J* **87**: 258-263.
- Pidoux, G. and K. Tasken (2010). "Specificity and spatial dynamics of protein kinase A signaling organized by A-kinase-anchoring proteins." *J Mol Endocrinol* **44**(5): 271-284.

- Pinkse, M. W., P. M. Uitto, et al. (2004). "Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns." *Anal Chem* **76**(14): 3935-3943.
- Polevoda, B. and F. Sherman (2000). "N-alpha -terminal acetylation of eukaryotic proteins." *J Biol Chem* **275**(47): 36479-36482.
- Polevoda, B. and F. Sherman (2002). "The diversity of acetylated proteins." *Genome Biol* **3**(5): reviews0006.
- Polevoda, B. and F. Sherman (2003). "Composition and function of the eukaryotic N-terminal acetyltransferase subunits." *Biochem Biophys Res Commun* **308**(1): 1-11.
- Polevoda, B. and F. Sherman (2003). "N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins." *J Mol Biol* **325**(4): 595-622.
- Rabilloud, T., M. Chevallet, et al. (2010). "Two-dimensional gel electrophoresis in proteomics: Past, present and future." *J Proteomics* **73**(11): 2064-2077.
- Rath, P., O. Saurel, et al. (2013). "NMR localization of the O-mycoloylation on PorH, a channel forming peptide from *Corynebacterium glutamicum*." *FEBS Lett* **587**(22): 3687-3691.
- Rigby, A. C., E. Lucas-Meunier, et al. (1999). "A conotoxin from *Conus textile* with unusual posttranslational modifications reduces presynaptic Ca²⁺ influx." *Proc Natl Acad Sci U S A* **96**(10): 5758-5763.
- Robbe, C., C. Capon, et al. (2004). "Diagnostic ions for the rapid analysis by nano-electrospray ionization quadrupole time-of-flight mass spectrometry of O-glycans from human mucins." *Rapid Commun Mass Spectrom* **18**(4): 412-420.
- Roepstorff, P. and J. Fohlman (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." *Biomed Mass Spectrom* **11**(11): 601.
- Ros, A., M. Faupel, et al. (2002). "Protein purification by Off-Gel electrophoresis." *Proteomics* **2**(2): 151-156.
- Rudd, P. M., T. Elliott, et al. (2001). "Glycosylation and the immune system." *Science* **291**(5512): 2370-2376.
- Santoni, V., M. Molloy, et al. (2000). "Membrane proteins and proteomics: un amour impossible?" *Electrophoresis* **21**(6): 1054-1070.
- Sarioglu, H., F. Lottspeich, et al. (2000). "Deamidation as a widespread phenomenon in two-dimensional polyacrylamide gel electrophoresis of human blood plasma proteins." *Electrophoresis* **21**(11): 2209-2218.
- Sato, T., T. Fukui, et al. (2003). "Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1." *J Bacteriol* **185**(1): 210-220.
- Schilling, B., M. J. Rardin, et al. (2012). "Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation." *Mol Cell Proteomics* **11**(5): 202-214.
- Schroeder, M. J., J. Shabanowitz, et al. (2004). "A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry." *Anal Chem* **76**(13): 3590-3598.
- Shaw, P. G., R. Chaerkady, et al. (2011). "Monoclonal antibody cocktail as an enrichment tool for acetylome analysis." *Anal Chem* **83**(10): 3623-3626.
- Soppa, J. (2010). "Protein acetylation in archaea, bacteria, and eukaryotes." *Archaea* **2010**.
- Soufi, B., N. C. Soares, et al. (2012). "Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation." *Curr Opin Microbiol* **15**(3): 357-363.
- Spange, S., T. Wagner, et al. (2009). "Acetylation of non-histone proteins modulates cellular signalling at multiple levels." *Int J Biochem Cell Biol* **41**(1): 185-198.
- Sprung, R., Y. Chen, et al. (2008). "Identification and validation of eukaryotic aspartate and glutamate methylation in proteins." *J Proteome Res* **7**(3): 1001-1006.

- Stadtman, E. R. and R. L. Levine (2003). "Free radical-mediated oxidation of free amino acids and amino acid residues in proteins." *Amino Acids* **25**(3-4): 207-218.
- Stapels, M. D. and D. F. Barofsky (2004). "Complementary use of MALDI and ESI for the HPLC-MS/MS analysis of DNA-binding proteins." *Anal Chem* **76**(18): 5423-5430.
- Steen, H., B. Kuster, et al. (2001). "Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode." *Anal Chem* **73**(7): 1440-1448.
- Steinberg, T. H., B. J. Agnew, et al. (2003). "Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology." *Proteomics* **3**(7): 1128-1144.
- Steinberg, T. H., K. Pretty On Top, et al. (2001). "Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots." *Proteomics* **1**(7): 841-855.
- Syka, J. E., J. J. Coon, et al. (2004). "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry." *Proc Natl Acad Sci U S A* **101**(26): 9528-9533.
- Tan, F., K. Zhang, et al. (2011). "Differential histone modification and protein expression associated with cell wall removal and regeneration in rice (*Oryza sativa*)." *J Proteome Res* **10**(2): 551-563.
- Tang, J., Y. Liu, et al. (2010). "Concanavalin A-immobilized magnetic nanoparticles for selective enrichment of glycoproteins and application to glycoproteomics in hepatocellular carcinoma cell line." *Proteomics* **10**(10): 2000-2014.
- Tapias, A., C. Leplat, et al. (2009). "Recovery of ionizing-radiation damage after high doses of gamma ray in the hyperthermophilic archaeon *Thermococcus gammatolerans*." *Extremophiles* **13**(2): 333-343.
- Taugog, A., M. L. Dorris, et al. (1996). "Mechanism of simultaneous iodination and coupling catalyzed by thyroid peroxidase." *Arch Biochem Biophys* **330**(1): 24-32.
- Terzic, J. and I. Dikic (2011). "Interplay between ubiquitin networks and NF- κ B signaling." *Periodicum Biologorum* **113**(1): 1-6.
- Thakur, S. S., T. Geiger, et al. (2011). "Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation." *Mol Cell Proteomics* **10**(8): M110 003699.
- Thao, S. and J. C. Escalante-Semerena (2011). "Control of protein function by reversible Nvarepsilon-lysine acetylation in bacteria." *Curr Opin Microbiol* **14**(2): 200-204.
- Thingholm, T. E., M. R. Larsen, et al. (2008). "TiO₂-based phosphoproteomic analysis of the plasma membrane and the effects of phosphatase inhibitor treatment." *J Proteome Res* **7**(8): 3304-3313.
- Toueille, M., B. Mirabella, et al. (2012). "A comparative proteomic approach to better define *Deinococcus* nucleoid specificities." *J Proteomics* **75**(9): 2588-2600.
- Tran, J. C., L. Zamdborg, et al. (2011). "Mapping intact protein isoforms in discovery mode using top-down proteomics." *Nature* **480**(7376): 254-258.
- Upreti, R. K., M. Kumar, et al. (2003). "Bacterial glycoproteins: functions, biosynthesis and applications." *Proteomics* **3**(4): 363-379.
- van Noort, V., J. Seebacher, et al. (2012). "Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium." *Mol Syst Biol* **8**: 571.
- Venne, A. S., L. Kollipara, et al. (2014). "The next level of complexity: crosstalk of posttranslational modifications." *Proteomics* **14**(4-5): 513-524.
- Villen, J. and S. P. Gygi (2008). "The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry." *Nat Protoc* **3**(10): 1630-1638.
- Voller, A., A. Bartlett, et al. (1978). "Enzyme immunoassays with special reference to ELISA techniques." *J Clin Pathol* **31**(6): 507-520.
- Vosseller, K., J. C. Trinidad, et al. (2006). "O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry." *Mol Cell Proteomics* **5**(5): 923-934.
- Wagner, P. D. and N. D. Vu (2000). "Histidine to aspartate phosphotransferase activity of nm23 proteins: phosphorylation of aldolase C on Asp-319." *Biochem J* **346 Pt 3**: 623-630.

- Wang, Q., Y. Zhang, et al. (2010). "Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux." *Science* **327**(5968): 1004-1007.
- Wang, W., H. Zhou, et al. (2003). "Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards." *Anal Chem* **75**(18): 4818-4826.
- Washburn, M. P., D. Wolters, et al. (2001). "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." *Nat Biotechnol* **19**(3): 242-247.
- Whitehouse, C. M., R. N. Dreyer, et al. (1985). "Electrospray interface for liquid chromatographs and mass spectrometers." *Anal Chem* **57**(3): 675-679.
- Wilkinson, K. D. (1997). "Regulation of ubiquitin-dependent processes by deubiquitinating enzymes." *FASEB J* **11**(14): 1245-1256.
- Witze, E. S., W. M. Old, et al. (2007). "Mapping protein post-translational modifications with mass spectrometry." *Nat Methods* **4**(10): 798-806.
- Wolters, D. A., M. P. Washburn, et al. (2001). "An automated multidimensional protein identification technology for shotgun proteomics." *Anal Chem* **73**(23): 5683-5690.
- Wu, X., A. Vellaichamy, et al. (2013). "Differential lysine acetylation profiles of *Erwinia amylovora* strains revealed by proteomics." *J Proteomics* **79**: 60-71.
- Xing, G., J. Zhang, et al. (2008). "Identification of four novel types of in vitro protein modifications." *J Proteome Res* **7**(10): 4603-4608.
- Yan, W. and S. S. Chen (2005). "Mass spectrometry-based quantitative proteomic profiling." *Brief Funct Genomic Proteomic* **4**(1): 27-38.
- Yang, Y. Y. and H. C. Hang (2011). "Chemical approaches for the detection and synthesis of acetylated proteins." *Chembiochem* **12**(2): 314-322.
- Yarbrough, M. L. and K. Orth (2009). "AMPylation is a new post-translational modification." *Nat Chem Biol* **5**(6): 378-379.
- Yi, C. H., H. Pan, et al. (2011). "Metabolic regulation of protein N-alpha-acetylation by Bcl-xL promotes cell survival." *Cell* **146**(4): 607-620.
- Yoshihara, A., T. Hara, et al. (2012). "Regulation of dual oxidase (DUOX) expression and H₂O₂ production by thyroglobulin." *Thyroid*.
- Young, N. L., P. A. DiMaggio, et al. (2009). "High throughput characterization of combinatorial histone codes." *Mol Cell Proteomics* **8**(10): 2266-2284.
- Yu, B. J., J. A. Kim, et al. (2008). "The diversity of lysine-acetylated proteins in *Escherichia coli*." *J Microbiol Biotechnol* **18**(9): 1529-1536.
- Zanivan, S., M. Krueger, et al. (2012). "In vivo quantitative proteomics: the SILAC mouse." *Methods Mol Biol* **757**: 435-450.
- Zeidan, Q. and G. W. Hart (2010). "The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways." *J Cell Sci* **123**(Pt 1): 13-22.
- Zhang, J., Y. Chen, et al. (2010). "MS/MS/MS reveals false positive identification of histone serine methylation." *J Proteome Res* **9**(1): 585-594.
- Zhang, J., R. Sprung, et al. (2009). "Lysine acetylation is a highly abundant and evolutionarily conserved modification in *Escherichia coli*." *Mol Cell Proteomics* **8**(2): 215-225.
- Zhang, K., P. M. Yau, et al. (2004). "Differentiation between peptides containing acetylated or trimethylated lysines by mass spectrometry: an application for determining lysine 9 acetylation and methylation of histone H3." *Proteomics* **4**(1): 1-10.
- Zhang, K., S. Zheng, et al. (2013). "Comprehensive profiling of protein lysine acetylation in *Escherichia coli*." *J Proteome Res* **12**(2): 844-851.
- Zhang, X., H. Wen, et al. (2012). "Lysine methylation: beyond histones." *Acta Biochim Biophys Sin (Shanghai)* **44**(1): 14-27.
- Zhang, Z., M. Tan, et al. (2011). "Identification of lysine succinylation as a new post-translational modification." *Nat Chem Biol* **7**(1): 58-63.
- Zhao, S., W. Xu, et al. (2010). "Regulation of cellular metabolism by protein lysine acetylation." *Science* **327**(5968): 1000-1004.

- Zivanovic, Y., J. Armengaud, et al. (2009). "Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea." *Genome Biol* **10**(6): R70.
- Zoumaro-Djayoon, A. D., A. J. Heck, et al. (2012). "Targeted analysis of tyrosine phosphorylation by immuno-affinity enrichment of tyrosine phosphorylated peptides prior to mass spectrometric analysis." *Methods* **56**(2): 268-274.
- Zubarev, R. A., D. M. Horn, et al. (2000). "Electron capture dissociation for structural characterization of multiply charged protein cations." *Anal Chem* **72**(3): 563-573.
- Zubarev, R. A. and A. Makarov (2013). "Orbitrap mass spectrometry." *Anal Chem* **85**(11): 5288-5296.

Résumé

Exploration des modifications post-traductionnelles des protéines : Nouvelles approches et nouveaux modèles biologiques

L'étude des modifications post-traductionnelles a connu au cours des dernières années un regain d'intérêt notable. Tout d'abord car elle s'effectue aujourd'hui au travers d'approches basées sur la spectrométrie de masse, technique qui pendant cette période a connu de profonds bouleversements, conduisant à des études plus aisées et systématiques. Mais aussi car tant par leur variété que par le rôle qu'elles jouent dans la vie et la régulation cellulaire, ces modifications ne peuvent plus être négligées. Par ailleurs au cours de ces quinze dernières années, nous avons assisté concernant les procaryotes à un changement total de paradigme. En effet à la fin des années 90, l'idée dominante était que ces modifications pouvaient exister chez ceux-ci mais de façon très partielle et/ou très particulière. Dans ce travail, les divers degrés d'iodation de la tyrosine ont été sondés par une approche de type «shotgun» sur un organe entier, la thyroïde de souris. L'efficacité de ce type d'approche démontrée, les modifications post-traductionnelles potentiellement présentes dans des organismes modèles radio-résistants, la bactérie *Deinococcus deserti* et l'archée *Thermococcus gammatolerans* ont été analysées. Dans le premier cas, les données de protéomique montrent que de nombreuses acétylations N-terminales portent sur un motif spécifique (essentiellement des thréonines et sérines), cas très atypique pour une bactérie. Chez *Thermococcus gammatolerans* les acétylations N-terminales sont rares, mais la présence d'acétylations sur les chaînes latérales des lysines est notable. La présence de phosphorylations sur ces mêmes protéines, laisse entrevoir un possible phénomène de «cross talk» entre les lysines acétylées et les sérines, thréonines phosphorylées. Ici, nous démontrons que la complexité du protéome chez les procaryotes par le biais des MPT est bien réelle et que de possibles interdépendances entre MPT mériteraient un regard nouveau.

Abstract

Exploration of protein post-translational modifications: New approaches and novel biological models

Recently, the study of post-translational modifications has greatly evolved, mainly because of crucial progresses in mass spectrometry methodology which have allowed high-throughput, high resolution analysis. Their variety and their role in the regulation of key molecular mechanisms are increasingly documented. In this work, the different degrees of iodination of tyrosine were probed with a "shotgun" approach carried out from an entire organ, the mice thyroid. Post-translational modifications present in two radioresistant organism models, the bacterium *Deinococcus deserti* and the archaeon *Thermococcus gammatolerans*, were analyzed. The large scale exploration of N-terminal acetylation in *D. deserti* indicates a specific pattern of this modification on serine and threonine, as well as an atypical, high propensity to acetylation with 50% of modified N-termini. In *T. gammatolerans*, N-terminal acetylation is rare, but the presence of acetylation on lysine side chains is significant. The presence of phosphorylation on these proteins suggests a potential "cross talk" between the acetylated lysine and phosphorylated serine or threonine residues. This work demonstrates that the complexity of the proteome in prokaryotes through post-translational modifications is higher than expected when extremophiles are scrutinized compared to classical prokaryote models. Interdependencies between post-translational modifications definitively deserve a fresher look.

Mots clés

Service de Biochimie et Toxicologie Nucléaire

Modifications post-traductionnelles (MPT),

CEA Marcoule

spectrométrie de masse, procaryote,

DSV – iBEB- SBTN - LBSP

radio-résistant, acétylations N-terminale,

BP 17171 / 30207 Bagnols sur Cèze cedex

acétylations des lysines, acétylome.

Discipline : Biochimie – Biologie moléculaire