



HAL
open science

Développements méthodologiques en analyse protéomique pour la découverte et la validation de biomarqueurs dans les hémopathies lymphoïdes B de l'adulte

Luc-Matthieu Fornecker

► **To cite this version:**

Luc-Matthieu Fornecker. Développements méthodologiques en analyse protéomique pour la découverte et la validation de biomarqueurs dans les hémopathies lymphoïdes B de l'adulte. Chimie analytique. Université de Strasbourg, 2016. Français. NNT : 2016STRAF015 . tel-01534680

HAL Id: tel-01534680

<https://theses.hal.science/tel-01534680>

Submitted on 8 Jun 2017

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

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



UNIVERSITÉ DE STRASBOURG

EDSC
Ecole Doctorale des
Sciences Chimiques

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

UMR7178 - IPHC

THÈSE

Présentée par :

Luc-Matthieu FORNECKER

Soutenue le **06 juin 2016**

pour obtenir le grade de

Docteur de l'Université de Strasbourg

Discipline / Spécialité : Chimie Analytique

Développements méthodologiques en analyse protéomique pour la découverte et la validation de biomarqueurs dans les hémopathies lymphoïdes B de l'adulte

THÈSE dirigée par :

Dr Sarah CIANFERANI

Dr Alain VAN DORSSELAER

Directeur de Recherche, CNRS, Université de Strasbourg

Directeur de Recherche, CNRS, Université de Strasbourg

RAPPORTEURS :

Pr Véronique LEBLOND

Dr Charles PINEAU

Professeur des Universités, Université Pierre et Marie Curie,
(Paris 6)

Directeur de Recherche, INSERM, Université de Rennes

EXAMINATEURS :

Pr Catherine THIEBLEMONT

Dr Thierry RABILLOUD

Professeur des Universités, Université Paris Diderot (Paris 7)

Directeur de Recherche, CNRS, Université de Grenoble

*« La théorie, c'est quand on sait tout et que rien ne fonctionne.
La pratique, c'est quand tout fonctionne et que personne ne sait pourquoi.
Ici, nous avons réuni théorie et pratique : rien ne fonctionne... et personne ne sait pourquoi ! »*

Albert Einstein

● Remerciements

Je tiens tout d'abord à remercier le Dr Sarah CIANFERANI et le Dr Alain VAN DORSSELAER pour m'avoir accueilli dans le laboratoire et pour la confiance qu'ils m'ont accordée. Vous m'avez permis de réaliser ce travail de thèse dans des conditions exceptionnelles. Merci pour votre patience et votre disponibilité tout au long de ces quatre années.

Je remercie le Pr Véronique LEBLOND, le Pr Catherine THIEBLEMONT, le Dr Charles PINEAU et le Dr Thierry RABILLOUD d'avoir accepté de juger ce travail. Veuillez trouver ici l'expression de ma respectueuse gratitude.

Je remercie le Pr Raoul HERBRECHT pour ses conseils et son soutien permanent. Merci de me guider avec bienveillance depuis de nombreuses années dans mon parcours professionnel. Qu'il me soit permis, au travers de ce travail, de vous exprimer ma plus profonde reconnaissance. Je remercie également toute l'équipe du service d'Hématologie, au sein de laquelle j'ai l'honneur de travailler, pour m'avoir donné la possibilité de réaliser ce travail.

Tout cela n'aurait pas été possible sans l'aide inestimable de Sarah LENNON et Leslie MULLER à qui j'adresse mes plus sincères remerciements. J'ai eu le plaisir et le privilège de pouvoir travailler à vos côtés pendant ces quatre années. Merci pour votre gentillesse, votre patience et votre disponibilité. Merci d'avoir pris le temps de partager avec moi votre savoir et vos compétences. De ces expériences professionnelles sont nées de réelles amitiés.

J'adresse également mes plus sincères remerciements à Christine CARAPITO. Merci pour ta patience, ta bonne humeur, ta disponibilité et pour tout ce que tu m'as appris pendant ces quelques années (et pour tout ce que tu m'apprendras encore !).

Je remercie le Pr Laurent MAUVIEUX et le Dr Laurent MIGUET pour leur aide précieuse dans la réalisation des projets et les conseils prodigués.

Je n'oublie pas bien sûr mes voisins de bureau—Charlotte, Magali, Alisson, Sebastian et Georg. Merci pour votre bonne humeur, pour votre aide et pour toutes les discussions que nous avons eues. Merci à Fabrice & Fabrice pour les multiples revues de projet, toutes plus passionnantes les unes que les autres...

Merci à l'ensemble des membres du laboratoire pour votre disponibilité : Laurence SABATIER, Danièle THIERSE, Christine SCHAEFFER, Jean-Marc STRUB, AGNES HOVASSE, François DELALANDE, Hélène DIEMER, Alexandre BUREL, Patrick GUTERL, Véronique DELVAL, Stella SUZANNE-OCHSENBEIN, Aymen ROMDHANI, Alfred PERRAUT.

Je n'oublie pas les plus « jeunes » avec qui j'ai eu le plaisir de partager ces quelques années : Nina, Aurélie, Margaux, Gauthier, Marianne, Maxime B., Maxime E., Pauline, Guillaume, Johann, Justine, Benoit, Thomas, Stéphane, Joanna, Paola, Blandine, Anthony, Maurane.

Et les plus anciens, déjà partis : Marine, Diego, Gilles, Jean-Michel, Amandine, François.

J'espère n'avoir oublié personne. Merci à toutes et à tous !

J'adresse enfin un immense merci à ma famille. Merci à Sandra pour m'accompagner depuis de nombreuses années et pour son soutien indéfectible sans lequel rien ne serait possible.

● Plan détaillé

ABREVIATIONS PRINCIPALES.....	11
INTRODUCTION GENERALE	13
PARTIE I : INTRODUCTION BIBLIOGRAPHIQUE	15
● CHAPITRE 1 LES HEMOPATHIES LYMPHOÏDES B : QUELS ENJEUX A L'HEURE ACTUELLE ?.....	17
1. <i>Classification des hémopathies lymphoïdes B et données épidémiologiques.....</i>	<i>18</i>
2. <i>La leucémie lymphoïde chronique</i>	<i>20</i>
2.1. Diagnostique et classification	20
2.2. Anomalies cytogénétiques	21
2.3. Statut mutationnel du gène codant pour la séquence variable des chaînes lourdes d'immunoglobulines	22
2.4. Séquençage de nouvelle génération et mutations génétiques récurrentes.....	24
2.5. Conclusion	25
3. <i>Lymphomes B diffus à grandes cellules.....</i>	<i>25</i>
4. <i>Lymphomes de la zone marginale.....</i>	<i>28</i>
5. <i>Conclusion.....</i>	<i>29</i>
● CHAPITRE 2 LA SPECTROMETRIE DE MASSE APPLIQUEE A LA PROTEOMIQUE POUR L'IDENTIFICATION ET LA	
VALIDATION DE BIOMARQUEURS.....	31
1. <i>Les progrès technologiques récents dans le domaine de la protéomique quantitative.....</i>	<i>32</i>
2. <i>Les différentes approches quantitatives par spectrométrie de masse pour la recherche et la</i>	
<i>validation de biomarqueurs.....</i>	<i>35</i>
2.1. Approches quantitatives globales pour la phase de découverte.....	35
2.1.1. Procédures expérimentales.....	35
2.1.1.1. Approches quantitatives avec marquages	36
2.1.1.2. Approches quantitatives sans marquage	37
2.1.1.2.1. Quantification par comptage de peptides ou de spectres.....	38
2.1.1.2.2. Quantification par extraction des courants d'ions (XIC).....	39
2.1.2. Instrumentations	40
2.1.2.1. Spectromètres hybrides quadripôle-Orbitrap	40
2.1.2.2. Spectromètres hybrides quadripôle-TOF (Q-TOF).....	41
2.1.3. Traitement des données.....	41
2.1.3.1. Recherche dans les banques de données.....	41
2.1.3.2. Logiciels pour la quantification par extraction des courants d'ions	43

2.2.	Approches ciblées pour la validation.....	44
2.2.1.	Procédures expérimentales.....	45
2.2.1.1.	Choix des peptides protéotypiques et des transitions	46
2.2.1.2.	Utilisation de peptides synthétiques marqués	47
2.2.2.	Instrumentations.....	48
2.2.3.	Traitement des données.....	49
3.	<i>Conclusion</i>	50
PARTIE II : PROJETS DE RECHERCHES ET RESULTATS OBTENUS		51
•	INTRODUCTION	53
•	CHAPITRE 1 IDENTIFICATION ET VALIDATION DES CD148 ET CD180 COMME MARQUEURS PHENOTYPIQUES DANS LES SYNDROMES LYMPHOPROLIFERATIFS B	55
1.	<i>Contexte scientifique et médical</i>	55
2.	<i>Etapas de découverte</i>	58
2.1.	Génération des microparticules induites et validation en tant que sous-protéome enrichi en protéines membranaires.....	58
2.2.	Identification du CD148.....	59
2.3.	Identification du CD180.....	61
3.	<i>Etapas de validations</i>	64
3.1.	Validation du CD148.....	64
3.2.	Validation du CD180.....	66
4.	<i>Conclusions</i>	68
5.	<i>Communications des résultats</i>	69
•	CHAPITRE 2 DEVELOPPEMENT DE STRATEGIES EN PROTEOMIQUE CIBLEE POUR UNE MEILLEURE CARACTERISATION DES SYNDROMES LYMPHOPROLIFERATIFS	71
1.	<i>Contexte médical et scientifique</i>	71
2.	<i>Mise au point de la procédure expérimentale</i>	72
2.1.	Préparation des échantillons.....	73
2.2.	Développement de la méthode LC-SRM	73
2.2.1.	Choix des peptides protéotypiques	73
2.2.2.	Validation des peptides marqués et choix des transitions	74
2.2.3.	Paramétrages de la méthode LC-SRM	75
3.	<i>Evaluation de la reproductibilité et résultats obtenus</i>	75
4.	<i>Constitution de la cohorte de patients</i>	78
•	CHAPITRE 3 RECHERCHE DE BIOMARQUEURS PRONOSTIQUES DU RISQUE DE RECHUTE NEURO-MENINGEE DANS LES LYMPHOMES B DIFFUS A GRANDES CELLULES : PROJET PROTELL.....	79
1.	<i>Contexte scientifique et médical</i>	79

2.	<i>Etapes préparatoires et stratégie analytique finale mise en place</i>	81
2.1.	Préparation des échantillons.....	81
2.1.1.	Collecte des MPPC.....	82
2.1.2.	Extraction et séparation des protéines.....	82
2.2.	Suivi des performances du couplage nanoLC-MS/MS	84
3.	<i>Résultats</i>	86
3.1.	Constitution des cohortes	86
3.2.	Suivi des performances du couplage nanoLC-MS/MS	87
3.3.	Analyse des données dans le cadre du projet PROTELL	87
4.	<i>Conclusions</i>	91
5.	<i>Communication des résultats</i>	92
•	CHAPITRE 4 VERS L'ANALYSE D'UN GRAND NOMBRE D'ECHANTILLONS : OPTIMISATION DES METHODES DE PREPARATION PAR UNE APPROCHE « TUBE-GELS »	93
1.	<i>Contexte scientifique</i>	93
1.1.	Préparation des échantillons : principes généraux.....	93
1.1.1.	Multiplicité de protocoles d'extraction des protéines et de digestion trypsique.....	93
1.1.2.	Influence du protocole de préparation sur les modifications des protéines et des peptides	95
1.2.	Méthode de préparation de type « tube-gel »	95
2.	<i>Procédure expérimentale</i>	96
3.	<i>Résultats</i>	98
3.1.	Identifications protéiques et peptidiques	98
3.1.1.	Identifications au niveau protéique.....	98
3.1.2.	Identifications au niveau peptidique.....	103
3.2.	Analyses quantitatives.....	104
3.2.1.	Linéarité de quantification	105
3.2.1.1.	Linéarité de quantification au niveau peptidique	105
3.2.1.2.	Linéarité de quantification au niveau protéique	106
3.2.2.	Evaluation de la reproductibilité	108
3.2.3.	Détection des protéines variantes.....	109
4.	<i>Conclusions</i>	113
5.	<i>Communication des résultats</i>	113
•	CHAPITRE 5 RECHERCHE DE BIOMARQUEURS PREDICTIFS DE LA REPONSE AU TRAITEMENT DANS LES LYMPHOMES B DIFFUS A GRANDES CELLULES	115
1.	<i>Contexte scientifique et médical</i>	115
2.	<i>Procédure expérimentale</i>	116
2.1.	Constitution des cohortes	116
2.2.	Extraction des protéines.....	117

2.3.	Analyses nanoLC-MS/MS et traitement des données	118
3.	Résultats	119
3.1.	Protéines uniquement quantifiées dans au moins quatre échantillons du groupe « sensible »	121
3.2.	Protéines uniquement quantifiées dans au moins quatre échantillons du groupe « réfractaire »	123
3.3.	Protéines quantifiées dans les deux groupes de patients	124
3.3.1.	Protéines surexprimées dans le groupe « réfractaire »	125
3.3.1.1.	Fibrinogène	126
3.3.1.2.	P-Rex1 (Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein)	127
3.3.1.3.	PLOD3 (Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3)	128
3.3.1.4.	Tensin-3	128
3.3.2.	Protéines surexprimées dans le groupe « sensible »	129
3.3.2.1.	CD20	129
3.3.2.2.	SYK (Spleen tyrosine kinase)	131
3.3.2.3.	LIMA1 (LIM domain and actin-binding protein 1)	131
4.	Conclusions et perspectives	132
PARTIE III : DISCUSSION ET SYNTHÈSE		135
1.	<i>Problématique générale des biomarqueurs</i>	<i>137</i>
2.	<i>Recherche de biomarqueurs par analyse protéomique : succès et difficultés rencontrées dans les différents projets</i>	<i>142</i>
3.	<i>Perspectives : protéogénomique et onco-protéogénomique</i>	<i>145</i>
CONCLUSION GÉNÉRALE		147
REFERENCES		151
PROCÉDURES EXPÉRIMENTALES		161
ANNEXES		165

● Abréviations principales

ABC	Activated B-Cells
ACN	Acétonitrile
BSA	Albumine de sérum bovin
CD	Cluster de Différenciation
CV	Coefficient de Variation
DDA	Data Dependant Acquisition
DL	Digestion Liquide
ESI	Electrospray Ionisation
FDR	False Discovery Rate
GC	Germinal Center
HPLC	High Performance Liquid Chromatography
IPI	Index Pronostique International
LBDGC	Lymphome B Diffus à Grandes Cellules
LC	Liquid Chromatography
LCM	Lymphome à Cellules du Manteau
LLC	Leucémie Lymphoïde Chronique
LZM	Lymphome de la Zone Marginale
m/z	Rapport masse sur charge
MP	Microparticules
MPPC	Microparticules Plasmatiques Circulantes
OMS	Organisation Mondiale de la Santé
PAGE	Polyacrylamide Gel Electrophoresis
Q	Quadripôle
SDS	Sodium Dodecyl Sulfate
SNC	Système Nerveux Central
SRM	Selected Reaction Monitoring
TG	Tuge-Gel
TOF	Time Of Flight
UPS	Universal Proteomics Standard
XIC	eXtracted Ion Chromatogram

● Introduction générale

La notion de biomarqueur renvoie à la définition très large d'une « caractéristique objectivement mesurée et évaluée comme un indicateur de processus biologiques normaux ou pathologiques, ou d'une réponse pharmacologique à une intervention thérapeutique ».¹ Cette définition explique l'utilisation des biomarqueurs dans des domaines aussi variés que la recherche fondamentale, l'industrie pharmaceutique ou la pratique médicale. Dans le domaine de la cancérologie, la recherche de nouveaux biomarqueurs est un enjeu crucial pour mieux comprendre la physiopathologie de ces maladies, améliorer le dépistage ou le diagnostic et, bien entendu, la prise en charge des patients.

L'essor considérable à l'heure actuelle des technologies « omiques » à haut débit (principalement génomique, transcriptomique et protéomique) génère depuis quelques années une multitude de biomarqueurs potentiels dans le domaine de la cancérologie. Les approches protéomiques par spectrométrie de masse ont pleinement bénéficié de progrès technologiques récents majeurs qui permettent d'étudier le protéome d'un point de vue qualitatif ou quantitatif avec des performances inégalées jusqu'à présent.

Les stratégies appliquées dans la recherche de biomarqueurs doivent néanmoins être optimisées puisque seule une minorité d'entre eux finit par être concrètement utilisée en pratique clinique. Ces stratégies de développement reposent sur deux étapes distinctes :

- une première étape de découverte basée sur des approches globales analysant simultanément un grand nombre de variables sur un nombre limité d'échantillons,
- une étape de validation, nécessaire pour confirmer les résultats à grande échelle sur des cohortes indépendantes de patients et basée idéalement sur une méthodologie différente de celle utilisée dans la phase de découverte.

Mon travail de thèse, centré sur la problématique des hémopathies lymphoïdes B, s'inscrit dans ce contexte avec pour objectif le développement de méthodologies en analyse protéomique pour la découverte et la validation de nouveaux biomarqueurs. Il s'articule dans la continuité des travaux initiés depuis 2004 sur cette thématique au Laboratoire de Spectrométrie de Masse Bio-Organique de Strasbourg (LSMBO, Institut Pluridisciplinaire Hubert Curien, UMR7178) en collaboration avec le Pr Laurent MAUVIEUX du Laboratoire d'Hématologie des Hôpitaux Universitaires de Strasbourg.

Ce mémoire de thèse est composé de trois parties :

- La première partie a pour objectif de souligner l'hétérogénéité clinique et biologique des hémopathies lymphoïdes. L'intérêt d'identifier de nouveaux biomarqueurs pour mieux individualiser des sous-groupes de patients aux profils évolutifs différents et pour améliorer les critères diagnostiques des hémopathies est illustré à travers trois exemples : la leucémie lymphoïde chronique, le lymphome B diffus à grandes cellules et le lymphome de la zone marginale. Les différentes approches protéomiques quantitatives utilisées dans ce travail de thèse et développées grâce aux progrès technologiques rapides de ces dernières années, seront également abordées dans cette partie.
- La deuxième partie présente les cinq projets de recherche auxquels j'ai participé et les résultats obtenus :
 - identifications et validations des CD148 et CD180 comme marqueurs phénotypiques pour améliorer le diagnostic des hémopathies lymphoïdes B chroniques,
 - développement de stratégies en protéomique quantitative ciblée pour la validation de biomarqueurs,
 - recherche de biomarqueurs pronostiques du risque de rechute neuro-méningée dans les lymphomes B diffus à grandes cellules,
 - optimisation des méthodes de préparation des échantillons par une approche « tube-gel »,
 - recherche de biomarqueurs prédictifs de la réponse au traitement dans les lymphomes B diffus à grandes cellules.
- La troisième partie aborde la problématique générale des biomarqueurs. Les principales raisons expliquant le paradoxe actuel entre le nombre croissant de biomarqueurs proposés et la très faible proportion utilisée au final en pratique clinique sont discutées. Cette dernière partie permet également de présenter les enseignements pouvant être tirés des projets développés dans mon travail de thèse et qui pourront être mis à profit pour des travaux futurs, ainsi que les nombreuses perspectives qu'offre aujourd'hui l'analyse protéomique par spectrométrie de masse dans la recherche de biomarqueurs.

Partie I

Introduction bibliographique

• Chapitre 1

Les hémopathies lymphoïdes B : quels enjeux à l'heure actuelle ?

Les hémopathies lymphoïdes B représentent un groupe hétérogène de maladies malignes développées à partir des lymphocytes B. Les principales entités se distinguent par des caractéristiques cliniques, morphologiques et biologiques qui sont à la base de leur classification. Ces données évoluent constamment grâce à la recherche fondamentale, la recherche clinique et les progrès technologiques. Cette amélioration des connaissances permet une meilleure compréhension de la physiopathologie de ces maladies, la mise à disposition de nouveaux outils à visée diagnostique ou pronostique et le développement de nouveaux traitements.

Les études génomiques et transcriptomiques occupent actuellement une place centrale. Le développement des techniques de séquençage de nouvelle génération a permis de mettre en évidence des mutations récurrentes de gènes impliqués dans différents processus de la cancérogénèse. Ces travaux ont permis de mieux caractériser les hémopathies lymphoïdes B sur le plan moléculaire, et plus largement l'ensemble des pathologies cancéreuses. La mise en évidence de dérégulations de certaines voies de signalisation permet actuellement le développement rapide de thérapies ciblées.

Cependant, l'utilisation en pratique clinique quotidienne des données issues des approches génomiques et transcriptomiques est souvent difficile. Deux raisons principales peuvent être évoquées :

- L'absence de disponibilité des tests moléculaires développés en dehors des laboratoires de recherche.
- L'absence d'impact réel dans certains cas sur la prise en charge thérapeutique du patient. Ce point, longtemps valable pour la plupart des chimiothérapies conventionnelles, est remis en cause actuellement avec l'arrivée de nouvelles molécules ciblant certaines voies de signalisation, ou le développement de nouveaux anticorps monoclonaux.

Néanmoins, d'importants progrès restent encore à réaliser pour permettre :

- d'obtenir une caractérisation plus précise des différentes entités (identification de biomarqueurs à visée diagnostique),
- d'individualiser précocement les patients à haut risque pour pouvoir, idéalement, adapter rapidement les stratégies thérapeutiques (identification de biomarqueurs à visée pronostique ou prédictive de la réponse au traitement),
- de mettre en évidence de nouvelles cibles thérapeutiques.

Après avoir rappelé les principes de la classification actuelle des hémopathies lymphoïdes B, trois entités seront développées plus en détail : la leucémie lymphoïde chronique (LLC), le lymphome B diffus à grandes cellules (LBDGC) et le lymphome de la zone marginale (LZM). Ces trois exemples illustrent l'hétérogénéité des hémopathies lymphoïdes B, les progrès accomplis sur le plan de leur caractérisation et le besoin actuel d'identifier de nouveaux biomarqueurs à visée diagnostique ou pronostique.

1. CLASSIFICATION DES HEMOPATHIES LYMPHOÏDES B ET DONNEES EPIDEMIOLOGIQUES

L'absence de consensus international a longtemps représenté un obstacle important pour la coopération des communautés scientifiques et médicales. Ce n'est qu'à partir des années 1990 qu'ont émergé les principes d'une classification internationale des hémopathies lymphoïdes.^{2,3} Ces travaux ont abouti en 2001 à la classification des hémopathies malignes établie par l'Organisation Mondiale de la Santé (OMS).⁴ La version publiée en 2008 a été utilisée dans ce travail de thèse.⁵ Une actualisation de la classification des hémopathies lymphoïdes vient d'être publiée très récemment en 2016.⁶ Elle intègre les dernières données issues des travaux génomiques et transcriptomiques pour préciser les critères diagnostiques et permettre de mieux définir les différentes sous-entités.

La première stratification des hémopathies malignes se base sur la lignée cellulaire concernée (lignée lymphoïde ou lignée myéloïde). Au sein de chaque lignée sont distinguées les hémopathies développées à partir de précurseurs immatures (telles que les leucémies aiguës myéloblastiques ou lymphoblastiques) ou à partir de cellules dites matures. Le tableau 1 expose les différentes entités constituant le groupe des hémopathies lymphoïdes B matures dans la classification OMS 2008 qui a été utilisée dans ce travail de thèse. Ces différentes entités sont définies par une combinaison de critères morphologiques, phénotypiques, génétiques, moléculaires et cliniques.

Les termes utilisés pour désigner ces maladies peuvent prendre en compte :

- la cellule d'origine à partir de laquelle est présumée se développer l'hémopathie : lymphomes développés à partir des cellules de la zone marginale ou de la zone du manteau des follicules lymphoïdes,
- les caractéristiques morphologiques de la prolifération tumorale : préservation des structures folliculaires des ganglions lymphatiques ou, au contraire, destruction de l'architecture ganglionnaire avec envahissement diffus par des cellules lymphomateuses de grande taille,
- l'aspect cytologique des cellules : dans la leucémie à tricholeucocytes les cellules présentent de multiples villosités cytoplasmiques leur donnant un aspect « chevelu »,
- la présentation clinico-biologique : le terme de « leucémie » est par exemple employé en cas d'atteinte médullo-sanguine prédominante. Les lymphomes spléniques de la zone marginale se caractérisent quant à eux par une augmentation importante du volume de la rate liée à l'infiltration tumorale.

Leucémie lymphoïde chronique / lymphome lymphocytaire
Leucémie prolymphocytaire B
Lymphome splénique de la zone marginale
Leucémie à tricholeucocytes
Lymphomes spléniques inclassables (lymphome diffus de la pulpe rouge splénique, forme variante de leucémie à tricholeucocytes)
Lymphome lymphoplasmocytaire / maladie de Waldenström
Maladie des chaînes lourdes
Myélome multiple
Plasmocytome solitaire osseux
Plasmocytome extra-osseux
Lymphome de la zone marginale du tissu lymphoïde associé aux muqueuses
Lymphome ganglionnaire de la zone marginale
Lymphome folliculaire
Lymphome centro-folliculaire cutané primitif
Lymphome à cellules du manteau
Lymphome B diffus à grandes cellules sans autre spécificité
Lymphome B diffus à grandes cellules associé aux inflammations chroniques
Granulomatose lymphomatoïde
Lymphome à grandes cellules primitif du médiastin
Lymphome intra-vasculaire à grandes cellules
Lymphome à grandes cellules ALK positif
Lymphome plasmablastique
Lymphome à grandes cellules survenant sur une maladie de Castleman multicentrique associée au virus HHV8
Lymphome primitif des séreuses
Lymphome de Burkitt
Lymphome B inclassable avec des caractéristiques intermédiaires entre le lymphome B diffus à grandes cellules et le lymphome de Burkitt
Lymphome B inclassable avec des caractéristiques intermédiaires entre le lymphome B diffus à grandes cellules et le lymphome de Hodgkin

Tableau 1

Classification OMS 2008 des hémopathies lymphoïdes B matures (d'après référence n° 5).

Les profils évolutifs permettent de distinguer des **formes indolentes** ou des **formes agressives**. Les formes indolentes (LLC ou LZM par exemple) ne nécessitent pas systématiquement l'instauration d'un traitement et peuvent simplement justifier d'une surveillance clinique et biologique. A l'opposé, les formes agressives (LBDGC ou lymphome de Burkitt par exemple) mettent en jeu le pronostic vital à court terme et nécessitent toujours la mise en place d'un traitement.

D'après les données épidémiologiques disponibles en France, le nombre de nouveaux cas d'hémopathies malignes a été estimé à environ 35000 en 2012, soit globalement 10 % des nouveaux cas de cancers. Les hémopathies lymphoïdes représentaient environ deux tiers des hémopathies. Parmi les quatre hémopathies malignes les plus fréquentes, trois étaient des hémopathies lymphoïdes : myélome (4888 nouveaux cas), leucémie lymphoïde chronique (4464 nouveaux cas) et lymphome B diffus à grandes cellules (4096 nouveaux cas).⁷

2. LA LEUCEMIE LYMPHOÏDE CHRONIQUE

2.1. Diagnostic et classification

La LLC est une hémopathie indolente qui s'accompagne d'une augmentation du taux de lymphocytes dans le sang (hyperlymphocytose). Elle est souvent associée également à un syndrome tumoral (ganglionnaire, splénique ou hépatique). Le diagnostic repose sur l'immunophénotypage des lymphocytes sanguins par cytométrie en flux. Cet examen évalue l'expression de cinq antigènes de surface : CD5, CD23, CD79b, FMC7 et l'immunoglobuline de surface (sIg). Le score de Matutes, également appelé score RMH (*Royal Marsden Hospital*) prend en compte le niveau d'expression de ces antigènes.^{8,9} Le diagnostic de LLC est posé quand ce score est égal à 4 ou 5 (tableau 2).

Antigène	Expression	
CD5	Positive	1 point
CD23	Positive	1 point
CD79b	Faible ou négative	1 point
FMC7	Négative	1 point
sIG	Faible	1 point

Tableau 2

Score de Matutes établi par cytométrie en flux. Le diagnostic de LLC est posé si le score obtenu est supérieur à 3 points (d'après références n° 8 et 9).

Le profil évolutif des patients atteints d'une LLC est extrêmement hétérogène. Environ un tiers des patients ne sera jamais traité et, parmi les deux tiers de patients traités, certains le sont dès le diagnostic et d'autres après une période de surveillance initiale.

La classification de Binet, publiée au début des années 1980, distingue dès le diagnostic trois groupes de patients de pronostics différents. Elle est établie à partir de deux paramètres simples : les données de l'hémogramme (taux d'hémoglobine et taux de plaquettes) et la présence ou non d'un syndrome tumoral à l'examen clinique (tableau 3).¹⁰ Cette classification est toujours utilisée actuellement. Elle fait partie des critères permettant de décider de l'instauration ou non d'un traitement (de façon schématique : surveillance simple des patients stade A et mise en place d'un traitement pour les stades B et C).

Stade (% de patients)	Taux d'Hb et de plaquettes	Nombre d'aires ganglionnaires	Médiane de survie globale
A (55 %)	Hb ≥ 10 g/dL Plaq ≥ 100 G/L	< 3	> 10 ans
B (30 %)	Hb ≥ 10 g/dL Plaq ≥ 100 G/L	≥ 3	7 ans
C (15 %)	Hb < 10 g/dL ou Plaq < 100 G/L	indifférent	2 ans

Tableau 3

Classification de Binet. La présence d'une atteinte hépatique ou splénique est comprise dans le décompte des aires ganglionnaires (d'après référence n° 10). Hb : hémoglobine.

L'évolution des patients ayant un stade identique de la maladie est cependant très hétérogène. Des facteurs pronostiques complémentaires sont progressivement apparus pour identifier des sous-groupes de patients ayant des profils évolutifs différents. A côté de facteurs pronostiques simples comme le temps de doublement des lymphocytes (un temps de doublement inférieur à 1 an étant associé à un pronostic plus défavorable¹¹), des marqueurs cytogénétiques et moléculaires ont progressivement émergé.

2.2. Anomalies cytogénétiques

Environ 80 % des patients présentent des anomalies chromosomiques. Quatre anomalies récurrentes sont décrites : délétion 13q14, délétion 11q, trisomie 12 et délétion 17p. Ces anomalies sont recherchées en réalisant le caryotype des cellules tumorales, ou à l'aide de sondes fluorescentes par FISH (*Fluorescence In Situ Hybridization*).

Le pronostic des patients diffère significativement selon ces anomalies cytogénétiques. Les courbes de survie représentées sur la figure 1 indiquent la proportion de patients survivants en fonction du temps et du type d'anomalie chromosomique présente. Les patients porteurs d'une délétion 17p ont le pronostic le plus péjoratif avec une médiane de survie globale inférieure à trois ans.^{12,13}

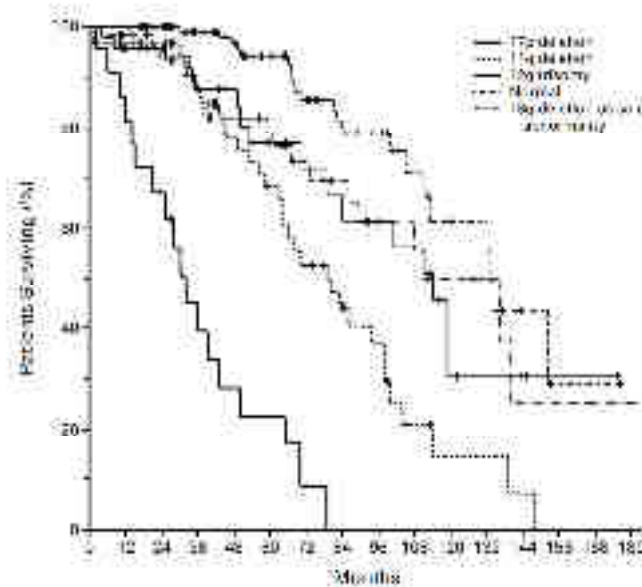


Figure 1
 Survie globale des patients selon le type d'anomalie cytogénétique présente (d'après référence n° 13).

La délétion du bras court du chromosome 17 entraîne une perte du gène suppresseur de tumeur *TP53*. Sa fréquence est inférieure à 10 % au diagnostic mais augmente significativement chez les patients en rechute. La présence de cette anomalie cytogénétique conditionne fortement le choix thérapeutique puisque ces patients sont réfractaires aux traitements basés sur l'utilisation des analogues des purines.¹⁴⁻¹⁶

2.3. Statut mutationnel du gène codant pour la séquence variable des chaînes lourdes d'immunoglobulines

Le récepteur des lymphocytes B (BCR, *B-Cell Receptor*) est une immunoglobuline de surface. La constitution du répertoire immunologique permettant à chaque lymphocyte B d'exprimer un récepteur de surface unique est un processus avec plusieurs étapes débutant par le réarrangement des gènes des immunoglobulines. Après stimulation antigénique, des mutations somatiques surviennent au niveau des séquences variables des chaînes lourdes (*IGHV*) et légères. Ce processus a lieu dans les centres germinatifs des follicules lymphoïdes secondaires.¹⁷ Ces mutations somatiques ont pour objectif d'augmenter l'affinité du récepteur à l'antigène. A la fin des années 1990, un taux plus élevé de mutations somatiques des gènes *IGHV* a été mis en évidence chez les patients de mauvais pronostic.¹⁸ Cette corrélation entre la survie des patients et le statut mutationnel a ensuite été confirmée par deux grandes études.^{19,20} Les patients ayant un statut non muté ont une médiane de survie globale significativement inférieure par rapport à celle des patients ayant un statut muté (117 mois versus 293 mois dans la publication de l'équipe anglaise, figure 2).¹⁹

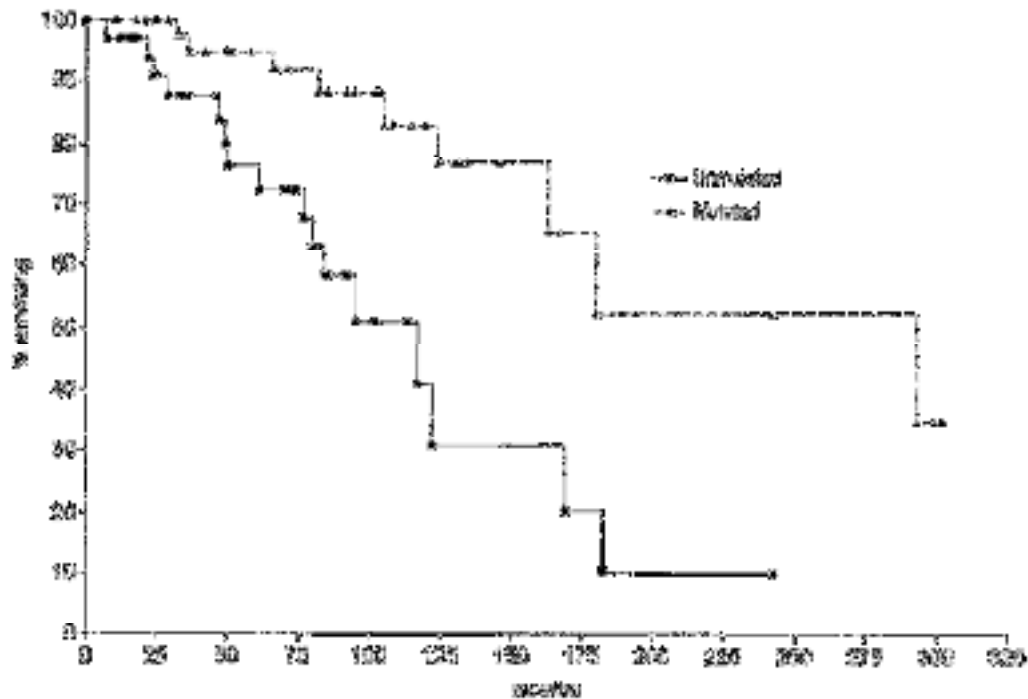


Figure 2
Survie globale des patients selon le statut mutationnel du gène IGHV (d'après référence n° 19).

Cependant la détermination du statut mutationnel des patients n'est pas réalisée en routine. La première raison est liée au coût et à la technicité nécessaire pour réaliser cet examen. La deuxième raison est l'absence de stratégie thérapeutique différente à proposer au patient selon le statut muté ou non muté. Des biomarqueurs de substitution, plus simples à obtenir, ont néanmoins été cherchés pour fournir une information équivalente. Les protéines CD38 ou ZAP-70, dont l'expression peut être déterminée par cytométrie en flux, ont été initialement proposées comme marqueurs de substitution au statut mutationnel. Une expression du CD38 et de ZAP-70 par les cellules tumorales est le plus souvent associée à un statut non muté.^{20,21}

Des études ultérieures ont cependant montré que cette corrélation n'est pas parfaite²² avec, de plus, des difficultés de standardisation pour la mesure de l'expression de ZAP-70.²³ Ces deux marqueurs, bien qu'ayant un impact pronostique, n'ont ainsi pas été adoptés universellement comme marqueurs de substitution au statut mutationnel.

2.4. Séquençage de nouvelle génération et mutations génétiques récurrentes

Les nouvelles techniques de séquençage ont permis de mettre en évidence ces dernières années l'existence de mutations somatiques acquises et récurrentes de certains gènes :

- *NOTCH1*, muté dans environ 10 % à 20 % des cas au diagnostic. Ces mutations sont associées à un pronostic plus défavorable. La fréquence de cette mutation semble plus élevée notamment dans les transformations de la maladie en lymphome agressif (syndrome de Richter)²⁴
- *BIRC3*, muté dans moins de 10 % des cas au diagnostic. La fréquence de cette mutation est plus élevée (supérieure à 20 %) chez les patients réfractaires aux traitements à base d'analogues des purines²⁵
- *SF3B1*, muté dans environ 10 % des cas au diagnostic avec une fréquence plus élevée dans les formes réfractaires²⁶
- *TP53* dont les mutations sont détectées dans moins de 10 % des cas au diagnostic. Tout comme la délétion 17p, les mutations du gène TP53 confèrent une résistance aux traitements par analogues des purines et sont associées à un mauvais pronostic.²⁷

La sensibilité actuelle des techniques de séquençage permet à présent d'appréhender l'évolution clonale et sous-clonale de la maladie et l'émergence de clone résistants. La recherche de ces mutations n'est cependant pas réalisée en routine dans tous les services d'hématologie.

Seule la recherche de la mutation du gène *TP53* a un réel impact sur la prise en charge des patients compte tenu de la résistance qu'elle confère aux traitements conventionnels à base d'analogues des purines. La prise en charge de ces patients a été récemment bouleversée par l'arrivée de thérapies ciblées inhibant la voie de signalisation du récepteur des lymphocytes B. Ces nouveaux médicaments (ibrutinib : inhibiteur de la tyrosine kinase de Bruton ; idélalisib : inhibiteur de l'isoforme delta de la PI3 kinase) ont obtenu une autorisation de mise sur le marché notamment pour le traitement des patients porteurs d'une délétion 17p ou d'une mutation du gène *TP53*.²⁸⁻³⁰

2.5. Conclusion

La leucémie lymphoïde chronique représente donc un modèle pour illustrer l'hétérogénéité du pronostic des patients au sein d'une même entité et les efforts accomplis grâce aux progrès techniques pour identifier de nouveaux biomarqueurs. Schématiquement, plusieurs périodes peuvent ainsi être identifiées :

- les années 1980 avec l'établissement de la classification de Binet, basée sur l'examen du patient et une simple numération des cellules sanguines,
- les années 1990 avec la découverte des anomalies cytogénétiques récurrentes et du statut mutationnel du gène codant pour la partie variable des chaînes lourdes d'immunoglobulines,
- les années 2000 avec l'arrivée des protocoles d'immunochimiothérapie et la confirmation du caractère réfractaire à ces traitements des patients porteurs d'une délétion 17p ou d'une mutation du gène *TP53*,
- les années 2010 avec la découverte de mutations somatiques récurrentes de certains gènes et leur lien avec la chimiorésistance ou la progression de la maladie. Cette période est également marquée par l'arrivée de thérapies ciblées qui ont bouleversé la prise en charge des patients à haut risque.

3. LYMPHOMES B DIFFUS A GRANDES CELLULES

Les LBDGC sont des lymphomes agressifs mettant en jeu le pronostic vital. Ils nécessitent toujours la mise en place d'un traitement. Ils représentent l'entité la plus fréquente parmi les lymphomes non-hodgkiniens (environ 30 % des cas). Le diagnostic repose sur l'examen anatomo-pathologique d'une biopsie tumorale. L'examen histologique est complété par des analyses immunohistochimiques permettant de confirmer l'origine lymphoïde B de la prolifération tumorale. La présence de cellules tumorales circulantes dans le sang est très rare dans les LBDGC contrairement à la LLC.

Les traitements de première ligne actuels permettent de guérir environ 60 % des patients. Une proportion encore importante de patients va donc être réfractaire au traitement de première ligne ou rechuter après une période de rémission. Le pronostic de ces patients en rechute ou réfractaires reste extrêmement péjoratif. La médiane de survie globale est souvent inférieure à 1 an pour les patients qui ne répondent pas aux traitements de rattrapage.³¹

Le premier index pronostic international (IPI) a été établi en 1993.³² Il prend en compte les paramètres suivants au diagnostic :

- Age : ≤ 60 ans versus > 60 ans
- Extension de la maladie : localisée (stades I et II de la classification d'Ann Arbor) versus disséminée (stades III et IV)
- Nombre de sites extra-ganglionnaires atteints : ≤ 1 versus > 1
- Etat général du patient : conservé versus altéré
- Taux de LDH (lactate déshydrogénase) : normal versus élevé

Quatre groupes pronostiques ont pu être établis et validés sur une cohorte de plus de 2000 patients (tableau 4 et figure 3) :

Groupe pronostique	Nbr de facteurs de risque	Pourcentage de patients	Taux de survie globale à 5 ans (%)
Faible risque	0 ou 1	35	73
Intermédiaire faible	2	27	51
Intermédiaire élevé	3	22	43
Haut risque	4 ou 5	16	26

Tableau 4

Distribution des patients et taux de survie globale à 5 ans selon l'index pronostic international (d'après référence n° 32).

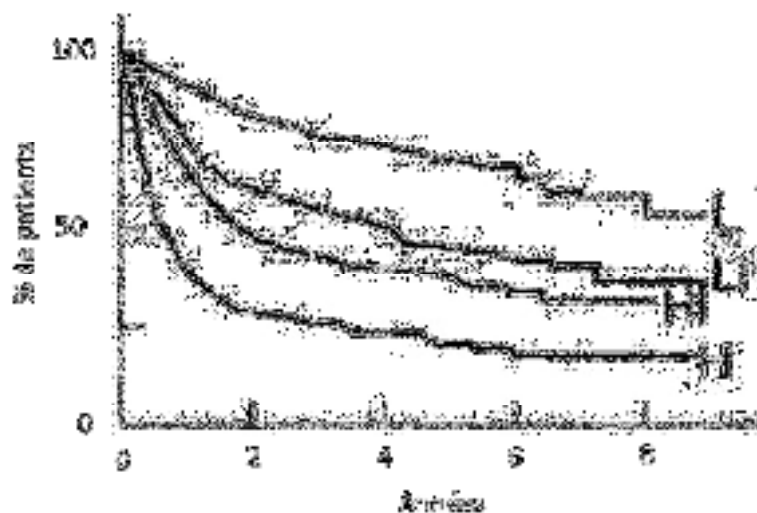


Figure 3

Courbes de survie globale des patients en fonction des groupes pronostiques définis par l'IPI (d'après référence n° 32). H : haut risque ; HI : intermédiaire élevé ; LI : intermédiaire faible ; L : faible risque.

Cet index pronostique a été établi avant l'ère du rituximab. Le rituximab est un anticorps monoclonal dirigé contre l'antigène CD20 exprimé à la surface des lymphocytes B. L'association du rituximab avec les protocoles de chimiothérapie a significativement amélioré la survie des patients traités pour un LBDGC.³³ Actuellement cet anticorps est quasi systématiquement associé aux protocoles de chimiothérapies pour le traitement des hémopathies lymphoïdes B. L'IPI conserve toutefois son impact pronostique à l'ère des traitements comportant du rituximab.³⁴

Au début des années 2000, les approches transcriptomiques basées sur les profils d'expression de gènes ont permis de mieux caractériser la diversité de ces lymphomes sur le plan biologique. Deux grands sous-types moléculaires ont ainsi pu être individualisés : un sous-type ABC (*Activated B Cell*) et un sous-type GC (*Germinal Center*). Ces sous-types moléculaires sont associés à des profils évolutifs différents avec un pronostic plus péjoratif pour les formes ABC (figure 4).³⁵

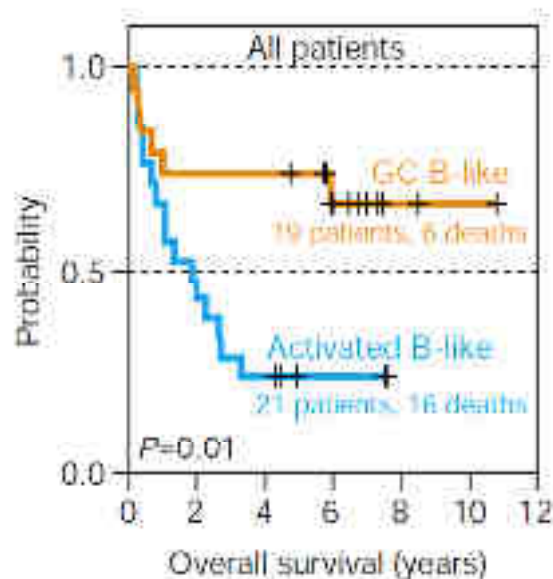


Figure 4
Survie globale des patients selon le sous-type ABC ou GC (d'après référence n° 35).

Cette détermination du sous-type moléculaire à partir des profils d'expression de gènes n'est cependant pas réalisable en pratique quotidienne compte tenu de la complexité des techniques à mettre en œuvre. Des travaux ultérieurs ont donc cherché à évaluer des biomarqueurs de substitution pour appréhender plus facilement le sous-type moléculaire. Différents score établis en immunohistochimie ont été proposés, le plus fréquemment utilisé étant l'algorithme de Hans évaluant l'expression de 3 protéines : CD10, BCL-6 et MUM1 (figure 5).³⁶



Figure 5
Algorithme de Hans utilisé en immunohistochimie (d'après référence n° 36).

D'autres facteurs pronostiques ont également émergés avec la mise en évidence de réarrangements des gènes *MYC*, *BCL2* et *BCL6* (LBDGC « double hit » ou « triple hit »). Ces réarrangement chromosomiques sont associés à des formes particulièrement agressives et de mauvais pronostic.^{37,38} La recherche de ces translocations n'est cependant pas effectuée systématiquement en routine et la meilleure stratégie thérapeutique à adopter pour ces patients n'est pas consensuelle.

Enfin les techniques de séquençage de nouvelle génération ont également permis récemment de mettre en évidence des gènes mutés de façon récurrente dans les LBDGC. Ces mutations concernent différentes voies de signalisation et ouvrent de nouvelles perspectives thérapeutiques avec le développement actuel des thérapies ciblées.³⁹

4. LYMPHOMES DE LA ZONE MARGINALE

Les lymphomes de la zone marginale sont des lymphomes indolents subdivisés en trois entités principales :

- les LZM spléniques caractérisés par une splénomégalie avec fréquemment présence de cellules lymphomateuses circulantes dans le sang périphérique,
- les LZM développés à partir du tissu lymphoïde associé aux muqueuses (MALT) dont l'exemple caractéristique est le lymphome du MALT gastrique associé le plus souvent à une infection par *Helicobacter pylori*,
- les LZM ganglionnaires.

Le diagnostic repose sur un faisceau d'arguments incluant les caractéristiques cliniques, l'examen anatomo-pathologique quand il est disponible, et les caractéristiques cytogénétiques et moléculaires.

On ne dispose pas actuellement de biomarqueurs diagnostiques spécifiques à ces lymphomes. Sur le plan phénotypique, ils expriment les marqueurs pan-lymphocytaires B (CD19, CD20, CD22) et n'expriment pas classiquement certains marqueurs plus « spécifiques » d'autres entités, comme le CD5 dont l'absence d'expression permet de les distinguer des formes classiques de LLC ou de lymphome à cellules du manteau.

Certaines anomalies génétiques ont pu être mises en évidence avec une fréquence plus élevée dans certains sous-types de LZM comme la délétion 7q dans les formes spléniques.^{40,41} Des mutations récurrentes de certains gènes ont également été décrites, surtout dans les formes spléniques : *NOTCH2* (10 %), *TP53* (16 %) ou *TNFAIP3* (7 %) par exemple.⁴² Ces anomalies ne sont cependant pas spécifiques et, le plus souvent, ne sont pas recherchées en pratique quotidienne.

La situation est rendue encore plus compliquée par l'existence de formes frontières entre ces LZM et d'autres syndromes lymphoprolifératifs actuellement considérées comme des entités provisoires dans la classification OMS : le lymphome de la pulpe rouge splénique ou la forme variante de leucémie à tricholeucocytes. Bien que ces entités partagent des caractéristiques morphologiques, phénotypiques ou moléculaires, le profil évolutif des patients diffère.

Aucun biomarqueur diagnostique n'est actuellement disponible pour distinguer facilement ces différentes entités. Ce groupe de syndromes lymphoprolifératifs illustre donc le besoin actuel d'identifier de nouveaux biomarqueurs à visée diagnostique.

5. CONCLUSION

Ces 3 exemples d'hémopathies lymphoïdes B sont représentatifs des principaux enjeux actuels dans la prise en charge des hémopathies lymphoïdes B :

- individualiser au mieux les différentes entités,
- identifier les sous-groupes de patients à haut risque,
- identifier de nouvelles cibles thérapeutiques et disposer de stratégies thérapeutiques adaptées aux différents profils de patients.

Les quantités considérables d'informations apportées ces dernières années par les données de génomique et de transcriptomique ont permis de mieux comprendre au niveau moléculaire ces maladies et d'établir des sous-groupes de patients ayant des profils évolutifs différents. Le développement rapide des thérapies ciblées permet déjà de proposer dans certains cas un traitement sur la base du profil moléculaire de certains patients.

Néanmoins beaucoup de ces biomarqueurs moléculaires n'ont pas été transposés en pratique clinique quotidienne. Un des obstacles réside dans la difficulté à appliquer en routine des techniques complexes nécessitant une importante expertise avec un coût financier souvent élevé. De plus, l'absence de conséquence pratique sur la prise en charge du patient limite dans certains cas le besoin de disposer impérativement de ces facteurs pronostiques.

Les protéines membranaires de surface sont des cibles particulièrement intéressantes pour la recherche de biomarqueurs. Elles sont en effet accessibles par des techniques telles que la cytométrie en flux ou l'immunohistochimie qui sont largement utilisées en pratique quotidienne dans les laboratoires d'hématologie ou les services de pathologie. Ces protéines membranaires sont aussi des cibles thérapeutiques potentielles pour le développement notamment des anticorps monoclonaux. La recherche de biomarqueurs au niveau protéique, et plus particulièrement au niveau des protéines membranaires de surface, reste donc un enjeu important à l'heure actuelle.

De façon similaire aux approches génomiques et transcriptomiques, les approches protéomiques par spectrométrie de masse ont également bénéficié de progrès technologiques considérables ces dernières années. Ces avancées offrent à présent de nouvelles perspectives pour la recherche de biomarqueurs. Dans le chapitre suivant j'aborderai ces principales avancées technologiques avant de présenter les approches protéomiques quantitatives globales et ciblées qui en découlent.

● Chapitre 2

La spectrométrie de masse appliquée à la protéomique pour l'identification et la validation de biomarqueurs

La spectrométrie de masse tient aujourd'hui une place centrale dans l'analyse protéomique suite aux progrès technologiques et bio-informatiques récents qui ont permis d'accroître considérablement les capacités d'identifications et d'évoluer vers des approches quantitatives applicables à des échantillons biologiques complexes. Ces développements méthodologiques ont ouvert de nouvelles perspectives pour l'analyse protéomique, en particulier pour la recherche et la validation de biomarqueurs.

La stratégie d'identification des protéines par spectrométrie de masse la plus employée (et la seule utilisée dans ce travail de thèse) repose sur une digestion enzymatique préalable des protéines. La trypsine est l'enzyme de référence car elle permet de générer des coupures spécifiques au niveau des liaisons peptidiques dans la partie C-terminale des résidus basiques lysine ou arginine permettant ensuite leur protonation pour générer principalement des espèces multi-chargées 2+ et 3+.⁴³ Ces peptides ont une taille moyenne d'environ 10 à 15 acides aminés, adaptée à la gamme de masse balayée lors de l'acquisition des spectres MS. Cette spécificité est essentielle pour la répétabilité des étapes de digestion enzymatique. Cette digestion peut être réalisée en solution (digestion liquide) ou en gel après séparation par électrophorèse des protéines.

L'analyse en spectrométrie de masse permet de caractériser les peptides obtenus par leur masse et leur spectre de fragmentation à partir duquel est déterminée la séquence en acides aminés. L'identification des peptides et l'inférence au niveau protéique sont réalisées à l'aide de moteurs de recherche utilisant des banques de données de séquences protéiques qui se sont considérablement enrichies ces dernières années. Cette stratégie d'analyse protéomique basée sur une digestion enzymatique des protéines est appelée **approche « bottom-up »** (figure 6). Elle s'oppose à la stratégie dite « top-down », qui ne sera pas abordée dans ce manuscrit, et qui repose sur l'analyse par spectrométrie de masse de protéines intactes non digérées.⁴⁴



Figure 6

Les différentes étapes d'une analyse protéomique selon une stratégie "bottom-up".

Après avoir évoqué les principales avancées technologiques récentes, nous aborderons dans ce chapitre les caractéristiques des deux grandes approches quantitatives, globales et ciblées, qui peuvent être utilisées pour la découverte et la validation de biomarqueurs potentiels. Les méthodes employées pour la réalisation de ce travail de thèse seront davantage détaillées.

1. LES PROGRES TECHNOLOGIQUES RECENTS DANS LE DOMAINE DE LA PROTEOMIQUE QUANTITATIVE

L'essor actuel de la spectrométrie de masse pour l'analyse protéomique est consécutif aux progrès technologiques réalisés dans quatre principaux domaines :

- amélioration des techniques séparatives au niveau protéique ou peptidique,
- progrès instrumentaux avec développement de spectromètres de masse de dernière génération caractérisés par leur haute résolution et une vitesse de séquençage élevée permettant de permettre d'obtenir un gain important en sensibilité,
- enrichissement des banques de données de séquences protéiques,
- progrès bio-informatiques avec le développement de logiciels dédiés à l'analyse des résultats.

Historiquement, la première grande étape qui a ouvert la voie pour l'analyse des protéines en spectrométrie de masse a été la mise au point de **sources d'ionisations douces** de type MALDI (*Matrix-Assisted Laser Desorption*)⁴⁵ ou électrospray (ESI, *Electrospray Ionisation*)⁴⁶ qui permettent de générer des ions tout en préservant l'intégrité des molécules protéiques ou peptidiques. Les sources ESI permettent d'ioniser des peptides en solution à pression atmosphérique sous l'effet d'un champ électrique.

L'approche protéomique par spectrométrie de masse a ensuite longtemps reposé sur un fractionnement extensif de l'échantillon à l'aide de gels bidimensionnels (gels 2D) séparant les protéines selon deux caractéristiques physico-chimiques majeures : le point isoélectrique et le poids moléculaire. Cette étape de fractionnement est suivie par la caractérisation des protéines d'intérêt (longtemps réalisée par détermination de l'empreinte peptidique massique obtenue avec des instruments de type MALDI-TOF).⁴⁷ Cette approche basée sur la détermination de l'empreinte peptidique massique nécessite d'isoler les protéines d'intérêt de façon très pure pour éviter la présence de protéines contaminantes dans le spot analysé.

Malgré le développement de méthodes quantitatives 2D-DIGE (*Difference In-Gel Expression*) qui permettent l'analyse simultanée de plusieurs échantillons,⁴⁸ les approches basées sur les gels 2D ne sont pas applicables à haut débit et ne sont pas adaptées pour l'analyse des protéines hydrophobes (cas notamment des protéines membranaires) ou des échantillons complexes avec une large gamme dynamique.⁴⁹ Cette notion de gamme dynamique peut être illustrée chez l'Homme par l'exemple du plasma dans lequel un ordre de grandeur de 10^{10} existe entre la protéine la plus abondante, l'albumine, et les protéines présentes en quantités les plus faibles.⁵⁰

D'importants progrès technologiques instrumentaux réalisés au cours de ces dernières années ont permis de révolutionner cette approche classique. Les spectromètres de masse de dernière génération ont apporté un gain considérable en termes de résolution, de précision de masse, de vitesse d'acquisition et de sensibilité. L'association de différents types d'analyseurs (quadripôles, temps de vol ou Orbitrap par exemple) dans un même spectromètre (instruments hybrides) a permis de développer des analyses par **spectrométrie de masse en tandem (MS/MS)**. Les analyseurs à trappe ionique permettent également de réaliser des analyses en tandem en permettant des fragmentations multiples au sein d'un analyseur unique.

Dans ces analyses en tandem, une première analyse MS permet de détecter les ions les plus intenses (ions précurseurs) sur la base de leur rapport masse sur charge (m/z) et de leur intensité. Chaque ion peptidique sélectionné est ensuite fragmenté et les différents ions fragments obtenus sont transmis dans un deuxième analyseur pour obtenir un spectre de fragmentation caractéristique de l'ion précurseur. La fragmentation de l'ion peptidique sélectionné est le plus souvent obtenue après collision avec un gaz inerte (hélium, azote ou argon par exemple) dans une cellule de collision (CID, *Collision-Induced Dissociation*). La première analyse MS renseigne sur l'intensité et la masse du précurseur déterminée à partir de son rapport m/z et de son état de charge. La deuxième analyse permet d'obtenir à partir du spectre MS/MS les informations sur la séquence en acides aminés du précurseur sélectionné pour la fragmentation. Cette information supplémentaire sur la séquence des peptides, par rapport à la seule empreinte de masse peptidique, permet d'augmenter considérablement la qualité des identifications.

La réalisation de ces analyses en tandem par cycles successifs (sélection des précurseurs dans un premier temps puis fragmentation et mesure de la masse des fragments) définit le **mode d'acquisition de type DDA (*Data-Dependent Acquisition*)**. Le mode DIA (*data-Independent Acquisition*), développé plus récemment, se distingue du mode DDA par la fragmentation systématique de tous les ions observés dans un spectre MS.^{51,52} Ce mode d'acquisition de type DIA, qui n'a pas été utilisé au cours de ce travail de thèse et qui ne permet encore que difficilement l'identification des protéines, ne sera pas traité davantage dans ce manuscrit.

En amont des progrès instrumentaux portant sur les spectromètres de masse, le développement de techniques de **chromatographie liquide à hautes performances** (HPLC, *High Performance Liquid Chromatography*) a permis d'obtenir une séparation à haute résolution du mélange peptidique obtenu après digestion. La méthode la plus communément employée est la chromatographie de partage de phases inversées. Elle repose sur une séparation des peptides en fonction de leur affinité entre une phase stationnaire apolaire composée de silice greffée de chaînes de carbone C18 et une phase mobile composée d'un mélange eau/acétonitrile. La proportion en acétonitrile augmente progressivement au cours du temps (gradient chromatographique) pour diminuer la polarité de la phase mobile et permettre l'élution des peptides des plus hydrophiles aux moins hydrophiles.

La réduction du diamètre interne des colonnes et du diamètre des particules composant la phase stationnaire a permis de réduire les débits (**nanoLC**) et donc d'utiliser un volume d'échantillon restreint tout en maintenant un pouvoir résolutif élevé. La taille des particules est habituellement de l'ordre de 5 µm dans les systèmes HPLC. Elle est actuellement inférieure à 2 µm dans les systèmes Ultra Performance LC (UPLC, Waters). Ce système chromatographique permet de travailler à des débits de 0,2 à 0,45 µL/min avec des colonnes ayant une longueur comprise entre 100 et 250 mm et un diamètre interne de 75 µm.

Le couplage en ligne d'un système chromatographique liquide à haute performance avec le spectromètre de masse permet la réalisation d'analyses **nanoLC-MS/MS**. La mise au point de la méthode d'analyse nécessite de trouver un compromis entre le temps de séparation chromatographique (qui détermine la durée d'une analyse) et la résolution de la séparation obtenue. Ces avancées technologiques peuvent permettre de s'affranchir d'un fractionnement extensif des protéines en amont de l'analyse en spectrométrie de masse. Il est à présent possible d'analyser un protéome complexe au cours d'une même analyse.⁵³ Ces approches sont dites **de type « shotgun »**.

L'application de ces stratégies a été à l'origine d'une publication récente dans laquelle environ 5000 protéines ont été identifiées à partir d'une lignée cellulaire humaine en utilisant des gradients chromatographiques de quelques heures.⁵⁴

2. LES DIFFERENTES APPROCHES QUANTITATIVES PAR SPECTROMETRIE DE MASSE POUR LA RECHERCHE ET LA VALIDATION DE BIOMARQUEURS

Deux grandes catégories de stratégies protéomiques quantitatives ont émergé :

- Des **approches globales** qui permettent d'obtenir dans un même échantillon une quantification relative simultanée d'un très grand nombre de protéines (plusieurs milliers).
- Des **approches ciblées**, extrêmement sensibles et spécifiques, mais ne concernant qu'un nombre limité de protéines. L'utilisation de standards internes marqués aux isotopes stables permet d'obtenir une quantification relative ou absolue des protéines ciblées au sein d'une matrice biologique complexe.

Ces deux types de stratégies sont donc particulièrement intéressants pour la recherche et la validation de biomarqueurs au sein d'une même plateforme de spectrométrie de masse :

- Les résultats issus des approches quantitatives globales permettent de comparer les profils d'expressions protéiques entre les différents échantillons pour sélectionner des biomarqueurs potentiels
- Validation de ces candidats sélectionnés par des approches ciblées

2.1. Approches quantitatives globales pour la phase de découverte

La recherche de biomarqueurs repose sur des comparaisons de protéomes pour mettre en évidence des différences d'expressions protéiques. Les enjeux inhérents à cette approche sont :

- Obtenir une couverture suffisamment large du protéome à explorer
- Avoir une limite de détection suffisamment basse pour mettre en évidence des variations d'expression sur des protéines faiblement exprimées dans des échantillons complexes avec une large gamme dynamique
- Avoir des temps d'analyses limités pour être compatibles avec l'exploration d'un nombre suffisant d'échantillons
- Obtenir une quantification relative des niveaux d'expression protéiques

Ces défis analytiques ont longtemps été considérés comme des facteurs limitants pour les analyses protéomiques. Ils sont maintenant accessibles par les approches « shotgun » qui ont émergé ces dernières années.

2.1.1. Procédures expérimentales

Une quantification relative peut être obtenue par deux grands types de stratégies expérimentales se distinguant par l'utilisation ou non d'un marquage préalable des échantillons.

Au cours de ce travail de thèse, je me suis concentré sur le développement d'approches sans marquage. Les principales stratégies avec marquage sont brièvement exposées avant de détailler celles sans marquage.

2.1.1.1. Approches quantitatives avec marquages

Le marquage d'un échantillon peut être réalisé métaboliquement et ont lieu dans ce cas dès l'étape de traduction, ou chimiquement et intervient alors après l'étape de synthèse protéique. Le principe général consiste à mélanger les deux échantillons, marqué et non marqué, pour effectuer une analyse simultanée. (figure 7).⁵⁵

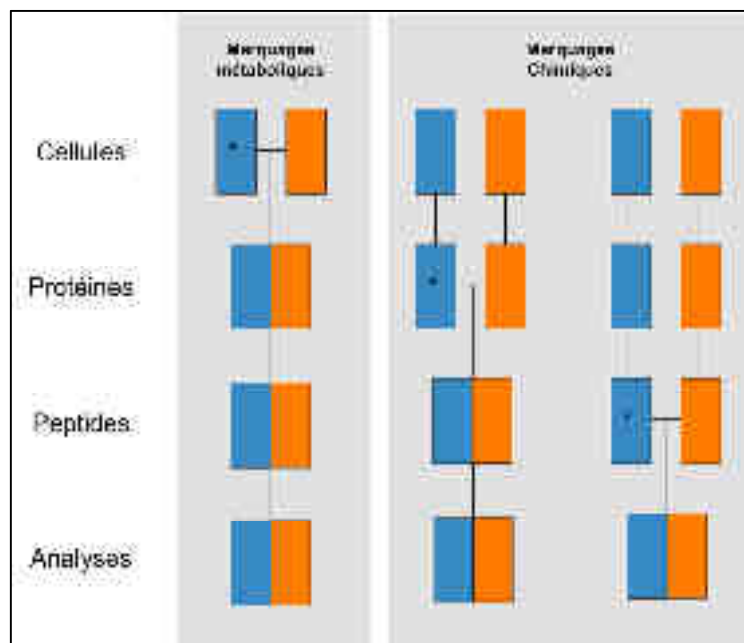


Figure 7
Représentations schématiques des stratégies avec marquage métabolique ou chimique. Chaque couleur représente une condition différente. L'étape de marquage est symbolisée par le signe * (adapté d'après la référence n° 55).

Les marquages métaboliques ont lieu au moment de la synthèse protéique. Dans le cas du marquage avec l'isotope ^{15}N , les cellules sont cultivées dans un milieu contenant comme unique source d'azote l'isotope ^{15}N , qui se retrouve ainsi intégré dans les protéines produites. Dans la technique SILAC (*Stable Isotope Labeling by Amino acids in Cell culture*), les cellules sont cultivées dans un milieu comportant des acides aminés essentiels marqués avec les isotopes ^{13}C et ^{15}N (principalement la lysine et l'arginine).⁵⁶ Le décalage de masse induit par ces isotopes stables permet ensuite de distinguer les formes légères non marquées des formes lourdes marquées. Avec des instruments de très haute résolution, jusqu'à douze échantillons peuvent être comparés en même temps.⁵⁷

Les marquages chimiques reposent sur l'utilisation de tags protéiques ciblant certains résidus d'acides aminés. Les différences de masse observées dans l'analyse MS et engendrées par ces tags permettent de distinguer l'échantillon marqué par l'isotope lourd de celui non marqué. La technique classique ICAT (*Isotope-Coded Affinity tags*) utilise les couples deutérium – hydrogène ou ^{12}C – ^{13}C .⁵⁸ Les techniques iTRAQ (*isobaric Tags for Relative and Absolute Quantitation*) ou TMT (*Tandem Mass Tags*) utilise quant à elle des tags isobariques chimiquement identiques qui se distinguent par des séquences reportrices pouvant être identifiées sur les spectres de fragmentation MS/MS. Jusqu'à huit conditions différentes peuvent être suivies.⁵⁹

2.1.1.2. Approches quantitatives sans marquage

Ces approches se distinguent par une plus grande facilité de mise en œuvre pré-analytique puisqu'elles ne nécessitent pas de manipulation supplémentaire des échantillons. Elles permettent plus facilement la comparaison d'un grand nombre d'échantillons entre eux. Elles se distinguent également des techniques avec marquages par une analyse indépendante des différents échantillons ce qui implique la nécessité d'une reproductibilité élevée sur le plan analytique.

L'aspect quantitatif peut être obtenu par le comptage du nombre de peptides ou de spectres assignés à une protéine, ou par l'extraction et l'intégration des aires sous la courbe des pics chromatographiques des peptides (XIC : *Extracted Ion Chromatograms*).⁵⁵

Ces deux modes de quantification globale sans marquage (quantification par comptage du nombre de peptides ou de spectres, ou quantification de type XIC) ont été utilisés dans ce travail de thèse.

La figure 8 schématise ces modes de quantification sans marquage comparés à une quantification avec marquage.⁶⁰

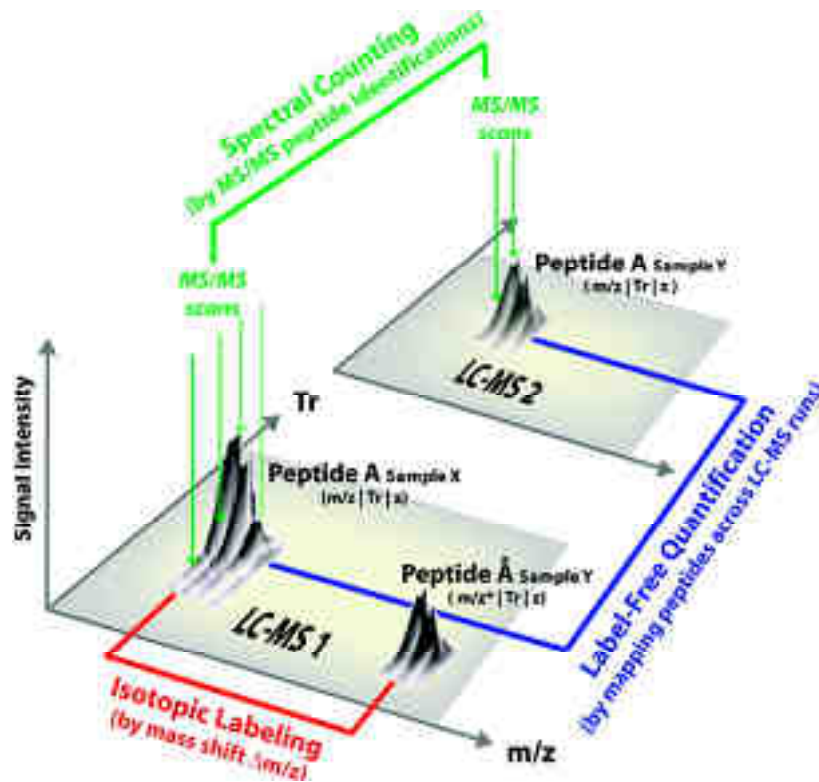


Figure 8

Illustration des principes de quantification avec ou sans marquage. En vert, comptage du nombre de spectres attribués à un peptide identique, présent dans deux échantillons différents analysés séparément. En bleu, intégration des aires des pics chromatographiques obtenus pour un même peptide, présent dans deux échantillons différents analysés séparément (XIC). En rouge, quantification avec marquage se base sur le décalage de masse lié à l'utilisation d'isotopes stables pour distinguer 2 peptides (léger et lourd) présents dans deux échantillons différents analysés simultanément (d'après référence n° 60).

2.1.1.2.1. Quantification par comptage de peptides ou de spectres

Le mode de quantification indirecte par comptage de peptides se base sur le nombre de peptides attribués à une protéine entre deux conditions différentes pour obtenir une information relative sur l'abondance de cette protéine dans chaque condition.⁶¹ De la même façon, le comptage par nombre de spectres utilise la relation existante entre la quantité de protéine présente et le nombre de spectres MS/MS acquis et assignés aux peptides de cette protéine.⁶²

Ces données sont facilement accessibles à partir des résultats fournis par les moteurs de recherche habituels et ne nécessitent donc pas de logiciel spécifique pour le traitement des données, ce qui en fait des méthodes facilement applicables pour l'analyse d'un grand nombre d'échantillons.

Ces approches sont cependant limitées par un certain nombre de facteurs :

- Les paramètres expérimentaux doivent permettre d'assurer qu'un nombre suffisant d'ions soit sélectionné au moment de l'analyse MS pour être fragmentés. Un temps d'exclusion dynamique élevé (temps pendant lequel un ion après avoir été sélectionné une première fois ne sera plus sélectionné pour l'analyse MS/MS) va ainsi sous-estimer l'abondance d'un peptide dans l'échantillon. Un effet de saturation doit aussi être évité par le contrôle de la quantité de protéines injectées et la largeur des pics chromatographiques.
- L'interprétation des résultats doit tenir compte de la taille des protéines identifiées, une protéine de petite taille générant moins de peptides après digestion enzymatique. Une protéine de petite taille faiblement abondante sera donc plus difficilement quantifiable par ces approches qu'une protéine abondante de haut poids moléculaire.
- Il faut également tenir compte pour le comptage du nombre de peptides du problème posé par les peptides partagés entre plusieurs protéines. Dans ces situations une solution consiste à répartir les spectres de ces peptides partagés proportionnellement au nombre de spectres uniques ayant permis l'identification de ces protéines.

2.1.1.2.2. Quantification par extraction des courants d'ions (XIC)

Ce mode de quantification se base sur la linéarité et la corrélation entre la concentration d'un peptide et le signal chromatographique obtenu.⁶³ Elle consiste à extraire et à intégrer à partir du profil d'élution chromatographique le signal (courant d'ion) d'un peptide déterminé par son rapport m/z et son temps de rétention. Cette approche nécessite une précision élevée dans la mesure du rapport m/z, associée à une résolution élevée pour isoler les ions d'intérêts et réduire les interférences (intégration d'un signal correspondant au bruit de fond ou d'un composé de rapport m/z identique éluant au même moment).

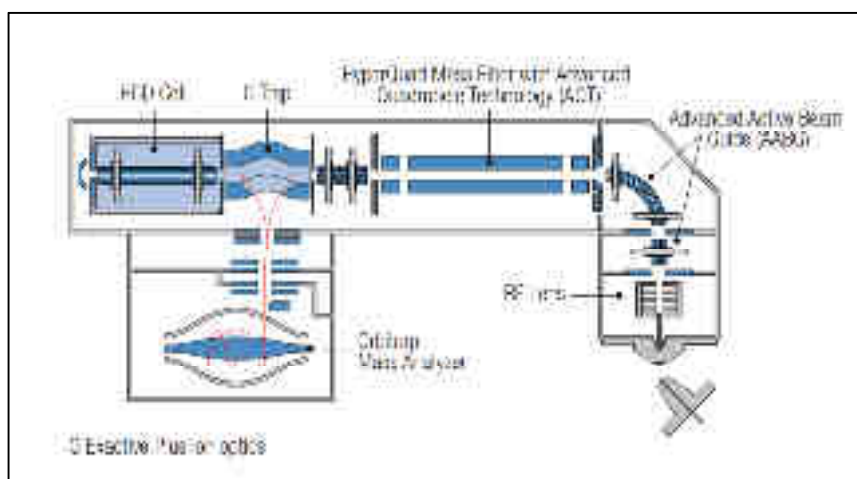
Ce type d'approche permet de plus d'extraire dans une analyse le signal d'un peptide non sélectionné pour la MS/MS, à condition d'avoir été identifié par MS/MS dans un autre échantillon au préalable, à partir de l'alignement des temps de rétention entre les analyses. Cet aspect nécessite sur le plan chromatographique une excellente stabilité et reproductibilité des temps de rétention entre les différentes analyses. Le mode d'acquisition des données DDA nécessite de s'assurer que la période de balayage en MS pour la sélection des ions soit suffisante pour acquérir suffisamment de points de mesure d'un ion et permettre ainsi la bonne intégration du signal sur le profil chromatographique.

Tout comme le comptage du nombre de peptides ou de spectres, cette méthode peut s'appliquer à un grand nombre d'échantillons mais nécessite l'utilisation de logiciels spécifiques pour l'extraction des aires des pics chromatographiques et le traitement des données.

2.1.2. Instrumentations

2.1.2.1. Spectromètres hybrides quadripôle-Orbitrap

Cette configuration est celle de l'Orbitrap Q Exactive Plus™ (Thermo Fisher Scientific). Les ions générés au niveau de la source ESI sont accélérés à l'entrée du spectromètre et focalisés au niveau du quadripôle avant d'être stockés dans la C-trap qui contrôle le nombre d'ions entrant dans la cellule Orbitrap. En mode MS la détection des ions précurseurs se fait dans l'Orbitrap. En mode MS/MS, la fragmentation a lieu dans la cellule de collision HCD (*Higher-energy Collisional Dissociation*), les fragments générés retournent ensuite dans la C-trap avant d'être analysés dans la cellule Orbitrap. (figure 9).



[Figure 9](#)

Représentation schématique de l'instrument Q Exactive Plus (Thermo Fisher Scientific).

Ce type d'instrument permet d'obtenir des vitesses de balayage élevées (réalisation de 10 MS/MS par seconde), une résolution élevée (Résolution = 70000 à $m/z = 200$ en MS ; Résolution = 17000 à $m/z = 200$ en MS/MS) et une grande précision de mesure de masse (< 3 ppm).

2.1.2.2. Spectromètres hybrides quadripôle-TOF (Q-TOF)

Ces instruments associent un analyseur de type quadripôle avec un analyseur à temps de vol (TOF, *Time Of Flight*) (figure 10).

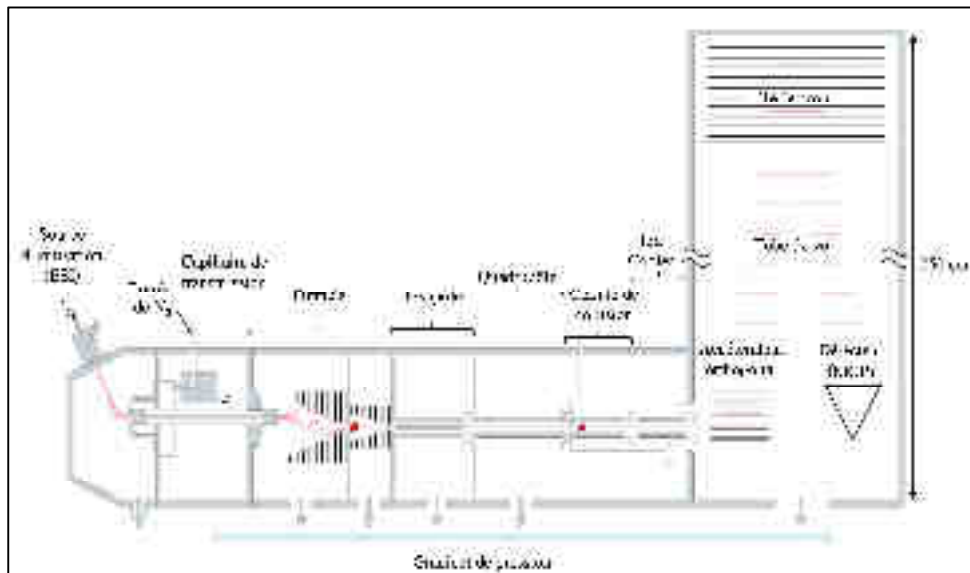


Figure 10
Représentation schématique de l'instrument Impact HD.

Pour l'acquisition des spectres MS, les ions transmis sont séparés selon leur rapport m/z par l'analyseur à temps de vol. En mode MS/MS, la sélection des précurseurs a lieu dans le quadripôle. Ces ions peptidiques sont ensuite fragmentés dans cellule de collision. L'analyseur à temps de vol permet de mesurer le rapport m/z des ions fragments et de générer ainsi le spectre de fragmentation.

Dans le cadre de ce travail, j'ai été amené à utiliser les Q-TOF Impact HD et Impact II (Bruker).

2.1.3. Traitement des données

2.1.3.1. Recherche dans les banques de données

Les banques de données de séquences protéiques découlent des banques génomiques générées à partir d'une transcription automatique des séquences codantes, mais la qualité des données diffère selon les critères d'exigences concernant la validation des données. Une des principales ressources disponibles de séquences protéiques annotées provient du consortium UniProt (*Universal Protein Resource*), créé en 2002, et sa base de données UniProtKB (*UniProt Knowledgebase*, <http://www.uniprot.org/uniprot>).

Cette base de données regroupe deux banques :

- La banque TrEMBL (Translated EMBL) contient des séquences protéiques dérivées automatiquement des séquences nucléotidiques regroupées au niveau de l'EMBL (*European Molecular Biology Laboratory*). Ces séquences protéiques générées automatiquement ne bénéficient d'aucune vérification manuelle.
- La banque Swiss-Prot qui fait aujourd'hui office de banque de référence car toutes les séquences qu'elle contient ont été vérifiées et validées manuellement à partir de données publiées. Le nombre d'entrées disponibles est donc beaucoup plus limité.

D'autres banques sont disponibles, comme la banque NCBI nr (*National Center for Biotechnology Information non redundant*), qui regroupe des données de différentes banques, dont Swiss-Prot, ainsi que des données générées automatiquement à partir des séquences nucléotidiques disponibles dans les banques génomiques. Des redondances persistent cependant car seules les séquences strictement identiques sont fusionnées sous un même numéro d'accession.

Lors de la recherche dans les banques de séquences protéiques, les données expérimentales (sous forme de listes contenant la masse des peptides précurseurs et de leurs fragments) sont comparées aux données théoriques obtenues par digestion et fragmentation *in silico* des séquences protéiques disponibles dans les banques. Cette première étape permet d'établir une liste de peptides identifiés. L'inférence au niveau protéique consiste ensuite à regrouper les différents peptides pouvant être attribués à une même séquence protéique.⁶⁴ Cette étape cruciale se base sur la relation complexe existant entre séquences protéiques et peptides : une séquence protéique contient plusieurs peptides et un peptide peut être attribué à différentes séquences protéiques. L'existence de peptides uniques à une séquence protéique au sein d'une banque permet d'identifier distinctement cette protéine. A l'opposé, différentes protéines peuvent être identifiées à partir d'un même ensemble de peptides.

Pour effectuer les recherches dans ces banques, différents moteurs de recherche sont disponibles (SEQUEST,⁶⁵ OMSSA⁶⁶, Mascot⁶⁷ ou Andromeda⁶⁸ par exemple). Au cours de mon travail de thèse j'ai principalement utilisé le logiciel Mascot et le moteur de recherche Androméda intégré au logiciel MaxQuant.

L'algorithme du logiciel Mascot attribue un score d'ion à chaque identification qui est fonction du risque que l'assignation faite soit liée au hasard. Plus ce score est élevé, plus la confiance dans l'assignation de la séquence est élevée. En plus de ce score d'ion, Mascot calcule également un score d'identité qui prend en compte la taille de l'espace de recherche.

Compte-tenu de la quantité importante de données générées par les analyses shotgun sur des échantillons complexes, il est nécessaire de valider les résultats fournis par les moteurs de recherche en estimant le taux de faux-positifs (assignation par erreur d'un spectre MS/MS à une séquence peptidique). Cette évaluation peut se faire par l'utilisation des stratégies cibles – leurres (« *target - decoy* ») : la banque contenant les séquences pour la taxonomie étudiée (cible, ou *target*) est concaténée avec une banque artificielle contenant les séquences inversées (leurres, ou *decoy*). Le nombre d'identifications faites à partir des séquences cibles et leurres permet de déterminer le FDR (*False Discovery Rate*) selon l'équation :

$$\text{FDR} = 2 \times \frac{\text{Nombre de decoy}}{(\text{nombre de decoy} + \text{nombre de target})}$$

Les paramètres de validation sont adaptés pour obtenir un FDR le plus souvent inférieur ou égal à 1 % au niveau des spectres assignés à un peptide ainsi qu'au niveau protéique. Différents logiciels ont été utilisés pour effectuer cette étape de validation des résultats : Scaffold,⁶⁹ MaxQuant⁷⁰ et Proline.⁷¹

2.1.3.2. Logiciels pour la quantification par extraction des courants d'ions

Ce mode de quantification nécessite l'utilisation de logiciels dédiés pour l'extraction et l'analyse des données. Ils permettent la détection des ions et le réalignement des temps de rétention entre les différentes analyses. La quantification au niveau protéique est obtenue à partir des intensités peptidiques.

Au cours de ce travail de thèse j'ai principalement utilisé le logiciel en accès libre MaxQuant⁷⁰ pour le traitement des données dans le cadre des approches globales. Il dispose de son propre moteur de recherche intégré (Andromeda). Les données sont extraites à partir des fichiers bruts (*raw data*). Le résultat final permet d'obtenir des valeurs d'intensités aux niveaux peptidiques et protéiques. Pour les analyses quantitatives sans marquage, MaxQuant applique un algorithme spécifique de quantification appelé MaxLFQ.⁷² Cet algorithme permet de générer des valeurs d'intensités protéiques normalisées à partir des intensités peptidiques obtenues dans chaque échantillon. Lorsqu'un peptide n'a pas été identifié par MS/MS dans une analyse, la fonction « *match between runs* » permet d'extraire son intensité à partir du profil chromatographique de cette analyse. Cette fonction nécessite le réalignement des temps de rétention entre les analyses. Le peptide est extrait du chromatogramme à partir de son temps de rétention et de son rapport m/z.

Le logiciel MaxQuant a été initialement développé pour traiter les données provenant des instruments Thermo Fisher Scientific. Il est maintenant aussi compatible avec les données provenant de spectromètres de masse de différents constructeurs. Le processus d'analyse est entièrement intégré et ne permet pas de contrôler certains paramètres comme, par exemple, l'intégration des pics sur le profil chromatographique.

Le logiciel Proline⁷¹ est en cours de développement dans le cadre de l'Infrastructure Nationale de Protéomique (ProFI) (<http://www.profi-proteomics.fr>). Cette infrastructure regroupe les laboratoires de Grenoble (EDyP), Toulouse (IPBS) et Strasbourg (LSMBO). Un des objectifs est la mise à disposition d'outils bio-informatiques dédiés à l'analyse des données de protéomique. L'objectif du logiciel Proline est de fournir une interface conviviale permettant de réaliser l'ensemble des étapes nécessaires au traitement des données. Le développement maîtrisé de ce logiciel entre les trois laboratoires permet de répondre au mieux aux besoins des utilisateurs, de connaître précisément les algorithmes utilisés et de pouvoir le faire évoluer selon les besoins. Il offre la possibilité de réaliser une quantification sans marquage par comptage du nombre de spectres ou par extraction des courants d'ions. Une version accessible gratuitement en ligne (<http://proline.profi-proteomics.fr>) est disponible depuis septembre 2015.

2.2. Approches ciblées pour la validation

L'approche ciblée de type SRM (*Selected Reaction Monitoring*) se distingue des approches globales par le suivi d'un nombre déterminé de protéines.⁷³ Chaque protéine est quantifiée à partir d'un nombre limité de peptides uniques (peptides protéotypiques). Pour chaque protéine, les peptides protéotypiques choisis (précurseurs) sont suivis à partir des fragments générés : un couple peptide précurseur – fragment est appelé « transition ». Plusieurs transitions peuvent être suivies pour un même peptide (plus le nombre de transitions suivies pour un peptide est élevé, plus la spécificité est élevée). Ces approches SRM permettent de suivre plusieurs peptides par protéines, et plusieurs protéines différentes dans une même analyse (multiplexage).

Les possibilités de multiplexage permettent actuellement de suivre en routine au laboratoire jusqu'à 30 protéines en suivant trois peptides par protéines. Chaque peptide peut être suivi avec trois transitions ce qui représente un total de 270 transitions suivies pour 30 protéines. Les transitions peuvent être suivies tout le long du gradient chromatographique ou uniquement sur une période déterminée (méthode dite « *scheduled* »).

Les transitions sont détectées avec une grande sensibilité sur des instruments de type triples quadripôles. La limite de quantification est située entre 500 amol et 1 fmol. La spécificité de cette approche est liée à la double sélection (ion précurseurs et fragments). L'absence d'interférence est vérifiée par le contrôle de l'intégration des pics chromatographiques.

Par rapport aux techniques immunologiques classiques, cette approche par SRM présente comme avantages :

- de ne pas être dépendante de l'existence d'un anticorps monoclonal dirigé contre la protéine cible,
- de pouvoir suivre plusieurs peptides d'une même protéine (un anticorps monoclonal ne reconnaissant qu'un seul épitope de la protéine),
- l'utilisation de standards internes (peptides marqués aux isotopes stables) permet d'obtenir une quantification relative ou absolue,
- se suivre des peptides modifiés et non modifiés.

2.2.1. Procédures expérimentales

La mise en place d'une analyse SRM suit différentes étapes (figure 11) :

- choix des peptides protéotypiques à suivre pour chaque protéine d'intérêt,
- sélection des transitions à suivre pour chaque peptide,
- optimisation des paramètres instrumentaux pour chaque transition (énergie de collision).

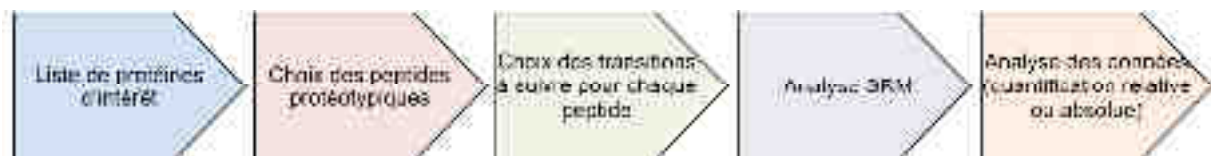


Figure 11

Schéma récapitulatif des différentes étapes pour la mise en place d'une analyse SRM.

2.2.1.1. Choix des peptides protéotypiques et des transitions

Un peptide protéotypique est un peptide unique à une protéine. Différents critères doivent être remplis par ces peptides protéotypiques pour être sélectionnés pour une analyse SRM :

- ne pas être spécifiques à une isoforme de la protéine,
- être détectables en analyse LC-MS/MS,
- ne pas contenir d'acides aminés pouvant être modifiés (exclusion des peptides contenant des méthionines pouvant être modifiées par oxydation, ou des cystéines qui sont carbamydométhylées lors de l'étape d'alkylation des protéines pendant la préparation de l'échantillon),
- ne pas contenir de sites de clivage enzymatique pouvant être manqués pendant l'étape de digestion (la trypsine n'hydrolyse pas les liaisons peptidiques lorsque la lysine ou l'arginine est suivie par une proline),
- ne pas être situés dans des régions transmembranaires de la protéine (peptides hydrophobes difficiles à détecter).

Différentes ressources sont maintenant disponibles pour aider au choix des peptides. La plateforme NextProt, accessible en ligne (<http://www.nextprot.org>), a été créée en 2011.⁷⁴ Cette base de données d'annotations des protéines a pour objectif de centraliser le maximum d'informations disponibles sur les protéines humaines. Elle intègre les données générées par les analyses transcriptomiques et protéomiques (fonction des protéines, expression tissulaire, localisation cellulaire, variants, modifications post-traductionnelles). La qualité des informations est assurée par une validation systématique des données. Une section « *peptide viewer* » permet de visualiser, pour une protéine donnée, les peptides protéotypiques utilisés dans des analyses SRM (peptides endogènes ou synthétiques). Le SRM Atlas, accessible depuis la plateforme NextProt, est une banque de données regroupant, pour chaque protéine, les différents peptides utilisés dans les analyses SRM colligées au niveau de cette banque. Ces outils apportent donc une aide considérable au choix des peptides protéotypiques pour la mise au point d'une analyse SRM.

Le choix des transitions à suivre pour chaque peptide est également une étape importante dans la mise au point de la procédure expérimentale. Leur choix se fait sur la qualité des spectres MS/MS obtenus par les analyses nanoLC-MS/MS réalisées au préalable (ayant permis de sélectionner les protéines d'intérêts) et sur la base de l'aspect du pic de chaque transition pouvant être extrait du profil chromatographique (intensité élevée avec des pics ayant une faible largeur à mi-hauteur). Pour assurer une sensibilité et une spécificité élevées, au moins trois peptides protéotypiques sont sélectionnés pour une protéine et chaque peptide est quantifié par le suivi d'au moins trois transitions.

Pour pouvoir suivre un grand nombre de peptides, des critères de « fenêtrage » sur le gradient chromatographique peuvent être utilisés afin de ne suivre les transitions d'un peptide que dans la période d'élution chromatographique de ce peptide. Cette méthode dite « *scheduled* » permet d'augmenter le nombre de peptides suivis sans compromettre la sensibilité.

2.2.1.2. Utilisation de peptides synthétiques marqués

L'utilisation de peptides marqués par des isotopes stables permet, selon leur degré de pureté, d'obtenir une quantification relative ou absolue. Des peptides synthétiques, identiques en séquence aux peptides protéotypiques sélectionnés, sont incorporés dans l'échantillon à analyser. Cette incorporation est réalisée juste avant l'analyse en LC-SRM, c'est-à-dire après les étapes d'extraction et de digestion des protéines endogènes. Ces peptides synthétiques sont marqués par les isotopes stables ^{13}C ou ^{15}N sur les lysines et arginines en position N-Ter, ce qui engendre un décalage du rapport m/z de 4 Da par rapport au peptide endogène. La quantification pour chaque peptide est obtenue par le ratio entre les formes lourdes (synthétiques) et légères (endogènes).

Différents types de peptides synthétiques sont disponibles et utilisés selon la méthode de dilution des isotopes stables (SID, *Stable Isotope Dilution*) :⁷⁵

- des peptides marqués mais non quantifiés de façon précise (PEPotec™, Thermo Fisher Scientific). Ils présentent l'avantage d'un coût moins important, mais ils ne permettent que d'obtenir une quantification relative entre les différents échantillons,
- des peptides AQUA™ (Thermo Fisher Scientific) purifiés et parfaitement quantifiés. Le mélange de ces peptides en quantité connue dans l'échantillon permet d'obtenir une quantification absolue.

Il faut cependant tenir compte du biais induit par le fait que ces peptides de référence ne sont pas soumis, comme indiqué précédemment, aux mêmes étapes de préparation que l'échantillon à analyser.

D'autres approches de quantification par méthode SRM ont également été développées :

- utilisation de standards QconCAT™ : une protéine chimérique formée par les peptides sélectionnés est produite par expression dans des cultures d'*E. Coli* en milieu contenant des arginines et lysines marquées⁷⁶,
- utilisations de standards PSAQ™ (*Protein Standard Absolute Quantification*) : la protéine marquée utilisée comme standard est identique en terme de séquence à la protéine endogène ciblée.⁷⁷

Ces standards protéiques marqués ont l'avantage d'être intégrés dès les premières étapes du traitement de l'échantillon, ce qui permet de limiter les biais liés aux différentes étapes de préparation des échantillons. Ce sont cependant des techniques beaucoup plus coûteuses, de mise en œuvre plus compliquée et qui ne sont applicables qu'à un nombre limité d'échantillons.

Dans mon travail de thèse j'ai utilisé une approche par SRM basée sur une quantification relative avec utilisation de peptides standards de type PEPotec™.

2.2.2. Instrumentations

L'approche SRM repose sur l'utilisation de spectromètres de type triple-quadripôles. Dans ce mode d'analyse, chaque quadripôle a une fonction distincte (figure 12) :

- le premier quadripôle permet la sélection du précurseur selon son rapport m/z ,
- le précurseur sélectionné est fragmenté dans le deuxième quadripôle (fragmentation de type CID (*Collision Induced Dissociation*)),
- le troisième quadripôle permet de sélectionner un fragment donné selon son rapport m/z pour suivre une transition déterminée.

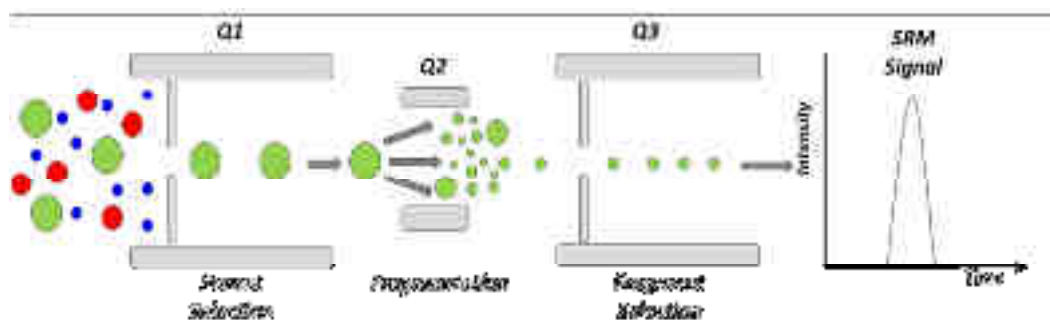


Figure 12

Sélection des peptides précurseurs et des transitions sur un instrument de type triple-quadripôles.

Ces analyseurs permettent de sélectionner spécifiquement les transitions à suivre au sein d'un échantillon avec une gamme dynamique élevée. Au cours de ce travail de thèse j'ai utilisé l'appareil TSQ Vantage (Thermo Scientific) couplé au système chromatographique Ultimate 3000.

Le paramétrage de la méthode permet d'établir le temps de cycle (durée des cycles pendant lesquels les transitions choisies sont acquises). Ce temps de cycle dépend donc du nombre de transitions à suivre et de la durée pendant laquelle chaque transition est suivie. Ce temps de cycle ne doit pas être trop long pour permettre d'acquérir suffisamment de points (idéalement une dizaine) pour définir précisément chaque pic sur le tracé chromatographique.

2.2.3. Traitement des données

L'analyse repose sur l'extraction des courants d'ions des peptides sélectionnés pour chaque transition (peptides endogènes et standards). Les aires sous la courbe des courants d'ions de chaque transition sont intégrées puis additionnées pour chaque peptide. Des logiciels permettent de réaliser ces étapes de façon automatisée, mais il est souvent nécessaire de vérifier manuellement la bonne intégration des pics. Dans ce travail de thèse j'ai utilisé le logiciel en accès libre Skyline⁷⁸ pour l'analyse des données de SRM.

Ce logiciel, dont l'utilisation n'est cependant pas restreinte aux approches ciblées, nécessite au préalable la constitution d'une librairie de spectres établie à partir des peptides sélectionnés. Il offre l'avantage de pouvoir vérifier, et corriger si besoin, l'intégration des pics. Cette vérification manuelle est cependant fastidieuse si un grand nombre de peptides doit être quantifié.

La quantification relative ou absolue, selon le standard interne utilisé, est obtenue par la mesure du ratio entre l'isotope lourd synthétique et l'isotope léger présent biologiquement dans l'échantillon. Ces ratios, calculés pour chaque échantillon, permettent la comparaison des échantillons (figure 13).

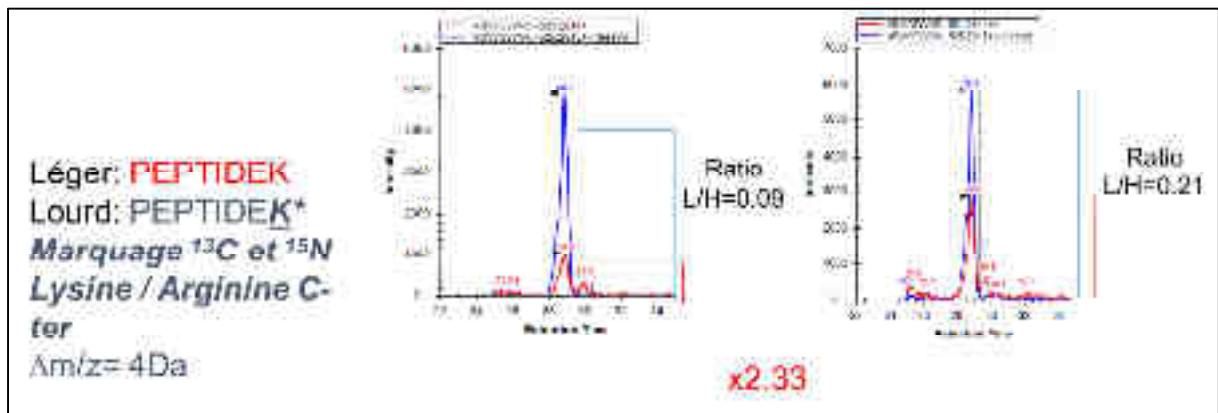


Figure 13

Exemple de quantification relative par SRM obtenue à l'aide du logiciel Skyline. Le ratio des intensités entre le peptide lourd et le peptide endogène est déterminé dans chaque échantillon. Ces ratios permettent de comparer les échantillons entre eux. Dans cet exemple le ratio calculé entre les deux échantillons est de 2,33.

3. CONCLUSION

Les approches protéomiques par spectrométrie de masse ont considérablement évolué ces dernières années. Une quantification relative sans marquage de plusieurs milliers de protéines peut être obtenue actuellement à partir d'échantillons complexes en quelques heures d'analyse.

Le développement des approches quantitatives ciblées permet une quantification extrêmement sensible et spécifique de protéines sélectionnées au préalable.

Ces deux types de quantification sans marquage permettent d'utiliser ces approches pour les phases de découvertes de biomarqueurs (approche quantitative globale) et pour leur validation ultérieure (approche quantitative ciblée).

Partie II

Projets de recherches et résultats obtenus

● Introduction

La thématique portant sur la mise en évidence et la validation de biomarqueurs dans les hémopathies lymphoïdes B a été initiée au laboratoire à partir de 2004. Les différents projets sont basés sur des approches quantitatives globales sans marquage et ont bénéficié au fil des années des progrès instrumentaux et bio-informatiques. Les premières approches globales ont reposé sur une quantification sans marquage par comptage de peptides et de spectres pour évoluer vers une quantification basée sur l'extraction des courants d'ions. Parallèlement aux stratégies globales, une méthode de quantification ciblée par SRM a été mise en place pour permettre la validation des biomarqueurs potentiels issus des approches globales.

Les progrès instrumentaux et méthodologiques ont permis en l'espace de dix ans de passer d'une liste de 400 protéines à plus de 4000 protéines dans les projets menés au laboratoire. Ils ont également permis d'évoluer dans le choix du matériel biologique initial. Les premiers travaux ont porté sur l'étude de cellules clonales isolées à partir de prélèvements sanguins, permettant d'avoir un matériel biologique très homogène.

Au cours de ma thèse j'ai également pu développer une approche basée sur l'utilisation de tissus tumoraux (prélèvements ganglionnaires obtenus chez des patients porteurs d'un lymphome B diffus à grandes cellules) qui représente un matériel complexe et hétérogène, associant les cellules tumorales, des cellules saines résiduelles et l'ensemble du micro-environnement tumoral. Cette évolution permet d'élargir le champ des possibilités concernant les différents types d'hémopathies accessibles à une étude protéomique par spectrométrie de masse.

- **Chapitre 1**

Identification et validation des CD148 et CD180 comme marqueurs phénotypiques dans les syndromes lymphoprolifératifs B

Les différents projets menés dans le laboratoire et qui ont permis d'aboutir à l'identification de ces deux biomarqueurs ont fait l'objet des travaux de doctorat du Dr Laurent MIGUET (2006), du Dr Guillaume BECHADE (2009) et du Dr Sarah LENNON (2014). Ils ont été réalisés en collaboration avec le Pr Laurent MAUVIEUX (Laboratoire d'Hématologie des Hôpitaux Universitaires de Strasbourg).

1. CONTEXTE SCIENTIFIQUE ET MEDICAL

La stratégie diagnostique devant un bilan d'hyperlymphocytose repose tout d'abord sur l'examen du frottis sanguin et la réalisation d'un immunophénotypage des lymphocytes circulants.

L'immunophénotypage est un examen fondamental dans ce contexte qui permet d'évaluer par cytométrie en flux l'expression par les cellules tumorales d'un panel d'antigènes. Les cibles antigéniques sont le plus souvent situées à la surface des cellules. Les antigènes de surface sont regroupés en clusters de différenciation (CD) définis par l'existence d'un anticorps monoclonal permettant de les cibler.

L'immunophénotypage des lymphocytes est suffisant pour poser le diagnostic de leucémie lymphoïde chronique (LLC) dans les formes typiques (score de Matutes). Le diagnostic précis des autres syndromes lymphoprolifératifs ne peut cependant pas être établi avec certitude sur la seule base de cet examen en raison de l'absence de marqueur phénotypique spécifique à chaque entité. La démarche diagnostique nécessite des examens complémentaires pour caractériser les cellules tumorales sur la base d'un faisceau d'arguments : cytogénétique (recherche de translocations chromosomiques récurrentes pouvant être plus ou moins spécifiques de certaines entités), biologie moléculaire (avec possibilité de techniques de séquençage de nouvelle génération) et histologie.

La cytométrie en flux a donc une place centrale dans la démarche diagnostique initiale. Elle offre l'avantage d'être réalisable rapidement et de pouvoir analyser simultanément différents antigènes à la surface des cellules. Elle reste néanmoins dépendante de l'existence d'anticorps monoclonaux dirigés contre les cibles antigéniques évaluées.

Les protéines de surface situées au niveau de la membrane cytoplasmique sont des cibles privilégiées pour la recherche de biomarqueurs. Elles sont en effet facilement accessibles pour une évaluation par cytométrie en flux. Elles occupent également d'un point de vue fonctionnel une place centrale dans les différents processus cellulaires (transmissions de signaux extra-cellulaires et transduction du signal, échanges de molécules entre les différents compartiments cellulaires, sites d'activités enzymatiques, adhésion cellulaire) et représentent donc fréquemment des cibles pharmacologiques.

L'analyse protéomique de ces protéines membranaires se heurte cependant à leur forte hydrophobicité et leur faible abondance qui les rendent difficiles à extraire et à solubiliser. Elles sont ainsi très largement sous-représentées dans les banques de données protéiques alors qu'environ 25 % des séquences codantes du génome codent pour des protéines ayant au moins un domaine transmembranaire.⁷⁹ Deux axes peuvent donc être développés pour l'analyse des protéines membranaires :

- obtenir une meilleure solubilisation de ces protéines membranaires lors des phases d'extraction et de digestion par l'utilisation de solvants organiques ou de détergents tels que le SDS,⁸⁰
- enrichir l'échantillon en protéines membranaires. La stratégie choisie dans les projets réalisés au laboratoire a été de travailler à partir d'un sous-protéome constitué de microparticules obtenues par stimulation des cellules tumorales.

Les cellules saines libèrent dans le milieu extracellulaire deux types de microvésicules (MV) : des exosomes et des microparticules (MP) (Figure 14).

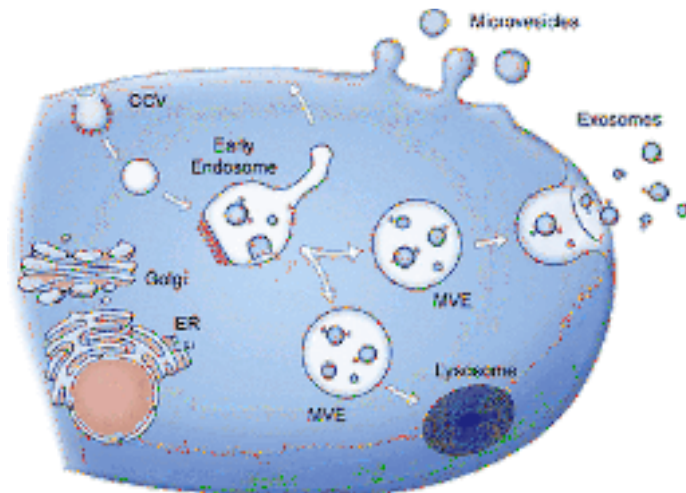


Figure 14
Les différentes structures vésiculaires (d'après référence n° 81).

Certaines caractéristiques distinguent ces deux types de MV :⁸¹

- Les exosomes ont une taille comprise entre 40 et 100 nm. Ils proviennent du système endosomal intracellulaire et sont formés à partir de vésicules intracellulaires qui s'accumulent dans des corps multi-vésiculaires. Une partie de ces corps multi-vésiculaires fusionne avec la membrane cytoplasmique pour permettre la libération des exosomes. Ils permettent le transport dans la cellule ou l'expulsion hors de la cellule de différents constituants protéiques ou d'acides nucléiques (ARNm).
- Les MP prennent naissance directement au niveau de la membrane cytoplasmique qui forme des évaginations à l'origine de ces vésicules. Ce « bourgeonnement » de la membrane cytoplasmique nécessite une réorganisation très localisée du cytosquelette et s'accompagne d'une modification de la répartition des phospholipides membranaires avec exposition de phosphatidylsérines sur le versant externe de la bicouche phospholipidique.⁸² Elles sont donc constituées par des fragments de membrane cytoplasmique et contiennent du matériel cytoplasmique provenant de la cellule d'origine. Leur taille, plus grande que celle des exosomes, est comprise entre 100 et 1000 nm.

Les MP sont produites dans le sang par différents types cellulaires (plaquettes, lymphocytes, polynucléaires, monocytes et les cellules endothéliales). Des modifications de ce taux de MP circulantes ont été rapportées dans différentes pathologies : lésions endothéliales liées au diabète,⁸³ hypertension artérielle pulmonaires,⁸⁴ pré-éclampsie,⁸⁵ athérosclérose et syndromes coronariens,⁸⁶ maladies auto-immunes⁸⁷ et cancers.⁸⁸

En raison de leur origine directement issue de la membrane cytoplasmique, l'hypothèse initiale était que les MP puissent constituer un matériel biologique de départ enrichi en protéines membranaires.

2. ETAPES DE DECOUVERTE

2.1. Génération des microparticules induites et validation en tant que sous-protéome enrichi en protéines membranaires

Les travaux initiaux ont permis de mettre au point la procédure expérimentale pour la génération et l'analyse protéomique des MP.⁸⁹

Les différentes étapes ont été :

- la génération *in vitro* de MP induites en grande quantité par stimulation des cellules avec de l'actinomycine D ou de la phyto-hémagglutinine en travaillant sur des lignées cellulaires de lymphocytes T (CEM).
- l'isolement des MP dans le surnageant du milieu de culture par ultra-centrifugations (45 minutes à 15000 g),
- l'extraction des protéines à l'aide de SDS et séparation par gels SDS-PAGE 1D (gels avec séparation sur 20 cm et gradient de polyacrylamide 8-15 %),
- la découpe systématique des gels en bandes de 2 mm, réduction/alkylation des cystéines puis digestion trypsique avant extraction des peptides,
- analyse nanoLC-MS/MS (CapLC (Waters) couplée à un Q-TOF II (Waters)).

Ces travaux ont permis d'identifier 390 protéines issues de ces préparations de MP induites avec 34 % de protéines membranaires. La répartition des différentes localisations cellulaires des protéines identifiées est représentée dans la figure 15. La répétabilité de cette procédure a également pu être établie avec 70 % d'identifications communes entre 3 préparations différentes de MP.

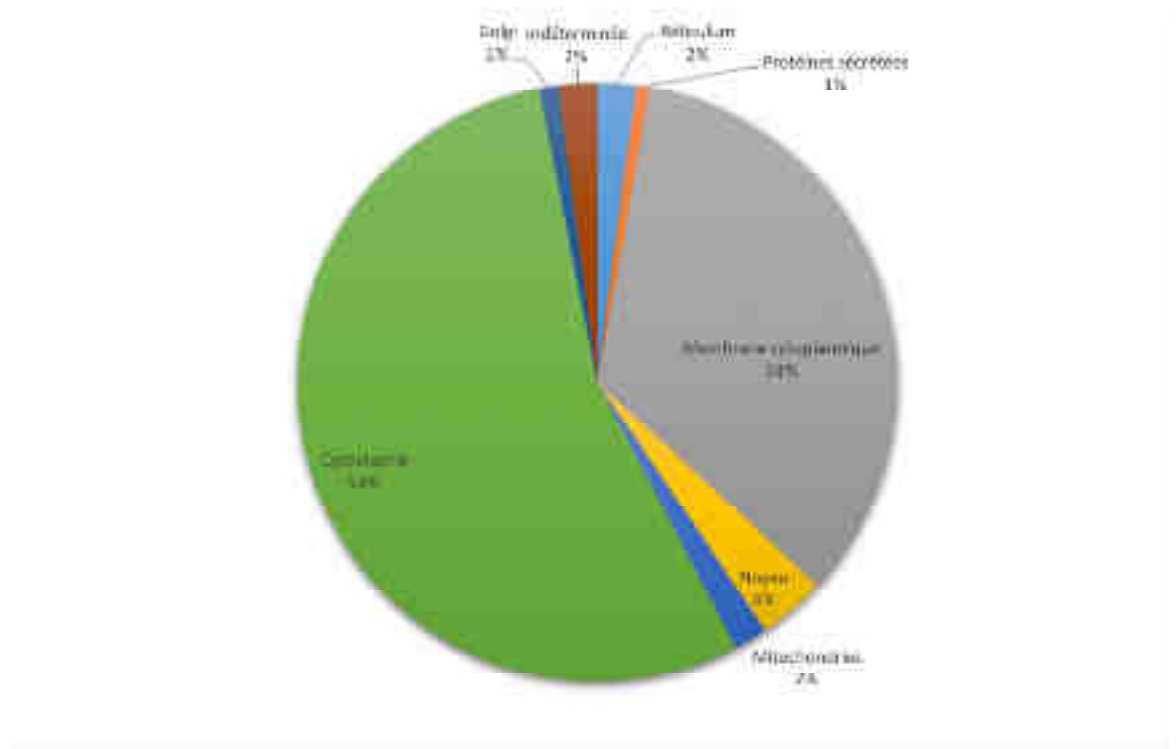


Figure 15
Répartition des localisations cellulaires des protéines identifiées à partir de MP (d'après référence n° 89).

2.2. Identification du CD148

La possibilité d'utiliser les MP induites comme sous-protéome enrichi en protéines membranaires ayant été démontrée, cette stratégie a été appliquée pour la comparaison du protéome des MP issues des différents types de syndromes lymphoprolifératifs : lymphome à cellules du manteau, leucémie lymphoïde chronique et lymphome de la zone marginale.

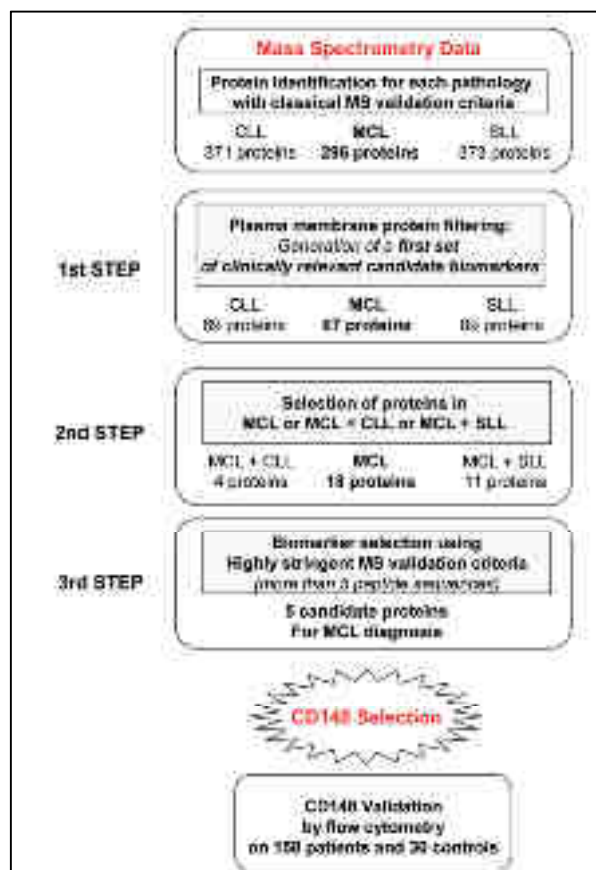
Au total 150 millions de lymphocytes ont été obtenus à partir du sang périphérique et stimulés en culture *in vitro* par l'actinomycine D pendant 18 heures. Après collecte des MP par centrifugations différentielles, les protéines ont été extraites et solubilisées dans un tampon à base de SDS et séparées sur gel SDS-PAGE 1D (gel de 20 cm avec gradient de polyacrylamide 8-15 %). Les mêmes étapes que précédemment ont été réalisées pour la découpe du gel, la digestion des protéines et l'extraction des peptides. Les analyses nanoLC-MS/MS ont été réalisées sur un couplage CapLC (Waters) Q-TOF II (Waters).

Une liste de 579 protéines non redondantes a pu être établie à partir de ces trois échantillons. Différents filtres ont été appliqués pour aboutir à une liste de candidats biomarqueurs : protéines associées à la membrane cytoplasmique et exprimées différemment selon les types d'hémopathies avec applications sur les candidats potentiels de critères stringents de validation (validation manuelle des spectres et au moins cinq séquences peptidiques différentes par protéine identifiée). Une liste finale de cinq protéines a ainsi pu être établie (tableau 5 et figure 16).⁹⁰

Protéine	Nombre de peptides validés		
	LLC	LZM	LCM
Receptor-type tyrosine-protein phosphatase eta (CD148)	0	4	10
Chloride intracellular channel protein 1	0	5	6
HLA class I histocompatibility antigen, A-24 alpha chain	0	0	6
Plexin-A1	0	0	9
Transgelin-2	0	0	6

[Tableau 5](#)

Nombre de peptides identifiés et validés pour les cinq protéines de la liste finale à partir de laquelle a été identifié le CD148.



[Figure 16](#)

Illustration des différents filtres appliqués à partir de la liste initiale de protéines identifiées pour aboutir à la sélection du CD148 (d'après référence n° 90).

La dernière étape a consisté à confronter ces biomarqueurs potentiels avec leurs fonctions biologiques et les données de la littérature. Cette étape a permis d'extraire de cette liste le CD148 (DEP-1/HPTP η) comme un biomarqueur d'intérêt. Il s'agit d'un récepteur membranaire à activité tyrosine phosphatase, caractérisé initialement en 1994.⁹¹ Le domaine extracellulaire de la protéine est composé d'une répétition de 8 à 10 domaines de type fibronectine III. Le domaine intracellulaire ne contient qu'un domaine catalytique à activité tyrosine phosphatase. Le poids moléculaire varie entre 220 et 250 kDa selon l'état de glycosylation. Le gène *PTPRJ*, localisé en 11p11.2, a une taille de 180 kb et comporte 25 exons.

Le CD148 est exprimé dans de nombreux tissus dont la majorité des épithéliums à différenciation glandulaire.⁹² Au sein de la lignée hématopoïétique, l'expression la plus forte est détectée au niveau de la lignée granuleuse, l'expression par les lymphocytes B étant intermédiaire et non modifiée après stimulation⁹³. *In vitro* ce récepteur intervient dans le contrôle de la prolifération, de la migration et de l'adhésion cellulaire.^{94,95} La fonction exacte du CD148 dans les lymphocytes B n'est pas connue, mais des études immunohistochimiques suggèrent qu'il pourrait être un marqueur des lymphocytes B mémoires.⁹⁶

L'expression du CD148 est conservée dans différents types d'adénocarcinomes (mammaire, gastrique ou colique par exemple).⁹² De nombreux travaux attribuent à ce récepteur un rôle suppresseur de tumeur dans le développement de pathologies malignes. Dans des lignées cellulaires humaines de cancer du côlon déficientes en CD148, la réexpression du gène entraîne une diminution de la croissance et de la migration des cellules tumorales.⁹⁷ Chez l'homme, une perte d'hétérozygotie du gène *PTPRJ* a été mise en évidence dans des cas de cancer du côlon, du sein ou de la thyroïde.^{98,99}

2.3. Identification du CD180

Une approche similaire à celle ayant permis l'identification du CD148 a été appliquée dans un deuxième projet de recherche mené dans le laboratoire. Le matériel biologique de départ était toujours constitué de MP induites générées à partir de cellules provenant de trois patients (lymphome à cellules du manteau, lymphome lymphoplasmocytaire et lymphome de la zone marginale).¹⁰⁰

Les différences par rapport au projet précédent ont porté sur :

- l'utilisation d'un couplage nanoLC-MS/MS de type nanoACQUITY UPLC (Waters) couplée à un spectromètre de masse de type Q-TOF Synapt G1 (Waters) plus récent,
- la réalisation de deux injections de chaque bande du gel pour augmenter le pourcentage de couverture du protéome,
- l'utilisation de deux moteurs de recherche (Mascot et OMSSA) en ne conservant que les protéines identifiées conjointement par les deux algorithmes.

Le nombre total de protéines ainsi identifiées était de 2016, 1476 et 1066 pour les échantillons LZM, LLP et LCM respectivement (1977 protéines non redondantes au total). Cette liste de protéines a ensuite été restreinte pour ne conserver que les protéines membranaires associées à un cluster de différenciation et identifiées avec au moins deux peptides. Au final une liste de 59 protéines a pu être établie (annexe 1). Ces résultats ont à nouveau été confrontés aux données biologiques et de la littérature afin de définir dans la liste de protéines les biomarqueurs potentiels les plus pertinents.

Le tableau 6 indique les résultats des comptages de peptides et de spectres obtenus pour le CD180.

	LZM	LLP	LCM
Nombre de peptides	9	0	2
Nombre de spectres	29	0	14

Tableau 6

Nombres de peptides et de spectres obtenus pour l'identification du CD180 dans chaque type de syndrome lymphoprolifératif.

Le CD180 (également appelé RP105 (*Radio Protective 105*)) est une protéine de 95 kDa. C'est un récepteur membranaire appartenant à la famille des récepteurs de type Toll-Like (TLR, *Toll-Like Receptors*), impliqué dans l'activation des lymphocytes B.^{101,102} Cette famille de récepteurs, composée de dix membres chez l'Homme (TLR1 à 10), joue un rôle majeur dans l'immunité innée et acquise. Les voies de signalisation impliquant les TLR font intervenir différentes protéines adaptatrices, dont MYD88 (Figure 17).¹⁰³

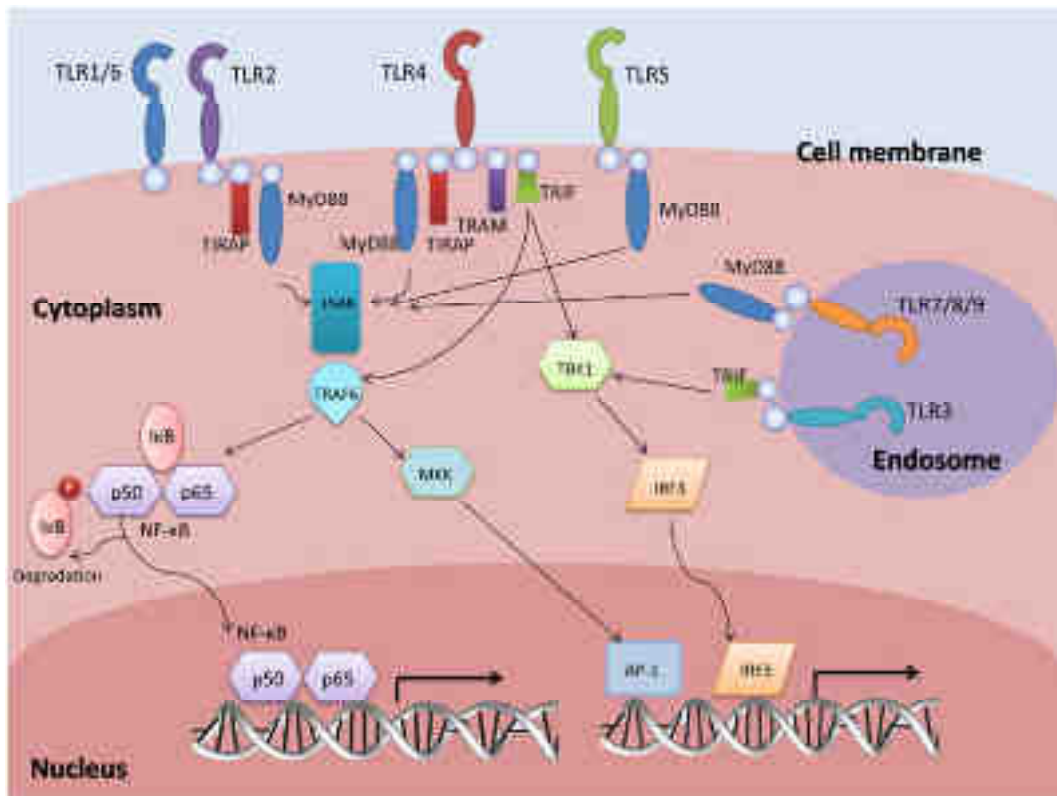


Figure 17
Voies de signalisations impliquant les récepteurs TLR (d'après référence n° 103).

Les lymphomes de la zone marginale, en particulier les formes spléniques, se caractérisent par une activation de la voie NF-κB suite à l'activation des voies de signalisation du récepteur des lymphocytes B (BCR) et des TLRs.¹⁰⁴ Des mutations récurrentes du gène *MYD88*, dont la protéine MYD88 est un adaptateur essentiel dans la voie de signalisation des TLR, ont été décrites dans les lymphomes spléniques de la zone marginale avec une fréquence d'environ 15 %.¹⁰⁵

Il existait donc un rationnel biologique fort pour poursuivre les investigations concernant le CD180 et évaluer son niveau d'expression dans les lymphomes de la zone marginale.

3. ETAPES DE VALIDATIONS

Le CD148 et le CD180 ont été proposés comme biomarqueurs potentiels pour améliorer le diagnostic des lymphomes à cellules du manteau et des lymphomes de la zone marginale respectivement. Les approches protéomiques globales mises en œuvre étaient des approches quantitatives sans marquage par comptage du nombre de peptides et de spectres. La mise en place d'une stratégie de validation était nécessaire puisque d'une part ces approches protéomiques ne permettent pas de quantifier avec précision l'expression de ces marqueurs à la surface des cellules, et, d'autre part, elles ont été réalisées sur un très faible nombre d'échantillons (comparaison dans les deux projets de trois patients différents). La stratégie de découverte s'étant focalisée sur des clusters de différenciation, une validation par cytométrie en flux a pu être réalisée pour ces deux marqueurs.

3.1. Validation du CD148

J'ai pu réaliser lors de mon travail de master la première étape de validation par cytométrie en flux. Les échantillons sélectionnés dans le cadre d'une étude prospective provenant de la banque de cellules congelées du laboratoire d'hématologie des Hôpitaux Universitaires de Strasbourg avaient tous bénéficié d'un phénotypage complet au diagnostic. Les cellules provenaient de prélèvements sanguins ou médullaires. Au total 63 échantillons pathologiques ont été sélectionnés se répartissant en 20 lymphomes de la zone marginale, 21 leucémies lymphoïdes chroniques et 22 lymphomes à cellules du manteau. Pour le groupe témoin, dix échantillons sanguins ont été sélectionnés de façon aléatoire et anonyme parmi du matériel restant d'analyses biologiques. Ces échantillons ne présentaient pas d'anomalie qualitative ou quantitative de l'hémogramme.

Les manipulations ont été effectuées sur un cytomètre de type FACScalibur (Becton Dickinson Bioscience). L'acquisition et l'analyse des données ont été réalisées à l'aide du logiciel CellQuest (Becton Dickinson Bioscience). Deux techniques principales ont été utilisées : des immunomarquages directs utilisant un anticorps anti-CD148 conjugué à la phycoérythrine, et une méthode quantitative basée sur une technique de marquages indirects (méthode Qifikit®).¹⁰⁶ Les deux méthodes utilisées en cytométrie en flux (immunomarquages directs et méthode quantitative indirecte) ont été appliquées à 46 échantillons pathologiques. La corrélation entre ces deux méthodes est très satisfaisante avec une droite de régression ayant un coefficient $R^2 = 0,92$. Ce résultat confirme que les variations d'intensités de fluorescence obtenues par marquage direct avec l'anticorps anti-CD148 couplé à la phycoérythrine sont liées au nombre de sites antigéniques présents à la surface des cellules et reflètent donc l'expression du CD148 à la surface des cellules (Figure 18). Les résultats ont confirmé que l'expression du CD148 est significativement plus élevée dans les lymphomes à cellules du manteau par rapport aux leucémies lymphoïdes chroniques ($p < 0,0001$) et aux lymphomes de la zone marginale ($p < 0,001$) (Figure 19).

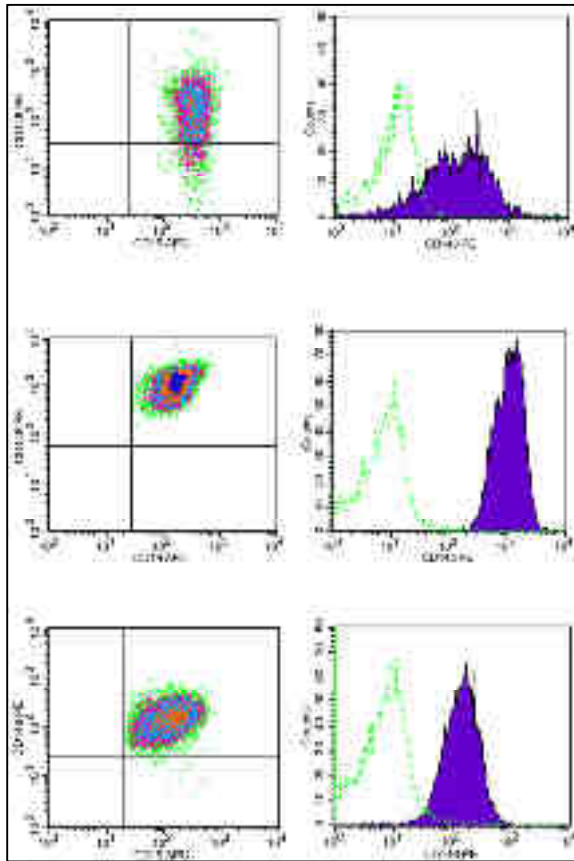


Figure 18 : profils d'expression du CD148 obtenus sur des lymphocytes B de sujets sains (en haut), LCM (au milieu) et L2M (en bas).

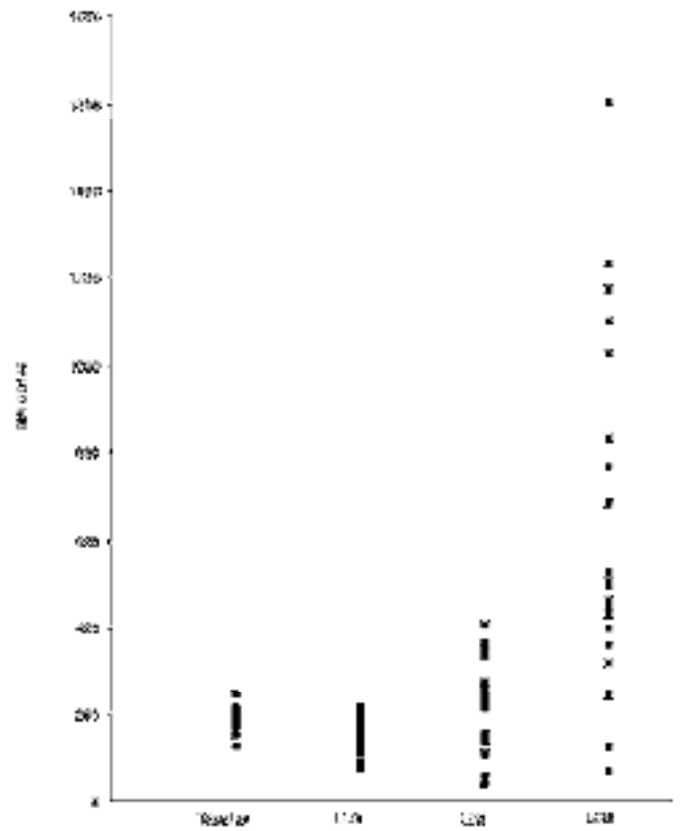


Figure 19 : intensités de fluorescence du CD148 obtenues par CMF chez des sujets sains ; LLC, LZM et L2M. CMF : cytométrie en flux.

Plus récemment, le CD148 a été évalué en association au CD200, également par cytométrie en flux, pour évaluer leurs expressions différentielles entre LLC et LCM. Le CD200 est un marqueur connu pour être surexprimé dans la LLC et sous-exprimé dans les LCM.¹⁰⁷ Les résultats de cette étude confirment la surexpression du CD148 dans les LCM.¹⁰⁸ Un ratio d'intensités de fluorescence CD148 / CD200 > 4,79 permet de discriminer les LCM par rapport aux LLC avec une spécificité de 94,5 % et une sensibilité de 100 %.

3.2. Validation du CD180

Une stratégie similaire de validation à grande échelle par cytométrie en flux a été réalisée pour le CD180. Elle a été réalisée conjointement entre les laboratoires d'Hématologie des CHU de Strasbourg et de Lyon, sur deux séries indépendantes de patients :¹⁰⁰

- une série strasbourgeoise de 96 patients (LLC : n = 30, LCM : n = 16, lymphome lymphoplasmocytaire: n = 15 ; LZM : n = 35),
- une série lyonnaise de 123 patients (LLC : n = 80, LZM : n = 25, LCM : n = 10, lymphomes de la pulpe rouge splénique : n = 8).

L'analyse de l'intensité de fluorescence sur ces deux séries a montré de façon concordante une surexpression du CD180 dans les lymphomes de la zone marginale (Figure 20). La série lyonnaise a pu montrer en particulier une surexpression du CD180 dans une entité rare de syndrome lymphoprolifératif B : le lymphome de la pulpe rouge splénique. Ce lymphome est considéré à l'heure actuelle comme une entité provisoire dans la classification OMS 2008. Son diagnostic reste difficile en l'absence de marqueur discriminant et peut nécessiter dans certains cas l'ablation de la rate pour un examen anatomo-pathologique.

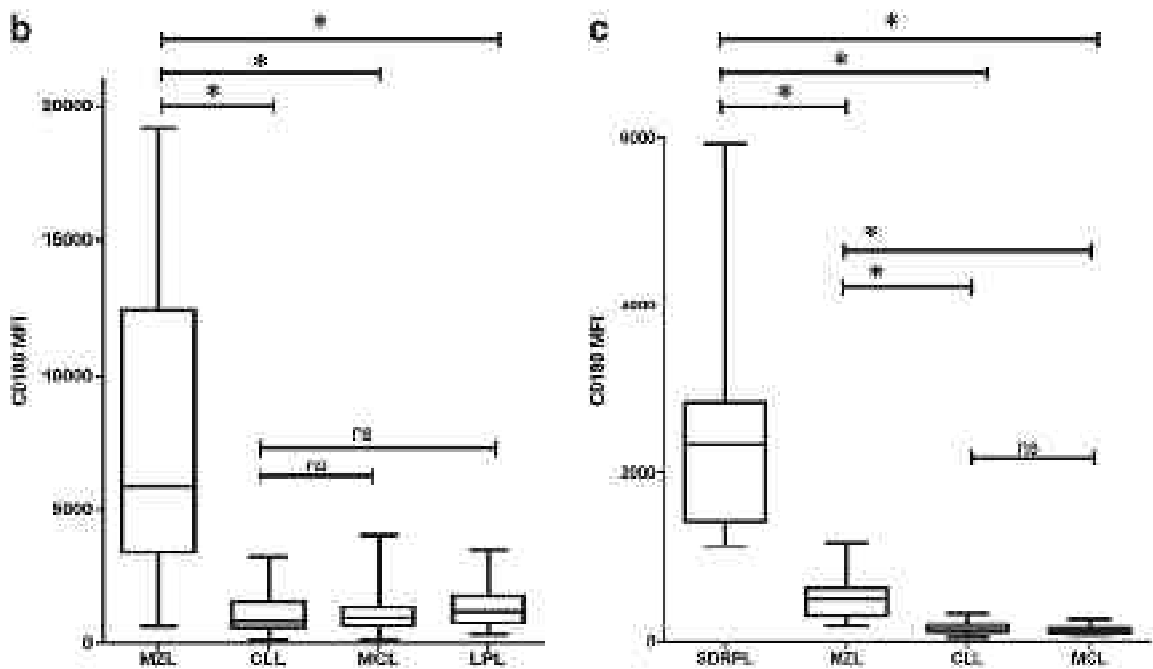


Figure 20

Expression du CD180 dans les différents types d'hémopathies dans la série strasbourgeoise (à gauche) et la série lyonnaise (à droite) (d'après référence n° 100).

MZL : lymphome de la zone marginale ; CLL : leucémie lymphoïde chronique ; MCL : lymphome à cellules du manteau ; LPL : lymphome lymphoplasmocytaire ; SDRPL : lymphome de la pulpe rouge splénique.

En complément de la cytométrie en flux, une quantification du transcrite du CD180 a été réalisée après extraction des ARN. Cette analyse a porté sur 22 patients de la cohorte strasbourgeoise et cinq témoins. Des niveaux d'expression similaires du transcrite ont été retrouvés entre les lymphocytes B pathologiques et les lymphocytes B sains. Compte tenu de ces résultats, une évaluation par cytométrie en flux de l'expression intracellulaire du CD180 a été réalisée sur 15 échantillons de syndromes lymphoprolifératifs et cinq témoins. Il est ainsi apparu que le marquage intracytoplasmique du CD180 ne diffère pas entre les différentes conditions, contrairement au marquage membranaire. Ces données de quantification du transcrite et de marquages intra-cytoplasmiques de la protéine semblent donc indiquer que la synthèse protéique du CD180 n'est pas modifiée, mais font suggérer que les différences observées dans les lymphomes de la zone marginale soient liée à une augmentation du processus d'adressage de la protéine à la membrane cytoplasmique.

Une validation multicentrique du CD180 par cytométrie en flux a également été réalisée au niveau national par l'intermédiaire du Groupe d'Etude Immunologique des Leucémies (GEIL). Cette évaluation a porté sur 236 patients analysés dans cinq laboratoires français avec une standardisation des méthodes de cytométrie. Cette évaluation multicentrique a permis de confirmer la surexpression du CD180 dans les LZM par rapport aux autres hémopathies. Une expression plus importante du CD180 dans les leucémies à tricholeucocytes a également été mise en évidence (Figure 21).¹⁰⁹

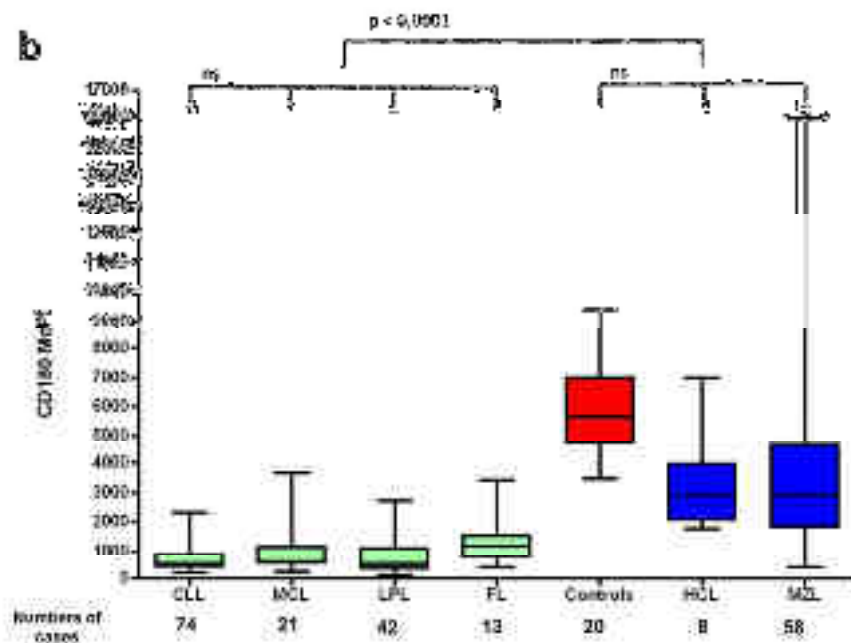


Figure 21

Expression du CD180 dans les différents types d'hémopathies dans l'étude réalisée par le GEIL (d'après référence n° 109). MZL : lymphome de la zone marginale ; CLL : leucémie lymphoïde chronique ; MCL : lymphome à cellules du manteau ; LPL : lymphome lymphoplasmocytaire ; FL : lymphome folliculaire ; HCL : leucémie à tricholeucocytes.

4. CONCLUSIONS

Ces deux projets ont permis d'identifier par des approches protéomiques globales deux biomarqueurs, le CD148 et le CD180, exprimés différemment dans les syndromes lymphoprolifératifs B chroniques. Ces deux biomarqueurs ont bénéficié de stratégies de validation à grande échelle par cytométrie en flux, qui ont permis de confirmer les résultats issus des approches protéomiques.

Les perspectives sont de combiner ces deux marqueurs avec le CD200 pour établir un score par cytométrie en flux permettant de distinguer les principaux syndromes lymphoprolifératifs (LLC, LZM, LCM et lymphome lymphoplasmocytaire). La figure 22 est une représentation de type « radar » qui montre la possibilité d'individualiser ces différentes entités en associant ces 3 marqueurs ainsi que le CD5 et le CD19.

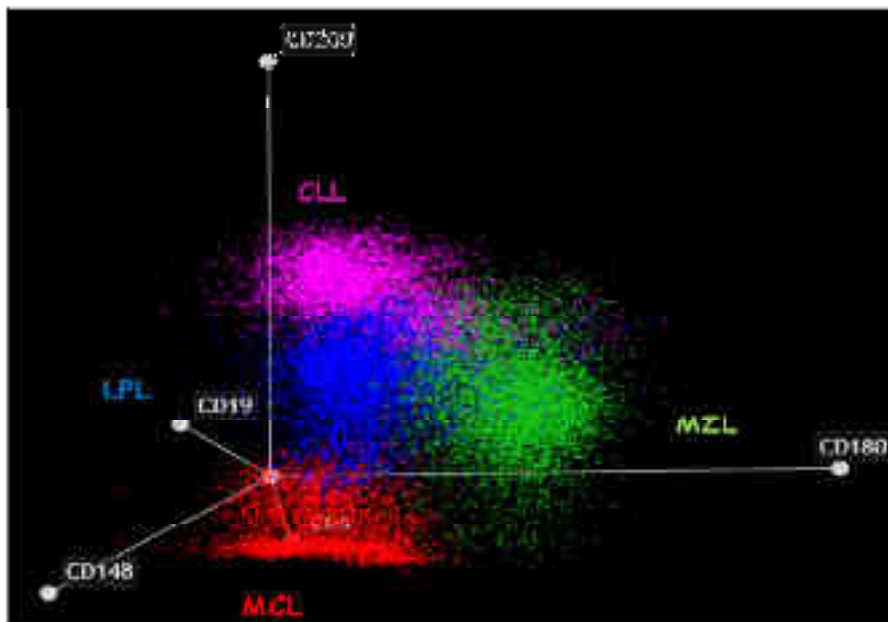


Figure 22

Représentation "radar" des profils d'expression du CD148, CD180, CD200, CD5 et CD19 dans la LLC, le LCM, les LZM et les lymphomes lymphoplasmocytaires (représentation obtenue avec le logiciel Kaluza Analysis 1.3).

MZL : lymphome de la zone marginale ; CLL : leucémie lymphoïde chronique ; MCL : lymphome à cellules du manteau ; LPL : lymphome lymphoplasmocytaire.

Un deuxième projet vise à évaluer des tubes prêts à l'emploi pour la cytométrie en flux contenant une combinaison de 8 anticorps incorporant le CD180 (Projet « DuraClone 180 »). Ce projet est en cours au niveau national par l'intermédiaire du Groupe d'Etude Immunologique des Leucémies (GEIL).

La stratégie de quantification sans marquage utilisée dans ces deux projets était basée sur un comptage du nombre de peptides et de spectres. Les listes de protéines ont ainsi été établies après validation des identifications dans chaque échantillon. Cependant, un biais important de cette méthode de quantification est lié au problème du sous-échantillonnage : une protéine peut ne pas être validée (et donc considérée à tort comme absente dans un échantillon) si les peptides permettant son identification n'ont pas été suffisamment sélectionnés en MS pour être fragmentés. Ce problème du sous-échantillonnage concerne tout particulièrement les protéines de faible poids moléculaire ou présentes en faible abondance. L'importance de ce biais a considérablement diminué avec les spectromètres de masse de dernière génération et le développement des algorithmes de quantification par extraction des courants d'ions.

5. COMMUNICATIONS DES RESULTATS

Ces travaux ont fait l'objet de :

Deux publications (annexes 2 et 3) :

- 1) **Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis.**
L. Miguët, G. Béchade, L. Fornecker, E. Zink, C. Felden, C. Gervais, R. Herbrecht, A. Van Dorsselaer, L. Mauvieux, S. Sanglier-Cianferani.
J. Proteome Res. 8, 3346–3354 (2009)
- 2) **Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas.**
L. Miguët, S. Lennon, L. Baseggio, A. Traverse-Glehen, F. Berger, N. Perrusson, M.P. Chenard, A.C. Galois, A. Eischen, C. Mayeur-Rousse, A. Maar, L. Fornecker, R. Herbrecht, P. Felman, A. Van Dorsselaer, C. Carapito, S. Cianferani, L. Mauvieux.
Leukemia. 2013 Aug;27(8):1748-50

Une communication orale :

- 1) **Les microparticules : un nouvel outil pour proposer par approche protéomique de nouveaux biomarqueurs pour le diagnostic des hémopathies lymphoïdes.**
L. Fornecker, S. Lennon, L. Miguët, C. Carapito, A. Van Dorsselaer, L. Mauvieux, S. Cianferani.
Premières journées scientifiques de la fédération de médecine translationnelle de Strasbourg (Strasbourg, 13-14 juin 2013)

Six communications par affiches (annexe 4) :

- 1) **Strong cell surface expression of the Toll-like receptor homolog CD180 identifies circulating cells of marginal zone lymphoma from other B-cell malignancies.**
L. Miguet, L. Baseggio, S. Lennon, P. Felman, L. Fornecker, A.C. Galois, C. Mayeur-Rousse, R. Herbrecht, A. Eischen, S. Cianferani, L. Mauvieux.
54th Meeting of the American Society of Hematology (Atlanta, USA, 8-11 décembre 2012)
- 2) **Expression du CD180, homologue des récepteurs Toll-like (TLR): marqueur immunologique des cellules circulantes de lymphomes de la zone marginale.**
L. Miguet, S. Lennon, L. Baseggio, A.C. Galois, A. Eischen, C. Mayeur-Rousse, A. Maar, L. Fornecker, R. Herbrecht, P. Felman, A. Van Dorsselaer, C. Carapito, S. Cianferani, L. Mauvieux
33^e Congrès annuel de la Société Française d'Hématologie (Paris, 27-29 mars 2013)
- 3) **Les microparticules : un nouvel outil pour proposer par approche protéomique de nouveaux biomarqueurs pour le diagnostic des hémopathies lymphoïdes.**
L. Fornecker, S. Lennon, L. Miguet, C. Carapito, A. Van Dorsselaer, L. Mauvieux, S. Cianferani.
7^{ème} forum du Cancéropôle Grand-Est (Strasbourg, 25-26 novembre 2013)
- 4) **Interest of the CD148, CD180 and CD200 combination in flow cytometry analyses for mature B-cell neoplasms diagnosis.**
L. Miguet, C. Mayeur-Rousse, S. Lennon, L. Fornecker, C. Gervais, A. Ittel, A.C. Galois, A. Eischen, J.F. Leseve, V. Latger-Cannard, R. Herbrecht, S. Cianferani, L. Mauvieux.
56th Meeting of the American Society of Hematology (San Francisco, USA, 6-9 décembre 2014)
- 5) **Intérêt de l'analyse du triple marquage CD148, CD180 et CD200 par cytométrie en flux dans le diagnostic des hémopathies lymphoïdes B chroniques.**
L. Miguet, C. Mayeur Rouse, L. Fornecker, V. Latger-Cannard, J.F. Leseve, R. Herbrecht, S. Cianferani, L. Mauvieux.
35^e Congrès annuel de la Société Française d'Hématologie (Paris, 01-03 avril 2015)
- 6) **Multicentric analysis of the CD148, CD180 and CD200 combination for the diagnosis of mature B-cell neoplasms using flow cytometry.**
L. Miguet, L. Fornecker, M. Wyrwas, S. Cianferani, R. Herbrecht, J.F. Leseve, V. Latger-Cannard, C. Mayeur-Rousse, L. Mauvieux.
57th Meeting of the American Society of Hematology (Orlando, USA, 5-8 décembre 2015)

• Chapitre 2

Développement de stratégies en protéomique ciblée pour une meilleure caractérisation des syndromes lymphoprolifératifs

1. CONTEXTE MEDICAL ET SCIENTIFIQUE

L'identification de nouveaux biomarqueurs par des approches globales nécessite dans un deuxième temps une étape de validation. Cette étape repose sur deux principes :

- utiliser une technique différente,
- être réalisée sur une cohorte indépendante de patients par rapport à la cohorte utilisée pour la phase de découverte.

D'autres paramètres sont également à prendre en compte pour optimiser cette étape de validation :

- utiliser une technique équivalente à celle qui sera utilisée pour ce biomarqueur en pratique quotidienne,
- une validation multicentrique permet de renforcer la validité de cette étape.

Ces différents critères ont été appliqués pour la validation par cytométrie en flux du CD148 et du CD180. La cytométrie en flux, technique utilisée en routine dans les laboratoires d'hématologie, est particulièrement adaptée pour la validation de biomarqueurs membranaires dans les syndromes lymphoprolifératifs chroniques. Elle présente l'avantage d'être applicable à haut débit avec un coût globalement limité. Certaines contraintes sont cependant inhérentes à cette technique : nécessité de disposer de cellules en suspension ainsi que d'un anticorps monoclonal dirigé contre le marqueur à évaluer. Comme toute approche basée sur une réaction antigène-anticorps, seul un épitope de la protéine cible est reconnu. Malgré le développement de cytomètres multi-couleurs, le suivi simultané de plusieurs marqueurs reste limité par les combinaisons de fluorochromes réalisables.

Dans ce contexte, le développement d'approches protéomiques quantitatives ciblées par SRM apparait comme une alternative pour la validation de biomarqueurs. Cette approche présente certains avantages par rapport à la cytométrie en flux :

- elle ne nécessite pas d'anticorps monoclonal,
- plusieurs épitopes peuvent être suivis pour une même protéine,
- les possibilités de multiplexage permettent de quantifier simultanément jusqu'à 30 protéines en routine.

Nous avons ainsi développé une méthode LC-SRM basée sur le suivi de 6 marqueurs pour évaluer la faisabilité d'une validation de biomarqueurs par ce type d'approche SRM. Les biomarqueurs sélectionnés sont le CD148 et le CD180, auxquels a été rajouté le CD200 sur les données de la littérature. Le CD45 est utilisé en tant que marqueur phénotypique pan-leucocytaire. Nous avons également incorporé le CD5 et le CD23 qui sont deux marqueurs phénotypiques classiquement utilisés dans le phénotypage des syndromes lymphoprolifératifs. Le tableau 7 récapitule le profil d'expression de ces différents marqueurs dans les 3 principaux syndromes lymphoprolifératifs chroniques évalués :

	CD148	CD180	CD200	CD45	CD5	CD23
LLC	+/-	+/-	+	+	+	+
LCM	+	+/-	+/-	+	+	-
LZM	+/-	+	+/-	+	-	-

[Tableau 7](#)

Profils d'expression des six biomarqueurs en fonction du type de syndrome lymphoprolifératif.

2. MISE AU POINT DE LA PROCEDURE EXPERIMENTALE

Toutes les analyses SRM ont été réalisées sur un couplage LC-SRM Ultimate 3000-TSQ Vantage (Thermo Fisher Scientific). L'analyse des données a été faite avec le logiciel Skyline.

2.1. Préparation des échantillons

Le mode de préparation des échantillons retenu pour ce projet est le suivant :

- collecte des lymphocytes à partir du sang total par séparation sur un gradient de densité. Un comptage du nombre de cellules isolées a été effectué à l'aide d'un automate de numération sanguine,
- extraction et solubilisation des protéines à partir du culot cellulaire dans un tampon de type Laemmli (SDS 5 %). Dosage des protéines par méthode RC-DC,
- Dépôt sur gel 1D SDS-PAGE de 50 µg de protéines avec gel de concentration à 4 %. Migration courte pour concentration dans une bande de stacking.

Les étapes de mise au point ont été effectuées à partir de sept prélèvements sanguins effectués chez des témoins sains. Un nombre de 1,5 millions de cellules permet d'extraire en moyenne 50 µg de protéines.

2.2. Développement de la méthode LC-SRM

2.2.1. Choix des peptides protéotypiques

Le choix des peptides protéotypiques a suivi les principes exposés précédemment :

- choix de peptides uniques dans la banque humaine Swiss-Prot,
- utilisation du SRM Atlas pour le choix des peptides lorsque le nombre de peptides déjà identifiés dans nos analyses nanoLC-MS/MS réalisées au cours des projets précédents était limité (cas du CD200 et du CD5) ou absent (CD23),
- utilisation des logiciels SOSUI,¹¹⁰ TMHMM,¹¹¹ PHOBIUS¹¹² et TOP CONS¹¹³ pour la prédiction des domaines trans-membranaires pour éviter de choisir des peptides situés dans ces domaines.

Cas particulier du CD200 : cette protéine présente trois isoformes avec un seul peptide partagé entre les trois. Les isoformes 1 et 3 possèdent 2 peptides uniques alors que l'isoforme 2 n'en possède pas. Il ne sera donc pas possible de quantifier par SRM l'isoforme 2. Les tableaux 8 et 11 récapitulent le nombre de peptides identifiés pour chaque protéine. Au total 49 peptides synthétiques marqués aux isotopes stables (PEPotec) ont été commandés.

	CD148	CD180	CD200_1	CD200_3	CD45	CD5	CD23
Nombre de peptides sélectionnés	10	7	2(1)	2(1)	10	10	7

Tableau 8

Nombres de peptides protéotypiques identifiés pour chaque protéine. Pour le CD200, le peptide partagé entre les deux isoformes est indiqué entre parenthèses.

2.2.2. Validation des peptides marqués et choix des transitions

Contrairement aux peptides AQUA, les PEPotecs ne sont pas quantifiés au moment de leur production. Après réception de ces peptides, une série d'analyses nanoLC-MS/MS est donc nécessaire pour évaluer la concentration de chaque peptide et adapter les dilutions lorsque les peptides sont trop concentrés. A l'issue de ces analyses, le mélange de PEPotecs peut être réalisé. Ces analyses nanoLC-MS/MS permettent également de vérifier la qualité des spectres pour chaque peptide.

Dans un deuxième temps le mélange de PEPotec est injecté sur le couplage LC-SRM pour sélectionner les différentes transitions à suivre pour chaque peptide. Une analyse manuelle et individuelle des pics chromatographiques des différentes transitions obtenues pour chaque peptide a été réalisée (Figure 23).

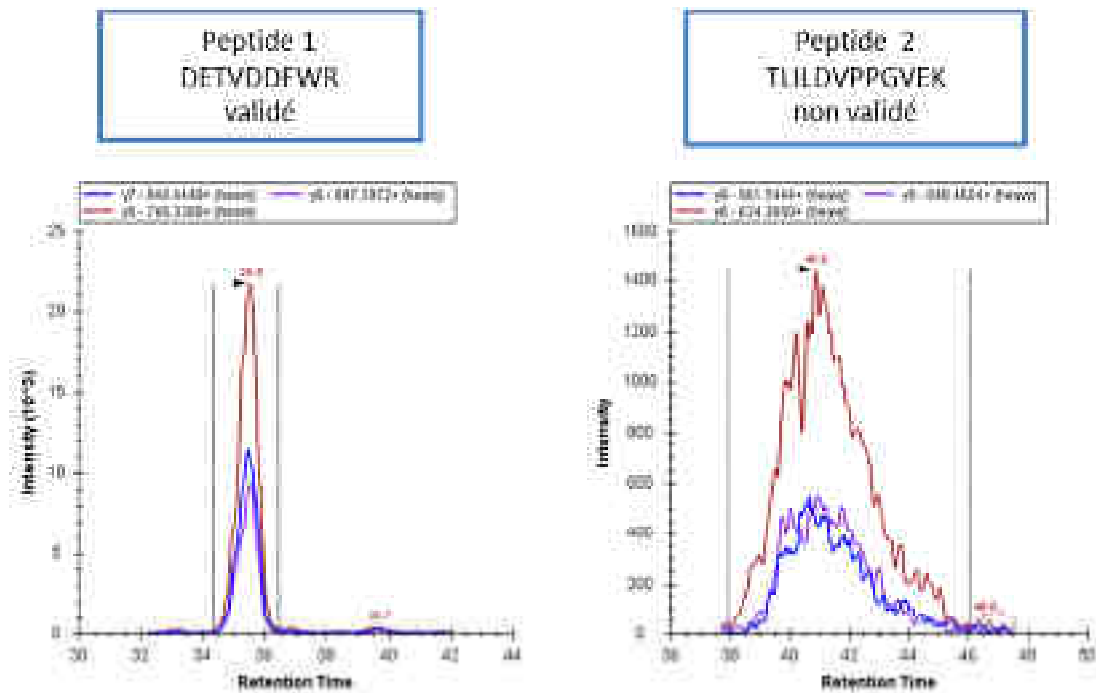


Figure 23
Exemple d'un peptide sélectionné (peptide 1) et d'un peptide non sélectionné (peptide 2).

Les tableaux 9 et 11 indiquent le nombre de peptides finalement validés pour chaque protéine. Au total 38 peptides ont donc été suivis, soit 228 transitions en additionnant les formes légères et lourdes de chaque peptide (les séquences peptidiques et la liste de transitions suivies sont indiquées en annexe 5).

	CD148	CD180	CD200_1	CD200_3	CD45	CD5	CD23
Nombre de peptides validés	8	4	1(1)	1(1)	10	8	5

[Tableau 9](#)

Nombres de peptides protéotypiques validés pour chaque protéine.

2.2.3. Paramétrages de la méthode LC-SRM

Un gradient chromatographique d'une durée de 70 minutes (5 à 35 % d'acétonitrile) a été choisi en raison de la complexité des échantillons.

Le suivi de chaque transition est réalisé de façon cyclique. Le nombre de transitions à suivre et le temps alloué au suivi de chaque transition (temps de scan) définissent le temps de cycle : temps de cycle = (nombre de transitions suivies) x (temps de scan). Deux paramètres sont à prendre en compte pour le paramétrage de la méthode :

- Obtenir idéalement pour chaque peptide dix points de mesure répartis sur la durée du pic chromatographique. La largeur des pics chromatographiques étant de trente secondes environ, le temps de cycle ne doit donc pas dépasser trois secondes
- Le temps de scan ne doit pas être < 10 ms avec l'utilisation du TSQ Vantage

Le mode *scheduled* permet de définir pour chaque peptide une fenêtre de temps sur le gradient chromatographique pendant laquelle les transitions correspondant à ce peptide doivent être suivies. Nous avons pu établir expérimentalement qu'en définissant une fenêtre d'analyse de dix minutes autour du temps de rétention de chaque peptide, environ 120 transitions étaient détectables simultanément.

En tenant compte de ces différents paramètres, pour pouvoir suivre jusqu'à 120 transitions au maximum pendant un cycle, le temps de scan a été fixé à 20 ms. Le temps de cycle est alors de 2,4 secondes.

3. EVALUATION DE LA REPRODUCTIBILITE ET RESULTATS OBTENUS

Pour évaluer la reproductibilité de cette méthode, nous avons utilisé une préparation cellulaire obtenue à partir d'un sujet témoin. Les protéines extraites ont été déposées sur trois gels différents avec trois dépôts par gel. Chacun de ces neuf dépôts a été injecté trois fois, soit un total de 27 analyses.

Un gel stacking à 4 % d'acrylamide a été utilisé avec découpe de deux bandes : bande de concentration et portion du gel située au-dessus de cette bande. Les tableaux 10 et 11 indiquent le nombre de peptides détectés pour chaque protéine. Seuls es peptides des protéines CD5, CD45 et CD148 ont été détectés. Ceci est lié au fait que l'échantillon analysé provenait d'un témoin sain avec donc une population lymphocytaire très majoritairement composée de lymphocytes T, expliquant l'absence de détection de peptides du CD180, du CD200 et du CD23 qui sont exprimés par les lymphocytes B.

	CD148	CD180	CD200_1	CD200_3	CD45	CD5	CD23
Nombre de peptides détectés	5	0	0	0	7	4	0

Tableau 10
Nombres de peptides détectés pour chaque protéine chez un sujet sain.

Le ratio des aires totales entre les formes légères et les formes lourdes a été calculé pour chaque peptide sur les 27 analyses. Un CV de ces ratios a été pour chaque peptide. Aucun peptide ne présente un CV supérieur à 20 %. La médiane de ces CV est de 12 % (figure 24). Ces résultats valident la reproductibilité de la méthode.

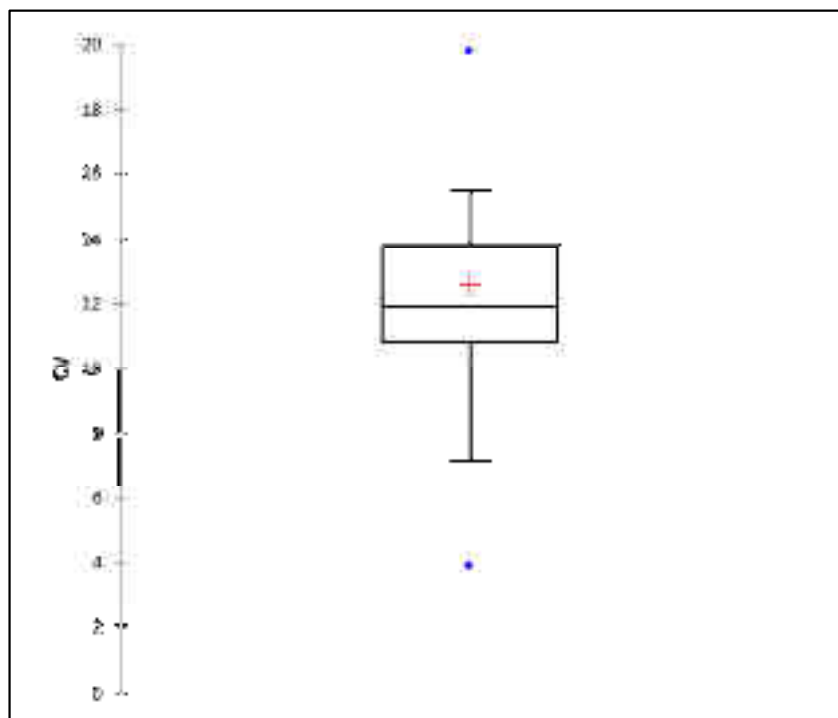


Figure 24
Distribution des CV des ratios des aires totales entre les formes légères et lourdes obtenus sur toutes les analyses pour les 16 peptides détectés.

Protéines	Peptides sélectionnés	Peptides validés	Peptides détectés
CD5	VLDAGDPTSR	X	X
CD5	HLPETEAGR	X	X
CD5	LSWYDPDFQAR	X	
CD5	EHQPLPIQWK	X	X
CD5	NSHLSAYPALEGALHR	X	
CD5	QWEDPSQASK	X	
CD5	VLALLCSGFQPK	X	X
CD5	LVGSSICEGTVEVR	X	
CD5	AQDPGEPR		
CD5	QGAQWAALCDSSAR		
CD23	LPTPSAPLHS	X	
CD23	NEASDLLER	X	
CD23	SQSTQISQELEELR	X	
CD23	HASHTGSWIGLR	X	
CD23	WNDAFCDR	X	
CD23	SQDLELSWNLNGLQADLSSFK		
CD23	SQGEDCVMMR		
CD45	TLILDVPPGVEK	X	X
CD45	VDVYGYVVK	X	X
CD45	NSNVIPYDYNR	X	X
CD45	IEFDNEVDK	X	
CD45	LNPTPGSNAISDVPGER	X	
CD45	YVDILPYDYNR	X	X
CD45	DPPSEPSPLEAEFQR	X	X
CD45	DETVDDFWR	X	X
CD45	YHLEVEAGNTLVR	X	X
CD45	LFLAEFQSIPR	X	
CD148	SQSLSEVLK	X	
CD148	YNNVLPYDISR	X	X
CD148	VITEPIPVSDLR	X	X
CD148	TGTFIAIDR	X	X
CD148	YAAELAENR	X	X
CD148	FSGFEASHGPIK	X	
CD148	GLIDGAESYVSFSR	X	
CD148	DTEVLLVGLLEPGTR	X	X
CD148	LIYQIENENTVDVYGIVYDLR		
CD148	AVSISPTNVILTWK		
CD180	VLDFQNNAIHYISR	X	
CD180	GNHFQDGTITK	X	
CD180	SLEQAINLSLNFNGNNVK	X	
CD180	NLSHLQTLNLSHNEPLGLSQAFK	X	
CD180	ECPQLELLDLAFTR		
CD180	LMNLTFLDLTR		
CD180	GIYLNLAANSINIISPR		
CD200	EQLYTPASLK	X	
CD200	TTINDYQVIR	X	
CD200	GELSQGVQK	X	
CD200	EVICQVLHLGTVTDFK		
CD200	TIGGPLLTATLLGK		

Tableau 11

Récapitulatif des peptides marqués (PEPotec) commandés : séquence, validation et détection chez un sujet sain.

4. CONSTITUTION DE LA COHORTE DE PATIENTS

Cet aspect s'inscrit dans le cadre du projet MALYZOMA qui a obtenu un financement du Cancéropôle Grand-Est. Ce projet a pour objectif de valider et caractériser de nouveaux marqueurs diagnostiques des lymphomes de la zone marginale. Il est coordonné par le Pr Laurent MAUVIEUX (Laboratoire d'Hématologie, Hôpitaux Universitaires de Strasbourg) et a pour objectif de caractériser et valider par cytométrie en flux de nouveaux biomarqueurs sur une série prospective de patients établie entre les CHU de Strasbourg et de Nancy. Le projet SRM développé pendant mon travail de thèse s'inscrit dans ce contexte. Le premier objectif est de valider sur cette cohorte de patients la méthode SRM développée pour la quantification du CD148, CD180 et CD200.

Les premières analyses sur des échantillons de patients ont été réalisées à partir de cellules congelées dans le cadre de la cohorte de patients du projet MALYZOMA. Après décongélation, les cellules sont lavées puis un culot cellulaire est obtenu après centrifugation. L'extraction et la solubilisation des protéines à partir de ce culot cellulaire en utilisant un tampon SDS sont apparues problématique. En effet, pour les premiers échantillons testés, la solubilisation du culot était incomplète avec persistance d'un résidu insoluble malgré de multiples étapes de vortex ou sonication. Ce problème d'extraction et de solubilisation des protéines n'avait pas été rencontré pour la mise au point de la méthode effectuée à partir de cellules non congelées obtenues chez des sujets sains.

Cette étape de préparation des échantillons étant primordiale pour une étude quantitative par SRM, un nouveau procédé de préparation des échantillons a dû être élaboré. L'extraction des protéines s'effectue au laboratoire d'Hématologie avant l'étape de congélation. Les cellules fraîches sont reprises dans le tampon échantillon puis soniquées. Le mélange protéique est ensuite congelé jusqu'à son utilisation pour l'analyse LC-SRM. Ce nouveau procédé de préparation des échantillons est schématisé dans la figure 25. La collecte des échantillons selon cette nouvelle procédure est actuellement en cours.



Figure 25

Schéma de la procédure expérimentale mise en place pour la collecte des échantillons.

- **Chapitre 3**

Recherche de biomarqueurs pronostiques du risque de rechute neuro-méningée dans les lymphomes B diffus à grandes cellules : projet PROTELL

1. CONTEXTE SCIENTIFIQUE ET MEDICAL

Les rechutes neuro-méningées des LBDGC systémiques sont des complications rares concernant environ 5 % des cas. Le pronostic est extrêmement péjoratif avec une médiane de survie globale de l'ordre de six mois.¹¹⁴ Ces atteintes secondaires du système nerveux central (SNC) dans un contexte LBDGC systémiques sont à distinguer des lymphomes primitifs du SNC dans lesquels la maladie est primitivement et uniquement localisée au SNC.

La physiopathologie de ces atteintes primitives ou secondaires du SNC n'est pas parfaitement connue. Le SNC est considéré comme un « sanctuaire immunologique » dépourvu physiologiquement de cellules lymphoïdes. Il est protégé du compartiment sanguin par les barrières hémato-encéphalique (BHE) et hémato-méningée (BHM). La prise en charge des atteintes du SNC reste à l'heure actuelle un défi thérapeutique. Elle est rendue compliquée par l'absence de pénétration dans le SNC des principaux médicaments de chimiothérapie actifs dans les LBDGC. Les données de la littérature sont essentiellement basées sur des études rétrospectives, souvent de faibles effectifs. Cette absence de données basées sur des essais prospectifs randomisés explique l'absence de consensus à l'heure actuelle sur la prise en charge thérapeutique des localisations au SNC des LBDGC. La prise en charge repose sur l'utilisation de certains médicaments (méthotrexate, aracytine) qui, utilisées par voie systémique à hautes doses, traversent ces barrières.

Il apparaît donc nécessaire de pouvoir individualiser précocement, dès le diagnostic, les patients à haut risque de rechutes neuro-méningées pour mettre en œuvre des stratégies de prophylaxie adaptées. Ces traitements prophylactiques reposent principalement sur des injections intra-thécales de méthotrexate réalisées concomitamment à l'administration de chimiothérapies systémiques. Il s'agit néanmoins d'un geste invasif nécessitant la réalisation d'une ponction lombaire.

L'existence de certains facteurs de risque aide le clinicien à mieux individualiser ces patients à haut risque et susceptibles de bénéficier des traitements prophylactiques. Ces facteurs de risque reflètent la prolifération et l'extension de la maladie :

- index pronostique international élevé avec notamment une élévation du taux de LDH, un état général altéré ou une maladie disséminée,
- l'existence d'atteintes lymphomateuses extra-ganglionnaires connues pour être associées à un risque plus élevé de rechutes neuro-méningées : testicules, ovaires, reins, glandes surrénales, glandes mammaires, sphère ORL et base du crâne par exemple.

Ces facteurs de risque divergent cependant selon les études. Ils ne permettent pas d'individualiser de façon optimale tous les patients à risque de rechute neuro-méningée. La mise en évidence de nouveaux biomarqueurs pronostiques du risque de rechute au SNC dans les LBDGC reste donc un enjeu important. C'est dans ce contexte qu'a été mis en place le protocole PROTELL en 2009 dont l'objectif était la recherche de biomarqueurs prédictifs de rechute cérébrale dans les LBDGC.

Ce protocole a combiné différentes approches :

- recherche en protéomique de biomarqueurs dans le liquide céphalo-rachidien (LCR) et le plasma. L'analyse protéomique du plasma était basée sur l'analyse des microparticules plasmatiques circulantes,
- analyse transcriptomique réalisée sur les prélèvements ganglionnaires correspondant à la tumeur primitive.

Deux cohortes de patients ont été réalisées :

- Cohorte 1 : patients avec un diagnostic de LBDGC systémique non antérieurement traité à haut risque de rechute neuro-méningée, défini par :
 - o Score IPI >1 ou présence d'au moins 3 facteurs parmi les suivants:
 - Taux de LDH supérieur à deux fois la valeur normale
 - Nombre de sites extra-ganglionnaires atteints > 1
 - Etat général altéré
 - Présence de signes généraux (fièvre, sueurs nocturnes, amaigrissement)
 - o Soit présence d'une localisation extra-ganglionnaire à risque
- Cohorte 2 (groupe contrôle) avec une atteinte avérée du SNC à l'inclusion :
 - o Lymphomes primitifs du SNC
 - o Rechute neuro-méningée documentée d'un LBDGC systémique

Ce protocole avait pour objectif d'inclure 125 patients. Tous les patients inclus devaient bénéficier d'un prélèvement sanguin (plasma, sérum) et d'un prélèvement de LCR. Ces prélèvements devaient être réalisés avant le début du traitement. Une deuxième série de prélèvements était programmée en fin de traitement.

Ce projet a bénéficié d'un financement INCa et a été coordonné par le Pr Catherine THIEBLEMONT (Hôpital Saint-Louis, Paris). Mon travail de thèse a consisté à réaliser l'analyse protéomique des MP plasmatiques circulantes.

Ce protocole se distingue des projets précédents par :

- une recherche de biomarqueurs à visée pronostique et non plus à visée diagnostique,
- l'utilisation du plasma en tant que matériel biologique de départ. Il s'agit d'un milieu complexe, dépourvu de cellules et contenant plusieurs milliers de protéines avec une gamme dynamique très large,
- l'analyse d'un grand nombre d'échantillons provenant de différents hôpitaux.

2. ETAPES PREPARATOIRES ET STRATEGIE ANALYTIQUE FINALE MISE EN PLACE

2.1. Préparation des échantillons

L'utilisation d'un milieu biologique dépourvu de cellules n'était pas compatible avec l'application d'une stratégie équivalente à celle utilisée pour l'identification du CD148 et du CD180. Ces stratégies étaient en effet basées sur l'utilisation de MP induites qui nécessitent une stimulation et une mise en culture des cellules tumorales pour être générées. Le choix s'est donc porté sur l'analyse des MP plasmatiques circulantes (MPPC).

Ces MPPC sont physiologiquement générées par les différentes cellules présentes dans le sang ou par les cellules endothéliales. Elles sont très majoritairement d'origine plaquettaire.¹¹⁵ La possibilité d'isoler dans le sang des microvésicules (exosomes) provenant de cellules tumorales a été démontrée dans certaines tumeurs cérébrales (glioblastomes).¹¹⁶ La présence dans le sang de MPPC provenant de cellules tumorales a aussi été démontrée au cours des hémopathies.¹¹⁷

L'utilisation des MPPC et le grand nombre d'échantillons à analyser ont nécessité différentes étapes préalables pour établir la procédure expérimentale applicable au projet PROTELL.

2.1.1. Collecte des MPPC

Les MPPC ont été isolées à partir du plasma par centrifugations différentielles. Une première centrifugation (18000 g pendant deux minutes) a été réalisée à partir d'1 mL de plasma pour obtenir du plasma pauvre en plaquettes. A partir de ce plasma pauvre en plaquettes, un culot de MP est obtenu par centrifugation à 18000 g pendant 45 min. Ce culot est ensuite lavé deux fois dans du tampon PBS afin de réduire au maximum la contamination plasmatique. Le culot est finalement repris dans 100 µL de PBS (Figure 26).

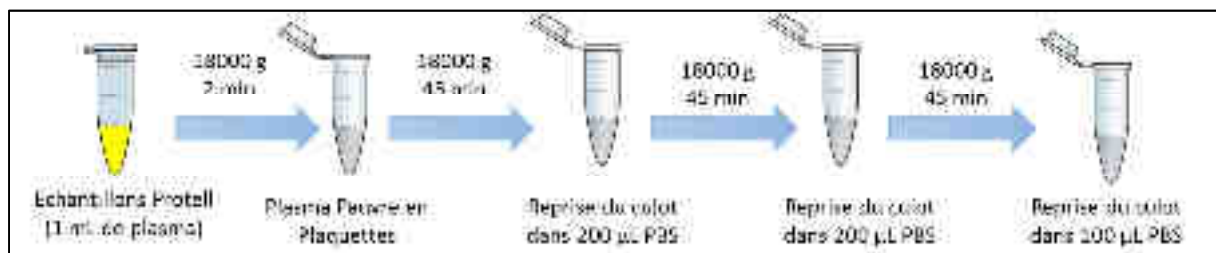


Figure 26
Protocole de collecte des MP plasmatiques circulantes.

2.1.2. Extraction et séparation des protéines

Après collecte des MPPC, les protéines ont été extraites et solubilisées à l'aide d'un tampon à base de SDS. L'étape de séparation des protéines a été considérablement réduite grâce à l'arrivée dans le laboratoire d'un spectromètre de masse de dernière génération de type Q-TOF (Impact HD, Bruker) couplé à un système chromatographique nanoAcquity UPLC (Waters). De plus, compte tenu du nombre d'échantillons, il n'était pas envisageable de réaliser une séparation extensive des protéines sur 90 bandes de gel 1D SDS-PAGE.

Nous avons ainsi évalué trois types de migrations sur gel 1D SDS-PAGE. Le gradient chromatographique a été adapté dans chaque cas pour rester compatible avec le nombre d'échantillons à analyser :

- migration et concentration des protéines dans une bande (gel stacking) avec un gradient de 120 minutes. La bande de concentration ainsi que la partie du gel située au-dessus de cette bande ont été découpées, soit deux bandes analysées au total,
- séparation 6 bandes avec un gradient de 90 minutes par bande,
- séparation 15 bandes avec un gradient de 60 minutes par bande.

Chaque protocole a été préparé en duplicata et deux injections ont été réalisées pour chaque analyse.

Le nombre de protéines identifiées était compris entre 1156 et 1507 avec le gel 6 bandes, entre 647 et 715 avec le gel stacking et entre 768 et 951 avec le gel 15 bandes (Figure 27). Les meilleurs résultats ont donc été obtenus avec une séparation en 6 bandes et un gradient de 90 minutes. Les résultats inférieurs obtenus avec une séparation 15 bandes peuvent s'expliquer par l'utilisation d'un gradient chromatographique plus court.

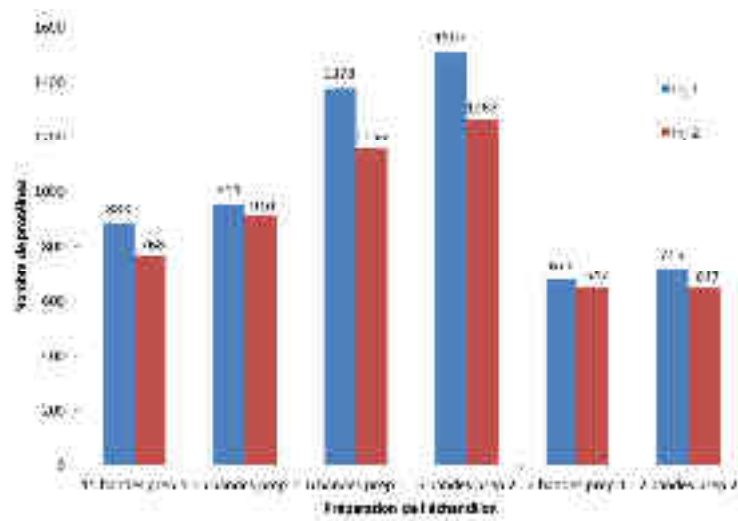


Figure 27
Nombres de protéines identifiées selon le type de séparation réalisée. Inj : injection.

Une analyse quantitative a été réalisée par extraction des courants d'ions à l'aide du logiciel Skyline pour évaluer la reproductibilité de la quantification. Le coefficient de variation médian des intensités obtenues pour chaque protéine sur les douze analyses est de 6 % pour la séparation 2 bandes et de 15 % pour la séparation 6 bandes.

Compte tenu de l'ensemble de ces résultats, une préparation de type stacking avec découpe du gel en 2 bandes et utilisation d'un gradient chromatographique de 90 minutes a semblé être le meilleur compromis entre la durée des analyses, le nombre de protéines identifiées et la reproductibilité de la quantification.

2.2. Suivi des performances du couplage nanoLC-MS/MS

La durée prévisible des analyses étant de plusieurs semaines, la mise en place d'un suivi des performances du couplage nanoLC-MS/MS a été nécessaire. Ce suivi des performances a été assuré par :

- Une injection de BSA après chaque analyse d'un échantillon, précédée d'une séquence de régénération de la colonne.
- La constitution d'un échantillon témoin de MPPC. Cet échantillon identique a été déposé une fois sur chaque gel. Il a suivi l'ensemble du mode opératoire de façon similaire aux échantillons de patients.
- L'ajout en même quantité, uniquement dans les échantillons témoins, d'un mélange commercial de 11 peptides standards (iRT, Biognosis). Ces 11 peptides ont été rajoutés à la fin du processus de préparation des échantillons, juste avant l'injection sur le couplage nanoLC-MS/MS.
- Le suivi d'une masse de référence (lock mass, $m/z = 922,0098$) ionisée en permanence dans l'instrument au niveau de la source.

Il a ainsi été possible de suivre différents paramètres :

- Suivi des paramètres chromatographiques par :
 - o Le suivi des temps de rétentions et des largeurs à mi-hauteur des pics chromatographiques de deux peptides de BSA ($m/z = 722,33$ et $m/z = 464,25$). La limite maximale tolérée pour la largeur à mi-hauteur était fixée à 0,10 min
 - o Le suivi des temps de rétentions des peptides iRT
- Suivi des paramètres de spectrométrie de masse par :
 - o Le suivi de l'intensité et de la résolution des spectres MS des deux peptides de BSA (nécessité d'un nettoyage si baisse de l'intensité en dessous de 10^6 ou de la résolution en dessous de 30000)
 - o Le suivi de l'intensité des peptides iRT
 - o Le suivi de l'intensité de 32 peptides issus du mélange témoin et répartis sur l'ensemble du gradient chromatographique.
 - o Le suivi de la lock mass avec étalonnage de l'instrument nécessaire en cas d'erreur sur la masse supérieur à 20 ppm

Chaque gel comportait cinq échantillons de patients et un échantillon témoin. La figure 28 résume l'ordre dans lequel les différentes injections ont été réalisées.

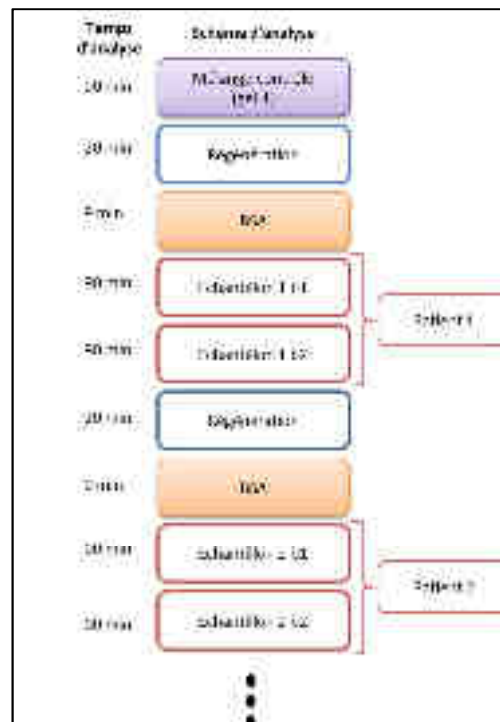


Figure 28
Séquence des différentes injections réalisées pour chaque gel.

La figure 29 résume la procédure expérimentale finalement appliquée à l'ensemble des échantillons. Compte tenu du nombre d'échantillons, une seule injection a été réalisée par analyse.

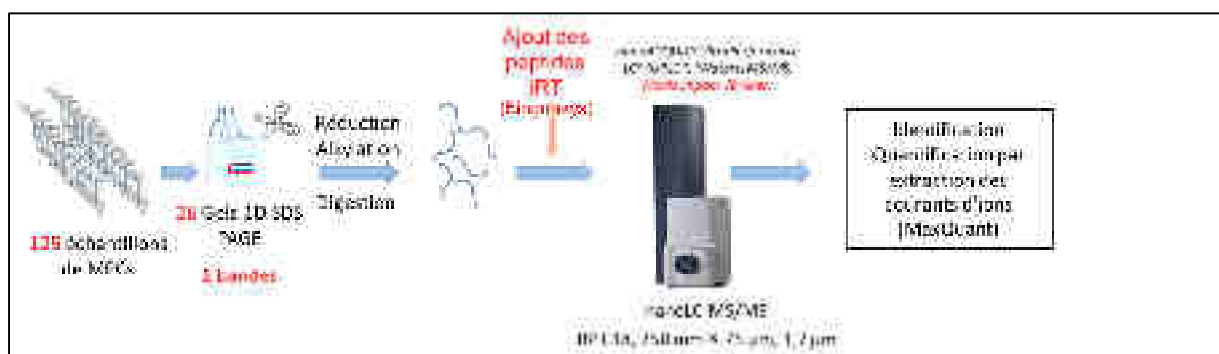


Figure 29
Schéma global de la procédure expérimentale mise en place.

3. RESULTATS

3.1. Constitution des cohortes

147 patients ont été inclus dans le protocole dans dix services cliniques entre décembre 2009 et mai 2012. Parmi ces patients, 19 ont été exclus en raison de données cliniques insuffisantes ou du non-respect des critères d'inclusions). Des MPPC ont pu être générées et analysées pour 102/128 patients. Ces 102 échantillons ont été déposés de façon aléatoire sur 26 gels. Quatre semaines ont été nécessaires pour la réalisation des analyses nanoLC-MS/MS (1 injection par bande). Le tableau 12 indique la répartition des patients en fonction des cohortes.

Cohortes	Patients inclus (n)	Analyse des MPPC (n)
Cohorte 1 : patients inclus au diagnostic du lymphome, sans atteinte neurologique	99	78
1.1 : Patients à haut risque de rechute neuroméningée	48	37*
1.2 : atteinte tumorale de la base du crâne	17	12
1.3 : atteinte d'organes à risque	34	29
Cohorte 2 : patients avec atteinte neurologique au moment de l'inclusion dans le protocole	29	24
2.1 : lymphomes primitifs du système nerveux central	21	21
2.2 : lymphomes systémiques avec rechutes neuro-méningées	8	3
Total	128	102

Tableau 12

Répartition des patients inclus et des patients analysés en MPPC selon les cohortes.

**dont 1 patient remplissant les critères 1.1, 1.2 et 1.3.*

L'actualisation des données cliniques a été faite en janvier 2014. Au total 15 évènements (rechutes cérébrales) ont été observés dans la cohorte 1, soit 15 % des patients. Seuls 9 de ces 15 patients avaient des échantillons disponibles pour l'analyse des MPPC (12 % des 78 patients analysés).

3.2. Suivi des performances du couplage nanoLC-MS/MS

Les différents paramètres analysés ont permis de mettre en évidence une très bonne stabilité du couplage pendant toute la durée des analyses.

Le suivi des paramètres chromatographiques a permis de montrer une stabilité des temps de rétention des deux peptides de la BSA avec des coefficients de variation de 1,2 % et 0,6 % (au total 143 injections de BSA ont été réalisées). A l'exception des 15 premières analyses pour lesquelles la BSA était injectée directement après l'échantillon, sans séquence de régénération de la colonne au préalable, les largeurs à mi-hauteur des pics chromatographiques sont toujours restées inférieures au seuil limite. Concernant les peptides iRTs, sur les neuf peptides détectés, les CV des temps de rétentions sont inférieurs à 1 % à l'exception du peptide le plus hydrophile (CV = 2,2 %).

Le suivi des paramètres de spectrométrie de masse par la surveillance des intensités et des résolutions des deux peptides de la BSA a permis de constater que les seuils limites d'intensités ou de résolutions n'ont jamais été atteints. Le suivi de l'erreur de masse sur la *lock mass* a permis de déterminer quatre périodes nécessitant un étalonnage de l'instrument. Ce point est en grande partie lié à des variations de température dans la pièce en raison de l'installation d'un nouveau spectromètre de masse. Les CV calculés sur les intensités des neuf peptides au cours de chaque semaine d'analyse sont très majoritairement inférieurs à 20 %, à l'exception du peptide le plus hydrophile et du peptide le plus hydrophobe.

Des résultats similaires ont été obtenus avec le suivi des intensités des 32 peptides sélectionnés dans l'échantillon témoin avec des CV très majoritairement inférieurs à 20. S'agissant du même échantillon déposé sur chaque gel, cela confirme la stabilité des échantillons durant toute la période d'analyse.

3.3. Analyse des données dans le cadre du projet PROTELL

Les données ont été traitées en utilisant le logiciel MaxQuant⁷⁰ (version 1.4.3.17) et son moteur de recherche intégré (Andromeda).⁶⁸ Les recherches ont été effectuées dans la banque humaine Swiss-Prot. Les résultats ont été filtrés avec un FDR inférieur ou égal à 1% au niveau peptidique et protéique. Une liste de 1397 protéines a été générée à partir des 102 échantillons analysés.

L'analyse statistique a été coordonnée par l'Institut de Mathématique de Toulouse (Institut National des Sciences Appliquées (INSA) de Toulouse, CNRS – UMR 5219).

Certains choix ont été effectués concernant la méthode de l'analyse statistique :

- Utilisation des valeurs de quantification non normalisées générées par le logiciel MaxQuant. L'objectif était de pouvoir choisir et maîtriser cette étape de normalisation au moment de l'analyse statistique.
- Répartition des patients en 4 groupes :
 - o Cohorte 1 sans rechute cérébrale (n = 69)
 - o Cohorte 1 avec rechute cérébrale (n = 9)
 - o Patients cohorte 2 avec lymphome cérébral primitif (n = 21)
 - o Patients cohorte 2 avec rechute cérébrale d'un lymphome systémique (n = 3)
- Prétraitement des données :
 - o Conservation uniquement des protéines quantifiées dans ≥ 50 % des échantillons d'au moins un groupe (le groupe 2.2 (cf tableau 12) a été exclu en raison du faible effectif) : 693 protéines sont conservées
 - o Normalisation des échantillons par ajustement des médianes : les médianes de chaque échantillon ont été ajustées à partir d'une médiane de référence
 - o Imputation des valeurs manquantes à partir de la distribution des minima de chaque échantillon

Trois comparaisons ont été effectuées

- Comparaison 1 : Cohorte 1 avec rechute cérébrale (n = 9) versus cohorte 1 sans rechute (n = 69)
- Comparaison 2 : Cohorte 2 globale (n = 24) versus cohorte 1 sans rechute (n = 69)
- Comparaison 3 : Cohorte 2.2 (n = 3) versus cohorte 1 sans rechute (n = 69)

En utilisant le test non paramétrique de Mann-Whitney (choisi en raison du faible échantillonnage lié à la rechute dans la cohorte 1), le nombre de protéines significatives ($p < 0.05$) par comparaison est :

- 12 pour la comparaison 1 (tableau 13)
- 63 pour la comparaison 2 (tableau 14)
- 12 pour la comparaison 3 (tableau 15)

Cependant, après ajustement des probabilités par la méthode de Benjamini-Hochberg,¹¹⁸ aucune protéine ne reste significative dans ces 3 comparaisons.

Protéine	p	Différence
2-oxoglutarate dehydrogenase complex component E2	0,0094	2,47
Protein FAM162A	0,0233	1,31
Reelin	0,0264	-0,89
Aconitase	0,0264	3,26
CDGSH iron-sulfur domain-containing protein 2	0,0297	3,31
Coronin-1B	0,0321	2,78
Serum albumin	0,0334	0,52
Ficolin-3	0,0334	0,77
Gelsolin	0,0453	0,18
Oncoprotein-induced transcript 3 protein	0,0453	-2,5
Ig heavy chain V-III region BRO	0,0488	0,54
Ig heavy chain V-III region HIL	0,0488	0,9

Tableau 13

Protéines significatives pour la comparaison 1.

Protéine	p	Différence
HLA class I histocompatibility antigen, Cw-12 alpha chain	0,0016	3,09
Antithrombin-III	0,0024	-0,35
Ig kappa chain V-I region Lay	0,0026	1,3
Transthyretin	0,0035	-0,29
Beta-Ala-His dipeptidase	0,0035	-0,95
Tryptophan--tRNA ligase, cytoplasmic	0,0061	2,02
Zinc-alpha-2-glycoprotein	0,0064	-0,35
Elongation factor Tu, mitochondrial	0,0064	-2,56
Angiotensinogen	0,0066	-0,42
Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	0,0069	-1,99
Biotinidase	0,0081	-0,72
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	0,0083	-0,81
Programmed cell death 6-interacting protein	0,009	0,86
Hemopexin	0,0097	-0,36
Alpha-2-antiplasmin	0,0108	-0,3
Ig kappa chain V-I region Wes	0,011	-0,55
Selenoprotein P	0,011	-0,58
Apolipoprotein A-II	0,0113	-0,6
Serum albumin	0,0119	-0,39
Plasma serine protease inhibitor	0,0122	-0,45
NAD-dependent malic enzyme, mitochondrial	0,0138	-2,41
Preylcysteine oxidase 1	0,0141	-1,02
NAD(P) transhydrogenase, mitochondrial	0,0152	-1,28
Plastin-2	0,0159	0,62
Ribonuclease inhibitor	0,0167	2,37
Thiosulfate sulfurtransferase	0,018	-2,99
Kallistatin	0,0184	-0,22
Cullin-associated NEDD8-dissociated protein 1	0,0188	0,56

Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	0,0193	0,91
Citrate synthase, mitochondrial	0,0197	-1,05
Gelsolin	0,0197	-0,14
Xaa-Pro aminopeptidase 1	0,0217	0,55
HLA class I histocompatibility antigen, B-46 alpha chain	0,0222	0,97
Retinol-binding protein 4	0,0227	-0,52
Kinesin-1 heavy chain	0,0249	-3,96
Cathelicidin antimicrobial peptide	0,026	0,75
Rho GTPase-activating protein 18	0,026	2,04
Apolipoprotein A-I	0,0272	-0,36
Receptor-type tyrosine-protein phosphatase C	0,0272	0,59
Voltage-dependent anion-selective channel protein 2	0,0291	-0,92
Myosin light chain kinase, smooth muscle	0,0298	2,39
Complement component C7	0,0304	-0,26
Alkyldihydroxyacetonephosphate synthase, peroxisomal	0,0311	-1,45
Apolipoprotein C-IV	0,0311	0,92
Solute carrier family 2, facilitated glucose transporter member 1	0,0325	1,01
Transferrin receptor protein 1	0,0347	0,56
Carboxypeptidase B2	0,0347	-0,32
FH1/FH2 domain-containing protein 1	0,0363	0,72
Putative heat shock protein HSP 90-beta 2	0,037	2,4
Aldehyde dehydrogenase family 16 member A1	0,037	0,95
Mannose-binding protein C	0,0378	-0,49
Ig heavy chain V-II region NEWM	0,0387	-0,51
Afamin	0,0395	-0,32
Annexin A5	0,0412	1,72
L-lactate dehydrogenase B chain	0,043	0,4
Pyruvate kinase PKM	0,0439	0,11
Microtubule-associated protein RP/EB family member 2	0,0439	0,89
Vitamin K-dependent protein C	0,0448	-0,37
Protein diaphanous homolog 1	0,0458	0,85
Phosphate carrier protein, mitochondrial	0,0467	-1,43
Low affinity immunoglobulin gamma Fc region receptor II-a	0,0477	2,79
N-acetylmuramoyl-L-alanine amidase	0,0477	-0,25
Staphylococcal nuclease domain-containing protein 1	0,0487	1,83

Tableau 14
Protéines significatives pour la comparaison 2.

Protéine	p	Différence
FH1/FH2 domain-containing protein 1	0,0166	3,32
Phosphatidylcholine-sterol acyltransferase	0,0179	-1,1
Mannose-binding protein C	0,0208	-1,28
Prenylcysteine oxidase 1	0,0208	-1,73
Plectin	0,026	3,19
Thrombospondin-1	0,03	2,13
Anoctamin-6	0,0322	3,5
Alpha-2-antiplasmin	0,0397	-0,77
PRA1 family protein 3	0,0425	2,54
Platelet glycoprotein Ib alpha chain	0,0425	2,27
MICOS complex subunit MIC60	0,0454	1,61
Ig kappa chain V-I region Wes	0,0485	-0,46

Tableau 15
Protéines significatives pour la comparaison 3.

4. CONCLUSIONS

Ce projet a montré pour la première fois au sein du laboratoire la possibilité de réaliser des analyses protéomiques quantitatives sans marquage par spectrométrie de masse sur une centaine d'échantillons. Les procédures mises en place pour le suivi des performances du couplage nanoLC-MS/MS ont démontré la stabilité de ce couplage sur une période de quatre semaines.

L'analyse statistique et l'interprétation des résultats sont néanmoins compliquées. Différentes protéines apparaissent significatives dans les trois comparaisons effectuées, mais cette significativité disparaît après ajustement par la méthode de Benjamini-Hochberg. La fragilité des résultats obtenus est illustrée par la mise en évidence dans la première comparaison de protéines significatives comme l'albumine ou des immunoglobulines.

Différents aspects peuvent expliquer ces résultats et seront discutés dans la troisième partie de ce manuscrit :

- Choix du matériel biologique de départ
- Constitution des cohortes avec un faible nombre de rechutes cérébrales observées dans la cohorte 1
- Difficultés rencontrées pour l'analyse statistique en raison de ces problèmes d'effectifs et des valeurs manquantes

5. COMMUNICATION DES RESULTATS

Ces travaux ont fait partie de deux communications orales :

- 1) **Increase of sensitivity with the nano-ESI CaptiveSpray source associated to nanoBooster (Bruker).**

Plumel, M., Bertaccini, D., Hovasse, A., Van Dorsselaer, A. & Schaeffer-Reiss, C.

13th HUPO World Congress : The proteome quest to understand biology and disease (2014).

- 2) **Evaluation of quantitative proteomics performance and robustness on a benchtop UHR-QTOF platform.**

A. Van Dorsselaer. Bruker Users Meeting (25 novembre 2014).

- **Chapitre 4**

Vers l'analyse d'un grand nombre d'échantillons : optimisation des méthodes de préparation par une approche « tube-gels »

1. CONTEXTE SCIENTIFIQUE

1.1. Préparation des échantillons : principes généraux

1.1.1. Multiplicité de protocoles d'extraction des protéines et de digestion trypsique

Le développement des différentes approches protéomiques quantitatives globales ou ciblées a été possible grâce aux progrès importants réalisés sur le plan instrumental. Cependant, en amont de l'analyse LC-MS/MS, la préparation des échantillons est une étape primordiale pour la réussite de ces approches. C'est pendant cette étape que sont extraites et solubilisées les protéines avant que soit réalisée la digestion enzymatique sur laquelle reposent les stratégies protéomiques de type « bottom-up ». Les deux méthodes les plus utilisées consistent à réaliser cette digestion enzymatique en solution (digestion liquide) ou dans un gel de polyacrylamide. Quelle que soit la méthode employée, l'objectif est d'obtenir une préparation d'échantillons reproductible pour garantir la justesse et la reproductibilité des résultats fournis.

Les détergents assurent la lyse des membranes cellulaires et la solubilisation des protéines. Ces détergents forment en solution des micelles dont la partie centrale permet d'établir des liaisons hydrophobes avec les protéines. Ils permettent ainsi la dénaturation des protéines et la rupture des interactions protéines-protéines. Différents types de détergents sont disponibles en fonction des caractéristiques de leur extrémité hydrophile : ioniques, non-ioniques, zwitterioniques ou chaotropiques.

Le SDS est un détergent ionique. Sa liaison aux protéines dénature leur conformation 3D et leur confère une charge négative expliquant son utilisation largement répandue pour les gels d'électrophorèse. Les fortes concentrations de SDS utilisées pour l'extraction et la solubilisation des protéines ne sont cependant pas compatibles avec le maintien de l'activité enzymatique de la trypsine et ne sont pas compatibles avec les systèmes de chromatographie liquide ou les analyses MS (suppression de signal). La migration de l'échantillon sur un gel de polyacrylamide et le lavage des bandes découpées permettent de retirer le SDS. Les protéines contenues dans les bandes de gel peuvent ensuite être soumises à la digestion enzymatique.

Les agents chaotropiques permettent de dénaturer les protéines en rompant les liaisons non covalentes. L'urée est un agent chaotrope fréquemment utilisé en amont d'une digestion tryptique en solution. Une haute concentration en urée (8 M) est généralement utilisée pour dénaturer les protéines extraites. La trypsine reste active jusqu'à une concentration en urée de 1 M. Une simple dilution des extraits avant ajout de la trypsine rend donc cette stratégie compatible avec une digestion tryptique en solution. Après digestion, les peptides en solution sont purifiés et concentrés par extraction en phase solide (SPE, *Solid-Phase Extraction*).

La littérature contient un grand nombre de protocoles variés de digestion tryptique liquide ou dans un gel, multiplicité qu'il est raisonnable d'expliquer par le fait qu'à chaque mélange protéique correspond certainement un protocole adapté et idéal. L'utilisation du SDS à forte concentration pour extraire un maximum de protéines, suivie d'une migration des protéines sur un gel de polyacrylamide (SDS-PAGE) est incontestablement une des méthodes prédominantes et très largement utilisée pour les analyses protéomiques par spectrométrie de masse. En parallèle de l'utilisation du gel 1D pour réaliser un fractionnement protéique en amont de l'analyse LC-MS/MS, l'utilisation de la pré-concentration d'extraits protéiques dans une seule bande de gel (gel stacking) s'est développée depuis plusieurs années.¹¹⁹ L'approche gel stacking permet de conserver les avantages d'une extraction réalisée avec du SDS, sans fractionner l'extrait protéique et donc compromettre la qualité de quantification qui suit. Néanmoins la préparation des gels stacking nécessite du temps pour la réalisation du gel de polyacrylamide suivie de l'électrophorèse. Ce temps de préparation peut représenter un obstacle pour l'analyse d'un grand nombre d'échantillons (par exemple, 26 gels 1D SDS-PAGE avaient été nécessaires pour l'analyse des échantillons du projet PROTELL).

1.1.2. Influence du protocole de préparation sur les modifications des protéines et des peptides

Pendant la préparation des échantillons, les protéines et les peptides peuvent être soumis à 3 principales modifications :

- 1) Les atomes de soufre des méthionines peuvent être oxydés pour former un sulfoxyde de méthionine. Cette modification est un processus physiologique *in vivo*¹²⁰ mais peut également apparaître *in vitro* et augmente avec la durée de conservation des échantillons. Une proportion variable de méthionines peut ainsi être oxydée d'un échantillon à un autre (modification qualifiée de variable). Du fait de cette variabilité aléatoire inter-échantillons, ces peptides modifiés ne sont généralement pas pris en compte pour la quantification.
- 2) Lors de la préparation des échantillons, les protéines sont dénaturées par réduction des ponts disulfures établis entre les résidus cystéines. Cette réduction est réalisée à l'aide de dithiothréitol (DTT) puis les résidus cystéines sont alkylés par l'iodoacétamide (IAA) pour éviter la reconstitution de ces ponts disulfures. Cette dénaturation des protéines, en supprimant leur conformation 3D, favorise leur migration dans un gel de polyacrylamide et leur digestion enzymatique. L'alkylation des groupements thiols des résidus cystéines par l'IAA provoque une carbamidométhylation de ces peptides. Cette modification concerne la quasi-totalité des résidus cystéines dans un échantillon : elle est donc souvent considérée comme une modification fixe pour la recherche dans les banques de données.
- 3) Une autre modification peut concerner les résidus cystéines si un gel de polyacrylamide est utilisé. Les monomères libres d'acrylamide réagissent avec les résidus cystéines et provoquent l'apparition d'une modification de type propionamide. Cette modification ne survient donc qu'en cas d'utilisation d'un gel de polyacrylamide et peut entrer en compétition avec la carbamidométhylation de ces mêmes résidus cystéines lors de l'étape d'alkylation réalisée dans un second temps. La proportion de modifications propionamides diminue si les gels de polyacrylamide sont réalisés la veille de leur utilisation. Cela permet d'assurer une polymérisation quasi-complète de l'acrylamide, ce qui diminue la quantité de monomères d'acrylamides libres résiduels dans le gel susceptibles de réagir avec les résidus cystéines. Il s'agit donc d'une modification variable puisqu'elle ne concerne pas l'ensemble des résidus cystéines.

1.2. Méthode de préparation de type « tube-gel »

Cette technique repose sur la polymérisation d'un gel de polyacrylamide directement dans l'échantillon en solution. Elle est réalisée après l'étape d'extraction et de solubilisation des protéines.

Cette technique présente différents avantages :

- Un gain de temps considérable pour l'analyse d'un grand nombre d'échantillons en raison de l'absence de nécessité de couler un gel au préalable et de réaliser une migration électrophorétique des protéines ;
- Une diminution du nombre d'étapes préparatoires et donc une diminution du risque de perte d'échantillon ;
- La possibilité d'utiliser le SDS comme détergent pour l'extraction et la solubilisation des protéines. Le SDS est retiré après polymérisation du gel par des étapes de nettoyage, de façon équivalente à ce qui est réalisé avec un gel 1D SDS-PAGE classique.

La première description dans la littérature de cette préparation par tube-gel (TG) date de 2005. Elle est apparue comme une technique particulièrement adaptée pour l'analyse des protéines membranaires en raison de sa compatibilité avec divers détergents.¹²¹ Par la suite différentes publications ont utilisé cette approche pour l'analyse de protéines membranaires¹²²⁻¹²⁶ ou des radeaux lipidiques.¹²⁷ Dans le cadre de stratégies protéomiques quantitatives, les tube-gels ont été utilisés avec des marquages iTRAQ,¹²² ou sans marquage par comptage de spectres.¹²⁷

L'objectif de ce projet était de comparer ce mode de préparation tube-gel avec les approches conventionnelles reposant sur un gel 1D SDS-PAGE (gel stacking) ou une préparation des échantillons en solution (digestion liquide).

2. PROCEDURE EXPERIMENTALE

Une expérience permettant de comparer trois protocoles de préparation des échantillons pour une analyse quantitative « shotgun » a été construite : un protocole de digestion liquide avec une solubilisation des protéines à base d'urée (classiquement utilisé au laboratoire), un protocole de gel stacking et un protocole optimisé de tube-gel. Ces trois modes de préparation ont été comparés en utilisant un mélange protéique standard de référence. Celui-ci est constitué de 48 protéines humaines mélangées en quantité équimolaires (*Universal Proteomics Standard*, UPS1, Sigma) qui ont été ajoutées dans un lysat total de protéines de levure.

Une gamme comportant six points de concentrations différentes en protéines UPS1 a été réalisée. Les protéines de levure constituent une matrice de fond complexe constante entre chaque point de la gamme. Ces six points de gamme ont été ajustés pour injecter sur le couplage nanoLC-MS/MS des quantités connues de protéines UPS1 : 500 amol, 1 fmol, 2,5 fmol, 5 fmol, 10 fmol et 25 fmol.

Cette comparaison des trois protocoles repose sur l'utilisation d'un mélange initial identique de protéines. L'avantage de l'utilisation du SDS pour une meilleure extraction des protéines de levure n'a donc pas été évalué dans cette configuration puisque cette utilisation aurait été incompatible avec le protocole de digestion liquide subséquent. Les protéines de levure ont ainsi été extraites dans un tampon urée (8 M) / bicarbonate d'ammonium (0,1 M).

Pour chaque point de la gamme, le même mélange initial de protéines UPS1 et de protéines de levure a été utilisé pour la réalisation des trois protocoles. Des triplicatas de préparations ont été réalisés pour chaque protocole.

L'utilisation de ce mélange standard de référence commun comportant une gamme de concentration en protéines UPS1 a permis de réaliser une comparaison qualitative et quantitative des 3 protocoles. Les analyses quantitatives ont été réalisées par extraction des courants d'ions MS1. La figure 30 schématise l'ensemble du protocole expérimental mis en place pour ce projet. Les gels 1D SDS-PAGE ont été préparés la veille du dépôt des échantillons.

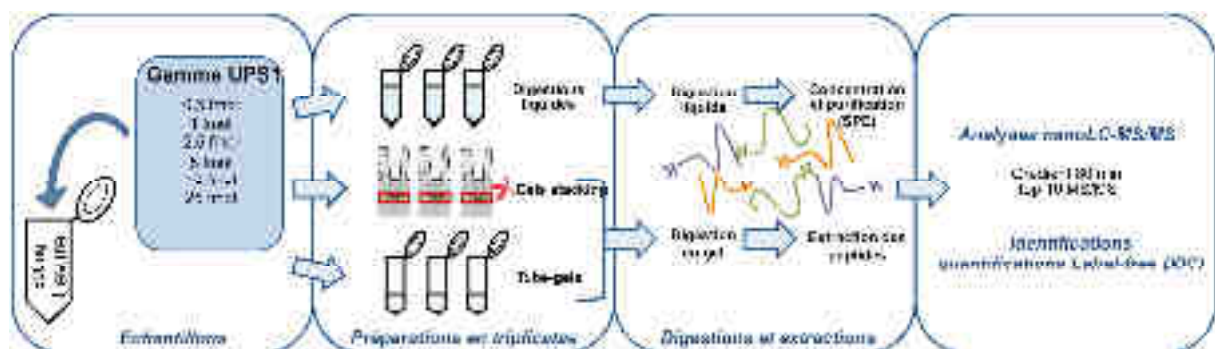


Figure 30
Schéma de la procédure expérimentale mise en place pour la comparaison des trois protocoles.

Les analyses ont été effectuées sur un instrument de type Q-Exactive Plus (Thermo Fisher Scientific). Les différents points de gamme, préparés en triplicatas pour chaque protocole, ont été injectés de façon aléatoire en monoplicat d'injection. Au total, 54 injections ont donc été réalisées. Les données ont été acquises en mode DDA. Pour chaque analyse MS, les dix ions les plus intenses ont été sélectionnés pour être fragmentés. Le temps d'exclusion dynamique était de 60 secondes. Les paramètres chromatographiques (nanoACQUITY UPLC, Waters) comportaient un gradient de 80 minutes de 1 à 35 % d'acétonitrile.

Les données de chaque protocole ont été analysées de façon indépendante avec le logiciel MaxQuant⁷⁰ et son moteur de recherche intégré (Andromeda).⁶⁸ Les séquences des 48 protéines UPS1 ont été concaténées à la banque Swiss-Prot restreinte à la taxonomie *Saccharomyces cerevisiae* (7806 entrées à la date du 16 avril 2015). Une recherche en banque target - decoy a été utilisée pour évaluer les taux de faux positifs lors de l'identification. Les intensités normalisées des protéines, générées par le logiciel MaxQuant, ont été utilisées pour la quantification.

Le tableau 16 indique le paramétrage des modifications prises en compte pour l'identification des peptides pour chaque protocole :

Modifications	Digestions liquides	Gels stacking et Tube-gels
Oxydation (M)	Variable	Variable
Carbamidométhylation (C)	Fixe	Variable
Propionamide (C)	/	Variable

[Tableau 16](#)

Paramétrage des modifications fixes et variables pour chaque protocole.

La quantification s'est faite dans tous les cas sur les peptides non modifiés. Les peptides comportant des méthionines oxydées ainsi que leur équivalent non modifié ont été systématiquement exclus de la quantification. La carbamidométhylation des résidus cystéines a été paramétrée en modification fixe dans le cas de la digestion liquide (du fait de l'absence de gel d'acrylamide, les résidus cystéines ne peuvent pas être sujets à une modification propionamide et sont donc considérés comme étant tous carbamidométhylés après l'étape d'alkylation). Les modifications propionamide et carbamidométhylation ont été paramétrées en modifications variables et utilisées pour la quantification dans les deux protocoles basés sur l'utilisation d'un gel de polyacrylamide. Ce paramétrage a permis d'évaluer la proportion de chacune de ces modifications selon le type de gel réalisé. Les résultats ont été validés avec FDR inférieur ou égal à 1 % au niveau peptidique et protéique en utilisant une approche target - decoy.

3. RESULTATS

3.1. Identifications protéiques et peptidiques

3.1.1. Identifications au niveau protéique

Le nombre de protéines de levure identifiées a été supérieur avec les préparations TG (en moyenne 1979 sur les 18 analyses, extrêmes de 1886 à 2057) par rapport aux digestions liquides (en moyenne 1323 sur les 18 analyses, extrêmes de 970 à 1440) et aux gels stacking (en moyenne 1788 sur les 18 analyses, extrêmes de 1699 à 1854).

La figure 31 indique le nombre de protéines de levure identifiées pour chaque réplicat de chaque protocole, aux six points de gamme réalisés. Cette figure illustre également la bonne reproductibilité des identifications du protéome de levure au sein de chaque mode de préparation, à l'exception des points de gamme à 25fmol préparés par digestion liquide. Il n'y a pas d'explication univoque concernant cette observation. A cette concentration élevée de protéines UPS1, la possibilité d'une compétition entre les protéines de levure et les protéines UPS1 pour la sélection des ions à fragmenter peut être évoquée. Le fait d'observer ce phénomène sur les trois répliques techniques et l'ordre d'injection aléatoire des échantillons peuvent permettre d'éliminer un artéfact lié à la préparation des échantillons.

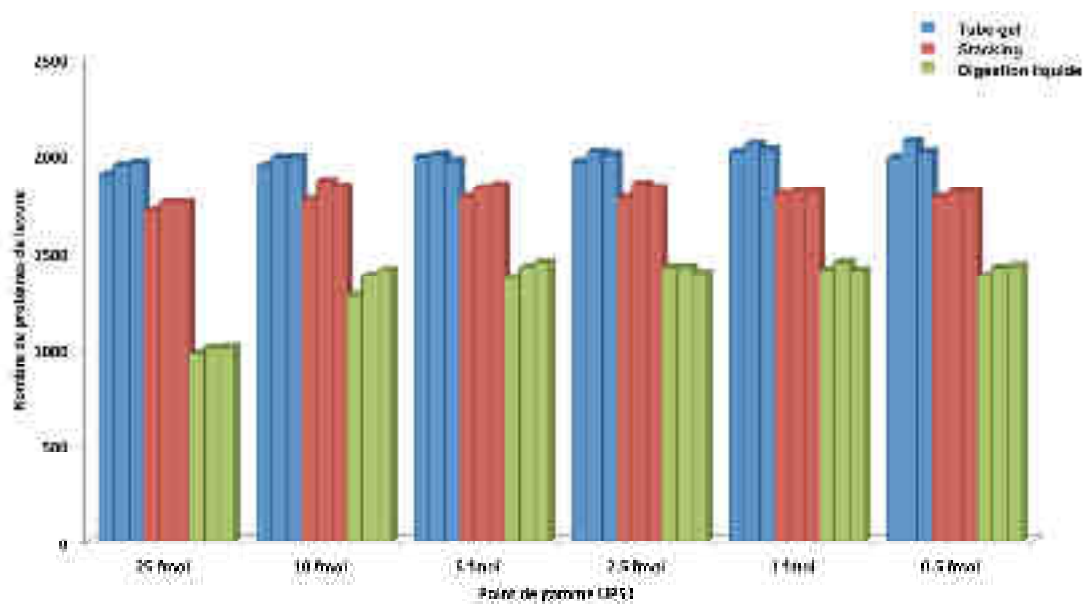


Figure 31
Nombres de protéines de levure identifiées dans chaque réplicat des différents protocoles.

Au total, sur l'ensemble des 54 injections, 2761 protéines de levures non redondantes ont été identifiées avec 55 % d'identifications communes aux 3 protocoles. Les deux protocoles basés sur un gel ont identifié 76 % de protéines communes. La proportion de protéines identifiées uniquement dans un des trois protocoles est la plus élevée avec la préparation TG (10 %) (Figure 32).

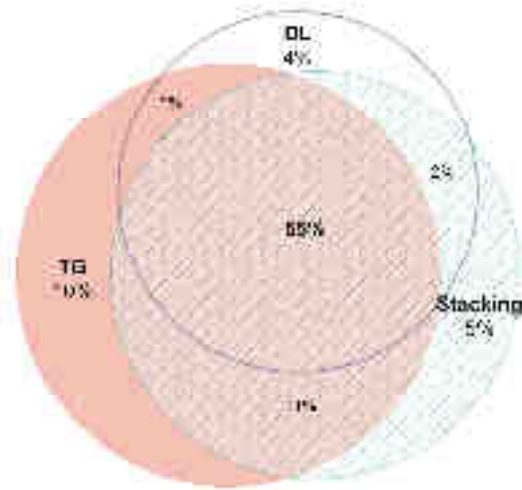


Figure 32

Diagramme de Venn indiquant la proportion de protéines de levure identifiées avec chaque protocole (nombre total de protéines = 2761).

TG : tube-gels, DL : digestions liquides.

L'analyse des caractéristiques des protéines identifiées uniquement dans 1 des 3 protocoles montre une proportion plus importante de protéines de petite taille avec la digestion liquide. La préparation TG permet quant à elle d'obtenir une proportion plus importante de protéines de grandes tailles (Figure 33). L'interprétation de ces résultats est cependant limitée par l'absence d'utilisation de SDS pour l'extraction et la solubilisation des protéines dans les deux protocoles à base de gel.

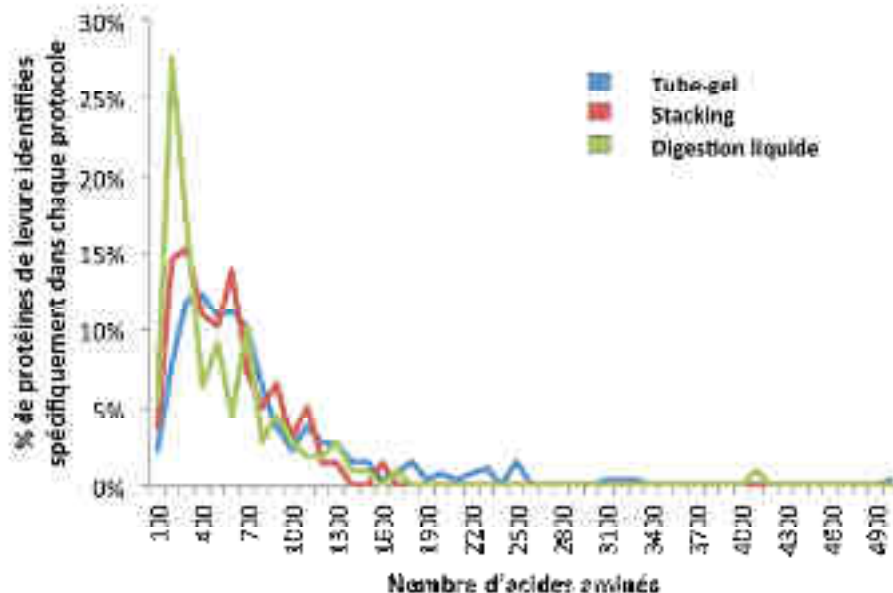


Figure 33

Pourcentage de protéines de levure identifiées spécifiquement dans chaque protocole en fonction de la taille de la séquence.

La distribution des localisations cellulaires des protéines de levure identifiées a été évaluée sur l'ensemble des 18 analyses réalisées pour chaque protocole à l'aide des annotations GO (*Gene Ontology*).¹²⁸ La grande majorité des protéines identifiées sont d'origine cytoplasmique ou nucléaire. La distribution est globalement similaire entre les 3 modes de préparation (Figure 34), cependant un nombre un peu plus important de protéines membranaires ou du réticulum endoplasmique est observé avec les préparations TG. Cette tendance est un peu plus prononcée en analysant la distribution cellulaire des protéines spécifiquement identifiées dans chacun des 3 protocoles (Figure 35).

Il faut cependant à nouveau tenir compte dans l'interprétation des résultats de l'absence d'utilisation de SDS pour l'extraction initiale des protéines : l'utilisation de SDS pour l'extraction des protéines aurait pu permettre d'obtenir une proportion plus importante de protéines membranaires (protéines hydrophobes) avec les protocoles à base de gel.

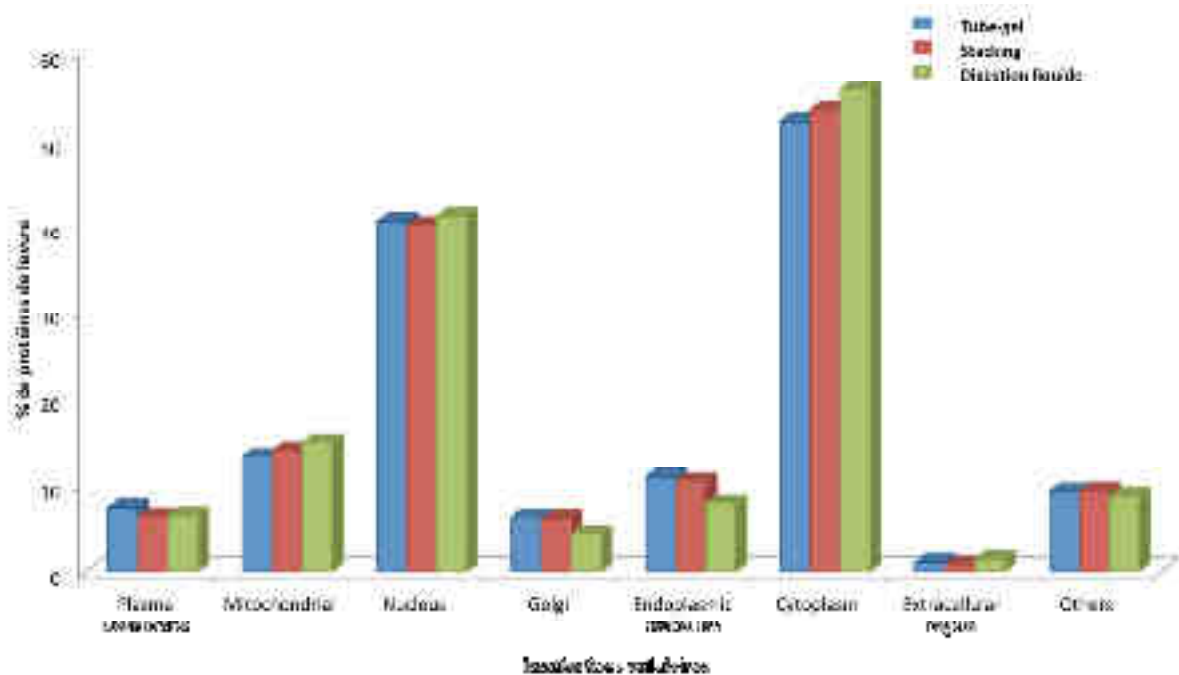


Figure 34
Distribution des localisations cellulaires pour l'ensemble des protéines de levure identifiées dans chaque protocole.

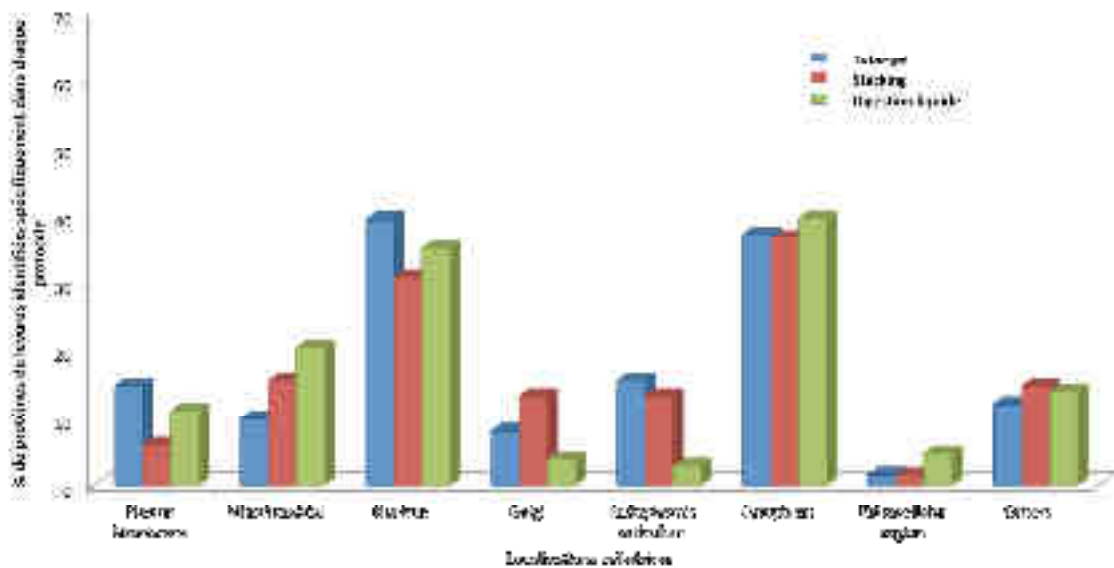


Figure 35
Distribution des localisations cellulaires des protéines de levure identifiées spécifiquement dans chaque protocole.

Presque toutes les protéines UPS1 ont été identifiées dans chaque protocole aux trois points de gamme les plus élevés (25, 10 et 5 fmol). L'ensemble des 48 protéines UPS1 n'a cependant été identifié que dans les protocoles TG et stacking au point de gamme 25 fmol. En dessous de 5 fmol, le nombre de protéines UPS1 identifiées diminue progressivement jusqu'à 12, 12 et 15 pour la digestion liquide, les gels stacking et les TG respectivement. (Figure 36).

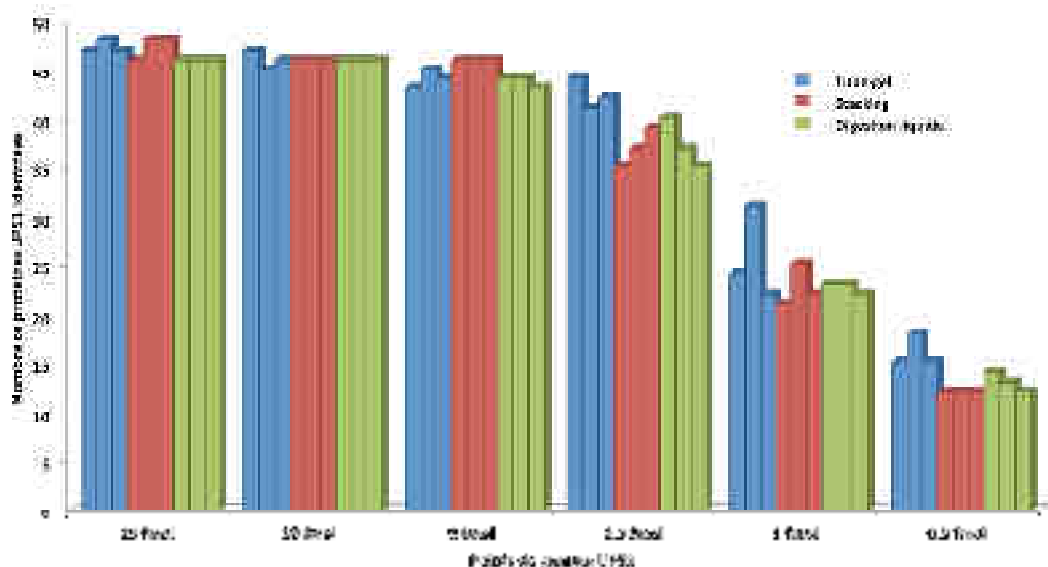


Figure 36
Nombre de protéines UPS1 identifiées dans chaque analyse réalisée.

Le nombre plus élevé de protéines UPS1 identifiées avec le protocole TG, en particulier aux points bas de la gamme, peut s'expliquer par la réduction du nombre d'étapes dans la préparation des échantillons. Cette réduction du nombre d'étapes diminue le risque de perte d'échantillon.

3.1.2. Identifications au niveau peptidique

Le nombre de peptides de levure identifiés dans les 18 analyses de chaque protocole était beaucoup moins important avec les digestions liquides (12528) par rapport aux préparations à base de gel (environ 20000 peptides). Seuls 21 % de tous les peptides identifiés sont communs aux trois protocoles. Les deux protocoles à base de gel ont 50 % de peptides identifiés en commun (Figure 37).

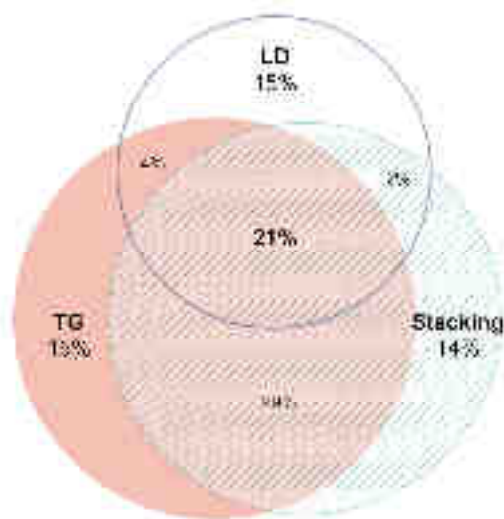


Figure 37

Diagramme de Venn indiquant le pourcentage de peptides de levure identifiés pour chaque protocole.

La répartition des modifications observées au niveau des peptides est représentée sur la figure 38. Un nombre plus important de résidus cystéines avec une modification propionamide a été observé dans les TG par rapport aux gels stacking (306 versus 119 respectivement). La proportion de peptides contenant un résidu cystéine avec une modification propionamide est de 5 % dans les gels stacking et 12 % dans les TG. Cette proportion plus importante dans les TG s'explique par le processus de polymérisation de l'acrylamide qui a lieu directement au contact de l'échantillon. Les monomères d'acrylamide sont donc directement au contact des peptides. Toutefois, cette proportion de peptides avec un résidu cystéine porteur d'une modification propionamide reste modérée et n'a pas constitué un obstacle pour leur quantification.

Une proportion plus importante de méthionines oxydées est également présente dans les TG par rapport aux gels stacking (27 % versus 17 % respectivement). Cette proportion plus importante de méthionines peut être expliquée par le contact direct entre les peptides et le persulfate d'ammonium, un agent oxydant fort utilisé pour la réaction de polymérisation (Figure 38).

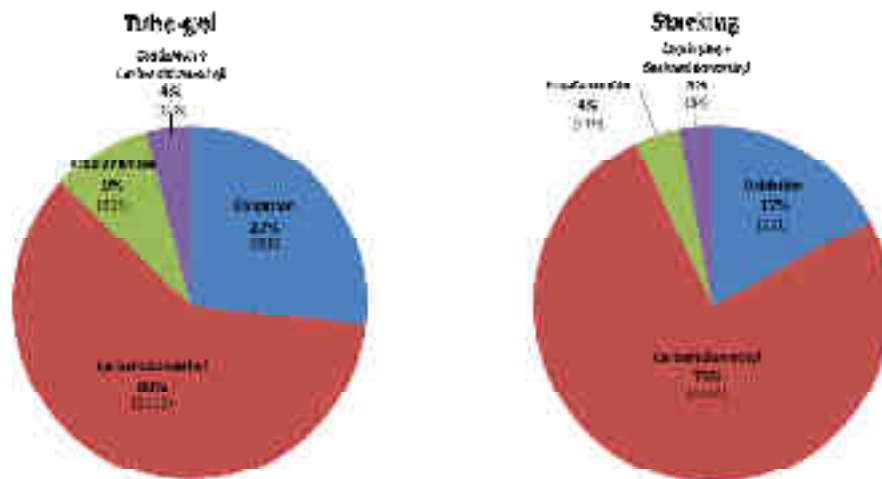


Figure 38
Fréquences des différentes modifications pour l'ensemble des modifiés.

3.2. Analyses quantitatives

L'utilisation d'une gamme de concentration de protéines UPS1 a permis de comparer ces trois protocoles pour évaluer leur linéarité de quantification, leur reproductibilité et leur capacité à détecter des protéines variantes (UPS1) au sein d'une matrice complexe.

Trois comparaisons de points de la gamme ont été réalisées, mimant des situations variées de ratios et d'abondances :

- Comparaison A : 25 fmol contre 2,5 fmol ; cette comparaison permet d'évaluer la quantification sur des protéines ayant de grandes variations de concentrations entre deux échantillons (ratio de concentration = 10)
- Comparaison B : 25 fmol contre 5 fmol ; cette comparaison permet d'évaluer la quantification sur des protéines ayant des variations de concentrations intermédiaires (ratio de concentration = 5)
- Comparaison C : 10 fmol contre 5 fmol ; cette comparaison permet d'évaluer la quantification sur des protéines ayant de faibles variations de concentrations (ratio de concentration = 2)

Les points de gamme à 1 fmol et 500 amol n'ont pas été utilisés pour établir ces comparaisons en raison du nombre plus faible de protéines UPS1 identifiées à ces concentrations.

3.2.1. Linéarité de quantification

3.2.1.1. Linéarité de quantification au niveau peptidique

La linéarité de quantification des peptides UPS1 a été évaluée par le calcul du coefficient de corrélation (R^2) entre l'intensité des peptides et la quantité de protéines UPS1 présente dans l'échantillon. Ce coefficient R^2 a été calculé pour tous les peptides ayant au moins deux valeurs d'intensités sur les trois réplicats dans au moins trois points de gamme. La figure 39 montre un exemple de coefficient R^2 obtenu pour un peptide UPS1 (LSVEALNSLTGEFK) avec le protocole TG.

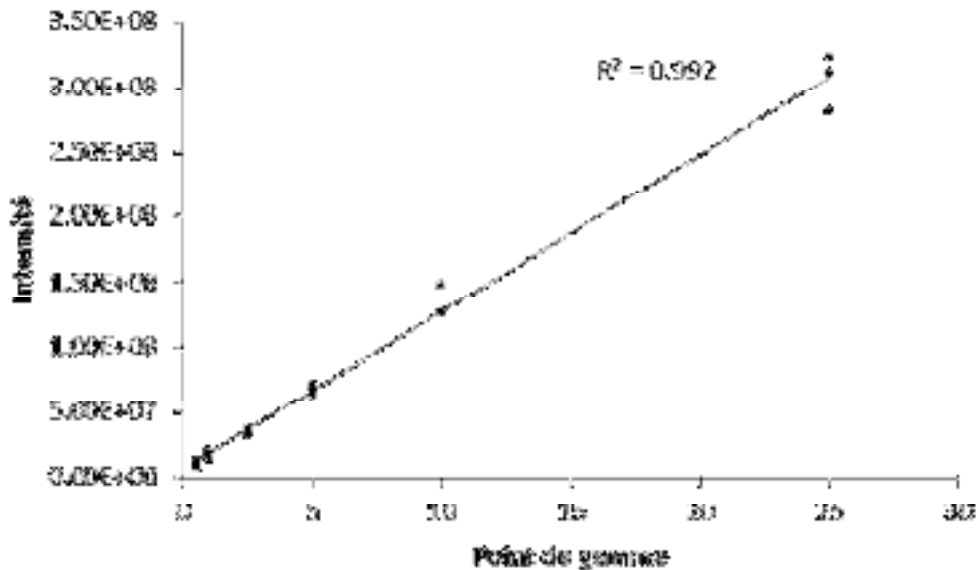


Figure 39

Exemple de coefficient R^2 obtenu pour le peptide LSVEALNSLTGEFK dans le protocole TG.

Le coefficient R^2 est $\geq 0,9$ pour 79 % des peptides avec la préparation TG contre 69 % et 64 % pour les gels stacking et les digestions liquides respectivement (figure 40). Une meilleure linéarité de quantification au niveau peptidique a donc été obtenue avec la préparation tube-gels.

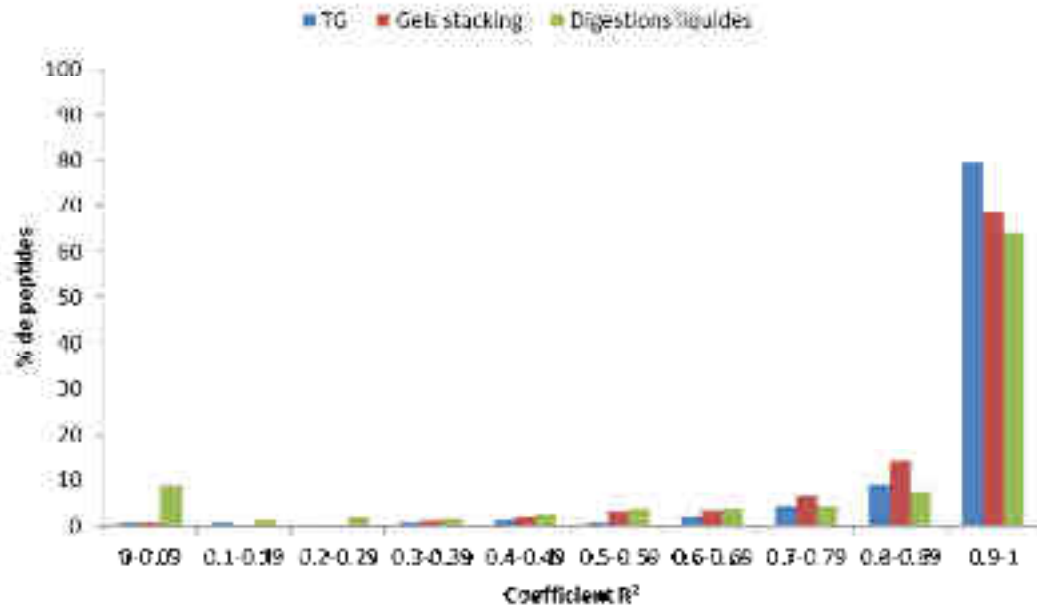


Figure 40
Distribution des coefficients de corrélation pour chaque protocole.

3.2.1.2. Linéarité de quantification au niveau protéique

La linéarité de quantification des protéines UPS1 a été évaluée par le calcul des ratios de concentrations obtenus dans chaque comparaison. Les protéines n'ayant pas au moins deux valeurs d'intensité normalisée pour chaque point de gamme n'ont pas été prises en compte pour ce calcul. Ce ratio de concentration a été calculé à partir du rapport des moyennes des intensités LFQ pour chaque point de gamme. Un exemple de calcul est donné dans le tableau 17 pour la protéine UBE2C avec le protocole TG (comparaison A : ratio de concentration théorique = 10)

	Point de gamme 25 fmol			Point de gamme 2,5 fmol		
	Réplikat 1	Réplikat 2	Réplikat 3	Réplikat 1	Réplikat 2	Réplikat 3
Intensités LFQ (x10 ⁷)	75,7	100,9	97,4	4,0	12,4	9,9
Moyennes des intensités (x10 ⁷)	91,3			8,8		
Ratio de concentration calculé	91,3 / 8,8 = 10,4					

Tableau 17
Exemple de calcul du ratio de concentration pour la protéine UBE2C dans le protocole TG pour la comparaison A (ratio de concentration théorique=10).

Les distributions des ratios de concentrations calculés pour les protéines UPS1 sont représentées dans la figure 41 pour chaque comparaison en fonction des trois protocoles. Pour la comparaison A (25 versus 2,5 fmol, ratio théorique = 10), la distribution des ratios calculés est davantage centrée autour de la valeur théorique pour le protocole TG. Pour les comparaisons B (25 versus 5 fmol, ratio théorique = 5) et C (10 versus 5 fmol, ratio théorique = 2), les ratios calculés sont distribués de façon équivalente et satisfaisante pour les trois protocoles.

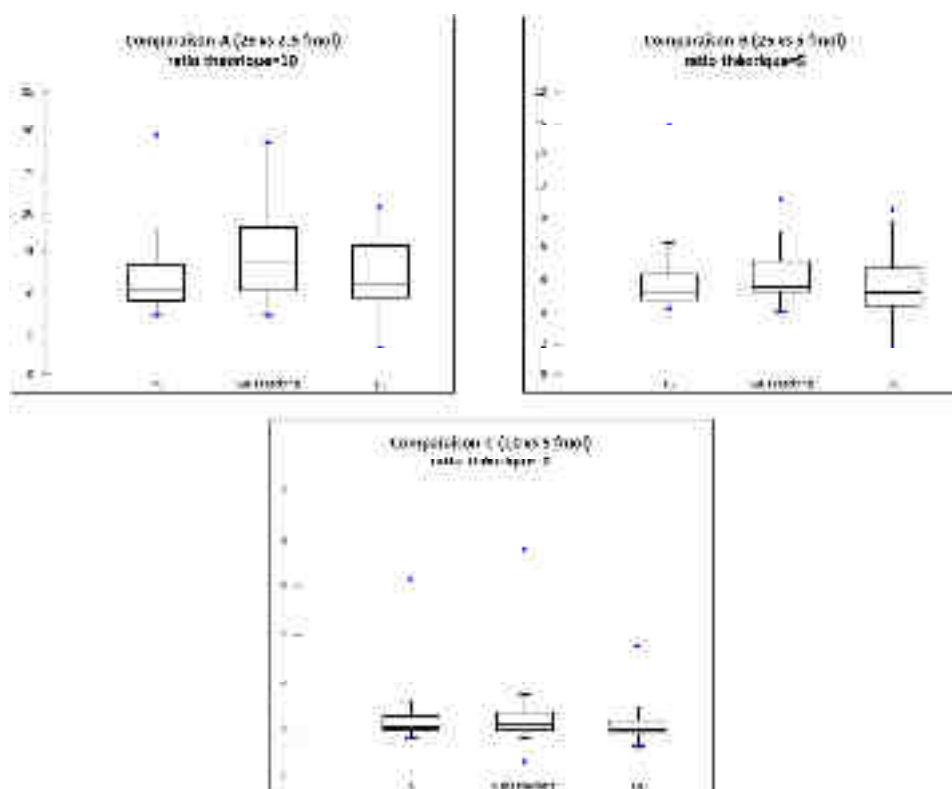


Figure 41
Distribution des ratios de concentration calculés pour les protéines UPS1 pour chaque comparaison.

Les valeurs médianes des ratios calculés sont reprises dans le tableau 18. Pour la comparaison A, la médiane des ratios calculés est la plus proche de la valeur théorique pour la préparation TG par rapport à la digestion liquide et au gel stacking (ratio médian = 10,4, 11,2 et 13,9 respectivement).

	Comparaison A (ratio théorique = 10)	Comparaison B (ratio théorique = 5)	Comparaison C (ratio théorique = 2)
Tube-gels	10,3	5,3	2,0
Gels stacking	13,9	5,6	2,1
Digestions liquides	11,2	5,2	2,0

Tableau 18
Médianes des ratios de concentrations calculés pour les protéines UPS1.

3.2.2. Evaluation de la reproductibilité

La reproductibilité de chaque protocole a été évaluée par le calcul du coefficient de variation (CV) des intensités normalisées des protéines.

Le CV de chaque protéine de levure a été calculé sur les valeurs d'intensités normalisées à partir des 18 analyses réalisées pour chaque protocole. Seules les protéines quantifiées dans au moins une analyse de chaque point de la gamme ont été considérées (soit 1780, 1612 et 1098 protéines pour les préparations tube-gels, gels stacking et digestions liquides respectivement). Les médianes des CV sont respectivement de 7 %, 9 % et 10 %.

Concernant les protéines UPS1, les CV ont été calculés pour chaque point de la gamme sur les valeurs d'intensités normalisées de chaque réplicat (Figure 42). Des CV médians inférieurs à 10 % ont été obtenus aux points 25, 10 et 5 fmol pour les trois protocoles. Les valeurs les plus faibles ont cependant toujours été obtenues avec les préparations tube-gels. Aux points 2,5 fmol et 1 fmol, les CV médians sont compris entre 10 % et 20 % pour les trois protocoles. Au point de gamme le plus bas (0,5 fmol), le CV médian obtenu avec la préparation TG était largement inférieur à celui obtenu avec les gels stacking ou les digestions liquides (22 %, 53 % et 48 % respectivement).

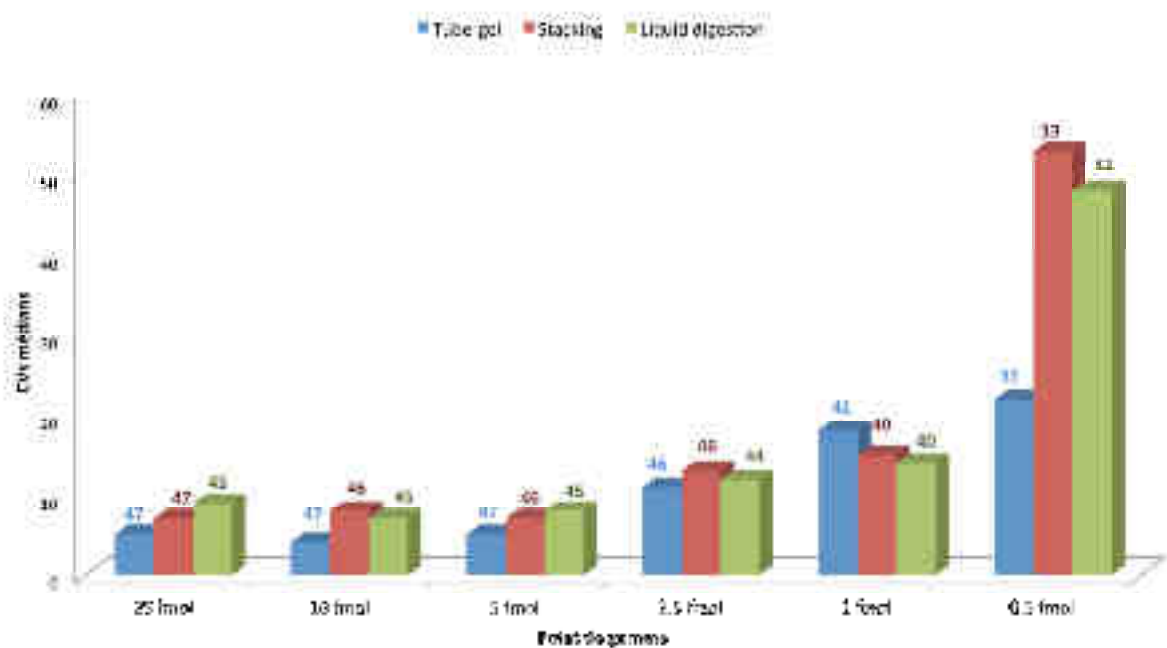


Figure 42
Médianes des CV calculés pour les protéines UPS1 entre les réplicats pour chaque point de gamme. Les chiffres situés au-dessus de chaque barre indiquent le nombre de protéines UPS1 pour lesquelles un CV a été calculé.

La reproductibilité a également été évaluée par le nombre de protéines UPS1 identifiées avec au moins 2 peptides uniques et des CV < 20 % entre les répliquats techniques. La préparation tube-gels a permis d'identifier avec ces critères stringents un nombre plus élevé de protéines aux points de gamme 25, 10 et 5 fmol. De meilleurs résultats sont obtenus avec la digestion liquide au point de gamme 2,5 fmol. Au point de gamme le plus bas (0,5 fmol), seuls les tube-gels ont permis d'identifier plus de dix protéines (Figure 43).

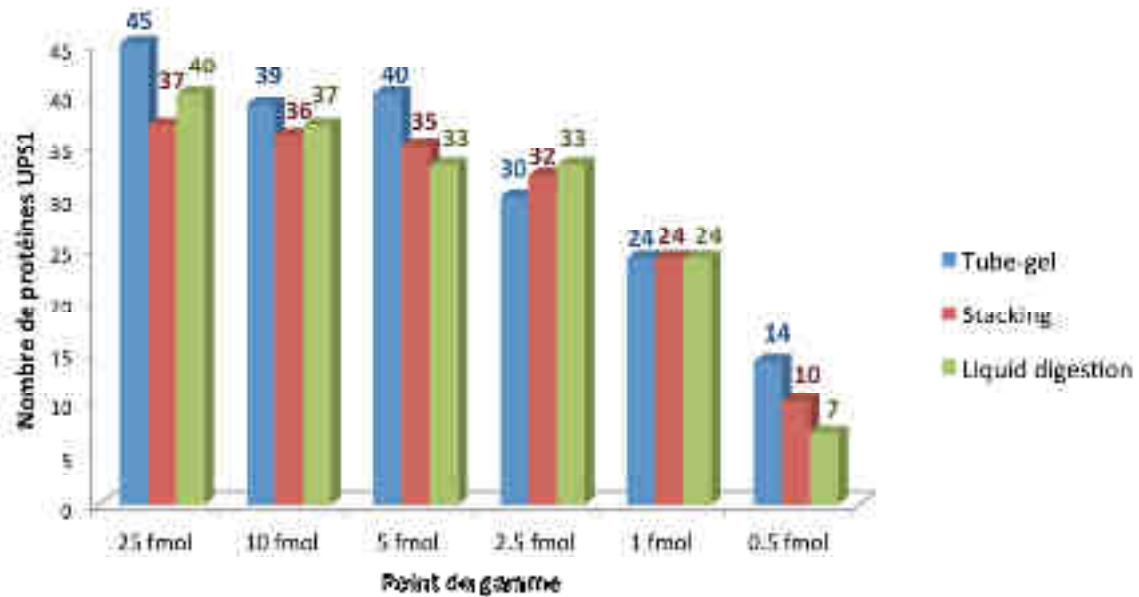


Figure 43
Nombre de protéines UPS1 identifiées avec au moins un peptide unique et un CV inférieur à 20 % entre les répliquats.

3.2.3. Détection des protéines variantes

Afin d'évaluer la capacité à pouvoir détecter des protéines d'abondances significativement différentes entre deux conditions, un test classique de Welch a été appliqué pour les trois comparaisons sélectionnées. Ce test devrait permettre d'évaluer la capacité à identifier l'ensemble des protéines UPS1 variantes connues pour les comparaisons choisies (estimation de la sensibilité) et à estimer les protéines variantes « fausses » de la matrice de levure (spécificité et taux de fausse découverte). Le test pour chacune des trois comparaisons sur les valeurs d'intensités normalisées. Seules les protéines quantifiées dans au moins deux répliquats sur trois ont été prises en compte. Les valeurs d'intensités normalisées ont été transformées en $\log(2)$ et les tests ont été réalisés à l'aide du logiciel Perseus.

La différence d'expression d'une protéine entre les deux conditions d'une comparaison est exprimée par la différence entre les moyennes des intensités de chaque condition transformées en $\log(2)$. Une protéine est considérée comme significative si $p < 0,05$ avec une différence ≥ 1 ou ≤ -1 . Une différence ≥ 1 correspond à un ratio de concentration ≥ 2 entre les deux conditions.

Ces résultats ont été représentés sous forme de « volcano plots » pour les trois comparaisons pour chaque protocole : gels stacking (figure 44), tube-gels (figure 45) et digestions liquides (figure 46). Chaque protéine est représentée en fonction de sa différence d'expression entre les deux conditions comparées et de la valeur p obtenue par le test de Welch. Les protéines UPS1 apparaissent bien distinctes des protéines de levure avec les trois protocoles. Un chevauchement plus important des ratios de concentration obtenus pour les trois comparaisons est observé pour les digestions liquides.

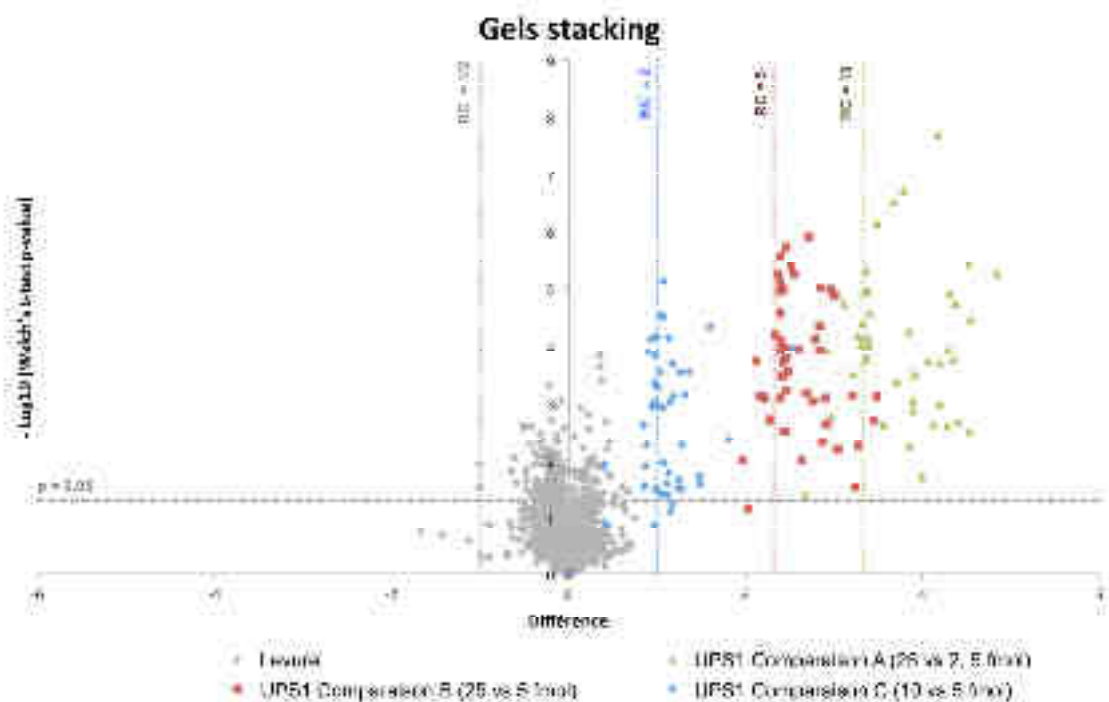


Figure 44
Représentation de type « volcano plot » obtenue avec les gels stacking.
RC : ratio de concentration théorique.

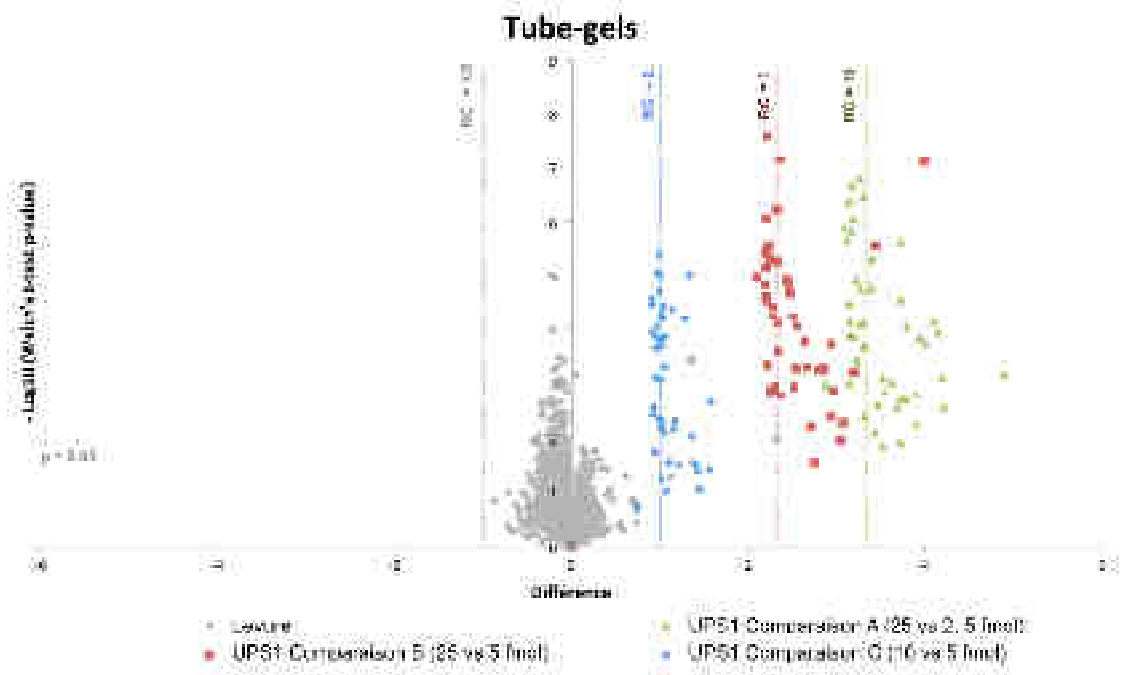


Figure 45
Représentation de type « volcano plot » obtenue avec les tube-gels.
RC : ratio de concentration théorique.

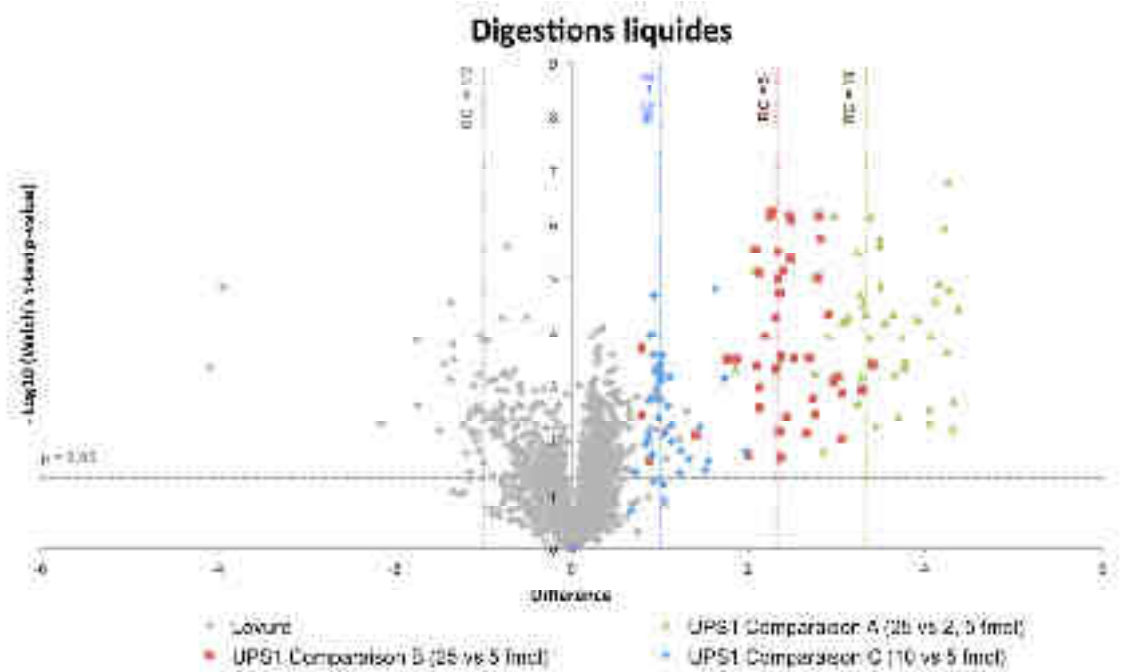


Figure 46
Représentation de type « volcano plot » obtenue avec les digestions liquides.
RC : ratio de concentration théorique.

Pour déterminer la sensibilité et la proportion de faux positifs pour chaque protocole, nous avons considéré les protéines comme pouvant être des évènements :

- vrais positifs (VP) : protéines UPS1 déterminées comme variantes avec les critères précédents ($p < 0,05$ avec une différence ≥ 1)
- faux négatifs : protéines UPS1 déterminées comme non variantes avec les critères précédents ($p < 0,05$ avec une différence ≥ 1)
- faux positifs (FP) : protéines de levure déterminées comme variantes avec les critères précédents ($p < 0,05$ avec une différence ≥ 1 ou ≤ 1)

La sensibilité et la proportion de faux positifs peuvent ainsi être définies de la façon suivante :

- sensibilité : $VP / (VP + FN)$. Dans notre cas, la somme $VP + FN$ est égale à 48 car elle correspond au nombre de protéines UPS1 attendues
- Taux de faux positifs : $FP / (VP + FP)$

Le tableau 19 indique les résultats obtenus pour chaque protocole.

	Sensibilité (%)			Proportion de faux positifs (%)		
	Comparaison	Comparaison	Comparaison	Comparaison	Comparaison	Comparaison
	A	B	C	A	B	C
Tube-gels	96	96	54	2	0	0
Stacking	92	94	56	2	2	0
Digestions liquides	90	83	40	30	27	0

[Tableau 19](#)

Sensibilité et proportion de faux positifs obtenues avec chaque comparaison pour les différents protocoles.

Pour les comparaisons A et B, le protocole TG a permis d'identifier comme variantes les protéines UPS1 avec une sensibilité supérieure aux gels stacking et aux digestions liquides. Pour ces deux comparaisons, la proportion de faux positifs obtenue avec les TG est inférieure ou égale à celle obtenue avec les gels stacking. La digestion liquide donne les moins bons résultats en terme de sensibilité et de proportion de faux positifs dans les deux comparaisons A et B.

Les sensibilités obtenues avec la comparaison C sont inférieures à celles obtenues avec les deux autres comparaisons. Ce point s'explique par l'utilisation d'un ratio seuil de concentration strictement supérieur à deux pour définir une protéine comme variante. Or ce ratio correspond à celui évalué par la comparaison C (10 versus 5 fmol). Les protéines ayant un ratio de concentration calculé strictement inférieur à deux dans la comparaison C n'ont donc pas été considérées comme variantes pour le calcul de la sensibilité.

Le protocole TG donne donc des résultats au moins équivalents en termes de sensibilité et de proportion de faux positifs par rapport aux gels stacking. En terme de sensibilité, les moins bons résultats ont été obtenus avec la digestion liquide.

4. CONCLUSIONS

Ce projet était basé sur la réalisation d'un mélange protéique parfaitement calibré comportant une gamme de concentration de protéines UPS1 dans un lysat protéique total de levure constituant une matrice complexe invariante. Cette stratégie a permis de comparer sur un plan qualitatif (identifications peptidiques et protéiques) et quantitatif (capacités à détecter des protéines variantes) une préparation d'échantillons de type tube-gel par rapport à une préparation 1D SDS-PAGE avec gel stacking ou par rapport à une digestion liquide.

Les résultats démontrent la possibilité d'utiliser ce protocole tube-gel pour l'identification de protéines au sein d'une matrice complexe et la réalisation d'analyses protéomiques différentielles. Ce mode de préparation a pour principaux avantages d'être compatible avec l'utilisation de SDS pour l'extraction et la solubilisation des protéines, et de réduire considérablement le temps de préparation par rapport aux gels d'acrylamide classiques et à la digestion liquide.

Ce mode de préparation tube-gel permet d'envisager plus facilement l'analyse d'un grand nombre d'échantillons compte tenu du gain de temps significatif obtenu lors de la préparation des échantillons.

5. COMMUNICATION DES RESULTATS

Un article soumis à la revue Journal of Proteome Research :

Tube-gel digestion: a fast and reproducible sample preparation protocol for label-free quantitative proteomics of complex biological samples.

L. Muller*, LM. Fornecker*, A. van Dorsselaer, S. Cianferani, C. Carapito.

* equally contributed to this work

Une communication par affiche (annexe 6) :

Evaluation de la préparation d'échantillons « tube-gel » pour l'analyse quantitative d'échantillons complexes. L. Muller, LM Fornecker, C. Carapito, S. Cianferani.

Congrès de Spectrométrie de Masse et d'Analyse Protéomique (SMAP) (Ajaccio, 15-18 septembre 2015)

• Chapitre 5

Recherche de biomarqueurs prédictifs de la réponse au traitement dans les lymphomes B diffus à grandes cellules

1. CONTEXTE SCIENTIFIQUE ET MEDICAL

Comme nous l'avons vu dans la première partie, les lymphomes B diffus à grandes cellules représentent l'entité la plus fréquente parmi les lymphomes non-hodgkiniens. L'hétérogénéité de cette maladie sur le plan biologique est illustrée par les profils d'expressions géniques qui permettent de distinguer deux principaux sous-types moléculaires : les formes ABC et GC.³⁵ Malgré l'amélioration de la survie des patients depuis l'introduction des anticorps monoclonaux anti-CD20, le sous-type ABC reste associé à un pronostic défavorable.¹²⁹ Ce sous-type ABC est prédominant chez les patients réfractaires ou en rechute.¹³⁰ La prise en charge de ces patients reste, à l'heure actuelle, un défi thérapeutique majeur. De même, une meilleure compréhension des mécanismes impliqués dans la chimiorésistance est nécessaire.

Une étude protéomique publiée en 2011 a comparé par 2D-DIGE (*two-dimensional differential in-gel electrophoresis*) des lignées cellulaires de LBDGC résistantes ou sensibles à une chimiothérapie de type CHOP. Une liste de dix protéines différentielles a pu être établie. Parmi cette liste, les protéines Akt, 14.3.3 et vimentine étaient surexprimées dans les cellules chimio-résistantes. Les auteurs ont pu montrer qu'une inhibition de la synthèse de la protéine 14.3.3 par siRNA ou qu'une inhibition d'Akt permettait de restaurer la sensibilité des cellules au traitement. L'inhibition de la synthèse de vimentine par siRNA était quant à elle associée à une diminution de la capacité des cellules à envahir une matrice de collagène.¹³¹

Une autre étude protéomique publiée en 2013 a utilisé une approche similaire.¹³² Des cultures cellulaires ont été réalisées à partir de tissus ganglionnaires prélevés au diagnostic chez 14 patients. Les cellules ont été exposées *in vitro* pendant 48 heures aux agents de chimiothérapie contenus dans le protocole CHOP (cyclophosphamide, doxorubicine, vincristine et prednisone). Deux groupes ont été constitués selon la viabilité cellulaire après exposition à la chimiothérapie : un groupe avec une faible sensibilité à la chimiothérapie (n = 7) et un groupe avec une forte sensibilité (n = 7). Après extraction des protéines, un fractionnement par gel 2D a été réalisé permettant de mettre en évidence 19 spots différentiels qui ont ensuite été analysés par MALDI TOF/TOF.

A partir des 19 protéines identifiées, une analyse par Western blot sur dix échantillons a confirmé une surexpression des protéines EZR et PLEK et une sous-expression des protéines GSTP1 et HSPB1 dans le groupe avec une sensibilité élevée par rapport au groupe avec une faible sensibilité au traitement. Ces résultats ont également été confirmés dans la même étude par des techniques immunohistochimiques sur 98 prélèvements.

Ces deux études publiées étaient basées sur des cultures cellulaires exposées *in vitro* aux agents de chimiothérapie avec une approche protéomique basée sur la réalisation de gels 2D. Notre projet se démarque par :

- l'utilisation du tissu tumoral (prélèvement ganglionnaire réalisé chez les patients) comme matériel biologique de départ, sans étape *in vitro* de culture cellulaire,
- une approche de type « shotgun » utilisant des spectromètres de masse de dernière génération. Ce type d'approche permet donc d'appliquer une stratégie protéomique quantitative globale sans marquage et d'augmenter significativement la couverture du protéome étudié,
- une cohorte de patients traités en première ligne de façon homogène par immuno-chimiothérapie (association du rituximab à la chimiothérapie pour tous les patients). Les données cliniques de ces patients sont connues ce qui a permis de déterminer deux groupes de patients : un groupe de patients « réfractaires » et un groupe de patients « sensibles ». Les patients du groupe « réfractaire » étaient soit réfractaires primaires au traitement de première ligne (absence de réponse), ou ont rechuté précocement après avoir obtenu une réponse complète initiale (durée de réponse inférieure à 12 mois). Tous les patients du groupe « sensible » ont obtenu une réponse complète au traitement de première ligne et n'ont pas rechuté.

2. PROCEDURE EXPERIMENTALE

2.1. Constitution des cohortes

La sélection des prélèvements a été faite à partir d'une liste de patients pour lesquels un prélèvement ganglionnaire congelé était disponible au Centre de Ressources Biologiques (CRB) des Hôpitaux Universitaires de Strasbourg. Les patients ont été sélectionnés sur la base des critères suivants :

- diagnostic d'un lymphome B diffus à grandes cellules selon la classification OMS 2008,
- prélèvement ganglionnaire congelé disponible,
- traitement de première ligne de type R-CHOP et données cliniques disponibles,
- patients informés de l'étude et n'ayant pas exprimé d'opposition à l'utilisation des prélèvements.

Deux groupes de six patients ont ainsi été constitués. Leurs caractéristiques sont résumées dans le tableau 20.

	Chimiosensibles (n = 6)	Rechute / Réfractaires (n = 6)
Âge médian (au diagnostic)	37 (18-72)	49 (34-79)
Sexe		
Homme	3	4
Femme	3	2
aalPI (au diagnostic)		
0-1	2	0
2-3	4	6
Stade d'Ann Arbor (au diagnostic)		
I-II	2	0
III-IV	4	6
LDH augmentées (au diagnostic)	4	6
Nombre de sites extra-ganglionnaires > 1	1	4
Nombre médian de lignes de traitement	1	4 (3-5)
Réponse au traitement de 1 ^{ière} ligne		
Rémission complète	6	2
Echec (réfractaires primaires)		4
Décès	0	4

[Tableau 20](#)
Caractéristiques des deux groupes de patients.

Les prélèvements ganglionnaires utilisés pour l'analyse protéomique étaient ceux ayant permis de poser le diagnostic de la maladie pour dix patients. Ces prélèvements ont donc été réalisés avant le début du traitement. Pour deux patients du groupe réfractaire, ces prélèvements initiaux n'étaient pas disponibles (absence de tissu congelé). Pour ces deux patients, les prélèvements ganglionnaires congelés utilisés pour l'analyse protéomique provenaient de nouvelles biopsies réalisées ultérieurement, après le début du traitement, pour confirmer l'absence de réponse ou la rechute précoce de la maladie.

2.2. Extraction des protéines

Tous les prélèvements ont été conservés à -80°C jusqu'à leur utilisation. Le matériel biologique étant constitué d'un fragment de tissu, nous avons réalisé une extraction par broyage mécanique à l'aide d'un potter dans un tampon de type Laemmli. Deux extractions successives ont été réalisées sur de la glace pour chaque prélèvement (première extraction avec 100 µL de tampon et deuxième extraction avec 85 µL de tampon). Après sonication pendant cinq minutes, un dosage DC™ (Bio Rad), compatible avec la présence de SDS, a été réalisé en duplicata après dilution de l'échantillon au 1/10^e.

La quantité de tissu utilisée était en moyenne de 9,4 mg (minimum : 1,5 mg ; maximum : 12,3 mg). La quantité de protéines obtenue après extraction était en moyenne de 166 µg par mg de ganglion (minimum : 120 µg ; maximum : 371 µg). Ce procédé a donc permis d'extraire et de solubiliser une quantité largement suffisante de protéines pour la réalisation des analyses protéomiques, y compris à partir de quantités de tissu très faible (au minimum 1,5 mg de tissu ganglionnaire).

Pour chaque échantillon, 20 µg de protéines ont été déposés sur gel 1D SDS-PAGE. Les protéines ont été concentrées dans une seule bande de stacking. La bande de stacking et la portion de gel située au-dessus ont été découpées. Les protéines ont été réduites et alkylées avant digestion trypsique. Après extraction, les peptides provenant de la portion de gel située au-dessus de la bande de stacking ont été ajoutés à ceux extraits de la bande de stacking.

2.3. Analyses nanoLC-MS/MS et traitement des données

Les analyses ont été réalisées sur un couplage nanoLC-QTOF (nanoAcquity (Waters), Impact II (Bruker)). Un gradient chromatographique de 120 minutes a été utilisé (5 à 35 % d'acétonitrile).

Par rapport à l'Impact HD utilisé dans le projet PROTELL, l'Impact II offre principalement un gain de résolution ($R = 45000$). Les améliorations techniques ont porté sur la cellule de collision, le réflectron et le détecteur.¹³³ Cet instrument a été mis en place dans le laboratoire en juillet 2015.

Le traitement des données a été réalisé à l'aide du logiciel MaxQuant. Les recherches ont été faites dans la banque humaine Swiss-Prot incluant les isoformes correspondant aux variants d'épissage annotés (42129 entrées à la date du 16 novembre 2015) avec le moteur de recherche Andromeda.⁶⁸ Les données ont été validées avec un FDR inférieur ou égal à 1 % au niveau peptidique et protéique en utilisant une stratégie target-decoy. La carbamidométhylation a été paramétrée en modification fixe. L'oxydation des méthionines et l'acétylation N-terminale des protéines ont été paramétrées en modifications variables. La quantification a été faite sur les peptides non modifiés sans utiliser de réplicat d'injection. Le logiciel Perseus a été utilisé pour l'analyse des résultats.

La figure 47 schématise la procédure expérimentale mise en place.



Figure 47
Récapitulatif global de la procédure expérimentale mise en place.

3. RESULTATS

Au total, 4251 protéines non redondantes ont pu être identifiées à partir des douze analyses. Le nombre de protéines identifiées par échantillon varie de 2445 à 2976.

Les comparaisons deux à deux des différents échantillons montrent un coefficient de corrélation de Pearson moyen à 0,92 pour les 132 comparaisons possibles. Ceci traduit une quantification reproductible et un protéome globalement similaire entre les échantillons. La figure 48 montre un exemple de corrélation obtenue entre deux échantillons ($r = 0,93$).

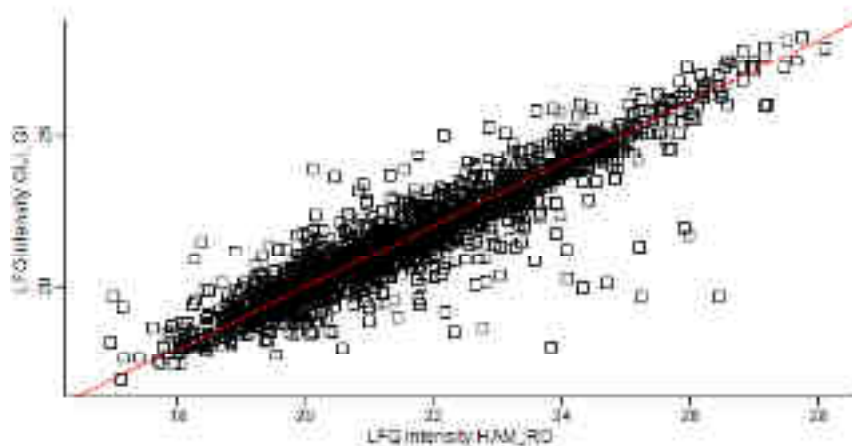


Figure 48
Comparaison des intensités normalisées des protéines entre deux échantillons (coefficient de Pearson $r=0.93$). Les valeurs d'intensités sont transformées en $\log(2)$.

La gamme dynamique couverte est proche de cinq ordres de grandeur avec 82 % des protéines ayant une intensité comprise entre deux ordres de grandeur (10^5 et 10^7) (Figure 49).

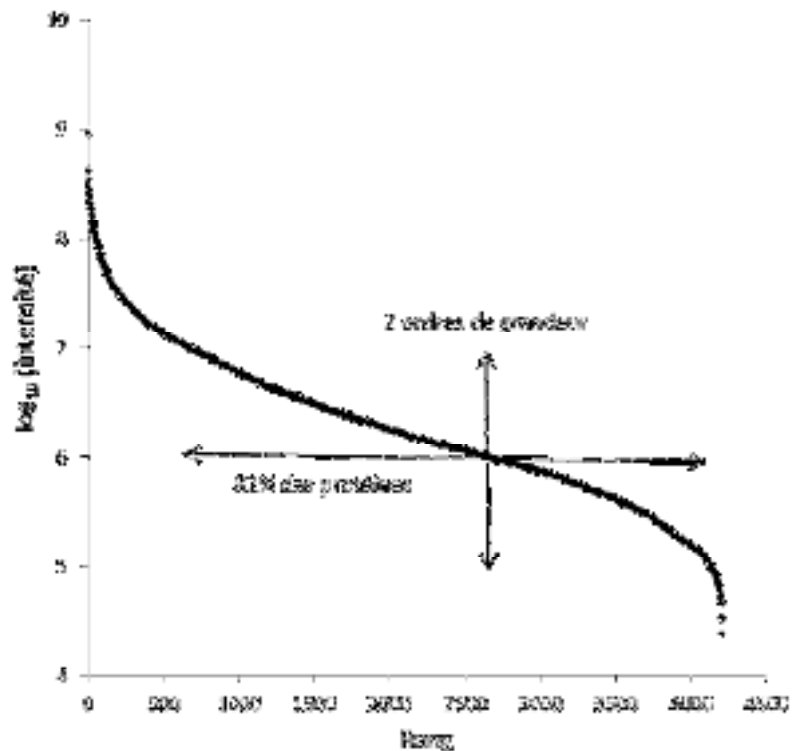


Figure 49

Représentation de la gamme dynamique pour l'ensemble des protéines quantifiées. Le rang des protéines a été attribué par ordre décroissant de leur intensité (rang 1 attribué à la protéine la plus intense).

La comparaison des deux groupes de patients a été faite en utilisant les valeurs normalisées d'intensité générées par le logiciel MaxQuant pour chaque protéine.

Parmi les 4251 protéines identifiées, 4221 l'ont été avec au moins un peptide unique. Parmi ces 4221 protéines, seules 47 n'ont pu être quantifiées dans aucun des douze échantillons. Sur les 4174 protéines restantes, 288 ont été quantifiées uniquement dans le groupe « sensible » (dont seules neuf dans au moins quatre échantillons sur les six de ce groupe) et 822 protéines uniquement dans le groupe « réfractaire » (dont seules 11 dans au moins quatre échantillons sur les six de ce groupe). Parmi les 3064 protéines quantifiées dans au moins un échantillon de chaque groupe, 2216 l'ont été dans au moins quatre échantillons sur les six de chaque groupe. Une analyse statistique avec un test de Welch a été faite sur ces 2216 protéines pour déterminer celles ayant une expression significativement différente entre les deux groupes. Le test de Welch a été réalisé à partir des valeurs normalisées d'intensité après transformation en (\log_2) . Ce test a été significatif ($p < 0,05$) pour 187 protéines : 106 surexprimées dans le groupe « réfractaire » et 81 surexprimées dans le groupe « sensible » (Figure 50).

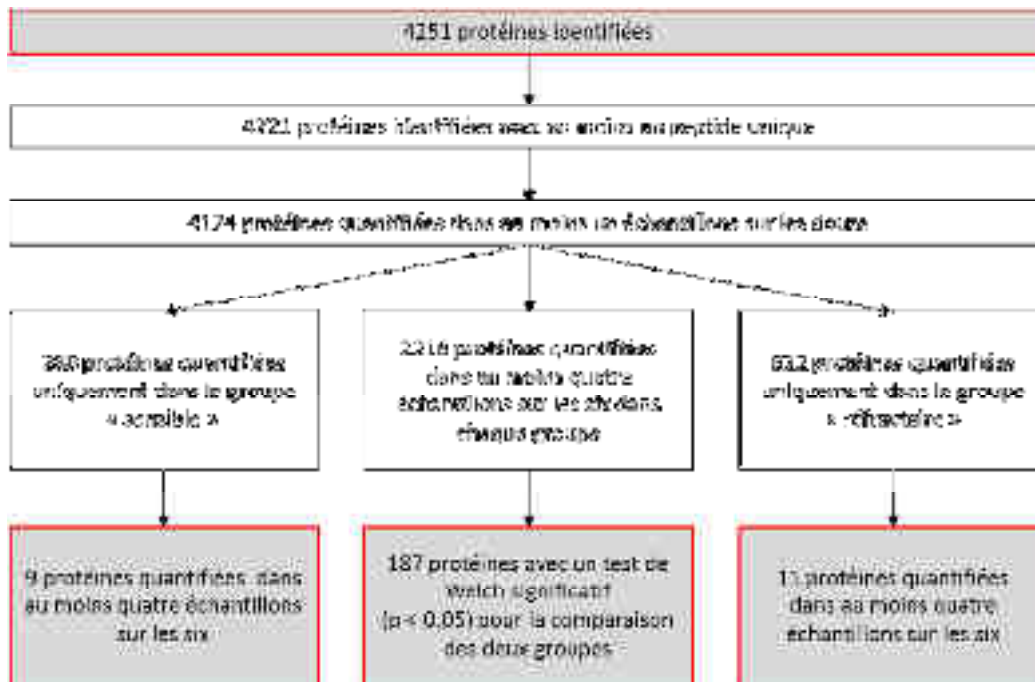


Figure 50
Filtres appliqués pour établir la liste de protéines d'intérêt.

3.1. Protéines uniquement quantifiées dans au moins quatre échantillons du groupe « sensible »

L'interprétation des résultats est faite en considérant ces neuf protéines comme étant présentes dans le groupe « sensible » et absentes dans le groupe « réfractaire » (tableau 21).

Protéine	Gène	Nombre total de peptides uniques	Intensité normalisée moyenne ($\times 10^5$)
UAP56-interacting factor	<i>FYTTD1</i>	3	7,48
Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial	<i>CPOX</i>	6	7,02
Paired box protein Pax-5	<i>PAX5</i>	3	6,00
B-lymphocyte antigen CD19	<i>CD19</i>	5	6,03
TBC1 domain family member 22A	<i>TBC1D22A</i>	3	5,21
CUGBP Elav-like family member 1	<i>CELF1</i>	2	5,44
Ubiquitin-conjugating enzyme E2 Z	<i>UBE2Z</i>	2	5,28
LIM and cysteine-rich domains protein 1	<i>LMCD1</i>	4	9,06
Protein NDRG3	<i>NDRG3</i>	3	5,39

Tableau 21
Liste des neuf protéines quantifiées dans au moins quatre échantillons sur six du groupe "sensible" et non quantifiées dans le groupe "réfractaire".

La protéine UAP56-interacting factor (UIF) est la seule parmi ces neuf à avoir été quantifiée dans l'ensemble des échantillons du groupe « sensible ». Cette protéine est codée par le gène *FYTTD1* et joue un rôle d'adaptateur dans l'export des ARN messagers du noyau vers le cytoplasme. Ce processus d'export est notamment médié par la protéine NXF1 et nécessite le recrutement au niveau des ARNm de l'adaptateur canonique REF/ALY. Ce recrutement est réalisé par l'intermédiaire de la protéine UAP56.¹³⁴ La protéine UIF se co-localise quant à elle sur les mêmes ARNm que l'adaptateur canonique REF/ALY.¹³⁵ D'autres protéines intervenant dans cette voie d'export des ARNm ont pu être identifiées et quantifiées dans les deux groupes de patients, mais sans différence significative d'expression :

- NXF1 a pu être quantifiée dans cinq échantillons (deux échantillons du groupe « réfractaire » et trois échantillons du groupe « sensible »),
- UAP56 et REF/ALY ont été quantifiées dans les douze échantillons.

La protéine PAX5 est un facteur de transcription qui joue un rôle majeur dans la lymphopoïèse. Il permet d'orienter les cellules vers la lignée lymphoïde B tout en réprimant l'expression des gènes impliqués dans la différenciation lymphoïde T.¹³⁶ Son rôle dans le développement des hémopathies lymphoïdes B est encore mal connu. Des anomalies du gène *PAX5* (délétions ou translocations) sont fréquemment décrites dans les leucémies aiguës lymphoblastiques.^{137,138} Des travaux récents réalisés sur des lignées cellulaires de lymphomes à cellules du manteau ont montré que l'inhibition de l'expression du gène *PAX5* était associée à un phénotype plus agressif avec augmentation de la prolifération cellulaire et des capacités d'infiltration des cellules tumorales. Cette inhibition du gène *PAX5* est également associée à une résistance au traitement plus importante ainsi qu'au développement de formes agressives de type blastoïdes.¹³⁹

Le CD19 appartient à la superfamille des immunoglobulines. Cette protéine est un marqueur des lymphocytes B. Elle participe à leur activation par la voie du BCR.¹⁴⁰ Son expression est notamment régulée par *PAX5*.¹⁴¹ Cette protéine est une cible thérapeutique¹⁴² avec, en particulier, le développement d'anticorps bi-spécifiques anti-CD19 / anti-CD3 (BiTE, *Bi-specific T-cell Engager*). Dans cette classe thérapeutique d'anticorps monoclonaux, le blinatumomab a récemment obtenu une autorisation de mise sur le marché pour le traitement des leucémies aiguës lymphoblastiques.¹⁴³ Un essai de phase II récemment publié a évalué cet anticorps dans le traitement des LBDGC réfractaires : un taux de réponse global de 43 % a pu être obtenu.¹⁴⁴ Dans cet essai il n'y a pas de donnée disponible concernant le niveau d'expression du CD19. Ce niveau d'expression pourrait éventuellement permettre de mieux identifier les patients éligibles à ce type de traitement.

3.2. Protéines uniquement quantifiées dans au moins quatre échantillons du groupe « réfractaire »

L'interprétation des résultats est faite en considérant les onze protéines comme étant présentes dans le groupe « réfractaire » et absentes dans le groupe « sensible » (tableau 22).

Protéine	Gène	Nombre total de peptides uniques	Intensité normalisée moyenne ($\times 10^5$)
stromal interaction molecule 1 (STIM1)	<i>STIM1</i>	4	10,8
DENN domain-containing protein 1C	<i>DENND1C</i>	2	3,70
Guanine nucleotide-binding protein subunit beta-4	<i>GNB4</i>	3	13,3
Very long-chain acyl-CoA synthetase	<i>SLC27A2</i>	5	7,53
Endonuclease domain-containing 1 protein	<i>ENDOD1</i>	3	7,65
Integrin alpha-5	<i>ITGA5</i>	5	5,94
Lysosomal acid phosphatase	<i>ACP2</i>	4	6,33
Tyrosine-protein kinase ZAP-70	<i>ZAP70</i>	5	6,81
Reticulocalbin-2	<i>RCN2</i>	3	11,1
Tripartite motif-containing protein 47	<i>TRIM47</i>	2	6,39
COMM domain-containing protein 3	<i>COMMD3</i>	3	8,54

[Tableau 22](#)

Liste des onze protéines quantifiées dans au moins quatre échantillons sur six du groupe "réfractaire" et non quantifiées dans le groupe "sensible".

La protéine stromal interaction molecule 1 (STIM1) est la seule à avoir été quantifiée dans les six échantillons. Elle est codée par le gène *STIM1*. Les protéines STIM 1 et 2 jouent un rôle important dans le contrôle de l'homéostasie intracellulaire du calcium et dans la régulation des signaux calciques. Elles sont localisées au niveau du réticulum endoplasmique et interagissent notamment avec les canaux calciques Orai1.^{145,146} Des travaux récents d'une équipe française ont montré que le rituximab pouvait entraîner une co-localisation au niveau de la membrane cytoplasmique du CD20, du CD95, des canaux calciques Orai1 et de la protéine STIM1. Le signal calcique ainsi généré pourrait inhiber l'apoptose induite par le CD20. L'inhibition de ces canaux calciques pourrait donc représenter une piste potentielle pour restaurer l'action pro-apoptotique du rituximab.¹⁴⁷

Les protéines de la famille COMMD (*Copper Metabolism gene MURR1 Domain-containing*) sont des protéines ubiquitaires intervenant dans la régulation de la voie NF-κB. L'activation de la voie NF-κB est décrite dans de nombreux types de lymphomes,¹⁴⁸ en particulier le sous-type ABC des LBDGC.¹⁴⁹ Parmi les protéines de la famille COMMD, COMMD1 est la mieux connue. Elle participe à l'inhibition de la voie NF-κB par l'ubiquitination de RELA et sa dégradation au niveau du protéasome.¹⁵⁰ Les mécanismes impliquant COMMD3 sont quant à eux moins bien connus. Son expression tissulaire la plus importante se trouve dans le thymus et COMMD3 peut s'associer à COMMD1.¹⁵¹ Contrairement à COMMD3, la protéine COMMD1 a été quantifiée dans les deux groupes de patients, mais uniquement dans deux échantillons du groupe « réfractaire » et trois du groupe « sensible ». Ces résultats suggèrent que la protéine COMMD3 pourrait donc un rôle spécifique dans la régulation de la voie NF-κB dans les LBDGC chimio-résistants.

La protéine ZAP70 (*Zeta-Chain Associated Protein kinase 70 kDa*) est une protéine cytoplasmique à activité tyrosine kinase qui intervient dans la voie de signalisation du récepteur des lymphocytes T.¹⁵² Son expression aberrante dans les hémopathies lymphoïdes B est connue, tout particulièrement dans la leucémie lymphoïde Chronique au cours de laquelle son expression est associée à un pronostic plus péjoratif.¹⁵³ Beaucoup moins de données sont disponibles concernant l'expression de ZAP70 dans les LBDGC, mais elle pourrait être plus importante dans le sous-type ABC.¹⁵⁴

3.3. Protéines quantifiées dans les deux groupes de patients

Parmi les 2216 protéines retenues pour l'analyse statistique, 187 ont un test significatif ($p < 0,05$). Une représentation de type « heatmap » basée sur ces 187 protéines permet de ségréguer sans ambiguïté les patients de chaque groupe. Cette liste de protéines permet donc d'établir une signature protéique distinguant les deux groupes de patients (Figure 51).

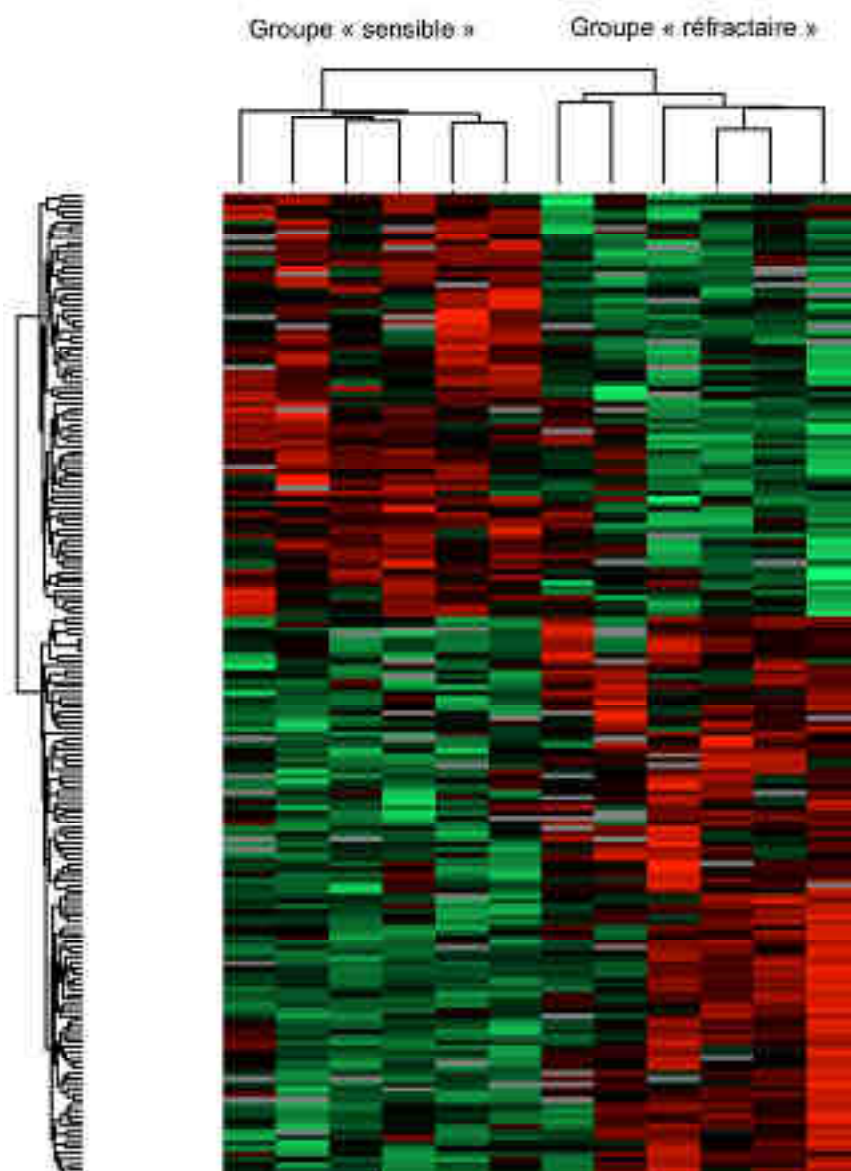


Figure 51

Représentation de type "heatmap" obtenue à partir des 187 protéines ayant un test de Welch significatif ($p < 0,05$) pour la comparaison des deux groupes.

3.3.1. Protéines surexprimées dans le groupe « réfractaire »

La liste complète de ces protéines figure en annexe 7. Certaines d'entre elles, ayant les différences d'expressions les plus importantes entre les deux groupes et un rationnel biologique potentiel, sont détaillées dans les paragraphes suivants.

3.3.1.1. Fibrinogène

Le fibrinogène est constitué par l'assemblage de trois protéines (FGA, FGB et FGG). Ces trois protéines sont surexprimées dans le groupe de patients « réfractaire » et ont été identifiées avec un nombre total de 30, 32 et 23 peptides uniques respectivement. Chacune de ces trois protéines a été quantifiée dans les douze échantillons. La figure 52 indique la distribution des intensités normalisées pour chacune d'entre elles, en fonction du groupe de patients.

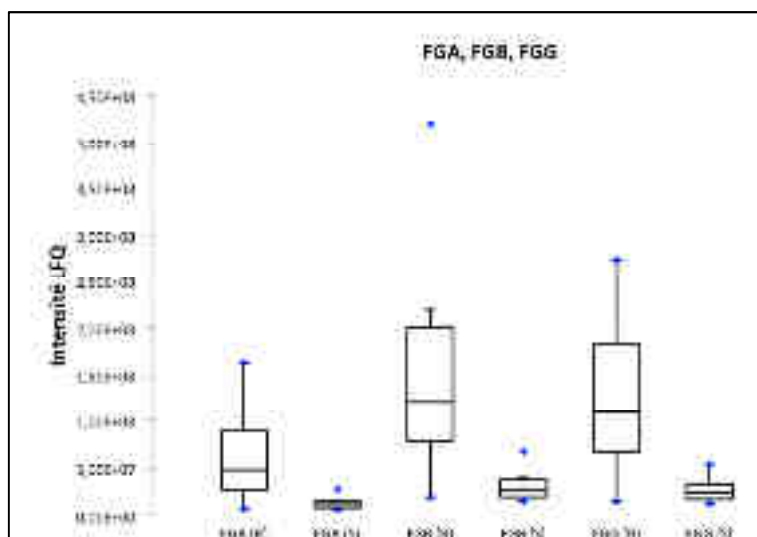


Figure 52
Distribution des intensités normalisées pour les trois protéines FGA, FGB et FGG.
R : réfractaire, S : sensible

Ces résultats sont en faveur d'une abondance de fibrinogène plus importante dans le tissu ganglionnaire des formes réfractaires de LBDGC. Peu de données existent sur un rôle éventuel du fibrinogène dans la lymphomagenèse. Néanmoins, des travaux *in vitro* ont montré que le fibrinogène pourrait stimuler la prolifération des cellules centroblastiques dans les follicules lymphoïdes par l'intermédiaire des cellules dendritiques folliculaires situées dans la zone sombre des centres germinatifs.¹⁵⁵

Le fibrinogène est également une protéine abondante dans le plasma. Pour évaluer la part d'une éventuelle contamination plasmatique, nous avons comparé les taux plasmatiques de fibrinogène entre les deux groupes de patients. Les taux rapportés correspondent aux dosages réalisés au moment du prélèvement ganglionnaire. Le taux plasmatique médian de fibrinogène est supérieur dans le groupe « sensible » par rapport au groupe « réfractaire » (4,7 g/L et 3,8 g/L respectivement) (figure 53). Ce résultat ne permet cependant pas d'exclure formellement une contamination des fragments de tissu ganglionnaire dans les six échantillons du groupe « réfractaire » plus importante par le plasma.

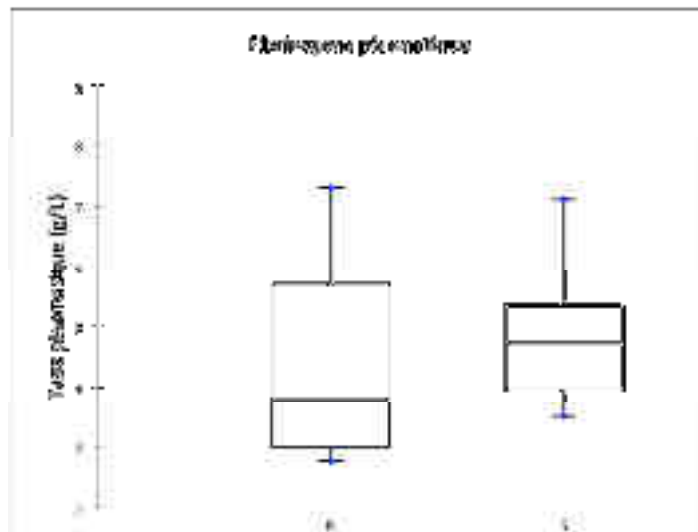


Figure 53
Distribution des taux plasmatiques de fibrinogène (R : réfractaire, S : sensible).

3.3.1.2. P-Rex1 (Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein)

Cette protéine a été identifiée avec 15 peptides uniques au total. Elle participe à l'activation des voies PI3K/AKT et MEK/ERK par l'intermédiaire de la protéine Rac. Son rôle potentiel dans la lymphomagenèse est inconnu, mais, dans le cancer du sein, elle pourrait participer à promouvoir la prolifération des cellules tumorales.¹⁵⁶ Dans le cancer de la prostate, P-Rex1 est associée à une résistance aux traitements anti-VEGF ou anti-VEGFR.¹⁵⁷ La distribution des intensités normalisées est indiquée dans la figure 54 pour chacun des deux groupes de patients.

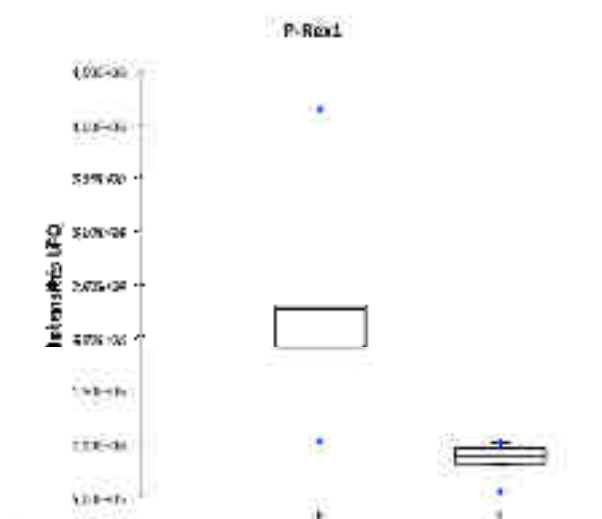


Figure 54
Distribution des intensités normalisées pour la protéine P-Rex1 (R : réfractaire, S : sensible).

3.3.1.3. PLOD3 (Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3)

La protéine PLOD3 a été identifiée avec 11 peptides uniques. Elle participe à l'organisation du réseau des fibres de collagène. Des travaux publiés en 2006 sur le cancer du sein ont montré que le gène *PLOD3* faisait partie d'une signature transcriptionnelle associée à une résistance aux anthracyclines et au cyclophosphamide.¹⁵⁸ Ces deux produits de chimiothérapies sont largement utilisés pour le traitement des lymphomes et font partie du protocole de chimiothérapie R-CHOP utilisé pour le traitement de première ligne des LBDGC. La figure 55 indique la distribution des valeurs d'intensités normalisées dans les deux groupes de patients.

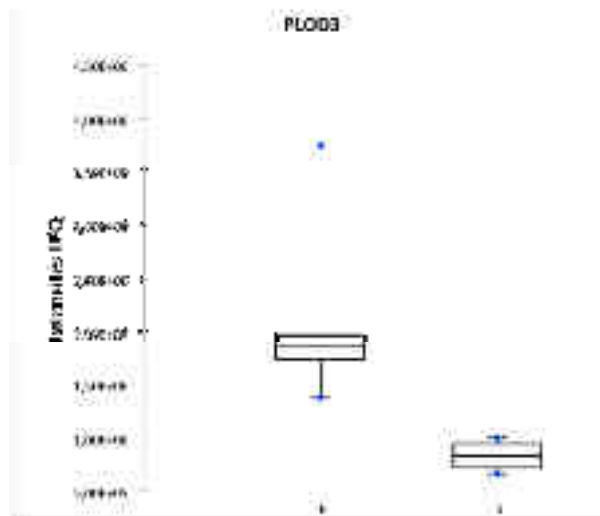


Figure 55
Distribution des intensités normalisées pour la protéine PLOD3.
R : réfractaire, S : sensible

3.3.1.4. Tensin-3

Cette protéine a été identifiée avec 13 peptides uniques. Des travaux réalisés sur des lignées de cancer du poumon, du sein ou de mélanome ont montré qu'elle pouvait être phosphorylée par l'intermédiaire de la voie Src. Elle contribue au développement tumoral et au pouvoir métastatique des cellules tumorales.¹⁵⁹ Il n'y a par contre pas de données disponibles dans la littérature concernant son rôle potentiel dans la lymphomagénèse.

La figure 56 indique la distribution des valeurs d'intensités normalisées dans les deux groupes de patients.

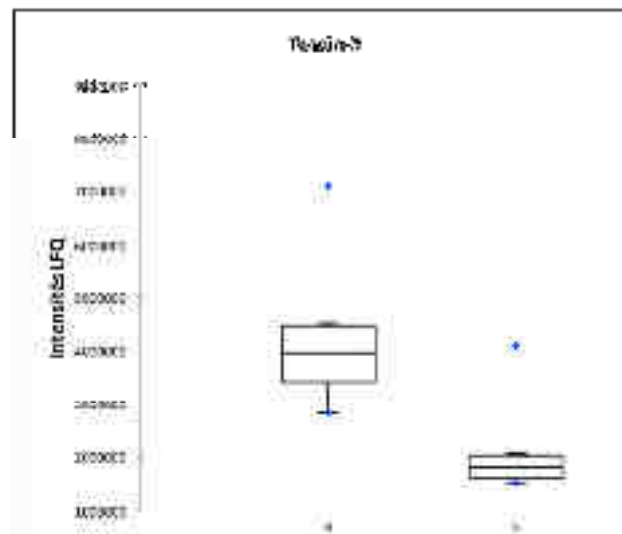


Figure 56
Distribution des intensités normalisées pour la protéine Tensin-3.
R : réfractaire, S : sensible

3.3.2. Protéines surexprimées dans le groupe « sensible »

La liste complète est donnée en annexe 7. Les exemples de certaines protéines ayant les différences d'expression les plus importantes sont détaillés ci-dessous.

3.3.2.1. CD20

Le CD20 a été identifié avec 11 peptides uniques. Cette protéine membranaire a une importance majeure dans le contexte des hémopathies lymphoïdes B. Elle constitue la cible des anticorps monoclonaux anti-CD20 dont le chef de file est le rituximab. L'association de ces anticorps monoclonaux aux polychimiothérapies conventionnelles a représenté un progrès thérapeutique majeur avec un bénéfice en survie globale obtenu notamment chez les patients traités pour un LBDGC. Les douze patients de cette série ont tous bénéficié d'un traitement par rituximab en première ligne.

La protéine CD20 a pu être quantifiée dans les six échantillons du groupe « sensible » et quatre du groupe « réfractaire ». Parmi les deux échantillons du groupe « réfractaire » pour lesquels la protéine n'a pas pu être quantifiée, un cas l'exprimait également faiblement en immunohistochimie (10 % des cellules) et l'autre cas correspondait à un sous-type de LBDGC « riche en cellules T » avec des cellules tumorales en proportion minoritaire.

Afin d'évaluer un éventuel biais lié à une mauvaise solubilisation des protéines membranaires, l'expression du CD45 (identifié sans différence significative entre les deux groupes avec 38 peptides uniques) a été comparée à celle du CD20. Le CD45 est une protéine membranaire à activité tyrosine phosphatase exprimée par toutes les cellules leucocytaires. La figure 57 indique les distributions des valeurs d'intensités normalisées obtenues dans chaque groupe de patients pour ces deux protéines.

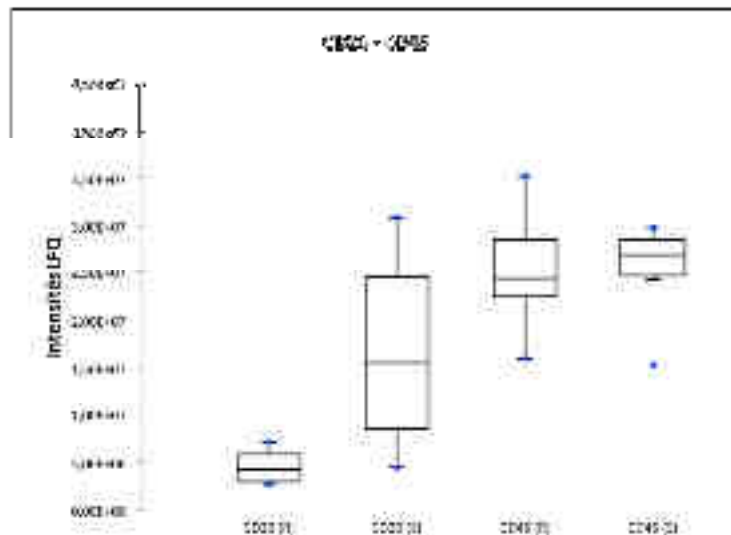


Figure 57
Distribution des intensités normalisées pour les protéines CD20 et CD45.
R : réfractaire, S : sensible

Dans la littérature, deux travaux ont évalué l'expression du CD20 dans les LBDGC par cytométrie en flux. De façon similaire aux résultats obtenus dans notre étude, une expression faible ou absente de CD20 était corrélée à un pronostic plus péjoratif.^{160,161}

3.3.2.2. SYK (Spleen tyrosine kinase)

La protéine SYK a été identifiée avec un total de 24 peptides uniques. Cette protéine, ayant une activité tyrosine kinase, a un rôle majeur dans les premières étapes de la voie de signalisation du récepteur des lymphocytes B.¹⁶² D'autres protéines sont également impliquées dans les premières étapes de cette voie de signalisation, telles que la tyrosine kinase de Bruton (BTK), LYN ou la PI3 kinase. Hormis SYK, il n'a pas été observé de différences significatives entre les deux groupes pour l'expression de BTK, LYN ou la sous-unité catalytique de PI3K delta.

Cette protéine constitue de plus une cible thérapeutique pour des inhibiteurs de tyrosine kinases.¹⁶³ Des inhibiteurs de SYK sont en cours de développement et d'évaluation dans les hémopathies lymphoïdes B (fostamatinib, entospletinib).^{164,165} La figure 58 indique les distributions des intensités normalisées pour ces différentes protéines impliquées dans la voie de signalisation du BCR.

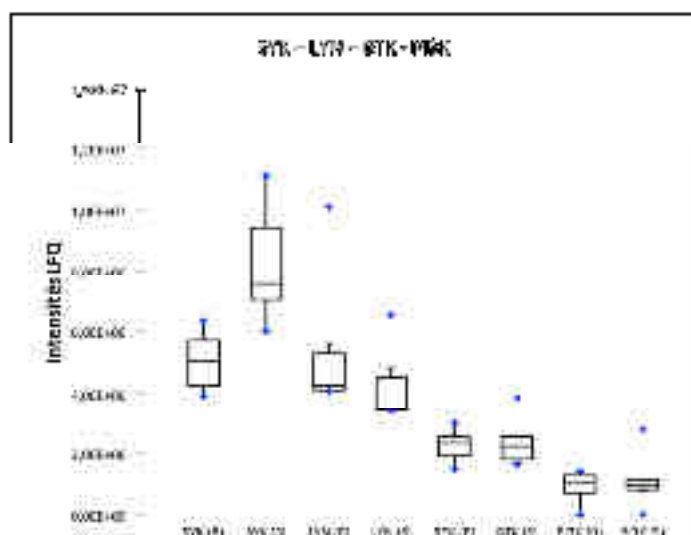


Figure 58

Distribution des intensités normalisées pour les protéines SYK, LYN, BTK et PI3K. La sous-unité catalytique de l'isoforme delta de la PI3K est représentée.

R : réfractaire, S : sensible

3.3.2.3. LIMA1 (LIM domain and actin-binding protein 1)

Cette protéine a été identifiée avec 18 peptides uniques. Elle est également connue sous le nom EPLIN (*Epithelial Protein Lost In Neoplasm*) et possède une fonction de suppresseur de tumeurs.¹⁶⁶ Son expression est dérégulée dans différents types de cancers.¹⁶⁷⁻¹⁶⁹

Son rôle potentiel dans la lymphomagénèse des LBDGC est inconnu. Cependant, dans les lymphomes du MALT, un clivage de LIMA1 a lieu en présence de la translocation t(11;18)(q21;q21) qui implique les gènes *API2* et *MALT1*.¹⁷⁰ Cette translocation induisant un clivage de LIMA1 est récurrente dans les lymphomes du MALT de l'estomac. Ces lymphomes se développent majoritairement en présence de la bactérie *Helicobacter pylori* au niveau de la muqueuse gastrique. Une guérison du lymphome est possible par un traitement antibiotique éradicateur d'*H. pylori*. Cependant, la présence de cette translocation t(11;14) est associée à une résistance du lymphome à ce traitement éradicateur.

La distribution des intensités normalisées obtenues pour les deux groupes de patients est indiquée dans la figure 59.

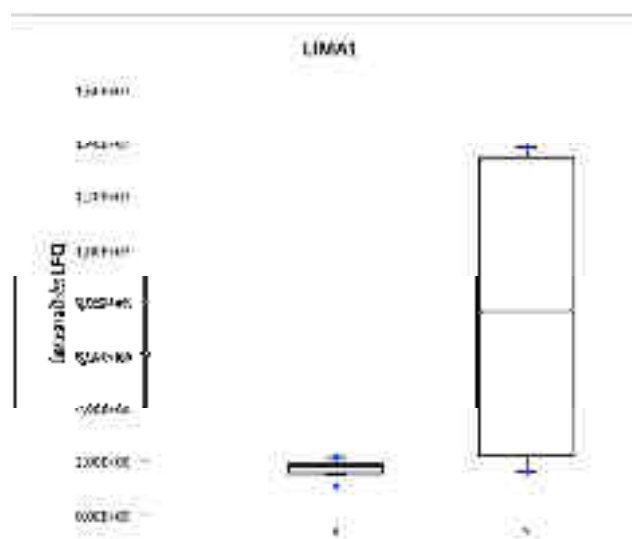


Figure 59
Distribution des intensités normalisées pour la protéine LIMA1.
R : réfractaire, S : sensible

4. CONCLUSIONS ET PERSPECTIVES

Ce projet a permis d'identifier et quantifier plus de 4000 protéines par une approche quantitative globale sans marquage à partir de tissus ganglionnaires. La constitution de deux cohortes distinctes de patients sur la base de leur chimiosensibilité a permis de mettre en évidence une liste de protéines exprimées différemment entre ces deux groupes de patients. Certaines de ces protéines pourraient permettre de mieux comprendre les processus impliqués dans la chimiorésistance ou de mieux définir des sous-groupes de patients pouvant bénéficier de traitements ciblés.

Les perspectives de ce travail sont:

- Etablir une nouvelle cohorte de patients chimiosensibles et chimioréfractaires pour confirmer ces résultats sur une deuxième série indépendante de patients.
- Valider les biomarqueurs potentiels par des approches différentes : immunohistochimie ou protéomique quantitative ciblée
- Corréler ces résultats de protéomique quantitative avec une analyse quantitative du transcriptome par RNAseq sur des échantillons identiques.

Ces perspectives ont fait l'objet d'une réponse, en cours d'évaluation, à un appel à projet du Cancéropôle Grand-Est, en collaboration avec le Laboratoire d'Immuno-Rhumatologie Moléculaire du Pr S. BAHRAM (INSERM UMRS_1109, Université de Strasbourg) pour l'analyse du transcriptome par RNAseq. Il a pour objectif d'évaluer la faisabilité d'une approche multidisciplinaire et complémentaire par analyses protéomique et transcriptomique reposant sur l'utilisation d'un matériel biologique identique pour ces deux approches. Cette utilisation d'un matériel biologique commun permettra de réaliser une analyse intégrée et multidimensionnelle des données de protéomique et de RNAseq.^{171,172}

Une autre perspective pour poursuivre l'étude protéomique des LBDGC est de mettre en place une méthode d'extraction des protéines à partir de lames histologiques. En effet, si le matériel tumoral utilisé dans ce projet provenait de tissu ganglionnaire congelé, cette étape de congélation n'est pas réalisée pour tous les patients (souvent en raison d'un manque de matériel après réalisation des analyses nécessaires pour poser le diagnostic). Des travaux publiés dans la littérature indiquent qu'il est cependant possible d'extraire des protéines et de réaliser des analyses protéomiques par spectrométrie de masse à partir de lames tissulaires incluses en paraffine comportant des protéines fixées.¹⁷³⁻¹⁷⁵ L'utilisation de ce type de méthode d'extraction pourrait permettre d'accroître considérablement le nombre d'échantillons disponibles.

Enfin, l'approche globale développée dans ce projet pourra être étendue à l'analyse d'autres types de lymphomes non-hodgkiniens ou hodgkiniens et ouvre ainsi encore plus de perspectives pour l'analyse protéomique des lymphomes.

Partie III

Discussion et synthèse

1. PROBLEMATIQUE GENERALE DES BIOMARQUEURS

Les applications concrètes des biomarqueurs dans le domaine médical sont multiples. Il est possible schématiquement de distinguer 1) les biomarqueurs utilisés pour le dépistage ou le diagnostic des maladies, 2) ceux servant à établir un pronostic, ou 3) ceux utilisés à visée prédictive de la réponse à un traitement. En parallèle de ces applications cliniques, les biomarqueurs sont également très largement utilisés dans le cadre de la recherche académique ou de l'industrie pharmaceutique.

Ces différents aspects relatifs à l'utilisation des biomarqueurs s'appliquent tout particulièrement au domaine de la cancérologie. Il existe en effet dans ce domaine un besoin majeur d'améliorer le diagnostic et le traitement des pathologies cancéreuses. Bien que de nombreux biomarqueurs encore utilisés aujourd'hui aient été découverts il y a plusieurs années voire décennies, les progrès technologiques récents ont permis de mettre en évidence un nombre important de nouveaux biomarqueurs potentiels. La figure 60 illustre ce phénomène en indiquant le nombre de publications indexées dans PubMed avec les deux mots clés « biomarkers » et « neoplasms » parues chaque année entre 2000 et 2014 (recherche effectuée avec les termes MeSH, *Medical Subject Headings*). Le nombre d'articles publiés chaque année a presque triplé pour atteindre au final plus de 167000 articles sur cette période.

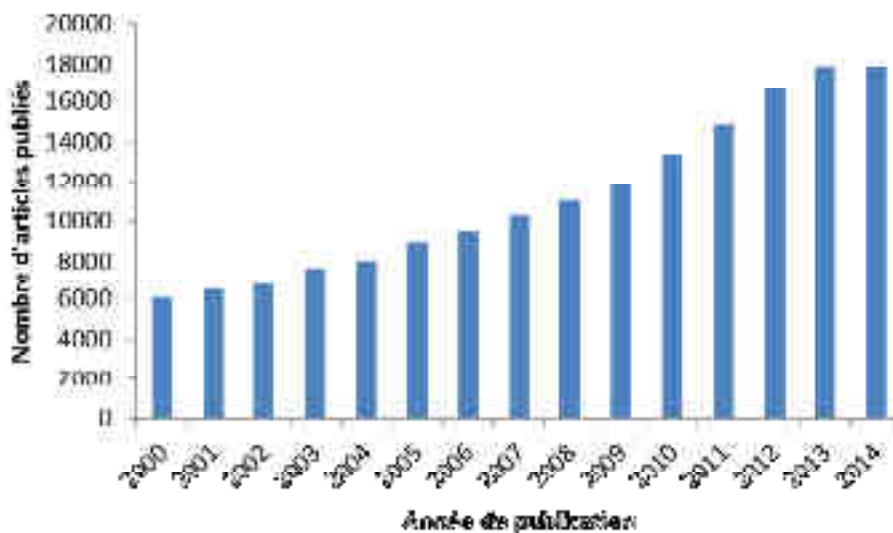


Figure 60
Nombre de publications indexées dans PubMed avec les deux mots clés « biomarkers » et « neoplasms » parues chaque année entre 2000 et 2014.

Certains exemples illustrent des succès liés à la découverte de biomarqueurs en lien direct avec la stratégie thérapeutique de certains cancers : le cetuximab est un anticorps monoclonal chimérique dirigé contre le récepteur du facteur de croissance épidermique (EGFR, *Epidermal Growth factor receptor*) indiqué dans le traitement du cancer colorectal métastatique. Cependant seuls les patients ne présentant pas de mutation du gène *KRAS* sont susceptibles de répondre à ce traitement. La recherche de cette mutation est donc nécessaire avant l'instauration du traitement. De façon similaire, le gefitinib est un inhibiteur de la tyrosine-kinase de l'EGFR, indiqué dans le traitement du cancer bronchique non à petites cellules. Il n'apporte cependant de bénéfice en survie sans progression que chez les patients ayant une mutation activatrice de l'EGFR.

Néanmoins, alors qu'un nombre croissant de biomarqueurs est proposé dans la littérature, très peu finissent par réellement trouver une place en pratique clinique. Dans le domaine des hémopathies lymphoïdes, l'exemple de la leucémie lymphoïde chronique illustre ce paradoxe. La classification de Binet a été publiée en 1981 à partir de deux séries comportant 99 et 196 patients.¹⁰ Depuis cette publication, quelques dates rendent compte des progrès exponentiels réalisés en biologie moléculaire :

- 1977 : Frederick Sanger décrit une nouvelle méthode de séquençage de l'ADN¹⁷⁶
- 1983-1985 : mise au point de la technique de PCR (*Polymerase Chain Reaction*)^{177,178}
- 2000-2001 : Le séquençage du génome humain dans le cadre du projet HGP (*Human Genome Project*) a été achevé en 2000 et les premiers résultats publiés en 2001^{179,180}
- 2007 : Le premier séquençage du génome d'un seul individu est publié (génome de James Watson qui a découvert dans les années 1950 la structure double-brin en hélice de l'ADN)^{181,182}
- Depuis la fin des années 2000, les techniques de séquençage de nouvelle génération (NGS) ont encore augmenté la vitesse d'acquisition des données et ouvert la voie à de nouvelles perspectives sur un plan biologique.¹⁸³ La base de données GOLD (*Genome Online Database*)¹⁸⁴ répertorie les différents projets de séquençage de génomes réalisés. La figure 61 montre l'augmentation rapide du nombre de projets ces dernières années. Plus de 7000 génomes sont annotés de façon complète à ce jour.

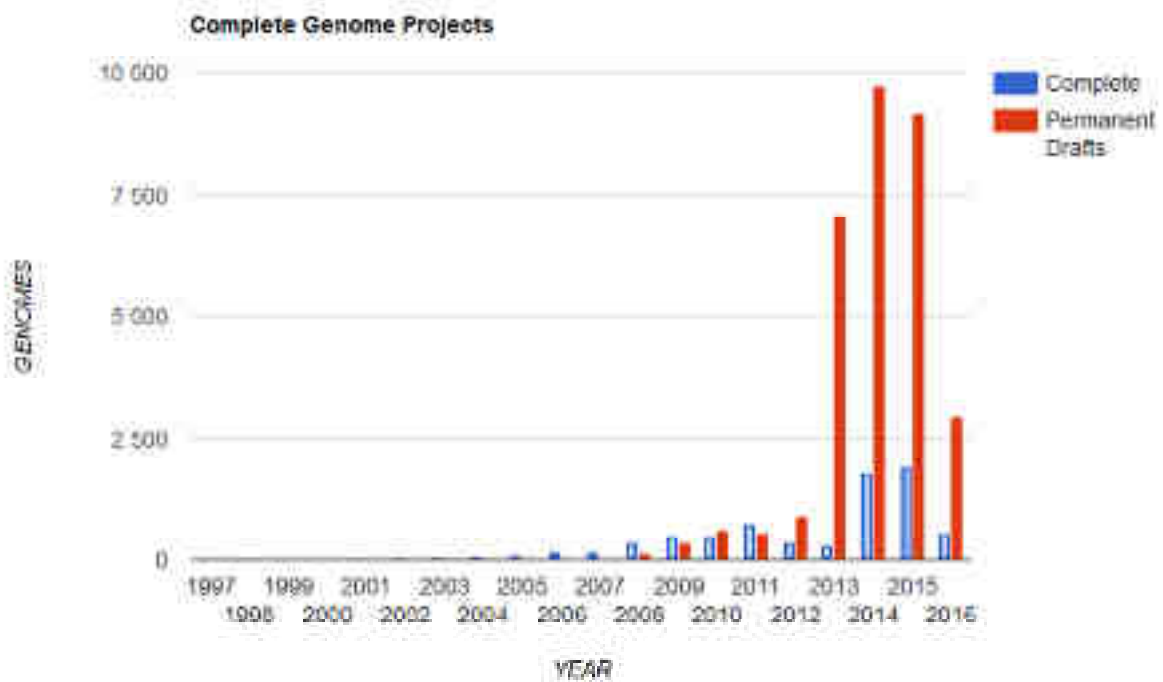


Figure 61
Nombre de projets de séquençage de génomes annotés de façon complète ou partielle (permanent drafts) par année recensés sur la base de données GOLD (d'après référence n° 184).

Appliqués au domaine de la LLC, tous ces progrès ont permis de mieux caractériser la maladie aux niveaux génétique et moléculaire à partir des années 1990. Des sous-groupes de patients à haut risque ont été clairement identifiés. Pour autant, malgré toutes ces avancées, les recommandations d'instauration d'un traitement dans la LLC en 2016 restent basées sur la classification de Binet :

- patients de stade A avec des critères de maladie active,
- patients de stade B avec un syndrome tumoral,
- patients de stade C.

Le seul facteur pronostique dont la recherche est « obligatoire » avant d'instaurer un traitement est la recherche de la délétion 17p par FISH et la recherche de mutations du gène *TP53* en biologie moléculaire. Ces patients bénéficient en effet de stratégies thérapeutiques différentes.

Il existe d'autres exemples de biomarqueurs encore largement utilisés aujourd'hui et qui ont été découverts bien avant l'ère des technologies « omiques » :

- 1980 : dosage du CA125 dans le cancer de l'ovaire¹⁸⁵ et du PSA (Prostatic Specific Antigen) dans le cancer de la prostate,^{186,187}
- dans les années 1960, découverte de l'alpha foeto-protéine (AFP)¹⁸⁸ et de l'antigène carcino-embryonnaire (ACE),¹⁸⁹
- il est même possible de remonter jusqu'à 1848 avec la découverte du dosage de la glycémie ou la découverte par Henry Bence Jones d'une protéine correspondant à une chaîne légère d'immunoglobuline. Le dosage de cette protéine de Bence Jones dans les urines est toujours largement utilisé actuellement pour le suivi des patients atteints de myélome.

Les raisons sont multiples pour expliquer cet écart important entre la pratique clinique quotidienne et les travaux considérables réalisés sur un plan fondamental pour identifier de nouveaux biomarqueurs. Ces raisons se trouvent dans chacune des deux grandes phases du développement d'un biomarqueur : phase de découverte et phase de validation.

Plusieurs sources d'erreurs sont possibles au moment de la phase de découverte, empêchant la validation ultérieure du biomarqueur :

- Etape pré-analytique : la question biologique de départ doit être clairement établie. De même, la sélection des échantillons et du matériel biologique doivent être rigoureux et permettre de répondre à la question posée. Le nombre d'échantillons utilisés doit être suffisant pour mettre en évidence une différence entre deux conditions. Une attention particulière est nécessaire sur les conditions de prélèvement et de stockage des échantillons. Toute modification d'un biomarqueur apparaissant *in vitro* altère sa capacité à refléter son état *in vivo*.
- Etape analytique : la technique employée et ses limites doivent être parfaitement maîtrisées. Il est fondamental à cette étape de pouvoir assurer la qualité des résultats générés par la mise en place de contrôles qualités. Une analyse protéomique, par exemple, ne se limite pas à fournir une liste de protéines.
- Etape post-analytique : l'analyse et l'interprétation des résultats fournis est une étape cruciale. Son importance est croissante avec les progrès technologiques actuels qui génèrent des quantités importantes de données (*big data*). La maîtrise des logiciels bio-informatiques et des tests statistiques permettant le traitement des données, en particulier avec les approches « omiques », est nécessaire. La pertinence biologique des résultats doit être vérifiée. La seule positivité d'un test statistique est insuffisante pour proposer un biomarqueur potentiel.

Deux exemples peuvent illustrer ce qui précède: l'acide lysophosphatidique a été proposé en 1998 comme biomarqueur plasmatique des cancers de l'ovaire.¹⁹⁰ Ce biomarqueur n'a pas été validé en raison notamment d'une sécrétion possible dans le plasma par différents types cellulaires, dont les plaquettes activées par le prélèvement sanguin ou la centrifugation du plasma.¹⁹¹ Le deuxième exemple est celui de la protéine EPCA-2 (Early Prostate Antigen-2) proposé en 2007 comme biomarqueur du cancer de la prostate avec des performances supérieures au dosage du PSA.¹⁹² Par la suite des erreurs analytiques dans l'utilisation des méthodes de dosage ont été mises en évidence et reconnues par les auteurs.¹⁹³ L'article initial a été rétracté en 2012.¹⁹⁴

Dans d'autres situations, certains biomarqueurs franchissent avec succès les étapes de découverte puis de validation, mais ne sont pas pour autant utilisés en pratique clinique. Plusieurs raisons peuvent à nouveau être évoquées :

- Absence de bénéfice apporté par les nouveaux biomarqueurs par rapport à ceux déjà existants. Selon le type de biomarqueur, le bénéfice en question peut être au niveau diagnostique (les performances d'un nouveau biomarqueur ne permettent pas d'améliorer le diagnostic par rapport à ceux déjà utilisés) mais également au niveau du bénéfice en survie des patients (l'utilisation d'un nouveau biomarqueur n'apporte au final aucun bénéfice en terme de survie pour les patients).
- Manque de sensibilité ou de spécificité pour une application pratique. Ce point est notamment fondamental si un biomarqueur est développé pour le dépistage d'une maladie. Dans ces situations, ces deux paramètres doivent être élevés pour permettre de réellement détecter les personnes malades (sensibilité) et d'écarter avec suffisamment de certitude une maladie si le test est négatif (spécificité). La marge d'erreur tolérée en pratique clinique est faible pour limiter le risque de décision médicale inadaptée.
- Difficultés à transposer en pratique la technique nécessaire à l'évaluation du biomarqueur. Une technique lourde et coûteuse sera un obstacle pour une utilisation à grande échelle d'un biomarqueur. Il s'agit sans doute là d'un des obstacles les plus importants ayant souvent été le facteur limitant pour le transfert des candidats identifiés par des approches protéomiques par spectrométrie de masse. Le coût et l'expertise nécessaire pour l'utilisation de ces instruments sont des obstacles à leur implantation large dans les hôpitaux.

Un exemple de la littérature est celui de B7-H4 proposé en 2006 comme biomarqueur du cancer de l'ovaire. Après mise en évidence d'une hyper-expression du gène dans les cancers de l'ovaire et du sein, une méthode ELISA a été développée pour mesurer l'expression de la protéine dans plus de 2500 prélèvements (sérum, ascite, tissus).¹⁹⁵ Une validation indépendante n'a cependant pas montré de bénéfice de ce biomarqueur par rapport à l'utilisation du CA125.¹⁹⁶

L'exemple de la LLC permet également d'illustrer ces situations. Le statut mutationnel du gène *IGHV* a montré son impact pronostique dans différentes études. Cependant, l'absence de conséquence sur le plan thérapeutique à ce jour explique que cette recherche ne soit pas nécessaire en pratique clinique. A l'opposé, toujours dans la LLC, la recherche d'une délétion 17p et/ou d'une mutation du gène *TP53* est devenue indispensable car les patients porteurs de ces anomalies bénéficient à présent de thérapies ciblées (inhibiteurs de la voie du BCR).

2. RECHERCHE DE BIOMARQUEURS PAR ANALYSE PROTEOMIQUE : SUCCES ET DIFFICULTES RENCONTREES DANS LES DIFFERENTS PROJETS

Les différents projets menés au laboratoire depuis 2004 ont progressivement bénéficié des progrès technologiques et bio-informatiques. Le nombre de protéines identifiées a été globalement multiplié par dix en l'espace d'une dizaine d'années (Figure 62).

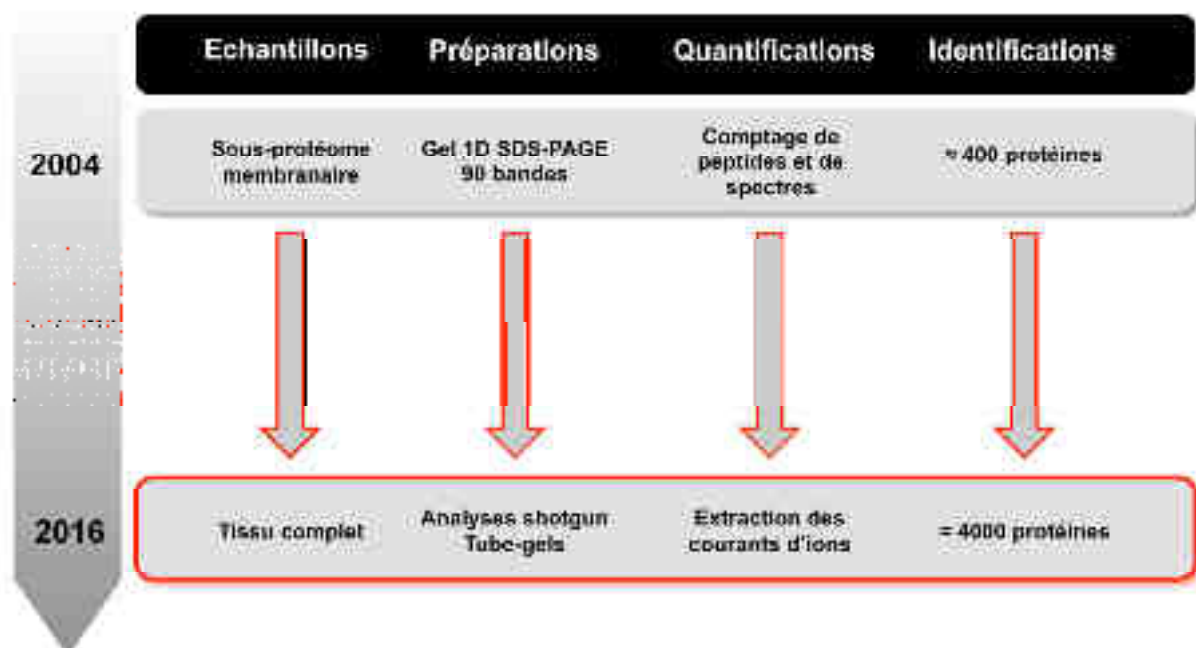


Figure 62
Impact des progrès technologiques récents sur les différentes étapes des procédures expérimentales utilisées dans le laboratoire entre 2004 et 2016.

La recherche par analyse protéomique de nouveau biomarqueurs se heurte cependant aux obstacles décrits précédemment.

L'utilisation du CD148 en tant que biomarqueur diagnostique pour les lymphomes à cellules du manteau rencontre plus de difficultés à trouver sa place en pratique clinique que celle du CD180 pour le diagnostic des lymphomes de la zone marginale. Malgré sa validation par cytométrie en flux, le bénéfice apporté par le CD148 par rapport aux biomarqueurs déjà disponibles (cytologie, histologie, cytogénétique avec présence d'une t(11;14), hyper-expression de la cycline D1 en biologie moléculaire ou immunohistochimie) est marginal. Ses performances en termes de sensibilité (78 %) et de spécificité (91 %) ne lui permettent pas de remplacer les biomarqueurs déjà établis.

Le projet Protell n'a quant à lui pas permis d'identifier à ce jour de biomarqueur pronostique du risque de rechute cérébrale dans les lymphomes B diffus à grandes cellules à partir des MP plasmatiques circulantes. L'analyse du projet a posteriori permet de mettre en évidence les difficultés rencontrées aux différentes étapes :

- Etape pré-analytique : le choix d'utiliser les MP plasmatiques circulantes a été fait dans la suite des projets CD148 et CD180. Ces derniers étaient cependant basés sur l'utilisation de MP induites à partir de cellules tumorales. Or, la composition des MPPC n'est pas équivalente à celle des MP induites. Les MPPC sont en effet très majoritairement d'origine plaquettaire et sont extrêmement sensibles aux conditions de prélèvement et de stockage des échantillons. En raison de ces modifications survenant *in vitro*, les variations mesurées expérimentalement risquent de ne pas refléter les variations survenues *in vivo*. Cet aspect nécessite une grande rigueur au moment de la collecte et de la conservation des échantillons. Malgré le fait que des consignes précises étaient cependant fixées dans le cadre du projet Protell, cet aspect ne peut pas être exclu.
- Etape analytique : une fois le culot de MPPC obtenu, des étapes de lavages sont réalisées pour diminuer la contamination plasmatique. Malgré ces lavages, une contamination plasmatique résiduelle est toujours présente et son importance peut varier aléatoirement d'un échantillon à un autre.
- Étapes post-analytiques : l'analyse statistique est rendue difficile par le nombre limité de patients de la cohorte 1 analysés en MPPC ayant eu une rechute cérébrale. L'hypothèse initiale utilisée pour le calcul des effectifs était basée sur un taux de rechute cérébrale estimé à 25 % dans cette population ciblée de patients à haut risque. Or, ce pourcentage s'est avéré inférieur avec, de plus, une collecte de plasma qui n'a pas pu être réalisée pour tous les patients ayant présenté une rechute cérébrale.

Cependant le projet Protell se distingue par le nombre important d'échantillons analysés dans le cadre d'une étude protéomique. Les différents critères de qualité mis en place ont permis de démontrer la stabilité du couplage nanoLC-MS/MS sur une période d'analyse de quatre semaines. La mise en place de tels critères qualité est indispensable pour des analyses s'étalant sur une période de temps aussi grande.

Pour la réalisation du projet Malyzoma, des difficultés ont également été rencontrées au niveau pré-analytique. L'extraction et la solubilisation des protéines à partir d'un culot cellulaire congelé ont été compliquées par la persistance de résidus non solubles. Or, une extraction et une solubilisation optimales des protéines sont nécessaires pour obtenir des résultats de quantification fiables et reproductibles. Un nouveau mode opératoire a été mis en place pour effectuer ces étapes à partir de cellules non congelées. Cette extraction « en temps réel » entraîne cependant un surcroît de travail au niveau du laboratoire d'hématologie où les échantillons sont préparés, ce qui ralentit la constitution d'une cohorte d'échantillons utiles.

Les différents enseignements tirés de ces différents projets ont été appliqués pour la recherche de biomarqueurs prédictifs de la réponse au traitement dans les LBDGC :

- utilisation du tissu tumoral à partir duquel le diagnostic a été posé. Ce choix permet également d'envisager après la phase de découverte une validation des biomarqueurs potentiels par des techniques d'immunohistochimie sur le même type de tissu tumoral,
- constitution de deux cohortes avec des effectifs identiques et pour lesquelles toutes les données cliniques et biologiques sont disponibles,
- choix des patients adapté à la question médicale posée : le seul critère utilisé pour la sélection des patients a été la réponse obtenue au traitement de première ligne,
- sélection d'un nombre suffisant d'échantillons pour permettre la réalisation des tests statistiques.

La mise en place de cette démarche expérimentale pour la phase de découverte a permis de mettre en évidence différents biomarqueurs potentiels associés à la chimiorésistance dans les LBDGC. Ces résultats doivent à présent être confirmés et validés sur des cohortes indépendantes de patients par des techniques d'immunohistochimie, de séquençage du transcriptome ou de protéomique quantitative ciblée.

3. PERSPECTIVES : PROTEOGENOMIQUE ET ONCO-PROTEOGENOMIQUE

Les différentes approches « omiques » à haut débit génèrent actuellement des quantités considérables d'informations. Le domaine de la protéogénomique se situe à l'interface de ces différentes approches (figure 63).¹⁹⁷ L'intégration des connaissances acquises par chacune d'elles représente actuellement un enjeu majeur pour permettre :

- de corriger ou compléter les annotations de gènes dans les banques de séquences nucléotidiques,
- de constituer dans le domaine de la cancérologie des banques de séquences protéiques spécifiques à partir de l'ensemble des anomalies génétiques des tumeurs acquises par les techniques de séquençage de nouvelle génération (onco-protéogénomique).

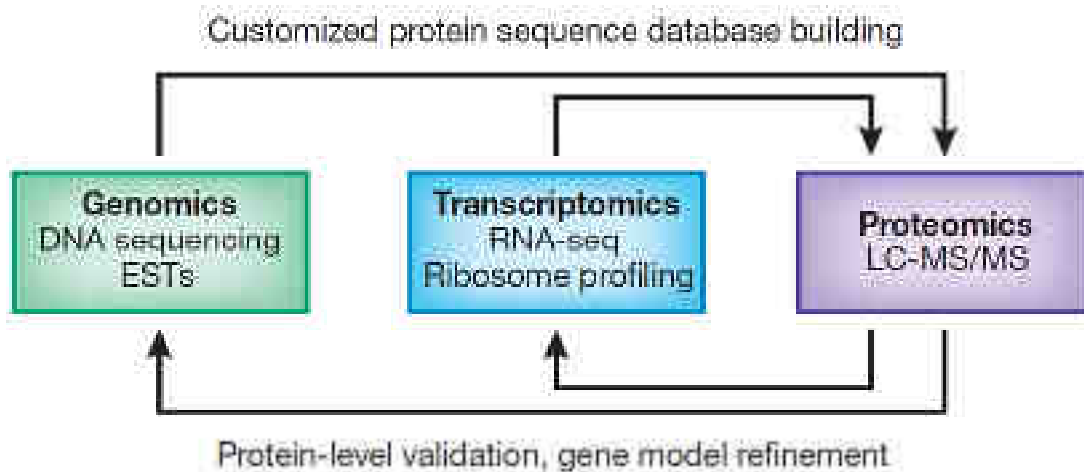


Figure 63
Interactions entre les différentes approches « omiques » (protéogénomique) (d'après référence n° 195).

La qualité et l'exhaustivité des banques de séquences protéiques utilisées pour interpréter les données MS/MS sont des éléments cruciaux pour les approches protéomiques, en particulier de type « bottom-up » puisque reposant sur l'identification des protéines à partir des séquences répertoriées dans les banques. Ces banques de données protéiques découlent directement des banques de séquences nucléotidiques. L'annotation des gènes dans ces banques génomiques repose sur des prédictions de séquences codantes générées par des logiciels bio-informatiques. Des erreurs d'annotations dans les séquences nucléotidiques ont donc un impact majeur sur les banques protéiques qui en découlent.

Plusieurs types d'erreurs sont possibles au moment de cette étape d'annotation des gènes :

- risque de sur-prédiction (cas d'une protéine figurant dans les banques mais qui n'est pas exprimée) ou de sous-prédiction (cas d'une protéine exprimée mais non prédite par le processus d'annotation de gènes),
- erreurs sur les séquences N-terminales ou C-terminales des protéines par erreur de prédiction des codons d'initiation ou des codons stop,
- erreurs de prédiction des décalages de cadres de lecture,
- erreurs de prédiction des épissages alternatifs.

Les progrès technologiques et bio-informatiques récents ainsi que la diversité des approches utilisées en spectrométrie de masse ont permis d'accroître considérablement la profondeur de couverture du protéome d'un organisme ou d'un tissu.¹⁹⁸ Ces données expérimentales acquises à partir des analyses protéomiques permettent à présent de valider les séquences codantes prédites (identification certaine d'une protéine par analyse protéomique), d'identifier de nouvelles séquences codantes ou de corriger certaines annotations de gènes à partir de la séquence protéique réelle (en utilisant des banques de séquences génomiques traduites dans les six cadres de lecture ou du séquençage *de novo*).

Parallèlement, le développement extrêmement rapide des techniques de séquençage de nouvelle génération¹⁹⁹ (séquençages du génomes, de l'exome ou du transcriptome) a largement contribué ces dernières années à mieux comprendre la complexité et l'hétérogénéité des pathologies cancéreuses au niveau génétique : mutations ponctuelles, fusions de gènes, insertions/délétions de bases ou présence de sites d'épissages aberrants. Toutes ces informations acquises à partir du séquençage des tumeurs ne sont cependant pas encore transposées dans les banques de séquences protéiques de référence utilisées par les moteurs de recherche en protéomique. Cet aspect constitue un obstacle majeur pour l'identification de biomarqueurs protéiques spécifiques aux tumeurs. Le développement de l'onco-protéogénomique a ainsi pour objectif de constituer des banques de données protéiques propres à chaque tumeur intégrant l'ensemble des informations fournies par les techniques de séquençage à haut débit.²⁰⁰

● Conclusion générale

La recherche de biomarqueurs reste un domaine d'investigation important dans le domaine de la cancérologie. Leur champ d'application est vaste, concernant aussi bien l'évaluation du risque de développer un cancer, le dépistage de la maladie, le diagnostic, l'évaluation du pronostic ou la réponse au traitement. L'analyse protéomique par spectrométrie de masse a bénéficié de progrès technologiques considérables ces dernières années, permettant d'obtenir des informations de plus en plus précises sur le protéome humain et d'en quantifier les variations. Malgré de nombreuses tentatives jusqu'à présent infructueuses dans la recherche en protéomique de biomarqueurs validés en pratique clinique, les progrès récents nécessitent et justifient la poursuite active des recherches dans ce domaine.

Centrés sur la thématique des hémopathies lymphoïdes B, les travaux menés dans le laboratoire ont permis d'identifier le CD148 et plus particulièrement le CD180 comme biomarqueurs diagnostiques dans ces pathologies. Mon travail de thèse, qui s'est inscrit dans la continuité de ces travaux, a pleinement bénéficié de l'ensemble des progrès technologiques et du développement de nouvelles approches quantitatives en analyse protéomique :

- 1) La mise au point d'une procédure expérimentale en protéomique quantitative ciblée permettra l'utilisation de ce type d'approche pour la validation de nouveaux biomarqueurs en s'affranchissant des contraintes imposées par les techniques immunologiques conventionnelles (nécessité de disposer d'un anticorps monoclonal qui ne cible qu'un épitope déterminé de la protéine).
- 2) Le projet PROTELL a démontré la possibilité d'appliquer des approches quantitatives globales sur un très grand nombre d'échantillons. La mise en place de critères pour le suivi de la qualité des analyses a permis de confirmer la stabilité d'un couplage nanoLC-MS/MS sur une période d'analyse de quatre semaines.
- 3) Le mode de préparation « tube-gel » facilite l'analyse de grandes séries d'échantillons en offrant un gain de temps important par rapport à la réalisation de gels classiques 1D SDS-PAGE. Cette méthode de préparation, compatible avec l'utilisation de détergents forts tels que le SDS (nécessaire pour obtenir extraction et une solubilisation efficaces des protéines), est particulièrement adapté pour des approches protéomiques quantitatives de type « shotgun ».

Malgré tout de nombreuses difficultés persistent : dans la cadre des projets développés dans mon travail de thèse, la validation de l'approche protéomique quantitative ciblée par SRM dans le cadre du projet Malyzoma se heurte au problème de la collecte des échantillons dont la préparation impose des critères contraignant pour être appliquée à grande échelle. L'analyse et l'interprétation des résultats du projet Protell sont compliquées du fait de multiples difficultés rencontrées aux niveaux pré-analytique, analytique et post-analytique (effectifs des cohortes, matériel biologique choisi pour l'analyse protéomique et choix de la méthode statistique à appliquer notamment).

Pour bénéficier pleinement des progrès technologiques récents, la recherche de biomarqueurs par analyse protéomique doit impérativement prendre en compte certains paramètres :

- 1) définition précise de la problématique scientifique et médicale,
- 2) choix d'un matériel biologique adapté à cette problématique, ainsi qu'à la validation des potentiels biomarqueurs et à leur utilisation en pratique clinique,
- 3) sélection rigoureuse des patients pour la collecte des échantillons biologiques,
- 4) maîtrise des outils bio-informatiques et des analyses statistiques,
- 5) confrontation et interprétation des résultats en fonction du rationnel biologique reliant les potentiels biomarqueurs à la question scientifique initiale.

Au cours de mon doctorat, j'ai eu pour objectif de mettre en application ces principes par la mise en place d'un projet pour la recherche de biomarqueurs associés à la chimiorésistance des lymphomes B diffus à grandes cellules. Différentes cibles potentielles ont pu être mises en évidence. Elles ouvrent des perspectives pour mieux comprendre la physiopathologie de ces maladies (notamment les interactions entre la cellule tumorale et son micro-environnement), les mécanismes moléculaires impliqués dans la chimiorésistance ainsi que l'importance de certaines cibles thérapeutiques. Tous ces résultats devront néanmoins faire l'objet d'une vérification et d'une validation rigoureuses. Le choix du tissu ganglionnaire comme matériel biologique initial permet d'envisager la validation de certaines protéines par des techniques immunohistochimiques qui sont parfaitement adaptées à une utilisation ultérieure en pratique clinique. L'extraction d'ADN et d'ARN à partir de ces mêmes tissus permet également d'envisager de corrélérer les résultats fournis par les approches de protéomique et de séquençage de nouvelle génération (séquençage d'exome et du transcriptome).

Un dernier enjeu majeur est directement lié à la quantité considérable d'informations générées actuellement par les nouvelles approches « omiques ». Leur partage au sein de la communauté scientifique doit permettre d'accroître significativement la quantité et la qualité des données constituant les différentes banques génomiques et protéiques. La protéogénomique émerge actuellement grâce à la corrélation de ces données génomiques, transcriptomiques et protéomiques. Elle permettra de faire évoluer l'analyse protéomique grâce à la constitution de banques de séquences protéiques « personnalisées » reflétant les profondes anomalies génétiques à l'origine des pathologies cancéreuses et permettant d'identifier des biomarqueurs protéiques spécifiques au tissu tumoral.

Une collaboration étroite entre les cliniciens et les équipes de recherche reste sans doute, au final, un des points les plus importants pour relever ces défis et mener à bien ces projets.

● Références

1. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69**, 89–95 (2001).
2. Harris, N. L. *et al.* A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood* **84**, 1361–1392 (1994).
3. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. *Blood* **89**, 3909–3918 (1997).
4. Jaffe, E. S., Harris, N. L., Stein, H. & Vardiman, J. . *Tumours of haematopoietic and lymphoid tissues*. (IARC : Lyon, 2001).
5. Swerdlow, S. H., Campo, E. & Harris, N. L. *WHO classification of tumours of haematopoietic and lymphoid tissues*. (IARC : Lyon, 2008).
6. Swerdlow, S. H. *et al.* The 2016 revision of the World Health Organization (WHO) classification of lymphoid neoplasms. *Blood* (2016). doi:10.1182/blood-2016-01-643569
7. Les cancers en France, Les Données, INCa, janvier 2014.
8. Matutes, E. *et al.* The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL. *Leukemia* **8**, 1640–1645 (1994).
9. Moreau, E. J. *et al.* Improvement of the chronic lymphocytic leukemia scoring system with the monoclonal antibody SN8 (CD79b). *Am. J. Clin. Pathol.* **108**, 378–382 (1997).
10. Binet, J. L. *et al.* A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer* **48**, 198–206 (1981).
11. Montserrat, E., Sanchez-Bisono, J., Viñolas, N. & Rozman, C. Lymphocyte doubling time in chronic lymphocytic leukaemia: analysis of its prognostic significance. *Br. J. Haematol.* **62**, 567–575 (1986).
12. Juliusson, G. *et al.* Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. *N. Engl. J. Med.* **323**, 720–724 (1990).
13. Döhner, H. *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* **343**, 1910–1916 (2000).
14. Gaidano, G. *et al.* p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5413–5417 (1991).
15. Dohner, H. *et al.* p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. *Blood* **85**, 1580–1589 (1995).
16. Hallek, M. *et al.* Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* **376**, 1164–1174 (2010).
17. Klein, U. & Dalla-Favera, R. Germinal centres: role in B-cell physiology and malignancy. *Nat. Rev. Immunol.* **8**, 22–33 (2008).
18. Oscier, D. G., Thompsett, A., Zhu, D. & Stevenson, F. K. Differential rates of somatic hypermutation in V(H) genes among subsets of chronic lymphocytic leukemia defined by chromosomal abnormalities. *Blood* **89**, 4153–4160 (1997).
19. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94**, 1848–1854 (1999).

20. Damle, R. N. *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94**, 1840–1847 (1999).
21. Crespo, M. *et al.* ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N. Engl. J. Med.* **348**, 1764–1775 (2003).
22. Thunberg, U. *et al.* CD38 expression is a poor predictor for VH gene mutational status and prognosis in chronic lymphocytic leukemia. *Blood* **97**, 1892–1894 (2001).
23. Vroblova, V., Smolej, L. & Krejsek, J. Pitfalls and limitations of ZAP-70 detection in chronic lymphocytic leukemia. *Hematology* **17**, 268–274 (2012).
24. Fabbri, G. *et al.* Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J. Exp. Med.* **208**, 1389–1401 (2011).
25. Rossi, D. *et al.* Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood* **119**, 2854–2862 (2012).
26. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2012).
27. Zenz, T. *et al.* TP53 mutation and survival in chronic lymphocytic leukemia. *J. Clin. Oncol.* **28**, 4473–4479 (2010).
28. Byrd, J. C. *et al.* Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia. *N. Engl. J. Med.* **371**, 213–223 (2014).
29. Farooqui, M. Z. H. *et al.* Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with TP53 aberrations: a phase 2, single-arm trial. *Lancet Oncol.* **16**, 169–176 (2015).
30. Furman, R. R. *et al.* Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N. Engl. J. Med.* **370**, 997–1007 (2014).
31. Van Den Neste, E. *et al.* Outcome of patients with relapsed diffuse large B-cell lymphoma who fail second-line salvage regimens in the International CORAL study. *Bone Marrow Transplant.* **51**, 51–57 (2016).
32. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N. Engl. J. Med.* **329**, 987–994 (1993).
33. Coiffier, B. *et al.* CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346**, 235–242 (2002).
34. Ziepert, M. *et al.* Standard international prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J. Clin. Oncol.* **28**, 2373–2380 (2010).
35. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
36. Hans, C. P. *et al.* Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **103**, 275–282 (2004).
37. Johnson, N. A. *et al.* Lymphomas with concurrent BCL2 and MYC translocations: the critical factors associated with survival. *Blood* **114**, 2273–2279 (2009).
38. Landsburg, D. J. *et al.* Impact of oncogene rearrangement patterns on outcomes in patients with double-hit non-Hodgkin lymphoma. *Cancer* **122**, 559–564 (2016).
39. Dubois, S. *et al.* Next Generation Sequencing in Diffuse Large B Cell Lymphoma Highlights Molecular Divergence and Therapeutic Opportunities: a LYSA Study. *Clin. Cancer Res.* (2016). doi:10.1158/1078-0432.CCR-15-2305
40. Rinaldi, A. *et al.* Genome-wide DNA profiling of marginal zone lymphomas identifies subtype-specific lesions with an impact on the clinical outcome. *Blood* **117**, 1595–1604 (2011).
41. Salido, M. *et al.* Cytogenetic aberrations and their prognostic value in a series of 330 splenic marginal zone B-cell lymphomas: a multicenter study of the Splenic B-Cell Lymphoma Group. *Blood* **116**, 1479–1488 (2010).

42. Parry, M. *et al.* Genetics and prognostication in splenic marginal zone lymphoma: Revelations from deep sequencing. *Clin. Cancer Res.* **21**, 4174–4183 (2015).
43. Olsen, J. V., Ong, S.-E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **3**, 608–614 (2004).
44. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C. & Yates, J. R. Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **113**, 2343–2394 (2013).
45. Hillenkamp, F. & Karas, M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol.* **193**, 280–295 (1990).
46. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).
47. Thiede, B. *et al.* Peptide mass fingerprinting. *Methods* **35**, 237–247 (2005).
48. Unlü, M., Morgan, M. E. & Minden, J. S. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071–2077 (1997).
49. Rabilloud, T., Chevallet, M., Luche, S. & Lelong, C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J. Proteomics* **73**, 2064–2077 (2010).
50. Anderson, N. L. & Anderson, N. G. The Human plasma proteome history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845–867 (2002).
51. Gillet, L. C. *et al.* 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**, O111.016717 (2012).
52. Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683–1695 (2009).
53. Wu, C. C. & MacCoss, M. J. Shotgun proteomics: tools for the analysis of complex biological systems. *Curr. Opin. Mol. Ther.* **4**, 242–250 (2002).
54. Pirmoradian, M. *et al.* Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol. Cell Proteomics* **12**, 3330–3338 (2013).
55. Bantscheff, M., Lemeer, S., Savitski, M. M. & Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404**, 939–965 (2012).
56. Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* **1**, 376–386 (2002).
57. Hebert, A. S. *et al.* Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat. Methods* **10**, 332–334 (2013).
58. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
59. Ross, P. L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics* **3**, 1154–1169 (2004).
60. Mueller, L. N., Brusniak, M.-Y., Mani, D. R. & Aebersold, R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J. Proteome Res.* **7**, 51–61 (2008).
61. Gao, J., Opiteck, G. J., Friedrichs, M. S., Dongre, A. R. & Hefta, S. A. Changes in the protein expression of yeast as a function of carbon source. *J. Proteome Res.* **2**, 643–649 (2003).
62. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).
63. Chelius, D. & Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **1**, 317–323 (2002).
64. Nesvizhskii, A. I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **4**, 1419–1440 (2005).

65. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
66. Geer, L. Y. *et al.* Open mass spectrometry search algorithm. *J. Proteome Res.* **3**, 958–964 (2004).
67. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
68. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
69. Searle, B. C. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **10**, 1265–1269 (2010).
70. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
71. Carapito, C. *et al.* Computational and mass-spectrometry-based workflow for the discovery and validation of missing human proteins: application to chromosomes 2 and 14. *J. Proteome Res.* **14**, 3621–3634 (2015).
72. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513–2526 (2014).
73. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555–566 (2012).
74. Lane, L. *et al.* neXtProt: a knowledge platform for human proteins. *Nucleic Acids Res.* **40**, D76–D83 (2012).
75. Gillette, M. A. & Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* **10**, 28–34 (2013).
76. Scott, K. B., Turko, I. V. & Phinney, K. W. in *Methods in Enzymology* (ed. Kelman, Z.) **566**, 289–303 (Academic Press, 2016).
77. Picard, G. *et al.* PSAQTM standards for accurate MS-based quantification of proteins: from the concept to biomedical applications. *J Mass Spectrom* **47**, 1353–1363 (2012).
78. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
79. Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M. & Berglund, L. Prediction of the human membrane proteome. *Proteomics* **10**, 1141–1149 (2010).
80. Zhang, N. *et al.* Comparison of SDS- and methanol-assisted protein solubilization and digestion methods for Escherichia coli membrane proteome analysis by 2-D LC-MS/MS. *Proteomics* **7**, 484–493 (2007).
81. Raposo, G. & Stoorvogel, W. Extracellular vesicles: Exosomes, microvesicles, and friends. *J. Cell Biol.* **200**, 373–383 (2013).
82. Bevers, E. M., Comfurius, P., Dekkers, D. W. & Zwaal, R. F. Lipid translocation across the plasma membrane of mammalian cells. *Biochim. Biophys. Acta* **1439**, 317–330 (1999).
83. Tramontano, A. F. *et al.* Circulating endothelial microparticles in diabetes mellitus. *Mediators Inflamm.* **2010**, 250476 (2010).
84. Amabile, N. *et al.* Increased CD62e(+) endothelial microparticle levels predict poor outcome in pulmonary hypertension patients. *J. Heart Lung Transplant.* **28**, 1081–1086 (2009).
85. González-Quintero, V. H. *et al.* Elevated plasma endothelial microparticles: preeclampsia versus gestational hypertension. *Am. J. Obstet. Gynecol.* **191**, 1418–1424 (2004).
86. Bernal-Mizrachi, L. *et al.* High levels of circulating endothelial microparticles in patients with acute coronary syndromes. *Am. Heart J.* **145**, 962–970 (2003).
87. Perez-Hernandez, J. & Cortes, R. Extracellular vesicles as biomarkers of systemic lupus erythematosus. *Dis. Markers* **2015**, 613536 (2015).

88. Gong, J., Jaiswal, R., Dalla, P., Luk, F. & Bebawy, M. Microparticles in cancer: A review of recent developments and the potential for clinical application. *Semin. Cell Dev. Biol.* **40**, 35–40 (2015).
89. Miguet, L. *et al.* Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. *Proteomics* **6**, 153–171 (2006).
90. Miguet, L. *et al.* Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J. Proteome Res.* **8**, 3346–3354 (2009).
91. Honda, H., Inazawa, J., Nishida, J., Yazaki, Y. & Hirai, H. Molecular cloning, characterization, and chromosomal localization of a novel protein-tyrosine phosphatase, HPTP *et al.* *Blood* **84**, 4186–4194 (1994).
92. Autschbach, F. *et al.* Expression of the membrane protein tyrosine phosphatase CD148 in human tissues. *Tissue Antigens* **54**, 485–498 (1999).
93. Tangye, S. G., Phillips, J. H., Lanier, L. L., de Vries, J. E. & Aversa, G. CD148: a receptor-type protein tyrosine phosphatase involved in the regulation of human T cell activation. *J. Immunol.* **161**, 3249–3255 (1998).
94. Ostman, A., Yang, Q. & Tonks, N. K. Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase, is enhanced with increasing cell density. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9680–9684 (1994).
95. Jandt, E., Denner, K., Kovalenko, M., Ostman, A. & Böhmer, F.-D. The protein-tyrosine phosphatase DEP-1 modulates growth factor-stimulated cell migration and cell-matrix adhesion. *Oncogene* **22**, 4175–4185 (2003).
96. Tangye, S. G., Liu, Y. J., Aversa, G., Phillips, J. H. & de Vries, J. E. Identification of functional human splenic memory B cells by expression of CD148 and CD27. *J. Exp. Med.* **188**, 1691–1703 (1998).
97. Balavenkatraman, K. K. *et al.* DEP-1 protein tyrosine phosphatase inhibits proliferation and migration of colon carcinoma cells and is upregulated by protective nutrients. *Oncogene* **25**, 6319–6324 (2006).
98. Ruivenkamp, C. A. L. *et al.* Ptp^{prj} is a candidate for the mouse colon-cancer susceptibility locus *Scc1* and is frequently deleted in human cancers. *Nat. Genet.* **31**, 295–300 (2002).
99. Ruivenkamp, C. *et al.* LOH of PTPRJ occurs early in colorectal cancer and is associated with chromosomal loss of 18q12-21. *Oncogene* **22**, 3472–3474 (2003).
100. Miguet, L. *et al.* Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas. *Leukemia* **27**, 1748–1750 (2013).
101. Nagai, Y. *et al.* The radioprotective 105/MD-1 complex links TLR2 and TLR4/MD-2 in antibody response to microbial membranes. *J Immunol* **174**, 7043–7049 (2005).
102. Miyake, K., Yamashita, Y., Ogata, M., Sudo, T. & Kimoto, M. RP105, a novel B cell surface molecule implicated in B cell activation, is a member of the leucine-rich repeat protein family. *J. Immunol.* **154**, 3333–3340 (1995).
103. Isaza-Correa, J. M., Liang, Z., van den Berg, A., Diepstra, A. & Visser, L. Toll-like receptors in the pathogenesis of human B cell malignancies. *J Hematol Oncol* **7**, 57 (2014).
104. Yan, Q. *et al.* BCR and TLR signaling pathways are recurrently targeted by genetic changes in splenic marginal zone lymphomas. *Haematologica* **97**, 595–598 (2012).
105. Martinez-Lopez, A. *et al.* MYD88 (L265P) somatic mutation in marginal zone B-cell lymphoma. *Am. J. Surg. Pathol.* **39**, 644–651 (2015).
106. Battle, R. & Clark, B. Quantitative analysis of human leucocyte antigen expression during culture of Epstein-Barr virus-transformed cell lines using the dako QIFIKIT. *Br. J. Biomed. Sci.* **64**, 32–34 (2007).
107. Palumbo, G. A. *et al.* CD200 expression may help in differential diagnosis between mantle cell lymphoma and B-cell chronic lymphocytic leukemia. *Leuk. Res.* **33**, 1212–1216 (2009).

108. Fan, L. *et al.* Expression patterns of CD200 and CD148 in leukemic B-cell chronic lymphoproliferative disorders and their potential value in differential diagnosis. *Leuk. Lymphoma* **56**, 3329–3335 (2015).
109. Mayeur-Rousse, C. *et al.* CD180 expression in B-cell lymphomas: a multicenter GEIL study. *Cytometry B Clin Cytom* (2015). doi:10.1002/cyto.b.21325
110. Hirokawa, T., Boon-Chieng, S. & Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinforma. Oxf. Engl.* **14**, 378–379 (1998).
111. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175–182 (1998).
112. Käll, L., Krogh, A. & Sonnhammer, E. L. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).
113. Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* **37**, W465–468 (2009).
114. Ferreri, A. J. M. Risk of CNS dissemination in extranodal lymphomas. *Lancet Oncol.* **15**, e159–169 (2014).
115. Berckmans, R. J. *et al.* Cell-derived microparticles circulate in healthy humans and support low grade thrombin generation. *Thromb Haemost* **85**, 639–46 (2001).
116. Skog, J. *et al.* Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol.* **10**, 1470–1476 (2008).
117. Savaşan, S., Büyükavcı, M., Buck, S. & Ravindranath, Y. Leukaemia/lymphoma cell microparticles in childhood mature B cell neoplasms. *J. Clin. Pathol.* **57**, 651–653 (2004).
118. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol* **57**, 289–300 (1995).
119. Gautier, V. *et al.* 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. *Mol. Cell Proteomics* **11**, 527–539 (2012).
120. Ghesquière, B. & Gevaert, K. Proteomics methods to study methionine oxidation. *Mass Spectrom. Rev.* **33**, 147–156 (2014).
121. Lu, X. & Zhu, H. Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. *Mol. Cell Proteomics* **4**, 1948–1958 (2005).
122. Han, C.-L. *et al.* A multiplexed quantitative strategy for membrane proteomics: opportunities for mining therapeutic targets for autosomal dominant polycystic kidney disease. *Mol. Cell Proteomics* **7**, 1983–1997 (2008).
123. Smolders, K., Lombaert, N., Valkenburg, D., Baggerman, G. & Arckens, L. An effective plasma membrane proteomics approach for small tissue samples. *Sci. Rep.* **5**, 10917 (2015).
124. Zhou, J. *et al.* Gel absorption-based sample preparation for the analysis of membrane proteome by mass spectrometry. *Anal. Biochem.* **404**, 204–210 (2010).
125. Cao, R. *et al.* High-throughput analysis of rat liver plasma membrane proteome by a nonelectrophoretic in-gel tryptic digestion coupled with mass spectrometry identification. *J. Proteome Res.* **7**, 535–545 (2008).
126. Cao, L., Clifton, J. G., Reutter, W. & Josic, D. Mass spectrometry-based analysis of rat liver and hepatocellular carcinoma Morris hepatoma 7777 plasma membrane proteome. *Anal. Chem.* **85**, 8112–8120 (2013).
127. Yu, H. *et al.* Quantifying raft proteins in neonatal mouse brain by ‘tube-gel’ protein digestion label-free shotgun proteomics. *Proteome Sci.* **5**, 17 (2007).
128. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

129. Fu, K. *et al.* Addition of rituximab to standard chemotherapy improves the survival of both the germinal center B-cell-like and non-germinal center B-cell-like subtypes of diffuse large B-cell lymphoma. *JCO* **26**, 4587–4594 (2008).
130. Thieblemont, C. *et al.* The germinal center/activated B-cell subclassification has a prognostic impact for response to salvage therapy in relapsed/refractory diffuse large B-cell lymphoma: a bio-CORAL study. *JCO* **29**, 4079–4087 (2011).
131. Maxwell, S. A., Cherry, E. M. & Bayless, K. J. Akt, 14-3-3 ζ , and vimentin mediate a drug-resistant invasive phenotype in diffuse large B-cell lymphoma. *Leuk. Lymphoma* **52**, 849–864 (2011).
132. Liu, Y. *et al.* Identification of differentially expressed proteins in chemotherapy-sensitive and chemotherapy-resistant diffuse large B cell lymphoma by proteomic methods. *Med. Oncol.* **30**, 1–10 (2013).
133. Beck, S. *et al.* The Impact II, a very high-resolution quadrupole time-of-flight instrument (QTOF) for deep shotgun proteomics. *Mol. Cell Proteomics* **14**, 2014–2029 (2015).
134. Delaleau, M. & Borden, K. L. B. Multiple export mechanisms for mRNAs. *Cells* **4**, 452–473 (2015).
135. Hautbergue, G. M. *et al.* UIF, a new mRNA export adaptor that works together with REF/ALY, requires FACT for recruitment to mRNA. *Current Biology* **19**, 1918–1924 (2009).
136. Busslinger, M. Transcriptional control of early B cell development. *Annu. Rev. Immunol.* **22**, 55–79 (2004).
137. Somasundaram, R., Prasad, M. A. J., Ungerback, J. & Sigvardsson, M. Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. *Blood* **126**, 144–152 (2015).
138. Coyaud, E. *et al.* Wide diversity of PAX5 alterations in B-ALL: a Groupe Francophone de Cytogénétique Hématologique study. *Blood* **115**, 3089–3097 (2010).
139. Teo, A. E. *et al.* Differential PAX5 levels promote malignant B-cell infiltration, progression and drug resistance, and predict a poor prognosis in MCL patients independent of CCND1. *Leukemia* **30**, 580–593 (2016).
140. Tedder, T. F., Inaoki, M. & Sato, S. The CD19–CD21 complex regulates signal transduction thresholds governing humoral immunity and autoimmunity. *Immunity* **6**, 107–118 (1997).
141. Xue, K. *et al.* PAX5 promotes pre-B cell proliferation by regulating the expression of pre-B cell receptor and its downstream signaling. *Mol. Immunol.* **73**, 1–9 (2016).
142. Naddafi, F. & Davami, F. Anti-CD19 monoclonal antibodies: a new approach to lymphoma therapy. *Int J Mol Cell Med* **4**, 143–151 (2015).
143. Topp, M. S. *et al.* Phase II trial of the anti-CD19 bispecific T cell-engager blinatumomab shows hematologic and molecular remissions in patients with relapsed or refractory B-precursor acute lymphoblastic leukemia. *J. Clin. Oncol.* **32**, 4134–4140 (2014).
144. Viardot, A. *et al.* Phase 2 study of the bispecific T-cell engager (BiTE) antibody blinatumomab in relapsed/refractory diffuse large B-cell lymphoma. *Blood* **127**, 1410–1416 (2016).
145. Soboloff, J., Rothberg, B. S., Madesh, M. & Gill, D. L. STIM proteins: dynamic calcium signal transducers. *Nat. Rev. Mol. Cell Biol.* **13**, 549–565 (2012).
146. Roos, J. *et al.* STIM1, an essential and conserved component of store-operated Ca²⁺ channel function. *J. Cell Biol.* **169**, 435–445 (2005).
147. Vacher, P. *et al.* Localized store-operated calcium influx represses CD95-dependent apoptotic effects of rituximab in non-hodgkin B lymphomas. *J Immunol* **195**, 2207–2215 (2015).
148. Nagel, D., Vincendeau, M., Eitelhuber, A. C. & Krappmann, D. Mechanisms and consequences of constitutive NF- κ B activation in B-cell lymphoid malignancies. *Oncogene* **33**, 5655–5665 (2014).

149. Davis, R. E., Brown, K. D., Siebenlist, U. & Staudt, L. M. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J. Exp. Med.* **194**, 1861–1874 (2001).
150. Bartuzi, P., Hofker, M. H. & van de Sluis, B. Tuning NF- κ B activity: a touch of COMMD proteins. *Biochim. Biophys. Acta* **1832**, 2315–2321 (2013).
151. Burstein, E. *et al.* COMMD proteins, a novel family of structural and functional homologs of MURR1. *J. Biol. Chem.* **280**, 22222–22232 (2005).
152. Wang, H. *et al.* ZAP-70: an essential kinase in T-cell signaling. *Cold Spring Harb Perspect Biol* **2**, a002279 (2010).
153. Chen, L. *et al.* Expression of ZAP-70 is associated with increased B-cell receptor signaling in chronic lymphocytic leukemia. *Blood* **100**, 4609–4614 (2002).
154. Fridberg, M. *et al.* Protein expression and cellular localization in two prognostic subgroups of diffuse large B-cell lymphoma: Higher expression of ZAP70 and PKC- β II in the non-germinal center group and poor survival in patients deficient in nuclear PTEN. *Leuk. Lymphoma* **48**, 2221–2232 (2007).
155. Lefevre, E. A. *et al.* Fibrinogen is localized on dark zone follicular dendritic cells in vivo and enhances the proliferation and survival of a centroblastic cell line in vitro. *J. Leukoc. Biol.* **82**, 666–677 (2007).
156. Dillon, L. M. *et al.* P-REX1 creates a positive feedback loop to activate growth factor receptor, PI3K/AKT and MEK/ERK signaling in breast cancer. *Oncogene* **34**, 3968–3976 (2015).
157. Goel, H. L. *et al.* P-Rex1 promotes resistance to VEGF/VEGFR-targeted therapy in prostate cancer. *Cell Rep* **14**, 2193–2208 (2016).
158. Cleator, S. *et al.* Gene expression patterns for doxorubicin (Adriamycin) and cyclophosphamide (Cytoxan) (AC) response and resistance. *Breast Cancer Res. Treat.* **95**, 229–233 (2005).
159. Qian, X. *et al.* The Tensin-3 protein, including its SH2 domain, is phosphorylated by Src and contributes to tumorigenesis and metastasis. *Cancer Cell* **16**, 246–258 (2009).
160. Suzuki, Y. *et al.* Association of CD20 levels with clinicopathological parameters and its prognostic significance for patients with DLBCL. *Ann. Hematol.* **91**, 997–1005 (2012).
161. Johnson, N. A. *et al.* Diffuse large B-cell lymphoma: reduced CD20 expression is associated with an inferior survival. *Blood* **113**, 3773–3780 (2009).
162. Seda, V. & Mraz, M. B-cell receptor signalling and its crosstalk with other pathways in normal and malignant cells. *Eur. J. Haematol.* **94**, 193–205 (2015).
163. Geahlen, R. L. Getting Syk: spleen tyrosine kinase as a therapeutic target. *Trends Pharmacol. Sci.* **35**, 414–422 (2014).
164. Sharman, J. *et al.* An open-label phase 2 trial of entospletinib (GS-9973), a selective spleen tyrosine kinase inhibitor, in chronic lymphocytic leukemia. *Blood* **125**, 2336–2343 (2015).
165. Friedberg, J. W. *et al.* Inhibition of Syk with fostamatinib disodium has significant clinical activity in non-Hodgkin lymphoma and chronic lymphocytic leukemia. *Blood* **115**, 2578–2585 (2010).
166. Maul, R. S. & Chang, D. D. EPLIN, epithelial protein lost in neoplasm. *Oncogene* **18**, 7838–7841 (1999).
167. Zhang, S. *et al.* Epidermal growth factor promotes protein degradation of epithelial protein lost in neoplasm (EPLIN), a putative metastasis suppressor, during epithelial-mesenchymal transition. *J. Biol. Chem.* **288**, 1469–1479 (2013).
168. Liu, Y., Sanders, A. J., Zhang, L. & Jiang, W. G. EPLIN- α Expression in Human Oesophageal Cancer and its Impact on Cellular Aggressiveness and Clinical Outcome. *Anticancer Res.* **32**, 1283–1289 (2012).
169. Sanders, A. J., Martin, T. A., Ye, L., Mason, M. D. & Jiang, W. G. EPLIN is a negative regulator of prostate cancer growth and invasion. *J. Urol.* **186**, 295–301 (2011).

170. Nie, Z. *et al.* Conversion of the LIMA1 tumour suppressor into an oncogenic LMO-like protein by API2–MALT1 in MALT lymphoma. *Nat. Commun.* **6**, 5908 (2015).
171. Meng, C., Kuster, B., Culhane, A. C. & Gholami, A. M. A multivariate approach to the integration of multi-omics datasets. *BMC Bioinformatics* **15**, 162 (2014).
172. Bady, P., Dolédec, S., Dumont, B. & Fruget, J.-F. Multiple co-inertia analysis: a tool for assessing synchrony in the temporal variability of aquatic communities. *C. R. Biol.* **327**, 29–36 (2004).
173. Ostasiewicz, P., Zielinska, D. F., Mann, M. & Wiśniewski, J. R. Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. *J. Proteome Res.* **9**, 3688–3700 (2010).
174. Gustafsson, O. J. R., Arentz, G. & Hoffmann, P. Proteomic developments in the analysis of formalin-fixed tissue. *Biochim. Biophys. Acta* **1854**, 559–580 (2015).
175. Wolff, C., Schott, C., Porschewski, P., Reischauer, B. & Becker, K.-F. Successful protein extraction from over-fixed and long-term stored formalin-fixed tissues. *PLoS One* **6**, e16353 (2011).
176. Sanger, F. *et al.* Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**, 687–695 (1977).
177. Saiki, R. K. *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354 (1985).
178. Mullis, K. B. & Faloona, F. A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**, 335–350 (1987).
179. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
180. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
181. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
182. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
183. Schuster, S. C. Next-generation sequencing transforms today's biology. *Nat. Methods* **5**, 16–18 (2008).
184. Reddy, T. B. K. *et al.* The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res.* **43**, D1099–1106 (2015).
185. Bast, R. C. *et al.* A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N. Engl. J. Med.* **309**, 883–887 (1983).
186. Kuriyama, M. *et al.* Quantitation of prostate-specific antigen in serum by a sensitive enzyme immunoassay. *Cancer Res.* **40**, 4658–4662 (1980).
187. Wang, M. C., Valenzuela, L. A., Murphy, G. P. & Chu, T. M. Purification of a human prostate specific antigen. *Invest Urol* **17**, 159–163 (1979).
188. Purves, L. R., MacNab, M., Geddes, E. W. & Bersohn, I. Serum-alpha-foetoprotein and primary hepatic cancer. *Lancet* **1**, 921–922 (1968).
189. Gold, P. & Freedman, S. O. Specific carcinoembryonic antigens of the human digestive system. *J. Exp. Med.* **122**, 467–481 (1965).
190. Xu, Y. *et al.* Lysophosphatidic acid as a potential biomarker for ovarian and other gynecologic cancers. *JAMA* **280**, 719–723 (1998).
191. Baker, D. L. *et al.* Plasma lysophosphatidic acid concentration and ovarian cancer. *JAMA* **287**, 3081–3082 (2002).
192. Leman, E. S. *et al.* EPCA-2: a highly specific serum marker for prostate cancer. *Urology* **69**, 714–720 (2007).
193. Hughes, V. Markers of dispute. *Nat. Med.* **15**, 1339–1343 (2009).

194. Retraction notice. "EPCA-2: a highly specific serum marker for prostate cancer" by Leman et al (Urology 2007;69:714-720). *Urology* **79**, 490 (2012).
195. Simon, I. *et al.* B7-h4 is a novel membrane-bound protein and a candidate serum and tissue biomarker for ovarian cancer. *Cancer Res.* **66**, 1570–1575 (2006).
196. Cramer, D. W. *et al.* Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer Prev Res (Phila)* **4**, 365–374 (2011).
197. Nesvizhskii, A. I. Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* **11**, 1114–1125 (2014).
198. Altelaar, A. F. M., Munoz, J. & Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14**, 35–48 (2013).
199. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol. Cell* **58**, 586–597 (2015).
200. Alfaro, J. A., Sinha, A., Kislinger, T. & Boutros, P. C. Onco-proteogenomics: cancer proteomics joins forces with genomics. *Nat. Methods* **11**, 1107–1113 (2014).

● Procédures expérimentales

1. COLLECTE DES MICROPARTICULES PLASMATIQUES CIRCULANTES ET EXTRACTION DES PROTEINES (PROJET PROTELL)

Centrifugation du plasma à 18 000 g pendant 2 min pour obtenir du plasma pauvre en plaquettes (PPP).

Centrifugation du PPP à 18 000 g pendant 45 min pour obtenir le culot de microparticules (MP).

Lavage du culot de MP avec du PBS, centrifugation à 18 000 g pendant 45 min (2 lavages).

Repris du culot de MP dans 15 µL de tampon de Laemmli (10 mM Tris pH 8, 1 mM EDTA, 5 % β-mercaptoéthanol, 5 % SDS, 10 % glycérol et 0,1 % bleu de bromophénol).

Stockage à -20 °C avant dépôt sur gel 1D SDS-PAGE.

2. EXTRACTION DES PROTEINES A PARTIR D'UN CULOT CELLULAIRE (ANALYSES SRM)

Reprise du culot de cellules dans un tampon Laemmli sans bleu ni glycérol (10 mM Tris pH 8, 1 mM EDTA, 5 % β-mercaptoéthanol, 5 % SDS).

Sonication à l'aiguille sur glace (3 cycles de 10 secondes à 20 W).

Précipitation des protéines à l'acétone glaciale (ajout de 6 volumes d'acétone glaciale avec incubation pendant une nuit à -20 °C).

Stockage à -80 °C avant dépôt sur gel.

3. PROJET TUBE-GELS

3.1. Préparation des tube-gels

Pour 20 µL d'échantillon contenant 20 µg de protéines dans un tampon Urée 8 M / NH₄HCO₃ 0,1 M, ajout de :

- 20 µL de SDS 10 % (5 % final)
- 32,2 µL d'eau
- 25 µL d'acrylamide / bis-acrylamide à 30 %
- 0,25 µL de Temed
- 2,5 µL d'APS 10 %

Lavages (3 fois) dans de l'eau.

Fixation avec une solution éthanol /acide phosphorique (50/3, v/v).

3.2. Protocole de digestion liquide

20 µg de protéines de levure sont reprises dans un volume final de 200 µL de tampon urée 8 M / bicarbonate d'ammonium (NH_4HCO_3) 0,1 M.

Réduction des protéines avec 3,5 µL de dithiothréitol (DTT) 700 mM (12 mM final) pendant 30 minutes à 37°C sous agitation.

Alkylation des protéines avec 12 µL IAA 700 mM (40 mM final) pendant 1h à température ambiante dans l'obscurité.

Dilution avec 1500 µL de tampon NH_4HCO_3 à 0,1 M (Urée 1 M final).

Digestion par ajout de 6,25 µL de trypsine à 0,04 µg/µL pendant une nuit à 37 °C avec agitation. Arrêt de la digestion par acidification avec 10 µL d'acide formique.

Concentration et lavage de l'échantillon par Sep-Pak (phases C18, 1 cc 50 mg, Waters) :

- Conditionnement : 500 µL méthanol (2 fois), 500 µL acétonitrile (ACN) (2 fois) et 500 µL acide formique (HCOOH) 0,1 % (2 fois)
- Chargement des échantillons
- Lavage : 750 µL HCOOH 0,1 % (3 fois)
- Elution : 200 µL ACN/ HCOOH (60/0,1 v/v)

3.3. Réduction, alkylation et digestion des protéines en gels

Décoloration des bandes pour les gels stacking avec un mélange ACN/ NH_4HCO_3 25 mM (75/25 v/v).

Déshydratation des bandes avec 50 µL ACN.

Réduction des protéines avec 50 µL de DTT à 10 mM 30 min à 60 °C et 30 min à température ambiante.

Alkylation des protéines avec 50 µL IAA à 55 mM 20 min à température ambiante dans l'obscurité.

Lavage (3 fois) des bouts de gels avec un mélange NH_4HCO_3 25 mM/ACN (50/50 v/v).

Déshydratation des bandes de gels avec 50 µL d'ACN.

Digestion des protéines par de la trypsine à 12,5 ng/µL, avec un rapport 1/100 (enzyme/protéine), pendant la nuit à 37 °C dans un tampon NH_4HCO_3 0,1 M.

4. EXTRACTION DES PROTEINES A PARTIR DE FRAGMENTS DE TISSU GANGLIONNAIRE

Broyage mécanique des fragments de tissus au potter, sur glace, dans un tampon de type Laemmli (10 mM Tris pH 8, 1 mM EDTA, 5 % SDS) : 1^{ère} extraction avec 100 µL de tampon puis 2^{ème} extraction avec 85 µL de tampon jusqu'à dissolution complète des tissus.

Sonication sur glace pendant 5 minutes.

Dosage DC (Bio-Rad) des protéines extraites.

5. COUPLAGE ULTIMATE 3000 – TSQ VANTAGE (PROJET SRM)

Le mélange peptidique est chargé pendant 3 min à 50 µL/min sur une précolonne d'enrichissement Agilent Zorbax 300SB-C18 (40 nL, 5 µm) et séparé sur une colonne analytique Agilent Zorbax 300SB-C18 (150 × 0,3 mm, 3,5 µm) à 5 µL/min. Le solvant A contient 98 % d'eau, 2 % d'acétonitrile et 0,1 % d'acide formique. Le solvant B contient 98 % d'acétonitrile, 2 % d'eau et 0,1 % d'acide formique. Le gradient comporte trois pentes : 8 à 15 % de B de 3 à 15 min, puis 15 à 22 % B de 15 à 65 min et de 22 à 40 % de B de 60 à 90 min. Le temps de cycle est de 2,4 s. 273 transitions sont suivies en mode *scheduled*. La tension appliquée sur le capillaire est de 3000 V et la température de la source de 300 °C. La résolution sur Q1 est de 0,7 Da. L'acquisition des données est réalisée avec le logiciel Xcalibur 2.1 (Thermo Fisher Scientific).

6. COUPLAGE NANOACQUITY – IMPACT HD (PROJET PROTELL)

Le système de chromatographie nanoAcquity est couplé à un spectromètre de masse hybride quadripôle – temps de vol (Q-TOF). La précolonne est une phase inverse C18 de dimension 180 µm × 20 mm nanoAcquity™ UPLC™ de porosité 5 µm (Waters). La colonne analytique est une phase inverse C18 de dimension 75 µm × 250 mm nanoAcquity™ UPLC™, de porosité 1,7 µm (Waters). Le gradient est réalisé par un mélange de deux solvants A et B. Le solvant A est une solution à 0,1 % en acide formique dans de l'eau et le solvant B est une solution à 0,1 % en acide formique dans de l'acétonitrile. Le gradient va de 1 à 33 % de B en 90 min. Le débit est fixé à 0,450 µL/min.

La source CaptiveSpray nanoBooster™ (Bruker) est utilisée. L'étalonnage interne est réalisé grâce à un système de *lock mass*. Celui-ci consiste en une infusion continue d'un composé de référence de masse $m/z = 922,09$. La tension du capillaire de l'électrospray est de 1,3 kV. La température de la source est de 150 °C. Le temps de cycle est de 3 s. Un scan MS dure 0,5 s. Le nombre et la durée des scans MS/MS sont établis suivant l'intensité des peptides en MS. L'appareil fonctionne à 4 Hz si l'intensité est de 3000 cps et à 25 Hz si l'intensité est de 25 000 cps. Les ions monochargés sont exclus de la sélection. L'exclusion dynamique est de 1 min. L'acquisition est réalisée à l'aide du logiciel HyStar v.3.2.

7. COUPLAGE NANOACQUITY – IMPACT II (PROJET D'ANALYSE DE GANGLIONS DE LBDGC)

Le système de chromatographie nanoAcquity est identique à celui du point 6. Le gradient va de 5 à 35 % de B en 120 min. Le débit est fixé à 0,450 µL/min.

La source CaptiveSpray nanoBooster™ (Bruker) est utilisée. L'étalonnage interne est réalisé grâce à un système de *lock mass*. Celui-ci consiste en une infusion continue d'un composé de référence de masse $m/z = 922,09$. La tension du capillaire de l'électrospray est de 1,3 kV. La température de la source est de 180 °C. Le temps de cycle est de 3 s. Un scan MS dure 0,5 s. Les ions monochargés sont exclus de la sélection. L'exclusion dynamique est de 1 min (sauf si un précurseur est revu avec une intensité trois fois supérieure pendant la période d'exclusion). L'acquisition est réalisée à l'aide du logiciel HyStar v.3.2.

8. COUPLAGE NANOACQUITY – Q EXACTIVE PLUS (PROJET TUBE-GELS)

Le système de chromatographie nanoAcquity est identique à celui du point 7. Le gradient va de 1 à 35 % de B en 79 min. Le débit est fixé à 0,450 $\mu\text{L}/\text{min}$.

La tension du capillaire de l'électrospray est de 1,8 kV. La température de la source est de 250 °C. Les données sont acquises en mode DDA avec une bascule automatique entre le mode MS (50 ms/scan sur une gamme de masse de 300 à 1800 m/z et une résolution à 70000) et le mode MS/MS (100 ms/scan sur une gamme de masse de 200 à 2000 m/z et une résolution à 17500). Les 10 ions les plus intenses sont sélectionnés pour la MS/MS (seuil d'intensité à 2×10^5). Les ions monochargés sont exclus de la sélection. L'exclusion dynamique est de 1 min. L'acquisition est réalisée à l'aide du logiciel Xcalibur v.3.0.63 (Thermo Fisher Scientific).

Annexe 1

Liste des 59 protéines parmi lesquelles a été identifié le CD180.

CD	Masse moléculaire (Da)	Nombre total de peptides			Nombre de spectres MS/MS			Couverture de séquence (%)		
		LPL	MCL	MZL	LPL	MCL	MZL	LPL	MCL	MZL
CD98	67978		2	9		7	67		4,29	20,5
CD233	101778	10	2		80	2		17,0	3,29	
CD79A	25021			3			31			31,0
CD79B	26031			5			98			31,0
CD72	40203			5			14			25,9
CD22	95329		12	13		25	56		13,1	20,4
CD19	61112	5		7	18		62	15,5		19,6
CD20	33061	6	7	9	40	84	215	22,2	29,0	29,3
CD277	57660	1		3	2		15	2,53		14,8
CD222	274355	4		13	7		22	1,77		7,19
CD196	42478		1	1		2	16		3,21	3,21
CD197	42857	1		1	13		7	2,91		2,91
CD180	74165		2	9		6	29		4,54	30,3
CD44	81520	7	4	8	37	22	110	9,43	6,87	10,6
CD48	27665	2		5	10		67	9,47		24,7
CD81	25792	4	4	4	6	15	21	28,4	28,4	28,4
CD82	29608	2	2	6	8	2	108	9,36	9,36	28,1
CD9	25399	2	1	3	4	4	13	15,4	4,39	20,6
CD99	18830	1	1		3	4		7,57	7,57	
CD55	41382			6			18			13,1
CD35	223642	1	19	28	4	49	103	0,589	12,2	18,0
CD21	112897	4		3	13		3	5,13		3,87
CD184	39729	2			4			8,52		
CD39	57948	2		8	3		30	4,71		20,2
CD74	33498	1	1	1	4	2	5	4,73	4,73	4,73
CD49B	129280	13			37			16,9		
CD49D	114884			12			40			16,2
CD49F	126618	13			27			17,0		

CD41	113359	32		12	335		20	40,6		15,1
CD11A	128754	9		23	16		157	12,0		26,4
CD11B	127161	11		28	26		139	12,2		32,6
CD29	88397	13		16	71		70	19,5		23,6
CD18	84764	9	6	31	20	10	106	11,6	8,06	53,2
CD161	87040	30		1	203		2	58,2		1,90
CD54	57807	3	2	4	9	4	7	7,52	5,64	9,59
CD102	30636	2		3	8		11	9,09		9,09
CD50	59521	6		10	87		93	15,7		19,0
CD225	13947	1			3			12,8		
CD321	32565	6			16			19,4		
CD37	31687	4	4	6	28	8	90	19,2	19,2	29,2
CD85J	70802			2			7			4,62
CD32	34026	2		3	9		19	9,68		8,71
CD107A	44865	3	1	1	12	3	8	8,63	3,84	3,84
CD31	82518	16	4	19	92	10	132	30,1	7,99	34,7
CD36	53036	8		1	37		5	21,2		3,60
CD42C	21701	3			14			15,5		
CD42A	19028	4			5			27,1		
CD232	175727			12			41			8,74
CD361	48650	2		2	2		9	5,80		5,80
CD62P	90815	6			19			7,83		
CD45	147238	45	42	60	525	182	1735	35,4	29,8	45,2
CD148	145924	17		32	57		202	19,0		33,8
CD352	37328	1		1	1		2	2,41		6,93
CD298	31496	1	1	4	8	4	16	5,02	5,02	12,5
CD6	71782	3		5	5		19	6,29		12,0
CD1C	37636			2			9			5,41
CD4	51093	1		1	1		2	2,84		2,84
CD5	54559	6		11	24		75	16,6		29,9
CD71	84856			9			20			16,2

Annexe 2

Publication dans la revue Journal of Proteome Research

Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis.

L. Miguët, G. Béchade, L. Fornecker, E. Zink, C. Felden, C. Gervais, R. Herbrecht, A. Van Dorselaer, L. Mauvieux, S. Sanglier-Cianferani.

J. Proteome Res. 8, 3346–3354 (2009)

Proteomic Analysis of Malignant B-Cell Derived Microparticles Reveals CD148 as a Potentially Useful Antigenic Biomarker for Mantle Cell Lymphoma Diagnosis

Laurent Miguet,^{†,‡,§,#} Guillaume Béchade,^{†,#} Luc Fornecker,^{‡,§} Estelle Zink,^{‡,§} Claire Felden,[‡] Carine Gervais,[§] Raoul Herbrecht,[§] Alain van Dorsselaer,^{*,†} Laurent Mauvieux,^{‡,§} and Sarah Sanglier-Cianferani^{*,†}

Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DSA, UDS, CNRS, UMR7178, ECPM 25 rue Becquerel, 67087 Strasbourg, France, Laboratoire d'Hématologie cellulaire EA 3948, Institut d'Hématologie, Strasbourg, France, and Pôle d'Onco-Hématologie - Hôpitaux Universitaires de Strasbourg, France

Received December 22, 2008

The diagnosis of mature B-cell neoplasms (MBCN) remains difficult in a number of cases, especially leukemic phases of non-Hodgkin lymphoma, for which discriminating criteria or biomarker are often lacking. To identify new surface biomarkers, we developed an original proteomic approach based on mass spectrometry analysis of plasma membrane microparticles derived from chronic B-cell lymphoproliferations of single patients: chronic lymphocytic leukemia (CLL), small cell lymphoma (SLL) and mantle cell lymphoma (MCL). A straightforward selection process for proteomic-based candidate biomarker identification was further constructed in order to propose potentially useful and relevant biomarkers. Comparison of the lists of the proteins identified in each pathology combined to highly stringent MS validation criteria for protein identification allowed to propose CD148, a membrane receptor with phosphatase activity, as a discriminating biomarker candidate. Flow cytometry analyses, performed on 158 patients and 30 controls, showed that an anti-CD148 antibody stained significantly higher MCL than CLL and SLL circulating cells ($p < 0.0001$), which validates CD148 overexpression in MCL. Our results indicate that a medium or high CD148 expression level may exclude the diagnosis of CLL and high CD148 expression levels (CD148 MFI equal or superior to 2 times the value obtained with CLL/SLL) allows MCL diagnosis to be suspected with 91% specificity (versus CLL and SLL) and 78% sensitivity. This study is one of the first where proteomic strategies allowed to identify a potentially useful biomarker.

Keywords: mass spectrometry • proteomics • plasma membrane microparticles • chronic B-cell malignancies • CD148 • mantle cell lymphoma • biomarker

Introduction

Mature B-cell neoplasms (MBCN) include several entities delimited within the WHO classification.¹ Facing a B-cell hyperlymphocytosis, morphological examination is the first step for diagnosis, which allows to distinguish well-defined entities such as chronic lymphocytic leukemia (CLL), small cell lymphoma (SLL) from hairy cell leukemia or splenic lymphoma with villous lymphocytes. However, due to their marked heterogeneity, many cases remain cytologically ambiguous or unclassified in overlapping entities. Immunophenotype classification is of great help for MBCN diagnosis² and investigation

of minimal residual disease,³ but remains nondiscriminating for some leukemic phases of non-Hodgkin lymphoma (NHL).

Cytogenetic abnormalities can be specific for some MBCN or define prognosis factors, mainly for CLL. New prognosis biomarkers, including CD38, Zap70 expression or mutation status of immunoglobulin heavy locus⁴ are extensively studied in this pathology, but are limited to CLL. However other lymphoma still raise routinely complicated diagnostic problems, for instance, marginal zone lymphoma, lymphoplasmacytic lymphoma, and mantle cell lymphoma (MCL).^{5–8} MCL, the most aggressive form of MBCN, can be particularly difficult to diagnose with a very variable cytology and ambiguous phenotypes. Moreover, if the translocation t(11;14) leading to a cyclin D1 hyperexpression is a hallmark of the disease, it is not MCL specific and is observed in other MBCN.^{8,9}

Despite the ongoing diagnostic advances, the management of NHL patients remains complex, with overlapping diseases and markedly disparate presentations and clinical outcomes. For all these reasons, there is a need for new biomarkers that

* Corresponding authors: Alain van Dorsselaer and Sarah Sanglier-Cianferani, Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DSA, UDS, CNRS, UMR7178, ECPM - 25, rue Becquerel, 67087 Strasbourg - France, email: vandors@chimie.u-strasbg.fr and sarah.sanglier@chimie.u-strasbg.fr, Tel: 00 33 (0)3 90 24 27 83 Fax: 00 33 (0)3 90 24 27 81.

[†] Laboratoire de Spectrométrie de Masse Bio-Organique.

[‡] Institut d'Hématologie.

[§] Pôle d'Onco-Hématologie - Hôpitaux Universitaires de Strasbourg.

[#] These authors equally contributed to this work.

could allow a more precise diagnostic/classification, and that could also represent therapeutic targets. The most valuable biomarkers would be surface antigens. So far, the search for new biomarkers has mostly been driven by the availability of monoclonal antibodies.¹⁰ Development of proteomics and mass spectrometry techniques has allowed the identification and characterization of hundreds of proteins constituting in a single sample.^{11–16} These proteomic techniques would therefore be of great interest for proliferative lymphoid cell antigen discovery. As surface biomarkers would be the most appropriate, plasma membrane proteomics is attractive. However, isolation of plasma membrane proteins is difficult because of a laborious, often imperfect purification. Moreover, the classical two-dimensional gel analysis approach was shown to be inappropriate for separation of hydrophobic proteins.^{17,18} To circumvent these drawbacks, we previously developed an original method to optimize the analysis of membrane proteins, using secreted plasma membrane microparticles (MPs) and 1D-gel separation.¹⁹ MPs are physiologically secreted in various situations and can be induced by actinomycin D or phytohemagglutinin. This specific methodology allowed the identification of several hundred proteins in the lymphocytic CEM cell line, even if samples were available in relatively small amounts. These proteins included numerous plasma membrane or plasma membrane associated proteins, CD antigens, annexins and signal transduction associated proteins.^{19,20}

These preliminary results prompted us to look for surface antigens as potential new biomarkers differentially expressed in various MBCN. In the present work, MPs from hyperlymphocytic patients suffering from CLL, SLL and MCL were generated and their protein contents were investigated with the previously developed proteomic approach.

A straightforward selection process based on membrane localization of the biomarker and highly stringent MS validation criteria for protein identifications²¹ was developed in order to propose potentially useful candidate biomarkers. Finally, a limited set of “realistic” candidate biomarkers for MCL was delivered and further submitted to quantitative immunological validation. Among them, we focus in the present study on CD148, a protein differentially identified in MCL by the productive but time-consuming proteomic approach. A differential expression of this protein was confirmed in a limited series of well-documented patients by flow cytometry analyses, suggesting that this two-step strategy is adapted and can be extended to different hematological malignancies.

Experimental Section

Patients. Fresh cells were isolated from blood samples from patients with chronic lymphocytic leukemia and small cell lymphoma (CLL $n = 93$, SLL, $n = 24$), mantle cell lymphomas (MCL, $n = 41$) and normal controls ($n = 30$). Initially, patients were included in a study approved by local ethic committee (CCPPRB Alsace, #0279, “Programme de recherche hospitalier des Hôpitaux Universitaires de Strasbourg” #2674). All patients were studied using morphology, immunophenotype and cytogenetics in the clinical hematology laboratory of Haute-pierre Hospital (Strasbourg, France). Cytological data were reviewed by at least two different biologists, as well as cytogenetics and molecular data. Diagnosis were assigned following World Health Organization (WHO) classification.¹ All CLL and SLL samples were cytologically univocal, with a Matutes score equal or superior to 4. The diagnosis of MCL was based on cell morphology, immunophenotype as well as the demonstration

of cyclin D1 expression and/or t(11;14) translocation, following WHO recommendations.¹ No follicular lymphoma nor prolymphocytic leukemia were included in this group. Thirty samples used for normal controls were randomly and anonymously selected within remaining blood samples after routine analyses for nonhematological disease.

Isolation of Membrane MPs from MBCN. An average of 200×10^6 cells were isolated from 3 mL of peripheral blood samples of well-documented hyperlymphocytic patients: one CLL, one SLL and one MCL. Ficoll isolated cells were treated with Actinomycin D ($0.5 \mu\text{g}/\text{mL}$, Sigma) for 18 h in serum-free X-VIVO 15 medium (Cambrex, East Rutherford, NJ). A supernatant was obtained by centrifugation at $750g$ for 15 min and then at $1500g$ for 5 min to remove cells and large fragments, respectively. The latter supernatant was recentrifuged for 45 min at $15\ 000g$. The pelleted MPs were resuspended in 15 mL of Hanks buffered saline solution (HBSS) and centrifuged again for 45 min at $15\ 000g$. The pellet was recovered in 1 mL of HBSS and stored at $-80\ ^\circ\text{C}$ until proteomic analysis.

Proteomic Studies. 1. 1D-SDS-PAGE. Crude MPs pellet were resuspended in $80\ \mu\text{L}$ of SDS-PAGE buffer, before being loaded onto a 8–15% acrylamide SDS-PAGE gel ($20 \times 20\ \text{cm}$). After 5 h of migration at 30 mA, the gel was stained with R250 coomassie blue (Sigma). Each electrophoretic lane was cut systematically at 2 mm intervals and the gel bands were transferred into a 96 well plate.

2. In-Gel Digestion. In-gel digestion was performed with an automated protein digestion system, a MassPrep Station (Waters, Manchester, U.K.). The gel plugs were washed twice with $50\ \mu\text{L}$ of 25 mM ammonium hydrogen carbonate (NH_4HCO_3) and $50\ \mu\text{L}$ of acetonitrile. The cysteine residues were reduced by addition of $50\ \mu\text{L}$ of 10 mM dithiothreitol at $57\ ^\circ\text{C}$ and alkylated by addition of $50\ \mu\text{L}$ of 55 mM iodoacetamide. After dehydration with acetonitrile, the proteins were cleaved in-gel with a $12.5\ \text{ng}/\mu\text{L}$ solution of modified porcine trypsin (Promega, Madison, WI) in 25 mM NH_4HCO_3 ($\sim 30\ \mu\text{L}$). The digestion was performed overnight at room temperature. Tryptic peptides were extracted twice: the first time with 60% acetonitrile in 5% formic acid for 1 h, and the second time with a 100% acetonitrile solution until the gel pieces were dehydrated. The collected extracts were pooled to a final volume of $60\ \mu\text{L}$. Excess of acetonitrile was evaporated at $37\ ^\circ\text{C}$ before analysis.

3. NanoLC-MS/MS Analysis. Nanocapillary liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) analyses were performed using a CapLC system (Waters, Manchester, U.K.), coupled to a hybrid Electrospray Quadrupole Time-Of-Flight mass spectrometer (Q-TOF II, Waters, Manchester, U.K.). A total of $6.4\ \mu\text{L}$ of each sample was loaded onto a precolumn, before chromatographic separation on a C_{18} column (LC packings C_{18} , $75\ \mu\text{m}$ i.d., 150 mm length). Separation was performed at 200 nL/min flow rate using a 35 min gradient from 5 to 45% B (eluent A, 0.1% (v/v) formic acid in water and eluent B, 0.1% (v/v) formic acid in acetonitrile). The ESI capillary voltage was set to 3500 V and cone voltage to 35 V. Mass calibration of the TOF was achieved using phosphoric acid (H_3PO_4) on the m/z 50–2000 range.²² The MS survey scan was acquired on the range m/z 300–1500 with a scan time of 1 s and an interscan time of 0.1 s. Automatic switching between MS and MS/MS modes was used. The 3 most intense ions (intensity threshold of 10 counts) were selected on each MS spectrum for further isolation and CID fragmentation. Fragmentation was performed using argon as collision gas and with

a collision energy profile optimized for various mass ranges of precursor ions. The scan range for MS/MS acquisition was from m/z 50 to 2000 with a scan time of 1 s and an interscan time of 0.1 s.

Data acquisition was piloted by MassLynx software V4.0 SP1. Mass data collected during a nanoLC-MS/MS analysis were automatically processed and converted into a .pkl using MassLynx software V4.0 SP1. Normal background subtraction type was used for both MS and MS/MS with 5% threshold, a polynomial correction of order 5 and deisotoping. Peak list files were submitted to Mascot search engine (Matrix Science, London, U.K.) installed in its 2.2.0 version on a local server. Searches were performed against a composite target-decoy database including both forward (target) and reversed (decoy) Swiss-Prot protein sequences of Human (9 July 2008, 39 862 total entries). Target-decoy database was generated using in-house Perl-CGI scripts. Searches were performed with a tolerance on mass measurement of 0.25 Da for both precursor and fragment ions. Carbamidomethylation of cysteine residues, oxidation of methionine residues and acetylation of protein N-term residues were searched as variable modifications. Up to one missed cleavage was allowed.

Scaffold V2.00.03 (Proteome Science, Portland, OR) was used for identification validation and false positive rate estimation for protein identification. For this, specific Mascot score cutoffs were empirically determined for each sample analysis to maximize the number of proteins identified, keeping the false positive rate <1% (number of decoy database assignments divided by the total number of protein identifications in the target-decoy data set). For each sample, peptides were filtered out according to the cutoff set for proteins hits with 2 or more peptides (Mascot Ion Score - Mascot Identity Score > -20, -32, and -33 for CLL, SSL and MCL samples, respectively). In case of single peptide hit, a different and more stringent criterion was used (Mascot Ion Score - Mascot Identity Score > -6, +9 and +3 for CLL, SSL and MCL samples, respectively). Further data manipulations were performed with Microsoft Excel to finally obtain a catalogue of proteins for each sample with an estimated false positive rate for protein identification below 1%. Subcellular localization and classification of the proteins was assigned using the information extractor Gofetcher (Mississippi Computational Biology Consortium, Hattiesburg, MS, <http://mcbc.usm.edu/gofetcher>) searching the GO EBI database (<http://ebi.ac.uk>, Cambridge, U.K.).

Flow Cytometry Immunophenotypic Studies. All studies were performed using a FACScalibur flow cytometer (Becton Dickinson Bioscience, Franklin Lakes, NJ). Calibrite (Becton Dickinson Bioscience) beads were used to monitor the instrument settings for both the intensity and color compensation. Acquisition and analysis were performed using CellQuest software program (Becton Dickinson Bioscience).

1. Quantitative Indirect Immunofluorescence CD148 Antigen Binding Assay. Quantitative analysis of indirect immunofluorescence staining was performed by using the QIFIKIT purchased from DakoCytomation. QIFIKIT allows expression of fluorescence intensity in antibody-binding capacity (ABC) which represents the number of primary monoclonal antibody per cell. Briefly, 5×10^5 cells were incubated for 30 min at 4 °C with CD148 nonconjugated primary mouse monoclonal antibody (clone 143-41, purchased from R&D Systems). After washing two times with phosphate buffer saline (PBS)/20% fetal calf serum (FCS), cells and beads (setup and calibration beads) were incubated for 45 min at 4 °C with FITC-conjugated anti-

mouse secondary antibody provided with the QIFIKIT. After washing 3 times with PBS/20% FCS, cells and beads were resuspended into 500 μ L of PBS. The lymphocyte populations were gated using side and forward scatter dot plots. Calibration beads were used for the construction of calibration curve (MFI versus ABC). For cell sample, ABC was determined by interpolation of its MFI on the calibration curve. Results were expressed into specific antibody-binding capacity (SABC), which was obtained after subtracting background antibody equivalent (BAE) due to background fluorescence (SABC = ABC - BAE).

Indirect quantitative assay was compared to the direct immunofluorescence staining that was realized as follows.

2. Direct Immunofluorescence Staining. Whole blood was stained with the following four-color combination of monoclonal antibodies CD45/CD19/CD3/CD148 directly conjugated with peridin chlorophyll protein (PerCP), allophycocyanin (APC), fluorescein isothiocyanate (FITC) and phycoerythrin (PE), respectively. Appropriate directly conjugated isotypic antibodies were used as nonspecific immunostaining controls in all experiments. CD45 (clone 2D1), CD19 (clone SJ25C1), CD3 (clone SK7) and APC-conjugated isotypic control (clone X40) were purchased from Becton Dickinson Bioscience, FITC-conjugated isotypic control (clone 679.1Mc7) from Beckman Coulter, CD148 (clone 143-41) from Santa Cruz Biotechnology and R&D Systems and PE-conjugated isotypic control from Monosan. A total of 5×10^5 cells were used per test and the appropriate saturating amounts of labeled monoclonal antibodies were added. Cell suspensions were incubated for 30 min at room temperature and were subsequently lysed. After washing with PBS, the obtained leukocytes were resuspended in PBS. The lymphocyte area was gated in a CD45 versus side scatter dot plot and 5000 CD19+ cells were acquired for each sample. The values of mean fluorescence intensity (MFI) were obtained after subtracting the autofluorescence values.

As a high correlation ($r^2 > 0.992$) between quantitative indirect SABC measurements and direct fluorescence MFI was observed, direct MFI was used for CD148 expression analysis using flow cytometry.

Results

Proteomic-Based Approach to Propose Potentially Relevant MCL Candidate Biomarkers. The previously published strategy for MP proteomic analysis^{19,20} was applied to three different lymphocytic malignancies: one CLL, one SLL and one MCL patient for which diagnosis was assessed as described in the Experimental Section. Briefly, for each pathology, MP proteins were first separated on a 1D-SDS-PAGE gel followed by systematic band cutting and nanoLC-MS/MS analysis.

A straightforward selection process for proteomic-based biomarker identification was further developed in order to propose potentially useful and relevant protein candidates (Figure 1). The first step of our biomarker identification workflow consists of the generation of lists of proteins detected in each sample (i.e., for each pathology) with classical MS validation criteria (see Experimental Section). With these criteria, between 300 and 400 unique proteins were identified in each pathology (296 for MCL, 371 for CLL and 373 for SLL), leading to 579 nonredundant protein identifications when combining results obtained for the three pathologies. Among the 579 unique proteins identified with classical MS validation criteria, roughly 30% were common to all pathologies (300 were detected in at least two of the three types of malignancies).

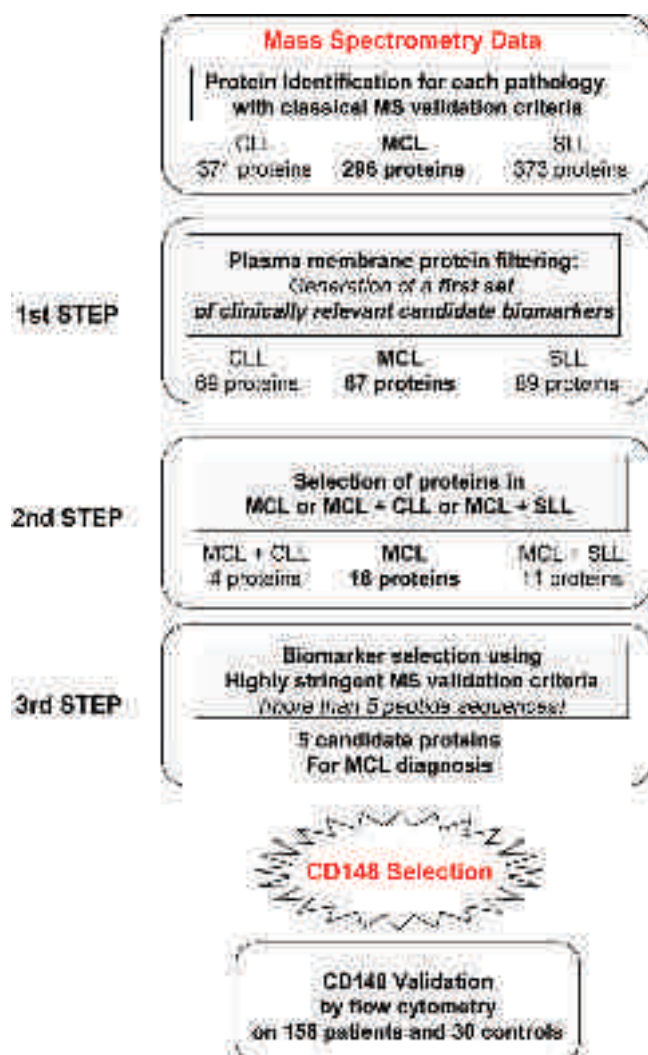


Figure 1. Description of the proteomic based strategy for the proposition of CD148 as a potentially useful candidate biomarker.

To reduce the number of proteins to those candidates that might be potentially useful, we first used a selection criterion based on the biological features of potential candidate proteins. We focused on plasma membrane proteins presenting extracellular domains which would be most appropriate for further flow cytometry analysis. This first selection step enabled to reduce the initial set of proteins to 67 for MCL, 69 for CLL and 89 for SLL (see Supporting Information S1 and S2). The next filter applied to search for relevant MCL biomarkers consisted of selecting proteins either uniquely identified in MCL or in MCL and another pathology (either CLL or SLL): all membrane proteins concomitantly present in the three pathologies were thus eliminated. This second step in the biomarker selection process allowed us to reduce the number of protein candidates from 67 to 33 for MCL (see Supporting Information S3). To provide the final limited set of “realistic” candidate biomarkers to be further validated in immunological studies, we applied highly stringent MS validation criteria in order to focus on proteins identified with a very high level of confidence in MCL, that is, with a large number of peptide sequences (more than 5 peptides) and high sequence coverage (see Supporting Information S3, proteins in bold boxes). Indeed, it has been shown that the amount of peptides used for the identification

of a protein is directly linked to the concentration of this protein in the sample.^{23,24} Among the 33 surface plasma membrane proteins, most of them, including CD20 (P11836), CD37 (P11049) or CD82 (P27701), were differentially identified but with less than 5 peptide sequences and did thus not fulfill our stringent requirements. This final selection step evidenced 5 candidate proteins (see Supporting Information S3, proteins in bold boxes). Among those, HLA (P05534) was not selected due to its lack of biological relevance. Chloride channel CLIC1 (O00299), which serves in signal transduction and chloride transport, was not retained because of its nondifferential expression in SLL versus MCL (5 peptides in SLL versus 6 in MCL). As cytoskeleton remodelling partly induces MP generation, Transgelin-2 (P37802) and Plexin-A1 (Q9UIW2), two proteins involved in the latter process, were also not considered as first-line candidate biomarkers. Thus, CD148 (Q12913), a plasma membrane protein receptor with tyrosine phosphatase activity, was identified as the best “realistic” candidate biomarker that fulfills the stringent MS validation criteria mentioned above. Indeed, this protein was unambiguously identified in the MCL sample in 3 consecutive 1D gel bands with 10 different peptides (Figure 2 and Supporting Information S4), while in SLL only 4 unique peptides were observed (no CD148 tryptic peptide was detected in the CLL sample). The strong confidence in CD148 identification and its differential detection in MCL versus SLL/CLL suggest that CD148 is more expressed in MCL compared to SLL and CLL lymphoma samples.

Validation of CD148 in 158 Patients and 30 Controls. We decided to look at CD148 expression in controls and patients with circulating pathologic B-cells of CLL, SLL and MCL by flow cytometry (see Experimental Section). Representative experiments are shown in Figure 3. CD148 MFI of 30 control cells was very homogeneous (mean = 168, SD = 31, Table 1 and Figure 4). In a total of 93 CLL (mean = 189, SD = 78) and 24 SLL cases (mean = 209, SD = 92), MFI was considered not significantly different from controls using nonparametric tests (controls versus CLL, $p = 0.59$; controls versus SLL, $p = 0.07$; and CLL versus SLL $p = 0.20$ Mann–Whitney Test). In these three groups (CLL, SLL and controls), it can thus be considered that CD148 is expressed at comparable levels. The results were strikingly different for the MCL group ($n = 41$), in which CD148 MFI was 613 (SD = 329) which is significantly different from both CLL and SLL groups ($p < 0.0001$, Mann–Whitney Test), and from controls ($p < 0.0001$, Mann–Whitney Test) (Figure 4). These results clearly support the differential expression of CD148, as suggested from the proteomic study. CD148 expression appeared to be highly elevated in MCL while not different from controls in CLL or SLL. Moreover, the realization of a Receiver Operator Characteristic test (ROC curve) between MCL and CLL/SLL shows that with a CD148 MFI equal or superior to 2 times the mean value obtained with CLL/SLL, MCL diagnosis can be suspected with 91% of specificity (versus CLL and SLL) and a sensitivity of 78% (Figure 5 and Supporting Information S5). This observation can be a rationale for its use for MCL diagnosis.

Discussion

In this work, we describe the selection of CD148 as a potential candidate biomarker for clinical use proposed from a proteomic approach, and a preliminary validation of its expression in different B-cell lymphoproliferations using classical immunological tools.

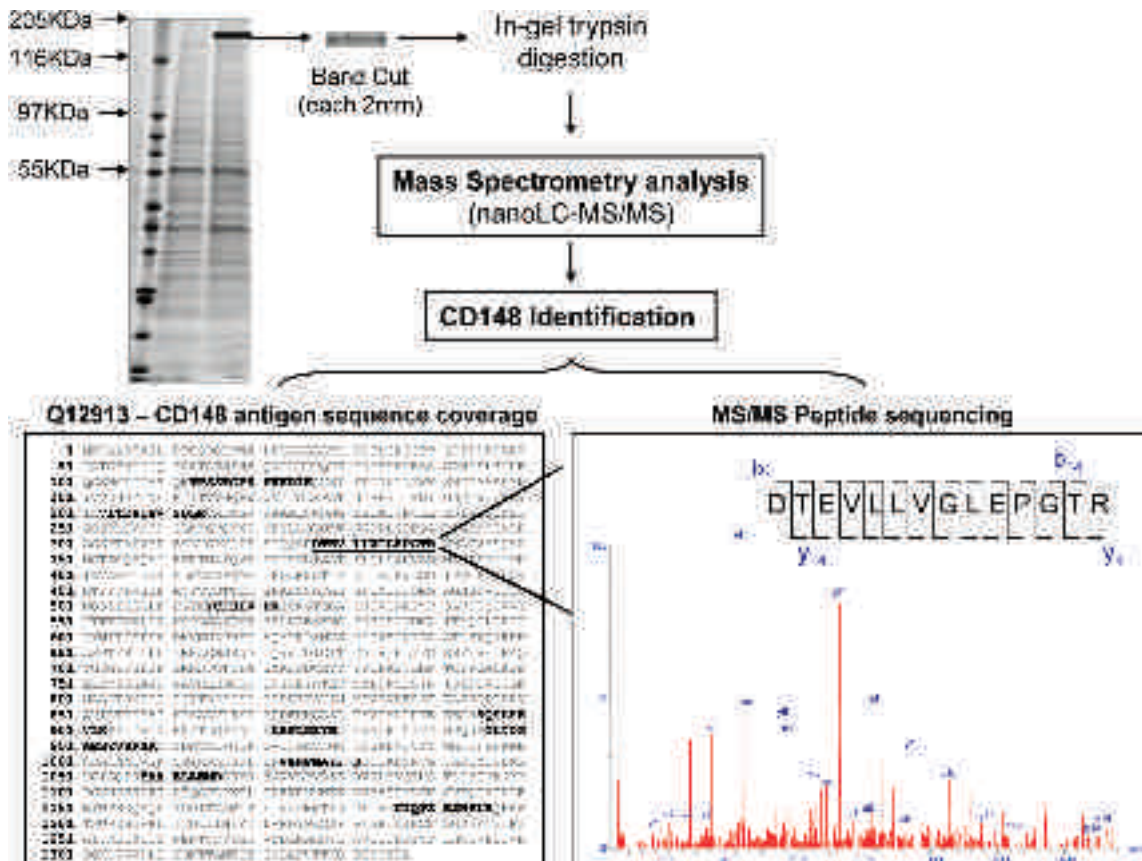


Figure 2. Identification of CD148 by mass spectrometry. Plasma membrane MPs are generated from pure cell populations. Proteins are extracted and fractionated using 1D SDS-PAGE electrophoresis. Gels bands are systematically cut into 2 mm pieces, followed by trypsin digestion and nanoLC-MS/MS analysis. In bold are shown the 10 different peptides sequenced by MS/MS during mass spectrometry analyses and used for CD148 identification. The fragmentation spectrum of the underlined peptide (measured monoisotopic mass of 1497.8 Da) is presented in the right panel of the figure. The fragmentation spectrum shows predominantly y and b ions (as indicated on the spectrum) allowing the determination of the DTEVLLVGLPEGTR amino acid sequence, permitting a clear identification of the CD148.

In the race for disease-specific clinically useful biomarker discovery, the use of proteomics has huge potential, although, to date, there are only a limited number of successful exploitations of proteomic experiments.²⁵ Recently, Lescuyer et al. have addressed this point of heated debate in a review paper.²⁶ These authors explain that several reasons can account for the relative failure of proteomic approaches to deliver new potential biomarkers. These include: no optimal definition of the clinical question, complexity of the different biological fluids chosen, inadequate proteomic strategies developed and underestimation of the necessity to validate proteomic-proposed biomarkers. However, a careful design of the proteomic experiment in the discovery phase can lead to successful identification of biomarkers for clinical purposes. Another crucial point is the fact that the proteomic-based discovery phase needs always to be followed by a separate clinical validation phase involving quantitative methodologies, often by immunological methods (mostly ELISA assays). Altogether, these points can explain the relative low yield of successful proteomic studies in the biomarker discovery field.

The study presented here relies on a well-defined question: how can we discover helpful biomarkers in MBCN, and in our case MCL, that could be easily detected using routine analysis in hospital, that is, membrane antigens by flow cytometry? To achieve this goal it was decided to work on MPs, which were shown to be an interesting starting material, affording the

combination of low sample consumption and enabling the identification of hundreds of plasma surface membrane proteins.^{19,20} At the proteomic level, the classical approach used for biomarker discovery is to compare protein/peptide profiles of normal versus diseased subjects.^{27,28} This approach was not applicable in our case, as sufficient amounts of MPs cannot be easily generated from healthy subjects (about 200×10^6 B-cells required). However, it can be applied to a wide range of lymphoma showing an accumulation of lymphocytes in the peripheral blood. Even if the limitation of our approach is to have enough circulating MPs in the concerned malignancy, other studies have already shown platelet MPs¹¹ and small vesicles²⁹ proteomic analyses in various biological fluids.³⁰ Moreover, which population would be explored: memory B-cells, naive B-cells? Also MBCN diagnosis issue is not to distinguish ill versus healthy patients but to precisely classify MBCN subtypes (CLL versus SLL versus MCL for instance). We therefore based our work on a differential proteomic study between CLL, SLL and MCL pathologies.

A proteomic experiment based on 1D-gel separation of proteins, nanoLC-MS/MS analyses of tryptic peptides combined with highly stringent MS validation criteria for protein identifications was subsequently designed. Although absolute quantification is not possible with this approach, comparison of MPs subproteomes originating from different pathologies indicate significant differences in protein expression levels.^{31,32}

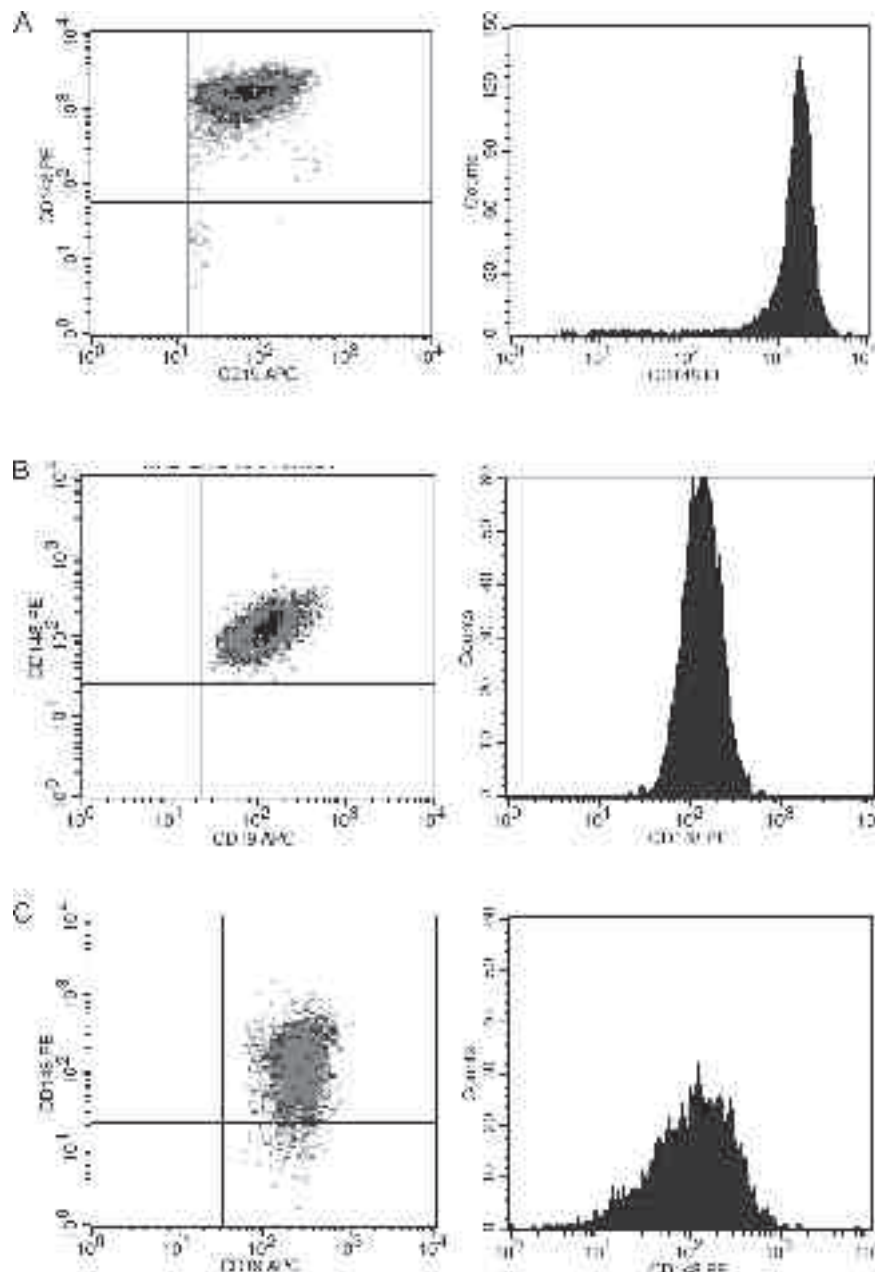


Figure 3. Anti-CD148 staining of three different representative samples: (A) mantle cell lymphoma (MCL), (B) chronic lymphocytic leukemia (CLL), and (C) control samples, showing a high CD148 expression in MCL samples. Left panel: CD19/CD148 dot blot (anti-CD19 allophycocyanin conjugated (CD19 APC) in X-axis, CD148 phycoerythrin conjugated in Y-axis). Right panel: Anti-CD148 phycoerythrin staining histogram.

Table 1. Anti-CD148 Staining Results and Cyclin D1 Expression of Chronic Lymphocytic Leukemias (CLL), Small Cell Lymphomas (SLL), Mantle Cell Lymphomas (MCL), and Control Groups

	controls	CLL	SLL	MCL
<i>n</i>	30	93	24	41
Cyclin D 1	Negative	Negative	Negative	Positive
CD148 MFI				
Mean	168	189	209	613
Min	104	80	74	70
Max	233	498	350	1584
SD	31	78	93	329

Cell surface CD148 quantification, which was achieved in our case by immunological methods, is thus necessary. The pro-

teomic selection process we developed, based on high numbers of peptides and high protein sequence coverages, allowed to focus not only on proteins differentially identified with a high confidence, but also proteins showing significant differences in their expression level.^{23,33} Even if the choice of CD148 can appear to be arbitrary as there are no quantitative criteria in the selection process, our proteomic data show that CD148 is identified with 10 different peptides in MCL, 4 in SLL and none in CLL, which strongly suggests that CD148 is overexpressed in MCL compared to the two other lymphomas. CD148 seemed thus to be the best protein candidate to be proposed for clinical validation. The presented strategy shows the ability of proteomics to propose strong protein candidates to be further confirmed as potential biomarkers. The quantification, the

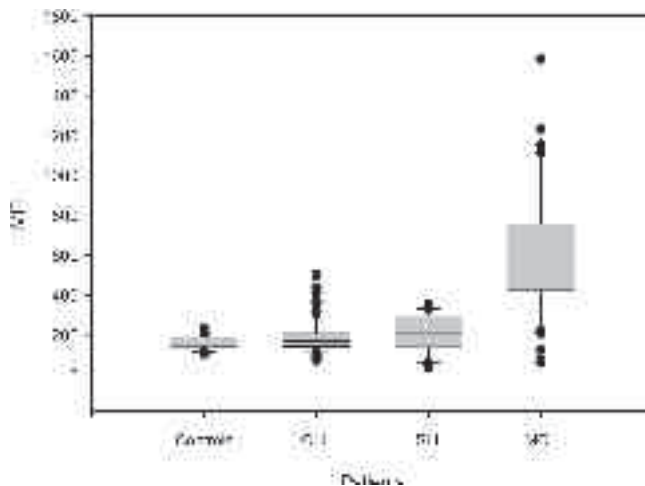


Figure 4. Comparison of CD148 median fluorescence intensity in controls ($n = 30$), chronic lymphocytic leukemias ($n = 93$), small cell lymphoma ($n = 24$) and mantle cell lymphomas (MCL, $n = 41$).

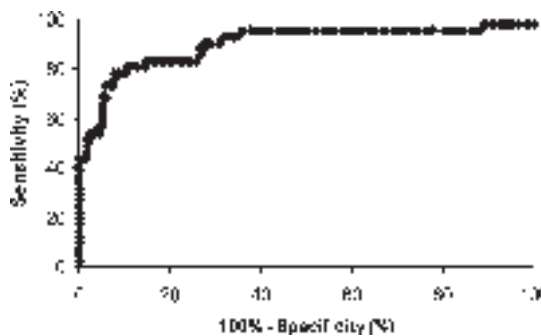


Figure 5. ROC Curve of MCL versus CLL/SLL group. A high CD148 (2-fold mean expression of normal B-cells) strongly indicates MCL diagnosis with 91% of specificity and a sensitivity of 78%.

biological and statistical relevance of CD148 as potential MCL biomarker, are then given by flow cytometry analyses on a larger set of patients.

CD148 is a plasma membrane receptor with phosphatase activity (type 3) also named DEP-1, PTPRJ or HPTPeta. This protein is composed of an extracellular domain containing 8 fibronectin type II-like domains, a transmembrane region and a single intracellular phosphatase domain. It was identified in human F-36 erythroleukemic³⁴ and HeLa cell lines, in which its expression was enhanced with increasing cell density.³⁵ It is expressed in various tissues, and particularly in hematopoietic cells. In peripheral blood cells, granulocytes display the highest intensity of anti-CD148 staining using flow cytometry, while lymphocytes are more weakly stained, CD19+ cells being slightly more stained than T cells.³⁶ The unique study published to date about CD148 expression in MBCN³⁷ described its expression in almost all the different types of the pathologies studied. On the contrary, our flow cytometry experiments revealed that CD148 is overexpressed in the large majority of MCL cases compared to CLL or SLL. Discrepancies between these results can be explained by several reasons: (i) Dong et al. use immunocytochemistry, a considerably less sensitive and nonquantitative technique compared to flow cytometry, and probably not helpful in the context of B-cell malignancies diagnosis described here; (ii) only flow cytometry techniques are able to discriminate variations of CD148 in the range of

what was detected in the present study. Thus systematic CD148 flow cytometry detection in MBCN could be very helpful in order to better diagnose MCL that harbor frequently misleading morphologic and immunophenotypic presentations. As a consequence, “gold standard” karyotype or cyclin D1 exploration, that are not systematically realized, may then be performed more rapidly when a high CD148 expression is detected by flow cytometry analysis. On the contrary, 30% of MCL patients display a CD148 below 380 counts. In these cases, CD148 does not help for MCL diagnosis. Other biomarkers are still needed and the current diagnosis relies on available biological tools. From our results, a high CD148 expression (superior to 2-fold the mean of normal B-cells) may be strongly in favor of MCL diagnosis with 91% of specificity (versus CLL and SLL) and a sensitivity of 78%. Interestingly, in the present study, for two cases where the diagnosis of MCL lymphoma was not evoked using conventional criteria (that do not include karyotype or cyclin D1), a high anti-CD148 staining led to CCND1 expression analysis. Both cases were positive and the diagnosis was reconsidered into MCL.

We have no clear explanation for the high level of expression of CD148 in MCL. *In vitro* activation of T cells induces an increase in CD148 expression, indicating a potential role in signal transduction on lymphocytes confirmed by other studies.^{36,38} However, for the moment, no CD148 ligand has been clearly identified even if a potential ligand may be present in the extracellular matrix.³⁹ Replacement of CD148 phosphatase domain with GFP was not viable in mice due to vascular abnormalities, thus, preventing any analysis in mature B-lymphocytes.⁴⁰ Mice with constitutively deleted transmembrane domain of CD148 were also generated.³⁸ In this model, no obvious defect in B-cell development was observed in bone marrow, but a partial block in peripheral B-cell development was detected. In CD45 (a crucial plasma membrane protein with phosphatase activity expressed in hematopoietic cells) and CD148 double KO mice,³⁸ a severe block in early B-cell development and impaired B-cell antigen receptor (BCR) mediated signaling in relation with impaired tyrosine phosphorylation was observed. Very recently, it was also shown that CD148 may interact with and dephosphorylate p85 and modulate the magnitude of PI3K activity.⁴¹ It is thus likely that CD148 participates in the regulation of signaling in B-cells.

In conclusion, the strategy described in this paper is one of the few successful proteomic studies that have led to the identification of a potentially useful biomarker. Indeed, CD148 may be a useful first-line biomarker for the diagnosis of MBCN using flow cytometry: (i) medium or high CD148 expression (defined by CD148 MFI > MFI of normal B-cells + 2 SD) may exclude the diagnosis for CLL or SLL samples; (ii) a high CD148 (2-fold mean expression of normal B-cells) may strongly indicate MCL diagnosis with 91% specificity and 78% sensitivity. As MCL assessment can still be ambiguous in several cases, an earlier, more precise and accurate diagnosis, using a combination of criteria (such as cyclinD1 activation and CD148 MFI, for example), would imply more aggressive chemo- or immuno-therapies. Further studies are now required to assess the clinical and prognosis strengths of this biomarker. In addition, the approach outlined here may be useful for the identification of potential biomarkers for other pathologies.

Abbreviations: MPs, microparticles; CLL, chronic lymphocytic leukemia; MCL, mantle cell lymphoma; SLL, small B-cell

lymphoma; CCND1, cyclin D1; NHL, non-Hodgkin lymphoma; MBCN, Mature B-cell neoplasms; MFI, mean fluorescence intensity; nanoLC-MS/MS, nanocapillary liquid chromatography-tandem mass spectrometry.

Acknowledgment. L. Miguet gratefully acknowledges the ARC foundation for financial support. G. Béchade was supported by the French Research Ministry and the European Postgraduate School (GRK 532). This study was also supported part by the CNRS, the Medical School of Strasbourg, and grants from the “Programme Hospitalier de Recherche Clinique des Hôpitaux de Strasbourg” # 2674, the ARECOH/ATGC association and the Institut National du Cancer. The FRM (Fondation pour la Recherche Médicale) is acknowledged for financing a mass spectrometer.

Supporting Information Available: S1, table of plasma membrane proteins identified per pathology with information on the total number of peptides involved in identification; S2, information relative to the identifications of the plasma membrane proteins; S3, table of proteins identified in MCL or in MCL and one of the two other pathologies and Biomarker candidate selection using highly stringent MS-validation criteria (more than 5 peptide sequences); S4, information on CD148 identification and MS/MS spectra of peptides involved in the identification; S5, raw data of ROC Curves. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Jaffe, E. S.; L., H. N.; Stein, H.; Vardiman, J. W. , *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*; IARC Press: Lyon, France, 2001; Vol. 1.
- Matutes, E.; Owusu-Ankomah, K.; Morilla, R.; Garcia Marco, J.; et al. The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL. *Leukemia* **1994**, *8*, 1640–1645.
- Rawstron, A. C.; Villamor, N.; Ritgen, M.; Böttcher, S.; et al. International standardized approach for flow cytometric residual disease monitoring in chronic lymphocytic leukaemia. *Leukemia* **2007**, *21*, 956–964.
- Montserrat, E. New prognostic markers in CLL. *Hematology* **2006**, 279–284.
- Ruchlemer, R.; Parry-Jones, N.; Brito-Babapulle, V.; Attolico, I.; et al. B-prolymphocytic leukaemia with t(11;14) revisited: a splenomegalic form of mantle cell lymphoma evolving with leukaemia. *Br. J. Haematol.* **2004**, *125*, 330–336.
- Nelson, B. P.; Variakojis, D.; Peterson, L. C. Leukemic phase of B-cell lymphomas mimicking chronic lymphocytic leukemia and variants at presentation. *Mod. Pathol.* **2002**, *15*, 1111–1120.
- Avet-Loiseau, H.; Garand, R.; Gaillard, F.; Daviet, A.; et al. Detection of t(11;14) using interphase molecular cytogenetics in mantle cell lymphoma and atypical chronic lymphocytic leukemia. *Genes Chromosomes Cancer* **1998**, *23*, 175–182.
- Troussard, X.; Mauvieux, L.; Radford-Weiss, I.; Rack, K.; et al. Genetic analysis of splenic lymphoma with villous lymphocytes: a Groupe Français d'Hématologie Cellulaire (GFHC) study. *Br. J. Haematol.* **1998**, *101*, 712–721.
- Resnitzky, P.; Matutes, E.; Hedges, M.; Morilla, R.; et al. The ultrastructure of mantle cell lymphoma and other B-cell disorders with translocation t(11;14)(q13;q32). *Br. J. Haematol.* **1996**, *94*, 352–361.
- Ek, S.; Hogerkerp, C. M.; Dictor, M.; Ehinger, M.; Borrebaeck, C. A. Mantle cell lymphomas express a distinct genetic signature affecting lymphocyte trafficking and growth regulation as compared with subpopulations of normal human B cells. *Cancer Res.* **2002**, *62*, 4398–4405.
- Garcia, B. A.; Smalley, D. M.; Cho, H.; Shabanowitz, J.; et al. The platelet microparticle proteome. *J. Proteome Res.* **2005**, *4*, 1516–1521.
- Pasini, E. M.; Kirkegaard, M.; Mortensen, P.; Lutz, H. U.; et al. In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood* **2006**, *108*, 791–801.
- Jeong, H. C.; Lee, S. Y.; Lee, E. J.; Jung, K. H.; et al. Proteomic analysis of peripheral T-lymphocytes in patients with asthma. *Chest* **2007**, *132*, 489–496.
- Bartee, E.; McCormack, A.; Fruh, K. Quantitative membrane proteomics reveals new cellular targets of viral immune modulators. *PLoS Pathog.* **2006**, *2*, e107.
- Salonen, J. M.; Valmu, L.; Ronnholm, G.; Kalkkinen, N.; Vihinen, M. Proteome analysis of B-cell maturation. *Proteomics* **2006**, *6*, 5152–5168.
- Barderas, M. G.; Darde, V. M.; Duran, M. C.; Egido, J.; Vivanco, F. Characterization of circulating human monocytes by proteomic analysis. *Methods Mol. Biol.* **2007**, *357*, 319–328.
- Santoni, V.; Kieffer, S.; Desclaux, D.; Masson, F.; Rabilloud, T. Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis* **2000**, *21*, 3329–3344.
- Galeva, N.; Altermann, M. Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins. *Proteomics* **2002**, *2*, 713–722.
- Miguet, L.; Pacaud, K.; Felden, C.; Hugel, B.; et al. Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. *Proteomics* **2006**, *6*, 153–171.
- Miguet, L.; Sanglier, S.; Schaeffer, C.; Potier, N.; et al. Microparticles: a new tool for plasma membrane sub-cellular proteomic. *Subcell Biochem.* **2007**, *43*, 21–34.
- Taylor, C. F.; Paton, N. W.; Lilley, K. S.; Binz, P. A.; et al. The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **2007**, *25*, 887–893.
- De Vos, R. C.; Moco, S.; Lommen, A.; Keurentjes, J. J.; et al. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2007**, *2*, 778–791.
- Paoletti, A. C.; Parmely, T. J.; Tomomori-Sato, C.; Sato, S.; et al. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18928–18933.
- Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; et al. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **2006**, *5*, 2339–2347.
- Hsich, G.; Kenney, K.; Gibbs, C. J.; Lee, K. H.; Harrington, M. G. The 14-3-3 brain protein in cerebrospinal fluid as a marker for transmissible spongiform encephalopathies. *N. Engl. J. Med.* **1996**, *335*, 924–930.
- Lescuyer, P.; Hochstrasser, D.; Rabilloud, T. How shall we use the proteomics toolbox for biomarker discovery. *J. Proteome Res.* **2007**, *6*, 3371–3376.
- Miguet, L.; Bogumil, R.; Decloquement, P.; Herbrecht, R.; et al. Discovery and identification of potential biomarkers in a prospective study of chronic lymphoid malignancies using SELDI-TOF-MS. *J. Proteome Res.* **2006**, *5*, 2258–2269.
- Sanchez, J. C.; Lescuyer, P.; Hochstrasser, D.; Allard, L. Detection of biomarkers of stroke using SELDI-TOF. *Methods Mol. Biol.* **2007**, *357*, 343–350.
- Gonzalez-Begne, M.; Lu, B.; Han, X.; Hagen, F. K.; et al. Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT). *J. Proteome Res.* **2009**, *8*, 1304–1314.
- Gonzales, P. A.; Pisitkun, T.; Hoffert, J. D.; Tchapyjnikov, D.; et al. Large-scale proteomics and phosphoproteomics of urinary exosomes. *J. Am. Soc. Nephrol.* **2009**, *20*, 363–379.
- Liu, H.; Sadygov, R. G.; Yates, J. R. 3rd A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76*, 4193–4201.
- Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **2005**, *4*, 1265–1272.
- de Godoy, L. M.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **2008**, *455*, 1251–1254.
- Honda, H.; Inazawa, J.; Nishida, J.; Yazaki, Y.; Hirai, H. Molecular cloning, characterization, and chromosomal localization of a novel protein-tyrosine phosphatase, HPTP eta. *Blood* **1994**, *84*, 4186–4194.

- (35) Ostman, A.; Yang, Q.; Tonks, N. K. Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase, is enhanced with increasing cell density. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9680–9684.
- (36) de la Fuente-Garcia, M. A.; Nicolas, J. M.; Freed, J. H.; Palou, E.; et al. CD148 is a membrane protein tyrosine phosphatase present in all hematopoietic lineages and is involved in signal transduction on lymphocytes. *Blood* **1998**, *91*, 2800–2809.
- (37) Dong, H. Y.; Shahsafaie, A.; Dorfman, D. M. CD148 and CD27 are expressed in B cell lymphomas derived from both memory and naive B cells. *Leuk. Lymphoma* **2002**, *43*, 1855–1858.
- (38) Zhu, J. W.; Brdicka, T.; Katsumoto, T. R.; Lin, J.; Weiss, A. Structurally distinct phosphatases CD45 and CD148 both regulate B cell and macrophage immunoreceptor signaling. *Immunity* **2008**, *28*, 183–196.
- (39) Sorby, M.; Sandstrom, J.; Ostman, A. An extracellular ligand increases the specific activity of the receptor-like protein tyrosine phosphatase DEP-1. *Oncogene* **2001**, *20*, 5219–5224.
- (40) Takahashi, T.; Takahashi, K.; St John, P. L.; Fleming, P. A.; et al. A mutant receptor tyrosine phosphatase, CD148, causes defects in vascular development. *Mol. Cell. Biol.* **2003**, *23*, 1817–1831.
- (41) Tsuboi, N.; Utsunomiya, T.; Roberts, R. L.; Ito, H.; et al. The tyrosine phosphatase CD148 interacts with p85 regulatory subunit of PI 3-kinase. *Biochem. J.* **2008**, *413*, 193–200.

PR801102C

Annexe 3

Publication dans la revue Leukemia

Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas.

L. Miguet, S. Lennon, L. Baseggio, A. Traverse-Glehen, F. Berger, N. Perrusson, M.P. Chenard, A.C. Galois, A. Eischen, C. Mayeur-Rousse, A. Maar, L. Fornecker, R. Herbrecht, P. Felman, A. Van Dorsselaer, C. Carapito, S. Cianferani, L. Mauvieux.

Leukemia. 2013 Aug;27(8):1748-50

REFERENCES

- Chaplin T, Ayton P, Bernard OA, Saha V, Della Valle V, Hillion J *et al.* A novel class of zinc finger/leucine zipper genes identified from the molecular cloning of the t(10;11) translocation in acute leukemia. *Blood* 1995; **85**: 1435–1441.
- Chaplin T, Bernard O, Beverloo HB, Saha V, Hagemeijer A, Berger R *et al.* The t(10;11) translocation in acute myeloid leukemia (M5) consistently fuses the leucine zipper motif of AF10 onto the HRX gene. *Blood* 1995; **86**: 2073–2076.
- Meyer C, Kowarz E, Hofmann J, Renneville A, Zuna J, Trka J *et al.* New insights to the MLL recombinome of acute leukemias. *Leukemia* 2009; **23**: 1490–1499.
- Morerio C, Rapella A, Tassano E, Rosanda C, Panarello C. MLL-MLLT10 fusion gene in pediatric acute megakaryoblastic leukemia. *Leuk Res* 2005; **29**: 1223–1226.
- Beverloo HB, Le Coniat M, Wijsman J, Lillington DM, Bernard O, de Klein A *et al.* Breakpoint heterogeneity in t(10;11) translocation in AML-M4/M5 resulting in fusion of AF10 and MLL is resolved by fluorescent *in situ* hybridization analysis. *Cancer Res* 1995; **55**: 4220–4224.
- DiMartino JF, Ayton PM, Chen EH, Naftzger CC, Young BD, Cleary ML. The AF10 leucine zipper is required for leukemic transformation of myeloid progenitors by MLL-AF10. *Blood* 2002; **99**: 3780–3785.
- Van Limbergen H, Poppe B, Janssens A, De Bock R, De Paepe A, Noens L *et al.* Molecular cytogenetic analysis of 10;11 rearrangements in acute myeloid leukemia. *Leukemia* 2002; **16**: 344–351.
- Chen J, Feng W, Jiang J, Deng Y, Huen MS. Ring finger protein RNF169 antagonises the ubiquitin-dependent signaling cascade at sites of dna damage. *J Biol Chem* 2012; **287**: 27715–27722.
- Poulsen M, Lukas C, Lukas J, Bekker-Jensen S, Mailand N. Human RNF169 is a negative regulator of the ubiquitin-dependent response to DNA double-strand breaks. *J Cell Biol* 2012; **197**: 189–199.

Supplementary Information accompanies the paper on the Leukemia website (<http://www.nature.com/leu>)

Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas

Leukemia (2013) **27**, 1748–1750; doi:10.1038/leu.2013.3

Marginal zone B-cell lymphomas (MZL) are classically characterized by an indolent proliferation of CD5[−], CD23[−], CD43[−], CCND1[−] and CD10[−] B cells that strongly express CD19, CD20, CD22, IgM and IgD, even if CD5⁺ cases have been characterized.¹ The World Health Organization (WHO)² classifies marginal B-cell lymphomas into splenic marginal zone lymphoma (SMZL), nodal MZL and extranodal MZL of mucosa-associated lymphoid tissue (MALT).² Among the lymphomas developed from the spleen, the splenic diffuse red pulp lymphoma with villous lymphocytes (SDRPLs) has been described as provisional entity close to but distinct from SMZL.^{2,3}

For small to intermediate-sized circulating abnormal lymphocytes, MZL diagnosis can be of great difficulty, particularly because there is no immunological-positive marker to distinguish MZL from atypical chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL) and lymphoplasmacytic lymphoma (LPL). Moreover, if translocations involving MALT1 (t(11;18), t(14;18)), BCL10 (t(1;14)) or FOXP1 (t(3;14)) are primarily associated with MALT, and del7q is frequent in SMZL, these cytogenetic markers are not always present. Therefore, there is no specific immunologic marker for splenic or nodal MZL.

To improve the diagnosis of chronic B-cell lymphoproliferative disease (B-LPD) with circulating cells, we have previously developed an innovative mass spectrometry-based proteomic workflow on plasma membrane microparticles derived from B-LPD⁴ and proposed CD148 as a potential marker for MCL diagnosis panel.⁵ Using again the proteomic toolbox, a differential proteomic analysis was applied here to compare the membrane proteomes originating from MPs derived from the following three different B-LPD: LPL, MCL and MZL.

A list of 388 unique membrane proteins was obtained including 59 CD annotated cell-surface proteins (Supplementary Table). Within this list of characterized cell-surface proteins, we focused on CD180, also called Radio-Protective 105 (RP105) or lymphocyte antigen 64, which was considered overexpressed in the MZL sample by our semi-quantitative mass spectrometry analysis (peptide and spectral counting). CD180 is a leucine-rich repeat type 1 membrane protein belonging to the Toll-like receptor (TLR)

family, with significant extracellular homology to the lipopolysaccharide receptor.⁶ Its expression is restricted to antigen presenting cells, such as B cells, monocytes, macrophages and dendritic cells, and anti-CD180 antibodies activate the majority of human and mouse B cells *in vitro*.^{7,8} The literature describes opposite functions for CD180; however, it has been shown that CD180 ligation induces significant B-cell proliferation and Ig production, in both humans and mice.^{7,8}

CD180 expression in B-CLL has been previously described to be different from that in normal CD19⁺ blood cells.⁹ In this paper, we evaluated CD180 expression using flow cytometry in two independent series of B-LPD patients from two French hematology departments, including 60 cases of MZL that were centrally reviewed.

Ninety-six blood samples from patients referred for investigation or follow-up of B-LPD in the Oncology and Hematology Department of the Strasbourg University Hospital were tested: 30 patients with CLL, 16 patients with MCL, 15 patients with LPL and 35 patients with MZL (25 splenic, 10 nodal without splenomegaly), in addition to 25 controls. Blood cells samples were collected after signed informed consent was obtained (Comité de protection des personnes EST-IV, # AC-2008-438) in accordance with the Declaration of Helsinki, and the Ethic Committee of the Faculty of Medicine of Strasbourg approved the study in 2011. Over 93% of the patients were included in the study at the time of diagnosis, and all patients presented a low count of circulating residual normal B cells. All cases of MCL displayed a t(11;14) and/or cyclin D1 hyperexpression. All cases of CLL had a Matutes score of 4 or 5. Only cases of MZL without plasma differentiation were selected in this study. Nine patients with SMZL displayed circulating villous lymphocytes.

As previously described, normal B cells displayed strong staining for CD180. CD27⁺ and CD27[−] cells were equally stained and only a small percentage of CD180[−] B cells were observed in the controls (mean 5.3%, range 0.4–14%, data not shown). In agreement with previous studies⁹ malignant B cells of CLL displayed weaker expression of CD180 ($P < 0.0001$ Mann-Whitney *U*-test) (Figure 1b).

We observed that CD180 staining was also weak in MCL ($P < 0.001$) and LPL ($P < 0.001$) B cells (Figures 1a and b). CD180 mean fluorescence intensities (MFIs) of CLL, MCL and LPL B cells

did not differ significantly ($P > 0.05$, one-way analysis of variance (ANOVA)) and were all significantly lower than the controls ($P < 0.001$, one-way ANOVA and Figure 1b).

By contrast, as hypothesized from the mass spectrometry studies, circulating MZL cells had a high CD180 MFI, which was not different from the controls ($P > 0.05$) but was significantly higher than the CLL, MCL and LPL B cells ($P < 0.001$, non-parametric one-way ANOVA). CD180 staining appeared more heterogeneous in the MZL cells than the control B cells (mean 5865, s.d. = 5643 vs mean 5671, s.d. = 1708). Within the MZL group, CD180 staining of the SMZL and the nodal MZL were not different from the controls ($P = 0.094$ and 0.162 , respectively, Figure 1d), and the SMZL displayed a higher MFI than nodal MZL (4941 vs 8708, $P = 0.0327$ Mann-Whitney U -test).

Similar results were obtained using a second anti-CD180 clone that was used to stain approximately one-third of the patients and control cells (clone MHR73-11, eBiosciences, San Diego, CA, USA; data not shown).

In the same way, the CD180 expression was performed in an independent series of 123 consecutive and prospective peripheral blood samples from patients with B-LPD diagnosed in the Hematology Department of Lyon-Sud University Hospital, according to 2008 WHO recommendations and after integration of all data (cytology, histology, flow cytometry immunophenotyping, conventional and molecular cytogenetics). These cases were classified as CLL ($n = 80$), MZL ($n = 25$), MCL ($n = 10$) and SDRPL ($n = 8$).

The MZL cases presented a strong CD180 expression (median = 537), whereas the MCL and CLL cases showed a very dim or negative CD180 expression (median = 126 and median = 171, respectively) (Table 1, Figures 1a and c). CD180 expression was statistically higher in the eight cases of SDRPL (median = 2 356) than in MZL (median = 537) (Table 1, Figure 1c). The difference of CD180 MFI values between Strasbourg and Lyon could be explained by lack

of standardization between both cytometers. In both centers CLL and MCL B cells displayed a low CD180 staining, always weaker than in those of MZL, whatever their mutational status.

To confirm the flow cytometric protein expression data, total RNA was extracted from highly purified B cells. CD180 RQ-PCR was performed on CD19+ and CD19- fractions of separated blood lymphocytes of 10 CLL, 6 MZL and 6 MCL patient samples, and 5 normal control individuals. As expected, the CD180 transcript was

Table 1. Median of CD180 expression in chronic B-cell lymphoproliferative diseases

	MFI	RFI
<i>Strasbourg series</i>		
CLL ($n = 30$)	906 (111–3226) ± 794.4	8.429 (2.8–35.5) ± 7.6
MCL ($n = 16$)	1028 (128–4124) ± 1084	8.610 (2.18–28.32) ± 8.8
LPL ($n = 15$)	1247 (316–3475) ± 905.5	7.188 (3.764–31.03) ± 10.1
MZL ($n = 30$)	5865 (608–19221) ± 5374	60.79 (7.753–220.7) ± 61.24
<i>Lyon series</i>		
CLL ($n = 80$)	171 (44–352) ± 117	3.4 (1.0–9.1) ± 2.2
MCL ($n = 10$)	126 (78–286) ± 80	2.4 (1.5–3.5) ± 0.7
MZL ($n = 25$)	537 (295–1197) ± 303	8.3 (4.2–18.7) ± 4.9
SDRPL ($n = 8$)	2 356 (1 145–5 920) ± 1 509	35.9 (16.7–103.9) ± 27.2

Abbreviations: CLL, chronic lymphocytic leukemia; LPL, lymphoplasmacytic lymphoma; MCL, mantle cell lymphoma; MFI, mean fluorescence intensity; MZL, marginal zone lymphoma; RFI, relative fluorescence intensity (over negative controls); SMZL, splenic MZL; SDRPL, splenic diffuse red pulp B-cell lymphoma with villous lymphocytes. Median, (range) and s.d. of CD180 expression. CD180 immunolabelling was performed with the phycoerythrin-conjugated RP105 (clone G28.8, BD Biosciences, Franklin Lakes, NJ, USA) and CD19 (clone 5J25C1, Per-CP, BD Biosciences) using standard procedures. An IgG1 isotype control was also tested for each patient, and the whole blood samples were analyzed within 24 h.

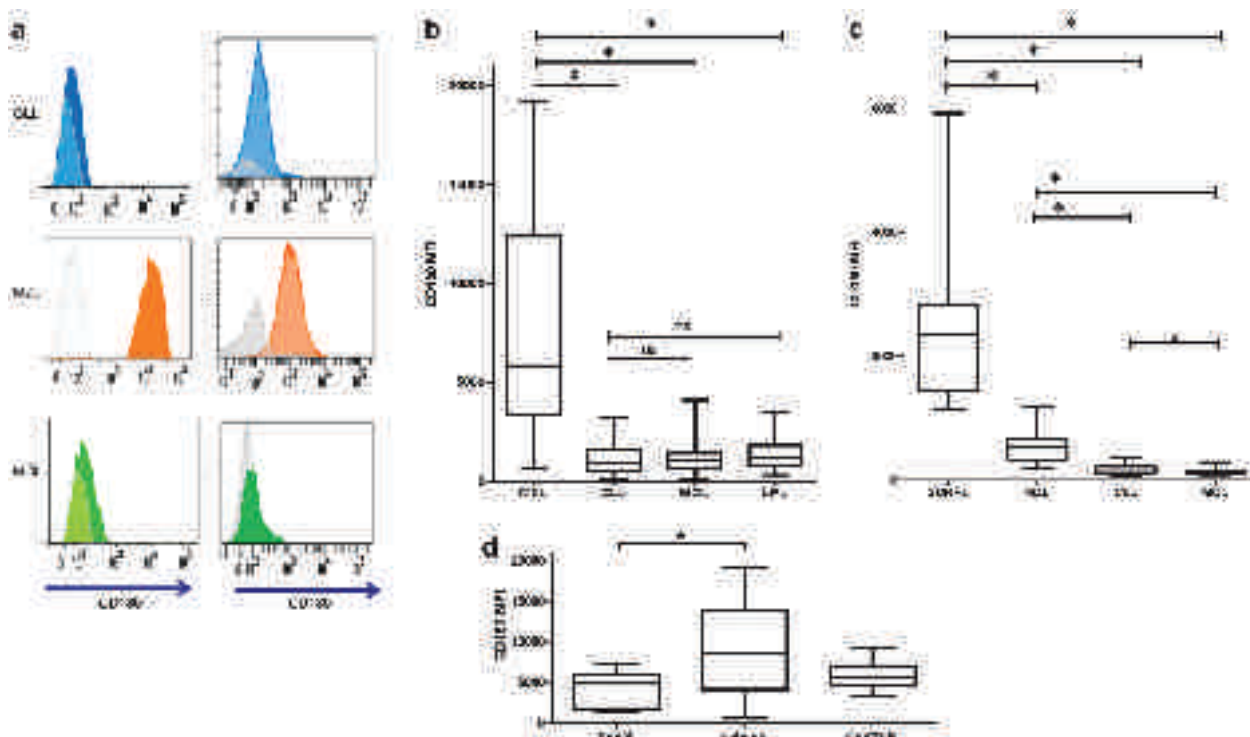


Figure 1. Fluorescence intensity of CD180 in normal and malignant B cells. (a) Representative histograms of CD180 staining of malignant lymphocytes. Right panel: Strasbourg series; left panel: Lyon series. (b–c) Fluorescence intensity of CD180 in malignant B cells. The results were calculated by the FACSDiva software (Becton-Dickinson, Franklin Lakes, NJ, USA) and are expressed as the MFI. The box plot represents the median, 25/75 percentiles and extreme values. * $P < 0.001$; ns, nonsignificant. (b) Strasbourg series, (c) Lyon series. (d) MFI of splenic and nodal MZL of the Strasbourg series.

detected in normal B cells. It was very weakly detected in the CD19[−] lymphocyte fraction, which primarily contains T cells that do not express CD180 at their surface. The CD180 transcript was also detected in all B-LPD tested, and was expressed at variable but comparable levels to the CD19⁺ lymphocytes of normal controls. Accordingly, CD180 expression was also observed in 56 samples of normal CD19⁺ B cells and B-CLP using microarray analyses (public NCI GEO dataset GSE6691¹⁰).

To test whether variant transcripts may explain the lower expression of CD180 at the cell surface, RT-PCR products covering the entire transcript were sequenced in four CLL and three MZL-purified CD19⁺ cells. According to whole genome sequencing studies of B-LPD,^{11,12} no somatic mutations to the coding sequence that was published previously (Genbank NM_005582.2) were observed. Only the rs169714 polymorphism (667A>G, Ile223Val) was detected in both CLL and MZL samples (data not shown). These results suggest that the lower expression of CD180 in CLL is not associated with the presence of an abnormal transcript.

To test the intracytoplasmic localization of the CD180 protein in B-LPD, FITC- and a PE-conjugated anti-CD180 antibodies were used alternatively for membrane and intracellular staining in CLL ($n=5$), MCL ($n=5$), MZL ($n=5$) and control blood samples ($n=5$). The two combinations revealed comparable intracytoplasmic staining of CD180 in all samples tested, which was not observed in control T cells, whereas the surface staining remained unchanged (data not shown). These findings suggest that the CD180 transcript and protein are present in normal and malignant B cells at similar levels. The area under the curve and receiver operating characteristic curve analysis were used to verify the accuracy of CD180 expression in MZL vs the other B-LPD (Supplementary Figure 1). It appeared that the CD180 MFI or RFI could positively discriminate MZL from CLL, LPL and MCL with more than 86% sensitivity and specificity. These results were obtained in two independent sets of patients, tested in two different laboratories.

To the best of our knowledge, no other available positive marker can be used for this type of characterization. Also, a high CD180 staining may be in favor of a splenic origin of the lymphoma, as SMZL and SDRPL display particularly high levels of CD180 expression. These results require however further confirmation in a larger cooperative study to assess its diagnostic potential.

In conclusion, this study shows the differential expression of CD180 in circulating cells of B-LPD, with high expression restricted to MZL and SDRPL B cells.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by research grants from the Institut National du Cancer (INCa PAIR) Lymphome 2008–2011, the Association pour la Recherche en Hématologie Biologique (ARHEB), the Ligue Régionale Contre le Cancer (Comité du Bas-Rhin), the Fondation pour la Recherche Médicale (FRM), the CNRS, the University of Strasbourg Uds, the Région Alsace and the GIS IBISA. SL was supported by a studentship from the French Ministère de la Recherche. This work is dedicated to the memory of Dr Estelle Zink.

L Miguet^{1,2,8}, S Lennon^{3,8}, L Baseggio⁴, A Traverse-Glehen⁵, F Berger⁵, N Perrusson², M-P Chenard⁶, A-C Galois¹, A Eischen¹, C Mayeur-Rousse¹, A Maar¹, L Fornecker^{3,8}, R Herbrecht⁷, P Felman⁴, A Van Dorsseleer³, C Carapito³, S Cianférani³ and L Mauvieux^{1,2}

¹Laboratoire d'Hématologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France;

²Laboratoire d'Hématologie cellulaire, EA 3430, Institut d'Hématologie et d'Immunologie, Faculté de Médecine de Strasbourg, Fédération de Médecine Translationnelle de Strasbourg, Université de Strasbourg, Strasbourg, France;

³Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Université de Strasbourg, IPHC, Strasbourg, France. CNRS, UMR7178, Strasbourg, France;

⁴Laboratoire d'Hématologie Cellulaire, Centre Hospitalier Lyon-Sud/Hospices Civils de Lyon, Université Claude Bernard Lyon 1, Strasbourg, France;

⁵Service d'Anatomie-pathologique, Centre Hospitalier Lyon-Sud/Hospices Civils de Lyon, Université Claude Bernard Lyon 1, Pierre-Bénite, France;

⁶Service d'Anatomie Pathologique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France and

⁷Département d'Oncologie et d'Hématologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

E-mail: laurent.mauvieux@chru-strasbourg.fr

⁸These authors contributed equally to this work.

REFERENCES

- Baseggio L, Traverse-Glehen A, Petinataud F, Callet-Bauchu E, Berger F, Ffrench M *et al*. CD5 expression identifies a subset of splenic marginal zone lymphomas with higher lymphocytosis: a clinico-pathological, cytogenetic and molecular study of 24 cases. *Haematologica* 2010; **95**: 604–612.
- Swerdlow SH, Campo E, Harris N, Jaffe ES, Pileri SA, Stein H *et al*. *World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues*. IARC, 2008.
- Traverse-Glehen A, Baseggio L, Bauchu EC, Morel D, Gazzo S, Ffrench M *et al*. Splenic red pulp lymphoma with numerous basophilic villous lymphocytes: a distinct clinicopathologic and molecular entity? *Blood* 2008; **111**: 2253–2260.
- Miguet L, Pacaud K, Felden C, Hugel B, Martinez MC, Freyssinet J-M *et al*. Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. *Proteomics* 2006; **6**: 153–171.
- Miguet L, Béchade G, Fornecker L, Zink E, Felden C, Gervais C *et al*. Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res* 2009; **8**: 3346–3354.
- Miura Y, Miyake K, Yamashita Y, Shimazu R, Copeland NG, Gilbert DJ *et al*. Molecular cloning of a human RP105 homologue and chromosomal localization of the mouse and human RP105 genes (Ly64 and LY64). *Genomics* 1996; **38**: 299–304.
- Miura Y, Shimazu R, Miyake K, Akashi S, Ogata H, Yamashita Y *et al*. RP105 is associated with MD-1 and transmits an activation signal in human B cells. *Blood* 1998; **92**: 2815–2822.
- Chaplin JW, Kasahara S, Clark EA, Ledbetter JA. Anti-CD180 (RP105) activates B cells to rapidly produce polyclonal Ig via a T cell and MyD88-independent pathway. *J Immunol* 2011; **187**: 4199–4209.
- Porakishvili N, Kulikova N, Jewell AP, Youinou PY, Yong K, Nathwani A *et al*. Differential expression of CD 180 and IgM by B-cell chronic lymphocytic leukaemia cells using mutated and unmutated immunoglobulin VH genes. *Br J Haematol* 2005; **131**: 313–319.
- Gutiérrez NC, Ocio EM, de Las Rivas J, Maiso P, Delgado M, Fermiñán E *et al*. Gene expression profiling of B lymphocytes and plasma cells from Waldenström's macroglobulinemia: comparison with expression patterns of the same cell counterparts from chronic lymphocytic leukemia, multiple myeloma and normal individuals. *Leukemia* 2007; **21**: 541–549.
- Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigel HG, Chiang MY *et al*. Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. *J Exp Med* 2012; **209**: 1553–1565.
- Rossi D, Trifonov V, Fangazio M, Brusca A, Rasi S, Spina V *et al*. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *J Exp Med* 2012; **209**: 1537–1551.

Supplementary Information accompanies the paper on the Leukemia website (<http://www.nature.com/leu>)

Annexe 4

Communications par affiches

- 1) **Strong cell surface expression of the Toll-like receptor homolog CD180 identifies circulating cells of marginal zone lymphoma from other B-cell malignancies.**
L. Miguet, L. Baseggio, S. Lennon, P. Felman, L. Fornecker, A.C. Galois, C. Mayeur-Rousse, R. Herbrecht, A. Eischen, S. Cianferani, L. Mauvieux.
54th Meeting of the American Society of Hematology (Atlanta, USA, 8-11 décembre 2012)
- 2) **Expression du CD180, homologue des récepteurs Toll-like (TLR): marqueur immunologique des cellules circulantes de lymphomes de la zone marginale.**
L. Miguet, S. Lennon, L. Baseggio, A.C. Galois, A. Eischen, C. Mayeur-Rousse, A. Maar, L. Fornecker, R. Herbrecht, P. Felman, A. Van Dorsselaer, C. Carapito, S. Cianferani, L. Mauvieux
33^e Congrès annuel de la Société Française d'Hématologie (Paris, 27-29 mars 2013)
- 3) **Les microparticules : un nouvel outil pour proposer par approche protéomique de nouveaux biomarqueurs pour le diagnostic des hémopathies lymphoïdes.**
L. Fornecker, S. Lennon, L. Miguet, C. Carapito, A. Van Dorsselaer, L. Mauvieux, S. Cianferani.
7^{ème} forum du Cancéropôle Grand-Est (Strasbourg, 25-26 novembre 2013)

4) **Intérêt de l'analyse du triple marquage CD148, CD180 et CD200 par cytométrie en flux dans le diagnostic des hémopathies lymphoïdes B chroniques.**

L. Miguet, C. Mayeur Rouse, L. Fornecker, V. Latger-Cannard, J.F. Leseve, R. Herbrecht, S. Cianferani, L. Mauvieux.

35^e Congrès annuel de la Société Française d'Hématologie (Paris, 01-03 avril 2015)

5) **Multicentric analysis of the CD148, CD180 and CD200 combination for the diagnosis of mature B-cell neoplasms using flow cytometry.**

L. Miguet, L. Fornecker, M. Wyrwas, S. Cianferani, R. Herbrecht, J.F. Leseve, V. Latger-Cannard, C. Mayeur-Rousse, L. Mauvieux.

57th Meeting of the American Society of Hematology (Orlando, USA, 5-8 décembre 2015)



Strong Cell Surface Expression of the Toll-Like Receptor Homolog CD180 Identifies Circulating Cells of Marginal Zone Lymphoma From Other B-Cell Malignancies

Laurent Miguet^{1,6}, Lucile Baseggio², Sarah Lennon³, Pascale Felman², Luc Fornecker⁴, Anne-Cecile Galois¹, Caroline Mayeur-Rousse¹, Raoul Herbrecht⁴, Alice Eischen¹, Sarah Cianferani³ and Laurent Mauvieux^{1,6}

¹Laboratoire d'hématologie, CHU Hautepierre, Stasbourg, France; ²Laboratoire d'Hématologie Cellulaire, Centre Hospitalier Lyon-Sud, Pierre-Bénite, France; ³Laboratoire de Spectrométrie de Masse BioOrganique, IPHC UMR 7178, Strasbourg, France; ⁴Oncologie et Hématologie, CHU Hautepierre, Strasbourg, France; ⁵Medical School of Strasbourg, Strasbourg, France



Introduction

Distinction of MZL from other mature B-cell lymphomas is hampered by the lack of specific cellular markers. Proteomic analyses on plasma membrane microparticles¹ derived from malignant cells identified the Toll-like receptor homolog CD180, as a candidate marker of MZL.

CD180 expression was analyzed in 96 patients and 25 controls by flow cytometry including marginal zone lymphomas, chronic lymphocytic leukemia, cell lymphoma, and lymphoplasmacytic lymphoma. CD180 was observed to be under-expressed at the cell surface in all B-cell malignancy subtypes except MZL B cells. These results were confirmed in an independent series of 123 patients.

Furthermore, ligation of this receptor, induced a greater increase of CD86 expression at the cell surface of MZL B, suggesting an activation of the NF- κ B pathway.

These results strongly suggest that CD180 may be the first positive immunological marker for MZL.

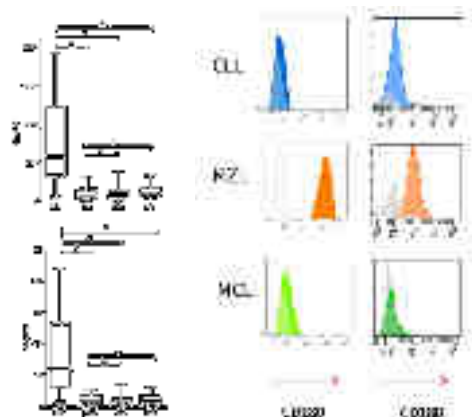


Figure1
Fluorescence intensity of CD180 in normal and malignant B-cells and Representative histograms of CD180 staining of malignant lymphocytes from both series

Materials & Methods

Patients and samples

25 controls, 30 patients with CLL, 16 patients with MCL, 15 patients with LPL and 35 patients with MZL (25 splenic, 10 nodal) were studied. The diagnosis was established following WHO guidelines.

In the same way, the CD180 expression was performed in an independent series of 123 consecutive and prospective patients with B-LPD diagnosed in the Hematology Department of Lyon-Sud University Hospital. These cases were classified as CLL (n=80), MZL (n=25), MCL (n=10), SDRPL (n=8).

Immunophenotypic analyses using CD180

CD180 immunolabelling was performed with the RP105 (clone G28.8. Blood samples were analyzed within 24 hours on a FACSCanto II flow cytometer (BD Biosciences). The mean fluorescence intensity (MFI) and relative fluorescence intensity (RFI, defined as the ratio of specific fluorescence over non-specific fluorescence) were measured.

	MFI	RFI
CLL (n=30)	906 +/- 795	8.43 +/- 7.6
MCL (n=16)	1028 +/- 1084	8.6 +/- 8.9
LPL (n=15)	1247 +/- 905	7.2 +/- 10
MZL (n=30)	5865 +/- 5374	61 +/- 61
sMZL (n=25)	8708 +/- 5706	89.8 +/- 59.5
Nodal MZL (n=10)	4941 +/- 2133	46.3 +/- 26.3

Table1
Median and standard deviation of CD180 expression. MFI: mean fluorescence intensity, RFI: relative fluorescence intensity (over negative controls),

Measurement of activation of CD19+ cells

CD180 ligation was realized similarly as described by Porakishvili et al². Briefly, 1 x 10⁶ leukocytes were stimulated for 48 hours at 37° C in 5% CO₂ with an anti-CD180 mAb (BD) final concentrations of 10 μ g/ml. Cells were stained with CD19 and CD86.

Results

CD180 expression in normal and malignant B-cells from the blood

results are summarized in table1 and figure1. Normal B-cells displayed strong staining for CD180. Malignant B-cells of CLL displayed weaker expression of CD180. CD180 staining was also weak in MCL and LPL B cells.

CD180 MFIs of CLL, MCL and LPL B cells did not differ significantly and were all significantly lower than the controls.

By contrast, circulating MZL cells had a high CD180 MFI, which was not different from the controls and was significantly higher than the CLL, MCL and LPL B cells. Similar results were obtained using a second anti-CD180 clone.

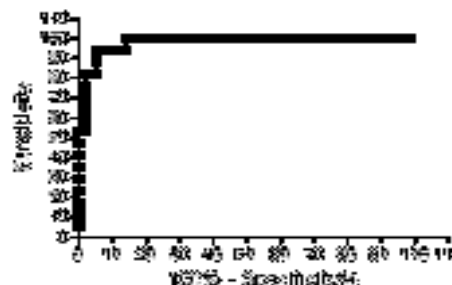


Figure2
ROC curves of sensitivity versus specificity of CD180 expression in B-LPD. The maximum sensitivity and specificity of 86.8% and 87.9%, respectively, was obtained for an MFI> 2122 threshold

The area under the curve and receiver operating characteristic curve analysis (figure2) was used to verify the accuracy of CD180 expression in MZL versus the other B-LPD.

A MFI> 2122 could discriminate MZL from the other B-LPD with 86.7% sensitivity and 87.9% specificity. For an RFI > 21, the results were 90% sensitivity and 86% specificity.

These results were confirmed with the Lyon series.

CD180 cross-linking

To test whether CD180 signaling is correlated with its cell surface expression, we studied the cell surface expression of CD86 following stimulation with an anti-CD180 mAb in 4 CLL, 4 MZL and 6 control samples (figure 3). Increased expression of CD86 after CD180 ligation has been previously shown in murine and human B-cells. It requires NF- κ B activation and is a hallmark of B-cell activation.

CD180 ligation induced a strong response in the MZL B-cells, as indicated by CD86 staining (MFI 5348 +/- 1483), and was higher than the normal B-cells (1550 +/- 259.9, p= 0,0087). No significant increase of the CD86 MFI was observed in the CLL B cells after CD180 ligation (135.5 +/- 239.4 vs 150 +/- 64.39, p= 0,4857).

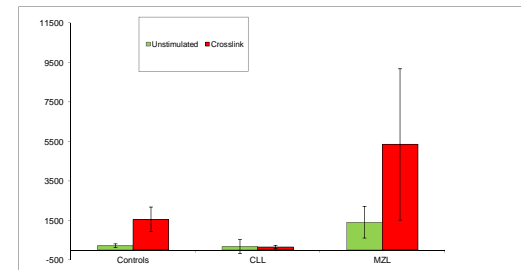


Figure3
CD86 expression following CD180 cross-linking stimulation. Normal, CLL and MZL blood samples were cultured 48 hours after CD180 cross-linking stimulation, and analyzed for the expression of the activation marker CD86.

Conclusion

CD180 MFI or RFI could positively discriminate MZL from CLL, LPL and MCL with more than 86% sensitivity and specificity. These results were obtained in 2 independent sets of patients, tested in 2 different laboratories.

Furthermore, we did not observe any variation of CD180 intensity in patients during the following two years.

In conclusion, this study shows the differential expression of CD180 in circulating cells of BLPD, with high expression in MZL that is associated with strong activation signals.

References

- Miguet et al. Journal of proteome research 2009 Jul;8(7):3346-54
- Porakishvili et al british journal of haematology 2011 May;153(4):486-98

Expression du CD180, homologue des récepteurs Toll-like (TLR): marqueur immunologique des cellules circulantes de lymphomes de la zone marginale



Laurent Miguet¹, Sarah Lennon², Lucile Baseggio³, Anne-Cécile Galois¹, Alice Eischen¹, Caroline Mayeur-Rousse¹, Aïda Maar¹, Luc Fornecker⁴, Raoul Herbrecht⁴, Pascale Felman³, Alain Van Dorselaer², Christine Carapito², Sarah Cianferani², Laurent Mauvieux¹



¹Laboratoire d'Hématologie, Hôpitaux Universitaires de Strasbourg, EA3430, France

²Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Université de Strasbourg, IPHC, CNRS, UMR7178, Strasbourg, France

³Laboratoire d'Hématologie Cellulaire, Centre Hospitalier Lyon-Sud/Hospices Civils de Lyon, Université Claude Bernard Lyon 1

⁴Département d'Oncologie et d'Hématologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

Introduction

Le diagnostic différentiel entre lymphomes de la zone marginale (LZM) et autres SLPB est difficile devant le manque de marqueurs biologiques spécifiques. Afin de répondre à cette question, nous avons développé une approche protéomique d'étude de microparticules membranaires issues de patients^{1,2}. Cette approche nous a permis de mettre en évidence une protéine homologue aux récepteurs Toll, le CD180, comme nouveau candidat spécifique des Lymphomes de la zone marginale (LZM). Ces résultats ont été confirmés par une deuxième étude indépendante.

	MFI	RFI
CLL (n=20)	421 ± 752	8.4 ± 7.6
MCL (n=14)	1228 ± 1384	11.5 ± 8.5
LLP (n=15)	1247 ± 960	7.2 ± 10
MZL (n=30)	5805 ± 2274	51 ± 61
MZL (n=20)	1724 ± 122	16.4 ± 11.1
MZL (n=10)	491 ± 27	17.4 ± 3.3

Tableau1

Moyenne et écart type de l'expression du CD180. MFI: moyenne des intensités, RFI: Ratio des intensités (spécifique/non spécifique).

Matériel et Méthode

Patients

25 contrôles, 30 LLC, 16 MCL, 15 LLP et 35 MZL (25 spléniques, 10 nodaux) ont été étudiés. Le diagnostic a été établi conformément aux recommandations de la WHO.

En parallèle, une deuxième étude indépendante a été menée par le laboratoire d'hématologie de l'hôpital de Lyon sud sur 123 patients (80 LLC; 25 MZL; 10 MCL et 8 SDRPL).

Marquage CD180

Le marquage CD180 a été réalisé avec le clone G28.8 marqué PE. Les échantillons ont été analysés dans les 24 heures après prélèvement sur un cytomètre BD FACSCanto II (BD Biosciences). La moyenne de fluorescence (MFI) ainsi que le ratio des intensités de fluorescence (RFI, fluorescence spécifique divisée par la fluorescence non spécifique) ont été calculés (BD FACSDiva software).

Résultats

Expression du CD180 dans les lymphocytes B contrôles et pathologiques (tableau 1 et Fig. 1).

- Les lymphocytes B normaux circulants expriment fortement le CD180 et les lymphocytes B pathologiques l'expriment très faiblement (LLC; MCL; LLP). Aucune différence n'a pu être observée entre les intensités de marquage du CD180 dans les LLC les MCL et les LLP, et toutes ces intensités étaient significativement inférieures à celle des contrôles.
- Par opposition, les lymphocytes B circulants des LZM expriment très fortement le CD180. Les MFI ainsi que les RFI obtenus dans les MZL ne sont pas significativement différents des valeurs obtenues dans le groupe contrôle, mais supérieurs aux groupes LLC, MCL et LLP. Des résultats similaires ont été obtenus en utilisant le deuxième clone anti-CD180.
- Une courbe de ROC (figure2) de l'expression du CD180 des LZM par rapport aux autres SLPB montre une sensibilité et une spécificité de discrimination des LZM respectivement de 87% et 88% avec une MFI>2122 (équivalent à une RFI>21).

Activation du CD180 par ligation

Afin de déterminer si l'expression du CD180 était corrélée à son niveau d'expression, nous avons étudié l'apparition du CD86 (stimulation de la voie NF-κB) à la surface des cellules après stimulation par ligation d'un anticorps anti-CD180 (figure 3). L'activation du CD180 entraîne une forte réponse des cellules issues de LZM (CD86 MFI 5348 +/- 1483) supérieure aux cellules contrôles (CD86 MFI 1550 +/- 259.9, p= 0,0087). Des tests réalisés sur des cellules de LLC ne montrent pas d'augmentation de l'expression du CD86 après activation (CD86 MFI 135.5 +/- 239.4 vs 150 +/- 64.39, p= 0,4857).

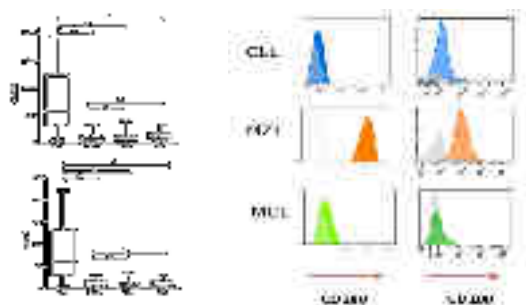


Figure1
MFI du CD180. Représentation des deux séries (Lyon et Strasbourg)

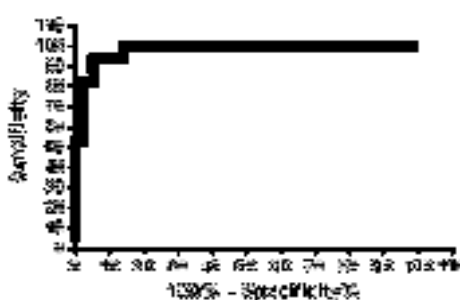


Figure2
Courbe de ROC du CD180 dans les SLPB. Spécificité et sensibilité maximum avec une MFI> 2122

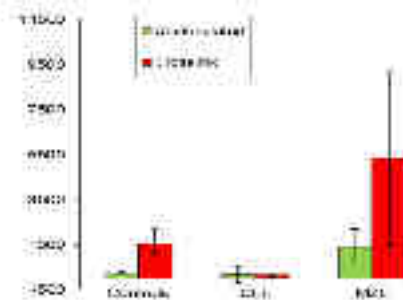


Figure3
Expression du CD86 après stimulation.

Conclusions

La MFI ou la RFI du CD180 peut permettre de discriminer les LZM des autres hémopathies lymphoïdes B chroniques avec plus de 86% de sensibilité et de spécificité. Ces résultats ont pu être reproduits dans deux études indépendantes. Aucune variation de l'expression du CD180 n'a été observée sur une période de 2 ans. En conclusion, cette étude montre la forte expression du CD180 dans les MZL par rapport aux autres syndromes lymphoprolifératifs B.

(1) Miguet, L. et al. Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res* 8, 3346-3354 (2009)

(2) Miguet L. et al. Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas. *Leukemia* 2013 Jan 10

LES MICROPARTICULES : UN NOUVEL OUTIL POUR PROPOSER PAR APPROCHE PROTEOMIQUE DE NOUVEAUX BIOMARQUEURS POUR LE DIAGNOSTIC DES HEMOPATHIES LYMPHOÏDES

L. FORNECKER¹, S. LENNON¹, L. MIGUET², C. CARAPITO¹, A. VAN DORSSELAER¹, L. MAUVIEUX², S. CIANFERANI¹

¹Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Institut pluridisciplinaire Hubert Curien (IPHC), Département des Sciences Analytiques – CNRS, UMR 7178. ²Laboratoire d'Hématologie, EA3430, Hôpitaux Universitaires de Strasbourg

INTRODUCTION

- La caractérisation des syndromes lymphoprolifératifs selon la classification OMS 2008 repose sur une combinaison d'analyses (cytométrie, immunophénotypage, immunohistochimie, cytogénétique, biologie moléculaire)
- Certains cas restent cependant mal catégorisés en l'absence de biomarqueurs spécifiques à chaque entité. L'hétérogénéité entre les différentes pathologies ou au sein d'une même entité rend nécessaire l'identification de nouveaux biomarqueurs en vue d'adapter idéalement la stratégie thérapeutique.
- Les protéines membranaires constituent une cible majeure pour cette recherche de nouveaux biomarqueurs en vue d'une utilisation ultérieure en routine par cytométrie en flux.
- L'approche protéomique par spectrométrie de masse permet d'identifier un grand nombre de protéines contenues dans un échantillon et ainsi de comparer le protéome d'échantillons différents provenant de pathologies distinctes. Cette approche se heurte cependant aux difficultés liées à l'identification de protéines peu abondantes dans des échantillons complexes.

Les microparticules (MPs)

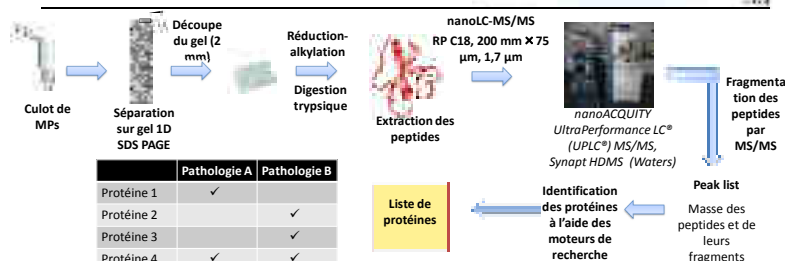
- Structures membranaires d'une taille comprise entre 50 nm et 1µm
- Obtenues après stimulation de cellules en cultures par de l'actinomycine D (MPs induites).
- Générées spontanément et physiologiquement par différents types cellulaires, notamment les cellules sanguines, et être isolées à partir du plasma (MPs plasmatiques).
- La production de MPs plasmatique est augmentée dans certaines situations pathologiques dont les processus néoplasiques.

	MP induites	MP plasmatiques
Avantages	<ul style="list-style-type: none"> Obtenues directement à partir des cellules tumorales des patients 	<ul style="list-style-type: none"> « facilité » d'obtention à partir du plasma Pas de culture cellulaire nécessaire Peuvent être obtenues en l'absence de cellules tumorales circulantes dans le sang
Inconvénients	<ul style="list-style-type: none"> Nécessité de collecter des cellules d'intérêt (patients présentant une hyperlymphocytose) Mise en culture des cellules Multiples centrifugations nécessaires pour isoler les MPs du milieu de culture 	<ul style="list-style-type: none"> Majoritairement d'origine plaquettaire Faible quantité de MPs issues de cellules tumorales

Approche protéomique par spectrométrie de masse

ETAPE 1 : - Identification des protéines exprimées par des cellules pathologiques.
- Comparaison avec le protéome de témoins et/ou le protéome d'autres pathologies.

ETAPE 2 : - Sélection du ou des candidats
- Validation à grande échelle.
- Analyse statistique de la force du marqueur (spécificité et sensibilité)



Applications des MPs induites à la caractérisation des syndromes lymphoprolifératifs : identification du CD148 surexprimé dans les lymphomes à cellules du manteau et du CD180 comme marqueur des lymphomes de la zone marginale

- Le CD148 (HPTPeta) est un récepteur membranaire à activité tyrosine phosphatase. Son ligand ainsi que sa fonction exacte restent inconnus.
- Le CD180 est une protéine membranaire appartenant à la famille des TLR (Toll-Like Receptor), exprimée par les cellules présentatrices d'antigènes

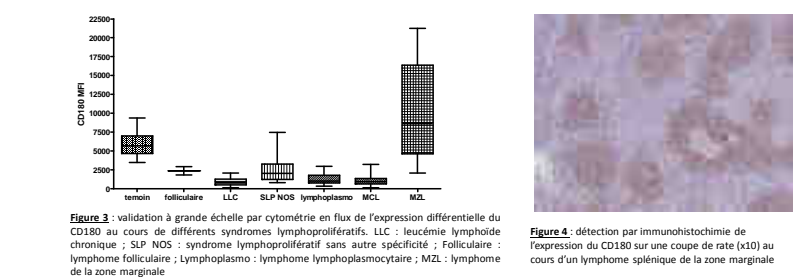
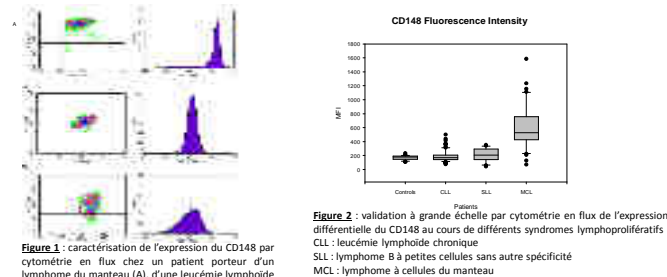
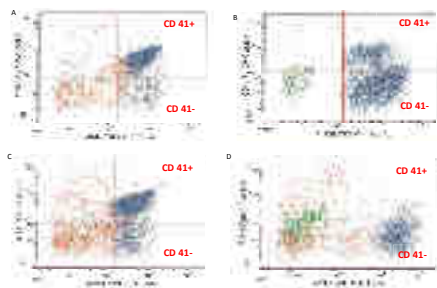


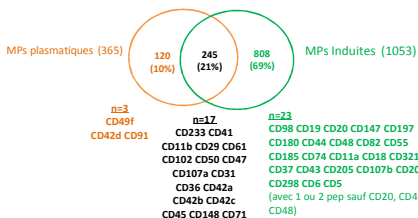
Figure 4 : détection par immunohistochimie de l'expression du CD180 sur une coupe de rate (x10) au cours d'un lymphome splénique de la zone marginale

Des MPs induites aux MPs plasmatiques

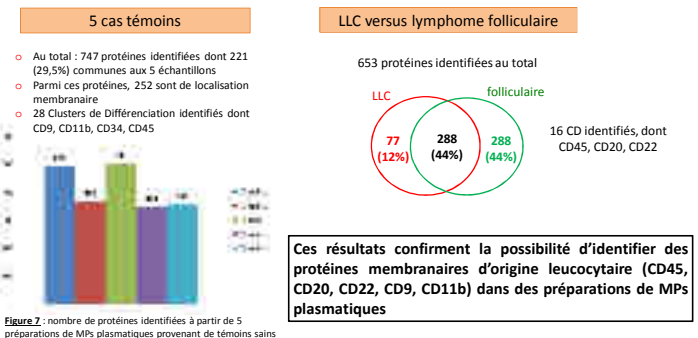
- Comparaison des MPs induites et plasmatiques par cytométrie en flux



- Comparaison des MPs induites et plasmatiques par analyse nanoLC-MS/MS



Microparticules plasmatiques : analyses de cas



CONCLUSION

- Les MPs représentent un sous-protéome enrichi en protéines membranaires
- Approche translationnelle validée pour l'identification de nouveaux biomarqueurs dans les syndromes lymphoprolifératifs à partir de MPs induites : utilisation en routine par cytométrie en flux du CD148 et du CD180 pour les phénotypes des hémopathies. Validation inter-régionale en cours (étude MALYZOMA).
- Optimisation de cette approche en cours à partir des MPs plasmatiques pour élargir le champ des applications pour la recherche de biomarqueurs en l'absence de cellules tumorales circulantes

Intérêt de l'analyse du triple marquage CD148, CD180 et CD200 par cytométrie en flux dans le diagnostic des hémopathies lymphoïdes B chroniques

Laurent Miguet^{1, 2}; Caroline Mayeur-Rousse¹; Luc Fornecker^{3, 4}; Jean-Francois Lesesve⁵; Véronique Latger-Cannard⁵; Raoul Herbrecht⁴; Sarah Sanglier³ et Laurent Mauvieux^{1, 2}



- 1) Laboratoire d'hématologie des hôpitaux universitaires de Strasbourg ;
- 2) Laboratoire d'hématologie cellulaire EA 3430 faculté de médecine de Strasbourg, fédération de médecine translationnelle de Strasbourg (FMTS) ;
- 3) Laboratoire de spectrométrie de masse bio-organique (LSMBO) CNRS UMR 7178 institut pluridisciplinaire Hubert Curien Strasbourg ;
- 4) Département d'oncohématologie hôpitaux universitaires de Strasbourg ;
- 5) Service d'Hématologie Biologique, CHU Nancy, avenue du Morvan, 54511 Vandoeuvre-les-Nancy

Introduction

Les hémopathies lymphoïdes B chroniques sont des affections en constante augmentation, et bien que le diagnostic soit facilement établi chez la majorité des patients, il subsiste un certain nombre de cas difficiles, surtout en cas de phase leucémique, pour lesquels nous ne disposons pas de marqueurs spécifiques. C'est particulièrement vrai dans certains cas de LLC atypiques, de lymphomes de la zone marginale (LZM), à cellules du manteau (LCM), notamment ceux CCND1 négatifs, ou lympho-plasmocytaires (LLP). Nous avons développé une stratégie basée sur l'analyse protéomique des microparticules issues de cellules lymphomateuses qui nous a permis de proposer deux nouveaux marqueurs: le CD148 et le CD180, associés aux LCM et des LZM respectivement. Nous avons étudié l'apport de ces deux marqueurs, couplés avec le CD200 pour le diagnostic des hémopathies lymphoïdes B chroniques.

	LLC	LCM	LZM	LLP
CD148	<690	>980	<1000	540 < MFI < 1350
CD180	<1150	<1500	>2700	500 < MFI < 1450
CD200	>500	<900	<2200	0 < MFI < 2700
Spécificité	95%	92%	99%	90%
Sensibilité	79%	64%	60%	54%

Tableau1: Seuils de fluorescence des marqueurs CD148, CD180 et CD200 dans les groupes de LLC, LZM LCM et LLP. L'application de ces seuils permet une classification des pathologies avec une spécificité et une sensibilité reportées ci-dessus.

Patients et Méthode

L'analyse de ces trois marqueurs a été réalisée par cytométrie en flux (CD148 (Clone 143-41 FITC); CD180 (Clone G28.8 PE) et CD200 (Clone OX104 APC)). Leur profil d'expression a été établi sur une cohorte de 114 patients bien caractérisés : LLC avec un score de Matutes > 3 (N=34); LCM caractérisés par une translocation t(11;14) et/ou une sur-expression de CCND1 (N=22); LLP (N=26) présentant une morphologie évocatrice, un pic IgM et une positivité du CD38 ; des LZM (N=32) avec un phénotype CD5- CD23- associé à une splénomégalie. Pour chaque groupe la moyenne des MFI et l'erreur standard ont été déterminées.

Résultats

L'analyse par cytométrie en flux des LCM révèle une forte expression du CD148 (MFI moyenne = 1498) associée à une faible expression des CD180 et CD200 (MFI moyenne = 888 et 426 respectivement). A l'inverse, une faible expression du CD148 et du CD180 (MFI moyenne = 495 et 754) associée à une forte expression du CD200 (MFI moyenne = 3750) est caractéristique des LLC. Une faible expression du CD148 et du CD200 (MFI moyenne = 640 et 1200) associée à une très forte expression du CD180 (MFI moyenne = 5300) est le profil observé dans les LZM et enfin, une expression modérée de ces trois marqueurs représente le groupe des LLP (figures 1 et 2). Un seuil correspondant à la moyenne des MFI +/- 4 SD a été calculé pour chacun des groupes. L'utilisation de ces seuils permet de classer correctement les LCM avec une spécificité de 96% et une sensibilité de 65%, les LLC avec une spécificité de 95% et une sensibilité de 79%, les LLP avec une spécificité de 95% et une sensibilité de 54 % et enfin les LZM avec une spécificité de 99% et une sensibilité de 60 % (tableau 1).

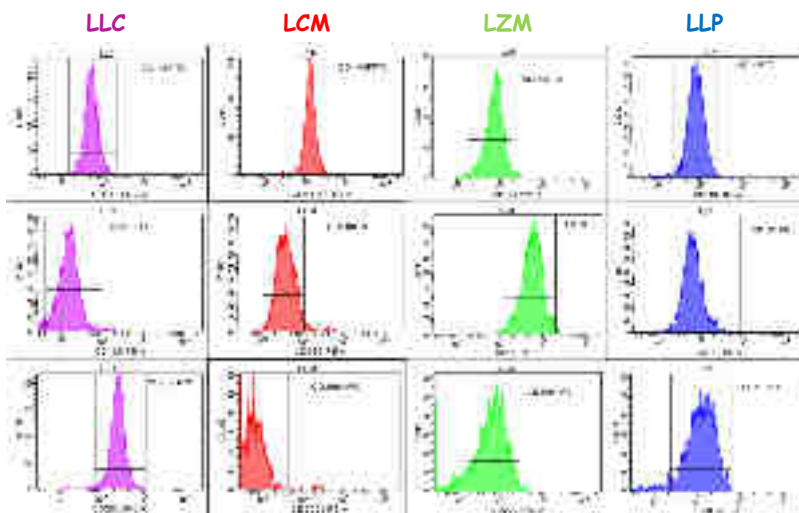


Figure1: Exemples de profils d'expression des marqueurs CD148, CD180 et CD200 dans les LLC (violet) LCM (rouge) LZM (vert) et LLP (bleu).

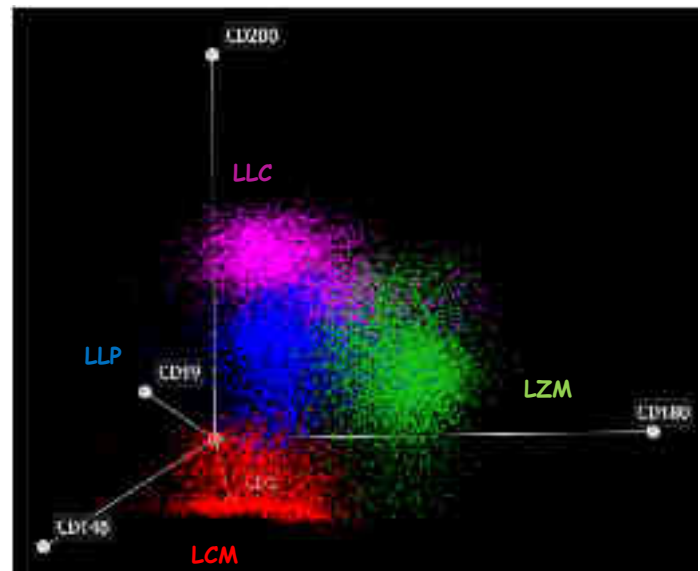


Figure2: Représentation « radar » sur le logiciel Kaluzia Analyses 1.3 de la combinaison de 10 LLC, 10 LLP, 10 LCM et 10 LZM pour l'expression des marqueurs CD148, CD180 et CD200.

Conclusions

Les seuils d'expression de ces marqueurs CD148, CD180 et CD200 peuvent être utilisés pour classer avec une très bonne précision les syndromes lymphoprolifératifs B (90-99% de spécificité). Associée aux marqueurs utilisés habituellement en routine cette combinaison d'anticorps peut être particulièrement utile dans les cas où la cytologie n'est pas spécifique, dans les cas où l'histologie n'est pas disponible ou ne peut être réalisée.

(1) Miguet, L. et al. Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res* 8, 3346-3354 (2009)

(2) Miguet L. et al. Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas. *Leukemia* 2013 Jan 10

Multicentric Analyses of the CD148, CD180, and CD200 Combination for the Diagnosis of Mature B-Cell Neoplasms Using Flow Cytometry

Laurent Miguet^{1,2,3}, Luc Fornecker⁴, Marie Wyrwas⁵, Sarah Cianferani⁶, Raoul Herbrecht⁴, Jean-Francois Lesesve⁵,
Véronique Latger-Cannard⁵, Caroline Mayeur-Rousse¹ and Laurent Mauvieux^{1,2,3}

¹Hôpital Hautepierre, Laboratoire d'hématologie, Strasbourg, France; ²Université de Strasbourg, EA3430, Strasbourg, France; ³FMTS, Strasbourg, France; ⁴Service d'Oncologie et d'Hématologie, Hôpital de Hautepierre, Strasbourg, France; ⁵Hôpital de Nancy, France; ⁶Laboratoire de spectrométrie de masse bioorganique (LSMBO) CNRS UMR 7178 institut pluridisciplinaire Hubert Curien Strasbourg, Strasbourg, France



ABSTRACT

Diagnosis of mature B-cells proliferations, do not always fulfill any of the WHO criteria of B-cell neoplasms using standard diagnosis tools. This situation is notably encountered in the case of the differential diagnosis of marginal zone lymphoma (MZL), atypical chronic lymphocytic leukemia (aCLL), mantle cell lymphoma (MCL), or lymphoplasmacytic lymphoma (LPL), mostly due to the lack of immunological positive markers.

In order to find new markers to discriminate between these different malignancies, we have previously developed a proteomic strategy based on the analyses of plasma membrane microparticles and proposed two new specific markers: CD148¹ and CD180^{2,3} for MCL and MZL respectively. The simultaneous use of these two markers, together with the CD200 that is positive in most cases of CLL and negative in MCL could be of great interest to better assess the differential diagnosis

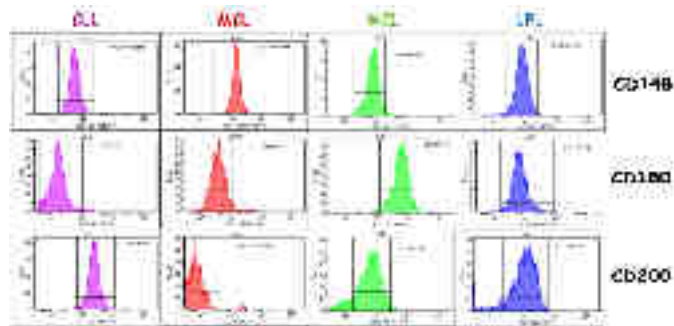


Figure 1
Fluorescence intensity of CD148, CD180 and CD200 in representative cases of CLL, MCL, MZL and LPL using BD FACSDiva® Software

REFERENCES

1. Miguet et al. Journal of proteome research 2009 Jul;8(7):3346-54
2. Miguet et al. Leukemia. 2013 Aug;27(8):1748-50
3. Mayeur-Rousse et al. Cytometry B Clin Cytom. 2015 Sep 18.

METHODS

Patients and samples

Blood or bone marrow samples of a well characterized set of patients (N=287) : CLL with a Matutes score > 3 (N=81); MCL harboring t(11;14) translocation or CCND1 overexpression (N=44); LPL (N=58) classified following cytological morphology, IgM peak and the positivity of CD38 and/or Myd88 mutation, MZL (N=84), displaying a CD5- CD23-immunophenotype associated to a splenomegaly and 20 controls.

Immunophenotypic analyses

Flow cytometry analyses have been realized in Nancy and Strasbourg hospitals by combining these three markers: CD148 (Clone 143-41 FITC, R&D system); CD180 (Clone G28.8 PE BD Biosciences) and CD200 (Clone OX104 APC, E Bioscience). For each group the Median of Fluorescence Intensity and Standard Error have been determined

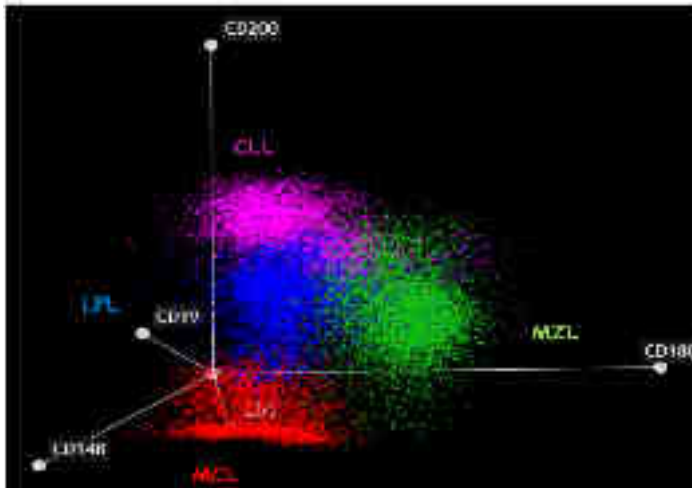


Figure 2
Radar representation of CD148, CD180 and CD200 expression levels on the merge of 10 CLL (purple), 10 LPL (blue), 10MCL (red) and 10MZL (green) (Kaluzza Analyses 1.3 ©software).

RESULTS

Discriminant MFI values of CD148, CD180 and CD200 obtained in the different sets of patients are summarized in table 1. MCL exhibited a strong expression of CD148 combined with a weak expression of CD180 and CD200. A weak expression of CD148 and CD180 coupled to a strong expression of CD200 was typical of the CLL group and a weak expression of CD148 and CD200 coupled to a strong expression of CD180 was observed in the MZL group. A moderate expression of these three markers was observed in the LPL group.

A threshold corresponding to MFI +/- 4 standard error was then calculated for each group, and patients were categorized following the expression profile of these 3 markers (see figures 1 and 2).

In this cohort, the above described profiles correctly identified MCL cases with a specificity of 92% and a sensitivity of 64%, aCLL cases with a specificity of 100% and a sensitivity of 47%, LPL cases with a specificity of 90% and a sensitivity of 54% and MZL cases with a specificity of 99% and a sensitivity of 60%.

CONCLUSIONS

These results strongly suggest that the incorporation of these three markers CD148 CD180 and CD200 in addition of the routinely used flow cytometry panel can be helpful in a number of cases for which the diagnosis is difficult

	CLL	MCL	MZL	LPL
CD 148 MFI	<690	>980	<1000	540<CD148 MFI<1350
CD 180 MFI	<1150	<1500	>2700	500<CD180 MFI<1450
CD 200MFI	>500	<900	<2200	<2700
Specificity	100%	92%	99%	90%
Sensitivity	47%	64%	60%	54%

Table 1
Discriminant fluorescence intensity threshold of CD148, CD180 and CD200 in CBCL

Annexe 5

Liste des transitions sélectionnées pour le suivi des protéines CD45,
CD148, CD5, CD200_1, CD200_3, CD180 et CD23

Les peptides marqués sont indiqués par le signe *

Protéine	Séquence peptidique	m/z précurseur	Temps de rétention (min)	m/z fragments
CD45	DETVDDFWR	591.757	37.16	938.4 738.3 837.4
	DETVDDFWR*	596.761	37.16	748.3 847.4 948.4
	DPPSEPSPLEAEFQR	849.902	32.59	175.1 1173.6 1486.7
	DPPSEPSPLEAEFQR*	854.906	32.59	185.1 1183.6 1496.7
	IEFDNEVDK	554.761	17.19	719.3 995.4 866.4
	IEFDNEVDK*	558.769	17.19	727.3 874.4 1003.4
	LFLAEFQSIPR	660.869	65	272.2 747.4 1060.6
	LFLAEFQSIPR*	665.873	65	282.2 1070.6 757.4
	LNPTPGSNAISDVPGER	862.434	19.86	458.2 1298.6 557.3
	LNPTPGSNAISDVPGER*	867.438	19.86	468.2 567.3 1308.6
	NSNVIPYDYNR	677.823	19.4	452.2 940.5 827.4
	NSNVIPYDYNR*	682.827	19.4	837.4 950.5 462.2
	VDVYGYVVK	521.284	29.46	565.3 827.5 573.3

	VDVYGYVVK*	525.291	29.46	736.4 835.5 728.4
	YHLEVEAGNTLVR	500.932	25.92	602.4 730.4 659.4
	YHLEVEAGNTLVR*	504.268	25.92	740.4 612.4 669.4
	TLILDVPPGVEK	640.877	42.87	840.4 626.4 953.5
	TLILDVPPGVEK*	644.884	42.87	961.5 634.4 848.5
	YVDILPYDYNR	715.851	44.06	452.2 940.5 827.4
	YVDILPYDYNR*	720.855	44.06	462.2 837.4 950.5
CD148	DTEVLLVGLPGTR	749.909	47.05	729.4 828.5 430.2
	DTEVLLVGLPGTR*	754.913	47.05	440.2 739.4 838.5
	FSGFEASHGPIK	638.82	19.21	1042.5 1129.6 838.4
	FSGFEASHGPIK*	642.827	19.21	846.5 1050.5 1137.6
	GLIDGAESYVSFSR	750.87	42	974.5 496.3 845.4
	GLIDGAESYVSFSR*	755.874	42	855.4 984.5 506.3
	SQSLSEVLK	495.777	20.66	147.1 575.3 775.5
	SQSLSEVLK*	499.784	20.66	155.1 583.4 783.5

	TGTFIAIDR	497.272	24.36	474.3 734.4 587.4
	TGTFIAIDR*	502.276	24.36	744.4 484.3 597.4
	VITEPIPVSDLR	674.889	39.58	1136.6 696.4 906.5
	VITEPIPVSDLR*	669.885	39.58	686.4 896.5 1126.6
	YAAELAENR	518.757	11.29	602.3 802.4 489.2
	YAAELAENR*	523.761	11.29	499.3 612.3 812.4
	YNNVLPYDISR	677.344	31.67	1076.6 750.4 863.5
	YNNVLPYDISR*	682.348	31.67	760.4 1086.6 873.5
CD5	HLPETEAGR	505.257	6.31	872.4 533.3 759.4
	HLPETEAGR*	510.261	6.31	543.3 769.4 882.5
	EHQPLPIQWK	638.346	29.5	1009.6 881.5 671.4
	EHQPLPIQWK*	642.353	29.5	889.5 679.4 1017.6
	LSWYDPDFQAR	699.328	47.08	733.4 848.4 1197.5
	LSWYDPDFQAR*	704.332	47.08	743.4 858.4 1207.5
	LVGSSIC[+57.0]EGTVEVR	781.896	20.5	660.4 949.4 1350.6

	LVGSSIC[+57.0]EGTVEVR*	786.9	20.5	1360.6 959.4 670.4
	NSHLSAYPALEGALHR	868.447	28.41	1284.7 1197.6 963.5
	NSHLSAYPALEGALHR*	873.451	28.41	973.5 1294.7 1207.6
	QWEDPSQASK	588.27	10.08	732.4 861.4 617.3
	QWEDPSQASK*	592.277	10.08	869.4 740.4 625.3
	VLALLC[+57.0]SGFQPK	666.871	50	823.4 936.5 1120.6
	VLALLC[+57.0]SGFQPK*	670.878	50	831.4 944.5 1128.6
	VLDAGDPTSR	515.762	9.2	460.3 703.3 818.4
	VLDAGDPTSR*	520.766	9.2	470.3 713.3 828.4
CD200	EQLYTPASLK	575.311	20.64	515.3 616.4 779.4
	EQLYTPASLK*	579.318	20.64	787.4 523.3 624.4
CD200_1	GELSQGVQK	473.254	6.48	646.4 759.4 431.3
	GELSQGVQK*	477.261	6.48	439.3 654.4 767.5
CD200_3	TTINDYQVIR	611.825	20.28	678.4 793.4 907.5
	TTINDYQVIR*	616.829	20.28	688.4 803.4 917.5

CD180	NLSHLQTLNLSHNEPLGLQSQAFK	1345.206	49	580.3 878.5 1088.6
	NLSHLQTLNLSHNEPLGLQSQAFK*	1349.214	49	588.3 886.5 1096.6
	GNHFQDGTITK	609.299	9.45	1046.5 909.5 248.2
	GNHFQDGTITK*	613.306	9.45	256.2 917.5 1054.5
	SLEQAINLSLNFNGNNVK	988.016	60	1106.6 906.4 1019.5
	SLEQAINLSLNFNGNNVK*	992.023	60	914.5 1027.5 1114.6
	VLDFQNNAIHYISR	845.439	42.27	675.4 538.3 1087.6
	VLDFQNNAIHYISR*	850.443	42.27	685.4 548.3 1097.6
	CD23	HASHTGSWIGLR	661.342	19.38
HASHTGSWIGLR*		666.346	19.38	798.5 1123.6 1194.6
LPTPSAPLHS		510.28	17.51	611.3 708.4 809.4
LPTPSAPLHS*		512.283	17.51	712.4 813.4 615.3
NEASDLLER		523.759	18.41	803.4 645.4 732.4
NEASDLLER*		528.763	18.41	655.4 742.4 813.4
SQSTQISQEEELR		824.413	42.47	788.4 546.3 1003.5

	SQSTQISQELEELR*	829.417	42.47	798.4
				556.3
				1013.5
	WNDAFC[+57.0]DR	542.219	17.99	450.2
				668.3
				783.3
	WNDAFC[+57.0]DR*	547.223	17.99	460.2
				607.3
				793.3

Annexe 6

Communication par affiche

Evaluation de la préparation d'échantillons « tube-gel » pour l'analyse quantitative d'échantillons complexes.

L. Muller, LM Fornecker, C. Carapito, S. Cianferani

Congrès de Spectrométrie de Masse et d'Analyse Protéomique (SMAP)

(Ajaccio, 15-18 septembre 2015)

Evaluation de la préparation d'échantillons « Tube-gel » pour l'analyse quantitative de mélanges protéiques complexes

Leslie MULLER, Luc-Matthieu FORNECKER, Christine CARAPITO, Sarah CIANFERANI

Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC-DSA, UMR 7178 CNRS-Uds, 67087 Strasbourg, FRANCE

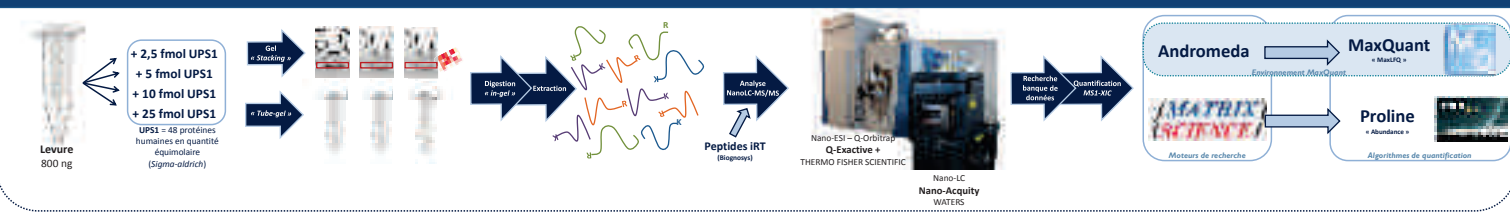
Introduction

En protéomique, la confiance dans l'identification des protéines est dépendante du nombre de peptides dérivant de chacune des protéines, elle-même particulièrement tributaire de l'efficacité de la digestion et par extension de la préparation d'échantillon. De plus, les analyses quantitatives requièrent une préparation d'échantillon la plus répétable possible, donc la plus simple possible, d'autant plus en amont d'approches quantitatives de type « label-free ».

Actuellement, ces stratégies de protéomique quantitative « label-free » s'effectuent via une digestion « in-solution », ou via une digestion « in-gel » des protéines. La digestion « in-gel » présente l'avantage d'être compatible avec une extraction des protéines en présence de détergents (tel que le SDS), composés qui ne sont pas compatibles avec la spectrométrie de masse (SM), mais également le désavantage d'être souvent chronophage.

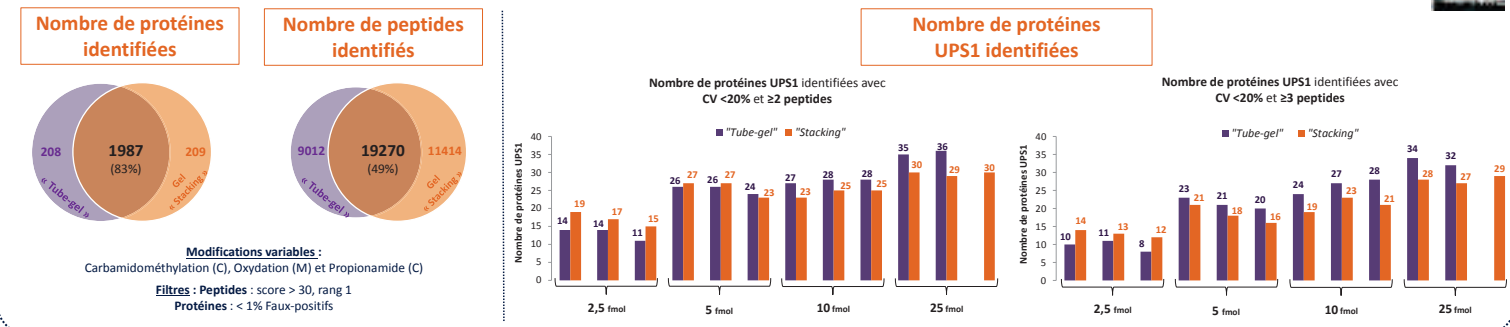
Dans cette optique de gain de temps et de compatibilité avec la SM, nous avons évalué la possibilité d'une **préparation par « Tube-gel »**¹, une technique qui consiste à faire polymériser un gel de polyacrylamide directement avec l'extrait protéique en solution, sans électrophorèse, en la comparant au gel « Stacking » classique.

Schéma analytique

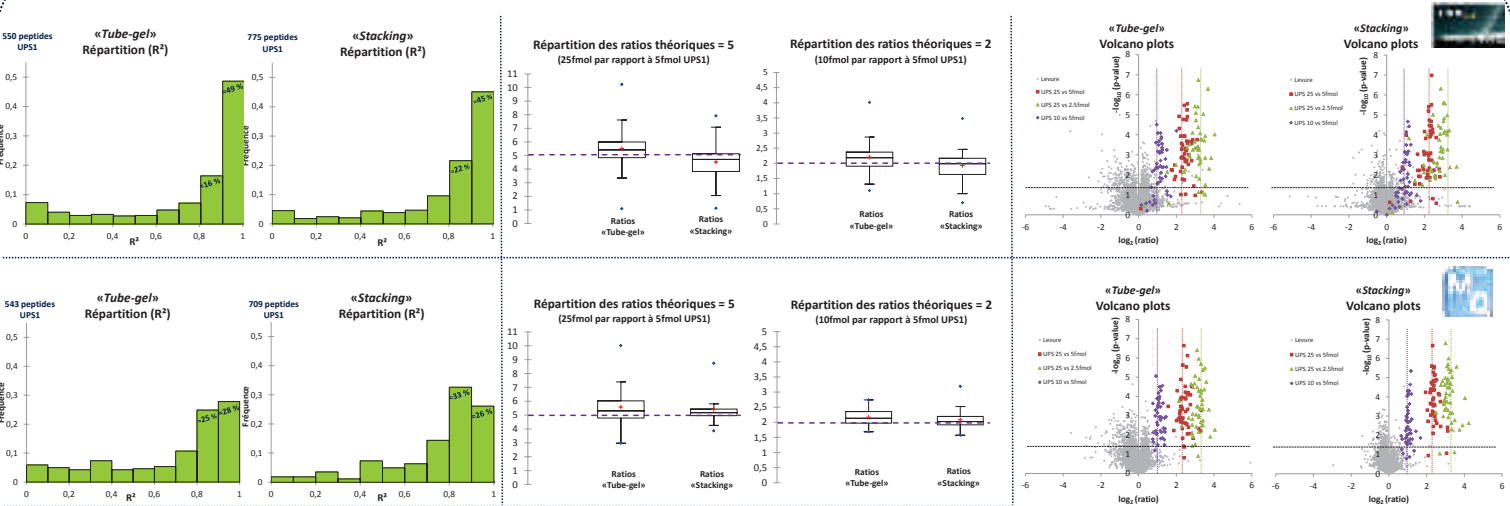


Résultats

Comparaisons Qualitatives



Comparaisons Quantitatives



	MaxQuant			Proline		
	« Tube-gel »	« Stacking »	« Tube-gel »	« Stacking »	« Tube-gel »	« Stacking »
Ratio 2 (10 vs 5)	73%	55%	60%	33%	0%	0%
Ratio 5 (25 vs 5)	98%	96%	90%	90%	10%	10%
Ratio 10 (25 vs 2.5)	98%	96%	90%	91%	4%	4%
Proportion de Faux-positifs	0%	4%	0%	0%	16%	23%

p-value < 0,05 et Ratio ≥ 2

Sensibilité
= Nombre de protéines UPS différentielles/Nombre de protéines UPS détectées

Proportion de Faux-positifs
= Nombre de Faux-positifs / Nombre total de protéines différentielles

Conclusion

La préparation d'échantillons « Tube-gel » :

- est un procédé **efficace** pour obtenir un aperçu de sa composition (>80% de protéines communes identifiées entre la méthode « Tube-gel » et la méthode « Stacking »),
- semble être un bon compromis entre **rapidité** de préparation et **répétabilité** pour les approches de protéomique quantitative de type « label-free »,
- permet une **quantification précise** et **sensible** de protéines faiblement abondantes au sein d'un échantillon complexe, avec des résultats équivalents à ceux d'une préparation de type gel « Stacking ».

¹ Lu, X. & Zhu, H. Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. Mol. Cell. Proteomics MCP 4, 1948-1958 (2005)

Annexe 7

Liste des protéines avec un test de Welch significatif entre les groupes « réfractaires » et « sensibles » de lymphomes B diffus à grandes cellules

Protéines surexprimées dans le groupe « réfractaire »

Proteine	p	Différence
Lipopolysaccharide-binding protein	0,006	4,3
Fibrinogen beta chain	0,025	4,0
Fibrinogen gamma chain	0,026	3,8
Fibrinogen alpha chain	0,037	3,6
Vitronectin	0,043	3,4
Hexokinase-3	0,015	2,7
C4b-binding protein alpha chain	0,009	2,7
Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1	0,011	2,5
Alpha-2-antiplasmin	0,045	2,5
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	0,003	2,4
Long-chain-fatty-acid--CoA ligase 1	0,014	2,4
Complement component C9	0,016	2,3
Erythrocyte band 7 integral membrane protein	0,033	2,2
Dipeptidyl peptidase 1	0,027	2,2
Interferon-induced guanylate-binding protein 2	0,045	2,1
Tensin-3	0,005	2,0
Heme-binding protein 1	0,014	2,0
Sepiapterin reductase	0,026	2,0
N-acetylneuraminatase lyase	0,019	1,9
Cathepsin B	0,009	1,9
Serum amyloid P-component	0,047	1,8
Procollagen galactosyltransferase 1	0,004	1,8
Cathepsin Z	0,007	1,8
Estradiol 17-beta-dehydrogenase 11	0,032	1,8
Neutral cholesterol ester hydrolase 1	0,002	1,8
V-type proton ATPase 116 kDa subunit a isoform 3	0,011	1,8
Inorganic pyrophosphatase	0,004	1,8
Unconventional myosin-1e	0,011	1,8
Glycerol-3-phosphate dehydrogenase, mitochondrial	0,013	1,8
Sulfide:quinone oxidoreductase, mitochondrial	0,038	1,8
Cytoplasmic FMR1-interacting protein 1	0,044	1,8
Interleukin-18	0,023	1,7
N-acylglucosamine 2-epimerase	0,048	1,7
Vesicle transport protein GOT1B	0,007	1,7
AMP deaminase 3	0,047	1,7
ERO1-like protein alpha	0,043	1,7
Thiopurine S-methyltransferase	0,010	1,7
EH domain-binding protein 1-like protein 1	0,028	1,6
Sodium/hydrogen exchanger 9	0,050	1,6
GDH/6PGL endoplasmic bifunctional protein	0,027	1,6
EF-hand domain-containing protein D2	0,014	1,6

Ganglioside GM2 activator	0,026	1,6
Pleckstrin homology domain-containing family O member 2	0,023	1,6
Leukotriene A-4 hydrolase	0,048	1,6
Reticulon-4	0,005	1,6
Monoacylglycerol lipase ABHD12	0,003	1,6
Epididymal secretory protein E1	0,015	1,6
Adipocyte plasma membrane-associated protein	0,027	1,5
PRA1 family protein 3	0,034	1,5
Cytoplasmic dynein 1 light intermediate chain 2	0,001	1,5
Pleckstrin	0,031	1,5
Apoptosis-associated speck-like protein containing a CARD	0,049	1,5
ADP-dependent glucokinase	0,014	1,5
V-type proton ATPase subunit d 1	0,022	1,5
Gamma-interferon-inducible lysosomal thiol reductase	0,050	1,5
Phosphatidate cytidyltransferase 2	0,007	1,5
Atlastin-3	0,012	1,5
Peroxisomal multifunctional enzyme type 2	0,005	1,5
Protein transport protein Sec24A	0,033	1,5
NADH-cytochrome b5 reductase 1	0,023	1,5
V-type proton ATPase subunit B, brain isoform	0,032	1,5
Ubiquitin-like modifier-activating enzyme ATG7	0,010	1,5
Twinfilin-1	0,021	1,5
Cytosolic acyl coenzyme A thioester hydrolase	0,043	1,5
V-type proton ATPase subunit F	0,044	1,5
Protein disulfide-isomerase A3	0,029	1,5
V-type proton ATPase subunit C 1	0,008	1,5
Acid ceramidase	0,020	1,4
Large neutral amino acids transporter small subunit 1	0,031	1,4
Hexokinase-1	0,035	1,4
V-type proton ATPase catalytic subunit A	0,036	1,4
Ras-related protein Rab-18	0,008	1,4
Phosphatidylinositol-binding clathrin assembly protein	0,038	1,4
Nicastrin	0,029	1,4
ER membrane protein complex subunit 1	0,034	1,4
Acetolactate synthase-like protein	0,022	1,4
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	0,017	1,4
Manganese-transporting ATPase 13A1	0,019	1,4
Calnexin	0,033	1,4
V-type proton ATPase subunit H	0,050	1,4
Long-chain-fatty-acid--CoA ligase 4	0,036	1,4
Proteasome-associated protein ECM29 homolog	0,021	1,4
GTP-binding protein SAR1a	0,018	1,3
Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	0,036	1,3
Unconventional myosin-IXb	0,002	1,3
UDP-glucose:glycoprotein glucosyltransferase 1	0,007	1,3
Signal recognition particle subunit SRP72	0,005	1,3

Signal recognition particle subunit SRP68	0,012	1,3
Cathepsin S	0,019	1,3
Constitutive coactivator of PPAR-gamma-like protein 1	0,040	1,3
ADP-ribosylation factor-like protein 8B	0,027	1,3
Ras GTPase-activating-like protein IQGAP1	0,021	1,3
Malectin	0,050	1,3
AP-3 complex subunit delta-1	0,001	1,2
Mannosyl-oligosaccharide glucosidase	0,029	1,2
Ras-related protein Rab-10	0,049	1,2
Protein transport protein Sec24B	0,034	1,2
Ras-related protein Rab-1A	0,036	1,2
Coiled-coil domain-containing protein 47	0,007	1,2
Vacuolar protein sorting-associated protein 35	0,006	1,2
Lysosome-associated membrane glycoprotein 1	0,031	1,2
Cytoplasmic dynein 1 intermediate chain 2	0,016	1,2
Guanine nucleotide-binding protein G(k) subunit alpha	0,008	1,2
Helicase SKI2W	0,049	1,2
AP-3 complex subunit beta-1	0,029	1,1
AP-3 complex subunit mu-1	0,024	1,1

Protéines surexprimées dans le groupe « sensible »

Protéine	p	Différence
Interferon-induced GTP-binding protein Mx2	0,041	-2,3
B-lymphocyte antigen CD20	0,019	-1,7
LIM domain and actin-binding protein 1	0,045	-1,6
Transgelin	0,008	-1,4
Lymphoid-restricted membrane protein	0,045	-1,4
Sorcin	0,042	-1,2
Phosphoribosyl pyrophosphate synthase-associated protein 2	0,029	-1,0
WD repeat- and FYVE domain-containing protein 4	0,026	-1,0
Rab GTPase-activating protein 1-like	0,005	-0,9
Calponin-2	0,012	-0,9
Aspartate aminotransferase, mitochondrial	0,047	-0,8
Protein TFG	0,018	-0,8
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1	0,037	-0,8
Glyoxalase domain-containing protein 4	0,007	-0,8
GRIP1-associated protein 1	0,034	-0,8
Deoxycytidine kinase	0,042	-0,7
Leucine-rich repeat and calponin homology domain-containing protein 3	0,008	-0,7
Haloacid dehalogenase-like hydrolase domain-containing protein 2	0,047	-0,7
Tyrosine-protein kinase SYK	0,003	-0,7
Cullin-4B	0,008	-0,7
High mobility group protein B2	0,034	-0,7
Synaptogyrin-2	0,021	-0,6
Rab3 GTPase-activating protein non-catalytic subunit	0,001	-0,6
Serine/threonine-protein kinase D2	0,030	-0,6
Lamina-associated polypeptide 2, isoform alpha	0,033	-0,6
Calcium-binding mitochondrial carrier protein Aralar2	0,018	-0,6
Spermine synthase	0,024	-0,6
Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1	0,008	-0,6
Protein FAM49B	0,031	-0,6
Cleavage stimulation factor subunit 2	0,038	-0,6
Lamina-associated polypeptide 2, isoforms beta/gamma	0,001	-0,6
Serine/threonine-protein phosphatase 5	0,045	-0,6
Ig kappa chain C region	0,022	-0,6
Lanosterol synthase	0,039	-0,5
Acidic leucine-rich nuclear phosphoprotein 32 family member E	0,029	-0,5
Palladin	0,042	-0,5
Myotrophin	0,022	-0,5
Heterogeneous nuclear ribonucleoprotein D0	0,016	-0,5
Phosphoribosylformylglycinamide synthase	0,046	-0,5
Translin-associated protein X	0,041	-0,5
Eukaryotic translation initiation factor 5A-1	0,022	-0,5
Testis-expressed sequence 10 protein	0,050	-0,5
DAZ-associated protein 1	0,020	-0,5
Polypyrimidine tract-binding protein 3	0,008	-0,5

Ubiquilin-2	0,030	-0,5
SUMO-conjugating enzyme UBC9	0,025	-0,4
Acyl-protein thioesterase 1	0,048	-0,4
14-3-3 protein epsilon	0,017	-0,4
Nucleosome assembly protein 1-like 4	0,015	-0,4
Transcription factor BTF3	0,031	-0,4
Protein-L-isoaspartate(D-aspartate) O-methyltransferase	0,004	-0,4
NEDD8-conjugating enzyme Ubc12	0,045	-0,4
Eukaryotic translation initiation factor 6	0,042	-0,4
Double-stranded RNA-specific adenosine deaminase	0,025	-0,4
Methionine--tRNA ligase, cytoplasmic	0,023	-0,4
60 kDa SS-A/Ro ribonucleoprotein	0,047	-0,4
Ras-related C3 botulinum toxin substrate 2	0,005	-0,4
Dr1-associated corepressor	0,048	-0,4
Small nuclear ribonucleoprotein Sm D3	0,031	-0,4
Eukaryotic translation elongation factor 1 epsilon-1	0,032	-0,4
BRO1 domain-containing protein BROX	0,023	-0,4
Adenylosuccinate lyase	0,038	-0,4
Proteasome subunit beta type-4	0,025	-0,4
Mitogen-activated protein kinase kinase kinase 4	0,040	-0,4
Methyl-CpG-binding domain protein 2	0,037	-0,3
Ubiquitin-like modifier-activating enzyme 5	0,031	-0,3
Thioredoxin domain-containing protein 17	0,044	-0,3
Putative oxidoreductase GLYR1	0,020	-0,3
Nucleolysin TIAR	0,013	-0,3
Small nuclear ribonucleoprotein-associated proteins B and B	0,045	-0,3
Microtubule-associated protein RP/EB family member 1	0,049	-0,3
C-terminal-binding protein 1	0,021	-0,3
Host cell factor 1	0,035	-0,3
Splicing factor 1	0,026	-0,3
ATP synthase subunit g, mitochondrial	0,021	-0,3
Serine--tRNA ligase, cytoplasmic	0,012	-0,3
60S ribosomal protein L38	0,011	-0,3
60S ribosomal protein L23	0,036	-0,3
Eukaryotic translation initiation factor 3 subunit F	0,021	-0,2
Eukaryotic translation initiation factor 3 subunit M	0,044	-0,2
Eukaryotic translation initiation factor 3 subunit E	0,041	-0,2



Luc-Matthieu FORNECKER



**Développements méthodologiques en analyse
protéomique pour la découverte et la validation
de biomarqueurs dans les hémopathies
lymphoïdes B de l'adulte**

Le développement actuel très rapide des différentes approches « omiques » génère un grand nombre de biomarqueurs potentiels, en particulier dans le domaine de la cancérologie. L'analyse protéomique par spectrométrie de masse a particulièrement bénéficié des progrès technologiques récents qui ont permis la mise au point de nouvelles approches quantitatives globales ou ciblées. Néanmoins, peu de biomarqueurs potentiels finissent par être concrètement utilisés en pratique clinique, nécessitant l'optimisation des différentes étapes de leur développement.

Dans la continuité des travaux ayant abouti à l'identification de biomarqueurs diagnostiques dans les hémopathies lymphoïdes B, ce travail de thèse a permis le développement d'une méthode d'analyse protéomique ciblée pour la vérification et la validation de nouveaux biomarqueurs potentiels. La possibilité d'appliquer des stratégies quantitatives globales à un très grand nombre d'échantillons a pu être démontrée. L'application de ces stratégies quantitatives globales à du tissu ganglionnaire provenant de lymphomes agressifs a permis d'identifier des biomarqueurs potentiellement associés à une résistance au traitement. Enfin, le mode de préparation tube-gel, facilitant l'étude d'un grand nombre d'échantillons, a été validé pour la réalisation d'analyses protéomiques différentielles.

Mots-clés : protéomique quantitative, spectrométrie de masse, biomarqueurs, hémopathies lymphoïdes B, tube-gel

The current development of new « omics » technologies has led to the discovery of a large number of potential biomarkers, particularly in the field of oncology. Proteomics analysis by mass spectrometry have particularly benefited from these technological advances with the development of new global or targeted quantitative approaches. Nevertheless, only a few numbers of potential biomarkers are finally used in clinical practice, requiring further optimization of the development process.

Following the initial identification of biomarkers in the diagnosis of lymphoid malignancies performed previously, this thesis has allowed the development of a targeted proteomics method that can be used for the validation of new potential biomarkers. The ability of analysing a large number of samples with a global quantitative approach has also been demonstrated. The application of these global quantitative strategies on lymph node tissue of aggressive lymphoma has permitted the identification of potential new biomarkers associated with chemorefractoriness. Lastly, a tube-gel protocol facilitating the analysis of a large number of samples has been validated for differential proteomics studies.

Key words: quantitative proteomics, mass spectrometry, biomarkers, B-cell lymphoma, tube-gel