

ETUDE DU PEPTIDOME DU STRATUM CORNEUM HUMAIN PAR UNE APPROCHE PROTEOMIQUE

Anne-Marie Hesse

▶ **To cite this version:**

Anne-Marie Hesse. ETUDE DU PEPTIDOME DU STRATUM CORNEUM HUMAIN PAR UNE APPROCHE PROTEOMIQUE. Ingénierie biomédicale. Université Pierre et Marie Curie - Paris VI, 2009. Français. NNT : . pastel-00558351

HAL Id: pastel-00558351

https://pastel.hal.science/pastel-00558351

Submitted on 21 Jan 2011

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.

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité

Chimie Analytique

Présentée par

M^{elle} **Anne-Marie HESSE**

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

ETUDE DU PEPTIDOME DU STRATUM CORNEUM HUMAIN PAR UNE APPROCHE PROTEOMIQUE

soutenue le 18 mai 2009

devant le jury composé de :

M. Jean ROSSIER, président du jury

Mme Virginie REDEKER et M. Christian ROLANDO, rapporteurs

Mme Sandrine SAGAN et M. Dominique BERNARD, examinateurs

M^{elle} Joëlle VINH, directrice de thèse

Université Pierre & Marie Curie - Paris 6
Bureau d'accueil, inscription des doctorants et base de données
Esc G, 2ème étage
15 rue de l'école de médecine
75270-PARIS CEDEX 06

Tél. Secrétariat : 01 42 34 68 35
Fax : 01 42 34 68 40
Tél. pour les étudiants de A à EL : 01 42 34 69 54
Tél. pour les étudiants de EM à ME : 01 42 34 68 41
Tél. pour les étudiants de MF à Z : 01 42 34 68 51
E-mail : scolarite.doctorat@upmc.fr

Remerciements

Je tiens tout d'abord à exprimer mes remerciements aux membres du jury qui ont accepté d'évaluer mon travail de thèse. En premier lieu, merci à Jean Rossier, directeur du laboratoire de Neurobiologie et Diversité Cellulaire à l'ESPCI, et à Dominique Bernard, responsable du pôle Peau et Aspects de Surface de L'Oréal à Clichy, de m'avoir fait confiance C'est le rapprochement de ces deux laboratoires qui a rendu possible ces travaux. Je suis très reconnaissante envers Virginie Redeker et Christian Rolando d'avoir été les rapporteurs de cette thèse. Leurs remarques et suggestions m'ont permis d'améliorer la qualité de ce mémoire. Merci également à Sandrine Sagan pour avoir accepté d'examiner mon mémoire et de faire partie de mon jury de thèse. Enfin, je remercie Joëlle Vinh pour m'avoir accueillie dans l'équipe de Spectrométrie de Masse et pour avoir encadré mon travail depuis le DEA et pendant ces (nombreuses !) années de thèse. Son aide précieuse sur le plan scientifique mais aussi sa confiance et sa sympathie m'ont permis de mener à terme cette thèse.

Je tiens à saluer ici les personnes qui, de près ou de loin, ont contribué à la concrétisation de ce travail de thèse. En particulier l'ensemble des membres de l'équipe de Spectrométrie de Masse. A Iman Haddad pour ses talents en informatique qui ont permis de ranimer Mascot plus d'une fois après mes recherches buggées… A Paulo Marcelo pour les heures passées avec moi à démonter, remonter, déboucher, reboucher la LC… A Emmanuelle Demey pour être toujours disponible pour un peu de biblio en biologie et pour garder le labo en ordre !! A Yann Verdier qui malgré sa discrétion a su porter son regard pertinent de biologiste sur ce travail. A Sega Ndiaye et Giovanni Chiappetta pour tous ces moments d'entraide entre thésards dans les moments difficiles… A Florian Sahnoun pour ces 5 x 6 x 2 x 3 analyses dépouillées. A Joëlle Vinh pour avoir toujours été là au bon moment avec la bonne idée ! Evidemment ces remerciements ne sont pas exhaustifs : merci à eux pour leur soutien et leur aide quotidiens. J'ai été heureuse de partager ces 5 années avec eux et de nouer plus que des relations de travail mais aussi des liens d'amitié… Nos japonais et nos sorties bowling me manquent déjà.

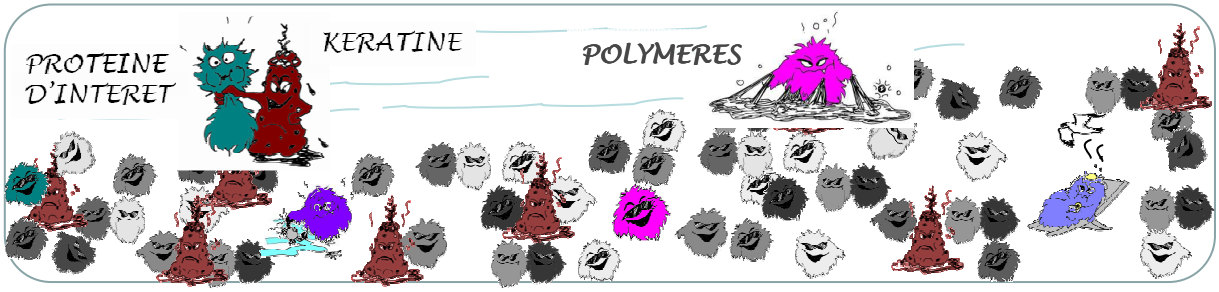
Merci à Dominique Bernard et Caroline Delattre de m'avoir initiée à la biologie du *stratum corneum* et à toute leur équipe pour leur accueil et l'aide qu'ils m'ont procurée : Mark, Lucie, Agnès, Géraldine, Françoise… J'en profite pour remercier tous les donneurs que j'ai pu martyriser pour obtenir des échantillons : Iman, Manu, Joëlle, Bob, Sylvie, Caroline mais aussi ceux que je ne connais pas !

Je tiens à remercier l'ensemble du laboratoire de Neurobiologie pour leurs encouragements et en particulier Sylvie Boulet, Robert Montagnat et Marcel Leopoldie pour la gestion administrative et logistique du labo qui permet aux chercheurs de travailler dans de bonnes conditions.

Merci à tous ceux qui ont relu articles et posters tout au long de ma thèse et cette « prose » en particulier : Joëlle, Dominique, Caroline, Iman, Manu, Paulo, Yann, Fred, Hélène, Bernadette et Magali.

Finalement j'adresse un grand merci à toute ma famille et mes amis pour leur soutien. Merci à mes parents et à mon frère de n'avoir jamais laissé Sébastien sur le pas de la porte quand nous rentrions fort tard… Merci à Fred d'avoir parcouru avec moi ce long chemin qu'est la thèse et à Sébastien de nous avoir pardonné les vacances et les WE sacrifiés. Enfin, merci à ma petite Maëlys d'avoir su rester calme tout au long de cette rédaction partagée et d'avoir attendu que la soutenance soit passée pour pointer le bout de son nez !

A mes enfants Sébastien et Maëlys


PROTEINE
D'INTERET
KERATINE
POLYMERES


THESARD
DIRECTEUR DE THESE

© E. Demey, 2009

Table des matières

PARTIE I- INTRODUCTION GÉNÉRALE 7
PARTIE II- INTRODUCTION 11
II-A. LES « OMICS »… 11
II-A.1. PROTÉOMIQUE 11
II-A.2. PEPTIDOMIQUE 12
II-A.2.1. Définition 12
II-A.2.2. Les difficultés de l'analyse : différences avec la protéomique 13
II-A.2.3. Les applications 14
II-A.3. DÉGRADOMIQUE 15
II-A.3.1. Définition 15
II-A.3.2. Les différentes approches 16
a La dégradomique transcriptionnelle : la technologie CLIP-CHIP™ 16
b La dégradomique traductionnelle : CLIP-PSP (Proteolytic Signature Peptide) 17
c Le dégradome actif 17
d Le dégradome du point de vue des substrats 18
II-B. LES TECHNIQUES ANALYTIQUES 18
II-B.1. LA SPECTROMÉTRIE DE MASSE 18
II-B.1.1. Les sources d'ionisation 19
a ESI 19
b MALDI 21
II-B.1.2. Les analyseurs 23
a Pourquoi a-t-on besoin de mesure de masse exacte ? 23
b La précision des mesures de masse 25
c Spectromètre de masse de type FTICR 26
d Spectromètre de masse de type Orbitrap 31
II-B.1.3. Fragmentation spécifique : Dissociation induite par collision 32
II-B.2. LES SÉPARATIONS CHROMATOGRAPHIQUES 33
II-B.2.1. Séparation par chromatographie à polarité de phases inversée 33
II-B.2.2. Capacité de pic 34
II-B.2.3. Augmenter la capacité de pic en RPLC 35
II-B.2.4. Augmenter la capacité de pic par couplage multidimensionnel 36
a La règle d'or 36
b Orthogonalité des deux dimensions 37
c Couplage SCX x RP 38
II-B.3. L'INTERPRÉTATION ET LA VALIDATION DES DONNÉES 40
II-B.3.1. Les identifications par recherche dans les banques 41
a BioWorks 42
b Mascot 43
II-B.3.2. L'identification par recherche dans les librairies de spectres 44
II-B.3.3. L'identification par séquençage *de novo* 45
II-B.3.4. L'identification par des approches hybrides 45
II-C. HISTOIRE DE LA PEAU 46
II-C.1. ORGANISATION DES CELLULES DE L'ÉPIDERME 46
II-C.1.1. Le *stratum* basal 47
II-C.1.2. Le *stratum spinosum* 47
II-C.1.3. Le *stratum granulosum* 48
II-C.1.4. Le *stratum lucidum* 48
II-C.1.5. Le *stratum corneum* 48

II-C.2. CONSTITUANTS DU *STRATUM CORNEUM* **49**
II-C.2.1. Les kératines 49
a Expression des kératines 49
b Les interactions protéines-kératines 50
c Les fonctions des kératines 50
II-C.2.2. Les jonctions cellulaires 50
a Généralités 50
b Structure et composition du desmosome et du cornéodesmosome 51
II-C.3. LA PEAU, UNE STRUCTURE VIVANTE : CAS PARTICULIER LA DESQUAMATION **53**
II-C.3.1. Généralités 53
II-C.3.2. Rôle des enzymes 54
a Les protéases à sérine 54
b Les protéases à acide aspartique 55
c Les protéases à cystéine 55
II-C.3.3. Régulation de la desquamation 55
II-C.4. DÉFINITION DES ÉCHANTILLONS **56**
PARTIE III- RESULTATS EXPERIMENTAUX et DISCUSSION **57**
III-A. MISES AU POINT ANALYTIQUES **57**
III-A.1. PRÉPARATION DES ÉCHANTILLONS **57**
III-A.1.1. Introduction 57
III-A.1.2. Reconcentration directe des peptides sur ZipTip™ 57
III-A.1.3. Fractionnement des peptides suivi de leur reconcentration sur ZipTip™ 60
III-A.1.4. De l'intérêt de la digestion des fractions 5-10kDa et sup 10kDa 62
III-A.1.5. Tampon d'extraction 64
III-A.1.6. Fractions de peptides utiles 65
III-A.2. CHROMATOGRAPHIE MONODIMENSIONNELLE : ELIMINATION DES POLYMÈRES **66**
III-A.2.1. Introduction 67
III-A.2.2. Caractérisation de la contamination 68
III-A.2.3. Elimination de la contamination 69
III-A.3. CHROMATOGRAPHIE MULTIDIMENSIONNELLE : OPTIMISATION **71**
III-A.3.1. Elimination des polymères 72
III-A.3.2. Mise au point du protocole *on-line* 74
a Influence de la concentration en sels du tampon salin 74
b Influence du gradient 75
c Influence de la concentration en acétonitrile 75
III-A.3.3. Mise au point du protocole *off-line* 76
a Influence de la concentration en acétonitrile 76
b Influence de la période d'échantillonnage 77
III-A.3.4. Comparaison des protocoles 78
a Caractéristiques physico-chimiques des peptides identifiés 78
b Orthogonalité 79
c Capacité de pic 80
d Identification peptidique et protéique 81
III-A.4. AUGMENTER LA COUVERTURE DE PROTÉOME AUTREMENT **83**
III-A.4.1. Travailler sur des spectromètres de masse différents 83
III-A.4.2. Utilisation de listes d'exclusions 85
III-A.4.3. Travailler sur des gradients de plus faible pente 86
III-A.5. MISE AU POINT DES ANALYSES PAR SPECTROMÉTRIE DE MASSE **88**
III-A.5.1. Configuration LIT-FTICR 88
a Cycle d'acquisition 88
b Gain en sensibilité et en gamme dynamique 89
c Comparaison des spectres MS/MS 90

III-A.5.2. Configuration LIT-Orbitrap 91
a Influence du seuil de déclenchement de la sélection des précurseurs 92
b Influence de la durée d'exclusion dynamique 96
c Influence du nombre de MS/MS 97
III-A.6. RECHERCHE DANS LES BANQUES DE DONNÉES : DE L'INTÉRÊT DES MODIFICATIONS 98
III-A.6.1. Protéines digérées 98
III-A.6.2. Peptides endogènes non digérés 101
a Cas de la pyroglutamisation 104
b Cas de l'acétylation N-terminale 105
c Données MS/MS 105
III-A.7. VALIDATION DES PROTÉINES IDENTIFIÉES PAR LC-MS/MS 107
III-B. APPLICATIONS BIOLOGIQUES 108
***III-B.1.* PEPTIDOME DE BASE DU *STRATUM CORNEUM* 108**
III-B.1.1. Identification des peptides 108
III-B.1.2. Comparaison topologique (PA/SCP) 112
III-B.1.3. Les protéines des cornéodesmosomes 115
a Les cadhérines 115
b La cornéodesmosine 118
c Les protéines de la plaque et les protéines armadillo 119
III-B.1.4. Les peptides antimicrobiens 120
a La dermcidine 121
b Les protéines de la famille S100 122
III-B.2. A LA RECHERCHE DES PROTÉASES IMPLIQUÉES : DONNÉES ISSUES DE LA LC-MS/MS 123
III-B.2.1. Tendance générale 123
III-B.2.2. Cas particulier : la cornéodesmosine 125
III-B.3. A LA RECHERCHE DES PROTÉASES IMPLIQUÉES : ANALYSES DE PEPTIDES FLUORESCENTS DE LA CORNÉODESMOSINE 126
III-B.3.1. Activité d'un extrait brut de protéases du *stratum corneum* 128
a Test enzymatique sur les peptides synthétiques de cornéodesmosine 128
b Vérification des sites de coupure in vitro par spectrométrie de masse 130
III-B.3.2. Activité d'un extrait fractionné par chromatographie 132
a Exemple d'un peptide dégradé en condition basique 134
b Exemple d'un peptide dégradé en condition acide 135
c Exemple d'un peptide dégradé de manière originale 136
PARTIE IV- CONCLUSION GENERALE 139
PARTIE V- PROTOCOLES EXPERIMENTAUX 145
V-A.1. PRÉPARATION DES ÉCHANTILLONS DE *STRATUM CORNEUM* NON DIGÉRÉ 145
V-A.1.1. Echantillons de peau 145
V-A.1.2. Extraction des peptides 145
V-A.1.3. Etape d'ultrafiltration 145
V-A.1.4. Etape de reconcentration des peptides 146
V-A.2. PRÉPARATION DES ÉCHANTILLONS DIGÉRÉS 147
V-A.3. SÉPARATION DES PEPTIDES PAR LC1D 148
V-A.3.1. Echantillons non digérés 148
a RPLC- nanoESI-QqTOF 148
b RPLC- nanoESI-FTMS 148
V-A.3.2. Echantillons digérés 149
a RPLC-nanoESI-FTMS 149
b RPLC-MALDI-TOF/TOF 149
V-A.4. SÉPARATION DES PEPTIDES PAR LC2D 150
V-A.4.1. Mode *on-line* 150
V-A.4.2. Mode *off-line* 152

V-A.5. ANALYSE DES PEPTIDES PAR SPECTROMÉTRIE DE MASSE 152
V-A.5.1. Analyse par spectrométrie de masse nanoESI-QqTOF 152
V-A.5.2. Analyse par spectrométrie de masse nanoESI-LIT-FTICR 153
V-A.5.3. Analyse par spectrométrie de masse nanoESI-LIT-Orbitrap 154
V-A.5.4. Analyse par spectrométrie de masse MALDI-TOF/TOF 154
V-A.6. RECHERCHE DANS LES BANQUES DE DONNÉES ET VALIDATION 154
V-A.6.1. Paramètres de recherche 154
a Echantillons non digérés 154
b Echantillons digérés 155
V-A.6.2. Paramètres de validation 155
a Analyses de peptides endogènes par nLC-nESI-QqTOF 155
b Analyses de peptides endogènes par nLC-nESI-LIT-FTICR 155
c Analyses de protéines digérées par nLC-nESI-LIT-FTICR 156
V-B. ACTIVITÉ DES PROTÉASES 156
V-B.1. EXTRACTION DES PROTÉINES DU *STRATUM CORNEUM* 156
V-B.2. PURIFICATION CHROMATOGRAPHIQUE 156
V-B.2.1. Chromatographie par exclusion stérique (SEC) 157
V-B.2.2. Echange de cations 157
V-B.3. TESTS ENZYMATIQUES 157
V-B.3.1. Test enzymatique générique 157
V-B.3.2. Test enzymatique spécifique des protéases à sérine 157
V-B.3.3. Test enzymatique sur les peptides de cornéodesmosine 158
V-B.4. ANALYSES PAR MALDI-TOF/TOF 158
PARTIE VI- TABLE DES FIGURES 159
PARTIE VII- TABLE DES TABLEAUX 164
PARTIE VIII- BIBLIOGRAPHIE 165

Publications

1. Articles scientifiques dans une revue internationale

[1] **Hesse A.M.**, Ndiaye S., Marcelo P., Rossier J., Vinh J., Optimized two-dimensional nano-liquid chromatography tandem mass spectrometry protocols for proteomics applications: *on-line* or *off-line* coupling?, en cours de rédaction.

[2] **Hesse A.M.**, Marcelo P., Delattre C., Rossier J., Bernard D., Vinh J., Analysis of the *stratum corneum* degradome by mass spectrometry : a peptidomics approach, en cours de rédaction.

[3] **Hesse A.M.**, Marcelo P., Rossier J., Vinh J., Simple and universal tool to remove *on-line* impurities in mono- or two-dimensional liquid chromatography–mass spectrometry analysis, J. Chromatogr. A (2008), Volume 1189, Issues 1-2.

2. Conférences orales

[1] **Hesse A.M.**, Marcelo P., Delattre C., Bernard D., Rossier J., Vinh J., Etude du dégradome de la peau par une approche protéomique, 2ème colloque Recherche de la Fédération Gay-Lussac, Rennes, France, 2008. **3ème prix de la FGL**.

[2] Ndiaye S., **Hesse A.M.**, Haddad I., Rossier J., Vinh J., Study of complex samples using multiple analytical dimensions: optimization of nanoLC-MALDI orbitrap and evaluation of complementary configurations, 25ème congrès de la Société Française d'Electrophorèse et d'Analyse Protéomique, Tours, France, 2008.

[3] Ndiaye S., **Hesse A.M.**, Rossier J., Vinh J., Improved knowledge of complex samples by multiplying dimensions of analyses: complementarity of 2D-LC-MALDI and 2D-LC-ESI, 25èmes Journées Françaises de Spectrométrie de Masse, Grenoble, France, 2008.

[4] **Hesse A.M.**, Marcelo P., Rossier J., Vinh J., Nanochromatographie et Protéomique, Journée sur l'innovation en protéomique et les nouvelles technologies pour la recherche en cancérologie, Grenoble, France, 2006.

[5] **Hesse A.M.**, Marcelo P., Rossier J., Vinh J., Peptidome et spectrométrie de masse : Application au dégradome du *stratum corneum*, Club Ile de France AFSEP, Paris, France, 2006.

3. Posters

[1] Verdier Y., Fukuyama H., **Hesse A.M.**, Haddad I., Makino C., Hoffmann J., Rossier J., Vinh J., Reliable interpretation from reliable data for complex proteomic characterization is a challenge, 25ème congrès de la Société Française d'Electrophorèse et d'Analyse Protéomique, Tours, France, 2008.

[2] Ndiaye S., **Hesse A.M.**, Rossier J., Vinh J., Improved knowledge of complex samples by multiplying dimensions of analyses: Complementarity of 2D-LC-MALDI and 2D-LC-ESI, 25ème congrès de la Société Française d'Electrophorèse et d'Analyse Protéomique, Tours, France, 2008.

[3] **Hesse A.M.**, Marcelo P., Rossier J., Vinh J., Optimized two-dimensional nano-liquid chromatography tandem mass spectrometry protocols for proteomics applications: *on-line* or *off-line* coupling?, 56th ASMS Conference on Mass Spectrometry, Denver, CO, USA, 2008.

[4] **Hesse A.M.**, Marcelo P., Rossier J., Vinh J., Improvement of sequence coverage and sensibility in two-dimensional nano-liquid chromatography tandem-mass spectrometry thanks to contaminants' removal, 56th ASMS Conference on Mass Spectrometry, Denver, CO, USA, 2008.

[5] Ndiaye S., **Hesse A.M.**, Rossier J., Vinh J., Complementarities of MALDI and ESI for 2DLC MS/MS: Expanding proteome coverage of complex samples with multiple dimensions of analysis, 56th ASMS Conference on Mass Spectrometry, Denver, CO, USA, 2008.

[6] Vinh J., Haddad I., Ndiaye S., **Hesse A.M.**, Rossier J., NanoLC-MALDI Orbitrap coupling evaluation: an attempt to optimized the acquisition strategy, 56th ASMS Conference on Mass Spectrometry, Denver, CO, USA, 2008.

[7] **Hesse A.M.**, Marcelo P., Rossier J., Vinh J., Optimized two-dimensional nano-liquid chromatography tandem mass spectrometry protocols for proteomics applications: *on-line* or *off-line* coupling?, 32nd International Symposium on High Performance Liquid Phase Separations and Related Techniques, Baltimore, MD, USA, 2008.

[8] Ndiaye S., **Hesse A.M.**, Rossier J., Vinh J., Complementarities of MALDI and ESI for 2DLC MS/MS: Expanding proteome coverage of complex samples with multiple dimensions of analysis, 32nd International Symposium on High Performance Liquid Phase Separations and Related Techniques, Baltimore, MD, USA, 2008.

[9] **Hesse A.M.**, Marcelo P., Rossier J., Vinh J., Un outil simple et universel pour éliminer les contaminants lors des analyses 1D- et 2D-LC-MS/MS, 24èmes Journées Françaises de Spectrométrie de Masse, Pau, France, 2007.

[10] **Hesse A.M.**, Marcelo P., Rossier J., Vinh J., A simple and universal tool to remove online impurities in LC-MS analyses, 31st International Symposium on High Performance Liquid Phase Separations and Related Techniques, Ghent, Belgium, 2007. **Sélectionné parmi les 25 meilleurs posters du congrès.**

[11] **Hesse A.M.**, Royer N., Marcelo P., Rossier J., Vinh J., Peptidome et Spectrométrie de masse, 22èmes Journées Françaises de Spectrométrie de Masse, Montpellier, France, 2005.

[12] **Hesse A.M.**, Labas V., Royer N., Marcelo P., Rossier J., Vinh J., Enrichissement et purification de peptides biologiques à partir de matrice biologique, 21èmes Journées Françaises de Spectrométrie de Masse, Strasbourg, France, 2004.

Abréviations et quelques définitions

a.a. : acide aminé

ACN : acétonitrile

ACTH : hormone corticotrope

AF : acide formique

ADN : acide désoxyribonucléique

AMP : peptide antimicrobien

AMT : *accurate mass and time* (étiquette masse précise/temps de rétention précis)

API : inhibiteur de pepsin sécrété par l'Ascaris

AQUA : *Absolute QUAntitation* (quantification absolue)

ARN : acide ribonucléique

ARN_m : ARN messager

BP : *Base Peak* (pic de base)

CAD : dissociation activée par collision

CDSN : cornéodesmosine

CHCA : acide α-cyano-4-hydroxy cinnamique

CID : dissociation induite par collision

cps : coups (mesure de l'intensité en masse)

d.i. : diamètre interne d'une colonne, d'un capillaire, *etc.*

DSC1 : desmocolline 1

DSG1 : desmogléine 1

DTT : dithiothreitol

ECD : dissociation par capture d'électron

EDTA : acide éthylène-diamine-tétraacétique

ESI : ionisation électrospray

EST : *Expressed Sequence Tag*

FI : filament intermédiaire

FTICR MS : Spectrométrie de masse par Transformée de Fourier à Résonance Cyclotronique d'Ions

HPLC : Chromatographie Liquide à Haute Pression

ICAT : *Isotope Coded Affinity Tag*

IPr : isopropanol

LIT : trappe d'ions linéaire

iTRAQ : *isobaric Tags for Relative and Absolute Quantification* (Quantification absolue et relative par étiquette isobarique)

L : longueur d'une colonne chromatographique ou d'un capillaire

LC : chromatographie en phase liquide

LC-MS/MS : couplage entre chromatographie en phase liquide et spectrométrie de masse en tandem

LC1D : chromatographie en phase liquide monodimensionnelle

LC2D : chromatographie en phase liquide bidimensionnelle

MALDI : désorption-ionisation laser assistée par matrice

MALDI-TOF/TOF : spectrométrie de masse en tandem de type MALDI couplé avec une analyse en temps de vol

MeOH : méthanol

MW : masse moléculaire d'une protéine ou d'un peptide

MMP-2 : métalloprotéinase matricielle 2 ou gélatinase

MPT : modification post-traductionnelle

MS : spectrométrie de masse

Mode MS : mode d'analyse par spectrométrie de masse dans lequel le rapport masse sur charge des analytes ionisés est déterminé

Mode MS/MS : mode d'analyse par spectrométrie de masse dans lequel un ion donné est sélectionné pour être fragmenté. Les ions fragments générés sont séparés en fonction de leur rapport masse sur charge et détectés.

m/z : rapport masse sur charge, où m est la masse en Dalton (Da) et z le nombre de charges de l'ion

N : nombre de plateaux théoriques caractérisant une séparation. $N = 16\left(\frac{t_R}{\omega}\right)^2$ (voir figure associée à la définition de R_s pour la description de t_R et ω)

NaCl : chlorure de sodium

nanoESI ou nESI : ionisation nanoélectrospray

nanoLC ou nLC : chromatographie en phase liquide utilisant des débits de l'ordre de quelques centaines-dizaines de nL/min

PA : poudre acétonique

P_c : capacité de pic d'une technique séparative

PBS : *Phosphate Buffer Saline* (tampon phosphate)

pI : point isoélectrique

ppm : partie par million

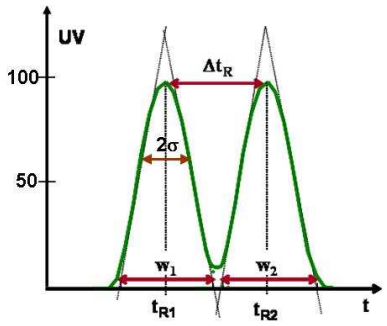
peptides protéotypiques : Les peptides générés à partir de la protéolyse d'une séquence protéique qui sont toujours détectés en MS sont appelés protéotypiques.

PS-DVB : poly-(styrène-divinylbenzène)

Q : quadripôle

R_s : résolution d'une technique séparative. La résolution R_s entre deux pics chromatographiques est donnée par :

$R_s = \frac{2(t_{R2} - t_{R1})}{\omega_2 - \omega_1}$, où t_{R1} et t_{R2} désignent les temps de rétention des deux analytes entre lesquels la résolution est calculée, et ω_2 et ω_1 désignent la largeur de leur pic chromatographique. σ désigne l'écart type mesuré à 60,6% de la hauteur du pic. Généralement, on prend 4σ comme définition de ω.



RPLC : chromatographie en phase liquide à polarité de phases inversée

SA : acide sinapinique

SC : *stratum corneum*

SCCE : kallikréine 7, *stratum corneum chymotryptic enzyme*

SCP : *stratum corneum* plantaire

SCTE : kallikréine 5, *stratum corneum tryptic enzyme*

SCCL : *stratum corneum cathepsin like*

SCTP : *stratum corneum thiol protease*

SCX-LC : Chromatographie d'échange de cations forts

SEC : chromatographie d'exclusion stérique

S/B : rapport signal sur bruit

SRM : *Selected Reaction Monitoring*

TFA : acide trifluoroacétique

Th : Thompson, unité des rapports m/z

TOF : analyseur à temps de vol

tpm : tour par minute

t_R : temps de rétention chromatographique

Tris : 2-amino-2-hydroxyméthyl-1,3-propanediol

UHPLC : Chromatographie Liquide à Ultra Haute Pression

Tampon, solvant, phase mobile : ces trois termes sont employés pour désigner les phases mobiles utilisées en chromatographie en phase liquide. Le mot tampon est utilisé de façon détournée, dans la mesure où la phase mobile aqueuse employée n'est pas un tampon de pH. Le pH est imposé par l'acide organique introduit dans la phase mobile (acide formique dans nos analyses), mais sa valeur n'a pas été ajustée pour que la solution soit au pKa de l'acide faible. Le terme solvant est employé en général par analogie aux publications en anglais (*solvent*).

PARTIE I- INTRODUCTION GÉNÉRALE

Comprendre le fonctionnement d'une cellule suppose de déchiffrer les mécanismes moléculaires complexes qui sous-tendent les différentes activités cellulaires. L'ensemble des gènes, ou génome, d'un organisme constitue une banque de données statique et caractéristique de cet organisme. A partir d'un génome unique, chaque type cellulaire d'un organisme va exprimer un ensemble de protéines, ou protéome, qui va varier en fonction de l'environnement de la cellule. Comme les protéines constituent les acteurs finaux des processus biologiques, leur étude est susceptible d'offrir la vision la plus pertinente du fonctionnement d'une cellule.

Le concept de **l'analyse protéomique** a été introduit en 1998 par Anderson et Anderson, qui le définissent comme l'analyse quantitative des niveaux d'expression des protéines, visant à déchiffrer les mécanismes du contrôle de l'expression génique [1]. Les études protéomiques peuvent être envisagées à des échelles variées :

- **Exhaustives** pour détecter le plus grand nombre possible de protéines d'un échantillon complexe,
- **Différentielles** pour comparer des extraits protéiques d'origine cellulaire différente et mettre en évidence par exemple des marqueurs de pathologies,
- **Ciblées** pour mettre en évidence les protéines impliquées dans une fonction cellulaire donnée.

Les applications de la protéomique sont aujourd'hui variées et ont donné naissance à de nouvelles disciplines. En effet, les approches associées à la protéomique sont, pour des raisons méthodologiques, limitées à l'analyse de protéines de masses moléculaires supérieures à 10kDa. Le développement de méthodologies couvrant les peptides et les protéines de plus bas poids moléculaire (0,5 à 15kDa) était donc nécessaire. Parmi ces peptides, se trouvent les familles des hormones, des cytokines et des facteurs de croissance, qui jouent un rôle central dans de nombreux processus biologiques. Pour récapituler les technologies utilisées pour identifier ces peptides, le terme **peptidomique** a été introduit (chapitre II-A.2). D'autre part, dans le génome humain, les protéases représentent 1,7% des gènes. Les protéases interagissent ou sont dynamiquement régulées par des milliers de protéines, d'inhibiteurs endogènes et de substrats. Même si de nombreuses voies biologiques impliquant une activité peptidase ont été caractérisées, la diversité des événements protéolytiques dans les systèmes biologiques, leur distribution spatiale ou temporelle, ainsi que leur rôle dans le développement de pathologies, sont loin d'être définis.

Pour décrire le réseau des protéases et leurs activités, de nouvelles approches à haut débit ont été développées en spectrométrie de masse dans un domaine appelé **dégradomique** (chapitre II-A.3).

L'analyse protéomique est une **science intégrative** qui est un **défi analytique** en raison de l'incroyable diversité chimique des protéines et de la très large gamme de concentrations auxquelles elles sont exprimées. Pour répondre à ce défi, des techniques puissantes de séparation ont été développées. En particulier, la chromatographie liquide à polarité de phases inversée s'est imposée pour l'analyse de mélanges complexes de protéines ou de peptides protéolytiques. Cependant, il est généralement accepté qu'aucune séparation chromatographique monodimensionnelle n'offre une **capacité de pic** suffisante pour résoudre un mélange complexe de peptides provenant de la digestion d'un protéome complexe. C'est pourquoi les efforts de développement ont été portés sur l'augmentation de la capacité de pic des séparations chromatographiques (chapitre II-B.2). Les techniques de séparation en phase liquide sont généralement associées à la spectrométrie de masse, qui fournit l'identification des protéines étudiées. Les outils de spectrométrie de masse biologique, utilisant le plus souvent des sources d'ionisation de type MALDI (*Matrix-Assisted Laser Desorption Ionization*) ou ESI (*Electrospray*), se sont rapidement imposés depuis leur introduction à la fin des années 1980, comme des techniques de caractérisation sensibles et universelles des protéines. Pour répondre aux exigences de plus en plus importantes en termes de **sensibilité**, de **gamme dynamique** et de **précision**, de nouveaux instruments de type FTMS ont fait leur apparition. Les principes de ces instruments sont détaillés dans le chapitre II-B.1. L'identification d'une protéine nécessite d'obtenir des informations de séquence par une analyse des peptides protéolytiques par spectrométrie de masse en tandem. Des logiciels puissants ont été développés, permettant d'interpréter automatiquement les spectres de fragmentation pour leur associer des séquences en acides aminés rendant compte des ions fragments observés (chapitre II-B.3).

Travail réalisé dans cette thèse

Cette thèse se place dans le **contexte de l'analyse protéomique et peptidomique**. D'un point de vue analytique, l'objectif de ma thèse a été d'implémenter au laboratoire des techniques capables de fournir rapidement des identifications fiables de protéines et de peptides en mélange complexe, quelles que soient leurs propriétés physico-chimiques. D'un point de vue biologique, nous souhaitions appréhender le **phénomène de desquamation du *stratum corneum*** chez l'Homme, en collaboration avec la société L'Oréal. Ce phénomène implique la dégradation par des enzymes spécifiques des protéines des cornéodesmosomes, jonctions cellulaires qui assurent la cohésion des cellules entre elles. Contrairement aux approches biochimiques classiques utilisées généralement dans ce type d'application (ciblage de protéases spécifiques), nous avons choisi d'adopter une méthodologie sans *a priori,* qui fait le lien entre la peptidomique et la dégradomique. Pour ce faire, nous avons centré notre étude sur les

peptides endogènes issus de la dégradation des protéines du *stratum corneum* afin d'identifier des peptides spécifiques de la desquamation.

L'approche développée, même si elle utilise les outils analytiques classiques de la protéomique, doit être optimisée à toutes les étapes dans le cadre de **l'étude de peptides endogènes**. En effet, les peptides endogènes recherchés sont des fragments de protéines générés par différentes voies métaboliques : ils sont donc extrêmement sensibles à la protéolyse lors des étapes d'extraction et de purification. La méthode de préparation des échantillons doit donc cibler spécifiquement ces peptides : nous avons opté pour une méthode de fractionnement des échantillons suivie d'une méthode d'enrichissement sur phase solide dont l'optimisation est présentée dans le chapitre III-A.1.

D'autre part, tout échantillon biologique est complexe. Cette complexité vient à la fois des **contaminants** de l'échantillon (constitutifs ou externes) mais également de la très large gamme dynamique de concentration des analytes. Afin de monter en sensibilité, nous nous sommes donc attachés à éliminer les contaminations externes auxquelles nous devions faire face lors des séparations chromatographiques. Il peut s'agir de composés polymériques hydrophobes qui sont élués dans la même gamme de temps de rétention que les peptides et masquent alors leur signal. Ils sont décrochés par les solvants sur le circuit chromatographique. Leur élimination est détaillée dans le chapitre III-A.2. En ce qui concerne les contaminants constitutifs (les kératines de la peau dans notre cas), la solution peut être de monter en gamme dynamique et de rendre notre analyse la plus exhaustive possible. Pour cela, l'approche couramment utilisée en protéomique est l'augmentation de la capacité de pic par couplage de plusieurs séparations chromatographiques orthogonales. L'optimisation et l'évaluation des différents modes de couplage possibles (*on-line* ou *off-line*) sur des échantillons de *stratum corneum* sont expliquées dans le chapitre III-A.3. Des solutions alternatives moins coûteuses en temps sont également décrites dans le chapitre III-A.4.

La différence principale dans les études peptidomiques par rapport aux études protéomiques classiques est **l'absence d'une étape de digestion *in vitro*** préalable. Or, cette absence supprime un paramètre important de la recherche dans les banques de données, à savoir la spécificité de l'enzyme utilisée. Pour pallier cette difficulté, le recours à de nouvelles instrumentations de type spectrométrie de masse à transformée de Fourier permet de diminuer le nombre de peptides candidats dans la banque de données pour un peptide fragmenté expérimentalement, grâce à **l'augmentation de la précision des mesures de masse**. L'optimisation des cycles d'acquisition sur ces instruments dans le cadre de l'étude des peptides endogènes ainsi que le gain en sensibilité et en gamme dynamique qui en découle sont présentés dans le chapitre III-A.5.

D'autre part, les potentialités du logiciel de recherche **Mascot** [2], utilisé tout au long de cette thèse, ont particulièrement été évaluées sur nos échantillons pour augmenter la couverture du protéome et du peptidome. Ainsi, dans le cas d'échantillons digérés, l'utilisation de certaines fonctions du logiciel permet d'augmenter la couverture de séquence des protéines identifiées grâce à la découverte par, exemple, de **peptides semi-tryptiques**. En ce qui concerne l'étude du peptidome du *stratum corneum,* certains **peptides endogènes** n'ont pu être identifiés que sous une forme modifiée post-traductionnelle. Tout ceci est détaillé dans le chapitre III-A.6.

Enfin, l'optimisation analytique du couplage de chromatographie en phase liquide monodimensionnelle ou bidimensionnelle avec la spectrométrie de masse tandem, et des paramètres utiles à l'exploitation des données, a été appliquée à l'analyse d'échantillons humains de ***stratum corneum***. Ceci nous a permis d'obtenir une meilleure connaissance du processus de desquamation en identifiant des peptides endogènes issus de la dégradation de protéines des **cornéodesmosomes** (chapitre III-B.1). Ce type d'approche sans *a priori* nous a permis de proposer des activités protéolytiques spécifiques. Les **protéases** impliquées sont associées à la libération des peptides endogènes identifiés (chapitre III-B.2) en corrélation avec le phénomène de desquamation étudié. Certaines de ces activités ont d'ailleurs pu être vérifiées par des approches biochimiques plus classiques sur le modèle de la dégradation de la **cornéodesmosine**, une protéine particulièrement importante dans les **structures jonctionnelles du *stratum corneum*** impliquées dans le **phénomène de desquamation** (chapitre III-B.3).

PARTIE II- INTRODUCTION

II-A. Les « Omics »…

Les Omics recouvrent différents champs d'investigation et nouvelles technologies dans plusieurs domaines : génomique, protéomique, métabolomique… Dans le cadre de ce travail de thèse, nous ne développerons dans cette partie que les Omics se rapportant au domaine des polypeptides et à la thématique des protéases.

II-A.1. Protéomique

Le concept de **l'analyse protéomique** a été introduit en 1998 par Anderson et Anderson, qui le définissent comme l'analyse quantitative des niveaux d'expression des protéines, visant à déchiffrer les mécanismes du contrôle de l'expression génique [1]. Les études protéomiques peuvent être envisagées à des échelles variées.

Avec l'achèvement du Projet de Génome Humain en 2003, la recherche scientifique sur la détermination des protéines du corps humain s'est accentuée. Cette étude de toutes les protéines produites par les cellules et l'organisme implique l'identification des protéines de l'organisme et la détermination de leurs rôles dans les fonctions physiologiques et pathologiques. Le séquençage de nombreux génomes entiers de procaryotes et d'eucaryotes a suscité un regain d'intérêt pour la protéomique. Le terme protéome désigne **l'ensemble des protéines exprimées par un génome donné dans un environnement donné, à un moment donné**. Tandis qu'un génome reste globalement inchangé, chaque cellule exprime des protéines particulières de façon radicalement différente en réponse à son environnement. Dans la mesure où ce sont les protéines qui sont directement impliquées dans la plupart des processus biochimiques à la fois normaux et pathologiques, une compréhension plus complète des maladies peut être obtenue en étudiant directement les protéines d'une cellule ou d'un tissu malade. Le protéome est par nature dynamique, ce qui a conduit certains chercheurs à utiliser le terme de **protéome fonctionnel pour décrire l'ensemble des protéines exprimées par une cellule spécifique à un temps donné**. Alors que le génome humain contient environ 22 000 gènes codant potentiellement pour 40 000 protéines différentes, l'épissage alternatif de l'ARN et les modifications post-traductionnelles augmentent le nombre de protéines ou de fragments de protéines jusqu'à 2 000 000 [3]. En conséquence, le protéome est beaucoup plus complexe que le génome.

La génomique a apporté une grande quantité d'informations liant l'activité d'un gène à une maladie, mais elle ne permet pas de prédire les **modifications post-traductionnelles** portées par de nombreuses protéines. Bien qu'il soit plus facile de travailler sur l'ADN ou l'ARN, les informations déduites de leurs analyses sont limitées. Pour de nombreuses raisons, la séquence d'un gène et son profil d'activité ne sont pas directement corrélés à l'image complète et précise de l'abondance d'une protéine, à sa structure moléculaire finale ou à son état d'activité. Le transcrit d'un gène peut être épissé de différentes manières avant la traduction en protéines. Après synthèse par les ribosomes, les protéines sont clivées pour éliminer les séquences d'initiation, de transfert et de signal puis subissent des modifications post-traductionnelles (MPTs), par exemple par l'ajout de sucres ou de phosphorylations. De telles modifications jouent un rôle vital dans la modulation de la fonction des protéines mais ne sont pas directement codées au niveau des gènes. Les MPTs sont nombreuses (plus de 300 ont été dénombrées). Elles peuvent être statiques ou dynamiques. En conséquence, un unique gène peut encoder jusqu'à une cinquantaine d'espèces protéiques. Ainsi, l'analyse de la séquence d'ADN ne peut pas prédire les formes actives d'une protéine. De même, la quantification des ARN ne reflète pas toujours les niveaux d'expression protéique correspondants. Au-delà de l'aspect génomique, il est clair que l'information apportée par le génome peut ne pas donner un profil précis de l'abondance des protéines, de leurs structures ou de leurs activités [4].

En réalité, la **génomique et la protéomique sont des domaines complémentaires**. Grâce à la génomique et à la protéomique, de nouveaux biomarqueurs de maladie et de nouvelles cibles thérapeutiques peuvent être identifiés, ce qui pourra aider à concevoir des produits pour prévenir, diagnostiquer et traiter ces maladies. La génomique et la protéomique pourraient donc influencer de façon notable l'avenir de la médecine et des biotechnologies, mais il y a encore beaucoup à faire pour en comprendre les bénéfices et passer aux applications cliniques.

II-A.2. Peptidomique

II-A.2.1. Définition

La protéomique s'est établie comme une méthodologie remarquable pour fournir des données sur l'identité des protéines et de leurs fonctions. Cependant, les approches associées sont, pour des raisons méthodologiques, limitées à l'analyse de protéines de masses moléculaires supérieures à 10kDa. Le développement de méthodologies couvrant les peptides et les protéines de plus bas poids moléculaire (0,5 à 15kDa) était nécessaire. En effet, parmi ces peptides, se trouve la famille des hormones, des cytokines et des facteurs de croissance, qui jouent un rôle central dans de nombreux processus biologiques. Pour récapituler les technologies utilisées pour identifier ces peptides, le terme

peptidomique (en anglais *peptidomics*) a été introduit. Cette notion a été utilisée pour la première fois en 2001 par Schulz-Knappe, et définie comme un **sous-ensemble de la protéomique pour la description des peptides**, considérés, eux aussi, comme produits de gènes [5, 6].

La peptidomique a pour but d'identifier tous les peptides endogènes d'une cellule, d'un organe ou d'un organisme (*i.e.* le peptidome). La peptidomique se définit comme une méthodologie permettant la description la plus complète possible des peptides d'un échantillon biologique, tant d'un point de vue qualitatif que quantitatif.

II-A.2.2. Les difficultés de l'analyse : différences avec la protéomique

Bien que la peptidomique et la protéomique partagent de nombreux objectifs scientifiques et de nombreux outils techniques, il y a quelques **différences majeures** entre ces approches. La plupart des études protéomiques utilisent des enzymes de digestion comme la trypsine pour générer des fragments qui peuvent être séquencés par MS/MS afin d'identifier les protéines présentes dans l'échantillon [7]. A l'inverse, la plupart des études en peptidomique n'utilisent **aucune étape de digestion *in vitro*,** afin d'identifier la forme native des peptides, en incluant les modifications post-traductionnelles [5, 8]. Bien que la suppression de cette étape de digestion simplifie la préparation des échantillons, l'analyse des données est généralement plus délicate pour les gros peptides. En protéomique, il n'est pas nécessaire d'identifier tous les fragments tryptiques pour valider l'identification d'une protéine. En revanche, en peptidomique, le but ultime est d'identifier le maximum de peptides possible même s'ils dérivent du même précurseur protéique, car chaque peptide peut avoir une fonction distincte [9]. Les techniques protéomiques et peptidomiques partagent le problème de gamme dynamique limitée de la plupart des spectromètres de masse par rapport à la gamme de concentration beaucoup plus large des protéines ou des peptides présents dans les échantillons biologiques. En général, les techniques de spectrométrie de masse (MS) peuvent aisément détecter des protéines et des peptides présents en quantité importante ou modérée mais s'adaptent mal à la détection des molécules présentes en faible quantité. Contrairement à la protéomique, la gamme dynamique des analyses peptidomiques est souvent limitée par la dégradation des protéines. Dans des conditions standard de dissection des tissus et d'extraction des peptides, il y a une petite fraction de protéines qui se dégradent par des processus enzymatiques ou non. La plupart des études protéomiques ne seront pas affectées si une faible proportion de protéines est dégradée pendant que la majorité reste intacte. Cependant, les peptides bioactifs sont généralement présents en faible quantité par rapport aux protéines majeures. Ainsi **la dégradation (même infime) de ces protéines majeures peut masquer les signaux plus faibles des peptides endogènes d'intérêt.** Il est donc nécessaire de s'affranchir de cette limitation en choisissant par exemple des tissus pauvres en enzymes [10] ou alors en enrichissant les peptides sur des colonnes d'affinité [11].

D'autre part, la séparation par **électrophorèse sur gel bidimensionnel** d'un mélange de peptides n'est pas satisfaisante. La coloration de peptides plus courts donne des intensités bien plus faibles que la coloration des protéines dans la mesure où cette coloration dépend à la fois de la quantité de peptides mais aussi du nombre d'acides aminés capables de lier les colorants [12] : de ce fait, cette coloration est insuffisante pour être détectée. Enfin, la **cartographie peptidique en mode MALDI** (désorption-ionisation laser assistée par matrice) n'est pas applicable directement pour l'identification de ces peptides endogènes, notamment puisque qu'aucune digestion enzymatique n'est réalisée. Les recherches dans les banques de données ne comportent dans ce cas pas assez de paramètres discriminants. Des données complémentaires de fragmentation par spectrométrie de masse en tandem (mode MS/MS) sont alors nécessaires [13].

En revanche, les peptides peuvent être séparés de manière très résolue par chromatographie, par exemple sur des colonnes à polarité de phases inversée. Lors d'une séparation chromatographique, les analytes restent en phase liquide, ce qui est un avantage notable. Il n'y a plus de perte de protéines lors des étapes d'extraction du gel. Cependant, l'identification des peptides uniquement par séparation chromatographique fait également face à certains problèmes spécifiques. En effet, les extraits peptidiques contiennent aussi une grande proportion de protéines comparée à la quantité de peptides récupérés. Les séparations chromatographiques sont alors compromises en termes de résolution et de capacité de chargement. Pour réduire la quantité de protéines chargées, la séparation des peptides des protéines de poids supérieur à 10 ou 20kDa s'avère être une étape nécessaire pour la préparation de l'échantillon.

La procédure classique pour une analyse peptidomique commence par la purification et l'extraction des peptides et des protéines. Des colonnes à exclusion de taille sont souvent utilisées pour séparer les peptides des protéines plus grosses. Si une approche par spectrométrie de masse tandem de type MALDI avec une analyse en temps de vol (MALDI TOF/TOF) est choisie, cette séparation n'est pas obligatoirement nécessaire. En effet, il est possible de s'appuyer sur la discrimination en termes d'efficacité d'ionisation qui favorise les espèces de faible rapport masse sur charge (m/z).

II-A.2.3. Les applications

Parmi les applications de la peptidomique, on peut citer **l'analyse des fluides extracellulaires** pour la comparaison de groupes de patients et la **détection de marqueurs diagnostic de pathologies** qui en découle. Récemment, des profils de spectres de masse peptidomiques ont été utilisés pour le diagnostic de maladies. Le diagnostic du cancer des ovaires a été poursuivi par exemple grâce à une

approche bioinformatique par reconnaissance des profils peptidiques sur les spectres de masse [14]. Cependant, ce type d'étude se fait sans identification des séquences spécifiques des biomarqueurs.

La caractérisation des peptides endogènes se fait essentiellement sans digestion *in vitro*. Il n'est donc pas possible de s'appuyer sur la spécificité d'endoprotéases pour l'identification des séquences d'intérêt. Afin de lever les ambiguïtés de séquence entre différentes espèces quasi isobares il est donc nécessaire de les résoudre (**travail à haute résolution**), et de mesurer les rapports m/z avec une haute précision (**travail à haute justesse**). De plus certains peptides peuvent avoir des masses relativement élevées (**travail à haute gamme de masse**). Enfin l'absence d'information sur les sites de clivage enzymatique implique des **ressources informatiques conséquentes** car les recherches sur les bases de données sont beaucoup plus exigeantes que les recherches de peptides tryptiques simples.

II-A.3. Dégradomique

II-A.3.1. Définition

Dans un génome typique, entre 2 et 4% de gènes codent pour des enzymes protéolytiques [15]. Ainsi, dans le génome humain, **les protéases représentent 1,7% des gènes**, et forment la plus grande famille d'enzymes humaines, avec 566 protéines [16]. Les protéases interagissent ou sont dynamiquement régulées par des milliers de protéines, de cofacteurs, de récepteurs, d'inhibiteurs endogènes et de substrats. Les enzymes protéolytiques agissent comme des effecteurs positifs ou négatifs de nombreux processus biologiques, comme des catalyseurs non spécifiques de la dégradation des protéines ou encore comme des agents très sélectifs contrôlant des événements physiologiques [17]. Même si de nombreuses voies biologiques impliquant une activité peptidase ont été caractérisées comme la cascade de coagulation ou le phénomène d'apoptose, la diversité des événements protéolytiques dans les systèmes biologiques, leur distribution spatiale ou temporelle, ainsi que leur rôle dans le développement de pathologies, sont loin d'être définis.

Pour décrire le réseau des protéases et leurs activités, de nouvelles approches à haut débit ont été développées en MS dans un domaine appelé **dégradomique**. Ce terme a été utilisé dès 2000 pour définir le répertoire de substrats d'une protéase à l'échelle du protéome [18]. Puis cette définition a été étendue, par Lopez-Otin *et al.,* aux techniques génomiques et protéomiques permettant l'identification et la caractérisation des protéases présentes dans un organisme, en incluant aussi leurs substrats et leurs inhibiteurs endogènes [19]. Ils ont également proposé le terme dégradome pour englober deux concepts :

- l'ensemble des protéases exprimées à un moment donné ou dans des circonstances données par une cellule (un tissu ou un organisme) comme le dégradome d'un modèle donné,
- l'ensemble des substrats naturels d'une enzyme dans cette cellule (tissu ou organisme) comme le dégradome d'une protéase donnée.

Il est nécessaire d'identifier tous les éléments du réseau des protéases avec l'analyse de leurs partenaires et de leurs substrats. Dans l'étude du dégradome, différentes questions peuvent être posées :

- Quelles sont les protéases exprimées par une cellule, un tissu ou un organe donné dans des conditions définies ?
- Quels sont, parmi elles, les espèces présentant une activité réelle ?
- Quels sont les substrats de ces protéases et leurs produits de dégradation ?
- Quels peuvent être les effets de ce potentiel protéolytique sur le système étudié ?

Pour répondre à ces questions, de nombreuses techniques ont été développées pour étudier les fonctions biologiques des protéases.

II-A.3.2. Les différentes approches

a La dégradomique transcriptionnelle : la technologie CLIP-CHIP™

Depuis une quinzaine d'années et l'invention des micro-puces, la **quantification des ARN$_m$** dans les cellules et les tissus s'est développée et a été fortement standardisée. L'analyse de l'abondance des ARN$_m$ entre des états sains et malades permet notamment l'**identification de signatures génétiques reflétant les changements importants de l'homéostasie des cellules et des tissus**. Mais même les dernières générations de puces commercialement disponibles et largement utilisées ne couvrent que 60% du dégradome humain et murin. La fin du séquençage de nombreux génomes a permis d'identifier le dégradome de plusieurs espèces comme celui de l'Homme, de la souris… La combinaison de ces informations avec les données de banques telles que MEROPS a conduit à l'annotation de 565 protéases et 157 inhibiteurs dans le génome humain [16]. Overall *et al.* ont conçu une puce ADN dédiée à la fois au dégradome humain et murin : la CLIP-CHIP™. Tous les inhibiteurs ou protéases connus sont spécifiquement représentés par des oligonucléotides uniques de 70 bases avec deux puces par lame [20] : au total 1561 protéases humaines et murines, homologues inactifs et inhibiteurs sont représentés. Cet outil présente un fort potentiel, par exemple dans l'étude des activités protéolytiques au cours de l'initiation ou de la progression de cancers.

b La dégradomique traductionnelle : CLIP-PSP (Proteolytic Signature Peptide)

Pour évaluer le potentiel protéolytique d'une cellule ou d'un tissu, il peut être intéressant d'évaluer la quantité absolue de protéases qu'il contient. Pour cela, la stratégie s'apparente à l'utilisation des peptides AQUA™ [21]. Après digestion *in vitro* par une enzyme, chaque protéine génère un ensemble de peptides protéotypiques spécifiques identifiables par MS. Un ou plusieurs de ces peptides sont alors synthétisés en remplaçant un acide aminé (a.a.) par le même, contenant des isotopes lourds stables (atomes ^{13}C ou ^{15}N). Ces peptides de synthèse peuvent alors être ajoutés en quantité connue au mélange de protéines à quantifier en tant que standard interne. L'ajout de ces peptides dès les premières étapes de l'étude (par exemple après la découpe des spots lorsqu'un gel est réalisé) permet de s'affranchir de tous les biais liés aux manipulations (extraction des peptides du gel, centrifugations sous vide…). Pour connaître le potentiel protéolytique d'une cellule ou d'un tissu, c'est la forme inactive de la protéase qu'il faut quantifier. En choisissant des peptides qui couvrent la zone de clivage du zymogène, la quantité absolue de pro-enzyme peut être déterminée.

c Le dégradome actif

Comme nous venons de le voir, la plupart des protéases sont synthétisées sous forme de proenzymes inactives et peuvent être soumises à l'action d'un inhibiteur. Ainsi, la quantité totale d'une enzyme n'est pas toujours reliée à sa quantité sous forme active ni à son activité. Pour **identifier, localiser et quantifier les protéases actives** d'un échantillon biologique, des marqueurs d'activité peuvent être utilisés. Il s'agit de molécules constituées d'un site réactif qui va former une liaison covalente avec l'enzyme cible et d'une étiquette pour la visualisation et/ou la purification du complexe créé (Figure 1). Le site de reconnaissance permet de cibler une enzyme selon sa classe et se lie au site actif de manière irréversible. L'étiquette peut être un fluorophore, contenir des isotopes radioactifs ou bien être une molécule de type biotine permettant une purification par la suite. L'utilisation de *tags* fluorescents ou radioactifs permet la visualisation des protéines sur un gel de séparation mono- ou bidimensionnel [22]. Les marqueurs d'activité enzymatique sont développés afin de réagir uniquement avec une protéase active et non avec son zymogène, ou avec la protéase inhibée. Ce type d'approche a été appliqué à différentes classes enzymatiques comme les protéases à serine ou à cystéine et les métalloprotéinases.

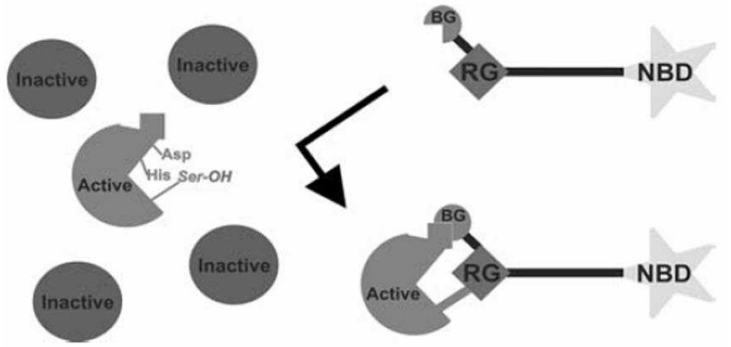


Figure 1 : Principe de la recherche protéomique d'activité enzymatique.
Ces marqueurs d'activité, qui contiennent un groupe réactif (RG) pour une liaison covalente avec un groupe nucléophile du site actif de l'enzyme, un groupe liant (BG) permettant la reconnaissance de l'enzyme et une étiquette pour la détection/isolation, capturent la protéine cible active même dans un protéome complexe. Figure d'après Schmidinger *et al.* [23].

d Le dégradome du point de vue des substrats

La tâche la plus difficile dans l'étude du potentiel protéolytique est la détermination des substrats d'une protéase. Pour ce faire, les approches impliquent en général la digestion *in vitro* des candidats possibles, ce qui introduit un biais dans la mesure où les substrats découverts peuvent ne pas être réels dans un contexte cellulaire. Pour mettre en évidence les activités *in vivo*, les chercheurs ont désormais recours à des approches de type quantification relative par spectrométrie de masse grâce aux marquages isotopiques de type ICAT ou iTRAQ.

II-B. Les techniques analytiques

II-B.1. La spectrométrie de masse

Afin d'étudier les profils protéiques exprimés dans différentes conditions physiologiques ou pathologiques, l'analyse par MS est couramment utilisée comme outil puissant sur les plateformes de protéomique. Cette technologie est appliquée pour la détermination de masses et peut être adaptée à l'identification des protéines. Un spectromètre de masse sépare les protéines (ou autres analytes) en fonction de leur rapport masse sur charge (m/z). Les molécules à analyser sont ionisées et les ions sont envoyés vers l'analyseur par un champ électrique qui sépare ces ions en fonction de leur rapport m/z. Le détecteur envoie alors l'information vers l'ordinateur pour être analysée. Les méthodes d'ionisation fréquemment utilisées en biologie sont l'électrospray (ESI) et la désorption-ionisation laser assistée par matrice (MALDI) car elles engendrent peu ou pas de fragmentation de la molécule durant le processus d'ionisation et/ou de désorption.

II-B.1.1. Les sources d'ionisation

a ESI

L'ionisation par électrospray a été développée par Fenn et ses collaborateurs, et décrite pour la première fois en 1985 [24]. L'ESI a été depuis très largement utilisée en protéomique (et pour d'autres applications) pour la caractérisation de biomolécules spécifiques. L'électrospray implique la production d'ions gazeux grâce à l'application d'un potentiel électrique à un liquide en mouvement, ce qui résulte en la formation d'un spray de petites gouttelettes, mélange de solvant et d'analytes. Le solvant est évaporé des gouttelettes par la chaleur ou par une autre forme d'énergie comme celle provenant de la collision avec un gaz (azote). La taille de la goutte diminue de plus en plus jusqu'à devenir instable (limite de Rayleigh), ce qui induit l'implosion en gouttelettes encore plus petites. Finalement les répulsions électrostatiques deviennent assez fortes pour provoquer la désorption des ions analytes, qui sont alors envoyés vers le spectromètre de masse (Figure 2). Les ions générés en ESI portent généralement plusieurs charges et sont de type $(M+nH)^{n+}$.

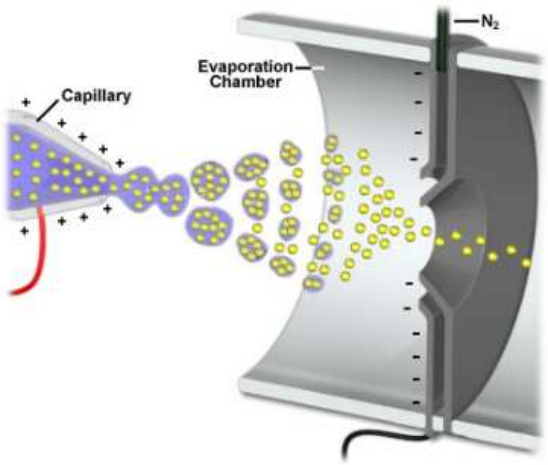


Figure 2 : Processus d'électrospray.
L'évaporation progressive du solvant contenu dans les gouttes permet la désorption de l'analyte et la formation d'ions multichargés.
(http://www.magnet.fsu.edu/education/tutorials/tools/ionization_esi.html).

Lors du processus d'ESI, des ions quasi-moléculaires sont créés par protonation ou déprotonation. Pour de larges biomolécules, des ions multichargés sont principalement observés et différents états de charge sont présents dans les spectres. Le nombre de charges dépend de la masse et des propriétés chimiques de la molécule, mais également du solvant utilisé. C'est pourquoi il y a souvent plusieurs ions pour un seul composé. Lorsque celui-ci est pur, le profil de ces multiples pics aide à déterminer de manière très précise sa masse. En revanche, pour des échantillons protéiques très complexes, ces multiples pics ne font qu'accroître la complexité des spectres de masse. C'est pourquoi une pré-

fractionnation, notamment par chromatographie liquide (HPLC), est souvent nécessaire pour les applications protéomiques utilisant l'ionisation par électrospray. Récemment, une ionisation électrospray par désorption de surface (source DESI, Figure 3) a été présentée par Cooks et ses collaborateurs [25, 26]. Ainsi l'ESI devient compatible avec l'analyse d'échantillons solides.

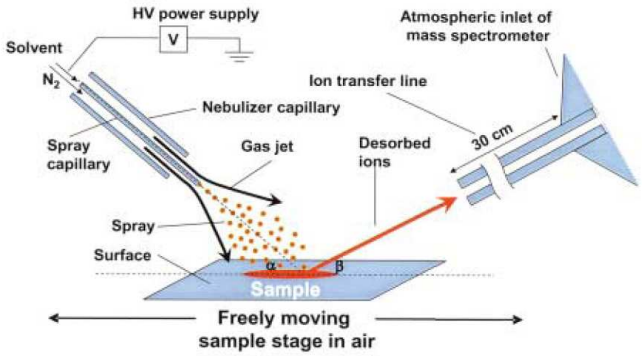


Figure 3 : Schéma d'une expérience DESI typique.
Un solvant est vaporisé sous l'influence d'une haute tension sur l'échantillon déposé sur une plaque. Figure d'après Takats *et al.* [25].

L'ESI est un processus d'ionisation douce où très peu de fragments sont observés. C'est pourquoi, il est bien adapté à l'analyse de biomolécules. Les avantages de l'électrospray incluent :

- Une haute reproductibilité puisque il n'y a pas de processus de cristallisation,
- Une haute flexibilité quant au type de spectromètre de masse en aval, du fait des rapports m/z plus faibles générés par la formation d'ions multichargés. Ceci rend possible le couplage avec les trappes ioniques et les quadripôles comme avec les cellules à résonance d'ion cyclotronique ou orbitalaire et les TOF.

En revanche, l'électrospray présente les inconvénients suivants :

- Des spectres de masses complexes dus aux multiples espèces ioniques pour un même analyte,
- Une consommation importante d'échantillon : ce problème a été relativement maîtrisé grâce à l'introduction des sources nanospray [27],
- L'électrospray ne peut pas être utilisé pour l'imagerie. En revanche, l'ionisation DESI peut l'être mais, pour l'instant, avec une résolution spatiale et une efficacité bien moindre que celles obtenues en MALDI.

b MALDI

En 1988, Hillenkamp et Karas découvrent que l'on peut produire des ions moléculaires de larges protéines par désorption laser sans fragmentation lorsque ces biomolécules sont mélangées avec une matrice [28]. Il s'agit d'une structure cristalline de petites molécules organiques chromophores. La longueur d'onde du laser utilisé pour l'ionisation doit être dans la gamme d'absorption des cristaux de matrice. Grâce à l'absorption des photons du laser, des petites molécules de matrice sont vaporisées et désorbent avec elles les plus grosses molécules de protéines. Aujourd'hui encore, l'ionisation MALDI n'est pas totalement bien décrite. Différents modèles ont été proposés :

- modèle d'ionisation photochimique [29],
- modèle d'ionisation par cluster [30],
- modèle par pseudo-transfert de proton [31].

Pour le **modèle d'ionisation photochimique**, l'ionisation multiphotonique suivie de collisions permet la production d'ions de matrice par photoionisation, protonation et déprotonation. Les ions biomoléculaires sont produits par protonation et déprotonation avec des ions de matrice et/ou d'autres biomolécules lors de processus de collisions en phase gazeuse. Pour le **modèle d'ionisation par cluster**, l'irradiation laser provoque la désorption de cluster d'ions qui produisent alors des ions biomoléculaires par expulsion des molécules de matrice. Pour le **modèle de pseudo transfert de proton**, les ions biomoléculaires sont produits pendant le processus de cristallisation, l'absorption de l'énergie laser permettant uniquement la désorption dans l'espace pour la détection par spectrométrie de masse. Une meilleure compréhension des mécanismes d'ionisation pourrait devenir essentielle à l'amélioration de la détection par spectrométrie de masse. Aujourd'hui, le MALDI est très efficace pour la détection de peptides : quelques dizaines d'attomoles à quelques femtomoles seulement sont nécessaires par spot [32]. En revanche, la détection de biomolécules plus grosses, en particulier de masse supérieure à 100 000Da devient plus difficile. Les protéines (ou les peptides) ionisées sont accélérées par un champ électrique et expulsées vers l'analyseur de masse. Il s'agit souvent d'un analyseur de type TOF. Accéléré par un potentiel constant, l'ion atteint le détecteur avec une vitesse qui dépend directement de son rapport m/z. Un schéma classique de spectromètre de masse MALDI-TOF est donné Figure 4. Selon le niveau de signal et la résolution nécessaire, les instrumentations TOF linéaire et réflectron sont largement utilisées pour les mesures de biomolécules.

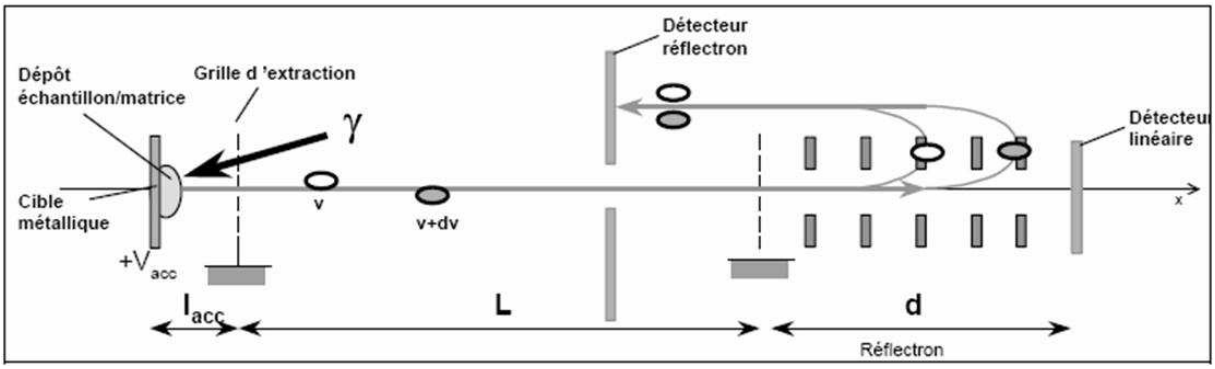


Figure 4 : Schéma de principe d'un MALDI-TOF en mode réflectron [33].

Typiquement, l'échantillon est dissous dans une solution puis la matrice est ajoutée. Elle consiste en un petit composé organique comme l'acide sinapinique (SA), l'acide α-cyano-4-hydroxy cinnamique (CHCA)… Ces deux solutions peuvent être mixées avant dépôt ou déposées successivement sur une cible métallique et laissées à sécher. Après cristallisation, la cible est placée dans le spectromètre de masse. En général, le rapport molaire matrice/biomolécule est supérieur à 1 000, ce en fonction de la masse de la biomolécule. La matrice SA est plus utilisée pour la détection des larges protéines et polypeptides, alors que la CHCA est au contraire utilisée pour la détection des peptides. Grâce au développement de l'extraction pulsée retardée, la résolution d'un MALDI-TOF linéaire peut atteindre 2 000. La résolution d'un MALDI-TOF réflectron peut, elle, atteindre 10 000 voire 20 000. Cependant il est difficile d'appliquer une détection réflectron pour des molécules de rapport m/z supérieur à 20 000Th.

Les avantages du MALDI-TOF incluent :

- la rapidité d'analyse : les lasers de fréquence 200Hz permettent de réaliser un spectre MS en quelques secondes (5 à 10 selon le nombre de tirs accumulés),
- sa gamme de masse théorique non limitée, s'étendant de peptides courts à de très grosses protéines (>100 000Da),
- des spectres de masse simples : la plupart des ions peptidiques détectés sont monochargés et seul un faible pourcentage est doublement chargé. Quelquefois, des ions trichargés sont observés,
- la possibilité de réaliser de l'imagerie moléculaire sur des tissus et donc de l'imagerie protéomique puisqu'on peut focaliser le rayonnement laser sur un très petit spot (moins de 10μm) selon la lentille de convergence utilisée et la divergence du laser [34],
- sa haute sensibilité : quelques attomoles pour des peptides courts.

En revanche, l'ionisation MALDI présente certains inconvénients :

- une faible reproductibilité due à la difficulté de contrôler le processus de cristallisation et la production d'ions (qui dépend de la fluence du laser, *i.e.* l'énergie reçue par unité de surface),
- une résolution faible pour les très grosses molécules due à la large distribution d'énergie cinétique des biomolécules pendant la désorption, à la formation éventuelle d'adduits de matrice et à la fragmentation métastable des molécules. Ceci est un inconvénient majeur pour l'identification directe de larges protéines,
- les ions de matrices interfèrent dans les basses masses et empêchent la détection de petites molécules. L'ionisation directe sur silice ou sur métal a été développée pour réduire ce problème [35, 36].

Cependant l'ionisation en mode MALDI reste un outil de choix pour la recherche rapide de biomarqueurs, notamment pour l'imagerie des larges biomolécules sans marquage préalable [37].

II-B.1.2. Les analyseurs

La demande en capacités analytiques puissantes dans le domaine de l'analyse biochimique des protéines étant de plus en plus importante, l'instrumentation en spectrométrie de masse a évolué très rapidement dans les dernières années. De nouveaux types d'analyseurs et des **instruments hybrides** ont fait leur apparition, fournissant ainsi de nouvelles opportunités dans l'étude des protéomes complexes [38, 39]. En particulier, des instruments permettant des **mesures de masse exacte** sont de plus en plus utilisés dans les études protéomiques, non seulement pour identifier des protéines mais aussi pour déterminer des états de modifications post-traductionnelles, ou encore des interactions entre des protéines et d'autres molécules de manière moins ambiguë qu'avant.

a Pourquoi a-t-on besoin de mesure de masse exacte ?

Certains défis significatifs en protéomique proviennent de la **grande complexité des systèmes biologiques et de la gamme de concentration des protéines dans ces systèmes**. L'exemple du protéome du plasma ou du sérum est souvent cité pour illustrer cette demande croissante en gamme dynamique. En effet, dans ces systèmes, toutes les protéines exprimées sont potentiellement présentes et couvrent une gamme de concentration d'au moins 10 ordres de grandeur, ce qui excède la gamme dynamique de tout instrument de spectrométrie de masse utilisé comme unique méthode analytique [40]. Après digestion enzymatique, chaque protéine est convertie en quelques dizaines de peptides, ce qui augmente encore la complexité du mélange. Si on ajoute à cela la présence de multiples formes d'une même protéine (isoformes, MPTs, formes tronquées par maturation), l'analyse protéomique pose de nombreux défis analytiques. Une **stratégie de fractionnement** peut être utilisée, comme par exemple le

couplage d'une séparation par chromatographie en phase liquide multidimensionnelle avec la MS [41, 42]. Ce type de couplage sera présenté plus en détail dans les chapitres suivants.

Quel que soit le degré de séparation, l'identification de peptides ou de protéines, soit par MS soit par MS/MS, dépend typiquement de la correspondance entre les masses des ions parents (ou les masses des ions fragments) et une banque de données dérivée des séquences des protéines d'un génome donné. Le niveau de confiance des identifications dépend fortement de l'exactitude des mesures de masse, surtout dans le cas d'échantillons très complexes provenant d'organismes développés (l'Homme par exemple). Il est bien connu que **le nombre de séquences d'acides aminés candidates pour une même masse décroît rapidement si la précision de mesure de masse augmente** [43-47]. Une précision de mesure de +/- 1ppm permet d'exclure 99% des peptides qui ont la même masse nominale mais une composition en éléments chimiques et/ou en acides aminés différente, ce qui résulte en une plus grande fiabilité des identifications des peptides [47]. Les trappes linéaires font partie des spectromètres de masse en tandem les plus populaires utilisés dans les études protéomiques notamment en raison de leur grande vitesse de balayage : elles sont capables d'acquérir des milliers de spectres MS/MS durant une séparation chromatographique d'une heure ; cependant, leur précision de mesure de masse est généralement faible [48]. Ainsi, une grande proportion de protéines peut être mal identifiée, selon les critères utilisés pour filtrer les données MS/MS lors de la recherche dans les banques de données [49-52]. Utiliser des filtres de score plus haut permet de diminuer significativement les taux de faux positifs mais entraîne également le risque de perdre une fraction des peptides identifiés, vrais positifs.

La spécificité des identifications peptidiques peut être significativement améliorée en utilisant de multiples étapes MS (MS^n) [53, 54], des techniques de fragmentation complémentaires (dissociation par capture d'électron ECD combinée à la dissociation induite par collision CID [55]) ou encore en faisant les mesures de masse avec une haute précision [43, 44, 47, 56, 57]. Bien que l'analyse MS/MS soit efficace dans l'identification de peptides et de protéines, le nombre de peptides détectables qui sont élués durant une analyse LC-MS/MS classique excède largement les capacités de fragmentation d'un spectromètre de masse en tandem pour réaliser des spectres CID sur chacun d'eux : **trop de peptides, trop peu de temps !** De plus une analyse complète du protéome demande souvent la collecte d'informations à grande échelle concernant les changements temporels dans l'expression protéique, ce qui implique des capacités d'analyse MS à haut débit pour une identification et une quantification sensibles et reproductibles des protéines d'échantillons globalement identiques. Ces besoins peuvent être traités en utilisant le couple **masse précise-temps de rétention précis** (*accurate mass and time*, AMT). Si la masse d'un peptide peut être mesurée avec une exactitude suffisante tout comme son temps de rétention, alors cette double étiquette est considérée comme unique et spécifique dans l'espace masse-

temps parmi tous les candidats possibles d'une banque de données, pré-établie par LC-MS/MS. Ainsi elle peut être utilisée comme une étiquette non ambiguë (appelée AMT) pour les identifications des peptides et des protéines à haut débit puisque l'étape de mesure MS/MS est dans ce cas contournée [58].

b La précision des mesures de masse

Une définition claire de la **précision de masse** et de **l'exactitude de mesure** en protéomique et de leur différence est nécessaire : les deux notions sont souvent confondues [59]. Dans la définition classique, la **justesse ou exactitude** est le degré de conformité de la quantité mesurée avec la vraie valeur. La **précision** détermine le degré avec lequel les quantités mesurées présentent les mêmes résultats ou tout du moins, des résultats similaires. En spectrométrie de masse biologique, l'un des objectifs est de déterminer précisément le rapport m/z des biomolécules d'intérêt, et d'obtenir ainsi leur masse exacte après déconvolution isotopique. Cependant, les spectromètres de masse ne mesurent pas directement le rapport m/z, mais d'autres paramètres comme la fréquence réduite cyclotronique (en spectrométrie de masse par résonance cyclotronique d'ions à transformée de Fourrier, FTICR MS) ou le temps de vol d'un ion pour atteindre le détecteur (TOF MS) : une procédure d'étalonnage est donc nécessaire pour convertir les quantités mesurées en rapports m/z. Puisque les paramètres expérimentalement mesurés sont souvent affectés par la complexité du système étudié, qui représente un ensemble de nombreuses particules, et en raison de la non idéalité des instruments, des routines de correction sophistiquées doivent généralement être utilisées dans des procédures d'étalonnage pour atténuer les imperfections expérimentales si une haute précision et une haute justesse doit être atteinte. Parmi les facteurs affectant le plus l'exactitude de la conversion des quantités expérimentalement mesurées en rapports m/z, on peut citer l'effet de charge spatiale, les effets de bord des champs, la dépendance du détecteur et du système d'acquisition à l'abondance des ions.

Les procédures d'**étalonnage** employées sur les instruments de spectrométrie de masse peuvent être séparées en deux catégories : externe et interne. L'**étalonnage externe** fixe certains coefficients durant toute la durée d'une expérience (ou d'une série d'expériences), soit des centaines de spectres. Ce type d'étalonnage compte sur la stabilité des paramètres instrumentaux et peut aboutir à des erreurs significatives si certains de ces paramètres sont affectés par l'environnement, comme par exemple la dérive de la température, des fluctuations de charge spatiale… L'**étalonnage interne** est basé sur le mélange d'un standard de rapport m/z connu avec les analytes et sur le calcul du rapport m/z de l'espèce inconnue à partir de l'équation d'étalonnage obtenue extemporanément grâce au standard. Bien que ce type d'étalonnage soit plus robuste aux variations des paramètres instrumentaux, certaines déviations expérimentales (par exemple un nombre excessif d'ions dans la cellule ICR) conduisent à des effets non linéaires qui réduisent la justesse de la mesure de masse.

Un haut **pouvoir de résolution** est nécessaire pour atteindre une précision suffisante et attribuer une masse exacte. Même si des spectromètres de masse de plus faible résolution peuvent atteindre une haute exactitude, leur application est limitée à l'analyse des composés cibles qui sont bien séparés dans le domaine m/z (utilisation de triple quadripôle pour une expérience de SRM, *Selected Reaction Monitoring*). Le besoin d'une instrumentation haute résolution est renforcé dans certaines expériences de protéomique et représente souvent un défi pour la détermination précise et exacte des masses de massifs isotopiques qui sont significativement différentes en termes d'abondance d'ions, mais qui se chevauchent dans le domaine des rapports m/z.

Les instrumentations de spectrométrie de masse capables d'atteindre des précisions de l'ordre de quelques ppm avec une haute résolution lors d'une expérience classique en protéomique se limitent à l'heure actuelle aux FTICR [60], aux TOF [61] et plus récemment, aux trappes tridimensionnelles électrostatiques (Orbitrap) [62].

c Spectromètre de masse de type FTICR

Le mouvement cyclotron a été utilisé pour la première fois en MS à la fin des années 40, avec l'introduction du premier spectromètre de masse à résonance d'ion cyclotronique, l'omégatron [63].

En pratique toute expérience de spectrométrie de masse par FTICR est basée sur le confinement des ions dans un **piège de Penning** (Figure 5), à savoir la combinaison d'un champ magnétique statique et spatialement uniforme **B** et d'un potentiel axial de piégeage électrostatique quadripolaire tridimensionnel [64]. Les pièges de Penning utilisent un champ magnétique homogène axial pour confiner les particules dans le plan médian, et un champ électrique quadripolaire pour les confiner selon l'axe de symétrie du piège. Le champ magnétique induit un mouvement des particules chargées en spirale, et le champ électrique permet d'éviter que les particules ne sortent du piège lorsqu'elles suivent les lignes de champ magnétique. La génération du champ magnétique uniforme est en général effectuée par un électro-aimant supraconducteur. Le champ créé est appliqué selon l'axe, ce qui oblige les particules chargées à parcourir des orbites circulaires, les maintenant ainsi à l'intérieur du piège.

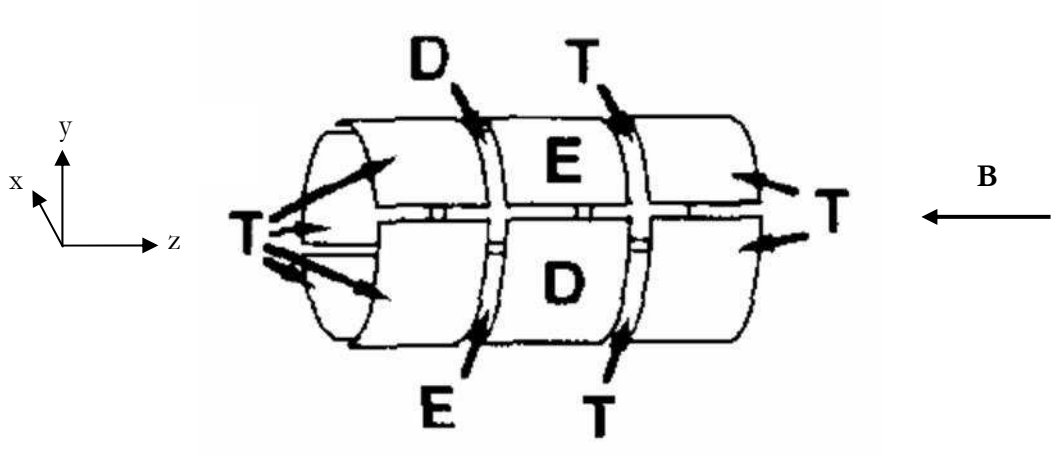


Figure 5 : Configuration d'une trappe d'ions à résonance cyclotronique ou piège de Penning. Elle est constituée de trois types d'électrodes. T : électrode de piégeage, E : électrode d'excitation, D : électrode de détection, B : champ magnétique. Figure d'après Guan *et al.* [64].

Les particules ainsi piégées subissent trois types d'oscillations :

- Une oscillation axiale,
- Deux oscillations dans le plan radial, similaires à celles d'une toupie en mouvement sur le sol :
  - une **oscillation magnétron** correspondant à l'orbite circulaire de la particule au sein du piège,
  - une **oscillation cyclotron** correspondant à la rotation de la particule « sur elle-même ».

De nombreux aspects fondamentaux de la FTICR peuvent être compris à partir de modèles très simples développés par la suite : le mouvement d'un ion dans un champ magnétique uniforme, sa cohérence spatiale par l'application d'un champ électrique radio-fréquence spatialement uniforme (excitation à la fréquence ICR pour rendre l'ion détectable) et son confinement par l'application d'un champ électrique continu tridimensionnel axial et quadripolaire [65].

Oscillation axiale :

Dans la mesure où les expériences sont réalisées dans une cellule, il est nécessaire d'y confiner les ions. Afin qu'ils ne s'échappent pas selon l'axe z, un potentiel électrostatique de faible amplitude V_{trap} est appliqué sur chaque électrode de piégeage. Ce potentiel a pour expression :

$$\varphi(x,y,z) = V_{trap}(\gamma + \frac{\alpha}{2a^2}(2z^2 - x^2 - y^2) \qquad \textbf{Equation 1}$$

où a est la longueur de la cellule, γ et α sont des constantes qui dépendent de la forme de la cellule [66].

La résolution de l'équation du mouvement selon l'axe z, $m\ddot{z} = -\frac{2qV_{trap}\alpha}{a^2}z$ nous montre que l'ion oscille suivant l'axe z selon :

$$z(t) = z_0 \cos(2\pi \nu_z t) \qquad \textbf{Equation 2}$$

Où $\nu_z = \frac{1}{2\pi}\sqrt{\frac{2qV_{trap}\alpha}{ma^2}}$ est la fréquence axiale et z_0 est une constante déterminée par les conditions initiales.

Rotation cyclotronique :

Un ion dans un champ magnétique vectoriel **B** uniforme est soumis à la force de Lorentz donnée par l'Equation 3 suivante:

$$\mathbf{F} = q\mathbf{v} \wedge \mathbf{B} = m\mathbf{a} \qquad \textbf{Equation 3}$$

Où m, q, **v** et **a** sont respectivement la masse de l'ion, sa charge, sa vitesse vectorielle et son accélération vectorielle. L'Equation 3 indique que la force de Lorentz est perpendiculaire au plan défini par **v** et **B**. Si l'ion évolue à vitesse constante (sans collision), alors le champ magnétique courbe la trajectoire de l'ion suivant un cercle (Figure 6) de rayon r donné par l'Equation 4.

$$r = \frac{mv_0}{qB_0} \qquad \textbf{Equation 4}$$

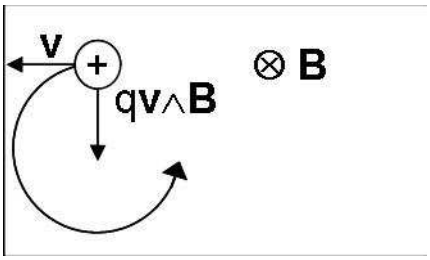


Figure 6 : Mouvement d'ion cyclotronique.
La trajectoire d'un ion se déplaçant dans le plan de la feuille est courbée suivant un cercle par la force de Lorentz produite par un champ magnétique perpendiculaire au plan de la feuille. Schéma dans le cas d'un ion chargé positivement. Pour un ion chargé négativement, la trajectoire se fait dans l'autre sens.

Si la vitesse des ions est faible et le champ intense, le rayon de courbure de la trajectoire devient petit.

La fréquence de rotation de l'ion ou fréquence cyclotronique peut alors s'exprimer ainsi :

$$\nu_c = \frac{eB_0}{2\pi m/z}$$ **Equation 5**

Avec ν_c la fréquence en Hz, B_0 le champ en tesla, m la masse en kg, z le nombre de charge de l'ion et -e la charge de l'électron (e=1,6.10^{-19}C). Si m est exprimé en Da alors $\nu_c = \frac{\lambda B_0}{m/z}$ où λ est une constante de valeur 1.535611 x 10^7 C.Da.s^{-1}. A champ uniforme et statique, pour tous les ions de même rapport m/z, cette fréquence est donc la même indépendamment de la vitesse initiale.

Rotation magnétron :

Cette rotation est due à la combinaison du champ **B** et de la composante radiale du potentiel électrostatique de piégeage. En effet les équations du mouvement sur les axes x et y s'écrivent :

$$m\frac{d^2x}{dt^2} = \frac{qV_{trap}\alpha}{a^2}x - qB_0\frac{dy}{dt}$$
$$m\frac{d^2y}{dt^2} = \frac{qV_{trap}\alpha}{a^2}y + qB_0\frac{dx}{dt}$$ **Equation 6**

Si on fait intervenir $\bar{r} = x + iy$, $\omega_z = 2\pi\nu_z = \sqrt{\frac{2qV_{trap}\alpha}{ma^2}}$, $\omega_c = 2\pi\nu_c = \frac{qB_0}{m}$ l'équation du mouvement en coordonnée radiale s'écrit :

$$\frac{d^2\bar{r}}{dt^2} = \frac{\omega_z^2}{2}\bar{r} + i\omega_c\frac{d\bar{r}}{dt}$$ **Equation 7**

Pour obtenir les pulsations propres, il suffit de résoudre l'équation du second degré en ω,

$$\omega^2 - \omega_c\omega + \frac{\omega_z^2}{2} = 0$$ **Equation 8**

.

On obtient alors :

La pulsation cyclotron réduite $\omega_+ = \frac{\omega_c}{2} + \sqrt{(\frac{\omega_c}{2})^2 - \frac{\omega_z^2}{2}}$

La pulsation magnétron $\omega_- = \frac{\omega_c}{2} - \sqrt{(\frac{\omega_c}{2})^2 - \frac{\omega_z^2}{2}}$ **Equation 9**

Au final, le mouvement des ions (Figure 7) est la superposition des trois modes (rotation cyclotron, rotation magnétron, oscillation de piégeage). La fréquence de chaque mode n'est fonction que de B_0, du

potentiel de piégeage V_{trap} et du rapport m/z. Les fréquences de piégeage et magnétron sont généralement beaucoup plus faibles que la fréquence cyclotron et ne sont donc pas détectées.

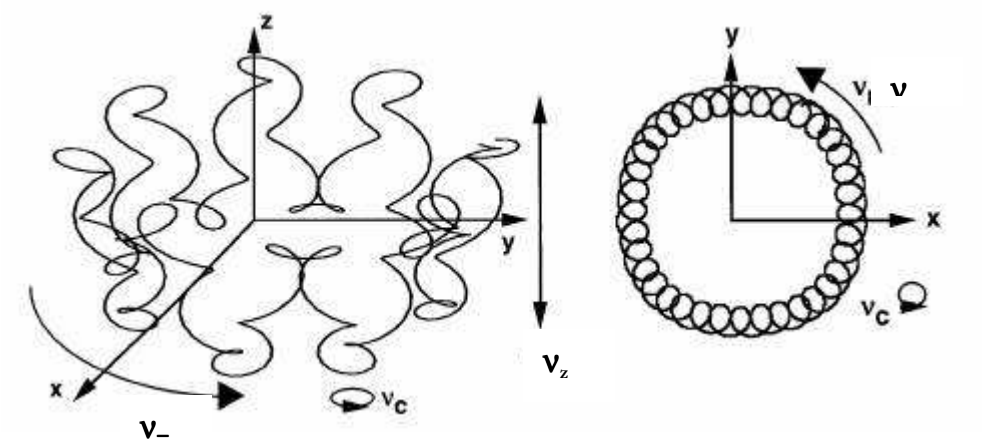


Figure 7 : Mouvement des ions dans une trappe de Penning. Figure d'après Marshall *et al.* [65].

Excitation et détection des ions

Pour être détectable, les ions sont accélérés de manière cohérente afin d'augmenter le rayon de leur trajectoire (Figure 7). Pour cela ils sont excités en appliquant un champ électrique uniforme oscillant à la fréquence cyclotronique particulière de chaque ion et perpendiculaire à **B**. Le rayon après excitation est donné par l'Equation 10 :

$$r = \frac{E_0 T_{excitation}}{2B_0}$$ **Equation 10**

Où E_0 est l'amplitude du champ électrique appliqué et $T_{excitation}$ la durée de l'excitation. On peut remarquer que ce rayon ne dépend pas du rapport m/z. Tous les ions sont donc excités sur la même trajectoire : il n'y a donc aucune discrimination induite par la masse des ions lors de la détection.

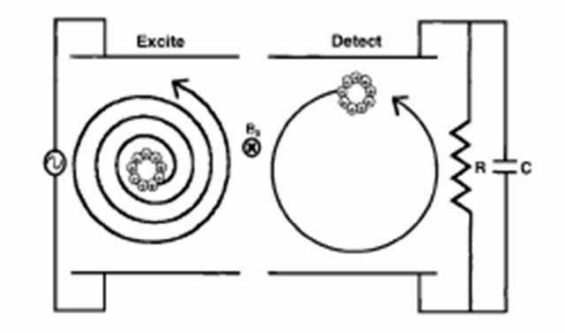


Figure 7 : La trajectoire cyclotron orbitale des ions est transformée d'un mouvement incohérent (à gauche) à un mouvement cohérent (à droite) par l'application d'un champ électrique tournant dans le même sens et à la même fréquence cyclotronique que les ions de rapport m/z. Figure d'après Marshall *et al.* [65].

d Spectromètre de masse de type Orbitrap

Depuis sa commercialisation en 2005, l'Orbitrap est devenu l'un des instruments de choix dans de nombreuses applications protéomiques. Le piégeage orbital fut mis en place pour la première fois en 1923 par Kingdon pour capturer les ions en les faisant tourner dans l'orbite d'une électrode centrale [67]. En 2000, Makarov a décrit l'Orbitrap comme un nouvel analyseur de spectrométrie de masse dérivé du piégeage orbital [62].

Géométrie de l'Orbitrap

Le piège consiste en une électrode externe semblable à un baril et en une électrode centrale de forme fuselée le long de l'axe (Figure 8). Les électrodes axiales symétriques de l'Orbitrap génèrent un potentiel électrostatique « quadralogarithmique».

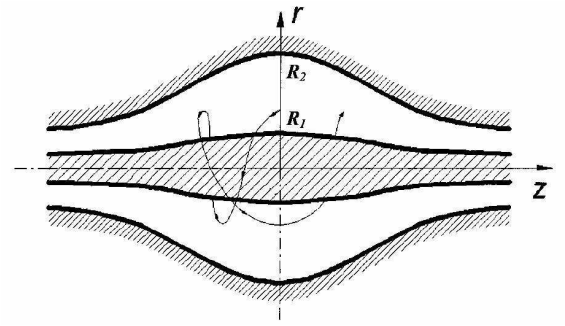


Figure 8 : Géométrie de l'Orbitrap.
Equipotentielles du champ quadralogarithmique ; exemple de trajectoire d'ion stable [62].

Trajectoire des ions

Le piégeage orbital des ions se fait dans un champ électrostatique donné par la distribution de potentiel suivante :

$$U(r,z) = \frac{k}{2}(z^2 - \frac{r^2}{2}) + \frac{k}{2}(R_m)^2 \ln(\frac{r}{R_m}) + C \qquad \textbf{Equation 11}$$

Où C est une constante, k la courbure du champ et R_m le rayon caractéristique. Il s'agit de la superposition d'un champ quadripolaire dû à la trappe ionique et d'un champ logarithmique dû au condensateur cylindrique. Dans ce champ, les trajectoires stables couplent la rotation autour de l'électrode centrale avec des oscillations le long de l'axe ce qui résulte en une spirale complexe.

D'après les équations du mouvement écrites en coordonnées polaires de $m\mathbf{a} = -q\nabla(U(r,z))$, on peut déduire que le mouvement selon l'axe z est indépendant du mouvement dans le plan (r, φ) (Equation 12).

$$\ddot{r} - r\dot{\varphi}^2 = -\frac{q}{m}\frac{k}{2}\left(\frac{R_m^2}{r} - r\right)$$

$$\frac{d}{dt}(r^2\dot{\varphi}) = 0 \qquad \textbf{Equation 12}$$

$$\ddot{z} = -\frac{q}{m}kz$$

Selon l'axe z, le mouvement est une simple oscillation de pulsation $\omega_z = \sqrt{k\frac{q}{m}}$.

Cette oscillation suit donc une fréquence qui est inversement proportionnelle à la racine carrée du rapport m/z de l'ion analysé. Ce rapport m/z peut ainsi être calculé par transformée de Fourier inverse à partir de l'enregistrement du signal.

II-B.1.3. Fragmentation spécifique : Dissociation induite par collision

Si la spectrométrie de masse ne permet pas ou peu la détermination de la structure moléculaire d'une biomolécule, elle est l'outil principal pour identifier une protéine. Afin d'obtenir la séquence en acides aminés, il est nécessaire d'obtenir les profils de fragmentation des peptides et/ou des protéines. Différentes méthodes ont été développées pour obtenir ces profils. Nous nous contenterons dans la suite de rappeler les principes de fragmentation des techniques utilisées au cours de cette thèse : les différentes formes de dissociation induite par collision (CID).

La fragmentation CID est la méthode de fragmentation des peptides la plus couramment utilisée [68]. La fragmentation a lieu lorsque les peptides entrent en collision avec des molécules de gaz neutre (Argon, Helium, Neon). Quand l'énergie de collision est assez haute pour convertir une partie de l'énergie cinétique en énergie vibrationnelle, la fragmentation de l'ion sélectionné peut être obtenue. Puisque la dissociation intervient lorsque l'énergie de collision est supérieure à l'énergie d'activation, on parle aussi de dissociation activée par collision (CAD). La fragmentation à basse énergie (10-100eV) intervient dans les quadripôles et les trappes ioniques, ce qui résulte en pertes de neutres et en ruptures des liaisons peptidiques [69]. La fragmentation à haute énergie (quelques keV) avec les configurations de type MALDI-TOF/TOF engendre la rupture d'autres liaisons du squelette peptidique et des chaînes latérales des acides aminés [70]. Lorsque le squelette peptidique est rompu, la charge peut être retenue du

côté C- ou N-terminal du site de coupure. La nomenclature des ions fragments a été proposée par Roepstorff *et al.* [71] puis modifiée par Biemann [72]. Le mécanisme CID à basse énergie est assez compliqué et on suppose que la fragmentation est liée à la délocalisation des charges [73, 74].

II-B.2. Les séparations chromatographiques

II-B.2.1. Séparation par chromatographie à polarité de phases inversée

La complexité des échantillons et leur gamme dynamique de concentrations constituent un challenge important dans les analyses protéomiques. Or la MS seule ne peut tolérer qu'une complexité relativement faible, présente une gamme dynamique limitée par spectre et est très sensible au phénomène de suppression spectrale. C'est pourquoi, en MS, **il est nécessaire de limiter le nombre de composés à analyser** en même temps.

La chromatographie en phase liquide est une méthode de séparation des composés d'un échantillon entre deux phases, une phase solide stationnaire et une phase liquide mobile, afin de pouvoir les identifier et les quantifier. Or les diverses propriétés physico-chimiques des peptides (charge, pI, hydrophobie, taille) leur permettent d'être séparés par presque tous les modes de séparation en phase liquide. La capacité de l'HPLC à séparer des biomolécules fragiles de hauts poids moléculaires rend son **application incontournable dans le domaine de la protéomique**. Grâce aux avancées des méthodes d'ionisation et de l'instrumentation, son couplage à la MS en fait une technologie puissante pour la caractérisation et l'identification des peptides et des protéines dans les mélanges complexes [75].

Le pouvoir de séparation par chromatographie à polarité de phases inversée (RPLC) et la très bonne compatibilité des solvants utilisés avec la spectrométrie de masse a conduit à son usage généralisé pour la séparation des peptides depuis le milieu des années 80. Grâce aux efforts de miniaturisation des colonnes et à l'introduction des techniques d'ionisation douce, la RPLC couplée à la MS est devenue la technique analytique principale dans le domaine de la protéomique. A partir d'une quantité d'échantillon limitée, des centaines de peptides peuvent être identifiés et quantifiés en routine avec une haute sensibilité.

L'électrospray reste la technique d'ionisation le plus souvent couplée à la LC en routine. Cependant, lors de l'analyse de mélange complexe, l'échelle de temps imposée par ce couplage en ligne peut limiter la quantité de données rassemblées. En effet, l'analyse alterne un *scan* MS sur l'ensemble de la gamme de masse avec des scans MS/MS des précurseurs les plus abondants. Plus la résolution de la séparation LC est grande, plus les pics sont fins. Le temps pour analyser les composants d'un pic est donc de plus en plus court : si la MS/MS dure trop longtemps, l'élution d'un pic peut être terminée au prochain *scan* MS

et tous ses composants n'auront pas pu être fragmentés. En revanche, si l'éluent LC est déposé sur une plaque MALDI, de tels problèmes ne se posent plus. En effet, l'ensemble des spots déposés peut être analysé plusieurs fois en changeant les paramètres d'acquisition. Depuis quelques années, l'apparition sur le marché de robots de dépôt comme le Probot (Dionex, Pays-Bas) permet d'automatiser les dépôts, ce qui est plus que nécessaire lors des séparations à nanodébit. D'autre part, par l'usage d'un *split* en sortie de colonne, il est possible de déposer une partie de l'éluent sur plaque, pendant que l'autre partie est analysée en ligne par électrospray. Bodnar *et al.* ont employé ce type de couplage pour évaluer la couverture de protéome et la complémentarité des deux approches [76].

II-B.2.2. Capacité de pic

Pour évaluer la qualité d'une séparation chromatographique, la capacité de pic (P_c) est une mesure simple et intuitive, applicable à toutes les séparations, même par gradient, ce qui n'est pas le cas pour la mesure du nombre de plateaux par exemple. Ce concept a d'abord été décrit par Giddings [77] puis rapidement appliqué par Horváth à la théorie des gradients [78]. Il s'agit d'une mesure du pouvoir de séparation en incluant l'ensemble de l'espace chromatographique tout en prenant en compte la variabilité des largeurs de pics. Le concept de la capacité de pic est encore plus important pour les séparations bidimensionnelles où l'usage de gradients est quasiment indispensable.

La capacité de pic mesure le **nombre de pics qui peuvent être séparés dans une fenêtre de temps avec une résolution donnée**. On considère généralement l'espace entre chaque pic égal à 4 fois la valeur de l'écart-type, ce qui permet de calculer la capacité de pic selon la formule $P_c = 1 + \frac{t_2 - t_1}{4\sigma}$. Le plus souvent, la capacité de pic est définie sur l'ensemble du chromatogramme : Dolan et Snyder définissent ainsi la capacité de pic P_c^{**} pour un échantillon donné entre le temps de rétention du premier pic d'intérêt et le temps de rétention du dernier [79].

Si la largeur des pics sur l'ensemble du chromatogramme ne change pas beaucoup comme c'est le cas dans la plupart des séparations par gradient sur colonne à polarité de phases inversée, et surtout dans les séparations de peptides, on peut calculer la capacité de pic par $P_c = 1 + \frac{t_g}{\overline{\omega}}$ où t_g est la durée du gradient et $\overline{\omega}$ la largeur de pic moyenne [80]. La mesure de la capacité de pic P_c^{**} d'un échantillon donné peut être plus informative surtout si une partie non négligeable de l'espace temporel n'est pas utilisée : $P_c^{**} = \frac{t_f - t_1}{\overline{\omega}}$, où t_1 est le temps de rétention de la première espèce éluée et t_f celui de la dernière.

II-B.2.3. Augmenter la capacité de pic en RPLC

Il est généralement accepté qu'aucune séparation chromatographique monodimensionnelle n'est capable de résoudre un mélange complexe de peptides provenant de la digestion d'un protéome complexe. Une séparation RPLC classique permet d'obtenir typiquement une capacité de pic de quelques centaines [81]. Cependant, au fil des ans, des efforts significatifs ont été faits pour augmenter le pouvoir de résolution des séparations monodimensionnelles [82], et ainsi augmenter le nombre d'analytes identifiés dans des échantillons toujours plus complexes. La capacité de pic peut notamment être améliorée en augmentant la **longueur** des colonnes ou en réduisant la **granulométrie** des particules. Cependant, dans les deux cas, la limitation provient de la pression engendrée, d'où le développement de chaînes chromatographiques pouvant résister à des **pressions** de quelques milliers de bars (Chromatographie Liquide à Ultra Haute Pression UHPLC).

Récemment, divers exemples ont été publiés rapportant des séparations efficaces de peptides sur de longues colonnes mais avec des instrumentations HPLC classiques (pression <600 bars). Pour cela, l'**élévation de la température** est exploitée afin de réduire la viscosité de la phase mobile et par conséquent la pression générée. Ainsi Wang et Carr ont présenté à température ambiante des capacités de pic comparables à l'UHPLC mais avec une longueur de colonne de 60cm obtenue en associant en série plusieurs colonnes de type 15cm x 2,1mm d.i. [83]. L'élévation de température de 25°C à 70°C leur a permis de réduire le temps d'analyse de 4 à 2h sans changer la capacité de pic.

Une autre façon d'augmenter la capacité de pic est de jouer sur la **pente du gradient**. En général, des gradients plus longs génèrent des P_c plus grandes bien que celles-ci atteignent un plateau. Ainsi Gilar *et al.* ont montré une augmentation de P_c de 157 à 351 en augmentant la durée du gradient de 25 à 100min (colonne C18 15cm x 4,6mm d.i., 5µm) [81]. Avec un système de type UHPLC, Shen *et al.* ont rapporté une augmentation du nombre de peptides et de protéines identifiés grâce à des gradients plus longs [84]. Cependant il faut noter que si le nombre de peptides et de protéines identifiés double en multipliant la durée de séparation par 3 (200 à 600min), la capacité de pic, elle, ne change pas beaucoup (420 à 450) car la largeur des pics augmente également.

Enfin, l'utilisation de **colonnes monolithiques** apparaît très prometteuse pour améliorer la séparation des peptides. Il s'agit de colonnes, consistant en un support de séparation en un seul bloc au lieu d'un ensemble de billes de phase. Elles sont caractérisées par une grande perméabilité et une faible résistance aux transferts de charge. Ainsi le groupe de Huber a montré qu'une colonne monolithique en poly-(styrène-divinylbenzène) (PS-DVB 6cm × 100µm d.i.) utilisée au débit de 500nL/min permettait

d'obtenir de meilleurs résultats en termes d'identification qu'une colonne de billes de silice (6.5cm x 75µm d.i.) utilisée au débit de 200nL/min [85]. La répétabilité des analyses apparaît également meilleure.

II-B.2.4. Augmenter la capacité de pic par couplage multidimensionnel

a La règle d'or

Le couplage bidimensionnel en chromatographie liquide (LC2D) fait référence à une technique analytique où deux systèmes de séparation en phase liquide sont couplés et appliqués à l'échantillon. Giddings [86] puis Guiochon [87] ont fait remarquer que l'intérêt d'un tel couplage réside dans le fait que dans certaines conditions **la capacité de pic globale n'est pas la somme mais le produit des capacités de pic** de chaque dimension : $P_c^{2D} = P_c^1 \times P_c^2$ comme illustré sur la Figure 9.

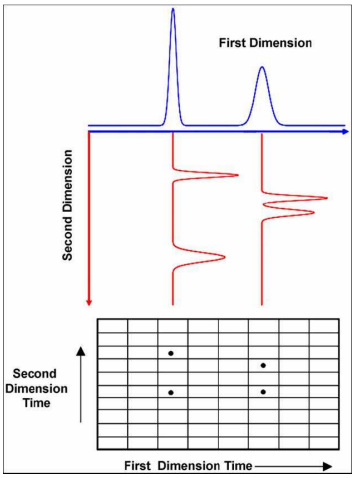


Figure 9 : Illustration de la capacité de pic d'une séparation bidimensionnelle et de sa dépendance en produit des capacités de pic des 1ère et 2ème dimension [88].

Cependant, pour que cela soit vérifié, **deux conditions** doivent être respectées. Tout d'abord, il ne faut pas que des pics séparés par la première dimension soient mélangés lors du transfert vers la deuxième dimension. Cela peut arriver lorsque la **période d'échantillonnage** des fractions de la première dimension n'est pas ajustée correctement (Figure 10). D'autre part, les mécanismes de rétention impliqués dans les deux dimensions successives doivent être **orthogonaux, *i.e.* indépendants** [86]. S'ils sont liés, la rétention des composés suivant les deux dimensions sera corrélée et donc l'espace de séparation, dans lequel les pics sont dispersés, plus petit que celui prévu lorsqu'ils le sont de manière aléatoire.

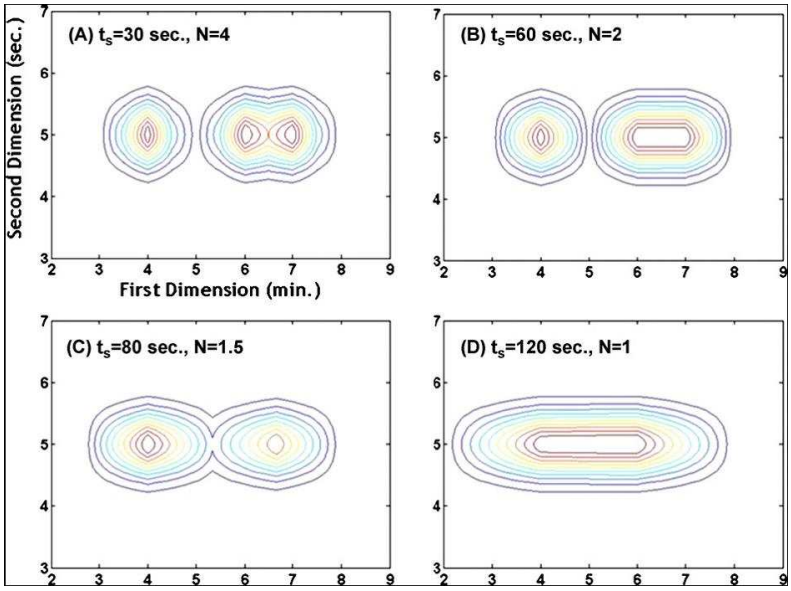


Figure 10 : Illustration de l'effet du temps de collecte (t_s) sur la capacité de pic de la 1ère dimension. n représente le nombre de fractions collectées sur la largeur 4σ d'un pic. Plus la fréquence d'échantillonnage diminue, plus la résolution des pics observés diminue jusqu'à la formation d'un seul et même pic contenant ici 3 espèces [88].

b Orthogonalité des deux dimensions

Alors que le couplage LC2D est de plus en plus utilisé dans l'analyse de peptides par LC-MS, très peu d'études ont été réalisées sur **l'orthogonalité des modes de séparation** choisis. Récemment Gilar *et al.* ont comparé plusieurs séparations monodimensionnelles utilisant différentes phases stationnaires (RP, exclusion stérique (SEC), échange de cations (SCX), hydrophile (HILIC)) avec différentes phases mobiles [89]. Pour cela, ils ont étudié un sous-ensemble de peptides d'un échantillon de protéines standard digérées. Pour chaque séparation monodimensionnelle réalisée, ils ont calculé le temps de rétention normalisé de chaque peptide (par rapport aux temps de rétention minimum et maximum de chaque chromatogramme). Pour chaque couplage envisagé, ils ont représenté le temps de rétention de la deuxième dimension hypothétique en fonction de la première dimension hypothétique. Selon le pourcentage d'espace couvert par les différents points, les auteurs évaluent l'orthogonalité des deux modes choisis, sachant qu'une surface remplie de façon aléatoire ne couvre que 63% de l'espace des temps de rétention mais correspond à une orthogonalité de 100%. La conclusion de leur étude montre que les séparations HILIC x RP sont les plus orthogonales. Cependant si on prend en compte les capacités de pic de chaque mode et qu'on calcule la capacité effective des couplages (*i.e.* P_c^1 x P_c^2 x % surface couverte), le couplage RP(pH=10) x RP (pH=2,6) permet d'obtenir la meilleure capacité de pic effective. De plus les séparations RP sont des séparations robustes et bien maîtrisées par les opérateurs, il est donc probable que les couplages RP x RP seront à l'avenir de plus en plus développés.

c Couplage SCX x RP

Même si les modes de séparation SCX et RP ne sont pas complètement orthogonaux notamment du fait que les peptides tryptiques sont principalement 2 fois et 3 fois chargés, le couplage SCX x RP reste le mode le plus largement utilisé en protéomique. En effet, les colonnes d'échange d'ions offrent généralement une capacité de chargement importante, la RPLC est parfaitement compatible avec la masse et les deux modes présentent une capacité de pic unidimensionnelle importante.

Il existe principalement deux approches pour coupler ces deux modes : *off-line* ou *on-line*.

Couplage on-line

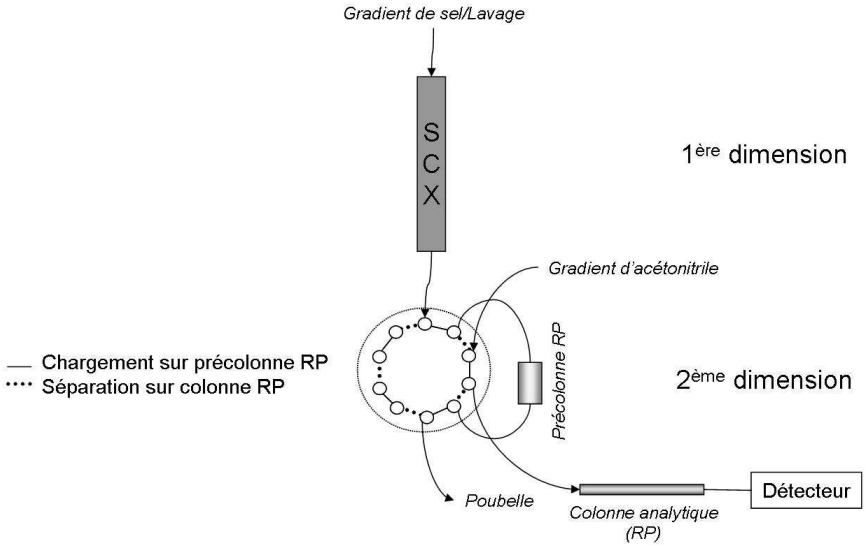


Figure 11 : Système de chromatographie multidimensionnelle *on-line*.
Les peptides sont élués de la colonne SCX en fonction de leur charge et retenus sur la précolonne RP. Dans cet exemple, le débit du gradient de sel doit être coupé après chaque fraction collectée, pendant l'élution et la séparation de cette fraction sur la colonne RP par un gradient d' acétonitrile.

Dans ce cas, l'effluent de la première dimension est directement élué vers la deuxième dimension sans collecter de fraction explicitement (Figure 11). Ce type de séparation est rapide et les pertes d'échantillons sont minimisées. Il est possible d'utiliser deux colonnes séparées et connectées par un jeu de vannes ou bien une seule colonne remplie à la fois des billes de phase SCX et RP.

Dans le premier cas, les deux colonnes sont connectées par au moins deux vannes. Les peptides sont d'abord chargés sur la colonne SCX en condition acide pour que les peptides soient tous chargés positivement. La concentration en sel est alors augmentée pour éluer les peptides en fonction de leur charge vers la précolonne RP. Pendant que les peptides sont dessalés sur cette précolonne, le débit sur la colonne SCX est arrêté. Après dessalage, les peptides sont élués vers la colonne analytique RP pour être séparés en fonction de leur hydrophobie. Cette approche a été utilisée avec succès par Tyan *et al.* dans l'étude de fluide pleural [90]. A partir de 25µg d'échantillon protéique, les auteurs ont pu identifier

1 800 peptides différents correspondant à 1 415 protéines. Cependant il faut noter que seules 124 protéines ont été identifiées avec au moins deux peptides uniques. D'autre part, ces identifications correspondent à la combinaison des données des trois réplicats et non à un consensus.

Dans le deuxième cas, la colonne contenant les deux types de phase, le montage ne nécessite aucune vanne. Le mélange de peptides est chargé sur la colonne biphasique qui est interfacée avec un système de pompe HPLC quaternaire. L'échantillon est soumis à une série d'impulsions salines ce qui permet d'éluer seulement une partie des peptides de la partie SCX vers la partie RP de la colonne. Après chaque impulsion, un gradient d'acétonitrile (ACN) est réalisé pour éluer les peptides vers le spectromètre de masse. Cette approche a été utilisée avec succès par Kislinger *et al.* dans l'étude de la localisation subcellulaire des protéines de 4 organelles de 6 organes chez la souris [91]. A partir de 100µg de protéines de chaque fraction sub-cellulaire, 4 768 protéines ont été identifiées avec au moins un peptide spécifique. Il s'agit des résultats correspondant à l'ensemble des fractions analysées plusieurs fois en LC2D (entre 7 et 9 fois). Chaque analyse LC2D dure environ 20h.

Couplage off-line

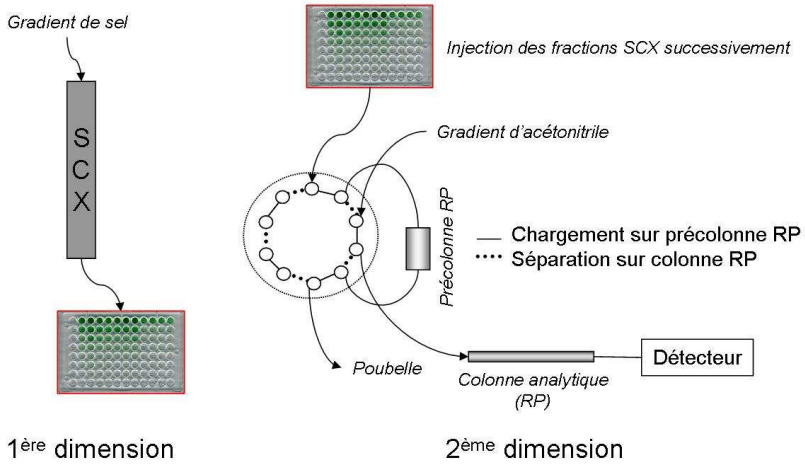


Figure 12 : Système de chromatographie multidimensionnelle *off-line*.
Les peptides sont élués de la colonne SCX par un gradient de sel en fonction de leur charge et différentes fractions sont collectées. Chaque fraction est ensuite réinjectée vers la 2ème dimension pour une séparation classique RPLC.

Lors d'une approche *off-line*, les peptides élués de la colonne de la première dimension sont collectés puis soumis plus tard à une séparation RPLC-MS/MS classique (Figure 12). A priori, ce type de couplage permet d'optimiser chaque dimension. Notamment, il est rapporté dans la littérature que les phases stationnaires SCX présentent des interactions secondaires hydrophobiques [92], qui peuvent réduire la récupération des peptides hydrophobes. Pour améliorer leur élution et la forme des pics, il est donc suggéré d'ajouter un solvant organique aux tampons de première dimension : généralement 25% d'ACN. Or en configuration *on-line*, ce fort pourcentage d'ACN est rédhibitoire pour la fixation des

peptides sur la phase RP C18. En revanche en configuration *off-line*, les deux dimensions étant découplées, il suffit de mettre à sec les fractions et de les resuspendre dans un solvant adéquat à la RPLC. Un avantage non négligeable de l'utilisation d'une configuration *off-line* réside également dans le fait qu'il est possible de travailler avec de véritables gradients salins plutôt qu'avec des impulsions, ce qui permet de ne pas rompre les équilibres chromatographiques et d'obtenir une élution continue des peptides. D'autre part, le découplage des deux dimensions permet de traiter les fractions de différentes manières : séchage, traitement chimique, mesure de la concentration en peptides des fractions pour adapter l'injection RP… Il est également possible d'analyser les fractions plusieurs fois : par exemple Chen *et al.* mélangent certaines fractions entre elles car elles contiennent peu de peptides, mais, en parallèle, en analysent certaines en duplicat en n'injectant que la moitié des fractions en RPLC [93]. Cependant le principal inconvénient de telles approches provient du temps pris pour l'optimisation des deux séparations. Finalement, la manipulation des fractions entre les deux séparations peut générer des pertes de matériel non négligeables. Wang *et al.* ont adopté l'approche *off-line* pour analyser le protéome de cerveau de souris à partir d'un digestat de 11mg de peptides. Séparé sur colonne SCX *narrow-bore*, cet extrait a permis d'analyser 60 fractions : 30 correspondant à l'extrait de protéines digérées et 30 à l'enrichissement en peptides cystéinylés. Au total, 48 328 peptides différents ont été validés permettant l'identification de 7792 protéines avec au moins un peptide spécifique [94].

II-B.3. L'interprétation et la validation des données

Aujourd'hui, la majorité des données protéomiques sont associées aux analyses en MS et en particulier *via* l'utilisation de spectromètres de masse en tandem dont les performances s'améliorent continuellement [39]. Ces instruments peuvent générer des centaines voire des milliers de spectres de fragmentation par heure d'acquisition. L'attribution d'une séquence peptidique à ces spectres, la déduction des protéines représentées par les peptides identifiés et la détermination de leur abondance dans l'échantillon étudié présentent des défis importants de type informatique et statistique [95].

L'attribution correcte d'une séquence peptidique à un spectre de fragmentation expérimental est la première étape, et l'une des plus cruciales, du traitement des données en protéomique. De nombreuses approches informatiques et de nombreux outils logiciels ont été développés pour attribuer automatiquement les séquences peptidiques aux spectres de fragmentation. Ces logiciels peuvent être classés en trois catégories [96] :

- Les recherches dans les banques : les peptides sont identifiés en corrélant les spectres acquis avec les spectres théoriques prédits pour chaque peptide contenu dans la banque

de séquence protéique ou en corrélant les spectres acquis avec des librairies de spectres MS/MS expérimentaux préalablement identifiés,

- Le séquençage *de novo* où la séquence peptidique est explicitement lue à partir du spectre de fragmentation,
- Des approches hybrides, comme celles basées sur l'extraction d'une information de séquences de 3 à 5 résidus (peptide sequence tag), suivie d'une recherche dans les banques en *error tolerant*.

Dans le cadre d'étude protéomique à grande échelle, la première stratégie reste la plus fréquemment utilisée.

II-B.3.1. Les identifications par recherche dans les banques

Différents programmes de recherche dans les banques MS/MS ont été développés et sont rappelés dans le Tableau 1.

Tableau 1 : Outils disponibles pour l'interprétation des spectres MS/MS (d'après [95]).

Programme	Référence	Site web
Outils de recherche dans les banques		
SEQUEST	[97]	http://www.thermo.com
Mascot	[2]	http://matrixscience.com
ProteinProspector	[43]	http://prospector.ucsf.edu
ProbID	[98]	http://tools.proteomecenter.org/wiki/index.php?title=Software:ProbID
Tandem	[99]	http://www.thegpm.org
SpectrumMill		http://www.chem.agilent.com
Phenyx	[100]	http://www.phenyx-ms.com
OMSSA	[101]	http://pubchem.ncbi.nlm.nih.gov/omssa
VEMS	[102]	http://personal.cicbiogune.es/rmatthiesen
MyriMatch	[103]	http://www.mc.vanderbilt.edu/msrc/bioinformatics
Outils de recherche dans les banques de spectres		
SpectraST	[104]	http://www.peptideatlas.org/spectrast
X ! P3	[105]	http://p3.thegpm.org/tandem/ppp.html
Biblispec	[106]	http://proteome.gs.washington.edu/bibliospec
Outils de séquençage de novo		
Lutefisk	[107]	http://www.hairyfatguy.com/lutefisk
PepNovo	[108]	http://peptide.ucsd.edu/pepnovo.py
PEAKS	[109]	http://www.bioinformaticssolutions.com
Sequit		http://www.proteomefactory.com
Approches hybrides ou par sequence tag		
GutenTag	[110]	http://fields.scripps.edu/GutenTag
Inspect	[111]	http://peptide.ucsd.edu/inspect.html
Popitam	[112]	http://www.expasy.org/tools/popitam
Paragon	[113]	https://licensing.appliedbiosystems.com/download/ProteinPilot/2.0/

En pratique, les logiciels prennent chaque spectre expérimental en entrée et lui attribuent un score en le comparant aux profils théoriques de fragmentation des peptides de la banque de données. Le nombre

de peptides candidats est réduit en fonction des critères spécifiés par l'utilisateur comme la tolérance de masse, l'enzyme de digestion, les modifications chimiques ou post-traductionnelles autorisées. En sortie les logiciels génèrent une liste de séquences peptidiques classées selon leur score de corrélation avec le spectre expérimental. Ce score mesure le degré de similitude entre le spectre expérimental et le spectre théorique et sert de premier paramètre pour discriminer les identifications correctes de celles qui sont fausses. Dans la suite, nous avons choisi de décrire un peu plus en détail les deux algorithmes des logiciels (BioWorks et Mascot) que nous avions au laboratoire et qui ont été utilisés au cours de cette thèse.

a BioWorks

Cet algorithme a été le premier décrit dans la littérature par Eng *et al.* en 1994 [97]. Il est vendu par Thermo Fisher Scientific. Dans la version originale, chaque spectre MS/MS était analysé séquentiellement ce qui impliquait des recherches très lentes. Afin de pallier ce problème de vitesse, l'algorithme a été modifié pour permettre les recherches dans des banques indexées (*i.e.* une banque de données protéiques prédigérées *in silico*). Une version *cluster* a également été introduite pour permettre la recherche en parallèle des différents spectres MS/MS sur plusieurs processeurs. Ces deux améliorations ont permis de réduire considérablement les temps de recherche. Cependant, les banques indexées présentent le désavantage d'être statiques : les conditions de digestion (enzyme, nombre de *missed cleavages*) et les modifications prises en compte sont déterminées une fois pour toutes. Depuis la version BioWorks 3.3, les modifications peuvent être spécifiées au moment de la recherche. Au cours des années, l'algorithme a été modifié notamment pour permettre l'inclusion de modifications [114] ou les recherches dans des banques ADN ou EST (*Expressed Sequence Tag)* [115].

BioWorks calcule deux types de scores. Après sélection des peptides dont la masse correspond à celle du précurseur, dans la fenêtre de tolérance spécifiée, le logiciel calcule un score préliminaire (Sp) basé sur les 200 ions les plus intenses du spectre. Comparé au spectre théorique, ce score prend en compte l'intensité des 200 ions gardés, un bonus lorsque plusieurs ions consécutifs d'une même série correspondent et également un bonus pour les ions immonium. Ce score permet de réduire la liste des peptides possibles à 500 avant de calculer le score de corrélation croisée (Xcorr). Ce deuxième score fait intervenir tous les ions du spectre après réduction des données. Le spectre expérimental est divisé en 10 segments chacun normalisé à 50 en intensité. Les spectres théoriques sont eux aussi réduits : les ions b et y ont pour intensité 50, les ions dans une fenêtre de 1Da par rapport aux ions y et b ont une intensité de 25 alors que les ions de type perte de neutre (par exemple les ions a) ont pour intensité 10.

Pour chaque peptide candidat, plusieurs scores et rangs correspondants sont donc déterminés. Une comparaison récente des différents algorithmes de recherche a montré que SEQUEST permettait d'identifier plus de peptides corrects même à partir de données pauvres en qualité [116] mais que sa spécificité était moins satisfaisante, *i.e.* avec une moins bonne capacité à séparer les faux positifs des vrais positifs. Plus récemment, un score basé sur la probabilité a été ajouté pour combler ce manque. Ce score représente la probabilité due au hasard de trouver une correspondance aussi bonne ou meilleure que le peptide envisagé.

b Mascot

L'algorithme de recherche Mascot est développé par Matrix Science et a fait l'objet d'une publication en 1999 par Perkins *et al.* [2]. Ce logiciel est extrêmement populaire notamment du fait de sa compatibilité avec les formats de données de tous les instruments de spectrométrie de masse, quel que soit leur fournisseur. Une version en ligne est disponible, restreinte au niveau du nombre de spectres MS/MS qu'il est possible d'envoyer.

Lors de la recherche, Mascot sélectionne de manière itérative des sous-groupes des ions les plus intenses du spectre, cherchant le sous-groupe qui permet de différencier et de choisir le plus clairement le score le plus haut pour le peptide correspondant. Les différentes séries d'ions (y ou b par exemple) et les états de charge considérés dans la procédure de correspondance sont testés indépendamment et en combinaison. Ainsi seules les séries d'ions contenant un nombre significatif d'ions fragments correspondants dans le spectre sont utilisées pour le score et testées indépendamment et en combinaison pour déterminer le plus haut score atteignable par le peptide. Ainsi avoir trop de séries d'ions n'affecte pas le score, cela réduit juste la spécificité [96].

D'autre part, la recherche d'un spectre particulier peut s'effectuer en supposant différents états de charge du précurseur et des fragments : cela est particulièrement intéressant dans le cas de données de faible résolution. Au final seul l'état de charge donnant le plus haut score est conservé ce qui permet de ne pas affecter les statistiques de la recherche et de ne pas contribuer à augmenter les faux positifs. Pour chaque peptide correspondant à un spectre, une probabilité basée sur le score MOWSE [117] est retournée et convertie en score (*Ions score* = $-10\log_{10}P$). Pour chaque spectre, deux scores supplémentaires sont indiqués et permettent d'évaluer les données : le seuil d'identité et le seuil d'homologie généralement plus faible. Le seuil d'identité est calculé en se basant sur le nombre d'ions précurseurs qui peuvent correspondre et du taux d'erreur spécifié par l'utilisateur (0,05 par défaut) :

$\text{seuil d'identité} = -10\log_{10}(\frac{\text{taux d'erreur}}{\text{précurseurs correspondants}})$. Ce seuil ne dépend donc pas de la qualité des spectres mais seulement de la tolérance MS. En revanche, le seuil d'homologie est directement lié à la qualité du spectre et est basé sur la distribution des scores peptidiques pour un spectre particulier. Pour un spectre médiocre où aucun peptide possible ne dépasse le seuil d'identité, ce seuil permet de valider tout de même le peptide de plus haut score en montrant qu'il se distingue des autres peptides possibles (Figure 13).

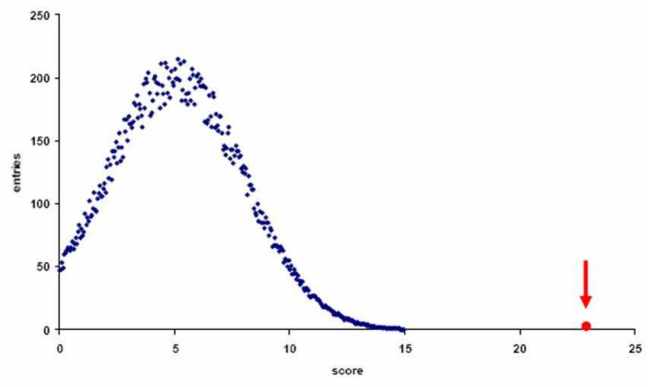


Figure 13 : Illustration de la signification du seuil d'homologie (ASMS 2005, user meeting Matrix Science). Le score du peptide marqué par une flèche se distingue de la distribution de scores des autres peptides.

II-B.3.2. L'identification par recherche dans les librairies de spectres

Une inefficacité notable des expériences de protéomique repose notamment sur le fait qu'on retrouve souvent des peptides déjà identifiés préalablement par recherche dans les banques de données, ce qui est souvent une perte de temps et une source d'erreur. Avec la disponibilité de grande quantité de données protéomiques, dont une partie est **organisée en banques** accessibles (Tableau 1), on peut imaginer que toutes les protéines d'une espèce, détectables par spectrométrie de masse, ont finalement déjà été découvertes. C'est notamment le cas dans le cadre de séquençage systématique des protéines produites par certaines bactéries ou espèces eucaryotes [118] qui ont atteint une remarquable couverture de protéome. De telles cartes protéomiques ouvrent maintenant la possibilité de déduire la séquence d'un peptide par correspondance de son profil de fragmentation avec une librairie de spectres représentant les séquences peptidiques contenues dans la carte protéomique [104, 106]. Dans ce type d'approche, la **librairie de spectres** est méticuleusement compilée à partir d'une large collection de spectres expérimentaux observés de peptides correctement identifiés. Un spectre d'une espèce inconnue peut être interprété en le comparant à tous les candidats de la librairie de spectres pour déterminer celui qui correspond avec le plus haut degré de similitude. Cette approche surpasse les recherches classiques dans les banques de données en termes de vitesse, de taux d'erreur et de sensibilité et présente l'avantage

que les outils développés pour évaluer les résultats des recherches dans les banques de données sont facilement adaptables à cette méthode [104, 106]. Cependant aucun peptide ne sera identifié s'il n'a pas été préalablement rentré dans la librairie. Tant que la carte protéomique de l'espèce étudiée n'a pas été complétée, ce type d'approche peut être utilisé comme première recherche rapide pour identifier des peptides et des protéines déjà présents dans la librairie avant une deuxième recherche classique dans les banques de données des spectres expérimentaux non attribués.

II-B.3.3. L'identification par séquençage *de novo*

Dans une approche par séquençage *de novo,* la séquence en acides aminés d'un peptide est lue de manière explicite à partir du spectre de fragmentation. Ce type d'approche était d'abord accompli manuellement. Mais plus récemment, des outils ont été développés pour aider le chercheur dans sa tâche (Tableau 1). Le principal avantage du séquençage *de novo* par rapport aux recherches dans les banques de données est qu'il permet d'identifier des spectres dont la séquence peptidique exacte n'est pas dans la banque de données, comme des peptides modifiés ou des isoformes. Pour cette raison, cette approche est principalement utilisée dans le cadre d'étude d'espèces dont le génome est peu ou pas décrit ou bien dans l'identification de peptides modifiés. Cependant, l'analyse *de novo* demande des ressources informatiques importantes et des spectres de fragmentation de haute qualité. De plus, le chercheur analysant des données protéomiques est d'abord soucieux de savoir quelles sont les protéines présentes dans son mélange. Ceci implique que les séquences déduites des spectres expérimentaux par analyse *de novo* doivent ensuite être attribuées par BLAST à une protéine connue présente dans les banques, ce qui dans le cas d'analyses protéomiques haut débit est relativement long. C'est pourquoi, une stratégie plus efficace combine une première recherche classique dans les banques de données avec des outils de séquençage *de novo* appliqués aux spectres de bonne qualité qui n'ont pas pu être attribués [119].

II-B.3.4. L'identification par des approches hybrides

Certaines approches sont basées à la fois sur l'analyse *de novo* et la recherche dans les banques de données. La première étape consiste en la déduction de courts *tags* peptidiques issus des spectres MS/MS, suivie par une recherche dans les banques en mode *error tolerant*. Cela signifie notamment que la recherche autorise une (ou plusieurs) erreur(s) entre la séquence peptidique réelle ayant produit le spectre expérimental et la séquence dans la banque. Ce type d'approche a d'abord été introduit par le groupe de Mann [120]. La limitation des recherches aux seuls peptides contenant le *tag* (ou les *tags*) permet de réduire considérablement la durée de recherche dans les banques. Ces approches hybrides présentent également un fort potentiel dans l'analyse systématique des peptides modifiés de manière artificielle ou post-traductionnelle. En effet si tous les types de modifications sont autorisés sur n'importe quel site, cela amène à une explosion du nombre des combinaisons possibles, ce qui est peu compatible avec les

recherches dans les banques de données. En revanche, l'utilisation de *tags* de séquences permet de réduire la taille de la banque et donc rend les recherches réalistes.

II-C. Histoire de la peau

Dans le cadre d'une collaboration avec L'Oréal, mes travaux portaient sur l'identification du dégradome de la couche la plus superficielle de la peau. La peau est un organe composé de l'épiderme, tissu majoritairement épithélial, et du derme, tissu de nature conjonctive [121]. Les kératinocytes sont les cellules les plus nombreuses de l'épiderme. Bien qu'il y ait des variations selon les régions anatomiques, la structure de base de la peau demeure toujours la même.

II-C.1. Organisation des cellules de l'épiderme

La fonction principale de la peau est de former une barrière efficace entre l'intérieur et l'extérieur de l'organisme. L'épiderme joue à la fois le rôle de barrière physique et de barrière biochimique. La barrière physique est principalement assurée par le *stratum corneum* (couche la plus superficielle de l'épiderme) alors que la barrière biochimique consiste en un ensemble de molécules comme des lipides, des peptides antimicrobiens ou encore des enzymes. Certains auteurs considèrent que la formation du *stratum corneum* est le but ultime de l'épiderme. L'épaisseur moyenne de l'épiderme varie entre 60 et 100µm et peut atteindre 600µm au niveau du *stratum corneum* de la plante des pieds ou de la paume des mains [122]. Il est constitué principalement (90%) de kératinocytes, mais également d'autres cellules (mélanocytes, cellules de Merkel, cellules de Langerhans…). L'épiderme est organisé en différentes couches de cellules dont la structure change en fonction de leur état de différenciation. Ces couches se nomment des plus profondes aux plus superficielles : basale, épineuse, granuleuse et cornée (Figure 14). Au terme de sa différenciation, la cellule est devenue de plus en plus spécialisée et limitée dans ses fonctions. Les cellules de la couche cornée (certes mortes mais protectrices) desquament : elles sont remplacées par d'autres cellules puisqu'en parallèle des cellules basales prolifèrent. Ainsi l'épaisseur du *stratum corneum* reste relativement constante du fait de ces phénomènes de desquamation et de multiplication bien équilibrés.

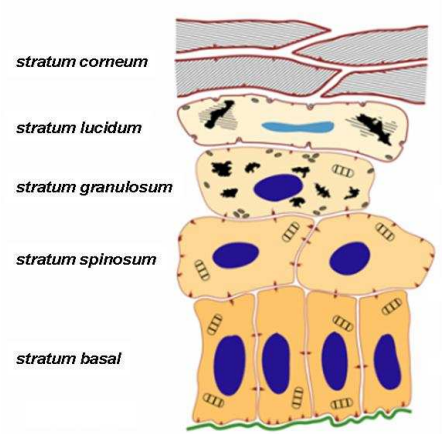


Figure 14 : Schéma de l'épiderme [123].

II-C.1.1. Le *stratum* basal

La **couche basale (*stratum* basal)** est formée de cellules en contact avec le derme, dont certaines prolifèrent et donnent naissance aux cellules des couches plus superficielles. La couche basale, composée majoritairement de kératinocytes, possède des cellules peu différenciées (10% des cellules) ainsi que d'autres types cellulaires. Une cellule souche, suite à la mitose, donne naissance à deux cellules identiques : l'une des cellules filles reste dans la couche basal et continue ses cycles de division, l'autre sort du cycle cellulaire pour se différencier. Une fois différenciée en kératinocyte, une cellule basale va, après un temps plus ou moins long, progresser vers la couche supérieure, à savoir la couche épineuse, puis compléter sa différenciation d'une couche à l'autre vers la surface.

La morphologie des cellules basales est d'aspect cubique avec un noyau volumineux. Leur cytosquelette renferme des filaments intermédiaires de kératines [124]. Les filaments de kératines sont reliés aux hémi-desmosomes qui ancrent les cellules à la jonction dermo-épidermique et aux desmosomes qui joignent les cellules entre elles.

II-C.1.2. Le *stratum spinosum*

La **couche épineuse** tire son nom des nombreux desmosomes qui joignent les cellules et qui leur donnent un aspect épineux [125]. Il s'agit d'une couche de plusieurs cellules. Des changements morphologiques sont observables entre les cellules situées plus profondément dans la couche épineuse et celles plus superficielles [121]. Celles localisées plus profondément sont polyédriques avec un gros noyau et peuvent présenter des mitoses. Celles en position plus superficielle, davantage différenciées, présentent un accroissement de leur taille ainsi qu'un aplatissement. Ces cellules sont en synthèse active de protéines en particulier de kératines de types 1 et 10 [124]. Ces cellules plus matures que les cellules basales progressent vers la troisième couche grâce à la poussée des cellules sous-jacentes.

II-C.1.3. Le *stratum granulosum*

La **couche granuleuse** moins épaisse est constituée de quelques cellules aplaties seulement et forme la dernière couche nucléée de l'épiderme. Les cornéocytes (kératinocytes en fin de différenciation) sont entourés d'une matrice extracellulaire, nommée enveloppe cornée. Les précurseurs de celle-ci font progressivement leur apparition dans le *stratum granulosum*. L'involucrine commence à apparaître dans la couche épineuse et s'accumule dans la couche granuleuse [121]. Il en est ainsi des transglutaminases, enzymes qui catalysent la formation de cette enveloppe protéique dans la couche cornée [126]. La loricrine est un autre constituant majeur de cette enveloppe qui apparaît dans la couche granuleuse [127].

II-C.1.4. Le *stratum lucidum*

La **couche claire** n'est retrouvée que dans l'épiderme épais de la paume des mains et de la plante des pieds. Il s'agit d'une couche de transition continue entre la couche granuleuse encore métaboliquement active et la couche cornée inerte. Dans les autres sites anatomiques, cette couche n'existe pas.

II-C.1.5. Le *stratum corneum*

La **couche cornée** est composée de cellules kératinisées (**cornéocytes**) qui proviennent des couches sous-jacentes et qui desquameront une fois leur rôle de protection joué. Avant de devenir inerte et adaptée pour remplir sa fonction de protection, la cellule subit donc une série de modifications. Les organelles cellulaires commencent à disparaître progressivement. La membrane devient perméable et une partie du contenu cytoplasmique fuit. L'espace intercellulaire s'élargit et le volume cytoplasmique décroît. Les organelles cellulaires, incluant le noyau, continuent à être dégradées. La **membrane cytoplasmique disparaît tandis que l'enveloppe cornée s'épaissit**.

Plusieurs protéines ou structures apparues dans la couche granuleuse deviennent fonctionnelles dans la couche cornée. Ces éléments contribuent alors au développement de la fonction barrière, soit le principal rôle de la couche cornée.

Les cornéocytes sont des cellules mortes. A l'exception des filaments de kératine, la presque totalité du contenu cytoplasmique, incluant le noyau et les organelles, est dégradée par diverses enzymes : protéases, hydrolases acides [121]. La membrane cytoplasmique devient perméable, puis le noyau se dégrade. Le cornéocyte se déshydrate et perd ses débris d'organelles dans l'espace intercellulaire par exocytose de sorte que les **amas de kératines** composent alors **80%** du contenu intercellulaire [128].

Avant de desquamer, les cornéocytes s'avèrent bien adaptés pour remplir la fonction barrière et réguler l'hydratation du *stratum corneum*. Ils sont inertes, dotés d'une bonne cohésion et entourés d'une

matrice lipidique : modèle des briques et du mortier [129]. Les lipides intercellulaires et les acides aminés résiduels de la filaggrine, en permettant à la couche cornée et à l'épiderme de conserver leur hydratation, contribuent à maintenir ses propriétés de distension, d'élasticité et, conséquemment, de fonction barrière efficace [130]. Des desmosomes, modifiés appelés cornéodesmosomes, demeurent présents [121]. Plus les **cornéocytes** sont proches de la **surface**, plus les **cornéodesmosomes** vont être **dégradés** sous l'action de différentes **enzymes**. Comme ceux-ci sont responsables en grande partie de la cohésion cellulaire, leur **disparition** conduit à la **desquamation** de ces résidus cellulaires [131].

Aplatis, les cornéocytes sont devenus les plus grandes cellules de l'épiderme.

II-C.2. Constituants du *stratum corneum*

II-C.2.1. Les kératines

Les filaments intermédiaires (FI) forment, avec les microtubules et les microfilaments d'actine, les différentes architectures du cytosquelette. Ces structures jouent toutes un rôle dans le maintien de la forme de la cellule, l'organisation des structures internes et la motilité. Les **kératines** forment une famille protéique qui **compose les FI** : elles sont exprimées exclusivement dans les épithéliums.

Dans cette partie sont rappelées quelques propriétés de ces protéines (profil d'expression, interaction avec les protéines, fonctions).

a Expression des kératines

Les kératines sont des **éléments structuraux majeurs** des cellules épithéliales. Chacun des 49 gènes code pour une protéine unique dont la classification est basée sur l'ordre de leur migration, dans un gel de polyacrylamide en deux dimensions [132].

On distingue **deux classes** de protéines :

- Les kératines de type I **acides** (pI<6) (K9–K23)
- Les kératines de type II **basiques** (pI>6) (K1–K8)

Les protéines d'un type s'associent obligatoirement avec l'autre type et forment des filaments hétérodimériques. Donc, pour qu'une cellule puisse produire des filaments de kératines, elle doit exprimer au moins une kératine de type I et une autre de type II [133].

Le **profil d'expression et de couplage** des différentes kératines est indicateur de fonctions spécialisées et complémentaires de ces protéines au sein des cellules. Par exemple, au niveau de

l'épiderme, les cellules de la couche basale expriment la paire K8/K18. La couche épineuse infrabasale se compose des K5/K14 et, lorsque qu'on remonte vers la couche épineuse suprabasale, il y a un remplacement graduel des K5/K14 par la paire K1/K10 [124, 134]. De plus, la K22 se rajoute au niveau de la couche granuleuse [135]. L'expression de kératine K9 est confinée dans l'épiderme palmo-plantaire [136]. Alors que dans la couche basale, les kératines représentent 30% des protéines totales, elles composent à 80% les cornéocytes du *stratum corneum* [128]. Ainsi, au sein d'un même tissu, les différentes cellules qui le composent peuvent exprimer diverses paires de kératines qui ont des fonctions différentes.

b Les interactions protéines-kératines

Les kératines peuvent interagir avec plusieurs protéines. Par exemple, la **famille des plakines**, relie les filaments intermédiaires aux microfilaments, aux microtubules et/ou à des complexes d'adhésion.

Les plakines incluent notamment la **desmoplakine**, la **plectine**, la **périplakine**, l'**envoplakine** et forment une famille de protéines cytoplasmiques servant de lien entre la membrane et le réseau de FI pour maintenir l'intégrité des tissus. Les **plakines lient les FI *via* leur extrémité globulaire C-terminale** [137], et sont exprimées plus particulièrement dans les tissus épithéliaux et musculaires subissant un stress mécanique. La desmoplakine peut lier diverses kératines (K5/K14 et K8/K18) et la vimentine [138, 139].

D'autres protéines peuvent aussi se lier aux FI, sans toutefois affecter leur organisation. Plusieurs d'entre elles sont des protéines impliquées dans une ou plusieurs voies de signalisation. Les kératines peuvent se lier directement entre elles, à la protéine kinase C γ [140], l'ARN polymérase II [141]…

c Les fonctions des kératines

Le rôle mécanique des kératines au niveau des épithéliums stratifiés est bien établi. Les mutations mineures dans les kératines ont comme conséquence la plupart des phénotypes ichtyosiques ou bulleux, ce qui prouvent le rôle essentiel de ces protéines dans l'intégrité et la cohésion de l'épiderme. Plusieurs maladies affectant l'épiderme sont dues à des mutations génétiques dans les kératines dont l'*epidermolysis bullosa simplex*. Cette maladie se traduit par une fragilité excessive de la peau à tout contact physique [142]. Les mutations majeures sont, elles, incompatibles avec la vie.

II-C.2.2. Les jonctions cellulaires

a Généralités

Les jonctions intercellulaires peuvent être classées en trois groupes [143] :

- Les jonctions étanches : sites où les membranes plasmiques de deux cellules adjacentes s'accolent,

- Les jonctions de communication : structures membranaires permettant une communication directe et la transmission de molécules entre cellules,

- Les jonctions d'adhésion : structures permettant l'ancrage des cellules entre elles.

Parmi ces dernières, on peut détailler trois types :

- Les jonctions adhérentes associées aux microfilaments,
- Les **desmosomes** : sites d'adhésion retrouvés principalement dans les cellules épithéliales. Les kératinocytes sont les cellules humaines qui contiennent la plus importante densité de desmosomes sur la surface membranaire,
- Les hémi-desmosomes : jonctions situées au niveau des contacts cellules basales-membrane basale. Ces jonctions ont une structure similaire à un demi-desmosome, bien qu'ayant des constituants biochimiques différents.

Dans la suite, nous nous sommes concentrés sur les desmosomes et les cornéodesmosomes qui assurent la cohésion épidermique en jouant le rôle de **Velcro intercellulaire** [144].

b Structure et composition du desmosome et du cornéodesmosome

Les desmosomes (Figure 15) sont des jonctions intercellulaires qui confèrent une **stabilité mécanique** à de nombreux tissus. Ces structures sont plus abondantes dans les zones soumises à un stress mécanique dont l'épiderme, le cœur… Les desmosomes se comportent comme des boutons ou des scratch qui joignent les faces latérales de cellules adjacentes. Le premier site d'adhésion est situé dans les **domaines extracellulaires** de protéines transmembranaires de la famille des **cadhérines**. Du côté intracellulaire des desmosomes, se trouve la plaque divisée en une région proche de la membrane (plaque dense extérieure) et en une région plus éloignée de la membrane (plaque dense intérieure). La zone extérieure contient les **domaines intracellulaires** de **cadhérines** et aussi deux protéines **armadillo** (**plakoglobine** et **plakophiline**). La zone intérieure est composée de la **desmoplakine** qui relie les sites d'adhésion intercellulaires au réseau des **FI**, ce qui renforce mécaniquement les sites d'adhésion intercellulaires [145].

L'aspect de cette structure change en fonction de la différenciation. Cette région transparente aux électrons devient **dense aux électrons** au niveau du *stratum corneum*. Pour souligner cette différence, le desmosome s'appelle alors **cornéodesmosome** [144, 146].

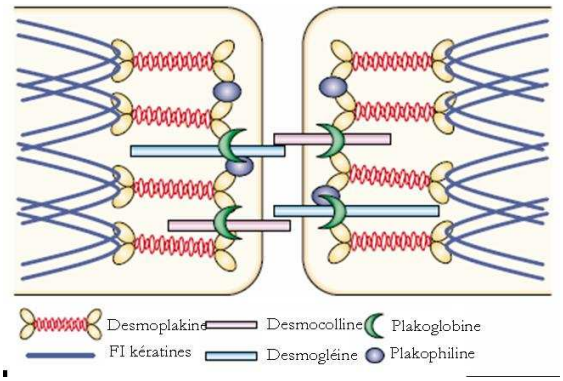


Figure 15 : Desmosome, mise en évidence de certaines des interactions principales protéines-protéines [147].

Quelques propriétés des principales protéines des cornéodesmosomes sont rappelées ici.

Desmoplakine

La desmoplakine interagit avec les filaments intermédiaires. La desmoplakine est une très grosse molécule en forme d'haltère, avec des domaines N-terminal et C-terminal d'un millier d'acides aminés, séparés par un domaine central d'à peu près la même taille [148]. Elle interagit avec la **plakophiline** par son domaine **N-terminal** [149]. La zone **C-terminale,** composée de trois domaines plakines [150], interagit avec les **FI** [151].

Protéines armadillo

Les protéines **armadillo** que l'on trouve dans les desmosomes regroupent les **plakophilines** et la **plakoglobine** [152]. Certaines données suggèrent que la plakoglobine pourrait être impliquée dans la régulation de l'association des autres composants des desmosomes, notamment à travers son domaine C-terminal [153]. Les plakophilines sont capables de lier de très nombreuses protéines des desmosomes par leur domaine N-terminal [149]. Une des hypothèses envisagées serait l'association des plakophilines aux cadhérines [154, 155] mais sans réels résultats définitifs au niveau structurel ou fonctionnel.

Cadhérines desmosomales

Les **cadhérines** desmosomales sont des constituants **transmembranaires** des desmosomes [156]. Il existe 3 isoformes de la desmocolline et 4 de la desmogléine. Leurs domaines extracellulaires peuvent être divisés en plusieurs sous-domaines appelés « **domaines cadhérines**» [157]. Les différentes isoformes présentent des profils de distribution et d'expression différentes dans l'épiderme. Les formes 2 et 3 sont présentes dans les couches plus profondes alors que les **formes 1** (et la desmogléine 4) sont des **marqueurs tardifs de la différenciation** épidermique [158-160]. Ces protéines présentent des **propriétés d'adhésion**. Les desmocollines et les desmogléines se lient de façon **hétérophile** [161, 162].

Cornéodesmosine

La **cornéodesmosine** est une glycoprotéine sécrétée **incorporée aux desmosomes** lors de leur conversion en cornéodesmosomes dans les **couches cornées** de l'épiderme. Sa fonction serait de stabiliser le cornéodesmosome contre les stress mécaniques. Cette protéine est une **protéine d'adhésion** [163], notamment par ses **boucles glycine N- et C-terminales** riches en glycines et en sérines [164] dont la **dégradation** protéolytique est un **prérequis à la desquamation** [165]. Sa protéolyse prématurée et la fragilité desmosomale associée est importante dans les pathologies liées au syndrome de Netherton [166]. Au contraire, la persistance de la cornéodesmosine conditionne les phénomènes de peau sèche (xérose) [167].

II-C.3. La peau, une structure vivante : cas particulier la desquamation

II-C.3.1. Généralités

La desquamation épidermique est un phénomène partagé par tous les mammifères. Cette activité du corps, continue mais invisible, est la plupart du temps ignorée tant que la production de squames ne devient pas anormale.

Sont rappelés par la suite les mécanismes de la desquamation, les protéases impliquées et les facteurs de régulation.

La desquamation est un processus actif et non passif comme supposé encore jusque dans les années 80. Dans l'une des premières expériences s'intéressant à la question et réalisée par Goldschmidt, la surface épidermique était couverte et protégée d'influences extérieures. La récupération de cornéocytes isolés et totalement matures a conduit à la conclusion que ni friction ni toilette n'était nécessaire à leur détachement [168]. De même, la perte spontanée de cellules différenciées des cultures stratifiées de kératinocytes humains a souligné que la desquamation était l'étape finale et active dans le processus de différenciation des kératinocytes [169].

Les observations ultrastructurales du *stratum corneum* montrent que la taille et la distribution des desmosomes diminuent d'autant plus que les cornéocytes s'approchent de la surface [170]. Comme les **desmosomes** ont un rôle central dans **l'adhésion intercellulaire**, leur **destruction** est l'hypothèse de base au mécanisme de **desquamation**.

II-C.3.2. Rôle des enzymes

Au-delà de cette disparition morphologique, de nombreuses observations mettent en évidence le rôle important de plus d'une classe d'enzymes. En plus de protéases, un traitement avec glycosidase est nécessaire pour réduire significativement la taille des desmosomes [171]. L'analyse des protéines du *stratum corneum* montrent une **dégradation progressive des cadhérines** [172-174], de la **desmoplakine** [175] et de la **cornéodesmosine** [165]. Des protéases ont été isolées du *stratum corneum* et plusieurs sont exprimées et activées dans l'épiderme dans les étapes finales de la différenciation : la kallikréine 7 (**SCCE**) [176], la kallikréine 5 (**SCTE**) [177], la **cathepsine D** [178], et des **protéases à cystéine** (cathepsine L2 [175]). De plus l'application d'inhibiteur de protéase permet d'augmenter l'épaisseur du *stratum corneum* chez la souris [179]. Les desmosomes pourraient ne pas être les seuls substrats de protéases importants dans la desquamation. Par exemple, l'absence de matriptase est associée avec l'accumulation de profilaggrine et une cohésion plus forte du *stratum corneum* [180].

a Les protéases à sérine

De nombreuses kallikréines sont localisées dans les couches supérieures du *stratum granulosum* et dans le *stratum corneum*. Encore récemment, l'activité de cette famille de protéases n'était attribuée dans la peau qu'à la SCCE et la SCTE. Cependant, d'autres rôles cutanés des kallikréines sont devenus évidents ces dernières années. Ces enzymes sont capables de dégrader différents substrats impliqués dans la desquamation, les défenses antimicrobiennes, l'inflammation de la peau, la pigmentation… Dans la suite, nous avons résumé le rôle des kallikréines dans le processus de desquamation, d'autres informations pourront être trouvées dans une revue récente [181].

La **SCCE** est responsable de la **dégradation** des structures cohésives comme la **desmogléine** [172, 182]. Il a également été démontré que cette protéase était associée avec la plaque cornéodesmosomale [183], ce qui pourrait être relié aux fragments de plakines mis en évidence au cours de la différenciation [175]. Son pH optimum se trouve autour de 7-8 tout en conservant une activité résiduelle à pH 5,5, pH mesuré du *stratum corneum*. La SCCE est activée par la SCTE une autre protéase à sérine. La SCCE recombinante est également capable de **dégrader** la **cornéodesmosine** ainsi que les **desmocollines** *in vitro* mais **pas la desmogléine** [184].

La **SCTE** est une enzyme capable de s'auto-activer et présentant une activité tryptique [185]. Son pH optimum se situe à pH 8 mais comme la SCCE, elle présente encore une activité importante à pH 5,5. Cette protéase est capable de **dégrader** seule la **desmocolline**, la **desmogléine** et la **cornéodesmosine** [165, 184].

b Les protéases à acide aspartique

En 1995, Rawlings *et al.* ont mis en évidence pour la première fois le rôle de la **cathepsine D** dans la desquamation [186]. Plus tard, Horikoshi *et al.* ont montré que la libération des cornéocytes *in vitro* était inhibée à 40% par la pepstatine, inhibiteur des protéases à acide aspartique [178, 187]. Cette activité est due, dans le *stratum corneum,* pour 80% à la cathepsine D et pour 20% à la cathepsine E. Cependant, le rôle de celle-ci dans la desquamation a été *a priori* écarté du fait de sa localisation dans les cornéocytes et non spécifiquement dans l'espace intercellulaire comme la cathepsine D [188].

c Les protéases à cystéine

En 1999, Watkinson *et al.* ont identifié une nouvelle protéase à cystéine nommée, *stratum corneum thiol protease* (SCTP) [189]. Son activité n'est présente que dans les kératinocytes différenciés ce qui en fait un bon candidat pour jouer un rôle dans la desquamation. De plus, les auteurs ont démontré sa capacité à dégrader la desmocolline *in vitro*. Plus tard, cette protéase a été caractérisée un peu plus finement par Bernard *et al.* comme étant la **cathepsine L2** mais son expression ne semble pas liée au processus de différenciation tardive [175]. Ces travaux ont également permis de mettre en évidence une autre protéase à cystéine nommée *stratum corneum cathepsine L like* (**SCCL**) de 28kDa capable de **dégrader** *in vitro* la **cornéodesmosine** à pH acide.

II-C.3.3. Régulation de la desquamation

Différents facteurs peuvent intervenir dans la régulation de la desquamation comme le pH, la présence d'inhibiteurs, la modification des substrats par le calcium, l'hydratation *etc.*

Certaines **protéases** sont synthétisées sous forme de **précurseurs inactifs** et doivent être **activées**. Par exemple, la pro-SCCE est activée par la SCTE, la SCTE s'auto-active et la pro-cathepsine D est activée par une protéase à cystéine non identifiée.

Le *lymphoepithelial Kazal-type 5 serine protease inhibitor* (LEKTI) est un exemple d'**inhibiteur** important dans le phénomène de **desquamation**. Des peptides dérivés de LEKTI ont une activité inhibitrice sur la SCCE et la SCTE. Dans le syndrome de Netherton, le gène qui code pour LEKTI subit des mutations qui créent une protéine tronquée et par conséquent moins de peptides inhibiteurs. De ce fait, le SCCE et la SCTE sont plus actives, ce qui conduit à une desquamation exacerbée caractéristique du

syndrome [190]. Ceci a notamment été confirmé par des travaux de Descargues *et al.* qui montrent que chez des souris KO, la desmogléine et la profilaggrine subissent une dégradation plus rapide [191].

Le **pH** du *stratum corneum* **diminue** des couches profondes **vers la surface**. Il paraît évident que cette variation **a des conséquences sur les activités enzymatiques**. La SCCE et la SCTE pourraient donc être actives en profondeur puis d'autres enzymes de pH optimum acide (comme la cathepsine D) pourraient prendre le relais [178, 187].

II-C.4. Définition des échantillons

Au cours de cette thèse, nous avons travaillé sur différents types d'échantillon de peau (pour les protocoles de préparation détaillés, voir la Partie V-A.1 Préparation des échantillons de *stratum corneum* non digéré).

Pour la mise au point des conditions d'analyses de peptides endogènes, nous avons réalisé nos expériences sur des extraits de **peptides endogènes issus de *stratum corneum* plantaire**. Il s'agit de cornes de pied mélangées correspondant à différents donneurs. Dans la suite du manuscrit, nous faisons référence à ces échantillons en utilisant généralement l'abréviation **SCP**. Pour la comparaison topologique du dégradome de différentes zones de peau, nous avons comparé ces échantillons de SCP à des échantillons de ***stratum corneum* non plantaire**. Ces échantillons sont préparés à partir de poudre acétonique obtenue à partir de différents donneurs. Ces poudres acétoniques consistent à récupérer les cornéocytes arrachés sur une zone de peau particulière par *stripping* [192]. Dans notre cas, nous avons travaillé sur des ***strippings* de la jambe**. Dans la suite du manuscrit, la référence à de tels échantillons se fait par l'abréviation **PA**.

Enfin la comparaison des protocoles de séparation en chromatographie liquide mono- et multidimensionnelle a été réalisée sur des échantillons digérés de SCP.

PARTIE III- RESULTATS EXPERIMENTAUX ET DISCUSSION

III-A. Mises au point analytiques

III-A.1. Préparation des échantillons

III-A.1.1. Introduction

Nous cherchons à analyser directement par spectrométrie de masse les molécules telles que protéines, peptides, neuropeptides ou neurotransmetteurs, présents dans le cytoplasme de cellules ou bien extraits à partir de tissus. Le problème essentiel qui se pose est la quantité de molécules disponibles à l'échelle de la cellule : si les quantités de matériel sont compatibles avec notre niveau de sensibilité, les concentrations sont bien inférieures à ce qui est requis pour une analyse, car les volumes d'extraction restent très importants. De plus ces peptides sont généralement des fragments de protéines générés par différentes voies métaboliques : ils sont donc extrêmement sensibles à la protéolyse lors des étapes d'extraction et de purification. Une méthode d'enrichissement et de purification de neuropeptides contenus dans une matrice biologique a été préalablement mise au point lors de mon stage de DEA, durant lequel nous avons choisi d'extraire les molécules d'intérêt sur phase solide puis de les concentrer dans un tampon adapté à l'étude par spectrométrie de masse. Afin d'améliorer notre étude des peptides endogènes, les mélanges sont d'abord simplifiés lors d'une étape de chromatographie liquide monodimensionnelle avant l'étape de spectrométrie de masse ce qui permet d'identifier plus de peptides minoritaires.

L'approche précédente est reprise pour l'étude du dégradome de la peau. Le dégradome regroupe à la fois l'ensemble des protéases d'un tissu mais également leurs substrats, leurs inhibiteurs et les produits des réactions. Dans notre cas, nous avons choisi l'approche peptidomique afin d'étudier les peptides générés lors des processus de protéolyse et de remonter aux protéases ayant permis de les obtenir.

Après avoir déterminer les étapes nécessaires lors de la préparation des échantillons (reconcentration, fractionnement, digestion), nous avons également choisi quel était le meilleur tampon d'extraction pour les peptides de la peau. Dans l'optique d'une étude d'un grand nombre d'échantillons en triplicat, nous avons par ailleurs décidé des fractions utiles ou non lors du fractionnement afin de limiter le nombre d'échantillons à passer en LC-MS et donc de limiter le nombre d'heures d'analyse !

III-A.1.2. Reconcentration directe des peptides sur ZipTip™

Protocoles de reference :

Préparation des échantillons : **V-A.1.2.** p.145, **V-A.1.4.** p.146

Expérience LC-MS/MS : **V-A.3.1. a** p.148, **V-A.5.1.** p.152
Recherche dans les banques et validation : **V-A.6.1. a** p.154, **V-A.6.2. a** p.155

Dans un premier temps, la préparation des échantillons et leur traitement ont été mis au point sur du *stratum corneum* plantaire (SCP) afin de privilégier l'étude des peptides et non pas des protéines comme on peut le faire classiquement. En effet, ce type de matrice est disponible en grande quantité. Il s'agit d'échantillons provenant de cornes de pieds obtenues par prélèvement par des pédicures sur des sujets sains. Pour extraire les peptides endogènes du tissu, différents tampons ou solvants ont été utilisés en parallèle afin de comparer leur efficacité d'extraction. Dans la mesure où l'on ne veut extraire que les peptides et non les protéines, les solvants choisis sont des solvants doux sans adjonction de composés dénaturants pour rester compatible avec l'analyse par spectrométrie de masse :

- Tampon phosphate (PBS) 10mM
- Tampon (EDTA) 5mM
- Tampon (Tris) 0.5M pH=6.8
- Solution aqueuse de chlorure de sodium (NaCl) 2M
- Solution aqueuse d'acide trifluoroacétique (TFA) 0.1% (v/v)
- Méthanol (MeOH) acidifié par du TFA (MeOH/H_2O/TFA 80/20/0.1, v/v/v)
- Solution aqueuse d'acide formique (AF) 1%
- Méthanol acidifié par de l'AF (MeOH/H_2O/AF 80/20/1, v/v/v)

Après extraction et reconcentration des peptides sur ZipTip^{TM}, les peptides ont été directement analysés par RPLC-MS/MS sur un spectromètre de masse de type QTof2 (Waters).

L'application de ce protocole à des échantillons de *stratum corneum* plantaire nous a permis de fragmenter 162 précurseurs dont les séquences ont abouti à l'identification de 10 protéines (Tableau 2).

Tableau 2 : Protéines identifiées dans les échantillons de SCP par reconcentration seule des peptides endogènes suivie de l'analyse par LC-MS/MS.
Le score protéique Mascot et le nombre de peptides identifiés sont précisés.

Accesion Number	Protéine	Score	# peptides identifiés
K1C9	Keratin, type I cytoskeletal 9	2342	52
K2C1	Keratin, type II cytoskeletal 1	1295	40
DSC1	Desmocollin-1 precursor	634	17
DSG1	Desmoglein-1 precursor	478	12
CDSN	Corneodesmosin precursor	307	10
G3P	Glyceraldehyde-3-phosphate dehydrogenase	301	9
HORN	Hornerin	299	8
CYTA	Cystatin-A	277	7
FABP5	Fatty acid-binding protein, epidermal	126	4
DMKN	Dermokine precursor	80	3

Le score moyen des peptides identifiés est de 38 ce qui indique que certains peptides ont un score de fragmentation médiocre : ils ont été conservés dans l'identification finale des protéines car d'autres peptides de scores d'identification plus grands permettent de confirmer et de valider la présence de ces protéines.

D'autre part, les spectres MS de ces extraits directs montrent la présence de gros peptides qui ont donc été élués du ZipTip™. Les ions multichargés générés sont intenses et suppriment le signal des ions moins chargés qui sont coélués (Figure 16).

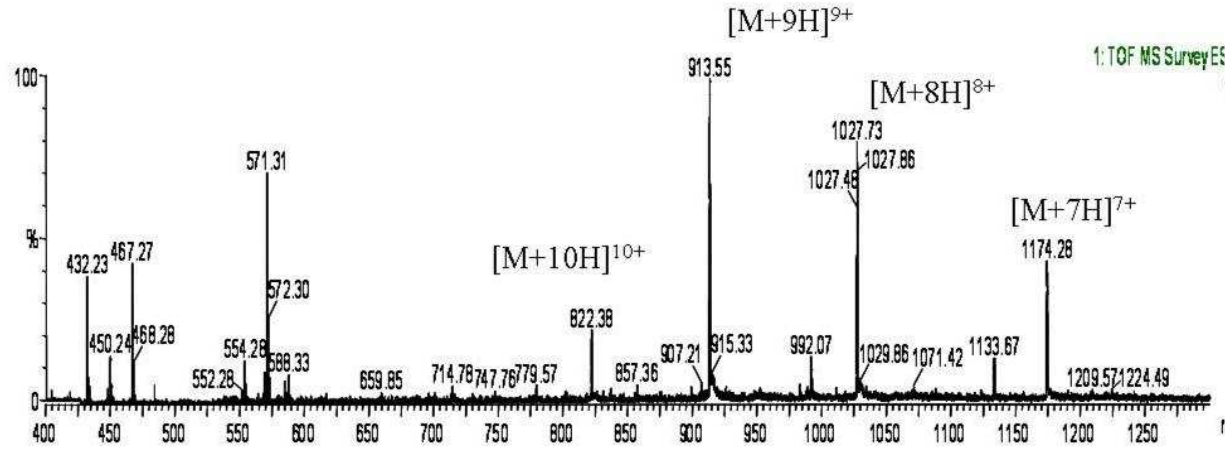


Figure 16 : Spectre MS correspondant à l'ionisation d'espèces peptidiques multi-chargées.
La plupart des ions les plus intenses (m/z 913,55Th, 1 027,73Th, 1 174,28Th) sont multichargés.

Lors de la recherche dans les banques, les ions à analyser considérés sont de charge 2+ ou 3+ puisque dans la méthode d'acquisition, la fragmentation est restreinte à ces mêmes états de charge. Or la résolution du QqTOF ne permet pas de déterminer l'état de charge d'un peptide 7 fois chargé : il est donc tout de même sélectionné mais la recherche dans les banques de données de façon automatisée ne peut être fructueuse. Dans ce cas, il faudrait éventuellement reprendre les analyses et effectuer une étape

de plus dans le traitement des données (déconvolution des spectres MS/MS). Cependant, même déconvolué, ce type de spectres reste difficilement interprétable : en effet le spectre d'une espèce 7 fois chargée contient évidemment des fragments eux-mêmes multichargés. Du fait de la résolution limitée du QqTOF, l'état de charge de ces fragments peut ne pas être déterminé correctement. D'autre part, la précision de masse de ce type de spectromètre de masse n'est également pas suffisante pour interpréter correctement les spectres. Enfin, ce type de gros peptides est très fortement retenu et nécessite un fort pourcentage d'ACN pour être élué de la colonne. Il n'est donc pas rare qu'il ne soit pas complètement élué et qu'un phénomène de pollution soit observé dans les échantillons suivants.

Afin de réduire cette contamination, nous avons choisi d'utiliser des concentrateurs par ultrafiltration. Deux types de concentrateurs ont été choisis : l'un avec un seuil de coupure de 10kDa qui permet d'éliminer les protéines de haut poids moléculaire (>20kDa) et l'autre avec un seuil de 5kDa qui permet de simplifier l'échantillon en le fractionnant en deux.

III-A.1.3. Fractionnement des peptides suivi de leur reconcentration sur ZipTip™

Protocoles de référence :

Préparation des échantillons : **V-A.1.2.** p.145, **V-A.1.3.** p.145, **V-A.1.4.** p.146

Expérience LC-MS/MS : **V-A.3.1. a** p.148, **V-A.5.1.** p.152

Recherche dans les banques et validation : **V-A.6.1. a** p.154, **V-A.6.2. a** p.155

Suite au fractionnement des peptides sur les concentrateurs, 3 fractions sont obtenues. La fraction inférieure à 5kDa (notée inf 5kDa dans la suite) est directement reconcentrée sur ZipTip™ par le protocole robotisé et la fraction entre 5 et 10kDa (notée 5-10kDa dans la suite) est digérée de façon à générer des peptides plus petits et plus facilement ionisables et surtout d'éviter les profils du spectre de la Figure 16 où le signal est supprimé par des ions multichargés correspondant au même peptide. On peut vérifier que l'étape de fractionnement assure bien sa fonction en représentant l'élution du peptide 913.7Th sur le chromatogramme d'un échantillon avec ou sans fractionnement. On constate alors que ce peptide a totalement disparu de la fraction inf 5kDa. Ainsi d'autres peptides sont détectables en MS dans la même fenêtre de temps et donc éventuellement sélectionnables pour être fragmentés.

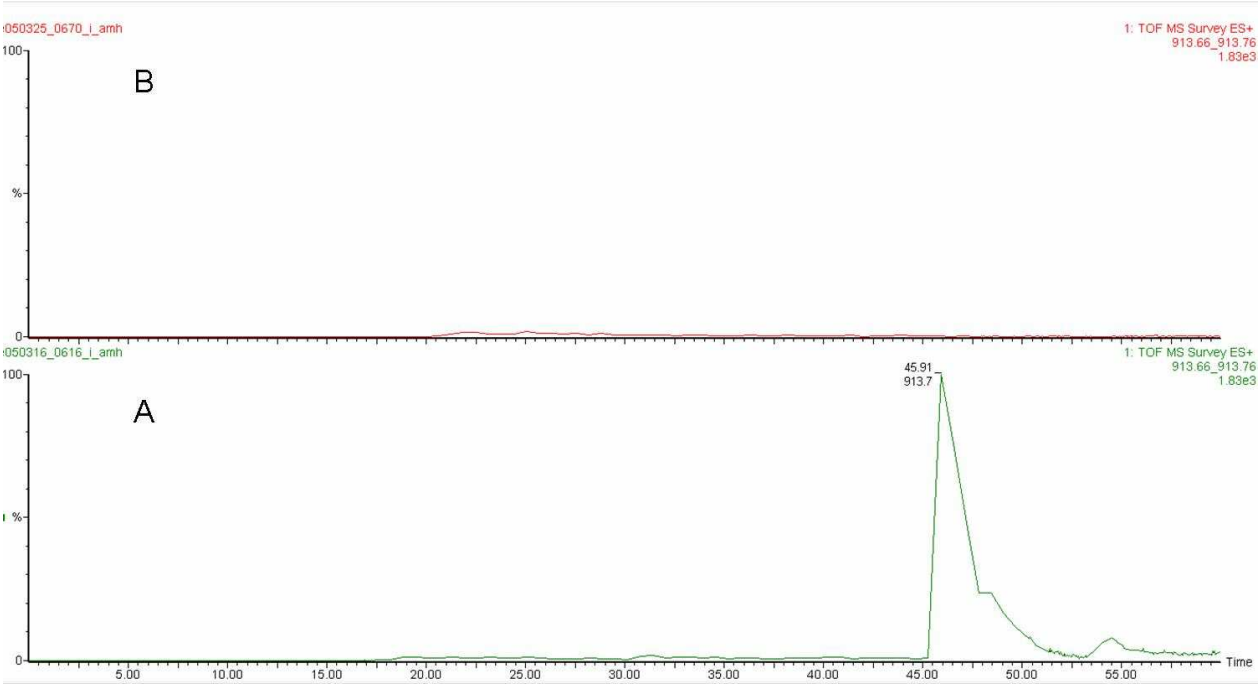


Figure 17 : Extraction du chromatogramme de l'espèce 913.7Th dont le spectre d'ionisation est représenté Figure 16.
A : Echantillon de SCP extrait dans du PBS 10mM avec reconcentration seule. B : Echantillon de SCP extrait dans du PBS 10mM avec fractionnement puis reconcentration (chromatogramme de la fraction inf 5kDa).

L'application de ce protocole à des échantillons de *stratum corneum* plantaire a permis l'identification de 14 protéines grâce à la fragmentation de 348 peptides, soit une multiplication par plus d'un facteur 2 par rapport à l'expérience sans l'étape de fractionnement par ultrafiltration (Tableau 3).

Tableau 3 : Protéines identifiées dans les fractions inf 5kDa des échantillons de SCP après fractionnement, reconcentration sur ZipTip™ et analyse par LC-MS/MS.
Le score protéique Mascot et le nombre de peptides identifiés sont précisés.

Accesion Number	Protéine	Score	# peptides identifiés
K1C9	Keratin, type I cytoskeletal 9	8132	135
K2C1	Keratin, type II cytoskeletal 1	2372	61
CYTA	Cystatin-A	1879	41
HORN	Hornerin	1084	25
DSC1	Desmocollin-1 precursor	710	22
CALL5	Calmodulin-like protein 5	444	13
S10A7	Protein S100-A7	411	9
S1A7A	Protein S100-A7A	411	9
K2C6A	Keratin, type II cytoskeletal 6A	355	8
K2C6C	Keratin, type II cytoskeletal 6C	355	8
K2C6B	Keratin, type II cytoskeletal 6B	327	8
DSG1	Desmoglein-1 precursor	301	10
CDSN	Corneodesmosin precursor	156	5
FABP5	Fatty acid-binding protein, epidermal	69	2

L'étape de fractionnement a donc permis d'augmenter significativement le nombre de peptides identifiés. Le score moyen d'identification de ces peptides est de 48 d'où des identifications plus sûres. En revanche, la proportion de peptides de kératines est toujours plus ou moins la même (57% sans fractionnement, 61% avec). Ceci est principalement dû au fait que ces protéines représentent 80% des protéines exprimées dans le *stratum corneum* [128] et sont donc la famille de « contaminants » constitutifs majoritaires de ce tissu. Il est assez logique que, si le spectromètre de masse a plus de temps pour fragmenter des peptides différents grâce à l'élimination des peptides multichargés qui supprimaient leurs signaux préalablement, les espèces nouvellement fragmentées soient encore majoritairement des peptides de kératines.

Nous verrons par la suite que l'analyse des échantillons par un spectromètre de masse de type LIT-FTICR permet de gagner en sensibilité et de monter en gamme dynamique.

III-A.1.4. De l'intérêt de la digestion des fractions 5-10kDa et sup 10kDa

Protocoles de référence :

Préparation des échantillons : **V-A.1.2.** p.145, **V-A.1.3.** p.145, **V-A.2** p.147

Expérience LC-MS/MS : **V-A.3.1. a** p.148, **V-A.5.1.** p.152

Recherche dans les banques et validation : **V-A.6.1. a** p.154, **V-A.6.2. a** p.155

Les fractions 5-10kDa et supérieure à 10kDa (notée sup 10kDa dans la suite) ont été digérées à la trypsine. Les peptides ainsi obtenus sont donc plus petits. D'autre part, cette digestion permet de placer un résidu basique en C-terminal des peptides ce qui permet à ces peptides d'être au minimum bichargés. Ainsi la présence minimum d'une charge en N-terminal et en C-terminal implique une fragmentation plus régulière des peptides et donc une interprétation plus aisée des spectres.

Dans ces fractions protéolysées, 848 peptides de 49 protéines différentes ont été identifiés avec un score moyen pour les spectres MS/MS de 52, ce qui là encore renforce la fiabilité des identifications (Tableau 4). Les recherches dans les bases de données en spécifiant *no enzyme* ont permis de mettre en évidence quelques peptides semi-tryptiques. Cependant ils ne représentent que 20% des peptides totaux. D'autre part, ces peptides semi-tryptiques n'apportent généralement pas d'information supplémentaire sur les sites de clivages endogènes car leur extrémité non tryptique a déjà été identifiée dans la fraction inf 5kDa.

Par la suite, nous avons donc décidé de ne plus digérer ces fractions et de les analyser directement par LC-MS/MS après dessalage et reconcentration sur ZipTip™. Notre stratégie générale est résumée Figure 18.

Tableau 4 : Protéines identifiées dans les fractions 5-10kDa et sup 10kDa des échantillons de SCP après digestion tryptique et analyse par LC-MS/MS.

Accesion Number	Protéine	Score	# peptides identifiés
K1C9	Keratin, type I cytoskeletal 9	14134	264
K2C1	Keratin, type II cytoskeletal 1	6115	97
K2C6A	Keratin, type II cytoskeletal 6A	1706	34
K2C6C	Keratin, type II cytoskeletal 6C	1706	34
K2C6B	Keratin, type II cytoskeletal 6B	1699	34
K1C16	Keratin, type I cytoskeletal 16	1575	29
K2C5	Keratin, type II cytoskeletal 5	1140	25
K22E	Keratin, type II cytoskeletal 2 epidermal	1054	19
K1C14	Keratin, type I cytoskeletal 14	920	18
CALL5	Calmodulin-like protein 5	879	17
K1C10	Keratin, type I cytoskeletal 10	873	14
DSC1	Desmocollin-1 precursor	869	18
ANXA2	Annexin A2	835	16
CYTA	Cystatin-A	749	18
K2C75	Keratin, type II cytoskeletal 75	644	12
K2C8	Keratin, type II cytoskeletal 8	599	12
K1C17	Keratin, type I cytoskeletal 17	580	12
K2C4	Keratin, type II cytoskeletal 4	580	12
K2C79	Keratin, type II cytoskeletal 79	555	10
K2C7	Keratin, type II cytoskeletal 7	516	10
GFAP	Glial fibrillary acidic protein	472	9
K2C80	Keratin, type II cytoskeletal 80	472	9
DSG1	Desmoglein-1 precursor	385	6
K1C15	Keratin, type I cytoskeletal 15	383	7
DESP	Desmoplakin	379	8
G3P	Glyceraldehyde-3-phosphate dehydrogenase	359	7
K1C13	Keratin, type I cytoskeletal 13	346	6
S10A7	Protein S100-A7	323	8
FABP5	Fatty acid-binding protein, epidermal	305	7
ALBU	Serum albumin precursor	300	7
K2C1B	Keratin, type II cytoskeletal 1b	289	5
K2C73	Keratin, type II cytoskeletal 73	256	7
TGM3	Protein-glutamine gamma-glutamyltransferase E precursor	254	6
S1A7A	Protein S100-A7A	243	6
UBIQ	Ubiquitin	213	5
CDSN	Corneodesmosin precursor	177	5
KRT81	Keratin type II cuticular Hb1	124	3
KRT82	Keratin type II cuticular Hb2	124	3
KRT83	Keratin type II cuticular Hb3	124	3
KRT84	Keratin type II cuticular Hb4	124	3
KRT85	Keratin type II cuticular Hb5	124	3
KRT86	Keratin type II cuticular Hb6	124	3

IL1F9	Interleukin-1 family member 9	123	4
PEBP1	Phosphatidylethanolamine-binding protein 1	112	2
K1C27	Keratin, type I cytoskeletal 27	99	3
PERI	Peripherin	88	2
VIME	Vimentin	88	2
K1C19	Keratin, type I cytoskeletal 19	82	2
HORN	Hornerin	70	2

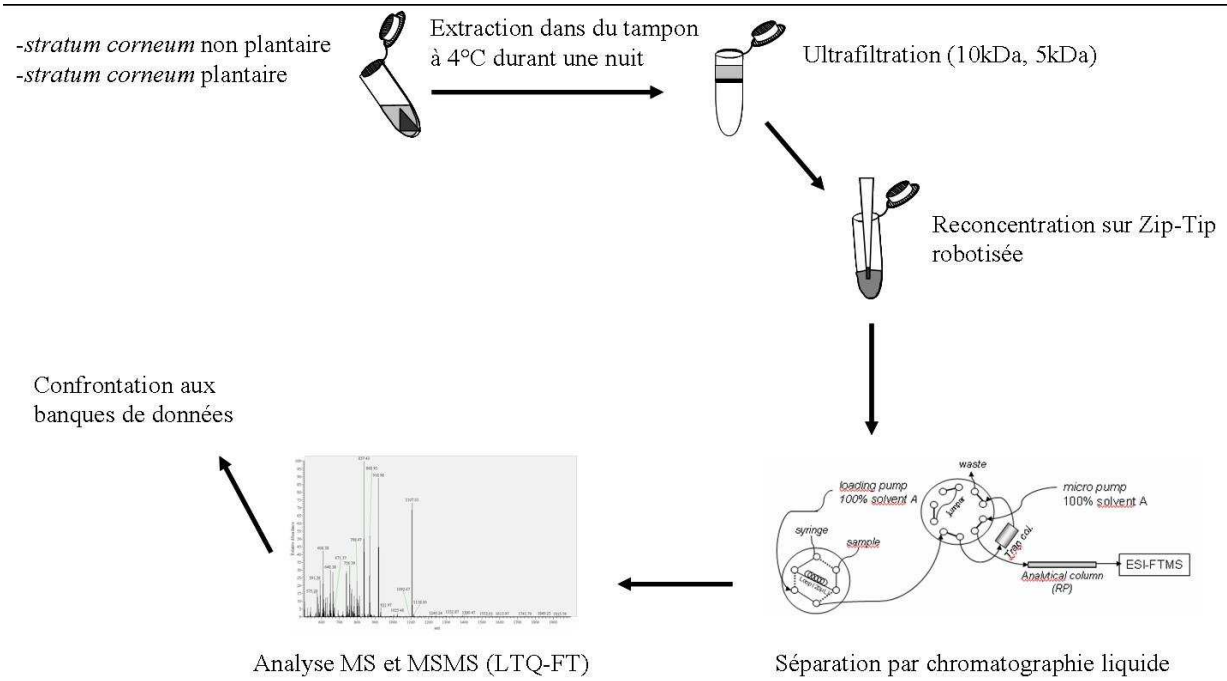


Figure 18 : Stratégie générale appliquée sur les échantillons de *stratum corneum* pour l'étude des peptides endogènes.

III-A.1.5. Tampon d'extraction

Protocoles de référence :

Préparation des échantillons : **V-A.1.1.** p.145, **V-A.1.2.** p.145, **V-A.1.3.** p.145, **V-A.1.4.** p.146

Expérience LC-MS/MS : **V-A.3.1. b** p.148, **V-A.5.2.** p.153

Recherche dans les banques et validation : **V-A.6.1. a** p.154, **V-A.6.2. b** p.155

A terme, il sera nécessaire de choisir un seul tampon d'extraction car si l'échantillon est disponible en quantité limitée, il n'est pas possible de réaliser plusieurs extractions. En effet dans le cadre d'étude de *stripping* itératif ou d'échantillons pathologiques, un seul essai est réalisable.

Pour le choix du tampon (ou du mélange de tampons), nous nous sommes appuyés sur les analyses d'échantillons de *stratum corneum* non plantaire par LIT-FTMS, puisque les échantillons de *stripping* itératif

sont préparés de la même façon : notamment l'étape d'extraction des peptides d'une poudre acétonique est commune.

Pour cela, nous avons repris les analyses de poudre acétonique provenant de prélèvements de différents donneurs et comparé l'apport des différents tampons d'extraction sur l'identification globale des peptides. En effet, les différents tampons permettent d'extraire un nombre de peptides différent mais sont également complémentaires, certains peptides n'étant retrouvés que grâce à l'un des tampons. Pour présenter les résultats (Figure 19), nous avons classé les tampons en ordre décroissant par rapport au nombre total de peptides différents qu'ils permettent d'identifier individuellement. Puis nous avons fait le rapport entre le nombre de peptides uniques utiles du tampon n et le nombre total de peptides uniques identifiés dans l'ensemble des analyses. On appelle peptide utile, un peptide qui n'a pas été identifié dans les tampons 1 à n-1 (exemple pour le tampon Tris pH=6,8 : peptide utile non identifié grâce à l'AF 1% ou au PBS 10mM). Pour la suite, nous avons donc décidé d'extraire les peptides dans un tampon PBS 10mM acidifié avec 1% d'AF ce qui devrait permettre de caractériser près de 70% des peptides identifiés globalement.

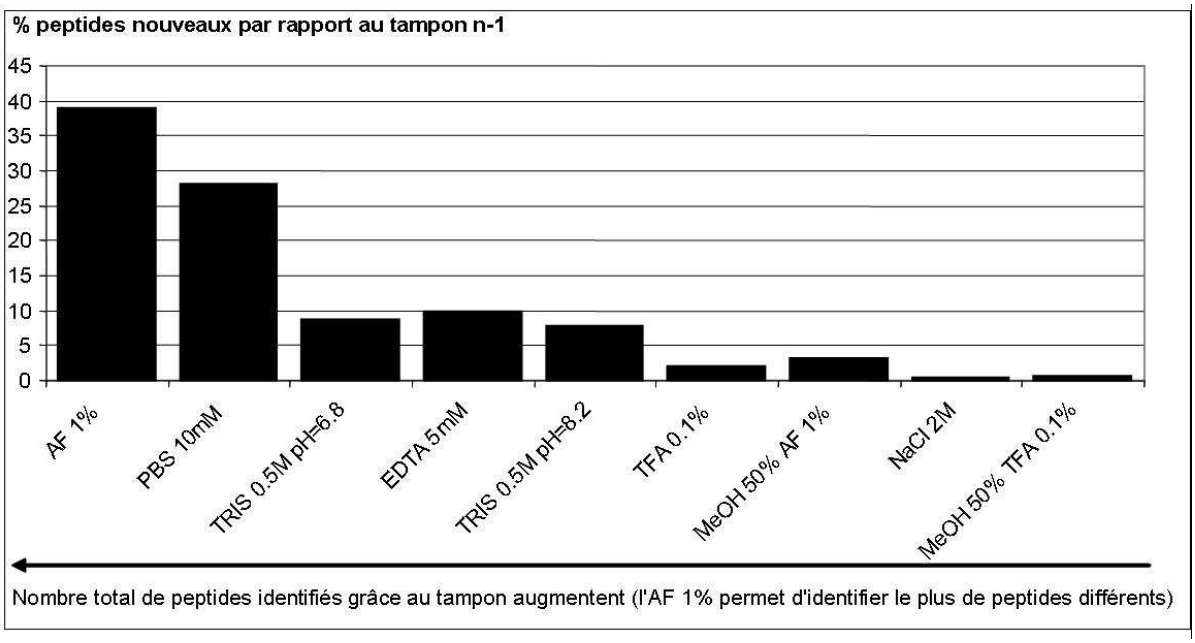


Figure 19 : Apport de chaque tampon dans les identifications.
Seuls les peptides nouvellement identifiés dans le tampon n par rapport aux tampons 1 à n-1 ont été comptabilisés.

III-A.1.6. Fractions de peptides utiles

Protocoles de référence :

Préparation des échantillons : **V-A.1.1.** p.145, **V-A.1.2.** p.145, **V-A.1.3.** p.145, **V-A.1.4.** p.146

Expérience LC-MS/MS : **V-A.3.1. b** p.148, **V-A.5.2.** p.153

Recherche dans les banques et validation : **V-A.6.1. a** p.154, **V-A.6.2. b** p.155

De la même façon, toutes les fractions de peptides n'apportent pas les mêmes informations. Globalement, la fraction de peptides 5-10kDa permet d'identifier un nombre de peptides différents plus important mais les autres fractions restent complémentaires. Ainsi, la fraction 5-10kDa permet de caractériser près de 60% des peptides identifiés par l'ensemble des analyses (Figure 20). Pour les échantillons issus de *stripping* itératif où un seul tampon pouvait être utilisé, nous avons donc choisi d'extraire les peptides dans un mélange PBS 10mM acidifié par de l'acide formique. D'autre part, dans la mesure où nous avions 30 échantillons à analyser en triplicat (soit au minimum 225h de LC-MS/MS = 30 x 3 x 1h30 + blancs intermédiaires), nous avons donc décidé d'analyser uniquement les fractions 5-10kDa et inf 5kDa. Globalement ces deux fractions doivent pouvoir nous permettre de caractériser 80% des peptides. De plus, l'analyse des échantillons de la fraction sup 10kDa génère très souvent un bouchage des précolonnes par la présence résiduelle de petites particules de peau.

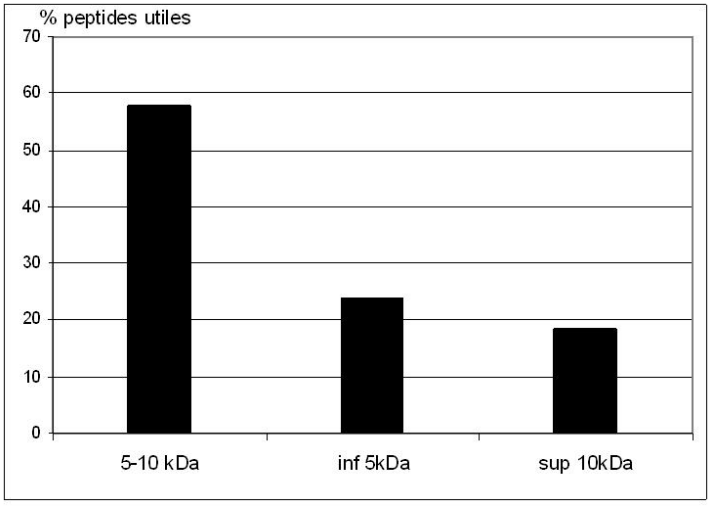


Figure 20 : Apport de chaque fraction dans les identifications.
Seuls les peptides nouvellement identifiés dans la fraction n par rapport aux fractions 1 à n-1 ont été comptabilisés.

III-A.2. Chromatographie monodimensionnelle : Elimination des polymères

Protocoles de référence :

Préparation des échantillons standard : **V-A.2** p.149

Expérience LC-MS/MS : **V-A.3.2. a** p.149**, V-A.5.3.** p.154

Recherche dans les banques et validation : **V-A.6.1. b** p.155, **V-A.6.2. c** p.156

III-A.2.1. Introduction

En routine, la spectrométrie de masse en tandem est classiquement utilisée pour identifier et séquencer des protéines et des peptides du fait de sa flexibilité, de sa sensibilité, de sa rapidité et de sa précision. Cependant les analyses de peptides peuvent être perturbées par certains composés connus [193]. Des études précédentes ont rapporté la détection de **pics interférents en HPLC**, provenant soit des produits chimiques [194], soit des tubes d'injection contenant l'échantillon ou des équipements [195], soit de problèmes de mélange des solvants ou de plastiques [196]. Même si les produits chimiques et les solvants sont choisis minutieusement pour être compatibles avec l'analyse par spectrométrie de masse, certains contaminants environnementaux peuvent provenir de l'équipement ou des matériaux de construction du laboratoire [197, 198] et interférer avec les analyses, spécialement lorsque l'on travaille avec de très faibles quantités de matériel de l'ordre de la femtomole ou moins [199]. Certains contaminants sont observés indépendamment de tout système chromatographique et sont attribués à des **contaminants volatils de l'air comme les polydimethylcyclosiloxanes** (PMC) [200].

Certaines impuretés, comme les **polyéthylène glycols** (PEGs) sont, au contraire, seulement détectées dans les analyses par LC-MS avec des temps de rétention spécifiques. Ces impuretés doivent être éliminées en priorité. La contamination aux PEGs est d'autant plus importante dans le cas d'étude à haute sensibilité et haute résolution. Les PEGs sont probablement enrichis sur la phase stationnaire et font diminuer la sensibilité de détection ou bien masquent l'élution des composés d'intérêt [194]. Ces contaminants de PEG sont observés indépendamment du système chromatographique utilisé, des solvants et des produits chimiques employés et même dans des laboratoires distincts. Nous nous sommes attachés à trouver une solution pour les éliminer. Notre montage expérimental inclut des spectromètres de masse sensibles et précis avec une haute résolution (LTQ-FT et/ou LTQ-Orbitrap, Thermo Fisher Scientific, CA, USA) couplés à des systèmes de nano-chromatographie (nanoLC). La solution la plus efficace a été l'utilisation de cartouche de purification de type micro-colonne de silice greffée C18 avec régénération en ligne. Cette modification simple et robuste nous a confirmé que ces impuretés étaient créées et enrichies au sein même du système nanoLC.

Cela nous a également montré que cette contamination est **nuisible à la qualité des analyses** du fait d'un fort effet de suppression d'ions. Ceci est un inconvénient majeur dans la mesure où le temps de rétention de ces polymères est dans la même gamme de temps de rétention que les peptides tryptiques d'intérêt. Après validation sur des échantillons standard, notre protocole a été appliqué sur tous les systèmes LC du laboratoire.

III-A.2.2. Caractérisation de la contamination

Comme indiqué ci-dessus, certains composés sont connus pour interférer avec le processus d'ionisation électrospray avec ou sans système chromatographique en amont. En effet les PMC (m/z 519.1389; $(Si(CH_3)_2O)_7$) sont observés indépendamment de toute séparation chromatographique et sont dus à des contaminants dans l'air. Ils peuvent être utilisés comme étalon interne pendant une analyse LC-MS en ligne [200].

En revanche, d'autres études récentes ont rapporté une contamination continue avec des impuretés qui interfèrent avec la qualité des profils LC et la qualité de la détection des ions en MS [194]. Cette contamination diminue la sensibilité de détection notamment du fait que ces polymères sont élués dans la même gamme de temps que les peptides d'intérêt et ceci indépendamment de l'instrument MS ou LC utilisé. Nous avons observé le même motif quel que soit l'échantillon analysé, avec nos systèmes Dionex et Agilent au laboratoire, mais également dans des analyses indépendantes dans un autre laboratoire en Allemagne avec un système Thermoelectron et des produits chimiques totalement différents. Comme il ne nous était pas possible de mettre en évidence l'origine de la contamination de façon claire, nous nous sommes attachés à éliminer efficacement ces impuretés. Les pics en question sont observés dans tous les chromatogrammes RPLC avec les rapports m/z 520.333, 564.359, 608.385, 652.411, 696.438 et un rapport S/B jusqu'à 2 000 (Figure 21A).

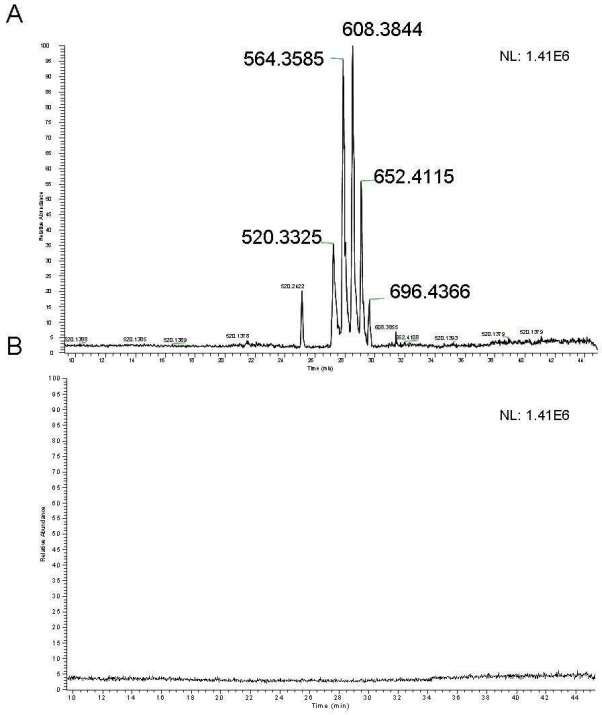


Figure 21 : Chromatogramme extrait des ions de rapport m/z 520,333, 564,359, 608,385, 652,411, 696,438 obtenu sur un blanc sans (A) ou avec (B) cartouche de purification sur un système 1D nanoLC Dionex.

Les pics détectés sont séparés de 44.026 Da ce qui permet d'identifier ces contaminants comme du PEG [201] et ils couvrent environ 4 minutes sur chaque chromatogramme. Ils sont élués par un pourcentage d'ACN entre 35 et 40% et leurs caractéristiques chromatographiques sont rappelées dans le Tableau 5.

Tableau 5 : Caractéristiques chromatographiques des contaminants de PEG sur un système 1D RPLC Dionex.
Les temps de rétention sont donnés à 0,2min près et le pourcentage d'ACN à 0,2% près.

MH+ des polymères	t_R (min)	k	% ACN	N(m^{-1})
520,333	27.2	1.95	36.2	405 000
564,358	27.8	2.02	36.9	793 000
608,385	28.5	2.09	37.8	750 000
652,412	29.1	2.16	38.6	1 080 000
696,438	29.6	2.21	39.2	1 260 000
740,465	30.1	2.26	39.8	925 000
784,491	30.6	2.32	40.5	1 350 000

Ces contaminants ne sont pas détectés directement en ESI-MS ni dans aucun solvant ou produit chimique ni lorsque les analyses sont réalisées sans colonne de préconcentration. Nous avons testé différents systèmes chromatographiques (Ultimate 3000 Dionex et 1200 LC Agilent), différents acides (type, lot, fournisseur) sans réel changement. Une purification hors ligne des solvants a été évaluée sans succès sur une colonne C18 capillaire (2,1mm d.i. 30cm de longueur). Il est difficile voire impossible de conclure quant à l'origine de cette contamination : *bleeding* dû au passage des solvants ou contamination commune due aux récipients de stockage ?

III-A.2.3. Elimination de la contamination

Comme rapporté dans des études récentes [194], ces polymères disparaissent lorsque la précolonne de concentration est enlevée : ces contaminants seraient donc enrichis sur la phase stationnaire de la précolonne à haut débit (15µL/min). Nous avons modifié notre montage LC en introduisant une cartouche de purification, *i.e.* une micro-précolonne C18 (C18 PepMap100, 5µm, 100 Å, 1 mm d.i., longueur 15 mm, Dionex) entre la sortie de la pompe de chargement et la vanne d'injection pour éliminer les polymères de façon continue (Figure 22) [202].

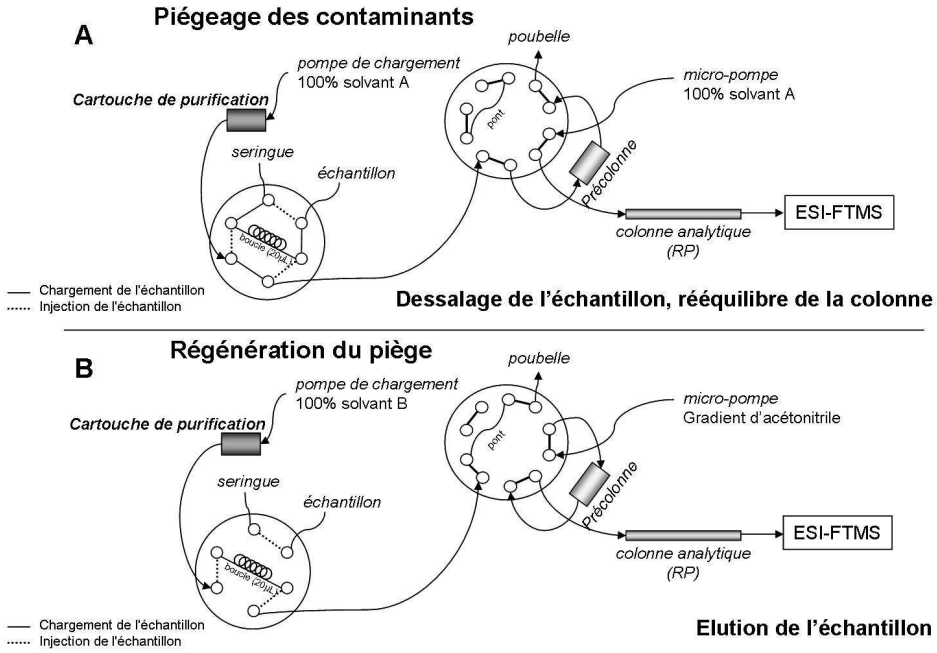


Figure 22 : Schéma du montage LC1D développé pour l'élimination des polymères.
A, les contaminants sont piégés sur la cartouche de purification pendant le chargement l'injection et le dessalage de l'échantillon. B, la cartouche de purification est régénérée avec du solvant B pendant l'élution de l'échantillon et le gradient RP.

Pour vérifier l'efficacité du système, des blancs ont été enregistrés après nettoyage de l'ensemble du système par un mélange ACN/MeOH/IPr (1/1/1). Pour éviter la saturation en contaminant, la cartouche de purification est nettoyée efficacement avec du solvant B pendant 25min en parallèle du gradient analytique. Ainsi, les contaminants sont éliminés de façon permanente (Figure 21B).

Cette modification a également été évaluée sur l'injection de 20fmol d'un mélange protéique standard (LCP) (Figure 23A et B). Avant tout changement, les pics contaminants sont observés sur le chromatogramme extrait en MS, chromatogramme normalisé sur le pic de base. La série des pics de PEG est observée avec un rapport S/B jusqu'à 4000. Le lysozyme C et le cytochrome C ne sont identifiés que par un seul peptide (Tableau 6). De plus, seulement 18 peptides ont été validés au total. En revanche, après introduction de la cartouche de purification, les pics interférents sont largement diminués à un rapport S/B inférieur au signal MS du bruit de fond dû aux ions PMC (Figure 23B). Comme décrit précédemment, ces contaminants sont enrichis sur la phase stationnaire et diminuent la sensibilité d'analyse notamment parce qu'ils sont élués en même temps que les peptides d'intérêt [194]. Mais ils peuvent également limiter la capacité de la précolonne. En effet, après introduction de la cartouche de purification, 5 des 6 protéines du mélange sont identifiées selon les standards de publication avec au moins 2 peptides différents en premier candidat et un score minimal de 30 [116] et 39 peptides différents sont validés pour l'ensemble des protéines (Tableau 6). Ce test a été répété

régulièrement pour vérifier l'efficacité et la robustesse de notre montage depuis maintenant plus d'un an sans changer la cartouche de purification.

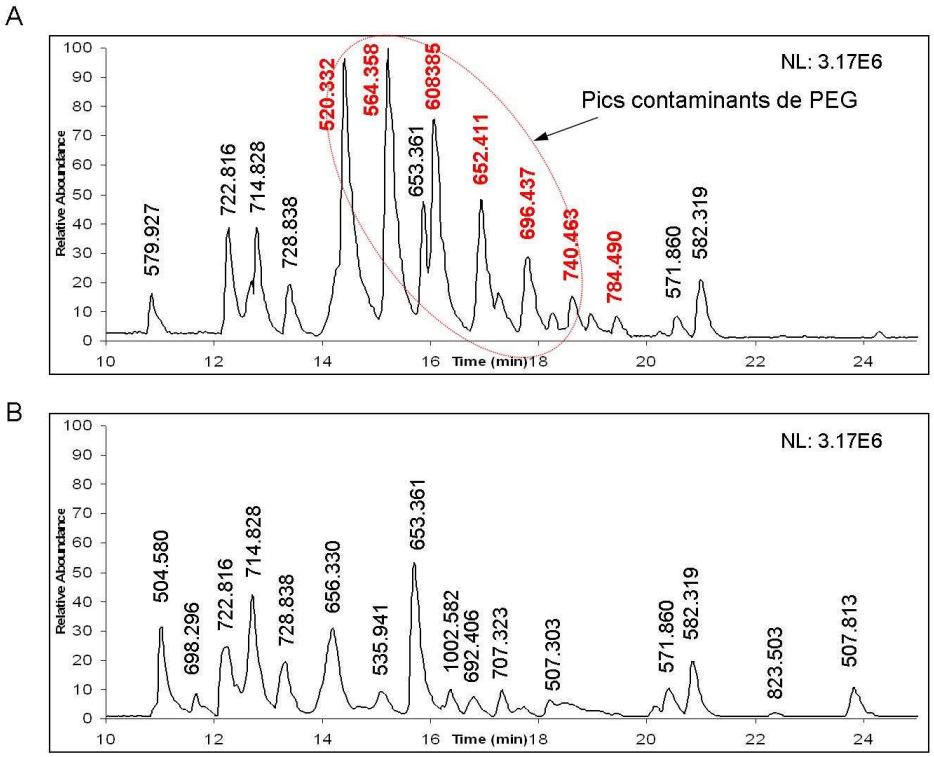


Figure 23 : Chromatogramme en pic de base avec extraction des scans en MS simple d'une injection de 20fmol de protéines sans (A) ou avec (B) cartouche de purification sur un système Dionex (colonne Agilent avec gradient approprié).

Tableau 6 : Identification en triplicat de 20fmol d'un mélange de protéines sans et avec cartouche de purification.

Protéine	Sans cartouche de purification		Avec cartouche de purification	
	Score Prot.	# peptides	Score Prot.	# peptides
ALBU_BOVIN	401 ± 20	7 ± 1	631 ± 4	14 ± 1
TRFE_BOVIN	236 ± 148	4 ± 2	632 ± 46	12 ± 1
BGAL_ECOLI	149 ± 21	3 ± 1	194 ± 27	4 ± 1
ADH1_YEAST	139 ± 19	2 ± 0	298 ± 10	5 ± 0
LYSC_CHICK	74 ± 3	1 ± 1	75 ± 3	1 ± 0
CYC_BOVIN	49 ± 43	1 ± 1	196 ± 46	3 ± 1

III-A.3. Chromatographie multidimensionnelle : optimisation

Protocoles de référence :

Préparation des échantillons digérés : **V-A.2** p.149

Expérience LC-MS/MS : LC2D *on-line* **V-A.4.1.** p.150, LC2D *off-line* **V-A.4.2.** p.152, LC1D **V-A.3.2. a** p.149, FT-MS **V-A.5.2.** p.153

Recherche dans les banques et validation : **V-A.6.1. b** p.155, **V-A.6.2. c** p.156

III-A.3.1. Elimination des polymères

Des études récentes ont également rapporté le même type de problème durant des analyses en LC2D MS qui pouvait être dû à du PEG contaminant les solutions commerciales [194]. Nous avons appliqué le même principe à notre montage de LC2D *on-line* et *off-line* car les pics interférents sont également observés dans les chromatogrammes RP de chaque fraction SCX. La série des PEGs est observée avec un rapport S/B jusqu'à 1 300 (Figure 24A). Nous avons donc commencé par insérer une cartouche de purification sur le circuit RP après la pompe de chargement 2 comme en LC1D (Figure 25). Cette cartouche est régulièrement régénérée à 40µL/min par du solvant B en parallèle de chaque gradient RP. Le montage LC2D entier est nettoyé avec un mélange ACN/MeOH/IPr (1/1/1) et des blancs sont réalisés. La série de PEGs est toujours observée mais beaucoup plus faiblement en intensité avec un rapport S/B de 300 (Figure 24B). Si les contaminants sont concentrés sur la phase stationnaire, il est également nécessaire de les éliminer de la première dimension du système LC2D. Une seconde cartouche a donc été insérée après la micro pompe 1 (Figure 25). Celle-ci est régulièrement régénérée après la dernière fraction SCX. Les PEGs sont encore observés mais avec un rapport S/B bien plus faible de 50 (Figure 24C) et bien en dessous du signal des ions PMC (m/z 536.170, S/B de 200). Des études récentes ont suggéré que la poudre de formiate d'ammonium commercial utilisée pour préparer le tampon salin pouvait contenir du PEG [194] venant du contenant en plastique. De ce fait nous avons changé la façon de préparer le tampon en utilisant de l'acide formique 99-100%, de l'ammoniaque en solution à 25% et du TFA. Ainsi les impuretés ne sont plus du tout détectées (Figure 24D) dans les chromatogrammes RP de n'importe quelle fraction SCX lorsque le nouveau tampon est utilisé combiné avec les cartouches de purification.

Sur le circuit *off-line*, dans la mesure où la micro pompe 1 assure à la fois le gradient salin et le chargement des fractions vers le montage RP, une seule cartouche de purification est nécessaire : celle placée au niveau de la vanne externe (Figure 76).

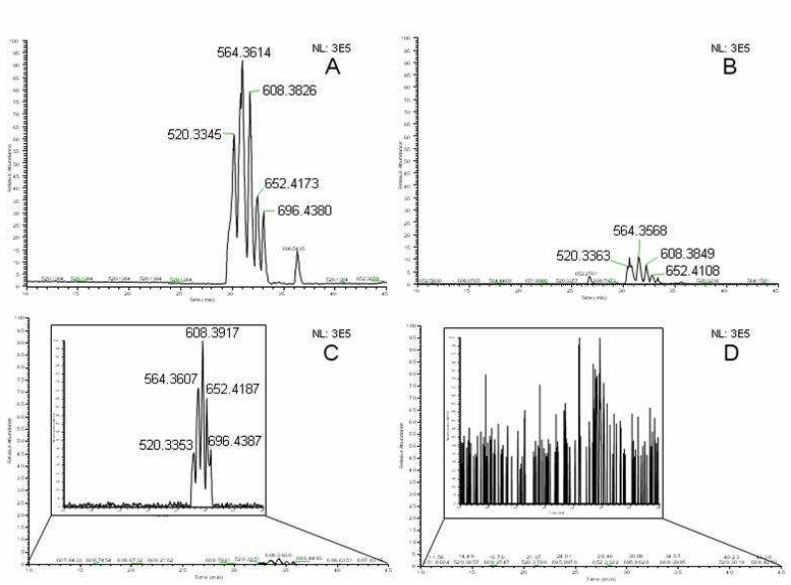


Figure 24 : Chromatogrammes extraits des ions de rapport m/z 520,333, 564,359, 608,385, 652,411, 696,438 obtenus sur un blanc en LC2D *on-line.*
Sans cartouche (A), avec une seule cartouche de purification sur la section RP (B), avec deux cartouches de purification sur la section RP et sur la section SCX (C), et avec une nouvelle solution de formiate d'ammonium combinée aux cartouches de purification (D).

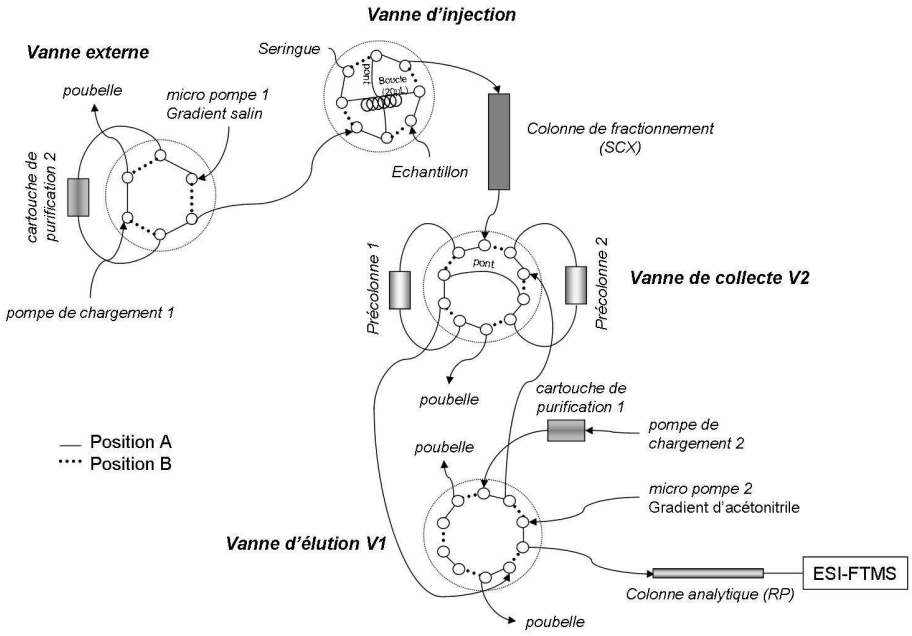


Figure 25 : Schéma du montage LC2D *on-line* développé pour éliminer les polymères.
Vanne d'injection : en position A, l'échantillon est chargé et en position B, injecté sur la colonne SCX. Vanne externe : en position A, la micro pompe 1 délivre le gradient salin pour séparer les peptides sur la colonne SCX et les contaminants sont piégés sur la cartouche de purification 2, en position B ; la cartouche est régénérée après la dernière fraction SCX avec un haut débit de solvant B venant de la pompe de chargement 1. Vanne de collecte V2 : en position A, les peptides sont collectés sur la précolonne 1 pendant le dessalage et l'élution de la précolonne 2 ; en position B c'est l'inverse. Vanne d'élution V1 : en position A, la pompe de chargement 2 est utilisée pour dessaler, la colonne RP est en équilibrage et les contaminants sont piégés sur la cartouche purification 1 ; en position B, la micro pompe 2 délivre le gradient d'ACN, la pompe de chargement 2 délivre du solvant B à haut débit pour régénérer la cartouche de purification 1.

III-A.3.2. Mise au point du protocole *on-line*

Comme nous l'avons décrit dans la partie introductive, il existe différentes configurations de couplage *on-line* utilisant soit une seule colonne biphasique soit deux colonnes avec un jeu de vannes. Dans la mesure où nous voulions comparer les protocoles *on-line* et *off-line*, nous avons opté pour la solution des deux colonnes découplées. Cependant, les protocoles décrits dans la littérature réalisent l'élution de la colonne SCX de manière discrète par impulsions salines et n'utilisent qu'une seule pompe gradient pour la 2ème dimension. Après chaque impulsion saline, l'élution est suspendue. Cela permet d'avoir moins de contrainte de temps en 2ème dimension qui peut ainsi durer le temps souhaité. En contrepartie, cela aboutit à des temps d'attente des analytes très long sur la colonne SCX, ce qui diminue l'efficacité de séparation de la 1ère dimension. Les pics risquent de s'élargir du fait de la diffusion des analytes le long de la colonne de 1ère dimension pendant cette pause.

Pour optimiser au mieux la séparation de 1ère dimension, nous avons donc choisi de travailler en condition de gradient sur les deux dimensions en utilisant un montage faisant intervenir deux pompes gradient, une vanne sur laquelle sont montées deux précolonnes RP C18 et une autre vanne pour le couplage avec la colonne analytique RP. Chaque fraction saline dure le temps du gradient RP soit 71 minutes mais les peptides élués de la colonne SCX sont focalisés sur la précolonne RP en ligne. Nous nous sommes contentés d'optimiser les protocoles de la littérature par impulsions salines pour optimiser notre gradient sur un échantillon de cils faciles à collecter au laboratoire. Cependant l'adaptation n'est pas directe dans la mesure où sont généralement précisés dans les protocoles la concentration en sels du tampon de la fraction n, le débit utilisé pour charger les sels mais pas la durée de l'impulsion [203] : il est donc difficile d'en déduire la quantité totale de sels qui passent à travers la colonne pendant la faction n.

a Influence de la concentration en sels du tampon salin

Tout d'abord, nous avons travaillé sur un gradient linéaire de 0 à 20% de tampon B' (1M de formiate d'ammonium) en 10 fractions de 71 minutes puis une montée à 50% de B', puis à 100% et enfin un plateau à 100% B'. Contrairement à ce à quoi nous nous attendions, les peptides n'ont pas été élués linéairement en fonction des fractions : au bout de 5 fractions, la quasi-totalité de l'échantillon est éluée. De fait, comme chaque fraction dure 71 minutes, la quantité totale de sels est un paramètre très important car les conditions d'élution se rapprochent du mode isocratique. Nous avons donc drastiquement diminué la concentration en sels du tampon B' à 100mM pour que la totalité du temps de séparation puisse être exploitée (Figure 26A et B).

b *Influence du gradient*

Afin d'optimiser le gradient, nous avons supposé qu'entre 2 fractions successives, le nombre de peptides élués était proportionnel à la quantité de sels qui avait traversé la colonne : ainsi nous avons calculé le débit et le pourcentage de tampon B' nécessaire à chaque fraction. Ainsi, le gradient optimisé permet d'éluer les peptides linéairement de manière très reproductible. Ce gradient n'est pas un gradient linéaire mais est composé d'une succession de gradients de différentes pentes ; le débit utilisé varie au cours du temps : élevé au départ, il diminue progressivement puis ré-augmente pendant le palier à 100% B' pour éluer les derniers peptides de la colonne (Figure 26B et C).

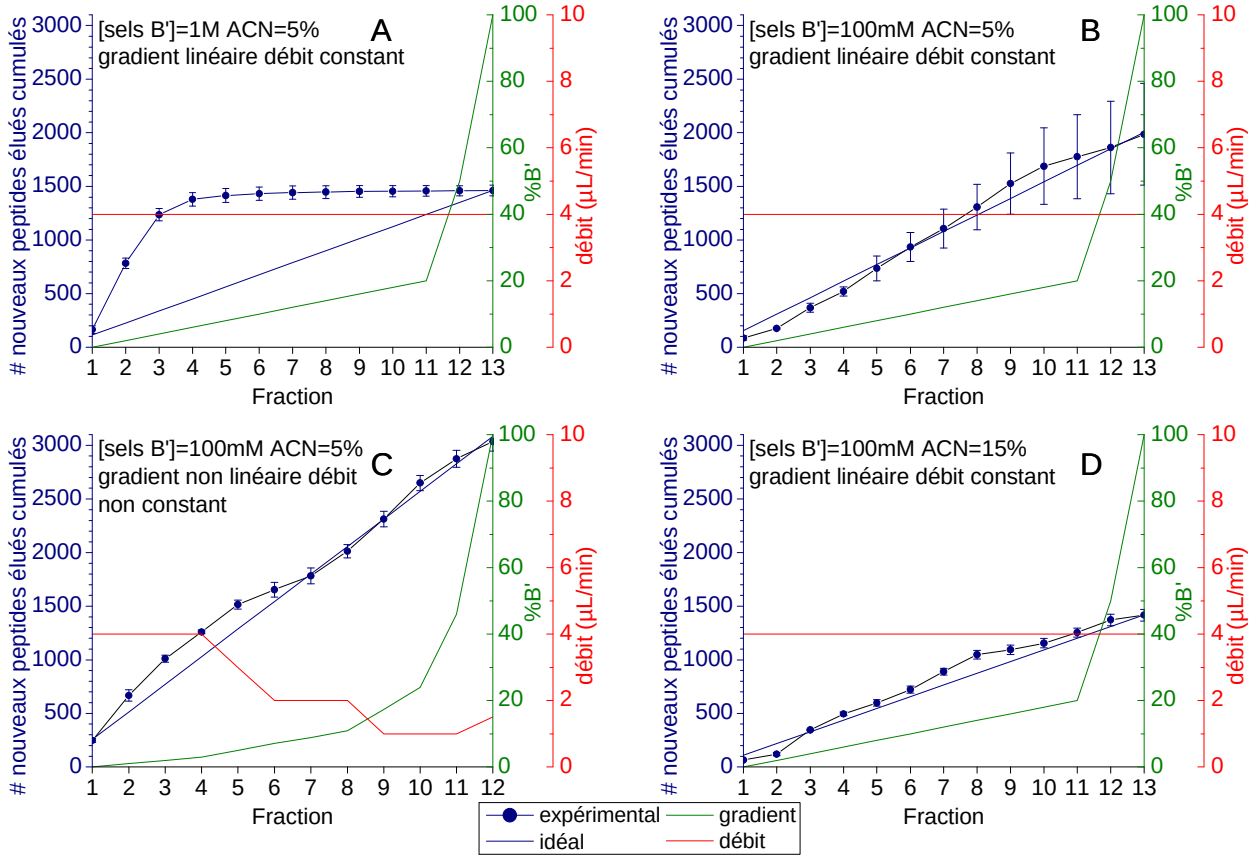


Figure 26 : Comparaison des différents protocoles sur un échantillon de cils digérés. Elution de nouveaux peptides cumulés en fonction des fractions SCX.
A : Concentration en sels et gradient linéaire directement déduits des protocoles d'impulsions salines décrits dans la littérature ; plus aucun peptide n'est élué après la 5ème fraction. B : Utilisation d'un tampon 10 fois moins concentré en sels ; l'ensemble des fractions est utile et 28% de peptides identifiés en plus ; manque de répétabilité. C : gradient optimisé où le débit varie au cours de l'élution SCX et le gradient n'est pas linéaire ; 54% de peptides identifiés en plus par rapport à B ; très bonne répétabilité. D : Augmentation de la concentration en ACN des tampons ; amélioration de la répétabilité par rapport à B mais perte de 29% des peptides identifiés par rapport à B.

c *Influence de la concentration en acétonitrile*

Comme nous l'avons évoqué préalablement, les colonnes d'échange de cations présentent des phénomènes d'interactions secondaires hydrophobes [92] que l'on peut limiter en ajoutant un solvant

organique aux tampons de 1ère dimension. Nous avons donc évalué si l'ajout d'ACN était compatible dans une configuration LC2D *on-line* et s'il permettait d'améliorer la séparation. Augmenter la concentration en ACN permet d'augmenter la résolution : ainsi avec 25% d'ACN, quasiment tous les peptides sont focalisés dans une seule fraction. Cependant, en augmentant cette concentration, la rétention des peptides sur les précolonnes RP C18 diminue. Avec 25% d'ACN, près de 85% des peptides sont perdus ! Même avec 15%, 29% des peptides ne sont pas retenus (Figure 26B et D). Au final, notre protocole optimisé en configuration *on-line* a été validé pour des tampons contenant 5% d'ACN. En 2006, Liu *et al.* ont décrit un protocole par impulsions salines où les tampons des dernières fractions contiennent jusqu'à 30% d'ACN [204]. Ce protocole leur permet de récupérer des peptides très hydrophobes non élués dans les fractions précédentes et pourtant de charges comparables. Dans notre cas, nous aurions pu utiliser la 3ème voie de la pompe réalisant le gradient salin pour augmenter la concentration en ACN en toute fin de gradient mais nous avons choisi de garder cette voie en eau pour nettoyer le système et éliminer les sels à la fin de chaque *run* LC2D, ce qui permet d'éviter la formation de cristaux de sels lorsque le système est à l'arrêt. En effet, ces cristaux peuvent rayer les stators et les rotors au niveau des vannes et créer des volumes morts importants par rapports aux débits utilisés (notamment nano dans la partie RP).

III-A.3.3. Mise au point du protocole *off-line*

La même démarche a été adoptée pour mettre au point les conditions d'élutions optimales en configuration *off-line*. A partir d'un gradient linéaire de 0 à 40% de tampon B' (1M de formiate d'ammonium) où les peptides étaient élués quasiment dans une seule fraction, nous avons diminué la concentration du tampon à 100mM et adopté un gradient de 20% B' à 70% B'. De plus l'injection est réalisée en condition légèrement saline (10mM) afin d'éluer les peptides dès les premières fractions. Ces conditions permettent de distribuer l'élution des peptides régulièrement en fonction des fractions.

A partir de ces conditions, nous avons modifié la fréquence d'échantillonnage des fractions et la concentration en ACN pour en mesurer l'influence.

a Influence de la concentration en acétonitrile

En configuration *off-line*, il est possible d'utiliser de l'ACN puisque entre la collecte et la réinjection des fractions sur le circuit RP, il est possible de retraiter les fractions et notamment d'évaporer l'ACN. Nous avons donc ajouté différentes concentrations d'ACN aux tampons A' et B' : 5%, 25%, 75%. L'utilisation de 25% d'ACN permet notamment d'éluer plus de peptides dans les dernières fractions alors qu'avec 5% d'ACN, les dernières fractions n'apportent rien (Figure 27). A même quantité de sels traversant la colonne, l'ACN doit donc permettre de décrocher des peptides plus hydrophobes de la

phase stationnaire. En revanche, pour une concentration de 75%, la pente d'élution des peptides est plus rapide sans gain au niveau du nombre de peptides total identifiés.

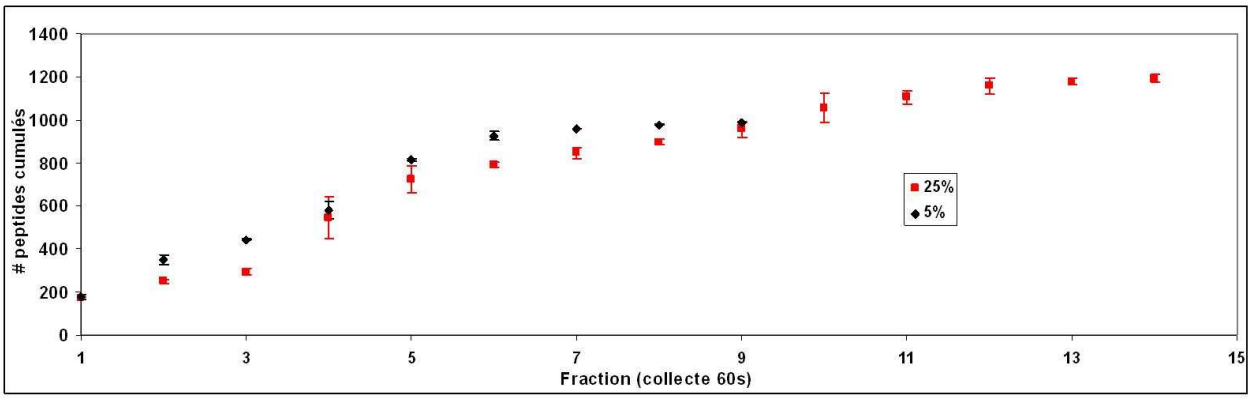


Figure 27 : Influence de la concentration en ACN sur l'évolution du nombre de peptides élués en fonction des fractions.

b Influence de la période d'échantillonnage

La période d'échantillonnage correspond à la durée de collecte de chaque fraction SCX. Les résultats de Horie [205] montrent que la durée optimale en LC2D doit être ajustée entre 2,2 et 4 fois l'écart-type des pics. Cette période d'échantillonnage peut être un facteur important dans l'évaluation de la capacité de pic de 1ère dimension. En effet, selon la fréquence d'échantillonnage choisie, des pics séparés en théorie peuvent être collectés dans la même fraction d'où un élargissement apparent des pics de 1ère dimension et donc une diminution de la capacité de pic réelle de 1ère dimension que l'on peut évaluer à $P_c^{D1'} = P_c^{D1}.\frac{\sigma}{\sigma^*}$ où σ^* est l'écart-type moyen des pics de 1ère dimension mais observé dans une configuration 2D [206]. Pour des pics d'environ 1min à la base, cela correspondrait donc à une durée de collecte entre 33 et 60s. Dans un premier temps, nous avons choisi une durée de collecte de 60s : nous nous sommes rendus compte qu'en moyenne chaque peptide était élué dans 2 fractions successives (Figure 28), ce qui impliquait *a priori* que notre résolution de 1ère dimension était surestimée. Ceci peut notamment être dû au fait que les volumes d'injection d'échantillon biologique sont importants : l'injection sur la colonne ne peut pas être considérée comme ponctuelle (5-10µL injecté sur une colonne de 3,8µL) d'où un élargissement des pics. Nous avons augmenté la durée de collecte à 120s sans modification notable du nombre de fractions successives où est élué chaque peptide (Figure 28). Il est donc plus probable que la présence du peptide dans la fraction n+1 soit due à une adsorption non spécifique de certains peptides en amont de la colonne de séparation RP (phénomène qui aurait d'ailleurs tendance à s'accroître avec la durée de collecte puisque chaque fraction est plus riche en peptides car provenant d'un volume de collecte plus important).

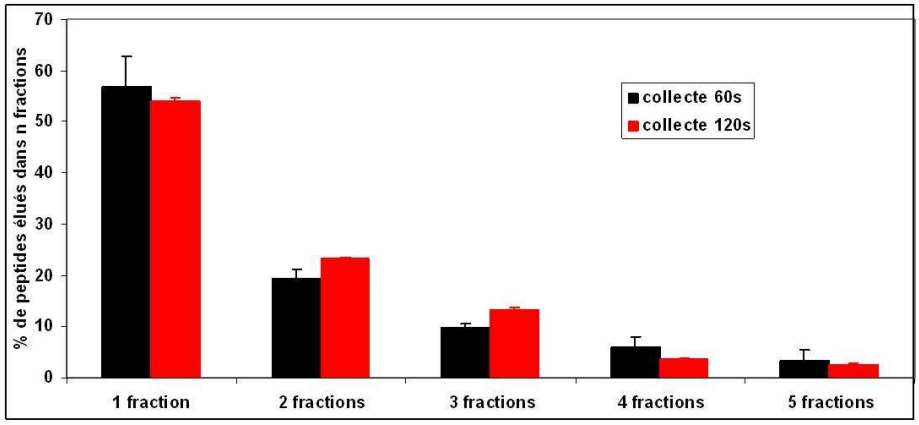


Figure 28 : Influence de la période d'échantillonnage sur l'élution des peptides en 1ère dimension. Plus de 50% des peptides sont élués en une fraction mais certains peptides parmi les plus abondants se retrouvent dans 5 fractions consécutives.

III-A.3.4. Comparaison des protocoles

Même si le couplage SCX-RP est largement utilisé dans les études protéomiques, très peu d'études ont été menées pour comparer de manière systématique les différentes approches que ce soit en termes de capacité de pic ou en termes d'identifications peptidiques ou protéiques. Afin de comparer de manière cohérente les données obtenues, la même quantité de protéines digérées a été analysée avec chaque protocole. De plus, chaque analyse a été réalisée avec un nombre comparable de fractions SCX (13-15), les mêmes conditions LC-MS/MS et des paramètres de recherche et de validation identiques. La capacité de pic de chaque configuration a été évaluée comme premier indicateur de l'efficacité des séparations. Nous avons également envisagé les divers paramètres pouvant expliquer les différences observées en termes de nombre total de protéines et de peptides identifiés : pouvoir de résolution, orthogonalité, perte d'échantillon, caractéristiques physico-chimiques des peptides identifiés par chaque méthode *etc.*

a Caractéristiques physico-chimiques des peptides identifiés

Dans un premier temps, nous avons examiné le profil d'élution des peptides de la colonne SCX en fonction de leur charge. Au pH utilisé pour la séparation (2,5), seules les lysines, arginines, histidines et le N-terminal du peptide contribuent à la charge des peptides. Comme le montre la Figure 29 (A1/B1), les peptides sont effectivement séparés en fonction de leur charge quelque soit la configuration utilisée : globalement les peptides bichargés sont élués dans les premières fractions alors que les peptides 4 et 5 fois chargés ne sont observés qu'à partir des dernières fractions. Si on observe la répartition de l'indice hydrophobique des peptides, calculé suivant l'échelle de Deber basée sur les temps de rétention RPLC de peptides hydrophobiques [207], en fonction des fractions, il n'apparaît pas de différence flagrante entre les configurations : des peptides d'hydrophobie très variable sont élués dans toutes les fractions

(Figure 29 A2/B2). Cependant, en configuration *off-line,* la moyenne de l'indice hydrophobique des peptides, dans une fraction pour chaque état de charge, ne dépend pas du n° de la fraction mais juste de l'état de charge considéré. En revanche en configuration *on-line*, l'indice moyen décroît en fonction des fractions. Dans cette configuration, il y aurait donc une contribution hydrophobique durant l'élution de la colonne SCX. Cette différence est probablement due à l'utilisation de 25% d'ACN dans la configuration *off-line* permettant de rompre les interactions hydrophobiques alors qu'en configuration *on-line*, nous n'utilisons que 5% d'ACN.

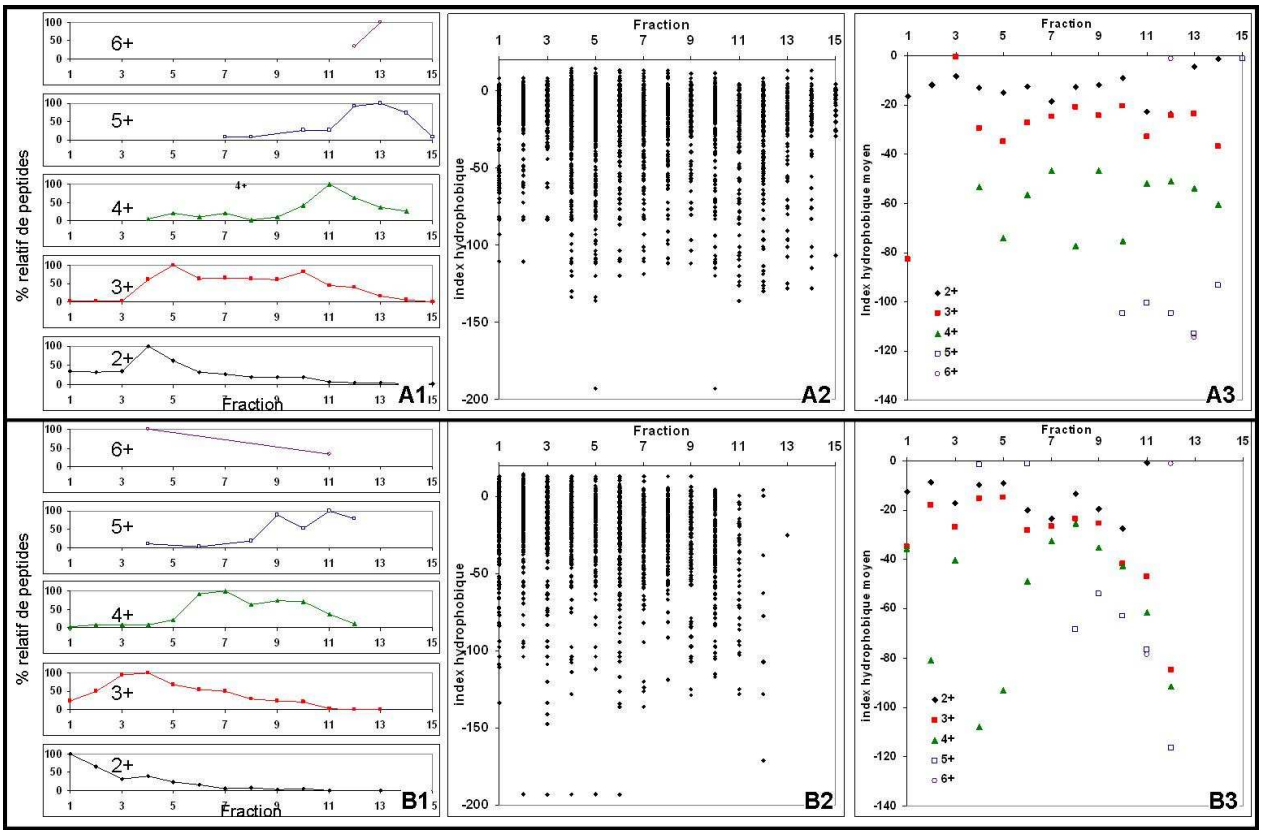


Figure 29 : Répartition des peptides en termes de charge (graphes 1), d'indice hydrophobique [207] (graphes 2) en fonction des fractions selon la configuration utilisée (A : *off-line*, B : *on-line*). L'indice hydrophobique moyen des peptides de chaque fraction triés d'après leur charge est également présenté (graphes 3).

b Orthogonalité

En termes d'orthogonalité, on peut représenter le temps de rétention normalisé ($\frac{t_R - t_{R\min}}{t_{R\max} - t_{R\min}}$) de 1ère dimension en fonction du temps de rétention normalisé de 2ème dimension. Si une relation quelconque, linéaire ou non, apparaît entre les deux dimensions, cela implique que les deux dimensions ne sont pas indépendantes. Que ce soit en configuration *off-line* ou *on-line*, il n'apparaît pas de lien entre les temps de rétention des deux dimensions : les deux dimensions sont donc relativement orthogonales même si l'ensemble de l'espace de rétention n'est pas utilisé (Figure 30). En particulier, le mode *off-line*

présente un vide plus important que le mode *on-line* dans la partie des temps de rétention élevés, c'est-à-dire des peptides les plus hydrophobiques. Il est possible que même si l'utilisation de 25% ACN permet d'éluer plus de peptides hydrophobiques de la colonne SCX, le mode *off-line* souffre d'un biais dû à la collecte et à l'étape d'évaporation : une grande partie des peptides hydrophobes élués sont retenus dans les puits de collecte par interactions non spécifiques et sont perdus par la suite. En configuration *on-line*, il semble que les peptides hydrophiles soient plus facilement perdus (en particulier dans les premières fractions). En effet, ces peptides peuvent être perdus de deux manières :

- du fait de la présence de 5% d'ACN dans le tampon salin, les peptides les plus hydrophiles pourraient être non retenus sur la précolonne RP,
- mais il est plus probable que des peptides, retenus sur la précolonne RP même avec 5% d'ACN, soient tout de même perdus par une élution isocratique de cette précolonne dans la mesure où la collecte dure 71min (volume de solvant aqueux qui traverse la précolonne = tampon salin + solvant A de lavage = 350µL soit 2 fois plus qu'en mode 1D classique).

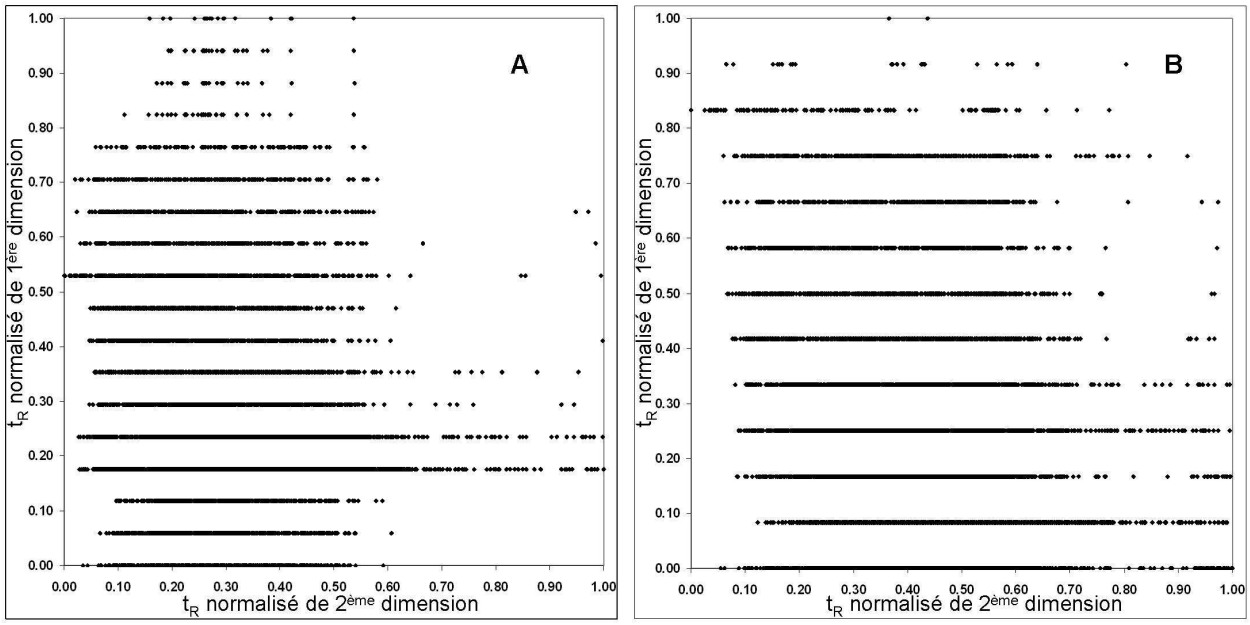


Figure 30 : Orthogonalité des deux dimensions en mode *off-line* (A) et en mode *on-line* (B).
Chaque peptide élué est représenté par un point de coordonnées son temps de rétention normalisé de 2ème dimension en abscisse et son temps de rétention normalisé de 1ère dimension en ordonnée. Aucune relation triviale n'apparaît entre les temps de rétention de 1ère et 2ème dimension ce qui met en évidence la bonne orthogonalité des deux modes de séparation.

c Capacité de pic

En mode *off-line*, le premier pic élué de la colonne SCX a pour temps de rétention $t_{R\,\min}^{D1} = 2\,\text{min}$ et le dernier pic $t_{R\,\max}^{D1} = 17\,\text{min}$. Plus de 60% des peptides ne sont identifiés que dans une seule fraction. Si on suppose que la largeur moyenne des pics chromatographiques est de 1min, la capacité de pic

obtenue en première dimension est donc de 15. En mode *on-line*, le premier pic est élué à $t_R^{D2}{}_{\min} = 17\,\text{min}$ et le dernier pic à $t_R^{D2}{}_{\max} = 869\,\text{min}$. Environ 80% des peptides sont présents dans une seule fraction qui résulte d'une collecte sur 71min, soit une capacité de pic en première dimension de 12.

La capacité de pic de deuxième dimension est la même quelle que soit la configuration, soit d'environ 80 (largeur de pic 30s, durée de gradient utile 40min). En théorie, il est donc possible d'obtenir des capacités de pic globales de 1 200 en configuration *off-line* et de 960 en *on-line*. Cependant, pour être rigoureux, il faut également tenir compte de l'orthogonalité des deux dimensions dans les deux configurations et notamment de l'utilisation de l'espace des temps de rétention. Comme nous l'avons représenté sur la Figure 30, les deux dimensions sont bien orthogonales mais l'espace des t_R n'est pas utilisé à 100% car les deux types de configurations souffrent de pertes de certains peptides spécifiquement. Selon Gilar [89], la capacité de pic effective correspond à la capacité de pic maximale (celle calculée par P^{D1} x P^{D2}) multipliée par le pourcentage d'espace des temps de rétention couvert. En configuration *on-line*, on peut évaluer géométriquement celui-ci à 81% alors qu'en configuration *off-line*, il n'est que de 66%. Les capacités de pic effectives ne sont alors plus que de 792 (*off-line*) et de 777 (*on-line*) ce qui est comparable.

d Identification peptidique et protéique

Cette similarité des capacités de pic se traduit, en termes d'identifications peptidiques par une multiplication par un facteur 2,5 en configuration *off-line* et 3,5 en configuration *on-line* du nombre de peptides identifiés en LC2D par rapport aux analyses en LC1D (Figure 31). Il est possible que cette différence, malgré des capacités de pic de même ordre de grandeur, soit due au fait qu'un peptide est cassé en moyenne 2 fois au cours de l'analyse LC2D *off-line* alors qu'il ne l'est que 1,38 fois en LC2D *on-line*. En mode *off-line,* l'instrument perd du temps MS/MS à fragmenter le même peptide alors qu'un peptide moins abondant élué dans une seule fraction SCX aurait pu être cassé, d'où des identifications peptidiques et protéiques moins performantes. Une manière de pallier cette difficulté serait d'exploiter le fait que les analyses sont faites en mode *off-line*. Il serait possible d'utiliser une liste d'exclusion de précurseurs faite à partir des peptides identifiés dans la fraction n et utilisée dans l'analyse de la fraction n+1. Cependant, cela rajouterait du temps d'interprétation entre les fractions et rallongerait donc la durée totale du *run* LC2D. On peut noter par ailleurs que seul 1/3 des peptides identifiés sont communs aux deux configurations *on-line* et *off-line* : 50% des peptides ne sont identifiés qu'en mode *on-line* et 17% en mode *off-line*. Considérant notre analyse de l'utilisation de l'espace des temps de rétention et de l'hydrophobie des peptides, cette différence peut être interprétée comme correspondant soit aux pertes en termes de peptides hydrophiles en mode *on-line*, soit aux pertes de peptides hydrophobes en mode *off-line*.

En termes d'identifications protéiques, les configurations LC2D permettent d'identifier de 2 à 3 fois plus de protéines qu'en mode 1D classique (Figure 31). Les résultats présentés correspondent à l'analyse d'un échantillon de *stratum corneum*. En LC1D, 60% des protéines identifiées sont des protéines du cytosquelette comme les kératines. Ces protéines constitutives diffèrent souvent des protéines de régulation moins abondantes. Avec la LC2D, la proportion des protéines du cytosquelette chute à 33% au profit d'autres catégories comme des enzymes ou des protéines impliquées dans les phénomènes d'adhésion des cornéocytes. L'utilisation de la LC2D permet ainsi d'augmenter la couverture du protéome de l'échantillon et d'identifier plus de protéines mineures. Elle permet aussi d'augmenter la couverture de séquence des protéines déjà identifiées en LC1D, en assurant une meilleure robustesse des identifications (Figure 32). Ainsi en moyenne, la couverture de séquence des protéines déjà identifiées en LC1D est multipliée par 2,4.

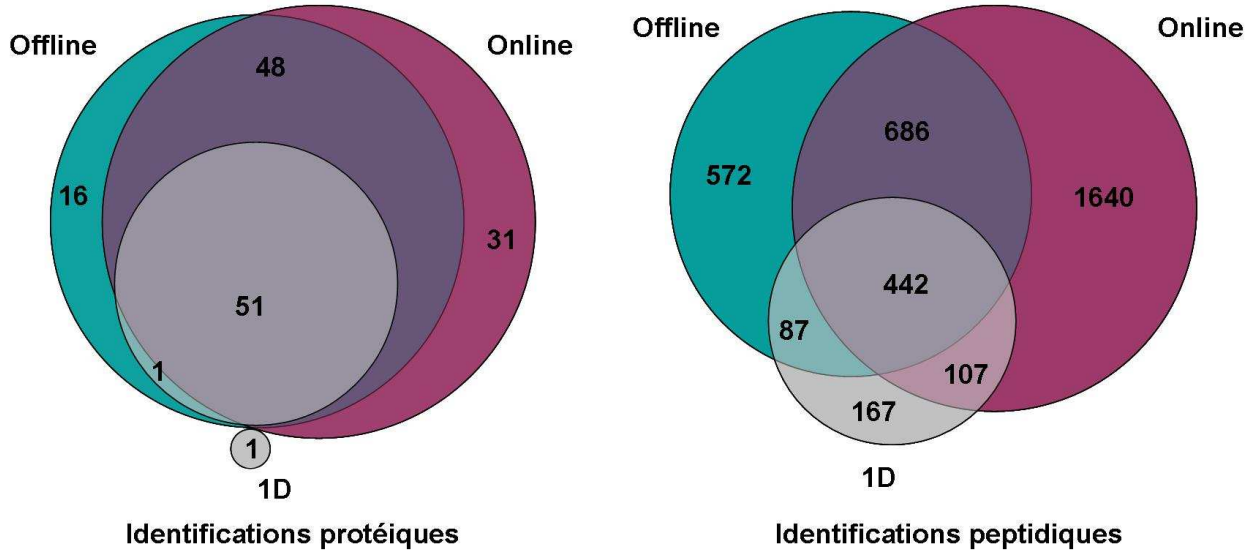


Figure 31 : Comparaison des identifications en termes de protéines et de peptides suivant les deux configurations LC2D et mise en évidence de l'intérêt par rapport à la LC1D.

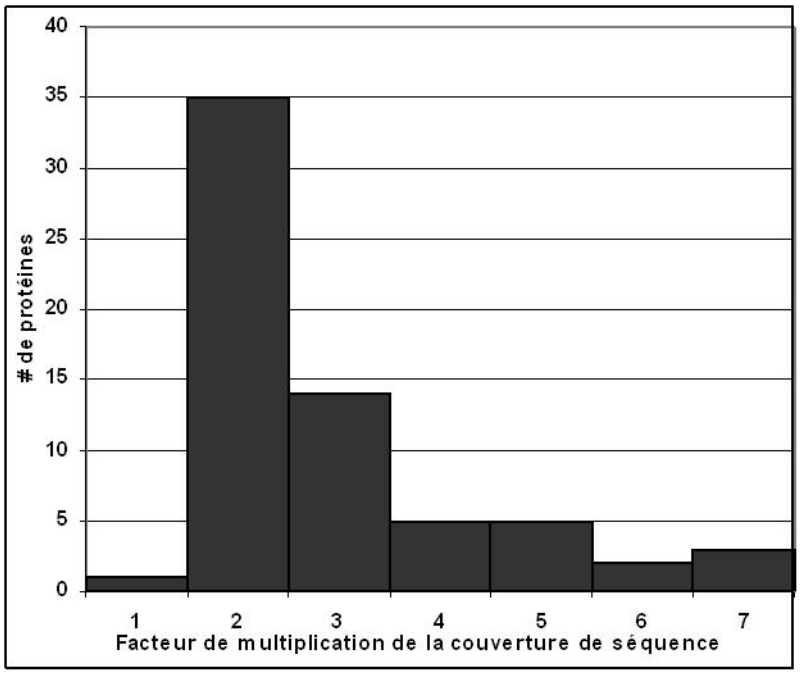


Figure 32 : Mise en évidence du gain en couverture de séquence des protéines déjà identifiées en LC1D.

III-A.4. Augmenter la couverture de protéome autrement

Pour augmenter la couverture du protéome de manière plus simple qu'en recourant à des protocoles de LC2D, d'autres possibilités, que nous avons également testées, s'offrent à nous. Le recours à une double analyse sur des spectromètres différents peut permettre par exemple d'augmenter le nombre de peptides et de protéines identifiés. D'autre part, il est également possible d'analyser un même échantillon une deuxième fois en spécifiant une liste d'exclusion de précurseurs à ne pas sélectionner pour la fragmentation ou bien de travailler sur des gradients plus longs et plus lents afin d'étaler les peptides.

III-A.4.1. Travailler sur des spectromètres de masse différents

Protocoles de référence :

Préparation des échantillons : **V-A.2** p.149

Expérience LC-MS/MS : LC2D *on-line* **V-A.4.1.** p.150, LC2D *off-line-ESI* **V-A.4.2.** p.152, FTMS **V-A.5.2.** p.153, LC2D *off-line-MALDI* **V-A.3.2. b** p.149, MALDI-TOF/TOF **V-A.5.4.** p.154

Recherche dans les banques et validation : **V-A.6.1. b** p.155, **V-A.6.2. c** p.156

Les sources d'ionisation ESI et MALDI sont utilisées en routine pour l'analyse de peptides. Si l'ESI peut être facilement interfacée avec une séparation en phase liquide, le MALDI a longtemps été utilisé

pour l'analyse directe de mélanges simples sans être couplé à l'HPLC. Avec l'apparition de robots de dépôt compatibles avec les séparations à nanodébit, de plus en plus de laboratoires ont mis en place des couplages LC-MALDI. Lorsque ces deux modes d'ionisation sont utilisés en parallèle sur les mêmes échantillons, ces techniques se montrent complémentaires [76, 208] du fait des différences en termes d'ionisation et de fragmentation des peptides et en termes d'acquisition des données. Cette complémentarité permet d'augmenter la couverture de protéome global tant en termes de protéines que de peptides identifiés.

Après validation de nos protocoles LC1D-MALDI, nous avons appliqué cette stratégie à une séparation LC2D. En effet, comme nous l'avons vu précédemment, l'approche LC2D *on-line* donne des résultats plus intéressants que la LC2D *off-line*, sans que la couverture de protéome soit totale. Il est possible d'utiliser les deux configurations LC2D de manière complémentaire, mais cela demande un changement de montage chromatographique. De façon plus simple, il est possible de réaliser une séparation LC2D-MALDI en complément d'une séparation LC2D-ESI. En configuration *on-line*, cela n'est possible qu'en ajoutant un *split* en sortie de colonne RP [76]. Dans ce cas, tous les appareils nécessaires (LC, ESI-MS, robot de dépôt) doivent être opérationnels en même temps et être placés dans le même espace de travail. Le couplage en configuration *off-line* laisse un peu plus de souplesse. En effet, une fois les fractions SCX collectées, il suffit de séparer leurs volumes en deux parties (égales ou non). Chaque fraction est ensuite séparée en RPLC sans qu'il soit nécessaire que le dépôt sur plaque MALDI soit réalisé en même temps que l'analyse ESI-MS, ni même directement à la suite de la séparation SCX. Nous avons donc opté pour le couplage LC2D*off-line*-MALDI pour mettre en évidence cette complémentarité sur nos échantillons.

En termes d'identifications protéiques et peptidiques (Figure 33), l'utilisation d'une configuration LC2D*off-line*-MALDI se révèle complémentaire de l'approche LC2D*offline*-ESI : quasiment autant de protéines ou de peptides sont identifiés par les deux méthodes. On peut noter que les protéines identifiées par l'approche MALDI seule ne sont pas les mêmes que les protéines vues uniquement pas l'approche *on-line*. D'autre part, alors que la moitié des 721 peptides vus uniquement par l'approche *off-line*ESI (en comparaison avec l'approche *off-line*MALDI) sont identifiés en mode *on-line,* le contraire n'est pas vrai : uniquement 44 des peptides vus en *off-line*MALDI sont identifiés en *on-line*ESI également. Ceci nous indique que les paramètres physico-chimiques guidant l'ionisation ESI et MALDI sont différents de ceux expliquant les pertes en *on-line* ou en *off-line*.

D'autre part, un des avantages de l'utilisation d'une source MALDI réside dans le fait que le processus d'ionisation est statique contrairement au processus d'ionisation ESI qui, lui, est dynamique.

Le temps d'acquisition en spectrométrie de masse et le temps d'élution chromatographique sont déconnectés. L'appareil sur lequel nous avons réalisé nos analyses MALDI (MALDI-TOF/TOF 4800, Applied Biosystems) utilise déjà cette différence dans le cas d'analyse en LC1D : ainsi, le logiciel est capable de repérer l'apex d'un pic chromatographique et de fragmenter le peptide dans le spot correspondant. Cependant, cette possibilité n'est pas mise à profit en LC2D. En effet, si un peptide est élué dans plusieurs fractions SCX, il sera cassé dans chaque fraction tout comme en ESI. Il pourrait être plus efficace de repérer dans quelle fraction le peptide est le plus intense et de le fragmenter uniquement dans le spot correspondant. Nous avons tenté de l'implémenter manuellement mais cette opération s'avère longue et fastidieuse : nous avons donc émis cette suggestion auprès des développeurs du logiciel afin que ce retraitement soit automatisé.

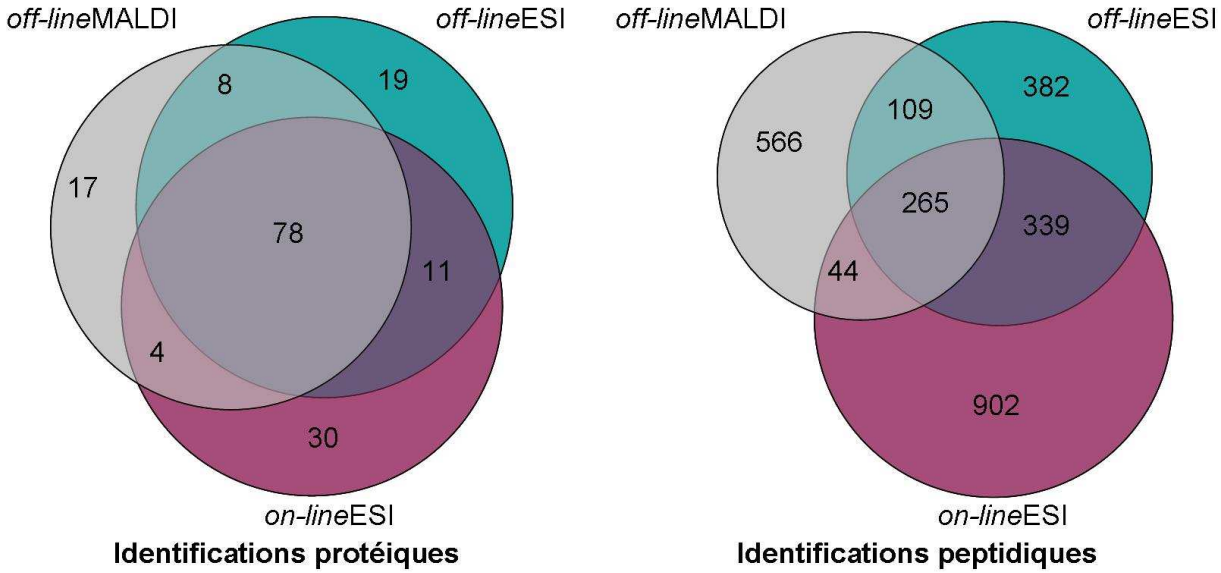


Figure 33 : Comparaison des identifications en termes de protéines et de peptides suivant les deux configurations LC2D-ESI et mise en évidence de l'intérêt l'utilisation d'une source d'ions complémentaire MALDI. Quasiment autant de protéines et de peptides sont identifiés par le mode *off-line*MALDI et par le mode *off-line*ESI. Cependant seuls 2/3 des protéines et 1/5 des peptides sont identifiés en commun par les deux méthodes. D'autre part, même si le mode *on-line*ESI est le plus performant, très peu de peptides identifiés en *off-line*MALDI le sont également en *on-line*ESI.

III-A.4.2. Utilisation de listes d'exclusions

Comme nous l'avons précédemment introduit, il est possible d'augmenter la couverture de l'échantillon en analysant une seconde fois l'échantillon en excluant les précurseurs précédemment sélectionnés. Ainsi les peptides moins intenses coélués avec des peptides déjà fragmentés peuvent à leur tour être sélectionnés pour fragmentation. Pour mettre en évidence ceci, nous avons analysé 3 fois un échantillon dilué 10 fois de SCP. Après validation des précurseurs ayant permis l'identification en triplicat

de protéines dans cet échantillon, 443 masses ont été exclues pour les analyses suivantes. Seuls 58 nouveaux précurseurs ont été fragmentés et ont permis de nouvelles identifications peptidiques et 2 protéines ont été validées en plus. L'utilisation de liste d'exclusion n'est donc pas suffisante pour rivaliser avec une approche LC2D : même en augmentant le nombre d'analyses successives et en agrandissant la liste d'exclusion à chaque *run*, il n'est pas envisageable que cela puisse augmenter le nombre de protéines identifiées par un facteur 2 ou 3 comme cela est le cas grâce à la LC2D. En effet, l'utilisation de listes d'exclusion n'a évidemment aucune incidence dans le cas d'effet de suppression de signal et certains peptides peu ionisés donneront des spectres MS/MS peu informatifs. De plus, certains peptides, quasi isobares d'espèces déjà fragmentées, risquent d'être rejetés par erreur. En effet, lorsque l'on travaille en balayage complet en FTMS, il faut ouvrir la fenêtre d'exclusion jusqu'à +/- 30ppm.

III-A.4.3. Travailler sur des gradients de plus faible pente

Protocoles de référence :

Préparation des échantillons : **V-A.2** p.149

Expérience LC-MS/MS : LC2D *on-line* **V-A.4.1.** p.150, LC2D *off-line* **V-A.4.2.** p.152, LC1D **V-A.3.2. a** p.149, FTMS **V-A.5.2.** p.153

Recherche dans les banques et validation : **V-A.6.1. b** p.155, **V-A.6.2. c** p.156

En travaillant sur des gradients de plus faible pente et sur des temps de séparation plus longs en configuration monodimensionnelle, il est possible d'augmenter la couverture du protéome tout en gardant un montage plus simple et plus robuste que les montages multidimensionnels. Pour pouvoir comparer les performances de la LC2D et de la LC1D à durée d'analyse égale, nous avons réalisé des gradients de 12h. Le temps d'analyse est alors augmenté d'un facteur 20 par rapport à une analyse LC1D classique. Ce type de gradient permet de répartir les peptides dans le temps : ainsi, si deux peptides sont élués dans le même pic chromatographique mais avec des intensités différentes par effet de suppression spectrale, ils ont une chance d'être élués séparément grâce à un gradient plus long. Cependant, il faut noter que l'augmentation de la durée du gradient sans augmentation de la longueur de la colonne de séparation implique un élargissement des pics chromatographiques : ainsi en moyenne la largeur des pics à leur base passe de 30s à 3min, voire 5min pour les peptides les plus abondants. Il est donc nécessaire d'augmenter la durée d'exclusion dynamique à 5min (1min en LC1D classique) pour éviter les fragmentations redondantes d'un même précurseur. D'autre part, et de manière plus préjudiciable pour les analyses, cet élargissement des pics s'accompagne *de facto* d'une diminution du rapport S/B des peptides élués (Figure 34) par effet de dilution. Pour les peptides les plus abondants, ce phénomène n'est pas limitant et ils restent identifiables ; en revanche, d'autres peptides minoritaires peuvent voir leur intensité chuter au-dessous du seuil de détection pour la fragmentation MS/MS ou être trop faible pour

donner un spectre MS/MS correctement interprétable. Dans tous les cas, ces peptides ne seront plus identifiés dans cette configuration alors qu'ils pouvaient l'être en LC1D classique (gradient de 35min) ou en LC2D. Ainsi, nous avons mis en évidence expérimentalement que certains peptides très bien identifiés en LC1D classique ne l'étaient plus en LC1D longue.

En termes d'identification protéique, l'utilisation d'une configuration gradient long 12h permet d'obtenir des résultats très proches de la LC2D *on-line* et même meilleurs qu'en LC2D *off-line* (Figure 35). En ce qui concerne les identifications peptidiques, 39% des peptides identifiés en LC1D long gradient ne sont identifiés ni en LC2D *off-line* ni en LC2D *on-line* : ce résultat met en évidence qu'il semble illusoire de tenter d'obtenir une image du protéome complète avec une seule méthode analytique. Ces peptides identifiés grâce à la LC1D longue seule correspondent notamment aux peptides longs et hydrophobes perdus en *off-line* par adsorption non spécifique sur les puits de collecte et également aux peptides courts et hydrophiles perdus en *on-line* du fait d'une élution isocratique des précolonnes de dessalage.

La configuration LC1D gradient long est cependant un excellent compromis : en effet, le montage utilisé reste le même qu'en LC1D classique avec des résultats d'identifications nettement meilleurs. Il faut noter que même si le nombre de peptides identifiés est multiplié par 3 par rapport à la configuration 1D classique, la capacité de pic, elle, n'est augmentée que d'un facteur 2, ce qui est dû à l'élargissement des pics chromatographiques. Il serait encore possible d'améliorer cette configuration en utilisant des colonnes plus longues afin de réduire cet élargissement : cependant, elles ne sont pas disponibles en version commerciale avec le même type de phase (greffon et granulométrie). Il serait possible de coupler plusieurs colonnes entre elles mais avec un risque de volume mort au niveau des jonctions. Des gradients de durée intermédiaire peuvent également être envisagés.

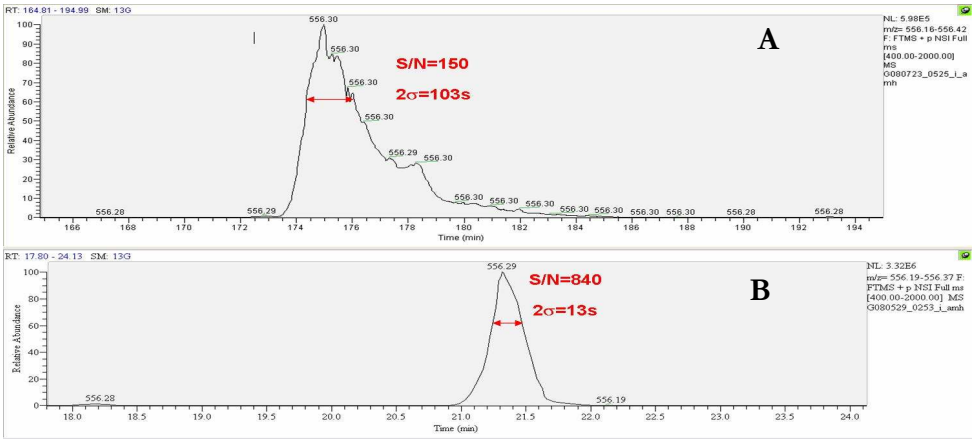


Figure 34 : Comparaison du profil d'élution chromatographique d'un même peptide en LC1D longue (A) et en LC1D classique (B).

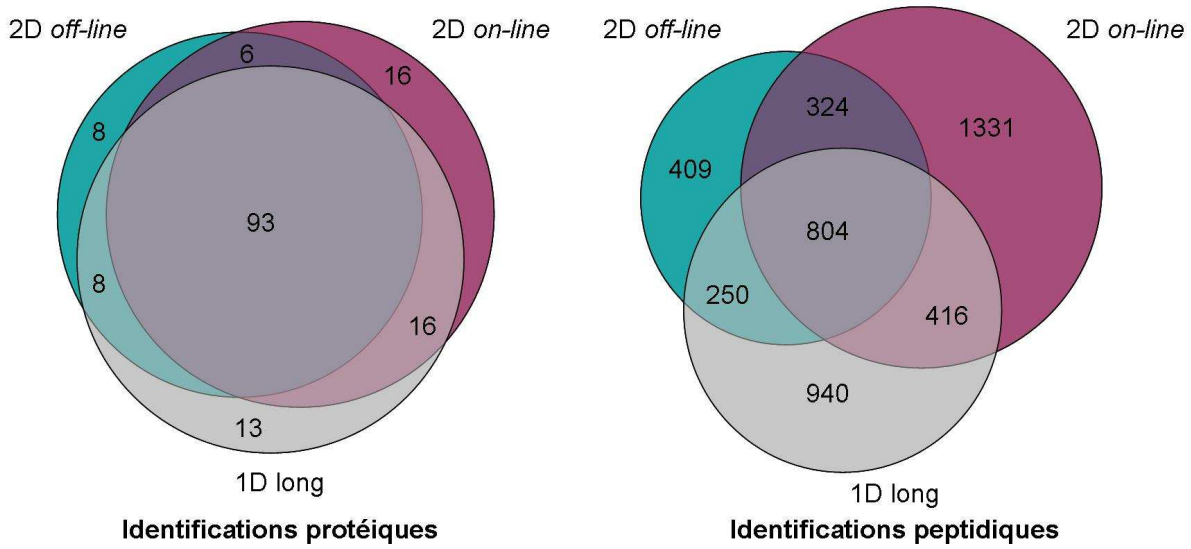


Figure 35 : Mise en évidence de l'intérêt d'utiliser des gradients longs. 88% des protéines identifiées grâce aux protocoles de LC2D (soit environ 6 x 12h) se retrouvent dans les identifications du protocole LC1D long (soit 3 x 12h). Cependant 39% des peptides identifiés par ce protocole ne le sont pas en LC2D.

III-A.5. Mise au point des analyses par spectrométrie de masse

Avant toute analyse d'échantillons réels, les conditions d'analyse sur les différents spectromètres de masse de type FTMS utilisés au cours de cette thèse ont été optimisés. Cette étude est présentée ici. Lors d'un couplage LC-MS/MS, il est possible de jouer sur différents paramètres pour améliorer l'analyse :

- le seuil de déclenchement de la sélection pour la fragmentation
- la durée d'exclusion dynamique durant laquelle un précurseur déjà fragmenté ne peut plus être sélectionné
- le nombre de MS/MS effectuées par cycle (correspond au nombre de précurseurs sélectionnés lors d'un *scan* MS)

III-A.5.1. Configuration LIT-FTICR

a Cycle d'acquisition

L'appareil de type **FTICR** utilisé au cours de ces travaux est un instrument hybride associant en série une trappe d'ions linéaire et une cellule FTICR (voir p.25). Afin d'exploiter au mieux les performances de cet analyseur il est préférable d'éviter une charge spatiale trop importante qui induirait une perte de justesse, tout en remplissant de manière optimale la cellule avec les ions d'intérêt. Pour optimiser les paramètres et la gamme dynamique de notre analyse, nous avons travaillé en alternant trois modes d'analyse, un mode en balayage complet en FTMS, un mode en SIM FTMS (*Single Ion Monitoring*) sur les

trois précurseurs les plus intenses, suivi d'une détection des fragments associés en mode LIT MS/MS [209].

b Gain en sensibilité et en gamme dynamique

Les mêmes échantillons de peptides endogènes de *stratum corneum* plantaire ont été analysés à la fois sur la configuration nanoLC-QqTOF (QTof2, Waters) et sur la configuration nanoLC-LIT-FTICR (LTQ-FT, ThermoFisher Scientific) afin de comparer les deux types de configuration et d'évaluer le gain en performance. Du point de vue de l'**exactitude**, nous sommes passés en routine d'une précision de 40ppm en moyenne à une précision légèrement inférieure au ppm. Cette précision nous permet d'obtenir des identifications plus sûres. De plus, la **gamme dynamique** a également été augmentée d'un ordre de grandeur environ. La montée en **résolution** permet une détermination non ambiguë de l'état de charge, ce qui est très important pour l'étude de peptides endogènes sur une gamme de masse de 1 000-10 000uma. Enfin, la **vitesse de balayage** très rapide du LTQ en MS/MS semble avoir un impact très important. En effet, un cycle complet sur le QTof2 (1MS + 3MS/MS) dure environ 39s alors que sur le LTQ-FT, le même cycle ne dure que 5s. Statistiquement, il pourrait donc y avoir 8 fois plus d'identifications grâce à la seconde configuration. Dans la pratique, pour un même échantillon de SCP, seulement 4 fois plus de spectres ont été interprétés avec succès.

La Figure 36 compare un spectre MS pris au hasard sur une analyse QqTOF avec la combinaison des MS sur une analyse LTQ-FT pendant la même durée soit environ 39s. Sur le LTQ-FT, 11 peptides ont été fragmentés et identifiés. En revanche, sur le QqTOF, 8 peptides sont détectés et seuls 3 ont été fragmentés et identifiés, ce qui correspond bien au rapport des identifications globales LTQ-FT/QqTOF. D'autre part il faut noter que suite à cette plus faible vitesse d'acquisition, au cycle suivant les peptides détectés sur le QqTOF mais qui n'ont pas été fragmentés ne peuvent plus être sélectionnés pour une fragmentation, car leur élution est terminée et ils ne sont plus détectables.

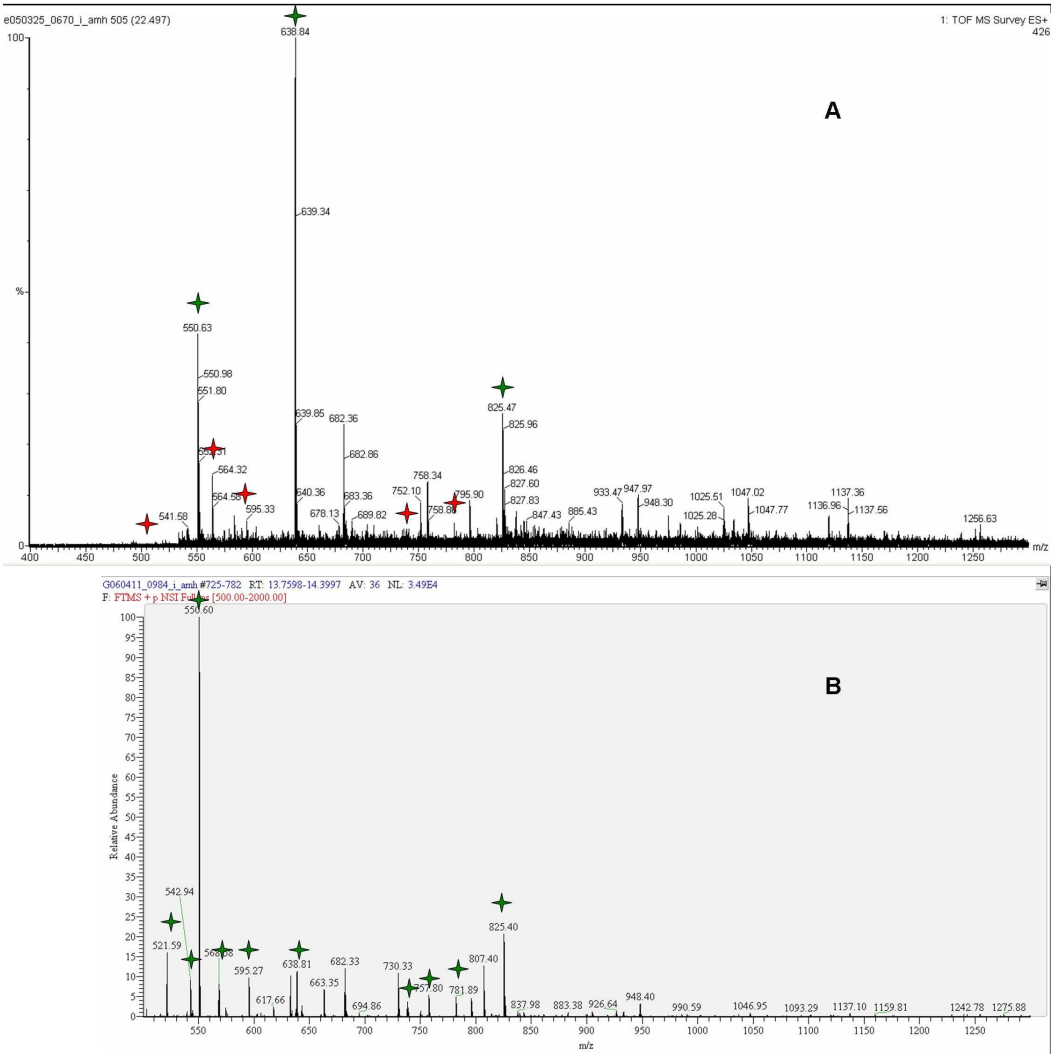


Figure 36 : Peptides fragmentés sur le QqTOF (A) pendant la même fenêtre temporelle que sur le LTQ-FT (B).
En vert correspondent les peptides détectés et fragmentés et en rouge les peptides détectés mais non sélectionnés.

c Comparaison des spectres MS/MS

De façon assez surprenante, malgré une précision de mesure bien moindre des données, **les résultats obtenus sur Mascot à partir des données QTof2 présentent un taux de faux positifs du même ordre de grandeur par rapport aux résultats obtenus à partir des données LTQ-FT** (environ 1,5%). Ceci peut s'expliquer par la très bonne qualité des spectres MS/MS acquis sur le QqTOF, et notamment grâce à l'observation des fragments de type y1/b1, qui ne sont pas détectés dans la trappe d'ions du fait de l'exclusion des fragments de basse masse. La Figure 37 présente deux spectres obtenus à partir du même précurseur dans les deux configurations différentes. L'interprétation par Mascot a permis d'obtenir un meilleur score d'identification grâce au QqTOF (Ions score 72) que sur le LTQ (Ions score 49) notamment grâce à l'observation d'une série d'ions y quasi complète (mise à part y6). D'autre part, la présence de plusieurs prolines dans la séquence semble moins gêner la fragmentation

dans le QqTOF que sur le LTQ. En effet, lorsque le peptide est fragmenté sur le LTQ, la rupture de liaison a lieu systématiquement non après la proline mais après l'acide aminé suivant (que ce soit dans la série b ou y).

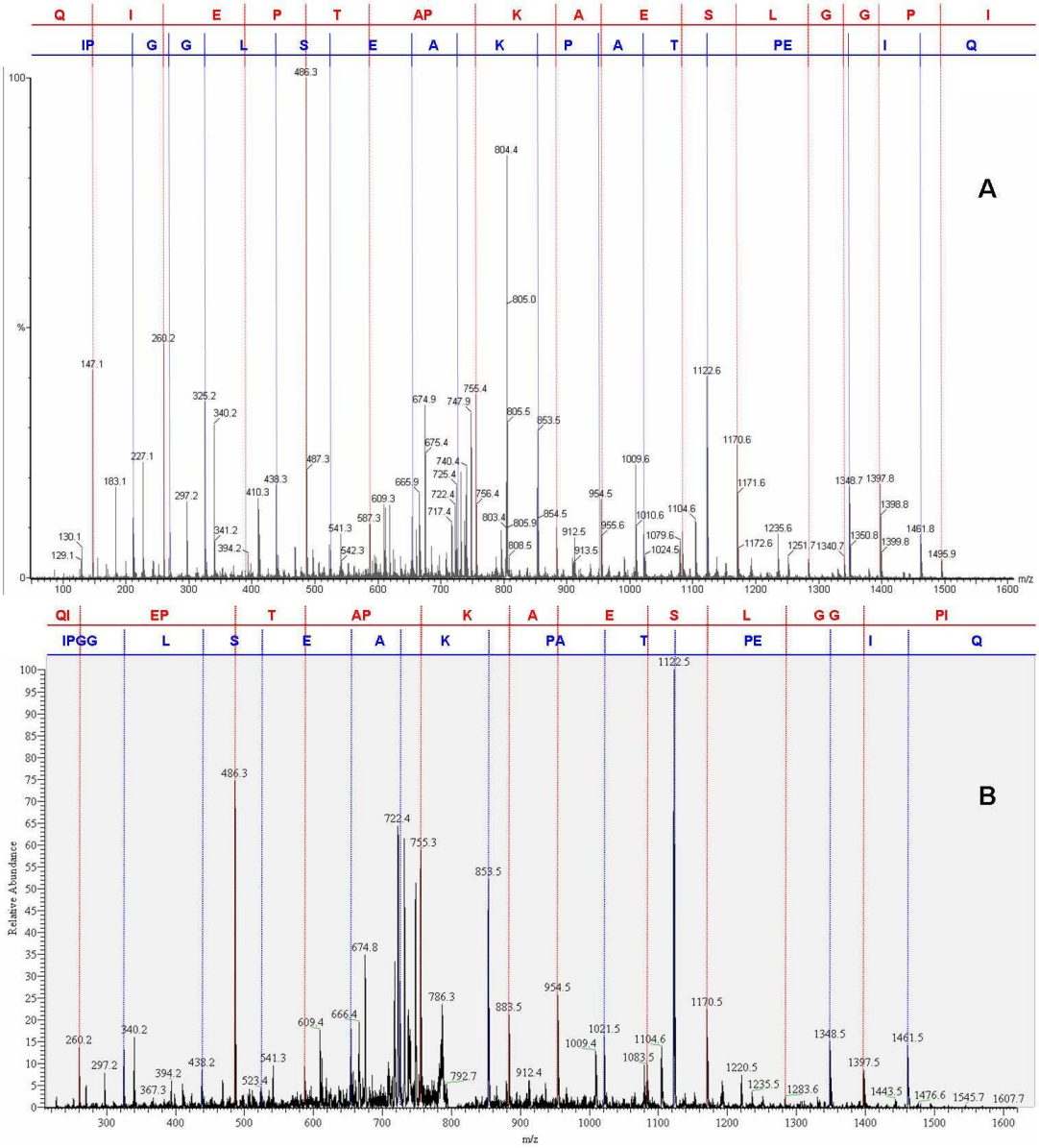


Figure 37 : Comparaison des spectres obtenus sur le même précurseur (804,4 Th) avec le QqTOF (A) et le LTQ (B).

III-A.5.2. Configuration LIT-Orbitrap

Protocoles de référence :

Préparation des échantillons digérés standard : **V-A.2** p.147

Préparation des échantillons non digérés : **V-A.1.2.** p.145, **V-A.1.3.** p.145, **V-A.1.4.** p.146

Expérience LC-MS/MS : **V-A.3.1. b** p.148, **V-A.5.3.** p.154

Recherche dans les banques et validation pour les échantillons digérés : **V-A.6.1. b** p.155, **V-A.6.2. c** p.156

Recherche dans les banques et validation pour les échantillons non digérés : **V-A.6.1. a** p.154, **V-A.6.2. b** p.155

Sur l'Orbitrap, l'optimisation des paramètres a d'abord été effectuée sur un échantillon standard (LCP) à différentes concentrations dans différentes conditions, puis sur un échantillon préparé dans les mêmes conditions que les échantillons réels que nous avions à analyser : un échantillon de *stratum corneum* plantaire repris dans 100µL après l'étape de reconcentration sur phase solide afin de pouvoir réaliser des analyses en triplicat dans différentes conditions.

a Influence du seuil de déclenchement de la sélection des précurseurs

Idéalement, le meilleur moment pour fragmenter un peptide correspond à l'apex de son pic d'élution chromatographique (Figure 38), c'est-à-dire là où il est le plus intense. En effet, dans ce cas, on peut supposer que le spectre de fragmentation aura le meilleur rapport signal sur bruit (S/B), ce qui permettra d'identifier des peptides moins abondants et moins intenses. Pour déterminer l'apex d'un pic, il suffit de calculer la dérivée de son intensité en fonction du temps : l'apex correspond au moment où cette dérivée s'annule. Cependant peu de logiciels travaillent de cette manière pour déterminer la sélection d'un peptide à la fragmentation : on peut notamment citer le logiciel Analyst (Applied Biosystems). La plupart des logiciels dont Xcalibur (ThermoFisher Scientific) se basent sur l'intensité des peptides. Il est donc nécessaire de déterminer un seuil de sélection pour la fragmentation qui ne soit ni trop faible (sélection trop précoce de l'espèce à fragmenter) ni trop élevé (fragmentation tardive de l'espèce à fragmenter induisant une chute de sensibilité).

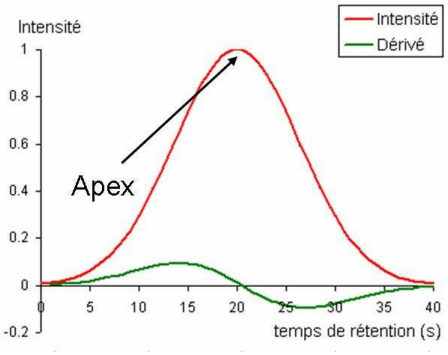


Figure 38 : Profil d'élution d'un pic chromatographique

Les méthodes par défaut avec lesquelles nous avons commencé à travailler sur l'Orbitrap étaient des méthodes où le seuil de détection absolue était baissé à 100cps. Or, le bruit électronique généré dans un *scan* FTMS dans une zone où il n'y a pas d'ions polydimethylcyclosiloxanes (Figure 39), est de l'ordre de 500cps au minimum.

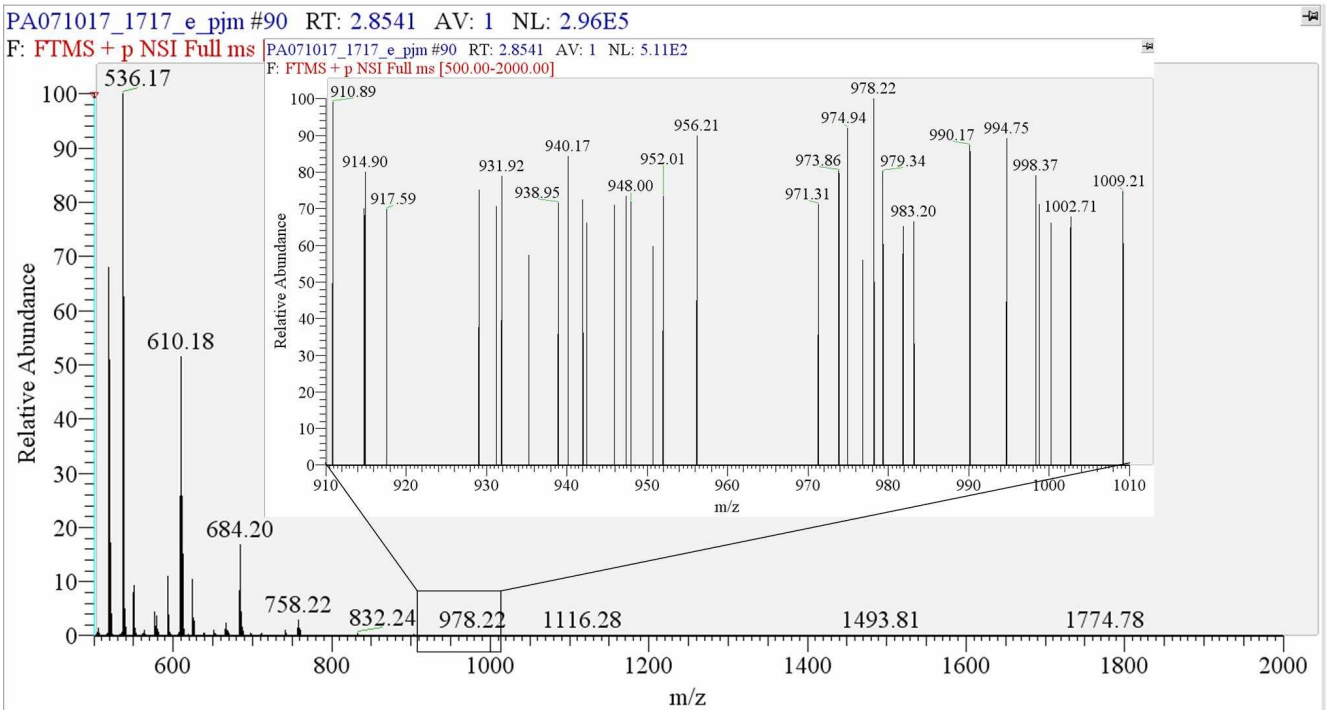


Figure 39 : Bruit électronique classiquement obtenu sur un *scan* MS avec détection dans l'Orbitrap.

Il semble donc que 100cps soit une valeur de seuil trop faible. En effet la limite de détection d'une espèce par une technique donnée peut être estimée comme correspondant à la concentration pour laquelle cette espèce est détectée à un rapport S/B ≥ 3 [210]. Cela correspondrait dans notre cas à une intensité de 1 500cps. Un seuil de 1 000cps devrait donc être suffisant pour déclencher la fragmentation. Ceci est confirmé lorsque l'élution d'un peptide peu abondant est suivie spécifiquement (Figure 40) : ce peptide n'est détectable que si son intensité est supérieure à 1 000cps et le massif isotopique de l'ion bichargé n'est vraiment reconnu qu'à partir de 4 000cps.

L'utilisation d'un seuil très faible permet *a priori* d'augmenter la sensibilité de l'analyse puisque théoriquement tous les peptides détectables ont la possibilité d'être sélectionnés pour la fragmentation. Cependant si un seuil trop faible est choisi, les ions les moins intenses sélectionnés ne donneront pas de spectres de fragmentation correctement interprétables.

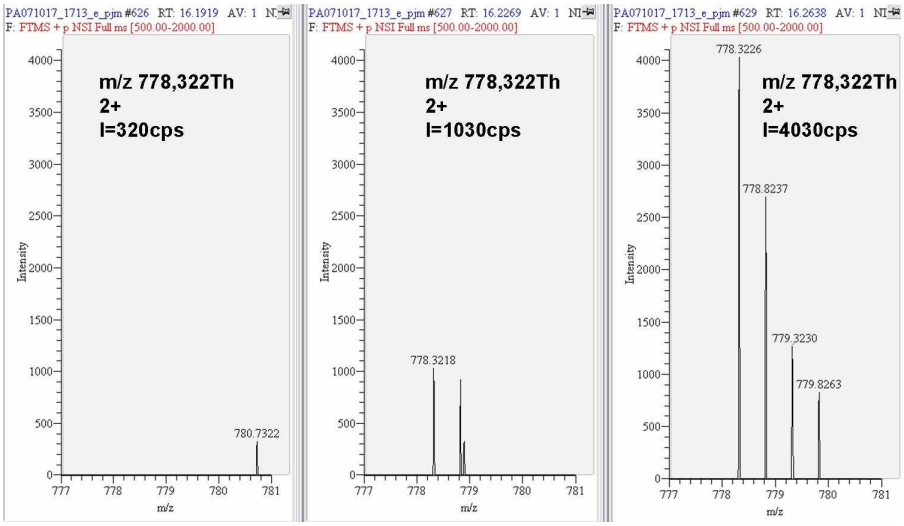


Figure 40 : Suivi dans 3 scans FTMS successifs de l'élution d'un peptide peu abondant de rapport m/z 778,322Th dans 20fmol de LCP.
Le massif isotopique ne devient clairement visible qu'à partir d'une intensité de 4 000cps.

Différents paramètres peuvent être utilisés pour évaluer l'influence du seuil de détection sur les résultats obtenus, notamment le score moyen des peptides identifiés, le nombre de peptides uniques identifiés et le rapport entre le nombre total de peptides identifiés et le nombre d'ions fragmentés.

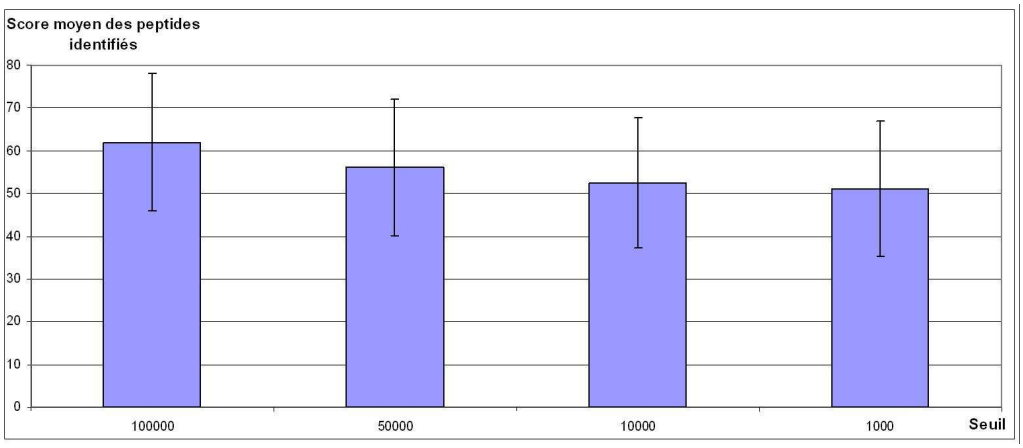


Figure 41 : Influence du seuil de sélection sur le score moyen des peptides identifiés pour un mélange de 20fmol de LCP

Comme prévu, le score moyen des peptides identifiés augmente si le seuil de détection augmente puisque qu'ils sont fragmentés lorsqu'ils sont plus intenses et donnent donc des spectres MS/MS de meilleur rapport S/B (Figure 41). Cependant, plus le seuil de détection augmente, plus le nombre de peptides uniques diminue (Figure 42), même si, sur ce point, les résultats obtenus avec un seuil de 1 000cps ne sont pas significativement différents de ceux obtenus avec un seuil de 10 000cps. Dans le cadre de l'analyse de 10fmol de mélange standard stoechiométrique, toutes les protéines sont identifiées mais dans le cadre de l'analyse d'échantillons complexes, il est probable que cela nuirait à l'identification des peptides les moins abondants et donc à l'identification des protéines associées.

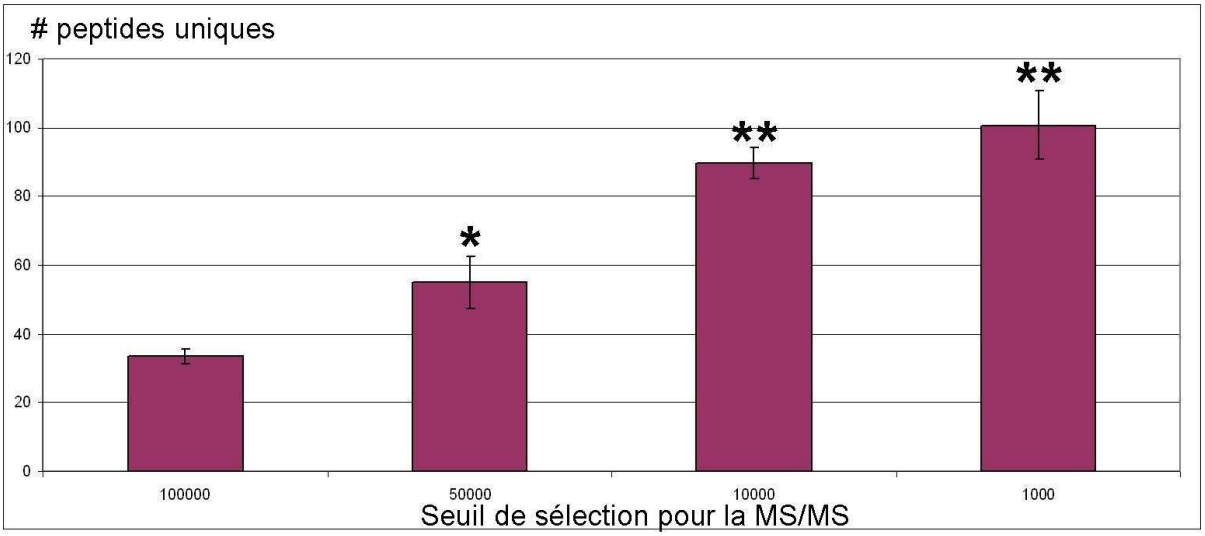


Figure 42 : Influence du seuil de sélection sur le nombre de peptides uniques identifiés sur un échantillon standard de 10fmol de LCP.
***le nombre de peptides uniques identifiés avec un seuil de 50 000cps est significativement supérieur au nombre de peptides uniques identifiés avec un seuil de 100 000cps. **le nombre de peptides uniques identifiés avec un seuil de 1 000cps ou 10 000cps est significativement supérieur au nombre de peptides uniques identifiés avec un seuil de 50 000cps ou de 100 000cps.**

Afin de vérifier cette hypothèse et afin d'effectuer la mise au point sur un échantillon plus proche de ceux à analyser sur l'Orbitrap, nous avons utilisé un extrait peptidique de *stratum corneum* plantaire. Dans ce cas, le nombre de peptides uniques identifiés au total ne varie pas significativement avec le seuil de détection utilisé car une forte majorité des peptides est issue de kératines très abondantes (Figure 43). Cependant, dans le cas spécifique des autres protéines, le nombre de peptides identifiés avec un seuil de 10 000cps est sensiblement le même que pour un seuil de 1 000cps mais significativement supérieur qu'avec un seuil de 50 000cps. En particulier pour les protéines du cornéodesmosome ou de la famille des plakines qui sont impliqués dans le phénomène de cohésion du *stratum corneum*, les protéines sont qualitativement mieux identifiées avec un seuil de détection de 10 000cps (mis à part la desmogléine) (Figure 44). La desmoplakine n'est par exemple identifiée qu'avec la méthode utilisant un seuil de détection de 10 000cps. La couverture de séquence de la cornéodesmosine est notamment significativement meilleure avec un seuil de 10 000cps qu'avec un seuil de 1 000cps. Quant à la desmocolline 1 et la desmogléine 1, leur couverture de séquence est sensiblement la même avec un seuil de 1 000cps ou de 10 000cps mais diminue avec un seuil de 50 000cps. Globalement, il semble donc que le seuil de 50 000cps favorise les faux négatifs bien que les scores peptidiques individuels soient meilleurs. Il y a moins de peptides identifiés, ce qui nuit aux couvertures de séquences des différentes protéines. Même si les résultats sont assez proches en termes de peptides identifiés et de couvertures de séquence avec un seuil de 1 000cps et de 10 000cps, nous avons tout de même observé une tendance favorisant le seuil de 10 000cps (en particulier pour l'identification de la cornéodesmosine), nous avons donc par la suite retenu ce seuil de détection.

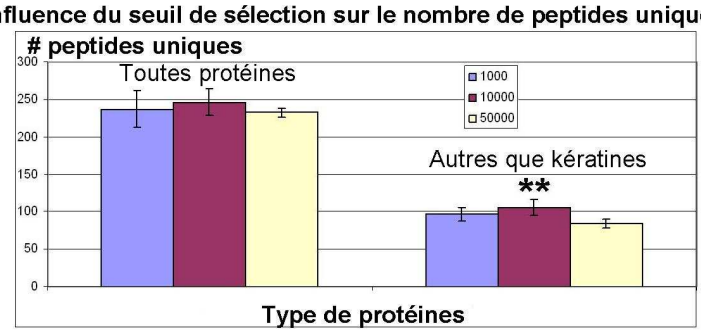


Figure 43 : Influence du seuil de détection sur le nombre de peptides uniques identifiés dans un échantillon complexe de *stratum corneum* plantaire, en particulier en ce qui concerne les protéines autres que les kératines.
****le nombre de peptides uniques identifiés avec un seuil de 10 000cps est significativement supérieur au nombre de peptides uniques identifiés avec un seuil de 50 000cps en ce qui concerne les protéines autres que les kératines.**

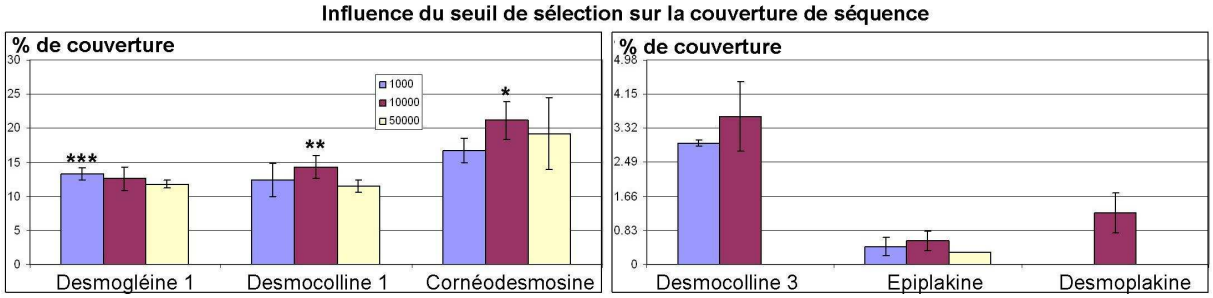


Figure 44 : Influence du seuil de détection sur la couverture de séquence des protéines du cornéodesmosome ou de la famille des plakines.
***seuil 10 000cps significativement différent du seuil 1 000cps. **seuil 10 000cps significativement différent du seuil 50 000cps. ***seuil 1 000cps significativement différent du seuil 50 000cps.**

b Influence de la durée d'exclusion dynamique

Ce paramètre correspond à la durée pendant laquelle un précurseur déjà fragmenté ne doit plus être sélectionné. Sa valeur résulte d'un compromis. En effet, si la durée d'exclusion est trop longue, un ion quasi isobare d'un précurseur de la liste d'exclusion risque de ne pas être sélectionné alors qu'il ne correspond pas au même peptide. Au contraire, si elle est trop courte, un même ion risque d'être fragmenté plusieurs fois sans information supplémentaire, notamment dans le cas d'ions très abondants. En théorie, il devrait suffire de déterminer cette durée par le profil d'élution des pics en fonction du seuil de sélection choisi, la durée d'exclusion devant correspondre à la durée durant laquelle l'ion est supérieur au seuil de sélection. Cependant en pratique, cette durée varie en fonction de l'abondance de l'ion. En effet, les pics chromatographiques ne sont pas parfaits et pour un ion peu intense, cette durée va généralement être assez courte, une vingtaine de secondes alors que pour un ion abondant et intense, qui va avoir tendance à traîner en fin d'élution, cette durée peut être beaucoup plus longue : jusqu'à 1min30 (Figure 45). Nous avons donc choisi d'exclure les ions durant 45s ce qui pour la plupart des peptides correspond à la largeur des pics à 5% de leur intensité maximale.

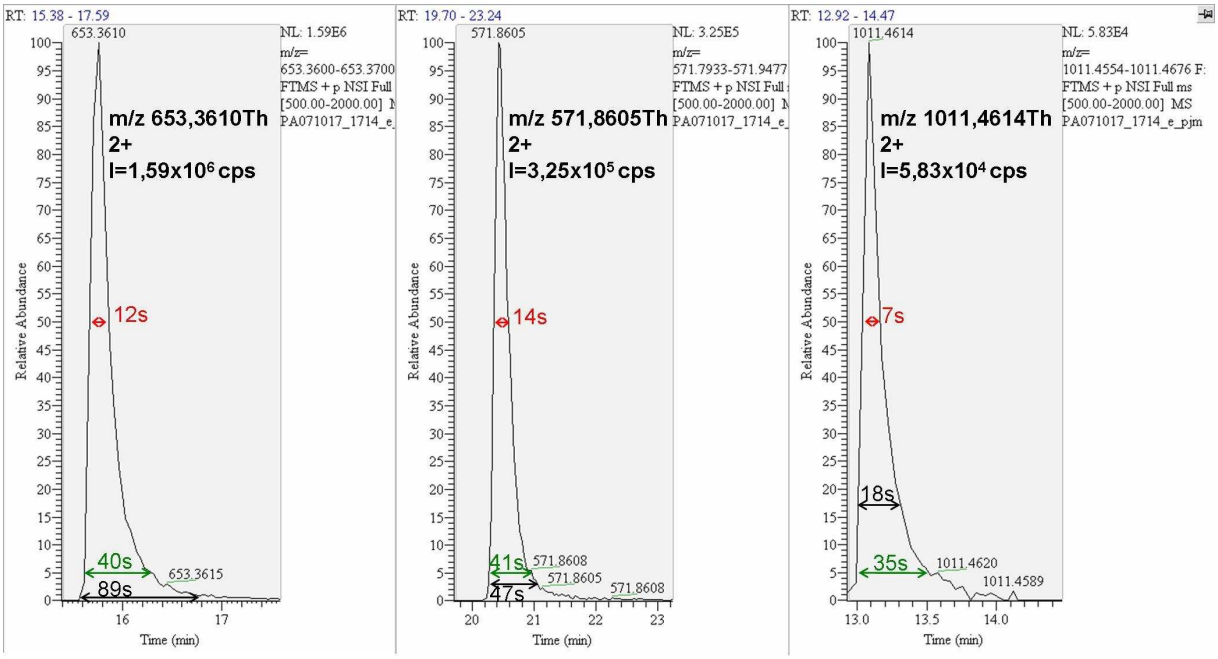


Figure 45 : Profils de différents peptides extraits d'une séparation de 20fmol de LCP.
Sont pointées différentes largeurs caractéristiques des pics : en rouge la largeur à mi-hauteur, en vert la largeur à 5% et en noir la largeur pour avoir une intensité supérieure à 10 000cps.

Dans la pratique, les résultats obtenus selon la durée d'exclusion sont relativement comparables : certaines protéines étant moins bien identifiées avec une durée de 45s, et d'autres mieux, comme la cornéodesmosine (Figure 46). Nous avons donc par la suite utilisé une durée d'exclusion dynamique de 45s.

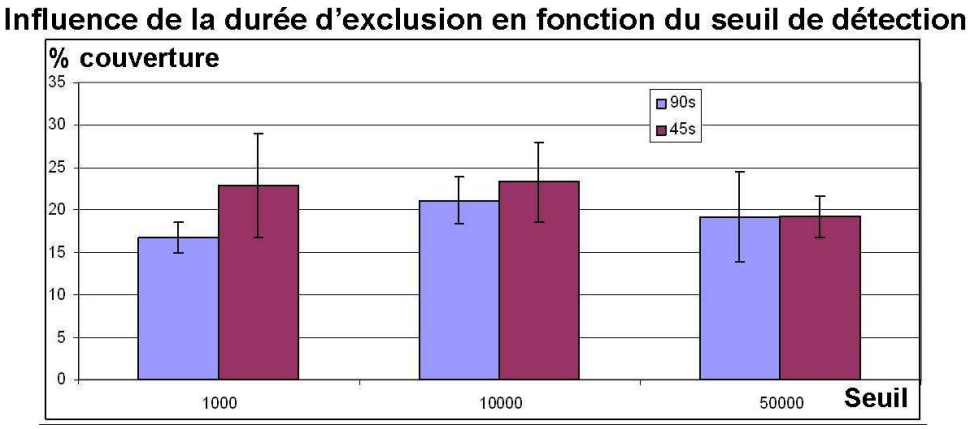


Figure 46 : Comparaison des couvertures de séquence de la cornéodesmosine en fonction du seuil de détection et de la durée d'exclusion dynamique.

c Influence du nombre de MS/MS

A chaque cycle d'analyse, est réalisée une analyse FTMS à la suite de laquelle n MS/MS sont réalisées dans le LTQ sur les n précurseurs les plus intenses. La co-sélection à la fragmentation d'un certain nombre de peptides est intéressante. Néanmoins, si trop d'espèces sont sélectionnées, alors même les espèces détectées très faiblement seront sélectionnées, et les précurseurs, de faible intensité en MS ne fourniront pas de spectres MS/MS exploitables. Dans ce cas, il serait alors plus intéressant de les ignorer

et de renouveler l'analyse MS pour rechercher des espèces de meilleur rapport S/B pour fournir un spectre MS/MS interprétable. Nous avons donc fait varier le nombre de MS/MS de 3 à 11. Même si, globalement, le nombre de peptides uniques identifiés selon la méthode ne varie pas beaucoup, la fragmentation de 5 précurseurs par cycle permet d'obtenir de meilleurs résultats que ce soit pour les échantillons standard commerciaux LCP ou pour les échantillons de SCP endogènes (Figure 47).

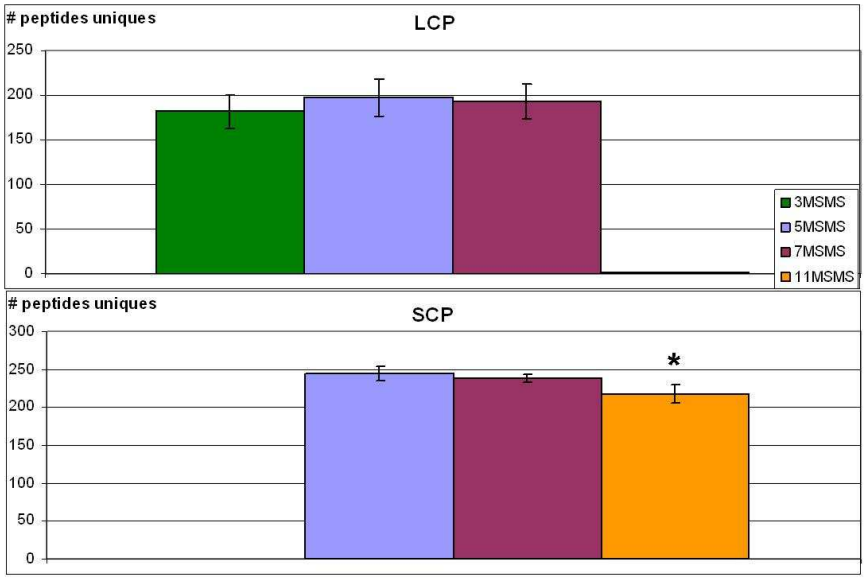


Figure 47 : Comparaison du nombre de peptides identifiés selon le nombre de MS/MS réalisées à chaque cycle.
***11 MS/MS significativement moins bon que 5 ou 7 MS/MS.**

La méthode optimisée que nous avons utilisée par la suite sur la plateforme a pour paramètres d'acquisition :

- Seuil de sélection à la fragmentation de 10 000cps,
- Durée d'exclusion dynamique de 45s,
- 5 MS/MS réalisées par cycle d'acquisition.

III-A.6. Recherche dans les banques de données : de l'intérêt des modifications

III-A.6.1. Protéines digérées

Protocoles de reference :

Préparation des échantillons digérés : **V-A.2** p.147

Expérience LC-MS/MS : **V-A.4.1.** p.150, **V-A.5.2.** p.153

Recherche dans les banques et validation : **V-A.6.1. b** p.155, **V-A.6.2. c** p.156

Pour les échantillons digérés de *stratum corneum*, les recherches par défaut ont été réalisées en spécifiant l'oxydation des méthionines et la carbamidométhylation des cystéines (traitement à l'iodoacétamide). Dans l'exemple d'un échantillon de SCP analysé en LC2D, 8349 spectres MS/MS ont été acquis mais seulement 22% ont été attribués à une protéine et validés, ce qui est la proportion classique de spectres identifiés dans les études protéomiques avec des paramètres de recherche standards : de nombreux spectres restent non interprétés. Si ces spectres contiennent suffisamment d'informations et de fragments avec un rapport S/B correct, il peut y avoir différentes raisons expliquant l'échec à l'interprétation lors de la recherche [211] :

- Mauvaise détermination de la charge du précurseur,
- Erreur de masse sous-estimée,
- Digestion non spécifique par l'enzyme,
- Modifications chimiques ou post-traductionnelles non suspectées,
- Séquence peptidique absente de la banque de données.

Pour pallier les trois dernières difficultés, il est possible de réaliser des recherches dans Mascot avec la fonction *error tolerant*. Lorsque cette fonction est activée, une première recherche standard est lancée, utilisant les paramètres spécifiés dans le formulaire. A partir de ces résultats, toutes les protéines identifiées avec au moins un peptide de score supérieur au seuil d'homologie sont sélectionnées pour une seconde recherche. Durant celle-ci, différents paramètres sont modifiés :

- L'enzyme spécifiée devient semi-spécifique et le nombre de *missed cleavages* est incrémenté de 1,
- La liste complète des modifications déclarées dans unimod est testée de manière séquentielle,
- L'ensemble des substitutions d'acides aminés, dues à la mutation ponctuelle d'une seule base, est testé.

Sachant toutefois que:

- Un peptide ne peut pas être identifié par le cumul de plusieurs de ces critères (peptide semi-spécifique **ou** modifié **ou** avec substitution d'acides aminés),

- Si les masses du peptide modifié et du peptide non modifié sont toutes deux dans la fenêtre de tolérance de masse, la modification est rejetée.

Dans notre cas, pour le même échantillon sur lequel nous avions réalisé une recherche de base, le nombre de spectres interprétés est augmenté de 50%. Parmi les spectres qui ont été interprétés grâce à cette fonction :

- 499 correspondent à des peptides semi-tryptiques,
- 699 correspondent à des peptides modifiés,
- 275 correspondent à une substitution d'acides aminés.

Les peptides semi-tryptiques proviennent de ruptures non spécifiques de liaisons de la chaîne protéique, ce qui devrait donc être défavorisé par rapport aux clivages tryptiques. Ils ne représentent en effet que 16% des spectres identifiés. Il est aussi possible que les peptides semi-tryptiques souffrent d'une moindre efficacité de fragmentation par rapport aux peptides tryptiques. Le fait qu'une très large majorité (81%) de ces peptides semi-tryptiques ait une structure analogue à un peptide tryptique, car ils résultent d'une digestion tryptique en C-terminal et non tryptique en N-terminal, va aussi dans ce sens. Seuls quelques peptides portent un autre résidu que K ou R en position C-terminale. Une autre hypothèse serait que ces peptides ne possédant pas de résidu basique C-terminal, se fragmentent moins régulièrement et sont donc moins informatifs.

Parmi ces peptides semi-tryptiques, 22% sont issus d'une protéolyse après un résidu aromatique (en particulier les résidus Y et F). Ce type de clivage semble indiquer une activité contaminante de type chymotryptique confirmée si les clivages après les résidus L, M, A, D et E sont également comptabilisés. Cette activité permet de justifier quasiment 60% des séquences semi-tryptiques.

De façon assez anecdotique, 2 peptides ont été identifiés, comme provenant d'une rupture de la liaison K\P ou R\P, généralement prohibée. La Figure 48 représente le spectre du peptide K\PFFHSLSEK de la thioredoxin où 7 fragments de type y consécutifs ont pu être repérés.

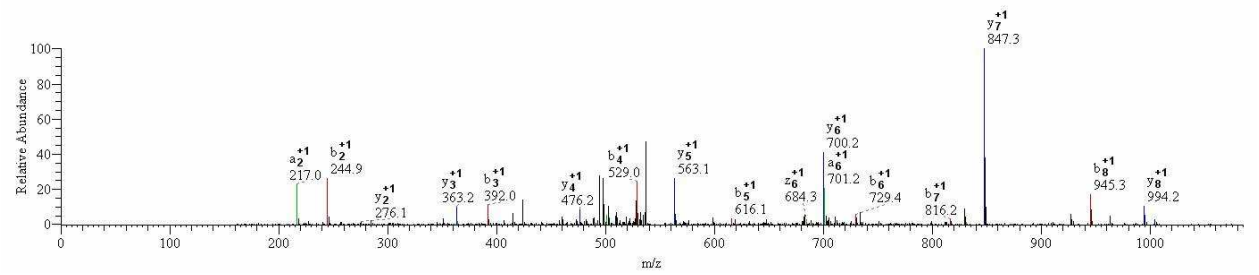


Figure 48 : Interprétation validée d'un spectre MS/MS lors de la confrontation dans la banque Swiss-Prot des données LC2D-MS/MS obtenues sur un échantillon de SCP, avec l'hypothèse que les peptides sont semi-tryptiques.
Mise en évidence de la séquence PFFHSLSEK, spécifique de la thioredoxin, qui provient d'un clivage K\P.

III-A.6.2. Peptides endogènes non digérés

Protocoles de reference :

Préparation des échantillons non digérés : **V-A.1.2.** p.145, **V-A.1.3.** p.145, **V-A.1.4.** p.146

Expérience LC-MS/MS : **V-A.3.1. b** p.148, **V-A.5.2.** p.153

Recherche dans les banques et validation : **V-A.6.1. a** p.154, **V-A.6.2. b** p.155

Après traduction, la plupart des protéines sont chimiquement modifiées par des MPTs, par exemple par l'addition d'un groupe sucre ou phosphate. De telles modifications jouent des rôles vitaux en modulant la fonction de nombreuses protéines mais elles ne sont pas directement codées dans les gènes. Les MPTs sont nombreuses (plus de 300 répertoriées) et peuvent être modulées de façon dynamique. Le changement de structure de l'acide aminé modifié est accompagné d'un changement de masse de cet acide aminé. La détection et la caractérisation des MPTs en MS exploitent cette variation de masse. Les MPTs peuvent être classées en trois catégories : clivage protéolytique d'une partie de la séquence (élimination de la méthionine d'initiation, de la séquence signal…), ajout d'un groupe chimique (acétylation, glycosylation, phosphorylation…) et formation d'un lien inter- ou intramoléculaires (pont dissulfure, thioether…). Le Tableau 7 présente les MPTs les plus courantes et les plus importantes [212].

Tableau 7 : Quelques MPT importantes et couramment observées.
Stabilité : + labile en spectrométrie de masse, ++ modérément stable, +++ stable. D'après Mann *et* Jensen [212].

Type de MPT	DeltaMasse (Da)	Stabilité	Fonctions
Phosphorylation			MPT réversible, activation/inactivation de l'activité d'enzymes, voie de signalisation
pTyr	+80	+++	
pSer, pThr	+80	+/++	
Acétylation	+42	+++	Stabilisation des protéines, protection N-ter ; régulation des interactions protéines/ADN
Methylation	+14	+++	Régulation de l'expression des gènes
Acylation			Localisation cellulaire, signaux cibles, attachement à la membrane, médiation des interactions protéine/protéine
Farnesyl	+204	+++	
Palmitoyl	+238	+/++	
Glycosylation			Protéines sécrétées, reconnaissance/signalisation cellule/cellule, fonction de régulation
N-linked	>800	+/++	
O-linked	203, >800	+/++	
Hydroxyproline	+16	+++	Stabilisation des protéines, interaction protéine/ligand
Sulfation (sTyr)	+80	+	Modulation des interactions protéine/protéine ou récepteur/ligand
Pont disulfures	-2	++	Crosslink inter/intra moléculaire, stabilisation des protéines
Déamidation	+1	+++	Régulation des interactions protéine/ligand et protéine/protéine ; artefact chimique commun
Pyroglutamisation	-17	+++	Stabilisation des protéines, blocage N-ter
Ubiquitination	>1000	+/++	Signal de destruction
Nitration des tyrosines	+45	+/++	Dommage oxydatif durant l'inflammation

Dans le cas de synthèse de neuropeptides (qui sont par définition des peptides endogènes), ils peuvent aussi être modifiés *via* des enzymes après l'action sélective de peptidases sur les protéines précurseurs [213]. Cependant bien que comptant parmi les modifications courantes des protéines, la phosphorylation des neuropeptides est rare. En revanche, d'autres MPTs qui protègent les peptides contre des dégradations protéolytiques sont plus fréquentes [214]. En position N-terminale, un grand nombre de neuropeptides porte un acide pyroglutamique, formé par la cyclisation d'une glutamine N-terminale (-17Da). Une autre protection possible des peptides est la formation d'un groupe amide en C-terminal à la place du groupe hydroxyl, ce qui constitue une modification portée par près de la moitié des peptides bioactifs [215]. D'autres modifications comme la sulfation des tyrosines, la methylation ou les glycosylation peuvent influer sur l'interaction des peptides avec les récepteurs et donc réguler leur activité [216].

Dans notre étude de peptides endogènes, les recherches dans les banques de données se font avec le paramètre *no enzyme*, il n'est donc pas envisageable, pour des raisons de temps, de faire une recherche en incluant de surcroît l'ensemble des modifications préalablement citées. En effet, lorsque les recherches

dans Mascot dépassent une semaine, elles s'arrêtent ; or une recherche avec 4 modifications différentes durent déjà plus de 2 jours (pour un fichier mgf de quelques centaines de Mo). Il faut également savoir que la durée de la recherche n'est proportionnelle ni à la taille du fichier ni au nombre de modifications, mais qu'elle augmente de manière assez exponentielle en bloquant le système pour les autres recherches menées en parallèle. Nous avons donc testé sur les échantillons de *stratum corneum* plantaire (données FTMS) quelles pouvaient être les modifications possibles dans nos échantillons en réalisant plusieurs recherches successives (Tableau 8).

Tableau 8 : Paramètres de recherche testés sur les échantillons non digérés de SCP.

Recherche 1	Aucune modification
Recherche 2	Phospho (ST) ; Phospho (Y)
Recherche 3	Sulfo (S) ; Sulfo (T) ; Sulfo (Y)
Recherche 4	Gln->pyro-Glu (N-term Q) ; Oxidation (M)
Recherche 5	Deamidated (NQ) ; Deamidated (R)
Recherche 6	Acetyl (K) ; Acetyl (N-term)
Recherche 7	Amidated (C-term) ; Amidated (Protein C-term)
Recherche 8	Methyl (C-term) ; Methyl (DE) ; Methyl (C) ; Methyl (H)
Recherche 9	Methyl (I) ; Methyl (K) ; Methyl (L) ; Methyl (N)
Recherche 10	Methyl (N-term) ; Methyl (Protein N-term) ; Methyl (Q) ; Methyl (R)
Recherche 11	Methyl (S) ; Methyl (T) ; Methyl+Deamidated (N) ; Methyl+Deamidated (Q)

Parmi les recherches réalisées, la plupart n'apportent que très peu d'information et ne permettent d'identifier que moins de 1% de spectres supplémentaires (Figure 49). Cependant les recherches 4 et 6 permettent d'augmenter significativement le nombre de spectres attribués à des séquences peptidiques avec succès. En particulier, la prise en compte de la pyroglutamisation des glutamines ainsi que l'acétylation des peptides en N-terminal permet d'identifier de nouveaux peptides (Tableau 9). Dans la suite de nos analyses, nous avons donc choisi de réaliser les recherches dans les banques de données en spécifiant ces modifications partielles : Gln->pyro-Glu (N-term Q), Acetyl (N-term) et Oxidation (M). En ce qui concerne les échantillons de *stratum corneum* plantaire, ces paramètres permettent d'augmenter de 23% le nombre de spectres correctement interprétés. Pour augmenter encore la couverture de peptidome, nous avons également utilisé la fonction *error tolerant* dans le cas des échantillons de peptides endogènes non digérés. Cependant cette option n'est utilisable que si une enzyme est déjà spécifiée. Comme nous l'avons vu dans les chapitres d'introduction, les protéines du cornéodesmosome sont dégradées par la SCCE (*stratum corneum chymotryptic enzyme*) et la SCTE (*stratum corneum tryptic enzyme)*. Les recherches en *error tolerant* ont donc été menées en spécifiant soit la trypsine soit la chymotrypsine

comme enzyme de digestion. Quelques peptides supplémentaires modifiés ou avec une substitution d'acide aminé ont pu ainsi être identifiés (Tableau 9).

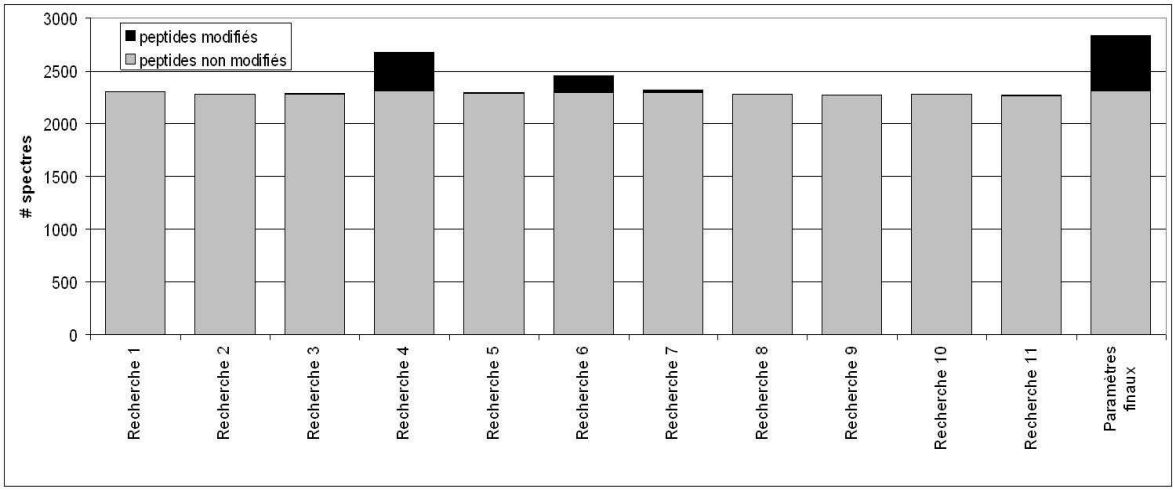


Figure 49 : Apport de la prise en compte des différentes modifications.

Tableau 9 : Distribution des identifications peptidiques parmi les différentes catégories en termes de spectres MS/MS et de peptides uniques.
Sont aussi comptés les nouveaux peptides [a]par rapport aux peptides non modifiés ou [b]par rapport à la recherche initiale (avec oxydation des méthionines) ainsi que les identifications protéiques.

	Catégories	Spectres	Peptides	Nouveaux peptides	Nouvelles protéines
SC plantaire	2 666 spectres attribués en recherche initiale				
	sans modifications	2568	628	628	53
	Met +16	98	35	18[a]	0
	480 spectres additionnels				
	N-terminal pyro-Glu de Gln	267	54	47[b]	0
	N-terminal acétyl	164	38	35[b]	0
	autres modifications	49	16	9[b]	0
SC non plantaire	1 863 spectres attribués en recherche initiale				
	sans modifications	1768	574	574	51
	Met +16	96	31	12[a]	0
	233 spectres additionnels				
	N-terminal pyro-Glu de Gln	120	30	28[b]	0
	N-terminal acétyl	91	23	23[b]	0
	autres modifications	22	18	12[b]	1[b]

a Cas de la pyroglutamisation

Cette modification peut se former artificiellement à partir d'une glutamine si les peptides sont conservés plus de 15min dans des conditions moyennement acides. Cependant, l'identification des

peptides contenant cette modification reste informative, même s'ils ne permettent pas de mettre en évidence de nouvelles protéines. En effet, dans nos études peptidomiques, tous les peptides identifiés sont importants. Dans nos analyses, un nombre significatif des peptides identifiés (7% pour les échantillons de SCP, 4% pour les échantillons de PA) contiennent une glutamine N-terminale convertie en acide pyroglutamique (267 spectres correspondant à 54 peptides pour les échantillons de SCP et 120 spectres correspondant à 30 peptides pour les échantillons de PA). Parmi ces peptides, quasiment tous (47 dans le cas du SCP, 28 dans les PA) n'avaient pas été vus dans leur version non modifiée dans nos données.

b Cas de l'acétylation N-terminale

Sur les 44 peptides uniques (255 spectres MS/MS, SCP et PA combinés) identifiés contenant une acétylation N-terminale, quasiment tous (42) n'ont été vus que sous cette forme modifiée, 4 sont localisés en N-terminal de leurs protéines correspondantes et 38 commencent par l'acide aminé juste après la méthionine d'initiation. Les séquences de ces peptides commencent soit par une alanine (21 peptides), une serine (16), une thréonine (1). Comme cela a été rapporté par Polevoda, ces résidus font partie des résidus les plus fréquemment acétylés avec une fréquence proche de 100% pour l'acétylation des alanines et des sérines [217].

c Données MS/MS

La détermination de ces modifications par spectrométrie de masse ne se fait pas de la même façon selon la modification. En effet, certaines sont labiles (sulfation, glycosylation) et seront perdues avant la fragmentation du peptide lui-même. Dans ce cas, le peptide et la modification sont identifiés selon la différence de masse théorique calculée d'après la séquence obtenue et la masse expérimentale du parent fragmenté. Cependant, la localisation exacte de la modification ne peut pas être déterminée [218]. D'autres au contraire sont stables (acétylation, méthylation, amidation, formation d'un acide pyroglutamique) : dans ce cas, le spectre de fragmentation est similaire à celui du peptide non modifié, à l'exception de l'acide aminé, dont la masse est changée en fonction de la MPT. Dans ce cas, la localisation de la modification peut être déterminée de manière exacte. Les Figure 50 et Figure 51 présentent des spectres caractéristiques où ces deux modifications ont pu être localisées.

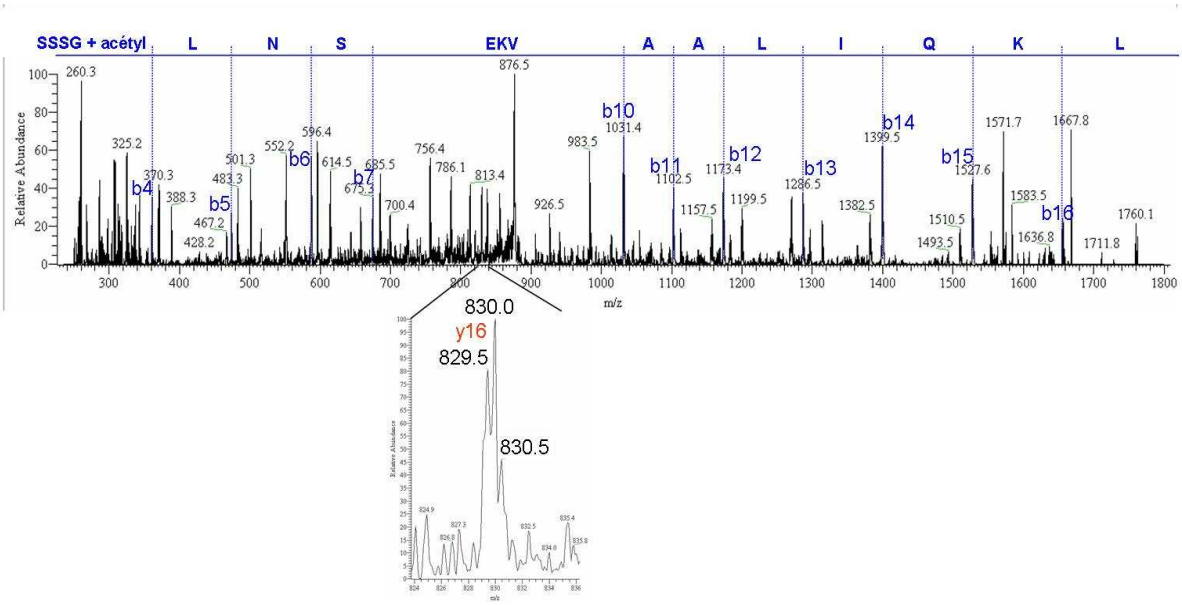


Figure 50 : Interprétation validée d'un spectre MS/MS lors de la confrontation dans la banque Swiss-Prot des données LC1D-FTMS obtenues sur un échantillon de SCP, avec l'hypothèse que les peptides sont acétylés.
Mise en évidence de la séquence SSSGLNSEKVAALIQKL, spécifique de l'hydrolase bleomycine. L'observation de la série des fragments b jusqu'à b4 permet de valider la présence d'une acétylation sur la séquence SSSG. La présence du fragment y16 permet de valider l'acétylation de la sérine N-terminale et non d'une sérine interne.

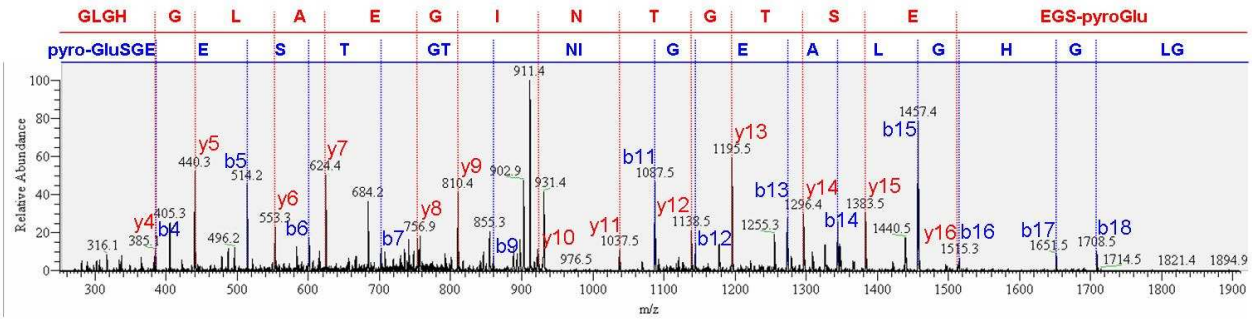


Figure 51 : Interprétation validée d'un spectre MS/MS lors de la confrontation dans la banque Swiss-Prot des données LC1D-FTMS obtenues sur un échantillon de SCP, avec l'hypothèse d'une pyroglutamisation N-terminale.
Mise en évidence de la séquence QSGEESTGTNIGEALGHGLG, spécifique de la dermokine, par l'observation d'une large série de fragments y complétée par la série des fragments b.

Pour améliorer la couverture du protéome et du peptidome, il est donc nécessaire de prendre en compte les modifications possibles. Dans les études d'échantillons digérés, cela a été réalisé en utilisant la fonction *error tolerant* et pour les échantillons de peptides endogènes en prenant en compte l'oxydation des méthionines, l'acétylation N-terminale et la pyroglutamisation des glutamines N-terminales en modifications partielles.

III-A.7. Validation des protéines identifiées par LC-MS/MS

Dans le cadre d'analyse protéomique par spectrométrie de masse, une très grande quantité de données sont générées. L'interprétation correcte des spectres MS/MS et la déduction des protéines représentées par les peptides identifiés représentent un challenge d'un point de vue notamment statistique.

La plupart de nos analyses, et en particulier toutes celles concernant des échantillons de peau, ont été réalisées en triplicat. Se pose alors la question des paramètres de validation à appliquer en termes de nombre de peptides par protéines, de score protéique et de nombre de réplicats où la protéine a été identifiée.

Classiquement, les paramètres utilisés au laboratoire tendaient à valider les protéines identifiées avec au moins deux séquences d'acides aminés différentes et un score protéique supérieur à 60 (soit deux peptides de score supérieur à 30). Cependant, lorsque nous avons réalisé nos analyses en triplicat, nous nous sommes rendu compte que certaines protéines attendues dans nos échantillons n'étaient pas validées avec ce paramètre. Par exemple, la dermcidine est une protéine exprimée au niveau des glandes sudoripares qui est sécrétée dans la sueur et transportée jusqu'à la surface de l'épiderme [219, 220]. Dans nos analyses de peau en LC1D-FTMS/MS, cette protéine n'était pas validée avec deux peptides de score supérieur à 30, mais pourtant était identifiée de façon répétable en triplicat avec 1 peptide supérieur à 60. D'autre part, même si avec ce paramètre, 80% des protéines validées sont identifiées en triplicat, nous avons choisi de valider également les protéines identifiées seulement en duplicat. En effet, ces protéines sont également exprimées dans la peau et lorsque les analyses sont menées en LC2D, aucun doute n'est plus possible sur leur présence du fait du grand nombre de peptides permettant leur identification. En revanche, les protéines identifiées avec un seul peptide, même avec un score de 60 dans un seul des triplicats, sont considérées comme des faux positifs et sont donc éliminées.

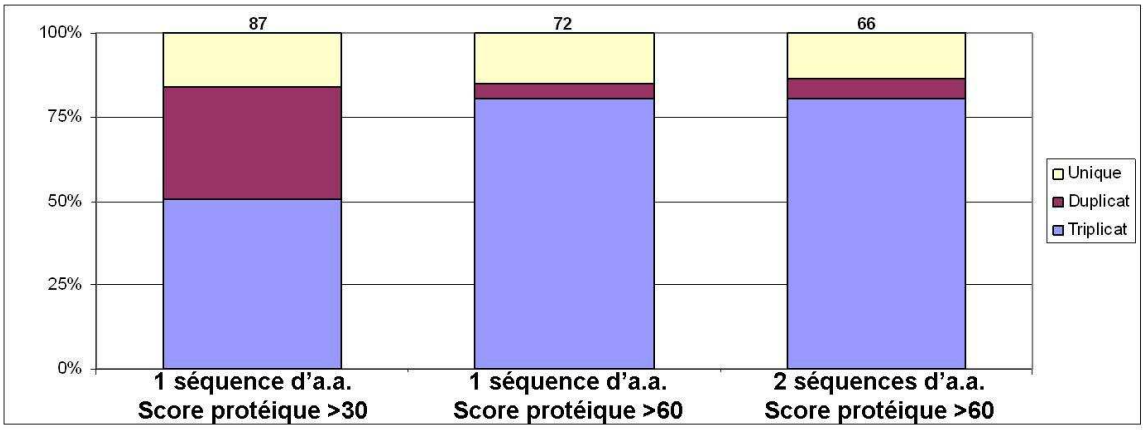


Figure 52 : Répétabilité des identifications protéiques sur l'analyse d'un échantillon de SCP digéré par LC1D-FTMS/MS.
Pour chaque paramètre, est rappelé le nombre total de protéines différentes validables tout réplicat confondu.

Il faut noter que même si 80% des protéines sont validées au moins en triplicat (et 85% en duplicat), il n'en est pas de même pour les peptides. En effet, seuls 58% des peptides (=séquences d'a.a.) sont identifiés en triplicat et 77% en duplicat. Lorsque les applications sont plus spécifiques qu'une analyse protéomique qualitative (recherche de modifications, recherche de peptides endogènes), il est donc intéressant de prendre en compte les informations peptidiques apportées par les différents réplicats.

Ainsi, pour les échantillons non digérés, nous avons validé les protéines identifiées avec au moins un peptide identifié en triplicat et avec un score protéique supérieur à 60. Une fois la protéine validée, tous les peptides identifiés, en triplicat ou non, ont été pris en compte.

III-B. Applications biologiques

III-B.1. Peptidome de base du *stratum corneum*

Protocoles de reference :
Préparation des échantillons non digérés : **V-A.1.2.** p.145, **V-A.1.3.** p.145, **V-A.1.4.** p.146
Expérience LC-MS/MS : **V-A.3.1. b** p.148, **V-A.5.2.** p.153
Recherche dans les banques et validation : **V-A.6.1. a** p.154, **V-A.6.2. b** p.155

III-B.1.1. Identification des peptides

Pour étudier le dégradome de base du *stratum corneum*, nous avons développé une stratégie basée sur l'ultrafiltration, la reconcentration sur phase solide et l'analyse par LC-MS/MS des échantillons. Dans les études peptidomiques, **l'ultrafiltration** par centrifugation est la méthode la plus utilisée pour extraire des

peptides et éliminer les protéines de haut poids moléculaire [6, 221, 222]. Cette technique exploite un mécanisme d'exclusion stérique grâce à des membranes de seuil d'exclusion spécifique. Ainsi Zheng *et al.* ont utilisé cette procédure sur des échantillons de sérum et ont identifié 350 peptides endogènes uniques provenant de 61 protéines [223]. Nous avons également utilisé l'ultrafiltration en tant qu'étape de pré-fractionnement. Combiner des membranes d'ultrafiltration de différents seuils (10kDa et 5kDa) nous permet de séparer les peptides en différentes fractions, ce qui augmente la sensibilité de nos identifications. Cependant, de la même manière que cette étape permet de concentrer les composés d'intérêt, les contaminants sont eux aussi concentrés. Leur présence peut réduire drastiquement les performances d'analyse des peptides par spectrométrie de masse. C'est pourquoi, nous avons ajouté une étape de **reconcentration** des peptides sur phase solide de type ZipTip^{TM} C18, ce qui permet également de dessaler l'échantillon. Enfin l'utilisation d'une **séparation chromatographique** préliminaire avant l'analyse par spectrométrie de masse est nécessaire pour les études d'échantillons complexes en particulier dans les approches peptidomiques : comme l'échantillon n'est pas digéré à la trypsine, certaines protéines ne seront identifiées que grâce à très peu de peptides, voire un seul. Les analyses MS ont été réalisées sur un instrument hybride LIT-FTICR. Ce type d'appareil nous a permis à la fois de monter en sensibilité et en gamme dynamique afin d'identifier des peptides minoritaires. De plus, la masse des précurseurs est mesurée précisément dans la cellule ICR pendant le *scan* SIM. Les recherches dans les banques sans digestion *in silico* sont compensées en diminuant la tolérance de masse à 10ppm. Ainsi le nombre de faux positifs est significativement réduit et les identifications sont plus sûres.

Nous avons analysé des **extraits de peptides endogènes** de deux types d'échantillons différents : du *stratum corneum* plantaire où le phénomène de desquamation est moins complet et du *stratum corneum* non plantaire obtenu à partir de prélèvements non invasifs (*stripping)* au niveau de la jambe. Au total, **1098 séquences d'acides aminés différentes** ont été identifiées, appartenant à **71 protéines** (voir le Tableau 10). 21 de ces protéines sont de masse inférieure à 30kDa et peuvent être classifiées comme des protéines de bas poids moléculaire, comme par exemple la cystatine A, la caspase 14, la suprabasine ou encore la calmoduline-like 5. La **composition en peptides** des deux types d'échantillons est relativement différente. Ainsi certaines protéines ne sont détectées que dans les échantillons de *stratum* non plantaire comme la filaggrine 2 et au contraire d'autres comme la *fatty acid-binding protein* ne sont détectées que dans les échantillons de *stratum* plantaire. De plus lorsqu'une protéine est présente dans les deux échantillons, les peptides identifiés peuvent être partiellement différents.

La plupart des protéines identifiées peuvent être divisées en **4 classes fonctionnelles** : les protéines structurelles de l'épiderme, les enzymes, les inhibiteurs de protéases et les protéines d'adhésion.

Parmi les **protéines structurelles**, se trouvent bien évidemment les **kératines** qui représentent 80% des protéines totales des cornéocytes [128].

Parmi les **enzymes** mises en évidence, se trouve par exemple la **caspase 14** impliquée dans les phénomènes d'inflammation ou d'apoptose mais aussi dans la prolifération, la différenciation et la cornification [224]. Cette enzyme présente un profil d'expression unique parmi les membres de la famille des caspases puisqu'elle est principalement présente dans les épithéliums cornifiés [225]. Dans la peau, la caspase 14 est uniquement exprimée dans le couches différenciées et cornifiées et dans le follicule pileux [226]. Les procaspases sont constituées d'un prodomaine, d'une grande sous-unité p20 et d'une petite sous-unité p10. Leur activation est induite par dimérisation et (auto)-clivage protéolytique. Contrairement aux autres caspases qui sont processées au niveau d'un résidu aspartate, la caspase 14 l'est au niveau d'un site atypique entre l'Ile152 et la Lys153 [227]. Or nos données ont mis en évidence plusieurs peptides endogènes de caspase 14 avec la Lys153 en N-terminal. En particulier le peptide ^{153}KDSPQTIPTYTDALHV^{168} est le peptide le plus souvent identifié dans les échantillons de *stratum* non plantaire ce qui n'est pas le cas dans les échantillons de *stratum* plantaire où le peptide le plus identifié est le peptide ^{87}MAHGREGFLKGEDGEMVKLENL^{108} du prodomaine ou de la grande sous-unité. Cette différence pourrait être due à une **activation de la caspase 14 différente** selon le type d'échantillon.

Parmi les **inhibiteurs** de protéases identifiés, on trouve la **cystatine A**. Inhibiteur de protéases à cystéine, il s'agit d'une des protéines précurseurs de l'enveloppe cornifiée des kératinocytes et cette protéine est exprimée durant la dernière étape de différenciation des kératinocytes. Dans nos analyses, nous avons mis en évidence différents peptides tout au long de la séquence. En particulier, l'un des peptides endogènes les plus observés correspond au fragment ^{2}IPGGLSEAKPATPEIQE^{18}. Or il a été montré que la délétion des 8 premiers acides aminés de la cystatine amenait à sa complète inactivation [228]. Il est donc possible que la présence du peptide 2-18 dans nos extraits soit la trace de l'inactivation de cet inhibiteur.

Tableau 10 : Protéines dont des peptides endogènes ont été identifiés dans des échantillons de peau par nanoLC-MS/MS. Combinaison des résultats des fractions inf 5kDa et 5-10kDa.
SCP : *stratum corneum* plantaire ; PA : poudre acétonique de *stratum corneum* non plantaire. # peptides et score en gras : au moins un peptide spécifique de la protéine dans l'échantillon. # peptides et score en italique : aucun peptide identifié en triplicat dans l'échantillon.

N°Swissprot	Nom de la protéine	MW kDa	Score		# séquences a.a. uniques	
			SCP	PA	SCP	PA
K1C9	Keratin, type I cytoskeletal 9	62	**4745**	**568**	**67**	**7**
DMKN	Dermokine	47	**4030**	**964**	**66**	**13**
K2C1	Keratin, type II cytoskeletal 1	66	**3014**	**2058**	**46**	**31**
DESP	Desmoplakin	332	**2698**	**2537**	**50**	**46**

DSG1	Desmoglein-1	114	**2517**	**2687**	**49**	**47**
DSC1	Desmocollin-1	100	**2213**	**2679**	**45**	**52**
G3P	Glyceraldehyde-3-phosphate dehydrogenase	36	**1874**	**2394**	**36**	**45**
CDSN	Corneodesmosin	51	**2358**	**1690**	**40**	**28**
K22E	Keratin, type II cytoskeletal 2 epidermal	66	**153**	**2112**	**4**	**37**
K1C10	Keratin, type I cytoskeletal 10	59	**300**	**2050**	**5**	**27**
CYTA	Cystatin-A	11	**1831**	**951**	**36**	**19**
FILA2	Filaggrin-2	248	0	**1532**	0	**26**
CALL5	Calmodulin-like protein 5	16	**1485**	**383**	**26**	**7**
K2C6A	Keratin, type II cytoskeletal 6A	60	**1442**	***159***	**25**	***3***
K2C6C	Keratin, type II cytoskeletal 6C	60	**1298**	***159***	**23**	***3***
K2C5	Keratin, type II cytoskeletal 5	62	**1264**	**831**	**22**	**15**
DSC3	Desmocollin-3	100	**515**	**1210**	**11**	**27**
K2C6B	Keratin, type II cytoskeletal 6B	60	**1094**	*81*	**19**	*2*
ARGI1	Arginase-1	35	**439**	**1041**	**9**	**21**
FILA	Filaggrin	435	**287**	**985**	**8**	**27**
SBSN	Suprabasin	25	**958**	**160**	**19**	**3**
CASPE	Caspase-14	28	**441**	**945**	**11**	**21**
TGM3	Protein-glutamine gamma-glutamyltransferase E	77	**368**	**922**	**9**	**23**
KPYM	Pyruvate kinase isozymes M1/M2	58	**911**	**161**	**21**	**3**
K1C16	Keratin, type I cytoskeletal 16	51	**874**	***150***	**14**	***3***
ANXA2	Annexin A2	39	**234**	**857**	**5**	**22**
CATA	Catalase	60	**384**	**680**	**8**	**14**
BLMH	Bleomycin hydrolase	53	**303**	**653**	**4**	**12**
HORN	Hornerin	282	**587**	0	**9**	0
CYTM	Cystatin-M	17	0	**478**	0	**9**
FABP5	Fatty acid-binding protein, epidermal	15	**439**	0	**9**	0
ALBU	Serum albumin	69	**373**	0	**8**	0
HUTH	Histidine ammonia-lyase	73	0	**336**	0	**7**
PEBP1	Phosphatidylethanolamine-binding protein 1	21	**301**	**302**	**7**	**6**
GRP78	78 kDa glucose-regulated protein	72	***297***	**117**	***7***	**2**
PLEC1	Plectin-1	531	0	234	0	4
PSA5	Proteasome subunit alpha type-5	26	***104***	**229**	***2***	**3**
TPIS	Triosephosphate isomerase	27	0	**208**	0	**4**
K1C17	Keratin, type I cytoskeletal 17	48	**199**	*54*	**5**	*1*
ZA2G	Zinc-alpha-2-glycoprotein	34	0	**194**	0	**4**
DCD	Dermcidin	11	**180**	**186**	**4**	**5**
PNPH	Purine nucleoside phosphorylase	32	**182**	**61**	**3**	**1**
ACTB	Actin, cytoplasmic 1	42	**163**	0	**5**	0
ACTG	Actin, cytoplasmic 2	42	**163**	0	**5**	0
S10A7	Protein S100-A7	11	**161**	0	**4**	0
K1C14	Keratin, type I cytoskeletal 14	52	***161***	***98***	***4***	***2***
SPB3	Serpin B3	45	**159**	0	**4**	0
PSA7	Proteasome subunit alpha type-7	28	0	**152**	0	**3**
S10A8	Protein S100-A8	11	**148**	0	**3**	0
CAN1	Calpain-1 catalytic subunit	82	0	**138**	0	**3**
A26CB	ANKRD26-like family C member 1B	121	133	0	4	0
PARK7	Protein DJ-1	20	**119**	***89***	**3**	***2***
MDHM	Malate dehydrogenase, mitochondrial	35	**119**	**94**	**2**	**2**
SPB12	Serpin B12	46	**101**	***108***	**2**	***2***
GLRX1	Glutaredoxin-1	12	***42***	**106**	***1***	**2**
HSPB1	Heat shock protein beta-1	23	**100**	0	**2**	0
SPB7	Serpin B7	43	0	**98**	0	**2**
GSHB	Glutathione synthetase	52	**96**	**98**	**1**	**1**

SODC	Superoxide dismutase [Cu-Zn]	16	**82**	**98**	**2**	**2**
S1A7A	Protein S100-A7A	11	97	0	2	0
HSP7C	Heat shock cognate 71 kDa protein	71	**97**	*47*	**2**	*1*
PPGB	Lysosomal protective protein	54	0	**90**	0	**2**
STRC	Stereocilin	193	**90**	***34***	**2**	***1***
STRCL	Putative stereocilin-like protein	192	**90**	***34***	**2**	***1***
THIO	Thioredoxin	12	0	**89**	0	**2**
PSB6	Proteasome subunit beta type-6	25	***76***	**86**	***1***	**1**
APRV1	Retroviral-like aspartic protease 1	37	0	**83**	0	**1**
PK1L3	Polycystic kidney disease protein 1-like 3	196	**82**	***72***	**2**	***2***
A26CA	ANKRD26-like family C member 1A	121	66	0	2	0
HSP72	Heat shock-related 70 kDa protein 2	70	65	*45*	2	*1*
PSA7L	Proteasome subunit alpha type-7-like	29	0	62	0	1

III-B.1.2. Comparaison topologique (PA/SCP)

La **structure et les fonctions** de la peau palmo-plantaire sont assez différentes en comparaison de la peau plus fine des autres parties du corps. Nous avons comparé des échantillons de différentes zones de peau à savoir de la jambe (PA) qui est une zone où la desquamation s'effectue normalement et de la plante de pied (SCP) où au contraire la desquamation est modifiée avec un épaississement de la couche de *stratum corneum*. Cette épaisseur peut être reliée au degré de stress mécanique subi par les zones palmo-plantaires [229].

L'analyse des deux types d'échantillons a mis en évidence des protéines plus ou moins dégradées selon le type de peau étudié. Les protéines identifiées peuvent être classées en **4 groupes selon leur état de dégradation protéolytique** (Figure 53). Des protéines peu dégradées avec l'identification de 1 ou 2 peptides différents, des protéines dégradées de manière plus importante au niveau du *stratum* plantaire comme certaines kératines, des protéines au contraire plus dégradées dans les zones de peau non palmo-plantaire et enfin des protéines qui présentent le même profil de dégradation quelle que soit la zone.

Parmi ces protéines, se trouvent notamment les protéines du **cornéodesmosome**, ce qui semble montrer que le phénomène de desquamation a un effet qualitativement identique sur ces protéines. En revanche, l'accumulation de peptides endogènes issus de la dégradation des kératines dans les zones plantaires pourrait indiquer une dégradation exacerbée de ces protéines, ce qui protégerait d'un point de vue quantitatif la dégradation des protéines cornéodesmosomales et retarderait la desquamation des couches supérieures de *stratum*. Au contraire, si on associe l'épaississement de la peau à un stress mécanique plus important, le processus de desquamation étant ralentie, toutes les protéines sont plus longtemps en contact avec les protéases de la peau ; les kératines, quoique plus résistantes, finissent par

être dégradées elles aussi avec une accumulation de peptides endogènes qu'il nous ait alors possible d'analyser.

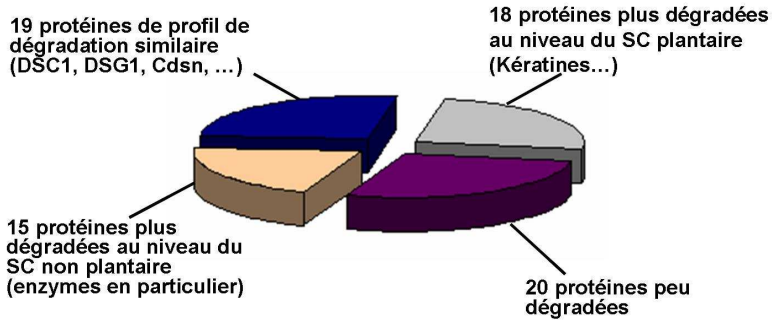


Figure 53 : Comparaison topologique des protéines identifiées dans le *stratum corneum* plantaire ou non.

Pour les protéines communes aux deux types d'échantillons, nous avons représenté le rapport du nombre de spectres MS/MS de ces deux types d'échantillons PA/SCP. Si un peptide est fragmenté plusieurs fois, toutes ses contributions ont été comptabilisées. En effet, il se pourrait que les **produits de dégradation** d'une protéine soient les mêmes dans les deux échantillons mais que leurs **quantités** soient différentes. Ceci s'apparente aux méthodes de quantification sans marquage isotopique par ***spectral counting*** [230] : le compte du nombre de spectres permet d'estimer les changements relatifs d'abondance de chaque protéine. Cela présuppose que la couverture d'une protéine augmente avec son abondance. Une augmentation du nombre de MS/MS traduit donc que la protéine est plus présente dans l'échantillon. Dans notre cas, cela traduit que la protéine en question est **soit plus présente soit plus dégradée** puisque nous ne nous intéressons qu'aux peptides endogènes, produits de la protéolyse des protéines étudiées. En ce qui concerne les protéines identifiées dans un seul des groupes d'échantillons, nous avons considéré un nombre minimum de peptides fictifs de 3 pour l'échantillon où la protéine n'a pas été identifiée. En effet, l'un des critères de validation que nous avons choisis est l'identification d'au moins un peptide de la protéine en triplicat.

Parmi les protéines communes aux deux types d'échantillons (Figure 54A) mais identifiées de façon différentielle, on peut noter par exemple le cas de deux kératines dont 7 à 10 fois plus de peptides endogènes sont identifiés dans les échantillons de *stratum corneum* non plantaire. Il s'agit des **kératines K1C10 et K22E**. Or cette paire devient majoritaire dans les couches les plus externes et les plus différenciées de l'épiderme [134]. *A contrario*, 30 fois plus de peptides de la **K1C9** sont identifiés dans le *stratum corneum* plantaire ; or l'expression de cette protéine est confinée dans l'épiderme palmo-plantaire [136]. Il est donc possible que la surabondance de peptides endogènes de ces deux protéines ne soit liée qu'à la **surexpression des protéines** elle-même et non à leur dégradation exacerbée.

Lors de cette comparaison topologique, nous avons identifié près de 5 fois plus de peptides endogènes pour la **filaggrine** dans les échantillons de *stratum* non plantaire que plantaire. De plus, des peptides de filaggrine 2 n'ont été identifiés que dans les échantillons non plantaires. Or, la filaggrine est aussi exprimée au niveau du *stratum corneum* plantaire [231]. L'identification d'un plus grand nombre de peptides endogènes dans la zone non plantaire pourrait donc être directement liée à une **dégradation plus importante** au niveau de cette zone. Cette protéine est synthétisée sous forme de pro-protéine et progressivement protéolysée : d'abord dégradée en monomère de filaggrine puis complètement en acides aminés dans les couches les plus superficielles du *stratum corneum* ce qui permet de maintenir l'hydratation de l'épiderme [130]. Cette variation, en termes de dégradation, que nous avons observée au niveau des peptides endogènes, peut donc être réliée à l'hydratation amoindrie des zones de peau palmo-plantaires. De plus, il a récemment été montré que **l'hydrolase bleomycine** et dans une moindre mesure la **calpaïne-1** jouaient un rôle important dans cette dégradation [232]. Or, nous avons identifié plus de peptides de ces deux protéases dans les échantillons de *stratum corneum* non plantaire : ceci pourrait traduire uniquement une surexpression (et non une surdégradation) de ces protéases et donc expliquer la dégradation des filaggrines dans ces échantillons.

Dans les échantillons de *stratum corneum* plantaire, se trouve la **dermokine** dont la dégradation est au moins cinq fois plus importante dans ce tissu que dans le *stratum corneum* non plantaire. Cette protéine est une protéine sécrétée dont la fonction n'est pas encore clairement définie. Il a déjà été décrit une transcription abondante au niveau de l'épiderme mais sans qu'une étude topologique n'ait été réalisée [233].

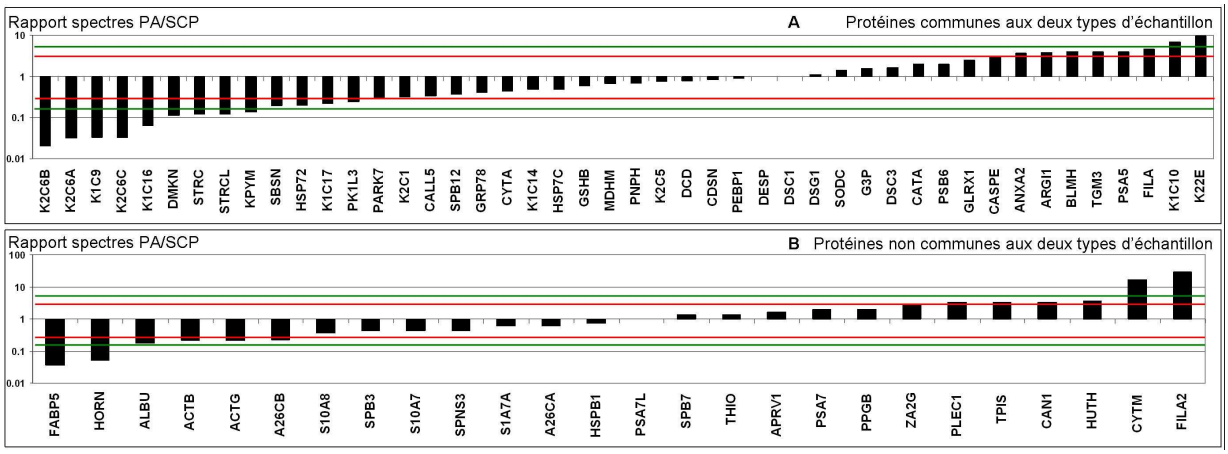


Figure 54 : Comparaison des protéines identifiées selon le type de peau étudié.
Rapport du nombre de peptides identifiés dans les échantillons de SC non plantaire sur le nombre de peptides identifiés dans les échantillons de SC plantaire. A : Protéines identifiées dans les deux échantillons. B : Protéines identifiées dans un seul des deux types d'échantillons. En vert, rapport 5 ou 1/5. En rouge 3 ou 1/3.

III-B.1.3. Les protéines des cornéodesmosomes

Ce type d'approche peptidomique nous a permis d'obtenir des informations sur le **profil de dégradation des protéines cornéodesmosomales**. En effet 26% des peptides identifiés correspondent de manière inattendue à des protéines impliquées dans les propriétés adhésives des cornéodesmosomes comme les **cadhérines, la cornéodesmosine, la desmoplakine ou la pkakoglobine**. Ceci est assez remarquable car lors de nos analyses classiques protéomiques sur des échantillons de *stratum corneum* plantaire digéré, assez peu de peptides de ces protéines ont été identifiés. Ainsi par exemple, en analyse protéomique LC1D, nous n'avons identifié que **9 peptides tryptiques** de desmocolline 1 alors que nous avons pu identifier **41 peptides endogènes**. Cette différence peut donc mettre en évidence une dégradation exacerbée de ces protéines par rapport à d'autres familles de protéines. Il faut tout de même noté que les échantillons ne sont pas préparés de manière identique. En effet, dans le cas des échantillons de peptides endogènes, les échantillons de peau subissent une première étape d'extraction dans un tampon durant une nuit. En revanche, pour les échantillons digérés, aucune étape d'extraction n'a été utilisée : la peau est directement digérée en la considérant comme un morceau de gel. Ainsi, il est possible que l'étape d'extraction permette d'accumuler plus de peptides endogènes de toutes les familles de protéines dégradées alors que la digestion directe ne conduit qu'à l'obtention de peptides protéolytiques à partir des protéines les plus accessibles en surface.

a Les cadhérines

Nous avons découvert **141 peptides uniques** provenant de 3 protéines de la famille des cadhérines : **la desmocolline 1 (DSC1), la desmocolline 3 (DSC3) et la desmogléine 1 (DSG1)**. Dans l'épiderme, les isoformes des cadhérines sont exprimées de manière différente : les formes 2 et 3 sont exprimées dans les couches profonde

s alors que les formes 1 le sont dans les couches supérieures [158-160]. Il n'est donc pas surprenant d'identifier principalement des peptides endogènes des formes 1. L'observation de peptides de DSC3 est plus singulière même si ces peptides peuvent correspondre à une dégradation de la protéine dans les couches plus profondes mais avec subsistance et accumulation jusqu'aux couches supérieures auxquelles nous avons accès. Il faut noter qu'environ deux fois plus de peptides de DSC3 (uniques ou redondants) ont été identifiés dans les échantillons de *stratum corneum* non plantaire, ce qui impliquerait donc une accumulation plus importante depuis les couches profondes au niveau d'un épiderme où la desquamation a lieu normalement. (Tableau 10 et Figure 54). Cette disproportion mise en parallèle d'une identification semblable au niveau des formes DSC1 et DSG1 pourrait mettre en évidence un rôle plus important de la DSC3 au niveau de la desquamation.

Sur les figures suivantes (Figure 55), nous avons représenté en couleurs les peptides identifiés en fonction du domaine auquel ils appartiennent. Pour chaque protéine de cette famille, nous avons observé des peptides endogènes dans chacun de ses **« domaines cadhérines »** et aucun dans le domaine transmembranaire ni dans le domaine intracellulaire. Or ces protéines sont impliquées dans **l'adhésion entre desmosomes** [145] et les deux types de protéines (desmocolline et desmogléine) sont nécessaires dans des interactions hétérotypiques [161, 234]. Ces cadhérines sont également engagées dans des interactions homotypiques au niveau des jonctions d'adhérence [235]. L'observation de peptides endogènes uniquement dans les domaines cadhérines semble donc indiquer une **dégradation préférentielle** de ces domaines, ce qui vient bien confirmer le rôle prépondérant de ces domaines dans l'adhésion entre cellules.

Cependant, chaque domaine ne semble pas dégradé de la même manière. Par exemple, pour la desmocolline 1, nous avons identifié 27 peptides différents provenant du premier domaine cadhérine alors que nous n'avons identifié qu'entre 5 à 12 peptides des autres domaines. Le premier domaine semble donc être la cible préférentielle de dégradation. Ceci est intéressant : en effet, les liaisons adhésives des cadhérines impliquent la formation de dimères de feuillets entre les domaines 1 des cadhérines des cellules opposées [236]. Une dégradation visant préférentiellement le premier domaine pourrait donc être reliée à un rôle plus important de ce domaine dans l'adhésion entre cadhérines.

Cependant, d'un point de vue quantitatif, si on compte combien de fois chaque peptide a été identifié dans les différentes analyses, en moyenne les peptides du dernier domaine cadhérine sont plus souvent identifiés. Cela impliquerait que le premier domaine est dégradé en de plus nombreux sites mais de manière moins préférentielle. Cependant cette différence peut être simplement due à une meilleure ionisation de certains peptides endogènes du 5ème domaine comparativement aux peptides du 1er domaine. Il est difficile de tirer une conclusion sur ces données dans la mesure où l'efficacité d'ionisation en électrospray dépend de différents paramètres physico-chimiques, qui peuvent favoriser l'identification de certains peptides par rapport à d'autres en fonction de leur séquence en acides aminés.

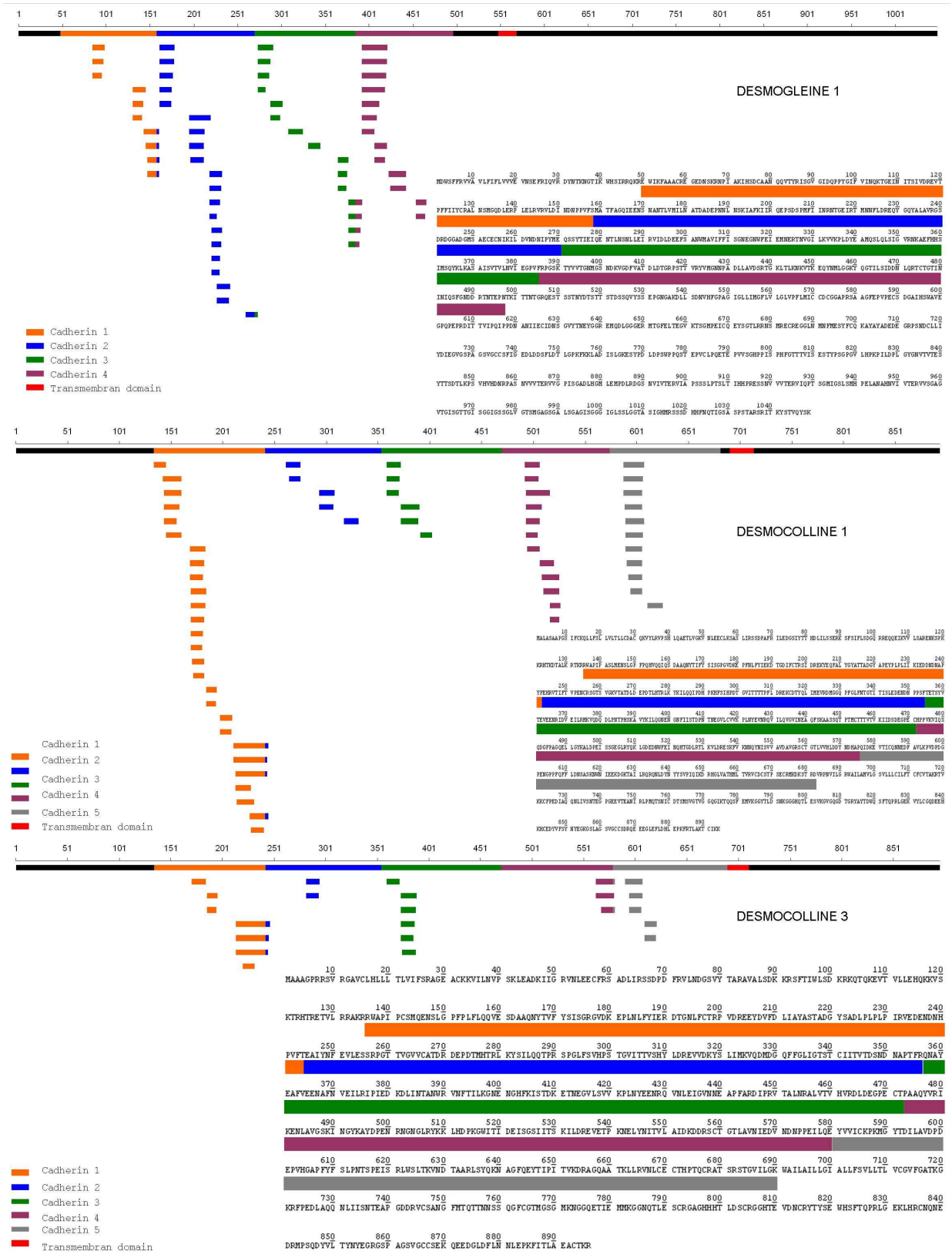


Figure 55 : Profil de dégradation des cadhérines identifiées.
Les peptides identifiés dans les échantillons de *stratum corneum* plantaire et non plantaire ont été regroupés.

b La cornéodesmosine

Parmi les autres protéines des cornéodesmosomes identifiées, de nombreux peptides de la cornéodesmosine ont été caractérisés. Ceci est également surprenant car cette protéine est très peu représentée dans les échantillons de *stratum corneum* plantaire digéré où seuls **3 peptides tryptiques** ont été séquencés en LC1D. En revanche, lors de nos analyses de peptides endogènes de *stratum corneum*, **43 peptides endogènes** ont été identifiés.

Les peptides que nous avons caractérisés ne couvrent pas l'intégralité de la séquence de la protéine. Comme le montre la Figure 56, des peptides ont été caractérisés dans la **boucle glycine N-terminale** ainsi que dans une **zone centrale de la protéine riche en sérines**. Aucun peptide de la boucle glycine C-terminale n'a été observé. Or il a été montré que cette protéine agit comme une molécule d'adhésion et que sa dégradation est un prérequis au phénomène de desquamation [146]. Ainsi les données de Jonca *et al.* ont permis de confirmer les propriétés d'interactions homotypiques de la cornéodesmosine. La boucle glycine N-terminale est suffisante mais non nécessaire pour la mise en place de ces interactions [163]. Notamment, il a été montré par Simon *et al.* que la protéine est d'abord dégradée au niveau des extrémités C- et N-terminales, puis elle perd la boucle glycine N-terminale et enfin une partie de la boucle glycine C-terminale ainsi qu'une zone centrale de la protéine. Après ces étapes de dégradation, ne subsiste dans les couches superficielles qu'une seule partie centrale de la protéine (entre les acides aminés 300 et 425 selon les auteurs) qui ne présente plus de caractère adhésif [165]. Nos observations de peptides uniquement entre les acides aminés 69 et 367 sont bien en accord avec ces données. Tant en termes qualitatifs que quantitatifs, un plus grand nombre de peptides ont été identifiés dans la zone centrale de la cornéodesmosine que dans la boucle N-terminale. Même si la boucle glycine semble plus importante pour conférer à la protéine ses qualités adhésives, nos résultats ne sont pas contradictoires : en effet, si la dégradation de la boucle glycine a eu lieu plus profondément que la dégradation de la zone centrale, il peut être normal d'avoir « perdu » les peptides endogènes qui nous permettraient de tracer cette première dégradation. Finalement, il faut tout de même noter que nous avons identifié 20 peptides différents entre les acides aminés 297 et 367 alors que cette zone est censée encore appartenir à la dernière forme observée par Simon *et al.*. Il pourrait donc être envisagé l'existence d'une **forme plus légère** que 15kDa (environ 5kDa entre 368-423) : forme si légère qu'elle sortirait des gels dans les conditions où ils ont été réalisés dans les travaux de Simon *et al.*

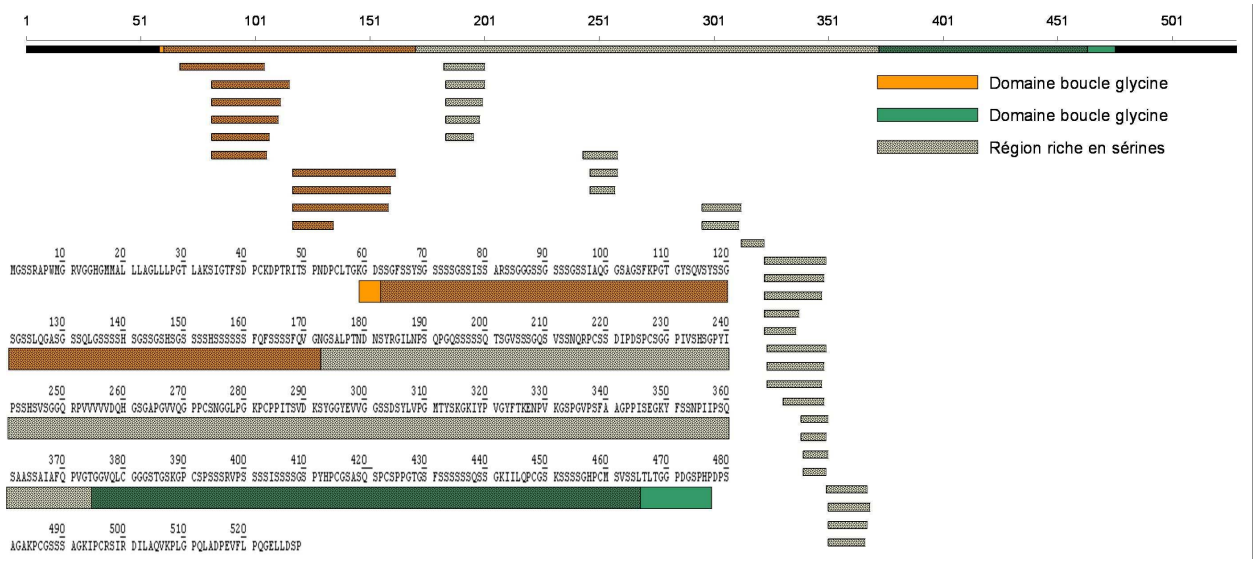


Figure 56 : Profil de dégradation de la cornéodesmosine.
Les peptides identifiés dans les échantillons de SC plantaire et non plantaire ont été regroupés.

c Les protéines de la plaque et les protéines armadillo

Parmi les autres protéines des cornéodesmosomes assurant une fonction importante d'adhésion, nous avons identifié une soixantaine de peptides différents de la **desmoplakine**. Les peptides endogènes caractérisés font partie d'un domaine particulier de la protéine situé dans sa zone C-terminale, noté **DPCT**, et qui interagit avec les **filaments intermédiaires (FI)** [137]. Ce domaine présente trois sous-domaines homologues A, B et C qui contiennent chacun 4,5 copies d'un motif plakin [151] (Figure 57). Au cours de nos analyses, nous avons identifié 2 peptides dans la zone C-terminale du domaine C et tous les autres sur l'intégralité du domaine B. Or, Choi *et al.* ont montré que tous les sous-domaines pris individuellement sont capables de lier la **vimentine** (membre de la famille des filaments intermédiaires) mais que le motif B-C présente une liaison plus forte que chaque motif B et C pris séparément [150]. Une dégradation préférentielle du domaine B pourrait indiquer qu'il suffit de protéolyser ce domaine pour **casser l'interaction desmoplakine-FI**.

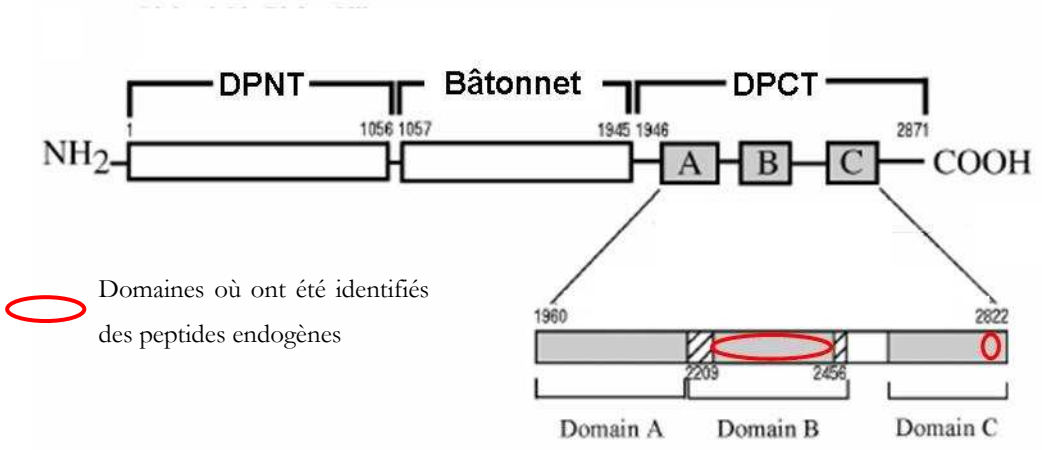


Figure 57 : Structure primaire de la desmoplakine (d'après [150]) montrant les différents domaines et sous-domaines de la protéine ainsi que les zones où nous avons identifié des peptides endogènes.
Les acides aminés de début et fin de domaine sont indiqués. DPNT : domaine N-terminal de la desmoplakine ; DPCT : domaine C-terminal de la desmoplakine.

De manière surprenante, nous n'avons identifié quasiment **aucun peptide des protéines armadillo** (comme la plakoglobine ou les plakophilines) alors que ces protéines interviennent également dans l'édifice global des interactions des protéines des cornéodesmosomes [237] et ont, de surcroît, des profils d'expression dans l'épiderme semblable aux cadhérines 1 ou à la desmoplakine [149, 237]. En effet, les cadhérines interagissent avec la plakoglobine au niveau de leur domaine intracellulaire [238], domaine dans lequel nous n'avons identifié aucun peptide non plus. La plakoglobine interagit également avec le domaine N-terminal de la desmoplakine [239] où de la même manière, nous n'avons identifié aucun peptide. Les plakophilines, au même titre que la plakoglobine, semblent jouer un rôle de jonction en interagissant à la fois avec les cadhérines, la desmoplakine et la plakoglobine [154].

Dans leur ensemble, nos résultats en termes de peptides endogènes des cornéodesmosomes montrent que la **desmoplakine et les cadhérines sont les protéines les plus dégradées**. Cela pourrait indiquer que les interactions intracellulaires desmoplakine-FI jouent un rôle tout aussi important que les interactions extracellulaires cadhérine-cadhérine dans le complexe cornéodesmosome-FI. Ainsi **la desquamation qui passe par la dégradation des cornéodesmosomes pourrait ne pas faire intervenir que la protéolyse des cadhérines seules mais également celle de la desmoplakine**.

III-B.1.4. Les peptides antimicrobiens

L'étude des peptides endogènes de la peau nous a également permis d'obtenir des informations sur les peptides antimicrobiens présents au niveau de ce tissu.

L'épiderme humain forme un **bouclier protecteur** consistant en une barrière mécanique dense composée de l'enveloppe cornée et de couches de lipides et de kératines pour empêcher l'invasion des microbes. Cependant, les cellules de l'épiderme ne sont pas seulement une barrière mécanique passive mais elles peuvent aussi sécréter des molécules pour combattre les agents pathogènes externes. Cette défense initiale est définie comme la réponse immunitaire innée et présente **3 fonctions majeures** : 1) reconnaissance des agents pathogènes, 2) induction de molécules pour activer les cellules dont le but est d'éliminer ces agents et 3) sécrétion de molécules pour combattre directement ces agents. Ces deux dernières fonctions sont notamment assurées par des molécules appelées peptides antimicrobiens (AMPs) [240]. La séquence de ces peptides est variable mais de nombreux AMPs présentent une activité antimicrobienne contre un large spectre de bactéries, de champignons et de virus. La peau est une source importante d'AMPs. Parmi eux, nous avons identifié, dans nos analyses, la dermcidine ainsi que des protéines de la famille S100.

a La dermcidine

La **dermcidine (DCD)** est une protéine de 110 acides aminés, identifiée dans les glandes sudoripares de type eccrine [220]. La DCD est sécrétée dans la sueur et n'est pas inductible par une blessure ou une inflammation de la peau [219]. Dans la sueur, le processus protéolytique génère un peptide de 47 acides aminés **(dermcidin-1, DCD-1)** qui présente une large **activité antimicrobienne** [220] mais avec un mécanisme restant à élucider. Cette protéine est sécrétée en tant que proprotéine puis processée dans la partie C-terminale de la proprotéine : les peptides de DCD sont alors distribués à la surface de la peau par la sueur [241]. Il a été montré par digestion *in vitro* que la cathepsine D était capable de générer certains fragments observés : nombre d'entre eux sont stables et quelques-uns présentent une plus forte activité antimicrobienne que le peptide DCD-1 [241, 242]. En 2002 puis en 2006, Flad puis Rieg ont identifié différents peptides dérivés de DCD-1 par SELDI-TOF et séquençage d'Edman [242, 243]. Rieg *et al.* ont également montré des différences topologiques dans la concentration de la sueur en peptides dérivés de la DCD avec des **quantités plus grandes dans la sueur des paumes, des bras et du front**. Cela souligne que la sécrétion de DCD est plus prononcée dans des zones du corps plus souvent exposées à des agents pathogènes ou à des traumatismes mineurs [242]. Au cours de nos analyses, nous n'avons identifié des peptides que dans la partie DCD-1 de la protéine et notamment le peptide LEK-26 (^{66}LEKGLDGAKKAVGGLGKLGKDAVEDL^{91}) déjà décrit. Nous avons notamment identifié deux peptides mettant en évidence le clivage entre la Leu91 et le Glu92, clivage par la cathepsine D décrit *in vitro* [241]. D'autre part, nous n'avons pas noté de différences majeures entre les zones étudiées puisque nous avons identifié 7 spectres MS/MS correspondants à des peptides de DCD dans le *stratum corneum* non plantaire contre 9 dans le *stratum corneum* plantaire même si un seul peptide a été identifié en commun dans les deux échantillons (Figure 58). Aucun de nos peptides

ne possède les résidus SSL (en N-terminal du peptide DCD-1). Or il a été montré que ces résidus semblaient importants pour assurer la fonction antimicrobienne [244]. Il pourrait donc être intéressant de tester l'activité antimicrobienne des nouveaux peptides que nous avons découverts pour confirmer ou infirmer ces observations.

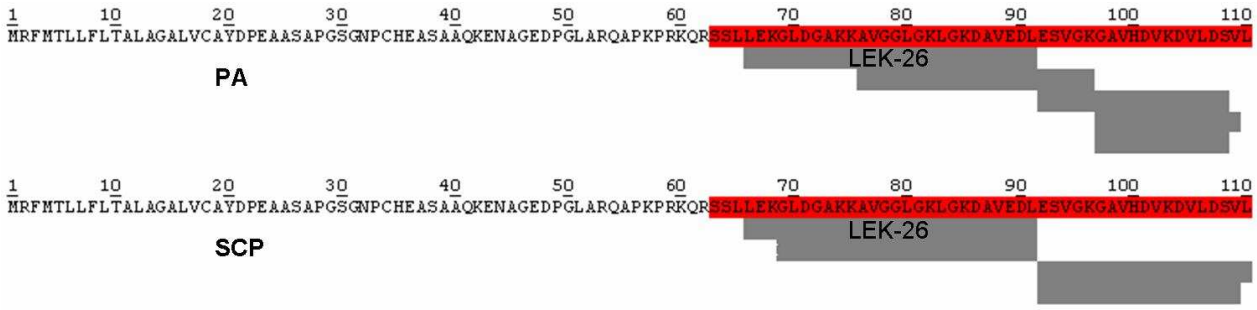


Figure 58 : Peptides dérivés de la DCD identifiés dans les échantillons de *stratum corneum* non plantaire (PA) et plantaire (SCP). En rouge est rappelé le peptide DCD-1.

b Les protéines de la famille S100

Les **protéines S100** forment une famille de protéines de bas poids moléculaire qui lient le calcium et qui partagent des motifs structurels communs. Parmi eux, se trouvent notamment **2 domaines hélice-boucle-hélice de liaison du calcium**, plusieurs domaines type hélice α et un domaine central flexible. La psoriasine (S100A7), membre de la famille des protéines S100 qui lient le calcium, est synthétisée abondamment par les patients souffrant de psoriasis [245]. Gläser *et al.* ont montré que la psoriasine possédait une **activité antibactérienne** et protégeait la peau humaine contre les infections par *E. coli* [246] en envisageant que sa capacité à séquestrer le zinc était impliquée dans ce phénomène. Cette protéine est présente en fortes concentrations au niveau du visage, du cuir chevelu, des paumes et des plantes des pieds [246]. De plus, elle est produite de manière constitutive par les **kératinocytes**. Des expériences de mutagenèse ont permis de montrer que l'inactivation du motif liant le zinc réduisait sa fonction antibactérienne mais que le motif N-terminal de liaison du calcium n'était pas nécessaire à cette activité [247]. En particulier la région centrale de la psoriasine incluant les acides aminés 35-80 semble suffire à la pleine activité antibactérienne de la protéine.

Au cours de nos analyses, nous avons identifié des peptides issus de la protéine S100A7 et S100A8 **uniquement dans les échantillons de *stratum corneum* plantaire**. Ces peptides sont distribués sur toute la séquence des protéines (Figure 59). Cependant, il faut noter que le peptide le plus souvent identifié pour la S100A7 est le peptide N-terminal acétylé alors que pour la S100A8, il s'agit d'un peptide central situé au niveau du motif C-terminal de liaison du calcium. L'observation des séquences

LADVFEKKDKNEDKKIDFSEF et LSLLGDIATDYHKQSHGAAPCSGGSQ qui se suivent dans la séquence protéique permet de mettre en évidence un clivage (F/L) qui pourrait être dû à la **cathepsine D**. Nous avons notamment découvert 2 peptides situés dans la zone potentiellement responsable de l'activité antimicrobienne : un test d'activité de ces peptides pourrait être intéressant pour confirmer cette hypothèse. Nos identifications viennent en tout cas confirmer l'observation en ESI-MS par Schröder de formes tronquées de la protéine S100A7A [248].

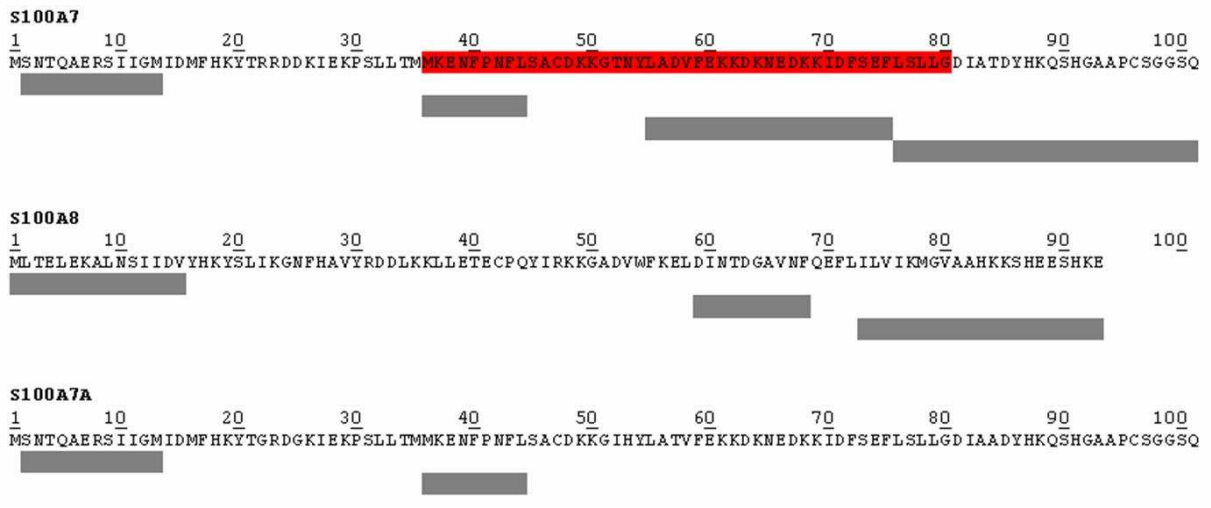


Figure 59 : Peptides dérivés des protéines de la famille S100, identifiés dans les échantillons de *stratum corneum*. En rouge est rappelée la zone centrale nécessaire à l'activité antibactérienne de la S100A7.

III-B.2. A la recherche des protéases impliquées : données issues de la LC-MS/MS

Protocoles de reference :

Préparation des échantillons non digérés : **V-A.1.2.** p.145, **V-A.1.3.** p.145, **V-A.1.4.** p.146

Expérience LC-MS/MS : **V-A.3.1. b** p.148, **V-A.5.2.** p.153

Recherche dans les banques et validation : **V-A.6.1. a** p.154, **V-A.6.2. b** p.155

III-B.2.1. Tendance générale

A partir des peptides identifiés en LC-MS/MS, il est possible de mettre en évidence quelles sont **les activités protéolytiques** les plus représentées dans nos échantillons. Pour cela, nous avons observé les extrémités N- et C-terminales des peptides endogènes identifiés ainsi que les acides aminés des protéines dont ils sont issus qui sont situés avant et après ces peptides.

Pour identifier les protéases susceptibles d'être impliquées, il est possible de travailler par des approches avec et sans *a priori*. En effet, d'après les données de la littérature, il est déjà connu que la **SCCE (***stratum corneum chymotryptic enzyme)***, la SCTE (***stratum corneum tryptic enzyme)* **et la cathepsine D** sont des protéases fortement engagées dans la dégradation des protéines des cornéodesmosomes. Connaissant les sites de coupure de ces protéases (Tableau 11), la proportion de sites de coupure observés en LC-MS/MS pouvant leur correspondre peut être déduite.

Tableau 11 : Sites de coupure des protéases décrites dans la dégradation des cornéodesmosomes.

Protéases	Sites de coupures
SCTE	K/ ; R/
SCCE	F/ ; L/ ; Y/ ; W/
Cathepsine D	F/ ; L/ ; /V ; /L ; /F ; /Y

Mais pour les sites de coupure différents de ces protéases, il est également possible de faire des recherches dans les banques de données notamment sur le site MEROPS (http://merops.sanger.ac.uk/index.htm). Cette base de données est une ressource d'informations en ce qui concerne les protéases mais également leurs inhibiteurs. Un outil de recherche permet notamment de rechercher à partir des acides aminés situés avant et après un site de coupure observé (P4P3P2P1/P1'P2'P3'P4'), quelles sont les protéases capables de casser une telle liaison.

Nous avons donc quantifié les différentes activités en comptant les sites de coupure observés correspondant à chaque activité. Nous avons compté **toutes les occurrences** de ces sites que ce soit dans le même peptide, mais observé de façon redondante, ou dans des peptides différents. Que ce soit pour les échantillons de *stratum corneum* plantaire ou non plantaire, l'activité principalement observée est une activité **chymotryptique** (Figure 60), ce qui correspond aux données bibliographiques. En effet, la **SCCE** est une protéase impliquée dans la desquamation et capable de dégrader la cornéodesmosine et la desmocolline 1 directement *in vitro* [184]. La seconde activité qui ressort est l'activité **tryptique** également décrite dans la littérature comme impliquée dans la dégradation des protéines cornéodesmosomales. En effet, la **SCTE** permet notamment d'**activer la SCCE** afin que celle-ci puisse dégrader la desmogléine 1 [184] et permet également de dégrader la cornéodesmosine [165]. Enfin, dans une moindre mesure, des sites pouvant correspondre à une activité de type **cathepsine D**, protéase impliquée dans **l'étape finale de la desquamation** [178], sont observés.

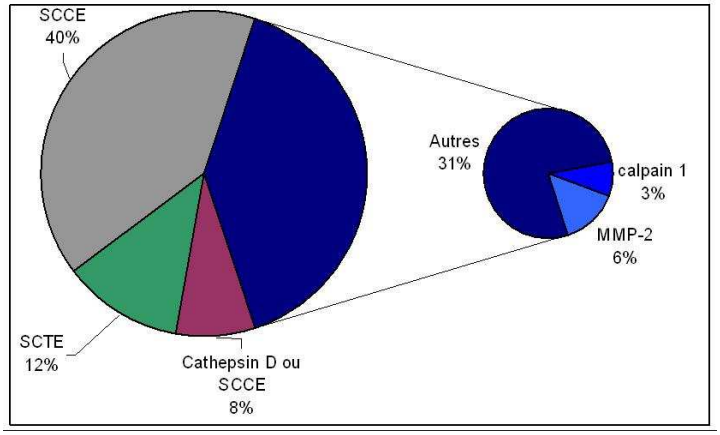


Figure 60 : Activités protéolytiques principalement observées d'après les données LC-MS/MS.

Cependant, une très grande partie des sites de coupure observés (40%) ne correspond à **aucune des trois activités** préalablement citées. Grâce aux recherches dans la banque de données MEROPS, nous avons pu retrouver que certains sites P1/P1' (représentant au moins 1% des sites observés) pouvaient correspondre à des activités type **calpaïne-1** (localisée dans le cytoplasme des kératinocytes de couches supérieures de l'épiderme et capable de dégrader notamment la filaggrine *in vitro* [232]) ou type gélatinase A (MMP-2, exprimée par les fibroblastes dermiques [249]).

III-B.2.2. Cas particulier : la cornéodesmosine

Dans le cas de la cornéodesmosine et pour les peptides identifiés avec un score supérieur à 30, le nombre de sites de coupure **chymotryptiques** est comparable aux résultats globaux sur toutes les protéines (38% pour la cornéodesmosine contre 40% en moyenne sur l'ensemble des protéines sont des sites chymotryptiques). En revanche, les sites **tryptiques** représentent environ 25% des sites identifiés sur la cornéodesmosine, soit plus de deux fois plus que pour les échantillons globalement. Cette protéine semble donc être une **cible privilégiée de la SCTE** par rapport à la moyenne. En ce qui concerne les autres sites, il faut constater une nette diminution de la proportion des sites type cathepsine D (de 8 à 1%) au profit principalement de l'activité SCTE déjà mentionnée et surtout d'une activité type **cathepsine L** (moins de 1% pour l'ensemble des protéines et 14% pour la cornéodesmosine). Ceci est extrêmement intéressant dans la mesure où Bernard *et al.* ont déjà mis en évidence une protéase nommée ***stratum corneum* cathepsin-L-like (SCCL)** capable de dégrader la cornéodesmosine *in vitro* [175]. Nos données qui proviennent de l'*in vivo* viennent ici confirmer ces informations. Le reste des sites observés pour la cornéodesmosine (soit encore 15%) ne peut être attribué à aucune protéase de la peau (Figure 61).

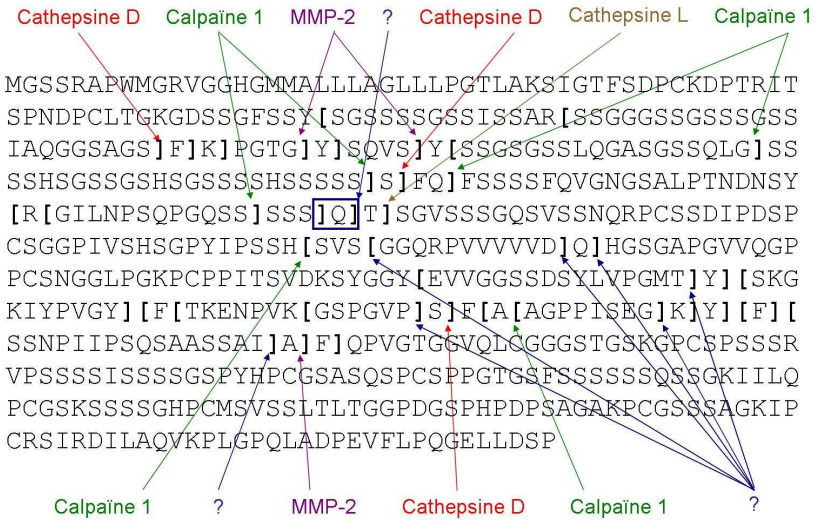


Figure 61 : Sites de coupure autres que SCCE et SCTE identifiés sur la cornéodesmosine.

III-B.3. A la recherche des protéases impliquées : analyses de peptides fluorescents de la cornéodesmosine

Afin de cibler les protéases impliquées grâce à une approche indépendante, nous avons fait synthétiser des **peptides fluorescents**. Il s'agit d'octopeptides synthétisés suivant la séquence de la cornéodesmosine et en regard des sites de clivages observés ou non en LC-MS/MS. Neuf peptides ont été synthétisés spécifiquement autour de sites observés et 4 autres dans des zones où aucun site n'avait été identifié. La séquence de la cornéodesmosine, les sites identifiés ainsi que les peptides synthétisés sont présentés sur la Figure 62.

MGSSRAPWMGRVGGHGMMALLLAGLLLPGTLAKSIGTFSDPCKDPTRIT
SPNDPCLTGKGDSSGFSSY[SGSSSSGSSISSAR[SSGGGSSGSSSGSS
IAQGGSAGS]F]K]PGTG]Y]SQVS]Y[SSGSGSSLQGASGSSQLG]SS
SSHSGSSGSHSGSSSSHSSSSS]S]FQ]FSSSSFQVGNGSALPTNDNSY
[R[GILNPSQPGQSS]SSS]Q]T]SGVSSSGQSVSSNQRPCSSDIPDSP
CSGGPIVSHSGPYIPSSH[SVS[GGQRPVVVVVD]Q]HGSGAPGVVQGP
PCSNGGLPGKPCPPITSVDKSYGGY[EVVGGSSDSYLVPGMT]Y][SKG
KIYPVGY][F[TKENPVK[GSPGVP]S]F[A[AGPPISEG]K]Y][F][
SSNPIIPSQSAASSAI]A]F]QPVGTGGVQLCGGGSTGSKGPCSPSSSR
VPSSSSISSSSGSPYHPCGSASQSPCSPPGTGSFSSSSSSQSSGKIILQ
PCGSKSSSSGHPCMSVSSLTLTGGPDGSPHPDPSAGAKPCGSSSAGKIP
CRSIRDILAQVKPLGPQLADPEVFLPQGELLDSP

Figure 62 : Séquence de la cornéodesmosine ; les crochets correspondent aux sites de clivage identifiés en LC-MS/MS.
En rouge sont représentés les peptides synthétisés aspécifiquement et en vert ceux synthétisés spécifiquement.

Ces peptides sont modifiés en N-terminal par un groupe **chromophore Dabcyl** et en C-terminal par un groupe fluorophore **Edans**. Le spectre d'absorption de celui-ci est centré sur 336nm et son spectre d'émission sur 490nm comme le spectre d'absorption du groupe Dabcyl. Ainsi, lorsque les deux groupes sont proches aucune fluorescence n'est mesurable alors que lorsqu'ils sont séparés cela devient possible. Cette propriété permet de suivre le clivage des peptides synthétisés lors des tests enzymatiques.

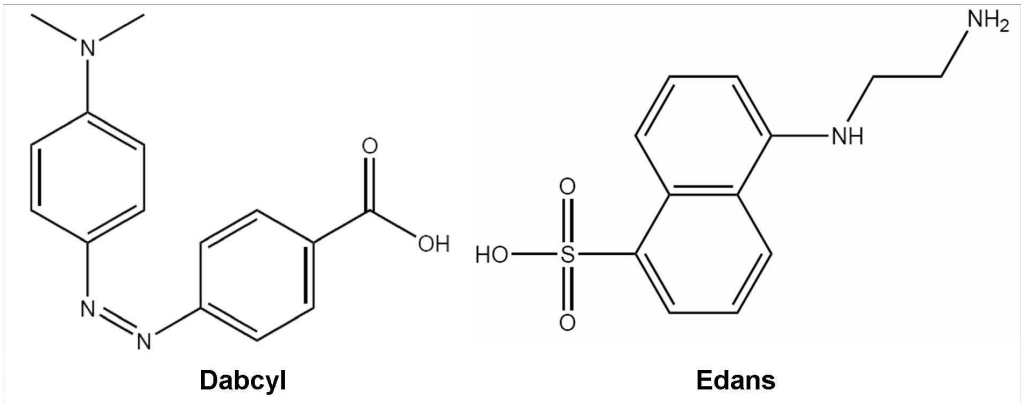


Figure 63 : Formule des groupes Dabcyl et Edans.

Les séquences des peptides, leurs masses et la nomenclature utilisées par la suite sont rappelées dans le Tableau 12.

Tableau 12 : Caractéristiques des peptides Dabcyl-Edans synthétisés.

Sigle	Séquence	Masse Théorique (MH^+)
S1 ou CDSN1	Dabcyl-QVSYSSGS-Glu(Edans)	1442,5Da
S2 ou CDSN2	Dabcyl-SSSSFQFS-Glu(Edans)	1504,8Da
S3 ou CDSN3	Dabcyl-NSYRGILN-Glu(Edans)	1564,7Da
S4 ou CDSN4	Dabcyl-SSQTSGVS-Glu(Edans)	1379,8Da
S5 ou CDSN5	Dabcyl-YGGYEVVG-Glu(Edans)	1473,1Da
S6 ou CDSN6	Dabcyl-PGMTYSKG-Glu(Edans)	1468,5Da
S7 ou CDSN7	Dabcyl-PVGFTKE-Glu(Edans)	1568,7Da
S8 ou CDSN8	Dabcyl-GVPSFAAG-Glu(Edans)	1334,9Da
S9 ou CDSN9	Dabcyl-SEGKYFSS-Glu(Edans)	1532,7Da
NS1	Dabcyl-GSSLQGAS-Glu(Edans)	1334,6Da
NS2	Dabcyl-SSSFQVGN-Glu(Edans)	1452,8Da
NS3	Dabcyl-GSSISSAR-Glu(Edans)	1392,6Da
NS4	Dabcyl-PITSVDKS-Glu(Edans)	1473,8Da

III-B.3.1. Activité d'un extrait brut de protéases du *stratum corneum*

Protocoles de reference :

Préparation d'un extrait de protéase : **V-B.1** p.156

Tests d'activité : **V-B.3** p.157

Analyse par MALDI-TOF/TOF des sites de coupures *in vitro* : **V-B.4** p.158

Un extrait de protéines total a été obtenu à partir d'une poudre acétonique de *strippings* de la jambe. Pour vérifier que l'extrait de protéines présente bien une activité enzymatique, il a été testé avec succès sur des substrats commerciaux :

Le **test Enzcheck** générique réalisé à la fois à pH=5 et pH=8 permet de vérifier les activités globales de l'extrait protéique. Dans le cadre de protéases de la peau, il permet principalement de mettre en évidence les activités cathepsine à pH=5 et protéases à sérine à pH=8.

Le **test S2288** réalisé à pH=8 permet de tester l'activité des protéases à sérine et plus particulièrement l'activité tryptique.

Le **test S2586** réalisé à pH=8 permet de tester spécifiquement l'activité chymotryptique.

a Test enzymatique sur les peptides synthétiques de cornéodesmosine

Une fois vérifiée l'activité enzymatique de l'extrait par ces trois tests, nous avons envisagé son activité sur les peptides de CDSN synthétiques. Les peptides synthétisés spécifiquement (**peptides S**) et aspécifiquement (**peptides NS**) se comportent différemment. En effet, pour ces derniers quasiment aucune variation de fluorescence n'est observée quand le peptide est mis en présence de l'extrait ou non et ce quelque soit le pH auquel est réalisé le test.

En revanche, mis à part pour le peptide S4, pour chaque peptide S, une variation spécifique de la fluorescence est mise en évidence seulement en présence de l'extrait (Figure 64) et particulièrement à l'un des pH d'essai. Cette variation est notée au bout **d'une heure seulement** d'incubation à 37°C, ce qui est relativement surprenant dans la mesure où pour les tests commerciaux, une nuit d'incubation est nécessaire pour observer une variation d'absorbance ou de fluorescence. Cette différence de comportement entre les peptides S et les peptides NS permet donc de conclure **à une dégradation enzymatique des « peptides S » 10 à 50 fois plus importante que pour les « peptides NS »**. Ceci vient donc confirmer nos résultats d'identification de peptides endogènes par LC-MS/MS. Les peptides identifiés en LC-MS/MS correspondent bien à une dégradation spécifique dans des zones privilégiées des protéines par des protéases endogènes et non des exoprotéases. D'autre part, chaque peptide S est

principalement dégradé soit à pH=5 soit à pH=8, ce qui vient là aussi, confirmer que les dégradations observées sont des protéolyses particulières due à des protéases différentes. Pour certains peptides, l'activité mise en évidence correspond bien aux clivages mis en évidence en LC-MS/MS. Ainsi nous avions identifié 2 sites tryptique (R/G) et chymotryptique (Y/R) sur le peptide CDSN3, ce qui est bien en accord avec la dégradation observée à pH basique. Cependant, il faut noter que pour d'autres peptides, l'activité principale n'est pas en accord avec nos suppositions : par exemple pour le peptide CDSN5, l'activité mise en évidence est une activité à pH acide alors que le seul site identifié sur ce peptide est un site Y/E *a priori* chymotryptique. Or, une recherche dans MEROPS montre que ce type de clivage peut effectivement correspondre à une activité chymotryptique mais également à une activité cathepsine (D ou L), ce qui serait donc le cas ici.

Il faut enfin remarquer que pour certains peptides S, la dégradation la plus faible observée est tout de même un peu plus importante que la dégradation des peptides NS ce qui impliquerait que plusieurs protéases seraient mises en jeu. Or, ceci peut être mis en corrélation avec nos données LC-MS/MS. Ainsi pour le peptide CDSN6, nous avons identifié 2 sites de clivages : un site Y/S de type chymotryptique pouvant donc correspondre à l'activité à pH=8 et un site T/Y que nous avons supposé correspondre à une activité de type cathepsine L, ce qui serait plutôt en accord avec l'activité à pH=5.

En ce qui concerne, le peptide CDSN4 ne présentant aucune dégradation *in vitro,* il est possible que la protéase (ou les protéases) impliquée(s) soit en présence de son inhibiteur dans l'extrait total et ne puisse donc pas dégrader le peptide. En outre, il ne faut pas oublier que ces peptides ne sont que des modèles d'étude. Il présente une structure 3D et une accessiblité différentes par rapport à la protéine entière ce qui peut expliquer que ces peptides ne soient pas dégradés *in vitro* de la même manière que la protéine *in vivo*.

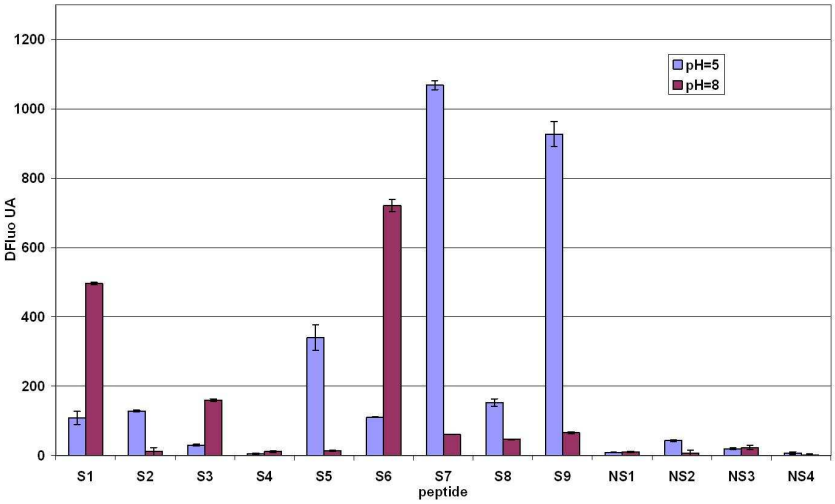


Figure 64 : Activité enzymatique d'un extrait total de protéines de la peau sur les peptides de cornéodesmosine de synthèse.
Les mesures correspondent à la variation de fluorescence des peptides en présence ou non de l'extrait, mesurée après une heure d'incubation à 37°C.

b Vérification des sites de coupure in vitro par spectrométrie de masse

Afin de vérifier que les activités mises en évidence généraient les mêmes sites de coupures *in vitro* que ceux mis en évidence *in vivo* par LC-MS/MS, chaque peptide mis en présence ou non de l'extrait de protéases a été analysé en MS par MALDI-TOF/TOF. Les profils de masses observés montrent que quand il n'y a pas d'extrait protéique ajouté, les formes majoritaires de peptides correspondent soit au peptide entier soit au peptide dont le groupe Dabcyl ou le groupe Edans a été dégradé. En présence de l'extrait de protéases, une différence de profil en fonction du pH auquel a été réalisé le test est observée. Ainsi pour le peptide CDSN1, les masses les plus intenses à pH acide restent des formes correspondant au peptide entier sans clivage interne alors qu'à pH basique, ces formes disparaissent au profit de peptides dans les basses masses qui correspondent au peptide clivé par une ou des protéases (Figure 65).

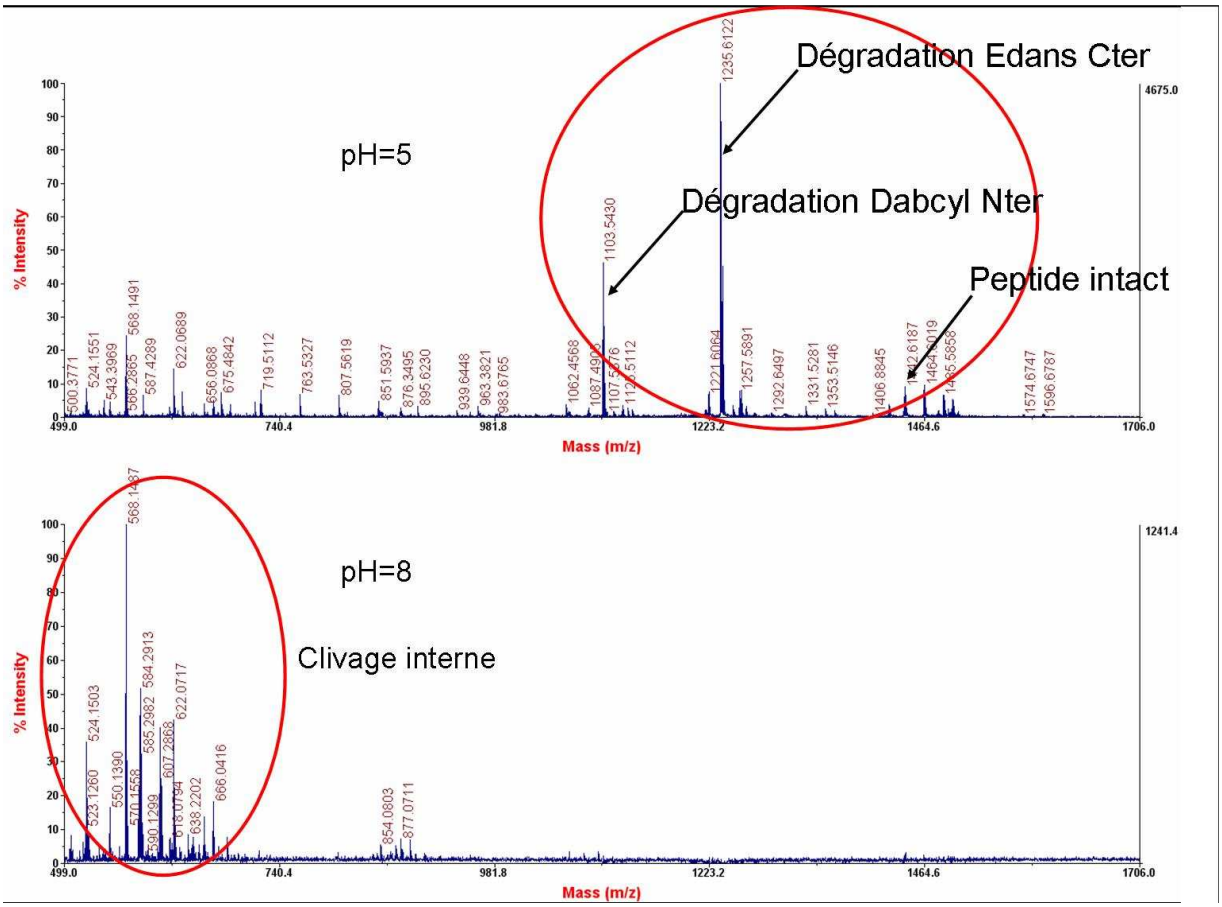


Figure 65 : Analyse MS du mélange CDSN1+extrait protéique à pH=5 et pH=8 en MALDI-TOF/TOF.

Pour identifier la séquence des différentes espèces peptidiques observées, nous avons réalisé une analyse en MS/MS de ces pics. Ainsi, les formes sans clivage interne, mais de masse plus faible que le peptide synthétisé, correspondent en fait à des formes où les groupes Dabcyl ou Edans sont dégradés. De plus, la fragmentation des espèces de plus basse masse nous a permis de mettre en évidence quels étaient les clivages enzymatiques mis en jeu et ainsi de vérifier qu'ils étaient en accord avec ceux identifiés par LC-MS/MS.

Pour le peptide CDSN1, la fragmentation du peptide de masse 584,29Da a permis de mettre en évidence un clivage interne entre l'acide aminé S et Y comme ce que nous avions identifié en LC-MS/MS (Figure 66). En revanche le deuxième clivage entre Y et S n'a pas pu être confirmé.

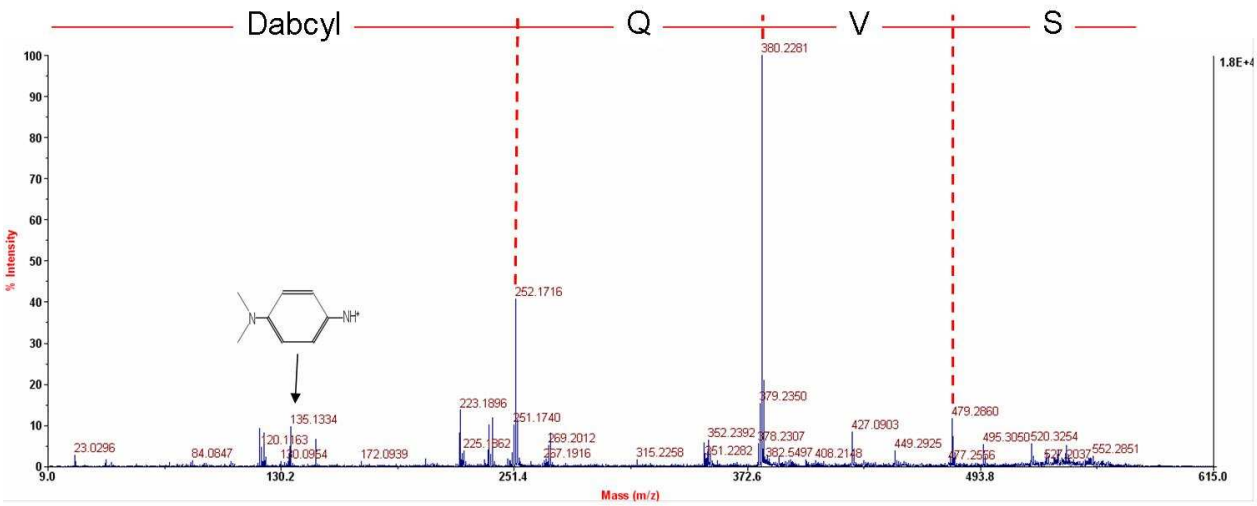


Figure 66 : Spectre de fragmentation de l'espèce 584,29Da qui apparaît dans les spectres MS du peptide CDSN1 lors de sa dégradation par un extrait de protéines du *stratum corneum* à pH basique.

III-B.3.2. Activité d'un extrait fractionné par chromatographie

Protocoles de reference :

Préparation d'un extrait de protéase : **V-B.1** p.156

Fractionnement de l'extrait de protéase : **V-B.2** p.156

Tests d'activité : **V-B.3** p.157

Afin de séparer les protéases mises en jeu et/ou de les séparer de leurs inhibiteurs, il est possible de réaliser un fractionnement de l'extrait de protéines. Nous avons réalisé ce fractionnement, d'une part en fonction de leur masse sur une colonne d'exclusion stérique et d'autre par en fonction de leur charge (et donc de leur pI) sur une colonne d'échange de cations. Au cours de la séparation, différentes fractions de protéines ont été collectées et leur activité enzymatique a été testée sur les peptides modèles suivant le même protocole que l'extrait total. En parallèle, l'activité des fractions sur les substrats commerciaux a également été contrôlée.

Pour la plupart des peptides synthétisés, l'activité des fractions SCX et/ou SEC permet de vérifier une activité enzymatique préférentielle à pH acide ou basique et éventuellement de cibler la ou les protéases impliquées.

Globalement, les activités mises en évidence correspondent aux mêmes fractions que les fractions réactives sur les substrats commerciaux (Figure 67 et Figure 68).

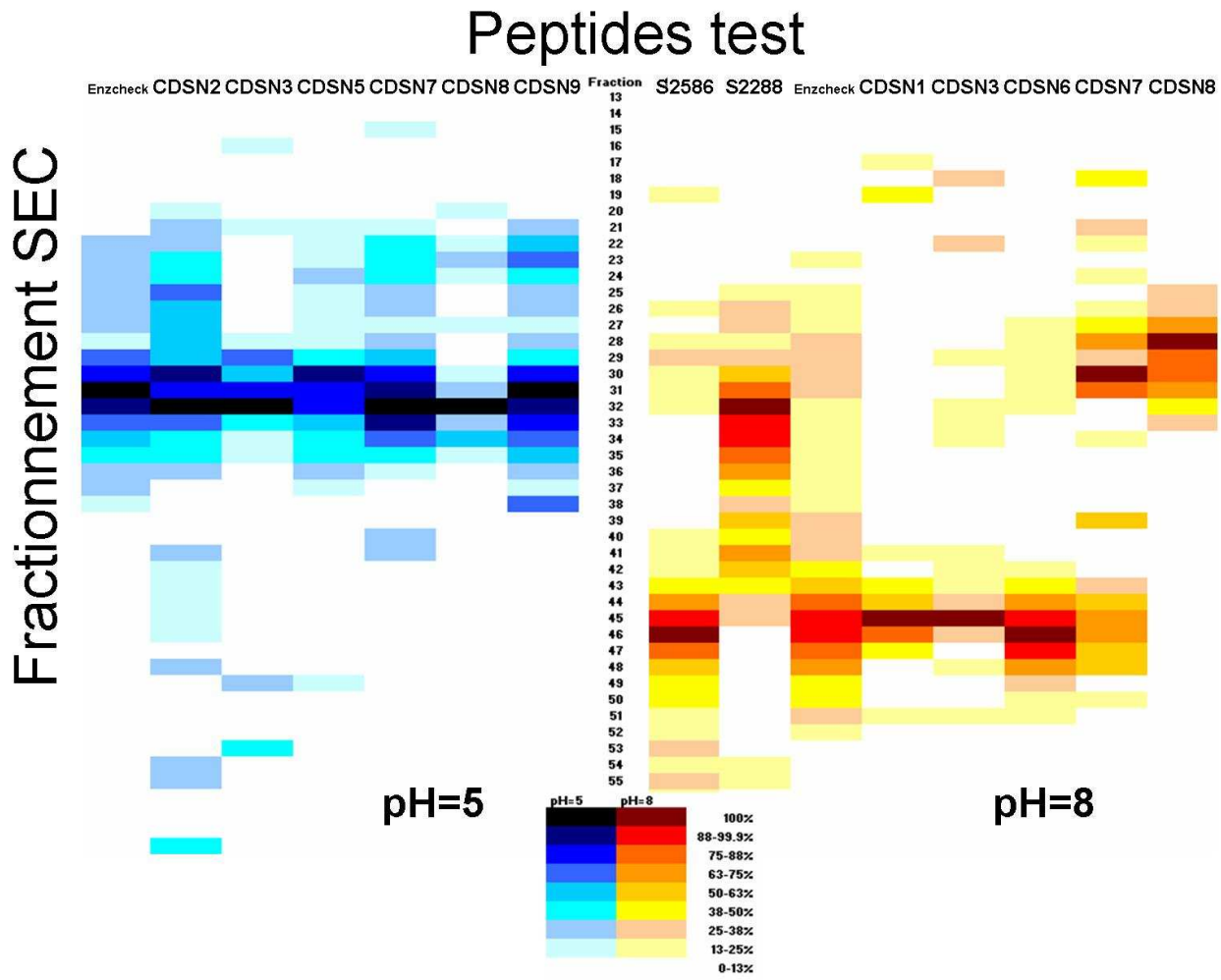


Figure 67 : Profil d'activité enzymatique des fractions obtenues après séparation par SEC d'un extrait total de protéines de *stratum corneum.*
Seuls les peptides dégradés significativement sont présentés. Chaque série de données est normalisée.

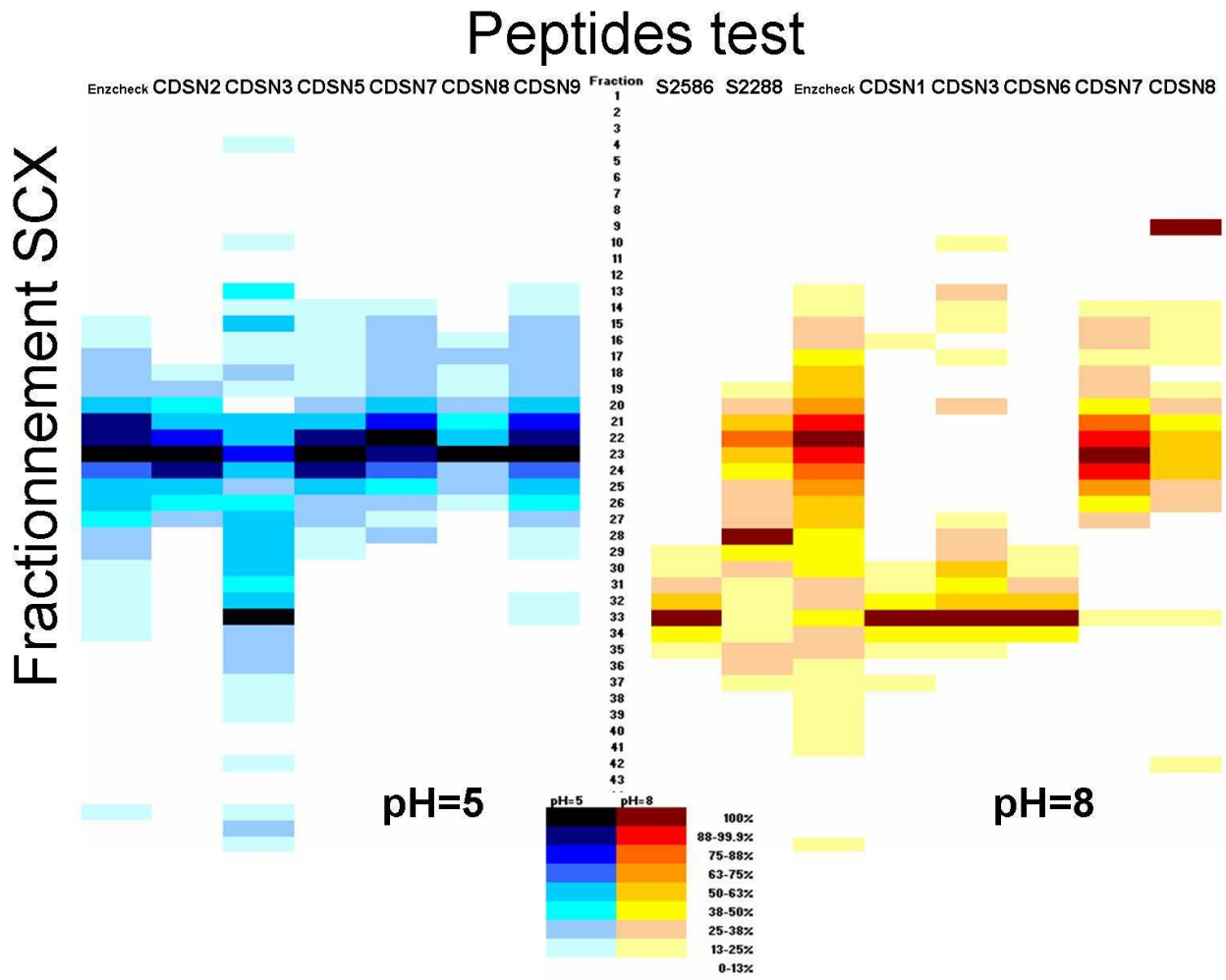


Figure 68 : Profil d'activité enzymatique des fractions obtenues après séparation par SCX d'un extrait total de protéines de *stratum corneum.*
Seuls les peptides dégradés significativement sont présentés. Chaque série de données est normalisée.

a Exemple d'un peptide dégradé en condition basique

Comme nous l'avons vu, le peptide CDSN1 est dégradé par l'extrait brut à pH basique. En comparant l'activité des fractions SCX ou SEC à pH basique sur le peptide CDSN1 et sur les peptides commerciaux, il apparaît une grande similitude avec l'activité mise en évidence par le test S2586. En particulier pour les fractions SCX, les deux profils sont assez superposables (Figure 69B). La protéase capable de dégrader le peptide CDSN1 serait donc la SCCE ce qui est assez original puisque le site mis en évidence par MALDI-TOF/TOF est le site S/Y, *a priori* non chymotryptique. L'ajout d'aprotinine (inhibiteur spécifique des protéases à sérine) sur les fractions SEC 43-48 actives a permis d'inhiber leurs activités et de confirmer cette activité chymotryptique.

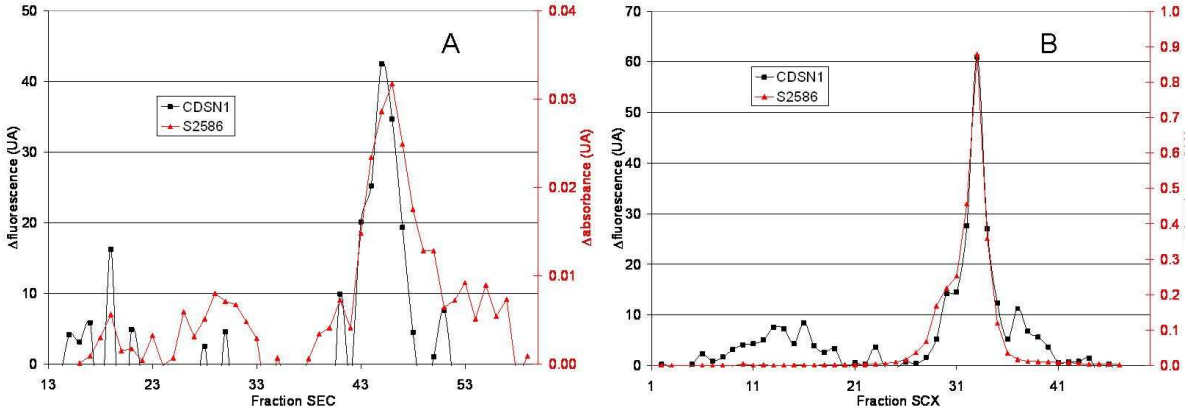


Figure 69 : Profil d'activité enzymatique des fractions obtenues après séparation d'un extrait de protéines de peau sur le peptide CDSN1 en condition basique.
A : Fractions SEC ; B : Fractions SCX. Dans les deux cas, l'activité mise en évidence présente une forte similitude avec l'activité sur le substrat spécifique de l'activité chymotryptique (S2586).

b Exemple d'un peptide dégradé en condition acide

Le peptide CDSN7 présente une dégradation surprenante à pH acide par l'extrait brut alors que nous avions dans un premier temps supposé que les sites de clivages (Y/F et F/T) mis en évidence par LC-MS/MS correspondaient à une activité chymotryptique. L'activité des fractions SEC et SCX à pH acide montre qu'il s'agirait plutôt d'une activité de type cathepsine (Figure 70), ce qui est confirmé par le fait que ces activités sont inhibées par l'ajout de pepstatine (inhibiteur des protéases à acide aspartique). Cependant, le test Enzcheck sur les fractions SCX reste assez aspécifique et ne permet pas de distinguer les différentes isoformes. Les sites mis en évidence par LC-MS/MS pourraient donc correspondre à une activité de type cathepsine D originale. Cependant, si on fait une recherche dans la banque de données MEROPS, ces deux sites de clivage correspondraient plutôt à un clivage par la cathepsine E. Or ceci est aussi très original car cette protéase n'est pas supposée intervenir dans le phénomène de desquamation même si elle est présente dans l'épiderme. En effet, Igarashi *et al.* ont montré par immunofluorescence que la cathepsine D était présente au niveau des desmosomes contrairement à la cathepsine E : ils en ont déduit que seule la cathepsine D était impliquée dans la dégradation des desmosomes pendant le phénomène de desquamation [188]. Nos données ne permettent pas de trancher la question. D'autres expériences pourraient être menées avec un inhibiteur spécifique de la cathepsine E (API, inhibiteur de pepsine sécrété par l'Ascaris) par exemple ou en étudiant la dégradation du peptide CDSN7 par les deux protéases recombinantes.

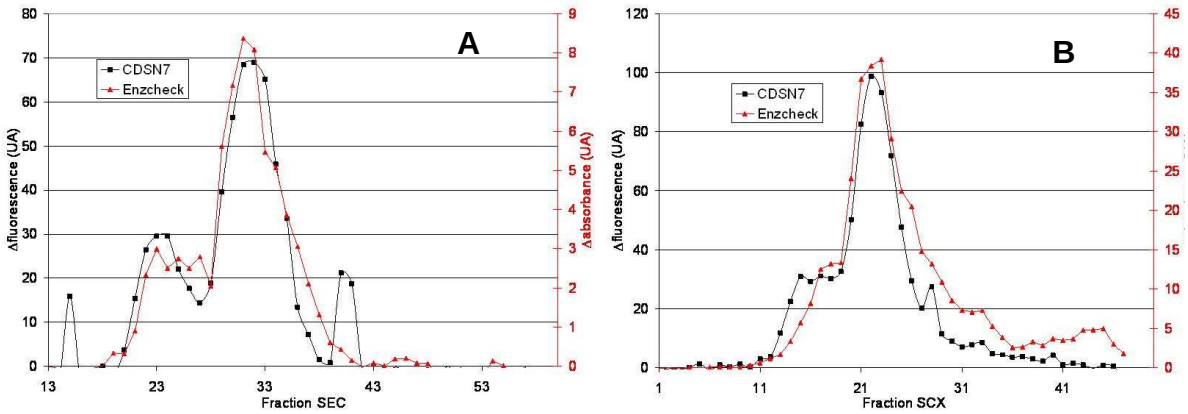


Figure 70 : Profil d'activité enzymatique des fractions obtenues après séparation d'un extrait de protéines de peau sur le peptide CDSN7 en condition acide.
A : Fractions SEC ; B : Fractions SCX. Dans les deux cas, l'activité mise en évidence est une activité de type cathepsine.

c Exemple d'un peptide dégradé de manière originale

Le peptide CDSN3 présente un profil de dégradation assez particulier. Les peptides identifiés en LC-MS/MS sont *a priori* tryptique (R/G) et chymotryptique (Y/R). L'extrait brut permet de dégrader préférentiellement le peptide à pH basique, ce qui serait donc en accord avec les sites observés. Dans l'étude par MALDI-TOF/TOF des produits de dégradation de CDSN3, 3 sites de clivages ont été identifiés : 2 correspondent aux sites observés en LC-MS/MS et un site supplémentaire correspond à un clivage G/I. Ce site peut correspondre à la calpaïne-1 (tout comme le site Y/R). Les fractions SEC et SCX ne présentent pas des profils qui permettent de décider clairement quelles sont les protéases impliquées (Figure 71). En effet, les fractions SEC présentent une assez faible fluorescence qui pourrait tout de même correspondre à une activité de type SCCE (Figure 71A). Les fractions SCX, quant à elles, présentent un profil atypique (Figure 71B). En effet, plusieurs activités semblent être superposées : on observe un pic à la fraction 33 pouvant correspondre à la SCCE mais l'épaulement en fraction 30 est plus important que pour le test S2586. D'autre part, on observe également une activité au niveau de la fraction 22.

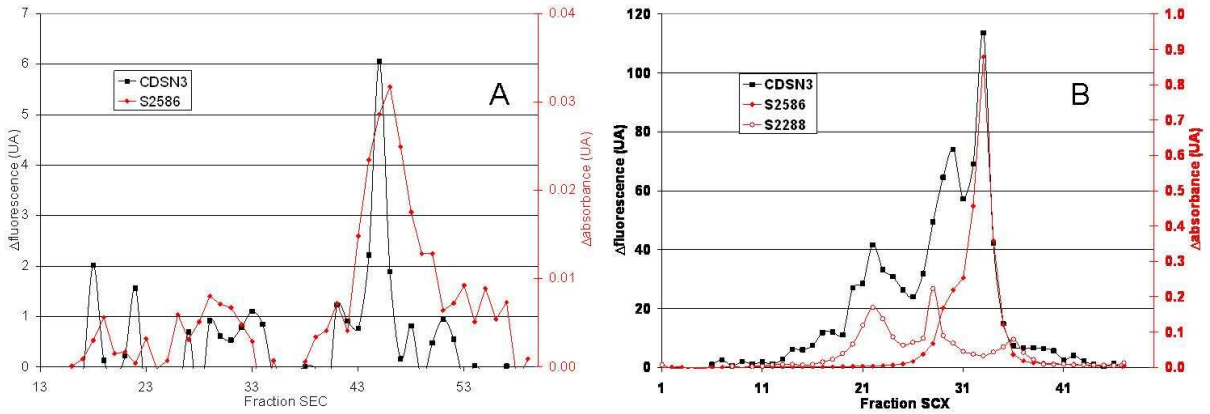


Figure 71 : Profil d'activité enzymatique des fractions obtenues après séparation d'un extrait de protéines de peau sur le peptide CDSN3 en condition basique.
A : Fractions SEC ; B : Fractions SCX. En SEC, aucune activité importante n'est observée. En SCX, plusieurs activités semblent se superposer.

Pour caractériser ces activités, nous avons testé l'extrait brut et les fractions avec ou sans cystéine, pour tester l'activité des protéases à cystéine. On constate alors que la dégradation du peptide CDSN3 par l'extrait brut est activée à pH basique mais aussi à pH acide par la cystéine ce qui met en évidence l'activité d'une protéase à cystéine (Figure 72) qui pourrait donc être la calpaïne-1. Ce peptide est le seul, parmi les peptides synthétisés, dont la dégradation soit activée par la cystéine.

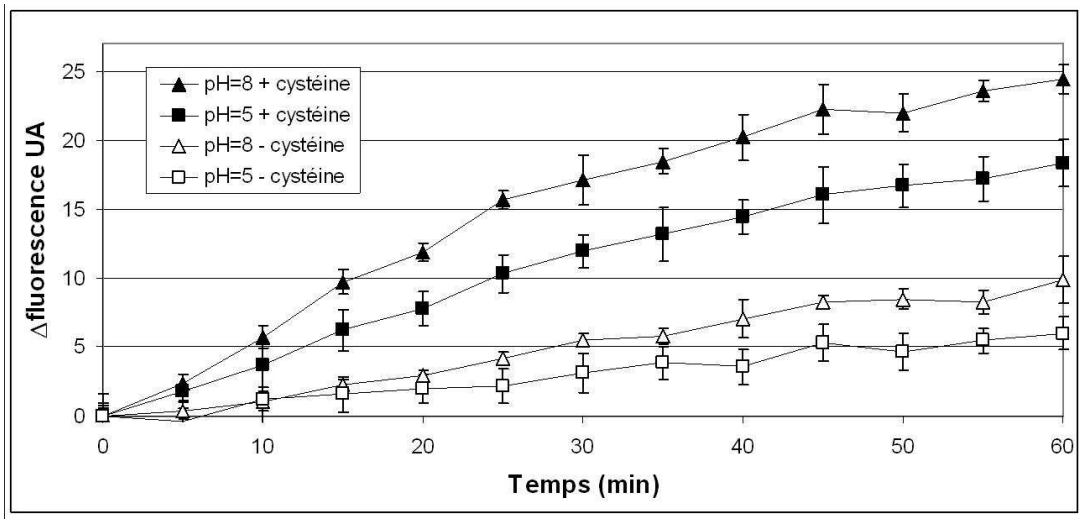


Figure 72 : Influence de la cystéine sur la dégradation du peptide CDSN3 par un extrait brut de protéines du SC. Quelque soit le pH, une dégradation accélérée est observée en présence de cystéine.

En ce qui concerne les fractions SCX, on montre que certaines fractions ne réagissent pas à l'ajout de cystéine (fractions 30 et 33) alors que d'autres sont cystéine-dépendantes (fractions 28, 32, 35) (Figure 73). Les fractions 30 et 33 correspondent aux fractions actives du test S2586 et confirmeraient donc l'activité SCCE alors que les autres fractions mettraient en évidence la (ou les) protéase(s) à cystéine.

Pour mieux caractériser ces activités, il serait nécessaire de réitérer le test avec et sans cystéine sur les fractions SEC.

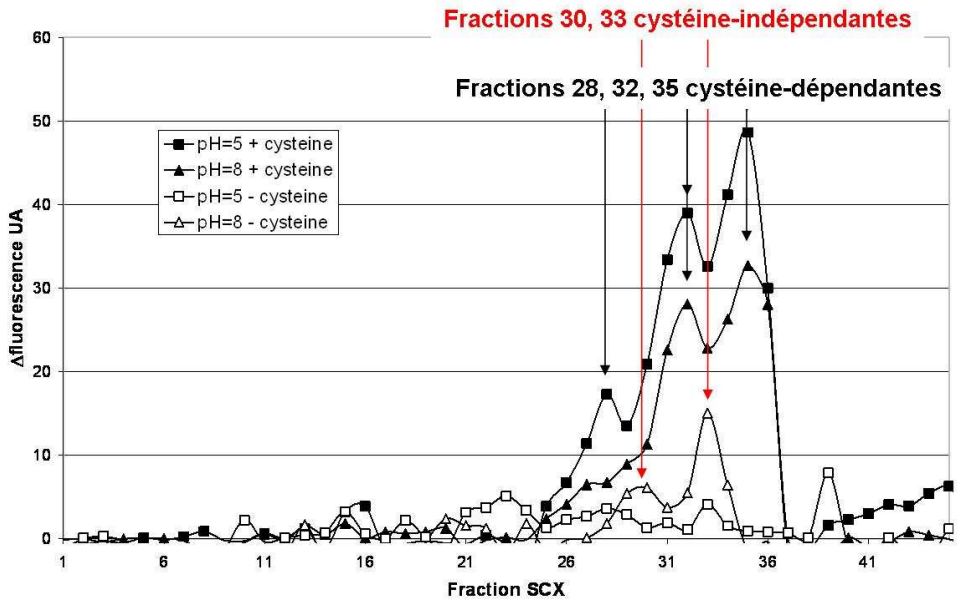


Figure 73 : Influence de la cystéine sur la dégradation du peptide CDSN3 par les fractions SCX. On constate un comportement différent selon les fractions : certaines sont cystéine-dépendantes, d'autres non.

Ces différents tests enzymatiques sur les peptides synthétiques de cornéodesmosine nous ont permis de confirmer que cette protéine est une protéine « hyper hydrolysable ». Plusieurs protéases semblent capables de la dégrader. Certaines activités, comme l'activité chymotryptique, ont été confirmées et d'autres hypothèses ont pu être formulées concernant des sites de clivage atypiques comme l'activité de la calpaïne-1 ou de la cathepsine E. De manière générale, ce type de stratégie peut être appliquée à d'autres protéines en synthétisant des peptides à partir des données LC-MS/MS. Ces peptides sont alors des outils de recherche permettant la découverte de nouvelles protéases ou d'envisager de nouvelles activités de protéases connues de la peau.

PARTIE IV- CONCLUSION GENERALE

Au cours de cette thèse, nous avons développé des méthodologies innovantes tant d'un point de vue des séparations chromatographiques que des analyses en spectrométrie de masse.

Afin de gagner en **sensibilité**, nous avons dans un premier temps optimisé notre montage RPLC-MS afin d'éliminer des contaminants de polymères qui sont élués dans la même gamme de temps que les peptides d'intérêt et qui masquent donc leur signal. Pour cela nous avons ajouté une cartouche de purification sur le circuit de chargement de l'échantillon afin de retenir ces contaminants. D'autre part, afin d'obtenir des cartographies peptidiques plus exhaustives mais aussi afin d'augmenter la **gamme dynamique** de nos analyses, nous avons mis en place et optimisé la chromatographie bidimensionnelle. En effet, si trop de peptides sont co-élués de la colonne analytique pour être ionisés, tous ne peuvent être sélectionnés pour être fragmentés en mode MS/MS, d'où l'absence de détection d'un certain nombre de protéines dans le mélange. Le recours à la nano-chromatographie bidimensionnelle permet de remédier à cette limitation. Le couplage entre la chromatographie d'échange de cations forts et la chromatographie à polarité de phase inversée a été privilégié. Il est possible de réaliser ce couplage soit en mode *on-line* soit en mode *off-line*. Nous avons optimisé nos protocoles dans les deux configurations afin d'améliorer la répartition des peptides dans les différentes fractions salines. Ces méthodologies ont été comparées de manière systématique entre elles et par rapport à une séparation RPLC classique sur un même échantillon digéré de *stratum corneum*. La **capacité de pic** de chaque configuration a été évaluée comme premier indicateur de l'efficacité des séparations. Nous avons également envisagé les divers paramètres pouvant expliquer les différences observées en termes de nombre total de protéines et de peptides identifiés : pouvoir de résolution, orthogonalité, perte d'échantillon, caractéristiques physico-chimiques des peptides identifiés par chaque méthode. Ainsi, nous avons pu montrer que quel que soit le mode de couplage choisi, la capacité de pic globale était multipliée par environ 10, ce qui se traduit en termes d'identifications protéiques et peptidiques par un gain d'un facteur 3. Cependant, nous avons également montré que ces deux modes de couplage étaient complémentaires. En effet, seul un tiers des peptides identifiés est mis en évidence par les deux méthodes. Ceci est principalement dû au fait que chaque méthodologie présente des biais différents : ainsi, en condition *on-line*, les peptides les plus hydrophiles peuvent être perdus alors qu'en condition *off-line*, ce sont plutôt les peptides hydrophobes qui sont retenus par interactions non spécifiques sur les puits de collecte. D'autres configurations de couplages multidimensionnels peuvent encore être envisagées et comparées. Par exemple, le couplage RP x RP présenterait l'avantage d'utiliser des séparations robustes dont les solvants seraient directement compatibles puisque seul le pH change (la première dimension est réalisée en conditions basiques alors

que la deuxième l'est en conditions acides). Reste à savoir, si l'orthogonalité et la capacité de pic obtenues seraient aussi performantes que celles obtenues avec l'association SCX x RP.

D'autre part, nous avons également comparé ces approches avec d'autres, plus simples à mettre en place. Si l'utilisation de liste d'exclusion lors de l'analyse MS ne permet pas de gagner en identification peptidique et protéique, en revanche, l'utilisation de gradient de pente plus faible en RPLC s'est montrée très performante et permet d'obtenir le même type de résultats qu'avec le **couplage multidimensionnel** sans toutefois augmenter la capacité de pic du système. Cette approche semble très prometteuse. En effet, elle fait appel aux séparations classiques par RPLC qui sont des séparations robustes et maîtrisées dans tous les laboratoires d'analyses protéomiques. Des optimisations peuvent être établies dans cette voie : il est probable qu'une durée de gradient intermédiaire, inférieure à 12h permettra d'obtenir des résultats sans doute identiques en limitant les pertes en signal sur bruit dues à l'élargissement des pics. L'utilisation de colonnes plus longues, de même phase stationnaire que les colonnes utilisées au cours de nos expériences, pourra compenser la perte en résolution due à l'allongement de la séparation. Ceci est maintenant rendu possible grâce à la commercialisation par Dionex de colonnes de 25cm.

Enfin, même si le couplage multidimensionnel *off-line* présente des résultats légèrement inférieurs au couplage *on-line,* il permet d'utiliser les fractions collectées de différentes manières. Ainsi chaque fraction est divisée en deux pour subir deux analyses : une séparation couplée à un spectromètre de masse équipé d'une source ESI et une autre couplée avec un spectromètre muni d'une source MALDI. Là encore, un quart seulement des peptides identifiés est commun aux deux approches. En effet, de par leurs propriétés physico-chimiques (masse, hydrophobie, acido-basicité, etc), certains peptides sont ionisés préférentiellement par l'un ou l'autre des processus. Pouvoir utiliser des sources d'ions variées en sortie de séparation chromatographique permet donc d'obtenir des résultats d'identification complémentaires et ainsi d'augmenter la couverture globale du protéome étudié. En termes d'efficacité, la meilleure stratégie pourrait être de séparer les peptides au cours d'une seule et unique analyse RPLC – gradient lent. Grâce à un *split* en sortie de colonne, une partie de l'éluat serait collectée sur une plaque MALDI tandis que l'autre partie serait ionisée en ligne en ESI. Ce montage permettrait d'économiser à la fois de l'échantillon et du temps.

Globalement, tous nos résultats montrent qu'il semble illusoire de tenter d'obtenir une image complète du protéome avec une seule méthode analytique. Chaque technique présente des biais qui vont rendre l'identification de certains peptides préférentielle. Une étude statistique est en cours afin de montrer quels sont les paramètres physico-chimiques des peptides qui permettraient de prévoir la

technique analytique à privilégier pour les identifier. En effet, pour certaines applications (recherche de peptides modifiés spécifiques par exemple), il peut ne pas être judicieux d'avoir recours à l'ensemble des techniques présentées ici. En revanche, dans le cas d'une étude peptidomique, il faut se rappeler que le but n'est pas d'identifier des protéines mais des peptides. Les identifications doivent donc être le plus exhaustives possible : chaque peptide peut apporter une information utile. Dans ce cas, la multiplication des analyses avec des techniques de séparation différentes et des sources d'ionisation variées est loin d'être inutile.

Du point de vue de l'analyse par spectrométrie de masse, nous avons pu montrer que le recours à des spectromètres de masse hybrides de type LIT-FTMS permettait à la fois de gagner en sensibilité, en gamme dynamique et en précision. Or, dans le cadre de l'étude de peptides endogènes de la peau, tous ces paramètres ont leur importance. En effet, sans digestion *in vitro,* le gain en précision de mesure permet de réduire le taux de peptides faux positifs identifiés et ainsi de rendre les identifications plus sures. Les mélanges étudiés sont des mélanges complexes où les peptides d'intérêt sont minoritaires par rapport aux peptides issus de la dégradation sans doute aspécifique des kératines qui représentent les contaminants constitutifs de ces échantillons : d'où un besoin important en gamme dynamique et en sensibilité.

En ce qui concerne l'**application biologique**, nos travaux avaient pour objectifs d'étudier le phénomène de desquamation du *stratum corneum* chez l'Homme, en collaboration avec la société L'Oréal. Pour cela, nous avons développé une stratégie d'analyse spécifique basée sur le fractionnement, l'enrichissement sur phase solide et l'analyse par LC-MS/MS des peptides endogènes de ce tissu.

Dans la mesure où notre approche était sans *a priori,* notre démarche est plus globale et s'oppose aux démarches plus tâtonnantes utilisées par le passé dans l'étude des protéases du *stratum corneum* et de leurs substrats. En effet, dans de nombreux laboratoires, l'approche classique consiste à étudier *in vitro* chaque protéase une à une par purification et à chercher à savoir si cette protéase est capable de dégrader un substrat en particulier. Grâce à notre stratégie plus générale basée sur l'extraction de peptides endogènes générés *in vivo*, nous avons pu accumuler rapidement un très grand nombre de données venant par exemple conforter ou éclairer sous un nouvel angle des résultats déjà acquis mais au cours de dizaines d'années d'expériences. Des résultats originaux ont également pu être obtenus en annexe de la thématique de desquamation. En effet, nous avons également identifié des peptides endogènes qui peuvent être importants dans d'autres processus. Par exemple, nous avons identifié de nombreux peptides issus de protéines à activité antimicrobienne. Ces peptides pourraient être la trace de cette activité ou au contraire de sa régulation. Ainsi, des peptides issus de la séquence de la dermcidine ou de

protéines de la famille S100 ont été identifiés et correspondent à des peptides qui n'ont pas encore été décrits dans la littérature comme AMP. Suite à nos travaux, le laboratoire Peau et Aspects de Surface de L'Oréal a fait synthétiser certains de ces nouveaux peptides afin de réaliser des tests simples de culture bactérienne en leur présence. Certains d'entre eux se sont révélés plus antimicrobiens que les peptides de dermcidine déjà décrits dans la littérature.

L'analyse par RPLC-FTMS d'échantillons de *stratum corneum* a permis l'identification d'un millier de peptides endogènes. Parmi eux, nous avons identifié de très nombreux peptides issus de la dégradation des protéines des cornéodesmosomes dont la protéolyse est un facteur clef de la desquamation. Ainsi, des fragments de cornéodesmosine, de desmogléine 1, et des desmocollines 1 et 3 ont été observés. Ces dernières font partie de la famille des cadhérines desmosomales, dont les domaines cadhérines permettent l'adhésion entre cellules voisines. Les peptides endogènes identifiés au cours de notre étude appartiennent tous à ces régions extracellulaires d'adhésion. En ce qui concerne la cornéodesmosine, les peptides identifiés font notamment partie du domaine N-terminal responsable des propriétés d'adhésion de la protéine.

De l'étude des peptides endogènes identifiés, nous avons pu émettre des hypothèses concernant les protéases mises en jeu dans les processus de dégradation des protéines des cornéodesmosomes en particulier. En effet, l'observation des sites de clivage identifiés par LC-MS/MS permet d'envisager que les activités principalement mises en évidence sont des activités chymotryptiques et tryptiques, pouvant correspondre à la SCCE et à la SCTE, protéases déjà caractérisées dans la dégradation des cornéodesmosomes. Pour confirmer ces hypothèses et mettre un nom sur la protéase permettant de générer tel ou tel peptide, nous avons travaillé en particulier sur la cornéodesmosine. Pour cela, nous avons fait synthétiser le long de la séquence de cette protéine des octopeptides de type Dabcyl-Edans pour suivre leur dégradation par observation de leur fluorescence. Neuf peptides ont été synthétisés spécifiquement de part et d'autre de sites de coupure observés par LC-MS/MS, quatre autres dans des zones où aucun peptide endogène n'avait été identifié. Nous avons suivi par fluorescence l'activité d'un extrait de protéases de la peau sur ces peptides. Les peptides synthétisés de façon spécifique sont 10 à 50 fois plus hydrolysables que les peptides synthétisés non spécifiquement. Ainsi ces neuf peptides fluorescents peuvent aisément remplacer les substrats commerciaux utilisés jusqu'à maintenant au sein du laboratoire Peau et Aspects de Surface de L'Oréal. Dans le cadre de l'étude de la desquamation, ces peptides permettront de caractériser plus finement les protéases de leurs échantillons et leur capacité de protéolyse puisqu'ils sont plus spécifiques de ce phénomène. Pour affiner notre étude, nous avons séparé l'extrait protéique en différentes fractions par chromatographie. Ainsi, nous avons pu mettre en évidence que les fractions actives correspondent à des protéases impliquées dans la desquamation, telles

que la cathepsine D, voire même à des protéases connues pour protéolyser la cornéodesmosine et les cadhérines desmosomales, telles que les protéases à sérine mais au niveau de sites de coupure originaux.

Une fois chaque peptide relié plus finement à l'activité d'une protéase particulière, il pourrait être intéressant de suivre l'activité de cette protéase en suivant la présence du peptide endogène en question. En effet, la dégradation des protéines des cornéodesmosomes est progressive en fonction des couches de l'épiderme. En particulier, la cornéodesmosine perd ses domaines et ses propriétés d'adhésion progressivement. Suivre les différents peptides endogènes en fonction des couches peut permettre de donner un ordre aux protéases dans les cascades protéolytiques mises en jeu. Une tentative a été menée sur six couches successives du *stratum corneum* analysées en LC-MS/MS : cependant, s'il est clair que la dégradation s'accentue dans les couches supérieures de l'épiderme avec l'identification de plus en plus de peptides endogènes, il ne nous a pas été possible de réellement suivre la présence d'un peptide en particulier en fonction des couches. En effet, contrairement à une étude protéomique, ce type d'étude a besoin d'être très exhaustif. Or, comme nous l'avons vu, aucune technique analytique ne permet d'avoir une image du protéome complète. Cependant, avec les développements analytiques en LC2D ou en LC1D gradient lent que nous avons réalisés, ce type d'étude pourrait être renouvelé avec plus de succès. Il pourrait même être envisagé de suivre la présence des peptides par des expériences de type AMT dans la mesure où les peptides ont déjà été caractérisés par MS/MS. Ainsi la quantification des peptides en fonction des couches pourrait être réalisée précisément par une approche sans marquage isotopique basée sur l'aire des pics chromatographiques.

Notre stratégie peptidomique sans *a priori* nous a permis d'obtenir une base de données conséquente de peptides endogènes du *stratum corneum* provenant de sujets sains de peau « normale ». Sur ces mêmes sujets, la comparaison des peptidomes avec ou sans application de produits cosmétiques permettrait de démontrer le mode d'action de ces produits. Par la suite, cette base de données pourra être utilisée pour identifier l'origine des troubles de la desquamation. L'étude du peptidome de sujets souffrant par exemple de xérose (peau sèche liée au froid, à l'âge…) ou d'ichtyose (peau sèche liée à des facteurs héréditaires) permettrait de mettre en évidence des peptides endogènes spécifiques de ces pathologies et ainsi de mettre en place des traitements adéquats. L'étude du peptidome permettrait de mieux comprendre les désordres de la desquamation impliqués dans des maladies plus graves. Par exemple, les enfants atteints du syndrome de Netherton subissent un phénomène de desquamation exacerbée du fait d'une mutation sur LEKTI, inhibiteur endogène des protéases à sérine. Une activité de type trypsine anormalement élevée a été mise en évidence dans la couche cornée de patients, mais les fonctions de

LEKTI restent encore mal comprises [190]. Il est probable que les peptides endogènes identifiés dans ce type de pathologie soient fortement différents, en termes qualitatifs ou quantitatifs.

A partir de tous les résultats issus des analyses en LC2D d'échantillons de peau et de cil, une autre base de données pourra également être utilisée au sein de la plateforme de protéomique. En effet, ces deux tissus sont des contaminants que les expérimentateurs redoutent fortement. Ainsi, un échantillon contaminé par un cil ou un squame est très difficilement exploitable. De nombreux peptides vont provenir de la digestion des protéines contaminantes et masqueront le signal des peptides des protéines d'intérêt du mélange. La connaissance des protéines provenant spécifiquement de la peau (palmo-plantaire ou non), de la sueur ou d'un cil peut permettre d'identifier la cause de contamination et ainsi de conseiller le collaborateur sur les précautions supplémentaires à prendre. D'autre part, si les kératines sont des contaminants classiques et sont donc éliminées des résultats, cette base de données permettra d'écarter d'autres protéines (comme la desmoplakine, la dermcidine ou l'hornerine…) des résultats d'identifications, et d'éviter ainsi de proposer au collaborateur de mauvaises pistes de recherches ultérieures.

Pour terminer, l'ensemble des techniques mises en place au cours de nos travaux en peptidomique et en protéomique sont désormais disponibles sur la plateforme de protéomique de l'ESPCI et proposées aux collaborateurs de la communauté scientifique.

PARTIE V- PROTOCOLES EXPERIMENTAUX

V-A.1. Préparation des échantillons de *stratum corneum* non digéré

V-A.1.1. Echantillons de peau

Les échantillons de *stratum corneum* plantaire correspondent à de la corne de pieds obtenue par prélèvement par des pédicures sur des sujets sains.

Les extraits de *stratum corneum* non plantaire ont été obtenus sur des sujets volontaires sains suivant le protocole publié dans la littérature [192]. Brièvement, le mollet est recouvert d'une bande de nylon (60cm^2) puis de vernis. Après dix minutes de séchage, la bande est arrachée. Le vernis est dissous avec 60mL d'acétone ce qui libère les cornéocytes. Ceux-ci sont lavés par 180mL d'acétone. Enfin la poudre acétonique est récupérée après filtration et séchage.

V-A.1.2. Extraction des peptides

Différents solvants ont été utilisés sans adjonction de composés dénaturants :

- Tampon phosphate (PBS) 10mM
- Tampon (EDTA) 5mM
- Tampon (Tris) 0,5M pH=6,8
- Tampon (Tris) 0,5M pH=8,2
- Solution aqueuse de chlorure de sodium (NaCl) 2M
- Solution aqueuse d'acide trifluoroacétique (TFA) 0,1% (v/v)
- Méthanol (MeOH) en présence de TFA (MeOH/H_2O/TFA 80/20/0,1, v/v/v)
- Solution aqueuse d'acide formique (AF) 1%
- Méthanol en présence d'AF (MeOH/H_2O/AF 80/20/1, v/v/v)

5 mg de peau (SCP ou PA) sont extraits dans 500µL de solvant/tampon durant une nuit à 4°C, sous agitation douce (800tpm). Après agitation, l'échantillon est soniqué dix minutes dans un bain à ultrasons.

V-A.1.3. Etape d'ultrafiltration

L'échantillon est fractionné par gamme de masse sur des micro-colonnes d'ultrafiltration Vivaspin 500 (Vivascience) selon le protocole suivant :

Equilibre de la micro colonne : Déposer 600µL de PBS 10mM sur les colonnes (seuil 5kDa et seuil 10kDa) et centrifuger 10 minutes à 13 000 tpm à température ambiante.
Jeter le PBS présent dans le tube et le volume restant dans la colonne.
Dépôt de l'échantillon : Déposer le volume d'échantillon sur la micro colonne de seuil 10kDa.
Centrifuger environ 10 minutes à 13 000 tpm à température ambiante.
Conserver à 4°C le volume restant au-dessus de la colonne (fraction sup 10kDa). Déposer l'éluat sur la micro colonne de seuil 5kDa.
Centrifuger la colonne environ 10 minutes à 13 000 tpm à température ambiante.
Conserver à 4°C le volume restant sur la colonne (fraction 5-10kDa) Conserver l'éluat. (fraction inf 5kDa) à 4°C.
A la fin de cette étape, trois types d'échantillons sont récupérés :

- une fraction de protéines et de peptides de masse supérieure à 10kDa (environ 50µL).
- une fraction de peptides de masse comprise entre 5 et 10kDa (environ 50µL).
- une fraction de peptides de masse inférieure à 5kDa (environ 400µL).

V-A.1.4. Etape de reconcentration des peptides

Afin d'éliminer les sels et reconcentrer les échantillons, les différentes fractions de peptides subissent une étape de reconcentration sur des micro-colonnes de silice greffée octadécyl (ZipTip™ C18, Millipore, St Quentin en Yvelines, France). Cette étape est réalisée grâce au robot Genesis Proteam 150 (Tecan) (Figure 74).



Figure 74 : Robot pipeteur/manipulateur Genesis Proteam 150

Lorsqu'il a été utilisé pour l'extraction des peptides, le MeOH est d'abord évaporé au *Speedvac*.

Les échantillons extraits dans des solutions/tampons non acides sont acidifiés par de l'AF afin d'ajuster le pH aux alentours de 2,5.

Pour les fractions 5-10kDa et sup 10kDa, le volume d'échantillon étant aux alentours de 50µL, un protocole classique de reconcentration sur ZipTip™ C18 (Millipore, St Quentin en Yvelines, France) est utilisé. Les ZipTip™ sont d'abord conditionnés par un mélange d'acétonitrile (ACN) et d'eau (ACN/H_2O/AF, 50/50/1, v/v/v) puis équilibrés par de l'AF 1% en solution aqueuse. Les peptides sont fixés sur le ZipTip™ par une série de 10 aspirations/refoulements puis lavés par de l'AF 1% en solution aqueuse. Enfin les peptides sont élués par 10µL d'ACN/$H_2$0/AF (50/50/1, v/v/v) et encore 10µL d'une solution plus concentrée en ACN (80/20/1, v/v/v). L'éluat final est séché au *speedvac* et repris dans 10µL de solvant A (H_2O/ACN/AF, 98/2/0,1, v/v/v).

Pour la fraction de peptides inf 5kDa, le volume d'échantillon est trop important pour réaliser un ZipTip™ « simple ». Après conditionnement et équilibrage du ZipTip™, le robot prélève 100µL d'échantillon, qu'il dépose et pousse au dessus du ZipTip™. Les peptides fixés sont lavés puis élués par 10µL d'ACN/$H_2$0/AF (50/50/1, v/v/v) puis 10µL d'une solution plus concentrée en ACN (80/20/1, v/v/v). Après rééquilibrage du ZipTip™, 100µL d'échantillon sont à nouveau fixés sur le ZipTip™ et ainsi de suite. L'éluat final est ensuite séché et repris dans 10µL de solvant A.

V-A.2. Préparation des échantillons digérés

Pour la mise au point des conditions de spectrométrie de masse sur l'Orbitrap, un mélange commercial de protéines standard digérées par la trypsine a été utilisé (référence 161088, Dionex, Pays-Bas). Ce mélange, noté LCP, est composé de six protéines de poids moléculaires variant de 11 à 135kDa (*cytochrome C, lysozyme C, alcohol dehydrogenase, bovine serum albumin, serotransferrin and β-galactosidase*).

Pour les échantillons humains digérés, le protocole suivant a été utilisé.

Solutions utilisées :
DTT : 10mM dans 50mM de bicarbonate d'ammonium aqueux
Iodoacétamide : 55mM dans 50mM de bicarbonate d'ammonium aqueux
Préparation de la trypsine : solution finale à 50ng/µL de trypsine Roche (Trypsin, modified, Sequencing Grade, 11418025001), 20mM de bicarbonate d'ammonium, 0,04% CaCl2. Cette solution est préparée le jour de la digestion et conservée à 4°C avant usage. La trypsine est d'abord activée selon le protocole du fournisseur : 25µg de trypsine sont dilués dans 250µL de TFA 0,01% et soniqués 10min.

Protocole de digestion :

Echantillons : 16mg de SCP

Une cinquantaine de cils

Incuber dans 300µL de DTT pendant 30min à 56°C. Enlever l'excès de solution et le conserver de côté.

Ajouter 1 volume d'ACN pour baigner le culot et remplacer rapidement par 300µL d'iodoacétamide.

Incuber pendant 20min à température ambiante. Enlever l'excès de solution et le conserver avec le DTT.

Laver avec un volume de NH_4HCO_3 50mM pendant 10 min. sous agitation

Ajouter 1 volume d'ACN 100% puis remplacer par 240µL de trypsine.

Ajouter 240µL de trypsine au mélange DTT/Iodo.

Laisser digérer une nuit à 37°C sous agitation douce.

Acidifier avec de l'acide formique : concentration finale 1%.

Les digestats des solutions de lavages et de l'échantillon sont mélangés et le volume final est réduit à 250µL par évaporation. Pour l'optimisation de la LC2D, 5µL de l'échantillon de cils ont été utilisés ou 0,5µL d'échantillon de SCP.

V-A.3. Séparation des peptides par LC1D

V-A.3.1. Echantillons non digérés

a RPLC- nanoESI-QqTOF

Les analyses ont été réalisées sur un système chromatographique Famos-Switchos-UltiMate (LC Packings, Pays-Bas) couplé au nanoESI-QqTOF (Micromass, Royaume-Uni). Les peptides sont chargés sur une précolonne (C18 PepMap100, 5µm, 100Å, 300µm d.i., 5mm de longueur, Dionex, Pays-Bas), dessalés et concentrés par du solvant A (H_2O/ACN/AF, 98/2/0,1, v/v/v) au débit de 15µL/min pendant 15 min. Ils sont alors élués vers la colonne analytique à polarité de phases inversée (C18 PepMap100, 3µm, 100Å, 75µm d.i., 15cm de longueur, Dionex) avec un débit de 220nL/min. Le gradient utilisé est un gradient linéaire de 0 à 50% de solvant B (ACN/H_2O/AF, 90/10/0,1, v/v/v) en 35min, puis un palier isocratique de 15min à 100% B et enfin 15min de rééquilibrage de la colonne en solvant A.

b RPLC- nanoESI-FTMS

Les analyses ont été réalisées sur un système chromatographique Agilent 1200 (Agilent Technologie, Etats-Unis) couplé au nanoESI-LIT-FTICR ou au nanoESI-LIT-Orbitrap (Thermo Fisher Scientific,

Etats-Unis). Les peptides sont chargés sur une précolonne (C18 Zorbax300SB, 5µm, 300Å, 300µm d.i., 5mm de longueur, Agilent), dessalés et concentrés par du solvant A (H_2O/ACN/AF, 98/2/0,1, v/v/v) au débit de 15µL/min pendant 5 min. Ils sont alors élués vers la colonne analytique à polarité de phases inversée (C18 Zorbax 300SB, 3µm, 300Å, 75µm d.i., 15cm de longueur, Agilent) avec un débit de 300nL/min. Le gradient utilisé est un gradient bilinéaire de 0 à 13% de solvant B (ACN/H_2O/AF, 90/10/0,1, v/v/v) en 3 min puis de 13 à 44% de B en 42 min, un palier isocratique de 15min à 100% B et enfin 15min de rééquilibrage de la colonne en solvant A.

V-A.3.2. Echantillons digérés

a RPLC-nanoESI-FTMS

Les analyses ont été réalisées sur un système chromatographique Ultimate 3000 (Dionex, Pays-Bas) couplé au nanoESI-LIT-FTICR ou au nanoESI-LIT-Orbitrap (Thermo Fisher Scientific, Etats-Unis). Les peptides sont chargés sur une précolonne (C18 PepMap100, 5µm, 100Å, 300µm d.i., 5mm de longueur, Dionex, Pays-Bas), dessalés et concentrés par du solvant A (H_2O/ACN/AF, 98/2/0,1, v/v/v) au débit de 15µL/min pendant 15 min. Ils sont alors élués vers la colonne analytique à polarité de phases inversée (C18 PepMap100, 3µm, 100Å, 75µm d.i., 15cm de longueur, Dionex) avec un débit de 220nL/min. Le gradient utilisé est un gradient linéaire de 0 à 50% de solvant B (ACN/H_2O/AF, 90/10/0.1, v/v/v) en 35min, puis un palier isocratique de 15min à 100% B et enfin 15min de rééquilibrage de la colonne en solvant A.

b RPLC-MALDI-TOF/TOF

Les analyses ont été réalisées sur un système chromatographique Ultimate 3000 (Dionex, Pays-Bas) couplé au MALDI-TOF/TOF 4800 (Applera Applied Biosystems, Framingham, MA). Les peptides sont chargés sur une précolonne (C18 PepMap100, 5µm, 100Å, 300µm d.i., 5mm de longueur, Dionex, Pays-Bas), dessalés et concentrés par du solvant A (H_2O/ACN/AF, 98/2/0,1, v/v/v) au débit de 20µL/min pendant 15 min. Ils sont alors élués vers la colonne analytique à polarité de phases inversée (C18 PepMap100, 3µm, 100Å, 75µm d.i., 15cm de longueur, Dionex) avec un débit de 220nL/min. Le gradient utilisé est un gradient linéaire de 0 à 50% de solvant B (ACN/H_2O/AF, 90/10/0.1, v/v/v) en 35min, puis un palier isocratique de 15min à 100% B et enfin 15min de rééquilibrage de la colonne en solvant A. L'éluat est déposé sur une plaque MALDI vierge en utilisant le robot de dépôt Probot (Dionex) à partir de t=15min (par rapport au début du gradient) pendant 40min. Pour chaque *run,* 240 spots (1 toutes les 10s) sont collectés avec l'ajout coaxial de matrice au débit de 442nL/min (CHCA 5mg/mL, 6:4:0,01 ACN/H2O/TFA v/v/v, 10mM citrate d'ammonium). Dans le cas du dépôt

d'un *run* LC2D complet, l'ensemble des fractions SCX x RP (soit 15*240=3600 spots) sont déposés sur la même plaque.

V-A.4. Séparation des peptides par LC2D

Dans cette partie, sont décrits les protocoles finaux.

V-A.4.1. Mode *on-line*

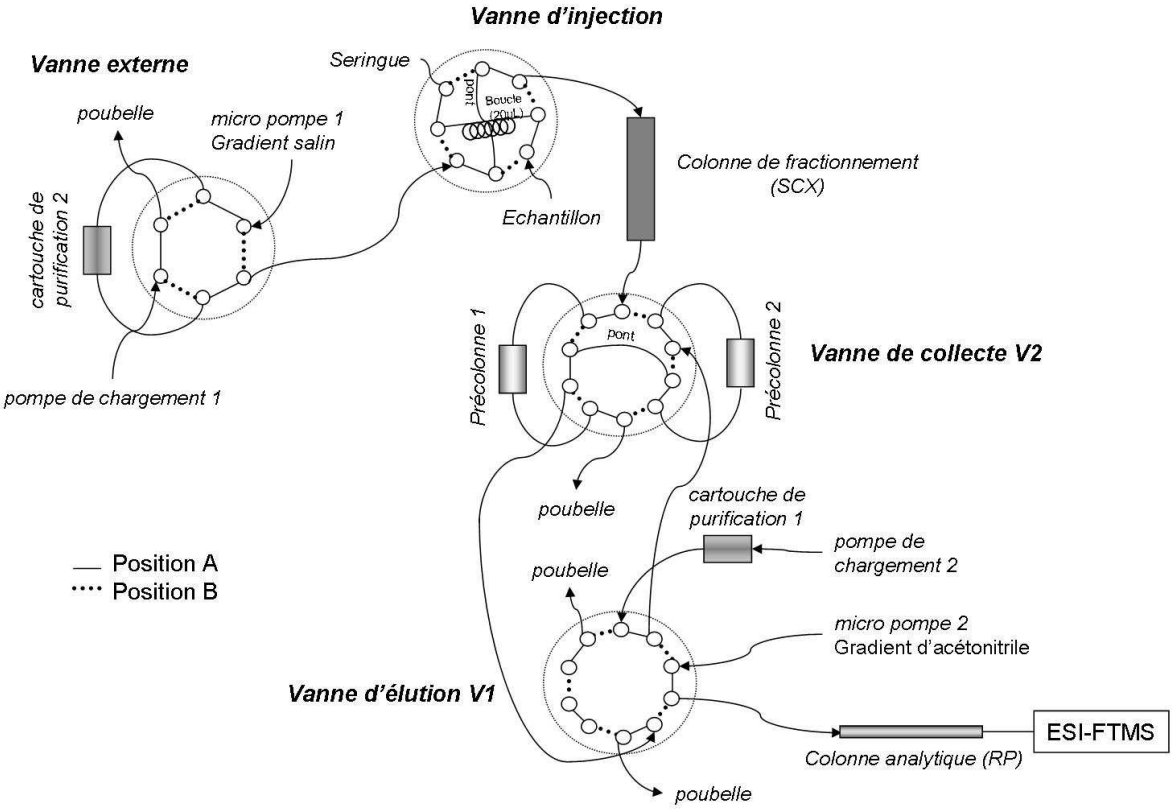


Figure 75 : Schéma du montage LC2D *on-line*.

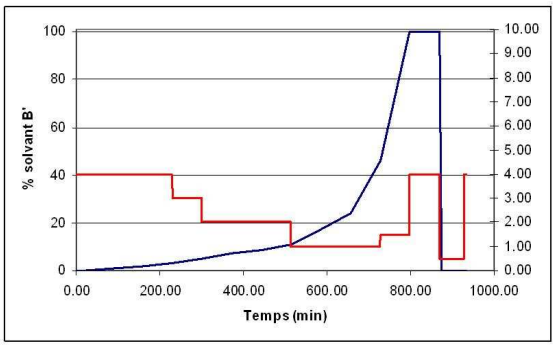
En mode *on-line*, les peptides sont d'abord retenus sur la colonne capillaire SCX (BioBasic SCX 5µm, 300 Å, 320µm d.i. 15cm de long, ThermoFisher Scientific). Les tampons utilisés sont un tampon A' aqueux ($ACN/H_2O/AF$, 5/95/0.1, v/v/v) et un tampon salin B' (100mM formiate d'ammonium, tamponné à pH=2,5 par du TFA, 5% ACN) préparé directement à partir d'acide formique et d'ammoniaque. Pour la séparation des peptides, le gradient est multilinéaire et le débit de la micro pompe 1 qui délivre le gradient salin varie au cours de la séparation.

Les peptides élués sont alternativement retenus sur la précolonne 1 (C18 PepMap100, 5µm, 100Å, 300µm d.i., 5mm de longueur, Dionex) ou sur la précolonne 2 puis dessalés durant 15min par du solvant A ($H_2O/ACN/AF$, 98/2/0,1, v/v/v) (Figure 1). Chaque cycle dure 71min. Après l'étape de piégeage, la

vanne V2 est basculée et les peptides sont élués vers la colonne analytique pour une séparation RP-HPLC (protocole V-A.3.2.). Pendant le dernier gradient RP, la colonne SCX est équilibrée en 100% A' pendant 1h. Pour l'élimination des polymères sur le circuit SCX, une cartouche de purification (C18 PepMap100, 5µm, 100 Å, 1 mm d.i., 15 mm de longueur, Dionex) est insérée après la micro pompe 1 (III-A.3.1.) [202].

Tableau 13 : Gradient optimisé en condition LC2D *on-line*.

Temps	% de tampon B'		Débit
	initial	final	
De 0 à 17min	0	0	4µL/min
De 17 à 88 min	0	1	4µL/min
De 88 à 159 min	1	2	4µL/min
De 159 à 230 min	2	3	4µL/min
De 230 à 301 min	3	5	3µL/min
De 301 à 372 min	5	7,1	2µL/min
De 372 à 443 min	7,1	8,8	2µL/min
De 443 à 514 min	8,8	10,9	2µL/min
De 514 à 585 min	10,9	17,3	1µL/min
De 585 à 656 min	17,3	24	1µL/min
De 656 à 727 min	24	46	1µL/min
De 727 à 798 min	46	100	1,5µL/min
De 798 à 869 min	100	100	4µL/min
De 869 à 929 min	0	0	0,5µL/min
De 929 à 934 min	0	0	4µL/min



V-A.4.2. Mode *off-line*

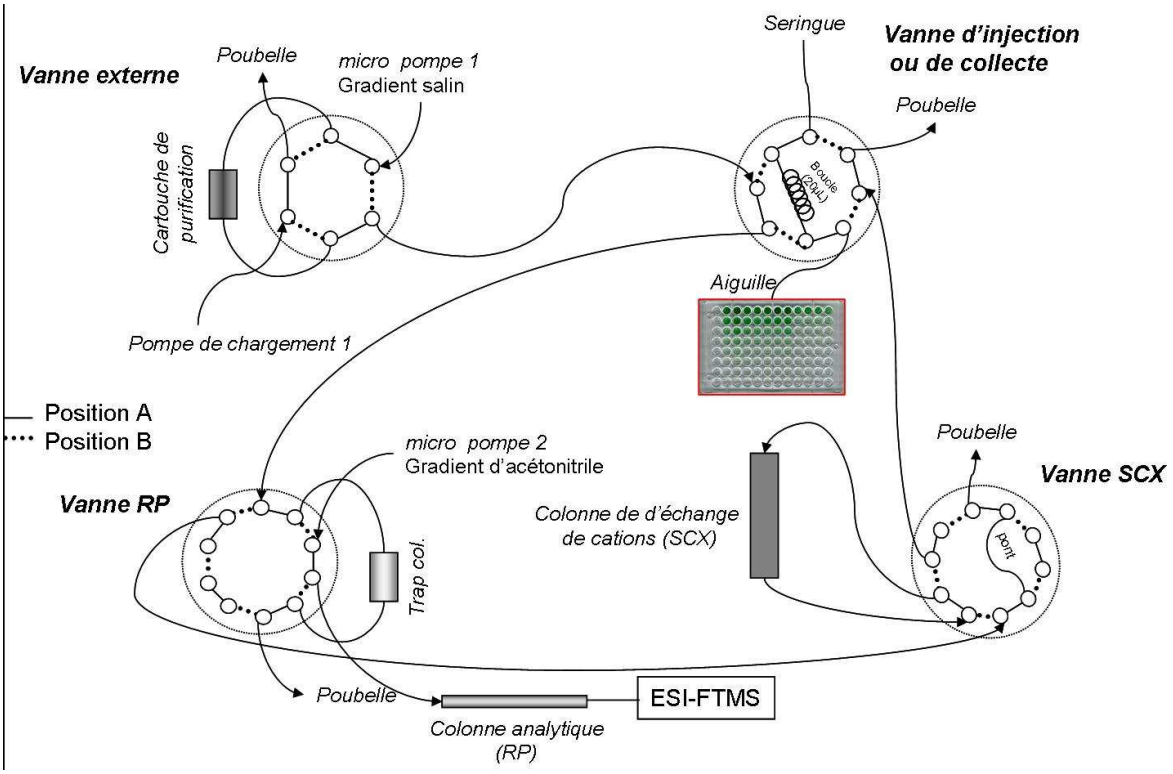


Figure 76 : Schéma du montage LC2D *off-line*.

En mode *off-line*, les peptides sont injectés et retenus sur une colonne capillaire SCX (BioBasic SCX 5µm, 300 Å, 320µm d.i., 15cm de long, ThermoFisher Scientific) en condition légèrement saline (10% solvant B' (100mM formiate d'ammonium, tamponné à pH=2,5 par du TFA, 25% ACN), 90% solvant A' (ACN/H_2O/AF, 5/95/0.1, v/v/v)). Pour la collecte des peptides, le gradient est linéaire de 20% à 70% de solvant B' en 12min à 4µL/min puis un palier à 100% B' de 15min. Les peptides élués sont automatiquement collectés dans une plaque multi-puits au niveau de l'autosampler toutes les 60s (Figure 76). Après évaporation de l'ACN, chaque fraction est reprise dans 10µL de solvant A, transférée dans un tube LC classique et réinjectée successivement pour une séparation RP-HPLC. Pour l'élimination des contaminants sur le circuit SCX, une cartouche de purification (C18 PepMap100, 5µm, 100 Å, 1 mm d.i.,15 mm de longueur, Dionex) est insérée après la micro pompe 1 [202].

V-A.5. Analyse des peptides par spectrométrie de masse

V-A.5.1. Analyse par spectrométrie de masse nanoESI-QqTOF

Les peptides élués sont détectés en ligne sur un spectromètre de masse hybride de type QqTOF (QTof2, Micromass Waters) équipé d'une source nanospray avec un capillaire d'ionisation en verre (New Objective, Etats-Unis). L'instrument est utilisé en mode positif sur la gamme de masse 400-1 300Th en

mode MS et 50-2 600Th en mode MS/MS. Les données sont acquises automatiquement avec alternance du mode MS et du mode MS/MS (Figure 77). A chaque cycle complet, les 3 peptides les plus intenses (deux fois ou trois fois chargés) sont sélectionnés pour être fragmentés dans la cellule de collision.

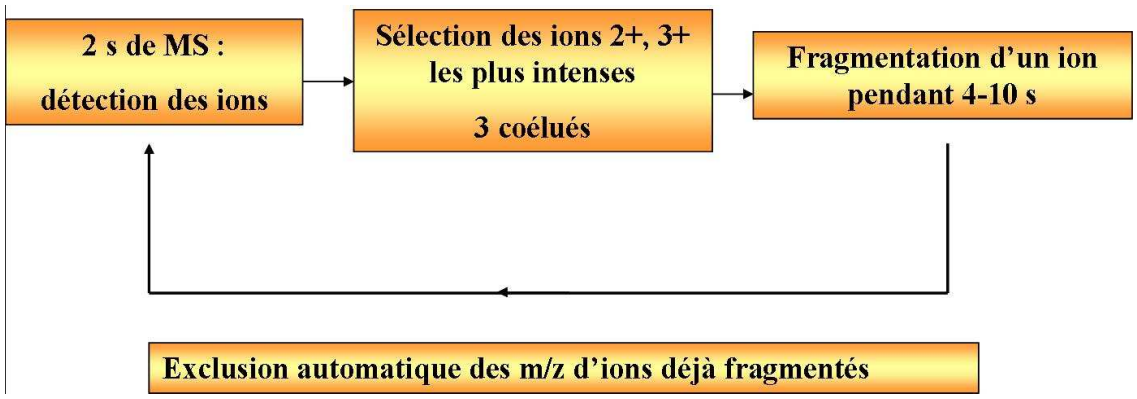


Figure 77 : Cycle d'analyse automatique sur le QqTOF

V-A.5.2. Analyse par spectrométrie de masse nanoESI-LIT-FTICR

Les peptides élués sont détectés en ligne sur un spectromètre de masse hybride de type LIT-FTMS (LTQ-FT, Thermo Fischer) équipé d'une source nanospray avec un capillaire d'ionisation en verre (New Objective). L'instrument est utilisé en mode positif sur la gamme de masse 400-2 000Th (échantillons digérés) ou 500-2 000Th (échantillons non digérés) en mode MS. Les données sont acquises automatiquement en mode *high dynamic* avec alternance d'un *full scan* en mode MS dans la cellule ICR (résolution 25 000), de trois SIM *scan* dans la cellule ICR (+/- 5Th autour du précurseur, résolution 50 000) en parallèle des trois MS/MS réalisées dans le LTQ (Figure 78). A chaque cycle complet, les 3 peptides les plus intenses (au moins deux fois chargés) sont sélectionnés (fenêtre de sélection 4Th) pour être fragmentés dans la trappe avec une énergie de 35%.

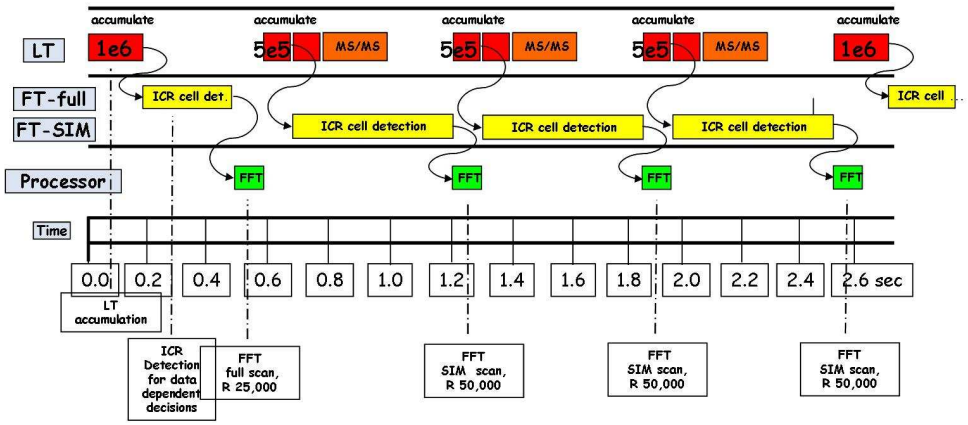


Figure 78 : Cycle d'acquisition automatique utilisé sur le LTQ-FTCIR afin d'augmenter la gamme dynamique d'après [209].

V-A.5.3. Analyse par spectrométrie de masse nanoESI-LIT-Orbitrap

Les peptides élués sont détectés en ligne sur un spectromètre de masse hybride de type LIT-FTMS (LTQ-Orbitrap, Thermo Fischer) équipé d'une source nanospray avec un capillaire d'ionisation en verre (New Objective). L'instrument est utilisé en mode positif sur la gamme de masse 500-2 000Th en mode MS. Les données sont acquises automatiquement avec alternance entre le mode full MS (résolution 30 000) et la fragmentation de n précurseurs dans le LTQ. La fenêtre de sélection dans le LTQ est de 4Th et l'énergie de collision de 35%. Afin d'améliorer la précision de masse, les spectres MS sont calibrés en ligne par étalonnage interne sur un ion PMC (m/z 519.1389; $(Si(CH_3)_2O)_7$) observé tout au long du chromatogramme [200].

V-A.5.4. Analyse par spectrométrie de masse MALDI-TOF/TOF

Les peptides déposés sur une plaque MALDI sont analysés sur un spectromètre de masse MALDI-TOF/TOF 4800 (Applera Applied Biosystems, Framingham, MA). Les analyses MS ont été réalisées en mode réflectron positif (gamme de masse 700-2 000), intensité du laser 3 100, 1 800 tirs cumulés). Les données MS/MS ont été acquises en utilisant de l'air comme gaz de collision et avec une énergie de 2keV (intensité du laser 4 000, 2400 tirs cumulés). La MS/MS est réalisée dans chaque spot sur 7 précurseurs de S/B>40 de telle sorte que le spot choisi corresponde à leur apex chromatographique. L'étalonnage est obtenu de façon externe en réalisant un modèle de plaque à partir de 6 spots sur lesquels est déposé un mélange de peptides standards (Proteomix Peptide calibration Mix 4, LaserBio Labs, Sophia Antipolis, France) contenant le fragment 1-5 de la bradykinine (573,315 Da), l'angiotensine humaine II (1 046,542 Da), la neurotensine (1 672,918 Da), le fragment 18-39 de l'ACTH (2 464,199) et la chaîne B de l'insuline (3 494,651 Da).

V-A.6. Recherche dans les banques de données et validation

V-A.6.1. Paramètres de recherche

a Echantillons non digérés

Pour les échantillons de *stratum corneum* non digérés, les recherches ont été menées avec le logiciel de recherche Mascot 2.2 (Matrixscience) via l'utilitaire Mascot Daemon contre la banque de données Swissprot (version 13.6) contenant uniquement les séquences humaines. Une première recherche est menée en *no enzyme* en précisant comme modifications partielles l'oxydation des méthionines, l'acétylation des peptides en N-terminal et la pyroglutamisation des glutamines en N-terminal. Les résultats de cette recherche sont complétés par ceux des recherches avec digestion *in silico* par la trypsine ou la chymotrypsine (9 *missed cleavages* autorisés) avec les mêmes modifications et la fonction *error tolerant*.

Dans ce cas, seuls les peptides identifiés pour des protéines présentes dans la recherche *no enzyme* sont pris en compte. Selon le spectromètre de masse utilisé, la tolérance de masse en MS et en MS/MS n'est pas la même et est rappelée dans le Tableau 14.

Tableau 14 : Paramètres de recherche qui différent selon le spectromètre de masse utilisé.

Appareil utilisé	Tolérance MS	Tolérance MS/MS
ESI-QqTOF	0,3Da	0,3Da
ESI-LIT-FTMS	10ppm	1Da

b Echantillons digérés

Pour les échantillons de *stratum corneum* digérés, les recherches ont été menées avec le logiciel de recherche Mascot 2.2 (Matrixscience) via l'utilitaire Mascot Daemon contre la banque de données Swissprot (version 13.6) contenant uniquement les séquences humaines. Les recherches sont menées avec digestion *in silico* à la trypsine (2 *missed cleavages* autorisés). En ce qui concerne les modifications, l'oxydation des méthionines et la carbamidométhylation des cystéines sont spécifiées en modifications partielles et la fonction *error tolerant* est activée. Les analyses de spectrométrie de masse ont été faites sur le LTQ-FT donc les tolérances de masse utilisées correspondent aux tolérances LIT-FTMS du Tableau 14.

V-A.6.2. Paramètres de validation

a Analyses de peptides endogènes par nLC-nESI-QqTOF

Pour l'analyse de peptides endogènes, nous avons combiné les résultats des analyses *no enzyme* avec les résultats *error tolerant* + trypsine et *error tolerant* + chymotrypsine, en ne conservant que les protéines dont des peptides avaient été validés lors de la recherche *no enzyme* seule [212]. En ce qui concerne les peptides non modifiés, les peptides de score inférieur à 20 ont été éliminés. Pour les peptides modifiés, nous avons pris en compte le nombre de modifications n permettant l'attribution du spectre : seuls les peptides d'*Ions score* supérieur à 20 + 10 x n sont conservés. **Seules les protéines identifiées grâce à 2 séquences d'acides aminés différentes et avec un score protéique supérieur à 60 sont validées** ($\text{score protéique} = \sum \text{Ions score (séquence a.a. non redondants)}$, pour chaque séquence d'a.a., seul le meilleur *Ions score* est pris en compte).

b Analyses de peptides endogènes par nLC-nESI-LIT-FTICR

Pour l'analyse de peptides endogènes, nous avons combiné les résultats des analyses *no enzyme* avec les résultats *error tolerant* + trypsine et *error tolerant* + chymotrypsine, en ne conservant que les protéines dont des peptides avaient été validés lors de la recherche *no enzyme* seule [212]. En ce qui concerne les peptides

non modifiés, les peptides de score inférieur à 20 ont été éliminés. Pour les peptides modifiés, nous avons pris en compte le nombre de modifications n permettant l'attribution du spectre : seuls les peptides d'*Ions score* supérieur à 20 + 10 x n sont conservés. **Les protéines identifiées grâce à 1 séquence d'acides aminés vue en triplicat et de score protéique supérieur à 60 ont été validées** ($\text{score protéique} = \sum \text{Ions score (séquence a.a. non redondants)}$, pour chaque séquence d'a.a., seul le meilleur *Ions score* est pris en compte)**.**

Dans le cadre de l'étude topologique PA/SCP, si une protéine était validée dans un seul type d'échantillon, elle a tout de même été conservée dans les résultats de l'autre type d'échantillon pour mener la comparaison.

c Analyses de protéines digérées par nLC-nESI-LIT-FTICR

Pour l'analyse d'échantillons digérés en LC1D ou en LC2D, seuls les peptides d'*Ions score* supérieur à 20 + 10 x n sont conservés (n nombre de modifications). **Seules les protéines identifiées au moins en duplicat avec un score protéique supérieur à 60 ont été validées** ($\text{score protéique} = \sum \text{Ions score (séquence a.a. non redondants)}$, pour chaque séquence d'a.a., seul le meilleur *Ions score* est pris en compte).

V-B. ACTIVITÉ DES PROTÉASES

V-B.1. Extraction des protéines du *stratum corneum*

Pour extraire les protéines, 64mg de poudre acétonique sont réhydratés avec 6,4mL de tampon (PBS, 1M NaCl, 0,1% Triton X100 v/v). Le mélange est homogénéisé dans un potter en verre. Après centrifugation (10min, 10 000g, 4°C), le surnageant est récupéré et passé à travers un filtre 0,45µm puis 0,22µm. La quantité de protéines du surnageant est mesurée par un dosage Bio-Rad.

L'ensemble des étapes est réalisé sur la glace.

V-B.2. Purification chromatographique

Toutes les étapes de purification ont été réalisées à 4°C sur une station de chromatographie biologique (Bio-Rad). Les extraits de protéines ont été concentrés de 4mL à 1mL par ultrafiltration sur une membrane de seuil 5kDa (Vivaspin 6, Vivascience).

V-B.2.1. Chromatographie par exclusion stérique (SEC)

L'extrait de protéines a été séparé par chromatographie d'exclusion stérique sur une colonne G75HR10/30, préalablement équilibrée par du tampon de migration (PBS, 1M NaCl, Triton X100 0,1%). La séparation a été réalisée à 400µL/min et le volume des fractions collectées est de 150µL. Les fractions de deux injections successives de 250µL ont été réunies.

V-B.2.2. Echange de cations

Un autre extrait de protéines a été séparé par échange de cations sur une colonne Mono S 5/50 GL (Tricorn, GE Healthcare) utilisée en tampon acétate 50mM pH=3,5 (tampon A ; tampon B=tampon A + 1M NaCl). Avant injection, le tampon de l'extrait a été remplacé par du tampon A avec une cartouche EconoPack 10DG (Bio-Rad) : 9mL d'extrait protéique ont été chargés et élués avec 12mL de tampon A. 11mL ont été injectés sur la colonne d'échange de cations. La séparation est réalisée avec un gradient linéaire de 0 à 1M NaCl à 300µL/min en 50min. Le volume des fractions collecté est de 450µL.

V-B.3. Tests enzymatiques

L'activité des protéases a été testée à la fois sur les extraits de protéines bruts et sur les fractions provenant de la séparation SEC et SCX.

Les mesures de fluorescence et d'absorbance ont été réalisées sur un instrument SpectraMax M5e instrument (Molecular Devices, Sunnyvale, CA).

V-B.3.1. Test enzymatique générique

Le kit d'essai Enzcheck (Molecular Probes) contenant de la caséine BODIPY FL comme substrat générique a été utilisé en suivant la procédure du fabricant, *i.e.* en maintenant un rapport extrait/substrat dilué de 1:10 (v/v). L'incubation dure une nuit à 37°C. Cette étude a été faite à différents pH : tampon acétate 0,1M pH=5 et tampon Tris 0,1M pH=8. La fluorescence du substrat est mesurée avec une longueur d'onde d'excitation à 485nm et d'émission de 535nm.

V-B.3.2. Test enzymatique spécifique des protéases à sérine

Les substrats S2288 et S2586 ont été utilisés respectivement pour tester l'activité des sérine protéases (et plus particulièrement de la trypsine) et de la chymotrypsine. S2288 a été employé à une concentration finale de 2,5nmol/µL et avec un rapport extrait/substrat dilué de 5:200 (v/v). S2586 a été utilisé à une concentration finale de 250pmol/µL et avec un rapport extrait/substrat dilué de 5:200 (v/v). L'incubation dure une nuit à 37°C à pH=8. L'absorbance est mesurée à la longueur d'onde de 405nm.

V-B.3.3. Test enzymatique sur les peptides de cornéodesmosine

Pour les tests enzymatiques spécifiques de la cornéodesmosine, 13 peptides synthétiques (JPT Peptides Technologies, Allemagne) ont été utilisés à la concentration finale de 10pmol/µL et avec un rapport extrait/substrat dilué de 1:20 (v/v). L'incubation à 37°C a duré au moins une nuit et l'étude a été faite à la fois à pH=5 et pH=8 (cf V-B.3.1.). La fluorescence des peptides est mesurée avec une longueur d'onde d'excitation de 336nm et une longueur d'émission de 490nm.
Pour étudier les profils d'inhibition des activités détectées, différents inhibiteurs ont été utilisés :

- pepstatine à la concentration finale de 1pmol/µL
- aprotinine à la concentration finale de 15pmol/µL
- cocktail d'inhibiteur complete EDTA-free (Roche Molecular) utilisé selon le protocole du fournisseur

L'activité des protéases a aussi été étudiée en présence de cystéine (5nmol/µL final).

V-B.4. Analyses par MALDI-TOF/TOF

0,3µL d'échantillon (mélange dessalé d'un peptide CDSN avec ou sans extrait de protéases) ont été déposés avec 0,6µL de matrice CHCA (4mg/mL, 6:4:0,01 ACN/H2O/TFA v/v/v, 10mM citrate d'ammonium). Les analyses MS ont été réalisées sur un spectromètre de masse MALDI-TOF/TOF 4800 (Applera Applied Biosystems, Framingham, MA) en mode réflectron positif (gamme de masse 500-2 000), intensité du laser 2 100, 1 000 tirs cumulés). Les données MS/MS ont été acquises en utilisant de l'air comme gaz de collision et avec une énergie de 2keV (fenêtre de sélection de 5Da, intensité du laser 4 500, 2 400 tirs cumulés). L'étalonnage est obtenu de façon externe en réalisant un modèle de plaque à partir de 13 spots sur lesquels est déposé un mélange de peptides standards (Proteomix Peptide calibration Mix 4, LaserBio Labs, Sophia Antipolis, France) contenant le fragment 1-5 de la bradykinine (573,315Da), l'angiotensine humaine II (1 046,542Da), la neurotensine (1 672,918Da), le fragment 18-39 de l'ACTH (2 464,199Da) et la chaîne B de l'insuline (3 494,651Da).

PARTIE VI- TABLE DES FIGURES

Figure 1 : Principe de la recherche protéomique d'activité enzymatique. 18
Figure 2 : Processus d'électrospray. 19
Figure 3 : Schéma d'une expérience DESI typique. 20
Figure 4 : Schéma de principe d'un MALDI-TOF en mode réflectron [33]. 22
Figure 5 : Configuration d'une trappe d'ions à résonance cyclotronique ou piège de Penning. 27
Figure 6 : Mouvement d'ion cyclotronique. 28
Figure 7 : La trajectoire cyclotron orbitale des ions est transformée d'un mouvement incohérent (à gauche) à un mouvement cohérent (à droite) par l'application d'un champ électrique tournant dans le même sens et à la même fréquence cyclotronique que les ions de rapport m/z. Figure d'après Marshall *et al.* [65]. 30
Figure 8 : Géométrie de l'Orbitrap. 31
Figure 9 : Illustration de la capacité de pic d'une séparation bidimensionnelle et de sa dépendance en produit des capacités de pic des 1ère et 2ème dimension [88]. 36
Figure 10 : Illustration de l'effet du temps de collecte (t_s) sur la capacité de pic de la 1ère dimension. 37
Figure 11 : Système de chromatographie multidimensionnelle *on-line*. 38
Figure 12 : Système de chromatographie multidimensionnelle *off-line*. 39
Figure 13 : Illustration de la signification du seuil d'homologie (ASMS 2005, user meeting Matrix Science). Le score du peptide marqué par une flèche se distingue de la distribution de scores des autres peptides. 44
Figure 14 : Schéma de l'épiderme [123]. 47
Figure 15 : Desmosome, mise en évidence de certaines des interactions principales protéines-protéines [147]. 52
Figure 16 : Spectre MS correspondant à l'ionisation d'espèces peptidiques multi-chargées. 59
Figure 17 : Extraction du chromatogramme de l'espèce 913.7Th dont le spectre d'ionisation est représenté Figure 16. 61
Figure 18 : Stratégie générale appliquée sur les échantillons de *stratum corneum* pour l'étude des peptides endogènes. 64
Figure 19 : Apport de chaque tampon dans les identifications. 65
Figure 20 : Apport de chaque fraction dans les identifications. 66

Figure 21 : Chromatogramme extrait des ions de rapport m/z 520,333, 564,359, 608,385, 652,411, 696,438 obtenu sur un blanc sans (A) ou avec (B) cartouche de purification sur un système 1D nanoLC Dionex. 68
Figure 22 : Schéma du montage LC1D développé pour l'élimination des polymères. 70
Figure 23 : Chromatogramme en pic de base avec extraction des scans en MS simple d'une injection de 20fmol de protéines sans (A) ou avec (B) cartouche de purification sur un système Dionex (colonne Agilent avec gradient approprié). 71
Figure 24 : Chromatogrammes extraits des ions de rapport m/z 520,333, 564,359, 608,385, 652,411, 696,438 obtenus sur un blanc en LC2D *on-line*. 73
Figure 25 : Schéma du montage LC2D *on-line* développé pour éliminer les polymères. 73
Figure 26 : Comparaison des différents protocoles sur un échantillon de cils digérés. Elution de nouveaux peptides cumulés en fonction des fractions SCX. 75
Figure 27 : Influence de la concentration en ACN sur l'évolution du nombre de peptides élués en fonction des fractions. 77
Figure 28 : Influence de la période d'échantillonnage sur l'élution des peptides en 1ère dimension. 78
Figure 29 : Répartition des peptides en termes de charge (graphes 1), d'indice hydrophobique [207] (graphes 2) en fonction des fractions selon la configuration utilisée (A : *off-line*, B : *on-line*). L'indice hydrophobique moyen des peptides de chaque fraction triés d'après leur charge est également présenté (graphes 3). 79
Figure 30 : Orthogonalité des deux dimensions en mode *off-line* (A) et en mode *on-line* (B). 80
Figure 31 : Comparaison des identifications en termes de protéines et de peptides suivant les deux configurations LC2D et mise en évidence de l'intérêt par rapport à la LC1D. 82
Figure 32 : Mise en évidence du gain en couverture de séquence des protéines déjà identifiées en LC1D. 83
Figure 33 : Comparaison des identifications en termes de protéines et de peptides suivant les deux configurations LC2D-ESI et mise en évidence de l'intérêt l'utilisation d'une source d'ions complémentaire MALDI. Quasiment autant de protéines et de peptides sont identifiés par le mode *off-line*MALDI et par le mode *off-line*ESI. Cependant seuls 2/3 des protéines et 1/5 des peptides sont identifiés en commun par les deux méthodes. D'autre part, même si le mode *on-line*ESI est le plus performant, très peu de peptides identifiés en *off-line*MALDI le sont également en *on-line*ESI. 85
Figure 34 : Comparaison du profil d'élution chromatographique d'un même peptide en LC1D longue (A) et en LC1D classique (B). 87

Figure 35 : Mise en évidence de l'intérêt d'utiliser des gradients longs. 88% des protéines identifiées grâce aux protocoles de LC2D (soit environ 6 x 12h) se retrouvent dans les identifications du protocole LC1D long (soit 3 x 12h). Cependant 39% des peptides identifiés par ce protocole ne le sont pas en LC2D. 88
Figure 36 : Peptides fragmentés sur le QqTOF (A) pendant la même fenêtre temporelle que sur le LTQ-FT (B)..... 90
Figure 37 : Comparaison des spectres obtenus sur le même précurseur (804,4 Th) avec le QqTOF (A) et le LTQ (B). 91
Figure 38 : Profil d'élution d'un pic chromatographique..... 92
Figure 39 : Bruit électronique classiquement obtenu sur un *scan* MS avec détection dans l'Orbitrap..... 93
Figure 40 : Suivi dans 3 scans FTMS successifs de l'élution d'un peptide peu abondant de rapport m/z 778,322Th dans 20fmol de LCP..... 94
Figure 41 : Influence du seuil de sélection sur le score moyen des peptides identifiés pour un mélange de 20fmol de LCP..... 94
Figure 42 : Influence du seuil de sélection sur le nombre de peptides uniques identifiés sur un échantillon standard de 10fmol de LCP..... 95
Figure 43 : Influence du seuil de détection sur le nombre de peptides uniques identifiés dans un échantillon complexe de *stratum corneum* plantaire, en particulier en ce qui concerne les protéines autres que les kératines. 96
Figure 44 : Influence du seuil de détection sur la couverture de séquence des protéines du cornéodesmosome ou de la famille des plakines. 96
Figure 45 : Profils de différents peptides extraits d'une séparation de 20fmol de LCP. 97
Figure 46 : Comparaison des couvertures de séquence de la cornéodesmosine en fonction du seuil de détection et de la durée d'exclusion dynamique..... 97
Figure 47 : Comparaison du nombre de peptides identifiés selon le nombre de MS/MS réalisées à chaque cycle..... 98
Figure 48 : Interprétation validée d'un spectre MS/MS lors de la confrontation dans la banque Swiss-Prot des données LC2D-MS/MS obtenues sur un échantillon de SCP, avec l'hypothèse que les peptides sont semi-tryptiques..... 100
Figure 49 : Apport de la prise en compte des différentes modifications. 104
Figure 50 : Interprétation validée d'un spectre MS/MS lors de la confrontation dans la banque Swiss-Prot des données LC1D-FTMS obtenues sur un échantillon de SCP, avec l'hypothèse que les peptides sont acétylés. 106

Figure 51 : Interprétation validée d'un spectre MS/MS lors de la confrontation dans la banque Swiss-Prot des données LC1D-FTMS obtenues sur un échantillon de SCP, avec l'hypothèse d'une pyroglutamisation N-terminale. 106
Figure 52 : Répétabilité des identifications protéiques sur l'analyse d'un échantillon de SCP digéré par LC1D-FTMS/MS. 108
Figure 53 : Comparaison topologique des protéines identifiées dans le *stratum corneum* plantaire ou non. 113
Figure 54 : Comparaison des protéines identifiées selon le type de peau étudié. 114
Figure 55 : Profil de dégradation des cadhérines identifiées. 117
Figure 56 : Profil de dégradation de la cornéodesmosine. 119
Figure 57 : Structure primaire de la desmoplakine (d'après [150]) montrant les différents domaines et sous-domaines de la protéine ainsi que les zones où nous avons identifié des peptides endogènes. 120
Figure 58 : Peptides dérivés de la DCD identifiés dans les échantillons de *stratum corneum* non plantaire (PA) et plantaire (SCP). En rouge est rappelé le peptide DCD-1. 122
Figure 59 : Peptides dérivés des protéines de la famille S100, identifiés dans les échantillons de *stratum corneum*. En rouge est rappelée la zone centrale nécessaire à l'activité antibactérienne de la S100A7. 123
Figure 60 : Activités protéolytiques principalement observées d'après les données LC-MS/MS. 125
Figure 61 : Sites de coupure autres que SCCE et SCTE identifiés sur la cornéodesmosine. 126
Figure 62 : Séquence de la cornéodesmosine ; les crochets correspondent aux sites de clivage identifiés en LC-MS/MS. 126
Figure 63 : Formule des groupes Dabcyl et Edans 127
Figure 64 : Activité enzymatique d'un extrait total de protéines de la peau sur les peptides de cornéodesmosine de synthèse 130
Figure 65 : Analyse MS du mélange CDSN1+extrait protéique à pH=5 et pH=8 en MALDI-TOF/TOF 131
Figure 66 : Spectre de fragmentation de l'espèce 584,29Da qui apparaît dans les spectres MS du peptide CDSN1 lors de sa dégradation par un extrait de protéines du *stratum corneum* à pH basique 132
Figure 67 : Profil d'activité enzymatique des fractions obtenues après séparation par SEC d'un extrait total de protéines de *stratum corneum*. 133
Figure 68 : Profil d'activité enzymatique des fractions obtenues après séparation par SCX d'un extrait total de protéines de *stratum corneum*. 134

Figure 69 : Profil d'activité enzymatique des fractions obtenues après séparation d'un extrait de protéines de peau sur le peptide CDSN1 en condition basique. 135
Figure 70 : Profil d'activité enzymatique des fractions obtenues après séparation d'un extrait de protéines de peau sur le peptide CDSN7 en condition acide. 136
Figure 71 : Profil d'activité enzymatique des fractions obtenues après séparation d'un extrait de protéines de peau sur le peptide CDSN3 en condition basique. 137
Figure 72 : Influence de la cystéine sur la dégradation du peptide CDSN3 par un extrait brut de protéines du SC. Quelque soit le pH, une dégradation accélérée est observée en présence de cystéine. 137
Figure 73 : Influence de la cystéine sur la dégradation du peptide CDSN3 par les fractions SCX. 138
Figure 74 : Robot pipeteur/manipulateur Genesis Proteam 150 146
Figure 75 : Schéma du montage LC2D *on-line*. 150
Figure 76 : Schéma du montage LC2D *off-line*. 152
Figure 77 : Cycle d'analyse automatique sur le QqTOF 153
Figure 78 : Cycle d'acquisition automatique utilisé sur le LTQ-FTCIR afin d'augmenter la gamme dynamique d'après [209]. 153

PARTIE VII- TABLE DES TABLEAUX

Tableau 1 : Outils disponibles pour l'interprétation des spectres MS/MS (d'après [95]). 41
Tableau 2 : Protéines identifiées dans les échantillons de SCP par reconcentration seule des peptides endogènes suivie de l'analyse par LC-MS/MS. 59
Tableau 3 : Protéines identifiées dans les fractions inf 5kDa des échantillons de SCP après fractionnement, reconcentration sur ZipTip™ et analyse par LC-MS/MS. 61
Tableau 4 : Protéines identifiées dans les fractions 5-10kDa et sup 10kDa des échantillons de SCP après digestion tryptique et analyse par LC-MS/MS.....63
Tableau 5 : Caractéristiques chromatographiques des contaminants de PEG sur un système 1D RPLC Dionex.....69
Tableau 6 : Identification en triplicat de 20fmol d'un mélange de protéines sans et avec cartouche de purification. 71
Tableau 7 : Quelques MPT importantes et couramment observées. 102
Tableau 8 : Paramètres de recherche testés sur les échantillons non digérés de SCP. 103
Tableau 9 : Distribution des identifications peptidiques parmi les différentes catégories en termes de spectres MS/MS et de peptides uniques. Sont aussi comptés les nouveaux peptides [a]par rapport aux peptides non modifiés ou [b]par rapport à la recherche initiale (avec oxydation des méthionines) ainsi que les identifications protéiques. 104
Tableau 10 : Protéines dont des peptides endogènes ont été identifiés dans des échantillons de peau par nanoLC-MS/MS. Combinaison des résultats des fractions inf 5kDa et 5-10kDa. 110
Tableau 11 : Sites de coupure des protéases décrites dans la dégradation des cornéodesmosomes. 124
Tableau 12 : Caractéristiques des peptides Dabcyl-Edans synthétisés. 127
Tableau 13 : Gradient optimisé en condition LC2D *on-line*. 151
Tableau 14 : Paramètres de recherche qui différent selon le spectromètre de masse utilisé. 155

PARTIE VIII- BIBLIOGRAPHIE

1. Anderson, N.L. and N.G. Anderson, *Proteome and proteomics: new technologies, new concepts, and new words.* Electrophoresis, 1998. **19**(11): p. 1853-61.
2. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
3. Kosak, S.T. and M. Groudine, *Gene order and dynamic domains.* Science, 2004. **306**(5696): p. 644-7.
4. Haynes, P.A., N. Fripp, and R. Aebersold, *Identification of gel-separated proteins by liquid chromatography-electrospray tandem mass spectrometry: comparison of methods and their limitations.* Electrophoresis, 1998. **19**(6): p. 939-45.
5. Schrader, M. and P. Schulz-Knappe, *Peptidomics technologies for human body fluids.* Trends Biotechnol, 2001. **19**(10 Suppl): p. S55-60.
6. Schulz-Knappe, P., et al., *Peptidomics: the comprehensive analysis of peptides in complex biological mixtures.* Comb Chem High Throughput Screen, 2001. **4**(2): p. 207-17.
7. Gygi, S.P. and R. Aebersold, *Mass spectrometry and proteomics.* Curr Opin Chem Biol, 2000. **4**(5): p. 489-94.
8. Clynen, E., A. De Loof, and L. Schoofs, *The use of peptidomics in endocrine research.* Gen Comp Endocrinol, 2003. **132**(1): p. 1-9.
9. Baggerman, G., et al., *Peptidomic analysis of the larval Drosophila melanogaster central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry.* J Mass Spectrom, 2005. **40**(2): p. 250-60.
10. Desiderio, D.M., *Mass spectrometric analysis of neuropeptidergic systems in the human pituitary and cerebrospinal fluid.* J Chromatogr B Biomed Sci Appl, 1999. **731**(1): p. 3-22.
11. Che, F.Y., et al., *Identification of peptides from brain and pituitary of Cpe(fat)/Cpe(fat) mice.* Proc Natl Acad Sci U S A, 2001. **98**(17): p. 9971-6.
12. Georgiou, C.D., et al., *Mechanism of Coomassie brilliant blue G-250 binding to proteins: a hydrophobic assay for nanogram quantities of proteins.* Anal Bioanal Chem, 2008. **391**(1): p. 391-403.
13. Chen, C.H., *Review of a current role of mass spectrometry for proteome research.* Anal Chim Acta, 2008. **624**(1): p. 16-36.
14. Petricoin, E.F., et al., *Use of proteomic patterns in serum to identify ovarian cancer.* Lancet, 2002. **359**(9306): p. 572-7.
15. Puente, X.S., et al., *A genomic view of the complexity of mammalian proteolytic systems.* Biochem Soc Trans, 2005. **33**(Pt 2): p. 331-4.
16. Puente, X.S., et al., *Human and mouse proteases: a comparative genomic approach.* Nat Rev Genet, 2003. **4**(7): p. 544-58.
17. Neurath, H., *Evolution of proteolytic enzymes.* Science, 1984. **224**(4647): p. 350-7.
18. McQuibban, G.A., et al., *Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3.* Science, 2000. **289**(5482): p. 1202-6.
19. Lopez-Otin, C. and C.M. Overall, *Protease degradomics: a new challenge for proteomics.* Nat Rev Mol Cell Biol, 2002. **3**(7): p. 509-19.
20. Overall, C.M. and R.A. Dean, *Degradomics: systems biology of the protease web. Pleiotropic roles of MMPs in cancer.* Cancer Metastasis Rev, 2006. **25**(1): p. 69-75.

21. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.
22. Birner-Gruenberger, R., et al., *The lipolytic proteome of mouse adipose tissue.* Mol Cell Proteomics, 2005. **4**(11): p. 1710-7.
23. Schmidinger, H., A. Hermetter, and R. Birner-Gruenberger, *Activity-based proteomics: enzymatic activity profiling in complex proteomes.* Amino Acids, 2006. **30**(4): p. 333-50.
24. Whitehouse, C.M., et al., *Electrospray Interface for Liquid Chromatographs and Mass Spectrometers.* Analytical Chemistry, 1985. **57**(3): p. 675-679.
25. Takats, Z., et al., *Mass spectrometry sampling under ambient conditions with desorption electrospray ionization.* Science, 2004. **306**(5695): p. 471-473.
26. Whitehouse, C.M., et al., *Electrospray interface for liquid chromatographs and mass spectrometers.* Anal Chem, 1985. **57**(3): p. 675-9.
27. Chatman, K., et al., *Nanoelectrospray mass spectrometry and precursor ion monitoring for quantitative steroid analysis and attomole sensitivity.* Anal Chem, 1999. **71**(13): p. 2358-63.
28. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal Chem, 1988. **60**(20): p. 2299-301.
29. Vertes, A., G. Irinyi, and R. Gijbels, *Hydrodynamic Model of Matrix-Assisted Laser-Desorption Mass-Spectrometry.* Analytical Chemistry, 1993. **65**(17): p. 2389-2393.
30. Karas, M. and R. Kruger, *Ion formation in MALDI: the cluster ionization mechanism.* Chem Rev, 2003. **103**(2): p. 427-40.
31. Chang, W.C., et al., *Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited.* Anal Chim Acta, 2007. **582**(1): p. 1-9.
32. Nordhoff, E., H. Lehrach, and J. Gobom, *Exploring the limits and losses in MALDI sample preparation of attomole amounts of peptide mixtures.* International Journal of Mass Spectrometry, 2007. **268**(2-3): p. 139-146.
33. Vinh, J., *Etude et caractérisation structurale de modifications post-traductionnelles de la tubuline par spectrométrie de masse*. 1999, Université Pierre et Marie Curie - Paris 6: Paris, France.
34. Chaurand, P., K.E. Schriver, and R.M. Caprioli, *Instrument design and characterization for high resolution MALDI-MS imaging of tissue sections.* J Mass Spectrom, 2007. **42**(4): p. 476-89.
35. Lee, J.C., et al., *Reactivity-based one-pot synthesis of the tumor-associated antigen N3 minor octasaccharide for the development of a photocleavable DIOS-MS sugar array.* Angew Chem Int Ed Engl, 2006. **45**(17): p. 2753-7.
36. Hsu, N.Y., et al., *Desorption ionization of biomolecules on metals.* Anal Chem, 2008. **80**(13): p. 5203-10.
37. Reyzer, M.L. and R.M. Caprioli, *MALDI-MS-based imaging of small molecules and proteins in tissues.* Curr Opin Chem Biol, 2007. **11**(1): p. 29-35.
38. Baldwin, M.A., *Mass spectrometers for the analysis of biomolecules.* Methods Enzymol, 2005. **402**: p. 3-48.
39. Domon, B. and R. Aebersold, *Mass spectrometry and protein analysis.* Science, 2006. **312**(5771): p. 212-7.
40. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.
41. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.

42. Wolters, D.A., M.P. Washburn, and J.R. Yates, 3rd, *An automated multidimensional protein identification technology for shotgun proteomics.* Anal Chem, 2001. **73**(23): p. 5683-90.
43. Clauser, K.R., P. Baker, and A.L. Burlingame, *Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching.* Anal Chem, 1999. **71**(14): p. 2871-82.
44. Conrads, T.P., et al., *Utility of accurate mass tags for proteome-wide protein identification.* Anal Chem, 2000. **72**(14): p. 3349-54.
45. Sleno, L., D.A. Volmer, and A.G. Marshall, *Assigning product ions from complex MS/MS spectra: the importance of mass uncertainty and resolving power.* J Am Soc Mass Spectrom, 2005. **16**(2): p. 183-98.
46. Takach, E.J., et al., *Accurate mass measurements using MALDI-TOF with delayed extraction.* J Protein Chem, 1997. **16**(5): p. 363-9.
47. Zubarev, R.A., P. Hakansson, and B. Sundqvist, *Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements.* Anal Chem, 1996. **68**(22): p. 4060-4063.
48. Gorshkov, M.V. and R.A. Zubarev, *On the accuracy of polypeptide masses measured in a linear ion trap.* Rapid Commun Mass Spectrom, 2005. **19**(24): p. 3755-8.
49. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.
50. Qian, W.J., et al., *Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome.* J Proteome Res, 2005. **4**(1): p. 53-62.
51. Liu, T., et al., *High dynamic range characterization of the trauma patient plasma proteome.* Mol Cell Proteomics, 2006. **5**(10): p. 1899-913.
52. Peng, J., et al., *Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome.* J Proteome Res, 2003. **2**(1): p. 43-50.
53. Beausoleil, S.A., et al., *Large-scale characterization of HeLa cell nuclear phosphoproteins.* Proc Natl Acad Sci U S A, 2004. **101**(33): p. 12130-5.
54. Olsen, J.V. and M. Mann, *Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation.* Proc Natl Acad Sci U S A, 2004. **101**(37): p. 13417-22.
55. Nielsen, M.L., M.M. Savitski, and R.A. Zubarev, *Improving protein identification using complementary fragmentation techniques in fourier transform mass spectrometry.* Mol Cell Proteomics, 2005. **4**(6): p. 835-45.
56. Haas, W., et al., *Optimization and use of peptide mass measurement accuracy in shotgun proteomics.* Mol Cell Proteomics, 2006. **5**(7): p. 1326-37.
57. He, F., et al., *Theoretical and experimental prospects for protein identification based solely on accurate mass measurement.* J Proteome Res, 2004. **3**(1): p. 61-7.
58. Smith, R.D., et al., *An accurate mass tag strategy for quantitative and high-throughput proteome measurements.* Proteomics, 2002. **2**(5): p. 513-23.
59. Zubarev, R. and M. Mann, *On the proper use of mass accuracy in proteomics.* Mol Cell Proteomics, 2007. **6**(3): p. 377-81.
60. Comisarow, M.B. and A.G. Marshall, *Theory of Fourier-Transform Ion-Cyclotron Resonance Mass-Spectroscopy .1. Fundamental Equations and Low-Pressure Line-Shape.* Journal of Chemical Physics, 1976. **64**(1): p. 110-119.
61. Dawson, J. and M. Guilhaus, *Orthogonal-acceleration time-of-flight mass spectrometer.* Rapid Commun Mass Spectrom, 1989. **3**(5): p. 155-159.

62. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* Anal Chem, 2000. **72**(6): p. 1156-62.
63. Hipple, J.A., H. Sommer, and H.A. Thomas, *A Precise Method of Determining the Faraday by Magnetic Resonance.* Physical Review, 1949. **76**(12): p. 1877-1878.
64. Guan, S.H. and A.G. Marshall, *Ion Traps for Fourier-Transform Ion-Cyclotron Resonance Mass-Spectrometry - Principles and Design of Geometric and Electric Configurations.* International Journal of Mass Spectrometry and Ion Processes, 1995. **146**: p. 261-296.
65. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, *Fourier transform ion cyclotron resonance mass spectrometry: A primer.* Mass Spectrometry Reviews, 1998. **17**(1): p. 1-35.
66. Dehmelt, H., *Experiments with an Isolated Subatomic Particle at Rest.* Reviews of Modern Physics, 1990. **62**(3): p. 525-530.
67. Kingdon, K.H., *A method for the neutralization of electron space charge by positive ionization at very low gas pressures.* Physical Review, 1923. **21**(4): p. 408-418.
68. Hunt, D.F., et al., *Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer.* Biomed Mass Spectrom, 1981. **8**(9): p. 397-408.
69. Papayannopoulos, I.A., *The Interpretation of Collision-Induced Dissociation Tandem Mass-Spectra of Peptides.* Mass Spectrometry Reviews, 1995. **14**(1): p. 49-73.
70. Johnson, R.S., et al., *Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine.* Anal Chem, 1987. **59**(21): p. 2621-5.
71. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom, 1984. **11**(11): p. 601.
72. Biemann, K., *Appendix 5. Nomenclature for peptide fragment ions (positive ions).* Methods Enzymol, 1990. **193**: p. 886-7.
73. Polce, M.J., D. Ren, and C. Wesdemiotis, *Dissociation of the peptide bond in protonated peptides.* J Mass Spectrom, 2000. **35**(12): p. 1391-8.
74. Wysocki, V.H., et al., *Mobile and localized protons: a framework for understanding peptide dissociation.* J Mass Spectrom, 2000. **35**(12): p. 1399-406.
75. Chen, G., et al., *Applications of LC/MS in structure identifications of small molecules and proteins in drug discovery.* J Mass Spectrom, 2007. **42**(3): p. 279-87.
76. Bodnar, W.M., et al., *Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage.* J Am Soc Mass Spectrom, 2003. **14**(9): p. 971-9.
77. Giddings, J.C., *Maximum Number of Components Resolvable by Gel Filtration and Other Elution Chromatographic Methods.* Analytical Chemistry, 1967. **39**(8): p. 1027-&.
78. Horvath, C.G. and S.R. Lipsky, *Peak Capacity in Chromatography.* Analytical Chemistry, 1967. **39**(14): p. 1893.
79. Dolan, J.W., et al., *Reversed-phase liquid chromatographic separation of complex samples by optimizing temperature and gradient time I. Peak capacity limitations.* J Chromatogr A, 1999. **857**(1-2): p. 1-20.
80. Neue, U.D. and J.R. Mazzeo, *A theoretical study of the optimization of gradients at elevated temperature.* Journal of Separation Science, 2001. **24**(12): p. 921-929.
81. Gilar, M., et al., *Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography.* J Chromatogr A, 2004. **1061**(2): p. 183-92.

82. Guiochon, G., *The limits of the separation power of unidimensional column liquid chromatography.* J Chromatogr A, 2006. **1126**(1-2): p. 6-49.
83. Wang, X., W.E. Barber, and P.W. Carr, *A practical approach to maximizing peak capacity by using long columns packed with pellicular stationary phases for proteomic research.* J Chromatogr A, 2006. **1107**(1-2): p. 139-51.
84. Shen, Y., et al., *Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics.* Anal Chem, 2005. **77**(10): p. 3090-100.
85. Toll, H., et al., *Comparing monolithic and microparticular capillary columns for the separation and analysis of peptide mixtures by liquid chromatography-mass spectrometry.* J Sep Sci, 2005. **28**(14): p. 1666-74.
86. Giddings, J.C., *Two-dimensional separations: concept and promise.* Anal Chem, 1984. **56**(12): p. 1258A-1260A, 1262A, 1264A passim.
87. Guiochon, G., et al., *Theoretical Investigation of the Potentialities of the Use of a Multidimensional Column in Chromatography.* Journal of Chromatography, 1983. **255**(JAN): p. 415-437.
88. Stoll, D.R., et al., *Fast, comprehensive two-dimensional liquid chromatography.* J Chromatogr A, 2007. **1168**(1-2): p. 3-43.
89. Gilar, M., et al., *Orthogonality of separation in two-dimensional liquid chromatography.* Anal Chem, 2005. **77**(19): p. 6426-34.
90. Tyan, Y.C., et al., *Proteomic profiling of human pleural effusion using two-dimensional nano liquid chromatography tandem mass spectrometry.* J Proteome Res, 2005. **4**(4): p. 1274-86.
91. Kislinger, T., et al., *Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling.* Cell, 2006. **125**(1): p. 173-86.
92. Alpert, A.J. and P.C. Andrews, *Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica.* J Chromatogr, 1988. **443**: p. 85-96.
93. Chen, M., et al., *Analysis of human liver proteome using replicate shotgun strategy.* Proteomics, 2007. **7**(14): p. 2479-88.
94. Wang, H., et al., *Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment.* J Proteome Res, 2006. **5**(2): p. 361-9.
95. Nesvizhskii, A.I., O. Vitek, and R. Aebersold, *Analysis and validation of proteomic data generated by tandem mass spectrometry.* Nat Methods, 2007. **4**(10): p. 787-97.
96. Kapp, E. and F. Schutz, *Overview of tandem mass spectrometry (MS/MS) database search algorithms.* Curr Protoc Protein Sci, 2007. **Chapter 25**: p. Unit25 2.
97. Eng, J.K., A.L. McCormack, and J.R. Yates, *An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database.* Journal of the American Society for Mass Spectrometry, 1994. **5**(11): p. 976-989.
98. Zhang, N., R. Aebersold, and B. Schwikowski, *ProbID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data.* Proteomics, 2002. **2**(10): p. 1406-12.
99. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra.* Bioinformatics, 2004. **20**(9): p. 1466-7.
100. Colinge, J., et al., *OLAV: towards high-throughput tandem mass spectrometry data identification.* Proteomics, 2003. **3**(8): p. 1454-63.
101. Geer, L.Y., et al., *Open mass spectrometry search algorithm.* J Proteome Res, 2004. **3**(5): p. 958-64.

102. Matthiesen, R., et al., *VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins.* J Proteome Res, 2005. **4**(6): p. 2338-47.
103. Tabb, D.L., C.G. Fernando, and M.C. Chambers, *MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis.* J Proteome Res, 2007. **6**(2): p. 654-61.
104. Lam, H., et al., *Development and validation of a spectral library searching method for peptide identification from MS/MS.* Proteomics, 2007. **7**(5): p. 655-67.
105. Craig, R., J.P. Cortens, and R.C. Beavis, *The use of proteotypic peptide libraries for protein identification.* Rapid Commun Mass Spectrom, 2005. **19**(13): p. 1844-50.
106. Frewen, B.E., et al., *Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries.* Anal Chem, 2006. **78**(16): p. 5678-84.
107. Johnson, R.S. and J.A. Taylor, *Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry.* Mol Biotechnol, 2002. **22**(3): p. 301-15.
108. Frank, A. and P. Pevzner, *PepNovo: de novo peptide sequencing via probabilistic network modeling.* Anal Chem, 2005. **77**(4): p. 964-73.
109. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.* Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.
110. Tabb, D.L., A. Saraf, and J.R. Yates, 3rd, *GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model.* Anal Chem, 2003. **75**(23): p. 6415-21.
111. Tanner, S., et al., *InsPecT: identification of posttranslationally modified peptides from tandem mass spectra.* Anal Chem, 2005. **77**(14): p. 4626-39.
112. Hernandez, P., et al., *Popitam: towards new heuristic strategies to improve protein identification from tandem mass spectrometry data.* Proteomics, 2003. **3**(6): p. 870-8.
113. Shilov, I.V., et al., *The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra.* Mol Cell Proteomics, 2007. **6**(9): p. 1638-55.
114. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.
115. Yates, J.R., 3rd, J.K. Eng, and A.L. McCormack, *Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases.* Anal Chem, 1995. **67**(18): p. 3202-10.
116. Kapp, E.A., et al., *An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis.* Proteomics, 2005. **5**(13): p. 3475-90.
117. Pappin, D.J., P. Hojrup, and A.J. Bleasby, *Rapid identification of proteins by peptide-mass fingerprinting.* Curr Biol, 1993. **3**(6): p. 327-32.
118. Brunner, E., et al., *A high-quality catalog of the Drosophila melanogaster proteome.* Nat Biotechnol, 2007. **25**(5): p. 576-83.
119. Nesvizhskii, A.I., et al., *Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides.* Mol Cell Proteomics, 2006. **5**(4): p. 652-70.
120. Mann, M. and M. Wilm, *Error-tolerant identification of peptides in sequence databases by peptide sequence tags.* Anal Chem, 1994. **66**(24): p. 4390-9.
121. Holbrook, K.A. and K. Wolff, *The structure and development of skin*, in *Dermatology in General Medecine*, McGraw-Hill, Editor. 1987: New-York. p. 92-131.
122. Odland, G.F., *Histology and fine structure of the epidermis*, in *The skin: by 30 authors*, W. Wilkins, Editor. 1971: Baltimore. p. 28-46.

123. Laplante, A., *Mécanismes de réépithéliation des plaies cutanées : expression des protéines de stress chez la souris et analyse à l'aide d'un nouveau modèle tridimensionnel humain développé par génie tissulaire*. 2002, Université Laval: Sainte-Foy, Canada.
124. Fuchs, E. and H. Green, *Changes in keratin gene expression during terminal differentiation of the keratinocyte.* Cell, 1980. **19**(4): p. 1033-42.
125. Eckert, R.L., *Structure, function, and differentiation of the keratinocyte.* Physiol Rev, 1989. **69**(4): p. 1316-46.
126. Thacher, S.M. and R.H. Rice, *Keratinocyte-specific transglutaminase of cultured human epidermal cells: relation to cross-linked envelope formation and terminal differentiation.* Cell, 1985. **40**(3): p. 685-95.
127. Mehrel, T., et al., *Identification of a major keratinocyte cell envelope protein, loricrin.* Cell, 1990. **61**(6): p. 1103-12.
128. Green, H., *Terminal differentiation of cultured human epidermal cells.* Cell, 1977. **11**(2): p. 405-16.
129. Elias, P.M., *Epidermal lipids, barrier function, and desquamation.* J Invest Dermatol, 1983. **80 Suppl**: p. 44s-49s.
130. Rawlings, A.V. and C.R. Harding, *Moisturization and skin barrier function.* Dermatol Ther, 2004. **17 Suppl 1**: p. 43-8.
131. Chapman, S.J. and A. Walsh, *Desmosomes, corneosomes and desquamation. An ultrastructural study of adult pig epidermis.* Arch Dermatol Res, 1990. **282**(5): p. 304-10.
132. Moll, R., M. Divo, and L. Langbein, *The human keratins: biology and pathology.* Histochem Cell Biol, 2008. **129**(6): p. 705-33.
133. Hesse, M., T.M. Magin, and K. Weber, *Genes for intermediate filament proteins and the draft sequence of the human genome: novel keratin genes and a surprisingly high number of pseudogenes related to keratin genes 8 and 18.* J Cell Sci, 2001. **114**(Pt 14): p. 2569-75.
134. Fuchs, E., *Epidermal differentiation and keratin gene expression.* J Cell Sci Suppl, 1993. **17**: p. 197-208.
135. Roop, D.R., et al., *Regulated expression of differentiation-associated keratins in cultured epidermal cells detected by monospecific antibodies to unique peptides of mouse epidermal keratins.* Differentiation, 1987. **35**(2): p. 143-50.
136. Moll, I., et al., *Distribution of a special subset of keratinocytes characterized by the expression of cytokeratin 9 in adult and fetal human epidermis of various body sites.* Differentiation, 1987. **33**(3): p. 254-65.
137. Stappenbeck, T.S. and K.J. Green, *The desmoplakin carboxyl terminus coaligns with and specifically disrupts intermediate filament networks when expressed in cultured cells.* J Cell Biol, 1992. **116**(5): p. 1197-209.
138. Meng, J.J., et al., *Two-hybrid analysis reveals fundamental differences in direct interactions between desmoplakin and cell type-specific intermediate filaments.* J Biol Chem, 1997. **272**(34): p. 21495-503.
139. Stappenbeck, T.S., et al., *Functional analysis of desmoplakin domains: specification of the interaction with keratin versus vimentin intermediate filament networks.* J Cell Biol, 1993. **123**(3): p. 691-705.
140. Ku, N.O., et al., *The cytoskeleton of digestive epithelia in health and disease.* Am J Physiol, 1999. **277**(6 Pt 1): p. G1108-37.
141. Bruno, T., et al., *The RNA polymerase II core subunit 11 interacts with keratin 19, a component of the intermediate filament proteins.* FEBS Lett, 1999. **453**(3): p. 273-7.

142. Bonifas, J.M., A.L. Rothman, and E.H. Epstein, Jr., *Epidermolysis bullosa simplex: evidence in two families for keratin gene abnormalities.* Science, 1991. **254**(5035): p. 1202-5.
143. Bernard, D., *Etude du mécanisme de la desquamation et des processus enzymatiques qui lui sont associés au cours de la différentiation tardive de l'épiderme humain.* 2005, Université Toulouse III Paul Sabatier: Toulouse, France.
144. Serre, G., et al., *Identification of late differentiation antigens of human cornified epithelia, expressed in re-organized desmosomes and bound to cross-linked envelope.* J Invest Dermatol, 1991. **97**(6): p. 1061-72.
145. Stokes, D.L., *Desmosomes from a structural perspective.* Curr Opin Cell Biol, 2007. **19**(5): p. 565-71.
146. Lundstrom, A., et al., *Evidence for a role of corneodesmosin, a protein which may serve to modify desmosomes during cornification, in stratum corneum cell cohesion and desquamation.* Arch Dermatol Res, 1994. **286**(7): p. 369-75.
147. Fuchs, E. and S. Raghavan, *Getting under the skin of epidermal morphogenesis.* Nat Rev Genet, 2002. **3**(3): p. 199-209.
148. O'Keefe, E.J., H.P. Erickson, and V. Bennett, *Desmoplakin I and desmoplakin II. Purification and characterization.* J Biol Chem, 1989. **264**(14): p. 8310-8.
149. Hatzfeld, M., *Plakophilins: Multifunctional proteins or just regulators of desmosomal adhesion?* Biochim Biophys Acta, 2007. **1773**(1): p. 69-77.
150. Choi, H.J., et al., *Structures of two intermediate filament-binding fragments of desmoplakin reveal a unique repeat motif structure.* Nat Struct Biol, 2002. **9**(8): p. 612-20.
151. Green, K.J., et al., *Structure of the human desmoplakins. Implications for function in the desmosomal plaque.* J Biol Chem, 1990. **265**(5): p. 2603-12.
152. Chitaev, N.A., et al., *The binding of plakoglobin to desmosomal cadherins: patterns of binding sites and topogenic potential.* J Cell Biol, 1996. **133**(2): p. 359-69.
153. Palka, H.L. and K.J. Green, *Roles of plakoglobin end domains in desmosome assembly.* J Cell Sci, 1997. **110 (Pt 19)**: p. 2359-71.
154. Kowalczyk, A.P., et al., *The head domain of plakophilin-1 binds to desmoplakin and enhances its recruitment to desmosomes. Implications for cutaneous disease.* J Biol Chem, 1999. **274**(26): p. 18145-8.
155. South, A.P., et al., *Lack of plakophilin 1 increases keratinocyte migration and reduces desmosome stability.* J Cell Sci, 2003. **116**(Pt 16): p. 3303-14.
156. Goodwin, L., et al., *Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.* Biochem Biophys Res Commun, 1990. **173**(3): p. 1224-30.
157. Nilles, L.A., et al., *Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.* J Cell Sci, 1991. **99 (Pt 4)**: p. 809-21.
158. Arnemann, J., et al., *Stratification-related expression of isoforms of the desmosomal cadherins in human epidermis.* J Cell Sci, 1993. **104 (Pt 3)**: p. 741-50.
159. King, I.A., et al., *Hierarchical expression of desmosomal cadherins during stratified epithelial morphogenesis in the mouse.* Differentiation, 1997. **62**(2): p. 83-96.
160. North, A.J., et al., *Distinct desmocollin isoforms occur in the same desmosomes and show reciprocally graded distributions in bovine nasal epidermis.* Proc Natl Acad Sci U S A, 1996. **93**(15): p. 7701-5.
161. Chitaev, N.A. and S.M. Troyanovsky, *Direct Ca2+-dependent heterophilic interaction between desmosomal cadherins, desmoglein and desmocollin, contributes to cell-cell adhesion.* J Cell Biol, 1997. **138**(1): p. 193-201.

162. Marcozzi, C., et al., *Coexpression of both types of desmosomal cadherin and plakoglobin confers strong intercellular adhesion.* J Cell Sci, 1998. **111 (Pt 4)**: p. 495-509.
163. Jonca, N., et al., *Corneodesmosin, a component of epidermal corneocyte desmosomes, displays homophilic adhesive properties.* J Biol Chem, 2002. **277**(7): p. 5024-9.
164. Caubet, C., et al., *Homo-oligomerization of human corneodesmosin is mediated by its N-terminal glycine loop domain.* J Invest Dermatol, 2004. **122**(3): p. 747-54.
165. Simon, M., et al., *Refined characterization of corneodesmosin proteolysis during terminal differentiation of human epidermis and its relationship to desquamation.* J Biol Chem, 2001. **276**(23): p. 20292-9.
166. Yang, T., et al., *Epidermal detachment, desmosomal dissociation, and destabilization of corneodesmosin in Spink5-/- mice.* Genes Dev, 2004. **18**(19): p. 2354-8.
167. Simon, M., et al., *Persistence of both peripheral and non-peripheral corneodesmosomes in the upper stratum corneum of winter xerosis skin versus only peripheral in normal skin.* J Invest Dermatol, 2001. **116**(1): p. 23-30.
168. Goldschmidt, H. and A.M. Kligman, *Desquamation of the human horny layer.* Arch Dermatol, 1967. **95**(6): p. 583-6.
169. Milstone, L.M., J. McGuire, and J.F. LaVigne, *Retinoic acid causes premature desquamation of cells from confluent cultures of stratified squamous epithelia.* J Invest Dermatol, 1982. **79**(4): p. 253-60.
170. Allen, T.D. and C.S. Potten, *Desmosomal form, fate, and function in mammalian epidermis.* J Ultrastruct Res, 1975. **51**(1): p. 94-105.
171. Walsh, A. and S.J. Chapman, *Sugars protect desmosome and corneosome glycoproteins from proteolysis.* Arch Dermatol Res, 1991. **283**(3): p. 174-9.
172. Lundstrom, A. and T. Egelrud, *Evidence that cell shedding from plantar stratum corneum in vitro involves endogenous proteolysis of the desmosomal protein desmoglein I.* J Invest Dermatol, 1990. **94**(2): p. 216-20.
173. Suzuki, Y., et al., *The role of two endogenous proteases of the stratum corneum in degradation of desmoglein-1 and their reduced activity in the skin of ichthyotic patients.* Br J Dermatol, 1996. **134**(3): p. 460-4.
174. King, I.A., et al., *Expression of distinct desmocollin isoforms in human epidermis.* J Invest Dermatol, 1993. **100**(4): p. 373-9.
175. Bernard, D., et al., *Analysis of proteins with caseinolytic activity in a human stratum corneum extract revealed a yet unidentified cysteine protease and identified the so-called "stratum corneum thiol protease" as cathepsin l2.* J Invest Dermatol, 2003. **120**(4): p. 592-600.
176. Egelrud, T. and A. Lundstrom, *A chymotrypsin-like proteinase that may be involved in desquamation in plantar stratum corneum.* Arch Dermatol Res, 1991. **283**(2): p. 108-12.
177. Brattsand, M. and T. Egelrud, *Purification, molecular cloning, and expression of a human stratum corneum trypsin-like serine protease with possible function in desquamation.* J Biol Chem, 1999. **274**(42): p. 30033-40.
178. Horikoshi, T., et al., *Role of endogenous cathepsin D-like and chymotrypsin-like proteolysis in human epidermal desquamation.* Br J Dermatol, 1999. **141**(3): p. 453-9.
179. Sato, J., et al., *Cholesterol sulfate inhibits proteases that are involved in desquamation of stratum corneum.* J Invest Dermatol, 1998. **111**(2): p. 189-93.
180. List, K., et al., *Loss of proteolytically processed filaggrin caused by epidermal deletion of Matriptase/MT-SP1.* J Cell Biol, 2003. **163**(4): p. 901-10.
181. Eissa, A. and E.P. Diamandis, *Human tissue kallikreins as promiscuous modulators of homeostatic skin barrier functions.* Biol Chem, 2008. **389**(6): p. 669-80.

182. Lundstrom, A. and T. Egelrud, *Cell shedding from human plantar skin in vitro: evidence that two different types of protein structures are degraded by a chymotrypsin-like enzyme.* Arch Dermatol Res, 1990. **282**(4): p. 234-7.
183. Watkinson, A. and C. Smith. *The role of pro-SCCE and SCCE in desquamation*. in *Proceedings of the 21st IFSCC Congress*. 2000.
184. Caubet, C., et al., *Degradation of corneodesmosome proteins by two serine proteases of the kallikrein family, SCTE/KLK5/hK5 and SCCE/KLK7/hK7.* J Invest Dermatol, 2004. **122**(5): p. 1235-44.
185. Brattsand, M., et al., *A proteolytic cascade of kallikreins in the stratum corneum.* J Invest Dermatol, 2005. **124**(1): p. 198-203.
186. Rawlings, A., et al., *The effect of glycerol and humidity on desmosome degradation in stratum corneum.* Arch Dermatol Res, 1995. **287**(5): p. 457-64.
187. Horikoshi, T., et al., *Isoforms of cathepsin D and human epidermal differentiation.* Biochimie, 1998. **80**(7): p. 605-12.
188. Igarashi, S., et al., *Cathepsin D, but not cathepsin E, degrades desmosomes during epidermal desquamation.* Br J Dermatol, 2004. **151**(2): p. 355-61.
189. Watkinson, A., *Stratum corneum thiol protease (SCTP): a novel cysteine protease of late epidermal differentiation.* Arch Dermatol Res, 1999. **291**(5): p. 260-8.
190. Komatsu, N., et al., *Elevated stratum corneum hydrolytic activity in Netherton syndrome suggests an inhibitory regulation of desquamation by SPINK5-derived peptides.* J Invest Dermatol, 2002. **118**(3): p. 436-43.
191. Descargues, P., et al., *Spink5-deficient mice mimic Netherton syndrome through degradation of desmoglein 1 by epidermal protease hyperactivity.* Nat Genet, 2005. **37**(1): p. 56-65.
192. Guerrin, M., et al., *Expression cloning of human corneodesmosin proves its identity with the product of the S gene and allows improved characterization of its processing during keratinocyte differentiation.* J Biol Chem, 1998. **273**(35): p. 22640-7.
193. Emmett, M.R. and R.M. Caprioli, *Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins.* Journal of the American Society for Mass Spectrometry, 1994. **5**(7): p. 605-613.
194. Mihailova, A., E. Lundanes, and T. Greibrokk, *Determination and removal of impurities in 2-D LC-MS of peptides.* J Sep Sci, 2006. **29**(4): p. 576-81.
195. Strasser, G.R. and I. Varadi, *Investigations of artifact peaks in sensitive high-performance liquid chromatography methods.* J Chromatogr A, 2000. **869**(1-2): p. 85-90.
196. Williams, S., *Ghost peaks in reversed-phase gradient HPLC: a review and update.* J Chromatogr A, 2004. **1052**(1-2): p. 1-11.
197. Verge, K.M. and G.R. Agnes, *Plasticizer contamination from vacuum system O-rings in a quadrupole ion trap mass spectrometer.* J Am Soc Mass Spectrom, 2002. **13**(8): p. 901-5.
198. Schlosser, A. and R. Volkmer-Engert, *Volatile polydimethylcyclosiloxanes in the ambient laboratory air identified as source of extreme background signals in nanoelectrospray mass spectrometry.* J Mass Spectrom, 2003. **38**(5): p. 523-5.
199. Zhang, X.K., R.C. Dutky, and H.M. Fales, *Rubber stoppers as sources of contaminants in electrospray analysis of peptides and proteins.* Anal Chem, 1996. **68**(18): p. 3288-9.
200. Olsen, J.V., et al., *Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap.* Mol Cell Proteomics, 2005. **4**(12): p. 2010-21.

201. Fenn, J.B., J. Rosell, and C.K. Meng, *In electrospray ionization, how much pull does an ion need to escape its droplet prison?* Journal of the American Society for Mass Spectrometry, 1997. **8**(11): p. 1147-1157.
202. Hesse, A.M., et al., *Simple and universal tool to remove on-line impurities in mono- or two-dimensional liquid chromatography-mass spectrometry analysis.* J Chromatogr A, 2008. **1189**(1-2): p. 175-82.
203. McCarthy, F.M., et al., *Differential detergent fractionation for non-electrophoretic eukaryote cell proteomics.* J Proteome Res, 2005. **4**(2): p. 316-24.
204. Liu, H., et al., *Development of an online two-dimensional nano-scale liquid chromatography/mass spectrometry method for improved chromatographic performance and hydrophobic peptide recovery.* J Chromatogr A, 2006. **1135**(1): p. 43-51.
205. Horie, K., et al., *Calculating optimal modulation periods to maximize the peak capacity in two-dimensional HPLC.* Anal Chem, 2007. **79**(10): p. 3764-70.
206. Davis, J.M., D.R. Stoll, and P.W. Carr, *Effect of first-dimension undersampling on effective peak capacity in comprehensive two-dimensional separations.* Anal Chem, 2008. **80**(2): p. 461-73.
207. Liu, L.P. and C.M. Deber, *Guidelines for membrane protein engineering derived from de novo designed model peptides.* Biopolymers, 1998. **47**(1): p. 41-62.
208. Stapels, M.D. and D.F. Barofsky, *Complementary use of MALDI and ESI for the HPLC-MS/MS analysis of DNA-binding proteins.* Anal Chem, 2004. **76**(18): p. 5423-30.
209. Olsen, J.V., S.E. Ong, and M. Mann, *Trypsin cleaves exclusively C-terminal to arginine and lysine residues.* Mol Cell Proteomics, 2004. **3**(6): p. 608-14.
210. Shabir, G.A., *Validation of high-performance liquid chromatography methods for pharmaceutical analysis. Understanding the differences and similarities between validation requirements of the US Food and Drug Administration, the US Pharmacopeia and the International Conference on Harmonization.* J Chromatogr A, 2003. **987**(1-2): p. 57-66.
211. Creasy, D.M. and J.S. Cottrell, *Error tolerant searching of uninterpreted tandem mass spectrometry data.* Proteomics, 2002. **2**(10): p. 1426-34.
212. Mann, M. and O.N. Jensen, *Proteomic analysis of post-translational modifications.* Nat Biotechnol, 2003. **21**(3): p. 255-61.
213. Fricker, L.D., et al., *Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues.* Mass Spectrom Rev, 2006. **25**(2): p. 327-44.
214. Baggerman, G., et al., *Peptidomics.* J Chromatogr B Analyt Technol Biomed Life Sci, 2004. **803**(1): p. 3-16.
215. Schoofs, L., et al., *Peptides in the locusts, Locusta migratoria and Schistocerca gregaria.* Peptides, 1997. **18**(1): p. 145-56.
216. Dias, J.A., *Is there any physiological role for gonadotrophin oligosaccharide heterogeneity in humans? II. A biochemical point of view.* Hum Reprod, 2001. **16**(5): p. 825-30.
217. Polevoda, B. and F. Sherman, *N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins.* J Mol Biol, 2003. **325**(4): p. 595-622.
218. Torfs, P., et al., *Isolation, identification, and synthesis of a disulfated sulfakinin from the central nervous system of an arthropods the white shrimp Litopenaeus vannamei.* Biochem Biophys Res Commun, 2002. **299**(2): p. 312-20.

219. Rieg, S., et al., *Dermcidin is constitutively produced by eccrine sweat glands and is not induced in epidermal cells under inflammatory skin conditions.* Br J Dermatol, 2004. **151**(3): p. 534-9.
220. Schittek, B., et al., *Dermcidin: a novel human antibiotic peptide secreted by sweat glands.* Nat Immunol, 2001. **2**(12): p. 1133-7.
221. Jiang, X., M. Ye, and H. Zou, *Technologies and methods for sample pretreatment in efficient proteome and peptidome analysis.* Proteomics, 2008. **8**(4): p. 686-705.
222. Schrader, M. and H. Selle, *The process chain for peptidomic biomarker discovery.* Dis Markers, 2006. **22**(1-2): p. 27-37.
223. Zheng, X., H. Baker, and W.S. Hancock, *Analysis of the low molecular weight serum peptidome using ultrafiltration and a hybrid ion trap-Fourier transform mass spectrometer.* J Chromatogr A, 2006. **1120**(1-2): p. 173-84.
224. Denecker, G., et al., *Caspase-14 reveals its secrets.* J Cell Biol, 2008. **180**(3): p. 451-8.
225. Lippens, S., et al., *Caspase-14 is expressed in the epidermis, the choroid plexus, the retinal pigment epithelium and thymic Hassall's bodies.* Cell Death Differ, 2003. **10**(2): p. 257-9.
226. Alibardi, L., E. Tschachler, and L. Eckhart, *Distribution of caspase-14 in epidermis and hair follicles is evolutionarily conserved among mammals.* Anat Rec A Discov Mol Cell Evol Biol, 2005. **286**(2): p. 962-73.
227. Chien, A.J., R.B. Presland, and M.K. Kuechle, *Processing of native caspase-14 occurs at an atypical cleavage site in normal epidermal differentiation.* Biochem Biophys Res Commun, 2002. **296**(4): p. 911-7.
228. Shibuya, K., et al., *Human cystatin A is inactivated by engineered truncation. The NH2-terminal region of the cysteine proteinase inhibitor is essential for expression of its inhibitory activity.* Biochemistry, 1995. **34**(38): p. 12185-92.
229. Egelrud, T. and A. Lundstrom, *Immunochemical analysis of the distribution of the desmosomal protein desmoglein I in different layers of plantar epidermis.* Acta Derm Venereol, 1989. **69**(6): p. 470-6.
230. Pang, J.X., et al., *Biomarker discovery in urine by proteomics.* J Proteome Res, 2002. **1**(2): p. 161-9.
231. Simon, M., et al., *Evidence that filaggrin is a component of cornified cell envelopes in human plantar epidermis.* Biochem J, 1996. **317 (Pt 1)**: p. 173-7.
232. Kamata, Y., et al., *Neutral cysteine protease bleomycin hydrolase is essential for the breakdown of deiminated filaggrin into amino acids.* J Biol Chem, 2009.
233. Toulza, E., et al., *The human dermokine gene: description of novel isoforms with different tissue-specific expression and subcellular location.* J Invest Dermatol, 2006. **126**(2): p. 503-6.
234. Syed, S.E., et al., *Molecular interactions between desmosomal cadherins.* Biochem J, 2002. **362**(Pt 2): p. 317-27.
235. Nose, A., A. Nagafuchi, and M. Takeichi, *Expressed recombinant cadherins mediate cell sorting in model systems.* Cell, 1988. **54**(7): p. 993-1001.
236. Boggon, T.J., et al., *C-cadherin ectodomain structure and implications for cell adhesion mechanisms.* Science, 2002. **296**(5571): p. 1308-13.
237. Garrod, D. and M. Chidgey, *Desmosome structure, composition and function.* Biochim Biophys Acta, 2008. **1778**(3): p. 572-87.
238. Troyanovsky, R.B., N.A. Chitaev, and S.M. Troyanovsky, *Cadherin binding sites of plakoglobin: localization, specificity and role in targeting to adhering junctions.* J Cell Sci, 1996. **109 (Pt 13)**: p. 3069-78.

239. Kowalczyk, A.P., et al., *The amino-terminal domain of desmoplakin binds to plakoglobin and clusters desmosomal cadherin-plakoglobin complexes.* J Cell Biol, 1997. **139**(3): p. 773-84.
240. Yamasaki, K. and R.L. Gallo, *Antimicrobial peptides in human skin disease.* Eur J Dermatol, 2008. **18**(1): p. 11-21.
241. Baechle, D., et al., *Cathepsin D is present in human eccrine sweat and involved in the postsecretory processing of the antimicrobial peptide DCD-1L.* J Biol Chem, 2006. **281**(9): p. 5406-15.
242. Rieg, S., et al., *Generation of multiple stable dermcidin-derived antimicrobial peptides in sweat of different body sites.* J Invest Dermatol, 2006. **126**(2): p. 354-65.
243. Flad, T., et al., *Detection of dermcidin-derived peptides in sweat by ProteinChip technology.* J Immunol Methods, 2002. **270**(1): p. 53-62.
244. Steffen, H., et al., *Naturally processed dermcidin-derived peptides do not permeabilize bacterial membranes and kill microorganisms irrespective of their charge.* Antimicrob Agents Chemother, 2006. **50**(8): p. 2608-20.
245. Madsen, P., et al., *Molecular cloning, occurrence, and expression of a novel partially secreted protein "psoriasin" that is highly up-regulated in psoriatic skin.* J Invest Dermatol, 1991. **97**(4): p. 701-12.
246. Glaser, R., et al., *Antimicrobial psoriasin (S100A7) protects human skin from Escherichia coli infection.* Nat Immunol, 2005. **6**(1): p. 57-64.
247. Lee, K.C. and R.L. Eckert, *S100A7 (Psoriasin)--mechanism of antibacterial action in wounds.* J Invest Dermatol, 2007. **127**(4): p. 945-57.
248. Schroder, J.M. and J. Harder, *Antimicrobial skin peptides and proteins.* Cell Mol Life Sci, 2006. **63**(4): p. 469-86.
249. Venencie, P.Y., et al., *Increased expression of gelatinases A and B by skin explants from patients with anetoderma.* Br J Dermatol, 1997. **137**(4): p. 517-25.