THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*
Discipline : Biochimie

Présentée et soutenue par **Violette GAUTIER**
Le 18 décembre 2012

Développement de méthodes quantitatives sans marquage pour l'étude protéomique des cellules endothéliales

---

JURY

Dr. Philippe Marin, IGF/CNRS (Montpellier) – Rapporteur
Dr Thierry Rabilloud, iRTSV/LCBM/CNRS (Grenoble) - Rapporteur
Dr Sarah Cianferani-Sanglier, IPHC/CNRS (Strasbourg) - Examinatrice
Dr. Bernard Monsarrat, IPBS/CNRS (Toulouse) - Directeur de thèse
Dr. Anne Gonzalez de Peredo, IPBS/CNRS (Toulouse) – Co-directrice de thèse
Pr. Jean-François Arnal, Université Paul Sabatier (Toulouse) - Président

---

Ecole doctorale : *Biologie - Santé - Biotechnologies*
Unité de recherche : *Institut de Pharmacologie et de Biologie Structurale (UMR5089, CNRS)*
Directeurs de Thèse : *Bernard Monsarrat, Anne Gonzalez de Peredo*
Rapporteurs : *Philippe Marin, Thierry Rabilloud*

RESUME

La compréhension du fonctionnement des systèmes biologiques, dont les protéines sont les principaux effecteurs, est un défi majeur en biologie. La protéomique est aujourd'hui l'outil incontournable pour l'étude des protéines. Au cours de ma thèse, j'ai donc utilisé différentes approches protéomiques pour répondre à plusieurs questions biologiques autour des cellules endothéliales, concernant l'étude de mécanismes fonctionnels de protéines d'intérêt ainsi que des processus inflammatoires au sein de ces cellules. Ces différentes études ont nécessité la mise en place et l'optimisation de méthodes de quantification sans marquage (« label free ») essentielles à la fois pour la caractérisation de complexes protéiques et pour l'analyse de protéomes entiers. Cette thèse décrit ainsi dans un premier temps l'utilisation de telles approches pour l'analyse de complexes immunopurifiés dans laquelle un enjeu important consiste souvent à discriminer de façon non ambiguë les composants *bona fide* du complexe par rapport aux contaminants non-spécifiques. J'ai ainsi notamment pu identifier certains partenaires spécifiques d'une nouvelle famille de facteurs de transcription humains, les protéines THAP, qui jouent un rôle clé dans la prolifération des cellules endothéliales. Dans un second temps, les processus activés par les cellules endothéliales en condition inflammatoire ont été étudiés au niveau de sous-protéomes ou à l'échelle de protéomes entiers, faisant appel à des méthodes de protéomique globale associées à des stratégies de quantification sans marquage. Le glycoprotéome des cellules endothéliales a ainsi d'une part été étudié lors la réponse inflammatoire, grâce à la mise en place d'une méthode d'enrichissement du protéome de surface des cellules. D'autre part, une analyse du protéome entier de ces cellules et de ses modulations lors de la stimulation par des cytokines pro-inflammatoires a également été réalisée. De façon à obtenir une couverture du protéome la plus profonde possible, cette étude a nécessité la mise en place d'une stratégie quantitative impliquant un fractionnement de l'échantillon sur gel 1D. Enfin, une troisième partie s'intéresse plus spécifiquement aux rôles et aux mécanismes d'action de l'interleukine-33 au sein des cellules endothéliales et a requis l'utilisation des méthodes quantitatives précédemment optimisées.

ABSTRACT

Understanding biological systems, in which proteins are the main effectors, is a major challenge in biology. Proteomics is now an indispensable tool for the study of proteins. During my PhD, I used different proteomic approaches to address several biological questions about endothelial cells for the study of functional mechanisms of proteins of interest as well as inflammatory processes in these cells. These studies involved the development and the optimization of label-free quantitative methods, essential both for the characterization of protein complexes and for the analysis total proteome. This thesis describes first the use of such approaches for the analysis of immunopurified complexes, for which an important issue is often to discriminate unambiguously *bona fide* components of the complex from non-specific proteins. I could identify specific partners of a new family of human transcription factors, the THAP proteins, which play a key role in endothelial cells proliferation. Then, the processes activated in endothelial cells under inflammatory condition were studied at sub-proteome or entire proteome level, using global proteomics strategies associated with label-free quantification. On the one hand, the glycoproteome has been studied under inflammatory conditions, through the establishment of a method for cell surface proteome enrichment. On the other hand, a whole proteome analysis of these cells was performed after stimulation with pro-inflammatory cytokines. To obtain deep proteome coverage, this study required the implementation of a quantitative strategy involving sample fractionation by 1D gel. Finally, the third section focuses specifically on the roles and mechanisms of action of interleukin-33 in endothelial cells, and required the use of quantitative methods previously optimized.

REMERCIEMENTS

Ma thèse se termine, j'en profite donc pour remercier toutes les personnes avec lesquelles j'ai travaillé ou que j'ai côtoyées durant ces 4 années, et qui ont participé à rendre ces années très enrichissantes et passionnantes.

Pour commencer, je souhaite remercier l'ensemble des membres du jury, le président Jean-François Arnal, les rapporteurs Philippe Marin et Thierry Rabilloud et également Sarah Cianferani-Sanglier pour avoir accepté d'évaluer mon travail de thèse et pour l'intérêt qu'ils y ont porté.

Bernard Monsarrat et Odile Schiltz, je vous remercie sincèrement de m'avoir permis de réaliser ma thèse dans votre équipe et de participer à différents projets. Merci aussi de votre soutien.

Je tiens ensuite tout particulièrement à remercier Anne Gonzalez. Un grand merci pour ta disponibilité, ton soutien, ta grande implication, tes conseils toujours très avisés tout au long de ma thèse. Ils m'ont permis de garder espoir malgré les difficultés rencontrées parfois, notamment avec les protéines THAP. Je suis ravie d'avoir réalisé ma thèse sous ta direction et celle de Bernard, et au final, d'avoir eu l'opportunité de travailler sur des projets très différents. Ces années ont été très intéressantes et très formatrices. J'en garde un très bon souvenir.

Au cours de ma thèse, j'ai eu le plaisir de travailler avec de nombreuses personnes, d'abord au sein de l'équipe. Merci à David pour ta disponibilité et tes solutions à tout, à Emmanuelle pour ton aide et ton soutien dans tout moment, à Florence pour ton initiation à la MRM, ainsi qu'à Nicolas Delcourt. Et un merci tout particulier à Mathilde, merci pour toute ton aide sur différents projets et ta bonne humeur, c'était super de travailler avec toi !

Je remercie également les collaborateurs avec qui mes projets de thèse ont été réalisés et tout d'abord, Jean-Philippe Girard et son équipe, en particulier Raoul Mazars qui m'a co-encadrée avec Anne durant mon Master 2 Recherche, Corinne Cayrol et Emma Lefrançais pour les projets IL-33. Merci aussi à l'équipe d'Ambra Mari et particulièrement à Sophies Mourgues pour les travaux sur TFIIH. Merci à l'équipe de Jean-Claude Florent et à Michel Azoulet de l'Institut Curie à Paris, pour la synthèse des sondes chimiques pour l'analyse des glycoprotéines.

Ça a été un réel plaisir de côtoyer l'ensemble de l'équipe au quotidien pendant ces années. Merci à tous pour vos conseils, les discussions et la bonne ambiance. Un grand merci à Bertrand (j'aurais pas réussi à te battre finalement, « abusé » !!), Karima (« Pouah », surtout à toi !!), Luc (toi par contre, j'ai réussi à te battre ! «Vachement » facile !), Marlène (Surtout à toi aussi « Charlie » !!) et Mathilde (« Qui ?? ») pour tous les très bons moments qu'on a passés, pour les nombreux fous-rire, et les très nombreux goûters, les grands moments de créativité et aussi pour vos conseils et votre aide. Merci également aux autres sportifs, Renaud (Merci pour les multiples installations d'Endnote !!) et Marc, à Alex, Chrystelle, Thomas, Roxana, Carine et Carole pour les discussions et les pauses, à Manue et Marie-Pierre pour votre gentillesse, vos encouragements et pour les cours à la fac, ainsi qu'à Esthelle, Laure, Nicolas, Stéphane, Sandrine, Michel, François. Merci également à Emilie, Fabien, Lucie et en particulier à Nawel pour les très bons moments passés et leur soutien.

Enfin un grand merci à toute ma famille et à tous mes amis pour leurs encouragements et leur soutien sans faille !

SOMMAIRE

LISTE DES PRINCIPALES ABREVIATIONS **1**

INTRODUCTION **3**

PARTIE I. L'ANALYSE PROTEOMIQUE NANOLC-MS/MS PAR SPECTROMETRIE DE MASSE 5

I. Principe général de la spectrométrie de masse *5*

II. Stratégie générale de l'analyse protéomique nanoLC-MS/MS *6*

II-1. Préparation de l'échantillon peptidique 7

II-2. Fractionnement peptidique par chromatographie liquide 8

II-3. La spectrométrie de masse en tandem pour l'identification des protéines 8

II-4. L'analyse nanoLC-MS/MS 10

III. Analyse bioinformatique des données nanoLC-MS/MS pour l'identification des protéines *11*

III-1. Interprétation des données MS/MS pour l'identification des peptides 11

III-2. Validation des résultats d'identification peptidiques 13

III-3. Des peptides aux protéines 14

IV. L'analyse quantitative des données LC-MS/MS *16*

IV-1. Méthodes de protéomique quantitative 16

IV-2. Analyse bioinformatique des données protéomiques quantitatives 24

PARTIE II. ETUDE DE COMPLEXES PROTEIQUES 27

I. Méthode générale pour l'analyse de complexes protéiques et défis associés *27*

II. Isolement des complexes protéiques *29*

II-1. Lyse cellulaire et extraction des protéines 29

II-2. Enrichissement des complexes protéiques 31

III. Identification des partenaires protéiques bona fide *35*

III-1. Constitution et utilisation de banques de contaminants 36

III-2. Comparaison avec un échantillon contrôle 36

PARTIE III. L'ANALYSE DE PROTEOMES COMPLEXES 43

I. Stratégie générale pour l'étude de mélanges protéiques complexes et défis associés *43*

II. Complexité des mélanges protéiques analysés *43*

III. Amélioration de la séparation LC pour une meilleure couverture du protéome *45*

IV. Simplification des mélanges protéiques complexes pour une meilleure couverture du protéome *47*

IV-1. Fractionnement protéique et peptidique 47

IV-2. Déplétion et « égalisation » 49

IV-3. Enrichissement de sous-protéomes d'intérêt 49

V. Quantification des mélanges protéiques complexes et couverture du protéome *50*

PRESENTATION GENERALE DES TRAVAUX **51**

RESULTATS **53**

PARTIE I. ETUDE DE COMPLEXES PROTEIQUES : IDENTIFICATION DE NOUVEAUX PARTENAIRES D'INTERACTION 55

I. Identification des partenaires protéiques des protéines THAP humaines *55*

I-1. Contexte biologique 55

I-2. Objectifs et stratégie mise en place 57

I-3. Analyse quantitative « label-free » des complexes protéiques THAP humains 59

I-4. Discussion et conclusion 76

I-5. Article *Mazars et al., JBC, 2010* 81

II. Recherche de partenaires protéiques de TFIIH dans les cellules ES murines et dans les cerveaux de souris *91*

II-1. Contexte biologique 91

II-2. Objectifs et stratégie mise en place 94

II-3. Etude des partenaires protéiques de TFIIH dans les cellules souches embryonnaires murines 96
II-4. Discussion et conclusion 98
II-5. Article en préparation 99
PARTIE II. DEVELOPPEMENT DE METHODES POUR L'ETUDE QUANTITATIVE SANS MARQUAGE DE PROTEOMES 109
I. Développement de stratégies de protéomique quantitative pour l'étude de protéomes de surface 109
I-1. Contexte général : méthodes pour l'analyse du protéome de surface 109
I-2. Analyse quantitative du glycoprotéome de surface des cellules endothéliales en conditions inflammatoires 112
I-3. Etude comparative de différentes sondes de biotinylation 117
II. Développement de stratégies de protéomique quantitative pour l'étude de protéomes entiers 123
II-1. Objectifs 123
II-2. Evaluation de la méthode 123
II-3. Application à l'analyse protéomique de la réponse inflammatoire dans les cellules endothéliales 124
II-4. Article *Gautier et al., MCP, 2012* 125
PARTIE III. ETUDE DU ROLE DE L'INTERLEUKINE-33 DANS LES CELLULES ENDOTHELIALES PAR PROTEOMIQUE FONCTIONNELLE 147
I. Contexte biologique 147
II. Caractérisation fonctionnelle de l'IL-33 dans les cellules endothéliales 149
II-1. Etude du rôle intracellulaire de l'IL-33 endogène 149
II-2. Etude du rôle extracellulaire de l'IL-33 152
III. Discussion et conclusion 173
CONCLUSION GENERALE ET PERSPECTIVES 175
MATERIEL ET METHODES 181
I. Préparation et analyse protéomique des complexes THAP 181
II. Préparation et analyse protéomique des complexes TFIIH 184
III. Enrichissement des glycoprotéines de surface et analyse protéomique 184
IV. Préparation et analyse protéomique quantitative des protéomes entiers des cellules endothéliales 185
LISTE DES PUBLICATIONS 187
BIBLIOGRAPHIE 189
ANNEXES 203

LISTE DES PRINCIPALES ABREVIATIONS

AP-MS	Affinity Purification coupled with Mass Spectrometry
CAK	CDK-Activating Kinase
CID	Collision Induced Dissociation
DDA	Data Dependent Acquisition
ESI	Electro Spray Ionization
FRAP	Fluorescence Recovery After Photobleaching
GFP	Green Fluorescent Protein
GGR	Global Genome Repair
HCF-1	Host Cell Factor-1
HEV	High Endothelial Veinules
HPLC	High Performance Liquid Chromatography
ICAT	Isotope-Coded Affinity Tag
IFNγ	InterFeroN gamma
IL-1β	Interleukine-1 beta
IL-33	Interleukine-33
iTRAQ	Isobaric tag for relative and absolute quantitation
LC	Liquid Chromatrography
LC-MS	Liquid Chromatrography coupled to Mass Spectrometry
LEC	Little Elongation Complex
MALDI	Matrix Assisted Laser Desorption Ionization
MAP-SILAC	Mixing After Purification - Stable Isotope Labeling by Amino acids in Cell culture
MPT	Modifications Post-Traductionnelles
MRM	Multiple Reaction Monitoring
MudPIT	Multidimensional chromatography Peptide Identification Technology
NER	Nucleotide Excision Repair
OGE	Off-Gel isoElectrofocusing
OGT	O-GlcNAc GlycosylTransférase
PAI	Protein Abundance Index
PAM-SILAC	Purification After Mixing - Stable Isotope Labeling by Amino acids in Cell culture
PC	Peak Capacity
pI	Point Isoélectrique
RP-LC	Reverse Phase - Liquid Chromatography
RT	Retention Time
SAX	Strong Anion eXchange
SCX	Strong Cation eXchange
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate - PolyAcrylamide Gel Electrophoresis
SEC	Super Elongation Complex
SILAC	Stable Isotope Labeling by Amino acids in Cell culture

TAP	Tandem Affinity Purification
TCR	Transcription-Coupled Repair
TFP	Taux de Faux-Postitifs
THABS	THAP1 Binding Sequence
THAP	THanatos-Associated Protein
TNFα	Tumor Necrosis Factor alpha
UPLC	Ultra Performance Liquid Chromatography
XIC	eXtracted Ion Chromatogram
YFP	Yellow Fluorescent Protein

INTRODUCTION

La connaissance des mécanismes biologiques complexes et du fonctionnement des cellules est un défi majeur et essentiel en biologie. Dans les cellules, tous les processus biologiques impliquent des protéines qui sont les principaux effecteurs des fonctions biologiques. II est donc important d'étudier le « protéome » qui désigne l'ensemble des protéines exprimées par un génome pour une espèce, un organe, une cellule ou un compartiment cellulaire à un moment défini (Patterson and Aebersold 2003). La protéomique est la technique haut débit ou « -Omique » qui permet l'étude du protéome. Ce terme exprime donc l'ambition d'obtenir une vue globale au niveau protéique en analogie avec ce qui est réalisé au niveau de l'ADN en génomique, et des ARNm en transcriptomique. Alors que la génomique mesure un génotype qui est universel et statique, la transcriptomique et la protéomique étudient des « entités » dynamiques issues de la « lecture » du génome dans une cellule donnée, à un moment précis et dans une condition physiologique particulière. Un même génome peut en effet produire des transcriptomes et des protéomes différents en fonction des étapes du cycle cellulaire ou de la différenciation, de la réponse à des signaux biologiques ou physiques, et de l'état physiopathologique de la cellule (Lottspeich 1999). Entre transcriptome et protéome, un niveau de complexité supplémentaire est ajouté puisque l'expression, l'abondance et la diversité des protéines dépendent des niveaux d'expression des ARNm mais également de la régulation de leur traduction, de leur épissage différentiel, de leur dégradation, de l'ajout de modifications post-traductionnelles des protéines (Gygi, Rochon et al. 1999). Ainsi, la protéomique mesure réellement un phénotype à un moment donné dans une condition donnée. Elle apparait à ce titre la méthode la plus adaptée pour l'étude des systèmes biologiques.

La protéomique consiste donc en l'analyse des protéines produites dans une cellule et s'intéresse principalement aujourd'hui à l'étude des interactions entre ces protéines et la caractérisation de complexes protéiques, à l'analyse des modifications post-traductionnelles des protéines et à l'étude de protéomes complexes et de leurs variations d'expression dans des conditions cellulaires données. Ces études font appel à un ensemble de techniques performantes qui sont aujourd'hui capables d'être appliquées à large échelle. Cela est devenu possible grâce aux évolutions et développements simultanés de différents domaines, impliquant la préparation de l'échantillon (fractionnement, gel d'électrophorèse, chromatographie liquide…), les analyses bioinformatiques (banques de données génomiques et protéiques, conception et amélioration des logiciels de traitement des données brutes protéomiques - interrogation en banque de données, logiciels de quantification…), ainsi que la spectrométrie de masse (amélioration de la sensibilité, de la résolution, de la vitesse de séquençage des spectromètres de masse…).

En dépit de ces avancées, la protéomique fait aujourd'hui encore face à de nombreux enjeux, à la fois au niveau technique, instrumental et de l'analyse bioinformatique des données générées. Au cours de ma thèse, j'ai été confrontée à certains de ces enjeux, en essayant de répondre à des questions biologiques précises, en particulier ceux rencontrés lors de l'étude de complexes protéiques ou encore de l'analyse de protéomes complexes et de leurs réponses à des stimuli donnés. Ils ont alors nécessité la mise en place et l'optimisation de méthodes protéomiques

adaptées. Avant de présenter les résultats obtenus lors de ce doctorat qui vise à mieux comprendre certains mécanismes cellulaires des cellules endothéliales humaines et les stratégies mises en place pour y parvenir, les principes généraux de la protéomique ainsi que son application pour l'étude de complexes protéiques et de protéomes complexes sont abordés.

Partie I. L'analyse protéomique nanoLC-MS/MS par spectrométrie de masse

I. Principe général de la spectrométrie de masse

La spectrométrie de masse est une méthode analytique qui permet de déterminer la masse moléculaire précise d'un composé chimique ou biologique en mesurant son rapport masse sur charge (m/z). Un spectromètre de masse est classiquement composé de trois modules : une source d'ionisation, un analyseur de masse et un détecteur de courant d'ions (Figure 1). Les molécules à analyser sont dans un premier temps ionisées et amenées en phase gazeuse au niveau de la source. Les ions ainsi formés sont ensuite focalisés et transmis à l'analyseur où ils sont séparés en fonction de leur rapport m/z, qui leur confère un comportement différent au sein d'un champ électrostatique ou électromagnétique (Domon and Aebersold 2006). Ils sont enfin collectés par le détecteur qui quantifie leur intensité et amplifie le signal. Ces dernières étapes se déroulent dans un vide poussé afin d'éviter toute collision entre les ions et les molécules de gaz. Après la détection, un système informatique effectue le traitement des données et génère un spectre de masse qui reflète la variation du courant ionique observé en fonction du rapport m/z et permet de déterminer la masse moléculaire des espèces ionisées.

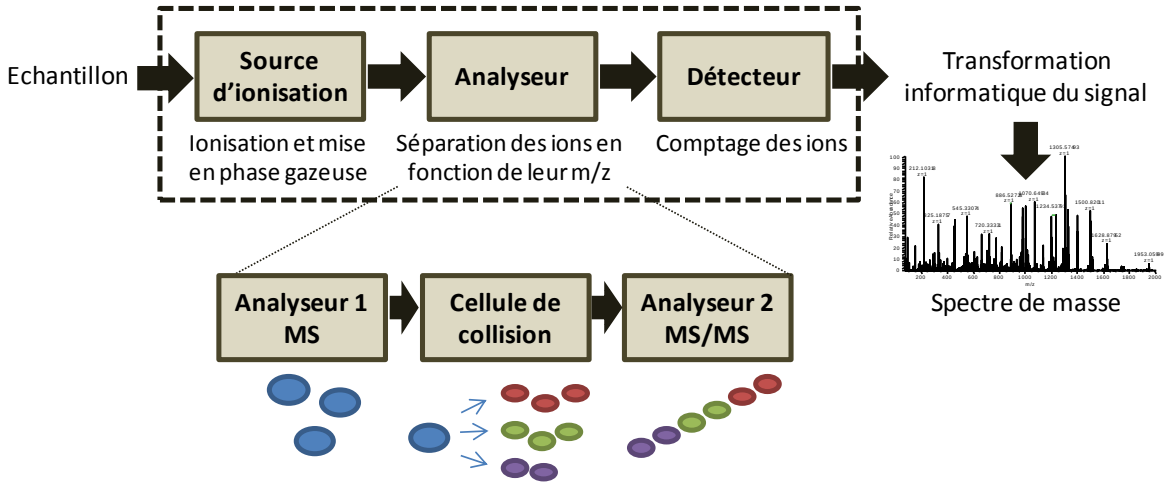


Figure 1 : Principe de fonctionnement d'un spectromètre de masse

La spectrométrie de masse a longtemps été réservée à l'analyse de petits composés thermostables à cause de l'absence de techniques d'ionisation permettant l'ionisation et le transfert de macrobiomolécules intactes en phase gazeuse sans fragmentation excessive. Les protéines et les peptides sont en effet non volatiles et thermiquement instables, et les techniques existantes n'étaient donc pas adaptées pour leur étude. Le développement de deux méthodes d'ionisation douce à la fin des années 1980 - le MALDI (« Matrix Assisted Laser Desorption Ionization ») (Takana,

Waki et al. 1988) et l'électrospray (ou ESI, « Electro Spray Ionization ») (Fenn, Mann et al. 1989)- a donc révolutionné l'analyse des molécules biologiques, en particulier des protéines et des peptides. Ces découvertes, récompensées en 2002 par un prix Nobel de chimie, ont ainsi ouvert la voie à l'analyse protéomique par spectrométrie de masse. Quatre types d'analyseurs sont couramment utilisés : l'analyseur quadripolaire (Q), l'analyseur à temps de vol (TOF), les trappes ioniques (IT) et les analyseurs à transformée de Fourier (FT-ICR, Orbitrap). Les sources, analyseurs et détecteurs existants peuvent être associés de différentes façons et offrent donc une grande variété de géométries d'appareils. Les spectromètres de masse peuvent utiliser un analyseur pour simplement mesurer la masse d'une molécule (spectrométrie de masse conventionnelle) mais peuvent également associer deux analyseurs (appareils hybrides) et ainsi offrir la possibilité de réaliser de la MS/MS pour séquencer les molécules analysées en MS (spectrométrie de masse en tandem). Dans ce cas, après l'analyse MS par le premier analyseur, des ions spécifiques sont fragmentés dans la cellule de collision, et les masses des fragments résultants sont mesurées par le second analyseur (Figure 1).

La spectrométrie de masse a connu un essor considérable dans le domaine de la biologie depuis le développement des sources d'ionisation douce (Patterson and Aebersold 2003). Elle est devenue la méthode de choix pour l'identification à haut débit et la quantification de mélanges protéiques. Cet essor est dû aux développements conjoints de quatre domaines : (1) le développement d'instruments de plus en plus performants notamment au niveau sensibilité, résolution, vitesse de séquençage, (2) le développement de méthodes séparatives des protéines et des peptides, (3) les projets de séquençage de génomes à haut débit qui alimentent les banques de séquences essentielles pour l'identification des protéines ainsi que (4) le développement de logiciels capables d'analyser les données produites. L'étude et la compréhension de phénomènes biologiques nécessitent généralement l'analyse de mélanges protéiques relativement complexes. Grâce à ces évolutions, la protéomique permet aujourd'hui l'identification de plusieurs milliers de protéines au sein d'un échantillon protéique complexe. Elle peut utiliser une séparation des protéines sur gel d'électrophorèse bi-dimentionnelle (gel 2D) suivie de leur analyse MS ou bien consister en une séparation des peptides issus des protéines par chromatographie liquide (LC) en amont de leur analyse MS/MS (approche nanoLC-MS/MS).

II. Stratégie générale de l'analyse protéomique nanoLC-MS/MS

Les échantillons étudiés en protéomique peuvent être de différente nature et représenter un protéome entier (organisme, cellule), un sous-protéome (organite, compartiment sub-cellulaire, phosphoprotéome) ou encore un complexe protéique, mais ils sont toujours constitués de protéines en mélange. Leur complexité est plus ou moins importante. Elle peut varier d'une centaine de protéines s'il s'agit de complexes protéiques purifiés, à plusieurs milliers voire dizaines de milliers de protéines dans le cas de protéomes entiers de mammifères. L'analyse protéomique doit donc être capable d'identifier un maximum de protéines pour couvrir et caractériser au mieux le protéome étudié. Pour cela, l'approche en chromatographie liquide (LC) couplée à la spectrométrie de masse en tandem (ou MS/MS) (Figure 2) représente une méthode largement répandue, dont l'efficacité a été démontrée dans de nombreuses études (Selbach and Mann 2006; de Godoy, Olsen et al. 2008; Hubner, Bird et al. 2010; Thakur, Geiger et al. 2011; Nagaraj, Kulak et al. 2012).

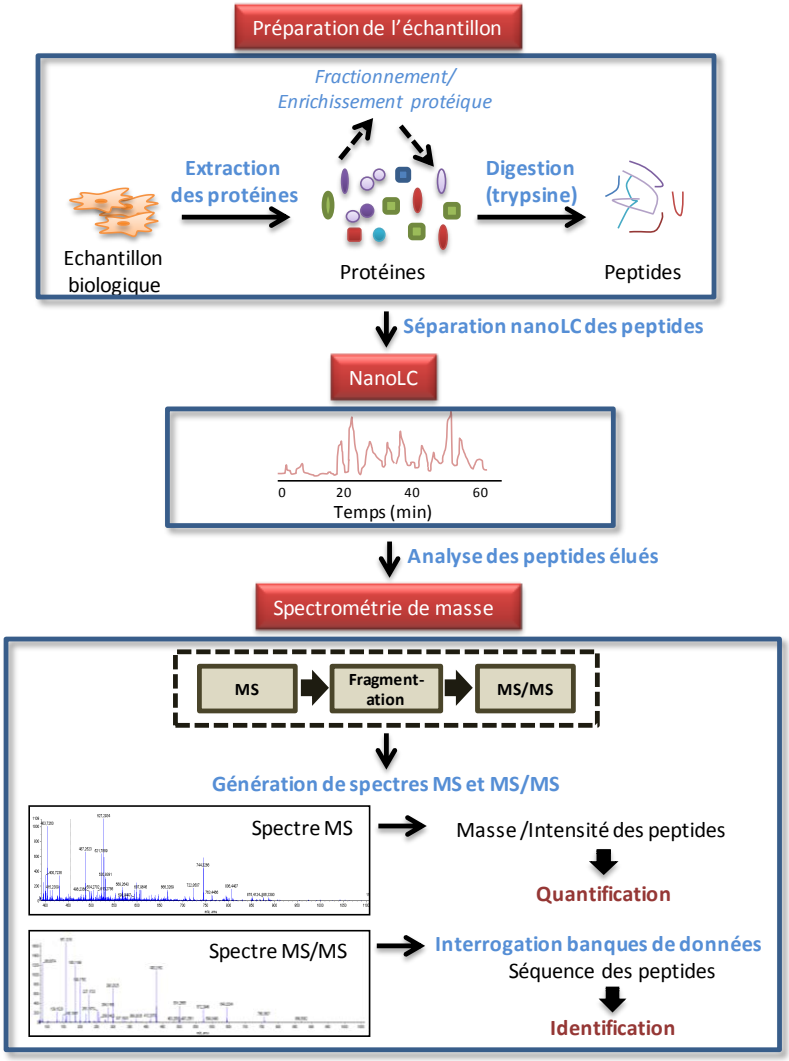


Figure 2 : Stratégie générale de l'analyse nanoLC-MS/MS

Dans cette stratégie, les échantillons sont analysés au niveau peptidique ce qui nécessite une digestion préalable du mélange protéique. Les peptides sont ensuite séparés par chromatographie liquide avant d'être analysés par spectrométrie de masse en tandem. Les informations de séquence ainsi obtenues sont ensuite utilisées pour identifier les peptides et donc les protéines présentes dans l'échantillon.

II-1. Préparation de l'échantillon peptidique

Dans cette approche, l'identification des protéines repose donc sur l'analyse de leurs peptides. Elle est qualifiée d'approche « bottom-up » en opposition à l'approche dite « top-down » qui analyse des protéines entières. La première étape consiste donc à préparer des mélanges peptidiques qui seront par la suite analysés par spectrométrie de masse. Les protéines sont d'abord extraites de l'échantillon biologique, qui peut correspondre à des cellules procaryotes ou eucaryotes, des tissus ou encore des fluides biologiques. Selon la nature de l'échantillon et l'objectif de l'analyse, les protéines extraites peuvent ensuite être fractionnées (par exemple sur gel 1D SDS-PAGE), des protéines spécifiques peuvent être enrichies à partir de l'extrait (analyse de sous-protéomes ou de

complexes protéiques), ou l'extrait total peut être analysé directement. Dans tous les cas, les protéines sont au préalable soumises à une digestion enzymatique qui va générer des fragments peptidiques de taille réduite accessibles à l'analyse par LC-MS/MS. L'enzyme la plus couramment utilisée est la trypsine. C'est une protéase stable qui clive les protéines en peptides de manière très spécifique du côté C-terminal des résidus lysine et arginine (Olsen, Ong et al. 2004), ce qui facilite par la suite l'identification de ces peptides par recherche en bases de données.

L'analyse de peptides trypsiques est techniquement beaucoup plus simple que celle des protéines entières. Cependant, comme chaque protéine digérée par la trypsine produit de nombreux peptides (une cinquantaine en moyenne), la complexité de l'échantillon est considérablement augmentée. Pour faire face à celle-ci, le mélange peptidique est séparé sur une chromatographie liquide (LC) en phase inverse avant d'être injecté dans le spectromètre de masse.

II-2. Fractionnement peptidique par chromatographie liquide

La chromatographie liquide (LC) est la technique de choix pour la séparation de mélanges peptidiques grâce à sa résolution, sa flexibilité et ses possibilités de couplage direct aux spectromètres via la source ESI. Il existe différentes techniques de LC en fonction de la phase stationnaire utilisée, chacune permettant de séparer les analytes en fonction de paramètres physico-chimiques particuliers. La chromatographie liquide classiquement utilisée pour le couplage avec la spectrométrie de masse est la LC en phase inverse qui permet la séparation des peptides en fonction de leur hydrophobicité et de leur taille. La séparation est effectuée grâce à un gradient d'élution qui met en jeu une variation continue de l'hydrophobicité de l'éluant (mélange eau/acétonitrile/acide généralement), compatible avec une infusion directe dans la source électrospray. Au cours de ce gradient chromatographique, les peptides sont retenus dans la colonne pendant un certain temps en fonction de leur hydrophobicité, puis élués à un temps de rétention donné (RT) sous forme d'un pic chromatographique détecté par le spectromètre de masse.

L'utilisation de la LC permet un fractionnement de l'échantillon en amont de l'analyse MS, limitant ainsi la compétition entre les peptides au moment de l'ionisation ainsi qu'au moment de la sélection des ions pour la fragmentation MS/MS. L'effet de suppression d'ion, où le signal d'un ion majoritaire supprime celui d'un autre ion minoritaire, est diminué, ce qui permet d'augmenter la gamme dynamique observable. De plus, le nombre d'ions analysés simultanément par le spectromètre de masse est réduit, permettant, grâce aux appareils à haute vitesse de séquençage actuels, d'identifier un très grand nombre de peptides tout au long du gradient chromatographique.

II-3. La spectrométrie de masse en tandem pour l'identification des protéines

Comme mentionné plus haut, la stratégie nanoLC-MS/MS est décrite comme une approche « bottom-up », et ce sont donc des informations sur les peptides qui sont acquises (masse et séquence) et qui permettent ensuite de remonter à l'identification de chaque protéine parente.

L'apparition des spectromètres de masse hybrides, et donc de la spectrométrie de masse en tandem (MS/MS) a été une révolution pour l'identification des protéines puisqu'elle permet d'obtenir des informations sur la séquence des acides aminés qui composent un peptide. La MS/MS

repose sur un processus de fragmentation d'une ou plusieurs liaisons de la molécule étudiée. Pour cela, il faut transférer aux ions stables produits lors de l'ionisation, l'énergie nécessaire à leur fragmentation. La méthode la plus utilisée permettant ce transfert d'énergie est la dissociation induite par collision (CID, « collision-induced dissociation »). D'une manière générale, dans un premier analyseur, un ion précurseur est sélectionné de façon spécifique en fonction de son rapport masse sur charge (m/z) et transféré dans une cellule de collision où il percute un flux d'atomes de gaz inerte (azote, hélium ou argon). Dès lors, son énergie cinétique est transformée en partie en énergie vibrationnelle nécessaire à sa fragmentation. Dans le cas d'un ion peptidique, cette fragmentation intervient principalement au niveau de la liaison amide entre les acides aminés, générant des séries de fragments N-terminaux et C-terminaux (appelés respectivement ions b et y dans la nomenclature conventionnelle) qui sont ensuite analysés par un second analyseur, d'où le terme de spectrométrie de masse en tandem (Figure 3).

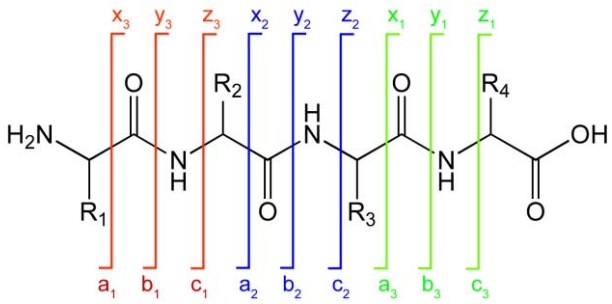


Figure 3 : Nomenclature des différentes séries d'ions peptidiques qui peuvent être générées lors de la fragmentation MS/MS. *Proposée initialement par Roepstorff et Fohlman (Roepstorff and Fohlman 1984), elle fut modifiée par Johnson et al., (Johnson, Martin et al. 1987) avant d'être définitivement adoptée. La molécule représentée en noir correspond à un peptide théorique comportant quatre résidus (R représentant la chaîne latérale). Les fragments ne seront détectés que s'ils portent au moins une charge. Si la charge est retenue du côté N-terminal alors l'ion est dit de série a, b ou c. Si la charge est retenue du côté C terminal alors l'ion est dit de série x, y ou z. Dans le cas d'une fragmentation CID, on observe majoritairement des séries b et y.*

Le balayage des rapports m/z des ions fragments par le détecteur génère un spectre de masse de second niveau, appelé spectre MS/MS ou MS2. C'est l'analyse de ce dernier qui permet de remonter à la structure de l'ion parent. En plus de la mesure de la masse des peptides d'une protéine, la MS/MS apporte ainsi des données de séquence sur ces peptides, qui sont ensuite utilisées par des logiciels dédiés afin de réaliser les recherches en banques de données protéiques. Ces banques de données, comme par exemple SwissProt (http://www.uniprot.org/) regroupent les séquences en acides aminés des protéines de divers organismes. Elles ont été largement alimentées grâce aux récentes campagnes de séquençage du génome de différents organismes, dont l'humain. Connaissant la séquence de ces protéines, il est possible de prédire la liste des peptides attendus par une digestion protéolytique *in silico* de chaque protéine avec une enzyme spécifique telle que la trypsine. La liste de fragments théoriques obtenus après une fragmentation CID peut également être générée pour chaque peptide, ce qui constitue la base du processus d'identification par recherche en base de données (voir ci-dessous).

Cette approche est ainsi adaptée à l'étude de mélanges de protéines. Celles-ci n'ont en effet pas besoin d'être isolées une à une, puisque le séquençage MS/MS d'un peptide individuel est

possible même quand d'autres peptides issus de protéines différentes sont présents dans l'échantillon. Des logiciels adaptés identifient par la suite les protéines du mélange grâce à l'ensemble des spectres MS/MS enregistrés.

II-4. L'analyse nanoLC-MS/MS

L'association de la spectrométrie de masse en tandem avec en amont une séparation chromatographique des peptides est donc une méthode de choix pour l'étude de mélanges protéiques. Le couplage de la LC et de l'ESI-MS se fait « en ligne » grâce à la compatibilité de la source ESI avec des échantillons en solution. La miniaturisation de la LC (nanoLC) et de la source ESI (nanoESI) a encore permis d'améliorer la sensibilité de cette technique et de diminuer la quantité de matériel nécessaire pour l'analyse.

Au cours du gradient chromatographique, les peptides, au fur et à mesure de leur élution, sont ionisés et injectés dans le spectromètre de masse pour être analysés en MS et ainsi générer des spectres MS qui représentent l'empreinte massique de l'ensemble des peptides élués à un instant donné (Figure 4).

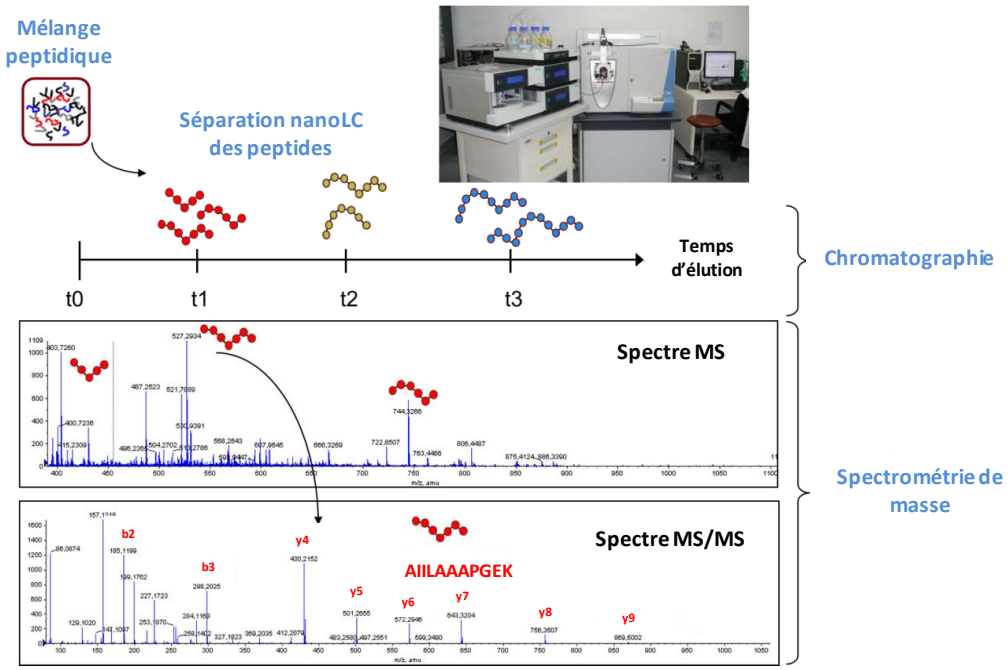


Figure 4 : Vue d'ensemble du couplage nanoLC-MS/MS.

Les ions les plus intenses détectés lors d'un scan MS sont sélectionnés individuellement et fragmentés pour être séquencés. Des spectres MS/MS contenant les différents fragments de l'ion peptidique parent sont alors obtenus. Le passage du mode MS au mode MS/MS est automatiquement réalisé par le logiciel de pilotage du spectromètre de masse, en fonction de critères prédéfinis au départ par l'utilisateur dans la méthode d'acquisition, et grâce à une analyse en temps réel des signaux détectés dans un scan MS à un instant t. Par exemple, le spectromètre peut être configuré de façon à déclencher une analyse MS/MS sur les N ions les plus intenses du scan MS en cours, s'ils dépassent un seuil d'intensité de signal donné et présentent un état de charge

conforme aux critères fixés. Ces ions sont alors successivement sélectionnés, isolés et fragmentés en MS/MS. Une fois ce cycle de MS/MS terminé, un nouveau scan MS est enregistré afin de détecter les peptides en cours d'élution de la colonne, et d'enchaîner un nouveau cycle de MS/MS sur les ions les plus intenses. Sur certains appareils, comme le LTQ-Orbitrap (Olsen, Schwartz et al. 2009), ces cycles de MS et MS/MS sont réalisés en parallèle : pendant l'analyse MS haute résolution dans l'Orbitrap, la trappe linéaire (LTQ), plus rapide, acquiert les spectres MS/MS des peptides les plus intenses qui ont été définis lors d'un « pré-scan » MS rapide de plus faible résolution (Figure 5). Ce mode d'acquisition des données est appelé « DDA » pour « data dependent acquisition ». Des paramètres d'exclusion dynamique peuvent aussi être ajustés afin d'éviter la sélection et l'acquisition répétées de multiples scans MS/MS sur le même ion.

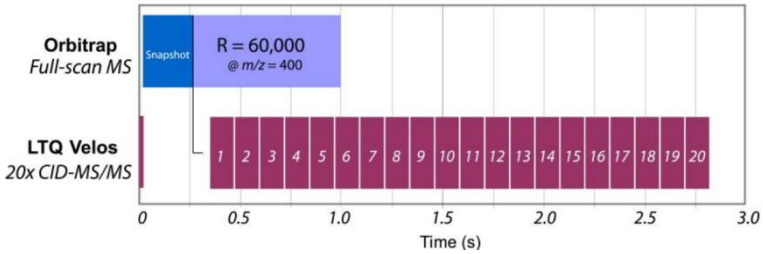


Figure 5 : Représentation schématique d'un cycle MS et MS/MS (Top 20) sur un LTQ-Orbitrap Velos (d'après (Olsen, Schwartz et al. 2009)).

Le couplage nanoLC-MS/MS permet ainsi l'analyse d'un grand nombre de peptides à partir d'un échantillon biologique complexe dans un mode automatisé et à haut débit. Cette approche génère une grande quantité de données MS qui donnent accès aux masses des peptides et à leur intensité, et de données MS/MS dont l'interprétation est à la base de l'identification de l'ensemble des peptides constituant l'échantillon. Au regard du volume considérable de données à traiter, de nombreux moteurs de recherche ont été développés permettant d'automatiser l'identification de peptides et de protéines à partir des spectres MS et MS/MS.

III. Analyse bioinformatique des données nanoLC-MS/MS pour l'identification des protéines

III-1. Interprétation des données MS/MS pour l'identification des peptides

Durant la dernière décennie, les spectromètres de masses en tandem ont été continuellement améliorés au niveau de la sensibilité, de la précision, de la résolution et de la vitesse de séquençage. A titre d'exemple, il est désormais possible de générer plus de 30000 spectres MS/MS au cours d'une analyse LC-MS/MS de 2h sur un appareil de type LTQ-Orbitrap. Les moteurs de recherche dans les banques de séquences protéiques ont ainsi évolué afin de répondre à cette montée en puissance de la spectrométrie de masse en tandem, pour permettre un traitement haut-débit des données MS et MS/MS générées.

Il existe un grand nombre de logiciels disponibles pour réaliser l'identification de peptides à partir de spectres MS/MS en utilisant différentes approches (Nesvizhskii 2010). Certains effectuent une interprétation partielle des spectres MS/MS (« Sequence Tag ») en calculant les différences de masse entre les fragments mesurés pour en extraire une séquence peptidique partielle (« Tag »), en se basant sur le principe qu'une courte série d'acides aminés, associée à la masse du peptide, suffit à l'identification non ambiguë de ce peptide (Mann and Wilm 1994). D'autres logiciels vont quant à eux réaliser un séquençage *de novo*, c'est-à-dire interpréter en totalité les spectres MS/MS en calculant les différences de masse entre chaque fragment mesuré et en extraire ainsi la séquence peptidique. Ils sont particulièrement utiles quand les protéines sont absentes des bases de données, notamment lors d'études réalisées à partir d'organismes dont le génome n'est pas séquencé. La déduction de la séquence peptidique à partir du spectre expérimental permet en effet, dans ces cas-là, d'effectuer des recherches par homologie à partir de banques protéiques constituées sur des organismes parents. Ils nécessitent cependant des spectres MS/MS de bonne qualité ainsi qu'un temps de calcul important et ne sont donc pas utilisés en routine pour les analyses à haut-débit dans des organismes bien caractérisés. Pour ce type d'analyse, d'autres logiciels plus adaptés et plus performants existent. Contrairement aux précédents, ils ne sont pas basés sur une interprétation des spectres MS/MS mais utilisent directement les listes de pics brutes extraites des spectres MS/MS.

De nombreux moteurs de recherche utilisent cette approche, parmi lesquels on peut citer par exemple Mascot (Perkins, Pappin et al. 1999), SEQUEST (Eng, McCormack et al. 1994), OMSSA (Geer, Markey et al. 2004) ou X!Tandem (Craig and Beavis 2004). Ils effectuent une comparaison entre les spectres MS/MS expérimentaux bruts (peaklist) et des spectres MS/MS générés *in silico* à partir de banques de séquences protéiques (Figure 6). La première étape de l'identification consiste en la digestion trypsique théorique de toutes les protéines présentes dans la banque interrogée, grâce à la connaissance de la spécificité de clivage de la trypsine. Le programme recherche dans toutes les séquences peptidiques ainsi obtenues celles qui correspondent au poids moléculaire des ions précurseurs fragmentés, avec une certaine tolérance d'erreur de masse (Steen and Mann 2004). Cette opération restreint l'espace de recherche à un petit nombre de candidats. Ensuite, l'information contenue dans le spectre MS/MS issu de chaque ion permet de trouver la séquence peptidique dont le spectre de fragmentation théorique présente la meilleure corrélation avec le spectre expérimental. En effet, comme le processus de fragmentation CID suit des règles bien particulières, l'algorithme est capable, pour chaque peptide candidat, de construire un modèle représentant la liste de fragments MS/MS théoriques. Ces derniers sont comparés avec les données expérimentales : plus la série d'ions fragments théoriques se rapproche de la série expérimentale, plus l'identification est probable et donc plus le score du peptide est élevé. La séquence du peptide correspondant à la meilleure superposition (meilleur score) est généralement attribuée à l'ion précurseur analysé (interprétation de rang1).

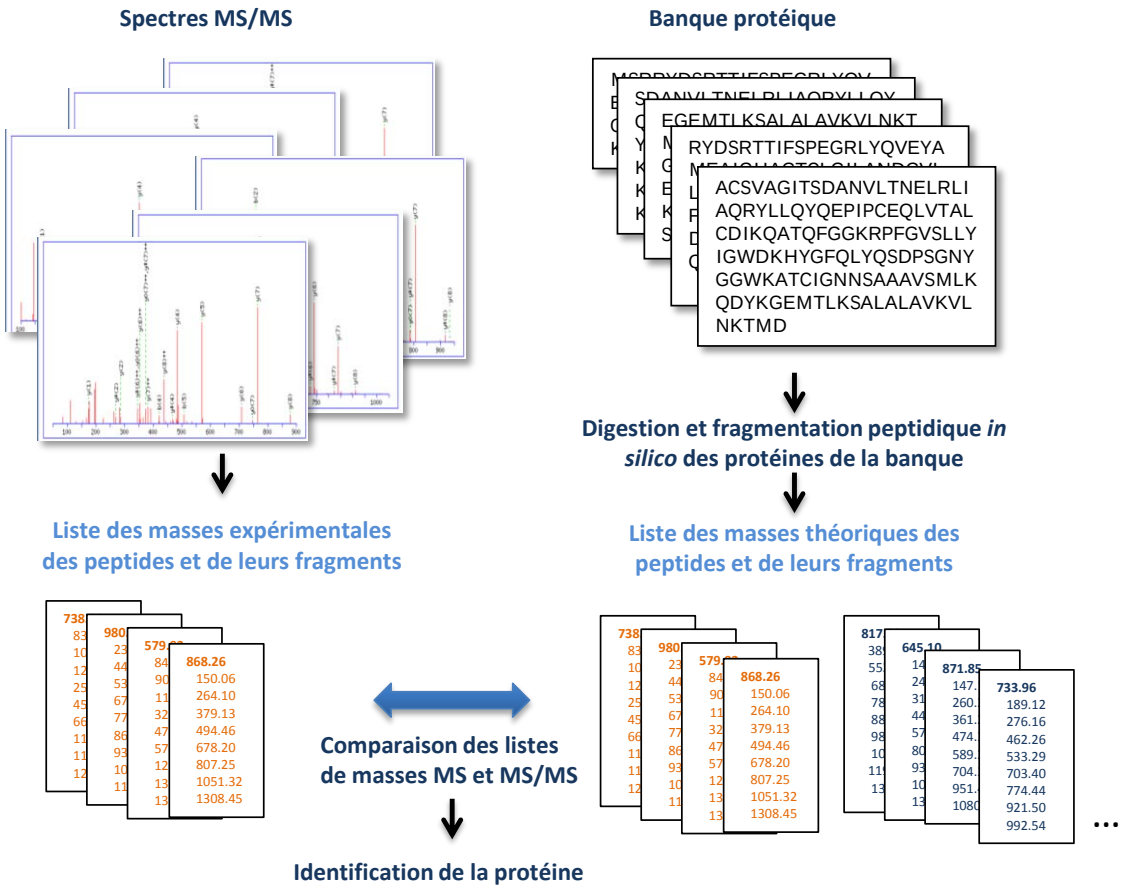


Figure 6 : Principe d'identification des protéines par interrogation de banques protéiques

La principale différence entre les différents logiciels existants réside dans la fonction de corrélation utilisée pour déterminer le degré de similarité entre un spectre MS/MS théorique et un spectre MS/MS expérimental et donc la vraisemblance de l'identification. Ces fonctions sont plus ou moins sophistiquées selon les programmes. Certains fournissent uniquement un score de corrélation ou de similarité peptide-spectre MS/MS. D'autres associent à ce score une valeur de probabilité. C'est le cas de Mascot, le moteur de recherche que j'ai utilisé au cours de mon doctorat. Son algorithme est basé sur une implémentation probabilistique de l'algorithme MOWSE (Perkins, Pappin et al. 1999). Dans Mascot, le score d'une assignation peptidique pour un ion donné est défini par la formule :

$$Ion\ score = -10 \times LOG10\ (P)$$

où P représente la probabilité qu'un assignement donné soit un événement dû au hasard. Une probabilité de 1/10000 donne donc un score de 40. Ainsi, à l'inverse de la p-value, le score est proportionnel à la « véracité » de l'assignement et représente donc une valeur plus intuitive pour l'utilisateur.

III-2. Validation des résultats d'identification peptidiques

Pour tous les spectres MS/MS enregistrés, une correspondance peptidique est généralement trouvée, mais certaines d'entre elles ne sont pas justes, soit parce que le spectre est de mauvaise

qualité et peu informatif, soit parce que la séquence protéique n'est pas présente dans la base de données, soit parce que les paramètres de recherche spécifiés ne sont pas suffisamment larges pour inclure le peptide réellement mesuré comme candidat potentiel. Il convient donc de filtrer et valider les résultats d'identification afin d'exclure les faux-positifs. L'approche la plus simple consiste à fixer une valeur seuil de score pour les peptides identifiés. Dans Mascot, un seuil d'identité ou « identity threshold », propre à chaque identification peptidique, est défini par la formule :

$$Identity\ Threshold = -10 \times LOG10(\frac{p}{N})$$

où p est le seuil de significativité (p-value) et N correspond au nombre de peptides candidats. Ainsi, le seuil d'identité reflète à la fois le niveau de significativité que l'on veut se fixer (p=0.05 par exemple) et également la taille de l'espace de recherche, qui est directement liée aux paramètres de recherche spécifiés : taille de la banque de données, spécificité de l'enzyme, nombre de modifications post-traductionnelles recherchées, valeur de la tolérance d'erreur de masse de l'ion précurseur. Il est important de noter que la variation de la taille de l'espace de recherche modifie la valeur du seuil d'identité mais qu'elle n'a pas d'influence sur la valeur du score Mascot. En fait, le seul paramètre de recherche qui a une influence réelle sur le score peptidique est la tolérance d'erreur de masse appliquée sur les fragments MS/MS car elle affecte le nombre de pics expérimentaux pouvant être corrélés aux données théoriques. En revanche, tous les paramètres spécifiés par l'utilisateur lors de la recherche affectent le seuil de score : plus l'espace de recherche est vaste, plus la probabilité d'une assignation due au hasard augmente, et plus la stringence de la validation augmente. Il est donc important que l'espace de recherche soit correctement défini, et reflète autant que possible la réalité biochimique de l'échantillon (espèce, clivages, modifications potentielles) et les conditions analytiques (tolérance de masse liée à la fois à la précision et la calibration du spectromètre).

Des méthodes de validation ont par ailleurs été développées ces dernières années afin d'estimer et de contrôler les taux de faux positifs (TFP) dans les jeux de données fournis par les moteurs de recherche (Nesvizhskii and Aebersold 2004). La plus populaire est la stratégie « target-decoy » qui permet d'avoir accès à une estimation globale du TFP du jeu de données, après validation des identifications en fonction d'un seuil de score donné. Elle utilise les séquences inversées ou aléatoires (ou « decoy ») des séquences protéiques présentes dans la banque de données cible (ou « target ») (Peng, Elias et al. 2003). Dans les banques inversées, les séquences des entrées sont lues du C-ter vers le N-ter et dans les banques « aléatoires », les acides aminés sont mélangés mais le nombre d'entrées est conservé. Le TFP peut alors être estimé en comparant le nombre d'identifications issues de la banque decoy, dans laquelle on ne s'attend à obtenir que de fausses identifications, avec celui obtenu dans la banque cible.

III-3. Des peptides aux protéines

L'identification des peptides issus de la digestion trypsique des protéines n'est qu'une étape intermédiaire dans la stratégie nanoLC-MS/MS et plus globalement dans les approches bottom-up. L'objectif de ces études est en effet d'identifier les protéines présentes dans l'échantillon étudié. L'étape suivante du processus informatique consiste donc à assembler les séquences peptidiques qui correspondent à une même protéine. Cette étape est complexe puisque plusieurs protéines peuvent

être associées au même ensemble de séquences peptidiques (protéines homologues, isoformes, séquences redondantes de la base de données) et un peptide peut correspondre à plusieurs séquences protéiques. Cette problématique est décrite dans la littérature sous le terme de « protein inference problem» (Nesvizhskii and Aebersold 2005). Elle nécessite des algorithmes spécifiques et dans l'approche la plus couramment utilisée, les moteurs de recherche, dont Mascot, utilisent une méthode dite « parcimonieuse » qui consiste à trouver la liste de protéines la plus petite capable d'expliquer l'ensemble des séquences peptidiques identifiées. Ainsi, à l'issue de la recherche en base de données, l'ensemble des numéros d'accession protéiques associés exactement au même ensemble de peptides (peptides same-set) sont regroupés et restitués dans la liste comme une identification. Les numéros d'accession qui auraient pu être associés uniquement à un sous-ensemble des mêmes peptides sont généralement éliminés (peptide sub-sets).

Pour que ce regroupement soit correctement effectué, il est important que les peptides considérés aient été préalablement validés. En effet, en fonction de la stringence de la validation appliquée sur les identifications peptidiques, la présence de peptides « faux » peut entraîner un gonflement artificiel du nombre de groupes protéiques, identifiés à partir d'ensemble de peptides très similaires mais pas complètement identiques. De plus, la façon dont les groupes de protéines sont constitués est également déterminante lorsqu'il s'agit de comparer des listes de protéines associées à des échantillons différents, ou de regrouper des listes de protéines issues de l'analyse séparée de plusieurs fractions d'un échantillon donné. Par exemple, lors de l'analyse comparée de deux échantillons, l'identification d'ensembles de peptides légèrement différents peut donner lieu à la création de groupes de protéines différents, et conduire à des artéfacts lors d'une comparaison automatique basée sur les numéros d'accession.

Une fois réalisé l'assemblage des peptides en protéines, une étape supplémentaire de validation des groupes de protéines est souvent nécessaire. En effet, dans la mesure où le résultat final est une liste de protéines identifiées, le TFP doit être estimé au niveau protéique. Si l'on décide simplement de conserver toutes les protéines contenant au moins un peptide validé, alors le TFP protéique est généralement supérieur à celui fixé au niveau peptidique. Ceci est lié au fait que les peptides corrects appartenant à une même protéine se regroupent de façon non aléatoire sur les séquences de la banque « target » contrairement aux peptides incorrects qui ont tendance à ne pas être identifiés sur les mêmes séquences protéiques de la banque « decoy ». Ainsi, même si le TFP peptidique est faible, il est possible d'obtenir un TFP protéique élevé si aucun filtrage supplémentaire n'est appliqué sur les protéines. Les résultats présentés dans ce manuscrit sont basés sur des validations dans lesquelles le TFP est contrôlé au niveau protéique, réalisées soit via le module de validation du logiciel MFPaQ (Mascot File Parsing and Quantification, (Bouyssie, Gonzalez de Peredo et al. 2007)), soit plus récemment via le logiciel Prosper, deux programmes développés dans l'équipe. MFPaQ réalise la validation protéique sur la base de schémas de sélections, c'est-à-dire que des règles de validation différentes peuvent être appliquées en fonction du nombre de peptides assignés à la protéine (une protéine est validée si elle comporte au moins P peptides de longueur de séquence supérieure ou égale à L, de score supérieur ou égal à S, et au plus de rang R), ce qui permet de requérir par exemple un score peptidique élevé pour valider les protéines « one-hit », identifiées sur la base d'un seul peptide. Prosper calcule en revanche un score protéique global (apparenté au score protéique Mascot dit « Mudpit »), combinant les valeurs individuelles (écart du score peptidique par rapport à son seuil d'identité) obtenues pour les différents peptides assignés à la protéine, préalablement validés. Dans les deux cas, le TFP au niveau protéique est généralement ajusté à 1%.

IV. L'analyse quantitative des données LC-MS/MS

La spectrométrie de masse permet ainsi l'identification de protéines présentes au sein d'échantillons biologiques complexes. Bien que l'obtention de ces données qualitatives soit essentielle, la simple identification de protéines présentes dans un échantillon n'est pas suffisante pour la compréhension de mécanismes biologiques complexes. Leur caractérisation nécessite en effet souvent l'analyse des variations des systèmes biologiques étudiés après un changement de leur environnement et fait donc appel à des méthodes de protéomique quantitative. L'exploitation des données quantitatives nécessite des logiciels adaptés.

IV-1. Méthodes de protéomique quantitative

La quantification à grande échelle des différences d'expression de protéines est aujourd'hui l'un des principaux enjeux de la protéomique. Ces dernières années, différentes approches de protéomique quantitative ont été développées pour faire face à ce besoin. Elles consistent soit à déterminer la quantité absolue de protéines dans un ou plusieurs échantillons, soit à comparer les quantités relatives de protéines présentes dans différentes conditions (par exemple malade/sain ; avec ou sans traitement…) en observant les variations des signaux associés aux protéines (Ong and Mann 2005).

IV-1.1 Méthodes de quantification absolue – Protéomique ciblée

Les méthodes de quantification absolue ont pour objectif de déterminer les concentrations de protéines dans des échantillons biologiques. La quantification absolue est souvent associée à un mode d'analyse ciblé dit MRM (« Multiple Reaction Monitoring ») dans lequel un nombre limité de peptides d'intérêt sont spécifiquement suivis. Il repose sur la capacité de spectromètres de masse dédiés de type triple quadripôle à agir comme des filtres de masse, en sélectionnant spécifiquement un ion peptidique dans un premier analyseur, puis un ou plusieurs ions fragments générés par la fragmentation de ce dernier dans un second analyseur (Figure 7) (Picotti and Aebersold 2012).

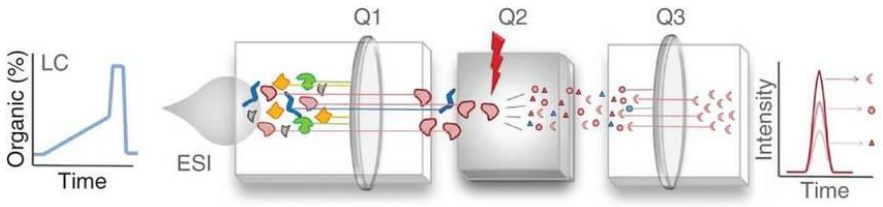


Figure 7 : Stratégie de protéomique ciblée, MRM*. Les ions peptidiques d'intérêt sont spécifiquement sélectionnés par le premier analyseur (Q1) et fragmenté dans Q2. Un ion fragment spécifique de l'ion parent (transition) est ensuite sélectionné par le second analyseur (Q3) et guidé jusqu'au détecteur. Les intensités de ces fragments sont suivies au cours du temps générant un profil MRM. D'après (Picotti and Aebersold 2012).*

Ces filtres successifs permettent de considérablement limiter le bruit de fond des ions et apportent une grande sensibilité à la méthode. La quantification est réalisée sur ces fragments spécifiques en intégrant leur profil chromatographique. La quantification absolue des protéines d'un

échantillon peut être réalisée par MRM grâce à l'addition dans l'échantillon à doser de quantités définies de standards peptidiques marqués avec des isotopes stables, correspondants aux peptides d'intérêt suivis, qui sont quant à eux présents dans l'échantillon sous forme non marquée. Selon la théorie de dilution des isotopes stables, un peptide marqué par un isotope stable est chimiquement identique à son homologue non marqué. Par conséquent, les deux peptides se comportent de manière identique lors de la séparation chromatographique ainsi que pendant l'analyse par spectrométrie de masse (de l'ionisation à la détection). La quantification est réalisée en intégrant et comparant le profil MRM du peptide endogène et celui du peptide standard marqué avec des isotopes lourds.

L'utilisation de standards peptidiques synthétiques marqués correspondants au peptide endogène à doser est la méthode la plus répandue et est appelée méthode « AQUA » (« Absolute QUAntification ») (Gerber, Rush et al. 2003). D'autres types de standards sont cependant également utilisés. La stratégie « QconCAT » (« Quantification CONCATemer ») (Pratt, Simpson et al. 2006) est basée quant à elle sur l'ajout dans l'échantillon à doser d'une protéine artificielle, un concatémère de peptides lourds, composée des différents peptides standard nécessaires à la quantification des protéines. Le standard interne de quantification peut également correspondre à une version alourdie de la protéine à doser comme proposé dans la méthode « PSAQ » (« Protein Standard Absolute Quantification ») (Brun, Dupuis et al. 2007).

Ces stratégies de protéomique ciblée sont sensibles (grâce au gain en gamme dynamique apporté par le double filtre quadripôlaire) et reproductibles (grâce à l'adjonction de standards internes marqués). Elles permettent donc de quantifier des protéines très minoritaires, et sont particulièrement utiles pour la validation d'un certain nombre de protéines candidates sur de grandes séries d'échantillons (Jovanovic, Reiter et al. 2010; Kim, Kim et al. 2010; Narumi, Murakami et al. 2012). Les méthodes MRM associées à la détection de chaque protéine sont cependant assez longues à optimiser, l'approche pouvant souffrir d'un manque de spécificité, notamment dans le cas de matrices complexes. Cette approche est généralement restreinte à l'analyse d'un nombre limité de protéines cibles. Elle n'est donc pas adaptée à l'analyse quantitative sans *a priori* de protéomes complexes.

IV-1.2 Méthodes de quantification relative pour l'analyse protéomique globale

Contrairement aux méthodes de protéomique ciblée, les approches comparatives globales ont pour objectif d'identifier et de comparer l'ensemble d'un protéome, sans hypothèse de départ sur les espèces potentiellement variantes, et utilisent généralement des méthodes de quantification relative (Figure 8).

Une première approche d'analyse protéomique comparative et quantitative est basée sur la séparation des mélanges protéiques à comparer sur gels d'électrophorèse bidimensionnelle (gels 2D). Les variations d'expression des protéines sont identifiées grâce à une différence d'intensité du spot correspondant sur le gel 2D. L'identification des protéines variantes est réalisée dans un second temps par une analyse MS. Cette stratégie présente l'avantage de réaliser la quantification relative avant digestion à la trypsine, sur les formes entières des protéines, et peut permettre par exemple de

repérer plus facilement des variations de modifications post-traductionnelles, affectant la masse et le point isoélectrique des protéines.

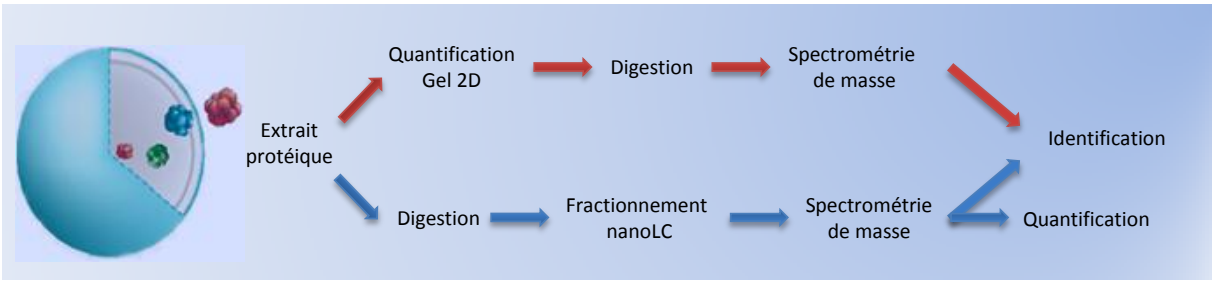


Figure 8 : Méthodes de quantification relative des protéines*. La comparaison d'images de gels d'électrophorèse bi-dimensionnelle et l'analyse par nanoLC-MS/MS sont les deux grandes méthodes qui peuvent être mises en œuvre afin de réaliser une étude protéomique différentielle et quantitative. D'après (Patterson and Aebersold 2003).*

La montée en puissance des approches haut-débit d'identification des protéines basées sur la nanoLC-MS/MS a ouvert depuis quelques années une voie alternative pour la comparaison de protéomes. Ces approches génèrent un volume important de données de structure complexe (jeux de spectres MS et MS/MS acquis cycliquement au cours du gradient chromatographique), et la bioinformatique a pendant plusieurs années représenté un verrou pour l'exploitation de ces données et l'optimisation des méthodes comparatives. Initialement, l'analyse différentielle était principalement basée sur la comparaison de listes de protéines identifiées dans différents échantillons. L'exploitation des jeux de données MS/MS par les moteurs de recherche, et les méthodes de validation associées, ont en effet fait l'objet de nombreux développements bioinformatiques, permettant de produire et de manipuler facilement ces données qualitatives. Cependant, ces comparaisons sont peu efficaces, dans la mesure où l'identification par MS/MS dans les approches « Data Dependant Acquisition » (DDA) classiques sont intrinsèquement peu reproductibles. Basées sur un processus stochastique de sélection automatique des précurseurs à séquencer par le spectromètre de masse, la nature des spectres MS/MS générés peut varier de façon significative pour un même échantillon complexe analysé successivement sur un même système (typiquement 20%/30% d'identifications protéiques différentes), en raison de la variabilité associée à la chromatographie et à la source (Liu, Sadygov et al. 2004). Pour être relativement fiables, ces comparaisons doivent porter sur des listes « exhaustives » obtenues après de nombreuses analyses répétées ou un fractionnement extensif de l'échantillon, représentant la couverture analytique maximale avec le système utilisé, et évitant de ce fait les biais liés à un sous-échantillonnage en MS/MS dans le cas de mélanges complexes analysés sur des appareils de vitesse de séquençage limitée.

Pour réaliser une comparaison efficace des mélanges complexes analysés par nanoLC-MS/MS, il était donc nécessaire d'extraire des jeux de données produits des métriques quantitatives ou semi-quantitatives associées aux peptides trypsiques, puis aux protéines. Cela a été rendu possible par le développement, aux cours de ces dernières années, d'outils logiciels de plus en plus performants. Ces métriques sont essentiellement liées soit à l'intensité du signal MS mesuré pour chaque peptide, soit au nombre de spectres MS/MS enregistrés pour chaque protéine (Figure 9) (voir ci-dessous).

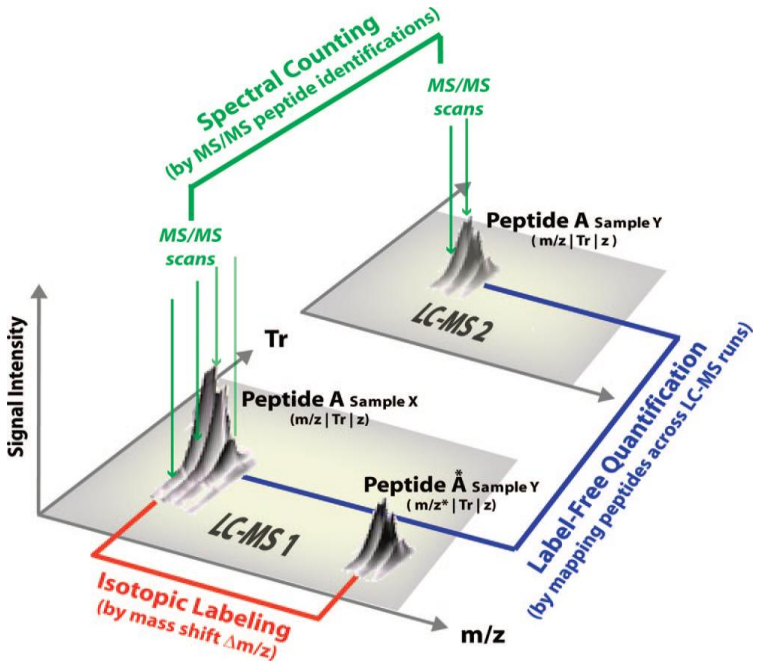


Figure 9 : Différentes approches pour la quantification des peptides selon la stratégie de protéomique quantitative utilisée. *Leur quantification est réalisée au sein d'une même acquisition nanoLC-MS/MS lorsqu'un marqueur isotopique est introduit. Elle est réalisée entre deux acquisitions nanoLC-MS/MS dans le cas de quantification sans marquage. Elle peut être basée sur la comparaison du nombre de spectres MS/MS (spectral counting) ou sur la comparaison des intensités MS des peptides. D'après (Mueller, Brusniak et al. 2008).*

Par ailleurs, l'analyse du signal MS a traditionnellement été couplée à des méthodes de marquage isotopique, permettant *in fine* de repérer facilement sur le spectre MS les signaux associés à un même peptide issu des échantillons à comparer. On distingue donc généralement les approches basées sur le marquage isotopique des peptides ou des protéines qui reposent principalement sur la comparaison des signaux MS, et celles réalisées sans marquage qui utilisent soit l'intensité des signaux MS mesurés soit le nombre de MS/MS effectué sur chaque protéine (Figure 10) (Bantscheff, Schirle et al. 2007; Bantscheff, Lemeer et al. 2012).

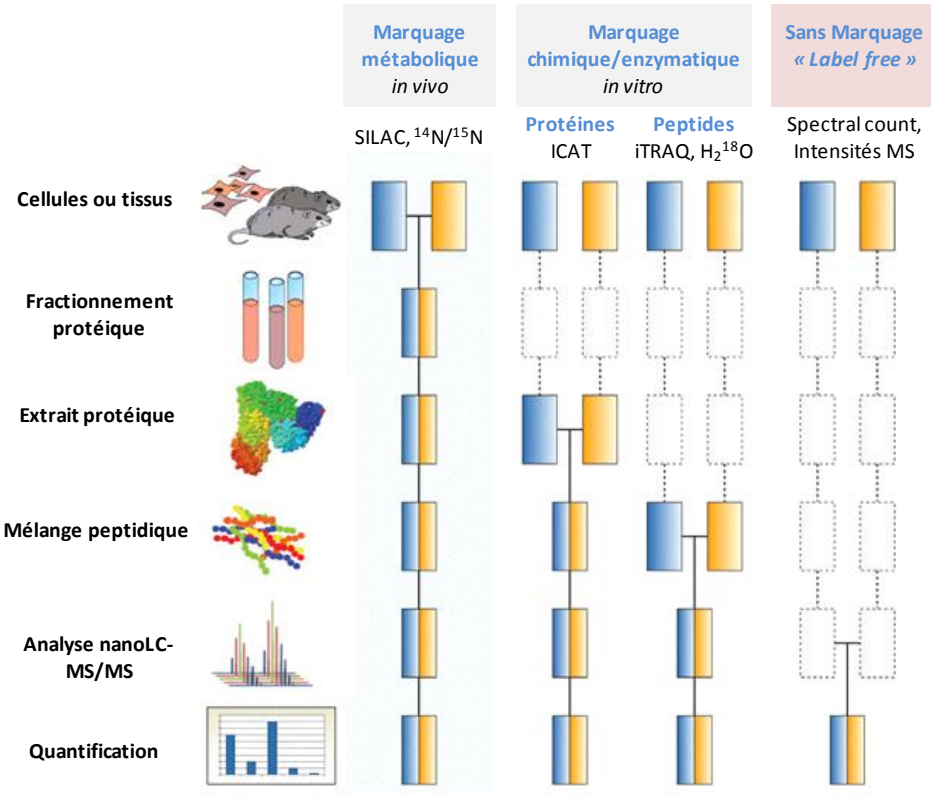


Figure 10 : Différentes méthodes de protéomiques quantitatives par nanoLC-MS/MS avec ou sans marquage isotopique*. Les marqueurs isotopiques sont introduits de différentes façons et à différents niveaux du processus selon la stratégie employée. D'après (Bantscheff, Schirle et al. 2007; Bantscheff, Lemeer et al. 2012)*

a. *Les méthodes avec marquage isotopique*

Les premières analyses quantitatives nanoLC-MS/MS globales ont utilisé des stratégies mettant en œuvre un marquage isotopique. L'introduction d'un marqueur sur les peptides ou les protéines facilite en effet la quantification relative de deux conditions au sein de la même acquisition nanoLC-MS/MS, la différence de masse entre un peptide marqué et un peptide non marqué étant mesurable en spectrométrie de masse (Figure 11). Ces approches consistent donc à marquer l'un des deux échantillons à comparer (échantillon lourd) avec un isotope stable (D, ^{13}C, ^{15}N, ^{18}O), puis à le rassembler avec l'échantillon léger, les deux échantillons étant ensuite traités et analysés simultanément. Les paires peptidiques ainsi formées se différencient uniquement par un écart de masse Δm au sein des spectres de masse, et permettent de quantifier de façon relative les peptides correspondants par comparaison de leur signal MS (Figures 9, 11). Plus le marquage est introduit tôt dans le processus, plus les échantillons peuvent être rassemblés précocement, évitant ainsi au mieux les différentes sources de variabilité liées au traitement parallèle des échantillons (Figure 10).

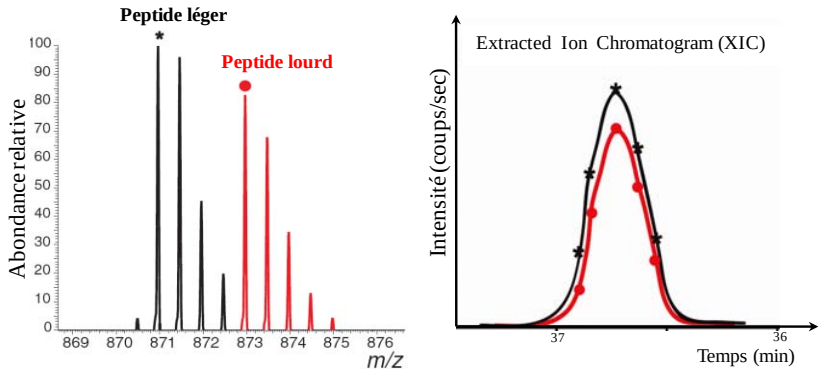


Figure 11 : Extraction des données quantitatives à partir d'un spectre de masse. *A gauche, après acquisition du spectre de masse, visualisation des massifs isotopiques pour chaque peptide, marqué (rouge) ou non marqué (noir). A droite, au cours de l'élution chromatographique, le signal du peptide est suivi par la courbe dont l'aire correspond au courant mesuré (XIC), une mesure proportionnelle à l'abondance du peptide. D'après (Ong and Mann 2005).*

Les marqueurs isotopiques peuvent être introduits de différentes façons et à différentes étapes du processus analytique (Figure 10). Ils peuvent d'une part être introduits *in vitro* par voie chimique grâce à différents réactifs qui ciblent les fonctions réactives des peptides/protéines comme les groupements amine et thiol. Ce marquage peut se faire au niveau protéique, c'est le cas du marquage ICAT (« Isotope-Coded Affinity Tag ») pour lequel les résidus cystéine sont marqués avec un réactif comprenant soit des ^{12}C soit des ^{13}C (Hansen, Schmitt-Ulms et al. 2003). Il peut aussi se faire au niveau peptidique comme dans l'approche iTRAQ (« Isobaric tag for relative and absolute quantitation ») (Thompson, Schafer et al. 2003) qui est basée sur le marquage des peptides trypsiques au niveau des groupements amines par un réactif associé à des étiquettes isobariques qui génèrent des ions rapporteurs de masses spécifiques lors de la fragmentation MS/MS des ions parents (Griffin, Xie et al. 2007). Le marquage des peptides par des isotopes stables peut d'autre part être réalisé par incorporation de ^{18}O au moment du processus de digestion enzymatique des protéines en présence de $H_2{}^{18}O$ (Yao, Freas et al. 2001; Reynolds, Yao et al. 2002). Enfin, un marquage métabolique peut également être réalisé, principalement dans le cas de plantes ou de cellules en culture. Il consiste à ajouter dans le milieu de culture un élément isotopiquement alourdi qui sera ensuite intégré au niveau des protéines par la cellule. L'un des premiers décrit dans la littérature met en œuvre un marquage total de bactéries en utilisant un milieu de culture enrichi en azote ^{15}N (Oda, Huang et al. 1999). Peu après, le marquage SILAC (« Stable Isotope Labeling by Amino acids in Cell culture ») (Ong, Blagoev et al. 2002) a été développé. Dans ce cas, cet élément correspond à un ou plusieurs acides aminés pour lesquels certains atomes ont été remplacés par une forme plus lourde mais stable. Ainsi, au cours de la croissance cellulaire et du « turn-over » protéique, les acides aminés marqués sont incorporés dans toutes les chaînes protéiques néosynthétisées par la cellule. En pratique, on utilise souvent un double marquage des lysines et des arginines (Graumann, Hubner et al. 2008), assurant ainsi à l'ensemble des peptides issus du clivage trypsique d'une protéine de posséder au moins un acide aminé marqué. L'écart de masse entre la forme légère et lourde d'un peptide donné est fonction de l'incrément de masse du marqueur présent dans sa séquence (+8 Da pour une lysine marquée $^{13}C_6{}^{15}N_2$ et +10 Da pour une arginine marquée $^{13}C_6{}^{15}N_4$) et du nombre total de marqueurs présents, nombre qui est normalement égal à 1 si le peptide ne présente pas de coupure manquée. Le principal avantage des stratégies de marquage

métabolique réside dans le fait que les échantillons sont combinés en amont des traitements biochimiques nécessaires à l'analyse protéomique. Ceci exclut donc des biais potentiels introduits lors de ces différentes étapes et permet théoriquement d'obtenir des données quantitatives plus fiables. Cette technique est particulièrement bien adaptée aux cellules immortalisées dont la majorité arrive à incorporer plus de 97% du marquage après 5 cycles de division cellulaire (Ong and Mann 2006). Elle ne permet cependant pas de quantifier des tissus ou des fluides biologiques.

L'approche SILAC a remporté un franc succès ces dernière années et de nombreux travaux très intéressants sont retrouvés dans la littérature (Graumann, Hubner et al. 2008; Hanke, Besir et al. 2008; Kruger, Moser et al. 2008). Cette technique a été utilisée pour détecter de faibles variations quantitatives ainsi que pour la caractérisation et la dynamique des modifications post traductionnelles (Blagoev, Kratchmarova et al. 2003). La stratégie est bien adaptée à l'étude quantitative des profils d'expression protéique à grande échelle, mais elle a également été décrite pour l'étude d'interactomes afin de distinguer les vrais partenaires protéique des faux positifs (Selbach and Mann 2006; Vinther, Hedegaard et al. 2006; Dengjel, Kristensen et al. 2008; Trinkle-Mulcahy, Boulon et al. 2008; Hayashi, Kim et al. 2009). Parmi les inconvénients associés au SILAC, on peut cependant citer le coût des marqueurs isotopiques, la difficulté de mise en œuvre dans le cas de tissus ou de certaines cellules cultivées dans des conditions non standard, et le nombre limité de conditions généralement comparées. Pour surmonter ces problèmes, des approches de « spike-in » SILAC ou super-SILAC ont été récemment décrites (Geiger, Cox et al. 2010; Geiger, Wisniewski et al. 2011; Deeb, D'Souza et al. 2012). Elles consistent à rajouter dans chacun des échantillons à comparer, un lysat de différentes lignées cellulaires marquées SILAC apparentées au tissu ou au système biologique à étudier, qui est utilisé comme un super standard interne représentant l'ensemble des espèces du protéome. Une autre variante du SILAC, le « pulsed-SILAC » a également été développée (Lam, Lamond et al. 2007; Schwanhausser, Gossen et al. 2009; Cambridge, Gnad et al. 2011). Elle permet la mesure du turnover et du taux de traduction protéique grâce à une quantification des protéines néo-synthétisées qui sont marquées spécifiquement avec des isotopes lourds. Ce marquage est obtenu en transférant les cellules initialement cultivées dans un milieu SILAC léger (ou dans un milieu composé d'un mélange 1 : 1 SILAC léger/SILAC lourd) dans un milieu SILAC lourd pendant un temps donné.

La possibilité de rassembler les échantillons à comparer est à l'origine du succès de toutes ces méthodes de marquage isotopique puisqu'elle permet une quantification très précise même à partir de données MS de faible résolution. Le multiplexage de l'analyse au sein d'un même spectre MS génère cependant des spectres de plus grande complexité. Cela peut conduire à une plus faible couverture analytique du protéome étudié, en raison de la vitesse de séquençage limitée des appareils qui passent plus de temps à séquencer les deux ions de la paire peptidique lourd-léger. De plus, l'augmentation de la complexité de l'échantillon peut accentuer les problèmes de superposition de massifs isotopiques issus de peptides différents. Les cartes peptidiques LC-MS sont généralement très encombrées sur les mélanges complexes (Michalski, Cox et al. 2011), et ces problèmes d'interférence entre différents massifs, apparaissant sur une seule des espèces de la paire peptidique, et peuvent entraîner des erreurs de quantification.

b. *Méthodes sans marquage ou « label free »*

Grâce aux développements instrumentaux, les nouvelles générations de spectromètre de masse comme le LTQ-Orbitrap sont capables de réaliser des analyses MS à haute résolution (grâce à l'analyseur Orbitrap) et disposent par ailleurs d'une vitesse de séquençage élevée (grâce à la trappe linéaire LTQ). Cette démocratisation de la haute-résolution a permis l'émergence de méthodes d'analyse quantitative alternatives basées notamment sur le traitement du signal MS sans utilisation de marquage isotopique. Cette nouvelle approche sans marquage ou « label-free » est plus facile à mettre en œuvre sur des données à haute-résolution (permettant d'obtenir une très grande précision de masse sur les ions peptidiques) que sur des données issues d'appareils à basse résolution, où la comparaison entre différentes analyses peut générer un grand nombre de faux appariements de signaux peptidiques.

Deux types de stratégies label free existent et reposent sur l'exploitation des données LC-MS/MS à deux niveaux différents (Figure 9) : l'une mesure et compare le nombre de spectres MS/MS identifiant les peptides d'une protéine (« spectral counting »), la seconde est basée sur la mesure et la comparaison des intensités des signaux MS des peptides d'une protéine. Elles suscitent un grand intérêt pour leur facilité de mise en œuvre, leur application pour l'étude d'un nombre non limité d'échantillons de tous types, leur moindre coût et leurs performances en terme de couverture de protéomes, tout en permettant d'obtenir une quantification fiable des protéines contenues dans un mélange (Olsen, Nielsen et al. 2007). Elles sont de plus en plus utilisées et s'appliquent à de nombreux domaines en biologie (Bodenmiller, Wanka et al. 2010; Luber, Cox et al. 2010; Bildl, Haupt et al. 2012).

Spectral counting

Au cours de l'analyse nanoLC-MS/MS, l'utilisation du mode DDA permet la sélection par le spectromètre de masse des espèces les plus intenses pour les séquencer en MS/MS. L'approche « spectral counting » est basée sur l'hypothèse que l'échantillonnage MS/MS des peptides d'une protéine, c'est-à-dire la fréquence de fragmentation de ces peptides, est directement lié à l'abondance de la protéine au sein d'un échantillon. Il existe en effet une corrélation entre l'abondance d'une protéine et le nombre d'événements MS/MS que l'on peut observer pour cette même protéine (Liu, Sadygov et al. 2004). Ainsi une quantification relative peut être effectuée en comparant le nombre de ces spectres MS/MS acquis pour une protéine donnée entre deux séries d'expériences. L'analyse par « spectral counting » constitue une méthode simple pour réaliser une quantification sans marquage mais elle souffre cependant d'un certain nombre de limitations. Elle n'est en effet valide que pour une gamme de concentration limitée (Bildl, Haupt et al. 2012). En particulier, aux faibles concentrations, le nombre de MS/MS associé à une protéine devient très peu reproductible, et ne permet pas de réaliser une quantification relative fiable. Une étude comparative portant sur une évaluation de la quantification par « spectral counting » par rapport au SILAC a par exemple montré qu'un seuil d'au moins 5 MS/MS par protéine devait être fixé pour obtenir une précision quantitative équivalente à l'échelle de la population totale de protéines analysées (Collier, Sarkar et al. 2010). De fait, avec cette méthode, un pourcentage significatif des protéines identifiées dans un protéome complexe est donc difficilement quantifiable.

Analyse des données MS

La seconde méthode de quantification sans marquage est basée sur l'analyse des signaux MS acquis au cours des analyses nanoLC-MS/MS. Elle nécessite l'obtention de données MS haute-résolution et le développement des instruments de dernière génération incluant un analyseur de type FT-ICR, FT-Orbitrap ou TOF à haute résolution, a donc constitué une étape cruciale pour cette stratégie récente. Contrairement à l'approche « spectral count » qui utilise les résultats d'identification produits par les moteurs de recherche à partir des données MS/MS, elle cherche à exploiter les données MS contenues dans les fichiers bruts issus des spectromètres de masse. En ce sens, elle se rattache aux approches basées par exemple sur un marquage isotopique de type ICAT ou SILAC, dans lesquelles l'information quantitative doit également être extraite des spectres MS bruts. Ces approches nécessitent donc des outils bioinformatiques plus élaborés.

IV-2. Analyse bioinformatique des données protéomiques quantitatives

IV-2.1 Méthodes d'analyse des données MS.

La première approche (historiquement mise en œuvre dans les premiers logiciels consacrés à l'analyse des données ICAT) consiste à utiliser comme point de départ les résultats d'identification MS/MS. A partir de ces résultats d'identification validés, les signaux MS des peptides sont extraits sous forme d'un appel d'ion ou XIC (« Extracted ion chromatogram ») (Figure 12). En effet, dans une expérience LC-MS, l'intensité du signal MS d'un peptide qui élue de la colonne chromatographique peut être suivie au cours du temps. L'aire sous la courbe du pic chromatographique est le courant ionique extrait (XIC) et elle est proportionnelle à l'abondance du peptide dans l'échantillon (Ong and Mann 2005). Cette méthode consiste donc à extraire l'intensité du signal pour chaque peptide identifié, de m/z connu. En fonction de la tolérance de masse autorisée pour réaliser cette extraction, plusieurs pics d'élution potentiels peuvent être extraits sur l'ensemble de l'échelle de temps du gradient chromatographique. Pour identifier le XIC correspondant au peptide recherché, il est donc également possible d'utiliser l'information sur le temps auquel la MS/MS a été réalisée pour ce peptide, ou sur son état de charge. La quantification est réalisée en comparant les XIC des peptides au sein des différents échantillons.

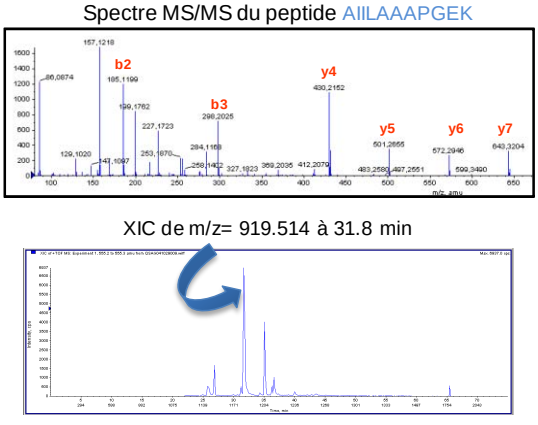


Figure 12 : Extraction du signal MS d'un peptide séquencé et identifié.

L'autre approche est centrée sur l'analyse des cartes LC-MS. Dans ce cas, c'est l'ensemble des signaux MS bruts qui est traité et interprété, sans *a priori* sur la masse, la charge ou le temps d'élution des peptides identifiés. Dans cette approche, les logiciels sont conçus pour définir, à partir des cartes LC-MS brutes, les empreintes MS des peptides (ou « features ») sur la base de la modélisation théorique des massifs isotopiques peptidiques et des profils d'élution de ces peptides sur colonne de phase inverse. Les signaux associés à un même peptide sur l'échelle des masses (isotopes) et de temps (mesures successives d'un même ion sur le pic d'élution) sont donc regroupés et intégrés pour fournir une mesure quantitative d'une même entité chimique. La quantification relative est réalisée en alignant les cartes et en comparant les aires de ces empreintes peptidiques dans les différents échantillons. Le lien avec les identifications peptidiques n'est réalisé que dans un second temps. L'identité des empreintes peptidiques peut être déterminée en les reliant avec les scans MS/MS ayant généré une identification, ou bien en réalisant une identification ciblée lors d'une seconde acquisition, ou enfin à partir d'une banque de données d'identifications, comme dans l'approche AMT (« Accurate Mass and Time ») (Smith, Anderson et al. 2002). Cette méthode est en principe plus exhaustive puisque l'ensemble des signaux est analysé, et non pas seulement ceux qui ont donné lieu à une identification de peptide par MS/MS.

IV-2.2 Outils bioinformatiques

Des outils bioinformatiques de plus en plus performants ont été développés ces dernières années pour permettre l'analyse quantitative des données protéomiques. Historiquement, ils ont tout d'abord été conçus pour traiter les données issues d'approches de quantification avec marquage isotopique. Dans ces approches, l'évaluation de l'abondance relative des peptides est réalisée en comparant les signaux des deux formes légères et lourdes d'un même peptide, au sein d'un même spectre MS. Selon la méthode de marquage utilisée, la valeur de l'écart de masse existant entre les peptides marqués et non marqués doit être configurée et calculée par le logiciel. Le traitement informatique consiste alors en l'extraction des signaux correspondant aux paires isotopiques, les peptides de chaque doublet étant repérés par leur différence de masse caractéristique. L'automatisation de la quantification a été possible grâce à l'apparition de logiciels qui se basent sur l'identification des peptides par MS/MS pour reconstituer le XIC des paires peptidiques, comme XPRESS (Han, Eng et al. 2001) ou ASAPratio pour le SILAC, ou encore msQuant (Schulze and Mann 2004) qui prend en charge les données d'identification Mascot et permet de réaliser la quantification relative de paires SILAC. Ces logiciels de première génération étaient pour la plupart difficiles à prendre en main et pouvaient présenter des problèmes de compatibilité à différents niveaux. A la même époque, le logiciel MFPaQ a donc été développé au sein de l'équipe pour répondre aux besoins propres de validation et de quantification des analyses protéomiques basées sur des marquages isotopiques. Il utilise lui aussi une approche basée sur l'extraction de XIC et gère la quantification de données marquées SILAC, ^{14}N-^{15}N, ICAT. L'apparition des instruments haute-résolution et à haute vitesse de séquençage s'est ensuite accompagnée d'efforts de plus en plus importants pour fournir des outils bioinformatiques adaptés au traitement de données de plus en plus volumineuses, mais également de meilleure qualité. La haute résolution a notamment été l'élément déclencheur de la mise au point de logiciels de quantification basée sur la comparaison de cartes LC-MS, donnant ainsi accès à la quantification d'espèces non séquencées et donc à une plus grande profondeur du protéome étudié. Ce type d'algorithme appliqué à l'analyse des données par

marquage isotopique SILAC a été implémenté dans le logiciel MaxQuant (Cox and Mann 2008) développé par le groupe de Mathias Mann qui constitue la référence dans le domaine.

Beaucoup d'outils bioinformatiques ont également été développés pour les stratégies « label-free ». Dans ces approches, la totalité du processus, de la préparation de l'échantillon à l'analyse MS est réalisé indépendamment pour chacun des échantillons à comparer, ce qui peut introduire des biais quantitatifs à chaque étape. Un retraitement des données après leur acquisition est donc nécessaire pour corriger les variabilités : (-) afin de pouvoir correctement extraire et associer les signaux, les variabilités d'élution chromatographique sont corrigées en alignant les données MS en temps de rétention (-) la normalisation des données MS est également nécessaire pour corriger la variabilité éventuelle issue de la réponse en signal MS du spectromètre de masse ou de la préparation des échantillons. La plupart des logiciels dédiés aux approches « label-free » sont basés sur une approche qui consiste à générer puis à comparer des cartes LC-MS. On peut citer parmi les plus connus MaxQuant (Cox and Mann 2008), MSInspect (Bellew, Coram et al. 2006), OpenMS (Sturm, Bertsch et al. 2008), Decon2LS (Jaitly, Mayampurath et al. 2009), ou le logiciel commercial Progenesis LC-MS. Bien qu'ils offrent une solution attractive pour l'analyse exhaustive des données disponibles, ces algorithmes basés sur la détection d'empreintes peptidiques et l'alignement de cartes LC-MS nécessitent des temps de calcul importants, ce qui rend encore pour l'instant difficile l'analyse de grandes séries de fichiers. Par ailleurs, la mise en relation des empreintes peptidiques détectées avec les données d'identification est plus ou moins bien gérée en fonction des logiciels. Enfin, dans la mesure où les cartes LC-MS sont souvent générées de façon individuelle, en utilisant des valeurs seuil pour la reconnaissance des empreintes peptidiques, les signaux de faible intensité sont souvent détectés de façon non reproductible dans différents échantillons, ce qui entraîne la présence de nombreuses valeurs manquantes et complique l'analyse statistique des données. D'autres outils ont par ailleurs continué à implémenter une approche basée sur l'identification MS/MS des peptides pour extraire les valeurs de XIC pour chaque pic d'élution peptidique. Cette fois, les signaux ne sont pas extraits dans le même scan MS, pour deux m/z différents correspondants aux deux membres de la paire isotopique, mais pour un même m/z dans des scans issus de plusieurs analyses différentes. Cette méthode, en principe plus simple et rapide, est par exemple utilisée par Ideal-Q (Tsou, Tsai et al. 2010) ou dans le logiciel commercial PeakView (AB Sciex). C'est également la méthode de quantification implémentée dans le logiciel MFPaQ, que j'ai utilisé au cours de ma thèse. Très récemment, elle a aussi été décrite pour le traitement des données sans marquage dans les approches globales par le logiciel Skyline, initialement dédié aux approches ciblées par MRM (Schilling, Rardin et al. 2012).

Les approches quantitatives sont essentielles en protéomique pour étudier et comprendre les mécanismes biologiques. Elles sont en particulier utiles pour l'analyse de complexes protéiques et d'interactions protéine-protéine (protéomique d'interaction) pour lesquelles elles permettent de d'identifier les protéines partenaires spécifiques. Elles sont également indispensables pour l'étude de protéomes complexes et de leurs variations dans un contexte environnemental donné (protéomique d'expression). Ces deux aspects de la protéomique sont abordés dans les deux parties suivantes.

Partie II. Etude de complexes protéiques

La plupart des processus biologiques de la cellule, notamment les processus essentiels comme la réplication de l'ADN, la transcription, la traduction, la dégradation des protéines ou encore le contrôle du cycle cellulaire, sont régis par des protéines qui interagissent et s'assemblent pour former des complexes multi-protéiques. Ces complexes évoluent au cours du temps et en fonction de leur environnement. Ils forment ainsi des réseaux dynamiques d'interactions protéine-protéine stables et/ou transitoires, qui coordonnent les machineries cellulaires et permettent le bon fonctionnement de la cellule. Les interactions protéine-protéine sont donc essentielles. La perturbation de certaines d'entre elles est d'ailleurs responsable de dysfonctionnements cellulaires et par conséquent de diverses pathologies humaines dont certains cancers (Charbonnier, Gallego et al. 2008; Kar, Gursoy et al. 2009). L'étude de ces complexes est donc importante pour pouvoir avancer dans la compréhension des mécanismes moléculaires de la cellule et de la pathophysiologie de certaines maladies. Pour les caractériser, il existe différentes techniques, basées sur des approches génétiques (double hybride) ou biochimiques (pull-down) (Suter, Kittanakom et al. 2008). Parmi celles-ci, la combinaison de la purification par affinité et de l'analyse protéomique par spectrométrie de masse (« Affinity purification coupled with mass spectrometry », AP-MS) constitue notamment une méthode de choix.

I. Méthode générale pour l'analyse de complexes protéiques et défis associés

L'AP-MS consiste à purifier par affinité la protéine d'intérêt afin d'enrichir les complexes protéiques dans lesquels elle est impliquée, puis à identifier ses interactants par une analyse protéomique (généralement en nanoLC-MS/MS). Contrairement à des techniques comme le double hybride, elle repose sur l'analyse de complexes issus d'un environnement cellulaire physiologique, et permet donc d'avoir accès à des complexes « natifs ». Les interactions protéiques qui dépendent de modifications post-traductionnelles (MPT) sont par conséquent conservées et peuvent être identifiées. Les MPT des composants des complexes peuvent elles aussi être identifiées et localisées en parallèle. Les complexes à étudier peuvent être purifiés à partir de n'importe quel type de matériel biologique, de cellules en culture à des organes entiers (Ranish, Yi et al. 2003; Bai, Markham et al. 2008; Ho, Ronan et al. 2009; Tai, Besche et al. 2010; Li, Yau et al. 2011).

Le principe de l'AP-MS (Figure 13) consiste à enrichir le complexe d'intérêt par chromatographie d'affinité avant de l'analyser par nanoLC-MS/MS. Pour cela, un extrait des protéines de l'échantillon doit dans un premier temps être préparé. La protéine appât est ensuite purifiée à partir de cet extrait grâce à une matrice présentant une forte affinité pour celle-ci. Des lavages des complexes purifiés sur la matrice sont alors nécessaires pour éliminer les protéines se liant de façon non spécifique au système. Les complexes sont ensuite élués de la matrice avant d'être

analysés par nanoLC-MS/MS. L'éluat peut être éventuellement fractionné sur gel 1D SDS-PAGE avant l'analyse MS.

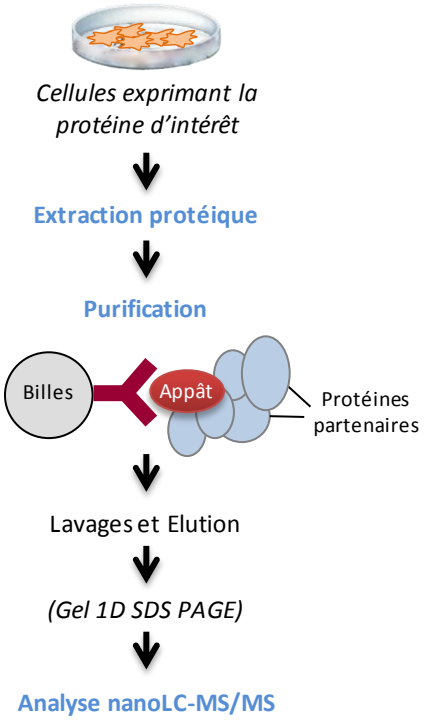


Figure 13 : Stratégie générale pour l'analyse de complexes protéiques par nanoLC-MS/MS

Cette stratégie est conceptuellement simple mais s'avère en pratique parfois complexe à mettre en œuvre. Elle implique en effet deux défis majeurs. La bonne réussite de la méthode repose d'abord sur la capacité à correctement isoler les complexes protéiques. La préparation de l'échantillon est donc une étape clé. Elle implique tout d'abord une lyse cellulaire, parfois un fractionnement subcellulaire de l'échantillon puis une purification des complexes protéiques. Les conditions expérimentales utilisées influencent fortement l'efficacité de ces différentes étapes. Il faut parvenir à lyser correctement les cellules et à conserver la solubilité des protéines, tout en préservant le complexe d'intérêt et en maintenant toutes les protéines associées à la protéine appât, ce qui implique la préservation d'interactions même transitoires et labiles. Les différentes approches envisageables pour l'extraction et la purification de complexes sont abordées ci-dessous.

Les complexes ainsi isolés sont ensuite analysés par spectrométrie de masse. Grâce à la grande sensibilité des spectromètres de masse dernière génération, l'analyse nanoLC-MS/MS des échantillons purifiés peut conduire à l'identification de centaines de protéines. Elles ne sont pas pour autant toutes des interactants potentiels. La majorité correspond en effet à des protéines qui se fixent de façon non spécifique à la matrice d'affinité, aux billes, à l'anticorps ou encore au complexe lui-même pendant la préparation de l'échantillon (Figure 14).

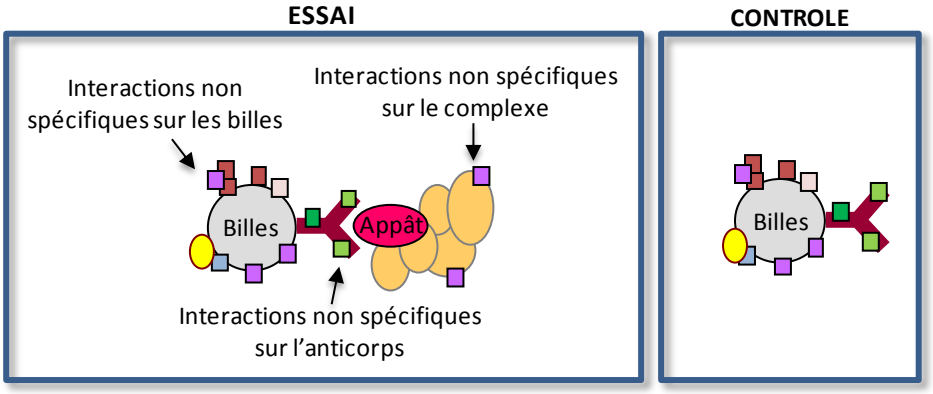


Figure 14 : Protéines contaminantes pouvant être retrouvées lors de la purification de complexes protéines.

Le second défi majeur de ce type d'analyse est donc de distinguer les vrais partenaires spécifiques de la protéine appât, parfois peu abondants et labiles, des protéines contaminantes, souvent des protéines majoritaires de la cellule, co-purifiées de façon non spécifique. Pour minimiser ce problème, il est possible d'essayer d'optimiser les étapes biochimiques en amont et de réaliser une purification poussée des complexes incluant par exemple plusieurs étapes de purification et/ou des conditions de lavage stringentes. Cette approche n'est cependant pas toujours idéale, en particulier si le complexe étudié est fragile ou fait intervenir des interactions protéine-protéine faibles ou labiles qui risquent d'être perdues. Une autre approche pour tenter d'identifier de façon non ambiguë les vrais partenaires consiste à comparer les protéines identifiées dans l'échantillon immunopurifié avec celles identifiées dans un échantillon contrôle adéquat, dans lequel le complexe n'est pas enrichi. Cette comparaison peut être réalisée grâce à une analyse protéomique quantitative entre l'échantillon IP et l'échantillon contrôle (cf Partie II-III. *Identification des partenaires spécifiques bona fide)*).

II. Isolement des complexes protéiques

La préparation de l'échantillon pour l'analyse des complexes protéiques se déroule en deux étapes : (1) préparation d'un extrait protéique à partir de l'échantillon biologique, (2) enrichissement des complexes. A chaque étape, les conditions expérimentales doivent être optimisées pour préserver les interactions protéine-protéine spécifiques.

II-1. Lyse cellulaire et extraction des protéines

Lors de cette première étape, les complexes macromoléculaires vont être extraits de leur contexte cellulaire, brutalement transférés dans des conditions de tampon différentes de celles de leur milieu d'origine, et mis en présence de tout un ensemble de protéines issues de localisations cellulaires différentes. Il convient alors de trouver les conditions permettant une extraction optimale des protéines, n'entrainant pas la dissociation des complexes, et limitant les interactions non-spécifiques. Elles doivent de plus être compatibles avec la purification en aval.

Une lyse totale des cellules peut être réalisée. Il est pour cela nécessaire de casser les membranes de la cellule de façon à relarguer son contenu protéique. Différentes méthodes sont communément utilisées. Certaines reposent sur une action mécanique plus ou moins poussée pour casser les membranes plasmiques et/ou intracellulaires à l'aide de différents types d'homogénéisateurs (par exemple homogénéisateurs manuels à piston de type « Dounce », homogénéisateurs mécaniques rotatifs à haute vitesse de type « Ultra-turrax » ou encore appareils de sonication). Des cycles de congélation/décongélation permettent également de casser la membrane plasmique. D'autres protocoles utilisent des détergents pour fragiliser ou solubiliser les membranes cellulaires. Il s'agit généralement de détergents non-ioniques relativement compatibles avec les interactions protéine-protéine (NP-40, Triton X100, lauryl maltoside, etc) (Selbach and Mann 2006; Tsai, Greco et al. 2012). Pour améliorer l'extraction des protéines de l'échantillon, l'utilisation d'une faible concentration de détergents peut être associée à une étape de casse mécanique des cellules (Trinkle-Mulcahy, Boulon et al. 2008).

Il est par ailleurs possible d'enrichir l'échantillon en une fraction cellulaire particulière (fraction cytoplasmique, fraction nucléaire…) contenant la protéine d'intérêt, en réalisant un fractionnement subcellulaire des cellules eucaryotes. En plus de l'enrichissement en protéine d'intérêt apporté, ce fractionnement peut permettre l'étude de complexes protéiques spécifiques d'un compartiment cellulaire donné. Par ailleurs, il évite aussi la mise en présence du complexe protéique d'intérêt avec de nombreuses protéines provenant d'autres organelles, ce qui diminue au final le nombre de contaminants fixés de façon non-spécifique au complexe. La fraction cytoplasmique est communément obtenue en réalisant une lyse de la membrane plasmique des cellules grâce à l'utilisation d'un tampon hypotonique qui provoque l'éclatement des cellules. Cette lyse osmotique est généralement assistée d'une casse mécanique de la membrane plasmique (Dignam, Lebovitz et al. 1983; Andersen, Lyon et al. 2002; Boisvert, Lam et al. 2010). Une centrifugation à basse vitesse permet ensuite de séparer les noyaux (culot) du reste du contenu cellulaire (surnageant post-nucléaire). Les protéines nucléaires peuvent ensuite être obtenues à partir des noyaux ainsi isolés par extraction saline (Dignam, Lebovitz et al. 1983; Nakatani and Ogryzko 2003; Groth, Corpet et al. 2007). Cette technique permet majoritairement l'extraction des protéines nucléaires solubles, mais il existe également des protocoles permettant d'extraire les protéines liées à la chromatine, qu'il faut généralement solubiliser en cassant l'ADN génomique en courts fragments par des moyens mécaniques (sonication) (Dejardin and Kingston 2009) ou enzymatiques (DNAse I, MNase, benzonase…) (Fujita, Kiyono et al. 1997; Remboutsika, Lutz et al. 1999).

De nombreux protocoles ont ainsi été décrits pour la préparation d'extraits protéiques. Il n'existe cependant pas de méthode idéale et les conditions d'extraction optimales sont souvent complexe-dépendantes et donc à définir empiriquement. Comme évoqué précédemment, le principal défi au cours de cette étape est d'extraire au mieux les protéines de l'échantillon tout en préservant les interactions protéine-protéine. Cette étape nécessite l'utilisation de détergents et/ou de fortes concentrations salines qui peuvent provoquer la perte de certaines interactions protéine-protéine, en particulier des interactions faibles ou labiles. Un compromis doit donc être trouvé pour une extraction efficace et la préservation des complexes protéiques.

Afin de stabiliser les interactions protéine-protéine même labiles, transitoires ou dynamiques, il est possible d'exploiter la grande proximité qui existe entre les protéines constituant

un complexe pour figer les interactions qu'elles nouent à l'aide d'agents chimiques pontants. Ces agents, généralement homo/hétéro-bi-fonctionnels, sont capables, grâce à la présence de deux groupes réactifs, de former des liaisons covalentes entre les molécules. Ces groupes réactifs sont séparés par un bras espaceur d'une longueur variable (généralement entre 2-15 Å) qui détermine la longueur maximum entre deux molécules pouvant être réticulées. La nature de l'agent de réticulation ainsi que les conditions de réaction de la réticulation (concentration du réactif, temps de réaction) doivent être choisies afin de stabiliser les interactions protéine-protéine spécifiques, de minimiser les pontages de protéines contaminantes et d'éviter les précipitations protéiques dues à une réticulation trop importante. De nombreux agents de réticulation sont disponibles, et certains ont été appliqués avec succès pour l'analyse MS de complexes protéiques solubles. Par exemple, le formaldéhyde est souvent utilisé et permet d'induire une liaison covalente entre une fonction amine et plusieurs autres types d'acides aminés situés à faible distance (2-3 Å) (Sutherland, Toews et al. 2008; Bousquet-Dubouch, Baudelet et al. 2009; Knobbe, Revett et al. 2011; Muller, Jungblut et al. 2011; Klockenbusch, O'Hara et al. 2012). Des pontages chimiques peuvent également être utilisés pour l'étude de complexes protéiques au voisinage de l'ADN, en réalisant une réticulation des protéines et de l'ADN. Ainsi, grâce à un autre agent chimique, le DSP (dithiobis(succinimidyl)propionate), un réactif homo-bifonctionnel possédant un bras espaceur plus long (12 Å), les protéines associées à l'ADN télomérique ont pu être mises en évidence (Nittis, Guittat et al. 2010).

II-2. Enrichissement des complexes protéiques

Suite à l'extraction des protéines de l'échantillon, la protéine appât est enrichie avec ses protéines partenaires par purification d'affinité grâce à une matrice d'affinité. Pour cela, plusieurs méthodes peuvent être envisagées (Figure 15).

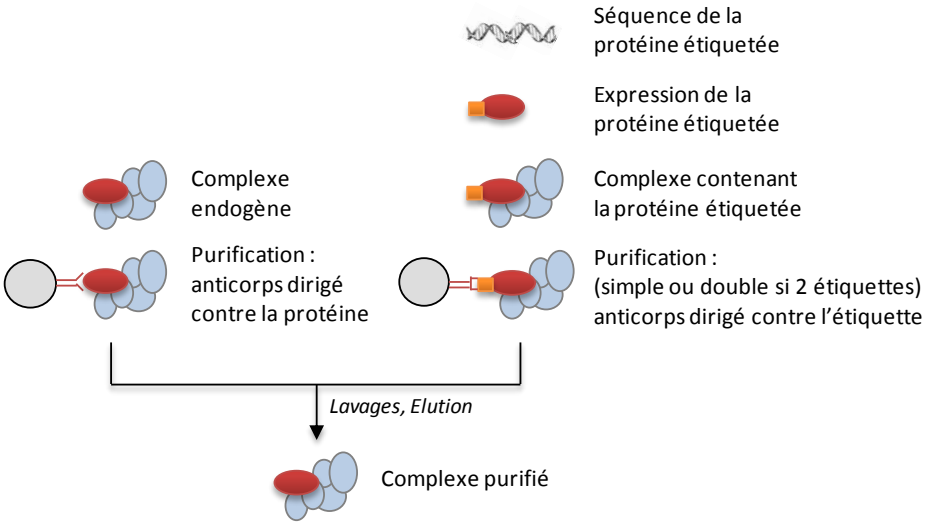


Figure 15 : Différentes stratégies d'enrichissement des complexes protéiques selon la nature de la protéine l'appât. *La purification peut être réalisée grâce à un anticorps ciblant la protéine endogène, permettant ainsi l'enrichissement de complexes endogènes. Par ailleurs, la protéine d'intérêt peut être exprimée ou surexprimée en fusion avec une ou plusieurs étiquettes d'affinité. La purification est dans le cas réalisée en tirant partie de l'affinité d'une matrice donnée (généralement un anticorps) pour cette/ces étiquettes.*

II-2.1 Enrichissement de protéines endogènes

L'approche la plus directe consiste à réaliser un enrichissement de la protéine endogène à l'aide d'un anticorps dirigé contre la protéine elle-même, ce qui permet de cibler et d'isoler les complexes endogènes natifs en conditions physiologiques. Cet anticorps doit présenter une bonne affinité ainsi qu'une bonne spécificité vis-à-vis de l'épitope ciblé afin d'enrichir au mieux la protéine d'intérêt sans présenter de réaction croisée avec d'autres protéines et ainsi éviter l'identification de protéines non spécifiques. Pour la purification, cet anticorps doit être immobilisé sur une matrice. Les méthodes classiques d'immunoprécipitation sont basées sur l'affinité de la partie Fc de l'anticorps pour la protéine A ou G, et utilisent des billes polymériques ou magnétiques sur lesquelles ces protéines sont greffées de façon covalente. Dans cette approche, le complexe anticorps- protéine appât-protéines partenaires peut être formé au préalable par incubation de l'anticorps avec l'extrait protéique, puis immunoprécipité en présence des billes de protéine A ou G. Alternativement, il est possible de préparer d'abord la matrice de protéine A ou G couplée à l'anticorps, puis d'immunoprécipiter les complexes d'intérêt présents dans l'extrait. Un avantage de ces méthodes est que l'anticorps, fixé via sa partie Fc, est correctement orienté pour interagir avec l'épitope ciblé, ce qui permet d'optimiser le rendement d'immunoprécipitation. En revanche, un inconvénient courant de ces protocoles pour les études par spectrométrie de masse est que l'anticorps est fixé de façon non covalente à la matrice, et souvent relargué par la suite en quantité importante lors de l'élution du complexe, ce qui peut perturber l'analyse et la détection de protéines peu abondantes. Pour remédier à ce problème, il existe des méthodes permettant de coupler de façon covalente l'anticorps à la matrice, soit par pontage chimique avec la protéine A ou G du support (dans ce cas l'anticorps est toujours correctement orienté, mais l'agent réticulant mis en solution peut diminuer l'efficacité de l'anticorps, par exemple en modifiant des lysines du site actif), soit en greffant directement l'anticorps lui-même sur une matrice polymérique activée (dans ce cas, n'importe quel acide aminé réactif présent à la surface de l'anticorps peut être engagé dans le pontage avec la matrice, et une partie des molécules d'anticorps sera potentiellement mal orientée pour la capture par affinité de la protéine cible). En fonction de l'approche utilisée, l'élution des complexes immunopurifiés peut ensuite être réalisée dans des conditions plus ou moins stringentes. La méthode la plus douce repose sur une élution spécifique à l'aide d'un antigène compétiteur, ce qui permet de limiter à la fois le décrochage des anticorps et celui de protéines contaminantes fixées sur la matrice, la protéine A ou G, ou l'anticorps. D'autres types de tampons d'élution plus puissants peuvent être employés pour assurer une élution totale du complexe, la rupture de la liaison anticorps-protéine appât pouvant être induite par exemple par une forte concentration saline, par un saut de pH, ou par des agents chaotropes ou dénaturants tels que l'urée ou le SDS. Une méthode courante consiste à faire bouillir les billes dans du tampon Laemmli, entraînant une élution très efficace de tous les complexes mais également des protéines non spécifiques. Cette approche de purification de complexes endogènes a été appliquée avec succès dans différentes études d'identification de partenaires protéiques à plus ou moins large échelle (Malovannaya, Li et al. 2010; Malovannaya, Lanz et al. 2011).

II-2.2 Enrichissement de protéines étiquetées

Malgré les efforts fournis pour produire des anticorps contre chaque protéine du protéome humain en particulier (Berglund, Bjorling et al. 2008), des anticorps ne sont cependant pas systématiquement disponibles ou lorsqu'ils le sont, pas toujours efficaces pour immunopurifier correctement les protéines étudiées. Pour pallier ce problème, une alternative largement répandue consiste à étiqueter la protéine d'intérêt en réalisant une fusion génique entre celle-ci et une séquence en acides aminés permettant ensuite une purification par affinité. La purification de la protéine chimérique résultante est alors réalisée grâce à une matrice d'affinité spécifique. Cette stratégie permet généralement un enrichissement efficace du complexe, mais il est nécessaire de prêter attention à certains critères. L'étiquette, placée en N-terminal ou C-terminal de la protéine, ne doit pas en effet modifier la structure, le repliement ou encore la stabilité de la protéine ni gêner ses interactions avec d'autres protéines pour ne pas altérer sa fonction. De nombreuses étiquettes d'affinité ont été développées (Li 2010). Elles diffèrent notamment au niveau de leur taille, pouvant aller d'un petit motif peptidique jusqu'à une protéine de plusieurs kDa, ainsi qu'au niveau de la matrice d'affinité permettant la purification de la protéine de fusion (Table 1).

Table 1 : Exemples d'étiquettes d'affinité couramment utilisées pour la purification de complexes protéiques.

Etiquette d'affinité	Description	Taille	Matrice d'affinité	Référence
FLAG	Peptide (DYKDDDDK)	~ 1 kDa	Anticorps anti-FLAG	Hopp, Prickett et al. 1988
HA	Peptide (YPYDVPDYA)	~ 1 kDa	Anticorps anti-HA	Field, Nikawa et al. 1988
Hexahistidine 6xHis	HHHHHH	< 1kDa	Ni^{2+} Co^{2+}	Hochuli, Bannwarth et al. 1988
StrepTagII	Peptide (WSHPQFEK)	~ 1 kDa	Streptavidine (StrepTactin)	Schmidt, Skerra et al. 1993
GFP	Protéine d'Aequoria victoria	~27 kDa	Anticorps anti-GFP	Chalfie, Euskirchen et al. 1994
GST	Enzyme	~ 26 kDa	Glutathion	Smith, Johnson et al. 1988

Certaines correspondent à des épitopes peptidiques qui sont enrichis grâce à des anticorps spécifiques, commerciaux et généralement disponibles sous forme greffée à différents types de matrices. C'est le cas des courts peptides FLAG (Hopp, Prickett et al. 1988; Olma, Roy et al. 2009) et HA (haemagglutinin) (Field, Nikawa et al. 1988; Sowa, Bennett et al. 2009) dont l'utilisation est très répandue. L'hexahistidine 6xHis (Hochuli, Bannwarth et al. 1988) et le StrepTagII (Schmidt and Skerra 1993) sont également de petits peptides mais font quant à eux appel à une stratégie de purification basée sur leur forte affinité pour une matrice particulière, constituée respectivement d'ions métalliques (nickel, cobalt) ou de streptavidine (StrepTactin). Ces étiquettes, grâce à leur petite taille, sont souvent considérées comme moins susceptibles de perturber la fonction des protéines auxquelles elles sont fusionnées. Cependant, des étiquettes de taille plus importante sont également utilisées. Il s'agit alors de domaines protéiques structurés, qui sont souvent décrits comme favorisant la solubilité de la molécule de fusion formée avec la protéine appât, facilitant ainsi sa purification. Certaines impliquent une purification par immunoaffinité, comme la GFP (« green fluorescent protein ») qui présente l'avantage de permettre en parallèle le suivi de la localisation et de la dynamique de la protéine d'intérêt par imagerie (Chalfie, Tu et al. 1994; Cristea, Williams et al. 2005; Hubner, Bird et al. 2010). La protéine GST (Glutathion-S-transférase) est également souvent employée et permet l'enrichissement de la protéine de fusion grâce à son affinité pour le glutathion (Smith and Johnson 1988; Jones, Wu et al. 2008).

Cette stratégie implique donc l'expression d'une protéine ectopique en fusion avec une étiquette, pour laquelle il est important de contrôler le niveau d'expression. Idéalement, celui-ci doit correspondre au niveau d'expression de la protéine endogène afin de mimer au mieux cette dernière et ne pas perturber les complexes protéiques dans lesquels elle est impliquée. Il est pour cela parfois possible d'étiqueter directement le gène codant pour la protéine d'intérêt dans le génome. Chez la levure et *E. coli*, la recombinaison homologue est ainsi utilisée en routine pour permettre cette expression à un niveau physiologique grâce au contrôle de promoteurs endogènes (Gavin, Aloy et al. 2006; Hu, Janga et al. 2009). Elle est aussi appliquée chez la souris par la technologie « Knock-in » mais la génération de cellules souches embryonnaires et d'animaux transgéniques est laborieuse. L'expression à un niveau très proche du niveau endogène est devenue également possible dans les cellules de mammifères via la méthode BAC TransgeneOmics basée sur l'utilisation de chromosomes artificiels de bactéries, les BAC (« bacterial artificial chromosome ») (Poser, Sarov et al. 2008). Ces larges transgènes contenant la totalité du locus génomique, incluant le gène ainsi que ses séquences régulatrices endogènes et ses promoteurs naturels sont transfectés de façon stable dans les cellules de mammifères. Cependant, le plus couramment, l'expression de la protéine étiquetée est sous contrôle de promoteurs non endogènes via l'utilisation de plasmides, de vecteurs rétroviraux ou lentiviraux. Les niveaux d'expression ainsi que la localisation doivent alors être vérifiés, une surexpression massive pouvant parfois entrainer une localisation aberrante, une toxicité ou encore une altération des complexes protéiques. Il est également possible de contrôler l'expression de la protéine d'intérêt grâce à l'emploi de promoteurs inductibles, comme le promoteur inductible à la tétracycline (système Tet Off/Tet On).

Les différentes étiquettes peuvent être utilisées seules et permettre de réaliser la purification de la protéine d'intérêt en une seule étape. Elles peuvent également être associées et conduire à l'expression d'une protéine appât avec une double étiquette. Son enrichissement est alors réalisé grâce à deux étapes de purification d'affinité successives, constituant une purification en tandem (ou TAP pour « tandem affinity purification »). Historiquement, cette méthode a été développée par le groupe de B. Séraphin chez la levure (Rigaut, Shevchenko et al. 1999). La double étiquette mise en place, le TAP tag, est composée d'une étiquette Protéine A et d'une étiquette CBP (« calmodulin-binding peptide ») séparées par un court espaceur incluant le site de clivage de la protéase TEV (« Tabacco etch virus ») (Figure 16). La protéine d'intérêt est alors purifiée avec ses interactants dans un premier temps grâce à des billes d'immunoglobuline IgG qui lient la protéine A. Les complexes enrichis sont élués spécifiquement par clivage du TAP tag par la protéase TEV. Une seconde étape de purification d'affinité est enfin réalisée grâce à la seconde étiquette CBP sur des billes greffées avec de la calmoduline en présence de calcium, et le matériel lié est finalement élué avec de l'EGTA. Cette stratégie permet une purification très spécifique du complexe d'intérêt avec un très faible bruit de fond chez la levure (Rigaut, Shevchenko et al. 1999). Elle a été appliquée avec succès pour l'étude de multiples complexes protéiques de levure mais également de complexes issus de nombreux autres organismes dont des bactéries, des plantes, des insectes, des cellules de mammifères (Xu, Song et al. 2010). Le TAP tag présente cependant certaines limitations, en particulier pour l'étude de complexes issus de cellules d'eucaryotes supérieurs bien que certains groupes l'aient utilisé avec succès (Li 2010). Le rendement de purification y est d'une part beaucoup plus faible que celui obtenu chez la levure et la quantité de matériel de départ plus limitée. D'autre part, de nombreuses protéines endogènes de cellules de mammifères se lient avec une forte affinité à la calmoduline (Head 1992; Terpe 2003), rendant l'utilisation des billes de calmoduline peu recommandée pour éviter la

présence d'un important bruit de fond. Par ailleurs, la taille relativement importante du TAP tag (environ 21 kDa) peut perturber la fonction de certaines protéines appât et la formation de leurs complexes. Pour répondre à ces contraintes, des variantes du TAP tag utilisant différentes étiquettes ont alors été développées. Certaines remplacent uniquement le CPB par d'autres étiquettes (FLAG (Knuesel, Wan et al. 2003), StrepII (Forsman, Ruetschi et al. 2008), His (Crawford, Yang et al. 2009)…), d'autres remplacent la protéine A (FLAG (Liang, Yu et al. 2009), His HA (Auty, Steen et al. 2004)…). De nombreux variants de plus petite taille associant plusieurs étiquettes peptidiques ont également été utilisés (FLAG-HA (Ye, Donigian et al. 2004), His-FLAG (Saade, Mechold et al. 2009), FLAG-Strep II (Gloeckner, Boldt et al. 2007)…).

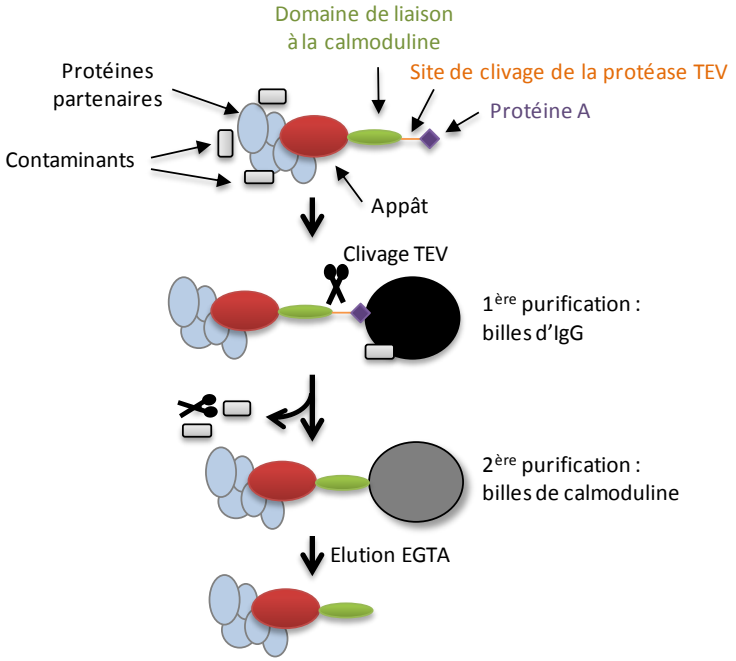


Figure 16 : Structure du TAP-TAG et stratégie de purification. *D'après (Rigaut, Shevchenko et al. 1999).*

III. Identification des partenaires protéiques *bona fide*

Une méthode classique pour l'identification des partenaires protéiques consiste à séparer les composants des complexes protéiques sur gel 1D SDS-PAGE pour réaliser une quantification relative basée sur la coloration du gel de la piste correspondante avec une piste contrôle. Elle permet de visualiser directement les bandes différentielles entre ces deux pistes qui représentent les partenaires protéiques spécifiques et de ce fait de les distinguer des protéines contaminantes. Les bandes de gel correspondantes sont ensuite spécifiquement découpées puis analysées. Cette méthode simple n'est cependant pas idéale. Les complexes protéiques ne sont en effet pas toujours suffisamment abondants et enrichis pour être distingués sur le gel, en particulier lorsqu'un bruit de fond important de protéines contaminantes est présent. Grâce aux développements instrumentaux, les mélanges protéiques issus de ce type d'expériences sont par ailleurs aujourd'hui accessibles aux spectromètres de masse. Il est ainsi possible d'analyser directement les complexes isolés par

spectrométrie de masse via une analyse nanoLC-MS/MS. La quantification est dans ce cas réalisée au niveau MS et permet plus efficacement, grâce à une gamme dynamique plus importante, d'identifier les partenaires protéiques spécifiques même au sein d'un bruit de fond important.

Différentes stratégies ont été mises en place pour analyser les complexes protéiques en nanoLC-MS/MS et identifier les partenaires protéiques spécifiques. Elles ont évoluées parallèlement aux développements des spectromètres de masse et des outils bioinformatiques permettant de traiter les données générées. Des efforts pour réduire ou distinguer les contaminants peuvent être fournis à différents niveaux. On peut d'une part essayer de réduire les contaminants au niveau expérimental dans le but d'obtenir des complexes les plus purs possibles tout en préservant leur intégrité. Pour cela, il est possible de jouer sur la stringence de la purification, en réalisant une purification plus ou moins poussée (en une ou deux étapes) et/ou en modifiant la stringence des lavages (détergents, sels). Le choix de la matrice d'affinité peut de plus aider à réduire le bruit de fond, puisqu'il a été montré que selon sa nature, elle engendre plus ou moins de fixation non spécifiques (Trinkle-Mulcahy, Boulon et al. 2008). Il est d'autre part possible d'intervenir suite à l'acquisition des données pour identifier les interactants spécifiques de la protéine appât. Plusieurs méthodes de traitement des données peuvent alors être envisagées.

III-1. Constitution et utilisation de banques de contaminants

Une méthode pour éliminer le bruit de fond des protéines identifiées lors d'une purification d'affinité consiste à soustraire de la liste de protéines les contaminants fréquemment retrouvés pour une matrice d'affinité donnée. Ces contaminants sont définis comme le protéome des billes ou « bead proteome » (Boulon, Ahmad et al. 2010). Cependant, en filtrant de cette façon les données, seules les protéines se liant de façon non spécifique aux billes peuvent être éliminées, celles qui se fixant à l'anticorps par exemple ne le sont pas. Par ailleurs, une partie du « bead proteome » peut correspondre à des vrais partenaires dans certaines expériences et ils seront écartés via cette méthode (Oeljeklaus, Meyer et al. 2009). Cette stratégie n'est donc pas idéale.

III-2. Comparaison avec un échantillon contrôle

Pour distinguer correctement les vrais partenaires parmi les protéines non spécifiques co-enrichies, il paraît plus rigoureux de réaliser une comparaison entre l'échantillon purifié et un échantillon contrôle, obtenus à partir de deux purifications réalisées en parallèle. Ainsi, les protéines contaminantes présentes sur les billes et sur l'anticorps devraient être retrouvées en quantité similaire dans les deux échantillons, alors que les partenaires spécifiques seront uniquement enrichis dans l'essai. Le choix du contrôle est donc crucial, il doit se rapprocher le plus possible de l'essai afin que le bruit de fond de protéines contaminantes soit identique dans ces deux conditions. Le type de contrôle envisageable dépend du contexte de l'expérience. D'une façon générale, la même matrice doit être employée pour l'enrichissement. Dans le cas de protéines étiquetées, le contrôle idéal correspond à une purification réalisée avec le même type de matrice d'affinité (billes + anticorps) sur des cellules, des tissus ou des organismes n'exprimant pas cette protéine ectopique. En revanche, un contrôle adéquat est plus difficile à trouver dans les cas d'étude de protéines endogènes. Souvent, la purification contrôle est effectuée à l'aide de billes de même nature, mais greffées avec des anticorps non spécifiques (anticorps isotypiques non relevants, sérum pré-immun,..) et à partir du

même extrait cellulaire que l'essai. Dans ce cas, il est cependant possible que le bruit de fond diffère légèrement entre les purifications essai et contrôle puisque des anticorps différents sont utilisés. Le contrôle le plus approprié consiste pour ce type d'expériences à réaliser encore une fois une purification avec le même type de matrice d'affinité (billes + anticorps), mais à partir d'organismes Knock-Out pour la protéine d'intérêt ou de cellules dans lesquelles l'expression de la protéine a été éteinte par siRNA comme proposé dans la méthode QUICK (Selbach and Mann 2006).

La comparaison entre l'échantillon immunopurifié et l'échantillon contrôle peut se faire de différentes façons, par simple comparaison des listes des protéines identifiées dans chacun d'entre eux, ou bien grâce à une analyse protéomique différentielle quantitative.

III-2.1 Comparaison de listes de protéines

Une comparaison des protéines identifiées dans l'échantillon essai et l'échantillon contrôle peut être réalisée. Dans ce cas, les protéines communes aux deux listes peuvent être considérées comme des protéines contaminantes, et celles spécifiques de l'IP comme de probables partenaires spécifiques de l'appât. Cette méthode simple peut être suffisante pour l'étude d'un complexe abondant fortement enrichi, où peu de contaminants sont co-purifiés et où les partenaires spécifiques sont généralement plus intenses que le bruit de fond. Cependant, ce type d'analyse différentielle présente des limites, liées principalement au caractère stochastique du séquençage MS/MS au sein d'un échantillon complexe. Comme décrit précédemment, le caractère partiellement aléatoire de la sélection des pics de faible intensité pour la fragmentation MS/MS entraîne une reproductibilité limitée des listes de protéines identifiées, notamment au niveau des composants minoritaires. De ce fait, la non-identification d'une protéine contaminante au sein d'un échantillon contrôle n'implique pas forcément son absence, entraînant la présence de nombreux faux-positifs au sein de la liste de partenaires potentiels. Inversement, une protéine peut être identifiée par MS/MS dans les deux expériences mais présenter néanmoins un fort enrichissement dans l'IP par rapport au contrôle, et être classée de façon erronée comme protéine contaminante.

III-2.2 Analyse protéomique quantitative différentielle

L'approche la plus performante et la plus largement utilisée aujourd'hui pour réaliser cette comparaison repose sur une analyse protéomique quantitative différentielle des deux échantillons (Dunham, Mullin et al. 2012). Elle peut s'appliquer à des purifications en tandem mais également à des simples purifications, puisque qu'elle permet de gérer un bruit de fond même important. La quantification permet en effet de distinguer les contaminants, qui s'associent de manière similaire à la matrice d'affinité dans les deux conditions (ratio proche de 1), des interactants spécifiques qui eux sont enrichis (ratio > 1) voire détectés comme spécifiques de l'essai (Figure 17).

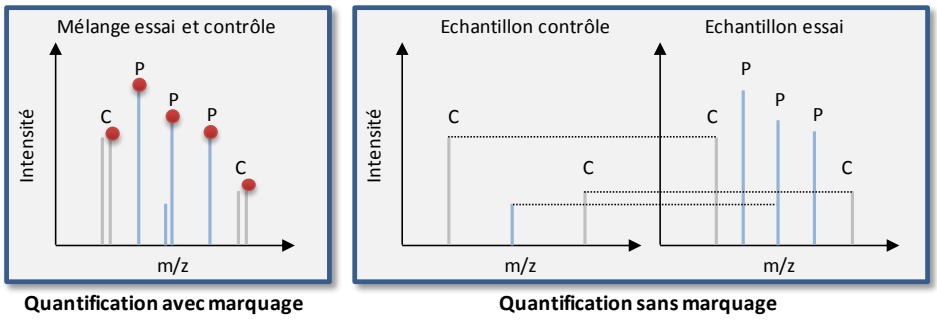


Figure 17 : Quantification avec ou sans marquage entre l'échantillon essai et l'échantillon contrôle pour la discrimination des interactants des protéines contaminantes*. Les protéines contaminantes (C) présentent un ratio de 1 alors que les partenaires de l'appât sont spécifiques ou enrichis (ratio >1) dans l'essai. Le point rouge représente le marquage isotopique (condition lourde) de l'essai.*

Nous avons vu précédemment les différentes stratégies développées pour la quantification relative des protéines, utilisant ou non un marquage isotopique des molécules. Ces différentes méthodes ont été appliquées pour l'étude de nombreux complexes protéiques et ont permis l'identification de protéines partenaires spécifiques.

Dans les études basées sur un marquage isotopique, l'échantillon de départ utilisé pour la purification essai est généralement marqué avec des isotopes lourds et le contrôle avec des isotopes légers. Le marquage est réalisé avant ou après la purification selon le type de marquage utilisé (Figure 18).

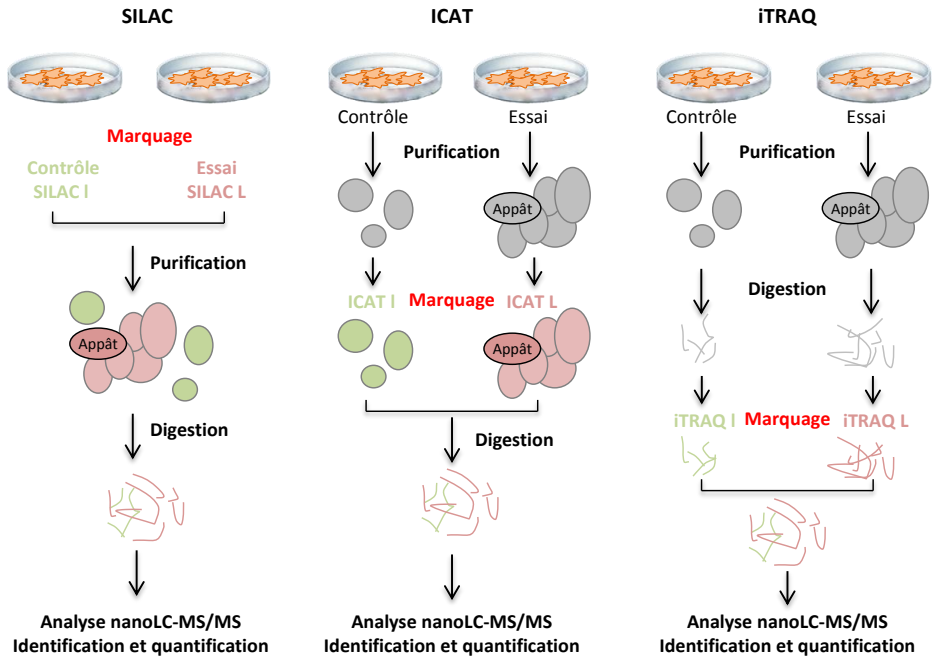


Figure 18 : Stratégies de quantification avec marquage isotopique pour l'analyse nanoLC-MS/MS de complexes protéiques. *L'échantillon contrôle est marqué avec l'isotope léger (l) et l'échantillon essai avec l'isotope lourd (L). La purification est réalisée à différents niveaux selon le marquage isotopique utilisé.*

Grâce à une quantification ICAT, Ranish et al. ont par exemple pu caractériser le complexe de pré-initiation de la transcription de l'ARN polymérase II chez la levure et ainsi identifier 49 partenaires spécifiques, dont 45 sont des composants connus du complexe, parmi plus de 300 protéines (Ranish, Yi et al. 2003). Une quantification basée sur le marqueur iTRAQ a par ailleurs permis l'analyse des interactomes de précurseurs de protéines amyloïdes dans le cerveau de souris (Bai, Markham et al. 2008). Le SILAC constitue également une stratégie quantitative très répandue pour l'analyse de complexes protéiques. Blagoev et al., ont été les premiers à la mettre en place pour la caractérisation de complexes de signalisation impliquant la voie de l'EGFR (« epidermal growth factor receptor »). Une purification par affinité de la forme activée phosphorylée de l'EGFR à partir de cellules HeLa stimulées ou non à l'EGF et différentiellement marquées a permis l'identification de composants spécifiques du complexe de signalisation EGF-dépendant parmi un large bruit de fond (Blagoev, Kratchmarova et al. 2003). Ce type de quantification a par ailleurs été associé à un enrichissement en une étape de protéines appât étiquetées GFP exprimées à un niveau endogène dans la méthode QUBIC (Quantitative BAC Interactomics). Elle a notamment permis la mise en évidence d'interactants des complexes TREX impliqués dans l'export des ARNm, ainsi que ceux de la protéine CDC23, un composant du complexe « anaphase-promoting complex », APC (Hubner, Bird et al. 2010). Selbach et al., ont quant à eux développé la stratégie QUICK (« quantitative immunoprecipitation combined with knockdown ») pour l'immunopurification de protéines endogènes, utilisant le SILAC pour réaliser une quantification différentielle entre un échantillon essai et un échantillon contrôle dans lequel l'expression de la protéine appât a été éteinte par RNA interference. Ses performances ont été démontrées à travers l'étude des complexes impliquant la β-caténine et la protéine Cbl dans des cellules de mammifères (Selbach and Mann 2006).

Dans ces stratégies de quantification, les échantillons sont rassemblés avant ou après la purification selon le niveau auquel est réalisé le marquage. Le marquage SILAC présente l'avantage de permettre la réunion des échantillons en amont de l'isolement des complexes protéiques (au niveau cellulaire). Ainsi, une extraction unique et une purification unique sont réalisées permettant de limiter les artéfacts liés à un traitement parallèle des échantillons. Cependant, au cours de cette purification, un échange des partenaires lourds et légers peut avoir lieu et induire la perte du ratio différentiel des protéines interagissant de façon dynamique et transitoire au sein des complexes (Wang and Huang 2008). Dans ce type d'expérience SILAC (PAM-SILAC, « Purification after mixing ») (Figure 19A), des complexes lourds sont enrichis et les protéines lourdes nouant des interactions dynamiques se trouvent impliquées dans un équilibre avec le complexe au sein d'un mélange où se trouve également présente la version légère de ces protéines. Un échange entre ces deux formes a ainsi lieu, ce qui peut conduire à terme à un ratio lourd/léger de 1. Ces partenaires dynamiques sont donc quantifiés parmi les protéines contaminantes. Pour parvenir à les identifier, il est nécessaire d'utiliser des approches SILAC dans lesquelles les échantillons sont cette fois rassemblés seulement après purification (MAP-SILAC, « mixing after purification »), où aucun mélange protéique ne peut avoir lieu (Figure 19B). Wang et al., ont de cette façon pu mettre en évidence 14 partenaires potentiels supplémentaires du protéasome 26S humain, préalablement désignés comme protéines non spécifiques par le PAM-SILAC (Wang and Huang 2008). Mousson et al., ont également pu par cette méthode souligner le caractère dynamique de la protéine BTAF1 au sein des complexes de la TATA-binding protein (TBP) (Mousson, Kolkman et al. 2008). La comparaison des partenaires protéiques identifiés par le PAM-SILAC et le MAP-SILAC permet ainsi de mettre en évidence les

propriétés dynamiques de certains interactants permettant d'avancer encore dans la compréhension du fonctionnement des complexes protéiques.

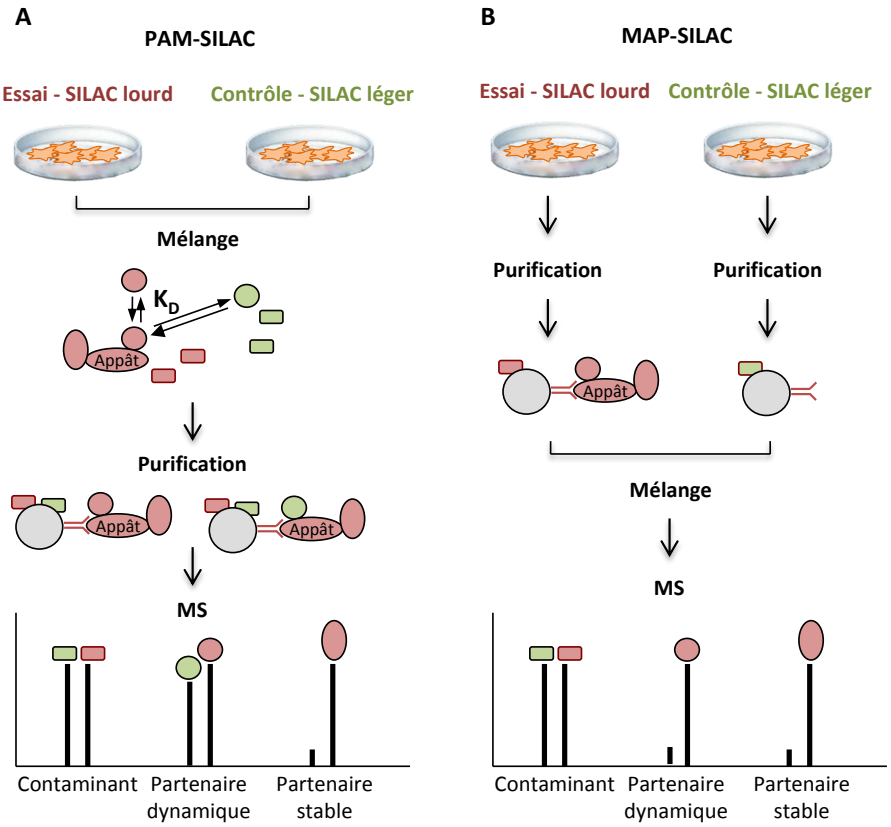


Figure 19 : Approches PAM-SILAC et MAP-SILAC : identification des partenaires protéiques dynamiques*. D'après (Mousson, Kolkman et al. 2008).*

Les stratégies de quantification sans marquage, basées sur le « spectral counting » ou la mesure des intensités des signaux MS des peptides d'une protéine, sont de plus en plus utilisées pour l'étude de complexes protéiques afin de discriminer les vrais partenaires d'interaction des protéines du bruit de fond. En effet, un avantage majeur de ces approches est leur facilité de mise en œuvre, évitant l'introduction parfois délicate de marqueurs isotopiques. Leur principal inconvénient réside dans le traitement parallèle des échantillons à comparer, qui peut conduire à l'introduction d'artéfacts. Cependant, comme vu précédemment, même dans le cas d'un marquage SILAC, ce traitement parallèle doit souvent être malgré tout réalisé au moins jusqu'au stade de l'obtention de complexes purifiés pour éviter de perdre les différentiels sur les interactions dynamiques. L'utilisation de méthodes quantitatives sans marquage apparaît d'autant plus simple et directe, même si celles-ci ne permettent pas de réduire la variabilité potentiellement présente au cours de la phase analytique. En effet, dans ce cas les échantillons sont traités en parallèle tout au long du protocole d'étude, de l'isolement des complexes protéiques jusqu'à leur analyse nanoLC-MS/MS (Figure 20).

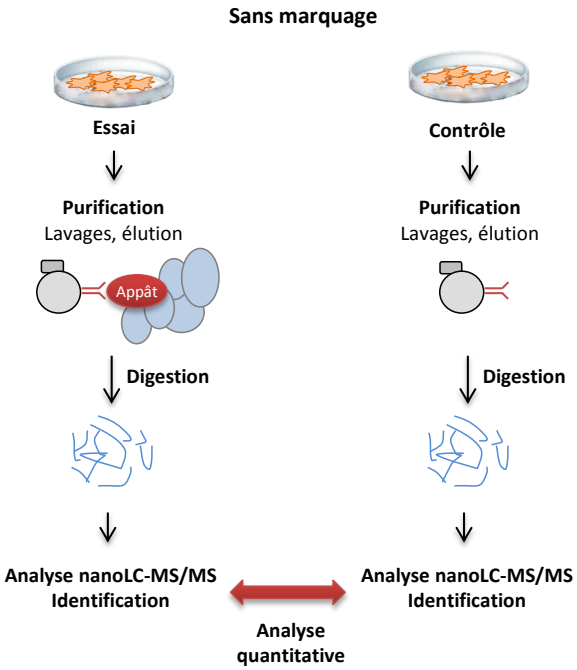


Figure 20 : Stratégie de quantification sans marquage pour l'analyse nanoLC-MS/MS de complexes protéiques.

La quantification des protéines dans une analyse AP-MS peut être effectuée par « spectral counting ». Florens et al., utilisent par exemple une valeur de nombre de spectres normalisée, le NSAF (« Normalized spectral abundance factor ») correspondant au nombre total de MS/MS réalisées sur une protéine normalisé par la longueur de la protéine et le nombre total de spectres MS/MS réalisés lors de l'analyse (Florens, Carozza et al. 2006). Mosley et al., ont quant à eux pu distinguer les partenaires spécifiques des ARN polymérases de levure du bruit de fond en calculant une variante du NSAF, le dNSAF (distributed NSAF) (Mosley, Sardiu et al. 2011). La quantification sans marquage utilisant la mesure des intensités des signaux MS des peptides a été également utilisée pour l'analyse de complexes protéiques. Elle a été par exemple introduite dans une méthode développée par Rinner et al., permettant l'identification d'interactants spécifiques (Rinner, Mueller et al. 2007). Elle est basée sur une dilution séquentielle des complexes immunopurifiés avec l'échantillon contrôle. Dans cette approche, quatre conditions correspondant à des dilutions successives contenant 0%, 10%, 20% et 100% d'échantillons immunopurifiés sont soumises à une analyse quantitative sans marquage à l'aide du logiciel SuperHirn : l'appât ainsi que ses partenaires spécifiques apparaissent ainsi progressivement enrichis dans la série, alors que les protéines contaminantes, présentes en quantité identique dans l'échantillon immunopurifié et dans l'échantillon contrôle, ont en revanche une intensité constante dans les différentes dilutions. Le suivi des profils des intensités peptidiques mesurées sur la série de dilutions permet ainsi la discrimination des interactants spécifiques. Ces auteurs ont de cette façon pu caractériser des partenaires déjà connus ainsi que de nouveaux partenaires de la protéine FoxO3A (Rinner, Mueller et al. 2007). La méthode QUBIC a également été adaptée pour la quantification sans marquage. La comparaison quantitative des intensités des signaux MS issus d'échantillons immunopurifiés et de contrôles à l'aide du logiciel MaxQuant permet d'après Hubner et al., de discriminer les partenaires des contaminants de façon aussi efficace que le QUBIC-SILAC (Hubner, Bird et al. 2010). Ils ont, par cette

méthode, étudié différents complexes protéiques dont le complexe APC (anaphase-promoting complex), pour lequel de nouveaux composants potentiels ont été mis en évidence.

Ces différentes études soulignent la grande efficacité de la quantification différentielle, avec ou sans marquage, pour distinguer les vrais partenaires protéiques spécifiquement enrichis même au sein d'un important bruit de fond de protéines contaminantes co-purifiées. Il n'est donc plus absolument nécessaire d'obtenir des complexes très purs pour identifier de nouveaux partenaires protéiques, et des conditions d'enrichissement plus douces, comme la purification en une seule étape ou l'utilisation de conditions de lavages non stringentes, peuvent être envisagées afin d'augmenter les chances de conserver des interactions biologiquement pertinentes mais faibles, transitoires ainsi que des interactants faiblement abondants.

Partie III. L'analyse de protéomes complexes

Pour répondre à certaines questions biologiques, il est parfois nécessaire d'obtenir une vision globale des protéines présentes au sein d'un échantillon, et de suivre leurs variations sous certaines conditions données. Cette étude globale et sans *a priori* de protéomes entiers ainsi que de leurs variations d'abondance constitue un autre domaine de la protéomique, la protéomique d'expression. Elle permet de comprendre le fonctionnement de systèmes biologiques et de mécanismes cellulaires complexes, de caractériser la réponse cellulaire à certains stimuli ou perturbations environnementales ou encore d'identifier des biomarqueurs protéiques signant une pathologie.

I. Stratégie générale pour l'étude de mélanges protéiques complexes et défis associés

L'étude protéomique de ces protéomes complexes peut être réalisée grâce à analyse nanoLC-MS/MS. Les échantillons biologiques sont pour cela lysés, les mélanges protéiques ainsi obtenus sont ensuite digérés par une protéase spécifique puis les mélanges peptidiques résultants sont analysés par nanoLC-MS/MS. Pour mettre en évidence les variations d'expression des protéines, une comparaison de deux ou plusieurs échantillons doit ensuite être réalisée via une analyse protéomique quantitative différentielle des données nanoLC-MS/MS obtenues, à l'aide de méthodes avec ou sans marquage isotopique.

Dans ce type d'étude, le défi majeur consiste à parvenir à couvrir la plus large partie du protéome possible, ce qui implique la détection des protéines faiblement abondantes et minoritaires. Ce défi est lié la très grande complexité des protéomes, en particulier des protéomes des eucaryotes supérieurs qui sont par conséquent difficiles à analyser.

II. Complexité des mélanges protéiques analysés

Le nombre d'espèces protéiques présentes dans une cellule peut être très important, en particulier chez les eucaryotes supérieurs. Le séquençage de génome humain a permis d'identifier environ 20 000 gènes chez l'homme mais le nombre de formes protéiques associées est potentiellement beaucoup plus important. Il est difficile de l'estimer correctement mais on considère qu'il peut aller de 20 000 protéines si l'on considère qu'un gène code pour une protéine, à plusieurs millions si l'on prend en compte les isoformes protéiques (épissage alternatif), le polymorphisme peptidique ainsi que les modifications post-traductionnelles des protéines (Cox and Mann 2011). Toutes ces formes ne sont cependant pas exprimées en même temps dans une cellule donnée, et les protéomes associés aux différents types cellulaires sont donc très divers et hétérogènes. Le protéome présente par ailleurs une très large gamme dynamique d'expression des protéines. Certaines sont en effet présentes à raison de quelques copies dans la cellule lorsque d'autres sont très abondantes. Le rapport entre les protéines les moins abondantes et les plus abondantes dans

une cellule dépasse 6 ordres de grandeur et atteint 10^{10}-10^{12} dans certains fluides biologiques comme le plasma (Corthals, Wasinger et al. 2000; Anderson and Anderson 2002).

Au vu de cette complexité, l'étude de protéomes entiers par nanoLC-MS/MS représente un véritable défi pour identifier le plus grand nombre de protéines et détecter les protéines plus faiblement représentées. La capacité à y parvenir est largement dépendante des développements des technologies instrumentales. Malgré les avancées très importantes de ces dernières années qui ont permis de considérablement augmenter le nombre de protéines identifiables, ces technologies présentent une gamme dynamique et une capacité de détection encore limitées, en tout cas insuffisantes pour caractériser de façon exhaustive des protéomes très complexes.

Il est traditionnellement admis que les approches nanoLC-MS/MS sont capables de détecter des protéines dans une gamme dynamique autour de 10^{2}-10^{4} (comme montré en 2006 par de Godoy et al., avec un spectromètre de masse FT-ICR (de Godoy, Olsen et al. 2006)). Elle est inférieure à la gamme dynamique d'expression des protéines observée dans les échantillons biologiques et pénalise par conséquent la couverture analytique du protéome. Cette couverture est dépendante de différents paramètres incluant la sensibilité, la vitesse de séquençage et la gamme dynamique propre des instruments MS utilisés ainsi que du pouvoir séparatif de la LC (Figure 21).

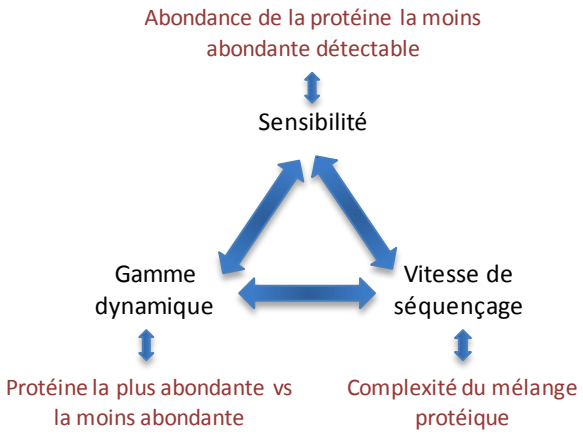


Figure 21 : Paramètres des spectromètres de masse affectant la profondeur d'analyse du protéome. *Ces trois paramètres sont interdépendants et affectent à plusieurs niveaux l'analyse de l'échantillon biologique. D'après (de Godoy, Olsen et al. 2006).*

La sensibilité intrinsèque d'un spectromètre de masse peut être définie comme la concentration limite à laquelle un peptide donné, analysé seul, est détecté. Elle atteint des niveaux très bas, de l'ordre de la dizaine d'attomoles sur les appareils récents comme le LTQ-Orbitrap Velos (Olsen, Schwartz et al. 2009). En pratique cependant, des peptides présents à ces concentrations au sein de mélanges complexes ne seront pas détectés. La sensibilité effective de la technique sur un mélange complexe (définie comme la concentration des protéines les plus faiblement abondantes réellement détectées) est généralement moins bonne, et dépendante des autres paramètres de gamme dynamique et vitesse de séquençage. Dans une étude réalisée en 2006 en couplage nanoLC-MS/MS sur un appareil de type FT-ICR, de Godoy et al., ont analysé précisément l'influence de ces trois paramètres lors de l'analyse d'un lysat de levure (de Godoy, Olsen et al. 2006). Lors de l'analyse

de mélanges peptidiques complexes, de nombreux peptides co-éluent de la colonne chromatographique et sont par conséquent analysés simultanément en MS, puis seulement certains d'entre eux sont sélectionnés pour être fragmentés en MS/MS. Pour évaluer l'impact de la vitesse de séquençage du spectromètre de masse sur la couverture du protéome, la fragmentation de paires peptidiques SILAC léger/lourd d'abondance égale (qui devraient donc être fragmentées de la même façon) a été suivie. Dans 40% des cas, seule l'une des deux formes a été séquencée, attestant d'une vitesse de séquençage insuffisante pour fragmenter tous les signaux des paires peptidiques. De plus, il a été montré que l'abondance des peptides influence cette fragmentation. En effet, 60% des peptides de faible abondance (retrouvés dans moins de 2 scans MS consécutifs) ne sont pas sélectionnés pour être fragmentés et seuls 20 % d'entre eux sont séquencés sur leurs 2 formes SILAC. En revanche, pour des peptides abondants (retrouvés dans plus de 5 scans MS consécutifs), une part plus importante (44%) est séquencée dans les deux formes et seuls 38 % ne sont pas séquencés du tout. Par ailleurs, il apparaît que la gamme dynamique effective de la méthode (10^2 pour l'expérience SILAC) est inférieure à la gamme dynamique d'expression des protéines de la levure (10^4) et pénalise de fait la profondeur d'analyse. Cette gamme réduite s'explique par un effet de suppression du signal, qui conduit à l'incapacité de détecter le signal MS d'espèces faiblement abondantes en présence de peptides abondants, empêchant donc l'identification de nombreuses protéines minoritaires. La vitesse de séquençage et la gamme dynamique constituent ainsi les deux paramètres qui limitent la couverture analytique du protéome et conduisent à une sensibilité effective de l'instrument très inférieure à sa sensibilité intrinsèque. Dans cette étude, la gamme dynamique de l'analyse a par contre été améliorée (10^3) en préfractionnant le lysat de levure (non marqué) sur un gel 1D SDS-PAGE avant l'analyse nanoLC-MS/MS.

Pour contourner ces limites et améliorer la couverture du protéome, il est donc nécessaire de réduire le nombre de peptides analysés simultanément par le spectromètre de masse. Plusieurs solutions peuvent alors être envisagées, à la fois au niveau instrumental et au niveau de la préparation des échantillons. Les performances séparatives de la chromatographie liquide nanoLC couplée au spectromètre de masse peuvent d'une part être optimisées. Il est d'autre part possible de réduire la complexité et la gamme dynamique des mélanges à étudier, par exemple en les fractionnant, afin qu'ils soient plus facilement analysables.

III. Amélioration de la séparation LC pour une meilleure couverture du protéome

Pour augmenter la couverture du protéome, la séparation des peptides par la LC peut être améliorée afin de diminuer le nombre de peptides élués et analysés en simultané dans le spectromètre de masse. Traditionnellement, les séparations chromatographiques en couplage avec l'analyse MS/MS sont réalisées sur des colonnes de phase inverse assez courtes (15cm en routine), habituellement remplies de particules de 3 à 5 µm, sur des systèmes HPLC conventionnels dont la pression d'opération classique se situe aux alentours de 100 à 200 bars (pression d'opération limite de 400 bars). Différentes études ont étudié l'influence de ces paramètres sur le pouvoir résolutif de la chromatographie (Liu, Finch et al. 2007; Eeltink, Dolman et al. 2009; Kocher, Swart et al. 2011; Iwasaki, Sugiyama et al. 2012). La métrique communément utilisée pour cette évaluation est la capacité de pic (PC, « Peak Capacity ») qui mesure le nombre de pics maximum pouvant être séparés

dans le temps chromatographique sur une colonne donnée. Köcher et al. ont mis en évidence une relation linéaire entre la capacité de pic et le nombre de peptides (et de protéines) identifiés en nanoLC-MS/MS (Kocher, Swart et al. 2011). L'amélioration de la PC permet donc d'approfondir la couverture du protéome obtenue. Le temps de gradient est l'un des paramètres influant sur la PC. Dans cette étude, il a en effet été montré, en étudiant un extrait total de cellules HeLa, qu'elle augmentait avec la longueur du gradient (1h à 10h) suivant une courbe logarithmique et est donc améliorée jusqu'à un certain niveau (8h). Cependant, il conduit en parallèle à un élargissement des pics d'élution qui deviennent moins intenses, ce qui affecte par conséquent la sensibilité analytique. Un compromis doit donc être trouvé entre sensibilité et efficacité de séparation. L'utilisation de colonnes plus longues (50 cm vs 15 cm ou 30 cm) permet d'augmenter encore la capacité de pic (Liu, Finch et al. 2007) et présente de plus l'avantage de permettre l'application de gradients longs sans compromettre la résolution des pics chromatographiques et donc la sensibilité analytique (Kocher, Swart et al. 2011). Enfin, la séparation des peptides peut être améliorée (PC augmentée de 50%) en diminuant la taille des particules de silice de la colonne de 3 µm à 1,7 µm (Liu, Finch et al. 2007). Ainsi, des gradients plus longs, des colonnes plus longues remplies de particules plus fines peuvent être utilisés et combinés pour augmenter le pouvoir séparatif de la chromatographie qui résulte à terme en une augmentation de la gamme dynamique d'analyse et de fait de la couverture du protéome (Kocher, Swart et al. 2011). Cependant, cela résulte également en une augmentation de la pression de la colonne incompatible avec les systèmes chromatographiques classiques. Pour contourner ce problème, Thakur et al., ont, dans une étude récente, réalisé la séparation à une température supérieure à la température ambiante. Sur une HPLC conventionnelle, ces auteurs ont utilisé une colonne de 50 cm remplie avec des particules de silice de 1,8 µm et un gradient de 8 h, chauffée de façon à diminuer la pression. Grâce au pouvoir résolutif de ce système, environ 2400 protéines ont pu être identifiées en une seule analyse sur un extrait de levure, et environ 4700 protéines sur une lignée cellulaire humaine avec un LTQ-Orbitrap Velos. Dans cette dernière analyse, une gamme dynamique estimée à 6 ordres de grandeur a été atteinte (Thakur, Geiger et al. 2011). Une autre approche consiste à utiliser ces colonnes longues sur des systèmes chromatographiques fonctionnant à haute pression (UPLC, « Ultra Performance Liquid Chromatography ») dont la pression d'opération limite se situe aux alentours de 800 bars. Très récemment, une couverture encore plus importante du protéome de levure a ainsi été obtenue (environ 3900 protéines identifiées par analyse nanoLC-MS/MS) par couplage d'une UPLC avec un Q-Exactive, en utilisant une colonne de 50cm remplie avec des particules de 1,8µm et un gradient de 4h (Nagaraj, Kulak et al. 2012). L'amélioration de la séparation des peptides en amont de la MS permet ainsi de simplifier efficacement le mélange peptidique et d'augmenter le nombre de protéines identifiées et par conséquent la couverture du protéome.

Ces stratégies récentes sont très prometteuses mais ne sont jusqu'à présent pas aisément applicables en routine. Elles nécessitent en effet d'une part des systèmes chromatographiques spécifiques fonctionnant à haute pression. L'utilisation des UPLC commence cependant à se démocratiser, bien que leur robustesse reste encore à être évaluée en routine. La reproductibilité d'acquisitions nanoLC-MS/MS aussi longues doit de plus être vérifiée. D'autre part, une difficulté majeure de ce type de stratégies pourrait être liée au traitement des données brutes très lourdes qui sont générées. L'analyse et le traitement bioinformatiques de celles-ci (interrogation en banque de données, quantification) peuvent alors être problématiques, les outils bioinformatiques

classiquement utilisés n'étant pas nécessairement adaptés et de fait capables de les supporter. Ils devront donc encore évoluer pour pouvoir facilement prendre en charge ce type de données.

IV. Simplification des mélanges protéiques complexes pour une meilleure couverture du protéome

Les approches classiques pour améliorer la profondeur d'analyse d'échantillons protéiques complexes consistent soit à le fractionner de façon à analyser un certain nombre de fractions simplifiées, soit à sélectionner par différents moyens biochimiques un sous-ensemble particulier de protéines. Cela peut être réalisé à différents stades de la préparation des échantillons :

- Au niveau protéique, en réalisant un fractionnement des protéines de l'échantillon par des techniques séparatives d'électrophorèse ou de chromatographie
- Au niveau peptidique, en utilisant des techniques séparatives dédiées au fractionnement du mélange peptidique
- Lors de la préparation de l'échantillon protéique, en diminuant sa complexité (déplétion des protéines très abondantes, enrichissement en protéines minoritaires, fractionnement subcellulaire, enrichissement d'un sous-protéome d'intérêt).

IV-1. Fractionnement protéique et peptidique

L'une des approches pour simplifier les mélanges à analyser consiste à fractionner celui-ci en plusieurs fractions. Chacune de ces fractions présente alors une complexité moindre, en termes de nombre d'espèces présentes, ainsi qu'une gamme dynamique réduite, ce qui permet *in fine* l'identification de protéines faiblement abondantes. La plus grande profondeur d'analyse obtenue se fait cependant au dépend du temps d'analyse qui augmente en effet proportionnellement avec le nombre de fractions à analyser. Ce fractionnement peut être réalisé à différents niveaux, peptidique ou protéique. Plusieurs méthodes sont dans ces deux cas disponibles.

IV-1.1 Fractionnement au niveau peptidique

Les peptides présentent une grande diversité physico-chimique (charge, point isoélectrique, hydrophobicité, taille) qui peut être utilisée pour les fractionner. En plus de la LC en phase inverse (RP-LC) systématiquement mise en œuvre dans l'analyse par couplage LC-MS/MS, il est possible de fractionner au préalable le mélange peptidique. Les études protéomiques à large échelle ont débuté avec la mise en place de la stratégie MudPIT (« Multidimensional chromatography Peptide Identification Technology ») qui consiste en une séparation multidimensionnelle orthogonale des peptides, faisant intervenir une double chromatographie en tandem, une phase échangeuse de cations (SCX, « Strong cation exchange ») en ligne avec la phase inverse (Washburn, Wolters et al. 2001; Wolters, Washburn et al. 2001). Ce couplage peut également être réalisé « off-line ». D'autres configurations couplent, en ligne ou non, différents types de chromatographie avec la phase inverse (Fournier, Gilmore et al. 2007). Une phase échangeuse d'anions, SAX (« Strong anion exchange ») peut par exemple être utilisée en amont de la RP-LC (Wisniewski, Zougman et al. 2009). Les peptides

peuvent également être séparés en milieu liquide en fonction de leur point isoélectrique, grâce à un gradient de pH immobilisé sur une bandelette de gel (strip) ou par « Off-Gel isoelectrofocusing » (OGE) (Horth, Miller et al. 2006). Ces méthodes de fractionnement peptidique nécessitent une digestion préalable des mélanges protéiques en milieu liquide et la reprise de l'échantillon protéique dans un milieu compatible avec l'activité de la trypsine. Dans ces conditions, certaines protéines comme les protéines hydrophobes (protéines membranaires) peuvent être insolubles et donc perdues.

IV-1.2 Fractionnement au niveau protéique

Les échantillons complexes à analyser peuvent par ailleurs être fractionnés au niveau protéique. Les approches électrophorétiques sont probablement pour cela les plus communément utilisées. Le gel SDS-PAGE ou gel monodimensionnel (gel 1D) constitue l'une des méthodes les plus robustes généralement associée avec la LC-MS/MS (approche GeLC-MS/MS, (de Godoy, Olsen et al. 2006; Dejardin, Guillou et al. 2009; von Pfeil, Decamp et al. 2009). Il permet la séparation des protéines en fonction de leur poids moléculaire quelques soient leurs propriétés physico-chimiques. Grâce à la présence du SDS (« sodium dodecyl sulfate »), les protéines, même les plus hydrophobes comme les protéines membranaires, sont en effet correctement solubilisées et donc fractionnées. Dans cette stratégie simple et facile à mettre en œuvre, le gel 1D est découpé en différentes bandes, dont le nombre dépend de l'importance du fractionnement désiré (généralement une dizaine de bandes). Les protéines concentrées dans une bande de gel sont ensuite analysées indépendamment par LC-MS/MS. D'autres méthodes ont été développées comme l'OGE qui permet la séparation des protéines en milieu liquide en fonction de leur pI sans passer par un gel.

Les performances de ces différents systèmes de fractionnement au niveau protéique ou peptidique ont été évaluées dans un certain nombre d'études (Fang, Robinson et al. 2010; Antberg, Cifani et al. 2012). Fang et al., ont comparé les trois méthodes de fractionnement les plus employées ces dernières années en protéomique nanoLC-MS/MS, le gel SDS-PAGE (protéines), la SCX (peptides) et l'OGE (protéines et peptides). Dans cette étude, le fractionnement protéique sur gel SDS-PAGE a conduit à l'identification d'un plus grand nombre de protéines (Fang, Robinson et al. 2010). Des résultats similaires ont été obtenus dans l'étude réalisée par Antberg et al., (Antberg, Cifani et al. 2012). L'étude la plus complète et la plus profonde de protéomes humains, qui a abouti à l'identification de 7000 à 8500 protéines dans une lignée cellulaire, est quant à elle basée sur un fractionnement par SAX des mélanges peptidiques en 6 fractions. Chaque fraction a été analysée par nanoLC-MS/MS sur un LTQ-Orbitrap Velos avec un analyseur Orbitrap dernière génération et un gradient chromatographique de 200min. Au total, onze lignées ont été analysées et ont permis l'identification de plus de 11000 protéines (Geiger, Wehner et al. 2012).

A l'heure actuelle, les approches nanoLC-MS/MS impliquant un fractionnement préalable de l'échantillon par gel 1D, SCX ou SAX, semblent ainsi performantes et permettent l'obtention d'une couverture de protéome relativement importante.

IV-2. Déplétion et « égalisation »

Un autre moyen de diminuer la complexité des échantillons et de réduire leur gamme dynamique consiste à déplèter sélectivement les protéines majoritaires présentes, afin d'identifier les plus faiblement représentées. Cela est en particulier utile lors d'études de fluides biologiques, comme le plasma qui présente une gamme dynamique de concentrations protéiques très large et pour lequel les 22 protéines les plus abondantes représentent 99% de la masse protéique totale de l'échantillon (Ray, Reddy et al. 2011). Des méthodes de déplétion par immunoaffinité de ces protéines sont pour cela généralement utilisées. Elles impliquent l'utilisation d'anticorps immobilisés sur billes, dirigés contre un certain nombre de protéines très abondantes du plasma (entre 1 et 20). Plusieurs études comparatives de l'effet de différentes colonnes d'immunodéplétion commerciales ont souligné le bénéfice apporté par celles-ci pour l'identification d'un plus grand nombre de protéines, dont des protéines minoritaires (Roche, Tiers et al. 2009; Tu, Rudnick et al. 2010). Dans ces méthodes de déplétion, certaines protéines minoritaires risquent cependant d'être perdues, co-déplétées avec les protéines abondantes.

La gamme dynamique de mélanges complexes peut également être diminuée grâce à l'application de la technique « Proteominer » (Thulasiraman, Lin et al. 2005; Guerrier, Thulasiraman et al. 2006), basée sur l'utilisation d'une banque de ligands peptidiques hexamériques obtenus par chimie combinatoire et fixés sur billes. Etant donné la très grande diversité de ligands présents, chaque protéine d'un mélange complexe peut théoriquement interagir avec une ou plusieurs billes. Les protéines abondantes, qui saturent rapidement leurs billes de reconnaissance, sont ainsi en partie déplétées, contrairement aux protéines minoritaires qui sont progressivement enrichies sur leurs billes respectives. Cette stratégie a permis de réduire efficacement la gamme dynamique et ainsi d'approfondir la couverture de protéomes du sérum (Sennels, Salek et al. 2007), de globules rouges (Roux-Dalvai, Gonzalez de Peredo et al. 2008) ou encore du fluide cérébrospinal (Mouton-Barbosa, Roux-Dalvai et al. 2010).

IV-3. Enrichissement de sous-protéomes d'intérêt

Afin de simplifier l'échantillon, il est par ailleurs possible de restreindre l'analyse à un sous-protéome particulier d'intérêt. Leur composition protéique est largement moins complexe en comparaison des protéomes entiers de cellules ou de tissus et ils sont donc plus facilement analysables en MS. Ils peuvent d'une part correspondre à des structures sub-cellulaires comme des compartiments de la cellule (noyaux, cytoplasmes) ou des organelles (mitochondries, réticulum endoplasmique…), qui sont généralement séparés en fonction de leur densité par centrifugation différentielle ou centrifugation en gradient de densité (Lee, Tan et al. 2010). Ils peuvent d'autre part être constitués d'une classe particulière de protéines comme les phosphoprotéines ou les glycoprotéines, si l'objectif de l'étude implique l'identification à large échelle de protéines modifiées. Une grande diversité de méthodes existe pour l'étude des modifications post-traductionnelles à large échelle. Elles ne seront pas détaillées ici car elles n'ont pas été généralement utilisées dans le cadre de cette thèse. Nous nous sommes cependant spécifiquement intéressés à l'analyse des protéines glycosylées. Les méthodes dédiées à leur enrichissement sont abordées dans la partie Résultats de ce manuscrit (Partie II-I). D'une façon générale, l'enrichissement de protéines modifiées est réalisé en exploitant l'affinité que présentent ces protéines pour une matrice particulière. Des anticorps dirigés

contre certaines modifications (lysines acétylées, tyrosines phosphorylées) peuvent ainsi être utilisés. Par ailleurs, les modifications post-traductionnelles sont parfois modifiées, étiquetées *in vitro* par dérivation chimique ou bien *in vivo* par marquage métabolique, afin d'être ensuite purifiées par chromatographie d'affinité. Elles peuvent enfin être enrichies grâce à leurs caractéristiques chimiques. L'enrichissement des phosphopeptides utilise par exemple les interactions ioniques que le groupement phosphate peut nouer grâce à sa charge négative (Zhao and Jensen 2009).

Ces différentes stratégies de simplification de l'échantillon permettent ainsi d'augmenter la gamme dynamique de l'analyse nanoLC-MS/MS, le nombre de protéines identifiées et ainsi d'obtenir une meilleure couverture des protéomes complexes. Elles peuvent être associées afin de simplifier davantage les mélanges protéiques, les méthodes de déplétion et d'égalisation pouvant en particulier être couplées à des méthodes de fractionnement protéique et peptidiques.

V. Quantification des mélanges protéiques complexes et couverture du protéome

Comme mentionné précédemment, la protéomique d'expression s'intéresse aux variations d'abondance des protéines engendrées par différents stimuli ou conditions environnementales, au sein de protéomes souvent complexes. Une analyse protéomique quantitative doit alors être réalisée pour mettre en évidence ces variations, à l'aide d'une méthode basée sur un marquage isotopique des échantillons, ou sans marquage. Le choix de la stratégie a un impact sur la profondeur du protéome obtenu. En effet, la quantification « label-free », moins précise que les méthodes de quantification avec marquage isotopique, semble cependant présenter un avantage en termes de couverture analytique (Bantscheff, Schirle et al. 2007). Comme décrit précédemment, les approches avec marquage impliquent en effet un rassemblement des divers échantillons à comparer, plus ou moins tôt dans le processus, ce qui résulte en une complexification de l'échantillon. De plus nombreuses espèces peptidiques sont alors éluées et analysées dans une même fenêtre de temps puisque la très grande majorité des peptides sont présents sous deux formes isotopiques d'abondance équivalente. Le spectromètre de masse passe alors plus de temps à fragmenter ces formes, au détriment d'autres peptides pouvant provenir d'espèces minoritaires. La quantification sans marquage dans laquelle les échantillons peptidiques sont analysés séparément peut conduire de ce fait à l'identification d'un plus grand nombre de protéines. L'étude du protéome de cellules humaines réalisée à la fois en « label-free » et en SILAC par Thakur at al., abonde dans ce sens, le « label-free » permettant l'identification de plus de 20 % de protéines supplémentaires (Thakur, Geiger et al. 2011). En revanche, la quantification sans marquage est plus sensible à des biais qui peuvent être introduits lors des différentes étapes du processus d'analyse, et influer sur la précision de la quantification. Il est donc crucial de limiter au maximum ces artéfacts, et de traiter les données post-acquisition afin de corriger les variations éventuelles dues à la préparation de l'échantillon ou à l'analyse nanoLC-MS/MS.

PRESENTATION GENERALE DES TRAVAUX

La protéomique est aujourd'hui l'outil incontournable pour l'étude des protéines, aussi bien pour l'analyse de modifications post-traductionnelles que pour la caractérisation de complexes protéiques et l'étude de protéomes complexes et de leurs variations. Au cours de ma thèse, j'ai donc utilisé différentes approches protéomiques pour répondre à plusieurs questions biologiques.

D'un point de vue biologique, la thématique de recherche principale de ce manuscrit concerne l'étude des cellules endothéliales et des processus inflammatoires au sein de ces cellules. Cette thématique regroupe différents projets en collaboration avec l'équipe « Biologie Vasculaire : cellules endothéliales, inflammation et cancer » dirigée par Jean-Philippe Girard à l'IPBS. Ce groupe s'intéresse depuis plusieurs années à la biologie et à la différenciation des cellules endothéliales, et en particulier à la fonctionnalité particulière des cellules endothéliales dites « cuboïdales ». Présentes dans les vaisseaux post-capillaires de type HEV (« high endothelial veinules »), elles assurent le recrutement des lymphocytes au niveau des organes lymphoïdes et des tissus subissant une inflammation chronique. Plusieurs gènes d'intérêt ont été découverts à partir de ces cellules HEV, codant notamment pour de nouveaux facteurs nucléaires comme la famille des facteurs THAP ou la cytokine interleukine-33, IL-33/NF-HEV. Lors de mon doctorat, j'ai pris notamment en charge différentes études protéomiques visant à avancer dans la compréhension du rôle et des mécanismes associés à ces nouvelles protéines. Les travaux réalisés concernent donc la recherche des partenaires de ces nouveaux facteurs de transcription THAP, l'étude des processus de maturation de l'interleukine-33, mais également l'analyse protéomique quantitative à grande échelle des cellules endothéliales lors de la stimulation par différentes cytokines pro-inflammatoires.

Ces différentes études ont nécessité des optimisations de méthodes. Au cours de ces quatre années de doctorat, les stratégies protéomiques, les instruments de spectrométrie de masse, les outils bioinformatiques ont évolué et ouvert de nouvelles possibilités. Je me suis particulièrement impliquée dans le développement des stratégies de protéomique quantitative sans marquage, qui représentaient un outil attractif pour l'analyse de complexes protéiques, mais aussi par la suite de protéomes entiers. C'est pourquoi ce document suit délibérément un plan méthodologique, et tente d'illustrer les possibilités des approches protéomiques récentes basées sur la nanoLC-MS/MS, au travers de différents exemples concrets d'application.

Ainsi, dans une première partie, je présenterai des résultats obtenus sur la caractérisation de complexes protéiques. Cette partie est consacrée notamment à la caractérisation des complexes associés aux facteurs THAP. Lors de cette étude, nous avons été amenés à optimiser les protocoles d'analyse quantitative différentielle qui permettent d'identifier les partenaires spécifiques des protéines d'intérêt dans des approches d'immunopurification. Ces méthodes ont été mises en œuvre par la suite pour la caractérisation d'autres complexes, dans le cadre de projets collaboratifs (complexes impliquant le facteur TFIIH).

Nous nous sommes dans un second temps intéressés à l'étude à large échelle des processus engagés dans cellules endothéliales en conditions inflammatoires. Dans la seconde partie de ce

manuscrit sont donc présentées des analyses protéomiques globales portant sur des protéomes entiers ou des sous-protéomes, associées à des méthodes quantitatives sans marquage. Le glycoprotéome des cellules endothéliales en condition inflammatoire a d'une part été étudié, grâce à la mise en place d'une méthode d'enrichissement du protéome de surface des cellules. D'autre part, une analyse du protéome entier des cellules endothéliales et de ses variations en conditions inflammatoires a été réalisée. Dans ce contexte, et pour obtenir une couverture du protéome la plus large possible, nous avons mis en place et optimisé une méthode de protéomique quantitative sans marquage impliquant un fractionnement sur gel 1D SDS-PAGE de l'échantillon protéique.

Enfin, la troisième partie présente plus spécifiquement les résultats obtenus sur la cytokine IL-33. La fonction de l'interleukine-33 au sein des cellules endothéliales, et les mécanismes d'action associés, ont été étudiés. Nous nous sommes d'abord penchés sur le processus de clivage de cette cytokine, afin de caractériser ses formes maturées. Le rôle d'une de ces formes maturées au sein des cellules endothéliales a ensuite été analysé en appliquant la stratégie d'analyse protéomique quantitative à large échelle optimisée précédemment.

RESULTATS

Nous avons vu dans la partie introductive le rôle central et l'importance de la quantification aussi bien dans la protéomique d'interaction, pour l'identification non ambiguë des partenaires spécifiques d'une protéine d'intérêt, que dans la protéomique d'expression, afin de mettre en évidence les variations d'abondance de protéines dans différentes conditions au sein de protéomes complexes. Dans ce travail de thèse, nous avons choisi d'utiliser et d'optimiser des méthodes sans marquage qui présentent certains avantages sur les méthodes impliquant un marquage isotopique des échantillons. Dans ces stratégies, les échantillons à comparer sont traités en parallèle tout au long du processus, engendrant une certaine variabilité issue de leur préparation et de leur analyse MS. Il faut donc être particulièrement attentif à les traiter exactement de la même façon afin minimiser au maximum les biais pouvant être introduits. Enfin, l'analyse bioinformatique de ce type de données MS est plus complexe et spécifique puisque la comparaison se fait entre différentes acquisitions LC-MS/MS.

Au cours de cette thèse, une stratégie générale a été suivie pour préparer, analyser en nanoLC-MS/MS, et quantifier les différents types d'échantillons. Les échantillons protéiques à comparer, issus d'une même quantité de matériel biologique à l'origine, sont réduits et alkylés à l'iodoacétamide avant d'être déposés sur un gel 1D SDS-PAGE (Figure 22). Selon la complexité de l'échantillon et la profondeur analytique désirée, ils peuvent ou non être fractionnés sur le gel d'électrophorèse. Même lorsqu'aucun fractionnement n'est nécessaire et que l'échantillon est analysé en une acquisition MS unique, nous avons choisi de passer par un gel SDS-PAGE et ne découper alors qu'une seule bande contenant la totalité de l'échantillon. Cette approche permet en effet de solubiliser de façon optimale toutes les protéines grâce à l'utilisation de SDS (« Sodium dodecyl sulfate ») et aboutit au final dans nos mains à l'identification d'un plus grand nombre de protéines par rapport à des méthodes de digestion liquide ou de type FASP (Wisniewski, Zougman et al. 2009), incluant une solubilisation des protéines en SDS puis une élimination du détergent par échange extensif avec du tampon urée sur membrane d'ultracentrifugation. Les protéines sont ensuite digérées in-gel par la trypsine, et les peptides résultants sont extraits des bandes de gel. Les mélanges peptidiques sont analysés en nanoLC-MS/MS sur une chromatographie Ultimate 3000 (Dionex) en phase inverse C18 couplée à un LTQ-Orbitrap (XL ou Velos) (Thermo Scientific). Les données brutes sont ensuite soumises à une recherche en banque de données protéiques, à l'aide du moteur de recherche Mascot. Les protéines identifiées sont validées par le logiciel MFPaQ selon des seuils fixes de scores peptidiques Mascot, ajustés pour obtenir un FDR protéique de 1%, ou bien par le logiciel Prosper en fixant un seuil de FRD peptidique (5%) et protéique (1%).

A partir des données validées, la quantification relative des échantillons est également réalisée via MFPaQ, qui effectue une analyse comparative des signaux MS. Il apparie dans un premier temps les peptides identifiés de façon commune dans les différentes acquisitions MS, puis les utilise pour aligner celles-ci en temps de rétention. Une caractéristique intéressante de MFPaQ est de pouvoir réaliser une mise en relation croisée des signaux détectés entre différentes analyses, même lorsque certains de ces signaux n'ont pas déclenché de MS/MS, ou n'ont pas abouti à une identification peptidique dans une analyse donnée. A partir de la matrice d'alignement en temps de rétention, le logiciel permet de prédire le temps de rétention des ions peptidiques dans les analyses

où ils n'ont pas été identifiés par MS/MS, et d'extraire le signal correspondant à partir du temps de rétention prédit et de la masse exacte du peptide. Les signaux MS des peptides identifiés sont ainsi extraits dans chacun des fichiers bruts. MFPaQ procède ensuite à l'assignement des peptides aux protéines correspondantes, et restitue dans l'interface, pour chacun des peptides d'une protéine, son profil d'élution (XIC) dans chacun des échantillons comparés, et l'aire sous le pic associée. Le logiciel calcule en parallèle pour chaque protéine un indice d'abondance protéique, le PAI (« Protein Abundance Index ») qui correspond à la moyenne des intensités des trois peptides les plus intenses d'une protéine (définis sur l'ensemble des conditions comparées, et identiques entre ces différentes conditions). Ces PAI peuvent ensuite être utilisés pour calculer le ratio d'abondance relative d'une même protéine dans les différentes conditions, mais également pour estimer de façon approximative l'abondance absolue de différentes protéines identifiées dans une même condition (Silva, Gorenstein et al. 2006).

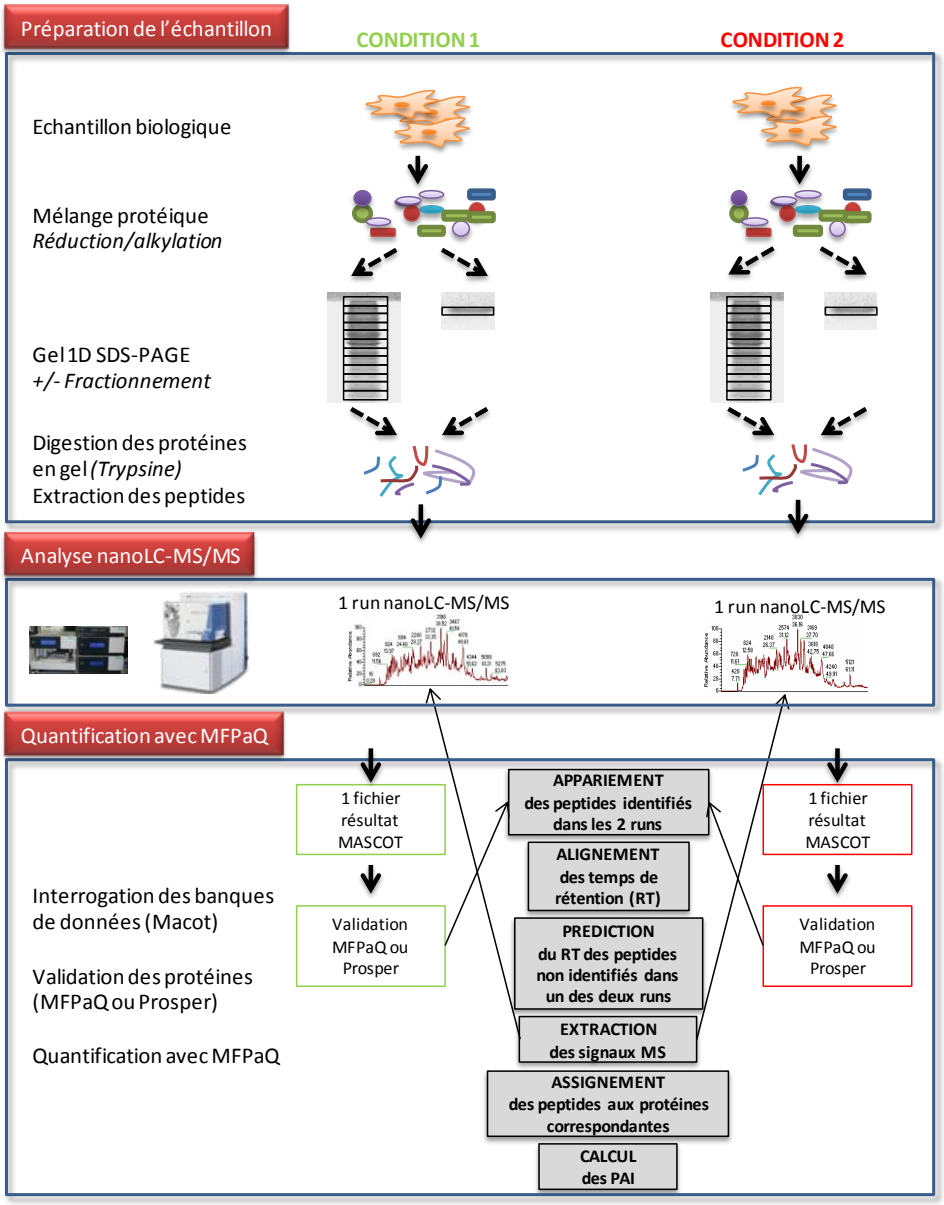


Figure 22 : Stratégie générale utilisée pour l'analyse nanoLC-MS/MS et la quantification sans marquage des échantillons protéiques.

Partie I. Etude de complexes protéiques : identification de nouveaux partenaires d'interaction

Une première partie de mon travail de thèse a porté sur l'étude de facteurs nucléaires identifiés au sein des cellules endothéliales, les protéines THAP, dont les fonctions et les mécanismes d'action sont peu connus. Afin d'appréhender leur fonctionnement, les complexes protéiques dans lesquels elles sont engagées ont été étudiés et les partenaires protéiques avec lesquels elles interagissent ont été identifiés. Nous avons pour cela mis en place des méthodes d'immunopurification et couplé celles-ci à une stratégie de quantification sans marquage de façon à parvenir à distinguer les partenaires protéiques *bona fide* des protéines contaminantes non spécifiques. Une stratégie similaire a été appliquée pour l'étude des partenaires d'interaction d'autres complexes protéiques impliquant le facteur général de transcription TFIIH.

I. Identification des partenaires protéiques des protéines THAP humaines

I-1. Contexte biologique

La famille des protéines THAP (Thanatos Associated Protein) est une nouvelle famille de facteurs nucléaires mis en évidence dans des cellules endothéliales par l'équipe de J-P Girard (IPBS). Ces protéines sont caractérisées par la présence d'un motif protéique conservé d'environ 90 acides aminés situé à l'extrémité N-terminale, le domaine THAP, présentant une signature de type C2CH (Cys-Xaa_{2-4}-Cys-Xaa_{35-50}-Cys-Xaa_2-His) (Figure 23A). Une centaine de protéines THAP ont à ce jour été identifiées à la fois chez les vertébrés (du poisson-zèbre à l'homme) et chez les invertébrés (drosophile, *C. elegans*). Ce domaine est en revanche absent chez les bactéries, les levures ou les plantes (Roussigne, Kossida et al. 2003). La famille des protéines THAP humaines comporte 12 membres nommés THAP0 à THAP11 (Figure 23B).

Certains éléments ont permis de mieux appréhender le rôle fonctionnel de cette nouvelle famille. Initialement, le motif THAP a été identifié dans le domaine de liaison à l'ADN de la transposase de l'élément P de *Drosophilia melanogaster* (Dejardin and Kingston 2009), ce qui laissait envisager que ces protéines pouvaient constituer une nouvelle classe de protéines de liaison à l'ADN (Roussigne, Kossida et al. 2003). Cette capacité du domaine THAP à lier l'ADN a par la suite été confirmée expérimentalement et le motif d'ADN reconnu par la protéine THAP1 humaine a été identifié. THAP1 se fixe sur une séquence d'ADN de 11 nucléotides nommée THABS (« THAP1 binding sequence ») de façon zinc-dépendante (Clouaire, Roussigne et al. 2005). Des études structurales par RMN ont par ailleurs permis de mieux caractériser ce domaine, montrant qu'il s'agit d'un domaine à doigt de zinc atypique (Bessiere, Lacroix et al. 2008) capable de reconnaitre et de lier sa séquence

d'ADN cible grâce aux interactions qu'il noue à la fois avec le grand sillon et le petit sillon d'ADN (Campagne, Saurel et al. 2010).

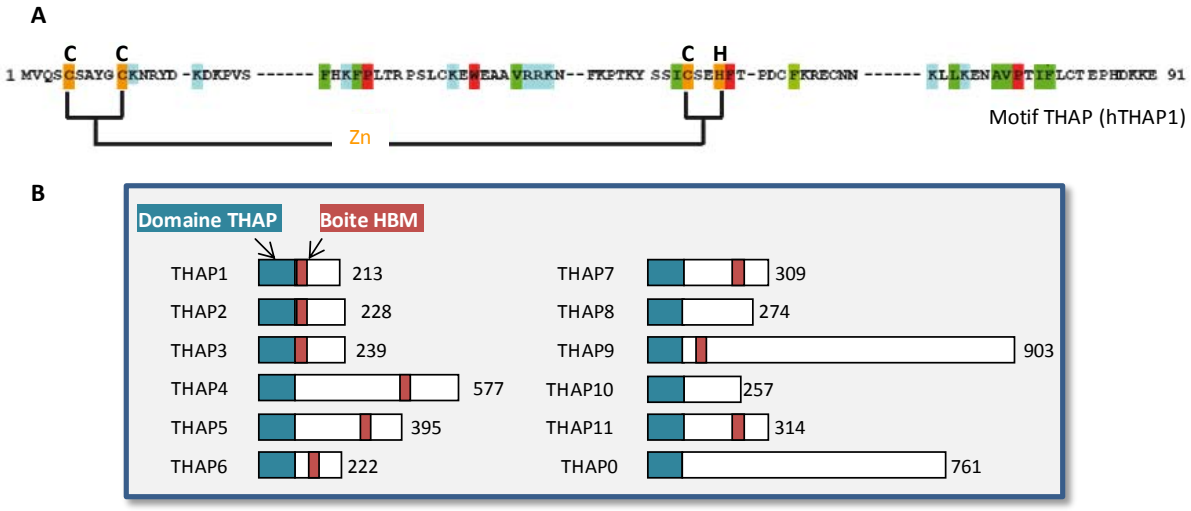


Figure 23 : Famille des protéines THAP. *(A) Domaine THAP de type C2CH caractérisant la famille des protéines THAP. Le motif présenté est celui de THAP1 humaine. (B) Famille des protéines THAP humaine, constituée de 12 membres, THAP0 à THAP11. Le domaine THAP est indiqué en bleu et la boite HBM en rouge.*

Il existe des données fonctionnelles pour certaines protéines homologues des THAP humaines chez des organismes modèles. Chez le poisson-zèbre, l'orthologue du facteur de transcription E2F6, impliqué dans la régulation du cycle cellulaire, contient en N-terminal un domaine présentant une homologie avec le domaine THAP (Clouaire, Roussigne et al. 2005). Ce facteur fonctionne comme un répresseur transcriptionnel durant la phase S, se fixant spécifiquement sur les promoteurs des gènes cibles E2F régulés en phase G1/S (Giangrande, Zhu et al. 2004). D'autre part, plusieurs études montrent que chez *C. elegans*, la protéine LIN-35, homologue de la protéine du rétinoblastome Rb humaine, interagit avec 4 protéines possédant des domaines THAP (LIN-36, LIN-15B, LIN-15A, et HIM-17) (Thomas and Horvitz 1999; Boxem and van den Heuvel 2002; Scholtissen, Guillemin et al. 2009), renforçant l'idée d'un lien entre les protéines THAP et la voie pRb/E2F. Parmi celles-ci, il a été montré que LIN-36 et LIN15-B agissent comme inhibiteurs de la transition G1/S (Boxem and van den Heuvel 2002). Par ailleurs, GON-14, une autre protéine à domaine THAP de *C. elegans*, joue un rôle clé dans la prolifération cellulaire et le développement (Chesney, Kidd et al. 2006). Ces protéines orthologues permettent d'envisager des fonctions biologiques potentielles pour les THAP humaines dans la régulation du cycle cellulaire.

Par ailleurs, quelques données fonctionnelles sont également disponibles sur les protéines THAP humaines. Il a été montré que la protéine THAP7 humaine est associée à la chromatine *in vivo* et est capable d'interagir avec les extrémités N-terminales des histones H3 et H4 *in vitro* (Macfarlan, Kutney et al. 2005). De plus, THAP7 interagit avec l'histone déacétylase HDAC3 et le co-répresseur NcoR, suggérant un rôle de répresseur transcriptionnel influant sur la structure de la chromatine *in vitro*. Cela a été effectivement vérifié lorsqu'elle est ciblée artificiellement au niveau d'un promoteur par le domaine de liaison à l'ADN de Gal4 (Macfarlan, Kutney et al. 2005). Concernant la protéine THAP1, des études réalisées dans l'équipe de J-P Girard ont montré que son inhibition par siRNA

provoque la répression de plusieurs gènes cibles de la voie pRB/E2F impliqués dans la progression du cycle cellulaire, et aboutit à l'arrêt du cycle cellulaire en phase G1/S, et à l'inhibition de la prolifération cellulaire (Cayrol, Lacroix et al. 2007). L'un des gènes cibles de THAP1 identifiés, *RRM1*, est par ailleurs requis pour la synthèse de l'ADN en phase S. THAP1 semble ainsi constituer un régulateur endogène majeur de la prolifération cellulaire ainsi que de la progression du cycle à la transition G1/S dans les cellules endothéliales. Il a également été montré que des mutations dans le gène THAP1 sont responsables d'une maladie neurologique, la dystonie de type 6, DYT6 (Hochuli 1988; Fuchs, Gavarini et al. 2009; Kathiresan, Willer et al. 2009; Liang, Yu et al. 2009; Stephenson, Gregory et al. 2009). THAP11 jouerait par ailleurs un rôle essentiel dans la pluripotence et l'auto-renouvellement des cellules souches embryonnaires murines (Dejosez, Krumenacker et al. 2008).

I-2. Objectifs et stratégie mise en place

Très peu de données fonctionnelles sont ainsi disponibles pour les protéines THAP humaines et leurs fonctions restent encore mal connues. L'objectif de ce projet réalisé en collaboration avec l'équipe de J-P Girard, est donc d'avancer dans la compréhension du rôle et des mécanismes d'action de certaines protéines THAP humaines. Nous avons pour cela étudié les complexes protéiques dans lesquels elles sont impliquées et identifié leurs partenaires protéiques afin d'appréhender leurs fonctions. Différents membres de la famille THAP ont été étudiés par cette approche (THAP3, THAP11, THAP7 et THAP1).

Les protéines THAP sont très peu représentées dans la cellule et ces faibles niveaux d'expression compliquent l'étude de leurs complexes endogènes. Nous nous sommes donc appuyés sur un modèle cellulaire permettant de les surexprimer de façon stable. Comme évoqué dans l'introduction, la surexpression de protéines peut cependant induire certains problèmes (toxicité, perturbation des complexes…), notamment dans le cas de protéines potentiellement impliquées dans la régulation du cycle et de la prolifération. Pour essayer de limiter ces problèmes au cours de la culture, nous avons opté pour un système permettant une surexpression conditionnelle de la protéine. Par ailleurs, aucun anticorps permettant d'immunoprécipiter efficacement les protéines THAP n'était disponible pour cette étude. Les protéines ont donc été exprimées en fusion avec une double étiquette FLAG-HA située en C-terminal. Des lignées stables de cellules HeLa exprimant de façon stable et inductible les protéines THAP étiquetées FLAG-HA sous le contrôle d'un système Tet-Off ont été pour cela développées dans l'équipe de J-P Girard. Ce système (Figure 24) permet de réprimer en présence de doxycycline l'expression de la protéine d'intérêt placée sous le contrôle d'un élément de réponse à la tétracycline (TRE) et d'un promoteur minimum. La protéine tTA (« tetracycline-controlled transactivator) est également exprimée, et est capable de se fixer au niveau du TRE en absence de doxycyline. Au contraire, après ajout de doxycycline, tTA est incapable de se fixer sur TRE et donc d'induire la transcription de la protéine d'intérêt. Les cellules sont donc cultivées et multipliées en présence de doxycycline (cellules non induites = contrôle), puis sur une partie de la culture, l'expression de la protéine étiquetée est induite par passage des cellules dans un milieu sans doxycycline (cellules induites = essai).

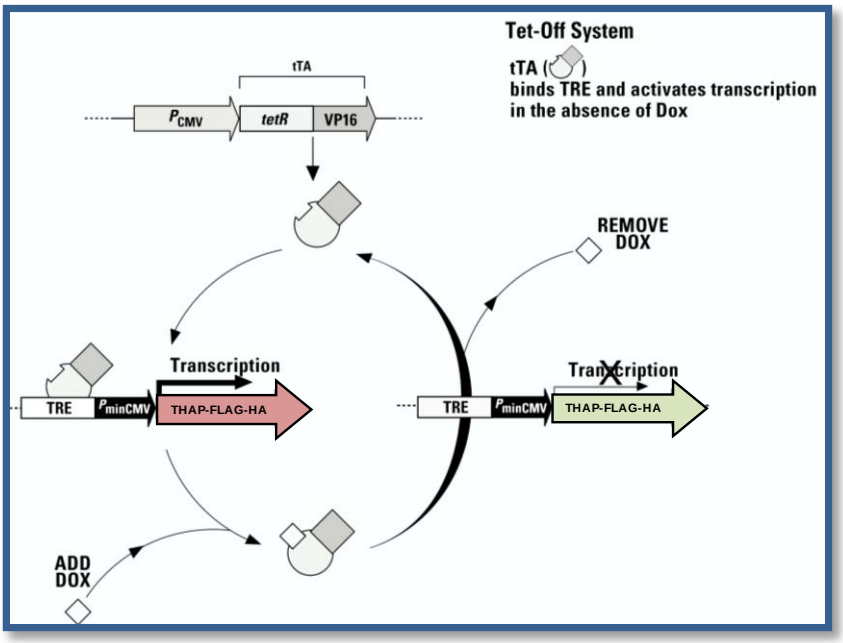


Figure 24 : Système d'expression conditionnelle de la protéine THAP étiquetée FLAG-HA (BD^{TM} Tet-Off). *La lignée cellulaire stable HeLa Tet-Off contient un plasmide régulateur (pTet-Off) et un plasmide de réponse (pTRE-THAP). Le plasmide pTet-Off comporte un gène codant pour la protéine tTA (tetracycline-controlled transactivator) qui résulte de la fusion du répresseur TetR (E. Coli) et du domaine C-terminal du domaine d'activation du virus Herpes simplex VP16. Le plasmide pTRE-THAP comporte le gène codant pour la protéine THAP étiquetée FLAG-HA en aval de la séquence TRE (tetracycline-response element) constitué d'une répétition de la séquence tetO et d'un promoteur minimum. L'expression de THAP est ainsi sous le contrôle de la protéine tTA. En présence de doxycycline, la protéine tTA est incapable de se fixer au niveau de l'opérateur tetO et par conséquent incapable d'initier la transcription du gène codant pour THAP. Au contraire, l'absence de doxycycline permet la surexpression de la protéine THAP.*

Le processus de préparation des complexes est ensuite réalisé de la même façon sur ces deux échantillons cellulaires. Ce processus inclut une préparation d'extraits nucléaires puis la purification des complexes. Les extraits nucléaires sont obtenus par extraction saline, suivant un protocole de base classiquement utilisé pour l'extraction et la purification de facteurs de transcription (Dignam, Martin et al. 1983) dans lequel nous avons ajouté l'utilisation d'une faible concentration de détergent (0,5 % NP-40). Dans le protocole originel décrit par Dignam et al., pour l'isolement de facteurs de transcription RNA polymérase de type TFII et TFIII (Dignam, Martin et al. 1983), la préparation des extraits nucléaires est réalisée en incubant les noyaux dans un tampon à 0 ,42M en NaCl. Dans cette étude, la concentration saline a été déterminée pour extraire efficacement les composants nucléaires tout en gardant une activité optimale des facteurs de transcription. Dans ce protocole que nous avons légèrement modifié, les cellules sont reprises dans un tampon hypotonique en présence de 0,5 % NP-40, puis cassées par casse mécanique (ultraturax). Une centrifugation à basse vitesse permet de séparer l'extrait cytoplasmique (surnageant) des noyaux (culot). Ceux-ci sont ensuite re-suspendus dans un tampon hypertonique contenant 0,42 M NaCl, 0,5 % NP-40. L'extrait nucléaire est récupéré après centrifugation à basse vitesse (surnageant) et dilué pour se ramener à une concentration saline physiologique (0,15 M NaCl). Ce protocole utilisé initialement a permis de caractériser avec succès certains complexes THAP. En revanche, dans

certains cas, nous avons tenté de modifier ces conditions biochimiques pour préserver des interactions fragiles (voir ci-dessous). A partir de l'extrait nucléaire, l'enrichissement des complexes est ensuite réalisé en deux étapes. Là encore, nous avons testé et utilisé différents protocoles pour améliorer l'enrichissement des complexes d'intérêt, basés soit sur un gradient de glycérol préparatif suivi d'une immunopurification anti-FLAG, soit sur une immunopurification en tandem (anti-FLAG puis anti-HA). Les complexes immunopurifiés sont ensuite fractionnés sur gel 1D puis analysés par nanoLC-MS/MS et les partenaires identifiés grâce à une comparaison quantitative entre essai et contrôle, comme décrit dans la partie précédente.

I-3. Analyse quantitative « label-free » des complexes protéiques THAP humains

I-3.1 Complexes protéiques de THAP3

a. Préparation des complexes THAP3 (Protocole 1)

Dans un premier temps, des extraits nucléaires ont été préparés à partir de cellules HeLa Tet-Off exprimant ou non THAP3 étiquetée, en présence 0,42 M de NaCl final et 0,5% NP40 (Figure 25). A partir de cet extrait, les complexes d'intérêt associés à la protéine appât ont été purifiés en deux étapes. La première a consisté en un fractionnement sur gradient de glycérol préparatif 10-40% permettant de séparer les complexes en fonction de leur densité : plus un complexe a un poids moléculaire élevé, plus il migrera dans le gradient. La présence de la protéine THAP dans les différentes fractions du gradient a été testée par western blot, ce qui nous a permis de vérifier l'existence de complexes contenant THAP (migrant dans des fractions haute densité du gradient), de distinguer différents complexes THAP éventuellement présents (de composition et de fonction potentiellement différentes), et de sélectionner au final les fractions du gradient correspondant à un même complexe. Celles-ci ont été rassemblées, et dans une deuxième étape, une immunopurification anti-FLAG a été réalisée. Les complexes immunopurifiés ont ensuite été élués de l'anticorps greffé sur billes de sépharose par ajout de peptide compétiteur FLAG avant d'être déposés sur gel SDS-PAGE et analysés en nanoLC-MS/MS.

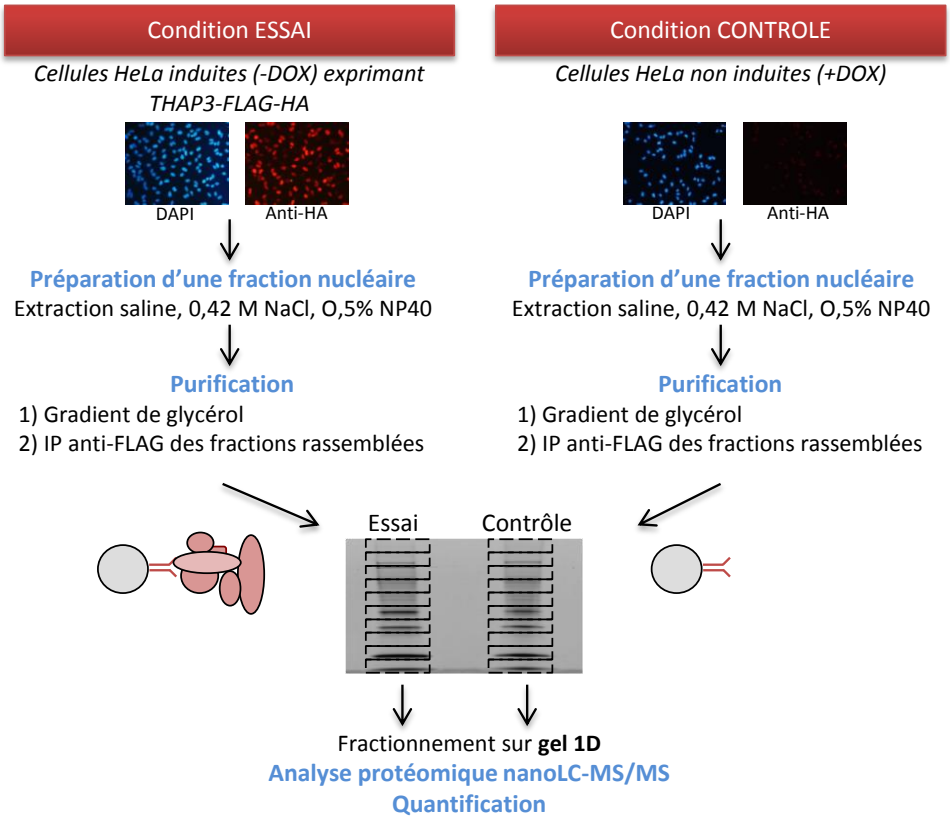


Figure 25 : Stratégie de préparation et d'analyse quantitative sans marquage des complexes THAP3. *Les complexes ont été purifiés à partir d'extraits nucléaires de cellules HeLa exprimant ou non THAP3, en deux étapes (gradient de glycérol puis immunopurification anti-FLAG). Les complexes purifiés ont ensuite été fractionnés sur gel SDS-PAGE avant d'être analysés par nanoLC-MS/MS. La quantification est réalisée avec MFPaQ entre l'échantillon essai et l'échantillon contrôle.*

b. *Analyse des complexes THAP3*

Analyse des complexes sur gradient de glycérol

L'analyse des complexes sur gradient de glycérol par western blot anti-HA (révélant THAP3) a montré la présence d'un complexe majoritaire d'environ 600 kDa impliquant la protéine THAP3 (Figure 26A). Les fractions correspondantes (2 à 5) ont été rassemblées et une immunopurification anti-FLAG a été réalisée. Les éluats de purification ont ensuite été déposés sur gel SDS-PAGE, et le profil du gel 1D obtenu est présenté dans la figure 26B. Sur ce dernier, THAP3 apparait comme la protéine majoritaire et la plus enrichie de l'essai, attestant de l'efficacité de la purification. Des bandes supplémentaires apparaissent clairement spécifiques de la piste essai correspondant aux partenaires potentiels de THAP3. Ainsi, dans cette expérience, les partenaires protéiques de THAP3 peuvent directement être visualisés sur le gel. Il apparait par ailleurs un très faible bruit de fond de protéines contaminantes. La purification a ainsi permis l'obtention de complexes très enrichis et « propres ».

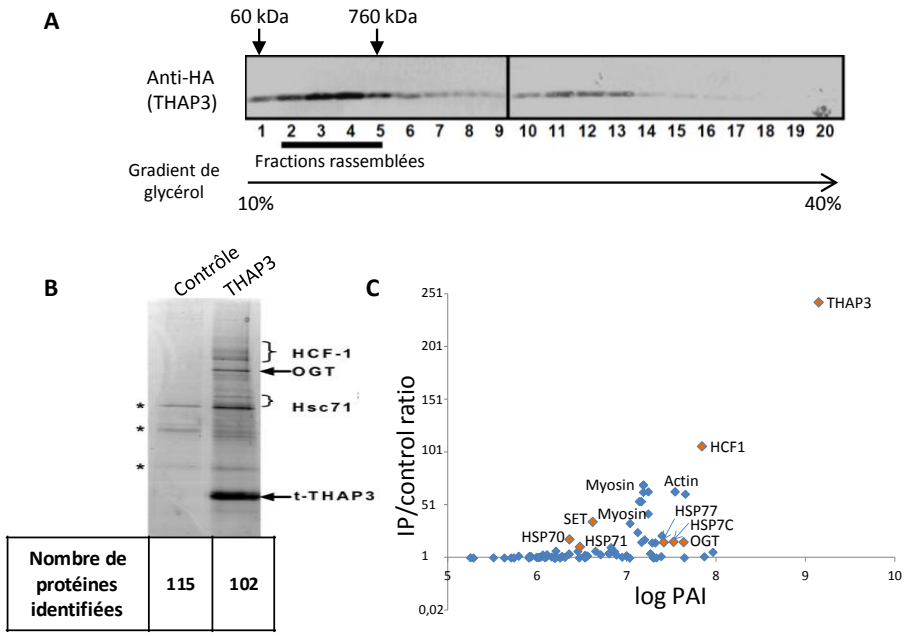


Figure 26 :: Analyse des complexes protéiques THAP3*. (A) Fractionnement sur gradient de glycérol 10-40% de l'extrait nucléaire des cellules induites pour l'expression de THAP3 et visualisation par western blot anti-HA (THAP3) des complexes THAP3 contenus dans les différentes fractions du gradient. (B) Analyse par gel 1D SDS-PAGE et par spectrométrie de masse du complexe THAP3 de 600kDa isolé sur gradient de glycérol et des fractions contrôle correspondantes. Les nombres de protéines indiqués correspondent aux protéines identifiées par Mascot après validation MFPaQ sur l'ensemble de la piste de migration (C) Représentation graphique (Ratio essai/contrôle versus log PAI essai) de l'analyse protéomique quantitative des complexes THAP3.*

Analyse protéomique des complexes par nanoLC-MS/MS

Dans cette étude, les pistes de migration ont été découpées en 30 bandes, en suivant le profil de coloration de la piste essai et en découpant des bandes similaires en vis-à-vis sur la piste contrôle. Chaque bande a été analysée en nanoLC-MS/MS sur un appareil LTQ-Orbitrap XL (Thermo Fisher Scientific). L'analyse nanoLC MS/MS des bandes des pistes essai et contrôle a confirmé la bonne purification de ces complexes puisque au total, seules 102 et 115 protéines ont été identifiées dans les pistes essai et contrôle respectivement (Figure 26B). Bien qu'il puisse apparaître élevé, ce nombre de protéines est cependant relativement réduit par rapport à ce qui peut être obtenu sur un appareil à haute vitesse de séquençage lors de l'analyse d'une immunoprécipitation contenant un bruit de fond important, comme cela sera illustré par la suite. Dans ce cas « idéal », une simple analyse différentielle des listes des protéines identifiées dans chacun des deux échantillons est suffisante pour mettre en évidence les partenaires protéiques de l'appât. En effet, lorsqu'on classe la liste des protéines identifiées dans l'essai en fonction d'une métrique semi-quantitative (score Mascot ou somme des MS/MS par protéine), THAP3 et ses partenaires potentiels apparaissent clairement en tête de liste comme les protéines majoritaires de l'échantillon immunoprécipité. Une analyse quantitative peut cependant être également réalisée, en calculant pour chaque protéine une valeur de PAI (moyenne des intensités des trois peptides les plus intenses). Le résultat de la quantification peut être représenté en traçant pour chaque protéine le ratio du PAI essai/contrôle en fonction de son PAI dans l'essai (Figure 26C). Nous avons choisi cette représentation dans le cas des complexes

THAP3, car le système Tet-Off présente souvent de légères « fuites », entraînant la présence d'une faible quantité de protéine appât dans les cellules contrôle, même en présence de doxycycline. Ainsi, il est possible dans ce cas de calculer un ratio de PAI pour la protéine appât et ses partenaires potentiels, car un signal associé, même faible, est généralement retrouvé dans le contrôle. On schématise ainsi vers la droite les protéines les plus abondantes de l'échantillon et vers le haut les plus enrichies dans l'essai par rapport au contrôle. Sont ainsi retrouvées, représentées en rouge, parmi quelques protéines contaminantes correspondant à des protéines majeures de la cellule, la protéine THAP3 (la plus enrichie et la plus abondante) ainsi que ses partenaires potentiels qui ressortent du faible bruit de fond constitué par les protéines quantifiées avec un ratio de 1.

Cette analyse a donc permis l'identification de plusieurs protéines partenaires regroupées dans la Table 2, qui ont été identifiées avec de bons scores et présentent un important ratio essai/contrôle : le facteur de prolifération cellulaire HCF-1 (Host Cell Factor 1), la O-GlcNAc glycosyltransférase OGT, des protéines chaperonnes comme HSP70 et HSP71, et la protéine SET impliquée dans la régulation de la transcription. Ces protéines ont été identifiées de façon reproductible dans trois réplicats biologiques différents.

Table 2 : Protéines d'intérêt identifiées comme partenaires potentiels de THAP3 après analyse protéomique quantitative différentielle de la piste IP vs la piste CT. (AC : numéro d'accession).

Gène	AC	Protéine - Description	Score	Ratio IP/Contrôle
THAP3	Q8WTV1	THAP domain-containing protein 3	904	240
OGT	O15294	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase	618	15
HCF1C1	P51610	Host cell factor	429	110
HSPA9	Q53FA3	Heat shock 70kDa protein 1-like variant	97	20
HSPA8	P11142	Heat shock cognate 71 kDa protein	743	15
SET	A5A5H4	Protein SET	62	35

L'association de THAP3 avec HCF-1 et OGT a par la suite été confirmée. L'analyse par western blot du gradient de glycérol réalisé sur les extraits nucléaires a d'une part montré une co-sédimentation des protéines HCF-1 et OGT avec THAP3 (Figure 27A). Ces interactions ont d'autre part été validées par des expériences de co-immunoprécipitation suivies en western blot (Figure 27B). Par ailleurs, l'interaction de HCF-1 avec THAP3 et THAP1 endogène a été validée *in vivo* au niveau du promoteur de gènes cibles par immunoprécipitation de la chromatine, et des expériences fonctionnelles ont permis de montrer que HCF-1 intervient dans la régulation des gènes cibles de THAP1 dans les cellules endothéliales. Un descriptif détaillé de ces expériences de validation peut être trouvé dans la publication (Mazars, Gonzalez-de-Peredo et al. 2010) présentée à la fin de ce chapitre.

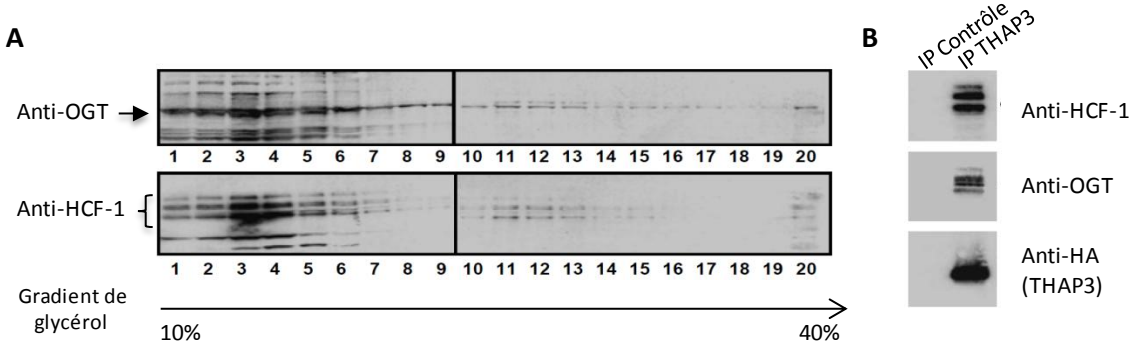


Figure 27. Validation des interactions de THAP3 avec HCF-1 et OGT. *(A) Co-sédimentation des protéines HCF-1 et OGT avec THAP3 sur gradient de glycérol. L'analyse du gradient de glycérol a été réalisée par western blot avec des anticorps dirigés contre HCF-1 et OGT. (B) Co-immunoprécipitation de HCF-1 et OGT révélée par analyse en western blot des éluats de purification essai (THAP3) et contrôle.*

La protéine THAP3 semble ainsi former un complexe avec le facteur de prolifération cellulaire HCF-1, la glycosyltransférase OGT et des protéines chaperonnes. Comme détaillé plus bas, HCF-1 est un facteur qui joue un rôle clé dans le contrôle du cycle cellulaire et la régulation de la prolifération (Reilly, Wysocka et al. 2002), et est capable de se fixer à différents facteurs de transcription via un motif de type D/EHxY (boite HBM, « HCF-1 binding motif ») (Freiman and Herr 1997). Un motif de ce type est retrouvé dans la plupart des membres de la famille THAP. Par ailleurs, HCF-1 a été décrit comme une protéine d'assemblage, se fixant sur certains facteurs de transcription et permettant de recruter, au voisinage des gènes cibles de ces facteurs, des protéines impliquées dans la modification des histones et le remodelage de la chromatine, à la fois de type HDAC (Histone deacetylase, répresseur transcriptionnel) et HMT (histone methyltransferase, activateur transcriptionnel) (Wysocka, Myers et al. 2003). Dans la suite de cette étude, nous avons donc cherché (1) à confirmer l'association de différents membres de la famille THAP avec HCF-1, via des analyses protéomiques comparables à celles menées sur THAP3, et (2) à identifier de nouveaux partenaires protéiques membres des complexes THAP, qui pourraient correspondre à des protéines recrutées par HCF-1 et possédant une activité enzymatique de modification de la chromatine.

La suite de ce chapitre détaille donc les résultats non publiés d'analyses protéomiques réalisées sur différents membres de la famille THAP. Au cours de ces expériences menées au début de mon doctorat, nous avons été confrontés à plusieurs difficultés : problèmes relatifs à l'extraction de complexes liés à l'ADN, instabilité des complexes protéiques et perte des partenaires labiles, présence dans certains échantillons purifiés d'un bruit de fond important. Dans ce dernier cas, comme nous le verrons, l'utilisation de méthodes quantitatives pour détecter les partenaires enrichis dans l'échantillon immunopurifié s'est révélée réellement nécessaire. Nous avons donc été amenés à évaluer différentes méthodes protéomiques quantitatives. Les parties suivantes illustrent donc, au travers de l'analyse de différents complexes THAP, des variations sur les méthodes biochimiques de préparation des complexes et/ou sur les techniques de quantification.

I-3.2 Complexes protéiques de THAP11

Au vu des résultats obtenus lors de l'étude du complexe THAP3 et de l'efficacité de la méthode de préparation des complexes, nous avons décidé dans un premier temps d'appliquer la même stratégie pour caractériser les partenaires d'interaction d'une autre protéine de la famille THAP, la protéine THAP11. Nous avons par la suite cherché à optimiser la préparation des complexes en testant des conditions d'extraction et de purification différentes.

a. Analyse des complexes THAP11 avec le protocole standard (Protocole 1)

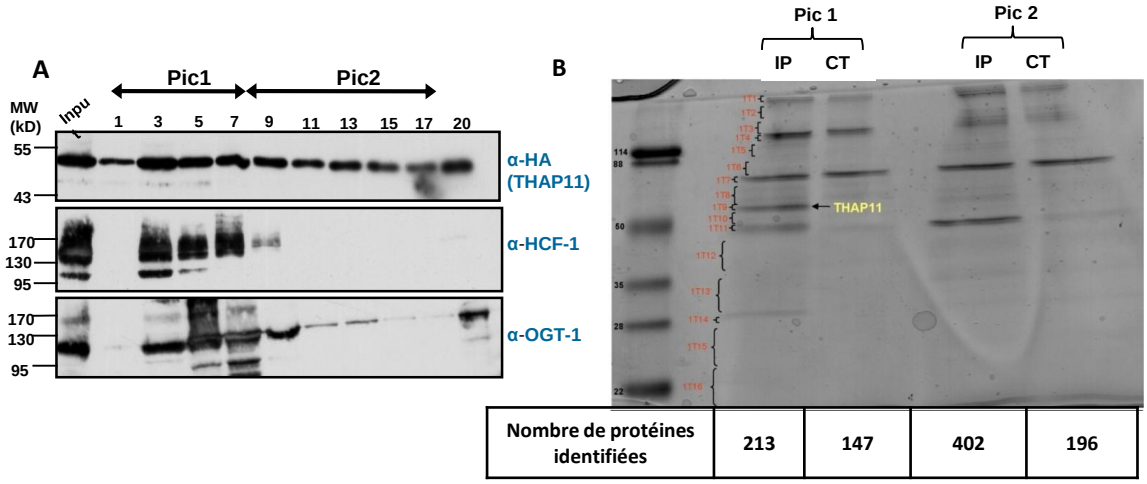
Nous avons dans un premier temps préparé les complexes protéiques impliquant la protéine THAP11 selon le protocole utilisé pour l'analyse des complexes THAP3, décrit plus haut. Pour cette étude, nous sommes partis de 5.10^8 cellules HeLa induites (-dox) et non induites (+dox). L'induction de THAP11 a été estimée par immunofluorescence à plus de 65 % des cellules HeLa. Les cellules non induites n'expriment pas ou très peu la protéine THAP11 et constituent ainsi un bon contrôle.

Analyse des complexes sur gradient de glycérol

Les extraits nucléaires de cellules HeLa ont ensuite été déposés sur gradient de glycérol (Figure 28A). L'analyse en western blot du gradient de glycérol montre que la séparation de différents complexes est moins nette que dans le cas de THAP3, mais laisse envisager la présence de deux complexes THAP11 de poids moléculaires différents, assez mal résolus mais pouvant correspondre aux fractions 2-8 (Pic 1) et 9-18 (Pic 2). A ce stade de la purification, nous avons cherché à tester la présence dans les fractions du gradient de glycérol de protéines candidates qui pourraient éventuellement interagir avec THAP11, tels que les partenaires déjà mis en évidence pour THAP3 (le facteur HCF-1 ou la O-GlcNAc transférase OGT-1). Ces protéines ont été retrouvées majoritairement dans le complexe THAP11 de plus faible poids moléculaire (Pic 1) suggérant leur interaction avec THAP11. Il est donc possible que les deux pics du gradient de glycérol contiennent des complexes THAP11 de nature distincte.

Analyse protéomique des complexes par nanoLC-MS/MS

Les deux complexes identifiés sur gradient de glycérol ont été immunopurifiés sur billes anti-FLAG, puis élués à l'aide de peptide compétiteur FLAG, et déposés sur gel 1D (Figure 28B). Le profil de migration des différentes pistes permet de faire ressortir une bande nettement spécifique dans le cas de l'analyse du premier complexe, qui correspond à la protéine appât THAP11. En revanche, la coloration (réalisée au bleu de Coomassie colloïdal) ne permet pas de faire clairement ressortir d'autres bandes spécifiques qui pourraient correspondre à des partenaires de THAP11, comme c'était le cas pour THAP3. Ici, le rendement de l'immunopurification paraît faible, et l'enrichissement réalisé sur les interactants endogènes potentiels de THAP11 est insuffisant pour qu'ils soient repérables via la coloration. Le bruit de fond de protéines contaminantes, visible dans la piste contrôle, est relativement limité, comme dans le cas de l'expérience précédente sur THAP3, mais la quantité de complexe immunopurifié est également réduite.



Nombre de protéines identifiées	213	147	402	196

Figure 28 : Analyse des complexes THAP11 purifiés avec le protocole standard. *(A) Fractionnement sur gradient de glycérol 10%-40% de l'extrait nucléaire des cellules induites pour l'expression de THAP11, et analyse sur western blot du gradient de glycérol. La présence de différents partenaires envisagés de THAP11 (HCF-1, OGT) a été testée. Input = extrait nucléaire avant fractionnement sur gradient de glycérol. Le gradient de glycérol est fractionné en 20 fractions. Une fraction sur deux est analysée, la fraction 1 correspondant à 10% glycérol et la fraction 20 à 40% glycérol. (B) Analyse par gel 1D (fractionnement) et spectrométrie de masse des complexes THAP11 isolés sur gradient de glycérol et des fractions contrôle correspondantes (Pic 1 : fractions 1-8 du gradient de glycérol ; Pic 2 : fractions 9-18). Les nombres totaux de protéines indiqués correspondent aux protéines identifiées par Mascot après validation par MFPaQ sur l'ensemble de la piste.*

De façon à identifier malgré tout les partenaires potentiels, l'ensemble des pistes de migration correspondant à chaque échantillon et à son contrôle ont été découpées en 16 bandes, chacune étant digérée à la trypsine et analysée par nanoLC-MS/MS. Pour le complexe « Pic 1 », 213 protéines ont été identifiées dans la piste IP et 143 dans la piste contrôle. Pour le « Pic 2 », 402 et 196 protéines ont été identifiées respectivement dans les pistes IP et contrôle.

Pour essayer de dégager les vrais partenaires de la protéine appât, nous avons réalisé une analyse quantitative différentielle entre la piste IP et la piste contrôle. Dans la figure 29, nous avons représenté, pour les deux pics du gradient de glycérol, le PAI des protéines identifiées dans l'essai versus leur PAI dans le contrôle. Cette représentation permet de visualiser plus facilement les protéines spécifiques de la piste essai (pour lesquelles aucun signal n'a pu être extrait dans le contrôle). Ces graphes indiquent bien que la grande majorité des protéines détectées sont non spécifiques (un signal quasiment équivalent peut être retrouvé dans la piste contrôle, même dans les cas où la protéine n'a pas été séquencée par MS/MS dans cet échantillon). Il existe cependant un certain nombre de protéines spécifiques de la piste IP, mais toutes ne représentent clairement pas des partenaires potentiels de THAP11. Beaucoup d'entre elles sont détectées avec un signal très faible, sur un seul peptide, potentiellement présent dans le contrôle mais en dessous du seuil de détection (ces cas « limites », inversement, sont également associés à des protéines qui apparaissent comme spécifiques du contrôle par rapport à l'essai). Les seuls candidats qui semblent réellement fixés à la protéine appât après immunopurification, pourraient être ceux qui sont spécifiques de la

piste essai, et suffisamment abondants. Pour les classer par ordre d'abondance, il est possible d'utiliser des métriques approximatives semi-absolues comme le score Mascot, le nombre de peptides identifiés, ou de façon plus précise, le nombre total de spectres MS/MS enregistrés pour la protéine, ou enfin le PAI comme calculé ici. D'après les graphes obtenus pour THAP11, il est intéressant de voir que le partenaire potentiel qui ressort nettement de l'analyse du pic 1 est HCF1, présentant un PAI très élevé parmi les protéines spécifiques. D'autres protéines sont spécifiquement détectées dans la piste IP et relativement abondantes, mais il s'agit essentiellement de composants majoritaires de la cellule (isoformes de tubuline).

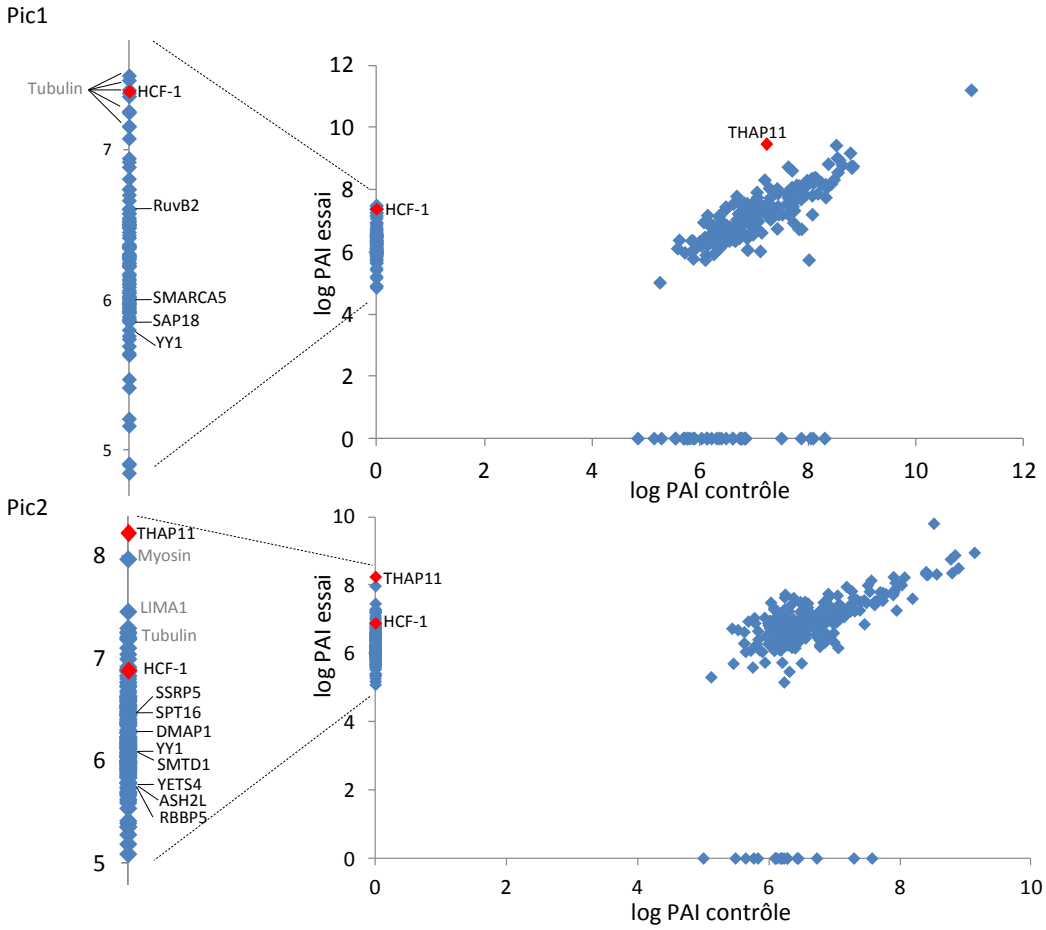


Figure 29 : Représentation graphique (PAI essai vs PAI contrôle) de l'analyse protéomique quantitative des complexes THAP11 pour le pic 1 et pour le pic 2

Dans l'analyse quantitative du pic 1, d'autres protéines biologiquement intéressantes sont détectées spécifiquement dans la piste essai. Elles présentent cependant de faibles PAI et sont pour la plupart détectées avec un faible nombre de peptides. Ces protéines partenaires potentiels participent à la régulation de la transcription : SMARCA5 (ou SNF2h/ISWI) impliquée dans le remodelage des nucléosomes, RuvB2 impliqué dans le complexe NuA4 HAT (Histone Acetyl Transférase) activateur de la transcription et SAP18, appartenant au complexes SIN3-HDAC1 répresseur de la transcription. La protéine YY1 (facteur de transcription intervenant dans la prolifération et le cycle cellulaire) est également identifiée avec un faible score. Des partenaires potentiels biologiquement intéressants sont également identifiés spécifiques du pic 2 dans l'analyse

quantitative mais, excepté HCF1, ils sont là encore peu abondants (faibles PAI, faibles scores) : YY1, certaines protéines du complexe NuA4 HAT (DMAP1, YETS4), du complexe histone methyltransferase Set1/Ash-2 (RBBP5, ASH2L), ou d'autres complexes impliqués dans le remodelage de la chromatine ou des nucléosomes (complexe FACT, SWI/SNF). Pour toutes ces protéines, les données protéomiques ne sont pas suffisamment probantes pour confirmer leur interaction avec THAP11.

b. Analyse des complexes THAP11 par double immunopurification (protocole 2)

Au vu des résultats de l'analyse protéomique précédente, bien que certains partenaires potentiels de THAP11 biologiquement pertinents aient pu être mis en évidence, certains sont identifiés avec des PAI et des scores très bas traduisant une abondance très faible dans l'échantillon final. Excepté HCF1, aucun partenaire spécifique de THAP11 n'a pu être identifié sans ambiguïté. Il est donc possible que l'interaction de THAP11 avec certains de ses partenaires soit diminuée ou même totalement perdue dans les conditions d'extraction nucléaire utilisées. Nous avons donc cherché à rendre cette extraction moins stringente en utilisant une force saline réduite par rapport au protocole-type de Dignam et al., et en éliminant la présence de détergent dans le tampon d'extraction. D'autre part, il apparaît qu'au moins dans le cas de THAP11, le gradient de glycérol n'a pas apporté un enrichissement suffisant des complexes. Par conséquent, nous avons tenté de rendre la purification plus spécifique en tirant parti de la double étiquette FLAG-HA dans la protéine THAP et en réalisant deux étapes consécutives d'immunopurification.

Préparation des complexes THAP11

Nous avons adapté le protocole 1 précédemment utilisé pour réduire la concentration en sel et éviter la dissociation éventuelle d'interactions faibles entre partenaires physiologiques. Dans le protocole 2 (Figure 30), l'extraction saline est réalisée par resuspension du culot de noyaux, puis ajout d'un faible volume (environ 2/3 du volume du culot) de tampon à 0,42M NaCl, aboutissant à une concentration saline finale de 0,17M NaCl pour le nucléoplasme extrait. Après centrifugation et prélèvement de l'extrait, une deuxième extraction a été effectuée sur le culot résiduel de noyaux pour aboutir à une concentration de 0,4M NaCl final. La concentration saline des deux extraits a ensuite été ramenée à 0,15M NaCl avant de les rassembler. Par ailleurs, aucun détergent n'a été rajouté dans les tampons d'extraction des noyaux.

Par la suite, la purification des complexes a été menée exclusivement par immunopurification. Deux immunopurifications anti-FLAG et anti-HA ont été successivement réalisées dans les conditions décrites par Nakatani et al (Nakatani and Ogryzko 2003) afin d'éliminer au maximum les protéines non spécifiques. L'élution des billes anti-FLAG a été effectuée comme précédemment par ajout de peptide compétiteur FLAG. L'élution des billes anti-HA a été effectuée par ajout de peptide compétiteur HA. Pour chacune de ces deux immunopurifications, plusieurs lavages de stringence modérée avec un tampon à 0.15M NaCl contenant 10% glycérol et 0,1% Tween ont été réalisés pour limiter la fixation non spécifique.

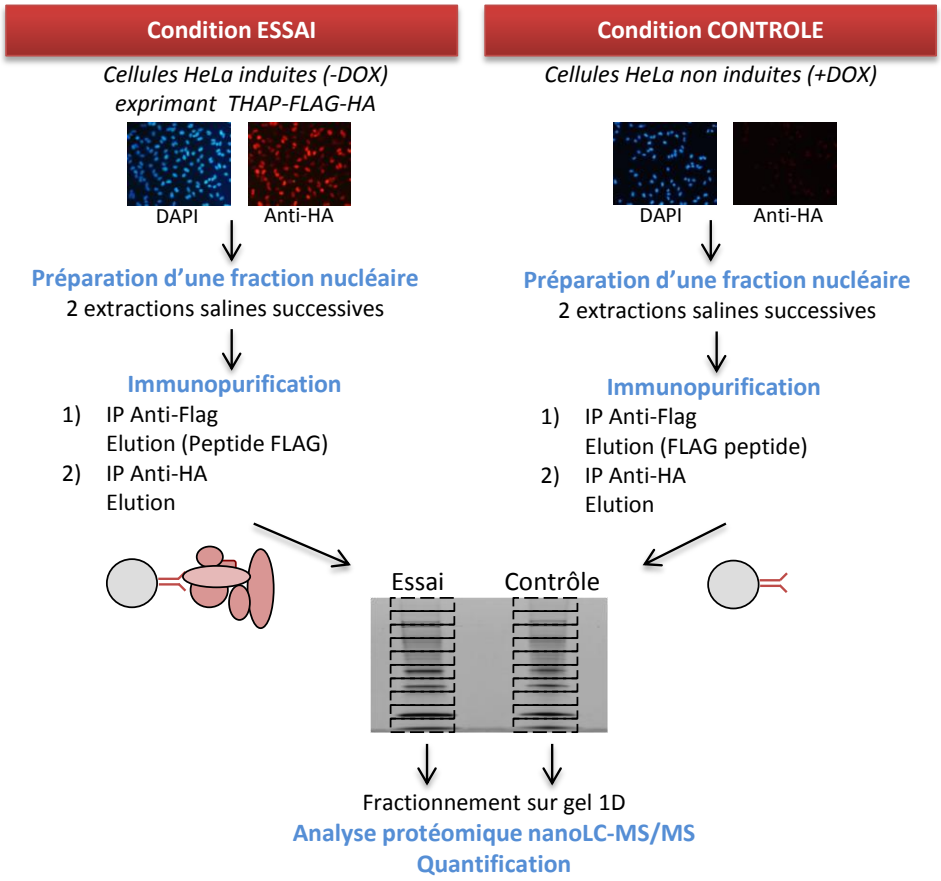


Figure 30 : Stratégie de préparation et d'analyse quantitative sans marquage des complexes THAP11 (Protocole 2 peu stringent). *Les complexes ont été purifiés à partir d'extraits nucléaires de cellules HeLa exprimant ou non THAP11, en deux étapes (double immunopurification anti-FLAG puis anti-HA). Les complexes purifiés ont ensuite été fractionnés sur gel SDS-PAGE avant d'être analysés par nanoLC-MS/MS. La quantification est réalisée entre l'échantillon essai et l'échantillon contrôle.*

Analyse protéomique des complexes par nanoLC-MS/MS

Pour cette étude, nous sommes partis de 6.10^8 cellules HeLa induites (-dox) et non induites (+dox). Les extraits nucléaires ont été soumis à une double immunoprécipitation comme décrit plus haut, puis les éluats ont été déposés sur gel 1D (Figure 31A). A l'issue de cette double immunopurification, nous avons obtenu très peu de matériel final, et peu de bandes sont visibles avec la coloration Coomassie. Un différentiel n'est pas visible sur gel sur la base de la coloration. Le bruit de fond semble cependant moins important. Chaque piste de migration (échantillon vs contrôle) a été découpée en plusieurs bandes qui ont ensuite été analysées en protéomique. 80 et 59 protéines ont été identifiées dans la piste IP et la piste contrôle respectivement. Ces nombres plus réduits de protéines illustrent la plus grande spécificité du protocole utilisé.

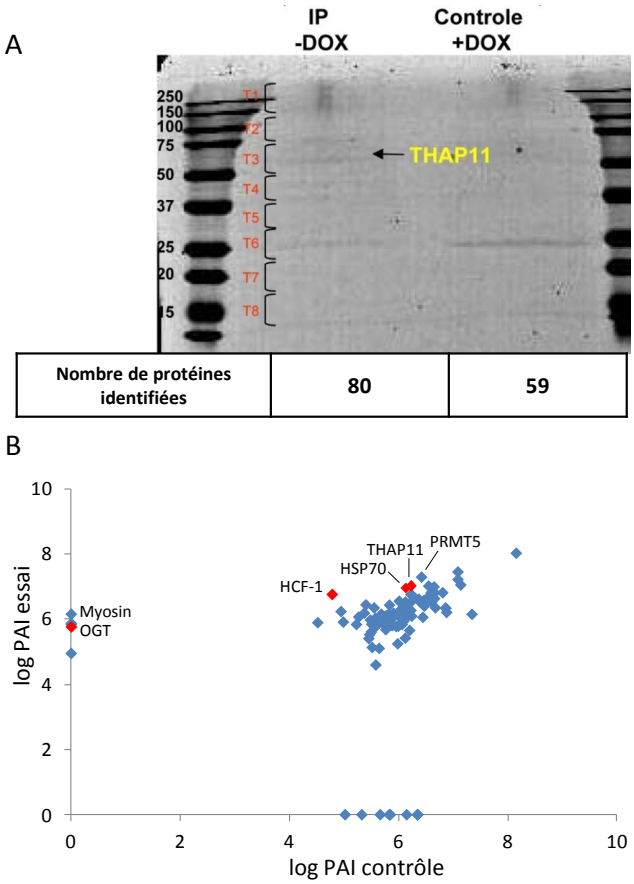


Figure 31 : Analyse des complexes THAP11 purifiés avec le protocole peu stringent (2) par gel1D (fractionnement) et nanoLC-MS/MS. *(A) Gel 1D SDS-PAGE. Les nombres totaux de protéines indiqués correspondent aux protéines identifiées par Mascot après validation par MFPaQ sur l'ensemble de la piste. (B) Représentation graphique (PAI essai vs PAI contrôle) de l'analyse protéomique quantitative des complexes THAP11.*

L'analyse quantitative réalisée (Figure 31B) a permis de mettre en évidence 12 partenaires potentiels spécifiques de l'essai ou présentant un ratio d'intensité supérieur à 10, parmi lesquels on retrouve, en plus de THAP11, les partenaires connus de THAP3, les protéines HCF-1, HSP70, et OGT-1 (Table 3). D'autres protéines pourraient être biologiquement intéressantes comme PRMT5, une arginine méthyltransférase impliquée dans la méthylation des histones, la répression de la transcription et s'associant aux complexes de remodelage de la chromatine SWI/SNF. D'un point de vue méthodologique, le protocole basé sur la double immunopurification nous a donc permis de limiter le bruit de fond de protéines contaminantes, au prix d'une perte de matériel aboutissant à une détection limite de l'appât et de ses partenaires. Les modifications apportées au protocole de préparation des complexes nous ont permis d'extraire un complexe similaire à celui qui avait été identifié pour THAP3 (HCF-1, HSP70, OGT-1), sans pour autant caractériser nettement d'autres protéines de modification de la chromatine potentiellement recrutées par HCF-1.

Table 3 : Protéines partenaires potentielles d'intérêt identifiées après analyse protéomique des complexes THAP11 purifiés avec le protocole 2 peu stringent*. (AC : numéro d'accession).*

Gène	AC	Protéine - Description	Score	Ratio IP/CT
HSP7C	P11142	Heat shock cognate 71 kDa protein	360,14	11
THAP11	A8K002	cDNA FLJ76985, highly similar to Homo sapiens THAP domain containing 11	108,36	11
PRMT5	O14744	Protein arginine N-methyltransferase 5	80,99	16
HCF1C1	P51610	Host cell factor	63,49	33
OGT	O15294	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit	55,63	Spécifique IP

I-3.3 Complexes protéiques de THAP7

Nous avons par la suite cherché à identifier les partenaires de la protéine THAP7. Les complexes ont été préparés selon le même protocole peu stringent (protocole 2). Pour cette étude, nous sommes partis de 6.10^8 cellules HeLa induites (-dox) et non induites (+dox) pour l'expression de THAP7-FLAG-HA. Les extraits nucléaires ont été soumis à une double immunoprécipitation comme décrit plus haut. Dans cette expérience, afin d'éviter toute perte de matériel, l'élution finale des complexes fixés sur les billes greffées anti-HA a été réalisée en tampon de type Laemmli (contenant 2% de SDS au final, mais sans réducteur pour éviter le décrochage des chaînes d'anticorps), et les éluats ont ensuite été déposés sur gel 1D (Figure 32A). Au vu des profils de coloration obtenu, un bruit de fond important (probablement lié à l'élution en SDS) est présent pour chaque échantillon, et malgré la double immunopurification, l'enrichissement de l'appât et de ses partenaires par rapport au bruit de fond est insuffisant pour visualiser un différentiel sur le gel 1D. Cependant, ce différentiel est clairement visible par western-blot pour la protéine appât THAP7, indiquant que l'immunopurification est néanmoins efficace. Il est donc dans ce cas essentiel de réaliser une analyse quantitative différentielle pour discriminer les vrais partenaires protéiques des contaminants. Les pistes IP et contrôle ont pour cela été découpées de façon systématique en 10 bandes, et analysées en nanoLC-MS/MS.

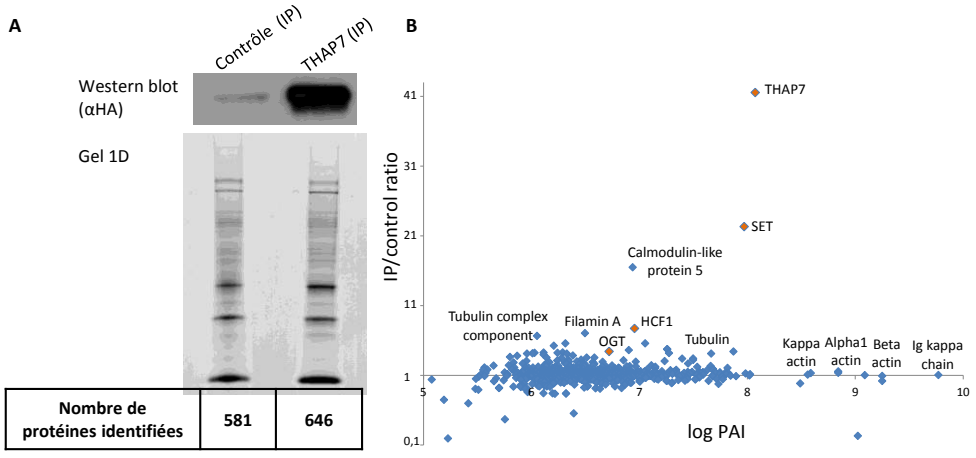


Figure 32 : Analyse des complexes THAP7 purifiés avec le protocole 2*. (A) Analyse et fractionnement des éluats immunoprécipités par gel 1D SDS-PAGE. (B) Représentation graphique (Ratio essai/contrôle vs log PAI essai) de l'analyse protéomique quantitative des complexes THAP7.*

a. Analyse protéomique quantitative des complexes

Comme attendu, l'analyse protéomique des différentes fractions a conduit à l'identification d'un nombre important de protéines dans les deux pistes (646 protéines dans l'IP et de 581 protéines dans le contrôle). Les protéines contaminantes sont quantifiées avec un ratio proche de 1 (Figure 32B) et quelques protéines qui apparaissent enrichies ont pu être mises en évidence (Table 4). On retrouve HCF-1 et OGT, qui semblent donc également interagir avec THAP7. La protéine SET dont l'interaction avec THAP7 a déjà été décrite (Macfarlan, Parker et al. 2006) apparait également comme interactant. Elle est membre du complexe INHAT (« inhibitor of acetyltransferase ») qui joue un rôle dans la modification de la chromatine et la régulation de la transcription.

Table 4 : Protéines partenaires potentielles d'intérêt identifiées après analyse protéomique des complexes THAP7 purifiés avec le protocole 2 peu stringent (*AC : numéro d'accession).*

Gène	AC	Protéine - Description	Score	Ratio IP/Contrôle
THAP7	**Q9BT49**	THAP domain-containing protein 7	157	40
SET	A5A5H4	Protein SET	59	20
HCF1C1	P51610	Host cell factor	73	8
OGT	O15294	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase	120	5

b. Comparaison des approches quantitatives label free et SILAC pour la caractérisation de complexes protéiques THAP7

Nous avons également souhaité évaluer la quantification sans marquage en la comparant avec une méthode de quantification utilisant un marquage isotopique métabolique SILAC pour l'analyse quantitative de complexes protéiques. Le SILAC permet *a priori* une quantification plus précise et pourrait ainsi aider à dégager plus clairement les partenaires *bona fide* du bruit de fond de protéines contaminantes.

Nous avons pour cela cultivé des cellules HeLa de la condition essai, dans lesquelles l'expression de THAP7 est induite, en présence d'acides aminés isotopiquement alourdis et réalisé une double marquage lysine (lysine 6, K6) et arginine (arginine 6, R6). Les cellules contrôle ont été cultivées en présence de lysine et d'arginine légères (K0, R0). Afin de s'assurer du marquage complet des protéines, 6 passages des cellules ont été réalisés. La stratégie utilisée est résumée dans la figure 33. Les cellules essai et contrôle ont été rassemblées et les différentes étapes (de la préparation des extraits protéiques à l'analyse MS) ont donc été réalisées sur un seul échantillon, permettant ainsi d'éviter l'introduction de biais expérimentaux à chacune des étapes du protocole.

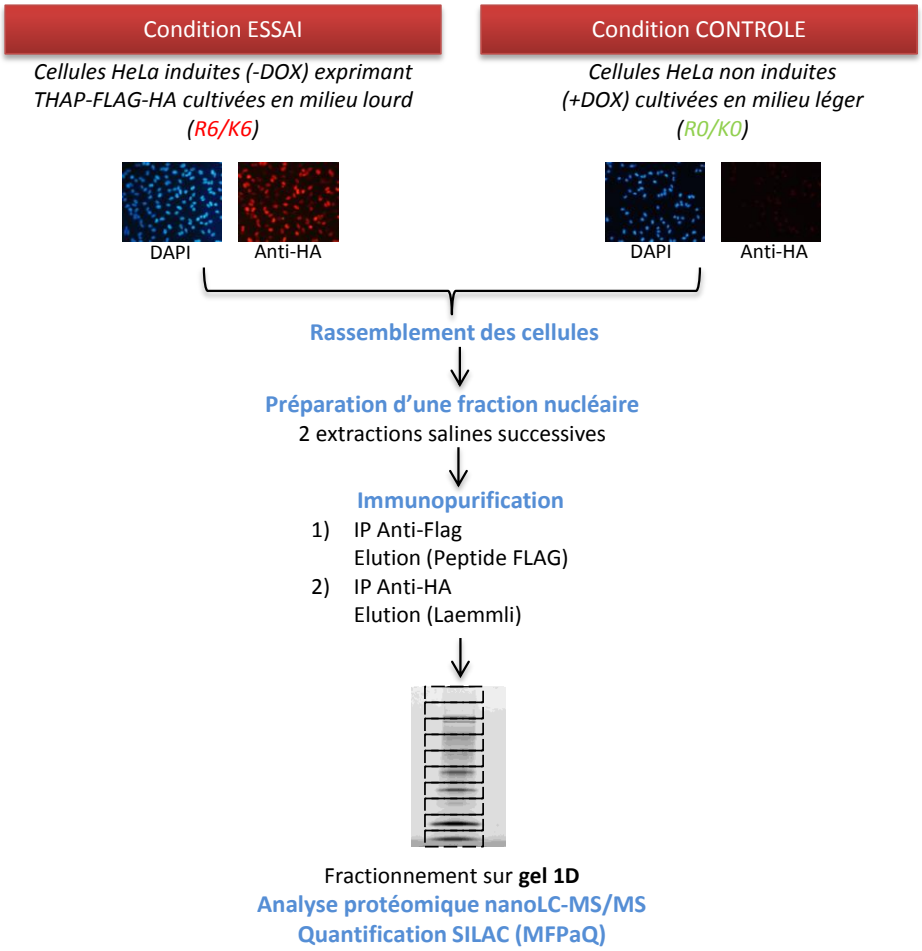


Figure 33 : Stratégie de préparation et d'analyse quantitative SILAC pour l'étude des complexes THAP7. *Les cellules HeLa exprimant ou non THAP7 et marquées ou non ont été rassemblées. Les complexes ont été purifiés à partir de l'extrait en deux étapes (double immunopurification anti-FLAG puis anti-HA). Les complexes purifiés ont ensuite été fractionnés sur gel SDS-PAGE avant d'être analysés par nanoLC-MS/MS. La quantification a été réalisée avec MFPaQ.*

Les complexes protéiques ont été isolés de la même façon que lors de l'expérience « label free » précédente (protocole 2 peu stringent). L'éluat unique obtenu au terme de la purification a été déposé sur gel 1D (figure 34A), 10 bandes ont été découpées et les extraits peptidiques issus de la digestion trypsine des protéines analysés en nanoLC-MS/MS.

Le résultat de l'analyse quantitative des données SILAC, réalisée avec MFPaQ, est représentée figure 34B. On remarque que les protéines du bruit de fond sont quantifiées avec un ratio très proche de 1 et que leur ratio présente une plus faible dispersion en comparaison des ratios « label-free ».

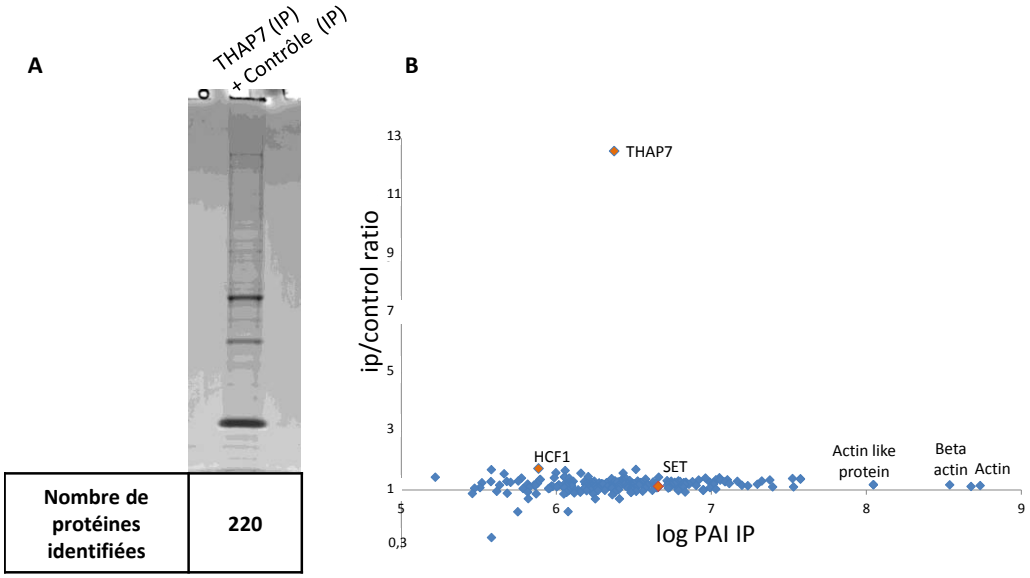


Figure 34 : Analyse quantitative SILAC des complexes THAP7 purifiés avec le protocole 2. *(A) Analyse par gel 1D SDS-PAGE. Les nombres de protéines indiqués correspondent aux protéines identifiées par Mascot après validation MFPaQ sur l'ensemble de la piste de migration (B) Représentation graphique (Ratio essai/contrôle vs log PAI essai) de l'analyse protéomique quantitative SILAC des complexes THAP7.*

Cette distribution des ratios est représentée sous forme d'histogramme dans la figure 35 où l'on observe clairement une distribution plus étroite des ratios de quantification SILAC par rapport à la quantification sans marquage, traduisant une meilleure précision de quantification.

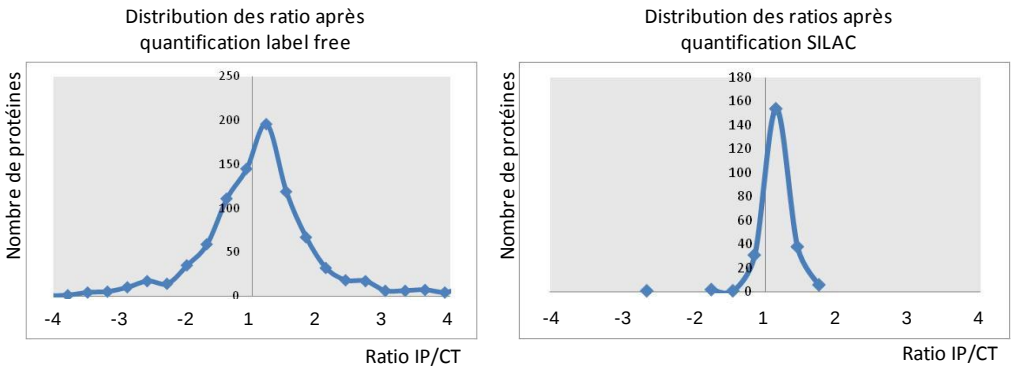


Figure 35 : Précision de la quantification sans marquage et de la quantification SILAC réalisées par MFPaQ. *Représentation graphique de la répartition des ratios des protéines quantifiées (Nombre de protéines vs ratio essai/contrôle).*

Cependant, dans cette expérience, bien que THAP7 apparaisse fortement enrichie, ses partenaires biologiques, dont HCF-1 et SET, sont quant à eux quantifiés avec un ratio proche de 1 au milieu des contaminants (Table 5 et figure 34B).

***Table 5 : Protéines partenaires potentielles d'intérêt identifiées après analyse protéomique SILAC des complexes THAP7 purifiés avec le protocole 2 peu stringent*.(AC : numéro d'accession).**

Gène	AC	Protéine - Description	Score	Ratio IP/Contrôle
THAP7	**Q9BT49**	THAP domain-containing protein 7	165	12,6
SET	A5A5H4	Protein SET	175	1,1
HCF1C1	P51610	Host cell factor	68	1,7
OGT	O15294	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase	-	-

Ce phénomène traduit probablement le caractère dynamique et labile de l'interaction de ces protéines avec THAP7 (cf figure 19 de la partie II de l'introduction). Dans cette étude, nous avons en effet mélangé les cellules essai et contrôle, marquées respectivement avec des isotopes lourds et légers (Protocole A de la figure 19). L'extrait nucléaire est alors composé des versions à la fois marquées et non marquées de toutes les protéines (mise à part la protéine appât qui n'est surexprimée que dans la condition essai). Lors de l'immunopurification, un échange peut avoir lieu entre les protéines dynamiques lourdes de la condition essai et les mêmes protéines légères de la condition contrôle. Ce type de protéine présentera donc un ratio de quantification de 1 (tout comme une protéine contaminante) alors qu'un interactant stable sera enrichi dans la condition essai. Nous avons ainsi grâce à ce protocole mis en évidence le caractère dynamique des partenaires de THAP7.

Pour mieux caractériser ces interactions labiles, il devrait être possible de combiner la grande précision de quantification du SILAC avec un protocole de purification alternatif (protocole B de la figure 19) dans lequel les échantillons ne sont rassemblés qu'après immunopurification, permettant ainsi d'éviter tout échange entre protéines lourdes et légères. Les protéines labiles devraient alors présenter un ratio important, pour autant que l'interaction soit conservée au cours des étapes biochimiques précédant la mise en commun des échantillons lourds et légers. Nous avons tenté de mettre en œuvre un protocole de ce type, mais nous n'avons pas pu reproduire les résultats de l'expérience « label-free », et mettre en évidence un enrichissement net des partenaires déjà identifiés (SET, HCF1, OGT1). Les difficultés expérimentales auxquelles nous avons été confrontés lors de l'extraction et la purification de ces complexes reflètent probablement leur nature instable *in vitro*, en dehors de leur contexte physiologique chromatinien, ou bien simplement la faible abondance des partenaires endogènes de THAP7 dans le modèle cellulaire utilisé.

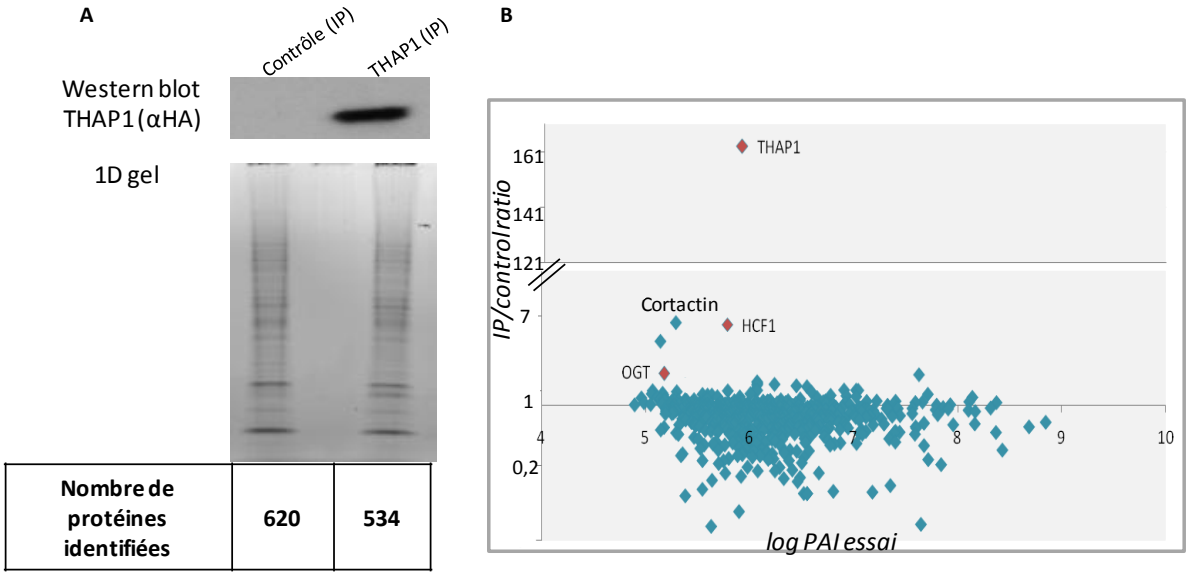
D'un point de vue méthodologique, le SILAC semble apporter une meilleure précision de quantification que la méthode « label-free », au moins dans le contexte d'une analyse après immunoprécipitation (réalisée sans réplicat technique de nanoLC-MS/MS). Néanmoins, la précision de la quantification « label-free » nous a tout de même permis de mettre en évidence les interactants de la protéine THAP7, et paraît suffisante dans ce type d'étude où des ratios importants sont attendus.

I-3.4 Complexes protéiques de THAP1

L'étude des complexes THAP1 s'est révélée encore plus délicate que celle des autres protéines THAP. L'application des protocoles 1 et 2 pour l'enrichissement des complexes n'a pas permis de détecter de partenaires spécifiques de THAP1 suite à l'analyse protéomique. De nombreuses expériences ont pourtant été réalisées mais sans succès. Le complexe THAP1 semble

ainsi peu stable et très fragile par rapport aux complexes THAP3, THAP7 ou encore THAP11. Il est également possible que l'interaction de THAP1 avec ses partenaires protéiques n'ait lieu qu'au voisinage de la chromatine, et que ces complexes n'existent pas du tout sous forme soluble. Les conditions d'extraction des protéines nucléaires utilisées ne permettent peut-être pas d'extraire des protéines fortement associées à la chromatine, mais majoritairement des complexes solubles. Pour tenter d'avoir accès à ce type de complexe, des approches de type ChIP (chromatine immunoprecipitation) utilisant une réticulation au formaldéhyde ou des approches impliquant un autre agent réticulant, le DSP (Dithiobis-succinimidyl propionate) ont alors été testées. Elles n'ont cependant pas non plus permis l'identification de partenaires spécifiques.

De précédentes études ont mis en évidence que THAP1 intervient dans la progression du cycle cellulaire et que l'un de ses gènes cibles, *RRM1*, joue un rôle important dans lors de la phase S du cycle cellulaire (Cayrol, Lacroix et al. 2007). L'association de THAP1 avec ses partenaires est donc peut-être dépendante du cycle cellulaire et pourrait avoir préférentiellement lieu lors de la phase S. Nous avons alors cherché à enrichir l'échantillon en complexes THAP1 en synchronisant les cellules HeLa en phase S du cycle cellulaire grâce à un double blocage à la thymidine. A partir de près de 10^9 cellules induites et non induites pour l'expression de la protéine d'intérêt, nous avons préparé des extraits nucléaires et purifié les complexes par une double immunopurification FLAG-HA selon le protocole 2 utilisé pour la recherche des partenaires protéiques de THAP11 et THAP7. Les complexes obtenus après l'élution finale réalisée en tampon Laemmli ont été fractionnés sur gel 1D (Figure 36A).



Nombre de protéines identifiées	620	534

Figure 36 : Analyse quantitative sans marquage des complexes THAP1 en phase S du cycle cellulaire purifiés avec le protocole 2 peu stringent*. (A) Analyse et fractionnement des complexes par gel 1D SDS-PAGE. Les nombres de protéines indiqués correspondent aux protéines identifiées par Mascot après validation MFPaQ sur l'ensemble de la piste de migration (B) Représentation graphique (Ratio essai/contrôle vs log PAI essai) de l'analyse protéomique quantitative sans marquage des complexes THAP1.*

Le différentiel entre les piste essai et contrôle n'est pas visible sur gel malgré la présence spécifique de la protéine THAP1 dans l'échantillon essai vérifiée par western blot anti-HA (Figure 36A). De nombreuses protéines contaminantes sont présentes dans chacune des pistes (534 protéines dans l'essai et 620 protéines dans le contrôle). Cet important bruit de fond apparait de façon claire sur le graphique représentant l'analyse protéomique quantitative (Figure 36B). La protéine THAP1 n'est cette fois pas parmi les protéines majoritaires de l'échantillon malgré la double immunopurification réalisée. Une quantification précise entre les deux échantillons est dans ce cas primordiale et nous a permis d'identifier deux partenaires spécifiques de THAP1, à nouveau HCF-1 et OGT. L'interaction de THAP1 avec ces deux partenaires a été validée par des expériences de co-immunoprécipitation par l'équipe de Biologie vasculaire (Figure 37).

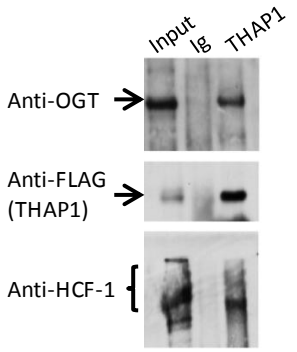


Figure 37 : Validation de l'interaction de THAP1 avec HCF-1 et OGT par co-immunoprécipitation*. L'éluat d'immunoprécipitation anti-THAP1 (anti-FLAG) a été* analysé *par western blot pour vérifier la présence de HCF-1 et OGT (anticorps anti-HCF-1 et anti-OGT) au sein des complexes THAP1. Input = extrait protéiques avant immunoputification. Ig = immunopurification contrôle.*

I-4. Discussion et conclusion

Les analyses protéomiques quantitatives des complexes THAP3, THAP11, THAP7 et THAP1 ont ainsi permis de mettre en évidence certains partenaires protéiques et en particulier deux principaux partenaires communs : le facteur de prolifération cellulaire HCF-1 et la glycosyltransférase OGT. Ces 4 protéines semblent ainsi former un complexe de base similaire.

La protéine HCF-1 joue un rôle clé dans le contrôle du cycle cellulaire et dans la régulation de la prolifération cellulaire. Elle a été initialement mise en évidence comme une protéine importante pour la transcription du virus de l'herpès dans la cellule hôte (Stern and Herr 1991). Lors de l'infection, la protéine virale VP16, un activateur transcriptionnel, lie HCF-1 de l'hôte et la recrute au niveau des promoteurs viraux pour activer leur transcription. Différentes études ont par la suite mis en évidence le rôle de cette protéine dans la régulation de la transcription et dans le contrôle de la croissance et de la division cellulaire (Wysocka, Reilly et al. 2001; Reilly, Wysocka et al. 2002). De façon intéressante, une étude protéomique visant à identifier les partenaires de HCF-1 a montré que ce facteur recrute différents complexes de modification de la chromatine, à la fois de type HDAC (histone déacétylase, répresseur transcriptionnel) et HMT (histone méthyltransférase, activateur transcriptionnel) (Wysocka, Myers et al. 2003). HCF-1 est structuré en différents domaines parmi lesquels on trouve un domaine basique et un domaine Kelch. Ce dernier est localisé dans sa partie N-

terminale et permet l'interaction de la protéine avec VP-16 ainsi qu'avec différents autres facteurs cellulaires (Freiman and Herr 1997). Par comparaison *in silico* des séquences des protéines connues pour se lier au domaine Kelch de HCF-1, un motif de 4 acides aminés (D/EHxY) impliqué dans l'interaction a pu être défini et désigné comme motif HBM (« HCF-1 binding motif »). Il est par exemple retrouvé chez des facteurs de transcription comme les protéines E2F qui assurent des fonctions importantes dans la régulation du cycle cellulaire. Il a été montré que ces facteurs interagissent avec HCF-1, qui joue un rôle de plateforme de recrutement de protéines de modification de la chromatine comme les histone H3K4 méthyltransférases MLL et Set1 (Tyagi, Chabes et al. 2007).

L'alignement des séquences des protéines de la famille THAP avait parallèlement conduit à la mise en évidence d'une séquence conservée de type D/EHSY. Grâce aux résultats obtenus à la suite des analyses protéomiques, celle-ci a pu être identifiée comme étant le motif HBM de liaison à HCF-1 (Figure 38). On le retrouve au sein de nombreuses des protéines THAP humaines (8 sur les 12 membres de la famille), dont THAP1, THAP3, THAP7 et THAP11. Les protéines THAP semblent ainsi former une grande famille de facteurs de transcription interagissant avec HCF-1, via lequel elles pourraient recruter des partenaires supplémentaires pour réguler la transcription de leurs gènes cibles.

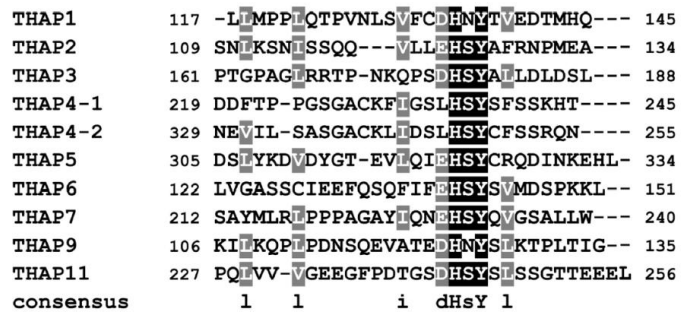


Figure 38 : Boite HBM (D/EHSY)* présente *dans la majorité des protéines THAP humaines mise en évidence par un alignement multiple des séquences des protéines THAP humaines

Des expériences fonctionnelles réalisées par l'équipe de Biologie vasculaire ont permis de valider l'interaction de HCF-1 avec THAP3 et THAP1 *in vivo* et de caractériser l'effet d'HCF-1 sur les mécanismes de régulation de la transcription assurés par THAP1 ou THAP3 (voir article *Mazars et al, JBC, 2010* à la fin de la partie). Il a ainsi été montré par des expériences d'immunoprécipitation de la chromatine (ChIP, « chromatin immunopurification ») que la protéine de fusion Gal4-THAP3 est capable de recruter HCF-1 au niveau de séquences promotrices d'un gène rapporteur Gal4-luciférase grâce à sa boite HBM, HCF-1 n'étant pas recruté lorsque cette boite est mutée. Ce recrutement engendre une activation de la transcription du gène rapporteur. De la même façon, des expériences de ChIP ont montré que THAP1 endogène recrute HCF-1 au niveau du promoteur de *rrm1*, l'un de ses gènes cibles qui joue un rôle dans la phase S du cycle cellulaire, dans les cellules endothéliales. De plus, l'extinction de l'expression de HCF-1 par siRNA conduit à une diminution des niveaux d'ARNm de *rrm1*, donc de sa transcription. HCF-1 joue ainsi, comme pour THAP3, un rôle fonctionnel dans l'activation de la transcription du gène cible de THAP1.

Le facteur HCF-1 est connu pour être fortement glycosylé. Effectivement, lors de l'analyse des complexes THAP par protéomique, de nombreux peptides portant des résidus de type N-acétyl hexosamine ont pu être identifiés (Figure 39). Les données obtenues par spectrométrie de masse montrent que la très grande majorité de ces peptides sont situés au niveau du domaine basique de la protéine (Figure 40), connu pour recruter des enzymes de modifications de la chromatine. Il a été suggéré que les modifications de type O-GlcNAc pouvaient moduler l'activité des protéines (Wells, Vosseller et al. 2001). On peut donc envisager que la présence de ces glycosylations sur HCF1 joue un rôle au niveau du recrutement de ces protéines. Elles peuvent par ailleurs expliquer la présence de la glycosyltransférase OGT au sein du complexe. Celle-ci pourrait de cette façon modifier l'état de glycosylation du domaine basique d'HCF-1, en fonction de certaines conditions cellulaires, de façon à recruter différentes enzymes de modifications de la chromatine afin de moduler la transcription de gènes cibles des protéines THAP.

A

Position	m/z	z	Theo. Mass	Exp. Mass	delta (ppm)	Score	Rank	Sequence	PTM	Peak Area	elution time
400-426	959.8239	3	2876.4444	2876.4499	1.91	30	1	YDIPATAATATSPTPNPVPSVPANPPK	HexNAc	64176	25.2
	892.1309	3	2673.3650	2673.3709	2.22	N.S.	N.S.	YDIPATAATATSPTPNPVPSVPANPPK		978981	26.8
	1337.6920	2	2673.3650	2673.3694	1.67	63	1	YDIPATAATATSPTPNPVPSVPANPPK			
489-511	830.0963	3	2487.2639	2487.2671	1.26	29	1	VTGPQATTGTPLVTMoxRPASQAGK	HexNAc	142617	19.8
	762.4044	3	2284.1846	2284.1866	0.88	18	1	VTGPQATTGTPLVTMoxRPASQAGK		375210	20.3
512-524	735.9127	2	1469.8090	1469.8108	1.27	9	1	APVTVTSLPAGVR	HexNAc	302271	21.5 / 21.8
	634.3732	2	1266.7296	1266.7302	0.51	51	1	APVTVTSLPAGVR		1121128	23.1
579-594	1006.0110	2	2010.0079	2010.0074	-0.22	9	1	TMoxAVTPGTTTLPATVK	2 HexNAc	191864	21.7 / 22.7
	904.4720	2	1806.9285	1806.9294	0.52	46.01	1	TMoxAVTPGTTTLPATVK	HexNAc	641204	23.3
	802.9332	2	1603.8491	N.D.	N.D.	N.S.	N.S.	TMoxAVTPGTTTLPATVK		N.D.	N.D.
612-637	1002.1888	3	3003.5361	3003.5452	3.03	9	4	TAAAQVGTSVSSATNTSTRPIITVHK	2 HexNAc	138768	18.1 / 19.2
	934.4946	3	2800.4567	2800.4620	1.89	N.S.	N.S.	TAAAQVGTSVSSATNTSTRPIITVHK	HexNAc	160607	18.8 / 19.8
	866.8010	3	2597.3773	2597.3812	1.49	N.S.	N.S.	TAAAQVGTSVSSATNTSTRPIITVHK		57376	20.6
638-659	1268.6730	2	2535.3280	2535.3314	1.35	45	1	SGTVTVAQQAQVVTTVVGGVTK	2 HexNAc	31763	21.9
	1167.1316	2	2332.2486	N.D.	N.D.	N.S.	N.S.	SGTVTVAQQAQVVTTVVGGVTK	HexNAc	N.D.	N.D.
	1065.5919	2	2129.1693	N.D.	N.D.	N.S.	N.S.	SGTVTVAQQAQVVTTVVGGVTK		N.D.	N.D.
683-713	1135.2730	3	3402.7917	3402.7972	1.62	58	1	VMoxSVVQTKPVQTSAVTGQASTGPVTQIIQTK	HexNAc	66822	24.9
	1067.5810	3	3199.7123	3199.7212	2.78	58	1	VMoxSVVQTKPVQTSAVTGQASTGPVTQIIQTK		3782	25.8
771-793	1210.6340	2	2419.2516	2419.2534	0.76	57	1	TIPMoxSAIITQAGATGVTSSPGIK	HexNAc	44428	24.9
	1109.0934	2	2216.1722	N.D.	N.D.	N.S.	N.S.	TIPMoxSAIITQAGATGVTSSPGIK		N.D.	N.D.
794-802	588.8403	2	1175.6649	1175.6660	0.96	5	4	SPITIITTK	HexNAc (T)	1854574	25.0
	487.2999	2	972.5855	972.5852	-0.30	44	1	SPITIITTK		114340	27.4
856-875	1113.1740	2	2224.3294	2224.3334	1.81	17	1	LVTPVTVSAVKPAVTTLVVK	HexNAc	182824	28
	1011.6345	2	2021.2500	2021.2544	2.18	N.S.		LVTPVTVSAVKPAVTTLVVK		39126	30.5
	674.7590	3	2021.2500	2021.2552	2.54	44	1	LVTPVTVSAVKPAVTTLVVK			
1233-1244	496.5697	3	1486.6834	1486.6873	2.58	9	2	HSHAVSTAAMoxTR	HexNAc	42294	8.9
	428.8763	3	1283.6041	1283.6071	2.34	N.S.		HSHAVSTAAMoxTR		86875	9.3
	642.8110	2	1283.6041	1283.6074	2.64	46	1	HSHAVSTAAMoxTR			
1483-1500	985.0272	2	1968.0416	1968.0398	-0.88	34	1	AVTTVTQSTPVPGPSVPK	HexNAc (S)	107656	22.7
	883.4901	2	1764.9622	1764.9656	1.92	38	1	AVTTVTQSTPVPGPSVPK		300096	24.2

Figure 39 : Glycosylations de la protéine HCF-1. *Les peptides portant une modification post-traductionnelle de masse 203.07937 correspondant à une N-acétylhexosamine, ont été détectés par Mascot. Dans la table sont rassemblés la position dans HCF-1 du peptide modifié identifié, son m/z, sa charge (z), les masses expérimentale et théorique, la* différence *de masse en ppm, le score Mascot, le rang Mascot, la séquence du peptide, la/les modifications post-traductionnelles qu'il porte, l'aire du pic d'élution, et son temps de rétention. Pour chaque peptide glycosylé identifié, le signal MS de la version non modifiée du peptide a été extrait (NS : non séquencé, ND : non détecté en MS).*

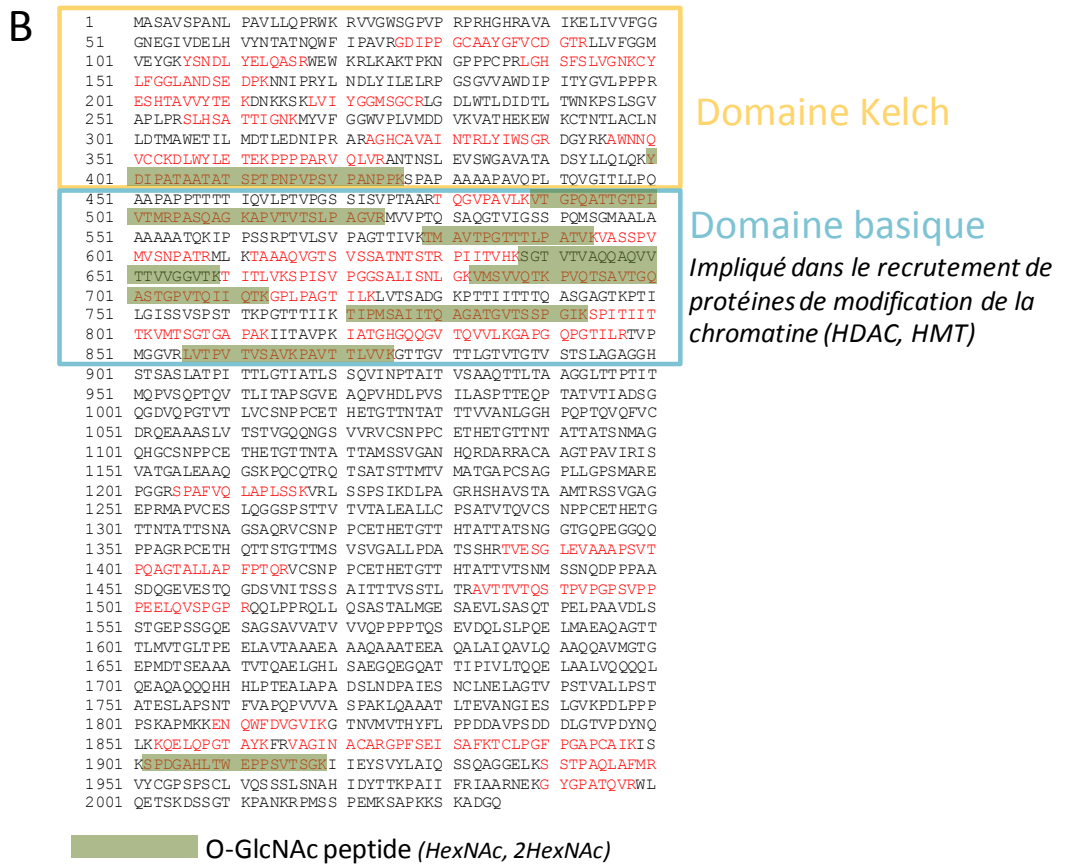


Figure 40 : Glycosylations de la protéine HCF-1*. Localisation des peptides O-GlcNAc identifiés sur la séquence d'HCF-1.*

Par ailleurs, différentes mutations au sein de THAP1 ont été associées à la dystonie primaire DYT6 (Bressman, Raymond et al. 2009; Fuchs, Gavarini et al. 2009; Xiao, Zhao et al. 2010; Song, Chen et al. 2011; Lohmann, Uflacker et al. 2012). Les dystonies sont des maladies neurologiques génétiques qui se manifestent par des troubles moteurs caractérisés par des contractions musculaires involontaires (mouvements désordonnés et répétitifs, mouvements de torsion et postures anormales) (Breakefield, Blood et al. 2008; Muller 2009). Il est donc possible que THAP1 joue un rôle dans la régulation transcriptionnelle au sein des neurones. Parmi les différentes mutations identifiées, une grande partie est retrouvée au niveau du domaine THAP de liaison à l'ADN, suggérant qu'elles pourraient perturber son activité de liaison à l'ADN (Bressman, Raymond et al. 2009; Fuchs, Gavarini et al. 2009). D'autres sont situées au niveau de la boite HBM qui permet l'interaction de THAP1 avec HCF-1. Il a été proposé que ces mutations pourraient conduire à la production d'une protéine incapable d'assurer la régulation transcriptionnelle, et par conséquent à une dérégulation de la transcription des gènes cibles de THAP1 (Tamiya 2009). Une dérégulation transcriptionnelle est d'ailleurs à l'origine d'une autre dystonie, la dystonie DYT3 associée à la maladie de Parkinson. Elle est en effet liée à une diminution de la quantité de TAF-1 (RNA polymerase 2 TATA box binding protein-associated factor 1), l'un des constituant du facteur général de la transcription TFIID, et de son isoforme spécifique dans les neurones (Makino, Kaji et al. 2007; Tamiya 2009). De façon intéressante, les données obtenues lors de nos études indiquent une interaction entre THAP1/DYT6 et OGT. Or le gène *ogt* se situe, comme le gène *taf1*, dans la région

chromosomique Xq13.1 critique pour la dystonie de type DYT3. Il est donc possible que les mutations identifiées dans cette région chromosomique pourraient non seulement affecter TAF-1, mais également OGT. OGT pourrait ainsi jouer un rôle à la fois dans la dystonie DYT6 et la dystonie DYT3.

En conclusion, l'application de méthodes de protéomique quantitative suffisamment précises s'est ainsi révélée très utile dans l'étude des complexes protéiques THAP, afin de discriminer efficacement les interactions spécifiques, et elle apparait nécessaire pour l'étude de complexes fragiles. Elle nous a permis de mettre en évidence deux partenaires communs à 4 protéines THAP, HCF-1 et OGT et ainsi de proposer un modèle fonctionnel dans lequel la protéine THAP se lie à sa séquence ADN cible et recrute HCF-1, qui recruterait ensuite lui-même des protéines de modification de la chromatine, régulant ainsi la transcription des gènes cibles de THAP (Figure 41). Ces protéines de modification de la chromatine n'ont pas été clairement identifiées dans les complexes étudiés même si de telles protéines sont apparues enrichies, mais avec de faibles scores, pour certaines THAP. Il est d'une part possible que les conditions d'isolement des complexes utilisées ne permettent pas d'avoir accès aux complexes dans leur intégralité ou que certaines interactions plus labiles soient perdues. Le complexe complet n'existe d'autre part peut-être qu'au voisinage de la chromatine, alors que nous travaillons sur des complexes solubles. Pour avoir accès à ce type de complexes, il serait peut-être nécessaire d'inclure dans le protocole d'isolement des complexes des étapes de digestion de la chromatine ou des réticulations chimiques. L'absence de ces protéines peut par ailleurs être liée au modèle cellulaire utilisé et à la surexpression de la protéine THAP. Bien que cette stratégie nous soit apparue indispensable pour pouvoir étudier ces protéines très peu abondantes, elle n'est pas forcément idéale. Les protéines sont en effet étudiées dans un modèle cellulaire différent (cellules HeLa) du modèle dans lequel elles ont été identifiées et où leur activité biologique a été mise en évidence (cellules endothéliales). Par ailleurs, contrairement à l'appât, les protéines partenaires ne sont pas surexprimées et sont probablement faiblement abondantes dans la cellule, et difficiles à identifier au sein des complexes.

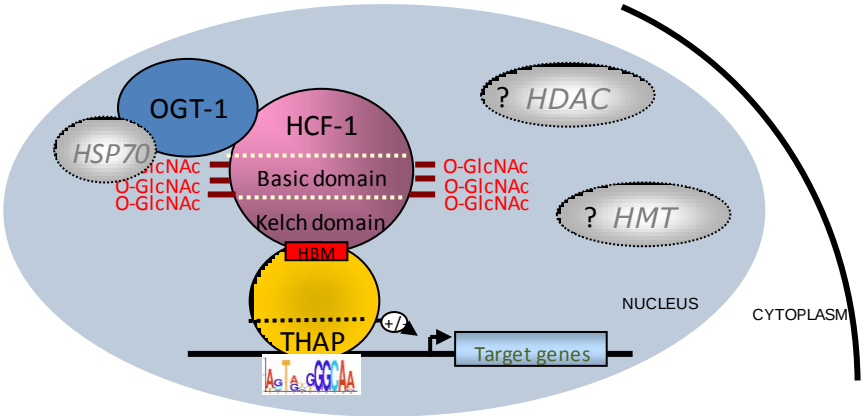


Figure 41 : Modèle fonctionnel envisagé du mécanisme d'action des protéines THAP humaines

I-5. Article *Mazars et al., JBC, 2010*

«The THAP-Zinc Finger Protein THAP1 Associates with Coactivator HCF-1 and O-GlcNAc Transferase»

A LINK BETWEEN DYT6 AND DYT3 DYSTONIAS

Raoul Mazars, Anne Gonzalez-de-Peredo, Corinne Cayrol, Anne-Claire Lavigne, Jodi L. Vogel, Nathalie Ortega, Chrystelle Lacroix, Violette Gautier, Gaelle Huet, Aurélie Ray, Bernard Monsarrat, Thomas M. Kristie, Jean-Philippe Girard

J Biol Chem. 2010 Apr 30;285(18):13364-71

Les supplementary data de l'article sont disponibles en Annexe 1 (p205)

Supplemental Material can be found at:
http://www.jbc.org/content/suppl/2010/03/03/M109.072579.DC1.html

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 18, pp. 13364–13371, April 30, 2010
Printed in the U.S.A.

The THAP-Zinc Finger Protein THAP1 Associates with Coactivator HCF-1 and *O*-GlcNAc Transferase

A LINK BETWEEN DYT6 AND DYT3 DYSTONIAS[*][S]

Received for publication, October 5, 2009, and in revised form, February 26, 2010 Published, JBC Papers in Press, March 3, 2010, DOI 10.1074/jbc.M109.072579

Raoul Mazars[‡§1], Anne Gonzalez-de-Peredo[‡§1], Corinne Cayrol[‡§1], Anne-Claire Lavigne[‡§], Jodi L. Vogel[¶], Nathalie Ortega[‡§], Chrystelle Lacroix[‡§], Violette Gautier[‡§], Gaelle Huet[‡§], Aurélie Ray[‡§], Bernard Monsarrat[‡§], Thomas M. Kristie[¶], and Jean-Philippe Girard[‡§2]

From the [‡]CNRS, Institut de Pharmacologie et de Biologie Structurale (IPBS), 205 route de Narbonne, F-31077 Toulouse, France, the [§]Université de Toulouse, UPS, IPBS, F-31077 Toulouse, France, and the [¶]Laboratory of Viral Diseases, NIAID, National Institutes of Health, Bethesda, Maryland 20892

THAP1 is a sequence-specific DNA binding factor that regulates cell proliferation through modulation of target genes such as the cell cycle-specific gene *RRM1*. Mutations in the THAP1 DNA binding domain, an atypical zinc finger (THAP-zf), have recently been found to cause *DYT6* dystonia, a neurological disease characterized by twisting movements and abnormal postures. In this study, we report that THAP1 shares sequence characteristics, *in vivo* expression patterns and protein partners with THAP3, another THAP-zf protein. Proteomic analyses identified HCF-1, a potent transcriptional coactivator and cell cycle regulator, and *O*-GlcNAc transferase (OGT), the enzyme that catalyzes the addition of *O*-GlcNAc, as major cellular partners of THAP3. THAP3 interacts with HCF-1 through a consensus HCF-1-binding motif (HBM), a motif that is also present in THAP1. Accordingly, THAP1 was found to bind HCF-1 *in vitro* and to associate with HCF-1 and OGT *in vivo*. THAP1 and THAP3 belong to a large family of HCF-1 binding factors since seven other members of the human THAP-zf protein family were identified, which harbor evolutionary conserved HBMs and bind to HCF-1. Chromatin immunoprecipitation (ChIP) assays and RNA interference experiments showed that endogenous THAP1 mediates the recruitment of HCF-1 to the *RRM1* promoter during endothelial cell proliferation and that HCF-1 is essential for transcriptional activation of *RRM1*. Together, our findings suggest HCF-1 is an important cofactor for THAP1. Interestingly, our results also provide an unexpected link between *DYT6* and *DYT3* (X-linked dystonia-parkinsonism) dystonias because the gene encoding the THAP1/DYT6 protein partner OGT maps within the *DYT3* critical region on Xq13.1.

Dystonias are neurological diseases characterized by involuntary muscle contractions which result in twisting, repetitive movements, and abnormal postures (1, 2). At least six primary forms, where dystonia is the only neurologic feature, have been described (1, 2), but the disease gene has been discovered for only two of these, *DYT1* and *DYT6* (3, 4). Early onset *DYT1* generalized dystonia is caused by mutations in the gene encoding TorsinA (TOR1A), a member of the AAA+ family of ATPases, which may function as a chaperone in the nuclear envelope and endoplasmic reticulum (3). Recently, mutations in *THAP1* have been identified as a cause of mixed-onset *DYT6* primary torsion dystonia (4–6). THAP1 is the prototype of a previously uncharacterized family of cellular factors (> 100 distinct members in the animal kingdom), defined by the presence at their amino terminus of the THAP-zinc finger (THAP-zf),[3] an atypical zinc-dependent sequence-specific DNA binding domain (7–9). THAP1 recognizes a consensus DNA target sequence of 11 nucleotides (THABS for THAP1 binding sequence) considerably larger than the 3–4 nucleotide motif typically recognized by classical C2H2 zinc fingers (8). Although THAP1 biological roles are not completely understood, data supporting an important function in cell proliferation and cell cycle pathways have been provided (10). THAP1 was found to be involved in the regulation of endothelial cell proliferation and G1/S cell cycle progression, and *RRM1*, a pRb/E2F cell cycle target gene involved in S-phase DNA synthesis, was identified as the first direct target gene of THAP1 (10). In addition to cell proliferation, THAP1 may also play roles in cell survival and/or apoptosis because it has been shown to interact with Par-4, a proapoptotic factor linked to prostate cancer and neurodegenerative diseases, including Parkinson disease (11).

Several distinct mutations in THAP1 were found in *DYT6* dystonia patients and most of them mapped to the DNA binding THAP-zf, suggesting the mutations may cause disease by disrupting the DNA binding activity of THAP1 (4, 5). Indeed, one of the THAP1 missense mutant proteins (F81L) was functionally analyzed and found to exhibit strongly reduced DNA binding affinity (4). Together, these findings supported the possibility that transcriptional dysregulation, because of mutations

* The work was supported by grants from Ligue Nationale contre le Cancer (Equipe labellisée Ligue 2009), INCA, and ANR-Programme Blanc "Regulome." This work was also supported, in part, by the Intramural Research Program of the National Institutes of Health, Laboratory of Viral Diseases, NIAID.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Methods and Figs. S1–S6.

[1] These authors contributed equally to this work.

[2] To whom correspondence should be addressed: IPBS-CNRS, 205 route de Narbonne, F-31077 Toulouse, France. Tel.: 33-5-61-17-59-67; Fax: 33-5-61-17-59-94; E-mail: Jean-Philippe.Girard@ipbs.fr.

[3] The abbreviations used are: THAP-zf, THAP-zinc finger; ChIP, chromatin immunoprecipitation; HBM, HCF-1 binding motif; HCF-1, host cell factor-1; HUVECs, human umbilical vein endothelial cells; OGT, *O*-GlcNAc transferase; HA, hemagglutinin; GST, glutathione *S*-transferase.

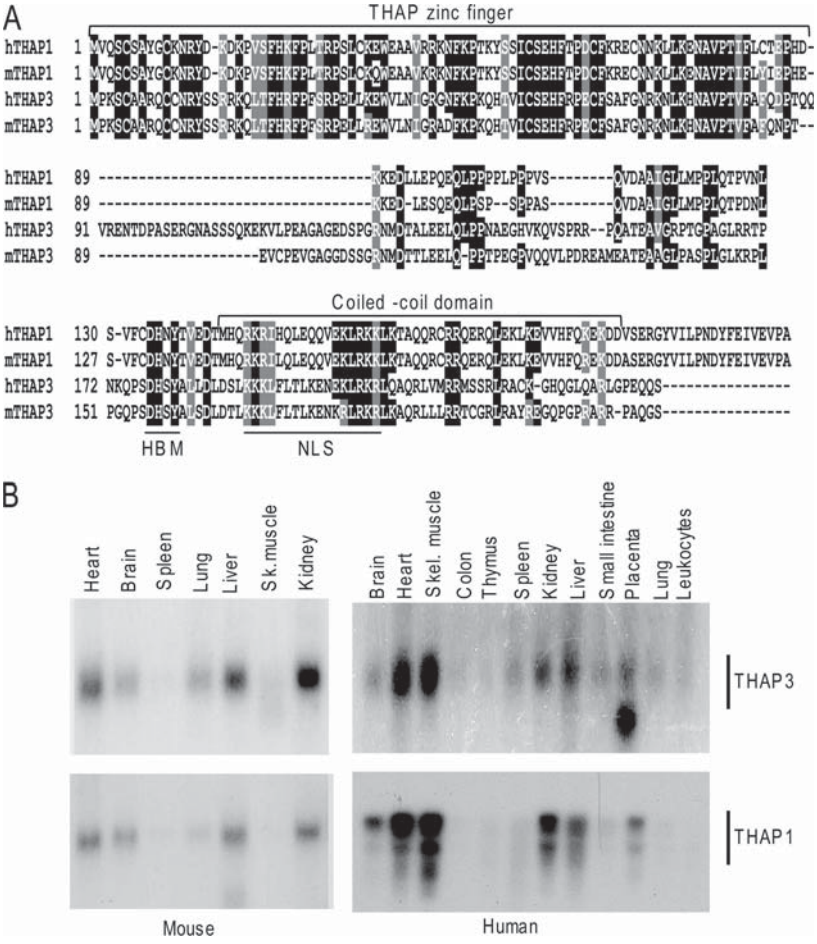


FIGURE 1. **THAP3 shares primary structure and *in vivo* expression pattern with THAP1/DYT6.** *A*, multiple sequence alignment of human THAP3 and THAP1 and their mouse orthologues. *Black boxes* indicate identical residues, whereas *shaded boxes* show similar amino acids. *B*, Northern blots containing 2 μg/lane poly(A)$^+$ RNA from adult tissues were probed with ^{32}P-labeled THAP3 and THAP1 cDNA probes. Major transcripts of about 1.3 kb (THAP3) and 2.4 kb (THAP1) were detected in both mouse and human tissues. The human THAP1 Northern blot has previously been presented (10) and is shown here only for comparison with THAP3.

in *THAP1*, might contribute to the phenotype of *DYT6* dystonia. Interestingly, transcriptional dysregulation because of reduced expression of TAF-1 (RNA polymerase II TATA-box-binding protein-associated factor 1) has previously been proposed to cause another form of dystonia, *DYT3* or X-linked dystonia-parkinsonism (12, 13).

In this study, we report that THAP1 shares sequence characteristics, *in vivo* expression profiles and cellular partners with THAP3, a previously uncharacterized member of the THAP-zf protein family. Using functional proteomics, we show that HCF-1, a potent transcriptional coactivator and cell cycle regulator (14–19), and OGT, the enzyme that catalyzes the addition of *O*-GlcNAc (20), are major cellular partners of THAP3. THAP1 also associates with HCF-1 and OGT *in vivo* and mediates recruitment of HCF-1 to the *RRM1* promoter during endothelial cell proliferation. Our results provide an unexpected link between THAP1/DYT6 and OGT, the product of a gene (*OGT*) mapped within the *DYT3* critical interval on Xq13.1 (12, 21), suggesting an unexpected link between *DYT6* and *DYT3* dystonias.

EXPERIMENTAL PROCEDURES

Cell Culture and RNA Interference—Hela cells, Hela Tet-Off cells (Clontech), and Hela HLR cells (Stratagene) were grown in Dulbecco's modified Eagle's medium. Knockdown of THAP1 and HCF-1 expression in primary human endothelial cells (HUVECs) was performed using ON-TARGET plus SMARTpool and individual siRNA duplexes (Dharmacon, Lafayette, CO) as previously described (10).

Plasmid Constructions—The full-length coding region of human THAP3 was cloned in-frame with an 82-bp linker encoding Flag/HA tags in pTRE-Tight expression vector (Clontech). The resulting vector was co-transfected with pRep-Hygro vector into Hela Tet-Off cells and stable transformants were obtained after 4 weeks of selection in 200 μg/ml hygromycin. Human THAP1 ORF was cloned into expression vector pFlag-CMV5a (Sigma) to generate pFlag-THAP1 expression vector. GST, GST-THAP1$_{1–213}$, GST-THAP3$_{162–239}$ fusion proteins were produced using pGEX-2T prokaryotic expression vector (Amersham Biosciences), and purified as previously described (11). GST-THAP3$_{H178A}$, -THAP3$_{Y180A}$, and -THAP3$_{HBM\ (DHSY/AAAA)}$ mutants were obtained from wild-type GST-THAP3$_{162–239}$ using PCR. Gal4-THAP3 and Gal4-THAP3$_{HBM}$ mutant expression vectors were generated by inserting the corresponding full-length THAP3 fragments, generated by PCR, into pCMVGT vector downstream of the Gal4-DB (amino acids 1–147). The THAP3$_{HBM}$ mutant was also cloned into pTRE-Tight expression vector. For two-hybrid assays, the open-reading frames of human THAP-zf proteins were amplified by PCR from full-length clones (Open Biosystems) and inserted into the pGADT7 expression vector (Clontech).

Sequence Analysis—Multiple sequence alignments of the THAP-zf proteins and their HBMs were generated with ClustalW according to the Blosum matrix and colored with Boxshade. Prediction of coiled-coil domains was performed with COILS and PAIRCOIL programs.

Immunoaffinity Purification and Mass Spectrometry Analysis—Nuclear extracts from induced Hela-t-THAP3 cells were prepared as previously described (22) with minor modifica-

THAP1/DYT6 Associates with HCF-1 and OGT

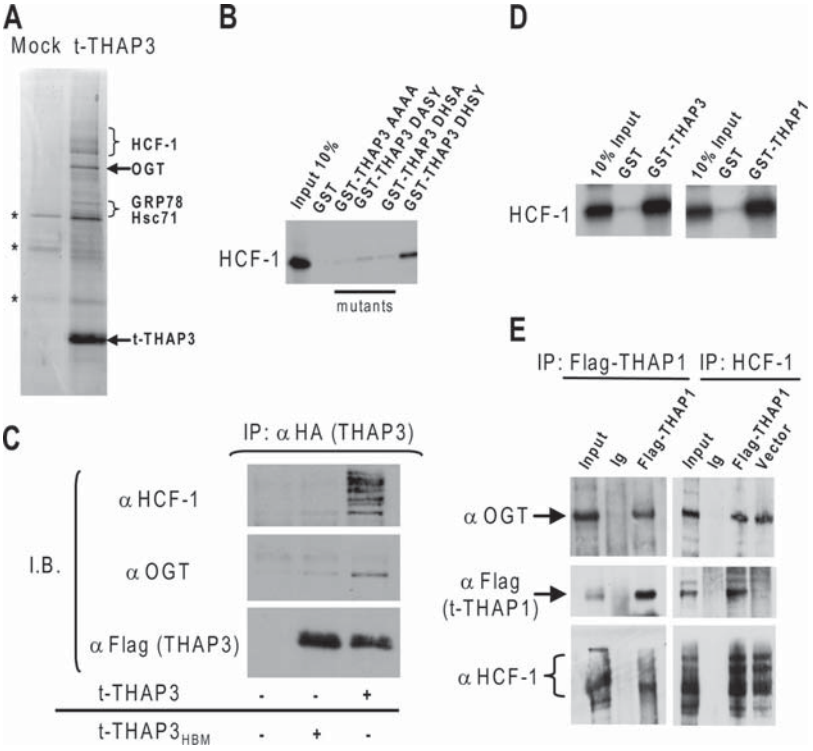


FIGURE 2. **THAP3 shares protein partners with THAP1/DYT6.** *A*, analysis of affinity-purified t-THAP3 complex. Proteins purified from induced (*t-THAP3*) or uninduced (*mock*) HeLa cells were separated on SDS-PAGE and stained with Coomassie Blue. Proteins identified by mass spectrometry as THAP3-specific partners are indicated on the *right*. Nonspecific proteins are denoted with an *asterisk* (*). *B*, wild-type THAP3 and $THAP3_{HBM}$ mutants fused with GST were incubated with *in vitro* translated [^{35}S]methionine labeled HCF-1 kelch domain, and HCF-1 binding was analyzed by autoradiography. *C*, THAP3 HBM is essential for THAP3-HCF-1 interaction *in vivo*. Flag/HA-tagged THAP3 and $THAP3_{HBM}$ mutant proteins, were expressed in Hela cells, and nuclear extracts were immunoprecipitated with anti-HA antibody. Immunoblotting was performed with anti-HCF-1, anti-OGT and anti-Flag antibodies. *D*, THAP1 interacts with HCF1 kelch domain *in vitro*. GST, GST-$THAP1_{1-213}$, and GST-$THAP3_{162-239}$ fusion proteins were incubated with *in vitro* translated [^{35}S]methionine-labeled HCF1 kelch domain and binding was revealed by autoradiography. *E*, THAP1 associates with HCF-1 and OGT *in vivo*. Nuclear extracts from Hela cells transfected with pFlag-THAP1 or pFlag-CMV5a empty vector were immunoprecipitated with anti-Flag, anti-HCF1 antibodies, or rabbit immunoglobulins (Ig). Immunoprecipitates were analyzed by Western blotting with anti-OGT, anti-Flag, or anti-HCF-1 antibodies.

tions. Nuclei were incubated for 30 min in buffer B: 20 mM Tris pH 7.4; 400 mM NaCl, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0,5% Nonidet, 1 mM phenylmethylsulfonyl fluoride, and Complete protease inhibitor mixture (Roche). The salt concentration was adjusted to 150 mM NaCl, and nuclear extracts were loaded onto a 4-ml 10–40% glycerol gradient in 150 mM NaCl buffer B, centrifuged at 50,000 rpm (200,000 × *g*) for 4 h. Fractions corresponding to a peak determined after immunoblotting were pooled and purified by immunoprecipitation with anti-FlagM2 antibody-conjugated agarose beads (Sigma), which were washed with 150 mM NaCl-containing IP buffer and eluted with a solution at 500 μg/ml of Flag peptide. As a control, a mock purification was performed from uninduced HeLa cells. Samples from induced and non-induced HeLa cells were separated by SDS-PAGE, and submitted for proteomic analysis (see supplemental information).

Immunoprecipitation and Western Blot Analyses—Nuclear extracts prepared in buffer B (20 mM Tris, pH 7,4; 400 mM NaCl, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0,5% Nonidet, 1 mM phenylmethylsulfonyl fluoride, and Complete protease inhibitor mixture, Roche), were incubated for 16 h at 4 °C with anti-HA mAb (clone HA-7; Sigma; A 2095), control Ig or polyclonal anti-HCF-1 antibodies (15), and precipitated proteins were captured with Protein G-Sepharose beads (eBioscience). Alternatively, Flag-tagged protein complexes were captured with anti-Flag (M2) agarose beads (Sigma; A 2220). After extensive washing, bound proteins were eluted in 2× SDS loading buffer and analyzed by Western blot with mAbs to HA or Flag epitope tags (Sigma), or rabbit antiserum to HCF-1 (15) or OGT (AL28; a generous gift of Dr. Gerald Hart), followed by horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit Ig (1/10000; Promega). Blots were developed with an enhanced chemiluminescence kit (GE Healthcare). Rabbit anti-THAP1 (10) and mouse anti-tubulin-α (Sigma) antibodies (1/1000) were used for some Western blot analyses.

Yeast Two-hybrid and GST Pulldown Assays—HF7c and Y190 strains were transformed with pGAL29 encoding the HCF-1 kelch domain (amino acids 3–455) fused to the GAL4 DNA binding domain, and subsequently retransformed with pGADT7 or the GADT7-THAP clones. HF7c cotransformant strains were patched on SD-Leu, Trp, and SD-Leu/Trp/His media. Y190 cotransformant strains were assayed for β-galactosidase using the Gal-Screen System (Applied Biosystems). GST pulldown assays were performed as previously described (11) using immobilized GST, GST-THAP1, GST-THAP3, or GST-$THAP3_{HBM}$ mutant proteins, and ^{35}S-labeled HCF1 kelch domain (amino acids 3–455) generated *in vitro* with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI).

Reporter Gene Assays—Hela HLR cells, which contain a Firefly luciferase chromatin-integrated reporter gene under the control of five GAL4 binding sites, were co-transfected with 50 ng of pCMVGT-Gal4-THAP3 or -Gal4-$THAP3_{HBM}$ constructs and 50 ng of *Renilla* luciferase construct (pRL-CMV, Promega), with or without 1 μg of HCF-1-V5 expression vector (15) using JetPEI reagent (2 μl/well, Polyplus Transfection). After 48h, Firefly and *Renilla* luciferase activities were assayed with the Dual luciferase assay system (Promega). Firefly luciferase activities were normalized to *Renilla* luciferase to control for transfection efficiency.

TABLE 1

THAP3-associated proteins identified by mass spectrometry

Proteins were identified with the Mascot software by database search in Uniprot. The table shows the protein name, Uniprot accession number (AC), gene name, molecular weight (MW), best Mascot score (if the protein was identified in several bands of the migration lane), number of identified peptides (No. Sequences), total number of MS/MS sequencing scans performed (No. Total MS/MS, for all the peptide ions of the protein, and all the gel bands in which it was identified, reflecting the abundance of the protein in the migration lane), and mean ratio of peptides intensity signal between the lane of immunopurified complex and control lane (Ratio ip/control). Three independent experiments were performed, and proteins displayed in the table were classified as potential specific partners (identified only in the complex lane, or with a Ratio ip/control>10) in at least 2 of 3 experiments.

Protein name	AC	Gene name	MW	Best score	No. of Sequences	No. of Total MS/MS	Ratio ip/control	No. of Experiments
Host cell factor	P51610	*HCFC1*	210,707	1180	43	734	78	3
THAP domain-containing protein 3	Q8WTV1	*THAP3*	27,384	904	20	229	191	3
UDP-*N*-acetylglucosamine–peptide *N*-acetylglucosaminyltransferase 110-kDa subunit	O15294	*OGT*	118,104	618	35	160	17	3
Heat shock cognate 71 kDa protein (Hsc71)	P11142	*HSPA8*	71,082	743	28	93	16	3
78-kDa glucose-regulated protein precursor (GRP-78)	P11021	*HSPA5*	72,402	709	31	53	24	2
Stress-70 protein, mitochondrial precursor	P38646	*HSPA9*	73,920	606	27	52	14	2

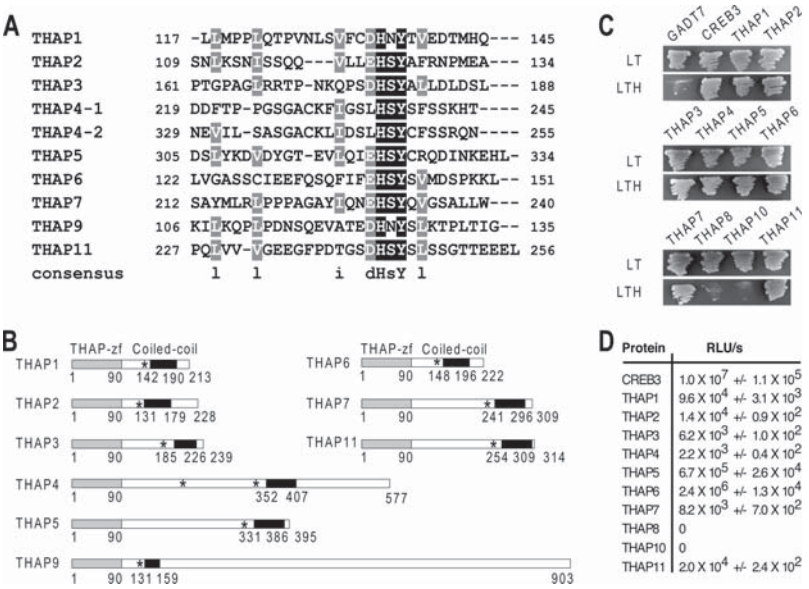


FIGURE 3. **THAP1 and THAP3 belong to a large family of HCF-1-binding THAP-zf proteins.** *A*, multiple alignment of the HBMs found in human THAP-zf proteins. A consensus is shown. *B*, primary structure of the nine human THAP-zf proteins containing consensus HBMs or HBM-like motifs. The *asterisk* indicates the position of the HBM, which is always located 4–9 amino acids upstream the predicted coiled-coil domain. *C*, two-hybrid assay. A yeast strain expressing the GAL4DB-HCF-1 kelch domain was transformed with constructs expressing the indicated human THAP-zf proteins or control protein CREB3. Transformants were patched onto SD-Leu/Trp (LT) and SD-Leu/Trp/His (LTH) plates. *D*, β-galactosidase assays. The β-galactosidase activity of each strain is expressed as relative light units per second (RLU/s). The result shown for each strain is the average of two independent colonies after subtraction of the background level obtained from the GAL4DB-HCF-1 kelch strain transformed with pGADT7 vector alone.

Chromatin Immunoprecipitations (ChIPs), qPCR Assays, and Northern Blot—ChIP-qPCR assays were performed as previously described (10) using 5 μg of antibodies specific for THAP1 (10) or HCF-1 (15). Sequences of the primers are available in the supplemental information. QuantiTect Primer Assays (Qiagen) were used for analysis of *HCF-1*, *RRM1*, and *GAPDH* expression by qPCR. For Northern blot analysis, blots of poly(A)$^+$ RNA from adult mouse and human tissues were hybridized according to the manufacturer's instructions (Clontech), using random-primed human THAP3 and THAP1 cDNA probes.

RESULTS

THAP3 Shares Similarities with THAP1—Human THAP-zf proteins are not known to share homologies outside the THAP-zf. However, we observed that, among the 12 human THAP-zf proteins (7), THAP3, shares sequence similarities with THAP1 not only within the THAP-zf (7) but also within the carboxyl-terminal domain, which contains in both proteins, a conserved bipartite nuclear localization sequence (NLS), and a coiled-coil domain predicted by COILS and PAIRCOIL programs (Fig. 1*A*). Northern blot analyses revealed that THAP3 and THAP1 also share striking similarities in their expression profile both in mouse and human tissues (Fig. 1*B*). Together, these findings indicated THAP3 and THAP1 exhibit similarities not only in their primary structure and sequence but also in their expression pattern *in vivo*.

THAP1 and THAP3 Associate with HCF-1 and OGT—To identify THAP3-interacting partners *in vivo*, nuclear extracts were prepared from Hela-t-THAP3 cells, conditionally expressing a carboxyl-terminal Flag/HA-tagged THAP3, and fractionated in a glycerol gradient. A predominant THAP3-containing complex of ~0.6 MDa (supplemental Fig. S1) was identified, immunoprecipitated from the corresponding fractions with anti-Flag antibody, and analyzed by nanoLC-MS/MS mass spectrometry (Fig. 2*A*). The purification was reproducible, as several THAP3-interacting proteins were repeatedly identified in three independent affinity purification/proteomic analyses (Table 1). Strikingly, many of the THAP3-interacting proteins corresponded to differentially processed forms of HCF-1. Numerous peptides were identified within both N- and C-terminal parts of HCF-1, including many peptides bearing *O*-glycosylated residues within the HCF-1 basic domain (supplemental Fig. S2). Interestingly, the human OGT enzyme (20) and protein chaperones with demonstrated *O*-GlcNAc lectin activity, such as the HSP70-like

THAP1/DYT6 Associates with HCF-1 and OGT

proteins Hsc71 and GRP-78 (23), were also reproducibly identified in the THAP3-associated protein complexes (Fig. 2*A* and Table 1). The association of HCF-1 and OGT with THAP3 was further confirmed by co-immunoprecipitation assays and immunoblot analyses of nuclear extracts fractionated in glycerol gradients (supplemental Fig. S1). Together, these proteomic analyses indicated that *O*-GlcNAcylated-HCF-1, *O*-GlcNAc transferase OGT, and *O*-GlcNAc lectins constitute the core components of the THAP3-associated protein complexes in human Hela cells.

Numerous viral and cellular factors bind the HCF-1 amino-terminal kelch domain, via a tetrapeptide motif ((D/E)H*X*Y), designated the HCF-1-binding motif (HBM) (19, 24, 25). A consensus HBM (DHSY) is also present in THAP3 (Fig. 1*A*), and GST pulldown assays revealed that wild-type THAP3 binds to HCF-1 *in vitro*. Mutations of the THAP3-HBM strongly reduced (single amino acid mutants) or completely abrogated (quadruple mutant) the THAP3-HCF-1 interaction *in vitro* (Fig. 2*B*). Moreover, co-immunoprecipitation assays in human cells revealed that THAP3, but not the $THAP3_{HBM}$ mutant, is able to precipitate HCF-1 and OGT (Fig. 2*C*). The association of $THAP3_{HBM}$ with OGT was also significantly reduced, suggesting THAP3 associates with OGT *in vivo* mainly through its interaction with HCF-1. We also observed that THAP1 contains a consensus HBM (DHNY) (Fig. 1*A*) and interacts with HCF-1 in two-hybrid (see below) and GST pulldown assays *in vitro* (Fig. 2*D*). In addition, THAP1 associates with HCF-1 and OGT in co-immunoprecipitation assays *in vivo* (Fig. 2*E*). These later findings indicated THAP3 shares its two major cellular partners with THAP1.

THAP1 and THAP3 Belong to a Large Family of HCF-1 Binding Factors—Comparison of the THAP1 and THAP3 sequences with other human THAP-zf proteins showed that a consensus HBM is also present in human THAP2 (EHSY), THAP5 (EHSY), THAP6 (EHSY), THAP7 (EHSY), THAP9 (DHNY), and THAP11 (DHSY), whereas THAP4 contains two HBM-like sequences (LHSY) (Fig. 3*A*). The HBMs are always located upstream predicted coiled-coil domains (Fig. 3*B*) and are found at similar positions in the orthologues of THAP-zf proteins in other species, including zebrafish and xenopus, indicating strong evolutionary conservation (supplemental Fig. S3).

In agreement with the presence of the HBM, several members of the human THAP-zf protein family were identified in large scale two-hybrid screens with the HCF-1 amino-terminal kelch domain (supplemental Fig. S4). Additional two-hybrid assays (Fig. 3, *C* and *D*) revealed that the HCF-1 kelch domain binds to human THAP-zf proteins that contain an HBM (THAP1 to THAP7 and THAP11) or HBM-like motif (THAP4) but not to those which do not possess HBMs (THAP8 and THAP10). Based on our bioinformatics analyses and two-hybrid results, we concluded that THAP-zf proteins represent a large family of HCF-1 binding factors.

THAP3 Recruits HCF-1 to a Chromatin-integrated Promoter for Transcriptional Activation—We asked if interaction of THAP3 with HCF-1 might play a role in transcriptional regulation. Because THAP3 target genes have not yet been identified, we used a chromatin-integrated Gal4-dependent luciferase reporter construct (Hela-HLR cell line). Wild-type THAP3

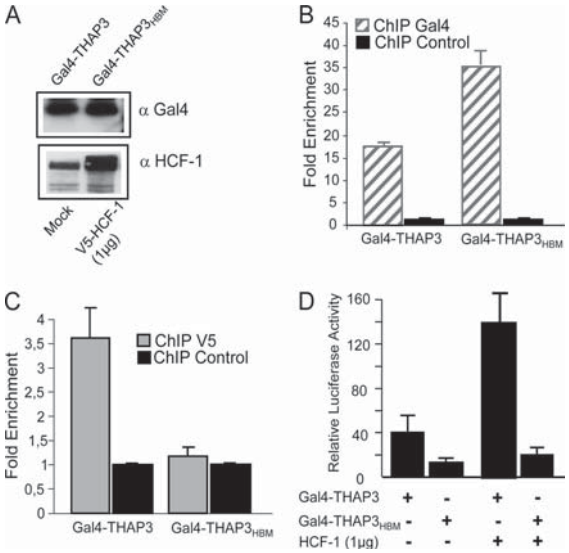


FIGURE 4. **THAP3 recruits HCF1 to a chromatin-integrated promoter for transcriptional activation.** *A*, Hela HLR cells harboring a Gal-4-dependent chromatin-integrated luciferase reporter gene were transfected with Gal4-THAP3 or Gal4-$THAP3_{HBM}$ mutant expression vectors, with or without V5-HCF-1 expression vector. Expression of HCF-1 and Gal4-THAP3 fusion proteins was analyzed by Western blotting. *B* and *C*, Gal4-THAP3, but not Gal4-$THAP3_{HBM}$ mutant, mediates recruitment of V5-HCF-1 to the chromatin-integrated Gal4-luciferase reporter gene. ChIP assays were performed with anti-Gal4 (*B*), anti-V5 (*C*), or irrelevant control antibodies, and the amount of DNA precipitated was quantified by qPCR. Fold enrichment was calculated by dividing the amount of Gal4 promoter precipitated by the different antibodies to the amount of DNA precipitated from the control U2 snRNA gene. Results are shown as means and SDs of three separate data points and are representative of at least two independent experiments. *D*, recruitment of HCF-1 by Gal4-THAP3 results in enhanced transcriptional activity of the Gal4-luciferase reporter, whereas no effect was found with the Gal4-$THAP3_{HBM}$ mutant. Normalized luciferase activities are shown as means and SDs of three independent transfection experiments.

or $THAP3_{HBM}$ mutant was expressed as Gal4 DNA binding domain-fusion proteins in Hela HLR cells, with or without V5-epitope-tagged HCF-1 (Fig. 4*A*). ChIP assays revealed that both Gal4-THAP3 and Gal4-$THAP3_{HBM}$ were significantly enriched at the Gal4 binding sites in the Gal4 reporter gene (Fig. 4*B*). In contrast, ChIP assays with anti-V5 antibodies showed that the V5-epitope-tagged HCF-1 was recruited to the Gal4 binding sites when co-expressed with Gal4-THAP3 but not when co-expressed with Gal4-$THAP3_{HBM}$ (Fig. 4*C*). Moreover, recruitment of HCF-1 by Gal4-THAP3 resulted in enhanced transcriptional activity of the Gal4-luciferase reporter gene (Fig. 4*D*), whereas no effect was found with the Gal4-$THAP3_{HBM}$ mutant. We conclude that THAP3, through its HBM, is able to recruit HCF-1 to a chromatin-integrated promoter for transcriptional activation.

Endogenous THAP1 Recruits HCF-1 to the RRM1 Promoter—We next asked if HCF-1 might play a role in transcriptional regulation by endogenous THAP1. Using primary human endothelial cells (HUVECs), we have recently shown that THAP1 associates *in vivo* with consensus THAP1 binding sites found in the *RRM1* promoter (10). Because HCF-1 has previously been found to be required for multiple stages of cell cycle progression (18), we determined whether binding of HCF-1 to

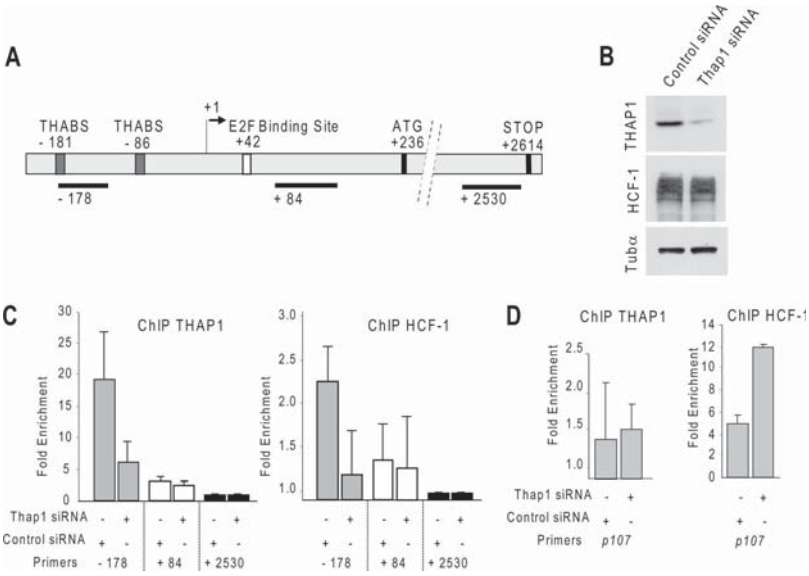


FIGURE 5. **Endogenous THAP1 recruits endogenous HCF-1 to the promoter of cell cycle-specific gene *RRM1* during cell proliferation *in vivo*.** *A*, schematic representation of the human *RRM1* promoter. The binding sites for THAP1 (THABS) and E2Fs are indicated. The position of the DNA fragments analyzed in ChIP-qPCR assays are shown. *B*, knockdown of endogenous THAP1 in primary HUVECs. THAP1, HCF-1, and Tubα (loading control) expression levels were analyzed by Western blot. Down-regulation of THAP1 does not modify HCF-1 levels. *C*, knockdown of THAP1 inhibits recruitment of THAP1 and HCF1 to the *RRM1* promoter *in vivo*. ChIP-qPCR assays with anti-THAP1 or anti-HCF1 antibodies were performed using proliferating primary HUVECs treated with control or THAP1 siRNAs. Immunoprecipitated DNA was quantified in triplicate by qPCR. Fold enrichment was calculated by dividing the amount of the *RRM1* promoter −178 and +84 DNA fragments precipitated to the amount of *RRM1* + 2530 DNA fragment precipitated (negative control). Results are representative of at least two independent experiments. *D*, knockdown of THAP1 results in increased recruitment of HCF1 to the *p107* promoter *in vivo*. ChIP-qPCR assays with anti-THAP1 or anti-HCF1 antibodies were performed as described in *C*.

THAP1 may play a role in transcriptional regulation of *RRM1*. The *RRM1* promoter (Fig. 5*A*) contains binding sites for both THAP1 and E2F transcription factors. ChIP assays were performed in HUVECs after knockdown of endogenous THAP1 expression with a pool of specific siRNAs (Fig. 5*B*). ChIP experiments performed in the presence of control siRNAs revealed that THAP1 and HCF-1 are recruited to the promoter region containing the THAP1 binding sites (−178) but not to the 5′-untranslated region containing the E2F site (+84) nor to a site close to the STOP codon (+2630) (Fig. 5*C*). In cells treated with THAP1 siRNAs, there was a significant reduction of THAP1 association. Interestingly, the reduction in THAP1 binding correlated with a parallel reduction in HCF-1 association with the *RRM1* promoter (Fig. 5*C* and supplemental Fig. S5). In contrast, the association of HCF-1 with the *p107* promoter, a promoter bound by E2Fs (19) but not by THAP1, was increased after knockdown of THAP1 (Fig. 5*D*). These experiments performed in primary human cells indicated that THAP1, rather than E2Fs, is important for HCF-1 recruitment to the *RRM1* promoter during endothelial cell proliferation.

Endogenous HCF-1 Regulates RRM1 mRNA Levels—We have previously shown that silencing of endogenous THAP1 in primary HUVECs led to a significant decrease in *RRM1* mRNA levels (10). Therefore, we asked whether HCF-1, a potent transcriptional coactivator (24), may also play a role in the regulation of *RRM1* expression. Knockdown of endogenous HCF-1 expression with a pool of siRNAs optimized to reduce off-target effects (ON-TARGET-plus SMARTpool siRNAs) resulted in down-regulation of *RRM1* mRNA levels (Fig. 6*A*), indicating HCF-1 is essential for the activation of *RRM1* during cell cycle progression. This result was confirmed using 4 individual siRNAs, which similarly reduced HCF-1 levels in HUVECs, as shown by Western blot (Fig. 6*B*) and qPCR (Fig. 6*C*) analyses. Importantly, the 4 HCF-1 siRNAs reduced *RRM1* mRNA levels (~50% inhibition), but had no significant effects on mRNA levels of *actin, GAPDH*, and *THAP1*. Together, these data provide strong evidence that endogenous HCF-1 modulates *RRM1* mRNA levels in primary human endothelial cells.

Finally, we performed knockdown experiments to address the potential role of OGT in *RRM1* regulation. As shown in supplemental Fig. S6, reduction of OGT levels did not affect RRM1 expression in HUVECs. However, this does not preclude a role for OGT in *RRM1* regulation in other cell types and for the THAP1/HCF-1/OGT complex in the regulation of other promoters.

DISCUSSION

In this study, we report that dystonia protein THAP1/DYT6 exhibits striking similarities with THAP3, another member of the THAP-zf protein family, in terms of primary structure, *in vivo* expression pattern and cellular partners. Similarly to THAP3, THAP1 associates with HCF-1, a cell cycle factor and potent transcriptional coactivator, and OGT, the enzyme that mediates *O*-GlcNAcylation of proteins. THAP3 binding to HCF-1 is mediated by a consensus HBM, a motif that is also found in THAP1. In addition to THAP1 and THAP3, seven other members of the human THAP-zf protein family, contain evolutionary conserved HBMs and possess HCF-1 binding properties, indicating THAP-zf proteins represent a large family of HCF-1-binding factors.

THAP1/HCF-1 Interaction and Transcriptional Regulation during the Cell Cycle—We found that THAP3 and THAP1 are both able to recruit HCF-1 to chromatin-integrated promoters for transcriptional activation. HCF-1 was recruited to the THAP1 binding sites, rather than the E2F site, in the promoter of cell cycle-specific gene *RRM1*, and endogenous THAP1 was found to be essential for HCF-1 recruitment during endothelial cell proliferation. This is an important result because, although recruitment of HCF-1 on cell cycle-specific promoters has been

THAP1/DYT6 Associates with HCF-1 and OGT

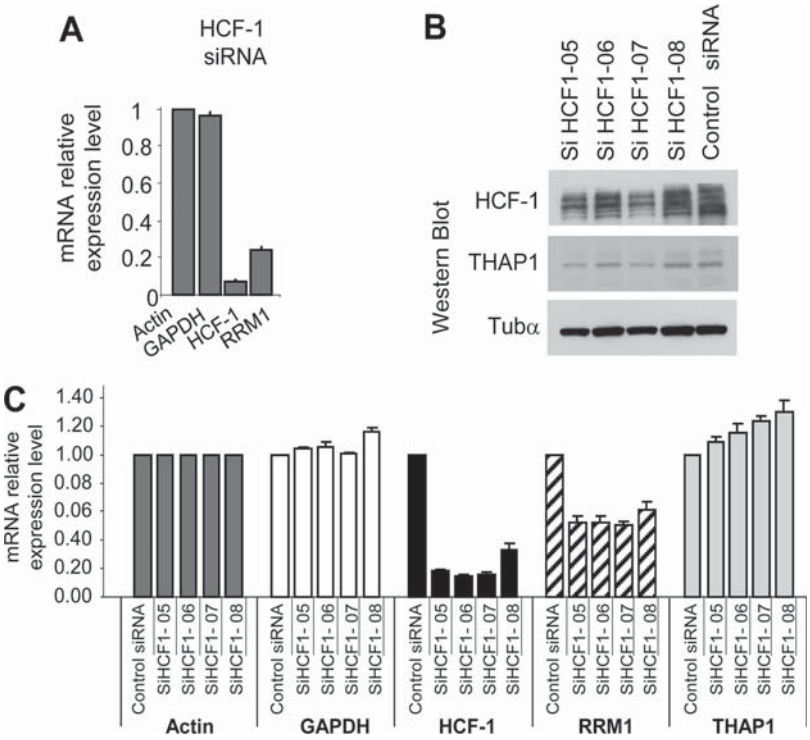


FIGURE 6. **Endogenous HCF-1 regulates RRM1 expression in primary human endothelial cells.** *A*, knockdown of endogenous HCF-1 with a pool of specific siRNAs reduces *RRM1* mRNA levels. RNA was isolated from cells transfected with ON-TARGET-plus SMARTpool HCF-1 siRNAs, 48 h after siRNA transfection, and used for qPCR analysis with the indicated human gene primers (*RRM1*, *HCF-1*, and control gene *GAPDH*). *Actin* was used as a control gene for normalization. Results are shown as means with S.D. from three separate data points. *B* and *C*, knockdown of endogenous HCF-1 with 4 individual ON-TARGET-plus HCF-1 siRNAs. *B*, HCF-1, THAP1 and Tubα (loading control) expression levels were analyzed by Western blot. *C*, knockdown of HCF-1 with individual HCF-1 siRNAs results in down-regulation of *RRM1* mRNA levels. qPCR analyses were performed as described in *A*.

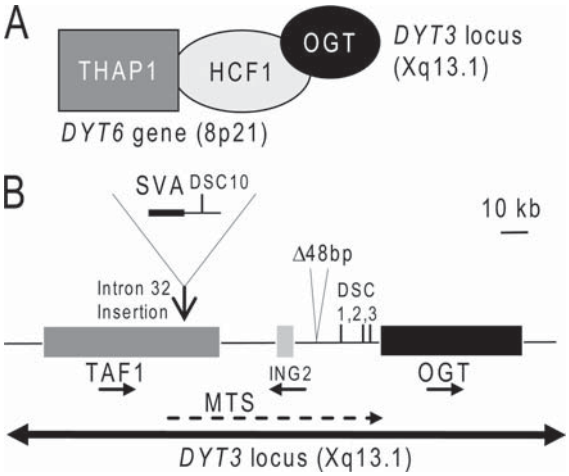


FIGURE 7. **A potential link between *DYT6* and *DYT3* dystonias.** *A*, THAP1/DYT6 associates with HCF-1 and OGT. *B*, map of the *DYT3* critical region on Xq13.1. The different genes (*TAF1*, *ING2*, *OGT*), transcripts (*MTS*) and mutations (SVA retrotransposon, DSCs, and 48-bp deletion) are indicated.

found to be correlated with the presence of E2F transcription factors (19), the absolute requirement of E2Fs for HCF-1 recruitment has not yet been shown. Knockdown of HCF-1 expression with siRNAs resulted in down-regulation of *RRM1* mRNA levels, indicating HCF-1 is essential for the activation of *RRM1* during cell cycle progression. Together, these findings suggest THAP1 and E2Fs may cooperate for HCF-1 recruitment on cell cycle specific promoters, with THAP1 playing important roles on some promoters (*i.e. RRM1*) and E2Fs (19) on other promoters (*i.e. p107, cyclin A*). They also support the possibility the critical role of HCF-1 in G1/S cell cycle progression (14, 18) may be caused by its interaction with both E2Fs and THAP1.

HCF-1 was initially discovered as a cellular cofactor in the induction of herpes simplex virus transcription and is a potent coactivator for viral VP16 and numerous transcription factors (reviewed in (24, 26)). Accordingly, our results indicate that interaction of THAP1 and THAP3 with HCF-1 plays a role in transcriptional activation. However, the interaction of HCF-1 with Ronin (the mouse orthologue of human THAP11) has recently been proposed to be involved in transcriptional repression in mouse embryonic stem cells (27). Therefore, interaction of THAP-zf proteins with HCF-1 may be important for both transcriptional activation and repression, depending on the THAP-zf protein involved, the cellular context and/or the cell cycle status. This would be similar to what has previously been proposed for the E2F family (19).

An Unexpected Link between DYT6 and DYT3 Dystonias—Our data also link THAP1 to OGT, an enzyme that plays an essential role in glucose metabolism by transferring *O*-GlcNAc to many nucleocytoplasmic proteins (20). The identification of a physical link between THAP1/DYT6 and OGT (Fig. 7*A*) supports the possibility OGT may play a role in *DYT3* dystonia/parkinsonism. Indeed, the *OGT* gene is mapped within the *DYT3* critical interval on Xq13.1 and mutations have been identified in *DYT3* patients that may affect the *OGT* regulatory regions (Fig. 7*B*). These include 4 disease-specific single nucleotide changes (DSCs), a 48-bp deletion and a SVA retrotransposon insertion in intron 32 of the *TAF1* gene (12, 21). The DSCs have been proposed to alter a multiple transcript system (MTS) expressed within the *DYT3* critical interval (21). In addition, the retrotransposon insertion has been shown to correlate with reduced TAF1 mRNA levels (12). However, the molecular pathological mechanism in *DYT3* dystonia remains unknown and the different mutations within the *DYT3* locus may also affect expression of OGT. An intriguing possibility is that OGT

and glucose metabolism may play a role in the pathology of both *DYT6* and *DYT3* dystonias.

Acknowledgments—We thank Dr. G. Hart (Baltimore, MD) for the gift of anti-OGT antibodies and S. Assbaghi for help with generation of THAP3$_{HBM}$ mutants.

REFERENCES

1. Breakefield, X. O., Blood, A. J., Li, Y., Hallett, M., Hanson, P. I., and Standaert, D. G. (2008) *Nat. Rev. Neurosci.* **9,** 222–234
2. Müller, U. (2009) *Brain* **132,** 2005–2025
3. Ozelius, L. J., Hewett, J. W., Page, C. E., Bressman, S. B., Kramer, P. L., Shalish, C., de Leon, D., Brin, M. F., Raymond, D., Corey, D. P., Fahn, S., Risch, N. J., Buckler, A. J., Gusella, J. F., and Breakefield, X. O. (1997) *Nat. Genet.* **17,** 40–48
4. Fuchs, T., Gavarini, S., Saunders-Pullman, R., Raymond, D., Ehrlich, M. E., Bressman, S. B., and Ozelius, L. J. (2009) *Nat. Genet.* **41,** 286–288
5. Bressman, S. B., Raymond, D., Fuchs, T., Heiman, G. A., Ozelius, L. J., and Saunders-Pullman, R. (2009) *Lancet Neurol* **8,** 441–446
6. Djarmati, A., Schneider, S. A., Lohmann, K., Winkler, S., Pawlack, H., Hagenah, J., Brüuggemann, N., Zittel, S., Fuchs, T., Raković, A., Schmidt, A., Jabusch, H. C., Wilcox, R., Kostić, V. S., Siebner, H., Altenmüller, E., Münchau, A., Ozelius, L. J., and Klein, C. (2009) *Lancet Neurol.* **8,** 447–452
7. Roussigne, M., Kossida, S., Lavigne, A. C., Clouaire, T., Ecochard, V., Glories, A., Amalric, F., and Girard, J. P. (2003) *Trends Biochem. Sci.* **28,** 66–69
8. Clouaire, T., Roussigne, M., Ecochard, V., Mathe, C., Amalric, F., and Girard, J. P. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 6907–6912
9. Bessière, D., Lacroix, C., Campagne, S., Ecochard, V., Guillet, V., Mourey, L., Lopez, F., Czaplicki, J., Demange, P., Milon, A., Girard, J. P., and Gervais, V. (2008) *J. Biol. Chem.* **283,** 4352–4363
10. Cayrol, C., Lacroix, C., Mathe, C., Ecochard, V., Ceribelli, M., Loreau, E., Lazar, V., Dessen, P., Mantovani, R., Aguilar, L., and Girard, J. P. (2007) *Blood* **109,** 584–594
11. Roussigne, M., Cayrol, C., Clouaire, T., Amalric, F., and Girard, J. P. (2003) *Oncogene* **22,** 2432–2442
12. Makino, S., Kaji, R., Ando, S., Tomizawa, M., Yasuno, K., Goto, S., Matsumoto, S., Tabuena, M. D., Maranon, E., Dantes, M., Lee, L. V., Ogasawara, K., Tooyama, I., Akatsu, H., Nishimura, M., and Tamiya, G. (2007) *Am. J. Hum. Genet.* **80,** 393–406
13. Tamiya, G. (2009) *Lancet Neurol.* **8,** 416–418
14. Goto, H., Motomura, S., Wilson, A. C., Freiman, R. N., Nakabeppu, Y., Fukushima, K., Fujishima, M., Herr, W., and Nishimoto, T. (1997) *Genes Dev.* **11,** 726–737
15. Narayanan, A., Ruyechan, W. T., and Kristie, T. M. (2007) *Proc. Natl. Acad. Sci. U.S.A.* **104,** 10835–10840
16. Luciano, R. L., and Wilson, A. C. (2003) *J. Biol. Chem.* **278,** 51116–51124
17. Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N., and Herr, W. (2003) *Genes Dev.* **17,** 896–911
18. Julien, E., and Herr, W. (2003) *EMBO J.* **22,** 2360–2369
19. Tyagi, S., Chabes, A. L., Wysocka, J., and Herr, W. (2007) *Mol. Cell* **27,** 107–119
20. Hart, G. W., Housley, M. P., and Slawson, C. (2007) *Nature* **446,** 1017–1022
21. Nolte, D., Niemann, S., and Müller, U. (2003) *Proc. Natl. Acad. Sci. U.S.A.* **100,** 10347–10352
22. Nakatani, Y., and Ogryzko, V. (2003) *Methods Enzymol.* **370,** 430–444
23. Lefebvre, T., Cieniewski, C., Lemoine, J., Guerardel, Y., Leroy, Y., Zanetta, J. P., and Michalski, J. C. (2001) *Biochem. J.* **360,** 179–188
24. Kristie, T. M., Liang, Y., and Vogel, J. L. (2010) *Biochim Biophys Acta* **1799,** 257–265
25. Freiman, R. N., and Herr, W. (1997) *Genes Dev.* **11,** 3122–3127
26. Wysocka, J., and Herr, W. (2003) *Trends Biochem. Sci* **28,** 294–304
27. Dejosez, M., Krumenacker, J. S., Zitur, L. J., Passeri, M., Chu, L. F., Songyang, Z., Thomson, J. A., and Zwaka, T. P. (2008) *Cell* **133,** 1162–1174

II. Recherche de partenaires protéiques de TFIIH dans les cellules ES murines et dans les cerveaux de souris

Une stratégie d'AP-MS similaire a été appliquée pour étudier chez la souris les protéines partenaires d'un complexe protéique impliqué dans l'initiation de la transcription et la réparation de l'ADN, le complexe TFIIH. Cette étude a été réalisée en collaboration avec l'équipe « Transcription et Réparation de l'ADN » dirigée par le Dr Giuseppina Giglia-Mari à l'IPBS.

II-1. Contexte biologique

Le facteur de transcription II H (TFIIH), un facteur général de la transcription, est un complexe multi-protéique composé de dix sous-unités regroupées en deux sous-complexes : le complexe cœur ou « core-complex » composé de 7 protéines fortement liées, les hélicases XPB et XPD, p62, p52, p44, p34 et TTD-A et le complexe CAK (« CDK-activating kinase ») comprenant CDK7, MAT1 et la Cycline H, qui est relié au complexe cœur via XPD (Giglia-Mari, Coin et al. 2004) (Figure 42). TFIIH est impliqué dans deux processus cellulaires primordiaux (Zhovmer, Oksenych et al. 2010). Il est d'une part essentiel dans l'initiation de la transcription médiée principalement par l'ARN polymérase II (RNAP2) mais aussi par l'ARN polymérase I (RNAP1) et ainsi dans la production d'ARN messagers (Hoogstraten, Nigg et al. 2002). Il joue d'autre part un rôle clé dans la réparation de lésions à l'ADN par excision de nucléotides (NER, « nucleotide excision repair ») (Hanawalt 2002). Dans ces deux cas, son activité majeure consiste à ouvrir la double hélice d'ADN grâce aux activités enzymatiques de ses deux hélicases XPB et XPD, afin de permettre l'initiation de la transcription par la RNAP2 sur les promoteurs ou de permettre aux protéines du NER d'accéder aux lésions de l'ADN pour les réparer (Fan, Arvai et al. 2006; Coin, Oksenych et al. 2007). Les autres composants de TFIIH permettent de réguler les activités enzymatiques du complexe (comme les activités hélicases et ATPases de XPB et XPD, et l'activité kinase de CDK7) et de le stabiliser (Coin, Marinoni et al. 1998; Kainov, Vitorino et al. 2008; Zhovmer, Oksenych et al. 2010).

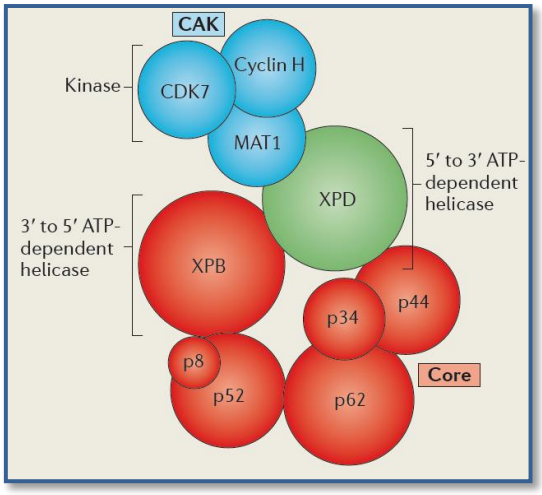


Figure 42 : Composition du facteur multiprotéique, TFIIH. D'après (Compe and Egly 2012)

Initiation de la transcription

La transcription des gènes chez les eucaryotes fait intervenir différentes protéines comme la RNAP2, les facteurs généraux de la transcription (TFIIA, B, D, E, F et H) ainsi que des protéines auxiliaires et régulatrices qui peuvent jouer des rôles d'activateurs ou de répresseurs (Thomas and Chiang 2006). La transcription de gènes sous le contrôle de la RNAP2 implique 3 étapes majeures : 1) l'initiation de la transcription au cours de laquelle ont lieu successivement (-) l'assemblage des facteurs généraux de la transcription avec la RNAP2 au niveau des promoteurs eucaryotes pour former un complexe de pré-initiation (PIC, « pre-initiation complex »), (-) la transition d'une forme inactive du PIC en un complexe d'initiation actif avec l'ouverture de l'ADN autour du site d'initiation, (-) l'initiation de la transcription où la première liaison phospho-diester est formée, puis (-) l'échappée de la RNAP2 des promoteurs qui laisse alors place à 2) l'élongation du transcrit puis à 3) la terminaison de la transcription. Au cours de l'initiation, le facteur TFIID (contenant la TBP, « TATA binding protein ») se lie aux promoteurs et forme un site d'assemblage pour les autres facteurs. TFIIB qui joue un rôle de pontage, se lie à la fois à TFIID et à la forme inactive de la RNAP2. Ce complexe protéique est alors stabilisé par l'entrée de TFIIF qui recrute TFIIE et TFIIH sur le site d'initiation de la transcription. L'activité hélicase de XPB de TFIIH permet l'ouverture du sillon de l'ADN autour du site d'initiation (Holstege, van der Vliet et al. 1996; Coin, Bergmann et al. 1999). Enfin, la phosphorylation du domaine C-terminal de la RNAP2 par l'une des sous-unités du complexe CAK, CDK7 (régulée par la cycline H et MAT1), permet l'échappée de la RNAP2 du promoteur puis l'élongation de la transcription (Rossignol, Kolb-Cheynel et al. 1997; Yankulov and Bentley 1997; Glover-Cutter, Larochelle et al. 2009). TFIIH est ainsi impliqué dans l'initiation de la transcription, la libération de la RNAP2 du promoteur et dans les phases précoces de l'élongation (Dvir, Conaway et al. 2001). Il joue également un rôle dans la ré-initiation de la transcription après les pauses de la RNAP2 (Yudkovsky, Ranish et al. 2000).

Réparation de l'ADN de type NER

TFIIH est également impliqué dans le processus de réparation de l'ADN de type NER qui permet le maintien de l'intégrité du génome en réparant un large spectre de lésions de l'ADN induites de façon endogène par le métabolisme cellulaire ou par des facteurs environnementaux telles que les radiations ultraviolettes (UV) ou les adduits chimiques. Le NER est divisé en deux voies selon le mode de détection des dommages (Figure 43). L'une est directement couplée à l'élongation de la transcription, la TCR (« transcription-coupled repair »). Elle répare les lésions du brin transcrit de gènes actifs bloquant l'élongation de la RNAP2, et permet une reprise rapide de la transcription. Les lésions sont reconnues par la RNAP2 et le facteur CSB. L'autre voie, la GGR (« global genome repair ») répare quant à elle les lésions localisées dans le reste du génome, incluant celles situées sur le brin non transcrit de gènes actifs (Hanawalt 2002). Ces lésions sont reconnues par la protéine XPC. Suite à la reconnaissance de la lésion qui fait intervenir des protéines spécifiques, les voies TCR et GGR suivent globalement un mécanisme commun. Il a cependant été décrit que le complexe CAK ne serait associé au complexe cœur de TFIIH que lors de la TCR, et que les deux sous-complexes se dissocient lors de la GGR (Coin, Oksenych et al. 2008). Au cours du NER, l'activité ATPase de XPB (stimulée par p52 et par TTD-A, qui est essentiel au NER) permet d'ancrer TFIIH au niveau de la chromatine (Coin, Oksenych et al. 2007; Kainov, Vitorino et al. 2008; Oksenych, Bernardes de Jesus et al. 2009). La protéine XPA est ensuite recrutée avec différents facteurs du NER et provoque, lors de la GGR, la dissociation du complexe CAK et du complexe cœur. Cette dissociation est dans ce cas

requise pour permettre une bonne ouverture du segment d'ADN lésé grâce à l'hélicase XPD (stimulée par p44 qui est elle-même stabilisée par p34) (Coin, Marinoni et al. 1998; Coin, Oksenych et al. 2008). Après dissociation des deux sous-complexes de TFIIH, la machinerie de réparation NER peut accéder à la lésion. Des endonucléases (XPG, XPF) sont ensuite recrutées pour exciser le morceau d'ADN contenant la lésion qui est ensuite remplacé à l'aide de la machinerie classique de réplication de l'ADN et scellé par une ADN ligase. Les facteurs du NER sont alors relargués du complexe.

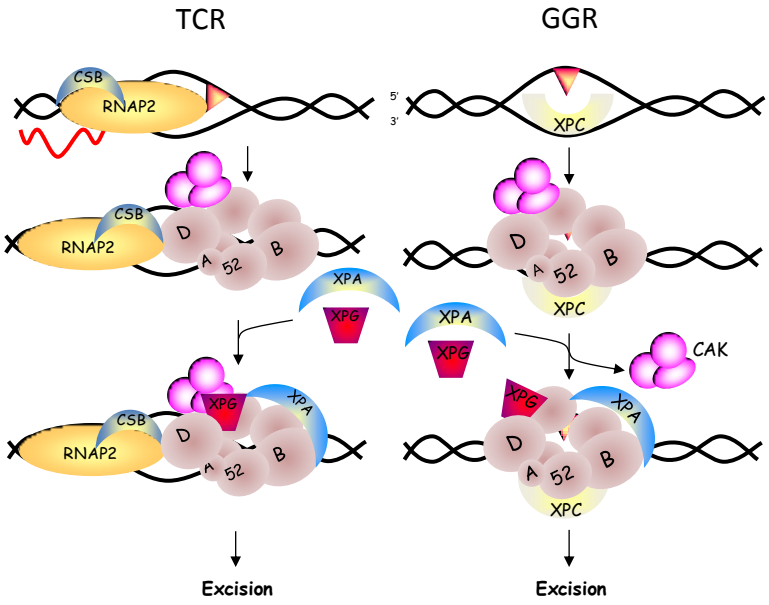


Figure 43 : Représentation schématique de l'implication de TFIIH et de certaines protéines de réparation dans les mécanismes de réparation de l'ADN de type NER, TCR et GGR.

TFIIH participe ainsi à diverses fonctions cellulaires fondamentales et des mutations au sein de ce facteur engendrent de ce fait des conséquences cliniques importantes qui sont très hétérogènes (Compe and Egly 2012). Des mutations de XPB, XPD et TTD-A sont ainsi associées à des maladies génétiques humaines telles que le xeroderma pigmentosum (XP), la trichothiodystrophie (TTD) et des formes combinées (XP/CS (« Cockayne Syndrome »), XP/TTD) (Scharer 2008). Les fonctions de transcription et/ou de réparation sont touchées dans ces syndromes, et les patients atteints présentent de très nombreux phénotypes. Le XP peut se traduire par une prédisposition au cancer de la peau, une dégénérescence neurologique progressive, un développement sexuel immature, un nanisme. Les patients atteints de XP/CS peuvent quant à eux développer un nanisme, un retard mental et une microcéphalie et ceux présentant une TTD possèdent typiquement des cheveux cassants, développent une stérilité et sont sujets à des défauts d'ordre neurologique comme un retard mental, des tremblements et une hypertonie musculaire. La très large hétérogénéité des manifestations cliniques des mutations de TFIIH suggère que chacune affecte différemment les propriétés de ce complexe (Ueda, Compe et al. 2009). Il est également envisagé que TFIIH établisse des interactions avec différents facteurs protéiques en fonction du contexte cellulaire. Une mutation donnée de TFIIH pourrait alors affecter spécifiquement l'interaction avec certains de ces facteurs influençant ainsi l'expression de gènes particuliers et expliquer les nombreux phénotypes observés.

II-2. Objectifs et stratégie mise en place

En collaboration avec l'équipe « Transcription et Réparation de l'ADN » dirigée par Giuseppina Giglia-Mari à l'IPBS, nous avons cherché à identifier les partenaires protéiques de TFIIH afin de mieux comprendre les mécanismes d'action de ce facteur. Cette équipe a généré un modèle de souris Knock-in (*xpb*$^{y/y}$) exprimant la sous-unité XPB de TFIIH étiquetée en C-terminal avec la YFP (« yellow fluorescent protein »), et une double étiquette HIS_6-HA sous le contrôle de son promoteur endogène. Grâce à cet outil, une étude de la cinétique d'association du facteur TFIIH avec la chromatine dans différents types cellulaires et différents tissus a été réalisée (Giglia-Mari, Theil et al. 2009). Il a été mis en évidence par des expériences de FRAP (« Fluorescence recovery after photobleaching ») une différence de mobilité de TFIIH en fonction du type cellulaire. Dans des cellules prolifératives, des cellules en culture (kératinocytes, cellules souches embryonnaires), la majorité de TFIIH diffuse librement et ne reste associée au niveau des promoteurs qu'une dizaine de secondes, reflétant un comportement hautement dynamique. Au contraire, dans des cellules post-mitotiques très différenciées comme les neurones, les hépatocytes ou encore les myocytes, il reste longuement immobilisé sur la chromatine lors de la transcription (Figure 44). Ces résultats suggèrent une organisation différente de l'initiation de la transcription dans ces types de cellules, qui pourrait être liée au recrutement de partenaires spécifiques. Dans un premier temps, nous avons cherché à caractériser ces partenaires éventuels au sein de cellules prolifératives comme les cellules souches embryonnaires (ES) murines.

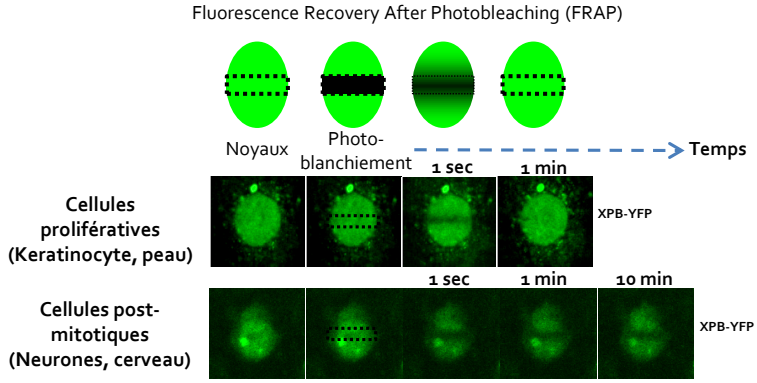


Figure 44 : Mobilité différente de TFIIH dans les cellules prolifératives et les cellules post-mitotiques. *La mobilité de TFIIH (XPB-YFP) a été mesurée par FRAP (Fluorescence Recovery after photobleaching) en évaluant le temps de retour de fluorescence, par microscopie confocale, après photoblanchiement d'une zone située au milieu de noyau. La fluorescence est rapidement retrouvée dans des cellules prolifératives (moins de 10sec) alors qu'elle n'est toujours pas totale après 10min dans les cellules post-mitotiques (Giglia-Mari, Theil et al. 2009).*

Pour cette étude, nous sommes partis de cellules ES murines exprimant de façon stable la sous-unité XPB étiquetée en C-terminal par la YFP (« yellow fluorescent protein »), un variant de la GFP, et par un tag HIS_6-HA, sous le contrôle de son promoteur endogène (Figure 45). Des tests préliminaires ont montré une plus grande efficacité de purification des complexes TFIIH en réalisant une immunopurification anti-YFP en comparaison avec une immunopurification anti-HA. En effet, cette immunopurification anti-YFP a été effectuée à l'aide d'anticorps de type GFP-trap (Chromotek) issus de camélidés, qui fixent avec une très forte affinité les protéines de type GFP ou YFP. De plus,

pour cette étude, nous avons décidé de ne pas réaliser de fractionnement de l'échantillon immunopurifié sur gel 1D SDS-PAGE. Comme décrit plus haut pour les expériences réalisées sur les complexes THAP, le bruit de fond de protéines contaminantes dans ces approches d'immunoprécipitation aboutit à des échantillons dont la complexité est de l'ordre de quelques centaines de protéines (600 espèces identifiées environ dans le cas d'une élution stringente). Etant donné la vitesse de séquençage des appareils récents (comme l'Orbitrap Velos utilisé dans le cadre de cette étude sur TFIIH), de tels mélanges peuvent désormais être caractérisés de façon exhaustive sans pré-fractionnement. Nous avons donc choisi d'analyser l'échantillon en une acquisition unique avec un gradient de 2h, permettant une analyse plus rapide et plus précise d'un point de vue quantitatif. Les complexes TFIIH ont ainsi été immunopurifiés en tirant partie de l'étiquette YFP grâce à un anticorps GFP-trap couplé de façon covalente à des billes d'agarose, puis élués en les faisant bouillir dans du tampon SDS. Les éluats n'ont pas été fractionnés, mais ont cependant été soumis à une migration rapide sur gel SDS-PAGE, afin d'éliminer le SDS et de réaliser la digestion trypsique des protéines in-gel. La migration électrophorétique a été arrêtée à l'entrée du gel de séparation, de façon à ne découper ensuite qu'une seule bande de gel de 0,5cm environ, contenant la totalité de l'échantillon. Les échantillons ont été traités puis analysés et quantifiés comme décrit plus haut dans la stratégie générale.

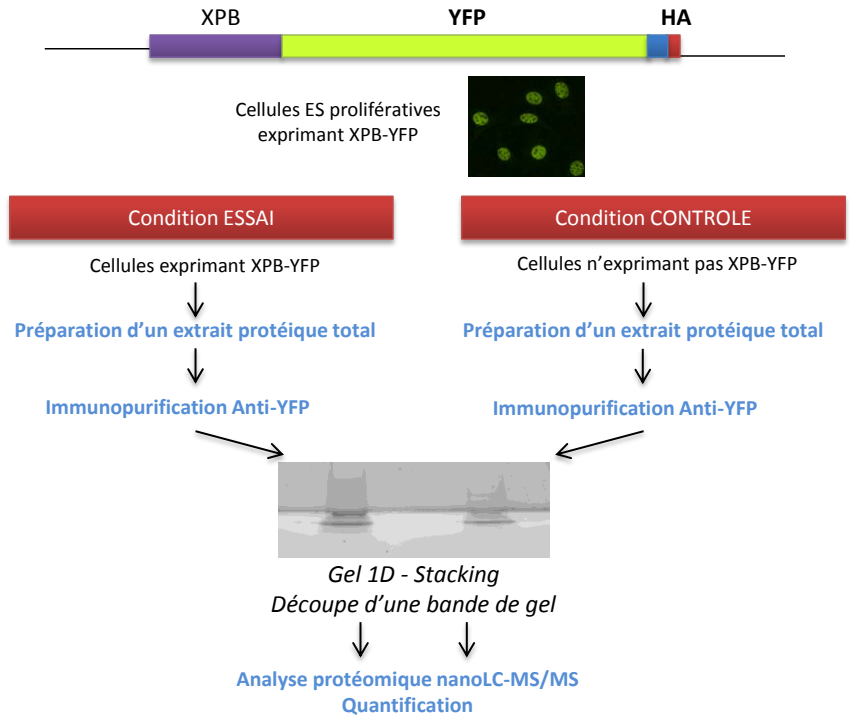


Figure 45 : Stratégie d'isolement et d'analyse quantitative sans marquage pour l'étude des partenaires d'interaction de TFIIH. *Un extrait protéique est préparé à partir de cellules ES exprimant ou non la protéine XPB étiquetée YFP-His6-HA. Les complexes TFIIH sont ensuite isolés en une étape grâce à une immunopurification anti-YFP. Les éluats de purification sont déposés sur gel 1S SDS-PAGE et concentrés en une seule bande de gel. Ils sont ensuite analysés par nanoLC-MS/MS et l'analyse quantitative entre les échantillons essai et contrôle est réalisée avec MFPaQ.*

II-3. Etude des partenaires protéiques de TFIIH dans les cellules souches embryonnaires murines

Pour tenter d'identifier des partenaires protéiques du facteur TFIIH dans des cellules prolifératives, 3.10^7 cellules souches embryonnaires (cellules ES) murines exprimant XPB-YFP-His-HA et de cellules ES sauvages ont été utilisées pour l'essai et pour le contrôle respectivement. Un extrait protéiques total a été obtenu par des cycles successifs de congélation/décongélation des cellules permettant de briser les membranes cellulaires, suivis d'une casse mécanique des noyaux (dounce) dans un tampon contenant 150mM de NaCl et 0,1% de NP-40. Les complexes TFIIH ont ensuite été enrichis et analysés comme décrit précédemment.

L'analyse quantitative de ces immunopurifications a conduit à la quantification de 433 protéines (identifiées et quantifiées avec plus d'un peptide). Le résultat de cette analyse quantitative est représenté figure 46 et table 6.

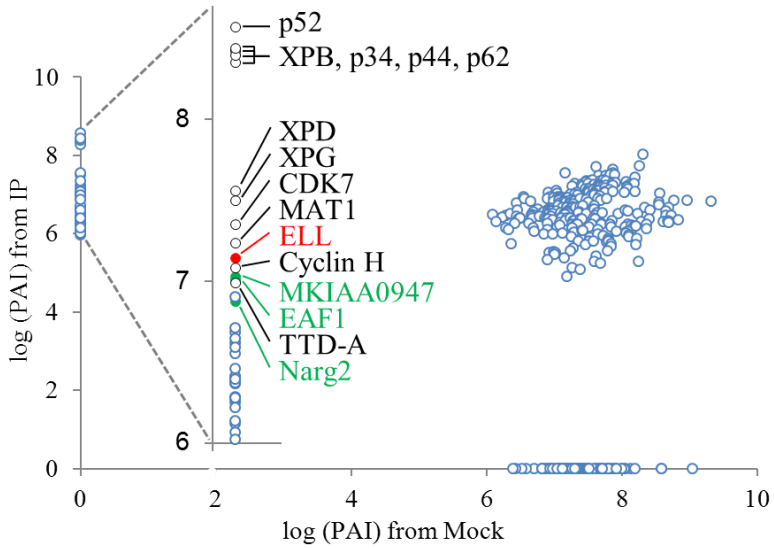


Figure 46 : Analyse quantitative sans marquage des complexes TFIIH des cellules ES murines représentée graphiquement en traçant les valeurs des log de PAI des protéines quantifiées de l'essai en fonction du contrôle. *Les protéines de TFIIH sont représentées par des ronds noirs vides, ELL par un rond rouge plein, et les partenaires potentiels supplémentaires, constituant le complexe LEC, par des ronds verts plein. Les protéines du bruit de fond sont représentées en bleu.*

Parmi les protéines spécifiques de l'essai (situées sur l'axe des ordonnées), on retrouve l'appât XPB identifiée avec un très bon score et quantifiée avec 61 peptides. Les 9 autres sous-unités de TFIIH ont également pu être identifiées et ont été quantifiées comme spécifiques de l'essai. Les trois composants du complexe CAK (MAT1, CDK7 et cycline H) sont très bien identifiés, mais présentent des scores et des PAI légèrement moins élevés que les protéines du cœur (ce qui paraît cohérent puisque le CAK est connu pour s'associer de façon plus transitoire au complexe cœur dont fait partie l'appât). Au final, les dix sous-unités de TFIIH ont été identifiées et quantifiées spécifiquement dans l'essai avec de très hauts scores et de très nombreux peptides, et font partie des protéines les plus abondantes de l'échantillon. Seule TTD-A, la dixième sous-unité, a été identifiée avec un plus faible score qui peut s'expliquer par la petite taille de la protéine (8kDa) mais

également par son interaction dynamique avec TFIIH (Giglia-Mari, Miquel et al. 2006). L'immunopurification réalisée a donc permis un très bon enrichissement du complexe, qui peut être facilement distingué des protéines contaminantes quantifiées avec un ratio centré autour de 1.

Table 6 : Protéines d'intérêt identifiées après analyse protéomique quantitative des complexes TFIIH au sein des cellules ES murines*. 4 réplicats ont été réalisés. (AC : numéro d'accession, MM(Da) : masse moléculaire en dalton, QPep : nombre de peptides quantifiés).*

AC	Protéine	Description de la protéine	Gène	Score essai (4 réplicats)	Score CT (4 réplicats)	Qpep (4réplicats)	Ratio IP/CT
P49135	XPB	TFIIH basal transcription factor complex helicase XPB subunit	ERCC3	6682;5845;1848;2886	-;-;-;-	61;60;16;34	Spécifique IP
Q8C487	XPD	TFIIH basal transcription factor complex helicase XPD subunit	ERCC2	1624;1074;1394;1496	-;-;-;-	32;24;23;35	Spécifique IP
Q9DBA9	p62	General transcription factor IIH subunit 1	Gtf2h1	3426;2422;6116;2086	-;-;-;-	40;36;21;30	Spécifique IP
O70422	p52	General transcription factor IIH subunit 4	Gtf2h4	3191;2045;9115;2741	-;-;-;-	26;27;17;21	Spécifique IP
Q9JIB4	p44	General transcription factor IIH subunit 2	Gtf2h2	2113;1691;8476;2268	-;-;-;-	30;25;16;25	Spécifique IP
Q8VD76	p34	General transcription factor IIH subunit 3	Gtf2h3	1886;1424;1408;1473	-;-;-;-	16;16;13;15	Spécifique IP
Q8K2X8	TTD-A	General transcription factor IIH subunit 5	Gtf2h5	140;179;-;-	-;-;-;-	1;2;-;-	Spécifique IP
P51949	MAT1	CDK-activating kinase assembly factor MAT1	Mnat1	874;582;2480;1174	-;-;-;-	16;11;13;17	Spécifique IP
Q03147	CDK7	Cell division protein kinase 7	Cdk7	707;185;2411;860	-;-;-;-	11;3;15;15	Spécifique IP
Q3UUW5	Cyclin H	Cyclin-H	Ccnh	557;181;1027;583	-;-;-;-	11;4;12;18	Spécifique IP
Q3UV64	XPG	DNA repair protein complementing XP-G cells homolog	ERCC5	2016;1196;550;793	-;-;-;-	36;26;18;27	Spécifique IP
O08856	ELL	RNA polymerase II elongation factor ELL	ELL	853;427;101;119	-;-;-;-	19;12;5;4	Spécifique IP
Q9D4C5	EAF1	ELL-associated factor 1	Eaf1	658;146;-;-	-;-;-;-	3;4;-;-	Spécifique IP
Q6ZQ20	KIAA0947	KIAA0947 protein	mKIAA0947	878;224;79;272	-;-;-;-	17;5;3;4	Spécifique IP
Q3UZ18	NARG2	NMDA receptor-regulated protein 2	Narg2	703;168;-;-	-;-;-;-	17;5;-;-	Spécifique IP

Par ailleurs, parmi les protéines spécifiques de la condition essai, certaines sont identifiées avec de bons scores et sont relativement abondantes (PAI élevés) et représentent ainsi des partenaires potentiels de TFIIH. On retrouve dans ces protéines candidates, la protéine XPG, une endonucléase connue pour interagir avec TFIIH et jouant un rôle dans la réparation de l'ADN de type NER (Ito, Kuraoka et al. 2007). Une autre protéine d'intérêt biologique semble interagir de façon spécifique avec TFIIH, le facteur ELL (Eleven-nineteen lysine-rich leukemia protein). C'est un facteur d'activation d'élongation de l'ARN polymérase II qui permet d'augmenter son activité d'élongation en réduisant le taux de RNAP2 en pause au niveau des promoteurs (Shilatifard, Lane et al. 1996). Elle a été identifiée comme composant d'un complexe d'élongation de la transcription des ARNm, le SEC (« super elongation complex ») (Lin, Smith et al. 2010). Parmi les protéines du SEC, seule Eaf1, un partenaire connu de ELL (Simone, Polak et al. 2001), a été identifiée dans nos analyses. ELL a également récemment été décrit comme membre d'un autre complexe, le LEC (« little elongation complex »), qui jouerait un rôle dans l'élongation de la transcription des snRNA (« small nuclear RNA ») (Smith, Lin et al. 2011; Takahashi, Parmely et al. 2011) Le LEC est composé de ELL, Eaf1 et de deux protéines de fonction jusque-là inconnue, KIAA0947 et Narg2, que nous avons également identifiées ici comme spécifiques de l'essai. Pour conforter les résultats obtenus, trois réplicats biologiques supplémentaires ont été réalisés. Seules XPG, ELL et KIAA0947 sont retrouvées systématiquement comme partenaires de TFIIH (Table 6). Cependant, certaines protéines comme Narg2 et Eaf1, bien que non identifiées dans les 4 réplicats, apparaissent spécifiques de l'essai lorsqu'elles sont identifiées, avec un PAI proche de certains composants de TFIIH et présentent de très bons scores. Elles sont de fait des partenaires potentiels de TFIIH.

Etude du rôle de ELL au sein du complexe TFIIH

Des études plus approfondies ont été réalisées dans l'équipe de G. Gigli-Mari pour comprendre le rôle joué par le facteur d'élongation de la transcription ELL en association avec TFIIH. Pour confirmer et préciser l'interaction d'ELL avec TFIIH, des expériences complémentaires de GST pull-down ont tout d'abord été effectuées à l'aide de protéines recombinantes, montrant qu'ELL interagit spécifiquement *in vitro* avec CDK7, et pourrait donc représenter un nouveau constituant du complexe CAK. Grâce à des expériences de FRAP réalisées sur des fibroblastes humains exprimant ELL-GFP, XPB-GFP et CDK7-GFP, la dynamique d'association à la chromatine de ces différentes protéines a été mesurée, à la fois après inhibition de la transcription ou en réponse à un dommage de l'ADN induit par UV. Ces mesures indiquent que la protéine ELL présente une mobilité et une dynamique d'association à l'ADN similaire au complexe CAK, suggérant qu'elle pourrait appartenir à ce complexe *in vivo* et jouer un rôle dans la réparation de l'ADN de type NER. Enfin, des expériences fonctionnelles réalisées après répression d'ELL ou CDK7 par siRNA ont effectivement montré *in vivo* un nouveau rôle de la protéine ELL dans la réponse aux dommages de l'ADN au cours de la TCR. Plus particulièrement, l'équipe de G. Gigli-Mari a montré que ELL, associée à TFIIH, facilite la reprise de la transcription par la RNAP2 après réparation des lésions. L'ensemble de ces études est détaillé dans une publication en préparation intégrée à la fin de ce chapitre.

II-4. Discussion et conclusion

Nous avons au cours de cette étude étudié les partenaires protéiques du facteur TFIIH dans les cellules ES murines. L'analyse du complexe TFIIH s'est révélée plus facile que celle des complexes THAP, traduisant probablement une plus grande stabilité du complexe, mais également peut-être une meilleure pertinence du modèle biologique utilisé, et une meilleure efficacité de l'immunopurification à l'aide des anticorps GFP-trap. Du point de vue de l'analyse protéomique, l'élimination du préfractionnement par gel SDS-PAGE a permis de réduire considérablement le temps d'analyse, et de tester plus facilement différentes conditions expérimentales de purification des complexes. L'analyse quantitative s'est révélée utile pour mettre facilement en évidence les dix sous-unités de TFIIH, et dégager des partenaires spécifiques. L'un d'entre eux est la protéine XPG, dont l'interaction avec TFIIH a déjà été plusieurs fois reportée (Ito, Kuraoka et al. 2007) corroborant ainsi nos résultats. Un autre partenaire potentiel apparaissant clairement d'après les données protéomiques est la protéine ELL, et des expériences biologiques menées par la suite ont permis de valider ELL comme un partenaire fonctionnel *in vivo*. La fonction des autres partenaires potentiels identifiés reste à déterminer.

L'un des objectifs initial de l'étude était par ailleurs de comparer les interactants de TFIIH dans différents contextes cellulaires, notamment dans des cellules prolifératives et dans des cellules post-mitotiques différenciées, de façon à expliquer la différence de mobilité de TFIIH sur la chromatine. Des premières analyses protéomiques ont été réalisées sur des complexes TFIIH immunopurifiés à partir de tissus (cerveaux de souris). Dans ces échantillons, nous avons pu à nouveau identifier la totalité des sous-unités du facteur TFIIH et mettre en évidence son interaction avec XPG, mais aucun autre partenaire potentiel n'a pu être clairement caractérisé grâce à ces expériences. Typiquement, la protéine ELL n'a pas été identifiée. Des expériences complémentaires devront être réalisées pour confirmer éventuellement que la protéine ELL n'est pas recrutée dans ce type de cellules. Par ailleurs, les conditions de préparation des complexes à partir de tissus devront

éventuellement être adaptées pour identifier des interactants interagissant plus faiblement ou de façon plus transitoire avec TFIIH.

II-5. Article en préparation

« ELL, a novel TFIIH partner is involved in transcription restart after DNA repair. »

Sophie Mourgues, Violette Gautier, Joris Slingerland, Christine Bordier, Amandine Mourcet, Fréderic Coin, Wim Vermeulen, Anne Gonzalez de Peredo, Bernard Monsarrat, Pierre-Olivier Mari, Giuseppina Giglia-Mari.

En préparation

Les supplementary data de l'article sont disponibles en Annexe 2 (p211)

ELL, a novel TFIIH partner is involved in transcription restart after DNA repair.

Sophie Mourgues[1,2], Violette Gautier[1,2], Joris Slingerland[1,2], Christine Bordier[1,2], Amandine Mourcet[1,2], Fréderic Coin[3], Wim Vermeulen[4], Anne Gonzales de Peredo[1,2], Bernard Monsarrat[1,2], Pierre-Olivier Mari[1,2], Giuseppina Giglia-Mari[1,2#]*.

1.CNRS; IPBS (Institut de Pharmacologie et de Biologie Structurale) ; 205 route de Narbonne, F-31077 Toulouse, France.
2.Université de Toulouse; UPS; IPBS; F-31077 Toulouse, France.
3.Department of Functional Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, Illkirch Graffenstaden, France.
4.Department of Genetics, Erasmus MC, GE Rotterdam, The Netherlands.

Keywords: Transcription resumption, CAK, ELL elongation factor.

DNA lesions that block transcription cause cell death. After repair of these lesions, transcription is expected to restart in order to reestablish cellular metabolism. However, transcription restart after DNA repair remains poorly described in mechanistic terms and its players are largely unknown. The transcription-coupled Repair (TCR) process highlights the interplay between DNA repair and transcription. A major actor of TCR is the general transcription factor II H (TFIIH). Using an unbiased proteomic approach, we have identified ELL (eleven nineteen lysine rich leukemia) as a novel TFIIH partner. We show here that ELL is an essential and specific factor for RNA Pol II transcription restart after DNA repair and its absence reduces the release of RNA Pol II from the chromatin during this process. These findings constitute a major stepping stone for broader investigations directed to unveil the mechanisms of transcription resumption.

UV-induced DNA damages are repaired by the Nucleotide Excision Repair (NER) system. Three DNA repair-deficient disorders emphasize the importance of NER in genome stability[1]. In eukaryotic cells, NER can be divided into two pathways: global genome repair (GGR), repairing lesions throughout the genome, and transcription-coupled repair (TCR), which repairs lesions on the transcribed strand of active genes[2]. After damage detection, the basal transcription/repair factor TFIIH, containing the XPB and XPD ATPase/helicases, is needed to locally unwind the DNA double helix around the lesion. Furthermore, UV irradiation elicits a change in the composition of TFIIH *in vitro*. Indeed, the majority of TFIIH present on the UV-damaged chromatin does not contain the ternary cyclin-activating kinase CAK complex. The CAK is only targeted to TCR and is not implicated during GGR, where it appears to be released from the core concomitantly to the recruitment of subsequent NER factors[3]. In order to improve our understanding of TFIIH functions *in vivo*, using the combination of an improved immuno-precipitation protocol and a sophisticated proteomic analysis, we have identified a new TFIIH interacting partner.

Our study highlights ELL: (eleven-nineteen lysine-rich leukemia), an RNA Polymerase II (RNA Pol II) Elongation Factor, as a new partner of TFIIH. The best characterized function of ELL is to help RNA Pol II to bypass transcription arrest sites during elongation[4]. In this study, we discovered an unexpected role of ELL in transcription coupled repair (TCR). We show that ELL facilitates resumption of transcription after removal of the transcription blocking DNA lesions by the repair machinery.

Identification of ELL as a new TFIIH partner

In order to efficiently immuno-purify the TFIIH complex and associated protein partners for proteomic analysis, we used embryonic stem cell (ES) expressing a tagged version of the helicase XPB fused to the Yellow Fluorescent Protein (YFP)[5]. A quantitative proteomic approach[6] was applied to compare immunopurified complexes and identify proteins specifically interacting with TFIIH. In our analysis, the 10 sub-units of TFIIH, of which seven (XPB, XPD, p62, p52, p44, p34, and TTDA) form a tight "core" complex, and the CAK complex (Cdk7, Mat1 and cyclinH) were identified together with the XPG endonuclease[7]. Interestingly, amongst the group of 10 new potential interacting partners, we identified ELL (Fig. 1a, Supplementary Table 1). Immunoblotting using a polyclonal antibody raised against this 68 kDa polypeptide confirmed that it co-purified

with TFIIH (Fig. 1b). These experiments unambiguously identify ELL as a new TFIIH partner. Recent studies have also identified ELL as part of the "super elongation complex" (SEC)[8,9], as well as a distinct ELL-containing complex named "little elongation complex" (LEC)[10,11]. Whereas no interacting proteins of the SEC complex have been identified in this study; we found an interaction of the LEC components (ELL-associated factor 1 (Eaf1), KIAA0947, and NARG2) with TFIIH. Although it is possible that ELL is contained in different complexes with distinct functional activities, here we only investigate the role of ELL when complexed with TFIIH.

To determine which TFIIH subunit interacts with ELL, recombinant TFIIH subunits were tested for their interaction with the bacterially produced and purified GST-ELL polypeptide. Although recombinant GST-ELL was able to pull-down the whole *in vitro* reconstituted TFIIH complex (Fig. 1c), when TFIIH subunits were individually tested, Cdk7 was the only one specifically pulled-down with GST-ELL (Fig. 1d, supplementary Fig.1). More than 10 years ago, a silver stained gel of the purified CAK complex showed the 3 polypeptides Cdk7, Cyclin H, Mat1 and a band at the same size of ELL, identified as a contaminant[12]. Presently, we postulate that Cdk7, CyclinH, Mat1 and ELL associate to form a quaternary complex rather than the ternary "CAK" complex as proposed before[13]. To investigate the *in vivo* role of ELL, we examined the dynamic behavior of ELL in living cells. For this purpose, we applied a tailor-made variant of fluorescence recovery after photo-bleaching (FRAP) assay[14,15] using cells expressing ELL-GFP, XPB-GFP[15] (representing a core TFIIH sub-unit) and Cdk7-GFP (representing the CAK sub-complex of TFIIH) (Fig.1e). Our results show that ELL-GFP mobility is similar to Cdk7-GFP mobility, whereas XPB-GFP appears to be significantly slower , supporting the idea that ELL belongs to the CAK sub-complex also *in vivo*.

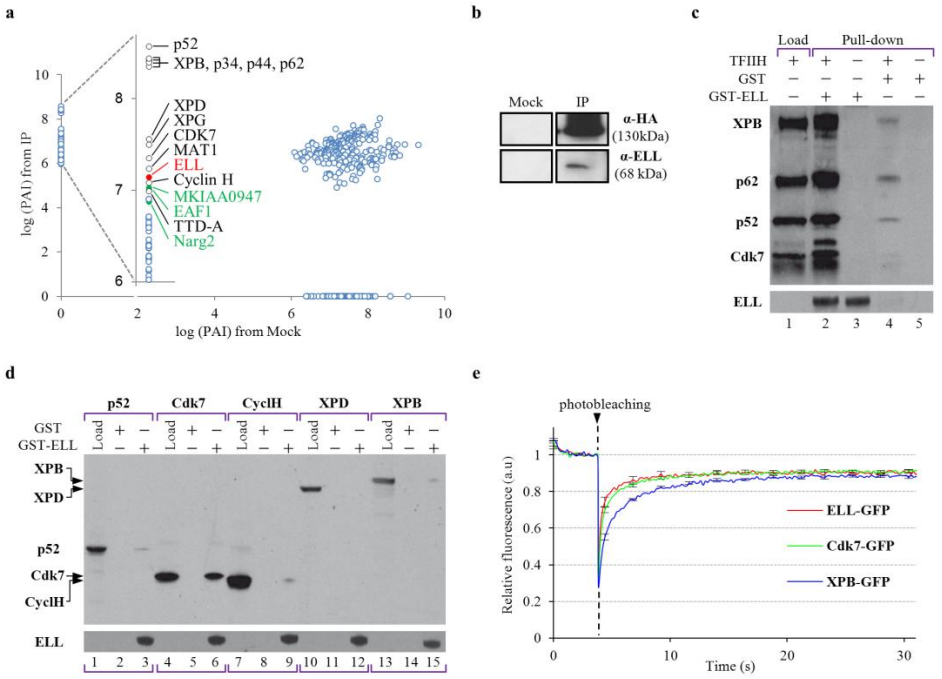


Figure 1 : Identification of a new interacting partner of TFIIH. ***a****, Quantitative proteomic analysis of proteins co-immunoprecipitated with XPB versus control sample. Data analysis was performed by plotting the logarithm of the protein abundance index (PAI) derived from mass spectrometry signal intensity measurements. In four independent experiments, all the TFIIH subunits as well as ELL were specifically detected in the immunopurified complex.* ***b****, western-blot showing a stable association between TFIIH and ELL in XPB-YFP expressing ES cells. The immunoprecipitate was analysed by western-blotting with HA and ELL antibodies.* ***c****, Bacterially expressed TFIIH subunits were tested for their ability to interact with GST-tagged ELL (lane 2) or GST alone (lane 4) coupled to glutathione-agarose beads. Proteins on the resin were resolved by SDS-PAGE and immunostained.* ***d****, SDS-PAGE analysis showing purified TFIIH subunits (lane 1, 4, 7, 10, 13), purified pull-down GST (lane 2, 5, 8, 11, 14), pull-down GST-ELL (lane 3, 6, 9, 12, 15). Cdk7 is the only sub-unit of TFIIH interacting with ELL (Supplementary Fig.1).* ***e****, Normalised relative fluorescence after photobleaching plotted against time (FRAP curves) measured in human fibroblasts stably expressing ELL-GFP (red line), Cdk7-GFP (green line) and XPB-GFP (blue line). Error bars represent the standard error of the mean (SEM) obtained from at least 15 cells.*

ELL dynamic during transcription and DNA repair.

In order to strengthen the hypothesis that ELL interacts with the CAK sub-complex *in vivo*, we investigated the behavior of ELL, XPB and Cdk7 after transcription inhibition as well as in response to DNA damage. We first compared their mobility during inhibition of transcription by DRB (5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole), a drug which is known to increase the overall mobility of XPB by decreasing its interaction with transcription initiation complexes[16,17]. All three proteins (ELL-GFP, XPB-GFP and Cdk7-GFP) exhibited a similar dynamic behavior in response to the DRB treatment (Fig. 2a-b-c), suggesting that binding of ELL to the chromatin is actually dependent on the transcriptional process. We then applied strip-FRAP on globally UV-irradiated cells and compared chromatin bound fractions of ELL, XPB and Cdk7. The immobile fractions of ELL-GFP and Cdk7-GFP were similar to each other, but significantly smaller compared to the immobile fraction of XPB-GFP (Fig. 2d-e-f). Because the core TFIIH is involved in both GGR and TCR, whereas the CAK sub-complex is only targeted to TCR sites[3,18], we hypothesized that, as the CAK sub-complex, ELL might play a role in TCR only. In order to further determine the dynamic properties of ELL while engaged in a DNA damage response mechanism, we compared the recruitment speed of ELL, XPB and Cdk7 proteins on locally micro-irradiated chromatin within cell nuclei. Indeed, we monitored protein redistribution for up to 6 minutes after local DNA damage induction with a pulsed 800 nm laser and compared the time-course recruitment of the proteins in the damaged areas. We found that ELL-GFP accumulated to a much lower extent than XPB-GFP in damaged areas. However, ELL-GFP accumulation was comparable to that of Cdk7-GFP (Fig. 2g-h). All these experiments show that ELL and CAK dynamics after DRB transcription inhibition and after DNA damage induction are closely related, suggesting that, *in vivo*, ELL is part of the CAK complex and might be involved in the TCR pathway.

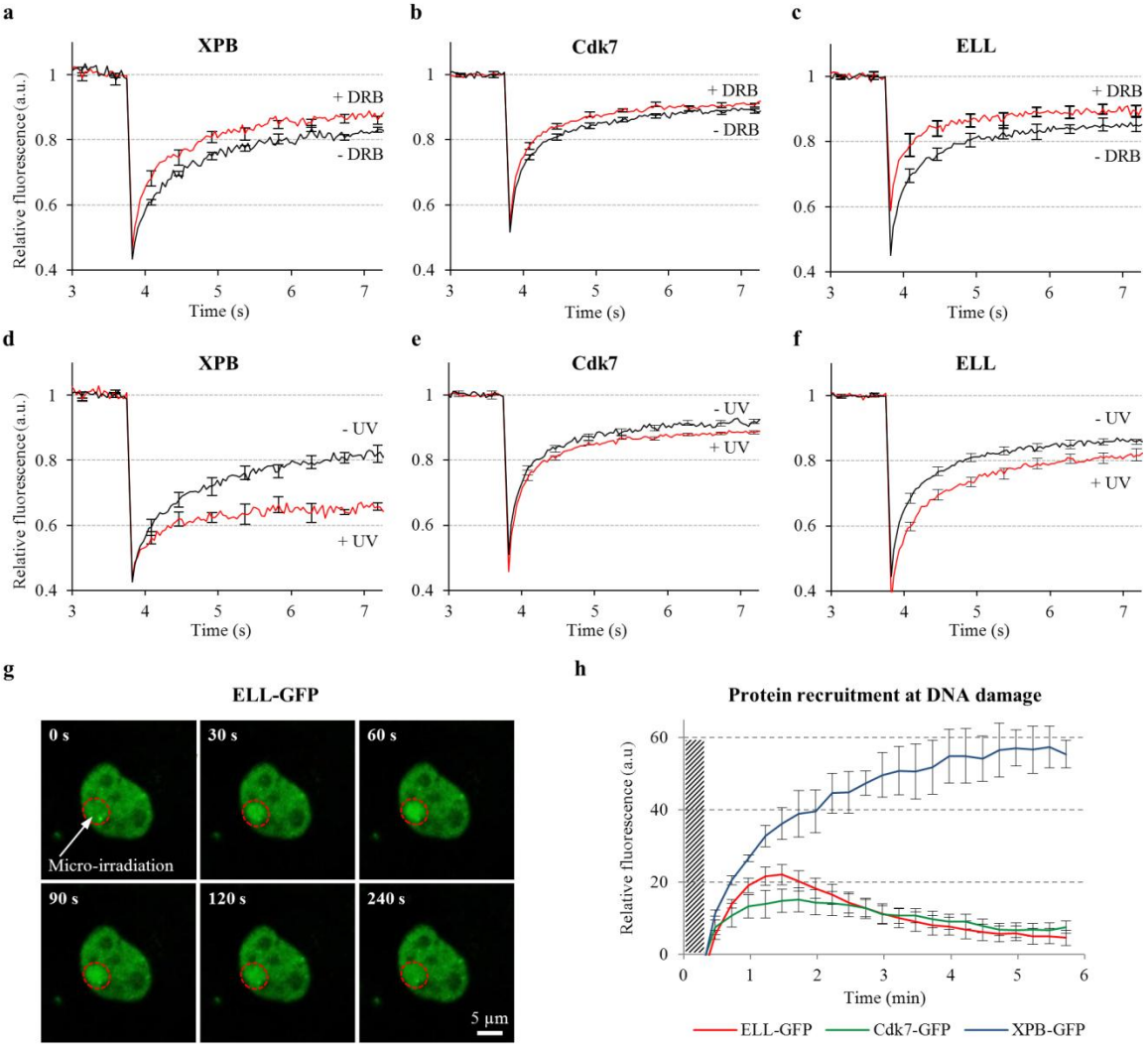


Figure 2 : ELL dynamics after transcription inhibition and during DNA repair.a, b and c, *FRAP data plotted for XPB-GFP* ***(a)****, Cdk7-GFP* ***(b)*** *and ELL-GFP* ***(c)*** *expressing human fibroblasts treated (red line) or not (black line) with DRB.* ***d, e and f,*** *FRAP data for XPB-GFP* ***(d)****, Cdk7-GFP* ***(e)*** *and ELL-GFP* ***(f)*** *expressing human fibroblasts treated (red line) or not*

(black line) with UV-C. ***g****, Confocal time-lapse images of a living human fibroblast expressing ELL-GFP seen accumulating within a circular micro-irradiated area (arrow and red circle).* ***h****, Time evolution of the relative fluorescence within the circular micro-irradiated area for XPB-GFP, Cdk7-GFP and ELL-GFP expressing human fibroblasts. Hatched region indicates the time window during which micro-irradiation induced photobleaching masks the accumulation signal leading to negative fluorescence values. Error bars represent the SEM obtained from at least 10 cells.*

ELL is involved in Transcription Coupled Repair

Because TFIIH mutants are UV-sensitive, we questioned whether ELL knock-down could have an effect on cell survival after UV irradiation. We thus repressed ELL as well as Cdk7 by siRNA in normal human fibroblasts (Supplementary Fig. 2 and Table 2), and showed that the knock-down of both proteins induces a moderate UV cytotoxicity, as measured by clonogenic survival (Fig. 3a). This result is indicative of a role in NER for both ELL and Cdk7. To further dissect the exact role of ELL and Cdk7 play during NER, we examined whether both ELL and Cdk7 knock-down affects DNA replication after repair. For this, we measured the refilling of ssDNA gaps generated by the processing of the DNA lesions (Unscheduled DNA synthesis (UDS)). NER-deficient cells, which are unable to process UV lesions, are UDS defective. As expected, we observed a reduction in UDS level in XPF-depleted cells compared to proficient cells. This is in accordance with the well-known role of the endonuclease XPF involved in processing of NER lesions. However, we did not find any difference in UDS level in the presence or in the absence of both Cdk7 and ELL suggesting that ELL, as well as Cdk7, has no role during GGR (Fig. 3b). This is in accordance with an *in vitro* repair assay[19], showing that TFIIH can support repair *in vitro* even in the absence of purified ELL (Supplementary Fig. 3). As no effect of the protein was found on UV-induced DNA repair synthesis (UDS), which is mainly a measure of GGR efficiency, we then investigated its specific role in TCR. Whereas transcription is not affected in cells depleted in ELL and Cdk7 proteins (Supplementary Fig. 4), knock-down of ELL as well as Cdk7 clearly resulted in a massive reduction of RNA synthesis after UV irradiation, a measure that is specific for TCR (Fig. 3c-d). Together, these data show that ELL, like Cdk7, is specifically involved in TCR either directly during the repair process or in the restart of transcription after completion of the repair reaction.

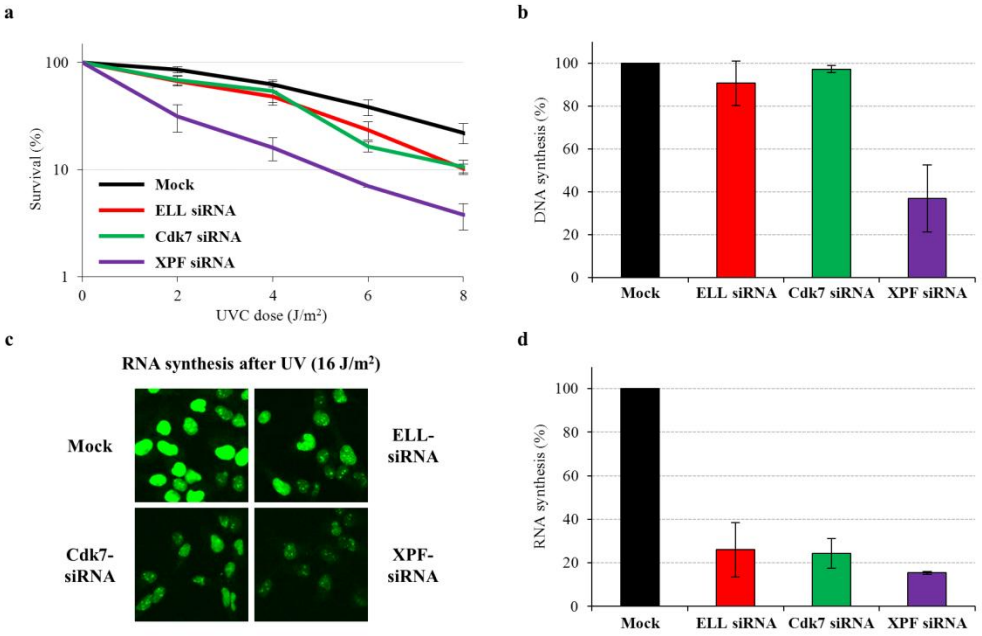


Figure 3 : Clonogenicity, Unscheduled DNA synthesis and RNA Recovery synthesis of ELL depleted cells. ***a****, UV sensitivity of immortalized human fibroblasts (MRC5) determined by colony-forming ability as a function of UV exposure. Cells were treated with siRNA against ELL (red line), Cdk7 (green line) and XPF (purple line). Black line indicates a mock siRNA treatment. The data is plotted as the logarithm of the mean colony number relative to the non-irradiated cases (in percent ± SEM)* ***b****, Unscheduled DNA Synthesis (UDS) determined by 5'ethynyl-2'-deoxyuridine (EdU) incorporation after UV-C exposure in MRC5 cells after siRNA against indicated factors. For each sample, at least 100 nuclei were analysed in total from 3 independent experiments. The data is plotted relative to the mock treated cells (as a percentage) with error bars representing the SEM.* ***c****, Confocal images of representative fields of cells analysed for RNA Recovery Synthesis (RRS). MRC5 cells were exposed to UV and incubated in the presence of 5-ethynyl uridine (EU), after siRNA-mediated knockdown of the indicated factors. Images were captured using the same parameters for all samples.* ***d****, Corresponding RRS data plotted relative to the mock. For each sample, at least 100 nuclei were analysed over 3 independent experiments; error bars represent the SEM.*

ELL is essential for transcription resumption after repair.

To discriminate whether ELL plays a role in either the repair process or in the restart of transcription after repair, we designed a new protocol to measure specifically repair replication during TCR. GGR-deficient cells (XPC negative cells) were used for this assay, to assure that the measure of replication repair (TCR-UDS) on locally damaged areas was exclusively due to TCR reactions. A rapid nonradioactive technique for measuring TCR-UDS using nucleoside analog EdU[20] was coupled with γH2AX immunofluorescence[21-23] labeling after local UV irradiation (Supplementary Fig. 5). With this new method, we were able to measure for the first time TCR repair replication in the absence of ELL, Cdk7, XPF and CSB by measuring EdU incorporation in UV-irradiated GGR-deficient cells (Fig. 4a-b). As expected, a repair defect, hence a low TCR-UDS, is observed in XPF and CSB knock-down cells. However, surprisingly, ELL knocked down cells have a normal level of TCR-UDS and do not show any inability of processing and completing the TCR repair reaction, whereas Cdk7 has a comparable deficiency as XPF and CSB knocked down cells. Interplay between transcription and repair during TCR remains poorly understood mainly because no test was available to specifically discriminate these two processes in cells. Our new assay, measuring specifically TCR-UDS, in combination with the results obtained by the classical recovery of RNA synthesis (RRS) assessment gives an unprecedented opportunity to differentiate and study both processes.

The absence of recovery of RNA synthesis (RRS) and the poor repair process (TCR-UDS) following UV-irradiation in knock-down Cdk7 cells are conclusive about the requirement of this kinase in the repair process during TCR *in vivo*. We found that Cdk7 is an essential kinase during repair and we propose that Cdk7 could be a potential candidate to catalyze the phosphorylation of NER factors[24]. Surprisingly, the absence of RRS and the unchanged TCR-UDS following UV-irradiation in knock-down ELL cells suggested a molecular scenario in which ELL is not directly involved during the repair reaction but might have a role exclusively and specifically during RNA Pol II transcription resumption after UV. Previous studies hypothesized that during TCR, stalled RNA Pol II has to be displaced by backtracking or degradation to allow access to the NER machinery[25]. Although there is no functional difference between these two options (regression or removal), it is intuitive that the regression mode might be preferable for mammalian genes, which are generally much larger than those of prokaryotes. In principle, TCR might be initiated by remodeling the RNA Pol II without removal from the arrested site[26,27]. To visualize the effect of ELL on RNA Pol II dynamic behavior on chromatin, we performed FRAP experiments on GFP-tagged RNA Pol II in MRC5 cells depleted or not for ELL protein. In the absence of UV damage, ELL protein has no effect on RNA Pol II mobility on the chromatin (Fig. 4c and Supplementary Fig. 6). Remarkably, UV-irradiation of ELL depleted cells induces RNA Pol II retention to the chromatin (Fig. 4d) strongly favoring a model in which functional TCR complex assembly does not require the dissociation of UV-stalled RNA Pol II from the chromatin.

Discussion

In this report, we describe ELL as a new previously unidentified partner of TFIIH and more specifically an interacting factor of the CAK sub-complex, which can hence be considered to be a quaternary complex. ELL, as well as the CAK sub-complex, plays an essential role during TCR but their functions are distinct. While Cdk7 plays a direct role in the repair reaction during TCR, ELL has an essential role during RNA Pol II transcription resumption after repair and its absence reduce the capacity of RNA Pol II to be released from the chromatin.

The positive transcriptional elongation factor ELL and Cdk9, as part of the SEC complex, have been found to play a central role in the resumption of arrested RNA Pol II molecules[28]. Nevertheless, we did not find any component of the SEC, including Cdk9, in our proteomic analysis. On the contrary, we show that together with TFIIH, ELL and the complete LEC complex was precipitated. The LEC complex have been known to be specialized for SnRNA transcription[10], however there is no evidence about its cellular functions. After repair completion, we might hypothesize two different scenarios to commit RNA Pol II restart, both implying a critical role of ELL protein. Either ELL acts together with the LEC to enhance the restarting of transcription or ELL enables the subsequent recruitment of positive elongation factor like Cdk9. Nevertheless, further investigation will be needed to discriminate between these two scenarios.

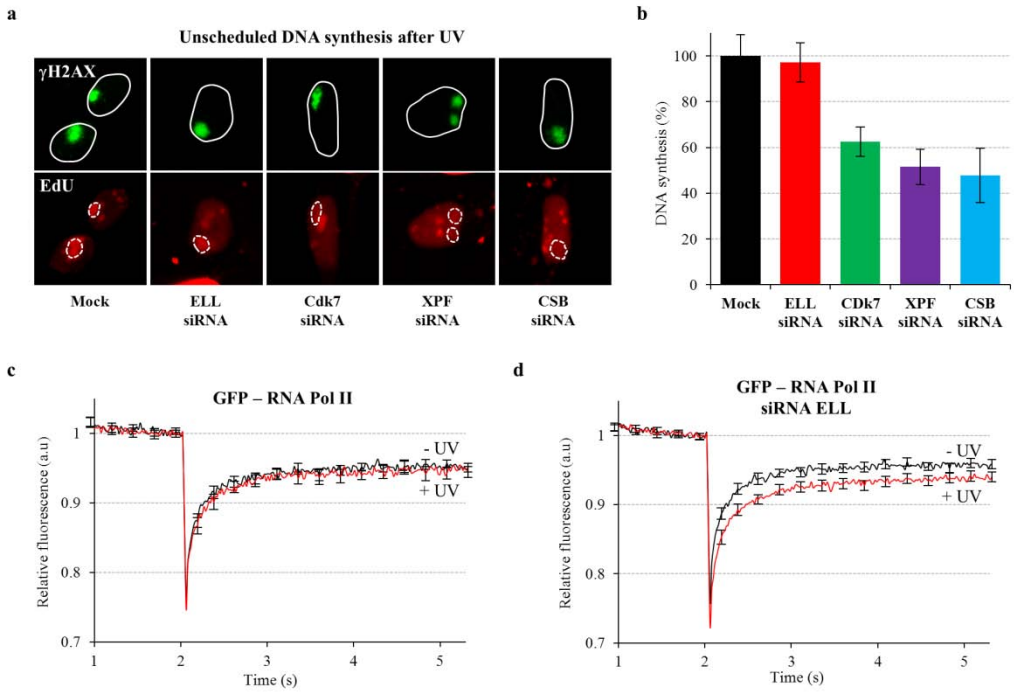


Figure 4 : ELL is essential in resumption of transcription during TCR. a*, Confocal images of cells scored for localised TCR-UDS. In green, γH2AX clearly indicates the partially UV-exposed cell nuclei (large white contours). In red, the EdU signal is measured within the identified local UV-exposed regions (small white contours). Images were captured using the same parameters for all samples.* ***b****, Quantified TCR-UDS from a GGR deficient cell line treated with siRNA against the indicated factors after local UV irradiation. The central tendency for each data set is robustly represented by its median (relative to the mock). 100 cells were imaged for each case except for siCSB treated cells for which 50 cells were taken. Error bars represent the standard error of the median.* ***c****, FRAP data of RNA Pol II-GFP expressing cells untreated (black line) or treated (red line) with UV-C.* ***d****, FRAP data of RNA Pol II-GFP expressing cells knocked down for ELL, then untreated (black line) or treated (red line) with UV-C. Error bars represent the SEM obtained from at least 30 independent frap curves.*

We believe the discovery of ELL, as both a TFIIH partner and a specific and exclusive protein implicated in transcription resumption, represents an essential advance in the knowledge of the still unclear transcription restart after DNA repair. Furthermore, finding a direct interaction between a transcription initiation complex (TFIIH) and an elongation factor (ELL) could suggest that in certain cellular circumstances these two processes could share the same protein machineries and a crosstalk should be expected between factors[29]. On the contrary, in the context of the TCR, our results clearly show that repair can be uncoupled from the transcription restart mechanism, although communicating probably via the TFIIH complex. This result adds a new layer of complexity in the knowledge of the crosstalk between repair and transcription processes. Last but not least, on the basis of this newly discovered ELL function, we propose that leukemia cells, in which ELL has been found as a translocation partner of mixed lineage leukemia (MLL) gene[8,30-33], would present a defect in transcription resumption after DNA repair. This assumption could lead to new therapeutic strategies for this kind of leukemia's which are particularly relapsing and require aggressive treatment.

METHODS SUMMARY

Immunoprecipitation and quantitative mass spectrometry TFIIH was immunoprecipitated from XPB-YFP expressing ES cells using GFP-trap antibodies. Protein samples (immunopurified complex and control samples obtained from wild-type ES cells) were digested with trypsin and the resulting peptides were analyzed by liquid nanochromatography using an Ultimate 3000 device (Dionex) coupled to a LTQ-Orbitrap Velos mass-spectrometer (Thermo Fisher Scientific). Proteins were identified by database search using the Mascot search engine, and quantified using the label-free module implemented in the MFPaQ v4.0.0 software.

Functional investigation in living cells. Stable expression of ELL-GFP, Cdk7-GFP and XPB-GFP fusion proteins was obtained in immortalized human fibroblasts. All the microscopy (FRAP, 2D and time-lapse imaging) was executed on an inverted LSM 710 NLO confocal system (Zeiss) with a 40x/1.3 oil objective. For FRAP experiments, transcription inhibition was obtained by treating cells with 150 µg/ml of 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole for 6h at 37°C. To create DNA damage (20 minutes prior to the FRAP measurements) cells were exposed to 16 J/m^2 of UV-C light (20 J/m^2 for MRC5 cells transiently expressing

RNA Pol II-GFP). Recruitment of GFP-tagged proteins was achieved by inducing DNA damage with a pulsed near infrared laser at 800 nm (Cameleon Vision II, Coherent). For cell survival, recovery of RNA synthesis (RRS), Unscheduled DNA synthesis (UDS) assays and TCR-UDS assays, cells were transfected with the indicated siRNA. mRNA was analyzed using RT-qPCR with the SyberGreen Gene expression assay, using a 7300 Real time PCR machine (Applied Biosystems). ELL, Cdk7, XPF and CSB expression levels were normalized to HPRT expression.

1 Lehmann, A. R. DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie* **85**, 1101-1111 (2003).
2 Hanawalt, P. C. Subpathways of nucleotide excision repair and their regulation. *Oncogene* **21**, 8949-8956 (2002).
3 Coin, F. *et al.* Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Mol Cell* **31**, 9-20 (2008).
4 Shilatifard, A., Lane, W. S., Jackson, K. W., Conaway, R. C. & Conaway, J. W. An RNA polymerase II elongation factor encoded by the human ELL gene. *Science* **271**, 1873-1876 (1996).
5 Giglia-Mari, G. *et al.* Differentiation driven changes in the dynamic organization of Basal transcription initiation. *PLoS Biol* **7**, e1000220, doi:10.1371/journal.pbio.1000220 (2009).
6 Gautier, V. *et al.* Label-free quantification and shotgun analysis of complex proteomes by 1D SDS-PAGE/nanoLC-MS: evaluation for the large-scale analysis of inflammatory human endothelial cells. *Molecular & cellular proteomics : MCP*, doi:10.1074/mcp.M111.015230 (2012).
7 Ito, S. *et al.* XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients. *Mol Cell* **26**, 231-243, doi:10.1016/j.molcel.2007.03.013 (2007).
8 Lin, C. *et al.* AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol Cell* **37**, 429-437, doi:10.1016/j.molcel.2010.01.026 (2010).
9 Mohan, M., Lin, C., Guest, E. & Shilatifard, A. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nat Rev Cancer* **10**, 721-728.
10 Smith, E. R. *et al.* The little elongation complex regulates small nuclear RNA transcription. *Mol Cell* **44**, 954-965, doi:10.1016/j.molcel.2011.12.008 (2011).
11 Takahashi, H. *et al.* Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146**, 92-104, doi:10.1016/j.cell.2011.06.005 (2011).
12 Rossignol, M., Kolb-Cheynel, I. & Egly, J. M. Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH. *The EMBO journal* **16**, 1628-1637, doi:10.1093/emboj/16.7.1628 (1997).
13 Nigg, E. A. Cyclin-dependent kinase 7: at the cross-roads of transcription, DNA repair and cell cycle control? *Current opinion in cell biology* **8**, 312-317 (1996).
14 Giglia-Mari, G. *et al.* Dynamic interaction of TTDA with TFIIH is stabilized by nucleotide excision repair in living cells. *PLoS Biol* **4**, e156, doi:10.1371/journal.pbio.0040156 (2006).
15 Hoogstraten, D. *et al.* Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. *Mol Cell* **10**, 1163-1174 (2002).
16 Yankulov, K., Yamashita, K., Roy, R., Egly, J. M. & Bentley, D. L. The transcriptional elongation inhibitor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription factor IIH-associated protein kinase. *J Biol Chem* **270**, 23922-23925 (1995).
17 Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2**, 103-108, doi:10.4161/trns.2.3.16172 (2011).
18 Svejstrup, J. Q. *et al.* Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. *Cell* **80**, 21-28 (1995).
19 Mu, D., Hsu, D. S. & Sancar, A. Reaction mechanism of human DNA repair excision nuclease. *J Biol Chem* **271**, 8285-8294 (1996).
20 Limsirichaikul, S. *et al.* A rapid non-radioactive technique for measurement of repair synthesis in primary human fibroblasts by incorporation of ethynyl deoxyuridine (EdU). *Nucleic Acids Res* **37**, e31, doi:gkp023 [pii]10.1093/nar/gkp023 (2009).
21 Matsumoto, M. *et al.* Perturbed gap-filling synthesis in nucleotide excision repair causes histone H2AX phosphorylation in human quiescent cells. *J Cell Sci* **120**, 1104-1112, doi:10.1242/jcs.03391 (2007).
22 Vrouwe, M. G., Pines, A., Overmeer, R. M., Hanada, K. & Mullenders, L. H. UV-induced photolesions elicit ATR-kinase-dependent signaling in non-cycling cells through nucleotide excision repair-dependent and -independent pathways. *J Cell Sci* **124**, 435-446, doi:10.1242/jcs.075325 (2011).
23 Godon, C. *et al.* Generation of DNA single-strand displacement by compromised nucleotide excision repair. *The EMBO journal*, doi:10.1038/emboj.2012.193 (2012).

24 Ariza, R. R., Keyse, S. M., Moggs, J. G. & Wood, R. D. Reversible protein phosphorylation modulates nucleotide excision repair of damaged DNA by human cell extracts. *Nucleic Acids Res* **24**, 433-440 (1996).
25 Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature reviews. Molecular cell biology* **9**, 958-970, doi:10.1038/nrm2549 (2008).
26 Fousteri, M. & Mullenders, L. H. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell research* **18**, 73-84, doi:10.1038/cr.2008.6 (2008).
27 Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249-253, doi:10.1038/nature09785 (2011).
28 Byun, J. S. *et al.* ELL facilitates RNA polymerase II pause site entry and release. *Nature communications* **3**, 633, doi:10.1038/ncomms1652 (2012).
29 Larochelle, S. *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature structural & molecular biology*, doi:10.1038/nsmb.2399 (2012).
30 Mueller, D. *et al.* Misguided transcriptional elongation causes mixed lineage leukemia. *PLoS Biol* **7**, e1000249, doi:10.1371/journal.pbio.1000249 (2009).
31 Bitoun, E., Oliver, P. L. & Davies, K. E. The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Human molecular genetics* **16**, 92-106, doi:10.1093/hmg/ddl444 (2007).
32 He, N. *et al.* HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Mol Cell* **38**, 428-438, doi:10.1016/j.molcel.2010.04.013 (2010).
33 Sobhian, B. *et al.* HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Mol Cell* **38**, 439-451, doi:10.1016/j.molcel.2010.04.012 (2010).

Partie II. Développement de méthodes pour l'étude quantitative sans marquage de protéomes

Une seconde partie de ce travail de thèse a été dédiée au développement de stratégies de quantification sans marquage pour étudier à grande échelle les mécanismes moléculaires au sein des cellules endothéliales humaines. Pour ces études, des cellules endothéliales humaines primaires de type HUVEC (« Human Umbilical Vein Endothelial Cells »), qui constituent un modèle classique, ont été utilisées. Pour évaluer les performances des méthodes mises en place, nous avons choisi de placer ces cellules dans un contexte de réponse inflammatoire forte. Nous les avons pour cela stimulées avec des cytokines pro-inflammatoires, le tumor necrosis factor alpha (TNFα), l'interféron gamma (IFNγ), ou l'interleukine-1β (IL-1β), qui déclenchent une réaction inflammatoire et immunologique, et analysé les variations protéiques induites. Nous avons vu dans la partie introductive que le principal défi dans l'analyse de mélanges protéiques complexes consiste à obtenir une couverture de protéome la plus large possible pour caractériser au mieux les processus biologiques étudiés. Pour y parvenir et faciliter la détection des protéines minoritaires, il est nécessaire de réduire la complexité des échantillons biologiques étudiés. Dans cette étude, nous avons dans un premier temps ciblé un sous-protéome particulier, le protéome de surface. Nous avons pour cela utilisé une méthode qui associe un enrichissement spécifique du glycoprotéome à une quantification sans marquage, permettant d'observer les variations des protéines de surface des cellules endothéliales induites par des conditions inflammatoires. Dans un second temps, nous nous sommes intéressés au protéome entier de ces cellules en conditions inflammatoires. Afin de caractériser en profondeur ce protéome total, nous avons mis en place une approche de quantification sans marquage associée à un fractionnement de l'échantillon au niveau protéique.

I. Développement de stratégies de protéomique quantitative pour l'étude de protéomes de surface

I-1. Contexte général : méthodes pour l'analyse du protéome de surface

I-1.1 Méthodes d'enrichissement des protéines membranaires

Le sous-protéome de la surface cellulaire est probablement l'un de ceux qui suscitent le plus d'intérêt, dans la mesure où il contient de nombreux composants impliqués dans des processus biologiques fondamentaux tels que la transduction du signal, le transport de molécules, le trafic membranaire, la migration des cellules et les interactions intercellulaires. De plus, ces protéines de surface sont directement accessibles aux médicaments et représentent des cibles pharmaceutiques

intéressantes. Cependant, le protéome de la membrane plasmique est aussi probablement l'un des plus difficiles à analyser, puisqu'il est composé de beaucoup de protéines très hydrophobes qui sont difficiles à solubiliser et à purifier, et qui représentent par ailleurs seulement une fraction très mineure du contenu protéique total de la cellule. L'étude du rôle et de la régulation des protéines de surface dans les processus physio-pathologiques nécessite donc le développement d'approches spécifiques pour enrichir ces protéines membranaires.

Ces approches reposent souvent sur le fractionnement subcellulaire, c'est-à-dire sur la préparation de membranes plasmiques (Wu, MacCoss et al. 2003; Nielsen, Olsen et al. 2005), mais les protocoles classiques ne permettent généralement pas d'éviter la contamination des préparations par des membranes d'autres organelles internes de la cellule, ni de s'affranchir des protéines semi-membranaires associées, largement majoritaires. Une autre technique conceptuellement plus intéressante repose sur le marquage des protéines localisées à la membrane plasmique, permettant ainsi leur détection et leur enrichissement spécifique. Différentes approches peuvent être utilisées pour cela.

Une méthode classique consiste à biotinyler les protéines localisées à la membrane plasmique à l'aide de sondes chimiques ciblant les fonctions amines présentes dans les chaînes latérales des lysines et en N-terminal des protéines, grâce à un groupement réactif comme un N-hydroxysuccinimide. Cependant, en pratique cette technique reste peu sélective, peu efficace sur de nombreuses protéines membranaires qui restent non détectées, et elle n'a pas permis la caractérisation d'un protéome de surface à grande échelle (Shin, Wang et al. 2003; Elia 2008).

D'autres stratégies consistent à cibler spécifiquement la chaîne glycosidique portée par la grande majorité des protéines de la membrane plasmique (Gahmberg and Tolvanen 1996). Le groupe de R. Aebersold a ainsi développé une méthode chimique de biotinylation des glycoprotéines, basée sur une oxydation au periodate des carbohydrates pour générer des fonctions aldéhydes capables de réagir avec un dérivé biotine-hydrazide, appelée méthode CSC (« Cell surface capture ») (Schiess, Mueller et al. 2008; Wollscheid, Bausch-Fluck et al. 2009). Cette approche permet de marquer et de purifier de façon très spécifique de nombreux marqueurs de surface cellulaire, mais il semble néanmoins que certaines glycoprotéines restent non détectées avec cette technique, par comparaison avec d'autres méthodes d'enrichissement telles qu'une purification par immunoaffinité sur colonne lectine (McDonald, Yang et al. 2009). De plus, elle reste générique et cible sans distinction l'ensemble des glycoprotéines de la cellule.

I-1.2 Marquage des glycoprotéines à l'aide de précurseurs azide

Parallèlement, le groupe de Carolyn Bertozzi a développé au cours de ces dix dernières années une approche particulièrement intéressante pour l'imagerie et l'enrichissement des glycoprotéines (Saxon and Bertozzi 2000). Cette approche en deux étapes repose tout d'abord sur l'incorporation métabolique dans le glycanne d'un précurseur monosaccharidique non-naturel comportant un groupement azoture (Figure 47). Ce groupement, du fait de sa petite taille, n'est pas métabolisé par les enzymes de la cellule, et présente des propriétés de bioorthogonalité, c'est-à-dire qu'il peut être très spécifiquement ciblé via différentes réactions chimiques, tout en restant inerte vis-à-vis des autres molécules de l'environnement biologique. Dans un deuxième temps, le groupement azoture va donc pouvoir réagir pour former une liaison covalente avec une sonde chimique portant un autre groupement fonctionnel bioorthogonal, comme une phosphine ou une

fonction alcyne. Ces sondes peuvent par ailleurs être fonctionnalisées avec un groupement biotine (Saxon and Bertozzi 2000), une étiquette peptidique de type FLAG (Prescher, Dube et al. 2004), ou un fluorophore (Baskin, Prescher et al. 2007), pour permettre la détection et/ou l'enrichissement des glycoprotéines marquées.

Plusieurs précurseurs métaboliques de glycannes comme la N-acétyl galactosamine (GalNAc) ou la N-acétyl mannosamine (ManNAc) sont disponibles commercialement sous forme modifiée par un groupement de type azoture, respectivement en GalNAc azido (GalNAz) et en ManNAc azido (ManNAz). Ces précurseurs azido, préalablement peracétylés (Ac4GalNAz et Ac4ManNAz) afin de faciliter leur incorporation dans la cellule, sont intégrés dans la machinerie de biosynthèse des glycannes (Figure 47). Ils sont ensuite convertis respectivement en GalNAz, un sucre qui se trouve en début de chaîne de la quasi-totalité des protéines O-glycosylées membranaires et sécrétées (Hang, Yu et al. 2003) et en acides sialiques azido (SialAz), qui se trouvent en bout de chaîne des glycannes de la majorité des glycoprotéines membranaires (Saxon and Bertozzi 2000).

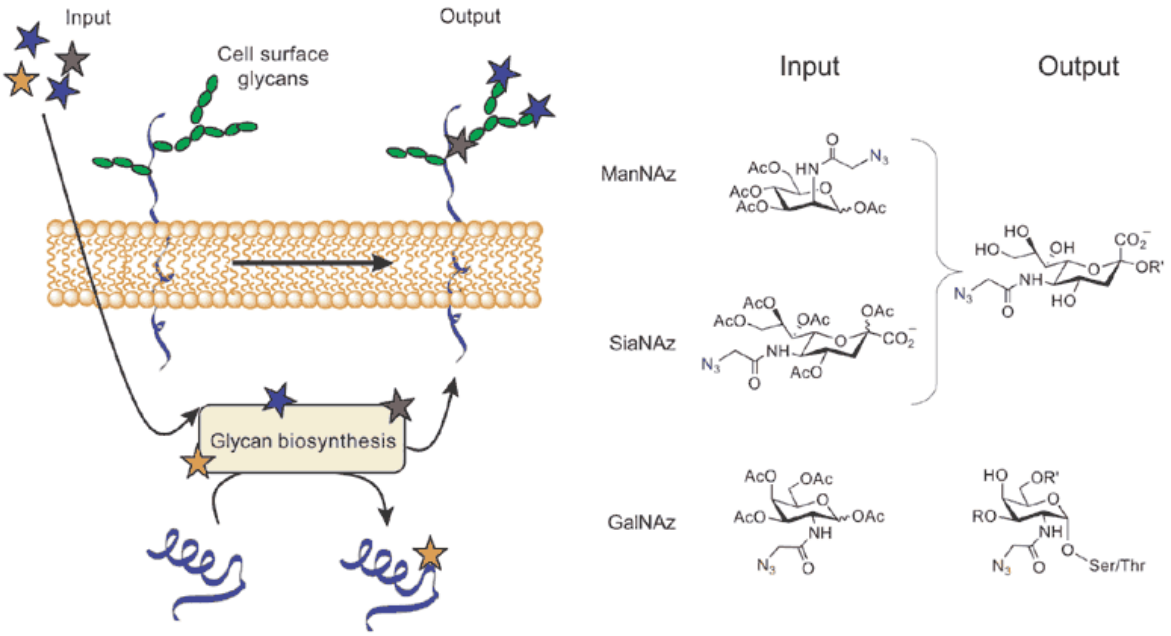


Figure 47 : Les dérivés azido de ManNAc (ManNAz) et d'acide sialique (SiaNAz) sont métabolisés dans la cellule et transformés en sialosides azido exposés à la surface de la cellule*. De même, l'analogue azido de GalNAc (GalNAz) peut être introduit par voie métabolique en position initiale des chaînes glycosidiques O-liées des glycoprotéines de type mucine.*

Différentes sondes et réactions chimiques ont été utilisées pour marquer les glycannes azido incorporés dans les protéines. Les travaux les plus aboutis sont certainement ceux qui ont été réalisés grâce à des sondes phosphines, réagissant sur le groupement azoture via une ligation de Staudinger. Cette méthode de couplage a initialement été utilisée sur des cellules en culture, et il a été montré que l'utilisation d'une sonde phosphine biotinylée, rendue soluble dans l'eau grâce à un bras de liaison de type tetraéthylèneglycol, permet un marquage particulièrement efficace des protéines sialylées présentes à la surface des cellules (Saxon and Bertozzi 2000) (Figure 48A). De même, cette approche utilisée chez la souris avec cette fois une phosphine étiquetée à l'aide d'un

peptide FLAG, a permis un marquage de ces glycoprotéines dans de nombreux organes comme le cœur, le foie ou encore la rate, sans aucune toxicité pour l'animal (Prescher, Dube et al. 2004; Dube, Prescher et al. 2006) (Figure 48B).

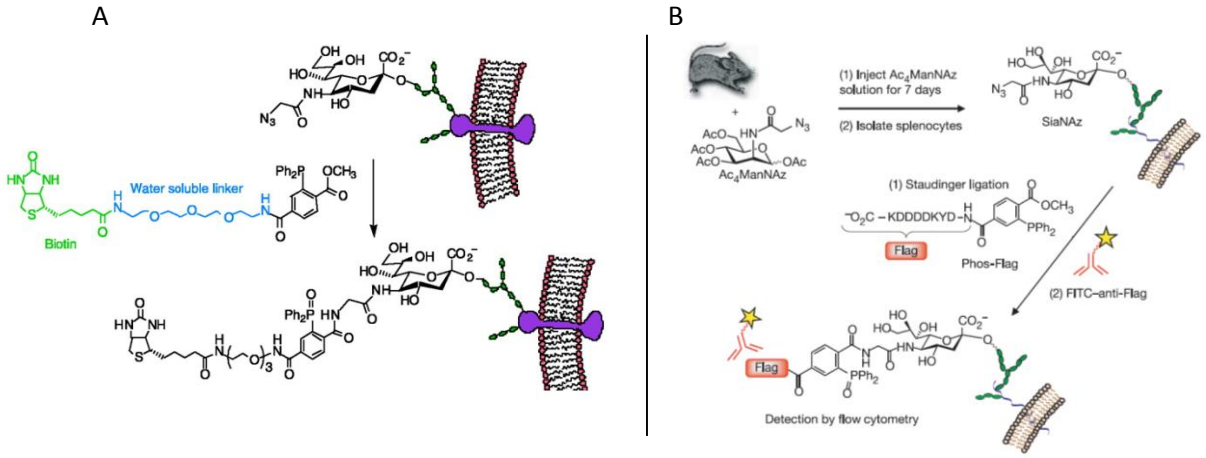


Figure 48 : Marquage de glycoprotéines par des sondes phosphine via une réaction de Staudinger. *(A) Biotinylation des glycoprotéines sialylées par une phosphine biotine sur cellules en culture (Saxon and Bertozzi 2000). (B) Marquage in vivo des glycoprotéines sialylées par une phosphine-FLAG et détection des cellules marquées à l'aide d'un anticorps fluorescent (Prescher, Dube et al. 2004).*

La stratégie de marquage des glycoprotéines par la méthode d'incorporation de sucres azido a donc largement démontré son efficacité, à la fois *in vitro* et *in vivo*. Elle a néanmoins jusqu'à présent été surtout utilisée pour détecter les protéines à l'aide de marquages fluorescents, par des méthodes de cytométrie de flux ou de microscopie, ou par western-blot. Dans notre étude sur la réponse inflammatoire des cellules endothéliales, nous avons cherché à l'adapter pour l'enrichissement et l'analyse protéomique à grande échelle du glycoprotéome de surface par spectrométrie de masse.

I-2. Analyse quantitative du glycoprotéome de surface des cellules endothéliales en conditions inflammatoires

Pour réaliser cette étude, nous avons utilisé un protocole expérimental préalablement optimisé (Delcourt et al, données non publiées). Des cellules HUVEC primaires ont été cultivées dans du milieu contenant du N-azido-acetyl-mannosamine peracetylaté (Ac4ManNAz), de façon à incorporer le sucre modifié dans les chaînes oligosaccharidiques des sialo-glycoprotéines (Figure 49). Celles-ci ont ensuite été marquées chimiquement avec de la triarylphosphine biotinylée, directement ajoutée à la suspension de cellules, afin de marquer les glycannes azido à la surface cellulaire. Les cellules ont ensuite été lysées dans un tampon détergent (1% SDS) afin de solubiliser efficacement les protéines membranaires, puis les glycoprotéines biotinylées ont été enrichies sur billes de streptavidine, en conditions dénaturantes fortes (0,2 % SDS, NaCl 3M, Urée 6M), afin de réduire la fixation non-spécifique sur les billes et obtenir une bonne sélectivité pour les protéines biotinylées. Après lavage, les protéines fixées sur les billes de streptavidine ont été digérées à la trypsine directement sur billes, et le mélange de peptides trypsiques correspondant a été analysé par nanoLC-

MS/MS pour identifier et quantifier les protéines enrichies. A la suite de cette étape de digestion, seuls les peptides porteurs des chaînes glycannes restent en principe capturés sur les billes streptavidine. Pour caractériser les sites de N-glycosylation sur la population de glycoprotéines enrichies, les billes ont donc été traitées à la peptide N-glycosidase F (PNGase F), afin de cliver spécifiquement la liaison amide entre le glycanne biotinylé et le résidu asparagine (Asn) du peptide, et libérer ainsi les N-glycopeptides. Ceux-ci ont été soumis à une analyse nanoLC-MS/MS séparée, en spécifiant lors de la recherche en base de données la modification possible de l'Asn en acide aspartique (Asp) causée par la déglycosylation à la PNGaseF, caractéristique des sites de glycosylation.

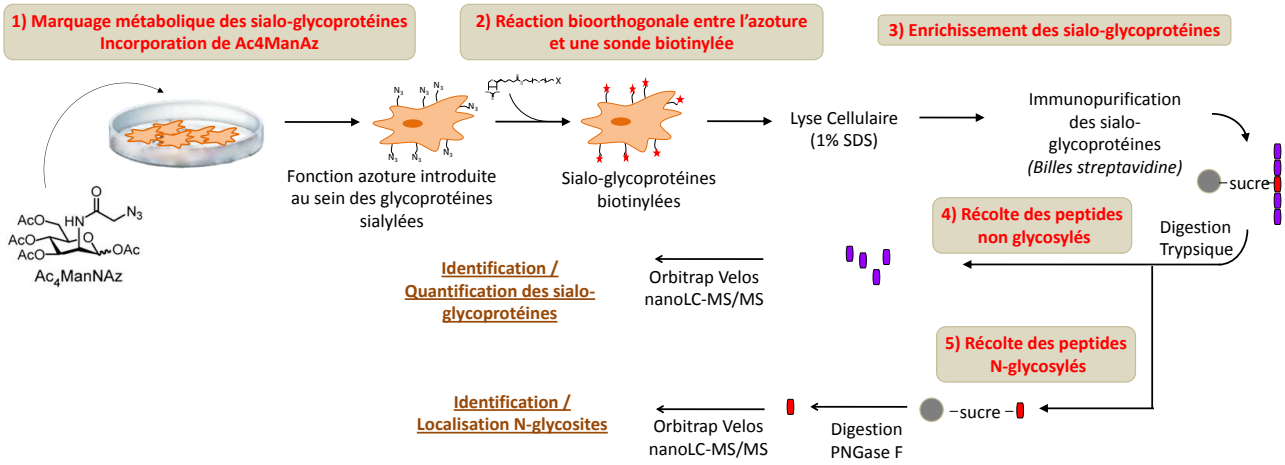


Figure 49: Stratégie d'analyse protéomique du glycoprotéome de surface

I-2.1 Cartographie des N-glycopeptides issus des glycoprotéines sialylées des cellules HUVEC

Pour évaluer la technique, nous avons tout d'abord analysé le mélange issu de cette digestion à la PNGase, afin de vérifier si la méthode permettait de capturer efficacement les N-glycoprotéines membranaires. Trois expériences indépendantes ont été réalisées sur des cellules HUVEC stimulées ou non avec un mélange TNFα/INFγ. Sur la première expérience, la purification streptavidine a été effectuée à partir de 1 mg de lysat cellulaire, les deux suivantes ont été effectuées à plus grande échelle à partir de 4,5mg de lysat. Après analyse des peptides libérés par clivage PNGase, nous avons pu identifier respectivement 319, 733 et 690 N-glycopeptides, après validation à 5% de FDR (séquences peptidiques uniques caractérisées par la présence d'au moins un résidu asparagine déamidé au sein d'un motif consensus de type NX(S/T). Au total, 889 N-glycopeptides ont donc pu être caractérisés à partir des cellules HUVEC, correspondant à 1106 sites de N-glycosylation indépendants, et à 362 glycoprotéines non-redondantes (Annexe 3). Cette analyse constitue donc la caractérisation la plus extensive à ce jour du N-glycoprotéome de cellules endothéliales humaines. D'un point de vue méthodologique, la technique basée sur le marquage métabolique et la biotinylation de glycannes-azidos apparaît comme une technique intéressante pour l'analyse des protéines de surface. Bien qu'une étude comparative plus poussée devrait être réalisée pour établir ses performances par rapport à d'autre types d'approches, elle semble permettre d'atteindre une profondeur d'analyse du même ordre que la méthode CSC basée sur l'oxydation des carbohydrates et leur marquage via des sondes hydrazide, dont l'application a été décrite sur d'autre types

cellulaires et permet typiquement d'identifier plusieurs centaines de glycopeptides (Wollscheid, Bausch-Fluck et al. 2009).

Parmi les 362 N-glycoprotéines identifiées dans cette étude, 148 (41%) ont été identifiées à partir d'un N-glycopeptide unique, alors que 214 (58%) ont été identifiées grâce à 2 N-glycopeptides (23%) ou plus (36%) (Figure 50). Les protéines les plus fortement glycosylées sont par exemple la multimérine-1 (MMRN1), l'intégrine beta1 (ITGB1/CD29) ou la fibronectine (FN1) qui sont de grosses protéines impliquées dans les processus d'adhésion cellulaire, connues pour être fortement N-glycosylées.

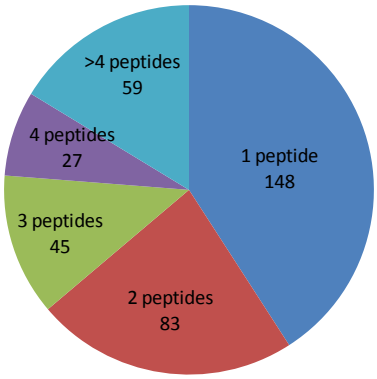


Figure 50 : Représentation schématique de la proportion de N-glycoprotéines identifiées avec 1,2,3,4 ou plus de 4 N-glycopeptides*. Les nombres de N-glycoprotéines identifiées dans chacun des cas sont indiqués.*

Pour vérifier la localisation membranaire des protéines identifiées, le logiciel Protein Center a été utilisé pour classifier celles-ci en fonction de leurs annotations GO de compartiment subcellulaire ou de fonction moléculaire, et comparer la représentation de certaines catégories par rapport à celles d'un protéome total, caractérisé indépendamment à partir d'un lysat brut de cellules HUVEC (Figure 51). Comme attendu, cette analyse montre un net enrichissement de la catégorie «membrane » dans le N-glycoprotéome par rapport au protéome total (81 % vs. 38 %), mais plus clairement encore des protéines annotées comme intrinsèques à la membrane (69 % vs. 16%). Ces chiffres correspondent à des pourcentages rapportés au nombre total de protéines identifiées (362 N-glycoprotéines versus 1068 protéines identifiées par analyse directe d'un lysat total). Cependant, même en nombres absolus de protéines, le nombre de protéines intrinsèques à la membrane et à la membrane plasmique est supérieur dans le N-glycoprotéome, indiquant que la méthode est réellement efficace pour identifier des protéines de surface non détectables à partir d'un lysat total. C'est le cas par exemple de protéines très hydrophobes multi-passages transmembranaires, comme la protéine Piezo-1, qui contient 30 domaines transmembranaires prédits et a été identifiée à partir de 3 séquences peptidiques glycosylées, mais aussi de très nombreux récepteurs de surface comme le récepteur de l'EGF, de l'IFNγ, de l'IFNα/β, de l'interleukine 18 ou du TGFβ, qui ne sont pas détectés dans une analyse « shotgun » classique sur un lysat de cellules HUVEC. Effectivement, l'analyse via Protein Center des protéines en termes de fonctions moléculaires montre que des catégories typiquement associées à des protéines de surface sont également fortement enrichies, comme les termes « receptor activity » (18% vs 1.5%), « signal transducer activity » (16% vs 3%), ou « transporter activity » (11% vs 6%). Les nombres absolus de protéines identifiées notamment dans

les catégories associées à des récepteurs confirment que la technique utilisée permet de mettre à jour de nombreux récepteurs de surface peu abondants et très difficilement détectables via une analyse directe.

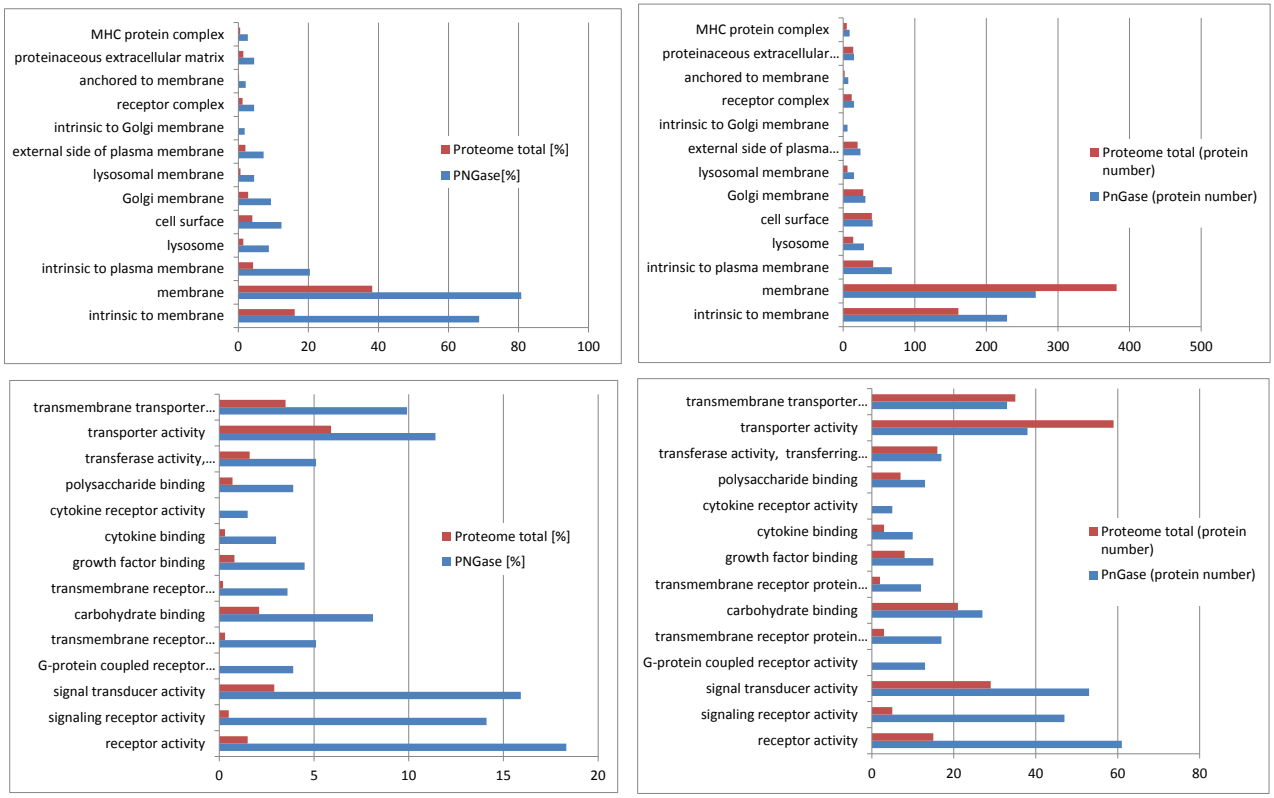


Figure 51 : Classification des N-glycoprotéines identifiées à partir des N-glycopeptides en fonction de leurs annotations GO par Protein Center*, correspondant au compartiment subcellulaire (en haut) ou à leur fonction moléculaire (en bas). Cette classification est représentée en termes de pourcentage de protéines identifiées par rapport à la totalité des protéines de l'échantillon (à gauche) ou en termes de nombre absolu de protéines identifiées (à droite). La classification selon les mêmes termes GO a également été réalisée en parallèle pour les protéines identifiées dans un lysat total de cellules HUVEC.*

I-2.2 Analyse quantitative des glycoprotéines régulées pendant l'inflammation

Dans le protocole que nous avons utilisé, l'enrichissement sur billes de streptavidine est réalisé sur les glycoprotéines entières, et non sur les glycopeptides biotinylés (comme c'est le cas par exemple dans la méthode CSC, où l'enrichissement est réalisé après la digestion trypsique). Une fois les glycoprotéines piégées par la streptavidine, la digestion trypsique directe sur billes permet de générer, pour chaque protéine, un ensemble de peptides contenus dans des domaines non-glycosylés, qui ne sont pas directement liés à la streptavidine. Cette approche permet donc en principe d'identifier et de quantifier les protéines enrichies à partir de plusieurs peptides trypsiques (et pas seulement à partir d'un nombre limité de peptides N-glycosylés). De plus, dans une approche de purification basée sur une interaction strepatvidine-glycopeptide, la libération finale des peptides d'intérêt ne peut se faire efficacement que par une digestion à la PNGaseF, et est donc

spécifiquement biaisée vers la détection des N-glycopeptides. En revanche, la digestion trypsique directe des protéines piégées sur billes de streptavidine permet théoriquement d'identifier aussi des protéines portant des glycannes O-liés sur sérine ou thréonine, ou des N-glycannes non clivables par la PNGase F.

Dans notre étude, nous avons donc cherché à caractériser la réponse inflammatoire des HUVEC en nous basant sur l'étude quantitative des protéines identifiées à partir du produit de la digestion trypsique des glycoprotéines enrichies sur billes (et non pas seulement à partir des N-glycopeptides décrits précédemment). Bien que dans cette approche le bruit de fond soit plus important (un nombre important de protéines non glycosylées fixées non spécifiquement sur billes sont également identifiées, comme en témoigne l'analyse de mélanges contrôles obtenus à partir de cellules non marquées métaboliquement), elle est en principe plus exhaustive et devrait permettre de caractériser l'ensemble des glycoprotéines enrichies sur la base de plusieurs peptides. Pour identifier les glycoprotéines de surface modulées en conditions inflammatoires, les cellules HUVEC ont été stimulées pendant 12h avec le mélange TNFα/ IFNγ, induisant une réponse pro-inflammatoire forte. Les peptides résultants de la digestion trypsique ont été analysés par nanoLC-MS/MS, et quantifiés par extraction de XIC avec le logiciel MFPaQ. Des triplicats techniques ont été réalisés sur les mélanges stimulés et non-stimulés, conduisant à l'identification et à la quantification de plus de 1000 protéines. L'analyse statistique a été effectuée sur les valeurs de PAI calculées pour les protéines après normalisation dans MFPaQ, et les protéines ont été définies comme variantes si elles présentaient un ratio d'expression supérieur à 2 et une p-value<0.05 (test de Student). Sur la base de ces valeurs seuil, près d'une cinquantaine de protéines ont été identifiées comme sur-exprimées en réponse à la stimulation pro-inflammatoire sur deux expériences indépendantes (Annexe 4). Parmi les protéines les plus nettement induites, on trouve toutes les protéines de surface connues pour leur rôle dans la reconnaissance et le recrutement des lymphocytes, notamment ICAM-1, VCAM-1 et la sélectine E, trois molécules d'adhésion jouant un rôle clé dans l'interaction leukocytes-endothélium lors de l'inflammation (Springer 1994). D'autres ligands de surface potentiellement impliqués dans l'adhésion et l'activation des lymphocytes sont également retrouvés, comme le ligand ICOS (ICOSLG/CD275), CD47, ou PD-L1/CD274. Plusieurs protéines jouant un rôle dans la présentation des antigènes via le complexe majeur d'histocompatibilité de classe I ou II apparaissent également surexprimées dans les cellules HUVEC suite à la stimulation (Antigen peptide transporter 1 TAP1, HLA-A, HLA-E**,** CD74)**.** Par ailleurs, les données obtenues montrent une surexpression de nombreux récepteurs ou protéines de surface, connues pour leur implication dans la réponse immunitaire (récepteur de chimiokines CXCR7, récepteur de l'interleukine-6 IL6ST) ou moins décrites dans ce contexte (P2X Purinoceptor 4 (P2XR4), Anthrax Toxin Receptor 2 (ANTRX2), Low Density Lipoprotein-receptor (LDL-R), Receptor-type tyrosine-protein phosphatase alpha (PTPRA), Neuropilin-2 (NRP2)). Des études plus poussées seraient nécessaires pour confirmer la régulation de ces protéines et valider leur implication fonctionnelle dans la réponse inflammatoire des cellules endothéliales. Cependant, les résultats de cette étude suggèrent que le protocole expérimental décrit ci-dessus permet effectivement de caractériser de façon détaillée les modulations d'un protéome de surface cellulaire.

I-3. Etude comparative de différentes sondes de biotinylation

I-3.1 Contexte : types de sondes pour le marquage des glycoprotéines

Les phosphines généralement utilisées pour le marquage des glycannes substitués par un groupement azoture s'oxydent rapidement à l'air, et leur optimisation en vue d'une meilleure solubilité ou d'une meilleure réactivité, s'est révélée délicate du point de vue de leur synthèse organique. D'autres sondes exploitant un autre type de réactivité de la fonction azoture, la cycloaddition [3 + 2], ont donc été développées. Le groupement azoture peut en effet réagir avec des alcynes linéaires en présence de cuivre, à température physiologique (chimie « click ») (Rostovtsev, Green et al. 2002). Cette voie a été exploitée pour modifier des protéines à partir de lysats cellulaires, mais le cuivre est toxique pour les cellules de mammifères vivantes, limitant son utilisation. En revanche, des composés cycliques contenant une fonction alcyne sous contrainte, de réactivité accrue par rapport à des alcynes simples (Figure 52), ont été utilisés pour réaliser une réaction de cycloaddition sans cuivre et marquer sélectivement les glycoprotéines de cellules en culture sans toxicité apparente (Agard, Prescher et al. 2004).

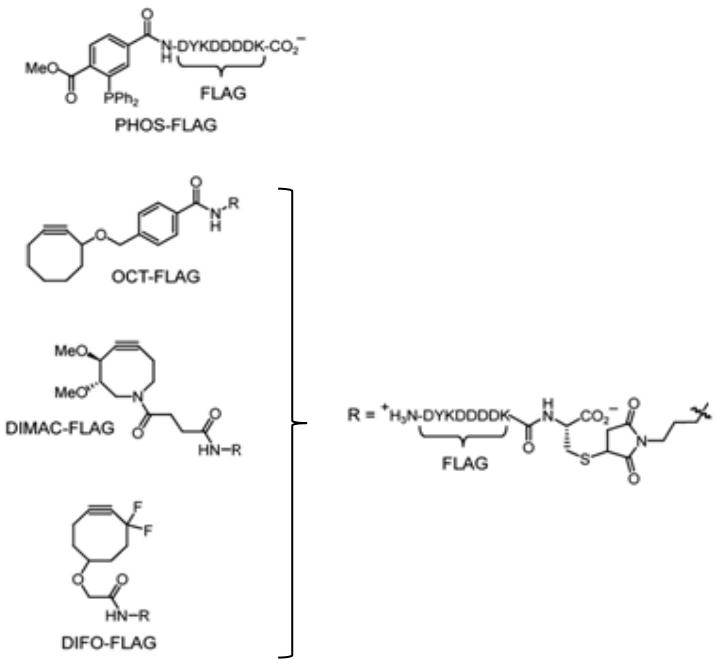


Figure 52 : Différentes sonde permettant de réaliser le marquage des sucres azido*, via une ligation de Staudinger (composé A, phosphine-FLAG), ou via une cycloaddition sans cuivre (composé B, cyclooctyne-FLAG ; C, DIMAC-FLAG ; D, DIFO-FLAG) (Chang, Prescher et al.)*

Néanmoins, ces sondes cyclooctynes sont relativement hydrophobes et assez peu solubles, ce qui réduit leur biodisponibilité pour des applications chez l'animal. Des composés modifiés, contenant un atome d'azote dans le cycle ainsi que deux groupements méthoxy pour améliorer la polarité de la molécule (sondes DIMAC), ont été synthétisés, et se sont révélés plus solubles et efficaces pour biotinyler les glycoprotéines de cellules en culture (Sletten and Bertozzi 2008). Par ailleurs, des cyclooctynes modifiées contenant deux atomes de fluor adjacents à la fonction alcyne

(sondes DIFO) ont également été décrites, et leur vitesse de réaction avec les groupements azoture est environ 50 fois supérieure à celle des cyclooctynes simples ou des phosphines, et qui permettent ainsi d'envisager des études cinétiques de la régulation du glycoprotéome. Ces sondes DIFO greffées à des fluorophores ont ainsi permis d'étudier par des techniques d'imagerie la dynamique du turn-over des glycannes dans des cellules en culture (Baskin, Prescher et al. 2007) et elles ont été également utilisées pour visualiser les glycannes *in vivo* (Laughlin, Baskin et al. 2008). Ainsi, des expériences de marquages multicolores ont été réalisées sur des O-glycannes (après incorporation d'Ac4GalNAz) produits au cours du développement chez le zebrafish et *C. elegans*, permettant d'étudier la régulation de leur synthèse ainsi que leur localisation cellulaires et tissulaires (Laughlin, Baskin et al. 2008; Laughlin and Bertozzi 2009). Une étude comparée de l'efficacité de ces différentes sondes greffées avec une étiquette peptidique FLAG, a été décrite à la fois sur cellules en culture et chez l'animal (Chang, Prescher et al.). La sonde DIFO montre une efficacité de marquage nettement supérieure sur cellules en culture, due à sa meilleure réactivité. En revanche, elle reste légèrement inférieure à la sonde DIMAC ou même à la phosphine chez l'animal, probablement du fait de sa moins bonne biodisponibilité, qui pourrait s'expliquer par la liaison non spécifique importante de la sonde avec des protéines majeures du sérum comme l'albumine.

I-3.2 Synthèse et évaluation de deux sondes chimiques pour l'analyse des glycoprotéines des cellules HUVEC

Dans le cadre de notre projet, nous avons cherché à tester l'efficacité de sondes cyclooctynes pour la purification et l'analyse protéomique des glycoprotéines de surface sur les cellules HUVEC. En collaboration avec l'équipe de chimie organique de l'Institut Curie (équipe Jean-Claude Florent « Pharmacochimie, Chimie Bioorganique, Vectorisation », UMR 176, Paris) une nouvelle sonde de type DIFO, greffée à de la biotine via un bras polyéthylèneglycol, a été synthétisée, ainsi qu'une sonde classique de type phopsphine-biotine (Figure 53). Les deux composés ont été testés en parallèle pour le marquage et l'enrichissement du glycoprotéome membranaire de cellules HUVEC, après incorporation en culture de N-azido-acetyl-mannosamine peracetylé (Ac4ManNAz).

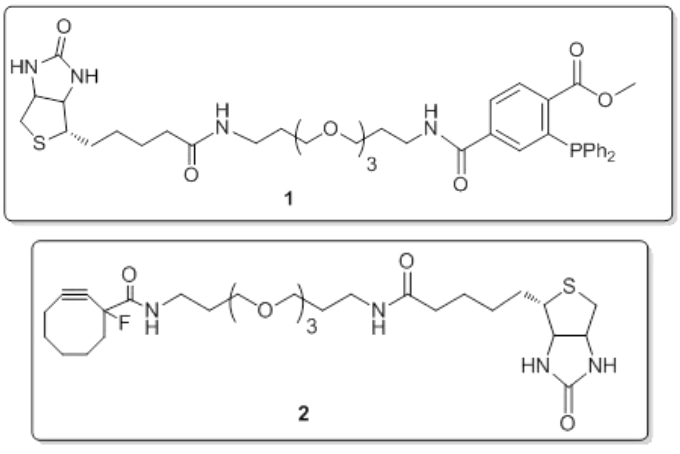


Figure 53: Sondes chimiques synthétisées pour la biotinylation des glycoprotéines de surface *préalablement marquées par incorporation métabolique de sucres comportant un groupement azido. (1) triarylphosphine-biotine, rendement global = 22% sur 4 étapes, (2) monofluorocyclooctyne-biotine, rendement global = 9% sur 5 étapes.*

a. *Comparaison de la dose-réponse et de la cinétique de marquage des deux sondes*

Dans un premier temps, une étude comparative de dose-réponse et de cinétique de marquage a été réalisée entre les deux sondes, en suivant le rendement de marquage à l'aide de western-blot basés sur la détection de protéines biotinylées par de la streptavidine-HRP. Pour chaque composé, un marquage efficace des sialo-glycoprotéines a été observé avec 250 µM de sonde (Figure 54).

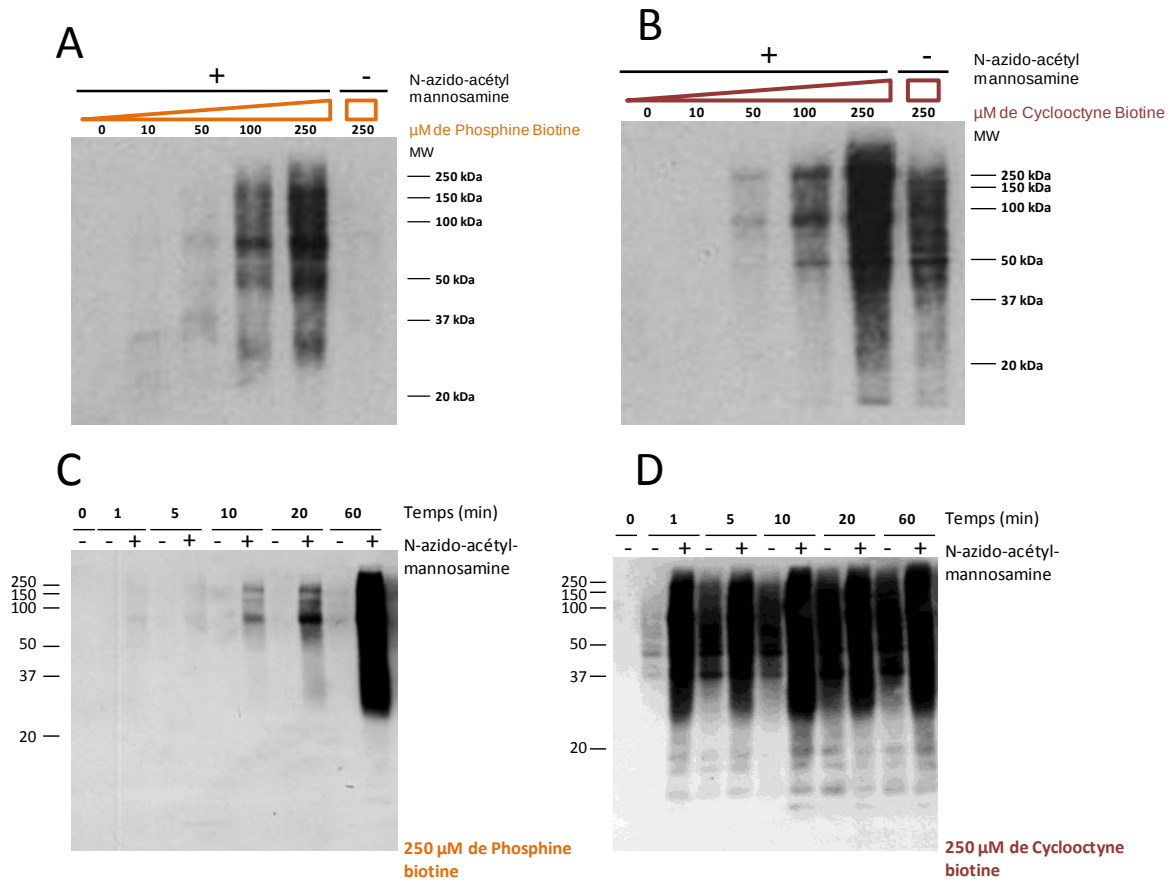


***Figure 54: Etudes de dose-réponse et cinétiques de marquage des deux sondes chimiques**. L'efficacité de marquage du protéome de surface des cellules endothéliales a été évaluée en fonction de la concentration de sonde de type triarylphosphine-biotine (A) ou monofluorocyclooctyne-biotine (B). Un marquage métabolique des glycoprotéines a été réalisé par ajout de N-azido-acetyl-mannosamine lors de la culture cellulaire, et les cellules ont ensuite été mises en présence de quantités croissantes de sondes pendant 1h. Le glycoprotéome biotinylé a été détecté par western-blot à l'aide de streptavidine-HRP. La vitesse de réaction des sondes a également été évaluée en suivant la cinétique de biotinylation des glycoprotéines pour une concentration de 250µM de triarylphosphine-biotine (C) ou monofluorocyclooctyne-biotine (D).*

A cette concentration, il semble que la cyclooctyne permet d'obtenir un meilleur rendement de marquage après 1h de réaction, au vu de l'intensité de signal détecté sur western-blot. Cependant, elle produit également du marquage non spécifique sur des contrôles réalisés sans incorporation de sucres modifiés, alors que la phosphine est réellement bioorthogonale et semble réagir très spécifiquement sur les glycoprotéines ayant incorporé les glycannes azido. Un suivi cinétique de la réaction de marquage des cellules a par ailleurs été réalisé avec les deux sondes, pour

une concentration fixe de 250 µM de chacun des composés. Comme le montre la Figure 54, la cyclooctyne réagit très rapidement, puisque dès 1min de réaction, un signal de biotinylation très intense peut être observé sur blot. Par comparaison, la phosphine est beaucoup plus lente et nécessite 1h de réaction pour obtenir un signal optimal. Cependant, là encore elle se révèle très spécifique, dans la mesure où très peu de signal apparaît pour les échantillons non marqués métaboliquement. Dans le cas de la cyclooctyne en revanche, un signal non spécifique significatif est encore une fois observé pour tous les temps de la cinétique, excepté à 1min de réaction où il reste modéré.

b. Comparaison des deux sondes pour l'analyse quantitative du glycoprotéome de surface en conditions inflammatoires

Au vu de ces résultats, une étude protéomique a été réalisée sur une culture de cellules à plus grande échelle, en réalisant le marquage avec la cyclooctyne à 250 µM pendant 1min sur la suspension de cellules, de façon à éviter les réactions non-spécifiques. En parallèle, une étude protéomique utilisant la phosphine biotine a été réalisée à partir de la même quantité d'HUVEC (8.10^7 cellules, soit 4,5 mg de lysat), dans les conditions décrites plus haut, pour permettre une étude comparative des deux sondes. Dans les deux cas, les HUVEC ont été stimulées ou non avec le mélange des cytokines TNFα/IFNγ. Après marquage, les cellules ont été rapidement lavées puis lysées, et les protéines biotinylées ont été purifiées et analysées suivant un protocole expérimental identique à celui décrit précédemment.

Cartographie des N-glycopeptides

Les mélanges peptidiques issus de la digestion PNGase ont dans un premier temps été analysés pour ces deux marquages (Table 7). Nous avons pu identifier 607 N-glycopeptides après marquage avec la cyclooctyne, correspondant à 661 sites de N-glycosylation et à 280 glycoprotéines. Le marquage avec la phosphine biotine a quant à lui aboutit à l'identification d'un nombre un peu plus important de N-glycopeptides (690) et donc de N-glycosites (726) et de N-glycoprotéines (311). Les deux sondes permettent ainsi la caractérisation d'un nombre important et relativement comparable de glycopeptides. Sur cette expérience, les résultats semblent légèrement meilleurs pour la phosphine biotine, mais ils indiquent que globalement, les deux composés permettent de piéger efficacement les glycoprotéines. Dans les deux cas, un nombre très réduit de peptides N-glycosylés ont été identifiés pour le contrôle réalisé sans marquage des glycoprotéines avec le sucre azido (4 ou 1), ce qui montre la bonne spécificité de la méthode pour la détection des glycosylations.

Table 7 : Comparaison de l'efficacité des deux sondes phosphine biotine et cyclooctyne pour la caractérisation du N-glycoprotéome*. Les nombre de peptides N-glycosylés identifiés avec ou sans marquage azide (correspondants au nombre de séquences peptidiques uniques caractérisés par la présence d'au moins un résidu asparagine déamidé dans le motif consensus NxS/T et validés à 5% de FDR) sont indiqués, ainsi que le nombre de protéines associées et le nombre de N-glycosites identifiés.*

	Marquage Azido			Contrôle sans marquage
	Peptides N-glycosylés	Protéines N-glycosylées	Nombre de sites N-glycosylation	Peptides N-glycosylés
Cyclo-octyne	607	280	661	4
Phosphine biotine	690	311	726	1

Analyse quantitative des glycoprotéines en conditions inflammatoires

Les échantillons trypsiques issus de la digestion des protéines piégées sur billes, préparés à partir des cellules stimulées ou non aux TNFα/IFNγ, ont été ensuite analysés en triplicat en nanoLC-MS/MS puis quantifiés avec MFPaQ en suivant le protocole expérimental précédemment décrit. Le marquage cyclooctyne a permis l'identification et la quantification de 942 protéines avec plus d'un peptide, contre 1200 pour la phosphine biotine. Les protéines ont été définies comme variantes après stimulation si elles présentaient un ratio d'expression supérieur à 2 et une p-value de test de Student inférieure à 5%. Ainsi, environ 70 protéines ont été identifiées comme surexprimées dans ces conditions inflammatoires. Dans les échantillons issus des marquages avec les deux sondes, on retrouve parmi les espèces très induites, les principales protéines d'adhésion impliquées dans le recrutement des leucocytes au niveau du site inflammatoire (ICAM-1, VCAM-1 et E-sélectine), les ligands de surface ICOSLG et Fractalkine/ CX3CL1, le récepteur de chimiokines CXCR7 ou le récepteur à l'IL3 et à l'EGF. De façon intéressante, on retrouve également une induction du récepteur purinergique P2X4R, dont le rôle important en tant que médiateur de l'inflammation et de la douleur a été souligné dans plusieurs études récentes (Ulmann, Hirbec et al. 2010; Trang and Salter 2012). Après liaison avec l'ATP extracellulaire, ce récepteur semble induire un flux calcique activant la synthèse de protaglandine E2, la molécule créant la sensation de douleur via excitation des neurones sensoriels. L'expression et le rôle fonctionnel de P2X4R dans la voie de synthèse des prostaglandines ont été mis en évidence essentiellement dans les macrophages (Ulmann, Hirbec et al. 2010). Les données d'expression obtenues ici sur les HUVEC suggèrent qu'il pourrait également être induit à la surface de l'endothélium dans un contexte inflammatoire. Par ailleurs, on retrouve également surexprimées plusieurs protéines impliquées dans la présentation des antigènes via le CMH de classe I (TAP1, TAP2, tapasine, HLA-A, HLA-B, HLA-C, HLA-E, CD74). Enfin, on peut noter que ces expériences ciblées sur le sialo-protéome de surface permettent de détecter des sialomucines comme la podocalyxine ou CD34. Ces protéines exprimées à la surface de l'endothélium comportent un domaine extracellulaire N-terminal de type mucine, fortement substitué par des O-glycannes sialylés. Il a été démontré que ces sialomucines ne sont reconnues par la L-sélectine, présente à la surface des lymphocytes, que si leurs chaînes oligosaccharidiques sont décorées par des motifs particuliers, de type sialyl-Lewis-X sulfatés. Ces glycoformes reconnues par la L-sélectine sont particulièrement exprimées à la surface des cellules HEV de type « cuboïdal », assurant un recrutement accentué des lymphocytes au niveau des organes lymphoïdes ou des tissus enflammés (Rosen 1999; Hemmerich, Bistrup et al. 2001). Les données protéomiques obtenues dans cette étude indiquent que la podocalyxine est une sialomucine relativement abondante, qui a été identifiée avec de très bons scores peptidiques en ciblant le protéome de surface, mais qui peut être également mise en évidence dans une analyse du protéome total des HUVEC (cf Résultats partie II-II, Annexes 5,6). Cependant, elle n'a été caractérisée comme variante que dans les expériences basées sur un enrichissement du sialoprotéome, et non dans les expériences mesurant son niveau d'expression de base à partir d'un lysat total. Ceci suggère que le protocole expérimental basé sur le marquage et l'enrichissement des glycoprotéines permet également de mesurer des variations sur la modification post-traductionnelle glycosidique elle-même, en plus des variations d'expression. On peut noter également que le marquage avec la cycloctyne a permis spécifiquement de mettre en évidence une autre sialomucine, le marqueur de surface CD34, qui est très faiblement exprimé à la surface des HUVEC, et ne semble pas en revanche être induit ou modifié en réponse au traitement pro-inflammatoire. Parmi les autres protéines spécifiquement identifiées dans l'expérience basée sur le marquage cyclooctyne, on peut

également citer le récepteur à l'interleukine 18 (IL18R), détecté comme surexprimé suite à l'analyse quantitative.

Ces données, obtenues sur des cellules endothéliales modèle, illustrent les potentialités de la méthode pour caractériser des protéines impliquées dans des maladies inflammatoires chroniques, comme les sialomucines qui jouent un rôle clé au sein des vaisseaux HEV dans la mise en place de processus inflammatoires pathologiques (Uchimura and Rosen 2006), ou des récepteurs de cytokines faiblement abondants comme le récepteur de l'IL18 (également impliqué dans plusieurs pathologies inflammatoires (Volin and Koch 2011; Kawayama, Okamoto et al. 2012). La méthode pourrait être applicable chez l'animal, ou sur des cellules purifiées à partir de tissus de patients, et conduire à une caractérisation fine du protéome de surface de l'endothélium *in vivo*. Enfin, les résultats de notre étude comparative indiquent que les deux sondes de marquage sont performantes pour l'identification de nombreux N-glycopeptides et pour mettre en évidence les modulations du protéome de surface. La cyclooctyne présente l'avantage d'être très rapide (réaction en 1min) et serait ainsi particulièrement adaptée pour réaliser des études cinétiques du glycoprotéome de surface. De plus, elle est significativement plus stable que la phosphine-biotine et moins sensible à l'oxydation, ce qui rend son utilisation plus facile et devrait améliorer la reproductibilité des analyses.

II. Développement de stratégies de protéomique quantitative pour l'étude de protéomes entiers

II-1. Objectifs

L'étude différentielle de protéomes entiers dans différentes conditions est essentielle en biologie puisque, grâce au suivi des variations d'abondance des protéines, elle permet de caractériser la réponse cellulaire à des signaux environnementaux donnés et ainsi de comprendre le fonctionnement de systèmes biologiques. Par rapport aux méthodes ciblant un sous-protéome, comme celle présentée ci-dessus pour le glycoprotéome, l'analyse du protéome entier permet d'avoir une vision plus globale et exhaustive des processus biologiques engagés dans la cellule, et nous avons donc souhaité tester ce type de caractérisation globale pour l'étude des cellules endothéliales.

La mise en place d'une stratégie de protéomique quantitative « label-free » pour l'étude à large échelle de protéomes entiers a été initiée pour répondre à une question biologique précise, à savoir caractériser les mécanismes d'action d'une nouvelle cytokine, l'interleukine 33 (IL-33) au sein des cellules endothéliales (cf Résultats, Partie III). Nous avions pour cela besoin d'une méthode nous permettant d'analyser en profondeur le protéome de cellules primaires comme les cellules endothéliales modèles humaines de type HUVEC, difficiles à maintenir longtemps en culture et à marquer isotopiquement. Nous avons donc testé une méthode de protéomique quantitative sans marquage basée sur l'intensité des signaux MS qui inclut un fractionnement des échantillons protéiques sur gel 1D SDS-PAGE. Nous l'avons évaluée pour l'étude du protéome entier des HUVEC et validée en réalisant une analyse quantitative globale des variations protéiques induites par stimulation avec des cytokines pro-inflammatoires déjà bien caractérisées (traitement combiné TNFα et IFNγ , ou traitement IL-1β, la cytokine type de la famille IL-1 à laquelle appartient IL-33).

II-2. Evaluation de la méthode

Nous avons dans un premier temps mesuré l'impact du fractionnement sur la précision de la quantification « label-free » réalisée avec MFPaQ. Nous avons pour cela évalué les performances du processus analytique en termes de répétabilité et de nombre de protéines identifiées, avec ou sans fractionnement SDS-PAGE. Afin de corriger les erreurs expérimentales systématiques liées par exemple à des quantités légèrement différentes d'échantillon déposé sur gel, ou encore à la variabilité de la réponse MS du spectromètre de masse lors d'analyses consécutives, nous avons introduit une procédure de normalisation des données MS, réalisée automatiquement par le logiciel. Dans le cas d'une expérience réalisée avec fractionnement des protéines par gel 1D, cette opération est effectuée sur les bandes en vis-à-vis, de façon à normaliser entre elles toutes les fractions situées sur un même niveau de découpe, contenant des mélanges protéiques de composition similaire.

Par ailleurs, la migration électrophorétique en parallèle des échantillons et la découpe des bandes de gel peuvent représenter en elles-mêmes des sources importantes de variabilité (par exemple, d'une piste de migration à l'autre, une protéine peut être partitionnée différemment sur plusieurs fractions consécutives). Pour tenter de corriger ces biais, nous avons ajouté au processus

analytique un traitement bioinformatique supplémentaire permettant d'intégrer au niveau protéique les données quantitatives issues de fractions consécutives du gel.

L'ensemble de cette procédure et les résultats d'évaluation de la méthode sont détaillés dans la publication insérée à la fin de ce chapitre. En résumé, nous avons montré que l'approche quantitative basée sur l'extraction des XIC peptidiques par MFPaQ permet une quantification précise et qu'une bonne répétabilité est obtenue sur les protéines quantifiées en une acquisition unique, sans fractionnement (le CV médian des PAI déterminés pour les protéines est de 5% sur trois réplicats, et 99% des protéines ont un CV inférieur à 48%). Le fractionnement protéique par gel SDS-PAGE entraîne une diminution modérée de la répétabilité des mesures quantitatives. Nous avons tenté d'évaluer la variabilité supplémentaire liée à la migration sur gel en caractérisant le profil de migration des protéines sur les pistes parallèles, et en vérifiant, sur l'ensemble de la population, si ces profils sont alignés. Sur l'ensemble des 35425 signaux peptidiques extraits, près de 96% présentent leur apex de profil de migration dans la même bande de gel 1D sur 3 pistes réplicats ayant migré en parallèle, et pour plus de 3% d'entre eux, cet apex est n'est décalé que d'une seule fraction. Pour la très grande majorité des protéines identifiées, la migration électrophorétique et le fractionnement en bandes de gel semblent donc reproductibles. Au final, après avoir appliqué les procédures de normalisation bande à bande et d'intégration du signal (réalisée sur trois bandes adjacentes), le CV médian obtenu sur le PAI des protéines identifiées et quantifiées à partir de l'expérience en gel SDS-PAGE est de 7%, et 99% des protéines ont un CV inférieur à 62%. De plus, le fractionnement apporte une augmentation très importante de la profondeur d'analyse du protéome. Le fractionnement sur gel 1D semble ainsi compatible avec une quantification sans marquage, et permet une analyse extensive de protéomes complexes.

II-3. Application à l'analyse protéomique de la réponse inflammatoire dans les cellules endothéliales

Nous avons donc dans un second temps testé cette méthode pour étudier les variations protéiques associées à la réponse inflammatoire de cellules endothéliales humaines. Nous avons ainsi pu quantifier autour de 5000 protéines et caractériser les mécanismes biologiques impliqués dans la réponse aux cytokines pro-inflammatoires TNFα/IFNγ ou IL-1β. Globalement, l'ensemble des voies de signalisation connues et des processus biologiques attendus sont retrouvés dans cette analyse protéomique différentielle. On détecte tout d'abord la surexpression de certains composants de la voie JAK-STAT induite par l'interféron (en particulier le facteur de transcription STAT1 qui est le médiateur principal de la réponse interféron) et de la voie de signalisation induite par le récepteur du TNF, convergeant vers l'activation du facteur de transcription NF-kB. Puis, en réponse à l'activation de ces facteurs, les données quantitatives permettent de mesurer la surexpression de nombreuses protéines de l'inflammation, notamment celles impliquées dans une des fonctions majeures des cellules endothéliales, consistant à recruter les lymphocytes au niveau du site d'inflammation. Nous avons ainsi pu mesurer une forte surexpression de tout un ensemble de cytokines chemo-attractantes et de molécules de signalisation sécrétées comme l'interleukine 27 ou l'interleukine 32, mais aussi l'expression de plusieurs molécules d'adhésion de la surface cellulaire impliquée dans l'interaction avec les lymphocytes circulants (E-sélectine, ICAM1, VCAM1…), l'expression des molécules du CMH ainsi que tout un ensemble de protéines activées par l'interféron jouant un rôle dans la réponse immunitaire et antivirale. Cette étude a donc permis de disséquer au niveau

protéomique plusieurs processus bien caractérisés de la réponse des cellules endothéliales au stimulus pro-inflammatoire, ce qui valide la méthodologie analytique utilisée. Par ailleurs, certaines protéines qui n'avaient pas été décrites comme activées dans les cellules endothéliales ont pu également être mises en évidence (cf publication jointe).

En conclusion, cette étude nous a permis de mettre en place une méthode efficace pour caractériser à grande échelle les variation protéiques sur des cellules primaires comme les cellules endothéliales HUVEC, et de produire des jeux de données protéomiques de référence caractérisant une réponse inflammatoire forte de ces cellules (stimulation TNF/INF) ou la réponse à la cytokine IL-1b, représentant la molécule type de la famille IL-1. Les listes des protéines majeures surexprimées en réponse à ces deux expériences de stimulation sont données en annexe de ce manuscrit. L'obtention de ces données types est importante pour des études à venir visant à caractériser le phénotype spécifique induit par d'autres molécules stimulatrices, comme par exemple IL-33.

II-4. Article Gautier et al., MCP, 2012

« Label-free Quantification and Shotgun Analysis of Complex Proteomes by One-dimensional SDS-AGE/NanoLC-MS »

EVALUATION FOR THE LARGE SCALE ANALYSIS OF INFLAMMATORY HUMAN ENDOTHELIAL CELLS

Violette Gautier, Emmanuelle Mouton-Barbosa, David Bouyssie, Nicolas Delcourt, Mathilde Beau, Jean-Philippe Girard, Corinne Cayrol, Odile Burlet-Schiltz, Bernard Monsarrat, and Anne Gonzalez de Peredo.

Mol Cell Proteomics. 2012 Aug;11(8):527-39.

Technological Innovations and Resources

Label-free Quantification and Shotgun Analysis of Complex Proteomes by One-dimensional SDS-PAGE/NanoLC-MS

EVALUATION FOR THE LARGE SCALE ANALYSIS OF INFLAMMATORY HUMAN ENDOTHELIAL CELLS*[S]

Violette Gautier‡§¶, Emmanuelle Mouton-Barbosa§¶, David Bouyssié‡§¶, Nicolas Delcourt§, Mathilde Beau‡§, Jean-Philippe Girard‡§‖, Corinne Cayrol‡§‖, Odile Burlet-Schiltz‡§, Bernard Monsarrat‡§, and Anne Gonzalez de Peredo‡§‡‡**

To perform differential studies of complex protein mixtures, strategies for reproducible and accurate quantification are needed. Here, we evaluated a quantitative proteomic workflow based on nanoLC-MS/MS analysis on an LTQ-Orbitrap-VELOS mass spectrometer and label-free quantification using the MFPaQ software. In such label-free quantitative studies, a compromise has to be found between two requirements: repeatability of sample processing and MS measurements, allowing an accurate quantification, and high proteomic coverage of the sample, allowing quantification of minor species. The latter is generally achieved through sample fractionation, which may induce experimental bias during the label-free comparison of samples processed, and analyzed independently. In this work, we wanted to evaluate the performances of MS intensity-based label-free quantification when a complex protein sample is fractionated by one-dimensional SDS-PAGE. We first tested the efficiency of the analysis without protein fractionation and could achieve quite good quantitative repeatability in single-run analysis (median coefficient of variation of 5%, 99% proteins with coefficient of variation <48%). We show that sample fractionation by one-dimensional SDS-PAGE is associated with a moderate decrease of quantitative measurement repeatability while largely improving the depth of proteomic coverage. We then applied the method for a large scale proteomic study of the human endothelial cell response to inflammatory cytokines, such as TNFα, interferon γ, and IL1β, which allowed us to finely decipher at the proteomic level the biological pathways involved in endothelial cell response to proinflammatory cytokines. ***Molecular & Cellular Proteomics 11: 10.1074/mcp.M111.015230, 527–539, 2012.***

From ‡Centre National de la Recherche Scientifique, Institut de Pharmacologie et de Biologie Structurale, F-31077 Toulouse, France, and §Université de Toulouse, Université Paul Sabatier, Institut de Pharmacologie et de Biologie Structurale, F-31077 Toulouse, France

Received October 21, 2011, and in revised form, April 10, 2012
Published, MCP Papers in Press, April 19, 2012, DOI 10.1074/mcp.M111.015230

With recent advances in mass spectrometry, label-free quantitative proteomic approaches have progressed and are now considered as reliable and efficient methods to study protein expression level changes in complex mixtures. These approaches, which have been reviewed recently (1, 2), are based on the measurement either of the MS/MS sampling rate of a particular peptide or of its MS chromatographic peak area, these values being directly related to peptide abundance. The increase of instrument sequencing speed has benefited MS/MS spectral counting approaches by improving MS/MS sampling of peptide mixtures, whereas the introduction of high resolution analyzers such as FT-Orbitrap has boosted the use of methods based on peptide intensity measurements by greatly facilitating the matching of peptide peaks in different complex maps acquired independently. The most obvious advantage of these methods over isotopic labeling techniques is their ease of use at the sample preparation step, because they do not require any preliminary treatment to introduce a label into peptides or proteins. Being more straightforward, they also do not present the classical drawbacks of labeling methods, *i.e.*, cost, applicability to limited types of samples (mostly cultured cells in the case of metabolic labeling) and the limited number of conditions that can be compared. On the other hand, the use of label-free strategies is hampered by two main difficulties: 1) the variability of all sample processing steps before MS analysis and of the analytical measurement itself, because the samples to be compared are processed and analyzed individually, and 2) the complexity of the MS data analysis step, which requires proper realignment, normalization, and peptide peaks matching across different nanoLC-MS runs.

Many bioinformatic tools have been developed in recent years for the quantification of MS data generated in label-free experiments, either by spectral counting (3–5) or by peptide MS signal intensity measurement (6–9). In the later field, a lot of emphasis has been put on peptide pattern-based methods, in which the software performs feature detection in LC-MS maps through analysis of the characteristic isotopic pattern of a peptide ion in the *m/z* dimen-

sion, and on its chromatographic elution peak in the retention time (RT)[1] dimension. The total ion current integrated under this MS feature can then be used as a quantitative measurement of the peptide concentration. The primary advantage of this approach is that any signal detected by the mass spectrometer in the MS survey scan can be in principle analyzed and quantified, whether or not the peak has been selected for MS/MS sequencing. Bioinformatic programs based on peptide feature detection as the starting step for label-free analysis include among others SuperHirn (8), MSInspect (6), OpenMS (9), Decon2LS (7), or the commercial software Progenesis LC-MS. Although they offer an attractive and powerful analysis of the data, algorithms based on recognition of peptide features and LC-MS maps alignment require intensive computer calculation, making the quantification time-consuming and difficult to perform on a large number of LC-MS files. In addition, integration of MS features quantitative data with MS/MS identification results from search engines occurs as a second step, and depending on the bioinformatic tool used, retrieving quantitative values for the list of identified and validated peptides, and then for the associated list of proteins, can be difficult to implement. Finally, because the LC-MS maps are usually analyzed individually, low intensity features near the cut-off value set for the recognition process are detected in an irreproducible way, and most of the available software generates quantitative data sets containing many missing values, which complicates further statistical analysis of the results.

On the other hand, another approach to extract quantitative data from MS survey scans is based on the reverse process, *i.e.*, making use initially of peptide identification results to go back in the MS scans to obtain peptide intensity values. For each peptide ion identified from MS/MS sequencing, experimentally measured RT and monoisotopic *m/z* values can be used as a starting point to retrieve the associated extracted ion chromatograms (XIC) of this ion. In that case, confident extraction of a peptide signal (*versus* chemical noise) is supported by the identification result, and because the charge state of the ion is known, definition of isotopic patterns and extraction of intensity values for the different isotopes of a same peptide ion is facilitated. Such a method, which is in principle more simple and rapid, has been used in a few software packages such as Serac (10), Quoil (11), Ideal-Q (12), and MFPaQ (13, 14). A drawback of this method, however, is that only identified peptides can be quantified. For analysis of highly complex peptide mixtures, MS/MS undersampling thus limits the number of identified and quantified proteins. Depending on the software, this problem can be alleviated by a cross-assignment of peptide signals across different replicate LC-MS/MS runs: if a peptide ion is identified in only one or a few runs, its signal can be extracted in the other analytical runs by using a predicted RT value, even if identification results are missing for this particular peptide in these runs because of MS/MS undersampling. Thus, acquisition of multiple replicate runs allows to increase the number of identified and thus quantified peptides and proteins. Nevertheless, the performances of identity-based methods are still strictly linked to the number of identifications and to the depth of the proteomic analysis on highly complex samples.

In a study focusing on label-free quantitative analysis of clinical samples (14), we previously described an approach based on the use of the MFPaQ software to circumvent the undersampling problem. Following extensive proteomic analysis of cerebrospinal fluid after treatment with combinatorial libraries of peptide ligands and one-dimensional SDS-PAGE fractionation, we generated an identification database containing sequences of identified peptides, along with their *m/z* and retention time-associated values that were then used to extract the XIC of these peptides in the one-shot analytical runs of unfractionated samples. This method was well suited for the analysis of clinical series in which very limited or no fractionation at all is performed on the samples, because of the large number of analyses (number of patients and technical replicates), and we showed that it indeed allowed significant increase of the number of proteins correctly quantified in replicate runs of individual samples. However, not all of the peptides from the database could be retrieved in the individual runs, because of the limited dynamic range of the instrument during the one-shot analysis of complex peptide mixtures. To overcome also the dynamic range limitation, a commonly used and efficient approach is to prefractionate individually the samples to be compared and perform nanoLC-MS/MS analysis of each fraction separately. Although it requires longer analytical time, this shotgun type of analysis clearly offers an improved coverage of the sample and allows the detection of low abundance proteins that remain undetected when the whole sample is analyzed in one run. To that aim, one-dimensional SDS-PAGE is often selected as a robust and simple method to fractionate most kinds of protein samples, even membrane ones, and is particularly used on SILAC or ICAT labeled proteomes, because the two samples to be compared are gathered and can be processed simultaneously. However, when label-free quantification is to be performed, parallel processing steps such as electrophoretic migration, gel cutting, and in-gel digestion represent different sources of variability that may alter the final quantitative comparison of the samples.

In the present study, our objective was to perform an in-depth quantitative analysis of the endothelial cell (EC) proteome using a label-free approach. First, we thus checked whether SDS-PAGE fractionation of the individual samples, which gives the best dynamic range on a global analysis, is

[1] The abbreviations used are: RT, retention time; XIC, extracted ion chromatogram; EC, endothelial cells; IFNγ, interferon γ; HUVEC, human umbilical vein endothelial cells; FDR, false discovery rate; PAI, protein abundance index; CV, coefficient of variation; MHC, major histocompatibility complex.

compatible with accurate label-free quantitation based on peptide signal intensity measurement. We evaluated the performances of a label-free quantitative workflow in terms of repeatability and number of quantified proteins, with or without protein fractionation by one-dimensional SDS-PAGE, for the analysis of a complex cellular proteome. We applied the MFPaQ software, which uses an identity-based extraction approach, to quantify the data obtained from the nanoLC-MS/MS analysis of a total lysate of primary cultured human vascular ECs. New data normalization and integration procedure dedicated to shotgun experiments were introduced in the software, allowing integration at the protein level the quantitative data from different fractions and correction of errors related to nonreproducible electrophoretic migration of proteins. We showed that the approach based on peptide XIC extraction provides good quality quantitative data on the identified proteome and that high repeatability is obtained on proteins quantified in single run analysis (median CV of 5%, 99% proteins with CV values of $<$48%). When the protein sample is fractionated by one-dimensional SDS-PAGE, the repeatability of the quantitative measurement decreases, although in a moderate way (median CV of 7%, 99% proteins with CV values of $<$62%), and concomitantly the depth of proteomic coverage is largely increased. We then applied the method for a large scale proteomic study of the response of ECs to proinflammatory treatments with TNFα/IFNγ or IL1β. It allowed us to identify and quantify more than 5400 unique proteins, providing an in-depth analysis of the endothelial cell proteome and a detailed characterization of the proteomic variations associated with the inflammatory response.

MATERIALS AND METHODS

EC Culture and Cytokine Stimulation—Primary human umbilical vein ECs (HUVECs) were purchased from Clonetics, grown in ECGM medium (Promocell, Heidelberg, Germany), and used after four passages for proteomic analyses. Cytokine treatment was performed by incubating the ECs for 12 h in OptiMEM medium (Invitrogen) with a combination of TNFα (25 ng/ml; R & D Systems) and IFNγ (50 ng/ml; R & D Systems) or with IL1β (5 ng/ml; R & D Systems).

Protein Sample Processing—The cells were lysed in a buffer containing 2% of SDS and sonicated, and protein concentration was determined by detergent-compatible assay (DC assay; Bio-Rad). Protein samples were reduced in Laemmli buffer (final composition: 25 mM DTT, 2% SDS, 10% glycerol, 40 mM Tris, pH 6.8) for 5 min at 95 °C. Cysteine residues were alkylated by addition of iodoacetamide at a final concentration of 90 mM and incubation for 30 min at room temperature in the dark. During the alkylation reaction, the pH of the samples was adjusted using small amounts of 1 M Tris, pH 8. Protein samples were loaded on a homemade one-dimensional SDS-PAGE gel (separating gel 1.5 mm $\times$ 5 cm, 12% acrylamide polymerized in SDS 0.1%, 375 mM Tris, pH 8.8, and stacking gel 1.5 mm $\times$ 1.5 cm, 4% acrylamide polymerized in 0.1% SDS, 125 mM Tris. For one-shot analysis of the entire mixture, no fractionation was performed, and the electrophoretic migration was stopped as soon as the protein sample (15 μg) entered the separating gel. The gel was briefly stained with Coomassie Blue, and a single band, containing the whole sample, was cut. For shotgun analysis, electrophoretic migration was performed to fractionate the protein sample (100 μg) into 12 gel bands. For replicate and comparative analyses, the samples were processed on adjacent migration lanes that were cut simultaneously with a long razor blade. To evaluate gel to gel repeatability, different gels were prepared and migrated in parallel, and the same number of homogeneous gel slices were cut successively on the separate gels, following the same cutting pattern. Gel slices were washed by two cycles of incubation in 100 mM ammonium bicarbonate for 15 min at 37 °C, followed by 100 mM ammonium bicarbonate/acetonitrile (1:1) for 15 min at 37 °C. The proteins were digested by 0.6 μg of modified sequencing grade trypsin (Promega) in 50 mM ammonium bicarbonate, overnight at 37 °C. The resulting peptides were extracted from the gel by incubation in 50 mM ammonium bicarbonate for 15 min at 37 °C and twice in 10% formic acid/acetonitrile (1:1) for 15 min at 37 °C. The three collected extractions were pooled with the initial digestion supernatant, dried in a SpeedVac, and resuspended with 17 μl of 5% acetonitrile, 0.05% TFA.

NanoLC-MS/MS Analysis—The Resulting peptides were analyzed by nanoLC-MS/MS using an Ultimate3000 system (Dionex, Amsterdam, The Netherlands) coupled to an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Five μl of each sample were loaded on a C-18 precolumn (300-μm inner diameter $\times$ 5 mm; Dionex) at 20 μl/min in 5% acetonitrile, 0.05% TFA. After 5 min of desalting, the precolumn was switched online with the analytical C-18 column (75 μm inner diameter $\times$ 15 cm; PepMap C18, Dionex) equilibrated in 95% solvent A (5% acetonitrile, 0.2% formic acid) and 5% solvent B (80% acetonitrile, 0.2% formic acid). The peptides were eluted using a 5 to 50% gradient of solvent B during 80 min at 300 nl/min flow rate. The LTQ-Orbitrap Velos was operated in data-dependent acquisition mode with the XCalibur software. Survey scan MS were acquired in the Orbitrap on the 300–2000 *m/z* range with the resolution set to a value of 60,000. The 10 most intense ions per survey scan were selected for CID fragmentation, and the resulting fragments were analyzed in the linear trap (LTQ). Dynamic exclusion was employed within 60 s to prevent repetitive selection of the same peptide.

Database Search and Data Validation—The Mascot Daemon software (version 2.3.2; Matrix Science, London, UK) was used to perform database searches, using the Extract_msn.exe macro provided with Xcalibur (version 2.0 SR2; Thermo Fisher Scientific) to generate peaklists. The following parameters were set for creation of the peaklists: parent ions in the mass range 400–4500, no grouping of MS/MS scans, and threshold at 1000. A peaklist was created for each analyzed fraction (*i.e.*, gel slice), and individual Mascot (version 2.3.01) searches were performed for each fraction. The data were searched against *Homo sapiens* entries in Uniprot protein database (release 2010_09, September 21, 2010; 1,215,533 sequences). Carbamidomethylation of cysteines was set as a fixed modification, and oxidation of methionine and protein N-terminal acetylation were set as a variable modifications. Specificity of trypsin digestion was set for cleavage after Lys or Arg, and two missed trypsin cleavage sites were allowed. The mass tolerances in MS and MS/MS were set to 5 ppm and 0.6 Da, respectively, and the instrument setting was specified as "ESI-Trap." To calculate the false discovery rate (FDR), the search was performed using the "decoy" option in Mascot. Peptide identifications extracted from Mascot result files were validated at a final peptide FDR of 5%. Peptide matches were validated if their score was greater than the Mascot homology threshold (when available, otherwise the Mascot identity threshold was used) for a given Mascot *p* value. The FDR at the peptide level was calculated as described in Navarro and Vázquez (15). Using this method, the *p* value was automatically adjusted to obtain a FDR of 5% at the peptide level. Validated peptides were assembled into proteins groups following the principle of parsimony (Ocam's razor), which involves the creation of the minimal list of protein groups explaining the list of peptide spec-

trum matches. Protein groups were then rescored for the protein validation process. For each peptide match belonging to a protein group, the difference between its Mascot score and its homology threshold (or identity threshold) was computed for a given *p* value (automatically adjusted to increase the discrimination between target and decoy matches), and these "score offsets" were then summed to obtain the protein group score. Protein groups were validated based on this score to obtain a FDR of 1% at the protein level (FDR = number of validated decoy hits/(number of validated target hits + number of validated decoy hits) × 100). In the case of sample fractionation on one-dimensional SDS-PAGE, the MFPaQ software was used to create a unique nonredundant protein list from the identification results of each fraction by clustering protein groups containing sequences matching the same set of peptides. If a final group was composed of several TrEMBL and SwissProt entry names, a SwissProt entry was singled out, and the associated protein description was reported in the final lists (supplemental Tables I and II).

Data Quantification—Quantification of proteins was performed using the label-free module implemented in the MFPaQ v4.0.0 software (http://mfpaq.sourceforge.net/). For each sample, the software uses the validated identification results and XICs of the identified peptide ions in the corresponding raw nanoLC-MS files, based on their experimentally measured RT and monoisotopic *m/z* values The time value used for this process is retrieved from Mascot result files, based on an MS2 event matching to the peptide ion. If several MS2 events were matched to a given peptide ion, the software checks the intensity of each corresponding precursor peak in the previous MS survey scan. The time of the MS scan that exhibits the highest precursor ion intensity is attributed to the peptide ion and then used for XIC extraction as well as for the alignment process. Peptide ions identified in all the samples to be compared were used to build a retention time matrix to align LC-MS runs. If some peptide ions were sequenced by MS/MS and validated only in some of the samples to be compared, their XIC signal was extracted in the nanoLC-MS raw file of the other samples using a predicted RT value calculated from this alignment matrix by a linear interpolation method. Quantification of peptide ions was performed based on calculated XIC areas values. To perform normalization of a group of comparable runs, the software computed XIC area ratios for all the extracted signals between a reference run and all the other runs of the group and used the median of the ratios as a normalization factor. To perform protein relative quantification in different samples, a protein abundance index was calculated, defined as the average of XIC area values for at most three intense reference tryptic peptides identified for this protein (the three peptides exhibiting the highest intensities across the different samples were selected as reference peptides, and these same three peptides were used to compute the PAI of the protein in each sample; if only one or two peptides were identified and quantified in the case of low abundant proteins, the PAI was calculated based on their XIC area values). In the case of SDS-PAGE fractionation, integration of quantitative data across the fractions was performed as indicated in the text, by summing the PAI values for fractions adjacent to the fraction with the best PAI (the same three consecutive fractions for all the samples to be compared). For differential studies, a Student's *t* test on the PAI values was used for statistical evaluation of the significance of expression level variations. For proteins specifically detected in one condition and not in the other, the *t* test *p* value was calculated by assigning a noise background value to the missing PAI values. A 2-fold change and *p* value of 0.05 were used as combined thresholds to define biologically regulated proteins.

Quantitative PCR Experiments—Total RNA from HUVEC cells (mock treated, TNFα + IFNγ-treated, or IL1β-treated) was isolated using the Absolute RNA kit from Stratagene (Agilent Technologies, Santa Clara, CA), and cDNAs were synthesized using SuperSript III First strand cDNA synthesis system for RT-PCR (Invitrogen) according to the manufacturer's instructions. Quantitative PCR was performed using the ABI7300 Prism SDS real time PCR detection system (Applied Biosystems, Foster City, CA) with a SYBR Green PCR Master Mix kit (Applied Biosystems) and a standard temperature protocol. The results are expressed as relative quantities and calculated by the 2-ΔΔCT method. *Actin* was used as a control gene for normalization. Three separate experiments were performed. Primers used were purchased from Qiagen (QuantiTect primer assay), except Actin, GAPDH, NFKB2, ICAM1, and VCAM1 (from Sigma Genosys).

RESULTS

Analytical Workflow—A total lysate of cultured primary human vascular ECs was used for all experiments and processed in all cases through one-dimensional SDS-PAGE, as shown in Fig. 1. When the samples were to be analyzed in one analytical nanoLC-MS/MS run (no fractionation), the electrophoretic migration was stopped immediately after the protein samples entered the separating part of the gel, so that the whole sample was isolated into a unique gel band and subsequently in-gel digested. In our hands, processing in this way, the total cell lysate for tryptic digestion gave slightly better proteomic coverage than digestion in solution. For sample fractionation and shotgun analysis, migration was performed so that 12 gel bands could be cut afterward along the migration lanes. Gel cutting was performed systematically with a long razor blade to simultaneously cut all the corresponding gel bands for the different samples to be compared, perpendicularly to the migration direction. All in gel digestion steps were manually performed in parallel. The resulting tryptic digests were analyzed by nanoLC-MS/MS on an Orbitrap-Velos instrument with high sequencing speed to improve the MS/MS sampling and analytical coverage of the samples. MS scans were recorded in the Orbitrap, and MS/MS CID spectra were recorded in the ion trap using a classical parallel acquisition mode to obtain high resolution MS^1 data for peptide quantification while optimizing the number of MS^2 sequencing events to increase peptide identifications. Database searches using MS/MS sequencing data were performed through Mascot, and the results files were parsed and validated based on target decoy calculated FDRs, set at 5% for peptides and 1% for proteins. After realignment in time of nanoLC-MS runs, the software uses the *m/z* and time values associated to validated peptides ions of validated proteins, to extract the XIC of each of them. If some peptide ions were sequenced by MS/MS and validated only in some of the samples to be compared, their XIC signal was extracted in the nanoLC-MS raw file of the other samples using a predicted RT value and a time tolerance window. For protein quantification, a PAI was calculated, defined as the average of XIC area values of at most three intense reference tryptic peptides identified for this protein.

Repeatability of the Label-free Quantification without Sample Fractionation—We first evaluated the repeatability of the label-free analytical workflow by comparing replicate LC-MS

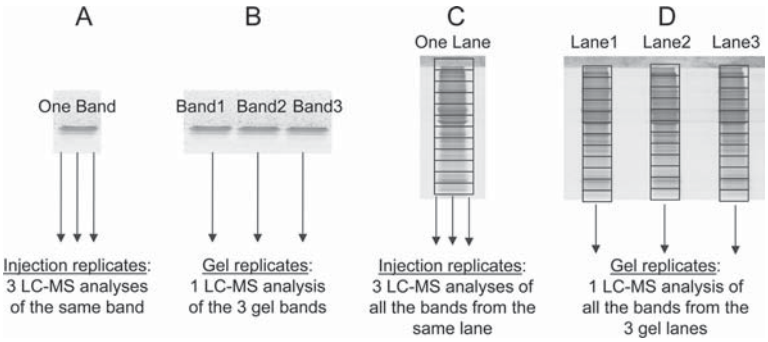
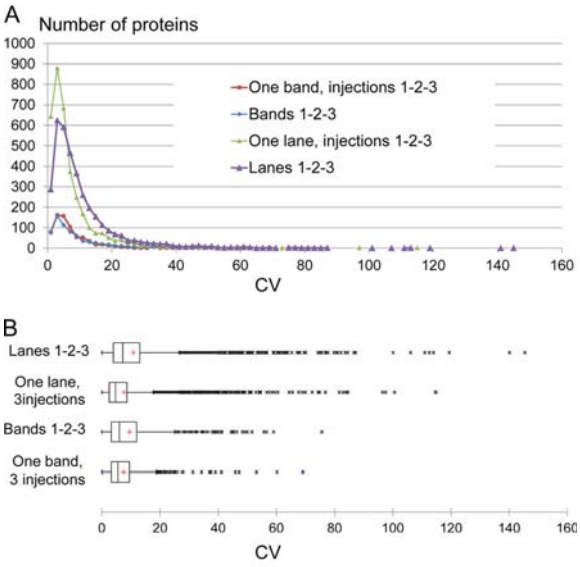


FIG. 1. **Experimental design to estimate the accuracy of label-free quantification with or without sample fractionation.** The same endothelial cell lysate was loaded on a one-dimensional SDS gel and either collected in a single band or fractionated into 12 gel bands cut along the migration lane, and to assess how repeatability is affected by each step of the analytical process, we compared either nanoLC-MS/MS injection replicates or gel replicates. Four experiments were performed. *A*, the protein sample (15 μg) was collected in a single band and digested, and the corresponding peptide digest was analyzed three times by nanoLC-MS/MS. *B*, three identical protein samples (15 μg each) were loaded on the gel and collected in three bands, and after digestion, one-third of each corresponding peptide digest was analyzed once by nanoLC-MS/MS. *C*, the protein sample (100 μg) was fractionated by electrophoresis into 12 gel fractions, the 12 bands were digested, and each of the corresponding peptide digests was consecutively analyzed three times by nanoLC-MS/MS. *D*, three identical protein samples (150 μg each) were loaded on the gel and fractionated into 12 gel bands, and after digestion, the corresponding molecular weight bands of each gel lane were consecutively analyzed once by nanoLC-MS/MS (one-third of resulting peptide digests for each band).

analyses of the same sample, without any fractionation. The first experiment consisted in triplicate nanoLC-MS/MS injections of the tryptic digest prepared from one gel band containing the whole protein mixture (Fig. 1*A*). In that case, sources of errors in the final quantitative results include only the variability of the nanoLC separation, of the mass spectrometry measurement, as well as of potential inconsistencies related to bioinformatic extraction of peptide XICs by the software. To evaluate the additional variability related to upstream sample processing steps (gel loading, gel migration, manual band cutting, in-gel trypsin digestion and peptide extraction), three nanoLC-MS/MS analyses were then performed on the tryptic digests obtained from triplicate gel bands containing each the same sample loaded on the gel (Fig. 1*B*). In both cases, the number of proteins identified from the three analytical runs was very similar (respectively 718 and 715 proteins for injection replicates or gel replicates; supplemental data 1). Although some of these proteins were identified by MS/MS in only one or two of the triplicates, the cross-assignment procedure used in MFPaQ allowed extraction of their MS signal in the runs, which did not contain any identification data for these particular proteins. As shown in supplemental data 2, this method generated a very modest number of missing values for quantification, at both the peptide and protein levels, leading to quantification of 715 and 686 proteins in these two experiments. To evaluate repeatability, the CVs of the PAI values obtained for these proteins were calculated. As shown in Fig. 2, the distribution of CVs for proteins quantified in the three gel replicates is very similar to that of CVs obtained with three injection replicates. The median CV is 5 and 6%, respectively, for the two experiments, and the interquartile range of the CV distribution is slightly increased in the case of gel replicates compared with injection replicates. Experimental steps such as gel migration or gel band processing may account for this little decrease of quantification accuracy observed for gel replicates. However, when the sample is isolated in only one band, such processes are supposed to be quite reproducible. Indeed, they seem to bring only a little additional variability, because the results show that a high percentage of the protein population is still correctly quantified (99% of proteins have CVs under 50%), with a relatively small absolute number of outlier proteins with extreme CV values. These results also confirm that label-free quantification using the identity-based signal extraction procedure in MFPaQ allows an accurate quantification of more than 600 proteins on a complex sample analyzed in a single run. This can also be seen from the correlation plots and the distribution of protein PAI ratios calculated between replicate nanoLC-MS/MS analyses (supplemental data 3).

Label-free Quantification after One-dimensional SDS-PAGE Shotgun Analysis—The sample was then submitted to one-dimensional SDS-PAGE and fractionated into 12 gel bands. Again, to assess how repeatability is affected by each step of the analytical process, we performed either three LC-MS/MS analyses of the 12 gel bands from the same migration lane or LC-MS/MS analyses of the gel bands from three replicate migration lanes of the same sample loaded on the gel. In this latter case, the bands within a particular molecular weight from the three lanes were analyzed successively, and peptide identifications from each of them were used to extract XICs in the corresponding bands, by cross-assignment of peptide signals in replicate LC-MS/MS runs. As expected, after fractionation the analytical coverage of the protein mixture was greatly improved, because more than 3500 unique proteins



C

	One band, no fractionation		Fractionation, 12 gel bands	
	injection replicates (1band)	gel replicates (3x1band)	injection replicates (12 bands)	gel replicates (3x12 bands)
Total number of identified proteins	725	696	3553	3590
Number of proteins with CV>50%	3	5	36	66
Number of proteins with CV>100%	0	0	3	8
Maximum CV	69,1	75,6	114,8	145,4
Median CV	5,5	6,0	4,7	7,2
99% CV distribution	37,4	48,4	49,9	61,2
95% CV distribution	19,6	30,8	24,3	32,9
75% CV distribution	9,4	11,9	8,6	13,1

FIG. 2. **Coefficients of variation for protein PAI values between triplicate LC-MS measurements.** The results are shown for experiments without fractionation (one gel band analyzed three times or three replicate gel bands analyzed once by LC-MS) or with one-dimensional SDS-PAGE fractionation (each of the 12 molecular weight fractions from one gel lane analyzed three times consecutively or analysis of three gel lanes). *A*, histogram of CVs distribution in the different experiments. *B*, box and whisker plots showing the dispersion of protein CVs near the median value. The *bottoms* and *tops* of the *boxes* correspond to the 25th and 75th percentiles of the CV distribution, and *whiskers* correspond to the lowest and highest values within 1.5× interquartile range of these limits. Extreme values falling out of the box plots correspond to outliers. *C*, number of proteins and quantitative repeatability in each experiment.

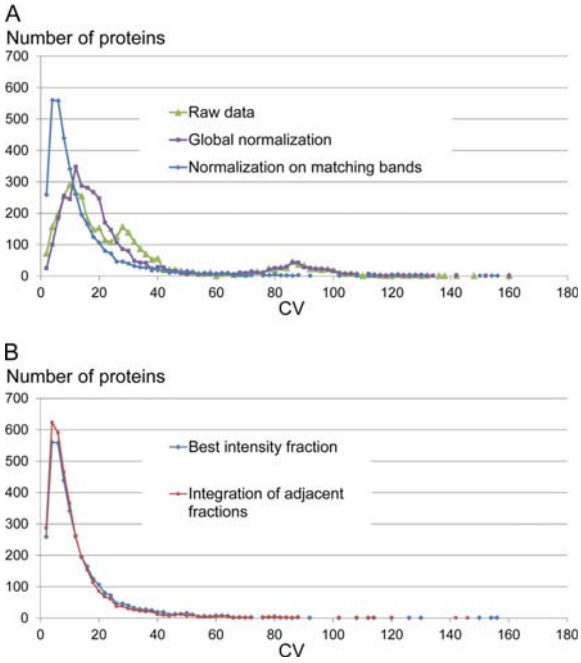


FIG. 3. **Effect of normalization and integration procedures on the quantitative results after SDS-PAGE fractionation.** *A*, histograms of CVs for 1) nontransformed protein PAI values calculated on raw data, 2) normalized protein PAI values transformed with a global correction factor (ratio of summed PAI values for all proteins along each migration lane in replicate experiments), or 3) protein PAI values calculated from normalized XIC signals for each group of matching gel bands. *B*, histograms of CVs for protein PAI values calculated after normalization of XIC signals in matching gel bands, taking into account only the fraction with the best PAI for proteins identified in several gel bands, or after integration by summing PAI values on three consecutive gel bands near the best intensity fraction.

were identified in both experiments (supplemental data 1). Even on this larger population, the signal extraction performed by the software allowed retrieval of quantitative data for almost 99% of the proteins after triplicate sample fractionation through one-dimensional gel (supplemental data 2). Preprocessing of the raw quantitative data was performed to remove the systematic effects and variations because of the measurement process. As for one-shot analysis, a normalization step was used to take into account variability of the nanoLC ESI-MS signal, and in the case of gel replicates, unequal amounts of protein were loaded on the gel and gel processing variability. When this normalization procedure was performed at the scale of the whole experiment (*i.e.*, by comparing signal intensity of all the peptides detected all along the migration lane in replicate experiments), a global correction factor was calculated and used to correct the protein PAI values of replicates experiments against a reference. As shown in Fig. 3*A*, this process improves to some extent the CVs of PAI values for proteins detected in replicate gel lanes but only in a limited way. A significantly better correction was achieved by comparing intensities of peptides detected in matching gel bands, allowing derivation of 12 different normalization factors applied separately to correct quantitative values in each group of molecular weight gel fractions replicates. Obviously, this approach is best suited to correct LC-MS variability, because it compares samples that were measured within a shorter lapse of time and that contain similar protein subpopulations. An automatic normalization

TABLE I
Statistics of peptide migration across the SDS-PAGE fractions

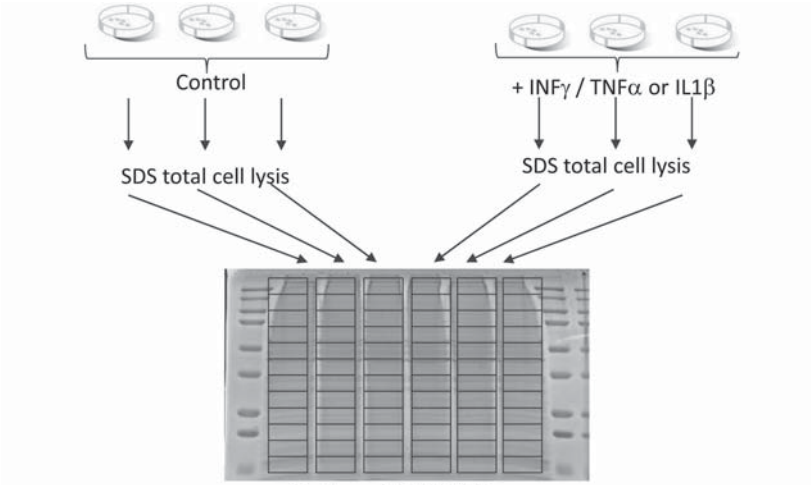
Peptide apex count distribution indicates the number of peptide precursor ions that were detected at their maximal intensity in the same matching fractions across all three replicates (one fraction), in the same fraction in only two replicates (two fractions), or in different fractions in the three replicates (three fractions). Peptide apex distance distribution illustrates the maximal gap between apex fractions when peptides were detected in nonmatching fractions across the replicates.

Peptide apex	No. of precursor ions	
	Three migration lanes	Three injections of one lane
Count distribution		
One fraction	33,932	34,769
Two fractions	1,402	364
Three fractions	91	0
Distance distribution		
0	33,932	34,769
1	1,269	312
2	53	22
3	21	10
4	12	4
5	11	1
6	10	3
7	9	6
8	54	1
9	38	4
10	15	1
11	1	0

procedure to correct peptide intensities of matching fractions in the case of fractionation experiments was thus included in MFPaQ and used in this study.

In addition, to correct for gel migration variability from lane to lane, an integration procedure was also included to sum up the signal of proteins detected in several fractions over the SDS-PAGE lane. However, bands along the migration lane will be corrected with different normalization factors, and integration of signal from very distant gel bands can generate quantitative errors. To evaluate gel migration variability, MFPaQ was used to retrieve the apex of the electrophoretic gel migration pattern for each peptide identified in each replicate experiment. Table I shows the apex count distribution, reflecting the number of peptides that were detected at their maximal intensity in nonmatching fractions in the three different replicates. As expected, the apex of the vast majority of peptide ions was detected in the same gel band, but in the case of replicate gel lanes, for 1402 of the peptide ions (~4% of the total number of precursors) the "best" gel band is identical in only two replicates of three, and for a small minority of them (91 peptide ions, 0.25%), it is different in all three replicates. To some extent, these figures include cases that may be explained by LC-MS variability: indeed, in the case of LC-MS replicates, there is also a small degree of disparity between apex fractions (364 peptide ions, 1% of total, for which the maximal intensity is measured in a nonreproducible way in one injection replicate). However, variability of the electrophoretic migration of proteins along the gel lanes and manual cutting of the fractions account for the majority of the discrepancies in the case of replicate gel lanes. Of the 1493 peptide ions for which conflicts were detected, as shown in Table I, the maximal distance between apex fractions is 1 for 1269 precursors, *i.e.*, the maximal intensity is measured in matching gel bands or in an adjacent band for all three replicates. In many cases, this is probably due to gel cutting inside protein migration patterns and unequal partitioning of these proteins into adjacent gel bands depending on the migration lane. In a small number of other cases, the apex fractions are more distant, probably because of migration problems, irreproducible degradation or precipitation of some proteins, or wrong signal extraction by the software. To correct the most frequent artifacts associated with the SDS-PAGE fractionation process, without introducing additional errors, we thus decided to integrate quantitative data by summing the PAI values for fractions adjacent to the fraction with the best PAI (the same three consecutive fractions for all the replicates to be compared). Fig. 3*B* shows the result of this integration procedure on the CVs of PAI values for proteins detected in replicate gel lanes, compared with CVs calculated by retrieving only the best PAI in one fraction (identified across all the replicates, and the same matching fraction for all of them). Although the distributions are globally very similar, integration brings a small improvement on the CVs, in particular by reducing the number of extreme values (89 proteins of 3585 are measured with a CV higher than 50% when the PAI is retrieved from the best intensity fraction, *versus* 66 proteins when integration is performed). Thus, PAI values were summed up from three consecutive fractions in the case of fractionation experiments.

Finally, as shown in Fig. 2, in the case of sample fractionation, the number of quantified proteins clearly increases, but this is associated with a higher number of extreme values falling of the normal distribution of CVs for both experiments, a significant number of proteins quantified with CVs above 50%, and a higher interquartile range for gel replicates. In the case of replicate injections, quantitative errors occur again from the same causes than in the first one-shot experiment (variations in the nanoLC peptide separation, MS analysis, and bioinformatic processing), and globally, the accuracy of the quantification is thus similar (median CV of ~5% and comparable interquartile ranges). However, the presence of extreme values can be explained by the higher number of low abundance species that are quantified compared with one-shot measurements. Indeed, by increasing the depth of proteome analysis, the fractionation strategy generates quantitative data on low intensity signals that may be subject to larger fluctuations from one run to another or that may be incorrectly extracted by the software. This is illustrated by CV to PAI plots, which reflect a significant decrease of quantitative repeatability for lower PAI values (supplemental data 4). On the other hand, when the 12 gel bands from the three different migration lanes are analyzed independently, additional errors



	INFγ / TNFα experiment						IL1β experiment					
	control1	control2	control3	assay1	assay2	assay3	control1	control2	control3	assay1	assay2	assay3
Number of identified proteins /lane	4044	3881	4095	4029	3981	3907	4471	4506	4542	4691	4533	4387
Total number of quantified proteins	4842						5477					
Variant proteins (fold>2, p<0.05)	207						153					

FIG. 4. **Experimental design and identification results of the large scale quantitative proteomic study of endothelial cells.** Three independent biological experiments were performed by stimulating HUVECs with inflammatory cytokines in culture (either a combination of TNFα and IFNγ or IL1β). Total cell lysates from control and stimulated samples were loaded and fractionated on six parallel gel lanes and cut into 12 gel [GRAPHIC]bands. The table indicates the number of proteins identified for each gel lane, and the total number of proteins identified and quantified in each experiment, as well as the number of proteins detected as differentially expressed.

related to the one-dimensional SDS-PAGE fractionation process (migration and gel band cutting), which was expected to be the most important source of variability, are introduced. The distribution of CVs is shifted compared with what was obtained for injection replicates and now has a median of 7%. Thus, the gel fractionation process contributes to the variability of the measurement. However, even in that case, still 99% of the protein population has CVs for PAI values under 70%. These values illustrate the variability of the gel fractionation process when samples are loaded on adjacent lanes, on the same gel. To further evaluate gel to gel repeatability, which may be an important parameter when numerous samples have to be processed, we also performed triplicate fractionation experiments on different gels. As shown in supplemental data 5, the median CV of proteins PAI shifts from 7% when they are fractionated and quantified on one gel, to 9% when they are quantified from samples fractionated on different gels, and the distribution of CVs is slightly broader. In conclusion, sample fractionation largely improves the depth of proteome coverage, although this is obtained at the expense of quantification accuracy. However, the repeatability of the method is still acceptable for a differential quantitative study, performed with statistical analysis of replicate gel migration lanes.

Large Scale Label-free Quantitative Proteomic Analysis of Human Primary ECs under Inflammatory Conditions—The workflow was then used in the context of a real differential biological analysis, in which we stimulated primary HUVECs with TNFα/IFNγ or IL1β, which represent potent proinflammatory cytokines that trigger inflammatory and immunological responses. The cells were lysed directly in SDS buffer and sonicated, and the resulting protein extract was loaded on a one-dimensional gel. Three biological experiments were performed, with three control samples and three stimulated samples fractionated independently on six migration lanes (Fig. 4). Using the fractionation workflow, we could identify and quantify 4842 and 5477 proteins, respectively, from ECs in the TNFα/IFNγ and IL1β experiments (supplemental data 6 and 7). Statistical analysis was performed on protein PAI values calculated after normalization and integration, as described above. For defining expression changes, two criteria were applied to derive confident data sets of modulated proteins: Student's *t* test *p* value <0.05 and expression fold change >2, as described in previous studies (16). Based on these cut-off values, 207 proteins were found to exhibit a significant variation following TNFα/IFNγ stimulation (175 up-regulated and 32 down-regulated) (supplemental data 6). Endothelial cell response to IL1β stimulation was slightly more restricted, because we measured 153 modulated proteins (119 over-regulated and 34 down-regulated) (supplemental data 7). Functional analysis of modulated proteins using the Protein-

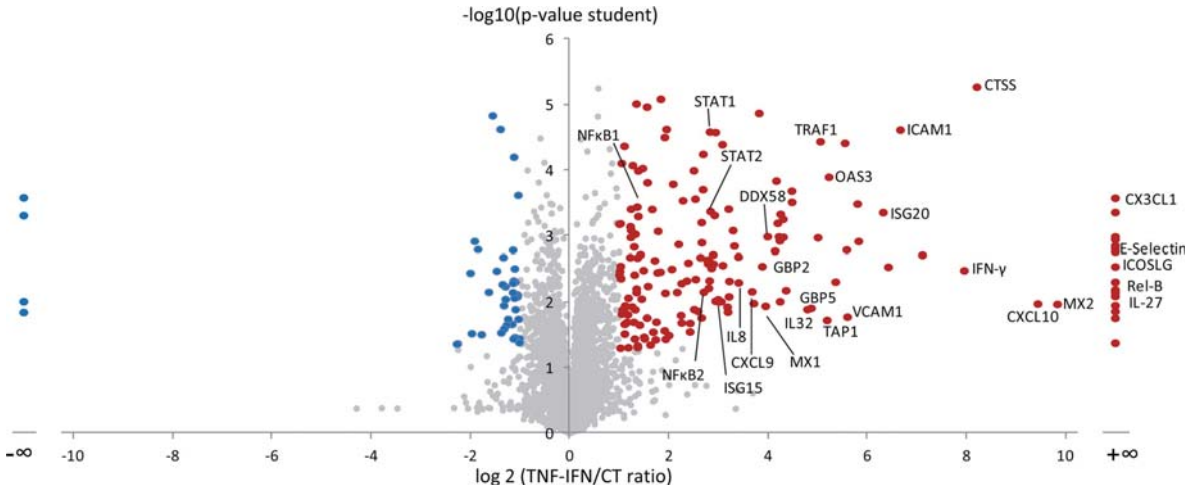


FIG. 5. **Quantitative analysis of endothelial cells proteome following TNFα-IFNγ stimulation.** The volcano plot of the statistical significance of expression level changes (*t* test *p* value) as a function of protein expression ratio between control and inflamed endothelial cells. The *red* and *blue dots* indicate up-regulated and down-regulated genes, respectively.

Center software shows an important enrichment of functional categories related to inflammation and immune response (supplemental data 8). Fig. 5 shows the volcano plot representing statistical significance in function of protein variation between treated and control ECs in the case of TNFα/IFNγ stimulation. Among the most induced proteins, we found many well known cell surface membrane proteins involved in leukocyte recognition and recruitment (E-selectin, ICAM-1, V-CAM1, and ICOSLG), proteins involved in antigen processing and presentation through the class I major histocompatibility complex, but also inflammatory mediators, such as signaling molecules and transcription factors downstream the TNFα pathway (TRAF1, NF-κB, and RELB) or IFNγ pathway (JAK1 and STAT transcription factors), as well as many characteristic interferon-induced proteins involved in antiviral response, as illustrated in Fig. 6. Interestingly, 42 proteins were found to be up-regulated both by IL1β and TNFα/IFNγ (supplemental data 9). Most of them have been described as NF-κB target genes, confirming the role of NF-κB as a key mediator of IL1 and TNFα pathways. To corroborate the results obtained using our quantitative workflow, we tested by quantitative PCR the up-regulation of a series of genes corresponding to modulated proteins identified by the proteomic approach. All of the genes tested confirmed the results of the proteomic study, including strongly induced genes coding for proteins well known to be involved in the inflammatory process and also other genes moderately up-regulated, corresponding to proteins less described in the literature to be part of endothelial cell response to cytokines, such as ROBO1 (supplemental data 10). Altogether, this study shows that the quantitative label-free workflow used here can successfully identify the pathways activated under inflammatory conditions, and it provided a detailed proteomic characterization of the response triggered by inflammatory cytokines in ECs.

DISCUSSION

Global analysis and quantitative comparison of large proteomes is a fruitful approach to get insights into molecular mechanisms of complex biological systems. To obtain a comprehensive picture of such systems, proteomic analysis must be as deep as possible, to map and quantify a large range of protein species, even low abundant ones. Although they have been greatly improved in recent years, the dynamic range and the sequencing speed of mass spectrometers still represent limiting factors for discovery-based proteomics, and in classical experimental LC-MS designs, they restrict the list of proteins that can be detected and quantified in a single-run analysis. To extend the list of identified proteins and obtain quantitative data on minor species, sample prefractionation is thus generally combined to nanoLC-MS analysis, either at the protein level (mainly by SDS-PAGE) or at the peptide level (often by SCX or isoelectric focusing). In recent studies, several thousand proteins could be identified from eukaryotic cells following sample fractionation (1, 17–19). This upstream separation step is often performed on isotopically labeled and mixed samples, ensuring accurate quantification. Here, we evaluated the repeatability of an analytical workflow combining SDS-PAGE fractionation and label-free quantification based on MS signal analysis. Some features of the label-free quantification performed through the MFPaQ software in this study were 1) extraction in raw MS files of XICs from identified and carefully validated peptides, 2) use of a global index for relative quantification at the protein level, derived from the intensity values of at most three intense peptides, and 3)

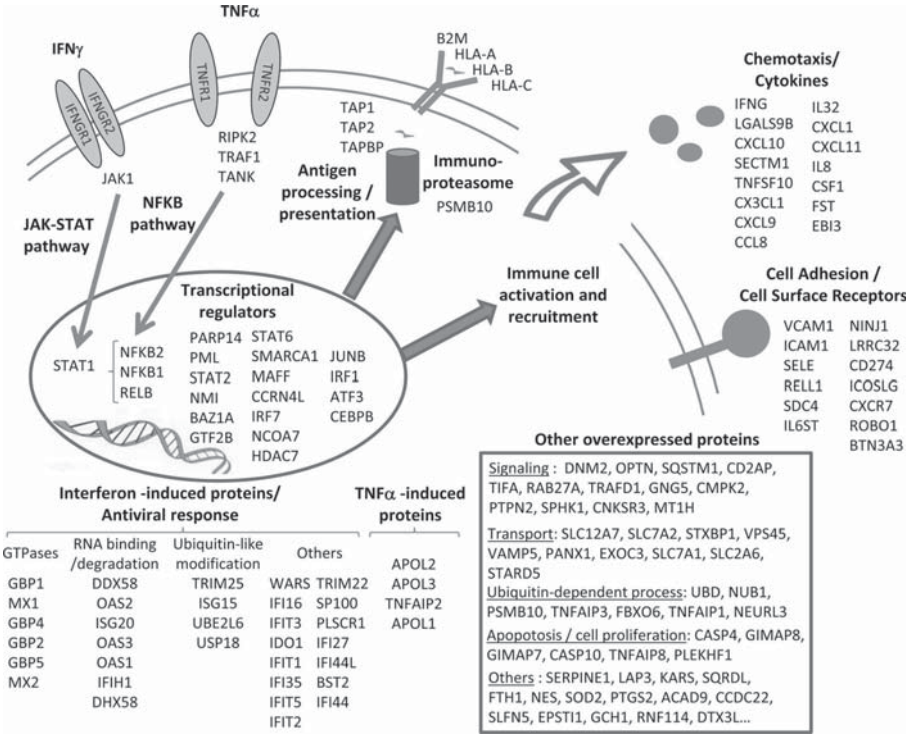


FIG. 6. **Biological pathways activated upon inflammatory response of endothelial cells after TNFα/IFNγ stimulation.** Proteins found to be up-regulated from the proteomic analysis are illustrated and classified in function of the biological processes in which they are involved or their subcellular location, according to literature data.

integration of the quantitative data from different fractions and overview of the shotgun experiment through the MFPaQ interface. The identity-based approach allowed extraction of signal for confidently identified peptides in an automated batch mode, directly on the 72 raw files of the comparative experiment (two conditions, three replicates, and twelve fractions). Quantitative data can be viewed at the peptide level through the MFPaQ interface, which displays the XICs of the peptide ions in all of the raw files corresponding to matched fractions but was also directly integrated by the software at the protein level, by computing the mean value of at most three intense peptides per protein. In the case of relatively abundant proteins identified with more than three peptides, this allowed calculation of PAI values on the highest quality signals, for a more accurate quantification. However, minor species identified with less than three peptides were also quantified based on the available peptide signals. Finally, data normalization and integration procedures were used in MFPaQ to correct LC-MS variability and errors related to nonreproducible electrophoretic migration of proteins in the case of sample fractionation.

Overall, our approach proved to behave in a robust way for the quantification of a complex proteome. Our results show that for one-shot analysis, label-free quantification can be achieved with good accuracy (median CV of 5%, 99% proteins with CV values $<$ 48%). Sample fractionation largely improved the depth of proteomic coverage, and this was associated with a moderate decrease of quantitative measurement repeatability (median CV of 7%, 99% proteins with CV values of $<$62%). Thus, prefractionation by SDS-PAGE appears to be compatible with label-free quantitation for the extensive analysis of complex proteomes. In the present study, it provided a detailed characterization of the proteomic variations associated with the inflammatory response in human primary ECs. Although they can be maintained for some time in culture, these primary cells are not easily amenable to SILAC labeling, and label-free methods were particularly convenient for their quantitative analysis. For each condition (control or stimulated cells), triplicate samples were fractionated by SDS-PAGE, and analysis of each gel lane led to the identification of up to 4600 unique proteins, based on a protein FDR of 1%. Globally, analysis of the six different gel lanes by nanoLC-MS/MS and cross-assignment of peptide signals between samples led to the identification and quantitation of more than 5400 unique proteins in the IL1β experiment. In a recent study, the use of very long LC-MS gradients on 50-cm-long columns was described for in-depth analysis of complex proteomes without prefractionation (20). Such experi-

mental strategies are probably not yet routinely applicable because they generally require high pressure chromatography devices and generate very large raw files that may be difficult to process with most current bioinformatic tools. However, they appear to be a promising approach that in principle could combine the advantages of extensive proteomic coverage and quantitative accuracy that can be obtained in single-run analysis. However, the quantitative repeatability of these several-hours-long LC-MS gradients is still to be assessed on replicate analytical runs. Regarding analytical time, analysis of 12 fractions of a one-dimensional gel lane in 2-h-long LC-MS gradients on conventional LC systems is three times longer, but technically easier to implement, than analysis of the whole sample on a long column with an 8-h gradient. On the other hand, sample prefractionation still probably represents up to now the most efficient way to get the deepest analytical coverage of a complex proteome, and the present study shows that the additional variability associated with this upstream process does not preclude quantitative analysis. Thus, although there is a trade-off between analytical time, quantitative accuracy, and proteomic coverage, putting the emphasis on this last parameter would probably require both sample fractionation and extensive peptide separation with long gradients for very extensive characterization of complex proteomes like the human one. Here, by using only a shotgun approach based on one-dimensional protein fractionation, sufficient depth was obtained here to detect changes on very low abundance proteins such as some transcription factors and signaling molecules. Although this label-free approach requires more analytical time than a SILAC-based experiment, because the samples are injected separately, it also avoids possible quantitation errors caused by superposition of one peptide of the SILAC pair with other different isotopic peptide patterns. In our hands, it also yielded a higher number of identified proteins, because the same MS/MS sequencing time is spent on less complex mixtures, containing half the number of peaks compared with isotopically labeled and mixed samples. Thus, MS intensity-based label-free quantification associated with SDS-PAGE fractionation appears as a valuable strategy for the differential analysis of complex proteomes.

The shotgun approach used in our study provided an in-depth characterization of the EC proteome and the label-free quantitative proteomic workflow allowed deciphering the inflammatory response of these cells. TNFα and IFNγ are potent pleiotropic cytokines that exert a number of biological effects and trigger a set of complex molecular programs in response to microbial or viral infection. IFNγ is produced mainly by NK cells and T helper type I cells and, through binding to its specific type II IFN receptor, activates the JAK-STAT signaling pathway, to induce the expression of a large number of genes (21, 22). In this large scale proteomic experiment, we measured an up-regulation of the JAK1 kinase and STAT1 transcription factor, which are known to mediate IFNγ response and regulate genes downstream of γ-activated sequence elements. We could also detect an increase of proteins involved in the TNFα signal transduction pathway, such as TRAF1, TANK, and RIPK2, converging to the activation of the NF-κB transcription factor (23). Accordingly, we measured overexpression of NF-κB subunits (NFKB2, NFKB1, and RELB) and a decrease of the inhibitor of NF-κB (IKBB), which controls nuclear translocation of NF-κB and undergoes proteasomal degradation upon TNFα signaling (24). In addition, several other proteins involved in transcriptional regulation were shown to be up-regulated after stimulation by the two cytokines (Fig. 6), such as members of the STAT family (STAT2 and STAT6); the PARP-14 protein, which enhances STAT6-dependent transcription (25); or the IRF1 secondary transcription factor, which is induced by STAT1 and plays a key role in orchestrating the IFN-induced inflammatory response (26).

One major biological process that makes part of this response in ECs is recruitment and activation of leukocytes to the inflammatory site. ECs line the blood vessel walls, and upon stimulation by cytokines, they secrete chemokines, which are chemoattractants for lymphocytes and monocytes, and express at their surface adhesion molecules that capture circulating leukocytes. We measured in our analysis the strong up-regulation of a panel of chemokines, such as Fractalkine, IL8, CXCL10, CXCL11, CXCL9, CXCL1, or CCL8, and of other secreted signaling molecules such as IL27b, or IL32. Indeed, IL32 was recently shown to be a critical regulator of EC function, which is strongly increased upon IL1β or TNFα stimulation and mediates in particular the expression of cell surface adhesion molecules involved in lymphocytes binding such as VCAM1 (27). Although our analysis was performed on a whole cell lysate and not focused on membrane proteins, we could clearly measure the overexpression of several cell surface proteins involved in cell-cell interactions. Leukocyte adhesion molecules such as E-selectin, ICAM1, and VCAM1 were the among the most strongly induced gene products and represent major players in the initial rolling and arrest step of leukocytes-EC interaction along the blood vessel walls (28). Simultaneously, molecules known to promote procoagulant activity at the EC surface such as plasminogen activator inhibitor 1 (Serpine 1) were also induced (29). Other cell surface proteins were shown to be overexpressed in response to TNFα/IFNγ treatment, such as the ICOS-ligand protein, which is an important costimulator in EC-mediated T cell activation (30); the ROBO1 receptor that may play a role in leukocyte migration (31); or the programmed cell death 1 receptor ligand PDL1, involved in immune regulation (32). Additionally, the expression of cell surface class I MHC molecules was also increased upon stimulation by inflammatory cytokines. ECs constitutively express class I MHC molecules *in vivo*, which are significantly decreased during cell culture but can be restored upon IFNγ or TNFα treatment (33). Following stimulation, we observed concomitantly the induction of all the machinery for antigen processing and presentation, including

the immunoproteasome responsible for degradation of cytoplasmic endogenous or viral proteins; TAP proteins involved in antigenic peptide transport to the endoplasmic reticulum; and Tapasin, which binds to the TAP complex and allows antigen loading to assembled MHC molecules (34). Finally, a wide range of interferon-induced proteins were detected as strongly up-regulated, such as small GTPases (guanylate-binding proteins, Mx1, and Mx2) and the 2′-5′-oligoadenylate synthase family, which play an essential role on viral RNA degradation and the innate immune response to viral infection (21).

The EC response to IL1β stimulation, as characterized from the second large scale proteomic experiment, shared many features with the response induced by the TNFα/IFNγ treatment. Major biological processes of EC inflammatory activation were again highlighted, *i.e.*, secretion of chemoattractant molecules and other cytokines (CXCL6, interleukin 8, CXCL1, CXCL2, CCL2, granulocyte colony-stimulating factor, macrophage colony-stimulating factor, interleukin 27, and interleukin 32), expression of cell surface leukocyte ligands (ICAM1, VCAM1, selectin, ICOS ligand, Syndecan-4), as well as antigen processing and presentation through MHC class I molecules (immunoproteasome subunits, TAP1, Tapasin, and HLA molecules). As IL1β signals through the NF-κB pathway, many proteins induced by IL1β were also induced by TNFα (see supplemental data 9). Both cytokine are, for example, endogenous pyrogens that cause fever, and in both experiments, we found an up-regulation of the prostaglandin G/H synthase 2 of the prostaglandin G/H synthase 2 (cyclooxygenase-2, COX2), which is responsible for synthesis of prostaglandin E_2 prostaglandin, the key molecule for activation of thermosensitive neurons in the hypothalamus (35, 36). Additionally, in the IL1β experiment, we detected the induction of phospholipase A_2, which hydrolyzes glycerophospholipids to produce arachidonic acid, the rate-limiting step in the synthesis of prostaglandin E_2 by COX2. In this experiment, we could also specifically detect the induction of the cysteine protease caspase 1, which is directly involved in cleavage of proactive IL1β into its mature form, as well new regulatory molecules such as TC1, which has been described as a novel endothelial inflammatory regulator that is up-regulated by IL1β and amplifies NF-κB signaling via a positive feedback (37).

Many proteins could be identified that were not previously described as activated in ECs, deserving further studies to determine their exact function in the inflammatory process. For example, the ROBO1 receptor protein has been described to be involved in axon guidance and neuronal precursor cell migration (38), but its potential role in mediating cell-cell interactions at the endothelial surface under inflammatory conditions has been poorly described (39). Here, we show that this protein is overexpressed in HUVECs after TNFα/IFNγ stimulation, and the induction of the corresponding gene was confirmed by quantitative PCR for both TNFα/IFNγ and IL1β treatment. Another example is the circadian deadenylase Nocturnin, which was found to be significantly induced by both TNFα/IFNγ and IL1β stimulations. This protein, that is under circadian regulation, can also mediate immediate early gene responses, and it has been hypothesized that it could be involved in the post-transcriptional regulation of both rhythmic and acutely inducible mRNAs, by controlling mRNA decay through poly(A) tail removal (40). Indeed, very recently it was shown that Nocturnin can be induced by endotoxin lipopolysaccharide and that it stabilizes the proinflammatory transcript inducible nitric-oxide synthase (40), suggesting that Nocturnin could play a role in the circadian response to inflammatory signals. The proteomic data obtained here indicate that it is also induced in endothelial cells upon stimulation with TNFα/IFNγ and IL1β and thus support the idea that this protein could play a general role in the regulation of cytokine-induced inflammatory response.

In conclusion, this is the most extensive proteomic study of EC to date, performed on the widely used *in vitro* primary endothelial cell model HUVEC. It allowed identification in these endothelial cells of more than 5400 proteins, adding some more depth to a large scale data set previously published (41), in which ~3800 proteins were identified and 1300 proteins could be quantified by ^{18}O labeling, following treatment with the proangiogenic factor vascular endothelial growth factor. The present study provides the first complete characterization at the proteomic level of the EC response to inflammatory cytokines such as TNFα, IFNγ, and particularly IL1β. The list of proteins modulated by these factors, as characterized here in a global way, can thus represent a reference to study the function of other newly discovered interleukins of the IL1 family that may trigger similar responses but also some specific pathways.

* This work was supported by grants from the Agence Nationale de la Recherche (Programme Plates-formes Technologiques du Vivant); Fondation pour la Recherche Médicale (Programme Grands Equipements); Ibisa (Infrastructures en Biologie, Santé, et Agronomie); and FEDER (Fonds Européen de Développement Régional); and a fellowship from Région Midi-Pyrénées (to N. D.).

S This article contains supplemental material.

¶ These authors contributed equally to this study.

‖ Supported by the "Ligue Nationale contre le Cancer" (LIGUE 2009).

** To whom correspondence may be addressed: Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, 31077 Toulouse cedex 4, France. E-mail: bernard.monsarrat@ipbs.fr.

‡‡ To whom correspondence may be addressed: Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, 31077 Toulouse cedex 4, France. E-mail: gonzalez@ipbs.fr.

REFERENCES

1. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362
2. Vaudel, M., Sickmann, A., and Martens, L. (2010) Peptide and protein quantification: A map of the minefield. *Proteomics* **10,** 650–670
3. Park, S. K., Venable, J. D., Xu, T., and Yates, J. R., 3rd (2008) A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat. Methods* **5,** 319–322

4. Searle, B. C. (2010) Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **10,** 1265–1269
5. Heinecke, N. L., Pratt, B. S., Vaisar, T., and Becker, L. (2010) PepC: Proteomics software for identifying differentially expressed proteins based on spectral counting. *Bioinformatics* **26,** 1574–1575
6. Bellew, M., Coram, M., Fitzgibbon, M., Igra, M., Randolph, T., Wang, P., May, D., Eng, J., Fang, R., Lin, C., Chen, J., Goodlett, D., Whiteaker, J., Paulovich, A., and McIntosh, M. (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics* **22,** 1902–1909
7. Jaitly, N., Mayampurath, A., Littlefield, K., Adkins, J. N., Anderson, G. A., and Smith, R. D. (2009) Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics* **10,** 87
8. Mueller, L. N., Rinner, O., Schmidt, A., Letarte, S., Bodenmiller, B., Brusniak, M. Y., Vitek, O., Aebersold, R., and Müller, M. (2007) SuperHirn: A novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* **7,** 3470–3480
9. Sturm, M., Bertsch, A., Gröpl, C., Hildebrandt, A., Hussong, R., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., Zerck, A., Reinert, K., and Kohlbacher, O. (2008) OpenMS: An open-source software framework for mass spectrometry. *BMC Bioinformatics* **9,** 163
10. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., and Ahn, N. G. (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **4,** 1487–1502
11. Hoffert, J. D., Wang, G., Pisitkun, T., Shen, R. F., and Knepper, M. A. (2007) An automated platform for analysis of phosphoproteomic datasets: Application to kidney collecting duct phosphoproteins. *J. Proteome Res.* **6,** 3501–3508
12. Tsou, C. C., Tsai, C. F., Tsui, Y. H., Sudhir, P. R., Wang, Y. T., Chen, Y. J., Chen, J. Y., Sung, T. Y., and Hsu, W. L. (2010) IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. *Mol. Cell. Proteomics* **9,** 131–144
13. Bouyssié, D., Gonzalez de Peredo, A., Mouton, E., Albigot, R., Roussel, L., Ortega, N., Cayrol, C., Burlet-Schiltz, O., Girard, J. P., and Monsarrat, B. (2007) Mascot file parsing and quantification (MFPaQ), a new software to parse, validate, and quantify proteomics data generated by ICAT and SILAC mass spectrometric analyses: Application to the proteomics study of membrane proteins from primary human endothelial cells. *Mol. Cell. Proteomics* **6,** 1621–1637
14. Mouton-Barbosa, E., Roux-Dalvai, F., Bouyssié, D., Berger, F., Schmidt, E., Righetti, P. G., Guerrier, L., Boschetti, E., Burlet-Schiltz, O., Monsarrat, B., and Gonzalezde Peredo, A. (2010) In-depth exploration of cerebrospinal fluid by combining peptide ligand library treatment and label-free protein quantification. *Mol. Cell. Proteomics* **9,** 1006–1021
15. Navarro, P., and Vázquez, J. (2009) A refined method to calculate false discovery rates for peptide identification using decoy databases. *J. Proteome Res.* **8,** 1792–1796
16. Sun, N., Pan, C., Nickell, S., Mann, M., Baumeister, W., and Nagy, I. (2010) Quantitative proteome and transcriptome analysis of the archaeon *Thermoplasma acidophilum* cultured under aerobic and anaerobic conditions. *J. Proteome Res.* **9,** 4839–4850
17. Graumann, J., Hubner, N. C., Kim, J. B., Ko, K., Moser, M., Kumar, C., Cox, J., Schöler, H., and Mann, M. (2008) Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* **7,** 672–683
18. Geiger, T., Cox, J., and Mann, M. (2010) Proteomic changes resulting from gene copy number variations in cancer cells. *PLoS Genet.* **6,** pii
19. Luber, C. A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keeffe, M., and Mann, M. (2010) Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* **32,** 279–289
20. Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Froehlich, F., Cox, J., and Mann, M. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without pre-fractionation. *Mol. Cell. Proteomics* 10.1074/mcp.M110.003699
21. Boehm, U., Klamp, T., Groot, M., and Howard, J. C. (1997) Cellular responses to interferon-γ. *Annu. Rev. Immunol.* **15,** 749–795
22. Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H., and Schreiber, R. D. (1998) How cells respond to interferons. *Annu. Rev. Biochem.* **67,** 227–264
23. Baud, V., and Karin, M. (2001) Signal transduction by tumor necrosis factor and its relatives. *Trends Cell Biol.* **11,** 372–377
24. Traenckner, E. B., Pahl, H. L., Henkel, T., Schmidt, K. N., Wilk, S., and Baeuerle, P. A. (1995) Phosphorylation of human IκB-α on serines 32 and 36 controls IκB-α proteolysis and NF-κB activation in response to diverse stimuli. *EMBO J.* **14,** 2876–2883
25. Mehrotra, P., Riley, J. P., Patel, R., Li, F., Voss, L., and Goenka, S. (2011) PARP-14 functions as a transcriptional switch for Stat6-dependent gene activation. *J. Biol. Chem.* **286,** 1767–1776
26. Taniguchi, T., Ogasawara, K., Takaoka, A., and Tanaka, N. (2001) IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.* **19,** 623–655
27. Nold-Petry, C. A., Nold, M. F., Zepp, J. A., Kim, S. H., Voelkel, N. F., and Dinarello, C. A. (2009) IL-32-dependent effects of IL-1β on endothelial cell functions. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 3883–3888
28. Springer, T. A. (1994) Traffic signals for lymphocyte recirculation and leukocyte emigration: The multistep paradigm. *Cell* **76,** 301–314
29. van Hinsbergh, V. W., Kooistra, T., van den Berg, E. A., Princen, H. M., Fiers, W., and Emeis, J. J. (1988) Tumor necrosis factor increases the production of plasminogen activator inhibitor in human endothelial cells *in vitro* and in rats *in vivo*. *Blood* **72,** 1467–1473
30. Khayyamian, S., Hutloff, A., Büchner, K., Gräfe, M., Henn, V., Kroczek, R. A., and Mages, H. W. (2002) ICOS-ligand, expressed on human endothelial cells, costimulates Th1 and Th2 cytokine secretion by memory CD4+ T cells. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 6198–6203
31. Prasad, A., Qamri, Z., Wu, J., and Ganju, R. K. (2007) Slit-2/Robo-1 modulates the CXCL12/CXCR4-induced chemotaxis of T cells. *J. Leukocyte Biol.* **82,** 465–476
32. Singh, A. K., Stock, P., and Akbari, O. (2011) Role of PD-L1 and PD-L2 in allergic diseases and asthma. *Allergy* **66,** 155–162
33. Lapierre, L. A., Fiers, W., and Pober, J. S. (1988) Three distinct classes of regulatory cytokines control endothelial cell MHC antigen expression. Interactions with immune γ interferon differentiate the effects of tumor necrosis factor and lymphotoxin from those of leukocyte α and fibroblast β interferons. *J. Exp. Med.* **167,** 794–804
34. Li, S., Paulsson, K. M., Chen, S., Sjögren, H. O., and Wang, P. (2000) Tapasin is required for efficient peptide binding to transporter associated with antigen processing. *J. Biol. Chem.* **275,** 1581–1586
35. Dinarello, C. A. (1999) Cytokines as endogenous pyrogens. *J. Infect. Dis.* **179,** (Suppl. 2) S294–S304
36. Dinarello, C. A., Gatti, S., and Bartfai, T. (1999) Fever: Links with an ancient receptor. *Current biology* **9,** R147–R150
37. Kim, J., Kim, Y., Kim, H. T., Kim, D. W., Ha, Y., Kim, J., Kim, C. H., Lee, I., and Song, K. (2009) TC1(C8orf4) is a novel endothelial inflammatory regulator enhancing NF-κB activity. *J. Immunol.* **183,** 3996–4002
38. Wong, K., Park, H. T., Wu, J. Y., and Rao, Y. (2002) Slit proteins: Molecular guidance cues for cells ranging from neurons to leukocytes. *Curr. Opin. Genet. Dev.* **12,** 583–591
39. Legg, J. A., Herbert, J. M., Clissold, P., and Bicknell, R. (2008) Slits and roundabouts in cancer, tumour angiogenesis and endothelial cell migration. *Angiogenesis* **11,** 13–21
40. Niu, S., Shingle, D. L., Garbarino-Pico, E., Kojima, S., Gilbert, M., and Green, C. B. (2011) The circadian deadenylase Nocturnin is necessary for stabilization of the iNOS mRNA in mice. *PLoS One* **6,** e26954
41. Jorge, I., Navarro, P., Martínez-Acedo, P., Núñez, E., Serrano, H., Alfranca, A., Redondo, J. M., and Vázquez, J. (2009) Statistical model to analyze quantitative proteomics data obtained by $^{18}O/^{16}O$ labeling and linear ion trap mass spectrometry: Application to the study of vascular endothelial growth factor-induced angiogenesis in endothelial cells. *Mol. Cell. Proteomics* **8,** 1130–1149

<u>Supplemental data 2</u>: Assessment of the number of missing values for peptide and protein quantification in triplicate analysis

Results are shown for experiments without fractionation (one gel band analyzed 3 times, or 3 replicate gel bands analyzed once by LC-MS) or with 1D SDS-PAGE fractionation (in that case, the results are shown at the peptide level for the first gel fraction, whereas at the protein level, the MFPaQ software was used to integrate quantitative data

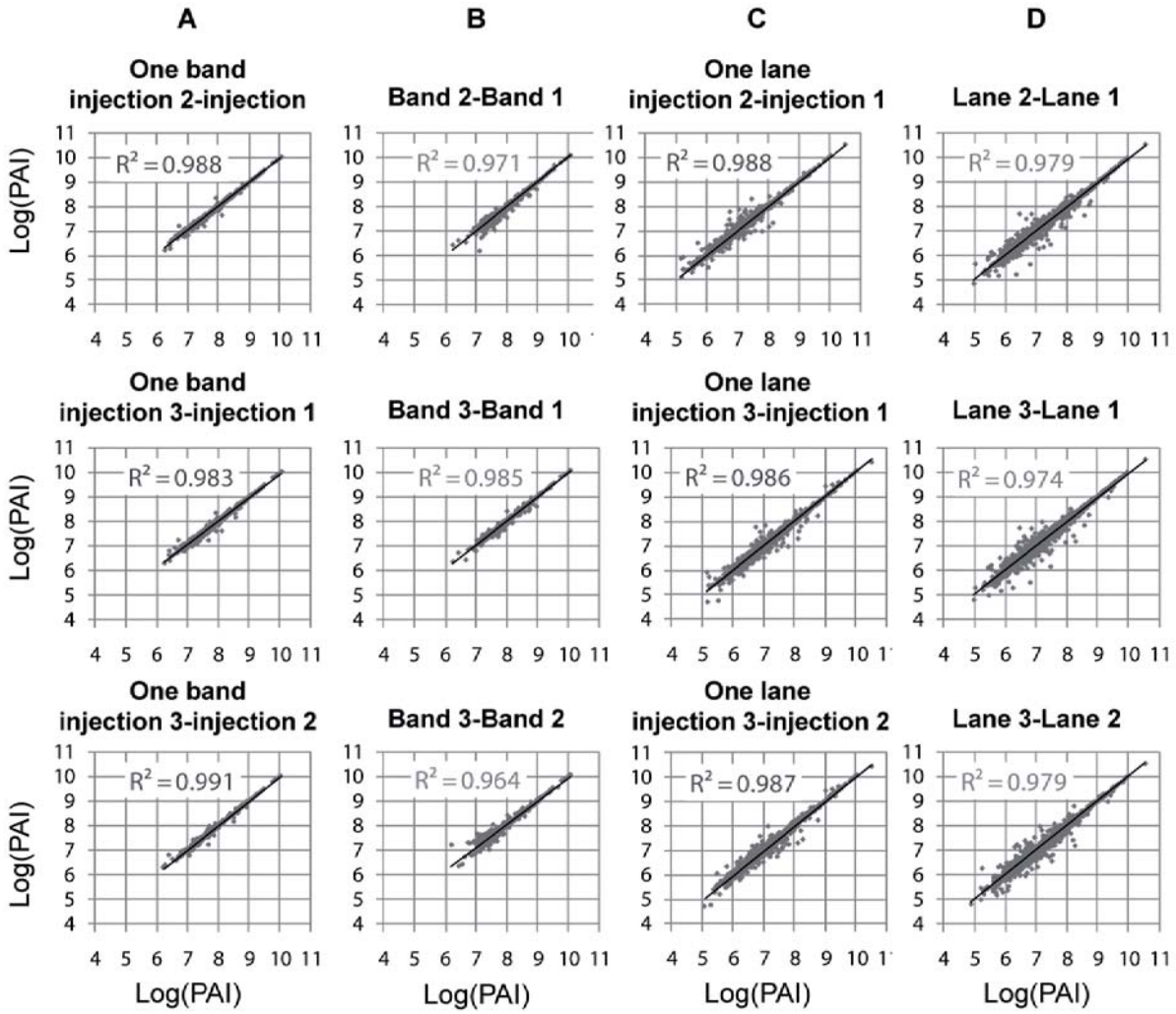
PEPTIDE IONS

	Total number of peptide ions identified	MS/MS performed in all 3 triplicates	MS/MS performed in 2 of the triplicates	MS/MS performed in only 1 triplicate	XIC extracted in all 3 triplicates	XIC extracted in 2 of the triplicates	XIC extracted in only 1 triplicate	XIC not extracted at all	% full extraction	% 1missing value	% 2missing values	% 3missing values
1band injected 3X	5209	2778	1114	1317	5137	49	19	4	98,62	0,94	0,36	0,08
3 bands	5203	2658	1155	1390	4929	256	14	4	94,73	4,92	0,27	0,08
1lane, fraction #1 injected 3X	5222	2965	1105	1152	5054	116	36	16	96,78	2,22	0,69	0,31
fraction#1 from the 3 lanes	5212	2723	1090	1399	4888	258	52	14	93,78	4,95	1,00	0,27

PROTEINS

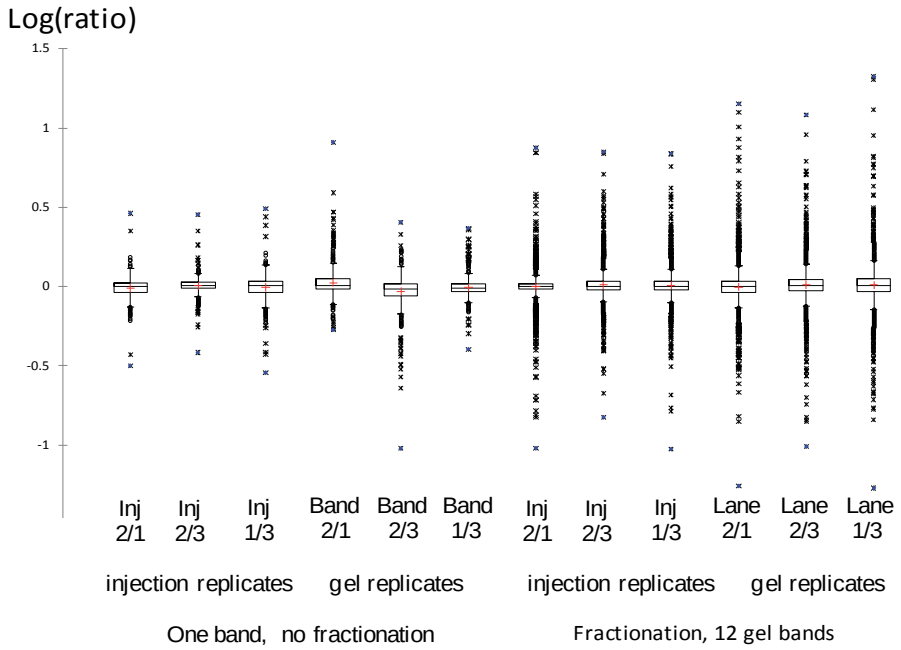
	Total number of proteins identified	MS/MS performed in all 3 triplicates	MS/MS performed in 2 of the triplicates	MS/MS performed in only 1 triplicate	Protein quantified in all 3 triplicates	Protein quantified in 2 of the triplicates	Protein quantified in only 1 triplicate	Protein quantified at all	% full extraction	% 1missing value	% 2missing values	% 3missing values
1band injected 3X	718	487	109	122	715	3	0	0	99,58	0,42	0,00	0,00
3 bands	715	468	118	129	686	28	1	0	95,94	3,92	0,14	0,00
1lane, each band injected 3X	3563	2695	391	477	3526	27	10	0	98,96	0,76	0,28	0,00
3 lanes	3614	2611	439	564	3564	32	18	0	98,62	0,89	0,50	0,00

<u>Supplemental data 3</u>: Reproducibility of label-free quantitative analysis between two replicate nanoLC-MS experiments (with or without sample fractionation)



***Supplemental data 3A**: Correlation plots showing the reproducibility of label-free quantitative analysis from run to run.* PAI values of proteins quantified in two replicate analysis were plotted against each other after $\log_{10}$ transformation. Correlation profiles are shown for experiments without fractionation: **panel A**, LC-MS injection replicates (one gel band analyzed 3 times), and **panel B**, LC-MS analysis of gel band replicates (3 replicate gel bands analyzed once by LC-MS), or with 1D SDS-PAGE fractionation: **panel C**, LC-MS injection replicates of one migration lane (each of the 12 MW fractions from one gel lane analyzed 3 times), and **panel C**, gel migration lanes replicates (fractions within a particular MW from each of the 3 gel lanes analyzed successively by LC-MS).

A



B

	One band, no fractionation						Fractionation, 12 gel bands					
	injection replicates			gel replicates			injection replicates			Gel replicates		
	inj 1/2	inj 2/3	inj 1/3	band 1/2	band 2/3	band 1/3	inj 1/2	inj 2/3	inj 1/3	lane 1/2	lane 2/3	lane 1/3
Minimum ratio	0.32	0.38	0.29	0.53	0.10	0.40	0.10	0.15	0.09	0.06	0.10	0.05
Max ratio	2.90	2.84	3.09	8.12	2.54	2.33	7.48	7.04	6.87	14.15	12.07	20.98
Mean ratio	0.99	1.02	1.00	1.09	0.96	1.00	1.01	1.04	1.04	1.03	1.06	1.07
Mean $\log_{10}$(ratio)	-0.01	0.01	0.00	0.02	-0.03	-0.01	0.00	0.01	0.01	0.00	0.01	0.01
Standard deviation $\log_{10}$(ratio)	0.06	0.05	0.08	0.09	0.10	0.07	0.08	0.08	0.08	0.11	0.11	0.12
Maximum fold (100%)	3.17	2.84	3.50	8.12	10.52	2.49	10.46	7.04	10.59	18.10	12.07	20.98
99% fold distribution	1.53	1.51	2.01	2.26	2.66	1.79	2.16	2.27	2.24	2.75	2.73	2.94
95% fold distribution	1.35	1.26	1.41	1.59	1.66	1.40	1.34	1.45	1.45	1.57	1.58	1.63
75% fold distribution	1.13	1.09	1.15	1.16	1.19	1.12	1.09	1.12	1.12	1.17	1.18	1.20

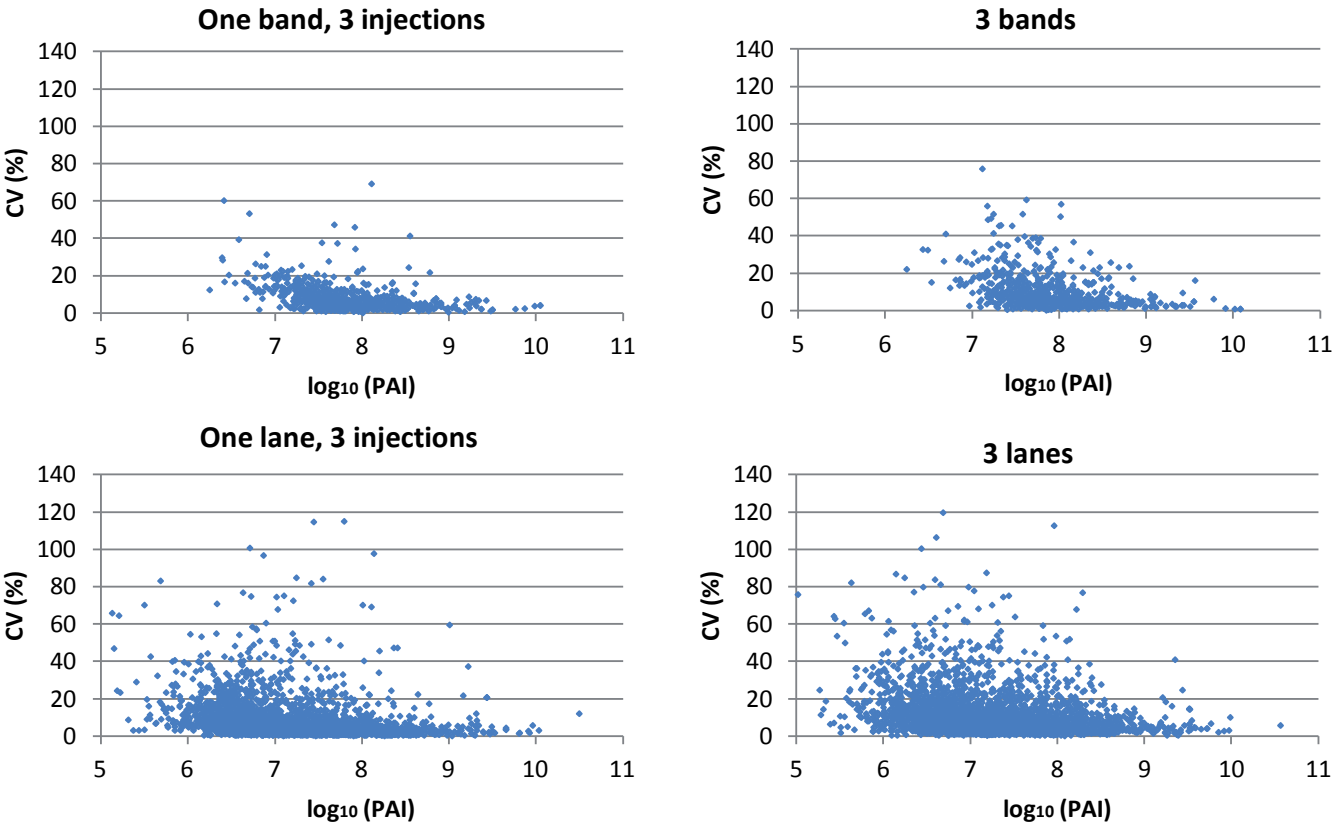
Supplemental data 3B: *Distribution of protein PAI ratios from run to run.*
Ratios of PAI values were calculated for proteins quantified in two individual LC-MS replicate measurements and $\log_{10}$ transformed. **Panel A**: Box-and-whisker diagrams are shown for experiments without fractionation (LC-MS injection replicates or gel band replicates) or with 1D SDS-PAGE fractionation (LC-MS injection replicates or gel migration lanes replicates), and illustrate the dispersion of protein ratios around the median value: bottom and top of the boxes correspond to the 25th and 75th percentile of the $\log_{10}$(ratio) distribution, and whiskers to the lowest and highest values within 1.5 x interquartile range of these limits. Extreme values falling out of the box-plots correspond to outliers. **Panel B:** The table indicates for each experiment some numerical values associated to the distribution of protein PAI ratios and PAI folds. Min and Max ratios indicate, for each comparison, the extreme values that can be obtained on absolute PAI ratios for relative quantification of some proteins, which increase in the case of sample fractionation. Mean ratios are close to 1 (or close to 0 after log-transformation), indicating that the distribution of ratios is correctly centered following data normalization. Standard deviations were calculated on the symmetrical distributions of log-transformed ratios, and increase slightly in the case of sample fractionation, showing higher dispersion of the values. To illustrate the values obtained for different percentiles of the total population, the ratios were then converted into fold changes (x->1/x for ratios <1), and the table indicates maximal fold values obtained either for the total population (100%), or 99%, 95%, and 75% of the population.

<u>Supplemental data 4</u>: Variability of the quantitative measurement as a function of protein abundance

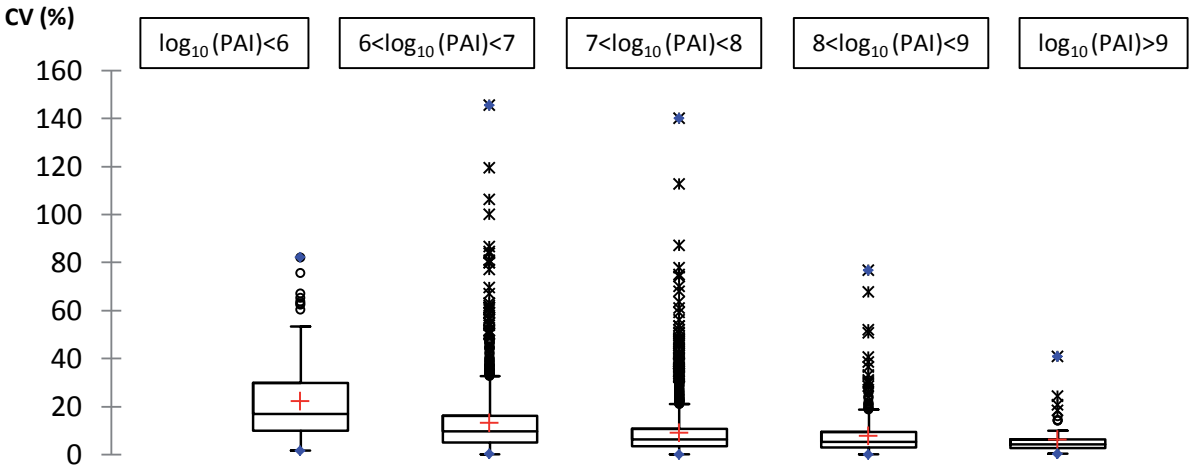
A, CVs-to-PAI plots : CVs of protein PAI values for triplicate measurements were plotted against the PAI value measured in one of the replicate, as a measure of protein abundance. Plots are shown for experiments without fractionation (one gel band analyzed 3 times, or 3 replicate gel bands analyzed once by LC-MS), or with 1D SDS-PAGE fractionation (each fraction from one gel lane analyzed 3 times, or gel migration lanes triplicates)

B, Box-plots of the CVs obtained in the case of triplicate gel lanes, for different classes of protein abundances, showing a decrease in quantitative repeatability for low-intensity proteins

A

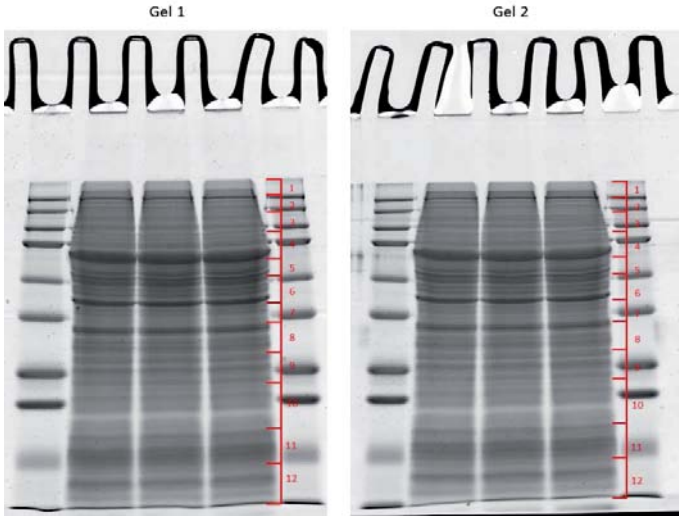


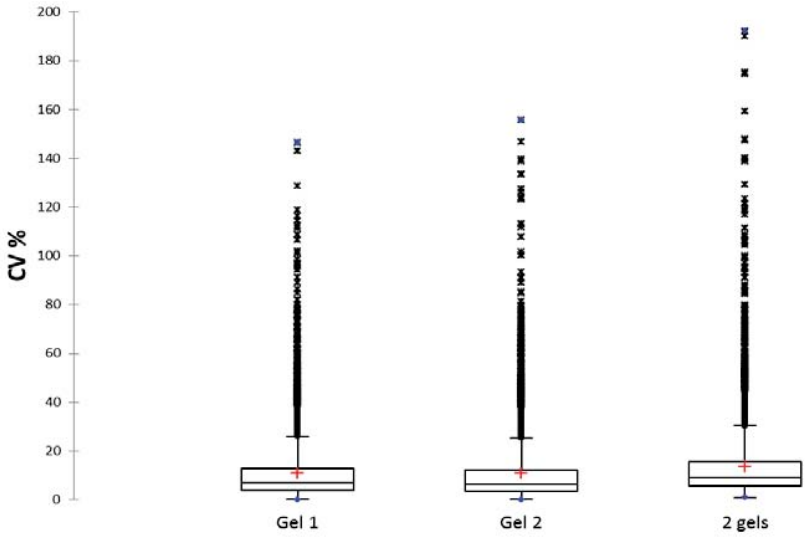
B



<u>Supplemental data 5</u>: Gel to gel repeatability

In order to evaluate gel to gel variations, different gels were prepared, loaded with the same total HUVEC lysate on 3 gel lanes, and migrated in parallel. Twelve gel bands were cut successively on the gels, following the same cutting pattern. Box-plots and table below illustrate the distribution of CVs obtained on protein PAI values when the comparison was performed on the 3 samples fractionated on gel1, the 3 samples fractionated on gel 2, or on the 6 samples fractionated on the 2 gels.



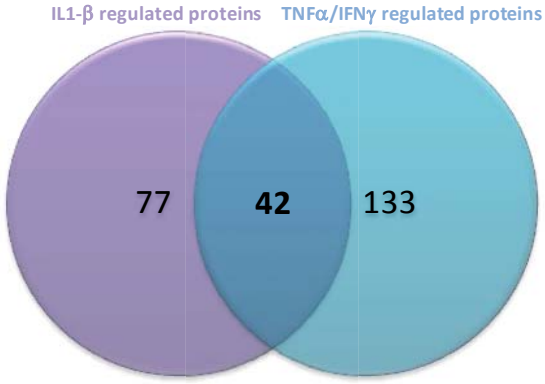


	Gel 1 (3X12 gel bands)	Gel 2 (3X12 gel bands)	2 gels (2 X 3X12 gel bands)
Total number of identified proteins	4439	4570	4857
Maximum CV	146.39	155.56	192.21
Median CV	6.87	6.27	8.99
99% CV distribution	70.68	71.24	78.88
95% CV distribution	33.20	34.97	39.80
75% CV distribution	12.71	12.18	15.48

<u>Supplemental data 9</u>: Proteins up-regulated by TNFα/IFNγ and IL1β

Venn diagram (A) and table (B) showing the proteins detected as up-regulated in both experiments. Known NF-kappa-B target genes are indicated.

A

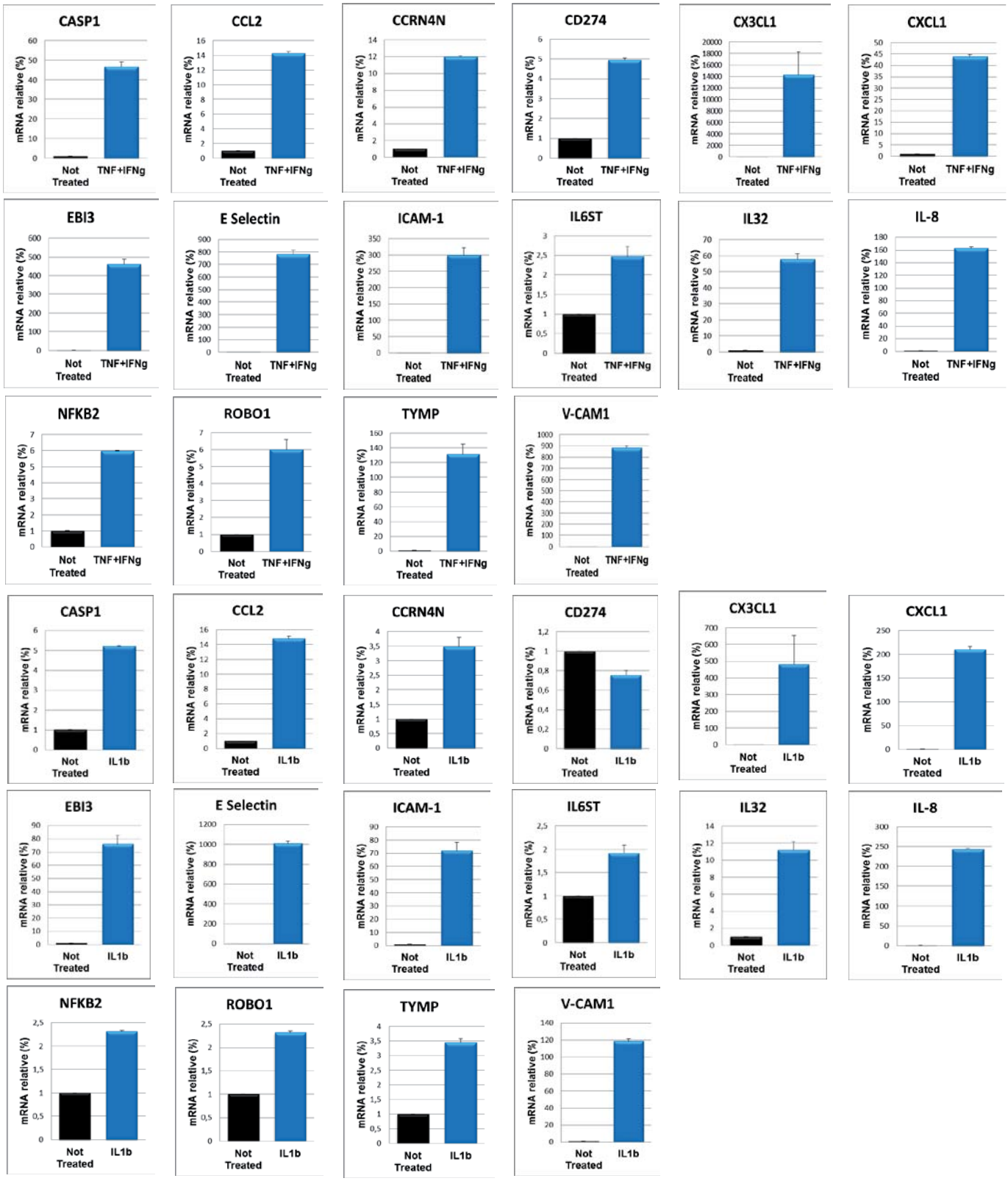


B

Protein AC	Protein ID	Protein Name	Gene Name	Known NF-kappa-B target gene
CSF1_HUMAN	P09603	Macrophage colony-stimulating factor 1 precursor	CSF1	yes
IL27B_HUMAN	Q14213	Interleukin-27 subunit beta precursor	EBI3	yes
ICOSL_HUMAN	O75144	ICOS ligand precursor	ICOSLG	
RELB_HUMAN	Q01201	Transcription factor RelB	RELB	yes
LYAM2_HUMAN	P16581	E-selectin precursor	SELE	yes
TNAP3_HUMAN	P21580	Tumor necrosis factor alpha-induced protein 3	TNFAIP3	yes
VCAM1_HUMAN	P19320	Vascular cell adhesion protein 1	VCAM1	yes
ICAM1_HUMAN	P05362	Intercellular adhesion molecule 1 precursor	ICAM1	yes
CATS_HUMAN	P25774	Cathepsin S precursor	CTSS	
ISG20_HUMAN	Q96AZ6	Interferon-stimulated gene 20 kDa protein	ISG20	
TNAP2_HUMAN	Q03169	Tumor necrosis factor alpha-induced protein 2	TNFAIP2	yes
APOL3_HUMAN	O95236	Apolipoprotein L3	APOL3	
IL32_HUMAN	P24001	Interleukin-32 precursor	IL32	
SODM_HUMAN	P04179	Superoxide dismutase [Mn], mitochondrial precursor	SOD2	yes
NOCT_HUMAN	Q9UK39	Nocturnin	CCRN4L	
TRAF1_HUMAN	Q13077	TNF receptor-associated factor 1	TRAF1	yes
JUNB_HUMAN	P17275	Transcription factor jun-B	JUNB	yes
C7FDR3_HUMAN	C7FDR3	MHC class I antigen (Fragment)	HLA-B	yes
APOL2_HUMAN	Q9BQE5	Apolipoprotein L2	APOL2	
SDC4_HUMAN	P31431	Syndecan-4 precursor	SDC4	yes
JAG1_HUMAN	P78504	Protein jagged-1 precursor	JAG1	
RIPK2_HUMAN	O43353	Receptor-interacting serine/threonine-protein kinase 2	RIPK2	yes
TPSN_HUMAN	O15533	Tapasin precursor	TAPBP	yes
IFIH1_HUMAN	Q9BYX4	Interferon-induced helicase C domain-containing protein 1	IFIH1	
B4E2H0_HUMAN	B4E2H0	cDNA FLJ58148, highly similar to Poly (ADP-ribose) polymerase 14 (EC 2.4.2.30) (Fragment)	PARP14	
CTHR1_HUMAN	Q96CG8	Collagen triple helix repeat-containing protein 1 precursor	CTHRC1	
PSB10_HUMAN	P40306	Proteasome subunit beta type-10 precursor	PSMB10	
TAP1_HUMAN	Q03518	Antigen peptide transporter 1	TAP1	yes
PGH2_HUMAN	P35354	Prostaglandin G/H synthase 2 precursor	PTGS2	yes
1A30_HUMAN	P16188	HLA class I histocompatibility antigen, A-30 alpha chain precursor	HLA-A	
DTX3L_HUMAN	Q8TDB6	E3 ubiquitin-protein ligase DTX3L	DTX3L	
IL8_HUMAN	P10145	Interleukin-8 precursor	IL8	yes
TFIP8_HUMAN	O95379	Tumor necrosis factor alpha-induced protein 8	TNFAIP8	
GROA_HUMAN	P09341	Growth-regulated alpha protein	CXCL1	yes
GTR6_HUMAN	Q9UGQ3	Solute carrier family 2, facilitated glucose transporter member 6	SLC2A6	
B6ETL5_HUMAN	B6ETL5	Pannexin 1	PANX1	
IL6RB_HUMAN	P40189	Interleukin-6 receptor subunit beta precursor	IL6ST	yes
CTR2_HUMAN	P52569	Low affinity cationic amino acid transporter 2	SLC7A2	
NFKB1_HUMAN	P19838	Nuclear factor NF-kappa-B p105 subunit	NFKB1	yes
MAFF_HUMAN	Q9ULX9	Transcription factor MafF	MAFF	
LEG9B_HUMAN	Q3B8N2	Galectin-9B	LGALS9B	
PAI1_HUMAN	P05121	Plasminogen activator inhibitor 1 precursor	SERPINE1	yes

Supplemental data 10: Relative mRNA expression of different genes regulated by TNFα/IFNγ or IL1β treatment in HUVEC cells.

Relative mRNA levels in HUVEC cells either not treated or treated with TNFα/IFNγ or IL1β were determined by qPCR. Primer sets were specific for genes that correspond to modulated proteins identified by the proteomic approach. Relative mRNA levels were calculated by normalizing the signals to those of actin. Results are shown as means with s.d. from 3 separate datapoints.



Partie III. Etude du rôle de l'interleukine-33 dans les cellules endothéliales par protéomique fonctionnelle

I. Contexte biologique

Les cytokines de la famille de l'interleukine 1 (IL-1) jouent un rôle majeur dans de nombreuses maladies inflammatoires, infectieuses et auto-immunes (Dinarello 2009; Dinarello 2011). Les cytokines IL-1α, IL-1β et IL-18, membres de cette famille, sont des cytokines hautement inflammatoires. La perturbation de leur activité ou de leur production peut engendrer des conséquences pathologiques sévères. L'interleukine 33 (IL-33) est un membre récent de cette famille impliqué dans plusieurs maladies (Miller 2011) comme l'arthrite rhumatoïde (Xu, Jiang et al. 2008), l'athérosclérose (Miller, Xu et al. 2008), l'asthme (Moffatt, Gut et al. 2010) et certaines maladies cardiovasculaires (Sanada, Hakuno et al. 2007). C'est une protéine de 31 KDa capable d'activer plusieurs types de cellules immunitaires (lymphocytes de type Th2, basophiles, mastocytes…) via son récepteur ST2 et son co-récepteur IL-1RAcP, qu'elle partage avec les autres membres de la famille IL-1 (Schmitz, Owyang et al. 2005). Dans ces cellules cibles, elle induit la production de cytokines inflammatoires et de chimiokines. Cette protéine, aussi désignée sous le nom de NF-HEV (Nuclear Factor from HEV), a été initialement identifiée dans le groupe de J-P Girard comme un facteur nucléaire présent dans le noyau des cellules endothéliales de veinules post-capillaires de type HEV (High endothelial Venules) (Baekkevold, Roussigne et al. 2003). En plus de sa forme sécrétée, l'IL-33 peut donc avoir une localisation nucléaire, et semble être une cytokine associée à la chromatine assurant un double rôle fonctionnel comme d'autres cytokines en sont capables, telles que l'IL-1α et HMGB1. L'IL-1α joue en effet à la fois un rôle de cytokine et d'activateur transcriptionnel (Werman, Werman-Venkert et al. 2004). Elle possède un signal de localisation nucléaire (NLS) au niveau de son domaine N-terminal qui permet l'adressage de sa forme précurseur non mature au noyau. Plusieurs facteurs nucléaires interagissant avec ce domaine N-terminal ont été mis en évidence, parmi lesquels on retrouve une histone acétyl-transférase (Buryskova, Pospisek et al. 2004). Une double fonction similaire a également été démontrée pour l'alarmine HMGB1 (« high-mobility group box 1 »). Cette protéine nucléaire se lie à la chromatine et affecte ainsi la régulation de la transcription en ouvrant l'accès à l'ADN aux complexes de remodelage de la chromatine et en facilitant l'interaction de certains facteurs de transcription avec l'ADN. Elle est par ailleurs capable d'induire une réponse inflammatoire lorsqu'elle est relarguée dans l'espace extracellulaire par des cellules endommagées ou bien sécrétée par certaines cellules immunitaires (Scaffidi, Misteli et al. 2002; Klune, Dhupar et al. 2008; Haraldsen, Balogh et al. 2009).

L'IL-33 semble de même faire partie de ces cytokines possédant des fonctions multiples. Elle serait en effet d'une part impliquée dans la modulation de la transcription, en se liant aux dimères

d'histones H2A-H2B à la surface des nucléosomes grâce à un motif hélice-boucle-hélice homeodomain-like situé à son extrémité N-terminale et agirait ainsi sur la compaction de la chromatine (Roussel, Erard et al. 2008). Cet environnement caractéristique de la répression transcriptionnelle laisse envisager un rôle de la cytokine dans ce mécanisme. Des expériences ont en effet montré que la protéine de fusion IL-33-Gal4-BD agit comme répresseur de la transcription d'un rapporteur Gal4-luciférase (Carriere, Roussel et al. 2007). Cependant, une étude très récente a au contraire mis en évidence que l'IL-33 intracellulaire agirait comme un activateur transcriptionnel du facteur de transcription NF-κB p65 et permettrait ainsi le contrôle de l'expression des molécules d'adhésion ICAM-1 et VCAM-1 présentes à la surface des cellules endothéliales (Choi, Park et al. 2012). La fonction intracellulaire de la cytokine nucléaire n'est ainsi pas clairement définie. D'autre part, l'IL-33 joue un rôle plus classique de cytokine qui, une fois libérée dans le milieu extracellulaire, déclenche une réponse pro-inflammatoire en ciblant via son récepteur ST2 différentes cellules immunitaires (Schmitz, Owyang et al. 2005; Iikura, Suto et al. 2007; Cherry, Yoon et al. 2008; Choi, Choi et al. 2009). Les mécanismes par lesquels l'IL-33 est sécrétée sont encore mal connus, mais des données récentes suggèrent qu'elle est libérée dans le milieu extracellulaire par les cellules endothéliales dans des conditions de nécrose ou de dommage cellulaire (Cayrol and Girard 2009). Elle est en effet exprimée de façon abondante dans le noyau des cellules endothéliales des vaisseaux sanguins dans la plupart des organes humains, mais également dans le noyau des cellules épithéliales au niveau des tissus exposés à l'environnement et aux pathogènes, tels que la peau ou l'épithélium intestinal (Moussion, Ortega et al. 2008). Il a été proposé que cette protéine pourrait avoir un rôle d'alarmine, semblable à celui de HMGB1, et constituer un signal de danger lors de sa libération par des cellules nécrosées au cours d'un traumatisme ou d'une infection, permettant ainsi d'alerter et d'activer le système immunitaire (Moussion, Ortega et al. 2008).

Les processus de maturation et la nature des formes actives d'IL-33 restent encore mal connus. Initialement, il avait été décrit que l'activation de la protéine nécessitait un clivage par la caspase-1, comme dans le cas d'IL-1β et IL-18, libérant une forme mature correspondant au domaine C-terminal de type IL-1 (a.a 112-270) (Schmitz, Owyang et al. 2005). Il a cependant été récemment démontré que l'IL-33 pleine taille est biologiquement active et ne nécessite pas de maturation pour lier et activer son récepteur ST2 (Ali, Nguyen et al. ; Cayrol and Girard 2009; Luthi, Cullen et al. 2009; Talabot-Ayer, Lamacchia et al. 2009). A l'inverse, le clivage par la caspase 1 et par les caspases apoptotiques 3 et 7 entraîne l'inactivation de cette cytokine (Ali, Nguyen et al. ; Cayrol and Girard 2009; Luthi, Cullen et al. 2009). La protéine pourrait néanmoins être clivée après sa libération dans le milieu extracellulaire, par des protéases autres que des caspases et notamment par des protéases sécrétées en conditions pro-inflammatoires. Cette hypothèse est à la base des études sur la maturation d'IL-33, auxquelles j'ai participé en assurant l'analyse des formes clivées de la protéine qui sont décrites dans ce chapitre.

Les cellules endothéliales constituent la source majeure de l'IL-33 dans les tissus humains. La cytokine est donc produite de façon constitutive par ces cellules, mais elle semble également agir sur celles-ci. En effet, bien que l'action de l'IL-33 via l'activation de son récepteur ST2 ait été principalement décrite dans les cellules du système immunitaire inné comme les lymphocytes Th2, le récepteur ST2 est également fortement exprimé à la surface des cellules endothéliales (Kumar, Tzimas et al. 1997; Aoki, Hayakawa et al. 2010). Il a été récemment montré que ces cellules répondent à la stimulation par l'IL-33 en augmentant leur production d'IL-6, d'IL-8 et d'oxyde nitrique, et des molécules d'adhésion ICAM-1, VCAM-1 suggérant un rôle de l'IL-33 dans leur

physiologie (Choi, Choi et al. 2009; Demyanets, Konya et al. 2011). Ainsi, l'activation paracrine des cellules endothéliales par l'IL-33 relarguée des cellules endothéliales endommagées pourrait jouer un rôle important dans l'inflammation, l'angiogénèse et la re-endothélialisation des vaisseaux sanguins lésés. Les effets globaux de l'IL-33 sur ces cellules n'ont cependant pas été étudiés.

L'IL-33 semble ainsi jouer des rôles cellulaires divers et importants. Elle est par ailleurs impliquée dans plusieurs maladies et a donc été proposée comme cible thérapeutique potentielle. Les fonctions et mécanismes d'action de l'IL-33 sont malgré tout peu connus et doivent être caractérisés avant d'envisager des applications thérapeutiques.

Dans ce contexte, ce projet réalisé en collaboration avec l'équipe de Biologie Vasculaire de l'IPBS a pour objectif d'avancer dans la compréhension des fonctions et des mécanismes d'action de l'IL-33 au sein des cellules endothéliales humaines. Cette cytokine semble jouer un double rôle, à la fois intracellulaire en tant que facteur nucléaire pour réguler la transcription, et extracellulaire comme cytokine agissant sur le récepteur ST2 de cellules cibles. Afin de caractériser au mieux ses fonctions, nous avons étudié ces deux aspects. Les fonctions nucléaires de l'IL-33 dans les cellules endothéliales ont dans un premier temps été analysées en quantifiant les variations protéiques induites par l'extinction de l'expression de la protéine endogène par knock-down avec des siRNA. Dans un second temps, nous nous sommes intéressés aux effets de la cytokine exogène. Nous avons tout d'abord cherché à caractériser par spectrométrie de masse la nature de formes clivées de la protéine présentant une activité supérieure à celle de la protéine pleine taille. Par la suite, des protéines recombinantes correspondant à ces formes super-actives ont été produites, et nous avons analysé leurs effets sur les cellules endothéliales, en mesurant les modulations d'expression des protéines en appliquant de la stratégie de protéomique quantitative sans marquage précédemment développée.

II. Caractérisation fonctionnelle de l'IL-33 dans les cellules endothéliales

II-1. Etude du rôle intracellulaire de l'IL-33 endogène

Afin de comprendre plus précisément le rôle intracellulaire de la cytokine nucléaire, une étude à large échelle des protéines surexprimées et sous-exprimées suite à l'extinction de l'expression de l'IL-33 a été réalisée. Nous avons pour cela transfecté environ 8.10^6 HUVEC avec des siRNA ciblant l'IL-33 endogène (condition essai, 3 réplicats réalisés à partir de boîtes de cellules indépendantes) ou avec des siRNA contrôle (condition contrôle, 3 réplicats indépendants) et réalisé une analyse quantitative afin de mettre en avant les variations protéiques obtenues (Figure 56). Deux expériences indépendantes (et donc deux analyses quantitatives) ont été effectuées avec des pools de siRNA différents pour éviter l'obtention de faux-positifs dus à des effets non spécifiques éventuels des siRNA. L'extinction de l'expression de l'IL-33 a été vérifiée au niveau des ARNm et au niveau protéique par Q-PCR et western blot (Figure 55).

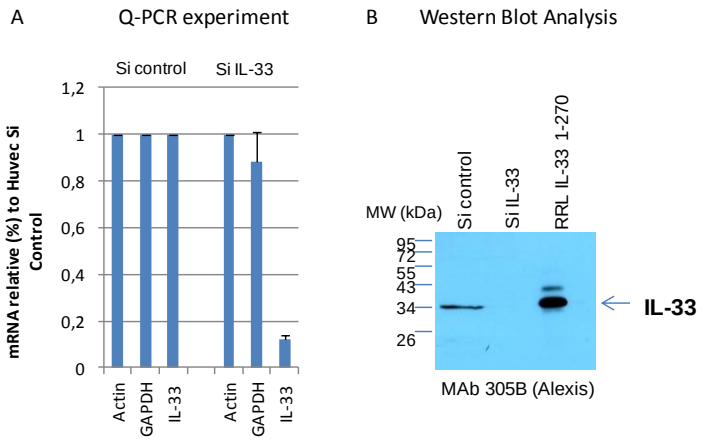
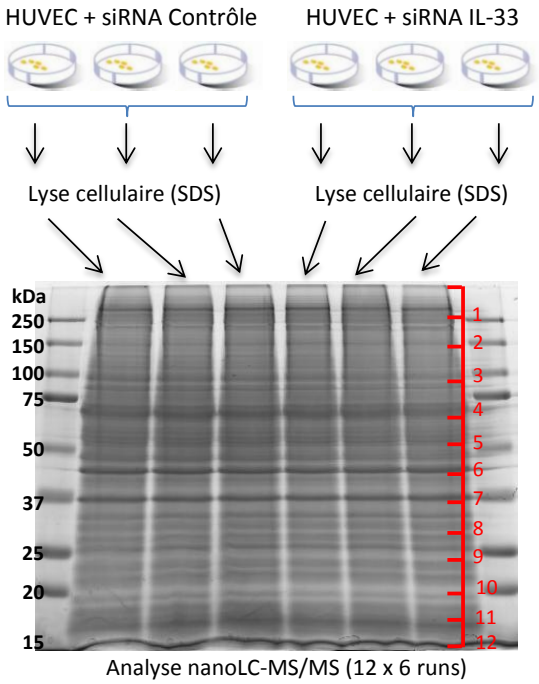


Figure 55 : Extinction de l'expression de l'IL-33 dans les cellules endothéliales primaires traitées siRNA IL-33. *Cette extinction a été vérifiée au niveau ARNm (Q-PCR) (A) et au niveau protéique (western blot anti-IL-33) (B). Les niveaux d'ARNm de protéines contrôle (Actine, GAPDH) ont été testés dans les deux conditions siRNA IL-33 et siRNA CT.*

L'analyse des deux expériences a permi l'identification et la quantification de 5615 et de 5230 protéines. Les protéines ont été décrites comme variantes si elles respectent les deux critères suivants : une p-value du test de Student inférieure à 0,05 et un ratio siRNA IL33/siRNA CT supérieur à 2. Ainsi dans la première expérience, seules 29 protéines sur les 5615 présentent une variation d'expression significative entre les deux conditions, dont 17 protéines apparaissent surexprimées et 12 sous-exprimées (Figure 56). Dans la seconde, 120 protéines sont considérées variantes (56 surexprimées, 64 sous-exprimées). Parmi ces protéines, très peu présentent de fortes variations comme on peut le visualiser sur le « volcano plot » représentant l'analyse quantitative de l'expérience 1 (Figure 57) et la majorité d'entre elles est quantifiée avec un ratio proche de la valeur seuil de 2. Par ailleurs, la plupart de ces protéines est identifiée et quantifiée à partir d'un seul peptide de faible score, donnant généralement un signal de faible intensité, proche du bruit de fond.

L'extinction de l'IL-33 intracellulaire ne semble ainsi pas avoir d'effets drastiques et nets dans ces expériences. Nous avons comparé les protéines désignées comme variantes dans les deux expériences. Toutes ne sont pas identifiées dans les deux cas, et c'est en particulier vrai pour les protéines faiblement abondantes quantifiées avec un seul peptide. Au final, aucune protéine considérée comme variante n'a été retrouvée en commun entre les deux expériences. Cela peut s'expliquer en partie par l'imprécision de la méthode au voisinage des valeurs seuil (ratio>2 et p-value<0.05), liée aux erreurs de quantification et à une mauvaise extraction du signal par le logiciel. Clairement, certains variants « faux positifs » sont présents à la suite de l'analyse quantitative, avec les critères de validation peu stringents choisis, et des optimisations du traitement bioinformatique devront être apportés pour éliminer automatiquement les signaux de mauvaise qualité. Par ailleurs, les variations de faible amplitude identifiées dans chaque expérience peuvent également être la conséquence d'effets non spécifiques des siRNA.



	Expérience 1						Expérience 2					
	siCT A	siCT B	siCT C	siIL33 A	siIL33 B	siIL33 C	siCT A	siCT B	siCT C	siIL33 A	siIL33 B	siIL33 C
Nombre de protéines identifiées par piste	4948	4897	5042	4925	4715	4877	4280	4183	4113	4215	4134	4307
Nombre total de protéines quantifiées	5615						5230					
Protéines variantes (Fold > 2 et p-value < 0,05)	29						120					

Figure 56 : Stratégie générale et résultats de l'étude protéomique à large échelle des cellules endothéliales HUVEC traitées ou non pour l'extinction de l'expression de l'IL-33 (par siRNA). *Trois réplicats biologiques indépendants ont été réalisés pour chaque condition. L'extrait protéique total des cellules obtenu par lyse des cellules au SDS puis sonication est déposé sur gel 1D SDS-PAGE. Chacune des 6 pistes est fractionnée en 12 bandes. Les extraits peptidiques issus de chaque bande de gel est analysé par nanoLC-MS/MS. Une analyse quantitative est ensuite réalisée par MFPaQ.*

Ainsi, l'inhibition de l'expression de l'IL-33 nucléaire ne conduit pas à une modification de l'expression des protéines dans les conditions expérimentales dans lesquelles nous nous sommes placés. Or, l'IL-33 a été récemment décrite comme activateur de la transcription de la protéine NF-κB p65 (Choi, Park et al. 2012) et dans cette étude, l'extinction de la cytokine (par siRNA) engendre une diminution de l'expression de NF-κB p65. Dans nos deux expériences, le facteur de transcription p65 a bien été identifié, et clairement quantifié avec un ratio d'expression de 1 à partir de plusieurs peptides dans les deux expériences. L'extinction de la cytokine, bien qu'elle semble pourtant efficace (10 fois moins d'ARNm dans l'essai), n'est peut-être pas suffisamment importante pour nous permettre de détecter les variations induites dans le protéome des cellules endothéliales. Dans cette étude, nous n'avons ainsi pas pu mettre en évidence un rôle intracellulaire de l'IL-33.

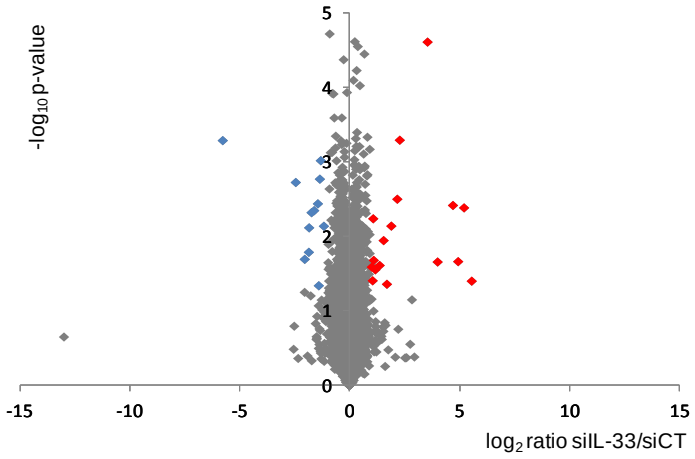


Figure 57 : Représentation graphique de l'analyse quantitative du protéome des cellules endothéliales après extinction de l'expression de l'IL-33. *Volcano plot représentant la significativité des variations d'expression des protéines (p-value du test de student) en fonction des ratios d'expression des protéines entre l'essai et le contrôle. Les points rouge indiquent les protéines surexprimées, les points bleus les protéines sous-exprimées et les points gris les protéines non variantes.*

II-2. Etude du rôle extracellulaire de l'IL-33

Le processus de maturation et la nature des formes actives de l'IL-33 extracellulaire sont encore mal connus. Il a été montré que la forme pleine taille est biologiquement active, et que contrairement à d'autres cytokines de la famille IL-1, comme l'IL-1β, le clivage par les caspases entraîne l'inactivation d'IL-33 plutôt que son activation. Dans la mesure où la protéine est libérée dans le milieu extracellulaire à partir de cellules endommagées ou nécrosées, il a été proposé qu'elle pourrait jouer un rôle d'alarmine ou signal de danger endogène, afin d'alerter le système immunitaire lors d'un traumatisme ou d'une infection. Les neutrophiles sont recrutés rapidement au niveau des tissus lésés pendant la phase inflammatoire, et après activation, ils libèrent dans l'espace extracellulaire des granules sécrétoires contenant diverses protéases à sérine. Il a donc été envisagé par l'équipe de J-P Girard que d'autres formes bioactives de la protéine pourraient exister, issues du clivage par des protéases autres que des caspases et notamment par des protéases sécrétées en conditions pro-inflammatoires, jouant un rôle clé dans la régulation de l'inflammation. Nous avons donc dans un premier temps étudié la maturation et la régulation de cette cytokine par des protéases de neutrophiles, l'élastase et la cathepsine G. Nous avons ensuite cherché à caractériser à large échelle l'effet de l'IL-33 maturée sur les cellules endothéliales afin d'avancer dans la compréhension de sa fonction.

II-2.1 Identification des formes maturées bioactives de l'IL-33

L'équipe de Biologie Vasculaire a mis en évidence que la forme pleine taille de l'IL-33 (1-270) est biologiquement active et entraine une activation de la sécrétion de l'IL-6 par des cellules immunitaires (mastocytes). Cette sécrétion est cependant significativement augmentée lorsque l'IL-33 pleine taille recombinante (produite *in vitro* à partir de lysat de réticulocytes de lapin) est préalablement incubée avec des protéases à sérine de neutrophiles, l'élastase et la cathepsine G

(Figure 58A). Ces résultats laissent donc envisager une maturation et une régulation de l'activité de l'IL-33 par ces protéases. Des produits de clivage d'environ 20 kDa ont en effet été identifiés par western blot (Figure 58B) : la digestion de l'IL-33 1-270 par la cathepsine G génère deux fragments C-terminaux visibles sur blot (détectés à l'aide d'anticorps dirigés contre la partie C-terminale de la protéine), et la digestion par l'élastase génère un seul fragment de taille intermédiaire.

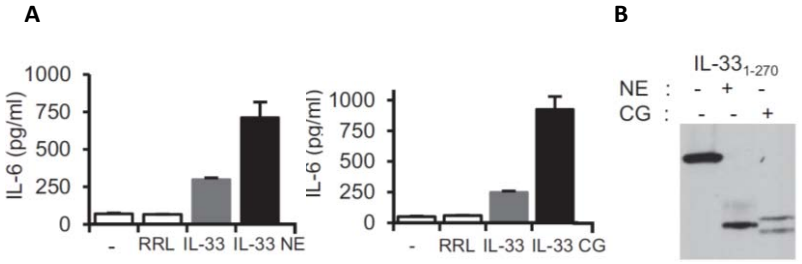


Figure 58 : Mise en évidence de l'existence de formes maturées très actives de l'IL-33. *(A) Augmentation de l'activité de l'IL-33 après incubation avec des protéases de neutrophiles (élastase, cathepsine G). L'activité de l'IL-33 est suivie par mesure de la sécrétion de l'IL-6 par des mastocytes (ELISA) suite à la stimulation par l'IL-33 incubé ou non avec les protéases. (B) Formes clivées issues du clivage de l'IL-33 pleine taille (produite in vitro à partir de lysat de réticulocytes de lapin) par les protéases de neutrophiles, élastase et cathepsine G visualisées par western blot à l'aide d'un anticorps dirigé contre la partie C-terminale de l'IL-33. NE : neutrophile elastase, CG : cathepsine G.*

Nous avons donc en collaboration avec l'équipe de J-P Girard, voulu localiser les sites de clivage de ces deux protéases sur l'IL-33 et ainsi déterminer les formes maturées de la cytokine. D'un point de vue méthodologique, cela revient à caractériser par spectrométrie de masse l'extrémité N-terminale des fragments formés après maturation *in vitro* de la protéine recombinante par les protéases de neutrophiles. L'analyse de ces fragments ainsi que de la forme entière a été faite par une approche classique de type « bottom-up », nécessitant de générer des peptides pour l'analyse MS. Il a donc été nécessaire pour cette étude 1) de définir un système d'expression permettant d'obtenir en quantité suffisante la protéine recombinante et ses formes maturées *in vitro*, afin d'atteindre une bonne couverture de séquence et de pouvoir détecter le peptide N-terminal et 2) de réaliser la digestion des formes protéiques avec différentes enzymes (trypsine et endoprotéase Glu-C), pour arriver à générer, en fonction des formes analysées, un peptide N-terminal de taille adéquate pour l'analyse MS.

Des premières mesures ont été réalisées à partir des formes recombinantes initialement produites à partir de lysat de réticulocytes de lapin. Cependant, cette analyse s'est révélée très difficile en raison de la faible abondance de l'IL-33 produite et par conséquent de ses formes maturées, et de la contamination importante par des protéines de lapin. La protéine IL-33 pleine taille recombinante a donc ensuite été exprimée chez *E. coli* en fusion avec la GST en N-terminal, permettant l'obtention d'une plus grande quantité de matériel de départ. Elle a ensuite été purifiée sur billes glutathion-sépharose. L'IL-33 pleine taille a été digérée directement sur billes par l'élastase ou la cathepsine G, et les produits de clivage ainsi que la forme non clivée ont été séparés sur gel 1D SDS-PAGE (Figure 59).

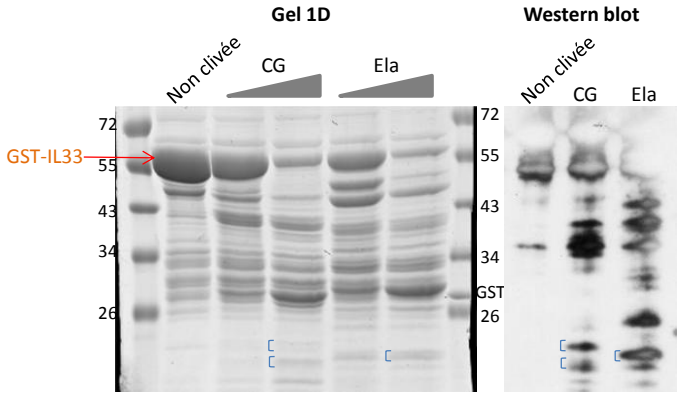


Figure 59 : Gel 1D SDS-PAGE et western blot correspondant de l'IL-33 pleine taille (GST-IL-33) et de l'IL-33 clivée par l'élastase et la cathepsine G*. Les bandes correspondant à l'IL-33 pleine taille et aux fragments générés par le clivage de l'élastase et de la cathepsine G ont été découpées pour être analysées en nanoLC-MS/MS. Western blot : révélé avec un anticorps dirigé contre la partie C-terminal de l'IL-33.*

Comme le montre le profil de coloration du gel 1D, l'IL-33 recombinante apparaît comme la bande majoritaire, grâce à l'étape d'enrichissement. Cependant, même dans la piste correspondant au contrôle non maturé, on détecte à des poids moléculaires inférieurs une série de bandes correspondant à des formes dégradées de la protéine, à des produits de clivage de celle-ci par des protéases d'*E.coli*, ou à des contaminants. Après analyse par western-blot à l'aide d'anticorps dirigés contre le C-ter d'IL-33, on détecte bien néanmoins les fragments C-terminaux générés par le clivage cathepsine G ou élastase. Les bandes correspondant à IL-33 pleine taille du contrôle et aux fragments spécifiques des deux protéases ont été découpées, digérées avec la trypsine ou la GluC, et les mélanges peptidiques générés ont été analysés en nanoLC-MS/MS. Les recherches en bases de données ont été effectuées en autorisant un clivage peptidique semi-spécifique par rapport à la spécificité de l'enzyme considérée (trypsine ou Glu-C), de façon à permettre pour les formes maturées l'identification de néo-extrémités N-teminales non définies dans les banques. Pour identifier les nouvelles extrémités N-terminales des fragments maturés, nous avons comparé systématiquement les peptides détectés dans la forme entière ou dans les fragments, par le biais d'une analyse quantitative différentielle. Les résultats de ces comparaisons, pour chacun des fragments, sont représentés figure 60, dans des graphes indiquant, pour tous les peptides identifiés, leur intensité dans la forme maturée par rapport à leur intensité dans la forme entière.

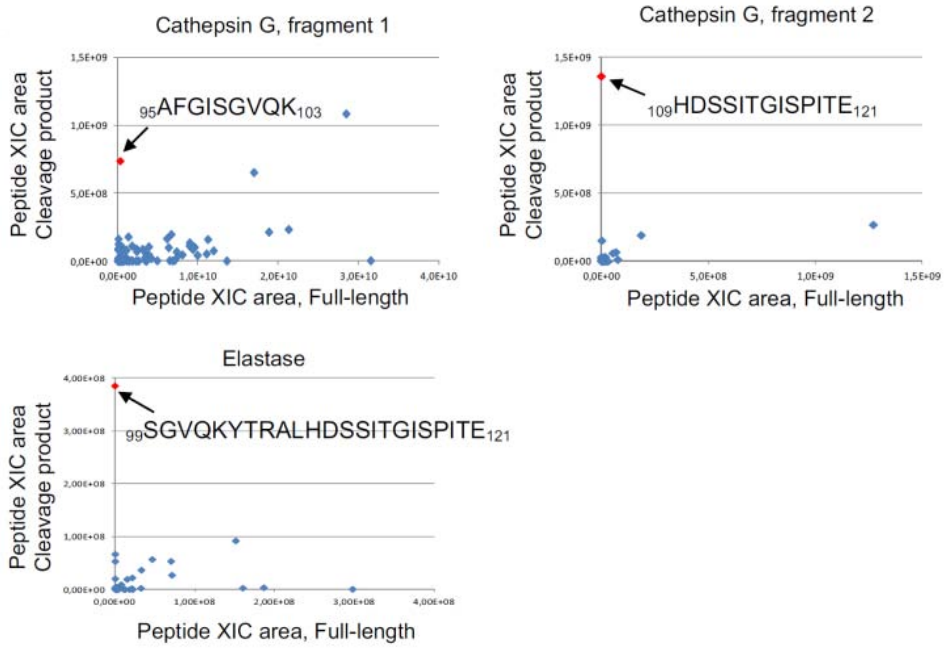


Figure 60 : Analyse protéomique des fragments issus du clivage de l'IL-33 par l'élastase et la cathepsine G. *Plot des intensités des peptides des formes clivées en fonction des intensités des peptides issus de l'IL-33 pleine taille. Les peptides générés par la digestion enzymatique à la trypsine ou à la GluC des produits de clivage ou de la forme pleine taille de l'IL-33 ont été analysés nanoLC-MS/MS et quantifiés avec MFPaQ. Les peptides intenses spécifiques des formes clivées observés seulement pour les formes maturées correspondant aux peptides N-terminaux et caractérisant le site de clivage sont représentés en rouge.*

Cette analyse nous a permis d'identifier des peptides communs aux formes entière et clivée, qui sont situés dans la région C-terminale de la protéine. En revanche, les peptides N-terminaux de la protéine entière sont spécifiques de la pleine taille et apparaissent sur l'axe des abscisses. Seul un peptide (point rouge situé sur l'axe des ordonnées sur la représentation graphique) a été identifié spécifiquement dans les formes C-terminales, correspondant à la nouvelle extrémité N-terminale de l'IL-33 maturée. Il correspond au peptide 95-AFGISGVQK-103 pour le premier fragment de la cathepsine G (obtenu après digestion trypsique), au peptide 109-HDSSITGISPITE-121 pour le second (obtenu après digestion GluC du fragment) et au peptide 99-SGVQKYTRALHDSSITGISPITE-121 pour le fragment généré par le clivage avec l'élastase (obtenu par digestion GluC). Ces différents peptides ont été séquencés plusieurs fois uniquement dans l'échantillon correspondant au fragment considéré, et pas pour l'IL-33 pleine taille, et constituent bien des peptides spécifiques des formes maturées (Table 8) En revanche, les longs peptides spécifiques de l'IL-33 1-270 couvrant le site de clivage n'ont pas été systématiquement identifiés lors de l'analyse de la forme entière, sans doute à cause d'une faible ionisation et/ou d'une mauvaise extraction du gel dues à leur taille importante.

Table 8 : Liste des peptides trypsiques ou Glu-C caractérisant les 2 sites de clivage de la cathepsine G, le site de clivage de l'élastase et la forme entière de l'IL-33 (*Nb MS/MS : nombre de MS/MS réalisées sur le peptide correspondant****).***

Protéase	Fragment IL-33	Peptide de l'IL-33	Séquence du peptide IL-33	Nb MS/MS
Cathepsin G, site 1	IL-33 pleine taille	Peptide 79-103	HLVLAACQQQSTVECFAFGISGVQK	80
		Peptide 95-103	AFGISGVQK	0
	Fragment 1 IL-33 après clivage par cathepsine G	Peptide 79-103	HLVLAACQQQSTVECFAFGISGVQK	2
		Peptide 95-103	AFGISGVQK	51
Cathepsin G, site 2	IL-33 pleine taille	Peptide 93-121	CFAFGISGVQKYTRALHD	0
		Peptide 109-121	HDSSITGISPITE	0
	Fragment 2 IL-33 après clivage par cathepsine G	Peptide 93-121	CFAFGISGVQKYTRALHD	0
		Peptide 109-121	HDSSITGISPITE	34
Elastase	IL-33 pleine taille	Peptide 93-121	CFAFGISGVQKYTRALHDSSITGISPITE	0
		Peptide 99-121	SGVQKYTRALHDSSITGISPITE	0
	Fragment IL-33 après clivage par élastase	Peptide 93-121	CFAFGISGVQKYTRALHDSSITGISPITE	0
		Peptide 99-121	SGVQKYTRALHDSSITGISPITE	11

Nous avons ainsi dans cette étude identifié plusieurs formes maturées de l'IL-33 : l'IL-$33_{95\text{-}270}$ et IL-$33_{109\text{-}270}$ issues du clivage par la cathepsine G aux sites F94 et L108, et le fragment IL-$33_{98\text{-}270}$ issu du clivage de l'IL-33 pleine taille au niveau de l'I98 par l'élastase (Figure 61).

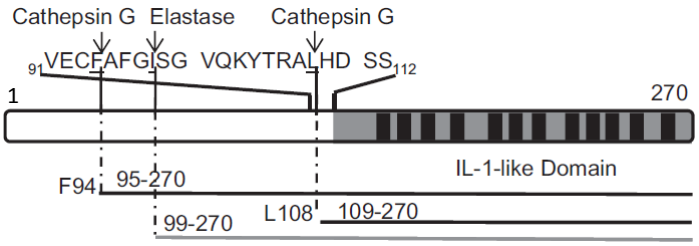


Figure 61: Structure primaire de l'IL-33 et sites de clivage identifiés de la cathepsine G et de l'élastase. *Le domaine IL-1-like est représenté avec les 12 brins β (en noir). Les sites de clivage de la cathepsine G et de l'élastase ainsi que les formes maturées résultantes (IL-*$33_{95\text{-}270}$*, IL-*$33_{99\text{-}270}$*, IL-*$33_{109\text{-}270}$*) sont représentées.*

Des expériences de mutagénèse dirigée réalisées par l'équipe de Biologie Vasculaire ont permis de confirmer les sites de clivage identifiés (voir publication jointe ci-dessous). Des expériences complémentaires réalisées dans cette équipe ont par ailleurs montré que ces trois formes maturées sont également générées par des neutrophiles humains activés *ex vivo*. Elles sont biologiquement actives *in vivo* chez la souris puisque leur injection entraine en particulier une augmentation de la concentration d'IL-5 dans le sérum. Des expériences supplémentaires ont également permis de mettre en évidence un clivage majoritaire de l'IL-33 murine (mIL-$33_{102\text{-}266}$) généré *in vivo*. Cette forme est en effet retrouvée dans le fluide broncho-alvéolaire de souris modèles qui présentent des lésions pulmonaires aiguës, entrainant un dommage de l'épithélium alvéolaire et un recrutement de neutrophiles au niveau de la paroi alvéolaire. Ces différents résultats indiquent donc que la maturation de l'IL-33 par des protéases à sérine de neutrophiles permet *in vivo* la génération de formes hyper actives de la cytokine au niveau extracellulaire.

Article *Lefrançais, Roga, Gautier et al., PNAS, 2012*

« IL-33 is processed into mature bioactive forms by neutrophil elastase and Cathepsin G »

Emma Lefrançais, Stephane Roga, Violette Gautier, Anne Gonzalez-de-Peredo, Bernard Monsarrat, Jean-Philippe Girard, Corinne Cayrol

Proc Natl Acad Sci U S A. 2012 Jan 31;109(5):1673-8.

IL-33 is processed into mature bioactive forms by neutrophil elastase and cathepsin G

Emma Lefrançais, Stephane Roga, Violette Gautier, Anne Gonzalez-de-Peredo, Bernard Monsarrat, Jean-Philippe Girard[1,2], and Corinne Cayrol[1,2]

Centre National de la Recherche Scientifique, Institut de Pharmacologie et de Biologie Structurale, F-31077 Toulouse, France; Université de Toulouse, Université Paul Sabatier, Institut de Pharmacologie et de Biologie Structurale, F-31077 Toulouse, France

Edited* by Charles A. Dinarello, University of Colorado Denver, Aurora, CO, and approved December 19, 2011 (received for review October 3, 2011)

Interleukin-33 (IL-33) (NF-HEV) is a chromatin-associated nuclear cytokine from the IL-1 family, which has been linked to important diseases, including asthma, rheumatoid arthritis, ulcerative colitis, and cardiovascular diseases. IL-33 signals through the ST2 receptor and drives cytokine production in type 2 innate lymphoid cells (ILCs) (natural helper cells, nuocytes), T-helper (Th)2 lymphocytes, mast cells, basophils, eosinophils, invariant natural killer T (iNKT), and natural killer (NK) cells. We and others recently reported that, unlike IL-1β and IL-18, full-length IL-33 is biologically active independently of caspase-1 cleavage and that processing by caspases results in IL-33 inactivation. We suggested that IL-33, which is released upon cellular damage, may function as an endogenous danger signal or alarmin, similar to IL-1α or high-mobility group box 1 protein (HMGB1). Here, we investigated the possibility that IL-33 activity may be regulated by proteases released during inflammation. Using a combination of in vitro and in vivo approaches, we demonstrate that neutrophil serine proteases cathepsin G and elastase can cleave full-length human IL-$33_{1–270}$ and generate mature forms IL-$33_{95–270}$, IL-$33_{99–270}$, and IL-$33_{109–270}$. These forms are produced by activated human neutrophils ex vivo, are biologically active in vivo, and have a ∼10-fold higher activity than full-length IL-33 in cellular assays. Murine IL-33 is also cleaved by neutrophil cathepsin G and elastase, and both full-length and cleaved endogenous IL-33 could be detected in the bronchoalveolar lavage fluid in an in vivo model of acute lung injury associated with neutrophil infiltration. We propose that the inflammatory microenvironment may exacerbate disease-associated functions of IL-33 through the generation of highly active mature forms.

innate immunity | inflammatory protease | serine protease inhibitor | alveolar epithelium

Cytokines of the IL-1 family (IL-1α, IL-1β, IL-18) play a major role in inflammatory, infectious, and autoimmune diseases (1–3). IL-33 [previously known as nuclear factor from high endothelial venule or NF-HEV (4, 5)], is a chromatin-associated nuclear cytokine from the IL-1 family (6, 7), which has been linked to important diseases (8–10), including asthma (11), rheumatoid arthritis (12, 13), ulcerative colitis (14), and cardiovascular diseases (15).

IL-33 signals through the ST2 receptor (4), a member of the IL-1 receptor family, which is expressed (or induced) on various immune cell types, including mast cells, basophils, eosinophils, T-helper (Th)2 lymphocytes, invariant natural killer T (iNKT) and natural killer (NK) cells, macrophages, dendritic cells, and neutrophils (8–10). IL-33 stimulation of ST2 on Th2 cells induces secretion of the Th2 cytokines IL-5 and IL-13 (4, 16). Recently, IL-33 has been shown to drive production of extremely high amounts of these Th2 cytokines by type 2 innate lymphoid cells (ILCs) (natural helper cells, nuocytes, innate helper 2 cells), which play important roles in innate immune responses, after helminth infection in the intestine (17–19) or influenza virus infection in the lungs (20, 21). An important role of IL-33 in innate rather than acquired immunity is also supported by observations in IL-33-deficient mice (22).

It was initially believed that, like IL-1β and IL-18, processing of IL-33 by caspase-1 to a mature form was required for biological activity (4). However, we (23) and others (24–26) demonstrated that full-length IL-33 is biologically active and that processing of IL-33 by caspases results in its inactivation, rather than its activation. Further analyses revealed that IL-33 is constitutively expressed to high levels in the nuclei of endothelial and epithelial cells in vivo (27) and that it can be released in the extracellular space after cellular damage (23, 24). IL-33 was, thus, proposed (23, 24, 27) to function as an endogenous danger signal or alarmin, similar to IL-1α and high-mobility group box 1 protein (HMGB1) (28–32), to alert cells of the innate immune system of tissue damage during trauma or infection.

An important question that has not yet been resolved is whether full-length IL-33 is the unique bioactive form of the cytokine or whether mature bioactive forms are also generated in vivo. Because IL-33 plays important roles in inflammatory diseases, we hypothesized that IL-33 could be cleaved by proteases released from innate effector cells during inflammation. In this report, we show, by using both in vitro and in vivo approaches, that IL-33 is processed into mature bioactive forms by neutrophil elastase and cathepsin G. Interestingly, the IL-33 mature forms generated by neutrophil serine proteases, IL-$33_{95–270}$, IL-$33_{99–270}$, and IL-$33_{109–270}$, have an increased biological activity compared with the full-length IL-$33_{1–270}$ protein. Proteolytic processing may, thus, play an important role in the regulation of IL-33 activity during inflammation.

Results

Biological Activity of Full-Length IL-33 Is Increased by Neutrophil Elastase and Cathepsin G. In situ processing of full-length IL-33 has been observed in some biological assays (25), and it has been questioned whether the full-length IL-33 form itself or one of its cleavage products represented the truly bioactive species. To address this issue, we tested the biological activity of full-length IL-$33_{1–270}$, produced in three different expression systems, using a cellular bioassay validated in our previous studies, IL-33-dependent secretion of IL-6 by MC/9 mast cells (23). As shown in Fig. 1*A*, full-length IL-33 proteins produced in rabbit reticulocyte lysate (RRL), wheat germ extract (WGE), or human in vitro protein expression system (HES) induced IL-6 secretion in mast cells. We then determined the size of the IL-33 proteins at the beginning and at the end of the assay (24-h incubation) by Western blot analysis (Fig. 1*B*). Whatever the expression system used, we found no evidence for IL-33 processing during the bioassay. We concluded that full-length IL-33 does not require processing for biological activity.

Author contributions: J.-P.G. and C.C. designed research; E.L., S.R., V.G., and C.C. performed research; E.L., V.G., A.G.-d.-P., B.M., J.-P.G., and C.C. analyzed data; and J.-P.G. and C.C. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]J.-P.G. and C.C. share senior authorship.

[2]To whom correspondence may be addressed. E-mail: jean-philippe.girard@ipbs.fr or corinne.cayrol@ipbs.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1115884109/-/DCSupplemental.

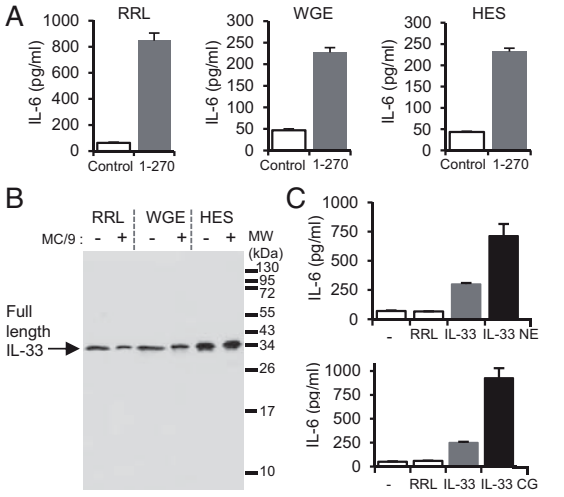


Fig. 1. The biological activity of full-length IL-$33_{1–270}$ is increased by neutrophil elastase and cathepsin G. (*A*) Capacity of full-length IL-$33_{1–270}$ produced in RRL, WGE, or HES (5 μL of lysate) to activate the IL-33-responsive mast cell line MC/9 (2×10^5 cells/well) was analyzed by determining IL-6 levels in supernatants using an ELISA. (*B*) Western blot analysis (305B mAb) of full-length IL-$33_{1–270}$ produced in RRL, WGE, or HES before or after incubation with MC/9 cells for 24 h. (*C*) Full-length IL-$33_{1–270}$ produced in RRL (5 μL of lysate) was preincubated with neutrophil elastase (NE) (30 mU/μL; 30 min at 37 °C) or cathepsin G (CG) (0.09 mU/μL; 1 h at 37 °C) before incubation with MC/9 cells (10^5 cells/well) for 24 h. IL-6 levels in supernatants were detected by ELISA. Results in *A* and *C* are shown as means and SD of three separate data points and are representative of two independent experiments.

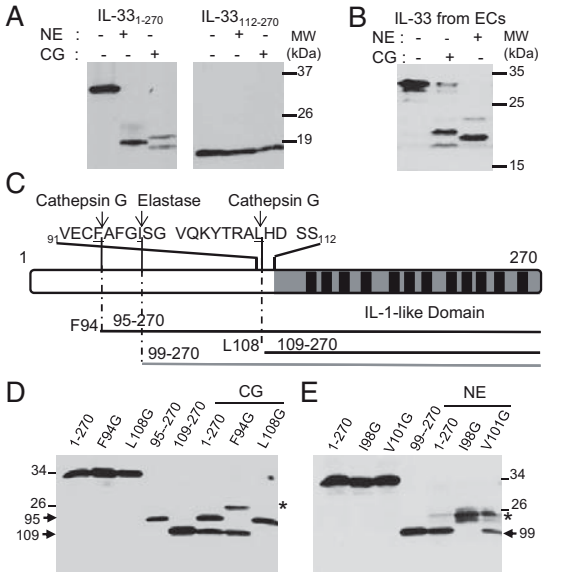


Fig. 2. Neutrophil elastase and cathepsin G generate IL-33 mature forms IL-$33_{95–270}$, IL-$33_{99–270}$, and IL-$33_{109–270}$. (*A* and *B*) IL-33 is a substrate for neutrophil elastase and cathepsin G. In vitro-translated IL-$33_{1–270}$ and IL-$33_{112–270}$ proteins (*A*) or endogenous native IL-33 isolated from necrotic endothelial cells (*B*) were incubated with purified neutrophil elastase (NE) (30 mU/μL; 30 min at 37 °C) or cathepsin G (CG) (0.09 mU/μL; 1 h at 37 °C). (*C*) Primary structure of human IL-33. The IL-1-like domain with its 12 β-strands (black boxes) is indicated. The sequence surrounding the cathepsin G and elastase cleavage sites is shown. (*D* and *E*) Identification of the cathepsin G and elastase cleavage sites using IL-33 single-point and deletion mutants. Mutation of F_{94} and L_{108} to glycine abrogates formation of the 21- and 18-kDa CG (0.06 mU/μL; 1 h at 37 °C) cleavage products, respectively (*D*). Mutation of I_{98} to glycine abrogates formation of the 20 kDa elastase (19 mU/μL; 30 min at 37 °C) cleavage product (*E*). In vitro-translated proteins IL-$33_{95–270}$, IL-$33_{109–270}$, and IL-$33_{99–270}$ comigrate on SDS/PAGE with the 21- and 18-kDa cathepsin G cleavage products (*D*) and the 20-kDa elastase cleavage product (*E*), respectively. *, secondary cleavage products. Proteins were separated by SDS/PAGE and revealed by Western blot with anti-IL-33-Cter mAb 305B (*A*–*E*). Blots are representative of at least three independent experiments.

The observation that full-length IL-33 is biologically active did not exclude the possibility that some proteases released during inflammation may cleave IL-33 and generate mature forms with increased biological activity. Because neutrophil serine proteases have been shown to play key roles in inflammatory processes and maturation of IL-1 family cytokines (1, 3, 33–37), we tested the possibility that neutrophil serine proteases may be involved in the processing and regulation of IL-33. We found that incubation of full-length IL-33 with neutrophil elastase and cathepsin G resulted in an increase of IL-6 secretion by mast cells, compared with cells treated with full-length IL-33 alone (Fig. 1*C*). These later results indicated that neutrophil elastase and cathepsin G can regulate IL-33 bioactivity.

Neutrophil Elastase and Cathepsin G Cleave IL-33 and Generate Mature Forms IL-$33_{95–270}$, IL-$33_{99–270}$, and IL-$33_{109–270}$. We then asked whether full-length IL-$33_{1–270}$ is cleaved by neutrophil serine proteases. Western blot analysis revealed that neutrophil elastase and cathepsin G process in vitro-translated full-length human IL-33 and generate cleavage products of ~18–21 kDa (Fig. 2*A*). In contrast, the C-terminal IL-1-like domain of IL-33 (IL-$33_{112–270}$) is not cleaved by neutrophil serine proteases. Importantly, neutrophil elastase and cathepsin G also cleaved endogenous native IL-33 released from human necrotic endothelial cells and generated similar cleavage products, one major product of ~20 kDa for elastase and two major products of ~21 and 18 kDa for cathepsin G (Fig. 2*B*). Based on size of the cleavage products, MS analyses (Fig. S1 and Table S1) and site-directed mutagenesis experiments (Fig. 2 *C*–*E*), we mapped the cathepsin G cleavage sites at F94 and L108 and the elastase cleavage site at I98 in the human IL-33 sequence. As shown in Fig. 2*D*, replacement of residue F94 by a glycine abrogated the formation of the larger cathepsin G cleavage product, whereas mutagenesis of residue L108 prevented formation of the smaller product. In addition, the cathepsin G cleavage products comigrated on SDS/PAGE with in vitro-translated IL-$33_{95–270}$ and IL-$33_{109–270}$ proteins. Similarly, the elastase cleavage product comigrated on gels with in vitro-translated IL-$33_{99–270}$, and this product was no longer observed when residue I98 was replaced by a glycine (Fig. 2*E*). Cleavage products of higher molecular mass were observed with the F94G and I98G mutants (Fig. 2 *D* and *E*), indicating that cathepsin G and elastase can cleave IL-33 at additional secondary sites, further upstream in the N-terminal part. Together, these findings indicated that IL-33 is a substrate for neutrophil elastase and cathepsin G, which generate mature forms, IL-$33_{95–270}$, IL-$33_{99–270}$, and IL-$33_{109–270}$.

IL-33 Mature Forms Are Generated by Activated Human Neutrophils ex Vivo. We observed that PMA-activated human neutrophils can process full-length IL-$33_{1–270}$ and generate three cleavage products, which comigrate on SDS/PAGE with IL-$33_{95–270}$, IL-$33_{99–270}$, and IL-$33_{109–270}$ (Fig. 3*A*). Cleavage of IL-$33_{1–270}$ was prevented when neutrophils were treated with the serine protease inhibitor 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF) (Fig. 3*B*). Western blot analysis revealed that the three cleavage products are detected with anti-IL-33-Cter antibodies directed against the IL-1-like domain but not detected with anti-IL-33-Nter antibodies

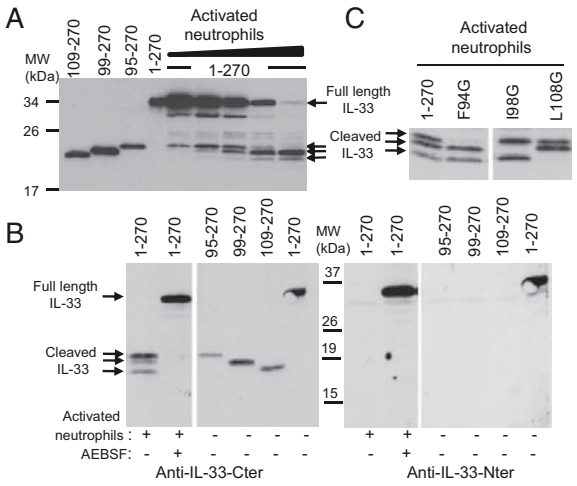


Fig. 3. Activated human neutrophils generate IL-33 mature forms IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$. (*A* and *B*) In vitro-translated full-length human IL-33$_{1–270}$ was incubated with PMA-activated human neutrophils. Proteins were comigrated on SDS/PAGE with in vitro-translated IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$ proteins and revealed by Western blot with anti-IL-33-Cter mAb 305B (*A* and *B*, *Left*) or anti-IL-33-Nter antibodies (*B*, *Right*). IL-33 was incubated with increasing amounts of supernatants from activated neutrophils (6×10^4–10^6 neutrophils; 15 min at 37 °C) (*A*). Cleavage of IL-33 by activated neutrophils (10^5 neutrophils; 2 h at 37 °C) was inhibited by serine protease inhibitor AEBSF (1 mM) (*B*). (*C*) In vitro-translated IL-33$_{1–270}$ or single-point mutants IL-33$_{F94G}$, IL-33$_{I98G}$, and IL-33$_{L108G}$ were incubated with PMA-activated human neutrophils (10^5 neutrophils; 2 h at 37 °C). Proteins were separated on SDS/PAGE and revealed by Western blot with anti-IL-33-Cter mAb 305B. Blots are representative of three independent experiments, with neutrophils isolated from three different donors.

recognizing the first 15 amino-terminal residues of IL-33, indicating that the cleaved forms contain the C-terminal IL-1-like domain of IL-33. Incubation of IL-33$_{1–270}$ single point mutants with activated neutrophils (Fig. 3*C*) indicated that: (*i*) the larger cleavage product corresponds to the IL-33$_{95–270}$ form, because this form is not observed in the Phe94 mutant; (*ii*) the second cleavage product corresponds to the IL-33$_{99–270}$ form, which is eliminated in the Ile98 mutant; (*iii*) the smaller cleavage product corresponds to the IL-33$_{109–270}$ form, which is not generated in the Leu108 mutant. These results thus demonstrate that activated human neutrophils can process full-length IL-33 into mature forms IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$.

IL-33 Mature Forms Have Increased Biological Activity Compared with the Full-Length IL-33$_{1–270}$ Protein. We tested the biological activity of IL-33 mature forms in two cellular bioassays (23, 38) and found that in vitro-translated human IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$ induced IL-6 secretion by MC/9 mast cells (Fig. 4*A*) and IL-5 secretion by KU812 basophil-like cells (Fig. 4*B*). We then compared the biological activity of the IL-33 mature forms to that of the full-length IL-33$_{1–270}$ protein. The four IL-33 forms were quantified by fluorescence (Fig. 4*C*), and secretion of IL-6 by MC/9 cells in response to different concentrations of the proteins was analyzed. These experiments revealed that ~10-fold higher concentrations of the full-length IL-33 protein were required to obtain similar levels of IL-6 secretion by MC/9 cells (Fig. 4*C*). We concluded that mature forms IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$ have a higher biological activity (~10-fold) than the full-length IL-33 protein.

IL-33 Mature Forms Are Biologically Active in Vivo. To determine whether IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$ are biologically

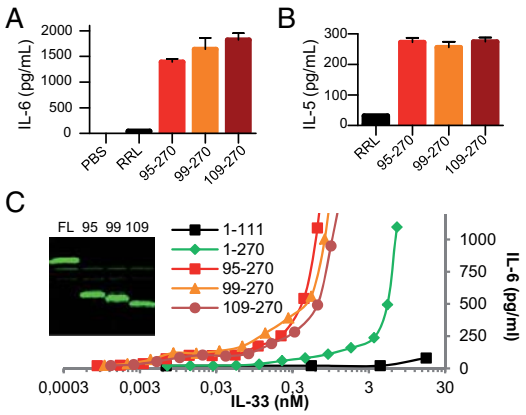


Fig. 4. IL-33 mature forms IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$ are biologically active. (*A* and *B*) The capacity of IL-33 mature forms produced in RRL (5 μL of lysate) to activate the IL-33-responsive MC/9 mast cells (10^5 cells/well; 24-h stimulation) (*A*) and KU812 basophil-like chronic myelogenous leukemia cells (5×10^5 cells/well; 24-h stimulation) (*B*) was analyzed by determining cytokine levels in supernatants [IL-6 (*A*); IL-5 (*B*)] using ELISAs. Results are shown as means and SD of three separate data points. (*C*) Comparison of the biological activity of IL-33 full-length and mature forms. The different proteins were quantified by fluorescence and various concentrations were used to stimulate MC/9 mast cells (10^5 cells/well; 24-h stimulation). IL-6 protein levels in supernatants were detected by ELISA. Results are representative of at least two independent experiments.

active in vivo, we produced recombinant proteins corresponding to these mature forms in *Escherichia coli*. We injected the IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$ recombinant proteins into wild-type mice, intraperitoneally (i.p.) every day over a 1-wk period and observed a striking increase in the size of the spleen (Fig. 5*A*), as previously observed with the artificially truncated form IL-33$_{112–270}$ (4, 24). Spleen weight increased from ~100 mg in PBS-treated mice to >200 mg in mice treated with IL-33$_{95–270}$, IL-33$_{99–270}$, or IL-33$_{109–270}$ recombinant proteins (Fig. 5*A*). Injection of IL-33 mature forms in vivo also increased the numbers of blood granulocytes and monocytes (Fig. 5 *B* and *C*) and the serum concentrations of IL-5 (Fig. 5*D*), a Th2 cytokine known to be produced by natural helper cells in vivo in response to IL-33 (17). In the intestine, IL-33 has previously been shown to induce goblet cell hyperplasia and mucus secretion, an effect which is mediated through up-regulation of the cytokine IL-13 (4). Periodic acid Schiff and alcian blue staining revealed that mucus secretion was highly increased in the jejunum of mice treated with the IL-33$_{95–270}$ and IL-33$_{99–270}$ recombinant proteins, compared with PBS-treated mice (Fig. 5*E*). Together, these results indicated that the mature forms IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$ are biologically active in vivo. As previously reported by another group (24), we had difficulties to obtain correctly folded recombinant full-length IL-33$_{1–270}$ in quantities compatible with in vivo assays, and this precluded the comparison of the bioactivity of full-length and mature IL-33 forms in vivo.

Full-Length and Cleaved Endogenous IL-33 Are Detected in Bronchoalveolar Lavage Fluid in an in Vivo Model of Acute Lung Injury Associated with Neutrophil Infiltration. We then determined whether cleavage of IL-33 by neutrophil elastase and cathepsin G also occurs in mice. Western blot analysis revealed that neutrophil cathepsin G processes in vitro-translated full-length murine IL-33$_{1–266}$ and generates one major cleavage product of ~20 kDa (Fig. 6*A*). This product comigrated on SDS/PAGE with in vitro-translated murine IL-33$_{102–266}$, indicating that cleavage by cathepsin G may occur after Phe101. Neutrophil elastase generated a similar major cleavage product of ~20 kDa and a second minor

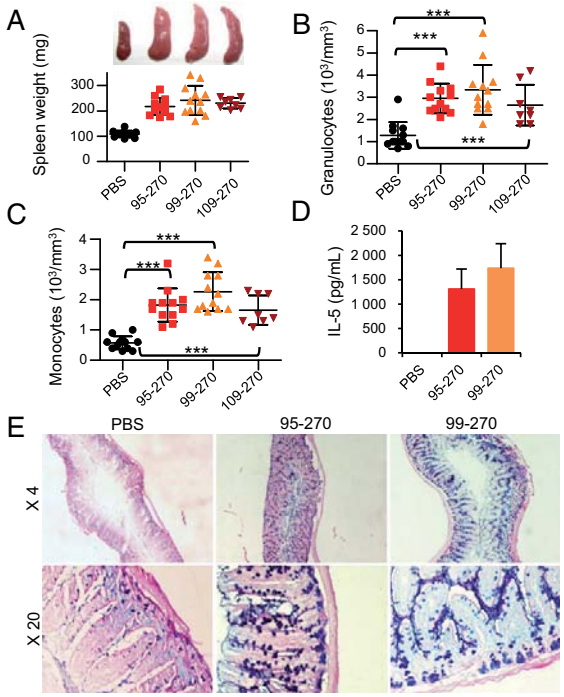


Fig. 5. In vivo activity of IL-33 mature forms IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$. (*A–E*) BALB/c mice were injected (i.p.) with recombinant proteins (4 μg) every day during 1 wk, and spleens (*A*), peripheral blood (*B* and *C*), serum (*D*), and jejunum (*E*) were collected at day 8 (n = 12/3 experiments for PBS, IL-33$_{95–270}$, and IL-33$_{99–270}$; n = 8/2 experiments for IL-33$_{109–270}$). Spleen weights and representative photographs for each group of mice are shown (*A*). Numbers of peripheral blood granulocytes (*B*), monocytes (*C*), and serum concentrations of IL-5 (*D*) were determined. ***$P < 0.001$ (Student unpaired and two-tailed t test). Epithelial changes in the jejunum are shown (*E*). Mucus was stained with periodic acid Schiff and alcian blue.

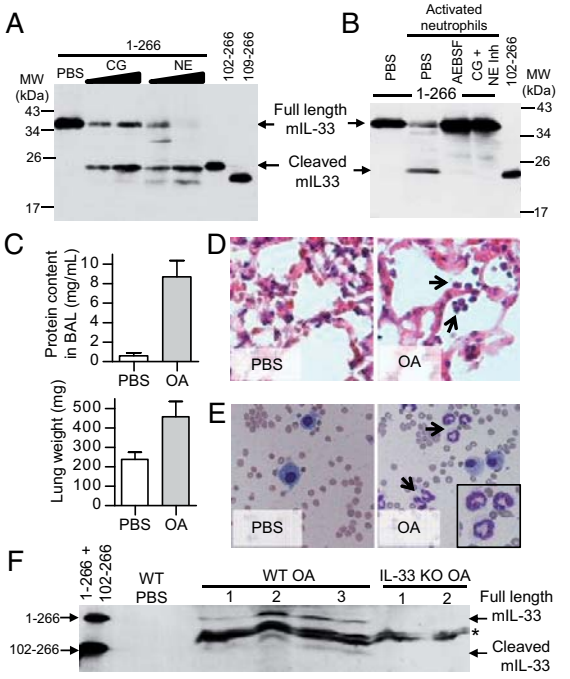


Fig. 6. Murine IL-33 is processed by neutrophil elastase and cathepsin G ex vivo, and cleaved endogenous IL-33 is detected in vivo in a model of acute lung injury associated with neutrophil infiltration. (*A* and *B*) In vitro-translated full-length murine IL-33$_{1–266}$ (2.5 μL) was incubated with purified neutrophil cathepsin G (CG) (14.7 and 29.4 μU/μL) or elastase (NE) (7.3 and 14.7 mU/μL) (*A*) or with supernatants of activated mouse bone marrow neutrophils (3 × 10^4 neutrophils) (*B*) for 1 h at 37 °C. Activated neutrophils were pretreated with PBS, serine protease inhibitor AEBSF (8 mM), or cathepsin G and elastase inhibitors (50 μM) for 1 h at 37 °C. Proteins were comigrated on SDS/PAGE with in vitro-translated murine IL-33$_{102–266}$ and IL-33$_{109–266}$ and revealed by Western blot with goat anti-mouse IL-33-Cter antibodies. Blots are representative of at least three independent experiments. (*C–F*) Full-length and cleaved endogenous IL-33 are released in BAL fluid following OA-induced acute lung injury. Lung injury was assessed by measuring lung weight and protein content in BAL fluid two hours after i.v. OA administration (*C*). Accumulation of neutrophils (black arrows) was analyzed by hematoxylin/eosin staining of lung tissue sections (*D*) and May–Grunwald–Giemsa staining of BAL fluids cytospins (*E*). Proteins from cell-free BAL fluids (30 μL) of OA-treated wild-type (three mice) or *IL-33*$^{-/-}$ mice (two mice) were comigrated on SDS/PAGE with in vitro-translated murine IL-33$_{1–266}$ and IL-33$_{102–266}$ and revealed by Western blot with goat anti-mouse IL-33-Cter antibodies (*F*). *Nonspecific bands detected in both wild-type and *IL-33*$^{-/-}$ mice.

product, which comigrated with in vitro-translated murine IL-33$_{102–266}$ and IL-33$_{109–266}$, respectively. A single major cleavage product of ~20 kDa comigrating with IL-33$_{102–266}$ was also generated when full-length murine IL-33$_{1–266}$ was incubated with supernatant of activated neutrophils prepared from murine bone marrow (Fig. 6*B*). Cleavage was abrogated in the presence of serine protease inhibitor AEBSF or neutrophil elastase and cathepsin G inhibitors. To determine whether cleaved IL-33 is generated in vivo, we used a mouse model of acute lung injury induced by i.v. administration of oleic acid (OA), which results in damage to the lung alveolar epithelium and recruitment of neutrophils in the injured lung (39). We selected this murine model of human ARDS (acute respiratory distress syndrome) because endogenous IL-33 is constitutively expressed in the alveolar epithelium (40) and can be released in the extracellular space after cellular damage (23, 24). Two hours after i.v. injection of OA, we observed an increase in lung weight and protein content in bronchoalveolar lavage (BAL) fluid (Fig. 6*C*) and an accumulation of neutrophils in the alveolar wall and in BAL fluid (Fig. 6 *D* and *E*). Western blot analysis with anti-IL-33-Cter antibodies revealed that full-length endogenous IL-33 is released in BAL fluid following acute lung injury (Fig. 6*F*). Interestingly, a cleaved IL-33 form comigrating with in vitro-translated murine IL-33$_{102–266}$ was also detected in BAL fluid from the injured lung. Full-length and cleaved IL-33 forms were detected with anti-IL-33 antibodies in OA-treated wild-type mice but not in OA-treated IL-33$^{-/-}$ mice (Fig. 6*F*), indicating they correspond to bona fide endogenous IL-33 forms. We concluded that full-length and cleaved endogenous IL-33 are detected in BAL fluid after alveolar epithelium damage and neutrophil recruitment in the alveolar wall.

Discussion

In this report, we describe mature bioactive forms of the IL-1 family cytokine IL-33. Because full-length IL-33 is biologically active and cleavage by caspases results in its inactivation, rather than its activation, it remained unclear whether mature bioactive forms of the cytokine could be generated. We now demonstrate that the bioactivity of human IL-33 is increased by neutrophil serine proteases cathepsin G and elastase and that full-length IL-33$_{1–270}$ is processed by these purified proteases or activated neutrophils, into mature forms IL-33$_{95–270}$, IL-33$_{99–270}$, and IL-33$_{109–270}$. These IL-33 mature forms contain an intact IL-1-like

cytokine domain and exhibit a ~10-fold higher activity than the full-length protein in cellular bioassays. They are very potent in vivo and induce striking increases in spleen weight, blood granulocyte and monocyte numbers, and serum concentrations of IL-5, as well as profound changes in the intestine. Murine IL-$33_{1–266}$ can also be processed by neutrophil elastase and cathepsin G or activated neutrophils, and both full-length and cleaved endogenous IL-33 were detected in vivo in BAL fluid in a mouse model of acute lung injury associated with high levels of neutrophil recruitment. Together, these findings bring important insights into the molecular mechanisms regulating the activity of IL-33. They provide experimental evidence that proteolytic processing of IL-33 and removal of the N-terminal part can greatly increase IL-33 bioactivity. These results, thus, support the possibility that proteolytic processing of IL-33 may be required for the extracellular generation of highly active cytokine in vivo. In addition, they suggest that the inflammatory microenvironment may exacerbate disease-associated functions of IL-33 through the generation of these highly active mature forms.

We have previously proposed (23, 27) that IL-33 may function as an endogenous danger signal or alarmin, similar to IL-1α and HMGB1 (28–32), to alert the immune system of cell or tissue damage during trauma or infection. In support of this model, we have shown that IL-33 is constitutively expressed to high levels in the nuclei of endothelial and epithelial cells in normal human tissues (27) and that it can be released in the extracellular space after cellular damage (23). Neutrophils are rapidly recruited into injured tissues during infection or in the absence of infection during "sterile inflammation," following the release of major alarmin molecules such as IL-1α and HMGB1 (28–32). After activation, they rapidly release serine proteases from cytoplasmic granules into the extracellular space (33). Neutrophil elastase and cathepsin G may, thus, process IL-33 released from damaged cells into highly active mature forms, soon after the tissue injury, in the early stages of inflammation. Regulation of IL-33 bioactivity by neutrophil serine proteases may be particularly important in sterile neutrophilic inflammation, which is thought to contribute to the pathogenesis of acute lung and liver injuries, acute ischemia-induced injuries, and chronic diseases affecting the lung, bowel, and joints (29). IL-33 has been shown to play important roles in mouse models of rheumatoid arthritis (12, 13) and to orchestrate neutrophil migration into the joints (41). Neutrophil serine proteases are critical for IL-1β processing in the acute phase of arthritis, characterized by a strong neutrophilic infiltrate (36, 37), and could also mediate IL-33 processing into mature bioactive forms in inflammatory arthritis. Finally, processing of IL-33 by neutrophils proteases may also occur during bacterial, fungal or viral infections. For instance, generation of IL-33 mature forms by airway neutrophils following influenza virus infection could modulate the activity of type 2 innate lymphoid cells in the lungs (20, 21).

Our results show that mature bioactive forms of IL-33 are generated by caspase-1-independent mechanisms. Interestingly, caspase-1-independent activation of IL-1β and IL-18 has been reported in several studies (1, 3, 34–37, 42, 43). Neutrophil serine proteases cathepsin G, elastase, and proteinase-3 (PR3) have been shown to cleave the IL-1β precursor a few residues upstream the caspase-1 maturation site and to produce mature bioactive forms of the cytokine (1, 3, 34–37). Extracellular processing of IL-33 into mature bioactive forms by neutrophil serine proteases is, thus, a mechanism shared with other IL-1 family members. In addition to cathepsin G and elastase, PR3 may also play a role in the regulation of IL-33 bioactivity because we observed that PR3 can process full-length IL-33 into cleavage products of ~18–20 kDa (Fig. S2). Moreover, other proteases, including granzymes, matrix metalloprotease 9, and mast cell chymase, have been shown to process IL-1α, IL-1β, and IL-18 precursors into active cytokines (1, 3, 36, 44). It remains to be seen whether these or other proteases may also play a role in the processing of IL-33 into mature bioactive forms. Furthermore, the IL-1α-processing protease calpain has been reported to cleave IL-33 (45), but neither the molecular nature nor the biological activity of the cleavage product has been characterized yet.

In summary, this study provides strong evidence that mature bioactive forms of the IL-1 family cytokine IL-33 can be generated by neutrophil serine proteases cathepsin G and elastase. This report describes a precise mechanism leading to the generation of highly active IL-33 mature forms. Further characterization of IL-33 processing could bring important insights in the regulation of IL-33 activity in inflammatory disease processes.

Materials and Methods

SI Materials and Methods provides details regarding plasmid constructions, protein production, Western blot analysis, histology, and analysis of blood, spleen, lung, and BAL samples.

Protein Cleavage Assays with Neutrophil Proteases and Isolated Neutrophils. In vitro-translated proteins (2–5 μL of lysate) were incubated with neutrophil elastase (0.3 U; Calbiochem) and cathepsin G (1 mU; Calbiochem) in 15 μL of assay buffer (2–5 μL of RRL lysate plus 10 μL of PBS) for 30 min to 1 h at 37 °C. Cleavage assays with activated neutrophils (6×10^4–10^6 neutrophils; 15 min to 2 h at 37 °C) were performed using human neutrophils from healthy blood donors (Etablissement Français du sang) or mouse neutrophils isolated from femur and tibia bone marrow. For a full description, see *SI Materials and Methods*.

IL-33 Activity Assays. In vitro-translated full-length IL-33 or mature forms (5 μL of lysate/well; 24-h treatment) were used to stimulate IL-33-responsive MC/9 mast cells (ATCC; 10^5 to 2×10^5 cells/well in 96-well plates) (23) and KU812 basophil-like chronic myelogenous leukemia cells (ATCC; 5×10^5 cells/well in 96-well plates) (38). Cytokine levels in supernatants were determined using DuoSet IL-6 and IL-5 ELISAs (R&D Systems).

Animals. Female BALB/c mice received daily i.p. injections of 4 μg of recombinant human IL-$33_{95–270}$, IL-$33_{99–270}$, IL-$33_{109–270}$, or saline for 7 d. Blood and histologic analyses were performed on day 8. Acute lung injury was induced in female C57BL/6 wild type or *IL-33*$^{-/-}$ mice (8–10 wk old) by i.v. injection of OA (0.8 μL/g body weight; Sigma) in a 15% solution with 0.1% BSA. Lung histology and BAL fluid were analyzed 2 h after OA injection. All mice were bred under specific pathogen-free conditions and handled according to institutional guidelines under protocols approved by the Institut de Pharmacologie et de Biologie Structurale (IPBS) and "Région Midi-Pyrénées" animal care committees.

ACKNOWLEDGMENTS. We thank the Infrastructures en Biologie, Santé et Agronomie (IBiSA) Toulouse Proteomics, Toulouse Réseau Imagerie (TRI)-IPBS and Anexplo-IPBS facilities. We are grateful to members of the J.-P.G. laboratory for help with animal experiments. This work was supported by Ligue Nationale contre le Cancer Equipe Labellisée "LIGUE 2009" (to J.-P.G.), an Association pour la Recherche sur le Cancer fellowship (to E.L.), and an Association pour la Recherche sur le Cancer Programme grant (to J.-P.G.).

1. Dinarello CA (2009) Immunological and inflammatory functions of the interleukin-1 family. *Annu Rev Immunol* 27:519–550.
2. Sims JE, Smith DE (2010) The IL-1 family: regulators of immunity. *Nat Rev Immunol* 10: 89–102.
3. Dinarello CA (2011) Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood* 117:3720–3732.
4. Schmitz J, et al. (2005) IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. *Immunity* 23:479–490.
5. Baekkevold ES, et al. (2003) Molecular characterization of NF-HEV, a nuclear factor preferentially expressed in human high endothelial venules. *Am J Pathol* 163:69–79.
6. Carriere V, et al. (2007) IL-33, the IL-1-like cytokine ligand for ST2 receptor, is a chromatin-associated nuclear factor in vivo. *Proc Natl Acad Sci USA* 104:282–287.
7. Roussel L, Erard M, Cayrol C, Girard JP (2008) Molecular mimicry between IL-33 and KSHV for attachment to chromatin through the H2A-H2B acidic pocket. *EMBO Rep* 9: 1006–1012.
8. Smith DE (2010) IL-33: a tissue derived cytokine pathway involved in allergic inflammation and asthma. *Clin Exp Allergy* 40:200–208.

9. Liew FY, Pitman NI, McInnes IB (2010) Disease-associated functions of IL-33: the new kid in the IL-1 family. *Nat Rev Immunol* 10:103–110.
10. Oboki K, Ohno T, Kajiwara N, Saito H, Nakae S (2010) IL-33 and IL-33 receptors in host defense and diseases. *Allergol Int* 59:143–160.
11. Moffatt MF, et al.; GABRIEL Consortium (2010) A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 363:1211–1221.
12. Xu D, et al. (2008) IL-33 exacerbates antigen-induced arthritis by activating mast cells. *Proc Natl Acad Sci USA* 105:10913–10918.
13. Palmer G, et al. (2009) Inhibition of interleukin-33 signaling attenuates the severity of experimental arthritis. *Arthritis Rheum* 60:738–749.
14. Pastorelli L, et al. (2010) Epithelial-derived IL-33 and its receptor ST2 are dysregulated in ulcerative colitis and in experimental Th1/Th2 driven enteritis. *Proc Natl Acad Sci USA* 107:8017–8022.
15. Sanada S, et al. (2007) IL-33 and ST2 comprise a critical biomechanically induced and cardioprotective signaling system. *J Clin Invest* 117:1538–1549.
16. Guo L, et al. (2009) IL-1 family members and STAT activators induce cytokine production by Th2, Th17, and Th1 cells. *Proc Natl Acad Sci USA* 106:13463–13468.
17. Moro K, et al. (2010) Innate production of T(H)2 cytokines by adipose tissue-associated c-Kit(+)Sca-1(+) lymphoid cells. *Nature* 463:540–544.
18. Neill DR, et al. (2010) Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* 464:1367–1370.
19. Price AE, et al. (2010) Systemically dispersed innate IL-13-expressing cells in type 2 immunity. *Proc Natl Acad Sci USA* 107:11489–11494.
20. Chang YJ, et al. (2011) Innate lymphoid cells mediate influenza-induced airway hyperreactivity independently of adaptive immunity. *Nat Immunol* 12:631–638.
21. Monticelli LA, et al. (2011) Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nat Immunol* 12:1045–1054 10.1038/ni.2131.
22. Oboki K, et al. (2010) IL-33 is a crucial amplifier of innate rather than acquired immunity. *Proc Natl Acad Sci USA* 107:18581–18586.
23. Cayrol C, Girard JP (2009) The IL-1-like cytokine IL-33 is inactivated after maturation by caspase-1. *Proc Natl Acad Sci USA* 106:9021–9026.
24. Lüthi AU, et al. (2009) Suppression of interleukin-33 bioactivity through proteolysis by apoptotic caspases. *Immunity* 31:84–98.
25. Talabot-Ayer D, Lamacchia C, Gabay C, Palmer G (2009) Interleukin-33 is biologically active independently of caspase-1 cleavage. *J Biol Chem* 284:19420–19426.
26. Ali S, Nguyen DQ, Falk W, Martin MU (2010) Caspase 3 inactivates biologically active full length interleukin-33 as a classical cytokine but does not prohibit nuclear translocation. *Biochem Biophys Res Commun* 391:1512–1516.
27. Moussion C, Ortega N, Girard JP (2008) The IL-1-like cytokine IL-33 is constitutively expressed in the nucleus of endothelial cells and epithelial cells in vivo: a novel 'alarmin'? *PLoS ONE* 3:e3331.
28. Scaffidi P, Misteli T, Bianchi ME (2002) Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature* 418:191–195.
29. Chen CJ, et al. (2007) Identification of a key pathway required for the sterile inflammatory response triggered by dying cells. *Nat Med* 13:851–856.
30. Bianchi ME (2007) DAMPs, PAMPs and alarmins: all we need to know about danger. *J Leukoc Biol* 81:1–5.
31. Cohen I, et al. (2010) Differential release of chromatin-bound IL-1alpha discriminates between necrotic and apoptotic cell death by the ability to induce sterile inflammation. *Proc Natl Acad Sci USA* 107:2574–2579.
32. Rider P, et al. (2011) IL-1α and IL-1β recruit different myeloid cells and promote different stages of sterile inflammation. *J Immunol* 187:4835–4843.
33. Pham CT (2006) Neutrophil serine proteases: specific regulators of inflammation. *Nat Rev Immunol* 6:541–550.
34. Hazuda DJ, Strickler J, Kueppers F, Simon PL, Young PR (1990) Processing of precursor interleukin 1 beta and inflammatory disease. *J Biol Chem* 265:6318–6322.
35. Coeshott C, et al. (1999) Converting enzyme-independent release of tumor necrosis factor alpha and IL-1beta from a stimulated human monocytic cell line in the presence of activated neutrophils or purified proteinase 3. *Proc Natl Acad Sci USA* 96: 6261–6266.
36. Guma M, et al. (2009) Caspase 1-independent activation of interleukin-1beta in neutrophil-predominant inflammation. *Arthritis Rheum* 60:3642–3650.
37. Joosten LA, et al. (2009) Inflammatory arthritis in caspase 1 gene-deficient mice: contribution of proteinase 3 to caspase 1-independent production of bioactive interleukin-1beta. *Arthritis Rheum* 60:3651–3662.
38. Tare N, et al. (2010) KU812 cells provide a novel in vitro model of the human IL-33/ST2L axis: functional responses and identification of signaling pathways. *Exp Cell Res* 316:2527–2537.
39. Zhou Z, Kozlowski J, Schuster DP (2005) Physiologic, biochemical, and imaging characterization of acute lung injury in mice. *Am J Respir Crit Care Med* 172:344–351.
40. Louten J, et al. (2011) Endogenous IL-33 enhances Th2 cytokine production and T-cell responses during allergic airway inflammation. *Int Immunol* 23:307–315.
41. Verri WA, Jr., et al. (2010) IL-33 induces neutrophil migration in rheumatoid arthritis and is a target of anti-TNF therapy. *Ann Rheum Dis* 69:1697–1703.
42. Fantuzzi G, et al. (1997) Response to local inflammation of IL-1 beta-converting enzyme- deficient mice. *J Immunol* 158:1818–1824.
43. van de Veerdonk FL, Netea MG, Dinarello CA, Joosten LA (2011) Inflammasome activation and IL-1β and IL-18 processing during infection. *Trends Immunol* 32:110–116.
44. Afonina IS, et al. (2011) Granzyme B-dependent proteolysis acts as a switch to enhance the proinflammatory activity of IL-1α. *Mol Cell* 44:265–278.
45. Hayakawa M, et al. (2009) Mature interleukin-33 is produced by calpain-mediated cleavage in vivo. *Biochem Biophys Res Commun* 387:218–222.

Supporting Information

Lefrançais et al. 10.1073/pnas.1115884109

SI Results

MS Analysis of IL-33 Processing by Neutrophil Serine Proteases. Recombinant IL-33, expressed in *E. coli* as a GST fusion, was purified on glutathione-Sepharose beads and digested on beads with either elastase or cathepsin G, and cleavage products were separated from full-length protein by 1D SDS/PAGE. To identify the new N terminus of these cleavage products, a quantitative proteomic mapping was performed: gel bands corresponding respectively to the full-length protein (control) and to the cleaved IL-33 were excised, further in-gel digested with a specific enzyme (trypsin or Glu-C), and the resulting smaller peptides were identified and quantified by MS. Peptides originating from the C-terminal part (IL-1 domain) were detected both in the full-length and cleaved proteins, and peptides originating from the N-terminal part of the protein (amino acids 1–90) were specifically detected in full-length IL-33. Only one peptide was specifically detected in the cleavage product (Fig. S1 *B*, *D*, and *F*, red spot on the intensity plots), and this peptide was, therefore, assigned as the new N terminus of the protein processed by elastase or cathepsin G. The sequences of these peptides were deduced from fragmentation MS/MS spectra acquired during the analysis (Fig. S1 *A*, *C*, and *E*).

Specific Tryptic or Glu-C Fragments from the Proteomic Mapping. Table S1 summarizes the results by listing the specific tryptic or Glu-C fragments expected from the proteomic mapping, encompassing the cleavage sites identified for the two neutrophil proteases. The number of MS/MS spectra acquired for each species reflects the abundance of the peptide. Short peptides corresponding to the new N terminus were specifically detected, with many MS/MS sequencing events only in the processed forms of IL-33.

SI Materials and Methods

Plasmid Construction and Protein Production. IL-33 deletion mutants were amplified by PCR using the human (NM_033439) and mouse (NM_133775) IL-33/NF-HEV cDNAs (1, 2) as templates. The PCR fragments, thus obtained, were cloned into plasmid pcDNA3.1 (Invitrogen). The human IL-33_{F94G}, IL-33_{I98G}, IL-33_{V101G}, and IL-33_{L108G} mutants were generated by PCR and cloned into the same expression vector. All primer sequences are available upon request. Wild-type and mutant IL-33 proteins were synthesized in vitro in RRL using the TNT-T7 kit (Promega). Human IL-$33_{1–270}$ was also synthesized using WGEs (Promega) and HES (Pierce). For expression in *Escherichia coli*, cDNAs encoding mature forms IL-$33_{95–270}$, IL-$33_{99–270}$, and IL-$33_{109–270}$ were subcloned into expression vector pET-15b (Novagen). Recombinant proteins were produced in *E. coli* BL21pLysS (Novagen) and purified using Ni-NTA agarose column (Qiagen). His tag was removed by cleavage with thrombin, and proteins were further purified by gel filtration (FPLC; GE Healthcare). Endotoxin levels were <0.01 EU/μg of protein, as determined by the *Limulus* amebocyte lysate QCL-1000 method (Lonza).

Protein Cleavage Assays with Neutrophil Proteases and Isolated Neutrophils. In vitro-translated proteins (2-5 μL lysate) were incubated with neutrophil elastase (0.3 U; Calbiochem), cathepsin G (1 mU; Calbiochem), or PR3 (70 μU; Calbiochem) in 15 μL of assay buffer (2–5 μL of RRL lysate plus 10 μL of PBS) for 30 min to 1 h at 37 °C. Endogenous native IL-33 protein isolated from human endothelial cell freeze-thaw extracts (3) was used in some experiments. Cleavage assays with activated neutrophils (6×10^4–10^6 neutrophils; 15 min to 2 h at 37 °C) were performed using human neutrophils from healthy blood donors (Etablissement Français du sang; Contract 21/PVNT/TOU/IPBS01/2009-0052), isolated using Polymorphprep cell separation media (Axis-Shield) and activated with PMA (25 mM; 2 h), or mouse neutrophils isolated by Percoll density gradient from femur and tibia bone marrow and activated by stimulation with cytochalasin B (5 μg/mL; 15 min) and fMLP (40 μM; 3 h). In some experiments, neutrophils or neutrophil supernatants were incubated with serine protease inhibitor AEBSF (1–8 mM; Calbiochem) or Cathepsin G Inhibitor I and Elastase Inhibitor IV (50 μM; Calbiochem). Cleavage products were analyzed by SDS/PAGE and Western blot.

Western Blot Analysis. Proteins were fractionated by SDS/PAGE, electroblotted, and detected with mAb to human IL-33-Cter (305B; 1/1,000; Alexis Biochemicals), rabbit antiserum to human IL-33-Nter [IL-$33_{1–15}$; 1/400 (1, 2)] or goat antiserum to mouse IL-33-Cter (AF326; 1/500; R&D Systems), followed by HRP-conjugated goat anti-mouse, goat anti-rabbit, or donkey anti-goat polyclonal antibodies (1/10,000; Promega), and finally an enhanced chemiluminescence kit (GE Healthcare). Quantitative IR Western blots were performed to quantify the four IL-33 forms used in cellular bioassays, using the IL-33-Cter mAb 305B, goat anti-mouse IgG IRDye800 secondary antibody (610-132-121; 1/10,000; Rockland Immunochemicals), and Odyssey IR Imager (LI-COR Biosciences). A standard curve was performed using purified recombinant IL-$33_{95–270}$ protein.

IL-33 Activity Assays. In vitro-translated full-length IL-33 or mature forms (5 μL lysate/well; 24-h treatment) were used to stimulate IL-33-responsive MC/9 mast cells (ATCC; 10^5–2×10^5 cells/well in 96-well plates) (3) and KU812 basophil-like chronic myelogenous leukemia cells (ATCC; 5×10^5 cells/well in 96-well plates) (4). Cytokine levels in supernatants were determined using DuoSet IL-6 and IL-5 ELISAs (R&D Systems).

Animals. BALB/c and C57BL/6 wild-type mice were purchased from Charles River Laboratories. Female BALB/c mice received daily i.p. injections of 4 μg of recombinant human IL-$33_{95–270}$, IL-$33_{99–270}$, IL-$33_{109–270}$, or saline for 7 d. Blood and histologic analyses were performed on day 8. Acute lung injury was induced in female C57BL/6 wild-type or *IL-33*$^{-/-}$ mice (8–10 wk old) by i.v. injection of OA (0.8 μL/g body weight; Sigma) in a 15% solution with 0.1% BSA. Lung histology and BAL fluid were analyzed 2 h after OA injection. All mice were bred under specific pathogen-free conditions and handled according to institutional guidelines under protocols approved by the IPBS and "Région Midi-Pyrénées" animal care committees.

Histology. Histological evaluation was performed on formalin-fixed mouse tissues. Five-micron paraffin-embedded tissue sections (jejunum) or cryosections (lung) were prepared and stained with hematoxylin and eosin for morphological evaluation. Periodic acid Schiff and alcian blue staining was used to detect the presence of mucus in jejunum tissue sections.

Analysis of Blood, Spleen, Lung, BAL, and Serum Samples. Peripheral blood was obtained by cardiac puncture and stored in EDTA-containing tubes at 4 °C until analysis with an automated hematological analyzer (ABX Micros 60). Spleens and lungs were collected, and the weights determined. Two hours after OA injection, mice were killed, and the lungs were lavaged in situ with 300 μL of PBS. The resultant BAL fluid was analyzed for protein content (Nanodrop 1000 Spectrophotometer; Thermo Fischer), leukocyte May–Grunwald–Giemsa staining of cytospins, and

presence of endogenous IL-33 forms (Western blot). IL-5 cytokine levels in serum were determined using a Cytometric Bead Array Kit (BD Biosciences).

Mapping of Cleavage Sites by MS. To map cleavage sites after processing by neutrophil proteases, recombinant IL-33 was expressed in *E. coli* Rosetta 2 (Novagen) as a GST fusion protein (pGEX-2T vector; GE Healthcare), purified on glutathione-Sepharose beads and digested on beads with either purified cathepsin G (0.25 mU) or neutrophil elastase (0.12 U). Beads were eluted with Laemli buffer, and the resulting fragments, as well as control nondigested GST-IL33, were analyzed by 1D SDS/PAGE. For the cathepsin G experiment, two cleavage products were detected specifically after digestion with the protease and excised from the gel. In the case of elastase digestion, one major cleavage product was detected, and the band was cut from the gel. To identify the corresponding cleavage sites, these three processed fragments were further in-gel digested with specific enzymes into small peptides that were then extracted from the gel and analyzed by MS. Mapping was performed either with trypsin or with endoproteinase Glu-C, to extend the protein sequence coverage and generate MS-detectable peptides in the cleavage region. To confidently assign the N terminus of each processed fragment, a comparative mapping was performed between each cleavage product and the full-length protein excised from the control gel lane. Before digestion, gel bands were washed by cycles of incubation in 100 mM ammonium bicarbonate/acetonitrile (1:1), and cysteine residues were reduced and alkylated in-gel with iodoacetamide. Proteins were then digested by 0.2 μg of modified sequencing grade trypsin (Promega) or 0.4 μg of Glu-C (Sigma) in 50 mM ammonium bicarbonate (5 h at 37 °C). The resulting peptides were extracted from the gel by 10% formic acid/acetonitrile (1:1), dried in a speed-vac (miVac sample concentrator, Genevac), and analyzed by nano-LC-MS/MS using an Ultimate3000 system (Dionex) coupled to an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Peptides were separated on C-18 column (75 μm i.d. × 15 cm; PepMap C18; Dionex) using a 60-min gradient of acetonitrile at 300 nL/min flow rate. The LTQ-Orbitrap Velos (Velos) was operated in data-dependent acquisition mode with the XCalibur software. Survey scan MS were acquired in the Orbitrap on the 300–2,000 *m/z* range with the resolution set to a value of 60,000. The 20 most intense ions per survey scan were selected for MS/MS fragmentation and the resulting fragments were analyzed in the linear trap (LTQ). Dynamic exclusion was used within 60 s to prevent repetitive selection of the same peptide. The Mascot Daemon software (Matrix Science) was used to perform searches against a database containing all *E. coli* entries from Uniprot and the GST-IL33 fusion protein sequence (63,080 total sequences). Carbamidomethylation of cysteines was set as a fixed modification, and oxidation of methionine and protein N-terminal acetylation were set as variable modifications. Specificity of digestion was set for cleavage after K or R for trypsin and after E for Glu-C. Two missed cleavage sites were allowed, as well as semispecific cleavages to enable identification of the N-terminal peptide from processed fragments. The mass tolerances in MS and MS/MS were set to 5 ppm and 0.8 Da, respectively. Only MS/MS peptide sequence matches with Mascot score higher than 20 were considered for identification of the peptides. Quantification of peptides from cleavage products and full-length IL33 was performed using the MFPaQ software.

1. Baekkevold ES, et al. (2003) Molecular characterization of NF-HEV, a nuclear factor preferentially expressed in human high endothelial venules. *Am J Pathol* 163: 69–79.
2. Carriere V, et al. (2007) IL-33, the IL-1-like cytokine ligand for ST2 receptor, is a chromatin-associated nuclear factor in vivo. *Proc Natl Acad Sci USA* 104:282–287.
3. Cayrol C, Girard JP (2009) The IL-1-like cytokine IL-33 is inactivated after maturation by caspase-1. *Proc Natl Acad Sci USA* 106:9021–9026.
4. Tare N, et al. (2010) KU812 cells provide a novel in vitro model of the human IL-33/ST2L axis: functional responses and identification of signaling pathways. *Exp Cell Res* 316: 2527–2537.

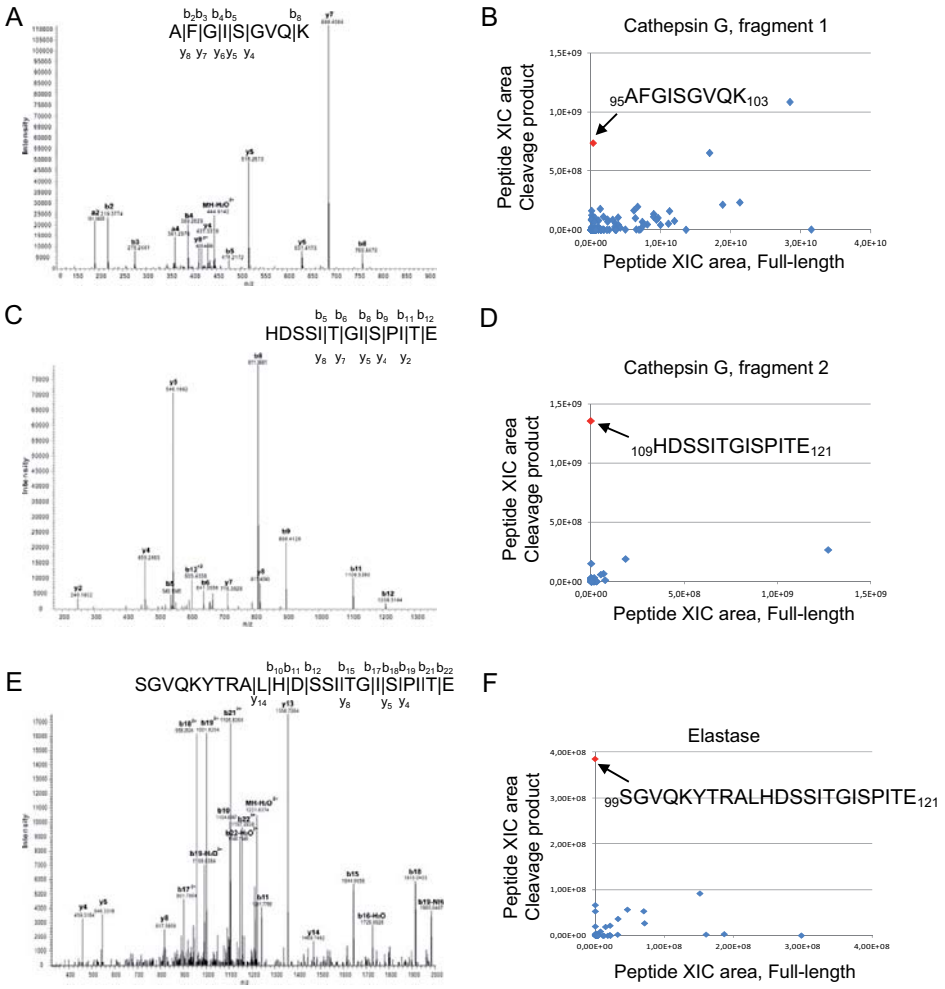


Fig. S1. MS analysis of IL-33 processing by neutrophil elastase and cathepsin G. (*A*, *C*, and *E*) Typical annotated MS/MS spectra identifying N-terminal peptides from IL-33 cleavage products after treatment of purified recombinant IL-33-GST fusion protein with cathepsin G (*A* and *C*) or elastase (*E*). (*B*, *D*, and F) Intensity plots of peptides from cleavage products versus full-length IL-33. Peptides derived from trypsin- or Glu-C-specific enzymatic cleavage of cleavage products or full-length protein were analyzed by MS and quantified based on their extracted ion chromatogram (XIC) area. Specific, highly intense peptides, observed only in the processed fragments and characteristic of the cleavage sites are indicated in red: (*i*) fragment 1 obtained after cathepsin G processing of IL-33, proteomic mapping with trypsin (*B*); (*ii*) fragment 2 obtained after cathepsin G processing of IL-33, proteomic mapping with Glu-C (*D*); and (*iii*) major fragment obtained after elastase processing of IL-33, proteomic mapping with Glu-C (*F*).

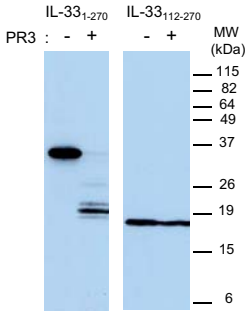


Fig. S2. IL-33 is a substrate for neutrophil proteinase 3. In vitro-translated IL-$33_{1–270}$ and IL-$33_{112–270}$ proteins were incubated with purified neutrophil proteinase 3 (PR3) (4.66 μU/μL; 30 min at 37 °C). Proteins were separated by SDS/PAGE and revealed by Western blot with anti-IL-33-Cter mAb 305B. Blots are representative of at least two independent experiments.

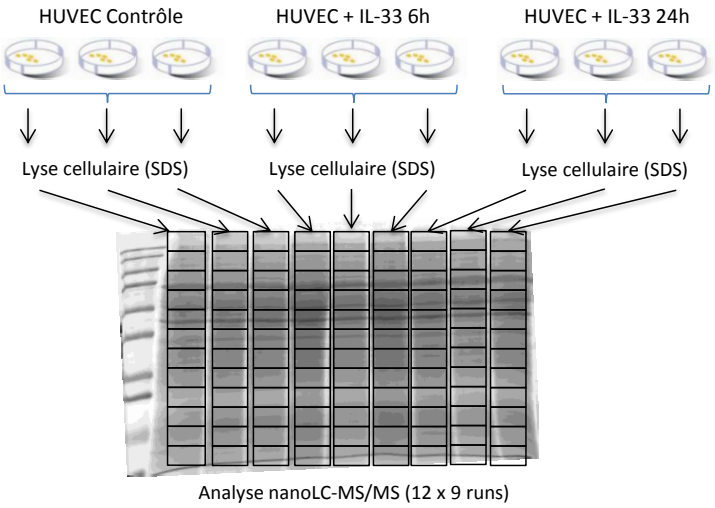
Table S1. List of the specific tryptic or Glu-C fragments expected from the proteomic mapping and containing respectively the two cleavage sites identified for cathepsin G and the main cleavage site identified for elastase

Protease	IL-33 fragment	IL-33 peptide	IL-33 peptide sequence	No. of MS/MS
Cathepsin G, site 1	Full-length IL-33	Peptide 79–103	HLVLAACQQQSTVECFAFGISGVQK	80
		Peptide 95–103	AFGISGVQK	0
	IL-33 cleaved by cathepsin G Fragment 1	Peptide 79–103	HLVLAACQQQSTVECFAFGISGVQK	2
		Peptide 95–103	AFGISGVQK	51
Cathepsin G, site 2	Full-length IL-33	Peptide 93–121	CFAFGISGVQKYTRALHDSSITGISPITE	0
		Peptide 109–121	HDSSITGISPITE	0
	IL-33 cleaved by cathepsin G Fragment 2	Peptide 93–121	CFAFGISGVQKYTRALHDSSITGISPITE	0
		Peptide 109–121	HDSSITGISPITE	34
Elastase	Full-length IL-33	Peptide 93–121	CFAFGISGVQKYTRALHDSSITGISPITE	0
		Peptide 99–121	SGVQKYTRALHDSSITGISPITE	0
	IL-33 cleaved by elastase	Peptide 93–121	CFAFGISGVQKYTRALHDSSITGISPITE	0
		Peptide 99–121	SGVQKYTRALHDSSITGISPITE	11

Although the long peptide encompassing the cleavage site and expected from the tryptic or Glu-C mapping was not always detected for the full-length protein (probably due to low ionization efficiency or poor extraction from the gel), the short peptide corresponding to the N terminus of the processed forms of IL-33 was identified with many MS/MS spectra, specifically during analysis of the cleavage products.

II-2.2 Etude de l'effet de l'IL-33 extracellulaire maturée sur les cellules endothéliales

Une fois relarguée dans le milieu extracellulaire puis maturée, l'IL-33 joue alors un rôle pro-inflammatoire en ciblant différentes cellules immunitaires et les cellules endothéliales. Pour appréhender le rôle extracellulaire de la cytokine et avancer dans la compréhension de ses mécanismes d'action au sein des cellules endothéliales, nous avons étudié l'effet global de l'IL-33 sur celles-ci. Pour cela, 8.10^6 HUVEC ont été stimulées (100 ng/mL) ou non avec l'une des formes maturées que nous avons précédemment mis en évidence, l'IL-$33_{95\text{-}270}$ (Figure 62). Deux stimulations ont été effectuées, pendant 6h et pendant 24h, afin de comprendre au mieux la réponse des cellules endothéliales à la cytokine, et d'identifier les protéines rapidement ou plus tardivement exprimées après cette stimulation. Trois réplicats biologiques issus de boites de culture indépendantes ont été réalisés pour chacune des conditions (8.10^6 cellules par réplicat). Les cellules contrôle, les cellules stimulées pendant 6h et celles stimulées pendant 24h ont été lysées dans un tampon contenant 2% de SDS, puis soniquées. Les extraits protéiques ont ensuite été déposés sur gel 1D SDS-PAGE et fractionnés en 12 bandes et chaque fraction analysée en nanoLC-MS/MS sur un LTQ-Orbitrap Velos (Figure 62). Une analyse quantitative a finalement été réalisée avec MFPaQ selon la méthode de quantification à large échelle optimisée précédemment afin de mettre en évidence les modulations d'expression de protéines induites par l'IL-33.



	CT A	CT B	CT C	IL33_6h_A	IL33_6h_B	IL33_6h_C	IL33_24h_A	IL33_24h_B	IL33_24h_C
Nombre de protéines identifiées par piste	3745	4009	4109	3940	4104	4158	4070	3949	3994
Nombre total de protéines quantifiées	5062								
Protéines variantes (Fold > 2 et p-value < 0,05)				Induction 6h : 32			Induction 24h : 49		

Figure 62 : Stratégie générale et résultats de l'étude protéomique à large échelle des cellules endothéliales HUVEC stimulées ou non par l'IL-3395-270 (stimulation de 6h et 24h). *Trois réplicats biologiques indépendants ont été réalisés pour chaque condition. L'extrait protéique total des cellules obtenu par lyse des cellules au SDS puis sonication est déposé sur gel 1D SDS-PAGE. Chacune des pistes est fractionnée en 12 bandes. Les extraits peptidiques issus de chaque bande de gel est analysé par nanoLC-MS/MS sur un LTQ-Orbitrap Velos. Une analyse quantitative est ensuite réalisée par MFPaQ.*

Cette analyse a permis l'identification et la quantification de plus de 5000 protéines (5062) au sein des cellules endothéliales (Figure 62). Comme précédemment, les protéines ont été définies comme variantes si à la fois leur fold était supérieur à 2 et leur p-value du test de Student inférieure à 0,05. Ainsi, 31 protéines sont apparues variantes après 6h de stimulation avec l'IL-33 dont 25 ont été surexprimées. Une réponse un peu plus importante des cellules endothéliales a été obtenue après 24h de stimulation. L'expression de 49 protéines a en effet été modulée dans ce cas, parmi lesquelles 48 ont été surexprimées. Les listes des protéines variantes dans ces deux conditions de stimulation sont présentées en annexe 7 Le résultat de la quantification obtenu après 24h de stimulation est figuré par un volcano plot (figure 63) représentant la significativité statistique de l'analyse en fonction des variations protéiques entre le contrôle et l'essai.

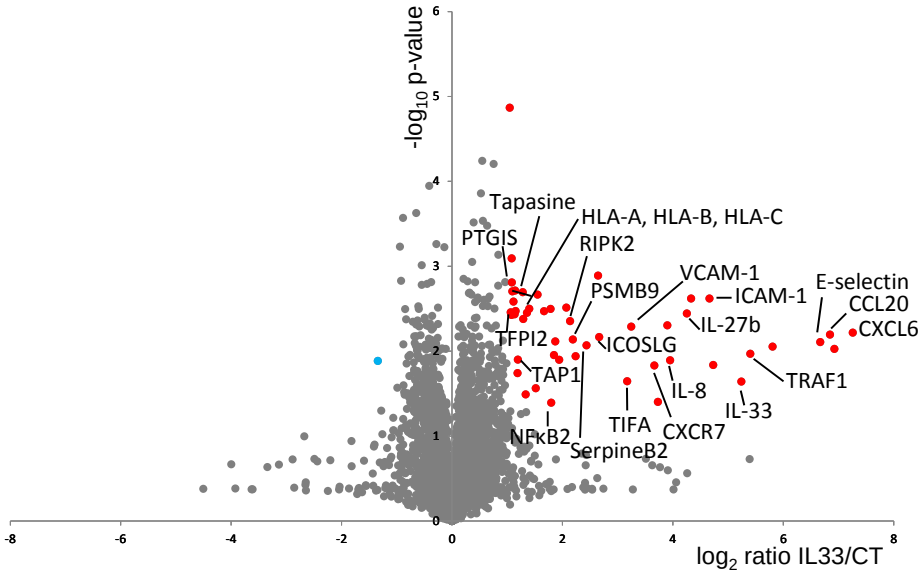


Figure 63 : Représentation graphique de l'analyse quantitative du protéome des cellules endothéliales après stimulation à l'IL-33 pendant 24h*. Volcano plot représentant la significativité des variations d'expression des protéines (p-value du test de student) en fonction des ratios d'expression des protéines entre l'essai et le contrôle. Les points rouge indiquent les protéines surexprimées, les points bleus les protéines sous-exprimées et les points gris les protéines non variantes.*

Parmi les protéines les plus induites par l'IL-33, comme nous l'avions déjà observé pour la stimulation TNFα et IL-1β, de nombreuses sont impliquées dans le recrutement de leucocytes circulants au niveau du site de l'inflammation. Elles présentent également à leur surface des molécules d'adhésion permettant le ralentissement et l'arrêt de ces cellules immunitaires qui sont alors être capables de traverser la barrière endothéliale et d'atteindre le site de l'inflammation. Ainsi, différentes chimiokines comme l'IL-8, CXCL6, CCL20, la galectin-9, sont surexprimées ainsi que des molécules d'adhésion comme la E-sélectine, VCAM-1, ICAM-1 qui jouent un rôle central pour l'interaction leucocytes/cellules endothéliales. D'autres protéines de surface comme les récepteurs CXCR7 et ICOSLG impliqués dans l'activation des cellules endothéliales sont également surexprimées. Par ailleurs, des protéines de la machinerie de maturation et de présentation des antigènes sont également retrouvées (HLA-A, HLA-B, HLA-C, HLA-E, TAP1, Tapasine) ainsi que des immuno-sous-

unités du protéasome (PSMB9, PSMB10) responsables de la génération des peptides cytosoliques présentés par le MHC. De plus, certaines protéines de la voie NFκB sont activées après stimulation à l'IL-33 (comme lors de la stimulation TNFα et IL-1β), telles que RIPK2, TRAF1 et TIFA impliquées dans la transduction du signal permettant l'activation du facteur de transcription NFκB, dont l'une des sous-unités, NFκB2, est également retrouvée induite dans cette étude. Il en effet connu que l'interaction de l'IL-33 avec son récepteur ST2 conduit à l'activation de la voie NFκB (Schmitz, Owyang et al. 2005). La serpine B2 (Plasminogen activator inhibitor 2) présentant une activité pro-coagulante est par ailleurs induite. La régulation de la coagulation sanguine constitue en effet l'un des processus biologiques essentiels assurés par les cellules endothéliales pour le maintien de l'homéostasie mais elle participe également à la défense de l'hôte en parallèle de la réaction inflammatoire (Delvaeye and Conway 2009). Des protéines connues pour être induites par le TNFα (ApoL3, TNFAIP2) sont également surexprimées.

La stimulation pendant 6h à l'IL-33 a quant à elle conduit à une réponse plus restreinte des cellules endothéliales puisqu'un nombre plus limité de protéines est surexprimé (25). La très grande majorité d'entre elles est également modulée après 24h. Ainsi, 22 protéines sont identifiées comme surexprimées dans les deux temps. La comparaison de leurs PAI entre les différentes conditions (Figure 64), permet d'avoir un aperçu de leurs profils d'expression au cours de la stimulation. Deux profils différents ont été observés, correspondant d'une part à des protéines présentant une expression maximum à 6h et d'autre part à des protéines dont l'induction est plus importante à 24h. Par exemple, les protéines impliquées dans la transduction du signal dans la voie NFκB, TIFA et TRAF1, les molécules d'adhésion E-sélectine et VCAM-1, et CXCR7 possèdent un PAI plus élevé à 6h qu'à 24h et semblent donc avoir un pic d'expression dans cette condition. L'expression est ensuite maintenue à un même niveau pour certaines à 24h (TIFA), ou diminuée (E-sélectine, VCAM-1). Au contraire, l'induction de l'expression d'autres protéines augmente au fur et à mesure de la stimulation et est maximum à 24h. C'est par exemple le cas d'ICAM-1 et de la serpine B2, PAI2, et des cytokines et chimiokines IL-27b, CXCL6, CCL20 peu induites après 6h de stimulation. Bien que ces deux uniques points de stimulation ne soient pas suffisants pour caractériser réellement la cinétique de la réponse des cellules endothéliales à l'IL-33, ils permettent d'avoir un premier aperçu des protéines rapidement induites (comme celles de la voie NFκB et certaines molécules d'adhésion permettant le recrutement de leucocytes, induites dès 6h de stimulation) et de celles participant à une réponse plus tardive (comme certaines cytokines et chimiokines dont l'expression parait plus importante au bout de 24h).

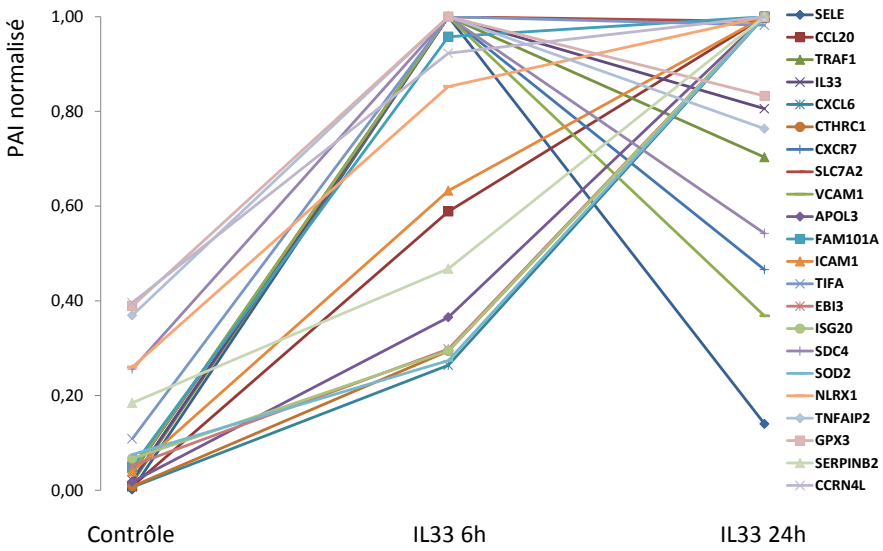


***Figure 64: Profils des PAI des protéines surexprimées à 6h et 24h**. Les valeurs moyennes des PAI des protéines dans chacune des trois conditions ont été normalisées puis représentées sur le graphique.*

Au vu de ces résultats, le phénotype inflammatoire des cellules endothéliales induit par l'IL-33 ressemble fortement au phénotype obtenu après stimulation de ces mêmes cellules avec l'IL-1β, une autre cytokine de la famille IL-1. En effet, 33 protéines ont été identifiées comme surexprimées à la fois par l'IL-33 et par l'IL-1β (cf Annexes 6 et 7). La grande majorité des protéines modulées par l'IL-33 sont de plus également induites par l'IL-1β. Parmi ces variants communs, on retrouve par exemple des cytokines et des chimiokines (IL27-b, CXCL6, CCL20, IL-8) et des molécules d'adhésion (ICAM-1, VCAM-1, E-sélectine, Galectine-9), des protéines de la voie de NFκB (TRAF1, RIPK2 , TIFA, NFκB2) ainsi que des composants de la machinerie de maturation et de présentation des antigènes (TAP1, Tapasine, HLA). Un grand nombre de ces protéines a par ailleurs également été identifié lors de la stimulation des HUVEC par le TNFα et l'IFNγ (cf Annexe 5). La réponse des cellules endothéliales à l'IL-33 semble à première vue plus restreinte que celle provoquée par l'IL1-β. En effet, un nombre beaucoup moins important de variants a été identifié au cours de cette étude (49 variants après 24h de stimulation contre 153 variants lors de l'analyse IL-1β). Les protéines variantes communes ont cependant des ratios d'expression comparables dans les deux expériences et semblent ainsi être induites de façon similaire et à un même niveau par les deux cytokines. Ainsi, après stimulation par l'IL-33 et l'IL-1β respectivement, ICAM-1 présente par exemple un ratio de 25 vs 38, la galectine-9 un ratio de 3 vs 3, l'IL-8 un ratio de 15 vs 10, RIPK2 de 4 vs 2, la tapasine et TAP1, des ratios de 2 vs 3. De plus, il faut noter que la couverture du protéome obtenue lors de l'étude IL-33 est un peu inférieure à celle atteinte lors de l'étude IL-1β (5062 protéines vs 5477 protéines respectivement). Environ 400 protéines supplémentaires, qui correspondent à des protéines très faiblement exprimées dans la cellule, ont pu être quantifiées dans le cas d'IL-1β. Or, de nombreuses protéines spécifiquement surexprimées par IL-1β sont identifiées avec très peu de peptides et représentent des protéines faiblement abondantes de la cellule. Il est donc possible que ces protéines puissent également être modulées par l'IL-33 mais que la profondeur d'analyse obtenue ne nous permette pas de les détecter. Les réponses respectives des cellules endothéliales à l'IL-33 et à l'IL-1β semblent ainsi comparables.

III. Discussion et conclusion

L'analyse de ces protéomes entiers avec la méthode de quantification sans marquage mise en place s'est ainsi révélée performante et utile pour avancer dans la compréhension du fonctionnement de l'IL-33 au sein des cellules endothéliales. Elle nous a permis d'accéder à une large couverture du protéome et de quantifier autour de 5000 protéines grâce au fractionnement de l'échantillon protéique sur gel 1D, tout en conservant une précision de quantification correcte. L'ensemble de cette étude a été réalisée via le logiciel MFPaQ, qui a démontré son efficacité pour quantifier de façon robuste et relativement rapide plusieurs milliers de protéines par extraction automatique des signaux peptidiques associés. Cependant, a posteriori, la comparaison des protéines variantes caractérisées dans le cas des stimulations IL-33, IL1β, et TNFα/INFγ a permis de mettre en évidence la présence de faux-positifs et de faux négatifs dans ces listes de variants. En effet, le logiciel permet à l'utilisateur, via une interface très synthétique, de visualiser les signaux extraits pour chaque ion peptidique dans les différentes conditions, et éventuellement de réaliser une validation manuelle de la quantification en dé-sélectionnant certains peptides mal quantifiés. Ces erreurs de quantification sont diverses et sont liées à la complexité des cartes peptidiques LC-MS (massifs isotopiques chevauchants, interférences, présence de peptides isobariques, non reproductibilité des temps de rétention, signaux de faible intensité…). Cependant, dans le cadre d'une étude à grande échelle portant sur plusieurs milliers de protéines, la validation manuelle des signaux associés aux protéines variantes devient rapidement impraticable. L'application d'un test statistique simple (test de Student) et d'un seuil de ratio pour la définition des protéines modulées nous a permis dans une certaine mesure d'éliminer ces faux-positifs. Malgré tout, un certain taux d'erreur persiste probablement dans les résultats, et il sera donc nécessaire d'optimiser encore le traitement bioinformatique des données, à la fois au niveau de l'extraction des signaux MS et au niveau du traitement statistique (cf Conclusion générale et Perspectives).

Cette méthode de quantification nous a tout de même permis de mettre en évidence plusieurs dizaines de protéines très clairement induites dans les cellules endothéliales suite à la stimulation IL-33. Par ailleurs, l'application de méthodes quantitatives nous a également aidés à caractériser les formes maturées d'IL-33 suite au clivage par des protéases de neutrophiles. Au final, ces données ont contribué à mieux comprendre le rôle et les mécanismes d'action de l'IL-33. Suite à un traumatisme ou à une infection, les cellules endothéliales lésées ou nécrosées pourraient libérer l'IL-33 présente dans leur noyau (Cayrol and Girard 2009). Les neutrophiles sont alors rapidement recrutés au niveau des tissus endommagés et libèrent dans l'espace extracellulaire différentes protéases à sérine, comme l'élastase et la cathepsine G. L'élastase clive alors l'IL-33 pleine taille en une forme maturée l'IL-$33_{99\text{-}270,}$ et la cathepsine G génère l'IL-$33_{95\text{-}270}$ et l'IL-$33_{109\text{-}270}$. Ces nouvelles formes présentent une activité supérieure à la forme pleine taille et permettraient ainsi d'activer fortement la réponse inflammatoire en activant différentes cellules de l'immunité innée et les cellules endothéliales. L'analyse de la réponse de ces cellules endothéliales à la stimulation par l'une des formes clivées de l'IL-33 (IL-$33_{95\text{-}270}$) a en effet permis de mettre en évidence l'induction d'un phénotype pro-inflammatoire très net. La cytokine semble pour cela stimuler la voie NFκB, permettant ainsi d'activer des gènes cibles de la réponse inflammatoire. Elle joue cette façon un rôle dans l'activation de ces cellules pour le recrutement de cellules immunitaires au niveau du site de l'inflammation grâce à la production de chimiokines et à l'expression de molécules d'adhésion. Les mécanismes de maturation et de présentation des antigènes sont également activés dans ces

conditions. Elle pourrait par ailleurs jouer un rôle dans la mise en place d'une régulation de la coagulation sanguine, qui accompagne la réponse inflammatoire. Ce phénotype est comparable à celui obtenu suite à la stimulation des cellules endothéliales avec l'IL-1β. L'IL-33 et l'IL-1β semblent ainsi induire les mêmes voies et les mêmes gènes dans les cellules endothéliales. Bien qu'il soit possible qu'elles jouent parfois un rôle redondant dans certaines conditions physiologiques, elles se distinguent par leur mode de libération et leur mode d'activation qui utilisent des mécanismes différents. Comme décrit précédemment, l'alarmine IL-33 est en effet relarguée dans l'espace extracellulaire suite à des dommages cellulaires des cellules endothéliales et est maturée en formes « super-actives » par des protéases de neutrophiles pour alerter de nombreux acteurs du système immunitaire inné. L'IL-1β est quant à elle produite par les cellules de l'immunité innée suite à la détection de motifs moléculaires portés par les pathogènes (PAMP, « pathogen associated molecular partterns ») ou de signaux de danger ou alarmines (DAMP, « danger associated molecular pattern »). Cette détection est assurée par un complexe multiprotéique de l'immunité innée, l'inflammasome (Gross, Thomas et al. 2011), qui contient la caspase-1. La caspase-1 assure alors la maturation de la forme pleine taille pro-IL-1β, non active, en IL-1β active qui va pouvoir être sécrétée par plusieurs voies non conventionnelles (incluant l'exocytose de lysosomes sécrétoires, le détachement de microvésicules de la membrane plasmique, le relargage d'exosomes et la libération directe au travers de pores (Lopez-Castejon and Brough 2011)) et induire une réponse inflammatoire. Les deux cytokines ne sont ainsi pas libérées ni activées dans les mêmes conditions cellulaires, et doivent de cette façon assurer chacune leur rôle pro-inflammatoire dans des conditions particulières.

Au cours de cette étude, nous n'avons en revanche pas mis en évidence un rôle intracellulaire de l'IL-33. Cela ne nous permet pas pour autant d'exclure totalement qu'elle joue un rôle à l'intérieur de la cellule. Il est en effet possible que les conditions expérimentales dans lesquelles nous nous sommes placées (niveau d'extinction de l'IL-33) ne soient pas idéales pour observer ces effets. De plus, l'analyse réalisée ne nous permet peut-être pas de mettre en évidence les variations d'expression engendrées qui peuvent être faibles et concerner des protéines très peu abondantes.

L'IL-33 est par ailleurs capable de cibler et d'activer de nombreuses cellules de l'immunité innée, comme les mastocytes, les basophiles ou encore les cellules lymphoïdes innées de type 2 (Schmitz, Owyang et al. 2005; Ali, Huber et al. 2007; Cherry, Yoon et al. 2008; Price, Liang et al. 2010; Chang, Kim et al. 2011). Il sera alors intéressant, en perspective de cette étude, d'étudier plus précisément le rôle de la cytokine dans l'immunité innée en mettant en évidence les protéines modulées par l'IL-33 dans ces cellules cibles.

CONCLUSION GENERALE ET PERSPECTIVES

Dans cette thèse, j'ai tenté d'illustrer au travers de différentes applications, l'intérêt des approches de protéomique globale par nanoLC-MS/MS de type « shotgun », permettant de caractériser sans *a priori* des mélanges de protéines. Associées à l'utilisation d'une méthode quantitative sans marquage basée sur l'extraction des signaux MS, ces approches analytiques nous ont permis d'obtenir des informations diverses, à l'échelle de la protéine unique (caractérisation des sites de clivages lors de la maturation d'une cytokine), à l'échelle de complexes protéiques (identification de partenaires d'une protéine appât d'intérêt), ou à l'échelle de protéomes entiers (étude des modulations d'expression des protéines à grande échelle). Tout au long de ma thèse, les méthodes protéomiques ont évolué, avec des instruments de plus en plus performants et des processus bioinformatiques plus élaborés. Malgré cela, les approches de protéomique globale ont à faire face à des défis analytiques énormes, en particulier dans le contexte de la caractérisation de protéomes entiers, ou de la recherche de biomarqueurs. Sur ces mélanges de très grande complexité et de gamme dynamique très large, les limitations de la protéomique « shotgun » ont largement été mises en avant dans la littérature récente, et les approches ciblées ont parallèlement gagné en popularité, principalement grâce à leur meilleure sensibilité. Cette partie a pour objectif de dresser un bilan des études réalisées et d'évoquer les enjeux techniques ainsi que les perspectives apportées par les développements récents dans le domaine de la protéomique.

Caractérisation de complexes et recherche de partenaires : vers des méthodes d'analyse simplifiées et robustes

Si les approches de protéomique globale peuvent être remises en question dans certains contextes d'analyse, un des domaines où elles sont probablement le plus utiles et efficaces concerne la caractérisation de complexes et les expériences d'AP-MS. En effet, ces expériences sont très souvent des études de « découverte » plutôt que des études de « validation », généralement entreprises pour découvrir de nouveaux partenaires non suspectés d'une protéine d'intérêt, et nécessitent donc une caractérisation sans *a priori* des échantillons. De plus, ces échantillons, bien que contenant souvent plusieurs centaines de protéines, sont néanmoins de complexité et de gamme dynamique limitées et accessibles à une analyse efficace par une approche « shotgun ». Il faut noter cependant que cela est directement lié au fait que l'approche comporte une composante biochimique cruciale, qui est l'étape de purification par affinité, et que le succès de ces expériences dépend autant de la qualité des étapes biochimiques que de celles de l'analyse protéomique en elle-même. On peut penser que l'une et l'autre sont complémentaires jusqu'à un certain point : une immunoprécipitation peu efficace n'aboutissant qu'à un enrichissement moyen avec un bruit de fond important demandera une analyse protéomique plus sensible et une quantification efficace pour définir les partenaires probables, alors qu'au contraire, une immunopurification très performante ne nécessitera qu'une analyse MS assez basique de l'échantillon. Ces deux étapes restent cependant essentielles et lorsqu' elles sont efficaces, les approches d'AP-MS sont souvent performantes d'un

point de vue biologique pour élucider des mécanismes et découvrir de nouveaux réseaux d'interactions.

Une partie de ma thèse a été consacrée à l'analyse de complexes impliquant des facteurs nucléaires, les protéines THAP, identifiées dans les cellules endothéliales humaines. Au cours de cette étude, l'emploi de méthodes de quantification sans marquage s'est avéré essentiel pour pouvoir identifier les partenaires spécifiques de plusieurs protéines THAP (THAP1, THAP3, THAP7 et THAP11). Des partenaires spécifiques communs (le facteur de prolifération cellulaire HCF-1 et la glycosyl-transférase OGT) ont pu être mis en évidence et ont à terme permis de mieux comprendre le mécanisme d'action de ces facteurs de transcription dans les cellules endothéliales humaines. Pour cette étude, les complexes immunopurifiés ont été fractionnés sur gel SDS-PAGE en amont de l'analyse nanoLC-MS/MS dans le but d'identifier correctement les protéines faiblement abondantes interagissant éventuellement avec les protéines THAP. A l'heure actuelle, réaliser un tel fractionnement dans ce type d'étude n'est souvent plus nécessaire. Il complique en effet le traitement des données post-acquisition, augmente le temps d'analyse, introduit des biais supplémentaires et donc potentiellement des erreurs dans la quantification, sans pour autant être largement bénéfique pour l'identification des protéines. Les spectromètres de masse récents (tel que le LTQ-Orbitrap Velos) sont en effet suffisamment performants pour identifier les quelques centaines de protéines présentes dans des échantillons immunopurifiés en une analyse unique. Pour ces raisons, l'étude des interactants du facteur général de la transcription TFIIH chez la souris a été réalisée en une seule acquisition nanoLC-MS/MS. Elle a abouti à la découverte de plusieurs nouveaux partenaires, dont la protéine ELL, ce qui a permis d'aller plus loin dans la compréhension des mécanismes cellulaires assurés par TFIIH. Au final, dans cette expérience, le très bon enrichissement obtenu pour le complexe TFIIH, probablement lié à la bonne stabilité du complexe ainsi qu'à l'immunopurification à l'aide d'anticorps très affins de type GFP-trap, a permis de caractériser efficacement les partenaires en un temps d'analyse relativement réduit. Le fait de pouvoir réduire les temps d'analyse a permis de mettre au point les conditions expérimentales de façon plus souple, et de réaliser facilement des réplicats biologiques. En revanche, les analyses nanoLC-MS/MS de ces différents échantillons biologiques ont été réalisées indépendamment, à des intervalles de temps assez longs. Cela pose le problème de la mauvaise reproductibilité des analyses nanoLC-MS/MS sur la durée, qui rend difficile la comparaison directe des signaux MS enregistrés pour ces différents réplicats biologiques. Pour chacun d'entre eux, des analyses quantitatives indépendantes immunopurification/contrôle ont été réalisées, et les listes de protéines variantes ont ensuite été recoupées entre les réplicats biologiques pour définir les protéines partenaires candidates. Il serait plus rapide et plus efficace de traiter ces réplicats d'expériences via des analyses nanoLC-MS/MS rapprochées, ce qui permettrait d'appliquer directement une méthode statistique sur les données quantitatives pour mesurer la significativité des résultats et aider à l'identification des partenaires protéiques *bona fide.* En conclusion, l'analyse nanoLC-MS/MS d'échantillon immunopurifiés apparaît comme un processus relativement simple, rapide, et robuste, qui doit pouvoir renseigner sur l'abondance relative immunopurification/contrôle de l'ensemble des protéines de l'échantillon, non pas seulement de la protéine appât ou de quelques protéines cibles comme un western-blot. L'application de méthodes sans marquage, faciles à mettre en œuvre et clairement suffisantes pour caractériser des variations de grande amplitude, semble idéale pour ce type d'expérience.

Analyse quantitative de protéomes complexes : vers une évaluation objective et une optimisation des méthodes bioinformatiques

Les méthodes de protéomique quantitative globale peuvent également être utilisées pour analyser des protéomes complexes et déterminer les modulations d'expression protéique engendrées par un stimulus donné. Elles doivent alors répondre à des enjeux plus complexes en termes de gamme dynamique et de quantification, et sont souvent concurrencées par d'autres approches globales (notamment transcriptomiques, de type séquençage RNA haut débit de nouvelle génération, offrant une bien meilleure couverture analytique) ou ciblées (notamment de type MRM, offrant une meilleure précision quantitative et une meilleure sensibilité). Elles restent cependant les seules à permettre de quantifier directement des protéines à très grande échelle, et méritent à ce titre d'être développées et optimisées. En dehors de la limitation en gamme dynamique qui demeure un problème important (cf ci-dessous), un défi important réside également dans le traitement bioinformatique des données pour réaliser la quantification.

Dans ce manuscrit, j'ai présenté des données quantitatives exclusivement obtenues à l'aide du logiciel MFPaQ. Au-delà du côté pratique associé à l'utilisation d'un logiciel « maison » (permettant l'organisation et le rapatriement automatique des données de spectrométrie de masse générées au laboratoire, la maîtrise des critères utilisés pour valider préalablement les protéines quantifiées, la possibilité d'implémenter des routines et des macros « à façon » répondant spécifiquement aux besoins de l'utilisateur, etc…), l'extraction de XIC via MFPaQ nous a semblé représenter une approche basique mais efficace pour la quantification des protéines identifiées par nanoLC-MS/MS. C'est donc cette méthode qui a été employée dans le but d'étudier à large échelle les cellules endothéliales dans des conditions inflammatoires induites par différentes cytokines, que ce soit via l'analyse du glycoprotéome ou celle du protéome total. Dans un premier temps, une stimulation « modèle » a été réalisée avec des cytokines bien caractérisées (TNFα/IFNγ), dans le but d'évaluer et de valider la méthode, avant de passer à l'étude du phénotype IL-33. L'analyse des données issues de plusieurs réplicats nanoLC-MS/MS réalisés sur le même échantillon montre que la méthode est globalement assez précise, puisque que sur toute la population de protéines quantifiées, le coefficient de variation médian se situe à 7%. Il faut cependant noter qu'il existe un nombre non négligeable de valeurs de CV extrêmes, correspondant à des protéines pour lesquelles le signal a été extrait de façon non reproductible dans un des réplicats.

Ce problème renvoie à celui des variants faux-positifs et faux-négatifs, évoqué précédemment. Même si le taux global d'erreur (« Family Wise Error Rate », FWER, soit le taux de protéines identifiées variantes à tort par rapport au nombre total de protéines étudiées) est relativement faible, le taux de fausse découverte (« False Discovery Rate », FDR, taux de variants faux-positifs par rapport au nombre de protéines déclarées variantes) est probablement non-négligeable. Il est certainement possible de contrôler ce FDR via l'application de méthodes statistiques plus élaborées et l'acquisition d'un nombre plus important de réplicats. Cependant, il est important de pouvoir l'estimer et d'identifier les sources d'erreur afin de pouvoir améliorer les méthodes. Si on considère les données brutes enregistrées par le spectromètre de masse, il est possible que de nombreux peptides soient non quantifiables, par exemple à cause d'interférences avec une d'autres espèces du fait de la grande complexité des mélanges peptidiques et de l'encombrement sur une carte LC-MS. Cependant, de nombreuses erreurs sont également dues au traitement bioinformatique lui-même, et il arrive que certaines protéines soient mal quantifiées alors

que les données existantes pourraient permettre une quantification correcte (typiquement lorsqu'une erreur d'extraction est commise sur un des peptides de la protéine, alors que les autres peptides sont correctement quantifiés). Il est difficile d'évaluer *a priori* l'étendue de ce type de problème sur une étude à grande échelle, dans la mesure où les variants biologiques réels ne sont pas strictement définis. La présence de faux-positifs ou de faux-négatifs a pu être mis en évidence lors de la comparaison des résultats obtenus sur la stimulation des HUVEC avec les différentes cytokines (TNFα/IFNγ, IL-1β, IL-33), où des résultats non cohérents sur certaines protéines caractérisées dans les différentes expériences ont été corrigés manuellement, après vérification visuelle des signaux enregistrés. De plus, la connaissance du contexte biologique et les données de la littérature ont permis de repérer des protéines de l'inflammation mal quantifiées et déclarées non-variantes, ainsi que des protéines *a priori* non pertinentes dans la réponse inflammatoire, et définies à tort comme variantes. Ce type d'erreur reste marginal, mais il est important de les détecter afin d'identifier les problèmes et d'améliorer les traitements bioinformatiques (extraction du signal, contrôle de la qualité des XIC, sélection des peptides quantifiables, élimination des cas aberrants, etc…). A ce titre, le travail réalisé sur ces études quantitatives nous a permis d'identifier certains défauts du logiciel MFPaQ, qui pourront donc être évités dans les nouveaux outils en développement dans l'équipe.

Enfin, au-delà de la détection d'erreurs ponctuelles propres à un outil bioinformatique particulier, il apparaît important de pouvoir mesurer précisément et objectivement les performances de différentes méthodes bioinformatiques, au travers de métriques quantifiables, comme la sensibilité ou le FDR. Une telle évaluation est rarement réalisée dans la littérature. Elle nécessite en effet d'une part de connaître précisément les protéines variantes de l'échantillon, et d'autre part, de disposer d'un échantillon contenant un nombre suffisamment élevé de variants pour avoir une évaluation statistique des erreurs possibles. Initialement, l'expérience réalisée en stimulation TNFα/IFNγ était censée représentée une situation modèle bien caractérisée, et nous a servi de support pour comparer sommairement MFPaQ avec d'autres logiciels d'analyse quantitative efficaces comme Progenesis LC-MS. Cependant, même sur une situation de ce type, si on peut facilement valider les variations majeures attendues (par exemple, VCAM1, CMH, SélectineE, etc…), il est difficile de définir avec certitude les vrais positifs et les vrais négatifs au voisinage des valeurs seuil, et de calculer précisément un FDR. Des études comparatives sont à présent en cours dans l'équipe pour comparer différentes méthodes et logiciels d'analyse quantitative sur la base de mesures fiables de FDR et de sensibilité, réalisées à partir de mélanges standards complexes et bien formatés. De telles évaluations paraissent importantes pour définir les meilleures méthodes bioinformatiques, mais aussi optimiser la performance des étapes analytiques elles-mêmes (fractionnement, HPLC, analyse MS), qui peuvent également influencer les données quantitatives finales.

En conclusion, le traitement bioinformatique est un facteur clé dans les approches protéomiques globales. Les outils peuvent encore être améliorés, et même si des logiciels paraissent vraiment efficaces (LC-Progenesis, MaxQuant, etc…), ils nécessitent une évaluation objective portant sur leur capacité à générer des résultats corrects dans une étude différentielle à grande échelle (sensibilité, FDR), mais également sur d'autres fonctionnalités importantes, comme la capacité à gérer les expériences en fractionnement, la vitesse d'exécution, ou la capacité à traiter des fichiers de plus en plus volumineux et nombreux.

L'enjeu de la couverture analytique des protéomes : vers une optimisation des conditions HPLC

Certaines erreurs de quantification évoquées ci-dessus peuvent en partie être liées à l'étape de fractionnement de l'échantillon au niveau protéique. Au début de cette étude, ce fractionnement nous est apparu indispensable et constituait probablement la solution la plus efficace pour couvrir au mieux le protéome des cellules endothéliales. Les erreurs quantitatives associées à la migration parallèle sur gel 1D ont en grande partie été corrigées via les procédures de normalisation et d'intégration implémentées dans MFPaQ. Au final, les résultats présentés ici montrent qu'il est possible par ces approches d'atteindre une profondeur d'analyse importante du protéome (5000 à 5500 protéines sont généralement quantifiées dans les études de stimulation HUVEC par les différentes cytokines) et de fait, d'accéder à la quantification d'un grand nombre de protéines minoritaires. Cependant, ces résultats ont été obtenus au prix d'un temps d'analyse très important (typiquement 8 jours pour l'analyse comparative de 2 conditions, avec 3 réplicats fractionnés sur des pistes parallèles en 12 bandes).

Des études récentes suggèrent par ailleurs qu'une alternative possible à ce pré-fractionnement extensif consiste à améliorer la séparation HPLC des peptides, grâce à la réalisation de la séparation chromatographique sur des colonnes longues (50 cm) avec des gradients longs (8h). Cela conduit à une meilleure distribution des peptides tout au long du gradient chromatographique, sans pour autant élargir les pics d'élution. Bien qu'elles ne constituent pas encore des méthodes de routine, ces approches deviennent plus faciles à mettre en œuvre grâce à la généralisation de pompes nano-débit de type UPLC. La robustesse et la reproductibilité de ces méthodes reste cependant encore à évaluer. Elles nécessitent également un certain savoir-faire pour préparer des colonnes chromatographiques de grande taille performantes et reproductibles. Cependant, elles apparaissent prometteuses et permettent aujourd'hui au laboratoire d'identifier près de 4500 protéines avec une analyse de 8h. Au final, cela permettrait de réduire par 3 ou 4 le temps d'analyse par rapport à une approche avec pré-fractionnement des protéines, et de réaliser en 2 jours environ une analyse quantitative portant sur deux conditions. Ce type de méthode génère en revanche des données très lourdes que les outils bioinformatiques communément utilisés peuvent avoir du mal à traiter. Pour contourner ce problème, des formats de fichiers optimisés, comme le format mzDB actuellement développé dans l'équipe, peuvent être utilisés. Ce format permet une diminution très significative du temps d'analyse informatique, et facilite considérablement le traitement des données volumineuses.

En conclusion, de telles approches semblent crédibles pour quantifier en une seule acquisition, de façon potentiellement assez robuste et rapide, des protéomes complexes à une profondeur d'environ 5000 protéines. De plus, l'utilisation dans ces conditions, de spectromètres de masse dernière génération (de type Q-Exactive par exemple) devrait permettre d'encore augmenter la couverture analytique. Ces performances apparaissent encore théoriquement inférieures à celles du séquençage d'ARN de nouvelle génération, mais permettent d'avoir un accès direct à l'expression des protéines. De plus, les études sur la réponse inflammatoire des cellules HUVEC présentées ici indiquent que même à une profondeur de « seulement » 5000 protéines, l'analyse protéomique permet de capturer une part non négligeable des événements biologiques à l'œuvre dans la cellule, et offre donc des possibilités intéressantes pour la caractérisation des systèmes biologiques.

Méthodes globales, méthodes ciblées, et méthodes hybrides

Les approches MS globales permettent ainsi d'accéder à une partie importante des protéomes et cela, sans *a priori*, justifiant donc leur utilisation pour la caractérisation de systèmes biologiques. Cependant, en dépit des améliorations constantes des spectromètres de masse la complexité et la gamme dynamique des protéomes humains excèdent encore à l'heure actuelle les capacités des systèmes LC-MS/MS actuels en mode « shotgun ». Ces limites sont en partie liées au mode d'acquisition des données employé qui est « data-dependent » (mode DDA, « data dependent acquisition »), dans lequel seuls les peptides les plus intenses sont sélectionnés de façon non reproductible pour être séquencés un à un. Au contraire, la protéomique ciblée est adaptée pour détecter et quantifier de façon reproductible et précise des protéines. Cependant, elle souffre d'un certain manque de spécificité et elle se limite à la mesure d'un nombre relativement réduit de protéines qui doivent de plus être définies au préalable. Alternativement, de nouvelles méthodes hybrides sont actuellement proposées et se basent sur un mode d'acquisition indépendant des données, appelé DIA (« data independent acquisition »), dans lequel plusieurs peptides sont séquencés simultanément (Venable, Dong et al. 2004; Plumb, Johnson et al. 2006; Panchaud, Scherl et al. 2009; Gillet, Navarro et al. 2012). Elles consistent à enchainer, sans aucune sélection, des cycles de MS et MS/MS de tous les peptides précurseurs co-éluant de la colonne et contenus dans une fenêtre de masse définie. Des données de séquençage sont ainsi en théorie disponibles pour toutes les espèces peptidiques présentes dans l'échantillon, ce qui semble permettre d'augmenter la gamme dynamique analytique. Cependant, les données résultant de ce type d'analyse consistent en des spectres MS/MS complexes, contenant les données de fragmentation de plusieurs précurseurs. Il faut alors encore parvenir à correctement relier ces informations au précurseur dont elles sont issues, ce qui nécessite des outils bioinformatiques performants et adaptés. Ces stratégies semblent néanmoins prometteuses et il sera intéressant de suivre leur évolution dans les années à venir. Ces différentes méthodes de protéomique (globales, ciblées ou hybrides) présentent ainsi chacune leurs spécificités, leurs avantages ainsi que leurs défauts, et sont plus ou moins bien adaptées pour des applications différentes. Elles doivent donc être exploitées judicieusement en tirant partie de leurs atouts respectifs afin de répondre à des questions biologiques diverses. Ces différentes méthodes vont continuer à évoluer en parallèle des instruments et des améliorations bioinformatiques, et ainsi permettre d'accéder à des protéines très minoritaires de la cellule et de les quantifier de façon plus reproductible et plus précise permettant *in fine* une meilleure compréhension des mécanismes moléculaires et des processus biologiques complexes des cellules .

MATERIEL ET METHODES

I. Préparation et analyse protéomique des complexes THAP

Cultures cellulaires et induction de l'expression de la protéine THAP

Culture cellulaire : Des cellules HeLa Tet-Off surexprimant de façon conditionnelle la protéine THAP-FLAG-HA d'intérêt sont cultivées en milieu DMEM Glutamax (4,5 g/mL de glucose) (Invitrogen), avec 10 % de Sérum de Veau Fœtal (SVF), 100 U/mL de pénicilline et 100 µg/mL de streptomycine, 100 µg/mL de G418 et 50 µg/mL d'hygromycine. 1 µg/mL de doxycycline est ajoutée toutes les 48h dans le milieu de culture pour bloquer le système de surexpression de THAP.

Synchronisation cellulaire : La synchronisation des cellules HeLa (réalisée uniquement pour l'étude des complexes THAP1) a été réalisée par double blocage thymidine de la façon suivante : 2 mM de thymidine sont ajoutés au milieu de culture sur les cellules (60% de confluence) pendant 18h (1^{er} blocage en G1/S). Les cellules sont lavées deux fois au PBS puis cultivées dans un milieu de culture frais pendant 12h. 2mM de thymidine sont à nouveau rajoutés au milieu pendant 18h (2^{nd} blocage en G1/S). Les cellules sont enfin relarguées, après 2 lavages au PBS, dans un nouveau milieu pendant 4h (cellules en phase S).

Induction de l'expression des protéines THAP et vérification de l'induction : L'expression de la protéine THAP est induite, pendant 96h, dans les cellules HeLa par levée du blocage doxycycline en réalisant deux séries de deux lavages successifs au moment de l'induction et 24h après l'induction. L'induction est ensuite vérifiée par immunofluorescence. Les lamelles préalablement introduites dans les boites de culture sont lavées au PBS, fixées dans 3.7% de formaldéhyde (20min), rincées au PBS, perméabilisées avec 0.2% triton X100 (5min), à nouveau rincées au PBS avant d'être incubées avec 1.5% BSA (2h). Les cellules sont ensuite incubées avec un anticorps primaire souris anti-HA (B12) ($1/1000^{ème}$ dans 1.5% BSA) pendant 2h, et après lavage, avec un anticorps secondaire anti-souris couplé Cy3 ($1/1000^{ème}$). Les lamelles sont marquées au Hoechst 0.2µg/mL, montées sur lame et observées au microscope.

Purification des complexes protéiques THAP

Protocole 1 :

Extraction nucléaire : Les culots de cellules congelés sont repris dans 1,5 mL de tampon B0 (20 mM Tris pH 7,4, 0,5 mM MgCl2, 0,5 % NP40, 10 mM 2-mercaptoéthanol, 1 mM PMSF, Complete 1X = inhibiteurs de protéases) puis les cellules sont cassées à l'ultraturax (position 1, 2 fois 8 sec avec 30s entre les deux pulses). Après centrifugation (10 min, 800 g, 4°C), le culot de noyaux est rincé avec 2 volumes de tampon B0. Il est ensuite repris dans le tampon B0,42 (Tampon B0 + 0,42 M NaCl + 20% glycérol) puis placé sur glace pendant 30 min et vortexé toutes les 5 min. Après centrifugation (800g, 10 min, 4°C), le surnagent constituant l'extrait nucléaire est récupéré puis ultracentrifugé 30 min à 40000 rpm.

Fractionnement des extraits nucléaires sur gradient de glycérol : Les gradients de glycérol sont coulés dans des tubes Beckman 4 mL à l'aide d'une pompe péristaltique alimentée par deux solutions : l'une à 10 % glycérol (20 mM Tris HCl pH 7,4, 0.15 M NaCl, 0,5 M MgCl2, 10 % glycérol, 10 mM 2-mercaptoéthanol, 0 ou 0,5 % NP40, 1 mM PMSF) et l'autre à 40 % glycérol (20 mM Tris HCl pH 7,4, 0,15 M NaCl, 0,5 M MgCl2, 10 % glycérol, 10 mM 2-mercaptoéthanol, 0 ou 0,5 % NP40, 1 mM PMSF). Les extraits sont chargés en haut du gradient de glycérol qui est ensuite soumis à une ultracentrifugation à 35000 rpm pendant 15h30 à 4°C. Des fractions de 200 µL sont récoltées par le haut du gradient et conservées à -80°C. Ces fractions sont ensuite analysées par Western Blot.

Immunoprécipitation anti-FLAG: 200 µL de billes agarose greffées avec l'anticorps murin anti-FLAG M2 (Sigma) sont lavées à deux reprises avec 1 mL de tampon B, puis avec 1 mL de 0,1 M glycine pH3, puis à nouveau à deux reprises par 1 mL de tampon B. Les billes sont mises en présence de l'extrait nucléaire et incubées par rotation sur roue sur la nuit à 4°C. Après deux lavages, deux élutions successives des protéines retenues sur les billes sont alors réalisées par incubation des billes pendant 30 min avec 100 µL de peptide compétiteur 3XFLAG 0,5 mg/mL. Les deux éluats sont rassemblés.

Protocole 2 :

Extraction nucléaire : Les culots de cellules sont repris dans 5 volumes de tampon A (10 mM HEPES pH 7,4, 1,5 mM MgCl2, 10 mM KCl, Complete 1X, PhoStop 1X), laissés sur glace 10 min, puis cassées à l'ultaturax (position 1, 3 fois 8 sec avec 30 sec entre chaque broyage). Les noyaux sont ensuite culotés par centrifugation (10 min, 1000g, 4°C) puis lavés successivement par 3 volumes puis 1 volume de tampon B (10 mM HEPES pH 7,4, 1,5 mM MgCl2, 10 mM KCl, Complete 1X, 25% glycérol). Les noyaux sont finalement culotés par centrifugation (10000g, 10min, 4°C) et une première extraction nucléaire est réalisée en incubant le culot de noyaux avec 2/3 de volume de tampon C (20 mM HEPES pH 7,4, 0,42 M NaCl, 1,5 mM MgCl2, Complete 1X, 25% glycérol) pendant 30 min sur vortex (1200 rpm, 4°C). L'extrait nucléaire 1 est alors récupéré après centrifugation (15000 g, 30 min, 4°C). Le culot de noyaux est ensuite repris dans 2 volumes de tampon D (20 mM HEPES pH 7,4, 0,5 M NaCl, 1,5 mM MgCl2, Complete 1X, 25% glycérol) pour une seconde extraction et incubé pendant 30 min à 4°C sur vortex à 1200 rpm. L'extrait nucléaire 2 est récupéré après centrifugation (15000 g, 30 min, 4°C). Les deux extraits nucléaires sont ensuite ultracentrifugés à 50000 rpm pendant 30 min à 4°C puis dilués dans de l'HEPES 50 mM de façon à ramener la concentration saline de chacun d'entre eux à 0,15 M NaCl final et rassemblés.

Double immunopurification FLAG-HA: 200 µL de billes agarose greffées avec l'anticorps murin anti-FLAG M2 (Sigma) sont lavées par 2mL de 0,1 M glycine pH 3 puis rincées par 2mL de tampon Wash (50 mM Tris pH 7,4, 0,15 M NaCl, 10% glycérol, Complete 1X). Les extraits nucléaires sont incubés avec les billes sur roue pendant 4h à 4°C. Après deux lavages des billes avec 2mL de tampon Wash, deux élutions successives sont alors réalisées par incubation des billes pendant 30 min avec 200 µL de peptide compétiteur 3XFLAG à 150 µg/mL dans le tampon Wash. Les éluats sont rassemblés puis déposés sur les billes anti-HA préalablement lavées (successivement : 1 lavage glycine 0,1 M, 1 lavage avec tampon Wash) pour une incubation de 4h sur roue à 4°C. Après deux lavages des billes au tampon Wash, les protéines retenues sont éluées soit par reprise des billes dans 2 volumes de peptide compétiteur HA (1mg/mL dans le tampon Wash) et incubation 15min à 37°C, soit par reprise des billes dans 2 volumes de tampon de type Laemmli sans réducteur (80mM Tris pH 6.8, 10% glycérol, 4% SDS) et chauffage à 90 °C pendant 5 min.

Analyse protéomique des complexes THAP immunopurifiés

Electrophorèse SDS-PAGE et digestion trypsique : Après réduction des ponts disulfures et alkylation des cystéines à l'iodoacétamide, les complexes protéiques purifiés sont déposés et fractionnés sur gel de polyacrylamide 12%. Le gel est coloré au bleu de Commassie et chaque piste est découpées en une dizaine de bandes puis découpées en cubes de 1 à 2 mm². Les morceaux de gel ainsi obtenus sont lavés à l'eau milliQ, décolorées par déshydratations/hydratations successives par de l'acétonitrile (ACN) puis par un mélange 50/50 (v/v) bicarbonate d'ammonium 100mM/ACN. Les bandes sont ensuite séchées au Speed-Vacum, puis digérées par réhydratation dans une solution de trypsine à 10 µg/mL dans du bicarbonate d'ammonium (BA) 100mM, à 37°C sur la nuit. Le surnageant de digestion est prélevé, et les peptides restants sont extraits du gel par extractions successives (suivies chacune d'une sonication) au BA 100mM, puis par un mélange 50/50 (v/v) acide formique 10%/ACN. Les différents extraits peptidiques sont réunis puis séchés au Speed-Vacuum et conservés à -20°C.

Analyse nanoLC-MS/MS : L'analyse protéomique des échantillons est réalisée par nano-LC-MS/MS grâce à un système de nano-HPLC (U3000, Dionex) couplé à un spectromètre de masse (LTQ-Orbitrap, Thermo Fischer Scientific). Les extraits peptidiques secs sont repris dans 14 µL de solvant A' (5% ACN, 0.05% TFA) et 5 µL sont injectés sur une pré-colonne de phase inverse C18 (300 µm ID x 5 mm, Dionex) à un débit de 20 µL/min. Après 5min de dessalage, la pré-colonne est basculée en ligne avec une colonne analytique de phase inverse C18 (75 µm ID x 15 cm PepMap C18, Dionex) équilibrée avec 95 % de solvant A (5% ACN, 0,2 % FA) et 5% de solvant B (80% ACN, 0,2% FA). L'élution des peptides est réalisée par un gradient croissant de 5% à 50 % de solvant B sur 60 min à un débit de 300nL/min. L'acquisition, pilotée par le logiciel Xcalibur, est réalisée en mode « Data Dependant Acquisition ». Les scans MS sont enregistrés dans l'analyseur Orbitrap avec une résolution de 60000 à m/z=400. Les 5 ions les plus intenses de chaque scan MS sont sélectionnés pour subir en parallèle une analyse MS/MS dans la trappe linéaire LTQ (fragmentation CID). Un temps d'exclusion dynamique de 60 sec a été utilisé.

Recherche dans les bandes de données et validation : Les recherches en banques de données ont été réalisées avec le logiciel Mascot Deamon (version 2.2.0, Matrix Science) en utilisant la macro Extract_msn.exe de Xcalibur (Thermo Fisher Scientific) pour générer les peaklists. L'interrogation a été réalisée pour le genre Homo sapiens dans la banque de données SwissProt-Trembl en incluant des modifications fixes (carbamidométhylation des cystéines) et variables (oxydation (méthionine), phosphorylations (sérine, thréonine), O-GlcNAc (sérine, thréonine)). Les tolérances de masse en mode MS et MS/MS ont été respectivement fixées à 5 ppm et 0,8 Da. La recherche a été effectuée avec la possibilité de deux clivages manqués et pour des peptides de charges 2+ et 3+. Les protéines identifiées ont ensuite été validées automatiquement avec MFPaQ si elle possède au moins un peptide (séquence d'au moins 6 acides aminés) de rang 1 présentant un p-value Mascot < 0,01 (score Mascot > 31) ou deux peptides (séquence d'au moins 6 acides aminés) de rang 1 présentant une p-value Mascot < 0,05 (score Mascot > 24). *Quantification avec MFPaQ* : Une analyse quantitative différentielle des pistes « Essai » versus « Contrôle » a été réalisée avec le module de quantification sans marquage de MFPaQ qui part des résultats d'identification validés pour aller extraire les XICs des peptides correspondants dans les fichiers bruts à comparer. Il apparie dans un premier temps les peptides identifiés validés dans toutes les différentes acquisitions MS, puis les utilise pour aligner ces acquisitions MS en temps de rétention. A partir de la matrice d'alignement en temps de rétention

obtenu, le logiciel prédit le temps de rétention des ions peptidiques dans les analyses où ils n'ont pas été identifiés par MS/MS, et extrait le signal correspondant à partir du temps de rétention prédit, et de la masse exacte du peptide. Les XICs des peptides identifiés sont ainsi extraits dans chacun des fichiers bruts et sont utilisés pour réaliser la quantification de chaque peptide. MFPaQ calcule ensuite pour chaque protéine un indice d'abondance protéique, le PAI (« Protein Abundance Index ») qui correspond à la moyenne des intensités des trois peptides les plus intenses d'une protéine (définis sur l'ensemble des conditions comparées, et identiques entre ces différentes conditions). Ces PAI sont ensuite utilisés pour calculer le ratio d'abondance relative d'une même protéine dans les différentes conditions.

II. Préparation et analyse protéomique des complexes TFIIH

Les conditions et les méthodes de préparation et d'analyse protéomique des complexes TFIIH sont décrites dans l'article Mourgues, Gautier et al., en préparation (cf Résultats, Partie I.II et Annexe 2).

III. Enrichissement des glycoprotéines de surface et analyse protéomique

Culture cellulaire, marquage et enrichissement des glycoprotéines

Les HUVEC (Primary human umbilical vein ECs) (Clonetics) ont été cultivées dans un milieu ECGM (Promocell) en présence de 20 % de sérum de veau fétal. Les cellules ont été marquées par ajout dans le milieu de culture de Ac_4ManNAz (25 µM) pendant 48h et stimulées ou non avec du TNF-α (25 ng/ml, R&D Systems) et IFNγ (50 ng/ml, R&D Systems) pendant 12h. Les cellules ont ensuite été récoltées en grattant les boites en présence de PBS, lavées deux fois avec 10 mL de PBS puis incubées avec 250 µM de phosphine biotine sur roue pendant 1h à température ambiante. Après deux lavages au PBS, le culot de cellules est repris dans 2 mL de SDS 1% et soniqué. Les extraits protéiques sont dilués à 0,2% SDS et incubés 3h sur roue à température ambiante en présence de billes de streptavidine (200µL) préalablement lavées dans un tampon 0,2% SDS. Les billes sont ensuite lavées pendant 15min 3 fois avec un tampon 0,2% SDS, 3 fois avec 3M NaCl et 3 fois avec du PBS.

Digestion protéique

Les glycoprotéines liées aux billes de streptavidine ont été réduites en présence de 25 mM DTT, 50 mM ammonium bicarbonate, 6M urée pendant 30min à 56°C puis alkylées avec 90mM d'idoacétamide pendant 30min à température ambiante, dans le noir. Les billes sont lavées dans un tampon ammonium bicarbonate 50mM avant d'être digérées dans une solution de trypsine à 0,1 µg/µL sur la nuit à 37°C. Le surnagent contenant les peptides trypsiques est récupéré. Les billes sont à nouveau lavées 2 fois avec un tampon 50mM ammonium bicarbonate. Les N-glycopeptides encore liés aux billes sont élués grâce à une digestion PNGase F (Sigma) (5 unités) pendant 3h à 37°C.

Analyse protéomique

Analyse nanoLC-MS/MS : Les mélanges peptidiques (trypsiques ou N-glycopeptides) ont été analysés par nanoLC-MS/MS sur un système Ultimate3000 (Dionex) couplé à un LTQ-Orbitrap Velos (Thermo Fisher Scientific). 5 µL de chaque échantillon sont injectés sur une pré-colonne de phase inverse C18 (300 µm ID x 5 mm, Dionex) dans 5% acétonitrile et 0,05% acide trifluoroacétique à un débit de 20 µL/min. Après 5min de dessalage, la pré-colonne est basculée en ligne avec une colonne analytique de phase inverse C18 (75 µm ID x 15 cm PepMap C18, Dionex) équilibrée avec 95 % de solvant A (5% ACN, 0,2 % FA) et 5% de solvant B (80% ACN, 0,2% FA). L'élution des peptides est réalisée par un gradient croissant de 5% à 50 % de solvant B sur 120 min à un débit de 300nL/min. L'acquisition, pilotée par le logiciel Xcalibur, est réalisée en mode « Data Dependant Acquisition ». Les scans MS sont enregistrés dans l'analyseur Orbitrap avec une résolution de 60000 à m/z=400 sur la gamme de masse m/z 300-2000. Les 20 ions les plus intenses de chaque scan MS sont sélectionnés pour subir en parallèle une analyse MS/MS dans la trappe linéaire LTQ (fragmentation CID). Un temps d'exclusion dynamique de 60 sec a été utilisé.

Recherche en banque de données et validation des identifications : Les recherches en banques de données ont été réalisées avec le logiciel Mascot Deamon (version 2.2.0, Matrix Science) en utilisant la macro Extract_msn.exe de Xcalibur (Thermo Fisher Scientific) pour générer les peaklists. L'interrogation a été réalisée dans la banque IPI human v3.72 (86392 sequences) en incluant la carbamidométhylation des cystéines, l'oxydation des méthionines, l'acétylation N-terminale des protéines (et la déamidation des asparagines dans le cas de l'interrogation des peptides glycosylés) comme modifications variables. La spécificité de clivage de la trypsine a été fixée après K ou R sauf avant P, et un clivage manqué a été autorisé. Les tolérances de masse en mode MS et MS/MS ont été respectivement fixées à 2 ppm et 0,6 Da. Les recherches ont été réalisées en utilisant l'option « decoy » de Mascot pour pouvoir calculer un taux de faux-positifs. La validation des peptides et protéines a été réalisée avec le logiciel Prosper, développé dans l'équipe. Les identifications peptidiques issues des résultats Mascot ont été validées à un FDR peptidique final de 5% et les protéines de 1%. Un filtre supplémentaire a été appliqué aux N-glycopeptides, validés s'ils contenaient au moins 1 asparagine déamidée dans le motif consensus NXS/T où X n'est pas une proline.

Quantification avec MFPaQ : la quantification sans marquage a été réalisée avec le module de quantification sans marquage de MFPaQ de la même façon que décrit dans la partie I du Matériel et Méthodes. De plus, pour chaque acquisition, un facteur de normalisation a été appliqué sur les valeurs de PAI, basé sur la somme de tous les signaux identifiés dans un run par rapport à un run de référence.

IV. Préparation et analyse protéomique quantitative des protéomes entiers des cellules endothéliales

Culture et différents traitements des HUVEC

Culture cellulaire : Les conditions de culture des HUVEC sont décrites dans la section Matériels et Méthodes de l'article (Gautier, Mouton-Barbosa et al. 2012) (cf Résultats, Partie II-II).

Stimulation au TNFα/IFNγ et à l'IL-1β : Les conditions de stimulation des cellules endothéliales avec le TNFα/IFNγ, et l'IL-1β sont également détaillées dans l'article (Gautier, Mouton-Barbosa et al. 2012) (cf Résultats, Partie II-II).

Extinction de l'expression de l'IL-33 par siRNA : Deux transfections successives de 6h avec 50 nM d'un pool de siRNA IL-33 ou de siRNA contrôle (ON-TARGET*plus* SMARTpool, Thermo Scientific/Dharmacon) ont été réalisées à 24h d'intervalle sur les HUVEC en présence d'oligofectamine (Invitrogen) dans un milieu de culture sans sérum OptiMEM (Invitrogen).

Stimulation à l'IL-$33_{95\text{-}270}$: Les HUVEC ont été stimulées avec 100 ng/mL de l'IL-$33_{95\text{-}270}$ recombinante purifiée produite dans *E.coli* pendant 6h ou 24h dans un milieu de culture ECGM (Promocell).

Préparation des extraits protéiques : les cellules ont été lysées directement sur boite de culture par ajout de 2% SDS puis ont été soniquées.

Gel 1D SDS-PAGE et analyse protéomique des protéomes entiers des cellules endothéliales

Les extraits protéiques issus de la lyse des cellules endothéliales après les différents traitements (stimulation par le TNFα/IFNγ, après stimulation avec l'IL-1β, après extinction de l'expression de l'IL-33 par siRNA et après stimulation avec l'IL-$33_{95\text{-}270}$) ont été traitées et analysés selon un même protocole (fractionnement sur gel 1D, digestion trypsique, analyse nanoLC-MS/MS sur un LTQ-Orbitrap Velos (Thermo Fisher Scientific) couplé à une nanoHPLC (U3000, Dionex), interrogation des banques de données, analyse quantitative sans marquage avec MFPaQ) décrit dans la section Matériels et Méthodes de l'article (Gautier, Mouton-Barbosa et al. 2012) (cf Résultats, Partie II-II).

LISTE DES PUBLICATIONS

1. Raoul Mazars, Anne Gonzalez-de-Peredo, Corinne Cayrol, Anne-Claire Lavigne, Jodi L. Vogel, Nathalie Ortega, Chrystelle Lacroix, **Violette Gautier**, Gaelle Huet, Aurélie Ray, Bernard Monsarrat, Thomas M. Kristie, Jean-Philippe Girard The THAP-Zinc Finger Protein THAP1 Associates with Coactivator HCF-1 and O-GlcNAc Transferase. A LINK BETWEEN DYT6 AND DYT3 DYSTONIAS. *J Biol Chem. 2010 Apr 30;285(18):13364-71*

2. Emma Lefrançais, Stephane Roga, **Violette Gautier**, Anne Gonzalez-de-Peredo, Bernard Monsarrat, Jean-Philippe Girard, Corinne Cayrol. IL-33 is processed into mature bioactive forms by neutrophil elastase and Cathepsin G . *Proc Natl Acad Sci U S A. 2012 Jan 31;109(5):1673-8*.

3. **Violette Gautier**, Emmanuelle Mouton-Barbosa, David Bouyssie, Nicolas Delcourt, Mathilde Beau, Jean-Philippe Girard, Corinne Cayrol, Odile Burlet-Schiltz, Bernard Monsarrat, and Anne Gonzalez de Peredo. Label-free Quantification and Shotgun Analysis of Complex Proteomes by One-dimensional SDS-AGE/NanoLC-MS - EVALUATION FOR THE LARGE SCALE ANALYSIS OF INFLAMMATORY HUMAN ENDOTHELIAL CELLS. *Mol Cell Proteomics. 2012 Aug;11(8):527-39*.

4. Sophie Mourgues, **Violette Gautier**, Joris Slingerland, Christine Bordier, Amandine Mourcet, Fréderic Coin, Wim Vermeulen, Anne Gonzalez de Peredo, Bernard Monsarrat, Pierre-Olivier Mari, Giuseppina Giglia-Mari. ELL, a novel TFIIH partner is involved in transcription restart after DNA repair. *En préparation*

BIBLIOGRAPHIE

Agard, N. J., J. A. Prescher, et al. (2004). "A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems." J Am Chem Soc **126**(46): 15046-15047.

Ali, S., M. Huber, et al. (2007). "IL-1 receptor accessory protein is essential for IL-33-induced activation of T lymphocytes and mast cells." Proc Natl Acad Sci U S A **104**(47): 18660-18665.

Ali, S., D. Q. Nguyen, et al. "Caspase 3 inactivates biologically active full length interleukin-33 as a classical cytokine but does not prohibit nuclear translocation." Biochem Biophys Res Commun **391**(3): 1512-1516.

Andersen, J. S., C. E. Lyon, et al. (2002). "Directed proteomic analysis of the human nucleolus." Curr Biol **12**(1): 1-11.

Anderson, N. L. and N. G. Anderson (2002). "The human plasma proteome: history, character, and diagnostic prospects." Mol Cell Proteomics **1**(11): 845-867.

Antberg, L., P. Cifani, et al. (2012). "Critical comparison of multidimensional separation methods for increasing protein expression coverage." J Proteome Res **11**(5): 2644-2652.

Aoki, S., M. Hayakawa, et al. (2010). "ST2 gene expression is proliferation-dependent and its ligand, IL-33, induces inflammatory reaction in endothelial cells." Mol Cell Biochem **335**(1-2): 75-81.

Auty, R., H. Steen, et al. (2004). "Purification of active TFIID from Saccharomyces cerevisiae. Extensive promoter contacts and co-activator function." J Biol Chem **279**(48): 49973-49981.

Baekkevold, E. S., M. Roussigne, et al. (2003). "Molecular characterization of NF-HEV, a nuclear factor preferentially expressed in human high endothelial venules." Am J Pathol **163**(1): 69-79.

Bai, Y., K. Markham, et al. (2008). "The in vivo brain interactome of the amyloid precursor protein." Mol Cell Proteomics **7**(1): 15-34.

Bantscheff, M., S. Lemeer, et al. (2012). "Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present." Anal Bioanal Chem **404**(4): 939-965.

Bantscheff, M., M. Schirle, et al. (2007). "Quantitative mass spectrometry in proteomics: a critical review." Anal Bioanal Chem **389**(4): 1017-1031.

Baskin, J. M., J. A. Prescher, et al. (2007). "Copper-free click chemistry for dynamic in vivo imaging." Proc Natl Acad Sci U S A **104**(43): 16793-16797.

Bellew, M., M. Coram, et al. (2006). "A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS." Bioinformatics **22**(15): 1902-1909.

Berglund, L., E. Bjorling, et al. (2008). "A genecentric Human Protein Atlas for expression profiles based on antibodies." Mol Cell Proteomics **7**(10): 2019-2027.

Bessiere, D., C. Lacroix, et al. (2008). "Structure-function analysis of the THAP zinc finger of THAP1, a large C2CH DNA-binding module linked to Rb/E2F pathways." J Biol Chem **283**(7): 4352-4363.

Bildl, W., A. Haupt, et al. (2012). "Extending the dynamic range of label-free mass spectrometric quantification of affinity purifications." Molecular & cellular proteomics : MCP **11**(2): M111 007955.

Bildl, W., A. Haupt, et al. (2012). "Extending the dynamic range of label-free mass spectrometric quantification of affinity purifications." Mol Cell Proteomics **11**(2): M111 007955.

Blagoev, B., I. Kratchmarova, et al. (2003). "A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling." Nat Biotechnol **21**(3): 315-318.

Bodenmiller, B., S. Wanka, et al. (2010). "Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast." Sci Signal **3**(153): rs4.

Boisvert, F. M., Y. W. Lam, et al. (2010). "A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage." Mol Cell Proteomics **9**(3): 457-470.

Boulon, S., Y. Ahmad, et al. (2010). "Establishment of a protein frequency library and its application in the reliable identification of specific protein interaction partners." Mol Cell Proteomics **9**(5): 861-879.

Bousquet-Dubouch, M. P., E. Baudelet, et al. (2009). "Affinity purification strategy to capture human endogenous proteasome complexes diversity and to identify proteasome-interacting proteins." Mol Cell Proteomics **8**(5): 1150-1164.

Bouyssie, D., A. Gonzalez de Peredo, et al. (2007). "Mascot file parsing and quantification (MFPaQ), a new software to parse, validate, and quantify proteomics data generated by ICAT and SILAC mass spectrometric analyses: application to the proteomics study of membrane proteins from primary human endothelial cells." Mol Cell Proteomics **6**(9): 1621-1637.

Boxem, M. and S. van den Heuvel (2002). "C. elegans class B synthetic multivulva genes act in G(1) regulation." Curr Biol **12**(11): 906-911.

Breakefield, X. O., A. J. Blood, et al. (2008). "The pathophysiological basis of dystonias." Nat Rev Neurosci **9**(3): 222-234.

Bressman, S. B., D. Raymond, et al. (2009). "Mutations in THAP1 (DYT6) in early-onset dystonia: a genetic screening study." Lancet Neurol **8**(5): 441-446.

Brun, V., A. Dupuis, et al. (2007). "Isotope-labeled protein standards: toward absolute quantitative proteomics." Mol Cell Proteomics **6**(12): 2139-2149.

Buryskova, M., M. Pospisek, et al. (2004). "Intracellular interleukin-1alpha functionally interacts with histone acetyltransferase complexes." J Biol Chem **279**(6): 4017-4026.

Cambridge, S. B., F. Gnad, et al. (2011). "Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover." J Proteome Res **10**(12): 5275-5284.

Campagne, S., O. Saurel, et al. (2010). "Structural determinants of specific DNA-recognition by the THAP zinc finger." Nucleic Acids Res **38**(10): 3466-3476.

Carriere, V., L. Roussel, et al. (2007). "IL-33, the IL-1-like cytokine ligand for ST2 receptor, is a chromatin-associated nuclear factor in vivo." Proc Natl Acad Sci U S A **104**(1): 282-287.

Cayrol, C. and J. P. Girard (2009). "The IL-1-like cytokine IL-33 is inactivated after maturation by caspase-1." Proc Natl Acad Sci U S A **106**(22): 9021-9026.

Cayrol, C. and J. P. Girard (2009). "The IL-1-like cytokine IL-33 is inactivated after maturation by caspase-1." Proceedings of the National Academy of Sciences of the United States of America **106**(22): 9021-9026.

Cayrol, C., C. Lacroix, et al. (2007). "The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes." Blood **109**(2): 584-594.

Chalfie, M., Y. Tu, et al. (1994). "Green fluorescent protein as a marker for gene expression." Science **263**(5148): 802-805.

Chang, P. V., J. A. Prescher, et al. "Copper-free click chemistry in living animals." Proc Natl Acad Sci U S A **107**(5): 1821-1826.

Chang, Y. J., H. Y. Kim, et al. (2011). "Innate lymphoid cells mediate influenza-induced airway hyper-reactivity independently of adaptive immunity." Nat Immunol **12**(7): 631-638.

Charbonnier, S., O. Gallego, et al. (2008). "The social network of a cell: recent advances in interactome mapping." Biotechnol Annu Rev **14**: 1-28.

Cherry, W. B., J. Yoon, et al. (2008). "A novel IL-1 family cytokine, IL-33, potently activates human eosinophils." J Allergy Clin Immunol **121**(6): 1484-1490.

Chesney, M. A., A. R. Kidd, 3rd, et al. (2006). "gon-14 functions with class B and class C synthetic multivulva genes to control larval growth in Caenorhabditis elegans." Genetics **172**(2): 915-928.

Choi, Y. S., H. J. Choi, et al. (2009). "Interleukin-33 induces angiogenesis and vascular permeability through ST2/TRAF6-mediated endothelial nitric oxide production." Blood **114**(14): 3117-3126.

Choi, Y. S., J. A. Park, et al. (2012). "Nuclear IL-33 is a transcriptional regulator of NF-kappaB p65 and induces endothelial cell activation." Biochem Biophys Res Commun **421**(2): 305-311.

Clouaire, T., M. Roussigne, et al. (2005). "The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity." Proc Natl Acad Sci U S A **102**(19): 6907-6912.

Coin, F., E. Bergmann, et al. (1999). "Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH." The EMBO journal **18**(5): 1357-1366.

Coin, F., J. C. Marinoni, et al. (1998). "Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH." Nature genetics **20**(2): 184-188.

Coin, F., V. Oksenych, et al. (2007). "Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair." Molecular cell **26**(2): 245-256.

Coin, F., V. Oksenych, et al. (2008). "Nucleotide excision repair driven by the dissociation of CAK from TFIIH." Mol Cell **31**(1): 9-20.

Collier, T. S., P. Sarkar, et al. (2010). "Direct comparison of stable isotope labeling by amino acids in cell culture and spectral counting for quantitative proteomics." Analytical chemistry **82**(20): 8696-8702.

Compe, E. and J. M. Egly (2012). "TFIIH: when transcription met DNA repair." Nat Rev Mol Cell Biol **13**(6): 343-354.

Corthals, G. L., V. C. Wasinger, et al. (2000). "The dynamic range of protein expression: a challenge for proteomic research." Electrophoresis **21**(6): 1104-1115.

Cox, J. and M. Mann (2008). "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification." Nat Biotechnol **26**(12): 1367-1372.

Cox, J. and M. Mann (2011). "Quantitative, high-resolution proteomics for data-driven systems biology." Annu Rev Biochem **80**: 273-299.

Craig, R. and R. C. Beavis (2004). "TANDEM: matching proteins with tandem mass spectra." Bioinformatics **20**(9): 1466-1467.

Crawford, N. P., H. Yang, et al. (2009). "The metastasis efficiency modifier ribosomal RNA processing 1 homolog B (RRP1B) is a chromatin-associated factor." J Biol Chem **284**(42): 28660-28673.

Cristea, I. M., R. Williams, et al. (2005). "Fluorescent proteins as proteomic probes." Mol Cell Proteomics **4**(12): 1933-1941.

de Godoy, L. M., J. V. Olsen, et al. (2008). "Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast." Nature **455**(7217): 1251-1254.

de Godoy, L. M., J. V. Olsen, et al. (2006). "Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system." Genome Biol **7**(6): R50.

Deeb, S. J., R. C. D'Souza, et al. (2012). "Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles." Mol Cell Proteomics **11**(5): 77-89.

Dejardin, J. and R. E. Kingston (2009). "Purification of proteins associated with specific genomic Loci." Cell **136**(1): 175-186.

Dejardin, L. M., R. P. Guillou, et al. (2009). "Effect of bending direction on the mechanical behaviour of interlocking nail systems." Vet Comp Orthop Traumatol **22**(4): 264-269.

Dejosez, M., J. S. Krumenacker, et al. (2008). "Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells." Cell **133**(7): 1162-1174.

Delvaeye, M. and E. M. Conway (2009). "Coagulation and innate immune responses: can we view them separately?" Blood **114**(12): 2367-2374.

Demyanets, S., V. Konya, et al. (2011). "Interleukin-33 induces expression of adhesion molecules and inflammatory activation in human endothelial cells and in human atherosclerotic plaques." Arteriosclerosis, thrombosis, and vascular biology **31**(9): 2080-2089.

Dengjel, J., A. R. Kristensen, et al. (2008). "Ordered bulk degradation via autophagy." Autophagy **4**(8).

Dignam, J. D., R. M. Lebovitz, et al. (1983). "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei." Nucleic Acids Res **11**(5): 1475-1489.

Dignam, J. D., P. L. Martin, et al. (1983). "Eukaryotic gene transcription with purified components." Methods Enzymol **101**: 582-598.

Dinarello, C. A. (2009). "Immunological and inflammatory functions of the interleukin-1 family." Annu Rev Immunol **27**: 519-550.

Dinarello, C. A. (2011). "Interleukin-1 in the pathogenesis and treatment of inflammatory diseases." Blood **117**(14): 3720-3732.

Domon, B. and R. Aebersold (2006). "Mass spectrometry and protein analysis." Science **312**(5771): 212-217.

Dube, D. H., J. A. Prescher, et al. (2006). "Probing mucin-type O-linked glycosylation in living animals." Proc Natl Acad Sci U S A **103**(13): 4819-4824.

Dunham, W. H., M. Mullin, et al. (2012). "Affinity-purification coupled to mass spectrometry: Basic principles and strategies." Proteomics **12**(10): 1576-1590.

Dvir, A., J. W. Conaway, et al. (2001). "Mechanism of transcription initiation and promoter escape by RNA polymerase II." Current opinion in genetics & development **11**(2): 209-214.

Eeltink, S., S. Dolman, et al. (2009). "Optimizing the peak capacity per unit time in one-dimensional and off-line two-dimensional liquid chromatography for the separation of complex peptide samples." J Chromatogr A **1216**(44): 7368-7374.

Elia, G. (2008). "Biotinylation reagents for the study of cell surface proteins." Proteomics **8**(19): 4012-4024.

Fan, L., A. S. Arvai, et al. (2006). "Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair." Molecular cell **22**(1): 27-37.

Fang, Y., D. P. Robinson, et al. (2010). "Quantitative analysis of proteome coverage and recovery rates for upstream fractionation methods in proteomics." J Proteome Res **9**(4): 1902-1912.

Fenn, J. B., M. Mann, et al. (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.

Field, J., J. Nikawa, et al. (1988). "Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method." Mol Cell Biol **8**(5): 2159-2165.

Florens, L., M. J. Carozza, et al. (2006). "Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors." Methods **40**(4): 303-311.

Forsman, A., U. Ruetschi, et al. (2008). "Identification of intracellular proteins associated with the EBV-encoded nuclear antigen 5 using an efficient TAP procedure and FT-ICR mass spectrometry." J Proteome Res **7**(6): 2309-2319.

Fournier, M. L., J. M. Gilmore, et al. (2007). "Multidimensional separations-based shotgun proteomics." Chem Rev **107**(8): 3654-3686.

Freiman, R. N. and W. Herr (1997). "Viral mimicry: common mode of association with HCF by VP16 and the cellular protein LZIP." Genes Dev **11**(23): 3122-3127.

Fuchs, T., S. Gavarini, et al. (2009). "Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia." Nat Genet **41**(3): 286-288.

Fujita, M., T. Kiyono, et al. (1997). "In vivo interaction of human MCM heterohexameric complexes with chromatin. Possible involvement of ATP." J Biol Chem **272**(16): 10928-10935.

Gahmberg, C. G. and M. Tolvanen (1996). "Why mammalian cell surface proteins are glycoproteins." Trends Biochem Sci **21**(8): 308-311.

Gautier, V., E. Mouton-Barbosa, et al. (2012). "Label-free quantification and shotgun analysis of complex proteomes by 1D SDS-PAGE/nanoLC-MS: evaluation for the large-scale analysis of inflammatory human endothelial cells." Mol Cell Proteomics.

Gavin, A. C., P. Aloy, et al. (2006). "Proteome survey reveals modularity of the yeast cell machinery." Nature **440**(7084): 631-636.

Geer, L. Y., S. P. Markey, et al. (2004). "Open mass spectrometry search algorithm." J Proteome Res **3**(5): 958-964.

Geiger, T., J. Cox, et al. (2010). "Super-SILAC mix for quantitative proteomics of human tumor tissue." Nat Methods **7**(5): 383-385.

Geiger, T., A. Wehner, et al. (2012). "Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins." Mol Cell Proteomics **11**(3): M111 014050.

Geiger, T., J. R. Wisniewski, et al. (2011). "Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics." Nat Protoc **6**(2): 147-157.

Gerber, S. A., J. Rush, et al. (2003). "Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS." Proc Natl Acad Sci U S A **100**(12): 6940-6945.

Giangrande, P. H., W. Zhu, et al. (2004). "A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription." Genes Dev **18**(23): 2941-2951.

Giglia-Mari, G., F. Coin, et al. (2004). "A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A." Nat Genet **36**(7): 714-719.

Giglia-Mari, G., C. Miquel, et al. (2006). "Dynamic interaction of TTDA with TFIIH is stabilized by nucleotide excision repair in living cells." PLoS Biol **4**(6): e156.

Giglia-Mari, G., A. F. Theil, et al. (2009). "Differentiation driven changes in the dynamic organization of Basal transcription initiation." PLoS Biol **7**(10): e1000220.

Gillet, L. C., P. Navarro, et al. (2012). "Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis." Mol Cell Proteomics **11**(6): O111 016717.

Gloeckner, C. J., K. Boldt, et al. (2007). "A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes." Proteomics **7**(23): 4228-4234.

Glover-Cutter, K., S. Larochelle, et al. (2009). "TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II." Molecular and cellular biology **29**(20): 5455-5464.

Graumann, J., N. C. Hubner, et al. (2008). "Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins." Mol Cell Proteomics **7**(4): 672-683.

Griffin, T. J., H. Xie, et al. (2007). "iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer." J Proteome Res **6**(11): 4200-4209.

Gross, O., C. J. Thomas, et al. (2011). "The inflammasome: an integrated view." Immunological reviews **243**(1): 136-151.

Groth, A., A. Corpet, et al. (2007). "Regulation of replication fork progression through histone supply and demand." Science **318**(5858): 1928-1931.

Guerrier, L., V. Thulasiraman, et al. (2006). "Reducing protein concentration range of biological samples using solid-phase ligand libraries." J Chromatogr B Analyt Technol Biomed Life Sci **833**(1): 33-40.

Gygi, S. P., Y. Rochon, et al. (1999). "Correlation between protein and mRNA abundance in yeast." Molecular and cellular biology **19**(3): 1720-1730.

Han, D. K., J. Eng, et al. (2001). "Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry." Nat Biotechnol **19**(10): 946-951.

Hanawalt, P. C. (2002). "Subpathways of nucleotide excision repair and their regulation." Oncogene **21**(58): 8949-8956.

Hang, H. C., C. Yu, et al. (2003). "A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation." Proc Natl Acad Sci U S A **100**(25): 14846-14851.

Hanke, S., H. Besir, et al. (2008). "Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level." J Proteome Res **7**(3): 1118-1130.

Hansen, K. C., G. Schmitt-Ulms, et al. (2003). "Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography." Mol Cell Proteomics **2**(5): 299-314.

Haraldsen, G., J. Balogh, et al. (2009). "Interleukin-33 - cytokine of dual function or novel alarmin?" Trends Immunol **30**(5): 227-233.

Hayashi, K., S. Y. Kim, et al. (2009). "Evaluation of a collagenase generated osteoarthritis biomarker in naturally occurring canine cruciate disease." Vet Surg **38**(1): 117-121.

Head, J. F. (1992). "A better grip on calmodulin." Curr Biol **2**(11): 609-611.

Hemmerich, S., A. Bistrup, et al. (2001). "Sulfation of L-selectin ligands by an HEV-restricted sulfotransferase regulates lymphocyte homing to lymph nodes." Immunity **15**(2): 237-247.

Ho, L., J. L. Ronan, et al. (2009). "An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency." Proc Natl Acad Sci U S A **106**(13): 5181-5186.

Hochuli, E., W. Bannwarth, et al. (1988). "Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent." Nat Biotech **6**(11): 1321-1325.

Hochuli, V. K. (1988). "Orthopaedic waiting list reduction through a review of service provision: the problems encountered." J R Soc Med **81**(8): 445-447.

Holstege, F. C., P. C. van der Vliet, et al. (1996). "Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH." The EMBO journal **15**(7): 1666-1677.

Hoogstraten, D., A. L. Nigg, et al. (2002). "Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo." Mol Cell **10**(5): 1163-1174.

Hopp, T. P., K. S. Prickett, et al. (1988). "A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification." Nat Biotech **6**(10): 1204-1210.

Horth, P., C. A. Miller, et al. (2006). "Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis." Mol Cell Proteomics **5**(10): 1968-1974.

Hu, P., S. C. Janga, et al. (2009). "Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins." PLoS Biol **7**(4): e96.

Hubner, N. C., A. W. Bird, et al. (2010). "Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions." J Cell Biol **189**(4): 739-754.

Iikura, M., H. Suto, et al. (2007). "IL-33 can promote survival, adhesion and cytokine production in human mast cells." Lab Invest **87**(10): 971-978.

Ito, S., I. Kuraoka, et al. (2007). "XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients." Mol Cell **26**(2): 231-243.

Iwasaki, M., N. Sugiyama, et al. (2012). "Human proteome analysis by using reversed phase monolithic silica capillary columns with enhanced sensitivity." J Chromatogr A **1228**: 292-297.

Jaitly, N., A. Mayampurath, et al. (2009). "Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data." BMC Bioinformatics **10**: 87.

Johnson, R. S., S. A. Martin, et al. (1987). "Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine." Anal Chem **59**(21): 2621-2625.

Jones, J., K. Wu, et al. (2008). "A targeted proteomic analysis of the ubiquitin-like modifier nedd8 and associated proteins." J Proteome Res **7**(3): 1274-1287.

Jovanovic, M., L. Reiter, et al. (2010). "A quantitative targeted proteomics approach to validate predicted microRNA targets in C. elegans." Nat Methods **7**(10): 837-842.

Kainov, D. E., M. Vitorino, et al. (2008). "Structural basis for group A trichothiodystrophy." Nature structural & molecular biology **15**(9): 980-984.

Kar, G., A. Gursoy, et al. (2009). "Human cancer protein-protein interaction network: a structural perspective." PLoS Comput Biol **5**(12): e1000601.

Kathiresan, S., C. J. Willer, et al. (2009). "Common variants at 30 loci contribute to polygenic dyslipidemia." Nat Genet **41**(1): 56-65.

Kawayama, T., M. Okamoto, et al. (2012). "Interleukin-18 in pulmonary inflammatory diseases." J Interferon Cytokine Res **32**(10): 443-449.

Kim, K., S. J. Kim, et al. (2010). "Verification of biomarkers for diabetic retinopathy by multiple reaction monitoring." J Proteome Res **9**(2): 689-699.

Klockenbusch, C., J. E. O'Hara, et al. (2012). "Advancing formaldehyde cross-linking towards quantitative proteomic applications." Anal Bioanal Chem.

Klune, J. R., R. Dhupar, et al. (2008). "HMGB1: endogenous danger signaling." Mol Med **14**(7-8): 476-484.

Knobbe, C. B., T. J. Revett, et al. (2011). "Choice of biological source material supersedes oxidative stress in its influence on DJ-1 in vivo interactions with Hsp90." J Proteome Res **10**(10): 4388-4404.

Knuesel, M., Y. Wan, et al. (2003). "Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system." Mol Cell Proteomics **2**(11): 1225-1233.

Kocher, T., R. Swart, et al. (2011). "Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides." Anal Chem **83**(7): 2699-2704.

Kruger, M., M. Moser, et al. (2008). "SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function." Cell **134**(2): 353-364.

Kumar, S., M. N. Tzimas, et al. (1997). "Expression of ST2, an interleukin-1 receptor homologue, is induced by proinflammatory stimuli." Biochem Biophys Res Commun **235**(3): 474-478.

Lam, Y. W., A. I. Lamond, et al. (2007). "Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins." Curr Biol **17**(9): 749-760.

Laughlin, S. T., J. M. Baskin, et al. (2008). "In vivo imaging of membrane-associated glycans in developing zebrafish." Science **320**(5876): 664-667.

Laughlin, S. T. and C. R. Bertozzi (2009). "Imaging the glycome." Proc Natl Acad Sci U S A **106**(1): 12-17.

Lee, Y. H., H. T. Tan, et al. (2010). "Subcellular fractionation methods and strategies for proteomics." Proteomics **10**(22): 3935-3956.

Li, B., P. Yau, et al. (2011). "Identification of cytochrome P450 2C2 protein complexes in mouse liver." Proteomics **11**(16): 3359-3368.

Li, Y. (2010). "Commonly used tag combinations for tandem affinity purification." Biotechnol Appl Biochem **55**(2): 73-83.

Liang, S., Y. Yu, et al. (2009). "Analysis of the protein complex associated with 14-3-3 epsilon by a deuterated-leucine labeling quantitative proteomics strategy." J Chromatogr B Analyt Technol Biomed Life Sci **877**(7): 627-634.

Lin, C., E. R. Smith, et al. (2010). "AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia." Mol Cell **37**(3): 429-437.

Liu, H., J. W. Finch, et al. (2007). "Effects of column length, particle size, gradient length and flow rate on peak capacity of nano-scale liquid chromatography for peptide separations." J Chromatogr A **1147**(1): 30-36.

Liu, H., R. G. Sadygov, et al. (2004). "A model for random sampling and estimation of relative protein abundance in shotgun proteomics." Anal Chem **76**(14): 4193-4201.

Lohmann, K., N. Uflacker, et al. (2012). "Identification and functional analysis of novel THAP1 mutations." Eur J Hum Genet **20**(2): 171-175.

Lopez-Castejon, G. and D. Brough (2011). "Understanding the mechanism of IL-1beta secretion." Cytokine & growth factor reviews **22**(4): 189-195.

Lottspeich, F. (1999). "Proteome Analysis: A Pathway to the Functional Analysis of Proteins." Angewandte Chemie **38**(17): 2476-2492.

Luber, C. A., J. Cox, et al. (2010). "Quantitative proteomics reveals subset-specific viral recognition in dendritic cells." Immunity **32**(2): 279-289.

Luthi, A. U., S. P. Cullen, et al. (2009). "Suppression of interleukin-33 bioactivity through proteolysis by apoptotic caspases." Immunity **31**(1): 84-98.

Macfarlan, T., S. Kutney, et al. (2005). "Human THAP7 is a chromatin-associated, histone tail-binding protein that represses transcription via recruitment of HDAC3 and nuclear hormone receptor corepressor." J Biol Chem **280**(8): 7346-7358.

Macfarlan, T., J. B. Parker, et al. (2006). "Thanatos-associated protein 7 associates with template activating factor-Ibeta and inhibits histone acetylation to repress transcription." Mol Endocrinol **20**(2): 335-347.

Makino, S., R. Kaji, et al. (2007). "Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism." Am J Hum Genet **80**(3): 393-406.

Malovannaya, A., R. B. Lanz, et al. (2011). "Analysis of the human endogenous coregulator complexome." Cell **145**(5): 787-799.

Malovannaya, A., Y. Li, et al. (2010). "Streamlined analysis schema for high-throughput identification of endogenous protein complexes." Proc Natl Acad Sci U S A **107**(6): 2431-2436.

Mann, M. and M. Wilm (1994). "Error-tolerant identification of peptides in sequence databases by peptide sequence tags." Anal Chem **66**(24): 4390-4399.

Mazars, R., A. Gonzalez-de-Peredo, et al. (2010). "The THAP-zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase: a link between DYT6 and DYT3 dystonias." J Biol Chem **285**(18): 13364-13371.

McDonald, C. A., J. Y. Yang, et al. (2009). "Combining results from lectin affinity chromatography and glycocapture approaches substantially improves the coverage of the glycoproteome." Mol Cell Proteomics **8**(2): 287-301.

Michalski, A., J. Cox, et al. (2011). "More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS." J Proteome Res **10**(4): 1785-1793.

Miller, A. M. (2011). "Role of IL-33 in inflammation and disease." J Inflamm (Lond) **8**(1): 22.

Miller, A. M., D. Xu, et al. (2008). "IL-33 reduces the development of atherosclerosis." J Exp Med **205**(2): 339-346.

Moffatt, M. F., I. G. Gut, et al. (2010). "A large-scale, consortium-based genomewide association study of asthma." N Engl J Med **363**(13): 1211-1221.

Mosley, A. L., M. E. Sardiu, et al. (2011). "Highly reproducible label free quantitative proteomic analysis of RNA polymerase complexes." Mol Cell Proteomics **10**(2): M110 000687.

Moussion, C., N. Ortega, et al. (2008). "The IL-1-like cytokine IL-33 is constitutively expressed in the nucleus of endothelial cells and epithelial cells in vivo: a novel 'alarmin'?" PLoS One **3**(10): e3331.

Mousson, F., A. Kolkman, et al. (2008). "Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes." Mol Cell Proteomics **7**(5): 845-852.

Mouton-Barbosa, E., F. Roux-Dalvai, et al. (2010). "In-depth exploration of cerebrospinal fluid by combining peptide ligand library treatment and label-free protein quantification." Mol Cell Proteomics **9**(5): 1006-1021.

Mueller, L. N., M. Y. Brusniak, et al. (2008). "An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data." J Proteome Res **7**(1): 51-61.

Muller, U. (2009). "The monogenic primary dystonias." Brain **132**(Pt 8): 2005-2025.

Muller, V. S., P. R. Jungblut, et al. (2011). "Membrane-SPINE: an improved method to identify protein-protein interaction partners of membrane proteins in vivo." Proteomics **11**(10): 2124-2128.

Nagaraj, N., N. A. Kulak, et al. (2012). "System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap." Mol Cell Proteomics **11**(3): M111 013722.

Nakatani, Y. and V. Ogryzko (2003). "Immunoaffinity purification of mammalian protein complexes." Methods Enzymol **370**: 430-444.

Narumi, R., T. Murakami, et al. (2012). "A strategy for large-scale phosphoproteomics and SRM-based validation of human breast cancer tissue samples." J Proteome Res.

Nesvizhskii, A. I. (2010). "A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics." J Proteomics **73**(11): 2092-2123.

Nesvizhskii, A. I. and R. Aebersold (2004). "Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS." Drug Discov Today **9**(4): 173-181.

Nesvizhskii, A. I. and R. Aebersold (2005). "Interpretation of shotgun proteomic data: the protein inference problem." Mol Cell Proteomics **4**(10): 1419-1440.

Nielsen, P. A., J. V. Olsen, et al. (2005). "Proteomic mapping of brain plasma membrane proteins." Mol Cell Proteomics **4**(4): 402-408.

Nittis, T., L. Guittat, et al. (2010). "Revealing novel telomere proteins using in vivo cross-linking, tandem affinity purification, and label-free quantitative LC-FTICR-MS." Mol Cell Proteomics **9**(6): 1144-1156.

Oda, Y., K. Huang, et al. (1999). "Accurate quantitation of protein expression and site-specific phosphorylation." Proc Natl Acad Sci U S A **96**(12): 6591-6596.

Oeljeklaus, S., H. E. Meyer, et al. (2009). "New dimensions in the study of protein complexes using quantitative mass spectrometry." FEBS Lett **583**(11): 1674-1683.

Oksenych, V., B. Bernardes de Jesus, et al. (2009). "Molecular insights into the recruitment of TFIIH to sites of DNA damage." The EMBO journal **28**(19): 2971-2980.

Olma, M. H., M. Roy, et al. (2009). "An interaction network of the mammalian COP9 signalosome identifies Dda1 as a core subunit of multiple Cul4-based E3 ligases." J Cell Sci **122**(Pt 7): 1035-1044.

Olsen, J. V., P. A. Nielsen, et al. (2007). "Quantitative proteomic profiling of membrane proteins from the mouse brain cortex, hippocampus, and cerebellum using the HysTag reagent: mapping of neurotransmitter receptors and ion channels." Brain Res **1134**(1): 95-106.

Olsen, J. V., S. E. Ong, et al. (2004). "Trypsin cleaves exclusively C-terminal to arginine and lysine residues." Mol Cell Proteomics **3**(6): 608-614.

Olsen, J. V., J. C. Schwartz, et al. (2009). "A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed." Molecular & cellular proteomics : MCP **8**(12): 2759-2769.

Ong, S. E., B. Blagoev, et al. (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics." Mol Cell Proteomics **1**(5): 376-386.

Ong, S. E. and M. Mann (2005). "Mass spectrometry-based proteomics turns quantitative." Nat Chem Biol **1**(5): 252-262.

Ong, S. E. and M. Mann (2006). "A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC)." Nat Protoc **1**(6): 2650-2660.

Panchaud, A., A. Scherl, et al. (2009). "Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean." Anal Chem **81**(15): 6481-6488.

Patterson, S. D. and R. H. Aebersold (2003). "Proteomics: the first decade and beyond." Nat Genet **33 Suppl**: 311-323.

Peng, J., J. E. Elias, et al. (2003). "Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome." J Proteome Res **2**(1): 43-50.

Perkins, D. N., D. J. Pappin, et al. (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-3567.

Picotti, P. and R. Aebersold (2012). "Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions." Nat Methods **9**(6): 555-566.

Plumb, R. S., K. A. Johnson, et al. (2006). "UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation." Rapid Commun Mass Spectrom **20**(13): 1989-1994.

Poser, I., M. Sarov, et al. (2008). "BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals." Nat Methods **5**(5): 409-415.

Pratt, J. M., D. M. Simpson, et al. (2006). "Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes." Nat Protoc **1**(2): 1029-1043.

Prescher, J. A., D. H. Dube, et al. (2004). "Chemical remodelling of cell surfaces in living animals." Nature **430**(7002): 873-877.

Price, A. E., H. E. Liang, et al. (2010). "Systemically dispersed innate IL-13-expressing cells in type 2 immunity." Proc Natl Acad Sci U S A **107**(25): 11489-11494.

Ranish, J. A., E. C. Yi, et al. (2003). "The study of macromolecular complexes by quantitative proteomics." Nat Genet **33**(3): 349-355.

Ray, S., P. J. Reddy, et al. (2011). "Proteomic technologies for the identification of disease biomarkers in serum: advances and challenges ahead." Proteomics **11**(11): 2139-2161.

Reilly, P. T., J. Wysocka, et al. (2002). "Inactivation of the retinoblastoma protein family can bypass the HCF-1 defect in tsBN67 cell proliferation and cytokinesis." Mol Cell Biol **22**(19): 6767-6778.

Remboutsika, E., Y. Lutz, et al. (1999). "The putative nuclear receptor mediator TIF1alpha is tightly associated with euchromatin." J Cell Sci **112 (Pt 11)**: 1671-1683.

Reynolds, K. J., X. Yao, et al. (2002). "Proteolytic 18O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent." J Proteome Res **1**(1): 27-33.

Rigaut, G., A. Shevchenko, et al. (1999). "A generic protein purification method for protein complex characterization and proteome exploration." Nat Biotechnol **17**(10): 1030-1032.

Rinner, O., L. N. Mueller, et al. (2007). "An integrated mass spectrometric and computational framework for the analysis of protein interaction networks." Nat Biotechnol **25**(3): 345-352.

Roche, S., L. Tiers, et al. (2009). "Depletion of one, six, twelve or twenty major blood proteins before proteomic analysis: the more the better?" J Proteomics **72**(6): 945-951.

Roepstorff, P. and J. Fohlman (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed Mass Spectrom **11**(11): 601.

Rosen, S. D. (1999). "Endothelial ligands for L-selectin: from lymphocyte recirculation to allograft rejection." Am J Pathol **155**(4): 1013-1020.

Rossignol, M., I. Kolb-Cheynel, et al. (1997). "Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH." The EMBO journal **16**(7): 1628-1637.

Rostovtsev, V. V., L. G. Green, et al. (2002). "A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes." Angew Chem Int Ed Engl **41**(14): 2596-2599.

Roussel, L., M. Erard, et al. (2008). "Molecular mimicry between IL-33 and KSHV for attachment to chromatin through the H2A-H2B acidic pocket." EMBO Rep **9**(10): 1006-1012.

Roussigne, M., S. Kossida, et al. (2003). "The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase." Trends Biochem Sci **28**(2): 66-69.

Roux-Dalvai, F., A. Gonzalez de Peredo, et al. (2008). "Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry." Mol Cell Proteomics **7**(11): 2254-2269.

Saade, E., U. Mechold, et al. (2009). "Analysis of interaction partners of H4 histone by a new proteomics approach." Proteomics **9**(21): 4934-4943.

Sanada, S., D. Hakuno, et al. (2007). "IL-33 and ST2 comprise a critical biomechanically induced and cardioprotective signaling system." J Clin Invest **117**(6): 1538-1549.

Saxon, E. and C. R. Bertozzi (2000). "Cell surface engineering by a modified Staudinger reaction." Science **287**(5460): 2007-2010.

Scaffidi, P., T. Misteli, et al. (2002). "Release of chromatin protein HMGB1 by necrotic cells triggers inflammation." Nature **418**(6894): 191-195.

Scharer, O. D. (2008). "Hot topics in DNA repair: the molecular basis for different disease states caused by mutations in TFIIH and XPG." DNA Repair (Amst) **7**(2): 339-344.

Schiess, R., L. N. Mueller, et al. (2008). "Analysis of cell surface proteome changes via label-free, quantitative mass spectrometry." Mol Cell Proteomics.

Schilling, B., M. J. Rardin, et al. (2012). "Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation." Molecular & cellular proteomics : MCP **11**(5): 202-214.

Schmidt, T. G. and A. Skerra (1993). "The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment." Protein Eng **6**(1): 109-122.

Schmitz, J., A. Owyang, et al. (2005). "IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines." Immunity **23**(5): 479-490.

Scholtissen, S., F. Guillemin, et al. (2009). "Assessment of determinants for osteoporosis in elderly men." Osteoporos Int **20**(7): 1157-1166.

Schulze, W. X. and M. Mann (2004). "A novel proteomic screen for peptide-protein interactions." J Biol Chem **279**(11): 10756-10764.

Schwanhausser, B., M. Gossen, et al. (2009). "Global analysis of cellular protein translation by pulsed SILAC." Proteomics **9**(1): 205-209.

Selbach, M. and M. Mann (2006). "Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK)." Nat Methods **3**(12): 981-983.

Sennels, L., M. Salek, et al. (2007). "Proteomic analysis of human blood serum using peptide library beads." J Proteome Res **6**(10): 4055-4062.

Shilatifard, A., W. S. Lane, et al. (1996). "An RNA polymerase II elongation factor encoded by the human ELL gene." Science **271**(5257): 1873-1876.

Shin, B. K., H. Wang, et al. (2003). "Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function." J Biol Chem **278**(9): 7607-7616.

Silva, J. C., M. V. Gorenstein, et al. (2006). "Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition." Molecular & cellular proteomics : MCP **5**(1): 144-156.

Simone, F., P. E. Polak, et al. (2001). "EAF1, a novel ELL-associated factor that is delocalized by expression of the MLL-ELL fusion protein." Blood **98**(1): 201-209.

Sletten, E. M. and C. R. Bertozzi (2008). "A hydrophilic azacyclooctyne for Cu-free click chemistry." Org Lett **10**(14): 3097-3099.

Smith, D. B. and K. S. Johnson (1988). "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase." Gene **67**(1): 31-40.

Smith, E. R., C. Lin, et al. (2011). "The little elongation complex regulates small nuclear RNA transcription." Mol Cell **44**(6): 954-965.

Smith, R. D., G. A. Anderson, et al. (2002). "The use of accurate mass tags for high-throughput microbial proteomics." OMICS **6**(1): 61-90.

Song, W., Y. Chen, et al. (2011). "Novel THAP1 gene mutations in patients with primary dystonia from southwest China." J Neurol Sci **309**(1-2): 63-67.

Sowa, M. E., E. J. Bennett, et al. (2009). "Defining the human deubiquitinating enzyme interaction landscape." Cell **138**(2): 389-403.

Steen, H. and M. Mann (2004). "The ABC's (and XYZ's) of peptide sequencing." Nat Rev Mol Cell Biol **5**(9): 699-711.

Stephenson, J. J., C. Gregory, et al. (2009). "An open-label clinical trial evaluating safety and pharmacokinetics of two dosing schedules of panitumumab in patients with solid tumors." Clin Colorectal Cancer **8**(1): 29-37.

Stern, S. and W. Herr (1991). "The herpes simplex virus trans-activator VP16 recognizes the Oct-1 homeo domain: evidence for a homeo domain recognition subdomain." Genes Dev **5**(12B): 2555-2566.

Sturm, M., A. Bertsch, et al. (2008). "OpenMS - an open-source software framework for mass spectrometry." BMC Bioinformatics **9**: 163.

Suter, B., S. Kittanakom, et al. (2008). "Interactive proteomics: what lies ahead?" Biotechniques **44**(5): 681-691.

Sutherland, B. W., J. Toews, et al. (2008). "Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions." J Mass Spectrom **43**(6): 699-715.

Tai, H. C., H. Besche, et al. (2010). "Characterization of the Brain 26S Proteasome and its Interacting Proteins." Front Mol Neurosci **3**.

Takahashi, H., T. J. Parmely, et al. (2011). "Human mediator subunit MED26 functions as a docking site for transcription elongation factors." Cell **146**(1): 92-104.

Talabot-Ayer, D., C. Lamacchia, et al. (2009). "Interleukin-33 is biologically active independently of caspase-1 cleavage." J Biol Chem **284**(29): 19420-19426.

Tamiya, G. (2009). "Transcriptional dysregulation: a cause of dystonia?" Lancet Neurol **8**(5): 416-418.

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems." Appl Microbiol Biotechnol **60**(5): 523-533.

Thakur, S. S., T. Geiger, et al. (2011). "Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation." Mol Cell Proteomics **10**(8): M110 003699.

Thomas, J. H. and H. R. Horvitz (1999). "The C. elegans gene lin-36 acts cell autonomously in the lin-35 Rb pathway." Development **126**(15): 3449-3459.

Thomas, M. C. and C. M. Chiang (2006). "The general transcription machinery and general cofactors." Crit Rev Biochem Mol Biol **41**(3): 105-178.

Thompson, A., J. Schafer, et al. (2003). "Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS." Anal Chem **75**(8): 1895-1904.

Thulasiraman, V., S. Lin, et al. (2005). "Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands." Electrophoresis **26**(18): 3561-3571.

Trang, T. and M. W. Salter (2012). "P2X4 purinoceptor signaling in chronic pain." Purinergic Signal **8**(3): 621-628.

Trinkle-Mulcahy, L., S. Boulon, et al. (2008). "Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes." J Cell Biol **183**(2): 223-239.

Tsai, Y. C., T. M. Greco, et al. (2012). "Functional proteomics establishes the interaction of SIRT7 with chromatin remodeling complexes and expands its role in regulation of RNA polymerase I transcription." Mol Cell Proteomics **11**(5): 60-76.

Tsou, C. C., C. F. Tsai, et al. (2010). "IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation." Mol Cell Proteomics **9**(1): 131-144.

Tu, C., P. A. Rudnick, et al. (2010). "Depletion of abundant plasma proteins and limitations of plasma proteomics." J Proteome Res **9**(10): 4982-4991.

Tyagi, S., A. L. Chabes, et al. (2007). "E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases." Mol Cell **27**(1): 107-119.

Uchimura, K. and S. D. Rosen (2006). "Sulfated L-selectin ligands as a therapeutic target in chronic inflammation." Trends Immunol **27**(12): 559-565.

Ueda, T., E. Compe, et al. (2009). "Both XPD alleles contribute to the phenotype of compound heterozygote xeroderma pigmentosum patients." J Exp Med **206**(13): 3031-3046.

Ulmann, L., H. Hirbec, et al. (2010). "P2X4 receptors mediate PGE2 release by tissue-resident macrophages and initiate inflammatory pain." EMBO J **29**(14): 2290-2300.

Venable, J. D., M. Q. Dong, et al. (2004). "Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra." Nat Methods **1**(1): 39-45.

Vinther, J., M. M. Hedegaard, et al. (2006). "Identification of miRNA targets with stable isotope labeling by amino acids in cell culture." Nucleic Acids Res **34**(16): e107.

Volin, M. V. and A. E. Koch (2011). "Interleukin-18: a mediator of inflammation and angiogenesis in rheumatoid arthritis." J Interferon Cytokine Res **31**(10): 745-751.

von Pfeil, D. J., C. E. Decamp, et al. (2009). "Does Osgood-Schlatter disease exist in the dog?" Vet Comp Orthop Traumatol **22**(4): 257-263.

Wang, X. and L. Huang (2008). "Identifying dynamic interactors of protein complexes by quantitative mass spectrometry." Mol Cell Proteomics **7**(1): 46-57.

Washburn, M. P., D. Wolters, et al. (2001). "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." Nat Biotechnol **19**(3): 242-247.

Wells, L., K. Vosseller, et al. (2001). "Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc." Science **291**(5512): 2376-2378.

Werman, A., R. Werman-Venkert, et al. (2004). "The precursor form of IL-1alpha is an intracrine proinflammatory activator of transcription." Proc Natl Acad Sci U S A **101**(8): 2434-2439.

Wisniewski, J. R., A. Zougman, et al. (2009). "Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome." J Proteome Res **8**(12): 5674-5678.

Wisniewski, J. R., A. Zougman, et al. (2009). "Universal sample preparation method for proteome analysis." Nat Methods **6**(5): 359-362.

Wollscheid, B., D. Bausch-Fluck, et al. (2009). "Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins." Nature biotechnology **27**(4): 378-386.

Wolters, D. A., M. P. Washburn, et al. (2001). "An automated multidimensional protein identification technology for shotgun proteomics." Anal Chem **73**(23): 5683-5690.

Wu, C. C., M. J. MacCoss, et al. (2003). "A method for the comprehensive proteomic analysis of membrane proteins." Nat Biotechnol **21**(5): 532-538.

Wysocka, J., M. P. Myers, et al. (2003). "Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1." Genes Dev **17**(7): 896-911.

Wysocka, J., P. T. Reilly, et al. (2001). "Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells." Mol Cell Biol **21**(11): 3820-3829.

Xiao, J., Y. Zhao, et al. (2010). "Novel THAP1 sequence variants in primary dystonia." Neurology **74**(3): 229-238.

Xu, D., H. R. Jiang, et al. (2008). "IL-33 exacerbates antigen-induced arthritis by activating mast cells." Proc Natl Acad Sci U S A **105**(31): 10913-10918.

Xu, X., Y. Song, et al. (2010). "The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification." Protein Expr Purif **72**(2): 149-156.

Yankulov, K. Y. and D. L. Bentley (1997). "Regulation of CDK7 substrate specificity by MAT1 and TFIIH." The EMBO journal **16**(7): 1638-1646.

Yao, X., A. Freas, et al. (2001). "Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus." Anal Chem **73**(13): 2836-2842.

Ye, J. Z., J. R. Donigian, et al. (2004). "TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres." J Biol Chem **279**(45): 47264-47271.

Yudkovsky, N., J. A. Ranish, et al. (2000). "A transcription reinitiation intermediate that is stabilized by activator." Nature **408**(6809): 225-229.

Zhao, Y. and O. N. Jensen (2009). "Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques." Proteomics **9**(20): 4632-4641.

Zhovmer, A., V. Oksenych, et al. (2010). "Two sides of the same coin: TFIIH complexes in transcription and DNA repair." TheScientificWorldJournal **10**: 633-643.

ANNEXES

- Annexe 1 : Supplementary data de l'article Mazars et al., JBC, 2010 : The THAP-Zinc Finger Protein THAP1 Associates with co-activator HCF-1 and O-GlcNAc Transferase

- Annexe 2 Supplementary data de l'article Mourgues, Gautier et al., "ELL, a novel TFIIH partner is involved in transcription restart after DNA repair" en preparation

- Annexe 3 : Table des N-glycopeptides identifiés dans les trois réplicats des expériences d'enrichissement des N-glycoprotéines de surface avec la phosphine biotine après digestion PNGase.

- Annexe 4 : Table des protéines variantes identifiées après digestion trypsique dans les deux réplicats biologiques des expériences d'enrichissement des glycoprotéines avec la phosphine biotine après stimulation TNF/IFN des HUVEC. (AC : numéro d'accession, MW : molecular weight, Score = score Mascot)

- Annexe 5 : Table des protéines variantes identifiées après analyse quantitative du protéome total des HUVEC suite à la stimulation TNFα-IFNγ Surlignées en bleu : protéines identifiées variantes à la fois après stimulation TNFα-IFNγ, IL-1β et IL-33 ; surlignées en jaune : protéines variantes après stimulation TNFα-IFNγ et IL-33. (AC : numéro d'accession, ID : identifiant de la protéine, MW : molecular weight)

- Annexe 6 : Table des protéines variantes identifiées après analyse quantitative du protéome total des HUVEC suite à la stimulation IL-1β des HUVEC. Surlignées en bleu : protéines identifiées variantes à la fois après stimulation TNFα-IFNγ, IL-1β et IL-33 ; surlignées en vert : protéines variantes après stimulation IL-1β et IL-33. (AC : numéro d'accession, ID : identifiant de la protéine, MW : molecular weight)

- Annexe 7 : Table des protéines variantes identifiées après analyse quantitative du protéome total des HUVEC suite à la stimulation IL-33 des HUVEC. Surlignées en bleu : protéines identifiées variantes à la fois après stimulation TNFα-IFNγ, IL-1β et IL-33 ; surlignées en vert : protéines variantes après stimulation IL-1β et IL-33 ; surlignées en jaune : protéines variantes après stimulation TNFα-IFNγ et IL-33. (AC : numéro d'accession, ID : identifiant de la protéine, MW : molecular weight).

Annexe 1 : Supplementary data de l'article *Mazars et al., JBC, 2010 : The THAP-Zinc Finger P THAP1 Associates with co-activator HCF-1 and O-GlcNAc Transferase*

Supplemental Methods

Mass spectrometry analysis
In order to identify the components of the purified complexes, bands were cut on the SDS-PAGE gel all along the migration lane of the immunoprecipitated complex, as well as on the control lane. Proteins of the different gel slices were treated with 10mM DTT and 55 mM iodoacetamide for cysteine alkylation, and digested with trypsin (Promega). The resulting peptides were extracted from the gel by successive incubations in 10% formic acid/acetonitrile (1/1) and dried in a speed-vac. Peptides were reconstituted in 5% acetonitrile, 0.05% trifluoroacetic acid and analyzed by nanoLC-MS/MS using an Ultimate 3000 system (Dionex) coupled to a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). Elution from the analytical column (75-µm inner diameter x 15-cm PepMap C18, Dionex) was performed using a 60min acetonitrile gradient at a flow rate of 300 nL/min. The LTQ-Orbitrap was operated in information-dependant acquisition mode with the XCalibur software. Survey scan MS were acquired in the Orbitrap on the 300-2000 m/z range with the resolution set to a value of 60000. The five most intense ions per survey scan were selected for CID fragmentation and the resulting fragments were analyzed in the linear trap (LTQ). Dynamic exclusion was employed within 60 seconds to prevent repetitive selection of the same peptide. The Mascot Daemon software (version 2.2.0, Matrix Science) was used to perform database searches against human entries in Uniprot. Peaklists were created with ExtractMSN (provided with Xcalibur version 2. 0 SR2, Thermo Fisher Scientific) with the following parameters: parent ions in the mass range 400-4500, no grouping of MS/MS scans, threshold at 1000. For database search, carbamidomethylation of cysteines was set as a fixed modification, oxidation of methionines, O-GlcNAc glycosylation and phosphorylation of serines and threonines were set as variable modifications, and the mass tolerances in MS and MS/MS were set to 10 ppm and 0.6 Da, respectively. Mascot results were parsed with the in-house developed software MFPaQ version 4.0 and protein hits were automatically validated if were assigned at least one top ranking peptide with a p-value < 0.05. The software was also used to extract MS signal of identified proteins from raw files, and to calculate mean ratio of peptides intensity signal between the lane of immunopurified complex and the control lane. Proteins were considered as potential specific partners if they were exclusively identified by MS/MS sequencing in the lane of the complex with a significant score, or if the ratio of their signal intensity between the complex and control lanes was above 10.

Primer sequences
1) Primers used in ChIP-QPCR experiments :
- -178 □ Forward : 5□-CCCACACAAAACATGGTAGCA-3□
  Reverse : 5□-ATGGGCCCCATTGGATATGGACATG -3□
- +84 □ Forward : 5□-AAAGTGCTGTCTGGCTCCAACT-3□
  Reverse : 5□-GACAGAGTGGGAAGGGTTAGGTT-3□
- +2530 □ Forward : 5□-GAGGAAGAAGAGAAGGAGAGGAACA-3□
  Reverse : 5□-ATCCACACATCAGACATTCATCTCTAT-3□
- p107 □ Forward : 5□-CCGGAGGAAAAACGGACTTT-3□
  Reverse : 5□-CTGCGGGACGTGTTGTCAT -3□
- Gal4UAS □ Forward : 5□-TCATCAATGTATCTTATCATGTCTGGAT-3□
  Reverse : 5□-CGGAATGCCAAGCTGGAA -3□

2) Primers used in Q-PCR experiments :
- actin □ Forward : 5□-TCCCTGGAGAAGAGCTACGA-3□
  Reverse : 5□-AGGAAGGAAGGCTGGAAGAG -3□
- GAPDH □ Forward : 5□-GAGTCAACGGATTTGGTCGT-3□

  Reverse : 5• -GACAAGCTTCCCGTTCTCAG -3•

Thap1, RRM1, HCF-1, OGT : Quantitect primers assays have been purchased from Qiagen (sequences are not provided by Qiagen)

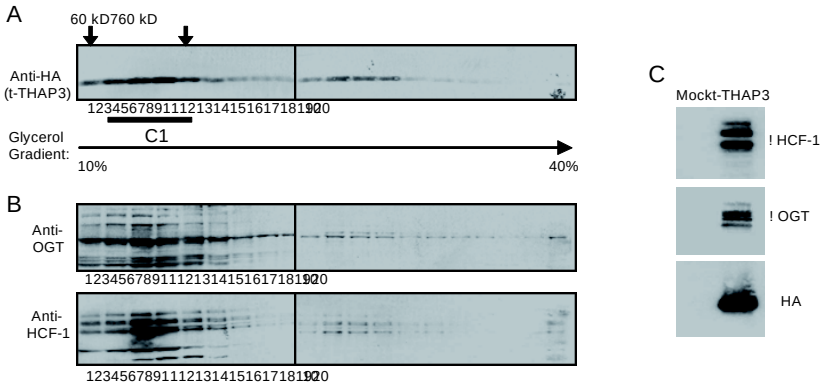


Fig. S1. THAP3, HCF-1 and OGT co-sediment on glycerol gradients. (A) Nuclear extracts from Hela cells expressing the human THAP3-Flag/HA protein (t-THAP3) were fractionated on a 10 to 40% glycerol gradient. Fractions were analyzed by SDS-PAGE and western blotting with anti-HA mAb. The migration of molecular weights markers bovine serum albumin (60 kDa) and thyroglobulin (670 kDa) is indicated. Fractions 3-7 corresponding to complex 1 (C1) were pooled for subsequent immunoprecipitation with anti-Flag antibody. (B) HCF-1 and OGT co-sediment with t-THAP3 in low (about 0,6 MDa) and high (about 2MDa) molecular-weight complexes. Nuclear extracts from t-THAP3 Hela cells were fractionated on a 10 to 40% glycerol gradient and collected fractions were immunobloted with anti-OGT and anti-HCF-1 antibodies. (C) t-THAP3 co-immunoprecipitated proteins from uninduced (mock) and induced cells were separated by SDS-PAGE and immunoblotted with the indicated antibodies.

Position	m/z	z	Theo. Mass	Exp. Mass	delta (ppm)	Score	Rank	Sequence	PTM	Peak Area	elution time
400-426	959.8239	3	2876.4444	2876.4499	1.91	30	1	YDIPATAATATSPTPNPVPSVPANPPK	HexNAc	64176	25.2
	892.1309	3	2673.3650	2673.3709	2.22	N.S.	N.S.	YDIPATAATATSPTPNPVPSVPANPPK		978981	26.8
	1337.6920	2	2673.3650	2673.3694	1.67	63	1	YDIPATAATATSPTPNPVPSVPANPPK			
489-511	830.0963	3	2487.2639	2487.2671	1.26	29	1	VTGPQATTGTPLVTMoxRPASQAGK	HexNAc	142617	19.8
	762.4044	3	2284.1846	2284.1866	0.88	18	1	VTGPQATTGTPLVTMoxRPASQAGK		375210	20.3
512-524	735.9127	2	1469.8090	1469.8108	1.27	9	1	APVTVTSLPAGVR	HexNAc	302271	21.5 / 21.8
	634.3732	2	1266.7296	1266.7302	0.51	51	1	APVTVTSLPAGVR		1121128	23.1
579-594	1006.0110	2	2010.0079	2010.0074	-0.22	9	1	TMoxAVTPGTTTLPATVK	2 HexNAc	191864	21.7 / 22.7
	904.4720	2	1806.9285	1806.9294	0.52	46.01	1	TMoxAVTPGTTTLPATVK	HexNAc	641204	23.3
	802.9332	2	1603.8491	N.D.	N.D.	N.S.	N.S.	TMoxAVTPGTTTLPATVK		N.D.	N.D.
612-637	1002.1888	3	3003.5361	3003.5452	3.03	9	4	TAAAQVGTSVSSATNTSTRPIITVHK	2 HexNAc	138768	18.1 / 19.2
	934.4946	3	2800.4567	2800.4620	1.89	N.S.	N.S.	TAAAQVGTSVSSATNTSTRPIITVHK	HexNAc	160607	18.8 / 19.8
	866.8010	3	2597.3773	2597.3812	1.49	N.S.	N.S.	TAAAQVGTSVSSATNTSTRPIITVHK		57376	20.6
638-659	1268.6730	2	2535.3280	2535.3314	1.35	45	1	SGTVTVAQQAQVVTTVVGGVTK	2 HexNAc	31763	21.9
	1167.1316	2	2332.2486	N.D.	N.D.	N.S.	N.S.	SGTVTVAQQAQVVTTVVGGVTK	HexNAc	N.D.	N.D.
	1065.5919	2	2129.1693	N.D.	N.D.	N.S.	N.S.	SGTVTVAQQAQVVTTVVGGVTK		N.D.	N.D.
683-713	1135.2730	3	3402.7917	3402.7972	1.62	58	1	VMoxSVVQTKPVQTSAVTGQASTGPVTQIIQTK	HexNAc	66822	24.9
	1067.5810	3	3199.7123	3199.7212	2.78	58	1	VMoxSVVQTKPVQTSAVTGQASTGPVTQIIQTK		3782	25.8
771-793	1210.6340	2	2419.2516	2419.2534	0.76	57	1	TIPMoxSAIITQAGATGVTSSPGIK	HexNAc	44428	24.9
	1109.0934	2	2216.1722	N.D.	N.D.	N.S.	N.S.	TIPMoxSAIITQAGATGVTSSPGIK		N.D.	N.D.
794-802	588.8403	2	1175.6649	1175.6660	0.96	5	4	SPITIITTK	HexNAc (T)	1854574	25.0
	487.2999	2	972.5855	972.5852	-0.30	44	1	SPITIITTK		114340	27.4
856-875	1113.1740	2	2224.3294	2224.3334	1.81	17	1	LVTPVTVSAVKPAVTTLVVK	HexNAc	182824	28
	1011.6345	2	2021.2500	2021.2544	2.18	N.S.		LVTPVTVSAVKPAVTTLVVK		39126	30.5
	674.7590	3	2021.2500	2021.2552	2.54	44	1	LVTPVTVSAVKPAVTTLVVK			
1233-1244	496.5697	3	1486.6834	1486.6873	2.58	9	2	HSHAVSTAAMoxTR	HexNAc	42294	8.9
	428.8763	3	1283.6041	1283.6071	2.34	N.S.		HSHAVSTAAMoxTR		86875	9.3
	642.8110	2	1283.6041	1283.6074	2.64	46	1	HSHAVSTAAMoxTR			
1483-1500	985.0272	2	1968.0416	1968.0398	-0.88	34	1	AVTTVTQSTPVPGPSVPK	HexNAc (S)	107656	22.7
	883.4901	2	1764.9622	1764.9656	1.92	38	1	AVTTVTQSTPVPGPSVPK		300096	24.2

Fig. S2. O-glycosylation of HCF1. Peptides ions carrying a labile post-translational modification of mass 203.07937, potentially corresponding to an N-acetyl-hexosamine, were detected by Mascot database searching for the protein HCF-1 (Uniprot accession number P51610). The table shows the position of the detected peptides in the protein sequence, and for each peptide ion: the mass to charge ratio (m/z), charge (z), theoretical mass, experimental mass, calculated mass deviation in ppm, Mascot identification score, Mascot rank (sequences of rank 1correspond to the best matching sequence for a given MS/MS spectrum), peptide sequence, posttraductional modification assigned by Mascot, integrated elution peak area (reflecting the abundance of the peptide ion), and elution time from the nanoLC column. As HCF-1 was identified in various bands cut from the migration lanes of the THAP3 immunopurified complexes, MS/MS identification data and peak area are given only for the band in which a given glycosylated peptide was identified with the best Mascot score. For each detected glycosylated peptide ion, the MS signal of its non-glycosylated counterpart ion was also extracted to evaluate approximately the level of glycosylation. In some cases, the corresponding non-glycosylated peptide ion of same charge was not sequenced (N.S.), but MS/MS identification data is given for ions of lower or higher charge state, if available. Moreover, in some cases, no MS signal could be detected at all for non-glycosylated species (N.D.). Although some species assigned by Mascot as glycosylated peptides were identified with a very low score, sequences shown in this table are proposed to be glycosylated based on the following criteria: i) detection of a peptide ion bearing a modification of 203.07937 Da with high mass accuracy (mass deviation <4ppm) on a high resolution Orbitrap mass spectrometer, ii) detection in the MS/MS spectrum of an intense ion corresponding to a loss of 203.1 Da from the parent ion, characteristic of a labile O-glycosylation lost under CID (collision-induced dissociation) MS/MS sequencing, and iii) detection of the nonglycosylated peptide ion with a shift in elution time of about 1-2 min from the glycosylated peptide, due to different migration behaviour on the reverse-phase nanoLC column. (N.B.: for species assigned by Mascot as doubly glycosylated peptides, two elution times are indicated in some cases due to the detection of two very close but distinct elution peaks, possibly related to heterogeneous patterns of glycosylation on the peptide). Doubly-glycosylated peptides are shown in red, singly-glycosylated peptides in orange, non-glycosylated peptides in yellow.

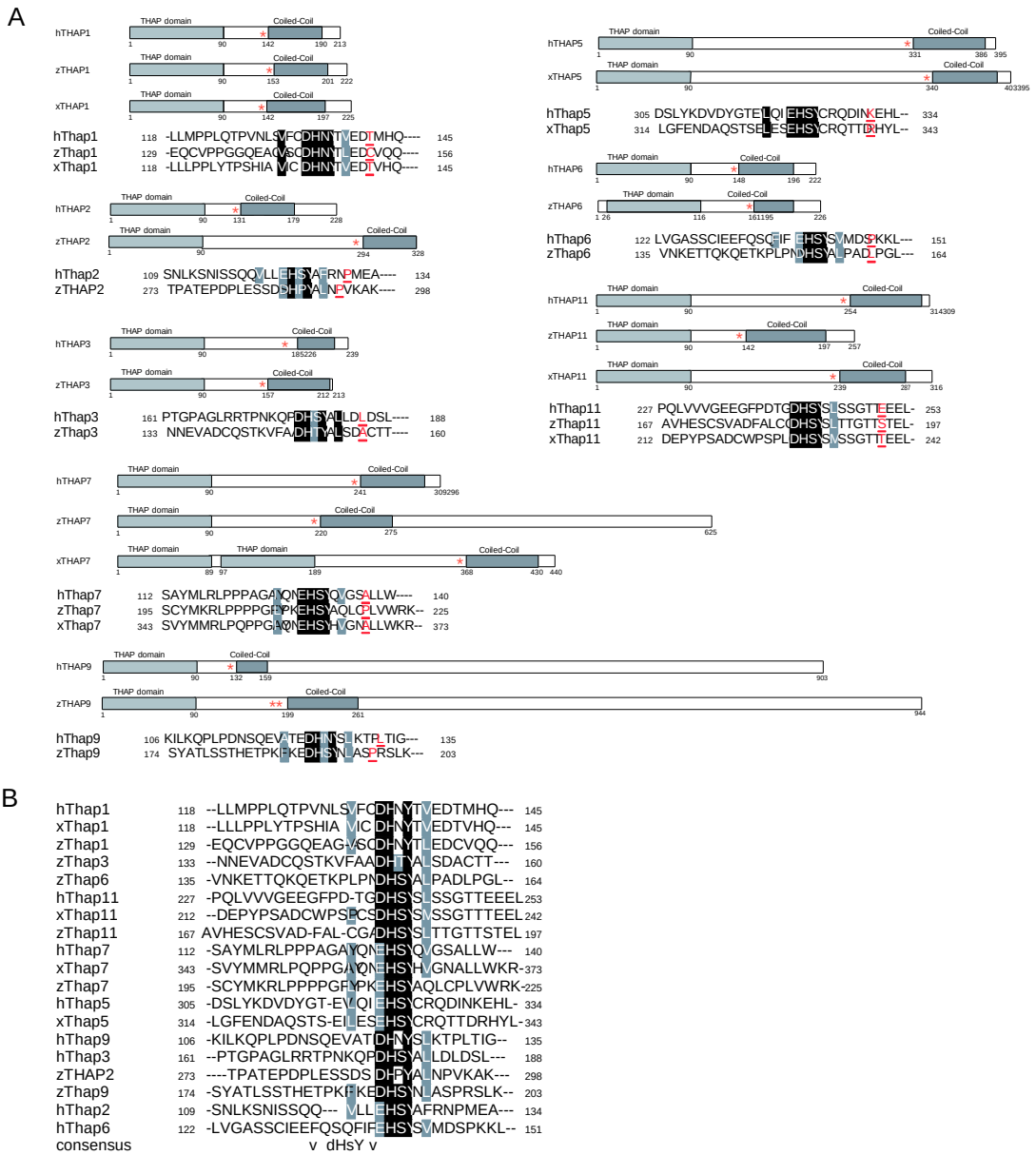


Fig S3. Evolutionary conservation of the HBMs in THAP-zf proteins. (A) The primary structure of the eight human THAP-zf proteins containing consensus HBMs and their putative orthologues in zebrafish and xenopus is shown above the alignment of their HBMs. The THAP-zf is indicated in grey and the predicted coiled-coil domains in dark grey. Red asterisk indicates the position of the HBM in each primary sequence. Residues underlined in red correspond to the first residues of the predicted coiled-coil domains. (B) Multiple alignment of the HBMs found in THAP-zf proteins. The alignment was generated with Clustal W (http://www.ebi.ac.uk/cgibin/clustalw) according to the Blosum matrix and colored with Boxshade (http://www.ch.embnet.org). Black boxes indicate identical residues, whereas shaded boxes show similar amino-acids. Dashed lines represent gaps introduced

Protein	#	amino acids
THAP3	3	128-239 153-239
THAP4	1	291-577
THAP7	5	223-309
THAP11	5	73-313 203-313 206-313 216-313

Fig. S4. Identification of human THAP-zf protein cDNAs in large scale two-hybrid screens with an HCF-1 bait. HF7c yeast strain was transformed with pGAL29 encoding the HCF-1 kelch domain (amino acids 3-455) fused to the GAL4 DNA binding domain. The resulting strain was transformed with a HeLa Matchmaker library (Clontech) and plated on SDLeu/Trp/His media. DNA was isolated from His+/β-galactosidase+ clones and the nucleotide sequence was determined. Of the positive cDNA clones, 14 encoded THAP proteins as detailed in the Table. The number of times a clone was isolated encoding the indicated THAP protein is shown (#). The portion of the THAP-zf protein encoded in various isolated clones is indicated.

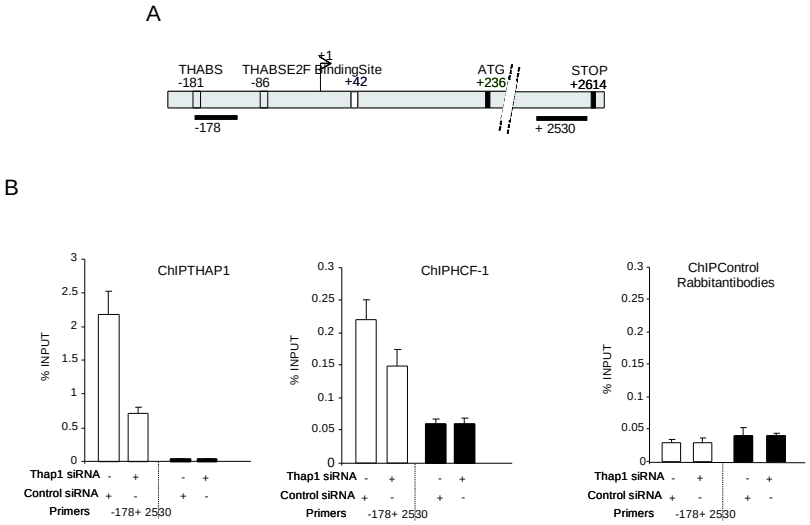


Fig. S5. Endogenous THAP1 recruits endogenous HCF-1 to the *RRM1* promoter inproliferating HUVECs. (A) Schematic representation of the human *RRM1* promoter. The bindingsites for THAP1 (THABS) and E2Fs are indicated. The position of the DNA fragments analyzedin ChIP-qPCR assays are shown. (B) Knockdown of THAP1 reduces recruitment of THAP1 and HCF1 to the *RRM1* promoter *in vivo*. ChIP-qPCR assays with anti-THAP1, anti-HCF1 or ChIP control (Abcam) rabbit antibodies were performed using proliferating primary HUVECs treated with control or THAP1 siRNAs. Immunoprecipitated DNA was quantified in triplicate by qPCR using the % of input method. In brief, the amount of genomic DNA co-precipitated with anti-THAP1 antibodies was calculated as a % of total input the following way: ΔCT=CT(input)-CT(THAP1-IP), %input=$2_{\Delta CT}$x0.25 (0.25% of total input was used). These independent experiments provide additional evidence for an important role of THAP1 in HCF-1 recruitment to the *RRM1* promoter *in vivo* (see Figure 5 of the manuscript)

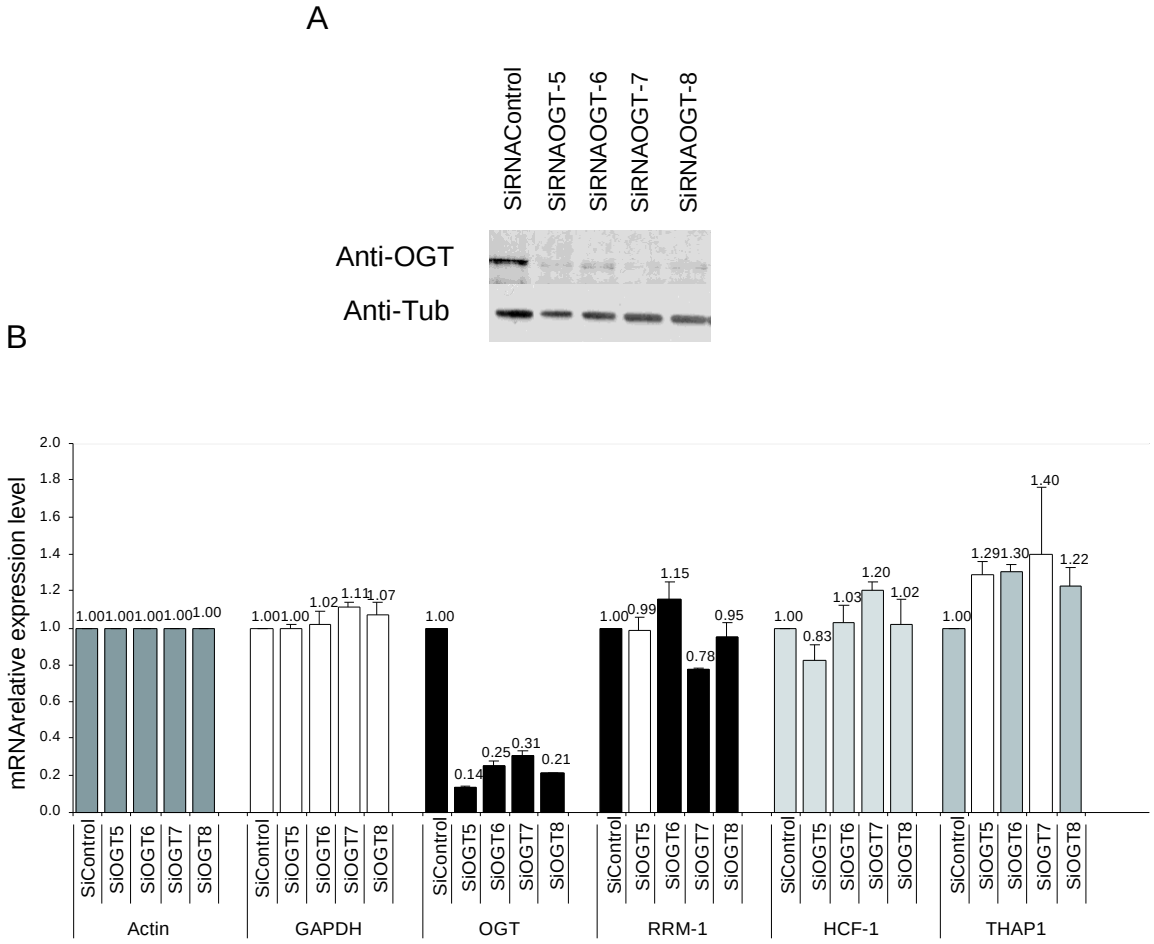


Fig. S6. Knockdown of endogenous OGT in primary HUVECs does not modify *RRM1* mRNA levels. (A,B) Knockdown of endogenous OGT was performed with 4 individual ONTARGET-plus OGT siRNAs. (A) OGT and Tubα (loading control) expression levels were analyzed by western blot. (B) RNA was isolated from cells transfected with individual OGT siRNAs, 48h after siRNA transfection, and used for qPCR analysis with the indicated human gene primers (*OGT, RRM1, HCF-1, THAP1* and control gene *GAPDH*). *Actin* was used as a control gene for normalization. Results are shown as means with s.d. from 3 separate datapoints.

Annexe 2 : Supplementary data de l'article Mourgues, Gautier et al., "ELL, a novel TFIIH partner is involved in transcription restart after DNA repair" en préparation

METHODS

Construction and expression of ELL-GFP fusion protein
Full length ELL cDNA was cloned in-frame into pEGFP-N1 vector (clontech, Heidelberg, Germany). Construct was sequenced prior to transfection. Transfection in MRC5-SV40 transformed human fibroblasts was performed using Fugene transfection reagent (Roche). Stably expressing cells were isolated after selection with G418 (Gibco) and single cell sorting using FACS (FACScalibur, Beckton Dickinson).

Cell culture and specific treatment
Cell strains used were: *(i)* XPB-YFP expressing embryonic stem (ES) cells[5], *(ii)* wild-type embryonic stem cells (IB10), *(iii)* wild type SV40-immortalized human fibroblasts (MRC5) stably expressing ELL-GFP, *(iv)* MRC5 stably expressing Cdk7-GFP; *(v)* MRC5 transiently expressing RNA Pol II-GFP[34], *(vi)* XPB deficient SV40-immortalized human fibroblasts (XPCS2BA) stably expressing XPB-GFP, *(vi)* XPC deficient SV40-immortalized human fibroblasts (XP4PA). ES cell lines were cultured in BRL-conditioned medium supplemented with 1000U/ml leukemia inhibitor factor. Human fibroblasts were cultured in a 1:1 mixture of Ham's F10 and DMEM (Lonza) supplemented with antibiotics and 10% fetal calf serum, at 37°C, 20 % O_2 and 5% CO_2. Cells were treated with 150 µg/ml of 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB, sigma) for 6h at 37°C. DNA damage was inflicted by UV-C light (254 nm, 6W UVC lamp). For UV survival experiments, cells were exposed to different UV-C doses, 1 day after plating. Survival was determined by clones counting 10 days after UV irradiation, as described previously[35]. For RNA recovery synthesis (RRS) and Unscheduled DNA synthesis (UDS), cells were either globally irradiated to 16 J/m^2 of UV-C or locally irradiated with 100 J/m^2 of UV-C through a 5 µm pore polycarbonate membrane filter (Millipore).

Immunoprecipitations, western blot and mass spectrometry analysis
Whole-cell extracts of wild type (IB10, for the mock) and XPB-YFP expressing ES cells were prepared. Pellets of scrapped cells (at least 20 to 30 million cells) were washed with phosphate-buffered saline (PBS) then snap-freezed in liquid nitrogen, and lysed using a Dounce homogeneizer (Bellco Glass, 20 strokes using large pestle A and 20 strokes using small pestle B) in 2 ml of immunoprecipitation (IP) buffer (50 mM Tris pH 7.9, 150 mM NaCl, 20% glycerol, 0.1% Nonidet-P40, 5 mM β-mercaptoethanol), supplemented with anti-proteases (Complete, Roche) and 0.1mM PMSF. Cellular extracts were incubated overnight at 4°C in IP buffer with mouse GFP trap antibody covalently coupled to sepharose beads (Chromotek). After immunoprecipitation, beads were washed three times with 20 volumes of IP buffer. Subsequently, beads were boiled in Laemmli SDS-PAGE sample buffer, and eluted proteins were separated on 8% SDS-PAGE, blotted onto nitrocellulose and analyzed using either anti-HA (Roche) or anti-ELL (abcam ab64824) antibodies. For proteomic analysis, immunopurified and control protein samples were also eluted from the beads with Laemmli buffer, reduced and alkylated with iodoacetamide, and processed through 1D SDS-PAGE without fractionation of the proteins. A very short migration was performed and a single gel band-containing the whole sample was cut. Proteins were in-gel digested by modified sequencing-grade trypsin (Promega). Resulting peptides were extracted and analyzed by nanoLC-MS/MS using an Ultimate3000 system (Dionex) coupled to an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) as previously described[6]. The Mascot software (Matrix Science) was used to perform database searches against mouse Uniprot protein database (56291 sequences). Relative quantification of proteins from immunopurified and control samples was performed with the MFPaQ software by computing for each protein a protein abundance index (PAI), defined as the average of the XIC area values for the three most intense identified peptides. Four independent experiments were performed, in which the ELL protein was repeatedly quantified as a specific partner of the immunopurified complex.

GST pull down
GST or GST-ELL polypeptides were expressed in E. coli and incubated with glutathione agarose beads. The various subunits of the core TFIIH were expressed in E. coli (1×10^6 cells), and cell lysates were incubated with 5 µg of GST or GST-Ell proteins bound to beads at 4°C for 1 h. Pull downs were analyzed by Western blotting.

Fluorescence Recovery after Photobleaching (FRAP)
FRAP experiments were performed as described before[14,15] on a Zeiss LSM 710 NLO confocal laser scanning microscope (Zeiss), using a 40x/1.3 oil objective, under a controlled environment (37°C, 5% CO_2). Briefly, a narrow region of interest (ROI) centered across the nucleus of a living cell was monitored every 20 or 200 ms (1% laser intensity of the 488 nm line of a 25 mW Argon laser) until the fluorescence signal reached a steady state level (after circa 2 s). The same strip was then photobleached for 20 ms at 100% laser intensity. Recovery of fluorescence in the strip was then monitored (1% laser intensity) every 20 or 200 ms for about 20 or 40 seconds respectively. Analysis of raw data was

performed with the ZEN software (Zeiss). All FRAP data were normalized to the average prebleached fluorescence after background removal. Every plotted FRAP curve is an average of at least ten measured cells.

Laser micro-irradiation
In order to locally induce DNA damage in living cells, we used a tunable near-infrared pulsed laser (Cameleon Vision II, Coherent Inc.) directly coupled to an inverted confocal microscope equipped with a 40x/1.3 oil objective and a thermostatic chamber maintained at 37°C with 5% CO_2 (LSM 710 NLO, Zeiss). Typically, a small circular area (~2.5 µm in diameter) within the nucleus of a living cell was targeted for ~35 ms (one iteration at 800 nm, 25% power output). Subsequent time-lapse imaging of targeted cells was performed every 15 s for 345 s (1% laser intensity, 488 nm Argon line). Image analysis was performed using ImageJ (Rasband, W.S., National Institutes of Health, USA) and a custom-built macro as follows: (i) the time series image stack was adjusted to compensate for cell movement (StackReg plugin), (ii) a ROI spanning the total nucleus was defined to compensate for unwanted photobleaching during the acquisition of images, (iii) a 'local damage' ROI was specified to quantify the fluorescence increase due to (GFP tagged) protein recruitment at the laser induced DNA damage area. At least ten cells were measured for all three cell lines.

RNA interference
Short interfering RNA (siRNA) used in this study are: *(i)* ELL siRNA (Santa Cruz, Sc-38041); *(ii)* Cdk7 siRNA (Dharmacon, L-003241-00-0005); *(iii)* XPF siRNA (Dharmacon, M-019946-00);*(iv)* CSB siRNA (pool of 2 séquences : UGAAGCAUCAGGCUUCGAAdTdT and AGAGAAACGUCUGAAGCUGdTdT)[36] and *(v)* non coding siRNA (Eurogentec) (n°SR-CL000-005). Cells were transfected using GenJET siRNA transfection reagent (Tebu-Bio) according to the manufacturer's protocol. Transfection complexes were formed by 15 min incubation at room temperature using buffer provided. Briefly, 100.000 cells were seeded per 3 cm dish and allowed to attach overnight. siRNAs were added 24h (5 nM for ELL, 10 nM for Cdk7, XPF and CSB) after seeding, and cells were grown confluent. Experiments were carried out 24 or 48h after seeding. Proteins knock down was confirmed by quantitative RT-PCR.

Quantitative RT-PCR (RT-qPCR)
Total RNA was isolated from siRNA-transfected cells using RNEasy mini kit (Qiagen). cDNA was synthetized using random hexamer primers and SuperScript II Reverse Transcriptase (Invitrogen). ELL, Cdk7, XPF and CSB expression levels were analyzed using RT-qPCR with the SyberGreen Gene expression assay, using 7300 Real time PCR machine (Applied Biosystems). ELL, Cdk7, XPF, CSB expression levels were normalized to HPRT expression.

Recovery of RNA synthesis (RRS) assays.
MRC5-SV40 cells were grown on 24 mm coverslips. siRNA transfections were performed 24h before RRS assays. RNA detection was performed using a Click-iT RNA Alexa Fluor Imaging kit (Invitrogen), according to the manufacturer's instructions. Briefly, cells were UV irradiated (16 J/m^2 of 254 nm UV-C) and incubated for 16 h at 37°C. Then, cells were incubated for 2 hours with 5-ethynyl uridine, fixed and permeabilized. Cells were incubated for 30 min with the Click-iT reaction cocktail containing Alexa Fluor Azide 488. After washing, the coverslips were mounted with Vectashield (Vector). Images of the cells were obtained with the same setup (see FRAP methods section) and constant acquisition parameters, then the average fluorescence intensity per nucleus was estimated after background subtraction (using ImageJ) and normalized to the mock treated cells. For each sample, at least 100 nuclei were analyzed from three independent experiments.

Unscheduled DNA synthesis (UDS) assays.
MRC5-SV40 cells were grown on 24 mm coverslips. siRNA transfections were performed 24h before UDS assays. *De novo* DNA synthesis detection was performed using a Click-iT DNA Alexa Fluor Imaging kit (Invitrogen), according to the manufacturer's instructions. Briefly, after global irradiation cells were incubated for 3 hours with ethynyldeoxyuridine, then cells are washed with PBS, fixed and permeabilized. Fixed cells were incubated for 30 min with the Click-iT reaction cocktail containing Alexa Fluor Azide 594. After washing, the coverslips were mounted with Vectashield (Vector). Images of the cells were acquired and analyzed in the same way as for the RRS assay (see previous paragraph). For each sample, at least 100 nuclei were analyzed from three independent experiments.

TCR-UDS assays: UDS measurement during TCR
XPC deficient SV40-immortalized human fibroblasts (XP4PA-GGR deficient cell line), were grown on 24 mm coverslips. siRNA transfections were performed 24h before UDS assays. After local irradiation (100 J/m^2 of 254 nm UV-C) through a 5 µm pore polycarbonate membrane filter, cells were incubated for 8 hours with ethynyldeoxyuridine, washed, fixed and permeabilized. Fixed cells were treated with a PBS-blocking solution (PBS+: PBS containing 0.15% glycine and 0.5% bovine serum albumin) for 30 min, subsequently incubated with primary antibodies mouse monoclonal anti-γH2AX (Ser139) (Upstate, clone JBW301) 1/500 diluted in PBS+ for 1h, followed by extensive PBST washes. Cells were then incubated for 1h with secondary antibodies conjugated with Alexa Fluor 488 fluorescent dyes (Molecular Probes, 1:400 dilution in PBS+). Then, cells were incubated for 30 min with the Click-iT reaction cocktail containing

Alexa Fluor Azide 594. After washing, the coverslips were mounted with Vectashield (Vector). Images of the cells were obtained with the same microscopy system and constant acquisition parameters. Images were analyzed using ImageJ as follows: (i) a ROI outlining the locally damaged area was defined by using the γH2AX staining, (ii) a second ROI of comparable size was defined in the nucleus (avoiding nucleoli and other non-specific signals) to estimate background signal, (iii) the 'local damage' ROI was then used to measure the average fluorescence correlated to the EdU incorporation, which is an estimate of DNA replication after repair once the nuclear background signal obtained during step (ii) is subtracted. For each sample, between 90 and 100 nuclei were analyzed from three independent experiments, except for cells treated with CSB siRNA (50 cells from two experiments).

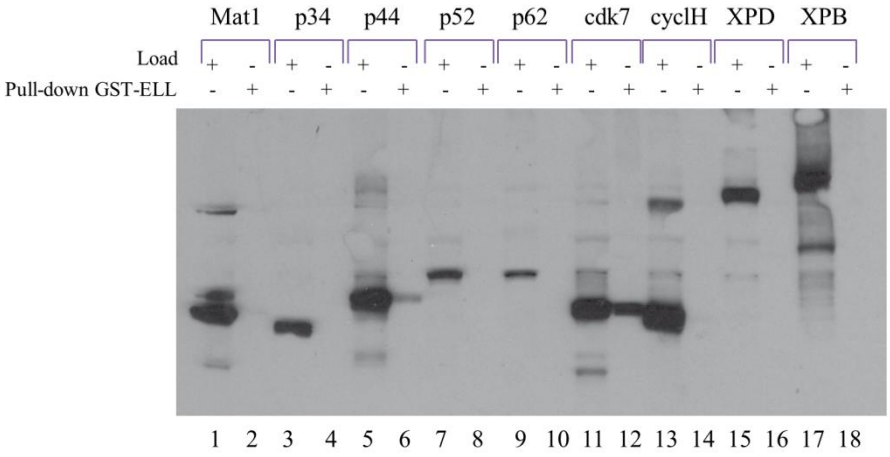
34 Mari, P. O. *et al.* Influence of the live cell DNA marker DRAQ5 on chromatin-associated processes. *DNA repair* **9**, 848-855, doi:10.1016/j.dnarep.2010.04.001 (2010).
35 Jensen, A. & Mullenders, L. H. Transcription factor IIS impacts UV-inhibited transcription. *DNA repair* **9**, 1142-1150, doi:10.1016/j.dnarep.2010.08.002 (2010).
36 Liu, F., Yu, Z. J., Sui, J. L., Bai, B. & Zhou, P. K. siRNA-mediated silencing of Cockayne Cyndrome group B gene potentiates radiation-induced apoptosis and antiproliferative effect in HeLa cells. *Chinese medical journal* **119**, 731-739 (2006).

Supplementary information

Supplementary Table 1 XPB-associated proteins of interest identified by quantitative proteomic Analysis.

Proteins were identified with the Mascot software and the quantitative analysis was performed with the MFPaQ software. Supplementary Table 1 shows protein name and description, Uniprot accession number (AC), gene name, molecular weight (MW) in dalton (Da), and for each of the 4 replicates, the Mascot score of the protein identified in the assay (IP score) and in the control (CT score), number of quantified peptides (Qpep) and the ratio IP/CT. Four independant experiments were performed and proteins displayed in the table were quantified as specific proteins of the assay and classified as potential specific partners.

Protein name	Protein description	AC	Gene name	MW (Da)	IP score (4 replicates)	CT score (4 replicates)	Qpep (4 replicates)	IP/CT ratio
XPB	TFIIH basal transcription factor complex helicase XPB subunit	P49135	*ERCC3*	89925	6682;5845;1848;2886	-;-;-;-	61;60;16;34	IP specific
XPD	TFIIH basal transcription factor complex helicase XPD subunit	Q8C487	*ERCC2*	87585	1624;1074;1394;1496	-;-;-;-	32;24;23;35	IP specific
p62	General transcription factor IIH subunit 1	Q9DBA9	*Gtf2h1*	62155	3426;2422;6116;2086	-;-;-;-	40;36;21;30	IP specific
p52	General transcription factor IIH subunit 4	O70422	*Gtf2h4*	52362	3191;2045;9115;2741	-;-;-;-	26;27;17;21	IP specific
p44	General transcription factor IIH subunit 2	Q9JIB4	*Gtf2h2*	45742	2113;1691;8476;2268	-;-;-;-	30;25;16;25	IP specific
p34	General transcription factor IIH subunit 3	Q8VD76	*Gtf2h3*	34906	1886;1424;1408;1473	-;-;-;-	16;16;13;15	IP specific
TTD-A	General transcription factor IIH subunit 5	Q8K2X8	*Gtf2h5*	8089	140;179;-;-	-;-;-;-	1;2;-;-	IP specific
MAT1	CDK-activating kinase assembly factor MAT1	P51949	*Mnat1*	36338	874;582;2480;1174	-;-;-;-	16;11;13;17	IP specific
CDK7	Cell division protein kinase 7	Q03147	*Cdk7*	39229	707;185;2411;860	-;-;-;-	11;3;15;15	IP specific
Cyclin H	Cyclin-H	Q3UUW5	*Ccnh*	38537	557;181;1027;583	-;-;-;-	11;4;12;18	IP specific
XPG	DNA repair protein complementing XP-G cells homolog	Q3UV64	*ERCC5*	131215	2016;1196;550;793	-;-;-;-	36;26;18;27	IP specific
ELL	RNA polymerase II elongation factor ELL	O08856	*ELL*	67146	853;427;101;119	-;-;-;-	19;12;5;4	IP specific
EAF1	ELL-associated factor 1	Q9D4C5	*Eaf1*	29120	658;146;-;-	-;-;-;-	3;4;-;-	IP specific
MKIAA0947	MKIAA0947 protein	Q6ZQ20	*mKIAA0947*	187648	878;224;79;272	-;-;-;-	17;5;3;4	IP specific
NARG2	NMDA receptor-regulated protein 2	Q3UZ18	*Narg2*	111007	703;168;-;-	-;-;-;-	17;5;-;-	IP specific

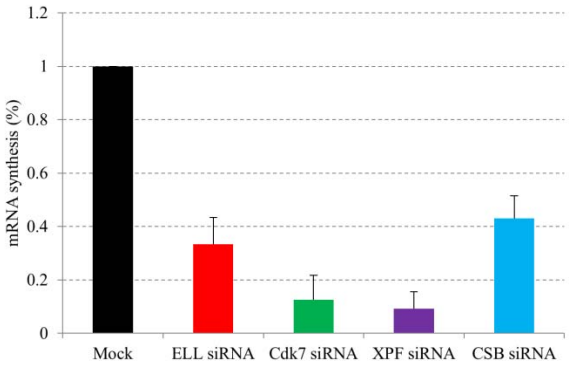


Supplementary Figure 1 ELL interact with Cdk7 subunit of TFIIH

SDS-PAGE analysis showing purified TFIIH sub-units (lane 1, 3, 5, 7, 9, 11, 13, 15, 17), pull-down GST-ELL (lane 2, 4, 6, 8, 10, 12, 14, 16, 18). Cdk7 is the only sub-unit of TFIIH interacting with ELL (lane12).

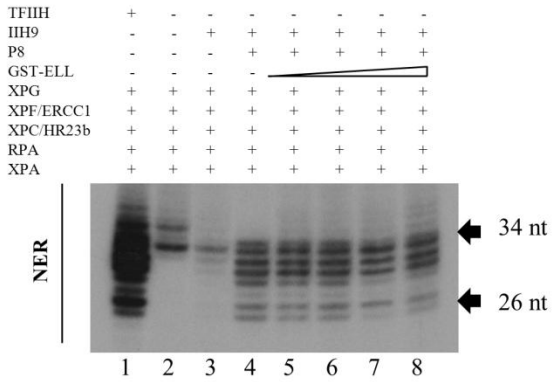
Supplementary Table 2 : gene specific primers used for RT-QPCR.

gene	5'-primer	3'-primer
ELL	ACCCCAGGTTTAAACGGAAC	TGTACTCGGCATTGAAGTCG
Cdk7	GGCCGGACATGTGTAGTCTT	CAGCTGACATCCAGGTGTTG
XPF	CGACACTGACGGGCTAGTAG	CGAGGGAGGTGTTCAACTC
CSB	AGTGAGGCCAAGGAACAGAG	TGATGGCATCGTGCTTCAT
HPRT	TGACACTGGCAAAACAATGCA	GTCTGCGACCTTGACCATCT



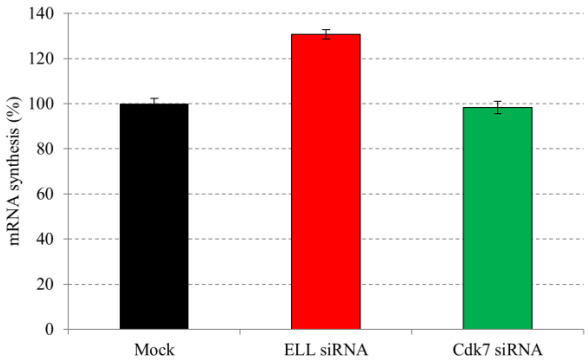
Supplementary Figure 2 qRT-PCR showing the mRNA levels of ELL in MRC5 cells.

ELL, Cdk7, XPF and CSB expression levels were analyzed using qRT-PCR with the SyberGreen Gene expression assay, using 7300 Real time PCR machine (Applied Biosystems). ELL, Cdk7, XPF, CSB expression levels were normalized to HPRT expression.



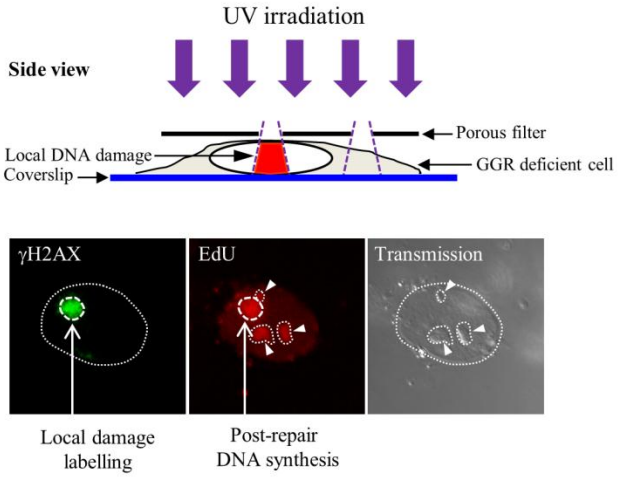
Supplementary Figure 3 ELL does not activate DNA repair *in vitro*

One hundred nanograms of recombinant TFIIH (IIH9 (lanes 3-8)) together with recombinant TTDA (20ng, lanes 4-8) was tested in a dual incision assay using Pt-DNA template in the presence of increasing amount (50 to 200ng) of GST-ELL (lane 5-8). Lane 1 contains TFIIH from HeLa immunoprecipitated with Ab-p44. Lane2 contains all the NER factors except TFIIH.



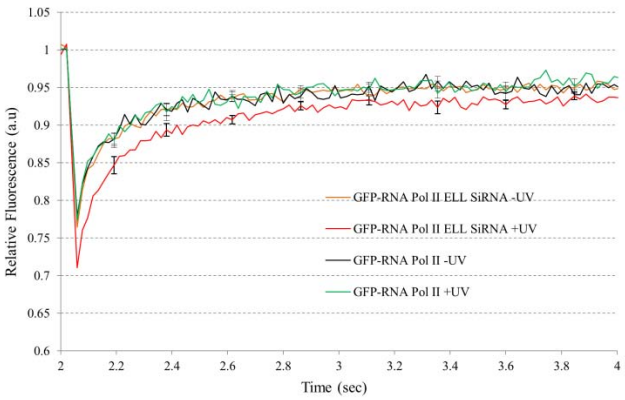
Supplementary figure 4 Transcription is not affected in cells depleted in ELL or Cdk7 proteins.

RRS graph in MRC5 cells with siRNA-mediated knockdown of the indicated factors. At least 50 cells were analyzed (mean ± s.e.m).



Supplementary figure 5 Scheme of Unscheduled DNA synthesis after local damage induction by UV

Representative GGR deficient cells after local UV exposure (100 J/m2) through a 5 µm pore polycarbonate membrane filter. 8h after irradiation, cells were incubated with ethynyldeoxyuridine, fixed and permeabilized. Cells were then immunostained with γH2AX to visualize local damage. Arrow heads indicate nucleoli (transmission Image) and their corresponding non specific EdU signal. Replication after repair is visualized using Zeiss LSM 710 inverted confocal microscope.



Supplementary figure 6 RNA Pol II binding on the chromatin during transcription coupled repair

FRAP recovery curves of GFP-RNA Pol II (UC4) transiently expressing cells and GFP-RNA Pol II with siRNA knockdown of ELL untreated (black lane and blue lane respectively) and treated with 20 J/m2 UV-C (green lane and red lane respectively). For each line at least 30 different cells were measured. Error bars are included in the curves and represent s.e.m.

Annexe 3 : Table des N-glycopeptides identifiés dans les trois réplicats des expériences d'enrichissement des N-glycoprotéines de surface avec la phosphine biotine après digestion PNGase.

Description	Gène	AC	Nombre de séquences uniques	Nombre MS/MS total	Signal sécrétion prédit (ProteinCenter)	Segments transmembranaires prédis (TMHMM)
Multimerin-1	MMRN1	IPI00012269	29	332	S	0
Integrin beta-1	ITGB1	IPI00217563	27	476	S	1
Isoform 4 of Fibronectin	FN1	IPI00339224	20	247	S	0
Isoform A of Endothelin-converting enzyme 1	ECE1	IPI00216758	18	356		1
Intercellular adhesion molecule 2	ICAM2	IPI00009477	15	510	S	1
Integrin alpha-5	ITGA5	IPI00306604	15	251	S	1
Lysosome-associated membrane glycoprotein 2	LAMP2	IPI00009030	14	240	S	1
Lysosome-associated membrane glycoprotein 1	LAMP1	IPI00884105	14	255	S	2
Lysosome membrane protein 2	SCARB2	IPI00217766	14	121	S	2
von Willebrand factor	VWF	IPI00023014	13	248	S	0
Laminin subunit gamma-1	LAMC1	IPI00298281	13	98	S	0
Vascular cell adhesion protein 1	VCAM1	IPI00018136	12	80		1
Cadherin-5	CDH5	IPI00012792	12	109	S	2
Platelet endothelial cell adhesion molecule	PECAM1	IPI00295618	12	158	S	1
Aminopeptidase N	ANPEP	IPI00221224	12	217	S	1
Intercellular adhesion molecule 1	ICAM1	IPI00008494	12	279	S	1
CD63 antigen	CD63	IPI00749070	11	258	S	4
Integrin alpha-V	ITGAV	IPI00027505	11	93	S	1
CD166 antigen	ALCAM	IPI00015102	11	187	S	1
Integrin alpha-2	ITGA2	IPI00013744	10	73	S	1
N-acetylglucosamine-1-phosphotransferase subunits alpha/beta	GNPTAB	IPI00382432	10	115	S	2
Urokinase plasminogen activator surface receptor	PLAUR	IPI00010676	10	87	S	0
Stabilin-1	STAB1	IPI00419565	10	104	S	2
Isoform 3 of Choline transporter-like protein 2	SLC44A2	IPI00645589	8	243	S	11
Isoform 1 of Neuroplastin	NPTN	IPI00011578	8	160	S	1
Isoform 2 of Golgi apparatus protein 1	GLG1	IPI00414717	8	222	S	1
Isoform 3 of Tetraspanin-3	TSPAN3	IPI00910611	8	81	S	2
Proactivator polypeptide	PSAP	IPI00873020	8	107	S	0
Angiotensin-converting enzyme	ACE	IPI00437751	7	77	S	1
Prostaglandin F2 receptor negative regulator	PTGFRN	IPI00022048	7	38	S	1
Carboxypeptidase D	CPD	IPI00027078	7	76	S	1
Leucyl-cystinyl aminopeptidase	LNPEP	IPI00307017	7	71		1
Vascular endothelial growth factor receptor 2	KDR	IPI00021396	7	56	S	2
Cation-independent mannose-6-phosphate receptor	IGF2R	IPI00289819	7	65	S	1
Cation-dependent mannose-6-phosphate receptor	M6PR	IPI00025049	7	68	S	1
Interleukin-6 receptor subunit beta	IL6ST	IPI00297124	7	72	S	1
Isoform 2 of Laminin subunit alpha-4	LAMA4	IPI00735310	7	50	S	0
Isoform 2 of Angiopoietin-2	ANGPT2	IPI00793960	7	71	S	0
Zinc transporter ZIP10	SLC39A10	IPI00008085	6	23		7
TMEM181 protein	TMEM181	IPI00166790	6	68		7
CD109 antigen	CD109	IPI00152540	6	58	S	0
Isoform 2 of Plexin-D1	PLXND1	IPI00472139	6	80	S	1
Basement membrane-specific heparan sulfate proteoglycan core protein	HSPG2	IPI00024284	6	82	S	0
Isoform 2 of Integrin alpha-3	ITGA3	IPI00290043	6	63	S	2
Isoform B of Disintegrin and metalloproteinase domain-containing protein 17	ADAM17	IPI00029606	6	63		1
CUB domain-containing protein 1	CDCP1	IPI00290039	6	55	S	1
neuronal cell adhesion molecule isoform D precursor	NRCAM	IPI00983847	6	52	S	1
Receptor-type tyrosine-protein phosphatase beta	PTPRB	IPI01021143	6	86		0
Multimerin-2	MMRN2	IPI00015525	6	83	S	0
Isoform 2 of Collectin-12	COLEC12	IPI00885028	6	51	S	1
Cadherin-13	CDH13	IPI01018976	6	45	S	0
Nicastrin	NCSTN	IPI00021983	5	25	S	1
Isoform 4 of Scavenger receptor class B member 1	SCARB1	IPI00447020	5	20		1
Prolow-density lipoprotein receptor-related protein 1	LRP1	IPI00020557	5	17	S	1
Isoform 3 of Angiopoietin-1 receptor	TEK	IPI00979987	5	82	S	2
E-selectin	SELE	IPI00296542	5	50	S	1
cDNA FLJ52545, highly similar to Dickkopf-related protein 3	DKK3	IPI00002714	5	59	S	0
Glycosyltransferase 8 domain-containing protein 1	GLT8D1	IPI00020470	5	48	S	1
Bone marrow stromal antigen 2	BST2	IPI00026241	5	81	S	1
High affinity cationic amino acid transporter 1	SLC7A1	IPI00027728	4	72		14
Sodium bicarbonate cotransporter 3	SLC4A7	IPI00021058	4	42		11
Isoform 2 of Transmembrane protein 87B	TMEM87B	IPI00847605	4	55	S	6
Transmembrane protein 87A	TMEM87A	IPI01013197	4	52		6
CD151 antigen	CD151	IPI00298851	4	71	S	4
Isoform 2 of 4F2 cell-surface antigen heavy chain	SLC3A2	IPI00027493	4	90		1
Isoform Placental I of Ectonucleoside triphosphate diphosphohydrolase 1	ENTPD1	IPI00220852	4	33	S	2
C-type mannose receptor 2	MRC2	IPI00005707	4	97	S	1
Isoform 2 of Basigin	BSG	IPI00019906	4	234	S	2
Isoform Delta of Poliovirus receptor	PVR	IPI00219427	4	41	S	1
Discoidin, CUB and LCCL domain-containing protein 1	DCBLD1	IPI00337612	4	20	S	1
Isoform 2 of Clusterin	CLU	IPI00400826	4	62		0
Heparan sulfate 2-O-sulfotransferase 1	HS2ST1	IPI00549891	4	59		0
HLA class I histocompatibility antigen, A-24 alpha chain	HLA-A	IPI00742968	4	69	S	1
V-type proton ATPase subunit S1	ATP6AP1	IPI00784119	4	15	S	2
Isoform 11 of CD44 antigen	CD44	IPI00827893	4	56	S	1
Plexin-B2	PLXNB2	IPI00853369	4	29	S	0
Inactive tyrosine-protein kinase 7	PTK7	IPI00298292	4	38	S	2
Ephrin type-A receptor 2	EPHA2	IPI00021267	4	46	S	1
Transferrin receptor protein 1	TFRC	IPI00022462	4	58		1
5'-nucleotidase	NT5E	IPI00009456	4	48	S	2
CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1	ST3GAL1	IPI00009629	4	42	S	1
Thrombospondin-1	THBS1	IPI00296099	4	64	S	0
Sodium/potassium-transporting ATPase subunit beta-1	ATP1B1	IPI00747849	4	62	S	1
Isoform 2 of Golgi membrane protein 1	GOLM1	IPI00759659	4	79	S	1
Serine protease 23	PRSS23	IPI00910747	4	37	S	0
glucosylceramidase isoform 2	GBA	IPI01011948	4	68		0
Piezo-type mechanosensitive ion channel component 1	PIEZO1	IPI00917650	3	28	S	31
Signal peptide peptidase-like 2A	SPPL2A	IPI00154588	3	21	S	9
Anoctamin-6	ANO6	IPI00917081	3	81		8
Probable G-protein-coupled receptor 116	GPR116	IPI00513936	3	5		7
Isoform 3 of Epidermal growth factor receptor	EGFR	IPI00221347	3	5	S	0
Isoform OA3-312 of Leukocyte surface antigen CD47	CD47	IPI00216516	3	69	S	6
Solute carrier family 43 member 3	SLC43A3	IPI00978999	3	61	S	7
Epithelial membrane protein 1	EMP1	IPI00008880	3	73		4
Teneurin-3	ODZ3	IPI00398020	3	6		1
Lysosome-associated membrane glycoprotein 3	LAMP3	IPI00004307	3	11		1
Junctional adhesion molecule 3 precursor	JAM3	IPI00152850	3	8		1
Isoform Alpha-6X1A of Integrin alpha-6	ITGA6	IPI00216221	3	26	S	1
Alpha-1,4-galactosyltransferase		IPI00433956	3	12		1

Endoglin	ENG	IPI00017567	3	24	S	1
EGF, latrophilin and seven transmembrane domain-containing protein 1	ELTD1	IPI00848229	3	28	S	7
Lysosomal acid phosphatase	ACP2	IPI00003807	3	16	S	1
HLA class I histocompatibility antigen, alpha chain E	HLA-E	IPI00010362	3	15	S	1
Fukutin-related protein	FKRP	IPI00013281	3	40		1
Disintegrin and metalloproteinase domain-containing protein 10	ADAM10	IPI00013897	3	28	S	1
Cell surface glycoprotein MUC18	MCAM	IPI00016334	3	114	S	1
Platelet endothelial aggregation receptor 1	PEAR1	IPI00154858	3	24	S	1
Isoform 2 of Scavenger receptor class A member 3	SCARA3	IPI00396540	3	28		1
Isoform 2 of Kin of IRRE-like protein 1	KIRREL	IPI00470361	3	34	S	1
NEGR1 protein	NEGR1	IPI00645710	3	7		0
Major prion protein	PRNP	IPI00646788	3	32	S	0
Isoform 2 of Myelin protein zero-like protein 1	MPZL1	IPI00760547	3	76	S	2
CD83 antigen isoform b	CD83	IPI00845356	3	17	S	1
Beta-galactosidase	GLB1	IPI01012577	3	51	S	0
Integrin beta-5	ITGB5	IPI00029741	3	31	S	1
Niemann-Pick C1 protein	NPC1	IPI00005107	3	43	S	12
Isoform 3 of Probable lysosomal cobalamin transporter	LMBRD1	IPI00640628	3	51		7
P2X purinoceptor	P2RX4	IPI01013466	3	63		1
Sodium/potassium-transporting ATPase subunit beta-3	ATP1B3	IPI00008167	3	24	S	1
Laminin subunit beta-1	LAMB1	IPI00013976	3	7	S	0
Dipeptidyl peptidase 4	DPP4	IPI00018953	3	22	S	1
ADP-ribosyl cyclase 2	BST1	IPI00026240	3	28	S	1
Follistatin-related protein 1	FSTL1	IPI00029723	3	34	S	0
Neprilysin	MME	IPI00247063	3	38	S	1
Granulins	GRN	IPI00296713	3	29	S	0
Carbohydrate sulfotransferase 12	CHST12	IPI00299758	3	43	S	1
Isoform 3 of Protein CASC4	CASC4	IPI00304892	3	31	S	1
Phospholipase D3	PLD3	IPI00328243	3	54	S	1
C-type lectin domain family 1 member A	CLEC1A	IPI00789445	3	42		0
HLA class I histocompatibility antigen, A-3 alpha chain	LOC100507680	IPI00892768	3	37	S	0
MHC class I polypeptide-related sequence A	MICA	IPI00942026	3	33	S	1
Isoform 3 of Solute carrier family 12 member 6	SLC12A6	IPI00783274	2	33		12
Natural resistance-associated macrophage protein 2	SLC11A2	IPI00793358	2	61		9
Multidrug resistance-associated protein 4	ABCC4	IPI00006675	2	10		11
Equilibrative nucleoside transporter 1	SLC29A1	IPI00939219	2	61	S	11
Transmembrane 9 superfamily member 3	TM9SF3	IPI00030847	2	59	S	9
G-protein coupled receptor 124	GPR124	IPI00743456	2	2	S	6
Membrane-bound transcription factor site-2 protease	MBTPS2	IPI00328263	2	19		6
Protein GPR107	GPR107	IPI00413788	2	20	S	7
Neutral amino acid transporter A	SLC1A4	IPI00913845	2	45	S	8
Glycerophosphodiester phosphodiesterase domain-containing protein 5	GDPD5	IPI00302491	2	15	S	6
Isoform 2 of Transmembrane and coiled-coil domain-containing protein 3	TMCO3	IPI00787600	2	21	S	6
Isoform 2 of Sialin	SLC17A5	IPI00411564	2	16	S	4
highly similar to Receptor-type tyrosine-protein phosphatase F (EC 3.1.3.48)	PTPRF	IPI00853550	2	13		1
Similar to Plexin A2	PLXNA2	IPI00872947	2	5	S	1
Isoform 3 of Protocadherin Fat 4	FAT4	IPI00888207	2	2	S	1
Dermatan-sulfate epimerase	DSE	IPI01013307	2	3		3
Tyrosine-protein kinase Mer	MERTK	IPI00029756	2	11	S	1
Epithelial membrane protein 3	EMP3	IPI00008901	2	28	S	4
Tetraspanin-31	TSPAN31	IPI00030418	2	25	S	4
Hedgehog-interacting protein precursor	HHIP	IPI00045106	2	7	S	1
Transmembrane protein 179B	TMEM179B	IPI00334453	2	27	S	3
Protein tweety homolog 3	TTYH3	IPI00749429	2	32	S	5
Isoform 3 of Voltage-dependent calcium channel subunit alpha-2/delta-1	CACNA2D1	IPI00953206	2	27	S	0
Tumor-associated calcium signal transducer 2	TACSTD2	IPI00297910	2	14	S	1
TGF-beta receptor type-2	TGFBR2	IPI00020431	2	16		1
Cell cycle control protein 50A	TMEM30A	IPI00019381	2	11		2
Proteinase-activated receptor 1	F2R	IPI00296869	2	17	S	7
Programmed cell death 1 ligand 1	CD274	IPI00023021	2	21		1
Transmembrane protein 2	TMEM2	IPI00170706	2	3		1
Type 2 lactosamine alpha-2,3-sialyltransferase	ST3GAL6	IPI00184851	2	39		1
Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase	GCNT1	IPI00295368	2	7		1
Isoform 2 of Vascular endothelial growth factor receptor 3	FLT4	IPI00337568	2	12	S	1
TM2 domain-containing protein 1	TM2D1	IPI00646722	2	11	S	3
Protein ITFG3	ITFG3	IPI00658094	2	33		1
Isoform 2 of Lymphocyte antigen 75	LY75-CD302	IPI00789911	2	5	S	1
Isoform 2 of Transmembrane protein 206	TMEM206	IPI00980690	2	21		1
Sortilin	SORT1	IPI01008999	2	11		1
Fibroblast growth factor receptor-like 1	FGFRL1	IPI00296561	2	17	S	1
Neuropilin-1	NRP1	IPI00299594	2	18	S	1
Gamma-aminobutyric acid type B receptor subunit 2	GABBR2	IPI00027250	2	22	S	7
Integrin beta-3	ITGB3	IPI00303283	2	26	S	1
C-X-C chemokine receptor type 7	CXCR7	IPI00012733	2	28		7
Thrombomodulin	THBD	IPI00010737	2	28	S	1
Isoform 3 of CD97 antigen	CD97	IPI00397230	2	29	S	7
HLA class I histocompatibility antigen, Cw-2 alpha chain	HLA-C	IPI00472605	2	29	S	1
Macrosialin	CD68	IPI00002252	2	6	S	1
Ectonucleotide pyrophosphatase/phosphodiesterase family member 4	ENPP4	IPI00007249	2	14	S	1
N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase	NAGPA	IPI00008303	2	26	S	1
Ephrin type-B receptor 2	EPHB2	IPI00021275	2	30	S	1
Isoform 2 of TM2 domain-containing protein 3	TM2D3	IPI00018607	2	14	S	1
Ephrin-B1	EFNB1	IPI00024307	2	21	S	1
Transmembrane emp24 domain-containing protein 7	TMED7	IPI00032825	2	9	S	1
Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 3	ST6GALNAC3	IPI00166295	2	10	S	1
Isoform 2 of UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 2	B3GNT2	IPI00217345	2	13		0
Isoform Beta of Tissue factor pathway inhibitor	TFPI	IPI00218184	2	22	S	0
Isoform Alpha of Poliovirus receptor-related protein 1	PVRL1	IPI00218887	2	9	S	2
ICOS ligand	ICOSLG	IPI00219131	2	26	S	1
C-type lectin domain family 14 member A	CLEC14A	IPI00240345	2	13	S	1
Membrane protein FAM174A	FAM174A	IPI00290826	2	4	S	2
Lysosomal protein NCU-G1	C1orf85	IPI00647672	2	31	S	2
Junctional adhesion molecule B	JAM2	IPI00299083	2	11	S	1
Tumor necrosis factor receptor superfamily member 5	CD40	IPI00018282	2	36	S	1
Sulfhydryl oxidase 2	QSOX2	IPI00376394	2	22	S	1
Isoform 5 of Disintegrin and metalloproteinase domain-containing protein 15	ADAM15	IPI00420067	2	20	S	1
HLA class I histocompatibility antigen, B-42 alpha chain	HLA-B	IPI00472676	2	40	S	1
Phospholipid transfer protein	PLTP	IPI00643034	2	37	S	0
Isoform 4 of CD276 antigen	CD276	IPI00719044	2	68	S	1
Beta-1,4-galactosyltransferase 3	B4GALT3	IPI00746600	2	10		1

Isoform 2 of Protocadherin-1	PCDH1	IPI00872579	2	15	S	1
Isoform 2 of Cytokine receptor common subunit beta	CSF2RB	IPI00879222	2	15	S	1
HLA class I histocompatibility antigen, Cw-4 alpha chain	MICB	IPI00892868	2	8	S	1
cDNA FLJ50175, highly similar to Tetraspanin-6	TSPAN6	IPI00909503	2	30	S	3
Autophagy-related protein 9A	ATG9A	IPI00917846	2	52		1
Isoform 4 of Low-density lipoprotein receptor	LDLR	IPI00983910	2	7	S	1
Glycosaminoglycan xylosylkinase	FAM20B	IPI00006657	2	10	S	1
Biglycan	BGN	IPI00010790	2	35	S	0
Lysosomal alpha-mannosidase	MAN2B1	IPI00012989	2	26	S	0
UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 1	B3GALNT1	IPI00021552	2	4	S	1
Tumor necrosis factor receptor superfamily member 10C precursor	TNFRSF10C	IPI00021970	2	7	S	2
Dipeptidyl peptidase 1	CTSC	IPI00022810	2	9	S	0
Growth/differentiation factor 15	GDF15	IPI00306543	2	20	S	0
TNFSF4 protein	TNFSF4	IPI00647536	2	25		0
Zinc transporter ZIP14	SLC39A14	IPI01013731	2	15	S	0
Low affinity cationic amino acid transporter 2	SLC7A2	IPI00003819	1	13		14
Sphingosine 1-phosphate receptor 1	S1PR1	IPI00015343	1	1		7
Solute carrier family 15 member 4	SLC15A4	IPI00465259	1	51		13
270 kDa protein	ABCA2	IPI00873151	1	2		13
Solute carrier family 22 member 4	SLC22A4	IPI00171334	1	1		11
Sodium- and chloride-dependent taurine transporter	SLC6A6	IPI00783565	1	1		12
ATP-binding cassette sub-family A member 8	ABCA8	IPI00952884	1	1	S	13
Sodium/hydrogen exchanger 7	SLC9A7	IPI00045928	1	19	S	12
Adenylate cyclase type 9	ADCY9	IPI00030099	1	6		8
Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1	IPI00220194	1	28		12
Solute carrier family 12 member 9	SLC12A9	IPI00387077	1	18		10
Isoform 3 of Latrophilin-2	LPHN2	IPI00410234	1	1	S	7
V-type proton ATPase 116 kDa subunit a isoform 2	ATP6V0A2	IPI00000425	1	1		6
Sugar phosphate exchanger 3	SLC37A3	IPI00301460	1	3	S	12
Isoform 2 of Proton-coupled amino acid transporter 4	SLC36A4	IPI00869176	1	2		9
cDNA FLJ34730 fis, clone MESAN2006580, highly similar to Chloride channel protein 5	CLCN5	IPI01010436	1	4	S	8
Basal cell adhesion molecule	BCAM	IPI00002406	1	1	S	1
Cysteine-rich motor neuron 1 protein	CRIM1	IPI00009294	1	1	S	1
cDNA FLJ58315, highly similar to Heme carrier protein 1	SLC46A1	IPI00909052	1	9		8
Neutral amino acid transporter	SLC1A5	IPI01010276	1	20	S	9
Sphingosine 1-phosphate receptor 3	S1PR3	IPI00015983	1	2		7
Insulin-like growth factor 1 receptor	IGF1R	IPI00027232	1	2	S	1
Plexin-A1	PLXNA1	IPI00552671	1	3	S	2
Activin receptor type-1	ACVR1	IPI00029219	1	5		1
Choline transporter-like protein 5	SLC44A5	IPI00168443	1	9	S	8
Zinc transporter ZIP6	SLC39A6	IPI00298702	1	15	S	6
Frizzled-4 precursor	FZD4	IPI00977659	1	18	S	7
G-protein coupled receptor 56	GPR56	IPI00412420	1	6	S	7
Latent-transforming growth factor beta-binding protein 1	LTBP1	IPI00784258	1	6	S	0
Interferon alpha/beta receptor 1	IFNAR1	IPI00012877	1	8	S	2
Lipid phosphate phosphohydrolase 3	PPAP2B	IPI00021453	1	10		6
Mucolipin-1	MCOLN1	IPI00452161	1	19		5
Isoform 4 of G-protein coupled receptor 126	GPR126	IPI00651770	1	9	S	8
Aquaporin-1	AQP1	IPI00024689	1	16	S	6
Lipid phosphate phosphohydrolase 2	PPAP2C	IPI00216620	1	6	S	6
Lymphatic vessel endothelial hyaluronic acid receptor 1	LYVE1	IPI00290856	1	9	S	1
Interleukin-18 receptor 1	IL18R1	IPI00021999	1	10	S	1
Leucine-rich repeat-containing protein 8A	LRRC8A	IPI00002070	1	13		4
Isoform 2 of Attractin	ATRN	IPI00162735	1	12		0
Isoform 2 of Receptor-type tyrosine-protein phosphatase gamma	PTPRG	IPI00796281	1	4		1
Heparan-alpha-glucosaminide N-acetyltransferase	HGSNAT	IPI00975914	1	14		1
FAS variant	FAS	IPI00985095	1	13	S	1
Neuropilin-2	NRP2	IPI00029693	1	14	S	1
Tetraspanin-13	TSPAN13	IPI00000735	1	13	S	4
Tetraspanin-15	TSPAN15	IPI00000736	1	14	S	4
Leucine-rich repeat transmembrane protein FLRT2	FLRT2	IPI00001633	1	1		1
Voltage-dependent calcium channel gamma-6 subunit	CACNG6	IPI00011072	1	1	S	4
CD82 antigen	CD82	IPI00020446	1	6	S	4
Bone morphogenetic protein receptor, type II (Serine/threonine kinase), isoform CRA_a	BMPR2	IPI00793420	1	15	S	1
Isoform 3 of Gamma-glutamyltransferase 5	GGT5	IPI00855794	1	9	S	1
Tetraspanin-9	TSPAN9	IPI00880170	1	1	S	3
Transmembrane 9 superfamily member 1	TM9SF1	IPI00981444	1	24	S	3
Interferon gamma receptor 1	IFNGR1	IPI00010808	1	16	S	1
Low-density lipoprotein receptor-related protein 10	LRP10	IPI00414231	1	16	S	1
Alpha-mannosidase 2	MAN2A1	IPI00003802	1	13		1
Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	MGAT4A	IPI00016743	1	26		1
Isoform 1 of Alpha-mannosidase 2x	MAN2A2	IPI00027703	1	3		1
Protein disulfide-isomerase TMX3	TMX3	IPI00064193	1	9		1
Protocadherin-16	DCHS1	IPI00064262	1	2	S	2
Isoform 2 of Protocadherin gamma-B7	PCDHGB7	IPI00100413	1	13	S	1
Isoform 2 of Signal peptide peptidase-like 2B	SPPL2B	IPI00304345	1	25	S	5
V-set and immunoglobulin domain-containing protein 10	VSIG10	IPI00386831	1	2	S	2
Neurogenic locus notch homolog protein 1 precursor	NOTCH1	IPI00412982	1	4	S	0
Dyslexia-associated protein KIAA0319-like protein	KIAA0319L	IPI00844309	1	5	S	1
cDNA FLJ56074, highly similar to 150 kDa oxygen-regulated protein (Orp150)	HYOU1	IPI00922838	1	3		2
Endoplasmic reticulum-Golgi intermediate compartment protein 2	ERGIC2	IPI01021557	1	5		0
IgG receptor FcRn large subunit p51	FCGRT	IPI00026646	1	30	S	1
Hyaluronidase-2	HYAL2	IPI00021531	1	35	S	0
EF-hand domain-containing protein KIAA0494	KIAA0494	IPI00006130	1	2		1
Ectonucleoside triphosphate diphosphohydrolase 7	ENTPD7	IPI00006476	1	5	S	2
Trophoblast glycoprotein	TPBG	IPI00009111	1	26	S	1
Endothelial protein C receptor precursor	PROCR	IPI00009276	1	31	S	1
Transmembrane protein 9B	TMEM9B	IPI00009880	1	14	S	2
Vesicular integral-membrane protein VIP36	LMAN2	IPI00009950	1	8	S	1
Beta-1,4-galactosyltransferase 5	B4GALT5	IPI00011656	1	1	S	1
Beta-galactoside alpha-2,6-sialyltransferase 1	ST6GAL1	IPI00013887	1	1	S	0
Transmembrane emp24 domain-containing protein 9	TMED9	IPI00023542	1	27	S	2
Myelin protein zero-like protein 2	MPZL2	IPI00024811	1	17	S	2
Membrane protein FAM174B	FAM174B	IPI00168255	1	23	S	1
Isoform 2 of Membrane cofactor protein	CD46	IPI00173903	1	4	S	1
Cell adhesion molecule 4	CADM4	IPI00176427	1	18	S	1
Isoform B of Protocadherin-7	PCDH7	IPI00215946	1	2	S	2
Isoform 2 of HLA class II histocompatibility antigen gamma chain	CD74	IPI00217775	1	7		1
Isoform 3 of Amyloid-like protein 2	APLP2	IPI00220978	1	26	S	2
Cadherin-2	CDH2	IPI00290085	1	1	S	1

Extracellular sulfatase Sulf-1	SULF1	IPI00293203	1	3	S	0
Transmembrane emp24 domain-containing protein 4	TMED4	IPI00296259	1	24	S	2
Protein HEG homolog 1	HEG1	IPI00297263	1	24	S	1
Endothelial cell-selective adhesion molecule	ESAM	IPI00303161	1	26	S	1
Osteopetrosis-associated transmembrane protein 1	OSTM1	IPI00329054	1	10	S	2
Isoform 2 of Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	PLOD2	IPI00337495	1	2	S	0
Isoform 2 of Discoidin, CUB and LCCL domain-containing protein 2	DCBLD2	IPI00433138	1	4	S	1
D-glucuronyl C5-epimerase	GLCE	IPI00433284	1	6	S	1
Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase B	MGAT4B	IPI00477905	1	7	S	1
Uncharacterized protein C1orf159	C1orf159	IPI00515131	1	5	S	1
HSD-43	LRRC8B	IPI00643826	1	1	S	2
fukutin isoform b	FKTN	IPI00645946	1	8		1
Inositol monophosphatase 3	IMPAD1	IPI00787853	1	25		1
Isoform 2 of Tyrosine-protein phosphatase non-receptor type substrate 1	SIRPA	IPI00848309	1	16	S	1
Isoform 2 of Transmembrane protein 248	TMEM248	IPI00852976	1	2		1
Isoform 2 of Nidogen-2	NID2	IPI00955436	1	1		0
Ceroid-lipofuscinosis neuronal protein 5	CLN5	IPI01011689	1	9		0
Beta-1,3-N-acetylglucosaminyltransferase radical fringe	RFNG	IPI01012125	1	7		0
Podocalyxin	PODXL	IPI01012180	1	23	S	1
Tetraspanin-8	TSPAN8	IPI01022643	1	1	S	2
cDNA FLJ59612, highly similar to Lactadherin	MFGE8	IPI00002236	1	1	S	0
Palmitoyl-protein thioesterase 1	PPT1	IPI00002412	1	2	S	0
Carbohydrate sulfotransferase 7	CHST7	IPI00008403	1	5	S	1
Tissue factor pathway inhibitor 2	TFPI2	IPI00009198	1	7	S	0
N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase	B3GNT1	IPI00009997	1	8	S	1
Cathepsin D	CTSD	IPI00011229	1	8	S	0
Ribonuclease pancreatic	RNASE1	IPI00014048	1	3	S	1
UPF0454 protein C12orf49	C12orf49	IPI00015479	1	5	S	1
Lysosomal protective protein	CTSA	IPI00021794	1	15	S	0
Beta-glucuronidase	GUSB	IPI00027745	1	1	S	1
Protein-tyrosine sulfotransferase 1	TPST1	IPI00030106	1	2	S	0
Metalloproteinase inhibitor 1	TIMP1	IPI00032292	1	27	S	0
Carbohydrate sulfotransferase 14	CHST14	IPI00044326	1	9	S	0
32 kDa protein	LOC285984	IPI00176692	1	3		0
A disintegrin and metalloproteinase with thrombospondin motifs 18	ADAMTS18	IPI00291532	1	3	S	0
Urokinase-type plasminogen activator	PLAU	IPI00296180	1	3	S	0
Chondroitin sulfate synthase 1	CHSY1	IPI00329141	1	2	S	1
cDNA FLJ53119, highly similar to ADP-ribosyl cyclase 1 (EC 3.2.2.5)	CD38	IPI00395006	1	16	S	1
Isoform 4 of Protocadherin gamma-C3	PCDHGC3	IPI00409668	1	2	S	0
Huntingtin interacting protein-1-related	NAGLU	IPI00555656	1	18		0
Deleted in autism-related protein 1	CXorf36	IPI00642861	1	1	S	0
Trem-like transcript 3 protein	TREML3P	IPI00742113	1	1		0
Isoform 3 of ATP-binding cassette sub-family A member 6	ABCA6	IPI00787668	1	1	S	1
Secreted glypican-1	GPC1	IPI00892905	1	5		0
Exostosin-1	EXT1	IPI00893515	1	7		0
cDNA FLJ56673, highly similar to Homo sapiens adipocyte-specific adhesion molecule (ASA	CLMP	IPI00908324	1	1	S	0
N-acetylglucosamine-6-sulfatase	GNS	IPI00908404	1	14	S	1
cDNA FLJ58558, highly similar to Tripeptidyl-peptidase 1 (EC 3.4.14.9)	TPP1	IPI00909516	1	69	S	0
cDNA FLJ50146	KIAA1467	IPI00909946	1	4		0
N-acetylgalactosaminidase, alpha-	NAGA	IPI00917374	1	4		0
Semaphorin-4C	SEMA4C	IPI00924493	1	1	S	0
Ecto-ADP-ribosyltransferase 4	ART4	IPI00939614	1	22	S	0
Claudin domain-containing protein 1	CLDND1	IPI00965136	1	17	S	1
Tetraspanin-4	TSPAN4	IPI00976403	1	10	S	1
N-acetylglucosamine-1-phosphotransferase subunit gamma	GNPTG	IPI00977899	1	6		0
CD59 glycoprotein	CD59	IPI00983417	1	823	S	0
Isoform 2 of Lymphocyte function-associated antigen 3	CD58	IPI00984332	1	3	S	2
SID1 transmembrane family member 2	SIDT2	IPI00985165	1	9		0
SPARC	SPARC	IPI01014139	1	13		0
Tumor necrosis factor receptor superfamily member 3	LTBR	IPI01014258	1	31		0
Killer cell lectin-like receptor subfamily G member 1	KLRG1	IPI01014981	1	2		0
Soluble calcium-activated nucleotidase 1	CANT1	IPI01018673	1	14	S	0
Sodium/potassium/calcium exchanger 6	SLC24A6	IPI01022026	1	11	S	0

Annexe 4 : Table des protéines variantes identifiées après digestion trypsique dans les deux réplicats biologiques des expériences d'enrichissement des glycoprotéines avec la phosphine biotine après stimulation TNF/IFN des HUVEC. (AC : numéro d'accession, MW : molecular weight, Score = score Mascot)

						Réplicat 1				Réplicat 2			
AC	MW	Gene Name	Description	Signal sécrétion prédit (ProteinCenter)	Segments transmembranaires prédis (TMHMM)	Number of quantified peptides	Score CT1 ; CT2 ; CT3 ; TNF1 ; TNF2 ; TNF3	Ratio PAI TNF/CT	p-value Student	Number of quantified peptides	Score CT1 ; CT2 ; CT3 ; TNF1 ; TNF2 ; TNF3	Ratio PAI TNF/CT	p-value Student
IPI00217775	26399	CD74	Isoform 2 of HLA class II histocompatibility antigen gamma chain	1		1	- ; - ; - ; 31 ; 61 ; 51	3,67	0,005746	2	- ; - ; - ; 23 ; 40 ; 45	#DIV/0!	0,000421
IPI00019771	42203	CX3CL1	Fractalkine	1	S	6	- ; - ; - ; 79 ; 52 ; 63	38,18	0,000895	2	- ; - ; - ; 166 ; 209 ; 96	#DIV/0!	0,000007
IPI00296542	66655	SELE	E-selectin	1	S	14	- ; - ; - ; 261 ; 129 ; 292	160,57	0,001486	3	- ; - ; - ; 84 ; 85 ; 38	#DIV/0!	0,002616
IPI00909039	33701	ICAM1	cDNA FLJ53671, highly similar to Intercellular adhesion molecule 1	1	S	8	- ; - ; - ; 277 ; 182 ; 269	118,10	0,000196	25	666 ; 700 ; 473 ; 4974 ; 5747 ; 5582	66,19	0,000385
IPI00012733	41493	CXCR7	C-X-C chemokine receptor type 7	7		2	64 ; - ; - ; - ; 34 ; 55	2,27	0,003330	2	33 ; - ; - ; 254 ; 404 ; 266	19,68	0,020927
IPI00003819	71673	SLC7A2	Low affinity cationic amino acid transporter 2	14		11	160 ; 120 ; 111 ; 296 ; 148 ; 288	11,00	0,000570	7	- ; - ; - ; 169 ; 257 ; 331	15,46	0,000225
IPI00010362	40058	HLA-E	HLA class I histocompatibility antigen, alpha chain E	1	S	11	- ; - ; - ; 189 ; 150 ; 156	10,51	0,004519	11	- ; - ; - ; 389 ; 530 ; 444	11,94	0,004262
IPI00023021	33275	CD274	Programmed cell death 1 ligand 1	1		4	- ; - ; - ; 50 ; 24 ; 57	3,86	0,000064	4	- ; - ; - ; 180 ; 122 ; 118	11,03	0,000297
IPI00293327	43369	P2RX4	P2X purinoceptor 4	2		7	- ; - ; 38 ; 82 ; 94 ; 46	5,64	0,000048	8	- ; 58 ; 33 ; 197 ; 286 ; 277	10,50	0,000705
IPI00887461	40479	HLA-B	MHC class I antigen	1	S	12	- ; - ; - ; 279 ; - ; -	14,34	0,007845	13	- ; - ; - ; 966 ; 1071 ; 1332	9,95	0,001096
IPI00219131	33349	ICOSLG	ICOS ligand	1	S	2	- ; - ; - ; 51 ; 61 ; 55	#DIV/0!	0,000054	2	58 ; 75 ; - ; 117 ; 148 ; 95	8,71	0,000699
IPI00892868	41466	MICB	HLA class I histocompatibility antigen, Cw-4 alpha chain	1	S	12	- ; - ; - ; 432 ; - ; -	11,02	0,000000	13	- ; - ; - ; - ; 757 ; 703	8,44	0,000725
IPI00550571	29235	C5orf15	Keratinocyte-associated transmembrane protein 2	1		2	33 ; - ; - ; 50 ; - ; 56	3,44	0,000813	2	- ; - ; - ; - ; - ; 43 ; 48	8,34	0,000614
IPI00743503	41055	HLA-A	HLA class I histocompatibility antigen, A-34 alpha chain	1	S	10	- ; - ; - ; - ; - ; 255	11,42	0,000251	14	- ; - ; - ; 768 ; 712 ; 816	7,01	0,001155
IPI00892768	34146	LOC100507	HLA class I histocompatibility antigen, A-3 alpha chain	1	S	12	- ; - ; - ; 327 ; - ; -	9,78	0,000479	15	- ; - ; - ; 1207 ; - ; -	7,01	0,001155
IPI00026241	19769	BST2	Bone marrow stromal antigen 2	1	S	7	80 ; 91 ; 123 ; 145 ; 67 ; 158	4,14	0,000669	5	133 ; 204 ; 75 ; 188 ; 239 ; 315	6,72	0,001509
IPI00018136	81276	VCAM1	Vascular cell adhesion protein 1	1		36	- ; - ; - ; 832 ; 575 ; 697	8,06	0,002100	39	- ; - ; - ; 2720 ; 2762 ; 2578	6,15	0,003895
IPI00894417	87176	TAP1	Antigen peptide transporter 1	7		13	- ; - ; - ; 281 ; 157 ; 200	22,26	0,000001	2	- ; - ; - ; 147 ; 153 ; 154	4,86	0,001145
IPI00936772	45333	GOLM1	Golgi membrane protein 1	1	S	17	111 ; 25 ; 63 ; 301 ; 157 ; 280	5,88	0,016152	12	95 ; 135 ; 217 ; 401 ; 652 ; 678	4,37	0,000776
IPI00843944	90719	PTPRA	Isoform 3 of Receptor-type tyrosine-protein phosphatase alpha	2	S	1	- ; - ; - ; - ; 44 ; -	2,46	0,006615	8	- ; - ; - ; 327 ; 227 ; 196	3,98	0,000535
IPI00872082	33192	CD47	Isoform OA3-305 of Leukocyte surface antigen CD47	6	S	3	10 ; - ; - ; - ; - ; 39	2,23	0,006660	3	- ; 26 ; - ; 61 ; 77 ; 49	3,98	0,000516
IPI00894466	44841	ANTXR2	ANTXR2 protein	1	S	4	- ; - ; - ; 62 ; - ; 34	2,55	0,008033	5	- ; 55 ; - ; 73 ; 96 ; 180	3,88	0,000483
IPI00000070	95376	LDLR	Low-density lipoprotein receptor	1	S	13	- ; - ; - ; 234 ; 63 ; 141	9,07	0,000014	2	- ; 43 ; - ; 37 ; 67 ; 69	3,78	0,001463
IPI00909245	39043	CANT1	cDNA FLJ53010, highly similar to Soluble calcium-activated nucleotidase 1	0		4	- ; 37 ; - ; 89 ; - ; -	2,76	0,000755	8	- ; 58 ; - ; 102 ; 239 ; 274	3,30	0,031914
IPI00914842	102243	NRP2	Isoform A0 of Neuropilin-2	1	S	9	- ; - ; - ; 125 ; 80 ; 112	9,38	0,000008	13	190 ; 221 ; 116 ; 335 ; 267 ; 342	3,24	0,000007
IPI00375688	76744	NCSTN	Isoform 2 of Nicastrin	1		11	108 ; 132 ; 86 ; 213 ; 132 ; 118	2,64	0,002102	8	244 ; 223 ; 299 ; 304 ; 406 ; 275	3,16	0,000007
IPI00002526	111824	DSE	Dermatan-sulfate epimerase	3		1	- ; - ; - ; 51 ; - ; -	3,05	0,000346	2	44 ; 44 ; 190 ; 172 ; 231 ; 119	3,07	0,000589
IPI00152505	56320	SLC2A14	Solute carrier family 2, facilitated glucose transporter member 14	10		1	- ; - ; - ; - ; - ; 37	12,39	0,016924	3	50 ; 106 ; 58 ; 84 ; 119 ; 172	2,92	0,006029
IPI00007118	45060	SERPINE1	Plasminogen activator inhibitor 1	0	S	18	312 ; 280 ; 285 ; 355 ; 311 ; 428	2,41	0,001887	16	652 ; 540 ; 677 ; 888 ; 1206 ; 939	2,84	0,000114
IPI00295542	53879	NUCB1	Nucleobindin-1	0	S	23	201 ; 184 ; 255 ; 459 ; 304 ; 325	3,62	0,005715	25	1604 ; 1760 ; 1485 ; 2124 ; 1818 ; 1782	2,73	0,000226
IPI00658202	97040	CDH2	Cadherin-2	1		14	118 ; 72 ; 63 ; 293 ; 271 ; 150	2,98	0,000007	2	- ; - ; - ; 58 ; 37 ; 124	2,60	0,001015
IPI00383014	43333	CASC4	Isoform 2 of Protein CASC4	1	S	12	126 ; 142 ; 179 ; 281 ; 143 ; 178	2,84	0,000000	8	122 ; 135 ; 106 ; 396 ; 190 ; 300	2,53	0,001562
IPI00760547	28981	MPZL1	Isoform 2 of Myelin protein zero-like protein 1	2	S	4	- ; - ; - ; 63 ; 30 ; 40	2,54	0,002677	6	231 ; 208 ; 338 ; 459 ; 551 ; 326	2,52	0,004337
IPI00023542	27277	TMED9	Transmembrane emp24 domain-containing protein 9	2	S	9	103 ; 129 ; 118 ; 160 ; 110 ; 95	2,14	0,000012	5	151 ; 177 ; 106 ; 568 ; 581 ; 536	2,47	0,002972
IPI00942795	44229	ANO6	ANO6 44 kDa protein	8		2	32 ; 31 ; - ; 32 ; - ; 36	2,21	0,026391	11	132 ; 111 ; 68 ; 132 ; 155 ; 126	2,40	0,000025
IPI00165651	43494	ERGIC2	Cd002 protein	0		5	36 ; 35 ; 38 ; 80 ; 88 ; 104	2,59	0,000506	4	87 ; 90 ; 128 ; 87 ; 147 ; 143	2,40	0,000474
IPI00941900	37107	CALU	Calumenin	0	S	12	90 ; 109 ; 71 ; 190 ; 158 ; 162	3,24	0,033108	5	58 ; 26 ; - ; 35 ; 63 ; 43	2,33	0,004923
IPI00783698	63430	TMEM87A	Transmembrane protein 87A	7	S	11	91 ; 76 ; 108 ; 195 ; 120 ; 131	2,01	0,000046	12	397 ; 350 ; 259 ; 332 ; 258 ; 188	2,29	0,000101
IPI00470361	85049	KIRREL	Isoform 2 of Kin of IRRE-like protein 1	1	S	2	- ; - ; - ; 74 ; 51 ; 33	2,06	0,029487	6	89 ; - ; 47 ; 122 ; 210 ; 350	2,26	0,001015
IPI00917006	43062	RNF149	E3 ubiquitin-protein ligase RNF149	1	S	3	- ; - ; - ; 19 ; 38 ; 45	2,08	0,000570	5	38 ; 40 ; - ; 202 ; 138 ; 83	2,26	0,021426
IPI00009950	40229	LMAN2	Vesicular integral-membrane protein VIP36	1		7	50 ; 41 ; 57 ; 71 ; 77 ; 101	2,66	0,000350	11	148 ; 252 ; 115 ; 504 ; 444 ; 435	2,22	0,182831
IPI00030847	67888	TM9SF3	Transmembrane 9 superfamily member 3	9	S	15	237 ; 214 ; 165 ; 317 ; 110 ; 240	2,26	0,006190	19	244 ; 450 ; 408 ; 695 ; 831 ; 858	2,18	0,000224
IPI00297124	103537	IL6ST	Interleukin-6 receptor subunit beta	1	S	16	- ; - ; - ; 229 ; 95 ; 173	5,12	0,000001	12	150 ; 187 ; 235 ; 360 ; 295 ; 293	2,15	0,002920
IPI00217766	54290	SCARB2	Lysosome membrane protein 2	2	S	15	125 ; 197 ; 227 ; 267 ; 225 ; 241	2,41	0,000702	12	468 ; 413 ; 298 ; 741 ; 828 ; 643	2,13	0,001905
IPI00220303	130539	MAN2A2	Alpha-mannosidase 2x	0		10	- ; 67 ; 64 ; 166 ; 100 ; 166	2,41	0,009582	7	101 ; 125 ; 171 ; 167 ; 102 ; 126	2,12	0,005667
IPI00743100	69620	MAN1A2	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IB	1	S	13	77 ; 81 ; 10 ; 138 ; 139 ; 125	2,21	0,000003	9	136 ; 151 ; 113 ; 280 ; 289 ; 174	2,11	0,007701
IPI00026991	71159	GALNT6	Polypeptide N-acetylgalactosaminyltransferase 6	1	S	7	- ; 64 ; 61 ; 55 ; - ; 43	3,03	0,000588	11	270 ; 340 ; 298 ; 449 ; 303 ; 376	2,04	0,001134
IPI00026824	41669	HMOX2	Heme oxygenase 2	1		6	38 ; - ; 43 ; 67 ; 68 ; 26	2,75	0,000071	7	64 ; - ; 59 ; 219 ; 143 ; 198	2,04	0,009626

Annexe 5 : Table des protéines variantes identifiées après analyse quantitative du protéome total des HUVEC suite à la stimulation TNFα-IFNγ Surlignées en bleu : protéines identifiées variantes à la fois après stimulation TNFα-IFNγ, IL-1β et IL-33 ; surlignées en jaune : protéines variantes après stimulation TNFα-IFNγ et IL-33. (AC : numéro d'accession, ID : identifiant de la protéine, MW : molecular weight)

Protein AC	Protein ID	MW	Protein Name	Gene Name	Number quantified peptides	Coverage (%)	Mascot score TNF1; TNF2; TNF3; CT1; CT2; CT3	PAI TNF/IFN 1	PAI TNF/IFN 2	PAI TNF/IFN 3	PAI CT 1	PAI CT 2	PAI CT 3	Ratio TNF-IFN/CT	Student
P14902	I23O1_HUMAN	45326	Indoleamine 2,3-dioxygenase 1	IDO1	23	49,88	459 ; 554 ; 560 ; - ; - ; -	2,42E+08	2,50E+08	2,08E+08	1,00E+05	1,00E+05	1,00E+05	2332,89	0,00299
O15205	UBD_HUMAN	18473	Ubiquitin D	UBD	11	66,67	- ; 398 ; - ; - ; - ; -	1,91E+08	2,15E+08	2,08E+08	1,00E+05	1,00E+05	1,00E+05	2045,60	0,00116
P16581	LYAM2_HUMAN	66629	E-selectin	SELE	5	6,72	180 ; 102 ; 135 ; - ; - ; -	9,77E+07	1,12E+08	9,96E+07	1,00E+05	1,00E+05	1,00E+05	1029,60	0,00179
P78423	X3CL1_HUMAN	42203	Fractalkine	CX3CL1	5	10,83	52 ; 92 ; 78 ; - ; - ; -	7,11E+07	7,38E+07	7,14E+07	1,00E+05	1,00E+05	1,00E+05	720,94	0,00015
P20592	MX2_HUMAN	82089	Interferon-induced GTP-binding protein Mx2	MX2	18	24,06	309 ; 367 ; 377 ; - ; - ; -	3,59E+08	2,47E+08	3,02E+08	1,01E+06			298,19	0,01141
P21580	TNAP3_HUMAN	61248	Tumor necrosis factor alpha-induced protein 3	TNFAIP3	8	3,57	38 ; 55 ; 76 ; - ; - ; -	2,87E+07	3,12E+07	2,79E+07	1,00E+05	1,00E+05	1,00E+05	292,63	0,00112
P02778	CXL10_HUMAN	10881	C-X-C motif chemokine 10	CXCL10	8	47,96	140 ; 155 ; 175 ; - ; - ; -	4,64E+08	7,01E+08	6,85E+08			2,93E+06	210,79	0,01524
P19971	TYPH_HUMAN	46087	Thymidine phosphorylase	TYMP	8	19,55	135 ; 177 ; 190 ; - ; - ; -	1,78E+07	2,46E+07	1,88E+07	1,00E+05	1,00E+05	1,00E+05	204,19	0,01058
P01579	IFNG_HUMAN	15576	Interferon gamma	IFNG	12	70,68	346 ; 248 ; 313 ; - ; - ; -	7,71E+07	9,35E+07	1,27E+08	7,35E+05		2,98E+05	192,05	0,02146
P09603	CSF1_HUMAN	28004	Macrophage colony-stimulating factor 1	CSF1	1	4,94	39 ; - ; - ; - ; - ; -	1,38E+07	1,52E+07	1,55E+07	1,00E+05	1,00E+05	1,00E+05	148,47	0,00121
P19320	VCAM1_HUMAN	81264	Vascular cell adhesion protein 1	VCAM1	38	55,89	930 ; 932 ; 982 ; - ; - ; -	5,98E+08	6,00E+08	5,91E+08			4,04E+06	147,59	5,1E-07
O75144	ICOSL_HUMAN	33349	ICOS ligand	ICOSLG	1	3,31	57 ; 61 ; 73 ; - ; - ; -	1,36E+07	1,23E+07	1,13E+07	1,00E+05	1,00E+05	1,00E+05	124,09	0,00303
Q9NZQ7	PD1L1_HUMAN	20246	Programmed cell death 1 ligand 1	CD274	1	9,09	- ; 62 ; - ; - ; - ; -	1,44E+07	1,36E+07	6,66E+06	1,00E+05	1,00E+05	1,00E+05	115,74	0,04328
Q01201	RELB_HUMAN	62134	Transcription factor RelB	RELB	1	2,94	42 ; 59 ; 59 ; - ; - ; -	1,26E+07	9,47E+06	1,12E+07	1,00E+05	1,00E+05	1,00E+05	110,77	0,00669
P05362	ICAM1_HUMAN	57880	Intercellular adhesion molecule 1	ICAM1	25	9,59	848 ; 841 ; 887 ; 92 ; 104 ; 116	2,31E+09	2,28E+09	2,32E+09	2,43E+07	2,49E+07	2,08E+07	98,82	2,5E-05
P25774	CATS_HUMAN	31626	Cathepsin S	CTSS	5	20,64	128 ; 125 ; 138 ; - ; - ; -	4,18E+07	4,06E+07	4,12E+07	4,24E+05			97,09	7,5E-05
Q53G44	IF44L_HUMAN	51322	Interferon-induced protein 44-like	IFI44L	4	7,08	40 ; 58 ; - ; - ; - ; -	1,00E+07	7,73E+06	7,94E+06	1,00E+05	1,00E+05	1,00E+05	85,68	0,00751
Q96AZ6	ISG20_HUMAN	20363	Interferon-stimulated gene 20 kDa protein	ISG20	12	4,42	522 ; 551 ; 478 ; - ; - ; 40	7,06E+07	7,59E+07	7,40E+07	1,05E+06	7,77E+05	9,39E+05	79,86	0,00044
P25106	CXCR7_HUMAN	41493	C-X-C chemokine receptor type 7	CXCR7	1	3,87	- ; 42 ; - ; - ; - ; -	8,70E+06	5,24E+06	7,00E+06	1,00E+05	1,00E+05	1,00E+05	69,80	0,02035
O14625	CXL11_HUMAN	10365	C-X-C motif chemokine 11	CXCL11	2	22,34	- ; 50 ; 53 ; - ; - ; -	2,42E+08	2,81E+08	3,05E+08			3,99E+06	69,16	0,00447
P10914	IRF1_HUMAN	36502	Interferon regulatory factor 1	IRF1	2	4,62	93 ; 55 ; - ; - ; - ; -	6,42E+06	7,39E+06	5,93E+06	1,00E+05	1,00E+05	1,00E+05	65,81	0,00435
P18847	ATF3_HUMAN	20576	Cyclic AMP-dependent transcription factor ATF-3	ATF3	1	7,18	43 ; 41 ; - ; - ; - ; -	5,74E+06	6,02E+06	5,63E+06	1,00E+05	1,00E+05	1,00E+05	57,95	0,0004
Q03169	TNAP2_HUMAN	72661	Tumor necrosis factor alpha-induced protein 2	TNFAIP2	12	16,06	151 ; 103 ; 179 ; - ; - ; -	1,88E+08	2,15E+08	1,98E+08	2,97E+06	5,81E+06	1,88E+06	56,40	0,00134
O00478	BT3A3_HUMAN	31789	Butyrophilin subfamily 3 member A3	BTN3A3	1	5,71	- ; - ; 36 ; - ; - ; -	5,73E+06	5,61E+06	5,31E+06	1,00E+05	1,00E+05	1,00E+05	55,52	0,00052
Q8NI08	NCOA7_HUMAN	96746	Nuclear receptor coactivator 7	NCOA7	3	1,51	55 ; 33 ; 106 ; - ; - ; -	4,54E+06	5,11E+06	5,93E+06	1,00E+05	1,00E+05	1,00E+05	51,94	0,00624
P20591	MX1_HUMAN	75520	Interferon-induced GTP-binding protein Mx1	MX1	47	75,53	985 ; 910 ; 1085 ; - ; - ; -	5,15E+08	5,59E+08	6,08E+08	9,23E+06	1,23E+07	1,20E+07	50,16	0,00237
O95236	APOL3_HUMAN	44278	Apolipoprotein L3	APOL3	17	37,56	258 ; 374 ; 323 ; - ; - ; -	7,80E+07	7,84E+07	8,92E+07	1,93E+06	1,81E+06	1,36E+06	48,14	0,00203
O14791	APOL1_HUMAN	47474	Apolipoprotein L1	APOL1	3	5,12	54 ; 114 ; 66 ; - ; - ; -	2,04E+07	2,41E+07	2,21E+07	4,74E+05		4,96E+05	45,85	0,00241
Q14213	IL27B_HUMAN	25396	Interleukin-27 subunit beta	EBI3	1	5,24	43 ; 51 ; 53 ; - ; - ; -	3,78E+06	4,67E+06	5,08E+06	1,00E+05	1,00E+05	1,00E+05	45,10	0,00742
Q8WVN6	SCTM1_HUMAN	27012	Secreted and transmembrane protein 1	SECTM1	8	24,19	228 ; 127 ; 282 ; - ; - ; -	8,40E+07	8,71E+07	8,03E+07	3,62E+06	7,30E+05		38,57	5,9E-05
O14879	IFIT3_HUMAN	55985	Interferon-induced protein with tetratricopeptide repeats 3	IFIT3	28	71,43	636 ; 708 ; 813 ; - ; - ; -	1,46E+08	1,76E+08	1,42E+08	3,12E+06	7,03E+06	4,51E+06	31,70	0,00473
Q96PP8	GBP5_HUMAN	66617	Guanylate-binding protein 5	GBP5	32	49,32	688 ; 581 ; 560 ; - ; - ; -	2,25E+08	2,73E+08	2,79E+08	8,98E+06	8,36E+06	7,30E+06	31,56	0,00452
Q96EH8	NEUL3_HUMAN	28789	E3 ubiquitin-protein ligase NEURL3	NEURL3	2	8,78	- ; - ; 37 ; - ; - ; -	2,39E+06	3,37E+06	3,48E+06	1,00E+05	1,00E+05	1,00E+05	30,82	0,01332
P50591	TNF10_HUMAN	32509	Tumor necrosis factor ligand superfamily member 10	TNFSF10	8	29,18	116 ; 162 ; 194 ; - ; - ; -	1,58E+07	1,47E+07	1,64E+07	8,29E+05	7,86E+05	4,40E+05	22,85	0,00064
P17676	CEBPB_HUMAN	36106	CCAAT/enhancer-binding protein beta	CEBPB	1	5,22	- ; 50 ; - ; - ; - ; -	2,56E+06	2,28E+06	1,77E+06	1,00E+05	1,00E+05	1,00E+05	22,02	0,0121
P32456	GBP2_HUMAN	67209	Interferon-induced guanylate-binding protein 2	GBP2	40	57,19	1035 ; 962 ; 987 ; 137 ; 99 ; 91	7,69E+08	8,68E+08	9,10E+08	4,17E+07	3,82E+07	3,65E+07	21,88	0,00264
Q96S86	HPLN3_HUMAN	40924	Hyaluronan and proteoglycan link protein 3	HAPLN3	4	12,78	91 ; 80 ; 75 ; - ; - ; -	1,61E+07	1,70E+07	1,71E+07	5,99E+05	1,03E+06		20,45	3,3E-05
Q9UGS9	Q9UGS9_HUMAN	8482	Putative uncharacterized protein DKFZp564M2422	SOD2	5	67,11	- ; - ; 141 ; - ; - ; -	2,69E+08	3,03E+08	3,50E+08	1,51E+07	1,54E+07	1,48E+07	20,31	0,00638
Q96CG3	TIFA_HUMAN	21445	TRAF-interacting protein with FHA domain-containing protein A	TIFA	9	30,98	96 ; 120 ; 115 ; - ; - ; -	1,83E+07	1,83E+07	2,01E+07	1,08E+06	8,20E+05	9,30E+05	19,38	0,00097
P24001	IL32_HUMAN	21737	Interleukin-32	IL32	3	13,3	46 ; 58 ; 53 ; - ; - ; -	1,75E+07	1,29E+07	1,96E+07	3,21E+05	9,77E+05	1,31E+06	19,19	0,01392
P04179	SODM_HUMAN	23673	Superoxide dismutase [Mn], mitochondrial	SOD2	17	59,15	400 ; 444 ; 439 ; 179 ; 237 ; 249	1,17E+09	1,25E+09	1,25E+09	7,02E+07	6,01E+07	6,10E+07	19,18	0,00043
P19883	FST_HUMAN	38007	Follistatin	FST	1	4,94	91 ; 69 ; 77 ; - ; - ; -	1,36E+07	1,61E+07	1,58E+07		8,03E+05		18,87	0,00294
P30793	GCH1_HUMAN	27903	GTP cyclohydrolase 1	GCH1	11	46	123 ; 117 ; 203 ; - ; - ; -	2,16E+07	2,37E+07	2,30E+07	1,15E+06	1,49E+06	1,03E+06	18,64	0,00052
Q9Y6K5	OAS3_HUMAN	121239	2'-5'-oligoadenylate synthase 3	OAS3	11	6,99	127 ; 111 ; 117 ; - ; - ; -	2,63E+07	2,60E+07	2,96E+07		1,53E+06		17,84	0,00196
Q05DN2	Q05DN2_HUMAN	48070	IFIT2 protein	IFIT2	11	30,1	208 ; 250 ; 226 ; - ; - ; -	3,22E+07	3,57E+07	3,12E+07		5,61E+05	3,23E+06	17,43	0,00092
O60645	EXOC3_HUMAN	85567	Exocyst complex component 3	EXOC3	2	2,28	32 ; - ; 40 ; 45 ; 45 ; 49	2,10E+08	1,53E+08	1,91E+08	1,13E+07	1,10E+07	9,49E+06	17,43	0,00926
Q96PP9	GBP4_HUMAN	73124	Guanylate-binding protein 4	GBP4	42	62,19	1026 ; - ; 1086 ; 188 ; 72 ; 248	4,04E+08	4,46E+08	4,36E+08	2,68E+07	2,46E+07	2,53E+07	16,77	0,00095
O95786	DDX58_HUMAN	98058	Probable ATP-dependent RNA helicase DDX58	DDX58	38	2,46	621 ; 630 ; 682 ; - ; - ; 31	1,24E+08	1,31E+08	1,36E+08	6,78E+06	8,65E+06	8,80E+06	16,16	0,00063
Q96J88	ESIP1_HUMAN	36793	Epithelial-stromal interaction protein 1	EPSTI1	12	33,02	140 ; 133 ; 126 ; - ; - ; -	3,27E+07	3,19E+07	3,14E+07	2,08E+06			15,39	0,00016
P23381	SYWC_HUMAN	53165	Tryptophanyl-tRNA synthetase, cytoplasmic	WARS	46	58,81	1820 ; 1633 ; 1738 ; 667 ; 675 ; 77	1,55E+09	1,58E+09	1,73E+09	1,07E+08	1,02E+08	1,10E+08	15,27	0,00135
Q07325	CXCL9_HUMAN	13920	C-X-C motif chemokine 9	CXCL9	4	35,2	96 ; 173 ; 138 ; - ; - ; -	2,74E+08	2,49E+08	3,45E+08	2,84E+07	2,09E+07	1,17E+07	14,23	0,00977
Q9NSY2	STAR5_HUMAN	23794	StAR-related lipid transfer protein 5	STARD5	1	4,23	- ; - ; 36 ; - ; - ; -	1,25E+06	1,03E+06	1,62E+06	1,00E+05	1,00E+05	1,00E+05	13,00	0,02045
Q9UK39	NOCT_HUMAN	48196	Nocturnin	CCRN4L	3	3,25	42 ; 103 ; 54 ; - ; - ; -	2,36E+07	2,34E+07	1,94E+07	1,15E+06	3,22E+06	1,30E+06	11,71	0,00106
Q13077	TRAF1_HUMAN	46164	TNF receptor-associated factor 1	TRAF1	9	22,12	175 ; 174 ; 232 ; - ; - ; -	8,95E+07	7,87E+07	8,67E+07			7,32E+06	11,61	0,00174
O14545	TRAD1_HUMAN	64841	TRAF-type zinc finger domain-containing protein 1	TRAFD1	8	7,22	104 ; 120 ; 96 ; - ; - ; -	2,58E+07	2,24E+07	2,32E+07	2,70E+06		1,89E+06	10,37	0,00083
P09914	IFIT1_HUMAN	55360	Interferon-induced protein with tetratricopeptide repeats 1	IFIT1	20	38,49	352 ; 383 ; 361 ; - ; - ; -	4,26E+07	4,81E+07	4,25E+07		2,38E+06	6,40E+06	10,10	0,0017
P32455	GBP1_HUMAN	67917	Interferon-induced guanylate-binding protein 1	GBP1	59	69,59	1580 ; 1430 ; 1579 ; - ; - ; -	1,72E+09	2,03E+09	2,20E+09	1,98E+08	2,00E+08	1,97E+08	10,00	0,00614
P17275	JUNB_HUMAN	35879	Transcription factor jun-B	JUNB	2	7,49	45 ; 40 ; - ; - ; - ; -	1,20E+07	1,43E+07	1,46E+07		1,43E+06	1,50E+06	9,29	0,00459
Q9BQE5	APOL2_HUMAN	48885	Apolipoprotein L2	APOL2	17	34,74	477 ; 469 ; 540 ; 175 ; 160 ; 141	3,46E+08	3,67E+08	3,64E+08	4,57E+07	4,22E+07	3,90E+07	8,49	0,00017
Q92982	NINJ1_HUMAN	16389	Ninjurin-1	NINJ1	2	13,16	46 ; 54 ; 56 ; - ; - ; -	7,58E+06	7,00E+06	7,12E+06	1,04E+06	6,24E+05	9,01E+05	8,46	2,2E-05
P31431	SDC4_HUMAN	21642	Syndecan-4	SDC4	4	17,68	93 ; 99 ; 83 ; - ; - ; -	3,24E+07	3,10E+07	3,76E+07	3,68E+06		4,31E+06	8,43	0,00386
P05161	ISG15_HUMAN	17888	Ubiquitin-like protein ISG15	ISG15	11	41,82	422 ; 404 ; 367 ; 161 ; 255 ; 182	4,10E+08	5,19E+08	4,30E+08	5,91E+07	3,87E+07	6,43E+07	8,38	0,00492
P78504	JAG1_HUMAN	133799	Protein jagged-1	JAG1	1	0,9	- ; 46 ; - ; - ; - ; -	1,74E+07	1,81E+07	1,37E+07	1,90E+06		2,09E+06	8,22	0,0085
P18463	1B37_HUMAN	41466	HLA class I histocompatibility antigen, B-37 alpha chain	HLA-B	12	43,28	- ; - ; 376 ; - ; - ; -	7,06E+08	6,73E+08	7,38E+08	7,25E+07	8,70E+07	9,85E+07	8,21	0,00019
O43353	RIPK2_HUMAN	61066	Receptor-interacting serine/threonine-protein kinase 2	RIPK2	16	31,3	353 ; 239 ; 396 ; - ; - ; -	6,48E+07	6,95E+07	5,73E+07	7,77E+06	8,65E+06	7,75E+06	7,93	0,00387
Q96RQ9	OXLA_HUMAN	65328	L-amino-acid oxidase	IL4I1	3	4,24	70 ; 86 ; 61 ; - ; - ; -	1,28E+07	1,30E+07	1,37E+07		1,62E+06	1,79E+06	7,71	0,00025
Q14392	LRC32_HUMAN	71979	Leucine-rich repeat-containing protein 32	LRRC32	2	5,89	37 ; 47 ; 89 ; - ; - ; -	5,90E+06	4,34E+06	5,15E+06	6,43E+05	6,72E+05	6,87E+05	7,68	0,00994
O15533	TPSN_HUMAN	47626	Tapasin	TAPBP	9	27,01	189 ; 201 ; 127 ; - ; - ; -	1,14E+08	9,96E+07	9,87E+07	1,59E+07	1,39E+07	1,23E+07	7,42	0,00209
P52630	STAT2_HUMAN	97916	Signal transducer and activator of transcription 2	STAT2	20	5,76	351 ; 359 ; 366 ; 92 ; 75 ; 90	1,13E+08	1,11E+08	1,21E+08	1,51E+07	1,47E+07	1,74E+07	7,32	0,00038
Q6P9H4	CNKR3_HUMAN	52644	Connector enhancer of kinase suppressor of ras 3	CNKSR3	2	5,47	64 ; 66 ; - ; - ; - ; -	3,18E+06	3,13E+06	3,55E+06	4,51E+05			7,30	0,00212
Q9BYX4	IFIH1_HUMAN	116689	Interferon-induced helicase C domain-containing protein 1	IFIH1	7	6,24	163 ; 184 ; 110 ; - ; - ; -	9,77E+07	9,97E+07	1,03E+08	1,54E+07	1,33E+07	1,25E+07	7,28	1,1E-05
Q96C10	DHX58_HUMAN	65250	Probable ATP-dependent RNA helicase DHX58	DHX58	7	15,4	99 ; 109 ; 167 ; - ; - ; -	1,62E+07	1,63E+07	1,81E+07		2,32E+06		7,27	0,00187
Q8IXQ6	PARP9_HUMAN	96223	Poly [ADP-ribose] polymerase 9	PARP9	3	5,15	59 ; 87 ; 63 ; - ; - ; -	8,05E+06	7,85E+06	7,87E+06	1,15E+06	9,34E+05	1,22E+06	7,19	9,9E-07
O60826	CCD22_HUMAN	65156	Coiled-coil domain-containing protein 22	CCDC22	12	8,62	164 ; 284 ; 170 ; 189 ; 131 ; 95	9,58E+07	7,44E+07	1,33E+08	1,44E+07	1,43E+07	1,40E+07	7,10	0,03706
O14933	UB2L6_HUMAN	17769	Ubiquitin/ISG15-conjugating enzyme E2 L6	UBE2L6	9	40,52	150 ; 140 ; 89 ; 39 ; 43 ; 41	3,87E+07	3,23E+07	4,10E+07	5,64E+06	5,98E+06	4,80E+06	6,82	0,0059
Q00653	NFKB2_HUMAN	96749	Nuclear factor NF-kappa-B p100 subunit	NFKB2	16	15,56	231 ; 316 ; 227 ; - ; - ; -	4,88E+07	4,95E+07	3,82E+07	5,30E+06	6,85E+06	8,31E+06	6,67	0,00645
Q460N5	PAR14_HUMAN	202800	Poly [ADP-ribose] polymerase 14	PARP14	42	27,21	849 ; 658 ; 920 ; - ; - ; -	8,85E+07	8,30E+07	9,26E+07	1,81E+07	1,44E+07	7,51E+06	6,61	6,2E-05
Q96CG8	CTHR1_HUMAN	26224	Collagen triple helix repeat-containing protein 1	CTHRC1	5	22,63	141 ; 97 ; 120 ; - ; - ; -	2,67E+07	2,79E+07	2,68E+07	4,76E+06	3,26E+06	4,41E+06	6,56	3,6E-06
P80217	IN35_HUMAN	31546	Interferon-induced 35 kDa protein	IFI35	13	44,76	361 ; 302 ; 270 ; 53 ; 139 ; 132	6,32E+07	6,47E+07	7,27E+07	1,05E+07	1,10E+07	9,51E+06	6,47	0,00227
P40306	PSB10_HUMAN	28936	Proteasome subunit beta type-10	PSMB10	8	45,42	85 ; 174 ; 87 ; - ; - ; -	3,38E+07	3,99E+07	3,84E+07	6,53E+06	4,37E+06	6,72E+06	6,36	0,00109
O15162	PLS1_HUMAN	35049	Phospholipid scramblase 1	PLSCR1	6	21,38	123 ; 107 ; 130 ; - ; - ; -	4,29E+07	4,02E+07	4,42E+07	7,08E+06	6,14E+06	7,07E+06	6,28	0,00057
A6NLJ0	C2C4B_HUMAN	38769	C2 calcium-dependent domain-containing protein 4B	C2CD4B	2	3,02	66 ; 55 ; 62 ; - ; - ; -	6,83E+06	7,93E+06	8,56E+06	1,24E+06			6,26	0,00591
P00973	OAS1_HUMAN	52166	2'-5'-oligoadenylate synthase 1	OAS1	8	19,61	163 ; 213 ; 187 ; - ; - ; -	5,70E+07	6,34E+07	6,36E+07	1,49E+07	1,11E+07	4,01E+06	6,13	0,00039
Q9UMW8	UBP18_HUMAN	43011	Ubl carboxyl-terminal hydrolase 18	USP18	7	16,13	77 ; 70 ; 85 ; - ; - ; -	2,38E+07	2,18E+07	2,24E+07	3,48E+06	4,14E+06	3,67E+06	6,02	0,00034
P61769	B2MG_HUMAN	13946	Beta-2-microglobulin	B2M	6	41,8	143 ; 168 ; 152 ; 123 ; 118 ; 137	4,82E+08	6,21E+08	4,69E+08	9,77E+07	6,76E+07	1,10E+08	5,70	0,00891
O14495	LPP3_HUMAN	35119	Lipid phosphate phosphohydrolase 3	PPAP2B	2	7,72	58 ; - ; - ; - ; - ; -	8,16E+06	8,48E+06	9,08E+06	1,44E+06	2,00E+06	1,15E+06	5,61	4,5E-05
Q03518	TAP1_HUMAN	87218	Antigen peptide transporter 1	TAP1	22	40,84	536 ; 622 ; 518 ; - ; 49 ; -	2,41E+08	2,54E+08	2,67E+08	6,27E+07	6,78E+07	7,22E+06	5,53	0,00389
P29728	OAS2_HUMAN	82431	2'-5'-oligoadenylate synthase 2	OAS2	24	38,53	347 ; 605 ; 467 ; - ; - ; -	4,21E+07	6,03E+07	5,62E+07	8,65E+06	9,64E+06	1,04E+07	5,53	0,01533
P35354	PGH2_HUMAN	68996	Prostaglandin G/H synthase 2	PTGS2	14	23,68	200 ; 166 ; 238 ; 45 ; 63 ; 65	2,31E+07	2,39E+07	2,36E+07	4,17E+06	2,97E+06	5,83E+06	5,45	0,00089
Q8TCB0	IFI44_HUMAN	50491	Interferon-induced protein 44	IFI44	3	1,8	51 ; 55 ; 47 ; - ; - ; -	1,48E+07	1,52E+07	1,46E+07	1,14E+06	6,40E+06	7,47E+05	5,38	0,02091
P28838	AMPL_HUMAN	56166	Cytosol aminopeptidase	LAP3	39	59,34	1125 ; 1136 ; 1166 ; 639 ; 653 ; 60	5,35E+08	6,18E+08	5,54E+08	1,03E+08	1,17E+08	1,02E+08	5,31	0,00222
Q9P2E3	ZNFX1_HUMAN	107926	NFX1-type zinc finger-containing protein 1	ZNFX1	9	1,28	184 ; 148 ; 178 ; - ; - ; 59	1,79E+07	1,87E+07	1,75E+07	4,14E+06	9,24E+05	5,62E+06	5,07	0,00629
P80294	MT1H_HUMAN	6039	Metallothionein-1H	MT1H	1	19,67	31 ; 40 ; - ; - ; - ; -	6,68E+06	7,28E+06		1,63E+06	9,08E+05	1,66E+06	4,99	0,00289
P48681	NEST_HUMAN	177427	Nestin	NES	18	10,49	243 ; 264 ; 256 ; 392 ; 478 ; 267	4,59E+08	5,28E+08	5,51E+08	1,14E+08	1,09E+08	8,55E+07	4,98	0,00227
Q14CV6	Q14CV6_HUMAN	58868	Chromosome 6 open reading frame 150	MB21D1	1	3,64	- ; - ; 50 ; - ; - ; -	1,12E+07	1,08E+07	9,84E+06	2,84E+06	2,06E+06	1,69E+06	4,82	9,8E-05
P80075	CCL8_HUMAN	11246	C-C motif chemokine 8	CCL8	3	25,25	- ; 59 ; 59 ; - ; - ; -	2,82E+07	3,64E+07	4,04E+07		3,89E+06	1,10E+07	4,71	0,01584
Q10589	BST2_HUMAN	19769	Bone marrow stromal antigen 2	BST2	3	8,33	126 ; 94 ; 127 ; 53 ; 63 ; -	7,31E+07	9,29E+07	8,04E+07	1,86E+07	1,82E+07	1,76E+07	4,53	0,00794
C5IWY5	C5IWY5_HUMAN	40832	MHC class I antigen	HLA-A	10	42,19	263 ; - ; 288 ; 67 ; 100 ; 49	7,85E+07	7,81E+07	8,33E+07	1,79E+07	2,02E+07	1,88E+07	4,22	0,00015

Q9NRW7	VPS45_HUMAN	61107	Vacuolar protein sorting-associated protein 45	VPS45	5	8,61	85 ; 102 ; 120 ; 82 ; 91 ; 57	5,63E+07	6,37E+07	6,48E+07	1,59E+07	1,63E+07	1,29E+07	4,09	0,00105
Q96CV9	OPTN_HUMAN	65922	Optineurin	OPTN	21	34,14	366 ; 308 ; 392 ; 165 ; 110 ; 173	6,30E+07	6,65E+07	7,71E+07	1,81E+07	1,75E+07	1,59E+07	4,02	0,00575
Q9NRD1	FBX6_HUMAN	33933	F-box only protein 6	FBXO6	4	17,06	96 ; 53 ; 82 ; - ; - ; -	5,11E+06	3,58E+06	6,27E+06	1,47E+06	1,53E+06	7,73E+05	3,97	0,03154
Q8TDB6	DTX3L_HUMAN	83554	E3 ubiquitin-protein ligase DTX3L	DTX3L	9	12,57	191 ; 151 ; 233 ; - ; - ; -	4,04E+07	4,02E+07	4,18E+07	7,30E+06	1,58E+07	7,98E+06	3,94	0,00651
P29590	PML_HUMAN	97551	Protein PML	PML	21	20,41	334 ; 273 ; 203 ; 48 ; - ; -	1,47E+08	1,47E+08	1,38E+08	3,83E+07	3,20E+07	3,95E+07	3,92	1,7E-05
P42224	STAT1_HUMAN	87335	Signal transducer and activator of transcription 1-alpha/beta	STAT1	40	51,2	908 ; 953 ; 963 ; 558 ; 508 ; 597	3,97E+08	3,80E+08	3,80E+08	8,77E+07	1,07E+08	1,08E+08	3,82	6,2E-06
P10145	IL8_HUMAN	10697	Interleukin-8	IL8	2	26,32	50 ; 48 ; 49 ; - ; - ; -	3,40E+07	3,84E+07	3,84E+07	1,03E+07			3,59	0,00302
P30825	CTR1_HUMAN	67650	High affinity cationic amino acid transporter 1	SLC7A1	2	5,09	69 ; 97 ; 72 ; - ; 71 ; 94	4,96E+07	5,76E+07	5,45E+07	1,44E+07	1,50E+07	1,58E+07	3,58	0,00273
Q9NYA1	SPHK1_HUMAN	42518	Sphingosine kinase 1	SPHK1	2	2,08	44 ; 71 ; 71 ; - ; - ; -	1,27E+07	1,21E+07	1,19E+07	3,01E+06	3,69E+06	3,58E+06	3,57	1,1E-05
Q9Y508	RN114_HUMAN	25694	RING finger protein 114	RNF114	10	23,25	202 ; 216 ; 221 ; 133 ; 135 ; 127	4,66E+07	3,60E+07	4,44E+07	1,18E+07	1,14E+07	1,27E+07	3,54	0,01021
P61764	STXB1_HUMAN	63629	Syntaxin-binding protein 1	STXBP1	8	6,82	91 ; 136 ; 111 ; 49 ; 83 ; 37	9,00E+07	1,01E+08	1,09E+08	6,24E+07	1,19E+07	1,10E+07	3,52	0,04085
Q8TEW0	PARD3_HUMAN	110000	Partitioning defective 3 homolog	PARD3	2	2	- ; 53 ; - ; - ; - ; -	3,74E+06	4,78E+06	4,03E+06			1,19E+06	3,50	0,01054
Q96S99	PKHF1_HUMAN	40679	Pleckstrin homology domain-containing family F member 1	PLEKHF1	2	6,04	- ; - ; 56 ; - ; - ; -	7,07E+06	6,56E+06	7,05E+06	2,05E+06	2,05E+06	1,95E+06	3,42	0,00077
Q9Y6N5	SQRD_HUMAN	49950	Sulfide:quinone oxidoreductase, mitochondrial	SQRDL	22	10,67	382 ; 438 ; 422 ; 80 ; 140 ; 188	9,96E+07	8,50E+07	9,25E+07	2,59E+07	2,77E+07	2,75E+07	3,41	0,00359
Q08AF3	SLFN5_HUMAN	100983	Schlafen family member 5	SLFN5	12	12,57	280 ; 236 ; 229 ; 55 ; - ; 44	4,60E+07	4,62E+07	4,73E+07	1,39E+07	1,32E+07	1,42E+07	3,38	9E-07
Q9NRR3	C42S2_HUMAN	9223	CDC42 small effector protein 2	CDC42SE2	2	22,62	85 ; 67 ; 71 ; 48 ; - ; 73	1,09E+07	1,21E+07	1,25E+07	4,12E+06	2,93E+06	3,61E+06	3,32	0,00028
Q9Y666	S12A7_HUMAN	119106	Solute carrier family 12 member 7	SLC12A7	10	6,56	88 ; 78 ; 144 ; - ; - ; 28	2,50E+07	3,37E+07	2,51E+07	9,73E+06	7,94E+06	7,91E+06	3,28	0,01797
A8K319	A8K319_HUMAN	56892	cDNA FLJ76655, highly similar to Homo sapiens tripartite motif-c	TRIM22	11	13,86	146 ; 296 ; 193 ; 111 ; 135 ; 82	2,87E+07	3,51E+07	2,34E+07	9,39E+06	8,10E+06	9,31E+06	3,25	0,02565
Q9NUT2	ABCB8_HUMAN	75664	ATP-binding cassette sub-family B member 8, mitochondrial	ABCB8	15	27,41	353 ; 265 ; 346 ; 36 ; 42 ; 54	1,09E+08	1,20E+08	1,38E+08	3,39E+07	4,63E+07	3,49E+07	3,18	0,00339
Q9UBQ6	EXTL2_HUMAN	37466	Exostosin-like 2	EXTL2	3	11,82	- ; 40 ; 49 ; - ; - ; -	1,30E+07	1,25E+07	1,25E+07	4,06E+06	4,00E+06	4,01E+06	3,15	0,00041
Q13287	NMI_HUMAN	35085	N-myc-interactor	NMI	16	55,7	254 ; 315 ; 222 ; 52 ; 67 ; 44	3,35E+07	3,32E+07	3,12E+07	1,14E+07	8,82E+06	1,11E+07	3,13	3,8E-05
Q9Y6N7	ROBO1_HUMAN	175830	Roundabout homolog 1	ROBO1	1	0,75	46 ; - ; - ; - ; - ; -	6,00E+06	5,92E+06	5,63E+06			1,94E+06	3,01	0,00081
Q8WUI4	HDAC7_HUMAN	102927	Histone deacetylase 7	HDAC7	2	2,52	- ; - ; - ; 43 ; - ; 39	3,94E+07	3,96E+07	4,14E+07	1,31E+07	1,39E+07	1,32E+07	2,99	0,00014
Q13325	IFIT5_HUMAN	55946	Interferon-induced protein with tetratricopeptide repeats 5	IFIT5	11	21,78	162 ; 208 ; 146 ; 107 ; 76 ; 88	3,27E+07	3,10E+07	3,19E+07	1,04E+07	1,03E+07	1,14E+07	2,98	6,9E-06
O95379	TFIP8_HUMAN	23003	Tumor necrosis factor alpha-induced protein 8	TNFAIP8	4	17,17	139 ; 84 ; 107 ; 110 ; 91 ; 104	2,26E+07	1,74E+07	1,41E+07	5,28E+06	6,73E+06	6,16E+06	2,98	0,03719
Q8N128	F177A_HUMAN	23757	Protein FAM177A1	FAM177A1	3	16,9	87 ; 96 ; 111 ; - ; - ; -	8,95E+06	8,77E+06	8,42E+06	1,93E+06	3,20E+06	3,72E+06	2,95	0,00527
Q9NRL2	BAZ1A_HUMAN	178702	Bromodomain adjacent to zinc finger domain protein 1A	BAZ1A	11	5,53	150 ; 175 ; 141 ; - ; - ; -	1,80E+07	2,03E+07	1,93E+07	9,21E+06	2,72E+06	8,06E+06	2,88	0,01688
Q8ND71	GIMA8_HUMAN	74890	GTPase IMAP family member 8	GIMAP8	10	12,48	138 ; 144 ; 172 ; - ; - ; -	2,16E+07	2,38E+07	2,27E+07	8,59E+06	9,15E+06	5,92E+06	2,88	0,00063
Q9Y5A7	NUB1_HUMAN	70538	NEDD8 ultimate buster 1	NUB1	9	5,69	108 ; 192 ; 124 ; 84 ; 72 ; -	2,86E+07	2,29E+07	2,51E+07	8,92E+06	9,11E+06	8,87E+06	2,85	0,01007
P49590	SYHM_HUMAN	54115	Probable histidyl-tRNA synthetase, mitochondrial	HARS2	6	5,82	- ; - ; - ; 114 ; 64 ; 124	3,02E+07	3,20E+07	4,55E+07	1,31E+07	1,02E+07	1,48E+07	2,83	0,03272
Q92844	TANK_HUMAN	47816	TRAF family member-associated NF-kappa-B activator	TANK	3	9,88	56 ; 48 ; 55 ; - ; - ; -	9,66E+06	1,17E+07	9,46E+06	3,58E+06	3,40E+06	4,17E+06	2,77	0,00722
O60507	TPST1_HUMAN	39553	Protein-tyrosine sulfotransferase 1	TPST1	2	8,62	- ; 18 ; - ; - ; - ; -	2,20E+06	3,59E+06	2,49E+06	1,19E+06	1,04E+06	8,17E+05	2,72	0,0462
Q16666	IF16_HUMAN	82876	Gamma-interferon-inducible protein 16	IFI16	34	29,33	601 ; 718 ; 718 ; 365 ; 344 ; 462	2,66E+08	2,51E+08	3,30E+08	9,19E+07	1,21E+08	1,00E+08	2,70	0,01051
P50570	DYN2_HUMAN	98064	Dynamin-2	DNM2	27	31,49	418 ; 513 ; 441 ; 418 ; 459 ; 556	2,55E+08	2,61E+08	2,53E+08	1,39E+08	8,87E+07	6,49E+07	2,63	0,01733
Q9P2N5	RBM27_HUMAN	88802	RNA-binding protein 27	RBM27	3	4,29	74 ; 48 ; - ; 71 ; - ; -	2,71E+07	3,03E+07	2,49E+07	8,19E+06	1,04E+07	1,27E+07	2,63	0,00129
P09341	GROA_HUMAN	11301	Growth-regulated alpha protein	CXCL1	3	14,95	42 ; 66 ; 44 ; - ; - ; -	7,82E+06	1,10E+07	1,25E+07	3,30E+06		4,69E+06	2,61	0,02842
P40305	IFI27_HUMAN	11268	Interferon alpha-inducible protein 27, mitochondrial	IFI27	4	37,82	155 ; 174 ; 172 ; - ; - ; -	4,91E+07	5,72E+07	5,44E+07	2,01E+07	2,46E+07	1,68E+07	2,61	0,00055
Q92851	CASPA_HUMAN	58951	Caspase-10	CASP10	6	2,69	- ; 179 ; 37 ; - ; - ; -	8,39E+06	1,31E+07	9,10E+06	3,68E+06	3,81E+06	4,24E+06	2,61	0,04817
Q14258	TRI25_HUMAN	70989	E3 ubiquitin/ISG15 ligase TRIM25	TRIM25	32	59,05	839 ; 703 ; 738 ; 689 ; 676 ; 681	1,71E+08	1,68E+08	1,90E+08	7,05E+07	7,28E+07	6,13E+07	2,58	0,00079
Q9UGQ3	GTR6_HUMAN	48041	Solute carrier family 2, facilitated glucose transporter member 6	SLC2A6	1	2,25	66 ; 66 ; 64 ; 66 ; 55 ; 65	1,38E+07	1,41E+07	1,62E+07	6,05E+06	5,11E+06	5,94E+06	2,58	0,00292
P17706	PTN2_HUMAN	41097	Tyrosine-protein phosphatase non-receptor type 2	PTPN2	3	5,1	83 ; 59 ; 52 ; - ; - ; -	6,04E+06	6,05E+06	6,21E+06	2,35E+06	1,97E+06	2,78E+06	2,58	0,00266
B6ETL5	B6ETL5_HUMAN	47937	Pannexin 1	PANX1	3	5,41	81 ; 73 ; 66 ; 38 ; 12 ; 71	7,37E+06	7,77E+06	8,63E+06	2,84E+06	3,34E+06	3,12E+06	2,55	0,00236
Q8NHV1	GIMA7_HUMAN	34509	GTPase IMAP family member 7	GIMAP7	6	24	149 ; 179 ; 169 ; 45 ; - ; 48	2,42E+07	2,50E+07	2,51E+07	1,08E+07	8,44E+06	9,87E+06	2,55	0,00055
P40189	IL6RB_HUMAN	23034	Interleukin-6 receptor subunit beta	IL6ST	2	12,81	19 ; - ; - ; - ; - ; -	4,36E+07	3,93E+07	4,28E+07	1,76E+07	1,42E+07	1,79E+07	2,53	0,00016
Q16739	CEGT_HUMAN	44854	Ceramide glucosyltransferase	UGCG	2	7,87	100 ; 98 ; 107 ; 80 ; - ; -	8,02E+06	6,72E+06	6,91E+06	2,86E+06	2,75E+06	2,95E+06	2,53	0,00754
P52569	CTR2_HUMAN	71673	Low affinity cationic amino acid transporter 2	SLC7A2	9	17,48	249 ; 274 ; 241 ; 73 ; - ; -	1,63E+08	1,67E+08	1,60E+08	6,33E+07	6,85E+07	6,36E+07	2,51	4,9E-06
P19838	NFKB1_HUMAN	21870	Nuclear factor NF-kappa-B p105 subunit	NFKB1	6	11,62	59 ; 91 ; 44 ; - ; - ; -	2,05E+07	2,24E+07	2,02E+07	1,00E+07	8,69E+06	6,62E+06	2,49	0,00091
O95183	VAMP5_HUMAN	12805	Vesicle-associated membrane protein 5	VAMP5	4	32,76	133 ; 133 ; 126 ; - ; 67 ; 70	2,19E+07	2,94E+07	3,33E+07	1,17E+07	1,10E+07	1,14E+07	2,48	0,03695
P28370	SMCA1_HUMAN	113358	Probable global transcription activator SNF2L1	SMARCA1	8	8,29	- ; 156 ; - ; - ; - ; -	2,98E+07	3,09E+07	2,51E+07	9,50E+06		1,40E+07	2,43	0,02201
O60825	F262_HUMAN	47426	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	PFKFB2	1	2,22	30 ; - ; - ; - ; - ; -	3,03E+06	2,51E+06	2,56E+06		1,11E+06	1,12E+06	2,43	0,01047
Q9ULX9	MAFF_HUMAN	17760	Transcription factor MafF	MAFF	5	9,76	113 ; 113 ; 120 ; 49 ; 50 ; 83	1,72E+07	1,74E+07	1,61E+07	7,50E+06	6,43E+06	7,01E+06	2,42	6,4E-05
P42226	STAT6_HUMAN	94134	Signal transducer and activator of transcription 6	STAT6	10	6,02	82 ; 62 ; 135 ; - ; 49 ; -	3,06E+07	2,93E+07	3,24E+07	1,49E+07	1,09E+07	1,30E+07	2,38	0,00037
Q9H0J9	PAR12_HUMAN	79064	Poly [ADP-ribose] polymerase 12	PARP12	8	3,85	166 ; 112 ; 161 ; 128 ; 88 ; 83	1,83E+07	1,84E+07	1,97E+07	8,05E+06	8,41E+06	7,28E+06	2,38	7,3E-05
Q3B8N2	LEG9B_HUMAN	35918	Galectin-9B	LGALS9B	9	11,15	308 ; 288 ; 269 ; 85 ; 97 ; 58	1,64E+08	1,77E+08	1,69E+08	7,54E+07	6,99E+07	7,07E+07	2,36	0,00034
Q8IUW5	RELL1_HUMAN	29340	RELT-like protein 1	RELL1	5	24,72	150 ; 111 ; 157 ; 60 ; - ; 26	2,50E+07	2,67E+07	2,89E+07	1,28E+07	9,43E+06	1,20E+07	2,35	0,00056
Q86XL3	ANKL2_HUMAN	104114	Ankyrin repeat and LEM domain-containing protein 2	ANKLE2	1	1,92	- ; 42 ; 38 ; - ; - ; -	6,93E+06	8,17E+06	8,29E+06	2,49E+06	3,73E+06	3,74E+06	2,35	0,00176
Q13501	SQSTM_HUMAN	47687	Sequestosome-1	SQSTM1	16	50,91	384 ; 413 ; 140 ; 54 ; 66 ; 96	1,76E+08	2,07E+08	1,48E+08	7,72E+07	7,58E+07	7,42E+07	2,34	0,02629
P23497	SP100_HUMAN	96549	Nuclear autoantigen Sp-100	SP100	10	5,46	106 ; 130 ; 152 ; - ; 61 ; -	1,54E+08	1,47E+08	1,43E+08	7,03E+07	5,04E+07	6,90E+07	2,34	0,00171
O75173	ATS4_HUMAN	90225	A disintegrin and metalloproteinase with thrombospondin moti	ADAMTS4	2	3,58	- ; 46 ; 52 ; - ; - ; -	8,65E+06	8,55E+06	8,50E+06	3,83E+06	2,61E+06	4,89E+06	2,27	0,01795
P49662	CASP4_HUMAN	43262	Caspase-4	CASP4	12	24,14	157 ; 81 ; 160 ; 66 ; 58 ; 125	1,90E+07	2,48E+07	2,13E+07	9,74E+06	8,96E+06	1,02E+07	2,25	0,01556
Q9H845	ACAD9_HUMAN	68760	Acyl-CoA dehydrogenase family member 9, mitochondrial	ACAD9	13	18,2	198 ; 238 ; 126 ; 147 ; 198 ; 276	9,74E+08	1,03E+09	1,39E+09	5,16E+08	5,46E+08	4,67E+08	2,22	0,037
Q92985	IRF7_HUMAN	54278	Interferon regulatory factor 7	IRF7	3	8,55	37 ; 53 ; 60 ; - ; - ; -	1,48E+07	1,57E+07	1,88E+07	8,40E+06	7,41E+06	6,80E+06	2,19	0,01016
Q9UPU9	SMAG1_HUMAN	79415	Protein Smaug homolog 1	SAMD4A	1	2,09	35 ; - ; 39 ; - ; - ; -	2,42E+07	1,90E+07	2,13E+07	1,05E+07	9,30E+06	9,95E+06	2,17	0,01322
P63218	GBG5_HUMAN	7318	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gam	GNG5	5	73,53	34 ; 102 ; 78 ; 51 ; - ; 42	2,65E+07	3,84E+07	3,93E+07	2,06E+07	1,09E+07	1,65E+07	2,17	0,02535
B4DNN6	B4DNN6_HUMAN	14233	cDNA FLJ52003, highly similar to Homo sapiens nucleoredoxin (N	NXN	1	13,49	- ; - ; 27 ; - ; - ; -	3,43E+07	2,24E+07	3,17E+07	1,19E+07	1,70E+07	1,25E+07	2,14	0,03278
P02794	FRIH_HUMAN	21226	Ferritin heavy chain	FTH1	20	74,86	602 ; 566 ; 614 ; 367 ; 354 ; 419	5,53E+08	5,64E+08	5,52E+08	2,59E+08	2,47E+08	2,82E+08	2,12	0,00031
Q13829	BACD2_HUMAN	36204	BTB/POZ domain-containing adapter for CUL3-mediated RhoA de	TNFAIP1	4	6,96	75 ; 57 ; 100 ; - ; - ; -	5,09E+06	4,26E+06	4,90E+06	1,46E+06	2,26E+06	3,01E+06	2,12	0,01472
P23458	JAK1_HUMAN	131690	Tyrosine-protein kinase JAK1	JAK1	8	1,31	162 ; 107 ; 125 ; 62 ; 55 ; 80	2,00E+07	1,79E+07	1,91E+07	8,00E+06	9,71E+06	9,26E+06	2,11	0,00026
Q5EBM0	CMPK2_HUMAN	49448	UMP-CMP kinase 2, mitochondrial	CMPK2	4	10,24	93 ; 90 ; 120 ; - ; - ; -	1,17E+07	1,47E+07	1,43E+07			6,50E+06	2,09	0,01747
Q9NZ45	CISD1_HUMAN	12199	CDGSH iron-sulfur domain-containing protein 1	CISD1	7	25,93	124 ; 146 ; 138 ; 124 ; 116 ; 144	3,79E+07	4,29E+07	3,94E+07	1,80E+07	1,48E+07	2,49E+07	2,08	0,00876
P51159	RB27A_HUMAN	24868	Ras-related protein Rab-27A	RAB27A	8	28,05	131 ; 115 ; 116 ; 109 ; 92 ; 113	4,11E+07	4,74E+07	4,26E+07	2,10E+07	2,05E+07	2,14E+07	2,08	0,00622
Q15046	SYK_HUMAN	68048	Lysyl-tRNA synthetase	KARS	31	41,54	566 ; 487 ; 563 ; 492 ; 421 ; 510	2,15E+08	2,51E+08	2,28E+08	1,14E+08	1,20E+08	9,96E+07	2,08	0,00169
Q00403	TF2B_HUMAN	34833	Transcription initiation factor IIB	GTF2B	11	16,77	321 ; 388 ; 358 ; 156 ; 241 ; 208	1,49E+07	1,22E+07	1,65E+07	5,99E+06	8,32E+06	6,74E+06	2,07	0,01203
P05121	PAI1_HUMAN	45060	Plasminogen activator inhibitor 1	SERPINE1	40	70,65	1200 ; 1006 ; 1319 ; 823 ; 855 ; 98	1,73E+09	1,96E+09	2,27E+09	9,52E+08	9,87E+08	9,90E+08	2,03	0,02276
Q7Z7L1	SLN11_HUMAN	102837	Schlafen family member 11	SLFN11	9	9,21	133 ; 160 ; 135 ; 46 ; - ; 48	3,07E+07	3,07E+07	3,05E+07	1,62E+07	1,31E+07	1,62E+07	2,02	0,0044
Q9Y5K6	CD2AP_HUMAN	71451	CD2-associated protein	CD2AP	16	25,67	276 ; 268 ; 244 ; 192 ; 179 ; 243	8,05E+07	7,40E+07	7,69E+07	3,77E+07	3,61E+07	4,11E+07	2,01	0,00012
Q9UKV5	AMFR2_HUMAN	72996	E3 ubiquitin-protein ligase AMFR	AMFR	1	2,95	- ; - ; - ; 59 ; - ; 48	1,31E+06	6,00E+05	6,63E+05	1,80E+06	1,44E+06	2,02E+06	0,49	0,03782
P52732	KIF11_HUMAN	119159	Kinesin-like protein KIF11	KIF11	1	1,04	- ; 61 ; - ; 66 ; 54 ; 62	1,87E+06	2,67E+06	3,13E+06	4,97E+06	4,97E+06	5,80E+06	0,49	0,0054
O95864	FADS2_HUMAN	48883	Fatty acid desaturase 2	FADS2	7	11,14	63 ; - ; - ; 66 ; 71 ; 64	1,05E+07	1,00E+07	1,22E+07	2,12E+07	2,33E+07	2,26E+07	0,49	0,00021
Q9NUW8	TYDP1_HUMAN	68420	Tyrosyl-DNA phosphodiesterase 1	TDP1	1	3,13	- ; - ; - ; 37 ; - ; -	1,58E+06	1,99E+06	1,11E+06	3,39E+06	2,98E+06	3,41E+06	0,48	0,00883
Q99865	SPIN3_HUMAN	29601	Spindlin-2A	SPIN2A	2	11,45	66 ; - ; - ; 50 ; 77 ; 63	4,80E+06	4,07E+06	4,47E+06	1,02E+07	8,85E+06	9,20E+06	0,47	0,00179
P25116	PAR1_HUMAN	37397	Proteinase-activated receptor 1	F2R	1	5,36	- ; - ; - ; - ; - ; 65	3,41E+06	4,70E+06	4,41E+06	1,10E+07	7,85E+06	7,76E+06	0,47	0,03594
O43155	FLRT2_HUMAN	73958	Leucine-rich repeat transmembrane protein FLRT2	FLRT2	5	10,15	57 ; - ; 66 ; 76 ; 91 ; 107	9,80E+06	1,10E+07	9,74E+06	2,15E+07	2,18E+07	2,26E+07	0,46	3,8E-05
O60443	DFNA5_HUMAN	47735	Non-syndromic hearing impairment protein 5	DFNA5	1	2,52	- ; - ; - ; 51 ; 54 ; 41	2,61E+06	3,23E+06	3,43E+06	6,07E+06	6,92E+06	7,20E+06	0,46	0,00144
Q9H2C0	GAN_HUMAN	67638	Gigaxonin	GAN	1	3,69	- ; - ; - ; 47 ; 101 ; -	1,42E+06	1,58E+06	7,97E+05	2,66E+06	3,15E+06	2,49E+06	0,46	0,00924
Q53F19	CQ085_HUMAN	70593	Uncharacterized protein C17orf85	C17orf85	2	4,19	- ; - ; - ; - ; - ; 44	5,09E+06	7,27E+06	5,64E+06	1,34E+07	1,31E+07	1,29E+07	0,46	0,00651
Q96EA4	SPDLY_HUMAN	61015	Protein Spindly	CCDC99	1	1,9	- ; - ; - ; 38 ; - ; -		1,61E+06	1,70E+06	3,17E+06	4,08E+06	3,66E+06	0,46	0,01506
Q9UNN5	FAF1_HUMAN	73954	FAS-associated factor 1	FAF1	1	2	62 ; - ; - ; 49 ; 50 ; 44	5,12E+06	3,58E+06	4,21E+06	9,66E+06	9,62E+06	9,40E+06	0,45	0,0058
Q8IWB1	IPRI_HUMAN	62060	Inositol 1,4,5-triphosphate receptor-interacting protein	ITPRIP	1	1,83	- ; - ; - ; 43 ; 48 ; -		1,33E+06	8,33E+05	2,71E+06	1,86E+06	2,70E+06	0,45	0,04024
O75976	CBPD_HUMAN	152931	Carboxypeptidase D	CPD	8	3,84	89 ; 141 ; 83 ; 104 ; 58 ; 129	6,90E+06	6,24E+06	6,98E+06	1,67E+07	1,64E+07	1,24E+07	0,44	0,02378
P23743	DGKA_HUMAN	82630	Diacylglycerol kinase alpha	DGKA	7	2,59	47 ; - ; 54 ; 175 ; 99 ; 95	1,47E+07	6,15E+06	9,94E+06	2,49E+07	2,26E+07	2,43E+07	0,43	0,02417
Q2T9J0	TYSD1_HUMAN	59309	Peroxisomal leader peptide-processing protease	TYSND1	1	1,94	- ; - ; - ; - ; - ; 56	4,03E+05	3,99E+05	2,11E+05	7,78E+05	6,82E+05	9,41E+05	0,42	0,01008
Q8IWF6	F116A_HUMAN	69575	Protein FAM116A	FAM116A	1	3,29	- ; - ; - ; 38 ; - ; -	7,94E+05	9,18E+05	5,56E+05	2,13E+06	1,84E+06	1,46E+06	0,42	0,01575
P10589	COT1_HUMAN	45571	COUP transcription factor 1	NR2F1	8	12,56	66 ; - ; 30 ; 146 ; 139 ; 201	6,27E+06	4,55E+06	5,22E+06	1,15E+07	1,15E+07	1,57E+07	0,41	0,02232
P11308	ERG_HUMAN	53857	Transcriptional regulator ERG	ERG	7	13,78	- ; - ; - ; 79 ; 101 ; 138	1,08E+07	1,01E+07	1,07E+07	2,57E+07	2,76E+07	2,34E+07	0,41	0,00525
Q8N8R3	MCATL_HUMAN	32062	Mitochondrial carnitine/acylcarnitine carrier protein CACL	SLC25A29	7	37,29	72 ; - ; - ; 118 ; 79 ; 179	3,31E+06	2,40E+06	2,82E+06	6,71E+06	6,24E+06	8,09E+06	0,41	0,00763
Q9P2E7	PCD10_HUMAN	97002	Protocadherin-10	PCDH10	4	4,46	- ; - ; - ; 83 ; 78 ; 77	4,92E+06	6,87E+06	5,93E+06	1,67E+07	1,54E+07	1,23E+07	0,40	0,01089
Q12894	IFRD2_HUMAN	65598	Interferon-related developmental regulator 2	IFRD2	1	2,8	- ; - ; - ; 52 ; - ; 42	1,47E+06	1,46E+06	5,25E+05	2,80E+06	2,99E+06	2,90E+06	0,40	0,02769
O14786	NRP1_HUMAN	68376	Neuropilin-1	NRP1	1	2,13	- ; - ; 47 ; 61 ; - ; 56	5,69E+06	6,09E+06	4,79E+06	1,29E+07	1,68E+07	1,37E+07	0,38	0,0112
Q8N699	MYCT1_HUMAN	26593	Myc target protein 1	MYCT1	3	17,02	- ; - ; - ; 65 ; 64 ; 70	8,87E+06	7,41E+06	8,37E+06	2,10E+07	2,21E+07	2,21E+07	0,38	2,2E-05
Q9BT88	SYT11_HUMAN	48297	Synaptotagmin-11	SYT11	2	3,25	- ; 52 ; 57 ; - ; - ; 40	1,31E+07	8,37E+06	1,12E+07	3,05E+07	2,62E+07	3,38E+07	0,36	0,00338
Q53EL6	PDCD4_HUMAN	50562	Programmed cell death protein 4	PDCD4	10	3,71	- ; 96 ; - ; 159 ; 172 ; 219	8,58E+06	8,36E+06	9,94E+06	2,62E+07	2,52E+07	2,63E+07	0,35	2,5E-05
P00352	AL1A1_HUMAN	54862	Retinal dehydrogenase 1	ALDH1A1	31	61,88	704 ; 652 ; 643 ; 806 ; 739 ; 826	2,38E+08	3,69E+08	1,89E+08	8,34E+08	7,81E+08	8,05E+08	0,33	0,00629
Q14690	RRP5_HUMAN	208701	Protein RRP5 homolog	PDCD11	22	4,38	172 ; 115 ; 142 ; 142 ; 63 ; 192	4,06E+07	6,03E+07	2,38E+07	1,28E+08	1,97E+08	1,13E+08	0,28	0,0407
Q8ND90	PNMA1_HUMAN	39761	Paraneoplastic antigen Ma1	PNMA1	3	6,52	- ; - ; - ; - ; 41 ; 46	1,28E+06	1,07E+06	1,11E+06	3,93E+06	4,52E+06	4,08E+06	0,28	0,00123
Q8WVD5	RN141_HUMAN	25535	RING finger protein 141	RNF141	1	5,22	- ; - ; - ; 47 ; - ; -	2,12E+05		1,73E+05	9,23E+05	7,71E+05	5,55E+05	0,26	0,03162
P49796	RGS3_HUMAN	35108	Regulator of G-protein signaling 3	RGS3	2	7,07	- ; - ; - ; 53 ; 60 ; 50	1,45E+06	8,05E+05	5,51E+05	3,79E+06	3,59E+06	3,98E+06	0,25	0,00361
O00767	ACOD_HUMAN	41505	Acyl-CoA desaturase	SCD	1	3,34	- ; - ; - ; 38 ; - ; -	1,00E+05	1,00E+05	1,00E+05	2,36E+06	2,39E+06	2,21E+06	0,04	0,00061

Annexe 6 : Table des protéines variantes identifiées après analyse quantitative du protéome total des HUVEC suite à la stimulation IL-1β des HUVEC. Surlignées en bleu : protéines identifiées variantes à la fois après stimulation TNFα-IFNγ, IL-1β et IL-33 ; surlignées en vert : protéines variantes après stimulation IL-1β et IL-33. (AC : numéro d'accession, ID : identifiant de la protéine, MW : molecular weight).

Protein AC	Protein ID	MW	Protein name	Genename	Number quantified	Coverage (%)	Mascot score IL1b-1; IL1b-2; IL1b-3; CT1; CT2; CT3	PAI IL1b 1	PAI IL1b 2	PAI IL1b 3	PAI CT 1	PAI CT 2	PAI CT 3	Ratio	Student
P16581	LYAM2_HUMAN	66629	E-selectin precursor	SELE	12	20	159 ; 215 ; 154 ; - ; - ; -	1,42E+08	1,73E+08	9,85E+07	1,00E+05	1,00E+05	1,00E+05	1379,98	0,023923706
Q13077	TRAF1_HUMAN	46164	TNF receptor-associated factor 1	TRAF1	7	14,42	156 ; 260 ; 146 ; - ; - ; -	7,02E+07	7,52E+07	7,65E+07	1,00E+05	1,00E+05	1,00E+05	739,52	0,000672405
P49716	CEBPD_HUMAN	14056	CCAAT/enhancer-binding protein delta	CEBPD	2	12,03	40 ; - ; 35 ; - ; - ; -	2,45E+07	2,07E+07	1,96E+07	1,00E+05	1,00E+05	1,00E+05	215,94	0,004788061
P25774	CATS_HUMAN	37496	Cathepsin S precursor	CTSS	1	4,83	75 ; 75 ; 80 ; - ; - ; -	2,08E+07	2,03E+07	2,13E+07	1,00E+05	1,00E+05	1,00E+05	207,90	0,000200602
Q14213	IL27B_HUMAN	25396	Interleukin-27 subunit beta precursor	EBI3	1	5,24	51 ; 61 ; - ; - ; - ; -	1,49E+07	1,67E+07	1,40E+07	1,00E+05	1,00E+05	1,00E+05	152,18	0,00278978
P13500	CCL2_HUMAN	11025	C-C motif chemokine 2 precursor	CCL2	3	11,11	- ; 33 ; 41 ; - ; - ; -	1,18E+07	1,23E+07	1,81E+07	1,00E+05	1,00E+05	1,00E+05	140,98	0,0203126
D7RVA3	D7RVA3_HUMAN	14813	Granulocyte colony-stimulating factor (Fragment)	CSF3	1	11,59	51 ; 66 ; 82 ; - ; - ; -	1,04E+07	1,17E+07	1,04E+07	1,00E+05	1,00E+05	1,00E+05	108,22	0,001786216
Q9UP65	PA24C_HUMAN	60949	Cytosolic phospholipase A2 gamma precursor	PLA2G4C	2	2,96	- ; 30 ; 64 ; - ; - ; -	6,87E+06	1,03E+07	1,53E+07	1,00E+05	1,00E+05	1,00E+05	108,22	0,048292983
Q9C0I1	MTMRC_HUMAN	86148	Myotubularin-related protein 12	MTMR12	1	2,54	- ; 62 ; - ; - ; - ; -	8,49E+06	1,33E+07	9,30E+06	1,00E+05	1,00E+05	1,00E+05	103,58	0,020146298
Q9UGQ3	GTR6_HUMAN	54539	Solute carrier family 2, facilitated glucose transporter member 6	SLC2A6	1	1,97	40 ; - ; - ; - ; - ; -	1,23E+07	7,10E+06	8,03E+06	1,00E+05	1,00E+05	1,00E+05	91,52	0,030138727
P17275	JUNB_HUMAN	35879	Transcription factor jun-B	JUNB	1	4,61	36 ; - ; - ; - ; - ; -	7,97E+06	7,26E+06	8,62E+06	1,00E+05	1,00E+05	1,00E+05	79,50	0,002499254
B2R8U9	B2R8U9_HUMAN	45171	cDNA, FLJ94074, highly similar to Homo sapiens caspase 1, apoptos	CASP1	1	2,48	39 ; - ; - ; - ; - ; -	6,41E+06	6,09E+06	6,46E+06	1,00E+05	1,00E+05	1,00E+05	63,23	0,000350048
P80162	CXCL6_HUMAN	11897	C-X-C motif chemokine 6 precursor	CXCL6	3	21,93	84 ; 90 ; 98 ; - ; - ; -	1,97E+08	1,46E+08	2,59E+08	1,42E+06	4,41E+06	4,72E+06	57,06	0,026108844
P09603	CSF1_HUMAN	49932	Macrophage colony-stimulating factor 1 precursor	CSF1	1	3,9	49 ; 52 ; 66 ; - ; - ; -	5,49E+06	5,84E+06	5,09E+06	1,00E+05	1,00E+05	1,00E+05	54,75	0,001642813
P19320	VCAM1_HUMAN	81264	Vascular cell adhesion protein 1 precursor	VCAM1	2	2,57	- ; 41 ; - ; - ; - ; -	2,19E+08	2,57E+08	2,46E+08	8,43E+06	6,38E+06	0,00E+00	48,83	0,001479461
P05362	ICAM1_HUMAN	57825	Intercellular adhesion molecule 1 precursor	ICAM1	21	2,63	582 ; 489 ; 553 ; - ; 32 ; 42	1,53E+09	1,65E+09	1,54E+09	2,93E+07	5,18E+07	4,23E+07	38,22	0,000441085
Q9NWX5	ASB6_HUMAN	22373	Ankyrin repeat and SOCS box protein 6	ASB6	1	5,08	58 ; 44 ; 43 ; - ; - ; -	4,24E+06	2,63E+06	3,20E+06	1,00E+05	1,00E+05	1,00E+05	33,60	0,020297509
Q03169	TNAP2_HUMAN	72661	Tumor necrosis factor alpha-induced protein 2	TNFAIP2	26	43,43	549 ; 496 ; 460 ; - ; - ; -	1,25E+08	1,10E+08	1,12E+08	3,25E+06	4,26E+06	5,91E+06	25,85	0,001268659
P19876	CXCL3_HUMAN	11342	C-X-C motif chemokine 3 precursor	CXCL3	4	14,95	74 ; 52 ; 64 ; - ; - ; -	2,80E+07	3,39E+07	4,02E+07	1,37E+06			24,86	0,010118912
P78556	CCL20_HUMAN	10762	C-C motif chemokine 20 precursor	CCL20	3	33,33	58 ; - ; 35 ; - ; - ; -	3,88E+07	1,57E+08	1,41E+08	1,10E+07	4,56E+06	3,70E+06	17,49	0,103142169
Q96AZ6	ISG20_HUMAN	20363	Interferon-stimulated gene 20 kDa protein	ISG20	10	49,17	167 ; 273 ; 151 ; - ; - ; -	2,98E+07	3,53E+07	3,41E+07	1,16E+06	2,22E+06	2,88E+06	15,82	0,001494986
P52569	CTR2_HUMAN	71673	Low affinity cationic amino acid transporter 2	SLC7A2	7	11,09	209 ; 112 ; 202 ; - ; - ; -	1,58E+08	1,42E+08	1,35E+08	4,86E+06	1,59E+07	9,07E+06	14,58	0,00051274
O95236	APOL3_HUMAN	44278	Apolipoprotein L3	APOL3	11	19,65	244 ; 274 ; 199 ; - ; - ; -	8,07E+07	6,95E+07	3,80E+07	4,91E+06	4,43E+06	5,48E+06	12,70	0,04553321
P04179	SODM_HUMAN	23673	Superoxide dismutase [Mn], mitochondrial precursor	SOD2	12	32,39	209 ; 213 ; 292 ; 138 ; 119 ; 181	1,89E+09	1,96E+09	2,02E+09	1,25E+08	2,06E+08	1,47E+08	12,30	1,23089E-05
P10145	IL8_HUMAN	10697	Interleukin-8 precursor	IL8	3	26,32	52 ; 48 ; 40 ; - ; - ; -	6,44E+07	7,07E+07	8,93E+07	6,41E+06	6,99E+06	7,39E+06	10,79	0,011812991
O75144	ICOSL_HUMAN	51925	ICOS ligand precursor	ICOSLG	2	4,23	51 ; 57 ; 57 ; - ; - ; -	4,36E+07	3,13E+07	4,10E+07		3,62E+06	4,21E+06	9,86	0,005959815
Q9BYX4	IFIH1_HUMAN	116689	Interferon-induced helicase C domain-containing protein 1	IFIH1	1	0,88	40 ; - ; - ; - ; - ; -	9,05E+06	1,20E+07	1,15E+07	1,26E+06	1,26E+06	1,07E+06	9,04	0,00837288
B4DJ74	B4DJ74_HUMAN	65221	cDNA FLJ58398, highly similar to A-kinase anchor protein 8	0	6	11,62	33 ; 92 ; 122 ; 32 ; - ; -	1,49E+08	1,33E+08	1,60E+08	1,76E+07	1,81E+07	1,72E+07	8,38	0,003661517
D3U3W1	D3U3W1_HUMAN	20979	MHC class I antigen (Fragment)	HLA-B	7	44,75	180 ; - ; - ; - ; - ; -	1,65E+08	1,40E+08	1,42E+08	1,76E+07	1,74E+07	2,60E+07	7,35	0,001577274
Q01201	RELB_HUMAN	62134	Transcription factor RelB	RELB	2	6,56	49 ; - ; 41 ; - ; - ; -	3,95E+07	3,01E+07	3,69E+07	5,43E+06	5,07E+06	4,17E+06	7,27	0,007457688
Q9UK39	NOCT_HUMAN	48196	Nocturnin	CCRN4L	10	23,2	232 ; 246 ; 181 ; - ; - ; -	2,89E+07	3,26E+07	3,30E+07	5,75E+06	4,55E+06	3,94E+06	6,64	0,00065802
Q96CG3	TIFA_HUMAN	21445	TRAF-interacting protein with FHA domain-containing protein A	TIFA	5	19,57	82 ; 57 ; 46 ; - ; - ; 38	1,41E+07	3,74E+07	1,40E+07	6,66E+06	1,29E+06	2,83E+06	6,09	0,138607615
Q96CG8	CTHR1_HUMAN	26224	Collagen triple helix repeat-containing protein 1 precursor	CTHRC1	4	8,23	71 ; 82 ; 79 ; - ; - ; -	4,84E+07	4,80E+07	5,72E+07	5,93E+06	1,43E+07	5,46E+06	5,97	0,000510091
O75191	XYLB_HUMAN	58382	Xylulose kinase	XYLB	1	2,43	49 ; - ; - ; - ; - ; -	6,12E+07	4,77E+07	2,54E+07		8,21E+06		5,45	0,048991934
P31431	SDC4_HUMAN	21642	Syndecan-4 precursor	SDC4	8	7,07	226 ; 249 ; 191 ; - ; 45 ; 49	3,74E+07	4,64E+07	3,70E+07	8,05E+06	5,95E+06	8,28E+06	5,42	0,006155173
Q9ULX9	MAFF_HUMAN	14285	Transcription factor MafF	MAFF	3	7,14	77 ; 54 ; 24 ; - ; 50 ; -	3,49E+07	5,11E+07	4,39E+07	1,17E+07	8,44E+06	4,33E+06	5,31	0,007910434
Q8IZV5	RDH10_HUMAN	38087	Retinol dehydrogenase 10	RDH10	2	4,99	- ; 47 ; 38 ; - ; - ; -	7,29E+06	4,13E+06	6,04E+06	1,11E+06			5,25	0,018243471
A8K9D9	A8K9D9_HUMAN	96678	Nuclear factor of kappa light polypeptide gene enhancer in B-cells	NFKB2	28	28,92	354 ; 370 ; 375 ; - ; - ; -	7,32E+07	6,85E+07	6,20E+07	1,31E+07	1,48E+07	1,29E+07	5,00	0,871023212
C6H0P0	C6H0P0_HUMAN	31561	MHC class I antigen (Fragment)	HLA-A	9	37,73	- ; 262 ; - ; - ; - ; -	2,10E+08	2,14E+08	1,77E+08	4,17E+07	3,40E+07	4,48E+07	4,99	0,00318515
P24001	IL32_HUMAN	20777	Interleukin-32 precursor	IL32	2	10,06	- ; 33 ; 34 ; - ; - ; -	1,92E+07	2,00E+07	2,13E+07	4,15E+06	3,35E+06	4,91E+06	4,87	5,27195E-05
Q9NR00	CH004_HUMAN	12337	Uncharacterized protein C8orf4	C8orf4	2	20,75	82 ; 86 ; 69 ; - ; - ; -	1,10E+07	1,09E+07	9,72E+06	2,68E+06		1,93E+06	4,58	0,001722631
P02795	MT2_HUMAN	6042	Metallothionein-2	MT2A	3	21,31	281 ; 291 ; 298 ; 120 ; 101 ; 155	2,65E+08	2,72E+08	2,82E+08	5,51E+07	5,97E+07	6,89E+07	4,45	7,02284E-06
Q8IV20	CM031_HUMAN	47780	UPF0124 protein C13orf31	C13orf31	1	2,79	70 ; 61 ; 68 ; - ; - ; -	1,15E+07	1,87E+07	1,61E+07	2,52E+06	2,63E+06	5,55E+06	4,33	0,015864768
P98172	EFNB1_HUMAN	38007	Ephrin-B1 precursor	EFNB1	2	8,96	- ; 57 ; 59 ; - ; - ; -	2,45E+07	2,83E+07	3,37E+07	3,67E+06	7,92E+06	9,07E+06	4,18	0,004436004
P55011	S12A2_HUMAN	131447	Solute carrier family 12 member 2	SLC12A2	27	3,38	362 ; 359 ; 336 ; 83 ; 90 ; 85	1,14E+08	1,22E+08	1,44E+08	3,07E+07	3,10E+07	3,15E+07	4,08	0,008948279
Q15025	TNIP1_HUMAN	71788	TNFAIP3-interacting protein 1	TNIP1	1	1,89	- ; 37 ; - ; - ; - ; -	1,77E+07	2,56E+07	1,76E+07		5,15E+06		3,95	0,006909657
P24386	RAE1_HUMAN	57265	Rab proteins geranylgeranyltransferase component A 1	CHM	3	6,34	- ; 53 ; 46 ; - ; - ; -	4,93E+06	8,33E+06	9,65E+06	2,59E+06	1,34E+06	2,02E+06	3,85	0,048962556
Q9NQ30	ESM1_HUMAN	20095	Endothelial cell-specific molecule 1 precursor	ESM1	4	23,91	49 ; 102 ; 66 ; 35 ; - ; 38	6,73E+06	8,93E+06	7,72E+06	2,81E+06	9,85E+05	2,53E+06	3,70	0,002773729
Q9BYK8	PR285_HUMAN	294697	Peroxisomal proliferator-activated receptor A-interacting complex	PRIC285	2	0,79	28 ; - ; - ; - ; - ; -	6,93E+06	7,24E+06	4,61E+06	1,81E+06	2,09E+06	1,35E+06	3,58	0,025985979
B4E2H0	B4E2H0_HUMAN	175162	cDNA FLJ58148, highly similar to Poly (ADP-ribose) polymerase 14 (	0	10	6,47	151 ; 137 ; - ; - ; - ; -	1,29E+07	1,46E+07	9,61E+06	3,87E+06	2,62E+06	4,27E+06	3,44	0,018247459
Q53GA4	PHLA2_HUMAN	17092	Pleckstrin homology-like domain family A member 2	PHLDA2	2	26,97	- ; 56 ; - ; - ; - ; -	6,06E+06	6,68E+06	5,16E+06	2,07E+06	5,33E+05	2,61E+06	3,43	0,007046051
P23511	NFYA_HUMAN	36877	Nuclear transcription factor Y subunit alpha	NFYA	1	5,48	55 ; 54 ; - ; - ; - ; -	2,60E+07	2,42E+07	1,27E+07	6,27E+06	1,05E+07	1,80E+06	3,39	0,049484864
P78504	JAG1_HUMAN	133859	Protein jagged-1 precursor	JAG1	7	5,25	126 ; 30 ; 69 ; - ; - ; -	1,87E+07	2,44E+07	2,04E+07		2,87E+06	9,66E+06	3,38	0,011790526
O15533	TPSN_HUMAN	49492	Tapasin precursor	TAPBP	6	15,24	125 ; 112 ; 87 ; 38 ; 57 ; -	9,63E+07	8,66E+07	1,05E+08	3,20E+07	3,23E+07	2,20E+07	3,34	0,001027166
P05120	PAI2_HUMAN	46611	Plasminogen activator inhibitor 2 precursor	SERPINB2	8	4,82	168 ; 131 ; 165 ; - ; 107 ; 65	7,97E+07	7,95E+07	8,45E+07	3,10E+07	1,21E+07	3,39E+07	3,17	0,011476205
B4E1T5	B4E1T5_HUMAN	48885	cDNA FLJ55764, highly similar to Apolipoprotein-L2	0	14	26,95	317 ; 309 ; 226 ; 133 ; 140 ; 181	2,04E+08	2,10E+08	2,16E+08	6,24E+07	5,53E+07	8,17E+07	3,16	0,000743723
P13747	HLAE_HUMAN	40157	HLA class I histocompatibility antigen, alpha chain E precursor	HLA-E	13	31,01	306 ; 344 ; 364 ; 42 ; 83 ; 87	2,37E+08	2,44E+08	2,33E+08	1,15E+08	4,72E+07	6,44E+07	3,15	0,013700053
Q6P3W7	SCYL2_HUMAN	103709	SCY1-like protein 2	SCYL2	11	2,48	113 ; 130 ; 124 ; 52 ; - ; -	2,72E+07	3,31E+07	2,75E+07	1,47E+07	6,19E+06	7,30E+06	3,12	0,005038803
P28065	PSB9_HUMAN	13900	Proteasome subunit beta type-9 precursor	PSMB9	3	7,63	93 ; 76 ; 83 ; - ; - ; -	9,64E+06	7,36E+06	8,52E+06	2,90E+06	2,42E+06	3,07E+06	3,04	0,008669311
Q9BQE5	APOL2_HUMAN	37092	Apolipoprotein L2	APOL2	12	35,01	319 ; - ; 223 ; 155 ; 155 ; -	1,93E+08	2,07E+08	1,81E+08	5,71E+07	6,67E+07	6,85E+07	3,02	0,000726088
Q8TDB6	DTX3L_HUMAN	83554	E3 ubiquitin-protein ligase DTX3L	DTX3L	6	2,84	87 ; 110 ; 33 ; - ; - ; -	1,26E+07	1,32E+07	7,72E+06	4,00E+06	2,67E+06	4,43E+06	3,02	0,040162668
O95379	TFIP8_HUMAN	23257	Tumor necrosis factor alpha-induced protein 8	TNFAIP8	5	16	114 ; 92 ; 71 ; 121 ; 69 ; 55	2,51E+07	2,64E+07	2,57E+07	7,87E+06	1,07E+07	7,26E+06	2,99	0,00155316
Q8WV24	PHLA1_HUMAN	45016	Pleckstrin homology-like domain family A member 1	PHLDA1	1	2,24	62 ; 56 ; 62 ; - ; - ; -	1,62E+07	1,39E+07	1,45E+07	5,66E+06	5,25E+06	4,40E+06	2,92	0,000954551
P15090	FABP4_HUMAN	14719	Fatty acid-binding protein, adipocyte	FABP4	4	22,73	89 ; 120 ; 86 ; - ; - ; -	2,67E+07	2,62E+07	2,42E+07	8,78E+06	9,07E+06	8,79E+06	2,89	0,001765712
O43795	MYO1B_HUMAN	131985	Myosin-Ib	MYO1B	11	5,55	144 ; 102 ; 127 ; - ; - ; -	3,13E+07	3,22E+07	2,90E+07	1,10E+07	1,20E+07	1,03E+07	2,78	0,000358038
O43353	RIPK2_HUMAN	61195	Receptor-interacting serine/threonine-protein kinase 2	RIPK2	11	6,3	190 ; 172 ; 116 ; 34 ; 28 ; 39	9,11E+07	7,26E+07	8,58E+07	3,30E+07	3,70E+07	2,02E+07	2,77	0,00213873
Q9Y4D1	DAAM1_HUMAN	123473	Disheveled-associated activator of morphogenesis 1	DAAM1	6	1,95	56 ; 25 ; 40 ; - ; - ; -	1,35E+07	1,44E+07	1,04E+07	4,03E+06	5,39E+06	4,46E+06	2,76	0,013892397
O94808	GFPT2_HUMAN	76906	Glucosamine--fructose-6-phosphate aminotransferase [isomerizin	GFPT2	5	9,24	- ; 100 ; - ; - ; - ; -	1,26E+08	1,67E+08	1,61E+08	7,16E+07	4,60E+07	4,73E+07	2,75	0,005343378
P03956	MMP1_HUMAN	53982	Interstitial collagenase precursor	MMP1	16	26,44	251 ; 290 ; 237 ; 41 ; 108 ; 135	1,55E+08	1,37E+08	1,48E+08	5,54E+07	5,31E+07	5,27E+07	2,73	0,002649228
P05121	PAI1_HUMAN	45060	Plasminogen activator inhibitor 1 precursor	SERPINE1	37	52,74	1268 ; 1285 ; 1187 ; 781 ; 959 ;	5,63E+09	4,52E+09	7,49E+09	2,87E+09	1,71E+09	1,88E+09	2,73	0,034823913
C9JI18	C9JI18_HUMAN	39853	Putative uncharacterized protein MAP2K3	MAP2K3	14	22,79	189 ; 185 ; 229 ; 169 ; 161 ; 183	1,30E+08	1,18E+08	1,30E+08	5,90E+07	4,26E+07	3,96E+07	2,68	0,000849109
Q03518	TAP1_HUMAN	80893	Antigen peptide transporter 1	TAP1	9	9,76	246 ; 110 ; 171 ; - ; - ; -	6,63E+07	5,16E+07	5,51E+07	2,17E+07	2,38E+07	1,97E+07	2,65	0,010690214
P35354	PGH2_HUMAN	68656	Prostaglandin G/H synthase 2 precursor	PTGS2	12	22,87	370;359;349;-;-;-	3,73E+08	2,52E+08	3,70E+08	1,38E+08	1,41E+08	9,67E+07	2,64	0,024842981
B6ETL5	B6ETL5_HUMAN	47937	Pannexin 1	PANX1	2	7,29	- ; 57 ; - ; - ; - ; -	1,93E+07	2,17E+07	1,65E+07	5,22E+06	8,36E+06	8,96E+06	2,55	0,004319587
P22830	HEMH_HUMAN	32283	Ferrochelatase, mitochondrial precursor	FECH	5	16,01	89 ; 108 ; 83 ; - ; - ; 75	2,23E+07	2,26E+07	1,88E+07	6,95E+06	6,27E+06	1,18E+07	2,55	0,005349861
P04150	GCR_HUMAN	82904	Glucocorticoid receptor	NR3C1	4	5,46	43 ; 75 ; 99 ; - ; - ; 31	1,09E+07	1,46E+07	1,36E+07	6,80E+06	4,49E+06	4,11E+06	2,54	0,005618045
Q5T7W0	ZN618_HUMAN	104956	Zinc finger protein 618	ZNF618	2	2,41	52 ; - ; - ; - ; - ; -	1,60E+07	1,19E+07	1,03E+07	5,60E+06	4,69E+06	4,80E+06	2,54	0,041473532
Q3B8N2	LEG9B_HUMAN	39518	Galectin-9B	LGALS9B	6	10,99	207 ; 119 ; 145 ; 79 ; 131 ; 131	9,96E+07	9,84E+07	1,07E+08	3,75E+07	4,44E+07	3,84E+07	2,53	7,31629E-05
P29083	T2EA_HUMAN	40118	General transcription factor IIE subunit 1	GTF2E1	2	8,68	- ; - ; 78 ; 50 ; - ; -	1,61E+07	2,02E+07	1,53E+07	7,59E+06	7,61E+06	5,47E+06	2,50	0,010715394
P40306	PSB10_HUMAN	28936	Proteasome subunit beta type-10 precursor	PSMB10	2	7,33	- ; 34 ; - ; - ; 30 ; -	4,27E+07	3,87E+07	2,63E+07	1,30E+07	1,62E+07	1,44E+07	2,46	0,04526375
P09238	MMP10_HUMAN	54151	Stromelysin-2 precursor	MMP10	3	5,88	40 ; 40 ; 40 ; - ; - ; -	1,66E+07	1,68E+07	1,93E+07			7,14E+06	2,46	0,014515931
P30740	ILEU_HUMAN	38685	Leukocyte elastase inhibitor	SERPINB1	6	9,97	105 ; 105 ; 147 ; 55 ; - ; 137	2,70E+07	2,73E+07	2,33E+07	8,33E+06	6,75E+06	1,66E+07	2,45	0,024548567
P26022	PTX3_HUMAN	41976	Pentraxin-related protein PTX3 precursor	PTX3	16	14,7	300 ; 341 ; 254 ; 127 ; 196 ; 141	3,01E+08	2,78E+08	3,15E+08	9,63E+07	1,48E+08	1,24E+08	2,43	0,001098467
Q8WVT3	TTC15_HUMAN	79345	Tetratricopeptide repeat protein 15	TTC15	6	10,2	136 ; 86 ; 145 ; 92 ; - ; -	9,32E+06	1,25E+07	1,16E+07	7,98E+06	2,67E+06	3,26E+06	2,40	0,040410691
B4DGC1	B4DGC1_HUMAN	22241	cDNA FLJ59877, highly similar to LAG1 longevity assurance homolo	0	1	9,74	37 ; - ; - ; - ; - ; -	3,30E+06	3,61E+06	2,43E+06		1,30E+06		2,39	0,008374483
P19838	NFKB1_HUMAN	105369	Nuclear factor NF-kappa-B p105 subunit	NFKB1	11	9,91	296 ; 275 ; 208 ; 61 ; 99 ; 187	5,90E+07	5,09E+07	5,78E+07	2,50E+07	2,53E+07	2,02E+07	2,38	0,000918969
O15321	TM9S1_HUMAN	68861	Transmembrane 9 superfamily member 1 precursor	TM9SF1	2	5,61	19 ; - ; - ; - ; 34 ; -	2,04E+08	2,37E+08	1,98E+08	8,25E+07	8,77E+07	1,01E+08	2,36	0,003749728
O75953	DNJB5_HUMAN	47145	DnaJ homolog subfamily B member 5	DNAJB5	1	4,29	43 ; - ; - ; - ; - ; -	9,15E+06	7,45E+06	1,00E+07	3,83E+06	3,44E+06	4,02E+06	2,36	0,017295279
P21580	TNAP3_HUMAN	89614	Tumor necrosis factor alpha-induced protein 3	TNFAIP3	7	7,59	54 ; 62 ; 43 ; - ; - ; -	1,22E+07	1,95E+07	1,78E+07	6,17E+06	6,67E+06	8,47E+06	2,33	0,040090432
Q5VW38	GP107_HUMAN	61932	Protein GPR107 precursor	GPR107	1	1,99	47 ; 52 ; - ; - ; - ; -	1,91E+07	1,86E+07	2,45E+07			8,96E+06	2,32	0,010999883
Q9Y376	CAB39_HUMAN	39869	Calcium-binding protein 39	CAB39	22	49,56	411 ; 422 ; 356 ; 243 ; 314 ; 336	2,68E+08	2,40E+08	1,96E+08	9,57E+07	1,21E+08	8,76E+07	2,32	0,011699108
Q86SQ9	DHDDS_HUMAN	34849	Dehydrodolichyl diphosphate synthase	DHDDS	1	3,68	49 ; - ; - ; - ; - ; -	3,97E+06	3,78E+06	2,96E+06	1,09E+06		2,14E+06	2,21	0,033807677
Q9BQE4	SELS_HUMAN	21163	Selenoprotein S	SELS	9	17,46	114 ; 58 ; 47 ; 45 ; 40 ; 101	3,28E+07	3,53E+07	2,92E+07	1,76E+07	1,02E+07	1,65E+07	2,20	0,004541841
B4DLD1	B4DLD1_HUMAN	110634	cDNA FLJ57342, highly similar to Laminin gamma-2 chain	0	2	2,67	- ; - ; 32 ; - ; - ; -	2,34E+07	2,36E+07	3,04E+07		9,33E+06	1,44E+07	2,17	0,031413662
Q8N4S0	CCD82_HUMAN	64002	Coiled-coil domain-containing protein 82	CCDC82	1	2,02	35 ; - ; - ; - ; - ; -	1,91E+06	2,72E+06	1,98E+06	1,42E+06	7,86E+05	8,54E+05	2,16	0,025312362
O15446	RPA34_HUMAN	46024	DNA-directed RNA polymerase I subunit RPA34	CD3EAP	4	6,06	71 ; 68 ; 145 ; 61 ; - ; -	1,83E+07	1,54E+07	1,92E+07	1,16E+07	7,84E+06	5,25E+06	2,14	0,018025167
Q9UBV7	B4GT7_HUMAN	37406	Beta-1,4-galactosyltransferase 7	B4GALT7	4	11,01	55 ; 38 ; - ; - ; - ; -	1,34E+07	1,15E+07	1,26E+07	5,92E+06	5,10E+06	6,78E+06	2,11	0,000850505
P40189	IL6RB_HUMAN	103537	Interleukin-6 receptor subunit beta precursor	IL6ST	2	3,27	- ; - ; 40 ; - ; - ; -	2,92E+07	2,19E+07	2,16E+07	8,36E+06	1,59E+07	1,04E+07	2,10	0,019447527
Q9UBS3	DNJB9_HUMAN	25518	DnaJ homolog subfamily B member 9	DNAJB9	3	12,56	91 ; 82 ; - ; - ; - ; 49	8,52E+06	6,55E+06	5,36E+06	2,87E+06	2,44E+06	4,59E+06	2,06	0,04134073
Q8WVP5	TP8L1_HUMAN	20827	Tumor necrosis factor alpha-induced protein 8-like protein 1	TNFAIP8L1	5	11,29	99 ; 80 ; 75 ; 108 ; 105 ; 110	2,21E+07	2,13E+07	2,22E+07	1,16E+07	8,46E+06	1,17E+07	2,06	0,005835414
Q9Y6K5	OAS3_HUMAN	121239	2'-5'-oligoadenylate synthase 3	OAS3	1	1,66	31 ; - ; - ; - ; - ; -	1,17E+07	1,42E+07	1,01E+07	9,94E+06	3,00E+06	4,60E+06	2,06	0,078485456
O60675	MAFK_HUMAN	17523	Transcription factor MafK	MAFK	4	19,87	62 ; 73 ; 47 ; - ; - ; 31	9,80E+06	9,09E+06	8,55E+06	3,89E+06	5,49E+06	3,98E+06	2,06	0,002715541
P02788	TRFL_HUMAN	33025	Lactotransferrin precursor	LTF	2	7	45 ; 58 ; 23 ; 47 ; 39 ; 15	2,49E+08	2,03E+08	2,45E+08	7,85E+07	1,40E+08	1,24E+08	2,03	0,008440959
O95361	TRI16_HUMAN	63971	Tripartite motif-containing protein 16	TRIM16	20	28,72	351 ; 337 ; 329 ; 166 ; 162 ; 174	1,18E+08	1,29E+08	1,11E+08	5,89E+07	5,68E+07	6,07E+07	2,03	0,005372103
P01130	LDLR_HUMAN	92915	Low-density lipoprotein receptor precursor	LDLR	3	3,58	70 ; 27 ; - ; - ; - ; -	1,55E+07	1,54E+07	1,35E+07	5,28E+06	8,26E+06	8,64E+06	2,01	0,006987004
P40337	VHL_HUMAN	15219	Von Hippel-Lindau disease tumor suppressor	VHL	2	17,56	52 ; 45 ; 52 ; - ; - ; 44	2,17E+07	2,47E+07	2,88E+07	1,13E+07	1,27E+07	1,36E+07	2,00	0,017929185
P51808	DYLT3_HUMAN	13062	Dynein light chain Tctex-type 3	DYNLT3	1	8,62	40 ; 50 ; 40 ; - ; - ; -	2,65E+07	3,39E+07	4,51E+07	6,36E+07	6,97E+07	8,60E+07	0,48	0,012749917
Q58WW2	DCAF6_HUMAN	103070	DDB1- and CUL4-associated factor 6	DCAF6	1	1,63	- ; - ; - ; - ; 59 ; -	3,08E+06	3,19E+06	1,51E+06	4,93E+06	6,09E+06	5,80E+06	0,46	0,014014593

Q9BR61	ACBD6_HUMAN	31137	Acyl-CoA-binding domain-containing protein 6	ACBD6	2	6,74	-;-;-;-;15;-	1,93E+06	3,25E+06	3,35E+06	5,90E+06	6,89E+06	6,09E+06	0,45	0,005280926
B5BUB8	B5BUB8_HUMAN	44285	Mitogen-activated protein kinase 8 isoform JNK1 alpha1 (Fragmen	MAPK8	3	8,59	85;52;-;-;-;-	2,76E+07	2,66E+07	2,32E+07	5,80E+07	4,89E+07	6,54E+07	0,45	0,016370621
C9JDE9	C9JDE9_HUMAN	40269	Putative uncharacterized protein ACAA1	ACAA1	2	4,96	-;70;-;-;-;-	1,25E+07	7,60E+06	9,38E+06	2,59E+07	1,93E+07	2,11E+07	0,44	0,00949878
P27986	P85A_HUMAN	83598	Phosphatidylinositol 3-kinase regulatory subunit alpha	PIK3R1	5	8,7	36;-;-;26;39;-	5,43E+06	2,33E+06	3,46E+06	6,99E+06	1,02E+07	8,08E+06	0,44	0,02309642
Q14204	DYHC1_HUMAN	532408	Cytoplasmic dynein 1 heavy chain 1	DYNC1H1	74	10,33	178;372;507;773;712;860	4,24E+08	4,30E+08	7,12E+08	1,24E+09	9,51E+08	1,39E+09	0,44	0,016703971
Q2NKY5	Q2NKY5_HUMAN	50090	TUBB6 protein	TUBB6	6	16,55	-;-;-;-;170;-;-	3,36E+07	3,09E+07	4,31E+07	8,84E+07	6,48E+07	9,46E+07	0,43	0,023336783
P07205	PGK2_HUMAN	44796	Phosphoglycerate kinase 2	PGK2	8	20,14	-;-;-;-;-;288;-	5,27E+08	4,95E+08	6,94E+08	1,63E+09	1,09E+09	1,35E+09	0,42	0,0245754
Q5VZL5	ZMYM4_HUMAN	172788	Zinc finger MYM-type protein 4	ZMYM4	1	0,78	38;-;-;-;-;-	7,37E+06	3,77E+06		1,00E+07	1,84E+07	1,15E+07	0,42	0,047955089
Q8NBF2	NHLC2_HUMAN	79444	NHL repeat-containing protein 2	NHLRC2	10	10,47	149;141;152;106;46;80	1,17E+08	9,91E+07	1,11E+08	2,51E+08	3,04E+08	2,37E+08	0,41	0,012504466
O00400	ACATN_HUMAN	60909	Acetyl-coenzyme A transporter 1	SLC33A1	2	2,37	79;79;38;38;58;49	7,27E+07	6,86E+07	9,96E+07	2,24E+08	2,14E+08	1,44E+08	0,41	0,032672889
Q9BVS4	RIOK2_HUMAN	54508	Serine/threonine-protein kinase RIO2	RIOK2	2	2,53	-;-;-;-;-;45;31	4,92E+06	2,13E+06	2,14E+06	5,78E+06	8,41E+06	8,52E+06	0,40	0,025086492
P62068	UBP46_HUMAN	42442	Ubiquitin carboxyl-terminal hydrolase 46	USP46	1	4,92	-;-;-;-;75;-;-	1,15E+07	1,63E+06	1,18E+07	2,35E+07	1,78E+07	2,11E+07	0,40	0,045743014
Q9NSP4	CENPM_HUMAN	16148	Centromere protein M	CENPM	1	6,85	-;-;-;-;36;-;-	2,47E+06	2,34E+06		4,99E+06	8,99E+06	5,30E+06	0,37	0,043749297
Q9NR33	DPOE4_HUMAN	12225	DNA polymerase epsilon subunit 4	POLE4	1	9,4	-;-;-;-;-;44;54	1,35E+06	1,75E+06		4,16E+06	4,61E+06	3,88E+06	0,37	0,013325548
B7Z8P5	B7Z8P5_HUMAN	100143	cDNA FLJ51438, highly similar to Probable ATP-dependent RNA hel	0	16	10,19	212;153;77;95;216;156	2,62E+07	2,77E+07	2,93E+07	7,82E+07	9,27E+07	6,61E+07	0,35	0,020616147
Q9HBU6	EKI1_HUMAN	50968	Ethanolamine kinase 1	ETNK1	4	6,86	54;-;-;-;46;-	7,94E+07	1,55E+07	3,15E+07	1,09E+08	1,35E+08	1,36E+08	0,33	0,031291735
Q5T8C6	Q5T8C6_HUMAN	54704	Cell division cycle 16 homolog (S. cerevisiae)	CDC16	6	5,89	52;35;48;44;47;58	8,44E+07	3,85E+07	7,58E+07	2,09E+08	2,30E+08	1,92E+08	0,32	0,001640203
Q6N069	NAA16_HUMAN	101462	N-alpha-acetyltransferase 16, NatA auxiliary subunit	NAA16	3	3,13	-;-;-;-;61;-;-	5,16E+06	6,35E+06	6,78E+06	1,50E+07	2,40E+07	1,92E+07	0,31	0,032031683
Q9BU76	MMTA2_HUMAN	29412	Multiple myeloma tumor-associated protein 2	MMTAG2	1	3,8	-;-;-;-;38;45;-	1,37E+06		3,26E+06	7,87E+06	7,92E+06	6,94E+06	0,31	0,014584597
Q9ULE0	WWC3_HUMAN	65586	Protein WWC3	WWC3	1	1,7	-;-;-;-;-;-;46	1,50E+07	1,47E+07		3,25E+07	5,98E+07	5,42E+07	0,30	0,023698206
Q8TAF3	WDR48_HUMAN	76210	WD repeat-containing protein 48	WDR48	2	1,92	-;-;-;-;38;58;52	5,83E+06	2,39E+06	1,61E+06	8,71E+06	1,37E+07	1,07E+07	0,30	0,016862469
B4DM52	B4DM52_HUMAN	94020	DNA ligase	0	3	2,7	92;90;60;77;113;71	1,02E+07	7,07E+06	1,06E+07	3,03E+07	3,31E+07	3,24E+07	0,29	0,000132953
P23229	ITA6_HUMAN	121665	Integrin alpha-6 precursor	ITGA6	18	17,77	223;185;226;276;220;343	6,35E+07	5,92E+07	1,53E+08	4,55E+08	4,14E+08	3,47E+08	0,23	0,002042906
D3DVZ9	D3DVZ9_HUMAN	13188	Chromosome 20 open reading frame 30, isoform CRA_a	C20orf30	2	20,83	-;-;32;81;33;-	1,81E+06	2,00E+06	2,25E+06	1,14E+07	9,32E+06	6,41E+06	0,22	0,039157142
Q6ZTE7	Q6ZTE7_HUMAN	62277	cDNA FLJ44731 fis, clone BRACE3025719, highly similar to Homo sap	0	3	7,42	-;-;-;80;-;-	2,02E+06	5,28E+05	1,23E+06	5,08E+06	7,91E+06	5,88E+06	0,20	0,013318606
Q5JPH6	SYEM_HUMAN	35377	Probable glutamyl-tRNA synthetase, mitochondrial precursor	EARS2	1	3,19	-;-;43;-;40;55	1,00E+05	1,00E+05	1,00E+05	1,74E+06	2,84E+06	2,62E+06	0,04	0,020714774
Q9NYL2	MLTK_HUMAN	91154	Mitogen-activated protein kinase kinase kinase MLT	MLTK	1	1,88	-;-;-;59;-;-	1,00E+05	1,00E+05	1,00E+05	4,51E+06	4,47E+06	4,40E+06	0,02	5,69103E-05
Q7Z570	Z804A_HUMAN	136888	Zinc finger protein 804A	ZNF804A	1	0,74	-;-;39;-;-;-	1,00E+05	1,00E+05	1,00E+05	8,23E+06	9,67E+06	5,36E+06	0,01	0,02637656

Annexe 7 : Table des protéines variantes identifiées après analyse quantitative du protéome total des HUVEC suite à la stimulation IL-33 des HUVEC. Surlignées en bleu : protéines identifiées variantes à la fois après stimulation TNFα-IFNγ, IL-1β et IL-33 ; surlignées en vert : protéines variantes après stimulation IL-1β et IL-33 ; surlignées en jaune : protéines variantes après stimulation TNFα-IFNγ et IL-33. (AC : numéro d'accession, ID : identifiant de la protéine, MW : molecular weight).

IL-33_Stimulation 24h

Protein AC	Protein ID	MW	Protein name	Genename	Number quantified	Coverage (%)	Mascot score CT1; CT2; CT3 ; IL33-1; IL33-2; IL33-3	PAI IL33 1	PAI IL33 2	PAI IL33 3	PAI CT 1	PAI CT 2	PAI CT 3	Ratio	Student
P80162	CXCL6_HUMAN	11897	C-X-C motif chemokine 6	CXCL6	1	10,53	0; 0; 0; 0; 39; 0	9,23E+04	9,85E+04	8,88E+04	1,23E+07	1,61E+07	1,44E+07	153,3642881	0,0060481
Q96CG8	CTHR1_HUMAN	26224	Collagen triple helix repeat-containing protein 1	CTHRC1	2	8,23	0;0;0;98;0;0	1,05E+05	9,94E+04	9,38E+04	1,31E+07	1,34E+07	9,75E+06	121,8077922	0,0094124
P78556	CCL20_HUMAN	10762	C-C motif chemokine 20	CCL20	2	17,71	0;0;0;0;38;30	9,77E+04	1,07E+05	9,29E+04	9,85E+06	1,30E+07	1,14E+07	115,1361325	0,0063915
P16581	LYAM2_HUMAN	66655	E-selectin	SELE	4	9,67	0;0;0;0;0;0	9,23E+04	9,85E+04	8,88E+04	1,10E+07	9,43E+06	8,09E+06	101,9500629	0,0078302
O95236	APOL3_HUMAN	44278	Apolipoprotein L3	APOL3	7	6,22	0;0;0;89;122;29	1,05E+05	9,94E+04	9,38E+04	5,54E+06	6,47E+06	4,68E+06	56,02886722	0,0088554
Q13077	TRAF1_HUMAN	46163	TNF receptor-associated factor 1	TRAF1	2	2,64	0;0;0;45;0;0	1,02E+05	9,59E+04	9,93E+04	4,67E+06	3,34E+06	4,58E+06	42,33165734	0,0107239
O95760	IL33_HUMAN	30759	Interleukin-33	IL33	5	16,67	0;0;0;77;72;141	1,24E+06	1,20E+06	1,23E+06	3,46E+07	4,57E+07	5,85E+07	37,83822629	0,0227639
P52569	CTR2_HUMAN	71673	Low affinity cationic amino acid transporter 2	SLC7A2	7	7,14	0;0;0;144;0;254	7,48E+05	6,91E+05	6,35E+05	2,13E+07	1,96E+07	1,42E+07	26,56545909	0,014522
P05362	ICAM1_HUMAN	57825	Intercellular adhesion molecule 1	ICAM1	20	36,65	51;51;51;925;788;772	1,21E+07	1,65E+07	1,04E+07	3,53E+08	3,40E+08	2,99E+08	25,36014658	0,0023981
Q6ZTI6	F101A_HUMAN	23610	Protein FAM101A	FAM101A	1	5,09	0;0;0;0;0;31	1,05E+05	1,01E+05	9,14E+04	1,97E+06	1,85E+06	2,18E+06	20,16082273	0,0024016
Q14213	IL27B_HUMAN	25396	Interleukin-27 subunit beta	EBI3	1	5,24	0;0;0;100;120;0	1,03E+05	9,45E+04	1,12E+05	2,13E+06	1,75E+06	2,02E+06	19,10237908	0,003594
P10145	IL8_HUMAN	11098	Interleukin-8	IL8	1	16,16	0;0;0;39;40;0	8,55E+04	1,08E+05	1,03E+05	1,86E+06	1,34E+06	1,39E+06	15,46317323	0,0128123
Q96AZ6	ISG20_HUMAN	20363	Interferon-stimulated gene 20 kDa protein	ISG20	5	30,39	69;69;69;84;79;91	1,28E+06	1,07E+06	9,96E+05	1,49E+07	1,88E+07	1,64E+07	14,95781807	0,0049631
P04179	SODM_HUMAN	24722	Superoxide dismutase [Mn], mitochondrial	SOD2	12	10,36	92;92;92;1169;669;1748	7,62E+07	7,29E+07	7,10E+07	9,15E+08	6,89E+08	1,32E+09	13,29153122	0,0396088
P25106	CXCR7_HUMAN	41493	C-X-C chemokine receptor type 7	CXCR7	1	3,87	0;0;0;0;0;0	1,00E+05	9,31E+04	9,67E+04	1,25E+06	9,76E+05	1,46E+06	12,70158236	0,0147323
P19320	VCAM1_HUMAN	81276	Vascular cell adhesion protein 1	VCAM1	15	7,04	0;0;0;77;116;118	2,16E+06	1,54E+06	3,10E+06	2,44E+07	2,17E+07	1,85E+07	9,50280095	0,0051382
Q96CG3	TIFA_HUMAN	21445	TRAF-interacting protein with FHA domain-containing protein	TIFA	4	11,96	0;0;0;52;71;53	1,02E+05	8,96E+04	1,04E+05	9,37E+05	1,07E+06	6,58E+05	9,036255178	0,0226137
O75144	ICOSL_HUMAN	33349	ICOS ligand	ICOSLG	1	3,31	32;32;32;163;143;217	3,41E+06	3,30E+06	3,19E+06	1,81E+07	2,20E+07	2,29E+07	6,349633763	0,0068176
P25774	CATS_HUMAN	37496	Cathepsin S	CTSS	3	4,83	0;0;0;95;174;79	2,56E+06	2,74E+06	4,35E+06	2,11E+07	2,16E+07	1,78E+07	6,260376187	0,0012875
P05120	PAI2_HUMAN	46596	Plasminogen activator inhibitor 2	SERPINB2	8	19,52	0;0;0;141;88;135	1,79E+06	1,69E+06	1,62E+06	8,53E+06	8,46E+06	1,06E+07	5,422269264	0,0085513
Q96RN5	MED15_HUMAN	86753	Mediator of RNA polymerase II transcription subunit 15	MED15	2	1,4	0;0;0;54;66;0	1,03E+05	8,93E+04	9,77E+04	5,17E+05	3,82E+05	4,74E+05	4,729224362	0,011463
P28065	PSB9_HUMAN	23264	Proteasome subunit beta type-9	PSMB9	3	6,85	0;0;0;79;68;55	1,65E+06	1,73E+06	1,71E+06	8,47E+06	8,07E+06	6,74E+06	4,56769713	0,0072738
O43353	RIPK2_HUMAN	61194	Receptor-interacting serine/threonine-protein kinase 2	RIPK2	1	2,41	0;0;0;39;121;0	2,43E+06	3,06E+06	3,44E+06	1,29E+07	1,48E+07	1,16E+07	4,406771999	0,0044189
P49427	UB2R1_HUMAN	26737	Ubiquitin-conjugating enzyme E2 R1	CDC34	1	6,36	0;0;0;58;72;65	4,34E+05	1,20E+06	1,51E+06	3,65E+06	4,66E+06	4,94E+06	4,207632886	0,0030643
Q86UT6	NLRX1_HUMAN	107616	NLR family member X1	NLRX1	1	1,03	0;0;0;32;27;0	1,01E+05	1,16E+05	1,09E+05	3,73E+05	4,89E+05	3,88E+05	3,845417199	0,012607
Q13751	LAMB3_HUMAN	129572	Laminin subunit beta-3	LAMB3	1	2,05	0;0;0;63;0;0	9,84E+04	9,98E+04	1,10E+05	3,90E+05	3,27E+05	4,12E+05	3,66124568	0,0076714
Q96RQ9	OXLA_HUMAN	62881	L-amino-acid oxidase	IL4I1	1	1,76	0;0;0;102;110;101	5,90E+05	5,42E+05	8,51E+05	2,64E+06	1,89E+06	2,61E+06	3,603445962	0,0111252
Q00653	NFKB2_HUMAN	96749	Nuclear factor NF-kappa-B p100 subunit	NFKB2	9	3,44	0;0;0;143;139;94	9,70E+05	1,05E+07	2,36E+06	1,68E+07	1,24E+07	1,89E+07	3,475201067	0,0404158
O00182	LEG9_HUMAN	39518	Galectin-9	LGALS9	4	7,89	0;0;0;70;129;70	3,42E+06	3,74E+06	3,83E+06	1,31E+07	1,33E+07	1,15E+07	3,45285726	0,0031765
Q9UGQ3	GTR6_HUMAN	54539	Solute carrier family 2, facilitated glucose transporter membe	SLC2A6	1	1,97	0;0;0;0;138;0	1,11E+06	1,19E+06	1,15E+06	3,41E+06	3,58E+06	3,94E+06	3,177850879	0,0033901
Q9BXJ8	T120A_HUMAN	40610	Transmembrane protein 120A	TMEM120A	7	6,12	71;71;71;97;234;147	4,96E+06	5,54E+06	4,68E+06	1,36E+07	1,60E+07	1,49E+07	2,932282087	0,0021619
P78504	JAG1_HUMAN	133798	Protein jagged-1	JAG1	1	0,9	0;0;0;0;27;63	8,86E+05	3,13E+06	2,04E+06	5,73E+06	5,68E+06	5,92E+06	2,862295676	0,0272722
P40305	IFI27_HUMAN	11268	Interferon alpha-inducible protein 27, mitochondrial	IFI27	1	10,92	0;0;0;0;0;43	5,34E+05	6,18E+05	4,01E+05	1,55E+06	1,22E+06	1,33E+06	2,638018422	0,0031623
P30483	1B45_HUMAN	40414	HLA class I histocompatibility antigen, B-45 alpha chain	HLA-B	6	21,82	110;110;110;231;0;0	5,32E+07	3,78E+07	2,77E+07	1,13E+08	9,03E+07	1,01E+08	2,566095599	0,0035422
Q9UK39	NOCT_HUMAN	48196	Nocturnin	CCRN4L	1	4,41	0;0;0;0;0;0	1,45E+06	8,65E+05	1,39E+06	3,03E+06	3,90E+06	2,43E+06	2,524636455	0,0322747
P02794	FRIH_HUMAN	21226	Ferritin heavy chain	FTH1	12	21,31	57;57;57;323;285;180	3,55E+07	4,30E+07	4,04E+07	9,01E+07	9,29E+07	1,08E+08	2,447132409	0,0041721
O15533	TPSN_HUMAN	47625	Tapasin	TAPBP	5	10,71	0;0;0;30;57;21	5,93E+06	6,64E+06	6,63E+06	1,53E+07	1,68E+07	1,46E+07	2,43111529	0,0020191
Q03518	TAP1_HUMAN	87218	Antigen peptide transporter 1	TAP1	5	3,09	34;34;34;180;39;82	4,63E+06	6,21E+06	5,40E+06	1,03E+07	1,38E+07	1,31E+07	2,286937413	0,0125605
Q9H7F0	AT133_HUMAN	138043	Probable cation-transporting ATPase 13A3	ATP13A3	5	3,43	0;0;0;104;59;35	5,61E+05	1,38E+06	1,22E+06	1,92E+06	2,67E+06	2,62E+06	2,279124779	0,0181405
P40306	PSB10_HUMAN	28936	Proteasome subunit beta type-10	PSMB10	1	7,33	104;104;104;112;132;24	3,32E+06	4,29E+06	4,20E+06	9,53E+06	7,74E+06	9,00E+06	2,224152331	0,0033704
P16188	1A30_HUMAN	40904	HLA class I histocompatibility antigen, A-30 alpha chain	HLA-A	7	29,04	0;0;0;404;507;0	6,73E+07	5,37E+07	4,18E+07	1,31E+08	1,05E+08	1,22E+08	2,198299388	0,0036733
Q24JP5	T132A_HUMAN	110110	Transmembrane protein 132A	TMEM132A	2	3,03	0;0;0;0;35;40	1,69E+06	1,56E+06	2,18E+06	3,48E+06	4,27E+06	4,02E+06	2,166614451	0,0026101
Q16647	PTGIS_HUMAN	57103	Prostacyclin synthase	PTGIS	1	3,4	0;0;0;59;0;55	6,82E+05	6,38E+05	8,47E+05	1,70E+06	1,42E+06	1,50E+06	2,134430196	0,0019733
P31431	SDC4_HUMAN	21641	Syndecan-4	SDC4	6	5,05	51;51;51;145;147;113	4,53E+06	3,38E+06	3,27E+06	8,17E+06	8,28E+06	7,26E+06	2,120615127	0,0015489
P13747	HLAE_HUMAN	40156	HLA class I histocompatibility antigen, alpha chain E	HLA-E	5	16,2	0;0;0;149;208;245	1,47E+07	1,51E+07	1,82E+07	3,42E+07	3,46E+07	3,28E+07	2,1158804	0,0008077
P48307	TFPI2_HUMAN	26934	Tissue factor pathway inhibitor 2	TFPI2	4	8,51	0;0;0;130;139;171	8,06E+06	8,04E+06	8,36E+06	1,72E+07	1,83E+07	1,61E+07	2,111657726	0,0037491
Q9Y6K5	OAS3_HUMAN	121170	2'-5'-oligoadenylate synthase 3	OAS3	2	1,38	0;0;0;41;0;45	1,90E+06	1,63E+06	2,02E+06	3,95E+06	3,46E+06	4,19E+06	2,090049634	0,0034858
Q03169	TNAP2_HUMAN	72661	Tumor necrosis factor alpha-induced protein 2	TNFAIP2	7	12,54	0;0;0;214;274;164	6,40E+06	6,30E+06	5,98E+06	1,28E+07	1,30E+07	1,28E+07	2,067093512	1,357E-05
Q9BT17	MTG1_HUMAN	37237	Mitochondrial GTPase 1	MTG1	1	5,69	0;0;0;0;0;0	1,73E+06	2,10E+06	2,44E+06	9,23E+05	6,30E+05	9,20E+05	0,39459955	0,0130541

Développement de méthodes quantitatives sans marquage pour l'étude protéomique des cellules endothéliales

Violette GAUTIER

Directeurs de thèse : Bernard Monsarrat, Anne Gonzalez de Peredo

Soutenance : le 18 décembre 2012, Auditorium F. Gallais – 205, route de Narbonne 31077 Toulouse

---

Résumé

La compréhension du fonctionnement des systèmes biologiques, dont les protéines sont les principaux effecteurs, est un défi majeur en biologie. La protéomique est aujourd'hui l'outil incontournable pour l'étude des protéines. Au cours de ma thèse, j'ai donc utilisé différentes approches protéomiques pour répondre à plusieurs questions biologiques autour des cellules endothéliales, concernant l'étude de mécanismes fonctionnels de protéines d'intérêt ainsi que des processus inflammatoires au sein de ces cellules. Ces différentes études ont nécessité la mise en place et l'optimisation de méthodes de quantification sans marquage (« label free ») essentielles à la fois pour la caractérisation de complexes protéiques et pour l'analyse de protéomes entiers. Cette thèse décrit ainsi dans un premier temps l'utilisation de telles approches pour l'analyse de complexes immunopurifiés dans laquelle un enjeu important consiste souvent à discriminer de façon non ambiguë les composants *bona fide* du complexe par rapport aux contaminants non-spécifiques. J'ai ainsi notamment pu identifier certains partenaires spécifiques d'une nouvelle famille de facteurs de transcription humains, les protéines THAP, qui jouent un rôle clé dans la prolifération des cellules endothéliales. Dans un second temps, les processus activés par les cellules endothéliales en condition inflammatoire ont été étudiés au niveau de sous-protéomes ou à l'échelle de protéomes entiers, faisant appel à des méthodes de protéomique globale associées à des stratégies de quantification sans marquage. Le glycoprotéome des cellules endothéliales a ainsi d'une part été étudié lors la réponse inflammatoire, grâce à la mise en place d'une méthode d'enrichissement du protéome de surface des cellules. D'autre part, une analyse du protéome entier de ces cellules et de ses modulations lors de la stimulation par des cytokines pro-inflammatoires a également été réalisée. De façon à obtenir une couverture du protéome la plus profonde possible, cette étude a nécessité la mise en place d'une stratégie quantitative impliquant un fractionnement de l'échantillon sur gel 1D. Enfin, une troisième partie s'intéresse plus spécifiquement aux rôles et aux mécanismes d'action de l'interleukine-33 au sein des cellules endothéliales et a requis l'utilisation des méthodes quantitatives précédemment optimisées.

Abstract

Understanding biological systems, in which proteins are the main effectors, is a major challenge in biology. Proteomics is now an indispensable tool for the study of proteins. During my PhD, I used different proteomic approaches to address several biological questions about endothelial cells for the study of functional mechanisms of proteins of interest as well as inflammatory processes in these cells. These studies involved the development and the optimization of label-free quantitative methods, essential both for the characterization of protein complexes and for the analysis total proteome. This thesis describes first the use of such approaches for the analysis of immunopurified complexes, for which an important issue is often to discriminate unambiguously *bona fide* components of the complex from non-specific proteins. I could identify specific partners of a new family of human transcription factors, the THAP proteins, which play a key role in endothelial cells proliferation. Then, the processes activated in endothelial cells under inflammatory condition were studied at sub-proteome or entire proteome level, using global proteomics strategies associated with label-free quantification. On the one hand, the glycoproteome has been studied under inflammatory conditions, through the establishment of a method for cell surface proteome enrichment. On the other hand, a whole proteome analysis of these cells was performed after stimulation with pro-inflammatory cytokines. To obtain deep proteome coverage, this study required the implementation of a quantitative strategy involving sample fractionation by 1D gel. Finally, the third section focuses specifically on the roles and mechanisms of action of interleukin-33 in endothelial cells, and required the use of quantitative methods previously optimized.

Mots-clés : Protéomique, Spectrométrie de masse, nanoLC-MS/MS, Cellules endothéliales, Complexes protéiques, Protéomes entiers, Glycoprotéomes, Protéines THAP, TFIIH, Interleukine-33

Discipline : Biochimie

Laboratoire : Institut de Pharmacologie et de Biologie Structurale, CNRS UMR 5089 – 205, route de Narbonne 31077 Toulouse

IL-33_Stimulation 6h

ID	MW	AC	Protein name	Gene name	Number quantified peptides	Coverage (%)	PAI CT_A	PAI CT_B	PAI CT_C	PAI 6h_A	PAI 6h_B	PAI 6h_C	ratio PAI 6h	student 6h
LYAM2_HUMAN	66655	P16581	E-selectin	SELE	4	9,67	9,23E+04	9,85E+04	8,88E+04	6,77E+07	7,96E+07	5,61E+07	790,7	0,05136
CCL20_HUMAN	10762	P78556	C-C motif chemokine 20	CCL20	2	17,71	9,77E+04	1,07E+05	9,29E+04	7,15E+06	7,43E+06	5,57E+06	73,55	0,01225
TRAF1_HUMAN	46163	Q13077	TNF receptor-associated factor 1	TRAF1	2	2,64	1,02E+05	9,59E+04	9,93E+04	5,95E+06	6,31E+06	5,65E+06	61,82	0,01933
IL33_HUMAN	30759	O95760	Interleukin-33	IL33	5	16,67	1,24E+06	1,20E+06	1,23E+06	7,53E+07	5,07E+07	4,63E+07	51,51	0,12517
CXCL6_HUMAN	11897	P80162	C-X-C motif chemokine 6	CXCL6	1	10,53	9,23E+04	9,85E+04	8,88E+04	4,16E+06	3,25E+06	3,91E+06	39,77	0,08012
CTHR1_HUMAN	26224	Q96CG8	Collagen triple helix repeat-containing protein 1	CTHRC1	2	8,23	1,05E+05	9,94E+04	9,38E+04	3,87E+06	3,34E+06	3,45E+06	36,32	0,04737
CXCR7_HUMAN	41493	P25106	C-X-C chemokine receptor type 7	CXCR7	1	3,87	1,00E+05	9,31E+04	9,67E+04	1,51E+06	3,14E+06	3,25E+06	24,1	0,22274
CTR2_HUMAN	71673	P52569	Low affinity cationic amino acid transporter 2	SLC7A2	7	7,14	7,48E+05	6,91E+05	6,35E+05	1,67E+07	1,93E+07	1,96E+07	26,06	0,04845
VCAM1_HUMAN	81276	P19320	Vascular cell adhesion protein 1	VCAM1	15	7,04	2,16E+06	1,54E+06	3,10E+06	5,36E+07	6,13E+07	6,04E+07	25,38	0,04157
DESP_HUMAN	331774	P15924	Desmoplakin	DSP	2	1,25	1,00E+05	9,63E+04	8,46E+04	2,92E+06	1,58E+06	2,23E+06	24,02	0,19216
APOL3_HUMAN	44278	O95236	Apolipoprotein L3	APOL3	7	6,22	1,05E+05	9,94E+04	9,38E+04	1,58E+06	1,84E+06	2,67E+06	17,25	0,05174
F101A_HUMAN	23610	Q6ZTI6	Protein FAM101A	FAM101A	1	5,09	1,05E+05	1,01E+05	9,14E+04	1,58E+06	2,00E+06	2,16E+06	18,06	0,07961
ICAM1_HUMAN	57825	P05362	Intercellular adhesion molecule 1	ICAM1	20	36,65	1,21E+07	1,65E+07	1,04E+07	1,95E+08	2,30E+08	2,02E+08	16,32	0,05451
COX1_HUMAN	57041	P00395	Cytochrome c oxidase subunit 1	MT-CO1	1	6,04	9,99E+04	1,02E+05	1,03E+05	1,27E+06	8,00E+05	1,28E+06	10,17	0,15569
TIFA_HUMAN	21445	Q96CG3	TRAF-interacting protein with FHA domain-containing protein	TIFA	4	11,96	1,02E+05	8,96E+04	1,04E+05	1,04E+06	6,97E+05	9,73E+05	8,854	0,14026
IL27B_HUMAN	25396	Q14213	Interleukin-27 subunit beta	EBI3	1	5,24	1,03E+05	9,45E+04	1,12E+05	6,95E+05	5,01E+05	5,62E+05	5,811	0,12265
ISG20_HUMAN	20363	Q96AZ6	Interferon-stimulated gene 20 kDa protein	ISG20	5	30,39	1,28E+06	1,07E+06	9,96E+05	4,64E+06	5,14E+06	4,95E+06	4,387	0,02507
SDC4_HUMAN	21641	P31431	Syndecan-4	SDC4	6	5,05	4,53E+06	3,38E+06	3,27E+06	1,37E+07	1,50E+07	1,50E+07	3,859	0,00808
SODM_HUMAN	24722	P04179	Superoxide dismutase [Mn], mitochondrial	SOD2	12	10,36	7,62E+07	7,29E+07	7,10E+07	2,73E+08	2,58E+08	2,70E+08	3,619	0,02125
NLRX1_HUMAN	107616	Q86UT6	NLR family member X1	NLRX1	1	1,03	1,01E+05	1,16E+05	1,09E+05	3,63E+05	3,53E+05	3,49E+05	3,307	0,00017
TNAP2_HUMAN	72661	Q03169	Tumor necrosis factor alpha-induced protein 2	TNFAIP2	7	12,54	6,40E+06	6,30E+06	5,98E+06	1,79E+07	1,66E+07	1,62E+07	2,764	0,03188
GPX3_HUMAN	25552	P22352	Glutathione peroxidase 3	GPX3	2	5,31	3,12E+06	3,19E+06	2,68E+06	6,65E+06	7,52E+06	8,95E+06	2,363	0,04243
PAI2_HUMAN	46596	P05120	Plasminogen activator inhibitor 2	SERPINB2	8	19,52	1,79E+06	1,69E+06	1,62E+06	5,00E+06	4,03E+06	3,88E+06	2,66	0,10573
PO2F2_HUMAN	51209	P09086	POU domain, class 2, transcription factor 2	POU2F2	1	1,88	1,01E+06	7,73E+05	8,58E+05	1,77E+06	1,98E+06	2,62E+06	2,135	0,01759
NOCT_HUMAN	48196	Q9UK39	Nocturnin	CCRN4L	1	4,41	1,45E+06	8,65E+05	1,39E+06	3,26E+06	2,37E+06	3,00E+06	2,281	0,13435
SFR1_HUMAN	28262	Q86XK3	Swi5-dependent recombination DNA repair protein 1 homolog	MEIR5	1	4,08	1,79E+07	2,53E+07	2,18E+07	1,20E+07	1,04E+07	6,19E+06	0,518	0,0293
S4A8_HUMAN	122938	Q2Y0W8	Electroneutral sodium bicarbonate exchanger 1	SLC4A8	2	2,47	3,45E+06	3,09E+06	3,71E+06	1,39E+06	9,14E+05	1,70E+06	0,337	0,01431
TSN6_HUMAN	27563	O43657	Tetraspanin-6	TSPAN6	1	5,31	2,81E+06	2,36E+06	2,91E+06	1,00E+05	1,06E+05	1,01E+05	0,038	0,00427
PP1RB_HUMAN	13953	O60927	Protein phosphatase 1 regulatory subunit 11	PPP1R11	1	19,84	3,19E+06	2,42E+06	3,75E+06	1,10E+05	1,05E+05	9,69E+04	0,035	0,01591
MYCBP_HUMAN	11967	Q99417	C-Myc-binding protein	MYCBP	1	8,74	3,63E+06	3,90E+06	2,85E+06	9,50E+04	1,10E+05	1,03E+05	0,03	0,00865
WASF1_HUMAN	61652	Q92558	Wiskott-Aldrich syndrome protein family member 1	WASF1	2	4,65	6,63E+06	7,39E+06	4,92E+06	1,07E+05	1,07E+05	9,60E+04	0,017	0,01356